การศึกษาเปรียบเทียบการพยากรณ์ระดับน้ำด้วยอัลกอริทึม ANN และ XGBoost กรณีศึกษาแม่น้ำป่าสัก

Comparative study of ANN and XGBoost in Water Level Prediction: A Case Study of the Pasak River in Thailand

สุรศักดิ์ พบวันดี [1] และ ณัฏฐ์ โอธนาทรัพย์[2]

[1]นักศึกษาหลักสูตรวิทยาศาสตรมหาบัณฑิต การจัดการนวัตกรรมเทคโนโลยีดิจิทัล บัณฑิตวิทยาลัย มหาวิทยาลัยเอเชียอาคเนย์ popwandee@gmail.com

[2]อาจารย์ประจำหลักสูตรวิทยาศาสตรมหาบัณฑิต การจัดการนวัตกรรมเทคโนโลยีดิจิทัล บัณฑิตวิทยาลัย มหาวิทยาลัยเอเชียอาคเนย์ nuto@sau.ac.th

บทคัดย่อ

ประเทศไทยประสบปัญหาภัยธรรมชาติขาดแคลนน้ำในฤดูแล้ง ในขณะที่น้ำท่วมในฤดูมรสุม การคาดการณ์ระดับน้ำล่วงหน้าได้อย่างแม่นยำจึงเป็นความท้าทายในกระบวนการตัดสินใจ เพื่อควบคุมระดับน้ำในแม่น้ำสายต่าง ๆ ให้อยู่ในระดับที่เหมาะสม รวมถึงการแจ้งเตือนภัยล่วงหน้า เพื่อลดผลกระทบที่จะเกิดจากภัยน้ำแล้งและน้ำท่วมได้อย่างมีประสิทธิภาพ อันจะนำไปสู่การบริหารจัดการทรัพยากรน้ำอย่างยั่งยืน

งานวิจัยนี้ได้มุ่งศึกษาเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) ด้วยการใช้อัลกอริทึมที่ได้รับความนิยม ได้แก่ Artificial Neuron Network (ANN) และ Extreme Gradient Boost (XGBoost) โดยใช้ข้อมูลสภาพอากาศจากสถานีวัดค่าบริเวณลุ่มแม่น้ำป่าสักตอนบน เพื่อพยากรณ์ระดับน้ำริมตลิ่งบริเวณสถานีวัดระดับน้ำ PAS004 ซึ่งในห้วงเวลาทดลองมีระดับน้ำเฉลี่ย 61.89 ม. สูงสุด 68.23 ม. และต่ำสุด 58.96 ม. การเปรียบเทียบประสิทธิภาพใช้ค่าความคลาดเคลื่อนเฉลี่ยสัมบูรณ์ Mean Absolute Error (MAE) ผลการทดลองพบว่าอัลกอริทึม ANN และอัลกอริทึม XGBoost สามารถพยากรณ์ได้ใกล้เคียงกัน ผลการพยากรณ์ระดับน้ำ ณ เวลาที่พยากรณ์ (เวลา t) แบบจำลองที่สร้างด้วยอัลกอริทึม ANN มีค่าความคลาดเคลื่อนเฉลี่ยสัมบูรณ์ MAE เท่ากับ 2.42 ซม. อัลกอริทึม XGBoost มีค่า MAE เท่ากับ 3.14 ซม. โดยที่อัลกอริทึม ANN ใช้หน่วยความจำของเครื่องคอมพิวเตอร์และใช้เวลาในกระบวนการเรียนรู้มากกว่าอัลกอริทึม XGBoost

คำสำคัญ: การเรียนรู้ของเครื่อง การพยากรณ์ระดับน้ำ แม่น้ำป่าสัก

Abstract

Flooding and drought are some of the natural disasters in Thailand. It is a challenge to manage water resources to control the water level. However, many dams have been built to sufficiently hold Thailand's water resource reservation in the dry season. Oversupply water retention lead to flooding in the rainy season. Accurate forecasting of water level is, therefore, one of the important factors to make a decision-making process in controlling water level. This research aims to study machine learning techniques for water level prediction in the Pasak River basin. Extreme Gradient Boosting, a useful machine learning technique, has been implemented. Water level and precipitation of stations PAS001, PAS002, PAS003, and PAS004 are variables, and water level in PAS004 is the target prediction. The study found that ANN and XGBoost algorithm are accurate in predicting water levels. The evaluation means absolute error (MAE) by ANN is 2.42 cm. and XGBoost is 3.14 cm. The ANN algorithm uses computer's memory much more and takes time for learning process longer than XGBoost.

