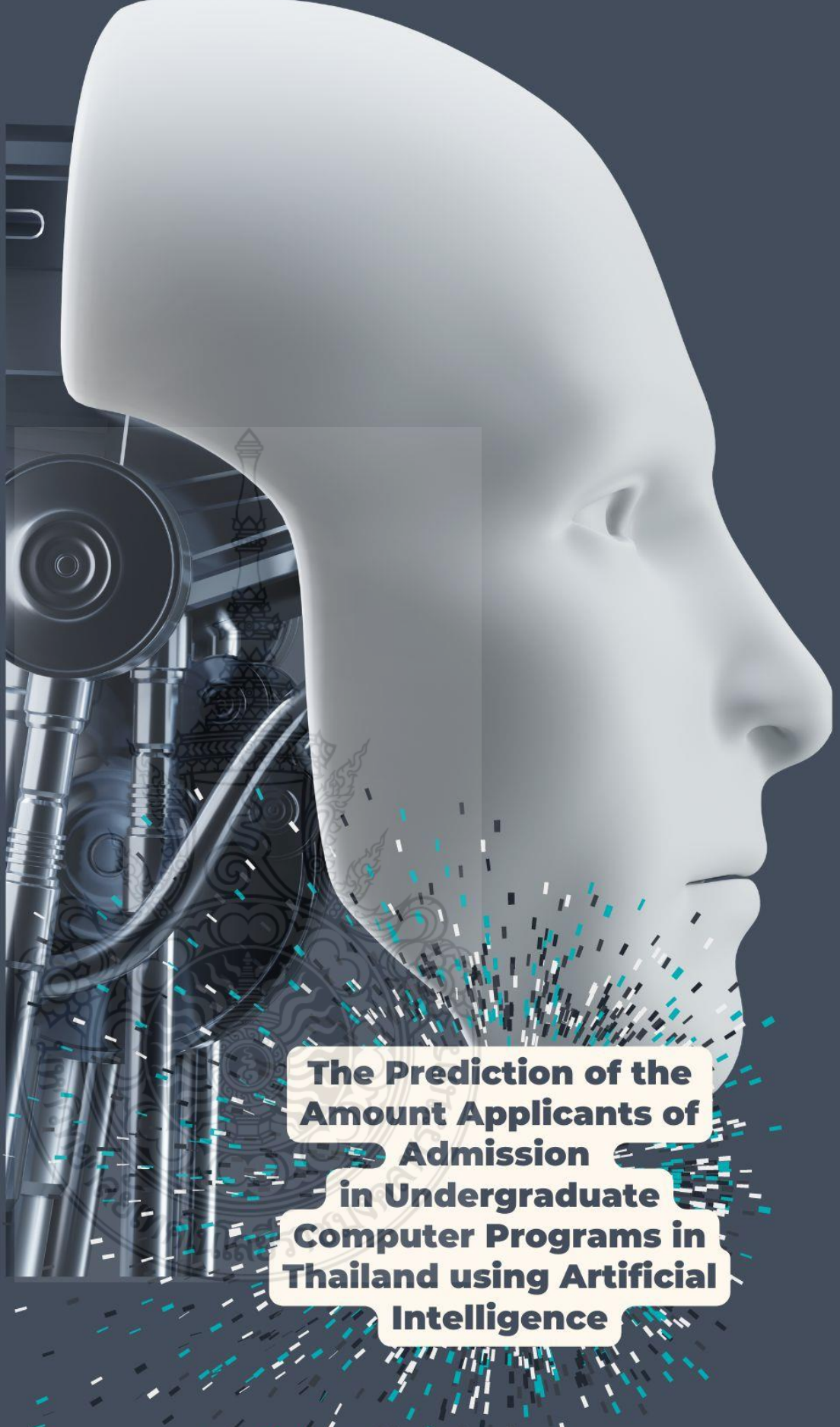
รายงานวิจัยฉบับสมบูรณ์

การทำนายจำนวนผู้สมัครเข้าศึกษาในหลักสูตรปริญญาตรี ด้านคอมพิวเตอร์ในประเทศไทยโดยใช้ปัญญาประดิษฐ์



The Prediction of the Amount Applicants of Admission in Undergraduate Computer Programs in Thailand using Artificial Intelligence

วีรวรรณ ฉันทนะทรัพย์

งานวิจัยนี้ได้รับทุนสนับสนุนจากเงินกองทุนวิจัย
ประจำปีงบประมาณ พ.ศ. 2567
คณะวิทยาศาสตร์และเทคโนโลยี
มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร

การทำนายจำนวนผู้สมัครเข้าศึกษา ในหลักสูตรปริญญาตรีด้านคอมพิวเตอร์ในประเทศไทยโดยใช้ปัญญาประดิษฐ์

The Prediction of the Amount Applicants of Admission in Undergraduate Computer Programs in Thailand using Artificial Intelligence

วีรวรรณ จันทนะทรัพย์

งานวิจัยนี้ได้รับทุนสนับสนุนจากเงินกองทุนวิจัย ประจำปีงบประมาณ พ.ศ. ๒๕๖๗
คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร

ชื่อเรื่อง : การทำนายจำนวนผู้สมัครเข้าศึกษาในหลักสูตรปริญญาตรีด้านคอมพิวเตอร์
ในประเทศไทยโดยใช้ปัญญาประดิษฐ์
ผู้วิจัย : ผศ.ดร.วีรวรรณ จันทนะทรัพย์
พ.ศ. : 2567

บทคัดย่อ

งานวิจัยนี้ดำเนินการวิจัยภายใต้การบูรณาการศาสตร์ด้านปัญญาประดิษฐ์โดยใช้การเรียนรู้ของเครื่องบนพื้นฐานการเรียนรู้แบบมีผู้สอนในงานทำนายข้อมูลจำนวนผู้สมัครเข้าศึกษาในหลักสูตรปริญญาตรีด้านคอมพิวเตอร์จากระบบการคัดเลือกกลางบุคคลเข้าศึกษาในสถาบันอุดมศึกษาของไทย หรือระบบ TCAS จำนวน 6 ปีการศึกษาระหว่างปี พ.ศ. 2561-2566 โดยใช้ข้อมูลปี พ.ศ. 2561-2565 เป็นข้อมูลสำหรับการเรียนรู้ และข้อมูลปี พ.ศ. 2566 เป็นข้อมูลสำหรับทดสอบประสิทธิภาพการทำงาน พัฒนาแบบจำลองด้วยภาษาไพธอนร่วมกับไลบรารีการเรียนรู้ของเครื่องและการเรียนรู้เชิงลึก ดำเนินการเปรียบเทียบประสิทธิภาพผลการทำนายข้อมูล จำนวน 5 แบบจำลอง ประกอบด้วย 1) แบบจำลองสุ่มป่าไม้ 2) แบบจำลองเพื่อนบ้านใกล้สุด 3) แบบจำลองซัพพอร์ตเวกเตอร์รีเกรสชัน 4) แบบจำลองโครงข่ายประสาทเทียมเพอร์เซฟตรอนแบบหลายชั้น และ 5) แบบจำลองโครงข่ายประสาทเทียมแบบความจำขนาดสั้นระยะยาว

ผลการทดสอบแสดงให้เห็นว่า แบบจำลองสุ่มป่าไม้มีประสิทธิภาพในการทำนายสูงสุดในทุกมิติของการทดสอบ โดยมีค่าสัมประสิทธิ์การทำนาย (Coefficient of Determination; R^2) ร้อยละ 93 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) เท่ากับ 944 ราย แบบจำลองเพื่อนบ้านใกล้สุดแสดงให้เห็นถึงประสิทธิภาพรองลงมา โดยมีค่า R^2 ร้อยละ 90 และค่า MSE เท่ากับ 1,414 ราย ส่วนแบบจำลองซัพพอร์ตเวกเตอร์รีเกรสชันมีประสิทธิภาพรองลงมามีค่า R^2 ร้อยละ 89 และค่า MSE เท่ากับ 1,614 ราย ในส่วนประสิทธิภาพการทำนายของแบบจำลองโครงข่ายประสาทเทียมแบบความจำขนาดสั้นระยะยาว และแบบจำลองโครงข่ายประสาทเทียมเพอร์เซฟตรอนแบบหลายชั้นแสดงให้เห็นถึงผลลัพธ์ที่ค่อนข้างต่ำโดยมีค่า R^2 ร้อยละ 88 และร้อยละ 82 มีค่า MSE เท่ากับ 1,688 ราย และ 2,495 ราย ตามลำดับ

คำสำคัญ : การทำนาย, ปัญญาประดิษฐ์

Title : The Prediction of the Amount Applicants of Admission in Undergraduate Computer Programs in Thailand using Artificial Intelligence
Researcher : Asst. Prof. Veerawan Janthanasub, Ph.D.
Year : 2024

Abstract

This research integrates artificial intelligence and machine learning, specifically utilizing supervised learning techniques, to predict the number of applicants for undergraduate computer science programs within Thailand's central university admission system (TCAS). The study covers six academic years, from 2018 to 2023, utilizing data from 2018 to 2022 for model training and data from 2023 for model performance testing. The model was developed using Python, incorporating both machine learning and deep learning libraries. A comparative analysis was conducted on the predictive performance of five models: 1) Random Forest (RF), 2) K-Nearest Neighbors (KNN), 3) Support Vector Regression (SVR), 4) Multi-Layer Perceptron neural network (MLP) and 5) Long Short-Term Memory neural network (LSTM).

The test results revealed that the Random Forest (RF) model achieved the highest predictive performance across all testing dimensions, evidenced by the coefficient of determination (R^2) value of 93% and a mean squared error (MSE) of 944. The K-Nearest Neighbors (KNN) model followed, with an R^2 of 90% and MSE of 1,414. The Support Vector Regression (SVR) model was slightly lower at R^2 of 89% and MSE of 1,614. In terms of predictive performance, the Long Short-Term Memory (LSTM) neural network and the Multi-Layer Perceptron (MLP) neural network showed relatively low results, with R^2 values of 88% and 82%, and MSE values of 1,688 and 2,495, respectively.

Keywords: Prediction; Artificial Intelligence

กิตติกรรมประกาศ

โครงการวิจัยเรื่อง การทำนายจำนวนผู้สมัครเข้าศึกษาในหลักสูตรปริญญาตรีด้านคอมพิวเตอร์ในประเทศไทยโดยใช้ปัญญาประดิษฐ์ ได้รับการสนับสนุนงบประมาณจากมหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร ประจำปีงบประมาณ พ.ศ. 2567 ซึ่งในการดำเนินงานวิจัยในครั้งนี้ ผู้วิจัยขอขอบพระคุณทุกหน่วยงานที่มีส่วนให้การสนับสนุนทำให้ผลการวิจัยสำเร็จลุล่วงไปด้วยดี ได้แก่ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร ซึ่งเป็นหน่วยงานต้นสังกัดของผู้วิจัย และเอื้อเฟื้อสถานที่ อุปกรณ์ต่าง ๆ ในการดำเนินงานวิจัยในครั้งนี้ รวมทั้งหน่วยงานภายในคณะที่เกี่ยวข้อง ผู้วิจัยขอขอบคุณเป็นอย่างสูง

ขอขอบคุณสมาคมที่ประชุมอธิการบดีแห่งประเทศไทยหน่วยงานพัฒนาระบบการคัดเลือกกลางบุคคลเข้าศึกษาในสถาบันอุดมศึกษา (Thai University Central Admission System; TCAS) และเป็นเจ้าของข้อมูลสถิติการรับสมัคร สถิติการสอบ ซึ่งงานวิจัยนี้นำข้อมูลมาใช้ในการสร้างแบบจำลองการเรียนรู้ และใช้เป็นข้อมูลสำหรับทดสอบประสิทธิภาพผลการทำนายข้อมูล และขอขอบคุณเจ้าของผลงานที่ได้ใช้เป็นแหล่งอ้างอิงข้อมูลที่ปรากฏในเล่มรายงานวิจัยฉบับนี้ ผู้วิจัยรู้สึกซาบซึ้งเป็นอย่างยิ่งจึงใคร่ขอขอบพระคุณเป็นอย่างสูง

คุณค่าและประโยชน์อันพึงมีจากงานวิจัยนี้ ผู้วิจัยขอมอบบูชาแด่ บิดา มารดา ที่ให้การอบรมสั่งสอนเลี้ยงดู และบูชาแด่คณาจารย์ทุกท่านที่ประสาทวิชาความรู้แก่ผู้วิจัย

วีรวรรณ จันทนะทรัพย์

สารบัญ

สารบัญตาราง

สารบัญภาพ

สารบัญภาพ (ต่อ)

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การจัดการศึกษาในระดับอุดมศึกษาของไทยในปัจจุบัน จัดการศึกษาตามบริบทความสอดคล้องกับเกณฑ์มาตรฐานหลักสูตร ซึ่งประกอบด้วย (1) ทิศทางนโยบายและยุทธศาสตร์การพัฒนากำลังคนของประเทศ และพันธกิจหลักและยุทธศาสตร์ของสถาบันการศึกษา (2) ความเสี่ยงและผลกระทบจากภายนอก อาทิ การเปลี่ยนแปลงทางเทคโนโลยี นโยบาย และสิ่งแวดล้อมอื่น ๆ ในบริบทโลก (3) ผลการสำรวจการรับฟังความคิดเห็นจากผู้ที่มีส่วนเกี่ยวข้อง อาทิ ผู้ใช้งานบัณฑิต ผู้เรียน และนักเรียนที่ต้องการเข้าเรียนในหลักสูตรการศึกษา (4) ผลลัพธ์การเรียนรู้ การออกแบบ การจัดการเรียนการสอน และระบบการบริหารจัดการการศึกษา ตามที่กำหนดในเกณฑ์มาตรฐานการศึกษาหลักสูตร รวมทั้งกรอบมาตรฐานคุณวุฒิสาขาหรือสาขาวิชา (ถ้ามี) (5) ผลการดำเนินงานของหลักสูตรที่ผ่านมาสำหรับกรณีปรับปรุงหลักสูตรเดิม (6) ผลการประเมินความพึงพอใจของผู้เรียน บัณฑิต ผู้ใช้บัณฑิต ศิษย์เก่า ตลอดจนข้อร้องเรียนจากบุคคลหรือหน่วยงานภายนอก และบุคคลภายในสถาบันอุดมศึกษา และ (7) ผลการประเมินคุณภาพภายนอกระดับหลักสูตร โดยรายการความสอดคล้องข้างต้นได้กำหนดไว้ในประกาศคณะกรรมการมาตรฐานการอุดมศึกษา เรื่อง หลักเกณฑ์ วิธีการ และเงื่อนไขในการแต่งตั้งหรือมอบหมายผู้ตรวจสอบและการตรวจสอบการดำเนินการจัดการศึกษาของสถาบันอุดมศึกษา พ.ศ. 2565 (ราชกิจจานุเบกษา, 2565) ดังนั้นการบริการจัดการหลักสูตรสาขาวิชาของสถาบันอุดมศึกษาจึงต้องดำเนินการให้สอดคล้องกับบริบทเกณฑ์ข้างต้น อย่างไรก็ตามการได้มาซึ่งข้อมูลต่าง ๆ ข้างต้น ต้องใช้เวลา งบประมาณ และกำลังคนในการปฏิบัติงาน โดยเฉพาะอย่างยิ่งความต้องการศึกษาต่อของนักเรียนในสาขาวิชาต่าง ๆ ในหลักสูตรที่เปิดสอนในปัจจุบัน หรือหลักสูตรใหม่ที่กำลังจะเปิดการเรียนการสอนภายใต้บริบทการเปลี่ยนแปลงของประเทศและของโลก อาทิ หลักสูตรในสาขาวิชาที่เกี่ยวข้องการเรียนการสอนด้านคอมพิวเตอร์ เพราะเนื่องจากในปัจจุบันเทคโนโลยีสารสนเทศ คอมพิวเตอร์ เครือข่าย และวิทยาการข้อมูลเป็นเทคโนโลยีที่ยอมรับกันว่าขับเคลื่อนอุตสาหกรรมธุรกิจ และการพัฒนานวัตกรรมสิ่งใหม่ของประเทศ และของโลก ดังนั้นจึงทำให้เกิดความต้องการกำลังคนที่มีความรู้ความสามารถปฏิบัติงานด้านคอมพิวเตอร์อย่างมากในตลาดแรงงาน ส่งผลให้มีนักเรียนที่มีความต้องการเรียนในหลักสูตรด้านคอมพิวเตอร์มากตามไปด้วย ดังจะเห็นได้จากข้อมูลจำนวนผู้สมัครเรียนในสาขาวิชาวิทยาการคอมพิวเตอร์ในปีการศึกษา 2565 ของมหาวิทยาลัยเกษตรศาสตร์ มีจำนวนทั้งสิ้น 2,815 ราย โดยมีจำนวนผู้ที่ได้เข้าเรียนเพียง 24 ราย หากคิดอัตราการแข่งขันการสอบเข้าเรียนได้ต่ำมากร้อยละ 0.85 (สมาคมที่ประชุมอธิการบดีแห่งประเทศไทย, 2567)

ผู้วิจัยจึงมีแนวคิดนำเสนอการนำข้อมูลจำนวนผู้สมัครเข้าศึกษาต่อในระดับปริญญาตรีในระบบการคัดเลือกกลางบุคคลเข้าศึกษาในสถาบันอุดมศึกษาของไทย หรือระบบ TCAS ในปีการศึกษา 2561-2566 รวมจำนวนทั้งสิ้น 6 ปีการศึกษา มาประยุกต์กับศาสตร์ด้านสถิติพยากรณ์และศาสตร์ด้านปัญญาประดิษฐ์ เพื่อสร้างแบบจำลองพยากรณ์จำนวนความต้องการเข้าศึกษาต่อในระดับปริญญาตรีในสาขาวิชาด้านคอมพิวเตอร์ ซึ่งเป็นสาขาวิชาที่กำลังได้รับความสนใจของนักเรียน เนื่องจากเป็นสาขาวิชาที่ตลาดแรงงานมีความต้องการ โดยมีจุดมุ่งหวังใช้เป็นเครื่องมือสำหรับทำนายจำนวนความต้องการเข้าศึกษาต่อในระดับอุดมศึกษาของนักเรียน และนำผลการพยากรณ์ไปใช้ในการบริหารจัดการการศึกษาของหลักสูตรด้านคอมพิวเตอร์ของสถาบันอุดมศึกษา

1.2 วัตถุประสงค์การวิจัย

เพื่อศึกษารูปแบบการทำนายจำนวนผู้สมัครเข้าศึกษาต่อในระดับปริญญาตรีด้านคอมพิวเตอร์ของไทยที่เหมาะสมในการนำไปใช้เป็นเครื่องมือสนับสนุนการตัดสินใจในการบริหารจัดการหลักสูตรด้านคอมพิวเตอร์ของสถาบันอุดมศึกษาในประเทศไทย

1.3 ขอบเขตของการวิจัย

ในการศึกษาวิจัยในครั้งนี้เป็นการศึกษาวิธีการค้นหาแบบจำลองด้านปัญญาประดิษฐ์สำหรับทำนายข้อมูลจำนวนผู้สมัครเข้าเรียนในระดับชั้นปริญญาตรีในสาขาวิชาด้านคอมพิวเตอร์จากระบบคัดเลือกกลางบุคคลเข้าศึกษาในสถาบันอุดมศึกษาของไทย (TCAS) กำหนดขอบเขตการวิจัยดังนี้

1.3.1 กลุ่มตัวอย่างที่ในใช้ในการดำเนินงานวิจัยในครั้งนี้ คือ ข้อมูลสถิติการสมัคร สถิติการสอบ จำนวน 6 ปีการศึกษาระหว่าง พ.ศ. 2561-2566 จากระบบการคัดเลือกกลางบุคคลเข้าศึกษาในสถาบันอุดมศึกษา (Thai University Central Admission System; TCAS) เข้าถึงผ่านลิงก์เว็บ https://www.mytcas.com/stat/ กำหนดรายละเอียดข้อมูลดังนี้

- ชุดข้อมูลสำหรับการฝึกสอนและสร้างแบบจำลองใช้ข้อมูลปีการศึกษา 5 ปี คือ ปี พ.ศ. 2561-2565
- ชุดข้อมูลสำหรับการทดสอบประสิทธิภาพของแบบจำลองใช้ข้อมูลปีการศึกษา พ.ศ. 2566

1.3.2 แบบจำลองทำนายข้อมูลในงานวิจัยนี้ศึกษา พัฒนา และเปรียบแบบจำลองด้วยวิธีการเรียนรู้ของเครื่องจำนวน 5 อัลกอริทึม ดังนี้

- แบบจำลองสุ่มป่าไม้ (Random Forest Regression Model; RFR Model)
- แบบจำลองเพื่อนบ้านใกล้สุด (K-Nearest Neighbors Model; KNN Model)
- แบบจำลองซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression Model; SVR Model)
- โครงข่ายประสาทเทียมเพอร์เซฟตรอนแบบหลายชั้น (Multi-Layer Perceptron; MLP)
- โครงข่ายประสาทเทียมแบบความจำขนาดสั้นระยะยาว (LSTM Neural Networks)

1.3.3 การวัดประสิทธิภาพของแบบจำลองทำนายข้อมูล เพื่อหาแบบจำลองที่เหมาะสมใช้ค่าสัมประสิทธิ์การทำนาย (Coefficient of Determination; R^2) และค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared Error; MSE)

1.4 นิยามศัพท์

1.4.1 การเรียนรู้ของเครื่อง (Machine Learning) เป็นสาขาหนึ่งของปัญญาประดิษฐ์ที่เน้นการพัฒนาและการออกแบบระบบที่สามารถเรียนรู้และปรับปรุงการปฏิบัติตามข้อมูลได้โดยอัตโนมัติโดยไม่ต้องโปรแกรมให้แก่คอมพิวเตอร์อย่างชัดเจน การเรียนรู้ของเครื่องสามารถใช้ข้อมูลในการเรียนรู้และสร้างแบบจำลองหรือโมเดลทางคณิตศาสตร์เพื่อพยากรณ์ผลลัพธ์หรือการกระทำในอนาคต โดยการเรียนรู้ของเครื่องสามารถแบ่งออกเป็นหลายลักษณะ เช่น การเรียนรู้แบบมีผู้สอน (Supervised Learning) การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) และการเรียนรู้แบบเสริมแรง (Reinforcement Learning) เป็นต้น การเรียนรู้ของเครื่องถูกนำไปประยุกต์ใช้งานอย่างกว้างขวางในหลายอุตสาหกรรม เช่น การจัดการข้อมูล การวิเคราะห์ข้อมูล การทำนายและการตัดสินใจ และการปรับปรุงการกระทำอัตโนมัติในงานซอฟต์แวร์และหุ่นยนต์ (OpenAI, 2024)

1.4.2 แบบจำลองการทำนายข้อมูล (Prediction Models) เป็นโมเดลทางคณิตศาสตร์หรืออัลกอริทึมที่ใช้ในการทำนายค่าหรือสถานะของข้อมูลที่ไม่ทราบล่วงหน้าโดยใช้ข้อมูลที่มีอยู่เป็นพื้นฐาน บทบาทหลักของแบบจำลองทำนายข้อมูล คือ การใช้ข้อมูลอดีตเพื่อสร้างโมเดลทางคณิตศาสตร์ที่สามารถทำนายผลลัพธ์ของข้อมูลใหม่ได้โมเดลเหล่านี้สามารถใช้กับข้อมูลที่มีหลายแบบ เช่น:

- การทำนายข้อมูลตามกาลเวลา (Time Series Forecasting) ใช้ข้อมูลที่มีลำดับเวลาเพื่อพยากรณ์ค่าในอนาคต เช่น การพยากรณ์ยอดขายของสินค้าในอนาคต
- การทำนายจำนวนจริง (Regression) ใช้เพื่อทำนายค่าต่อเนื่อง โดยปกติใช้กับข้อมูลที่เป็นตัวเลข เช่น การพยากรณ์ราคาของอสังหาริมทรัพย์ในอนาคต
- การจำแนกประเภท (Classification) ใช้เพื่อแยกแยะข้อมูลออกเป็นกลุ่มต่าง ๆ ตัวอย่างเช่น การจำแนกอีเมลล์เป็นกลุ่มของอีเมลล์ขยะหรือไม่ใช่ขยะ

การสร้างแบบจำลองทำนายข้อมูลจะเกี่ยวข้องกับการเลือกแบบจำลองที่เหมาะสม การปรับพารามิเตอร์ การตรวจสอบและปรับปรุงโมเดลเพื่อให้มีประสิทธิภาพมากที่สุดต่อข้อมูลที่ใช้ในการฝึกสอนและทดสอบโมเดล (OpenAI, 2024)

1.4.3 ระบบการคัดเลือกกลางบุคคลเข้าศึกษาในสถาบันอุดมศึกษาในประเทศไทย (Thai University Central Admission System) หรือเรียกชื่อย่อว่าระบบ TCAS คือ ระบบกลางที่ออกแบบโดยที่ประชุมอธิการบดีแห่งประเทศไทย เริ่มต้นใช้งานในปีการศึกษา พ.ศ. 2561 บริหารจัดการรับสมัครเข้าศึกษาต่อในระดับอุดมศึกษาแบ่งออกเป็น 5 รอบ ประกอบด้วย รอบที่ 1 การรับด้วยแฟ้มสะสมผลงาน (Portfolio) รอบที่ 2 การรับแบบโควตาโดยมีการสอบข้อเขียนหรือข้อสอบปฏิบัติ รอบ 3 การรับตรงร่วมกัน รอบ 4 การรับแบบ Admission และรอบ 5 การรับตรงอิสระ (สมาคมที่ประชุมอธิการบดีแห่งประเทศไทย, 2567)

1.5 วิธีการวิจัย

วิธีการดำเนินการงานวิจัยเรื่อง การทำนายจำนวนผู้สมัครเข้าศึกษาในหลักสูตรปริญญาตรีด้านคอมพิวเตอร์ในประเทศไทยโดยใช้ปัญญาประดิษฐ์ มีขั้นตอนวิธีการในการดำเนินงานวิจัยดังนี้

1.5.1 ศึกษาปัญหาและงานวิจัยที่เกี่ยวข้องกับการสร้างแบบจำลองทำนายข้อมูล และข้อมูลต่าง ๆ ประกอบด้วย ข้อมูลหลักสูตรสาขาวิชาด้านคอมพิวเตอร์ของสถาบันอุดมศึกษา ข้อมูลจำนวนผู้สมัครเข้าศึกษาต่อของระบบ TCAS ในปีการศึกษา พ.ศ. 2561-2566 เพื่อใช้ในการออกแบบและพัฒนาแบบจำลองทำนายข้อมูล

1.5.2 เก็บรวบรวมข้อมูลจากระบบ TCAS

1.5.3 จัดเตรียมข้อมูลสำหรับการเรียนรู้ และข้อมูลสำหรับการทดสอบ

1.5.4 พัฒนาแบบจำลองทำนายข้อมูล

1.5.5 ทดสอบประสิทธิภาพแบบจำลองที่ได้พัฒนาขึ้น

1.5.6 สรุปผลและจัดทำรายงานผลการวิจัย

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1.6.1 แบบจำลองทำนายข้อมูลที่เหมาะสมสำหรับใช้ทำนายจำนวนผู้สมัครเข้าศึกษาในหลักสูตรปริญญาตรีด้านคอมพิวเตอร์ในประเทศไทย

1.6.2 ผลลัพธ์จากการทำนายจำนวนผู้สมัครเข้าศึกษาในหลักสูตรปริญญาตรีด้านคอมพิวเตอร์ที่แม่นยำซึ่งสามารถนำไปใช้ประโยชน์ในการตัดสินใจวางแผนการบริหารจัดการหลักสูตรด้านคอมพิวเตอร์ของสถาบันการศึกษาในประเทศไทยได้

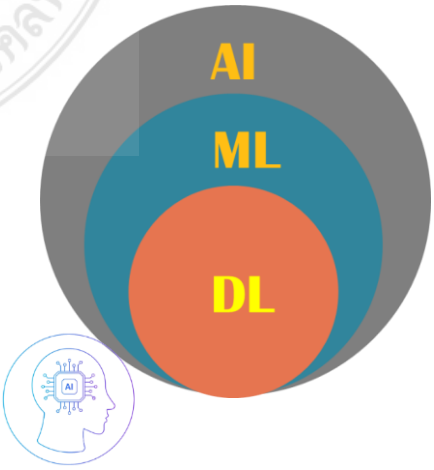
บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

ทฤษฎีและงานวิจัยที่เกี่ยวข้องกับการทำนายข้อมูลหรือพยากรณ์ข้อมูลในปัจจุบันมีอยู่หลากหลายวิธี โดยทั่วไปแบ่งออกเป็น 2 หลักการ คือ หลักการพยากรณ์โดยใช้สถิติ และวิธีการเรียนรู้ของเครื่อง (Machine Learning; ML) ซึ่งปัจจุบันวิธีการเรียนรู้ของเครื่องได้รับความนิยมและใช้กันมาก งานวิจัยนี้มุ่งเน้นการศึกษาพัฒนา และเปรียบเทียบประสิทธิภาพผลการทำนายข้อมูลของแบบจำลองการเรียนรู้ของเครื่องซึ่งเป็นศาสตร์ด้านปัญญาประดิษฐ์ (AI) มาประยุกต์ใช้ในการทำนายจำนวนผู้สมัครเข้าศึกษาต่อในระดับปริญญาตรีด้านคอมพิวเตอร์ของไทย ผ่านระบบการคัดเลือกกลางบุคคลเข้าศึกษาในสถาบันอุดมศึกษาของไทย หรือระบบ TCAS ซึ่งมีรายละเอียดเอกสาร ทฤษฎี หลักวิชา และงานวิจัยที่เกี่ยวข้องดังนี้

2.1 ปัญญาประดิษฐ์ (Artificial Intelligence; AI)

ปัญญาประดิษฐ์ หรือ AI เป็นเทคโนโลยีที่สำคัญในยุคดิจิทัล โดยมีนิยาม คือ เทคโนโลยีการสร้างเครื่องจักรให้มีความฉลาดเหมือนมนุษย์ ทั้งการคิดแบบมนุษย์ การกระทำแบบมนุษย์ การคิดอย่างมีเหตุผล และการกระทำอย่างมีเหตุผล (สำนักงานส่งเสริมเศรษฐกิจดิจิทัล, ม.ป.ป.) สำหรับความเป็นมาของปัญญาประดิษฐ์เริ่มต้นครั้งแรกในช่วงปี 1950-1980 โดยในปี ค.ศ. 1956 ได้มีการจัดงานประชุมสัมมนาเชิงปฏิบัติการสำหรับนักวิจัยที่สนใจให้ประเด็นหัวข้อที่เกี่ยวข้องกับ AI อาทิ ทฤษฎีออโตมาตา โครงข่ายประสาทเทียม และอัลกอริทึมที่เกี่ยวข้องกับการพัฒนาให้เครื่องจักร หรือคอมพิวเตอร์มีความชาญฉลาด (ปริญญา สงวนสัตย์, 2562) ต่อมาในช่วงปี ค.ศ. 1980-2000 วิวัฒนาการคอมพิวเตอร์ได้พัฒนาอย่างต่อเนื่องและถูกนำมาใช้ขับเคลื่อนเศรษฐกิจและสังคม ทฤษฎีและหลักการสำคัญที่ถูกนำมาใช้ในการพัฒนาศาสตร์ด้านปัญญาประดิษฐ์ได้ถูกพัฒนาขึ้นภายใต้ชื่อ การเรียนรู้ของเครื่อง (Machine Learning; ML) และนับเป็นหัวใจสำคัญของการพัฒนาผลงานด้านปัญญาประดิษฐ์ให้มีกว้างขวาง และได้รับการยอมรับอย่างแพร่หลาย และในห่วงเวลาปี ค.ศ. 2020 จนถึงปัจจุบัน ศาสตร์ด้านปัญญาประดิษฐ์ได้รับความสนใจเป็นอย่างมาก แนวทางการพัฒนาด้วยการเรียนรู้ของเครื่อง ได้ถูกพัฒนาให้มีประสิทธิภาพสูงขึ้นกับงานที่มีความซับซ้อนและยากต่อการประมวลผลและเข้าใจของเครื่องจักรก็ได้ถูกพัฒนาขึ้นภายในใต้ชื่อ การเรียนรู้เชิงลึก (Deep Learning) ด้วยแนวคิดที่ว่าเครื่องจักรจะเรียนรู้คุณลักษณะ รูปแบบ ความรู้ ได้เองโดยที่มนุษย์ไม่ต้องป้อนคำตอบหรือสอนให้กับเครื่องจักร เหมาะสมกับงานด้านการประมวลผลภาพดิจิทัล (Digital Image Processing) และด้านประมวลผลสัญญาณเสียง และอื่น ๆ ที่มีรูปแบบที่ซับซ้อนและยากต่อการเรียนรู้

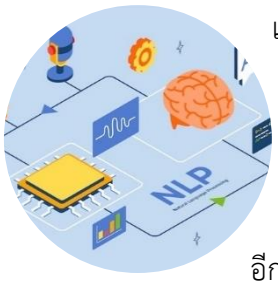
จากนิยาม และวิวัฒนาการความเป็นมาของปัญญาประดิษฐ์ ก็พอจะสรุปได้ถึงศาสตร์และหลักวิชาที่เกี่ยวข้องกับการพัฒนาให้เครื่องจักรให้มีความชาญฉลาดได้เหมือนมนุษย์ ประกอบด้วย ปัญญาประดิษฐ์ (AI) การเรียนรู้ของเครื่อง (ML) และการเรียนรู้เชิงลึก (DL) นิยามขอบเขตของศาสตร์แสดงดังภาพที่ 2-1



ภาพที่ 2-1 แบบจำลองความสัมพันธ์ของศาสตร์ด้าน AI

ปัจจุบันเทคโนโลยีปัญญาประดิษฐ์ นับเป็นหนึ่งในเทคโนโลยีที่ได้รับความสนใจมาก เนื่องจากเทคโนโลยี AI ก่อให้เกิดประโยชน์ในงานหลากหลายประเภท และช่วยเพิ่มขีดความสามารถในการแข่งขันให้แก่องค์กรได้ ทั้งยังมีแนวโน้มที่จะถูกใช้ร่วมกับเทคโนโลยี Cloud Computing และ Big Data ซึ่งทำให้เกิดการส่งถ่ายข้อมูลไปมาระหว่างเครื่องจักรในระบบเครือข่ายอินเทอร์เน็ต เมื่อเครื่องจักรหนึ่งสามารถสร้างองค์ความรู้ใหม่ขึ้นมาได้ จะสามารถถ่ายทอดองค์ความรู้นี้ไปยังเครื่องจักรอื่น ๆ ภายในเวลาอันรวดเร็ว จึงทำให้เกิดการพัฒนาความรู้ใหม่ที่ต่อยอดจากความรู้เดิมอยู่ตลอดเวลา (สำนักงานส่งเสริมเศรษฐกิจดิจิทัล, ม.ป.ป.) โดยแขนงวิชาที่นำปัญญาประดิษฐ์มาบูรณาการประกอบด้วย

ก. การประมวลผลภาษาธรรมชาติ (Natural Language Processing; NLP)



เป็นเทคโนโลยีด้านปัญญาประดิษฐ์ ที่นำหลักการของการเรียนรู้ของเครื่อง (ML) มาพัฒนาให้เครื่องจักรหรือคอมพิวเตอร์มีความชาญฉลาดและความสามารถในการตีความ จัดการ และทำความเข้าใจภาษามนุษย์ได้ ผลงานในยุคแรกของ NLP อาทิ Siri บนโทรศัพท์มือถือ iPhone ของของบริษัท Apple ผู้ใช้งานสามารถสนทนาผ่านอุปกรณ์ไมโครโฟนระบบจะรับข้อความเสียงแล้วนำมาประมวลผลและโต้ตอบ อีกหนึ่งตัวอย่างคือ Google Assistant ของบริษัท Google ก็เป็นอีกหนึ่งผลงานด้าน NLP ซึ่งการพัฒนางานวิจัยและนวัตกรรมมด้านการประมวลผลภาษาธรรมชาติมีมาอย่างต่อเนื่อง (ปริญญา สงวนสัตย์, 2562) และในเดือนมีนาคม ปี พ.ศ. 2566 การเปิดตัวแชทบอท (Chatbot) ที่พลิกโลกภายใต้ชื่อ ChatGPT ของบริษัท OpenAI แชทบอท ChatGPT พัฒนาจากโมเดลการเรียนรู้ทางภาษาขนาดใหญ่โดยเรียนรู้มากกว่า 1.76 ล้านล้านพารามิเตอร์ และผู้ใช้งานทั่วโลกยอมรับในความสามารถของแชทบอทอัจฉริยะ ChatGPT และนำไปประยุกต์ใช้งานอย่างแพร่หลาย และในเวลาต่อมาในเดือน กรกฎาคม ปี พ.ศ. 2566 Mark Zuckerberg ผู้บริหารและผู้ก่อตั้งโซเชียลมีเดีย Facebook ได้เปิดตัว Llama ซึ่งเป็นโมเดลภาษาขนาดใหญ่ เรียนรู้พารามิเตอร์ 70 ล้านพารามิเตอร์ ต่อมาในเดือนพฤศจิกายนในปีเดียวกันบริษัทไมโครซอฟต์เปิดตัว Copilot และในเดือนธันวาคมบริษัท Google ก็เปิดตัว Gemini ซึ่งทำให้คอมพิวเตอร์ในปัจจุบันของโลกถูกนำไปขับเคลื่อนธุรกิจอุตสาหกรรม การศึกษา และอื่น ๆ อยู่ภายใต้เครื่องมือที่เรียกว่า Generative AI (GEN AI) (อมรินทร์ทีวีออนไลน์, 2566) ตัวอย่างผลงาน GEN AI ของ 4 บริษัทใหญ่ของโลก แสดงดังภาพที่ 2-2

ภาพที่ 2-2 ตัวอย่างผลงานด้าน NLP ของ 4 บริษัทใหญ่ด้านการพัฒนา AI ของโลก

ข. คอมพิวเตอร์วิทัศน์ (Computer Vision)

เป็นเทคโนโลยี และศาสตร์ที่เกี่ยวข้องกับการพัฒนาให้เครื่องจักร หรือคอมพิวเตอร์ หรืออุปกรณ์อินเตอร์เน็ตแห่งสรรพสิ่ง (IoT) สามารถมองเห็น และมีความชาญฉลาดในการตรวจจับ จดจำว่าสิ่งที่มองเห็นนั้นคือวัตถุใด หรือเป็นอะไร ในศาสตร์ด้านนี้เกี่ยวข้องกับกล้องดิจิทัลซึ่งเป็นอุปกรณ์แทนตาของเครื่องจักรด้วยการรับสัญญาณภาพต่าง ๆ โดยรอบ และสัญญาณภาพเหล่านั้นจะถูกนำมาผ่านกระบวนการประมวลผลภาพด้วยศาสตร์ด้านการประมวลผลภาพดิจิทัล (Digital Image Processing) สัญญาณภาพที่ผ่านการประมวลผลแล้วจะถูกนำมาหาคุณลักษณะ (Feature Extraction) และข้อมูลคุณสมบัติเหล่านั้นก็จะถูกนำเข้าสู่การเรียนรู้ของเครื่อง (ML) เพื่อทำให้เครื่องจักรเรียนรู้ว่าคือวัตถุใด หรือภาพอะไร และนำไปใช้ประโยชน์ในด้านต่าง ๆ ต่อไป อาทิ การรู้จำใบหน้า การตรวจจับการหลับในของผู้ขับขี่ การตรวจจับวัตถุต่าง ๆ การรู้จำลายนิ้วมือ การจดจำป้ายทะเบียนรถยนต์ การตรวจความผิดปกติของผลิตภัณฑ์ การติดตามตำแหน่ง การติดตามผู้เล่นด้านกีฬา เป็นต้น ในศาสตร์ด้านคอมพิวเตอร์วิทัศน์มักนำหลักการของการเรียนรู้เชิงลึก (DL) เข้ามาช่วยหาคุณลักษณะของข้อมูลภาพ ทั้งนี้เพื่อช่วยลดเวลาและภาระของขั้นตอนการหาคุณลักษณะ เพราะการเรียนรู้เชิงลึกมนุษย์ไม่ต้องสอนเครื่องจักรแต่เครื่องจักรจะเรียนรู้จากข้อมูลนำเข้าที่ป้อนให้เรียนรู้เอง (ปริญญา สงวนสัตย์, 2562)

