

การพยากรณ์ข้อมูลอนาคตของข้อมูลอนุกรมเวลาด้วยเครื่องจักรการเรียนรู้

กรสิริณัฐ โรจนวรรณ^{1*} และ พยุง มีสัจ²

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์ เพื่อนำเสนอขั้นตอนวิธีการสร้างรูปแบบการพยากรณ์ข้อมูลอนุกรมเวลาโดยใช้ข้อมูลอนุกรมหลายตัวแปรสำหรับการสอนเครื่องจักรการเรียนรู้ ในงานวิจัยนี้คัดเลือกเครื่องจักรการเรียนรู้สำหรับเป็นตัวแบบพยากรณ์จาก 4 เทคนิค ได้แก่ เทคนิคค่าเฉลี่ยเคลื่อนที่ เพื่อนบ้านใกล้เคียง ระบบผสมโครงข่ายประสาทเทียม และตรรกคลุมเครือ โครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น ในการวิจัยใช้ข้อมูลอนุกรมเวลาปริมาณน้ำของกรมชลประทานที่ 15 โครงการพระราชดำริลุ่มน้ำปากพนัง ในช่วงระยะเวลา 7 ปี จำนวน 12 ตัวแปร โดยใช้ข้อมูลน้ำตัวแปรที่ 1 ถึง 11 เป็นอินพุตสำหรับพยากรณ์ตัวแปรที่ 12 ผลการวิจัยพบว่า โครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้นให้ผลการทำนายในชุดทดลองได้ดีที่สุด โดยมีค่าผิดพลาดในการพยากรณ์เท่ากับ 0.486 รองลงมาได้แก่ เทคนิคเพื่อนบ้านใกล้เคียง มีค่าผิดพลาดในการพยากรณ์เท่ากับ 1.376 เทคนิคค่าเฉลี่ยเคลื่อนที่ มีค่าผิดพลาดในการพยากรณ์เท่ากับ 2.169 และระบบผสมโครงข่ายประสาทเทียมตรรกคลุมเครือ มีค่าผิดพลาดในการพยากรณ์เท่ากับ 4.284 ตามลำดับ

คำสำคัญ: ข้อมูลอนุกรมเวลา; การพยากรณ์; เครื่องจักรการเรียนรู้

รับพิจารณา: 23 กรกฎาคม 2561

แก้ไข: 18 พฤษภาคม 2565

ตอบรับ: 27 พฤษภาคม 2565

¹ นักศึกษาระดับปริญญาเอก ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

² รองศาสตราจารย์ประจำภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

* ผู้นิพนธ์ประสานงาน โทร. +666 5348 6952 อีเมล: kornsirinut.r@email.kmutnb.ac.th

Forecasting the Future of Time Series Data Using Machine Learning

Kornsirinut Rothjanawan^{1*} and Phayung Meesud²

Abstract

The purpose of this presents a methodology for creating time series forecasting models using multivariable time series data for training machine learning. In this research, we propose a model for predicting future data for time series data. We select four machine learning techniques: Moving Average (MAV), K-Nearest Neighbor (KNN), Adaptive Neuro Fuzzy Inference Systems (ANFIS) and Multilayer Perceptron (MLP). This research uses water volume of 15 Royal Irrigation Project during 7-year period, 12 variables of time series water data, Variables 1 through 11 are used as the input for predicting Variable 12. The results show that the MLP performed best prediction with rmse equals 0.486. The other techniques yield rmse equal 1.376, 2.169, and 4.284, for KNN, Moving Average, and ANFIS, respectively.

Keywords: Time Series; Forecasting; Machine Learning

Received: July 23, 2018

Revised: May 18, 2022

Accepted: May 27, 2022

¹ Doctoral Degree Student, Department of Information Technology, Faculty of Information Technology, King Mongkut's University of Technology North Bangkok

² Associate Professor, Department of Information Technology, Faculty of Information Technology, King Mongkut's University of Technology North Bangkok

* Corresponding Author Tel. +666 5348 6952 e-mail: kornsirinut.r@email.kmutnb.ac.th

1. บทนำ

การพยากรณ์ข้อมูลอนุกรมเวลา [1] เป็นปัญหาที่ทำได้ไม่่ง่ายนัก เนื่องจากข้อมูลอนุกรมเวลามีความเป็นพลวัตและโกลาหล (dynamics and chaos) มีงานวิจัยจำนวนมากได้นำเสนอการพยากรณ์ข้อมูลอนุกรมเวลา [2] เทคนิคที่นิยม ได้แก่ เพื่อนบ้านใกล้เคียง โครงข่ายประสาทเทียม การถดถอยแบบไม่เป็นเชิงเส้น ซัพพอร์ตเวกเตอร์รีเกรสชัน และระบบฟuzzyลอจิก เป็นต้น การแทนค่าสูญหายจะถูกใช้ก่อนนำข้อมูลไปสร้างโมเดลสำหรับการพยากรณ์ค่าอนุกรมเวลา