Keywords: Machine learning, Water level prediction, Artificial Neuron Network, XGBoost

1. ข้อมูลทั่วไป

ประเทศไทยประสบปัญหาภัยธรรมชาติ ทั้งน้ำแล้งและน้ำท่วมในหลายพื้นที่ ส่วนหนึ่งมาจากสภาพภูมิประเทศ และการเปลี่ยนแปลงสภาพอากาศ ความท้าทายในการบริหารจัดการทรัพยากรน้ำตามแม่น้ำสายหลักคือการควบคุมปริมาณน้ำให้มีใช้เพียงพอในฤดูแล้งและไม่มากเกินไปในฤดูฝน โดยเฉพาะอย่างยิ่งในพื้นที่แม่น้ำป่าสัก ซึ่งมีลักษณะทอดตัวยาวจากต้นน้ำบริเวณจังหวัดเลย ไหลลงทางทิศใต้ผ่านจังหวัดเพชรบูรณ์ ลพบุรี สระบุรี และ พระนครศรีอยุธยา ในฤดูฝนหรือฤดูน้ำหลากจะมีน้ำท่วมริมตลิ่งสร้างผลกระทบต่อประชาชนในพื้นที่ลุ่มต่ำในหลายพื้นที่อย่างไรก็ตามน้ำจะไหลผ่านพื้นที่และทำให้ประสบปัญหาน้ำแล้งตามมา

การสร้างเขื่อนป่าสักชลสิทธิ์ช่วยให้สามารถกักเก็บน้ำสำรองไว้ใช้ในฤดูแล้งและลดผลกระทบจากน้ำท่วมในพื้นที่ใต้เขื่อนในฤดูมรสุม อย่างไรก็ตามการบริหารจัดการทรัพยากรน้ำให้สามารถมีน้ำใช้เพียงพอและไม่มากเกินไป จำเป็นต้องมีการคาดการณ์ระดับน้ำล่วงหน้าที่แม่นยำ การศึกษานี้ได้ศึกษาการพยากรณ์ระดับน้ำล่วงหน้าตามแนวแม่น้ำป่าสักตอนบน เพื่อเป็นทางเลือกในกระบวนการตัดสินใจสำหรับการบริหารจัดการทรัพยากรน้ำได้อย่างมีประสิทธิภาพ โดยใช้เทคนิคการเรียนรู้ของเครื่อง หรือ Machine Learning

รูปที่ 1 สภาพภูมิประเทศพื้นที่ศึกษาตามแนวแม่น้ำป่าสักตอนบน

available online May 1, 2023.
*Corresponding author: สุรศักดิ์ พบวันดี มหาวิทยาลัยเอเชียอาคเนย์ (E-mail: popwandee@gmail.com)

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

จากการศึกษาค้นคว้างานวิจัยที่เกี่ยวข้องพบว่ามีการวิจัยเพื่อพยากรณ์น้ำท่วมในลุ่มน้ำมูลตอนล่าง ด้วยเทคนิคโครงข่ายประสาทเทียมในกระบวนการเรียนรู้ด้วยเครื่อง โดยใช้ข้อมูลระดับน้ำรายชั่วโมงจากสถานีต่างๆ เป็นตัวแปรนำเข้า ผลการพยากรณ์ระดับน้ำล่วงหน้า 24 และ 48 ชั่วโมง มีค่าความคลาดเคลื่อนประมาณ 8.7 และ 12.7 เซ็นติเมตร[1]

สำหรับในลุ่มแม่น้ำเจ้าพระยามีผู้ศึกษาเพื่อพัฒนาและประยุกต์ใช้แบบจำลองโครงข่ายประสาทเทียมเช่นเดียวกัน โดยใช้ข้อมูลระดับน้ำและปริมาณน้ำฝนบริเวณเหนือชุมชนจากสถานีวัดระดับน้ำลุ่มน้ำเจ้าพระยาตอนล่าง และข้อมูลระดับน้ำจากเครื่องวัดระดับน้ำในคลองทับน้ำและเครื่องวัดปริมาณน้ำฝนที่ติดตั้งในพื้นที่ชุมชนเป็นข้อมูลนำเข้า ซึ่งจากผลการทดลองพบว่าแบบจำลองโครงข่ายประสาทเทียมสามารถพยากรณ์ระดับน้ำรายวันมีประสิทธิภาพ การพยากรณ์รายสัปดาห์และรายเดือนมีความแม่นยำน้อยลงตามระยะเวลาที่ยาวนานขึ้น[2]