ค. วิทยาการหุ่นยนต์ (Robotics)

เป็นเทคโนโลยีที่ถูกนำมาใช้ขับเคลื่อนเศรษฐกิจและสังคมบนพื้นฐานการปฏิวัติอุตสาหกรรม 4.0 หลักการคือ การทำให้เครื่องจักรมีลักษณะทางกายภาพรูปร่างคล้ายมนุษย์ หรือในรูปแบบอุปกรณ์เครื่องมือเครื่องใช้ในภาคอุตสาหกรรม ภาคธุรกิจต่าง ๆ หรือในรูปแบบเครื่องใช้ในชีวิตประจำวัน เช่น เครื่องดูดฝุ่นที่ใช้ทำความสะอาดบ้านหรือ ที่อยู่อาศัยให้สามารถทำงานด้วยความชาญฉลาดในงานเฉพาะด้านที่ได้ถูกโปรแกรมไว้ ปัจจุบันความก้าวหน้าด้านวิทยาการหุ่นยนต์ได้เป็นที่ยอมรับ และบริษัทด้านเทคโนโลยีหุ่นยนต์ได้ผลิตหุ่นยนต์ออกเป็นสินค้าจำหน่ายในเชิงพาณิชย์ อาทิ หุ่นยนต์ส่งอาหารในร้านอาหารต่าง ๆ หุ่นยนต์พนักงานต้อนรับ หุ่นยนต์ทำความสะอาดบ้าน เป็นต้น (ปริญญา สงวนสัตย์, 2562)

ง. ระบบผู้เชี่ยวชาญ (Expert System; ES)

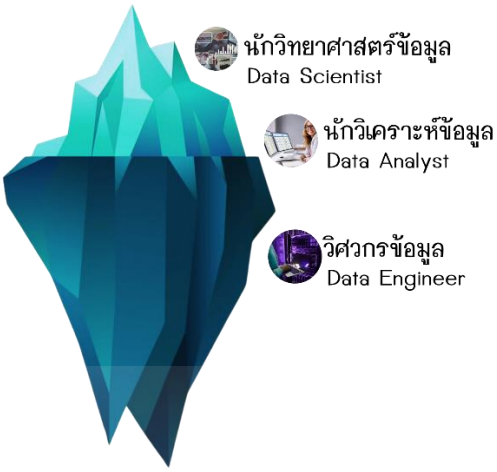
เป็นเทคโนโลยีที่พัฒนาให้เครื่องจักรมีความชาญฉลาด ในการคิดวิเคราะห์ โดยใช้ความรู้จากข้อมูลในอดีตและปัจจุบัน (Knowledge Base) และนำเสนอข้อมูลที่ผ่านการวิเคราะห์เพื่อช่วยในการตัดสินใจให้แก่มนุษย์ หรือตัดสินใจดำเนินการแบบอัตโนมัติแทนมนุษย์ หลักการพัฒนาระบบผู้เชี่ยวชาญจะนำศาสตร์ด้านปัญญาประดิษฐ์มาใช้ในการสกัดความรู้ หรือข้อมูลต่าง ๆ และนำไปพัฒนาทำให้ระบบสามารถคิดวิเคราะห์ แยกแยะข้อมูล สารสนเทศอย่างมีเหตุและผล จนได้ผลลัพธ์ที่ดีและเหมาะสมที่สุดในการช่วยตัดสินใจในงานเฉพาะด้าน ตัวอย่างระบบผู้เชี่ยวชาญ อาทิ ระบบการซื้อขายหุ้น ระบบการวินิจฉัยโรคในงานด้านการแพทย์ หรือระบบช่วยวิเคราะห์และตัดสินใจในการขับขี่แบบอัตโนมัติของรถยนต์ เครื่องบิน อากาศยานไร้คนขับ เป็นต้น (ปริญญา สงวนสัตย์, 2562)

จ. วิทยาการข้อมูล (Data Science)

วิทยาศาสตร์ข้อมูล คือ เทคโนโลยีสำหรับขับเคลื่อนองค์กรทั้งในภาครัฐ ภาคธุรกิจเอกชน ด้วยการนำศาสตร์วิชาด้านคณิตศาสตร์ สถิติ ปัญญาประดิษฐ์ วิศวกรรมคอมพิวเตอร์และวิทยาการคอมพิวเตอร์ มาวิเคราะห์ข้อมูลชนิดต่าง ๆ ที่มีจำนวนมหาศาล หรือที่เรียกว่าข้อมูลขนาดใหญ่ (Big Data) สกัดข้อมูลให้เห็นข้อมูลในเชิงลึก และนำไปใช้ประโยชน์ในด้านต่าง ๆ

โดยอาชีพที่เกี่ยวข้องกับศาสตร์ด้านวิทยาการข้อมูลประกอบด้วย 3 อาชีพหลัก คือ นักวิทยาศาสตร์ข้อมูล (Data scientist) นักวิเคราะห์ข้อมูล (Data Analyst) และวิศวกรข้อมูล (Data Engineer) ซึ่ง 3 อาชีพหลักข้างต้นมีความต้องการอย่างมากในตลาดแรงงาน

สำหรับความสัมพันธ์ของ 3 อาชีพแสดงดังภาพภูเขาน้ำแข็งภาพที่ 2-3



ภาพที่ 2-3 แบบจำลองภูเขาน้ำแข็งกับความสัมพันธ์ของ 3 อาชีพในศาสตร์วิทยาการข้อมูล

2.2 การเรียนรู้ของเครื่อง (Machine Learning; ML)

การเรียนรู้ของเครื่อง เป็นหัวใจสำคัญของการขับเคลื่อนปัญญาประดิษฐ์ให้ได้รับการยอมรับ และใช้งานอย่างกว้างขวางในปัจจุบัน สำหรับนิยามความหมาย หลักการ ประเภท และงานของการเรียนรู้ของเครื่องมีรายละเอียดดังนี้

นิยามความหมาย

การเรียนรู้ของเครื่อง หมายถึง การทำให้เครื่องจักร หรือคอมพิวเตอร์ หรืออุปกรณ์ประมวลผลอื่น ๆ เรียนรู้งานใดงานหนึ่ง (Task) จากตัวอย่างข้อมูลสำหรับการเรียนรู้ (Sample Data) หรือประสบการณ์ (Experience) จำนวนหนึ่งเพื่อให้ทำงานนั้นอย่างมีประสิทธิภาพ (Performance) และนำผลลัพธ์การเรียนรู้ที่ได้ไปใช้ประโยชน์ในงานด้านต่าง ๆ อาทิ แยกประเภท (Classification) จัดกลุ่ม (Clustering) พยากรณ์เหตุการณ์ หรือทำนายข้อมูลในอนาคต (Forecasting) เป็นต้น (ปริญญา สงวนสัตย์, 2562)

ประเภทและหลักการทำงาน

การเรียนรู้ของเครื่องปัจจุบันสามารถแบ่งออกได้ 3 ประเภท คือ การเรียนรู้แบบมีผู้สอน (Supervised Learning) การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Leaning) และการเรียนรู้แบบปรับตัว (Reinforcement Learning) โดยแต่ละประเภทมีหลักการขั้นตอนวิธี (Algorithm) หน้าที่การทำงาน (Task) ที่แตกต่างกัน รายละเอียดมีดังนี้ (Mark E. Fenner, 2020)

การเรียนรู้แบบมีผู้สอน
(Supervised Learning)

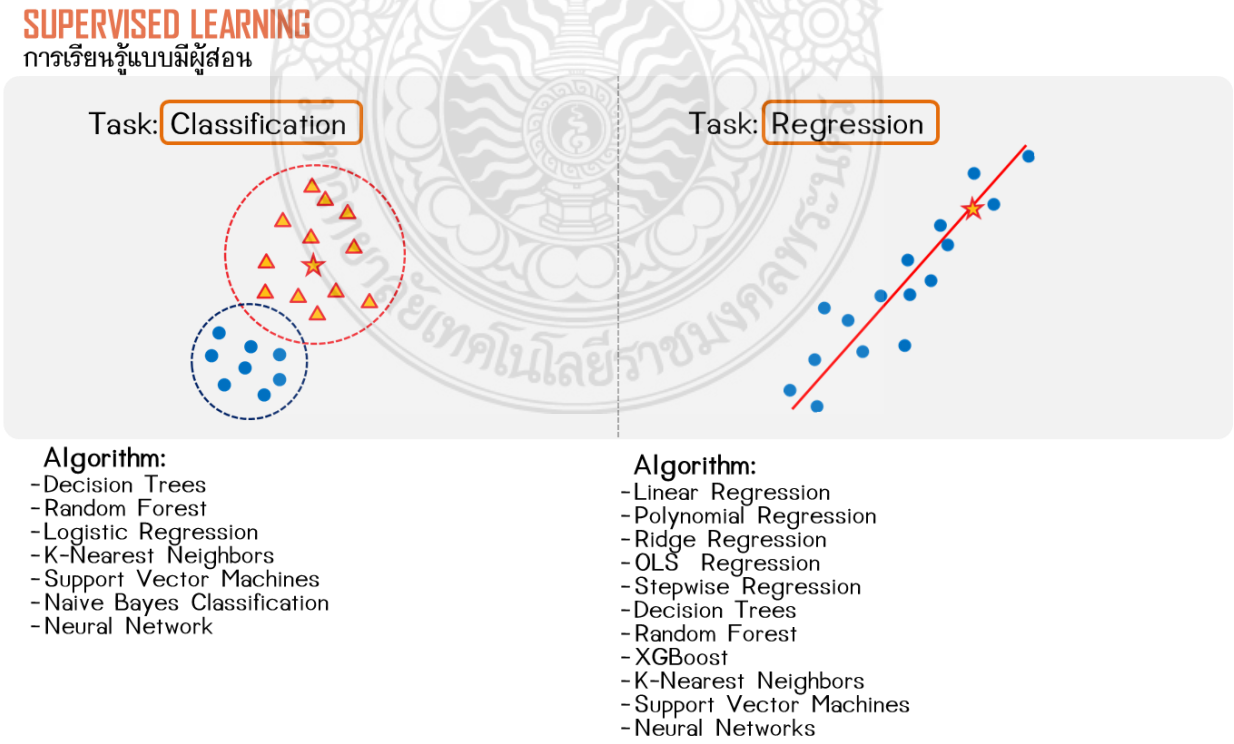
เป็นการทำให้เครื่องจักรเรียนรู้จากชุดข้อมูลตัวอย่างชุดหนึ่ง ซึ่งเรียกว่าชุดข้อมูลสำหรับการเรียนรู้ (Training Dataset) โดยชุดข้อมูลประกอบด้วยข้อมูลคุณลักษณะสำคัญของปัญหา และข้อมูลคำตอบ หรือป้ายกำกับ (Label) ผลลัพธ์ของการเรียนรู้ของเครื่องกับชุดข้อมูลสำหรับการฝึกการเรียนรู้ในระยะหนึ่ง เครื่องจักรจะสามารถหาคำตอบของปัญหาได้ด้วยตนเอง และเมื่อต้องการทดสอบประสิทธิภาพ (Performance) ผลการเรียนรู้ของเครื่องจะทดสอบกับข้อมูลอีกชุดที่เรียกว่าชุดข้อมูลสำหรับการทดสอบ (Testing Dataset) ผลการหาคำตอบจากโมเดลการเรียนรู้ของเครื่องกับชุดข้อมูลสำหรับการทดสอบจะบ่งบอกถึงประสิทธิภาพของผลการเรียนรู้ของเครื่องนั่นเอง (Mark E. Fenner, 2020)

สำหรับงาน (Task) ของการเรียนรู้แบบมีผู้สอนแบ่งออกเป็น 2 งาน คือ (Mark E. Fenner, 2020)

- งานจำแนกประเภท (Classification) ผลลัพธ์ของการจำแนกประเภทจะอยู่ในรูปแบบประเภทของกลุ่ม ซึ่งมีชนิดข้อมูลเป็น แคตตากอรี่ (Category) ตัวอย่างการนำไปใช้ อาทิ การวิเคราะห์ความรู้สึก (Sentiment Analysis) การจำแนกลายมือ (Handwritten Recognition) การตรวจจับวัตถุ (Object Detection) การรู้จำใบหน้า (Face Recognition) การกรองสแปม (Spam Filtering) เป็นต้น สำหรับแบบจำลองสำหรับงานจำแนกประเภทข้อมูล อาทิ แบบจำลองต้นไม้ตัดสินใจ (Decision Tree) แบบจำลองสุ่มป่าไม้ (Random Forest) แบบจำลองถดถอยแบบโลจิสติกส์ (Logistic Regression) แบบจำลองเพื่อนบ้านใกล้สุด (K-Nearest Neighbors) แบบจำลองซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine; SVM) และแบบจำลองโครงข่ายประสาทเทียม (Neural Network) แบบต่าง ๆ เป็นต้น

- งานทำนายหรือพยากรณ์ข้อมูล (Regression) ผลลัพธ์ของงานทำนายจะอยู่ในรูปแบบของข้อมูลต่อเนื่อง หรือข้อมูลในรูปแบบตัวเลข เช่น ปริมาณน้ำฝน ราคาน้ำมัน ราคาทองคำ ราคาบ้าน ราคาหุ้น จำนวนลูกค้า เป็นต้น อัลกอริทึมสำหรับงานทำนายข้อมูล อาทิ การถดถอยเชิงเส้น (Linear Regression) การถดถอยแบบเส้นโค้ง (Polynomial Regression) อัลกอริทึมต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึมสุ่มป่าไม้ (Random Forest) อัลกอริทึม XGBoost อัลกอริทึมเพื่อนบ้านใกล้สุด (K-Nearest Neighbors) อัลกอริทึมอัลกอริทึมซัพพอร์ตเวกเตอร์รีเกสชัน (Support Vector Regression; SVR) และโครงข่ายประสาทเทียม (Neural Network) แบบต่าง ๆ เป็นต้น

แบบจำลองงานในการเรียนรู้ของเครื่องแบบมีผู้สอนแสดงดังภาพที่ 2-4



ภาพที่ 2-4 แบบจำลองงาน (Task) ในการเรียนรู้ของเครื่องแบบมีผู้สอน

การเรียนรู้แบบไม่มีผู้สอน
(Unsupervised Learning)

เป็นการทำให้เครื่องจักรเรียนรู้จากชุดข้อมูลตัวอย่างชุดหนึ่ง โดยไม่ต้องกำหนดป้ายกำกับ (Label) หรือค่าเป้าหมาย (Target) ของแต่ละข้อมูล เครื่องจักรจะเรียนรู้ได้จากการวิเคราะห์คุณลักษณะของข้อมูล (Data Analytics) แล้วสร้างแบบแผนด้วยการจัดกลุ่มของข้อมูล สำหรับงาน (Task) ของการเรียนรู้แบบไม่มีผู้สอนประกอบด้วยงาน 3 งาน คือ (Mark E. Fenner, 2020)

- งานการจัดกลุ่ม (Clustering)

เป็นเทคนิควิธีสำหรับการจัดกลุ่มหรือจำแนกกลุ่มที่ไม่การกำหนดค่าเป้าหมาย (Target) หรือคลาส (Class) ไว้ล่วงหน้า แต่เครื่องจักรจะเรียนรู้จากคุณลักษณะของข้อมูลนำเข้าแล้วสร้างรูปแบบการจัดกลุ่มข้อมูลไว้ใน Cluster ต่าง ๆ ตัวอย่างการประยุกต์ใช้งาน อาทิ การแนะนำผลิตภัณฑ์ต่าง ๆ ให้กับลูกค้าที่ได้ดำเนินจัดกลุ่มไว้แล้ว การจัดกลุ่มลูกค้า หรือจัดกลุ่มลูกหนี้ เป็นต้น ตัวอย่างอัลกอริทึมสำหรับจัดกลุ่มข้อมูล อาทิ อัลกอริทึม K-Means Clustering, Hierarchical Clustering และ DBSCAN เป็นต้น

- งานหากฎความสัมพันธ์ (Association Rule)

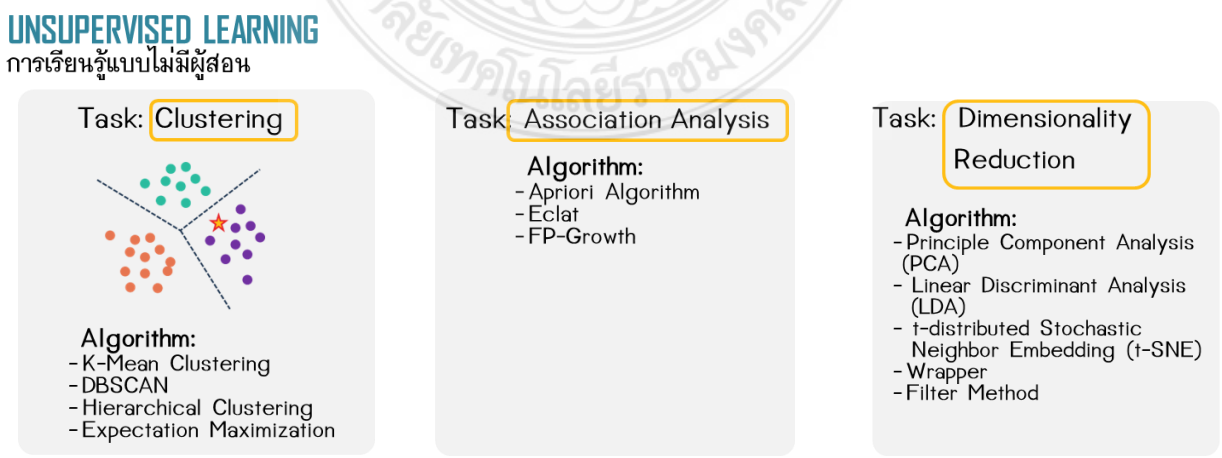
เป็นเทคนิคการหาความสัมพันธ์ของข้อมูลด้วยกฎ ผลลัพธ์ทำให้ทราบว่าข้อมูลต่าง ๆ มีความเกี่ยวข้องกันอย่างไร ด้วยการวัดค่าความสัมพันธ์ (Association Value) ค่าสนับสนุน (Support Value) และค่าความเชื่อมั่น (Confidence Value) โดยเทคนิคการหากฎความสัมพันธ์ถูกนำไปประยุกต์ใช้กับการคาดการณ์ยอดขายและส่วนลด การวิเคราะห์และออกแบบโปรโมชันสินค้าร่วมรายการ การจัดวางตำแหน่งสินค้าบนชั้นวาง เป็นต้น สำหรับอัลกอริทึม อาทิ Apriori, Eclat และ FP-Growth

- งานลดมิติของข้อมูล (Dimensionality Reduction)

เป็นเทคนิควิธีที่ใช้ในการลดจำนวนมิติของข้อมูล หรือลด Features ของชุดข้อมูลที่มีความซ้ำซ้อนลง โดยที่สารสนเทศ (Information) ไม่สูญเสีย โดยมีวัตถุประสงค์เพื่อให้ง่ายต่อการเรียนรู้ของเครื่อง โดยมีเป้าหมายประกอบด้วย

- การเลือกคุณลักษณะของข้อมูล (Feature Selection) ตัวอย่างอัลกอริทึม อาทิ วิธีการใช้ตัวกรอง (Filtering Methods) วิธีการ Wrapper และวิธีการ Embedded เป็นต้น
- การสกัดคุณสมบัติของข้อมูล (Feature Extraction) ตัวอย่างอัลกอริทึม อาทิ PCA, LDA, และ t-SNE เป็นต้น

งานของการเรียนรู้แบบไม่มีผู้สอนแสดงดังภาพที่ 2-5

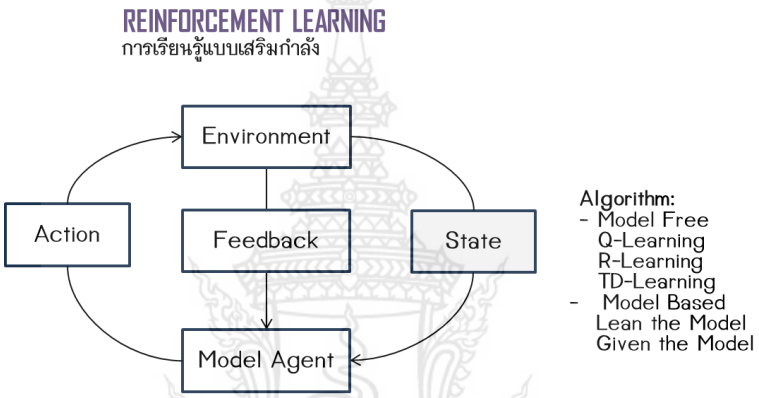


ภาพที่ 2-5 แบบจำลองงาน (Task) ในการเรียนรู้ของเครื่องแบบไม่มีผู้สอน

การเรียนรู้แบบเสริมกำลัง

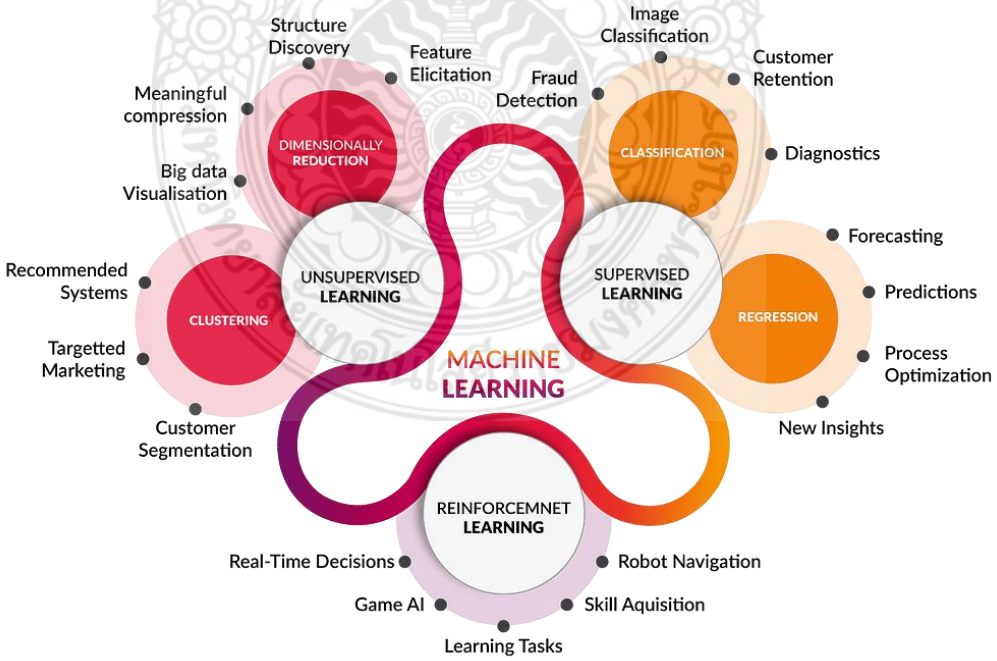
(Reinforcement Leaning; RL)

เป็นเทคนิคการเรียนรู้ของเครื่องโดยใช้หลักการให้ตัวแทน (Agent) ได้เรียนรู้และทำความเข้าใจถึงความสัมพันธ์ของสถานการณ์ (State) การกระทำ (Action) และรางวัล (Reward) ในสภาพแวดล้อม (Environment) ที่จำลองขึ้นโดยใช้หลักการ (Policy) ประเมินผลเพื่อที่จะสามารถวางแผนการตัดสินใจว่าควรกระทำสิ่งใดในสถานะการณ์ต่าง ๆ อันได้มาซึ่งรางวัลรวม (Cumulative Reward) ในระยะเวลาที่กำหนดได้มากที่สุด เทคนิคการเรียนรู้แบบปรับตัวเปรียบเสมือนการเรียนรู้จากการลองผิดลองถูกของมนุษย์ ตัวอย่างที่พบการนำเทคนิคการเรียนรู้แบบปรับตัวไปใช้ เช่น การปรับตัวเองของเอเจนท์เกมส์ (Adaptive Game Plying Agent) อาทิ AlphaGo, OpenAI Gym หรือนำมาใช้แก้ปัญหาการซื้อขายหุ้นให้ได้ค่าตอบแทนมากสุด (Stock Trading Optimization) การค้นพบสูตรการรักษาโรคแบบใหม่ (Drug Discovery) หรือการใช้ในการขับขี่รถยนต์แบบอัตโนมัติ (Self-Driving Car) เป็นต้น (Mark E. Fenner, 2020)



ภาพที่ 2-6 แบบจำลองหลักการทำงานของการเรียนรู้แบบปรับตัว

แผนภาพรวมแสดงรูปแบบการเรียนรู้และงานของการเรียนรู้ของเครื่องแสดงดังภาพที่ 2-7



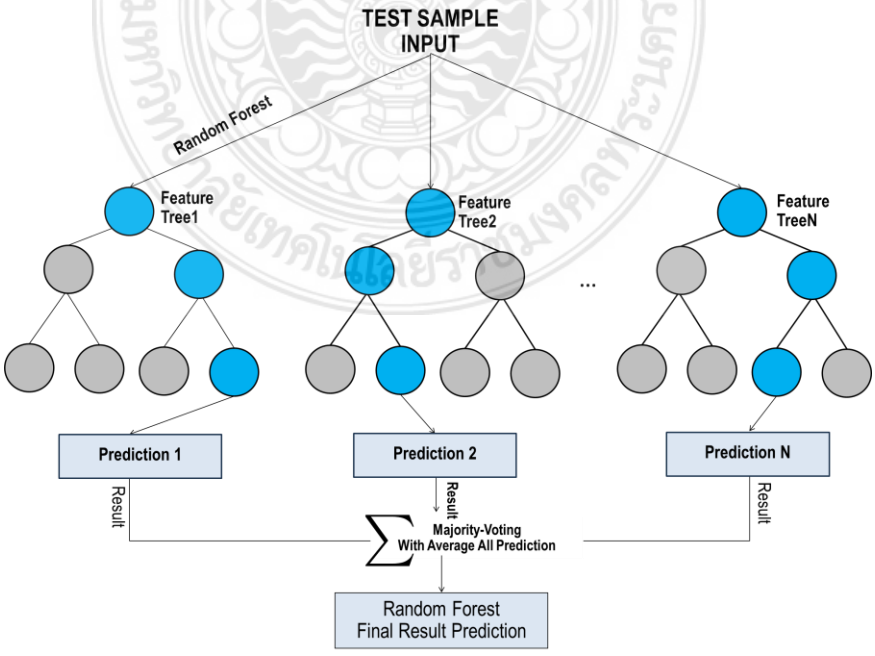
ภาพที่ 2-7 แบบจำลองภาพรวมรูปแบบการเรียนรู้ของเครื่อง

From: Coding Deep Learning for Beginners – Types of Machine Learning, KAMIL KRZYK. (2024). Experfy. (https://cdn-images-1.medium.com/max/1000/1*8wU0hfUY3UK_D8Y7tblyFQ.png)

สำหรับการดำเนินงานวิจัยเรื่องการทำนายจำนวนผู้สมัครเข้าศึกษาในหลักสูตรปริญญาตรีด้านคอมพิวเตอร์ในประเทศไทยโดยใช้ปัญญาประดิษฐ์ (The Prediction of the Amount Applicants of Admission in Undergraduate Computer Programs in Thailand using Artificial Intelligence) ดำเนินการวิจัยภายใต้การบูรณาการศาสตร์ด้านปัญญาประดิษฐ์ (AI) ด้วยการใช้การเรียนรู้ของเครื่อง (ML) บนพื้นฐานการเรียนรู้แบบมีผู้สอน (Supervised Learning) ในงาน (Task) งานทำนายข้อมูลจำนวนจริง หรือข้อมูลค่าต่อเนื่อง (Regression) เพื่อทำนายจำนวนผู้สมัครเข้าศึกษาในหลักสูตรปริญญาตรีด้านคอมพิวเตอร์ในระบบการคัดเลือกกลางบุคคลเข้าศึกษาในสถาบันอุดมศึกษาของไทย หรือระบบ TCAS โดยดำเนินการศึกษา พัฒนา และเปรียบเทียบประสิทธิภาพผลการทำนายข้อมูล จำนวน 5 อัลกอริทึม ประกอบด้วย (1) แบบจำลองเพื่อนบ้านใกล้สุด (K- Nearest Neighbors Model; KNN) (2) แบบจำลองซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression Model; SVR) (3) แบบจำลองสุ่มป่าไม้ (Random Forest Model; RF) (4) โครงข่ายประสาทเทียมเพอร์เซฟตรอนแบบหลายชั้น (Multi Leyer Perceptron Neural Network; MLP) และ (5) โครงข่ายประสาทเทียมแบบความจำขนาดสั้นระยะยาว (Long Short-Term Memory; LSTM) โดยหลักการทฤษฎีของอัลกอริทึมต่าง ๆ มีรายละเอียดพอสังเขปดังจะกล่าวต่อไป

2.3 แบบจำลองสุ่มป่าไม้ (Random Forest Model; RF)

เป็นเทคนิควิธีการเรียนรู้ของเครื่อง (Machine Learning) แบบมีผู้สอน (Supervised Learning; SL) สำหรับงานการทำนายหรือพยากรณ์ข้อมูลตัวเลข (Regression) มีชื่อเรียกย่อว่าเทคนิควิธี RFR นำเสนอครั้งแรกในปี พ.ศ. 2538 โดยทิม คาม (Tim Kam) ต่อมาวิธีนี้ถูกต่อยอดโดยลีโอ ไบรแมน (Leo Breiman) หลักการทำงานของเทคนิควิธี RF มีพื้นฐานจากเทคนิคต้นไม้ตัดสินใจ (Decision Tree) ด้วยการสร้างแบบจำลองต้นไม้ตัดสินใจหลาย ๆ แบบจำลองไม่ซ้ำกันโดยวิธีการสุ่มตัวแปรแล้วนำผลที่ได้จากแต่ละแบบจำลองมารวมกันแล้วทำการโหวตด้วยการนับจำนวนผลลัพธ์ที่มีจำนวนที่ซ้ำกันมากที่สุด หรืออาจใช้ค่าเฉลี่ย (Mean) จากผลลัพธ์ (Vivian Siahaan, Rismon H. Sianipar, 2023 : 83-87) แบบจำลองภาพรวมกระบวนการทำงานของอัลกอริทึมทำนายสุ่มป่าไม้แสดงดังภาพที่ 2-8



ภาพที่ 2-8 แบบจำลองภาพรวมกระบวนการทำงานของเทคนิควิธี RF

สามารถอธิบายขั้นตอนการทำงานของแบบจำลอง RF พอสังเขปดังนี้

1) สุ่มเลือกตัวอย่างข้อมูล (Bootstrap Sampling)
เป็นการสุ่มเลือกตัวอย่างข้อมูลย่อย (Data Subsets) จากชุดข้อมูลทั้งหมด โดยกระบวนการนี้เรียกว่า “Bootstrapping” หรือ “Bugging” ซึ่งช่วยในการลดความผันผวนในข้อมูล

2) สร้างต้นไม้ตัดสินใจ (Decision Tree)
เป็นกระบวนการที่นำข้อมูลย่อยที่สุ่มเลือกมาสร้างต้นไม้ด้วยขั้นตอนวิธีต้นไม้ตัดสินใจ (Decision Tree) ในการสร้างต้นไม้ตัดสินใจแต่ละต้น อัลกอริทึมจะสุ่มเลือกชุดตัวแปรย่อย (Subset of Features) ที่จะใช้ในการตัดสินใจที่แต่ละโหนด (Node) ของต้นไม้ โดยมีหลักการเลือกตัวแปรที่ใช้เป็นเงื่อนไขการตัดสินใจหรือที่เรียกว่า Node จะใช้ตัวแปรทุกตัวในการประมวลผล และไม่ต้องดำเนินการหาความสำคัญของตัวแปรก่อนแต่จะประมวลผลจากการสุ่มหาค่าตัวแปรที่สามารถจำแนกกลุ่มของข้อมูล วิธีการนี้เรียกว่า Minimal Depth เมื่อดำเนินการสุ่มแล้วจะใส่กลับเข้าไปเพื่อเรียนรู้ อันเป็นการเปิดโอกาสการถูกเลือกอีกครั้ง ซึ่งวิธีการนี้สามารถนำไปใช้ได้กับปัญหา และข้อมูลหลายประเภททั้งข้อมูลค่าไม่ต่อเนื่อง (Discrete Values) และข้อมูลค่าต่อเนื่อง (Continuous Value) (ปริญญา, 2562) จากนั้นดำเนินการทำนายค่าของตัวแปรตามจากข้อมูลนำเข้า โดยค่าทำนายของต้นไม้แต่ละต้นอาจมีค่าแตกต่างกันไปเนื่องจากการสุ่มเลือกตัวอย่างข้อมูลและตัวแปรที่ใช้ในการตัดสินใจ

3) รวมผลลัพธ์ (Aggregation Result)
เป็นกระบวนการโหวตผลลัพธ์ค่าทำนายของต้นไม้แต่ละต้น โดยค่าทำนายผลลัพธ์ที่ได้รับการโหวตมากที่สุด (Majority-Voting) จนสกัดได้ผลลัพธ์การทำนายสุดท้าย (Final Result) สำหรับข้อดีของขั้นตอนวิธีสุ่มป่าไม้ คือ ผลการจำแนกมีความแม่นยำที่สูง นอกจากนั้นยังสามารถช่วยแก้ปัญหา Overfitting ของการเรียนรู้ในกระบวนการสร้างแบบจำลองให้น้อยลง

ตัวอย่างการพัฒนาชุดคำสั่งภาษาไพธอนร่วมกับไลบรารี Scikit Learn สำหรับการทำนายข้อมูลด้วยหลักการถดถอยด้วยแบบจำลองสุ่มป่าไม้ (RF) มีรายละเอียดดังนี้ (Scikit Learn, 2024)

ชุดคำสั่งเรียกใช้ไลบรารี Scikit Learn

```
from sklearn.ensemble import RandomForestRegressor
```

ชุดคำสั่งสร้างแบบจำลองทำนายข้อมูลด้วยแบบจำลอง RF

```
RFmodel = RandomForestRegressor(random_state=1)
```

เมื่อกำหนดข้อมูลสำหรับการเรียนรู้ X และ y จากชุดข้อมูล **diabetes** ด้วยชุดคำสั่ง

```
from sklearn.datasets import load_diabetes
X, y = load_diabetes(return_X_y=True)
```

ดำเนินการเรียนรู้ข้อมูลจากชุดข้อมูลฝึกสอนด้วยชุดคำสั่ง

```
RFmodel.fit(X, y)
```

ดำเนินการตรวจสอบค่าสัมประสิทธิ์การทำนาย (R^2) เขียนชุดคำสั่งได้ดังนี้

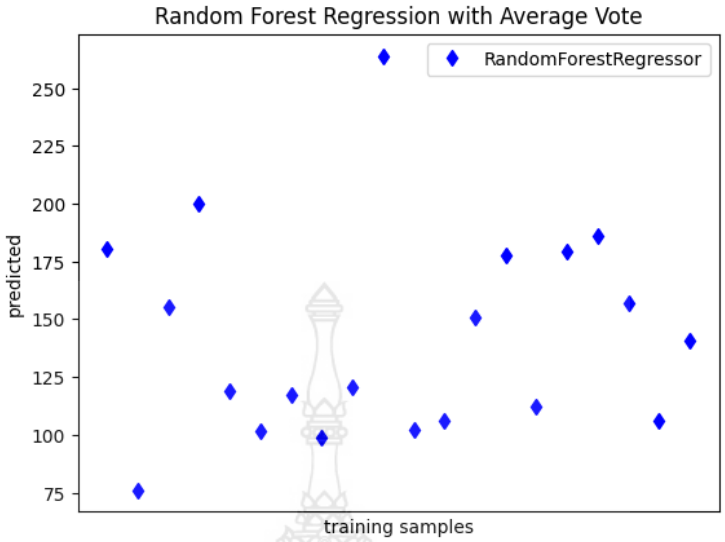
```
print('R-Squared of RF = %.2f'%RFmodel.score(X, y))
```

ซึ่งจะได้ค่าผลลัพธ์ดังนี้ *R-Squared of RF = 0.92*

ดำเนินการทดสอบการทำนายข้อมูล (Prediction) เขียนชุดคำสั่งได้ดังนี้

```
xt = X[:20]
pred = RFmodel.predict(xt)
```

ภาพแผนภูมิเส้นเปรียบเทียบการทำนายข้อมูลด้วยแบบจำลองสุ่มป่าไม้ รายละเอียดดังภาพที่ 2-9



ภาพที่ 2-9 แผนภูมิผลลัพธ์การทำนายข้อมูลด้วยแบบจำลอง (RF)

From : Plot individual and voting regression predictions. (2024). Skcikit-Learn. (https://scikit-learn.org/stable/auto_examples/ensemble/plot_voting_regressor.html#plot-individual-and-voting-regression-predictions)

2.4 แบบจำลองเพื่อนบ้านใกล้สุด (K-Nearest Neighbors Model; KNN Model)

แบบจำลองเพื่อนบ้านใกล้ที่สุด หรือเรียกสั้น ๆ แบบจำลอง KNN เป็นอัลกอริทึมการเรียนรู้ของเครื่อง (Machine Learning; ML) แบบมีผู้สอน (Supervised Learning) สามารถทำนายข้อมูลได้ทั้งจำแนกกลุ่มประเภทแคตตากอรี (Classification) และทำนายข้อมูลข้อมูลตัวเลข (Regression) ผลลัพธ์ที่ได้จากการเรียนรู้ของเครื่องกับข้อมูลชุดฝึกสอน (Training Dataset) คือ แบบจำลองทางคณิตศาสตร์สำหรับทำนายข้อมูลใหม่ (Unknown Data) หลักการทำงานของ KNN ในงานทำนายข้อมูลตัวเลขใช้การเฉลี่ยค่าผลลัพธ์ของตัวอย่างที่ใกล้เคียงที่สุด K ตัวอย่างใกล้เคียงกันในพื้นที่ข้างเคียง (Neighborhood) กับข้อมูลใหม่ (Current Data) มากที่สุดจากข้อมูลทั้งหมดในชุดข้อมูลฝึกสอน (Training Dataset) ในการทำนายค่าผลลัพธ์ใหม่เหมาะสมกับข้อมูลที่ไม่เป็นเชิงเส้น (Non-Linear) และง่ายต่อการเข้าใจและใช้งาน (Vivian Siahaan, Rismon H. Sianipar, 2023 : 92-95) รายละเอียดขั้นตอนการทำงานสรุปพอสังเขปดังนี้

1) เลือกค่า K โดย K คือ จำนวนของตัวอย่างข้อมูลที่ใกล้เคียงกับข้อมูลใหม่ที่สุดที่จะใช้ในการทำนาย
2) คำนวณค่าความใกล้เคียงกันของข้อมูลโดยใช้วิธีการหาระยะห่างระหว่าง (Distance) ข้อมูลแต่ละจุดด้วยเมตริกการวัดระยะทาง อาทิ Euclidean Distance, Cosine Distance, Manhattan Distance และ Correlation Distance เป็นต้น นิยามดังสมการที่ (2-1) ถึงสมการ (2-4) ตามลำดับ

$$d_1(\mathbf{p},\mathbf{q}) = \sqrt{\sum_{i=1}^{n}(\mathbf{q}_i - \mathbf{p}_i)^2} \tag{2-1}$$

$$d_2(\mathbf{p},\mathbf{q}) = 1 - \frac{\sum_{i=1}^{n}\mathbf{p}_i * \mathbf{q}_i}{\sqrt{\sum_{i=1}^{n}(\mathbf{p}_i)^2}\sqrt{\sum_{i=1}^{n}(\mathbf{q}_i)^2}} \tag{2-2}$$

$$d_3(\mathbf{p},\mathbf{q}) = \sum_{i=1}^{n}\left|\mathbf{p}_i - \mathbf{q}_i\right| \tag{2-3}$$

$$d_4(\mathbf{p},\mathbf{q}) = 1 - \frac{\mathrm{cov}(\mathbf{p},\mathbf{q})}{std(\mathbf{p}) * std(\mathbf{q})} \quad (2\text{-}4)$$

