รูปภาพประกอบด้วย ศิลปะ, เครื่องเซรามิก, ออกแบบ

คำอธิบายที่สร้างโดยอัตโนมัติ

**แบบจำลองการพยากรณ์ปริมาณน้ำฝนด้วยเทคนิคการเรียนรู้แบบเครื่อง กรณีศึกษา : จังหวัดนครพนม**

**นาย กฤษณกร เทพพิทักษ์**

**งานวิจัยนี้สำหรับการศึกษาตามหลักสูตรปริญญา**

**วิศวกรรมศาสตร์บัณฑิต สาขาวิศวกรรมคอมพิวเตอร์**

**มหาวิทยาลัยนครพนม**

**ปีการศึกษา 2567**

**แบบจำลองการพยากรณ์ปริมาณน้ำฝนด้วยเทคนิคการเรียนรู้แบบเครื่อง**

**กรณีศึกษา : จังหวัดนครพนม**

**นาย กฤษณกร เทพพิทักษ์**

**งานวิจัยนี้สำหรับการศึกษาตามหลักสูตรปริญญา**

**วิศวกรรมศาสตร์บัณฑิต สาขาวิศวกรรมคอมพิวเตอร์**

**มหาวิทยาลัยนครพนม**

**ปีการศึกษา 2567**

**Rainfall Prediction Model Using Machine Learning Techniques: A Case Study of Nakhon Phanom Province**

**Kritsanakorn Tappitak**

**This research is conducted as part of the requirements for the Bachelor of Engineering program  
in Computer Engineering**

**Nakhon Phanom University**

**2024**

ชื่อโครงงาน แบบจำลองการพยากรณ์ปริมาณน้ำฝนด้วยเทคนิคการเรียนรู้แบบเครื่อง กรณีศึกษา : จังหวัดนครพนม

โดย นาย กฤษณกร เทพพิทักษ์

สาขาวิชา วิศวกรรมคอมพิวเตอร์

อาจารย์ที่ปรึกษา ผศ.ดร.ชาญวิช สุวรรณพงศ์

สาขาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยนครพนม อนุมัติให้นับโครงงานนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิศวกรมมศาสตร์บัณฑิต

…………………………………………………………………

( )

หัวหน้าสาขา..................................................

คณะกรรมการสอบโครงงาน

................................................................ประธานกรรมการ

( )

..............................................................................กรรมการ

( )

..............................................................................กรรมการ

( )

...................................................................อาจารย์ที่ปรึกษา

( )

ลิขสิทธิ์ของสาขาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยนครพนม

ชื่อโครงงาน แบบจำลองการพยากรณ์ปริมาณน้ำฝนด้วยเทคนิคการเรียนรู้แบบเครื่อง กรณีศึกษา : จังหวัดนครพนม

โดย นาย กฤษณกร เทพพิทักษ์

สาขาวิชา วิศวกรรมคอมพิวเตอร์

อาจารย์ที่ปรึกษา ผศ.ดร.ชาญวิช สุวรรณพงศ์

ปีการศึกษา 2567

# **บทคัดย่อ**

งานวิจัยนี้มุ่งเน้นการพัฒนาและเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์ปริมาณน้ำฝนโดยใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) กรณีศึกษาจังหวัดนครพนม โดยเปรียบเทียบแบบจำลอง Long Short-Term Memory (LSTM) และ Support Vector Regression (SVR) วัตถุประสงค์หลักคือการระบุแบบจำลองที่ให้ความแม่นยำสูงสุดในการพยากรณ์ปริมาณน้ำฝนรายวัน

การวิจัยใช้ข้อมูลอุตุนิยมวิทยาย้อนหลัง 2 ปี 8 เดือน (มกราคม 2565 - กันยายน 2567) จากกรมอุตุนิยมวิทยาจังหวัดนครพนม ประกอบด้วยตัวแปรสำคัญ 10 ตัว ได้แก่ ความกดอากาศ ต่ำสุด สูงสุด เฉลี่ย8เวลา อุณหภูมิ ต่ำสุด สูงสุด เฉลี่ย ความชื้นสัมพัทธ์ ต่ำสุด สูงสุด เฉลี่ย และ น้ำระเหย โดยพัฒนาแบบจำลองผ่านกระบวนการทำความสะอาดข้อมูล การปรับขนาดข้อมูลด้วยวิธี Min-Max Scaling และประเมินประสิทธิภาพด้วยวิธี 5-Fold Cross Validation และ Confusion Matrix

ผลการวิจัยพบว่า แบบจำลอง LSTM ให้ประสิทธิภาพโดยรวมสูงกว่า SVR ในทุกมิติการวัดผล โดย LSTM มีค่าความถูกต้อง (Accuracy) ที่ 76.28% ในการทดสอบแบบ Cross Validation เทียบกับ SVR ที่ 75.32% เมื่อพิจารณาความสามารถในการจำแนกระดับปริมาณฝนตามเกณฑ์ของกรมอุตุนิยมวิทยา LSTM ให้ค่าความถูกต้อง 37.44% ขณะที่ SVR มีค่าความถูกต้อง 30.61% นอกจากนี้ LSTM ยังมีค่าความคลาดเคลื่อนเฉลี่ยสัมบูรณ์ (MAE) และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ต่ำกว่า SVR อย่างมีนัยสำคัญ

อย่างไรก็ตาม ประสิทธิภาพการพยากรณ์ของทั้งสองแบบจำลองยังมีข้อจำกัด โดยพบว่าค่าสัมประสิทธิ์การตัดสินใจ (R²) ในชุดข้อมูลทดสอบมีค่าเป็นลบ (-0.0832 สำหรับ LSTM และ -1.9577 สำหรับ SVR) ซึ่งสะท้อนให้เห็นถึงความท้าทายในการพยากรณ์ปริมาณน้ำฝนที่มีความผันผวนสูงและมีข้อมูลไม่สมดุล (67.09% ของข้อมูลในชุดฝึกสอนมีค่าฝนเป็นศูนย์) งานวิจัยนี้ชี้ให้เห็นถึงความสำคัญของการพัฒนาแบบจำลองที่สามารถจัดการกับข้อมูลที่มีความไม่สมดุลและมีความผันผวนสูง เพื่อเพิ่มประสิทธิภาพในการพยากรณ์ปริมาณน้ำฝนในอนาคต

**คำสำคัญ:** การเรียนรู้ของเครื่อง, หน่วยความจำระยะสั้นแบบยาว, การถดถอยแบบเวกเตอร์สนับสนุน, การพยากรณ์ปริมาณน้ำฝน, K-Fold Cross Validation, Confusion Matrix

Project Name Rainfall Prediction Model Using Machine Learning Techniques: A Case Study of Nakhon Phanom Province

By Kritsanakorn Tappitak

Major Field Computer Engineering

Advisor Asst. Prof. Dr. Chanwit Suwannapong

Academic Year 2024

# **Abstract**

This research focuses on developing and comparing the efficiency of rainfall prediction models using Machine Learning techniques in the case study of Nakhon Phanom Province. The study compares Long Short-Term Memory (LSTM) and Support Vector Regression (SVR) models, with the main objective of identifying the model that provides the highest accuracy in daily rainfall prediction.

The research utilized historical meteorological data spanning 2 years and 8 months (January 2022 – September 2024) from the Nakhon Phanom Provincial Meteorological Department. The dataset included 10 key variables: minimum, maximum, and 8-time average atmospheric pressure; minimum, maximum, and average temperature; minimum, maximum, and average relative humidity; and evaporation. The model was developed through a process involving data cleaning, data normalization using the Min-Max Scaling method, and performance evaluation using 5-Fold Cross Validation and a Confusion Matrix.Results showed that the LSTM model outperformed SVR in all measurement dimensions, with LSTM achieving an accuracy of 76.28% in Cross Validation testing compared to 75.32% for SVR. When considering the ability to classify rainfall levels according to Meteorological Department criteria, LSTM provided an accuracy of 37.44%, while SVR had an accuracy of 30.61%. Additionally, LSTM had significantly lower Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) than SVR.

However, the forecasting performance of both models still has limitations. The coefficient of determination (R²) in the test dataset was negative (-0.0832 for LSTM and -1.9577 for SVR), reflecting the challenges in predicting rainfall with high volatility and imbalanced data (67.09% of the training data had zero rainfall values). This research highlights the importance of developing models that can handle imbalanced data and high volatility to increase the efficiency of rainfall prediction in the future.

**Keywords:** Machine Learning, Long Short-Term Memory, Support Vector Regression, Rainfall Prediction, K-Fold Cross Validation, Confusion Matrix

**กิตติกรรมประกาศ**

การศึกษาวิจัยครั้งนี้ สำเร็จลุล่วงไปได้ด้วยความช่วยเหลือจากหลายฝ่าย ข้าพเจ้าขอขอบพระคุณคณะวิศวกรรมศาสตร์ มหาวิทยาลัยนครพนม ที่ให้โอกาสในการทำวิจัยและสนับสนุนการพัฒนาแบบจำลองการพยากรณ์ระดับน้ำด้วยเทคนิคเรียนรู้แบบเครื่อง ขอบคุณอาจารย์ที่ปรึกษา ที่ให้คำแนะนำตลอดจนให้ข้อเสนอแนะต่างๆ เพื่อให้ผลงานวิจัยมีความสมบูรณ์

สุดท้ายนี้ ข้าพเจ้าขอขอบคุณอาจารย์ที่ปรึกษา ครอบครัวและเพื่อน ที่ีคอยสนับสนุนและให้กำลังใจในการทำวิจัยนี้ข้าพเจ้าหวังเป็นอย่างยิ่ง ว่าผลงานชิ้นนี้จะเป็นประโยชน์ต่อการพัฒนาและเป็นแนวทางในการศึกษาในอนาคต

นาย กฤษณกร เทพพิทักษ์

# **สารบัญ**

หน้า

[บทคัดย่อ ข](#_Toc198002870)

[Abstract ง](#_Toc198002871)

[สารบัญ ช](#_Toc198002872)

[สารบัญตาราง ญ](#_Toc198002873)

[สารบัญภาพ ฎ](#_Toc198002874)

[บทที่ 1 1](#_Toc198002875)

[บทนำ 1](#_Toc198002876)

[1.1 ที่มาและความสำคัญของปัญหา 1](#_Toc198002877)

[1.2 วัตถุประสงค์ 2](#_Toc198002878)

[1.3 ขอบเขตของการทำโครงงาน 2](#_Toc198002879)

[1.4 นิยามศัพท์เฉพาะ 2](#_Toc198002880)

[1.5 ประโยชน์ที่คาดว่าจะได้รับ 4](#_Toc198002881)

[บทที่ 2 5](#_Toc198002882)

[ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 5](#_Toc198002883)

[2.1 การใช้แบบจำลองทางคณิตศาสตร์และเทคโนโลยีการจำลองข้อมูลในการพยากรณ์น้ำ 5](#_Toc198002884)

[2.1.1 การวิเคราะห์ข้อมูลและการพยากรณ์ (MATLAB) 5](#_Toc198002885)

[2.2 การพยากรณ์ด้วยเทคนิค Machine Learning 6](#_Toc198002886)

[2.2.1 แบบจำลองหน่วยความจำระยะสั้นแบบยาว (Long Short-Term Memory: LSTM) 6](#_Toc198002887)

[2.2.2 Support Vector Regression (SVR) 8](#_Toc198002888)

[2.3 การประเมินประสิทธิภาพตัวแบบ 10](#_Toc198002889)

[2.3.1 การประเมินประสิทธิภาพ K-Fold Cross Validation 10](#_Toc198002890)