นอกจากนี้การเรียนรู้ด้วยเครื่องยังถูกนำไปใช้สำหรับการพยากรณ์ระดับน้ำในต่างประเทศ อาทิ ประเทศเกาหลี Changhyun Choi และ Jungwook Kim ได้ทำการศึกษาเพื่อพยากรณ์ระดับน้ำในเขตชุ่มน้ำ Upo wetland โดยแบ่งชุดข้อมูลตามห้วงระยะเวลา กล่าวคือห้วงการฝึกสอนการเรียนรู้ด้วยเครื่องใช้ข้อมูลห้วงวันที่ 1 เม.ย. 2009 ถึง 31 มี.ค. 2013 และข้อมูลที่ใช้ในการทดสอบอยู่ในห้วงวันที่ 1 เม.ย. 2013 ถึงวันที่ 31 มี.ค. 2015[3]

รวมทั้งยังมีการศึกษาเปรียบเทียบการเรียนรู้ของเครื่องในงานลักษณะเดียวกัน ด้วยเทคนิค Artificial Neuron Network, Support Vector Regression และ Extreme Gradient Boost เพื่อพยากรณ์ระดับน้ำในอ่างเก็บน้ำในเมืองเซลังงอ ประเทศมาเลเซีย ด้วยข้อมูลปริมาณน้ำฝน อุณหภูมิ และการระเหยของน้ำ ผลการทดลองพบว่า XGBoost และการเพิ่มข้อมูลระดับน้ำย้อนหลัง 1 วันให้ผลการพยากรณ์ดีที่สุด ประเมินผลด้วย R-Square เท่ากับ 0.92[4]

2.1 การเรียนรู้ของเครื่อง

การเรียนรู้ของเครื่อง หมายถึงการพัฒนาอัลกอริทึมสำหรับคอมพิวเตอร์ในการประมวลผลข้อมูลให้นำไปสู่การตอบสนองที่มีความฉลาด ภายใต้สภาวะแวดล้อมที่มีพัฒนาการของข้อมูล กระบวนการทางสถิติ ด้วยพลังการคำนวณอย่างรวดเร็วไปด้วยกัน[5]

ในเชิงแนวความคิดในการเรียนรู้เกี่ยวข้องกับนามธรรมของการนำเสนอข้อมูลในเชิงโครงสร้าง และแยกแยะลักษณะทั่วไปของโครงสร้างนี้ให้สามารถประเมินค่าได้สำหรับการนำไปใช้ประโยชน์ส่วนในเชิงปฏิบัตินั้น การเรียนรู้ของเครื่องใช้ข้อมูลที่ประกอบด้วยตัวอย่างและคุณลักษณะของแนวความคิดที่จะเรียนรู้ และสรุปรวมข้อมูลในรูปแบบของแบบจำลอง ที่จะถูกนำไปใช้ในการพยากรณ์หรืออธิบายตามวัตถุประสงค์ โดยวัตถุประสงค์นี้สามารถจัดแบ่งเป็นกลุ่มตามลักษณะงานประกอบด้วยการจัดประเภท (Classification) การคาดการณ์เป็นตัวเลข (numeric prediction) การตรวจจับรูปแบบ (Pattern detection) และการจัดกลุ่ม (Clustering) การเลือกใช้จะขึ้นอยู่กับพื้นฐานของข้อมูลนำเข้าและลักษณะงาน[5]

2.2 โครงข่ายประสาทเทียม (Artificial Neuron Network: ANN)