โดยที่ $d_1(\mathbf{p},\mathbf{q})$ คือ ระยะทางระหว่างข้อมูล $\mathbf{p}$ ไปยังจุด $\mathbf{q}$ ด้วยมาตรวัด Euclidean
$d_2(\mathbf{p},\mathbf{q})$ คือ ระยะทางระหว่างข้อมูล $\mathbf{p}$ ไปยังจุด $\mathbf{q}$ ด้วยมาตรวัด Cosine
$d_3(\mathbf{p},\mathbf{q})$ คือ ระยะทางระหว่างข้อมูล $\mathbf{p}$ ไปยังจุด $\mathbf{q}$ ด้วยมาตรวัด Correlation
$d_4(\mathbf{p},\mathbf{q})$ คือ ระยะทางระหว่างข้อมูล $\mathbf{p}$ ไปยังจุด $\mathbf{q}$ ด้วยมาตรวัด Manhattan
$\mathbf{p}$ คือ คือจุดใด ๆ ของข้อมูล
$\mathbf{q}$ คือ คือจุดใด ๆ ของข้อมูล
n คือ จำนวนมิติของข้อมูล

3) เลือกตัวอย่างที่ใกล้ที่สุดจำนวน K ตัวอย่างที่มีระยะห่างน้อยที่สุด
4) ทำนายค่าผลลัพธ์ด้วยการคำนวณค่าผลลัพธ์จากค่าเฉลี่ยของค่าผลลัพธ์ของ K ตัวอย่างที่ใกล้ที่สุด นิยามดังสมการที่ (2-5) โดยค่าเฉลี่ยที่คำนวณได้จะถูกนำไปใช้ในการทำนายสำหรับข้อมูลใหม่

$$\hat{y} = \frac{1}{K}\sum_{i=1}^{K} y_i \quad (2\text{-}5)$$

โดยที่ $\hat{y}$ คือ ค่าทำนายสำหรับข้อมูลใหม่
y_i คือ ผลลัพธ์ของ K ตัวอย่างที่ใกล้ที่สุดตัวที่ i
K คือ จำนวนเพื่อนบ้านที่ใกล้ที่สุด

ปัจจุบันแบบจำลอง KNN ถูกนำไปใช้ในการทำนายข้อมูลราคาสินค้า ผลิตภัณฑ์ และบริการ รวมทั้งนำไปใช้ในการทำนายปริมาณข้อมูลทางภูมิศาสตร์ อาทิ ปริมาณน้ำฝน ปริมาณผลผลิตทางการเกษตร เป็นต้น โดยข้อดีของแบบจำลอง KNN คือ มีความง่าย ไม่ซับซ้อน มีความยืดหยุ่นกับข้อมูลที่มีมิติสูงได้ดี แต่ก็พบข้อด้อย อาทิ มีความช้าในการประมวลผลเมื่อมีจำนวนขนาดใหญ่ และหากข้อมูลมี outliers ก็จะส่งผลต่อประสิทธิภาพของการทำนาย และประเด็นที่สำคัญสุดของแบบจำลอง KNN คือ การกำหนดค่า K ต้องมีความเหมาะสม เพราะเนื่องจากจำนวน K มีผลต่อการคำนวณหาค่าผลลัพธ์การทำนายข้อมูลใหม่

ตัวอย่างการพัฒนาชุดคำสั่งภาษาไพธอนร่วมกับไลบรารี Scikit Learn สำหรับการทำนายข้อมูลด้วยหลักการถดถอยมีรายละเอียดดังนี้ (Scikit Learn, 2024)

ชุดคำสั่งเรียกใช้ไลบรารี Scikit Learn

```
from sklearn import neighbors
```

ชุดคำสั่งสร้างแบบจำลอง KNN

```
KNNmodel =  neighbors.KNeighborsRegressor(
            n_neighbors=5,
            weights='distance',
            algorithm='auto',
            leaf_size=30, p=2)
```

เมื่อกำหนดข้อมูลสำหรับการฝึกสอนและทดสอบด้วยชุดคำสั่ง

```
from sklearn.model_selection import train_test_split
```

```
X_train, X_test, y_train, y_test = train_test_split(X, df.tip, random_state=42)
```

ดำเนินการเรียนรู้ข้อมูลจากชุดข้อมูลฝึกสอนด้วยชุดคำสั่ง

```
model.fit(X_train, y_train)
```

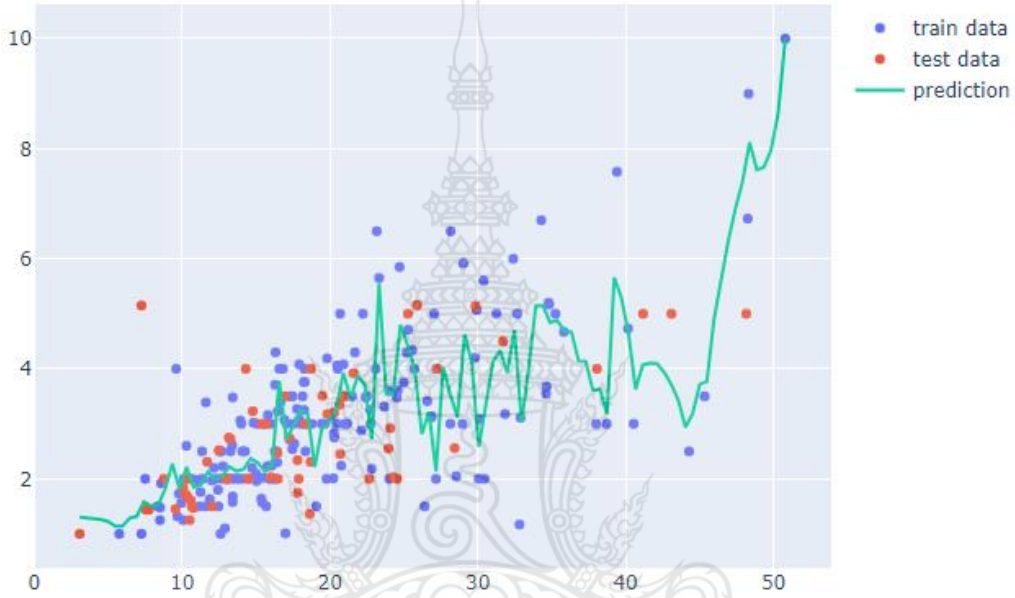
ดำเนินการตรวจสอบค่าสัมประสิทธิ์การทำนาย (R^2) เขียนชุดคำสั่งได้ดังนี้

```
print('R-Squared of KNN model   = %.2f'%KNNmodel.score(X_test, y_test))
```

ซึ่งจะได้ค่าผลลัพธ์ดังนี้

R-Squared of KNN model = 0.99

ภาพแผนภูมิจุดภาพทำนายข้อมูลด้วยแบบจำลอง KNN รายละเอียดดังภาพที่ 2-10

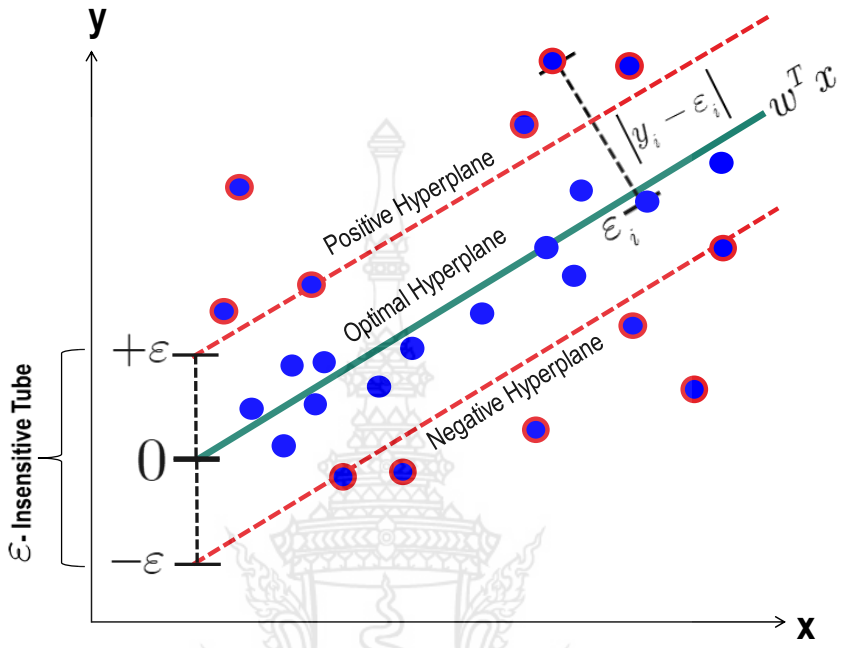


ภาพที่ 2-10 แผนภูมิผลลัพธ์การทำนายข้อมูลด้วยแบบจำลอง KNN

2.5 แบบจำลองซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression Model; SVR)

เป็นแบบจำลองเรียนรู้ของเครื่อง (Machine Learning) แบบมีผู้สอน (Supervised Learning) สามารถทำนายข้อมูลได้ทั้งจำแนกกลุ่มประเภทแคตตากอรี (Classification) และทำนายข้อมูลตัวเลข (Regression) ผลลัพธ์ที่ได้จากการเรียนรู้ของเครื่องกับข้อมูลชุดฝึกสอน (Training Dataset) คือ แบบจำลองทางคณิตศาสตร์สำหรับทำนายข้อมูลใหม่ (Unknown Data) โดยในงานการจำแนกกลุ่มเรียกเทคนิควิธีนี้ว่าซัพพอร์ทเวกเตอร์แมชชีน (Support Vector Machine; SVM) แต่สำหรับงานทำนายข้อมูลตัวเลขเรียกเทคนิควิธีซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression; SVR) โดยทั่วไปหลักการทำงานของ SVM และ SVR มีความคล้ายกันแต่ต่างกันที่เป้าหมาย กล่าวคือ เป้าหมายของ SVM คือ การทำนายข้อมูลด้วยการสร้างสมการเพื่อให้ได้เส้นแนวที่เหมาะสม (Optimal Hyperplane) สำหรับแยกข้อมูลออกเป็นกลุ่ม ๆ (Classification) ออกจากกัน ในขณะที่ SVR มีเป้าหมายทำนายข้อมูลจำนวนจริงด้วยการสร้างสมการเส้นแนวที่เหมาะสมเพื่อให้ได้ฟังก์ชันอธิบายความสัมพันธ์ระหว่างตัวแปรนำเข้า (Input Variable) กับตัวแปรผลลัพธ์ (Output Variable) ที่มีความแม่นยำและมีความคลาดเคลื่อนน้อยสุด เพื่อทำนายข้อมูลเชิงตัวเลข (Regression) ซึ่งฟังก์ชันที่หาได้จากการเรียนรู้จากข้อมูลฝึกสอนจะถูกนำไปใช้ในการทำนายค่าข้อมูลใหม่ (Unknown Data) ต่อไป โดยฟังก์ชันหรือแบบจำลอง SVR ที่ได้อาจอยู่ในรูปแบบฟังก์ชันคณิตศาสตร์แบบเชิงเส้น (Linear Function) และแบบเส้นโค้ง (Polynomial Function) หลักการทำงานในภาพรวมของเทคนิควิธี SVR จะดำเนินการวิเคราะห์ข้อมูลเพื่อหาเส้นทำนายข้อมูลที่เรียกว่า Optimal Hyperplane ดำเนินการคำนวณหาเส้นขอบสำหรับการตัดสินใจ (Decision Boundary) ทั้งในด้านบวก (Positive Hyperplane)

และด้านลบ (Negative Hyperplane) คำนวณค่าคลาดเคลื่อน (Error; ε) ต่ำสุดด้วยฟังก์ชันสูญเสียแบบ Epsilon Insensitive (ε-insensitive Loss Function) ในรูปแบบค่าสัมบูรณ์ (Absolute Term) ดำเนินการพิจารณาข้อมูลใดที่หลุดจากพื้นที่ที่สร้างขึ้นจะถูกกำหนดค่าเป็นค่าคลาดเคลื่อนของแบบจำลอง (Vivian Siahaan, Rismon H. Sianipar, 2023 : 118-121) นิยามการทำงานดังภาพที่ 2-11



ภาพที่ 2-11 แบบจำลองหลักการทำงานของเทคนิควิธี SVR

ขั้นตอนการทำงานของ SVR อธิบายพอสังเขปดังนี้

1) จัดเตรียมข้อมูลสำหรับการฝึกสอน (Training Dataset) และชุดข้อมูลสำหรับการทดสอบประสิทธิภาพของแบบจำลอง (Testing Dataset) ซึ่งประกอบด้วยเวกเตอร์สำหรับการฝึกสอนเครื่องจักร $\{(x_1, y_1), \ldots, (x_n, y_n)\} \subset X \times \mathbb{R}$ โดยที่ X คือ เวกเตอร์อินพุต (Input Vector) ในมิติขนาด $n(X \in \mathbb{R}^n)$ และ y คือตัวแปรผลลัพธ์ (Output Variable) เมื่อ $(y \in \mathbb{R})$
2) เลือกเคอร์เนลฟังก์ชัน (Kernel Function) ที่เหมาะสม ที่นิยมใช้มีดังนี้
   - ลีเนียร์เคอร์เนล (Linear Kernel)
     กำหนดเคอร์เนล $K(\mathrm{x}_i, \mathrm{x}) = \mathrm{x}_i \cdot \mathrm{x}_j$ นิยามดังสมการที่ (2-6)

$$f(x) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) \mathrm{x}_i \cdot \mathrm{x}_j + b \qquad (2\text{-}6)$$

   - โพลีโนเมียลเคอร์เนล (Polynomial Kernel)
     กำหนดเคอร์เนล $K(\mathrm{x}_i, \mathrm{x}) = (1 + \mathrm{x}_i \cdot \mathrm{x}_j)^d$ นิยามดังสมการที่ (2-7)

$$f(x) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*)(1 + \mathrm{x}_i \cdot \mathrm{x}_j)^d + b \qquad (2\text{-}7)$$

   - เกาส์เซียนเคอร์เนล (Gaussian Kernel) หรือ Radial-basis Function (RBF)
     กำหนดเคอร์เนล $K(\mathrm{x}_i, \mathrm{x}) = \exp\left(-\gamma \left\| \mathrm{x}_i - \mathrm{x}_j \right\|^2\right)$ นิยมดังสมการที่ (2-8)

$$f(x)=\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)\exp\left(-\gamma\left\|\mathrm{x}_i-\mathrm{x}_j\right\|^2\right)+b \tag{2-8}$$

3) หาสมการไฮเปอร์เพลนเพื่อสร้างเส้นแนวทำนายข้อมูลที่เหมาะสม (Optimal Hyperplane Equation) อธิบายความสัมพันธ์ระหว่างเวกเตอร์ข้อมูลนำเข้า (Input Vector) กับเวกเตอร์ข้อมูลผลลัพธ์ (Output Vector) ด้วยการคำนวณหาค่าสัมประสิทธิ์สำหรับแบบจำลองทำนายข้อมูลที่มีความคลาดเคลื่อนน้อยสุดภายในขอบเขตของ ε นิยามดังสมการที่ (2-9)

$$f(\mathrm{x})=\mathbf{w}\cdot\mathbf{x}+b \quad \begin{cases}\mathbf{w}\in X\\ b\in\mathbb{R}\end{cases} \tag{2-9}$$

โดยที่ $\mathbf{w}$ คือ เวกเตอร์ค่าน้ำหนักที่ทำการคูณอยู่กับตัวแปร $\mathbf{x}$ ใด ๆ
b คือ ค่าไบแอสซึ่งเป็นค่าคงที่

เมื่อ $\mathbf{w}$ และ b คือ ความชันของออฟเซต (Offset) ของเส้นแนวที่เหมาะสมสำหรับการถดถอย ที่มีเป้าหมายมีค่าคลาดเคลื่อนของคำตอบผลลัพธ์น้อยสุดภายในขอบเขตของ ε หรือท่อเอปซีลอน (Epsilon Tube) ซึ่งสามารถเขียนสมการหาค่าเป้าหมาย (Optimize) ได้จากสมการที่ (2-10)

$$\min_{\mathbf{w},b,\zeta_i,\zeta_i^*}=\frac{1}{2}\left\|\mathbf{w}\right\|^2+C\sum\nolimits_{i=1}^{n}\zeta_i+\zeta_i^* \tag{2-10}$$

โดยที่ C คือ จำนวนเต็มที่เป็นค่าคงที่ หรือค่าการสูญเสีย (Cost) โดยใช้ฟังก์ชันสูญเสีย (Loss Function) ที่มีหลากหลายรูปแบบตัวอย่างฟังก์ชันสูญเสียที่นิยมใช้ได้แก่ Insensitive Loss Function
ζ คือ ค่าระยะห่างระหว่าง ε กับข้อมูลตัวอย่างภายใต้เงื่อนไข

$$\begin{cases}y_i-(\left\langle\mathbf{w}\cdot\mathrm{x}_i\right\rangle+b)\le\varepsilon+\zeta_i\\ (\left\langle\mathbf{w}\cdot\mathrm{x}_i\right\rangle+b)-y_i\le\varepsilon+\zeta_i^*\\ \zeta_i\zeta_i^*\ge 0\end{cases}$$

จากสมการไฮเปอร์เพลนที่เหมาะสมที่ค่า $\mathbf{w}$ น้อยสุดดังสมการที่ (2-10) ข้างต้น สามารถหาค่า $\mathbf{w}$ ที่มีค่าน้อยสุดได้โดยใช้สมการลากรานจ์ (Lagrange Function) โดยการเพิ่มตัวแปร η,α ในสมการซึ่งเรียกว่าตัวคูณลากรานจ์ (Lagrange Multipliers) นิยามดังสมการที่ (2-11)

$$\begin{aligned}L=&\frac{1}{2}\left\|\mathbf{w}\right\|^2+C\sum\nolimits_{i=1}^{n}\ \zeta_i+\zeta_i^*\ -\sum\nolimits_{i=1}^{n}\eta_i\zeta_i+\eta_i^*\zeta_i^*\\ &-\sum\nolimits_{i=1}^{n}\alpha_i\ \varepsilon+\zeta_i-y_i+\left\langle\mathbf{w}\cdot\mathrm{x}_i\right\rangle+b\ -\sum\nolimits_{i=1}^{n}\alpha_i^*\ \varepsilon+\zeta_i^*+y_i-\left\langle\mathbf{w}\cdot x_i\right\rangle-b\end{aligned} \tag{2-11}$$

โดยที่ $\eta_i,\eta_i^*,\alpha_i,\alpha_i^*$ คือ ค่าตัวคูณลากรานจ์มีค่ามากกว่าศูนย์ (0) $\eta_i^{(*)},\alpha_i^{(*)}\ge 0$ เมื่อค่า $\eta_i^{(*)}$ เป็นค่าจาก η_i,η_i^* และ $\alpha_i^{(*)}$ เป็นค่าจาก α_i,α_i^* และเมื่อนำสมการลากรานจ์มาหาค่าอนุพันธ์เทียบกับตัวแปรที่ต้องการมีค่าเท่ากับศูนย์ (0) คำนวณได้ดังสมการที่ (2-12) ถึง (2-14) ตามลำดับ

$$\partial_b L=\sum\nolimits_{i=1}^{n}(\alpha_i^*-\alpha_i) \tag{2-12}$$

$$\partial_{\mathbf{w}} L = \mathbf{w} - \sum_{i=1}^{n} (\alpha_i - \alpha_i^*)\mathrm{x}_i \tag{2-13}$$

$$\partial_{\xi_i^{(*)}} L = C - \alpha_i^* - \eta_i^* \tag{2-14}$$

นำสมการที่ (2-12) แทนค่าลงในสมการที่ (2-11) จะได้ผลลัพธ์ดังสมการที่ (2-15)

$$\max \begin{cases} \frac{1}{2}\sum_{i=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_i^*)x_i \cdot y_j \\ -\varepsilon\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) + \sum_{i=1}^{n} y_i(\alpha_j - \alpha_i^*) \end{cases} \tag{2-15}$$

เมื่อ $\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0$ และ $\alpha_i, \alpha_i^* \in [0, C]$

และนำสมการที่ (2-13) สามารถแปลงได้ดังสมการที่ (2-16)

$$\mathbf{w} = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\mathrm{x}_i \tag{2-16}$$

และเมื่อแทนค่า $\mathbf{w}$ ลงใน $f(x)$ จะได้ผลลัพธ์ดังสมการที่ (2-17)

$$f(\mathrm{x}) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\mathrm{x}_i \cdot \mathbf{x} + b \tag{2-17}$$

ปัจจุบันแบบจำลอง SVR ถูกนำไปใช้ในการทำนายข้อมูลต่าง ๆ อาทิ ทำนายราคาหุ้น ทำนายยอดขายสินค้า ทำนายปริมาณน้ำฝน โดยข้อดีของแบบจำลอง SVR คือ ให้ประสิทธิภาพสูงในการทำนายข้อมูลตัวเลขหรือค่าต่อเนื่อง แม้มี Outliers ในข้อมูล สามารถจัดการกับข้อมูลที่มีมิติสูงได้ดี แปลผลลัพธ์ได้ง่าย สำหรับข้อด้อยของแบบจำลอง SVR คือ การประมวลผลช้าและต้องใช้เวลาเมื่อมีข้อมูลขนาดใหญ่ ประเด็นที่สำคัญคือการเลือกเคอร์เนลฟังก์ชันและการกำหนดค่าพารามิเตอร์ต่าง ๆ ที่เหมาะสม เพราะมีผลต่อประสิทธิภาพของการทำนาย

ตัวอย่างการพัฒนาชุดคำสั่งภาษาไพธอนร่วมกับไลบรารี Scikit Learn สำหรับการทำนายข้อมูลด้วยหลักการถดถอยมีรายละเอียดดังนี้ (Scikit Learn, 2024)

ชุดคำสั่งเรียกใช้ไลบรารี Scikit Learn

```
from sklearn.svm import SVR
```

ชุดคำสั่งสร้างแบบจำลอง SVR ด้วยฟังก์ชันแบบเชิงเส้น

```
svr_lin =     SVR(kernel="linear",
              C=100, gamma="auto")
```

ชุดคำสั่งสร้างแบบจำลอง SVR ด้วยฟังก์ชันพหุนามดีกรี 3 หรือ Cubic Function

```
svr_poly = SVR(kernel="poly",   C=100, gamma="auto",
           degree=3, epsilon=0.1,  coef0=1)
```

ชุดคำสั่งสร้างแบบจำลอง SVR ด้วยฟังก์ชันเกาส์เซียน หรือ RBF

```
svr_rbf = SVR(kernel="rbf", C=100,
               gamma=0.1, epsilon=0.1)
```

เมื่อกำหนดข้อมูลสำหรับการเรียนรู้ด้วยชุดคำสั่ง

```
import numpy as np
X = np.sort(5 * np.random.rand(40, 1), axis=0)
y = np.sin(X).ravel()
```

ดำเนินการเรียนรู้ข้อมูลจากชุดฝึกสอนด้วยชุดคำสั่ง

```
svr_lin.fit(X, y).predict(X)
svr_poly.fit(X, y).predict(X)
svr_rbf.fit(X, y).predict(X)
```

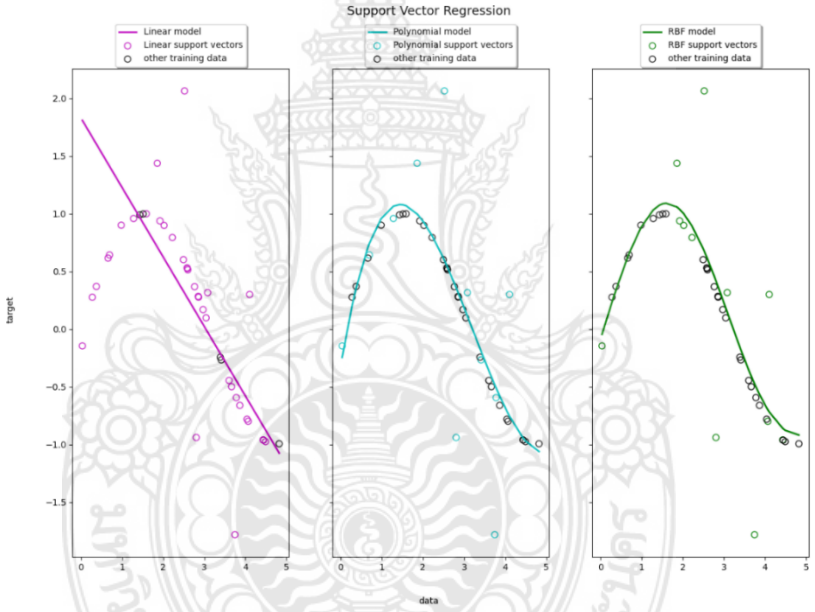
ดำเนินการตรวจสอบค่าสัมประสิทธิ์การทำนาย (R^2) เขียนชุดคำสั่งได้ดังนี้

```
print('R-Squared of SVM LR    = %.2f'%svr_lin.score(X, y))
print('R-Squared of SVM Cubic = %.2f'%svr_poly.score(X, y))
print('R-Squared of SVM RBF   = %.2f'%svr_rbf.score(X, y))
```

ซึ่งจะได้ค่าผลลัพธ์ดังนี้

```
R-Squared of SVM LR    = 0.4993
R-Squared of SVM Cubic = 0.9883
R-Squared of SVM RBF   = 0.9879
```

จากผลลัพธ์ค่า R^2 ข้างต้นพบว่าแบบจำลองแบบพหุนามกำลังสาม หรือ (Cubic Function) ให้ค่าสัมประสิทธิ์การทำนายร้อยละ 98.83 ซึ่งเป็นแบบจำลองทำนายข้อมูลที่ดีสุด ภาพแผนภูมิเส้นเปรียบเทียบการทำนายข้อมูลด้วยเทคนิคซัพพอร์ตเวกเตอร์กับการกำหนดรูปแบบของเคอร์เนลชนิดต่าง ๆ รายละเอียดดังภาพที่ 2-12

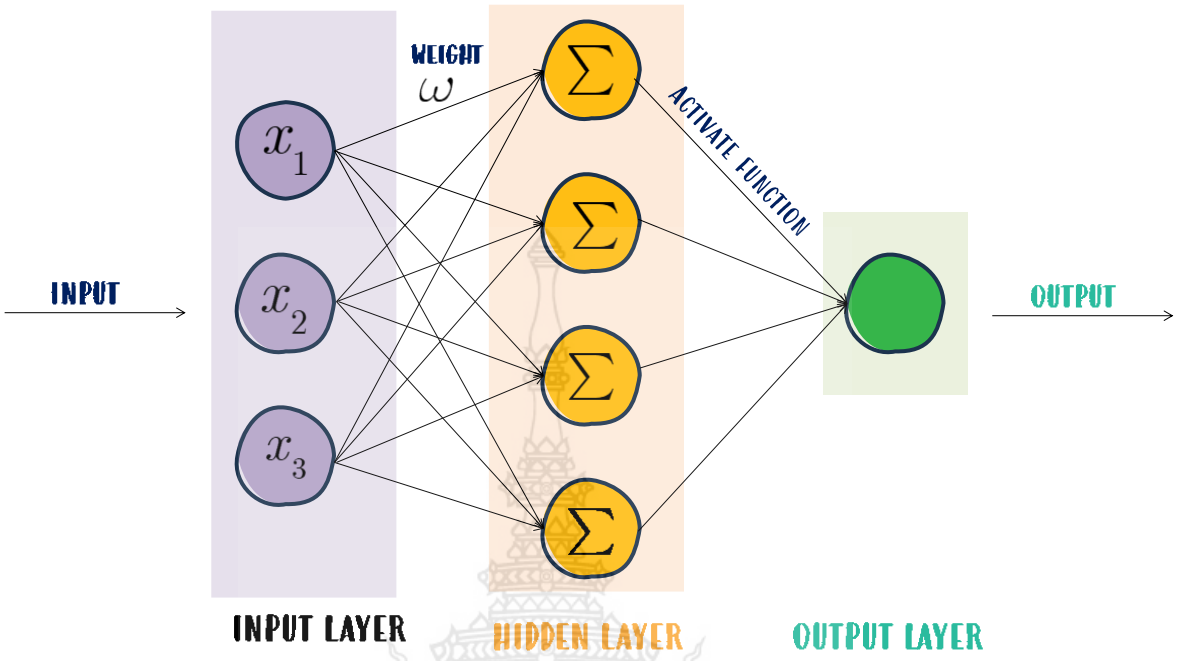


ภาพที่ 2-12 แผนภูมิเส้นเปรียบเทียบการทำนายข้อมูลด้วยแบบจำลอง SVR กับการกำหนดเคอร์เนลแบบต่าง ๆ
From: Support Vector Regression (SVR) using linear and non-linear kernels. (2024). Scikit Learn. (https://scikit-learn.org/stable/auto_examples/svm/plot_svm_regression.html#sphx-glr-auto-examples-svm-plot-svm-regression-py)

2.6 โครงข่ายประสาทเทียม (Artificial Neural Network; ANN)

โครงข่ายประสาทเทียม หรือ ANN เป็นแบบจำลองทางคณิตศาสตร์ที่มีแรงบันดาลใจมาจากโครงข่ายโหนดเซลล์ประสาทของสมองมนุษย์ (Biological Neural Network) กล่าวคือ สมองของมนุษย์จะที่มีเซลล์ประสาท หรือ นิวรอน (Neurons) เชื่อมต่อกันเป็นโครงข่ายประสาทเทียม (Neural Network) โดยนิวรอนในสมองมนุษย์มีจำนวนเฉลี่ยประมาณ 86 พันล้านเซลล์ทำหน้าที่ประมวลผลและส่งข้อมูลผ่านสัญญาณไฟฟ้าและเคมี ด้วยการส่งผ่านจุดประสานประสาท (Synapse) โดยโครงข่ายประสาทเทียมยังเป็นอัลกอริทึมที่จัดอยู่ในกลุ่มการเรียนรู้เชิงลึก (Deep Learning) ในศาสตร์ด้านปัญญาประดิษฐ์ (Artificial Intelligence; AI) ผลลัพธ์ที่ได้คือแบบจำลองคณิตศาสตร์เพื่อใช้ในงานทำนายกลุ่ม (Classification) และงานทำนายข้อมูลตัวเลข (Regression) โครงข่ายประสาทเทียมนิยมใช้กันมากในปัญหาที่มีความซับซ้อนและยากในการหาคำตอบ สถาปัตยกรรมโครงข่ายประสาทเทียม หรือ ANN

โดยทั่วไป ประกอบด้วย โหนด (Nodes) หรือ นิวรอน (Neural) ที่ถูกแบ่งออกเป็นระดับชั้น ซึ่งประกอบด้วย Input Layer, Hidden Layer และ Output Layer รายละเอียดดังภาพแบบจำลองที่ 2-13



ภาพที่ 2-13 แบบจำลองการทำงานโครงข่ายประสาทเทียมแบบชั้นเดียว

จากภาพแบบจำลองการทำงานโครงข่ายประสาทเทียมแบบชั้นเดียวข้างต้น มีโครงสร้างการทำงานประกอบด้วย Input Layer, Hidden Layer และ Output Layer รายละเอียดหน้าที่ของแต่ละเลเยอร์มีดังนี้

1) Input Layer เป็นเลเยอร์ที่ประกอบด้วยโหนดรับข้อมูลนำเข้าทั้งหมดที่ต้องการประมวลผล
2) Hidden Layers เป็นเลเยอร์ที่ประกอบด้วยโหนดที่ทำหน้าที่ประมวลผลข้อมูลโดยใช้ค่าถ่วงน้ำหนัก (Weight) และฟังก์ชันกระตุ้น (Activation Function)
3) Output Layer เป็นเลเยอร์ที่ประกอบด้วยโหนดผลลัพธ์ของการทำนาย หรือผลลัพธ์ของการประมวลผล และคำนวณค่าคลาดเคลื่อนจากการทำนายด้วยฟังก์ชันการสูญเสีย (Loss Function)

โดยมีลักษณะการทำงาน 3 ลักษณะ คือ

1) การส่งข้อมูลไปข้างหน้า (Feedforward) คือ ข้อมูลนำเข้าถูกส่งผ่านชั้นโหนดใน Hidden ไปยังชั้น Output โดยใช้น้ำหนัก (Weight) และฟังก์ชันกระตุ้น (Activate Function) ที่อยู่ในรูปแบบต่าง ๆ อาทิ Sigmoid / Logistic Activation Function, Hyperbolic Tangent (Tanh function), Rectified Linear Unit, (ReLU) function, Leaky ReLU Function, Parametric ReLU Function, Exponential Linear Units (ELUs) Function และ Softmax Function เป็นต้น
2) การถอยกลับ (Backpropagation) คือ กระบวนการปรับน้ำหนักในโครงข่ายเพื่อลดค่าความคลาดเคลื่อนระหว่างผลลัพธ์ที่ทำนายกับผลลัพธ์ที่ต้องการ (Ground Truth)
3) การฝึกสอน (Training) คือ การปรับน้ำหนักเป็นขั้นตอนการทำซ้ำเพื่อฝึกสอนจนกว่าโครงข่ายจะประสิทธิภาพในการทำนายที่ดี

ปัจจุบันมีการพัฒนาสถาปัตยกรรมโครงข่ายประสาทเทียมออกเป็นหลายรูปแบบ อาทิ โครงข่ายประสาทเทียมเพอร์เซฟตรอนแบบหลายชั้น (Multi Leyer Perceptron Neural Network; MLP) ซึ่งปัจจุบันได้รับความนิยมมาก เนื่องจาก MLP มีการทำงานแบบถอยกลับ (Backpropagation) เพื่อความซับซ้อนในการเรียนรู้ด้วยการเพิ่มชั้นของ Hidden Layer ออกเป็นหลายชั้น ทำให้สามารถเรียนรู้ข้อมูลที่มีความสัมพันธ์แบบไม่เชิงเส้น (Non-Linear) ได้ดี โครงข่ายประสาทเทียมแบบวกกลับ (Recurrent Neural Network; RNN) ซึ่งเป็นสถาปัตยกรรมโครงข่ายประสาทเทียมที่ความสามารถในการดูข้อมูลย้อนหลังได้ และโครงข่ายประสาทเทียมแบบความจำขนาดสั้นระยะยาว (LSTM Neural Networks) ที่พัฒนาขึ้นเพื่อแก้ปัญหาการดูข้อมูลก่อนหน้าของ RNN ได้ในระยะสั้น ๆ เป็นต้น

ตัวอย่างการพัฒนาชุดคำสั่งภาษาไพธอนร่วมกับไลบรารี Scikit Learn สำหรับการทำนายข้อมูลด้วยโครงข่ายประสาทเทียมแบบ MLP (Scikit Learn, 2024) มีรายละเอียดดังนี้

ชุดคำสั่งเรียกใช้ไลบรารี Scikit Learn ด้วย Multi-layer Perceptron Regressor

```
from sklearn.neural_network import MLPRegressor
```

ชุดคำสั่งสร้างแบบจำลองทำนายข้อมูลด้วยโครงข่ายประสาทเทียมแบบบ MLP

```
MLPmodel = make_pipeline(
           StandardScaler(),
           MLPRegressor(activation='relu',
           hidden_layer_sizes=(10, 100),
           alpha=0.001,   max_iter=500,
           solver='adam', random_state=2,
           early_stopping=False)
```

เมื่อกำหนดข้อมูลสำหรับการเรียนรู้ฝึกสอน

```
from sklearn import datasets
from sklearn.model_selection import train_test_split

diabetes = load_diabetes()
X = pd.DataFrame(diabetes.data, columns=diabetes.feature_names)
y = diabetes.target

X_train, X_test, y_train, y_true = train_test_split(X, y,
                                   test_size=0.3,
                                   random_state=1)
```

ดำเนินการเรียนรู้ข้อมูลจากชุดข้อมูลฝึกสอนด้วยชุดคำสั่ง

```
MPLmodel.fit(X_train, y_train)
```

ดำเนินการทดสอบการทำนายข้อมูล (Prediction) เขียนชุดคำสั่งได้ดังนี้

```
y_pred = MLPmodel.predict(X_test)
```

ดำเนินการตรวจสอบค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) เขียนชุดคำสั่งได้ดังนี้

```
from sklearn.metrics import mean_squared_error
mse = mean_squared_error(y_true, y_pred)
print('Mean Squared Error (MSE) = %.2f'%mse)
```

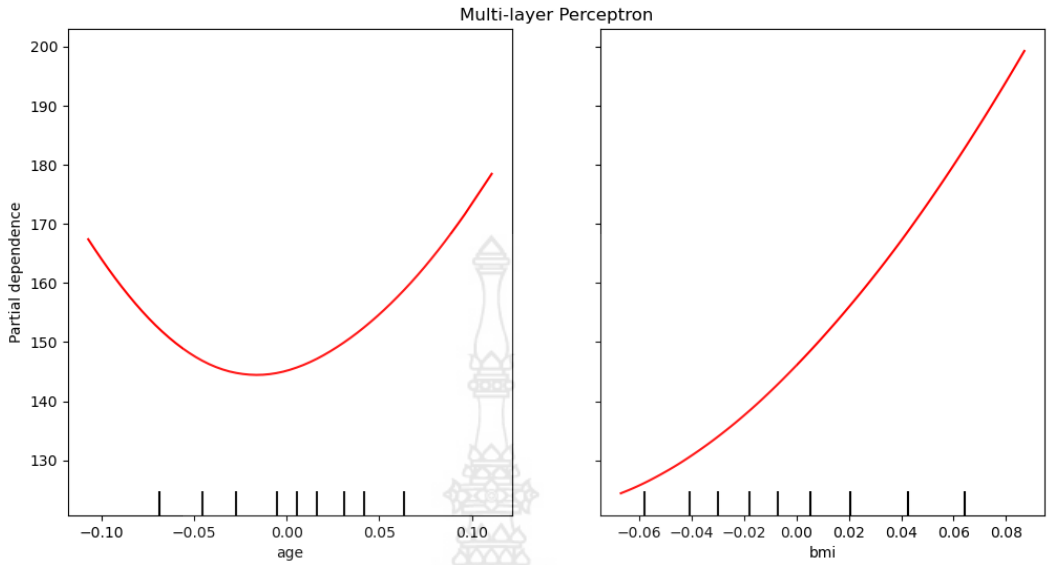
ซึ่งจะได้ค่าผลลัพธ์ดังนี้

Mean Squared Error (MSE) = 0.70

ดำเนินการทดสอบการทำนายข้อมูลกับชุดข้อมูลทดสอบเขียนชุดคำสั่งได้ดังนี้

```
y_predicted = MLPmodel.predict(X_test)
```

ภาพแผนภูมิเส้นเปรียบเทียบการทำนายข้อมูลด้วยแบบจำลอง MLP รายละเอียดดังภาพที่ 2-14



ภาพที่ 2-14 แผนภูมิเส้นผลลัพธ์การทำนายข้อมูลด้วยโครงข่ายประสาทเทียมแบบ MLP

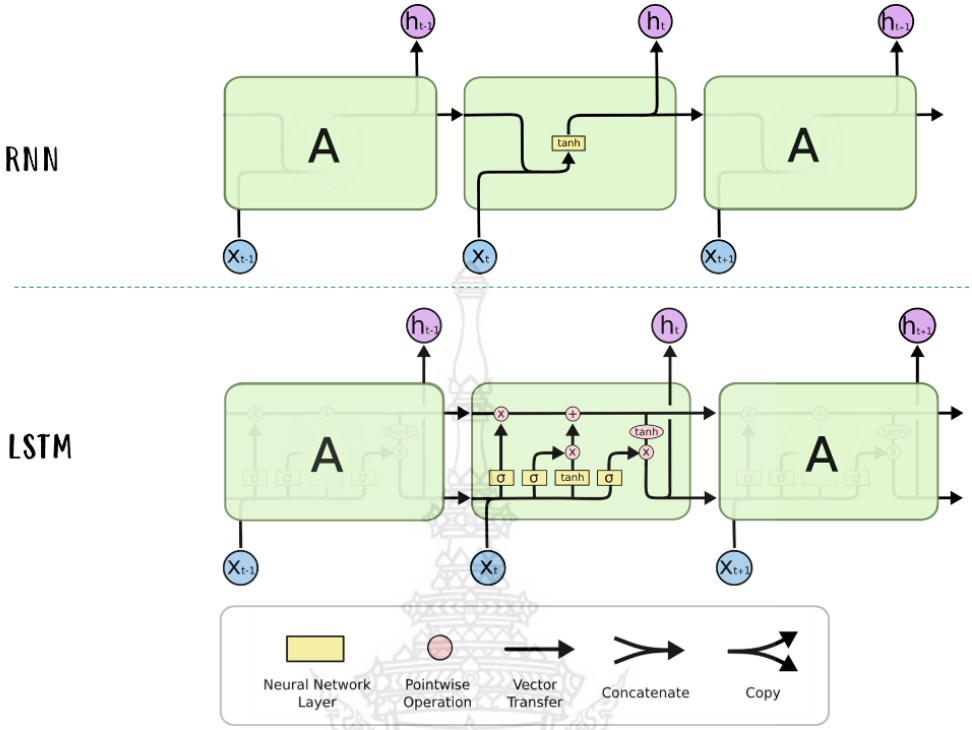
From: Advanced plotting with partial dependence. (2024). Scikit-learn. (https://scikit-learn.org/stable/auto_examples/miscellaneous/plot_partial_dependence_visualization_api.html)