[2.3.2 การประเมินประสิทธิภาพ Confusion Matrix 11](#_Toc198002891)

[2.4 งานวิจัยที่เกี่ยวข้อง 12](#_Toc198002892)

[บทที่ 3 16](#_Toc198002893)

[วิธีดำเนินการวิจัย 16](#_Toc198002894)

[3.1 การรวบรวมข้อมูล และเตรียมข้อมูล 18](#_Toc198002895)

[3.1.1 แหล่งที่มาและลักษณะของข้อมูล 18](#_Toc198002896)

[3.1.2 การทำความสะอาดข้อมูล (Data Cleaning) 20](#_Toc198002897)

[3.2 การพัฒนาแบบจำลองเพื่อการพยากรณ์ 23](#_Toc198002898)

[3.2.1 แบบจำลอง Long Short-Term Memory (LSTM) 24](#_Toc198002899)

[3.3 การทดสอบโมเดล 27](#_Toc198002900)

[3.3.1 วิธี K-Fold Cross Validation 27](#_Toc198002901)

[3.3.2 วิธี Confusion Matrix 27](#_Toc198002902)

[3.3.3 การวิเคราะห์ค่าความคลาดเคลื่อนของแบบจำลอง 28](#_Toc198002903)

[บทที่ 4 30](#_Toc198002904)

[ผลการวิจัย 30](#_Toc198002905)

[4.1 ผลลัพธ์จาก K-Fold Cross Validation 30](#_Toc198002906)

[4.1.1 ผลลัพธ์ 5-Fold Cross Validation 30](#_Toc198002907)

[4.2 ผลลัพธ์จาก Confusion Matrix 31](#_Toc198002908)

[4.2.1 ผลลัพธ์ Confusion Matrix ของ LSTM 31](#_Toc198002909)

[4.2.2 ผลลัพธ์ Confusion Matrix ของ SVR 32](#_Toc198002910)

[4.3 ผลการวิเคราะห์ค่าความคลาดเคลื่อนของโมเดล 33](#_Toc198002911)

[4.3.1 พารามิเตอร์ของโมเดล 34](#_Toc198002912)

[4.3.2 ผลลัพธ์ค่าความคลาดเคลื่อน 34](#_Toc198002913)

[4.3.3 การวิเคราะห์รูปแบบความคลาดเคลื่อน 35](#_Toc198002914)

[4.4.1 ด้านความแม่นยำในการพยากรณ์ 35](#_Toc198002915)

[4.4.2 ด้านความเสถียรในการพยากรณ์ 36](#_Toc198002916)

[4.4.3 ด้านความสามารถในการจับความสัมพันธ์ในข้อมูล 36](#_Toc198002917)

[บทที่ 5 38](#_Toc198002918)

[สรุปผล อภิปรายผล และข้อเสนอแนะ 38](#_Toc198002919)

[5.1 สรุปผลการวิจัย 38](#_Toc198002920)

[5.1.1 การพัฒนาแบบจำลองการพยากรณ์ปริมาณน้ำฝนที่มีความแม่นยำ 38](#_Toc198002921)

[5.1.2 การเปรียบเทียบประสิทธิภาพของแบบจำลอง LSTM และ SVR 39](#_Toc198002922)

[5.1.3 การวิเคราะห์ปัจจัยที่มีอิทธิพลต่อปริมาณน้ำฝนและข้อจำกัดของแบบจำลอง 39](#_Toc198002923)

[5.2 อภิปรายผล 40](#_Toc198002924)

[5.3 ข้อเสนอแนะ 40](#_Toc198002925)

[5.3.1 ข้อเสนอแนะสำหรับการนำผลการวิจัยไปใช้ประโยชน์ 41](#_Toc198002926)

[5.3.2 ข้อเสนอแนะสำหรับการวิจัยในอนาคต 41](#_Toc198002927)

[บรรณานุกรม 43](#_Toc198002928)

# **สารบัญตาราง**

หน้า

ตารางที่ 1 ตารางสรุปการแบ่งข้อมูลโดยละเอียด ……………………………………………………………….……23

ตารางที่ 2 ผลลัพธ์ของการทดสอบ 5 Fold Cross Validation ………………………………….................30

ตารางที่ 3 ตารางเปรียบเทียบประสิทธิภาพของโมเดล LSTM และ SVR ……………………………….…..31

ตารางที่ 4 ผลลัพธ์ LSTM Confusion Matrix ………………………………………………………………….…....31

ตารางที่ 5 ผลลัพธ์ SVR Confusion Matrix …………………………………………………………………….….…32

ตารางที่ 6 ตารางแสดงพารามิเตอร์โมเดล LSTM …………………………………………………………….………34

ตารางที่ 7 ตารางแสดงพารามิเตอร์โมเดล SVR …………………………………………………………….…………34

ตารางที่ 8 ผลลัพธ์ค่าความคลาดเคลื่อนของชุดข้อมูลทดสอบ ……………………………………….………….34

# **สารบัญภาพ**

หน้า

รูปที่ 1 แบบจำลองหน่วยความจำระยะสั้นแบบยาว (LSTM) …………………………………………………..…..8

รูปที่ 2 โครงสร้างของ Support Vector Regression (SVR) …………………………………………………..…10

รูปที่ 3 K-fold cross-validation ………………………………………………………………………………………….10

รูปที่ 4 ตาราง Confusion Matrix ……………………………………………………………………........................11

รูปที่ 5 ขั้นตอนการดำเนินงาน ……………………………………………………………………..............................17

รูปที่ 6 ตัวอย่างชุดข้อมูลที่ได้รับจากกรมอุตุนิยมวิทยาจังหวัดนครพนม ………………………………….....19

รูปที่ 7 ตัวอย่างข้อมูลที่ได้ทำความสะอาดข้อมูล ………………………………………………………………….….21

รูปที่ 8 ตัวอย่างข้อมูลที่ผ่านการทำ Standardization …………………………………………………………..…22

รูปที่ 9 กราฟแสดงสัดส่วนปริมาณน้ำฝน ………………………………………………………………………….…….23

รูปที่ 10 เกณฑ์การตรวจวัดฝน ปริมาณฝนรายวัน ……………………………………………………………….…..27

# **บทที่ 1**

# **บทนำ**

## **1.1 ที่มาและความสำคัญของปัญหา**

ปริมาณน้ำฝนเป็นปัจจัยสำคัญที่ส่งผลกระทบโดยตรงต่อ ภาคการเกษตร ระบบนิเวศ และการบริหารจัดการทรัพยากรน้ำ ของประเทศ โดยเฉพาะใน จังหวัดนครพนม ซึ่งตั้งอยู่ทางภาคตะวันออกเฉียงเหนือของประเทศไทย มีลักษณะภูมิอากาศแบบมรสุม ส่งผลให้มีปริมาณฝนตกชุกในช่วงฤดูฝน และในบางช่วงอาจเกิดภาวะฝนแล้งหรือฝนทิ้งช่วง การพยากรณ์ปริมาณน้ำฝนที่แม่นยำจึงมีความสำคัญอย่างยิ่งต่อการวางแผนด้านการเกษตร การบริหารจัดการน้ำ และการเตรียมความพร้อมรับมือกับภัยพิบัติทางธรรมชาติ เช่น น้ำท่วมและภัยแล้ง

ในอดีต การพยากรณ์ปริมาณน้ำฝนมักอาศัยแบบจำลองทางคณิตศาสตร์และสถิติ เช่น แบบจำลอง ARIMA และแบบจำลองเชิงเส้น (Linear Regression) อย่างไรก็ตาม แบบจำลองเหล่านี้อาจไม่สามารถจับความซับซ้อนของข้อมูลที่มีลักษณะเป็นอนุกรมเวลา (Time Series) ได้อย่างแม่นยำ เนื่องจากปริมาณฝนขึ้นอยู่กับปัจจัยทางอุตุนิยมวิทยาหลายประการ เช่น อุณหภูมิ ความชื้นสัมพัทธ์ ความกดอากาศ ความเร็วลม และลักษณะของฤดูกาล ซึ่งมีความสัมพันธ์กันอย่างซับซ้อน

ปัจจุบัน เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) และปัญญาประดิษฐ์ (Artificial Intelligence - AI) ได้รับความสนใจมากขึ้นในการพยากรณ์สภาพอากาศ โดยเฉพาะ โครงข่ายประสาทเทียมแบบลึก (Deep Learning) เช่น Long Short-Term Memory (LSTM) ซึ่งเป็นแบบจำลองที่ถูกออกแบบมาให้สามารถเรียนรู้ข้อมูลเชิงเวลาและแนวโน้มของปริมาณฝนตกได้อย่างมีประสิทธิภาพ ส่งผลให้ การพยากรณ์ฝนมีความแม่นยำมากกว่าวิธีการแบบดั้งเดิม

โครงการนี้มีเป้าหมายในการ พัฒนาแบบจำลองการพยากรณ์ปริมาณน้ำฝนโดยใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) และเปรียบเทียบประสิทธิภาพของแบบจำลองสองรูปแบบ ได้แก่ Long Short-Term Memory (LSTM) ซึ่งสามารถเรียนรู้แนวโน้มของข้อมูลอนุกรมเวลาและจับรูปแบบการเปลี่ยนแปลงของปริมาณฝนตกได้ดี Support Vector Regression (SVR) ซึ่งเป็นโมเดลที่สามารถวิเคราะห์ความสัมพันธ์เชิงซับซ้อนระหว่างปริมาณฝนกับปัจจัยทางอุตุนิยมวิทยา

การศึกษานี้จะช่วยให้สามารถ ประเมินและเปรียบเทียบประสิทธิภาพของทั้งสองโมเดล เพื่อพิจารณาว่าเทคนิคใดมีความเหมาะสมมากที่สุดสำหรับการพยากรณ์ฝนตกในพื้นที่ศึกษา ซึ่งจะเป็นประโยชน์ต่อหน่วยงานที่เกี่ยวข้องและผู้ที่ต้องพึ่งพาข้อมูลปริมาณน้ำฝนในการตัดสินใจ เช่น เกษตรกร นักวิทยาศาสตร์ด้านสภาพอากาศ หน่วยงานป้องกันภัยพิบัติ และนักวางแผนนโยบายด้านทรัพยากรน้ำ

## 

## **1.2 วัตถุประสงค์**

1.2.1 เพื่อพัฒนาแบบจำลองการพยากรณ์ปริมาณน้ำฝนรายวันสำหรับจังหวัดนครพนมที่มีความแม่นยำไม่ต่ำกว่าร้อยละ 75 โดยใช้เทคนิคการเรียนรู้ของเครื่อง

1.2.2 เพื่อเปรียบเทียบประสิทธิภาพและความแม่นยำของแบบจำลอง Long Short-Term Memory (LSTM) และ Support Vector Regression (SVR) ในการพยากรณ์ปริมาณน้ำฝนรายวัน

## **1.3 ขอบเขตของการทำโครงงาน**

1.3.1 ข้อมูลที่ใช้ในการวิจัย

1) ใช้ข้อมูลปริมาณน้ำฝนและตัวแปรทางอุตุนิยมวิทยารายวันที่เก็บรวบรวมจากกรมอุตุนิยมวิทยา จังหวัดนครพนม ย้อนหลัง 2 ปี 8 เดือน (ตั้งแต่ 1 มกราคม 2565 จนถึง 7 กันยายน 2567)

2) ตัวแปรที่ใช้ในการวิเคราะห์ประกอบด้วย ความกดอากาศ ต่ำสุด สูงสุด เฉลี่ย8เวลา อุณหภูมิ ต่ำสุด สูงสุด เฉลี่ย ความชื้นสัมพัทธ์ ต่ำสุด สูงสุด เฉลี่ย และ น้ำระเหย

