



การพัฒนาแบบจำลองคณิตศาสตร์และพัฒนาโปรแกรม  
เพื่อการพยากรณ์ค่า PM 2.5  
The Development of Mathematical Model  
and Program for Forecasting PM 2.5

กฤษณ์ เจ็ดวรรณะ  
KRIT JEDWANNA

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิศวกรรมศาสตรดุษฎีบัณฑิต  
สาขาวิชาวิศวกรรมการจัดการอุตสาหกรรมเพื่อความยั่งยืน คณะวิศวกรรมศาสตร์  
มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร

2566



การพัฒนาแบบจำลองคณิตศาสตร์และพัฒนาโปรแกรม  
เพื่อการพยากรณ์ค่า PM 2.5  
The Development of Mathematical Model  
and Program for Forecasting PM 2.5

กฤษณ์ เจ็ดวรรณะ  
KRIT JEDWANNA

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิศวกรรมศาสตรดุษฎีบัณฑิต  
สาขาวิชาวิศวกรรมการจัดการอุตสาหกรรมเพื่อความยั่งยืน คณะวิศวกรรมศาสตร์  
มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร

2566

ลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร

|                  |  |
|------------------|--|
| ชื่อวิทยานิพนธ์  | การพัฒนาแบบจำลองคณิตศาสตร์และพัฒนาโปรแกรมเพื่อการพยากรณ์ค่า PM 2.5     |
| ชื่อ นามสกุล     | กฤษณ์ เจ็ดวรรณะ  |
| ชื่อปริญญา       | วิศวกรรมศาสตรดุษฎีบัณฑิต (วิศวกรรมการจัดการอุตสาหกรรมเพื่อความยั่งยืน) |
| สาขาวิชา         | วิศวกรรมการจัดการอุตสาหกรรมเพื่อความยั่งยืน                            |
| คณะ              | วิศวกรรมศาสตร์   |
| อาจารย์ที่ปรึกษา | ผู้ช่วยศาสตราจารย์ ดร.สุวิทย์ แพงกันยา                                 |

คณะกรรมการสอบวิทยานิพนธ์ได้ให้ความเห็นชอบวิทยานิพนธ์ฉบับนี้แล้ว

..... ประธานกรรมการ  
(ผู้ช่วยศาสตราจารย์ ดร.วรวจน์ มีถม)

..... กรรมการ  
(ดร.ณัฐวรพล รัชสิริวัชรบุล)

..... กรรมการ  
(ผู้ช่วยศาสตราจารย์ ดร.ปริญญา บุญกนิษฐ)

..... กรรมการ  
(ผู้ช่วยศาสตราจารย์ ดร.เทอดพงษ์ แดงสี)

..... กรรมการ  
(ผู้ช่วยศาสตราจารย์ ดร.สุวิทย์ แพงกันยา)

คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร อนุมัติให้บัณฑิตวิทยานิพนธ์ฉบับนี้  
เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร วิศวกรรมศาสตรดุษฎีบัณฑิต  
สาขาวิชาวิศวกรรมการจัดการอุตสาหกรรมเพื่อความยั่งยืน มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร

..... คณบดีคณะวิศวกรรมศาสตร์  
(ผู้ช่วยศาสตราจารย์ ดร.ณัฐพงศ์ พันธนะ)

วันที่ 12 เดือน ธันวาคม พ.ศ. 2566

|                 |  |
|-----------------|--|
| ชื่อวิทยานิพนธ์ | การพัฒนาแบบจำลองคณิตศาสตร์และพัฒนาโปรแกรมเพื่อการพยากรณ์<br>ค่า PM 2.5 |
| ชื่อ นามสกุล    | กฤษณ์ เจ็ดวรรณะ  |
| ชื่อปริญญา      | วิศวกรรมศาสตรดุษฎีบัณฑิต (วิศวกรรมการจัดการอุตสาหกรรมเพื่อความยั่งยืน) |
| สาขาวิชา และคณะ | วิศวกรรมการจัดการอุตสาหกรรมเพื่อความยั่งยืน<br>คณะวิศวกรรมศาสตร์       |
| ปีการศึกษา      | 2566   |

### บทคัดย่อ

การศึกษานี้เป็นการพัฒนาแบบจำลองเพื่อการคาดการณ์ PM 2.5 โดยใช้ Machine Learning Model ทั้งหมด 6 แบบจำลอง ได้แก่ Decision Tree Model, Gradient Boosting Model, K-neighbors Model, MLP Model, Random Forest Model, และ Ridge Model จากชุดข้อมูลของ PM 2.5 จำนวน 2.3 ล้านข้อมูล โดยรวบรวมข้อมูลตั้งแต่เดือนมกราคม พ.ศ. 2558 - ธันวาคม พ.ศ. 2564 จากสถานีตรวจวัดจำนวน 63 สถานี ซึ่งอยู่ในพื้นที่กรุงเทพมหานคร ประเทศไทย ในการพัฒนาแบบจำลองได้เริ่มจากการตรวจสอบความถูกต้องและคัดกรองข้อมูล และนำข้อมูลมาจัดกลุ่มออกเป็น 2 กลุ่ม ได้แก่ ข้อมูลในช่วงก่อนการระบาดของ COVID-19 และข้อมูลในช่วงการแพร่ระบาดของ COVID-19 สำหรับใช้พัฒนาแบบจำลองและทดสอบประสิทธิภาพของแบบจำลอง ทั้งในช่วงสถานการณ์ปกติและสถานการณ์ที่มีการแพร่ระบาดของ COVID-19 และในงานวิจัยนี้ได้นำข้อมูลย้อนหลัง (Historic data) มาใช้ในการพัฒนาแบบจำลอง จากการวิเคราะห์ พบว่า ผลการพัฒนาแบบจำลองโดยใช้ข้อมูลในช่วงก่อนการระบาดของ COVID-19 และใช้ข้อมูลย้อนหลังรายชั่วโมง (Lag by Hour) ประกอบ ให้ผลการวิเคราะห์ที่ดีที่สุดด้วยแบบจำลอง Gradient Boosting Model มีค่า  $R^2 = 0.9533$ ,  $RMSE = 4.1344$  และ  $MAPE =$  ร้อยละ 17.13 ในขณะที่ผลการทดสอบประสิทธิภาพของแบบจำลองด้วยชุดข้อมูลทั้งช่วงก่อนการระบาดของ COVID-19 และในช่วงสถานการณ์ COVID-19 ก็พบว่า Gradient Boosting Model มีผลการทดสอบที่ดีที่สุด โดยมีค่า  $R^2 = 0.8790$ ,  $RMSE = 6.1375$  และ  $MAPE =$  ร้อยละ 20.68 และ  $R^2 = 0.8720$ ,  $RMSE = 5.1344$  และ  $MAPE =$  ร้อยละ 27.18 ตามลำดับ งานวิจัยนี้สรุปได้ว่า Gradient Boosting Model เป็นแบบจำลองที่เหมาะสมต่อการนำมาคาดการณ์ PM 2.5 ทั้งในช่วงสถานการณ์ปกติ และช่วงที่มีการแพร่ระบาดของ COVID-19 อย่างไรก็ตาม ประสิทธิภาพในการคาดการณ์โดยใช้ข้อมูลช่วงสถานการณ์แพร่ระบาดของ COVID-19 จะทำให้ประสิทธิภาพของแบบจำลองลดลงเล็กน้อย สำหรับแบบจำลองที่มีประสิทธิภาพรองลงมา ได้แก่ Random Forest Model, MLP Model, Decision Tree Model, K-neighbors Model, และ Ridge Model ตามลำดับ

**คำสำคัญ :** PM 2.5, COVID-19, แบบจำลอง, ปัญญาประดิษฐ์, Machine Learning

|                      |  |
|----------------------|--|
| <b>Thesis Title</b>  | The Development of Mathematical Model and Program for Forecasting PM 2.5 |
| <b>Author</b>        | Krit Jedwanna  |
| <b>Degree</b>        | Doctor of Engineering (Sustainable Industrial Management Engineering)    |
| <b>Major Program</b> | Sustainable Industrial Management Engineering<br>Faculty of Engineering  |
| <b>Academic Year</b> | 2023   |

## ABSTRACT

The study aimed to develop Machine Learning Models for fine particulate matter (PM 2.5) prediction. Six different machine learning models which included Decision Tree Model, Gradient Boosting Model, K-neighbors Model, MLP Model, Random Forest Model, and Ridge Model have been used in this study, based on 2.3 million data sets of PM 2.5 that were collected from January 2015 to December 2021, from 63 measuring stations in Bangkok, Thailand. The development of the Machine Learning Models started from checking the data accuracy and doing the data screening. Then the screened data were classified into two groups: the data before the COVID-19 pandemic and that after the COVID-19 pandemic. The obtained data were used to develop Machine Learning Models and to test their efficiency in both situations: in the normal situation and during the COVID-19 pandemic. In addition, the concerned historic data have been used to develop the Machine Learning Models in this research study as well. According to the analysis of the development of Machine Learning Models using the database before the COVID-19 pandemic together with the data of Lag by Hour, it was found that Gradient Boosting Model found the best analysis results:  $R^2 = 0.9533$ ,  $RMSE = 4.1344$  and  $MAPE = 17.13\%$ . When testing the efficiency of the learning models with the data sets before and during the COVID-19 pandemic, Gradient Boosting Model showed the best test results:  $R^2 = 0.8790$ ,  $RMSE = 6.1375$ ,  $MAPE = 20.68\%$  and  $R^2 = 0.8720$ ,  $RMSE = 5.1344$ ,  $MAPE = 27.18\%$  respectively. According to this research, it can be concluded that Gradient Boosting Model was appropriate for predicting PM 2.5 both in the normal situation and during the COVID-19 pandemic. However, the prediction efficiency using the database during the COVID-19 pandemic would slightly decrease the efficiency of the machine learning models. The less efficient models were Random Forest Model, MLP Model, Decision Tree Model, K-neighbors Model, and Ridge Model, respectively.

**Keywords :** PM 2.5, COVID-19, Models, AI, Machine Learning

## กิตติกรรมประกาศ

การศึกษาวิจัยเพื่อการจัดทำวิทยานิพนธ์ฉบับนี้ได้สำเร็จเสร็จได้ด้วยความอนุเคราะห์ของผู้มีพระคุณหลายท่าน ไม่สามารถนำมากล่าวได้ทั้งหมด ผู้วิจัยใคร่ขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.สุวิทย์ แพงกันยา และ ผู้ช่วยศาสตราจารย์ ดร.ปริญญ์ บุญกนิษฐ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ได้ให้ความรู้ คำแนะนำ ตรวจสอบและแก้ไขข้อบกพร่องต่าง ๆ ด้วยความเข้าใจและเอาใจใส่ในทุกขั้นตอน เพื่อให้การเขียนวิทยานิพนธ์ฉบับนี้มีความสมบูรณ์อย่างที่สุด

ขอขอบคุณ เจ้าหน้าที่จากห้องปฏิบัติการตรวจวัดมลพิษจากยานพาหนะ กรมควบคุมมลพิษ และเจ้าหน้าที่จากสำนักงานสิ่งแวดล้อม กรุงเทพมหานคร ในการอนุเคราะห์สถิติชุดข้อมูลด้านอุตุนิยมวิทยา และมลพิษทางอากาศในพื้นที่กรุงเทพมหานคร พร้อมทั้งให้ความรู้และความเข้าใจเกี่ยวกับฝุ่นละอองขนาดเล็ก การจัดเก็บข้อมูลและการวิเคราะห์ข้อมูลที่ได้มาจากสถานีสำรวจต่าง ๆ ตลอดจนความรู้อื่นๆ ที่ไม่ได้กล่าวมาทั้งหมด และขอบคุณเจ้าหน้าที่ของการทางพิเศษแห่งประเทศไทย ที่อนุเคราะห์ข้อมูลสถิติจราจรบนทางพิเศษสายต่าง ๆ สำหรับการนำมาศึกษาวิจัยในครั้งนี้ จนเป็นที่ลุล่วงและประสบผลสำเร็จดังผลงานวิจัยที่ปรากฏ

ขอขอบคุณพระคุณบิดา พ.ต.ท.สุเทพ เจ็ดวรรณะ มารดา นางจินตนา เจ็ดวรรณะ และครอบครัวของผู้วิจัยที่อยู่เบื้องหลังในความสำเร็จลุล่วงในครั้งนี้ ทั้งยังให้ความช่วยเหลือสนับสนุน และให้กำลังใจตลอดมาตราบจนถึงบุคคลต่าง ๆ ที่ให้ความช่วยเหลืออีกมาก ที่ผู้วิจัยไม่สามารถกล่าวชื่อได้ทั้งหมดในที่นี้ ผู้วิจัยจึงขอกราบขอบพระคุณเป็นอย่างสูง มา ณ โอกาสนี้

กฤษณ์ เจ็ดวรรณะ



## สารบัญ

|  | หน้า |
|--|------|
| บทคัดย่อ   | (ก)  |
| ABSTRACT   | (ข)  |
| กิตติกรรมประกาศ  | (ค)  |
| สารบัญ   | (ง)  |
| สารบัญตาราง  | (ฉ)  |
| สารบัญภาพ  | (ญ)  |
| 1. บทนำ  | 1    |
| 1.1 ความเป็นมาและความสำคัญของปัญหา                                       | 1    |
| 1.2 วัตถุประสงค์   | 4    |
| 1.3 ขอบเขตการศึกษา   | 4    |
| 1.4 สมมติฐาน   | 6    |
| 1.5 กรอบแนวคิด   | 6    |
| 1.6 ประโยชน์ที่คาดว่าจะได้รับ  | 8    |
| 1.7 นิยามศัพท์   | 9    |
| 2. แนวคิด ทฤษฎี และการทบทวนวรรณกรรม                                      | 11   |
| 2.1 สถานการณ์และผลกระทบของ PM 2.5  | 11   |
| 2.2 การตรวจวัดค่า PM 2.5 และมลพิษทางอากาศ                                | 26   |
| 2.3 แนวคิดและทฤษฎีที่เกี่ยวกับการพัฒนาแบบจำลอง                           | 35   |
| 2.4 การคาดการณ์มลพิษทางอากาศ ตั้งแต่อดีตจนถึงปัจจุบัน                    | 44   |
| 2.5 สรุปผลการทบทวนวรรณกรรม   | 49   |
| 3. วิธีการดำเนินการวิจัย (Methodology)                                   | 55   |
| 3.1 ขั้นตอนการพัฒนาแบบจำลอง  | 55   |
| 3.2 การรวบรวมข้อมูล (Data Collection)                                    | 57   |
| 3.3 การตรวจสอบและสำรวจข้อมูลเบื้องต้น (Data Exploration)                 | 65   |
| 3.4 การจัดเตรียมชุดข้อมูล (Data Preparation)                             | 81   |
| 3.5 แบบจำลอง (Modelling)   | 88   |
| 3.6 การวัดค่าประสิทธิภาพ (Performance) ของแบบจำลอง<br>(Model Evaluation) | 97   |
| 4. ผลการวิจัยและพัฒนา (Result)   | 100  |
| 4.1 ผลการทดสอบแบบจำลอง โดยไม่อาศัยข้อมูลในอดีต (No Historical Data)      | 100  |

## สารบัญ (ต่อ)

|  | หน้า |
|--|------|
| 4.2 ผลการทดสอบแบบจำลอง โดยอาศัยข้อมูลในอดีต (Model Results by Historical Data) | 106  |
| 4.3 ภาพรวมผลการทดสอบแบบจำลองในสถานการณ์ปกติ และเมื่อมีการแพร่ระบาดของ Covid-19 | 119  |
| 5. การอภิปรายผล (Discussion)   | 122  |
| 5.1 แบบจำลองที่มีประสิทธิภาพและความแม่นยำสูงสุด                                | 123  |
| 5.2 แนวทางการนำแบบจำลองไปใช้กรณีมีข้อจำกัดเกี่ยวกับข้อมูล                      | 135  |
| 6. สรุปผลการศึกษา (Conclusion)   | 137  |
| 6.1 สรุปผลการศึกษา   | 137  |
| 6.2 ข้อเสนอแนะ   | 140  |
| 6.3 การวิจัยในอนาคต (Future Study)   | 141  |
| เอกสารอ้างอิง  | 142  |
| ภาคผนวก  | 153  |
| ภาคผนวก ก เอกสารตีพิมพ์  | 154  |
| ภาคผนวก ข รายละเอียดชุดคำสั่ง (Code) ของการพัฒนาแบบจำลอง                       | 196  |
| ประวัติการศึกษาและการทำงาน   | 202  |



## สารบัญตาราง

| ตาราง |  | หน้า |
|-------|--|------|
| 1.1   | ชุดข้อมูลด้านอุตุนิยมวิทยาและมลพิษทางอากาศแบ่งตามชุดการทดสอบ   | 8    |
| 2.1   | ผลกระทบต่อสุขภาพจากการสัมผัส Particulate matter (PM) ในอากาศ   | 13   |
| 2.2   | มาตรการในการจัดการการแพร่ระบาดของ Covid-19 ของประเทศไทย  | 18   |
| 2.3   | ค่ามาตรฐานฝุ่นละอองในต่างประเทศ  | 26   |
| 2.4   | ค่ามาตรฐานฝุ่นละอองในบรรยากาศของประเทศไทย  | 26   |
| 2.5   | เกณฑ์ของดัชนีคุณภาพอากาศของประเทศไทย   | 27   |
| 2.6   | ค่าความเข้มข้นของสารมลพิษทางอากาศที่เทียบเท่ากับค่าดัชนี<br>คุณภาพอากาศ  | 28   |
| 2.7   | วิธีการตรวจวัดฝุ่นละอองในบรรยากาศของประเทศไทย  | 30   |
| 2.8   | สรุปการทบทวนงานวิจัยเกี่ยวกับการคาดการณ์ PM 2.5 ในช่วงที่ผ่านมา  | 51   |
| 3.1   | ข้อมูลที่ใช้ในการศึกษา (Raw Dataset)   | 59   |
| 3.2   | สรุปจำนวนชุดข้อมูลแยกรายสถานี  | 59   |
| 3.3   | สถิติด้านมลพิษทางอากาศช่วงก่อนการระบาดของ Covid-19<br>และระหว่างการแพร่ระบาด Covid-19  | 66   |
| 3.4   | ตัวอย่างการแยกข้อมูลค่า PM 2.5 แยกข้อมูลรายเดือนในแต่ละปี<br>ตั้งแต่ปี พ.ศ. 2558 ถึงปีพ.ศ. 2564                                      | 75   |
| 3.5   | ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร Pearson Correlation   | 76   |
| 3.6   | ผลการวิเคราะห์พารามิเตอร์ด้านมลพิษ<br>ด้วย Kruskal-wallis test (มกราคม 2558 - ธันวาคม 2564)  | 77   |
| 3.7   | ตัวแปรที่มาจากการปรับโครงสร้าง (Feature Engineering)   | 82   |
| 3.8   | ตัวแปรที่มาจากการแปลงข้อมูล (Data Transformation)  | 84   |
| 3.9   | ตัวแปรที่ไม่เกี่ยวข้องกับเวลา และข้อมูล PM 2.5 ในอดีต (10 กลุ่มตัวแปร)   | 85   |
| 3.10  | ตัวแปรที่เกี่ยวข้องกับเวลา และข้อมูล PM 2.5 ในอดีต (14 กลุ่มตัวแปร)  | 85   |
| 3.11  | การแปรผลระดับความถูกต้องแม่นยำของค่า MAPE  | 99   |
| 4.1   | ผลการทดสอบจำนวนตัวแปร (Feature) ที่ใช้ในการสร้างแบบจำลอง<br>กรณี trainset, Test set and Testout set<br>ของแบบจำลอง Decision Tree     | 101  |
| 4.2   | ผลการทดสอบจำนวนตัวแปร (Feature) ที่ใช้ในการสร้างแบบจำลอง<br>กรณี trainset, Test set and Testout set ของแบบจำลอง<br>Gradient Boosting | 101  |

## สารบัญตาราง (ต่อ)

| ตาราง |  | หน้า |
|-------|--|------|
| 4.3   | ผลการทดสอบจำนวนตัวแปร (Feature) ที่ใช้ในการสร้างแบบจำลอง<br>กรณี trainset, Test set and Testout set ของแบบจำลอง<br>K-neighbors.  | 102  |
| 4.4   | ผลการทดสอบจำนวนตัวแปร (Feature) ที่ใช้ในการสร้างแบบจำลอง<br>กรณี trainset, Test set and Testout set ของแบบจำลอง MLP  | 103  |
| 4.5   | ผลการทดสอบจำนวนตัวแปร (Feature) ที่ใช้ในการสร้างแบบจำลอง<br>กรณี trainset, Test set and Testout set ของแบบจำลอง<br>Random Forest   | 103  |
| 4.6   | ผลการทดสอบจำนวนตัวแปร (Feature) ที่ใช้ในการสร้างแบบจำลอง<br>กรณี trainset, Test set and Testout set ของแบบจำลอง Ridge  | 104  |
| 4.7   | สรุปผลการคัดเลือกตัวแปรของแบบจำลองชนิดต่าง ๆ<br>โดยไม่อาศัยข้อมูลในอดีต (No Historical Data)   | 104  |
| 4.8   | ผลการทดสอบความถูกต้องและความแม่นยำของแบบจำลอง<br>Training set และ Test Set โดยไม่อาศัยข้อมูลในอดีต (No Historical Data)  | 105  |
| 4.9   | ผลการทดสอบความถูกต้องและความแม่นยำของแบบจำลอง<br>Testout (Covid-19 Pandemic) โดยไม่อาศัยข้อมูลในอดีต (No Historical<br>Data)   | 105  |
| 4.10  | แสดงผลการทดสอบจำนวนตัวแปร (Feature) ที่ใช้ในการสร้างแบบจำลอง<br>กรณี trainset, Test set and Testout set ของแบบจำลอง<br>ทั้ง 6 ชนิด เมื่อเพิ่มตัวแปร (Feature) ที่เป็นข้อมูลค่า PM 2.5<br>เฉลี่ยรายเดือน (Lag by Month) | 108  |
| 4.11  | สรุปผลการคัดเลือกตัวแปรสำหรับนำไปวิเคราะห์แบบจำลอง<br>กรณีเพิ่มตัวแปรข้อมูลค่า PM 2.5 เฉลี่ยรายเดือน (Lag by Month)  | 110  |
| 4.12  | ผลการทดสอบความถูกต้องและความแม่นยำของแบบจำลอง Training set<br>และ Test Set โดยอาศัยข้อมูลค่า PM 2.5<br>เฉลี่ยรายเดือน (Lag by Month)   | 111  |
| 4.13  | ผลการทดสอบความถูกต้องและความแม่นยำของแบบจำลอง<br>Testout (COVID-19 Pandemic) โดยอาศัยข้อมูลค่า PM 2.5<br>เฉลี่ยรายเดือน (Lag by Month)   | 111  |

## สารบัญตาราง (ต่อ)

| ตาราง |   | หน้า |
|-------|---|------|
| 4.14  | แสดงผลการทดสอบจำนวนตัวแปร (Feature) ที่ใช้ในการสร้างแบบจำลอง<br>กรณี trainset, Test set and Testout set ของแบบจำลอง<br>ทั้ง 6 ชนิด เมื่อเพิ่มตัวแปร (Feature) ที่เป็นข้อมูลค่า PM 2.5<br>เฉลี่ยรายวัน (Lag by Day)      | 112  |
| 4.15  | สรุปผลการคัดเลือกตัวแปรสำหรับนำไปวิเคราะห์แบบจำลอง<br>กรณีเพิ่มตัวแปรข้อมูลค่า PM 2.5 เฉลี่ยรายวัน (Lag by Day)   | 114  |
| 4.16  | ผลการทดสอบความถูกต้องและความแม่นยำของแบบจำลอง<br>Training set และ Test Set โดยอาศัยข้อมูลค่า PM 2.5 เฉลี่ยรายวัน<br>(Lag by Day)  | 115  |
| 4.17  | ผลการทดสอบความถูกต้องและความแม่นยำของแบบจำลอง<br>Testout (COVID-19 Pandemic) โดยอาศัยข้อมูลค่า PM 2.5<br>เฉลี่ยรายวัน (Lag by Day)  | 115  |
| 4.18  | แสดงผลการทดสอบจำนวนตัวแปร (Feature) ที่ใช้ในการสร้างแบบจำลอง<br>กรณี trainset, Test set and Testout set ของแบบจำลอง<br>ทั้ง 6 ชนิด เมื่อเพิ่มตัวแปร (Feature) ที่เป็นข้อมูลค่า PM 2.5<br>เฉลี่ยรายชั่วโมง (Lag by Hour) | 116  |
| 4.19  | สรุปผลการคัดเลือกตัวแปรสำหรับนำไปวิเคราะห์แบบจำลอง<br>กรณีเพิ่มตัวแปรข้อมูลค่า PM 2.5 เฉลี่ยรายชั่วโมง (Lag by Hour)  | 118  |
| 4.20  | ผลการทดสอบความถูกต้องและความแม่นยำของแบบจำลอง<br>Training set และ Test Set โดยอาศัยข้อมูลค่า PM 2.5<br>เฉลี่ยรายชั่วโมง (Lag by Hour)   | 119  |
| 4.21  | ผลการทดสอบความถูกต้องและความแม่นยำของแบบจำลอง<br>Testout (COVID-19 Pandemic) โดยอาศัยข้อมูลค่า PM 2.5<br>เฉลี่ยรายชั่วโมง (Lag by Hour)   | 119  |
| 4.22  | ผลการวิเคราะห์ Train set และ Test Set ของแบบจำลองชนิดต่าง ๆ   | 120  |
| 4.23  | ผลการวิเคราะห์ Testout Set ของแบบจำลอง ชนิดต่าง ๆ   | 121  |
| 5.1   | ค่าเฉลี่ยของ PM 2.5 รายเดือนเปรียบเทียบระหว่างก่อนการระบาดของ<br>ของ COVID-19 และเมื่อมีการแพร่ระบาด COVID-19   | 122  |
| 5.2   | ผลการวิเคราะห์ประสิทธิภาพแบบจำลองชนิดต่าง ๆ ที่อาศัยข้อมูล PM 2.5<br>เฉลี่ยรายชั่วโมง (by hour)   | 123  |

## สารบัญตาราง (ต่อ)

| ตาราง |   | หน้า |
|-------|---|------|
| 5.3   | ตัวแปรที่เกี่ยวข้อง (Features) ที่ใช้ในการทดสอบแบบจำลอง Gradient Boosting ในกรณีชุดข้อมูลเฉลี่ยรายชั่วโมง (by hour) | 136  |
| 6.1   | สรุปผลความถูกต้องและแม่นยำ ( $R^2$ ) ของแบบจำลองที่ใช้ในงานศึกษา  | 140  |



## สารบัญภาพ

| ภาพ  |  | หน้า |
|------|--|------|
| 1.1  | ขอบเขตพื้นที่การศึกษา  | 4    |
| 1.2  | กระบวนการแบ่งชุดข้อมูลสำหรับแบบจำลอง   | 7    |
| 2.1  | ขนาดของฝุ่น PM 2.5   | 12   |
| 2.2  | กรุงเทพมหานครเป็นเมืองที่มีค่าเฉลี่ย PM 2.5 รายปี<br>สูงเป็นอันดับ 4 ของประเทศ   | 15   |
| 2.3  | เมืองที่มีรถติดมากที่สุดในโลก ปี พ.ศ. 2562<br>โดยกรุงเทพมหานครอยู่อันดับที่ 10   | 15   |
| 2.4  | Timeline การแพร่ระบาดของ COVID-19 ในประเทศไทย<br>และจำนวนผู้ติดเชื้อใหม่รายวัน   | 20   |
| 2.5  | การกระจายตัวการเดินทางบนทางด่วนในเมืองตั้งแต่ พ.ศ. 2558 – 2564<br>โดยเมื่อเกิดการแพร่ระบาดของ COVID-19 ทำให้การเดินทางบนถนน<br>ลดลงเฉลี่ยร้อยละ 23                                 | 21   |
| 2.6  | สถิติข้อมูลค่าความเข้มข้นเฉลี่ยของ PM 2.5 ก่อนที่จะมีการแพร่ระบาด<br>ของ COVID-19 (พ.ศ. 2558-2562) และเมื่อมีการ Lockdown<br>ในขณะที่มีการแพร่ระบาดของโรค COVID-19 (พ.ศ.2563-2564) | 24   |
| 2.7  | ข้อมูลแสดงค่าความเข้มข้นของ PM 2.5 เฉลี่ยรายเดือน ตั้งแต่เดือน<br>เมษายน – ธันวาคม   | 25   |
| 2.8  | วิธีการตรวจวัดแบบ Beta Radiation Attenuation หรือ<br>Beta Ray Attenuation  | 30   |
| 2.9  | ระบบการทำงานของ Tapered Element Oscillating<br>Microbalance: TEOM  | 30   |
| 2.10 | เครื่อง Light Scattering ของกรมควบคุมมลพิษ   | 31   |
| 2.11 | การเก็บอนุภาคฝุ่นด้วย Dichotomous Air Sample   | 31   |
| 2.12 | สถานีตรวจวัดคุณภาพอากาศของกรมควบคุมมลพิษ   | 33   |
| 2.13 | สถานีตรวจวัดคุณภาพอากาศของสำนักสิ่งแวดล้อม กรุงเทพมหานคร   | 34   |
| 2.14 | การวิเคราะห์พหุตัวแปร (Multivariate Analysis Model)  | 38   |
| 2.15 | ประเภทของ Machine Learning   | 39   |
| 2.16 | Machine Learning ประเภท Supervised Learning  | 40   |
| 2.17 | Machine Learning ประเภท Unsupervised Learning  | 41   |
| 2.18 | กระบวนการทำงานของ Deep Learning  | 42   |
| 2.19 | การแสดงผลภาพที่ได้ถูกรวบรวมโดย Deep Learning   | 42   |

## สารบัญภาพ (ต่อ)

| ภาพ  |   | หน้า |
|------|---|------|
| 2.20 | แสดงภาพรวมของชนิด AI  | 43   |
| 3.1  | ภาพรวมของขั้นตอนการดำเนินงานวิจัย   | 56   |
| 3.2  | ตัวอย่างชุดข้อมูล Dataset ที่ใช้ในการศึกษา  | 59   |
| 3.3  | การทำ histogram ของค่า PM 2.5   | 63   |
| 3.4  | ภาพรวมการแบ่งกลุ่มข้อมูล (Grouping Data)  | 64   |
| 3.5  | (ก) ค่าความเข้มข้นเฉลี่ยของ PM 2.5 ก่อนการระบาดของ Covid-19<br>และระหว่างการแพร่ระบาด Covid-19          | 68   |
| 3.5  | (ข) ค่าความเข้มข้นเฉลี่ยของ PM 10 ก่อนการระบาดของ Covid-19<br>และระหว่างการแพร่ระบาด Covid-19           | 68   |
| 3.5  | (ค) ค่าความเข้มข้นเฉลี่ยของ NOx ก่อนการระบาดของ Covid-19<br>และระหว่างการแพร่ระบาด Covid-19             | 70   |
| 3.5  | (ง) ค่าความเข้มข้นเฉลี่ยของ NO <sub>2</sub> ก่อนการระบาดของ Covid-19<br>และระหว่างการแพร่ระบาด Covid-19 | 70   |
| 3.5  | (จ) ค่าความเข้มข้นเฉลี่ยของ NO ก่อนการระบาดของ Covid-19<br>และระหว่างการแพร่ระบาด Covid-19              | 72   |
| 3.5  | (ฉ) ค่าความเข้มข้นเฉลี่ยของ CO ก่อนการระบาดของ Covid-19<br>และระหว่างการแพร่ระบาด Covid-19              | 72   |
| 3.5  | (ช) ค่าความเข้มข้นเฉลี่ยของ O <sub>3</sub> ก่อนการระบาดของ Covid-19<br>และระหว่างการแพร่ระบาด Covid-19  | 73   |
| 3.6  | ชุดข้อมูลเชิงประจักษ์ของมลพิษทางอากาศเฉลี่ยรายวัน ย้อนหลัง 6 ปี<br>(พ.ศ. 2558 - 2564)                   | 78   |
| 3.7  | สถิติข้อมูล PM 2.5 รายเดือน   | 79   |
| 3.8  | ผลกระทบของ PM 2.5 ต่ออุณหภูมิ   | 80   |
| 3.9  | ผลกระทบรายชั่วโมงของ PM 2.5   | 81   |
| 3.10 | กระบวนการนำข้อมูลชุดต่าง ๆ เข้าแบบจำลอง   | 87   |
| 3.11 | ภาพแสดงโครงสร้างของ Decision tree   | 88   |
| 3.12 | แสดงโครงสร้างของ Gradient boosting  | 90   |
| 3.13 | แสดงโครงสร้างของ K-nearest neighbor   | 91   |
| 3.14 | แสดงโครงสร้างของ MLP Regressor Model  | 93   |
| 3.15 | แสดงโครงสร้างของ Random forest Model  | 96   |
| 4.1  | การทำ Autocorrelation plot ของข้อมูลรายวัน  | 106  |

## สารบัญภาพ (ต่อ)

| ภาพ  |   | หน้า |
|------|---|------|
| 5.1  | ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง<br>Discision Tree Regression กรณี Training set : $R^2 = 0.8673$ ,<br>RMSE=6.3513, MAPE=31.58%  | 124  |
| 5.2  | ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง<br>Discision Tree Regression กรณี Test Set: $R^2=0.8400$ ,<br>RMSE=6.9425, MAPE=32.99%         | 124  |
| 5.3  | ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง<br>Discision Tree Regression กรณี Testout Set : $R^2=0.8507$ ,<br>RMSE=6.3513, MAPE=33.45%     | 125  |
| 5.4  | ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง<br>Gradient boosting กรณี Training set : $R^2 = 0.9533$ , RMSE=4.1344,<br>MAPE=17.13%          | 126  |
| 5.5  | ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง<br>Gradient boosting กรณี Test Set : $R^2=0.8790$ , RMSE=6.1375,<br>MAPE=20.68%                | 126  |
| 5.6  | ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง<br>Gradient boosting กรณี Testout Set: $R^2=0.8720$ , RMSE=5.1344,<br>MAPE=27.18%              | 127  |
| 5.7  | ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง<br>K-nearest neighbor กรณี Training set : $R^2=0.8603$ , RMSE=6.4414,<br>MAPE=30.33%           | 128  |
| 5.8  | ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง<br>K-nearest neighbor กรณี Test Set : $R^2=0.8383$ , RMSE=6.9019,<br>MAPE=31.88%               | 128  |
| 5.9  | ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง<br>K-nearest neighbor กรณี Testout Set : $R^2=0.8295$ , RMSE=6.4414,<br>MAPE=34.79%            | 129  |
| 5.10 | ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง<br>Multi Layer Perception (MLP) กรณี Training set : $R^2=0.8530$ ,<br>RMSE=6.6861, MAPE=32.20% | 130  |

## สารบัญภาพ (ต่อ)

| ภาพ  |  | หน้า |
|------|--|------|
| 5.11 | ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง Multi Layer Perception (MLP) กรณี Test Set : $R^2=0.8453$ , RMSE=6.8252, MAPE=32.49%    | 130  |
| 5.12 | ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง Multi Layer Perception (MLP) กรณี Testout Set : $R^2=0.8645$ , RMSE=6.6861, MAPE=30.79% | 131  |
| 5.13 | ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง Random Forest กรณี Training set : $R^2=0.8947$ , RMSE=5.6590, MAPE=28.50%               | 132  |
| 5.14 | ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง Random Forest กรณี Test Set : $R^2=0.8578$ , RMSE=6.5434, MAPE=31.55%                   | 132  |
| 5.15 | ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง Random Forest กรณี Testout Set : $R^2=0.8691$ , RMSE=5.6590, MAPE=30.87%                | 133  |
| 5.16 | ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง Ridge Regression กรณี Training Set : $R^2=0.7375$ , RMSE=8.9346, MAPE=39.69%            | 134  |
| 5.17 | ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง Ridge Regression กรณี Test Set : $R^2=0.7391$ , RMSE=8.8637, MAPE=40.06%                | 134  |
| 5.18 | ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง Ridge Regression กรณี Testout Set : $R^2=0.7654$ , RMSE=8.9346, MAPE=37.87%             | 135  |
| 6.1  | เปรียบเทียบความเข้มข้นของ PM 2.5 รายเดือน ระหว่างก่อนและหลัง   | 138  |



# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันปัญหามลพิษทางอากาศโดยเฉพาะ PM 2.5 นับว่าเป็นปัญหาหลักของโลก โดยกล่าวได้ว่าไม่มีที่ไหนในโลกปลอดภัยจาก PM 2.5 จากสถิติข้อมูล PM 2.5 ในปีค.ศ. 2000 - 2019 ที่วิเคราะห์โดย Yu et al. (2023) พบว่า พื้นที่ผิวโลกกว่าร้อยละ 99 มี PM 2.5 มากกว่า  $15 \mu\text{g}/\text{m}^3$  โดยมีเพียงร้อยละ 0.01 ของพื้นผิวโลกเท่านั้น ที่ประชากรสามารถสูดดมอากาศที่บริสุทธิ์และสะอาด ซึ่งผลการศึกษายังพบว่า ค่าเฉลี่ยของ PM 2.5 ทั่วโลกอยู่ที่  $32.8 \mu\text{g}/\text{m}^3$  โดยพื้นที่ผิวโลกที่ต้องเผชิญกับ PM 2.5 สูงที่สุดในโลก คือ ภูมิภาคเอเชียตะวันออก ซึ่งมีค่า PM 2.5 เฉลี่ยตลอดทั้งปีเท่ากับ  $50 \mu\text{g}/\text{m}^3$  ตามด้วยเอเชียใต้เท่ากับ  $37 \mu\text{g}/\text{m}^3$  และแอฟริกาเหนือเท่ากับ  $30 \mu\text{g}/\text{m}^3$

หน่วยงานป้องกันสิ่งแวดล้อม ประเทศสหรัฐอเมริกา US. EPA (EPA, 2016) ได้กำหนดค่ามาตรฐานของฝุ่นละอองขนาดเล็ก โดยใช้ค่า PM หรือย่อมาจาก Particulate Matter เป็นตัววัดคุณภาพของอากาศ ซึ่งฝุ่น PM2.5 หรือที่เรียกว่า “ฝุ่นละเอียด” (Fine Particles) คือ อนุภาคขนาดเล็กในอากาศที่มีขนาดเส้นผ่าศูนย์กลางอยู่ที่ 2.5 ไมครอน ซึ่งมีความอันตรายต่อสุขภาพมนุษย์ เนื่องจากสามารถเข้าสู่ระบบทางเดินหายใจได้อย่างง่าย และเป็นสาเหตุของปัญหาสุขภาพ เช่น โรคทางเดินหายใจ โรคหัวใจ และปัญหาสุขภาพอื่นๆ ซึ่งในประเทศไทยตั้งแต่ พ.ศ. 2554-2561 มีค่าฝุ่น PM 2.5 เฉลี่ยรายปีอยู่ที่  $24 \mu\text{g}/\text{m}^3$  โดยพบว่า ปัจจัยหลักที่ส่งผลให้เกิดฝุ่น PM 2.5 ในประเทศไทย คือ การคมนาคมขนส่ง เช่นเดียวกับภูมิภาคอื่นๆ ของโลก ซึ่งจากหลักฐานการศึกษาและวิจัยพบว่า การคมนาคมขนส่งเป็นตัวการสำคัญที่ส่งผลให้เกิด PM 2.5 มาโดยตลอด (Heydari et al., 2020; Hodan and Barnard, 2004; Li and Managi, 2021; Li et al., 2019; Sun et al., 2019) และรองลงมา คือ การเผาจากภาคการเกษตรที่เป็นชีวมวล ฝุ่นละอองทั่วไปและการก่อสร้าง และอุตสาหกรรมโรงงาน ตามลำดับ

ทุกวันนี้ปัญหาหมอกควันอันเกิดขึ้นจากฝุ่น PM 2.5 กลายเป็นปัญหาระดับชาติ ของประเทศไทย โดยเฉพาะอย่างยิ่ง ในพื้นที่กรุงเทพมหานครและปริมณฑล รวมทั้งพื้นที่ทางภาคเหนือหลายจังหวัดของประเทศไทยอันได้แก่ เชียงราย เชียงใหม่ ลำปาง เป็นต้น (Greenpeace, 2016) ทั้งประเทศไทยยังมีปัญหาการจราจรติดขัดอยู่ในอันดับโลก โดยเฉพาะอย่างยิ่ง กรุงเทพมหานครที่ติดอันดับเมืองที่รถติดที่สุดในโลกทุกปี (Pishue and Trepanier, 2021; Tomtom Traffic Index, 2021)

จนกระทั่ง ต้นปี 2020 ที่มีการระบาดของโรค Covid-19 ไปทั่วโลก ส่งผลให้รูปแบบการดำเนินชีวิตของผู้คนเปลี่ยนแปลงไปมีการเว้นระยะห่างทางสังคม โดยมีมาตรการที่เรียกว่า Lockdown ให้คนอยู่กับบ้าน ให้มีการเรียนและการทำงานจากที่บ้าน และเมื่อกิจกรรมที่เป็นปกติของผู้คนถูกจำกัดส่งผลให้การเดินทางลดน้อยลงเป็นอย่างมาก มีการเปลี่ยนรูปแบบการเดินทางอย่างเห็นได้ชัดในหลายประเทศ (Dasgupta and Srikanth, 2020; Venter et al., 2020) โดยเฉพาะอย่างยิ่ง ในเมืองหลวงอย่างกรุงเทพมหานคร ช่วงที่ Covid-19 ระบาดอย่างหนัก และมีการจำกัดกิจกรรมของประชาชนให้

อยู่แต่บ้าน ทำให้การเดินทางบนถนนลดลง โดยเฉพาะการจราจรบนทางด่วนลดลงไปถึงร้อยละ 23 และการเดินทางด้วยระบบขนส่งสาธารณะลดลงไปถึงร้อยละ 50 (การทางพิเศษแห่งประเทศไทย, 2021) เนื่องจากการเดินทางนั้นเป็นสาเหตุหลักของการเกิด PM 2.5 และปริมาณการเดินทางที่เปลี่ยนแปลงลดลงไปอย่างมากในช่วงที่มีการแพร่ระบาดของ Covid-19 การวิเคราะห์การเปลี่ยนแปลงความเข้มข้นของ PM 2.5 ในช่วงที่มีการแพร่ระบาดของ Covid-19 นั้น จะทำให้ทราบถึงพฤติกรรมความเข้มข้นของ PM 2.5 ที่เปลี่ยนแปลงไปเมื่อกิจกรรมต่างๆ ถูกจำกัด จะทำให้สามารถวิเคราะห์หาความสัมพันธ์ของสาเหตุและแนวทางในการแก้ปัญหา PM 2.5 ได้อย่างยั่งยืนในอนาคต

นอกจากการทราบถึงพฤติกรรมความเข้มข้นของ PM 2.5 ที่มีการเปลี่ยนแปลงไปเนื่องจากสาเหตุและปัจจัยต่างๆ แล้วนั้น หากสามารถพยากรณ์หรือคาดการณ์ PM 2.5 ที่จะเกิดขึ้นในอนาคต หรือสามารถทราบถึงความสัมพันธ์ของตัวแปรต่างๆ ที่ส่งผลต่อความเข้มข้นของ PM 2.5 ในอนาคตได้ จะทำให้ผู้ที่เกี่ยวข้องสามารถวางแผนมาตรการหรือแนวทางในการป้องกันและแก้ปัญหาได้อย่างมีประสิทธิภาพ

จากการศึกษาและทบทวนงานวิจัยที่ผ่านมา (รายละเอียดแสดงในตารางที่ 2.8) พบว่า การจัดทำแบบจำลองเพื่อการคาดการณ์การเกิด PM 2.5 มีการพัฒนามาตามลำดับ ซึ่งในการคาดการณ์ในอดีตเริ่มตั้งแต่การคาดการณ์โดยใช้แบบจำลองความสัมพันธ์แบบ Regression Model และความก้าวหน้าในปัจจุบันมีการพัฒนาแบบจำลองโดยประยุกต์ใช้ AI คือ Machine Learning Techniques และจากผลการศึกษาที่ผ่านมาพบว่ามีความแม่นยำในการคาดการณ์ดีขึ้นมากกว่าวิธีการพัฒนาแบบจำลองโดยใช้ Regression Model อย่างไรก็ตาม Machine Learning Techniques มีความหลากหลายของแบบจำลองที่สามารถนำมาใช้ได้ จากการทบทวนงานวิจัยที่ผ่านมาพบว่า แบบจำลอง Machine Learning Techniques ที่นิยมใช้กัน ได้แก่ 1) Decision Tree Regressor CART 2) Gradient Boosting Regression 3) K-neighbor Regressor 4) MLP Regressor 5) Random Forest Regressor และ 6) Ridge Model โดยในงานวิจัยนี้ได้ดำเนินการพัฒนาแบบจำลองและทำการทดสอบทุกแบบจำลองที่มีการใช้ที่ผ่านมา เพื่อทำนำมาทำการเปรียบเทียบประสิทธิภาพความแม่นยำของแต่ละแบบจำลอง เพื่อพิจารณาว่าแบบจำลองใดมีความเหมาะสมมากที่สุดในการนำมาใช้คาดการณ์

ประสิทธิภาพความแม่นยำของแบบจำลองจากผลการศึกษาในต่างประเทศ พบว่า การประยุกต์ใช้ Machine Learning Techniques มีความแม่นยำในระดับที่ดีถึงดีมาก มีค่า  $R^2$  ที่ดีที่สุด ตั้งแต่ 0.89 ถึง 0.95 (Pan, 2018) สำหรับการพัฒนาแบบจำลองในประเทศไทย จากการทบทวนผลการศึกษาที่ผ่านมา ยังพบว่าได้มีการพัฒนาแบบจำลองในระดับ Regression Model โดยใช้ Multivariate Linear Regression Model โดยมีความแม่นยำ  $R^2$  ที่ดีที่สุดที่ 0.60 (Amnuaylojaroen, 2022) ซึ่งยังอยู่ในระดับที่ไม่สูงเมื่อเทียบกับการพัฒนาแบบจำลองของต่างประเทศ โดยในงานวิจัยนี้จะเป็นการพัฒนาแบบจำลองเพื่อคาดการณ์การเกิด PM 2.5 จากข้อมูลที่มีการสำรวจในประเทศไทย โดยการประยุกต์ใช้ Machine Learning Techniques ซึ่งคาดว่าจะได้ผลความแม่นยำที่สูงขึ้นจากการพัฒนาแบบจำลองในระดับ Regression Model

การพัฒนาแบบจำลองเพื่อคาดการณ์ PM 2.5 โดย Machine Learning Techniques จากการทบทวนงานวิจัยที่ผ่านมา(รายละเอียดแสดงในตารางที่ 2.8)พบว่า ตัวแปรที่นำมาใช้ในการคาดการณ์ PM 2.5 จะประกอบด้วย ตำแหน่งภูมิประเทศ สารมลพิษทางอากาศ ข้อมูลด้านอุตุนิยมวิทยา และอื่นๆ

โดยข้อมูลสารมลพิษทางอากาศจะมีตัวแปรที่ใช้ในการพัฒนาแบบจำลองในแต่ละงานวิจัยตั้งแต่ 2-9 ตัวแปร และข้อมูลด้านอุตุนิยมวิทยาในแต่ละงานวิจัยจะใช้ตั้งแต่ 4-5 ตัวแปร ซึ่งในงานวิจัยนี้ได้ใช้ตัวแปรครบทุกตัวแปรตามที่พบจากการทบทวนการศึกษาที่ผ่านมาและจากข้อมูลที่มาการสำรวจไว้ และจากการทบทวนผลการศึกษาในอดีตยังพบอีกว่า ปริมาณความเข้มข้นของ PM 2.5 มีความสัมพันธ์กับช่วงเวลาในแต่ละช่วงฤดูกาล แต่จากผลการศึกษาที่ผ่านมายังไม่ม้งานวิจัยที่ใช้ข้อมูลในอดีตมาใช้เป็นข้อมูลในการคาดการณ์ PM 2.5 ซึ่งมีความเป็นไปได้ที่จะทำให้แบบจำลองมีความแม่นยำมากขึ้น ทำให้ในงานวิจัยนี้จึงได้นำข้อมูลในอดีตมาเป็นตัวแปรสำหรับคาดการณ์ PM 2.5 ด้วย

การพัฒนาแบบจำลองเพื่อคาดการณ์ PM 2.5 โดย Machine Learning Techniques ความต่อเนื่องของข้อมูลและจำนวนชุดของข้อมูลเป็นสิ่งสำคัญ จะส่งผลทำให้แบบจำลองมีความน่าเชื่อถือมากขึ้น จากการทบทวนผลการศึกษาที่ผ่านมาพบว่า งานวิจัยที่ผ่านมาส่วนใหญ่ข้อมูลที่มีการนำมาใช้ในการพัฒนาแบบจำลองอยู่ในช่วงเวลาค่อนข้างสั้น ประมาณน้อยกว่า 1-2 ปี ซึ่งอาจไม่ครอบคลุมทุกฤดูกาลของสภาพแวดล้อมในรอบปี หรือเหตุการณ์ต่างๆ ที่ส่งผลต่อการเกิด PM 2.5 เช่นในช่วงที่ผ่านมาเกิดการแพร่ระบาดของ Covid-19 สำหรับในงานวิจัยนี้ได้ทำการรวบรวมข้อมูลจากหน่วยงานที่เกี่ยวข้องในช่วงเวลาที่ยาวนานกว่าที่เคยมีการศึกษาในอดีต (รายละเอียดแสดงในตารางที่ 2.8) และครอบคลุมระยะเวลาที่เกิดสถานการณ์การแพร่ระบาดของ COVID-19 โดยมีการเก็บข้อมูลตั้งแต่ 1 มกราคม พ.ศ.2558 ถึง 31 ธันวาคม 2564 ครอบคลุมระยะเวลา 7 ปีเต็ม และมีจำนวนชุดข้อมูล 2,303,892 ชุดข้อมูล ซึ่งประกอบด้วยจำนวนชุดข้อมูลจำนวนมากและครอบคลุมระยะเวลาที่เกิดสถานการณ์การแพร่ระบาดของ COVID-19

สำหรับสถานการณ์ที่มีการแพร่ระบาดของ COVID-19 เกิดขึ้นในประเทศไทยประมาณปีพ.ศ. 2563 ในขณะนั้นมีมาตรการต่างๆ เกิดขึ้นมากมาย เพื่อจำกัดกิจกรรมลดโอกาสที่จะให้ผู้คนมาพบเจอกัน โดยเฉพาะมาตรการ Lockdown ส่งผลทำให้การเดินทางลดลงเป็นอย่างมาก จากการคมนาคมขนส่งเป็นสาเหตุสำคัญของการเกิด PM 2.5 ส่งผลทำให้ค่าความเข้มข้นของ PM 2.5 มีค่าลดลงอย่างเห็นได้ชัด และมีความผันผวนอย่างมากเมื่อเทียบกับช่วงสถานการณ์ปกติ จากการทบทวนงานวิจัยที่ผ่านมายังไม่ม้งานวิจัยใดได้ใช้ข้อมูลในช่วงสถานการณ์ที่มีการแพร่ระบาดของ COVID-19 มาทำการพัฒนาแบบจำลองหรือนำมาทดสอบประสิทธิภาพของแบบจำลอง ซึ่งเกิดข้อสงสัยว่าข้อมูลในช่วงเวลาดังกล่าวจะส่งผลต่อการพัฒนาและประสิทธิภาพของแบบจำลองอย่างไร

การศึกษาวิเคราะห์พฤติกรรมกรการเกิด PM 2.5 ทั้งในช่วงสถานการณ์ปกติ และช่วงที่มีการแพร่ระบาดของ Covid-19 ประกอบกับการพัฒนาแบบจำลองเพื่อพยากรณ์และตรวจสอบความสัมพันธ์ของตัวแปรที่ส่งผลต่อการเกิด PM 2.5 จะทำให้เข้าใจถึงพฤติกรรมและประสิทธิภาพของแบบจำลองต่างๆ ที่สามารถจะนำมาใช้คาดการณ์การเกิด PM 2.5 จะเป็นเครื่องมือที่สำคัญในการนำไปใช้ในการป้องกันและแก้ปัญหา PM 2.5 ในอนาคตต่อไป

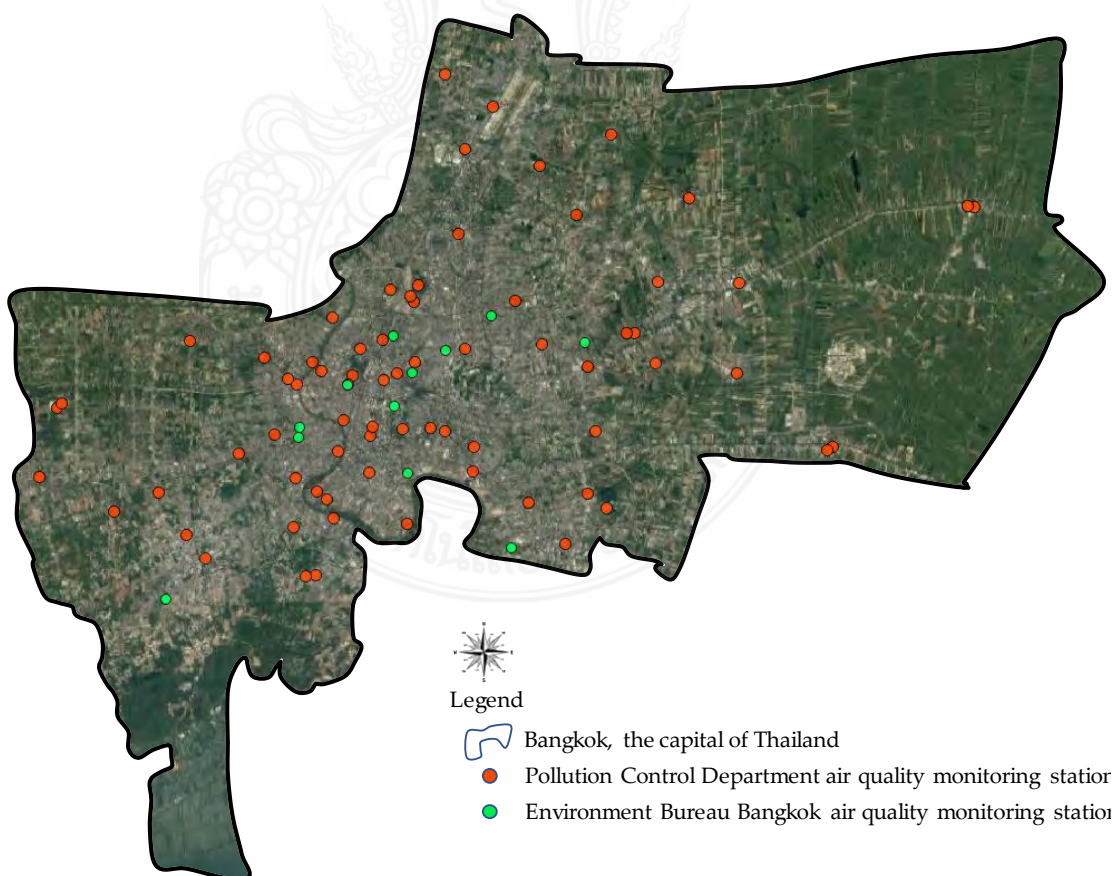
## 1.2 วัตถุประสงค์

งานวิจัยนี้ จะดำเนินการศึกษาและพัฒนาแบบจำลองการพยากรณ์การเกิด PM 2.5 โดยใช้ข้อมูลสถิติมลพิษทางอากาศ ในช่วงปี พ.ศ. 2558 – 2564 โดยมีรายละเอียดวัตถุประสงค์การศึกษา ดังนี้

- 1.2.1 ศึกษาพฤติกรรมและวิเคราะห์ความเข้มข้นของ PM 2.5 ที่เปลี่ยนแปลงไปเมื่อเกิดการแพร่ระบาดของ Covid-19
- 1.2.2 พัฒนาแบบจำลองคณิตศาสตร์ ด้วยการใช้ Machine Learning techniques ในการพยากรณ์การเกิด PM 2.5
- 1.2.3 เพื่อการสร้างแบบจำลองทางคณิตศาสตร์ที่ดีที่สุดที่นำไปใช้ในการทำนายค่า PM 2.5 ในอนาคต โดยสามารถใช้ได้ทั้งในช่วงเวลาปกติ และช่วงสถานการณ์ที่ไม่ปกติที่ส่งผลกระทบต่อสภาพแวดล้อมให้มีการเปลี่ยนแปลง เช่น สถานการณ์การแพร่ระบาดของ Covid-19

## 1.3 ขอบเขตการศึกษา

การดำเนินการศึกษาและวิจัยครั้งนี้ ได้นำข้อมูลด้านอุตุนิยมวิทยา (Meteorological Dataset) ในเขตพื้นที่กรุงเทพมหานคร มาใช้ในการศึกษาวิจัย โดยมีขอบเขตงานวิจัยดังนี้



ภาพ 1.1 ขอบเขตพื้นที่การศึกษา

### 1.3.1 ข้อมูลที่ใช้ในการวิเคราะห์ผลการศึกษา

ชุดข้อมูลที่นำมาใช้เป็นข้อมูลทุติยภูมิ (Secondary Data) จาก 2 แหล่งข้อมูล คือ ข้อมูลมลพิษจากจุดตรวจวัดที่ดำเนินการโดยกรมควบคุมมลพิษ ทั้งหมด 13 จุด (วงกลมสีเขียวในภาพ 1) และข้อมูลมลพิษจากจุดตรวจวัดที่ดำเนินการโดยกรุงเทพมหานคร ทั้งหมด 50 จุด (วงกลมสีแดงในภาพ 1) ซึ่งติดตั้งจุดตรวจวัดข้อมูลอุตุนิยมวิทยาตามมาตรฐานในประกาศกรมควบคุมมลพิษ (กรมควบคุมมลพิษ; คู่มือการควบคุมและประกันคุณภาพงานห้องปฏิบัติการ สิ่งแวดล้อม) (เรื่อง เครื่องวัดและวิธีตรวจวัดค่าเฉลี่ยของก๊าซหรือฝุ่นละอองในบรรยากาศโดยทั่วไป ระบบอื่นหรือวิธีอื่นที่กรมควบคุมมลพิษเห็นชอบ) ของกรมควบคุมมลพิษ และติดตั้งแบบถาวร ด้วยความสูงจากพื้นดิน 2 เมตร ชุดข้อมูลอุตุนิยมวิทยาที่ได้รับประกอบด้วย PM 2.5 ฝุ่นละอองขนาดเล็กไม่เกิน 2.5 ไมครอน (PM 2.5) ฝุ่นหยาบ (Course Particle) (PM 10) NO<sub>x</sub> (Nitrogen Oxide) ไนโตรเจนไดออกไซด์ (NO<sub>2</sub>) ไนตริกออกไซด์ (NO) คาร์บอนมอนอกไซด์ (CO) ความเร็วลม (WS, Wind Speed) ทิศทางลม (WD, Wind Direction) อุณหภูมิ (Temp) ความกดอากาศ (BP) ความชื้นสัมพัทธ์ (RH%, Relative Humidity) และโอโซน (O<sub>3</sub>) โดยเก็บข้อมูลที่ความละเอียดทุก ๆ 1 ชั่วโมง โดยผู้วิจัยได้รวบรวมข้อมูลตั้งแต่ปี 2015 – 2021 ทั้งหมด 1.6 ล้านข้อมูล และทำการ cleaning ข้อมูลจนได้ชุดข้อมูลอุตุนิยมวิทยาที่พร้อมสำหรับนำไปวิเคราะห์ต่อไป ทั้งนี้ ผู้วิจัยได้นำข้อมูลที่ 99<sup>th</sup> percentile ของข้อมูลดังกล่าวมาใช้ในการศึกษาครั้งนี้ อย่างไรก็ตาม ชุดข้อมูลอุตุนิยมวิทยา ไม่เพียงแต่บอกแหล่งกำเนิดของการปล่อยมลพิษที่มีอิทธิพลต่อคุณภาพอากาศเท่านั้น แต่ต้นเหตุปัจจัยของข้อมูลทางอุตุนิยมวิทยาที่ได้มานั้น ยังส่งผลอย่างมีนัยสำคัญต่อการเจือจางและการสะสมของสารมลพิษที่ปล่อยออกมา จากแหล่งต้นกำเนิดในท้องถิ่นนั้น ๆ ด้วย (Miao et al., 2019; Wetchayont, 2021)

### 1.3.2 การศึกษาค่ามลพิษทางการอากาศที่เปลี่ยนแปลง ตั้งแต่ พ.ศ. 2558 – 2564

- 1) มุ่งเน้นศึกษาข้อมูลการเกิด PM 2.5 ในช่วงสถานการณ์และเวลาที่แตกต่างกันไปในแต่ละฤดูกาล โดยใช้ข้อมูลจากปี พ.ศ. 2558 – 2564 ซึ่งข้อมูลที่นำมาใช้ในการวิเคราะห์ จะครอบคลุมข้อมูลในช่วงที่มีการแพร่ระบาดของ Covid-19 และช่วงก่อนหน้านั้น ที่ยังไม่มีมีการแพร่ระบาดของ Covid-19
- 2) วิเคราะห์ และเปรียบเทียบการเกิด PM 2.5 ในช่วงเวลาต่าง ๆ ได้แก่ เวลาในช่วงวัน วันที่ในช่วงเดือน เดือนในแต่ละปี

### 1.3.3 การพัฒนาแบบจำลอง Machine Learning ที่ใช้ในการคาดการณ์ผลการศึกษา

การศึกษาและวิจัยนี้ จะศึกษาและใช้ข้อมูลอัลกอริทึมของ Machine Learning จำนวนทั้งหมด 6 แบบจำลอง เพื่อเปรียบเทียบและตรวจสอบความถูกต้องของแบบจำลอง ได้แก่

- 1) Decision Tree Regressor CART
- 2) Gradient Boosting Regression
- 3) K-neighbor Regressor
- 4) MLP Regressor
- 5) Random Forest Regressor
- 6) Ridge Model

### 1.3.4 ชุดข้อมูลที่ใช้สำหรับการพัฒนาแบบจำลอง Machine Learning

ชุดข้อมูลที่ใช้ในการพัฒนาแบบจำลอง แบ่งได้เป็น 2 ส่วนหลัก คือ ส่วนที่นำไปใช้ในการพัฒนาแบบจำลอง ซึ่งเรียกว่า Training set และส่วนที่ใช้ในการทดสอบแบบจำลอง เรียกว่า Test set

อย่างไรก็ตาม ในส่วนของ Test set ได้แบ่งข้อมูลในช่วงที่มีการแพร่ระบาดของ Covid-19 ออกมา เพื่อตรวจสอบความแม่นยำของแบบจำลองในช่วงสถานการณ์ไม่ปกติ โดยชุดข้อมูลนี้เรียกว่า Testout set โดยมีรายละเอียดการแบ่งชุดข้อมูล ดังนี้

- 3) Training set : ข้อมูลมลพิษทางอากาศ และข้อมูลด้านอุตุนิยมวิทยา ตั้งแต่ พ.ศ. 2558 – 2562 จำนวน ร้อยละ 80 ของชุดข้อมูลทั้งหมด (ก่อนการระบาดของ Covid-19)
- 4) Test set: ข้อมูลมลพิษทางอากาศ และข้อมูลด้านอุตุนิยมวิทยา ตั้งแต่ พ.ศ. 2558 – 2562 จำนวน ร้อยละ 20 ของชุดข้อมูลทั้งหมด (ก่อนการระบาดของ Covid-19)
- 5) Testout set: ข้อมูลมลพิษทางอากาศ และข้อมูลด้านอุตุนิยมวิทยา ตั้งแต่ พ.ศ. 2563 – 2564 จำนวน ร้อยละ 100 ของชุดข้อมูลทั้งหมด (เมื่อมีการแพร่ระบาดของ Covid-19)

## 1.4 สมมติฐาน

การเปลี่ยนแปลงการใช้ชีวิตของประชาชนในกรุงเทพมหานครที่ลดลงไป (Human Activities) ส่งผลต่อค่าความเข้มข้นของ PM 2.5 โดยในห้วงที่มีการแพร่ระบาดของ Covid-19 ทำให้ค่า PM 2.5 ลดลงไปอย่างมีนัยสำคัญ เนื่องมาจากการเดินทางที่ลดลง โดยพิจารณาอ้างอิงจากผลลัพธ์การศึกษาระดับนานาชาติที่ผ่านมา

เมื่อนำข้อมูลด้านมลพิษทางอากาศและข้อมูลด้านอุตุนิยมวิทยาของช่วงก่อนการระบาด Covid-19 และช่วงที่เกิด Covid-19 ไปทำนายค่า PM 2.5 ในอนาคต จะมีแบบจำลองที่น่าเชื่อถือไม่น้อยกว่าร้อยละ 80 โดยพิจารณาอ้างอิงจากผลลัพธ์การศึกษาระดับนานาชาติที่ผ่านมา

## 1.5 กรอบแนวความคิด

ในการศึกษาวิจัยค่าความเข้มข้นของ PM 2.5 ที่เปลี่ยนแปลงไปในช่วงก่อนและระหว่างเกิดการแพร่ระบาดของ Covid-19 และการพัฒนาแบบจำลองคณิตศาสตร์ ด้วยการประยุกต์ใช้ Machine Learning ในการคาดการณ์ PM 2.5 ที่จะเกิดขึ้นในอนาคต พบว่า ในงานวิจัยที่ผ่านมา Doreswamy et al. (2020) ได้นำเสนอวิธีการและขั้นตอนที่มีลักษณะในการศึกษาและวิเคราะห์ผลการศึกษาคู่กัน โดยในงานวิจัยนี้จะนำกรอบแนวความคิดดังกล่าวมากำหนดขั้นตอนให้สอดคล้องต่อแนวทางการวิจัย 4 ขั้นตอน ประกอบไปด้วย

### 1.5.1 กระบวนการรวบรวมและคัดกรองข้อมูลจากแหล่งข้อมูล

โดยปกติงานวิจัยด้านอุตุนิยมวิทยาและมลพิษทางอากาศจะใช้ข้อมูลทุติยภูมิในการศึกษาเป็นหลัก โดยแต่ละประเทศจะมีหน่วยงานที่คอยจัดเก็บสถิติข้อมูลด้านอุตุนิยมวิทยาและค่ามลพิษทางอากาศและสามารถเผยแพร่แก่ผู้ที่นำไปใช้เพื่อสาธารณประโยชน์ต่อไปได้ โดยการรวบรวมข้อมูลและนำข้อมูลมาใช้ในการศึกษาและวิจัยนี้ ได้รับข้อมูลสถิติของกรมควบคุมมลพิษของประเทศไทย (กรมควบคุมมลพิษ, 2564) และ สำนักสิ่งแวดล้อม กรุงเทพมหานคร ทั้งหมด 63 สถานีตรวจวัดในรูปแบบไฟล์ Excel เพื่อนำมาใช้ในการศึกษาและวิจัยต่อไป

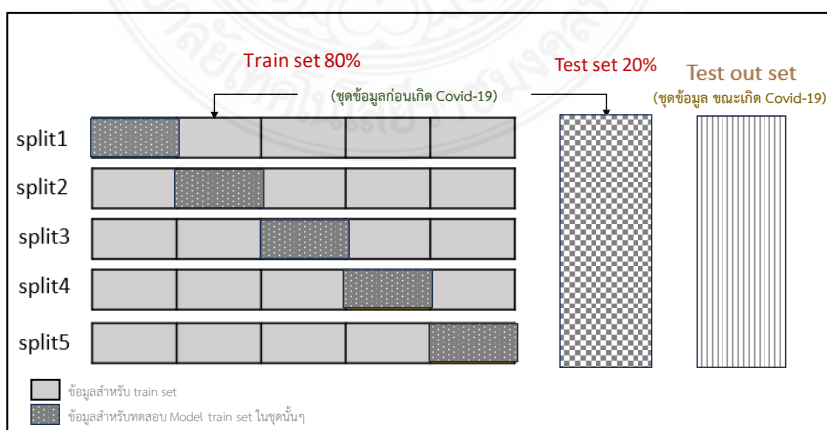
### 1.5.2 กระบวนการวิเคราะห์พฤติกรรมข้อมูลตามระยะเวลาที่เปลี่ยนแปลงไป (Time Series Analysis)

การวิเคราะห์อนุกรมเวลา (Time Series Analysis) เป็นกระบวนการที่ใช้ในการวิเคราะห์ข้อมูลที่มีลักษณะเป็นอนุกรมเวลา หรือการสร้างกราฟเพื่อตรวจสอบแนวโน้มของข้อมูล และใช้ในการจำแนกและพยากรณ์พฤติกรรมของชุดข้อมูลนั้นๆ อย่างชัดเจนมากยิ่งขึ้น ซึ่งการวิเคราะห์อนุกรมเวลาช่วยในการทำนายค่าที่อาจจะเปลี่ยนแปลงในอนาคต โดยการสังเกตพฤติกรรมและการเคลื่อนไหวของข้อมูลในอดีต

การวิเคราะห์ในกรณีที่มีชุดข้อมูลเป็นจำนวนมาก จะนิยมใช้โปรแกรมที่ซับซ้อนมากยิ่งขึ้น โดยงานวิจัยนี้มีชุดข้อมูลที่มากถึงล้านชุดข้อมูล ผู้วิจัยจึงใช้โปรแกรม R และ Python ช่วยในการวิเคราะห์ ทั้งนี้ ยังคงใช้หลักการตามทฤษฎีของการวิเคราะห์ Time Series Analysis ในการวิจัย

### 1.5.3 กระบวนการจัดเตรียมข้อมูลในการจัดเตรียมแบบจำลองคณิตศาสตร์

ในการจัดเตรียมชุดข้อมูลก่อนการจัดทำแบบจำลอง ชุดข้อมูลด้านอุตุนิยมวิทยาและมลพิษทางอากาศที่ได้รับมาจากหน่วยงานที่เกี่ยวข้อง จะถูกนำไปคัดกรองเอาข้อมูลที่ใช้งานไม่ได้ออก เช่น ข้อมูลซ้ำ ข้อมูลที่ผิดพลาด เป็นต้น เพื่อนำข้อมูลที่สมบูรณ์มาเตรียมเข้าแบบจำลองต่อไป โดยชุดข้อมูลจะถูกนำมาแบ่งออกตามชุดการทดสอบที่เตรียมไว้แสดงดังภาพ 1.2 และแสดงกระบวนการนำข้อมูลชุดต่าง ๆ เข้าแบบจำลอง ดังแสดงในตาราง 1.1



ภาพ 1.2 กระบวนการนำข้อมูลชุดต่าง ๆ เข้าแบบจำลอง

ตาราง 1.1 ชุดข้อมูลด้านอุตุนิยมวิทยาและมลพิษทางอากาศแบ่งตามชุดการทดสอบ

| ชุดที่ | ชื่อชุดข้อมูล | คำอธิบาย   |
|--------|---------------|--|
| 1      | Training set  | ข้อมูลมลพิษทางอากาศ และข้อมูลด้านอุตุนิยมวิทยา ตั้งแต่ พ.ศ. 2558 – 2562 จำนวน ร้อยละ 80 ของชุดข้อมูลทั้งหมด (ก่อนการระบาดของ Covid-19)         |
| 2      | Test set      | ข้อมูลมลพิษทางอากาศ และข้อมูลด้านอุตุนิยมวิทยา ตั้งแต่ พ.ศ. 2558 – 2562 จำนวน ร้อยละ 20 ของชุดข้อมูลทั้งหมด (ก่อนการระบาดของ Covid-19)         |
| 3      | Testout set   | ข้อมูลมลพิษทางอากาศ และข้อมูลด้านอุตุนิยมวิทยา ตั้งแต่ พ.ศ. 2563 – 2564 จำนวน ร้อยละ 100 ของชุดข้อมูลทั้งหมด (เมื่อมีการแพร่ระบาดของ Covid-19) |

#### 1.5.4 กระบวนการประยุกต์ใช้ Machine Learning ในการคาดการณ์ PM 2.5

กระบวนการใช้แบบจำลองในการคาดการณ์ PM 2.5 ของงานศึกษาวิจัยชิ้นนี้ คือ การเรียนรู้โดยมีข้อมูล (Data) มาสอน (Train) โดยการทำให้คอมพิวเตอร์สามารถหาคำตอบของปัญหาได้ด้วยตัวเอง หลังจากการเรียนรู้จากข้อมูลตัวอย่างที่ส่งเข้าไปแล้วระยะหนึ่ง โดยผู้วิจัยจะทำการสร้างแบบจำลอง (Model) ที่จะทำให้คอมพิวเตอร์รู้จักข้อมูลอุตุนิยมวิทยา และมลพิษทางอากาศในแต่ละปีเข้าไป (ปัจจัยที่ก่อให้เกิดมลพิษทางอากาศ) โดยแปลงให้เป็นภาษาคอมพิวเตอร์ (เป็นตัวเลข) หรือที่เรียกว่า Features พร้อมทำเฉลยไว้ให้โปรแกรม ว่านี่คือ PM 2.5 PM 10 O3 เป็นต้น (เป็นตัวเลข) เรียกว่า labels (โดยแบบจำลองของงานวิจัยนี้ จะสนใจเฉพาะค่าความเข้มข้นของ PM 2.5 ที่เกิดขึ้นเท่านั้น) เมื่อใส่ข้อมูลตั้งต้น (Input data) แล้วเสร็จก็จะทำการสอนแบบจำลอง (Train Model) เพื่อให้คอมพิวเตอร์แยกแยะค่าความเข้มข้นของ PM 2.5 ที่จะเกิดขึ้นได้ตามข้อมูล Features ที่ใส่เข้าไป

#### 1.5.5 กระบวนการทดสอบประสิทธิภาพแบบจำลองที่ใช้ในการคาดการณ์ PM 2.5

แบบจำลองจะมีค่าความคลาดเคลื่อนเกิดขึ้น โดยที่ความถูกต้องของค่าที่ทำนายได้จะมากหรือน้อยขึ้นอยู่กับค่าความคลาดเคลื่อนของการทำนาย ซึ่งสามารถตรวจสอบความถูกต้องและความแม่นยำของการทำนายได้จากหลักเกณฑ์ที่ใช้ตรวจสอบ (Performance Criteria) คือ  $R^2$  RMSE และ MAPE

### 1.6 ประโยชน์ที่คาดว่าจะได้รับ

- 1.6.1 ทำให้ทราบถึงพฤติกรรมการเกิดมลพิษทางอากาศในพื้นที่กรุงเทพมหานคร
- 1.6.2 ทำให้ทราบถึงการเปลี่ยนแปลงความเข้มข้นของ PM 2.5 เมื่อเกิดการแพร่ระบาดของ Covid-19
- 1.6.3 มีความรู้และเข้าใจในกระบวนการพัฒนาแบบจำลองที่เกี่ยวข้องกับการทำนายค่าความเข้มข้นของ PM 2.5
- 1.6.4 สามารถพยากรณ์ค่าความเข้มข้นของ PM 2.5 ได้แม่นยำมากขึ้น



- 1.6.5 ผลการศึกษาที่ได้จากการคาดการณ์ สามารถนำไปวางแผนการจัดการด้านสิ่งแวดล้อม
- 1.6.6 ช่วยให้หน่วยงานที่เกี่ยวข้องดำเนินการแก้ไขปัญหาสิ่งแวดล้อมได้อย่างถูกต้องและเหมาะสม

## 1.7 นิยามศัพท์

### 1.7.1 อนุภาคที่ลอยอยู่ในอากาศ หรือ Particulate Matter (PM)

อนุภาคที่ลอยอยู่ในอากาศ หรือ Particulate Matter (PM) คืออนุภาคเล็กที่ประกอบด้วยของแข็งและของเหลวที่ลอยอยู่ในอากาศที่เราหายใจเข้าไป ซึ่งมีความสำคัญในประเด็นคุณภาพอากาศและสุขภาพมนุษย์ เช่น PM<sub>2.5</sub> หมายถึง อนุภาคขนาดเล็ก 2.5 ไมครอน (micrometers) หรือ 2.5 ล้านหนึ่งมิลลิเมตร (micrometers) ซึ่งเป็นหนึ่งใน class ของอนุภาคขนาดเล็กที่อาจประสิทธิภาพที่สุดต่อสุขภาพของมนุษย์ และ PM<sub>10</sub> หมายถึงฝุ่นละอองขนาดเล็กกว่า 10 ไมครอน เป็นต้น อนุภาค PM สามารถมีต้นทางมาจากหลายแหล่ง เช่น การเผาไหม้, การจราจร, การผลิตอุตสาหกรรม, และกิจกรรมมนุษย์อื่น ๆ อนุภาคที่มีขนาดเล็กสามารถกระจายออกในอากาศและส่งผลกระทบต่อคุณภาพอากาศและสุขภาพมนุษย์เมื่อถูกหายใจเข้าไป (คู่มือปฏิบัติการในการป้องกันและแก้ไขปัญหาฝุ่นละออง PM<sub>2.5</sub> ระดับจังหวัด, 2562)

### 1.7.2 โควิด-19 (Covid-19)

คือ โรคติดต่ออุบัติใหม่ที่เกิดจากการติดเชื้อไวรัสโคโรนาสายพันธุ์ใหม่ชื่อ SARS-CoV-2 (Severe Acute Respiratory Syndrome-Coronavirus-2) Wu et al. (2020) พบการแพร่ระบาดครั้งแรกในเดือนธันวาคมปี 2019 ที่เมืองอู่ฮั่น ประเทศจีน ก่อนที่จะแพร่ระบาดไปทั่วโลก และส่งผลกระทบต่อทั้งทางเศรษฐกิจและทางด้านสาธารณสุขต่อเนื่องมาจนถึงปัจจุบัน

### 1.7.3 AI หรือ ปัญญาประดิษฐ์ (Artificial Intelligence: AI)

คือ เครื่องจักร (machine) ที่มีฟังก์ชันที่มีความสามารถในการทำความเข้าใจ เรียนรู้องค์ความรู้ต่างๆ อาทิเช่น การรับรู้ การเรียนรู้ การให้เหตุผล และการแก้ปัญหาต่างๆ เครื่องจักรที่มีความสามารถเหล่านี้ก็ถือว่าเป็นปัญญาประดิษฐ์ (Stuart and Peter, 2010; Winston, 1992) จึงอาจกล่าวได้ว่า AI ถือกำเนิดขึ้นเมื่อเครื่องจักรมีความสามารถที่จะเข้าใจและเรียนรู้ได้เอง โดย AI แบ่งออกเป็นหลายระดับตามความฉลาดและความสามารถ โดยจะวัดประสิทธิภาพของการใช้เหตุผล และทัศนคติของ AI เปรียบเทียบกับความสามารถและความฉลาดมนุษย์ เป็นต้น

### 1.7.4 Machine Learning

คือ การเรียนรู้ของเครื่องมือที่ถูกใช้งานตั้งเป็นสมองของ AI หมายถึงการที่ AI ใช้ Machine Learning สร้างความฉลาด ซึ่งส่วนใหญ่จะใช้เรียกแบบจำลองที่ได้มาจากการเรียนรู้ของปัญญาประดิษฐ์ และไม่ได้มาจากการเขียนโปรแกรมโดยมนุษย์ ทั้งนี้ มนุษย์มีหน้าที่เพียงแค่เขียนโปรแกรมให้ AI เรียนรู้จากข้อมูลเท่านั้น

Machine Learning (Biggio and Roli, 2018; Molnar, 2020) เรียนรู้ข้อมูลจากสิ่งที่ส่งเข้าไปกระตุ้น และจำไว้ในส่วนของสมอง จากนั้นส่งผลลัพธ์ออกมาเป็นตัวเลข หรือ code และส่งต่อไปแสดงผลหรือให้ AI แสดงการตอบสนอง Machine Learning โดยสามารถใช้งานได้หลายรูปแบบ แต่ต้องอาศัยกลไกที่เป็นโปรแกรม หรือ Algorithm ที่นักวิเคราะห์ข้อมูลเป็นผู้ออกแบบ โดย Algorithm ที่ได้รับความนิยมสูง คือ Deep Learning (Goodfellow et al., 2016; Zhang et al., 2021) เพราะสามารถใช้งานง่าย และประยุกต์ใช้ได้หลายด้าน เช่น ด้านการจราจรและขนส่ง ด้านสุขภาพ เป็นต้น ทั้งนี้ ในการทำงานของนักวิเคราะห์ข้อมูล ยังคงต้องออกแบบตัวแปร (Feature) ต่าง ๆ หา Algorithm อื่นๆ มาใช้ในการประยุกต์ใช้ Deep Learning โดยทำการเปรียบเทียบ เพื่อหา Algorithm ในการทำงานให้มีความเหมาะสมที่สุด



## บทที่ 2

### แนวคิด ทฤษฎี และการทบทวนวรรณกรรม

#### 2.1 สถานการณ์และผลกระทบของ PM 2.5

##### 2.1.1 ข้อมูลทั่วไปเกี่ยวกับฝุ่นละออง

ฝุ่นละออง หมายถึง อนุภาคของของเหลวและของแข็งที่ลอยลอยอยู่ในอากาศ มีทั้งอนุภาคขนาดเล็กและขนาดใหญ่ ทั้งที่ไม่สามารถมองเห็นได้ด้วยตาเปล่าและมองเห็นได้ด้วยตาเปล่า ซึ่งอนุภาคเหล่านี้จะรวมตัวกันเป็นมวลสารละเอียดแต่มีขนาดเบา จึงทำให้รวมตัวกันลอยอยู่ในชั้นบรรยากาศได้ โดยทั่วไปฝุ่นละอองจะมีอนุภาคขนาดไม่เกิน 100 ไมครอน สามารถทำให้เกิดผลกระทบต่อสุขภาพและอนามัยของประชาชน หรือในบางครั้งก็มีความหนาแน่นจนบดบังทัศนวิสัย ซึ่งฝุ่นละอองดังกล่าวนี้ มีทั้งที่เกิดขึ้นเองตามธรรมชาติ และเป็นฝุ่นละอองที่เกิดจากกิจกรรมต่างๆ ของมนุษย์ ไม่ว่าจะเป็น การเผาไหม้เชื้อเพลิงจากการขนส่ง การอุตสาหกรรมต่างๆ การก่อสร้าง ตลอดจนการเผาในที่โล่ง ซึ่งกิจกรรมเหล่านี้ ล้วนเป็นตัวการที่ทำให้เกิด “อากาศเสีย” ที่ทำให้ชั้นบรรยากาศเต็มไปด้วย คาร์บอน ก๊าซ เขม่า ฝุ่นละออง โดยเฉพาะอย่างยิ่งฝุ่นละอองขนาดเล็กที่ลอยลอยและกระจายอยู่ทั่วไปในอากาศ (กรมควบคุมมลพิษ, 2554) เมื่อสิ่งมีชีวิตสูดหายใจนำอากาศเข้าไปในปอด ฝุ่นละอองก็จะวิ่งเข้าไปอยู่ภายในระบบทางเดินหายใจส่วนล่าง ซึ่งทำให้เกิดปัญหาสุขภาพมากมาย เช่น โรคหอบหืด โรคปอด โรคทางเดินหายใจ เป็นต้น โดยฝุ่นละอองขนาดเล็ก (Chulalongkorn University, 2020) หรือที่รู้จักกันในชื่อ PM ซึ่งย่อมาจาก Particulate Matters ซึ่งใช้เรียกค่าฝุ่นละอองขนาดเล็กที่เป็นอันตรายต่อสุขภาพอนามัย ในต่างประเทศจึงได้มีการกำหนดมาตรฐานค่าฝุ่นละอองขนาดเล็ก โดยหน่วยงาน United State Environmental Protection Agency หรือ U.S. EPA ของสหรัฐอเมริกา ได้กำหนดมาตรฐานของค่าฝุ่นละอองขนาดเล็กแบ่งเป็น 2 ชนิด ตามขนาดของฝุ่น ได้แก่ 1) ฝุ่นหยาบ (Course Particle) หรือฝุ่นละอองอนุภาคเล็กขนาดน้อยกว่า 10 ไมครอน (PM 10) เป็นอนุภาคฝุ่นที่มีขนาดเส้นผ่าศูนย์กลางอยู่ระหว่าง 2.5 - 10 ไมครอน ซึ่งฝุ่นละอองขนาดเล็กเหล่านี้หากรวมตัวกันเป็นจำนวนมากแล้ว มักสามารถมองเห็นได้ง่าย เช่น ฝุ่นที่เกาะอยู่ตามพื้นหรือของใช้ต่างๆ หรือฝุ่นจากการก่อสร้าง เป็นต้น มีแหล่งต้นกำเนิดจากการจราจรขนส่งบนท้องถนน ตลอดจนฝุ่นจากกิจกรรมต่าง ๆ (EPA, 2016) ส่งผลกระทบต่อสุขภาพได้เช่นกัน อย่างไรก็ตาม ฝุ่นหยาบมักจะส่งผลกระทบต่อสุขภาพมากนักเมื่อเทียบกับฝุ่นขนาดเล็ก (PM 2.5) แต่ฝุ่น PM 10 ยังสามารถเข้าสู่ระบบทางเดินหายใจและไปสะสมได้ (กรมควบคุมมลพิษ, 2562) เมื่อได้รับในปริมาณหนึ่งจะทำให้เกิดโรคหอบหืด (กรมอนามัยและกรมควบคุมโรค, 2558) 2) ฝุ่นละเอียด (Fine Particle) หรือฝุ่นละอองอนุภาคขนาดเล็กกว่า 2.5 ไมครอน (PM 2.5) เป็นอนุภาคฝุ่นที่มีเส้นผ่าศูนย์กลางขนาดเล็กกว่า 2.5 ไมครอน หรือมีขนาดประมาณ 1 ใน 25 ส่วนของเส้นผ่าศูนย์กลางของขนาดเส้นผม ซึ่งขนจมูกไม่สามารถกรองได้ มีแหล่งต้นกำเนิดมาจากแหล่งต่างๆ เช่น คาร์บอนเสียของรถยนต์ และโรงงานอุตสาหกรรม นอกจากนี้ ก๊าซบางชนิด เช่น ก๊าซซัลเฟอร์ไดออกไซด์ (SO<sub>2</sub>) และสารอินทรีย์ระเหยง่าย (VOCs) สามารถทำปฏิกิริยากับสารอื่นๆ ในอากาศ กลายเป็นฝุ่น PM 2.5 ได้อีกด้วย (EPA, 2016)



ภาพ 2.1 แสดงขนาดของฝุ่น PM 2.5

ที่มา : EPA (2016)

ฝุ่น PM 2.5 สามารถเข้าสู่ระบบทางเดินหายใจ และลงไปถึงถุงลมในปอดได้ เป็นสาเหตุให้เกิดโรคที่เกี่ยวข้องกับระบบทางเดินหายใจ เมื่อรับฝุ่น PM2.5 เข้าไปในปริมาณมากและสะสมเป็นเวลานานจะส่งผลให้ปอดทำงานเสื่อมประสิทธิภาพ มีอาการหอบหืด และหลอดลมอักเสบได้ (กรมควบคุมมลพิษ, 2562) โดยฝุ่น PM 2.5 สามารถล่องลอยและกระจายอยู่ในอากาศได้เป็นเวลานาน ซึ่งหากมีปริมาณมากในอากาศ จะทำให้เห็นท้องฟ้าเป็นสีเทา หรือเกิดเป็นกลุ่มหมอกควัน ซึ่งจะเห็นได้ในเมืองใหญ่ที่มีประชากรและปริมาณจราจรหนาแน่น

### 2.1.2 ฝุ่น Particulate matter 2.5 (PM 2.5) ในอากาศ

อย่างไรก็ตาม ฝุ่นละออง (PM) เป็นส่วนผสมที่ต่างกันของอนุภาคซึ่งประกอบด้วยสิ่งสกปรก เขม่า ควัน และแม่แต่หยดของเหลว (Aerosols) ที่มาจากทั้งแหล่งธรรมชาติ เช่น ฝุ่นละอองเกสรดอกไม้ พืชพรรณต่าง ๆ และ ที่มาจากกิจกรรมของมนุษย์ เช่น ปล่องควันในโรงงาน เครื่องยนต์ดีเซล ตลอดจนการผลิตไฟฟ้า การก่อสร้าง การขุด และการเกษตร (Rosenlund et al., 2009; Sacks et al., 2011; Zhao et al., 2013) เนื่องจากอนุภาคของฝุ่น PM มีขนาดเล็กจนสามารถเจาะเข้าไปในปอดได้ง่าย จึงเพิ่มความเสี่ยงและความรุนแรงของการเกิดโรคหอบหืดได้ รวมไปถึงสามารถทำให้เกิดโรคหลอดลมอักเสบและโรคอื่น ๆ ที่เกี่ยวกับปอด ทั้งนี้ยังเข้าไปลดความสามารถของร่างกายในการต่อสู้กับเชื้อโรคหรือการติดเชื้อได้ (Trasande and Thurston, 2005) โดยสามารถสรุปผลกระทบด้านสุขภาพจากการได้รับ PM ในตาราง 2.1 ทั้งนี้ งานวิจัยในประเทศตุรกี Tecer et al. (2008) แสดงให้เห็นว่า ในช่วงเวลาที่มี PM สูง เด็ก ๆ มีความเสี่ยงสูงที่จะเป็นโรคภูมิแพ้ และหลอดลมอักเสบ ทั้งนี้ยังมีการศึกษาที่พบว่า PM สามารถดูดซับสารก่อภูมิแพ้ที่ปล่อยออกมาจากละอองเรณูของฝุ่น และพืชได้ จึงสามารถกักเก็บสารก่อภูมิแพ้ไว้เป็นเวลานานและส่งผลต่อคนที่สูดอนุภาคดังกล่าวเข้าไปในร่างกายได้โดยตรง (Zhou et al., 2019) ฝุ่นละอองขนาดเล็ก (PM 2.5) ในบรรยากาศเกิดจากการปล่อยอนุภาคมูลฐานและการเปลี่ยนแปลงของสารตั้งต้นของก๊าซ เช่น SO<sub>2</sub>, NO<sub>x</sub>, VOC<sub>s</sub> และ NH<sub>3</sub> ในการปล่อยแบบปฐมภูมิ (Delfino et al., 2008; Grundström et al., 2015; Kanakidou et al., 2005; Liu et al., 2015) ทั้งนี้ PM 2.5 ไม่เพียงแต่เต็มไปด้วยสารพิษและสารอันตรายจำนวนมาก แต่ยังสามารถอยู่ในบรรยากาศได้เป็นเวลานาน ยิ่งไปกว่านั้น PM 2.5 สามารถเคลื่อนย้ายได้ในระยะทางไกลผ่านการไหลเวียนของบรรยากาศซึ่งเป็นอันตรายต่อสุขภาพอย่างมาก (Harrison and Yin, 2000; Pöschl, 2005)

ตาราง 2.1 ผลกระทบต่อสุขภาพจากการสัมผัส Particulate matter (PM) ในอากาศ

| ความเข้มข้น<br>( $\mu\text{g}/\text{m}^3$ ) | ผลกระทบต่อสุขภาพ  |
|---|---|
| 10 $\mu\text{g}/\text{m}^3$ in same day     | มีการการรับเข้าโรงพยาบาลระบบทางเดินหายใจและหัวใจเพิ่มขึ้น 1.9% และ 3.3% ตามลำดับ (Delfino et al., 2008)             |
| 20 $\mu\text{g}/\text{m}^3$ in same day     | โรคทางเดินหายใจส่วนล่างในเด็กเพิ่มขึ้น 8% (Delfino et al., 2008)  |
| 120 $\mu\text{g}/\text{m}^3$                | มีจำนวนเด็กที่เป็นโรคทางเดินหายใจเข้าโรงพยาบาลเพิ่มขึ้น (Tecer et al., 2008)  |
| 150 $\mu\text{g}/\text{m}^3$                | พบโอกาสที่อาการโรคปอดและโรคทางเดินหายใจจะกำเริบเพิ่มมากขึ้น (Rosenlund et al., 2009)                                |
| 200 $\mu\text{g}/\text{m}^3$                | โรคเกี่ยวกับระบบทางเดินหายใจเพิ่มขึ้น เช่น โรคปอดบวม โรคปอดอุดกั้นเรื้อรัง (COPD) และโรคหอบหืด (Sacks et al., 2011) |
| 400 – 600 $\mu\text{g}/\text{m}^3$          | พบ Alveolar macrophages, neutrophils, and T-lymphocytes ใน Broncho alveolar lavage fluid (Maciejczyk et al., 2010)  |

### 2.1.3 สถานการณ์ฝุ่น PM 2.5 ในประเทศไทย

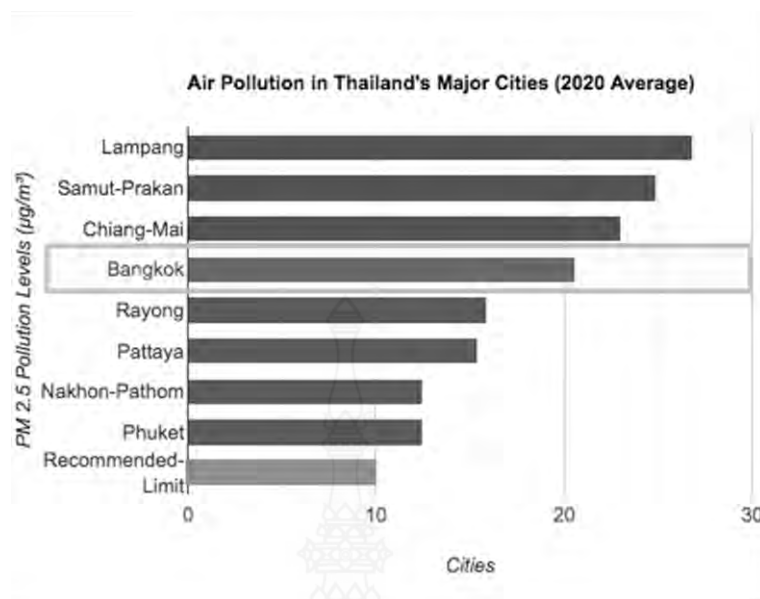
ปัญหาใหญ่ที่ท้าทายในปัจจุบัน คือ ปัญหาฝุ่นขนาดเล็กจิ๋ว หรือ PM 2.5 ที่ทำให้เกิดทัศนวิสัยไม่ดีในการมองเห็น โดยเฉพาะอย่างยิ่งในพื้นที่เขตเมือง เช่น กรุงเทพมหานคร รวมไปถึงในเขตปริมณฑล รวมถึงพื้นที่อีกหลายจังหวัด ได้แก่ เชียงราย เชียงใหม่ ลำปาง เป็นต้น

กรมควบคุมมลพิษ ได้ประมาณการว่า คนไทยอาจต้องเสียชีวิตเพราะฝุ่น PM 2.5 ในอากาศสูงถึง 22,000 คน (กรมควบคุมมลพิษ, 2564) และเมื่อดูค่า AQI ของประเทศไทยในช่วงต้นปีนี้ โดยสาเหตุที่ทำให้เกิดฝุ่น PM 2.5 ส่วนใหญ่มาจากกิจกรรมต่างๆ ของมนุษย์ ได้แก่

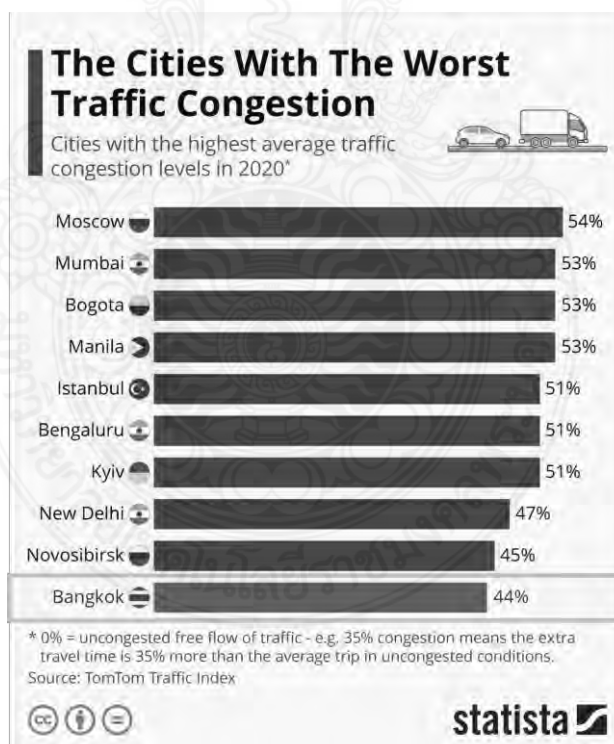
- 1) ไฟป่า/เผาป่า : ไฟป่าเกิดขึ้นเองตามธรรมชาติและจากฝีมือมนุษย์ โดยร้อยละ 90 ของไฟป่าในประเทศไทยเกิดจากการเผาโดยมนุษย์ ซึ่งเผาเพื่อหาของป่า และเพื่อทำการเกษตร
- 2) การก่อสร้าง : ฝุ่นที่เกิดจากการก่อสร้าง นับเป็นฝุ่นที่เป็นมลพิษอันดับต้นๆ ของบรรยากาศ โดยเฉพาะอย่างยิ่งในเมืองใหญ่ เช่น การสร้างรถไฟฟ้า และที่พักอาศัย
- 3) การผลิตไฟฟ้าและการทำอุตสาหกรรม : นิคมอุตสาหกรรมต่างๆ และการเผาถ่านหินเพื่อเป็นเชื้อเพลิงของการผลิตกระแสไฟฟ้า ทำให้เกิดควันและของเสียจากการเผาไหม้และปล่อยมลพิษออกมาในปริมาณมาก เช่น ก๊าซไนโตรเจนออกไซด์ ( $\text{NO}_x$ ) ก๊าซซัลเฟอร์ไดออกไซด์ ( $\text{SO}_2$ ) ก๊าซคาร์บอนมอนอกไซด์ (CO) รวมไปถึงเขม่าควัน และฝุ่นต่างๆ

- 4) การคมนาคม : คิวจากการเผาไหม้ที่ไม่สมบูรณ์ของเครื่องยนต์ ก็เป็นอีกแหล่งกำเนิดใหญ่ของ PM 2.5 ในเขตเมือง เช่น กรุงเทพฯ เชียงใหม่ เป็นต้น เพราะมีการเดินทางสูง รถติดหนัก จึงก่อให้เกิดการสะสมของฝุ่น PM 2.5 เป็นจำนวนมาก
- 5) กิจกรรมอื่นๆ ที่เป็นสาเหตุทำให้เกิดฝุ่นละอองในอากาศ เช่น การจตุรูบเทียน การสูบบุหรี่ การเผากระดาษ การจุดพลุ และการประกอบอาหาร เป็นต้น