2.7 โครงข่ายประสาทเทียมแบบความจำขนาดสั้นระยะยาว (LSTM Neural Networks)

โครงข่ายประสาทเทียมแบบความจำขนาดสั้นระยะยาว (Long Short-Term Memory Neural Network) หรือเรียกสั้น ๆ ว่า LSTM ถูกพัฒนาขึ้นเพื่อแก้ปัญหาการหายไปและการระเบิดของค่าความคลาดเคลื่อน (Vanishing & Exploding Gradient Problem) กล่าวคือ ในโครงข่ายประสาทเทียมแบบวกกลับ (Recurrent Neural Network) หรือ RNN มีความสามารถในการดูข้อมูลย้อนหลังได้เพียงระยะสั้นส่งผลให้เกิดปัญหาในการทำ Backpropagation นั้นจะต้องทำย้อนไปหลายขั้นตอนและหลายโหนด ดังนั้น LSTM จึงมีสถาปัตยกรรมแบบวกกลับ และเป็นหนึ่งในสถาปัตยกรรมของโครงข่ายประสาทเทียมแบบ RNN ซึ่งมีความเหมาะสมกับงานที่มีลักษณะลำดับ (Sequence) หรือข้อมูลที่ความต่อเนื่อง (Continuous Data) เช่น ข้อมูลอนุกรมเวลา (Time Series) ข้อมูลเสียง (Voice Data) ข้อมูลข้อความ (Text Data) ข้อมูลภาพและวิดิทัศน์ (Image & Video Data) โดยทั่วไปหลักการทำงานของ RNN นั้นประกอบด้วยข้อมูลที่นำเข้ามา 2 ส่วน คือ ข้อมูลนำเข้า (Input) ของโหนดนั้น ๆ กับผลลัพธ์ที่ผ่านการคำนวณจากโหนดก่อนหน้าที่ได้ทำการเก็บสถานะ (State) ไว้ โดยข้อมูลนำเข้าทั้ง 2 ชุดที่เข้ามาในโหนดจะถูกนำมารวมเข้าด้วยกัน ก่อนจะถูกแยกผลลัพธ์ออกเป็น 2 ส่วน คือ ผลลัพธ์ที่ได้จากโหนดนั้น ๆ และผลลัพธ์ที่ถูกนำไปใช้เป็นข้อมูลนำเข้า (Input) ของโหนดถัดไป ทั้งนี้ก็เพื่อให้สามารถนำข้อมูลก่อนหน้าหรือข้อมูลในอดีตมาใช้ในการทำนายสิ่งที่อาจจะเกิดขึ้นในอนาคตได้ แต่ก็พบปัญหาเรื่องความสามารถในการดูข้อมูลย้อนหลังได้แค่เพียงระยะสั้น ๆ เท่านั้น

สำหรับหลักการทำงานของโครงข่ายประสาทเทียมแบบ LSTM มีการทำงานที่คล้ายกับโครงข่ายประสาทเทียมแบบ RNN แต่จะมีเซลล์ (Cell) ความจำขนาดสั้นระยะยาวที่ประกอบด้วยฟังก์ชันพิเศษทำหน้าที่เสมือนเกตหรือประตู (Gate) คอยควบคุมข้อมูลที่จะเข้าไปในแต่ละโหนดประกอบด้วย 3 ชั้นคือ เกตลืมข้อมูล (Forget Gate Layer) เกตข้อมูลขาเข้า (Input Gate Layer) และ เกตข้อมูลขาออก (Output Gate Layer) ซึ่งจะทำให้การส่งผ่านและรับข้อมูลจากโหนดก่อนหน้าได้ ทำให้เกิดการเรียนรู้และส่งค่าความคลาดเคลื่อนไปยังโหนดก่อนหน้า

(Backpropagation through time) ได้ แบบจำลองเปรียบเทียบสถาปัตยกรรมการทำงานของ RNN และ LSTM (Oinkina, 2015) แสดงดังภาพที่ 2-15



ภาพที่ 2-15 แบบจำลองการเปรียบเทียบสถาปัตยกรรมโครงข่ายประสาทเทียมแบบ RNN และ LSTM
From: Understanding LSTM Networks. Oinkina. (2015). Colah's blog. (https://colah.github.io/ posts/2015-08-Understanding-LSTMs/)

จากภาพแบบจำลองสถาปัตยกรรมโครงข่ายประสาทเทียมแบบ LSTM ข้างต้น แสดงให้เห็นถึงฟังก์ชันพิเศษที่นำหน้าที่เสมือนเกต หรือประตูที่คอยควบคุมข้อมูลที่จะเข้าในแต่ละโหนด ประกอบด้วย Forget Gate Layer, Input Gate Layer และ Output Gate Layer สามารถอธิบายรายละเอียดหน้าที่และหลักการทำงานได้ดังนี้

1) เกตลืมข้อมูล (Forget Gate Layer)

เป็นเกตที่มีหน้าที่กำหนดว่าข้อมูลที่เข้ามาในเซลล์ (Cell State) ควรถูกจัดเก็บไว้หรือทิ้งไป โดยข้อมูล ที่พิจารณาว่าให้จัดเก็บนั้น ๆ รวมกับผลลัพธ์ที่ได้จากการคำนวณของโหนดก่อนหน้ามาผ่านฟังก์ชัน ReLU (Oinkina, 2015) นิยามดังสมการที่ (2-18)

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \qquad (2\text{-}18)$$

โดยที่ f_t คือ เกตลืมข้อมูล (Forget Gate)

σ คือ ฟังก์ชัน ReLU

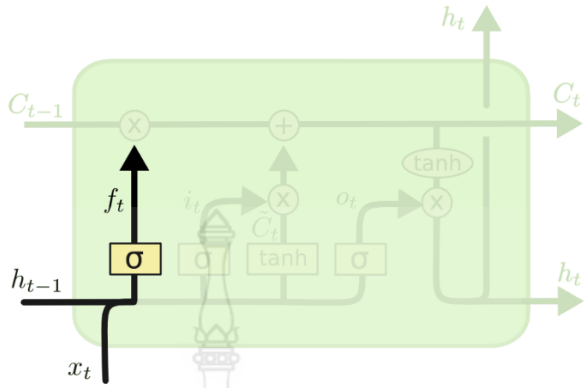
W_f คือ ค่าน้ำหนักของเมตริกซ์

h_{t-1} คือ ค่าผลลัพธ์ของโหนดก่อนหน้า

x_t คือ ค่าข้อมูลนำเข้าที่เข้ามาในโหนดปัจจุบัน

b_f คือ ค่าความเอนเอียง (Bias)

แบบจำลองการทำงานของ Forget Gate Layer แสดงดังภาพที่ 2-16



ภาพที่ 2-16 แบบจำลองการทำงานของ Forget Gate Layer

From: Understanding LSTM Networks. Oinkina. (2015). Colah's blog. (https://colah.github.io/ posts/2015-08-Understanding-LSTMs/)

ผลลัพธ์ที่ได้จาก Forget Gate Layer จะอยู่ระหว่างค่าศูนย์ (0) และหนึ่ง (1) โดยกรณีมีค่าเป็น 0 หมายถึงสถานะลบค่า Cell State เดิมออก และในกรณีมีค่าเป็น 1 หมายถึงสถานะเก็บค่า Cell State ไว้

2) เกตข้อมูลขาเข้า (Input Gate Layer)

เป็นเกตที่ทำหน้าที่รับข้อมูลอินพุตเข้ามาใหม่แล้วเขียน (Write) ข้อมูลลงไปในแต่ละโหนด ด้วยการทำงาน 2 ส่วน คือ Input Gate ส่วนควบคุมการ Update Cell State ว่าจะ Update หรือ ไม่ Update อีกส่วนหนึ่งคือ Input Gate ส่วนเลือกที่ทำการ Update Cell State นิยามการทำงานดังสมการที่ (2-19) โดยใช้ฟังก์ชัน tanh สร้าง Candidate Value ขึ้นมาใน State (Oinkina, 2015) นิยามดังสมการที่ (2-20)

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2\text{-}19)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (2\text{-}20)$$

โดยที่ i_t คือ เกตข้อมูลนำเข้า (Input Gate)

σ คือ ฟังก์ชัน ReLU

C_t คือ ค่า Candidate ของ Cell State ที่เวลา t

W_i, W_c คือ ค่าน้ำหนักของเมตริกซ์

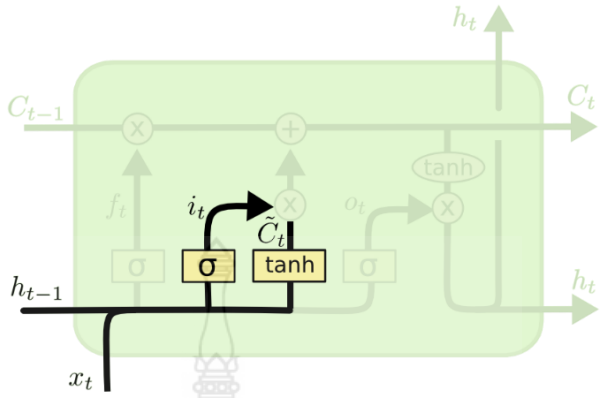
$\tanh$ คือ ฟังก์ชัน $\tanh(x) = \dfrac{1}{1+e^{-2x}} - 1$

h_{t-1} คือ ค่าผลลัพธ์ของโหนดก่อนหน้าที่เวลา $t-1$

x_t คือ ค่าข้อมูลนำเข้าที่เข้ามาในโหนดที่เวลา t

b_i, b_c คือ ค่าความเอนเอียง (Bias)

แบบจำลองการทำงานของ Input Gate Layer แสดงดังภาพที่ 2-17



ภาพที่ 2-17 แบบจำลองการทำงานของ Input Gate Layer

From: Understanding LSTM Networks. Oinkina. (2015). Colah's blog. (https://colah.github.io/ posts/2015-08-Understanding-LSTMs/)

3) เกตข้อมูลขาออก (Output Gate Layer)

เป็นเกตที่มีหน้าที่ตรวจสอบการส่งออกข้อมูล (Output Data) โดยพิจารณาข้อมูลผลลัพธ์จาก Cell State ที่ผ่านการคำนวณต่าง ๆ ด้วยฟังก์ชัน ReLU นิยามดังสมการที่ (2-21) โดยเลือกว่าข้อมูลของ Cell State ที่จะถูกส่งออก จากนั้นนำค่า Cell State ดังกล่าวผ่านฟังก์ชัน tanh เพื่อแปลงค่าข้อมูลให้อยู่ระหว่าง 1 หรือ -1 แล้วนำค่าผลลัพธ์ที่ได้มาคำนวณกับค่าผลลัพธ์ที่ได้จาก ReRU Gate ซึ่งจะได้ค่าผลลัพธ์ที่ต้องการ (Oinkina, 2015) นิยามดังสมการที่ (2-22)

$$O_t = \sigma(W_o \cdot \left[h_{t-1}, x_t\right] + b_o) \quad (2\text{-}21)$$

$$h_t = O_t * \tanh(C_t) \quad (2\text{-}22)$$

โดยที่ O_t คือ เกตข้อมูลนำออก (Output Gate)

σ คือ ฟังก์ชัน ReLU

W_o คือ ค่าน้ำหนักของเมตริกซ์

h_{t-1} คือ ค่าผลลัพธ์ของโหนดก่อนหน้าที่เวลา $t-1$

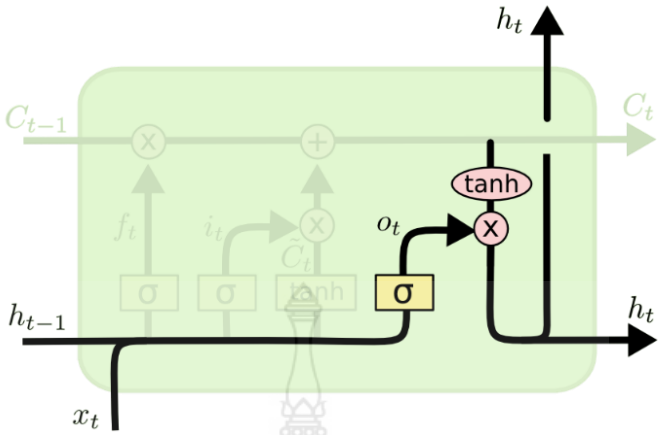
h_t คือ ค่าผลลัพธ์ของโหนดก่อนหน้าที่เวลา t

b_o คือ ค่าความเอนเอียง (Bias)

$\tanh$ คือ ฟังก์ชัน $\tanh(x) = \dfrac{1}{1+e^{-2x}} - 1$

C_t คือ ค่า Candidate ของ Cell State ที่เวลา t

แบบจำลองการทำงานของ Output Gate Layer แสดงดังภาพที่ 2-18



ภาพที่ 2-18 แบบจำลองการทำงานของ Output Gate Layer

From: Understanding LSTM Networks. Oinkina. (2015). Colah's blog. (https://colah.github.io/ posts/2015-08-Understanding-LSTMs/)

ตัวอย่างการพัฒนาชุดคำสั่งภาษาไพธอนร่วมกับ Keras Library สำหรับการทำนายข้อมูลจำนวนจริง หรือค่าต่อเนื่อง (Regression) ด้วยโครงข่ายประสาทเทียมแบบ LSTM มีรายละเอียดดังนี้ (Ihsncnkz, 2022)

ชุดคำสั่งเรียกใช้ไลบรารี Keras Library

```
from keras.models import Sequential
from keras.layers import Dense, LSTM
```

ชุดคำสั่งสร้างแบบจำลองทำนายข้อมูลด้วย LSTM

```
LSTMmodel = Sequential()
LSTMmodel.add(LSTM(10, input_shape = (1, timesteps)))
LSTMmodel.add(Dense(1))
LSTMmodel.compile(loss = "mean_squared_error", optimizer = "adam")
```

เมื่อกำหนดข้อมูลสำหรับการเรียนรู้ด้วยชุดคำสั่ง

```
inputs = data_set[len(data_set) - len(test) - timesteps:]
inputs = scaler.transform(inputs)
X_test = []
for i in range(timesteps, inputs.shape[0]):
    X_test.append(inputs[i - timesteps:i, 0])
testX = np.array(X_test)
testX = testX.reshape(testX.shape[0], 1, testX.shape[1])
```

ดำเนินการเรียนรู้ข้อมูลจากชุดข้อมูลฝึกสอนด้วยชุดคำสั่ง

```
LSTMmodel.fit(trainX, Y_train, epochs = 50, batch_size = 1)
```

ดำเนินการตรวจสอบค่าคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของแบบจำลอง ด้วยชุดคำสั่ง

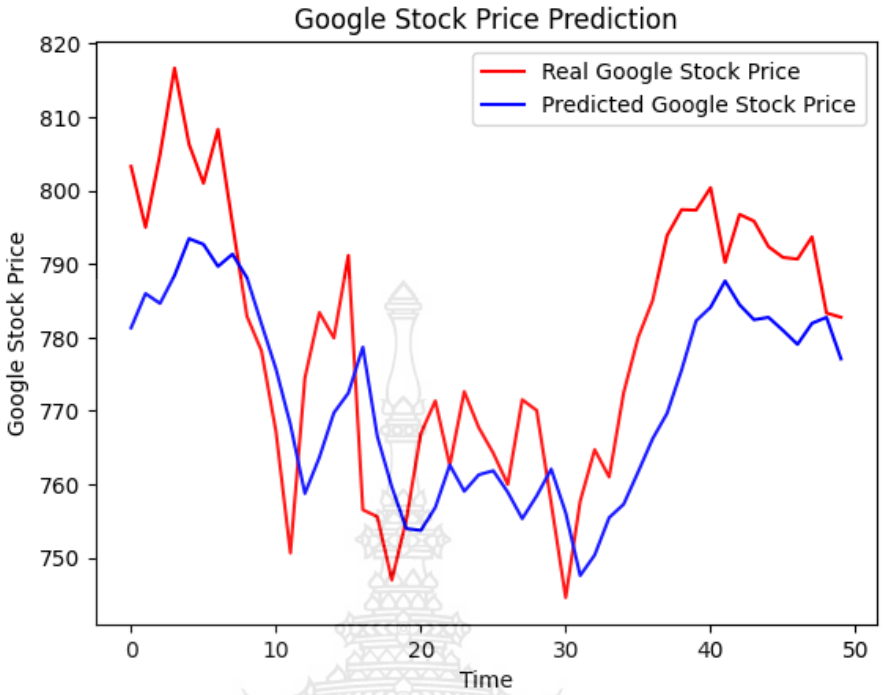
```
print('MSE of LSTM model = %.2f'%mean_squared_error(predict_lstm,y_true))
```

ซึ่งจะได้ค่าผลลัพธ์ดังนี้ *MSE of LSTM model = 2025.00*

ดำเนินการทดสอบการทำนายข้อมูลเขียนชุดคำสั่งได้ดังนี้

```
predict_lstm = model.predict(testX)
```

ภาพแผนภูมิเส้นเปรียบเทียบการทำนายข้อมูลด้วยแบบจำลอง LSTM รายละเอียดดังภาพที่ 2-19



ภาพที่ 2-19 แผนภูมิเส้นผลลัพธ์การทำนายข้อมูลด้วยโครงข่ายประสาทเทียมแบบ LSTM
From: RNN And LSTM With Keras. İHSAN. (2022). koggle. (https://www.kaggle.com/code/ihsncnkz/rnn-and-lstm-with-keras)

2.8 การวัดประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่อง

แบบจำลองการเรียนรู้ของเครื่อง (Machine Learning Model) ที่ได้จากการเรียนรู้ข้อมูลชุดฝึกสอน (Training Dataset) และถูกนำไปใช้ในการทำนายข้อมูลตัวเลข (Regression) ผลลัพธ์ที่ได้จากการทำนายข้อมูลใหม่ (Testing Dataset) จะมีความแม่นยำมากน้อยเพียงใดสามารถวัดจากค่าประสิทธิภาพ โดยค่าประสิทธิภาพสำหรับแบบจำลองการทำนายข้อมูลตัวเลขมีอยู่ด้วยกันหลายค่า (Mark E. Fenner, 2020) รายละเอียดค่าประสิทธิภาพต่าง ๆ มีดังต่อไปนี้

2.8.1 ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error; MAE)

เป็นค่าประสิทธิภาพของแบบจำลองด้วยการวัดค่าคลาดเคลื่อนจากค่าเฉลี่ยสัมบูรณ์ผลต่างระหว่างผลลัพธ์ค่าทำนาย $(\hat{Y}_t)$ กับค่าจริง (Y_t) เป็นการอธิบายถึงขนาดของค่าคลาดเคลื่อนรวมของการใช้แบบจำลองทำนายข้อมูล โดยเป้าหมายของการทำนายข้อมูลมุ่งหวังให้ค่า MAE มีค่าน้อยสุด เพราะแสดงให้เห็นว่าผลการทำนายมีค่าคลาดเคลื่อนน้อย แบบจำลองมีความแม่นยำมาก นิยามดังสมการที่ (2-23)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|(y_i - \hat{y}_i)\right| \qquad (2\text{-}23)$$

โดยที่ MAE คือ ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย
n คือ จำนวนข้อมูล
Y_t คือ ค่าข้อมูลจริงที่เวลา t ใด ๆ
$\hat{Y}_t$ คือ ค่าข้อมูลผลลัพธ์การทำนายที่เวลา t ใด

2.8.2 ค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared Error; MSE)

เป็นค่าประสิทธิภาพของแบบจำลองด้วยการวัดค่าผลต่างระหว่างผลลัพธ์ค่าทำนาย $(\hat{Y}_t)$ กับค่าจริง (Y_t) ทั้งหมดยกกำลังสอง โดยเป้าหมายการทำนายข้อมูลมุ่งหวังให้ค่า MSE มีค่าน้อยสุด เพราะแสดงให้เห็นว่าผลการทำนายคลาดเคลื่อนน้อยแบบจำลองที่ได้มีความแม่นยำมาก นิยามดังสมการที่ (2-24)

$$MSE = \frac{1}{n}\sum_{t=1}^{n}\left|(Y_t - \hat{Y}_t)^2\right| \qquad (2\text{-}24)$$

โดยที่ MSE คือ ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
n คือ จำนวนข้อมูล
Y_t คือ ค่าข้อมูลจริงที่เวลา t ใด ๆ
$\hat{Y}_t$ คือ ค่าข้อมูลผลลัพธ์การทำนายที่เวลา t ใด

2.8.3 ค่ารากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error; RMSE)

เป็นค่าประสิทธิภาพของแบบจำลองด้วยการวัดค่าคลาดเคลื่อนจากรากที่สองของค่าเฉลี่ยผลต่างระหว่างผลลัพธ์ค่าทำนาย $(\hat{Y}_t)$ กับค่าจริง (Y_t) ทั้งหมดยกกำลังสอง โดยเป้าหมายของการทำนายข้อมูลมุ่งหวังให้ค่า RMSE มีค่าน้อยสุด เพราะแสดงให้เห็นว่าผลการทำนายคลาดเคลื่อนน้อย แบบจำลองที่ได้มีความแม่นยำมาก นิยามดังสมการที่ (2-25)

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left|(Y_t - \hat{Y}_t)^2\right|} \qquad (2\text{-}25)$$

โดยที่ $RMSE$ คือ ค่ารากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ย
n คือ จำนวนข้อมูล
Y_t คือ ค่าข้อมูลจริงที่เวลา t ใด ๆ
$\hat{Y}_t$ คือ ค่าข้อมูลผลลัพธ์การทำนายที่เวลา t ใด ๆ

2.8.4 ค่าสัมประสิทธิ์การทำนาย (Coefficient of Determination; R^2)

ค่าสัมประสิทธิ์การทำนาย เป็นค่าที่แสดงให้เห็นถึงความสัมพันธ์ของตัวแปรข้อมูลนำเข้าทั้งหมดสามารถอธิบายการผันแปรของตัวแปรผลลัพธ์ หรือผลการทำนายได้มากน้อยเพียงใด มักใช้ในการวัดประสิทธิภาพของแบบจำลองแบบทำนายข้อมูลแบบเชิงเส้น (Linear Regressions) เรามักนำค่าสัมประสิทธิ์การทำนายมาคูณด้วย 100 ซึ่งจะอธิบายได้ถึงค่าร้อยละ สูตรการคำนวณหาค่าสัมประสิทธิ์การตัดสินใจพหุคูณนิยามดังสมการ (2-26)

$$R^2 = 1 - \frac{RSS}{TSS} \qquad (2\text{-}26)$$

โดยที่ R^2 คือ ค่าสัมประสิทธิ์การตัดสินใจพหุคูณ
RSS คือ ผลรวมค่าคลาดเคลื่อนจากการทำนายกำลังสอง (Residual Sun of Squares)
TSS ผลรวมค่าคลาดเคลื่อนของข้อมูลจริงกับค่าเฉลี่ย (Total Sum of Squares)

ค่า RSS คำนวณได้จากสูตรดังสมการที่ (2-27) และค่า TSS คำนวณได้จากสูตรนิยามดังสมการที่ (2-28) ตามลำดับ

$$RSS = \sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2 \quad (2\text{-}27)$$

$$TSS = \sum_{t=1}^{n}(Y_t - \bar{Y})^2 \quad (2\text{-}28)$$

โดยที่ Y_t คือ ค่าข้อมูลจริงที่เวลา t ใด ๆ

$\hat{Y}_t$ คือ ค่าข้อมูลผลลัพธ์การทำนายที่เวลา t ใด ๆ

$\bar{Y}$ คือ ค่าเฉลี่ยข้อมูลจริง

n คือ จำนวนข้อมูล

2.9 งานวิจัยที่เกี่ยวข้อง

สำหรับผลงานวิจัยที่นำการเรียนรู้ของเครื่อง (ML) และการเรียนรู้เชิงลึก (DL) ซึ่งเป็นศาสตร์ด้านปัญญาประดิษฐ์ (AI) มาสร้างแบบจำลองเพื่อการทำนายข้อมูล (Prediction) ทั้งในส่วนทำนายข้อมูลกลุ่ม (Classification) และทำนายข้อมูลจำนวนจริง (Regression) มีอยู่มากมาย เพราะเนื่องจากในปัจจุบันปัญหาต่าง ๆ ถูกขับเคลื่อนด้วยเทคโนโลยีปัญญาประดิษฐ์ ผู้วิจัยขออธิบายและยกตัวอย่างผลงานวิจัยที่เกี่ยวข้องการนำเทคโนโลยีปัญญาประดิษฐ์มาใช้ในการทำนายข้อมูลจำนวนจริง สรุปพอสังเขปดังนี้

งานวิจัยเรื่องการทำนายราคาขายทรัพย์สินรอการขายประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานครโดยใช้เทคนิคการเรียนรู้ของเครื่อง (พีรภัทร วัสแสง และ กองกูณฑ์, 2567) เป็นการนำเสนอศาสตร์ด้านปัญญาประดิษฐ์กับอุตสาหกรรมอสังหาริมทรัพย์ งานวิจัยนี้เปรียบเทียบประสิทธิภาพการทำนายราคาบ้านเดี่ยวในเขตพื้นที่กรุงเทพมหานครด้วยแบบจำลองการเรียนรู้ของเครื่อง 3 เทคนิควิธีคือ เทคนิควิธีซัพพอร์ทเวคเตอร์รีเกรสชัน (Support Vector Machine; SVR) เทคนิควิธีต้นไม้กาเดียนบูทเสต็ท (Gradient Boosted Trees; GBT) เทคนิควิธีโครงข่ายประสาทเทียม (Artificial Neural Network; ANN) และเทคนิควิธีการทำเหมืองข้อมูลแบบรวมกลุ่ม (Ensemble Vote) โดยใช้ข้อมูลตัวอย่างทรัพย์สินประเภทบ้านเดี่ยวในพื้นที่กรุงเทพมหานคร จำนวน 446 ตัวอย่าง ดำเนินการวัดประสิทธิภาพแบบจำลองด้วยค่าประสิทธิภาพค่าสัมประสิทธิ์การทำนาย (R^2) ค่า RMSE ผลการวิจัยพบว่าเทคนิควิธี Ensemble Vote ให้ค่าประสิทธิภาพการทำนายข้อมูลดีสุด

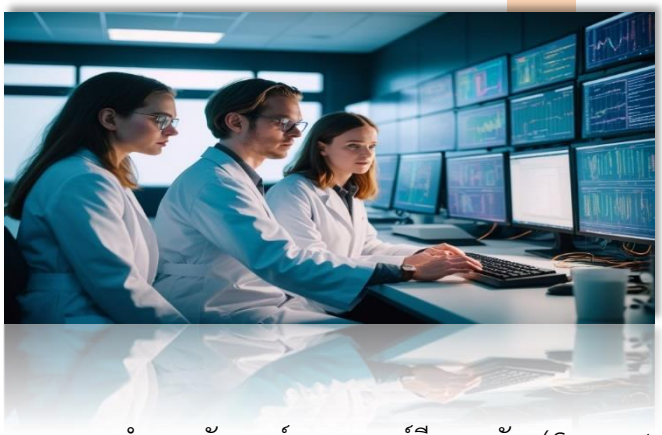
งานวิจัยการวิเคราะห์ข้อมูลเพื่อทำนายราคาข้าวโพดเลี้ยงสัตว์ด้วยขั้นตอนวิธีถดถอย นำเสนอโดย ภวิตา จันทันแก้ว และ นวลวรรณ สุนทรภิษัช (2566) งานวิจัยนี้เปรียบเทียบแบบจำลองการเรียนรู้ของเครื่องสำหรับทำนายข้อมูลราคาข้าวโพดเลี้ยงสัตว์ 3 แบบจำลอง ประกอบด้วย แบบจำลองวิธีถอดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression Model) แบบจำลองวิธีการถดถอยแบบลาซโซ (Lasso Regression Model) และแบบจำลองวิธีการถดถอยแบบริดจ์ (Ridge Regression Model) โดยใช้ข้อมูลจากสำนักงานเศรษฐกิจการเกษตรในช่วงเดือนมกราคม พ.ศ. 2545 ถึงเดือนพฤษภาคม 2562 ใช้ค่าวัดประสิทธิภาพการทำนายของแบบจำลองด้วยค่าประสิทธิภาพ MSE และ RMSE ผลการวิจัยนี้แบบจำลองวิธีการถดถอยเชิงเส้นแบบพหุคูณมีประสิทธิภาพในการทำงานที่ดีที่สุด

ผลงานวิจัยของ นิติกร จันหาญ และจารี ทองคำ (2565) นำเสนอการสร้างแบบจำลองการเรียนรู้ของเครื่องทำนายราคาบิตคอยน์ งานวิจัยเลือกใช้แบบจำลองการเรียนรู้ของเครื่อง 2 เทคนิควิธี คือ โครงข่ายประสาทเทียมแบบหลายชั้น (Multilayer Perceptron Neural Network) ซัพพอร์ทเวกเตอร์รีเกรสชัน (Support Vector Regression; SVR) และใช้วิธีทางสถิติด้วยวิธีอริมา (ARIMA) โดยใช้ข้อมูลราคาบิตคอยน์จากเว็บไซต์ www.coinmarketcap.com เก็บรวบรวมข้อมูลราคาบิตคอยน์ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2560 ถึง 31 ธันวาคม พ.ศ. 2562 วัดประสิทธิภาพการทำนายข้อมูลของแบบจำลองด้วยค่าประสิทธิภาพ RMSE และ MSE ผลการวิจัยพบว่าวิธีการทางสถิติด้วยวิธีคอริมาให้ประสิทธิภาพการทำนายดีสุด

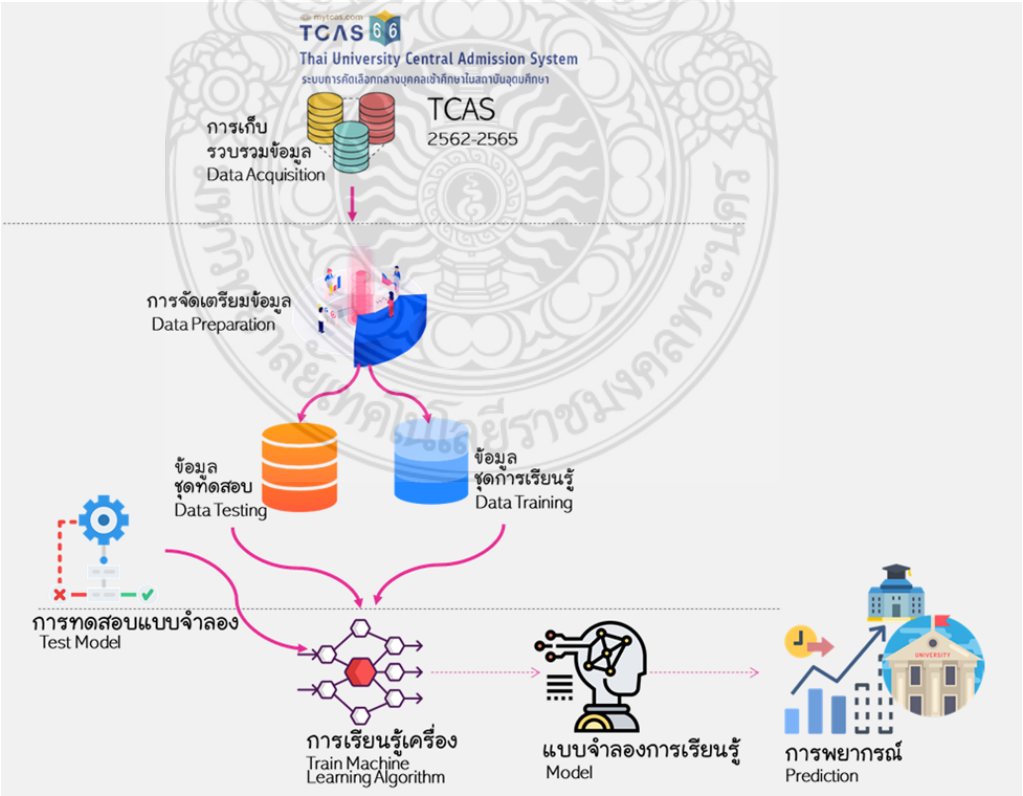
งานวิจัยเรื่อง แบบจำลองการทำนายระยะเวลาในการเข้าเทียบท่าของเรือโดยสารสาธารณะ (ชนะวิชญ พัชรเจริญวงษ์ และคณะ, 2563) นำเสนอการสร้างแบบจำลองเพื่อทำนายระยะเวลาการเข้าเทียบท่าของเรือโดยสารสาธารณะ โดยเก็บบันทึกข้อมูลจากอุปกรณ์ไอโอที (IoT) ตั้งแต่วันที่ 27 พฤศจิกายน 2561 ถึงวันที่ 31 พฤษภาคม 2562 ดำเนินการพัฒนาแบบจำลองการเรียนรู้ของเครื่อง 6 แบบจำลอง ประกอบด้วย Linear Regression, Random Forest, Gradient Boost, eXtreme Gradient Boost, Light Gradient Boost และ CatBoost วัดประสิทธิภาพแบบจำลองทำนายข้อมูลด้วยค่าประสิทธิภาพ RMSE ผลการวิจัยพบว่าแบบจำลองทำนายข้อมูล CatBoost มีความแม่นยำมากสุดที่ค่าคลาดเคลื่อนโดยเฉลี่ยในการเดินเรือที่ดีที่สุดที่เวลา 90.06 วินาที

ผลงานวิจัยของ รณชัย ชื่นธวัช และคณะ (2560) นำเสนอการทำนายความต้องการใช้หน่วยจำหน่ายไฟฟ้าด้วยแบบจำลองซัพพอร์ทเวกเตอร์รีเกรสชันแบบตรวจสอบสลับ 3 ส่วน โดยใช้ข้อมูลหน่วยจำหน่ายไฟฟ้าของการไฟฟ้านครหลวง (กฟน.) ที่เก็บรวบรวมจากรายงานสถานการณ์การจำหน่ายไฟฟ้าของการไฟฟ้านครหลวงของผู้บริโภคประเภทบ้านอยู่อาศัย ตั้งแต่เดือน มกราคม ถึงเดือนพฤษภาคม พ.ศ. 2558 ดำเนินการเรียนรู้ด้วยชุดข้อมูลฝึกสอน จากนั้นวัดประสิทธิภาพความแม่นยำของการทำนายข้อมูลด้วยค่าประสิทธิภาพ RMSE และ MAPE งานวิจัยนี้ยังเปรียบเทียบแบบจำลอง SVR กับวิธีการทางสถิติการถดถอยเชิงเส้นแบบพหุ (Multiple Linear Regression; MLR) ผลการวิจัยพบกว่าแบบจำลอง SVR มีความแม่นยำในการทำนายข้อมูลที่สูงกว่าวิธีการทางสถิติ

บทที่ 3
วิธีการดำเนินงานวิจัย



การศึกษาวิจัยเรื่อง การทำนายจำนวนผู้สมัครเข้าศึกษาในหลักสูตรปริญญาตรีด้านคอมพิวเตอร์ ในประเทศไทยโดยใช้ปัญญาประดิษฐ์ดำเนินการวิจัยภายใต้การบูรณาการศาสตร์ด้านปัญญาประดิษฐ์ (Artificial Intelligence) หรือ AI ด้วยการนำเทคนิคการเรียนรู้ของเครื่อง (Machine Learning; ML) และเทคนิคการเรียนรู้เชิงลึก (Deep Learning; DL) มาศึกษา พัฒนาแบบจำลองทำนายข้อมูล และเปรียบเทียบประสิทธิภาพผลลัพธ์การทำนายข้อมูลกับแบบจำลองจำนวน 5 แบบจำลอง ประกอบด้วย แบบจำลองสุ่มป่าไม้ (Random Forest Regression Model; RFR) แบบจำลองเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbors Model; KNN) แบบจำลองซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression Model; SVR) แบบจำลองโครงข่ายประสาทเทียมเพอร์เซฟตรอนแบบหลายชั้น (Multi-Layer Perceptron; MLP) และโครงข่ายประสาทเทียมแบบความจำขนาดสั้นระยะยาว (LSTM Neural Networks) โดยใช้ข้อมูลสถิติการสมัครเข้าเรียน จำนวน 6 ปีการศึกษาระหว่าง พ.ศ. 2561 ถึง พ.ศ. 2566 จากระบบการคัดเลือกกลางบุคคลเข้าศึกษาในสถาบันอุดมศึกษา (Thai University Central Admission System; TCAS) เข้าถึงผ่านลิงก์ https://www.mytcas.com/stat/ โดยมีกรอบแนวคิดในการดำเนินงานวิจัยดังภาพที่ 3-1



ภาพที่ 3-1 กรอบแนวคิดในการดำเนินงานวิจัย

จากภาพกรอบแนวคิดในการดำเนินงานวิจัยในครั้งนี้มุ่งเน้นศึกษาเกี่ยวกับการทำนายจำนวนผู้สมัครเข้าศึกษาในหลักสูตรปริญญาตรีด้านคอมพิวเตอร์ในประเทศไทย โดยข้อมูลหลักของงานวิจัยนี้ใช้ข้อมูลจากระบบ TCAS (Thai University Central Admission System) มีขั้นตอนวิธีการในการดำเนินงานวิจัยดังนี้

1) การเก็บรวบรวมข้อมูล (Data Acquisition)

2) การจัดเตรียมข้อมูล (Data Preparation)

3) การสร้างแบบจำลอง (Models Creation)

4) การทดสอบแบบจำลอง (Models Testing)

รายละเอียดการดำเนินงานในแต่ละขั้นตอนมีรายละเอียดดังต่อไปนี้

3.1 การเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูล (Data Acquisition) เป็นกระบวนการรวบรวมข้อมูลเป้าหมายสำหรับการสร้างแบบจำลองทำนายข้อมูลจำนวนผู้สมัครเข้าศึกษาในหลักสูตรปริญญาตรีด้านคอมพิวเตอร์ และข้อมูลสำหรับการทดสอบประสิทธิภาพผลการทำนายของแบบจำลองที่พัฒนาขึ้น งานวิจัยนี้เก็บรวบรวมข้อมูลสถิติการสมัครเข้าเรียนมาจากระบบ TCAS (Thai University Central Admission System) ลิงก์เว็บ https://www.mytcas.com/stat/ รายละเอียดหน้าเว็บเพจของ TCAS สถิติการสอบแสดงภาพที่ 3-2

ภาพที่ 3-2 รายละเอียดหน้าเว็บเพจ TCAS

จาก (ระบบการคัดเลือกกลางบุคคลเข้าศึกษาในสถาบันอุดมศึกษา, https://www.mytcas.com/stat/)

งานวิจัยนี้กำหนดข้อมูลเป้าหมายคือ ข้อมูลสถิติการสอบเข้าปี พ.ศ.2561-2566 รวมจำนวน 6 ปีการศึกษา โดยกำหนดให้ข้อมูลสถิติการสอบเข้าปี พ.ศ. 2561-2565 จำนวน 5 ปีการศึกษา เป็นข้อมูลสำหรับฝึกสอนในการสร้างแบบจำลอง และใช้ข้อมูลสถิติการสอบปี 2566 เป็นข้อมูลสำหรับการทดสอบประสิทธิภาพการทำนายข้อมูล โดยข้อมูล