เป็นวิธีการที่มีพื้นฐานมาจากการเลียนแบบการทำงานของสมองมนุษย์ ประกอบด้วยชั้นข้อมูลนำเข้า (Input Layer) หรือข้อมูลตัวแปรที่นำมาคำนวณ แต่ละข้อมูลจะมีค่าน้ำหนัก (weight) ของแต่ละข้อมูล ในชั้นซ่อน (Hidden Layer) ซึ่งสามารถมีได้หลายชั้น แต่ละชั้นจะมีจำนวนเซลล์ประสาท (Neuron) ทำหน้าที่ในการคำนวณค่าตัวแปรนำเข้า ค่าน้ำหนักและค่าความเบี่ยงเบน (bias) รวมกันเพื่อคำนวณค่าผลลัพธ์ที่เหมาะสมของแต่ละเซลล์ และชั้นผลลัพธ์ (Output Layer) นำข้อมูลที่ได้จากการคำนวณไปใช้เป็นผลลัพธ์ จำนวนผลลัพธ์ที่ต้องการสามารถกำหนดได้ตามลักษณะงาน ในกรณีการพยากรณ์แบบถดถอย (Regression) จะกำหนดให้มีชั้นผลลัพธ์เป็นเพียงคำตอบเดียว[6] ถึงแม้ว่าปัจจุบันโครงข่ายประสาทเทียมมีการพัฒนาเทคนิคการประมวลผลการเรียนรู้ใหม่ ๆ เช่น Convolutional Neuron Network (CNN) และ Recurrent Neuron Network (RNN) งานวิจัยนี้ยังคงใช้เทคนิค ANN เนื่องจากเป็นวิธีพื้นฐานที่ไม่ซับซ้อนมากแต่ยังคงมีประสิทธิภาพในการประมวลผล

2.3 เอ็กซ์ตรีมเกรเดียนบูส (Extreme Gradient Boost XGBoost)

เป็นเทคนิคการเรียนรู้ด้วยเครื่องที่ได้รับความนิยมมากในปัจจุบัน พัฒนามาจาก อัลกอริทึม Gradient Boosted decision trees เพื่อให้มีประสิทธิภาพและความเร็วในการประมวลผลมากขึ้น เป็นเทคนิคที่รวมแบบจำลองหลาย ๆ อันเข้ามาแก้ไขค่าความคลาดเคลื่อนโดยใช้เทคนิคการเรียนรู้ข้อผิดพลาดของการตัดสินใจครั้งก่อน นำมาแก้ไขในการตัดสินใจครั้งต่อไป ในการเรียนรู้แต่ละครั้งจะสร้างตัวเรียนรู้ที่อ่อนแอ (weak learner) หลาย ๆ ตัว เพื่อเรียนรู้และลดค่าความผิดพลาดจากการสอนครั้งก่อนหน้าให้น้อยลงและทำการเรียนรู้ซ้ำจนกว่าจะไม่มีข้อผิดพลาดแล้ว [7]

2.4 การประเมินผลประสิทธิภาพของแบบจำลอง

การศึกษานี้ใช้การประเมินผลแบบจำลองด้วยค่าความคลาดเคลื่อนเฉลี่ยสัมบูรณ์ (Mean Absolute Error) หรือ MAE ดังสมการที่ 1 และ ค่าสัมประสิทธิ์การตัดสินใจ หรือ R-Square ดังสมการที่ 2

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \hat{Y}_i| \quad (1)$$

โดยที่ n แทนจำนวนครั้งในการพยากรณ์, i แทนครั้งที่ทำการพยากรณ์
Y_i แทนค่าระดับน้ำจริง ณ การพยากรณ์ครั้งที่ i
$\hat{Y}_i$ แทนค่าระดับน้ำจากการพยากรณ์ครั้งที่ i

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \quad (2)$$

โดยที่ n แทนจำนวนครั้งในการพยากรณ์, i แทนครั้งที่ทำการพยากรณ์
Y_i แทนค่าระดับน้ำจริง ณ การพยากรณ์ครั้งที่ i
$\hat{Y}_i$ แทนค่าระดับน้ำจากการพยากรณ์ครั้งที่ i
$\bar{Y}$ แทนค่าเฉลี่ยระดับน้ำจริง

3. การดำเนินการทดลอง

3.1 การรวบรวมข้อมูล

การศึกษาครั้งนี้ได้รวบรวมข้อมูลสถานีวัดค่าตัวแปรสภาพน้ำและอากาศจากสถาบันสารสนเทศทรัพยากรน้ำ (องค์การมหาชน) จากเว็บไซต์ https://www.hii.or.th จำนวน 4 สถานี ประกอบด้วยสถานี PAS001 (P1), PAS002 (P2), PAS003 (P3) และ PAS004 (P4) เป็นสถานีที่วัดค่าระดับน้ำและสภาพอากาศบริเวณแม่น้ำป่าสักตอนบน โดยมีสภาพทางภูมิศาสตร์ของสถานีวัดค่าตัวแปร ตามที่แสดงในรูปที่ 1