นอกจากนี้ประเทศไทยยังติดอันดับ 5 ในประเทศอาเซียนที่มีค่า PM 2.5 มากที่สุด รองจากอินโดนีเซีย (อันดับ 1) พม่า (อันดับ 2) เวียดนาม (อันดับ 3) และลาว (อันดับ 4) ซึ่งในไทยนั้นพบว่า ลำปางเป็นเมืองที่มี PM 2.5 สูงที่สุดในปี 2021 ที่ผ่านมา อย่างไรก็ตาม ทั้งเมืองใหญ่อย่าง กรุงเทพฯ และเชียงใหม่ ก็เจอปัญหา PM 2.5 หนักมากตลอดในช่วงระยะเวลาที่ผ่านมา สาเหตุหลักๆ เกี่ยวกับปริมาณ PM 2.5 ในประเทศไทย จะมาจากมลภาวะและฝุ่นควันที่รุนแรงตามสภาพอากาศ โดยเฉพาะอย่างยิ่ง ในลุ่มน้ำเชียงใหม่ที่เป็นแหล่งการสะสมของฝุ่นละออง จากการเผาตอซังข้าว เผาป่า เผาหญ้า เพื่อเตรียมรับฝนและการปลูกข้าว ซึ่งพื้นที่ดังกล่าว เป็นหุบเขาแคบ ๆ ทำให้เกิดแอ่งที่เหมาะสมสำหรับมลพิษทางอากาศ (Amnuaylojaroen, 2022) จากการเปลี่ยนแปลงสภาพภูมิอากาศอย่างรวดเร็วและเฉียบพลัน รวมไปถึงอุณหภูมิโลกที่สูงขึ้นอย่างรวดเร็ว ส่งผลให้พื้นที่ชนบททางภาคเหนือของประเทศไทยมีความถี่และความรุนแรงของฝุ่นละอองขนาดเล็ก (PM 2.5) เพิ่มมากขึ้นทุกปี โดยเฉพาะอย่างยิ่งในช่วงฤดูแล้ง (Amnuaylojaroen, 2022; Chirasophon and Pochanart, 2020) ส่วนสาเหตุ PM 2.5 ในเมืองใหญ่ (Community Area) อย่างกรุงเทพมหานครนั้น พบว่าในช่วงฤดูหนาว จะมีมลพิษเข้มข้นสูงกว่าฤดูอื่น จากการศึกษาวิจัยก่อนหน้านี้ พบว่า ที่มาของการปล่อย PM 2.5 ที่พบมากที่สุดในการเผาไหม้ชีวมวล การจราจร และอุตสาหกรรม กิจกรรมต่าง ๆ ที่มีความเข้มข้นต่างกันไปตามปัจจัยของแต่ละฤดูกาล (Uttamang et al., 2018; Chirasophon and Pochanart, 2020; Narita et al., 2019) นอกจากนี้ ยังมีการศึกษาเกี่ยวกับความเข้มข้นของสารก่อมลพิษสำหรับ O<sub>3</sub>, NO<sub>x</sub>, CO และ SO<sub>2</sub> จากปี 1996 ถึงปี 2552 ในกรุงเทพมหานคร ระหว่างเขตที่อยู่อาศัยและริมถนน พบว่า ความเข้มข้นของมลพิษทางอากาศนั้น จะลดลงตั้งแต่เดือนมกราคมถึงสิงหาคม และจะเพิ่มขึ้นจากเดือนกันยายนถึงธันวาคม โดยความเข้มข้นของสารก่อมลพิษในอากาศที่เป็นก๊าซแสดงให้เห็นอย่างชัดเจนว่า มีความเข้มข้นสูงขึ้นระดับพื้นที่ริมถนนมากกว่าเขตที่อยู่อาศัย (Uttamang et al., 2018) ข้อมูลจาก Greenpeace ประเทศไทย (Greenpeace, 2016) พบว่า ในปี พ.ศ. 2562 กรุงเทพมหานครเป็นเมืองที่มีค่าเฉลี่ย PM 2.5 รายปีสูงเป็นอันดับ 4 ของประเทศ และกรุงเทพมหานครยังเมืองหลวงที่มีสภาพการจราจรที่แย่ที่สุดในโลกในอันดับที่ 10 ของโลก ดังแสดงรายการในภาพ 2.2 - 2.3



ภาพ 2.2 กรุงเทพมหานครเป็นเมืองที่มีค่าเฉลี่ย PM 2.5 รายปี  
สูงเป็นอันดับ 4 ของประเทศ  
ที่มา: Greenpeace (2016)



ภาพ 2.3 เมืองที่มีรถติดมากที่สุดในโลก ปี พ.ศ. 2562  
โดยกรุงเทพมหานครอยู่อันดับที่ 10  
ที่มา: Statista (2021)

#### 2.1.4 การปล่อยมลพิษที่เกี่ยวข้องกับภาคการขนส่งในประเทศไทย

โดยการขนส่งทางถนนเป็นที่มาของมลพิษทางอากาศ ก่อให้เกิด โอโซน CO, NO<sub>x</sub> และ NMVOCs ก๊าซเรือนกระจก CO<sub>2</sub>, CH<sub>4</sub> และ N<sub>2</sub>O สารที่เป็นกรด NH<sub>3</sub> และ SO<sub>2</sub> อนุภาคสารก่อมะเร็ง เช่น PAHs และ POPs รวมไปถึงพวกโลหะหนักที่เป็นพิษอีกด้วย จากคู่มือการปล่อยมลพิษของ European Monitoring and Evaluation Program (EMEP)/ (EEA) EPA. (2013) ซึ่งเป็นคู่มือทางเทคนิคในการพัฒนารายการการปล่อยก๊าซเรือนกระจก จากการขนส่งทางถนนที่ปล่อยออกมาจาก 4 แหล่ง ได้แก่ การปล่อยไอเสียที่เกิดจากกระบวนการเผาไหม้เชื้อเพลิงในเครื่องยนต์ของรถยนต์ การปล่อยไอระเหยที่เกิดจากการระเหยของน้ำมันเชื้อเพลิงในถังน้ำมันเชื้อเพลิง การปล่อยมลพิษจากยาง และการสึกหรอของเบรกที่เกิดจากกระบวนการสึกกร่อน และการปล่อยมลพิษจากพื้นผิวถนนที่เกิดจากการเสียดสีระหว่างยางกับพื้นถนน และสำหรับ การปล่อยมลพิษจากท่อไอเสีย (Exhaust) นั้น ประกอบไปด้วย 2 ประเภท คือ การปล่อยความร้อนที่ปล่อยออกมาจากยานพาหนะหลังจากที่เครื่องยนต์อุ่นเครื่องและในขั้นตอนการขับขึ้น และการปล่อยความเย็นที่ปล่อยออกมาจากยานพาหนะระหว่างขั้นตอนการอุ่นเครื่อง เป็นต้น

ตลอดระยะเวลาที่ผ่านมา การขนส่งและเดินทางมักจะเป็นสาเหตุหลักที่ก่อให้เกิด PM 2.5 ในพื้นที่เขตเมือง โดยเฉพาะอย่างยิ่ง ในเมืองหลวงของไทย (Chirasophon and Pochanart, 2020) ที่มีปริมาณจราจรสูงสุดติดอันดับโลกมาตลอดระยะเวลาหลายปี เช่นเดียวกันกับหลายพื้นที่ทั่วโลก มีงานวิจัยที่พบว่าภาคการขนส่งส่งผลต่อการเกิด PM 2.5 ในประเทศจีน ภาคการขนส่งมีบทบาทสำคัญต่อการเกิดมลพิษทางอากาศ ซึ่งมีลักษณะเฉพาะจากแหล่งกำเนิดมลพิษที่หลากหลายและการกระจายที่แตกต่างกัน โดยพบว่า กิจกรรมการขนส่งสินค้าระหว่างภูมิภาคต่าง ๆ the North China Plain (NCP), Yangtze River Delta (YRD), Pearl River Delta (PRD) และพื้นที่บริเวณ Chengyu นั้นมีความแตกต่างกันของมลพิษที่เกิดขึ้น โดยพบว่า ใน CNP จะมีความเข้มข้นของมลพิษตั้งต้นของ PM 2.5 สูงกว่าภูมิภาคอื่น แม้ว่ากิจกรรมการขนส่งต่างๆ จะเป็นกิจกรรมที่เกิดขึ้นภายในภูมิภาคเองก็ตาม ซึ่งส่งผลให้ทีมนักวิจัย จะต้องศึกษาวิจัยเพิ่มเติมต่อไปว่า ภาคการขนส่งที่เกิดขึ้นใน CNP นั้น มีปัจจัยและสาเหตุของการเกิด PM 2.5 มาจากการขนส่งจริงๆ หรือไม่ (Li et al., 2019) ยังมีการศึกษาถึงการเปลี่ยนแปลงรูปแบบการขนส่งสินค้าในประเทศจีน โดยจากการศึกษาดังกล่าว พบว่า การเปลี่ยนรูปแบบการขนส่งสินค้าที่เปลี่ยนรูปแบบจากการขนส่งสินค้าทางถนนไปสู่ระบบราง ช่วยลดความเข้มข้น PM 2.5 ที่เกิดขึ้นในภูมิภาคตอนกลางและตะวันตกของประเทศอีกด้วย (Wang and Yang, 2018) เช่นเดียวกันกับการศึกษาวิจัยเรื่องการเปลี่ยนรูปแบบการเดินทางโดยใช้แบบจำลองใน เมือง Adelaide ทางตอนใต้ของประเทศออสเตรเลีย พบว่า หากในอนาคต (2030) มีการปรับลดระยะการเดินทาง (vehicle kilometers travelled, VKT) ด้วยรถลง 40% สู่อุปกรณ์การเดินทางด้วยรูปแบบการเดินทางอื่นๆ ปรากฏว่าค่า PM 2.5 ในเขตเมืองเฉลี่ย จะลดลงไปได้ถึง 0.4 µg/m<sup>3</sup> (Xia et al., 2015) อย่างไรก็ตาม มีการการศึกษาถึงองค์ประกอบของมลพิษ (trace elements) ต่างๆ ที่เกิดขึ้นในช่วงชั่วโมงเร่งด่วน (Rush hour) เปรียบเทียบกับตัวอย่างในวันทำงาน (work-week) ในเมือง Pittsburgh ของประเทศสหรัฐอเมริกา พบว่า ในช่วงชั่วโมงเร่งด่วนจะมีค่าความเข้มข้นสูงกว่าในช่วงปกติ เช่น As, Se เป็นต้น ทั้งวันทำงานและวันหยุด (Tunno et al., 2018) จากงานวิจัยข้างต้น จะเห็นว่า การขนส่ง โดยเฉพาะทางถนน ที่ยิ่งมีความหนาแน่นมากเท่าไร ก็ยิ่งส่งผลต่อการเกิด PM 2.5 มากเท่านั้น



### 2.1.5 การแพร่ระบาดของโรค Covid-19 ในประเทศไทย

เมื่อไม่นานมานี้ เกิดเหตุการณ์ในลักษณะ Black Swan ซึ่งไม่มีใครคาดคิดมาก่อน ด้วยสถานการณ์การระบาดของโรค Covid-19 ซึ่งทำให้กิจกรรมต่าง ๆ บนโลกใบนี้แทบจะหยุดชะงัก โดยเฉพาะอย่างยิ่ง การเดินทางระหว่างทวีป ระหว่างภูมิภาค ระหว่างประเทศ หรือแม้แต่ระหว่างเมือง เพื่อหยุดยั้งการระบาดของโรคนี้อีก

ในเดือนธันวาคม 2019 ในเมืองอู่ฮั่น, ประเทศจีน พบกลุ่มผู้ป่วยที่มีอาการปอดบวมโดยไม่ทราบที่มา “pneumonia of unknown origin” ถูกรายงานซึ่งได้รับการยืนยันในภายหลัง (Varrassi et al., 2019) และตั้งชื่อว่า coronavirus disease 2019 (Covid-19) โดยองค์การอนามัยโลก (Varrassi et al., 2019) การระบาดใหญ่นี้นำไปสู่มาตรการด้านสาธารณสุขทั่วโลก เพื่อลดการแพร่กระจาย ในหลายประเทศ มาตรการดังกล่าวรวมถึงมาตรการเว้นระยะห่างทางกายภาพ เช่น การ “ล็อกดาวน์” สถาบันการศึกษา ร้านอาหาร บาร์ ร้านค้าปลีกและธุรกิจที่ไม่จำเป็นอื่นๆ การห้ามกิจกรรมสาธารณะ เช่น การแข่งขันกีฬา คอนเสิร์ต การแสดงละคร และการส่งเสริมหรือกำหนดให้ทำงานทางไกล จากบ้านและอยู่บ้าน (Vos, 2020)

สำหรับในประเทศไทย มีการเฝ้าระวังและติดตามสถานการณ์การแพร่ระบาดของโรค Covid-19 อย่างใกล้ชิดจากทุกภาคส่วนที่เกี่ยวข้อง จากรายงานของ Global Health Security Index, 2019 (GHS index, 2019) ระบุว่า ประเทศไทยถือว่าเป็นหนึ่งใน 10 อันดับแรกของประเทศที่มีการเตรียมพร้อมมากที่สุดและอยู่ในอันดับต้น ๆ ของประเทศที่มีรายได้ปานกลางระดับสูงโดยพิจารณาจากดัชนีความมั่นคงด้านสุขภาพโลกปี 2019 โดยประเทศไทยรายงานตัวเลขผู้ป่วยติดเชื้อรายแรกเมื่อวันที่ 12 มกราคม 2563 และหลังจากการแพร่ระบาดสูงสุดในช่วงเดือนมกราคม (188 รายต่อวัน) Covid-19 ถูกควบคุมได้ภายหลังจากมีการติดตามและกักตัวในสังคมอย่างเข้มงวด

ประเทศไทยได้ประกาศภาวะฉุกเฉินเมื่อวันที่ 26 มีนาคม 2563 และเริ่มมีการล็อกดาวน์บางส่วนทั้งในและนอกกรุงเทพฯ อย่างไรก็ตาม มีรายงานผู้ป่วยนอกเขตกรุงเทพฯ เพิ่มขึ้น หลังจากที่มีแรงงานจำนวนมากอพยพไปยังต่างจังหวัด แต่จำนวนผู้ป่วยยืนยันไม่เคยเกิน 200 รายต่อวัน แม้ว่าสถานการณ์การแพร่ระบาดดูเหมือนจะควบคุมได้ค่อนข้างดีในห้วงเวลานั้น แต่ก็ยังมีความไม่แน่นอนอยู่มากและมีความเสี่ยงที่จะเกิดการระบาดมากขึ้นในอนาคต ซึ่งทำให้ต้องมีการประกาศมาตรการในการจัดการแก้ไขปัญหาอย่างใกล้ชิด ทั้งนี้ ตลอดระยะเวลาที่เริ่มมีการระบาดในไทย (มกราคม 2562) จนถึงปัจจุบัน (สิงหาคม 2565) ประเทศไทยมีแนวทางการป้องกันและแก้ไขสถานการณ์การแพร่ระบาดของ Covid-19 ทั้งเชิงรุกและเชิงรับ จนสามารถเปิดประเทศให้มีการเดินทางข้ามพรมแดนได้อย่างเสรี พร้อมทั้งประกาศให้ Covid-19 เป็นโรคประจำถิ่น ทั้งนี้มาตรการต่าง ๆ จากทางภาครัฐตั้งแต่เริ่มพบผู้ติดเชื้อรายแรก จนถึงห้วงเวลาสิ้นปี พ.ศ. 2564 (สถานการณ์เริ่มผ่อนคลาย) มีรายละเอียดดังตาราง 2.2

ตาราง 2.2 มาตรการในการจัดการการแพร่ระบาดของ Covid-19 ของประเทศไทย

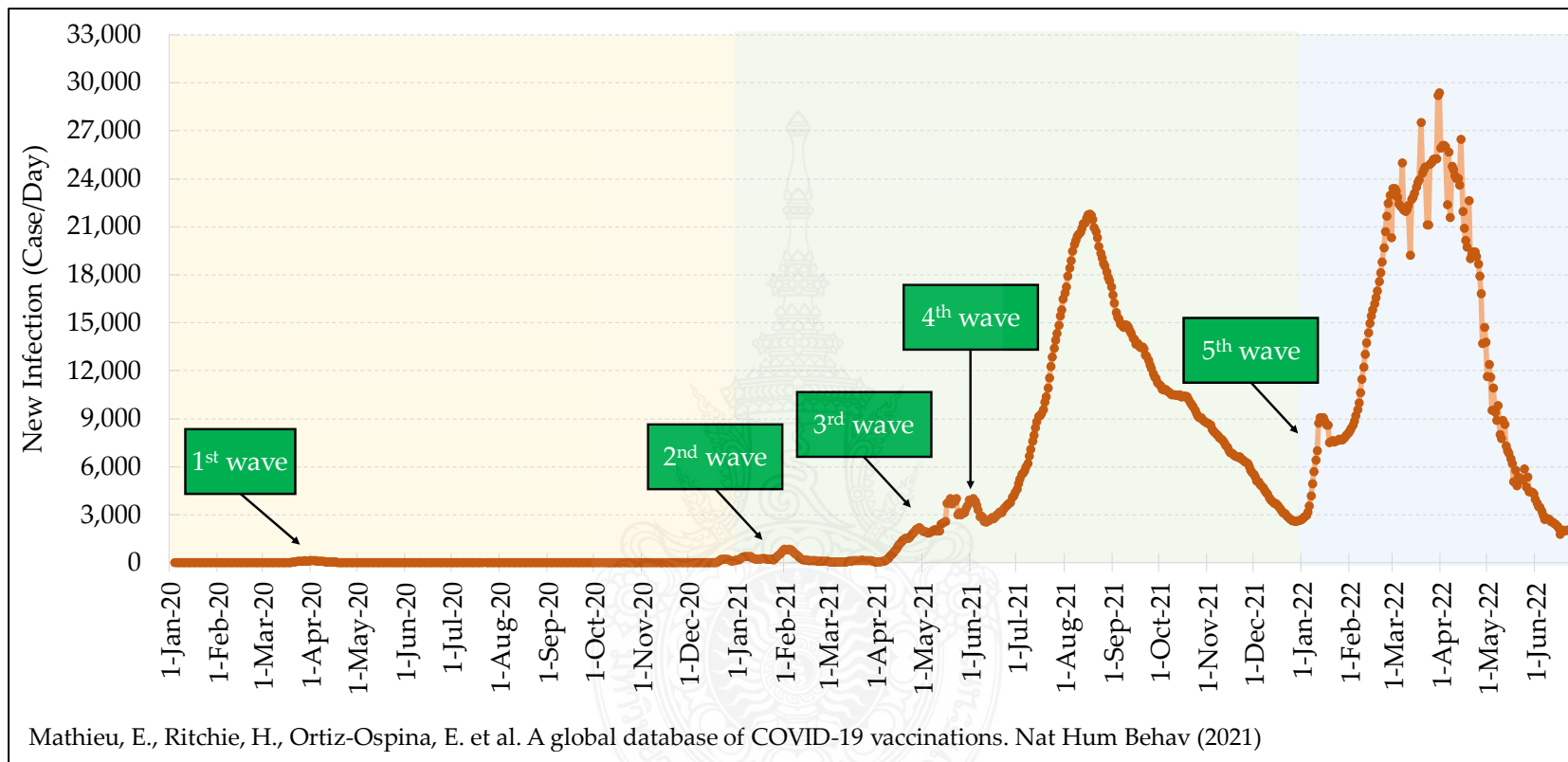
| ผลกระทบจากการแพร่ระบาดของ Covid-19  | จำนวนผู้ติดเชื้อใหม่ (Case/วัน) | มาตรการ  |
|---|---------------------------------|--|
| ระลอกที่ 1<br>(March 2020 – May 2020)<br><br>26 มีนาคม 63 – เมษายน 63         |                                 | <ul style="list-style-type: none"> <li>● การห้ามเข้าพื้นที่เสี่ยง ได้แก่ กทม.</li> <li>● Work from Home</li> <li>● ปิดห้างสรรพสินค้าให้ (เปิดเฉพาะศูนย์จำหน่ายอาหารที่ซื้อกลับบ้านเท่านั้น)</li> <li>● ตลาด/ตลาดนัด เปิดแค่จำหน่ายอาหารซื้อกลับบ้าน</li> <li>● ปิดพื้นที่ ที่มีคนพลุกพล่าน เช่น สถานที่แสดงมหรสพ สถานที่ที่มีการเล่นสากาธารณะ สถานที่จัดเลี้ยง จัดงานแต่ง เป็นต้น</li> <li>● ปิดมหาวิทยาลัยในพื้นที่ กทม. ให้เรียนออนไลน์</li> <li>● ปิดช่องทางเข้า-ออก ราชอาณาจักร</li> </ul>   |
| ระลอกที่ 2<br>(December 2020 – February 2021)<br><br>25 ธันวาคม 63- มกราคม 64 |                                 | <ul style="list-style-type: none"> <li>● การห้ามเข้าพื้นที่เสี่ยง ได้แก่ กทม.ที่ติดเขตจังหวัดสมุทรสาคร</li> <li>● Work from Home</li> <li>● จำกัดเวลาเปิด-ปิดห้างสรรพสินค้า(11.00-21.00 น.)</li> <li>● ตลาด/ตลาดนัด เปิดแค่จำหน่ายอาหารซื้อกลับบ้านหลัง 21.00 น.</li> <li>● ปิดพื้นที่ ที่มีคนพลุกพล่าน เช่น สถานที่แสดงมหรสพ สถานที่ที่มีการเล่นสากาธารณะ สถานที่จัดเลี้ยง จัดงานแต่ง เป็นต้น</li> <li>● ปิดมหาวิทยาลัยและโรงเรียนในพื้นที่ กทม. ให้เรียนออนไลน์</li> <li>● เดินทางเข้า-ออก ราชอาณาจักรได้ แต่กักตัว 14 วัน</li> <li>● ประชาชนทั่วไปเริ่มได้รับวัคซีน พฤษภาคม 2564</li> </ul> |
| ระลอกที่ 3<br>(April 2021 – May 2021)   |                                 | <ul style="list-style-type: none"> <li>● การห้ามเข้าพื้นที่เสี่ยง ได้แก่ กทม.</li> <li>● Work from Home</li> <li>● จำกัดเวลาเปิด-ปิดห้างสรรพสินค้า(11.00-21.00 น.)</li> <li>● ตลาด/ตลาดนัด เปิดแค่จำหน่ายอาหารซื้อกลับบ้านหลัง 21.00 น.</li> <li>● ปิดพื้นที่ ที่มีคนพลุกพล่าน เช่น สถานที่แสดงมหรสพ สถานที่ที่มีการเล่นสากาธารณะ สถานที่จัดเลี้ยง จัดงานแต่ง เป็นต้น</li> <li>● ปิดมหาวิทยาลัยและโรงเรียนในพื้นที่ กทม. ให้เรียนออนไลน์</li> <li>● เดินทางเข้า-ออก ราชอาณาจักรได้ แต่กักตัว 14 วัน</li> <li>● ประชาชนทยอยได้รับวัคซีน เข็ม 2</li> </ul>                                       |

ตาราง 2.2 (ต่อ)

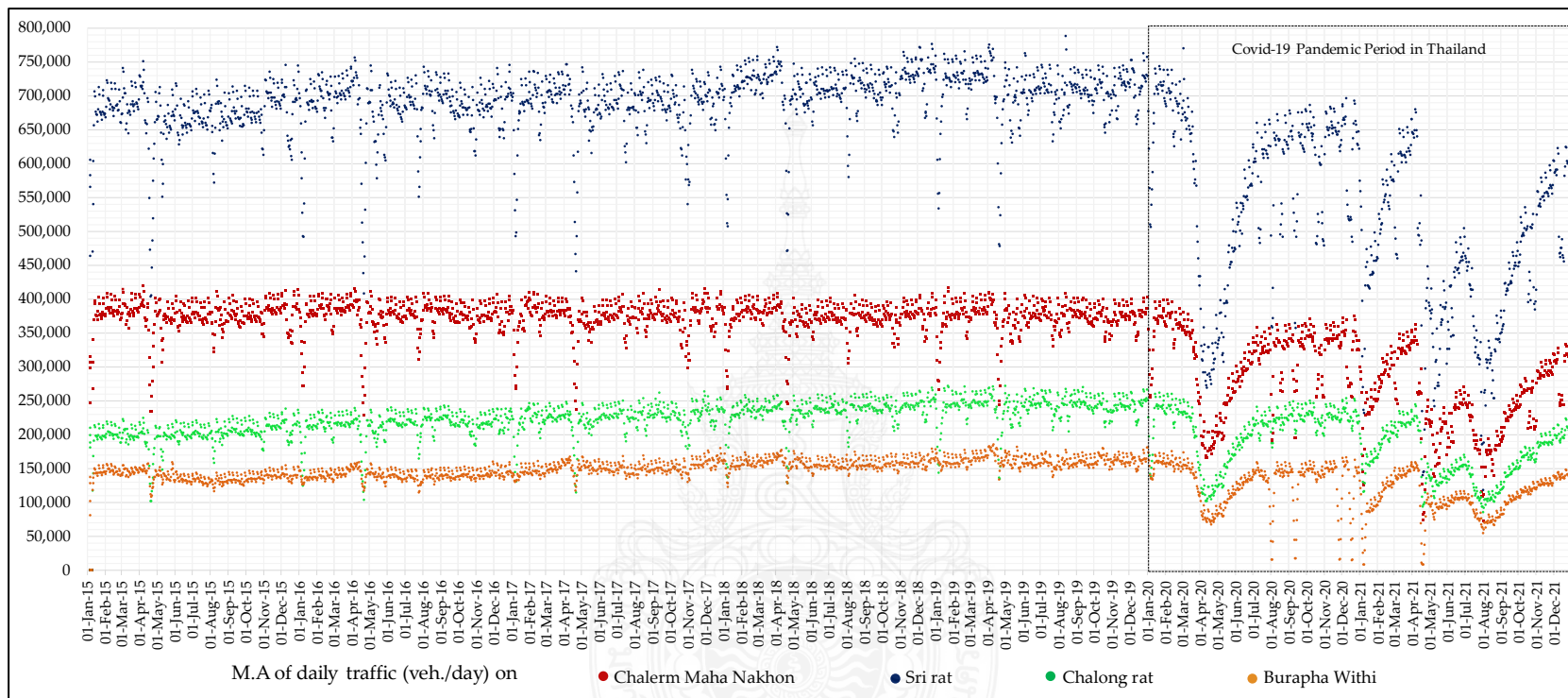
| ผลกระทบจากการแพร่ระบาดของโควิด 19         | จำนวนผู้ติดเชื้อใหม่ (Case/วัน) | มาตรการ  |
|---|---------------------------------|--|
| ระลอกที่ 4<br>(June 2021 – December 2021) |                                 | <ul style="list-style-type: none"> <li>● การห้ามเข้าพื้นที่เสี่ยง ได้แก่ กทม.</li> <li>● ห้ามเดินทางข้ามจังหวัด</li> <li>● ห้ามออกนอกเคสสถาน เวลา 21.00 – 04.00 น.</li> <li>● Work from Home</li> <li>● จำกัดเวลาเปิด-ปิดห้างสรรพสินค้า(11.00-20.00 น.)</li> <li>● ตลาด/ตลาดนัด เปิดแค่จำหน่ายอาหารซ็อกกลับบ้านหลัง 20.00 น.</li> <li>● ปิดพื้นที่ ที่มีคนพลุกพล่าน เช่น สถานที่แสดงมหรสพ สถานที่ที่มีการละเล่นสาธารณะ สถานที่จัดเลี้ยง จัดงานแต่ง เป็นต้น</li> <li>● ปิดมหาวิทยาลัยและโรงเรียนในพื้นที่ กทม. ให้เรียนออนไลน์</li> <li>● เดินทางเข้า-ออก ราชอาณาจักรได้ แต่กักตัว 14 วัน</li> <li>● ประชาชนทยอยได้รับวัคซีน เข็มกระตุ้น</li> </ul> |
| ระลอกที่ 5<br>(January 2022 – May 2022)   |                                 | <ul style="list-style-type: none"> <li>● เดินทางเข้า-ออก พื้นที่ กทม. ได้</li> <li>● หน่วยงานเอกชนเริ่มทยอยยกเลิก Work from Home</li> <li>● เปิด-ปิดห้างสรรพสินค้าเวลาปกติ</li> <li>● ตลาด/ตลาดนัด เปิดปกติ</li> <li>● เปิดพื้นที่ ที่มีคนพลุกพล่าน เช่น สถานที่แสดงมหรสพ สถานที่ที่มีการละเล่นสาธารณะ สถานที่จัดเลี้ยง จัดงานแต่ง เป็นต้น</li> <li>● เปิดมหาวิทยาลัยและโรงเรียนในพื้นที่ กทม.</li> <li>● เดินทางเข้า-ออก ราชอาณาจักรได้ ไม่ต้องกักตัวหากฉีดวัคซีนครบถ้วน และมี vaccine passport</li> <li>● ประชาชนทยอยได้รับวัคซีน เข็มกระตุ้น</li> </ul>   |

ที่มา: รวบรวมโดยผู้วิจัย

จากภาพ 2.4 ในช่วงที่มีการระบาด 2 ช่วงแรก หรือในปีแรกของการระบาดนั้น พบว่าจำนวนผู้ติดเชื้อต่อวันไม่สูงนัก เนื่องจากประชาชนก็ต่างกักตัวอยู่บ้าน (New Normal) ไม่ออกไปใช้ชีวิตตามปกติ เพราะ Covid-19 ยังถือเป็นโรคอุบัติใหม่ และยังไม่มียุทธศาสตร์ป้องกัน แต่เมื่อประชาชนต่างได้รับวัคซีน ก็เริ่มกลับมาใช้ชีวิตปกติ และคุ้นชินกับการแพร่ระบาดของ Covid-19 ประกอบกับหละหลวมต่อมาตรการป้องกัน จึงทำให้การระบาดของโรคในระลอกที่ 3 และระลอกที่ 4 เริ่มมีประชาชนติดเชื้อต่อวันสูงขึ้นในช่วงระยะเวลาที่สั้นลง ในขณะที่ตัวเชื้อไวรัส Covid-19 ก็วิวัฒนาการตัวเองให้สามารถแพร่กระจายเชื้อได้เร็วและรุนแรงมากยิ่งขึ้น จึงทำให้การแพร่ระบาดของไวรัสในระลอกที่ 4 และ 5 นั้น มีอัตราการติดเชื้อเฉลี่ยต่อวันสูงสุดถึงหลัก 10,000 คน และค่อยๆ ลดจำนวนลงมาในช่วงเดือนกรกฎาคม พ.ศ. 2565 ประเทศไทยจึงประกาศว่าสามารถควบคุมสถานการณ์การแพร่ระบาดได้ และประกาศให้ Covid-19 เป็นโรคประจำถิ่น



ภาพ 2.4 Timeline การแพร่ระบาดของ Covid-19 ในประเทศไทย และจำนวนผู้ติดเชื้อใหม่รายวันที่  
ที่มา: รวบรวมโดยผู้วิจัย



ภาพ 2.5 การกระจายตัวการเดินทางบนทางด่วนในเมืองตั้งแต่ พ.ศ. 2558 – 2564 โดยเมื่อเกิดการแพร่ระบาดของ Covid-19 ทำให้การเดินทางบนถนนลดลงเฉลี่ยร้อยละ 23

ที่มา: รวบรวมโดยผู้วิจัย

จากภาพ 2.5 สถานการณ์การแพร่ระบาดของโรค Covid-19 ส่งผลให้การเดินทางลดลงอย่างเห็นได้ชัด ซึ่งแน่นอนว่า การระบาดของ Covid-19 นี้ทำให้สถานการณ์การเดินทางเปลี่ยนไปทั่วโลก

โดยมีงานศึกษาวิจัยใน Texas (Jiao et al., 2021) ได้นำแบบจำลอง autoregressive distributed lag มาใช้เพื่อวิเคราะห์การเปลี่ยนแปลงรูปแบบการเดินทางใน Houston ช่วง Covid-19 โดยแบบจำลองนี้ แสดงวิธีการเฉพาะ (unique) ในการหาปริมาณและวิเคราะห์พฤติกรรมการเดินทางใน Houston ในช่วง Covid-19 เท่านั้น ซึ่งผลการวิจัยชี้ให้เห็นว่า รูปแบบการเข้าเยี่ยมผู้ป่วยและการเปลี่ยนแปลงจำนวนของผู้ป่วย Covid-19 ของหนึ่งสัปดาห์ก่อนหน้า จะมีอิทธิพลอย่างมากต่อพฤติกรรมเดินทางในสัปดาห์ถัดไป นอกจากนี้ ยังพบว่า ระบบขนส่งสาธารณะมีการใช้งานโดยรวมลดลง แต่ไม่มีนัยสำคัญต่อการประมาณจำนวนคนที่เดินเท้าที่เพิ่มขึ้นทั้งหมด ในประเทศฮังการี (Dingil and Esztergár-Kiss, 2021) มีงานศึกษาถึงความสัมพันธ์ระหว่างการระบาดของ Covid-19 และรูปแบบการเดินทางที่เกิดขึ้นในประเทศ โดยใช้ข้อมูลจากการสำรวจระหว่างประเทศ ซึ่งมีการวิเคราะห์ลักษณะทาง Socio economic ของผู้ใช้ระบบขนส่งในช่วงก่อนเกิดการระบาด และเมื่อเกิดการแพร่ระบาดของโรคจึงใช้วิธีมีการจำลองพฤติกรรมเดินทางแทน โดยใช้แบบจำลอง the multinomial จากผลการวิจัย พบว่า ผู้ใช้ระบบขนส่งสาธารณะมีแนวโน้มที่จะเปลี่ยนโหมดการเดินทางมากกว่าผู้ใช้รถยนต์ ผู้ใช้รถจักรยานยนต์ และคนเดินเท้า มากถึง 31.5 เท่า มากกว่าร้อยละ 10.6 และ 6.9 เท่า ตามลำดับ ผลการวิจัยยังแสดงให้เห็นว่าระยะทางที่ใช้ในการเดินทางและระดับรายได้เป็นสองปัจจัยที่มีอิทธิพลมากที่สุด ในการตัดสินใจเกี่ยวกับการเดินเลือกรูปแบบการเดินทางในช่วงการระบาดของโรค ซึ่งผลลัพธ์เหล่านี้สามารถยืนยันภาพสะท้อนของความไม่เท่าเทียมกันเชิงพื้นที่และเศรษฐกิจในระหว่างการแพร่ระบาด อีกทั้ง รถจักรยานยนต์และรถยนต์ส่วนบุคคลยังถูกมองว่าเป็นรูปแบบการขนส่งในเมืองที่มีความเสี่ยงน้อยที่สุด ในช่วงการแพร่ระบาด ดังนั้นการเติบโตของการเดินทางในรูปแบบดังกล่าวจึงเพิ่มขึ้นถึงร้อยละ 26 และ 15 ตามลำดับ หลังจากการระบาดของ Covid-19 ความต้องการเดินทางลดลง เนื่องจากการเดินทางด้วยระบบขนส่งมวลชนลดลง ซึ่งส่งผลกระทบต่อเป็นวงกว้างจึงมีการสำรวจการเปลี่ยนแปลงพฤติกรรม การเดินทางด้วยระบบขนส่งสาธารณะเนื่องจากการแพร่ระบาดของ Covid-19 ในประเทศสหรัฐอเมริกา (Cho and Park, 2021) โดยเปรียบเทียบจำนวนการเบียดของผู้โดยสาร (crowding impedances) ในเชิงปริมาณก่อนและหลังการระบาดของ Covid-19 ผ่านการประเมินด้วยแบบจำลอง random parameter mixed logit ผลการวิจัยแสดงให้เห็นว่า การเบียดของผู้โดยสารหลังจากการระบาดของ Covid-19 นั้นตัวเลขสถิติที่สูงกว่าก่อนการระบาดของ Covid-19 ประมาณ 1.04 ~ 1.23 เท่า ทั้งนี้ จึงมีการเสนอนโยบายการขนส่งสำหรับผู้กำหนดนโยบายการขนส่งและผู้ปฏิบัติงานเพื่อรับมือกับวิกฤตการณ์การขนส่งสาธารณะที่เกิดจากโรคติดเชื้อ

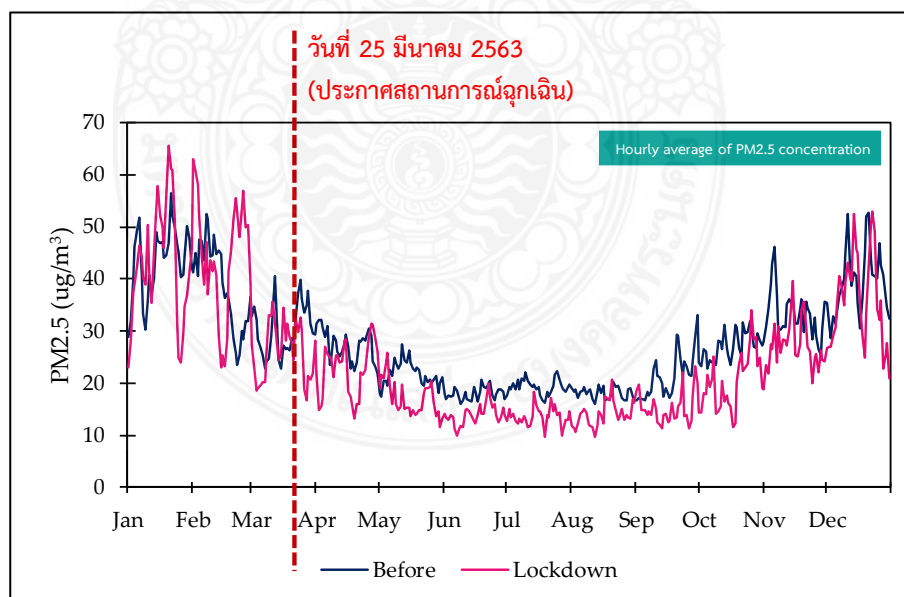
จากเหตุการณ์ Lockdown เนื่องจากการแพร่ระบาดของ Covid-19 มีการศึกษาที่พยายามจะทำความเข้าใจการเปลี่ยนแปลงของรูปแบบการจราจรบนท้องถนนในระยะสั้น โดยใช้ข้อมูลจากเครื่องสแกน MAC แบบ Wi-Fi สองเครื่องที่ติดตั้งที่จุดยุทธศาสตร์ใน Chennai ประเทศอินเดีย (Patra et al., 2021) ผลการวิจัยชี้ให้เห็นว่ากิจกรรมการจราจรบนท้องถนนลดลงอย่างมีนัยสำคัญเนื่องจากข้อจำกัดในการเดินทาง (ลดการเดินทางไม่จำเป็น) การไม่เดินทางไปสถานที่ทำงาน และการเฝ้าระวังอย่างเข้มงวดในระหว่าง Lockdown อย่างไรก็ตาม เมื่อมีการผ่อนคลาย Lockdown ลง

กิจกรรมการจราจรบนท้องถนนก็เริ่มฟื้นตัว พบว่า การ Lockdown อย่างสมบูรณ์มีประสิทธิภาพสูงสุดในการลดกิจกรรมการเดินทางบนท้องถนน แต่การ Lockdown ที่สมบูรณ์ในระยะเวลาสั้น ๆ แบบเฉพาะกิจอาจให้ผลประโยชน์เพียงชั่วคราวเท่านั้น นอกจากนี้ การ Lockdown โดยไม่มีการบังคับใช้ที่เหมาะสมอาจไม่ได้ผลเนื่องจากประชาชนดูเหมือนจะเพิกเฉยต่อคำแนะนำหลังจากนั้นไม่นาน

จากปฏิริยาตอบสนองต่อโรคโคโรนาไวรัสสายพันธุ์ใหม่ (Covid-19) ประเทศต่างๆ ทั่วโลกได้ดำเนินการมาตรการต่างๆ เพื่อลดการแพร่กระจายของไวรัส ภาคการขนส่งได้รับผลกระทบโดยตรง โดยเฉพาะอย่างยิ่งจากการแพร่ระบาดของสถานการณ์ดังกล่าว ในประเทศเยอรมนี (König and Dreßler, 2021) จึงได้ทำการศึกษาเกี่ยวกับผลกระทบของสถานการณ์โคโรนาไวรัส เกี่ยวกับความคล่องตัวในการเดินทางเพื่อขยายมุมมองไปยังผู้อยู่อาศัยในพื้นที่ชนบทที่มีการเดินทางและให้ข้อมูลเกี่ยวกับการรับรู้การเปลี่ยนแปลงของแนวปฏิบัติด้านการเดินทางของบุคคลที่ได้รับผลกระทบ โดยการสัมภาษณ์ทางโทรศัพท์และแบบสำรวจครัวเรือนเชิงปริมาณ ในเขตชนบทของ Altmarkkreis Salzwedel ทางตอนเหนือเยอรมนี พบว่า ร้อยละ 62 ไม่มีการเปลี่ยนแปลงการเดินทางหรือรูปแบบการเดินทางใด ๆ เนื่องจากสถานการณ์การแพร่ระบาดของ Covid-19 แต่ในภาพรวม พบว่า การเดินทางทั้งหมดเกือบหนึ่งในสามก็ถูกยกเลิก และยังพบข้อสังเกตว่า มีการเปลี่ยนรูปแบบการเดินทาง (mode-shift) ทางจากรถยนต์และรถประจำทาง ไปสู่การเดินทางด้วยจักรยาน ส่วน mode-shift ของการเดินทางไม่มีการเปลี่ยนแปลง ผู้ตอบแบบสอบถามส่วนใหญ่ไม่ได้คิดถึงผลกระทบในระยะยาวที่ชัดเจนของการแพร่ระบาดต่อพฤติกรรมการเดินทางของพวกเขา ยังมีการศึกษาเพิ่มเติม (Anke et al., 2021) โดยทำการวิเคราะห์ข้อมูลของผู้เข้าร่วมการสำรวจ 4,157 คน (2,512 จากพื้นที่ที่ไม่มีการ Lockdown และ 1,645 จากพื้นที่ Lockdown) ในช่วงสถานการณ์ Covid-19 พบว่า พฤติกรรมการเดินทางของผู้คนเปลี่ยนแปลงไป โดยเฉพาะอย่างยิ่ง การเดินทางจากระบบขนส่งสาธารณะไปสู่รถยนต์ส่วนบุคคล และการเดินกับปั่นจักรยานที่เพิ่มขึ้น ในขณะที่ประเทศฟิลิปปินส์ (Hasselwander et al., 2021) ทำการสำรวจพฤติกรรมของประชากรในพื้นที่กรุงมะนิลาโดยทำการวิเคราะห์ข้อมูลโทรศัพท์มือถือและ GPS แบบรวมจาก Google และ Apple ซึ่งให้ข้อมูลเกี่ยวกับพฤติกรรมการเดินทางทั้งก่อนและระหว่างการ Lockdown พบว่า การเดินทางทุกรูปแบบ ได้ลดลงอย่างมีนัยสำคัญ และระบบขนส่งสาธารณะก็ลดลงมากที่สุดเฉลี่ยถึงร้อยละ 74.5 ในประเทศสเปน (Awad-Núñez et al., 2021) ได้วิเคราะห์ข้อมูลที่มาจาก ITS ของระบบขนส่งสาธารณะและจากกล้องควบคุมการจราจรและเซ็นเซอร์สิ่งแวดล้อมเพื่อเปรียบเทียบการเดินทางระหว่างก่อน Lockdown และ หลัง Lockdown ซึ่งจากการวิเคราะห์ พบว่า การจราจรโดยรวมลดลงร้อยละ 76 จากผู้ใช้ระบบขนส่งสาธารณะลดลงถึงร้อยละ 93 การปล่อยก๊าซ NO<sub>2</sub> ลดลงได้ถึงร้อยละ 60 และอุบัติเหตุจากการจราจรลดลงได้ถึงร้อยละ 67 ในขณะที่ประเทศอินโดนีเซีย (Irawan et al., 2021) ได้มีการศึกษาการเปลี่ยนแปลงของกิจกรรมและการเดินทางที่เกี่ยวข้องในช่วงเริ่มต้นของการระบาดของ Covid-19 ผ่านการตอบแบบสอบถาม 1,062 คน โดยใช้แบบสอบถามบนเว็บไซต์ ซึ่งจากการศึกษานี้พบว่า การระบาดของ Covid-19 และ การ WFH ส่งผลต่อการเดินทางโดยตรง ซึ่งทำให้เกิดการเปลี่ยนแปลงพฤติกรรมการเดินทาง ในขณะที่หลายประเทศทั่วโลก (Cui et al., 2021; Lee et al., 2020; Muley et al., 2021; Politis et al., 2021) ต่างก็ศึกษาวิเคราะห์พฤติกรรมการเดินทางของประชาชนในขณะที่มีการแพร่ระบาดของโรค Covid-19 พบว่าการเดินทางลดลงอย่างเห็นได้ชัด ทั้งนี้ ในบางประเทศ รูปแบบการเดินทางจะยังคงเดิม และการละเมิดกฎจราจรและอุบัติเหตุ พบว่าลดลงถึงร้อยละ 73 และร้อยละ 37 ตามลำดับ

### 2.1.6 สถานการณ์ของ PM 2.5 หลังจกมีการแพร่ระบาดของ Covid-19

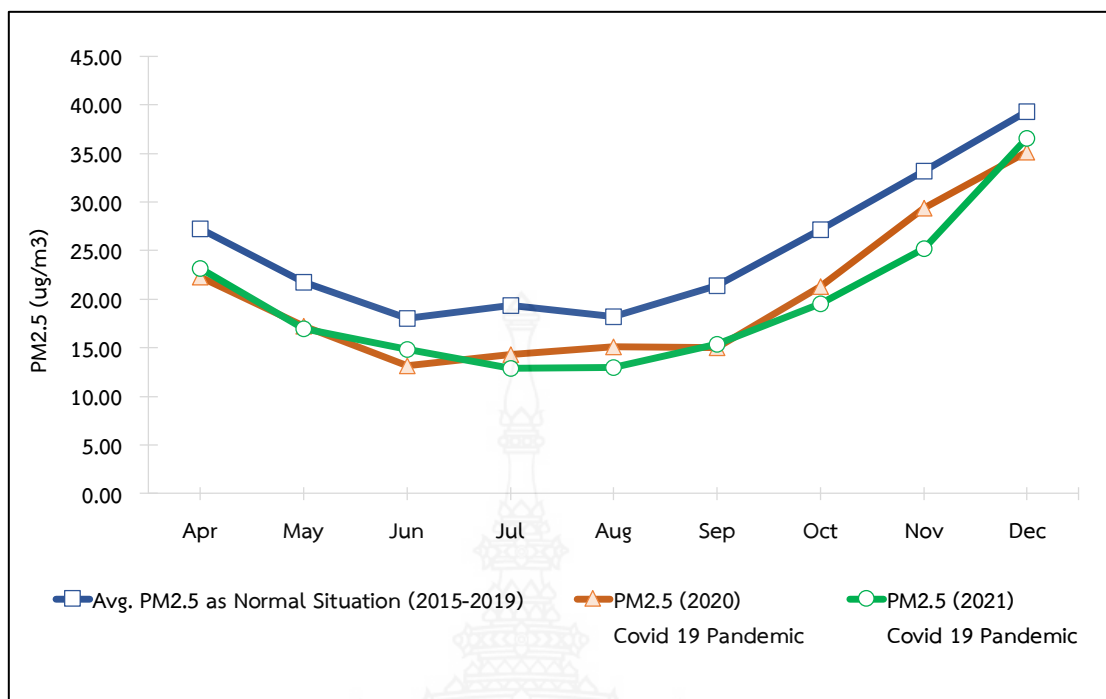
ในช่วง 2-3 ปีที่ผ่านมา มีการแพร่ระบาดของ Covid-19 ซึ่งลดกิจกรรมของมนุษย์ (human activities) ลงไปเยอะมาก และนั่นทำให้สภาพแวดล้อมโดยรวมของโลกดีขึ้น แต่หากการจะนำข้อมูล PM 2.5 หรือข้อมูลมลพิษทางอากาศต่าง ๆ ในห้วงเวลาดังกล่าว ไปคาดการณ์ในอนาคต ซึ่งอาจทำให้ปัจจัยที่ส่งผลต่อการเกิด PM 2.5 มีการเปลี่ยนแปลงไป จากข้อมูลความสัมพันธ์ของ PM 2.5 กับแก๊สมลพิษต่าง ๆ ในช่วงก่อนการเกิด PM 2.5 และช่วงที่มีการระบาดของ PM 2.5 นั้น มีความแตกต่างกันอย่างมีนัยสำคัญ โดยจากสถิติดังกล่าว พบว่า ค่าความเข้มข้นของ PM 2.5 จะเพิ่มสูงขึ้นในห้วงฤดูหนาว (Dry season) และจะลดต่ำลงในห้วงฤดูฝน (Wet season) แม้ว่าจะอยู่ในห้วงการระบาดของ Covid-19 ค่าความเข้มข้นของ PM 2.5 ก็มีรูปแบบพฤติกรรมความเข้มข้นไม่ต่างจากสถานการณ์ปกติ ก่อนที่จะมีการระบาดของ Covid-19 แต่อย่างไรก็ดี ซึ่งจะแตกต่างกันเพียงแต่ความเข้มข้นของ PM 2.5 ในห้วงที่มีการระบาดของ Covid-19 อาจจะมีค่าความเข้มข้นต่ำกว่าช่วงก่อนการระบาดของ Covid-19 (พ.ศ. 2558-2562) และเมื่อมีการ Lockdown ในขณะที่มีการแพร่ระบาดของโรค Covid-19 (พ.ศ.2563-2564) เมื่อพิจารณาเปรียบเทียบความเข้มข้นของ PM 2.5 ระหว่างสถานการณ์ปกติและสถานการณ์การแพร่ระบาดของ Covid-19 พบว่า ในช่วงฤดูหนาวค่าความเข้มข้น PM 2.5 ระหว่างการแพร่ระบาดของ Covid-19 มีสถิติสูงกว่าช่วงสถานการณ์ปกติ (เดือน มกราคม - มีนาคม) แต่ในช่วงฤดูฝนของห้วงการแพร่ระบาดของ Covid-19 กลับพบว่ามีค่าความเข้มข้น PM 2.5 ต่ำกว่าห้วงสถานการณ์ปกติ (เดือน พฤษภาคม - ตุลาคม) พบว่า ฤดูกาลเป็นหนึ่งในปัจจัยที่สำคัญต่อความเข้มข้น PM 2.5 ที่เกิดขึ้น



ภาพ 2.6 สถิติข้อมูลค่าความเข้มข้นเฉลี่ยของ PM 2.5 ในขณะก่อนที่จะมีการแพร่ระบาดของ Covid-19 (พ.ศ. 2558-2562) และเมื่อมีการ Lockdown ในขณะที่มีการระบาดของเชื้อ Covid-19 (พ.ศ. 2563-2564)

ที่มา: รวบรวมโดยผู้วิจัย





ภาพ 2.7 ข้อมูลแสดงค่าความเข้มข้นของ PM 2.5 เฉลี่ยรายเดือน ตั้งแต่เดือน เมษายน - ธันวาคม  
ที่มา: รวบรวมโดยผู้วิจัย

จากภาพ 2.7 เมื่อพิจารณาข้อมูลแสดงค่าความเข้มข้นของ PM 2.5 เฉลี่ยรายเดือน ตั้งแต่เดือน เมษายน - ธันวาคม (ที่เริ่มเดือนเมษายน เนื่องจากผู้วิจัยต้องการทราบการเปลี่ยนแปลงของการ Lockdown ของประเทศไทย ซึ่งเริ่มขึ้นในเดือนเมษายน) โดยเส้นกราฟสีฟ้าเป็นข้อมูลค่าความเข้มข้นของ PM 2.5 เฉลี่ยรายเดือนของช่วงที่มีสถานการณ์ปกติ (พ.ศ. 2558-2562) เส้นกราฟสีส้มเป็นข้อมูลค่าความเข้มข้นของ PM 2.5 เฉลี่ยรายเดือนของช่วงที่มีการแพร่ระบาดของ Covid-19 ในปี พ.ศ. 2563 เส้นกราฟสีเขียวเป็นข้อมูลค่าความเข้มข้นของ PM 2.5 เฉลี่ยรายเดือนของช่วงที่มีการแพร่ระบาดของ Covid-19 ในปี พ.ศ. 2564 นั้นหมายความว่า ค่าความเข้มข้นของ PM 2.5 ในช่วงที่มีการแพร่ระบาดของ Covid-19 มีสถิติลดลงอย่างชัดเจน เมื่อเปรียบเทียบกับสถานการณ์ปกติ แสดงว่าการ Lockdown ให้อยู่ที่บ้าน และการจำกัดพื้นที่ของประชาชนมีผลต่อค่าความเข้มข้นของ PM 2.5 ในกรุงเทพมหานคร

อย่างไรก็ตาม ในงานวิจัยของต่างประเทศ พบว่า เมื่อมีการแพร่ระบาดของ Covid-19 ทั่วโลกนั้น ส่งผลให้รูปแบบการดำเนินชีวิตของผู้คนเปลี่ยนแปลงไปมีการ Lockdown ให้คนอยู่แต่ในการ เรียนออนไลน์ Work from home ซึ่งส่งผลกระทบต่อเศรษฐกิจเป็นวงกว้าง และเมื่อกิจกรรมที่เป็นปกติของผู้คนถูกจำกัด ส่งผลให้การเดินทาง (Travel trip) ลดน้อยลงเป็นอย่างมาก และมีการเปลี่ยนรูปแบบการเดินทาง (Mode choice) อย่างเห็นได้ชัดในหลายประเทศ (Dang et al., 2020; Puliatti et al., 2020; Venter et al., 2020)

## 2.2 ข้อมูลการตรวจวัดค่า PM 2.5 และมลพิษทางอากาศ

### 2.2.1 ค่ามาตรฐานที่เกี่ยวข้องกับฝุ่นละอองของประเทศไทย

กระทรวงทรัพยากรธรรมชาติและสิ่งแวดล้อม โดยกรมควบคุมมลพิษ (กรมควบคุมมลพิษ , 2562) ได้กำหนดมาตรฐานเกี่ยวกับคุณภาพอากาศในบรรยากาศทั่วไป สรุปได้ ดังนี้

- 1) ประกาศคณะกรรมการสิ่งแวดล้อมแห่งชาติ ฉบับที่ 10 (พ.ศ. 2538)
- 2) ประกาศคณะกรรมการสิ่งแวดล้อมแห่งชาติ ฉบับที่ 24 (พ.ศ. 2547)
- 3) ประกาศคณะกรรมการสิ่งแวดล้อมแห่งชาติ ฉบับที่ 28 (พ.ศ. 2550)
- 4) ประกาศคณะกรรมการสิ่งแวดล้อมแห่งชาติ ฉบับที่ 36 (พ.ศ. 2553)

จากประกาศคณะกรรมการสิ่งแวดล้อมแห่งชาติทั้ง 4 ฉบับข้างต้น พบว่า มาตรฐานกำหนดให้มีค่าความเข้มข้นในบรรยากาศของฝุ่นขนาดเล็กกว่า 2.5 ไมครอน (PM 2.5) เฉลี่ย 24 ชั่วโมง ไม่เกิน 0.05 mg/m<sup>3</sup> และค่าเฉลี่ยในระยะเวลา 1 ปี มีค่าไม่เกิน 0.025 mg/m<sup>3</sup> โดยมีรายละเอียดค่ามาตรฐานฝุ่นละอองในต่างประเทศ และค่ามาตรฐานของฝุ่นละอองชนิดต่างๆ ในชั้นบรรยากาศของประเทศไทย ดังตาราง 2.3 – 2.4

ตาราง 2.3 ค่ามาตรฐานฝุ่นละอองในต่างประเทศ

| ประเทศ          | PM 10 (µg/m <sup>3</sup> ) |      | PM 2.5 (µg/m <sup>3</sup> ) |      |
|-----------------|----------------------------|------|-----------------------------|------|
|                 | 24 ชั่วโมง                 | 1 ปี | 24 ชั่วโมง                  | 1 ปี |
| WHO             | 50                         | 20   | 25                          | 10   |
| US.EPA          | 150                        | -    | 35                          | 12   |
| สหภาพยุโรป (EU) | 50                         | 40   | -                           | 25   |
| ออสเตรเลีย      | -                          | 50   | 25                          | 8    |
| อินเดีย         | 100                        | 60   | 60                          | 40   |
| สิงคโปร์        | 50                         | 20   | 37.5                        | 12   |
| จีน             | 150                        | 70   | 75                          | 35   |
| ไทย             | 120                        | 50   | 50                          | 25   |

ที่มา: กรมควบคุมมลพิษ (2564)

ตาราง 2.4 ค่ามาตรฐานฝุ่นละอองในบรรยากาศของประเทศไทย

| ฝุ่นละออง              | ค่ามาตรฐาน (µg/m <sup>3</sup> ) |             |
|------------------------|---------------------------------|-------------|
|                        | เฉลี่ย 24 ชั่วโมง               | เฉลี่ย 1 ปี |
| ขนาดไม่เกิน 100 ไมครอน | ไม่เกิน 330                     | ไม่เกิน 100 |
| ขนาดไม่เกิน 10 ไมครอน  | ไม่เกิน 120                     | ไม่เกิน 50  |
| ขนาดไม่เกิน 2.5 ไมครอน | ไม่เกิน 50                      | ไม่เกิน 25  |

ที่มา: กรมควบคุมมลพิษ (2564)

## 2.2.2 ดัชนีคุณภาพอากาศของประเทศไทย (Air Quality Index: AQI)

ประเทศไทยได้ใช้ดัชนีคุณภาพอากาศ (Air Quality Index: AQI) (กรมควบคุมมลพิษ, 2564) ในการติดตามสถานการณ์ปัญหาหมอกควัน เพื่อให้ประชาชนเข้าใจได้ง่าย และเพื่อให้คนในแต่ละพื้นที่ได้รับทราบถึงสถานการณ์มลพิษทางอากาศว่ารุนแรงแค่ไหน กระทบต่อสุขภาพร่างกายมากน้อยเพียงใด โดยดัชนีคุณภาพอากาศถูกใช้กันอย่างแพร่หลายทั่วโลก ทั้งในสหรัฐอเมริกา ออสเตรเลีย ในเอเชียอย่าง ประเทศสิงคโปร์ มาเลเซีย และไทย เป็นต้น

ดัชนีคุณภาพอากาศใช้แทนค่าความเข้มข้นของมลพิษที่กระจายอยู่ในอากาศทั้งสิ้น 6 ชนิด ด้วยกัน (กรมควบคุมมลพิษ, 2564) ได้แก่ 1) ฝุ่นละอองขนาดไม่เกิน 10 ไมครอน (PM 10) 2) ฝุ่นละอองขนาดไม่เกิน 2.5 ไมครอน (PM 2.5) 3) ก๊าซโอโซน ( $O_3$ ) 4) ก๊าซคาร์บอนมอนอกไซด์ (CO) 5) ก๊าซไนโตรเจนไดออกไซด์ ( $NO_2$ ) 6) ก๊าซซัลเฟอร์ไดออกไซด์ ( $SO_2$ )

ดัชนีคุณภาพอากาศสำหรับประเทศไทยสามารถแบ่งออกได้เป็น 5 ระดับ ได้แก่ ระดับอันตราย ระดับไม่ดี ระดับปานกลาง ระดับดี และระดับดีมาก โดยมีค่า AQI ตั้งแต่ 0 ถึง 201 ขึ้นไป ซึ่งแต่ละระดับจะใช้สีเป็นสัญลักษณ์เทียบกับระดับที่ส่งผลกระทบต่อสุขภาพ (ตาราง 2.5) โดยมาตรฐานคุณภาพอากาศในบรรยากาศโดยทั่วไปจะมีค่าดัชนีคุณภาพอากาศอยู่ที่ 100 ซึ่งหากดัชนีคุณภาพอากาศมีค่ามากกว่า 100 แสดงว่าเกินมาตรฐาน โดยจะเริ่มมีผลกระทบต่อสุขภาพ โดยมีรายละเอียดของเกณฑ์ของดัชนีคุณภาพอากาศของประเทศไทย ดังตาราง 2.5

ตาราง 2.5 เกณฑ์ของดัชนีคุณภาพอากาศของประเทศไทย

| AQI         | ระดับคุณภาพอากาศ | สีที่ใช้ | คำอธิบาย  |
|-------------|------------------|----------|---|
| 0-25        | ดีมาก            | ฟ้า      | คุณภาพอากาศดีมาก เหมาะสำหรับกิจกรรมที่อยู่กลางแจ้งและการท่องเที่ยว  |
| 26-50       | ดี               | เขียว    | คุณภาพอากาศดี สามารถทำกิจกรรมที่อยู่กลางแจ้งและการท่องเที่ยวได้ตามปกติ  |
| 51-100      | ปานกลาง          | เหลือง   | คนทั่วไปสามารถทำกิจกรรมในพื้นที่กลางแจ้งได้ตามปกติ  |
| 101-200     | ไม่ดี            | ส้ม      | คนทั่วไปต้องเฝ้าระวังสุขภาพ ถ้ามีอาการ ไอ หายใจลำบาก ระคายเคือง ควรงดกิจกรรมกลางแจ้ง หาเสื้อคลุม แว่นตา และผ้าปิดจมูก สำหรับผู้ป่วยโรคระบบทางเดินหายใจ และควรเลี่ยงกิจกรรมที่อยู่ในพื้นที่กลางแจ้ง ส่วนผู้สูงอายุและเด็ก ควรหลีกเลี่ยงการทำกิจกรรมที่อยู่ในพื้นที่กลางแจ้งเป็นเวลานาน |
| มากกว่า 200 | อันตราย          | แดง      | ทุกคนต้องหลีกเลี่ยงกิจกรรมที่อยู่กลางแจ้ง และควรสวมใส่ผ้าปิดจมูกป้องกันตนเอง  |

ที่มา: กรมควบคุมมลพิษ (2564)

การคำนวณดัชนีคุณภาพอากาศรายวันจะคำนวณจากค่าความเข้มข้นของสารมลพิษทางอากาศที่ได้จากข้อมูลผลการตรวจคุณภาพอากาศ จากสถานตรวจวัดของกรมควบคุมมลพิษ

โดยเทียบเท่ากับค่าดัชนีคุณภาพอากาศที่ระดับต่างๆ โดยสำนักจัดการคุณภาพอากาศและเสียง กรมควบคุมมลพิษ ในประเทศไทย ได้กำหนดเกณฑ์ของดัชนีชี้วัดคุณภาพอากาศ โดยมีการคำนวณเปรียบเทียบกับมาตรฐานคุณภาพบรรยากาศทั่วไปของสารมลพิษทางอากาศ 6 ประเภท ดังแสดงในตาราง 2.6

การคำนวณดัชนีย่อยคุณภาพของอากาศ ดังสมการ (2.1)

$$I = \frac{I_j - I_i}{X_j - X_i} (X - X_i) + I_i \quad (2.1)$$

โดยที่

- $I$  คือ ค่าดัชนีย่อยคุณภาพอากาศ  
 $X$  คือ ค่าความเข้มข้นสารมลพิษทางอากาศ  
 $X_i, X_j$  คือ ค่าต่ำสุด, สูงสุด ของช่วงความเข้มข้นสารมลพิษที่มีค่า  $X$   
 $I_i, I_j$  คือ ค่าต่ำสุด, สูงสุด ของช่วงดัชนีคุณภาพอากาศที่ตรงกับช่วงความเข้มข้นจากค่าดัชนีย่อยที่คำนวณได้ โดยที่สารมลพิษทางอากาศประเภทใดที่มีค่าดัชนีสูงสุด จะใช้เป็นดัชนีคุณภาพอากาศ (AQI) ในช่วงเวลานั้น

ตาราง 2.6 แสดงรายละเอียดของความเข้มข้นของมลพิษทางอากาศตามหน่วยวัด AQI ซึ่งผลการตรวจ AQI ที่ดีไม่ควรเกิน 50

ตาราง 2.6 ค่าความเข้มข้นของสารมลพิษทางอากาศที่เทียบเท่ากับค่าดัชนีคุณภาพอากาศ

| AQI       | PM 2.5<br>( $\mu\text{g}/\text{m}^3$ ) | PM 10<br>( $\mu\text{g}/\text{m}^3$ ) | O <sub>3</sub><br>(ppb)   | CO<br>(ppm) | NO <sub>2</sub><br>(ppb)  | SO <sub>2</sub><br>(ppb) |
|-----------|--|---------------------------------------|---------------------------|-------------|---------------------------|--------------------------|
|           | เฉลี่ย 24 ชั่วโมงต่อเนื่อง             |                                       | เฉลี่ย 8 ชั่วโมงต่อเนื่อง |             | เฉลี่ย 1 ชั่วโมงต่อเนื่อง |                          |
| 0 – 25    | 0 – 25                                 | 0 – 50                                | 0 – 35                    | 0 – 4.4     | 0 – 60                    | 0 – 100                  |
| 26 – 50   | 26 – 37                                | 51 – 80                               | 36 – 50                   | 4.5 – 6.4   | 61 – 106                  | 101 – 200                |
| 51 – 100  | 38 – 50                                | 81 – 120                              | 51 – 70                   | 6.5 – 9.0   | 107 – 170                 | 201 – 300                |
| 101 – 200 | 51 – 90                                | 121 – 180                             | 71 – 120                  | 9.1 – 30    | 171 – 340                 | 301 – 400                |
| > 200     | > 90                                   | > 180                                 | > 120                     | > 30        | > 340                     | > 400                    |

ที่มา: กรมควบคุมมลพิษ (2564)

### 2.2.3 วิธีการตรวจวัดคุณภาพอากาศในประเทศไทย

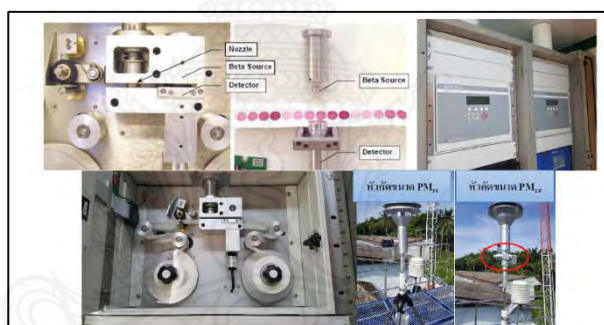
ประเทศไทยได้กำหนดค่ามาตรฐานฝุ่นละอองขนาดไม่เกิน 2.5 ไมครอน ในบรรยากาศทั่วไป ในปี พ.ศ. 2553 โดยค่าเฉลี่ยความเข้มข้นภายในระยะเวลา 24 ชั่วโมง จะต้องไม่เกิน 0.05 mg/m<sup>3</sup> และค่าเฉลี่ยภายในระยะเวลา 1 ปี จะต้องไม่เกิน 0.025 mg/m<sup>3</sup> ทั้งนี้ได้กำหนดวิธีตรวจวัดคุณภาพอากาศที่ยอมรับในระดับสากล และให้ค่าเฉลี่ยของฝุ่นละอองในบรรยากาศที่เป็นมาตรฐานเดียวกัน สำหรับวิธีการตรวจวัดด้วยวิธีอื่น ๆ นั้น กรมควบคุมมลพิษได้เห็นชอบและกำหนดไว้ในประกาศของกรมควบคุมมลพิษ มีรายละเอียดดังตาราง 2.7

จากประกาศคณะกรรมการสิ่งแวดล้อมแห่งชาติ เรื่องการกำหนดมาตรฐานฝุ่นละอองขนาดไม่เกิน 2.5 ไมครอน ในบรรยากาศทั่วไป กรมควบคุมมลพิษ, (2562) กรมควบคุมมลพิษ, (2565) ได้ระบุวิธีการตรวจวัดค่าเฉลี่ยของฝุ่น PM 2.5 ไว้ดังนี้ 1) Beta Radiation Attenuation หรือ Beta Ray Attenuation วิธีการตรวจวัดระบบดูดกลืนรังสีเบต้าเป็นการตรวจวัดอนุภาคฝุ่นอย่างต่อเนื่องเมื่อดูดอากาศเข้ามาในระบบอนุภาคฝุ่นจะตกลงบนกระดาษกรอง โดยมีรังสีเบต้าฉายผ่านกระดาษกรอง สามารถวิเคราะห์ปริมาณอนุภาคฝุ่นได้จากความสัมพันธ์ของปริมาณรังสีที่เดินทางผ่านกระดาษกรองเทียบกับรังสีที่ฉาย โดยภาพ 2.8 แสดงตัวอย่างของเครื่องตรวจวัด 2) Tapered Element Oscillating Microbalance: TEOM เป็นอุปกรณ์เก็บตัวอย่างที่กระดาษกรองจะถูกติดอยู่ที่ปลายของ Tapered Quartz Tube ซึ่งเป็นเหมือนส้อมเสียงที่ไวต่อการเปลี่ยนแปลงความถี่ในการสั่นของฝุ่นบนกระดาษกรอง โดยภาพ 2.9 เป็นการแสดงขั้นตอนการทำงานของระบบ 3) Light Scattering ซึ่งภายในเครื่อง Light Scattering จะมีแหล่งกำเนิดแสงเลเซอร์ทำหน้าที่เป็นส่วนยิงแสงให้กระทบกับอนุภาคขนาดต่างๆ โดยภาพ 2.10 แสดงเครื่อง Light Scattering ของกรมควบคุมมลพิษ 4) Dichotomus Air Sample การตรวจวัดฝุ่นด้วยอุปกรณ์เก็บตัวอย่างแบบ Dichotomus Air Sample เป็นอุปกรณ์ที่ใช้ในการเก็บฝุ่น แยกเก็บเป็น 2 ขนาด ซึ่งตรวจวัดขนาดฝุ่นที่เล็กกว่า 2.5 ไมครอน และอนุภาคฝุ่นระหว่าง 2.5-10 ไมครอน โดยใช้เครื่องมือเครื่องเดียวกัน เครื่องนี้มีขนาดเล็กกว่าเครื่อง High-Volume และใช้อากาศน้อยกว่า โดยใช้อากาศผ่านหัวคัดขนาดของฝุ่นไม่เกิน 10 ไมครอน ให้ตกกระทบกับอุปกรณ์คัดแยกฝุ่นละออง (Virtual Impactor) เพื่อแยกฝุ่นละอองออกเป็นสองแบบ คือ ขนาดไม่เกิน 2.5 ไมครอน และขนาดตั้งแต่ 2.5 ไมครอนขึ้นไป แต่ไม่เกิน 10 ไมครอน จากนั้น นำแผ่นกรองมาชั่งน้ำหนักของฝุ่นละออง โดย ภาพ 2.11 แสดงขั้นตอนการเก็บอนุภาคฝุ่นด้วยเครื่อง Dichotomus Air Sample อย่างไรก็ตาม ยังคงมีการตรวจวัดฝุ่นละอองในบรรยากาศ ระบบกราวิเมตริก (Gravimetric) ด้วยเครื่องเก็บตัวอย่าง High Volume Air Sampler) ซึ่งคือการวัดค่าฝุ่นโดยการดูดอากาศผ่านแผ่นกรอง ซึ่งสามารถกรองฝุ่นละอองขนาด 0.3 ไมครอน ได้ร้อยละ 99 แล้วหาน้ำหนักของฝุ่นละอองที่ค้างบนแผ่นกรองอีกเช่นกัน

ตาราง 2.7 วิธีการตรวจวัดฝุ่นละอองในบรรยากาศของประเทศไทย

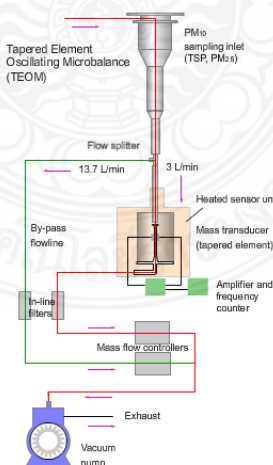
| สารมลพิษทางอากาศ                             | วิธีการวัด/วิธีตรวจวัดตามประกาศคณะกรรมการสิ่งแวดล้อมแห่งชาติ          | วิธีการตรวจวัดตามประกาศกรมควบคุมมลพิษ   |
|--|---|---|
| ฝุ่นละอองขนาดเล็กไม่เกิน 2.5 ไมครอน (PM 2.5) | วิธีการตรวจวัดมาตรฐาน Federal Reference Method (FRM) ตาม US.EPA กำหนด | วิธีการตรวจวัดค่าเฉลี่ยของฝุ่น PM 2.5 ดังนี้<br>1. Beta Radiation Attenuation: Beta Ray<br>2. Tapered Element Oscillating Microbalance: TEOM<br>3. Light Scattering<br>4. Dichotomus Air Sample |

ที่มา: กรมควบคุมมลพิษ (2564)



ภาพ 2.8 วิธีการตรวจวัดแบบ Beta Radiation Attenuation หรือ Beta Ray Attenuation

ที่มา: กรมควบคุมมลพิษ (2564)

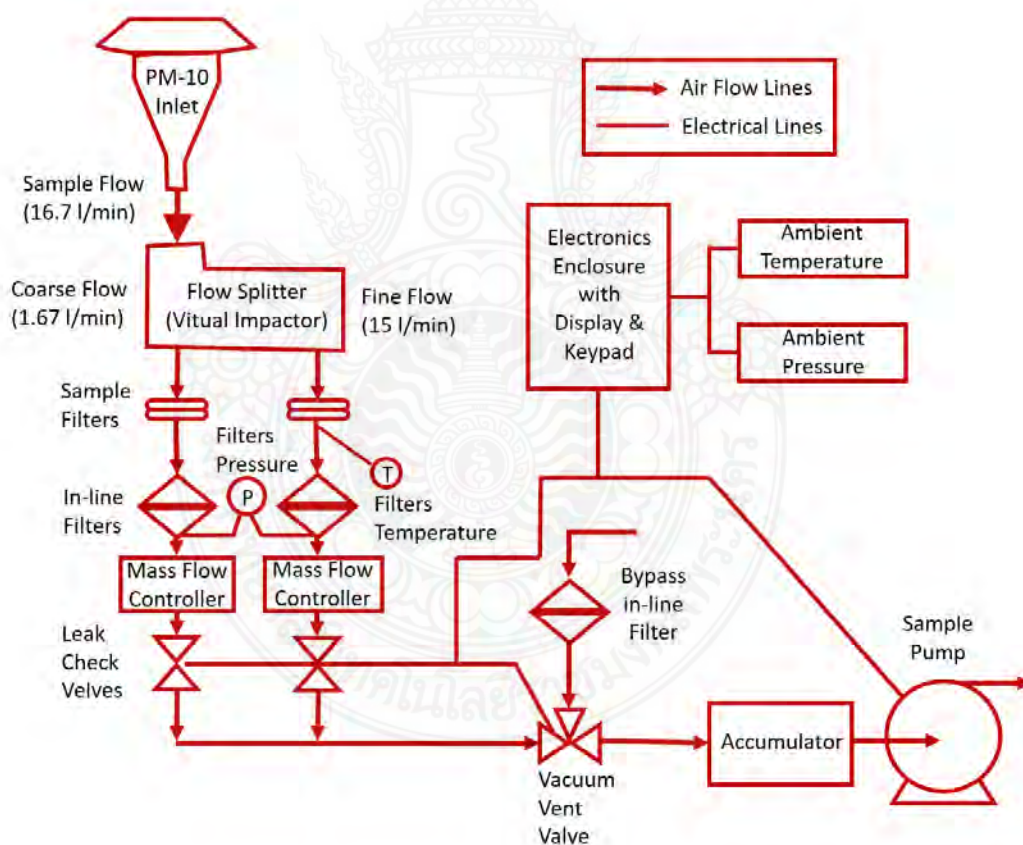


ภาพ 2.9 ระบบการทำงานของ Tapered Element Oscillating Microbalance

ที่มา: กรมควบคุมมลพิษ (2564)



ภาพ 2.10 เครื่อง Light Scattering ของกรมควบคุมมลพิษ  
ที่มา: กรมควบคุมมลพิษ (2564)



ภาพ 2.11 การเก็บอนุภาคฝุ่นด้วย Dichotomus Air Sample  
ที่มา: กรมควบคุมมลพิษ (2564)

## 2.2.4 สถานีตรวจวัดคุณภาพอากาศของกรมควบคุมมลพิษ

ปัจจุบันมีสถานีตรวจวัดทั้งหมด 64 สถานี ในพื้นที่ 34 จังหวัด ซึ่งกระจายอยู่ในเขตพื้นที่ กรุงเทพมหานครและปริมณฑล 23 สถานี ภาคเหนือ 15 สถานี ภาคตะวันออก 11 สถานี ภาคภาคใต้ 6 สถานี ตะวันออกเฉียงเหนือ 3 สถานี ภาคกลางและภาคตะวันตก 6 สถานี

สถานีตรวจวัดคุณภาพอากาศของกรมควบคุมมลพิษ มีโครงสร้างของสถานีเป็นตู้คอนเทนเนอร์ ขนาดโดยประมาณกว้าง 3 เมตร ยาว 4 เมตร สูง 2.4 เมตร ตั้งบนฐานคอนกรีตขนาด 5 เมตร ยาว 5 เมตร สถานีตรวจวัดที่ทำการติดตั้งมี 2 ลักษณะ คือ สถานีตรวจวัดคุณภาพอากาศริมเส้นทางจราจร (Road side) อยู่ริมถนนหรือห่างจากถนนหลักไม่เกิน 10 เมตร และสถานีตรวจวัดคุณภาพอากาศในชุมชน ซึ่งจะอยู่ห่างจากถนนหลักประมาณ 50 เมตรขึ้นไป และต้องเป็นพื้นที่เปิดเพื่อตรวจวัดคุณภาพในพื้นที่ทั่วไปหรือเป็นตัวแทนของแหล่งที่พกอาศัยของประชาชน

พื้นที่ข้างสถานีฯ จะติดตั้งเสาสำหรับติดตั้งเครื่องมือตรวจวัดสภาพอากาศทางอุตุนิยมวิทยา ซึ่งมีความสูงประมาณ 10 เมตร คอมพิวเตอร์ของเครื่องปรับอากาศ จำนวน 2 เครื่อง และจะติดตั้งรั้วโดยรอบสถานีฯ โดยมีการตรวจวัดสภาพอากาศด้านอุตุนิยมวิทยา ประกอบด้วย 1) อุณหภูมิของอากาศ 2) ทิศทางของกระแสลม 3) ความเร็วของกระแสลม 4) ความดันของบรรยากาศ 5) ปริมาณน้ำฝน และ 6) ปริมาณการแผ่รังสีของแสงจากดวงอาทิตย์ และแหล่งอื่นๆ

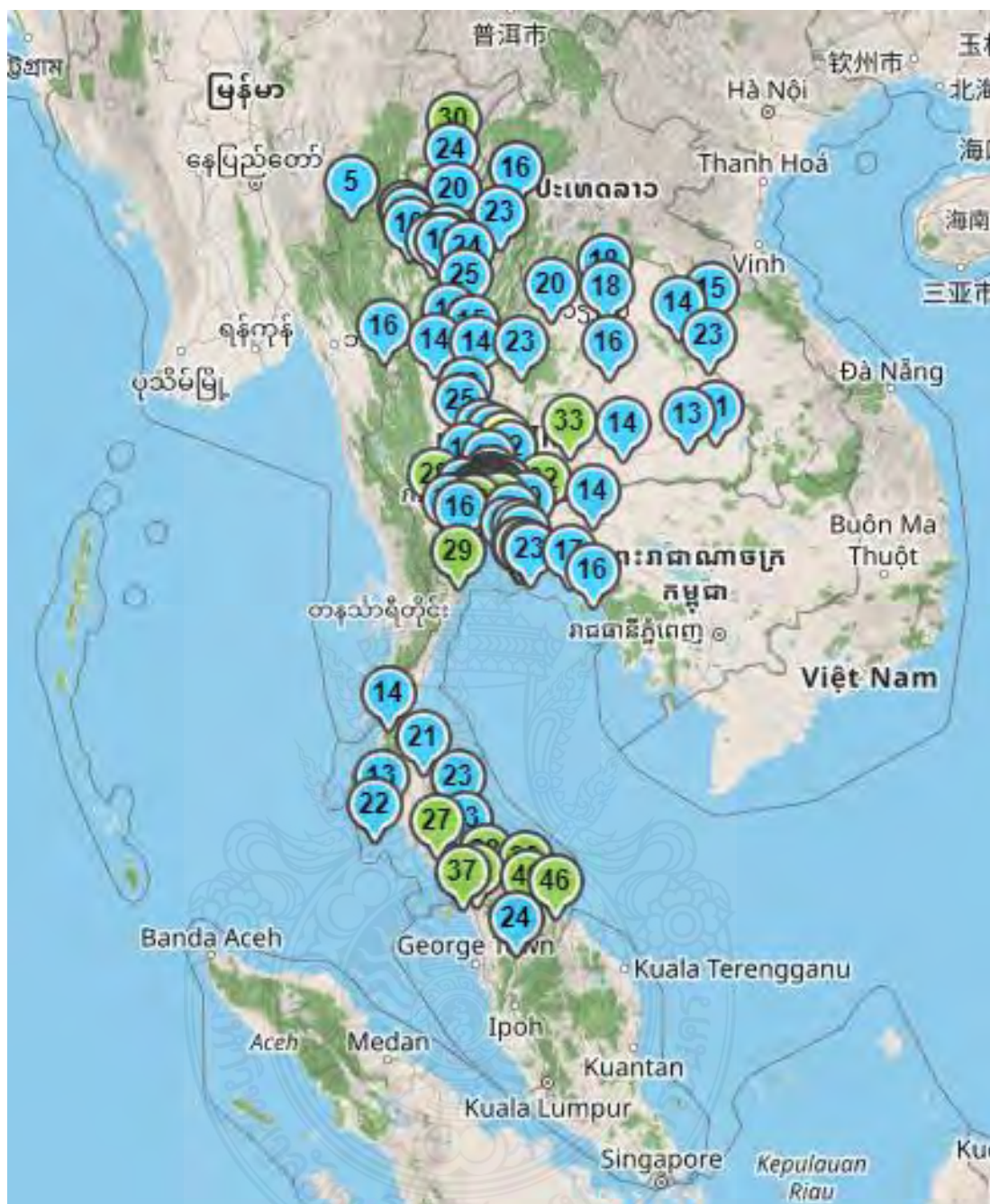
ภายในสถานีตรวจวัดมีการติดตั้งเครื่องตรวจวัดคุณภาพอากาศและอุปกรณ์เครื่องตรวจวัดสภาพอากาศทางอุตุนิยมวิทยา เครื่องวัดระดับเสียงบางสถานี และเครื่องวิเคราะห์รวมถึงบันทึกผลการตรวจวัด ซึ่งมีระบบการทำงานแบบอัตโนมัติ ซึ่งภายในสถานีต้องมีการควบคุมอุณหภูมิให้เหมาะสมต่อการทำงานของเครื่องมือ ตามคำแนะนำจากผู้ผลิต

สารมลพิษในบรรยากาศที่ทำการตรวจวัด ประจำสถานี ประกอบด้วย

- 1) ฝุ่นละออง ได้แก่
  - ฝุ่นละอองขนาดใหญ่ตั้งแต่ 100 ไมครอนลงมา
  - ฝุ่นละอองขนาดเล็กตั้งแต่ 10 ไมครอนลงมา (PM 10)
  - ฝุ่นละอองขนาดเล็กตั้งแต่ 2.5 ไมครอนลงมา (PM 2.5)
- 2) ก๊าซคาร์บอนมอนนอกไซด์ (CO)
- 3) ก๊าซซัลเฟอร์ไดออกไซด์ (SO<sub>2</sub>)
- 4) ก๊าซออกไซด์ของไนโตรเจน (NO NO<sub>2</sub> NO<sub>x</sub>)
- 5) ก๊าซไฮโดรคาร์บอน (CH)
- 6) ก๊าซโอโซน (O<sub>3</sub>)

โดยสถานีตรวจวัดคุณภาพอากาศของกรมควบคุมมลพิษ ทั่วประเทศแสดงดังภาพ 2.12





ภาพ 2.12 สถานีตรวจวัดคุณภาพอากาศของกรมควบคุมมลพิษ

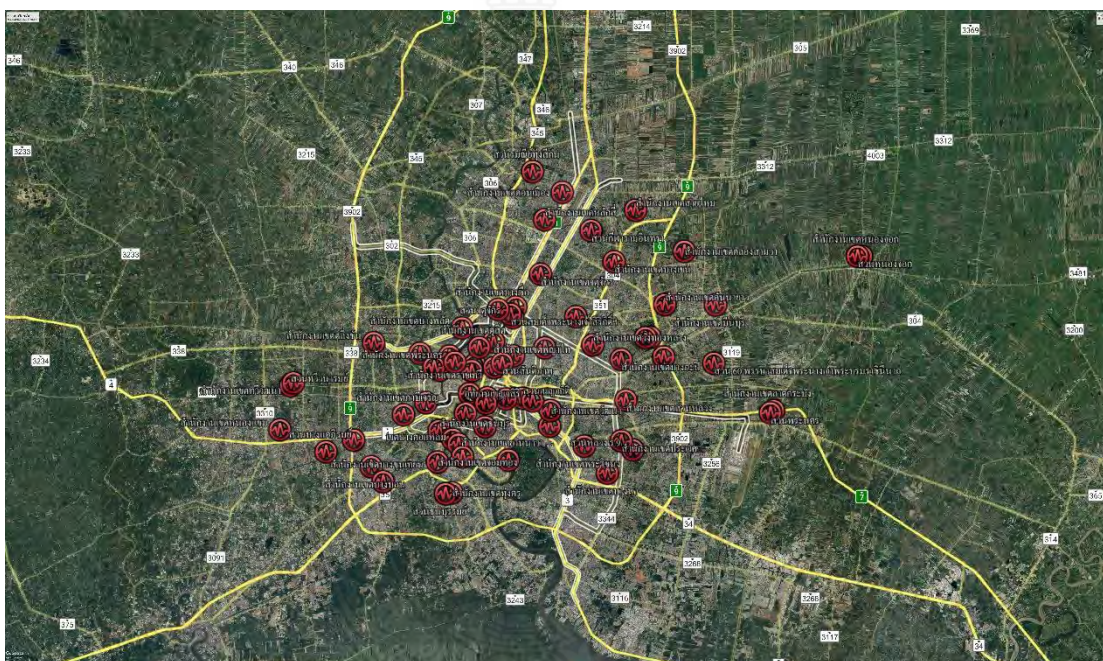
### 2.2.5 สถานีตรวจวัดของสำนักสิ่งแวดล้อม กรุงเทพมหานคร

ภายในสถานีฯ มีการติดตั้งเครื่องตรวจวัดคุณภาพอากาศและอุปกรณ์เครื่องตรวจวัดสภาพอากาศทางอุตุนิยมวิทยา เครื่องวัดระดับเสียงบางสถานี และเครื่องวิเคราะห์รวมถึงบันทึกผลการตรวจวัด ซึ่งใช้ระบบทำงานแบบอัตโนมัติ ซึ่งจะต้องมีการควบคุมอุณหภูมิภายในสถานีให้เหมาะสมต่อการทำงานของเครื่องมือ

สารมลพิษในอากาศที่จะทำการตรวจวัดของแต่ละสถานี ประกอบด้วย

- 1) ฝุ่นละออง ได้แก่
  - ฝุ่นละอองขนาดใหญ่ตั้งแต่ 100 ไมครอนลงมา
  - ฝุ่นละอองขนาดเล็กตั้งแต่ 10 ไมครอนลงมา (PM 10)
  - ฝุ่นละอองขนาดเล็กตั้งแต่ 2.5 ไมครอนลงมา (PM 2.5)
- 2) ก๊าซคาร์บอนมอนอกไซด์ (CO)
- 3) ก๊าซซัลเฟอร์ไดออกไซด์ (SO<sub>2</sub>)
- 4) ก๊าซออกไซด์ของไนโตรเจน (NO NO<sub>2</sub> NO<sub>x</sub>)
- 5) ก๊าซไฮโดรคาร์บอน (CH)
- 6) ก๊าซโอโซน (O<sub>3</sub>)

โดยสถานีตรวจวัดคุณภาพอากาศของกรมควบคุมมลพิษ ทั่วประเทศแสดงดังภาพ 2.13



ภาพ 2.13 สถานีตรวจวัดคุณภาพอากาศของสำนักสิ่งแวดล้อม กรุงเทพมหานคร

## 2.3 แนวคิดและทฤษฎีที่เกี่ยวกับการพัฒนาแบบจำลอง

แบบจำลองทางสถิติ (Statistical Model) เป็นการจำลองสถานการณ์จากหลักการและการคำนวณทางสถิติ โดยส่วนใหญ่แล้วจะนำไปใช้ในการวิเคราะห์ข้อมูลทั้งในอดีตจนถึงปัจจุบัน เพื่อทำนาย พยากรณ์ หรือคาดการณ์เหตุการณ์ที่เป็นไปได้ ที่จะเกิดขึ้นในอนาคต

แบบจำลองเชิงพยากรณ์ คือ การนำเอาคณิตศาสตร์มาช่วยทำนายและคาดการณ์สถานการณ์ที่จะเป็นไปในอนาคต ซึ่งจะมีส่วนช่วยในการตัดสินใจ จากผลลัพธ์ที่ได้จากเหตุการณ์จากต่าง ๆ โดยแบบจำลองที่นิยมใช้ในปัจจุบัน คือ แบบจำลองการวิเคราะห์การถดถอย

### 2.3.1 แบบจำลองการวิเคราะห์การถดถอย (Regression Analysis Model)

การวิเคราะห์การถดถอย (Regression Analysis) คือ การคำนวณสถานการณ์ทางสถิติผ่านการศึกษความสัมพันธ์ระหว่างตัวแปรต่าง ๆ โดยสมการที่ได้มาจะนำมาใช้สำหรับการประมาณค่าตัวแปรหนึ่งจากตัวแปรอื่น ๆ ที่ทราบค่า ตั้งแต่ 2 ตัวแปรขึ้นไป ประกอบด้วย

- 1) ตัวแปรอิสระหรือตัวแปรต้น (Independent Variable) เป็นตัวแปรที่สมมติว่ามีผลต่อตัวแปรที่ต้องการประมาณค่า แทนด้วย  $x$
- 2) ตัวแปรที่ต้องการทราบหรือตัวแปรตาม (Dependent Variable) คือตัวแปรที่ต้องการรู้หรือต้องการประมาณค่าขึ้นมา แทนด้วย  $y$

ซึ่งการวิเคราะห์ด้วย Regression จะแทนด้วยสมการหรือฟังก์ชันทางคณิตศาสตร์ ดังสมการ (2.2)

$$y = f(x) \quad (2.2)$$

หรือ

$$y = ax + b$$

โดยที่

- |     |  |
|-----|--|
| $x$ | คือ ตัวแปรต้น (Independent Variable)                               |
| $y$ | คือ ตัวแปรตาม (Dependent Variable)                                 |
| $a$ | คือ ค่าคงที่ของสมการถดถอย ซึ่งเป็นค่าจุดตัด (Intercept) แกนของ $y$ |
| $b$ | คือ ค่าสัมประสิทธิ์การถดถอย (Regression Coefficient) ของ $x$       |

การวิเคราะห์การถดถอยสามารถแบ่งออกเป็น 2 ลักษณะ คือ

- 1) การวิเคราะห์ถดถอยอย่างง่าย (Simple Regression Analysis)

การวิเคราะห์การถดถอยอย่างง่ายเป็นการศึกษาเพื่อหาสมการ ซึ่งแสดงความสัมพันธ์ระหว่างตัวแปรตาม ( $y$ ) 1 ตัวแปร กับตัวแปรอิสระ ( $x$ ) 1 ตัวแปร เท่านั้น ยกตัวอย่างเช่น ในการคาดการณ์ปริมาณฝุ่นละออง PM 2.5 ถ้าผู้วิจัยประเมินว่าปริมาณฝุ่นละออง PM 2.5 ขึ้นอยู่กับปริมาณจราจรเพียงอย่างเดียว โดยที่ปริมาณจราจรจะมีผลทำให้ค่า PM 2.5 เพิ่มขึ้นหรือต่ำลงได้ ดังนั้น จึง

เลือกใช้การวิเคราะห์การถดถอยอย่างง่าย โดยให้ปริมาณฝุ่นละออง PM 2.5 เป็นตัวแปรตาม ( $y$ ) และปริมาณจราจร เป็นตัวแปรอิสระ ( $x$ ) ดังสมการที่ 2.3

$$y = \beta_0 + \beta_1 x + u \quad (2.3)$$

โดยที่

|           |   |
|-----------|---|
| $x$       | คือ ตัวแปรต้น (Independent Variable)                |
| $y$       | คือ ตัวแปรตาม (Dependent Variable)                  |
| $\beta_0$ | คือ ค่าคงที่ (Intercept)                            |
| $\beta_1$ | คือ ค่าสัมประสิทธิ์การถดถอย (Coefficient Parameter) |
| $u$       | คือ ค่าความคลาดเคลื่อน                              |

## 2) การวิเคราะห์ถดถอยพหุคูณ (Multiple Regression Analysis)

การวิเคราะห์การถดถอยพหุคูณเป็นวิธีการหรือเครื่องมือที่ถูกใช้กันทั่วไป และนิยมใช้อย่างแพร่หลาย ซึ่งแสดงความสัมพันธ์ระหว่างตัวแปรตาม ( $y$ ) 1 ตัวแปร กับตัวแปรอิสระ ( $x$ ) มากกว่า 1 ตัวแปร ยกตัวอย่างเช่น ในการคาดการณ์ปริมาณฝุ่นละออง PM 2.5 ปริมาณฝุ่นละออง PM 2.5 ขึ้นอยู่กับหลายปัจจัย เช่น ปริมาณจราจร ความกดอากาศ อุณหภูมิ และความเร็วลม เป็นต้น ดังนั้นจึงเลือกใช้การวิเคราะห์แบบพหุคูณ โดยให้ปริมาณฝุ่นละออง PM 2.5 เป็นตัวแปรตาม ( $y$ ) ซึ่งขึ้นอยู่กับตัวแปรอิสระหลายตัวแปร ได้แก่ ปริมาณจราจร ( $x_1$ ) ความกดอากาศ ( $x_2$ ) อุณหภูมิ ( $x_3$ ) และความเร็วลม ( $x_4$ ) ดังสมการที่ 2.4-2.5

$$y = \beta_0 + \beta_1 x + u \quad (2.4)$$

โดยที่

|           |   |
|-----------|---|
| $x$       | คือ ตัวแปรต้น (Independent Variable)                |
| $y$       | คือ ตัวแปรตาม (Dependent Variable)                  |
| $\beta_0$ | คือ ค่าคงที่ (Intercept)                            |
| $\beta_1$ | คือ ค่าสัมประสิทธิ์การถดถอย (Coefficient Parameter) |
| $u$       | คือ ค่าความคลาดเคลื่อน                              |

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u \quad (2.5)$$

โดยที่

|                        |  |
|------------------------|--|
| $x_1, x_2, \dots, x_k$ | คือ ตัวแปรต้นต่าง ๆ (Independent Variable) |
| $y$                    | คือ ตัวแปรตาม (Dependent Variable)         |
| $\beta_0$              | คือ ค่าคงที่ (Intercept)                   |

$\beta_1, \beta_2, \dots, \beta_k$  คือ ค่าสัมประสิทธิ์การถดถอย (Coefficient Parameter)  
 $u$  คือ ค่าความคลาดเคลื่อน

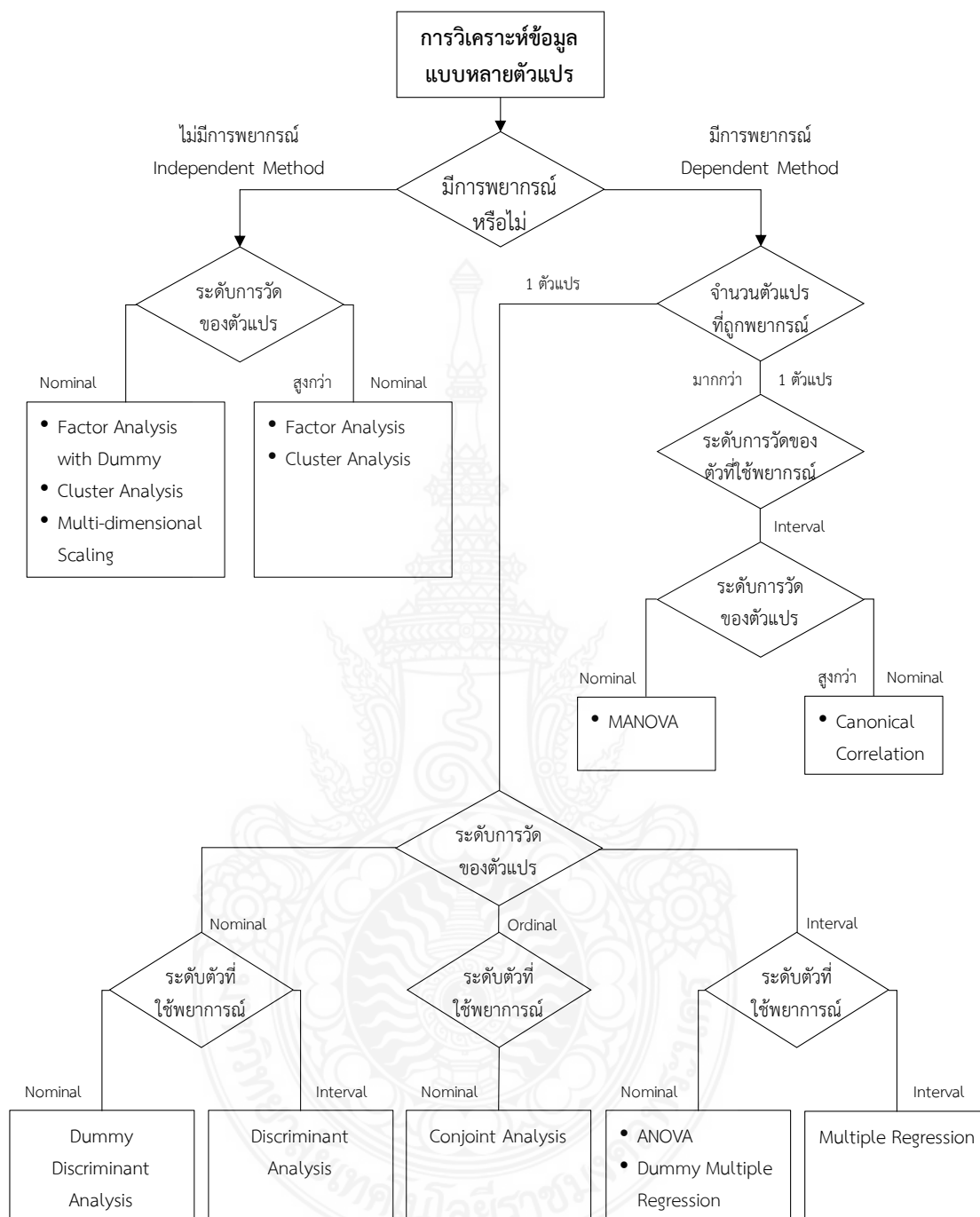
### 2.3.2 แบบจำลองการวิเคราะห์พหุตัวแปร (Multivariate Analysis Model)

การวิเคราะห์พหุตัวแปร (Multivariate Analysis) เป็นเครื่องมือทางสถิติที่สำคัญ เครื่องมือหนึ่งในการวิเคราะห์ข้อมูลเชิงปริมาณ (Quantitative Data Analysis) เพื่อให้เข้าใจความสัมพันธ์และความหมายของข้อมูลแต่ละชุดอย่างเป็นระบบ โดยการวินิจฉัยนั้นจะไม่ค่อนข้างเที่ยงตรงและมีความน่าเชื่อถือได้เป็นประจักษ์นิยม (Empirical Generalization)

การวิเคราะห์พหุตัวแปรเป็นการวิเคราะห์ตัวแปรตั้งแต่ 2 ตัว เพื่อที่จะต้องการหาความสัมพันธ์ของสถานการณ์ที่เกิดขึ้นในห้วงเวลาเดียวกัน ซึ่งนำไปใช้ในการทำความเข้าใจตัวแปรชุดนั้นๆ นอกจากนี้ ยังช่วยในการจัดกลุ่มชุดของตัวแปรที่มีความสัมพันธ์ต่อกัน และทำนายค่าตัวแปรตัวใหม่ โดยอาศัยความสัมพันธ์ที่มีอยู่จากการวิเคราะห์เดิม และค่าระดับนัยสำคัญ (Level of Significance) หรือโอกาสที่จะเกิดความผิดพลาด ซึ่งแปรผันตามความน่าเชื่อนั่นเอง

การวิเคราะห์ในลักษณะนี้ จะต้องคำนึงถึงความเหมาะสมของการตั้งสมมติฐาน ความพร้อมของข้อมูลและความเหมาะสมของตัวแปรที่สัมพันธ์กับสมมติฐานและวัตถุประสงค์ ดังแสดงภาพรวมในภาพ 2.14





ภาพ 2.14 การวิเคราะห์พหุตัวแปร (Multivariate Analysis Model)

### 2.3.3 แนวคิดและทฤษฎีเกี่ยวกับ Machine Learning

#### 1) การเรียนรู้ของเครื่อง (Machine Learning)

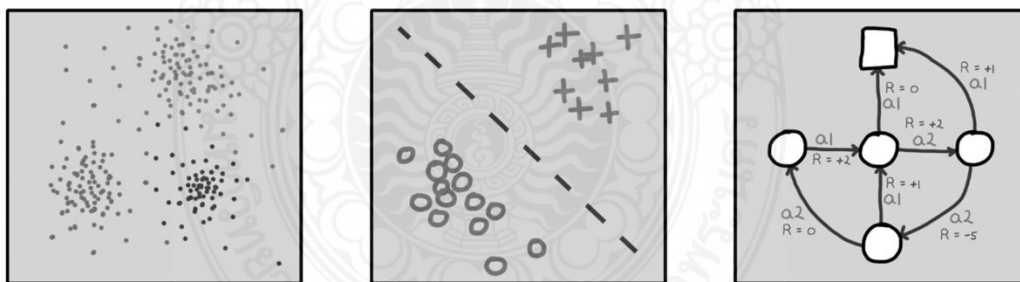
Machine Learning คือ การทำให้คอมพิวเตอร์เรียนรู้ได้ด้วยตนเองผ่านการป้อนข้อมูล ซึ่งจะแตกต่างกับการเขียนโปรแกรมคอมพิวเตอร์โดยทั่วไป เพราะการเขียนโปรแกรมจะต้องมีองค์ประกอบทั้งข้อมูลและตัวโปรแกรมที่ป้อนเข้าไป เพื่อให้ได้ค่าที่เป็นผลลัพธ์ออกมา แต่ Machine Learning ใส่ชุดข้อมูล และคำตอบไปให้ และจะให้ ML หาโปรแกรมที่สามารถคำนวณผลลัพธ์ นั้นๆ ได้ สำหรับนำไปใช้คาดการณ์ผลลัพธ์ในอนาคต

