

การประยุกต์ใช้เทคนิคเหมืองข้อมูลเพื่อพัฒนาระบบสารสนเทศ

พยากรณ์การใช้น้ำประปา

Application of Data Mining Techniques for Developing an Information System for Water Demand Forecasting

หฤทัย อาษากิจ^{1*} และ พัชรมัย ชาวสืบ²

Haruethai Asakit^{1*} and Patcharamai Saosueb²

^{1,2}คณะบริหารธุรกิจและศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา เชียงใหม่

^{1,2}Faculty of Business Administration and Liberal Arts, Rajamangala University of Technology Lanna,

Chiang Mai

aeharuethai@gmail.com

Received: February 3, 2026; Revised: June 26, 2026; Accepted: June 27, 2026; Published: June 30, 2026

ABSTRACT – This study compared the performance of water consumption forecasting models in Thailand using data mining techniques based on the Cross-Industry Standard Process for Data Mining (CRISP-DM). Four models were evaluated: Linear Regression (LR), Artificial Neural Network (ANN), Random Forest (RF), and Gradient Boosted Trees (GBT). Monthly water consumption data published by the Metropolitan Waterworks Authority (MWA) through Thailand's Government Open Data Center, covering 2014–2023, were divided into training, validation, and testing sets (80:10:10). The models were developed using RapidMiner and evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The results showed that the Linear Regression model achieved the best performance, with an MAE of 188,158.96 m³ and an RMSE of 310,809.74 m³, outperforming the other models while demonstrating high prediction stability. A web-based prototype decision support system was also developed to visualize forecasting results through an interactive dashboard, supporting water production planning and water distribution management in alignment with Sustainable Development Goal 6 (SDG 6).

KEY WORDS: Water demand forecasting, Data mining, Information system, Decision support system

บทคัดย่อ -- งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์ปริมาณการใช้น้ำประปาในประเทศไทย โดยประยุกต์ใช้เทคนิคเหมืองข้อมูลร่วมกับกระบวนการ CRISP-DM แบบจำลองที่ศึกษา ได้แก่ การถดถอยเชิงเส้น โครงข่ายประสาทเทียม แรนดอมฟอเรสต์ และต้นไม้เสริมกำลังแบบไล่ระดับ ข้อมูลปริมาณการใช้น้ำประปารายเดือนที่เผยแพร่โดยการประปานครหลวงผ่านศูนย์กลางข้อมูลเปิดภาครัฐ ช่วงปี พ.ศ.2557–2566 และแบ่งเป็นชุดฝึกสอน ชุดตรวจสอบ และชุดทดสอบในสัดส่วน 80:10:10 โดยใช้โปรแกรม RapidMiner ในการพัฒนาแบบจำลอง การประเมินผลใช้ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยและค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ผลการวิจัยพบว่าแบบจำลองการถดถอยเชิงเส้นให้ผลดีที่สุด โดยมีค่า MAE เท่ากับ 188,158.96 ลูกบาศก์เมตร และค่า RMSE เท่ากับ 310,809.74 ลูกบาศก์เมตร ซึ่งต่ำกว่าแบบจำลองอื่น และมีความเสถียรสูง นอกจากนี้ยังได้พัฒนาระบบต้นแบบบนแพลตฟอร์มเว็บเพื่อแสดงผลการพยากรณ์ในรูปแบบแดชบอร์ด สนับสนุนการวางแผนการผลิตและการกระจายน้ำ และสอดคล้องกับเป้าหมายการพัฒนาที่ยั่งยืนด้านทรัพยากรน้ำ (SDG 6)

คำสำคัญ: การพยากรณ์การใช้น้ำประปา เหมืองข้อมูล ระบบสารสนเทศ ระบบสนับสนุนการตัดสินใจ

1. บทนำ

น้ำเป็นทรัพยากรธรรมชาติที่มีบทบาทสำคัญต่อการดำรงชีวิตของมนุษย์ พืช และสัตว์ รวมถึงกิจกรรมทางเศรษฐกิจที่สำคัญ เช่น ภาคการเกษตร อุตสาหกรรม และชุมชนเมือง แม้ว่าน้ำจะเป็นทรัพยากรที่สามารถหมุนเวียนได้ตามวัฏจักรธรรมชาติ แต่การเปลี่ยนแปลงสภาพภูมิอากาศ การเพิ่มขึ้นของประชากร และการขยายตัวของเมืองได้ก่อให้เกิดความต้องการใช้น้ำที่เพิ่มสูงขึ้นอย่างต่อเนื่อง โดยเฉพาะในภาคครัวเรือนและอุตสาหกรรม ส่งผลให้หลายพื้นที่ทั่วโลกเผชิญกับความเสี่ยงด้านความมั่นคงทางน้ำ [1] สถานการณ์ดังกล่าวสะท้อนให้เห็นถึงความจำเป็นในการพัฒนาระบบบริหารจัดการน้ำที่มีประสิทธิภาพและยั่งยืนในยุคปัจจุบัน รายงานด้านการประเมินผลกระทบสิ่งแวดล้อมระบุว่า ความผันผวนของอุณหภูมิและปริมาณน้ำฝนมีความสัมพันธ์โดยตรงกับรูปแบบการใช้น้ำของชุมชนเมืองและภาคครัวเรือน [2] ซึ่งสอดคล้องกับรายงานของการประปานครหลวง (2564) ที่พบว่าอุณหภูมิเฉลี่ยรายเดือนมีความสัมพันธ์กับความต้องการใช้น้ำประปาอย่างมีนัยสำคัญ โดยอุณหภูมิที่เพิ่มขึ้น 1 องศาเซลเซียส ส่งผลให้ความต้องการใช้น้ำเพิ่มขึ้นเฉลี่ยประมาณร้อยละ 1.3 ต่อปี นอกจากนี้ รายงานสถานการณ์ทรัพยากรน้ำของประเทศไทยยังชี้ให้เห็นว่าแนวโน้มอุณหภูมิที่สูงขึ้นอย่างต่อเนื่องส่งผลให้ความต้องการใช้น้ำประปาในเขตเมืองเพิ่มสูงขึ้นอย่างมีนัยสำคัญ [3] ทำให้การวางแผนระบบผลิตและกระจายน้ำในอนาคตจำเป็นต้องอาศัยข้อมูลเชิงพยากรณ์ที่มีความแม่นยำและเชื่อถือได้

ในบริบทของประเทศไทย การบริหารจัดการทรัพยากรน้ำยังเผชิญกับความท้าทายทั้งด้านนโยบายและเทคนิค โดยเฉพาะการพัฒนาระบบประปาที่ยั่งยืนและการรองรับการขยายตัวของเมือง [4] ภายใต้แผนนโยบาย Thailand 4.0 และแนวคิดสมรรถนะประเทศไทย ได้เน้นย้ำถึงความสำคัญของการประยุกต์ใช้เทคโนโลยีดิจิทัลและระบบพยากรณ์แบบเรียลไทม์เพื่อสนับสนุนการตัดสินใจเชิงบริหาร [5] ขณะเดียวกัน ภาครัฐยังให้ความสำคัญต่อการบูรณาการข้อมูลและเทคโนโลยีเพื่อยกระดับประสิทธิภาพการบริหารจัดการน้ำในระดับประเทศ [6]

ปัจจุบัน เทคโนโลยีการวิเคราะห์ข้อมูลขนาดใหญ่ เทคนิคเหมืองข้อมูล และการเรียนรู้ของเครื่อง ถูกนำมาใช้ในการพัฒนาระบบพยากรณ์ความต้องการใช้น้ำมากขึ้น เนื่องจากสามารถรองรับข้อมูลจำนวนมากและสร้างแบบจำลองที่มีความแม่นยำสูง รวมทั้งสามารถพิจารณาปัจจัยด้านสภาพอากาศ ปริมาณฝน ความชื้น และจำนวนประชากรควบคู่กันได้ [7] อย่างไรก็ตาม การเลือกแบบจำลองที่เหมาะสมยังคงเป็นประเด็นสำคัญ เนื่องจากแบบจำลองแต่ละประเภทมีข้อจำกัดและประสิทธิภาพที่แตกต่างกัน

จากประเด็นดังกล่าว ผู้วิจัยจึงมีแนวคิดในการประยุกต์ใช้เทคนิคเหมืองข้อมูลเพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์การใช้น้ำประปา และคัดเลือกแบบจำลองที่เหมาะสมเพื่อนำไปพัฒนาเป็นระบบพยากรณ์เชิงสารสนเทศสำหรับสนับสนุนการวางแผนและการตัดสินใจด้านการจัดการทรัพยากรน้ำอย่างเป็นระบบและยั่งยืน โดยมุ่งเน้นการบูรณาการการเปรียบเทียบแบบจำลองเหมืองข้อมูลหลายประเภทเข้ากับการพัฒนาระบบสารสนเทศเชิงปฏิบัติการบนแพลตฟอร์มเว็บ ซึ่งยังพบว่ามีการศึกษาจำกัดในบริบทข้อมูลการใช้น้ำประปาระดับประเทศของประเทศไทยในลักษณะดังกล่าว

2. วัตถุประสงค์

2.1 เพื่อศึกษาและเปรียบเทียบประสิทธิภาพของแบบจำลองเหมืองข้อมูลสำหรับการพยากรณ์ปริมาณการใช้น้ำประปา

2.2 เพื่อพัฒนาและทดสอบแบบจำลองพยากรณ์ปริมาณการใช้น้ำประปาโดยใช้เทคนิคเหมืองข้อมูล

2.3 เพื่อพัฒนาระบบสารสนเทศเชิงสนับสนุนการตัดสินใจบนแพลตฟอร์มเว็บสำหรับการจัดการน้ำประปา

3. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

3.1 การทำเหมืองข้อมูลและกระบวนการ CRISP-DM

(Data Mining and CRISP-DM)

การทำเหมืองข้อมูล (Data Mining) เป็นกระบวนการค้นหาและสกัดองค์ความรู้ รูปแบบ และความสัมพันธ์ที่มีคุณค่าจากชุดข้อมูลขนาดใหญ่ โดยอาศัยองค์ความรู้จากหลากหลายศาสตร์ ได้แก่ สถิติ ปัญญาประดิษฐ์ และการเรียนรู้ของเครื่อง (Machine Learning) เพื่อวิเคราะห์ข้อมูลที่มีความซับซ้อนและ

สนับสนุนการตัดสินใจเชิงบริหาร [8] ในยุคดิจิทัลที่องค์กรและหน่วยงานภาครัฐมีการจัดเก็บข้อมูลจำนวนมากจากหลากหลายแหล่ง การทำเหมืองข้อมูลจึงมีบทบาทสำคัญในการแปลงข้อมูลให้เป็นสารสนเทศและองค์ความรู้ที่สามารถนำไปใช้ประโยชน์ได้จริง

การทำเหมืองข้อมูลถือเป็นส่วนหนึ่งของกระบวนการวิเคราะห์ข้อมูล (Data Analytics) ซึ่งเป็นกระบวนการตรวจสอบจัดการ และวิเคราะห์ข้อมูลเพื่อค้นหาสารสนเทศเชิงลึกที่ช่วยสนับสนุนการตัดสินใจและเพิ่มประสิทธิภาพการดำเนินงานขององค์กร [9] โดยการบูรณาการเทคนิคการวิเคราะห์ข้อมูลร่วมกับการทำเหมืองข้อมูลและปัญญาประดิษฐ์ ช่วยเพิ่มความสามารถในการค้นหารูปแบบที่ซับซ้อนของข้อมูล และสนับสนุนแนวคิดการบริหารจัดการองค์กรด้วยข้อมูล (Data-driven Decision Making) [10]

