

การเปรียบเทียบประสิทธิภาพของตัวแบบการพยากรณ์ปริมาณน้ำในเขื่อน ด้วยเทคนิคการทำเหมืองข้อมูล

An Efficiency Comparison of Forecasting Models for the Amount of Water in Dam by Data Mining Techniques

จิรโรจน์ ตอสะสุกุล* และสุพิชชา ขัดธิพงษ์²

Jiraroj Tosasukul* and Supitcha Kuttipong²

Received : October 15, 2021 Revised : February 9, 2022 Accepted : February 15, 2022

บทคัดย่อ

เขื่อนภูมิพลถือเป็นเขื่อนกักเก็บน้ำขนาดใหญ่ที่สุดของประเทศ นอกจากมีความสำคัญต่อการผลิตกระแสไฟฟ้าเพื่อสนองความต้องการของประเทศแล้ว ยังมีความสำคัญในการกักเก็บน้ำสำหรับการอุปโภคและบริโภค รวมถึงช่วยในการจัดสรรน้ำในการเกษตรให้เพียงพอกับความต้องการในช่วงฤดูแล้ง โดยงานวิจัยชิ้นนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพตัวแบบสำหรับการพยากรณ์ปริมาณน้ำในเขื่อนภูมิพล จังหวัดตาก โดยใช้ข้อมูลรายวันจากรายงานสถานการณ์น้ำในเขื่อนของคลังข้อมูลน้ำแห่งชาติและข้อมูลทางอุตุนิยมวิทยาของสถานีกรมอุตุนิยมวิทยาภาคเหนือ ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2554 ถึง 31 พฤษภาคม พ.ศ. 2564 จำนวนทั้งสิ้น 3,804 รายการ และวิเคราะห์ตามกระบวนการมาตรฐานในการทำเหมืองข้อมูล โดยเปรียบเทียบประสิทธิภาพในการพยากรณ์ของเทคนิคการทำเหมืองข้อมูล 6 เทคนิค ได้แก่ เทคนิคการถดถอยเชิงเส้นพหุคูณ เทคนิคแรนดอมฟอเรสต์ เทคนิคโครงข่ายประสาทเทียม เทคนิคจำลองต้นไม้เอมไพรี เทคนิคเพื่อนบ้านใกล้ที่สุด และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน และในแต่ละเทคนิคได้เปรียบเทียบการแบ่งข้อมูลด้วยวิธีการตรวจสอบไขว้ ซึ่งผลการวิจัยพบว่า ตัวแบบการพยากรณ์ปริมาณน้ำในเขื่อนภูมิพลของเทคนิคแรนดอมฟอเรสต์มีประสิทธิภาพมากที่สุด โดยให้ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยและค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด เท่ากับ 341.945 และ 593.790 ตามลำดับ ซึ่งตัวแบบการพยากรณ์ที่ได้ดังกล่าวสามารถนำไปใช้ในการประกอบการตัดสินใจและกำหนดแนวทางการพัฒนาระบบการบริหารจัดการน้ำสำหรับการผลิตกระแสไฟฟ้า การอุปโภค และบริโภคอย่างเพียงพอและเหมาะสม อีกทั้งยังสามารถช่วยในการวางแผนเพิ่มผลผลิตทางการเกษตรและกำหนดชนิดของพืชให้สอดคล้องกับปริมาณน้ำในเขื่อนได้ต่อไปในอนาคต

คำสำคัญ : การพยากรณ์ ปริมาณน้ำในเขื่อน การทำเหมืองข้อมูล

* อาจารย์ประจำภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยนเรศวร

^{*} Lecturer, Department of Mathematics, Faculty of Science, Naresuan University

² นิสิตระดับปริญญาตรี ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยนเรศวร

² Undergraduate student, Department of Mathematics, Faculty of Science, Naresuan University

Abstract

Bhumibol Dam, which is regarded as the largest barrage in Thailand, embraces a crucial role in generating electricity to accommodate the demand of the country and providing water supply for consumption as well as agriculture in a drought season. The objective of the research is to compare the efficiency of the forecasting models for the amount of water stored in Bhumibol Dam, Tak province, based on the daily report of the dam's water condition from the National Hydroinformatics Data Center (NHC) and information from the Northern Meteorological Center from 1 January 2011 to 31 May 2021, which makes a total of 3,804 items. The analysis has been conducted according to the standard process for data mining by making an efficiency comparison of forecasting with six data mining techniques including a Multiple Linear Regression (MLR), a Random Forest, an Artificial Neural Networks Technique (ANN), an M5P Model Tree Technique (M5P), K-Nearest Neighbor Algorithm Technique (K-NN), and a Support Vector Machines Technique (SVM). Each technique uses k-Fold Cross Validation to make a comparison of data splitting. As a result, the research has indicated that the forecasting model for the amount of water in Bhumibol Dam using Random Forest technique is the most effective of all, giving the lowest figures for Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) at 341.945 and 593.790 respectively. Accordingly, the forecasting model can be implemented to make a decision and define the regulation of water supply management for electricity generation and adequate consumption. It will also help plan to increase the productivity of agriculture and specify a species of plants that are determined by the amount of water in the dam in the future.

Keywords : Forecasting, Amount of water in the dam, Data Mining

1. บทนำ

1.1 ความเป็นมาและความสำคัญ

จากข้อมูลสถานการณ์น้ำของฝ่ายประมวลและวิเคราะห์สถานการณ์น้ำ กรมชลประทาน ณ วันที่ 7 มิถุนายน พ.ศ. 2564 พบว่า เขื่อนทั้งหมดทั่วประเทศปริมาณน้ำในเขื่อนรวมกันมีอยู่ทั้งสิ้น 32,339 ล้านลูกบาศก์เมตร โดยคิดเป็น 46% ของปริมาณความจุน้ำเก็บกักในเขื่อนทั้งหมด 70,926 ล้านลูกบาศก์เมตรทั่วประเทศ ซึ่งในจำนวนนี้มีปริมาณน้ำใช้การได้อยู่เพียง 8,797 ล้านลูกบาศก์เมตร โดยคิดเป็น 12% ของปริมาณความจุน้ำเก็บกักในเขื่อนทั้งหมดรวมกัน โดยเฉพาะในส่วนของเขื่อนและอ่างเก็บน้ำขนาดใหญ่ในพื้นที่ภาคเหนือและภาคตะวันตกหลายแห่งเหลือปริมาณน้ำใช้การได้อยู่ที่ 6% ของปริมาณความจุที่ระดับน้ำเก็บกักได้ ซึ่งเขื่อนภูมิพลเป็นเขื่อนที่เหลือปริมาณน้ำใช้การได้น้อยที่สุด คือ เหลืออยู่เพียง 4,334 ล้านลูกบาศก์เมตร คิดเป็นสัดส่วน 32% ของปริมาณความจุที่ระดับน้ำเก็บกักในเขื่อนทั้งหมด 13,462 ล้านลูกบาศก์เมตร จะเห็นได้ว่าสถานการณ์ของปริมาณน้ำไหลลงเขื่อนและอ่างเก็บน้ำต่าง ๆ นั้นลดน้อยลง ทำให้เหลือปริมาณน้ำใช้การได้น้อยซึ่งเป็นผลจากภาวะฝนทิ้งช่วงและน้ำต้นทุนในเขื่อนหลักอยู่ในเกณฑ์น้อย ส่งผลกระทบต่อการบริหารจัดการน้ำและเกิดปัญหาปริมาณน้ำไม่เพียงพอที่จะสนับสนุนเกษตรกรรมที่เพาะปลูกข้าวนาปี อีกทั้งในบางพื้นที่ขาดแคลนน้ำในการผลิตน้ำประปา (สำนักข่าวออนไลน์ไทยพับลิก้า, 2564)

เขื่อนภูมิพลเป็นเขื่อนคอนกรีตโค้งเพียงแห่งเดียวในประเทศไทย อีกทั้งยังเป็นเขื่อนกักเก็บน้ำขนาดใหญ่ และเขื่อนอเนกประสงค์แห่งแรกในประเทศไทย ซึ่งมีความจุสูงสุด 13,462 ล้านลูกบาศก์เมตร สร้างปิดกันแม่น้ำปิงที่บริเวณเขาแก้ว อำเภอสามเงา จังหวัดตาก อยู่ในความรับผิดชอบของการไฟฟ้าฝ่ายผลิตแห่งประเทศไทย (กฟผ.) โดยมีเป้าหมายคือ การระบายน้ำและการผลิตกระแสไฟฟ้า ซึ่งเขื่อนภูมิพลมีความสำคัญในการกักเก็บและบริหารจัดการควบคุมปริมาณน้ำให้เหมาะสมกับความต้องการของแต่ละพื้นที่ให้เพียงพอในการเกษตรช่วงฤดูแล้ง และการอุปโภคบริโภคในลุ่มน้ำเจ้าพระยา ภาคกลางลงมาจนถึงกรุงเทพมหานคร และปริมณฑล นอกจากนี้น้ำที่ถูกระบายออกจากเขื่อนภูมิพลยังเป็นประโยชน์ในการรักษาระบบนิเวศ เช่น ผลักดันน้ำเค็มในพื้นที่บริเวณแม่น้ำเจ้าพระยา ปากอ่าวไทย (การไฟฟ้าฝ่ายผลิตแห่งประเทศไทย, 2561) โดยจะเห็นได้ว่าทรัพยากรน้ำที่ไหลเข้าเขื่อนภูมิพลมีผลกระทบโดยตรงต่อปริมาณการใช้น้ำทำเขื่อนอย่างชัดเจน ทั้งในด้านการผลิตกระแสไฟฟ้า ด้านการเกษตร และด้านการผลิตน้ำประปา และจากข้อมูลในปี พ.ศ. 2550 ถึง ปี พ.ศ. 2559 พบว่า ทรัพยากรน้ำไหลเข้าในเขื่อนภูมิพลมีความเสื่อมโทรมลงอย่างต่อเนื่อง ซึ่งได้ส่งผลกระทบโดยตรงทางเศรษฐกิจโดยคิดเป็นมูลค่าระหว่าง 166,594.7 ถึง 233,155.48 ล้านบาท (นิพนธ์พันธุ์ ศิริธร, 2561) อีกทั้งเขื่อนภูมิพลยังเป็น 1 ใน 4 เขื่อนหลักที่มีผลต่อปริมาณน้ำต้นทุนสำหรับพื้นที่ลุ่มน้ำเจ้าพระยาตอนบน ซึ่งในปี พ.ศ. 2554 ปริมาณน้ำไหลเข้าสะสมของอ่างเก็บน้ำขนาดใหญ่ในประเทศไทยรวมถึงเขื่อนภูมิพลเป็นตัวแปรหนึ่งที่ส่งผลต่อการเกิดอุทกภัยในพื้นที่ภาคเหนือและภาคกลาง โดยในช่วงต้นเดือนเมษายน ปี พ.ศ. 2554 เริ่มมีปริมาณน้ำไหลเข้าเขื่อนภูมิพลสะสมเพิ่มขึ้นอย่างชัดเจนและมากกว่าในช่วงเวลาเดียวกันของปี พ.ศ. 2538 พ.ศ.2549 และ พ.ศ. 2553 ซึ่งเป็นปีที่เกิดอุทกภัยหนัก (สำนักบริหารจัดการน้ำและอุทกวิทยา, 2563)