1.3.2 แบบจำลองที่ใช้ในการวิจัย

1) พัฒนาและเปรียบเทียบแบบจำลอง 2 ประเภท ได้แก่ Long Short-Term Memory (LSTM) และ Support Vector Regression (SVR)

2) ใช้โปรแกรม MATLAB Version 23.2.0.2485118 (R2023b) ในการพัฒนาแบบจำลอง

1.3.3 การประเมินประสิทธิภาพแบบจำลอง

1) ใช้วิธี K-Fold Cross Validation โดยกำหนด K = 5 เพื่อประเมินความแม่นยำของแบบจำลอง

2) วัดประสิทธิภาพด้วยตัวชี้วัด Accuracy, Precision, Recall, F1-Score, Mean Absolute Error (MAE), Root Mean Square Error (RMSE) และ Coefficient of Determination (R²)

3) ประเมินความสามารถในการจำแนกระดับปริมาณฝนตามเกณฑ์ของกรมอุตุนิยมวิทยา

## **1.4 นิยามศัพท์เฉพาะ**

1. การเรียนรู้ของเครื่อง (Machine Learning) หมายถึง เทคนิคทางคอมพิวเตอร์ที่ช่วยให้ระบบสามารถเรียนรู้และปรับปรุงจากประสบการณ์โดยอัตโนมัติ โดยไม่ต้องเขียนโปรแกรมโดยตรง ในงานวิจัยนี้หมายถึงกระบวนการที่ให้คอมพิวเตอร์เรียนรู้รูปแบบความสัมพันธ์ระหว่างปัจจัยทางอุตุนิยมวิทยาและปริมาณน้ำฝน เพื่อนำไปใช้ในการพยากรณ์

2. หน่วยความจำระยะสั้นแบบยาว (Long Short-Term Memory; LSTM) หมายถึง รูปแบบของโครงข่ายประสาทเทียมแบบย้อนกลับ (Recurrent Neural Network) ที่สามารถเรียนรู้ความสัมพันธ์ระยะยาวในข้อมูลอนุกรมเวลา โดยมีโครงสร้างพิเศษที่ช่วยจดจำและลืมข้อมูลอย่างเหมาะสม ทำให้สามารถวิเคราะห์ข้อมูลที่มีความสัมพันธ์ต่อเนื่องในระยะเวลานานได้

3. การถดถอยแบบเวกเตอร์สนับสนุน (Support Vector Regression; SVR) หมายถึง เทคนิคการเรียนรู้ของเครื่องที่พัฒนามาจาก Support Vector Machine สำหรับปัญหาการถดถอย (Regression) โดยมีหลักการคือการหาฟังก์ชันที่สามารถทำนายค่าต่อเนื่องได้ ด้วยการสร้างช่องว่าง (ε-tube) รอบฟังก์ชันที่ให้ค่าความคลาดเคลื่อนน้อยที่สุด

4. การตรวจสอบแบบไขว้ K ส่วน (K-Fold Cross Validation) หมายถึง วิธีการประเมินประสิทธิภาพของแบบจำลองโดยการแบ่งข้อมูลออกเป็น K ส่วนเท่าๆ กัน จากนั้นทำการฝึกและทดสอบแบบจำลอง K รอบ โดยแต่ละรอบจะใช้ข้อมูล K-1 ส่วนในการฝึก และใช้ข้อมูล 1 ส่วนที่เหลือในการทดสอบ จากนั้นนำผลลัพธ์ทั้ง K รอบมาเฉลี่ยกัน เพื่อให้ได้ผลการประเมินที่น่าเชื่อถือ

5. ตารางความสับสน (Confusion Matrix) หมายถึง เครื่องมือที่ใช้วัดประสิทธิภาพของแบบจำลองในการจำแนกประเภท (Classification) โดยแสดงความสัมพันธ์ระหว่างค่าจริงและค่าที่ทำนายได้ในรูปแบบตาราง ช่วยให้คำนวณค่าวัดประสิทธิภาพต่างๆ เช่น Accuracy, Precision, Recall และ F1-Score

6. การปรับขนาดแบบ Min-Max (Min-Max Scaling) หมายถึง เทคนิคการปรับขนาดข้อมูลให้อยู่ในช่วงที่กำหนด (โดยทั่วไปคือ [0, 1]) เพื่อให้ตัวแปรทุกตัวมีระดับสเกลที่เทียบเคียงกันได้ ช่วยเพิ่มประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่อง

7. ความถูกต้อง (Accuracy) หมายถึง สัดส่วนของผลการทำนายที่ถูกต้องต่อผลการทำนายทั้งหมด ในงานวิจัยนี้คือร้อยละของการทำนายระดับปริมาณน้ำฝนได้ถูกต้องตามเกณฑ์ของกรมอุตุนิยมวิทยา

8. ความแม่นยำ (Precision) หมายถึง สัดส่วนของผลการทำนายเชิงบวกที่ถูกต้องต่อผลการทำนายเชิงบวกทั้งหมด ในงานวิจัยนี้คือความสามารถในการทำนายว่ามีฝนตกได้อย่างถูกต้อง

9. ความครบถ้วน (Recall) หมายถึง สัดส่วนของผลการทำนายเชิงบวกที่ถูกต้องต่อกรณีที่เป็นเชิงบวกทั้งหมด ในงานวิจัยนี้คือความสามารถในการตรวจจับวันที่มีฝนตกได้ทั้งหมด

10. ค่า F1 (F1-Score) หมายถึง ค่าเฉลี่ยฮาร์โมนิกของ Precision และ Recall ใช้เป็นตัวชี้วัดประสิทธิภาพของแบบจำลองที่คำนึงถึงทั้งความแม่นยำและความครบถ้วน

11. ค่าความคลาดเคลื่อนเฉลี่ยสัมบูรณ์ (Mean Absolute Error; MAE) หมายถึง ค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ระหว่างค่าที่ทำนายได้กับค่าจริง ค่า MAE ที่ต่ำแสดงถึงความแม่นยำที่สูงของแบบจำลอง

12. รากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error; RMSE) หมายถึง รากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองระหว่างค่าที่ทำนายได้กับค่าจริง ให้น้ำหนักกับความคลาดเคลื่อนขนาดใหญ่มากกว่า MAE

13. สัมประสิทธิ์การตัดสินใจ (Coefficient of Determination; R²) หมายถึง ตัวชี้วัดที่แสดงว่าแบบจำลองสามารถอธิบายความแปรปรวนของข้อมูลได้ดีเพียงใด มีค่าตั้งแต่ -∞ ถึง 1 โดยค่า 1 หมายถึงแบบจำลองสามารถทำนายได้อย่างสมบูรณ์แบบ

14. ชุดข้อมูลฝึกสอน (Training Dataset) หมายถึง ชุดข้อมูลที่ใช้ในการฝึกฝนแบบจำลองให้เรียนรู้ความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตาม ในงานวิจัยนี้คือข้อมูลอุตุนิยมวิทยาและปริมาณน้ำฝนตั้งแต่วันที่ 1 มกราคม 2565 ถึง 23 กุมภาพันธ์ 2567

15. ชุดข้อมูลทดสอบ (Test Dataset) หมายถึง ชุดข้อมูลที่ใช้ในการประเมินประสิทธิภาพของแบบจำลองหลังจากผ่านการฝึกฝนแล้ว ในงานวิจัยนี้คือข้อมูลอุตุนิยมวิทยาและปริมาณน้ำฝนตั้งแต่วันที่ 24 กุมภาพันธ์ 2567 ถึง 7 กันยายน 2567

## **1.5 ประโยชน์ที่คาดว่าจะได้รับ**

1.5.1 ได้แบบจำลองการพยากรณ์ปริมาณน้ำฝนที่มีความแม่นยำสูงสำหรับจังหวัดนครพนม ซึ่งสามารถนำไปใช้ในการวางแผนการเกษตร การบริหารจัดการน้ำ และการเตือนภัยน้ำท่วมหรือภัยแล้ง

1.5.2 ทราบถึงประสิทธิภาพและข้อจำกัดของแบบจำลอง LSTM และ SVR ในการพยากรณ์ปริมาณน้ำฝน รวมถึงปัจจัยที่ส่งผลต่อความแม่นยำของการพยากรณ์ในบริบทของจังหวัดนครพนม

1.5.3 สามารถนำองค์ความรู้และวิธีการที่ได้ไปประยุกต์ใช้ในการพัฒนาแบบจำลองการพยากรณ์สภาพอากาศในพื้นที่อื่นๆ ที่มีลักษณะภูมิอากาศคล้ายคลึงกัน

1.5.4 สร้างพื้นฐานสำหรับการพัฒนาระบบเตือนภัยล่วงหน้าด้านอุทกภัยและภัยแล้งในพื้นที่จังหวัดนครพนม ซึ่งจะช่วยลดความเสียหายทางเศรษฐกิจและสังคมจากภัยพิบัติทางธรรมชาติ

# **บทที่ 2**

# **ทฤษฎีและงานวิจัยที่เกี่ยวข้อง**

## **2.1 การใช้แบบจำลองทางคณิตศาสตร์และเทคโนโลยีการจำลองข้อมูลในการพยากรณ์น้ำ**

การใช้แบบจำลองทางคณิตศาสตร์เพื่อพยากรณ์ปริมาณน้ำฝนเป็นหนึ่งในเครื่องมือที่ได้รับความนิยมในงานวิจัยด้านการบริหารจัดการน้ำและการป้องกันน้ำท่วม งานวิจัยจำนวนมากได้นำเสนอการใช้ข้อมูลสถิติ ข้อมูลภูมิอากาศ และข้อมูลน้ำฝนในอดีตในการพัฒนาแบบจำลองเพื่อคาดการณ์ปริมาณน้ำ ซึ่งช่วยให้ผู้ที่รับผิดชอบสามารถเตรียมความพร้อมและจัดการน้ำได้อย่างมีประสิทธิภาพ โปรแกรม MATLAB เป็นหนึ่งในเครื่องมือที่นักวิจัยและวิศวกรนิยมใช้ในการพัฒนาและจำลองแบบจำลองทางคณิตศาสตร์ เนื่องจากความสามารถในการคำนวณทางคณิตศาสตร์ที่ซับซ้อน การวิเคราะห์ข้อมูล และการสร้างกราฟที่แสดงผลข้อมูลได้อย่างแม่นยำ งานวิจัยต่างๆ พบว่าแบบจำลองทางคณิตศาสตร์สามารถนำมาใช้ในการคาดการณ์ปริมาณน้ำฝนและระดับน้ำในพื้นที่เสี่ยงน้ำท่วมได้อย่างมีประสิทธิภาพ โดยเฉพาะเมื่อมีการปรับแต่งแบบจำลองให้เหมาะสมกับสภาพพื้นที่และข้อมูลในพื้นที่นั้นๆ

### 2.1.1 การวิเคราะห์ข้อมูลและการพยากรณ์ (MATLAB)

MATLAB เป็นซอฟต์แวร์ที่นิยมใช้ในการวิเคราะห์ข้อมูล เนื่องจากมีเครื่องมือที่สามารถช่วยในการประมวลผลข้อมูล การสร้างกราฟสถิติ และการพยากรณ์ข้อมูลเชิงเวลาได้อย่างมีประสิทธิภาพ ในโครงงานนี้ MATLAB จะถูกนำมาใช้ในหลายส่วน เช่น การนำเข้าข้อมูล การสร้างกราฟแสดงผลข้อมูล และการพยากรณ์ข้อมูลปริมาณน้ำ ซึ่งมีฟังก์ชันการใช้งานต่างๆ ของ MATLAB ดังนี้