ในการดำเนินงานด้านการทำเหมืองข้อมูลและการวิเคราะห์ข้อมูล กรอบแนวคิด CRISP-DM (Cross-Industry Standard Process for Data Mining) ได้รับการยอมรับอย่างแพร่หลาย เนื่องจากเป็นกระบวนการมาตรฐานที่ช่วยให้การพัฒนาแบบจำลองเป็นระบบและสามารถนำผลลัพธ์ไปประยุกต์ใช้ได้จริง [11] โดย CRISP-DM ประกอบด้วย 6 ขั้นตอน ได้แก่ (1) การทำความเข้าใจปัญหาทางธุรกิจ (Business Understanding) (2) การทำความเข้าใจข้อมูล (Data Understanding) (3) การเตรียมข้อมูล (Data Preparation) (4) การสร้างแบบจำลอง (Modeling) (5) การประเมินผล (Evaluation) และ (6) การนำผลไปใช้งาน (Deployment)

กระบวนการดังกล่าวได้รับการนำไปประยุกต์ใช้อย่างแพร่หลายในหลากหลายสาขา เช่น การแพทย์ การเงิน พลังงาน สิ่งแวดล้อม และการบริหารจัดการทรัพยากรน้ำ โดยเฉพาะงานด้านการพยากรณ์ความต้องการใช้น้ำ ซึ่งต้องอาศัยการวิเคราะห์ข้อมูลเชิงเวลาและการสร้างแบบจำลองเพื่อสนับสนุนการวางแผนและการตัดสินใจเชิงบริหาร [12] ดังนั้น งานวิจัยนี้จึงประยุกต์ใช้กรอบแนวคิด CRISP-DM เป็นแนวทางในการดำเนินงานวิจัย ตั้งแต่การรวบรวมและเตรียมข้อมูล การพัฒนาแบบจำลองพยากรณ์ ไปจนถึงการประเมินประสิทธิภาพของแบบจำลอง เพื่อคัดเลือกแนวทางที่เหมาะสมสำหรับการพัฒนาระบบพยากรณ์ปริมาณการใช้น้ำประปา

3.2 การแบ่งข้อมูล (Data Splitting)

การแบ่งชุดข้อมูลออกเป็น 3 ส่วน ได้แก่ ชุดข้อมูลฝึกสอน (Training set) ร้อยละ 80 สำหรับการเรียนรู้แบบจำลอง ชุดข้อมูลตรวจสอบ (Validation set) ร้อยละ 10 สำหรับการปรับแต่งค่าพารามิเตอร์ของแบบจำลอง และชุดข้อมูลทดสอบ (Testing set) ร้อยละ 10 สำหรับการประเมินประสิทธิภาพขั้นสุดท้ายของแบบจำลอง ถือเป็นแนวปฏิบัติมาตรฐานในงานด้านการเรียนรู้ของเครื่อง เพื่อช่วยลดปัญหาการจดจำรูปแบบเฉพาะของข้อมูลฝึกสอน (overfitting) และเพิ่มความสามารถในการอธิบายทั่วไปของแบบจำลอง (generalization) การใช้ชุดข้อมูลตรวจสอบช่วยให้สามารถคัดเลือกแบบจำลองและค่าพารามิเตอร์ที่เหมาะสมก่อนนำไปประเมินผลด้วยชุดข้อมูลทดสอบจริง ส่งผลให้ผลการประเมินมีความน่าเชื่อถือและสะท้อนประสิทธิภาพของแบบจำลองในสถานการณ์ใช้งานจริงได้ดียิ่งขึ้น [13] [14]

3.3 การวัดประสิทธิภาพ (Evaluation)

การประเมินประสิทธิภาพของแบบจำลองเป็นขั้นตอนสำคัญในกระบวนการวิเคราะห์ข้อมูลและการพยากรณ์ เนื่องจากช่วยให้สามารถตรวจสอบความแม่นยำ ความเสถียร และความเหมาะสมของแบบจำลองก่อนนำไปประยุกต์ใช้งานจริง โดยทั่วไปนิยมใช้ตัวชี้วัดทางสถิติเพื่อประเมินความคลาดเคลื่อนระหว่างค่าที่พยากรณ์และค่าจริง ซึ่งตัวชี้วัดที่ใช้กันอย่างแพร่หลาย ได้แก่ ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Root Mean Square Error: RMSE) และค่าเฉลี่ยของค่าความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Error: MAE) ค่า RMSE เป็นตัวชี้วัดที่ให้น้ำหนักกับค่าความคลาดเคลื่อนขนาดใหญ่เป็นพิเศษ จึงเหมาะสำหรับกรณีที่ต้องการลดผลกระทบจากความผิดพลาดที่มีขนาดสูง ขณะที่ค่า MAE เป็นตัวชี้วัดที่สะท้อนค่าความคลาดเคลื่อนเฉลี่ยในรูปแบบสัมบูรณ์ ทำให้สามารถตีความผลลัพธ์ได้ง่ายและมีความเสถียรต่อค่าผิดปกติ (Outliers) มากกว่า [15] ดังนั้นการใช้ตัวชี้วัดทั้ง RMSE และ MAE ร่วมกันจึงช่วยให้สามารถประเมินประสิทธิภาพของแบบจำลองได้อย่างรอบด้าน และสนับสนุนการเลือกแบบจำลองที่เหมาะสมสำหรับการนำไปใช้งานด้านการพยากรณ์เชิงบริหารและการตัดสินใจเชิงข้อมูล

3.4 งานวิจัยที่เกี่ยวข้อง

A. Antunes et al. [16] ศึกษาการพยากรณ์ความต้องการใช้น้ำประปาในระยะสั้นในเขตเมืองของประเทศโปรตุเกส โดยใช้ข้อมูลอนุกรมเวลารายวันร่วมกับตัวแปรด้านสภาพอากาศ เช่น อุณหภูมิและปริมาณน้ำฝน และเปรียบเทียบแบบจำลอง Machine Learning ได้แก่ Random Forest, Gradient Boosting และ Support Vector Regression (SVR) การประเมินประสิทธิภาพใช้ตัวชี้วัด RMSE, MAE และ R^2 ผลการวิจัยพบว่า Gradient Boosting สามารถลดค่าความคลาดเคลื่อนได้มากที่สุด และให้ค่าความแม่นยำสูงสุดเมื่อเทียบกับโมเดลอื่น เนื่องจากสามารถจัดการความสัมพันธ์เชิงไม่เชิงเส้นของข้อมูลได้อย่างมีประสิทธิภาพ ผู้วิจัยเสนอว่าแบบจำลองดังกล่าวเหมาะสำหรับการนำไปพัฒนาเป็นระบบสนับสนุนการตัดสินใจขององค์กรประปา เพื่อช่วยในการวางแผนกำลังการผลิตและการจัดสรรทรัพยากรน้ำในระยะสั้น

W. Hao et al. [17] ศึกษาการเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์ความต้องการใช้น้ำประปาในเมืองมิลาน ประเทศอิตาลี สำหรับช่วงระยะสั้น (1 วันล่วงหน้า) และระยะยาว (7 วันล่วงหน้า) โดยใช้ชุดข้อมูลรายวันจากระบบน้ำของเมือง งานวิจัยนี้ทดสอบทั้งแบบจำลองเชิงสถิติและวิธี Machine Learning ได้แก่ ARIMA, Artificial Neural Networks (ANN), Support Vector Regression (SVR), Light Gradient Boosting Machine (LightGBM) และ Long Short-Term Memory (LSTM) พร้อมพิจารณาการใช้กรอบ Wavelet Data-Driven Forecasting Framework (WDDFF) เพื่อเพิ่มประสิทธิภาพของโมเดล ผลการเปรียบเทียบชี้ให้เห็นว่าเทคนิค WDDFF ช่วยพัฒนาความสามารถในการพยากรณ์ของโมเดลทั้งหมด โดยเฉพาะเมื่อใช้ร่วมกับ LSTM ซึ่งให้ค่าความแม่นยำสูง ($R^2 > 0.9$) สำหรับทั้งการพยากรณ์ระยะสั้นและระยะยาว นอกจากนี้ LightGBM ยังแสดงศักยภาพในการลดจำนวนตัวแปรและเลือกคุณลักษณะสำคัญในบริบทของน้ำและทรัพยากรน้ำได้อย่างมีประสิทธิภาพ

N. Fumo et al. [18] ศึกษาการพยากรณ์ความต้องการใช้น้ำโดยใช้ชุดข้อมูลขนาดใหญ่จากระบบสาธารณูปโภคในสหรัฐอเมริกา และเปรียบเทียบแบบจำลอง Random Forest, XGBoost (ซึ่งอยู่ในกลุ่ม Gradient Boosted Trees) และ Linear Regression โดยใช้ตัวชี้วัด RMSE และ MAE เป็นเกณฑ์

ประเมินผล ผลการวิจัยพบว่า XGBoost ให้ประสิทธิภาพสูงสุดในการลดค่าความคลาดเคลื่อน เนื่องจากสามารถจัดการความสัมพันธ์เชิงซับซ้อนและข้อมูลที่มีความไม่เชิงเส้นได้ดี ขณะที่ Linear Regression ให้ผลลัพธ์ที่มีเสถียรภาพและสามารถตีความเชิงบริหารได้ง่าย ผู้วิจัยเสนอว่าการผสมผสานโมเดลขั้นสูงกับโมเดลเชิงสถิติสามารถช่วยเพิ่มความยืดหยุ่นในการพัฒนาเครื่องมือพยากรณ์สำหรับองค์กรด้านสาธารณูปโภค

Z. Pu et al. [19] นำเสนอโมเดล deep learning แบบผสมระหว่าง wavelet multi-resolution analysis, convolutional Artificial Neural Networks และ long short-term memory เพื่อพยากรณ์ปริมาณการใช้น้ำประปาในเขตเมืองระยะสั้น ผลการศึกษาแสดงให้เห็นว่าระบบ hybrid นี้สามารถให้ความแม่นยำในการพยากรณ์ได้ดีกว่าโมเดล ANN, LSTM, Conv1D และ GRUN แบบดั้งเดิม ซึ่งสอดคล้องกับแนวคิดการพัฒนาแบบจำลองเชิงดิจิทัลที่มีความสามารถตอบสนองต่อรูปแบบข้อมูลที่ซับซ้อนได้อย่างมีประสิทธิภาพ

I. I. Aminu [7] ได้ทบทวนเชิงระบบของงานวิจัยด้านการพยากรณ์และการจำลองความต้องการใช้น้ำในช่วง 10 ปีที่ผ่านมา โดยมุ่งเน้นการประยุกต์ใช้แบบจำลองเชิงคำนวณอัจฉริยะ (Computational Intelligence Models) เช่น Artificial Neural Networks (ANN), Support Vector Machines (SVM), Random Forest, Gradient Boosting, Deep Learning และ Hybrid Models ผลการทบทวนพบว่าแบบจำลองปัญญาประดิษฐ์และการเรียนรู้ของเครื่องสามารถจัดการข้อมูลที่มีความไม่เชิงเส้นและความผันผวนได้ดีกว่าแบบจำลองเชิงสถิติแบบดั้งเดิม อย่างไรก็ตาม ยังไม่มีแบบจำลองใดที่เหมาะสมกับทุกบริบท เนื่องจากประสิทธิภาพขึ้นอยู่กับลักษณะข้อมูล ความละเอียดเชิงเวลา และปัจจัยด้านสภาพแวดล้อม โดยเสนอแนวทางการพัฒนาในอนาคต ได้แก่ การใช้ Hybrid Models การบูรณาการข้อมูลจาก Internet of Things (IoT) การประยุกต์ใช้ Big Data Analytics และการพัฒนาระบบพยากรณ์อัจฉริยะเพื่อสนับสนุนการตัดสินใจด้านการบริหารจัดการทรัพยากรน้ำอย่างยั่งยืน