จากที่กล่าวมาข้างต้นจะเห็นได้ว่าการพยากรณ์ปริมาณน้ำในเขื่อนภูมิพลมีความสำคัญเป็นอย่างยิ่ง เพราะสามารถใช้ในการประกอบการตัดสินใจและกำหนดแนวทางการพัฒนาระบบการบริหารจัดการน้ำเพื่อการผลิตกระแสไฟฟ้า การอุปโภค บริโภค และการเกษตรให้เหมาะสมและเพียงพอโดยเฉพาะในช่วงฤดูแล้ง รวมถึงเพื่อเป็นข้อมูลใช้ประกอบการวางแผนเฝ้าระวังความเสี่ยงที่จะเกิดอุทกภัยในช่วงฤดูฝน ดังนั้น ผู้วิจัยจึงสนใจศึกษาการสร้างตัวแบบสำหรับการพยากรณ์ปริมาณน้ำในเขื่อนภูมิพล จังหวัดตาก ด้วยเทคนิคการทำเหมืองข้อมูล โดยใช้ข้อมูลรายวันจากรายงานสถานภาพน้ำในเขื่อนของคลังข้อมูลน้ำแห่งชาติและข้อมูลทางอุตุนิยมิวิทยาของสถานีกรมอุตุนิยมิวิทยาภาคเหนือ ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2554 ถึง 31 พฤษภาคม พ.ศ. 2564 เพื่อนำมาสร้างและเปรียบเทียบประสิทธิภาพของตัวแบบสำหรับการพยากรณ์ปริมาณน้ำในเขื่อนภูมิพล จังหวัดตาก โดยเทคนิคการทำเหมืองข้อมูลที่นำมาใช้ในการพยากรณ์ประกอบด้วย 6 เทคนิค ได้แก่ เทคนิคการถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression Technique) เทคนิคแรนดอมฟอเรสต์ (Random Forest Technique) เทคนิคโครงข่ายประสาทเทียม (Artificial Neural Networks Technique; ANN) เทคนิคจำลองต้นไม้เฒ่าไฟว์พี (M5P Model Tree Technique; M5P) เทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor Algorithm Technique; K-NN) และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines Technique; SVM) และในแต่ละเทคนิคได้ใช้วิธีแบ่งข้อมูลเพื่อเปรียบเทียบประสิทธิภาพของตัวแบบการพยากรณ์ด้วยวิธีการตรวจสอบไขว้ (k-Fold Cross Validation) ที่กำหนดให้ค่า k เท่ากับ 5 10 15 และ 20 (Dutta & Bandyopadhyay, 2020; Pires et al., 2020; Zhang et al., 2021) และสำหรับการเปรียบเทียบประสิทธิภาพของวิธีการตรวจสอบไขว้ ในแต่ละเทคนิคการทำเหมืองข้อมูล ผู้วิจัยใช้การเลือกเทคนิคที่เหมาะสมที่สุดจากเกณฑ์ค่าต่ำที่สุดของค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error; **MAE**) และในการหาตัวแบบการพยากรณ์ที่มีประสิทธิภาพมากที่สุดของเทคนิคการทำเหมืองข้อมูลทั้ง 6 เทคนิคดังกล่าว ผู้วิจัยใช้การพิจารณาจากเกณฑ์ค่าต่ำที่สุดของค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยและค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error; **RMSE**)

1.2 วัตถุประสงค์การวิจัย

ในการวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของตัวแบบการพยากรณ์ปริมาณน้ำในเขื่อนภูมิพล จังหวัดตาก

2. เอกสารงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

การพยากรณ์ คือ การคาดคะเนหรือการทำนายลักษณะของการเกิดเหตุการณ์หรือสภาพการณ์ในอนาคต โดยศึกษารูปแบบการเกิดเหตุการณ์หรือสภาพการณ์จากข้อมูลที่เก็บรวบรวมไว้อย่างมีระบบหรือจากความรู้ ความสามารถ ประสบการณ์ และวิจารณญาณของผู้พยากรณ์ (ทรงศิริ แต่สมบัติ, 2549)

เหมืองข้อมูล หมายถึง กระบวนการในการค้นหาความรู้และรูปแบบที่สนใจจากข้อมูลจำนวนมาก ที่แหล่งข้อมูลมาจากรูปร่างข้อมูล คลังข้อมูล เว็บไซต์ และแหล่งอื่น ๆ หรือข้อมูลที่ถูกจัดเก็บในระบบพลวัต (Han et al., 2011) โดยการทำเหมืองข้อมูลแบ่งออกเป็น 4 ด้าน คือ ด้านการจำแนกประเภท (Classification) ด้านการจัดกลุ่ม (Clustering) ด้านการพยากรณ์ (Prediction) ด้านกฎความสัมพันธ์ (Association Rule) ปัจจุบันการทำเหมืองข้อมูล ได้ถูกนำไปประยุกต์ใช้ในงานหลากหลายประเภทที่สามารถช่วยในการตัดสินใจ เช่น ปัญหาเกี่ยวกับการเงินธุรกิจ ปัญญา และเศรษฐศาสตร์ เป็นต้น (อัมเรศ ริทราณี และคณะ, 2563; รุ่งนภา สุทธิชัย และคณะ, 2561)

วิธีการตรวจสอบไขว้เป็นวิธีการแบ่งข้อมูลเพื่อนำมาใช้ทดสอบประสิทธิภาพของตัวแบบพยากรณ์ โดยเริ่มจากการแบ่งข้อมูลออกเป็นจำนวน k ส่วนเท่า ๆ กัน จากนั้นข้อมูลหนึ่งส่วน เรียกว่า ชุดข้อมูลทดสอบ (Test Data Set) จะถูกนำไปใช้เป็นตัวทดสอบประสิทธิภาพของตัวแบบการพยากรณ์ และข้อมูลจำนวน $k - 1$ ส่วนที่เหลือ เรียกว่า ชุดข้อมูลฝึกฝน (Training Data Set) จะถูกนำไปสร้างตัวแบบพยากรณ์ โดยจะนำข้อมูลหนึ่งส่วนนั้นไปทดสอบประสิทธิภาพจนครบจำนวนทั้งหมด k ส่วน สุดท้ายจะได้ตัวแบบพยากรณ์ครบ k ตัวแบบ แล้วนำค่าความถูกต้องเฉลี่ยจากตัวแบบทั้งหมด เพื่อใช้เป็นตัวแทนของค่าความถูกต้องของข้อมูลทั้งหมด (Zhang et al., 2021)

เทคนิคการถดถอยเชิงเส้นพหุคูณเป็นวิธีทางสถิติที่ใช้วิเคราะห์รูปแบบความสัมพันธ์เชิงเส้นระหว่างตัวแปรตาม (Dependent Variable) และตัวแปรอิสระ (Independent or Explanatory Variables) โดยมีตัวแบบดังนี้

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \dots + \beta_k x_{k,i} + e_i \quad (1)$$

โดย y_i แทน ตัวแปรตาม

$x_{1,i}, x_{2,i}, \dots, x_{k,i}$ แทน ตัวแปรอิสระ

$\beta_0, \beta_1, \beta_2, \dots, \beta_k$ แทน สัมประสิทธิ์การถดถอย (Regression Coefficients)

e_i แทน ความคลาดเคลื่อนสุ่ม (Random Error)

ในการประมาณค่าสัมประสิทธิ์การถดถอยนิยมให้วิธีกำลังสองน้อยสุดสามัญ (Ordinary Least Squares Method) ซึ่งเป็นวิธีการที่ทำให้ผลบวกกำลังสองของความคลาดเคลื่อนสุ่มมีค่าต่ำที่สุด (Makridakis et al., 2008)

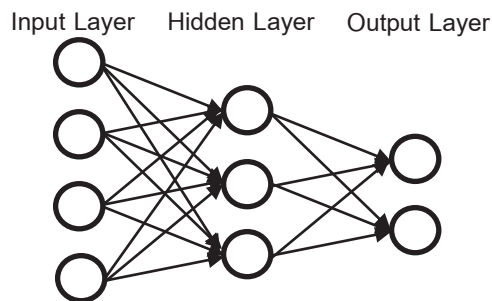
ผู้วิจัยได้วิเคราะห์โดยใช้อัลกอริทึม Linear Regression ในโปรแกรม Weka โดยกำหนดวิธีการเลือกตัวแปรอิสระเข้าในตัวแบบด้วยวิธีการเลือกแบบไปข้างหน้า (Forward Selection) และวิธีการกำจัดแบบถอยหลัง (Backward Elimination)

เทคนิคแรนดอมฟอเรสต์เป็นตัวแบบที่พัฒนามาจากตัวแบบต้นไม้ตัดสินใจ (Decision Tree) โดยจะใช้วิธีการแบ่งจำนวนต้นไม้ตัดสินใจออกเป็นหลาย ๆ ต้น (Tree) โดยแต่ละต้นสร้างมาจากการสุ่มคุณลักษณะ (Feature)

และข้อมูล (Data) บางส่วนด้วยวิธีบูตสแตรป์ (Bootstrap Method) จากคุณลักษณะและข้อมูลชุดฝึกฝนทั้งหมด ซึ่งทำให้ได้ต้นไม้ตัดสินใจที่มีความเป็นอิสระต่อกันมากขึ้น จากนั้นหาค่าพยากรณ์จากต้นไม้ตัดสินใจที่ได้ในแต่ละต้นทำซ้ำจนกว่าจะได้จำนวนต้นไม้ตัดสินใจครบตามที่กำหนด หากตัวแปรกลุ่มเป้าหมายที่สนใจศึกษาเป็นตัวแปรเชิงปริมาณซึ่งถือว่าเป็นปัญหาวิเคราะห์การถดถอย ในการหาค่าพยากรณ์สุดท้ายสามารถทำได้โดยการคำนวณหาค่าเฉลี่ยของค่าพยากรณ์ของต้นไม้ตัดสินใจทุกต้น (Granata et al., 2018) ซึ่งในการวิจัยครั้งนี้ ผู้วิจัยได้วิเคราะห์โดยใช้อัลกอริทึม Random Forest ในโปรแกรม Weka โดยกำหนดจำนวนต้นไม้ตัดสินใจ เท่ากับ 100 ต้น และจำนวนกิ่งของต้นไม้ตัดสินใจแต่ละต้นไม่เกิน 100 กิ่ง