การพยากรณ์ข้อมูลอนุกรมเวลาถูกนำไปประยุกต์ใช้กับงานหลายด้าน [3] เช่น การวิเคราะห์แนวโน้ม การวิเคราะห์การถดถอย ขั้นตอนวิธีทางพันธุกรรม การพยากรณ์การไหลของน้ำ เป็นต้น ความยุ่งยากในการพยากรณ์ข้อมูลอนุกรมเวลามีมากขึ้น

ในงานวิจัยครั้งนี้ได้นำเสนอวิธีการสร้างรูปแบบการพยากรณ์อนุกรมเวลาโดยใช้ข้อมูลหลายตัวแปรเป็นอินพุต สำหรับการพยากรณ์ข้อมูลสูญหายสำหรับข้อมูลอนุกรมเวลา เพื่อนำไปประยุกต์ใช้ในการพยากรณ์ระดับน้ำ [4] ซึ่งงานวิจัยนี้ใช้ข้อมูลระดับน้ำที่ประตูทกวิชาประสิทธิของโครงการพระราชดำริ ลุ่มน้ำปากพนัง จังหวัดนครศรีธรรมราช

เนื้อหาต่อจากนี้ถูกแบ่งเป็นส่วนดังนี้ ส่วนที่ 2 กล่าวถึงทฤษฎีที่เกี่ยวข้อง ส่วนที่ 3 วิธีการดำเนินการวิจัย ส่วนที่ 4 ผลการดำเนินงานวิจัย และส่วนที่ 5 กล่าวถึงการสรุปผล การอภิปรายผลและข้อเสนอแนะ

2. วัตถุประสงค์ของการวิจัย

เพื่อนำเสนอขั้นตอนวิธีการสร้างรูปแบบการพยากรณ์ข้อมูลอนุกรมเวลาโดยใช้ข้อมูลอนุกรมหลายตัวแปร

3. ทฤษฎีที่เกี่ยวข้อง

3.1 การทำเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูล [5] เป็นสกัดองค์ความรู้ที่ซ่อนอยู่ในข้อมูลเพื่อใช้ในการทำนายแนวโน้มและพฤติกรรม

โดยอาศัยข้อมูลในอดีตเพื่อค้นหารูปแบบความสัมพันธ์และองค์ความรู้ใหม่จากข้อมูล โดยมีขั้นตอนดังนี้

1) ทำความเข้าใจปัญหา โดยการเลือกข้อมูลให้มีความเหมาะสมกับอัลกอริธึมที่ใช้ งาน จำนวนที่ต้องการ และค่าเป้าหมายเพื่อให้ได้ผลลัพธ์ที่ต้องการ

2) ทำความเข้าใจข้อมูล โดยการรวบรวมตรวจสอบความถูกต้องและกำหนดคุณสมบัติที่ต้องการให้กับข้อมูล

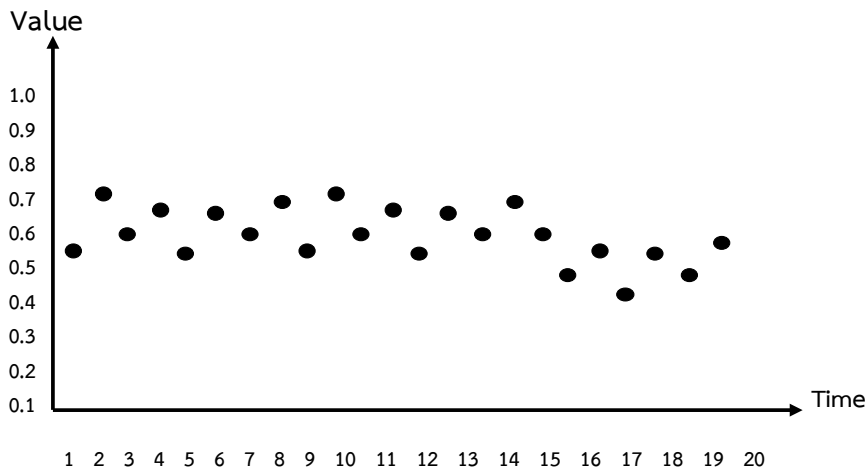
3) เตรียมข้อมูล โดยการคัดเลือกข้อมูลเพื่อทำการแปลงให้อยู่ในรูปแบบที่เหมาะสมต่อการวิเคราะห์ข้อมูลด้วยเทคนิคต่าง ๆ