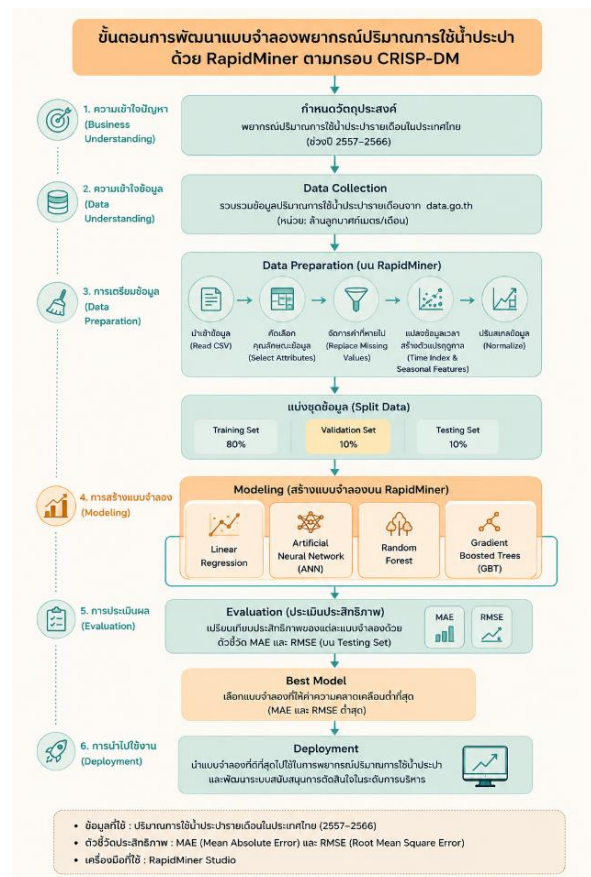
จากการทบทวนวรรณกรรมในช่วงปี ค.ศ. 2022–2025 พบว่า งานวิจัยด้านการพยากรณ์ปริมาณการใช้น้ำประปามีแนวโน้มประยุกต์ใช้เทคนิค Deep Learning และ Hybrid

Models มากขึ้น โดยเฉพาะ Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU) และการผสมเทคนิค Deep Learning ร่วมกับการประมวลผลสัญญาณหรือการคัดเลือกคุณลักษณะ ซึ่งสามารถเพิ่มความสามารถในการเรียนรู้รูปแบบข้อมูลอนุกรมเวลาและความสัมพันธ์ที่ซับซ้อนได้ อย่างไรก็ตาม แบบจำลองดังกล่าวมักมีความซับซ้อนในการพัฒนา ต้องการทรัพยากรด้านการคำนวณสูง และอาศัยข้อมูลจากหลายแหล่งร่วมกันเพื่อให้ได้ประสิทธิภาพสูงสุด

นอกจากนี้ งานวิจัยส่วนใหญ่ใช้ข้อมูลจากระบบประปาเฉพาะพื้นที่หรือข้อมูลที่มีตัวแปรสนับสนุนจำนวนมาก ขณะที่การเปรียบเทียบประสิทธิภาพของเทคนิคเหมือนข้อมูลภายใต้ข้อมูลเปิดภาครัฐที่สามารถเข้าถึงได้จริงยังมีจำนวนจำกัด ซึ่งสะท้อนให้เห็นถึงช่องว่างของงานวิจัย (Research Gap)

ดังนั้น งานวิจัยนี้จึงมุ่งเปรียบเทียบประสิทธิภาพของเทคนิคเหมือนข้อมูลที่ได้รับความนิยม ได้แก่ Artificial Neural Network (ANN), Linear Regression, Random Forest และ Gradient Boosted Trees บนแพลตฟอร์ม RapidMiner เพื่อประเมินความเหมาะสมในการพัฒนาระบบพยากรณ์ปริมาณการใช้น้ำประปาภายใต้บริบทของข้อมูลภาครัฐ พร้อมเสนอให้การประยุกต์ใช้แบบจำลอง Deep Learning เช่น LSTM, GRU และ Hybrid Models เป็นแนวทางสำหรับการศึกษาค้นคว้าในอนาคต

3.5 กรอบแนวคิดงานวิจัย



รูปที่ 1 กรอบแนวคิดและกระบวนการพัฒนาแบบจำลองพยากรณ์ปริมาณการใช้น้ำประปา ตามกรอบ CRISP-DM

จากรูปที่ 1 แสดงกรอบแนวคิดและกระบวนการดำเนินงานของงานวิจัยตามแนวทาง CRISP-DM โดยเริ่มจากการรวบรวมข้อมูลปริมาณการใช้น้ำประปารายเดือนจากศูนย์กลางข้อมูลเปิดภาครัฐ (data.go.th) จากนั้นดำเนินการเตรียมข้อมูล ประกอบด้วย การคัดเลือกตัวแปร การจัดการข้อมูลที่ไม่สมบูรณ์ การสร้างตัวแปรเชิงเวลา และการปรับสเกลข้อมูลก่อนแบ่งชุดข้อมูลออกเป็น Training Set, Validation Set และ Testing Set เพื่อใช้ในการพัฒนาและประเมินแบบจำลอง

ในขั้นตอนการสร้างแบบจำลอง ผู้วิจัยได้เปรียบเทียบประสิทธิภาพของแบบจำลอง 4 ประเภท ได้แก่ Linear Regression, Artificial Neural Network (ANN), Random Forest และ Gradient Boosted Trees โดยใช้ค่า Mean Absolute Error (MAE) และ Root Mean Squared Error (RMSE) เป็นตัวชี้วัดประสิทธิภาพ ก่อนคัดเลือกแบบจำลองที่ให้ผลการพยากรณ์ดีที่สุดเพื่อนำไปประยุกต์ใช้ในการพัฒนาระบบพยากรณ์ปริมาณการใช้น้ำประปา

4. วิธีดำเนินการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบประสิทธิภาพของแบบจำลองพยากรณ์ปริมาณการใช้น้ำประปาโดยประยุกต์ใช้เทคนิคเหมืองข้อมูล เพื่อพัฒนาแบบจำลองที่มีความเหมาะสมสำหรับการนำไปใช้งานเชิงบริหาร งานวิจัยได้ดำเนินการตามกรอบกระบวนการ CRISP-DM (Cross-Industry Standard Process for Data Mining) ซึ่งเป็นมาตรฐานสากลด้านการวิเคราะห์ข้อมูล ประกอบด้วย 6 ขั้นตอนหลัก ได้แก่ การทำความเข้าใจปัญหาทางธุรกิจ การทำความเข้าใจข้อมูล การเตรียมข้อมูล การสร้างแบบจำลอง การประเมินผล และการนำไปใช้งาน ดังนี้

4.1 การทำความเข้าใจปัญหา (Business Understanding)

กระบวนการทำความเข้าใจปัญหาทางธุรกิจ (Business Understanding) เป็นขั้นตอนแรกของกรอบการวิเคราะห์ข้อมูลตามแนวคิด CRISP-DM ซึ่งมุ่งเน้นการกำหนดขอบเขตของปัญหา วัตถุประสงค์ขององค์กร และความต้องการเชิงบริหาร เพื่อแปลงโจทย์ทางธุรกิจให้อยู่ในรูปแบบของปัญหาการวิเคราะห์ข้อมูลที่สามารถดำเนินการได้อย่างเป็นระบบ ในการวิจัยครั้งนี้ ผู้วิจัยได้กำหนดปัญหาหลักเป็นการพยากรณ์ปริมาณการใช้น้ำประปาในเขตเมือง เพื่อสนับสนุนการวางแผนการผลิตและการกระจายน้ำของหน่วยงานที่เกี่ยวข้อง ข้อมูลที่นำมาใช้ในการศึกษาได้ถูกรวบรวมจากแหล่งข้อมูลภาครัฐ ได้แก่ ศูนย์กลางข้อมูลเปิดภาครัฐ (Open Government Data) และหน่วยงานด้านการประปา ซึ่งเป็นแหล่งข้อมูลที่มีความน่าเชื่อถือและสามารถนำมาใช้ในการวิเคราะห์เชิงนโยบายและเชิงบริหารได้ ผู้วิจัยได้วิเคราะห์ความต้องการของผู้มีส่วนได้ส่วนเสีย เช่น หน่วยงานด้านการบริหารจัดการน้ำ ผู้กำหนดนโยบาย และผู้ใช้น้ำ เพื่อกำหนดเป้าหมายของระบบพยากรณ์ให้สามารถตอบสนองต่อการตัดสินใจด้านการจัดสรรทรัพยากร การวางแผนกำลังการผลิต และการบริหารความเสี่ยงจากภาวะขาดแคลนน้ำในอนาคต ผลจากขั้นตอนการทำความเข้าใจปัญหาทางธุรกิจถูกนำไปใช้เป็นกรอบกำหนดตัวแปรนำเข้า แนวทางการเลือกแบบจำลองพยากรณ์ และเกณฑ์การประเมินประสิทธิภาพของแบบจำลอง เพื่อให้กระบวนการวิเคราะห์ข้อมูลมีความสอดคล้องกับวัตถุประสงค์ของงานวิจัย

และสามารถนำผลลัพธ์ไปใช้ประโยชน์เชิงปฏิบัติได้อย่างเป็นรูปธรรม

4.2 การทำความเข้าใจข้อมูล (Data Understanding)

ผู้วิจัยได้รวบรวมข้อมูลปริมาณการใช้น้ำประปาจากศูนย์กลางข้อมูลเปิดภาครัฐ (Open Government Data of Thailand: data.go.th) ซึ่งเป็นแพลตฟอร์มกลางสำหรับการเผยแพร่ข้อมูลของหน่วยงานภาครัฐในประเทศไทย โดยข้อมูลชุดนี้เผยแพร่โดยการประสานครหลวง (Metropolitan Waterworks Authority: MWA) ครอบคลุมช่วงระยะเวลาตั้งแต่ปี พ.ศ. 2557–2566 รวมระยะเวลา 10 ปี และมีจำนวนข้อมูลทั้งสิ้น 257,046 ระเบียบ

ข้อมูลที่ใช้ในการวิจัยเป็นข้อมูลปริมาณการใช้น้ำประปารายเดือน (Monthly Water Consumption Data) ซึ่งอยู่ในรูปแบบข้อมูลสรุปรายเดือน (Monthly Aggregated Data) ที่หน่วยงานเจ้าของข้อมูลได้ดำเนินการรวบรวมและสรุปผลเรียบร้อยแล้ว ก่อนเผยแพร่ผ่านระบบ data.go.th โดยแต่ละระเบียนแสดงปริมาณการใช้น้ำประปารวมของพื้นที่และประเภทผู้ใช้น้ำในเดือนนั้น ๆ มีหน่วยวัดเป็นลูกบาศก์เมตร (m³) ดังนั้น ผู้วิจัยจึงไม่จำเป็นต้องดำเนินการรวมข้อมูล (Data Aggregation) เพิ่มเติมก่อนการวิเคราะห์

ข้อมูลต้นฉบับประกอบด้วยตัวแปรจำนวน 11 ตัวแปร ได้แก่ (1) ปี (2) เดือน (3) รหัสจังหวัด (4) ชื่อจังหวัด (5) รหัสอำเภอ (6) ชื่ออำเภอ (7) รหัสตำบล (8) ชื่อตำบล (9) รหัสประเภทผู้ใช้น้ำ (10) ชื่อประเภทผู้ใช้น้ำ และ (11) ปริมาณการใช้น้ำประปา ซึ่งเป็นตัวแปรเป้าหมาย (Target Variable) สำหรับการพยากรณ์

ข้อมูลดังกล่าวจัดเก็บในระดับพื้นที่เชิงภูมิศาสตร์ ประกอบด้วยข้อมูลระดับจังหวัด อำเภอ และตำบล รวมถึงประเภทผู้ใช้น้ำในแต่ละพื้นที่ ซึ่งช่วยสะท้อนลักษณะการใช้น้ำประปาในมิติเชิงพื้นที่และเชิงเวลาได้อย่างครอบคลุม เหมาะสำหรับการนำมาพัฒนาแบบจำลองพยากรณ์ปริมาณการใช้น้ำประปา