Machine Learning เรียนรู้จากสิ่งที่ส่งเข้าไป แล้วจดจำข้อมูลไว้ ส่งผลลัพธ์ออกมา หรือ code ที่ส่งต่อไปแสดงผล Machine Learning ถูกใช้งานในหลายรูปแบบ ต้องอาศัยโปรแกรม หรือ Algorithm หลายแบบ โดยมีนักวิเคราะห์ข้อมูลเป็นผู้ออกแบบ สร้างแบบจำลองเพื่อทำนายหรือพยากรณ์สิ่งที่จะเกิดขึ้นในอนาคต โดยใช้ข้อมูลเป็นเครื่องมือ

Machine Learning เป็นศาสตร์การเรียนรู้ของ AI กล่าวคือ AI สามารถทำงานได้โดยอาศัย Machine Learning เรียนรู้จากสิ่งที่ส่งไปให้และแสดงผลลัพธ์ออกมาเป็นชุดข้อมูล โดยอาศัยโปรแกรม Algorithm ในการประมวลผล

#### 2) ประเภทของ Machine Learning

Machine Learning แบ่งออกเป็นรูปแบบการเรียนรู้ 3 แบบ ได้แก่ การเรียนรู้แบบมีการสอน (Supervised Learning) การเรียนรู้แบบไม่ต้องสอน (Unsupervised Learning) และการเรียนรู้แบบเสริมแรง (Reinforcement Learning)



(ก.) Unsupervised Learning (ข.) supervised Learning (ค.) Reinforcement Learning

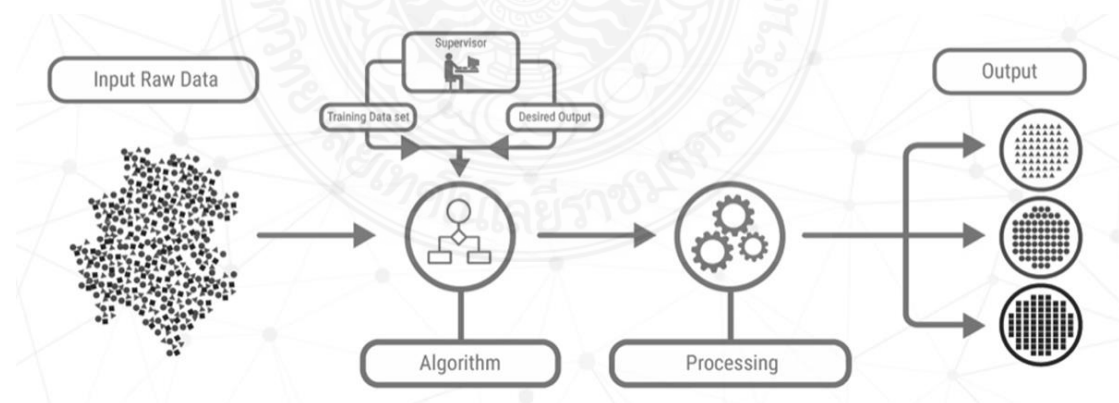
ภาพ 2.15 ประเภทของ Machine Learning

## 2.1) Supervised Learning

การเรียนรู้โดยมีข้อมูล (Data) มาสอน คือการทำให้คอมพิวเตอร์ค้นหาคำตอบของปัญหาได้ด้วยตัวเอง หลังจากการเรียนรู้จากข้อมูลตัวอย่างที่ส่งเข้าไปแล้วระยะหนึ่ง เช่น การแยกหมา แมว โดยนักวิเคราะห์ข้อมูลจะต้องสร้างแบบจำลอง ที่จะทำให้คอมพิวเตอร์รู้จักหมา แมวเข้าไป โดยเอาข้อมูลหมาแมวใส่เข้าไป เช่น ลักษณะของแมวแต่ละตัวเข้าไป สีของแมวแต่ละตัว โดยแปลงให้เป็นภาษาคอมพิวเตอร์ก่อน คือการแปลงเป็นตัวเลข เรียกว่า Features พร้อมเฉลยไว้เลยวันนี้คือ แมว โดยใส่ข้อมูลเป็นตัวเลข เรียกว่า labels และก็เอาหมาใส่เข้าไปพร้อมเฉลยวันนี้คือหมา เมื่อใส่ Input เสร็จแล้ว Data Scientist ก็จะ Train Model เพื่อให้คอมพิวเตอร์แยกแยะหมาแมวได้ตาม Features โดย Machine Learning ที่ใช้โดยทั่วไป ประกอบไปด้วย Support Vector Machine, Gradient Boosting (XGBoost), Naive Bayes, Classification Trees หรือ Random Forest เป็นต้น

หลักการของ Supervised Learning สามารถนำไปประยุกต์ใช้แก้ปัญหาได้ 2 รูปแบบ ดังนี้ 1) Classification เป็นการสอนให้คอมพิวเตอร์เรียนรู้จากการใส่ข้อมูลลงไป เช่น สมมติชี้ภาพสัตว์ให้เด็กที่ไม่เคยเห็นแล้วบอกว่าสัตว์ตัวไหน คือ แมว ตัวไหนไม่ใช่แมว หรือ บอกว่าสัตว์ตัวนั้นคือ แมว ตัวนั้น คือ หมา ให้เห็นหลายๆ ตัว ทำแบบนี้ไปเรื่อยๆ แล้วให้เด็กตอบเอง ถ้าเด็กตอบถูกก็แปลว่า การสอนมีประสิทธิภาพ เรียกกระบวนการสอนแบบนี้ว่า Classification และ 2) Regression เป็นการสอนให้คอมพิวเตอร์รู้จากการใส่ข้อมูลลงไป เช่น สมมติหีบเพชร สีเหลืองมา แล้วบอกเด็กว่าราคา 1 ล้านบาท และสีฟ้า ราคา 2 ล้านบาท ทำแบบนี้ไปหลายๆ เม็ด จนเด็กเกิด Model หรือ Logic ในการที่จะคาดเดาราคาเพชรได้เอง จนลองหีบเพชรเม็ดใหม่ขึ้น เด็กก็จะคาดเดาราคาได้เลย เรียกกระบวนการสอนแบบนี้ว่า Regression

การใช้ Supervised Learning มีรูปแบบการเขียนโปรแกรมที่แตกต่างออกไป ในช่วงแรกจะเขียนโปรแกรมให้คอมพิวเตอร์สร้าง Model หรือ Logic ของโปรแกรมขึ้นมาเอง จากข้อมูลขาเข้าและผลลัพธ์ที่ต้องการ จากนั้นจึงนำ Model มาใช้ ดังนั้นยังมีข้อมูลขาเข้าและผลลัพธ์ที่มีความหลากหลาย และมีปริมาณมาก ก็จะทำให้ได้ Model ที่มีความแม่นยำมากขึ้น



ภาพ 2.16 Machine Learning ประเภท Supervised Learning



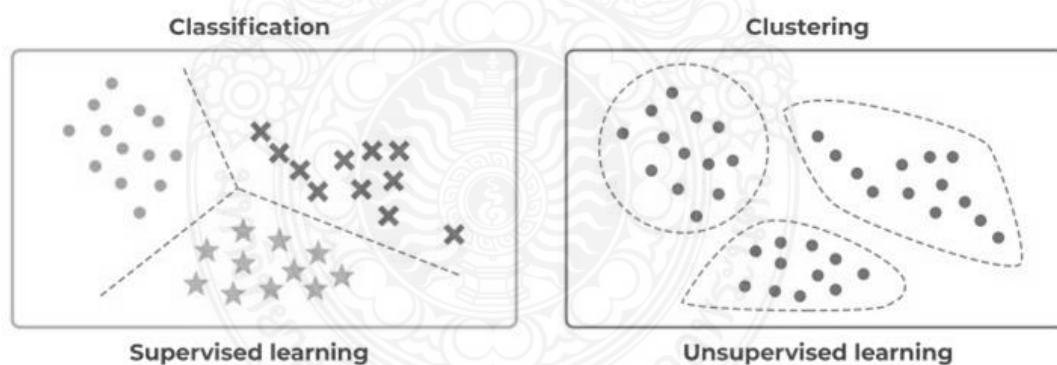
## 2.2) Unsupervised Learning

การเรียนรู้และพยากรณ์ผลได้จากการจำแนกและสร้างแพทเทิร์นของตัวเองจากข้อมูลที่ได้รับ จะตรงข้ามกับ Supervised Learning คือ ไม่มีข้อมูลมาสอน ไม่บอกว่า คือ หมาหรือแมว แต่จะให้แยกแยะเอง แบบจำลองที่ใช้กัน คือ K-nearest Neighbor, K-mean ดังแสดงความแตกต่างของ Supervised Learning และ Unsupervised Learning ในภาพ 2.17

## 2.3) Reinforcement Learning

Reinforcement Learning เป็น Machine Learning Algorithm แบบหนึ่ง ที่ทำงานคล้ายกับการที่มนุษย์เรียนรู้ด้วยการลองผิดลองถูก โดย Reinforcement Learning ประกอบด้วย 1) Agent – ตัวแทน Action 2) Action (a) – การทำของ Agent ที่ส่งผลกระทบต่อสิ่งแวดล้อม (Environment) 3) สิ่งแวดล้อม หรือ Environment (e) – ระบบที่ Agent ต้องสื่อสารระหว่างกัน 4) State (s) – หรือ สถานการณ์ของสิ่งแวดล้อม ที่ทาง เอเจนต์สามารถเข้าใจได้ 5) Policy ( $\pi$ ) – หลักการที่เอเจนต์ใช้ในการเลือกการกระทำ หรือ Action หลังจากทราบสถานการณ์แล้ว 6) Reward (R) – ตัวประเมินค่าตอบที่เกิดจากการกระทำของเอเจนต์ เช่น คะแนน รางวัล หรือ เหตุการณ์ที่เดาได้ถูกต้อง เป็นต้น

โดยวิธีการเรียนของ Reinforcement Learning คือ การทำความเข้าใจเอเจนต์ (Agent) ที่เกิดจากการสื่อสารระหว่างกันแบบลองผิดลองถูกระหว่างผู้กระทำ กับ สิ่งแวดล้อม โดยที่ทางตัวของเอเจนต์จะสามารถรู้ถึงสถานการณ์ของสิ่งแวดล้อม ผ่าน State และเลือกการกระทำที่ส่งผลต่อ สภาพแวดล้อม เพื่อให้ได้ Reward ที่ดีที่สุด



ภาพ 2.17 Machine Learning ประเภท UnSupervised Learning

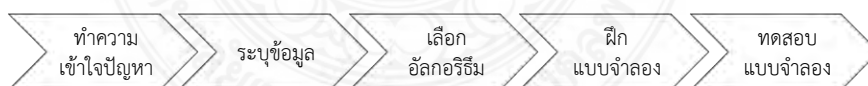
### 2.3.4 แนวคิดและทฤษฎีเกี่ยวกับ Deep Learning

ในส่วนของแนวคิดทฤษฎีเกี่ยวกับ Deep Learning ผู้วิจัยได้ทบทวนมาจากหนังสือ Deep Learning (Ian et al., 2016) ของสถาบัน MIT ในประเทศสหรัฐอเมริกา โดยมีรายละเอียดดังนี้

Deep Learning คือ การเรียนรู้แบบอัตโนมัติผ่านการเลียนแบบการทำงานของสมองมนุษย์ (Neurons) ซึ่งถือเป็นส่วนหนึ่งของ Machine Learning โครงข่ายประสาทเทียมเหล่านี้จะพยายามจำลองพฤติกรรมของสมอง แม้จะห่างไกลจากความสามารถของโครงข่ายประสาทเทียม แต่ก็ทำให้สมอง "เรียนรู้" จากข้อมูลจำนวนมากได้ แม้ว่าโครงข่ายประสาทเทียมที่มี layer เดียวยังสามารถคาดการณ์โดยประมาณได้ แต่ layer ที่ซ่อนอยู่เพิ่มเติมสามารถช่วยเพิ่มความสามารถหรือเพิ่มประสิทธิภาพและสามารถปรับแต่งเพื่อเพิ่มความแม่นยำได้ โดยแสดงกระบวนการ ดังภาพ 2.18

การเรียนรู้ของ Deep Learning การเรียนรู้ของแบบจำลองนั้น จะเกิดขึ้น 2 ระยะ ได้แก่ ระยะที่ 1 เป็นการประยุกต์ใช้การแปลงแบบไม่เป็นเชิงเส้น (Non-linear Transformation) กับข้อมูลที่ได้รับ (Input) ได้ผลลัพธ์ (Output) ออกมาในรูปแบบของแบบจำลองทางสถิติ (Statistical Model) และระยะที่ 2 เป็นการทำซ้ำแล้วซ้ำอีกจนกว่าจะได้แบบจำลองที่มีความแม่นยำ (Accuracy) ในระดับที่น่าพึงพอใจ ซึ่งการทำซ้ำ ๆ ในระยะที่ 2 นี้ เรียกว่า Iteration ปัจจุบันได้มีการนำ Deep Learning มาประยุกต์ใช้อย่างแพร่หลาย เช่น รถยนต์ไร้คนขับ (Driverless Car), สมาร์ทโฟน, การค้นหาของ Google, เครื่องจับเท็จ (Fraud Detector) และอื่น ๆ อีกมากมาย

ในปัจจุบันมีการนำ Deep Learning ไปประยุกต์ใช้อย่างแพร่หลายทั้งการจราจรและขนส่ง เช่น การนับปริมาณจราจร รถยนต์อัตโนมัติ เป็นต้น การแปลภาษา เช่น Google Translate เป็นต้น Chatbots สามารถแก้ปัญหาลูกค้าได้ภายในไม่กี่วินาที Chatbot คือแอปพลิเคชัน AI สำหรับแชทออนไลน์ผ่านข้อความหรือข้อความเป็นคำพูด มีความสามารถในการสื่อสารและดำเนินการคล้ายกับมนุษย์ การประยุกต์ใช้ Deep Learning ในภาคการดูแลสุขภาพ การตรวจหาโรคโดยใช้คอมพิวเตอร์ช่วยและการวินิจฉัยโดยใช้คอมพิวเตอร์ช่วยสามารถทำได้โดยใช้ Deep Learning มีการใช้กันอย่างแพร่หลายสำหรับการวิจัยทางการแพทย์ การค้นคว้ายา และการวินิจฉัยโรคที่คุกคามถึงชีวิต เช่น มะเร็งและเบาหวานขึ้นจอประสาทตา โดยผ่านกระบวนการสร้างภาพทางการแพทย์ เป็นต้น รวมไปถึงด้านความบันเทิง การรวมข่าวและการตรวจจับข่าวปลอม การแต่งเพลง เป็นต้น



ภาพ 2.18 กระบวนการทำงานของ Deep Learning



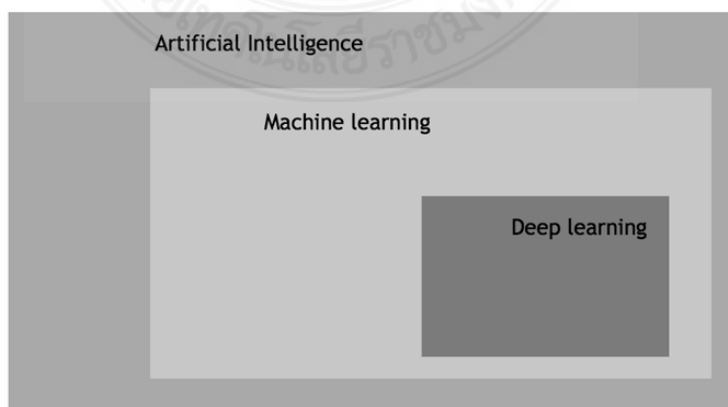
ภาพ 2.19 การแสดงภาพที่ได้ถูกรวบรวมโดย Deep Learning

### 2.3.5 แนวคิดและทฤษฎีเกี่ยวกับปัญญาประดิษฐ์ (Artificial Intelligence) หรือ AI

ในปีคริสต์ศักราช 1956 ศาสตราจารย์ จอห์น แมคคาธิ ได้ให้คำนิยามของปัญญาประดิษฐ์ไว้ว่า “ความมีเหตุผลเชิงคำนวณ” (Stuart and Peter, 2010; Winston, 1992) รากทางความคิดและวิทยาศาสตร์ AI ยังคงถูกพัฒนาต่อเนื่องตลอดประวัติศาสตร์ทางวิทยาศาสตร์ ต่อไปมาจากนั้นก็มีเหตุการณ์สำคัญดังนี้ 1) การท้าทายทางความคิด : การท้าทายในการสร้างเครื่องจักรที่สามารถคำนวณและคิดเลขอัตโนมัติเป็นเรื่องที่ได้รับความสนใจตั้งแต่สมัยคริสต์ศักราช 17 และ 18 โดยนักคณิตศาสตร์และนักวิทยาศาสตร์ 2) การพัฒนาโปรแกรมคอมพิวเตอร์ที่สามารถเล่นเกมซึ่งความสามารถ เช่น หมากรุก เป็นต้น นอกจากนี้ AI ยังหมายถึง ความฉลาดที่เสมือนสร้างขึ้นให้กับสิ่งที่ไม่มีชีวิต แต่สามารถเข้าใจกลไกเพียงบางส่วนที่เฉพาะเจาะจงเท่านั้น ซึ่ง AI ถูกจำแนกออกได้ดังนี้

- (1) ปัญญาประดิษฐ์เชิงแคบ (Narrow AI) หรือที่รู้จักกันในชื่อ Weak AI หมายถึงระบบปัญญาประดิษฐ์ที่ได้รับการออกแบบและฝึกอบรมสำหรับงานเฉพาะหรืองานเฉพาะด้าน ระบบ AI เหล่านี้มีความเป็นเลิศในการทำหน้าที่ที่กำหนดไว้ อย่างชัดเจนและเฉพาะเจาะจง แต่ขาดการเรียนรู้แบบกว้างๆ และความสามารถในการแก้ปัญหาทั่วไปที่เกี่ยวข้องกับความฉลาดของมนุษย์
- (2) ปัญญาประดิษฐ์ทั่วไป (General AI) คือ AI ทั่วไปมีความสามารถในการถ่ายทอดความรู้และทักษะจากโดเมนหนึ่งไปยังอีกโดเมนหนึ่ง แสดงทักษะการใช้เหตุผลและการแก้ปัญหาในบริบทต่างๆ ตลอดจนเข้าใจและประมวลผลภาษาธรรมชาติเหมือนมนุษย์ ไม่จำกัดเฉพาะชุดงานที่กำหนดไว้ล่วงหน้า และสามารถปรับให้เข้ากับสถานการณ์ใหม่ๆ ที่ไม่คุ้นเคยได้
- (3) ปัญญาประดิษฐ์แบบเข้ม (Strong AI) คือ หรือ Artificial General Intelligence (AGI) เป็นรูปแบบขั้นสูงของปัญญาประดิษฐ์ที่มีความสามารถด้านการรับรู้เหมือนมนุษย์ ต่างจาก Narrow AI ที่ได้รับการออกแบบมาเพื่องานเฉพาะ AI ทั่วไปโดดเด่นด้วยความสามารถในการเข้าใจ เรียนรู้ และปฏิบัติงานทางปัญญาที่หลากหลายในระดับที่เทียบเคียงได้กับความฉลาดของมนุษย์

โดย AI มีความสัมพันธ์กับการเรียนรู้ของเครื่อง (Machine Learning) และการเรียนรู้เชิงลึก (Deep Learning) ดังแสดงในภาพ 2.20



ภาพ 2.20 แสดงภาพรวมของชนิด AI

## 2.4 การคาดการณ์มลพิษทางอากาศ ตั้งแต่อดีตจนถึงปัจจุบัน

ในปี 1896, นักวิทยาศาสตร์ชาวสวีเดนชื่อ Svante Arrhenius ได้ทำการคาดการณ์ว่าการปล่อยก๊าซคาร์บอนไดออกไซด์ ( $\text{CO}_2$ ) จากการใช้เชื้อเพลิงในการผลิตพลังงานจะส่งผลให้สภาวะโลกร้อนเกิดขึ้น และมีผลกระทบต่ออุณหภูมิโลก (Arrhenius, 2017) ในปี 1976, สมาคมการธรณีวิทยาและมลพิษทางอากาศในสหราชอาณาจักรได้เผยแพร่รายงานแรกเกี่ยวกับมลพิษทางอากาศและผลกระทบต่อสิ่งแวดล้อม รายงานนี้มีบทสรุปเกี่ยวกับภาวะโลกร้อนและปัญหาอื่น ๆ ที่เกี่ยวข้อง (Silberston, 1993) ในปี 1990 หน่วยงานระดับโลกที่ชื่อว่า IPCC ได้เผยแพร่รายงานการประเมินครั้งแรกเกี่ยวกับการเปลี่ยนแปลงสภาวะอากาศ รายงานนี้รวบรวมข้อมูลวิทยาศาสตร์และประเมินผลกระทบของกิจกรรมมนุษย์ต่อสภาวะโลกร้อน (Schneider, 1991) ต่อมาในปี 1992, ประชุมสหประชาชาติเรื่องสภาวะโลกร้อน (UNFCCC) ได้สร้างข้อตกลงที่จะลดการปล่อยก๊าซเรือนกระจกและการเผยแพร่ข้อมูลเกี่ยวกับสภาวะโลกร้อน (United Nations, 1992) จนมาถึงปัจจุบัน การคาดการณ์มลพิษทางอากาศทั่วโลกเป็นกระบวนการที่ต้องพัฒนาอย่างต่อเนื่อง และมีการวิจัยและรายงานที่ติดตามเรื่องนี้อย่างใกล้ชิดเพื่อให้เกิดความเข้าใจที่ลึกซึ้งเกี่ยวกับผลกระทบของมลพิษทางอากาศต่อสิ่งแวดล้อมและมนุษย์

อย่างไรก็ตาม การคาดการณ์ PM 2.5 เริ่มด้วยการตรวจสอบข้อมูลปริมาณ PM 2.5 ปัจจุบันซึ่งมักมีจากสถานีตรวจวัดคุณภาพอากาศหรือหน่วยงานที่เกี่ยวข้องในพื้นที่ศึกษา โดยการคาดการณ์ PM 2.5 รวมถึงการสะสมข้อมูลปริมาณ PM 2.5 ในช่วงเวลาที่ผ่านมา เช่น 7 วันหรือ 30 วัน เพื่อให้ได้ข้อมูลประวัติและแนวโน้มมาก่อน บางพื้นที่ศึกษาจะใช้โมเดลทางสถิติหรือเทคนิคการเรียนรู้ของเครื่องเพื่อคำนวณการคาดการณ์ PM 2.5 ในอนาคต โมเดลเหล่านี้อาจใช้ข้อมูลปริมาณ PM 2.5 ปัจจุบันและข้อมูลประวัติเป็นพื้นฐาน บางครั้งการคาดการณ์ PM 2.5 อาจผนวกข้อมูลจากตัวแปรอื่น ๆ ร่วมด้วย เช่น สภาพอากาศปัจจุบัน, การกระจายของลม, และปัจจัยอื่น ๆ เพื่อให้มีความแม่นยำในการทำนายมากขึ้น

ในอดีต การคาดการณ์คุณภาพอากาศโดยเฉพาะ PM 2.5 มักใช้ข้อมูลจากสถานีวัดคุณภาพอากาศในพื้นที่เพื่อประมาณปริมาณ PM 2.5 ในสถานที่นั้นๆ แต่ในปัจจุบัน ตลอดเวลาการวิจัยและพัฒนาโมเดลการคาดการณ์คุณภาพอากาศด้วยโมเดลทางสถิติและโมเดลการเรียนรู้ของเครื่อง (Machine Learning) ได้รับความสนใจเป็นอย่างมาก โดยผู้วิจัยได้รวบรวมงานที่ศึกษาเกี่ยวกับการคาดการณ์ PM 2.5 ตั้งแต่อดีตจนถึงปัจจุบัน ไว้ดังนี้

โดยมีงานวิจัยในประเทศจีน Mao et al. (2022) ได้ทำการเก็บข้อมูลมลพิษทางอากาศตั้งแต่ปี 2013-2019 เพื่อทำการวิเคราะห์และทำนายการเปลี่ยนแปลงมลพิษต่าง ๆ ซึ่งการศึกษานี้ตรวจสอบความสามารถของแบบจำลอง WRF/CMAQ ในการจำลองมลพิษทางอากาศในประเทศจีนในระยะยาว และให้ข้อมูลเชิงลึกสำหรับการออกแบบกลยุทธ์การควบคุมการปล่อยก๊าซที่มีประสิทธิภาพทั่วประเทศจีนอีกด้วย และมีอีกหลายงานวิจัย ที่มีการศึกษาเกี่ยวกับการคาดการณ์ PM 2.5 ในหลายประเทศ โดยใช้แบบจำลองทางคณิตศาสตร์ เพื่อคาดการณ์การเกิด PM 2.5 ในอนาคต

ในไต้หวัน Doreswamy et al. (2020) ได้เสนอ งานวิจัยเกี่ยวกับแบบจำลองการเรียนรู้ Machine Learning เพื่อวิเคราะห์มลพิษทางอากาศ โดยนำข้อมูลจากในสถานีตรวจอากาศไต้หวันจำนวน 76 สถานี ปี 2555 ถึง 2560 โดยใช้พารามิเตอร์สองกลุ่ม คือ ข้อมูลเฉพาะและข้อมูลตามลำดับเวลา และสร้างแบบจำลองการทำนายที่แตกต่างกัน โดยแบบจำลองเหล่านี้สามารถทำงานได้ดีกว่า

แบบจำลองเดิมในการทำนายประสิทธิภาพ และสรุปได้ว่าแบบจำลอง Gradient boosting regressor มีประสิทธิภาพดีที่สุด โดยทำการประเมินประสิทธิภาพของแบบจำลอง ด้วยการคำนวณทางสถิติของ ตัวชี้วัด MAE, MSE, RMSE, และ  $R^2$

ในประเทศจีน Chen et al. (2022) เสนอการศึกษาการแปรผันสามมิติเพื่อตรวจสอบข้อมูล PM 2.5 แบบหลายแหล่งจากการสังเกตการณ์จากดาวเทียมและภาคพื้นดิน และการปรับการปล่อยมลพิษ ร่วมกันเพื่อการคาดการณ์ PM 2.5 ของแบบจำลอง WRF-Chem (แบบจำลองควบคุมอุตุนิยมวิทยา-เคมี ระบบการพยากรณ์คุณภาพอากาศ เป็นระบบการพยากรณ์คุณภาพอากาศแบบเวลาจริง (real-time forecast) ล่วงหน้า 3 วัน) ได้ทำการทดลองเหนือมณฑลหูเป่ย์ ประเทศจีน ในช่วงการระบาดของ Covid-19 ตั้งแต่วันที่ 21 มกราคม ถึง 20 มีนาคม 2563 ผลการวิจัย พบว่าการคาดการณ์ PM 2.5 ได้รับการปรับปรุงในการตรวจสอบเกือบทุกที่ การศึกษานี้ประเมินความเข้มข้นของ PM 2.5 โดยใช้ อัลกอริทึมสุ่ม Random Forest ผลการศึกษาพบว่าแบบจำลอง WRF-Chem ให้การคาดการณ์ PM 2.5 ที่แม่นยำยิ่งขึ้นซึ่งสามารถอยู่ได้นาน 36 ชม.

ในประเทศเกาหลีใต้ Ju et al. (2021) เสนอการศึกษาการเปลี่ยนแปลงของมลพิษทางอากาศ ในช่วงการเว้นระยะห่างทางสังคมหลังการระบาดของ Covid-19 ในเกาหลี โดยการเก็บรวบรวมข้อมูล มลพิษทางอากาศช่วงระยะเวลา 4 ปีที่ผ่านมารวบรวมจาก 446 สถานีตรวจวัดคุณภาพสิ่งแวดล้อมใน เกาหลี นำมาเปรียบเทียบกับข้อมูลในปัจจุบัน และนำไปวิเคราะห์ข้อมูลทางสถิติดำเนินการโดยใช้ โปรแกรม SAS จากการศึกษาพบว่า มาตรการในการจัดการแพร่กระจายของ Covid-19 ส่งผลดีต่อ ระดับมลพิษทางอากาศ จากกิจกรรมที่ลดลง โดยการศึกษามุ่งเน้นไปที่ PM 2.5, PM 10, NO<sub>2</sub>, และ CO ซึ่งเป็นมลพิษที่สำคัญกับการใช้พลังงานในอุตสาหกรรม และปริมาณการจราจร ทั้งหมดลดลงอย่างมาก ในช่วงระยะเวลาการเว้นระยะห่างทางสังคม PM 2.5 ลดลง 45.45%, PM 10 ลดลง 35.56%, NO<sub>2</sub> ลดลง 17.33%, และ CO ลดลง 17.33% ปริมาณการจราจรบนถนนและการขนส่งสาธารณะในโซลลดลง 7.2% และ 34.5% แผนการฟื้นฟูจาก Covid-19 สามารถใช้วางรากฐานไปสู่การเปลี่ยนแปลงที่ยั่งยืน ในด้านการลดการใช้เชื้อเพลิง และเสนอแนะปัญหาด้านสาธารณสุขอันเนื่องมาจากมลพิษทางอากาศ

ในขณะที่มีนักวิจัย Benchrif et al. (2021) เสนอการศึกษาดัชนีคุณภาพอากาศ AQI และความเข้มข้นของ PM 2.5 และ NO<sub>2</sub> ช่วงก่อน ระหว่างและหลังล็อกดาวน์ใน 21 เมืองทั่วโลก แบ่งตามภูมิศาสตร์ ได้ดังนี้ แอฟริกาจำนวน 6 เมือง ตะวันออกกลางจำนวน 3 เมือง อเมริกาจำนวน 2 เมืองและเอเชียจำนวน 10 เมือง ผลการศึกษาพบว่าจากข้อจำกัดในช่วงล็อกดาวน์ ดัชนีคุณภาพอากาศ ของ AQI ลดลง 2-42% โดยเมืองอาบูดาบีลดลงมากที่สุด (42%) และเมืองมานามา ลดลงน้อยสุด (2%) ค่าความเข้มข้น NO<sub>2</sub> ลดลง 3-58% ใน 18 เมือง แต่มีการเพิ่มขึ้นใน 3 เมือง ได้แก่ อาบีจาน (1%) โคนากรี (3%) ฉิงตู (10%) ค่าความเข้มข้นของ PM 2.5 ลดลง 46-60% โดยเมืองอาบูดาบีลดลงสูงสุด (42%) และเมืองมานามาลดลงต่ำสุด (2%)

ในประเทศจีน Ni (2019) ได้ศึกษาและทำนายความเข้มข้นของ PM 2.5 ในระยะสั้น ของกรุง ปักกิ่ง ประเทศจีน โดยใช้หลายแบบจำลอง (Multi-Source Data Mining) ในการวิเคราะห์ โดย Dataset ที่ใช้ในการทำแบบจำลองการวิเคราะห์ความสัมพันธ์ของการเกิด PM 2.5 มาจาก ข้อมูลทาง กายภาพ (ข้อมูลอุตุนิยมวิทยา รวมถึงปริมาณน้ำฝนเฉลี่ยในภูมิภาค อุณหภูมิเฉลี่ยรายวัน ความชื้นสัมพัทธ์เฉลี่ย ความเร็วลมเฉลี่ย ความเร็วลมสูงสุด และข้อมูลความเข้มข้นของสารก่อมลพิษอื่นๆ ซึ่ง

รวมถึง CO, NO<sub>2</sub>, SO<sub>2</sub>, PM10) และข้อมูลโซเซียลมีเดีย จากการศึกษาพบว่า ปัจจัยต่าง ๆ ที่มาจากข้อมูลด้านอุตุนิยมวิทยา ค่าความเร็วลมเฉลี่ย ความเข้มข้นของ CO NO<sub>2</sub> PM10 และเคียเวิร์ด ปักกิ่งจากโซเซียลมีเดีย; สัมพันธ์กับความเข้มข้นของ PM 2.5 ที่เกิดขึ้น มีการวิเคราะห์เพิ่มเติมด้วยชุดข้อมูล Bigdata โดยใช้ Back Propagation Neural Network (BPNN) พบว่า วิธีนี้ทำงานได้ดีกว่า (Performs) ในความสัมพันธ์ของชุดข้อมูลจากแบบจำลองอนุกรมเวลา Autoregressive Integrated Moving Average (hereinafter referred to as ARIMA) Time-Series ถูกนำมาใช้เพื่อสำรวจการทำนาย PM 2.5 ในอนุกรมเวลาระยะสั้น พบว่า ผลลัพธ์ที่คาดการณ์อยู่ช่ายกับข้อมูลที่เก็บมา

ในขณะที่อิหร่าน Mehdipour et al. (2018) ได้เสนอแบบจำลองเพื่อทำนายความเข้มข้นของ PM 2.5 3 แบบจำลอง คือ Decision tree (DT), Support vector machine (SVM) และ Bayesian network (BN) โดยรวบรวมข้อมูลเป็นเวลาสามปีติดต่อกัน (มกราคม 2013 ถึงมกราคม 2016) เพื่อใช้ในการพัฒนาแบบจำลอง ซึ่งพารามิเตอร์ซึ่งครอบคลุมตัวแปรอุตุนิยมวิทยาและความเข้มข้นของสารเคมีหลายชนิดเป็นตัวทำนายที่เป็นไปได้ของ PM 2.5 โดยจะใช้ PM 10, NO<sub>2</sub>, SO<sub>2</sub> และ O<sub>3</sub> ซึ่งเป็นแปรที่ใช้ในการทำนายการเกิดของ PM 2.5 ของทั้ง 3 แบบจำลอง ซึ่งจากการศึกษา พบว่า SVM มีผลงานที่ดีที่สุด (Best Performing) ตามเกณฑ์การประเมิน

ที่ประเทศมาเลเซีย Mahanta et al. (2019) เป็นการศึกษาประสิทธิภาพของแบบจำลองที่ใช้ในการคาดการณ์อยู่ในการคาดการณ์คุณภาพอากาศค่าดัชนี (AQI) ที่ให้ข้อมูลนำเข้าบางส่วน โดยขึ้นอยู่กับข้อมูลมลพิษและอุตุนิยมวิทยาในนิวเดลี ประเทศอินเดีย โดยทำการวิเคราะห์การถดถอยในชุดข้อมูลและผลลัพธ์ เพื่อหาว่าปัจจัยทางอุตุนิยมวิทยาส่งผลต่อค่า AQI มากขึ้นไหมและตัวแบบการทำนายจะมีประโยชน์เพียงใดในการพยากรณ์คุณภาพอากาศ ซึ่งจากการทดสอบประโยชน์ของตัวแบบการถดถอยที่มีอยู่ในไลบรารี ทำนายค่าดัชนีคุณภาพอากาศจากในอดีต ซึ่งแบบจำลองส่วนใหญ่มีความแม่นยำถึง 85% โดยแบบจำลองที่มีความแม่นยำสูงสุดคือ Extra Trees Regression Model ซึ่งหมายความว่า แบบจำลองดังกล่าว ค่อนข้างมีประโยชน์ในการทำเป็นตัวทำนาย

ในประเทศจีน Li et al. (2014) ยังได้ทำการศึกษา PM 2.5 และมลพิษที่เกี่ยวข้อง คือ SO<sub>2</sub>, NO<sub>2</sub>, PM 10, CO, และ O<sub>3</sub> ซึ่งวิเคราะห์โดยใช้วิธีทางสถิติ และมีการสร้างแบบจำลองการถดถอยที่แตกต่างกันสองแบบ เพื่อหาสาเหตุและปัจจัยของ PM 2.5 โดยใช้ข้อมูลตั้งแต่วันที่ 1 มกราคม ถึง 26 สิงหาคม 2013 ซึ่งได้มาจาก Wuhan Environmental ผลการวิจัยพบว่าค่าสัมประสิทธิ์สหสัมพันธ์ของ PM 2.5 และอากาศชนิดอื่นๆ (SO<sub>2</sub>, NO<sub>2</sub>, PM 10, CO) มีค่ามากกว่า 0.7 ซึ่งแสดงความสัมพันธ์เชิงเส้นที่มีนัยสำคัญ และแบบจำลองการถดถอยได้รับการพิสูจน์แล้วว่ามีความแม่นยำและน่าเชื่อถือ

ส่วนในประเทศไทย Amnuaylojaroen, (2022) ได้ทำการศึกษาของการเปลี่ยนแปลงครั้งใหญ่ในพื้นที่ชนบทในภาคเหนือของประเทศไทย ความถี่และความรุนแรงของการระบาดของหมอกควันจากมลภาวะที่เป็นอนุภาค โดยเฉพาะฝุ่นละอองขนาดเล็ก (PM 2.5) ได้เพิ่มขึ้นในภูมิภาคนี้ เนื่องจากการสังเกต PM 2.5 เป็นไปอย่างจำกัด การศึกษานี้จึงใช้แบบจำลองการถดถอยเชิงเส้นหลายตัวแปร (Applied A Multivariate Linear Regression Model) เพื่อทำนายความเข้มข้นของ PM 2.5 ในปี พ.ศ. 2563 โดยใช้ความลึกเชิงแสงของละอองลอย (AOD) พารามิเตอร์อุตุนิยมวิทยาของความเร็วลม อุณหภูมิ และความชื้นสัมพัทธ์ และก๊าซมลพิษ เช่น SO<sub>2</sub>, NO<sub>2</sub>, CO และ O<sub>3</sub> จากการวัดภาคพื้นดิน 3 แห่ง ได้แก่ เชียงใหม่, ลำปางและน่าน ในภาคเหนือของประเทศไทย ในการศึกษาที่ใช้ตัวแบบการ

ถดถอยเชิงเส้นพหุตัวแปร 2 ตัว โดยแบบจำลองที่ 1 เป็นแบบจำลองที่มีพารามิเตอร์ด้านอุตุนิยมวิทยาของความลึกเชิงแสงของละอองลอย (AOD) อุณหภูมิ ความชื้นสัมพัทธ์ และความเร็วลม ทั่วไป Model 2 ประกอบด้วยพารามิเตอร์อุตุนิยมวิทยาและสารมลพิษในก๊าซหลายชนิด เช่น  $\text{SO}_2$ ,  $\text{NO}_2$ ,  $\text{CO}$  และ  $\text{O}_3$  โดยใช้ข้อมูลรายชั่วโมงจากปี พ.ศ. 2563 ในจังหวัดนั้นๆ ระบุความเข้มข้นของ  $\text{PM}_{2.5}$  ได้อย่างเพียงพอจากผลการวิเคราะห์ พบว่า ประสิทธิภาพของแบบจำลองที่ 2 เป็นผลดีต่อการทำนายความเข้มข้น  $\text{PM}_{2.5}$  ที่เชียงใหม่ ( $R^2 = 0.52$ ) และลำปาง ( $R^2 = 0.60$ ) โดยแบบจำลองที่ 2 สามารถทำนายความเข้มข้นของ  $\text{PM}_{2.5}$  ได้ทั้งฤดูฝนและฤดูแล้งเมื่อเทียบกับแบบจำลองที่ 1 อย่างไรก็ตาม ยังมีความไม่แน่นอนของแบบจำลอง ซึ่งวางรากฐานสำหรับการศึกษาต่อไป ในปี พ.ศ. 2563 ยังมีงานวิจัยในกรุงเทพมหานคร Kanchanasuta et al. (2020) ใช้ข้อมูลจากสถานีวัดคุณภาพอากาศในกรุงเทพมหานคร เพื่อวิเคราะห์การกระจายของ  $\text{PM}_{2.5}$  และ  $\text{PM}_{10}$  ในพื้นที่และเวลาที่แตกต่างกัน โดยใช้โดยใช้เทคนิคทางสถิติ แผนภาพขั้วสองตัวแปร และฟังก์ชันความน่าจะเป็นแบบสองตัวแปรแบบมีเงื่อนไข (Bivariate Polar Plot และ Conditional Bivariate Probability Function) ซึ่งผลการศึกษาพบว่า จุดที่ติดตั้งสถานีตรวจวัด  $\text{PM}_{2.5}$  ที่อยู่ในบริเวณริมถนน จะมีค่า  $\text{PM}_{2.5}$  สูงกว่าจุดอื่นเช่นเดียวกับงานก่อนหน้านี

ในประเทศมาเลเซีย Sobri et al. (2021) การศึกษานี้แสดงลักษณะรูปแบบของความเข้มข้น  $\text{PM}_{2.5}$  ที่เกี่ยวข้องกับสถานี 7 แห่งในประเทศมาเลเซีย ด้วยตัวชี้วัด 7 ชนิด ได้แก่ คาร์บอนมอนนอกไซด์, โอโซน, ซัลเฟอร์ไดออกไซด์, ไนโตรเจนไดออกไซด์, ความชื้น, อุณหภูมิ, และ ความเร็วลม ในการทำนาย  $\text{PM}_{2.5}$  แต่ละสถานี ใช้วิธี Multiple Linear Regression (MLR) และ Artificial Neural Networks (ANN) โดยประสิทธิภาพของ MLR และ ANN ทั้งสอง ถูกเปรียบเทียบโดยอิงจาก  $R^2$  และ Mean Square Error (MSE) ทั้งนี้ ผลการศึกษาพบว่า ANN ทำงานได้ดีกว่า MLR โดยมีค่าสัมประสิทธิ์การตรวจวัด ( $R^2$ ) สูงและการวัดค่าความผิดพลาดต่ำกว่า MLR

ทั้งนี้ ยังมีการรวบรวมงานวิจัยที่ใช้วิธี Machine Learning ในการทำนายค่าความเข้มข้นของ  $\text{PM}_{2.5}$  ในประเทศมาเลเซีย Palanichamy et al. (2021) ด้วยวิธีต่าง ๆ โดยขอบเขตงานวิจัยที่ค้นหาตั้งแต่ปี 2017 ถึงปี 2021 ซึ่งรวบรวมโดยคำนึงถึงตัวแปรหลัก 5 ตัว ได้แก่ 1. งานวิจัยที่ใช้ Machine learning 2. วิธีการทำการวิจัยที่เลือกใช้ 3. สถานที่ที่ผู้วิจัยเลือกใช้ 4. วิธีการวัดผลที่เลือกใช้ ได้แก่ RMSE, MSE, MAE, และ  $R^2$  5. ประสิทธิภาพของงานวิจัยของอัลกอริทึม วิธีการวิจัยของ machine learning ที่นิยมใช้มากที่สุดในการทำนายค่าความเข้มข้นของ  $\text{PM}_{2.5}$  เรียงตามลำดับจากมากไปน้อย ได้แก่ Deep learning, neural network, decision tree, regression และ support vector หนึ่งในงานวิจัยที่บทความนี้ทำการศึกษาคือการทำนายค่าความเข้มข้นของ  $\text{PM}_{2.5}$  ในประเทศอินเดียด้วยวิธีผสมระหว่าง Gradient Boosting และ Decision Tree ซึ่งทำให้รองรับข้อมูลจำนวนมากโดยใช้ปริมาณความจำเพียงเล็กน้อยและมีความรวดเร็วในการประมวลผล ซึ่งวิธีการทำนายแบบผสมดังกล่าวเมื่อทำมาเปรียบเทียบกับวิธีอื่นๆ ได้แก่ Gradient Boosting และ Decision Tree และ Regression พบว่าวิธีการทำนายแบบผสมได้ผลลัพธ์ที่ดีกว่า โดยให้ค่า Root Mean Square Error และ ค่า Mean Absolute Error ที่ต่ำที่สุด

ในประเทศจีน Yin et al. (2021) ได้ทำงานวิจัยนี้แสดงถึงการทำนายความเข้มข้นรายชั่วโมงของค่า PM 2.5 ด้วยวิธี Novel Multi-Step Forecasting Method ซึ่งใช้ 2 อัลกอริทึม ได้แก่ Modified Adaboost.RT และ Gradient Boosting ในการทำ Extreme Learning Machine (ELM) เพื่อพยากรณ์ค่า PM 2.5 สถานที่ที่เลือกใช้ในงานวิจัยนี้ คือ 4 เมืองในประเทศจีน ได้แก่ ปักกิ่ง เซี่ยงไฮ้ วูฮั่น และกวางโจว ซึ่งอยู่ในภาคเหนือ ภาคตะวันออก ภาคกลางและภาคใต้ของประเทศจีนตามลำดับ จากผลงานวิจัยพบว่า Gradient Boosting ให้ผลลัพธ์จากการทำนายได้ดีกว่า Modified Adaboost.RT ในเกือบทุกกรณีศึกษา

ในประเทศสหรัฐอเมริกา Mahmud et al. (2022) งานนี้มุ่งเน้นไปที่การคาดการณ์ฝุ่นละออง (PM 2.5) ที่ภูมิภาค Paso Del Norte ของสหรัฐอเมริกา โดยการใช้อัลกอริทึมการเรียนรู้ของเครื่อง (Supervised Machine Learning Algorithms) ภายใต้การดูแลที่แตกต่างกัน โดยอิงตามการถดถอยและการจำแนกประเภท เพื่อคาดการณ์ค่า PM 2.5 รายวันอย่างแม่นยำ ซึ่งใช้ตัวแปรอุตุนิยมวิทยาและสภาพอากาศจากสถานีตรวจสอบของคณะกรรมการคุณภาพสิ่งแวดล้อมแห่งเท็กซัส ในปี 2014 – 2019 โดยทำแบบจำลองทั้งหมด 6 ชนิด โดยผลลัพธ์แสดงให้เห็นว่า random forest สามารถพยากรณ์ข้อมูลด้วยความแม่นยำ 92% ซึ่งทำงานได้ดีที่สุดจากแบบจำลองการเรียนรู้ทั้งหมด

ในประเทศมาเลเซีย Palanichamy et al. (2022) ยังได้ทำแบบจำลอง ML สำหรับการพยากรณ์ความเข้มข้นของ PM 2.5 ได้รับการตรวจสอบในชุดข้อมูลคุณภาพอากาศของมาเลเซีย ตั้งแต่ปี 2017 ถึง 2018 ชุดข้อมูลได้รับการประมวลผลล่วงหน้าโดยการล้างข้อมูลและกระบวนการทำให้เป็นมาตรฐาน จากนั้น ถูกย่อให้เป็นชุดข้อมูลที่มีปัจจัยด้านสถานที่และเวลาในกระบวนการแยกคุณลักษณะ (Feature Extraction Process) ชุดข้อมูลถูกป้อนเข้าไปในตัวแยกประเภท ML ที่ได้รับการดูแล 3 ตัว ซึ่งประกอบด้วย Random Forest (RF), Artificial Neural Network (ANN) และ Long Short-Term Memory (LSTM) จากนั้นทำการประเมินผลเปรียบเทียบเพื่อระบุแบบจำลองที่ดีที่สุดสำหรับการคาดการณ์ PM 2.5 ที่แม่นยำ ซึ่งจากผลการทดสอบพบว่า Random Forest (RF) มีความแม่นยำที่สุด ร้อยละ 97.7 ในการคาดการณ์ PM 2.5

ในประเทศจีน Pan (2018) ได้ใช้อัลกอริทึม XGBoost (Extreme Gradient Boosting) เพื่อทำนายความเข้มข้นของ PM 2.5 รายชั่วโมง ข้อมูลการติดตามคุณภาพอากาศในเมืองเทียนจิน ของประเทศจีน ได้รับการวิเคราะห์โดยใช้อัลกอริทึม XGBoost นอกจากนี้ วิธี XGBoost ยังถูกเปรียบเทียบกับ Random Forest Algorithm, Multiple Linear Regression, Decision Tree Regression And Support Vector Machines ที่รองรับสำหรับแบบจำลองการถดถอยโดยใช้ผลการคำนวณ โดยผลการศึกษาพบว่า อัลกอริทึม XGBoost มีประสิทธิภาพเหนือกว่าวิธีการอื่นๆ

ทั้งนี้ ยังมีอีกหลากหลายงานวิจัย (Guo et al., 2023; Yu et al., 2023) ที่แสดงให้เห็นว่า Gradient Boosting เป็นแบบจำลองที่สามารถพยากรณ์หรือคาดการณ์ PM 2.5 ได้แม่นยำมากที่สุดในปัจจุบัน



## 2.5 สรุปผลการทบทวนวรรณกรรม

สรุปการทบทวนงานวิจัย พบว่า ปัญหาใหญ่เรื่องหนึ่งของประเทศไทยในปัจจุบันคือปัญหาฝุ่น PM 2.5 ซึ่งสาเหตุส่วนหนึ่งปัญหาการจราจรติดขัด โดยเฉพาะอย่างยิ่ง กรุงเทพมหานคร ที่ติดอันดับเมืองที่รถติดที่สุดในโลกทุกปี (Pishue and Trepanier, 2021; Tomtom Traffic Index, 2021)

จนกระทั่ง ต้นปี พ.ศ. 2562 ที่มีการระบาดของโรค Covid-19 ทั่วโลก ส่งผลให้รูปแบบการดำเนินชีวิตของผู้คนเปลี่ยนแปลงไป มีการ Lockdown ให้คนอยู่แต่ในการ เรียนออนไลน์ Work from home ซึ่งส่งผลกระทบต่อเศรษฐกิจเป็นวงกว้าง และเมื่อกิจกรรมที่เป็นปกติของผู้คนถูกจำกัด ส่งผลให้การเดินทาง (Travel trip) ลดน้อยลงไปอย่างมาก และมีการเปลี่ยนรูปแบบการเดินทาง (Mode Choice) อย่างเห็นได้ชัดในหลายประเทศ (Cui et al., 2021; Hasselwander et al., 2021; König and Dreßler, 2021)

ในช่วง 2-3 ปีที่ผ่านมา การระบาดใหญ่ของ Covid-19 ทำให้กิจกรรมของมนุษย์ลดลงอย่างมาก ในขณะที่สิ่งแวดล้อมโดยรวมสามารถฟื้นตัวได้บ้าง (Dasgupta and Srikanth, 2020; Venter et al., 2020) ดังนั้น ผู้วิจัยจึงได้พิจารณาการเกิด PM 2.5 และมลพิษก๊าซในช่วงเวลานี้เพื่อคาดการณ์สถานการณ์ในอนาคตที่อาจเปลี่ยนแปลงปัจจัยหลักที่ทำให้เกิด PM 2.5

จากงานวิจัยที่ผู้วิจัยได้ทำการทบทวนมานั้น พบว่า ความเข้มข้นของ PM 2.5 จะขึ้นอยู่กับปัจจัยที่เกี่ยวข้องต่อการเกิด PM 2.5 ทั้งตัวแปรทั้งทางด้านข้อมูลอุตุนิยมวิทยา และข้อมูลสารประกอบมลพิษทางอากาศต่าง ๆ ล้วนเป็นปัจจัยที่ก่อให้เกิด PM 2.5 ทั้งสิ้น รวมไปถึง ฤดูกาล (วัน เดือน) ช่วงเวลาในรอบวัน (ชม.) ก็เกี่ยวเนื่องอย่างหลีกเลี่ยงไม่ได้ และพฤติกรรมของมนุษย์ก็ส่งผลต่อความเข้มข้นของ PM 2.5 เช่นเดียวกัน ซึ่งในประเทศไทย ระดับ PM 2.5 มีการกระจายตัวตามฤดูกาล โดยความเข้มข้นของมลพิษจะสูงสุดในช่วงฤดูแล้ง (พฤศจิกายน-มีนาคม) และลดลงในฤดูฝนและฤดูร้อน

ทั้งนี้ จากการทบทวนวรรณกรรมที่เกี่ยวข้องกับการคาดการณ์ PM 2.5 ทั้งในส่วนของ Model ที่ใช้ในการคาดการณ์ วิธีการดำเนินงานวิจัย การเก็บรวบรวมข้อมูลปีที่ใช้ข้อมูลในการวิจัย ปีที่ทำการวิจัย และผลลัพธ์ที่ได้ในอดีต ผู้วิจัยสามารถสรุปรายละเอียดดังตาราง 2.8

การคาดการณ์มลพิษทางอากาศทั่วโลกเป็นกระบวนการที่ต้องพัฒนาอย่างต่อเนื่อง และมีการวิจัยและรายงานที่ติดตามเรื่องนี้อย่างใกล้ชิดเพื่อให้เกิดความเข้าใจที่ลึกซึ้งซึ่งเกี่ยวกับผลกระทบของมลพิษทางอากาศต่อสิ่งแวดล้อมและมนุษย์ โดยในอดีตการคาดการณ์ PM 2.5 จะนำข้อมูลปริมาณ PM 2.5 ในช่วงเวลาที่ผ่านมา เช่น 7 วันหรือ 30 วัน เพื่อให้ได้ข้อมูลประวัติและแนวโน้มมาก่อนและมาวิเคราะห์อย่างง่ายตามวันเวลาที่มีชุดข้อมูลอยู่ จนกระทั่งบางพื้นที่ได้ทำการศึกษาจะใช้แบบจำลองทางสถิติหรือเทคนิคการเรียนรู้ของเครื่องเพื่อคำนวณการคาดการณ์ PM 2.5 ในอนาคต โดยแบบจำลองเหล่านี้จะใช้ข้อมูลปริมาณ PM 2.5 ปัจจุบันในการทำนายการเกิด PM 2.5 ในอนาคต จนมาถึงในปัจจุบัน เริ่มมีการคาดการณ์ PM 2.5 ผสมเข้ากับข้อมูลจากตัวแปรอื่น ๆ ร่วมด้วย เช่น สภาพอากาศปัจจุบัน, การกระจายของลม, และปัจจัยอื่น ๆ เพื่อให้มีความแม่นยำในการทำนายมากขึ้น เมื่อมีการพัฒนาทางเทคโนโลยีที่สูงขึ้น การทำแบบจำลองในปัจจุบันจึงมีการนำแบบจำลองที่มาจาก Machine Learning เข้ามาใช้ในการวิเคราะห์และทำนายการเกิด PM 2.5 มากยิ่งขึ้น โดยคำนึงถึงปัจจัยที่มีผลต่อการกระจายของ PM 2.5 เช่น ลักษณะการไหลของลม อุณหภูมิ ความชื้น และการเผาไหม้ และใช้ข้อมูลปัจจัยเหล่านี้ในแบบจำลอง ซึ่งให้ผลการศึกษาที่แม่นยำกว่า

การศึกษาที่ผ่านมา ดังที่สรุปไว้ในตาราง 2.8 ยังมีข้อจำกัดและช่องว่างของงานวิจัย ในเรื่องชุดข้อมูลที่นำมาใช้วิจัย โดยงานวิจัยที่ผ่านมาใช้ข้อมูล (Dataset) ย้อนหลังของแต่ละพื้นที่ ในช่วงเวลาปกติ (Normal Situation) และพบว่างานวิจัยบางส่วนใช้ชุดข้อมูลในช่วงเวลาที่สั้นเกินกว่าจะเห็นการเปลี่ยนแปลงของข้อมูลด้านอุตุนิยมวิทยาและข้อมูลด้านมลพิษทางอากาศในภาพกว้าง และแม้ว่าจะมีการนำแบบจำลองทางคณิตศาสตร์มาใช้ในการวิเคราะห์และคาดการณ์การเกิด PM 2.5 อย่างแพร่หลายทั่วโลก แต่ในประเทศไทย ยังไม่มีการนำแบบจำลองทางคณิตศาสตร์มาใช้ในการศึกษาวิจัยในด้านนี้ และหลังจากการแพร่ระบาดของ Covid-19 ซึ่งถือว่าเป็นช่วงสถานการณ์ที่ไม่ปกติ (Back Swan Situation) ยังไม่มีงานวิจัยที่นำชุดข้อมูลด้านอุตุนิยมวิทยาและข้อมูลด้านมลพิษทางอากาศนำมาวิเคราะห์หาแบบจำลองที่ดีที่สุดที่ใช้ในการวิเคราะห์คาดการณ์ PM 2.5 ด้วยชุดข้อมูลที่แตกต่างกัน ซึ่งในงานวิจัยชิ้นนี้ ผู้วิจัยจะใช้ข้อมูล (Dataset) ด้านอุตุนิยมวิทยาและข้อมูลด้านมลพิษทางอากาศย้อนหลังในกรุงเทพมหานคร ทั้งในสถานการณ์ปกติ (Normal Situation) และใช้ข้อมูล (Dataset) เมื่อมีการแพร่ระบาดของ Covid-19 มาทดสอบแบบจำลอง

อย่างไรก็ตาม ยังพบว่า ช่องว่างงานวิจัยในส่วนที่เป็นชุดข้อมูลด้านอุตุนิยมวิทยาและข้อมูลด้านมลพิษทางอากาศย้อนหลัง (Historical Data) ยังไม่ค่อยมีงานวิจัยที่นำช่วงเวลาต่าง ๆ (Time) มาเป็นปัจจัยในการวิเคราะห์ความแม่นยำของแบบจำลอง เช่น ข้อมูลในช่วงเดือน (Month) ข้อมูลในช่วงวัน (Day) และข้อมูลในแต่ละชั่วโมง (Hour) เป็นต้น ในงานวิจัยนี้ ผู้วิจัยจึงนำเอาข้อมูลด้านเวลา มาใช้เป็นตัวแปร (Feature) ในการทำนายแบบจำลองร่วมด้วย

จากการทบทวนวรรณกรรมและศึกษางานวิจัยที่เคยมีการดำเนินการ ผู้วิจัยจึงทำการวิเคราะห์ว่าแบบจำลองด้านคณิตศาสตร์ชนิดใดที่จะสามารถทำนายการเกิด PM 2.5 จากการใช้ชุดข้อมูลด้านอุตุนิยมวิทยาและข้อมูลด้านมลพิษทางอากาศย้อนหลัง (Historical Data) ในกรุงเทพมหานคร ทั้งในสถานการณ์ปกติ (Normal Situation) และ เมื่อมีการแพร่ระบาดของ Covid-19 ได้แม่นยำที่สุด และสารก่อมลพิษชนิดใดที่ส่งผลต่อการเกิด PM 2.5 พร้อมทั้งนำเอาข้อมูลด้านเวลาใช้เป็นตัวแปร (Feature) ในการทำนายแบบจำลอง เพื่อหาแบบจำลองที่สามารถทำนายการเกิด PM 2.5 ได้แม่นยำมากที่สุดในทุกสถานการณ์



ตาราง 2.8 (ต่อ)

| แหล่งอ้างอิง | แบบจำลอง (Model)   |                        |                         |                             |                       |               | ข้อมูล        |                      |             | ตัวแปร         |  |      |       |       |      |     |                |                         |    |                | ความแม่นยำ |       |                 |                 |            |                 |  |             |
|--------------|--|------------------------|-------------------------|-----------------------------|-----------------------|---------------|---------------|----------------------|-------------|----------------|--|------|-------|-------|------|-----|----------------|-------------------------|----|----------------|------------|-------|-----------------|-----------------|------------|-----------------|--|-------------|
|              | Linear Regression / Multivariate Linear Regression Model | Machine Learning Model |                         |                             |                       |               | จำนวนสถานีวัด | ปีที่สำรวจ (ปี ค.ศ.) | แหล่งข้อมูล | ตำแหน่งภูมิภาค | สารมลพิษทางอากาศ   |      |       |       |      |     |                | ข้อมูลด้านอุตุนิยมวิทยา |    |                |            | อื่นๆ |                 |                 |            |                 |  |             |
|              |  | Ridge Model            | Random forest regressor | Gradient boosting regressor | K neighbors regressor | MLP regressor |               |                      |             |                | Decision Tree regressor  | CART | อื่นๆ | PM2.5 | PM10 | NOx | O <sub>2</sub> | NO                      | CO | O <sub>3</sub> |            |       | SO <sub>2</sub> | NO <sub>2</sub> | ความเร็วลม | ทิศทางลม        | อุณหภูมิ   | ความกดอากาศ |
| Pan (2018)   | ✓  |                        | ✓                       | ✓                           |                       |               |               | 19                   | 2016        | จีน            |  | ✓    | ✓     |       |      |     | ✓              | ✓                       |    | ✓              | ✓          | ✓     | ✓               | ✓               | ✓          | SO <sub>3</sub> | XGBoost (R <sup>2</sup> ) = 0.9520, Random Forest (R <sup>2</sup> ) = 0.9426, Multiple Linear Regression (R <sup>2</sup> ) = 0.9234, Decision Tree Regression (R <sup>2</sup> ) = 0.8922 |             |
| Ni (2019)    | ✓  |                        |                         |                             |                       |               |               |                      |             |                | Back Propagation Neural Network (hereinafter referred to as BPNN), |      | 2014  | จีน   | ✓    | ✓   | ✓              |                         |    | ✓              |            | ✓     | ✓               | ✓               | ✓          | ✓               |  | -           |

ตาราง 2.8 (ต่อ)

| แหล่งอ้างอิง         | แบบจำลอง (Model)   |                        |                         |                             |                       |               | อื่นๆ                            | ข้อมูล     |                      | ตัวแปร      |                   |                         |      |       |      |     |                |    |    |                         |                 |                 |            | ความแม่นยำ |  |          |
|----------------------|--|------------------------|-------------------------|-----------------------------|-----------------------|---------------|----------------------------------|------------|----------------------|-------------|-------------------|-------------------------|------|-------|------|-----|----------------|----|----|-------------------------|-----------------|-----------------|------------|------------|--|----------|
|                      | Linear Regression / Multivariate Linear Regression Model | Machine Learning Model |                         |                             |                       |               |                                  | จำนวนสถานี | ปีที่สำรวจ (ปี ค.ศ.) | แหล่งข้อมูล | ตำแหน่งภูมิประเทศ | สารมลพิษทางอากาศ        |      |       |      |     |                |    |    | ข้อมูลด้านอุตุนิยมวิทยา |                 |                 |            |            | อื่นๆ  |          |
|                      |  | Ridge Model            | Random forest regressor | Gradient boosting regressor | K neighbors regressor | MLP regressor |                                  |            |                      |             |                   | Decision Tree regressor | CART | PM2.5 | PM10 | NOx | O <sub>2</sub> | NO | CO | O <sub>3</sub>          | SO <sub>2</sub> | NO <sub>2</sub> | ความเร็วลม |            |  | ทิศทางลม |
| Sobri et al. (2021)  | ✓  |                        |                         |                             |                       |               | Artificial Neural Networks (ANN) | 2018-2019  | มาเลเซีย             |             | ✓                 | ✓                       |      |       |      | ✓   | ✓              | ✓  | ✓  | ✓                       | ✓               | ✓               | ✓          | ✓          | ANN (avg. R <sup>2</sup> ) = 0.7658  |          |
| Yin et al. (2021)    | ✓  |                        | ✓                       |                             |                       |               | extreme learning machine (ELM)   | 2019       | จีน                  | ✓           | ✓                 |                         |      |       |      |     |                |    |    |                         |                 |                 |            |            | WRF-Chem (R <sup>2</sup> ) = 0.4489, Linear regression (R <sup>2</sup> ) = 0.3983, |          |
| Mahmud et al. (2022) | ✓  | ✓                      | ✓                       |                             |                       |               |                                  | 2014-2019  | สหรัฐอเมริกา         | ✓           | ✓                 | ✓                       | ✓    | ✓     | ✓    | ✓   | ✓              | ✓  | ✓  | ✓                       | ✓               | ✓               | ✓          | ✓          | Random forest ความแม่นยำ ดีสุดที่ 92%.   |          |

ตาราง 2.8 (ต่อ)

| แหล่งอ้างอิง              | แบบจำลอง (Model)   |                        |                         |                             |                       |               | ข้อมูล |                                 |                      | ตัวแปร      |                   |                         |      |       |      |     |                |    |                         |                |                 |                 | ความแม่นยำ |       |            |          |          |             |                  |   |   |  |
|---------------------------|--|------------------------|-------------------------|-----------------------------|-----------------------|---------------|--------|---------------------------------|----------------------|-------------|-------------------|-------------------------|------|-------|------|-----|----------------|----|-------------------------|----------------|-----------------|-----------------|------------|-------|------------|----------|----------|-------------|------------------|---|---|--|
|                           | Linear Regression / Multivariate Linear Regression Model | Machine Learning Model |                         |                             |                       |               | อื่นๆ  | จำนวนสถานีวัด                   | ปีที่สำรวจ (ปี ค.ศ.) | แหล่งข้อมูล | ตำแหน่งภูมิประเทศ | สารมลพิษทางอากาศ        |      |       |      |     |                |    | ข้อมูลด้านอุตุนิยมวิทยา |                |                 |                 |            | อื่นๆ |            |          |          |             |                  |   |   |  |
|                           |  | Ridge Model            | Random forest regressor | Gradient boosting regressor | K neighbors regressor | MLP regressor |        |                                 |                      |             |                   | Decision Tree regressor | CART | PM2.5 | PM10 | NOx | O <sub>2</sub> | NO | CO                      | O <sub>3</sub> | SO <sub>2</sub> | NO <sub>2</sub> |            |       | ความเร็วลม | ทิศทางลม | อุณหภูมิ | ความกดอากาศ | ความชื้นสัมพัทธ์ |   |   |  |
| Palanichamy et al. (2022) | ✓  |                        | ✓                       |                             |                       |               |        | artificial neural network (ANN) |                      | 2017-2018   | มาเลเซีย          |                         | ✓    | ✓     | ✓    | ✓   | ✓              | ✓  | ✓                       | ✓              | ✓               | ✓               | ✓          | ✓     | ✓          | ✓        | ✓        | ✓           | ✓                | ✓ | ✓ | Random forest ความแม่นยำดีที่สุดที่ 97.7%, ANN = 61.14%, |
| Amnuaylojaroen (2022)     | ✓  |                        |                         |                             |                       |               |        |                                 |                      | 2020        | ไทย               | ✓                       | ✓    | ✓     | ✓    | ✓   | ✓              | ✓  | ✓                       | ✓              | ✓               | ✓               | ✓          | ✓     | ✓          | ✓        | ✓        | ✓           | ✓                | ✓ | ✓ | R <sup>2</sup> = 0.60                                    |

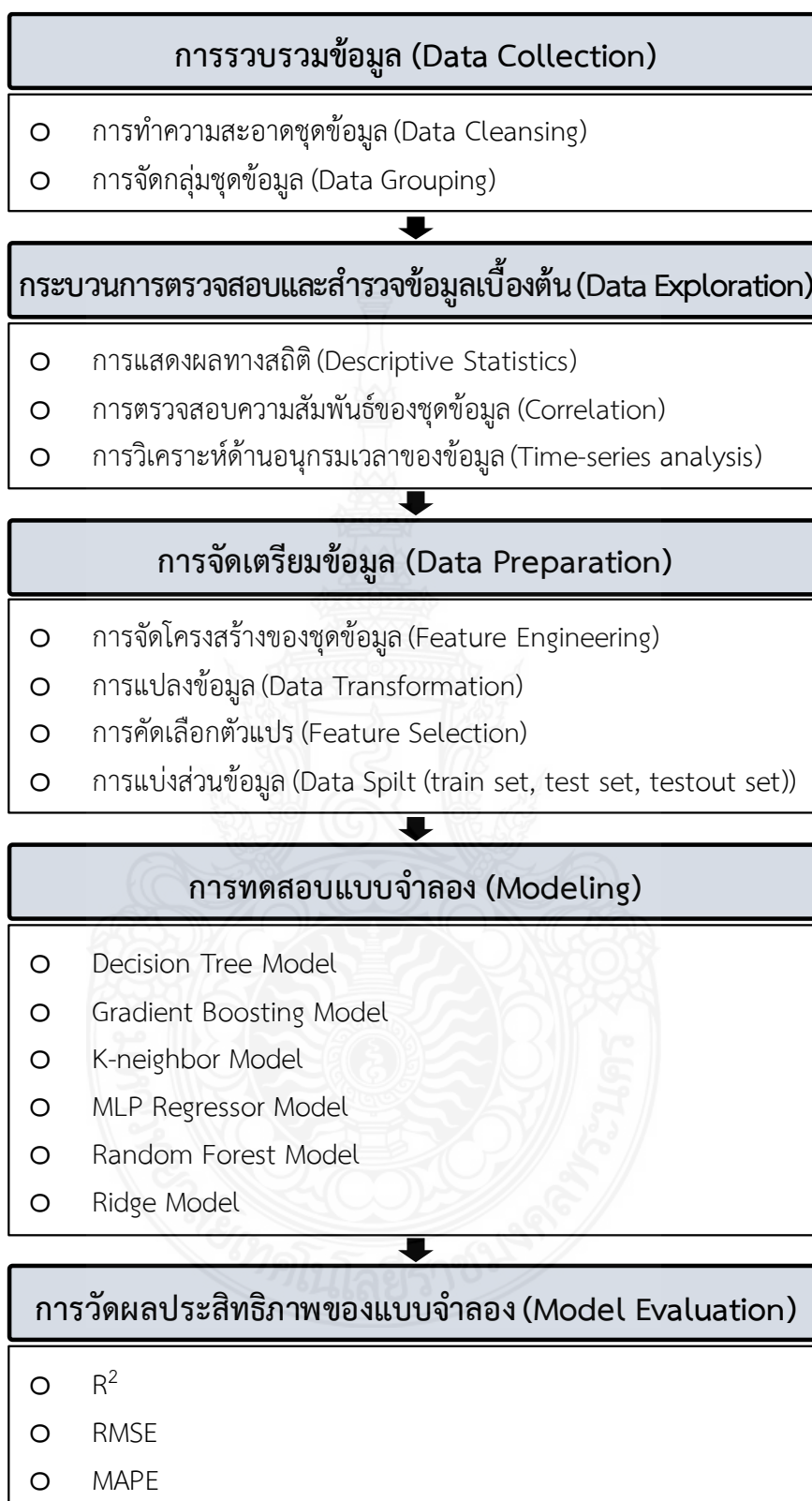
## บทที่ 3

### วิธีการดำเนินการวิจัย (Methodology)

ในการศึกษาวิเคราะห์พฤติกรรมของค่าความเข้มข้นของ PM 2.5 ที่เปลี่ยนแปลงไปตลอดในช่วงปี พ.ศ. 2558 - 2564 และนำไปสู่การประยุกต์ใช้แบบจำลองทางคณิตศาสตร์เพื่อพยากรณ์ PM 2.5 ในอนาคตนั้น เริ่มจากการสังเคราะห์ชุดสถิติข้อมูลให้อยู่ในระเบียบแบบแผน และวิเคราะห์ถึงพฤติกรรมที่เปลี่ยนแปลงไปด้วยปัจจัยที่เกี่ยวข้อง ทั้งยังพยากรณ์ค่า PM 2.5 เมื่อพฤติกรรมของ PM 2.5 เปลี่ยนแปลงไป จะส่งผลกระทบต่อการนำแบบจำลองที่เหมาะสมกับสถานการณ์ปัจจุบันไปใช้ได้หรือไม่ ท้ายที่สุดคือการหาว่าแบบจำลองใดที่เหมาะสมแก่การนำไปทำนายเพื่อหาค่า PM 2.5 ในอนาคต แม้จะเกิดสถานการณ์ที่ไม่คาดฝันที่ส่งผลต่อค่า PM 2.5 ในช่วงเวลานั้น ๆ

#### 3.1 ขั้นตอนการพัฒนาแบบจำลอง

ในการดำเนินการวิจัยนี้มีขั้นตอนการดำเนินงานวิจัย โดยเริ่มจากการรวบรวมข้อมูล (Collecting data) ด้านอุตุนิยมวิทยาและมลพิษทางอากาศย้อนหลัง การนำข้อมูลเข้าสู่กระบวนการตรวจสอบและสำรวจข้อมูลเบื้องต้น (Data Exploration) ซึ่งขั้นตอนนี้จะเป็นการวิเคราะห์ข้อมูลที่จำเป็นก่อนการนำข้อมูลไปใช้หรือนำไปวิเคราะห์เชิงลึก และทำการจัดเตรียมข้อมูล (Data Preparation) เพื่อวิเคราะห์ถึงโครงสร้างทางวิศวกรรมของข้อมูล การเปลี่ยนรูปข้อมูล สำหรับเตรียมการชุดข้อมูลนำเข้าระบบจำลองต่อไป หลังจากนั้น เมื่อได้ชุดข้อมูลที่เหมาะสม (fit) ต่อการนำเข้าแบบจำลองแล้ว จะทำการวิเคราะห์ข้อมูลด้วยแบบจำลองทั้ง 6 แบบจำลอง คือ 1) Ridge Model 2) Decision Tree Model 3) Random Forest Model 4) Gradient Boosting Model 5) K-neighbor Model และ 6) MLP Regressor Model ตามสถานการณ์ที่ต้องการทราบ โดยขั้นตอนสุดท้ายผู้วิจัย จะทำการประเมินผลแบบจำลอง (Model Evaluation) ด้วยวิธีการที่ใช้กันอย่างแพร่หลายในงานวิจัยด้วย  $R^2$  RMSE และ MAPE โดยมีภาพรวมของขั้นตอนการดำเนินงานวิจัยดังภาพ 3.1 และรายละเอียดการดำเนินการวิจัยดังนี้



ภาพ 3.1 ภาพรวมของขั้นตอนการดำเนินงานวิจัย



### 3.2 การรวบรวมข้อมูล (Data Collection)

จากชุดข้อมูลอุตุนิยมวิทยาในพื้นที่ศึกษา พบว่า ผลการตรวจวัดฯ จากสถานีตรวจวัด ถูกเก็บมาในรูปแบบรายชั่วโมง และรายงานผลตามค่ามาตรฐาน เช่น PM 2.5 เก็บค่าทุกๆ 1 ชั่วโมง และนำค่ามาเฉลี่ยเป็นค่าเฉลี่ย 24 ชั่วโมง เทียบกับค่ามาตรฐาน

วิธีการตรวจวัดรวมทั้งวิธีการในการเก็บตัวอย่างเป็นไปตามมาตรฐานของ Federal Equivalent Method (FEM) ตามที่องค์การพิทักษ์สิ่งแวดล้อมแห่งประเทศสหรัฐอเมริกา (United States Environmental Protection Agency; US EPA) กำหนด

โดยใช้ข้อมูลอุตุนิยมวิทยาตั้งแต่วันที่ 1 มกราคม 2558 – 31 ธันวาคม 2564 (ก่อนการระบาดของ Covid-19 = 1 มกราคม 2558 - 24 มีนาคม 2563 และสำหรับช่วง Covid-19 = 25 มีนาคม 2563 – 31 ธันวาคม 2564) ทั้งนี้ รายการผลการตรวจวัดจากเครื่องวัดคุณภาพอากาศ ณ สถานีต่างๆ จะถูกแบ่งออกเป็น 2 กลุ่ม คือ 1) กลุ่มชุดข้อมูลอุตุนิยมวิทยา และ 2) กลุ่มชุดข้อมูลด้านมลพิษทางอากาศ โดยมีรายละเอียดดังนี้

#### 3.2.1 ชุดข้อมูลด้านอุตุนิยมวิทยา

##### 1) ความเร็วลม (WS, Wind Speed)

ความเร็วลม (Wind speed) จะเคลื่อนตัวจากความกดอากาศสูงไปยังความดันต่ำ ซึ่งมาจากการเปลี่ยนแปลงของอุณหภูมิ ความเร็วลมสามารถวัดด้วยแอนนิโมมิเตอร์ (Anemometer) ความเร็วลมส่งผลต่อความสามารถของการพยากรณ์อากาศ

##### 2) ทิศทางลม (WD, Wind Direction)

การวัดทิศทางลมจะวัดตามทิศเข็มทิศ โดยแบ่งได้เป็น 4 ทิศหลัก ประกอบไปด้วย ทิศเหนือ ทิศตะวันออก ทิศใต้ และทิศตะวันตก และสามารถแบ่งย่อยลงไปได้ถึง 32 ทิศ แต่โดยทั่วไปจะใช้แค่ 8 ถึง 16 ทิศ

##### 3) อุณหภูมิ (Temp)

อุณหภูมิ คือ สื่อถึงปริมาณความร้อนและปริมาณความเย็น หรือระดับของความร้อน หรือระดับความเย็นของวัตถุ ที่ได้จากการสัมผัส

สามารถวัดระดับความร้อนวัดด้วยอุปกรณ์ที่เรียกว่า เทอร์โมมิเตอร์ โดยหน่วยการวัดที่นิยมใช้มากที่สุดคือ องศาเซลเซียส ( $^{\circ}\text{C}$ ) เคลวิน (K) และฟาเรนไฮต์ ( $^{\circ}\text{F}$ ) โดยส่วนใหญ่หน่วยเคลวินจะใช้ในทางวิทยาศาสตร์เป็นหลัก

##### 4) ความกดอากาศ

ความกดอากาศ (Atmospheric pressure) หรือ ความดันบรรยากาศ เป็นความดันภายใต้ชั้นบรรยากาศโลก โดยปกติความกดอากาศจะมีค่าต่ำลงเมื่อระดับความสูงของภูมิประเทศเพิ่มขึ้น

##### 5) ความชื้นสัมพัทธ์ (RH%, Relative Humidity)

ความชื้น (Humidity) คือ ปริมาณไอน้ำที่อยู่ในอากาศ ความชื้นของอากาศมีการเปลี่ยนแปลงโดยตลอด จะน้อยหรือมากขึ้นอยู่กับอุณหภูมิและความดัน

### 3.2.2 ชุดข้อมูลสารมลพิษทางอากาศ

โดยชุดข้อมูลสารมลพิษทางอากาศ ได้มาจากข้อมูลดัชนีคุณภาพอากาศ กองการจัดการคุณภาพอากาศและเสียง กรมควบคุมมลพิษ (2564) โดยมีรายละเอียด ดังนี้

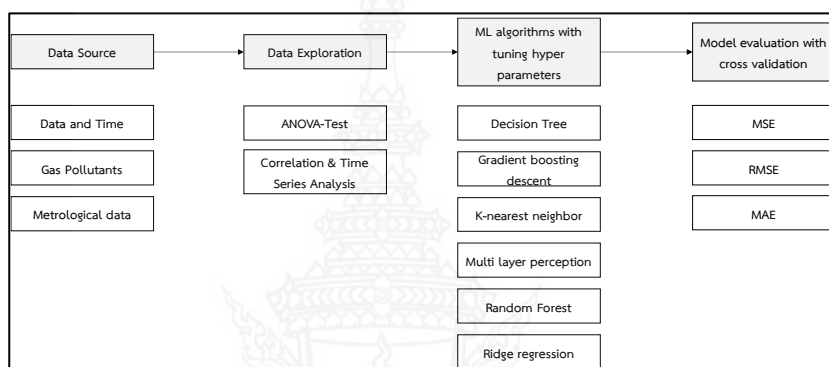
- 1) PM 2.5 ฝุ่นละอองขนาดเล็กไม่เกิน 2.5 ไมครอน (PM 2.5)  
ฝุ่นละอองขนาดเล็กผ่านศูนย์กลางไม่เกินขนาด 2.5 ไมครอน ซึ่งเกิดจากการเผาวัสดุ การเกษตร การเผาไหม้จากยานพาหนะ กระบวนการอุตสาหกรรม และไฟ สามารถเข้าสู่ระบบทางเดินหายใจได้ง่าย ส่งผลทำให้เกิดโรคเกี่ยวกับระบบทางเดินหายใจ
- 2) ฝุ่นหยาบ (Course Particle) (PM 10)  
ฝุ่นละอองขนาดเล็กผ่านศูนย์กลางไม่เกิน 10 ไมครอน เกิดจากการเผาในที่โล่ง กระบวนการอุตสาหกรรม การเผาไหม้เชื้อเพลิง หรือการทำให้เป็นผงจากการก่อสร้าง ส่งผลกระทบต่อสุขภาพและระบบทางเดินหายใจ
- 3) ไนโตรเจนออกไซด์ (NO<sub>x</sub>)  
NO<sub>x</sub> หรือ ไนโตรเจนออกไซด์เป็นก๊าซซึ่งเกิดจากการเผาไหม้ของเชื้อเพลิงฟอสซิล โดยมีรถยนต์ที่ใช้เครื่องยนต์ดีเซลเป็นตัวการสำคัญที่ก่อให้เกิดมลพิษ เนื่องจากการเผาไหม้ที่ไม่สมบูรณ์
- 4) ไนโตรเจนไดออกไซด์ (NO<sub>2</sub>)  
ก๊าซไนโตรเจนไดออกไซด์ (NO<sub>2</sub>) เป็นก๊าซที่ไม่มีกลิ่นและสี มีอยู่ทั่วไปในธรรมชาติ และบางส่วนเกิดจากการกระทำของมนุษย์ เช่น การเผาไหม้เชื้อเพลิงจากอุตสาหกรรมบางชนิด ก๊าซนี้มีผลต่อระบบการมองเห็นและระบบทางเดินหายใจ
- 5) ไนตริกออกไซด์ (NO)  
ก๊าซไนตริกออกไซด์ (NO) เป็นก๊าซที่ไม่มีกลิ่นและสี โดยทั่วไปในธรรมชาติพบในปริมาณน้อยกว่า 0.5 ppm. มีคุณสมบัติทำมห้สลบได้
- 6) คาร์บอนมอนอกไซด์ (CO)  
ก๊าซคาร์บอนมอนอกไซด์ (CO) ไม่มีกลิ่น สี และรส โดยทั่วไปเกิดจากการเผาไหม้ที่ไม่สมบูรณ์ของเชื้อเพลิงที่มีคาร์บอนเป็นส่วนประกอบ สามารถจับกับฮีโมโกลบินในเลือด ทำให้ประสิทธิภาพในการลำเลียงออกซิเจนลดลง
- 7) โอโซน (O<sub>3</sub>)  
ก๊าซโอโซน (O<sub>3</sub>) มีสี มีกลิ่นฉุน และละลายน้ำได้บ้างเล็กน้อย ก๊าซโอโซนที่เป็นสารมลพิษทางอากาศคือก๊าซโอโซนที่เกิดจากปฏิกิริยาระหว่างก๊าซออกไซด์ของไนโตรเจนและสารประกอบอินทรีย์ ซึ่งส่งผลกระทบต่อสุขภาพ

โดยข้อมูลข้างต้นที่ได้มาจากกรมควบคุมมลพิษ และสำนักสิ่งแวดล้อม กรุงเทพมหานคร มีทั้งหมด 2.3 ล้านข้อมูล รวมทั้งข้อมูลด้านอุตุนิยมวิทยา และข้อมูลมลพิษทางอากาศ ตามห้วงเวลาที่กำหนด ดังแสดงในตาราง 3.1 ทั้งนี้ ตัวอย่างชุดข้อมูล Dataset ที่ใช้ในการศึกษา ดังภาพ 3.2

ตาราง 3.1 Raw Dataset

| ที่มาของข้อมูล                | จำนวนสถานี | ข้อมูลทั้งหมด (แถว)<br>(ปีพ.ศ. 2558-2564) |
|-------------------------------|------------|---|
| กรมควบคุมมลพิษ                | 13         | 787,550                                   |
| สำนักสิ่งแวดล้อมกรุงเทพมหานคร | 50         | 1,516,342                                 |
| ทั้งหมด                       | 63         | 2,303,892                                 |

ที่มา: กรมควบคุมมลพิษ (2564)



ภาพ 3.2 ตัวอย่างชุดข้อมูล Dataset ที่ใช้ในการศึกษา

จากข้อมูลทั้ง 2.3 ล้านชุด เมื่อแบ่งข้อมูลตามสถานีตรวจวัดทั่วกรุงเทพมหานคร พบว่าได้ข้อมูลทั้งหมด ดังนี้

ตาราง 3.2 สรุปจำนวนชุดข้อมูลแยกรายสถานี

| หน่วยงาน       | ลำดับที่ | ชื่อสถานีตรวจวัด (Location)  | Location |           | จำนวนข้อมูล (แถว) |
|----------------|----------|--|----------|-----------|-------------------|
|                |          |  | Latitude | Longitude |                   |
| กรมควบคุมมลพิษ | 1        | (02t) มหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา แขวงหิรัญรูจี เขตธนบุรี, กรุงเทพฯ | 100.4893 | 13.73286  | 61,368            |
| กรมควบคุมมลพิษ | 2        | (03t) रिमถนนทางหลวงหมายเลข 3902 रिमถนนกาญจนาภิเษก เขตบางขุนเทียน, กรุงเทพฯ   | 100.4153 | 13.63742  | 18,960            |
| กรมควบคุมมลพิษ | 3        | (03t) มหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา แขวงหิรัญรูจี เขตธนบุรี, กรุงเทพฯ | 100.4153 | 13.63742  | 42,408            |
| กรมควบคุมมลพิษ | 4        | (05t) กรมอุตุนิยมวิทยาบางนา แขวงบางนา เขตบางนา, กรุงเทพฯ                     | 100.6066 | 13.66615  | 61,368            |
| กรมควบคุมมลพิษ | 5        | (10t) เคหะชุมชนคลองจั่น แขวงคลองจั่น เขตบางกะปิ, กรุงเทพฯ                    | 100.6451 | 13.77981  | 61,368            |

ตาราง 3.2 (ต่อ)

| หน่วยงาน              | ลำดับ<br>ที่ | ชื่อสถานีตรวจวัด (Location)   | Location |          | จำนวน<br>ข้อมูล (แถว) |
|-----------------------|--------------|---|----------|----------|-----------------------|
|                       |              |   |          |          |                       |
| กรมควบคุมมลพิษ        | 6            | (11t) การเคหะชุมชนห้วยขวาง แขวงดินแดง เขตดินแดง, กรุงเทพฯ                     | 100.5687 | 13.77681 | 61,368                |
| กรมควบคุมมลพิษ        | 7            | (12t) โรงเรียนนนทรีวิทยา แขวงช่องนนทรี เขตยานนาวา, กรุงเทพฯ                   | n/a      | n/a      | 61,368                |
| กรมควบคุมมลพิษ        | 8            | (50t) โรงพยาบาลจุฬาลงกรณ์ ริมถนนพระราม 4 เขตปทุมวัน, กรุงเทพฯ                 | 100.5409 | 13.74517 | 61,368                |
| กรมควบคุมมลพิษ        | 9            | (52t) การไฟฟ้าอ้อยอนบุรี ริมถนนอินทรพิทักษ์ เขตธนบุรี, กรุงเทพฯ               | 100.4874 | 13.72857 | 61,368                |
| กรมควบคุมมลพิษ        | 10           | (53t) สถานีตำรวจนครบาลโชคชัย ริมถนนลาดพร้าว เขตวังทองหลาง, กรุงเทพฯ           | 100.595  | 13.79604 | 61,368                |
| กรมควบคุมมลพิษ        | 11           | (54t) การเคหะชุมชนดินแดง ริมถนนดินแดง เขตดินแดง, กรุงเทพฯ                     | n/a      | n/a      | 61,368                |
| กรมควบคุมมลพิษ        | 12           | (59t) กรมประชาสัมพันธ์ แขวงพญาไท เขตพญาไท, กรุงเทพฯ                           | 100.5395 | 13.78311 | 61,368                |
| กรมควบคุมมลพิษ        | 13           | (61t) โรงเรียนบดินทรเดชา (สิงห์ สิงหเสนี) แขวงพลับพลา เขตวังทองหลาง, กรุงเทพฯ | 100.5134 | 13.7574  | 61,368                |
| สำนักสิ่งแวดล้อม กทม. | 14           | คลองสาน   | 100.5087 | 13.71977 | 32,904                |
| สำนักสิ่งแวดล้อม กทม. | 15           | คลองสามวา   | 100.7045 | 13.8602  | 32,904                |
| สำนักสิ่งแวดล้อม กทม. | 16           | คลองเตย   | 100.5838 | 13.7085  | 33,000                |
| สำนักสิ่งแวดล้อม กทม. | 17           | คันนายาว  | 100.6869 | 13.81396 | 32,904                |
| สำนักสิ่งแวดล้อม กทม. | 18           | จตุจักร   | 100.5759 | 13.84015 | 33,000                |
| สำนักสิ่งแวดล้อม กทม. | 19           | จอมทอง  | 100.4837 | 13.67752 | 33,320                |
| สำนักสิ่งแวดล้อม กทม. | 20           | ดอนเมือง  | 100.595  | 13.91087 | 34,344                |
| สำนักสิ่งแวดล้อม กทม. | 21           | ดุสิต   | 100.5208 | 13.77675 | 32,904                |
| สำนักสิ่งแวดล้อม กทม. | 22           | ตลิ่งชัน  | 100.4268 | 13.78083 | 32,904                |

ตาราง 3.2 (ต่อ)

| หน่วยงาน                 | ลำดับ<br>ที่ | ชื่อสถานีตรวจวัด (Location) | Location |          | จำนวน        |
|--------------------------|--------------|-----------------------------|----------|----------|--------------|
|                          |              |                             |          |          | ข้อมูล (แถว) |
| สำนักสิ่งแวดล้อม<br>กทม. | 23           | ทวีวัฒนา                    | 100.3551 | 13.74612 | 32,904       |
| สำนักสิ่งแวดล้อม<br>กทม. | 24           | ทุ่งครุ                     | 100.4966 | 13.65101 | 32,928       |
| สำนักสิ่งแวดล้อม<br>กทม. | 25           | ธนบุรี                      | 100.4847 | 13.70523 | 32,904       |
| สำนักสิ่งแวดล้อม<br>กทม. | 26           | บางกอกน้อย                  | 100.4856 | 13.75822 | 32,928       |
| สำนักสิ่งแวดล้อม<br>กทม. | 27           | บางกอกใหญ่                  | 100.4739 | 13.72893 | 32,594       |
| สำนักสิ่งแวดล้อม<br>กทม. | 28           | บางกะปิ                     | 100.6476 | 13.76641 | 33,000       |
| สำนักสิ่งแวดล้อม<br>กทม. | 29           | บางขุนเทียน                 | 100.4358 | 13.66045 | 32,904       |
| สำนักสิ่งแวดล้อม<br>กทม. | 30           | บางคอแหลม                   | 100.5028 | 13.69305 | 31,440       |
| สำนักสิ่งแวดล้อม<br>กทม. | 31           | บางซื่อ                     | 100.538  | 13.80955 | 32,928       |
| สำนักสิ่งแวดล้อม<br>กทม. | 32           | บางนา                       | 100.6361 | 13.66831 | 32,905       |
| สำนักสิ่งแวดล้อม<br>กทม. | 33           | บางบอน                      | 100.4247 | 13.67362 | 32,904       |
| สำนักสิ่งแวดล้อม<br>กทม. | 34           | บางพลัด                     | 100.5055 | 13.79374 | 32,904       |
| สำนักสิ่งแวดล้อม<br>กทม. | 35           | บางรัก                      | 100.527  | 13.7282  | 32,903       |
| สำนักสิ่งแวดล้อม<br>กทม. | 36           | บางเขน                      | 100.6413 | 13.85086 | 32,904       |
| สำนักสิ่งแวดล้อม<br>กทม. | 37           | บางแค                       | 100.3843 | 13.6864  | 32,904       |
| สำนักสิ่งแวดล้อม<br>กทม. | 38           | บึงกุ่ม                     | 100.6695 | 13.78535 | 32,904       |
| สำนักสิ่งแวดล้อม<br>กทม. | 39           | ปทุมวัน                     | 100.5284 | 13.73317 | 32,904       |

ตาราง 3.2 (ต่อ)

| หน่วยงาน                 | ลำดับ<br>ที่ | ชื่อสถานีตรวจวัด (Location) | Location |          | จำนวน        |
|--------------------------|--------------|-----------------------------|----------|----------|--------------|
|                          |              |                             |          |          | ข้อมูล (แถว) |
| สำนักสิ่งแวดล้อม<br>กทม. | 40           | ประเวศ                      | 100.6483 | 13.69602 | 32,915       |
| สำนักสิ่งแวดล้อม<br>กทม. | 41           | ป้อมปราบ                    | 100.5143 | 13.75657 | 8,760        |
| สำนักสิ่งแวดล้อม<br>กทม. | 42           | ป้อมปราบศัตรูพ่าย           | 100.5143 | 13.75657 | 24,144       |
| สำนักสิ่งแวดล้อม<br>กทม. | 43           | พญาไท                       | 100.5337 | 13.78165 | 32,928       |
| สำนักสิ่งแวดล้อม<br>กทม. | 44           | พระนคร                      | 100.4993 | 13.76402 | 32,928       |
| สำนักสิ่งแวดล้อม<br>กทม. | 45           | ภาษีเจริญ                   | 100.4536 | 13.71847 | 32,904       |
| สำนักสิ่งแวดล้อม<br>กทม. | 46           | มีนบุรี                     | 100.732  | 13.81344 | 32,904       |
| สำนักสิ่งแวดล้อม<br>กทม. | 47           | ยานนาวา                     | 100.547  | 13.67933 | 33,624       |
| สำนักสิ่งแวดล้อม<br>กทม. | 48           | ราชเทวี                     | 100.5346 | 13.75929 | 51,134       |
| สำนักสิ่งแวดล้อม<br>กทม. | 49           | ลาดกระบัง                   | 100.7833 | 13.72199 | 32,927       |
| สำนักสิ่งแวดล้อม<br>กทม. | 50           | ลาดพร้าว                    | 100.607  | 13.80338 | 32,904       |
| สำนักสิ่งแวดล้อม<br>กทม. | 51           | วังทองหลาง                  | 100.6224 | 13.77905 | 32,904       |
| สำนักสิ่งแวดล้อม<br>กทม. | 52           | วัฒนา                       | 100.3551 | 13.74612 | 32,893       |
| สำนักสิ่งแวดล้อม<br>กทม. | 53           | สวนหลวง                     | 100.652  | 13.7310  | 32,160       |
| สำนักสิ่งแวดล้อม<br>กทม. | 54           | สะพานสูง                    | 100.6857 | 13.76868 | 35,175       |
| สำนักสิ่งแวดล้อม<br>กทม. | 55           | สัมพันธวงศ์                 | 100.5121 | 13.73683 | 32,904       |
| สำนักสิ่งแวดล้อม<br>กทม. | 56           | สาทร                        | 100.5264 | 13.70804 | 32,904       |

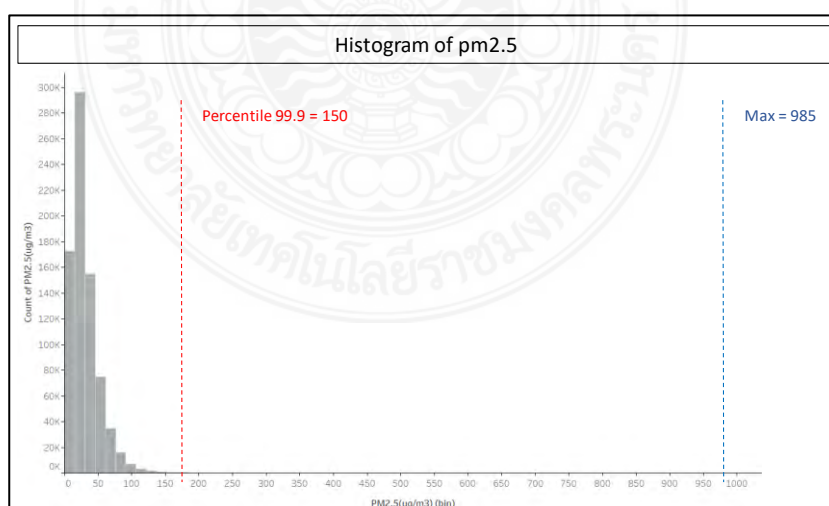
ตาราง 3.2 (ต่อ)

| หน่วยงาน                 | ลำดับ<br>ที่ | ชื่อสถานีตรวจวัด (Location) | Location |          | จำนวน<br>ข้อมูล (แถว) |
|--------------------------|--------------|-----------------------------|----------|----------|-----------------------|
|                          |              |                             |          |          |                       |
| สำนักสิ่งแวดล้อม<br>กทม. | 57           | สายไหม                      | 100.6603 | 13.8956  | 32,904                |
| สำนักสิ่งแวดล้อม<br>กทม. | 58           | หนองจอก                     | 100.8620 | 13.85533 | 32,904                |
| สำนักสิ่งแวดล้อม<br>กทม. | 59           | หนองแขม                     | 100.3431 | 13.70538 | 32,903                |
| สำนักสิ่งแวดล้อม<br>กทม. | 60           | หลักสี่                     | 100.5793 | 13.88734 | 32,904                |
| สำนักสิ่งแวดล้อม<br>กทม. | 61           | ห้วยขวาง                    | 100.5795 | 13.77651 | 32,903                |

### 3.2.3 การตรวจสอบแก้ไขข้อมูลเบื้องต้น (Data Cleansing)

การตรวจสอบแก้ไขข้อมูล (Editing and Cleaning) เป็นการตรวจสอบความถูกต้องและความเป็นไปได้ของข้อมูล ซึ่งกำหนดให้ข้อมูลทั่วไปที่นำมาใช้ได้มีค่ามากกว่าหรือเท่ากับ 0.01 (มีค่าไม่เป็นลบ) และปรับปรุงข้อมูลที่ขาดหายไป โดยเอาข้อมูลชั่วโมงก่อนและหลังมาเฉลี่ยในกรณีข้อมูลขาดหายไปบางส่วน ทั้งนี้ ข้อมูลแต่ละประเภทมีการตรวจสอบและแก้ไข โดยมีรายละเอียดดังต่อไปนี้

โดยหากพบค่า PM 2.5 มากกว่า  $985 \text{ ug/m}^3$  จะถูกลบออกไป เนื่องจากถือเป็น Error ของข้อมูล โดยจากการทำ histogram ของค่า PM 2.5 พบว่า ค่า Max ของ PM 2.5 เท่ากับ  $985 \text{ ug/m}^3$  ดังแสดงในภาพ 3.3 ในขณะที่ข้อมูลค่า PM 2.5 มีค่าน้อยกว่า 0.01 จะถูกลบออกไปเช่นเดียวกัน เนื่องจากถือเป็น Error ของการตรวจวัดจากสถานีตรวจฯ



ภาพ 3.3 การทำ histogram ของค่า PM 2.5

อย่างไรก็ตาม ก่อนนำข้อมูลชุดเรียนรู้เข้าแบบจำลอง จะทำการเอาข้อมูลที่มีค่าผิดปกติออก โดยจะพิจารณาข้อมูลที่มีค่ามากกว่าค่า percentile ที่ 99.9 เป็นข้อมูลผิดปกติ โดยค่า PM 2.5 กำหนดให้ข้อมูลผิดปกติมีค่ามากกว่า  $150 \text{ ug/m}^3$ , PM 10 กำหนดให้ข้อมูลผิดปกติมีค่ามากกว่า  $212 \text{ ug/m}^3$ ,  $\text{NO}_x$  กำหนดให้ข้อมูลผิดปกติมีค่ามากกว่า 337 ppb,  $\text{NO}_2$  กำหนดให้ข้อมูลผิดปกติมีค่ามากกว่า 131 ppb, NO กำหนดให้ข้อมูลผิดปกติมีค่ามากกว่า 287 ppb, CO กำหนดให้ข้อมูลผิดปกติมีค่ามากกว่า 4 ppm, ความเร็วลม กำหนดให้ข้อมูลผิดปกติมีค่ามากกว่า 4 m/s, ทิศทางลม กำหนดให้ข้อมูลผิดปกติมีค่ามากกว่า  $360 \text{ }^\circ\text{C}$ , อุณหภูมิ กำหนดให้ข้อมูลผิดปกติมีค่ามากกว่า  $45 \text{ }^\circ\text{C}$ , ความกดอากาศกำหนดให้ข้อมูลผิดปกติมีค่ามากกว่า 1019 mBar, ความชื้นสัมพัทธ์ กำหนดให้ข้อมูลผิดปกติมีค่ามากกว่า 100% และโอโซน กำหนดให้ข้อมูลผิดปกติมีค่ามากกว่า 137 ppb.