4) สร้างแบบจำลอง แบบจำลองแบ่งเป็น 2 ประเภทคือ (1) แบบจำลองเพื่อการทำนาย (Predictive model) เป็นการคาดคะเนลักษณะหรือประมาณค่าของข้อมูลที่จะเกิดขึ้นในอนาคต โดยใช้ข้อมูลในอดีต และ (2) แบบจำลองเชิงบรรยาย (Descriptive model) เพื่อหาแบบจำลองมาอธิบายลักษณะของข้อมูล

3.2 การวิเคราะห์ข้อมูลอนุกรมเวลา (Time Series Analysis)

เป็นชุดข้อมูลเชิงปริมาณที่มีการจัดบันทึก ช่วงเวลาใด เวลาหนึ่ง จะมีการจัดเรียงตามเวลาที่จัดบันทึกข้อมูล โดยมีเวลาห่างเท่า ๆ กัน ซึ่งอาจจะมีรูปแบบหรือไม่มีรูปแบบ เช่น รายวัน รายสัปดาห์ รายเดือน และรายปี ก็ได้ ตัวอย่างเช่น ข้อมูลราคาหุ้น ข้อมูลปริมาณน้ำฝนรายวัน ข้อมูลปริมาณแม่น้ำ เป็นต้น

ข้อมูลอนุกรมเวลา [6] ที่เกิดขึ้นอย่างต่อเนื่องกันเป็นระยะเวลาอันยาวนานวิเคราะห์หาความสัมพันธ์ เพื่อนำข้อมูลในอดีตมาใช้ในการพยากรณ์ข้อมูลที่จะเกิดขึ้นในอนาคตได้ เนื่องจากค่าของข้อมูลหรือลักษณะต่าง ๆ ที่เกิดขึ้นกับข้อมูลนั้น จะมีรูปแบบเฉพาะที่มีความสัมพันธ์กันในเชิงเวลา จึงมีการนำข้อมูลจากช่วงเวลาที่ผ่านมาในอดีตมาทำการวิเคราะห์ เพื่อให้สามารถพยากรณ์ข้อมูลที่จะเกิดขึ้นในอนาคต ซึ่งข้อมูลที่น่ามาใช้ในการวิเคราะห์หากมีจำนวนมากและมีความสมบูรณ์ ผลของการวิเคราะห์หรือการพยากรณ์ จะมีความถูกต้องใกล้เคียงค่าจริงมากขึ้น ลักษณะของข้อมูลที่แสดงความสัมพันธ์กับเวลาแสดงดังรูปที่ 1



รูปที่ 1 ตัวอย่างข้อมูลอนุกรมเวลา

3.3 ค่าเฉลี่ยเคลื่อนที่เคลื่อนที่ (Moving Average: MAV)

ค่าเฉลี่ยเคลื่อนที่เป็นเทคนิคง่ายในการทำเหมืองข้อมูลวิธีหนึ่ง ซึ่งทำได้โดยเลื่อนหน้าต่างในข้อมูลอนุกรมเวลาและทำการหาค่าเฉลี่ยในกรอบหน้าต่างสำหรับการเลื่อนข้อมูลแต่ละครั้งของตัวแปรที่ต้องการพยากรณ์

3.4 เทคนิคเพื่อนบ้านใกล้เคียง (K-Nearest Neighbor: KNN)

เทคนิคเพื่อนบ้านใกล้เคียงเป็นวิธีการทำเหมืองข้อมูลที่ไม่ต้องสอนโมเดล โดยเทคนิคนี้จะเก็บข้อมูลสอนไว้ทั้งหมด เมื่อต้องการพยากรณ์ค่าเอาต์พุต จะทำการวัดค่าความคล้ายของข้อมูลอินพุตเทียบกับข้อมูลชุดสอน และเลือกข้อมูลชุดสอนที่มีความคล้ายกับข้อมูลอินพุตมากที่สุดจำนวน K เรคคอร์ด ผลพยากรณ์เอาต์พุตจะได้จากค่าเฉลี่ย K เรคคอร์ดนั่นเอง

3.5 ระบบผสมโครงข่ายประสาทเทียมและตรรกคลุมเครือ (Adaptive Neuro Fuzzy Inference System: ANFIS)

ANFIS เป็นระบบผสมโครงข่ายประสาทเทียมและตรรกคลุมเครือ ซึ่งมีความสามารถในการเรียนรู้เช่นเดียวกับโครงข่ายประสาทเทียมแบบก้าวไปข้างหน้า (Feedforward network) และใช้หลักการตรรกศาสตร์คลุมเครือโดยแปลงเป็นกฎคล้ายการให้เหตุผลของมนุษย์ ซึ่งเป็นคุณลักษณะที่ดียิ่งหนึ่ง [7]

3.6 โครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น (Multilayer Perceptron: MLP) [8]