ในการพัฒนาแบบจำลอง ผู้วิจัยได้คัดเลือกตัวแปรจำนวน 6 ตัวแปร ได้แก่ ปี เดือน รหัสจังหวัด รหัสอำเภอ รหัสประเภทผู้ใช้น้ำ และปริมาณการใช้น้ำประปา เนื่องจากเป็นตัวแปรที่สะท้อนลักษณะเชิงเวลาและเชิงพื้นที่ที่มีผลต่อความต้องการใช้

น้ำประปาโดยตรง ขณะที่ตัวแปรประเภทชื่อจังหวัด ชื่ออำเภอ ชื่อตำบล และชื่อประเภทผู้ใช้น้ำเป็นข้อมูลเชิงข้อความที่ซ้ำซ้อนกับตัวแปรรหัส จึงไม่นำมาใช้ในการสร้างแบบจำลอง เพื่อลดความซ้ำซ้อนของข้อมูล

4.3 การเตรียมข้อมูล (Data Preparation)

การเตรียมข้อมูลเป็นขั้นตอนสำคัญในกระบวนการวิเคราะห์ข้อมูลและการพัฒนาแบบจำลองพยากรณ์ โดยมีวัตถุประสงค์เพื่อปรับปรุงคุณภาพของข้อมูลให้อยู่ในรูปแบบที่ถูกต้อง ครบถ้วน และเหมาะสมต่อการนำไปประมวลผลเชิงคำนวณ ในงานวิจัยนี้ ผู้วิจัยได้ดำเนินการกระบวนการเตรียมข้อมูลตามขั้นตอนดังต่อไปนี้

1) การคัดเลือกข้อมูล (Data Selection) ผู้วิจัยได้ทำการคัดเลือกเฉพาะข้อมูลที่มีความเกี่ยวข้องโดยตรงกับวัตถุประสงค์ของการวิจัย เพื่อลดความซ้ำซ้อนของข้อมูลและเพิ่มประสิทธิภาพในการวิเคราะห์ โดยเลือกใช้ข้อมูลปริมาณการใช้น้ำประปารายเดือนในช่วงปี พ.ศ. 2557–2566 ซึ่งรวบรวมจากศูนย์กลางข้อมูลเปิดภาครัฐ (Open Government Data of Thailand: data.go.th) และเผยแพร่โดยการประปานครหลวง (Metropolitan Waterworks Authority: MWA)

ข้อมูลต้นทางถูกจัดเก็บและเผยแพร่ในรูปแบบข้อมูลสรุปรายเดือน (Monthly Aggregated Data) อยู่แล้ว โดยแต่ละระเบียนแสดงปริมาณการใช้น้ำประปารวมของพื้นที่และประเภทผู้ใช้น้ำในแต่ละเดือน ดังนั้น ผู้วิจัยจึงนำข้อมูลมาใช้ในการวิเคราะห์โดยตรง โดยไม่ต้องดำเนินการรวมข้อมูลตามช่วงเวลาเพิ่มเติม (Temporal Aggregation) ทั้งนี้เพื่อรักษาค่าปริมาณการใช้น้ำประปาที่สะท้อนการใช้งานจริงในแต่ละเดือน และให้สอดคล้องกับวัตถุประสงค์ของการพยากรณ์ปริมาณการใช้น้ำประปารายเดือน

สำหรับการสร้างแบบจำลอง ผู้วิจัยได้คัดเลือกตัวแปรที่สำคัญจำนวน 6 ตัวแปร ได้แก่ (1) ปี พ.ศ. (2) เดือน (3) รหัสจังหวัด (4) รหัสอำเภอ (5) รหัสประเภทผู้ใช้น้ำ และ (6) ปริมาณการใช้น้ำประปา ซึ่งเป็นตัวแปรเป้าหมาย (Target Variable) สำหรับการพยากรณ์

2) การกลั่นกรองข้อมูล (Data Cleaning) ในขั้นตอนการกลั่นกรองข้อมูล ผู้วิจัยได้ดำเนินการตรวจสอบความถูกต้อง ความครบถ้วน และความสอดคล้องของข้อมูล เพื่อเพิ่มคุณภาพ

ของชุดข้อมูลก่อนนำไปสร้างแบบจำลอง โดยมีขั้นตอนสำคัญดังนี้

- การปรับรูปแบบภาษา (Data Standardization) เนื่องจากข้อมูลต้นทางจัดเก็บในรูปแบบภาษาไทย ผู้วิจัยได้ดำเนินการแปลงข้อมูลบางตัวแปร เช่น เดือน และประเภทผู้ใช้น้ำ ให้อยู่ในรูปแบบภาษาอังกฤษ เพื่อให้สามารถประมวลผลด้วยเครื่องมือวิเคราะห์ข้อมูลและซอฟต์แวร์ทางสถิติได้อย่างมีประสิทธิภาพ

- การจัดการค่าว่าง (Missing Values) ทำการตรวจสอบค่าที่ขาดหาย และแทนค่าที่สูญหายด้วยค่ากลางหรือค่าที่เหมาะสมตามลักษณะของข้อมูล เพื่อหลีกเลี่ยงผลกระทบต่อความแม่นยำของแบบจำลอง

- การตรวจสอบและลบข้อมูลซ้ำ (Duplicate Records) ดำเนินการตรวจสอบระเบียนข้อมูลที่ซ้ำซ้อนทั้งในด้านตัวแปรและค่า และลบข้อมูลที่ซ้ำออกจากชุดข้อมูล

- การตรวจสอบค่าผิดปกติ (Outliers Detection) วิเคราะห์ค่าที่มีการเบี่ยงเบนจากรูปแบบข้อมูลทั่วไปอย่างมีนัยสำคัญ และดำเนินการจัดการข้อมูลที่เป็นค่าผิดปกติเพื่อลดผลกระทบต่อประสิทธิภาพของแบบจำลอง

3) การแปลงข้อมูล (Data Transformation) ผู้วิจัยได้ดำเนินการแปลงข้อมูลให้อยู่ในรูปแบบที่เหมาะสมสำหรับการวิเคราะห์เชิงเวลาและการพัฒนาแบบจำลองพยากรณ์ โดยมีขั้นตอนสำคัญ ได้แก่

- แปลงข้อมูลเดือนให้อยู่ในรูปแบบเชิงตัวเลข (1–12) เพื่อใช้ในการจัดลำดับข้อมูลอนุกรมเวลา

- รวมข้อมูลปีและเดือนเป็นตัวแปรเชิงเวลาในรูปแบบ Year-Month และสร้างตัวแปรดัชนีเวลา (Time Index) จากลำดับของเดือนในชุดข้อมูล โดยกำหนดให้เดือนมกราคม พ.ศ. 2557 มีค่าเท่ากับ 1 และเพิ่มขึ้นทีละ 1 ตามลำดับเดือนจนถึงเดือนธันวาคม พ.ศ. 2566 ทั้งนี้ ระเบียนข้อมูลที่อยู่ในเดือนเดียวกันจะมีค่า Time Index เดียวกัน เพื่อใช้แทนแนวโน้มการเปลี่ยนแปลงของปริมาณการใช้น้ำประปาตามช่วงเวลา

- สร้างตัวแปรฤดูกาล (Seasonal Feature) จากตัวแปรเดือน (Month) โดยแบ่งข้อมูลออกเป็น 3 ฤดูกาล ได้แก่ ฤดูร้อน (มีนาคม–พฤษภาคม) ฤดูฝน (มิถุนายน–ตุลาคม) และฤดูหนาว (พฤศจิกายน–กุมภาพันธ์) จากนั้นแปลงข้อมูลให้อยู่ในรูปแบบ

One-Hot Encoding เพื่อให้แบบจำลองสามารถเรียนรู้ผลกระทบของฤดูกาลที่มีต่อปริมาณการใช้น้ำประปาได้อย่างเหมาะสม

4) การเลือกตัวแปร (Feature Selection) ผู้วิจัยได้ทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรต้น (Independent Variables) และตัวแปรตาม คือ ปริมาณการใช้น้ำประปา โดยใช้เครื่องมือทางสถิติและการวิเคราะห์ความสัมพันธ์เชิงสหสัมพันธ์ เพื่อคัดเลือกรูปแบบตัวแปรที่มีอิทธิพลต่อการพยากรณ์มากที่สุด ทั้งนี้เพื่อช่วยลดมิติของข้อมูลและเพิ่มประสิทธิภาพของแบบจำลอง

5) การแบ่งชุดข้อมูล (Data Splitting) เพื่อให้การประเมินประสิทธิภาพของแบบจำลองมีความถูกต้องและน่าเชื่อถือ งานวิจัยนี้ได้ดำเนินการแบ่งชุดข้อมูลออกเป็น 3 ส่วน ได้แก่ ชุดข้อมูลฝึกสอน (Training set) ร้อยละ 80 สำหรับการเรียนรู้แบบจำลอง ชุดข้อมูลตรวจสอบ (Validation set) ร้อยละ 10 สำหรับการปรับแต่งค่าพารามิเตอร์ของแบบจำลอง และชุดข้อมูลทดสอบ (Testing set) ร้อยละ 10 สำหรับการประเมินประสิทธิภาพขั้นสุดท้ายของแบบจำลอง ซึ่งเป็นแนวปฏิบัติมาตรฐานในงานด้านการเรียนรู้ของเครื่อง (Machine Learning) เพื่อหลีกเลี่ยงปัญหาการจดจำข้อมูลเฉพาะ (Overfitting) และเพิ่มความสามารถในการอธิบายทั่วไป (Generalization) ของแบบจำลอง ในกระบวนการแบ่งข้อมูล ผู้วิจัยได้ใช้วิธีการแบ่งข้อมูลตามลำดับเวลา (Time-based Splitting) โดยกำหนดให้ข้อมูลในช่วงเวลาก่อนหน้านี้เป็นชุดฝึกสอนและชุดตรวจสอบ ขณะที่ข้อมูลในช่วงเวลาล่าสุดถูกใช้เป็นชุดทดสอบ เพื่อป้องกันปัญหาการรั่วไหลของข้อมูล (Data Leakage) และสะท้อนลักษณะการใช้งานจริงของระบบพยากรณ์เชิงเวลา ชุดข้อมูล Validation ถูกนำมาใช้สำหรับการปรับแต่งพารามิเตอร์ของแบบจำลอง (Hyperparameter Tuning) และการคัดเลือกโครงสร้างแบบจำลองที่เหมาะสม ก่อนนำแบบจำลองที่ผ่านการปรับแต่งแล้วไปประเมินผลขั้นสุดท้ายด้วยชุดข้อมูล Testing ซึ่งช่วยให้ผลการประเมินมีความน่าเชื่อถือและสามารถสะท้อนประสิทธิภาพของแบบจำลองได้อย่างเหมาะสมตามบริบทการใช้งานจริง

4.4 การสร้างแบบจำลอง (Modeling)

ในขั้นตอนการสร้างแบบจำลอง ผู้วิจัยได้นำข้อมูลผ่านกระบวนการเตรียมข้อมูลและการแบ่งชุดข้อมูลมาใช้ในการ

พัฒนาแบบจำลองพยากรณ์ปริมาณการใช้น้ำประปา โดยเลือกใช้แบบจำลองจำนวน 4 ประเภท ได้แก่ Artificial Neural Network (ANN), Linear Regression (LR), Random Forest (RF) และ Gradient Boosted Trees (GBT) ซึ่งเป็นแบบจำลองที่ได้รับความนิยมในการวิเคราะห์ข้อมูลเชิงพยากรณ์และสามารถรองรับข้อมูลที่มีความซับซ้อนได้