-การนำเข้าข้อมูล ฟังก์ชัน readtable จะถูกใช้สำหรับนำเข้าข้อมูลปริมาณน้ำจากไฟล์ เช่นไฟล์ CSV หรือ Excel เพื่อให้สามารถนำข้อมูลไปประมวลผลต่อไปได้

-การสร้างกราฟ ใช้ฟังก์ชัน plot ในการแสดงผลข้อมูลในรูปแบบกราฟเส้น (Line Graph) เพื่อให้เห็นการเปลี่ยนแปลงของปริมาณน้ำในช่วงเวลาต่างๆ อย่างชัดเจน -การพยากรณ์ MATLAB มีฟังก์ชัน fitlm ที่ใช้ในการสร้างแบบจำลองเชิงเส้น (Linear Regression Model) เพื่อพยากรณ์ปริมาณน้ำในอนาคต โดยอิงจากข้อมูลในอดีตที่เก็บรวบรวมไว้

-การวิเคราะห์ความคลาดเคลื่อน การวิเคราะห์ความแม่นยำของการพยากรณ์สามารถทำได้โดยการคำนวณค่าความคลาดเคลื่อน เพื่อประเมินความแตกต่างระหว่างค่าจริงและค่าที่พยากรณ์ได้

## **2.2 การพยากรณ์ด้วยเทคนิค Machine Learning**

การพยากรณ์ (Forecasting) คือกระบวนการที่ใช้ในการคาดการณ์ข้อมูลในอนาคต โดยการวิเคราะห์ข้อมูลในอดีต การนำ Machine Learning (ML) มาใช้ในการพยากรณ์นั้นเริ่มจากการสร้างโมเดลที่เรียนรู้จากข้อมูลตัวอย่าง ซึ่งประกอบด้วยอินพุตและผลลัพธ์ที่ทราบอยู่แล้ว กระบวนการนี้อยู่ภายใต้แนวคิดของ Supervised Learning โมเดลจะพยายามหาความสัมพันธ์ระหว่างอินพุตและผลลัพธ์เพื่อใช้ในการทำนายข้อมูลใหม่ โดยมีโมเดลการพยากรณ์ดังนี้

### 2.2.1 แบบจำลองหน่วยความจำระยะสั้นแบบยาว (Long Short-Term Memory: LSTM)

แบบจำลองหน่วยความจำระยะสั้นแบบยาว (LSTM) เป็นรูปแบบหนึ่งของโครงข่ายระบบ ประสาทแบบย้อนกลับ (RNN) ถูกพัฒนาขึ้นโดย (Hochreiter & Schmidhuber, 1997) เพื่อ แก้ปัญหาการลดลงของเกรเดียนต์ (Vanishing Gradient) เมื่อลำดับของข้อมูลที่รับเข้ามามีจำนวน มากเกินไป โดยการทำงานของ LSTM นั้นจะมีความคล้ายคลึงกับ RNN แต่ในส่วนของชั้นซ่อน (Hidden state) ที่ใช้สำหรับจดจำลำดับของข้อมูลก่อนหน้านั้น LSTM จะมีการเรียนรู้ว่า เมื่อใดควร ลืม (Forget) เขียน (Write) หรืออนุญาตให้อ่าน (Read) ซึ่งทำให้สามารถรองรับข้อมูลที่มีลำดับ ปริมาณมากที่เข้ามาได้ โดยส่วนประกอบต่างๆ ของแบบจำลองหน่วยความจำระยะสั้นแบบยาวมีดังนี้

1) สถานะเซลล์ (Cell state) ทำหน้าที่เป็นตัวเก็บสถานะของเซลล์ความจำ (Memory cell)

2) ประตู (Gate) ทำหน้าที่เป็นตัวควบคุมการไหลของข้อมูลที่เข้ามาในแบบจำลอง ซึ่งประตูแต่ ละบานจะมีฟังก์ชันกระตุ้นเป็นของตัวเองเพื่อตัดสินใจว่าควรจะจัดการกับข้อมูลที่จะเข้ามา อย่างไร โดยประตูควบคุมการทำงานนั้น จะประกอบไปด้วย

3) ประตูลืม (Forget gate) ทำหน้าที่เป็นประตูที่จะตัดสินใจว่าควรจะเก็บสถานะเซลล์เดิม เอาไว้ หรือจะลบสถานะเซลล์เดิมออกไป เป็นเสมือนการเตรียมพื้นที่เพื่อรับข้อมูลใหม่ โดยฟังก์ชันกระตุ้นที่ควบคุมการทำงานประตูนี้อยู่คือ ฟังก์ชันซิกมอยด์ (Sigmoid Function: 𝜎) ผลลัพธ์ที่จะได้ออกมาจะมีค่าอยู่ระหว่างค่า 0 ถึง 1 ถ้าผลลัพธ์ที่ได้มีค่า เข้าใกล้ 0 หมายถึง การลบสถานะเซลล์เดิมออกไป แต่ถ้าผลลัพธ์ทีได้มีค่าเข้าใกล้ 1 หมายถึง การเก็บสถานะเซลล์เดิมเอาไว้

4) ประตูเขียน (Write gate) ในส่วนของประตูเขียนนั้น จะมีการทำงาน 2 ส่วน โดยในส่วน แรก คือการตัดสินใจว่าจะปรับสถานะเซลล์เมื่อมีข้อมูลใหม่เข้ามาหรือไม่ โดยส่วนนี้จะ ถูกควบคุมด้วยประตูนำเข้า (Input gate) ซึ่งมีฟังก์ชันซิกมอยด์ทำหน้าที่เป็นฟังก์ชัน กระตุ้น เพื่อทำการตัดสินใจว่า จะให้มีการปรับเปลี่ยนสถานะเซลล์หรือไม่ ในการ คำนวณส่วนนี้จะใช้ข้อมูลนำเข้าปัจจุบันที่เข้ามาพร้อมกับชั้นซ่อน (Hidden state) ก่อน หน้า ถ้าผลลัพธ์ที่ได้มีค่าเข้าใกล้ 0 ก็จะไม่มีการปรับเปลี่ยนสถานะของเซลล์ใดๆ แต่ถ้า ผลลัพธ์ที่ได้มีค่าเข้าใกล้ 1 ก็จะทำการปรับเปลี่ยนสถานะของเซลล์ใหม่ให้เป็น สถานะปัจจุบัน

หลังจากที่ทำการตัดสินใจไปแล้วในส่วนแรกว่าจะทำการปรับเปลี่ยนค่าสถานะเซลล์ใหม่ ให้เป็นปัจจุบัน ก็จะนำไปสู่การทำงานในส่วนที่สองของประตูเขียน นั่นคือการหาค่า สถานะเซลล์ใหม่เพื่อทำการปรับเปลี่ยน โดยในส่วนนี้จะถูกควบคุมด้วยประตูปรับค่า นำเข้า (Input moderation gate) ซึ่งมีฟังก์ชันไฮเพอร์โบลิกแทนเจนต์ (Hyperbolic Tangent: tanh) หรือ ฟังก์ชันแทน ทำหน้าที่เป็นฟังก์ชันกระตุ้น โดยผลลัพธ์ที่ได้ออกมา นั้น จะเป็นเหมือนกับผู้ท้าชิงในสนาม (Cell state candidate) ก่อนที่จะถูกนำไป ปรับเปลี่ยนอีกครั้งตามสถานะที่ถูกคำนวณมาก่อนหน้าเพื่อส่งเป็นข้อมูลขาออกต่อไป

จะเห็นว่าในตอนนี้มีข้อมูลจากทั้ง ประตูลืม ประตูนำเข้า และประตูปรับค่านำเข้า ซึ่ง ทั้งหมดนี้ก็เพียงพอสำหรับการปรับค่าสถานะเซลล์ (Update cell state) แล้ว

เมื่อพิจารณาจากฝั่งขวามือของสมการจะพบว่า ถ้าประตูลืม (𝑓𝑡) มีค่าเป็น 0 ค่าสถานะ เซลล์ก่อนหน้า (𝐶𝑡−1) ก็จะไม่ถูกนำมาพิจารณาประกอบการปรับเปลี่ยนค่าสถานะเซลล์ปัจจุบัน แต่ถ้าประตูลืมมีค่าเป็น 1 ค่าสถานะเซลล์ก่อนหน้า (𝐶𝑡−1) ก็จะถูกนำมาพิจารณาประกอบการปรับเปลี่ยนค่าสถานะเซลล์ปัจจุบัน และเมื่อพิจารณาฝั่งซ้ายมือ ของสมการจะพบว่า ถ้าประตูนำเข้า (𝑖𝑡) มีค่าเป็น 0 แสดงว่าค่าที่ได้จากการคำนวณใน ประตูปรับค่านำเข้า (C~𝑡) จะไม่ถูกนำมาใช้ประกอบการพิจารณาในการปรับเปลี่ยนค่า สถานะเซลล์ปัจจุบัน แต่ถ้าประตูนำเข้า (𝑖𝑡) มีค่าเป็น 1 แสดงว่าค่าที่ได้จากการคำนวณ ในประตูปรับค่านำเข้า (C~𝑡) จะถูกนำมาใช้ประกอบการพิจารณาในการปรับเปลี่ยนค่า สถานะเซลล์ปัจจุบัน และจากค่าทั้งหมดที่ได้มาก็จะทำให้ได้ผลลัพธ์ออกมาเป็นค่า สถานะเซลล์ในปัจจุบัน (𝐶𝑡 )

5) ประตูนำออก (Output gate) ทำหน้าที่เป็นประตูที่จะตัดสินใจว่าจะทำการส่งค่าสถานะ ชั้นซ่อน (Hidden state: ℎ𝑡 ) ออกไปหรือไม่ โดยฟังก์ชันกระตุ้นที่ควบคุมการทำงาน ประตูนี้อยู่คือ ฟังก์ชันซิกมอยด์ (Sigmoid Function: σ) ในการคำนวณส่วนนี้จะใช้ ข้อมูลนำเข้าปัจจุบันที่เข้ามาพร้อมกับชั้นซ่อน (Hidden state) ก่อนหน้า

เนื่องจากผลลัพธ์ที่จะต้องผลิตออกไปคือค่าสถานะชั้นซ่อนที่เวลาปัจจุบัน(ℎ𝑡 ) ดังนั้น สมการสำหรับคำนวณค่าดังกล่าว จึงแสดงได้ดังนี้

จากสมการดังกล่าวพบว่า ถ้าประตูนำออก (𝑜𝑡) มีค่าเป็น 0 ก็จะไม่มีการส่งค่าใดๆ ออกไป แต่ถ้าประตูนำออกมีค่าเป็น 1 ก็จะคำนวณค่า ℎ𝑡 แล้วส่งออกไป โดยที่ค่า ℎ𝑡 ก็ คือผลลัพธ์สุดท้ายที่ได้จากสถานะเซลล์และสถานะชั้นซ่อน เพื่อใช้กับหน่วยความจำ ระยะสั้นแบบยาวในลำดับถัดไป ซึ่งในส่วนประตูนำออกนี้ก็จะเป็นเสมือนประตูอนุญาต ให้อ่าน (Read gate) ที่ได้กล่าวถึงไปข้างต้น ที่ทำหน้าที่ตัดสินใจว่าจะอนุญาตให้นำ ข้อมูลผลลัพธ์สุดท้ายที่ได้ออกไปหรือไม่