สถิติการสอบเข้าปีการศึกษา พ.ศ. 2562-2566 จัดเก็บอยู่ในรูปแบบไฟล์ .XLSX และในปี พ.ศ. 2561 จัดเก็บในรูปแบบไฟล์ .PDF โดยเก็บแยกไฟล์เป็นรายปีการศึกษา ผู้วิจัยดำเนินการดึงข้อมูลจากเว็บสถิติการสอบเข้าในรูปแบบไฟล์ .XLSX จำนวน 5 ไฟล์ และไฟล์ .PDF จำนวน 1 ไฟล์ (สำหรับข้อมูลสถิติการสอบเข้าปีการศึกษา พ.ศ. 2561) รายละเอียดข้อมูลดังตารางที่ 3-1

ตารางที่ 3-1 รายละเอียดข้อมูลสถิติการสอบเข้าจากระบบ TCAS

ชื่อไฟล์	ลิงก์ไฟล์	รูปแบบไฟล์
TCAS61_maxmin.pdf	https://assets.mytcas.com/maxmin/TCAS61_maxmin.pdf	pdf
TCAS62_maxmin.xlsx	https://assets.mytcas.com/maxmin/TCAS62_maxmin.xlsx	xlsx
TCAS63_maxmin.xlsx	https://assets.mytcas.com/maxmin/TCAS63_maxmin.xlsx	xlsx
TCAS64_maxmin.xlsx	https://assets.mytcas.com/maxmin/TCAS64_maxmin.xlsx	xlsx
TCAS65_maxmin.xlsx	https://assets.mytcas.com/maxmin/TCAS65_maxmin.xlsx	xlsx
TCAS66_maxmin.xlsx	https://assets.mytcas.com/maxmin/TCAS66_maxmin.xlsx	xlsx

จากรายละเอียดข้อมูลสถิติการสอบเข้าจากระบบ TCAS ข้างต้น แสดงให้เห็นว่า ข้อมูลสถิติการสอบเข้า 5 ปี คือ พ.ศ.2562-2566 จัดเก็บในข้อมูลในรูปแบบตาราง .XLSX ซึ่งสามารถเขียนชุดคำสั่งภาษาไพธอนอ่านข้อมูลผ่านลิงก์ไฟล์เพื่อเก็บรวบรวมได้ทันที รายละเอียดชุดคำสั่งดังนี้

```
import requests
from google.colab import files

url =['https://assets.mytcas.com/maxmin/TCAS61_maxmin.pdf',
      'https://assets.mytcas.com/maxmin/TCAS62_maxmin.xlsx',
      'https://assets.mytcas.com/maxmin/TCAS63_maxmin.xlsx',
      'https://assets.mytcas.com/maxmin/TCAS64_maxmin.xlsx',
      'https://assets.mytcas.com/maxmin/TCAS65_maxmin.xlsx',
      'https://assets.mytcas.com/maxmin/TCAS66_maxmin.xlsx']
filename = ['TCAS61.pdf','TCAS62.xlsx',
            'TCAS63.xlsx','TCAS64.xlsx',
            'TCAS65.xlsx','TCAS66.xlsx']
for i in range(0,6):
  r = requests.get(url[i])
  with open(filename[i], "wb") as file:
    file.write(r.content)

filepath = ['/content/TCAS61.pdf',
            '/content/TCAS62.xlsx',
            '/content/TCAS63.xlsx',
            '/content/TCAS64.xlsx',
            '/content/TCAS65.xlsx',
            '/content/TCAS66.xlsx']

for i in range(0,6):
  files.download(filepath[i])
```

สำหรับข้อมูลสถิติการสอบเข้าปีการศึกษา 2561 จัดเก็บอยู่ในรูปแบบไฟล์ .pdf ดังนั้น ผู้วิจัยได้ดำเนินการดาวน์โหลดไฟล์แล้วนำไฟล์มาผ่านกระบวนการแปลงให้อยู่ในรูปแบบข้อมูลตาราง .XLSX ก่อน ดังนั้นสรุปไฟล์ข้อมูลในขั้นตอนการจัดเก็บและรวบรวมประกอบด้วย 6 ไฟล์

สำหรับรายละเอียดไฟล์ข้อมูลสถิติการสอบเข้าจากระบบ TCAS ในแต่ละปีการศึกษาที่ได้ดำเนินการเก็บรวบรวมตามวิธีการข้างต้น สามารถอธิบายรายละเอียดของข้อมูลได้ดังนี้

ข้อมูลสถิติการสอบเข้าปีการศึกษา 2561: มีจำนวนข้อมูลทั้งสิ้น 3,302 รายการ และมีมิติข้อมูลจำนวน 12 มิติข้อมูล รายละเอียดสารสนเทศของข้อมูลมีดังนี้

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3302 entries, 0 to 3301
Data columns (total 12 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   รหัสแถว             3302 non-null   int64
 1   คณะ                 3302 non-null   object
 2   หลักสูตรสาขาวิชา    3297 non-null   object
 3   ชื่อสถานบัน         3302 non-null   object
 4   รหัสรับร่วม         3302 non-null   int64
 5   จำนวนสมัคร          3302 non-null   int64
 6   จำนวนรับ            3302 non-null   int64
 7   จำนวนผ่าน           3302 non-null   int64
 8   คะแนนสูงสุด         2447 non-null   object
 9   คะแนนต่ำสุด         2447 non-null   object
 10  คะแนนเฉลี่ย         2447 non-null   object
 11  SD           2447 non-null   object
dtypes: int64(5), object(7)
memory usage: 309.7+ KB
```

ข้อมูลสถิติการสอบเข้าปีการศึกษา 2562: มีจำนวนข้อมูลทั้งสิ้น 4,084 รายการ และมีมิติข้อมูลจำนวน 22 มิติข้อมูล รายละเอียดสารสนเทศของข้อมูลมีดังนี้

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4084 entries, 0 to 4083
Data columns (total 22 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   รหัสแถว                         4084 non-null   int64
 1   ชื่อสถานบัน                     4084 non-null   object
 2   รหัสหลักสูตร                    4084 non-null   object
 3   ชื่อคณะ                         4084 non-null   object
 4   ชื่อหลักสูตร                    4084 non-null   object
 5   รหัสรับร่วม                      661 non-null    float64
 6   กลุ่มองค์ประกอบ                 4083 non-null   object
 7   PAT7                             691 non-null    object
 8   จำนวนรับ                        4084 non-null   int64
 9   จำนวนสมัคร                      3544 non-null   float64
 10  จำนวนผ่าน                       4084 non-null   int64
 11  อัตราการแข่งขัน                 4084 non-null   object
 12  ผ่านสัมภาษณ์                    4084 non-null   int64
 13  ไม่ประสงค์เข้าศึกษา/ตกสัมภาษณ์  4084 non-null   int64
 14  สละสิทธิ์                       4084 non-null   int64
 15  มีสิทธิ์เข้าศึกษา               4084 non-null   int64
 16  คะแนนสูงสุด                     3127 non-null   object
 17  คะแนนต่ำสุด                     3127 non-null   object
 18  คะแนนต่ำสุด(มีสิทธิ์เข้าศึกษา)  2715 non-null   object
 19  ความต่างของคะแนน                4084 non-null   object
 20  คะแนนเฉลี่ย                2997 non-null   object
 21  SD                   2596 non-null   object
dtypes: float64(2), int64(7), object(13)
memory usage: 702.1+ KB
```

ข้อมูลสถิติการสอบเข้าปีการศึกษา 2563: มีจำนวนข้อมูลทั้งสิ้น 4,120 รายการ และมีมิติข้อมูลจำนวน 18 มิติข้อมูล รายละเอียดสารสนเทศของข้อมูลมีดังนี้

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4120 entries, 0 to 4119
Data columns (total 18 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   ลำดับ             4120 non-null   int64
 1   รหัสหลักสูตร      4120 non-null   object
 2   สถาบัน            4120 non-null   object
 3   วิทยาเขต          4120 non-null   object
 4   คณะ               4120 non-null   object
 5   หลักสูตร          4120 non-null   object
 6   แขนง/วิชาเอก       513 non-null    object
 7   รหัสโครงการ       2105 non-null   object
 8   โครงการ/รูปแบบ    2237 non-null   object
 9   PAT               584 non-null    object
 10  รหัสรับร่วม        839 non-null    float64
 11  จำนวนรับ          4120 non-null   int64
 12  จำนวนสมัคร        4120 non-null   int64
 13  จำนวนผ่าน         4120 non-null   int64
 14  อัตราการแข่งขัน    3482 non-null   object
 15  คะแนนสูงสุด       3102 non-null   object
 16  คะแนนต่ำสุด       3102 non-null   object
 17  ความต่างของคะแนน  3102 non-null   object
dtypes: float64(1), int64(4), object(13)
memory usage: 579.5+ KB
```

ข้อมูลสถิติการสอบเข้าปีการศึกษา 2564: มีจำนวนข้อมูลทั้งสิ้น 8,371 รายการ และมีมิติข้อมูลจำนวน 18 มิติข้อมูล รายละเอียดสารสนเทศของข้อมูลมีดังนี้

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8371 entries, 0 to 8370
Data columns (total 18 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   ลำดับ             8371 non-null   int64
 1   รหัสหลักสูตร      8371 non-null   object
 2   สถาบัน            8371 non-null   object
 3   วิทยาเขต          8371 non-null   object
 4   คณะ               8371 non-null   object
 5   หลักสูตร          8371 non-null   object
 6   แขนง/วิชาเอก       1290 non-null   object
 7   โครงการ           3835 non-null   object
 8   รูปแบบ            8371 non-null   object
 9   PAT               659 non-null    object
 10  รหัสรับร่วม        1306 non-null   float64
 11  จำนวนรับ          8371 non-null   int64
 12  จำนวนสมัคร        8371 non-null   int64
 13  จำนวนผ่าน         8371 non-null   int64
 14  ยืนยันสิทธิ์       8371 non-null   int64
 15  อัตราการแข่งขัน    8371 non-null   object
 16  คะแนนสูงสุด       8371 non-null   float64
 17  คะแนนต่ำสุด       8371 non-null   float64
dtypes: float64(3), int64(5), object(10)
memory usage: 1.1+ MB
```

ข้อมูลสถิติการสอบเข้าปีการศึกษา 2565: มีจำนวนข้อมูลทั้งสิ้น 4,894 รายการ และมีมิติข้อมูลจำนวน 19 มิติข้อมูล รายละเอียดสารสนเทศของข้อมูลมีดังนี้

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4894 entries, 0 to 4893
Data columns (total 19 columns):
 #   Column                                  Non-Null Count  Dtype
---  ------                                  --------------  -----
 0   university_id                           4893 non-null   int64
 1   university_name                         4893 non-null   object
 2   program_id                              4893 non-null   object
 3   major_id                                4893 non-null   object
 4   project_id                              4893 non-null   object
 5   campus_name_th                          4893 non-null   object
 6   program_lookup_programs.faculty_name_th 4893 non-null   object
 7   program_name_th                         4893 non-null   object
 8   project_name_th                         2556 non-null   object
 9   major_name_th1                          4893 non-null   object
 10  major_name_th                           4229 non-null   object
 11  join_id                                 4726 non-null   float64
 12  จำนวนรับ                                4867 non-null   float64
 13  จำนวนสมัคร                              4776 non-null   float64
 14  จำนวนผ่าน                               4776 non-null   float64
 15  คะแนนต่ำสุด                             4776 non-null   float64
 16  คะแนนสูงสุด                             4776 non-null   float64
 17  คะแนนต่ำสุด หลังประมวลผลรอบ 2           4776 non-null   float64
 18  คะแนนสูงสุด หลังประมวลผลรอบ 2           4776 non-null   float64
dtypes: float64(8), int64(1), object(10)
memory usage: 726.4+ KB
```

ข้อมูลสถิติการสอบเข้าปีการศึกษา 2566: มีจำนวนข้อมูลทั้งสิ้น 4,670 รายการ และมีมิติข้อมูลจำนวน 14 มิติข้อมูล รายละเอียดสารสนเทศของข้อมูลมีดังนี้

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4670 entries, 0 to 4669
Data columns (total 14 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   สถาบัน          4670 non-null   object
 1   วิทยาเขต        4670 non-null   object
 2   รหัสหลักสูตร    4670 non-null   object
 3   คณะ/สำนักวิชา   4670 non-null   object
 4   ชื่อหลักสูตร    4670 non-null   object
 5   ชื่อหลักสูตร1   4670 non-null   object
 6   รายละเอียด      2600 non-null   object
 7   สาขาวิชา/วิชาเอก 840 non-null   object
 8   รหัสรับร่วม     4488 non-null   object
 9   จำนวนรับ        4670 non-null   int64
 10  จำนวนสมัคร      4670 non-null   object
 11  จำนวนผ่าน       4670 non-null   int64
 12  คะแนนต่ำสุด     4670 non-null   float64
 13  คะแนนสูงสุด     4670 non-null   float64
dtypes: float64(2), int64(2), object(10)
memory usage: 510.9+ KB
```

จากการเก็บรวบรวมข้อมูลสถิติการสอบเข้าทั้งสิ้น 6 ปีการศึกษา 2561-2566 มีจำนวนรายการข้อมูลทั้งสิ้น 29,441 รายการ ผู้วิจัยดำเนินการตรวจสอบข้อมูลสารสนเทศของไฟล์ข้อมูลสถิติการสอบเข้าในแต่ละปีการศึกษา พบว่าในแต่ละปีการศึกษาระบบ TCAS เก็บข้อมูลที่มีมิติ (คอลัมน์) ข้อมูลที่แตกต่างกัน และมีรูปแบบการบันทึกข้อมูลในมิติข้อมูลเดียวกันไม่เป็นไปตามแบบมาตรฐานเดียวกัน นอกจากนั้นในปีการศึกษา 2564 มีจำนวนรายการมากสุดสูงถึง 8,371 รายการ จากการสำรวจพบว่าปีการศึกษา 2564 จัดเก็บข้อมูลการสมัครเข้าด้วยการแจกแจงรอบการสมัคร ดังนั้นจึงดำเนินการสรุปยอดรวมการรับเข้าของแต่ละหลักสูตรในแต่ละรอบการรับสมัคร ดังนั้นจึงเป็นงานสำคัญของขั้นตอนการจัดเตรียมข้อมูลที่ต้องดำเนินการทำความสะอาดข้อมูล แปลงรูปแบบข้อมูล จัดรูปแบบ และเลือกมิติข้อมูลที่เกี่ยวข้องและสำคัญต่อผลการทำนายข้อมูลจำนวนผู้สมัครเข้าศึกษาต่อในสถาบันอุดมศึกษาระดับปริญญาตรีด้านคอมพิวเตอร์ ซึ่งผู้วิจัยได้อธิบายในหัวข้อการจัดเตรียมข้อมูลต่อไป

3.2 การจัดเตรียมข้อมูล

การจัดเตรียมข้อมูล (Data Preparation) เป็นขั้นตอนสำคัญที่ช่วยทำให้ข้อมูลพร้อมสำหรับการวิเคราะห์และสร้างแบบจำลอง โดยงานวิจัยนี้ดำเนินการจัดเตรียมข้อมูลดังนี้

ขั้นตอน: การทำความสะอาดข้อมูล (Data Cleaning)

1) ลบแถวข้อมูลจำนวนผู้สมัครที่เป็นศูนย์ (0) ลบข้อมูลค่าว่าง (Null)
2) แก้ไขข้อมูลที่พิมพ์ผิดพลาดเช่น สะกดผิด เว้นวรรค และอักษรพิเศษที่ไม่เกี่ยวข้องออก
3) จัดรูปแบบข้อมูลให้อยู่ในรูปแบบเดียวกัน ด้วยการแยกข้อมูลวิทยาเขตออกจากมิติข้อมูลชื่อสถาบัน ข้อมูลหลักสูตร และสาขาวิชา
4) จัดรูปแบบมิติข้อมูลชื่อหลักสูตรให้เป็นมาตรฐานเดียวกัน
5) สรุปรวมรายการรอบรับในแต่ละรอบของแต่ละหลักสูตรในปีการศึกษา 2564 เนื่องจากข้อมูลที่รวบรวมแจกแจงรายการรอบรับในแต่ละหลักสูตร
6) แปลงชื่อมิติข้อมูลเป็นภาษาอังกฤษเพื่อให้ง่ายต่อการเขียนชุดคำสั่ง
6) แปลงข้อมูลรูปแบบการจัดเก็บไฟล์ข้อมูลที่อยู่ในรูปแบบ .XLSX ไปเป็นรูปแบบ .CSV

ขั้นตอน: การสร้างมิติข้อมูลใหม่ (Feature Engineering)

1) สร้างมิติข้อมูลใหม่ ปีการศึกษา (year) ลงในแต่ละไฟล์เพราะในข้อมูลต้นฉบับ หรือข้อมูลดิบที่เก็บรวมรวมแยกไฟล์เป็นในแต่ละปีการศึกษาไว้ ดังนั้นเมื่อต้องการรวมข้อมูลสถิติการสอบเข้าในทุกปีการศึกษาเข้าด้วยกันจึงจำเป็นต้องเพิ่มมิติข้อมูลปีการศึกษาเพื่อแยกข้อมูลรายปีการศึกษาไว้
2) สร้างมิติข้อมูล ชื่อมหาวิทยาลัย/สถาบันการศึกษา (university_m) โดยดำเนินการแยก (split data) มิติข้อมูลชื่อสถาบันกับวิทยาเขตออกจากกันเพื่อให้ได้ชื่อสถาบันหรือชื่อมหาวิทยาลัย
3) สร้างมิติข้อมูล ชื่อย่อหลักสูตร (curriculum_c) ชื่อหลักสูตร (curriculum_m) และ ชื่อสาขาวิชา (major_m) โดยดำเนินการแยกข้อมูลจากมิติข้อมูลเดิมคือ ชื่อหลักสูตร

4) สร้างมิติข้อมูลใหม่ ชื่อกลุ่มตามระบบ ISCED (isced_m) โดย ISCED (International Standard Classification of Education) เป็นระบบมาตรฐานในการจำแนกสาขาวิชา โดยงานวิจัยนี้กำหนดเป้าหมายกลุ่ม ISCED จำนวน 3 กลุ่มคือ กลุ่มการศึกษา (Education) ที่เปิดการเรียนการสอนในสาขาวิชาด้านคอมพิวเตอร์ กลุ่มเทคโนโลยีสารสนเทศ และการสื่อสาร (Information and Communication Technology) กลุ่มวิศวกรรม อุตสาหกรรม และการก่อสร้าง (Engineering, manufacturing and construction) ในสาขาวิชาวิศวกรรมคอมพิวเตอร์ สาขาวิชาวิศวกรรมซอฟต์แวร์ เป็นต้น โดยงานวิจัยนี้ไม่นับรวมข้อมูลหลักสูตรศิลปกรรมศาสตรบัณฑิต (ศป.บ) หลักสูตรศิลปศาสตรบัณฑิต (ศล.บ) ที่เปิดการเรียนการสอนด้านคอมพิวเตอร์สำหรับงานออกแบบ ศิลปะ และอื่น ๆ

ขั้นตอน: การลดมิติของข้อมูล (Dimensionality Reduction)

จากการตรวจสอบข้อมูลพบกว่ามิติข้อมูลในแต่ละไฟล์ต้นฉบับมีจำนวนมิติข้อมูลที่แตกต่างกัน ดังนั้นผู้วิจัยดำเนินการลดมิติข้อมูลที่ไม่เกี่ยวข้องออกเพื่อให้ไฟล์สามารถรวมกันได้

ขั้นตอน: การรวมข้อมูล (Aggregate Data)

ดำเนินการรวมข้อมูลสถิติสอบเข้าทั้ง 6 ไฟล์ที่ผ่านกระบวนทำความสะอาด การสร้างมิติข้อมูลใหม่ การลดมิติข้อมูลเข้าด้วยกัน เขียนคำสั่งภาษาไพธอนได้ดังนี้

```
TCASdata = pd.concat([df2561,
                      df2562,
                      df2563,
                      df2564,
                      df2565,
                      df2566],
                      ignore_index=True)
TCASdata.to_csv('TCAS2561-2566_Clearn.csv',
                encoding='utf-8')
```

ขั้นตอน: การระบุมิติข้อมูล หรือตัวแปรอินพุตที่เกี่ยวข้อง (Feature Selection)

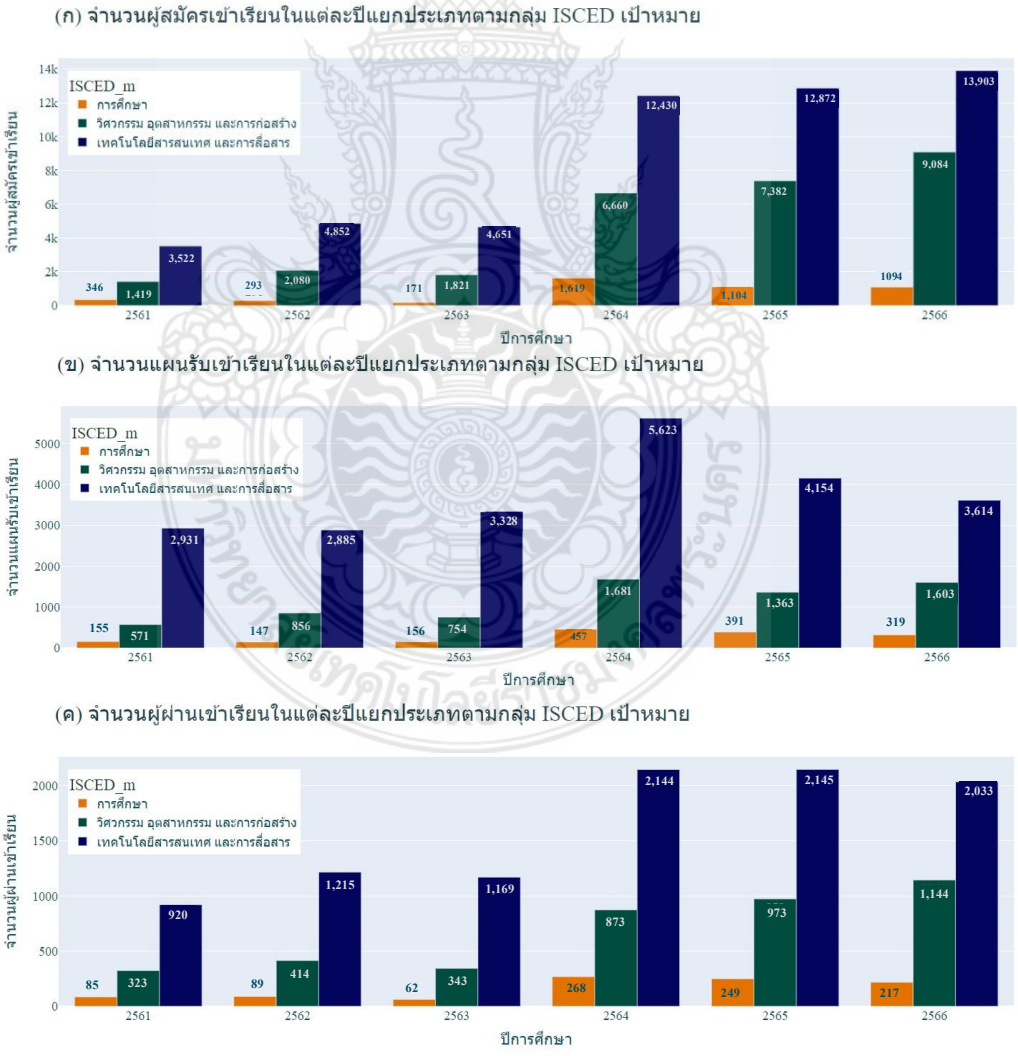
เป็นขั้นตอนการคัดกรองข้อมูลเพื่อให้การวิเคราะห์มีความแม่นยำและตรงประเด็นมากขึ้น โดยการคัดกรองข้อมูลเป็นขั้นตอนที่สำคัญในการเลือกเฉพาะข้อมูลที่เกี่ยวข้องกับการทำนายจำนวนผู้สมัครเข้าศึกษาในหลักสูตรปริญญาตรีด้านคอมพิวเตอร์ ด้วยการกำหนดคีย์ หรือคำศัพท์ที่เกี่ยวข้องกับชื่อสาขาวิชาที่เกี่ยวข้องกับคอมพิวเตอร์ เพื่อคัดกรองรายการข้อมูลสาขาวิชาที่ต้องการ คีย์กรองข้อมูล ประกอบด้วย “คอมพิวเตอร์” “เทคโนโลยีสารสนเทศ” “ซอฟต์แวร์” และ “วิทยาการข้อมูล” โดยเขียนชุดคำสั่งภาษาไพธอนดังนี้

```
newdf=df.query('major_m.str.contains("คอมพิวเตอร์|\
                                      เทคโนโลยีสารสนเทศ|\
                                      ซอฟต์แวร์|\
                                      วิทยาการข้อมูล",
                                       na=False)')
```

ผลลัพธ์จะได้ข้อมูลทั้งหมด 913 แถว 11 คอลัมน์ โดยชื่อสาขาวิชาที่ผ่านคีย์สำหรับกรองข้อมูลจำนวน 84 สาขาวิชา รายละเอียดสารสนเทศข้อมูลมีดังนี้

```
<<class 'pandas.core.frame.DataFrame'>
RangeIndex: 913 entries, 0 to 912
Data columns (total 11 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   row_n          913 non-null    int64
 1   year           913 non-null    int64
 2   university_m   913 non-null    object
 3   faculty_m      913 non-null    object
 4   curriculumn_m  913 non-null    object
 5   curriculumn_c  913 non-null    object
 6   major_m        913 non-null    object
 7   ISCED_m        913 non-null    object
 8   regis_n        913 non-null    int64
 9   plan_n         913 non-null    int64
 10  pass_n         913 non-null    int64
dtypes: int64(5), object(6)
memory usage: 78.6+ KB
```

รายละเอียดผลลัพธ์จำนวนผู้สมัคร (regis_n) จำนวนแผนรับ (plan_n) และจำนวนผู้ผ่านการเข้าเรียน (pass) ในหลักสูตรสาขาวิชาด้านคอมพิวเตอร์ของแต่ละปีการศึกษาในกลุ่ม ISCED เป้าหมายของงานวิจัยนี้ แสดงดังภาพที่ 3-3



ภาพที่ 3-3 แผนภูมิแท่งเปรียบเทียบข้อมูลจำนวนผู้สมัคร แผนรับ และจำนวนผู้ผ่านการเข้าเรียนในแต่ละปีการศึกษา

จากภาพแผนภูมิแท่งเปรียบเทียบข้อมูลจำนวนผู้สมัคร แผนรับ และจำนวนผู้ผ่านการเข้าเรียนในแต่ละปีการศึกษา โดยแยกประเภทตามกลุ่ม ISCED 3 กลุ่ม คือ กลุ่มการศึกษา กลุ่มวิศวกรรม อุตสาหกรรม และการก่อสร้าง และกลุ่มเทคโนโลยีสารสนเทศ และการสื่อสาร ที่เปิดการเรียนการสอนในหลักสูตรด้านคอมพิวเตอร์พบว่าในภาพรวมมีอัตราเพิ่มขึ้นในทุก ๆ ปี อย่างเป็นลำดับ โดยเฉพาะอย่างยิ่งในปีการศึกษา 2564 มีอัตราเพิ่มขึ้นเป็นอย่างมากจากปีการศึกษา 2563 โดยหลักสูตรในกลุ่มเทคโนโลยีสารสนเทศ และการสื่อสาร มีจำนวนผู้สมัครสนใจเข้าศึกษามากที่สุด รองลงมาคือกลุ่มวิศวกรรม อุตสาหกรรม และการก่อสร้าง และสุดท้ายคือกลุ่มการศึกษา และเมื่อพิจารณาแผนภูมิแท่งเปรียบเทียบภาพที่ 3-3(ข) จำนวนแผนรับและภาพที่ 3-3(ค) จำนวนผู้ผ่านการคัดเลือกมีแนวโน้มของข้อมูลในทิศทางเดียวกัน

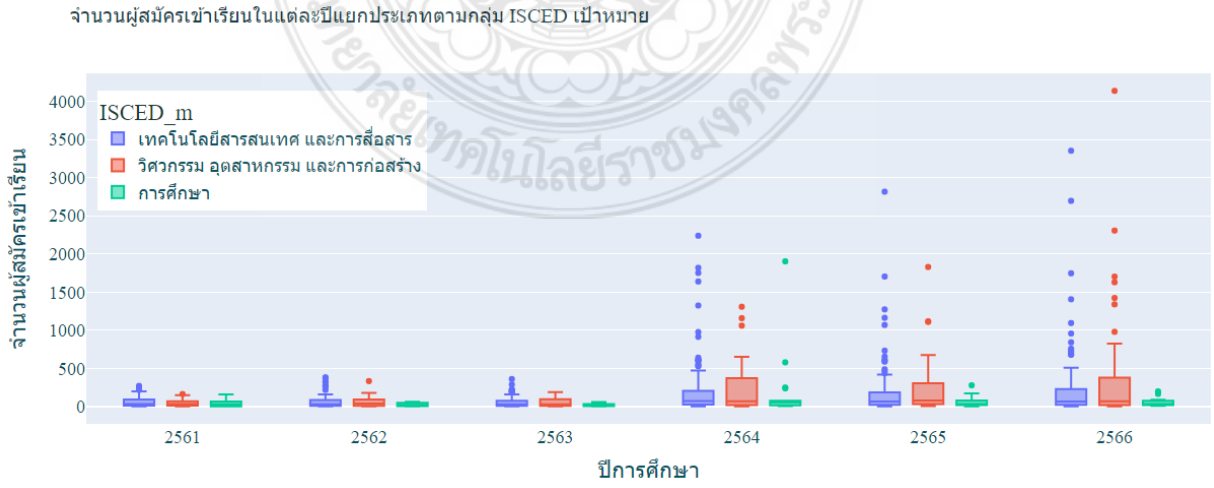
ขั้นตอน: ดำเนินการปรับข้อมูลส่วนผิดปกติ (Outlier)

Outlier คือ ข้อมูลที่มีค่าผิดปกติ หรือ ข้อมูลที่มีค่าสูงหรือต่ำกว่าจากข้อมูลส่วนใหญ่ในชุดข้อมูลหนึ่ง ๆ อย่างมาก สำหรับงานวิจัยนี้เลือกใช้วิธีการตรวจสอบข้อมูล Outlier ด้วยการตรวจสอบค่าทางสถิติของชุดข้อมูล และดำเนินการนำเสนอข้อมูลด้วยแผนภูมิ boxplot พัฒนาชุดคำสั่งด้วยภาษาไพธอน ร่วมกับไบรารี Plotly บนเครื่องมือ Colab ของ Google รายละเอียดค่าสถิติของข้อมูลดังนี้

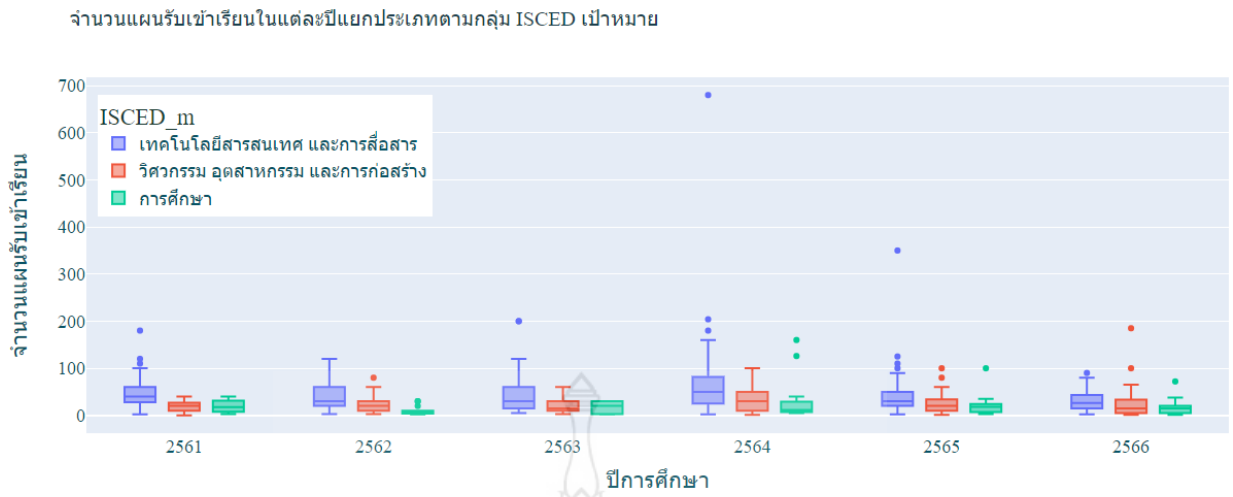
```
df.describe()
```

	row_n	year	regis_n	plan_n	pass_n
count	913.000000	913.000000	913.000000	913.000000	913.000000
mean	457.000000	2563.916758	154.947426	36.012048	17.405257
std	263.704696	1.658204	347.542607	39.115647	20.522347
min	1.000000	2561.000000	1.000000	0.000000	1.000000
25%	229.000000	2563.000000	14.000000	12.000000	4.000000
50%	457.000000	2564.000000	48.000000	29.000000	10.000000
75%	685.000000	2565.000000	138.000000	50.000000	24.000000
max	913.000000	2566.000000	4137.000000	680.000000	185.000000

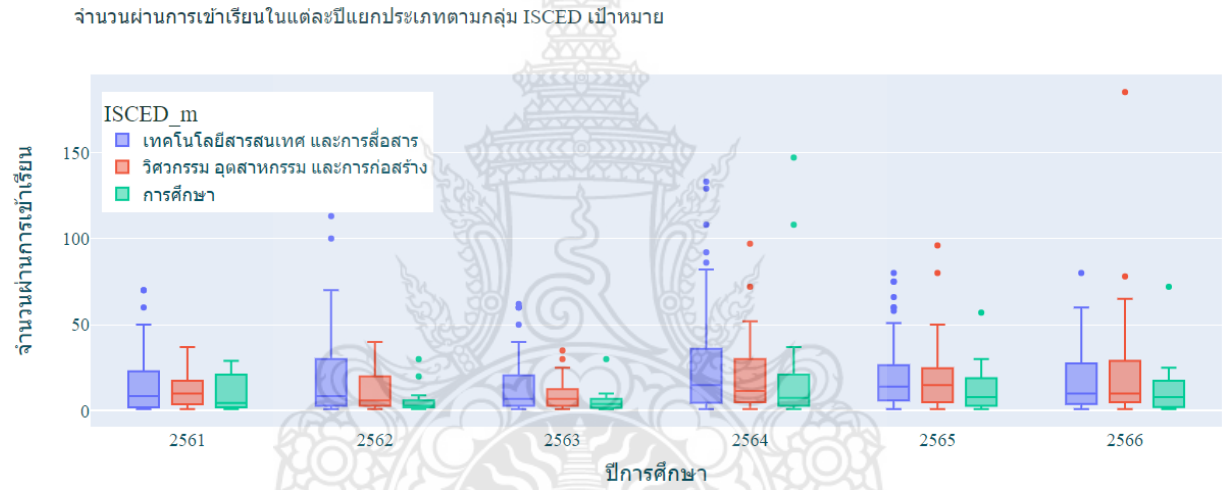
ผลลัพธ์การนำเสนอข้อมูลด้วย boxplot เพื่อตรวจสอบข้อมูล Outlier แสดงดังภาพที่ 3-4 ถึงภาพที่ 3-6 ตามลำดับ



ภาพที่ 3-4 แผนภูมิ boxplot ตรวจจับข้อมูล Outlier จำนวนผู้สมัคร



ภาพที่ 3-5 แผนภูมิ boxplot ตรวจจับข้อมูล Outlier จำนวนแผนรับ

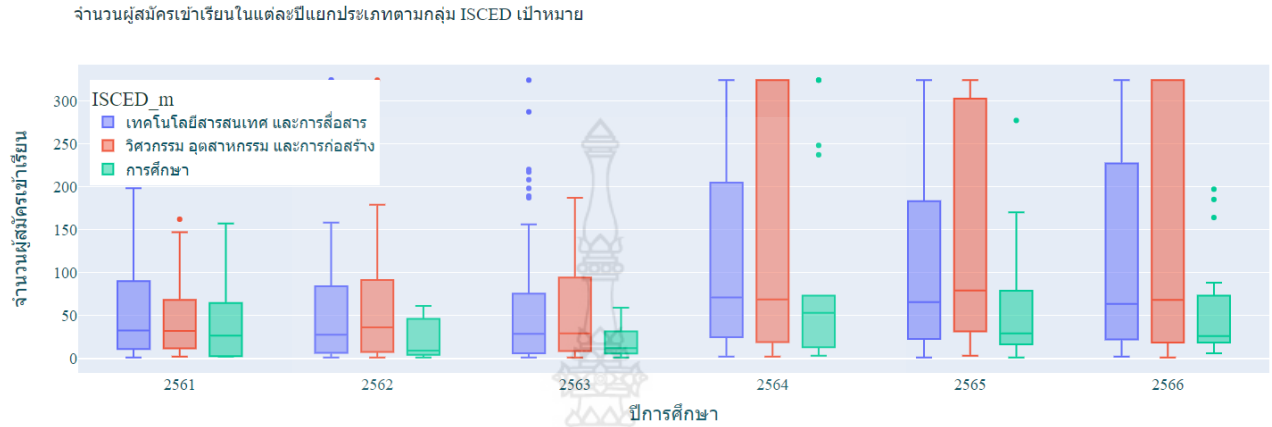


ภาพที่ 3-6 แผนภูมิ boxplot ตรวจจับข้อมูล Outlier จำนวนผู้ผ่านการเข้าเรียน

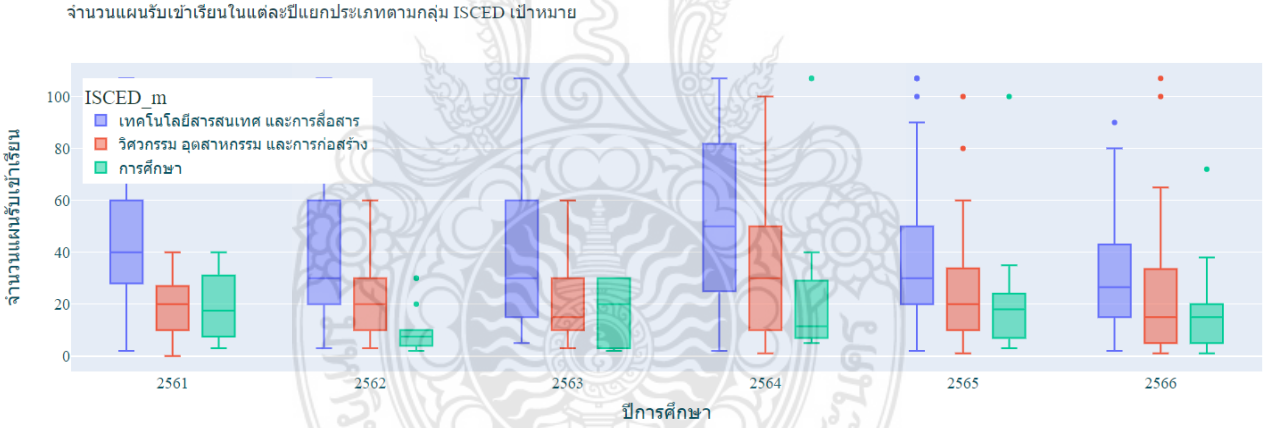
งานวิจัยนี้ดำเนินการปรับปรุงข้อมูล Outlier ด้วยเทคนิควิธีพิจารณาจากข้อมูลในควอนไทน์เริ่มต้น และควอนไทน์ระดับบน อาทิ เปอร์เซนไทล์ที่ 25 กับเปอร์เซ็นไทล์ที่ 75 ซึ่งหากข้อมูลที่มีค่าไม่อยู่ในช่วงที่พิจารณาจะถูกกำหนดค่าให้มีค่าเป็นค่ากลาง (Mean Value) ตัวอย่างชุดคำสั่งภาษาไพธอนปรับปรุงข้อมูล Outlier มิติข้อมูลจำนวนผู้สมัคร (regis_n)

```
import numpy as np

Q1  = df['regis_n'].quantile(0.25)
Q3  = df['regis_n'].quantile(0.75)
IQR = Q3 - Q1

whisker_width = 1.5
lower_whisker = Q1 -(whisker_width*IQR)
upper_whisker = Q3 + (whisker_width*IQR)

df['regis_n']=np.where(
                       df['regis_n']>upper_whisker,
                       upper_whisker,
                       np.where(df['regis_n']<lower_whisker,
                       lower_whisker,
                       df['regis_n'])
                      )
```