### 3.2.4 การตรวจสอบแก้ไขข้อมูลเบื้องต้น (Data Cleansing)

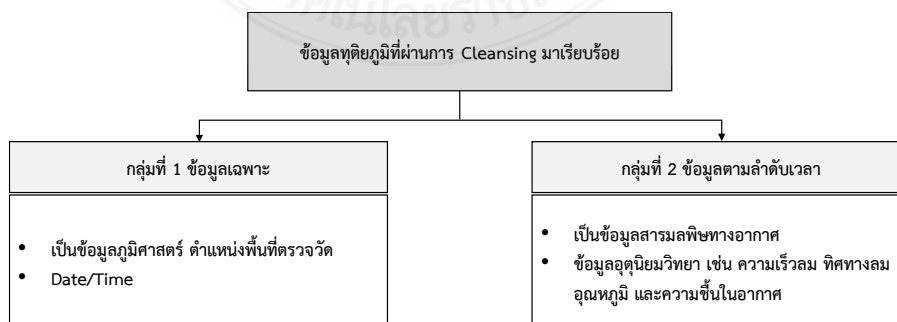
อย่างไรก็ตาม จากข้อมูลตั้งต้นทั้งหมด 2.3 ล้านข้อมูล ทั้งข้อมูลด้านอุตุนิยมวิทยา และข้อมูลมลพิษทางอากาศ เมื่อทำการตรวจสอบแก้ไขข้อมูลเบื้องต้นเสร็จเรียบร้อยแล้ว ทำให้เหลือชุดข้อมูลตั้งต้น (Data) ที่สามารถนำไปใช้งานในการวิเคราะห์ในขั้นตอนต่อไป โดยผู้วิจัยได้ทำการแบ่งชุดข้อมูลออกเป็นทั้งหมด 2 กลุ่ม คือ 1) ข้อมูลเฉพาะ และ 2) ข้อมูลตามลำดับเวลา ทั้งนี้ เพื่อให้รูปแบบของชุดข้อมูลเหมาะสมต่อการนำไปใช้ในการวิเคราะห์ในขั้นตอนต่อไป ซึ่งการแบ่งกลุ่มชุดข้อมูลมีรายละเอียดดังนี้

#### 1) กลุ่มที่ 1 ข้อมูลเฉพาะ

กลุ่มข้อมูลเฉพาะ หมายถึง ชุดข้อมูลที่มีลักษณะเฉพาะเจาะจง เป็นข้อมูลที่ใช้บ่งบอกคุณลักษณะของข้อมูลนั้น ๆ เช่น ข้อมูลภูมิศาสตร์ ตำแหน่งพื้นที่ตรวจวัด หมายเลขสถานีตรวจวัด วัดที่ เวลา เดือน ปี เป็นต้น

#### 2) กลุ่มที่ 2 ข้อมูลตามลำดับเวลา

กลุ่มข้อมูลตามลำดับเวลา หมายถึง ชุดข้อมูลที่แสดงการพัฒนาหรือวิวัฒนาการตามลำดับเวลา ที่ถูกเก็บย้อนหลังเป็นเวลาหลายๆ ปี เพื่อจะได้นำไปวิเคราะห์เปรียบเทียบหาแนวโน้ม ของข้อมูลเปรียบเทียบกับปีที่ผ่านมา ประกอบด้วย ข้อมูลอุตุนิยมวิทยา เช่น ความเร็วลม ทิศทางลม อุณหภูมิ และความชื้นในอากาศ เป็นต้น และ ข้อมูลสารมลพิษทางอากาศ เช่น PM 2.5 PM 10 และ CO เป็นต้น



ภาพ 3.4 ภาพรวมการแบ่งกลุ่มข้อมูล (Grouping Data)



### 3.3 การตรวจสอบและสำรวจข้อมูลเบื้องต้น (Data Exploration)

ในขั้นตอนการตรวจสอบและสำรวจข้อมูลเบื้องต้น (Data Exploration) ผู้วิจัยจะทำการวิเคราะห์ทางสถิติของชุดข้อมูล (Descriptive Statistics) การตรวจสอบความสัมพันธ์ของชุดข้อมูล (Correlation Analysis) และทำการวิเคราะห์ชุดข้อมูลตามลำดับเวลา (Time-Series Analysis) เพื่อให้สามารถเข้าใจพฤติกรรมของชุดข้อมูลเบื้องต้นโดยละเอียด

ผู้วิจัยได้นำชุดข้อมูลมาแบ่งออกตามช่วงเวลาออกเป็น 2 ช่วง ประกอบไปด้วย ช่วงสถานการณ์ปกติ (1 มกราคม 2558 - 24 มีนาคม 2563) และช่วงที่เกิดสถานการณ์การแพร่ระบาดของ Covid-19 (25 มีนาคม 2563 - 31 ธันวาคม 2564) โดยนำข้อมูลทั้งหมดมาทำการทดสอบว่าช่วงเหตุการณ์ Covid-19 มีผลต่อ PM 2.5 หรือไม่ ซึ่งได้แยกข้อมูลรายเดือนในแต่ละปี ตั้งแต่ปี พ.ศ. 2558 ถึงปี พ.ศ. 2564 การวิเคราะห์มีดังต่อไปนี้

#### 3.3.1 การวิเคราะห์สถิติเชิงพรรณนา (Descriptive Statistics)

- 1) การวิเคราะห์พฤติกรรมของ PM 2.5 ในสถานการณ์ปกติ และระหว่างการแพร่ระบาดของ Covid-19

ข้อมูลด้านอนุกรมวิธานและสารมลพิษต่าง ๆ ที่ถูกแบ่งออกเป็น 2 ช่วง คือ ก่อนการระบาดของของโรค COVID-19 (ชุดข้อมูลปี พ.ศ. 2558 - 2562) และระหว่างการแพร่ระบาดของโรค Covid-19 เมื่อมีการ Lockdown (ชุดข้อมูลปี พ.ศ. 2563-2564) ในตาราง 3.3 แสดงให้เห็นถึงภาพรวมของความเข้มข้นของมลพิษก่อนเกิดการแพร่ระบาดของ Covid-19 และระหว่าง Lockdown เมื่อผู้วิจัยทำการเปรียบเทียบข้อมูลความเข้มข้นของมลพิษทางอากาศรายชั่วโมง เฉลี่ยของทั้ง 63 สถานี ในพื้นที่กรุงเทพมหานคร ทำให้เห็นความแตกต่างและแนวโน้มของความเข้มข้นมลพิษทางอากาศที่เกิดขึ้นก่อนการระบาดของของ Covid-19 และระหว่าง Lockdown โดยพบว่า ช่วงระหว่างที่เกิดการระบาดของ Covid-19 ค่าเฉลี่ยของ PM 2.5 NO<sub>x</sub> NO<sub>2</sub> NO CO มีค่าลดลงจากช่วงก่อนการระบาดของ Covid-19 ในขณะที่ ค่าเฉลี่ยของ PM 10 และ O<sub>3</sub> นั้นกลับเพิ่มสูงขึ้น เมื่อมีการระบาดของ Covid-19

เมื่อพิจารณาในมิติของการแพร่ระบาดของ Covid-19 และมีการ Lockdown ตั้งแต่เดือน มีนาคม พ.ศ. 2563 เป็นต้นไป ซึ่งเป็นช่วงเวลาของการจำกัดการเดินทาง ไม่มีรถวิ่งบนถนน ไม่มีประชาชนออกไปใช้ชีวิตในสถานที่สาธารณะ ไม่มีกิจกรรมใดๆ นอกบ้าน พบว่า ช่วงเวลาดังกล่าวของปี ค่าความเข้มข้นของ PM 2.5, NO<sub>x</sub>, NO<sub>2</sub>, NO, และ CO ลดลงกว่าช่วงเวลาดังกล่าวของปีที่ไม่มีการ Lockdown เกิดขึ้น 14.38%, 12.93%, 18.43%, 2.65% และ 5.66% ตามลำดับ

ตาราง 3.3 สถิติด้านมลพิษทางอากาศช่วงก่อนการระบาดของ Covid-19 และระหว่างการแพร่ระบาดของ Covid-19

| พารามิเตอร์ด้านมลพิษ        | ก่อนการระบาดของ Covid-19 |                           | ระหว่างการแพร่ระบาดของ Covid-19 |                           | % การเปลี่ยนแปลง |
|-----------------------------|--------------------------|---------------------------|---------------------------------|---------------------------|------------------|
|                             | ค่ากลาง Mean             | ส่วนเบี่ยงเบนมาตรฐาน S.D. | ค่ากลาง Mean                    | ส่วนเบี่ยงเบนมาตรฐาน S.D. |                  |
| PM 2.5 (ug/m <sup>3</sup> ) | 27.93                    | 18.94                     | 23.92                           | 16.21                     | -14.38%          |
| PM 10 (ug/m <sup>3</sup> )  | 44.97                    | 26.19                     | 47.45                           | 27.94                     | 5.51%            |
| NO <sub>x</sub> (ppb)       | 42.15                    | 40.88                     | 36.70                           | 41.32                     | -12.93%          |
| NO <sub>2</sub> (ppb)       | 22.15                    | 15.10                     | 18.07                           | 14.35                     | -18.43%          |
| NO (ppb)                    | 23.00                    | 32.96                     | 22.39                           | 33.93                     | -2.65%           |
| CO (ppm)                    | 0.85                     | 0.55                      | 0.80                            | 0.55                      | -5.66%           |
| WS (m/s)                    | 0.84                     | 0.66                      | 0.65                            | 0.52                      | -23.40%          |
| WD (Deg)                    | 188.56                   | 92.74                     | 186.22                          | 97.33                     | -1.24%           |
| Temp (Deg.C)                | 29.42                    | 2.92                      | 29.18                           | 3.09                      | -0.82%           |
| BP (mBar)                   | 869.66                   | 59.13                     | 978.40                          | 83.61                     | 12.50%           |
| RH (%)                      | 67.42                    | 14.42                     | 69.97                           | 14.91                     | 3.78%            |
| O <sub>3</sub> (ppb)        | 19.42                    | 18.10                     | 22.24                           | 17.90                     | 14.47%           |

\*หมายเหตุ: 99 เปอร์เซ็นต์ไทล์ของชุดข้อมูล

ที่มา: รวบรวมและวิเคราะห์โดยผู้วิจัย

อย่างไรก็ตาม ผู้วิจัยได้นำค่าความเข้มข้นของมลพิษทางอากาศเฉลี่ย 24 ชั่วโมง ของค่ามลพิษต่าง ๆ มาทำการเปรียบเทียบเดือนต่อเดือนของช่วงก่อนการระบาดของ Covid-19 ระหว่างปี พ.ศ. 2558-2562 และระหว่างที่เกิดการระบาดของ Covid-19 ปี พ.ศ. 2563-2564 โดยแสดงรายละเอียดใน ภาพ 3.5 (ก) PM 2.5 (ข) PM 10 (ค) NO<sub>x</sub> (ง) NO<sub>2</sub> (จ) NO (ฉ) CO (ช) O<sub>3</sub> แสดงให้เห็นว่า ค่าความเข้มข้นของมลพิษทางอากาศในแต่ละชนิด จะสูงในช่วงหน้าหนาว (Dry Season) ทั้งก่อนการระบาดของ Covid-19 และระหว่างที่เกิดการระบาดของ Covid-19 และความเข้มข้นของมลพิษทางอากาศจะลดลงในช่วงฤดูฝน (Wet Season) ทั้งก่อนการระบาดของ Covid-19 และระหว่างที่เกิดการระบาดของ Covid-19 โดยมีสถิติเป็นไปในทิศทางเดียวกัน

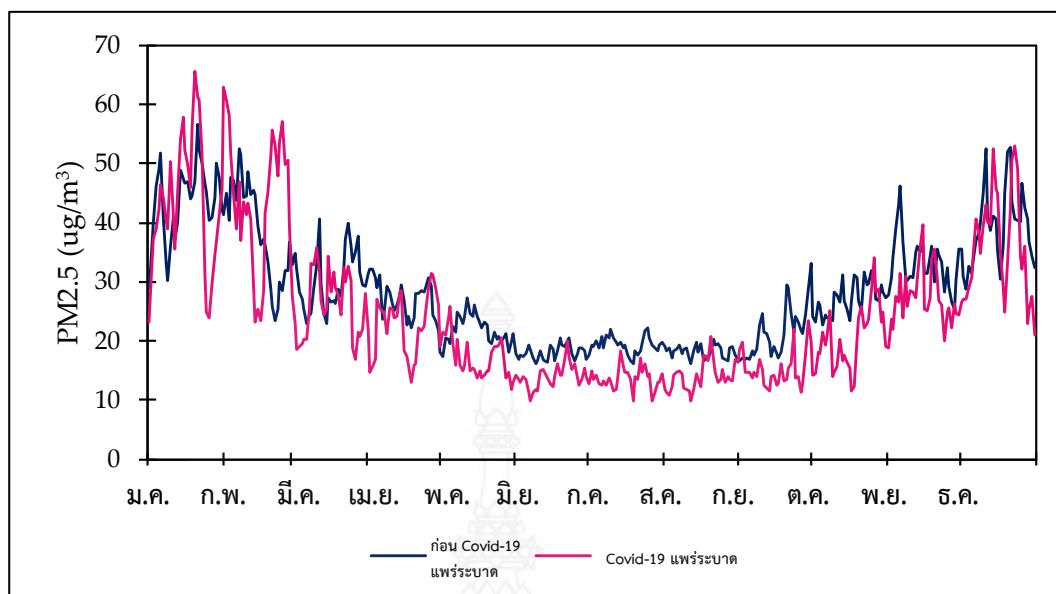
จากภาพ 3.5 (ก) ซึ่งแสดงรายละเอียดของความเข้มข้นเฉลี่ย PM 2.5 ในแต่ละเดือนของช่วงเวลาสถานการณ์ปกติ ตั้งแต่ปี พ.ศ. 2558 - 2562 (เส้นสีน้ำเงิน) และความเข้มข้นเฉลี่ย PM 2.5 ในแต่ละเดือน ของช่วงสถานการณ์การแพร่ระบาดของ Covid-19 ตั้งแต่ปี พ.ศ. 2563 - 2564 (เส้นสีแดง) พบว่า ค่า PM 2.5 ในกรุงเทพมหานครสูงสุดเฉลี่ยต่อเดือน ประมาณ 60 - 70 ug/m<sup>3</sup> และค่าต่ำที่สุดประมาณ 10 ug/m<sup>3</sup> ทั้งนี้ อาจแตกต่างกันไปในแต่ละบริเวณ เช่น พื้นที่บริเวณที่มีการจราจรหนาแน่นอาจมีความเข้มข้นของ PM 2.5 ที่สูงมาก ในขณะที่พื้นที่บริเวณสวนสาธารณะอาจจะมี PM 2.5 ต่ำ ทั้งนี้ ค่าความเข้มข้นที่นำมาวิเคราะห์ในการศึกษานี้ เป็นค่าความเข้มข้นของ PM 2.5 เฉลี่ยโดยทั่วบริเวณพื้นที่ของเขตกรุงเทพมหานคร

โดยจากสถิติดังกล่าว พบว่า ค่าความเข้มข้นของ PM 2.5 จะเพิ่มสูงขึ้นในห้วงฤดูหนาว (Dry Season) และจะลดต่ำลงในห้วงฤดูฝน (Wet Season) แม้ว่าจะอยู่ในห้วงการระบาดของ Covid-19 ค่าความเข้มข้นของ PM 2.5 ก็มีรูปแบบพฤติกรรมความเข้มข้นไม่ต่างจากสถานการณ์ปกติ ก่อนที่จะมีการระบาดของ Covid-19 แต่อย่างไร ซึ่งจะแตกต่างกันเพียงแค่ว่าความเข้มข้นของ PM 2.5 ในห้วงที่มีการระบาดของ Covid-19 อาจจะมีค่าต่ำกว่าช่วงก่อนการระบาดเท่านั้น

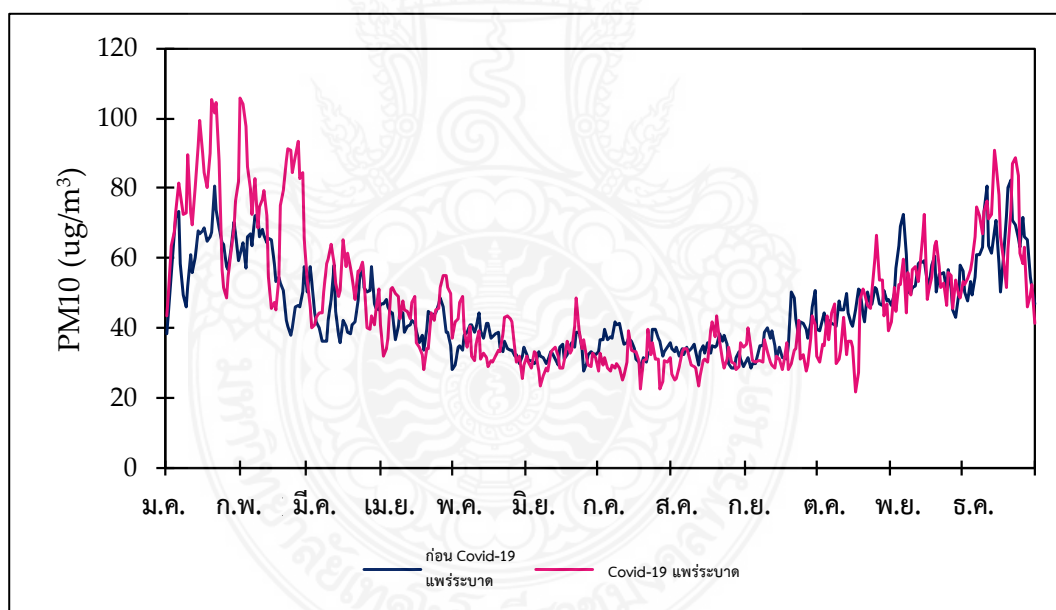
เมื่อพิจารณาเปรียบเทียบความเข้มข้นของ PM 2.5 ระหว่างสถานการณ์ปกติและสถานการณ์การแพร่ระบาดของ Covid-19 พบว่า ในช่วงฤดูหนาวค่าความเข้มข้น PM 2.5 ในฤดูหนาวระหว่างการแพร่ระบาดของ Covid-19 สูงกว่าช่วงสถานการณ์ปกติ (เดือน มกราคม - มีนาคม) แต่ในช่วงฤดูฝนของห้วงการแพร่ระบาดของ Covid-19 กลับพบว่ามีค่าความเข้มข้น PM 2.5 ต่ำกว่าห้วงสถานการณ์ปกติ (เดือน พฤษภาคม - ตุลาคม) พบว่า ฤดูกาลเป็นหนึ่งในปัจจัยที่สำคัญต่อความเข้มข้น PM 2.5 ที่เกิดขึ้น

จากภาพ 3.5 (ข) ซึ่งแสดงรายละเอียดของความเข้มข้นเฉลี่ย PM 10 ในแต่ละเดือนของช่วงเวลาสถานการณ์ปกติ ตั้งแต่ปี พ.ศ. 2558-2562 (เส้นสีน้ำเงิน) และความเข้มข้นเฉลี่ย PM 10 ในแต่ละเดือน ของช่วงสถานการณ์การแพร่ระบาดของ Covid-19 ตั้งแต่ปี พ.ศ. 2563-2564 (เส้นสีแดง) พบว่า ค่า PM 10 ในกรุงเทพมหานครสูงสุดเฉลี่ยต่อเดือน ประมาณ  $100 \text{ ug/m}^3$  และค่าต่ำที่สุดประมาณ  $20 \text{ ug/m}^3$  ทั้งนี้ อาจแตกต่างกันไปในแต่ละบริเวณ เช่น พื้นที่บริเวณที่มีการก่อสร้าง และมีอุตสาหกรรมหนาแน่น อาจมีความเข้มข้นของ PM 10 ที่ สูงมาก ในขณะที่พื้นที่บริเวณสวนสาธารณะอาจจะมี PM 10 ต่ำ ทั้งนี้ ค่าความเข้มข้นที่นำมาวิเคราะห์ในการศึกษานี้ เป็นค่าความเข้มข้นของ PM 10 เฉลี่ยโดยทั่วบริเวณพื้นที่ของเขตกรุงเทพมหานคร โดยจากสถิติดังกล่าว พบว่า ค่าความเข้มข้นของ PM 10 จะเพิ่มสูงขึ้นในห้วงฤดูหนาว (Dry Season) และจะลดต่ำลงในห้วงฤดูฝน (Wet Season) เช่นเดียวกับกับ PM 2.5

เมื่อพิจารณาเปรียบเทียบความเข้มข้นของ PM 10 ระหว่างสถานการณ์ปกติและสถานการณ์การแพร่ระบาดของ Covid-19 พบว่า ในช่วงเดือน เมษายน - ธันวาคม ค่าความเข้มข้น PM 10 ใกล้เคียงกันทั้งเหตุการณ์ในขณะที่มีการแพร่ระบาดของ Covid-19 เทียบกับช่วงสถานการณ์ปกติ แต่ในช่วง มกราคม - มีนาคม กลับพบว่า เมื่อมีการแพร่ระบาดของ Covid-19 ค่าความเข้มข้น PM 10 สูงกว่าห้วงสถานการณ์ปกติ พบว่า ฤดูกาลเป็นหนึ่งในปัจจัยที่สำคัญต่อความเข้มข้น PM 10 ที่เกิดขึ้น



ภาพ 3.5 (ก) ค่าความเข้มข้นเฉลี่ยของ PM 2.5 ก่อนการระบาดของ Covid-19 และระหว่างการแพร่ระบาด Covid-19



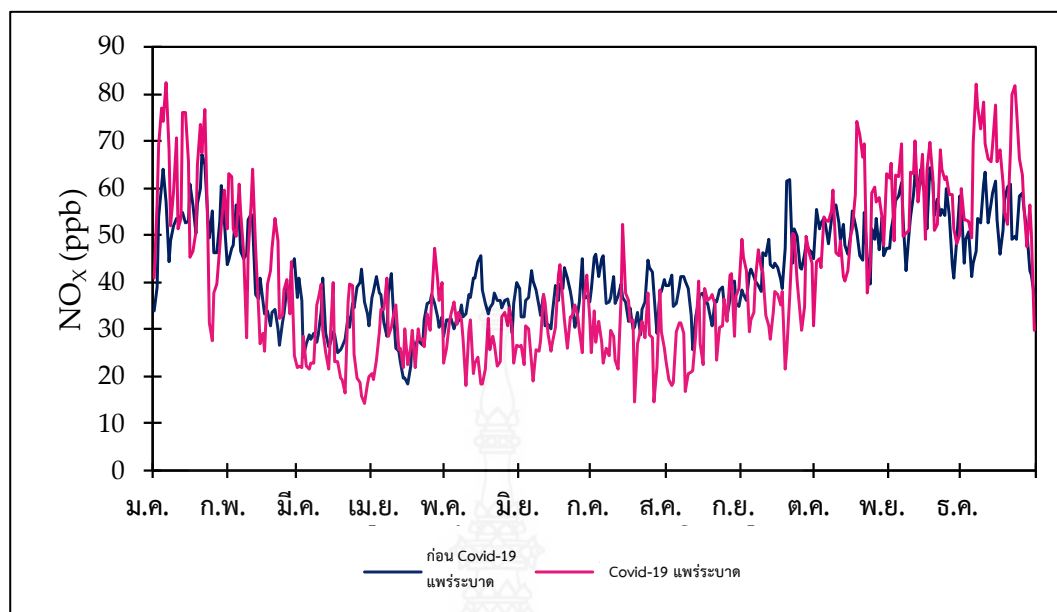
ภาพ 3.5 (ข) ค่าความเข้มข้นเฉลี่ยของ PM 10 ก่อนการระบาดของ Covid-19 และระหว่างการแพร่ระบาด Covid-19

จากภาพ 3.5 (ค) ซึ่งแสดงรายละเอียดของความเข้มข้นเฉลี่ยของ  $\text{NO}_x$  ในแต่ละเดือน ในช่วงเวลาสถานการณ์ปกติ ตั้งแต่ปี พ.ศ. 2558 - 2562 (เส้นสีน้ำเงิน) และความเข้มข้นเฉลี่ย  $\text{NO}_x$  ในแต่ละเดือน ของช่วงสถานการณ์การแพร่ระบาด Covid-19 ตั้งแต่ปี พ.ศ. 2563 - 2564 (เส้นสีแดง) พบว่า ค่า  $\text{NO}_x$  ในกรุงเทพมหานครสูงสุดเฉลี่ยต่อเดือน ประมาณ 80 - 90 ppb และค่าต่ำที่สุดประมาณ 10 ppb ทั้งนี้ อาจแตกต่างกันไปในแต่ละบริเวณ โดยทั่วไป  $\text{NO}_x$  จะเกิดขึ้นเองตามธรรมชาติแต่เกิดในปริมาณไม่มากนัก ซึ่งสาเหตุหลักที่ทำให้มี  $\text{NO}_x$  ในอากาศสูงคือ การเผาไหม้เชื้อเพลิงชนิดต่าง ๆ เช่น การเผาไหม้เชื้อเพลิงเครื่องยนต์ของรถยนต์ โรงไฟฟ้า อุตสาหกรรมต่าง ๆ การเผาไหม้เชื้อเพลิงในครัวเรือน ทั้งนี้ยังพบว่า สัดส่วนการเกิดจากรยนต์มีสูงถึงร้อยละ 55 ของการเกิดทั้งหมด ทั้งนี้ ค่าความเข้มข้นที่นำมาวิเคราะห์ในการศึกษา นี้ เป็นค่าความเข้มข้นของ  $\text{NO}_x$  เฉลี่ย โดยทั่วไปบริเวณพื้นที่ของเขตกรุงเทพมหานคร โดยจากสถิติดังกล่าว พบว่า ค่าความเข้มข้นของ  $\text{NO}_x$  จะเพิ่มสูงขึ้นในห้วงฤดูหนาว (Dry Season) และจะลดต่ำลงในห้วงฤดูฝน (Wet Season) เช่นเดียวกันกับมลพิษทางอากาศตัวอื่น

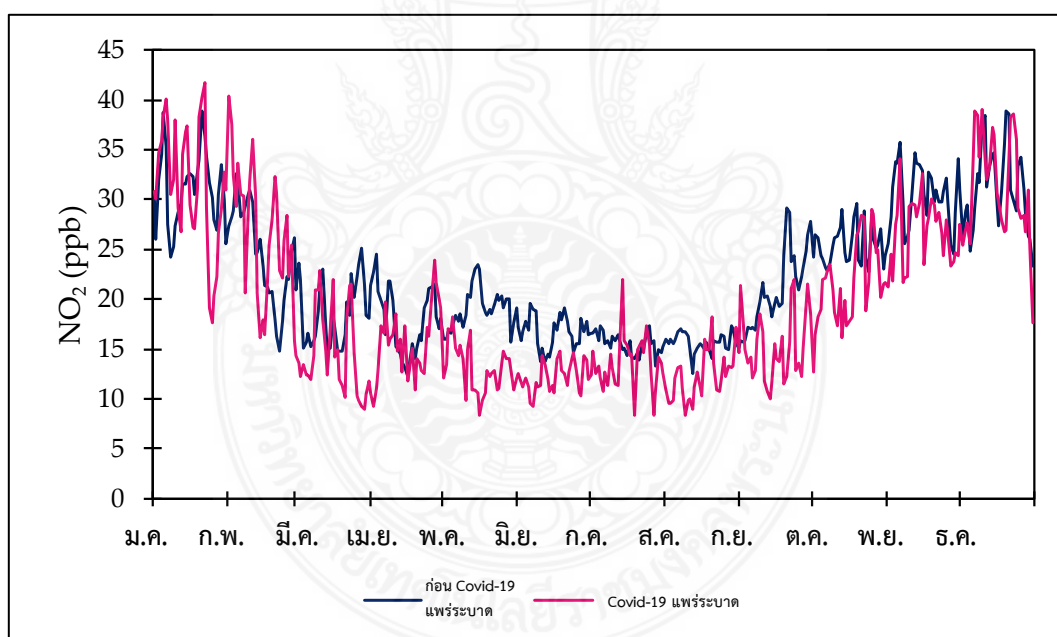
เมื่อพิจารณาเปรียบเทียบความเข้มข้นของ  $\text{NO}_x$  ระหว่างสถานการณ์ปกติและสถานการณ์การแพร่ระบาด Covid-19 พบว่า ตลอดทั้งปีค่าความเข้มข้น  $\text{NO}_x$  ใกล้เคียงกันทั้งเหตุการณ์ในขณะที่มีการแพร่ระบาด Covid-19 เทียบกับช่วงสถานการณ์ปกติ แต่ในช่วง ธันวาคม - มกราคม กลับพบว่า เมื่อมีการแพร่ระบาดของ Covid-19 ค่าความเข้มข้น  $\text{NO}_x$  สูงกว่าห้วงสถานการณ์ปกติ

จากภาพ 3.5 (ง) ซึ่งแสดงรายละเอียดของความเข้มข้นเฉลี่ยของ  $\text{NO}_2$  ในแต่ละเดือน ในช่วงเวลาสถานการณ์ปกติ ตั้งแต่ปี พ.ศ. 2558 - 2562 (เส้นสีน้ำเงิน) และความเข้มข้นเฉลี่ย  $\text{NO}_2$  ในแต่ละเดือน ของช่วงสถานการณ์การแพร่ระบาด Covid-19 ตั้งแต่ปี พ.ศ. 2563-2564 (เส้นสีแดง) ซึ่ง  $\text{NO}_2$  เป็นหนึ่งในบรรดามลพิษทางอากาศหลัก (Criteria Pollutants) โดย  $\text{NO}_2$  ยังเป็นสารตั้งต้น (Precursor) ตัวสำคัญของการเกิด PM 2.5 จากข้อมูลสถิติ พบว่า ค่า  $\text{NO}_2$  ในกรุงเทพมหานครสูงสุดเฉลี่ยต่อเดือน ประมาณ 40-45 ppb และค่าต่ำที่สุดประมาณ 10 ppb ทั้งนี้ อาจแตกต่างกันไปในแต่ละบริเวณ จากภาพ 3.5 (ง) ในช่วงเดือนเมษายน - เดือนตุลาคม จะเห็นว่าก่อนการระบาดของ Covid-19 พบค่า  $\text{NO}_2$  สูงกว่าขณะที่มีการระบาดของ Covid-19 และมีมาตรการให้อยู่บ้าน ซึ่งสอดคล้องกับข้อมูลของ PM 2.5 และ  $\text{NO}_x$  นอกจากนี้  $\text{NO}_2$  และ  $\text{NO}_x$  ในบรรยากาศยังสามารถนำไปสู่การก่อตัวของ PM 2.5 และ  $\text{O}_3$  ซึ่งถือว่าเป็นมลพิษทางอากาศหลัก ที่มีบทบาทสำคัญต่อผลกระทบด้านสาธารณสุขของประชาชนทั่วทุกมุมโลก โดยจากสถิติดังกล่าว พบว่า ค่าความเข้มข้นของ  $\text{NO}_2$  จะเพิ่มสูงขึ้นในห้วงฤดูหนาว (Dry Season) และจะลดต่ำลงในห้วงฤดูฝน (Wet Season) เช่นเดียวกันกับมลพิษทางอากาศตัวอื่น

เมื่อพิจารณาเปรียบเทียบความเข้มข้นของ  $\text{NO}_2$  ระหว่างสถานการณ์ปกติและสถานการณ์การแพร่ระบาด Covid-19 พบว่า ตลอดทั้งปีค่าความเข้มข้น  $\text{NO}_2$  ใกล้เคียงกันทั้งเหตุการณ์ในขณะที่มีการแพร่ระบาด Covid-19 เทียบกับช่วงสถานการณ์ปกติ แต่ในช่วง พฤษภาคม - ตุลาคม กลับพบว่า เมื่อมีการแพร่ระบาดของ Covid-19 ค่าความเข้มข้น  $\text{NO}_2$  ต่ำกว่าห้วงสถานการณ์ปกติ



ภาพ 3.5 (ค) ค่าความเข้มข้นเฉลี่ยของ  $\text{NO}_x$  ก่อนการระบาดของ Covid-19 และระหว่างการแพร่ระบาด Covid-19



ภาพ 3.5 (ง) ค่าความเข้มข้นเฉลี่ยของ  $\text{NO}_2$  ก่อนการระบาดของ Covid-19 และระหว่างการแพร่ระบาด Covid-19

จากภาพ 3.5 (จ) ซึ่งแสดงรายละเอียดของความเข้มข้นเฉลี่ยของ NO ในแต่ละเดือน ในช่วงเวลาสถานการณ์ปกติ ตั้งแต่ปี พ.ศ. 2558-2562 (เส้นสีน้ำเงิน) และความเข้มข้นเฉลี่ย NO ในแต่ละเดือน ของช่วงสถานการณ์การแพร่ระบาด Covid-19 ตั้งแต่ปี พ.ศ. 2563-2564 (เส้นสีแดง) ซึ่งมีต้นกำเนิดมาจาก เมื่อยานพาหนะปล่อยออกไซด์ของไนโตรเจน 90 ถึง 95 เปอร์เซ็นต์ของการปล่อย ก๊าซคือไนตริกออกไซด์ (NO) โดยจากสถิติดังกล่าว พบว่า ค่าความเข้มข้นของ NO จะเพิ่มสูงขึ้นในห้วง ฤดูหนาว (Dry Season) และจะลดต่ำลงในห้วงฤดูฝน (Wet season) เช่นเดียวกันกับมลพิษทาง อากาศตัวอื่น เมื่อพิจารณาเปรียบเทียบความเข้มข้นของ NO ระหว่างสถานการณ์ปกติและสถานการณ์ การแพร่ระบาด Covid-19 พบว่า ตลอดทั้งปีค่าความเข้มข้น NO ใกล้เคียงกันทั้งเหตุการณ์ในขณะที่มี การแพร่ระบาด Covid-19 เทียบกับช่วงสถานการณ์ปกติ แต่ในช่วง ตุลาคม - กุมภาพันธ์ (Dry Season) กลับพบว่า เมื่อมีการแพร่ระบาดของ Covid-19 ค่าความเข้มข้น NO สูงกว่าห้วงสถานการณ์ ปกติ

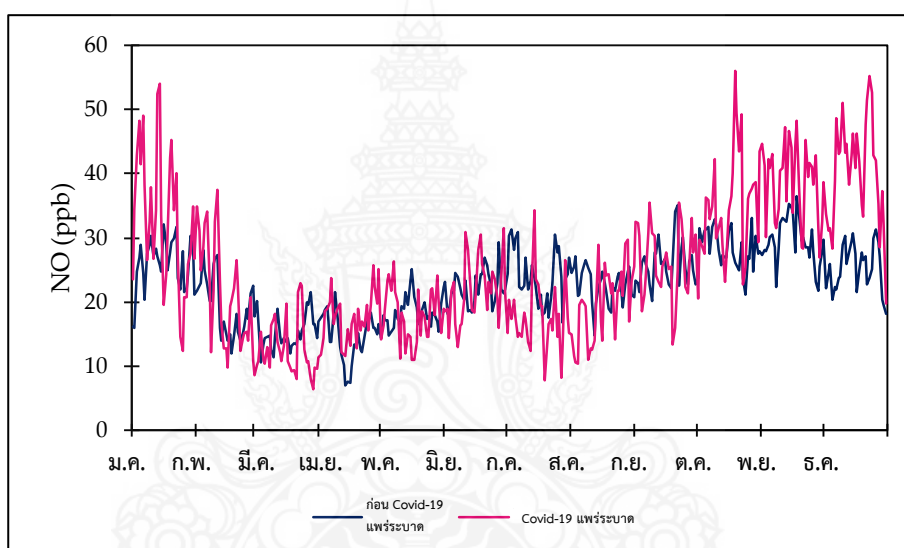
จากภาพ 3.5 (ฉ) ซึ่งแสดงรายละเอียดของความเข้มข้นเฉลี่ยของ CO ในแต่ละเดือน ในช่วงเวลาสถานการณ์ปกติ ตั้งแต่ปี พ.ศ. 2558 - 2562 (เส้นสีน้ำเงิน) และความเข้มข้นเฉลี่ย CO ในแต่ละเดือน ของช่วงสถานการณ์การแพร่ระบาด Covid-19 ตั้งแต่ปี พ.ศ. 2563 - 2564 (เส้นสีแดง) ซึ่ง CO หรือที่รู้จักกันในชื่อ Carbon Monoxide เป็นก๊าซที่ไม่มีกลิ่น ไม่มีสี เกิดจากการ เผาไหม้คาร์บอนที่ไม่สมบูรณ์ในเชื้อเพลิงต่างๆ เช่น ฟอสซิล การเผาถ่าน เต่าเผา ไฟไหม้ เครื่องยนต์ สารเคมีหรือปฏิกิริยาเคมีบางชนิด จากข้อมูลสถิติ CO ในกรุงเทพมหานคร พบว่า ค่า CO ใน กรุงเทพมหานครสูงสุดเฉลี่ยต่อเดือน ประมาณ 1.5 ppm และค่าต่ำที่สุดประมาณ 0.4 ppm ทั้งนี้ อาจแตกต่างกันไปในแต่ละบริเวณ เช่น พื้นที่บริเวณที่มีการเผาไหม้ ไม่เพียงแต่การเผาไหม้ของ เชื้อเพลิง แต่ยักรวมไปถึง พื้นที่ที่เป็นโรงงานอุตสาหกรรม บริเวณชานเมืองที่มีการเผาอันเนื่องมา จาก การเกษตร โดยบริเวณพื้นที่เหล่านี้ อาจจะมีค่าความเข้มข้นของ CO ที่สูงมาก ในขณะที่พื้นที่บริเวณ ใจกลางเมือง สวนสาธารณะอาจจะมี CO ต่ำ ทั้งนี้ ค่าความเข้มข้นที่นำมาวิเคราะห์ในการศึกษานี้ เป็น ค่าความเข้มข้นของ CO เฉลี่ยโดยทั่วบริเวณพื้นที่ของเขตกรุงเทพมหานคร โดยจากสถิติดังกล่าว พบว่า ค่าความเข้มข้นของ CO จะเพิ่มสูงขึ้นในห้วงฤดูหนาว (Dry season) และจะลดต่ำลงในห้วงฤดูฝน (Wet season) เช่นเดียวกันกับมลพิษทางอากาศตัวอื่น

เมื่อพิจารณาเปรียบเทียบความเข้มข้นของ CO ระหว่างสถานการณ์ปกติและ สถานการณ์การแพร่ระบาด Covid-19 พบว่า ตลอดทั้งปีค่าความเข้มข้น CO ใกล้เคียงกันทั้งเหตุการณ์ ในขณะที่มีการแพร่ระบาด Covid-19 เทียบกับช่วงสถานการณ์ปกติ

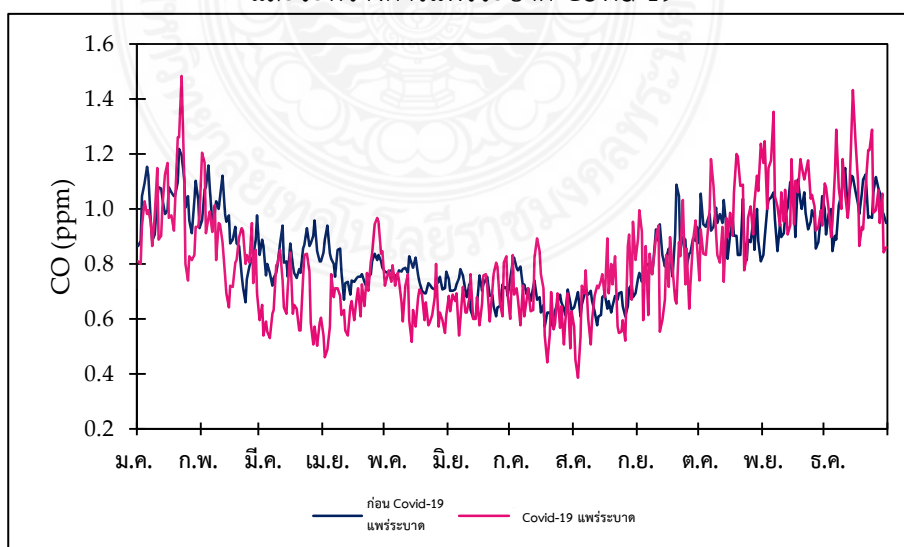
จากภาพ 3.5 (ข) ซึ่งแสดงรายละเอียดของความเข้มข้นเฉลี่ยของ O<sub>3</sub> ในแต่ละเดือน ในช่วงเวลาสถานการณ์ปกติ ตั้งแต่ปี พ.ศ. 2558 - 2562 (เส้นสีน้ำเงิน) และความเข้มข้นเฉลี่ย O<sub>3</sub> ในแต่ละเดือน ของช่วงสถานการณ์การแพร่ระบาด Covid-19 ตั้งแต่ปี พ.ศ. 2563 - 2564 (เส้นสีแดง) โดยเป็นก๊าซพิษที่อยู่ชั้นล่าง หมายถึง อยู่สูงจากพื้นดินไม่เกิน 2 กิโลเมตร โดยอาจเกิดขึ้นเองตาม ธรรมชาติโดยมีแสงสว่างเป็นตัวเร่ง หรือควันทนจากท่อไอเสียรถยนต์ เครื่องกำเนิดไฟฟ้า เครื่องกำเนิด ความร้อน จากกระบวนการผลิตในแหล่งอุตสาหกรรม เป็นต้น

โดยจากข้อมูลสถิติ O<sub>3</sub> ในกรุงเทพมหานคร พบว่า ปริมาณ O<sub>3</sub> สูงสุดเฉลี่ยต่อเดือน ประมาณ 40 ppb และค่าต่ำที่สุดประมาณ 10 ppb อย่างไรก็ตาม ปริมาณ O<sub>3</sub> ในประเทศไทยยัง

พบว่าไม่เกินค่ามาตรฐานคุณภาพอากาศที่กำหนดไว้ แต่ก็มีแนวโน้มที่จะก่อให้เกิดปัญหาสุขภาพทางอากาศในอนาคตได้ เนื่องจากปริมาณการเดินทางที่เพิ่มขึ้นอย่างรวดเร็วและต่อเนื่อง ทั้งนี้ ค่าความเข้มข้นที่นำมาวิเคราะห์ในการศึกษานี้ เป็นค่าความเข้มข้นของ  $O_3$  เฉลี่ยโดยทั่วบริเวณพื้นที่ของเขตกรุงเทพมหานคร ค่าความเข้มข้นของ  $O_3$  จะเพิ่มสูงขึ้นในห้วงฤดูร้อนมากกว่าฤดูอื่น โดยมีค่าสูงสุดในช่วงเดือนมีนาคม – พฤษภาคม ส่วนในฤดูฝนความเข้มข้นของ  $O_3$  ในบรรยากาศจะลดลงอย่างเห็นได้ชัด เนื่องจากแสงแดดซึ่งให้พลังงานต่อการเกิดก๊าซ  $O_3$  มีน้อย เมื่อพิจารณาเปรียบเทียบความเข้มข้นของ  $O_3$  ระหว่างสถานการณ์ปกติและสถานการณ์การแพร่ระบาด Covid-19 พบว่า ตลอดทั้งปีค่าความเข้มข้น  $O_3$  มีแนวโน้มใกล้เคียงกันทั้งเหตุการณ์ในขณะที่มีการแพร่ระบาด Covid-19 เทียบกับช่วงสถานการณ์ปกติ

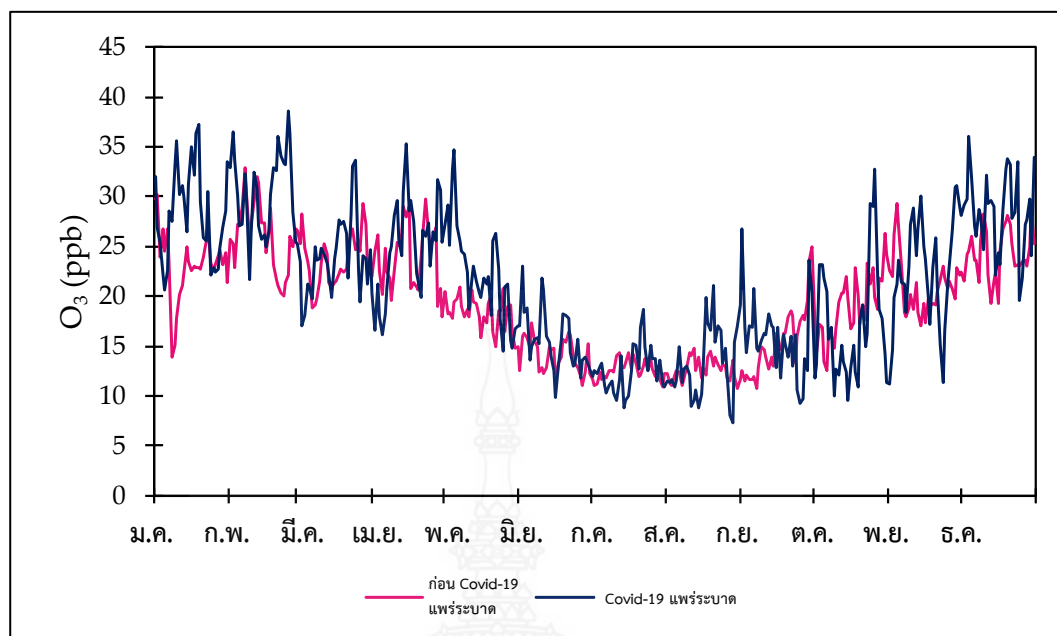


ภาพ 3.5 (จ) ค่าความเข้มข้นเฉลี่ยของ NO ก่อนการระบาดของ Covid-19 และระหว่างการแพร่ระบาด Covid-19



ภาพ 3.5 (ฉ) ค่าความเข้มข้นเฉลี่ยของ CO ก่อนการระบาดของ Covid-19 และระหว่างการแพร่ระบาด Covid-19





ภาพ 3.5 (ข) ค่าความเข้มข้นเฉลี่ยของ  $O_3$  ก่อนการระบาดของ Covid-19 และระหว่างการแพร่ระบาด Covid-19

จากงานวิเคราะห์พฤติกรรมของมลพิษทางอากาศในภาพรวมข้างต้น พบว่า ก๊าซต่างๆ มีลักษณะที่คล้ายคลึงกัน โดยจะมีค่าความเข้มข้นสูงในช่วงฤดูหนาว หรือ Dry season และจะลดลงในช่วงฤดูฝน Wet season ทั้งในสถานการณ์ปกติ และในขณะที่มีการแพร่ระบาดของ Covid-19 ซึ่งจะมีก๊าซมลพิษบางชนิดเท่านั้น ที่จะลดต่ำลงในห้วงการระบาด Covid-19 เช่น PM 2.5 เป็นต้น

## 2) One Way ANOVA Test

จากการตั้งสมมติฐานของผู้วิจัยที่ว่าปริมาณ PM 2.5 ที่เกิดขึ้นในช่วง Covid-19 จะมีความแตกต่างจากช่วงก่อนการระบาดของ Covid-19 หรือไม่ จึงทำการทดสอบด้วยวิธี One-way anova ซึ่งเป็นวิธีทางสถิติที่ใช้วิเคราะห์ความแปรปรวนทางเดียว เพื่อพิจารณาว่าค่าเฉลี่ยของปริมาณ PM 2.5 ในแต่ละช่วงเวลามีความแตกต่างกันอย่างมีนัยสำคัญหรือไม่ โดยมีการตั้งสมมติฐาน ดังนี้

$H_0$ : ค่าเฉลี่ยของปริมาณ PM 2.5 เท่ากัน สำหรับทุกกลุ่มช่วงเวลา

$H_1$ : ค่าเฉลี่ยของปริมาณ PM 2.5 ไม่เท่ากัน อย่างน้อยหนึ่งกลุ่มช่วงเวลา

โดยมีวิธีการคำนวณ ค่า F statistic ดังสมการที่ (3.6) – (3.8) และตัวอย่างการแยกข้อมูลค่า PM 2.5 แยกข้อมูลรายเดือนในแต่ละปีตั้งแต่ปี พ.ศ. 2558 ถึงปี พ.ศ. 2564 ในตาราง 3.4

$$F \text{ statistic} = \frac{MST}{MSE} = \frac{\frac{SST}{df_t}}{\frac{SSE}{df_e}} \quad (3.6)$$

โดยที่

|        |  |
|--------|--|
| $MST$  | คือ ค่า mean squared treatment             |
| $MSE$  | คือ ค่า mean squared error                 |
| $SST$  | คือ ค่า <i>sum of squares treatment</i>    |
| $SSE$  | คือ ค่า <i>sum of squares error</i>        |
| $df_t$ | คือ ค่า <i>degree of freedom treatment</i> |
| $df_e$ | คือ ค่า <i>degree of freedom error</i>     |

$$SST = \sum_{i=1}^c \sum_{j=1}^{r_i} (\bar{Y}_{i.} - \bar{Y}_{..})^2 \quad (3.7)$$

โดยที่

|                |   |
|----------------|---|
| $c$            | คือ จำนวนกลุ่มช่วงเวลา                              |
| $r_i$          | คือ จำนวนข้อมูลในกลุ่มช่วงเวลาที่เป็นที่ $i$        |
| $\bar{Y}_{i.}$ | คือ ค่าเฉลี่ยของข้อมูลในกลุ่มช่วงเวลาที่เป็นที่ $i$ |
| $\bar{Y}_{..}$ | คือ ค่าเฉลี่ยของข้อมูลทั้งหมด                       |

$$SSE = \sum_{i=1}^c \sum_{j=1}^{r_i} (Y_{ij} - \bar{Y}_{i.})^2 \quad (3.8)$$

โดยที่

|                |  |
|----------------|--|
| $c$            | คือ จำนวนกลุ่มช่วงเวลา                                       |
| $r_i$          | คือ จำนวนข้อมูลในกลุ่มช่วงเวลาที่เป็นที่ $i$                 |
| $Y_{ij}$       | คือ ค่าของข้อมูลในกลุ่มช่วงเวลาที่เป็นที่ $i$ ตำแหน่งที่ $j$ |
| $\bar{Y}_{i.}$ | คือ ค่าเฉลี่ยของข้อมูลในกลุ่มช่วงเวลาที่เป็นที่ $i$          |

ตาราง 3.4 ตัวอย่างการแยกข้อมูลค่า PM 2.5 แยกข้อมูลรายเดือนในแต่ละปีตั้งแต่ปี พ.ศ. 2558 ถึงปี พ.ศ. 2564

| ปี ค.ศ. | เดือน | PM 2.5 ( $\mu\text{g}/\text{m}^3$ ) |
|---------|-------|-------------------------------------|
| 2015    | 1     | 54.26849                            |
| 2015    | 2     | 43.15628                            |
| 2015    | 3     | 27.49932                            |
| 2015    | 4     | 26.01806                            |
| 2015    | 5     | 15.69897                            |
| ...     | ...   | ...                                 |
| 2021    | 8     | 12.95834                            |
| 2021    | 9     | 15.25815                            |
| 2021    | 10    | 19.48477                            |
| 2021    | 11    | 25.21595                            |
| 2021    | 12    | 36.82919                            |

ที่มา: วิเคราะห์โดยผู้วิจัย

จากการทดสอบด้วย One Way ANOVA ซึ่งผลการทดสอบให้ค่า F statistic มีค่าเท่ากับ 0.88 และให้ค่าทางสถิติ p-value เท่ากับ 0.51 โดยมีค่ามากกว่า 0.05 ซึ่งสามารถสรุปผลได้ว่ายอมรับสมมติฐานที่ว่าค่าเฉลี่ยของ PM 2.5 มีค่าไม่แตกต่างกันในแต่ละกลุ่มช่วงเวลาในระดับความเชื่อมั่น 95% ฉะนั้นช่วงเหตุการณ์โควิดไม่มีผลต่อปริมาณของ PM 2.5 ผู้วิจัยจึงสามารถนำข้อมูลชุดนี้ไปใช้ในการวิเคราะห์ต่อไปได้

### 3) ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร Pearson Correlation

ผู้วิจัยได้ทำการตรวจสอบความสัมพันธ์ของข้อมูลตัวแปรต่าง ๆ ในห้วงเวลาก่อนการเกิด Covid-19 และระหว่างการระบาดของ Covid-19 โดยเลือกใช้การทดสอบความสัมพันธ์แบบ Pearson Correlation (Shaun Turney, 2022) ซึ่งเป็นวิธีทางสถิติที่นำมาใช้ เพื่อหาความสัมพันธ์ของข้อมูลต่าง ๆ เพราะเป็นวิธีที่สามารถทำความเข้าใจได้ง่าย ทั้งนี้ค่า Pearson Correlation จะมีผลลัพธ์อยู่ระหว่างช่วง -1.0 ถึง +1.0 ซึ่งหากมีค่าเข้าใกล้ -1.0 หมายถึง ตัวแปรทั้งสองตัวมีความสัมพันธ์กันอย่างมากในทางตรงกันข้าม หากมีค่าใกล้ +1.0 คือ ตัวแปรทั้งสองมีความสัมพันธ์ซึ่งกันและกันมากที่สุดโดยเป็นไปในทิศทางเดียวกัน และถ้าหากมีค่าเท่ากับ 0 หมายถึง ตัวแปรทั้งสองตัวไม่มีความสัมพันธ์ใดๆ ซึ่งกันและกัน

ตาราง 3.5 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร Pearson Correlation

|                | WS<br>(m/s) | WD<br>(Deg) | Temp<br>(Deg.C) | RH<br>(%) | BP<br>(mBar) | CO<br>(ppm) | NO<br>(ppb) | NO2<br>(ppb) | NOX<br>(ppb) | O3<br>(ppb) | PM 10<br>(ug/m3) | PM 2.5<br>(ug/m3) |
|----------------|-------------|-------------|-----------------|-----------|--------------|-------------|-------------|--------------|--------------|-------------|------------------|-------------------|
| WS (m/s)       | 1.00        | 0.02        | 0.12            | -0.04     | -0.01        | -0.32       | -0.12       | -0.34        | -0.22        | 0.20        | -0.15            | -0.08             |
| WD (Deg)       | 0.02        | 1.00        | 0.04            | 0.07      | 0.04         | -0.08       | -0.02       | -0.05        | -0.02        | -0.09       | -0.11            | -0.10             |
| Temp (Deg.C)   | 0.12        | 0.04        | 1.00            | -0.60     | -0.02        | -0.08       | -0.12       | -0.10        | -0.11        | 0.39        | -0.03            | -0.08             |
| RH (%)         | -0.04       | 0.07        | -0.60           | 1.00      | 0.13         | -0.04       | -0.02       | -0.17        | -0.10        | -0.55       | -0.25            | -0.12             |
| BP (mBar)      | -0.01       | 0.04        | -0.02           | 0.13      | 1.00         | -0.12       | -0.14       | -0.10        | -0.10        | 0.13        | 0.10             | 0.08              |
| CO (ppm)       | -0.32       | -0.08       | -0.08           | -0.04     | -0.12        | 1.00        | 0.56        | 0.47         | 0.60         | -0.10       | 0.41             | 0.36              |
| NO (ppb)       | -0.12       | -0.02       | -0.12           | -0.02     | -0.14        | 0.56        | 1.00        | 0.46         | 0.93         | -0.20       | 0.41             | 0.27              |
| NO2 (ppb)      | -0.34       | -0.05       | -0.10           | -0.17     | -0.10        | 0.47        | 0.46        | 1.00         | 0.62         | -0.12       | 0.53             | 0.47              |
| NOX (ppb)      | -0.22       | -0.02       | -0.11           | -0.10     | -0.10        | 0.60        | 0.93        | 0.62         | 1.00         | -0.22       | 0.52             | 0.38              |
| O3 (ppb)       | 0.20        | -0.09       | 0.39            | -0.55     | 0.13         | -0.10       | -0.20       | -0.12        | -0.22        | 1.00        | 0.09             | 0.08              |
| PM 10 (ug/m3)  | -0.15       | -0.11       | -0.03           | -0.25     | 0.10         | 0.41        | 0.41        | 0.53         | 0.52         | 0.09        | 1.00             | 0.78              |
| PM 2.5 (ug/m3) | -0.08       | -0.10       | -0.08           | -0.12     | 0.08         | 0.36        | 0.27        | 0.47         | 0.38         | 0.08        | 0.78             | 1.00              |

ที่มา: วิเคราะห์โดยผู้วิจัย

### 3.3.2 การวิเคราะห์อนุกรมเวลา (Time Series Analysis)

1) การวิเคราะห์พฤติกรรมของ PM 2.5 ด้วย Kruskal-wallis test

ผู้วิจัยได้ทำการทดสอบ Kruskal-wallis One-Way of Variance โดยการทดสอบค่า H (Kruskal and Wallis, 1952; MacFarland and Yates, 2018) เพื่อทดสอบว่าชุดข้อมูลดังกล่าวเป็นลักษณะ seasonal data หรือไม่ โดยการใช้ Kruskal-Wallis H test เป็นวิธีการทดสอบที่นิยมใช้กันในงานด้านอุตุนิยมวิทยา และการวิเคราะห์ความสัมพันธ์ของมลภาวะต่างๆ เนื่องจาก Kruskal-Wallis-test สามารถตรวจวัดค่าต่าง ๆ ตามฤดูกาลได้อย่างเสถียรและแม่นยำ (Webel, 2021; Zhao et al., 2018) โดยสมมติฐานสำหรับการทดสอบ คือ

H0: ค่ามัธยฐานของค่ามลพิษทางอากาศแต่ละตัวมีค่าเท่ากันทุกเดือน

H1: ค่ามัธยฐานของค่ามลพิษทางอากาศแต่ละค่าจะแตกต่างกันอย่างน้อยหนึ่งคู่ของเดือน

โดยสามารถคำนวณค่า H ได้จากสมการ (3.9)

$$H = \left[ \frac{12}{n(n+1)} \sum_{j=1}^c \frac{T_j^2}{n_j} \right] - 3(n+1) \quad (3.9)$$

โดยที่

$n$  คือ ผลรวมของขนาดตัวอย่างสำหรับตัวอย่างทั้งหมด

$c$  คือ จำนวนตัวอย่าง

$j^{th}$  คือ ผลรวมของอันดับ

- $T_j$  คือ จำนวนตัวอย่างที่  $j$   
 $j^{th}$  คือ อันดับของตัวอย่าง  $j$   
 $n_j$  คือ จำนวนตัวอย่างที่  $j$

จากตาราง 3.6 พบว่า ค่ามัธยฐานของค่ามลพิษทางอากาศส่วนใหญ่แตกต่างกันอย่างน้อยหนึ่งคู่ของเดือนที่มีค่า  $p$  น้อยกว่า 0.05 ดังนั้นจึงยอมรับ  $H_1$  ซึ่งแสดงว่าค่ามลพิษเหล่านี้แตกต่างกันอย่างมีนัยสำคัญในเดือนต่างๆ ทั้งก่อนสถานการณ์ Covid-19 และขณะที่มีการระบาดของ Covid-19 ที่มีการล็อกดาวน์เกิดขึ้น

จากผลการวิเคราะห์ในตาราง 3.6 สามารถสรุปได้ว่าโดยทั่วไปค่าความเข้มข้นของมลพิษต่าง ๆ จะมีการเปลี่ยนแปลงตามฤดูกาล อย่างมีนัยสำคัญ โดยจากสถิติที่ถูกบันทึกไว้ในประเทศไทย พบว่า ในช่วงฤดูฝนค่ามลพิษจะลดต่ำลง และจะเพิ่มขึ้นในฤดูร้อนและหนาว โดยเฉพาะอย่างยิ่งค่า PM 2.5 ที่มาจากการเผาไหม้ไม่สมบูรณ์ เช่น การเผาป่า และจากการจราจร เป็นต้น โดยเฉพาะอย่างยิ่ง ฤดูหนาว ซึ่งเป็นช่วงเวลาที่มีความกดอากาศสูงในกรุงเทพมหานคร คาดว่าค่าความเข้มข้นของ PM 2.5 สูงมาจากปัญหาจราจรเป็นหลัก โดยเฉพาะอย่างยิ่งในช่วงฤดูหนาว (เดือน พฤศจิกายน - มีนาคม) ของทุกปี จะมีค่า PM 2.5 ที่สูงกว่ามาตรฐาน อย่างไรก็ตาม จะเห็นได้จากหลักฐานเชิงประจักษ์ของสารมลพิษทางอากาศตามอนุกรมเวลา (Empirical evidence of Gas pollutants by Time Series) เมื่อนำชุดข้อมูลพารามิเตอร์ต่าง ๆ ของมลพิษทางอากาศย้อนหลัง 6 ปี (พ.ศ. 2558 - 2564) มาพล็อตกราฟให้เห็นภาพรวมของมลพิษต่าง ๆ เฉลี่ยรายวันพบว่า ค่าความเข้มข้นของมลพิษทางอากาศจะเพิ่มสูงขึ้น และลดลงในแต่ละฤดูกาลที่ผันแปรอย่างประจักษ์เป็นรูปร่างที่ชัดเจนดังแสดงในภาพ 3.6 ที่สามารถบ่งชี้ได้ว่าการกระจายตัวในลักษณะที่เป็นฤดูกาลดังผลการทดสอบ Kruskal-wallis test ข้างต้น

ตาราง 3.6 ผลการวิเคราะห์พารามิเตอร์ด้านมลพิษด้วย Kruskal-wallis test (มกราคม 2558 – ธันวาคม 2564)

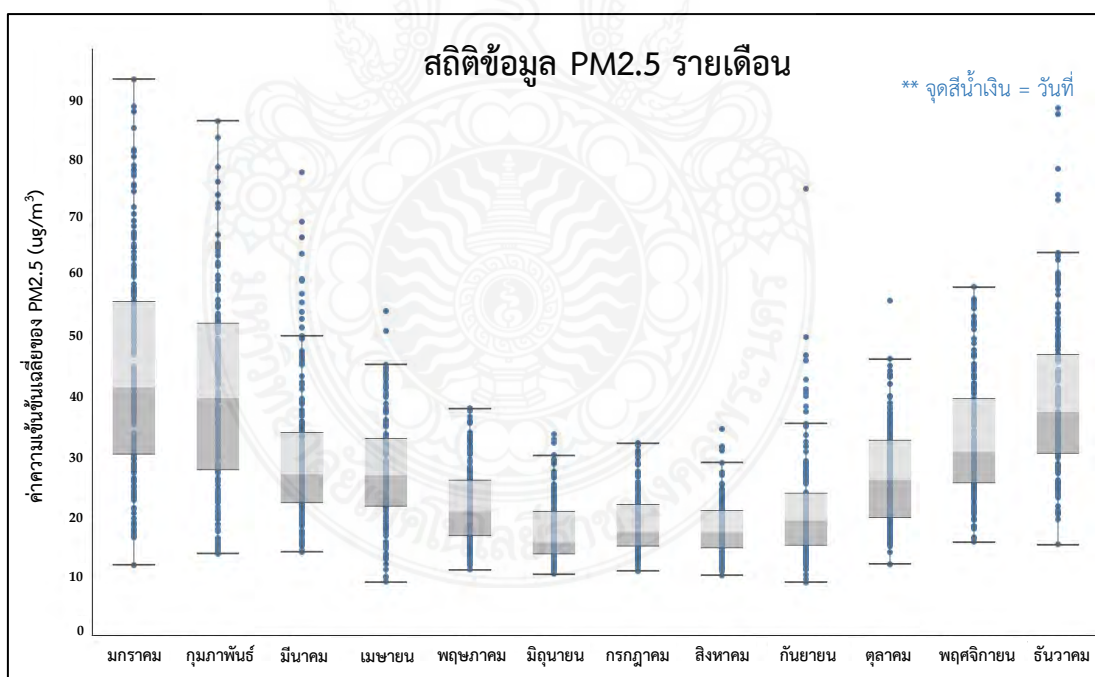
| พารามิเตอร์ด้านมลพิษ        | ผลการวิเคราะห์ |             |
|-----------------------------|----------------|-------------|
|                             | H statistics   | Asymp. Sig. |
| PM 2.5 (ug/m <sup>3</sup> ) | 48.57          | 0.000       |
| PM 10 (ug/m <sup>3</sup> )  | 49.37          | 0.000       |
| NO <sub>x</sub> (ppb)       | 28.62          | 0.000       |
| NO <sub>2</sub> (ppb)       | 28.94          | 0.000       |
| NO (ppb)                    | 44.48          | 0.000       |
| CO (ppm)                    | 36.69          | 0.000       |
| O <sub>3</sub> (ppb)        | 48.47          | 0.000       |

ที่มา: วิเคราะห์โดยผู้วิจัย

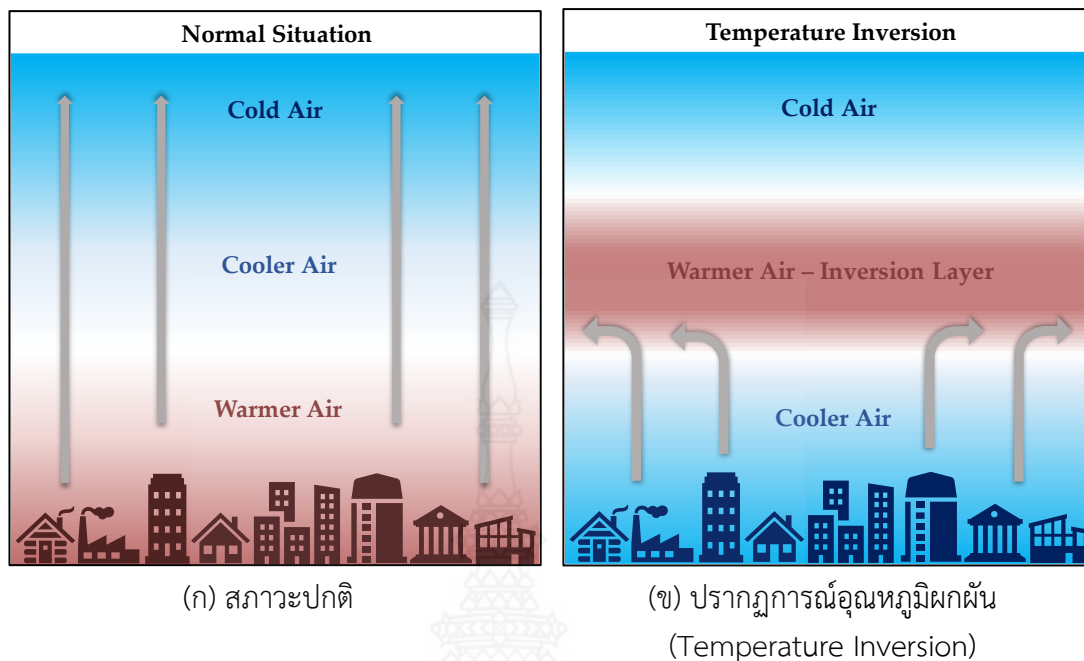


2) การวิเคราะห์ผลกระทบรายเดือนของ PM 2.5 การวิเคราะห์ผลกระทบรายเดือนของ PM 2.5

เมื่อพิจารณาผลกระทบรายเดือนของ PM 2.5 ในภาพ 3.7 ซึ่งเป็นการแสดงข้อมูล PM 2.5 ย้อนหลัง 5 ปี ตั้งแต่ปี 2558 - 2564 โดยได้นำมาแสดงแยกเป็นรายเดือน จะเห็นได้ว่าในช่วงปลายปีไปจนถึงต้นปี ตั้งแต่เดือนพฤศจิกายน - เดือนกุมภาพันธ์ ซึ่งอยู่ในช่วงฤดูหนาวของประเทศไทย ค่า PM 2.5 มีค่าสูงและมีความแปรปรวนสูงมาก เมื่อเทียบกับเดือนอื่นๆ ซึ่งสอดคล้องกับหลากหลายงานวิจัยทั้งในและต่างประเทศ ซึ่งหากอธิบายด้วยหลักการทางอุตุนิยมวิทยาเกี่ยวกับการหมุนเวียนอากาศ พบว่า โดยปกติแล้วอากาศที่อยู่บริเวณพื้นผิวดินจะมีอุณหภูมิที่สูงกว่าอากาศที่ลอยอยู่บนท้องฟ้า ซึ่งตามหลักวิทยาศาสตร์ อากาศจะเคลื่อนตัวจากที่มีอุณหภูมิสูงไปยังที่มีอุณหภูมิต่ำ ดังภาพ 3.8 (ก) คือ อากาศบริเวณพื้นดินจะไหลขึ้นสู่ชั้นบรรยากาศที่สูงขึ้นเรื่อย ๆ โดยจะพัดพานำฝุ่นละอองต่าง ๆ ลอยขึ้นไปด้วย ทำให้บริเวณพื้นดินมีฝุ่นละอองน้อยลง ส่งผลให้ PM 2.5 มีค่าต่ำและค่อนข้างคงที่ ในขณะที่ช่วงฤดูหนาว ประเทศไทยจะได้รับอิทธิพลของความกดอากาศสูง เนื่องจากมวลอากาศเย็นจากทางตอนเหนือมาปกคลุม ทำให้อุณหภูมิบริเวณพื้นดินเกิดการคายความร้อนอย่างรวดเร็วกว่าปกติ ความร้อนที่ถูกคายออกมาจึงถูกกักและแทรกอยู่ระหว่างมวลอากาศเย็นกับพื้นดินที่มีความเย็น เกิดเป็นภาวะอากาศปิด ดังภาพ 3.8 (ข) ซึ่งทำให้ฝุ่นละอองถูกกักกันไว้ในบริเวณดังกล่าวและไหลย้อนกลับสู่พื้นดิน ดังนั้น ในช่วงฤดูหนาวจึงมีปริมาณ PM 2.5 สูงกว่าในฤดูอื่นๆ



ภาพ 3.7 สถิติข้อมูล PM 2.5 รายเดือน  
ที่มา: วิเคราะห์โดยผู้วิจัย



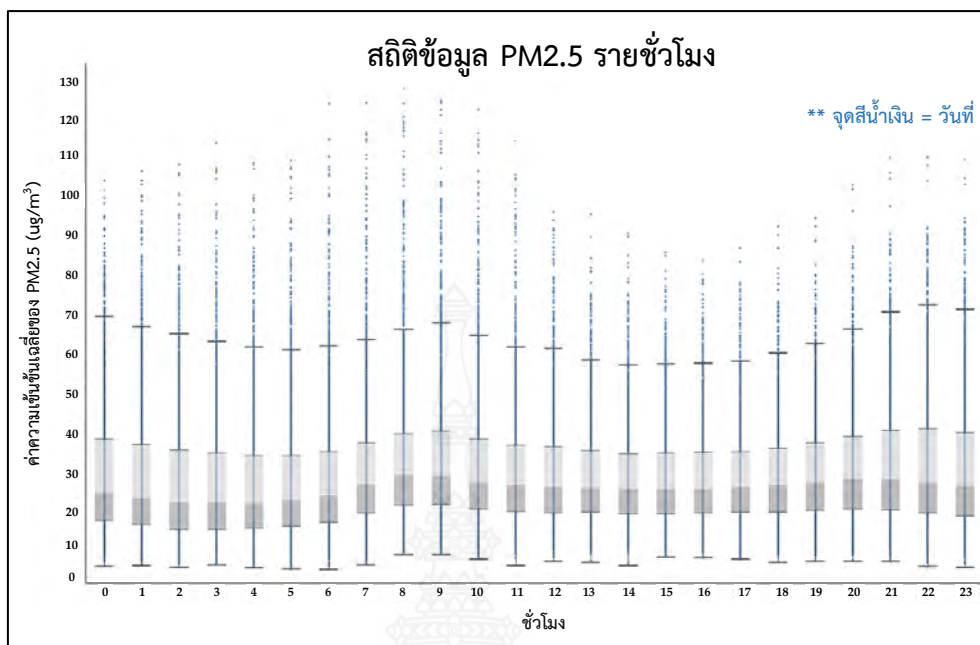
ภาพ 3.8 ผลกระทบของ PM 2.5 ต่ออุณหภูมิ  
ที่มา: วิเคราะห์โดยผู้วิจัย

### 3) การวิเคราะห์ผลกระทบรายชั่วโมงของ PM 2.5

เมื่อวิเคราะห์ผลกระทบ PM 2.5 รายเดือน จะเห็นภาพรวมของการเกิด PM 2.5 ขึ้นอยู่กับฤดูกาล ดังที่ผู้วิจัยในนำเสนอมาทั้งหมดในก่อนหน้านี้

อย่างไรก็ตาม เมื่อพิจารณาถึงรายละเอียดการเกิด PM 2.5 ในรายชั่วโมง พบว่าในแต่ละชั่วโมงความเข้มข้นของ PM 2.5 ก็จะแตกต่างกันอย่างมีนัยสำคัญ ซึ่งจากภาพ 3.9 เป็นการแสดงข้อมูล PM 2.5 ในกรุงเทพมหานครย้อนหลัง 5 ปี ตั้งแต่ปี 2015 - 2021 โดยแสดงแยกเป็นรายชั่วโมง จะเห็นได้ว่าค่า PM 2.5 ในแต่ละชั่วโมงของ 1 วัน PM 2.5 มีความเข้มข้นที่แตกต่างกัน ซึ่งความเข้มข้นดังกล่าวสอดคล้องกับปัจจัยในด้านความกดอากาศและกิจกรรมของมนุษย์ โดยจากการวิเคราะห์ข้อมูล พบว่า PM 2.5 จะมีความเข้มข้นสูงใน 2 ช่วงเวลา คือ ช่วงเช้า ระหว่าง 08:00 - 10:00 น. และ ช่วงกลางคืน 21:00 - 00:00 น. โดยในช่วงเช้าจะสอดคล้องกับกิจกรรมการเดินทางไปทำกิจกรรมในตอนเช้า (Working trip และ school trip) ส่วนในช่วงกลางคืนจะสอดคล้องกับกิจกรรมการเดินทางกลับบ้านและรถบรรทุกต่างๆ สามารถวิ่งเข้า-ออก ไซต์ก่อสร้างได้ในช่วงเวลานี้เช่นกัน ยังรวมไปถึงอากาศจะเย็นลงในกลางคืน ซึ่งอาจจะส่งผลต่อความเข้มข้นของ PM 2.5 ที่เกิดขึ้นอย่างมีนัยสำคัญ ทั้งนี้ ช่วงเวลาที่ PM 2.5 มีความเข้มข้นสูง อาจจะแตกต่างกันไปในแต่ละพื้นที่ เช่น พื้นที่เขตเมือง (CBD, urban area) ที่พักอาศัย แหล่งก่อสร้างขนาดใหญ่ใจกลางเมือง โรงงานอุตสาหกรรม หรือการทำการเกษตรและฟาร์ม เป็นต้น





ภาพ 3.9 ผลกระทบรายชั่วโมงของ PM 2.5

ที่มา: วิเคราะห์โดยผู้วิจัย

จากการทำสำรวจและวิเคราะห์พฤติกรรมของ PM 2.5 ข้างต้น จะเห็นว่า PM 2.5 มีพฤติกรรมผันแปรตามฤดูกาล ทั้งนี้ ยังพบว่าความเข้มข้นของ PM 2.5 นั้นผันแปรไปตามช่วงเวลา ตั้งแต่รายชั่วโมงไปจนถึงรายเดือนเป็น รูปแบบที่ชัดเจน

### 3.4 การจัดเตรียมชุดข้อมูล (Data Preparation)

การจัดเตรียมชุดข้อมูลในที่นี้ หมายถึงการเตรียมความพร้อมของข้อมูลก่อนที่จะนำข้อมูลเข้าสู่แบบจำลอง โดยหลังจากที่ทำการทดสอบพฤติกรรมชุดข้อมูลเรียบร้อยแล้ว ผู้วิจัยได้ทำการวิเคราะห์โครงสร้าง รูปแบบของชุดข้อมูล (Feature Engineering) การปรับรูปชุดข้อมูล (Data Transformation) และการคัดเลือก Feature ของข้อมูลก่อนที่จะทำการ แบ่งชุดข้อมูลตามสถานการณ์ที่ต้องการจะทดสอบด้วยแบบจำลองต่าง ๆ โดยมีรายละเอียดในขั้นตอนการจัดเตรียมชุดข้อมูล ดังนี้

#### 3.4.1 กระบวนการจัดการข้อมูลพื้นฐาน (Feature Engineering)

Feature Engineering เป็นกระบวนการจัดการข้อมูลพื้นฐานที่เรียกว่า data manipulation หรือ pre-processing หรือ normalization โดยการทำ Feature engineering จึงเป็นขั้นตอนแรกที่สำคัญมาก โดย Feature engineering คือ งานที่ต้องใช้ความรู้เกี่ยวกับกระบวนการ และผลลัพธ์ของมันในการแยกคุณสมบัติ (Properties) หรือคุณลักษณะ (Features) ต่าง ๆ ที่ทำให้แบบจำลองการคาดการณ์ทำงานได้ดีขึ้น

ตาราง 3.7 ตัวแปรที่มาจาก การแปลงข้อมูล (Data Transformation)

| Independent Feature |             |                 |           |              |             |             |              |              |             |                  |                       |       |      | Target            |
|---------------------|-------------|-----------------|-----------|--------------|-------------|-------------|--------------|--------------|-------------|------------------|-----------------------|-------|------|-------------------|
| WS<br>(m/s)         | WD<br>(Deg) | Temp<br>(Deg.C) | RH<br>(%) | BP<br>(mBar) | CO<br>(ppm) | NO<br>(ppb) | NO2<br>(ppb) | NOX<br>(ppb) | O3<br>(ppb) | PM 10<br>(ug/m3) | PM 2.5<br>(ug/m3)_lag | Month | Hour | PM 2.5<br>(ug/m3) |
| 0.2                 | 2           | 23.2            | 62        | 1007.7       | 0.82        | 1           | 12           | 13           | 2           | 53               | 26                    | 1     | 0    | 29                |
| 0.5                 | 29          | 20.5            | 67        | 1007.1       | 0.78        | 2           | 19           | 21           | 15          | 42               | 31                    | 1     | 5    | 23                |
| 0.5                 | 11          | 20.2            | 68        | 1007.6       | 0.79        | 1           | 24           | 26           | 6           | 36               | 23                    | 1     | 6    | 20                |
| 0.3                 | 26          | 22.7            | 64        | 1009.7       | 0.86        | 8           | 20           | 28           | 20          | 50               | 25                    | 1     | 10   | 28                |
| 0.3                 | 355         | 25.1            | 58        | 1009.2       | 0.84        | 4           | 17           | 21           | 20          | 47               | 28                    | 1     | 11   | 26                |
| 0.3                 | 13          | 26.7            | 55        | 1008.2       | 0.82        | 4           | 17           | 21           | 22          | 49               | 26                    | 1     | 12   | 27                |
| 0.3                 | 14          | 27.2            | 53        | 1007         | 0.77        | 4           | 14           | 17           | 3           | 41               | 27                    | 1     | 13   | 23                |
| 0.4                 | 16          | 27.6            | 52        | 1006.1       | 0.75        | 2           | 14           | 16           | 61          | 28               | 23                    | 1     | 14   | 15                |
| 0.3                 | 355         | 26.8            | 53        | 1005.7       | 0.75        | 2           | 14           | 16           | 35          | 40               | 15                    | 1     | 15   | 22                |
| 0.4                 | 7           | 26.9            | 53        | 1005.5       | 0.74        | 1           | 17           | 18           | 7           | 41               | 22                    | 1     | 16   | 23                |

ที่มา: วิเคราะห์โดยผู้วิจัย

โดยทั่วไปแล้ว คุณลักษณะ (Feature) จะอยู่ในรูปแบบของโครงสร้างของคอลัมน์ หรือ แอตทริบิวต์ (attributes) ซึ่งสามารถออกแบบได้โดยการแยกคุณลักษณะ (splitting Features), การรวมคุณลักษณะ (combining Features) หรือการสร้างคุณลักษณะขึ้นมาใหม่ (creating new Features) เพื่อให้ได้ผลลัพธ์ที่ดีที่สุดจากแบบจำลอง, การทำความสะอาด, คุณภาพของข้อมูล เป็นกุญแจสำคัญสู่กระบวนการ Feature engineering ที่เหมาะสมและประสิทธิภาพของแบบจำลองที่มีความแม่นยำ (Zheng et al., 2021)

นอกจากตัวแปรที่ได้มาจากข้อมูลตั้งต้น ได้ทำการเพิ่มตัวแปรอื่นอีก 2 ตัวแปร เพื่อนำมาใช้ในการทดสอบแบบจำลอง ซึ่งได้จากการคำนวณจากข้อมูลตั้งต้น คือ 1) ข้อมูลเดือน (Month) โดยทำการแปลงข้อมูลนี้จากข้อมูลวันที่ และ 2) ข้อมูล PM 2.5 ย้อนหลัง ทำการคำนวณจากข้อมูล PM 2.5 ของ 1 ปีก่อนหน้าในชั่วโมงเดียวกันนับจากวันที่จะทำนาย ดังแสดงในตาราง 3.7

### 3.4.2 การแปลงข้อมูล (Data Transformation)

ก่อนจะนำข้อมูลให้แบบจำลองเรียนรู้ จำเป็นต้องทำการแปลงข้อมูลก่อน เนื่องจากตัวแปรที่เป็นข้อมูลเชิงปริมาณแต่ละตัวจะมีค่าของช่วงข้อมูลที่แตกต่างกัน ซึ่งทำให้บางอัลกอริทึมให้ผลลัพธ์ที่ไม่สมเหตุผล รวมถึงตัวแปรที่เป็นข้อมูลเชิงคุณภาพ จำเป็นต้องทำการแปลงข้อมูลให้เป็นตัวเลข เนื่องจากอัลกอริทึมที่ใช้งานจะยอมให้เรียนรู้เฉพาะข้อมูลเชิงปริมาณเท่านั้น โดยวิธีการแปลงข้อมูล (Transform) แบ่งออกเป็น 2 ส่วน คือ

## 1) วิธี Weight of Evidence (WOE)

การแปลงข้อมูลตัวแปรอิสระที่เป็นข้อมูลเชิงคุณภาพ จะใช้วิธี Weight of Evidence (WOE) (Persson, 2021) เป็นการให้ค่าน้ำหนักของข้อมูลเพื่อให้เกิดความสัมพันธ์กันระหว่างตัวแปรอิสระ และตัวแปรตามมากที่สุด โดยมีวิธีการคำนวณ ดังสมการ (3.10)

$$WOE = \ln \left( \frac{\% \text{ target value}}{\% \text{ record}} \right) \quad (3.10)$$

โดยที่

$\% \text{ target value}$  คือ สัดส่วนของค่าตัวแปรตามในกลุ่มนั้นๆ เทียบกับค่าตัวแปรตามทั้งหมด  
 $\% \text{ record}$  คือ สัดส่วนของจำนวนข้อมูลในกลุ่มนั้นๆ เทียบกับจำนวนข้อมูลทั้งหมด

## 2) วิธี Min-Max Normalization

การแปลงข้อมูลตัวแปรอิสระที่เป็นข้อมูลเชิงปริมาณ จะใช้วิธี Min-Max Normalization (Patro and sahu, 2015) เป็นการทำให้ข้อมูลเป็นค่ามาตรฐานอยู่ในช่วง 0 ถึง 1 และนำข้อมูลมาปรับให้มีการกระจายตัวแบบปกติ (Gaussian Distribution) โดยมีวิธีการคำนวณ ดังสมการ (11)

$$Z_{norm} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (3.11)$$

โดยที่

$Z_{norm}$  คือ ข้อมูลที่ได้รับการปรับให้เป็นค่ามาตรฐานอยู่ในช่วง 0 ถึง 1  
 $x$  คือ ข้อมูลตั้งต้น  
 $\min(x)$  คือ ค่าต่ำสุดของข้อมูลชุดนั้นๆ  
 $\max(x)$  คือ ค่าสูงสุดของข้อมูลชุดนั้นๆ

ผลการแปลงข้อมูล แสดงดังตาราง 3.8 โดยผู้วิจัย จัดประเภทของข้อมูลออกเป็น 3 ส่วน ประกอบด้วย 1) ตัวแปรอิสระ (Independent Feature) 2) ข้อมูล PM 2.5 ย้อนหลัง (Historical Data) และ 3) ข้อมูลค่าเป้าหมาย (Target)

ตาราง 3.8 ตัวแปรที่มาจาก การแปลงข้อมูล (Data Transformation)

| Independent Feature |             |                 |           |              |             |             |              |              |             |                  |                   |                 |                | Target            |
|---------------------|-------------|-----------------|-----------|--------------|-------------|-------------|--------------|--------------|-------------|------------------|-------------------|-----------------|----------------|-------------------|
| WS<br>(m/s)         | WD<br>(Deg) | Temp<br>(Deg.C) | RH<br>(%) | BP<br>(mBar) | CO<br>(ppm) | NO<br>(ppb) | NO2<br>(ppb) | NOX<br>(ppb) | O3<br>(ppb) | PM 10<br>(ug/m3) | PM 2.5<br>(ug/m3) | X_Month<br>_WOE | X_Hour<br>_WOE | PM 2.5<br>(ug/m3) |
| -0.23               | 0.17        | -0.32           | 1.12      | 0.24         | 1.18        | 0.56        | -0.96        | -0.39        | -0.46       | -1.29            | -0.56             | -0.49           | 1.07           | 15                |
| -0.23               | 1.58        | -1.79           | -0.88     | 2.28         | -1.43       | 0.92        | 0.55         | 0.69         | -0.18       | 0.47             | 1.28              | 0.93            | 1.77           | 20                |
| -0.92               | -1.53       | -0.25           | -1.77     | 0.63         | -1.14       | -0.73       | 0.97         | 0.49         | 1.04        | 0.28             | -0.40             | 0.93            | -0.32          | 48                |
| -0.23               | -1.33       | -1.11           | -0.09     | 1.67         | 1.10        | 0.69        | -0.31        | 0.02         | -0.02       | 0.80             | -0.48             | 0.93            | 1.07           | 31                |
| 1.18                | 0.19        | 1.41            | -0.60     | -2.30        | -0.43       | -0.98       | 0.97         | 0.43         | 0.13        | 1.22             | -1.34             | -1.20           | -0.32          | 24                |
| 1.18                | 0.16        | 1.02            | -0.24     | -1.84        | -0.54       | -0.09       | -0.21        | -0.25        | -0.18       | 0.06             | -0.01             | -1.45           | -0.32          | 15                |
| -0.23               | -1.44       | -1.03           | -1.02     | 1.43         | 0.85        | -0.98       | -0.62        | -0.88        | 0.92        | 0.44             | -1.03             | 0.93            | 0.19           | 44                |
| 1.18                | 0.27        | -0.66           | 1.12      | 0.91         | -0.97       | -1.55       | -0.96        | -1.25        | 0.34        | 0.15             | -1.03             | -1.45           | 1.07           | 22                |
| -1.77               | -1.49       | -0.49           | 1.12      | 1.32         | 1.29        | 0.56        | -0.41        | -0.18        | 0.27        | -0.74            | 0.62              | -0.49           | -0.67          | 12                |
| 0.33                | 0.19        | -0.35           | 0.79      | -0.12        | 0.24        | 0.09        | -0.41        | -0.39        | -0.76       | -0.52            | -0.16             | 0.93            | 0.54           | 20                |

ที่มา: วิเคราะห์โดยผู้วิจัย

### 3.4.3 การคัดเลือกตัวแปร (Feature Selection)

ผู้วิจัยจะนำข้อมูลช่วงก่อนเกิดเหตุการณ์ Covid-19 ในส่วนที่จะนำมาให้โปรแกรมเรียนรู้ใช้ในการเลือกตัวแปรที่มีผลต่อค่า PM 2.5 มากที่สุด โดยวิธีที่นำมาใช้สำหรับการเลือกตัวแปร คือ Recursive Feature Elimination ซึ่งเป็นวิธีการเลือกกลุ่มตัวแปรอิสระที่มีผลต่อตัวแปรตาม ตามจำนวนตัวแปรอิสระที่กำหนดไว้ หลักการทำงาน จะเริ่มต้นจากการเรียนรู้แบบจำลองจาก algorithm ที่กำหนดไว้ โดยในงานวิจัยนี้ใช้ Random Forest Algorithm เพื่อกำจัดตัวแปรอิสระที่มีผลต่อตัวแปรตามน้อยที่สุดออกทีละตัว โดยพิจารณาจากค่าความสำคัญของตัวแปร (Feature importance) ที่มีค่าน้อยที่สุด หลังจากนั้นจะทำการเรียนรู้แบบจำลองใหม่จากตัวแปรอิสระที่เหลืออยู่วนไปจนกระทั่งเหลือจำนวนตัวแปรอิสระตามที่กำหนดไว้

โดยทำการทดสอบความแม่นยำของแบบจำลองออกเป็น 2 แบบ คือ 1) ไม่มีตัวแปรที่เกี่ยวข้องกับเวลา และข้อมูล PM 2.5 ในอดีต มีทั้งหมด 10 ตัวแปร ซึ่งสามารถจัดกลุ่มตัวแปรได้เป็น 10 กลุ่ม ตามตาราง 3.9

ในขณะที่แบบที่ 2) มีตัวแปรที่เกี่ยวข้องกับเวลา และข้อมูล PM 2.5 ในอดีต มีทั้งหมด 14 ตัวแปร ซึ่งสามารถจัดกลุ่มตัวแปรได้เป็น 14 กลุ่ม ตามตาราง 3.10

ตาราง 3.9 ตัวแปรที่ไม่เกี่ยวข้องกับเวลา และข้อมูล PM 2.5 ในอดีต (10 กลุ่มตัวแปร)

| Number of Features | Features  |
|--------------------|---|
| 1                  | ['PM 10(ug/m3)']  |
| 2                  | ['WD(Deg)', 'PM 10(ug/m3)']   |
| 3                  | ['WD(Deg)', 'CO(ppm)', 'PM 10(ug/m3)']  |
| 4                  | ['WD(Deg)', 'CO(ppm)', 'NO2(ppb)', 'PM 10(ug/m3)']  |
| 5                  | ['WD(Deg)', 'BP(mBar)', 'CO(ppm)', 'NO2(ppb)', 'PM 10(ug/m3)']  |
| 6                  | ['WD(Deg)', 'BP(mBar)', 'CO(ppm)', 'NO(ppb)', 'NO2(ppb)', 'PM 10(ug/m3)']   |
| 7                  | ['WD(Deg)', 'Temp(Deg.C)', 'BP(mBar)', 'CO(ppm)', 'NO(ppb)', 'NO2(ppb)', 'PM 10(ug/m3)']                                |
| 8                  | ['WD(Deg)', 'Temp(Deg.C)', 'BP(mBar)', 'CO(ppm)', 'NO(ppb)', 'NO2(ppb)', 'O3(ppb)', 'PM 10(ug/m3)']                     |
| 9                  | ['WD(Deg)', 'Temp(Deg.C)', 'RH(%)', 'BP(mBar)', 'CO(ppm)', 'NO(ppb)', 'NO2(ppb)', 'O3(ppb)', 'PM 10(ug/m3)']            |
| 10                 | ['WS(m/s)', 'WD(Deg)', 'Temp(Deg.C)', 'RH(%)', 'BP(mBar)', 'CO(ppm)', 'NO(ppb)', 'NO2(ppb)', 'O3(ppb)', 'PM 10(ug/m3)'] |

ที่มา: วิเคราะห์โดยผู้วิจัย

ตาราง 3.10 ตัวแปรที่เกี่ยวข้องกับเวลา และข้อมูล PM 2.5 ในอดีต (14 กลุ่มตัวแปร)

| Number of Features | Features   |
|--------------------|--|
| 1                  | ['PM 10(ug/m3)']   |
| 2                  | ['PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag']  |
| 3                  | ['WD(Deg)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag']   |
| 4                  | ['WD(Deg)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag', 'Month_WOE']  |
| 5                  | ['WD(Deg)', 'CO(ppm)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag', 'Month_WOE']   |
| 6                  | ['WD(Deg)', 'CO(ppm)', 'NO(ppb)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag', 'Month_WOE']  |
| 7                  | ['WD(Deg)', 'BP(mBar)', 'CO(ppm)', 'NO(ppb)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag', 'Month_WOE']                                    |
| 8                  | ['WD(Deg)', 'Temp(Deg.C)', 'BP(mBar)', 'CO(ppm)', 'NO(ppb)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag', 'Month_WOE']                     |
| 9                  | ['WD(Deg)', 'Temp(Deg.C)', 'BP(mBar)', 'CO(ppm)', 'NO(ppb)', 'O3(ppb)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag', 'Month_WOE']          |
| 10                 | ['WD(Deg)', 'Temp(Deg.C)', 'RH(%)', 'BP(mBar)', 'CO(ppm)', 'NO(ppb)', 'O3(ppb)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag', 'Month_WOE'] |

ที่มา: วิเคราะห์โดยผู้วิจัย

ตาราง 3.10 (ต่อ)

| Number of Features | Features  |
|--------------------|---|
| 11                 | ['WD(Deg)', 'Temp(Deg.C)', 'RH(%)', 'BP(mBar)', 'CO(ppm)', 'NO(ppb)', 'NO2(ppb)', 'O3(ppb)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag', 'Month_WOE']                                    |
| 12                 | ['WD(Deg)', 'Temp(Deg.C)', 'RH(%)', 'BP(mBar)', 'CO(ppm)', 'NO(ppb)', 'NO2(ppb)', 'O3(ppb)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag', 'Month_WOE', 'Hour_WOE']                        |
| 13                 | ['WS(m/s)', 'WD(Deg)', 'Temp(Deg.C)', 'RH(%)', 'BP(mBar)', 'CO(ppm)', 'NO(ppb)', 'NO2(ppb)', 'O3(ppb)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag', 'Month_WOE', 'Hour_WOE']             |
| 14                 | ['WS(m/s)', 'WD(Deg)', 'Temp(Deg.C)', 'RH(%)', 'BP(mBar)', 'CO(ppm)', 'NO(ppb)', 'NO2(ppb)', 'O3(ppb)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag', 'Month_WOE', 'Hour_WOE', 'NOX(ppb)'] |

ที่มา: วิเคราะห์โดยผู้วิจัย

### 3.4.4 การแบ่งข้อมูล (Data Split For Data Modeling)

การแบ่งชุดข้อมูลที่ใช้ในงานวิจัยจะแบ่งข้อมูลออกเป็น 2 ส่วน คือ ช่วงก่อนการระบาดของโควิด-19 ช่วงปี พ.ศ. 2558 - 2562 (Data Train) และ ช่วงการระบาดของโควิด-19 ช่วงปี พ.ศ. 2563 - 2564 (Data Testout)

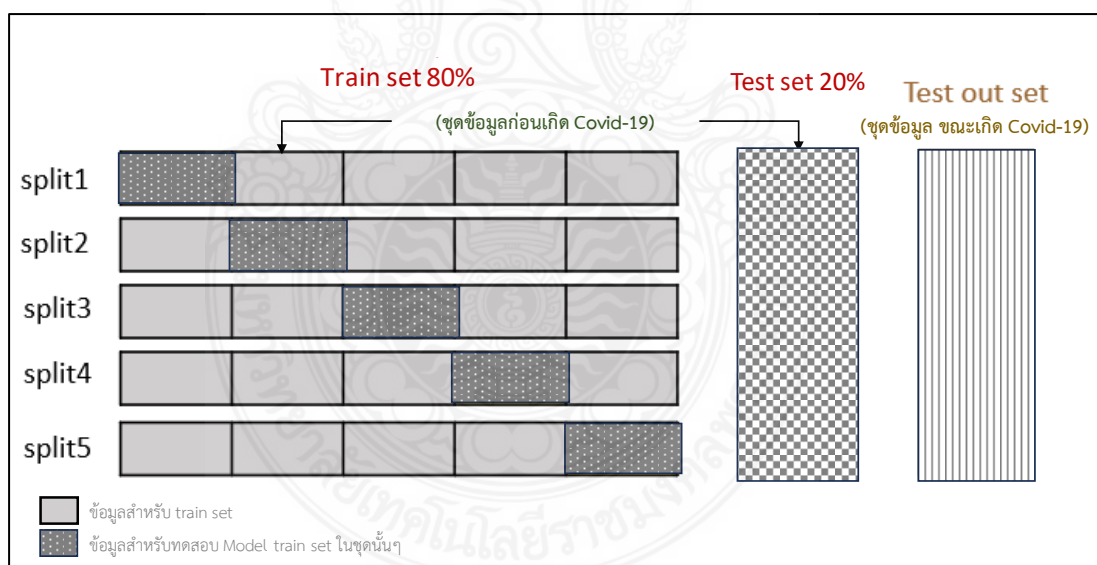
- 1) ช่วงก่อนการระบาดของโควิด-19 ช่วงระหว่าง 1 มกราคม 2558 - 24 มีนาคม 2563 (Data train)
  - แบ่งเป็นข้อมูล **Train Set** : 80% เพื่อใช้ในการเรียนรู้ข้อมูลให้ได้ค่า hyperparameter ที่เหมาะสมในแต่ละ algorithm มากที่สุด
  - แบ่งเป็นข้อมูล **Test Set** : 20% เพื่อใช้สำหรับการทดสอบประสิทธิภาพของแบบจำลองก่อนเอาไปใช้งานจริง
- 2) ช่วงช่วงการระบาดของโควิด-19 ช่วงระหว่าง 25 มีนาคม 2563 - 31 ธันวาคม 2564 (Data Testout)
  - แบ่งเป็นข้อมูล Test out set : 100% เพื่อใช้สำหรับการทดสอบประสิทธิภาพของแบบจำลองว่าสามารถนำมาประยุกต์ใช้กับข้อมูลในช่วงหลังเกิดการระบาดของโควิดได้อย่างแม่นยำหรือไม่

ในขั้นตอนการเรียนรู้ในแต่ละ algorithm ตัวโปรแกรมจะมีการเรียนรู้จากข้อมูล Train set เพื่อให้ได้ค่า Hyperparameter ที่ทำให้ได้ประสิทธิภาพของแบบจำลองที่มีประสิทธิภาพมากที่สุด โดยมีการประยุกต์ใช้วิธี Cross Validation ซึ่งมีขั้นตอนดังนี้ข้อมูลช่วงก่อนการระบาดของโควิด 19 Train set มาแบ่งออกเป็น 5 ส่วนเท่า ๆ กัน

- 1) ข้อมูลแต่ละส่วน (split) จะถูกแบ่งออกเป็น 2 ส่วน คือ 4 ส่วนสำหรับเป็น **Train set** เพื่อนำเข้าไปให้ Model ได้เรียนรู้ข้อมูลในการปรับค่าพารามิเตอร์ต่าง ๆ เพื่อให้ได้ Model ที่เรียนรู้ข้อมูลได้แม่นยำที่สุด และ 1 ส่วน สำหรับเป็น **Train validation set** เพื่อนำไปเป็นข้อมูลเพื่อทดสอบว่า Model มีความแม่นยำกับข้อมูลชุดอื่นในช่วงเวลาก่อนการระบาดโควิด 19 หรือไม่
- 2) หลังจากได้ Model จากการนำข้อมูลส่วน train set เข้าไปเรียนรู้แล้ว จะทำการวัดประสิทธิภาพของ Model ด้วยตัวชี้วัด ดังนี้ คือ R-squared adjust, MSE, RMSE และ MAE ของข้อมูลทั้งชุด train set, train validation set และชุดข้อมูลช่วงการระบาดของ Covid-19 Test Set
- 3) ทำขั้นตอนที่ 3 ซ้ำจนครบข้อมูลที่แบ่งไว้ทั้ง 5 ส่วน (split1-split5) และใช้ค่าเฉลี่ยของตัวชี้วัดที่ได้จากการข้อมูลแต่ละส่วน (Split) เป็นตัวชี้วัดในการประเมินประสิทธิภาพของ model โดยใช้ตัวชี้วัดของชุดข้อมูลทั้ง 3 ชุด (ชุด train set, train validation set และชุดข้อมูลช่วงการระบาดของโควิด 19 (Test Set) มาใช้ในการพิจารณาสำหรับการเลือก model ที่ดีที่สุด