เทคนิคโครงข่ายประสาทเทียมเป็นตัวแทนทางคณิตศาสตร์ที่จำลองจากการทำงานของเครือข่ายใยประสาทในสมองมนุษย์ซึ่งประกอบด้วยเซลล์ประสาท (Neuron) และจุดเชื่อมประสาท (Synapses) โดยเชื่อมต่อกันในรูปแบบโครงข่ายหลายชั้น แต่ละชั้นจะรับค่าที่อยู่ก่อนหน้าแล้วคำนวณด้วยฟังก์ชันแบบไม่เชิงเส้นแล้วส่งผลลัพธ์ออกไปยังเซลล์ประสาทชั้นถัดไปผ่านเส้นเชื่อมโยงจนได้ผลลัพธ์สุดท้ายของการคำนวณเครือข่าย โดยมีชั้นปกปิด (Hidden Layer) เป็นชั้นที่อยู่ตรงกลางระหว่างชั้นรับข้อมูลและส่งผลลัพธ์ ดังรูปภาพประกอบ 1 (สุรพงษ์ เอื้อวัฒนามงคล, 2562)

รูปภาพประกอบ 1
ข่ายงานของโครงข่ายประสาท



โดยกระบวนการทำงานของโครงข่ายประสาทเทียมเริ่มจากการนำเข้าสู่ข้อมูล ($X_1, X_2, \dots, X_n$) เข้ามาในโครงข่าย โดยข้อมูลที่ถูกนำเข้ามาจะถูกนำมาคูณกับค่าน้ำหนัก ($W_1, W_2, \dots, W_n$) ของแต่ละเส้นเชื่อม จากนั้นโหนดในแต่ละชั้นถัดไปจะรับค่าเพื่อนำไปคำนวณผลด้วยฟังก์ชันกระตุ้น (Activation function) แทนด้วย f ซึ่งโดยทั่วไปเป็นฟังก์ชันซิกมอยด์ (Sigmoid Function) แล้วส่งต่อผลลัพธ์ที่ได้ไปยังชั้นถัดไป ซึ่งจะเกิดขึ้นแบบนี้ซ้ำ ๆ ทีละชั้นจนถึงโหนดส่งผลลัพธ์ (Output) แทนด้วย y ซึ่งสามารถเขียนได้ดังสมการ (Vijai & Sivakumar, 2018)

$$y = f \left(\sum_{j=0}^{N_c} W_j \cdot X_j \right) \quad (2)$$

โดย W_j แทน ค่าน้ำหนักของเส้นเชื่อมที่ j

X_j แทน เวกเตอร์ของข้อมูลนำเข้า (Input vector) ของเส้นเชื่อมที่ j

N_c แทน จำนวนของเส้นเชื่อม (Input Connection)

ในโปรแกรม Weka ผู้วิจัยได้วิเคราะห์โดยใช้อัลกอริทึม Multilayer Perceptron โดยใช้ฟังก์ชันซิกมอยด์ และกำหนดพารามิเตอร์เป็นจำนวนชั้นปกปิด เท่ากับ จำนวนเต็มของผลรวมจำนวนปัจจัยกับจำนวนตัวแปร กลุ่มเป้าหมายหารด้วยสอง

เทคนิคจำลองต้นไม้เอนไซม์ไฟฟ้ถูกพัฒนามาจากต้นไม้ตัดสินใจ (Decision Tree) ซึ่งเป็นตัวแบบที่ใช้ในการพยากรณ์ตัวแปรเป้าหมายที่เป็นเชิงปริมาณและตัวแปรอื่น ๆ ที่ใช้อธิบายลักษณะได้ทั้งเชิงคุณภาพและเชิงปริมาณ ซึ่งต้นไม้เอนไซม์ไฟฟ้มีอัลกอริทึมเหมือนกับต้นไม้ตัดสินใจ แตกต่างตรงที่ต้นไม้เอนไซม์ไฟฟ้จะวิเคราะห์ข้อมูลจากโหนดราก (Root Nodes) ลงมาจนถึงโหนดใบ (Leaf Nodes) โดยในแต่ละโหนดใบแทนด้วยการถดถอยเชิงเส้น (Linear Regression) แต่ในส่วน of ต้นไม้ตัดสินใจที่โหนดใบแทนด้วยค่าคงที่ (Witten & Frank, 2005) โดยเริ่มจากขั้นตอนการจัดกลุ่มข้อมูลเชิงปริมาณจากค่าคุณลักษณะ (Attribute) โดยคำนวณหาค่าการเรียนรู้และเรียงลำดับตามค่าเฉลี่ยของกลุ่มข้อมูล ก่อนการสร้างตัวแบบการพยากรณ์ ซึ่งการเรียงลำดับจากค่าเฉลี่ยของกลุ่มข้อมูลด้วยค่าคุณลักษณะเป็นขั้นตอนการเรียงลำดับของข้อมูลที่ถูกเลือกในแต่ละโหนด จากนั้นสร้างต้นไม้ (Building Tree) โดยใช้เกณฑ์การเลือกค่าคุณลักษณะของตัวแบบเอนไซม์ไฟฟ้ที่ขึ้นอยู่กับการจัดการค่าเบี่ยงเบนมาตรฐานของกลุ่มข้อมูลที่พิจารณาจากการตัดเล็มต้นไม้จากแต่ละโหนดใบ ซึ่งจะวัดค่าความผิดพลาดที่เกิดขึ้นในแต่ละโหนด และทดสอบหาการลดลงของค่าความผิดพลาดที่เกิดขึ้นมาด้วยค่าคุณลักษณะในแต่ละโหนด ด้วยกระบวนการแบ่งข้อมูลที่จะช่วยลดข้อผิดพลาดของกลุ่มข้อมูลในแต่ละกิ่ง (Branch) ซึ่งสมการค่าเบี่ยงเบนมาตรฐาน (Standard Deviation Reduction: SDR) ที่ใช้คำนวณมีดังต่อไปนี้

$$SDR = sd(S) - \sum_i \frac{S_i}{|S|} \times sd(S_i) \quad (3)$$

โดยที่ S แทน ชุดข้อมูลตัวอย่างที่ไปยังโหนด

S_i แทน ชุดข้อมูลที่ได้จากการแบ่งโหนดตามค่าคุณลักษณะ

sd แทน ส่วนเบี่ยงเบนมาตรฐานของชุดตัวอย่าง

จากนั้นสร้างตัวแบบการถดถอยเชิงเส้นขึ้นในแต่ละกิ่งที่ได้ และพิจารณาการตัดเล็มต้นไม้ (Pruning Tree) เพราะในการสร้างต้นไม้อาจเกิดส่วนของโหนดใบมากเกินไป หรือตัวแบบสร้างขึ้นให้ผลการพยากรณ์ที่ตีมาก (Overfitting) จนไม่สามารถนำไปใช้กับชุดข้อมูลอื่นได้ เพื่อแก้ไขปัญหานี้จึงมีขั้นตอนของการตัดเล็มต้นไม้ที่จะใช้แบบจำลองเชิงเส้น (Linear Model) และขั้นตอนต่อไปเป็นการปรับเรียบ (Smoothing Process) เพื่อลดความไม่ต่อเนื่องของตัวแบบเชิงเส้นในส่วน of โหนดใบ หรือส่วนย่อยต่าง ๆ ที่เป็นผลมาจากขั้นตอนการตัดเล็มต้นไม้ และมีจำนวนข้อมูลการเรียนรู้ที่น้อย ซึ่งการปรับเรียบจะใช้ตัวแบบของโหนดใบในการคำนวณหาค่าประมาณ หลังจากนั้นกรอกข้อมูลจากเส้นทางไปยังโหนดราก ต่อจากนั้นปรับเรียบแต่ละโหนดด้วยการรวมค่าข้อมูลกับค่าที่ได้จากการประมาณ (Behnood et al., 2017) ด้วยสมการต่อไปนี้

$$E' = \frac{ne - ka}{n + k} \quad (4)$$

โดย E' แทน การประมาณกลับไปโหนดที่สูงกว่า

e แทน การประมาณกลับมาโหนดที่ต่ำกว่า

a แทน ค่าประมาณจากตัวแบบของโหนดชั้นเดียวกัน

n แทน จำนวนตัวอย่างการเรียนรู้ในโหนดที่ต่ำกว่า

k แทน ค่าคงที่

ผู้วิจัยได้วิเคราะห์โดยใช้อัลกอริทึม M5P ในโปรแกรม Weka โดยกำหนดจำนวนกิ่งไม่เกิน 100 กิ่งและจำนวนตัวอย่างการเรียนรู้ในแต่ละโหนดขั้นต่ำเท่ากับ 4 ตัวอย่าง

เทคนิคเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbors) เป็นเทคนิคการจำแนกที่จะพิจารณาจากชุดข้อมูลที่มีความคล้ายกัน ซึ่งเป็นวิธีที่เหมาะสมสำหรับข้อมูลเชิงปริมาณและสามารถใช้ค่ามาตรฐานที่แตกต่างกันตามความเหมาะสมของข้อมูล โดยอาศัยวัดเป็นระยะทาง (Distance) ถ้าระยะทางระหว่างข้อมูลที่ต้องการพิจารณากับข้อมูลชุดใดมีค่าน้อยที่สุด ก็จะพิจารณาว่าข้อมูลนั้นควรจัดอยู่ในหมวดหมู่ชนิดเดียวกัน ซึ่งเป็นการค้นหาข้อมูลในฐานที่มีลักษณะใกล้เคียงกับข้อมูลใหม่มากที่สุดจำนวน K ตัว โดยวัดจากความคล้ายหรือระยะทาง โดยปกติแล้วจำนวน K ที่มีค่ามากหรือน้อยเกินไปมีผลทำให้การตัดสินใจประเภทข้อมูลมีความผิดพลาดได้ การหาจำนวน K ที่เหมาะสมที่สุดจึงจำเป็นต้องทำเพื่อเพิ่มประสิทธิภาพในการจำแนก (Kantardzic, 2011) โดยระยะทางที่นิยมใช้มากที่สุดคือระยะทางแบบยูคลิด (Euclidean Distance) สามารถคำนวณได้ดังสมการ

$$dist(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (5)$$

โดย $dist(p, q)$ แทน ระยะทางระหว่างข้อมูล p และ q

p_i แทน ค่าของข้อมูลคุณสมบัติที่ i ของข้อมูล p

q_i แทน ค่าของข้อมูลคุณสมบัติ i ของข้อมูล q

n แทน จำนวนมิติข้อมูล

ในโปรแกรม Weka ผู้วิจัยได้วิเคราะห์โดยใช้อัลกอริทึม IBK โดยกำหนดค่าพารามิเตอร์ K เท่ากับ 1 ถึง 20 เพื่อหาค่า K ที่เหมาะสมที่สุดสำหรับการสร้างตัวแบบการพยากรณ์

เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines: SVM) อาศัยหลักการของการหาสมบัติของสมการเพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลชุดเรียนรู้ (Training Data) ที่ถูกป้อนเข้าสู่กระบวนการเพื่อสอนให้ระบบเรียนรู้ โดยเน้นไปยังเส้นแบ่งข้อมูลซึ่งสามารถใช้เป็นมาตรฐานในการแยกแยะกลุ่มข้อมูลได้ดีที่สุด มักใช้ในการแบ่งข้อมูลออกเป็น 2 กลุ่ม โดยจะใช้กับชุดข้อมูลที่มีมิติของข้อมูลที่สูง กล่าวคือเป็นการที่นำค่าของกลุ่มข้อมูลมาวางลงในฟีเจอรัสเปซ (Feature Space) จากนั้นจึงหาเส้นที่แบ่งข้อมูลทั้งสองออกจากกันและใช้ฟังก์ชันเคอร์เนล (Kernel Function) ในการวัดความคล้ายกันของข้อมูล จากนั้นจะสร้างเส้นแบ่ง (Hyper Plane) ที่เป็นเส้นตรงขึ้นมาและหาว่าเส้นตรงใดเป็นเส้นที่ดีที่สุด เพื่อใช้ในการแบ่งชุดข้อมูลสองกลุ่มออกจากกัน (Tan et al., 2016) กำหนดให้ข้อมูลชุดเรียนรู้ $I_i = \{(x_i, y_i) | i = 1, 2, \dots, N\}$ เมื่อ $x_i \in R^q$ และตัวแปรเป้าหมาย $y_i \in R$ เมื่อ N แทน จำนวนข้อมูลชุดเรียนรู้ และ q แทน จำนวนของมิติข้อมูลเข้า โดยตัวแบบการถดถอยสามารถคำนวณหาได้จากสมการนี้

$$y = W^T \theta(x) + b$$

โดย W แทน ค่าน้ำหนัก (Weight) ที่เชื่อมโยงจากฟีเจอรัสเปซไปยังเอาต์พุตสเปซ

b แทน ส่วนประกอบของความเอนเอียง (Bias Component)

$\theta(x)$ แทน ฟังก์ชันการส่งไม่เชิงเส้น (Nonlinear Mapping Function)

การจำแนกกลุ่มด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีนจะประกอบด้วย 2 ส่วน ได้แก่ ส่วนของการเพิ่มระยะแยกแยะที่มากที่สุด และส่วนของการแก้ปัญหาทั้งแบบเชิงเส้นและแบบไม่เชิงเส้นด้วยการลดข้อผิดพลาดให้เหลือน้อยที่สุด สามารถแสดงได้ดังนี้

$$\text{Min} \left\{ \frac{1}{2} W^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \right\} \quad (7)$$

โดยมีเงื่อนไข

$$y_i - \{(W^T \theta(x_i)) + b\} \leq 1 - \xi_i, \quad i = 1, 2, \dots, N$$

$$\xi_i, \xi_i^* \geq 0, \quad i = 1, 2, \dots, N$$

โดย W แทน ค่าน้ำหนัก (Weight) ที่เชื่อมโยงจากพีเจอร์สเปซไปยังเอาท์พุตสเปซ

C แทน ค่าพารามิเตอร์ที่ควบคุมความผิดพลาด (Penalty Parameter)

b แทน ส่วนประกอบของความเอนเอียง (Bias Component)

ξ_i และ ξ_i^* แทน ตัวแปรใช้วัดดีกรี (Slack Variable) ของการจำแนกข้อมูลที่ผิดพลาดของ x_i

สุดท้ายจะได้สมการถดถอยซัพพอร์ตเวกเตอร์ที่ใช้พยากรณ์ค่าตัวแปรเป้าหมาย คือ

$$f(x) = \sum_{i=1}^N (\delta_i + \delta_i^*) k(x_i, x_j) + b \quad (8)$$

โดย δ_i และ δ_i^* แทน สัมประสิทธิ์ลากรางจ์ (Lagrangian Coefficients)

$k(x_i, x_j)$ แทน ฟังก์ชันเคอร์เนลของข้อมูล x_i และ x_j โดยที่ x_i จะถูกส่งต่อไปยังสเปซที่มีมิติสูงกว่าด้วยฟังก์ชันการส่งไม่เชิงเส้น (Ahmadi & Khashei, 2021)

ในงานวิจัยนี้ใช้โปรแกรม Weka ด้วยอัลกอริทึม SMO โดยกำหนดฟังก์ชันเคอร์เนลพหุนาม (Polynomial Kernel Function) และกำหนดค่าพารามิเตอร์ที่ควบคุมความผิดพลาด (C) เท่ากับ 1

2.2 งานวิจัยที่เกี่ยวข้อง

การประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลสำหรับการพยากรณ์ปริมาณน้ำเป็นที่แพร่หลายในปัจจุบัน เช่น งานวิจัยของ Guyennon et al. (2021) ที่ประยุกต์ใช้เทคนิคการทำเหมืองข้อมูล พบว่า เทคนิคแรนดอมฟอเรสต์ เหมาะสมที่สุดสำหรับการพยากรณ์ปริมาณน้ำของทะเลสาบบราซโซ (Lake Bracciano) ในตอนกลางของประเทศ อิตาลี โดยใช้ปัจจัยในการศึกษา คือ ปริมาณน้ำฝน อุณหภูมิเฉลี่ย อุณหภูมิสูงสุดและต่ำสุด ความชื้นสัมพัทธ์ และความกดอากาศ รวมทั้งเทคนิคเดียวกันนี้มีประสิทธิภาพมากที่สุดในผลงานวิจัยของ Zang et al. (2021) ที่ศึกษาการพยากรณ์ปริมาณน้ำของทะเลสาบไท่ (Lake Tai) ในบริเวณสามเหลี่ยมปากแม่น้ำแยงซีของประเทศจีน โดยใช้ปัจจัยในการศึกษา คือ ความเร็วลม ทิศทางลม อุณหภูมิและปริมาณน้ำฝน งานวิจัยของ Wang และ Wang (2020) ที่ศึกษาการพยากรณ์ปริมาณน้ำของทะเลสาบอีรี (Lake Erie) ที่เป็นพรมแดนธรรมชาติระหว่างประเทศแคนาดา与美国 สหราชอาณาจักร โดยใช้ปัจจัยในการศึกษา คือ ความเร็วลม อุณหภูมิ รังสีคลื่นยาว รังสีคลื่นสั้น ความชื้นสัมพัทธ์และปริมาณน้ำฝน ผลการศึกษาพบว่า เทคนิคการถดถอยเชิงเส้นพหุคูณและเทคนิคจำลองต้นไม้เอนิมโฟว์ที่เป็นเทคนิคที่มีประสิทธิภาพในการพยากรณ์มากที่สุด

ในประเทศไทยได้มีการนำเทคนิคการทำเหมืองข้อมูลมาประยุกต์ใช้ในการพยากรณ์ปริมาณน้ำในเขื่อน เช่น งานวิจัยของ จิรโรจน์ ตอสะสุกุล และ สุพิชชา ชัดธิพงษ์ (2564) ได้ศึกษาการพยากรณ์ปริมาณน้ำในเขื่อนภูมิพล จังหวัดตาก โดยใช้ปัจจัยในการศึกษา คือ ค่าเฉลี่ยน้ำรวมทั้งปี ทิศทางลม ปริมาณน้ำไหลลงอ่าง ปริมาณน้ำระบาย ความกดอากาศ อุณหภูมิเฉลี่ย ความชื้นสัมพัทธ์ ปริมาณน้ำฝน ปริมาณน้ำระเหย และกำลังลมสูงสุด ผลการศึกษาพบว่า ตัวแบบการพยากรณ์ของเทคนิคแบบจำลองต้นไม้เอนิมโฟว์ที่เป็นเทคนิคที่เหมาะสมที่สุด และงานวิจัยของ อนุชิต เนาแสง และคณะ (2563) ได้ศึกษาการพยากรณ์ปริมาณน้ำในเขื่อนอุบลรัตน์ จังหวัดขอนแก่น โดยมีปัจจัยที่ใช้ในการศึกษา คือ ปริมาณน้ำในอ่าง ปริมาณน้ำปัจจุบัน ระดับน้ำเก็บกักของอ่างปัจจุบัน ปริมาณน้ำที่ใช้การได้จริงระดับ

น้ำเก็บกักของอ่างใช้การได้จริง ปริมาณน้ำลงอ่าง ปริมาณน้ำสะสมเทียบกับค่าเฉลี่ยทั้งปีสะสม ปริมาณน้ำระบายวันนี้ ค่าเฉลี่ยน้ำรวมทั้งปีและปริมาณน้ำระบายสะสม ผลการวิจัยพบว่า เทคนิคที่เหมาะสมสำหรับการสร้างตัวแบบการพยากรณ์ปริมาณน้ำในเขื่อนอุบลรัตน์ คือ เทคนิควิธีเพื่อนบ้านใกล้ที่สุด และรองลงมาคือ เทคนิคโครงข่ายประสาทเทียม รวมทั้งงานวิจัยของ วีรศักดิ์ ฟองเงิน และคณะ (2560) ได้ศึกษาการพยากรณ์ปริมาณน้ำในเขื่อนก๊วลม จังหวัดลำปาง โดยมีปัจจัยที่ใช้ในการศึกษา คือ ปริมาณน้ำไหลเข้าเขื่อน ปริมาณน้ำในเขื่อน ปริมาณการปล่อยน้ำและอัตราการละเหย ผลการวิจัยพบว่า วิธีแบบจำลองต้นไม้เอมไพร์เป็นแบบที่เหมาะสมที่สุด รองลงมาคือ เทคนิคชัพพอตเวกเตอร์แมชชีน

จากงานวิจัยที่กล่าวข้างต้นผู้วิจัยสนใจที่จะขยายผลการศึกษาของ จิรโรจน์ ตอสะสุกุล และ สุพิชชา ชัดธิพงษ์ (2564) ในการเปรียบเทียบประสิทธิภาพตัวแบบสำหรับการพยากรณ์ปริมาณน้ำในเขื่อนภูมิพล จังหวัดตาก โดยใช้เทคนิคการทำเหมืองข้อมูล 6 เทคนิค ได้แก่ เทคนิคการถดถอยเชิงเส้นพหุคูณ เทคนิคแรนดอมฟอเรสต์ เทคนิคโครงข่ายประสาทเทียม เทคนิคจำลองต้นไม้เอมไพร์ เทคนิคเพื่อนบ้านใกล้ที่สุด และเทคนิคชัพพอตเวกเตอร์แมชชีน และในแต่ละเทคนิคได้เพิ่มเติมในส่วนของการศึกษาเปรียบเทียบการแบ่งข้อมูลด้วยวิธีการตรวจสอบไขว้ เพื่อให้ได้ตัวแบบสำหรับการพยากรณ์ที่มีประสิทธิภาพมากที่สุด