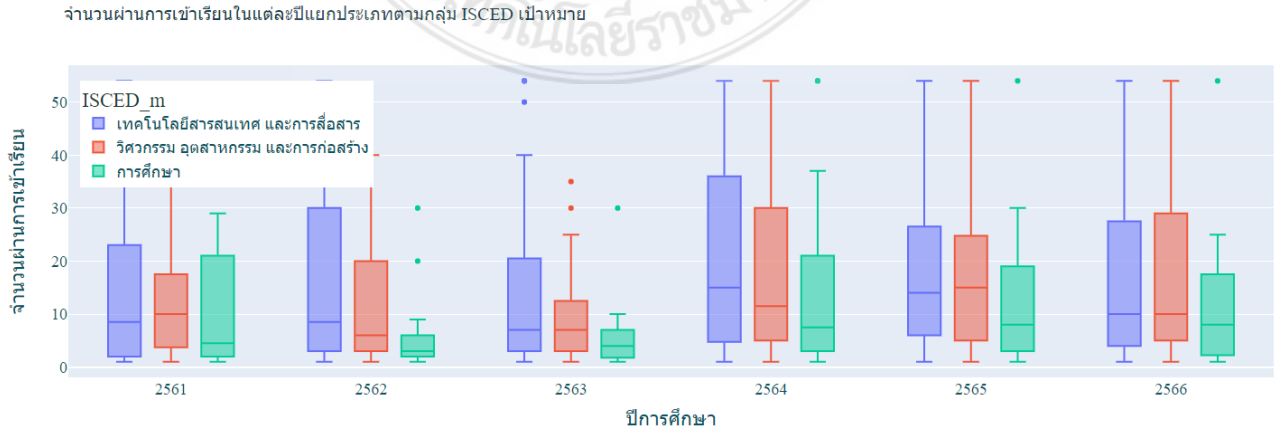
เมื่อดำเนินการนำเสนอข้อมูลด้วย boxplot เพื่อตรวจสอบข้อมูล Outlier อีกครั้งหลังการปรับปรุงข้อมูล Outlier แล้ว พบว่าจำนวนข้อมูลผู้สมัคร (regis_n) จำนวนแผนรับ (plan_n) และจำนวนผู้ผ่านเข้าเรียน (pass_n) มีจำนวน Outlier ที่น้อยลง รายละเอียดดังภาพที่ 3-7 ถึงภาพที่ 3-9 ตามลำดับ



ภาพที่ 3-7 แผนภูมิ boxplot หลังปรับค่าข้อมูล Outlier จำนวนผู้สมัคร



ภาพที่ 3-8 แผนภูมิ boxplot หลังปรับค่าข้อมูล Outlier จำนวนแผนรับ



ภาพที่ 3-9 แผนภูมิ boxplot หลังปรับค่าข้อมูล Outlier จำนวนผ่านการเข้าเรียน

ขั้นตอน: การแบ่งข้อมูล

งานวิจัยนี้ดำเนินการแบ่งข้อมูลออกเป็น 2 ส่วน คือ ชุดข้อมูลสำหรับการฝึกการเรียนรู้สำหรับสร้างแบบจำลองทำนายข้อมูลจำนวนผู้สมัครเข้าเรียนในสาขาวิชาด้านคอมพิวเตอร์ คือ ข้อมูลสถิติการสมัครเข้าเรียนปีการศึกษา 2561-2565 และชุดข้อมูลสำหรับการทดสอบประสิทธิภาพการทำนายข้อมูลของแบบจำลองที่พัฒนาขึ้น คือ ข้อมูลสถิติการสมัครเข้าเรียนปีการศึกษา 2566 ด้วยการเขียนชุดคำสั่งภาษาไพธอนดังนี้

```
dataTrain = df.loc[df['year'] != 2566]
dataTest  = df.loc[df['year'] == 2566]
```

ผลลัพธ์จากการแบ่งข้อมูลพบว่า ข้อมูลชุดฝึกสอนมีจำนวนรายการทั้งสิ้น 710 แถว และข้อมูลชุดทดสอบมีจำนวนรายการทั้งสิ้น แถว 203 แถว รายละเอียดสารสนเทศข้อมูลมีดังนี้

dataTrain

```
<class 'pandas.core.frame.DataFrame'>
Index: 710 entries, 0 to 709
Data columns (total 11 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   row_n           710 non-null    int64
 1   year            710 non-null    int64
 2   university_m    710 non-null    object
 3   faculty_m       710 non-null    object
 4   curriculumn_m   710 non-null    object
 5   curriculumn_c   710 non-null    object
 6   major_m         710 non-null    object
 7   ISCED_m         710 non-null    object
 8   regis_n         710 non-null    int64
 9   plan_n          710 non-null    int64
 10  pass_n          710 non-null    int64
dtypes: int64(5), object(6)
memory usage: 66.6+ KB
```

dataTest

```
<class 'pandas.core.frame.DataFrame'>
Index: 203 entries, 710 to 912
Data columns (total 11 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   row_n           203 non-null    int64
 1   year            203 non-null    int64
 2   university_m    203 non-null    object
 3   faculty_m       203 non-null    object
 4   curriculumn_m   203 non-null    object
 5   curriculumn_c   203 non-null    object
 6   major_m         203 non-null    object
 7   ISCED_m         203 non-null    object
 8   regis_n         203 non-null    int64
 9   plan_n          203 non-null    int64
 10  pass_n          203 non-null    int64
dtypes: int64(5), object(6)
memory usage: 19.0+ KB
```

สำหรับตัวอย่างข้อมูลแสดงดังภาพที่ 3-10 และภาพที่ 3-11 ตามลำดับ

	row_n	year	university_m	faculty_m	curriculumn_m	curriculumn_c	major_m	ISCED_m	regis_n	plan_n	pass_n
0	1	2561	จุฬาลงกรณ์มหาวิทยาลัย	คณะวิทยาศาสตร์	หลักสูตรวิทยาศาสตรบัณฑิต	วท.บ.	วิทยาการคอมพิวเตอร์	เทคโนโลยีสารสนเทศ และการสื่อสาร	131	60	60
1	2	2561	มหาวิทยาลัยเกษตรศาสตร์	คณะวิทยาศาสตร์	หลักสูตรวิทยาศาสตรบัณฑิต	วท.บ.	วิทยาการคอมพิวเตอร์	เทคโนโลยีสารสนเทศ และการสื่อสาร	233	40	40
2	3	2561	มหาวิทยาลัยเกษตรศาสตร์	คณะวิทยาศาสตร์	หลักสูตรวิทยาศาสตรบัณฑิต	วท.บ.	วิทยาการคอมพิวเตอร์	เทคโนโลยีสารสนเทศ และการสื่อสาร	114	50	50
3	4	2561	มหาวิทยาลัยเกษตรศาสตร์	คณะวิศวกรรมศาสตร์	หลักสูตรวิศวกรรมศาสตรบัณฑิต	วศ.บ.	วิศวกรรมคอมพิวเตอร์	วิศวกรรม อุตสาหกรรม และการก่อสร้าง	134	10	10
4	5	2561	มหาวิทยาลัยเกษตรศาสตร์	คณะวิศวกรรมศาสตร์	หลักสูตรวิศวกรรมศาสตรบัณฑิต	วศ.บ.	วิศวกรรมซอฟต์แวร์และความรู้	วิศวกรรม อุตสาหกรรม และการก่อสร้าง	20	10	10
...	...	...	...	...	...	...	...	...	...	...	...
705	706	2565	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ	คณะวิทยาศาสตร์และเทคโนโลยี	หลักสูตรวิทยาศาสตรบัณฑิต	วท.บ.	วิทยาการคอมพิวเตอร์	เทคโนโลยีสารสนเทศ และการสื่อสาร	37	20	11
706	707	2565	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน	คณะวิทยาศาสตร์และศิลปศาสตร์	หลักสูตรวิทยาศาสตรบัณฑิต	วท.บ.	วิทยาการคอมพิวเตอร์	เทคโนโลยีสารสนเทศ และการสื่อสาร	43	15	15
707	708	2565	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน	คณะวิศวกรรมศาสตร์และสถาปัตยกรรมศาสตร์	หลักสูตรวิศวกรรมศาสตรบัณฑิต	วศ.บ.	วิศวกรรมคอมพิวเตอร์	วิศวกรรม อุตสาหกรรม และการก่อสร้าง	79	10	11
708	709	2565	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน	คณะเกษตรศาสตร์และเทคโนโลยี	หลักสูตรวิทยาศาสตรบัณฑิต	วท.บ.	วิทยาการคอมพิวเตอร์	เทคโนโลยีสารสนเทศ และการสื่อสาร	4	5	2
709	710	2565	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน	คณะครุศาสตร์อุตสาหกรรม	หลักสูตรครุศาสตรอุตสาหกรรมบัณฑิต	ค.อ.บ.	เทคโนโลยีคอมพิวเตอร์	การศึกษา	21	3	3

710 rows × 11 columns

ภาพที่ 3-10 ตัวอย่างข้อมูลชุดสำหรับการฝึกสอนการเรียนรู้

	row_n	year	university_m	faculty_m	curriculumn_m	curriculumn_c	major_m	ISCED_m	regis_n	plan_n	pass_n
710	711	2566	จุฬาลงกรณ์มหาวิทยาลัย	คณะวิศวกรรมศาสตร์	หลักสูตรวิศวกรรมศาสตรบัณฑิต	วศ.บ.	วิศวกรรมคอมพิวเตอร์	วิศวกรรม อุตสาหกรรม และการก่อสร้าง	1422	65	65
711	712	2566	จุฬาลงกรณ์มหาวิทยาลัย	คณะวิศวกรรมศาสตร์	หลักสูตรวิศวกรรมศาสตรบัณฑิต	วศ.บ.	วิศวกรรมคอมพิวเตอร์และเทคโนโลยีดิจิทัล	วิศวกรรม อุตสาหกรรม และการก่อสร้าง	2305	185	185
712	713	2566	จุฬาลงกรณ์มหาวิทยาลัย	คณะวิทยาศาสตร์	หลักสูตรวิทยาศาสตรบัณฑิต	วท.บ.	วิทยาการคอมพิวเตอร์	เทคโนโลยีสารสนเทศ และการสื่อสาร	2695	24	24
713	714	2566	จุฬาลงกรณ์มหาวิทยาลัย	คณะพาณิชยศาสตร์และการบัญชี	หลักสูตรสถิติศาสตรบัณฑิต	สถ.บ.	สถิติและวิทยาการข้อมูล	เทคโนโลยีสารสนเทศ และการสื่อสาร	711	40	40
714	715	2566	จุฬาลงกรณ์มหาวิทยาลัย	คณะครุศาสตร์	หลักสูตรครุศาสตรบัณฑิต	ค.บ.	คอมพิวเตอร์ศึกษา	การศึกษา	88	10	10
...	...	...	...	...	...	...	...	...	...	...	...
908	909	2566	มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา	คณะวิทยาศาสตร์และเทคโนโลยีการเกษตร	หลักสูตรวิทยาศาสตรบัณฑิต	วท.บ.	วิทยาการคอมพิวเตอร์	เทคโนโลยีสารสนเทศ และการสื่อสาร	3	5	1
909	910	2566	มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย	คณะวิศวกรรมศาสตร์	หลักสูตรวิศวกรรมศาสตรบัณฑิต	วศ.บ.	วิศวกรรมคอมพิวเตอร์	วิศวกรรม อุตสาหกรรม และการก่อสร้าง	27	1	1
910	911	2566	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน	คณะวิทยาศาสตร์และศิลปศาสตร์	หลักสูตรวิทยาศาสตรบัณฑิต	วท.บ.	วิทยาการคอมพิวเตอร์	เทคโนโลยีสารสนเทศ และการสื่อสาร	49	5	5
911	912	2566	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน	คณะวิศวกรรมศาสตร์และเทคโนโลยี	หลักสูตรวิศวกรรมศาสตร์บัณฑิต	วศ.บ.	วิศวกรรมคอมพิวเตอร์	วิศวกรรม อุตสาหกรรม และการก่อสร้าง	77	5	5
912	913	2566	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน	คณะครุศาสตร์อุตสาหกรรม	หลักสูตรครุศาสตรอุตสาหกรรมบัณฑิต	ค.อ.บ.	เทคโนโลยีคอมพิวเตอร์	การศึกษา	18	1	1

203 rows × 11 columns

ภาพที่ 3-11 ตัวอย่างข้อมูลชุดสำหรับการทดสอบประสิทธิภาพผลการทำนาย

3.3 การสร้างแบบจำลอง

งานวิจัยนี้ดำเนินการศึกษา พัฒนา และเปรียบเทียบประสิทธิภาพแบบจำลองการเรียนรู้ของเครื่อง (Machine Learning Model) จำนวน 5 แบบจำลอง ประกอบด้วย แบบจำลองสุ่มป่าไม้ (Random Forest Model; RF) แบบจำลองเพื่อนบ้านใกล้สุด (K-Nearest Neighbors Model; KNN) แบบจำลองซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression Model; SVR) แบบจำลองโครงข่ายประสาทเทียมเพอร์เซฟตรอนแบบหลายชั้น (Multi Leyer Perceptron Neural Network Mode; MLP) และแบบจำลองโครงข่ายประสาทเทียมแบบความจำขนาดสั้นระยะยาว (Long Short Term Memory Neural Networks Model; LSTM) ดำเนินการพัฒนาชุดคำสั่งภาษาไพธอนร่วมกับไลบรารีสำหรับการเรียนรู้ของเครื่อง ประกอบด้วย ไลบรารี Scikit-learn (Sklearn) ไลบรารี TensorFlow และ Keras พัฒนาชุดคำสั่งบนเครื่องมือ colab รุ่น PRO ของบริษัทกูเกิ้ล (Google) รายละเอียดคุณสมบัติหน่วยประมวลผล (CPU) หน่วยความจำ (Ram) และหน่วยเก็บข้อมูล (Disk) และเวอร์ชันของไลบรารีการเรียนรู้ของเครื่อง มีรายละเอียดดังนี้

```
Python 3 Google Computer Engine backend
CPU:  GenuineIntel, Intel(R) Xeon(R) CPU @ 2.20GHz
RAM:  1.25GB/12.67GB
Disk: 30.82 GB/225.83 GB

The scikit-learn version is 1.3.2.
The Tensorflow version is 2.17.0.
The Keras version is 3.4.1.
```

สำหรับขั้นตอนการพัฒนาแบบจำลองการเรียนรู้ของเครื่องด้วยแบบจำลองทั้ง 5 ข้างต้น เริ่มต้นจากนำชุดข้อมูลสำหรับการฝึกสอนการเรียนรู้ (Training Dataset) ที่ผ่านกระบวนจัดเตรียมข้อมูลเรียบร้อยแล้ว โดยงานวิจัยกำหนดข้อมูลสำหรับฝึกสอนการเรียนรู้ คือ ข้อมูลสถิติการสมัครปีการศึกษา 2561-2565 ในสาขาวิชาที่เกี่ยวข้องกับคอมพิวเตอร์ จากนั้นดำเนินการหาค่าพารามิเตอร์ที่เหมาะสม เพื่อนำพารามิเตอร์ที่เหมาะสมมากำหนดค่าพารามิเตอร์ของแต่ละแบบจำลองในการสร้างแบบจำลองการเรียนรู้ของเครื่อง รายละเอียดในแต่ละขั้นตอนมีดังนี้

3.3.1 ขั้นตอนการหาพารามิเตอร์ที่เหมาะสม (Hyperparameter Optimization)

Hyperparameters คือ พารามิเตอร์ต่าง ๆ ที่ผู้พัฒนาแบบจำลองการเรียนรู้ของเครื่องสามารถกำหนดหรือตั้งค่าก่อนการเรียนรู้ของเครื่อง โดยในแต่ละแบบจำลองมีค่าพารามิเตอร์ที่แตกต่างกัน การหาค่าพารามิเตอร์ที่เหมาะสมสำหรับในแต่ละแบบจำลองนั้น จะมีผลให้ได้แบบจำลองการเรียนรู้ที่ได้มีความแม่นยำ (Accuracy) ที่ดี ปัจจุบันเทคนิควิธีการหาค่าพารามิเตอร์ที่เหมาะสมสามารถแบ่งออกได้ 2 ประเภทหลัก คือ (1) วิธีการดั้งเดิม (Traditional Hyperparameter Tuning) ใช้หลักการปรับค่าแล้วเทียบกับผลลัพธ์ไปเรื่อย ๆ จนกว่าจะพบค่าที่เหมาะสม ตัวอย่างวิธีการ คือ Manual Search, Grid Search และ Random Search และ (2) วิธีการปรับค่าพารามิเตอร์แบบอัตโนมัติ (Automated Hyperparameter Tuning) ใช้หลักการปรับค่าพารามิเตอร์แบบอัตโนมัติด้วยแบบจำลองต่าง ๆ ที่ถูกออกแบบมาเฉพาะงาน

สำหรับงานวิจัยนี้ดำเนินการหาค่าพารามิเตอร์ที่เหมาะสมของแบบจำลองทั้ง 5 ด้วยวิธีการดั้งเดิม โดยการใช้เทคนิควิธี Grid Search พัฒนาชุดคำสั่งภาษาไพธอน ร่วมกับไลบรารีการเรียนรู้ของเครื่อง รายละเอียดชุดคำสั่งมีดังนี้

```
modelName = ['K-Nearest Neighbors Model (KNN)',
             'Support Vector Regression Model (SVR)',
             'Random Forest Model (RF)',
             'Multi Layer Perceptron Neural Network (MLP)',
             'Long Short-Term Memory (LSTM)']

models = [neighbors.KNeighborsRegressor(), SVR(),
         RandomForestRegressor(),
         MLPRegressor(), KerasRegressor()]

hyperParameters =[{'n_neighbors' : [5,10,15,20,25,30,35,40,45,50],
                   'weights' : ['uniform','distance'],
                   'metric' : ['minkowski','euclidean','manhattan']},
                  {'C': [0.1, 1, 10, 100, 1000],
                   'gamma': [1, 0.1, 0.01, 0.001, 0.0001],
                   'kernel': ['rbf','linear']},
                  {'bootstrap': [True],
                   'max_depth': [80, 90, 100, 110],
                   'max_features': [2, 3],
                   'min_samples_leaf': [3, 4, 5],
                   'min_samples_split': [8, 10, 12],
                   'n_estimators': [100, 200, 300, 1000]},
                  {'hidden_layer_sizes': [(10,10), (100,100), (20,120)],
                   'max_iter': [50, 100],
                   'activation': ['tanh', 'relu'],
                   'solver': ['sgd', 'adam'],
                   'alpha': [0.001, 0.01, 0.5]},
                 {'neurons': [8, 16, 32, 64, 128],
                 'optimizer': ['SGD', 'RMSprop', 'Adam'],
                   'activation':['relu', 'tanh', 'sigmoid', 'linear','swish']}]

for i in range (0,m):
    grid_search.append(GridSearchCV(models[i],hyperParameters[i]))
    grid_search[i].fit(X_train,y_train)
    best_params = grid_search[i].best_params_
    best_score  = grid_search[i].best_score_
    test_score  = grid_search[i].score(X_test, y_test)
    results.append({
        'Model': modelName[i],
        'Best Parameters': best_params,
    })
```

ผลลัพธ์ค่าพารามิเตอร์ที่เหมาะสมในแต่ละแบบจำลองมีดังนี้

K-Nearest Neighbors Model (KNN)

```
Best Parameters: {'metric': 'manhattan',
                  'n_neighbors': 45,
                 'weights': 'uniform'}
```

Support Vector Regression Model (SVR)

```
Best Parameters: {'C': 0.1,
                  'gamma': 1,
                 'kernel': 'linear'}
```

Random Forest Model (RF)

```
Best Parameters: {'bootstrap': True,
                      'max_depth': 90,
                      'max_features': 3,
                      'min_samples_leaf': 3,
                      'min_samples_split': 10,
                      'n_estimators': 100}
```

Multi Leyer Perceptron Neural Network (MLP)

```
Best Parameters: {'hidden_layer_sizes': (100, 100,
                'max_iter':100,
                'activation': 'tanh',
                'solver': 'adam',
                'alpha': 0.05}
```

Long Short-Term Memory (LSTM)

```
Best Parameters: {'neurons': 32,
                  'optimizer': 'Adam',
                  'activation': 'tanh'}
```

3.3.2 ขั้นตอนการเรียนรู้ของเครื่อง (Leaning Model)

ขั้นตอนนี้ผู้วิจัยดำเนินการนำค่าพารามิเตอร์ที่เหมาะสมข้างต้นมากำหนดค่าให้กับแบบจำลองก่อนการเรียนรู้ของเครื่อง พัฒนาด้วยชุดคำสั่งภาษาไพธอน ร่วมกับไลบรารีการเรียนรู้ของเครื่องบนเครื่องมือ colab รายละเอียดชุดคำสั่งสำคัญมีดังนี้

- แบบจำลองสุ่มป่าไม้ (Random Forest Model; RF)

```
RFmodel = RandomForestRegressor(bootstrap = True,
                                max_depth = 90,
                                max_features = 3,
                                min_samples_leaf=3,
                                min_samples_split=10,
                                n_estimators=100,
                                random_state=1)

X_train = dataTrain[['plan_n'],['pass_n']]
y_train = dataTrain.regis_n

RFmodel.fit(X_train, y_train)
```

- แบบจำลองเพื่อนบ้านใกล้สุด (K-Nearest Neighbors Model; KNN)

```
KNNmodel =  neighbors.KNeighborsRegressor(
            n_neighbors=45,
            weights='uniform',
            algorithm='auto',
            metric='manhattan',
            leaf_size=50, p=5)

X_train = dataTrain[['plan_n'],['pass_n']]
y_train = dataTrain.regis_n

KNNmodel.fit(X_train, y_train)
```

- แบบจำลองซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression Model; SVR)

```
SVRmodel  = SVR(kernel="linear",
                C = 0.1,
                gamma = 1)

X_train = dataTrain[['plan_n'],['pass_n']]
y_train = dataTrain.regis_n

SVRmodel.fit(X_train, y_train)
```

- แบบจำลองโครงข่ายประสาทเทียมเพอร์เซฟตรอนแบบหลายชั้น (Multi Leyer Perceptron Neural Network Model; MLP)

```
MLPmodel = make_pipeline(StandardScaler(),
         MLPRegressor(activation=' tanh',
         hidden_layer_sizes=(100, 100),
         alpha=0.05,
         max_iter=100,
         solver='adam',
         random_state=2,
         early_stopping=False)

X_train = dataTrain[['plan_n'],['pass_n']]
y_train = dataTrain.regis_n

MPLmodel.fit(X_train, y_train)
```

- แบบจำลองโครงข่ายประสาทเทียมแบบความจำขนาดสั้นระยะยาว (LSTM Neural Networks Models)

```
LSTMmodel = Sequential()
LSTMmodel.add(LSTM(units=32,
              input_shape=(n_teps, n_features),
              return_sequences=True))
LSTMmodel.add(Dropout(0.2))
LSTMmodel.add(Dense(units=n_out))
LSTMmodel.add(Activation("relu"))
LSTMmodel.compile(loss='mse',
                  optimizer='rmsprop')

LSTMmodel.fit(X_train, y_train,
              epochs=30,
              batch_size=32,
              verbose=2)
```

3.4 การทดสอบประสิทธิภาพแบบจำลอง

งานวิจัยนี้เลือกใช้ค่าประสิทธิภาพ ค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared Error; MSE) และค่าสัมประสิทธิ์การทำนาย (Coefficient of Determination; R^2) วัดประสิทธิภาพผลการทำนายข้อมูลจำนวนผู้สมัครเข้าเรียนในหลักสูตรสาขาวิชาด้านคอมพิวเตอร์ โดยนำแบบจำลองทำนายข้อมูลที่ได้จากขั้นตอนการ สร้างแบบจำลองข้างต้น จำนวน 5 แบบจำลอง ได้แก่ แบบจำลองสุ่มป่าไม้ (Random Forest Model; RF) แบบจำลองเพื่อนบ้านใกล้สุด (K-Nearest Neighbors Model; KNN) แบบจำลองซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression Model; SVR) แบบจำลองโครงข่ายประสาทเทียมเพอร์เซฟตรอนแบบหลายชั้น (Multi-Layer Perceptron Neural Network Model; MLP) แบบจำลองโครงข่ายประสาทเทียมแบบความจำขนาดสั้นระยะยาว (Long Short Term Memory Neural Networks Model; LSTM) ใช้ชุดข้อมูลสำหรับการทดสอบประสิทธิภาพการทำนายของแบบจำลอง คือข้อมูลสถิติการสมัครเข้าเรียนปีการศึกษา 2566 พัฒนาชุดคำสั่งด้วยภาษาไพธอนร่วมกับไลบรารีการเรียนรู้ของเครื่องที่เกี่ยวข้อง โดยเขียนชุดคำสั่งบน colab ของ Google รายละเอียดชุดคำสั่งและขั้นตอนมีดังนี้

1) เรียกใช้ไลบรารีที่เกี่ยวข้องกับการวัดค่าประสิทธิภาพ

```
from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score
```

2) กำหนดชุดข้อมูลสำหรับการทดสอบ

```
X_test  = dataTest[['plan_n'],['pass_n']]
y_test  = dataTest.regis_n
```

3) ทำนายข้อมูลจำนวนผู้สมัครเข้าเรียนในหลักสูตรด้านคอมพิวเตอร์ทั้ง 5 แบบจำลอง

- กรณี RF Model

```
y_predict = RFmodel.predict(X_test)
```

- กรณี KNN Model

```
y_predict = KNNmodel.predict(X_test)
```

- กรณี SVR Model

```
y_predict = SVRmodel.predict(X_test)
```

- กรณี MLP Model

```
y_predict = MLPmodel.predict(X_test)
```

- กรณี LSTM Model

```
y_predict = LSTMmodel.predict(X_test)
```

4) วัดประสิทธิภาพผลการทำนายด้วยค่าประสิทธิภาพ MSE และ R^2

```
model_score = r2_score(y_test, y_predict)
model_mse   = mean_squared_error(y_test, y_predict)
print('R-Score of model = %.2f'%model_score)
print('MSE of model     = %.2f'%model_mse)
```

บทที่ 4
ผลการวิจัย

ในบทนี้ผู้วิจัยนำเสนอผลการดำเนินงานวิจัยเรื่องการทำนายจำนวนผู้สมัครเข้าศึกษาในหลักสูตรปริญญาตรีด้านคอมพิวเตอร์ในประเทศไทยโดยใช้ปัญญาประดิษฐ์ โดยมีวัตถุประสงค์เพื่อศึกษารูปแบบการทำนายจำนวนผู้สมัครเข้าศึกษาต่อในระดับปริญญาตรีด้านคอมพิวเตอร์ของไทยที่เหมาะสมในการนำไปใช้เป็นเครื่องมือสนับสนุนการตัดสินใจในการบริหารจัดการหลักสูตรด้านคอมพิวเตอร์ของสถาบันอุดมศึกษาในประเทศไทย ดำเนินการเก็บรวบรวมข้อมูลสถิติผู้สมัครเข้าเรียนจำนวน 6 ปีการศึกษาระหว่างปีการศึกษา พ.ศ. 2561-2566 จากระบบการคัดเลือกกลางบุคคลเข้าศึกษาในสถาบันอุดมศึกษา (Thai University Central Admission System; TCAS) โดยเข้าถึงผ่านลิงก์เว็บ https://www.mytcas.com/stat/ ผู้วิจัยดำเนินการจัดเตรียมข้อมูล และสร้างแบบจำลองการเรียนรู้ของเครื่องจำนวน 5 แบบจำลอง ประกอบด้วย แบบจำลองสุ่มป่าไม้ (Random Forest Model; RF) แบบจำลองเพื่อนบ้านใกล้สุด (K-Nearest Neighbors Model; KNN) แบบจำลองซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression Model; SVR) แบบจำลองโครงข่ายประสาทเทียมเพอร์เซฟตรอนแบบหลายชั้น (Multi-Layer Perceptron Neural Network; MLP) และแบบจำลองโครงข่ายประสาทเทียมแบบความจำขนาดสั้นระยะยาว (Long Short Term Memory Neural Networks; LSTM) รายละเอียดการทดสอบ และผลการทดสอบมีดังต่อไปนี้

4.1 การทดสอบ

งานวิจัยนี้ดำเนินการทดสอบประสิทธิภาพการทำนายจำนวนผู้สมัครเข้าศึกษาในหลักสูตรปริญญาตรีด้านคอมพิวเตอร์มีรายละเอียดดังนี้

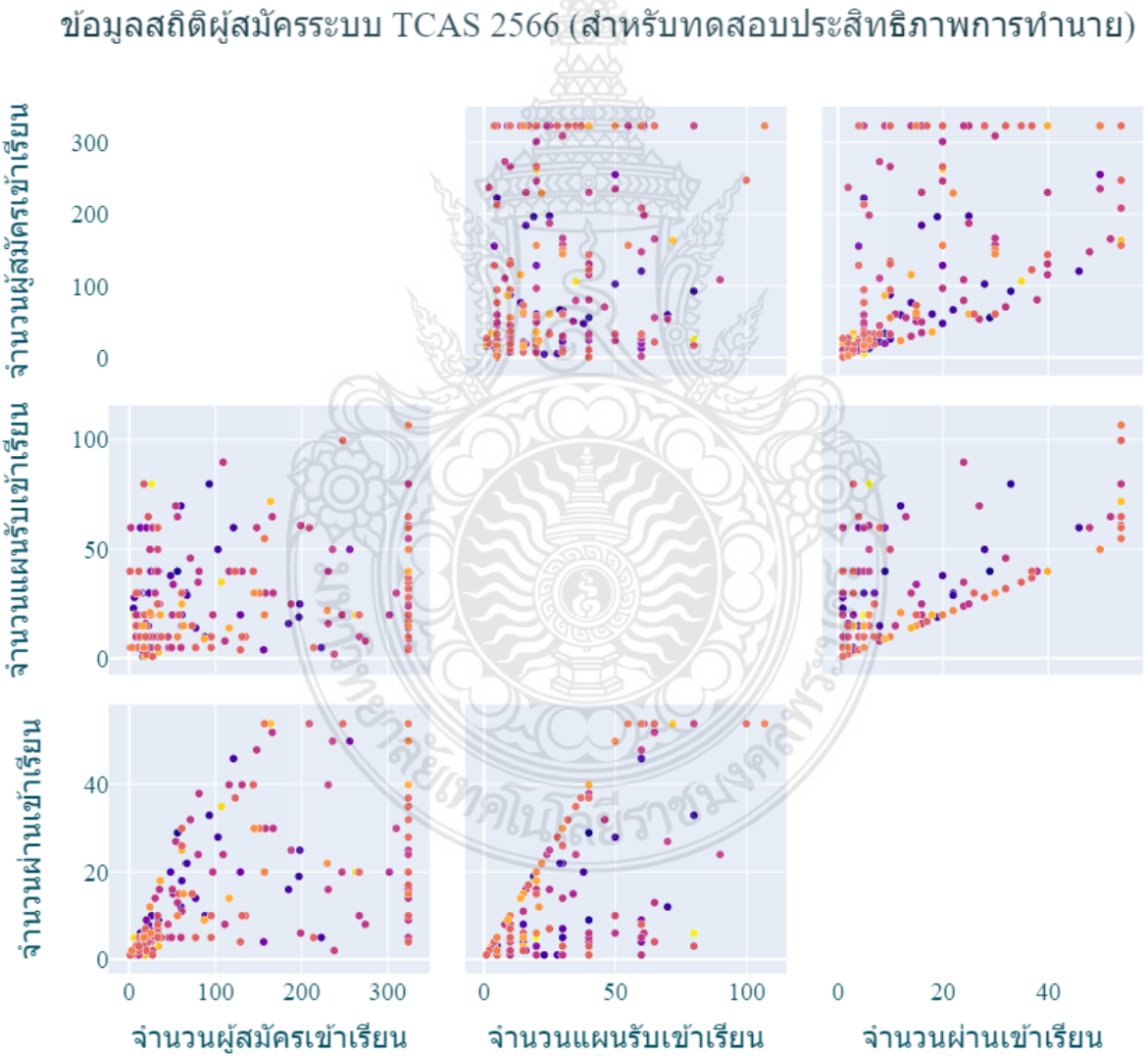
4.1.1 ข้อมูลสำหรับการทดสอบประสิทธิภาพการทำนายของแบบจำลองใช้ข้อมูลสถิติสมัครเข้าเรียนในหลักสูตรด้านคอมพิวเตอร์ ปีการศึกษา 2566 จำนวน 203 แถว 11 คอลัมน์ รายละเอียดสารสนเทศข้อมูลชุดทดสอบมีรายละเอียดดังนี้

```
dataTest
<class 'pandas.core.frame.DataFrame'>
Index: 203 entries, 710 to 912
Data columns (total 11 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   row_n          203 non-null    int64
 1   year           203 non-null    int64
 2   university_m   203 non-null    object
 3   faculty_m      203 non-null    object
 4   curriculumn_m  203 non-null    object
 5   curriculumn_c  203 non-null    object
 6   major_m        203 non-null    object
 7   ISCED_m        203 non-null    object
 8   regis_n        203 non-null    int64
 9   plan_n         203 non-null    int64
 10  pass_n         203 non-null    int64
dtypes: int64(5), object(6)
memory usage: 19.0+ KB
```

โดยข้อมูลค่าทางสถิติข้อมูลชุดทดสอบมีดังนี้

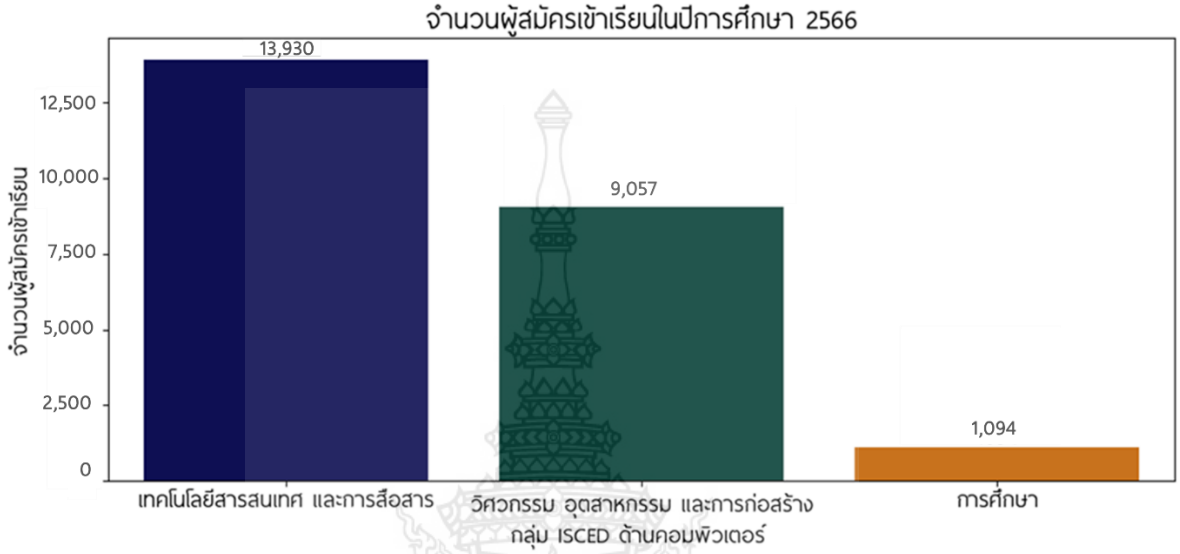
index	Unnamed: 0	row_n	year	regis_n	plan_n	pass_n
count	203.00	203.00	203.00	203.00	203.00	203.00
mean	811.00	812.00	2566.00	118.65	27.27	16.72
std	58.75	58.75	0.00	118.80	21.90	16.37
min	710.00	711.00	2566.00	1.00	1.00	1.00
25%	760.50	761.50	2566.00	20.00	10.00	5.00
50%	811.00	812.00	2566.00	61.00	20.00	10.00
75%	861.50	862.50	2566.00	218.50	40.00	25.00
max	912.00	913.00	2566.00	324.00	107.00	54.00

การกระจายของข้อมูลจำนวนผู้สมัครเข้าเรียน (regis_n) จำนวนแผนรับเข้าเรียน (plan_n) และจำนวนผู้ผ่านเข้าเรียน (pass_n) นำเสนอด้วยแผนภูมิ Scatter ดังภาพที่ 4-1



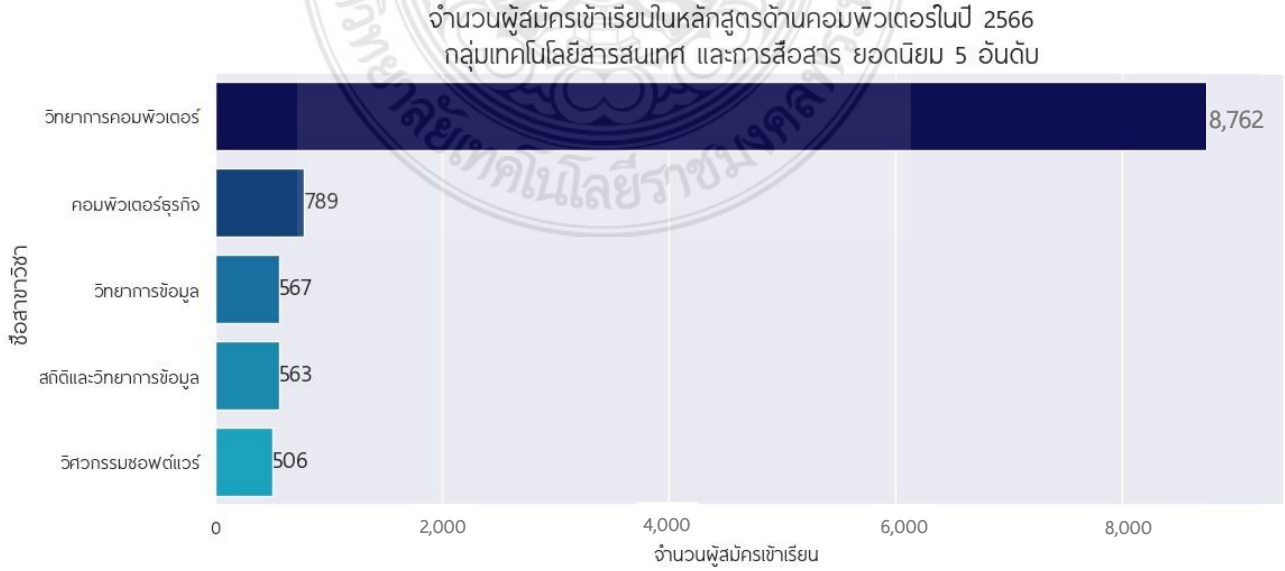
ภาพที่ 4-1 แผนภูมิ Scatter นำเสนอการกระจายข้อมูลชุดทดสอบประสิทธิภาพการพยากรณ์ของแบบจำลอง

จำนวนผู้สมัครเข้าเรียนในปีการศึกษา 2566 ในหลักสูตรด้านคอมพิวเตอร์ตามขอบเขตและเป้าหมายของงานวิจัยนี้มีจำนวนทั้งสิ้น 24,081 ราย โดยแยกตามกลุ่มมาตรฐาน ISCED ออกเป็น 3 กลุ่ม ประกอบด้วย กลุ่มการศึกษามีจำนวนทั้งสิ้น 1,094 ราย กลุ่มวิศวกรรม อุตสาหกรรม และการก่อสร้าง จำนวนทั้งสิ้น 9,057 ราย และกลุ่มเทคโนโลยีสารสนเทศและการสื่อสารจำนวน 13,930 ราย รายละเอียดจำนวนผู้สมัครเข้าเรียนปีการศึกษา 2566 ในหลักสูตรด้านคอมพิวเตอร์นำเสนอดังภาพที่ 4-2