สำหรับโครงสร้างภายในของแบบจำลองหน่วยความจำระยะสั้นแบบยาว สามารถแสดงได้ดัง รูปที่ 1

รูปภาพประกอบด้วย แผนภาพ, ภาพหน้าจอ, ไลน์, วางแผน

คำอธิบายที่สร้างโดยอัตโนมัติ

รูปที่ 1 แบบจำลองหน่วยความจำระยะสั้นแบบยาว (LSTM)

### 2.2.2 Support Vector Regression (SVR)

SVR เป็นเทคนิคการเรียนรู้ของเครื่องที่พัฒนามาจาก Support Vector Machine (SVM) ซึ่งเดิมใช้สำหรับปัญหาการจำแนกประเภท (Classification) โดย SVR ได้รับการปรับให้เหมาะกับปัญหาการถดถอย (Regression) เพื่อทำนายค่าต่อเนื่อง เช่น ปริมาณน้ำฝน (Qiu et al., 2020)

หลักการทำงานของ SVR คือการหาเส้นโค้งที่เหมาะสมสำหรับการพยากรณ์ค่าของข้อมูล โดยพยายามให้ค่าคลาดเคลื่อนอยู่ภายในค่าที่กำหนด (ε-insensitive loss function) ซึ่งช่วยให้โมเดลมีความสามารถในการทนทานต่อค่าผิดพลาดเล็กน้อย และสามารถพยากรณ์ข้อมูลที่มีลักษณะไม่เชิงเส้นได้ดีผ่านการใช้ Kernel Trick เช่น Linear Kernel, Polynomial Kernel, และ Radial Basis Function (RBF) Kernel

SVR เหมาะสำหรับการพยากรณ์ข้อมูลที่มีความซับซ้อนและไม่เป็นเชิงเส้น เช่น การพยากรณ์ปริมาณน้ำฝน ระดับน้ำในแม่น้ำ และตัวแปรทางอุตุนิยมวิทยา ซึ่งช่วยให้ได้ค่าพยากรณ์ที่แม่นยำกว่าวิธีการเชิงเส้นแบบดั้งเดิม

สมการที่เกี่ยวข้องกับ Support Vector Regression (SVR)

Support Vector Regression (SVR) มีพื้นฐานมาจาก Support Vector Machine (SVM) ซึ่งปรับใช้สำหรับปัญหาการพยากรณ์ค่าต่อเนื่อง โดยใช้แนวคิด ε-insensitive loss function เพื่อกำหนดขอบเขตความผิดพลาดที่ยอมรับได้

1. ฟังก์ชันพื้นฐานของ SVR

SVR พยายามหาฟังก์ชันเชิงเส้นที่เหมาะสมกับข้อมูลในรูปแบบของ

(2.1)

โดยที่

คือ เวกเตอร์น้ำหนักของฟังก์ชัน

คือ ฟังก์ชันแปลงข้อมูลไปยังมิติที่สูงขึ้น (ใช้ Kernel Trick)

คือ ค่าคงที่ (bias)

1. ฟังก์ชันวัตถุประสงค์ของ SVR

SVR ใช้แนวคิดการหาค่า **w** และ **b** ที่เหมาะสม โดยลดค่าผลรวมของระยะห่างข้อมูลจากฟังก์ชันที่พยากรณ์ได้ และควบคุมค่าคลาดเคลื่อนตาม ε-insensitive loss function ดังนี้

(2.2)

โดยที่

คือ ค่าความซับซ้อนของโมเดลที่ต้องการลด

คือ ค่าพารามิเตอร์ที่ใช้ควบคุมการเทรดออฟระหว่างค่าผิดพลาดกับความซับซ้อนของโมเดล

คือ ค่าคลาดเคลื่อนของข้อมูลที่อยู่เกินช่วง **ε**

1. Kernel Trick ใน SVR

เพื่อให้ SVR สามารถพยากรณ์ข้อมูลที่ไม่เป็นเชิงเส้นได้ สามารถใช้ Kernel Trick เพื่อแปลงข้อมูลไปยังมิติที่สูงขึ้น โดยใช้ Kernel Function แทนฟังก์ชัน ดังนี้

1. Polynomial Kernel

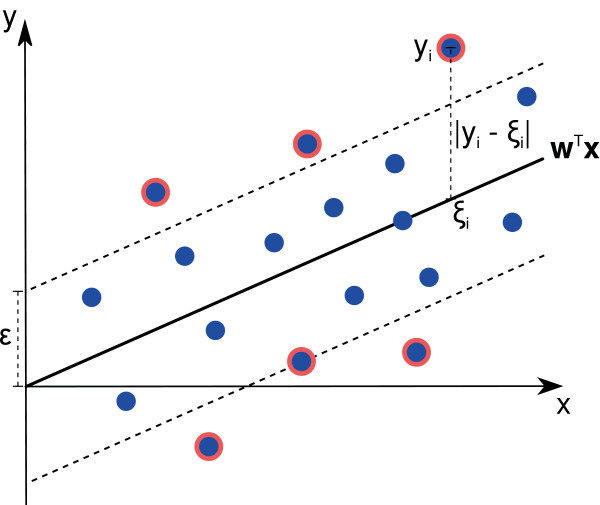
(2.4)

1. Radial Basis Function (RBF) Kernel

(2.5)

โดย Kernel Function ช่วยให้โมเดลสามารถจับรูปแบบของข้อมูลที่มีความซับซ้อนและไม่เป็นเชิงเส้นได้

สรุป SVR ใช้ฟังก์ชันพยากรณ์ (2.6) และหาค่าที่เหมาะสมโดยลดค่าความคลาดเคลื่อนที่เกินขอบเขต **ε** ผ่านฟังก์ชันวัตถุประสงค์ที่ควบคุมความสมดุลระหว่างค่าผิดพลาดและความซับซ้อนของโมเดล โดยสามารถเลือก **Kernel Trick** เพื่อให้โมเดลสามารถทำงานได้ดียิ่งขึ้นกับข้อมูลที่ไม่เป็นเชิงเส้น



รูปที่ 2 โครงสร้างของ Support Vector Regression (SVR)

## **2.3 การประเมินประสิทธิภาพตัวแบบ**

2.3.1 การประเมินประสิทธิภาพ K-Fold Cross Validation เป็นหนึ่งในเทคนิคที่ได้รับความนิยมในการประเมินประสิทธิภาพของแบบจำลอง Machine Learning โดยมีจุดมุ่งหมายเพื่อลดปัญหาการเกิด Overfitting และเพิ่มความแม่นยำของการพยากรณ์ข้อมูล เทคนิคนี้ทำให้แบบจำลองสามารถใช้ข้อมูลได้อย่างมีประสิทธิภาพมากขึ้น โดยการแบ่งชุดข้อมูลออกเป็นหลายส่วน (Folds) และทำการฝึกและทดสอบซ้ำๆ

หลักการทำงานคือแบ่งชุดข้อมูลออกเป็น K ส่วนที่มีขนาดเท่ากัน (Folds) จากนั้นทำการฝึกโมเดล K รอบ โดยใช้ K-1 Folds เป็น Training Set และใช้ 1 Fold เป็น Validation Set ในแต่ละรอบ เมื่อครบ K รอบ จะนำค่าความแม่นยำของแต่ละรอบมาคำนวณค่าเฉลี่ยเพื่อให้ได้ผลลัพธ์ที่เสถียร วิธีนี้ช่วยให้โมเดลใช้ข้อมูลได้อย่างเต็มที่ ลดปัญหา Overfitting และเพิ่มความแม่นยำของการประเมินผลลัพธ์

รูปภาพประกอบด้วย ภาพหน้าจอ, สี่เหลี่ยม, ข้อความ, ไลน์

เนื้อหาที่สร้างโดย AI อาจไม่ถูกต้อง

*รูปที่ 3 K-fold cross-validation*

ตัวอย่างโครงสร้างกระบวนการทำงานของ K-Fold Cross Validation (K=5) เริ่มจากการแบ่งชุดข้อมูลออกเป็น 5 ส่วน (Folds) ที่มีขนาดเท่ากัน โดยในแต่ละรอบของกระบวนการ 4 Folds จะถูกใช้เป็น Training Set และ 1 Fold จะถูกใช้เป็น Validation Set ซึ่งสลับกันไปจนครบทั้ง 5 รอบ เมื่อกระบวนการฝึกและทดสอบโมเดลเสร็จสิ้น ค่าความแม่นยำหรือค่าความผิดพลาดจากแต่ละรอบจะถูกนำมาคำนวณค่าเฉลี่ยเพื่อให้ได้ผลลัพธ์ที่เสถียร

K-Fold Cross Validation เหมาะกับ โมเดล Machine Learning ที่ต้องการ ประเมินความแม่นยำอย่างมีเสถียรภาพ และลดปัญหา Overfitting โดยเฉพาะในกรณีที่มี ชุดข้อมูลขนาดไม่ใหญ่มาก ซึ่งต้องใช้ข้อมูลให้เกิดประโยชน์สูงสุด นอกจากนี้ ยังเหมาะกับ ปัญหาการจำแนกประเภท (Classification) การพยากรณ์ค่าเชิงปริมาณ (Regression) และการพยากรณ์ข้อมูลอนุกรมเวลา (Time Series)

สมการที่เกี่ยวข้องกับ K-Fold Cross Validation การคำนวณค่าความแม่นยำเฉลี่ยของ K-Fold

(2.7)

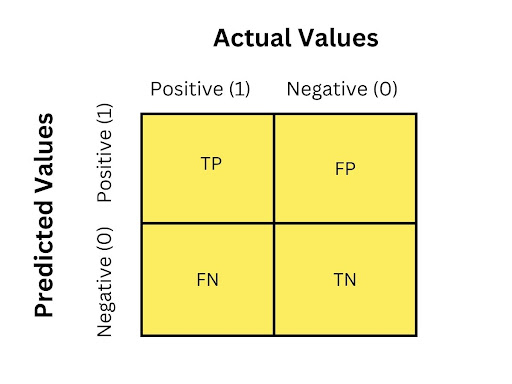
โดยที่

คือ จำนวนของ Folds

คือ ค่าความแม่นยำของโมเดล Fold ที่

สรุป K-Fold Cross Validation เป็นเทคนิคที่ช่วยประเมินประสิทธิภาพของโมเดลโดยแบ่งข้อมูลเป็น K ส่วน และทำการฝึกและทดสอบ K รอบ เพื่อลด Overfitting และเพิ่มความแม่นยำของการพยากรณ์

2.3.2 การประเมินประสิทธิภาพ Confusion Matrix เป็นเครื่องมือที่ใช้สำหรับวัดประสิทธิภาพของแบบจำลองการจำแนกประเภท (Classification Model) โดยเป็นตารางที่ใช้เปรียบเทียบค่าผลลัพธ์ที่ได้จากโมเดลกับค่าจริงของข้อมูล



รูปที่ 4 ตาราง Confusion Matrix

Confusion Matrix สามารถคำนวณค่าตัวชี้วัดสำคัญสำหรับการประเมินประสิทธิภาพของโมเดลได้ ดังนี้

1. Accuracy (ค่าความแม่นยำ) เป็นการวัดความแม่นยำของแบบจำลองโดยรวม ค่าความแม่นยำจะบอกได้ว่าแบบจำลองทำนายถูกกี่ครั้งจากจำนวนการทำนายทั้งหมด โดยมีสมการดังนี้

(2.8)

1. Precision (ค่าความเที่ยงตรง) เป็นการวัดว่าแบบจำลองสามารถทำนายคำตอบถูกต้องเป็นเท่าไหร่จากคลาสที่กำลังพิจารณาอยู่ คำนวณได้จากสมการ