โครงข่ายประสาทเทียมมีพื้นฐานมาจากการจำลองการทำงานของสมองมนุษย์ ด้วยโปรแกรมคอมพิวเตอร์ เนื่องจากต้องการให้คอมพิวเตอร์มีความชาญฉลาดในการเรียนรู้เหมือนมนุษย์ที่สามารถเรียนรู้จากข้อมูลได้ สามารถฝึกฝนได้ และสามารถนำความรู้และทักษะโครงข่ายประสาทเทียมเปอร์เซ็ปตรอนแบบหลายชั้น [9] เป็นที่นิยมใช้งานมากในการจำแนกข้อมูลแบบไม่เป็นเชิงเส้น งานมีความยากในการตัดสินใจหรือความซับซ้อนสูง และงานพยากรณ์ค่าจริง เป็นต้น

4. วิธีดำเนินการวิจัย

ในงานวิจัยนี้นำเสนอวิธีการสร้างตัวแบบการพยากรณ์ข้อมูลอนาคตของอนุกรมเวลาโดยใช้อนุกรมเวลาแบบหลายตัวแปรเป็นอินพุต สำหรับตัวแบบการพยากรณ์ผู้วิจัยเลือกจาก วิธีเฉลี่ยเคลื่อนที่ เพื่อนบ้านใกล้เคียง ระบบผสมโครงข่ายประสาทเทียมและตรรกคลุมเครือ และโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น (MLP) โดยดำเนินการวิจัยตามขั้นตอน 3 ขั้นตอน ดังนี้

4.1 การรวบรวมและวิเคราะห์ข้อมูล

ข้อมูลอนุกรมเวลาที่นำมาใช้เป็นข้อมูลน้ำจากกรมชลประทานที่ 15 โครงการพระราชดำริลุ่มน้ำปากพนัง ในช่วงระยะเวลา 7 ปี ข้อมูลที่ใช้ในการวิจัยประกอบด้วย ประกอบด้วย 12 ตัวแปร ได้แก่ A1:

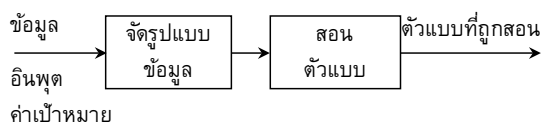
ข้อมูลระดับเหนือเวลา 6.00 น. A2: ข้อมูลระดับเหนือเวลา 12.00 น. A3: ข้อมูลระดับเหนือเวลา 18.00 น. A4: ข้อมูลระดับท้ายน้ำเวลา 6.00 น. A5: ข้อมูลระดับท้ายน้ำเวลา 12.00 น. A6: ข้อมูลระดับท้ายน้ำเวลา 18.00 น. A7: ระดับน้ำสูงสุดหน้าประตู A8: ระดับน้ำสูงสุดท้ายประตู A9: อัตราระบายน้ำ/วินาที A10: อัตราระบายน้ำ/วัน A11: ปริมาณน้ำฝน และ A12: ปริมาณน้ำเก็บกัก

ตัวแปรที่ 1 ถึง 11 เป็นตัวแปรอิสระ ส่วนตัวแปรที่ 12 เป็นตัวแปรตาม งานวิจัยนี้เน้นการใช้ตัวแปร 1 ถึง 11 เพื่อพยากรณ์ตัวแปรที่ 12 ซึ่งเป็นปริมาณน้ำเก็บกักเหนือเขื่อน

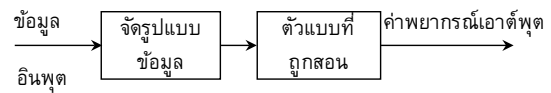
4.2 การออกแบบโมเดลการพยากรณ์ข้อมูลสูญหายสำหรับข้อมูลอนุกรมเวลา

ผู้วิจัยได้ออกแบบขั้นตอนวิธีการพยากรณ์ค่าในอนาคตของข้อมูลอนุกรมเวลาตัวแปรเป้าหมาย โดยเริ่มจากการออกแบบรูปแบบข้อมูลซึ่งแปลงจากอนุกรมเวลาเป็นเวกเตอร์และเมทริกซ์โดยการเลื่อนหน้าต่าง โดยการเลื่อนหนึ่งครั้งจะได้ข้อมูลหนึ่งเรคคอร์ดทั้งอินพุตและค่าเป้าหมาย การแปลงข้อมูลเพื่อให้ได้รูปแบบข้อมูลที่เหมาะสมกับการเรียนรู้ของเครื่อง ได้แก่ เทคนิคค่าเฉลี่ยเคลื่อนที่ เทคนิคเพื่อนบ้านใกล้เคียง ระบบผสมโครงข่ายประสาทเทียมและตรรกคลุมเครือ และโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น