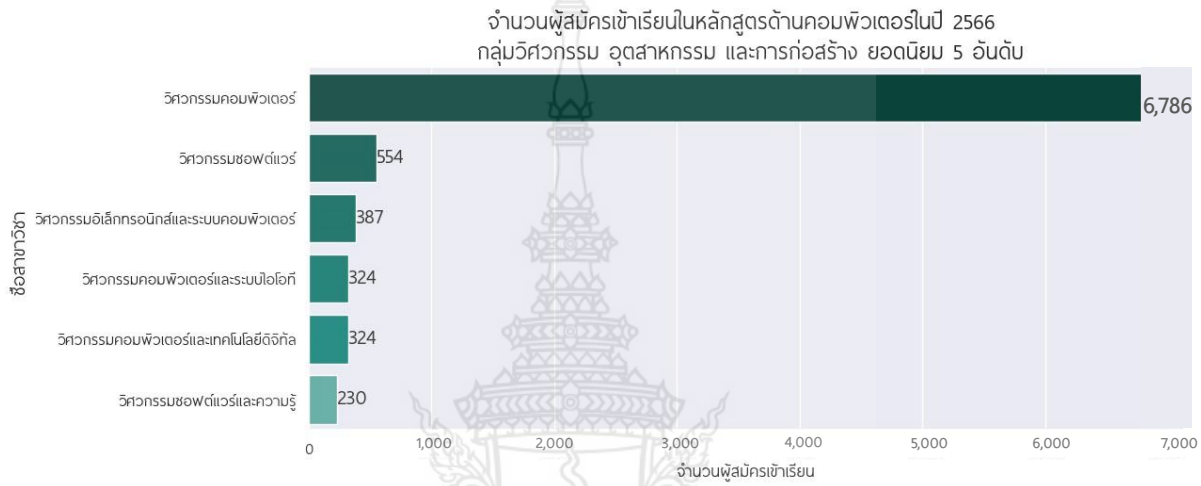
ภาพที่ 4-2 แผนภูมิแท่งเปรียบเทียบจำนวนผู้สมัครเข้าเรียนปี 2566 ด้านคอมพิวเตอร์ในแต่ละกลุ่ม ISCED

เมื่อพิจารณาข้อมูลผู้สมัครเข้าเรียนกลุ่มเทคโนโลยีสารสนเทศและการสื่อสารในแต่ละหลักสูตรสาขาวิชา พบว่าหลักสูตรสาขาวิชาที่มีผู้สมัครเข้าเรียนมากที่สุดในปีการศึกษา 2566 ใน 5 อันดับหลักสูตรสาขาวิชาที่มีผู้สมัครมากสุด คือ อันดับที่ 1 สาขาวิชาวิทยาการคอมพิวเตอร์จำนวนรวมทั้งสิ้น 8,762 ราย อันดับที่ 2 สาขาวิชาคอมพิวเตอร์ธุรกิจ จำนวน 789 ราย อันดับที่ 3 สาขาวิชาวิทยาการข้อมูล จำนวน 567 ราย อันดับที่ 4 สาขาวิชาสถิติและวิทยาการข้อมูล จำนวน 563 ราย และสาขาวิชาวิศวกรรมซอฟต์แวร์ (วท.บ.) จำนวน 506 ราย รายละเอียดดังแผนภูมิแท่งภาพที่ 4-3



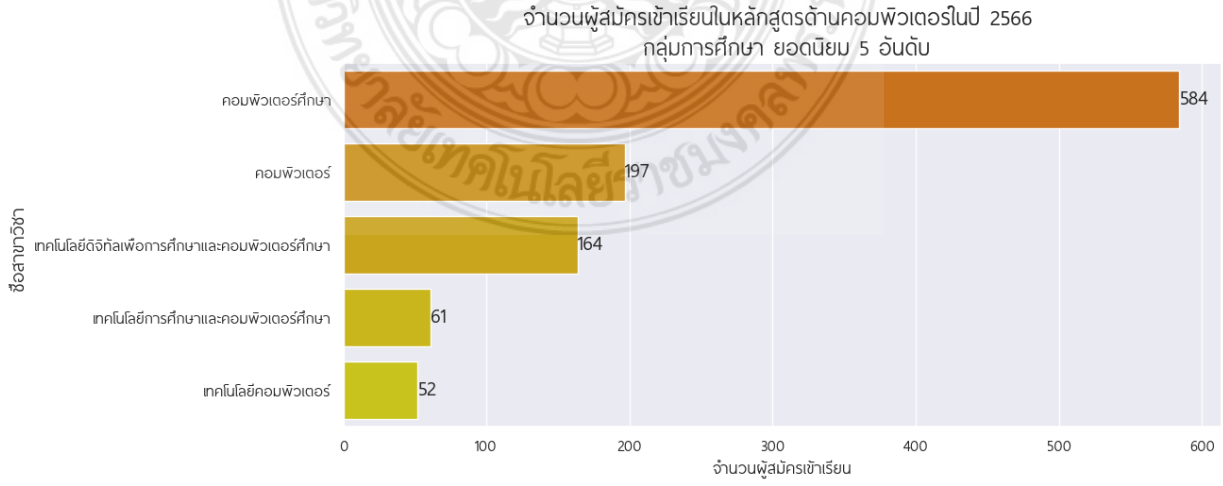
ภาพที่ 4-3 แผนภูมิแท่งเปรียบเทียบจำนวนผู้สมัครเข้าเรียนกลุ่มเทคโนโลยีสารสนเทศฯ ของแต่ละสาขาในปี 2566

สำหรับ ISCED ในกลุ่ม วิศวกรรม อุตสาหกรรม และการก่อสร้าง ที่เปิดการเรียนการสอนในหลักสูตรด้านคอมพิวเตอร์ ในปีการศึกษา 2566 มีผู้สนใจสมัครเข้าเรียนในสาขาวิชามากสุด 5 อันดับ ดังนี้ อันดับที่ 1 สาขาวิชาวิศวกรรมคอมพิวเตอร์มีจำนวนทั้งสิ้น 6,786 ราย อันดับที่ 2 สาขาวิชาวิศวกรรมซอฟต์แวร์ (วศ.บ.) จำนวน 554 ราย อันดับที่ 3 สาขาวิชาวิศวกรรมอิเล็กทรอนิกส์และระบบคอมพิวเตอร์ จำนวน 387 ราย อันดับที่ 4 มี 2 หลักสูตรสาขาวิชา คือ สาขาวิชาวิศวกรรมคอมพิวเตอร์และระบบไอโอที และสาขาวิชาวิศวกรรมคอมพิวเตอร์และเทคโนโลยีดิจิทัล จำนวน 324 ราย และอันดับที่ 5 สาขาวิชาวิศวกรรมซอฟต์แวร์และความรู้มีจำนวน 230 ราย รายละเอียดนำเสนอด้วยแผนภูมิแท่งดังภาพที่ 4-4



ภาพที่ 4-4 แผนภูมิแท่งเปรียบเทียบจำนวนผู้สมัครเข้าเรียนกลุ่มวิศวกรรมศาสตร์ฯ ของแต่ละสาขาในปี 2566

เมื่อพิจารณาข้อมูลผู้สมัครเข้าเรียน ISCED กลุ่มการศึกษา พบว่า 5 อันดับสาขาวิชาที่มีผู้สนใจสมัครเข้าเรียนมากสุดดังนี้ อันดับที่ 1 สาขาวิชาคอมพิวเตอร์ศึกษามีจำนวน 584 ราย อันดับที่ 2 สาขาวิชาคอมพิวเตอร์ จำนวน 197 และเทคโนโลยีดิจิทัลเพื่อการศึกษาและคอมพิวเตอร์ศึกษา สาขาวิชาเทคโนโลยีการศึกษาและคอมพิวเตอร์ศึกษา เทคโนโลยีคอมพิวเตอร์ มีจำนวนผู้สมัคร 387, 324, 324 และ 230 ตามลำดับ รายละเอียดดังภาพที่ 4-5



ภาพที่ 4-5 แผนภูมิแท่งเปรียบเทียบจำนวนผู้สมัครเข้าเรียนกลุ่มการศึกษาในแต่ละสาขาในปี 2566

4.1.2 แบบจำลองการเรียนรู้ของเครื่องจำนวน 5 แบบจำลอง ที่ได้ดำเนินการเรียนรู้จากชุดข้อมูลชุดฝึกสอน คือ สถิติข้อมูลการสมัครเข้าเรียนจำนวน 5 ปีการศึกษา ระหว่างปีการศึกษา พ.ศ. 2561-2565 โดยโครงสร้างสถาปัตยกรรมของแบบจำลองทั้ง 5 แบบจำลอง มีดังนี้

1) แบบจำลองสุ่มป่าไม้ (RF)

```
RandomForestRegressor
RandomForestRegressor(max_depth=90, max_features=3, min_samples_leaf=3,
                      min_samples_split=10, random_state=1)
```

ภาพที่ 4-6 สถาปัตยกรรมแบบจำลอง RF

2) แบบจำลองเพื่อนบ้านใกล้สุด (KNN)

```
KNeighborsRegressor
KNeighborsRegressor(leaf_size=50, metric='manhattan', n_neighbors=45, p=5)
```

ภาพที่ 4-7 สถาปัตยกรรมแบบจำลอง KNN

3) แบบจำลองซัพพอร์ตเวกเตอร์รีเกรสชัน (SVR)

```
SVR
SVR(C=0.1, gamma=1, kernel='linear')
```

ภาพที่ 4-8 สถาปัตยกรรมแบบจำลอง SVR

4) แบบจำลองโครงข่ายประสาทเทียมเพอร์เซฟตรอนแบบหลายชั้น (MLP)

```
Pipeline
Pipeline(steps=[('standardscaler', StandardScaler()),
                ('mlpregressor',
                 MLPRegressor(activation='tanh', alpha=0.05,
                              hidden_layer_sizes=(100, 100), max_iter=100,
                              random_state=2))])
StandardScaler
StandardScaler()
MLPRegressor
MLPRegressor(activation='tanh', alpha=0.05, hidden_layer_sizes=(100, 100),
             max_iter=100, random_state=2)
```

ภาพที่ 4-9 สถาปัตยกรรมแบบจำลอง MLP

5) แบบจำลองโครงข่ายประสาทเทียมแบบความจำขนาดสั้นระยะยาว (LSTM)

Layer (type)	Output Shape	Param #
lstm_9 (LSTM)	(None, 710, 32)	4,608
dense_11 (Dense)	(None, 710, 100)	3,300
dense_12 (Dense)	(None, 710, 1)	101

```
Total params: 8,009 (31.29 KB)
Trainable params: 8,009 (31.29 KB)
Non-trainable params: 0 (0.00 B)
```

ภาพที่ 4-10 สถาปัตยกรรมแบบจำลอง LSTM

4.1.3 ดำเนินการทดสอบประสิทธิภาพการทำนายข้อมูลผู้สมัครเข้าเรียนด้วยชุดข้อมูลสำหรับทดสอบคือข้อมูลสถิติผู้สมัครเข้าเรียนปีการศึกษา 2566 ตามรายละเอียดข้างต้น วัดประสิทธิภาพผลการทำนายข้อมูลด้วยค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared Error; MSE) และค่าสัมประสิทธิ์การทำนาย (Coefficient of Determination, R^2)

4.2 ผลการทดสอบ

ผลการทดสอบประสิทธิภาพการทำนายข้อมูลจำนวนผู้สมัครเข้าเรียนของแบบจำลองทั้ง 5 แบบจำลอง ประกอบด้วย แบบจำลองสุ่มป่าไม้ (RF) แบบจำลองเพื่อนบ้านใกล้สุด (KNN) แบบจำลองซัพพอร์ตเวกเตอร์รีเกรสชัน (SVR) แบบจำลองโครงข่ายประสาทเทียมเพอร์เซฟตรอนแบบหลายชั้น (MLP) และแบบจำลองโครงข่ายประสาทเทียมแบบความจำขนาดสั้นระยะยาว (LSTM) กับชุดข้อมูลทดสอบจำนวน 913 แถว ด้วยการวัดประสิทธิภาพผลการทำนายจากค่าคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) และค่าสัมประสิทธิ์การทำนาย (Coefficient of Determination; R^2) ดำเนินการเขียนชุดคำสั่งด้วยภาษาไพธอนร่วมกับไลบรารีจัดการข้อมูล และไลบรารีการเรียนรู้ของเครื่อง พัฒนาชุดคำสั่งบนเครื่องมือ colab รุ่น Pro ของ Google

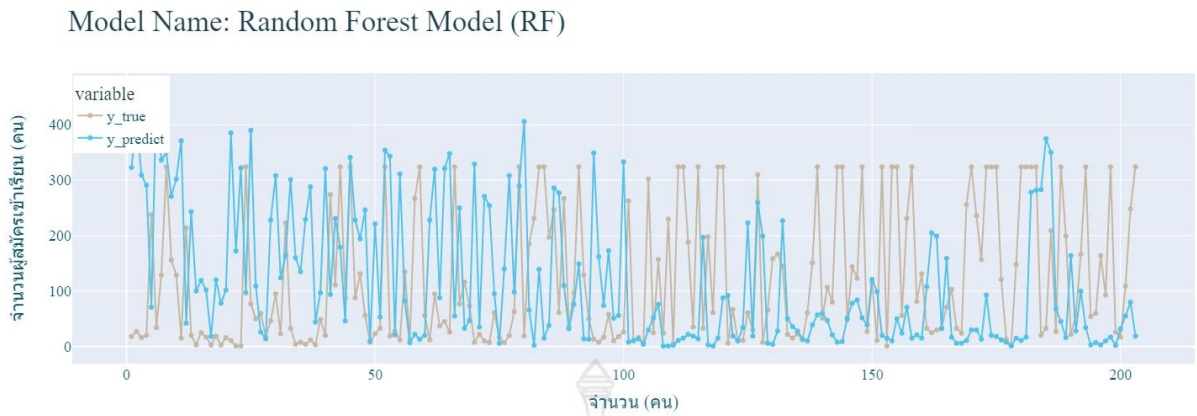
ในปีการศึกษา 2566 มีจำนวนผู้สมัครเข้าเรียนในหลักสูตรด้านคอมพิวเตอร์ตามขอบเขตของงานวิจัย จำนวน 24,081 ราย จากผลการทดสอบประสิทธิภาพการทำนายของแบบจำลองทั้ง 5 แบบจำลอง พบว่า ***แบบจำลองสุ่มป่าไม้ (RF) มีประสิทธิภาพในการทำนายที่ดีสุด***ในการทำนายจำนวนผู้สมัครเข้าเรียนในหลักสูตรด้านคอมพิวเตอร์ในปีการศึกษา 2566 รองลงมา คือ แบบจำลองเพื่อนบ้านใกล้สุด (KNN) และแบบจำลองซัพพอร์ตเวกเตอร์รีเกรสชัน (SVR) ตามลำดับ รายละเอียดข้อมูลผลการทำนายดังนี้

- แบบจำลองสุ่มป่าไม้ (RF) มีประสิทธิภาพการทำนายที่ดีสุด ทำนายจำนวนผู้สมัครเข้าเรียนในหลักสูตรด้านคอมพิวเตอร์ในปี 2566 จำนวน 23,798 ราย ที่ค่าสัมประสิทธิ์การทำนายร้อยละ 93 และมีค่าคลาดเคลื่อนกำลังสองเฉลี่ยจากการทำนายจำนวน 944 ราย
- แบบจำลองเพื่อนบ้านใกล้สุด (KNN) มีประสิทธิภาพการทำนายอันดับที่ 2 ทำนายจำนวนผู้สมัครเข้าเรียนจำนวน 20,568 ราย ที่ค่าสัมประสิทธิ์การทำนายร้อยละ 90 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย 1,414 ราย
- แบบจำลองซัพพอร์ตเวกเตอร์รีเกรสชัน (SVR) มีประสิทธิภาพการทำนายอันดับที่ 3 โดยทำนายจำนวนผู้สมัครเข้าเรียน 21,376 ราย ที่ค่าสัมประสิทธิ์การทำนายร้อยละ 89 มีค่าคลาดเคลื่อนกำลังสองเฉลี่ย 1,614 ราย
- แบบจำลองโครงข่ายประสาทเทียมแบบความจำขนาดสั้นระยะยาว (LSTM) มีประสิทธิภาพการทำนายอันดับที่ 4 ทำนายจำนวนผู้สมัครเข้าเรียน 22,923 ราย ที่ค่าสัมประสิทธิ์การทำนายร้อยละ 88 และมีค่าคลาดเคลื่อนกำลังสองเฉลี่ย 1,688 ราย
- แบบจำลองโครงข่ายประสาทเทียมเพอร์เซฟตรอนแบบหลายชั้น (MLP) มีประสิทธิภาพการทำนายต่ำสุด ทำนายจำนวนผู้สมัครเข้าเรียน 17,258 ที่ค่าสัมประสิทธิ์การทำนายร้อยละ 82 มีค่าคลาดเคลื่อนกำลังสองเฉลี่ย 2,495 ราย

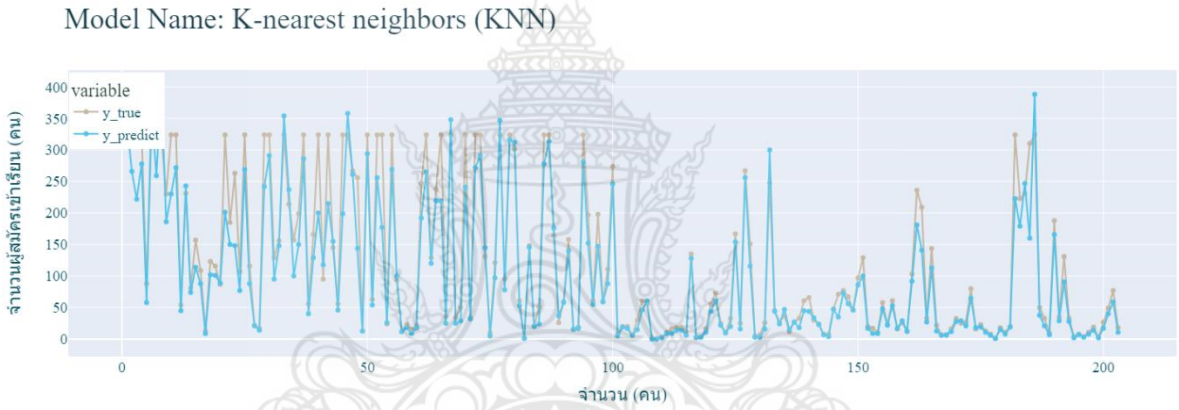
รายละเอียดการเปรียบเทียบค่าประสิทธิภาพการทำนายของแต่ละแบบจำลองแสดงดังตารางที่ 4-1 และการนำเสนอผลลัพธ์การทำนายแสดงดังภาพที่ 4-11 ถึงภาพที่ 4-15 ตามลำดับ

ตารางที่ 4-1 รายละเอียดผลการวัดประสิทธิภาพการทำนายข้อมูลของแบบจำลองทั้ง 5 แบบจำลอง

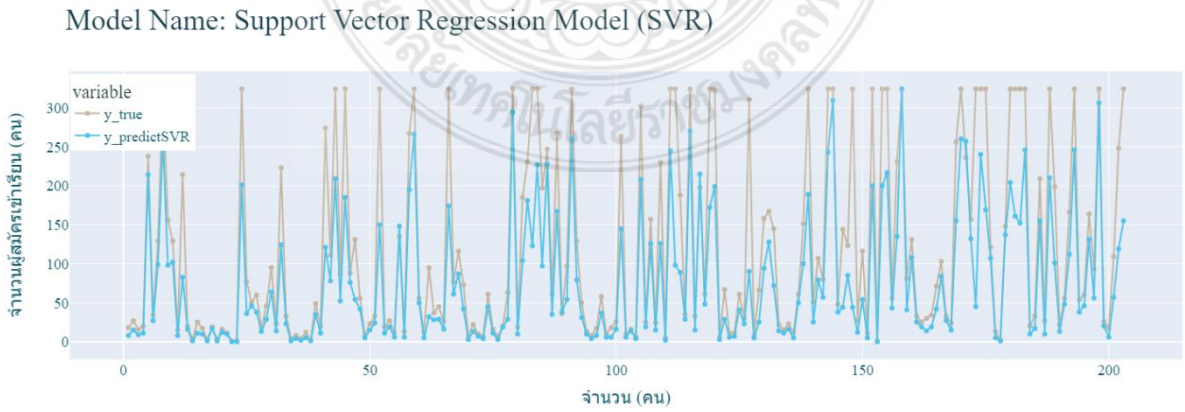
แบบจำลอง	MSE	R^2
1. แบบจำลองสุ่มป่าไม้ (RF)	944.83	93%
2. แบบจำลองเพื่อนบ้านใกล้สุด (KNN)	1,414.84	90%
3. แบบจำลองซัพพอร์ตเวกเตอร์รีเกรสชัน (SVR)	1,614.39	89%
4. แบบจำลองโครงข่ายประสาทเทียมเพอร์เซฟตรอนแบบหลายชั้น (MLP)	2,495.81	82%
5. แบบจำลองโครงข่ายประสาทเทียมแบบความจำขนาดสั้นระยะยาว (LSTM)	1,688.17	88%



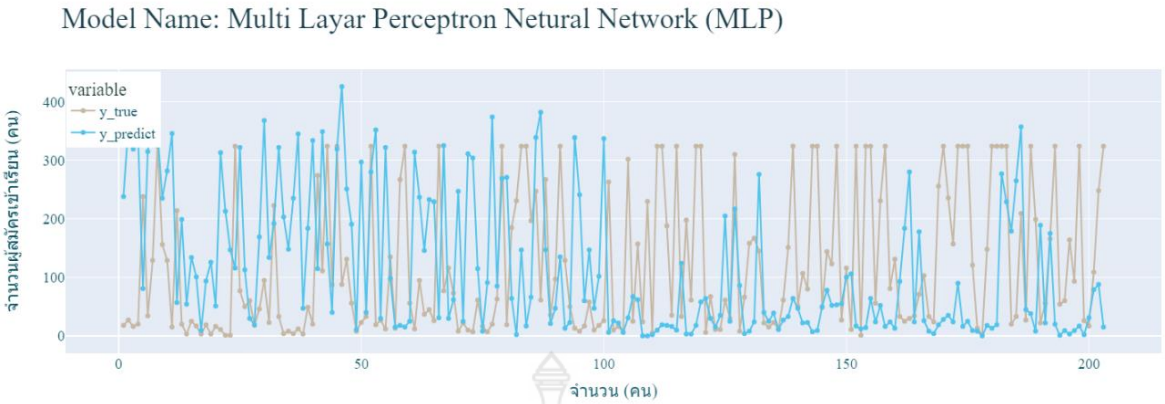
ภาพที่ 4-11 แผนภูมิเส้นผลลัพธ์การทำนายข้อมูลของแบบจำลองสุ่มป่าไม้ (RF)



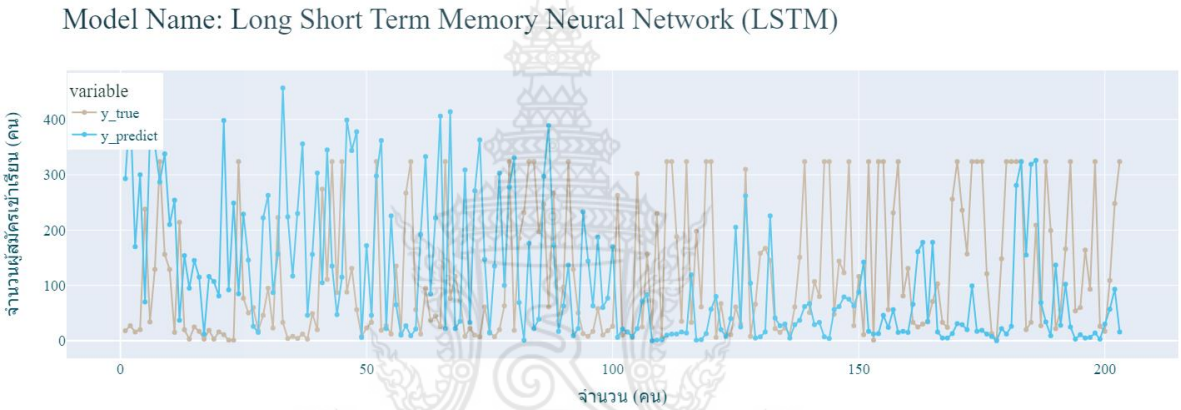
ภาพที่ 4-12 แผนภูมิเส้นผลลัพธ์การทำนายข้อมูลของแบบจำลองเพื่อนบ้านใกล้สุด (KNN)



ภาพที่ 4-13 แผนภูมิเส้นผลลัพธ์การทำนายข้อมูลของแบบจำลองซัพพอร์ตเวกเตอร์รีเกรสชัน (SVR)

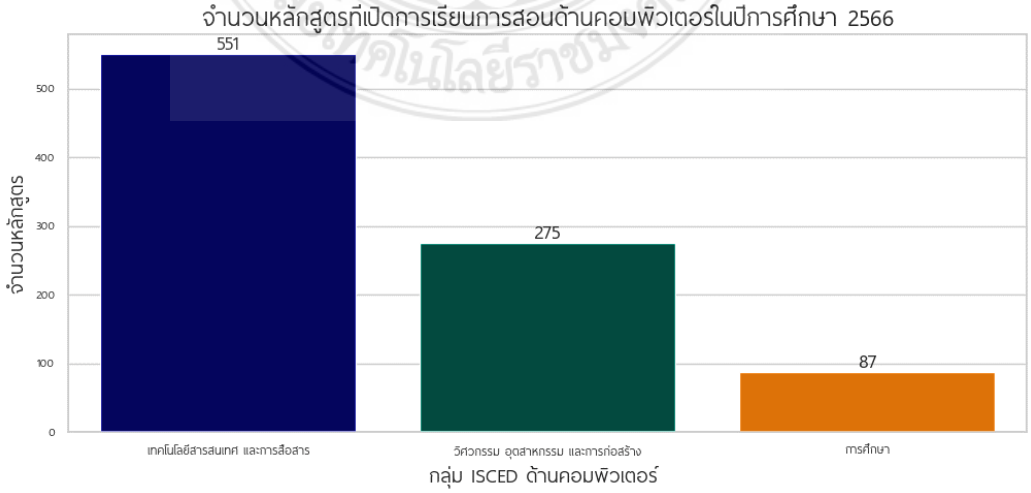


ภาพที่ 4-14 แผนภูมิเส้นผลลัพธ์การทำนายข้อมูลของโครงข่ายประสาทเทียมเพอร์เซฟตรอนแบบหลายชั้น (MLP)



ภาพที่ 4-15 แผนภูมิเส้นผลลัพธ์การทำนายข้อมูลของโครงข่ายประสาทเทียมแบบความจำขนาดสั้นระยะยาว (LSTM)

ผู้วิจัยได้พิจารณาข้อมูลจากการศึกษาข้อมูลสถิติสมัครเข้าเรียนในปี พ.ศ 2561-2566 หลังผ่านกระบวนการจัดเตรียมข้อมูลมีจำนวนทั้งสิ้น 913 รายการ โดยแยกตามกลุ่ม ISCED 3 กลุ่ม ประกอบด้วย กลุ่มเทคโนโลยีสารสนเทศ และการสื่อสาร กลุ่มวิศวกรรม อุตสาหกรรม และการก่อสร้าง และกลุ่มการศึกษา มีจำนวนรายการข้อมูล 551, 275, และ 87 รายการตามลำดับ โดยจำนวนรายการข้อมูลหมายถึงจำนวนหลักสูตรรายวิชาที่เปิดการเรียนการสอนที่เกี่ยวข้องกับคอมพิวเตอร์ รายละเอียดดังภาพที่ 4-16



ภาพที่ 4-16 แผนภูมิแท่งเปรียบเทียบจำนวนหลักสูตรที่เปิดการเรียนการสอนด้านคอมพิวเตอร์

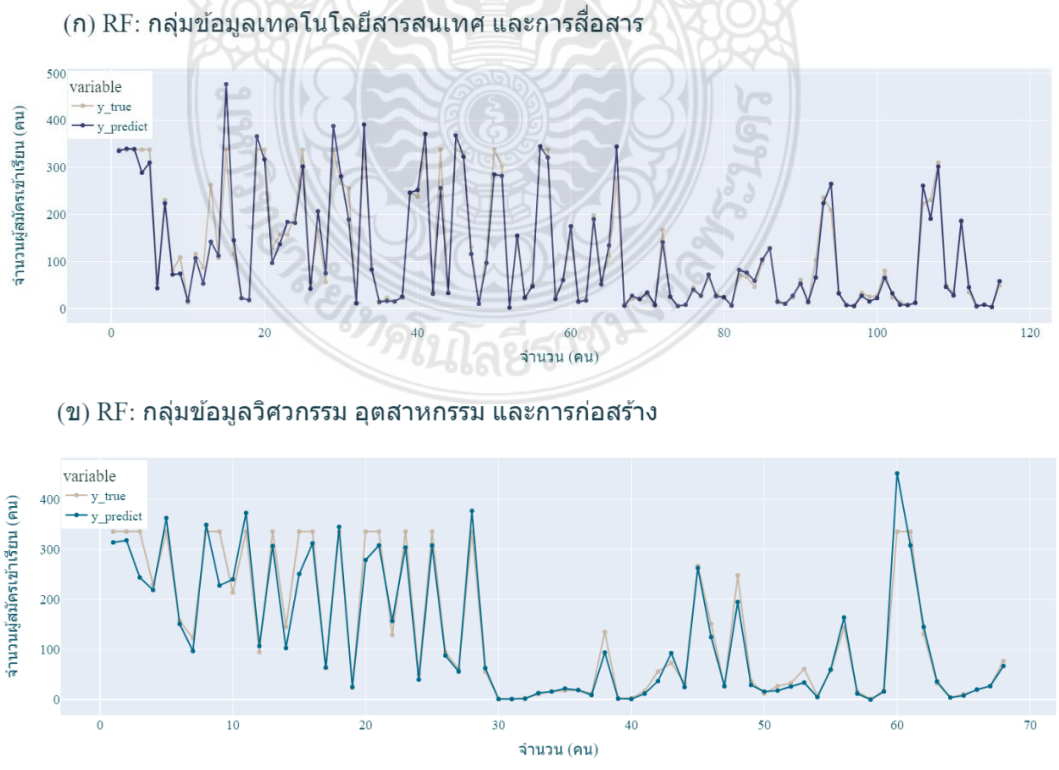
ผู้วิจัยดำเนินการสร้างแบบจำลองทำนายข้อมูลจำนวนผู้สมัครในกลุ่มมาตรฐาน ISCED 2 กลุ่ม คือ กลุ่มเทคโนโลยีสารสนเทศ และการสื่อสาร กลุ่มวิศวกรรม อุตสาหกรรม และการก่อสร้าง เพื่อเปรียบเทียบประสิทธิภาพผลการทำนายตามวิธีการและขั้นตอนข้างต้น สำหรับกลุ่มการศึกษาไม่ได้ถูกนำมาสร้างแบบจำลองเนื่องจากมีจำนวนข้อมูลน้อย ผลการวัดประสิทธิภาพการทำนายข้อมูลผู้สมัครเข้าเรียนรายกลุ่ม ISCED กลุ่ม มีรายละเอียดดังนี้

- กลุ่มเทคโนโลยีสารสนเทศ และการสื่อสาร พบว่า แบบจำลองสุ่มป่าไม้ (RF) มีประสิทธิภาพในการทำนายที่ดีสุด ทำนายจำนวนผู้สมัครเข้าเรียนในหลักสูตรด้านคอมพิวเตอร์ที่จัดอยู่ในกลุ่มเทคโนโลยีสารสนเทศจำนวน 13,943 ราย โดยมีค่าสัมประสิทธิ์การทำนายร้อยละ 95 ที่ค่าคลาดเคลื่อนกำลังสองเฉลี่ย 751 ราย
- กลุ่มวิศวกรรม อุตสาหกรรม และการก่อสร้าง พบว่า แบบจำลองที่มีประสิทธิภาพที่ดีที่สุดคือแบบจำลองสุ่มป่าไม้ (RF) เช่นกัน ทำนายจำนวนผู้สมัครเข้าเรียนจำนวน 8,724 ราย มีค่าสัมประสิทธิ์การทำนายร้อยละ 94 และมีค่าคลาดเคลื่อนกำลังสองเฉลี่ย 1,151 ราย ตามลำดับ

รายละเอียดการวัดประสิทธิภาพการทำนายข้อมูลจำนวนผู้สมัครเข้าเรียนด้านคอมพิวเตอร์ ดังตารางที่ 4-2 และภาพที่ 4-17 ตามลำดับ

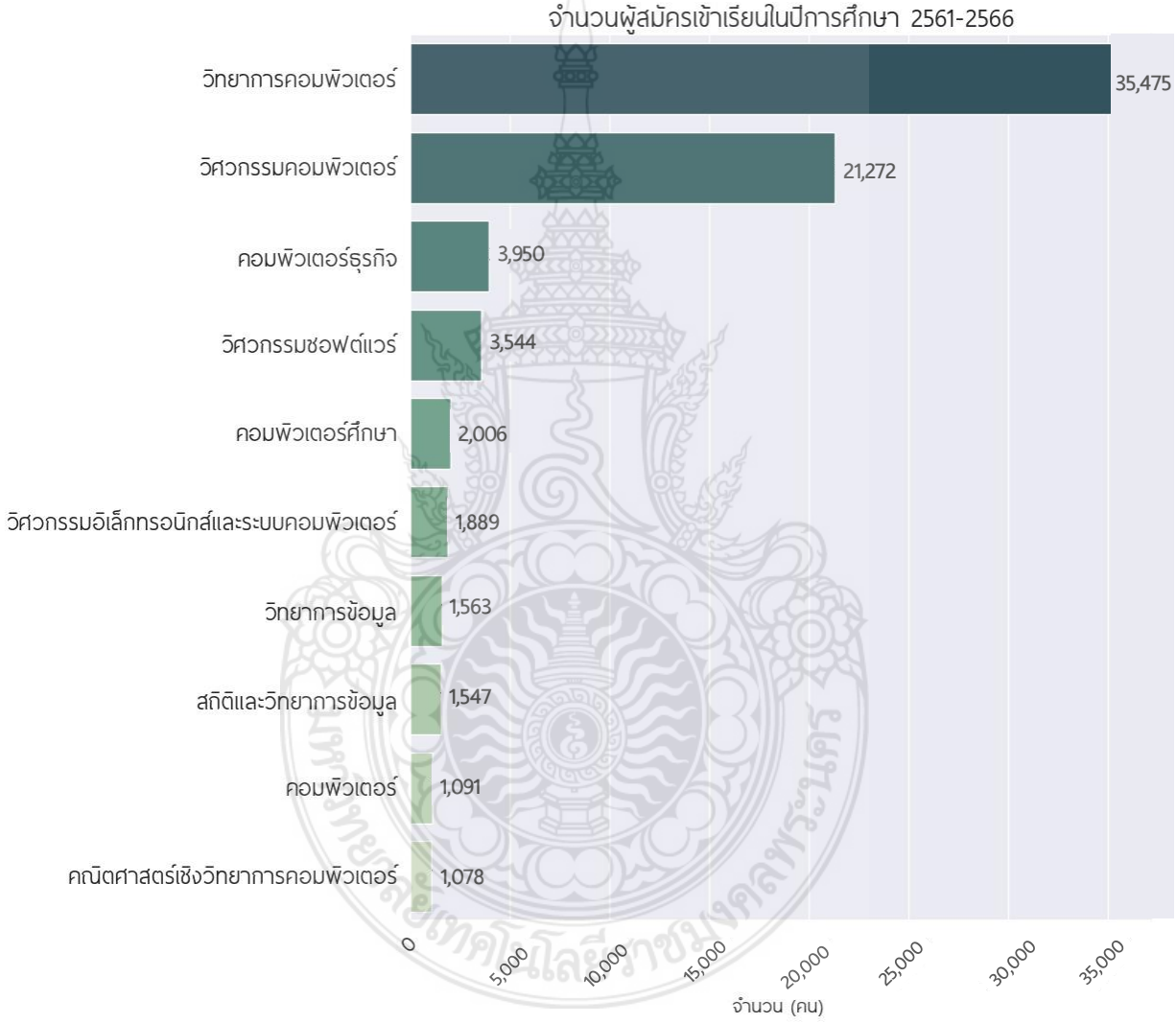
ตารางที่ 4-2 ประสิทธิภาพการทำนายข้อมูลจำนวนผู้สมัครเข้าเรียนด้านคอมพิวเตอร์แยกตามกลุ่ม ISCED

กลุ่มมาตรฐาน ISCED	y-true	y-predict	MSE	R^2
1. เทคโนโลยีสารสนเทศ และการสื่อสาร	14,145	13,943	751.52	95%
2. วิศวกรรม อุตสาหกรรม และการก่อสร้าง	9,295	8,724	1,151.37	94%



ภาพที่ 4-17 แผนภูมิเส้นผลลัพธ์การทำนายข้อมูลผู้สมัครแยกตามกลุ่ม ISCED ของแบบจำลองที่ดีสุด

นอกจากนั้นผู้วิจัยได้ดำเนินการพิจารณาข้อมูลจำนวนผู้สมัครเข้าเรียนใน 10 อันดับสาขาวิชาด้านคอมพิวเตอร์ จากข้อมูลสถิติการสมัครเข้าเรียน 6 ปีการศึกษา ระหว่างปี พ.ศ.2561-2566 พบว่า หลักสูตรวิทยาศาสตรบัณฑิต (วท.บ.) สาขาวิชาวิทยาการคอมพิวเตอร์เป็นสาขาวิชายอดนิยมอันดับที่ 1 มีจำนวนผู้สมัครเข้าเรียนรวมทั้งสิ้น 35,475 ราย หรือมีจำนวนผู้สมัครเฉลี่ยต่อปี 5,913 ราย สาขายอดนิยมอันดับที่ 2 คือ หลักสูตรวิศวกรรมศาสตรบัณฑิต (วศ.บ.) สาขาวิชาวิศวกรรมคอมพิวเตอร์ มีจำนวนผู้สมัครรวมทั้งสิ้น 21,272 ราย มีจำนวนผู้สมัครเข้าเรียนเฉลี่ยรายปี 3,545 ราย และสาขายอดนิยมอันดับที่ 3 คือ หลักสูตรบริหารธุรกิจบัณฑิต (บธ.บ.) สาขาวิชาคอมพิวเตอร์ธุรกิจ มีจำนวนผู้สมัครรวม 3,950 ราย มีจำนวนผู้สมัครเฉลี่ยรายปี 658 ราย รายละเอียดการนำเสนอข้อมูลสาขายอดนิยมที่มีจำนวนผู้สมัครเข้าเรียน 10 อันดับ แสดงดังภาพที่ 4-18



ภาพที่ 4-18 แผนภูมิแท่งเปรียบเทียบจำนวนผู้สมัครเข้าเรียนมากสุด 10 อันดับในสาขาวิชาด้านคอมพิวเตอร์

จากข้อมูลการจัดอันดับสาขาด้านคอมพิวเตอร์ยอดนิยมที่มีผู้สมัครเข้าเรียนสูงสุด 10 อันดับข้างต้น ผู้วิจัยได้นำข้อมูลในแต่ละสาขาวิชามาดำเนินการสร้างแบบจำลองทำนายจำนวนผู้สมัครเข้าเรียนใน 3 อันดับสาขาวิชา คือ สาขาวิชาวิทยาการคอมพิวเตอร์ สาขาวิชาวิศวกรรมคอมพิวเตอร์ และสาขาวิชาคอมพิวเตอร์ธุรกิจ โดยกำหนดให้ข้อมูลสถิติจำนวนผู้สมัครปี 2561-2565 เป็นข้อมูลสำหรับการฝึกสอนเพื่อการสร้างแบบจำลอง และใช้ข้อมูลสถิติจำนวนผู้สมัครปี พ.ศ. 2566 เป็นข้อมูลสำหรับการทดสอบประสิทธิภาพผลการทำนายข้อมูลของแบบจำลอง