3. วิธีการดำเนินงานวิจัย

เครื่องมือที่ใช้ในการศึกษาและสร้างตัวแบบสำหรับการพยากรณ์ปริมาณน้ำในเขื่อนภูมิพล จังหวัดตาก ประกอบด้วย Google Sheets ซึ่งใช้สำหรับจัดเก็บข้อมูลปริมาณน้ำในเขื่อนภูมิพลและแอตทริบิวต์ที่ใช้ในการศึกษาทั้งหมดจำนวน 12 แอตทริบิวต์ (Attribute) ทั้งจากคลังข้อมูลน้ำแห่งชาติและสถานีกรมอุตุนิยมวิทยาภาคเหนือ และโปรแกรม Weka เวอร์ชัน 3.9 ถูกใช้เพื่อการเตรียมข้อมูลและการวิเคราะห์สร้างตัวแบบในการพยากรณ์ปริมาณน้ำในเขื่อนภูมิพลรวมถึงการวิเคราะห์เปรียบเทียบประสิทธิภาพเทคนิคที่ใช้ในการพยากรณ์ โดยผู้วิจัยได้ดำเนินงานวิจัยวิเคราะห์ข้อมูลตามกระบวนการมาตรฐานในการทำเหมืองข้อมูล (Cross-Industry Standard Process for Data Mining หรือ CRISP-DM) โดยแบ่งออกเป็น 6 ขั้นตอนดังนี้

3.1 ขั้นตอนการทำความเข้าใจปัญหา (Business Understanding)

จากการศึกษางานวิจัยที่เกี่ยวข้องและแอตทริบิวต์ที่คาดว่าจะเกี่ยวกับปริมาณน้ำในเขื่อนภูมิพล พบว่าสถานะฝนทั้งช่วงและน้ำต้นทุนในเขื่อนน้อยเป็นสาเหตุให้เกิดปัญหาปริมาณน้ำไม่เพียงพอที่จะสนับสนุนเกษตรกรรมในการเพาะปลูกข้าวนาปีและพืชเศรษฐกิจอื่น ๆ ปัญหาการขาดแคลนน้ำในการผลิตกระแสไฟฟ้า การผลิตน้ำประปาเพื่ออุปโภคและบริโภค รวมไปถึงส่งผลกระทบต่อสถานการณ์น้ำเค็มรุกฉ่ำในแม่น้ำเจ้าพระยาอีกด้วย อีกทั้งเขื่อนภูมิพลยังมีผลต่อปริมาณน้ำต้นทุนสำหรับพื้นที่ลุ่มน้ำเจ้าพระยาตอนบนโดยปริมาณน้ำไหลเข้าสะสมของเขื่อนภูมิพลถือว่าเป็นตัวแปรหนึ่งที่มีผลต่อการเกิดอุทกภัยในพื้นที่ภาคเหนือและภาคกลาง

3.2 ขั้นตอนการทำความเข้าใจข้อมูล (Data Understanding)

ข้อมูลที่นำมาใช้ในการสร้างตัวแบบพยากรณ์เป็นการเก็บรวบรวมข้อมูลทุติยภูมิเกี่ยวกับปริมาณน้ำในเขื่อนภูมิพลและข้อมูลทางอุตุนิยมวิทยารายวัน ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2554 ถึง 31 พฤษภาคม พ.ศ. 2564 จากคลังข้อมูลน้ำแห่งชาติและสถานีกรมอุตุนิยมวิทยาภาคเหนือ ซึ่งถูกจัดเก็บอยู่ในรูปของ Google Sheets จำนวนทั้งหมด 3,804 รายการ และ 12 แอตทริบิวต์ แสดงในตาราง 1

3.3 ขั้นตอนการเตรียมข้อมูล (Data Preparation)

การจัดเตรียมและคัดเลือกแอตทริบิวต์ที่มีความสำคัญต่อการนำไปวิเคราะห์ข้อมูลเพื่อใช้ในการพยากรณ์ปริมาณน้ำในเขื่อนภูมิพลด้วยเทคนิคการทำเหมืองข้อมูล มีขั้นตอนการเตรียมข้อมูลดังต่อไปนี้

3.3.1 นำเข้าข้อมูลสำหรับการวิเคราะห์ที่เป็นข้อมูลดิบ (Raw Data) โดยเก็บรวบรวมข้อมูลทุติยภูมิจากคลังข้อมูลน้ำแห่งชาติและสถานีกรมอุตุนิยมวิทยาภาคเหนือ ให้อยู่ในรูปของ Google Sheets หลังจากนั้นส่งออกข้อมูลมาเก็บในชนิดไฟล์ csv เพื่อเตรียมข้อมูลให้พร้อมสำหรับการประมวลผลในโปรแกรม Weka

3.3.2 คัดเลือกแอตทริบิวต์ที่ใช้ในการสร้างตัวแบบสำหรับการพยากรณ์ปริมาณน้ำในเขื่อนภูมิพล โดยตัดแอตทริบิวต์ที่ไม่มีผลต่อการสร้างตัวแบบหรือที่ไม่มีความเคลื่อนไหวหรือเป็นค่าคงที่ และในการทำความสะอาดข้อมูลไม่พบข้อมูลสูญหาย (Missing Data) ของทุกแอตทริบิวต์ในช่วงเวลาที่ศึกษา

3.3.3 กำหนดและตรวจสอบความถูกต้องของประเภทข้อมูล (Data Type) ที่จัดเก็บในแต่ละแอตทริบิวต์แสดงในตาราง 1

3.3.4 กำหนดหน้าที่ (Set Role) ของแอตทริบิวต์ โดยกำหนดให้แอตทริบิวต์วันที่ (date) มีหน้าที่เป็นรหัส (ID) และกำหนดให้แอตทริบิวต์ปริมาณน้ำในอ่าง (amount_of_water) มีหน้าที่เป็นตัวแปรกลุ่มเป้าหมาย (Class) หรือตัวแปรตาม (Dependent Variable) สำหรับการสร้างตัวแบบพยากรณ์ซึ่งเป็นข้อมูลเชิงปริมาณโดยมีหน่วยเป็นล้านลูกบาศก์เมตร

3.3.5 สำหรับเทคนิคการทำเหมืองข้อมูล 6 เทคนิค ได้แก่ เทคนิคการถดถอยเชิงเส้นพหุคูณ เทคนิคแรนดอมฟอเรสต์ เทคนิคโครงข่ายประสาทเทียม เทคนิคจำลองต้นไม้เอมไพริคัล เทคนิคเพื่อนบ้านใกล้ที่สุด และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน ผู้วิจัยได้เลือกแอตทริบิวต์ที่มีความสัมพันธ์สูงกับตัวแปรกลุ่มเป้าหมายด้วยวิธี Classifier Subset Evaluator ซึ่งพิจารณาจากเกณฑ์ค่าน้ำหนักของแต่ละปัจจัยที่มีค่าตั้งแต่ 50 เปอร์เซนต์ขึ้นไป ในส่วนเทคนิคการถดถอยเชิงเส้นพหุคูณผู้วิจัยได้เลือกแอตทริบิวต์ด้วยวิธีการเลือกแบบไปข้างหน้าและวิธีการกำจัดแบบถอยหลัง

ตาราง 1 แสดงรายละเอียดแอตทริบิวต์ที่ใช้ในการศึกษาทั้งหมด

ลำดับ	แอตทริบิวต์	ประเภทข้อมูล	คำอธิบาย
1	date	วันที่ (Date)	วัน/เดือน/ปี
2	amount_of_water	จำนวนจริง (Real)	ปริมาณน้ำในอ่าง (ล้านลูกบาศก์เมตร)
3	average_peryear	จำนวนจริง (Real)	ค่าเฉลี่ยน้ำในเขื่อนรวมทั้งปี (ล้านลูกบาศก์เมตร)
4	water_flow	จำนวนจริง (Real)	ปริมาณน้ำไหลลงอ่าง (ล้านลูกบาศก์เมตร)
5	drainage_water	จำนวนจริง (Real)	ปริมาณน้ำระบาย (ล้านลูกบาศก์เมตร)
6	air_pressure	จำนวนจริง (Real)	ความกดอากาศเฉลี่ย
7	temperature	จำนวนจริง (Real)	อุณหภูมิเฉลี่ย (เซลเซียส)
8	moisture	จำนวนจริง (Real)	ความชื้นเฉลี่ย (%)
9	rain	จำนวนจริง (Real)	ปริมาณฝน (มิลลิเมตร)
10	evaporated_water	จำนวนจริง (Real)	น้ำระเหย (มิลลิเมตร)
11	wind_direction	จำนวนจริง (Real)	ทิศทางลมสูงสุด
12	wind_power	จำนวนจริง (Real)	กำลังลมสูงสุด (กิโลเมตรต่อชั่วโมง)

3.4 ขั้นตอนการสร้างตัวแบบ (Modeling)

สร้างตัวแบบพยากรณ์ด้วยเทคนิคการทำเหมืองข้อมูล 6 เทคนิค ได้แก่ เทคนิคการถดถอยเชิงเส้นพหุคูณ เทคนิคแรนดอมฟอเรสต์ เทคนิคโครงข่ายประสาทเทียม เทคนิคจำลองต้นไม้เอมไพร์ฟิ เทคนิคเพื่อนบ้านใกล้ที่สุด และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน ด้วยโปรแกรม Weka โดยแบ่งข้อมูลสำหรับเป็นชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบด้วยวิธีการตรวจสอบไขว้ของแต่ละเทคนิค โดยกำหนดให้ค่า k เท่ากับ 5 10 15 และ 20 เพื่อหาตัวแบบการพยากรณ์ที่เหมาะสมที่สุดในแต่ละเทคนิคการทำเหมืองข้อมูล จากนั้นเปรียบเทียบประสิทธิภาพของตัวแบบการพยากรณ์ของแต่ละเทคนิคในทุกกรณีที่ได้ เพื่อหาตัวแบบการพยากรณ์ที่มีประสิทธิภาพมากที่สุด

3.5 ขั้นตอนการประเมินผล (Evaluation)