ข้อมูลจากแต่ละสถานีประกอบด้วยข้อมูลรายชั่วโมงของปริมาณฝน (RN) อุณหภูมิ (TM) ความชื้นสัมพัทธ์ (HM) ความกดอากาศ (PS) และระดับน้ำริมตลิ่ง (WL) ซึ่งบันทึกในไฟล์รูปแบบ csv แยกไฟล์ตามประเภทของข้อมูล และแยกแฟ้มตามเดือน และปี นำมารวมให้อยู่ในไฟล์เดียวกัน โดยใช้วันที่และเวลารายชั่วโมงเป็นคีย์หลัก กำหนดประเภทของข้อมูลเป็นคอลัมน์ (variables)

3.2 การเตรียมข้อมูล

3.2.1 การสำรวจทำความเข้าใจข้อมูล พบว่ามีข้อมูลตั้งแต่วันที่ 1 ม.ค. 2012 ถึงวันที่ 31 ธ.ค. 2022 มีจำนวนข้อมูลรวมทั้งสิ้น 96,432 แถว (observations) และ 22 คอลัมน์ (variables) และพบว่ามีข้อมูลที่ขาดหายไปในห้วงปี 2014 ถึง 2018 จำนวนมาก จึงเลือกห้วงเวลาของข้อมูลตั้งแต่เดือน ก.ย. 2018 ถึงเดือน ธ.ค. 2022 เหลือข้อมูลรวมจำนวนทั้งสิ้น 37,968 แถว (observations) จากนั้นดำเนินการปรับปรุงข้อมูลและทดแทนข้อมูลที่สูญหายด้วยเทคนิคสมการเชิงเส้น (linear interpolation)

3.2.2 การวิเคราะห์ตัวแปร ด้วยวิธีสหสัมพันธ์ (Coefficient correlation) พบว่าตัวแปรที่มีค่าความสัมพันธ์กับระดับน้ำ ณ สถานี P4 ได้แก่ ระดับน้ำ WL_{P1}, WL_{P2}, WL_{P3} ค่าความสัมพันธ์เท่ากับ 0.71, 0.78 และ 0.83 ตามลำดับ และความชื้นสัมพัทธ์ HM_{P1}, HM_{P2}, HM_{P3} และ HM_{P4} ค่าความสัมพันธ์เท่ากับ 0.29, 0.27, 0.32 และ 0.30 ตามลำดับ สำหรับข้อมูลความกดอากาศ อุณหภูมิ และปริมาณฝนมีความสัมพันธ์กับระดับน้ำ ณ สถานี P4 ในระดับที่น้อยมาก

3.2.3 การปรับขนาดข้อมูล เนื่องจากข้อมูลตัวแปรมีช่วงของข้อมูลที่แตกต่างกันมากจึงต้องปรับข้อมูลให้อยู่ในขนาดเดียวกันไม่ให้ตัวแปรใดตัวแปรหนึ่งมีอิทธิพลมากเกินไป และช่วยให้การประมวลผลในการเรียนรู้ของเครื่องมีความเร็วมากขึ้นโดยใช้เทคนิค min-max normalization ดังสมการที่ 3

$$scale = \frac{(x - min(x))}{(max(x) - min(x))} \quad (3)$$

3.2.4 การปรับปรุงตัวแปร เพื่อให้การพยากรณ์มีความแม่นยำมากขึ้นได้นำข้อมูลจากคาบเวลาย้อนหลังของแต่ละข้อมูลมาใช้เป็นตัวแปรต้นด้วย ทั้งนี้ได้ทดลองเพิ่มตัวแปรย้อนหลัง 1, 2, และ 3 ชม. ผลการทดลองให้ผลลัพธ์ไม่ต่างกันมากนัก จึงเพิ่มข้อมูลย้อนหลังเพียง 1 ชม. เพื่อให้จำนวนตัวแปรนำเข้าไม่มากเกินความจำเป็น สำหรับการพยากรณ์ค่าระดับน้ำล่วงหน้า ได้นำข้อมูลจากคาบเวลาถัดไปมาเป็นตัวแปรตาม ดังแสดงในตารางที่ 1 ผลจากการปรับข้อมูลดังกล่าวทำให้ข้อมูลบางส่วนถูกตัดออกไปเหลือจำนวน 37,780 แถว (observations)