ผลทดสอบพบว่า ***แบบจำลองที่มีประสิทธิภาพผลการทำนายที่ดีสุดยังคงเป็นแบบจำลองจากอัลกอริทึมสุ่มป่าไม้ (Random Forest Algorithm; RF)*** โดยแต่ละสาขาวิชามีค่าสัมประสิทธิ์การทำนาย และค่าคลาดเคลื่อนกำลังสองเฉลี่ย ดังนี้

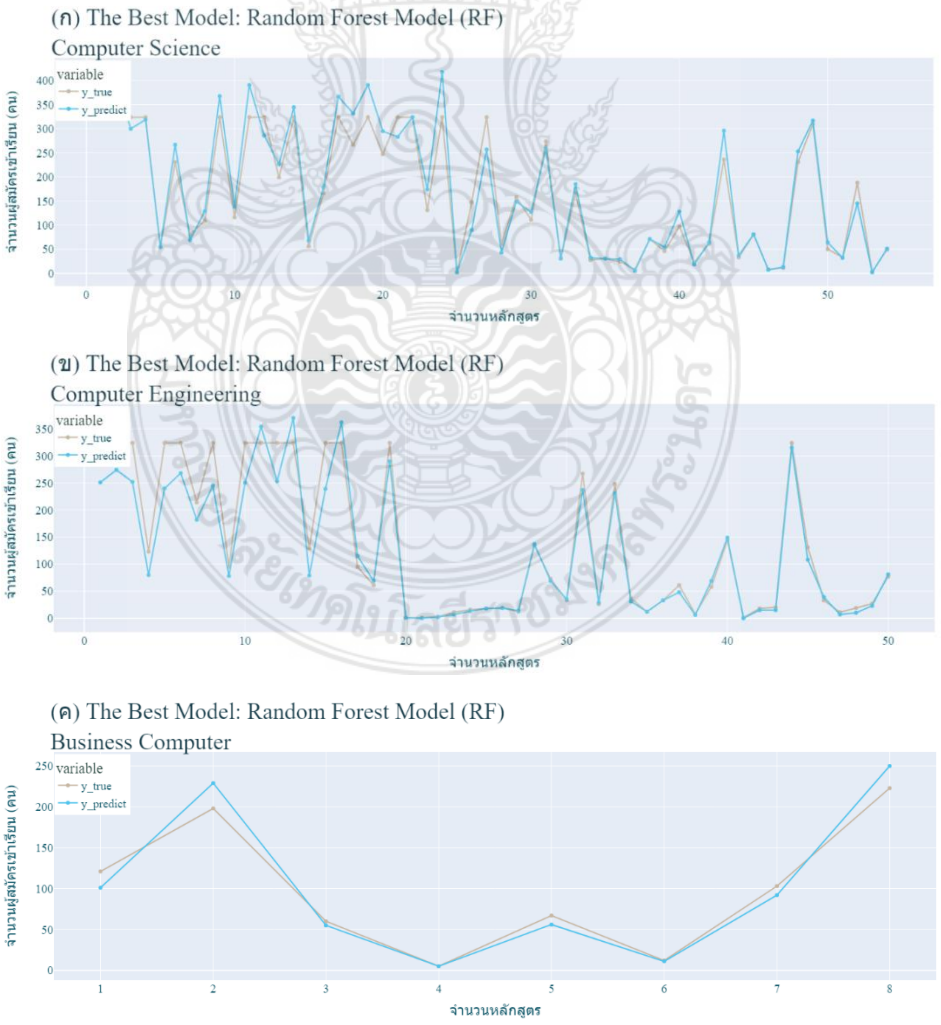
- สาขาวิชาวิทยาการคอมพิวเตอร์มีค่าสัมประสิทธิ์การทำนายร้อยละ 92 ที่ค่าคลาดเคลื่อนกำลังสองเฉลี่ย 1,144 ราย
- สาขาวิชาวิศวกรรมคอมพิวเตอร์มีค่าสัมประสิทธิ์การทำนายร้อยละ 93 ที่ค่าคลาดเคลื่อนกำลังสองเฉลี่ย 1,210 ราย
- สาขาวิชาคอมพิวเตอร์ธุรกิจมีค่าสัมประสิทธิ์การทำนายร้อยละ 95 ที่คลาดเคลื่อนกำลังสองเฉลี่ย 299 ราย

รายละเอียดผลการทดสอบดังตารางที่ 4-3

ตารางที่ 4-3 ประสิทธิภาพการทำนายข้อมูลจำนวนผู้สมัครเข้าเรียนด้านคอมพิวเตอร์ในสาขายอดนิยม 3 อันดับ

สาขาวิชา	y-true	y-predict	MSE	R^2
1. วิทยาการคอมพิวเตอร์	8,762	9,354	1,144.15	92%
2. วิศวกรรมคอมพิวเตอร์	6,806	6,024	1,210.89	93%
3. คอมพิวเตอร์ธุรกิจ	789	799	299.36	95%

รายละเอียดผลลัพธ์ประสิทธิภาพในการทำนายข้อมูลจำนวนผู้สมัครเข้าเรียนแสดงดังภาพที่ 4-19



ภาพที่ 4-19 แผนภูมิเส้นผลลัพธ์การทำนายจำนวนผู้สมัครเรียนในสาขายอดนิยม 3 อันดับของแบบจำลองสุ่มป่าไม้

จากผลการทดสอบประสิทธิภาพแบบจำลองทำนายข้อมูลจำนวนผู้สมัครเข้าเรียน โดยใช้ข้อมูลสถิติการสมัครเข้าเรียน 5 ปีการศึกษา คือ ระหว่างปี พ.ศ. 2561-2565 จากระบบการคัดเลือกกลางบุคคลเข้าศึกษาในสถาบันอุดมศึกษา (Thai University Central Admission System; TCAS) เป็นข้อมูลสำหรับฝึกสอนเพื่อให้ได้แบบจำลองทำนายข้อมูลจำนวน 5 แบบจำลอง ประกอบด้วย แบบจำลองสุ่มป่าไม้ (Random Forest Model; RF) แบบจำลองเพื่อนบ้านใกล้สุด (K-Nearest Neighbors Model; KNN) แบบจำลองซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression Model; SVR) แบบจำลองโครงข่ายประสาทเทียมเพอร์เซฟตรอนแบบหลายชั้น (Multi Leyer Perceptron Neural Network; MLP) และแบบจำลองโครงข่ายประสาทเทียมแบบความจำขนาดสั้นระยะยาว (Long Short Term Memory Neural Networks; LSTM) และใช้ข้อมูลสถิติสมัครเข้าเรียนปีการศึกษา 2566 เป็นข้อมูลสำหรับการทดสอบประสิทธิภาพการทำนายข้อมูลของแบบจำลองทั้ง 5 โดยทดสอบข้อมูลในมิติภาพรวมทั้งหมด มิติข้อมูลรายกลุ่มมาตรฐาน ISCED จำนวน 2 กลุ่ม และมิติข้อมูลรายสาขาวิชา โดยเลือกสาขาวิชายอดนิยมจำนวน 3 อันดับ ผลการทดสอบ พบว่า แบบจำลองสุ่มป่าไม้ (RF) มีประสิทธิภาพในการทำนายดีที่สุด รองลงมา คือ แบบจำลองเพื่อนบ้านใกล้สุด (KNN) และแบบจำลองซัพพอร์ตเวกเตอร์รีเกรสชัน (SVR) โดยมีค่าสัมประสิทธิ์การทำนาย (R^2) อยู่ในช่วงร้อยละ 92 ถึงร้อยละ 95

บทที่ 5
สรุป อภิปรายผล และข้อเสนอแนะ

5.1 สรุป อภิปรายผล

รายงานการวิจัยฉบับนี้นำเสนอการสร้างแบบจำลองการเรียนรู้ของเครื่อง (Machine Learning Model) และเปรียบเทียบประสิทธิภาพผลการทำนายข้อมูลจำนวนผู้สมัครเข้าเรียนในหลักสูตรด้านคอมพิวเตอร์ของประเทศไทย โดยใช้ศาสตร์ในศตวรรษที่ 21 ด้านปัญญาประดิษฐ์ (AI) มาบูรณาการกับศาสตร์ด้านวิทยาการข้อมูล (Data Science) โดยมีจุดมุ่งหวังให้ได้เครื่องมือสำหรับช่วยตัดสินใจในการบริหารจัดการหลักสูตรสาขาวิชาที่เกี่ยวข้องกับศาสตร์ด้านคอมพิวเตอร์ งานวิจัยนี้เลือกใช้แบบจำลองการเรียนรู้ของเครื่องจำนวน 5 แบบจำลอง ประกอบด้วย แบบจำลองสุ่มป่าไม้ (Random Forest; RF) แบบจำลองเพื่อนบ้านใกล้สุด (K-Nearest Neighbors; KNN) แบบจำลองซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression; SVR) แบบจำลองโครงข่ายประสาทเทียมเพอร์เซฟตรอนแบบหลายชั้น (Multi Leyer Perceptron Neural Network; MLP) และแบบจำลองโครงข่ายประสาทเทียมแบบความจำขนาดสั้นระยะยาว (Long Short Term Memory Neural Networks; LSTM) ดำเนินการวิจัยตามหลักการในศาสตร์ด้านวิทยาการข้อมูล ประกอบด้วย 4 ขั้นตอนหลัก คือ (1) การจัดเก็บและรวบรวมข้อมูล (Data Acquisition)งานวิจัยนี้ดำเนินการจัดเก็บข้อมูลสถิติสมัครเข้าเรียนจำนวน 6 ปีการศึกษา กล่าวคือ ปีการศึกษา 2561 ถึง 2566 จากระบบการคัดเลือกกลางบุคคลเข้าศึกษาในสถาบันอุดมศึกษา(TCAS) เข้าถึงได้จากลิงก์เว็บ https://www.mytcas.com/stat/ (2) การจัดเตรียมข้อมูล (Data Preparation) ในขั้นตอนนี้เริ่มต้นจากนำข้อมูลที่จัดเก็บและรวบรวมได้ถูกนำมาทำความสะอาดข้อมูล (Data Cleaning) สร้างมิติข้อมูลใหม่ (Feature Engineering) ลดมิติข้อมูลที่ไม่เกี่ยวข้อง (Dimensionality Reduction) รวมข้อมูลเข้าด้วยกัน (Aggregate Data) ระบุมิติข้อมูลที่เกี่ยวข้อง (Feature Selection) จัดการกับข้อมูลส่วน Outlier และดำเนินการแบ่งกลุ่มข้อมูล โดยแบ่งข้อมูลสถิติสมัครเข้าเรียนปีการศึกษา 2561-2565 เป็นข้อมูลชุดฝึกสอนในการเรียนรู้ของเครื่อง (Training Data) และข้อมูลสถิติสมัครเข้าเรียน ปีการศึกษา 2566 เป็นข้อมูลชุดทดสอบประสิทธิภาพการทำนายข้อมูลของแบบจำลอง (Testing Data) (3) การสร้างแบบจำลอง (Models Creation) ด้วยการนำข้อมูลชุดฝึกสอนการเรียนรู้ของเครื่องที่ได้แบ่งไว้มาฝึกสอนกับแบบจำลองทั้ง 5 ข้างต้น โดยงานวิจัยนี้ดำเนินการหาพารามิเตอร์ที่เหมาะสม (Hyperparameter Optimization) ของแต่ละแบบจำลองโดยใช้วิธีการค้นหาแบบกริด (Grid Search Method) ขั้นตอนสุดท้าย (4) การทดสอบแบบจำลอง (Models Testing) แบบจำลองการเรียนรู้ของเครื่องทั้ง 5 แบบจำลองที่ได้จากขั้นตอนการสร้างแบบจำลองถูกนำมาทดสอบประสิทธิภาพการทำนายจำนวนผู้สมัครเข้าเรียนกับชุดข้อมูลทดสอบ งานวิจัยนี้ใช้ค่าประสิทธิภาพ 2 ค่า คือ ค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared Error; MSE) และค่าสัมประสิทธิ์การทำนาย (Coefficient of Determination; R^2) หรือค่าความเชื่อมั่นในการทำนายข้อมูล งานวิจัยนี้ดำเนินการทดสอบประสิทธิภาพการทำนายข้อมูลจำนวนผู้สมัครเข้าเรียนในหลักสูตรด้านคอมพิวเตอร์ที่ครอบคลุมมิติของข้อมูลที่หลากหลาย คือ ทำนายในภาพรวม ทำนายแยกตามประเภทของกลุ่มมาตรฐานการศึกษา ISCED และทำนายผลแยกตามสาขาวิชายอดนิยมที่มีผู้สมัครเข้าเรียนสูงสุด 3 อันดับ

งานวิจัยนี้ใช้เครื่องมือสำหรับพัฒนาชุดคำสั่งต่าง ๆ ด้วยภาษาไพธอน ร่วมกับไลบรารีจัดการข้อมูล คือ Pandas Library และ NumPy Library ไลบรารีการเรียนรู้ของเครื่อง ประกอบด้วย Scikit Learn, TensorFlow และ Keas ไลบรารีสำหรับแสดงผลข้อมูล ประกอบด้วย Matplotlib, Seaborn และ Plotly พัฒนาชุดคำสั่งบนเครื่องมือ colab รุ่น Pro ของ Google ด้วยหน่วยประมวลผล Intel ® ความเร็ว 2.20 GHz. ที่หน่วยความจำ (Ram) 1.25GB./12.67GB.

ผลการทดสอบประสิทธิภาพการทำนายจำนวนผู้สมัครเข้าเรียนในหลักสูตรด้านคอมพิวเตอร์ในภาพรวมปีการศึกษา 2566 พบว่า แบบจำลองสุ่มป่าไม้ (RF) มีประสิทธิภาพในการทำนายที่ดีที่สุดในทุกมิติของการทดสอบ โดยทำนายจำนวนผู้สมัครเข้าเรียนจำนวน 23,798 ราย ที่ค่าสัมประสิทธิ์การทำนาย (R^2) เท่ากับร้อยละ 93 และมีค่าคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) จากการทำนาย 944 ราย ทำนายจำนวนผู้สมัครเข้าเรียนในกลุ่มมาตรฐานการศึกษา ISCED กลุ่มเทคโนโลยีสารสนเทศ และการสื่อสารจำนวน 13,943 ราย มีค่าสัมประสิทธิ์การทำนาย ร้อยละ 95 ที่ค่าคลาดเคลื่อนกำลังสองเฉลี่ย 751 ราย ทำนายจำนวนผู้สมัครเข้าเรียนในกลุ่มวิศวกรรม อุตสาหกรรม และการก่อสร้างจำนวน 8,724 ราย ที่ค่าสัมประสิทธิ์การทำนายร้อยละ 94 มีค่าคลาดเคลื่อนกำลังสองเฉลี่ย 1,151 ราย สำหรับผลการทำนายรายสาขาวิชา พบว่า แบบจำลองสุ่มป่าไม้ทำนายจำนวนผู้สมัครเข้าเรียนในสาขาวิชาวิทยาการคอมพิวเตอร์ซึ่งเป็นสาขาวิชาด้านคอมพิวเตอร์ที่มีจำนวนผู้สมัครมากที่สุดจำนวน 9,354 ราย มีค่าสัมประสิทธิ์การทำนายร้อยละ 92 ในสาขาวิชาวิศวกรรมคอมพิวเตอร์ทำนายจำนวน 6,024 ราย ที่ค่าสัมประสิทธิ์การทำนายร้อยละ 93 และในสาขาวิชาคอมพิวเตอร์ธุรกิจทำนายจำนวนผู้สมัคร 799 ราย ที่ค่าสัมประสิทธิ์การทำนายร้อยละ 95 สำหรับแบบจำลองที่มีค่าประสิทธิภาพอันดับรองลงมาคือ แบบจำลองเพื่อนบ้านใกล้สุด (KNN) และแบบจำลองซัพพอร์ตเวกเตอร์รีเกรสชัน ตามลำดับ โดยประสิทธิภาพการทำนายข้อมูลด้วยแบบจำลองโครงข่ายประสาทเทียมเพอร์เซฟตรอนแบบหลายชั้น (MLP) มีค่าประสิทธิภาพน้อยสุด

สำหรับปัญหาและอุปสรรคที่พบในการดำเนินงานวิจัยในครั้งนี้พบว่า ข้อมูลสถิติการสมัครเข้าเรียนทั้ง 6 ปีการศึกษาจากระบบการคัดเลือกกลางบุคคลเข้าศึกษาในสถาบันอุดมศึกษา (Thai University Central Admission System; TCAS) มีการจัดรูปแบบ และมิติข้อมูลที่แตกต่างกันมาก เนื่องจากข้อมูลถูกจัดเก็บตามเงื่อนไขการรับเข้าเรียนในแต่ละปี ทำให้ต้องใช้เวลาในการจัดเตรียมข้อมูลมาก นอกจากนั้นยังพบว่าจำนวนผู้สมัคร จำนวนแผนรับ และจำนวนผู้ผ่านเข้าเรียนมีข้อมูลส่วน Outlier จำนวนมากเช่นกัน ซึ่งในงานวิจัยนี้เลือกวิธีสำหรับปรับปรุงข้อมูล Outlier เป็นค่ากลาง (Mean Value) ด้วยการพิจารณาจากขอบข้อมูลในควอนไทน์ล่างและควอนไทน์ระดับบน นอกจากนั้นในขั้นตอนการหาพารามิเตอร์ที่เหมาะสม (Hyperparameter Optimization) ของแต่ละแบบจำลอง งานวิจัยเลือกใช้วิธีพื้นฐานคือการค้นหาแบบกริด (Grid Search Method) ซึ่งอาจได้ค่าพารามิเตอร์ยังไม่เหมาะสม จากปัญหาและอุปสรรคที่กล่าวข้างต้นส่งผลต่อประสิทธิภาพการทำนายข้อมูลของแบบจำลอง ซึ่งต้องใช้เวลาในการศึกษาค้นคว้าต่อไป

5.2 ข้อเสนอแนะ

จากปัญหาและอุปสรรคในการดำเนินงานวิจัยตามบทสรุปข้างต้น ผู้วิจัยมีข้อเสนอแนะในการดำเนินงานปรับปรุงแก้ไขที่สำคัญดังนี้

1) ดำเนินการศึกษาเกณฑ์การให้คะแนนการรับสมัครของระบบ TCAS ในแต่ละปี เพื่อนำมาใช้ในการปรับสเกลข้อมูลในมิติคะแนนต่ำสุด มิติคะแนนสูงสุด และมิติคะแนนเฉลี่ย ซึ่งในการดำเนินงานวิจัยในครั้งนี้ได้ตัดมิติข้อมูลเหล่านี้ออกเนื่องจากข้อจำกัดด้านระยะเวลาในการดำเนินงานวิจัยที่ต้องใช้ในการศึกษาเกณฑ์คะแนนในแต่ละปีการศึกษา

2) ศึกษาวิธีการตรวจจับข้อมูลผิดปกติ หรือ ข้อมูลส่วน Outlier ด้วยวิธีการที่หลากหลาย และดำเนินการปรับปรุงข้อมูลเหล่านั้นอย่างเหมาะสม

3) ศึกษาวิธีการหาค่าพารามิเตอร์ที่เหมาะสม (Hyperparameter Optimization) อาทิ วิธีการ Random Search และวิธีการปรับค่าพารามิเตอร์แบบอัตโนมัติ (Automated Hyperparameter Tuning) เป็นต้น

4) เพิ่มส่วนการนำเสนอข้อมูลในรูปแบบอินเตอร์แอคทีฟแดชบอร์ด (Interactive Dashboard) เพื่อนำเสนอข้อมูลภาพรวมการจำนวนผู้สมัคร จำนวนแผนรับ และจำนวนผู้ผ่านการคัดเลือกเข้าเรียน เพื่อใช้เป็นประโยชน์กับนักเรียนที่ต้องการสมัครเข้าเรียนต่อในระดับอุดมศึกษาผ่านระบบ TCAS

5) เพิ่มส่วนการวิเคราะห์ข้อมูล และทำนายข้อมูลรายสถาบันการศึกษา เพื่อใช้ประโยชน์ในการบริหารหลักสูตรขององค์กรแบบเฉพาะ เนื่องจากผลการวิเคราะห์และทำนายข้อมูลในงานวิจัยอยู่ในภาพรวมระดับประเทศ

บรรณานุกรม

ภาษาไทย

นิติกร จันหาญ และจารี ทองคำ. (2565) การเปรียบเทียบประสิทธิภาพของเทคนิคอริมาและเทคนิคถดถอยการเรียนรู้เครื่องในการพยากรณ์ราคาบิตคอนยน์. วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยอุบลราชธานี, 24(2), 62-75.

ปริญญา สงวนสัตย์. (2562). Artificial Intelligence with Machine Learning AI สร้างด้วยแมชชีนเลิร์นนิ่ง Python Edition. บริษัทไอดีซี พรีเมียร์ จำกัด.

พีรภัทร วัสแสง, กองกูณฑ์ โตชัยวัฒน์. (2567). การพยากรณ์ราคาขายทรัพย์สินรอการขายประเภทบ้านเดี่ยว ในพื้นที่กรุงเทพมหานครโดยใช้เทคนิคการเรียนรู้ของเครื่อง. วารสารนวัตกรรมการศึกษาและการวิจัย, 8(1), 423–439. https://doi.org/10.14456/jeir.2024.26

ภวิตา จันทันแก้ว และ นวลวรรณ สุนทรภิษัช. (2565). การวิเคราะห์ข้อมูลเพื่อพยากรณ์ราคาข้าวโพดเลี้ยงสัตว์ด้วยขั้นตอนวิธีถดถอย. วารสารวิจัย มข., 23(2), 92-106.

รณชัย ชื่นธวัช และคณะ. (2560). การพยากรณ์ความต้องการใช้หน่วยจำหน่ายไฟฟ้าด้วยแบบจำลองซัพพอร์ทเวกเตอร์รีเกรสชันแบบตรวจสอบสลับ 3 ส่วน. วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยอุบลราชธานี, 19(1), 215-232.

วิชญ พัชรเจริญวงษ์ และคณะ. (2563). แบบจำลองการทำนายระยะเวลาในการเข้าเทียบท่าของเรือโดยสารสาธารณะ. วารสารวิทยาศาสตร์ลาดกระบัง, 29(2), 31-44.

สำนักงานส่งเสริมเศรษฐกิจดิจิทัล. (ม.ป.ป.). เทคโนโลยีที่สำคัญในยุคดิจิทัล: เทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence: AI). [บทความออนไลน์]. สืบค้นจาก https://www.depa.or.th/th/article-view/tech-series-artificial-intelligence-ai

สมาคมที่ประชุมอธิการบดีแห่งประเทศไทย. สถิติการสอบเข้าระบบการคัดเลือกกลางบุคคลเข้าศึกษาในสถาบันอุดมศึกษา. [ออนไลน์]. สืบค้นจาก https://www.mytcas.com/stat/

อมรินทร์ทีวีออนไลน์. (2023). 4 AI ตัวเทพจากค่ายดัง Microsoft - Google - Meta - OpenAI เก่งอะไรบ้าง. [ข่าวออนไลน์]. สืบค้นจาก https://www.amarintv.com/spotlight/ technology/detail/57039

ภาษาอังกฤษ

Ihsncnkz. (2022). RNN and LSTM with Keras. [online]. https://www.kaggle.com/code/ihsncnkz/rnn-and-lstm-with-keras

Kamil Krzyk. (2023). *Coding Deep Learning for Beginners - Types of Machine Learning.* [Image]. Retrieved from https://cdn-images-1.medium.com/max/1000/1*8wU0hfUY3UK_D8Y7tblyFQ.png

Mark E. Fenner. (2020). Machine Learning with Python for Everyone. Pearson Education, Inc.

Oinkina. (2015). Understanding LSTM Networks. [Blog online]. colah's blog. https://colah.github.io/posts/2015-08-Understanding-LSTMs/

OpenAI. (2024). *ChatGPT* (May 13 version) [Large language model]. https://chat.openai.com/chat

Scikit Learn. (2024). Nearest Neighbors regression. [Online]. https://scikit-learn.org/stable/auto_examples/neighbors/plot_regression.html#sphx-glr-auto-examples-neighbors-plot-regression-py

_____. (2024). Support Vector Regression (SVR) using linear and non-linear kernels. https://scikit-learn.org/stable/auto_examples/svm/plot_svm_regression.html #sphx-glr-auto-examples-svm-plot-svm-regression-py

_____. (2024). Advanced plotting with partial dependence. (n.d.). Scikit-learn. https://scikit-learn.org/stable/auto_examples/miscellaneous/plot_partial_dependence_visualization_api.html.

Vivian Siahaan, Rismon H. Sianipar. (2023). Amazon Stock Price: Visualization, Forecasting, and Prediction using Machine Learning with Python. 2nd Edition. Balige Publishing.

ประวัติผู้วิจัย

ผศ.ดร.วีรวรรณ จันทนะทรัพย์
Asst.Prof.Dr.Veerawan Juntanasub

อาจารย์ประจำสาขาวิชาวิทยาการคอมพิวเตอร์
อาจารย์ผู้รับผิดชอบหลักสูตร วท.บ. วิทยาการข้อมูลและเทคโนโลยีสารสนเทศ
อาจารย์ผู้สอนหลักสูตร วท.บ. สาขาวิชาวิทยาการคอมพิวเตอร์
อาจารย์ผู้สอนหลักสูตร วท.บ. สาขาวิชาวิทยาการข้อมูลและเทคโนโลยีสารสนเทศ

Contact

(+66) 0-2836-3000 ต่อ 4211
veerawan.j@rmutp.ac.th
https://sci.rmutp.ac.th
คณะวิทยาศาสตร์และเทคโนโลยี
มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร

ตำแหน่งงาน : ข้าราชการพลเรือนในสถาบันอุดมศึกษา
จำนวนปีที่ทำงาน : 25 ปี

Education

- **2015** Ph.D. (Information Technology) King Mongkut's University of Technology North Bangkok, Thailand.
- **2005** M.S.Tech.Ed. (Computer Technology) King Mongkut's University of Technology North Bangkok, Thailand.
- **1997** B.B.A (Business Computer) Sripatum University, Thailand.

Language

Thai English

Awards

Jan 2024 | RMUTP
บุคคลดีเด่นด้านการจัดการองค์ความรู้แห่งปี
เจ้าของผลงานวิจัยชนะเลิศด้านวิทยาศาสตร์เทคโนโลยีฯ

Jan 2023 | RMUTP
บุคคลดีเด่นด้านการจัดการองค์ความรู้แห่งปี

PROFILE

ผศ.ดร.วีรวรรณ จันทนะทรัพย์ อาจารย์ประจำสาขาวิชาวิทยาการคอมพิวเตอร์ และเป็นอาจารย์ผู้รับผิดชอบหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาวิทยาการข้อมูลและเทคโนโลยีสารสนเทศ มีประสบการสอนกว่า 25 ปี มีความเชี่ยวชาญในศาสตร์ด้านการประมวลผลภาพ การนำข้อมูลด้วยภาพ และอื่น ๆ

ปฏิบัติหน้าที่ความรับผิดชอบในหลักสูตร:

- ประชุม ร่วมบริหารหลักสูตรทั้งในระดับหลักสูตร คณะ และมหาวิทยาลัย
- ดูแลประสานงานกลุ่มการเขียนโปรแกรม และแสดงผลข้อมูลด้วยภาพ
- การกำกับ ติดตามและประเมินคุณภาพในการจัดการเรียนการสอน
- การประชาสัมพันธ์หลักสูตรและแนวแนะทางการศึกษา
- จัดทำสถิติประเมินผล และจัดทำเอกสาร
- อาจารย์ที่ปรึกษานักศึกษา
- อาจารย์ที่ปรึกษาโครงงานนักศึกษา
- อาจารย์นิเทศสหกิจศึกษา

EXPERTISE

- การประมวลผลภาพดิจิทัล (Digital Image Processing)
- การคำนวณเชิงวิวัฒนาการ (Evolutionary Computation)
- การนำเสนอข้อมูลด้วยภาพ (Data Visulization)
- การทำเหมืองข้อมูล (Data mining)
- ปฏิสัมพันธ์ระหว่างมนุษย์กับคอมพิวเตอร์ (HCI)
- เทคโนโลยีอากาศยานไร้คนขับ (UAV Technology)

COURSE

- การเขียนโปรแกรมสำหรับวิทยาการข้อมูล
- ปฏิสัมพันธ์ระหว่างมนุษย์กับคอมพิวเตอร์
- การวิเคราะห์ข้อมูลภาพดิจิทัล
- การนำเสนอข้อมูลด้วยภาพ

ORCID
http://orcid.org
0000-0002-9474-0573

RESEARCH / PAPERS / BOOK

2 H-INDEX GOOGLE SCHOLAR CITATIONS
1 H-INDEX SCOPUS CITATIONS

2023

วีรวรรณ จันทนะทรัพย์, นริศรา นาคเมธี และสยาม ลางคุลเสน. (2566). การแสดงข้อมูลด้วยภาพสำหรับวัดไทยในจังหวัดนนทบุรี. การประชุมวิชาการวิศวกรรมศาสตร์และเทคโนโลยี มทร.พระนคร ครั้งที่ 7. คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร. (น.208-213), 19 พฤษภาคม 2566

นริศรา นาคเมธี, ชาญวิทย์ ปราบพยัคย์, สยาม ลางคุลเสน และวีรวรรณ จันทนะทรัพย์. (2566). สหกิจศึกษาบนแพลตฟอร์มออนไลน์. การประชุมวิชาการวิศวกรรมศาสตร์และเทคโนโลยี มทร.พระนคร ครั้งที่ 7. คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร. (น.262-267), 19 พฤษภาคม 2566.

2022

วีรวรรณ จันทนะทรัพย์. การเขียนโปรแกรมคอมพิวเตอร์ด้วยภาษาไพธอน ฉบับปรับปรุงใหม่. พิมพ์ครั้งที่ 2. กรุงเทพฯ : บริษัทแดเน็กซ์อินเตอร์คอร์ปอเรชั่นจำกัด. 2565. 360 หน้า.

วีรวรรณ จันทนะทรัพย์ และสุรีโย ก่อวานิชกุล. (2565). การเพิ่มประสิทธิภาพงานทรัพยากรบุคคลด้วยแดชบอร์ด เทคโนโลยี. รายงานสืบเนื่องการประชุมวิชาการระดับชาติด้านวิทยาศาสตร์ เทคโนโลยี และนวัตกรรม ครั้งที่ 5. คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร. (น. 692-701), 19 สิงหาคม 2565.

2021

ศุภธีนี สุวรรณพงศ์, ชนากานต์ ครูวิชา และวีรวรรณ จันทนะทรัพย์. (2564). สื่อการเรียนรู้วิธีใช้งานเครื่องออกกำลังกายสาธารณะสำหรับผู้สูงอายุด้วยเทคโนโลยีความจริงเสริม.รายงานสืบเนื่องการประชุมวิชาการระดับชาติด้านวิทยาศาสตร์ เทคโนโลยี และนวัตกรรม ครั้งที่ 4 "วิทยาศาสตร์ เทคโนโลยี และนวัตกรรมสร้างสรรค์ เพื่อก้าวผ่านสถานการณ์ COVID-19". คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร. (น.135-142), 30 สิงหาคม 2564.

คณะวิทยาศาสตร์และเทคโนโลยี
มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร (ศูนย์พระนครเหนือ)
เลขที่ 1381 ถนนประชาราษฎร์ 1 แขวงวงศ์สว่าง เขตบางซื่อ กรุงเทพ 10800
โทรศัพท์ : 02-836-3000 ต่อ 4194 ,4155
อีเมล : infosci@rmutp.ac.th

มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร
เลขที่ 399 ถ.สามเสน แขวงวชิรพยาบาล เขตดุสิต กรุงเทพฯ 10300
โทรศัพท์ : 02-665-3777, 02-665-3888
โทรสาร : 02-665-3722
One Stop Service ต่อ 6636
อีเมล : oci@rmutp.ac.th

Contact

(+66) 0-2836-3000 ต่อ 4211

veerawan.j@rmutp.ac.th

https://sci.rmutp.ac.th

คณะวิทยาศาสตร์และเทคโนโลยี
มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร

ตำแหน่งงาน : ข้าราชการพลเรือน ในสถาบันอุดมศึกษา

จำนวนปีที่ทำงาน : 25 ปี

Education

- **2015** Ph.D. (Information Technology) King Mongkut's University of Technology North Bangkok, Thailand.
- **2005** M.S.Tech.Ed. (Computer Technology) King Mongkut's University of Technology North Bangkok, Thailand.
- **1997** B.B.A (Business Computer) Sripatum University, Thailand.

Language

Thai English

Awards

Jan 2024 | RMUTP
บุคคลดีเด่นด้านการจัดการองค์ความรู้แห่งปี
เจ้าของผลงานวิจัยชนะเลิศด้านวิทยาศาสตร์เทคโนโลยีฯ
บุคคลดีเด่นด้านการจัดการองค์ความรู้แห่งปี

Jan 2023 | RMUTP
บุคคลดีเด่นด้านการจัดการองค์ความรู้แห่งปี

RESEARCH / PAPERS

2021

ศุทธินี สุวรรณพงศ์, ชนากานต์ ครูวิชา และวีรวรรณ จันทนะทรัพย์. (2564). สื่อการเรียนรู้วิธีใช้งานเครื่องออกกำลังกายสาธารณะสำหรับผู้สูงอายุด้วยเทคโนโลยีความจริงเสริม.รายงานสืบเนื่องการประชุมวิชาการระดับชาติด้านวิทยาศาสตร์ เทคโนโลยี และนวัตกรรม ครั้งที่ 4 "วิทยาศาสตร์ เทคโนโลยี และนวัตกรรมสร้างสรรค์ เพื่อก้าวผ่านสถานการณ์ COVID-19". คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร. (น.135-142), 30 สิงหาคม 2564.

2019

ณัฐชา สกุณา, กฤษณะ ธรรมนิตยกุล และวีรวรรณ จันทนะทรัพย์. (2562). ระบบกลอนประตูอัจฉริยะด้วยการรู้จำใบหน้า. การประชุมวิชาการวิศวกรรมศาสตร์และเทคโนโลยี มทร.พระนคร ครั้งที่ 4, คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร, กรุงเทพมหานคร, 31 พฤษภาคม 2562, 363-366.

2016

วีรวรรณ จันทนะทรัพย์ และ พยุง มีสัจ. (2016). การพัฒนาและวัดประสิทธิภาพระบบติดตามดวงตาต้นทุนต่ำ. วารสารวิชาการและวิจัย มทร.พระนคร, 10(1), 141-154. doi: https://doi.org/10.14456/jrmutp.2016.12

Janthanasub, V. (2016). Ophapasai: Augmentative and Alternative Communication Based on Video-Occulography Control Interface. In Applied Mechanics and Materials. 848, 60-63. doi: 10.4028/www.scientific.net/AMM.848.60

2015

Janthanasub, V., and Meesad, P. (2015). Improving the Evolutionary Computation for General Keyboard Arrangement Problem. In Applied Mechanics and Materials. 804, 337-340. doi: 10.4028/www.scientific.net/AMM.804.337

Janthanasub, V., and Meesad, P. (2015). Evaluation of a Low-Cost Eye Tracking System for Computer Input. In King Mongkut's University of Technology North Bangkok International Journal of Applied Science and Technology KMUTNB: IJAST. 8(3), 1-12. doi: http://dx.doi.org/10.14416/j.ijast.2015.07.001

2014

Veerawan J., Phayung M., "Improving the Evolutionary Computation for General Keyboard Arrangement Problem". The 5th RMUTP International Conference on Science, Technology and Innovation for Sustainable Development: the road towards a green future. Pullman Bangkok King Power, Bangkok, Thailand, 17 – 18 July 2014.

วีรวรรณ จันทนะทรัพย์, พยุง มีสัจ และเธรา ศิริกุล. "การออกแบบโครงสร้างแป้นพิมพ์เสมือนด้วยการคำนวณเชิงวิวัฒนาการ". การประชุมวิชาการระดับชาติด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ครั้งที่ 10, 8-9 พฤษภาคม 2557, ภูเก็ต. หน้าที่ 804-809.

2013

วีรวรรณ จันทนะทรัพย์ และคณะ."การเปรียบเทียบประสิทธิภาพการติดตามดวงตาด้วยอัลกอริทึมมีนชิปและลูคัส-คานาเดะ". การประชุมวิชาการระดับชาติด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ครั้งที่ 9, 9-10 พฤษภาคม 2556, กรุงเทพ.

วีรวรรณ จันทนะทรัพย์ และภัสสร สิงหธรรม."การวิจัยและพัฒนาระบบจัดการเนื้อหาเว็บศิลปวัฒนธรรมและสิ่งแวดล้อมของชุมชน". การประชุมทางวิชาการระดับชาติเทคโนโลยีราชมงคล ครั้งที่ 5. 15-16 กรกฎาคม 2556. กรุงเทพฯ.

2012

วิภาวรรณ บัวทอง, กิตติพงศ์ สุวรรณราช และวีรวรรณ จันทนะทรัพย์. "การวิเคราะห์ผลกระทบการลดมิติของข้อมูลกับประสิทธิภาพในการจำแนกข้อมูล แบบต้นไม้ตัดสินใจ ซัพพอร์ตเวกเตอร์แมชชีน และนาอีฟเบย์." การประชุมทางวิชาการระดับชาติด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ครั้งที่ 8. 9-10 พฤษภาคม 2555, พัทยา ชลบุรี.

2011

วีรวรรณ จันทนะทรัพย์. "การพยากรณ์ผลการเรียนด้วยการแบบจำลองถดถอยโพลิโนเมียลเชิงพหุ" การประชุมวิชาการนำเสนอผลงานนานาชาติ ครั้งที่ 2. 6-7 ธันวาคม 2554, ภูเก็ต.

2007

ราตรี จันทนะทรัพย์."การจำแนกลายนิ้วมือด้วยเทคนิคเฮาดอรฟดิสแทนซ์." การประชุมวิชาการบทความของมหาวิทยาลัยต่อการศึกษาที่เน้นการปฏิบัติงานจริงครั้งที่ 1 มหาวิทยาลัย เทคโนโลยีราชมงคลล้านนา. 23-28 สิงหาคม 2552. เชียงใหม่.

2005

ราตรี จันทนะทรัพย์."การจำแนกลายนิ้วมือด้วยเทคนิคเฮาดอรฟดิสแทนซ์." การประชุมวิชาการบทความของมหาวิทยาลัยต่อการศึกษาที่เน้นการปฏิบัติงานจริงครั้งที่ 1 มหาวิทยาลัย เทคโนโลยีราชมงคลล้านนา. 23-28 สิงหาคม 2552. เชียงใหม่.

ราตรีจันทนะทรัพย์ และ นิดาพรรณสุรีรัตนันท์. "การรู้จำป้ายทะเบียนรถยนต์ไทยโดยใช้เทคนิคเฮาดอรฟดิสแทนซ์." วารสารเทคโนโลยีสารสนเทศ ปีที่ 1 ฉบับที่ 2 กรกฎาคม-ธันวาคม, 2548 : 27-31.

Ratree Jannasub and Nidapan Sureerattanan. "Car License Plate Recognition through Hausdorff Distance.", in Proceedings of The 17st IEEE International Conference on Toolswith Artificial Intelligence."November 14-16, 2005. Hong Kong.

BOOK

2022

วีรวรรณ จันทนะทรัพย์. การเขียนโปรแกรมคอมพิวเตอร์ด้วยภาษาไพธอน ฉบับปรับปรุงใหม่. พิมพ์ครั้งที่ 2. กรุงเทพฯ : บริษัทแดเน็กซ์อินเตอร์คอร์ปอเรชั่นจำกัด. 2565. 360 หน้า.

2018

วีรวรรณ จันทนะทรัพย์. 2561. ตำราเรียนการประมวลผลภาพดิจิทัล. บริษัท แดเน็กซ์ อินเตอร์คอร์ปอเรชัน จำกัด. กรุงเทพมหานคร. 320 หน้า.

PETTY PATENT

พยุง มีสัจ และวีรวรรณ จันทนะทรัพย์. (2559). แป้นพิมพ์เสมือนสำหรับการพิมพ์ด้วยสายตา. เลขที่อนุสิทธิบัตร 11651, 23 มิถุนายน 2559.