ในการแบ่งข้อมูลเพื่อหาตัวแบบการพยากรณ์ที่เหมาะสมที่สุดในแต่ละเทคนิคการทำเหมืองข้อมูลด้วยวิธีการตรวจสอบไขว้ โดยกำหนดให้ค่า k เท่ากับ 5 10 15 และ 20 ผู้วิจัยจะพิจารณาจากเกณฑ์ค่าต่ำที่สุดของค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) จากนั้นใช้เกณฑ์ค่าต่ำที่สุดของทั้งค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยและค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย ($RMSE$) เพื่อเปรียบเทียบประสิทธิภาพของเทคนิคการทำเหมืองข้อมูลในการพยากรณ์ทั้ง 6 เทคนิค และให้ได้ตัวแบบสำหรับการพยากรณ์ปริมาณน้ำในเขื่อนภูมิพลที่เหมาะสมที่สุด (Granata & Nunno, 2021) โดยมีการคำนวณค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยและค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยดังต่อไปนี้

ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย

$$MAE = \frac{\sum_{i=1}^m |f_i - y_i|}{m} \quad (9)$$

ค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย

$$RMSE = \sqrt{\frac{\sum_{i=1}^m (f_i - y_i)^2}{m}} \quad (10)$$

โดย y_i แทน ค่าข้อมูลจริง ณ เวลา i

f_i แทน ค่าพยากรณ์ ณ เวลา i

m แทน จำนวนของข้อมูล

3.6 ขั้นตอนการนำไปใช้งาน (Deployment)

ตัวแบบสำหรับการพยากรณ์ปริมาณน้ำในเขื่อนภูมิพลที่เหมาะสมที่สุดที่ได้มาด้วยเทคนิคเหมืองข้อมูลสามารถนำไปใช้ประโยชน์ในการประกอบการตัดสินใจวางแผนและการกำหนดนโยบายในการบริหารจัดการน้ำในการผลิตกระแสไฟฟ้าเพื่อรองรับความต้องการของประชาชนและการขยายของทางเศรษฐกิจของประเทศ อีกทั้งยังช่วยเป็นข้อมูลสนับสนุนในการพัฒนาระบบการบริหารจัดการสำรองน้ำในด้านการเกษตร การผลิตน้ำประปาเพื่อใช้อุปโภคและบริโภค และการอุตสาหกรรมในช่วงฤดูแล้ง รวมถึงกำหนดชนิดของพืชและการจัดหาพันธุ์พืชเชิงเศรษฐกิจให้สอดคล้องเหมาะสมกับปริมาณน้ำในเขื่อน อีกทั้งยังสามารถเป็นข้อมูลช่วยในการรักษาระบบนิเวศ ธรรมชาติและสิ่งแวดล้อมในเขตพื้นที่รอบ ๆ เขื่อนได้ต่อไปในอนาคต อีกทั้งยังสามารถใช้เป็นข้อมูลช่วยในการติดตามและเฝ้าระวังสถานการณ์การเกิดอุทกภัย เพื่อที่จะสามารถรวบรวมแผนเตรียมความพร้อมต่าง ๆ ทั้งในด้านความพร้อมของเครื่องมือเครื่องจักรและเจ้าหน้าที่ของหน่วยงานที่เกี่ยวข้องในการบรรเทาสาธารณภัยในกรณีที่มีแนวโน้มของการเกิดอุทกภัยในอนาคต

4. ผลลัพธ์การวิจัย

ผลการเปรียบเทียบประสิทธิภาพในการแบ่งข้อมูลในแต่ละเทคนิคการทำเหมืองข้อมูลที่น่าสนใจในการพยากรณ์ทั้งหมด 6 เทคนิค ได้แก่ เทคนิคการถดถอยเชิงเส้นพหุคูณ เทคนิคแรนดอมฟอเรสต์ เทคนิคโครงข่าย

ประสาทเทียม เทคนิคจำลองต้นไม้เอนไซม์ไฟว์พี เทคนิคเพื่อนบ้านใกล้ที่สุด และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน ด้วยวิธีการตรวจสอบไขว้ โดยกำหนดให้ค่า k เท่ากับ 5 10 15 และ 20 โดยใช้เกณฑ์พิจารณาจากค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) ได้ผลปรากฏในตาราง 2

ตาราง 2 แสดงการเปรียบเทียบประสิทธิภาพของเทคนิคที่ใช้ในการพยากรณ์

เทคนิคที่ใช้ในการพยากรณ์	k	MAE
เทคนิคการถดถอยเชิงเส้นพหุคูณ	5*	1,038.825 **
	10	1,040.295
	15	1,039.828
	20	1,040.167
เทคนิคแรนดอมฟอเรสต์	5	352.233
	10	347.639
	15	344.524
	20*	341.945**
เทคนิคโครงข่ายประสาทเทียม	5*	992.815**
	10	1,180.250
	15	1,036.630
	20	1,043.640
เทคนิคจำลองต้นไม้เอนไซม์ไฟว์พี	5	470.037
	10*	461.519**
	15	461.804
	20	470.293
เทคนิคเพื่อนบ้านใกล้ที่สุด***	5	1,327.061
	10*	1,323.321**
	15	1,324.153
	20	1,324.624
เทคนิคซัพพอร์ตเวกเตอร์แมชชีน	5*	1,002.811**
	10	1,004.292
	15	1,003.350
	20	1,003.619

หมายเหตุ: * คือ ค่า k ที่เหมาะสมที่สุดในการแบ่งข้อมูลด้วยวิธีการตรวจสอบไขว้

** คือ ค่าต่ำที่สุดของค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE)

*** สำหรับเทคนิคเพื่อนบ้านใกล้ที่สุดใช้จำนวนข้อมูลมากสุดในข้อมูลเพื่อนบ้าน K เท่ากับ 18 ซึ่งเหมาะสมที่สุดในการสร้างตัวแบบการพยากรณ์ของเทคนิคเพื่อนบ้านใกล้ที่สุด

จากตาราง 2 แสดงการเปรียบเทียบประสิทธิภาพในการแบ่งข้อมูลในแต่ละเทคนิคการทำเหมืองข้อมูล ที่นำมาใช้ในการพยากรณ์ด้วยวิธีการตรวจสอบไขว้ โดยกำหนดให้ค่า k เท่ากับ 5 10 15 และ 20 โดยใช้เกณฑ์ พิจารณาจากค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย พบว่า วิธีการตรวจสอบไขว้ เมื่อกำหนดให้ค่า k เท่ากับ 5 มี ประสิทธิภาพสูงที่สุดสำหรับเทคนิคการถดถอยเชิงเส้นพหุคูณ เทคนิคโครงข่ายประสาทเทียม และเทคนิคซัพพอร์ต เวกเตอร์แมชชีน โดยมีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยต่ำที่สุด คือ 1,038.825 992.815 และ 1,002.811 ตามลำดับ ในส่วนของวิธีการตรวจสอบไขว้ เมื่อกำหนดให้ค่า k เท่ากับ 10 มีประสิทธิภาพสูงที่สุดสำหรับเทคนิคจำลองต้นไม้เอมไพร์ฟิ และเทคนิคเพื่อนบ้านใกล้ที่สุด โดยมีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยต่ำที่สุด คือ 461.519 และ 1,323.321 ตามลำดับ และวิธีการตรวจสอบไขว้ เมื่อกำหนดให้ค่า k เท่ากับ 20 มีประสิทธิภาพสูงที่สุดสำหรับเทคนิคแรนดอมฟอเรสต์

ผลการเปรียบเทียบประสิทธิภาพเทคนิคการทำเหมืองข้อมูลนำมาใช้ในการพยากรณ์ปริมาณน้ำในเขื่อน ภูมิพล โดยใช้เกณฑ์พิจารณาจากค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยและค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยที่ได้มาจากตัวแบบการพยากรณ์ปริมาณน้ำทั้ง 6 เทคนิค ได้แก่ เทคนิคการถดถอยเชิงเส้นพหุคูณ เทคนิคแรนดอมฟอเรสต์ เทคนิคโครงข่ายประสาทเทียม เทคนิคจำลองต้นไม้เอมไพร์ฟิ เทคนิคเพื่อนบ้านใกล้ที่สุด และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน ได้ผลปรากฏผลดังตาราง 3

ตาราง 3 แสดงการเปรียบเทียบค่าทดสอบประสิทธิภาพของการพยากรณ์ปริมาณน้ำในเขื่อนภูมิพล

เทคนิคที่ใช้ในการพยากรณ์	MAE	RMSE
เทคนิคแรนดอมฟอเรสต์*	341.945**	593.790**
เทคนิคจำลองต้นไม้เอมไพร์ฟิ	461.519	754.416
เทคนิคโครงข่ายประสาทเทียม	992.815	1,294.753
เทคนิคการถดถอยเชิงเส้นพหุคูณ	1,038.825	1366.960
เทคนิคซัพพอร์ตเวกเตอร์แมชชีน	1,002.811	1,408.007
เทคนิคเพื่อนบ้านใกล้ที่สุด	1,323.321	1,690.211

หมายเหตุ: * คือ เทคนิคที่เหมาะสมในการสร้างตัวแบบการพยากรณ์

** คือ ค่าทดสอบประสิทธิภาพของการพยากรณ์ที่มีค่าต่ำที่สุด

จากตาราง 3 แสดงการเปรียบเทียบค่าทดสอบการทดสอบประสิทธิภาพในการสร้างตัวแบบสำหรับการพยากรณ์ปริมาณน้ำในเขื่อนภูมิพล จังหวัดตาก ของทั้ง 6 เทคนิค พบว่า เทคนิคแรนดอมฟอเรสต์ เป็นเทคนิคที่มี ประสิทธิภาพมากที่สุด เพราะมีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยและค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำ ที่สุด เท่ากับ 341.945 และ 593.790 ตามลำดับ