ตารางที่ 1 การปรับปรุงข้อมูลด้วยคาบเวลาย้อนหลังและคาบเวลาล่วงหน้า

วันที่	เวลา	ตัวแปรต้น ณ เวลา t และเวลาย้อนหลัง				ตัวแปรตาม ณ เวลา t และเวลาล่วงหน้า		
		$WL_{P4}(t)$	$WL_{P4}(t-1)$	$HM_{P4}(t)$	$HM_{P4}(t-1)$	$WL_{P4}(t)$	$WL_{P4}(t+1)$	$WL_{P4}(t+2)$
2022-08-31	18:00:00	142.42	142.47	72.6	66.5	64.48	64.47	64.46
2022-08-31	19:00:00	142.38	142.42	71.1	72.6	64.47	64.46	64.45
2022-08-31	20:00:00	142.33	142.38	74.2	71.1	64.46	64.45	64.45
2022-08-31	21:00:00	142.27	142.33	79.4	74.2	64.45	64.45	64.44

3.3 การสร้างแบบจำลอง

สร้างแบบจำลองด้วยภาษา R และใช้โปรแกรม R studio นำเข้าไลบรารี neuralnet และ xgboost สำหรับการสร้างแบบจำลองการเรียนรู้ของเครื่องด้วยเทคนิค ANN, และ XGBoost ตามลำดับ โดยการค้นหาค่าพารามิเตอร์ที่เหมาะสมที่จะสามารถให้ผลการพยากรณ์ที่ดีที่สุดด้วยการใช้ไลบรารี caret::train ค้นหาและเปรียบเทียบค่าพารามิเตอร์แบบกำหนดค่าตัวเลือกล่วงหน้า (Grid search Hyperparameter Tuning)

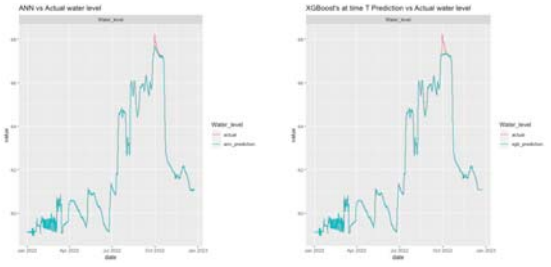
สำหรับข้อมูลได้แบ่งออกเป็น 2 ส่วนประกอบด้วย ส่วนแรกใช้ข้อมูลห้วงเดือน ก.ย. 2018 ถึง ธ.ค.2021 สำหรับการฝึกสอนจำนวน 29,188 แถว (77.25%) ดำเนินการฝึกสอนการเรียนรู้ด้วยเครื่องและการปรับปรุงค่าพารามิเตอร์ให้มีความเหมาะสม และส่วนที่สองห้วงเดือน ม.ค. 2021 ถึง ธ.ค. 2022 จำนวน 8,592 แถว (22.75%) ใช้สำหรับการทดสอบแบบจำลอง จากนั้นบันทึกผลและประเมินประสิทธิภาพการพยากรณ์ด้วยค่าความคลาดเคลื่อนเฉลี่ยสัมบูรณ์ (MAE) และค่าสัมประสิทธิ์การตัดสินใจ(R^2)

3.4 ผลการทดลอง

สถานี P4 ในห้วงเวลาทดสอบมีระดับน้ำเฉลี่ย 61.89 เมตร สูงสุด 68.23 ม. และต่ำสุด 58.96 ม. ผลการพยากรณ์ ณ เวลาที่พยากรณ์ (t) อัลกอริทึม ANN มีผลการประเมินประสิทธิภาพด้วยค่าความคลาดเคลื่อนเฉลี่ยสัมบูรณ์ (MAE) เท่ากับ 2.42 ซม. และค่าสัมประสิทธิ์การตัดสินใจ (R^2) เท่ากับ 0.9993 สำหรับอัลกอริทึม XGBoost มีผลการประเมินประสิทธิภาพด้วยค่าความคลาดเคลื่อนเฉลี่ยสัมบูรณ์เท่ากับ 3.14 ซม. และค่าสัมประสิทธิ์การตัดสินใจเท่ากับ 0.9987 ผลการพยากรณ์ล่วงหน้าดังแสดงในตารางที่ 2