โดยรายละเอียดแสดงดังภาพ 3.10 ซึ่งแสดงกระบวนการนำข้อมูลชุดต่าง ๆ เข้า

แบบจำลอง



ภาพ 3.10 กระบวนการนำข้อมูลชุดต่าง ๆ เข้าแบบจำลอง

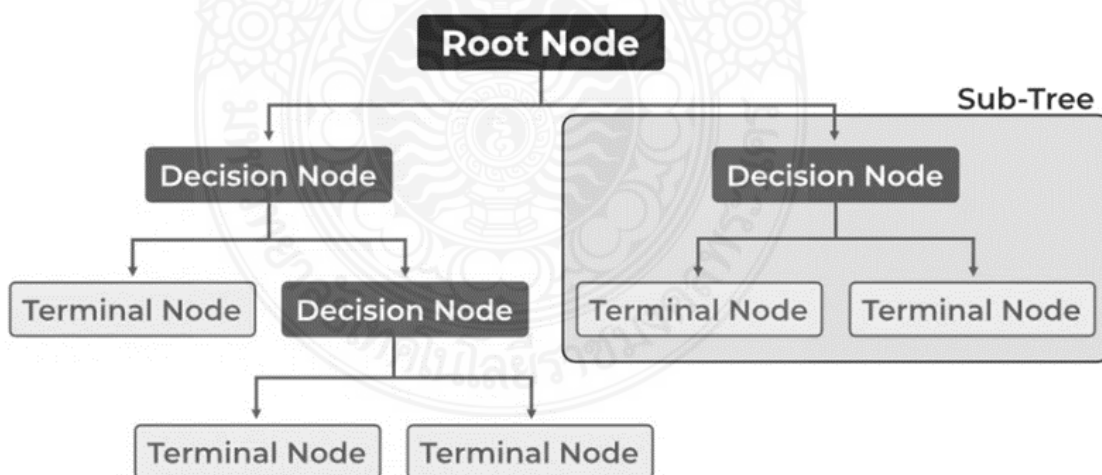
ที่มา: วิเคราะห์โดยผู้วิจัย

### 3.5 แบบจำลอง (Modelling)

สำหรับแบบจำลองที่ใช้ในการศึกษาวิจัยครั้งนี้ ผู้วิจัยเลือกใช้แบบจำลอง Machine Learning (Ethem, 2010; Biggio and Roli, 2018; Friedman, 2001; Goodfellow et al., 2016; Molnar, 2020; Stuart and Peter, 2010; Winston, 1992; Zhang et al., 2021) ชนิด Supervised Learning ทั้งหมด 6 แบบจำลอง ประกอบด้วย แบบจำลอง 1) Decision Tree Model 2) Gradient Boosting Model 3) K-neighbor Model 4) MLP Regressor Model 5) Random Forest Model และ 6) Ridge Model โดยมีรายละเอียดดังนี้

#### 3.5.1 Decision Tree Model

เป็นวิธีการทำนายได้ทั้งตัวแปรตามเชิงปริมาณ และตัวแปรตามเชิงคุณภาพ โดยภาพ 3.11 เป็นแบบจำลองที่มีการทำงาน โดยการแบ่งชุดข้อมูลทั้งหมดเป็นโครงสร้างต้นไม้ตามเงื่อนไข ซึ่งประกอบด้วย Root node เป็นโหนดชั้นบนสุดของต้นไม้ Decision node เป็นโหนดตัวเลือกที่เกี่ยวข้องกับคุณลักษณะของ input ซึ่งสามารถแยกออกเป็นโหนดย่อย ๆ เพิ่มเติม และ Terminal node คือโหนดสุดท้ายของต้นไม้ที่จะไม่มีโหนดย่อย ๆ ต่อท้าย ซึ่งแต่ละโหนด คือ เงื่อนไขจากตัวแปรอิสระ เพื่อใช้ในการแบ่งข้อมูลชุดเรียนรู้ (training data) ให้มีลักษณะของตัวแปรตามที่แตกต่างกันมากที่สุด ฉะนั้นในแต่ละโหนด แบบจำลองจะเลือกเงื่อนไขจากตัวแปรอิสระที่ให้ค่าความไม่บริสุทธิ์ของกลุ่มตัวแปรตาม (Impurity) ต่ำที่สุด โดยจะพิจารณาจากค่า entropy หรือ information gain สำหรับการทำนายตัวแปรตามเชิงคุณภาพ อย่างไรก็ตามจะพิจารณาค่า mean squared error (MSE) สำหรับการทำนายตัวแปรตามเชิงปริมาณ โดยมีรายละเอียด ดังนี้



ภาพ 3.11 ภาพแสดงโครงสร้างของ Decision tree



## 1) Entropy

Entropy เป็นตัววัดความไม่บริสุทธิ์ของกลุ่มตัวแปรตาม (Impurity) โดยถ้ายิ่ง entropy มีค่าน้อย แสดงว่าเงื่อนไขจากตัวแปรอิสระที่เลือกมาใช้ในการแบ่งข้อมูลได้ค่อนข้างดีใน โหนดนั้นๆ ซึ่งสามารถคำนวณ entropy ได้ดังสมการ (3.12)

$$E(S) = \sum_{i=1}^n -p_i \log_i p_i \quad (3.12)$$

โดยที่

|        |  |
|--------|--|
| $E(S)$ | คือ ค่า <i>entropy</i>   |
| $n$    | คือ จำนวนกลุ่มข้อมูลของตัวแปรตาม   |
| $p_i$  | คือ ค่าต่ำสุดของข้อมูลชุดนั้นๆ คือ สัดส่วนของกลุ่มข้อมูลที่ $i$ ของตัวแปรตาม |

## 2) Information gain (IG)

Information gain (IG) เป็นตัววัดความลดลงใน entropy ซึ่งใช้ในการพิจารณาว่าจะแยกคุณลักษณะใดในแต่ละลำดับขั้นในการสร้างแผนผังต้นไม้ โดยถ้ายิ่ง Information gain มีค่ามาก แสดงว่าเงื่อนไขจากตัวแปรอิสระที่เลือกมาใช้ในการแบ่งข้อมูลได้ค่อนข้างดีในโหนดนั้นๆ ดังสมการ 3.13

$$\text{Information Gain} = \text{Entropy}_{\text{parent}} - \text{Entropy}_{\text{children}} \quad (3.13)$$

โดยที่

|                                    |  |
|------------------------------------|--|
| $\text{Entropy}_{\text{parent}}$   | คือ ค่า <i>entropy</i> ของโหนดนั้นๆ  |
| $\text{Entropy}_{\text{children}}$ | คือ ค่าเฉลี่ย <i>entropy</i> ของแต่ละกลุ่มตัวแปรตามในโหนดนั้นๆ ซึ่งมีสูตรคำนวณ ดังสมการ (3.14) |

$$\text{Entropy}_{\text{children}} = \sum_{i=1}^n \frac{\text{Number of subnode } i}{\text{Total number of parent node}} \times \text{Entropy of subnode } i \quad (3.14)$$

## 3) mean squared error (MSE)

mean squared error (MSE) เป็นตัววัดความผิดพลาดของการทำนายจากแบบจำลอง ซึ่งมีสูตรคำนวณ ดังสมการ (3.15)

$$MSE = \frac{1}{m} \sum_{j=1}^m (\hat{x}_i - x_i)^2 \quad (3.15)$$

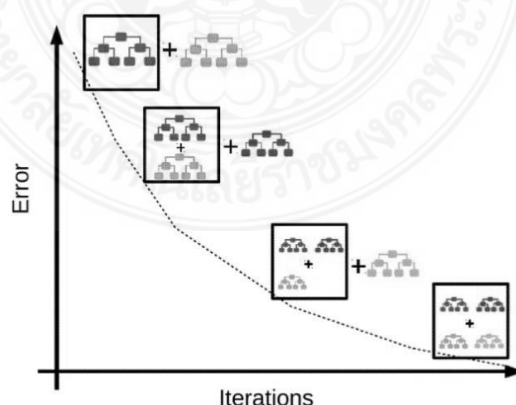
โดยที่

|             |                            |
|-------------|----------------------------|
| $M$         | คือ จำนวนข้อมูล            |
| $\hat{x}_i$ | คือ ค่าที่ได้มาจากการคำนวณ |
| $x_i$       | คือ ค่าที่ได้มาจากการสำรวจ |

อย่างไรก็ตามสามารถปรับค่าจากส่วนประกอบต้นไม้ (Hyperparameter) เพื่อให้ได้แบบจำลองให้มีประสิทธิภาพมากที่สุด คือ 1. ความลึกของต้นไม้ หากยิ่งลึกแบบจำลองก็จะมี ความซับซ้อนเพิ่มขึ้น ซึ่งสามารถส่งผลให้เกิด Overfitting ได้ 2. จำนวนข้อมูลขั้นต่ำสำหรับการแบ่ง decision node ขึ้นต่อไป 3. จำนวนข้อมูลขั้นต่ำสำหรับ leaf node 4. จำนวน leaf node ที่ต้องการให้สามารถเกิดขึ้นได้มากที่สุด

### 3.5.2 Gradient Boosting Model

เป็นวิธีการทำนายได้ทั้งตัวแปรตามเชิงปริมาณ และตัวแปรตามเชิงคุณภาพ มีหลักการ ทำงานเช่นเดียวกับ Random Forest ซึ่งเป็นการเรียนรู้จากต้นไม้หลายต้นอย่างเป็นอิสระต่อกัน ในขณะที่ Gradient Boosting (Friedman, 2001) จะเริ่มต้นเรียนรู้จากต้นไม้ที่ละต้น โดยต้นไม้ต้นแรกจะเรียนรู้จากข้อมูลชุดเรียนรู้ที่ได้จากการสุ่ม และจะทำนายค่าของตัวแปรตามด้วยค่าเฉลี่ยของข้อมูล แล้วจึงมีการคำนวณค่าความผิดพลาดของแบบจำลอง ดังแสดงโครงสร้างของแบบจำลอง Gradient Boosting ในภาพ 3.12 โดยจะให้ค่าน้ำหนักสำหรับข้อมูลที่ทำนายผิด (Weak Learners) ในการคำนวณค่าความผิดพลาดของแบบจำลอง เพื่อเป็นการเพิ่มโอกาสให้ต้นไม้ต้นถัดไปได้เรียนรู้ข้อมูลที่ทายผิดเหล่านี้ กระบวนการนี้จะทำวนไปจนกว่าค่าความผิดพลาดของแบบจำลองไม่เปลี่ยนแปลง



ภาพ 3.12 แสดงโครงสร้างของ Gradient boosting

สามารถกำหนดค่า learning rate เพื่อให้แบบจำลองได้เรียนรู้การสร้างเงื่อนไขที่ให้ค่าความผิดพลาดของแบบจำลองน้อยที่สุด หรือจนถึงเงื่อนไขที่กำหนดไว้ โดยมีสูตรการคำนวณความผิดพลาดของแบบจำลอง ดังสมการ (3.16)

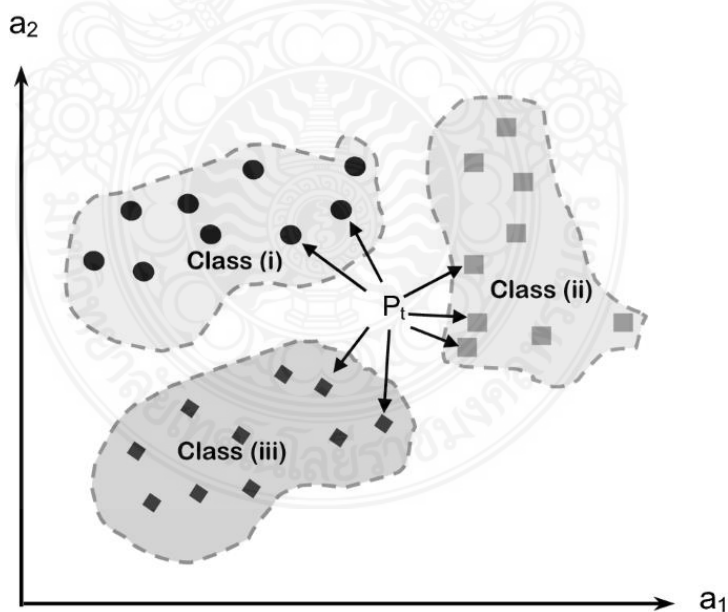
$$Residue_i = \bar{y} + \gamma * \hat{e}_i \quad (3.16)$$

โดยที่

- $\bar{y}$  คือ ค่าเฉลี่ยของตัวแปรตามจากชุดข้อมูลเรียนรู้จากต้นไม้มันแรก
- $\gamma$  คือ ค่า learning rate
- $\hat{e}_i$  คือ ค่าความผิดพลาดระหว่างค่าจริง และค่าทำนายของต้นไม้มันที่  $i$

### 3.5.3 K-nearest neighbor

เป็นวิธีการทำนายตัวแปรเชิงปริมาณ โดยใช้หลักการเลือกกลุ่มข้อมูลชุดเรียนรู้ (training data) ที่มีลักษณะใกล้เคียงกับข้อมูลชุดทดสอบ (test data) ที่จะทำนายมากที่สุดมาหาค่าเฉลี่ยของข้อมูลชุดเรียนรู้ ดังแสดงโครงสร้างของ K-nearest neighbor ในภาพ 3.13 โดยสามารถกำหนดจำนวนข้อมูลชุดเรียนรู้ เพื่อเป็นแบบจำลองของข้อมูลชุดทดสอบ (test data) โดยวิธีการวัดความคล้ายคลึงของข้อมูล จะทำการคำนวณระยะห่างระหว่างข้อมูลในแต่ละตัวแปรอิสระ ซึ่งมีวิธีทั่วไปที่ใช้กัน ดังนี้



ภาพ 3.13 แสดงโครงสร้างของ K-nearest neighbor

## 1) Euclidean Distance

Euclidean Distance เป็นวิธีการคำนวณสำหรับตัวแปรเชิงปริมาณด้วยการถอดรากที่สองของผลรวมของผลต่างยกกำลังสองระหว่างข้อมูล 2 ชุดใดๆของทุกตัวแปรอิสระ โดยมีสูตรการคำนวณ ดังสมการ (3.17)

$$\text{Euclidean Distance} = \sqrt{\sum_{i=1}^k (x_i - y_i)^2} \quad (3.17)$$

โดยที่

$k$  คือ จำนวนตัวแปรอิสระ  
 $x_i, y_i$  คือ ข้อมูลตัวแปรอิสระที่  $i$

## 2) Manhattan Distance

Manhattan Distance เป็นวิธีการคำนวณสำหรับตัวแปรเชิงปริมาณด้วยผลรวมของผลต่างค่าสัมบูรณ์ระหว่างข้อมูล 2 ชุดใดๆของทุกตัวแปรอิสระ โดยมีสูตรการคำนวณ ดังสมการ (3.18)

$$\text{Manhattan Distance} = \sum_{i=1}^k |x_i - y_i| \quad (3.18)$$

โดยที่

$k$  คือ จำนวนตัวแปรอิสระ  
 $x_i, y_i$  คือ ข้อมูลตัวแปรอิสระที่  $i$

## 3) Hamming Distance

Hamming Distance เป็นวิธีการคำนวณสำหรับตัวแปรเชิงคุณภาพ ถ้าข้อมูลมีค่าแตกต่างกันจะให้ค่าระยะห่างระหว่างข้อมูล 2 ชุดใดๆมีค่าเท่ากับ 1 ในขณะที่ถ้าข้อมูลมีค่าเหมือนกันจะให้ค่าระยะห่างมีค่าเท่ากับ 0 ดังสมการ (3.19)

$$\text{Hamming Distance} = \sum_{i=1}^k f(x_i, y_i) \quad (3.19)$$

โดยที่

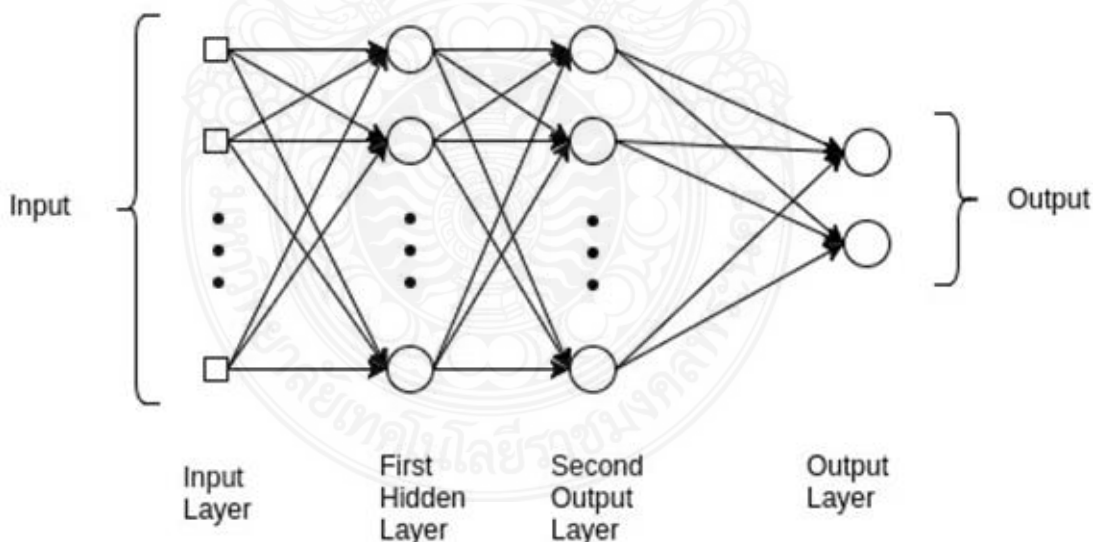
$k$  คือ จำนวนตัวแปรอิสระ

$x_i, y_i$  คือ ข้อมูลตัวแปรอิสระที่  $i$

$$f(x_i, y_i) = \begin{cases} 1; & x_i = y_i \\ 0; & x_i \neq y_i \end{cases}$$

### 3.5.4 MLP Regressor Model

เป็นวิธีการทำนายได้ทั้งตัวแปรตามเชิงปริมาณ และตัวแปรตามเชิงคุณภาพ ซึ่งเป็นโครงข่ายประสาทเทียม ประเภทหนึ่ง (Neural network) ที่มีโครงสร้างแบบหลายชั้น ซึ่งประกอบด้วย ชั้นขาเข้า (input layer) ชั้นประมวลผลที่ซ่อนอยู่ (hidden layer) อย่างน้อย 1 ชั้น และชั้นผลลัพธ์ (output layer) โดย input layer จะเป็นชั้นแรกสำหรับนำข้อมูลเข้า ซึ่งจำนวนโหนด (node) จะขึ้นอยู่กับจำนวนตัวแปรอิสระที่นำเข้ามาเรียนรู้ ในขณะที่ hidden layer เป็นชั้นถัดมาที่เชื่อมกับ input layer โดยโหนดทุกตัวของ input layer จะเชื่อมกับโหนดของ hidden layer ทุกตัว โดยแสดงโครงสร้าง MLP ดังภาพ 3.14



ภาพ 3.14 โครงสร้างของ MLP Regressor Model

โดยสามารถกำหนดจำนวนชั้น และจำนวนโหนดของแต่ละชั้นได้ หลังจากที Hidden Layer รับข้อมูลมาจาก Input Layer แล้ว จะทำการประมวลผลด้วยฟังก์ชันเส้นตรง (Linear Function) แล้วจึงนำผลลัพธ์นี้ไปคำนวณใน Activation function ก่อนที่จะส่งผลลัพธ์ต่อยังชั้นถัดไป โดยจะวนชั้นตอนเหล่านี้จนครบจำนวน hidden layer ที่กำหนดไว้ รวมถึง Output Layer ดังสมการต่าง ๆ ดังนี้

1) Linear function

$$Z^{[l]} = W^{[l]}A^{[l-1]} + b^{[l]} \quad (3.20)$$

โดยที่

|             |                               |
|-------------|-------------------------------|
| $Z^{[l]}$   | คือ ผลลัพธ์ของชั้นที่ $l$     |
| $W^{[l]}$   | คือ ค่าน้ำหนักของชั้นที่ $l$  |
| $A^{[l-1]}$ | คือ ผลลัพธ์ของชั้นที่ $l - 1$ |
| $b^{[l]}$   | คือ ค่าคงที่ของชั้นที่ $l$    |

2) Activation function

$$A^{[l]} = f(Z^{[l]}) \quad (3.21)$$

โดยที่

|             |   |
|-------------|---|
| $A^{[l]}$   | คือ ผลลัพธ์ของชั้นที่ $l$                                 |
| $Z^{[l]}$   | คือ ค่าน้ำหนักของชั้นที่ $l$                              |
| $A^{[l-1]}$ | คือ ผลลัพธ์ที่ได้จากการคำนวณฟังก์ชันเส้นตรงของชั้นที่ $l$ |

โดยสามารถเลือก Activation function ได้ดังต่อไปนี้

(1) *Rectified Linear Units (ReLU)* เป็นฟังก์ชันที่เป็นที่นิยมมากที่สุด โดยจะให้ค่ามากกว่าหรือเท่ากับ 0 ขึ้นไป

$$\phi(z) = \max(0, z) \quad (3.22)$$

โดยที่

|           |   |
|-----------|---|
| $\phi(z)$ | คือ คือ ผลลัพธ์ที่ได้จาก <i>ReLU function</i> |
| $z$       | คือ ค่าของข้อมูล                              |

(2) Sigmoid function เป็นฟังก์ชันที่เหมาะสมสำหรับแบบจำลองที่ต้องการทำนายความน่าจะเป็นที่จะเกิดขึ้นของตัวแปรตาม โดยจะให้ค่าอยู่ตั้งแต่ 0 ถึง 1 ดังสูตรด้านล่าง

$$\phi(z) = \frac{1}{1 + e^{-z}} \quad (3.23)$$

โดยที่

$\phi(z)$  คือ คือ ผลลัพธ์ที่ได้จาก sigmoid function

$z$  คือ ค่าของข้อมูล

(3) Tanh function เป็นฟังก์ชันที่มีลักษณะคล้ายกับ sigmoid function แต่จะให้ค่าอยู่ตั้งแต่ -1 ถึง 1

$$\phi(z) = \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \quad (2.24)$$

โดยที่

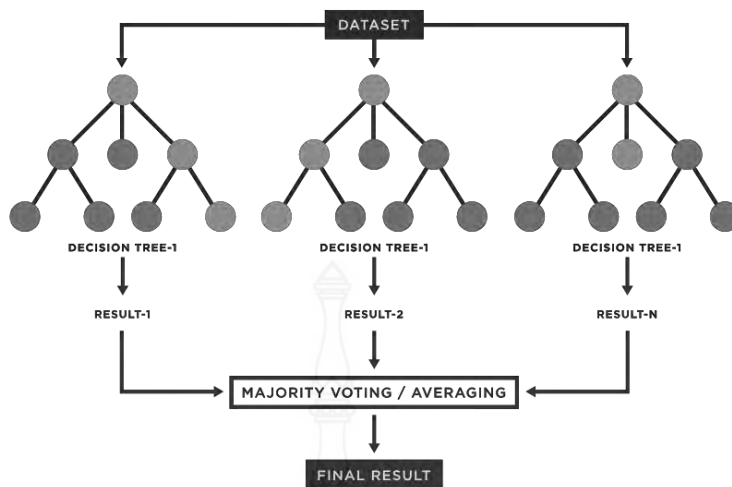
$\phi(z)$  คือ คือ ผลลัพธ์ที่ได้จาก tanh function

$z$  คือ ค่าของข้อมูล

จากกระบวนการทำงานทั้งหมดที่กล่าวมานี้ เรียกกระบวนการนี้ว่า Forward Propagation ยิ่งไปกว่านั้นอัลกอริทึมนี้จะใช้กระบวนการที่เรียกว่า Backward Propagation สำหรับการปรับพารามิเตอร์ หรือค่าน้ำหนักในแต่ละโหนดต่างๆ เพื่อลดค่าความผิดพลาดของแบบจำลอง โดยใช้วิธี Gradient Descent

### 3.5.5 Random forest

เป็นวิธีการทำนายได้ทั้งตัวแปรตามเชิงปริมาณ และตัวแปรตามเชิงคุณภาพ มีหลักการทำงานเช่นเดียวกับ Decision Tree ซึ่งมีการเรียนรู้จากต้นไม้เพียงต้นเดียวเท่านั้น ในขณะที่ Random Forest จะมีการเรียนรู้แบบ Ensemble ซึ่งจะเรียนรู้จากต้นไม้หลายต้นพร้อมกัน เพื่อลดความแปรปรวน และเพิ่มประสิทธิภาพที่สูงขึ้นกว่าการเรียนรู้จากต้นไม้เพียงต้นเดียว โดยผลลัพธ์สุดท้ายที่ได้จากการทำนายมาจากการคำนวณค่าเฉลี่ยจากต้นไม้ทั้งหมดที่เรียนรู้ สำหรับการทำนายตัวแปรตามเชิงปริมาณ หรือมาจากการเลือกค่าตัวแปรตามที่มีมากที่สุดจากต้นไม้ทั้งหมดที่เรียนรู้ สำหรับการทำนายตัวแปรตามเชิงคุณภาพ โดยสามารถกำหนดจำนวนต้นไม้ และค่าอื่นๆ เช่นเดียวกับ Decision Tree เพื่อให้ได้ผลลัพธ์ของแบบจำลองที่ดีที่สุด ดังแสดงโครงสร้างของ Random Forest Model ในภาพ 3.15



ภาพ 3.15 แสดงโครงสร้างของ Random forest Model

### 3.5.6 Ridge Model

เป็นวิธีการทำนายตัวแปรเชิงปริมาณ ซึ่งมีหลักการทำงานเหมือนกับ Linear Regression โดยมีหลักการ คือ การหาค่าพารามิเตอร์ต่างๆ เพื่อให้ได้ค่าที่คำนวณได้จาก Cost Function มีค่าน้อยที่สุด ซึ่ง Cost Function เป็นตัววัดในการประเมินความผิดพลาดของแบบจำลอง ระหว่างค่าจริงกับค่าที่ทำนายได้จากแบบจำลอง ยิ่งไปกว่านั้น Ridge จะมีการประยุกต์ใช้เทคนิค Regularization โดยการเพิ่ม Regularized Term ในการคำนวณ Cost Function เพื่อช่วยลดค่าสัมประสิทธิ์ของตัวแปรอิสระ ซึ่งช่วยลดความซับซ้อนของแบบจำลอง และป้องกันการเกิด Overfitting โดยการเพิ่ม L2 Regularization ให้กับสมการเส้นตรง (Linear Equation) สำหรับการคำนวณ Cost Function ดังสมการ (3.25)

$$cost\ function = \sum_{i=1}^n (y_i - \sum_{j=0}^m w_j x_{ij})^2 + \lambda \sum_{j=0}^m w_j^2 \tag{3.25}$$

โดยที่

- $n$  คือ จำนวนข้อมูล
- $m$  คือ จำนวนตัวแปรอิสระ
- $y_i$  คือ ข้อมูลตัวแปรตามลำดับที่  $i$
- $x_{ij}$  คือ ข้อมูลตัวแปรอิสระที่  $j$  ลำดับที่  $i$
- $w_j$  คือ ค่าน้ำหนักของตัวแปรอิสระที่  $j$  (ค่าสัมประสิทธิ์)
- $\lambda$  คือ ค่าคงที่แลมดา



โดยการปรับค่า  $\lambda$  จะมีผลต่อค่าสัมประสิทธิ์ของตัวแปรอิสระแต่ละตัว ถ้ากำหนดให้มีค่าเป็น 0 หมายถึง ค่าสัมประสิทธิ์ของแบบจำลองจะเป็นไปตามแบบจำลอง Linear Regression ซึ่งสามารถนำไปสู่การเกิด Overfitting ได้ แต่ถ้ากำหนดค่า  $\lambda$  สูงเกินไป จะส่งผลให้ค่าสัมประสิทธิ์ของแบบจำลองมีค่าน้อยเกินไป ซึ่งสามารถนำไปสู่การเกิด Underfitting ด้วยเหตุนี้จึงจำเป็นต้องเลือกค่า  $\lambda$  ที่เหมาะสม เพื่อให้แน่ใจว่าความซับซ้อนของแบบจำลองจะลดลงอย่างมีประสิทธิภาพ ซึ่งทำได้โดยการปรับแต่งไฮเปอร์พารามิเตอร์ (Hyperparameter Tuning)

### 3.6 การวัดค่าประสิทธิภาพ (Performance) ของแบบจำลอง (Model Evaluation)

โดยทั่วไปแล้ว Machine Learning (ML) ประเภท Regression models จะให้ข้อมูลผลลัพธ์ที่ต่อเนื่อง (continuous) สามารถเขียนสมการคำนวณค่า error ได้แบบนี้  $\text{error} = \text{prediction} - \text{actual}$  ซึ่ง ML เรียก error function นี้ว่า “Loss function” และเป้าหมายของการ train แบบจำลอง คือการ minimize หรือลดค่า loss ให้มีค่าต่ำที่สุด ดังนั้น จึงต้องการหน่วยวัดเมตริกที่สามารถคำนวณระยะห่างระหว่างความจริงที่คาดการณ์ไว้ (prediction) และความจริงที่เกิดขึ้น (actual) ถึงจะได้ error ของแบบจำลองออกมา ซึ่งตัววัดประสิทธิภาพที่ใช้โดยทั่วไปของ Machine Learning (ML) ประเภท Regression models

งานวิจัยชิ้นนี้ ได้ใช้ evaluation indicators ที่ใช้โดยทั่วไป เพื่อประเมินความแม่นยำ (Accuracy) ในการทำนายผลลัพธ์ของแต่ละแบบจำลอง ซึ่งผู้วิจัยใช้อยู่ทั้งหมด 3 ตัว ได้แก่ ประกอบไปด้วย  $R^2$  (R-Squared), Root Mean Squared Error (RMSE) และ Mean Absolute Percentage Error (MAPE) โดยมีรายละเอียด ดังนี้

#### 3.6.1 $R^2$ (R-Squared)

ค่า R-Squared เป็นผลทางสถิติที่ใช้วัดว่าสมการทางคณิตศาสตร์ที่มีอยู่ มีเหมาะสมกับข้อมูลมากน้อยเพียงใด หรือ ค่าสัมประสิทธิ์แสดงการตัดสินใจ (Coefficient of Determination) หรือ ค่าสัมประสิทธิ์แสดง การตัดสินใจในเชิงซ้อน (Coefficient of Multiple Determination) โดยการวิเคราะห์การถดถอยแบบพหุคูณ (Multiple Regression) ซึ่งความหมายของค่า R-Squared คือค่าความผันแปรของตัวแปรตามซึ่งอธิบายได้มีอยู่ในสมการที่เปอร์เซ็นต์ ค่า R-Squared จะอยู่ในช่วง 0% - 100% ซึ่งค่า R-Squared ยิ่งสูง หมายถึง สมการนั้นเหมาะสมกับข้อมูล แต่ก็มีเงื่อนไขบางประการที่ต้องมีพร้อมกับค่า R-Squared สูงๆ นี้ด้วยเช่นกัน

$$R^2 = \left[ \frac{1}{M} \frac{\sum_{j=1}^M [(Y_j - \bar{Y})(X_j - \bar{X})]}{\sigma_y \sigma_x} \right]^2 \quad (3.26)$$

โดยที่

|            |  |
|------------|--|
| $M$        | คือ จำนวนข้อมูล                                      |
| $\sigma_x$ | คือ ค่าเบี่ยงเบนมาตรฐานของการสำรวจ $X$               |
| $\sigma_y$ | คือ ค่าเบี่ยงเบนมาตรฐานของค่าที่ได้มาจากการคำนวณ $Y$ |
| $X_j$      | คือ ค่าที่ได้มาจากการสำรวจ                           |
| $\bar{X}$  | คือ ค่าเฉลี่ยของค่าที่ได้มาจากการสำรวจ               |
| $Y_j$      | คือ ค่าที่ได้มาจากการคำนวณ                           |
| $\bar{Y}$  | คือ ค่าเฉลี่ยของค่าที่ได้มาจากการคำนวณ               |

### 3.6.2 Mean Squared Error (RMSE)

RMSE เป็นอีกหนึ่งตัวบ่งชี้ทั่วไปที่ใช้ในการประเมินความแม่นยำของแบบจำลองการคาดการณ์ ซึ่ง RMSE เป็นรากที่สองของ MSE และมีความไวต่อข้อผิดพลาดขนาดใหญ่มากกว่า MAPE โดยมีสูตรการคำนวณ ดังสมการที่ (3.27)

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (\hat{x}_i - x_i)^2} \quad (3.27)$$

โดยที่

|             |                            |
|-------------|----------------------------|
| $M$         | คือ จำนวนข้อมูล            |
| $\hat{x}_i$ | คือ ค่าที่ได้มาจากการคำนวณ |
| $x_i$       | คือ ค่าที่ได้มาจากการสำรวจ |

### 3.6.3 Mean Absolute Percentage Error (MAPE)

MAPE เป็นการวัดความแม่นยำในการทำนายของวิธีการพยากรณ์ และโดยทั่วไปถือว่าเป็นตัวบ่งชี้ที่ยุติธรรมที่สุด MAPE เป็นตัวบ่งชี้การประเมินความแม่นยำที่สำคัญที่สุด โดยนำแต่ละข้อผิดพลาดมาหารด้วยมูลค่าจริงเพื่อให้ได้ผลลัพธ์โดยเฉลี่ยเป็นเปอร์เซ็นต์ (%) โดยมีสูตรการคำนวณ ดังสมการที่ (3.28)

$$MAPE = \frac{1}{m} \sum_{i=1}^m \left| \frac{\hat{x}_i - x_i}{x_i} \right| \cdot 100\% \quad (3.28)$$

โดยที่

- $M$  คือ จำนวนข้อมูล  
 $\hat{x}_i$  คือ ค่าที่ได้มาจากการคำนวณ  
 $x_i$  คือ ค่าที่ได้มาจากการสำรวจ

ตาราง 3.11 การแปรผลระดับความถูกต้องแม่นยำของค่า MAPE

| MAPE      | ระดับความถูกต้องแม่นยำ   |
|-----------|--|
| <10%      | ผลของการคาดการณ์มีความแม่นยำสูงมาก (Highly Accurate Forecasting)     |
| 10% ~ 20% | ผลของการคาดการณ์อยู่ในเกณฑ์ที่ดี (Good Forecasting)                  |
| 20% ~ 50% | ผลของการคาดการณ์สมเหตุสมผล (Reasonable Forecasting)                  |
| >50%      | ผลของการคาดการณ์ไม่ค่อยน่าเชื่อถือ (Weak and Inaccurate Forecasting) |

ที่มา: Lewis (1982)



## บทที่ 4

### ผลการวิจัยและพัฒนา (Result)

จากการศึกษาขั้นตอนและวิธีการดำเนินการวิเคราะห์คาดการณ์ PM 2.5 ในเนื้อหาส่วนก่อนหน้า ในบทที่ 4 จะเป็นการวิเคราะห์ผลการศึกษาและพัฒนาแบบจำลองที่เกิดขึ้น ซึ่งผลการศึกษาจะถูกดำเนินการตั้งแต่การคัดเลือกตัวแปร (Feature) ที่เหมาะสมของแต่ละแบบจำลองสำหรับนำมาสร้างแบบจำลอง (Training set) และทำการทดสอบ (Test set) และนำแบบจำลองแต่ละชนิดไปทดสอบกับชุดข้อมูลในห้วงเวลาที่มีการแพร่ระบาดของ Covid-19 (Testout Set) แบบจำลองแต่ละชนิด โดยทำการทดสอบแบบจำลองด้วยชุดข้อมูลในห้วงเวลาที่แตกต่างกัน ประกอบด้วย ชุดข้อมูลเฉลี่ยรายเดือน เฉลี่ยรายวัน และเฉลี่ยรายชั่วโมง เพื่อวิเคราะห์ถึงความแม่นยำของแบบจำลองเมื่อความละเอียดของห้วงเวลาแตกต่างกัน ทั้งนี้ แบบจำลองแต่ละชนิดและทุกการทดสอบจะถูกตรวจสอบความถูกต้องและความแม่นยำจากหลักเกณฑ์ที่ใช้ตรวจสอบแบบจำลอง (Performance Criteria) ด้วย  $R^2$  RMSE และ MAPE ซึ่งเป็นสถิติที่ใช้กันเป็นสากล โดยรายละเอียดการวิเคราะห์ผลการศึกษามีดังต่อไปนี้

#### 4.1 ผลการทดสอบแบบจำลอง โดยไม่อาศัยข้อมูลในอดีต (No Historical Data)

จากการแบ่งกลุ่มตัวแปรต่าง ๆ ในบทที่ 3 โดยใช้ Random Forest Algorithm เพื่อกำจัดตัวแปรอิสระที่มีผลต่อตัวแปรตามน้อยที่สุดออกทีละตัว และได้กลุ่มตัวแปรมาทั้งหมด 10 กลุ่มนั้น ในส่วนนี้ ผู้วิจัยได้ดำเนินการทดสอบแบบจำลองด้วยตัวแปรกลุ่มต่าง ๆ โดยผลการทดสอบ จะใช้  $R^2$  ในการประเมินผลแบบจำลองของตัวแปรแต่ละกลุ่ม ซึ่งได้ผลการทดสอบมาดังนี้ ตาราง 4.1 แสดงผลการทดสอบแบบจำลอง DecisionTree โดยไม่อาศัยข้อมูลในอดีต (No Historical Data) ซึ่งจากกลุ่มตัวแปรทั้งหมด 10 กลุ่มที่ใช้ทดสอบ พบว่า กลุ่มตัวแปรที่ 4 ซึ่งประกอบไปด้วยทิศทางลม (Wind Direction) คาร์บอนมอนอกไซด์ (CO) ไนโตรเจนไดออกไซด์ ( $NO_2$ ) และฝุ่นหยาบ (PM 10) เป็นกลุ่มตัวแปรที่ดีที่สุดในการนำไปทดสอบแบบจำลองทั้ง 3 แบบ (Training set, Test set, Testout set) ซึ่งมีค่า  $R^2$  เฉลี่ยของทั้ง 3 แบบจำลองเท่ากับ 0.7683

ตาราง 4.2 แสดงผลการทดสอบแบบจำลอง Gradient Boosting โดยไม่อาศัยข้อมูลในอดีต (No Historical Data) ซึ่งจากกลุ่มตัวแปรทั้งหมด 10 กลุ่มที่ใช้ทดสอบ พบว่า กลุ่มตัวแปรที่ 7 ซึ่งประกอบไปด้วยทิศทางลม (Wind Direction) อุณหภูมิ (Temp) ความกดอากาศ (BP) คาร์บอนมอนอกไซด์ (CO) ไนตริกออกไซด์ (NO) ไนโตรเจนไดออกไซด์ ( $NO_2$ ) และฝุ่นหยาบ (PM 10) เป็นกลุ่มตัวแปรที่ดีที่สุดในการนำไปทดสอบแบบจำลองทั้ง 3 แบบ (Training set, Test set, Testout set) ซึ่งมีค่า  $R^2$  เฉลี่ยเท่ากับ 0.8088

ตาราง 4.1 ผลการทดสอบจำนวนตัวแปร (Feature) ที่ใช้ในการสร้างแบบจำลอง กรณี Trainset, Test set and Testout set ของแบบจำลอง DecisionTree

| จำนวน Feature<br>ที่ใช้ทดสอบแบบจำลอง<br>DecisionTree | Performance Criteria ( $R^2$ ) ของชุดข้อมูล |          |             | Avg. $R^2$ |
|--|---|----------|-------------|------------|
|  | Training set                                | Test set | Testout set |            |
| 1  | 0.6810                                      | 0.6595   | 0.5233      | 0.6212     |
| 2  | 0.7655                                      | 0.7507   | 0.5016      | 0.6726     |
| 3  | 0.7604                                      | 0.7427   | 0.7595      | 0.7542     |
| 4  | 0.7759                                      | 0.7674   | 0.7615      | 0.7683     |
| 5  | 0.7676                                      | 0.7260   | 0.7294      | 0.7410     |
| 6  | 0.7763                                      | 0.7307   | 0.7340      | 0.7470     |
| 7  | 0.7880                                      | 0.7508   | 0.7112      | 0.7500     |
| 8  | 0.6192                                      | 0.5558   | 0.6788      | 0.6179     |
| 9  | 0.6902                                      | 0.5435   | 0.6028      | 0.6122     |
| 10   | 0.6943                                      | 0.5782   | 0.5903      | 0.6209     |

ตาราง 4.2 ผลการทดสอบจำนวนตัวแปร (Feature) ที่ใช้ในการสร้างแบบจำลอง กรณี Trainset, Test set and Testout set ของแบบจำลอง Gradient Boosting

| จำนวน Feature<br>ที่ใช้ทดสอบแบบจำลอง<br>Gradient Boosting | Performance Criteria ( $R^2$ ) ของชุดข้อมูล |          |             | Avg. $R^2$ |
|---|---|----------|-------------|------------|
|   | Training set                                | Test set | Testout set |            |
| 1   | 0.7209                                      | 0.6534   | 0.4892      | 0.6212     |
| 2   | 0.7986                                      | 0.7577   | 0.4809      | 0.6791     |
| 3   | 0.7840                                      | 0.7545   | 0.7513      | 0.7633     |
| 4   | 0.8202                                      | 0.7853   | 0.7591      | 0.7882     |
| 5   | 0.8750                                      | 0.7661   | 0.7259      | 0.7890     |
| 6   | 0.8822                                      | 0.7853   | 0.7382      | 0.8019     |
| 7   | 0.8688                                      | 0.8214   | 0.7962      | 0.8088     |
| 8   | 0.8327                                      | 0.7002   | 0.6833      | 0.7387     |
| 9   | 0.9134                                      | 0.6759   | 0.6510      | 0.7468     |
| 10  | 0.9246                                      | 0.7016   | 0.6530      | 0.7598     |

ตาราง 4.3 แสดงผลการทดสอบแบบจำลอง K-neighbors โดยไม่อาศัยข้อมูลในอดีต (No Historical Data) ซึ่งจากกลุ่มตัวแปรทั้งหมด 10 กลุ่มที่ใช้ทดสอบ พบว่า กลุ่มตัวแปรที่ 4 ซึ่งประกอบไปด้วยทิศทางลม (Wind Direction) คาร์บอนมอนอกไซด์ (CO) ไนโตรเจนไดออกไซด์ (NO<sub>2</sub>) และฝุ่นหยาบ (PM 10) เป็นกลุ่มตัวแปรที่ดีที่สุดในการนำไปทดสอบแบบจำลองทั้ง 3 แบบ (Training set, Test set, Testout set) ซึ่งมีค่า  $R^2$  เฉลี่ยของทั้ง 3 แบบจำลองเท่ากับ 0.7732

ตาราง 4.4 แสดงผลการทดสอบแบบจำลอง MLP โดยไม่อาศัยข้อมูลในอดีต (No Historical Data) ซึ่งจากกลุ่มตัวแปรทั้งหมด 10 กลุ่มที่ใช้ทดสอบ พบว่า กลุ่มตัวแปรที่ 4 ซึ่งประกอบไปด้วย ทิศทางลม (Wind Direction) คาร์บอนมอนอกไซด์ (CO) ไนโตรเจนไดออกไซด์ (NO<sub>2</sub>) และฝุ่นหยาบ (PM 10) เป็นกลุ่มตัวแปรที่ดีที่สุดในการนำไปทดสอบแบบจำลองทั้ง 3 แบบ (Training set, Test set, Testout set) ซึ่งมีค่า R<sup>2</sup> เฉลี่ยของทั้ง 3 แบบจำลองเท่ากับ 0.7769

ตาราง 4.5 แสดงผลการทดสอบแบบจำลอง Random Forest โดยไม่อาศัยข้อมูลในอดีต (No Historical Data) ซึ่งจากกลุ่มตัวแปรทั้งหมด 10 กลุ่มที่ใช้ทดสอบ พบว่า กลุ่มตัวแปรที่ 6 ซึ่งประกอบไปด้วยทิศทางลม (Wind Direction) ความกดอากาศ (BP) คาร์บอนมอนอกไซด์ (CO) ไนตริกออกไซด์ (NO) ไนโตรเจนไดออกไซด์ (NO<sub>2</sub>) และฝุ่นหยาบ (PM 10) เป็นกลุ่มตัวแปรที่ดีที่สุดในการนำไปทดสอบแบบจำลองทั้ง 3 แบบ (Training set, Test set, Testout set) ซึ่งมีค่า R<sup>2</sup> เฉลี่ยของทั้ง 3 แบบจำลองเท่ากับ 0.7892

ตาราง 4.6 แสดงผลการทดสอบแบบจำลอง Ridge โดยไม่อาศัยข้อมูลในอดีต (No Historical Data) ซึ่งจากกลุ่มตัวแปรทั้งหมด 10 กลุ่มที่ใช้ทดสอบ พบว่า กลุ่มตัวแปรที่ 4 ซึ่งประกอบไปด้วย ทิศทางลม (Wind Direction) คาร์บอนมอนอกไซด์ (CO) ไนโตรเจนไดออกไซด์ (NO<sub>2</sub>) และฝุ่นหยาบ (PM 10) เป็นกลุ่มตัวแปรที่ดีที่สุดในการนำไปทดสอบแบบจำลองทั้ง 3 แบบ (Training set, Test set, Testout set) ซึ่งมีค่า R<sup>2</sup> เฉลี่ยของทั้ง 3 แบบจำลองเท่ากับ 0.7004

**ตาราง 4.3** ผลการทดสอบจำนวนตัวแปร (Feature) ที่ใช้ในการสร้างแบบจำลอง กรณี Trainset, Test set and Testout set ของแบบจำลอง K-neighbors

| จำนวน Feature<br>ที่ใช้ทดสอบแบบจำลอง<br>K-neighbors | Performance Criteria (R <sup>2</sup> ) ของชุดข้อมูล |          |             | Avg. R <sup>2</sup> |
|---|---|----------|-------------|---------------------|
|   | Training set  | Test set | Testout set |                     |
| 1   | 0.6736  | 0.6632   | 0.5343      | 0.6237              |
| 2   | 0.7664  | 0.7574   | 0.5151      | 0.6796              |
| 3   | 0.7627  | 0.7486   | 0.7392      | 0.7502              |
| 4   | 0.7915  | 0.7768   | 0.7515      | 0.7732              |
| 5   | 0.7910  | 0.7552   | 0.7233      | 0.7565              |
| 6   | 0.8083  | 0.7694   | 0.7281      | 0.7686              |
| 7   | 0.8238  | 0.7858   | 0.7043      | 0.7713              |
| 8   | 0.7977  | 0.6884   | 0.6133      | 0.6998              |
| 9   | 0.7901  | 0.6653   | 0.6207      | 0.6920              |
| 10  | 0.8001  | 0.6672   | 0.5754      | 0.6809              |

ตาราง 4.4 ผลการทดสอบจำนวนตัวแปร (Feature) ที่ใช้ในการสร้างแบบจำลอง กรณี Trainset, Test set and Testout set ของแบบจำลอง MLP

| จำนวน Feature<br>ที่ใช้ทดสอบแบบจำลอง<br>MLP | Performance Criteria ( $R^2$ ) ของชุดข้อมูล |          |             | Avg. $R^2$ |
|---|---|----------|-------------|------------|
|   | Training set                                | Test set | Testout set |            |
| 1   | 0.4241                                      | 0.4296   | 0.1153      | 0.3230     |
| 2   | 0.7263                                      | 0.7232   | 0.4865      | 0.6453     |
| 3   | 0.7199                                      | 0.7152   | 0.7538      | 0.7296     |
| 4   | 0.7645                                      | 0.7703   | 0.7505      | 0.7618     |
| 5   | 0.7334                                      | 0.7243   | 0.7470      | 0.7349     |
| 6   | 0.7480                                      | 0.7499   | 0.6823      | 0.7267     |
| 7   | 0.7507                                      | 0.7601   | 0.7369      | 0.7492     |
| 8   | 0.6537                                      | 0.6266   | 0.6607      | 0.6470     |
| 9   | 0.6522                                      | 0.5993   | 0.6782      | 0.6433     |
| 10  | 0.6520                                      | 0.6249   | 0.6138      | 0.6302     |

ตาราง 4.5 ผลการทดสอบจำนวนตัวแปร (Feature) ที่ใช้ในการสร้างแบบจำลอง กรณี Trainset, Test set and Testout set ของแบบจำลอง Random Forest

| จำนวน Feature<br>ที่ใช้ทดสอบแบบจำลอง<br>Random Forest | Performance Criteria ( $R^2$ ) ของชุดข้อมูล |          |             | Avg. $R^2$ |
|---|---|----------|-------------|------------|
|   | Training set                                | Test set | Testout set |            |
| 1   | 0.6829                                      | 0.6655   | 0.5120      | 0.6268     |
| 2   | 0.7768                                      | 0.7605   | 0.5353      | 0.6975     |
| 3   | 0.7851                                      | 0.7536   | 0.7239      | 0.7609     |
| 4   | 0.8114                                      | 0.7831   | 0.7441      | 0.7862     |
| 5   | 0.8123                                      | 0.7583   | 0.7296      | 0.7734     |
| 6   | 0.8335                                      | 0.7757   | 0.7384      | 0.7892     |
| 7   | 0.8277                                      | 0.7857   | 0.7325      | 0.7886     |
| 8   | 0.7428                                      | 0.6616   | 0.6858      | 0.7034     |
| 9   | 0.7377                                      | 0.6151   | 0.6837      | 0.6855     |
| 10  | 0.7351                                      | 0.6439   | 0.6683      | 0.6891     |

ตาราง 4.6 ผลการทดสอบจำนวนตัวแปร (Feature) ที่ใช้ในการสร้างแบบจำลอง กรณี Trainset, Test set and Testout set ของแบบจำลอง Ridge

| จำนวน Feature<br>ที่ใช้ทดสอบแบบจำลอง<br>Ridge | Performance Criteria ( $R^2$ ) ของชุดข้อมูล |          |             | Avg. $R^2$ |
|---|---|----------|-------------|------------|
|   | Training set                                | Test set | Testout set |            |
| 1   | 0.6186                                      | 0.6138   | 0.4770      | 0.5698     |
| 2   | 0.6798                                      | 0.6784   | 0.5248      | 0.6277     |
| 3   | 0.6676                                      | 0.6645   | 0.7179      | 0.6833     |
| 4   | 0.6883                                      | 0.6933   | 0.7196      | 0.7004     |
| 5   | 0.6567                                      | 0.6497   | 0.6832      | 0.6632     |
| 6   | 0.6654                                      | 0.6717   | 0.6882      | 0.6751     |
| 7   | 0.6723                                      | 0.6809   | 0.6861      | 0.6798     |
| 8   | 0.5573                                      | 0.5395   | 0.6433      | 0.5800     |
| 9   | 0.5641                                      | 0.5237   | 0.6357      | 0.5745     |
| 10  | 0.5572                                      | 0.5425   | 0.6224      | 0.5740     |

ตาราง 4.7 สรุปผลการคัดเลือกตัวแปรของแบบจำลองชนิดต่าง ๆ โดยไม่อาศัยข้อมูลในอดีต (No Historical Data)

| แบบจำลอง         | Feature<br>ที่เหมาะสม | ตัวแปรที่เกี่ยวข้อง  | Avg. $R^2$ |
|------------------|-----------------------|--|------------|
| DecisionTree     | 4                     | ['WD(Deg)', 'CO(ppm)', 'NO2(ppb)', 'PM 10(ug/m3)']                                       | 0.7683     |
| GradientBoosting | 7                     | ['WD(Deg)', 'Temp(Deg.C)', 'BP(mBar)', 'CO(ppm)', 'NO(ppb)', 'NO2(ppb)', 'PM 10(ug/m3)'] | 0.8088     |
| K-neighbors      | 4                     | ['WD(Deg)', 'CO(ppm)', 'NO2(ppb)', 'PM 10(ug/m3)']                                       | 0.7732     |
| MLP              | 4                     | ['WD(Deg)', 'CO(ppm)', 'NO2(ppb)', 'PM 10(ug/m3)']                                       | 0.7618     |
| RandomForest     | 6                     | ['WD(Deg)', 'BP(mBar)', 'CO(ppm)', 'NO(ppb)', 'NO2(ppb)', 'PM 10(ug/m3)']                | 0.7892     |
| Ridge            | 4                     | ['WD(Deg)', 'CO(ppm)', 'NO2(ppb)', 'PM 10(ug/m3)']                                       | 0.7004     |

จากผลการทดสอบตัวแปรที่ใช้ในการสร้างแบบจำลอง สามารถสรุปผลการทดสอบที่เหมาะสมที่สุดของแต่ละแบบจำลอง ดังตาราง 4.7

โดยจากผลการทดสอบข้างต้น พบว่า ภาพรวมของผลการศึกษาจากแบบจำลอง GradientBoosting มีความแม่นยำ ( $R^2$ ) สูงที่สุด ตามด้วย Random Forest, K-neighbors, DecisionTree, MLP และ Ridge

ทั้งนี้ ผู้วิจัยได้ทำการตรวจสอบความถูกต้องและความแม่นยำจากหลักเกณฑ์ที่ใช้ตรวจสอบแบบจำลอง (Performance Criteria) ที่เหลือด้วย RMSE และ MAPE โดยได้ผลการทดสอบจากตัวแปรที่เหมาะสมในแต่ละแบบจำลอง ดังนี้



เมื่อพิจารณาผลการทดสอบของแบบจำลองในกรณี Training Set และ Test Set ในตาราง 4.8 พบว่า ความแม่นยำของแบบจำลองของทั้ง 2 ส่วนมีความแม่นยำสูงที่สุดเมื่อทดสอบด้วยแบบจำลอง Gradient Boosting ซึ่งจากชุดข้อมูล Training Set ได้  $R^2 = 0.8688$  สูงกว่าแบบจำลอง Test Set ซึ่งได้  $R^2 = 0.8214$  ตามด้วยแบบจำลอง Random Forest และ แบบจำลอง MLP มีประสิทธิภาพต่ำที่สุดเมื่อเทียบกับแบบจำลองชนิดอื่น

เมื่อพิจารณาผลการทดสอบของแบบจำลองที่ใช้ชุดข้อมูล PM 2.5 ในห้วงที่มีการแพร่ระบาดของ Covid-19 หรือเรียกว่า กรณี Testout Set ในตาราง 4.9 พบว่า ความแม่นยำของแบบจำลองจะมีความแม่นยำสูงที่สุดเมื่อทดสอบด้วยแบบจำลอง Gradient Boosting ซึ่งได้ผลการทดสอบ  $R^2 = 0.7962$  ตามด้วยแบบจำลอง Random Forest และ แบบจำลอง Ridge มีประสิทธิภาพต่ำที่สุดเมื่อเทียบกับแบบจำลองชนิดอื่น

**ตาราง 4.8** ผลการทดสอบความถูกต้องและความแม่นยำของแบบจำลอง Training set และ Test Set โดยไม่อาศัยข้อมูลในอดีต (No Historical Data)

| Model                    | Training      |               |               | Test          |               |               |
|--------------------------|---------------|---------------|---------------|---------------|---------------|---------------|
|                          | $R^2$         | RMSE          | MAPE          | $R^2$         | RMSE          | MAPE          |
| Decision Tree            | 0.7759        | 8.4434        | 35.15%        | 0.7674        | 8.6631        | 35.34%        |
| <b>Gradient Boosting</b> | <b>0.8688</b> | <b>6.5459</b> | <b>21.53%</b> | <b>0.8214</b> | <b>8.1684</b> | <b>26.79%</b> |
| K-neighbors              | 0.7915        | 8.1444        | 33.40%        | 0.7768        | 8.4879        | 34.14%        |
| MLP                      | 0.7645        | 8.6563        | 35.68%        | 0.7703        | 8.6089        | 35.11%        |
| Random Forest            | 0.8335        | 7.3898        | 25.08%        | 0.7757        | 8.6443        | 28.52%        |
| Ridge                    | 0.6883        | 9.9582        | 40.03%        | 0.6933        | 9.9480        | 39.63%        |

**ตาราง 4.9** ผลการทดสอบความถูกต้องและความแม่นยำของแบบจำลอง Testout (Covid-19 Pandemic) โดยไม่อาศัยข้อมูลในอดีต (No Historical Data)

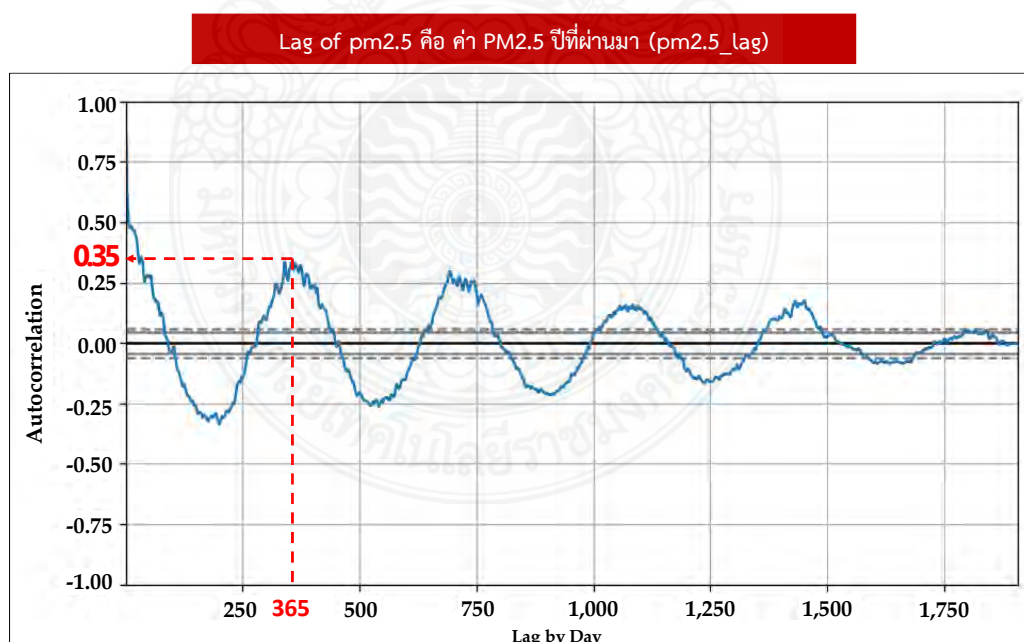
| Model                    | Testout (Covid-19 Pandemic) |               |               |
|--------------------------|-----------------------------|---------------|---------------|
|                          | $R^2$                       | RMSE          | MAPE          |
| Decision Tree            | 0.7615                      | 7.6303        | 36.77%        |
| <b>Gradient Boosting</b> | <b>0.7962</b>               | <b>7.0337</b> | <b>29.92%</b> |
| K-neighbors              | 0.7515                      | 7.7893        | 35.85%        |
| MLP                      | 0.7505                      | 7.8567        | 31.73%        |
| Random Forest            | 0.7384                      | 7.9845        | 34.85%        |
| Ridge                    | 0.7196                      | 8.2743        | 35.22%        |

## 4.2 ผลการทดสอบแบบจำลอง โดยอาศัยข้อมูลในอดีต (Model Results by Historical Data)

จากการวิเคราะห์ข้อมูลเบื้องต้นในบทที่ 3 เกี่ยวกับค่าความเข้มข้นของ PM 2.5 ตามช่วงเวลาต่าง ๆ พบว่า ในแต่ละช่วงเวลา ค่าความเข้มข้นของ PM 2.5 จะไม่เท่ากัน แม้ในแต่ละวัน หรือเดือน ซึ่งตรงกับงานวิจัย หรือข้อมูลการศึกษาในก่อนหน้านี้ ซึ่งแสดงให้เห็นว่าพฤติกรรมของ PM 2.5 และสารมลพิษทางอากาศค่าต่าง ๆ จะขึ้นอยู่กับช่วงเวลาต่าง ๆ ด้วย

สำหรับการนำข้อมูล PM 2.5 ทั้ง 2.3 ล้านชุดข้อมูลที่เป็นอนุกรมเวลาไปใช้ในการทดสอบแบบจำลอง ผู้วิจัยได้ทำการสัมพันธ์อัตโนมัติ (Autocorrelation) เพื่อหาความสัมพันธ์ของช่วงข้อมูลในอดีตที่มีต่อข้อมูล ณ ปัจจุบัน โดยพบว่า ชุดข้อมูลในอดีตที่จะนำมาใช้ในการทดสอบแบบจำลอง เพื่อให้ได้ผลการศึกษาที่มีประสิทธิภาพมากที่สุด คือ ชุดข้อมูลในห้วง 365 วัน หรือ 1 ปี โดยพบว่า มีความสัมพันธ์ต่อข้อมูลในปัจจุบันอยู่ที่ 0.35 แต่หากจะนำชุดข้อมูลในห้วงเวลาที่มากกว่า 365 วันไปใช้ในการทดสอบแบบจำลอง ความสัมพันธ์ของชุดข้อมูลในอดีตต่อชุดข้อมูลในปัจจุบัน ก็ยิ่งลดน้อยลง ตามลำดับ ดังแสดงในภาพ 4.1

ทั้งนี้ ผู้วิจัยจึงได้ทำการสังเคราะห์ข้อมูล PM 2.5 ตามช่วงเวลาต่าง ๆ เช่น เฉลี่ยรายชั่วโมง เฉลี่ยรายวัน และเฉลี่ยรายเดือน เพื่อที่จะสามารถวิเคราะห์ถึงความแม่นยำของแบบจำลองแต่ละชนิดในกรณีที่มีข้อมูลมีความละเอียดแตกต่างกัน โดยรายละเอียดของการแบ่งข้อมูล PM 2.5 ตามช่วงเวลาเป็นดังนี้



**Autocorrelation plot.**

ภาพ 4.1 การทำ Autocorrelation plot ของข้อมูลรายวัน

#### 4.2.1 ข้อมูลค่า PM 2.5 เฉลี่ยรายเดือน (Lag by Month)

โดยข้อมูลค่า PM 2.5 เฉลี่ยรายเดือน (Lag by Month) คำนวณจากข้อมูล PM 2.5 โดยเฉลี่ย ในเดือนเดียวกันของปีก่อนหน้า แยกคำนวณตามแต่ละสถานี และนำมาเป็น 1 ในตัวแปร (Feature) สำหรับทดสอบแบบจำลอง

จากการแบ่งกลุ่มตัวแปรต่างๆ ในบทที่ 3 โดยใช้ Random Forest Algorithm เพื่อกำจัดตัวแปรอิสระที่มีผลต่อตัวแปรตามน้อยที่สุดออกทีละตัว โดยการวิเคราะห์นี้ ผู้วิจัยได้เพิ่มตัวแปร ซึ่งเป็นข้อมูลค่า PM 2.5 เฉลี่ยรายเดือน (Lag by Month) และได้กลุ่มตัวแปรมาทั้งหมด 14 กลุ่ม นำไปทดสอบแบบจำลองทั้ง 3 แบบ (Training set, Test set, Testout set) โดยผลการทดสอบ จะใช้  $R^2$  ในการประเมินผลแบบจำลองของตัวแปรแต่ละกลุ่ม ซึ่งได้ผลการทดสอบดังแสดงในตาราง 4.10 แสดงผลการทดสอบแบบจำลองทั้ง 6 ชนิด



ตาราง 4.10 ผลการทดสอบจำนวนตัวแปร (Feature) ที่ใช้ในการสร้างแบบจำลอง กรณี Training set, Test set, Testout set ของแบบจำลอง ทั้ง 6 ชนิด เมื่อเพิ่มตัวแปร (Feature) ที่เป็นข้อมูลค่า PM 2.5 เฉลี่ยรายเดือน (Lag by Month)

| Feature Group | R <sup>2</sup> of Decision Tree |        |        |                     | R <sup>2</sup> of Gradient Boosting |        |        |                     | R <sup>2</sup> of K-neighbors |        |        |                     | R <sup>2</sup> of MLP |        |        |                     | R <sup>2</sup> of Random Forest |        |        |                     | R <sup>2</sup> of Ridge |        |        |                     |
|---------------|---------------------------------|--------|--------|---------------------|-------------------------------------|--------|--------|---------------------|-------------------------------|--------|--------|---------------------|-----------------------|--------|--------|---------------------|---------------------------------|--------|--------|---------------------|-------------------------|--------|--------|---------------------|
|               | Train                           | Test   | Testo  | Avg. R <sup>2</sup> | Train                               | Test   | Testo  | Avg. R <sup>2</sup> | Train                         | Test   | Testo  | Avg. R <sup>2</sup> | Train                 | Test   | Testo  | Avg. R <sup>2</sup> | Train                           | Test   | Testo  | Avg. R <sup>2</sup> | Train                   | Test   | Testo  | Avg. R <sup>2</sup> |
| 1             | 0.7410                          | 0.7338 | 0.5913 | 0.6887              | 0.7505                              | 0.7326 | 0.5789 | 0.6873              | 0.7384                        | 0.7353 | 0.6106 | 0.6947              | 0.6789                | 0.6787 | 0.2941 | 0.5506              | 0.7420                          | 0.7371 | 0.6063 | 0.6951              | 0.6655                  | 0.6647 | 0.5501 | 0.6267              |
| 2             | 0.8488                          | 0.8270 | 0.7643 | 0.8134              | 0.8592                              | 0.8442 | 0.7746 | 0.8260              | 0.8466                        | 0.8325 | 0.7651 | 0.8147              | 0.7904                | 0.7936 | 0.7258 | 0.7699              | 0.8478                          | 0.8355 | 0.7867 | 0.8233              | 0.7099                  | 0.7078 | 0.7378 | 0.7185              |
| 3             | 0.8684                          | 0.8444 | 0.7505 | <b>0.8211</b>       | 0.8877                              | 0.8667 | 0.7641 | 0.8395              | 0.8642                        | 0.8408 | 0.7567 | 0.8206              | 0.8236                | 0.8205 | 0.7629 | 0.8023              | 0.8763                          | 0.8557 | 0.7754 | 0.8358              | 0.7301                  | 0.7264 | 0.7253 | 0.7272              |
| 4             | 0.8765                          | 0.8449 | 0.7446 | <b>0.8220</b>       | 0.8945                              | 0.8670 | 0.7827 | 0.8480              | 0.8726                        | 0.8504 | 0.7750 | <b>0.8326</b>       | 0.8324                | 0.8232 | 0.7721 | 0.8092              | 0.8841                          | 0.8578 | 0.7838 | 0.8419              | 0.7362                  | 0.7316 | 0.7384 | <b>0.7354</b>       |
| 5             | 0.8621                          | 0.8344 | 0.7533 | 0.8166              | 0.8935                              | 0.8630 | 0.7911 | 0.8492              | 0.8624                        | 0.8390 | 0.7711 | 0.8242              | 0.8200                | 0.8218 | 0.8006 | 0.8141              | 0.8811                          | 0.8558 | 0.7954 | 0.8441              | 0.7251                  | 0.7309 | 0.7272 | 0.7277              |
| 6             | 0.8713                          | 0.8443 | 0.7183 | 0.8113              | 0.9221                              | 0.8930 | 0.7998 | <b>0.8716</b>       | 0.8659                        | 0.8344 | 0.7646 | 0.8217              | 0.8299                | 0.8292 | 0.7856 | <b>0.8149</b>       | 0.8973                          | 0.8663 | 0.7988 | <b>0.8541</b>       | 0.7297                  | 0.7281 | 0.7307 | 0.7295              |
| 7             | 0.8540                          | 0.8077 | 0.7154 | 0.7924              | 0.9235                              | 0.8452 | 0.7787 | 0.8491              | 0.8425                        | 0.8187 | 0.7516 | 0.8043              | 0.8140                | 0.8094 | 0.7311 | 0.7848              | 0.8824                          | 0.8387 | 0.7900 | 0.8370              | 0.7164                  | 0.7150 | 0.7143 | 0.7152              |
| 8             | 0.8459                          | 0.7929 | 0.7603 | 0.7997              | 0.9134                              | 0.8406 | 0.7746 | 0.8429              | 0.8692                        | 0.8114 | 0.7209 | 0.8005              | 0.8209                | 0.8083 | 0.7737 | 0.8010              | 0.8864                          | 0.8289 | 0.7891 | 0.8348              | 0.7203                  | 0.7202 | 0.7118 | 0.7174              |
| 9             | 0.7189                          | 0.6654 | 0.6752 | 0.6865              | 0.9743                              | 0.7535 | 0.6736 | 0.8005              | 0.8301                        | 0.7257 | 0.5994 | 0.7184              | 0.7396                | 0.7025 | 0.6442 | 0.6954              | 0.8334                          | 0.7304 | 0.7263 | 0.7634              | 0.6316                  | 0.6075 | 0.6319 | 0.6237              |
| 10            | 0.7746                          | 0.7016 | 0.5473 | 0.6745              | 0.8935                              | 0.7705 | 0.6723 | 0.7788              | 0.8328                        | 0.7406 | 0.5583 | 0.7106              | 0.7340                | 0.7369 | 0.6574 | 0.7095              | 0.8380                          | 0.7394 | 0.7129 | 0.7634              | 0.6301                  | 0.6347 | 0.6368 | 0.6338              |
| 11            | 0.7533                          | 0.7264 | 0.5445 | 0.6747              | 0.9497                              | 0.8028 | 0.6724 | 0.8083              | 0.8312                        | 0.7623 | 0.5788 | 0.7241              | 0.7417                | 0.7470 | 0.6703 | 0.7197              | 0.7990                          | 0.7610 | 0.7209 | 0.7603              | 0.6399                  | 0.6581 | 0.6337 | 0.6439              |
| 12            | 0.7552                          | 0.7250 | 0.5641 | 0.6814              | 0.9264                              | 0.8072 | 0.6755 | 0.8030              | 0.8114                        | 0.7700 | 0.5716 | 0.7177              | 0.7609                | 0.7557 | 0.6097 | 0.7088              | 0.8463                          | 0.7864 | 0.7280 | 0.7869              | 0.6486                  | 0.6444 | 0.6272 | 0.6401              |
| 13            | 0.7587                          | 0.7040 | 0.6141 | 0.6923              | 0.9757                              | 0.7851 | 0.6337 | 0.7982              | 0.8329                        | 0.7381 | 0.5640 | 0.7117              | 0.7760                | 0.7476 | 0.6099 | 0.7112              | 0.8541                          | 0.7617 | 0.7063 | 0.7740              | 0.6401                  | 0.6433 | 0.5873 | 0.6235              |
| 14            | 0.7627                          | 0.6976 | 0.6135 | 0.6913              | 0.9831                              | 0.8028 | 0.6462 | 0.8107              | 0.8306                        | 0.7487 | 0.5687 | 0.7160              | 0.7578                | 0.7466 | 0.6430 | 0.7158              | 0.8338                          | 0.7632 | 0.7104 | 0.7691              | 0.6415                  | 0.6336 | 0.6113 | 0.6288              |

หมายเหตุ : Train คือ Training set, Test คือ Test set และ Testo คือ testout-set data

จากตาราง 4.11 พบว่า แบบจำลอง DecisionTree ได้กลุ่มตัวแปรที่เหมาะสมที่สุดคือ กลุ่มตัวแปรที่ 4 ซึ่งประกอบไปด้วยทิศทางลม (Wind Direction) ฝุ่นหยาบ (PM 10) ฝุ่นละเอียด (PM 2.5) ย้อนหลัง และน้ำหนักรถของจำนวนเดือน (Month) ย้อนหลัง ซึ่งเป็นกลุ่มตัวแปรที่ดีที่สุดในการนำไปทดสอบแบบจำลองทั้ง 3 แบบ (Training set, Test set, Testout set) มีค่า  $R^2$  เฉลี่ยของทั้ง 3 แบบจำลองเท่ากับ 0.8211

แบบจำลอง GradientBoosting ได้กลุ่มตัวแปรที่เหมาะสมที่สุดคือ กลุ่มตัวแปรที่ 6 ซึ่งประกอบไปด้วยทิศทางลม (Wind Direction) คาร์บอนมอนอกไซด์ (CO) ไนตริกออกไซด์ (NO) ฝุ่นหยาบ (PM 10) ฝุ่นละเอียด (PM 2.5) ย้อนหลัง และน้ำหนักรถของจำนวนเดือน (Month) ย้อนหลัง ซึ่งเป็นกลุ่มตัวแปรที่ดีที่สุดในการนำไปทดสอบแบบจำลองทั้ง 3 แบบ (Training set, Test set, Testout set) มีค่า  $R^2$  เฉลี่ยของทั้ง 3 แบบจำลองเท่ากับ 0.8716

แบบจำลอง K-neighbors ได้กลุ่มตัวแปรที่เหมาะสมที่สุดคือ กลุ่มตัวแปรที่ 4 ซึ่งประกอบไปด้วยทิศทางลม (Wind Direction) ฝุ่นหยาบ (PM 10) ฝุ่นละเอียด (PM 2.5) ย้อนหลัง และน้ำหนักรถของจำนวนเดือน (Month) ย้อนหลัง ซึ่งเป็นกลุ่มตัวแปรที่ดีที่สุดในการนำไปทดสอบแบบจำลองทั้ง 3 แบบ (Training set, Test set, Testout set) มีค่า  $R^2$  เฉลี่ยของทั้ง 3 แบบจำลองเท่ากับ 0.8326

แบบจำลอง MLP ได้กลุ่มตัวแปรที่เหมาะสมที่สุดคือ กลุ่มตัวแปรที่ 6 ซึ่งประกอบไปด้วยทิศทางลม (Wind Direction) คาร์บอนมอนอกไซด์ (CO) ไนตริกออกไซด์ (NO) ฝุ่นหยาบ (PM 10) ฝุ่นละเอียด (PM 2.5) ย้อนหลัง และน้ำหนักรถของจำนวนเดือน (Month) ย้อนหลัง ซึ่งเป็นกลุ่มตัวแปรที่ดีที่สุดในการนำไปทดสอบแบบจำลองทั้ง 3 แบบ (Training set, Test set, Testout set) มีค่า  $R^2$  เฉลี่ยของทั้ง 3 แบบจำลองเท่ากับ 0.8149

แบบจำลอง Random Forest ได้กลุ่มตัวแปรที่เหมาะสมที่สุดคือ กลุ่มตัวแปรที่ 6 ซึ่งประกอบไปด้วยทิศทางลม (Wind Direction) คาร์บอนมอนอกไซด์ (CO) ไนตริกออกไซด์ (NO) ฝุ่นหยาบ (PM 10) ฝุ่นละเอียด (PM 2.5) ย้อนหลัง และน้ำหนักรถของจำนวนเดือน (Month) ย้อนหลัง ซึ่งเป็นกลุ่มตัวแปรที่ดีที่สุดในการนำไปทดสอบแบบจำลองทั้ง 3 แบบ (Training set, Test set, Testout set) มีค่า  $R^2$  เฉลี่ยของทั้ง 3 แบบจำลองเท่ากับ 0.8541

แบบจำลอง Ridge ได้กลุ่มตัวแปรที่เหมาะสมที่สุดคือ กลุ่มตัวแปรที่ 4 ซึ่งประกอบไปด้วยทิศทางลม (Wind Direction) ฝุ่นหยาบ (PM 10) ฝุ่นละเอียด (PM 2.5) ย้อนหลัง และน้ำหนักรถของจำนวนเดือน (Month) ย้อนหลัง ซึ่งเป็นกลุ่มตัวแปรที่ดีที่สุดในการนำไปทดสอบแบบจำลองทั้ง 3 แบบ (Training set, Test set, Testout set) มีค่า  $R^2$  เฉลี่ยของทั้ง 3 แบบจำลองเท่ากับ 0.7354

ตาราง 4.11 สรุปผลการคัดเลือกตัวแปรสำหรับนำไปวิเคราะห์แบบจำลอง กรณีเพิ่มตัวแปรข้อมูล  
มูลค่า PM 2.5 เฉลี่ยรายเดือน (Lag by Month)

| แบบจำลอง         | Feature<br>ที่เหมาะสม | ตัวแปรที่เกี่ยวข้อง  | Avg. R <sup>2</sup> |
|------------------|-----------------------|--|---------------------|
| DecisionTree     | 4                     | ['WD(Deg)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag']                                     | 0.8220              |
| GradientBoosting | 6                     | ['WD(Deg)', 'CO(ppm)', 'NO(ppb)', 'PM 10 (ug/m3)', 'PM 2.5(ug/m3)_lag', 'Month_WOE'] | 0.8716              |
| K-neighbors      | 4                     | ['WD(Deg)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag', 'Month_WOE']                        | 0.8326              |
| MLP              | 6                     | ['WD(Deg)', 'CO(ppm)', 'NO(ppb)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag', 'Month_WOE']  | 0.8149              |
| RandomForest     | 6                     | ['WD(Deg)', 'CO(ppm)', 'NO(ppb)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag', 'Month_WOE']  | 0.8541              |
| Ridge            | 4                     | ['WD(Deg)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag', 'Month_WOE']                        | 0.7354              |

โดยจากผลการทดสอบข้างต้น พบว่า ภาพรวมของผลการศึกษาจากแบบจำลอง Gradient Boosting มีความแม่นยำ (R<sup>2</sup>) สูงที่สุด ตามด้วย Random Forest, GradientBoosting, K-neighbors, DecisionTree, MLP และ Ridge ทั้งนี้ ผู้วิจัยได้ทำการตรวจสอบความถูกต้องและความแม่นยำจากหลักเกณฑ์ที่ใช้ตรวจสอบแบบจำลอง (Performance Criteria) ที่เหลือด้วย RMSE และ MAPE โดยได้ผลการทดสอบจากตัวแปรที่เหมาะสมในแต่ละแบบจำลอง ดังนี้

เมื่อพิจารณาผลการทดสอบของแบบจำลองในกรณี Training Set และ Test Set ในตาราง 4.12 พบว่า ความแม่นยำของแบบจำลองของทั้ง 2 ส่วนมีความแม่นยำสูงที่สุดเมื่อทดสอบด้วยแบบจำลอง Gradient Boosting ซึ่งจากชุดข้อมูล Training Set ได้ R<sup>2</sup> = 0.9221 สูงกว่าแบบจำลอง Test Set ซึ่งได้ R<sup>2</sup> = 0.8930 ตามด้วยแบบจำลอง Random Forest และ แบบจำลอง Ridge มีประสิทธิภาพต่ำที่สุดเมื่อเทียบกับแบบจำลองชนิดอื่น

เมื่อพิจารณาผลการทดสอบของแบบจำลองที่ใช้ชุดข้อมูล PM 2.5 ในห้วงที่มีการแพร่ระบาดของ Covid-19 หรือเรียกว่า กรณี Testout Set ในตาราง 4.13 พบว่า ความแม่นยำของแบบจำลองจะมีความแม่นยำสูงที่สุดเมื่อทดสอบด้วยแบบจำลอง Gradient Boosting ซึ่งได้ผลการทดสอบ R<sup>2</sup> = 0.7998 ตามด้วยแบบจำลอง Random Forest และ แบบจำลอง Ridge มีประสิทธิภาพต่ำที่สุดเมื่อเทียบกับแบบจำลองชนิดอื่น

ตาราง 4.12 ผลการทดสอบความถูกต้องและความแม่นยำของแบบจำลอง Training set และ Test Set โดยอาศัยข้อมูลมูลค่า PM 2.5 เฉลี่ยรายเดือน (Lag by Month)

| Model                    | Training       |               |               | Test           |               |               |
|--------------------------|----------------|---------------|---------------|----------------|---------------|---------------|
|                          | R <sup>2</sup> | RMSE          | MAPE          | R <sup>2</sup> | RMSE          | MAPE          |
| Decision Tree            | 0.8765         | 6.4229        | 27.42%        | 0.8444         | 6.9565        | 28.69%        |
| <b>Gradient Boosting</b> | <b>0.9221</b>  | <b>5.3487</b> | <b>25.76%</b> | <b>0.8930</b>  | <b>6.4423</b> | <b>27.44%</b> |
| K-neighbors              | 0.8726         | 6.3165        | 25.95%        | 0.8504         | 6.8415        | 28.83%        |
| MLP                      | 0.8299         | 7.4407        | 32.73%        | 0.8292         | 7.4717        | 32.61%        |
| Random Forest            | 0.8973         | 5.7815        | 27.44%        | 0.8663         | 6.6112        | 29.65%        |
| Ridge                    | 0.7362         | 9.0880        | 36.52%        | 0.7316         | 9.1631        | 36.47%        |

ตาราง 4.13 ผลการทดสอบความถูกต้องและความแม่นยำของแบบจำลอง Testout (Covid-19 Pandemic) โดยอาศัยข้อมูลมูลค่า PM 2.5 เฉลี่ยรายเดือน (Lag by Month)

| Model                    | Testout (Covid-19 Pandemic) |               |               |
|--------------------------|-----------------------------|---------------|---------------|
|                          | R <sup>2</sup>              | RMSE          | MAPE          |
| Decision Tree            | 0.7505                      | 7.4635        | 35.38%        |
| <b>Gradient Boosting</b> | <b>0.7998</b>               | <b>7.7946</b> | <b>30.14%</b> |
| K-neighbors              | 0.7750                      | 7.5931        | 33.04%        |
| MLP                      | 0.7856                      | 7.6058        | 31.68%        |
| Random Forest            | 0.7988                      | 7.3680        | 32.01%        |
| Ridge                    | 0.7384                      | 8.1863        | 35.85%        |

#### 4.2.2 ข้อมูลค่า PM 2.5 เฉลี่ยรายวัน (Lag by Day)

โดยข้อมูลค่า PM 2.5 เฉลี่ยรายวัน (Lag by Day) คำนวณจากข้อมูล PM 2.5 โดยเฉลี่ยในวันทีมนับย้อนหลัง 1 ปีก่อนหน้า แยกคำนวณตามแต่ละสถานี และนำมาเป็น 1 ในตัวแปร (Feature) สำหรับทดสอบแบบจำลอง

จากการแบ่งกลุ่มตัวแปรต่างๆ ในบทที่ 3 โดยใช้ Random Forest Algorithm เพื่อกำจัดตัวแปรอิสระที่มีผลต่อตัวแปรตามน้อยที่สุดออกทีละตัว โดยการวิเคราะห์นี้ ผู้วิจัยได้เพิ่มตัวแปรซึ่งเป็นข้อมูลค่า PM 2.5 เฉลี่ยรายวัน (Lag by Day) และได้กลุ่มตัวแปรมาทั้งหมด 14 กลุ่ม นำไปทดสอบแบบจำลองทั้ง 3 แบบ (Training set, Test set, Testout set) โดยผลการทดสอบ จะใช้ R<sup>2</sup> ในการประเมินผลแบบจำลองของตัวแปรแต่ละกลุ่ม ซึ่งได้ผลการทดสอบดังแสดงในตาราง 4.14 แสดงผลการทดสอบแบบจำลองทั้ง 6 ชนิด

ตาราง 4.14 ผลการทดสอบจำนวนตัวแปร (Feature) ที่ใช้ในการสร้างแบบจำลอง กรณี Training set, Test set, Testout set ของแบบจำลอง ทั้ง 6 ชนิด เมื่อเพิ่มตัวแปร (Feature) ที่เป็นข้อมูลค่า PM 2.5 เฉลี่ยรายวัน (Lag by Day)

| Feature Group | R <sup>2</sup> of Decision Tree |        |        |                     | R <sup>2</sup> of Gradient Boosting |        |        |                     | R <sup>2</sup> of K-neighbors |        |        |                     | R <sup>2</sup> of MLP |        |        |                     | R <sup>2</sup> of Random Forest |        |        |                     | R <sup>2</sup> of Ridge |        |        |                     |
|---------------|---------------------------------|--------|--------|---------------------|-------------------------------------|--------|--------|---------------------|-------------------------------|--------|--------|---------------------|-----------------------|--------|--------|---------------------|---------------------------------|--------|--------|---------------------|-------------------------|--------|--------|---------------------|
|               | Train                           | Test   | Testo  | Avg. R <sup>2</sup> | Train                               | Test   | Testo  | Avg. R <sup>2</sup> | Train                         | Test   | Testo  | Avg. R <sup>2</sup> | Train                 | Test   | Testo  | Avg. R <sup>2</sup> | Train                           | Test   | Testo  | Avg. R <sup>2</sup> | Train                   | Test   | Testo  | Avg. R <sup>2</sup> |
| 1             | 0.7431                          | 0.7280 | 0.5827 | 0.6846              | 0.7522                              | 0.7264 | 0.5761 | 0.6849              | 0.7399                        | 0.7281 | 0.6095 | 0.6925              | 0.6735                | 0.6673 | 0.3098 | 0.5502              | 0.7435                          | 0.7302 | 0.6051 | 0.6929              | 0.6659                  | 0.6623 | 0.5482 | 0.6255              |
| 2             | 0.8100                          | 0.7885 | 0.7465 | 0.7817              | 0.8941                              | 0.8352 | 0.7270 | 0.8188              | 0.8135                        | 0.7930 | 0.6506 | 0.7524              | 0.7656                | 0.7664 | 0.1374 | 0.5564              | 0.8351                          | 0.8034 | 0.7734 | 0.8040              | 0.7077                  | 0.7026 | 0.7036 | 0.7046              |
| 3             | 0.8396                          | 0.8144 | 0.7481 | 0.8007              | 0.9333                              | 0.8644 | 0.7267 | 0.8415              | 0.8458                        | 0.8211 | 0.6434 | 0.7701              | 0.8187                | 0.8132 | 0.4022 | 0.6780              | 0.8658                          | 0.8313 | 0.7706 | 0.8226              | 0.7295                  | 0.7232 | 0.7090 | 0.7206              |
| 4             | 0.8600                          | 0.8363 | 0.7645 | <b>0.8203</b>       | 0.9448                              | 0.8778 | 0.7910 | <b>0.8712</b>       | 0.8715                        | 0.8422 | 0.7592 | <b>0.8243</b>       | 0.8309                | 0.8312 | 0.7718 | <b>0.8113</b>       | 0.8802                          | 0.8525 | 0.7881 | <b>0.8403</b>       | 0.7355                  | 0.7312 | 0.7382 | <b>0.7350</b>       |
| 5             | 0.8459                          | 0.8246 | 0.7682 | 0.8129              | 0.9359                              | 0.8732 | 0.7574 | 0.8555              | 0.8667                        | 0.8302 | 0.7475 | 0.8148              | 0.8216                | 0.8224 | 0.7960 | 0.8133              | 0.8698                          | 0.8444 | 0.7779 | 0.8307              | 0.7247                  | 0.7268 | 0.7315 | 0.7277              |
| 6             | 0.8523                          | 0.8140 | 0.6705 | 0.7789              | 0.9510                              | 0.8699 | 0.7496 | 0.8568              | 0.8705                        | 0.8250 | 0.7442 | 0.8132              | 0.8312                | 0.8229 | 0.7993 | 0.8178              | 0.8901                          | 0.8457 | 0.7731 | 0.8363              | 0.7274                  | 0.7269 | 0.7380 | 0.7307              |
| 7             | 0.8469                          | 0.8111 | 0.6801 | 0.7794              | 0.9267                              | 0.8605 | 0.7564 | 0.8479              | 0.8724                        | 0.8167 | 0.7119 | 0.8003              | 0.8060                | 0.8109 | 0.7607 | 0.7926              | 0.8849                          | 0.8450 | 0.7622 | 0.8307              | 0.7102                  | 0.7073 | 0.7216 | 0.7130              |
| 8             | 0.8454                          | 0.8051 | 0.6651 | 0.7719              | 0.9658                              | 0.8735 | 0.7217 | 0.8536              | 0.8874                        | 0.8168 | 0.6889 | 0.7977              | 0.8120                | 0.8065 | 0.7598 | 0.7928              | 0.8688                          | 0.8330 | 0.7307 | 0.8108              | 0.7148                  | 0.7210 | 0.7120 | 0.7159              |
| 9             | 0.7432                          | 0.6333 | 0.5704 | 0.6489              | 0.9868                              | 0.7662 | 0.6073 | 0.7867              | 0.8285                        | 0.7247 | 0.5931 | 0.7154              | 0.7274                | 0.6832 | 0.6799 | 0.6968              | 0.8367                          | 0.7186 | 0.6457 | 0.7337              | 0.6176                  | 0.5826 | 0.6423 | 0.6142              |
| 10            | 0.7862                          | 0.6659 | 0.4974 | 0.6499              | 0.9642                              | 0.7723 | 0.5953 | 0.7772              | 0.8372                        | 0.7341 | 0.5243 | 0.6986              | 0.7364                | 0.7122 | 0.5978 | 0.6821              | 0.8626                          | 0.7334 | 0.6239 | 0.7399              | 0.6143                  | 0.6141 | 0.6332 | 0.6205              |
| 11            | 0.7310                          | 0.6764 | 0.6329 | 0.6801              | 0.9495                              | 0.7894 | 0.6313 | 0.7901              | 0.8081                        | 0.7392 | 0.5724 | 0.7065              | 0.7458                | 0.7197 | 0.6079 | 0.6911              | 0.8298                          | 0.7486 | 0.6927 | 0.7571              | 0.6233                  | 0.6221 | 0.6411 | 0.6289              |
| 12            | 0.7527                          | 0.6560 | 0.4334 | 0.6140              | 0.9703                              | 0.7716 | 0.5866 | 0.7761              | 0.8489                        | 0.7328 | 0.5261 | 0.7026              | 0.7509                | 0.6986 | 0.5766 | 0.6754              | 0.8360                          | 0.7148 | 0.6084 | 0.7197              | 0.6326                  | 0.6142 | 0.6332 | 0.6267              |
| 13            | 0.7611                          | 0.6798 | 0.5898 | 0.6769              | 0.9455                              | 0.7829 | 0.6111 | 0.7798              | 0.8367                        | 0.7501 | 0.5472 | 0.7114              | 0.7685                | 0.7177 | 0.6040 | 0.6967              | 0.8768                          | 0.7509 | 0.6376 | 0.7551              | 0.6213                  | 0.6187 | 0.6111 | 0.6171              |
| 14            | 0.7184                          | 0.7046 | 0.6160 | 0.6797              | 0.9677                              | 0.8152 | 0.6044 | 0.7957              | 0.8348                        | 0.7489 | 0.5512 | 0.7116              | 0.7491                | 0.7346 | 0.5904 | 0.6914              | 0.8642                          | 0.7752 | 0.6075 | 0.7490              | 0.6195                  | 0.6199 | 0.6154 | 0.6182              |

หมายเหตุ : Train คือ Training set, Test คือ Test set และ Testo คือ testout-set data



จากตาราง 4.15 พบว่า แบบจำลอง DecisionTree ได้กลุ่มตัวแปรที่เหมาะสมที่สุดคือ กลุ่มตัวแปรที่ 4 ซึ่งประกอบไปด้วยทิศทางลม (Wind Direction) ฝุ่นหยาบ (PM 10) ฝุ่นละเอียด (PM 2.5) ย้อนหลัง และน้ำหนักรถของจำนวนเดือน (Month) ย้อนหลัง ซึ่งเป็นกลุ่มตัวแปรที่ดีที่สุดในการนำไปทดสอบแบบจำลองทั้ง 3 แบบ (Training set, Test set, Testout set) มีค่า  $R^2$  เฉลี่ยของทั้ง 3 แบบจำลองเท่ากับ 0.8203

แบบจำลอง GradientBoosting ได้กลุ่มตัวแปรที่เหมาะสมที่สุดคือ กลุ่มตัวแปรที่ 4 ซึ่งประกอบไปด้วยทิศทางลม (Wind Direction) ฝุ่นหยาบ (PM 10) ฝุ่นละเอียด (PM 2.5) ย้อนหลัง และน้ำหนักรถของจำนวนเดือน (Month) ย้อนหลัง ซึ่งเป็นกลุ่มตัวแปรที่ดีที่สุดในการนำไปทดสอบแบบจำลองทั้ง 3 แบบ (Training set, Test set, Testout set) มีค่า  $R^2$  เฉลี่ยของทั้ง 3 แบบจำลองเท่ากับ 0.8712

แบบจำลอง K-neighbors ได้กลุ่มตัวแปรที่เหมาะสมที่สุดคือ กลุ่มตัวแปรที่ 4 ซึ่งประกอบไปด้วยทิศทางลม (Wind Direction) ฝุ่นหยาบ (PM 10) ฝุ่นละเอียด (PM 2.5) ย้อนหลัง และน้ำหนักรถของจำนวนเดือน (Month) ย้อนหลัง ซึ่งเป็นกลุ่มตัวแปรที่ดีที่สุดในการนำไปทดสอบแบบจำลองทั้ง 3 แบบ (Training set, Test set, Testout set) มีค่า  $R^2$  เฉลี่ยของทั้ง 3 แบบจำลองเท่ากับ 0.8243

แบบจำลอง MLP ได้กลุ่มตัวแปรที่เหมาะสมที่สุดคือ กลุ่มตัวแปรที่ 4 ซึ่งประกอบไปด้วยทิศทางลม (Wind Direction) ฝุ่นหยาบ (PM 10) ฝุ่นละเอียด (PM 2.5) ย้อนหลัง และน้ำหนักรถของจำนวนเดือน (Month) ย้อนหลัง ซึ่งเป็นกลุ่มตัวแปรที่ดีที่สุดในการนำไปทดสอบแบบจำลองทั้ง 3 แบบ (Training set, Test set, Testout set) มีค่า  $R^2$  เฉลี่ยของทั้ง 3 แบบจำลองเท่ากับ 0.8113

แบบจำลอง Random Forest ได้กลุ่มตัวแปรที่เหมาะสมที่สุดคือ กลุ่มตัวแปรที่ 4 ซึ่งประกอบไปด้วยทิศทางลม (Wind Direction) ฝุ่นหยาบ (PM 10) ฝุ่นละเอียด (PM 2.5) ย้อนหลัง และน้ำหนักรถของจำนวนเดือน (Month) ย้อนหลัง ซึ่งเป็นกลุ่มตัวแปรที่ดีที่สุดในการนำไปทดสอบแบบจำลองทั้ง 3 แบบ (Training set, Test set, Testout set) มีค่า  $R^2$  เฉลี่ยของทั้ง 3 แบบจำลองเท่ากับ 0.8403

แบบจำลอง Ridge ได้กลุ่มตัวแปรที่เหมาะสมที่สุดคือ กลุ่มตัวแปรที่ 4 ซึ่งประกอบไปด้วยทิศทางลม (Wind Direction) ฝุ่นหยาบ (PM 10) ฝุ่นละเอียด (PM 2.5) ย้อนหลัง และน้ำหนักรถของจำนวนเดือน (Month) ย้อนหลัง ซึ่งเป็นกลุ่มตัวแปรที่ดีที่สุดในการนำไปทดสอบแบบจำลองทั้ง 3 แบบ (Training set, Test set, Testout set) มีค่า  $R^2$  เฉลี่ยของทั้ง 3 แบบจำลองเท่ากับ 0.7350

ตาราง 4.15 สรุปผลการคัดเลือกตัวแปรสำหรับนำไปวิเคราะห์แบบจำลอง กรณีเพิ่มตัวแปรข้อมูล  
ค่า PM 2.5 เฉลี่ยรายวัน (Lag by Day)

| แบบจำลอง         | Feature<br>ที่เหมาะสม | ตัวแปรที่เกี่ยวข้อง  | Avg. R <sup>2</sup> |
|------------------|-----------------------|--|---------------------|
| DecisionTree     | 4                     | ['WD(Deg)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag',<br>'Month_WOE'] | 0.8203              |
| GradientBoosting | 4                     | ['WD(Deg)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag',<br>'Month_WOE'] | 0.8712              |
| K-neighbors      | 4                     | ['WD(Deg)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag',<br>'Month_WOE'] | 0.8243              |
| MLP              | 4                     | ['WD(Deg)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag',<br>'Month_WOE'] | 0.8113              |
| RandomForest     | 4                     | ['WD(Deg)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag',<br>'Month_WOE'] | 0.8403              |
| Ridge            | 4                     | ['WD(Deg)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag',<br>'Month_WOE'] | 0.7350              |

โดยจากผลการทดสอบข้างต้น พบว่า ภาพรวมของผลการศึกษาจากแบบจำลอง GradientBoosting มีความแม่นยำ (R<sup>2</sup>) สูงที่สุด ตามด้วย Random Forest, K-neighbors, DecisionTree, MLP, และ Ridge ทั้งนี้ ผู้วิจัยได้ทำการตรวจสอบความถูกต้องและความแม่นยำจากหลักเกณฑ์ที่ใช้ตรวจสอบแบบจำลอง (Performance Criteria) ที่เหลือด้วย RMSE และ MAPE โดยได้ผลการทดสอบจากตัวแปรที่เหมาะสมในแต่ละแบบจำลอง ดังนี้

เมื่อพิจารณาผลการทดสอบของแบบจำลองในกรณี Training Set และ Test Set ในตาราง 4.16 พบว่า ความแม่นยำของแบบจำลองของทั้ง 2 ส่วนมีความแม่นยำสูงที่สุดเมื่อทดสอบด้วยแบบจำลอง Gradient Boosting ซึ่งจากชุดข้อมูล Training Set ได้ R<sup>2</sup> = 0.9448 สูงกว่าแบบจำลอง Test Set ซึ่งได้ R<sup>2</sup> = 0.8778 ตามด้วยแบบจำลอง Random Forest และ แบบจำลอง Ridge มีประสิทธิภาพต่ำที่สุดเมื่อเทียบกับแบบจำลองชนิดอื่น

เมื่อพิจารณาผลการทดสอบของแบบจำลองที่ใช้ชุดข้อมูล PM 2.5 ในห้วงที่มีการแพร่ระบาดของ Covid-19 หรือเรียกว่า กรณี Testout Set ในตาราง 4.17 พบว่า ความแม่นยำของแบบจำลองจะมีความแม่นยำสูงที่สุดเมื่อทดสอบด้วยแบบจำลอง Gradient Boosting ซึ่งได้ผลการทดสอบ R<sup>2</sup> = 0.7910 ตามด้วยแบบจำลอง Random Forest และ แบบจำลอง Ridge มีประสิทธิภาพต่ำที่สุดเมื่อเทียบกับแบบจำลองชนิดอื่น

ตาราง 4.16 ผลการทดสอบความถูกต้องและความแม่นยำของแบบจำลอง Training set และ Test Set โดยอาศัยข้อมูลมูลค่า PM 2.5 เฉลี่ยรายวัน (Lag by Day)

| Model                    | Training       |               |               | Test           |               |               |
|--------------------------|----------------|---------------|---------------|----------------|---------------|---------------|
|                          | R <sup>2</sup> | RMSE          | MAPE          | R <sup>2</sup> | RMSE          | MAPE          |
| Decision Tree            | 0.8600         | 6.5859        | 28.18%        | 0.8363         | 7.1398        | 30.33%        |
| <b>Gradient Boosting</b> | <b>0.9448</b>  | <b>4.1346</b> | <b>17.80%</b> | <b>0.8778</b>  | <b>6.1684</b> | <b>26.24%</b> |
| K-neighbors              | 0.8715         | 6.3090        | 26.57%        | 0.8422         | 7.0105        | 29.18%        |
| MLP                      | 0.8309         | 7.2386        | 30.46%        | 0.8312         | 7.2501        | 31.07%        |
| Random Forest            | 0.8802         | 6.0928        | 26.91%        | 0.8525         | 6.7785        | 29.27%        |
| Ridge                    | 0.7355         | 9.0533        | 36.63%        | 0.7312         | 9.1508        | 37.09%        |

ตาราง 4.17 ผลการทดสอบความถูกต้องและความแม่นยำของแบบจำลอง Testout (Covid-19 Pandemic) โดยอาศัยข้อมูลมูลค่า PM 2.5 เฉลี่ยรายวัน (Lag by Day)

| Model                    | Testout (Covid-19 Pandemic) |               |               |
|--------------------------|-----------------------------|---------------|---------------|
|                          | R <sup>2</sup>              | RMSE          | MAPE          |
| Decision Tree            | 0.7645                      | 7.7771        | 31.79%        |
| <b>Gradient Boosting</b> | <b>0.7910</b>               | <b>7.6863</b> | <b>28.26%</b> |
| K-neighbors              | 0.7592                      | 7.8636        | 33.18%        |
| MLP                      | 0.7718                      | 7.1700        | 29.21%        |
| Random Forest            | 0.7881                      | 7.3777        | 30.27%        |
| Ridge                    | 0.7382                      | 8.1998        | 36.06%        |

#### 4.2.3 ข้อมูลค่า PM 2.5 เฉลี่ยรายชั่วโมง (Lag by Hour)

โดยข้อมูลค่า PM 2.5 เฉลี่ยรายชั่วโมง (Lag by Hour) คำนวณจากข้อมูล PM 2.5 โดยเฉลี่ย ในชั่วโมงของวันที่นับย้อนหลัง 1 ปีก่อนหน้า แยกคำนวณตามแต่ละสถานี และนำมาเป็น 1 ในตัวแปร (Feature) สำหรับทดสอบแบบจำลอง