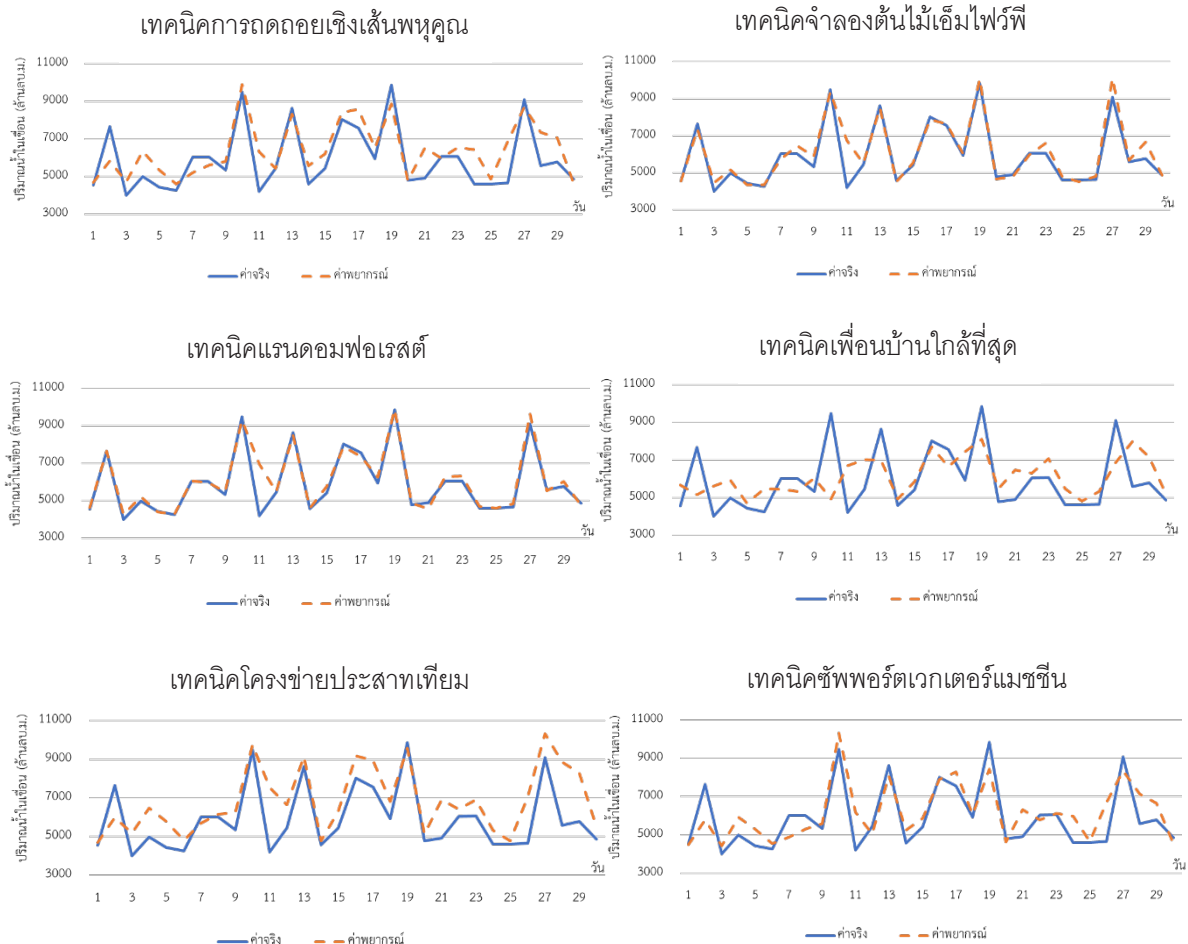
นอกจากนั้น เพื่อให้เห็นถึงความแตกต่างของค่าความคลาดเคลื่อนระหว่างค่าปริมาณน้ำในเขื่อนจริงกับ ปริมาณน้ำในเขื่อนพยากรณ์ที่ได้จากตัวแบบในแต่ละเทคนิคพยากรณ์ทั้ง 6 เทคนิค โดยนำเสนอเฉพาะข้อมูลของ ปริมาณน้ำในเขื่อนล่าสุดวันที่ 1 ถึง 30 มิถุนายน พ.ศ. 2564 จำนวน 30 วัน ได้ผลดังรูปภาพประกอบ 2

โดยจากรูปภาพประกอบ 2 จะเห็นได้ว่า เมื่อเปรียบเทียบปริมาณน้ำในเขื่อนทั้งค่าจริง (Actual) และค่า พยากรณ์ (Predicted) ของเทคนิคการทำเหมืองข้อมูลทั้ง 6 เทคนิค ของชุดข้อมูลทดสอบ (Testing Data Set) จำนวน 30 วันดังกล่าว โดยจะเห็นได้ค่อนข้างชัดเจนว่าตัวแบบพยากรณ์ด้วยเทคนิคแรนดอมฟอเรสต์มีลักษณะของแนวโน้ม ของค่าพยากรณ์ปริมาณน้ำในเขื่อนเข้าใกล้ค่าปริมาณน้ำในเขื่อนจริงมากที่สุด ซึ่งผลดังกล่าวสอดคล้องกับค่าความ

คลาดเคลื่อนสัมบูรณ์เฉลี่ยและค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยที่มีค่าต่ำที่สุดในตาราง 3 แสดงว่า เทคนิค แรนดอมฟอเรสต์เป็นเทคนิคที่เหมาะสมที่สุดในการสร้างตัวแบบสำหรับการพยากรณ์ปริมาณน้ำในเขื่อนภูมิพล จังหวัดตาก

รูปภาพประกอบ 2

เปรียบเทียบปริมาณน้ำในเขื่อนทั้งค่าจริงและค่าพยากรณ์ของเทคนิคการทำเหมืองข้อมูลทั้ง 6 เทคนิค



5. การอภิปรายผล

ในการวิจัยครั้งนี้ได้ พบว่า เทคนิคแรนดอมฟอเรสต์โดยใช้การแบ่งข้อมูลด้วยวิธีการตรวจสอบไขว้เมื่อกำหนดให้ค่า k เท่ากับ 20 เป็นเทคนิคที่มีประสิทธิภาพมากที่สุดสำหรับการพยากรณ์ปริมาณน้ำในเขื่อนภูมิพล จังหวัดตาก โดยมีแอตทริบิวต์ที่ใช้ในการสร้างตัวแบบ ได้แก่ ค่าเฉลี่ยน้ำในเขื่อนรวมทั้งปี ปริมาณน้ำไหลลงอ่าง ปริมาณน้ำระบาย ความกดอากาศเฉลี่ย อุณหภูมิเฉลี่ยและความชื้นเฉลี่ย ซึ่งผลสอดคล้องกับผลการศึกษากการพยากรณ์ปริมาณน้ำของ ทะเลสาบบราซโซ (Lake Bracciano) ในประเทศอิตาลี ของ Guyennon et al. (2021) โดยมีแอตทริบิวต์ที่ใช้ในการสร้างตัวแบบเหมือนกัน คือ ความกดอากาศ อุณหภูมิ และความชื้น และสอดคล้องกับผลงานวิจัยของ Zang et al. (2021) ที่ศึกษาการพยากรณ์ปริมาณน้ำของทะเลสาบไท่ (Lake Tai) ในประเทศจีน โดยมีแอตทริบิวต์ที่ใช้ในการสร้างตัวแบบเหมือนกัน คือ อุณหภูมิ และเมื่อพิจารณาแต่ไม่สอดคล้องกับงานวิจัยของ จิรโรจน์ ตอสะสุกุล และสุพิชชา

ชัดธิพงษ์ (2564) ได้ศึกษาการพยากรณ์ปริมาณน้ำในเขื่อนภูมิพล จังหวัดตาก ที่ผลการศึกษาพบว่า ตัวแบบการพยากรณ์ของเทคนิคแบบจำลองต้นไม้เอมไพร์เป็นเทคนิคที่เหมาะสมที่สุด เนื่องจากในงานวิจัยครั้งนี้ ผู้วิจัยได้เพิ่มเทคนิคการทำเหมืองข้อมูลอีก 2 เทคนิค ได้แก่ เทคนิคการถดถอยเชิงเส้นพหุคูณและเทคนิคแรนดอมฟอเรสต์ รวมทั้งเพิ่มจำนวนข้อมูลที่ใช้ในการศึกษามากขึ้นและเพิ่มเติมในส่วนของการศึกษาเปรียบเทียบการแบ่งข้อมูลด้วยวิธีการตรวจสอบไขว้จึงส่งผลทำให้ได้ตัวแบบสำหรับการพยากรณ์ที่มีประสิทธิภาพมากกว่างานวิจัยของ จิรโรจน์ ตอสะสุกุล และ สุพิชชา ชัดธิพงษ์ (2564) ที่ใช้วิธีการตรวจสอบไขว้โดยกำหนดให้ค่า k เท่ากับ 10 เพียงอย่างเดียว

6. สรุปผลการวิจัย

การวิจัยครั้งนี้ศึกษาเกี่ยวกับการสร้างและเปรียบเทียบประสิทธิภาพของตัวแบบสำหรับการพยากรณ์ปริมาณน้ำในเขื่อนภูมิพล จังหวัดตาก โดยใช้แอตทริบิวต์ในการศึกษาทั้งหมดจำนวน 12 แอตทริบิวต์ จากข้อมูลรายวันของรายงานสถานการณ์น้ำในเขื่อนของคลังข้อมูลน้ำแห่งชาติและข้อมูลทางอุตุนิยมหาวิทยาลัยเกษตรศาสตร์ภาคเหนือ ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2554 ถึง 31 พฤษภาคม พ.ศ. 2564 และใช้เทคนิคการทำเหมืองข้อมูล 6 เทคนิค ได้แก่ เทคนิคการถดถอยเชิงเส้นพหุคูณ เทคนิคแรนดอมฟอเรสต์ เทคนิคโครงข่ายประสาทเทียม เทคนิคจำลองต้นไม้เอมไพร์ เทคนิคเพื่อนบ้านใกล้ที่สุด และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน และในแต่ละเทคนิคได้ใช้วิธีแบ่งข้อมูลเพื่อเปรียบเทียบประสิทธิภาพของตัวแบบการพยากรณ์ด้วยวิธีการตรวจสอบไขว้ (k-Fold Cross Validation) ที่กำหนดให้ค่า k เท่ากับ 5 10 15 และ 20