4. อภิปรายผลการทดลอง

การพยากรณ์ระดับน้ำสถานี P4 ด้วยอัลกอริทึม ANN และ XGBoost ณ เวลาที่พยากรณ์ (t) ผลการทดลองปรากฏว่ามีผลการพยากรณ์ที่ใกล้เคียงกัน แบบจำลองที่สร้างด้วยอัลกอริทึม ANN มีความคลาดเคลื่อนน้อยกว่าแบบจำลองที่สร้างด้วยอัลกอริทึม XGBoost เมื่อนำผลการพยากรณ์มาสร้างเป็นกราฟเส้นเปรียบเทียบกับระดับน้ำจริง ตามรูปที่ 2



(a) การพยากรณ์ด้วยอัลกอริทึม ANN (b) การพยากรณ์ด้วยอัลกอริทึม XGBoost

รูปที่ 2 แสดงกราฟการพยากรณ์ด้วยอัลกอริทึม ANN และ XGBoost เทียบกับระดับน้ำจริง ณ เวลาที่พยากรณ์ (t)

พบว่าการพยากรณ์ด้วยอัลกอริทึม ANN ค่อนข้างถูกต้องแม่นยำอย่างสม่ำเสมอตลอดห้วงระยะเวลาของการทดสอบ แม้ว่าในห้วงเวลาที่มีการเปลี่ยนแปลงระดับน้ำอย่างรวดเร็วในห้วงต้นเดือน ต.ค. ปี 2022 แบบจำลอง ANN สามารถพยากรณ์ได้ใกล้เคียงมากกว่าอัลกอริทึม XGBoost

เมื่อทำการการพยากรณ์ระดับน้ำล่วงหน้า 1 ถึง 12 ชั่วโมง อัลกอริทึม ANN สามารถพยากรณ์ได้แม่นยำและมีค่าความคลาดเคลื่อนเฉลี่ยสัมบูรณ์น้อยกว่าอัลกอริทึม XGBoost แต่เมื่อทำการพยากรณ์ล่วงหน้านานขึ้นเป็น 24, 48 และ 72 ชั่วโมง พบว่าอัลกอริทึม XGBoost กลับมีผลการพยากรณ์ที่แม่นยำกว่า อัลกอริทึม ANN สำหรับด้านการประมวลผลพบว่าอัลกอริทึม ANN ใช้หน่วยความจำและเวลาในการประมวลผลมากกว่าอัลกอริทึม XGBoost

5. สรุป

การพยากรณ์ระดับน้ำบริเวณแม่น้ำป่าสักตอนบน ด้วยชุดข้อมูลที่ใช้ในการศึกษาครั้งนี้ กระบวนการเรียนรู้ของเครื่องด้วยอัลกอริทึม XGBoost และอัลกอริทึม ANN มีประสิทธิภาพมาก ผลการพยากรณ์มีความแม่นยำใกล้เคียงกัน สามารถนำไปใช้เพื่อเป็นข้อมูลประกอบการตัดสินใจบริหารจัดการทรัพยากรน้ำบริเวณเหนือเขื่อนป่าสักชลสิทธิ์ล่วงหน้าประมาณ 1 ถึง 24 ชั่วโมง

สำหรับอัลกอริทึม ANN มีความแม่นยำมากกว่าอัลกอริทึม XGBoost ในการพยากรณ์ ณ เวลา t รวมถึงการพยากรณ์ล่วงหน้า 1 ถึง 12 ชั่วโมง ในขณะที่อัลกอริทึม XGBoost มีความแม่นยำในการพยากรณ์ล่วงหน้า 24, 48 และ 72 ชม. มากกว่าอัลกอริทึม ANN

ตารางที่ 2 แสดงผลการทดลองพยากรณ์ระดับน้ำ ณ สถานี P4

เวลาที่พยากรณ์	ANN		XGBoost	
	MAE	R square	MAE	R square
ณ เวลาที่พยากรณ์ (t)	0.0242	0.9993	0.0314	0.9987
พยากรณ์ล่วงหน้า 1 ชม. (t+1)	0.0350	0.9991	0.0399	0.9981
พยากรณ์ล่วงหน้า 2 ชม. (t+2)	0.0411	0.9989	0.0473	0.9979
พยากรณ์ล่วงหน้า 3 ชม. (t+3)	0.0475	0.9986	0.0539	0.9977
พยากรณ์ล่วงหน้า 6 ชม. (t+6)	0.0616	0.9981	0.0727	0.9969
พยากรณ์ล่วงหน้า 12 ชม. (t+12)	0.1010	0.9856	0.1050	0.9951
พยากรณ์ล่วงหน้า 24 ชม. (t+24)	0.1961	0.9856	0.1660	0.9899
พยากรณ์ล่วงหน้า 48 ชม. (t+48)	0.2986	0.9675	0.2585	0.9766
พยากรณ์ล่วงหน้า 72 ชม. (t+72)	0.4648	0.9207	0.3470	0.9604