(2.9)

1. Recall(*ค่าความระลึก)* *เป็นการวัดความถูกต้องของแบบจำลองโดยพิจารณาแยกทีละคลาส กล่าวอีกนัยคือจำนวนที่ทายคลาส 1 ถูกต้องเมื่อเทียบกับคลาส 1ทั้งหมด* สามารถคำนวณได้จากสมการ

(2.10)

1. F1-Score(*ค่าความถ่วงดุลระหว่าง* Precision *และ* Recall) *เป็นการวัดค่าเที่ยงตรงและค่าระลึกของแบบจำลองไปพร้อมกัน โดยเขียนสมการได้ดังนี้*

(2.11)

True Positive (TP) เป็นจำนวนคำตอบที่มีผลลัพธ์ทำนายได้ 1 และคำตอบคือ 1

True Negative (TN) เป็นจำนวนคำตอบที่มีผลลัพธ์ทำนายได้ 0 และคำตอบคือ 0

False Negative (FN) เป็นจำนวนคำตอบที่มีผลลัพธ์ทำนายได้ 0 และคำตอบคือ 1

False Positive (FP) เป็นจำนวนคำตอบที่มีผลลัพธ์ทำนายได้ 1 และคำตอบคือ 0

## **2.4 งานวิจัยที่เกี่ยวข้อง**

ยุพิน ไชยสมภารและทวี ชัยพิมลผลิน (2560) ศึกษาเกี่ยวกับการพยากรณ์ระดับน้ำโดยใช้แบบจำลองโครงข่ายประสาทเทียมร่วมกับข้อมูลปริมาณน้ำฝนจากแบบจำลอง WRF-ECHAM5 มีวัตถุประสงค์เพื่อพัฒนาวิธีการพยากรณ์ที่มีประสิทธิภาพสูงในการรับมือกับการเปลี่ยนแปลงสภาพภูมิอากาศและภัยธรรมชาติเช่นน้ำท่วม โดยเฉพาะในพื้นที่ที่เคยได้รับผลกระทบอย่างรุนแรง เช่น จังหวัดเชียงใหม่ การวิจัยนี้มุ่งเน้นการใช้ข้อมูลน้ำฝนจากแบบจำลองภูมิอากาศ WRF-ECHAM5 ร่วมกับโครงข่ายประสาทเทียม (ANN) เพื่อพยากรณ์ระดับน้ำ โดยมีการเปรียบเทียบกระบวนการเรียนรู้ Levenberg-Marquardt (LM) และ Bayesian Regularization (BR) พร้อมกับการปรับปรุงจำนวนโหนดในชั้นซ่อนเร้นและวิธีการหาค่าเฉลี่ยเคลื่อนที่ (Moving Average) ของข้อมูลน้ำฝนเพื่อเพิ่มความแม่นยำในการพยากรณ์ ผลการศึกษาพบว่ากระบวนการ LM มีประสิทธิภาพดีกว่า BR และการใช้ข้อมูลน้ำฝนที่ผ่านการหาค่าเฉลี่ยเคลื่อนที่ช่วยให้การพยากรณ์ระดับน้ำมีความแม่นยำมากขึ้น

Pan et al (2020) พัฒนาแบบจำลองโครงข่ายประสาทเชิงลึก (Deep Neural Network) สำหรับการทำนายระดับน้ำ โดยพิจารณาปัจจัยด้านเวลาและตำแหน่งของเซ็นเซอร์ แบบจำลองที่พัฒนาคือ CNN-GRU ซึ่งผสมผสาน GRU สำหรับเรียนรู้แนวโน้มการเปลี่ยนแปลงระดับน้ำ และ CNN สำหรับจับความสัมพันธ์เชิงพื้นที่ระหว่างสถานีวัดน้ำ แบบจำลอง CNN-GRU มีความแม่นยำสูงกว่าแบบจำลองอื่นๆ เช่น ARIMA, WANN และ LSTM โดยการทดสอบใช้ข้อมูลระดับน้ำย้อนหลัง 30 ปีจากสถานีหลายแห่ง

Aderyani et al (2022) ศึกษาการพยากรณ์ปริมาณน้ำฝนระยะสั้นเพื่อสนับสนุนการบริหารจัดการน้ำและการเตือนภัยน้ำท่วม โดยเปรียบเทียบประสิทธิภาพของ PSO-SVR (Support Vector Regression ที่ปรับแต่งด้วย Particle Swarm Optimization), LSTM (Long Short-Term Memory) และ CNN (Convolutional Neural Network) โดยใช้ข้อมูลจากสถานี Niavaran ในกรุงเตหะราน ประเทศอิหร่าน สำหรับการพยากรณ์ล่วงหน้า 5 นาที และ 15 นาที ผลลัพธ์พบว่า PSO-SVR และ LSTM ให้ผลลัพธ์ที่ดีกว่า CNN และเมื่อนำข้อมูลปริมาณฝนมาจัดกลุ่มด้วย K-Nearest Neighbor (KNN) สามารถช่วยเพิ่มความแม่นยำของแบบจำลองได้ นอกจากนี้ การเพิ่มตัวแปรนำเข้า เช่น ค่าความแตกต่างของปริมาณฝนก่อนหน้า สามารถเพิ่มความแม่นยำของ PSO-SVR ได้สูงถึง 13% และช่วยปรับปรุงประสิทธิภาพของ PSO-SVR และ LSTM ได้ 3–15% และ 2–10% ตามลำดับ ทั้งนี้ พบว่า LSTM เหมาะสมกับการพยากรณ์ระยะสั้น 5 นาที ในขณะที่ PSO-SVR ให้ผลลัพธ์ที่ดีกว่าสำหรับการพยากรณ์ 15 นาที งานวิจัยนี้แสดงให้เห็นว่า LSTM และ PSO-SVR เป็นแบบจำลองที่มีศักยภาพสูงในการพยากรณ์ปริมาณน้ำฝนระยะสั้น และสามารถปรับปรุงความแม่นยำได้โดยการเลือกตัวแปรนำเข้าที่เหมาะสม

เจนจิรา มงคลเมือง ศศิธร พิศพล และ เฉลิมวุฒิ คาเมือง (2561) มีวัตถุประสงค์เพื่อพยากรณ์ปริมาณน้ำฝนรายเดือนของจังหวัดบุรีรัมย์ในปี 2562 โดยใช้ข้อมูลปริมาณน้ำฝนตั้งแต่ปี 2553-2562 และเปรียบเทียบการพยากรณ์ระหว่าง 3 วิธี ได้แก่ วิธีแยกส่วนประกอบ (decomposition method), วิธีปรับให้เรียบอย่างง่าย (simple exponential smoothing), และวิธีฟังก์ชันการถ่ายโอน (transfer function). ผลการวิจัยพบว่าวิธีฟังก์ชันการถ่ายโอนเป็นวิธีที่มีประสิทธิภาพที่สุด เนื่องจากมีค่าเปอร์เซ็นต์ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE) ต่ำสุดที่ 20.43% ซึ่งเหมาะสมที่สุดในการพยากรณ์ปริมาณน้ำฝนรายเดือนสำหรับพื้นที่นี้

จากผลการวิจัย วิธีฟังก์ชันการถ่ายโอนเป็นวิธีที่แม่นยำที่สุดในการคาดการณ์ปริมาณน้ำฝนรายเดือนของจังหวัดบุรีรัมย์ เมื่อเปรียบเทียบกับวิธีอื่นๆ เนื่องจากสามารถลดความคลาดเคลื่อนของข้อมูลได้มากที่สุด โดยมีค่าใกล้เคียงกับข้อมูลจริงที่สุด

รักษ์คณา ภูสีเขียว (2560) มุ่งเน้นการพัฒนาและเปรียบเทียบประสิทธิภาพของแบบจำลองต่างๆ ในการพยากรณ์ปริมาณน้ำฝนล่วงหน้าในระยะสั้น บริเวณสนามบินสุวรรณภูมิโดยใช้โครงข่ายประสาทเทียมแบบย้อนกลับ (RNNs) รวมถึง Long Short-Term Memory (LSTM) และ Gated Recurrent Unit (GRU) ซึ่งเป็นเทคนิคที่ใช้ในการจัดการข้อมูลอนุกรมเวลา ทั้งนี้ ยังเปรียบเทียบกับแบบจำลองทางสถิติดั้งเดิมอย่าง ARIMA และ ARIMAX เพื่อตรวจสอบประสิทธิภาพในการพยากรณ์ โดยใช้ข้อมูลสภาพอากาศรายชั่วโมงตั้งแต่ปี 2016 ถึงปี 2020 ซึ่งได้รับการสนับสนุนจากกรมอุตุนิยมวิทยา วัตถุประสงค์หลักของงานวิจัยนี้คือการเปรียบเทียบผลการพยากรณ์ระหว่างชุดข้อมูลที่ผ่านการเพิ่มคุณลักษณะพิเศษกับชุดข้อมูลดั้งเดิม รวมถึงการประเมินว่าแบบจำลองใดเหมาะสมที่สุดสำหรับการพยากรณ์ปริมาณน้ำฝน ด้วยเป้าหมายเพื่อพัฒนาการคาดการณ์สภาพอากาศที่แม่นยำและสามารถนำไปใช้ประโยชน์ในการวางแผนและการตัดสินใจได้

ทิพาภรณ์ หอมดี et al. (2561) ศึกษาเทคนิคการพยากรณ์ปริมาณน้ำฝนรายเดือนในจังหวัดนครพนม ณ สถานีโครงการน้ำก่ำตอนล่างโดยเปรียบเทียบวิธีการพยากรณ์ 2 วิธี ได้แก่ วิธีบอกซ์-เจนกินส์ (Box-Jenkins) และวิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังที่มีฤดูกาลอย่างง่าย (Simple Seasonal Exponential Smoothing) โดยใช้ข้อมูลปริมาณน้ำฝนรายเดือนตั้งแต่ปี 2552 ถึง 2561 ผลการวิจัยพบว่า วิธีบอกซ์-เจนกินส์ ซึ่งใช้แบบจำลอง ARIMA(0,0,0)(0,1,1) ให้ผลการพยากรณ์ที่แม่นยำกว่าวิธีการทำให้เรียบ โดยพิจารณาจากค่า MAPE และ RMSE ที่ต่ำกว่า ทำให้เหมาะสมในการใช้พยากรณ์ปริมาณน้ำฝนในพื้นที่ดังกล่าว

เฟื่องลดา มะโนกิจ (2559) มุ่งเน้นการพยากรณ์ปริมาณน้ำฝนในประเทศไทยโดยใช้วิธีการเรียนรู้เชิงลึก (Deep Learning) เพื่อเพิ่มความแม่นยำในการพยากรณ์ เนื่องจากประเทศไทยมีความแปรปรวนสูงในการเกิดฝนและข้อมูลที่ไม่สมดุล วิธีการที่ใช้ในการวิจัยประกอบด้วยการใช้แบบจำลองอนุกรมเวลา (Time-series model), การเรียนรู้ของเครื่อง (Machine Learning), และการเรียนรู้เชิงลึก (Deep Learning) โดยเฉพาะการใช้โครงข่ายประสาทเทียม (Artificial Neural Networks) ซึ่งช่วยในการวิเคราะห์ลักษณะสำคัญของข้อมูลและลดปัญหาข้อมูลที่ไม่สมดุล การวิจัยยังพัฒนาแบบจำลองหลายชนิด เช่น Gated Recurrent Unit (GRU) และ Autoencoder ร่วมกับ Multi-Layer Perceptron (MLP) เพื่อปรับปรุงประสิทธิภาพในการพยากรณ์ปริมาณน้ำฝนและแก้ไขปัญหาชั่วโมงที่ฝนไม่ตก โดยเปรียบเทียบผลลัพธ์กับแบบจำลองอนุกรมเวลาและแบบจำลองเรียนรู้เชิงลึกพื้นฐาน