ในการวิจัยครั้งนี้ได้ พบว่า ในการแบ่งข้อมูลในแต่ละเทคนิคการทำเหมืองข้อมูลที่นำมาใช้ในการพยากรณ์ด้วยวิธีการตรวจสอบไขว้ โดยกำหนดให้ค่า k เท่ากับ 5 มีประสิทธิภาพสูงที่สุดสำหรับเทคนิคการถดถอยเชิงเส้นพหุคูณ เทคนิคโครงข่ายประสาทเทียมและเทคนิคซัพพอร์ตเวกเตอร์แมชชีน และที่ค่า k เท่ากับ 10 มีประสิทธิภาพสูงที่สุดสำหรับเทคนิคจำลองต้นไม้เอมไพร์และเทคนิคเพื่อนบ้านใกล้ที่สุด และที่ค่า k เท่ากับ 20 มีประสิทธิภาพสูงที่สุดสำหรับเทคนิคแรนดอมฟอเรสต์ และจากการเปรียบเทียบประสิทธิภาพของเทคนิคการทำเหมืองข้อมูลทั้ง 6 เทคนิคในการสร้างตัวแบบสำหรับการพยากรณ์ปริมาณน้ำในเขื่อนภูมิพล จังหวัดตาก พบว่า เทคนิคแรนดอมฟอเรสต์เป็นเทคนิคที่มีประสิทธิภาพมากที่สุด โดยให้ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยและค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด เท่ากับ 341.945 และ 593.790 ตามลำดับ และแอตทริบิวต์ที่ใช้ในการสร้างตัวแบบของเทคนิคแรนดอมฟอเรสต์นี้มี 6 แอตทริบิวต์ คือ ค่าเฉลี่ยน้ำในเขื่อนรวมทั้งปี ปริมาณน้ำไหลลงอ่าง ปริมาณน้ำระบาย ความกดอากาศเฉลี่ย อุณหภูมิเฉลี่ยและความชื้นเฉลี่ย โดยถูกเลือกมาจากแอตทริบิวต์ที่มีความสัมพันธ์สูงกับตัวแปรกลุ่มเป้าหมายด้วยวิธี Classifier Subset Evaluator ที่ใช้เกณฑ์ค่าน้ำหนักของแต่ละปัจจัยที่มีค่าตั้งแต่ 50 เปอร์เซนต์ขึ้นไป อีกทั้งตัวแบบการพยากรณ์ของเทคนิคแรนดอมฟอเรสต์ถูกสร้างขึ้นจากแอตทริบิวต์จำนวนน้อยกว่าเมื่อเปรียบเทียบกับตัวแบบการพยากรณ์ของเทคนิคจำลองต้นไม้เอมไพร์ที่ถูกสร้างมาจากแอตทริบิวต์จำนวน 7 แอตทริบิวต์ โดยมีปริมาณฝนเพิ่มเข้ามาจาก 6 แอตทริบิวต์ของตัวแบบในเทคนิคแรนดอมฟอเรสต์ ซึ่งจะเห็นได้ว่า ถึงแม้ว่าตัวแบบการพยากรณ์ของเทคนิคแรนดอมฟอเรสต์ถูกสร้างขึ้นมาขึ้นจากแอตทริบิวต์จำนวนน้อยกว่าแต่มีประสิทธิภาพมากกว่า ซึ่งตัวแบบการพยากรณ์ที่ได้ดังกล่าวสามารถนำไปใช้ในการประกอบการตัดสินใจและกำหนดแนวทางการพัฒนาระบบการบริหารจัดการน้ำในแง่ของการจัดการปริมาณน้ำไหลลงอ่างและปริมาณน้ำระบาย รวมทั้งใช้ข้อมูลเกี่ยวกับค่าเฉลี่ยน้ำในเขื่อนรวมทั้งปี ความกดอากาศเฉลี่ย อุณหภูมิเฉลี่ยและความชื้นเฉลี่ย เพื่อรองรับการผลิตกระแสไฟฟ้า การอุปโภค และบริโภคอย่างเพียงพอและเหมาะสม ช่วยในการวางแผนเพิ่มผลผลิตทางการเกษตรและกำหนดชนิดของพืชให้สอดคล้องกับปริมาณน้ำในเขื่อนในช่วงฤดูแล้ง อีกทั้งยังสามารถนำไปใช้เป็นข้อมูลช่วยสนับสนุนการติดตามและเฝ้าระวังการเกิด

อุทกภัยแนวโน้มของเกิดอุทกภัยในช่วงฤดูฝน อันเป็นประโยชน์ต่อการวางแผนเตรียมความพร้อมทั้งในด้านเครื่องมือ เครื่องจักรและเจ้าหน้าที่ของหน่วยงานที่เกี่ยวข้อง

7. ข้อเสนอแนะ

ในการวิจัยครั้งต่อไปควรศึกษาเพิ่มเติมเกี่ยวกับปัจจัยที่อาจจะมีผลต่อการพยากรณ์ปริมาณน้ำในเขื่อน เช่น ปัจจัยในส่วนของอุณหภูมิสูงสุดและต่ำสุด ปริมาณฝนสะสมในเขตพื้นที่เหนือเขื่อน และปริมาณแสงแดด เป็นต้น รวมทั้งควรเพิ่มเติมในส่วนของการกำหนดค่าพารามิเตอร์ของแต่ละเทคนิคการทำเหมืองข้อมูลให้ครอบคลุมและเหมาะสมที่สุด

เอกสารอ้างอิง

- การไฟฟ้าฝ่ายผลิตแห่งประเทศไทย. (2561). *เขื่อนภูมิพล*. ค้นเมื่อ 26 กันยายน 2564, จาก <https://rb.gy/lgyg8>.
- จิรโรจน์ ตอสะสุกุล และสุพิชชา ชัดธิพงษ์. (2564). การประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลสำหรับพยากรณ์ ปริมาณน้ำในเขื่อนภูมิพล จังหวัดตาก. ใน *การประชุมสัมมนาวิชาการเพื่อนำเสนอผลงานวิจัยระดับชาติ* (หน้า 645-646). มหาสารคาม : คณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม.
- ทรงศิริ แด่สมบัติ. (2549). *การพยากรณ์เชิงปริมาณ*. กรุงเทพฯ : สำนักพิมพ์มหาวิทยาลัยเกษตรศาสตร์.
- นิพนธ์พันธุ์ ศศิธร. (2561). *มูลค่าทางเศรษฐกิจของความเสื่อมโทรมของทรัพยากรน้ำเข้าสู่เขื่อนภูมิพลและหลักการ และเหตุผลสำหรับการคุ้มครองธรรมชาติในพื้นที่ภาคเหนือของประเทศไทย*. วิทยานิพนธ์มหาบัณฑิต สถาบันบัณฑิตพัฒนบริหารศาสตร์.
- รุ่งนภา สุทธิชัย จรวัย สาวิณี และมณีนรัตน์ วงษ์ขี้ม. (2561). ความสัมพันธ์ระหว่างความชาญฉลาดทางธุรกิจกับ ความสำเร็จของธุรกิจพาณิชย์อิเล็กทรอนิกส์ในประเทศไทย. *วารสารการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม*, 10(1), 197-206.
- วีรศักดิ์ ฟองเงิน วรภา อาธิราษฎร์ และเผด็จ พรหมสาขา ณ สกลนคร. (2017). การพยากรณ์ปริมาณน้ำในเขื่อนกิ่วลม โดยใช้เทคนิคเหมืองข้อมูล. *วารสารวิทยาการจัดการสมัยใหม่ คณะวิทยาการจัดการมหาวิทยาลัยราชภัฏ ลำปาง*, 10(2), 121-131.
- สำนักข่าวออนไลน์ไทยพับลิก้า. (2564, 8 มิถุนายน). ฝนทิ้งช่วงเขื่อนใหญ่ภาคเหนือ-ตะวันตก เหลือน้ำใช้การได้ 6%. [ฉบับอิเล็กทรอนิกส์]. *ไทยพับลิก้า*. ค้นเมื่อ 16 สิงหาคม 2564, จาก <https://bit.ly/3lhYuuq>.
- สำนักบริหารจัดการน้ำและอุทกวิทยา. (2563). *แผนการป้องกันและบรรเทาภัยอันเกิดจากน้ำ (ฤดูฝน) พ.ศ.2563*. กรุงเทพฯ : กรมชลประทาน.
- สุรพงศ์ เชื้อวัฒนามงคล. (2562). *การทำเหมืองข้อมูล*. กรุงเทพฯ : สำนักพิมพ์สถาบันบัณฑิตพัฒนบริหารศาสตร์.
- อนุชิต เนาส่ง ชลธิชา วงษ์คะสุ่ม และคณะ. (2563). การพยากรณ์ปริมาณน้ำในเขื่อนอุบลรัตน์ จังหวัดขอนแก่น โดยใช้เทคนิคการทำเหมืองข้อมูล. ใน *การประชุมวิชาการระดับชาติวิทยาศาสตร์ เทคโนโลยีและนวัตกรรม (มหาวิทยาลัยแม่โจ้) ครั้งที่ 1* (หน้า 595-608). เชียงใหม่ : คณะวิทยาศาสตร์ มหาวิทยาลัยแม่โจ้.
- อัมเรศ รินทราณี รัตนาวดี สนธิประสาธ และณัฐภูมิ ตันติเศรษฐ. (2563). ผลกระทบของการทำเหมืองข้อมูล ทางการบัญชีที่มีต่อประสิทธิภาพการตัดสินใจขององค์กรของธุรกิจชั้นส่วนยานยนต์ในประเทศไทย. *วารสารการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม*, 11(4), 31-40.

- Ahmadi, M., & Khashei, M. (2021). Generalized support vector machines (GSVMs) model for Real-world time series forecasting. *Research Square*, 1, 1-22.
- Behnood, A., Behnood, V., Gharehveran, M. M., & Alyamac, K. E. (2017). Prediction of the compressive strength of normal and high-performance concretes using M5P model tree algorithm. *Construction and Building Materials*, 142, 199-207.
- Dutta, S., & Bandyopadhyay, S. K. (2020). Early lung cancer prediction using neural network with Cross-validation. *Asian Journal of Research in Infectious Diseases*, 4(4), 15-22.
- Granata, F., & Di Nunno, F. (2021). Artificial intelligence models for prediction of the tide level in Venice. *Stochastic Environmental Research and Risk Assessment*, 35(12), 1-12.
- Granata, F., Saroli, M., de Marinis, G., & Gargano, R. (2018). Machine learning models for spring discharge forecasting. *Geofluids*, 2018, 1-15.
- Guyennon, N., Salerno, F., Rossi, D., Rainaldi, M., Calizza, E., & Romano, E. (2021). Climate change and water abstraction impacts on the long-term variability of water levels in Lake Bracciano (Central Italy): A Random Forest approach. *Journal of Hydrology: Regional Studies*, 37, 1-13.
- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques* (3rd ed.). Amsterdam : Elsevier.
- Kantardzic, M. (2011). *Data mining: concepts, models, methods, and algorithms* (2nd ed.). Hoboken : John Wiley & Sons.
- Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (2008). *Forecasting methods and applications*. John Wiley & sons.
- Pires, I. M., Marques, G., Garcia, N. M., & Ponciano, V. (2020). Machine learning for the evaluation of the presence of heart disease. *Procedia Computer Science*, 177, 432-437.
- Quinlan, J. R. (1992). *Learning with continuous classes*. Paper presented at the 5th Australian joint conference on artificial intelligence, Tasmania, Australia.
- Tan, P. N., Steinbach, M., & Kumar, V. (2016). *Introduction to data mining*. New Delhi : Pearson Education.
- Vijai, P., & Sivakumar, P. B. (2018). Performance comparison of techniques for water demand forecasting. *Procedia computer science*, 143, 258-266.
- Wang, Q., & Wang, S. (2020). Machine learning-based water level prediction in Lake Erie. *Water*, 12(10), 2654.
- Witten, I. H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques* (2nd ed.). San Francisco : Morgan Kaufmann Publisher.
- Zhang, W., Wu, C., Zhong, H., Li, Y., & Wang, L. (2021). Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. *Geoscience Frontiers*, 12(1), 469-477.