เพื่อเพิ่มประสิทธิภาพของแบบจำลอง ผู้วิจัยได้ดำเนินการปรับแต่งค่าพารามิเตอร์ (Hyperparameter Tuning) โดยใช้ชุดข้อมูล Validation เป็นเกณฑ์ในการประเมินผล และทดลองปรับค่าพารามิเตอร์หลายชุดสำหรับแต่ละแบบจำลอง เช่น จำนวนชั้นซ่อนและจำนวนโหนดของ ANN จำนวนต้นไม้ของ Random Forest และจำนวนต้นไม้ร่วมกับอัตราการเรียนรู้ของ Gradient Boosted Trees จากนั้นเลือกชุดค่าพารามิเตอร์ที่ให้ค่าความคลาดเคลื่อนต่ำที่สุดบนชุดข้อมูล Validation เพื่อนำไปใช้ในการสร้างแบบจำลองขั้นสุดท้ายก่อนประเมินผลด้วยชุดข้อมูล Testing

ตารางที่ 1. ค่าพารามิเตอร์ที่ได้รับการคัดเลือกสำหรับแต่ละแบบจำลอง

แบบจำลอง	ค่าพารามิเตอร์ที่ได้รับการคัดเลือก
ANN	Hidden Layers = 2, Neurons = (64, 32), Learning Rate = 0.001, Batch Size = 32, Epochs = 100
Random Forest	Number of Trees = 100, Max Depth = 20
Gradient Boosted Trees	Number of Trees = 100, Learning Rate = 0.05, Max Depth = 5
Linear Regression	ใช้ค่าพารามิเตอร์เริ่มต้นของอัลกอริทึม

ค่าพารามิเตอร์ที่แสดงในตารางที่ 1 เป็นค่าที่ให้ผลการพยากรณ์ดีที่สุดบนชุดข้อมูล Validation และถูกนำมาใช้ในการสร้างแบบจำลองขั้นสุดท้ายก่อนประเมินประสิทธิภาพด้วยชุดข้อมูล Testing

กระบวนการ Grid Search ทำการทดสอบชุดค่าพารามิเตอร์ที่กำหนดไว้ล่วงหน้าหลายรูปแบบ และประเมินประสิทธิภาพของแต่ละชุดด้วยค่า Root Mean Squared Error (RMSE) บนชุด Validation จากนั้นเลือกชุดค่าพารามิเตอร์ที่ให้ผลการพยากรณ์ดีที่สุดเพื่อนำไปใช้ในการสร้างแบบจำลองขั้นสุดท้าย

สำหรับแบบจำลอง ANN ผู้วิจัยใช้โครงสร้างแบบ Feedforward Neural Network ประกอบด้วยชั้นซ่อนจำนวน 2 ชั้น โดยกำหนดจำนวนโหนดเท่ากับ 64 และ 32 โหนด ตามลำดับ ซึ่งเลือกจากผลการทดลองเบื้องต้นโดยพิจารณาค่า Validation Error ต่ำสุด ใช้ฟังก์ชันกระตุ้นแบบ ReLU และ อัลกอริทึม Adam สำหรับการปรับค่าน้ำหนัก กำหนดอัตราการเรียนรู้เท่ากับ 0.001 ขนาดชุดข้อมูลฝึกต่อรอบเท่ากับ 32 และ จำนวนรอบการฝึกเท่ากับ 100 รอบ โดยใช้ฟังก์ชันสูญเสียแบบ Mean Squared Error เพื่อเพิ่มความแม่นยำในการพยากรณ์

นอกจากตัวแปรเชิงพื้นที่และตัวแปรเชิงเวลาแล้ว ผู้วิจัยได้ สร้างตัวแปรฤดูกาล (Seasonal Feature) จากข้อมูลเดือนและ แปลงให้อยู่ในรูปแบบ One-Hot Encoding เพื่อสะท้อน ผลกระทบของฤดูกาลต่อปริมาณการใช้น้ำประปา โดยตัวแปร ดังกล่าวถูกนำไปใช้ร่วมกับตัวแปร ปีรหัสจังหวัด รหัสอำเภอ และรหัสประเภทผู้ใช้น้ำ เป็นข้อมูลนำเข้าสำหรับแบบจำลอง Artificial Neural Network, Random Forest และ Gradient Boosted Trees

สำหรับแบบจำลอง Linear Regression ผู้วิจัยใช้ตัวแปร ดัชนีเวลา (Time Index) เป็นตัวแปรอิสระในการอธิบาย แนวโน้มการเปลี่ยนแปลงของปริมาณการใช้น้ำประปาตาม เวลา โดยสมมติให้ตัวแปรเป้าหมายสามารถอธิบายได้ด้วย ความสัมพันธ์เชิงเส้นระหว่าง Time Index และปริมาณการใช้น้ำประปา ทั้งนี้ ก่อนการสร้างแบบจำลองได้ดำเนินการปรับ สเกลข้อมูลและแปลงตัวแปรเวลาให้อยู่ในรูปแบบดัชนีเวลา (Time Index) เพื่อเพิ่มเสถียรภาพของการประมาณค่า สัมประสิทธิ์และลดปัญหาความเอนเอียงเชิงตัวเลข (Numerical Instability) แบบจำลองถูกประเมินด้วยข้อมูล Validation Set เพื่อคัดเลือกพารามิเตอร์ที่เหมาะสมก่อนนำไปทดสอบ ประสิทธิภาพกับชุดข้อมูล Testing Set

สำหรับแบบจำลอง Random Forest ผู้วิจัยใช้เทคนิค Ensemble Learning โดยสร้างชุดต้นไม้ตัดสินใจหลายต้นและ รวมผลการพยากรณ์ด้วยวิธีการเฉลี่ยค่า (Averaging) เพื่อเพิ่ม เสถียรภาพของแบบจำลองและลดความแปรปรวนของผลลัพธ์ โดยกำหนดจำนวนต้นไม้ (Number of Trees) เท่ากับ 100 ต้น และกำหนดความลึกสูงสุดของต้นไม้ (Maximum Depth) เท่ากับ 20 ระดับ ทั้งนี้ ใช้วิธีสุ่มตัวอย่างข้อมูลแบบ Bootstrap

Sampling และการสุ่มเลือกตัวแปรย่อยในแต่ละโหนด เพื่อเพิ่ม ความสามารถในการเรียนรู้รูปแบบข้อมูลที่มีความไม่เชิงเส้น และลดปัญหา overfitting ของแบบจำลอง

สำหรับแบบจำลอง Gradient Boosted Trees ผู้วิจัยใช้ แนวคิด Boosting ซึ่งเป็นการสร้างต้นไม้ตัดสินใจแบบลำดับขั้น โดยแต่ละต้นจะเรียนรู้จากค่าความคลาดเคลื่อนของต้นก่อน หน้า เพื่อเพิ่มความแม่นยำของแบบจำลองโดยรวม โดยกำหนด จำนวนต้นไม้เท่ากับ 100 ต้น ค่าอัตราการเรียนรู้ (Learning Rate) เท่ากับ 0.05 และความลึกสูงสุดของต้นไม้เท่ากับ 5 ระดับ ทั้งนี้ การกำหนดพารามิเตอร์ดังกล่าวช่วยควบคุมความซับซ้อน ของโมเดล ลดความเสี่ยงของการเกิด overfitting และเพิ่ม เสถียรภาพของผลการพยากรณ์ในข้อมูลอนุกรมเวลา

โดยแบบจำลองทั้ง 4 เทคนิคถูกพัฒนาและปรับแต่ง พารามิเตอร์โดยอาศัยชุดข้อมูล Validation เพื่อคัดเลือกราคาที่ ให้ผลการเรียนรู้มีเสถียรภาพและลดความคลาดเคลื่อนในการ พยากรณ์ จากนั้นจึงนำแบบจำลองที่ได้ไปประเมิน ประสิทธิภาพขั้นสุดท้ายด้วยชุดข้อมูล Testing Set เพื่อสะท้อน สมรรถนะของโมเดลภายใต้สภาพการใช้งานจริง

4.5 การประเมินผล (Evaluation)

การประเมินประสิทธิภาพของแบบจำลองพยากรณ์ ผู้วิจัยได้ใช้ ตัวชี้วัดทางสถิติที่ได้รับการยอมรับอย่างแพร่หลายในงานด้าน การพยากรณ์และการเรียนรู้ของเครื่อง ได้แก่ ค่าเฉลี่ยของความ คลาดเคลื่อนสัมบูรณ์ (Mean Absolute Error: MAE) และ ค่ารากที่สองของความคลาดเคลื่อนกำลังสอง (Root Mean Square Error: RMSE) เพื่อเปรียบเทียบความแม่นยำของ แบบจำลองแต่ละประเภท โดยตัวชี้วัดทั้งสองสะท้อนระดับ ความคลาดเคลื่อนระหว่างค่าที่พยากรณ์และค่าจริง และสามารถประเมินประสิทธิภาพของแบบจำลองเชิงตัวเลขได้ อย่างมีประสิทธิภาพ

1) ค่ารากที่สองของความคลาดเคลื่อนกำลังสอง (Root Mean Squared Error: RMSE) ดังสมการที่ 1 ซึ่งใช้ วัดขนาดของความคลาดเคลื่อนโดยรวมของแบบจำลอง โดย ค่าที่ต่ำกว่าบ่งชี้ถึงความแม่นยำที่สูงกว่า

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1)$$

โดย n คือ จำนวนข้อมูลทั้งหมด

y_i คือ ค่าจริงของข้อมูลลำดับที่ i

$\hat{y}_i$ คือ ค่าที่โมเดลพยากรณ์ได้ของลำดับที่ i

$(y_i - \hat{y}_i)^2$ คือ ค่าคลาดเคลื่อนกำลังสองของแต่ละจุดข้อมูล

2) ค่าเฉลี่ยของผลต่างสัมบูรณ์ (Mean Absolute Error: MAE) ดังสมการที่ 2 ซึ่งใช้วัดความคลาดเคลื่อนโดยเฉลี่ยในหน่วยเดียวกับข้อมูลจริง ให้ภาพรวมที่ชัดเจนของค่าที่ทำนายผิดพลาด

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

โดย n คือ จำนวนข้อมูลทั้งหมด

y_i คือ ค่าจริงของข้อมูลลำดับที่ i

$\hat{y}_i$ คือ ค่าที่โมเดลพยากรณ์ได้ของลำดับที่ i

$|y_i - \hat{y}_i|$ คือ ค่าสัมบูรณ์ของค่าคลาดเคลื่อน

ในงานวิจัยนี้ แบบจำลองที่ให้ค่า MAE และ RMSE ต่ำที่สุดจะถูกพิจารณาว่าเป็นแบบจำลองที่มีประสิทธิภาพสูงสุด และเหมาะสมสำหรับการนำไปประยุกต์ใช้ในการพัฒนาระบบพยากรณ์การใช้น้ำประปาเชิงดิจิทัล

4.6 การนำไปใช้งาน (Deployment)

การนำแบบจำลองที่ผ่านกระบวนการพัฒนาและประเมินประสิทธิภาพไปใช้งานจริงเป็นขั้นตอนสำคัญของกรอบ CRISP-DM โดยมีเป้าหมายเพื่อแปลงผลลัพธ์เชิงเทคนิคให้อยู่ในรูปแบบที่สามารถนำไปใช้สนับสนุนการตัดสินใจเชิงบริหารและการกำหนดนโยบายด้านการจัดการทรัพยากรน้ำได้อย่างมีประสิทธิภาพ ในงานวิจัยนี้ กระบวนการนำแบบจำลองไปใช้งานประกอบด้วย 3 ขั้นตอนหลัก ดังนี้

1) ขั้นตอนการคัดเลือกแบบจำลองที่เหมาะสม ภายหลังจากการพัฒนาแบบจำลองทั้ง 4 ประเภท ได้แก่ Artificial Neural Network (ANN), Linear Regression (LR), Random Forest (RF) และ Gradient Boosted Trees (GB) ผู้วิจัยได้ประเมินประสิทธิภาพของแต่ละแบบจำลองด้วยตัวชี้วัดทางสถิติ ได้แก่ Mean Absolute Error (MAE) และ Root Mean Square Error (RMSE) โดยแบบจำลองที่ให้ค่าความคลาดเคลื่อนต่ำที่สุดและมีความเสถียรในการพยากรณ์ถูกเลือกเป็นแบบจำลองหลักสำหรับการนำไปใช้งานจริง เพื่อให้สามารถพยากรณ์ปริมาณการใช้น้ำประปาได้อย่างแม่นยำและเชื่อถือได้