การทำโมเดลเหมือนข้อมูลเริ่มต้นจาก 1) การเลือกตัวแปรเป้าหมาย 2) การจัดข้อมูลอินพุตและเอาต์พุต 3) การแบ่งข้อมูลสอน ข้อมูลตรวจสอบ และข้อมูลทดสอบ 4) การสร้างโมเดลการพยากรณ์ด้วยเครื่องจักรการเรียนรู้ และ 5) การประยุกต์ใช้โมเดลการพยากรณ์ไปใช้งาน แสดงดังรูปที่ 2 และรูปที่ 3



รูปที่ 2 โมเดลการสร้างตัวแบบการพยากรณ์ในอนุกรมเวลา



รูปที่ 3 โมเดลการใช้งานตัวแบบการพยากรณ์

สำหรับโมเดลการสร้างตัวแบบการพยากรณ์ข้อมูลในอนุกรมเวลา แสดงดังรูปที่ 2 และโมเดลการใช้งานตัวแบบการพยากรณ์ แสดงดังรูปที่ 3 ซึ่งมีขั้นตอนการทำงานดังนี้ 1) รับเข้าข้อมูลอนุกรมเวลาแบบหลายตัวแปร 2) เลือกตัวแปรเป้าหมายสำหรับการพยากรณ์ 3) สำหรับตัวแปรอินพุตทำการจัดข้อมูลอนุกรมเวลา [10] โดยเลื่อนหน้าต่างข้อมูลแต่ละตัวย้อนหลังจำนวน N วัน เพื่อจัดเป็นเวกเตอร์อินพุตและตั้งค่าเป้าหมายที่เวลา ล่วงหน้า 1 วัน ($t + 1$) จากตัวแปรที่ เลือกไว้เป็นค่าเป้าหมายสำหรับการพยากรณ์ค่าข้อมูลในอนาคต

ในการสอนเครื่องจักรการเรียนรู้ ผู้วิจัยนำเสนอวิธีการจัดรูปแบบข้อมูลโดยใช้ข้อมูลในอดีตเพื่อพยากรณ์ข้อมูลในอนาคตซึ่งต้องทำการเลื่อนหน้าต่างข้อมูล โดยจัดให้อยู่ในรูปเวกเตอร์ และเมทริกซ์ตามสมการ (1) – (6)

$$\mathbf{x}_{j,t} = [x_{j,t-N} \dots x_{j,t-1} x_{j,t}], j = 1, \dots, R \quad (1)$$

$$\mathbf{p}_t = [x_{1,t} \ x_{2,t} \dots x_{R,t}], t = 1, \dots, M \quad (2)$$

$$y_i = x_{Q,i}, Q \in \{j = 1, \dots, R\} \quad (3)$$

$$\mathbf{P} = [\mathbf{p}_1^T \ \mathbf{p}_2^T \dots \ \mathbf{p}_M^T] \quad (4)$$

$$\mathbf{y} = [y_1 \ y_2 \dots \ y_M] \quad (5)$$

$$\hat{y}_t = f(\mathbf{p}_t^T, \text{parameters}) \quad (6)$$

เมื่อ $\mathbf{x}_{j,t}$ คือ เวกเตอร์หน้าต่างข้อมูลที่มีตัวประกอบ

เป็นข้อมูลจากตัวแปร j ที่เวลา

$t - N, \dots, t - 1$ และ t

\mathbf{p}_t คือ เวกเตอร์ที่ประกอบด้วย $\mathbf{x}_{j,t}$

\mathbf{P} คือ เมทริกซ์ที่มีคอลัมน์เป็นอินพุตเวกเตอร์

\mathbf{y} คือ เป็นเวกเตอร์ค่าเป้าหมายตามคอลัมน์ของเมทริกซ์ \mathbf{P}

\hat{y} คือ ค่าพยากรณ์

f คือ ฟังก์ชันหรือโมเดลการพยากรณ์



- N คือ จำนวนวันย้อนหลังในอดีต
 - M คือ จำนวนเรคคอร์ด
 - R คือ จำนวนตัวแปรอินพุต
 - Q คือ ตัวแปรเป้าหมายซึ่งต้องการพยากรณ์
- $parameters$ คือ ค่าพารามิเตอร์ของโมเดลการพยากรณ์