จากการแบ่งกลุ่มตัวแปรต่างๆ ในบทที่ 3 โดยใช้ Random Forest Algorithm เพื่อกำจัดตัวแปรอิสระที่มีผลต่อตัวแปรตามน้อยที่สุดออกทีละตัว โดยการวิเคราะห์นี้ ผู้วิจัยได้เพิ่มตัวแปร ซึ่งเป็นข้อมูลค่า PM 2.5 เฉลี่ยรายชั่วโมง (Lag by Hour) และได้กลุ่มตัวแปรมาทั้งหมด 14 กลุ่ม นำไปทดสอบแบบจำลองทั้ง 3 แบบ (Training set, Test set, Testout set) โดยผลการทดสอบ จะใช้ R<sup>2</sup> ในการประเมินผลแบบจำลองของตัวแปรแต่ละกลุ่ม ซึ่งได้ผลการทดสอบดังแสดงในตาราง 4.18 แสดงผลการทดสอบแบบจำลองทั้ง 6 ชนิด

ตาราง 4.18 ผลการทดสอบจำนวนตัวแปร (Feature) ที่ใช้ในการสร้างแบบจำลอง กรณี Training set, Test set, Testout set ของแบบจำลอง ทั้ง 6 ชนิด เมื่อเพิ่มตัวแปร (Feature) ที่เป็นข้อมูลค่า PM 2.5 เฉลี่ยรายชั่วโมง (Lag by Hour)

| Feature Group | R <sup>2</sup> of Decision Tree |        |        |                     | R <sup>2</sup> of Gradient Boosting |        |        |                     | R <sup>2</sup> of K-neighbors |        |        |                     | R <sup>2</sup> of MLP |        |        |                     | R <sup>2</sup> of Random Forest |        |        |                     | R <sup>2</sup> of Ridge |        |        |                     |
|---------------|---------------------------------|--------|--------|---------------------|-------------------------------------|--------|--------|---------------------|-------------------------------|--------|--------|---------------------|-----------------------|--------|--------|---------------------|---------------------------------|--------|--------|---------------------|-------------------------|--------|--------|---------------------|
|               | Train                           | Test   | Testo  | Avg. R <sup>2</sup> | Train                               | Test   | Testo  | Avg. R <sup>2</sup> | Train                         | Test   | Testo  | Avg. R <sup>2</sup> | Train                 | Test   | Testo  | Avg. R <sup>2</sup> | Train                           | Test   | Testo  | Avg. R <sup>2</sup> | Train                   | Test   | Testo  | Avg. R <sup>2</sup> |
| 1             | 0.7412                          | 0.7329 | 0.5865 | 0.6869              | 0.7500                              | 0.7322 | 0.5796 | 0.6873              | 0.7387                        | 0.7336 | 0.608  | 0.6934              | 0.6902                | 0.6885 | 0.3833 | 0.5873              | 0.7424                          | 0.7353 | 0.608  | 0.6952              | 0.6657                  | 0.6635 | 0.5497 | 0.6263              |
| 2             | 0.8006                          | 0.7903 | 0.7644 | 0.7851              | 0.8137                              | 0.7961 | 0.7431 | 0.7843              | 0.7955                        | 0.7913 | 0.7407 | 0.7758              | 0.7837                | 0.7887 | 0.7008 | 0.7577              | 0.8015                          | 0.797  | 0.777  | 0.7918              | 0.704                   | 0.7065 | 0.728  | 0.7128              |
| 3             | 0.8317                          | 0.8101 | 0.7498 | 0.7972              | 0.8488                              | 0.8192 | 0.7379 | 0.8020              | 0.8298                        | 0.8106 | 0.7012 | 0.7805              | 0.799                 | 0.7964 | 0.6392 | 0.7449              | 0.8487                          | 0.8206 | 0.7542 | 0.8078              | 0.7277                  | 0.7189 | 0.7281 | 0.7249              |
| 4             | 0.8615                          | 0.8552 | 0.8129 | 0.8432              | 0.8765                              | 0.8674 | 0.8187 | 0.8542              | 0.872                         | 0.8506 | 0.8003 | 0.8410              | 0.8454                | 0.8476 | 0.8252 | 0.8394              | 0.878                           | 0.8655 | 0.8207 | 0.8547              | 0.7369                  | 0.7396 | 0.7155 | 0.7307              |
| 5             | 0.8641                          | 0.8401 | 0.8472 | 0.8505              | 0.8943                              | 0.8602 | 0.8583 | 0.8709              | 0.8603                        | 0.8383 | 0.8295 | <b>0.8427</b>       | 0.8416                | 0.8378 | 0.8595 | 0.8463              | 0.8895                          | 0.8567 | 0.8629 | 0.8697              | 0.7372                  | 0.7365 | 0.7623 | 0.7453              |
| 6             | 0.8673                          | 0.8400 | 0.8507 | <b>0.8527</b>       | 0.9533                              | 0.8790 | 0.8720 | <b>0.9014</b>       | 0.8715                        | 0.8369 | 0.8193 | 0.8426              | 0.853                 | 0.8453 | 0.8645 | <b>0.8543</b>       | 0.8947                          | 0.8578 | 0.8691 | <b>0.8739</b>       | 0.7375                  | 0.7391 | 0.7654 | <b>0.7473</b>       |
| 7             | 0.8412                          | 0.8107 | 0.7280 | 0.7933              | 0.9584                              | 0.8479 | 0.7636 | 0.8566              | 0.845                         | 0.8076 | 0.7206 | 0.7911              | 0.8066                | 0.8141 | 0.7744 | 0.7984              | 0.8849                          | 0.842  | 0.7736 | 0.8335              | 0.7084                  | 0.7156 | 0.7212 | 0.7151              |
| 8             | 0.8431                          | 0.8004 | 0.7223 | 0.7886              | 0.9613                              | 0.8523 | 0.7511 | 0.8549              | 0.8722                        | 0.8095 | 0.6979 | 0.7932              | 0.823                 | 0.8199 | 0.7774 | 0.8068              | 0.8753                          | 0.8406 | 0.7813 | 0.8324              | 0.7157                  | 0.7163 | 0.7233 | 0.7184              |
| 9             | 0.7288                          | 0.6729 | 0.4526 | 0.6181              | 0.9759                              | 0.7405 | 0.6292 | 0.7819              | 0.7982                        | 0.7122 | 0.5959 | 0.7021              | 0.7140                | 0.6870 | 0.6720 | 0.6910              | 0.8456                          | 0.7258 | 0.6563 | 0.7426              | 0.6115                  | 0.5849 | 0.6428 | 0.6131              |
| 10            | 0.7885                          | 0.7113 | 0.434  | 0.6446              | 0.9798                              | 0.7736 | 0.6053 | 0.7862              | 0.823                         | 0.7606 | 0.5362 | 0.7066              | 0.7129                | 0.7554 | 0.608  | 0.6921              | 0.8398                          | 0.771  | 0.6625 | 0.7578              | 0.5960                  | 0.6618 | 0.6419 | 0.6332              |
| 11            | 0.7492                          | 0.6539 | 0.5959 | 0.6663              | 0.9778                              | 0.7467 | 0.5712 | 0.7652              | 0.8377                        | 0.7203 | 0.5302 | 0.6961              | 0.7411                | 0.7253 | 0.6361 | 0.7008              | 0.8286                          | 0.7356 | 0.6524 | 0.7389              | 0.6183                  | 0.6256 | 0.646  | 0.6300              |
| 12            | 0.7279                          | 0.6379 | 0.4948 | 0.6202              | 0.967                               | 0.7400 | 0.5671 | 0.7580              | 0.842                         | 0.7223 | 0.5242 | 0.6962              | 0.7475                | 0.6977 | 0.577  | 0.6741              | 0.851                           | 0.7185 | 0.6222 | 0.7306              | 0.6267                  | 0.6124 | 0.6346 | 0.6246              |
| 13            | 0.8129                          | 0.6255 | 0.1494 | 0.5293              | 0.9793                              | 0.7661 | 0.5701 | 0.7718              | 0.8351                        | 0.7234 | 0.5551 | 0.7045              | 0.7486                | 0.7005 | 0.6133 | 0.6875              | 0.8348                          | 0.7134 | 0.6164 | 0.7215              | 0.6216                  | 0.6001 | 0.6206 | 0.6141              |
| 14            | 0.7051                          | 0.6745 | 0.6028 | 0.6608              | 0.9904                              | 0.7783 | 0.5739 | 0.7809              | 0.8343                        | 0.7273 | 0.5466 | 0.7027              | 0.7293                | 0.7165 | 0.5900 | 0.6786              | 0.8019                          | 0.7338 | 0.5965 | 0.7107              | 0.6124                  | 0.6358 | 0.6085 | 0.6189              |

หมายเหตุ : Train คือ Training set, Test คือ Test set และ Testo คือ testout-set data

จากตาราง 4.19 พบว่า แบบจำลอง DecisionTree ได้กลุ่มตัวแปรที่เหมาะสมที่สุดคือ กลุ่มตัวแปรที่ 6 ซึ่งประกอบไปด้วยทิศทางลม (Wind Direction) คาร์บอนมอนอกไซด์ (CO) ไนตริกออกไซด์ (NO) ฝุ่นหยาบ (PM 10) ฝุ่นละเอียด (PM 2.5) ย้อนหลัง และน้ำหนักรถของจำนวนเดือน (Month) ย้อนหลัง ซึ่งเป็นกลุ่มตัวแปรที่ดีที่สุดในการนำไปทดสอบแบบจำลองทั้ง 3 แบบ (Training set, Test set, Testout set) มีค่า  $R^2$  เฉลี่ยของทั้ง 3 แบบจำลองเท่ากับ 0.8527

แบบจำลอง GradientBoosting ได้กลุ่มตัวแปรที่เหมาะสมที่สุดคือ กลุ่มตัวแปรที่ 6 ซึ่งประกอบไปด้วยทิศทางลม (Wind Direction) คาร์บอนมอนอกไซด์ (CO) ไนตริกออกไซด์ (NO) ฝุ่นหยาบ (PM 10) ฝุ่นละเอียด (PM 2.5) ย้อนหลัง และน้ำหนักรถของจำนวนเดือน (Month) ย้อนหลัง ซึ่งเป็นกลุ่มตัวแปรที่ดีที่สุดในการนำไปทดสอบแบบจำลองทั้ง 3 แบบ (Training set, Test set, Testout set) มีค่า  $R^2$  เฉลี่ยของทั้ง 3 แบบจำลองเท่ากับ 0.9014

แบบจำลอง K-neighbors ได้กลุ่มตัวแปรที่เหมาะสมที่สุดคือ กลุ่มตัวแปรที่ 5 ซึ่งประกอบไปด้วยทิศทางลม (Wind Direction) คาร์บอนมอนอกไซด์ (CO) ฝุ่นหยาบ (PM 10) ฝุ่นละเอียด (PM 2.5) ย้อนหลัง และน้ำหนักรถของจำนวนเดือน (Month) ย้อนหลัง ซึ่งเป็นกลุ่มตัวแปรที่ดีที่สุดในการนำไปทดสอบแบบจำลองทั้ง 3 แบบ (Training set, Test set, Testout set) มีค่า  $R^2$  เฉลี่ยของทั้ง 3 แบบจำลองเท่ากับ 0.8427

แบบจำลอง MLP ได้กลุ่มตัวแปรที่เหมาะสมที่สุดคือ กลุ่มตัวแปรที่ 6 ซึ่งประกอบไปด้วยทิศทางลม (Wind Direction) คาร์บอนมอนอกไซด์ (CO) ไนตริกออกไซด์ (NO) ฝุ่นหยาบ (PM 10) ฝุ่นละเอียด (PM 2.5) ย้อนหลัง และน้ำหนักรถของจำนวนเดือน (Month) ย้อนหลัง ซึ่งเป็นกลุ่มตัวแปรที่ดีที่สุดในการนำไปทดสอบแบบจำลองทั้ง 3 แบบ (Training set, Test set, Testout set) มีค่า  $R^2$  เฉลี่ยของทั้ง 3 แบบจำลองเท่ากับ 0.8543

แบบจำลอง Random Forest ได้กลุ่มตัวแปรที่เหมาะสมที่สุดคือ กลุ่มตัวแปรที่ 6 ซึ่งประกอบไปด้วยทิศทางลม (Wind Direction) คาร์บอนมอนอกไซด์ (CO) ไนตริกออกไซด์ (NO) ฝุ่นหยาบ (PM 10) ฝุ่นละเอียด (PM 2.5) ย้อนหลัง และน้ำหนักรถของจำนวนเดือน (Month) ย้อนหลัง ซึ่งเป็นกลุ่มตัวแปรที่ดีที่สุดในการนำไปทดสอบแบบจำลองทั้ง 3 แบบ (Training set, Test set, Testout set) มีค่า  $R^2$  เฉลี่ยของทั้ง 3 แบบจำลองเท่ากับ 0.8739

แบบจำลอง Ridge ได้กลุ่มตัวแปรที่เหมาะสมที่สุดคือ กลุ่มตัวแปรที่ 6 ซึ่งประกอบไปด้วยทิศทางลม (Wind Direction) คาร์บอนมอนอกไซด์ (CO) ไนตริกออกไซด์ (NO) ฝุ่นหยาบ (PM 10) ฝุ่นละเอียด (PM 2.5) ย้อนหลัง และน้ำหนักรถของจำนวนเดือน (Month) ย้อนหลัง ซึ่งเป็นกลุ่มตัวแปรที่ดีที่สุดในการนำไปทดสอบแบบจำลองทั้ง 3 แบบ (Training set, Test set, Testout set) มีค่า  $R^2$  เฉลี่ยของทั้ง 3 แบบจำลองเท่ากับ 0.7473

ตาราง 4.19 สรุปผลการคัดเลือกตัวแปรสำหรับนำไปวิเคราะห์แบบจำลอง กรณีเพิ่มตัวแปรข้อมูลมูลค่า PM 2.5 เฉลี่ยรายชั่วโมง (Lag by Hour)

| แบบจำลอง         | Feature ที่เหมาะสม | ตัวแปรที่เกี่ยวข้อง   | Avg. R <sup>2</sup> |
|------------------|--------------------|---|---------------------|
| DecisionTree     | 6                  | ['WD(Deg)', 'CO(ppm)', 'NO(ppb)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag', 'Month_WOE'] | 0.8527              |
| GradientBoosting | 6                  | ['WD(Deg)', 'CO(ppm)', 'NO(ppb)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag', 'Month_WOE'] | 0.9014              |
| K-neighbors      | 5                  | ['WD(Deg)', 'CO(ppm)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag', 'Month_WOE']            | 0.8427              |
| MLP              | 6                  | ['WD(Deg)', 'CO(ppm)', 'NO(ppb)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag', 'Month_WOE'] | 0.8543              |
| RandomForest     | 6                  | ['WD(Deg)', 'CO(ppm)', 'NO(ppb)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag', 'Month_WOE'] | 0.8739              |
| Ridge            | 6                  | ['WD(Deg)', 'CO(ppm)', 'NO(ppb)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag', 'Month_WOE'] | 0.7473              |

โดยจากผลการทดสอบข้างต้น พบว่า ภาพรวมของผลการศึกษาจากแบบจำลอง GradientBoosting มีความแม่นยำ (R<sup>2</sup>) สูงที่สุด ตามด้วย RandomForest, MLP, DecisionTree, K-neighbors และ Ridge ทั้งนี้ ผู้วิจัยได้ทำการตรวจสอบความถูกต้องและความแม่นยำจากหลักเกณฑ์ที่ใช้ตรวจสอบแบบจำลอง (Performance Criteria) ที่เหลือด้วย RMSE และ MAPE โดยได้ผลการทดสอบจากตัวแปรที่เหมาะสมในแต่ละแบบจำลอง ดังนี้

เมื่อพิจารณาผลการทดสอบของแบบจำลองในกรณี Training Set และ Test Set ในตาราง 4.20 พบว่า ความแม่นยำของแบบจำลองของทั้ง 2 ส่วนมีความแม่นยำสูงที่สุดเมื่อทดสอบด้วยแบบจำลอง Gradient Boosting ซึ่งจากชุดข้อมูล Training Set ได้ R<sup>2</sup> = 0.9533 สูงกว่าแบบจำลอง Test Set ซึ่งได้ R<sup>2</sup> = 0.8790 ตามด้วยแบบจำลอง Random Forest และ แบบจำลอง Ridge มีประสิทธิภาพต่ำที่สุดเมื่อเทียบกับแบบจำลองชนิดอื่น

เมื่อพิจารณาผลการทดสอบของแบบจำลองที่ใช้ชุดข้อมูล PM 2.5 ในห้วงที่มีการแพร่ระบาดของ Covid-19 หรือเรียกว่า กรณี Testout Set ในตาราง 4.21 พบว่า ความแม่นยำของแบบจำลองจะมีความแม่นยำสูงที่สุดเมื่อทดสอบด้วยแบบจำลอง Gradient Boosting ซึ่งได้ผลการทดสอบ R<sup>2</sup> = 0.8720 ตามด้วยแบบจำลอง Random Forest และ แบบจำลอง Ridge มีประสิทธิภาพต่ำที่สุดเมื่อเทียบกับแบบจำลองชนิดอื่น

ตาราง 4.20 ผลการทดสอบความถูกต้องและความแม่นยำของแบบจำลอง Training set และ Test Set โดยอาศัยข้อมูลมูลค่า PM 2.5 เฉลี่ยรายชั่วโมง (Lag by Hour)

| Model                    | Training       |               |               | Test           |               |               |
|--------------------------|----------------|---------------|---------------|----------------|---------------|---------------|
|                          | R <sup>2</sup> | RMSE          | MAPE          | R <sup>2</sup> | RMSE          | MAPE          |
| Decision Tree            | 0.8673         | 6.3513        | 31.58%        | 0.8400         | 6.9425        | 32.99%        |
| <b>Gradient Boosting</b> | <b>0.9533</b>  | <b>4.1344</b> | <b>17.13%</b> | <b>0.8790</b>  | <b>6.1375</b> | <b>20.68%</b> |
| K-neighbors              | 0.8603         | 6.4414        | 30.33%        | 0.8383         | 6.9019        | 31.88%        |
| MLP                      | 0.8530         | 6.6861        | 32.20%        | 0.8453         | 6.8252        | 32.49%        |
| Random Forest            | 0.8947         | 5.6590        | 28.50%        | 0.8578         | 6.5434        | 31.55%        |
| Ridge                    | 0.7375         | 8.9346        | 39.69%        | 0.7391         | 8.8637        | 40.06%        |

ตาราง 4.21 ผลการทดสอบความถูกต้องและความแม่นยำของแบบจำลอง Testout (Covid-19 Pandemic) โดยอาศัยข้อมูลมูลค่า PM 2.5 เฉลี่ยรายชั่วโมง (Lag by Hour)

| Model                    | Testout (Covid-19 Pandemic) |               |               |
|--------------------------|-----------------------------|---------------|---------------|
|                          | R <sup>2</sup>              | RMSE          | MAPE          |
| Decision Tree            | 0.8507                      | 5.7591        | 33.45%        |
| <b>Gradient Boosting</b> | <b>0.8720</b>               | <b>5.5371</b> | <b>27.18%</b> |
| K-neighbors              | 0.8295                      | 6.0365        | 34.79%        |
| MLP                      | 0.8645                      | 5.4865        | 30.79%        |
| Random Forest            | 0.8691                      | 5.3928        | 30.87%        |
| Ridge                    | 0.7654                      | 7.2194        | 37.87%        |

### 4.3 ภาพรวมผลการทดสอบแบบจำลองในสถานการณ์ปกติ และเมื่อมีการแพร่ระบาดของ Covid-19

จากผลการวิเคราะห์แบบจำลองด้วยสถานการณ์ต่าง ๆ ทั้งกรณีสถานการณ์ปกติ และเมื่อมีการแพร่ระบาดของ Covid-19 ซึ่งผู้วิจัยได้ทำการวิเคราะห์โดยคำนึงถึงช่วงเวลาต่าง ๆ (Historical Data Feature) เฉลี่ยรายชั่วโมง (by Hour) เฉลี่ยรายวัน (by Day) และ เฉลี่ยรายเดือน (by Month) เพื่อให้เห็นถึงความแม่นยำของแบบจำลองต่าง ๆ โดยแบบจำลองได้ถูกตรวจสอบความถูกต้องและความแม่นยำจาก R<sup>2</sup> RMSE และ MAPE ซึ่งให้ผลการศึกษาของชุดข้อมูลสำหรับสร้างแบบจำลอง (Train Set Model) และผลการศึกษาของชุดข้อมูลที่ใช้ทดสอบแบบจำลอง (Test Set Model) โดยสรุปรายละเอียดดังตาราง 4.22 พบว่าแบบจำลอง Gradient Boosting เป็นแบบจำลองที่มีประสิทธิภาพและความแม่นยำสูงที่สุด โดยเมื่อนำข้อมูลเฉลี่ยรายชั่วโมง (by Hour) มาทำการวิเคราะห์แบบจำลอง

ตาราง 4.22 ผลการวิเคราะห์ Train Set และ Test Set ของแบบจำลองชนิดต่าง ๆ

| Model                    | Historical Data | Training       |               |               | Test           |               |               |
|--------------------------|-----------------|----------------|---------------|---------------|----------------|---------------|---------------|
|                          |                 | R <sup>2</sup> | RMSE          | MAPE          | R <sup>2</sup> | RMSE          | MAPE          |
| Decision Tree            | by hour         | 0.8673         | 6.3513        | 31.58%        | 0.8400         | 6.9425        | 32.99%        |
|                          | by day          | 0.8600         | 6.5859        | 28.18%        | 0.8363         | 7.1398        | 30.33%        |
|                          | by month        | 0.8765         | 6.4229        | 27.42%        | 0.8444         | 6.9565        | 28.69%        |
|                          | non             | 0.7759         | 8.4434        | 35.15%        | 0.7674         | 8.6631        | 35.34%        |
| <b>Gradient Boosting</b> | <b>by hour</b>  | <b>0.9533</b>  | <b>4.1344</b> | <b>17.13%</b> | <b>0.8790</b>  | <b>6.1375</b> | <b>20.68%</b> |
|                          | by day          | 0.9448         | 4.1346        | 17.80%        | 0.8778         | 6.1684        | 26.24%        |
|                          | by month        | 0.9221         | 5.3487        | 25.76%        | 0.8930         | 6.4423        | 27.44%        |
|                          | non             | 0.8688         | 6.5459        | 21.53%        | 0.8214         | 8.1684        | 26.79%        |
| K-neighbors              | by hour         | 0.8603         | 6.4414        | 30.33%        | 0.8383         | 6.9019        | 31.88%        |
|                          | by day          | 0.8715         | 6.3090        | 26.57%        | 0.8422         | 7.0105        | 29.18%        |
|                          | by month        | 0.8726         | 6.3165        | 25.95%        | 0.8504         | 6.8415        | 28.83%        |
|                          | non             | 0.7915         | 8.1444        | 33.40%        | 0.7768         | 8.4879        | 34.14%        |
| MLP                      | by hour         | 0.8530         | 6.6861        | 32.20%        | 0.8453         | 6.8252        | 32.49%        |
|                          | by day          | 0.8309         | 7.2386        | 30.46%        | 0.8312         | 7.2501        | 31.07%        |
|                          | by month        | 0.8299         | 7.4407        | 32.73%        | 0.8292         | 7.4717        | 32.61%        |
|                          | non             | 0.7645         | 8.6563        | 35.68%        | 0.7703         | 8.6089        | 35.11%        |
| Random Forest            | by hour         | 0.8947         | 5.6590        | 28.50%        | 0.8578         | 6.5434        | 31.55%        |
|                          | by day          | 0.8802         | 6.0928        | 26.91%        | 0.8525         | 6.7785        | 29.27%        |
|                          | by month        | 0.8973         | 5.7815        | 27.44%        | 0.8663         | 6.6112        | 29.65%        |
|                          | non             | 0.8335         | 7.3898        | 25.08%        | 0.7757         | 8.6443        | 28.52%        |
| Ridge                    | by hour         | 0.7375         | 8.9346        | 39.69%        | 0.7391         | 8.8637        | 40.06%        |
|                          | by day          | 0.7355         | 9.0533        | 36.63%        | 0.7312         | 9.1508        | 37.09%        |
|                          | by month        | 0.7362         | 9.0880        | 36.52%        | 0.7316         | 9.1631        | 36.47%        |
|                          | non             | 0.6883         | 9.9582        | 40.03%        | 0.6933         | 9.9480        | 39.63%        |



หลังจากนั้นผู้วิจัยได้นำแบบจำลองต่างๆ มาทดสอบกับชุดข้อมูลย้อนหลังในห้วงสถานการณ์การแพร่ระบาดของ Covid-19 ซึ่งอยู่ในช่วงปี พ.ศ. 2563-2564 โดยใช้ชื่อชุดข้อมูลว่า “Testout Set” โดยให้ผลการศึกษาของชุดข้อมูลสำหรับสร้างแบบจำลอง (Testout Set Model) สรุปได้ตาราง 4.23 ซึ่งพบว่า แบบจำลอง Gradient Boosting เป็นแบบจำลองที่มีประสิทธิภาพและความแม่นยำสูงที่สุด โดยเมื่อนำข้อมูลเฉลี่ยรายชั่วโมง (by Hour) มาทำการวิเคราะห์แบบจำลอง

ตาราง 4.23 ผลการวิเคราะห์ Testout Set ของแบบจำลอง ชนิดต่าง ๆ

| Model             | Historical Data | Testout (Covid-19 Pandemic) |        |        |
|-------------------|-----------------|-----------------------------|--------|--------|
|                   |                 | R <sup>2</sup>              | RMSE   | MAPE   |
| Decision Tree     | by hour         | 0.8507                      | 5.7591 | 33.45% |
|                   | by day          | 0.7645                      | 7.7771 | 31.79% |
|                   | by month        | 0.7505                      | 7.4635 | 35.38% |
|                   | non             | 0.7615                      | 7.6303 | 36.77% |
| Gradient Boosting | by hour         | 0.8720                      | 5.5371 | 27.18% |
|                   | by day          | 0.7910                      | 7.6863 | 28.26% |
|                   | by month        | 0.7998                      | 7.7946 | 30.14% |
|                   | non             | 0.7962                      | 7.0337 | 29.92% |
| K-neighbors       | by hour         | 0.8295                      | 6.0365 | 34.79% |
|                   | by day          | 0.7592                      | 7.8636 | 33.18% |
|                   | by month        | 0.7750                      | 7.5931 | 33.04% |
|                   | non             | 0.7515                      | 7.7893 | 35.85% |
| MLP               | by hour         | 0.8645                      | 5.4865 | 30.79% |
|                   | by day          | 0.7718                      | 7.1700 | 29.21% |
|                   | by month        | 0.7856                      | 7.6058 | 31.68% |
|                   | non             | 0.7505                      | 7.8567 | 31.73% |
| Random Forest     | by hour         | 0.8691                      | 5.3928 | 30.87% |
|                   | by day          | 0.7881                      | 7.3777 | 30.27% |
|                   | by month        | 0.7988                      | 7.3680 | 32.01% |
|                   | non             | 0.7384                      | 7.9845 | 34.85% |
| Ridge             | by hour         | 0.7654                      | 7.2194 | 37.87% |
|                   | by day          | 0.7382                      | 8.1998 | 36.06% |
|                   | by month        | 0.7384                      | 8.1863 | 35.85% |
|                   | non             | 0.7196                      | 8.2743 | 35.22% |

## บทที่ 5

### การอภิปรายผล (Discussion)

จากข้อมูลความเข้มข้นของ PM 2.5 รายเดือนระหว่างก่อนการระบาดของ Covid-19 ในพื้นที่กรุงเทพมหานครที่เป็นเมืองหลวงของประเทศไทย ซึ่งมีกิจกรรมต่าง ๆ มากมายเกิดขึ้น ทั้งทางด้านเศรษฐกิจ สังคม และการท่องเที่ยว เปรียบเทียบกับช่วงที่มีการแพร่ระบาดของ Covid-19 และมีการ Lockdown เมือง ทำให้กิจกรรมต่าง ๆ หายไป พบว่าค่าความเข้มข้นของ PM 2.5 ลดลงอย่างมีนัยสำคัญ เมื่อมีการแพร่ระบาดของ Covid-19 โดยลดลงเฉลี่ยถึงร้อยละ 21 ดังแสดงในตาราง 5.1

จากผลการเปรียบเทียบค่าความเข้มข้นของ PM 2.5 ระหว่างก่อนการระบาดของ Covid-19 และเมื่อมีการระบาดของ Covid-19 พบว่า ในช่วงเดือนพฤศจิกายน และธันวาคม ความเข้มข้นของ PM 2.5 ลดลงแต่น้อยกว่าเดือนอื่น เมื่อเทียบกับเดือนอื่นที่เป็น wet season เพราะสาเหตุที่ทำให้ PM 2.5 สูงจริง ๆ คือ ฤดูกาล แต่ในขณะที่การจราจรขนส่งก็เป็นปัจจัยสำคัญที่ทำให้เกิด PM 2.5 เช่นเดียวกัน จึงทำให้ค่า PM 2.5 ลดลงอย่างเห็นได้ชัด เมื่อมีการ Lockdown เกิดขึ้น

ตาราง 5.1 ค่าเฉลี่ยของ PM 2.5 รายเดือนเปรียบเทียบระหว่างก่อนการระบาดของ Covid-19 และเมื่อมีการแพร่ระบาดของ Covid-19

| Month     | Avg. PM 2.5 as Normal Situation (2015-2019) | Avg. PM 2.5 as Covid19 Pandemic (2020-2021) | %Diff.  |
|-----------|---|---|---------|
| เมษายน    | 27.26                                       | 22.76                                       | -16.49% |
| พฤษภาคม   | 21.79                                       | 17.11                                       | -21.47% |
| มิถุนายน  | 18.02                                       | 14.00                                       | -22.34% |
| กรกฎาคม   | 19.32                                       | 13.60                                       | -29.58% |
| สิงหาคม   | 18.16                                       | 14.00                                       | -22.91% |
| กันยายน   | 21.35                                       | 15.20                                       | -28.82% |
| ตุลาคม    | 27.15                                       | 20.39                                       | -24.90% |
| พฤศจิกายน | 33.23                                       | 27.27                                       | -17.93% |
| ธันวาคม   | 39.28                                       | 35.87                                       | -8.70%  |

หมายเหตุ : เริ่มตั้งแต่ เมษายน เนื่องจากเป็นเดือนที่เริ่มมีการ Lockdown เกิดขึ้น

## 5.1 แบบจำลองที่มีประสิทธิภาพและความแม่นยำสูงสุด

จากผลการวิเคราะห์แบบจำลองในบทที่ 4 พบว่าแบบจำลองที่อาศัยข้อมูลอนุกรมเวลาให้ผลการทดสอบที่มีความแม่นยำกว่าแบบจำลองที่ไม่ได้อาศัยข้อมูลด้านอนุกรมเวลา ซึ่งผลการทดสอบที่ให้ประสิทธิภาพสูงสุด คือผลการทดสอบที่อาศัยข้อมูล PM 2.5 เฉลี่ยรายชั่วโมง (by hour) ทั้งแบบจำลองที่ใช้ข้อมูล PM 2.5 ในห้วงสถานการณ์ปกติ และเมื่อมีการแพร่ระบาดของ Covid-19 ดังรายละเอียดในตาราง 5.2 พบว่า แบบจำลอง Gradient Boosting ให้ผลการทดสอบสูงสุด ทั้ง Train, Test และ Testout Set ตามด้วยแบบจำลอง Random Forest, Decision Tree, K-neighbors, MLP และแบบจำลอง Ridge ให้ผลการทดสอบต่ำที่สุด ทั้ง Train, Test และ Testout Set ตามลำดับ ซึ่งสอดคล้องกับผลการวิเคราะห์ประสิทธิภาพแบบจำลองชนิดต่าง ๆ ในงานวิจัยต่างประเทศ (Mahanta et al., 2019; Mahmud et al., 2022; Pan, 2018; Sobri et al., 2021) ที่ให้ผลการศึกษาล้ายคลึงกัน

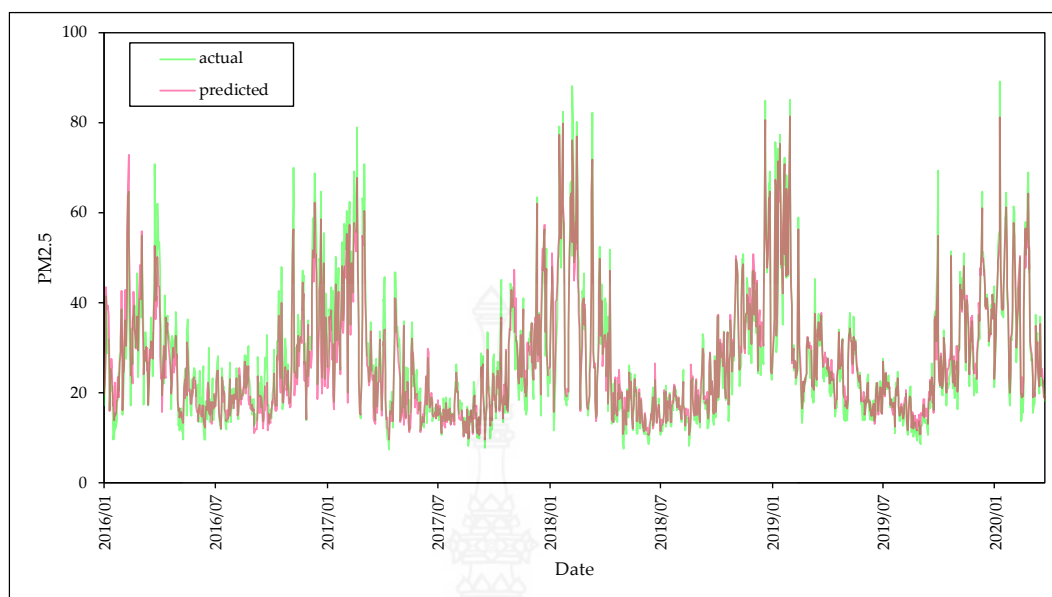
ผู้วิจัยยังได้อภิปรายให้เห็นรายละเอียดของผลการทดสอบแบบจำลองแต่ละชนิด โดยแสดงผลการเปรียบเทียบข้อมูลค่า PM 2.5 ที่สำรวจมา (Actual) กับข้อมูลค่า PM 2.5 ที่ได้มาจากการวิเคราะห์ในแบบจำลอง แสดงดังภาพ 5.1 – 5.18 โดยในแต่ละภาพ ได้วิเคราะห์และอภิปรายถึงการนำแบบจำลองแต่ละชนิดไปใช้งาน โดยเทียบผลการทดสอบประสิทธิภาพของแบบจำลอง ในแต่ละสถานการณ์ คือ Train set, Test Set และ Testout Set

ภาพ 5.1 แสดงผลการเปรียบเทียบค่าสถิติของ PM 2.5 และผลการคาดการณ์จากแบบจำลอง Discision Tree Regreesion โดยจากผลการวิเคราะห์ พบว่า ชุดข้อมูล Train set ( $R^2=0.8673$ ) มีความแม่นยำสูงสุด เมื่อเทียบกับชุดข้อมูล Test Set ( $R^2=0.8400$ ) ในภาพ 5.2 และ Testout Set ( $R^2=0.8507$ ) ในภาพ 5.3

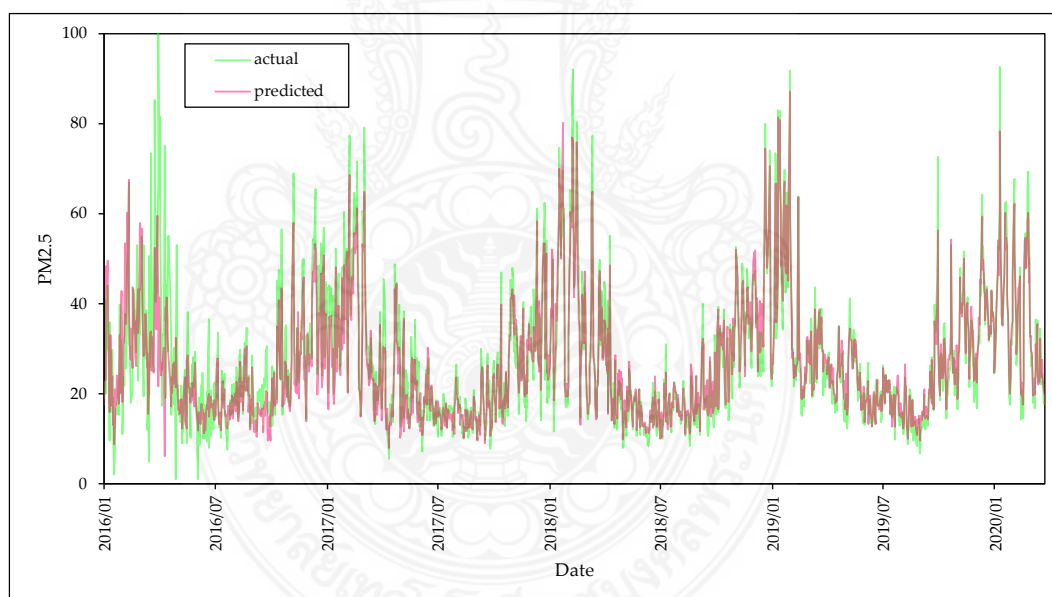
อย่างไรก็ตาม หากจะนำแบบจำลอง Discision Tree Regreesion มาใช้ในการวิเคราะห์ค่า PM 2.5 ควรเลือกใช้ชุดข้อมูลในช่วงก่อนการระบาดของ Covid-19 จะให้ผลการศึกษาค่าที่ดีที่สุด

**ตาราง 5.2** ผลการวิเคราะห์ประสิทธิภาพแบบจำลองชนิดต่าง ๆ ที่อาศัยข้อมูล PM 2.5 เฉลี่ยรายชั่วโมง (by hour)

| Model             | Training       |               |               | Test           |               |               | Testout        |               |               |
|-------------------|----------------|---------------|---------------|----------------|---------------|---------------|----------------|---------------|---------------|
|                   | R <sup>2</sup> | RMSE          | MAPE          | R <sup>2</sup> | RMSE          | MAPE          | R <sup>2</sup> | RMSE          | MAPE          |
| Decision Tree     | 0.8673         | 6.3513        | 31.58%        | 0.8400         | 6.9425        | 32.99%        | 0.8507         | 5.7591        | 33.45%        |
| Gradient Boosting | <b>0.9533</b>  | <b>4.1344</b> | <b>17.13%</b> | <b>0.8790</b>  | <b>6.1375</b> | <b>20.68%</b> | <b>0.8720</b>  | <b>5.5371</b> | <b>27.18%</b> |
| K-neighbors       | 0.8603         | 6.4414        | 30.33%        | 0.8383         | 6.9019        | 31.88%        | 0.8295         | 6.0365        | 34.79%        |
| MLP               | 0.8530         | 6.6861        | 32.20%        | 0.8453         | 6.8252        | 32.49%        | 0.8645         | 5.4865        | 30.79%        |
| Random Forest     | 0.8947         | 5.6590        | 28.50%        | 0.8578         | 6.5434        | 31.55%        | 0.8691         | 5.3928        | 30.87%        |
| Ridge             | 0.7375         | 8.9346        | 39.69%        | 0.7391         | 8.8637        | 40.06%        | 0.7654         | 7.2194        | 37.87%        |



ภาพ 5.1 ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง Discision Tree Regression กรณี Training set :  $R^2=0.8673$ ,  $RMSE=6.3513$ ,  $MAPE=31.58\%$

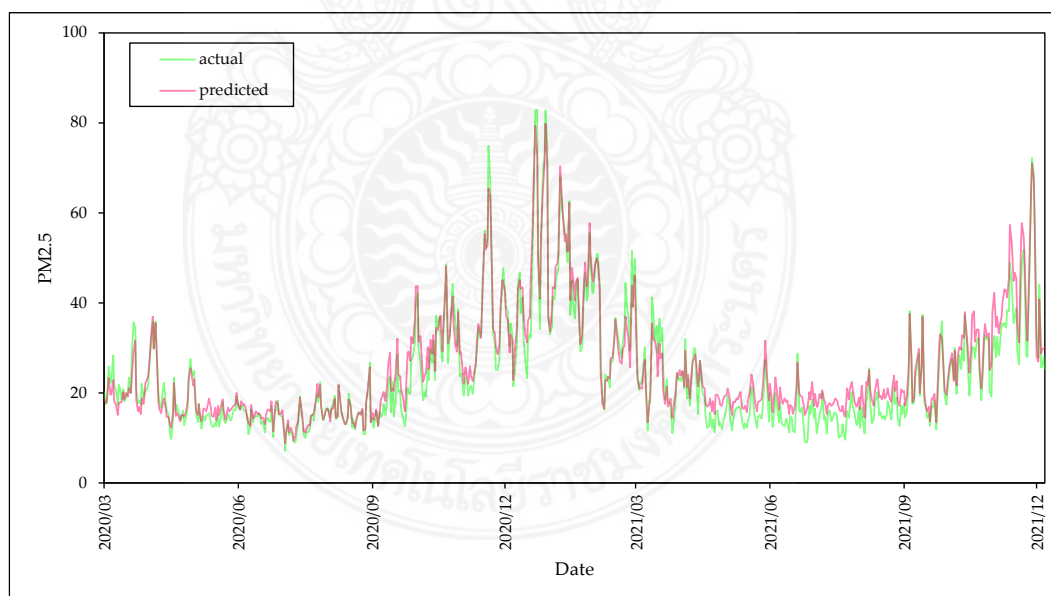


ภาพ 5.2 ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง Discision Tree Regression กรณี Test Set:  $R^2=0.8400$ ,  $RMSE=6.9425$ ,  $MAPE=32.99\%$

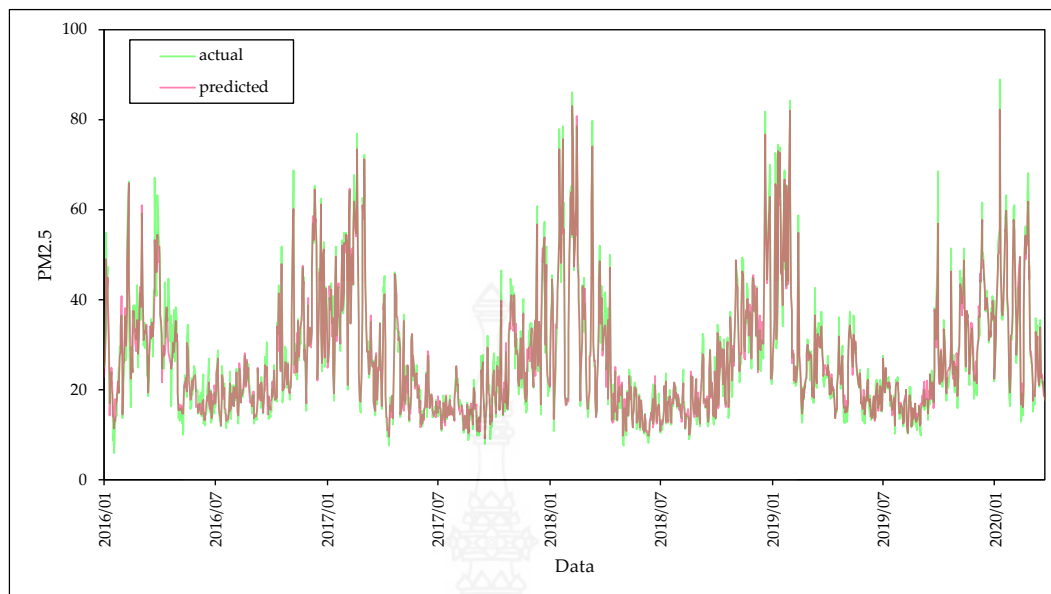
ภาพ 5.4 แสดงผลการเปรียบเทียบค่าสถิติของ PM 2.5 และผลการคาดการณ์จากแบบจำลอง Gradient boosting โดยจากผลการวิเคราะห์ พบว่า ชุดข้อมูล Train set ( $R^2=0.9533$ ) มีความแม่นยำสูงที่สุด เมื่อเทียบกับชุดข้อมูล Test Set ( $R^2=0.8790$ ) ในภาพ 5.5 และ Testout Set ( $R^2=0.8720$ ) ในภาพ 5.6 อย่างไรก็ตาม หากจะนำแบบจำลอง Gradient boosting มาใช้ในการทำนายการเกิด PM 2.5 เลือกใช้ชุดข้อมูลในช่วงก่อนการระบาดของ Covid-19 จะให้ผลการศึกษาค่าที่ดีที่สุด

เมื่อดูพฤติกรรมของค่า PM 2.5 เปรียบเทียบระหว่าง ข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง Gradient boosting กรณี Training set ในภาพ 5.4 พบว่า ในช่วงปี พ.ศ. 2558 แบบจำลองยังเคราะห์ข้อมูลออกมาได้ไม่แม่นยำ เมื่อเทียบกับช่วงเวลาอื่น แต่เมื่อแบบจำลองได้รับข้อมูลมากขึ้น คือตั้งแต่ปี พ.ศ. 2559 - 2562 พบว่า สามารถวิเคราะห์ข้อมูลออกมาได้แม่นยำมากขึ้น เมื่อเทียบกับสถิติที่เกิดขึ้นจริง และเช่นเดียวกันกับกรณี Test Set ภาพ 5.5 ซึ่งให้ผลการวิเคราะห์ไปในทิศทางเดียวกัน

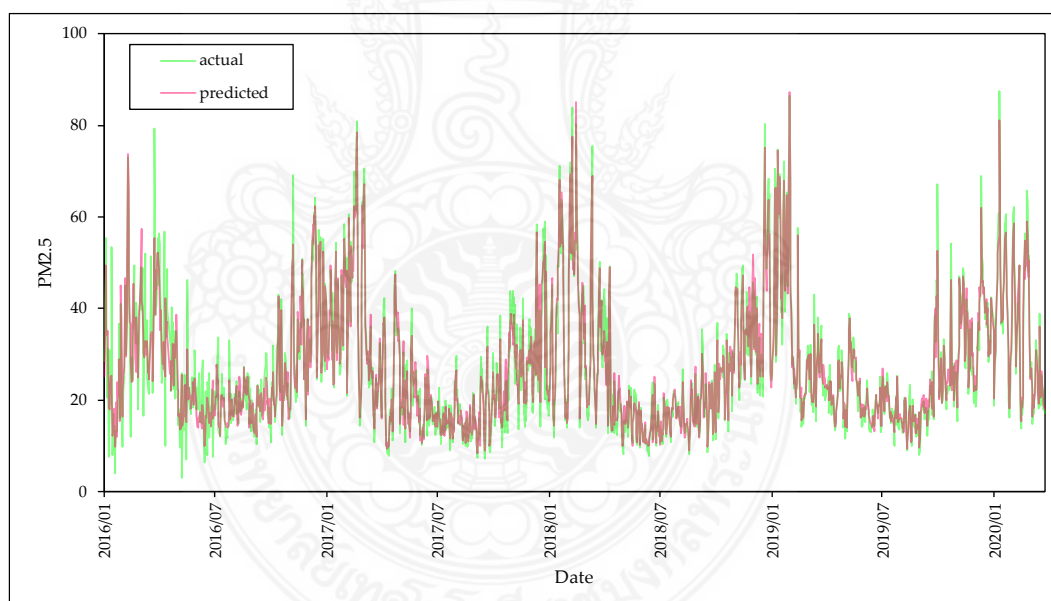
แต่เมื่อพิจารณาการวิเคราะห์แบบจำลองในกรณี Testout Set ในภาพ 5.6 กลับพบว่าแบบจำลองมีความแม่นยำน้อยกว่าข้อมูลจากสถิติจริง ตั้งแต่เริ่มวิเคราะห์ข้อมูลไปจนถึงชุดข้อมูลในปีสุดท้าย (พ.ศ.2562 - 2564) อาจเป็นเพราะชุดข้อมูลของ Testout Set พฤติกรรมของ PM 2.5 ในช่วงที่มีการแพร่ระบาดของ Covid-19 ซึ่งมีความแตกต่างจากชุดข้อมูลก่อนการระบาดของ รวมไปถึงชุดข้อมูลในช่วง Covid-19 เป็นข้อมูลในระยะเวลานั้น เพียงแค่ 2 ปี ซึ่งแตกต่างจากชุดข้อมูลในสถานการณ์ปกติ ที่มีปริมาณข้อมูลรวม 5 ปี



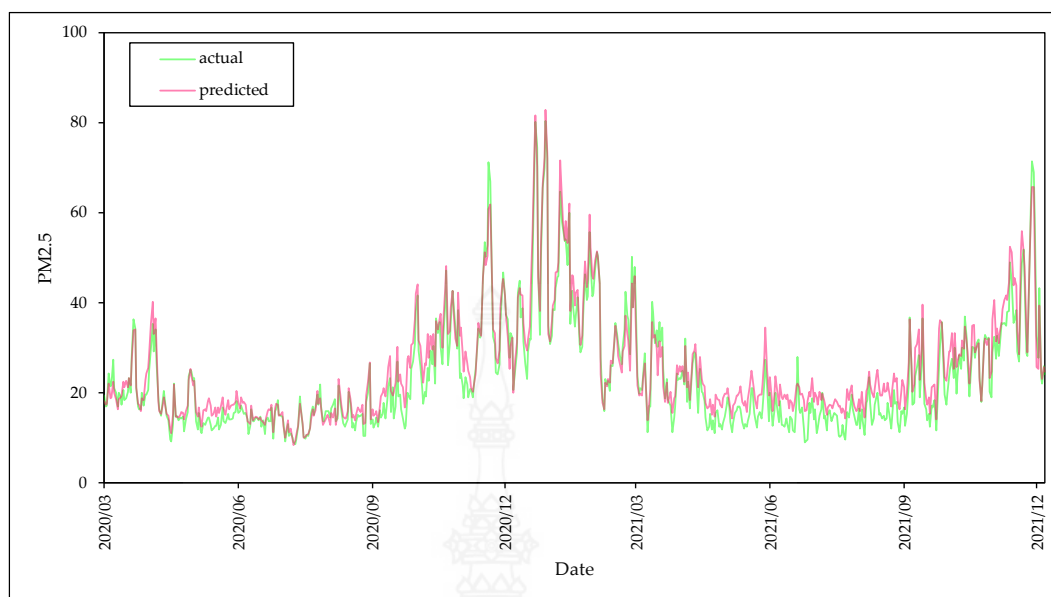
ภาพ 5.3 ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง Discision Tree Regression กรณี Testout Set :  $R^2=0.8507$ ,  $RMSE=6.3513$ ,  $MAPE=33.45\%$



ภาพ 5.4 ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง Gradient boosting กรณี Training set :  $R^2=0.9533$ ,  $RMSE=4.1344$ ,  $MAPE=17.13\%$



ภาพ 5.5 ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง Gradient boosting กรณี Test Set :  $R^2=0.8790$ ,  $RMSE=6.1375$ ,  $MAPE=20.68\%$

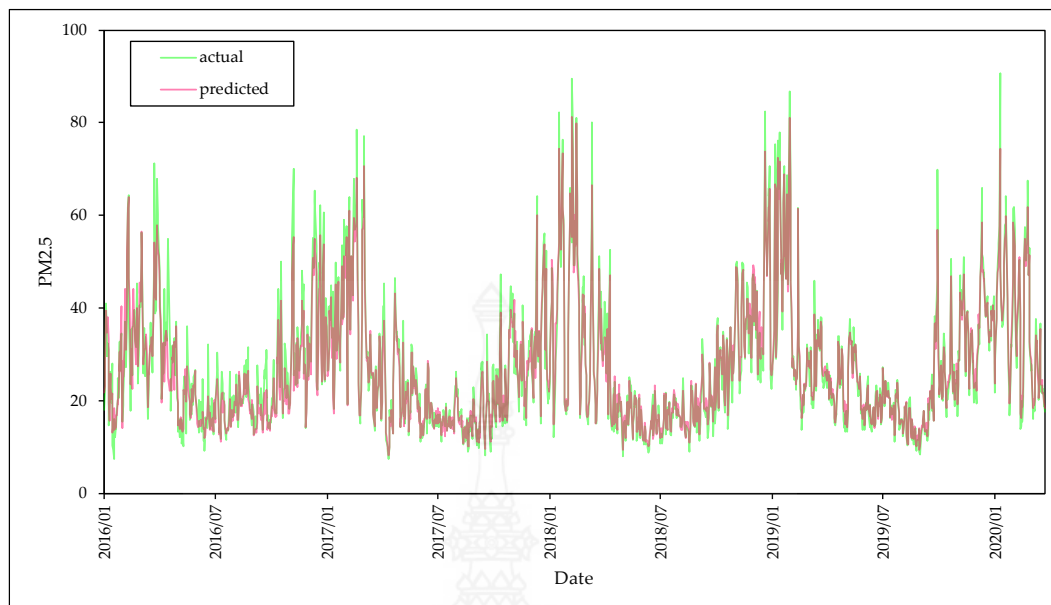


ภาพ 5.6 ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง Gradient boosting กรณี Testout Set:  $R^2=0.8720$ ,  $RMSE=5.1344$ ,  $MAPE=27.18\%$

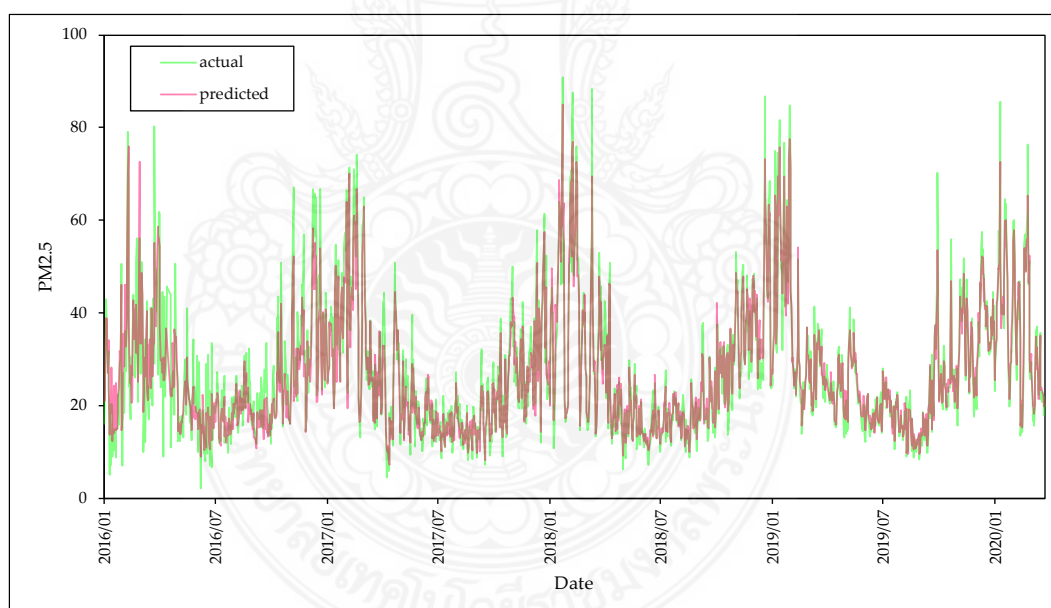
ภาพ 5.7 แสดงผลการเปรียบเทียบค่าสถิติของ PM 2.5 และผลการคาดการณ์จากแบบจำลอง K-nearest neighbor โดยจากผลการวิเคราะห์ พบว่า ชุดข้อมูล Train set ( $R^2=0.8603$ ) มีความแม่นยำสูงสุด เมื่อเทียบกับชุดข้อมูล Test Set ( $R^2=0.8383$ ) ในภาพ 5.8 และ Testout Set ( $R^2=0.8295$ ) ในภาพ 5.9 อย่างไรก็ตาม หากจะนำแบบจำลอง K-nearest neighbor มาใช้ในการทำนายการเกิด PM 2.5 เลือกใช้ชุดข้อมูลในช่วงก่อนการระบาดของ Covid-19 จะให้ผลการศึกษาค่าที่ดีที่สุด

เมื่อดูพฤติกรรมของค่า PM 2.5 ในภาพ 5.7 เปรียบเทียบระหว่าง ข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง K-nearest neighbor กรณี Training set พบว่า ในช่วงปี พ.ศ. 2558 - 2559 แบบจำลองยังวิเคราะห์ข้อมูลออกมาได้ไม่แม่นยำ เมื่อเทียบกับช่วงเวลาอื่น แต่เมื่อแบบจำลองได้รับข้อมูลมากขึ้น คือตั้งแต่ปี พ.ศ. 2559 - 2562 พบว่า แบบจำลองสามารถวิเคราะห์ข้อมูลออกมาได้แม่นยำมากขึ้นเมื่อเทียบกับสถิติที่เกิดขึ้นจริง และเช่นเดียวกันกับกรณี Test Set ในภาพ 5.8 ซึ่งให้ผลการวิเคราะห์ไปในทิศทางเดียวกัน

แต่เมื่อพิจารณาการวิเคราะห์แบบจำลองในกรณี Testout Set ในภาพ 5.9 กลับพบว่า แบบจำลองมีความแม่นยำน้อยกว่าข้อมูลจากสถิติจริงตั้งแต่เริ่มวิเคราะห์ข้อมูลไปจนถึงชุดข้อมูลในปีสุดท้าย (พ.ศ.2562 - 2564) โดยเฉพาะอย่างยิ่งในปี พ.ศ. 2564 อาจเป็นเพราะชุดข้อมูลของ Testout Set พฤติกรรมของ PM 2.5 ในช่วงที่มีการระบาดของ Covid-19 มีความแตกต่างจากชุดข้อมูลก่อนการระบาดของ รวมไปถึงชุดข้อมูลในห้วง Covid-19 เป็นข้อมูลในระยะเวลานั้น เพียงแค่ 2 ปี ซึ่งแตกต่างจากชุดข้อมูลในสถานการณ์ปกติ ที่มีปริมาณข้อมูลรวม 5 ปี

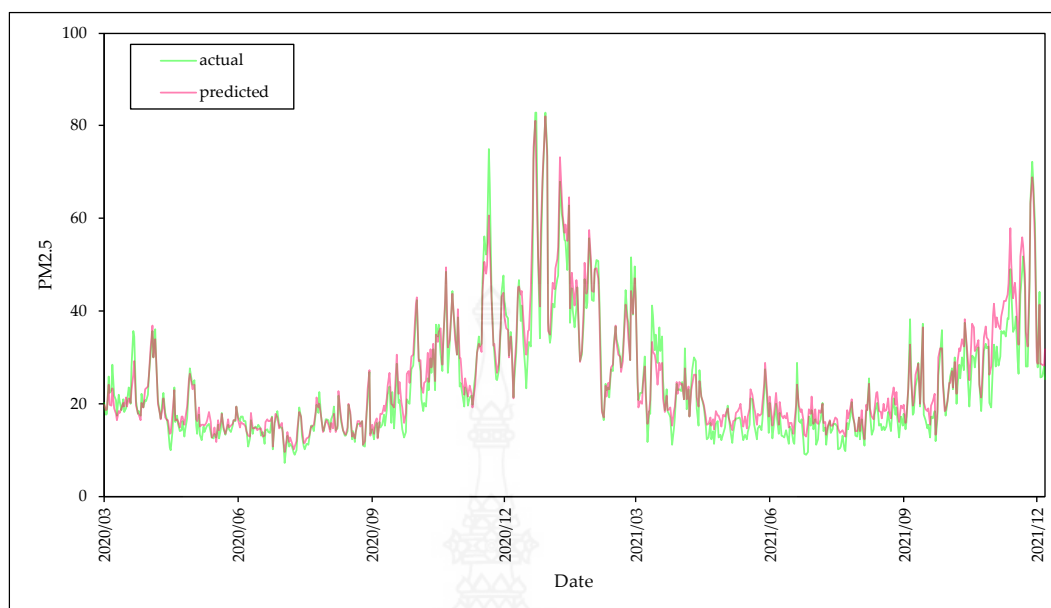


ภาพ 5.7 ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง K-nearest neighbor กรณี Training set :  $R^2=0.8603$ ,  $RMSE=6.4414$ ,  $MAPE=30.33\%$



ภาพ 5.8 ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง K-nearest neighbor กรณี Test Set :  $R^2=0.8383$ ,  $RMSE=6.9019$ ,  $MAPE=31.88\%$



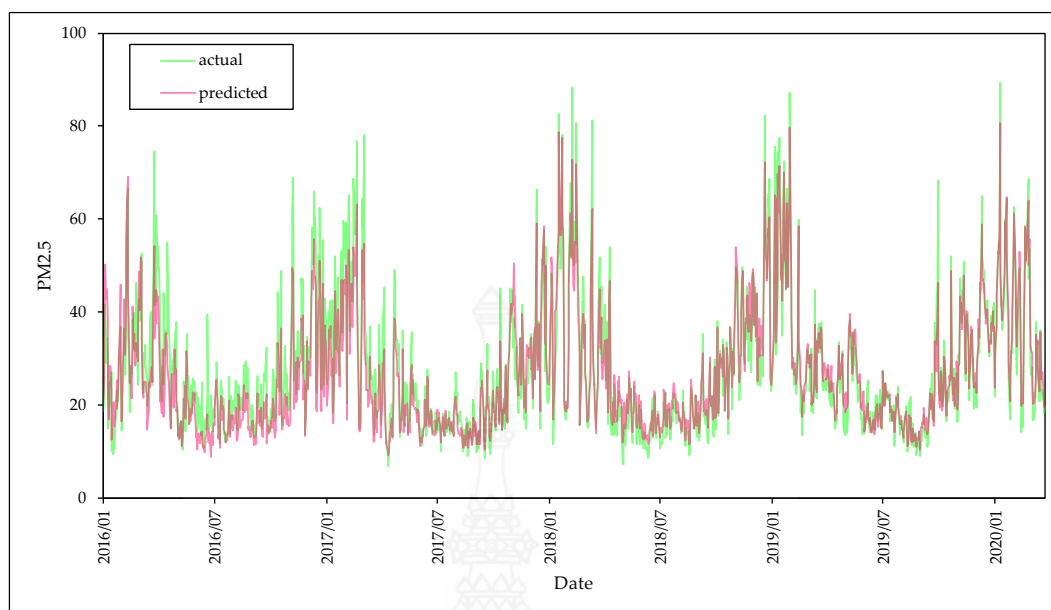


ภาพ 5.9 ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง K-nearest neighbor กรณี Testout Set :  $R^2=0.8295$ ,  $RMSE=6.4414$ ,  $MAPE=34.79\%$

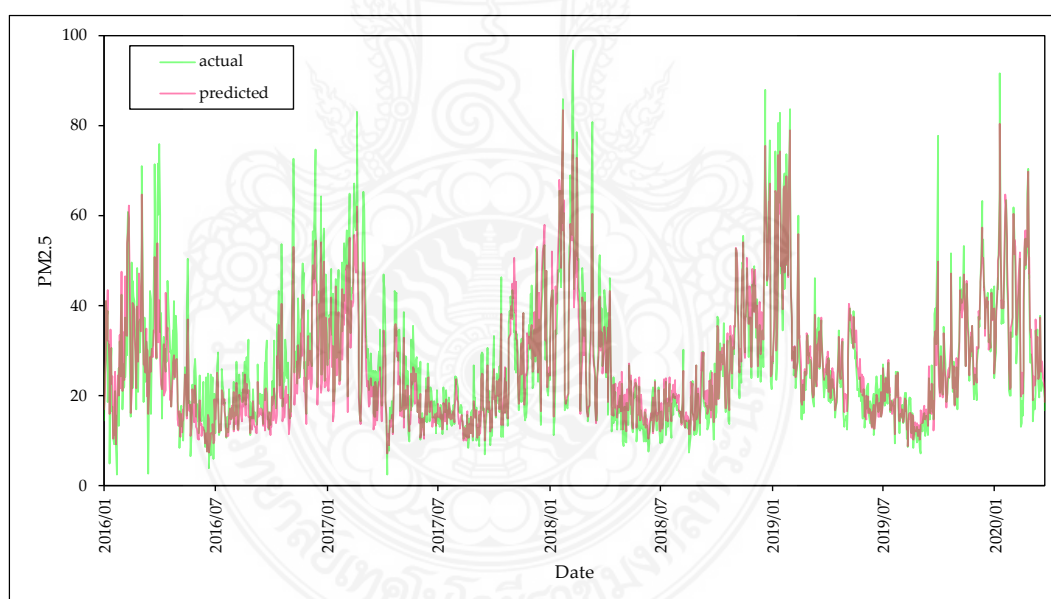
ภาพ 5.10 แสดงผลการเปรียบเทียบค่าสถิติของ PM 2.5 และผลการคาดการณ์จากแบบจำลอง Multi Layer Perception (MLP) โดยจากผลการวิเคราะห์ พบว่า ชุดข้อมูล Training set ( $R^2=0.8530$ ) มีความแม่นยำสูงที่สุด เมื่อเทียบกับชุดข้อมูล Test Set ( $R^2=0.8453$ ) ในภาพ 5.11 และ Testout Set ( $R^2=0.8645$ ) ในภาพ 5.12 อย่างไรก็ตาม หากจะนำแบบจำลอง Multi Layer Perception (MLP) มาใช้ในการทำนายการเกิด PM 2.5 เลือกใช้ชุดข้อมูลในช่วงก่อนการระบาดของ Covid-19 จะให้ผลการศึกษาคดีที่ดีที่สุด

เมื่อดูพฤติกรรมของค่า PM 2.5 ในภาพ 5.10 เปรียบเทียบระหว่าง ข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง Multi Layer Perception (MLP) กรณี Training set พบว่า ในช่วงปี พ.ศ. 2558 - 2559 แบบจำลองยังวิเคราะห์ข้อมูลออกมาได้ไม่แม่นยำ เมื่อเทียบกับช่วงเวลาอื่น แต่เมื่อแบบจำลองได้รับข้อมูลมากขึ้น คือตั้งแต่ปี พ.ศ. 2559 - 2562 พบว่า แบบจำลองสามารถวิเคราะห์ข้อมูลออกมาได้แม่นยำมากขึ้นเมื่อเทียบกับสถิติที่เกิดขึ้นจริง และเช่นเดียวกันกับกรณี Test Set ในภาพ 5.11 ซึ่งให้ผลการวิเคราะห์ไปในทิศทางเดียวกัน

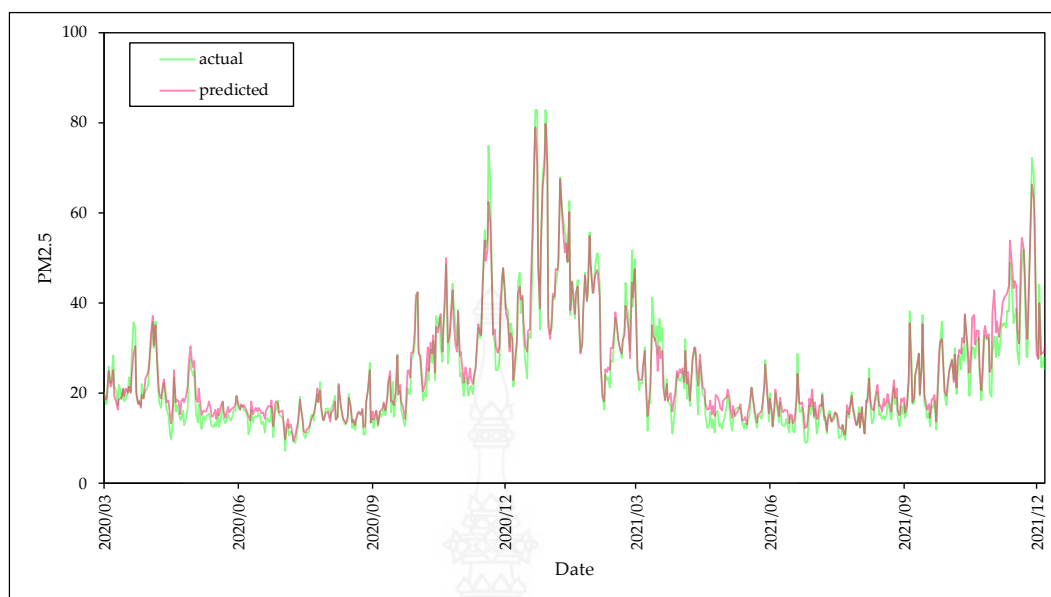
แต่เมื่อพิจารณาการวิเคราะห์แบบจำลองในกรณี Testout Set ในภาพ 5.12 กลับพบว่าแบบจำลองมีความแม่นยำน้อยกว่าข้อมูลจากสถิติจริงตั้งแต่ เริ่มวิเคราะห์ข้อมูลไปจนถึงชุดข้อมูลในปีสุดท้าย (พ.ศ.2562 - 2564) โดยเฉพาะอย่างยิ่งในปี พ.ศ. 2564 อาจเป็นเพราะชุดข้อมูลของ Testout Set พฤติกรรมของ PM 2.5 ในช่วงที่มีการระบาดของ Covid-19 มีความแตกต่างจากชุดข้อมูลก่อนการระบาดของ รวมไปถึงชุดข้อมูลในช่วง Covid-19 เป็นข้อมูลในระยะเวลานั้น เพียงแค่ 2 ปี ซึ่งแตกต่างจากชุดข้อมูลในสถานการณ์ปกติ ที่มีปริมาณข้อมูลรวม 5 ปี



ภาพ 5.10 ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง MLP  
กรณี Training set :  $R^2=0.8530$ ,  $RMSE=6.6861$ ,  $MAPE=32.20\%$



ภาพ 5.11 ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง MLP  
กรณี Test Set :  $R^2=0.8453$ ,  $RMSE=6.8252$ ,  $MAPE=32.49\%$

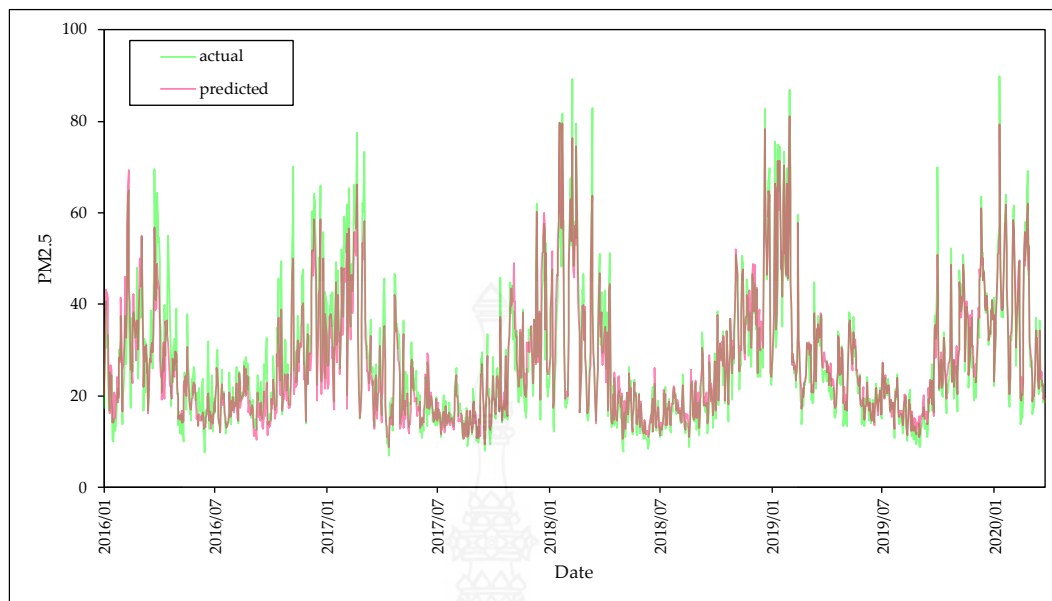


ภาพ 5.12 ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง MLP  
กรณี Testout Set :  $R^2=0.8645$ ,  $RMSE=6.6861$ ,  $MAPE=30.79\%$

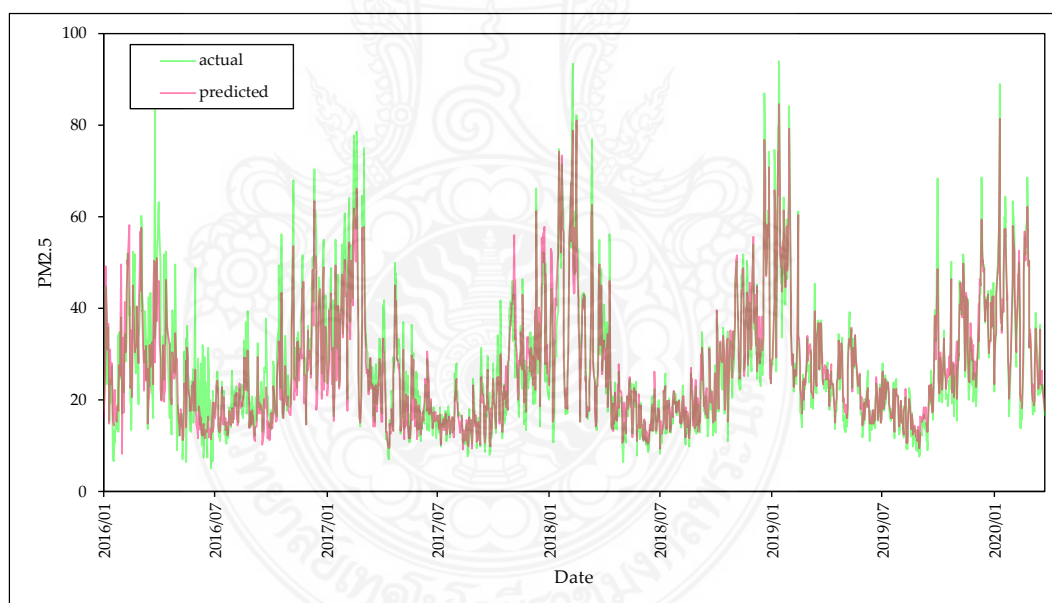
ภาพ 5.13 แสดงผลการเปรียบเทียบค่าสถิติของ PM 2.5 และผลการคาดการณ์จากแบบจำลอง Random Forest โดยจากผลการวิเคราะห์ พบว่า ชุดข้อมูล Training set ( $R^2=0.8947$ ) ในภาพ 5.14 มีความแม่นยำสูงที่สุด เมื่อเทียบกับชุดข้อมูล Test Set ( $R^2=0.8578$ ) และ Testout Set ( $R^2=0.8691$ ) ในภาพ 5.15 อย่างไรก็ตาม หากจะนำแบบจำลอง Random Forest มาใช้ในการทำนายการเกิด PM 2.5 เลือกใช้ชุดข้อมูลในช่วงก่อนการระบาดของ Covid-19 จะให้ผลการศึกษาค่าที่ดีที่สุด

เมื่อดูพฤติกรรมของค่า PM 2.5 ในภาพ 5.13 เปรียบเทียบระหว่าง ข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง Random Forest กรณี Training set พบว่า ในช่วงปี พ.ศ. 2558 - 2559 แบบจำลองยังวิเคราะห์ข้อมูลออกมาได้ไม่แม่นยำ เมื่อเทียบกับช่วงเวลาอื่น แต่เมื่อแบบจำลองได้รับข้อมูลมากขึ้น คือตั้งแต่ปี พ.ศ. 2559 - 2562 พบว่า แบบจำลองสามารถวิเคราะห์ข้อมูลออกมาได้แม่นยำมากขึ้นเมื่อเทียบกับสถิติที่เกิดขึ้นจริง และเช่นเดียวกันกับกรณี Test Set ในภาพ 5.14 ซึ่งให้ผลการวิเคราะห์ไปในทิศทางเดียวกัน

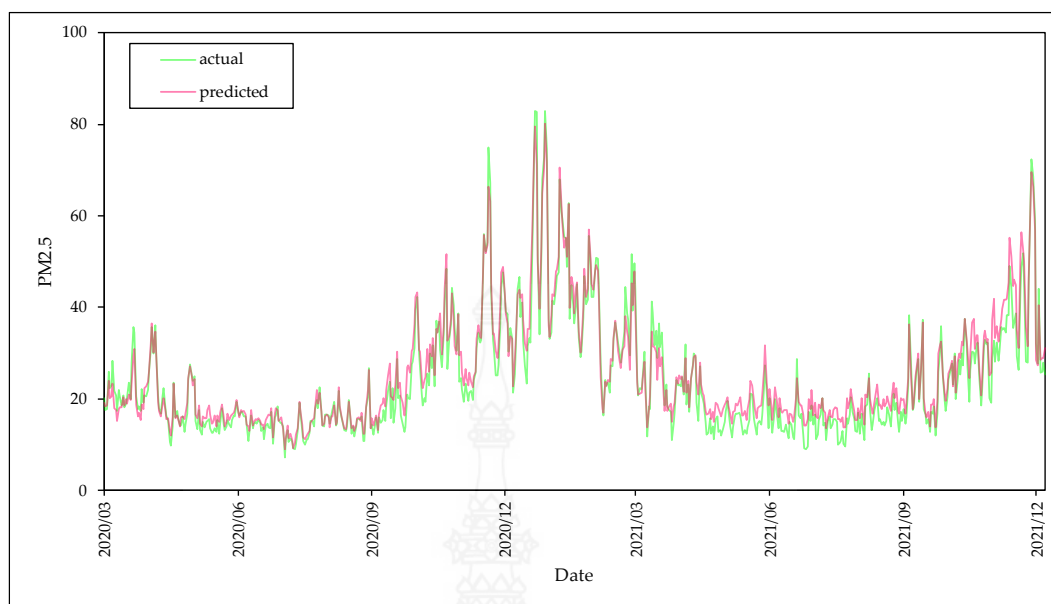
แต่เมื่อพิจารณาการวิเคราะห์แบบจำลองในกรณี Testout Set ในภาพ 5.15 กลับพบว่าแบบจำลองมีความแม่นยำน้อยกว่าข้อมูลจากสถิติจริงตั้งแต่ เริ่มวิเคราะห์ข้อมูลไปจนถึงชุดข้อมูลในปีสุดท้าย (พ.ศ.2562 - 2564) โดยเฉพาะอย่างยิ่งในปี พ.ศ. 2564 อาจเป็นเพราะชุดข้อมูลของ Testout Set พฤติกรรมของ PM 2.5 ในช่วงที่มีการระบาดของ Covid-19 มีความแตกต่างจากชุดข้อมูลก่อนการระบาดของ รวมไปถึงชุดข้อมูลในช่วง Covid-19 เป็นข้อมูลในระยะเวลานั้น เพียงแค่ 2 ปี ซึ่งแตกต่างจากชุดข้อมูลในสถานการณ์ปกติ ที่มีปริมาณข้อมูลรวม 5 ปี



ภาพ 5.13 ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง Random Forest  
กรณี Training set :  $R^2=0.8947$ ,  $RMSE=5.6590$ ,  $MAPE=28.50\%$



ภาพ 5.14 ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง Random Forest  
กรณี Test Set :  $R^2=0.8578$ ,  $RMSE=6.5434$ ,  $MAPE=31.55\%$

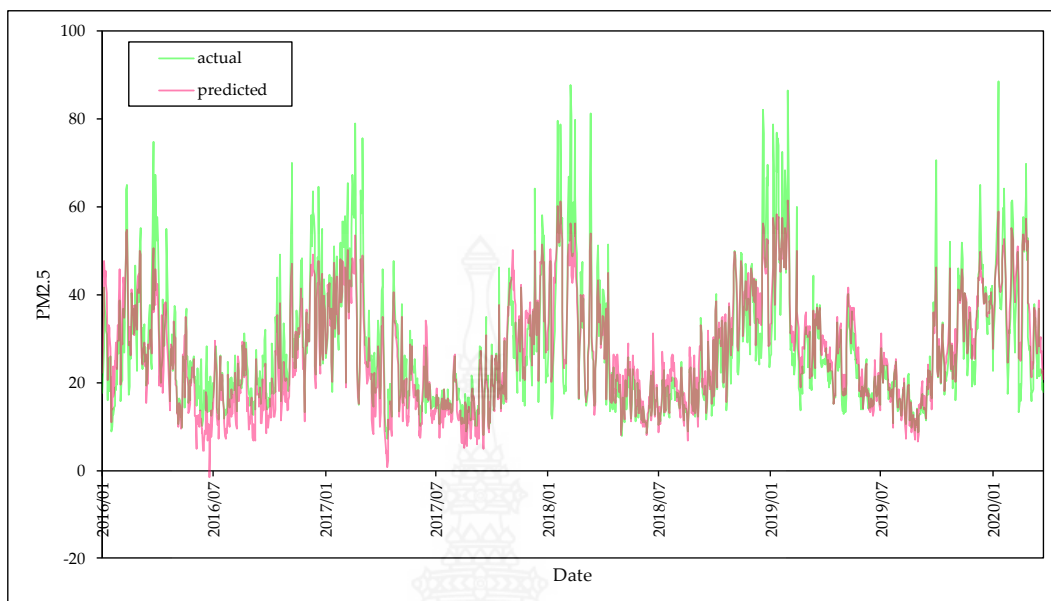


ภาพ 5.15 ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง Random Forest กรณี Testout Set :  $R^2=0.8691$ ,  $RMSE=5.6590$ ,  $MAPE=30.87\%$

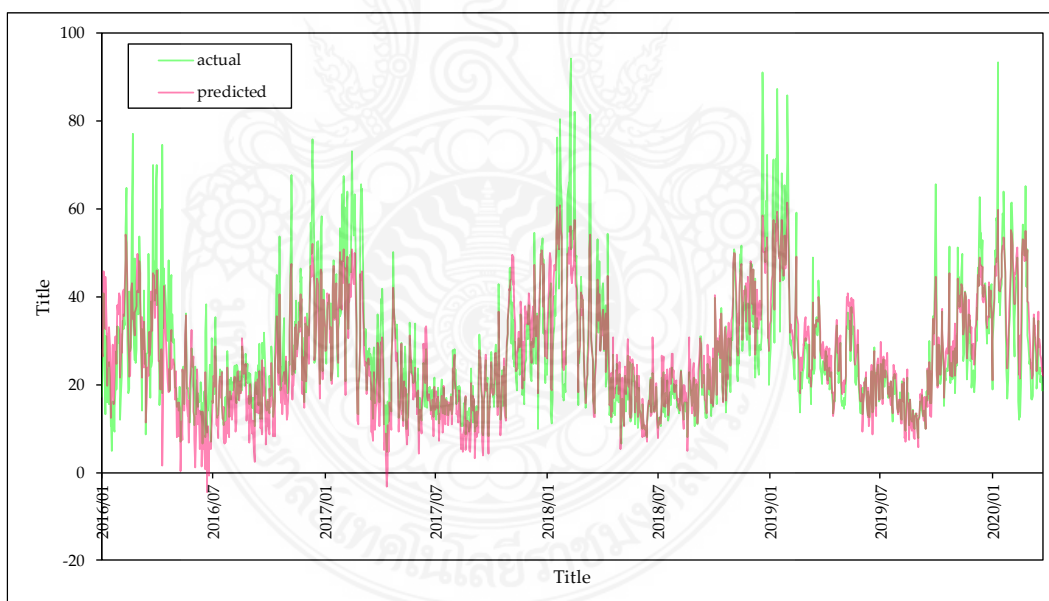
ภาพ 5.16 แสดงผลการเปรียบเทียบค่าสถิติของ PM 2.5 และผลการคาดการณ์จากแบบจำลอง Ridge Regression โดยจากผลการวิเคราะห์ พบว่า ชุดข้อมูล Training set ( $R^2=0.7375$ ) มีความแม่นยำสูงที่สุด เมื่อเทียบกับชุดข้อมูล Test Set ( $R^2=0.7391$ ) ในภาพ 5.17 และ Testout Set ( $R^2=0.7654$ ) ในภาพ 5.18 อย่างไรก็ตาม หากจะนำแบบจำลอง Ridge Regression มาใช้ในการทำนายการเกิด PM 2.5 เลือกใช้ชุดข้อมูลในช่วงการแพร่ระบาด Covid-19 จะให้ผลการศึกษาที่ดีที่สุด

เมื่อดูพฤติกรรมของค่า PM 2.5 ในภาพ 5.16 เปรียบเทียบระหว่าง ข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง Ridge Regression กรณี Training set พบว่า ในช่วงปี พ.ศ. 2558 - 2562 แบบจำลองยังวิเคราะห์ข้อมูลออกมาได้ไม่แม่นยำ เมื่อเทียบกับแบบจำลองชนิดอื่น โดยเฉพาะอย่างยิ่ง เมื่อค่า PM 2.5 มีปริมาณที่สูงในแต่ละปี โดยพบว่า แบบจำลอง Ridge Regression ให้ผลการศึกษาแบบค่าเฉลี่ย ไม่สอดคล้องกับพฤติกรรมจริงของ PM 2.5 ที่เกิดขึ้น และเช่นเดียวกันกับกรณี Test Set ในภาพ 5.17 ซึ่งให้ผลการวิเคราะห์ไปในทิศทางเดียวกัน

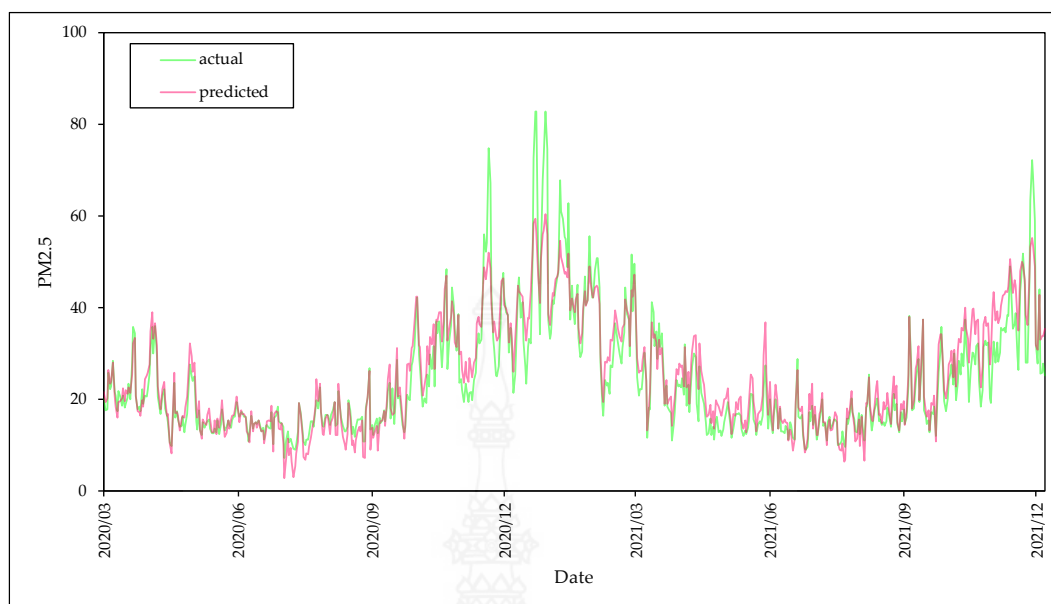
แต่เมื่อพิจารณาการวิเคราะห์แบบจำลองในกรณี Testout Set ในภาพ 5.18 กลับพบว่า แบบจำลองมีความแม่นยำน้อยกว่าข้อมูลจากสถิติจริงตั้งแต่เริ่มวิเคราะห์ข้อมูลไปจนถึงชุดข้อมูลในปีสุดท้าย (พ.ศ.2562 - 2564) โดยเฉพาะอย่างยิ่งในปี พ.ศ. 2564 อาจเป็นเพราะชุดข้อมูลของ Testout Set พฤติกรรมของ PM 2.5 ในช่วงที่มีการระบาดของ Covid-19 มีความแตกต่างจากชุดข้อมูลก่อนการระบาดของ รวมไปถึงชุดข้อมูลในช่วง Covid-19 เป็นข้อมูลในระยะเวลาสั้น เพียงแค่ 2 ปี ซึ่งแตกต่างจากชุดข้อมูลในสถานการณ์ปกติ ที่มีปริมาณข้อมูลรวม 5 ปี และสามารถวิเคราะห์ได้ว่าแบบจำลอง Ridge Regression ไม่เหมาะสมที่จะนำมาทำนายพฤติกรรมเกิด PM 2.5 ในอนาคต โดยการใช้สถิติชุดข้อมูลสะสม



ภาพ 5.16 ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง Ridge Regression กรณี Training Set :  $R^2=0.7375$ ,  $RMSE=8.9346$ ,  $MAPE=39.69\%$



ภาพ 5.17 ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง Ridge Regression กรณี Test Set :  $R^2=0.7391$ ,  $RMSE=8.8637$ ,  $MAPE=40.06\%$



ภาพ 5.18 ผลการเปรียบเทียบข้อมูลสถิติ PM 2.5 กับผลการวิเคราะห์แบบจำลอง Ridge Regression กรณี Testout Set :  $R^2=0.7654$ ,  $RMSE=8.9346$ ,  $MAPE=37.87\%$

เมื่อพิจารณาผลการทดสอบความแม่นยำของแบบจำลองจากการใช้ข้อมูล PM 2.5 ในห้วงสถานการณ์ปกติ และเมื่อมีการแพร่ระบาดของ Covid-19 ในตาราง 5.1 และเห็นหลักฐานเชิงประจักษ์ในภาพ 5.1-5.18 พบว่า แบบจำลองที่ใช้ชุดข้อมูลในห้วงสถานการณ์ Covid-19 ให้ผลการทดสอบ ไม่แตกต่างจากชุดข้อมูลในห้วงสถานการณ์ปกติ โดยผลการทดสอบจากแบบจำลอง Gradient Boosting ของทั้ง 2 ชุดข้อมูล มีความแตกต่างกัน (% difference) ไม่ถึงร้อยละ 10 ซึ่งอยู่ในเกณฑ์ที่รับได้โดยทั่วไป

ทั้งนี้ สามารถวิเคราะห์ได้ว่าชุดข้อมูล PM 2.5 ที่เกิดขึ้นในห้วงเวลาที่สถานการณ์ไม่ปกติ เช่น การเกิด Covid-19 ตลอดในห้วงหลายปีที่ผ่านมา ก็สามารถใช้แบบจำลองชนิดเดียวกันกับที่ใช้วิเคราะห์ชุดข้อมูล PM 2.5 ในสถานการณ์ปกติคาดการณ์การเกิด PM 2.5 ได้เช่นเดียวกัน ในกรณีใช้แบบจำลอง Gradient Boosting

## 5.2 แนวทางการนำแบบจำลองไปใช้กรณีมีข้อจำกัดเกี่ยวกับข้อมูล

จากการอภิปรายผลการศึกษาในข้างต้น พบว่าแบบจำลองที่มีประสิทธิภาพมากที่สุด ที่ใช้ในการวิเคราะห์ค่า PM 2.5 ทั้งในสถานการณ์ปกติ และสถานการณ์ไม่ปกติ (เกิดการแพร่ระบาดของโรค Covid-19) คือแบบจำลอง Gradient Boosting ซึ่งแบบจำลองนี้ ให้ผลการศึกษาที่ดีที่สุดในการศึกษาชุดข้อมูลในกรุงเทพมหานคร ประเทศไทย กลุ่มของตัวแปรที่เกี่ยวข้องกับการทดสอบแบบจำลอง Gradient Boosting คือกลุ่มตัวแปร (Feature) ที่ 6 ซึ่งประกอบไปด้วยทิศทางลม (Wind Direction) คาร์บอนมอนอกไซด์ (CO) ไนตริกออกไซด์ (NO) ฝุ่นหยาบ (PM 10) ฝุ่นละออง (PM 2.5) ย้อนหลัง

และน้ำหนักของจำนวนเดือน (Month) ย้อนหลัง โดยให้ผลการทดสอบประสิทธิภาพความแม่นยำ ( $R^2$ ) ทั้งการ train, test แบบจำลอง และรวมไปถึง Testout Set โดยพบว่า ค่า  $R^2$  เฉลี่ย 0.8801

ทั้งนี้ เมื่อพิจารณาผลการทดสอบประสิทธิภาพของแบบจำลอง Gradient Boosting ในกลุ่มตัวแปร (Feature) ที่มีจำนวนตัวแปรน้อยกว่าในกลุ่มที่ 6 พบว่า ตัวแปรที่สำคัญที่จะนำไปวิเคราะห์ค่า PM 2.5 ได้คือ ทิศทางลม (Wind Direction) ฝุ่นหยาบ (PM 10) และ ฝุ่นละเอียด (PM 2.5) ย้อนหลัง ก็ยังเพียงพอที่จะสามารถทำนายค่า PM 2.5 ได้ที่ความแม่นยำ  $R^2$  ประมาณ 0.75-0.80 ซึ่งความแม่นยำลดน้อยลงกว่าแบบจำลอง Gradient Boosting ที่ดีที่สุดเฉลี่ยร้อยละ 6.5 ในกรณีกลุ่มตัวแปรอยู่ที่ประมาณกลุ่มที่ 3-5

อย่างไรก็ตาม ถ้าต้องการให้แบบจำลอง Gradient Boosting มีความแม่นยำสูงที่สุด ตัวแปรที่เกี่ยวข้องที่นำไปทดสอบแบบจำลอง จำเป็นจะต้องมีชุดข้อมูล คาร์บอนมอนอกไซด์ (CO) ไนตริกออกไซด์ (NO) น้ำหนักของจำนวนเดือน (Month) ย้อนหลังร่วมอยู่ด้วย ดังแสดงในตาราง 5.3

ในขณะที่แบบจำลองชนิดอื่น ให้ผลการทดสอบที่ดีที่สุดอยู่ในกรณีที่มีตัวแปรกลุ่ม 3-6 โดยอย่างน้อยที่สุดการทดสอบแบบจำลองต่าง ๆ จะต้องประกอบไปด้วยชุดข้อมูลจากทิศทางลม (Wind Direction) ฝุ่นหยาบ (PM 10) และ ฝุ่นละเอียด (PM 2.5) ย้อนหลัง จึงจะให้ผลลัพธ์ของแบบจำลองมีความน่าเชื่อถือ ที่  $R^2$  มากกว่า 0.7000

**ตาราง 5.3** ตัวแปรที่เกี่ยวข้อง (Features) ที่ใช้ในการทดสอบแบบจำลอง Gradient Boosting ในกรณีชุดข้อมูลเฉลี่ยรายชั่วโมง (by hour)

| Feature Group | ตัวแปรที่เกี่ยวข้อง (Features)  | $R^2$ of Gradient Boosting |        |        |            |
|---------------|---|----------------------------|--------|--------|------------|
|               |   | Train                      | Test   | Testo  | Avg. $R^2$ |
| 1             | ['PM 10(ug/m3)']  | 0.7500                     | 0.7322 | 0.5796 | 0.6873     |
| 2             | ['PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag']   | 0.8137                     | 0.7961 | 0.7431 | 0.7843     |
| 3             | ['WD(Deg)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag']                                    | 0.8488                     | 0.8192 | 0.7379 | 0.8020     |
| 4             | ['WD(Deg)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag', 'Month_WOE']                       | 0.8765                     | 0.8674 | 0.8187 | 0.8542     |
| 5             | ['WD(Deg)', 'CO(ppm)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag', 'Month_WOE']            | 0.8943                     | 0.8602 | 0.8583 | 0.8709     |
| 6             | ['WD(Deg)', 'CO(ppm)', 'NO(ppb)', 'PM 10(ug/m3)', 'PM 2.5(ug/m3)_lag', 'Month_WOE'] | 0.9533                     | 0.8790 | 0.8720 | 0.9014     |

หมายเหตุ : Train คือ Training set, Test คือ Test set และ Testo คือ testout-set data



## บทที่ 6

### สรุปผลการศึกษา (Conclusion)

การศึกษานี้เป็นการศึกษาเกี่ยวกับข้อมูลมลพิษทางอากาศ โดยเฉพาะอย่างยิ่งความเข้มข้นของ PM 2.5 ที่เปลี่ยนแปลงไป ในช่วงที่เกิดการแพร่ระบาดของ Covid-19 ผู้วิจัยได้ใช้สถิติชุดข้อมูลค่ามลพิษต่าง ๆ ในช่วงปี พ.ศ. 2558 – 2564 ของกรมควบคุมมลพิษ และกรุงเทพมหานคร ทั้งหมด 63 สถานี มาใช้ในการดำเนินการศึกษาครั้งนี้

#### 6.1 สรุปผลการศึกษา

##### 6.1.1 พฤติกรรมมลพิษทางอากาศในพื้นที่กรุงเทพมหานคร (พ.ศ. 2558-2564)

จากผลการศึกษาข้อมูลมลพิษทางอากาศที่เกิดขึ้นตลอดระยะเวลา 7 ปี (พ.ศ. 2558 – 2564) พบว่า ในช่วงปี พ.ศ. 2558 – 2562 ซึ่งเป็นช่วงที่ยังไม่มีการแพร่ระบาดของ Covid-19 มลพิษทางอากาศแต่ละชนิดมีค่าความเข้มข้นใกล้เคียงกัน แต่ในปี พ.ศ. 2563-2564 พบว่าค่าความเข้มข้นของมลพิษทางอากาศมีการเปลี่ยนแปลงไปอย่างมีนัยสำคัญ โดย PM 2.5 NO<sub>x</sub> NO<sub>2</sub> NO และ CO ลดลงร้อยละ 14.38, 12.93, 18.43, 2.65 และ 5.66 ตามลำดับ ซึ่งมลพิษทางอากาศที่ลดลงมากที่สุดคือ NO<sub>2</sub> PM 2.5 NO<sub>x</sub> ตามลำดับ ทั้งนี้มลพิษทางอากาศทั้ง 3 ชนิดนี้ มีความเกี่ยวข้องกับการจราจรขนส่งเนื่องจากการเผาไหม้ที่ไม่สมบูรณ์ของเครื่องยนต์จะก่อให้เกิดมลพิษเหล่านี้ โดยเฉพาะอย่างยิ่งในเขตเมืองเช่นกรุงเทพมหานคร

อย่างไรก็ตาม ผู้วิจัยได้ทำการศึกษาพฤติกรรมการเกิดมลพิษทางอากาศตลอดช่วงเวลา 7 ปีย้อนหลัง (time series) เพื่อหารูปแบบของการกระจายตัวของมลพิษทางอากาศในกรุงเทพมหานครเมื่อระยะเวลาเปลี่ยนแปลงไป โดยผู้วิจัยได้ใช้การทดสอบด้วยวิธี Kruskal-wallis Analysis ซึ่งผลการทดสอบพบว่า การกระจายตัวของมลพิษทางอากาศตลอดระยะเวลา 7 ปี เป็นการกระจายตัวความเข้มข้นในรูปแบบฤดูกาล กล่าวคือ ค่าความเข้มข้นของมลพิษทางอากาศจะเพิ่มสูงขึ้นในช่วงหน้าหนาว และหน้าร้อน เพราะเป็นฤดูกาลที่อากาศแห้ง ตรงกันข้ามกับหน้าฝนซึ่งมีความชื้นในอากาศสูงจะทำให้ค่าความเข้มข้นของมลพิษทางอากาศลดลงในช่วงเวลานี้ ซึ่งเป็นพฤติกรรมเช่นนี้ตลอดทุกช่วงปี ทั้งก่อนการระบาดของ Covid-19 และขณะที่มีการแพร่ระบาดของ Covid-19

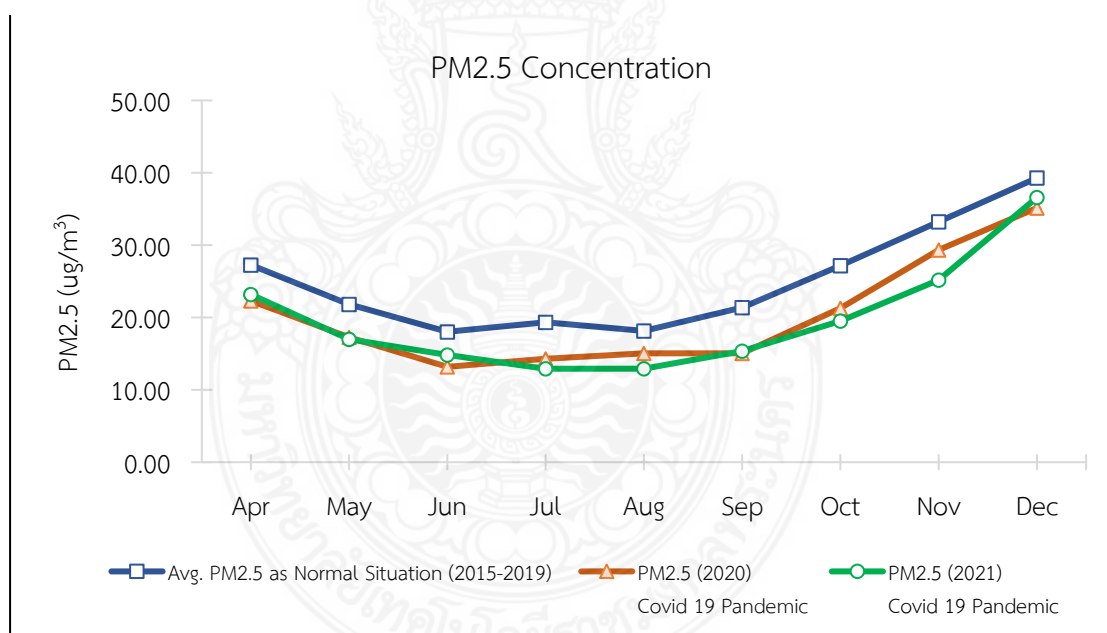
จากผลการศึกษาพบว่า ในห้วงของการแพร่ระบาดของ Covid-19 ซึ่งรัฐบาลไทยได้ประกาศให้หยุดทำงานอยู่กับบ้านส่งผลทำให้การเดินทางในพื้นที่กรุงเทพมหานครลดลง โดยปริมาณจราจรบนทางพิเศษลดลงถึงร้อยละ 40 ทั้งนี้ จากสถิติที่พบว่ามลพิษทางอากาศบางชนิดที่ความเข้มข้นลดลงไปเมื่อเกิดการแพร่ระบาดของ Covid-19 อย่างมีนัยสำคัญ ซึ่งสอดคล้องกับปริมาณจราจรในกรุงเทพมหานครที่ลดลง ผู้วิจัยจึงได้พิจารณาและเห็นว่าการเกิด PM 2.5 กับปริมาณการเดินทางใน