2) ขั้นตอนการนำแบบจำลองที่ผ่านการคัดเลือกไปใช้ในการพยากรณ์ข้อมูลในอนาคต โดยใช้ข้อมูลชุดใหม่ (New Incoming Data) เป็นข้อมูลนำเข้า เพื่อสร้างผลลัพธ์การพยากรณ์ที่สะท้อนสถานการณ์การใช้น้ำในช่วงเวลาถัดไป ผลลัพธ์ดังกล่าวสามารถนำไปใช้สนับสนุนการวางแผนเชิงบริหาร เช่น การเตรียมความพร้อมด้านกำลังการผลิตน้ำในช่วงฤดูแล้ง การบริหารจัดการน้ำในช่วงความต้องการสูง และการจัดสรรทรัพยากรน้ำในแต่ละพื้นที่ให้สอดคล้องกับความต้องการใช้งานจริง

3) ขั้นตอนการนำเสนอและสื่อสารผลลัพธ์การพยากรณ์ผ่านระบบสารสนเทศบนแพลตฟอร์มเว็บ โดยผู้วิจัยได้พัฒนาเว็บไซต์ด้วยเทคโนโลยี HTML, PHP และ JavaScript เพื่อแสดงผลลัพธ์ในรูปแบบกราฟเส้น (Line Chart) และตารางสรุปข้อมูล ซึ่งช่วยให้ผู้ใช้งานสามารถเข้าใจแนวโน้มการใช้น้ำได้อย่างชัดเจนและสามารถเข้าถึงข้อมูลได้แบบออนไลน์อย่างสะดวกและรวดเร็ว ระบบดังกล่าวสามารถนำไปใช้เป็นเครื่องมือสนับสนุนการตัดสินใจของหน่วยงานภาครัฐ องค์กรท้องถิ่น และผู้มีส่วนได้ส่วนเสียอื่น ๆ รวมถึงประชาชนทั่วไปที่ต้องการข้อมูลเพื่อวางแผนการใช้น้ำในระดับครัวเรือนและชุมชน

5. ผลการวิจัย

การวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์ปริมาณการใช้น้ำประปาในประเทศไทย โดยประยุกต์ใช้กระบวนการวิเคราะห์ข้อมูลตามกรอบ CRISP-DM และเทคนิคเหมืองข้อมูลจำนวน 4 แบบ ได้แก่ โครงข่ายประสาทเทียม (Artificial Neural Network: ANN) การถดถอยเชิงเส้น (Linear Regression) แรนดอมฟอเรสต์ (Random Forest) และต้นไม้เสริมกำลังแบบไล่ระดับ (Gradient Boosted Trees) ข้อมูลที่ใช้เป็นข้อมูลการใช้น้ำประปารายเดือนในช่วงปี พ.ศ. 2557–2566 ซึ่งรวบรวมจากศูนย์กลางข้อมูลเปิดภาครัฐ และผ่านกระบวนการเตรียมข้อมูลก่อนนำไปสร้างแบบจำลอง

ในการแบ่งชุดข้อมูล ผู้วิจัยได้ดำเนินการแบ่งข้อมูลออกเป็น 3 ส่วน ได้แก่ ชุดข้อมูลสำหรับการฝึกสอนแบบจำลอง (Training Set) ร้อยละ 80 ชุดข้อมูลตรวจสอบ (Validation Set) และปรับแต่งพารามิเตอร์ของแบบจำลอง ร้อยละ 10 และชุดข้อมูลสำหรับการทดสอบประสิทธิภาพขั้นสุดท้าย (Testing

Set) ร้อยละ 10 โดยใช้วิธีการแบ่งข้อมูลตามลำดับเวลา (Time-based Splitting) เพื่อป้องกันปัญหาการรั่วไหลของข้อมูล (Data Leakage) และสะท้อนลักษณะการใช้งานจริงของระบบพยากรณ์เชิงเวลา จากนั้นใช้โปรแกรม RapidMiner เป็นเครื่องมือในการพัฒนาแบบจำลอง การประเมินประสิทธิภาพของแบบจำลองดำเนินการโดยใช้ตัวชี้วัดทางสถิติ ได้แก่ ค่าเฉลี่ยของค่าความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Error: MAE) และค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Root Mean Square Error: RMSE) เพื่อเปรียบเทียบระดับความแม่นยำของแต่ละเทคนิค ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองแสดงดังตารางที่ 2

ตารางที่ 2. ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองพยากรณ์การใช้น้ำประปา

Model	MAE (ลบ.ม.)	RMSE (ลบ.ม.)
Artificial Neural Network	264,804.09	459,365.09
Linear Regression	188,158.96	310,809.74
Random Forest	543,430.74	806,258.97
Gradient Boosted Trees	1,631,260.09	2,609,506.10

จากตารางที่ 2 พบว่า แบบจำลอง Linear Regression ให้ค่าความคลาดเคลื่อนต่ำที่สุดทั้งในด้านค่า Mean Absolute Error (MAE) และ Root Mean Square Error (RMSE) เมื่อเปรียบเทียบกับแบบจำลองอื่นทั้งหมด แสดงให้เห็นว่าแบบจำลองดังกล่าวมีประสิทธิภาพสูงสุดในการพยากรณ์ปริมาณการใช้น้ำประปาภายใต้ชุดข้อมูลและบริบทของการศึกษานี้

เมื่อพิจารณาเปรียบเทียบกับปริมาณการใช้น้ำประปาสูงสุดในชุดข้อมูล ซึ่งมีค่าเท่ากับ 1,240,771 ลูกบาศก์เมตร พบว่า ค่า MAE ของแบบจำลอง Linear Regression คิดเป็นร้อยละ 15.16 และค่า RMSE คิดเป็นร้อยละ 25.04 ของค่าสูงสุด ผลการวิเคราะห์ดังกล่าวแสดงให้เห็นว่าแบบจำลอง Linear Regression มีประสิทธิภาพสูงสุดเมื่อเปรียบเทียบกับผลดังกล่าว สะท้อนว่าข้อมูลที่ใช้ในการศึกษามีแนวโน้ม (Trend) ชัดเจน และยังไม่มีรูปแบบฤดูกาลหรือความไม่เชิงเส้นที่ซับซ้อนมากนัก จึงทำให้แบบจำลอง Linear Regression สามารถอธิบายความสัมพันธ์ของข้อมูลได้ดีกว่าแบบจำลองอื่นภายใต้ชุดข้อมูลนี้

อย่างไรก็ตาม ผลการศึกษานี้ควรพิจารณาภายใต้ขอบเขตของชุดข้อมูลที่ใช้ในการวิจัย เนื่องจากตัวแปรที่นำมาวิเคราะห์ประกอบด้วยข้อมูลเชิงเวลาและข้อมูลเชิงพื้นที่ที่ได้จากฐานข้อมูลเปิดภาครัฐเป็นหลัก โดยยังไม่ได้รวมปัจจัยภายนอกที่อาจส่งผลกระทบต่อปริมาณการใช้น้ำประปา เช่น อุณหภูมิ ปริมาณน้ำฝน ความชื้น จำนวนประชากร การขยายตัวของเมือง และกิจกรรมทางเศรษฐกิจ ดังนั้น แม้แบบจำลอง Linear Regression จะให้ผลการพยากรณ์ที่ดีที่สุดภายใต้ชุดข้อมูลที่ใช้ศึกษา แต่ประสิทธิภาพของแบบจำลองอาจเปลี่ยนแปลงได้เมื่อมีการบูรณาการตัวแปรภายนอกเพิ่มเติมเข้าสู่กระบวนการวิเคราะห์

ผลการทดลองพบว่าแบบจำลอง Gradient Boosted Trees (GBT) ให้ค่าความคลาดเคลื่อนสูงกว่า Linear Regression และ Random Forest ทั้งในด้าน RMSE และ MAE ซึ่งอาจสะท้อนถึงแนวโน้มการเกิด Overfitting ของแบบจำลอง กล่าวคือแบบจำลองสามารถเรียนรู้รูปแบบเฉพาะของข้อมูลฝึกสอนได้ดี แต่ไม่สามารถอธิบายข้อมูลที่ไม่เคยพบมาก่อนได้อย่างมีประสิทธิภาพ

สาเหตุหนึ่งอาจเกิดจากลักษณะการทำงานของ GBT ที่สร้างต้นไม้ตัดสินใจหลายต้นแบบลำดับต่อเนื่อง (Sequential Learning) เพื่อแก้ไขข้อผิดพลาดของต้นไม้ก่อนหน้า หากกำหนดจำนวนต้นไม้ (Number of Trees) มากเกินไป หรือกำหนดความลึกของต้นไม้ (Tree Depth) สูงเกินความจำเป็น แบบจำลองอาจเรียนรู้รายละเอียดและความผันผวนเฉพาะของข้อมูลฝึกสอนมากเกินไป ส่งผลให้ความสามารถในการอธิบายทั่วไป (Generalization) ลดลงเมื่อทดสอบกับข้อมูลใหม่

แนวทางในการลดปัญหา Overfitting ของ GBT ได้แก่ การลดค่า Learning Rate การจำกัดความลึกของต้นไม้ การลดจำนวนต้นไม้ และการใช้เทคนิค Early Stopping เพื่อหยุดการเรียนรู้เมื่อประสิทธิภาพบนชุดข้อมูล Validation ไม่ปรับตัวดีขึ้น ซึ่งอาจช่วยเพิ่มประสิทธิภาพของแบบจำลองในการศึกษาครั้งต่อไป

สมการของแบบจำลอง Linear Regression สามารถเขียนได้ดังสมการที่ 3

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad (3)$$

เมื่อ

Y = ปริมาณการใช้น้ำประปาที่ต้องการพยากรณ์

β_0 = ค่าคงที่ (Intercept)

$\beta_1, \beta_2, \dots, \beta_n$ = ค่าสัมประสิทธิ์ของตัวแปรอิสระ

$X_1, X_2, \dots, X_n$ = ตัวแปรอิสระ

$\mathcal{E}$ = ค่าความคลาดเคลื่อน (Error Term)

จากผลการวิเคราะห์ด้วยโปรแกรม RapidMiner ดังแสดงในรูปสมการที่ 3 สามารถนำผลลัพธ์มาสร้างสมการพยากรณ์ปริมาณการใช้น้ำประปาโดยใช้เทคนิคการถดถอยเชิงเส้น (Linear Regression) ซึ่งใช้สำหรับอธิบายความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตาม โดยตัวแปรอิสระในงานวิจัยนี้คือดัชนีเวลา (Time Index) ที่แทนลำดับช่วงข้อมูลในช่วงปี พ.ศ. 2557–2566 ซึ่งผ่านกระบวนการปรับสเกลก่อนนำเข้าสู่แบบจำลองรูปแบบสมการถดถอยเชิงเส้นสามารถแสดงได้ดังสมการที่ 4

$$Y_t = \beta_0 + \beta_1 T_t \quad (4)$$

โดยที่

Y_t คือ ปริมาณการใช้น้ำประปาที่ถูกพยากรณ์ในช่วงเวลา t

T_t คือ ดัชนีเวลา (Time Index) ที่ผ่านการปรับสเกลแล้ว

β_0 คือ ค่าคงที่ของแบบจำลอง (Intercept)

β_1 คือ ค่าสัมประสิทธิ์ถดถอยของตัวแปรเวลา

ทั้งนี้ เพื่อหลีกเลี่ยงปัญหาที่เกิดจากการใช้ค่าปีจริงซึ่งมีขนาดตัวเลขสูงและอาจส่งผลต่อเสถียรภาพของการประมาณค่าสัมประสิทธิ์ ผู้วิจัยได้ทำการแปลงข้อมูลปีให้อยู่ในรูปแบบดัชนีเวลา (Time Index) และปรับสเกลข้อมูลก่อนนำเข้าสู่กระบวนการสร้างแบบจำลอง ซึ่งช่วยลดปัญหาความเอนเอียงเชิงตัวเลข (Numerical Instability) และเพิ่มความเสถียรของแบบจำลองในการเรียนรู้แนวโน้มเชิงเวลา