หลักการสร้างโมเดลพยากรณ์ด้วยวิธีเครื่องจักรการ เรียนรู้ โดยนำข้อมูลอนุกรมเวลามาใช้ในการสร้าง โมเดล เพื่อนำไปเป็นอินพุตและค่าเป้าหมาย สำหรับ สอนโมเดลการพยากรณ์ตัวแปรเป้าหมาย โดยแบ่ง ข้อมูลออกเป็นชุดสอน ชุดตรวจสอบ และชุดทดสอบ สำหรับเทคนิคการพยากรณ์ ด้วยการเปรียบเทียบ เทคนิค จำนวน 4 เทคนิค ได้แก่ ค่าเฉลี่ยเคลื่อนที่ (Moving Average) เพื่อนบ้านใกล้เคียง (K-Nearest Neighbor: KNN) ระบบผสมโครงข่ายประสาทเทียม และตรรกคลุมเครือ (Fuzzy Logic Systems) โครงข่าย ประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น (MLP) สำหรับการวัดประสิทธิภาพของโมเดล การพยากรณ์ค่า ในอนาคตของข้อมูลอนุกรมเวลาตัวแปรเป้าหมาย ผู้วิจัยใช้วิธีคำนวณรากที่สองของค่าเฉลี่ย ผิดพลาดยก กำลังสอง (root mean squared error: $rmse$) ดัง สมการ (7)

$$RMSE = \sqrt{\frac{1}{M} \sum_{i=1}^M (y_i - \hat{y}_i)^2} \quad (7)$$

- เมื่อ y_i คือ ค่าเป้าหมายจริงของข้อมูลทดสอบ
- \hat{y}_i คือ ค่าพยากรณ์
- M คือ จำนวนข้อมูลทดสอบ

5. ผลการศึกษา

ในการดำเนินงานวิจัยในครั้งนี้ ผู้วิจัยนำเสนอ ขั้นตอนวิธีการสร้างรูปแบบการพยากรณ์ข้อมูลอนุกรม เวลา โดยใช้ค่าในอนาคตของข้อมูลอนุกรมเวลาตัวแปร เป้าหมาย ด้วยการเลื่อนหน้าต่าง (Sliding Window) จากสมการที่ (1) – (6) เพื่อจัดรูปแบบข้อมูลและสอน โมเดลการเรียนรู้เครื่อง ได้แก่ เทคนิควิธีเฉลี่ยเคลื่อนที่ (Moving Average) วิธีสมาชิกใกล้เคียงที่สุด (KNN) ระบบ ผสมโครงข่ายประสาทเทียมและตรรกคลุมเครือ

(ANFIS) และโครงข่ายประสาทเทียมแบบเปอร์ เซ็ปตรอนหลายชั้น (MLP)

ในการทดลองได้ กำหนดขนาดหน้าต่างข้อมูลเท่า กับ 7 นั่นคือ จะใช้ข้อมูลวันปัจจุบันร่วมกับข้อมูล ใน อดีตอีก 6 วัน ย้อนหลัง ตั้งแต่ปี 2550 – 2556 จำนวน 2527 วัน (เรคคอร์ด) ข้อมูลถูกแบ่งเป็นชุดข้อมูลสอน 50% ชุดตรวจสอบ 25% และชุดทดสอบ 25%

ซอฟต์แวร์แมทแล็บถูกใช้ในการสร้างโมเดล ได้แก่ Moving Average, KNN, MLP และ ANFIS โดยที่ MLP เลือกใช้จาก Matlab Neural Network Toolbox ออกแบบเป็น 2 ชั้นซ่อน แต่ละชั้นซ่อนมี 15 นิวรอน ชนิด Hyperbolic Tangent Sigmoidal (tansig) และ ชั้นเอาต์พุตเป็นแบบเชิงเส้น ทำการสอนด้วยวิธี trainlm (Levenberg-Marquardt) ส่วน ANFIS ใช้ จาก Matlab Fuzzy System Toolbox ใช้ฟังก์ชันเซต แบบ Gaussian Bell Membership Function (gbellmf) จำนวน 5 ฟังก์ชันเซต และสอนด้วยวิธี Hybrid

เทคนิคการเรียนรู้เครื่องทั้งสี่แบบจะถูกสอนด้วย ข้อมูลสอน ตรวจสอบด้วยข้อมูลตรวจสอบ และทดสอบ ด้วยข้อมูลทดสอบ จะได้โมเดลพยากรณ์เพื่อนำไป ประยุกต์ใช้การพยากรณ์ข้อมูลในอนาคตของตัวแปร เป้าหมาย คือ ตัวแปรที่ 12 (A12) ปริมาณน้ำเก็บกัก เพื่อวิเคราะห์ประสิทธิภาพที่ได้จากเทคนิคทั้ง 4 เทคนิค โดยใช้เกณฑ์ในการวัดด้วยค่ารากที่สองของค่าเฉลี่ย ผิดพลาดยกกำลังสอง (root mean squared error: $rmse$) ซึ่งถ้าค่า RMSE มีค่าน้อย แสดงให้เห็นว่าเทคนิค ตัวนั้นมีค่าความผิดพลาดน้อย และมีค่าความถูกต้อง ใกล้เคียงกับค่าจริง ในการพยากรณ์ ถ้าแสดงดังตาราง ที่ 1