พื้นที่กรุงเทพมหานครที่เปลี่ยนแปลงไปมีความเกี่ยวข้องกันอย่างมีนัยสำคัญ จึงได้เลือกนำค่าความเข้มข้นของ PM 2.5 มาศึกษาและคาดการณ์การเกิดขึ้นต่อไป

### 6.1.2 พฤติกรรมการเกิด PM 2.5

จากผลการวิเคราะห์ในรายละเอียดของ PM 2.5 เฉลี่ยรายเดือนเปรียบเทียบกันระหว่างก่อนเกิด Covid-19 กับระหว่างการแพร่ระบาดของ Covid-19 พบว่า ในห้วงของการแพร่ระบาดของ Covid-19 ที่มีการ Lockdown นั้น PM 2.5 ลดลงจากเดิมถึงร้อยละ 21 นั้นหมายความว่า ในห้วงของการแพร่ระบาดของ Covid-19 ที่มีการ Lockdown เกิดขึ้น ส่งผลกระทบโดยตรงต่อกิจกรรมของมนุษย์ (Human behavior) โดยเฉพาะอย่างยิ่งในการเดินทางที่ลดลงถึงร้อยละ 40 และนั่นมีส่วนให้ PM 2.5 ลดลงอย่างมีนัยสำคัญ

แม้ว่าจะตอบไม่ได้ชัดเจนว่าปัจจัยใดระหว่างการเดินทาง หรือฤดูกาล ที่ส่งผลต่อความเข้มข้นของการเกิด PM 2.5 มากกว่ากัน แต่งานวิจัยชิ้นนี้ก็สามารถตอบได้ว่ากิจกรรมของมนุษย์ (Human Activity) ที่ลดลงในช่วงการแพร่ระบาดของ Covid-19 นั้น โดยเฉพาะอย่างยิ่งการเดินทางที่ลดลงนั้น ส่งผลต่อการเกิด PM 2.5 อย่างแท้จริง



ภาพ 6.1 เปรียบเทียบความเข้มข้นของ PM 2.5 รายเดือน ระหว่างก่อนและหลังการระบาดของ Covid-19 เนื่องจากการแพร่ระบาดของ Covid-19

### 6.1.3 การประยุกต์ใช้ Machine Learning techniques ในการวิเคราะห์ PM 2.5

ผู้วิจัยได้พัฒนาแบบจำลองคณิตศาสตร์ ด้วยการประยุกต์ใช้ Machine Learning techniques ในการคาดการณ์ PM 2.5 ที่เกิดขึ้นในอนาคต เพื่อสร้างแบบจำลองทางคณิตศาสตร์ที่ดีที่สุด ที่นำไปใช้ในการทำนายค่า PM 2.5 จากชุดข้อมูลทางสถิติที่มีอยู่ โดยได้แบ่งชุดข้อมูลออกเป็น 3 ชุดทดสอบซึ่งประกอบไปด้วย

- 1) Training set data คือ สถิติข้อมูล PM 2.5 ย้อนหลัง ตั้งแต่ พ.ศ. 2558-2562 ที่อยู่ในช่วงก่อนการระบาดของของ Covid-19 จำนวนร้อยละ 80
- 2) Test Set คือ สถิติข้อมูล PM 2.5 ย้อนหลัง ตั้งแต่ พ.ศ. 2558-2562 ที่อยู่ในช่วงก่อนการระบาดของของ Covid-19 จำนวนร้อยละ 20
- 3) Testout Set คือ สถิติข้อมูล PM 2.5 ย้อนหลัง ตั้งแต่ พ.ศ. 2563-2564 ที่อยู่ในช่วงก่อนการระบาดของของ Covid-19 จำนวนร้อยละ 20 ที่อยู่ในระหว่างการแพร่ระบาดของ Covid-19

ชุดข้อมูลต่าง ๆ ถูกนำมาวิเคราะห์คาดการณ์ด้วยแบบจำลอง 6 ชนิด ดังที่กล่าวมาในข้างต้น ได้แก่ Decision Tree, Gradient Boosting, K-neighbors, MLP, Random Forest, และ Ridge โดยแต่ละแบบจำลองจะทำการวิเคราะห์โดยคำนึงถึงช่วงเวลาต่าง ๆ (Historical Data Feature) ทุกๆ ชั่วโมง (by Hour) ทุกๆ วัน (by Day) และคำนึงถึงช่วงเวลาทุกๆ เดือน (by Month) ยังรวมไปถึงการวิเคราะห์แบบจำลอง โดยไม่คำนึงถึงช่วงเวลา (No Time) เพื่อให้เห็นถึงความแม่นยำของแบบจำลองต่าง ๆ

โดยจากผลการศึกษา พบว่าข้อมูลที่น่ามาวิเคราะห์นั้น คือ ข้อมูลช่วงเวลาที่ทำให้ความแม่นยำมากที่สุดคือ สถิติข้อมูลของ PM 2.5 รายชั่วโมง และแบบจำลองต่าง ๆ ถูกตรวจสอบความถูกต้องและความแม่นยำจากหลักเกณฑ์ที่ใช้ตรวจสอบ (Performance Criteria) คือ  $R^2$  RMSE และ MAE จากผลการวิเคราะห์ พบว่า แบบจำลอง Gradient Boosting, Random Forest, K-neighbors, Decision Tree, MLP, และ Ridge Model ซึ่ง Gradient Boosting by hour ให้ผลการศึกษาที่ดีที่สุดของ Train set โดยมีค่า  $R^2 = 0.9533$  และ Test Set ที่มีค่า  $R^2 = 0.8790$  ในขณะที่ทำการทดสอบด้วยชุดข้อมูล Testout Set Model ซึ่งเป็นชุดข้อมูลในช่วงสถานการณ์ Covid-19 Pandemic พบว่า Gradient Boosting โดยชุดข้อมูลที่มี Historical Data by Hour ให้ผลการวิเคราะห์ที่ดีที่สุดเช่นกัน โดยมีค่า  $R^2 = 0.8720$

จากตาราง 6.1 แสดงผลการวิเคราะห์  $R^2$  ของแต่ละแบบจำลอง ซึ่งสามารถสรุปได้ว่าแบบจำลองที่ดีที่สุดที่เหมาะสมต่อการนำมาวิเคราะห์ PM 2.5 ด้วย Historical Data Feature by Hour ทั้งในช่วงสถานการณ์ปกติที่ชุดข้อมูลไม่แตกต่างกันและเมื่อมีการแพร่ระบาดของ Covid-19 อันดับที่ 1 คือ แบบจำลอง Gradient Boosting, อันดับที่ 2 คือ แบบจำลอง Random Forest, อันดับที่ 3 คือ แบบจำลอง MLP ตามลำดับ

ตาราง 6.1 สรุปผลความถูกต้องและแม่นยำ ( $R^2$ ) ของแบบจำลองที่ใช้ในงานศึกษา

| แบบจำลอง          | $R^2$        |          |             |        | อันดับที่ดีที่สุด |
|-------------------|--------------|----------|-------------|--------|-------------------|
|                   | Training set | Test Set | Testout Set | Avg.   |                   |
| Decision Tree     | 0.8673       | 0.8400   | 0.8507      | 0.8527 | 4                 |
| Gradient Boosting | 0.9533       | 0.8790   | 0.8720      | 0.9014 | 1                 |
| K-neighbors       | 0.8603       | 0.8383   | 0.8295      | 0.8427 | 5                 |
| MLP               | 0.8530       | 0.8453   | 0.8645      | 0.8543 | 3                 |
| Random Forest     | 0.8947       | 0.8578   | 0.8691      | 0.8739 | 2                 |
| Ridge             | 0.7375       | 0.7391   | 0.7654      | 0.7473 | 6                 |

ทั้งนี้ ผลการศึกษาในงานวิจัยชิ้นนี้เป็นการคาดการณ์สถานการณ์ PM 2.5 ในพื้นที่กรุงเทพมหานคร ด้วยชุดข้อมูลด้านอุตุนิยมวิทยาและมลพิษทางอากาศของประเทศไทยของกรมควบคุมมลพิษ และ สำนักสิ่งแวดล้อม กรุงเทพมหานคร

อย่างไรก็ตาม จากผลการศึกษาที่การคาดการณ์สถานการณ์ PM 2.5 ในพื้นที่กรุงเทพมหานคร ด้วยการประยุกต์ใช้ Machine Learning Techniques ในงานวิจัยชิ้นนี้ ให้ผลการศึกษาที่มีความถูกต้องแม่นยำกว่าการใช้แบบจำลองทางคณิตศาสตร์ทั่วไป (Amnuaylojaroen, 2022) ซึ่งแบบจำลองที่ดีที่สุดของงานวิจัยนี้ให้  $R^2 \sim 0.9014$  ในขณะที่แบบจำลองทั่วไปได้  $R^2 \sim 0.600$  บนพื้นฐานชุดข้อมูลด้านอุตุนิยมวิทยาและมลพิษทางอากาศของประเทศไทย

## 6.2 ข้อเสนอแนะ

แบบจำลองในงานศึกษานี้ เป็นการใช้ข้อมูลอุตุนิยมวิทยาและข้อมูลมลภาวะทางอากาศที่รวบรวมในพื้นที่กรุงเทพมหานคร ซึ่งเป็นพื้นที่เขตเมือง (Urban) โดยมีปริมาณจราจรหนาแน่น ไม่มีโรงงานอุตสาหกรรม และการทำเกษตรกรรม อย่างไรก็ตาม หากมีการนำแบบจำลองเหล่านี้ไปทดสอบกับชุดข้อมูลทางสถิติในพื้นที่อื่น ที่มีลักษณะทางกายภาพและกิจกรรมของประชาชนในพื้นที่แตกต่างกันออกไป อาจส่งผลให้ผลการศึกษาที่ได้แตกต่างจากข้อมูลชุดนี้ก็เป็นได้

โดยจากผลการศึกษาข้างพบว่า ปัจจัยด้านปริมาณจราจรอาจเป็นปัจจัยหลักที่ส่งผลกระทบต่อความเข้มข้นของ PM 2.5 ในพื้นที่ศึกษา ซึ่งในอนาคตหากผู้กำหนดนโยบาย (Policy Maker) และหน่วยงานที่เกี่ยวข้อง มีการตั้งจุดตรวจวัดข้อมูลอุตุนิยมวิทยาและข้อมูลมลภาวะทางอากาศบริเวณริมถนน (Roadside Area) และมีการสำรวจปริมาณจราจร (Collecting Traffic Data) ไปพร้อมกัน ซึ่งจะทำให้ได้สถิติชุดข้อมูลมาจัดทำแบบจำลองให้มีความแม่นยำเพิ่มมากขึ้น และสามารถคาดการณ์ความเข้มข้นของ PM 2.5 ได้อย่างมีประสิทธิภาพมากยิ่งขึ้น

ทั้งนี้ อาจจะสามารถวิเคราะห์ถึงถึงต้นตอของปัญหา PM 2.5 ในเขตเมืองที่มีความอ่อนไหว (Sensitive) ด้วยการจราจรที่หนาแน่นซึ่งจะสามารถแก้ไขปัญหาได้อย่างถูกต้องและตรงจุด พร้อมทั้งสามารถวางแผนแก้ไขปัญหาได้อย่างมีประสิทธิภาพ

### 6.3 การวิจัยในอนาคต (Future Study)

เมื่อผลการวิเคราะห์ค่า PM 2.5 ด้วยแบบจำลองทั้ง 6 ชนิด โดยใช้ข้อมูลด้านอนุกรมเวลาในการทดสอบแบบจำลอง ซึ่งได้ผลการทดสอบประสิทธิภาพด้วย  $R^2$  มากกว่า 0.7 ดังที่กล่าวไปในบทที่ 5 ในหัวข้อการนำแบบจำลองไปใช้โดยทั่วไป (Model Results by General) พบว่า ตัวแปรที่เกี่ยวข้องกับการวิเคราะห์ PM 2.5 ให้เกิดความแม่นยำทั้งในสถานการณ์ปกติ และไม่ปกติ ได้แก่ ทิศทางลม (Wind Direction) ฝุ่นหยาบ (PM 10) และ ฝุ่นละเอียด (PM 2.5) ย้อนหลัง ซึ่งเป็นชุดข้อมูลที่เกี่ยวข้อง เช่น กรมอุตุนิยมวิทยา กรมควบคุมมลพิษ และหน่วยงานท้องถิ่น มีการเก็บสถิติอยู่ตลอดทั้งปี

โดยผู้วิจัยขอเสนอแนะว่า หากมีการศึกษาในอนาคตด้วยการจัดทำแบบจำลองเพื่อคาดการณ์ PM 2.5 หรือสารมลพิษต่าง ๆ ที่เกี่ยวข้อง โดยทำเป็น Dynamic Model กล่าวคือ แบบจำลองไม่เพียงแต่คาดการณ์เฉพาะ PM 2.5 เท่านั้น แบบจำลองยังสามารถคาดการณ์ค่าตัวแปรต่างๆ ที่เป็นปัจจัยของการเกิด PM 2.5 เช่น ทิศทางลม (Wind Direction) ฝุ่นหยาบ (PM 10) เป็นต้น จะยิ่งทำให้แบบจำลองมีความแม่นยำสูงมากยิ่งขึ้น

ทั้งนี้ หากประเทศไทยมีแบบจำลองที่สามารถคาดการณ์ PM 2.5 และสามารถนำไปต่อยอดโดยการพัฒนาขึ้นบน Web Application โดยแสดงผลการวิเคราะห์และคาดการณ์ผลการศึกษา นำเสนอข้อมูลเป็น Dashboard จะยังมีประโยชน์สูงสุด

## เอกสารอ้างอิง

- กรมควบคุมมลพิษ, กระทรวงทรัพยากรธรรมชาติและสิ่งแวดล้อม. 2562. **ประกาศกรมควบคุมมลพิษ เรื่อง เครื่องวัดและวิธีตรวจวัดค่าเฉลี่ยของก๊าซหรือฝุ่นละอองในบรรยากาศ โดยทั่วไป ระบบอื่นหรือวิธีอื่นที่กรมควบคุมมลพิษเห็นชอบ.** [ออนไลน์] เข้าถึงได้จาก : <https://www.pcd.go.th/laws/2768>, 30 ตุลาคม 2564.
- กรมควบคุมมลพิษ, กระทรวงทรัพยากรธรรมชาติและสิ่งแวดล้อม. 2547. **คู่มือการควบคุมและประกันคุณภาพงานห้องปฏิบัติการ สิ่งแวดล้อม.** [ออนไลน์] เข้าถึงได้จาก : <https://www.pcd.go.th/publication/5420>, 30 ตุลาคม 2564.
- กรมควบคุมมลพิษ, กระทรวงทรัพยากรธรรมชาติและสิ่งแวดล้อม. 2562. **คู่มือปฏิบัติการในการป้องกันและแก้ไขปัญหาฝุ่นละออง PM2.5 ระดับจังหวัด.** [ออนไลน์] เข้าถึงได้จาก : <https://www.pcd.go.th/publication/13479>, 30 ตุลาคม 2564.
- กรมควบคุมมลพิษ, กระทรวงทรัพยากรธรรมชาติและสิ่งแวดล้อม. 2565. **ประกาศคณะกรรมการสิ่งแวดล้อมแห่งชาติ เรื่อง กำหนดมาตรฐานฝุ่นละอองขนาดไม่เกิน 2.5 ไมครอน ในบรรยากาศโดยทั่วไป.** [ออนไลน์] เข้าถึงได้จาก : <https://www.pcd.go.th/laws/26439>, 30 มีนาคม 2565.
- การทางพิเศษแห่งประเทศไทย. 2564. **รายงานจราจรประจำปี 2564.** [ออนไลน์] เข้าถึงได้จาก : <https://www.exat.co.th/statistics/>, 15 มกราคม 2565.
- Amnuaylojaroen, T. 2022. "Prediction of PM2.5 in an Urban Area of Northern Thailand Using Multivariate Linear Regression Model." **Advances in Meteorology.** 2022. Article ID 3190484 : 1-9.
- Anke, J., Francke, A., Schaefer, L. M., and Petzoldt, T. 2021. "Impact of SARS-CoV-2 on the mobility behaviour in Germany." **European Transport Research Review.** 2021. 13, 1 : 1-13.
- Arrhenius, S. A. 1824. "On the Influence of Carbonic Acid in the Air upon the Temperature of the Ground." **Philosophical Magazine and Journal of Science.** 41, 5 : 237-276.
- Awad-Núñez, S., Julio, R., Gomez, J., Moya-Gómez, B., and González, J. S. 2021. "Post-COVID-19 travel behaviour patterns: impact on the willingness to pay of users of public transport and shared mobility services in Spain." **European Transport Research Review.** 13, 20 : 1-18.

## เอกสารอ้างอิง (ต่อ)

- Benchrif, A., Wheida, A., Tahri, M., Shubbar, R. M., and Biswas, B. 2021. "Air quality during three covid-19 lockdown phases: AQI, PM2.5 and NO2 assessment in cities with more than 1 million inhabitants." **Sustainable Cities and Society**. 74, 103170 : 1-21
- Biggio, B., and Roli, F. 2018. "Wild patterns: Ten years after the rise of adversarial machine learning." **Pattern Recognition**. 84, 2018 : 317–331.
- Chen, L., Mao, F., Hong, J., Zang, L., Chen, J., Zhang, Y., Gan, Y., Gong, W., and Xu, H. 2022. "Improving PM2.5 predictions during Covid-19 lockdown by assimilating multi-source observations and adjusting emissions." **Environmental Pollution**. 297, 118783 : 1-10.
- Chirasophon, S., and Pochanart, P. 2020. "The long-term characteristics of PM10 and PM2.5 in Bangkok, Thailand." **Asian Journal of Atmospheric Environment**. 14, 1 : 73–83.
- Cho, S.-H., and Park, H.-C. 2021. "Exploring the Behaviour Change of Crowding Impedance on Public Transit due to Covid-19 Pandemic: Before and After Comparison." **Transportation Letters**. 13, 5–6, 2021 : 1–8.
- Chulalongkorn University. 2020. **Stay Safe In The PM 2.5**. Translated by Thanutra Teerasuphaset. Bangkok.
- Cui, Q., He, L., Liu, Y., Zheng, Y., Wei, W., Yang, B., and Zhou, M. 2021. "The impacts of COVID-19 pandemic on China's transport sectors based on the CGE model coupled with a decomposition analysis approach." **Transport Policy**. 103 (January) : 103–115.
- Dang, S., Penney, L. S., Trivedi, R., Noel, P. H., Pugh, M. J., Finley, E., Pugh, J. A., Van Houtven, C. H., and Leykum, L. 2020. "Caring for Caregivers During COVID-19". In **Journal of the American Geriatrics Society**. 68, 10 : 2197–2201.
- Dasgupta, P., and Srikanth, K. 2020. "Reduced air pollution during COVID-19: Learnings for sustainability from Indian Cities." **Global Transitions**, 2 : 271–282.
- Delfino, R. J., Staimer, N., Tjoa, T., Gillen, D., Kleinman, M. T., Sioutas, C., and Cooper, D. 2008. "Personal and ambient air pollution exposures and lung function decrements in children with asthma." **Environmental Health Perspectives**. 116, 4 : 550–558.

## เอกสารอ้างอิง (ต่อ)

- Dingil, A. E., and Esztergár-Kiss, D. 2021. "The Influence of the Covid-19 Pandemic on Mobility Patterns: The First Wave's Results." **Transportation Letters**. 13, 5-6 : 434-446.
- Doreswamy, Harishkumar, K. S., Km, Y., and Gad, I. 2020. "Forecasting Air Pollution Particulate Matter (PM<sub>2.5</sub>) Using Machine Learning Regression Models." **Procedia Computer Science**. 171 (2019) : 2057–2066.
- Ethem Alpaydin. 2010. **Introduction to Machine Learning**. 2<sup>nd</sup> ed. The MIT Press. Cambridge, Massachusetts London, England.
- European Environment Agency. 2013. "EMEP/EEA air pollutant emission inventory guidebook 2013: Technical guidance to prepare national emission inventories." **EEA Technical Report**. 12 : 1-26.
- EPA, U. S. 2016. "Criteria Air Pollutants." **Encyclopedia of Immunotoxicology**, October : 218–218.
- Friedman, J. H. 2001. "Greedy function approximation: A gradient boosting machine." **Annals of Statistics**. 29, 5 : 1189–1232.
- ghsindex.org. 2019. **2019 Global Health Security Index**. [Online] Available : <https://www.ghsindex.org/wp-content/uploads/2019/10/2019-Global-Health-Security-Index.pdf>, October 17, 2021.
- Goodfellow, I., Bengio, Y., and Courville, A. 2016. **Deep Learning**. MIT Press. [Online] Available : <https://books.google.co.th/books?id=Np9SDQAAQBAJ>, June 12, 2021.
- Greenpeace Thailand. 2016. **Greenpeace's City Rankings for PM<sub>2.5</sub> in Thailand**. [Online] Available : <https://greenpeace.or.th/s/right-to-clean-air/PM2.5CityRankingsREV.pdf>, March 20, 2021.
- Grundström, M., Hak, C., Chen, D., Hallquist, M., and Pleijel, H. 2015. "Variation and co-variation of PM<sub>10</sub>, particle number concentration, NO<sub>x</sub> and NO<sub>2</sub> in the urban air - Relationships with wind speed, vertical temperature gradient and weather type." **Atmospheric Environment**. 120, 2 : 317–327.
- Guo, Z., Wang, X., and Ge, L. 2023. "Classification prediction model of indoor PM<sub>2.5</sub> concentration using CatBoost algorithm." **Frontiers in Built Environment**. 9(July) : 1–10.



## เอกสารอ้างอิง (ต่อ)

- Hai-Anh H. Dang. and Trong-Anh Trinh. 2020. **Does the COVID-19 Pandemic Improve Global Air Quality? New Cross-National Evidence on Its Unintended Consequences.** DISCUSSION PAPER SERIES. IZA DP No. 13480. The IZA Institute of Labor Economics. Schaumburg-Lippe-Straße 5–9 53113 Bonn, Germany.
- Harrison, R. M., and Yin, J. 2000. "Particulate matter in the atmosphere: which particle properties are important for its effects on health?" **The Science of the Total Environment.** 249, 1–3 : 85–101.
- Hasselwander, M., Tamagusko, T., Bigotte, J. F., Ferreira, A., Mejia, A., and Ferranti, E. J. S. 2021. "Building back better: The COVID-19 pandemic and transport policy implications for a developing megacity." **Sustainable Cities and Society.** 69(October) : 1-12.
- Heydari, S., Tainio, M., Woodcock, J., and de Nazelle, A. 2020. "Estimating traffic contribution to particulate matter concentration in urban areas using a multilevel Bayesian meta-regression approach." **Environment International.** 141(December) : 1-8.
- Hodan, W. M., and Barnard, W. R. 2004. "Evaluating the Contribution of PM2.5 Precursor Gases and Re-entrained Road Emissions to Mobile Source PM2.5 Particulate Matter Emissions." **MACTEC Federal Programs, Research Triangle Park.** : 1-58.
- Ian, J., Goodfellow., Yoshua, Bengio., and Aaron, Courville. 2016. **Deep Learning.** MIT Press. Cambridge, MA, USA.
- IPCC 1992. **Climate Change: The IPCC 1990 and 1992 Assessments.** Intergovernmental Panel on Climate Change 1992, Canada.
- IQAir. 2014. **World air quality report 2021.** [Online] Available : [https://www.iqair.com/th-en/newsroom/WAQR\\_2021\\_PR](https://www.iqair.com/th-en/newsroom/WAQR_2021_PR), April 30, 2022.
- Irawan, M. Z., Belgiawan, P. F., Joewono, T. B., Bastarianto, F. F., Rizki, M., and Ilahi, A. 2021. "Exploring activity-travel behavior changes during the beginning of Covid-19 pandemic in Indonesia." **Transportation (Amst).** 49, 2 : 529-553.
- Jiao, J., Bhat, M., and Azimian, A. 2021. "Measuring travel behavior in Houston, Texas with mobility data during the 2020 Covid-19 outbreak." **Transportation Letters.** 13, 5-6 : 461-472.

## เอกสารอ้างอิง (ต่อ)

- Ju, M. J., Oh, J., and Choi, Y. H. 2021. "Changes in air pollution levels after Covid-19 outbreak in Korea." **Science of the Total Environment**. 750, 141521 : 1-9.
- Kanakidou, M., Seinfeld, J. H., Pandis, S. N., Barnes, I., Dentener, F. J., Facchini, M. C., Van Dingenen, R., Ervens, B., Nenes, A., Nielsen, C. J., Swietlicki, E., Putaud, J. P., Balkanski, Y., Fuzzi, S., Horth, J., Moortgat, G. K., Winterhalter, R., Myhre, C. E. L., Tsigaridis, K., Wilson, J. 2005. "Organic aerosol and global climate modelling: a review." **Atmospheric Chemistry and Physics**. 5, 4 : 1053–1123.
- Kanchanasuta, S., Sooktawee, S., Patpai, A., and Vatanasomboon, P. 2020. "Temporal Variations and Potential Source Areas of Fine Particulate Matter in Bangkok, Thailand." **Air, Soil and Water Research**. 13 : 1-10.
- König, A., and Dreßler, A. 2021. "A mixed-methods analysis of mobility behavior changes in the COVID-19 era in a rural case study." **European Transport Research Review**. 13, 1 : 1-13.
- Kruskal, W. H., and Wallis, W. A. 1952. "Use of Ranks in One-Criterion Variance Analysis." **Journal of the American Statistical Association**. 47, 260 : 583–621.
- Lee, H., Park, S. J., Lee, G. R., Kim, J. E., Lee, J. H., Jung, Y., and Nam, E. W. 2020. "The relationship between trends in Covid-19 prevalence and traffic levels in South Korea." **International Journal of Infectious Diseases**. 96 : 399–407.
- Lewis, C. D. 1982. **Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting**. Butterworth Scientific. Lancashire.
- Li, C., and Managi, S. 2021. "Contribution of on-road transportation to PM<sub>2.5</sub>." **Scientific Reports**. 11, 1 : 1–12.
- Li, L. N., Gong, X. P., Dai, L. C., and Zhan, X. H. 2014. "The regression models of PM<sub>2.5</sub> and other air pollutants in Wuhan." **Advanced Materials Research**. 864–867(June) : 1356–1359.
- Li, R., Mei, X., Wei, L., Han, X., Zhang, M., and Jing, Y. 2019. "Study on the contribution of transport to PM<sub>2.5</sub> in typical regions of China using the regional air quality model RAMS-CMAQ." **Atmospheric Environment**. 214 : 1-10.
- Liu, C., Hsu, P. C., Lee, H. W., Ye, M., Zheng, G., Liu, N., Li, W., and Cui, Y. 2015. "Transparent air filter for high-efficiency PM 2.5 capture." **Nature Communications**. 6, 6205 : 1-9.

## เอกสารอ้างอิง (ต่อ)

- MacFarland, T. W., and Yates, J. M. 2018. **Introduction to Nonparametric Statistics for the Biological Sciences Using R. 1<sup>st</sup> ed.** Springer Publishing Company.
- Maciejczyk, P., Zhong, M., Lippmann, M., and Chen, L.-C. 2010. "Oxidant generation capacity of source-apportioned PM<sub>2.5</sub>." **Inhalation Toxicolog.** 2, 2 : 1-9.
- Mahanta, S., Ramakrishnudu, T., Jha, R. R., and Tailor, N. 2019. "Urban Air Quality Prediction Using Regression Analysis." **EEE Region 10 Annual International Conference, Proceedings/TENCON** : 1118–1123.
- Mahmud, S., Ridi, T. B. I., Miah, M. S., Sarower, F., and Elahee, S. 2022. "Implementing Machine Learning Algorithms to Predict Particulate Matter (PM<sub>2.5</sub>): A Case Study in the Paso del Norte Region." **Atmosphere.** 13, 12 : 1-16.
- Mao, J., Li, L., Li, J., Sulaymon, I. D., Xiong, K., Wang, K., Zhu, J., Chen, G., Ye, F., Zhang, N., Qin, Y., Qin, M., and Hu, J. 2022. "Evaluation of Long-Term Modeling Fine Particulate Matter and Ozone in China During 2013–2019." **Frontiers in Environmental Science.** 10, 4 : 1–13.
- Mehdipour, V., Stevenson, D., Memarianfard, M., and Sihag, P. 2018. "Comparing different methods for statistical modeling of particulate matter in Tehran, Iran." **Air Quality, Atmosphere & Health.** 11 : 1155–1165.
- Miao, Y., Liu, S., Sheng, L., Huang, S., and Li, J. 2019. "Influence of Boundary Layer Structure and Low-Level Jet on PM<sub>2.5</sub> Pollution in Beijing: A Case Study." **International Journal of Environmental Research and Public Health.** 16, 4 : 1-14.
- Ministry of Transport. 2021. **People's travel During the COVID-19 viral epidemic - dataset - MOT Data Catalog.** [Online] Available : <https://datagov.mot.go.th/dataset/greenhouse-emissions-in-the-road-transport-sector>, January 22, 2022.
- Molnar, C. 2020. **Interpretable Machine Learning. A Guide for Making Black Box Models Explainable.** Leanpub Book. Canada.
- Muley, D., Ghanim, M. S., Mohammad, A., and Kharbeche, M. 2021. "Quantifying the impact of Covid–19 preventive measures on traffic in the State of Qatar." **Transport Policy.** 103, : 45–59.

## เอกสารอ้างอิง (ต่อ)

- Narita, D., Oanh, N.T., Sato, K., Huo, M., Permadi, D.A., Chi, N.N., Ratanajatroj, T., and Paw-armart, I. 2019. "Pollution Characteristics and Policy Actions on Fine Particulate Matter in a Growing Asian Economy: The Case of Bangkok Metropolitan Region." **Atmosphere**. 10, 227 : 1-18.
- Nevada Department of Environmental Protection. n.d. **Particulate Matter Pollution Fact Sheet**. [Online] Available : [https://ndep.nv.gov/baqp/monitoring/docs/particulate\\_matter.pdf](https://ndep.nv.gov/baqp/monitoring/docs/particulate_matter.pdf), January 10, 2021.
- Ni, X. 2019. "Relevance analysis and short-term prediction of PM2.5 concentrations in Beijing based on multi-source data." **Atmospheric Environment**. 150(February) : 146-161.
- Palanichamy, N., Haw, S.-C., S, S., Govindasamy, K., and Murugan, R. 2021. "Prediction of PM2.5 concentrations in Malaysia using machine learning techniques: a review." **F1000Research** : 1-10.
- Palanichamy, N., Haw, S.-C., S, S., Murugan, R., and Govindasamy, K. 2022. "Machine learning methods to predict particulate matter PM (2.5)." **F1000Research**. 11(76798) : 1-10.
- Pan, B. 2018. "Application of XGBoost algorithm in hourly PM2.5 concentration prediction." **IOP Conference Series: Earth and Environmental Science**. 113, 1 : 1-8.
- Patra, S., Chilukuri, B., and Vanajakshi, L. 2021. "Analysis of road traffic pattern changes due to activity restrictions during Covid-19 pandemic in Chennai." **Transportation Letters**. 13, 5-6 : 473-481.
- Patro, S. G. K., and sahu, K. K. 2015. Normalization: A Preprocessing Stage. **Iarjset**. 1-4.
- Pérez-Campos Mayoral, L., Hernández-Huerta, M. T., Mayoral-Andrade, G., Pérez-Campos Mayoral, E., and Pérez-Campos, E. 2020. "World Health Organization declares global emergency: A review of the 2019 novel Coronavirus (Covid-19)." **International Journal of Surgery**. 79(February) : 163–164.
- Persson, R. 2021. "Weight of evidence transformation in credit scoring models: How does it affect the discriminatory power?" **School of Economics & Management, LUND UNIVERSITY**. 1–5.

## เอกสารอ้างอิง (ต่อ)

- Pishue, B., and Trepanier, T. 2021. 2020 INRIX Global Traffic Scorecard. **INRIX** : 1-23.
- Politis, I., Georgiadis, G., Nikolaidou, A., Kopsacheilis, A., Fyrogenis, I., Sdoukopoulos, A., Verani, E., and Papadopoulos, E. 2021. "Mapping travel behavior changes during the Covid-19 lock-down: a socioeconomic analysis in Greece." **European Transport Research Review**. 13, 1 : 1-19.
- Pollution Control Department. 2021. **Bangkok Air Pollution: Real-time Air Quality Index (AQI)**. [Online] Available : <https://aqicn.org/city/bangkok>, January 15, 2021.
- Pöschl, U. 2005. "Atmospheric aerosols: composition, transformation, climate and health effects." **Angewandte Chemie (International Ed. in English)**. 44, 46 : 7520–7540.
- Puliatti, S., Eissa, A., Eissa, R., Amato, M., Mazzone, E., Dell'Oglio, P., Sighinolfi, M. C., Zoeir, A., Micali, S., Bianchi, G., Patel, V., Wiklund, P., Coelho, R. F., Bernhard, J. C., Dasgupta, P., Mottrie, A., and Rocco, B. 2020. "Covid-19 and urology: a comprehensive review of the literature." **BJU International**. 125, 6 : E7–E14.
- Rosenlund, M., Forastiere, F., Porta, D., De Sario, M., Badaloni, C., and Perucci, C. A. 2009. "Traffic-related air pollution in relation to respiratory symptoms, allergic sensitisation and lung function in schoolchildren." **Thorax**. 64, 7 : 573–580.
- Sacks, J. D., Stanek, L. W., Luben, T. J., Johns, D. O., Buckley, B. J., Brown, J. S., and Ross, M. 2011. "Particulate matter-induced health effects: who is susceptible?" **Environmental Health Perspectives**. 119, 4 : 446–454.
- Schneider, S. H. 1991. "Report on reports: Three reports of the Intergovernmental Panel on Climate Change." **Environment: Science and Policy for Sustainable Development**. 33, 1 : 25-30.
- Silberston, A. 1993. "The Royal Commission on Environmental Pollution." **Proceedings of a conference held by the Confederation of European Economic Associations at Oxford, 1993**. : 129–148.
- Shaun, T. 2022. **Pearson Correlation Coefficient (r) | Guide & Examples**. [Online] Available : <https://www.scribbr.com/statistics/pearson-correlation-coefficient/>, October 20, 2022.

## เอกสารอ้างอิง (ต่อ)

- Sobri, N. M., Wan Yaacob, W. F., Ismail, N. A., Malik, M. A. A., Rahman, R. A., Baser, N. A., and Sukhairi, S. A. M. 2021. "Predicting particulate matter (PM<sub>2.5</sub>) in Malaysia using Multiple Linear Regression and Artificial Neural Network." **Journal of Physics: Conference Series**. 2084. : 1-11.
- Statista. 2021. **Chart : The Cities With The Worst Traffic Congestion | Statista. (2021)**. [Online] Available : <https://www.statista.com/chart/18281/percentage-of-extra-travel-time-due-to-congestion/>, January 15, 2021.
- Stuart, J., Russell. and Peter, Norvig. 2010. **Artificial intelligence : A Modern Approach Third edition**. Pearson Education, Inc., Upper Saddle River, New Jersey 07458.
- Sun, R., Zhou, Y., Wu, J., and Gong, Z. 2019. "Influencing factors of PM<sub>2.5</sub> pollution: Disaster points of meteorological factors." **International Journal of Environmental Research and Public Health**. 16, 20 : 1-31.
- Tecer, L. H., Alagha, O., Karaca, F., Tuncel, G., and Eldes, N. 2008. "Particulate matter (PM<sub>2.5</sub>), PM<sub>10-2.5</sub>, and PM<sub>10</sub>) and children's hospital admissions for asthma and respiratory diseases: a bidirectional case-crossover study." **Journal of Toxicology and Environmental Health**. Part A, 71, 8 : 512-520.
- Tomtom Traffic Index. (2021). **Traffic Index Traffic Index results 2018 - 2021**. [Online] Available : <https://www.tomtom.com/traffic-index/ranking/>, February 23, 2022.
- Trasande, L., and Thurston, G. D. 2005. "The role of air pollution in asthma and other pediatric morbidities." **The Journal of Allergy and Clinical Immunology**. 115, 4 : 689-699.
- Tunno, B. J., Michanowicz, D. R., Shmool, J. L. C., Tripathy, S., Kinnee, E., Cambal, L., Chubb, L., Roper, C., and Clougherty, J. E. 2018. "Spatial patterns in rush-hour vs. Work-week diesel-related pollution across a downtown core." **International Journal of Environmental Research and Public Health**. 15, 1968 : 1-15.
- United Nations Framework Convention on Climate Change. 1992**. S. Treaty Doc No. 102-38, 1771 U.N.T.S. 107.
- Uttamang, P., Aneja, V. P., and Hanna, A. F. 2018. "Assessment of gaseous criteria pollutants in the Bangkok Metropolitan Region, Thailand." **Atmospheric Chemistry and Physics**. 18, 16 : 12581-12593.

## เอกสารอ้างอิง (ต่อ)

- Varrassi, G., Alon, E., Bagnasco, M., Lanata, L., Mayoral-Rojals, V., Paladini, A., Pergolizzi, J.V., Perrot, S., Scarpignato, C., and Tölle, T.R. 2019. "Towards an Effective and Safe Treatment of Inflammatory Pain: A Delphi-Guided Expert Consensus." **Advances in Therapy**. 36 : 2618 - 2637.
- Venter, Z. S., Aunan, K., Chowdhury, S., and Lelieveld, J. 2020. "Covid-19 lockdowns cause global air pollution declines." **Proceedings of the National Academy of Sciences of the United States of America**. 117, 32 : 18984–18990.
- Vos, J. De. 2020. "The effect of Covid-19 and subsequent social distancing on travel behavior." **Transportation Research Interdisciplinary Perspectives**. 5 : 1-5.
- Wang, Y., and Yang, D. 2018. "Impacts of freight transport on PM2.5 concentrations in China: A spatial dynamic panel analysis." **Sustainability**. 10, 8 : 1-16.
- Webel, K. 2021. "A Data-Driven Selection of an Appropriate Seasonal Adjustment Approach." **SSRN Electronic Journal**. 07. : 1-40.
- Wetchayont, P. 2021. "Investigation on the Impacts of Covid-19 Lockdown and Influencing Factors on Air Quality in Greater Bangkok, Thailand." **Advances in Meteorology**. 2 : 1-11.
- Winston, P. H. 1992. **Artificial Intelligence** 3<sup>rd</sup> ed. Addison-Wesley Longman Publishing Co., Inc.
- Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y., Yuan, M.L., Zhang, Y.L., Dai, F. H., and Liu, Y., Wang, Q.M., Zheng, J.J., Xu, L., Holmes, E.C. and Zhang, Y. Z. 2020. "Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, co - Nucleotide - NCBI." **Nature**. 579, 7798 : 265–269.
- Xia, T., Nitschke, M., Zhang, Y., Shah, P., Crabb, S., and Hansen, A. 2015. "Traffic-related air pollution and health co-benefits of alternative transport in Adelaide, South Australia." **Environment International**. 74 : 281–290.  
<https://doi.org/10.1016/j.envint.2014.10.004>
- Yin, S., Liu, H., and Duan, Z. 2021. "Hourly PM2.5 concentration multi-step forecasting method based on extreme learning machine, boosting algorithm and error correction model." **Digital Signal Processing: A Review Journal**. 118 : 103221.

## เอกสารอ้างอิง (ต่อ)

- Yu, W., Ye, T., Zhang, Y., Xu, R., Lei, Y., Chen, Z., Yang, Z., Zhang, Y., Song, J., Yue, X., and Li, S. 2023. "Articles Global estimates of daily ambient fine particulate matter concentrations and unequal spatiotemporal distribution of population exposure : a machine learning modelling study." **The Lancet Planetary Health**. 7, 3 : e209–e218.
- Yu, X., Xi, M., Wu, L., and Zheng, H. 2023. "Spatiotemporal Weighted for Improving the Satellite-Based High-Resolution Ground PM2.5 Estimation Using the Light Gradient Boosting Machine." **Remote Sensing**. 15, 16 : 1-21.
- Zhang, A., Lipton, Z. C., Li, M., and Smola, A. J. 2021. **Dive into Deep Learning**. Amazon Science.
- Zhao, H., Che, H., Zhang, X., Ma, Y., Wang, Y., Wang, H., and Wang, Y. 2013. "Characteristics of visibility and particulate matter (PM) in an urban area of Northeast China." **Atmospheric Pollution Research**. 4, 4 : 427–434.
- Zhao, X., Gao, Q., Sun, M., Xue, Y., Ma, R. J., Xiao, X., and Ai, B. 2018. "Statistical analysis of spatiotemporal heterogeneity of the distribution of air quality and dominant air pollutants and the effect factors in Qingdao Urban Zones." **Atmosphere**. 9, 4 : 12–16.
- Zheng, A., and Casari, A. 2018. **Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists**. O'Reilly Media Inc.
- Zhou, L., Zhou, C., Yang, F., Che, L., Wang, B., and Sun, D. 2019. "Spatio-temporal evolution and the influencing factors of PM2.5 in China between 2000 and 2015." **Journal of Geographical Sciences**. 29, 2 : 253–270.



## ภาคผนวก

ภาคผนวก ก เอกสารตีพิมพ์

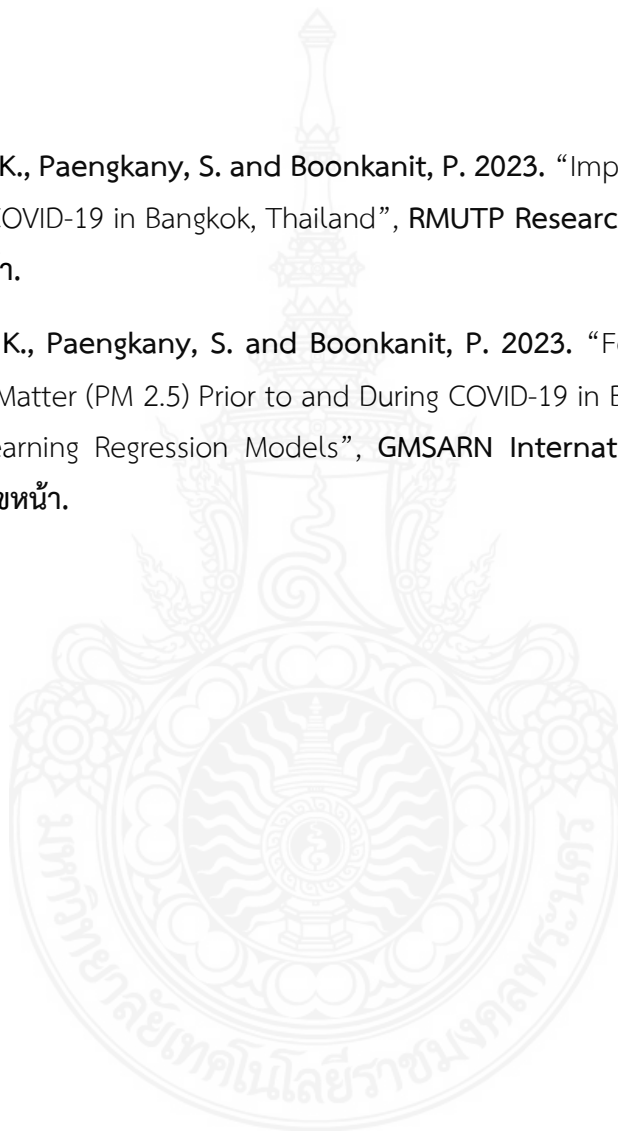
ภาคผนวก ข รายละเอียดชุดคำสั่ง (Code) ของการพัฒนาแบบจำลอง



## ภาคผนวก ก

### เอกสารตีพิมพ์

1. Jedwanna, K., Paengkany, S. and Boonkanit, P. 2023. “Impact of Particulate Matter 2.5 During COVID-19 in Bangkok, Thailand”, *RMUTP Research Journal*. ปีที่ 18, ฉบับที่ 2 : เลขหน้า.
2. Jedwanna, K., Paengkany, S. and Boonkanit, P. 2023. “Forecasting Air Pollution Particulate Matter (PM 2.5) Prior to and During COVID-19 in Bangkok, Thailand Using Machine Learning Regression Models”, *GMSARN International Journal*. Vol.19, Issue 4 : เลขหน้า.



ภาพ ก-1 แบบตอบรับการตีพิมพ์บทความในวารสารวิชาการและวิจัย มทร. พระนคร เรื่อง Impact of Particulate Matter 2.5 During COVID-19 in Bangkok, Thailand



## บันทึกข้อความ

**ส่วนราชการ** สถาบันวิจัยและพัฒนา งานวารสารและเผยแพร่งานวิจัย

โทรศัพท์ ๐ ๒๖๖๕ ๓๗๗๗ ต่อ ๘๒๐๗

ที่ อว.๑๖๕๒.๑๑.๑๒๒ วันที่ ๑๐ พฤษภาคม ๒๕๖๖

**เรื่อง** ...ตอบรับการตีพิมพ์บทความในวารสารวิชาการและวิจัย มทร.พระนคร...

**เรียน** ผู้ช่วยศาสตราจารย์กฤษณ์ เจ็ดวรรณะ ดร.สุวิทย์ แพ่งกันยา และ ดร.ปริญญ์ บุญนิษฐ

ตามที่ท่านได้ส่งบทความวิจัย เรื่อง Impact of Particulate Matter 2.5 During Covid-19 in Bangkok, Thailand เพื่อตีพิมพ์ลงในวารสารวิชาการและวิจัย มทร.พระนคร นั้น

กองบรรณาธิการวารสารวิชาการและวิจัย มทร.พระนคร ขอแจ้งให้ทราบว่าบทความวิจัยดังกล่าว ได้ผ่านการประเมินจากกรรมการผู้ทรงคุณวุฒิ ๓ ท่าน และได้รับการตีพิมพ์ลงในวารสารวิชาการและวิจัย มทร.พระนคร สาขาวิทยาศาสตร์และเทคโนโลยี ปีที่ ๑๘ ฉบับที่ ๒ เดือนกรกฎาคม-ธันวาคม ๒๕๖๖ อนึ่งบทความดังกล่าวถือเป็นลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร

จึงเรียนมาเพื่อโปรดทราบ

  
(นายชลากร อุตมรักษาสกุล)  
ผู้อำนวยการสถาบันวิจัยและพัฒนา



<http://journal.rmutp.ac.th/>

## The Impact of Particulate Matter (PM2.5) During the Covid-19 in Bangkok, Thailand

Krit Jedwanna<sup>1\*</sup> Suwit Paengkanya<sup>2</sup> and Prin Boonkanit<sup>3</sup>

<sup>1,2,3</sup>Sustainable Industrial Management Engineering, Faculty of Engineering, Rajamangala University of Technology Phra Nakhon

<sup>1,2,3</sup>No.1381 Pracharat 1 Road, Wong Sawang, Bang Sue, Bangkok.

---

*Received* .....; *Revised* .....; *Accepted* .....

### Abstract

Bangkok is one of cities which has severe PM2.5 and traffic congestion problem. In the past decades, traffic is the main cause of PM2.5. Due to the pandemic of Covid-19 in 2020, Bangkok had been lockdown, this affect to significantly decrease of traffic volume during that period. In this research, author has studied about how the change of traffic volume affect to the concentration of pollution especially PM2.5. By collecting data of 63 points in Bangkok and 4 highways during January 2018 to December 2023, the data shown that, during lockdown period, the concentration of PM2.5, NOX, NO2, NO and CO decreased by 14.38%, 12.93%, 18.43%, 2.65% and 5.66%, consecutively. Meanwhile traffic volume decreased about 40%, the concentration of PM2.5 has no significant reduction when compared to decreasing of traffic volume. Therefore, due to pandemic of covid-19, this study found that traffic volume may be a cause of PM2.5, but it is not a main cause. The effect of season on PM2.5 outweigh traffic volume.

**Keywords :** Traffic; PM2.5; Covid-19 pandemic; Bangkok – Thailand.

---

\* *Corresponding Author. Tel.:* +66-89-777-1654, *E-mail Address:* [krit.j@rmutp.ac.th](mailto:krit.j@rmutp.ac.th)

## 1. Introduction

US. EPA (United State Environmental Protection Agency) [1] has set a standard of small particle which is harmful to human health by using PM value (Particulate matters). For PM 10 or course particles is defined as a small particle found in dust and smoke. It has diameter from 2.5 to 10-micron diameter and can be obviously seen. For PM2.5 or fine particle is tiny particle which has diameter from 2.5 micron and below.

There are loads of evidence and studies over the world has shown that traffic is the main factor which contribute PM2.5 [2–7]. Currently, a problem of pollution of PM2.5 has become a severe problem in Thailand especially in metropolitan areas such as Bangkok, Chiangmai, Chiang Rai and Lampang [8,9]. Moreover, traffic jam in Thailand is still the concerned problem as Bangkok has been on top-ranking of traffic jam over long period [10,11]. From beginning of 2020, the pandemic of Covid-19 had changed people behavior. In many countries, there are cities lock down policy which catalyst of online learning and working. Due to the policy, it not only affects economic but also restrict activities in communities. This cause significant reduction of travel trip and changing of transportation mode in many countries [12–14]. Especially on Bangkok highway, during seriously outbreak of covid-19 period, traffic volume decreased about 40 percent and the number of people which use public transportation fall by 50 percent [15,16]. Authors questioned on the relation between the decreasing of traveling and air pollution.

In this research, the statistics data between 2015 to 2021 was separated into 2 periods for comparing data. First set is data before pandemic of Covid-19 (from 2015 to 2019) and second set is data during pandemic of covid-19 (2020 to 2021) [17]. Statistical method was applied to compare the data in same period range [16] as well as to study on dimension of how traffic volume and air pollution related especially the PM2.5 concentration level on changing of traffic.

## 2. Other Studies

Particulate matter (PM) is made up of particles (tiny pieces) of solids or liquids that are in the air. It may include dust, dirt, Soot, Smoke and drops of liquid (Aerosols). It may occur from natural source such as dust, pollen, etc. Also, human activities such as industrial process, engine working, electrical generation process, construction mining and agriculture [18–20], lead to releasing of PM2.5 in atmosphere occur from releasing elementary particle and changing of initial substance of SO<sub>2</sub>, NO<sub>X</sub>, VOCs and NH<sub>3</sub> on primary releasing stage [21–24]. In addition, PM2.5 not only fully contain of toxic but also long-lasting exist in environment. Furthermore, PM2.5 can flow in long distance through atmospheric circulation. This is extremely harmful to human health [25,26].

### 2.1 PM2.5 in Thailand

According to Pollution Control Department has estimated that the number of dead Thai people causes from PM2.5 may reach 22,000 in 2021 [27]. Considering

Thailand air quality index (AQI) from beginning of 2021 [28], the concentration level of PM<sub>2.5</sub> is exceedingly high. Main factor of PM<sub>2.5</sub> are from human activities especially from transportation – smoke from exhaust pipe, emission from incomplete combustion of engine. Those are the main sources of PM<sub>2.5</sub>. For instance, Bangkok, Chiangmai, Khon Kaen, etc [9].

Moreover, high density traffic leads to countless of PM<sub>2.5</sub>. Also, Thailand was ranked fifth in Asian region after Indonesia (rank 1), Myanmar (rank 2), Vietnam (rank 3), and Laos (rank 4) [8]. In 2021, Lampang is the city which has highest level of PM<sub>2.5</sub> in Thailand. The main reason of PM<sub>2.5</sub> in Thailand is dust and air pollution which related to seasoning especially river basin in Chiangmai. As river basin is a source that gather dust particles from burning agriculture method to preparing for next rice harvest in rainy season. These selected areas for agriculture are valley which its landscape is suitable for accumulation of PM<sub>2.5</sub> [29]. Due to rapidly climate change, PM<sub>2.5</sub> problem, especially during dry season of north rural area such as Chiangmai, Lampang and Nan etc., become increasingly severe over period [30,31]. Research studies on the causes of PM<sub>2.5</sub> in community areas such as Bangkok, found that has higher levels of PM<sub>2.5</sub> in the dry season compared to wet seasons. The most common sources of PM<sub>2.5</sub> emissions in Bangkok are biomass combustion, traffic, and industry activities with varying concentrations based on seasonal factors [32–34]. There was also a study on pollutant concentrations for O<sub>3</sub>, NO<sub>x</sub>, CO, and SO<sub>2</sub>

from 1996 to 2009 in Bangkok, between the residential area and the roadside, it was found that winter air pollution concentrations are higher than in summer and rainy seasons. Also found that the concentration of pollutants in roadside areas was higher than in residential areas in Bangkok [35].

## 2.2 Transportation and PM<sub>2.5</sub>

Road Transportation is a source of pollution such as O<sub>3</sub>, CO, NO<sub>x</sub>, and NMVOC; greenhouse gas, CO<sub>2</sub>, CH<sub>4</sub>, and N<sub>2</sub>O; acidic substances NH<sub>3</sub> and SO<sub>2</sub>; carcinogens, for instance, PAHs and POPs as well as toxic heavy metals [36]. Over past period, transportation and traveling is main cause of PM<sub>2.5</sub> in urban areas. According to the study about impacts of freight transport on PM<sub>2.5</sub> concentrations in China, it was shown switching mode of goods transportation from road freight to sea freight significantly help in reducing of the PM<sub>2.5</sub> concentration level in rural area [37]. In accordance with another research study, alternative transport in Adelaide, South Australia using modeling, found that PM<sub>2.5</sub> was directly proportional to VKT, when VKT decreased by 40%, PM<sub>2.5</sub> decreased by 0.4 µg/m<sup>3</sup>. In other words, shifting transportation modes from road to others could help in reducing PM<sub>2.5</sub> level [38]. Moreover, the study about trace elements, which occurs in rush hour compared to the workweek in Pittsburgh, USA., found that the concentration of pollution substances, for instance, As and Se during rush hours is higher than during other periods in each week [39]. From the above research, it can be seen

that in the transport sector, especially road transport, the higher the traffic density, the more it affects the occurrence of PM2.5.

### 2.3 Transportation and PM2.5 during Covid-19 pandemic

In Wuhan, China on December 2020, the WHO China Country office have been reported cases of pneumonia of unknown etiology (unknown case) [40]. Later confirmed by WHO as an emerging infectious disease defined as coronavirus disease 2019 (COVID-19) [41]. Thailand also declared a state of emergency on 26th Mar 2021. This leads to a partial lockdown in Bangkok and surrounding areas. Although the number of infected cases in other areas also continuously increases as a result of the immigration of workers, the number of confirmed cases never exceeds 200 per day. The situation of covid-19 in Thailand is still under control; however, there are uncertain factors, and it is risky to the outbreak. Therefore, the government issued prevention measures and closely monitor the situation [9]. Throughout the pandemic of Covid-19, January 2019 to August 2022, Thailand had taken preventive and proactive measures until the situation unfolded. Then, covid-19 has been declared endemic. Nowadays Thailand has reopened its borders to tourists after a long period of closing. For the countermeasure, from the first infection was found till 2021 (when the situation unfolded), as **Table 1**.

**Table 1.** Preventive measure for handling Covid-19 in Thailand.

| Measure   | Wave            |                 |                 |                 |                 |
|---|-----------------|-----------------|-----------------|-----------------|-----------------|
|   | 1 <sup>st</sup> | 2 <sup>nd</sup> | 3 <sup>rd</sup> | 4 <sup>th</sup> | 5 <sup>th</sup> |
| Restriction on entering high risk infected areas            | •               | •               | •               | •               |                 |
| Curfew 9.00pm.–04.00am.                                     |                 |                 |                 | •               |                 |
| Working from home and online learning                       | •               | •               | •               | •               |                 |
| Restriction on opening times for shopping malls and markets | •               | •               | •               | •               |                 |
| Parks and entertainment places closed                       | •               | •               | •               | •               |                 |
| Border closed   | •               |                 |                 |                 |                 |
| 14 days quarantine on international arrival                 |                 | •               | •               | •               | •               |
| Availability of 1 <sup>st</sup> dose vaccine (Aug. 2021)    |                 |                 |                 | •               |                 |
| Availability of 2 <sup>nd</sup> dose vaccine                |                 |                 |                 |                 | •               |

1<sup>st</sup> wave, Mar 2020–Jun 2020; 2<sup>nd</sup> wave, Dec 2020–Jan 2021; 3<sup>rd</sup> wave, Apr 2021–May 2021; 4<sup>th</sup> wave Jun 2021–Dec 2021; and 5<sup>th</sup> wave, Jan 2022–May 2022.

**Fig. 1.** shows the number of infected cases from January 2020 to June 2021. During the 1st and 2nd waves of Covid-19, there were low numbers of new infected cases per day because of the quarantine policy when Covid-19 was an emerging disease, with no developed vaccine. However, after some people received a vaccination and returned to living a normal life, they became familiar with the Covid-19 situation and were less careful and ignored some of the preventative measures. This led to new waves (3rd and 4th) of Covid-19, resulting in increased numbers of infected cases over a short time interval. Concurrently, Covid-19 had evolved to produce mutations that were more severe and more rapidly transmitted. Therefore, during the 4th and 5th waves, the numbers of infected cases increased to 10,000 per day before continuously

declining. Finally in July 2022, the Ministry of Public Health Thailand declared the Covid-19 situation to be fully controlled and Covid-19 was downgraded from pandemic to endemic status. However, Covid-19 pandemic changed the travel behavior of people globally. In Europe, research in Spain (Mobility Patterns: The First Wave’s Results [42], provided evidence of changing travel behavior.



Fig 1. Overview of number of Covid-19 infections (cases/day) in Thailand January 2020–June 2021.

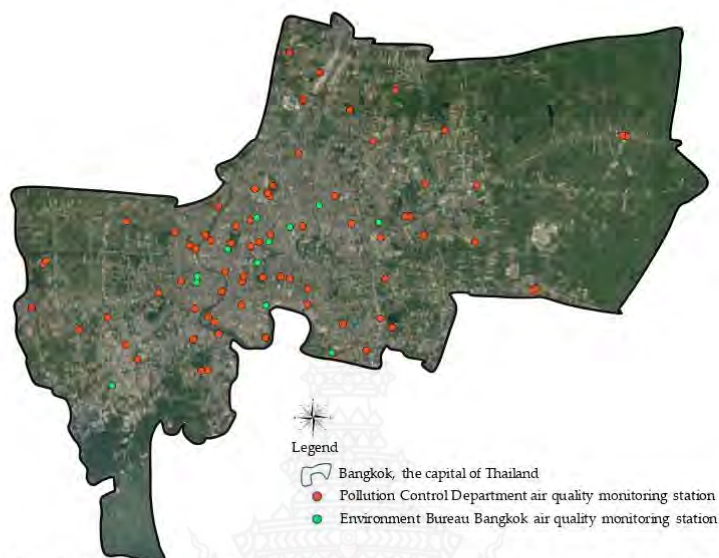
### 3. Methodology and Motivation

The meteorological dataset used in this research contained air pollution and statistical data of expressway traffic volumes in Bangkok. These data were used to identify the correlation between traffic volume and the PM2.5 level during the Covid-19 pandemic. Daily data were obtained from the Bangkok Pollution Control Department and the Expressway Authority, from January 2015 to December 2021.

#### 3.1. Statistic data of meteorology and air pollution

In this research, the meteorological and pollution data of Bangkok areas are collected as shown in Figure 2. Those data are collected from 2 sources which totals 66 stations; 12 stations are from Pollution Control Department (green points in Figure 2) and 50 stations are from Environment Bureau Bangkok (red points in Figure 2). Those inspection tools are aligned to announcement of pollution control department [43] (Measuring Instruments and Methods for Measuring Average Gases or Particulates in the Atmosphere in General Other systems or other methods approved). The instruments are permanently installed two-meter above ground level.





**Fig. 2. Air monitoring stations of Pollution Control Department (50 red points) and Environment Bureau Bangkok (12 green points)**

The meteorological dataset consisted of more than 2.3 million measured values for: PM2.5 (fine dust particles with a size less than 2.5 microns), coarse particles (PM10), nitrogen oxide (NO<sub>x</sub>), nitrogen dioxide (NO<sub>2</sub>), nitric oxide (NO), carbon monoxide (CO), wind speed (WS), wind direction (WD), temperature (Temp), barometric pressure (BP), relative humidity (RH%), and ozone (O<sub>3</sub>) from 1 January 2015 to 31 December 2021. Hourly measurements were taken using appropriate instrumentation and

the results were reported according to standard values. For example, PM2.5 values were collected every hour and then averaged to provide a 24-hour average that was compared with standard values. The measurement and sampling methods followed the Federal Equivalent Method standards. **Fig. 3** is an example some of the dataset used in this study. **Table 2** provides a summary of the raw data obtained from various organization during the study period.

| Key_mth    | Hour | WS(m/s) | WD(Deg) | Temp(Deg.C) | RH(%) | BP(mBar) | CO(ppm) | NO(ppb) | NO2(ppb) | NOX(ppb) | O3(ppb) | PM10(ug/m3) | PM2.5(ug/m3) |
|------------|------|---------|---------|-------------|-------|----------|---------|---------|----------|----------|---------|-------------|--------------|
| 20210101.0 | 1.0  | 1.2     | 81.0    | 22.8        | 57.0  | 763.0    | 0.85    | 7.0     | 20.0     | 27.0     | 22.0    | 43.0        | 15.0         |
| 20210101.0 | 2.0  | 0.8     | 78.0    | 22.2        | 59.0  | 762.0    | 0.99    | 14.0    | 22.0     | 36.0     | 18.0    | 45.0        | 23.0         |
| 20210101.0 | 3.0  | 0.6     | 83.0    | 21.8        | 62.0  | 762.0    | 0.98    | 9.0     | 21.0     | 30.0     | 18.0    | 45.0        | 21.0         |
| 20210101.0 | 5.0  | 0.6     | 32.0    | 20.9        | 62.0  | 762.0    | 0.92    | 11.0    | 21.0     | 32.0     | 15.0    | 46.0        | 33.0         |
| 20210101.0 | 6.0  | 1.2     | 342.0   | 20.2        | 61.0  | 762.0    | 0.86    | 18.0    | 24.0     | 42.0     | 13.0    | 42.0        | 32.0         |
| 20210101.0 | 7.0  | 0.5     | 6.0     | 19.9        | 63.0  | 763.0    | 1.01    | 40.0    | 33.0     | 73.0     | 7.0     | 36.0        | 26.0         |
| 20210101.0 | 8.0  | 0.6     | 6.0     | 20.0        | 64.0  | 764.0    | 1.06    | 58.0    | 33.0     | 91.0     | 6.0     | 41.0        | 27.0         |
| 20210101.0 | 9.0  | 0.6     | 46.0    | 20.5        | 64.0  | 765.0    | 1.04    | 49.0    | 33.0     | 60.0     | 10.0    | 47.0        | 28.0         |
| 20210101.0 | 10.0 | 0.9     | 13.0    | 21.8        | 60.0  | 765.0    | 1.02    | 44.0    | 28.0     | 71.0     | 14.0    | 52.0        | 33.0         |
| 20210101.0 | 11.0 | 1.1     | 22.0    | 23.3        | 56.0  | 765.0    | 1.01    | 32.0    | 28.0     | 60.0     | 18.0    | 52.0        | 33.0         |

Fig. 3 Example of dataset used in this study.

Table 2 Summary of data sources.

| Data Source   | Number of stations | Data records (2015–2021) |
|---|--------------------|--------------------------|
| Pollution Control Department air quality monitoring station | 13                 | 787,550                  |
| Environment Bureau Bangkok air quality monitoring station   | 46                 | 1,516,342                |
| <b>Total</b>  | <b>59</b>          | <b>2,303,892</b>         |

Then, data cleansing process was conducted to obtain the data used in the analysis. Data within the 99<sup>th</sup> percentile were used. This meteorological dataset was used to not only help to identify influential factors of pollution source regarding air quality but also to help directly reduce pollution from the origin source [44], [45].

The meteorological data were separated into 2 datasets: 1) before the Covid-19 pandemic (2015–2019); and 2) during the Covid-19 lockdown. (2020–2021) Table 3 provides an overview of the various pollution concentrations before the Covid-19 outbreak and during the lockdown. The authors compared the hourly average pollution concentration data from the 63 monitoring stations in Bangkok. There were differences and trends in the air pollution concentrations before the Covid-19 epidemic

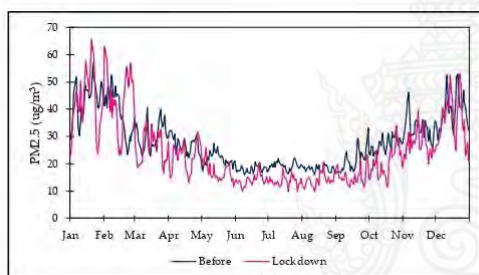
and during the lockdown. During the Covid-19 outbreak, the average values for PM2.5, NOx, NO2, NO, and CO were lower than before the Covid-19 outbreak, while the average PM10 and O3 values increased during the Covid-19 outbreak.

During the lockdown period, there were fewer cars on roads. Mobility was restricted and most people stayed at home. During this time, the concentrations of PM2.5, NOx, NO2, NO, and CO decreased 14.38, 12.93, 18.4, 2.65, and 5.66, respectively. In addition, we compared the average hourly concentration of air pollution before the pan-demic and during the lockdown period.

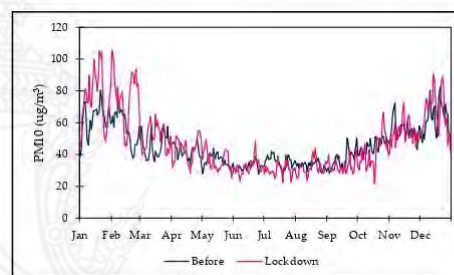
Figure 3 (A) PM2.5 (B) PM10 (C) NOx (D) NO2 (E) NO (F) CO (G) O3 shown the concentration value of each parameter. after analyzed and compared on the concentration of pollution in dry season, from November to march both before pandemic and during lock down period, found that concentration level of pollution in both periods is correspondingly in same direction

Table 3 Statistic data on air pollution used in this study.

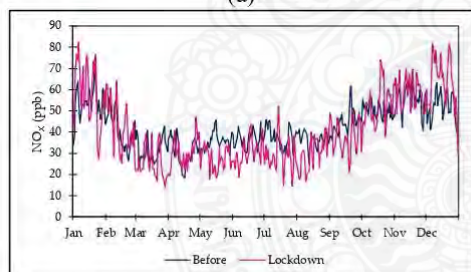
| Factor                     | Before Covid-19 pandemic |       | During lockdown period |       | Difference (%) |
|----------------------------|--------------------------|-------|------------------------|-------|----------------|
|                            | Mean                     | SD    | Mean                   | SD    |                |
| PM2.5 (ug/m <sup>3</sup> ) | 27.93                    | 18.94 | 23.92                  | 16.21 | -14.38         |
| PM10 (ug/m <sup>3</sup> )  | 44.97                    | 26.19 | 47.45                  | 27.94 | 5.51           |
| NO <sub>x</sub> (ppb)      | 42.15                    | 40.88 | 36.70                  | 41.32 | -12.93         |
| NO <sub>2</sub> (ppb)      | 22.15                    | 15.10 | 18.07                  | 14.35 | -18.43         |
| NO (ppb)                   | 23.00                    | 32.96 | 22.39                  | 33.93 | -2.65          |
| CO (ppm)                   | 0.85                     | 0.55  | 0.80                   | 0.55  | -5.66          |
| WS (m/s)                   | 0.84                     | 0.66  | 0.65                   | 0.52  | -23.40         |
| WD (Deg)                   | 188.56                   | 92.74 | 186.22                 | 97.33 | -1.24          |
| Temp (Deg.C)               | 29.42                    | 2.92  | 29.18                  | 3.09  | -0.82          |
| BP (mBar)                  | 869.66                   | 59.13 | 978.40                 | 83.61 | 12.50          |



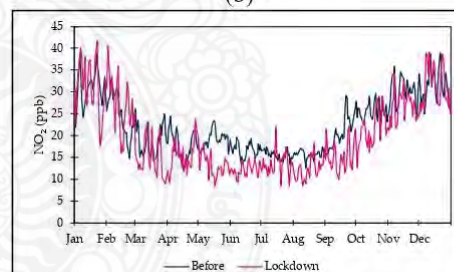
(a)



(b)



(c)



(d)

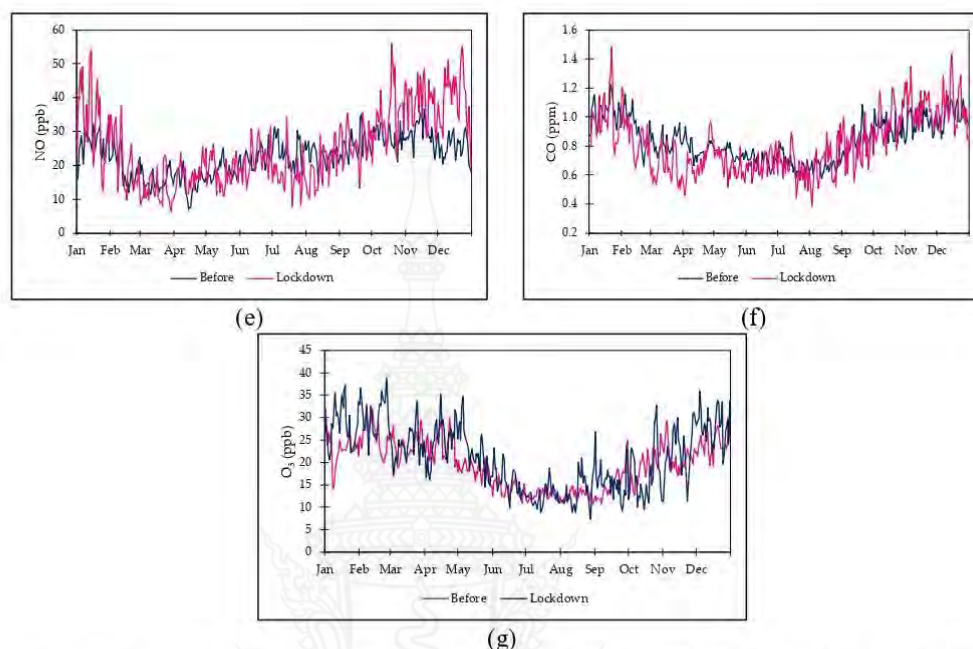


Figure 1. Hourly average concentration level of each parameter before period of Covid-19 pandemic and during period of Lockdown (A) PM<sub>2.5</sub> (B) PM<sub>10</sub> (C) NO<sub>x</sub> (D) NO<sub>2</sub> (E) NO (F) CO (G) O<sub>3</sub>.

### 3.2. Kruskal-Wallis Test

The Kruskal-wallis one-way analysis of variance method was used to identify the characteristics of the distribution of the time series data [46], [47]. As the Kruskal-Wallis method is commonly used for identifying the distribution of air pollution data, it has been proven effective at monitoring the seasonal distribution data of air quality and dominant air pollutants [48], [49]. Initially in the characteristic distribution testing, the pollution data characteristic was assumed to be seasonal using the hypotheses:

H<sub>0</sub>: Median value of each pollution parameter is equal every month

H<sub>1</sub>: Median value of each pollution parameter differs for at least 1 month.

H was calculated using equation (1):

$$H = \left[ \frac{12}{n(n+1)} \sum_{j=1}^c \frac{T_j^2}{n_j} \right] - 3(n+1) \quad (1)$$

based on the sum of sample sizes for all samples ( $n$ ), the number of samples ( $c$ ), the sum of ranks in the  $j^{\text{th}}$  sample ( $T_j$ ), and the size of the  $j^{\text{th}}$  sample ( $n_j$ ).

Table 4 shows the results, indicating that the median of each pollution parameter differed in at least 1 pair in each monthly

result, with a P-value lower than 0.05. Therefore, hypothesis H1 was accepted that pollution parameters each month before the pandemic and during the lockdown period were significantly different.

**Table 4** Results of Kruskal-Wallis test on pollution parameters (January 2015 to December 2021).

| Parameter                  | Statistic   |             |
|----------------------------|-------------|-------------|
|                            | H statistic | H statistic |
| PM2.5 (ug/m <sup>3</sup> ) | 48.57       | 0.000       |
| PM10 (ug/m <sup>3</sup> )  | 49.37       | 0.000       |
| NOx (ppb)                  | 28.62       | 0.000       |
| NO <sub>2</sub> (ppb)      | 28.94       | 0.000       |
| NO (ppb)                   | 44.48       | 0.000       |
| CO (ppm)                   | 36.69       | 0.000       |
| O <sub>3</sub> (ppb)       | 48.47       | 0.000       |

In summary, the concentration of PM2.5 was significantly influenced by seasonal factors.

Some research has shown that decreasing road traffic volume is not the major cause of decreasing air pollution level, with general human activity more directly related to the pollution level [50].

Nonetheless, research in Italy [51], studying the correlation between road traffic and air pollution, indicated the effect of traffic on the concentration of air pollution; however, the research focus was on NO, NO<sub>2</sub>, and NO<sub>x</sub> (PM2.5 was not mentioned). Therefore, it was not possible to conclusively determine the relationship between traffic and the concentration of PM2.5.

### 3.3. The relationship between traffic density and PM2.5

The daily traffic data in the metropolitan area, shown in Fig. 4., covered the Chalerm Maha Nakhon Expressway, the Si Rat Expressway, Chalong Rat Expressway, and the Burapha Withi Expressway. The travel data showed under normal conditions, the traffic volume on expressways reached 1.78 million vehicles per day. However, during the lockdown period (2020–2021), the traffic volume on expressways dropped to 1.36 million vehicles per day.

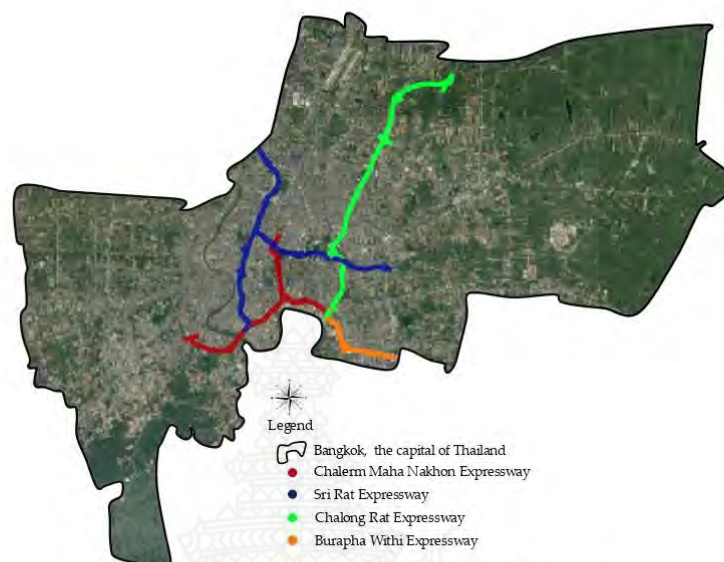


Fig. 4. Overview of 3 urban expressways in Bangkok.

In Fig. 5., the daily volume of vehicles are separated into 2 periods: before the Covid-19 pandemic (January 2015 to December 2019), and during the Covid-19 outbreak (January 2020 to December 2021). In Thailand, the Covid-19 pandemic began in January 2020, and the lockdown that

prohibited people from going out for non-essential activities was implemented from April to June 2020 and was considered to be most effective as indicated by the approximately 40% reduction in travel volume that occurred during this time.

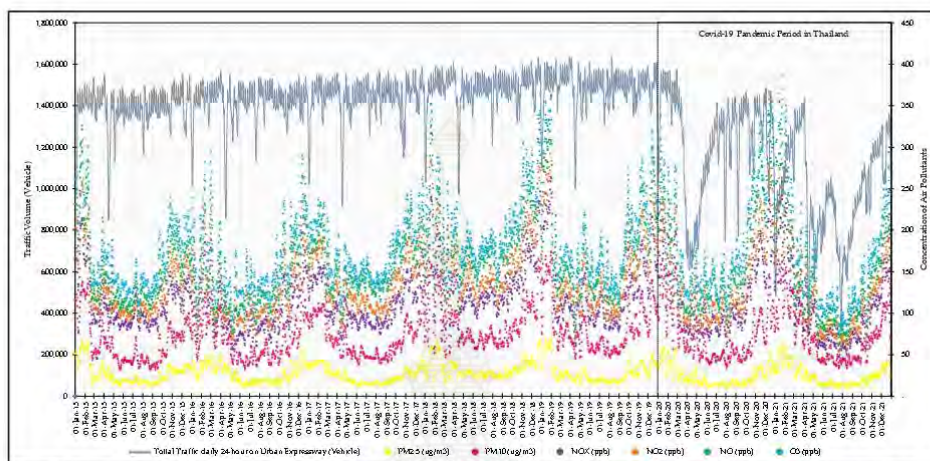


Fig. 5. shown the daily traffic volume and concentration of air pollution of 2 periods; Before pandemic of Covid-19 (January 2015 to December 2019), and During outbreak of Covid- (January 2020 to De-cember 2021).

**4. Results and Discussion**

Considering relationship between traffic volume and PM2.5 in Figure 6, before pandemic of covid-19 (January 2015 to December 2019), the traffic volume each day is on the same range and PM2.5 concentration level depend on season factor. As every year, PM2.5 is highest in dry season (from November to March).

During pandemic of Covid-19, lockdown caused traffic volume rapidly decreased (from March 2020). In case, considering only the data before March 2020, Author might inevitably mislead by the data and conclude that PM2.5 and traffic are significantly related. However, considering all data in this research (from 2015 to January 2021), which is dry season and there was 5th wave Covid-19 which affect to reduction of traffic volume again. The analysis data shown that PM2.5 concentration level is highest which its result inversing traffic

volume. Especially, during the dry season (January) when traffic volumes are reduced but PM2.5 is noticeably higher (square box highlighted in yellow in Figure 6).

In addition, the statistical data shown that the correlation coefficient between traffic volume and concentration of PM2.5 is at 0.206.

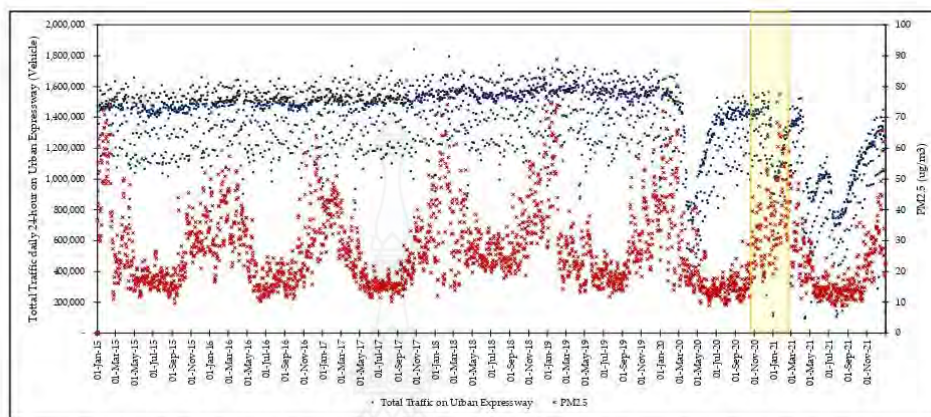


Figure 6. shown daily traffic volume and concentration of PM2.5 in 2 periods; Before pandemic of covid-19 (January 2020 to December 2019), and outbreak of covid-19 (January 2020 to December 2021).

### 5. Conclusions

This study examined the changes in PM2.5 levels as a result of the Covid-19 outbreak. The researchers used data on the levels of PM2.5 and other air pollution parameters gathered at 63 points in addition to traffic volume data on 4 expressways from January 2015 to December 2021.

The results showed that during the city lockdown due to the Covid-19 pandemic, the percentage of concentration of the air pollution parameters PM2.5 NOX NO2 NO, and CO, decreased by 14.38, 12.93, 18.43, 2.65, and 5.66 percent, respectively. In addition, testing the time series air pollution dataset using the by Kruskal-Wallis method indicated that the data directly related to seasonal distribution, with the concentrations of pollution being highest during the dry season every year.

In addition, the city lockdown forced people to stay at home and to work and learn online. This directly reduced the traffic volume by 40 percent. The average monthly PM2.5 level dropped by 21 percent from

before the Covid-19 outbreak compared to during the lockdown period. This indicated that during the outbreak, the lockdown had a direct impact on human behavior, particularly transportation, which decreased by 40%. These changes played a substantial role in the reduction of PM2.5 concentrations.

While it was not possible to clearly determine which factors (traffic or seasonal) had the greater impact on the PM2.5 concentration, this research provided evidence that human activities, specifically traffic in urban regions, had a notable effect on PM2.5 levels.

The PM2.5 problem has been of concern in Thailand, with government and academic studies attributing the main cause of increased PM2.5 levels to traffic. The current research results showed that decreasing the traffic volume significantly reduced PM2.5 levels. The authors believe that this research can contribute to policy and stakeholder reviews on how to sustainably solve the PM2.5 problem in Thailand in a more



appropriate way. However, the data collected represent only a sample from a part of Thailand.

## 6. Acknowledgement

The completion of this study would not have been possible without the assistance of the Rajamagala University of Technology Phra Nakhon. The authors also thank the Pollution Control Department, the Environment Bureau Bangkok, and the Expressway Authority of Thailand for providing support in data collection.

## 7. References

- [1] U. S. Epa, "Criteria Air Pollutants," *Encycl. Immunotoxicol.*, no. October, pp. 218–218, 2016, doi: 10.1007/978-3-642-54596-2\_200326.
- [2] V. Ratanavaraha and S. Jomnonkwao, "Trends in Thailand CO2 emissions in the transportation sector and Policy Mitigation," *Transp. Policy*, vol. 41, pp. 136–146, 2015, doi: 10.1016/j.tranpol.2015.01.007.
- [3] C. Li and S. Managi, "Contribution of on-road transportation to PM2.5," *Sci. Rep.*, vol. 11, no. 1, pp. 1–12, 2021, doi: 10.1038/s41598-021-00862-x.
- [4] S. Heydari, M. Tainio, J. Woodcock, and A. de Nazelle, "Estimating traffic contribution to particulate matter concentration in urban areas using a multilevel Bayesian meta-regression approach," *Environ. Int.*, vol. 141, no. December 2019, p. 105800, 2020, doi: 10.1016/j.envint.2020.105800.
- [5] W. M. Hodan and W. R. Barnard, "Evaluating the Contribution of PM2.5 Precursor Gases and Re-entrained Road Emissions to Mobile Source PM2.5 Particulate Matter Emissions. MACTEC Under Contract to the Federal Highway Administration," *MACTEC Fed. Programs, Res. Triangle Park. NC*, p. 58 pp, 2004, [Online]. Available: <https://www3.epa.gov/ttnchie1/conference/ei13/mobile/hodan.pdf>.
- [6] R. Sun, Y. Zhou, J. Wu, and Z. Gong, "Influencing factors of PM2.5 pollution: Disaster points of meteorological factors," *Int. J. Environ. Res. Public Health*, vol. 16, no. 20, 2019, doi: 10.3390/ijerph16203891.
- [7] R. Li, X. Mei, L. Wei, X. Han, M. Zhang, and Y. Jing, "Study on the contribution of transport to PM2.5 in typical regions of China using the regional air quality model RAMS-CMAQ," *Atmos. Environ.*, vol. 214, no. 40, p. 116856, 2019, doi: 10.1016/j.atmosenv.2019.116856.
- [8] Engel, "World air quality report 2021," *Pap. Knowl. . Towar. a Media Hist. Doc.*, 2014.
- [9] M. Station, K. Findings, K. Kaen, C. Mai, and C. Mai, "Greenpeace 's City Rankings for PM2.5 in Thailand," no. 2013, pp. 1–6, 2016.
- [10] Tomtom Traffic Index, "Traffic Index Traffic Index results 2018 - 2021," *Www.Tomtom.Com*, p. 21, 2021, [Online]. Available: <https://www.tomtom.com/traffic-index/ranking/>.
- [11] B. Pishue and T. Trepanier, "2020 INRIX Global Traffic Scorecard and latest travel trends," no. March, 2021.
- [12] A. König and A. Dreßler, "A mixed-methods analysis of mobility behavior changes in the COVID-19 era in a rural case study," *Eur. Transp. Res.*

- Rev., vol. 13, no. 1, 2021, doi: 10.1186/s12544-021-00472-8.
- [13] M. Hasselwander, T. Tamagusko, J. F. Bigotte, A. Ferreira, A. Mejia, and E. J. S. Ferranti, "Building back better: The COVID-19 pandemic and transport policy implications for a developing megacity," *Sustain. Cities Soc.*, vol. 69, no. October 2020, 2021, doi: 10.1016/j.scs.2021.102864.
- [14] Q. Cui *et al.*, "The impacts of COVID-19 pandemic on China's transport sectors based on the CGE model coupled with a decomposition analysis approach," *Transp. Policy*, vol. 103, no. January, pp. 103–115, 2021, doi: 10.1016/j.tranpol.2021.01.017.
- [15] "People's travel During the COVID-19 viral epidemic - dataset - MOT Data Catalog." <https://datagov.mot.go.th/dataset/greenhouse-emissions-in-the-road-transport-sector>.
- [16] "Annual Traffic Report 2021, EXAT." <https://www.exat.co.th/statistics/>.
- [17] "Air quality and noise statistics, Pollution Control Department." <http://air4thai.pcd.go.th/webV3/#/History>.
- [18] J. D. Sacks *et al.*, "Particulate matter-induced health effects: who is susceptible?," *Environ. Health Perspect.*, vol. 119, no. 4, pp. 446–454, Apr. 2011, doi: 10.1289/ehp.1002255.
- [19] M. Rosenlund, F. Forastiere, D. Porta, M. De Sario, C. Badaloni, and C. A. Perucci, "Traffic-related air pollution in relation to respiratory symptoms, allergic sensitisation and lung function in schoolchildren," *Thorax*, vol. 64, no. 7, pp. 573–580, 2009, doi: 10.1136/thx.2007.094953.
- [20] G. Chen *et al.*, "Short-term effects of ambient gaseous pollutants and particulate matter on daily mortality in Shanghai, China.," *J. Occup. Health*, vol. 50, no. 1, pp. 41–47, 2008, doi: 10.1539/joh.50.41.
- [21] J. Cao, J. C. Chow, F. S. C. Lee, and J. G. Watson, "Evolution of PM2.5 measurements and standards in the U.S. And future perspectives for China," *Aerosol Air Qual. Res.*, vol. 13, no. 4, pp. 1197–1211, 2013, doi: 10.4209/aaqr.2012.11.0302.
- [22] M. Kanakidou *et al.*, "Organic aerosol and global climate modelling: A review," *Atmos. Chem. Phys.*, vol. 5, no. 4, pp. 1053–1123, 2005, doi: 10.5194/acp-5-1053-2005.
- [23] M. Hallquist *et al.*, "The formation, properties and impact of secondary organic aerosol: Current and emerging issues," *Atmos. Chem. Phys.*, vol. 9, no. 14, pp. 5155–5236, 2009, doi: 10.5194/acp-9-5155-2009.
- [24] X. Liu *et al.*, "Secondary formation of sulfate and nitrate during a haze episode in megacity Beijing, China," *Aerosol Air Qual. Res.*, vol. 15, no. 6, pp. 2246–2257, 2015, doi: 10.4209/aaqr.2014.12.0321.
- [25] R. M. Harrison and J. Yin, "Particulate matter in the atmosphere: Which particle properties are important for its effects on health?," *Sci. Total Environ.*, vol. 249, no. 1–3, pp. 85–101, 2000, doi: 10.1016/S0048-9697(99)00513-6.
- [26] U. Pöschl, "Atmospheric aerosols: Composition, transformation, climate and health effects," *Angew. Chemie - Int. Ed.*, vol. 44, no. 46, pp. 7520–7540, 2005, doi: 10.1002/anie.200501122.

- [27] I. Z. A. D. P. No, H. H. Dang, T. Trinh, and H. H. Dang, "DISCUSSION PAPER SERIES Does the COVID-19 Pandemic Improve Global Air Quality? New Cross-National Evidence on Its Unintended Consequences," no. 13480, 2020.
- [28] P. Dasgupta and K. Srikanth, "Reduced air pollution during COVID-19: Learnings for sustainability from Indian Cities," *Glob. Transitions*, vol. 2, pp. 271–282, 2020, doi: 10.1016/j.glt.2020.10.002.
- [29] Z. S. Venter, K. Aunan, S. Chowdhury, and J. Lelieveld, "COVID-19 lockdowns cause global air pollution declines," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 117, no. 32, pp. 18984–18990, 2020, doi: 10.1073/pnas.2006853117.
- [30] A. Farrow and E. Newport, "The Burden of Air Pollution in Thailand 2021 The Burden of Air Pollution in Thailand 2021," no. May, 2022.
- [31] P. C. Department, "Bangkok Air Pollution: Real-time Air Quality Index (AQI)." pp. 1–5, 2021, [Online]. Available: <https://aqicn.org/city/bangkok>.
- [32] T. Amnuaylojaroen and J. Kreasuwun, "Investigation of fine and coarse particulate matter from burning areas in Chiang Mai, Thailand using the WRF/CALPUFF," *Chiang Mai J. Sci.*, vol. 39, no. 2, pp. 311–326, 2012.
- [33] S. Khodmanee and T. Amnuaylojaroen, "Impact of Biomass Burning on Ozone, Carbon Monoxide, and Nitrogen Dioxide in Northern Thailand," *Front. Environ. Sci.*, vol. 9, no. April, pp. 1–6, 2021, doi: 10.3389/fenvs.2021.641877.
- [34] T. Amnuaylojaroen, "Prediction of PM<sub>2.5</sub> in an Urban Area of Northern Thailand Using Multivariate Linear Regression Model," *Adv. Meteorol.*, vol. 2022, p. 3190484, 2022, doi: 10.1155/2022/3190484.
- [35] S. Chirasophon and P. Pochanart, "The long-term characteristics of PM10 and PM2.5 in Bangkok, Thailand," *Asian J. Atmos. Environ.*, vol. 14, no. 1, pp. 73–83, 2020, doi: 10.5572/AJAE.2020.14.1.073.
- [36] D. Narita *et al.*, "Pollution characteristics and policy actions on fine particulate matter in a growing Asian economy: The case of Bangkok Metropolitan Region," *Atmosphere (Basel)*, vol. 10, no. 5, pp. 1–18, 2019, doi: 10.3390/atmos10050227.
- [37] P. Uttamang, V. P. Aneja, and A. F. Hanna, "Assessment of gaseous criteria pollutants in the Bangkok Metropolitan Region, Thailand," *Atmos. Chem. Phys.*, vol. 18, no. 16, pp. 12581–12593, 2018, doi: 10.5194/acp-18-12581-2018.
- [38] P. Watcharaviton, C. P. Chio, and C. C. Chan, "Temporal and spatial variations in ambient air quality during 1996-2009 in Bangkok, Thailand," *Aerosol Air Qual. Res.*, vol. 13, no. 6, pp. 1741–1754, 2013, doi: 10.4209/aaqr.2012.11.0305.
- [39] Eea, "EMEP/EEA air pollutant emission inventory guidebook 2013: Technical guidance to prepare national emission inventories," *EEA Tech. Rep.*, no. 12/2013, p. 23, 2013, [Online]. Available: <http://www.eea.europa.eu/publications/emep-eea-guidebook-2013>.
- [40] Y. Wang and D. Yang, "Impacts of freight transport on PM2.5 concentrations in China: A spatial

- dynamic panel analysis,” *Sustain.*, vol. 10, no. 8, 2018, doi: 10.3390/su10082865.
- [41] T. Xia, M. Nitschke, Y. Zhang, P. Shah, S. Crabb, and A. Hansen, “Traffic-related air pollution and health co-benefits of alternative transport in Adelaide, South Australia,” *Environ. Int.*, vol. 74, pp. 281–290, 2015, doi: 10.1016/j.envint.2014.10.004.
- [42] B. J. Tunno *et al.*, “Spatial patterns in rush-hour vs. Work-week diesel-related pollution across a downtown core,” *Int. J. Environ. Res. Public Health*, vol. 15, no. 9, 2018, doi: 10.3390/ijerph15091968.
- [43] D. S. Hui *et al.*, “The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health — The latest 2019 novel coronavirus outbreak in Wuhan, China,” *Int. J. Infect. Dis.*, vol. 91, pp. 264–266, 2020, doi: 10.1016/j.ijid.2020.01.009.
- [44] L. Pérez-Campos Mayoral, M. T. Hernández-Huerta, G. Mayoral-Andrade, E. Pérez-Campos Mayoral, and E. Pérez-Campos, “World Health Organization declares global emergency: A review of the 2019 novel Coronavirus (COVID-19),” *Int. J. Surg.*, vol. 79, no. February, pp. 163–164, 2020, doi: 10.1016/j.ijisu.2020.05.066.
- [45] A. E. Dingil and D. Esztergár-Kiss, “The Influence of the Covid-19 Pandemic on Mobility Patterns: The First Wave’s Results,” *Transp. Lett.*, pp. 1–13, Mar. 2021, doi: 10.1080/19427867.2021.1901011.
- [46] “Instruments and methods for measuring the average value of gas or dust in the atmosphere in general - Pollution Control Department.” <https://www.pcd.go.th/publication/5420/>.
- [47] P. Wetchayont, “Investigation on the Impacts of COVID-19 Lockdown and Influencing Factors on Air Quality in Greater Bangkok, Thailand,” *Adv. Meteorol.*, vol. 2021, no. 2, 2021, doi: 10.1155/2021/6697707.
- [48] Y. Miao, S. Liu, L. Sheng, S. Huang, and J. Li, “Influence of Boundary Layer Structure and Low-Level Jet on PM<sub>2.5</sub> Pollution in Beijing: A Case Study,” *Int. J. Environ. Res. Public Health*, vol. 16, no. 4, 2019, doi: 10.3390/ijerph16040616.
- [49] W. H. Kruskal and W. A. Wallis, “Use of Ranks in One-Criterion Variance Analysis,” *J. Am. Stat. Assoc.*, vol. 47, no. 260, pp. 583–621, 1952, doi: 10.1080/01621459.1952.10483441.
- [50] T. W. MacFarland and J. M. Yates, *Introduction to Nonparametric Statistics for the Biological Sciences Using R*, 1st ed. Springer Publishing Company, Incorporated, 2018.
- [51] X. Zhao *et al.*, “Statistical analysis of spatiotemporal heterogeneity of the distribution of air quality and dominant air pollutants and the effect factors in Qingdao Urban Zones,” *Atmosphere (Basel)*, vol. 9, no. 4, pp. 12–16, 2018, doi: 10.3390/atmos9040135.
- [52] K. Webel, “A Data-Driven Selection of an Appropriate Seasonal Adjustment Approach,” *SSRN Electron. J.*, no. 07, 2021, doi: 10.2139/ssrn.2797077.
- [53] S. Dursun, M. Sagdic, and H. Toros, “The impact of COVID-19 measures on air quality in Turkey,” *Environ. Forensics*, vol. 23, no. 1–2, pp. 47–59,

*RMUTP Research Journal, Vol. ...., No. ....*

- 2022, doi:  
10.1080/15275922.2021.1892876.
- [54] R. Rossi, R. Ceccato, and M. Gastaldi,  
“Effect of Road Traffic on Air  
Pollution. Experimental Evidence  
from COVID-19 Lockdown,”  
*Sustainability*, vol. 12, no. 21, 2020,  
doi: 10.3390/su12218984.



ภาพ ก-2 แบบตอบรับการตีพิมพ์บทความใน GMSARN International Journal เรื่อง Forecasting Air Pollution Particulate Matter (PM 2.5) Prior to and During COVID-19 in Bangkok, Thailand Using Machine Learning Regression Models

Greater Mekong Subregion Academic and Research Network (GMSARN)





Asian Institute of Technology



Hanoi University of Technology



Ho Chi Minh City University of Technology



Institute of Technology Cambodia



Khon Kaen University



Kunming University of Science and Technology



Chitralada Technology Institute

September 22, 2023 No. GMSARNJ 2023/20

**Krit Jedwanna**  
Department of Civil Engineering, Faculty of Engineering,  
Rajamangala University of Technology Phra Nakhon,  
Bangkok, Thailand

Dear Krit Jedwanna,

**Subject: Acceptance Letter (Paper ID. GMSARN-23Env-04)**

Title of the Journal: GMSARN International Journal  
Paper Title: Forecasting Air Pollution Particulate Matter (PM2.5) Prior to and During Covid-19 in Bangkok, Thailand Using Machine Learning Regression Models

Authors: Krit Jedwanna, Suwit Paengkanya, Prin Boonkanit  
Corresponding Author: Krit Jedwanna

Thank you very much for your submission to our journal. We are pleased to inform you that your paper has been reviewed and accepted with revision for publication in Volume 19 Issue 4, 2025. In case you have not submitted copyright form; please send scanned copy shortly through e-mail: [gmsarn@ait.ac.th](mailto:gmsarn@ait.ac.th). Thank you for making the journal a vehicle for your research interests.

Best wishes,



**Prof. Weerakorn Ongsakul, PhD, CFA**  
Editor-in-Chief  
GMSARN International Journal

GMSARN Office: Asian Institute of Technology

|   |   |
|---|---|
| Postal Address:<br>P.O. Box 4<br>Klong Luang<br>Pathumthani 12120<br>Thailand | Street Address:<br>Km. 42 Paholyothin Highway<br>Klong Luang<br>Pathumthani 12120<br>Thailand |
|---|---|

Tel: (66-2) 524-6537  
(for local calls, dial 02 before the tel no.)  
E-mail: [gmsarn@ait.ac.th](mailto:gmsarn@ait.ac.th)  
Internet: <http://www.gmsarn.com>



National University of Laos



Royal University of Phnom Penh



Thammasat University



Yongon Technological University



Yunnan University



Naresuan University



Surathani Rajabhat University



## Forecasting Air Pollution Particulate Matter (PM<sub>2.5</sub>) Prior to and During Covid-19 in Bangkok, Thailand Using Machine Learning Regression Models

Krit Jedwanna<sup>1\*</sup>, Suwit Paengkanya<sup>2</sup>, and Prin Boonkanit<sup>3</sup>

### ARTICLE INFO

#### Article history:

Received:

Revised:

Accepted:

#### Keywords:

Modeling

AI

Prediction PM<sub>2.5</sub>

Covid-19 Pandemic

### ABSTRACT

Particulate Matter (PM<sub>2.5</sub>) was predicted based on 6 machine-learning models (Decision Tree, Gradient Boosting, K-neighbors, Multi-layer Perception, Random Forest, and Ridge) using a PM<sub>2.5</sub> time-series dataset, consisting of 1.6 million records from 1 January 2015 to 31 December 2021. The study compared 2 testout datasets for each model, with one being for the normal period before the Covid-19 pandemic (2015–2019) and the other for the Covid-19 pandemic period (2020–2021). All models were based on historical data features. The results showed that Gradient Boosting by the hour provided the best results for the training dataset ( $R^2=0.8284$ ), the test dataset ( $R^2=0.8051$ ), and the testout dataset ( $R^2=0.8053$ ). Therefore, the Gradient Boosting model was considered as the best choice for the prediction of PM<sub>2.5</sub> using historical data by the hour before and during the Covid-19 pandemic, despite its variable data. However, this research was based only on a dataset for Bangkok, Thailand.