Salman et al.(2018) ได้พัฒนาแบบจำลอง LSTM ผสมผสานกับ ARIMA และพบว่าโมเดลผสมสามารถเพิ่มประสิทธิภาพของการพยากรณ์ได้ดีกว่า LSTM หรือ ARIMA เพียงตัวเดียว ขณะที่ Fan et al. (2020) ได้นำแบบจำลอง LSTM มาเปรียบเทียบกับ ARIMA และพบว่า LSTM มีความแม่นยำในการพยากรณ์ที่สูงกว่า

Aurnhammer & Frank. (2019) และ Yang et al. (2020) ได้ทำการเปรียบเทียบประสิทธิภาพของ RNN, LSTM และ GRU พบว่า LSTM และ GRU มีความโดดเด่นในการพยากรณ์ โดย GRU ให้ผลลัพธ์ที่ดีกว่าเมื่อนำไปใช้กับชุดข้อมูลขนาดเล็กและมีข้อมูลก่อนหน้าไม่มากนัก ขณะที่ LSTM มีความเหมาะสมกับชุดข้อมูลขนาดใหญ่ที่ต้องใช้หน่วยความจำระยะยาว

งานวิจัยส่วนใหญ่มีข้อมูลที่มากและให้ความสำคัญกับการ ปรับจูนไฮเปอร์พารามิเตอร์ หรือพัฒนาแบบจำลองให้เหมาะสมกับข้อมูล โดยยังไม่ได้มีการทดลองโมเดล SVR และข้อมูลที่ไม่มาก ดังนั้น จุดมุ่งหมายหนึ่งของงานวิจัยนี้ คือการเปรียบเทียบความแตกต่างระหว่างการใช้ชุดข้อมูลและวิธีการที่แตกต่างกันเพื่อให้ได้ค่าความแม่นยำมากที่สุด

นอกจากนี้ ยังมีการศึกษาเพื่อเปรียบเทียบความแม่นยำของโครงข่ายประสาทเทียมแบบย้อนกลับ LSTM กับแบบจำลอง SVR เพื่อหาว่าแบบจำลองใดมีแม่นยำมากที่สุด ซึ่งเป็นแนวทางในการเลือกแบบจำลองที่เหมาะสมสำหรับงานพยากรณ์ปริมาณน้ำฝนกับชุดข้อมูลที่มีน้อยในอนาคต

# **บทที่ 3**

# **วิธีดำเนินการวิจัย**

การวิจัยครั้งนี้ เป็นการศึกษาเพื่อหาความแม่นยำของแบบจำลองการพยากรณ์ปริมาณน้ำฝนด้วยเทคนิคการเรียนรู้แบบเครื่อง กรณีศึกษาของจังหวัดนครพนม โดยแบบจำลอง Long Short-Term Memory (LSTM) และ Support Vector Regression (SVR) ที่พัฒนาโดยใช้โปรแกรม MATLAB Version 23.2.0.2485118 (R2023b) ในบทนี้จึงเป็นการเสนอวิธีการดำเนินการวิจัย วิธีการดำเนินการวิจัยแบ่งออกเป็น 3 ส่วนหลัก ดังนี้

3.1 การรวบรวมข้อมูล และเตรียมข้อมูล

3.1.1 แหล่งที่มาและลักษณะของข้อมูล

3.1.2 การทำความสะอาดข้อมูล (Data Cleaning)

3.2 การพัฒนาแบบจำลองเพื่อพยากรณ์

3.2.1 การพัฒนาแบบจำลอง Long Short-Term Memory (LSTM)

3.2.2 การพัฒนาแบบจำลอง Support Vector Regression (SVR)

3.2.3 การปรับแต่งพารามิเตอร์ของแบบจำลอง (Hyperparameter Tuning)

3.3 การทดสอบและประเมินประสิทธิภาพแบบจำลอง

3.3.1 การประเมินด้วยวิธี K-Fold Cross Validation

3.3.2 การประเมินด้วยวิธี Confusion Matrix

3.3.3 การวิเคราะห์ค่าความคลาดเคลื่อนของแบบจำลอง

ในส่วนของลำดับขั้นตอนในการดำเนินการของงานวิจัยชิ้นนี้ สามารถเขียนเป็นแผนผังโดย คร่าวได้ดังต่อไปนี้

การพัฒนาแบบจำลอง

การรวบรวมข้อมูล

การทดสอบโมเดล

Cleaning data

SVR

LSTM

K-Fold Cross Validation

Confusion Matrix

รูปที่ 5 ขั้นตอนการดำเนินงาน

## **3.1 การรวบรวมข้อมูล และเตรียมข้อมูล**

การเก็บรวบรวมข้อมูลที่มีคุณภาพและการเตรียมข้อมูลอย่างเหมาะสมเป็นขั้นตอนสำคัญที่ส่งผลต่อความแม่นยำของแบบจำลองการพยากรณ์ (Hassan et al., 2023) ในการวิจัยนี้ ข้อมูลที่ใช้ประกอบด้วยข้อมูลอุตุนิยมวิทยาที่เก็บรวบรวมจากกรมอุตุนิยมวิทยาจังหวัดนครพนม โดยมีรายละเอียดดังนี้

### 3.1.1 แหล่งที่มาและลักษณะของข้อมูล

งานวิจัยนี้ได้รับการสนับสนุนข้อมูลสภาพอากาศจากกรมอุตุนิยมวิทยาจังหวัดนครพนมย้อนหลัง 2 ปี 8 เดือน (ตั้งแต่ 1 มกราคม 2565 จนถึง 7 กันยายน 2567) โดยเป็นข้อมูลที่บันทึกรายวันประกอบด้วยตัวแปรทางอุตุนิยมวิทยา 13 ตัวแปร ได้แก่

1) ความกดอากาศรายวัน (สูงสุด ต่ำสุด เฉลี่ย 8 เวลา) หน่วยเป็น hPa

2) อุณหภูมิอากาศรายวัน (สูงสุด ต่ำสุด เฉลี่ย) หน่วยเป็นองศาเซลเซียส

3) ความชื้นสัมพัทธ์ (สูงสุด ต่ำสุด เฉลี่ย) หน่วยเป็นเปอร์เซ็นต์

4) ปริมาณน้ำฝน หน่วยเป็นมิลลิเมตร

5) ปริมาณน้ำระเหย หน่วยเป็นมิลลิเมตร

6) ชั่วโมงแสงแดด หน่วยเป็นชั่วโมง

7) ความเร็วลม หน่วยเป็นน็อต

ลักษณะของข้อมูลดิบที่ได้รับจากกรมอุตุนิยมวิทยาจังหวัดนครพนม แสดงดังรูปที่ 6

รูปภาพประกอบด้วย ข้อความ, จำนวน, ภาพหน้าจอ, ขนาน

คำอธิบายที่สร้างโดยอัตโนมัติ

รูปที่ 6 ตัวอย่างชุดข้อมูลที่ได้รับจากกรมอุตุนิยมวิทยาจังหวัดนครพนม

จากรูปที่ 6 จะเห็นว่าชุดข้อมูลที่ได้รับมานั้น ยังไม่อยู่ในสภาพที่สามารถจะนำไปใช้ในการคำนวณ หรือ นำเข้าแบบจำลองได้ ซึ่งการจะนำข้อมูลทั้งหมดที่ได้รับมานั้นไปใช้งานต่อ จำเป็นที่จะต้องทำความสะอาดข้อมูล (Data cleaning) ใหม่เสียก่อน แล้วจึงนำข้อมูลที่ผ่านการทำความสะอาดแล้วไป เข้าสู่ขั้นตอนพัฒนาโมเดล ในกรณีของปริมาณน้ำฝนสะสมนั้น จะมีสัญลักษณ์พิเศษเพิ่มขึ้นมาคือ สัญลักษณ์ T ซึ่งมีความหมายว่า ปริมาณน้ำฝนวัดค่าไม่ได้หรือน้อยกว่า 0.1 มิลลิเมตร ซึ่งเราได้แทนค่า T ดังกล่าว ด้วยค่า 0 เนื่องจากในทางอุตุนิยมวิทยานั้น ไม่สามารถที่จะวัดปริมาณน้ำฝนที่มีค่าน้อยกว่า 0.1 มิลลิเมตรได้ เราจึงจะให้ความหมายค่า T นี้ว่าเป็นช่วงเวลาที่ฝนไม่ตก

### 3.1.2 การทำความสะอาดข้อมูล (Data Cleaning)

ในการเตรียมข้อมูลสำหรับการพยากรณ์ปริมาณน้ำฝน ขั้นตอนการทำความสะอาดข้อมูล(Data Cleaning) เป็นกระบวนการสำคัญที่ช่วยให้แบบจำลองสามารถเรียนรู้จากข้อมูลที่มีคุณภาพและลดข้อผิดพลาดที่อาจเกิดขึ้นในการวิเคราะห์ โดยขั้นตอนที่ดำเนินการมีดังนี้

3.1.2.1 การจัดการข้อมูลที่ไม่สมบูรณ์และค่าผิดปกติ

1. การลบค่าผิดปกติ (Outlier Removal) ข้อมูลที่มีค่าผิดแปลกเกินความเป็นจริง เช่น ค่าปริมาณฝนที่สูงผิดปกติ หรือค่าลบที่ไม่สมเหตุสมผล จะถูกวิเคราะห์และตัดออก เพื่อให้โมเดลสามารถเรียนรู้จากข้อมูลที่เป็นตัวแทนของความเป็นจริงได้ดีขึ้น
2. การตรวจสอบค่าความผิดพลาดในการบันทึกข้อมูล ในบางกรณีอาจเกิดข้อผิดพลาดจากเซ็นเซอร์หรือกระบวนการบันทึกข้อมูล เช่น การกรอกค่าผิดพลาด หรือหน่วยที่ไม่ตรงกัน ซึ่งต้องมีการปรับแก้เพื่อให้แน่ใจว่าข้อมูลมีความถูกต้องและสามารถนำไปใช้งานได้ ตัวอย่างข้อมูลที่ได้ทำความสะอาดข้อมูล สามารถแสดงได้ดังรูปที่ 7

*รูปภาพประกอบด้วย ข้อความ, จำนวน, ขนาน, ภาพหน้าจอ

เนื้อหาที่สร้างโดย AI อาจไม่ถูกต้อง* 3.1.2.2 การปรับขนาดและการแปลงค่าข้อมูล

รูปที่  ตัวอย่างข้อมูลที่ได้ทำความสะอาดข้อมูล

เพื่อลดความซับซ้อน และหลีกเลี่ยงความผิดปกติของข้อมูลในแบบจำลอง โดยจะทำการปรับปรุง โครงสร้างข้อมูลด้วยวิธี Standardization หรือ Scaling ข้อมูลโดยใช้ Min-Max Scaling เพื่อให้ค่าของตัวแปรทั้งหมดอยู่ในช่วงที่ ปรับขนาดข้อมูลให้อยู่ในช่วง 0,1 หรือ −1,1 เพื่อทำให้ค่าของตัวแปร อยู่ในช่วงที่เทียบเคียงกัน สามารถแสดงสมการได้ดังนี้

(3.1)