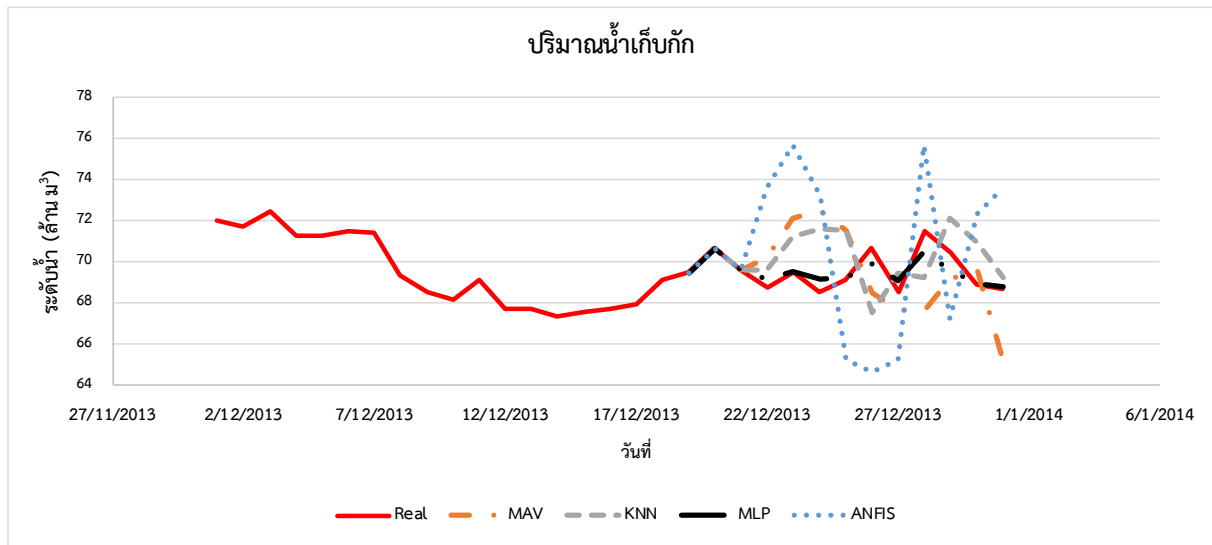
ตารางที่ 1 การเปรียบเทียบค่า RMSE ของโมเดลต่าง ๆ

	ROW	KNN	MLP	ANFIS
RMSE	2.169	1.376	0.486	4.284

จากตารางที่ 1 เป็นการแสดงค่าความผิดพลาดของ เทคนิคต่าง ๆ ที่นำไปใช้ในโมเดลการพยากรณ์ใน อนุกรมเวลาแบบหลายตัวแปร จะเห็นว่าโครงข่าย ประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น (MLP) ให้ ค่าผิดพลาดต่ำกว่าวิธีอื่น ๆ ได้แก่ สมาชิกใกล้เคียงที่สุด (KNN) วิธีเฉลี่ยเคลื่อนที่ (Moving Average) และระบบ

ผสมโครงข่ายประสาทเทียมและตรรกคลุมเครือ (ANFIS) ตารางที่ 2 และรูปที่ 4 แสดงผลการพยากรณ์

ด้วยเทคนิคต่าง ๆ จะเห็นว่า MLP พยากรณ์ข้อมูลทดสอบมีค่าใกล้เคียงค่าจริงกว่าเทคนิคอื่น ๆ



รูปที่ 4 ผลการพยากรณ์ด้วยเทคนิคต่าง ๆ

ตารางที่ 2 ตารางเปรียบเทียบผลการพยากรณ์ด้วยเทคนิคต่าง ๆ

real	MAV	KNN	MLP	ANFIS
68.758	70.191	69.560	69.120	73.569
69.475	72.102	71.256	69.549	75.695
68.520	72.46	71.598	69.156	73.226
69.117	71.505	71.505	69.223	65.235
70.669	68.520	67.541	69.895	64.654
68.520	67.684	69.415	69.056	65.235
71.505	67.564	69.248	70.489	75.653
70.430	69.117	72.125	69.589	67.233
68.878	69.594	70.950	68.958	72.365
68.639	65.14311	69.235	68.769	73.596

6. อภิปรายผลและสรุปผลการศึกษา

ในการพยากรณ์ข้อมูลอนุกรมเวลา ข้อมูลที่ถูกนำมาใช้ในการวิเคราะห์ข้อมูลอนุกรมเวลา เพื่อให้การพยากรณ์มีความถูกต้อง และเกิดประโยชน์ในการวางแผน ด้านเวลา ด้านทรัพยากรหรือบรรเทาความเสียหายที่อาจเกิดขึ้นได้ในอนาคต ข้อมูลที่มีความสมบูรณ์จะยิ่งส่งผลต่อการพยากรณ์มากขึ้น ซึ่งงานวิจัยนี้ นำเสนอขั้นตอนวิธีการสร้างรูปแบบการพยากรณ์