การศึกษาครั้งต่อไปควรปรับปรุงข้อมูลปริมาณน้ำฝนรายชั่วโมงเป็นข้อมูลปริมาณน้ำฝนสะสม รวมถึงข้อมูลอัตราการระเหยของน้ำ และข้อมูลอัตราการระบายน้ำออกจากเขื่อน มาใช้ในการพยากรณ์ร่วมด้วย อาจจะทำให้การพยากรณ์มีความแม่นยำมากขึ้น

ทั้งนี้ข้อมูลที่นำมาใช้ในการศึกษาครั้งนี้แม้จะมีช่วงระยะเวลาการบันทึกข้อมูลจำนวนมาก แต่ก็มีข้อมูลที่สูญหายมากเช่นกัน ดังนั้นควรมีการพัฒนาระบบเซ็นเซอร์ในการวัดค่าสภาพอากาศและระดับน้ำที่มีความถูกต้อง แม่นยำ และเชื่อถือได้ รวมถึงการพัฒนากระบวนการจัดเก็บข้อมูล เพื่อลดโอกาสของการเกิดข้อมูลสูญหายให้น้อยลง จะช่วยเพิ่มประสิทธิภาพการพยากรณ์ให้มากขึ้น

เอกสารอ้างอิง

[1] ทวี ชัยพิมลผลิน และ ทวีศักดิ์ วังไพศาล, “แบบจำลองโครงข่ายประสาทเทียมสำหรับการพยากรณ์น้ำท่วมในลุ่มน้ำมูลตอนล่าง Artificial Neural Network Models for Flood Forecasting in Lower Mun Catchment”, *J Sci Technol MSU*, ปี 35, ฉบับที่ 5, น. 587–595, 2016.

[2] สาลินันท์ บุญมี, แสงทอง บุญยิ่ง, และ อเนก พุทธิเดช, “การพยากรณ์ปริมาณน้ำหลากเขตพื้นที่ทำเกษตรของชุมชนในเขตรับน้ำจังหวัดพระนครศรีอยุธยา โดยแบบจำลองโครงข่ายประสาทเทียม”, *วารสารวิทยาศาสตร์เกษตร*, ปี 53, ฉบับที่ 2, น. 116–128, 2022.

[3] C. Choi, J. Kim, H. Han, D. Han, & H. S. Kim, “Development of water level prediction models using machine learning in wetlands: A case study of Upo wetland in South Korea”, *Water (Switzerland)*, vol. 12, no.1, Jan., 2020, doi: 10.3390/w12010093.

[4] A. Ibrahem Ahmed Osman, A. Najah Ahmed, M. F. Chow, Y. Feng Huang, & A. El-Shafie, “Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia”, *Ain Shams Engineering Journal*, vol. 12, no. 2, June, pp. 1545–1556, 2021, doi: 10.1016/j.asej.2020.11.011.

[5] B. Lantz, *Machine learning with R : discover how to build machine learning algorithms, prepare data, and dig deep into data prediction techniques with R*, Second edition. Packt Publishing, 2015. สืบค้น: 28 พฤษภาคม 2023. [ออนไลน์]. Available at: www.packtpub.com

[6] ทวีศักดิ์ วังไพศาล และ จักรฤทธิ์ ตรีนาจ, “ปัจจัยที่มีผลต่อความแม่นยำในการพยากรณ์ระดับน้ำหลากที่สถานีวัดระดับน้ำ M.7 โดยใช้แบบจำลองโครงข่ายประสาทเทียม Factors Affecting the Accuracy of Water Level Forecasting at M.7 Gauge Station Using Artificial Neural Network Model”.

[7] T. Chen & C. Guestrin, “XGBoost: A Scalable Tree Boosting System”, Mar., 2016, doi: 10.1145/2939672.2939785..