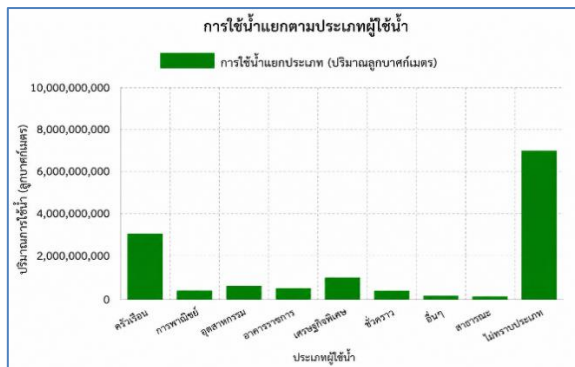
สมการพยากรณ์ดังกล่าวถูกนำไปใช้ในการคาดการณ์ปริมาณการใช้น้ำประปาในอนาคต โดยการแทนค่าดัชนีเวลาของช่วงเวลาที่ต้องการพยากรณ์ลงในสมการ จากนั้นทำการคำนวณเพื่อให้ได้ค่าพยากรณ์ Y_t ซึ่งใช้เป็นข้อมูลสนับสนุนการแสดงผลบนระบบสารสนเทศเชิงพยากรณ์บนแพลตฟอร์มเว็บ

ภายหลังจากการคัดเลือกแบบจำลองที่ให้ประสิทธิภาพดีที่สุด ผู้วิจัยได้นำแบบจำลองดังกล่าวไปพัฒนาเป็นระบบสนับสนุนการตัดสินใจบนเว็บ (Web-based Decision Support System: DSS) เพื่อสนับสนุนการติดตาม วิเคราะห์ และพยากรณ์ปริมาณการใช้น้ำประปา โดยระบบได้รับการออกแบบให้แสดงผลข้อมูลในรูปแบบ Dashboard เพื่อช่วยให้ผู้ใช้งานสามารถติดตามแนวโน้มการใช้น้ำประปาและใช้ข้อมูลประกอบการวางแผนด้านการบริหารจัดการน้ำได้อย่างสะดวก

ระบบต้นแบบได้รับการพัฒนาในลักษณะ Client–Server Architecture โดยสามารถเชื่อมโยงฐานข้อมูลกับแบบจำลองพยากรณ์และแสดงผลผ่านเว็บไซต์ ตัวอย่างดังรูปที่ 2(a) ซึ่งแสดงกราฟแนวโน้มปริมาณการใช้น้ำประปารายปี ช่วยให้ผู้ใช้งานสามารถติดตามการเปลี่ยนแปลงของปริมาณการใช้น้ำในแต่ละปีและวิเคราะห์แนวโน้มของข้อมูลได้อย่างชัดเจน ขณะที่รูปที่ 2(b) แสดงปริมาณการใช้น้ำจำแนกตามประเภทผู้ใช้น้ำ ซึ่งช่วยให้สามารถเปรียบเทียบลักษณะการใช้น้ำของผู้ใช้น้ำแต่ละประเภท และใช้ประกอบการวางแผนการผลิตและการจัดสรรน้ำให้เหมาะสมกับความต้องการของแต่ละกลุ่มผู้ใช้น้ำ นอกจากนี้ การนำเสนอข้อมูลในรูปแบบแดชบอร์ดยังช่วยให้ผู้ใช้งานสามารถวิเคราะห์แนวโน้มและเปรียบเทียบข้อมูลได้อย่างรวดเร็ว ลดความซับซ้อนในการตีความข้อมูล และสนับสนุนการตัดสินใจเชิงบริหารได้อย่างมีประสิทธิภาพ



รูปที่ 2 (a) Dashboard แสดงแนวโน้มปริมาณการใช้น้ำประปารายปี



รูปที่ 2 (b) Dashboard แสดงปริมาณการใช้น้ำประจำวันแยกตามประเภทผู้ใช้น้ำ

ผลการพัฒนาระบบแสดงให้เห็นว่า เว็บไซต์สามารถทำหน้าที่เป็นศูนย์กลางข้อมูลด้านการใช้น้ำประปา (Water Information Dashboard) ที่สนับสนุนการจัดเก็บ การประมวลผล การพยากรณ์ และการแสดงผลข้อมูลผ่านเว็บได้อย่างครบถ้วน อีกทั้งยังรองรับการขยายระบบในอนาคต เช่น การเชื่อมต่อกับฐานข้อมูลภายนอกผ่าน API หรือข้อมูลจากอุปกรณ์ Internet of Things (IoT) เพื่อเพิ่มศักยภาพของระบบในการรองรับข้อมูลแบบเรียลไทม์ และสนับสนุนการพัฒนาระบบบริหารจัดการน้ำประปาเชิงอัจฉริยะ

6. สรุปและอภิปรายผล

ผลการวิจัยแสดงให้เห็นว่าแบบจำลอง Linear Regression ให้ประสิทธิภาพในการพยากรณ์ปริมาณการใช้น้ำประปาสูงที่สุดเมื่อเปรียบเทียบกับแบบจำลองอื่น ได้แก่ Artificial Neural Network, Random Forest และ Gradient Boosted Trees ภายใต้อัตราการผิดพลาดค่าเฉลี่ย (MAE) และค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ที่ต่ำที่สุด ผลลัพธ์ดังกล่าวสะท้อนถึงลักษณะของข้อมูลที่มีแนวโน้มเชิงเวลาและรูปแบบเชิงเส้นค่อนข้างสม่ำเสมอ ส่งผลให้แบบจำลองเชิงสถิติสามารถอธิบายความสัมพันธ์ระหว่างตัวแปรได้อย่างมีประสิทธิภาพและเหมาะสมต่อการนำไปประยุกต์ใช้งานจริง

จากผลการเปรียบเทียบประสิทธิภาพของแบบจำลองพบว่าแบบจำลอง Linear Regression ให้ค่าความคลาดเคลื่อนต่ำที่สุดเมื่อเปรียบเทียบกับ Artificial Neural Network (ANN),

Random Forest และ Gradient Boosted Trees ซึ่งสามารถอธิบายได้จากลักษณะของข้อมูลที่ใช้ในการศึกษา

เมื่อพิจารณาข้อมูลปริมาณการใช้น้ำประจำวันรายเดือนในช่วงปี พ.ศ. 2557–2566 พบว่าข้อมูลมีแนวโน้ม (Trend) ที่เปลี่ยนแปลงอย่างค่อยเป็นค่อยไปตามช่วงเวลา และมีรูปแบบฤดูกาล (Seasonality) ที่สัมพันธ์กับเดือนของปี อันเนื่องมาจากการเปลี่ยนแปลงของสภาพภูมิอากาศและพฤติกรรมการใช้น้ำในแต่ละฤดูกาล นอกจากนี้ ความสัมพันธ์ระหว่างตัวแปรเชิงเวลา ได้แก่ ปี เดือน และดัชนีเวลา (Time Index) กับปริมาณการใช้น้ำประปามีลักษณะใกล้เคียงเชิงเส้น (Linear Pattern) ในหลายช่วงของข้อมูล ส่งผลให้แบบจำลอง Linear Regression สามารถอธิบายแนวโน้มของข้อมูลได้อย่างมีประสิทธิภาพ

ในทางกลับกัน แบบจำลองที่มีความซับซ้อนมากกว่า เช่น Artificial Neural Network, Random Forest และ Gradient Boosted Trees แม้จะสามารถเรียนรู้ความสัมพันธ์แบบไม่เชิงเส้นได้ แต่ภายใต้ชุดข้อมูลที่ใช้ในการศึกษาซึ่งมีรูปแบบแนวโน้มและฤดูกาลค่อนข้างชัดเจน ประสิทธิภาพของแบบจำลองที่ซับซ้อนกว่าไม่ได้เหนือกว่า Linear Regression อย่างมีนัยสำคัญ อีกทั้ง Gradient Boosted Trees ยังมีแนวโน้มเกิด Overfitting จากการเรียนรู้รายละเอียดของข้อมูลฝึกสอนมากเกินไป ทำให้ประสิทธิภาพในการพยากรณ์ข้อมูลใหม่ลดลงเมื่อเปรียบเทียบกับแบบจำลองเชิงเส้น

ผลการศึกษาสอดคล้องกับงานวิจัยของ N. Fumo et al. [18] ซึ่งรายงานว่า Linear Regression สามารถให้ผลการพยากรณ์ที่มีเสถียรภาพในกรณีที่ข้อมูลความต้องการใช้น้ำมีแนวโน้มเชิงเส้น ขณะเดียวกัน งานของ A. Antunes et al. [16] และ I. I. Aminu [7] ชี้ให้เห็นว่าแบบจำลองในกลุ่ม Gradient Boosting และ Artificial Neural Network มีประสิทธิภาพสูงในบริบทของข้อมูลที่มีความไม่เชิงเส้นและความซับซ้อนสูง ผลการเปรียบเทียบดังกล่าวสนับสนุนข้อค้นพบของงานวิจัยนี้ว่าความเหมาะสมของแบบจำลองพยากรณ์ขึ้นอยู่กับลักษณะของข้อมูล โครงสร้างเชิงเวลา และบริบทการนำไปใช้งาน มากกว่าการเลือกใช้เทคนิคที่มีความซับซ้อนเพียงอย่างเดียว

ในเชิงการบริหารจัดการทรัพยากรน้ำ ผลลัพธ์จากแบบจำลอง Linear Regression ที่มีค่าความคลาดเคลื่อนอยู่ในระดับที่ยอมรับได้สะท้อนถึงศักยภาพในการนำไปประยุกต์ใช้

เป็นเครื่องมือสนับสนุนการตัดสินใจ (Decision Support Tool) สำหรับหน่วยงานที่เกี่ยวข้อง เช่น การวางแผนกำลังการผลิตน้ำ การบริหารการกระจายน้ำในช่วงฤดูแล้ง และการจัดลำดับความสำคัญของพื้นที่เสี่ยงต่อการขาดแคลนน้ำ นอกจากนี้ ความเรียบง่ายของแบบจำลองยังช่วยให้ผู้ใช้งานสามารถตีความผลลัพธ์ได้ง่ายและสามารถนำไปใช้ประกอบการตัดสินใจเชิงบริหารได้อย่างมีประสิทธิภาพ

เมื่อพิจารณาพร้อมกับผลการพัฒนาเว็บไซต์ ระบบสารสนเทศที่พัฒนาขึ้นสามารถเชื่อมโยงผลลัพธ์จากแบบจำลองพยากรณ์เข้ากับการแสดงผลในรูปแบบกราฟและแดชบอร์ดออนไลน์ ช่วยเพิ่มประสิทธิภาพในการสื่อสารข้อมูลและการเข้าถึงสารสนเทศเชิงพยากรณ์ของผู้มีส่วนได้ส่วนเสีย ระบบดังกล่าวจึงสามารถมองได้ว่าเป็นต้นแบบของเทคโนโลยีการจัดการที่บูรณาการข้อมูล การวิเคราะห์ และการนำเสนอผลลัพธ์ไว้ในแพลตฟอร์มเดียว และสามารถนำไปต่อยอดสู่ระบบบริหารจัดการน้ำอัจฉริยะในอนาคต