ข้อมูลอนุกรมเวลา โดยเน้นหารูปแบบการพยากรณ์ที่เหมาะสมสำหรับใช้สอนตัวแบบการพยากรณ์ ด้วยการเปรียบเทียบ 4 เทคนิค คือ เทคนิค ค่าเฉลี่ยเคลื่อนที่ (Moving Average) เพื่อนบ้านใกล้เคียง (KNN) ระบบผสมโครงข่ายประสาทเทียมและตรรกคลุมเครือ (Fuzzy Logic Systems) และโครงข่ายประสาทเทียม (Artificial Neural Network) ซึ่งในการวิจัยครั้งนี้ นำไปประยุกต์ใช้กับข้อมูลอนุกรมเวลาปริมาณน้ำของกรมชลประทานที่ 15 โครงการพระราชดำริลุ่มน้ำปากพนัง ในช่วงระยะเวลา 7 ปี จำนวน 12 ตัวแปร ข้อมูลน้ำตัวแปรที่ 1 ถึง 11 จะให้เป็นอินพุตสำหรับพยากรณ์ตัวแปรที่ 12 วัดประสิทธิภาพของโมเดลด้วยค่ารากที่สองของค่าเฉลี่ยผิดพลาดยกกำลังสอง พบว่า โครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น (MLP) ให้ผลการทำนายในชุดทดสอบได้ดีที่สุด และเมื่อทำไปใช้ในการพยากรณ์ค่าในอนาคตของข้อมูลอนุกรมเวลาตัวแปรเป้าหมายให้ผลลัพธ์ใกล้เคียงค่าจริง สำหรับโมเดลการพยากรณ์ข้อมูลในอนาคตของปริมาณน้ำเก็บกัก ของกรมชลประทานที่ 15 โครงการพระราชดำริลุ่มน้ำปากพนัง ยังซึ่งจะช่วยในการวางแผน ป้องกัน

และบรรเทา พื้นที่การเกษตรและการอุปโภคบริโภค
ของประชาชนในพื้นที่ดังกล่าว

7. เอกสารอ้างอิง

- [1] M. Albayrak, K. Turhan and B. Kurt, "A missing data imputation approach using clustering and maximum likelihood estimation," in *Medical Technologies National Congress (TIPTEKNO)*, Turkey, 2017.
- [2] P. Jangyodsuk, D.-J. Seo, R. Elmasr and J. Gao, "Flood Prediction and Mining Influential Spatial Features on Future Flood with Causal Discovery," in *IEEE 1^{5th} International Conference on Data Mining Workshops (ICDMW)*, NEW YORK, 2015.
- [3] S. A. Rahman, Y. Huang, J. Claassen and S. Kleinberg, "Imputation of Missing Values in Time Series with Lagged Correlations," in *4 IEEE International Conference on Data Mining Workshop*, USA, 2014.
- [4] M. Khairalla, X. Ning and N. AL-Jallad, "Modelling and optimisation of effective hybridisation model for time-series," *The Journal of Engineering*, vol. 2018, no. 2, p. 117-122, 2017.
- [5] I. R. Widiyari and L. E. Nugroho, "Deep learning multilayer perceptron (MLP) for flood prediction model using wireless sensor network based hydrology time series data mining," in *International Conference on Innovative and Creative Information Technology (ICITech)*, Indonesia, 2017.
- [6] A. Sangsongfa and P. Meesud, "Interval Type-2 Fuzzy Systems Trained by Hybrid Heuristic Algorithm for Stock Exchange of Thailand," *Information Technology Journal King Mongkut's University of Technology North Bangkok*, vol. 11, no. 2, pp. 49-55, 2015. (in Thai)
- [7] P. Meesud, *Fuzzy Systems and Neural Network*, Bangkok: King Mongkut's University of Technology North Bangkok, 2007. (in Thai)
- [8] Z. Liu, W. Zhang, T. Q. Quek and S. Lin, "Deep Fusion of Heterogeneous Sensor Data," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, USA, 2017.
- [9] C. M. Ennett, M. Frize and C. R. Walker, "Imputation of missing values by integrating neural networks and case-based reasoning," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Canada, 2008.
- [10] K. Rothjanawan and P. Meesud, "The Imputation Many Missing Value in Time Series Data Use Multivariate Relationships," *Journal of Information Science and Technology (JIST)*, vol. 8, no. 1, pp. 16-26, 2018. (in Thai)