<sup>1,2,3</sup>Sustainable Industrial Management Engineering, Faculty of Engineering, Rajamangala University of Technology Phra Nakhon, Bangkok, Thailand.  
\*Corresponding author: Krit Jedwanna, Phone: +66-897-771-654; Email: krit.j@mutp.ac.th

## 1. INTRODUCTION

The USEPA (United States Environmental Protection Agency) [1] has set a standard for small particles considered harmful to human health using a PM (particulate matter) value. PM10 (coarse particles) defines the small particles found in dust and smoke having a diameter in the range 2.5–10  $\mu\text{m}$  which can be clearly seen. PM2.5 (fine particles) are not visible to the naked eye and have a diameter of 2.5  $\mu\text{m}$  or less.

There is much published evidence from studies over the world showing that traffic is the main factor contributing to PM2.5 levels [2–7]. Currently, pollution caused by PM2.5 has become a severe problem in Thailand, especially in metropolitan areas such as Bangkok, Chiang Mai, Chiang Rai and Lampang [8,9]. Traffic jams in Thailand remain a major concern, with Bangkok being among the top ranking cites for traffic congestion over a long period [10,11]. Researchers have investigated this problem over the past decades, with many studies and analytical model being developed to identify workable solutions in specific areas. As a traditional method, analytical models were developed using available statistical data. Currently, thanks to technology advancement, re-searchers can apply large datasets of situations and use machine learning in modeling. Therefore, air quality prediction models have become more accurate and efficient in estimating the impact of future problems.

During the last 2–3 years, the Covid-19 pandemic has substantially reduced human activities, while overall the environment has been able to recover somewhat [12–14]. Therefore, applying PM2.5 and gaseous pollution during this period to predict future situations may change the main factors causing PM2.5, since there are major differences based on comparing the data on the relationship between PM2.5 and gaseous pollution during the Covid-19 pandemic with levels before the pandemic.

Correlation between PM2.5 and gaseous pollutants: Pearson's correlation coefficient [15] has been applied to determine the relationship between PM2.5 and gaseous pollution before and during the Covid-19 pandemic. This method is commonly applied in such investigations because it is easily understood and readily calculated. Its value is within the range from -1.0 to 1.0, with values close to -1.0 indicating an inverse correlation, while values close to +1.0 indicating a direct correlation. A Pearson's correlation value equal to 0 means there is no correlation. The higher the value, the stronger the linear relationship between the 2 variables.

Table 1 shows the correlation between PM2.5 and gas pollutant pre and post the Covid-19 pandemic. Both before and during the Covid-19 pandemic, PM2.5 and PM10 were most closely related, followed by NO<sub>2</sub>, CO, and NO<sub>x</sub>, respectively. However, there was significant relationship between from the reduction in human activities between PM2.5 and gaseous pollutants regardless of whether before

or during the Covid-19 pandemic. With reference to gaseous pollutants, human activities may affect the concentration of PM2.5; however, seasonal factors had a greater significant impact based on analysis using Kruskal-Wallis one-way analysis.

**Table 1. Correlation between PM2.5 and gaseous pollutants**

| Feature           | Pearson's correlation |            |
|-------------------|-----------------------|------------|
|                   | (Normal Situation)    | (Covid-19) |
|                   | PM2.5                 | PM2.5      |
| PM10              | 0.792                 | 0.906      |
| CO                | 0.355                 | 0.455      |
| NO                | 0.268                 | 0.318      |
| NO <sub>2</sub>   | 0.470                 | 0.563      |
| NO <sub>x</sub>   | 0.380                 | 0.429      |
| O <sub>3</sub>    | 0.082                 | 0.149      |
| Wind Speed        | -0.138                | -0.115     |
| Wind Direction    | -0.103                | -0.084     |
| Temperature       | -0.079                | -0.141     |
| Relative Humidity | -0.224                | -0.271     |
| Air Pressure      | -0.073                | -0.002     |

The meteorological database for Bangkok, Thailand, contains the concentration of gaseous pollutants for 7 years (2015–2022), with the beginning of 2021 coinciding with the start of the Covid-19 outbreak. Thus, machine learning techniques were applied to compare the analyzed data during both periods (normal situation and during Covid-19 pandemic) to predict particulate matter (PM2.5) in the future.

## 2. LITERATURE REVIEW

### 2.1. PM 2.5 emissions in Thailand

Thailand Pollution control department has estimated the number of human deaths in Thailand due to PM2.5 may be as high as 22,000 persons [16]. Considering the air quality index (AQI) values in Thailand [17] during early 2021 was undoubtedly the reason for the predicted number of deaths being is high, with human activities the major driver of the high AQI values.

The main sources of PM2.5 are from transportation emissions, especially smoke from exhaust pipes due to the incomplete combustion of the fuel in major cities, such as Bangkok, Chiang Mai, and Khon Kaen, where the high traffic density leads to especially high levels of PM2.5 [9]. In 2021, Thailand was ranked fifth in the Asian region after Indonesia (ranked first), followed by Myanmar, Vietnam, and Laos, respectively [8]. In 2021, Lampang recorded the highest peak level of PM2.5 in Thailand. However, the big cities, such as Bangkok and Chiang Mai, face severe PM2.5 problems throughout extended periods. Furthermore, the frequency and severity of PM2.5 levels have been constantly increasing.



especially in the dry season [18,19].

The PM<sub>2.5</sub> problem in areas such as Bangkok is worse during the dry season. Other studies study have shown that the main sources of high PM<sub>2.5</sub> levels in Bangkok were from biomass burning, transportation, and industrial activities, with the concentration levels varying due to seasonal factors [16,20,21]. A study on the concentrations of gaseous pollutants (O<sub>3</sub>, NO<sub>x</sub>, CO, and SO<sub>2</sub>) from 1996 to 2009 in Bangkok in residential areas and along roads found that air pollution levels rose during the dry season and dropped during the rainy and summer seasons. Furthermore, the roadside concentration levels were clearly greater than for residential areas [22].

## 2.2. Analysis and prediction of air pollutants from past to present

In the past, the prediction and analysis of air pollutants commonly involved combining statistical data with traditional based mathematical methods. However, such methods have some restrictions, including on the number of factors used for analysis and the dataset must not be too large. Therefore, the accuracy of the modeling may not have adequately represented the actual situation [23–27]. Kalman filtering and single variable linear regression is the primary method for predicting air quality and pollution levels [28,29].

Currently, air quality analysis tends to be more complex due to the increasing population and greater numbers of social activities. Many effects on air quality from increasing biological and physical diversity are still in doubt, as they may cause meteorological changes through releasing greenhouse gas. Therefore, machine learning and neural network analysis may be applied to predict PM<sub>2.5</sub> levels, especially as new machine learning models can easily analyze many aspects jointly [30–34]. Additionally, the results from these new methods have improved the prediction accuracy [35]. Combining a large dataset of PM<sub>2.5</sub> and time series data in a machine learning model could produce the highest level of accuracy predicting PM<sub>2.5</sub> value [29].

Applying a machine learning (ML) technique, such as Multilayered Perceptron Regression [36] and regression trees [37] based methods such as Decision Tree Regression [38], Random Forest Regression [39], and Lasso, has become the common method today.

In Australia, a machine learning model has been applied successfully to attain a higher accuracy level in forecasting PM<sub>2.5</sub> [40]. Other research has compared existing traditional

models with a machine learning regression model using TAQMN data in Taiwan from 2012–2017 to predict the concentration level of PM in the air. The results showed that the machine learning regression model was more accurate and efficient than the existing traditional models, with the gradient boosting regressor model being better for forecasting air pollution using the TAQMN data [32]. A comparison of 2 ML approaches an artificial neural network (ANN) and a support vector model (SVM)—the prediction PM<sub>2.5</sub> in Delhi, India using meteorological and pollutant parameters over 2 years (2016–2018), it was found that the ANN model performed better and was more accurate than the SVM [41]. In China, Random Forest Regression (RF), support vector regression (SVR), and ANN were applied to measure the accuracy of multiple ML regression models on the prediction of PM<sub>2.5</sub>. The results showed that RF was more efficient than SVR and ANN in every aspect (mean absolute error (MAE) and R<sup>2</sup>) [42].

Furthermore, applying an ML approach to predict PM<sub>2.5</sub>, most research used the boosting technique together with another type of model, such as XGBoost [43] for which LSTM (Long Short-Term Memory) proved to be efficient in detecting PM<sub>2.5</sub>. This was supported by research from Iran, where ML-based models were applied using 3 different models (XGBoost, Random Forest, and deep learning for PM<sub>2.5</sub> prediction). The results showed that XGBoost was the best performing model compared to the other 2 in terms of the R<sup>2</sup>, MAE, and root mean square error (RMSE) values [44].

It seems that the PM<sub>2.5</sub> level has a seasonal distribution, with the concentration of pollution peaking during the dry season (November–March) and reducing in the rainy and summer seasons. The current study predicted PM<sub>2.5</sub> levels in Bangkok by applying different ML techniques to forecast the level of particulate matter (PM<sub>2.5</sub>).

## 3. STUDY AREA AND DATA

### 3.1. Study area

The meteorological dataset was collected from 2 sources: Pollution Control Department air quality monitoring stations (13 points) [45] and Environment Bureau Bangkok air quality monitoring stations (46 points) [17], as shown in Figure 1. The monitoring stations were permanently installed 2 meters above ground level and the instruments were calibrated based on guidelines of the Pollution Control Department [46].

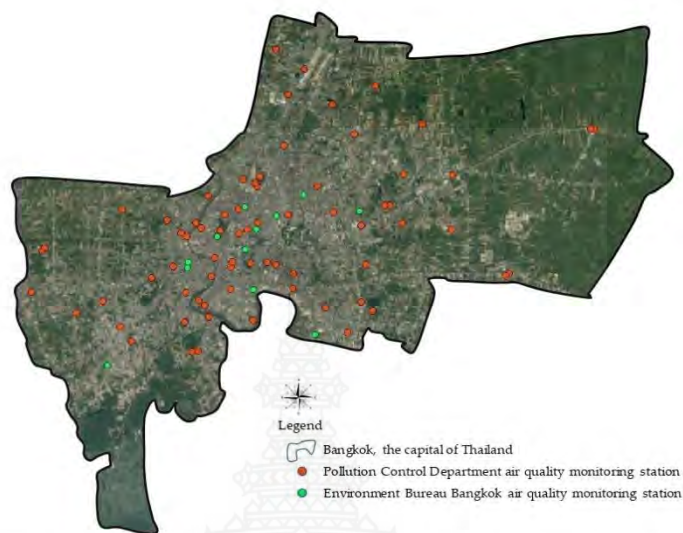


Fig. 1. Air monitoring stations of Pollution Control Department (red) and Environment Bureau Bangkok (green).

**3.2. Meteorological datasets**

The meteorological datasets contained: PM2.5 (particulate matter with a diameter lower than 2.5 μm, coarse particles (PM10), nitrogen oxide (NOx), nitrogen dioxide (NO2), nitric oxide (NO), carbon monoxide (CO), windspeed (WS), wind direction (WD), temperature (Temp), barometric pressure (BP), relative humidity (RH%), and ozone (O3). The data were collected at 1-hour intervals from 1 January 2015

to 31 December 2021. These data were averaged for 24-hour periods and compared with the standard values. The collection method was based on the standard federal equivalent method (FEM) of the United States Environmental Protection Agency (US EPA). Figure 2 shows a sample of the dataset used in the current study. Table 2 shows a summary of the dataset sources.

| Key_mth    | Hour | WS(m/s) | WD(Deg) | Temp(Deg.C) | RH(%) | BP(mBar) | CO(ppm) | NO(ppb) | NO2(ppb) | NOX(ppb) | O3(ppb) | PM10(ug/m3) | PM2.5(ug/m3) |
|------------|------|---------|---------|-------------|-------|----------|---------|---------|----------|----------|---------|-------------|--------------|
| 20210101.0 | 1.0  | 1.2     | 81.0    | 22.8        | 57.0  | 763.0    | 0.85    | 7.0     | 20.0     | 27.0     | 22.0    | 43.0        | 15.0         |
| 20210101.0 | 2.0  | 0.8     | 78.0    | 22.2        | 59.0  | 762.0    | 0.99    | 14.0    | 22.0     | 36.0     | 18.0    | 45.0        | 23.0         |
| 20210101.0 | 3.0  | 0.6     | 83.0    | 21.6        | 62.0  | 762.0    | 0.98    | 9.0     | 21.0     | 30.0     | 18.0    | 45.0        | 21.0         |
| 20210101.0 | 5.0  | 0.6     | 32.0    | 20.9        | 62.0  | 762.0    | 0.92    | 11.0    | 21.0     | 32.0     | 15.0    | 46.0        | 33.0         |
| 20210101.0 | 6.0  | 1.2     | 342.0   | 20.2        | 61.0  | 762.0    | 0.86    | 18.0    | 24.0     | 42.0     | 13.0    | 42.0        | 32.0         |
| 20210101.0 | 7.0  | 0.5     | 8.0     | 19.9        | 63.0  | 763.0    | 1.01    | 40.0    | 33.0     | 73.0     | 7.0     | 38.0        | 28.0         |
| 20210101.0 | 8.0  | 0.6     | 6.0     | 20.0        | 64.0  | 764.0    | 1.06    | 59.0    | 33.0     | 91.0     | 6.0     | 41.0        | 27.0         |
| 20210101.0 | 9.0  | 0.8     | 46.0    | 20.5        | 64.0  | 765.0    | 1.04    | 40.0    | 33.0     | 80.0     | 10.0    | 47.0        | 28.0         |
| 20210101.0 | 10.0 | 0.9     | 13.0    | 21.8        | 60.0  | 765.0    | 1.02    | 44.0    | 26.0     | 71.0     | 14.0    | 52.0        | 33.0         |
| 20210101.0 | 11.0 | 1.1     | 22.0    | 23.3        | 56.0  | 765.0    | 1.01    | 32.0    | 28.0     | 60.0     | 18.0    | 52.0        | 33.0         |

Fig. 2. Sample of dataset used in this study

Table 2. Summary of datasets and source

| Data Source   | Number of Stations | Data Records (2015–2021) |
|---|--------------------|--------------------------|
| Pollution Control Department air quality monitoring station | 13                 | 787,550                  |
| Environment Bureau Bangkok air quality monitoring station   | 46                 | 1,516,342                |
| <b>Total</b>  | <b>59</b>          | <b>2,303,892</b>         |

4. MODELING

After decades of data collection, a large amount of data has been collected. ML involves a combination of statistics, data science, and mathematics. Many ML models have been developed to solve problems in a variety of applications and situations [47–50]. The current study utilized the models detailed below.

4.1. Machine learning regression models

4.1.1. Decision Tree Model

A decision tree is a commonly used algorithm because it is readily understandable, provides good results and is user friendly. It is the foundation of the random forest approach, which is one of the best algorithms currently in use. The basis of the decision tree is easily understood, with the model visualized as an upside-down tree. At the top is the root node that is divided into child nodes, with leaf nodes the smallest (not splittable further). At the root node, if the data condition is fulfilled, then the route proceeds to successive child nodes until the final leaf node is reached [38]. The decision tree can be defined using Equations 1 and 2 below,

$$\text{Information Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v) \quad (1)$$

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v) \quad (2)$$

where  $S$  is the set of instances,  $A$  is an attribute,  $S_v$  is a subset of  $S$ , and  $(A)$  is the set of all possible values of  $A$ .

4.1.2. Gradient Boosting Descent Model

Gradient boosting is a method of learning regression and classification that converts poorer accuracy into better accuracy. It is a kind of multi decision tree that builds a new tree by learning from the errors of past trees based on a level-wise algorithm [43].

4.1.3. Gradient Boosting Descent Model

K-Nearest Neighbor (K-NN) is a method for classifying data through supervision to develop an initial model. Then, the model is classified from the current solution. K-NN is

used to analyze new data that is related to the old data [51].

4.1.4. Multilayer perception Model

Multi-Layer Perception (MLP) is a standard structure of NN. Generally, the number of input nodes is equal to the number of features, with the output node equal to 1 where there are 2 classes, or the number is equal to the available classes. The number of hidden nodes is variable and is selected by the user [36].

4.1.5. Random Forest Model

Random Forest (RF) is a machine learning method that is commonly used to solve problems involving regression or classification. RF is based on a decision tree [37,39,51]. By combining many decisions trees, RF is more efficient and accurate. Multiple decision trees are grown, with each tree having different features and data to contribute to variety and independence. The algorithm has several steps:

1. Random feature and data sampled from available dataset
2. A decision tree is constructed from all training data to find a prediction value
3. The number of selected decision trees is chosen and steps 1 and 2 are repeated to construct the new decision tree
4. The result is the prediction value of each tree

The final decision tree from the classification is used based on a majority vote that becomes the final value. However, in a regression problem, the mean value is calculated from every decision tree and presented as the final value.

4.1.6. Ridge regression Model

Ridge Regression (RR) involves analysis of a regression model in which the parameters have a strong correlation. With this method, there is no need to remove the independent correlation variable from the model. The fundamental approach is to increase the regression coefficient value. This may cause bias on the estimator but with a lower variance that results in a lower variance of the mean square of error [51].

4.2. Model evaluation

General evaluation indicators were used in this research to estimate the accuracy of prediction model, with the research focus on 3 parameters: coefficient of determination ( $R^2$ ), root mean square error (RMSE), and mean absolute error (MAE). The formulas are shown in Equations 3–6, respectively:

$$R^2 = \left[ \frac{1}{M} \sum_{j=1}^M \frac{(y_j - \bar{y})(\hat{y}_j - \bar{y})}{\sigma_y \sigma_{\hat{y}}} \right]^2 \quad (3)$$

where  $M$  is the number of observations,  $\sigma_x$  is the standard deviation of the observation  $X$ ,  $\sigma_y$  is the standard deviation of  $Y$ ,  $X_j$  is the observation value,  $\bar{X}$  is the mean of the observed values,  $Y_j$  is the calculated value, and  $\bar{Y}$  is the mean of the calculated values [52].

$$MSE = \frac{1}{m} \sum_{j=1}^m (x_i - \hat{x}_i)^2 \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{j=1}^m (x_i - \hat{x}_i)^2}{m}} \quad (5)$$

$$MAE = \frac{\sum_{j=1}^m |x_i - \hat{x}_i|}{m} \quad (6)$$

Where  $m$  is the number of observations,  $\hat{x}_i$  is the predicted value, and  $x_i$  is the actual value.

## 5. METHODOLOGY

The meteorological dataset for the Bangkok area was compiled from 2 sources: the Pollution Control Department air quality monitoring stations and Environment Bureau Bangkok air quality monitoring stations from 1 January 2015 to 31 December, 2021. The total dataset consisted of 2.3 million values.

### 5.1. Model Procedure

The raw data were divided into 2 groups. The first group consisted of specific geographic data at measurement locations. The second data consisted of a time-series involving rapid data changes, such as the PM 2.5 data, wind speed and direction, temperature, and humidity. The dataset was built based on several steps:

1. Coding was assigned to each data to make the computation more efficient in terms of time and size. Codes were either numeral or letters.
2. Editing was applied to correct the dataset involved removing junk data and representing blank data with a 0. Generally, data values were required to be equal to or more than 0.01.
3. Classifying data into types of computation improved convenience. In this research, specific data, such as the measuring location, were classified. Time series data, such as PM2.5 data, wind speed and direction, temperature, and humidity, were classified as shown in Figure 3.

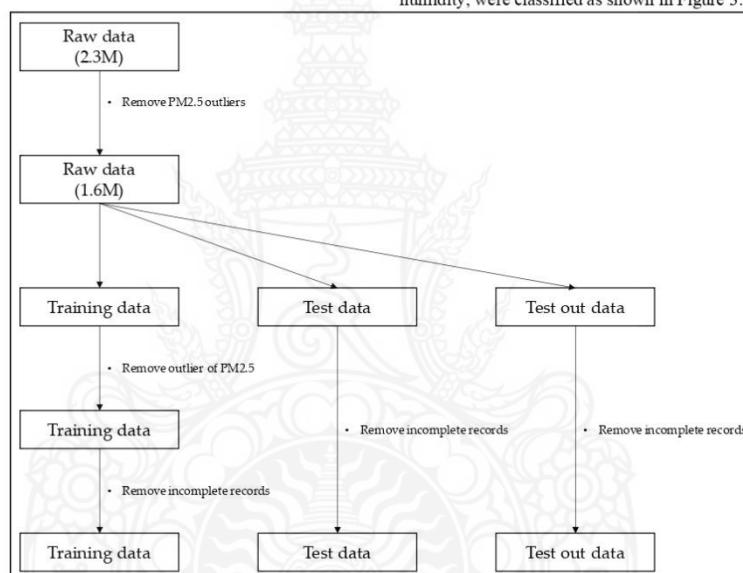


Fig. 3. Data processing and outlier removal.

In addition, developing a 1.6 million subset of the raw data was divided into 4 parts: data sourcing, data preparation, ML algorithms with tuning hyper parameters, and model evaluation with cross validation, to provide the highest analysis accuracy of the model.

In summary, the research compiled a 1.6 million subset of

raw data based on time, gas pollutants and metrological information from 1 January 2015 to 31 December 2021. The data had been cleaned and formatted before being imported into the model and used for validation analysis. The overall process is shown in Figure 4. The details of the procedure are mentioned in next section.

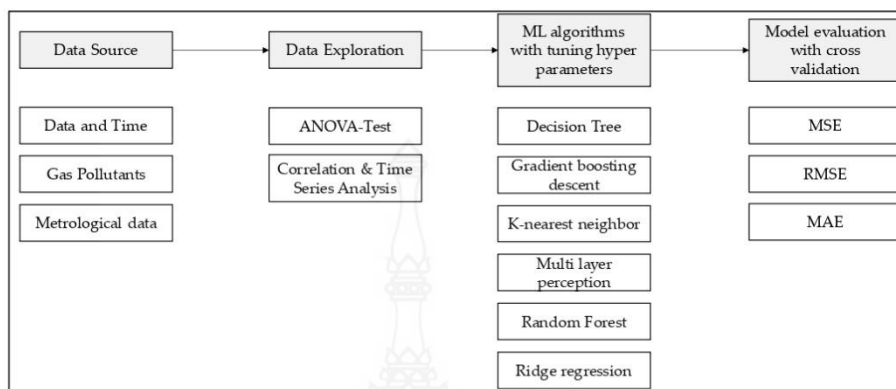


Fig. 4. Workflow of research.

5.2. Data exploration

The Kruskal-Wallis One-Way Analysis of Variance method was selected to identify the characteristics of the distribution of the time series data [53,54]. As the Kruskal-Wallis approach is commonly used to identify the distribution of air pollution data, its has been proven effective for monitoring seasonal distribution data regarding air quality and dominant air pollutants [52,55]. To conduct characteristic distribution testing, first, the concentration of pollution data characteristics was assumed using the seasonal distribution characteristic by setting the hypothesis as:

- H0: Median values of each pollution parameter are equal every month
- H1: Median values of each pollution parameter for each parameter differ at least once per month

This was tested based on H from Equation 7,

$$H = \left[ \frac{12}{n(n+1)} \sum_{j=1}^c \frac{T_j^2}{n_j} \right] - 3(n+1) \tag{7}$$

Where n is the sum of sample sizes for all samples, c is the number of samples,  $T_j$  is the sum of ranks in the  $j^{th}$  sample, and  $n_j$  is the size of the  $j^{th}$  sample.

Table 3 shows that the median values for each pollution parameter were different for at least 1 pair in each month, with P-values lower than 0.05. Therefore, hypothesis H1 was

supported and so it was correct to consider for each month that the pollution parameters before and during the lockdown period were significantly different.

Table 3. Kruskal-Wallis test results of pollution parameters (January 2015 to December 2021)

| Parameter                  | Statistics  |             |
|----------------------------|-------------|-------------|
|                            | H statistic | Asymp. Sig. |
| PM2.5 (ug/m <sup>3</sup> ) | 48.57       | 0.000       |
| PM10 (ug/m <sup>3</sup> )  | 49.37       | 0.000       |
| NO <sub>x</sub> (ppb)      | 28.62       | 0.000       |
| NO <sub>2</sub> (ppb)      | 28.94       | 0.000       |
| NO (ppb)                   | 44.48       | 0.000       |
| CO (ppm)                   | 36.69       | 0.000       |
| O <sub>3</sub> (ppb)       | 48.47       | 0.000       |

In summary, the testing results showed that the seasonal factors on pollution significantly outweighed the influence of road traffic volume. Furthermore, other research has shown that decreasing the road traffic volume was related to air pollution levels, with human activities being more directly related to the pollution level [56].

5.3. Time Series Analysis

Empirical evidence of PM2.5 and gas pollutants based on time series: Plotting the air pollutants parameters from 1 January 2015 to 31 December 2021 as a time series by day clearly showed that gas pollutant levels were dependent on seasonal factors, including PM2.5, as shown in Figure 5.

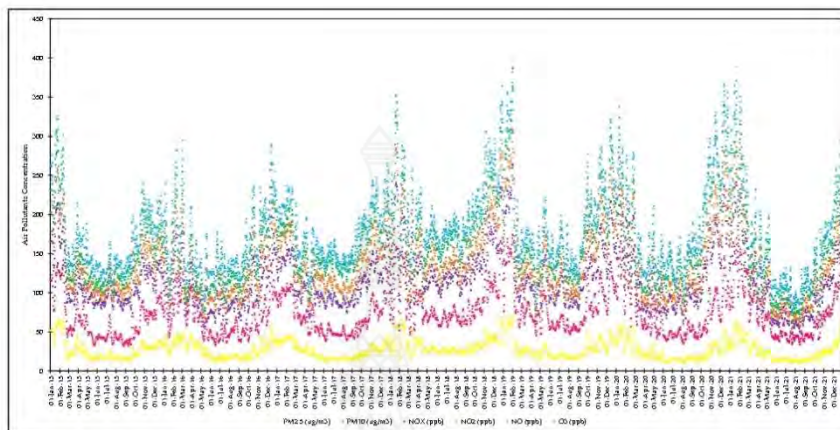


Fig. 5. Air pollutant parameters dataset from January 2015 to December 2021 showing seasonal characteristics.

5.3.1. Monthly effects on PM2.5

Figure 6 shows PM2.5 levels on a monthly basis over the past 5 years (2015-2021), indicating that PM2.5 peaked during the dry season every year (from November to December). In addition, the variation in the PM2.5 level was also high compared to other months of the year, which was consistent with others research. This could be explained by the meteorological data on air circulation as well. Generally, the air at ground level (troposphere layer), is higher than for other layers and gas moves from a high temperature to a lower temperature level as shown in Figure 7(a) Therefore,

air and gas from ground level will move to a higher level. Normally, dust and particles are blow away to a higher layer without any obstacle. However, during the dry season, Thailand is influenced by a high-pressure mass of cold air from the north. This rapidly cools the air temperature at ground level while the next layer is also still cold. Therefore, the released hot air and gas cannot pass through to a higher layer. This phenomenon is called an inversion trap. Figure 7(b) shows that the dust and particles that cannot be transported by convection to a higher layer return are trapped at ground level. Consequently, PM2.5 levels in the dry season are higher than in other seasons.

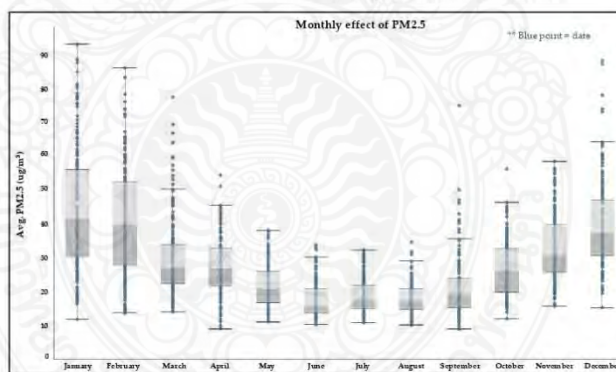


Fig. 6. Monthly effect of PM2.5.

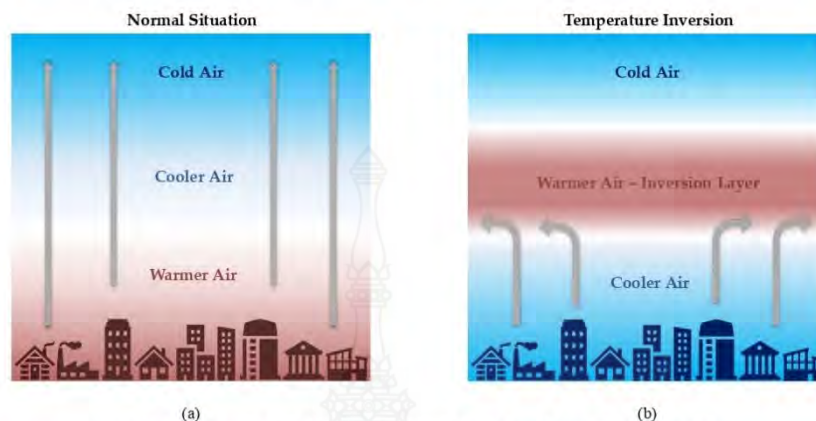


Fig.7. Temperature effect on PM2.5: (a) normal situation; (b) temperature inversion.

5.3.2. Hourly effects on PM2.5

Figure 8. shows the PM2.5 data from 2015 to 2021 on an hourly basis, with PM2.5 levels varying hourly during the day based on human activities and other factors. Every day, the PM2.5 level peaked during 2 periods: in the morning (08:00-10:00 AM) and then again in the evening (21:00-00:00 PM).

The morning peak was related to vehicular traffic taking people to workplaces and schools, while in the evening most people went home and trucks were more active during this period. However, the PM2.5 hourly rate may also vary depending on the area, such as urban or residential, factory zone, agriculture, or farming.

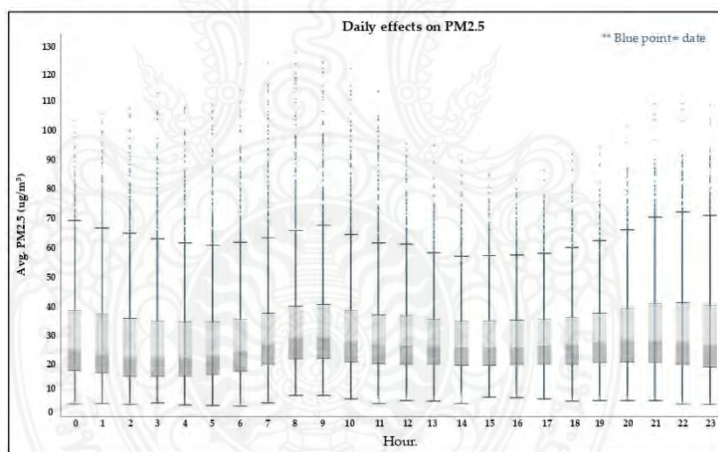


Fig. 8. Daily effects on PM2.5.

The investigation of the results showed that the PM2.5 level displayed seasonal characteristics and was related to gas pollutants. As mentioned, there was a clear dependence on a time pattern (hourly to monthly). Therefore, this result was included in building the predictive PM2.5 model.

The meteorology data was divided into 3 sets: a training set for building the model, a correction set data for cross checking the suitability of the model to predict the PM2.5 level and a testout set to model data during the Covid-19 pandemic and for subsequent testing.

The training set used train from 2015 to 2019 before the pandemic, split into 80% used to build the model and the other 20% for use in the training set. The testout set was built using the data during the pandemic and used in modelling and testing the model.

The model analysis and prediction considered historical data features with and without the time frame (by hour, by

day, by month) to demonstrate the accuracy of each type of model.

An autocorrelation function (ACF) was applied using an autocorrelation plot, with the result showing the relation between data and the time series based on the R. By analyzing the average PM2.5 from the previous year for the same month and day, as shown in Figure 9, the correlation was investigated between the observation at the current time and at the previous time spot. Most correlations were statistically significant because they were outside the gray line. Thus, the data used in the model with historical data feature lags by the hour, day and month should exceed 365 days. This was considered appropriate for the prediction of PM2.5 based on the correlation between the present and past. Notably, the acceptable correlation value for each study may depend on the study's purpose and so may not necessarily agree with other studies.

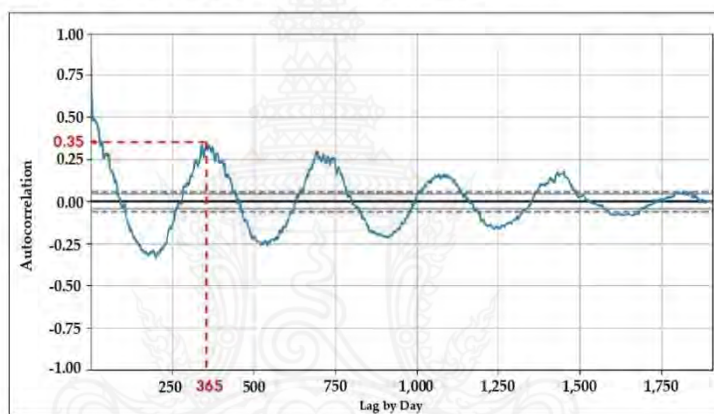


Fig. 9. Autocorrelation plot.

## 6. Results and Discussion

The corrections were investigated based on the 1.6 million time series PM2.5 data-set from January 2015 to December 2021, using the training dataset (80% of the PM2.5 data from 2015 to 2019, before the Covid-19 pandemic) the test set (use the remaining 20% of the training dataset) and the testout data (the complete PM2.5 data from 2020 to 2021, during the Covid-19 pandemic, for verifying the model correction).

This approach was used for the 6 tested models (Decision Tree, Gradient Boosting, K-neighbors, MLP, Random Forest, and Ridge). Each model was analyzed with and without the history data feature, including by the hour, by the day, by the

month, and with no time constraint to demonstrate the accuracy of those models. After scoring using  $R^2$ , RMSE, and MAE as the performance criteria, in Figure 10 to Figure 15 in actual and predicted.

Moreover, coefficient of determination values is shown that the historical data by the hour produced the best performance using in the descending order: Gradient Boosting, Random Forest, K-neighbors, Decision Tree, MLP, and Ridge Model, respectively.

For Gradient Boosting by the hour, the  $R^2$  value of the training dataset was 0.8284 and for the test set was 0.8051. Details of the test result are shown in Table 4.



Table 4. Results of training and test values for different machine learning models forecasting PM2.5

| Model             | Historical Data Feature | Training       |          |         |        | Test           |          |         |        |
|-------------------|-------------------------|----------------|----------|---------|--------|----------------|----------|---------|--------|
|                   |                         | R <sup>2</sup> | MSE      | RMSE    | MAE    | R <sup>2</sup> | MSE      | RMSE    | MAE    |
| Decision Tree     | by hour                 | 0.8017         | 65.7543  | 8.1089  | 5.8145 | 0.7890         | 70.0052  | 8.3669  | 5.9728 |
|                   | by day                  | 0.7618         | 80.9123  | 8.9951  | 6.0930 | 0.7241         | 92.6223  | 9.6240  | 6.5184 |
|                   | by month                | 0.7921         | 67.3461  | 8.2065  | 5.8507 | 0.7827         | 72.7251  | 8.5279  | 6.0172 |
|                   | no time                 | 0.7544         | 81.2886  | 9.0160  | 6.3955 | 0.7418         | 86.8981  | 9.3219  | 6.5325 |
| Gradient Boosting | by hour                 | 0.8284         | 55.6621  | 7.4607  | 5.4034 | 0.8051         | 64.9063  | 8.0564  | 5.7150 |
|                   | by day                  | 0.7824         | 73.2535  | 8.5588  | 6.0007 | 0.7645         | 78.3282  | 8.8503  | 6.1122 |
|                   | by month                | 0.8642         | 43.9988  | 6.6332  | 4.7948 | 0.8329         | 55.9262  | 7.4784  | 5.2317 |
|                   | no time                 | 0.7848         | 71.2274  | 8.4396  | 6.0626 | 0.7588         | 81.1613  | 9.0090  | 6.3240 |
| K-neighbors       | by hour                 | 0.8302         | 56.3061  | 7.5037  | 5.3432 | 0.7901         | 69.6302  | 8.3445  | 5.9256 |
|                   | by day                  | 0.8077         | 65.0308  | 8.0642  | 5.3802 | 0.7653         | 80.6869  | 8.9826  | 6.0057 |
|                   | by month                | 0.8624         | 45.7893  | 6.7668  | 4.7886 | 0.7898         | 68.7357  | 8.2907  | 5.8749 |
|                   | no time                 | 0.8325         | 56.8526  | 7.5401  | 5.0911 | 0.7465         | 87.0013  | 9.3274  | 6.2412 |
| MLP               | by hour                 | 0.7142         | 96.9269  | 9.8451  | 6.9650 | 0.7073         | 96.7124  | 9.8342  | 6.8934 |
|                   | by day                  | 0.7005         | 99.0589  | 9.9528  | 6.9428 | 0.6913         | 109.4271 | 10.4607 | 7.2350 |
|                   | by month                | 0.7059         | 99.5924  | 9.9796  | 7.0319 | 0.6843         | 102.5848 | 10.1284 | 7.1108 |
|                   | no time                 | 0.6542         | 126.7605 | 11.2588 | 8.1284 | 0.6386         | 130.5280 | 11.4249 | 8.3154 |
| Random Forest     | by hour                 | 0.8205         | 57.5180  | 7.5841  | 5.2128 | 0.8010         | 62.6875  | 7.9175  | 5.4309 |
|                   | by day                  | 0.7006         | 110.0523 | 10.4906 | 7.4418 | 0.6714         | 121.3946 | 11.0179 | 7.7173 |
|                   | by month                | 0.8166         | 58.4333  | 7.6442  | 5.2637 | 0.7997         | 64.5855  | 8.0365  | 5.5144 |
|                   | no time                 | 0.6907         | 114.5693 | 10.7037 | 7.6976 | 0.6578         | 125.3576 | 11.1963 | 7.9611 |
| Ridge             | by hour                 | 0.6983         | 100.0441 | 10.0022 | 7.2036 | 0.6979         | 100.2079 | 10.0104 | 7.1910 |
|                   | by day                  | 0.6637         | 113.5240 | 10.6548 | 7.5389 | 0.6714         | 110.1318 | 10.4944 | 7.4766 |
|                   | by month                | 0.6983         | 100.3902 | 10.0195 | 7.2072 | 0.6970         | 99.0929  | 9.9545  | 7.2097 |
|                   | no time                 | 0.6670         | 110.2289 | 10.4990 | 7.5299 | 0.6640         | 113.0522 | 10.6326 | 7.5495 |

After testing the models with the historical data in the testout dataset (during the Covid-19 Pandemic), the result also showed that the historical data by the hour provided the best performance in the descending order : Gradient

Boosting, Decision Tree, Random Forest, K-neighbors, Ridge, and MLP Model, respectively. The R<sup>2</sup> value for the Gradient Boosting model by the hour was 0.8053, as shown in Table 5. The result was significantly related to the test set.

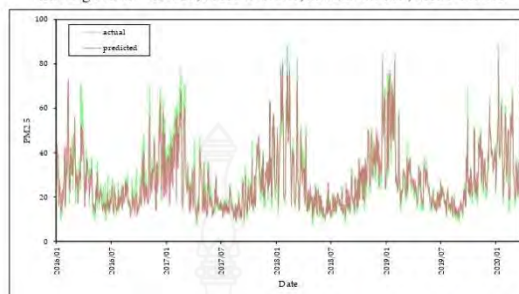
Table 5. Resulting testout values for different machine learning models forecasting PM2.5 during Covid-19 pandemic

| Model             | Time Series | Testout (Covid-19 Pandemic) |         |        |        |
|-------------------|-------------|-----------------------------|---------|--------|--------|
|                   |             | R <sup>2</sup>              | MSE     | RMSE   | MAE    |
| Decision Tree     | by hour     | 0.8035                      | 50.1563 | 7.0821 | 4.6952 |
|                   | by day      | 0.7501                      | 69.3920 | 8.3302 | 5.5310 |
|                   | by month    | 0.7918                      | 51.8494 | 7.2007 | 4.7252 |
|                   | no time     | 0.7642                      | 59.8631 | 7.7371 | 5.1764 |
| Gradient Boosting | by hour     | 0.8053                      | 48.5086 | 6.9648 | 4.5551 |
|                   | by day      | 0.7998                      | 49.8644 | 7.0615 | 4.5519 |
|                   | by month    | 0.7818                      | 54.3374 | 7.3714 | 4.8408 |
|                   | no time     | 0.7678                      | 58.9524 | 7.6780 | 5.0122 |
| K-neighbors       | by hour     | 0.7809                      | 55.9070 | 7.4771 | 5.0635 |
|                   | by day      | 0.7318                      | 74.4811 | 8.6302 | 5.9177 |
|                   | by month    | 0.7477                      | 64.3791 | 8.0237 | 5.4863 |

| Model         | Time Series | Testout (Covid-19 Pandemic) |         |        |        |
|---------------|-------------|-----------------------------|---------|--------|--------|
|               |             | R <sup>2</sup>              | MSE     | RMSE   | MAE    |
|               | no time     | 0.6986                      | 83.7123 | 9.1494 | 6.3331 |
| MLP           | by hour     | 0.7017                      | 81.4719 | 9.0262 | 6.2649 |
|               | by day      | 0.7052                      | 80.4965 | 8.9720 | 6.4183 |
|               | by month    | 0.7010                      | 81.6593 | 9.0366 | 6.2794 |
|               | no time     | 0.6782                      | 86.0405 | 9.2758 | 6.4872 |
| Random Forest | by hour     | 0.7936                      | 57.3293 | 7.5716 | 4.9910 |
|               | by day      | 0.7003                      | 81.8452 | 9.0468 | 6.4941 |
|               | by month    | 0.7965                      | 56.5099 | 7.5173 | 4.9020 |
|               | no time     | 0.7044                      | 80.7073 | 8.9837 | 6.3562 |
| Ridge         | by hour     | 0.7073                      | 74.6906 | 8.6424 | 5.9407 |
|               | by day      | 0.7159                      | 70.7681 | 8.4124 | 5.7173 |
|               | by month    | 0.7052                      | 75.2117 | 8.6725 | 5.9715 |
|               | no time     | 0.6841                      | 80.2058 | 8.9558 | 6.2944 |

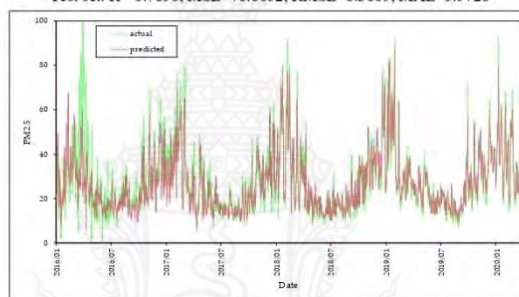


Training set:  $R^2=0.8017$ , MSE=65.7543, RMSE=8.1089, MAE=5.8145



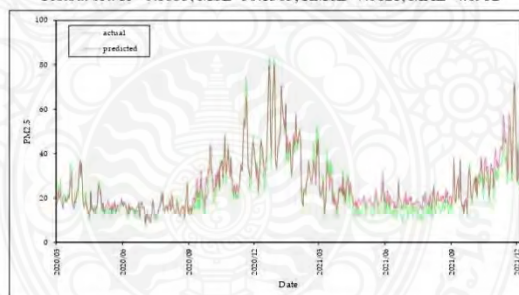
(a) Training set results

Test set:  $R^2=0.7890$ , MSE=70.0052, RMSE=8.3669, MAE=5.9728



(b) Test set results

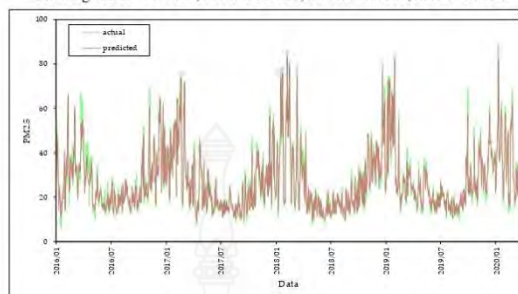
Testout set:  $R^2=0.8035$ , MSE=50.1563, RMSE=7.0821, MAE=4.6952



(c) Testout set results

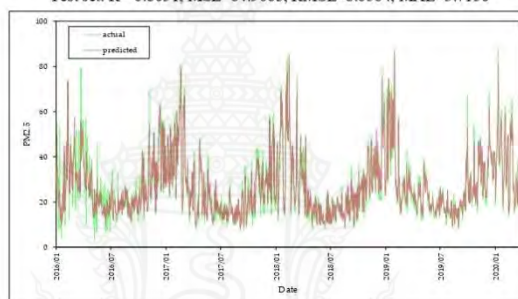
Fig. 10. Prediction models for decision tree regression: (a) training set; (b) test set; (c) testout set.

Training set:  $R^2=0.8284$ , MSE=55.6621, RMSE=7.4607, MAE=5.4034



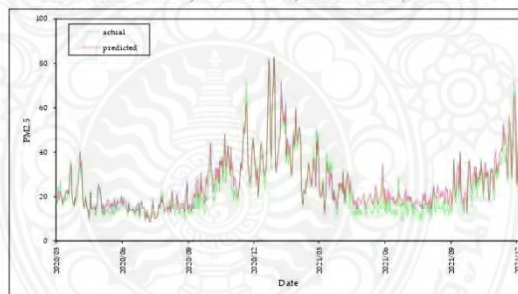
(a) training set

Test set:  $R^2=0.8051$ , MSE=64.9063, RMSE=8.0564, MAE=5.7150



(b) test set

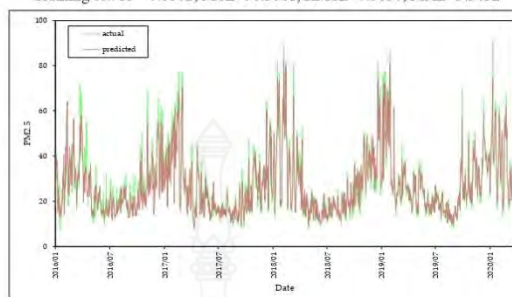
Testout set:  $R^2=0.8053$ , MSE=48.5086, RMSE=6.9648, MAE=4.5551



(c) testout set

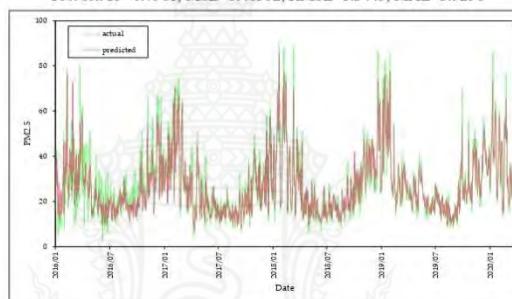
Fig. 11. Prediction models for Gradient Boosting: (a) training set; (b) test set; (c) testout set.

Training set:  $R^2=0.8302$ ,  $MSE=56.3061$ ,  $RMSE=7.5037$ ,  $MAE=5.3432$



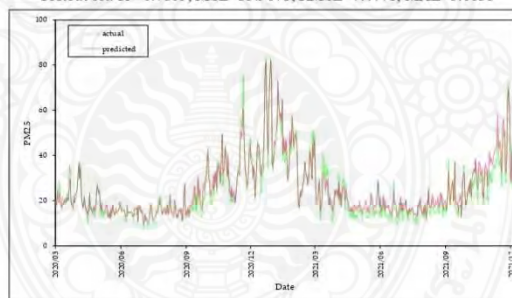
(a) training set

Test set:  $R^2=0.7901$ ,  $MSE=69.6302$ ,  $RMSE=8.3445$ ,  $MAE=5.9256$



(b) test set

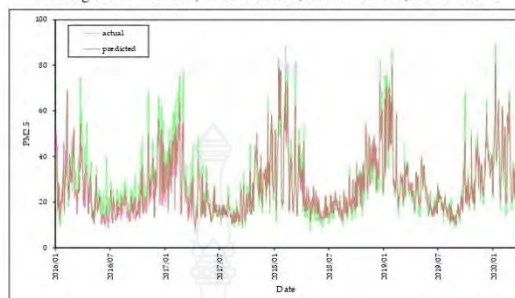
Testout set:  $R^2=0.7809$ ,  $MSE=55.9070$ ,  $RMSE=7.4771$ ,  $MAE=5.0635$



(c) testout set

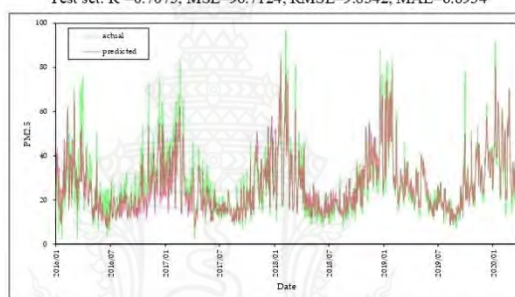
Fig. 12. Prediction models for K-Nearest Neighbor: (a) training set; (b) test set; (c) testout set.

Training set:  $R^2=0.7142$ ,  $MSE=96.9269$ ,  $RMSE=9.8451$ ,  $MAE=6.9650$



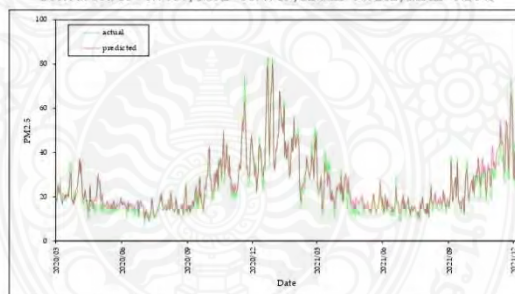
(a) training set

Test set:  $R^2=0.7073$ ,  $MSE=96.7124$ ,  $RMSE=9.8342$ ,  $MAE=6.8934$



(b) test set

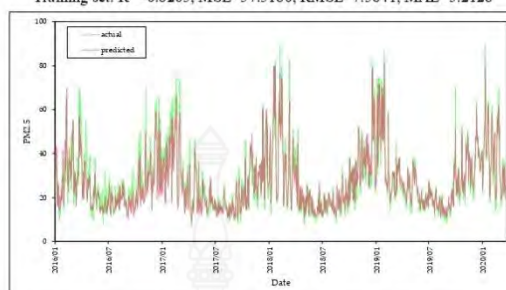
Testout set:  $R^2=0.7017$ ,  $MSE=81.4719$ ,  $RMSE=9.0262$ ,  $MAE=6.2649$



(c) testout set

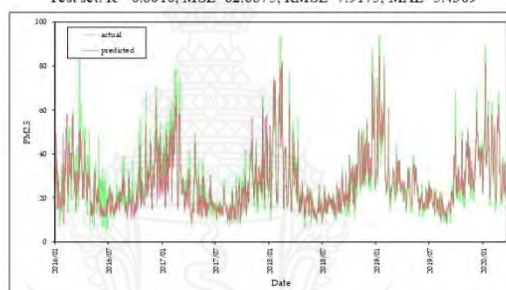
Fig. 13. Prediction models for Multi Layer Perception: (a) training set; (b) test set; (c) testout set.

Training set:  $R^2=0.8205$ , MSE=57.5180, RMSE=7.5841, MAE=5.2128



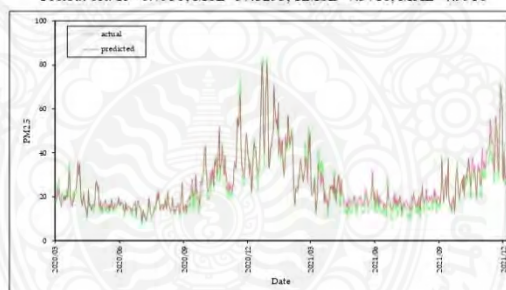
(a) training set

Test set:  $R^2=0.8010$ , MSE=62.6875, RMSE=7.9175, MAE=5.4309



(b) test set

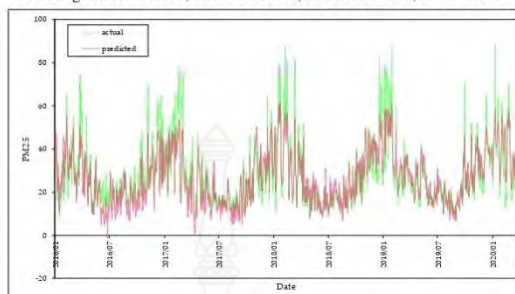
Testout set:  $R^2=0.7936$ , MSE=57.3293, RMSE=7.5716, MAE=4.9910



(c) testout set

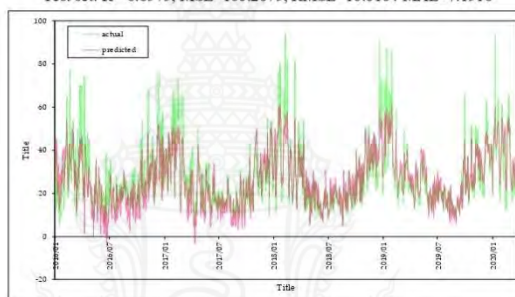
Fig. 14. Prediction models for Random Forest: (a) training set; (b) test set; (c) testout set.

Training set:  $R^2=0.6983$ ,  $MSE=100.0441$ ,  $RMSE=10.0022$ ,  $MAE=7.2036$



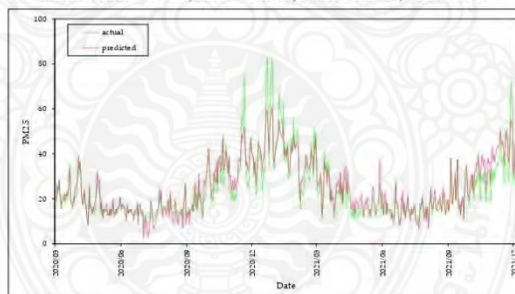
(a) training set

Test set:  $R^2=0.6979$ ,  $MSE=100.2079$ ,  $RMSE=10.0104$ ,  $MAE=7.1910$



(b) test set

Testout set:  $R^2=0.7073$ ,  $MSE=74.6906$ ,  $RMSE=8.6424$ ,  $MAE=5.9407$



(c) testout set

Fig. 15. Prediction models for Ridge Regression : (a) training set; (b) test set; (c) testout set.



## 7. CONCLUSIONS

The main objective of this research was to predict the PM2.5 level using 6 different machine learning approaches based on 1.6 million records of PM2.5 time-series data from 1 January 2015 to 31 December 2021.

The dataset was divided into 3 groups: a training dataset for model building, consisting of 80% of the PM2.5 data from 2015 to 2019; a test dataset to check the accuracy of the predictive model, consisting of the remaining 20% of the training set; and a testout dataset based on the complete data covering the period during the Covid-19 pandemic from 2020 to 2021.

Six models were investigated: Decision Tree, Gradient Boosting, K-neighbors, MLP, Random Forest, and Ridge. Model accuracy was investigated using historical data features by the hour, by the day, by the month and with no time to determine the contribution of time related data to model accuracy based on the values for R2 RMSE, and MAE.

The results showed that historical data by the hour deliver the best performance with the descending ranking of the models being: Gradient Boosting, Random Forest, K-neighbors, Decision Tree, MLP, and Ridge Model, respectively. Gradient Boosting by the hour provided the best results for both the training dataset (R2 =0.8284) and the test dataset (R2 =0.8051). In addition, based on the R2 values of the testout dataset Gradient Boosting provided the same result (R2 =0.8053).

In summary, predicting the PM2.5 level in the Bangkok area with meteorological data and data from the Bangkok Air Pollution Department, the Gradient Boosting model using historical data by the hour was the most appropriate model for PM2.5 prediction regardless of the timing (before or during the Covid-19 pandemic).

## 8. RECOMMENDATION

The model was based on meteorological and pollution data for Bangkok, which has a high traffic density but low industrial agricultural activities. The results should be applied taking this information into account. The results may vary depending on different physical characteristics.

The results indicated that the traffic volume directly affected the PM2.5 concentration level in the study area. In future, policy maker should install monitoring stations to collected meteorological and pollution data from roadside areas as well as collecting traffic data. This could help to improve the prediction of PM2.5 levels and lead to more efficient analysis of the root causes of the PM2.5 problem in urban areas with large traffic volumes that vary over time during the day and week. This could help policy makers to gather more useful data to effectively address the problem.

## ACKNOWLEDGEMENTS

The completion of this study would not have been possible without assistance from Rajamagala University of Technology Phra Nakhon, Thailand. The authors also thank the Pollution Control Department, Environment Bureau Bangkok, and the Expressway Authority of Thailand for providing support with data collection.

## REFERENCES

- [1] U. S. EPA, "Criteria Air Pollutants," *Encycl. Immunotoxicol.*, no. October, pp. 218–218, 2016, doi: 10.1007/978-3-642-54596-2\_200326.
- [2] V. Ratanavaraha and S. Jomnonkwo, "Trends in Thailand CO2 emissions in the transportation sector and Policy Mitigation," *Transp. Policy*, vol. 41, pp. 136–146, 2015, doi: 10.1016/j.tranpol.2015.01.007.
- [3] C. Li and S. Managi, "Contribution of on-road transportation to PM2.5," *Sci. Rep.*, vol. 11, no. 1, pp. 1–12, 2021, doi: 10.1038/s41598-021-00862-x.
- [4] S. Heydari, M. Tainio, J. Woodcock, and A. de Nazelle, "Estimating traffic contribution to particulate matter concentration in urban areas using a multilevel Bayesian meta-regression approach," *Environ. Int.*, vol. 141, no. December 2019, p. 105800, 2020, doi: 10.1016/j.envint.2020.105800.
- [5] W. M. Hodan and W. R. Barnard, "Evaluating the Contribution of PM2.5 Precursor Gases and Re-entrained Road Emissions to Mobile Source PM2.5 Particulate Matter Emissions. MACTEC Under Contract to the Federal Highway Administration," MACTEC Fed. Programs, Res. Triangle Park, NC, p. 58 pp, 2004, [Online]. Available: <https://www3.epa.gov/ttnchie1/conference/ei13/mobile/hodan.pdf>.
- [6] R. Sun, Y. Zhou, J. Wu, and Z. Gong, "Influencing factors of PM2.5 pollution: Disaster points of meteorological factors," *Int. J. Environ. Res. Public Health*, vol. 16, no. 20, 2019, doi: 10.3390/ijerph16203891.
- [7] R. Li, X. Mei, L. Wei, X. Han, M. Zhang, and Y. Jing, "Study on the contribution of transport to PM2.5 in typical regions of China using the regional air quality model RAMS-CMAQ," *Atmos. Environ.*, vol. 214, no. 40, p. 116856, 2019, doi: 10.1016/j.atmosenv.2019.116856.
- [8] Engel, "World air quality report 2021," *Pap. Knowl. Towar. a Media Hist. Doc.*, 2014.
- [9] M. Station, K. Findings, K. Kaen, C. Mai, and C. Mai, "Greenpeace's City Rankings for PM2.5 in Thailand," no. 2013, pp. 1–6, 2016.
- [10] Tomtom Traffic Index, "Traffic Index Traffic Index results 2018 - 2021," *Www.Tomtom.Com*, p. 21, 2021, [Online]. Available: <https://www.tomtom.com/traffic-index/ranking/>.
- [11] B. Pishue and T. Trepanier, "2020 INRIX Global Traffic Scorecard and latest travel trends," no. March, 2021.
- [12] I. Z. A. D. P. No, H. H. Dang, T. Trinh, and H. H. Dang, "DISCUSSION PAPER SERIES Does the COVID-19 Pandemic Improve Global Air Quality? New Cross-National Evidence on Its Unintended Consequences Does the COVID-19 Pandemic Improve Global Air Quality? New Cross-National Evidence on Its Unintended Consequences," no. 13480, 2020.
- [13] P. Dasgupta and K. Srikanth, "Reduced air pollution during COVID-19: Learnings for sustainability from Indian Cities," *Glob. Transitions*, vol. 2, pp. 271–282, 2020, doi:

- 10.1016/j.glt.2020.10.002.
- [14] Z. S. Venter, K. Atunan, S. Chowdhury, and J. Lelieveld, "COVID-19 lockdowns cause global air pollution declines," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 117, no. 32, pp. 18984–18990, 2020, doi: 10.1073/pnas.2006853117.
- [15] "Pearson Correlation Coefficient (r) \_ Guide & Examples." .
- [16] P. Uttamang, V. P. Aneja, and A. F. Hanna, "Assessment of gaseous criteria pollutants in the Bangkok Metropolitan Region, Thailand," *Atmos. Chem. Phys.*, vol. 18, no. 16, pp. 12581–12593, 2018, doi: 10.5194/acp-18-12581-2018.
- [17] P. C. Department, "Bangkok Air Pollution: Real-time Air Quality Index (AQI)." pp. 1–5, 2021, [Online]. Available: <https://aqicn.org/city/bangkok>.
- [18] T. Amnuaylojaroen, "Prediction of PM<sub>2.5</sub> in an Urban Area of Northern Thailand Using Multivariate Linear Regression Model," *Adv. Meteorol.*, vol. 2022, p. 3190484, 2022, doi: 10.1155/2022/3190484.
- [19] S. Khodmanee and T. Amnuaylojaroen, "Impact of Biomass Burning on Ozone, Carbon Monoxide, and Nitrogen Dioxide in Northern Thailand," *Front. Environ. Sci.*, vol. 9, no. April, pp. 1–6, 2021, doi: 10.3389/fenvs.2021.641877.
- [20] S. Chirasophon and P. Pochanart, "The long-term characteristics of PM<sub>10</sub> and PM<sub>2.5</sub> in Bangkok, Thailand," *Asian J. Atmos. Environ.*, vol. 14, no. 1, pp. 73–83, 2020, doi: 10.5572/AJAE.2020.14.1.073.
- [21] D. Narita et al., "Pollution characteristics and policy actions on fine particulate matter in a growing Asian economy: The case of Bangkok Metropolitan Region," *Atmosphere (Basel)*, vol. 10, no. 5, pp. 1–18, 2019, doi: 10.3390/atmos10050227.
- [22] P. Watcharaviton, C. P. Chio, and C. C. Chan, "Temporal and spatial variations in ambient air quality during 1996–2009 in Bangkok, Thailand," *Aerosol Air Qual. Res.*, vol. 13, no. 6, pp. 1741–1754, 2013, doi: 10.4209/aaqr.2012.11.0305.
- [23] I. G. Zurbenko, "Detecting and tracking changes in ozone air quality," *Air Waste*, vol. 44, no. 9, pp. 1089–1092, 1994, doi: 10.1080/10473289.1994.10467303.
- [24] C. Libiseller, A. Grimvall, and H. Saari, "Meteorological Normalisation and Non-Parametric Smoothing For Quality Assessment And Trend Analysis Of Tropospheric ozone Data," pp. 33–52, 2005.
- [25] E. K. Wise, A. C. Comrie, E. K. Wise, and A. C. Comrie, "Extending the Kolmogorov – Zurbenko Filter : Application to Ozone , Particulate Matter , and Meteorological Trends Extending the Kolmogorov – Zurbenko Filter : Application to Ozone , Particulate Matter , and Meteorological Trends," vol. 2247, 2012, doi: 10.1080/10473289.2005.10464718.
- [26] S. C. Pryor, I. G. McKendry, and D. G. Steyn, "Synoptic-Scale Meteorological Variability and Surface Ozone Concentrations in Vancouver, British Columbia," *J. Appl. Meteorol.*, vol. 34, no. 8, pp. 1824–1833, Oct. 1995, [Online]. Available: <http://www.jstor.org/stable/26187067>.
- [27] C. Libiseller and A. Grimvall, "Model selection for local and regional meteorological normalisation of background concentrations of tropospheric ozone," vol. 37, pp. 3923–3931, 2003, doi: 10.1016/S1352-2310(03)00502-8.
- [28] Y. Guo, Q. Tang, D. Gong, and Z. Zhang, "Estimating ground-level PM<sub>2.5</sub> concentrations in Beijing using a satellite-based geographically and temporally weighted regression model," *Remote Sens. Environ.*, vol. 198, pp. 140–149, 2017, doi: 10.1016/j.rse.2017.06.001.
- [29] S. Kumar, S. Mishra, and S. K. Singh, "A machine learning-based model to estimate PM<sub>2.5</sub> concentration levels in Delhi ' s atmosphere," *Heliyon*, vol. 6, no. November, p. e05618, 2020, doi: 10.1016/j.heliyon.2020.e05618.
- [30] A. Azid, H. Juahir, and M. E. Toriman, "Prediction of the Level of Air Pollution Using Principal Component Analysis and Artificial Neural Network Techniques : a Case Study in Malaysia," 2014, doi: 10.1007/s11270-014-2063-1.
- [31] H. Karimian et al., "Evaluation of Different Machine Learning Approaches to Forecasting PM<sub>2.5</sub> Mass Concentrations," pp. 1400–1410, 2019, doi: 10.4209/aaqr.2018.12.0450.
- [32] Y. Km, I. Gad, K. S. Harishkumar, Y. Km, and I. Gad, "Forecasting Air Pollution Particulate Matter (PM<sub>2.5</sub>) Using Machine Learning Regression Models," *Procedia Comput. Sci.*, vol. 171, no. 2019, pp. 2057–2066, 2020, doi: 10.1016/j.procs.2020.04.221.
- [33] T. M. T. Lei, S. W. I. Siu, J. Monjardino, and L. Mendes, "Using Machine Learning Methods to Forecast Air Quality: A Case Study in Macao," pp. 1–14, 2022.
- [34] I. Gutiérrez-avila et al., "Prediction of daily mean and one-hour maximum PM<sub>2.5</sub> concentrations and applications in Central Mexico using satellite-based machine-learning models," no. March, 2022, doi: 10.1038/s41370-022-00471-4.
- [35] Y. Km, I. Gad, K. S. Harishkumar, Y. Km, and I. Gad, "Using Machine Forecasting Air Air Pollution Pollution Particulate Particulate Matter Matter (PMLearning Learning Regression Regression Models Models," *Procedia Comput. Sci.*, vol. 171, no. 2019, pp. 2057–2066, 2020, doi: 10.1016/j.procs.2020.04.221.
- [36] Q. Zhou, H. Jiang, J. Wang, and J. Zhou, "A hybrid model for PM<sub>2.5</sub> forecasting based on ensemble empirical mode decomposition and a general regression neural network," *Sci. Total Environ.*, vol. 496, pp. 264–274, 2014, doi: 10.1016/j.scitotenv.2014.07.051.
- [37] Leo Breiman, *Classification And Regression Trees*, 1st Editio. New York: Routledge, 1984.
- [38] L. E. O. Bbeiman, "Bagging Predictors," vol. 140, pp. 123–140, 1996.
- [39] L. E. O. Breiman, "Random Forests," pp. 5–32, 2001.
- [40] R. O. Sinnott and Z. Guan, "Prediction of Air Pollution through Machine Learning Approaches on the Cloud," in *2018 IEEE/ACM 5th International Conference on Big Data Computing Applications and Technologies (BDCAT)*, 2018, pp. 51–60, doi: 10.1109/BDCAT.2018.00015.
- [41] A. Masood and K. Ahmad, "A model for particulate matter ( PM<sub>2.5</sub> ) prediction for Delhi based on A model for particulate matter ) prediction for Delhi based on machine learning machine learning approaches," *Procedia Comput. Sci.*, vol. 167, no. 2019, pp. 2101–2110, 2020, doi: 10.1016/j.procs.2020.03.258.
- [42] D. Lu, W. Mao, L. Zheng, W. Xiao, L. Zhang, and J. Wei, "Ambient pm<sub>2.5</sub> estimates and variations during covid-19 pandemic in the yangtze river delta using machine learning and big data," *Remote Sens.*, vol. 13, no. 8, 2021, doi: 10.3390/rs13081423.
- [43] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," pp. 785–794, 2016.

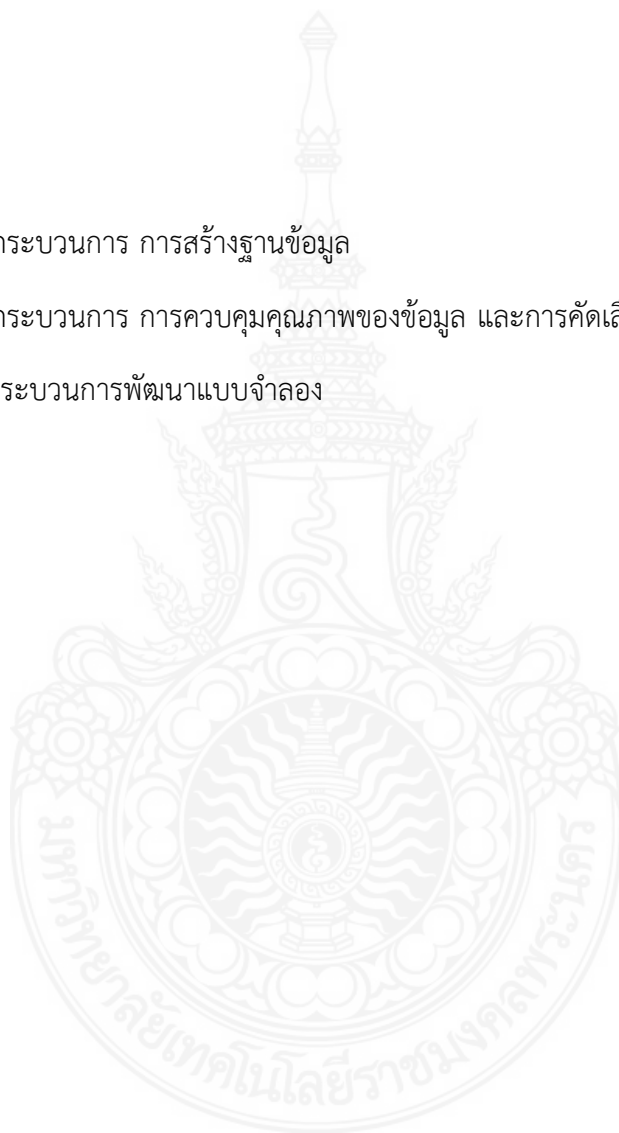
- [44] A. X. V. I. Simp and S. Remoto, "PM2.5 Prediction Based on Random Forest, XGBoost, and Deep Learning Using Multisource Remote Sensing Data Mehdi," no. 1992, pp. 6425–6432, 2013.
- [45] "Air quality and noise statistics, Pollution Control Department." <http://air4thai.pcd.go.th/webV3/#/History>.
- [46] "Instruments and methods for measuring the average value of gas or dust in the atmosphere in general - Pollution Control Department." <https://www.pcd.go.th/publication/5420/>.
- [47] S. K. Grange, D. C. Carslaw, A. C. Lewis, E. Boleti, and C. Hueglin, "Random forest meteorological normalisation models for Swiss PM10 trend analysis," *Atmos. Chem. Phys.*, vol. 18, no. 9, pp. 6223–6239, 2018, doi: 10.5194/acp-18-6223-2018.
- [48] A. Smola and S. V. N. Vishwanathan, "Introduction to Machine Learning," *Mach. Learn. Refin.*, 2020.
- [49] M. Kuhn et al., "useR! 2013 Modeling Conventions in R," 2013.
- [50] J. Xie, X. Wang, Y. Liu, and Y. Bai, "Autoencoder-based deep belief regression network for air particulate matter concentration forecasting," *J. Intell. Fuzzy Syst.*, vol. 34, pp. 3475–3486, 2018, doi: 10.3233/JIFS-169527.
- [51] D. W. Gareth James Trevor Hastie, Robert Tibshirani, An introduction to statistical learning : with applications in R. New York : Springer, [2013] ©2013.
- [52] K. Weibel, "A Data-Driven Selection of an Appropriate Seasonal Adjustment Approach," *SSRN Electron. J.*, no. 07, 2021, doi: 10.2139/ssrn.2797077.
- [53] W. H. Kruskal and W. A. Wallis, "Use of Ranks in One-Criterion Variance Analysis," *J. Am. Stat. Assoc.*, vol. 47, no. 260, pp. 583–621, 1952, doi: 10.1080/01621459.1952.10483441.
- [54] T. W. MacFarland and J. M. Yates, *Introduction to Nonparametric Statistics for the Biological Sciences Using R*, 1st ed. Springer Publishing Company, Incorporated, 2018.
- [55] X. Zhao et al., "Statistical analysis of spatiotemporal heterogeneity of the distribution of air quality and dominant air pollutants and the effect factors in Qingdao Urban Zones," *Atmosphere (Basel)*, vol. 9, no. 4, pp. 12–16, 2018, doi: 10.3390/atmos9040135.
- [56] S. Dursun, M. Sagdic, and H. Toros, "The impact of COVID-19 measures on air quality in Turkey," *Environ. Forensics*, vol. 23, no. 1–2, pp. 47–59, 2022, doi: 10.1080/15275922.2021.1892876.



## ภาคผนวก ข

## รายละเอียดชุดคำสั่ง (Code) ของการพัฒนาแบบจำลอง

1. Coding ของกระบวนการ การสร้างฐานข้อมูล
2. Coding ของกระบวนการ การควบคุมคุณภาพของข้อมูล และการคัดเลือกตัวแปร
3. Coding ของกระบวนการพัฒนาแบบจำลอง



## 1. Coding ของกระบวนการ การสร้างฐานข้อมูล

### Load Data

```
In [6]: data64 = pd.read_csv('Data_2564.csv'); print('2564', data64.shape)
data63 = pd.read_csv('Data_2563.csv'); print('2563', data63.shape)
data62 = pd.read_csv('Data_2562.csv'); print('2562', data62.shape)
data61 = pd.read_csv('Data_2561.csv'); print('2561', data61.shape)

data1 = pd.read_csv('Data_13stations_2560_2564.csv'); print('กรมมลพิษ', data1.shape)
data2 = pd.read_csv('Data_13stations_2558_2564.csv'); print('งานพร', data2.shape)

all_data = pd.concat([data64,data63,data62,data61,data1,data2])
all_data = all_data[all_data['PM2.5(ug/m3)']>0]
all_data.shape; all_data.head()

print('pm2.5:',
      all_data[all_data['PM2.5(ug/m3)']>0].shape
      )

print('Gas:',
      all_data[(all_data['PM2.5(ug/m3)']>0)&(all_data['CO(ppm)']>0)&(all_data['NO(ppb)']>0)&(all_data['NO2(ppb)']>0)&(all_data['NOX(ppb)']>0)]
      )

print('Atmosphere:',
      all_data[(all_data['PM2.5(ug/m3)']>0)&(all_data['WS(m/s)']>0)&(all_data['WD(Deg)']>0)&(all_data['Temp(Deg.C)']>0)&(all_data['RH(%)']>0)]
      )

all_data.describe()

2564 (402959, 10)
2563 (403512, 10)
2562 (406271, 10)
2561 (303600, 10)
กรมมลพิษ (736416, 15)
งานพร (51134, 15)

Out[6]: (1646180, 15)
```

### Add More Features

```
In [9]: # Load original data
all_data['Year'] = all_data.apply(lambda x: str(x['Key_mth'])[0:4], axis=1)
all_data['Year_prev'] = all_data.apply(lambda x: str(int(x['Year'])-1), axis=1)
all_data['Month'] = all_data.apply(lambda x: str(x['Key_mth'])[4:6], axis=1)
all_data['Day'] = all_data.apply(lambda x: str(x['Key_mth'])[6:8], axis=1)
all_data['Hour'] = all_data.apply(lambda x: np.nan if (x['Hour']=='-1') or (x['Hour']=='nan') else str(x['Hour']), axis=1)
all_data['data_split'] = all_data.apply(lambda x: 'test' if x['Key_mth']>=20200325 else 'train', axis=1)
all_data['Date'] = pd.to_datetime(all_data['Key_mth'], format='%Y%m%d')
all_data['Weekday'] = all_data.apply(lambda x: x['Date'].weekday(), axis=1)
all_data['Weekyear'] = all_data.apply(lambda x: int(x['Date'].strftime('%W'))+1 if x['Year']!='2018' else int(x['Date'].strftime('%W')), axis=1)
print('All data:', all_data.shape, all_data.columns)

# Add pm2.5 Lag
def lag(data, target='PM2.5(ug/m3)', time_prev=['Year'], time_cur=['Month','Day','Hour']):
    time = time_prev + time_cur
    col_used = time + [target]
    col_final = [target] + [target + '_lag']
    master_lag = data[col_used].groupby(time).agg('mean').reset_index()
    col_join = list(map(lambda x: x+'_prev', time_prev)) + time_cur
    master_lag.columns = col_join + [target]
    all_data2 = data.merge(master_lag, how='left', on=col_join, suffixes=('', '_lag'))

    return all_data2

all_data_lag = lag(all_data, target='PM2.5(ug/m3)', time_prev=['Year'], time_cur=['Month'])
print('All data (Add pm2.5_lag):', all_data_lag.shape, all_data_lag.columns)

all_data_train = all_data_lag[all_data_lag['data_split']=='train']
all_data_test = all_data_lag[all_data_lag['data_split']=='test']
print('All train data:', all_data_train.shape, all_data_train.columns)
print('All test out data:', all_data_test.shape, all_data_test.columns)

# remove outlier of pm2.5 (For all data train)
all_data_train1 = all_data_train[(all_data_train['PM2.5(ug/m3)']<=165)]
print('All train data (pm2.5<=165):', all_data_train1.shape, all_data_train1.columns)

# set primary key
all_data_train1['Primary_key'] = all_data_train1.apply(lambda x: x['File']+str(x['Key_mth']), axis=1)
all_data_train1.set_index('Primary_key', inplace=True)

all_data_test['Primary_key'] = all_data_test.apply(lambda x: x['File']+str(x['Key_mth']), axis=1)
all_data_test.set_index('Primary_key', inplace=True)
```

## Data Transformation

```
In [11]: # Data preparation
x_time = ['Month', 'Hour']; x_time_woe = ['Month_WOE', 'Hour_WOE']
x_atm = ['WS(m/s)', 'WD(Deg)', 'Temp(Deg.C)', 'RH(%)', 'BP(mBar)']
x_gas = ['CO(ppm)', 'NO(ppb)', 'NO2(ppb)', 'O3(ppb)', 'PM10(ug/m3)'] #, 'PM2.5(ug/m3)_Lag', 'NOX(ppb)'
col_x = x_atm + x_gas #+ x_time
col_x_selected = x_atm + x_gas #+ x_time_woe
col_y = ['PM2.5(ug/m3)']

def convert_time_period(x):
    if x in [0,1,2,3,4,5]: return 'Dawn'
    elif x in [6,7,8,9,10]: return 'Morning'
    elif x in [11,12,13,14,15,16,17]: return 'Evening'
    elif x in [18,19,20,21,22,23]: return 'Night'
    else: return np.nan

def WOE_continuous(data, col_x, col_y):
    data[col_x] = data[col_x].fillna('NoData')
    k = data[[col_x, col_y]].groupby(col_x)[col_y].agg(['count', 'sum']).reset_index()
    k.columns = [col_x, 'Count', 'Sum']
    k['Sum %'] = (k['Sum'] / k['Sum'].sum()*100).round(2)
    k['Count %'] = (k['Count'] / k['Count'].sum()*100).round(2)
    k[col_x+' _WOE'] = np.log(k['Sum %'] / k['Count %']).round(2)
    k = k.sort_values(by=col_x+' _WOE')

    dict_woe = {}; new_col_woe = col_x+' _WOE'
    for row in k[col_x]:
        dict_woe[row]=k[k[col_x]==row][new_col_woe].values[0]

    return k, dict_woe
```

```
def prep_data(data_train, data_test, col_x, col_y, col_cat):

    data_train_tmp = data_train.copy()[col_x+col_y]
    data_test_tmp = data_test.copy()[col_x+col_y]

    data_train_tmp.dropna(inplace=True)
    data_test_tmp.dropna(inplace=True)

    X_train, X_test, y_train, y_test = train_test_split(data_train_tmp[col_x], data_train_tmp[col_y[0]], test_size=0.2)

    X_test_out = data_test_tmp[col_x]; y_test_out = data_test_tmp[col_y[0]]
    print('Train set', X_train.shape)
    print('Test set', X_test.shape)
    print('Test out set', X_test_out.shape)

    new_col_woe_ls = []
    if len(col_cat)>0:
        train_tmp = X_train.copy()
        train_tmp[col_y[0]] = y_train
        for col_woe in col_cat:
            df_woe, dict_woe = WOE_continuous(train_tmp, col_woe, col_y[0])
            new_col_woe = col_woe+' _WOE'
            new_col_woe_ls.append(new_col_woe)

            X_train[new_col_woe] = X_train[col_woe].map(dict_woe)
            X_test[new_col_woe] = X_test[col_woe].map(dict_woe)
            X_test_out[new_col_woe] = X_test_out[col_woe].map(dict_woe)

    # scale data
    col_scale = col_x + new_col_woe_ls #list(set(col_x)-set(col_cat))
    pd.options.mode.chained_assignment = None
    scaler = MinMaxScaler()
    scaler.fit(X_train.loc[:, col_scale])
    X_train.loc[:, col_scale] = scaler.transform(X_train.loc[:, col_scale].values)
    X_test.loc[:, col_scale] = scaler.transform(X_test.loc[:, col_scale].values)
    X_test_out.loc[:, col_scale] = scaler.transform(X_test_out.loc[:, col_scale].values)

    # - Log
    normal_scaler = PowerTransformer(standardize=True)
    normal_scaler.fit(X_train.loc[:, col_x])
    X_train.loc[:, col_x] = normal_scaler.transform(X_train.loc[:, col_x].values)
    X_test.loc[:, col_x] = normal_scaler.transform(X_test.loc[:, col_x].values)
    X_test_out.loc[:, col_x] = normal_scaler.transform(X_test_out.loc[:, col_x].values)

    return X_train, X_test, X_test_out, y_train, y_test, y_test_out
```

## 2. Coding ของกระบวนการ การควบคุมคุณภาพของข้อมูล และการคัดเลือกตัวแปร

### Feature selection

```
In [10]: def feature_selection(X_train, y_train, n_features = [4,5,6,7,8]):
result=pd.DataFrame()
for v_features in n_features:
rfe = RFE(estimator=RandomForestRegressor(), n_features_to_select=v_features)
rfe.fit(X_train, y_train)

model = pd.DataFrame()
model['model_rfe'] = ['RF']; model['n_features'] = [v_features];
model['features'] = [list(X_train.columns[rfe.support_])]
model['features_score'] = [list(rfe.estimator_.feature_importances_)]

result = pd.concat([result, model]).reset_index(drop = True)

return result
```

```
In [8]: tune_result = feature_selection(all_data_train2[col_x_selected], all_data_train2[col_y[0]], n_features = np.arange(1,14,1))
tune_result

# tune_result = pd.read_csv('feature_selection_lag_hour.csv')
# tune_result
```

```
Out[8]:
```

|    | model_rfe | n_features | features                                      | features_score                                      |
|----|-----------|------------|---|---|
| 0  | RF        | 1          | [PM10(ug/m3)]                                 | [1.0]   |
| 1  | RF        | 2          | [PM10(ug/m3), PM2.5(ug/m3)_lag]               | [0.6808810056407285, 0.3191189943592716]            |
| 2  | RF        | 3          | [WD(Deg), PM10(ug/m3), PM2.5(ug/m3)_lag]      | [0.14558324200418699, 0.6738364671832584, 0.18...   |
| 3  | RF        | 4          | [WD(Deg), PM10(ug/m3), PM2.5(ug/m3)_lag, ...] | [0.1155923411336168, 0.6579678662609816, 0.128...   |
| 4  | RF        | 5          | [WD(Deg), CO(ppm), PM10(ug/m3), PM2.5(u...    | [0.0899066022927673, 0.0833413133265937, 0.645...   |
| 5  | RF        | 6          | [WD(Deg), CO(ppm), NO(ppb), PM10(ug/m3)...    | [0.07201725834212863, 0.07123863923463042, 0.0...   |
| 6  | RF        | 7          | [WD(Deg), BP(mBar), CO(ppm), NO(ppb), ...]    | [0.06255349159664944, 0.051872187704887066, 0.0...  |
| 7  | RF        | 8          | [WD(Deg), Temp(Deg C), BP(mBar), CO(ppm)...   | [0.05196840274883894, 0.04507192142820617, 0.0...   |
| 8  | RF        | 9          | [WD(Deg), Temp(Deg C), BP(mBar), CO(ppm)...   | [0.04662895843710494, 0.037812136617411236, 0.0...  |
| 9  | RF        | 10         | [WD(Deg), Temp(Deg C), RH(%), BP(mBar)...     | [0.04203210752967176, 0.03270204001294556, 0.0...   |
| 10 | RF        | 11         | [WD(Deg), Temp(Deg C), RH(%), BP(mBar)...     | [0.03798967897617767, 0.02997621120083328, 0.0...   |
| 11 | RF        | 12         | [WD(Deg), Temp(Deg C), RH(%), BP(mBar)...     | [0.035969333908372605, 0.026912895268851128, 0.0... |
| 12 | RF        | 13         | [WS(m/s), WD(Deg), Temp(Deg C), RH(%), ...]   | [0.01815196168604188, 0.03292611006085553, 0.0...   |

```
In [207]: feature_selection = pd.DataFrame()
for i in range(len(tune_result)):
features_score = tune_result.loc[i,'features_score']
# features = tune_result.loc[i,'features']
features = ast.literal_eval(tune_result.loc[i,'features'])
n_features = len(features)
print('n_features=',n_features, features)

tmp_data = pd.DataFrame({'n_features':[str(n_features)]*n_features, 'features':features, 'features_score': features_score})
feature_selection = pd.concat([feature_selection,tmp_data])

n_features= 1 ['PM10(ug/m3)']
n_features= 2 ['WD(Deg)', 'PM10(ug/m3)']
n_features= 3 ['WD(Deg)', 'CO(ppm)', 'PM10(ug/m3)']
n_features= 4 ['WD(Deg)', 'CO(ppm)', 'PM10(ug/m3)', 'Month_WOE']
n_features= 5 ['WD(Deg)', 'CO(ppm)', 'PM10(ug/m3)', 'PM2.5(ug/m3)_lag', 'Month_WOE']
n_features= 6 ['WD(Deg)', 'BP(mBar)', 'CO(ppm)', 'PM10(ug/m3)', 'PM2.5(ug/m3)_lag', 'Month_WOE']
n_features= 7 ['WD(Deg)', 'Temp(Deg.C)', 'BP(mBar)', 'CO(ppm)', 'PM10(ug/m3)', 'PM2.5(ug/m3)_lag', 'Month_WOE']
n_features= 8 ['WD(Deg)', 'Temp(Deg.C)', 'BP(mBar)', 'CO(ppm)', 'NO(ppb)', 'PM10(ug/m3)', 'PM2.5(ug/m3)_lag', 'Month_WOE']
n_features= 9 ['WD(Deg)', 'Temp(Deg.C)', 'RH(%)', 'BP(mBar)', 'CO(ppm)', 'NO(ppb)', 'PM10(ug/m3)', 'PM2.5(ug/m3)_lag', 'Month_WOE']
n_features= 10 ['WD(Deg)', 'Temp(Deg.C)', 'RH(%)', 'BP(mBar)', 'CO(ppm)', 'NO(ppb)', 'O3(ppb)', 'PM10(ug/m3)', 'PM2.5(ug/m3)_lag', 'Month_WOE']
n_features= 11 ['WD(Deg)', 'Temp(Deg.C)', 'RH(%)', 'BP(mBar)', 'CO(ppm)', 'NO(ppb)', 'NO2(ppb)', 'O3(ppb)', 'PM10(ug/m3)', 'PM2.5(ug/m3)_lag', 'Month_WOE']
n_features= 12 ['WD(Deg)', 'Temp(Deg.C)', 'RH(%)', 'BP(mBar)', 'CO(ppm)', 'NO(ppb)', 'NO2(ppb)', 'O3(ppb)', 'PM10(ug/m3)', 'PM2.5(ug/m3)_lag', 'Month_WOE', 'Hour_WOE']
n_features= 13 ['WS(m/s)', 'WD(Deg)', 'Temp(Deg.C)', 'RH(%)', 'BP(mBar)', 'CO(ppm)', 'NO(ppb)', 'NO2(ppb)', 'O3(ppb)', 'PM10(ug/m3)', 'PM2.5(ug/m3)_lag', 'Month_WOE', 'Hour_WOE']
```

## 3. Coding ของกระบวนการพัฒนาแบบจำลอง

```

Model

In [13]: def score(model, X_train, y_train, X_test, y_test, X_test_out, y_test_out):
    y_pred_train = model.predict(X_train)
    # y_pred_train_valid = model.predict(X_train_valid)
    y_pred_test = model.predict(X_test)
    y_pred_test_out = model.predict(X_test_out)

    def r2_adj(y, y_pred, X):
        n, p = X.shape
        return 1 - ((1 - r2_score(y, y_pred)) * (n - 1)) / (n - p - 1)

    result = pd.DataFrame(data =
        {
            'R2-adj_train': [r2_adj(y_train, y_pred_train, X_train)], 'R2-adj_test': [r2_adj(y_test, y_pred_test, X_test)],
            'MSE_train': [mean_squared_error(y_train, y_pred_train, squared=True)], 'MSE_test': [mean_squared_error(y_test, y_pred_test, squared=True)],
            'RMSE_train': [mean_squared_error(y_train, y_pred_train, squared=False)], 'RMSE_test': [mean_squared_error(y_test, y_pred_test, squared=False)],
            'MAE_train': [mean_absolute_error(y_train, y_pred_train)], 'MAE_test': [mean_absolute_error(y_test, y_pred_test)]
        }
    )

    return result

def model_tune_params(X, y, model, param_grid, n_iter=5):

    def r2_adj(y, y_pred, X):
        n, p = X.shape
        return 1 - ((1 - r2_score(y, y_pred)) * (n - 1)) / (n - p - 1)

    scoring = {"R2-adj": make_scorer(r2_adj, X=X),
               "MSE": make_scorer(mean_squared_error, squared=True),
               "RMSE": make_scorer(mean_squared_error, squared=False),
               "MAE": make_scorer(mean_absolute_error)}

    gs = RandomizedSearchCV(model, param_grid, n_iter=n_iter, cv=5, scoring=scoring, refit="R2-adj", return_train_score=True, n_jobs=-1)
    gs.fit(X, y)
    gs_result = pd.DataFrame(gs.cv_results_)
    col = [x for x in gs_result.columns if x.startswith(('mean_', 'std_', 'params')) and not x.endswith(('time'))]
    gs_all = gs_result[col].sort_values(by='mean_test_R2-adj', ascending=False).reset_index(drop=True)
    gs_best = pd.DataFrame(gs_all.iloc[0,:]).transpose().reset_index(drop=True)

    return gs_all, gs_best, gs.best_params_, gs.best_estimator_

```

```

In [14]: def model_tune_params_by_n_features(X_train, y_train, X_test, y_test, X_test_out, y_test_out, n_iter=5):
    #dict_params = {}
    # k-neighbor
    param_grid = {'n_neighbors': [5,10,50,100,300,500,1000], 'weights': ['uniform']}
    model = KNeighborsRegressor()
    kn_gs_all, kn_gs_best, kn_param, kn_model = model_tune_params(X_train, y_train, model, param_grid, n_iter=n_iter)
    kn_score = score(kn_model, X_train, y_train, X_test, y_test, X_test_out, y_test_out)
    kn_gs_best = pd.concat([kn_gs_best, kn_score], axis=1)
    kn_gs_best['algorithm'] = 'K-neighbors'; kn_gs_best['best_param'] = [kn_param]; kn_gs_best['best_model'] = [kn_model]
    #dict_params['KN']['best_param'] = kn_param; dict_params['KN']['best_model'] = kn_model

    # Linear regression
    param_grid = {'alpha': [0,0.1,0.3,0.5,1,3,5,10]}
    model = Ridge()
    lr_gs_all, lr_gs_best, lr_param, lr_model = model_tune_params(X_train, y_train, model, param_grid, n_iter=n_iter)
    lr_score = score(lr_model, X_train, y_train, X_test, y_test, X_test_out, y_test_out)
    lr_gs_best = pd.concat([lr_gs_best, lr_score], axis=1)
    lr_gs_best['algorithm'] = 'Ridge'; lr_gs_best['best_param'] = [lr_param]; lr_gs_best['best_model'] = [lr_model]
    #dict_params['LR']['best_param'] = lr_param; dict_params['LR']['best_model'] = lr_model

    # Decision tree
    param_grid = {"max_depth": [3,4,5,6,7,8,9,10],
                  "min_samples_split": [100,300,500]}
    model = DecisionTreeRegressor()
    dt_gs_all, dt_gs_best, dt_param, dt_model = model_tune_params(X_train, y_train, model, param_grid, n_iter=n_iter)
    dt_score = score(dt_model, X_train, y_train, X_test, y_test, X_test_out, y_test_out)
    dt_gs_best = pd.concat([dt_gs_best, dt_score], axis=1)
    dt_gs_best['algorithm'] = 'DecisionTree'; dt_gs_best['best_param'] = [dt_param]; dt_gs_best['best_model'] = [dt_model]
    #dict_params['DT']['best_param'] = dt_param; dict_params['DT']['best_model'] = dt_model

    # Random forest
    param_grid = {"n_estimators": [100,300,500],
                  "max_depth": [3,4,5,6,7,8,9,10],
                  "min_samples_split": [100,300,500]}
    model = RandomForestRegressor()
    rf_gs_all, rf_gs_best, rf_param, rf_model = model_tune_params(X_train, y_train, model, param_grid, n_iter=n_iter)
    rf_score = score(rf_model, X_train, y_train, X_test, y_test, X_test_out, y_test_out)
    rf_gs_best = pd.concat([rf_gs_best, rf_score], axis=1)
    rf_gs_best['algorithm'] = 'RandomForest'; rf_gs_best['best_param'] = [rf_param]; rf_gs_best['best_model'] = [rf_model]
    #dict_params['RF']['best_param'] = rf_param; dict_params['RF']['best_model'] = rf_model

```



```

# Gradient boosting
param_grid = {"learning_rate": np.arange(0.05,0.4,0.05), # Learning rate
             "n_estimators": [100,300,500],
             "max_depth": [3,4,5,6,7,8,9,10],
             "min_samples_split": [100,300,500]
            }

model = GradientBoostingRegressor()
gb_gs_all, mlp_gs_best, gb_param, gb_model = model_tune_params(X_train, y_train, model, param_grid, n_iter=n_iter)
gb_score = score(gb_model, X_train, y_train, X_test, y_test, X_test_out, y_test_out)
gb_gs_best = pd.concat([gb_gs_best, mlp_score], axis=1)
gb_gs_best['algorithm'] = 'GradientBoosting'; gb_gs_best['best_param'] = [gb_param]; gb_gs_best['best_model'] = [gb_model]
#dict_params['GB']['best_param'] = gb_param; dict_params['GB']['best_model'] = gb_model

# mlp
param_grid = {'hidden_layer_sizes': np.arange(1, len(X_train.columns)+1), 'activation': ['tanh', 'relu']}
model = MLPRegressor(max_iter=2000)
mlp_gs_all, mlp_gs_best, mlp_param, mlp_model = model_tune_params(X_train, y_train, model, param_grid, n_iter=n_iter)
mlp_score = score(mlp_model, X_train, y_train, X_test, y_test, X_test_out, y_test_out)
mlp_gs_best = pd.concat([mlp_gs_best, mlp_score], axis=1)
mlp_gs_best['algorithm'] = 'MLP'; mlp_gs_best['best_param'] = [mlp_param]; mlp_gs_best['best_model'] = [mlp_model]
#dict_params['MLP']['best_param'] = mlp_param; dict_params['MLP']['best_model'] = mlp_model

gs_result = pd.concat([kn_gs_best, lr_gs_best, dt_gs_best, rf_gs_best, gb_gs_best, mlp_gs_best])
tuple_model = (kn_model, lr_model, dt_model, rf_model, gb_model, mlp_model)

n_features = len(X_train.columns); col_x = list(X_train.columns)

gs_result['n_features'] = n_features
gs_result['features'] = str(col_x)
gs_result['data_record (train)'] = X_train.shape[0]
gs_result['data_record (test)'] = X_test.shape[0]
gs_result['data_record (test out)'] = X_test_out.shape[0]

return gs_result, tuple_model

```

```

In [17]: def model_result_by_n_features(features):

        col_x_selected_tmp = features
        col_x_tmp = [col.replace('_WOE','') if col.endswith('_WOE') else col for col in col_x_selected_tmp]
        col_cat = [col.replace('_WOE','') for col in col_x_selected_tmp if col.endswith('_WOE')]
        col_y_tmp = ['PM2.5(ug/m3)']

        print(col_x_tmp)
        X_train, X_test, X_test_out, y_train, y_test, y_test_out = prep_data(all_data_train1, all_data_test, col_x_tmp, col_y_tmp, cc

        final, tuple_model = model_tune_params_by_n_features(X_train[col_x_selected_tmp], y_train, X_test[col_x_selected_tmp], y_test
        display(final[['algorithm','params','mean_train_R2-adj','mean_test_R2-adj','R2-adj_train','R2-adj_test','R2-adj_test_out'],'de

        # final = model_final_by_n_features(X_train[col_x_selected_tmp], y_train, X_test_out[col_x_selected_tmp], y_test_out, dict_pa
        # display(final)

        return final, tuple_model

```

```

In [43]: ### End
start = datetime.now()
features = tune_result.loc[0,'features']
# features = ast.literal_eval(tune_result.loc[0,'features'])
final1, tuple_model1 = model_result_by_n_features(features)
final1.to_csv('1.model_evaluation_1.csv', index=False)
# pickle.dump(tuple_model1, open("1.model_1.pkl", 'wb'))
end = datetime.now()
print('total time:', (end-start).total_seconds()/60, 'min')

['PM10(ug/m3)']
Train set (256632, 1)
Test set (64159, 1)
Test out set (537928, 1)

/usr/local/lib64/python3.6/site-packages/sklearn/model_selection/_search.py:282: UserWarning: The total space of parameters 2 is
s smaller than n_iter=5. Running 2 iterations. For exhaustive searches, use GridSearchCV.
% (grid_size, self.n_iter, grid_size), UserWarning)

```

|   | algorithm        | params  | mean_train_R2-adj | mean_test_R2-adj | R2-adj_train | R2-adj_test | R2-adj_test_out | data_record (train) | data_record (test) | data_record (test out) |
|---|------------------|---|-------------------|------------------|--------------|-------------|-----------------|---------------------|--------------------|------------------------|
| 0 | K-neighbors      | {'weights': 'uniform', 'n_neighbors': 50}         | 0.687069          | 0.683997         | 0.687173     | 0.682570    | 0.620115        | 256632              | 64159              | 537928                 |
| 0 | Ridge            | {'alpha': 0.5}                                    | 0.618619          | 0.618607         | 0.618619     | 0.617681    | 0.524772        | 256632              | 64159              | 537928                 |
| 0 | DecisionTree     | {'min_samples_split': 300, 'max_depth': 10}       | 0.693062          | 0.687877         | 0.692724     | 0.685619    | 0.614255        | 256632              | 64159              | 537928                 |
| 0 | RandomForest     | {'n_estimators': 300, 'min_samples_split': 100... | 0.692269          | 0.689256         | 0.691975     | 0.686744    | 0.623888        | 256632              | 64159              | 537928                 |
| 0 | GradientBoosting | {'n_estimators': 300, 'min_samples_split': 500... | 0.699129          | 0.68652          | 0.698017     | 0.684910    | 0.619469        | 256632              | 64159              | 537928                 |
| 0 | MLP              | {'hidden_layer_sizes': 1, 'activation': 'tanh'}   | 0.676166          | 0.676162         | 0.676350     | 0.673000    | 0.600919        | 256632              | 64159              | 537928                 |

total time: 68.33293675 min

## ประวัติการศึกษาและการทำงาน

ชื่อ นามสกุล นาย กฤษณ์ เจ็ดวรรณะ

วัน เดือน ปีเกิด 21 มีนาคม 2519

ภูมิลำเนา พัทลุง

### ประวัติการศึกษา

| วุฒิการศึกษา                              | ชื่อสถาบัน  | ปีที่สำเร็จการศึกษา |
|---|---|---------------------|
| วิศวกรรมศาสตรมหาบัณฑิต<br>(วิศวกรรมขนส่ง) | มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี                     | 2546                |
| วิศวกรรมศาสตรบัณฑิต<br>(วิศวกรรมโยธา)     | สถาบันเทคโนโลยีพระจอมเกล้า<br>เจ้าคุณทหารลาดกระบัง (สจล.) | 2542                |

### ตำแหน่งและสถานที่ทำงานปัจจุบัน

รองอธิการบดีฝ่ายวางแผนและกายภาพ

มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร

399 ถนนสามเสน แขวงวชิรพยาบาล เขตดุสิต กรุงเทพมหานคร 10300

โทรศัพท์ 02-665-3777