อย่างไรก็ตาม งานวิจัยนี้ยังมีข้อจำกัด ได้แก่ การใช้ข้อมูลเชิงเวลาระดับรายเดือน ซึ่งอาจไม่สามารถสะท้อนความผันผวนของปริมาณการใช้น้ำในระยะสั้นได้อย่างละเอียด นอกจากนี้ แบบจำลองที่พัฒนาขึ้นยังอาศัยข้อมูลเชิงเวลาและข้อมูลเชิงพื้นที่จากฐานข้อมูลเปิดภาครัฐเป็นหลัก โดยยังไม่ได้บูรณาการปัจจัยภายนอกที่อาจส่งผลกระทบต่อความต้องการใช้น้ำ เช่น อุณหภูมิ ปริมาณน้ำฝน ความชื้น จำนวนประชากร การขยายตัวของเมือง และปัจจัยด้านเศรษฐกิจและสังคม

แม้ว่างานวิจัยนี้จะมุ่งเน้นการเปรียบเทียบประสิทธิภาพของเทคนิคเหมืองข้อมูลบนแพลตฟอร์ม RapidMiner ได้แก่ Linear Regression, Artificial Neural Network, Random Forest และ Gradient Boosted Trees แต่แบบจำลองอนุกรมเวลา เช่น ARIMA, SARIMA, Holt-Winters และ Prophet ก็เป็นแนวทางที่ได้รับความนิยมในการพยากรณ์ข้อมูลอนุกรมเวลา เนื่องจากสามารถอธิบายองค์ประกอบของแนวโน้ม (Trend) และฤดูกาล (Seasonality) ได้อย่างมีประสิทธิภาพ อย่างไรก็ตาม เพื่อให้การเปรียบเทียบเป็นไปอย่างเป็นธรรมภายใต้ชุดข้อมูลและสภาพแวดล้อมเดียวกัน งานวิจัยนี้จึงมุ่งเปรียบเทียบเฉพาะเทคนิคเหมืองข้อมูลที่สามารถพัฒนาได้บนแพลตฟอร์ม RapidMiner

สำหรับการวิจัยในอนาคต ควรพิจารณาบูรณาการข้อมูลด้านสภาพภูมิอากาศ เช่น อุณหภูมิ ปริมาณน้ำฝน และความชื้น รวมถึงข้อมูลด้านประชากร การขยายตัวของเมือง และปัจจัยทางเศรษฐกิจและสังคม เพื่อเพิ่มความสามารถในการอธิบายความสัมพันธ์ของข้อมูลและปรับปรุงความแม่นยำของการพยากรณ์ นอกจากนี้ ควรเปรียบเทียบประสิทธิภาพของเทคนิคเหมืองข้อมูลกับแบบจำลองอนุกรมเวลา เช่น ARIMA, SARIMA, Holt-Winters และ Prophet รวมถึงศึกษาการประยุกต์ใช้แบบจำลองเชิงลึก (Deep Learning) เช่น Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU) และ Hybrid Models เพื่อประเมินศักยภาพในการเพิ่มประสิทธิภาพของระบบพยากรณ์ปริมาณการใช้น้ำประปาในอนาคต

7. ข้อเสนอแนะ

งานวิจัยการประยุกต์ใช้เทคนิคเหมืองข้อมูลเพื่อพัฒนาระบบพยากรณ์การใช้น้ำประปานี้ ได้นำเสนอข้อเสนอแนะสำหรับการปรับปรุงและพัฒนางานในอนาคต รวมถึงแนวทางเชิงนโยบาย โดยแบ่งออกเป็น 2 ประเด็น ดังนี้

7.1 ข้อเสนอแนะการวิจัยในอนาคต

1) ควรเพิ่มเติมแปรด้านสภาพภูมิอากาศ เช่น อุณหภูมิ ปริมาณน้ำฝน ความชื้น และความเร็วลม รวมถึงปัจจัยทางเศรษฐกิจและสังคม เช่น จำนวนประชากร การขยายตัวของเมือง และกิจกรรมทางอุตสาหกรรม เพื่อเพิ่มความสามารถของแบบจำลองในการอธิบายความผันผวนของความต้องการใช้น้ำประปา

2) ควรศึกษาการประยุกต์ใช้แบบจำลองขั้นสูง เช่น Long Short-Term Memory (LSTM), Hybrid Model และ Ensemble Learning เพื่อเปรียบเทียบประสิทธิภาพกับแบบจำลองเชิงสถิติและ Machine Learning แบบดั้งเดิม โดยเฉพาะในกรณีที่ข้อมูลรายวันหรือรายชั่วโมง ซึ่งมีความละเอียดเชิงเวลาสูง

3) ควรพัฒนาระบบให้รองรับข้อมูลแบบเรียลไทม์จากอุปกรณ์ Internet of Things (IoT) เช่น มาตรวัดน้ำอัจฉริยะ (Smart Meter) และเซนเซอร์สิ่งแวดล้อม เพื่อเพิ่มความแม่นยำของการพยากรณ์และสนับสนุนการบริหารจัดการน้ำแบบอัจฉริยะ

4) การศึกษาการประยุกต์ใช้เทคนิคการเรียนรู้เชิงลึก (Deep Learning) ร่วมกับข้อมูลเชิงพื้นที่ (Spatial Data) เพื่อพัฒนาแบบจำลองพยากรณ์ระดับพื้นที่ เช่น ระดับเขตหรือชุมชน ซึ่งจะช่วยเพิ่มความละเอียดของผลการพยากรณ์

สำหรับการวิจัยในอนาคต ควรพิจารณาบูรณาการข้อมูลปัจจัยภายนอกที่อาจส่งผลกระทบต่อปริมาณการใช้น้ำประปา เช่น ข้อมูลสภาพอากาศ (อุณหภูมิ ปริมาณน้ำฝน ความชื้น และความเร็วลม) ข้อมูลด้านประชากร การขยายตัวของเมือง และกิจกรรมทางเศรษฐกิจและสังคม เพื่อเพิ่มความสามารถในการอธิบายความสัมพันธ์ของข้อมูลและปรับปรุงประสิทธิภาพของแบบจำลองพยากรณ์ให้มีความแม่นยำมากยิ่งขึ้น นอกจากนี้ ควรศึกษาการประยุกต์ใช้แบบจำลอง Deep Learning สำหรับข้อมูลอนุกรมเวลา เช่น Long Short-Term Memory (LSTM) และ Gated Recurrent Unit (GRU) รวมถึง Hybrid Models เพื่อเปรียบเทียบประสิทธิภาพกับเทคนิคเหมืองข้อมูลที่ใช้ในการศึกษาครั้งนี้ เนื่องจากแบบจำลองดังกล่าวสามารถเรียนรู้ความสัมพันธ์เชิงเวลาและรูปแบบข้อมูลที่ซับซ้อนได้ดีกว่าแบบจำลอง ANN แบบดั้งเดิม อันจะช่วยเพิ่มศักยภาพของระบบพยากรณ์ปริมาณการใช้น้ำประปาในอนาคต

7.2 ข้อเสนอแนะเชิงนโยบาย

1) หน่วยงานที่เกี่ยวข้องกับการบริหารจัดการน้ำประปา เช่น การประปานครหลวง การประปาส่วนภูมิภาค และองค์กรปกครองส่วนท้องถิ่น ควรนำระบบพยากรณ์เชิงดิจิทัลไปใช้เป็นเครื่องมือสนับสนุนการตัดสินใจเชิงบริหาร เพื่อเพิ่มประสิทธิภาพในการวางแผนกำลังการผลิตและการกระจายน้ำ

2) ภาครัฐควรส่งเสริมการบูรณาการข้อมูลจากหลายหน่วยงานเข้าสู่แพลตฟอร์มกลางด้านการจัดการน้ำ เพื่อพัฒนาเป็นระบบสารสนเทศเชิงบูรณาการสำหรับการบริหารจัดการน้ำในระดับประเทศ ซึ่งสอดคล้องกับแนวทาง Smart City และเป้าหมาย SDG 6 และ SDG 11

3) ควรสนับสนุนการเปิดเผยข้อมูลภาครัฐ (Open Data) อย่างเป็นระบบและมีมาตรฐาน เพื่อเอื้อต่อการนำข้อมูลไปใช้ประโยชน์เชิงวิชาการและเชิงนโยบาย รวมถึงการพัฒนาเครื่องมือพยากรณ์เชิงดิจิทัล

4) ควรส่งเสริมความร่วมมือระหว่างภาครัฐ ภาคเอกชน และสถาบันการศึกษาในการพัฒนาเทคโนโลยีและนวัตกรรมด้านการจัดการน้ำ เพื่อสนับสนุนการบริหารจัดการทรัพยากรน้ำอย่างยั่งยืนในระยะยาว

เอกสารอ้างอิง

- [1] The Water Project, “Water scarcity and global water crisis,” 2025. [Online]. Available: <https://thewaterproject.org/water-scarcity-water-crisis>
- [2] World Bank, “Climate change impact assessment on urban water demand,” 2025. [Online]. Available: <https://www.worldbank.org/en/topic/water>
- [3] UNICEF Thailand, “Water, sanitation and climate resilience in Thailand,” 2024. [Online]. Available: <https://www.unicef.org/thailand/wash>
- [4] Pollution Control Department, “Thailand environmental situation report,” 2025. [Online]. Available: <https://www.pcd.go.th>
- [5] Mordor Intelligence, “Smart irrigation market – growth, trends and forecast,” 2025. [Online]. Available: <https://www.mordorintelligence.com/industry-reports/smart-irrigation-market>
- [6] สำนักงานปลัดกระทรวงมหาดไทย, “สุนทรพจน์วันน้ำโลก ประจำปี พ.ศ. 2567,” 2567. [ออนไลน์]. แหล่งที่มา: <https://www.moi.go.th>
- [7] I. I. Aminu, “State-of-art implementation of computing intelligent models for water demand modelling: A decade review and future direction,” Journal of Water and Land Development, 2025. [Online]. Available: <https://www.jwld.pl/files/2025-03-JWLD-22.pdf>
- [8] J. Han, J. Pei, and M. Kamber, Data Mining: Concepts and Techniques, 4th ed. San Francisco, CA, USA: Morgan Kaufmann, 2022.

- [9] B. Marr, *Data Strategy: How to Profit from a World of Big Data, Analytics and Artificial Intelligence*. London, UK: Kogan Page, 2021.
- [10] Y. Zhang, X. Wang, J. Chen, and H. Li, "Application of data mining techniques in water resources management: A review," *Environmental Modelling & Software*, vol. 143, pp. 105104, 2021.
- [11] F. Martínez-Plumed, L. Contreras-Ochando, C. Ferri, J. Hernández-Orallo, M. Kull, N. Lachiche, M. J. Ramírez-Quintana, and P. Flach, "CRISP-DM twenty years later: From data mining processes to data science trajectories," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 8, pp. 3048–3061, 2021.
- [12] T. H. Davenport and S. Miller, "Working with AI: Real stories of human-machine collaboration," *MIT Sloan Management Review*, 2020.
- [13] S. Raschka, Y. Liu, V. Mirjalili, and D. Dzhulgakov, *Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2*. Birmingham, UK: Packt Publishing, 2020.
- [14] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 3rd ed. Sebastopol, CA, USA: O'Reilly Media, 2022.
- [15] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Research*, vol. 30, pp. 79–82, 2021.
- [16] A. Antunes, J. P. Santos, and R. Martins, "Short-term water demand forecasting using machine learning techniques," *Water Resources Management*, vol. 35, no. 4, pp. 1327–1343, 2021.
- [17] W. Hao, A. Cominola, and A. Castelletti, "Comparing predictive machine learning models for short- and long-term urban water demand forecasting in Milan, Italy," *IFAC-PapersOnLine*, vol. 55, no. 33, pp. 92–98, 2022.
- [18] N. Fumo and M. R. Biswas, "Water demand forecasting using machine learning algorithms," *Applied Water Science*, vol. 13, no. 1, pp. 12, 2023.
- [19] Z. Pu, J. Yan, L. Chen, Z. Li, W. Tian, T. Tao, and K. Xin, "A hybrid Wavelet-CNN-LSTM deep learning model for short-term urban water demand forecasting," *Frontiers of Environmental Science & Engineering*, vol. 17, no. 2, pp. 22, 2023.