โดยที่

คือ ค่าของข้อมูลที่ต้องปรับ

คือ ค่าจริงของตัวแปร

คือ ค่าต่ำสุดและค่าสูงสุดของตัวแปร

รูปภาพประกอบด้วย ข้อความ, จำนวน, ขนาน

เนื้อหาที่สร้างโดย AI อาจไม่ถูกต้อง

รูปที่ 8 ตัวอย่างข้อมูลที่ผ่านการทำ Standardization

3.1.2.3 การแบ่งข้อมูลให้กับการฝึกและทดสอบโมเดล

การแบ่งข้อมูลเป็นส่วนสำคัญในกระบวนการพยากรณ์ เพื่อทำการสร้างแบบจำลอง และการทดสอบ จะทำการแบ่งชุดข้อมูลออกเป็นกลุ่มๆ เพื่อประกอบการพิจารณาในการสร้างตัวแบบ

*รูปภาพประกอบด้วย ข้อความ, ภาพหน้าจอ, แผนภาพ, พล็อต

เนื้อหาที่สร้างโดย AI อาจไม่ถูกต้อง*

รูปที่ 9 กราฟแสดงสัดส่วนปริมาณน้ำฝน

ทำการแบ่งข้อมูลออกเป็น 2 ส่วน ดังนี้ 1) ข้อมูลตั้งแต่วันที่ 1 มกราคม 2565 ถึง 23 กุมภาพันธ์ 2567 เป็นกลุ่มฝึกสอน (Training data set) 2) ข้อมูลวันที่ 24 กุมภาพันธ์ 2567 ถึง 7 กันยายน 2567 เป็นกลุ่มทดสอบ (Test data set) โดยสรุปได้ดังนี้

**ตารางที่ 1** ตารางสรุปการแบ่งข้อมูลโดยละเอียด

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ชุดข้อมูล | จำนวนวัน | ฝนตก (วัน) | ฝนไม่ตก (วัน) | ฝนตก (%) | ฝนไม่ตกตก (%) |
| ชุดฝึกสอน (80%) | 784 วัน | 258 วัน | 526 วัน | 32.91% | 67.09% |
| ชุดทดสอบ (20%) | 196 วัน | 114 วัน | 82 วัน | 58.16% | 41.84% |

จากตารางที่ 1 จะเห็นได้ว่าข้อมูลมีความไม่สมดุลระหว่างวันที่มีฝนตกและวันที่ไม่มีฝนตก โดยในชุดข้อมูลฝึกสอนมีวันที่ไม่มีฝนตกถึง 67.09% ขณะที่ในชุดข้อมูลทดสอบมีวันที่มีฝนตกมากกว่า (58.16%) เนื่องจากช่วงเวลาของชุดข้อมูลทดสอบอยู่ในช่วงฤดูฝน ความไม่สมดุลนี้เป็นความท้าทายสำคัญในการพัฒนาแบบจำลองที่มีความแม่นยำสูง และได้ถูกนำมาพิจารณาในการปรับแต่งพารามิเตอร์ของแบบจำลองและการประเมินประสิทธิภาพ

## **3.2 การพัฒนาแบบจำลองเพื่อการพยากรณ์**

หลังจากการเตรียมข้อมูลเรียบร้อยแล้ว ขั้นตอนต่อไปคือการพัฒนาแบบจำลองเพื่อพยากรณ์ปริมาณน้ำฝน งานวิจัยนี้ได้พัฒนาและเปรียบเทียบแบบจำลอง 2 ประเภท ได้แก่ Long Short-Term Memory (LSTM) และ Support Vector Regression (SVR) โดยใช้โปรแกรม MATLAB Version 23.2.0.2485118 (R2023b) ซึ่งมีความสามารถในการทำงานกับแบบจำลอง Machine Learning และ Deep Learning อย่างมีประสิทธิภาพ

### **3.2.1 แบบจำลอง Long Short-Term Memory (LSTM)**

LSTM เป็นแบบจำลองประเภท Deep Learning ที่ออกแบบมาเพื่อเรียนรู้ความสัมพันธ์ระยะยาวในข้อมูลอนุกรมเวลา ซึ่งเหมาะสมกับการพยากรณ์ปริมาณน้ำฝนที่มีความผันผวนสูงและขึ้นอยู่กับปัจจัยทางอุตุนิยมวิทยาในช่วงก่อนหน้า (Fan et al., 2020) กระบวนการพัฒนาแบบจำลอง LSTM มีดังนี้

3.2.1.1 การออกแบบโครงสร้างของแบบจำลอง LSTM โครงสร้างของแบบจำลอง LSTM ประกอบด้วย

1) ชั้นนำเข้า (Input Layer): รับข้อมูลตัวแปรทางอุตุนิยมวิทยาจำนวน 10 ตัวแปร ในรูปแบบลำดับเวลา (Time Sequence) ย้อนหลัง 7 วัน

2) ชั้น LSTM (LSTM Layer): ประกอบด้วยหน่วย LSTM จำนวน 50 หน่วย ซึ่งได้จากการทดลองหาค่าที่เหมาะสม

3) ชั้นเชื่อมต่อเต็มรูปแบบ (Fully Connected Layer): ทำหน้าที่แปลงผลลัพธ์จากชั้น LSTM เป็นค่าพยากรณ์ปริมาณน้ำฝนรายวัน

4) ชั้นถดถอย (Regression Layer): ใช้สำหรับการพยากรณ์ค่าต่อเนื่องของปริมาณน้ำฝน

3.2.1.2 การกำหนดพารามิเตอร์ของแบบจำลอง LSTM กระบวนการปรับแต่งพารามิเตอร์ของแบบจำลอง LSTM มีดังนี้

1) จำนวน Epoch กำหนดให้เท่ากับ 1,500 ซึ่งเพียงพอสำหรับการเรียนรู้รูปแบบในข้อมูลโดยไม่เกิด Overfitting

2) อัตราการเรียนรู้ (Learning Rate): เริ่มต้นที่ 0.001 และลดลงด้วยอัตรา 0.1 หลังจากทุก 200 Epoch เพื่อให้การเรียนรู้ค่อยๆ ละเอียดขึ้น

3) Optimizer ใช้ Adam (Adaptive Moment Estimation) ซึ่งเหมาะสมกับแบบจำลอง Deep Learning เนื่องจากสามารถปรับอัตราการเรียนรู้อัตโนมัติตามลักษณะของข้อมูล

4) Activation Function ใช้ Tanh (Hyperbolic Tangent) ในชั้น LSTM เนื่องจากช่วยให้การเรียนรู้มีความเสถียรและลดปัญหา Vanishing Gradient

5) Batch Size กำหนดให้เท่ากับ 32 ซึ่งเป็นค่าที่สมดุลระหว่างการเรียนรู้ที่เสถียรและการใช้ทรัพยากรคอมพิวเตอร์

3.2.1.3 การฝึกแบบจำลอง LSTM

แบบจำลอง LSTM ถูกฝึกด้วยชุดข้อมูลฝึกสอนในรูปแบบ Time Series โดยใช้ Sliding Window ขนาด 7 วัน เลื่อนทีละ 1 วัน เพื่อให้แบบจำลองเรียนรู้ความสัมพันธ์ระหว่างข้อมูลย้อนหลังและปริมาณน้ำฝนในวันถัดไป กระบวนการฝึกแบบจำลองประกอบด้วย:

1) การจัดรูปแบบข้อมูล: แปลงข้อมูลให้อยู่ในรูปแบบ 3D Array ขนาด [numFeatures × sequenceLength × numObservations] ซึ่งเหมาะสมกับการนำเข้าแบบจำลอง LSTM

2) การแบ่งข้อมูลสำหรับ Cross-Validation: ใช้วิธี 5-Fold Cross Validation เพื่อประเมินประสิทธิภาพของแบบจำลองระหว่างการฝึก

**3.2.2 แบบจำลอง Support Vector Regression (SVR)**

SVR เป็นแบบจำลองประเภท Machine Learning ที่มีความสามารถในการจับความสัมพันธ์ไม่เชิงเส้นระหว่างตัวแปรต่างๆ โดยใช้ Kernel Trick เพื่อแปลงข้อมูลไปยังมิติที่สูงขึ้น (Qiu et al., 2020) กระบวนการพัฒนาแบบจำลอง SVR มีดังนี้

3.2.2.1 การเลือก Kernel และการปรับแต่งพารามิเตอร์

การเลือก Kernel ที่เหมาะสมเป็นปัจจัยสำคัญที่ส่งผลต่อประสิทธิภาพของแบบจำลอง SVR (Aderyani et al., 2022) ในงานวิจัยนี้ ได้ทดลองใช้ Kernel 3 ประเภท ได้แก่ Linear, Polynomial และ Radial Basis Function (RBF) และพบว่า RBF Kernel ให้ผลลัพธ์ที่ดีที่สุดสำหรับข้อมูลปริมาณน้ำฝนที่มีความซับซ้อนและไม่เชิงเส้น

พารามิเตอร์สำคัญของแบบจำลอง SVR ที่ต้องปรับแต่งมีดังนี้

1) C (Box Constraint): ควบคุมความสมดุลระหว่างความซับซ้อนของแบบจำลองและความผิดพลาดในการฝึก กำหนดให้เท่ากับ 180,000 ซึ่งได้จากการทดลองหาค่าที่เหมาะสม

2) Epsilon (ε): กำหนดความกว้างของช่องว่าง (ε-tube) ที่ยอมรับความผิดพลาดได้ กำหนดให้เท่ากับ 0.05

3) Gamma (γ): พารามิเตอร์ของ RBF Kernel ที่ควบคุมอิทธิพลของแต่ละตัวอย่างฝึก กำหนดให้เท่ากับ 0.1

3.2.2.2 การฝึกแบบจำลอง SVR

แบบจำลอง SVR ถูกฝึกด้วยชุดข้อมูลฝึกสอนที่ประกอบด้วยตัวแปรทางอุตุนิยมวิทยา 10 ตัวแปรในวันปัจจุบัน (ไม่รวมลำดับเวลาย้อนหลังเหมือน LSTM) กระบวนการฝึกแบบจำลองประกอบด้วย:

1) การจัดรูปแบบข้อมูล: แปลงข้อมูลให้อยู่ในรูปแบบ 2D Array ขนาด [numObservations × numFeatures] ซึ่งเหมาะสมกับการนำเข้าแบบจำลอง SVR

2) การแบ่งข้อมูลสำหรับ Cross-Validation: ใช้วิธี 5-Fold Cross Validation เช่นเดียวกับแบบจำลอง LSTM

3) การแก้ไขปัญหาข้อมูลไม่สมดุล: ใช้เทคนิค SMOTE (Synthetic Minority Over-sampling Technique) เพื่อสร้างตัวอย่างเพิ่มเติมสำหรับวันที่มีฝนตก ทำให้ข้อมูลมีความสมดุลมากขึ้น

3.2.2.3 การใช้เทคนิค Feature Selection สำหรับแบบจำลอง SVR

เนื่องจาก SVR อาจมีประสิทธิภาพลดลงเมื่อมีตัวแปรนำเข้าจำนวนมาก จึงได้ใช้เทคนิค Feature Selection เพื่อคัดเลือกตัวแปรที่มีอิทธิพลต่อปริมาณน้ำฝนมากที่สุด (P Sai et al., 2024) โดยใช้วิธี Feature Importance ด้วยเทคนิค Permutation Importance ซึ่งวัดการลดลงของประสิทธิภาพเมื่อสลับค่าของตัวแปรแต่ละตัว

ผลการวิเคราะห์พบว่า ตัวแปรที่มีอิทธิพลต่อปริมาณน้ำฝนมากที่สุด 5 อันดับแรก ได้แก่

1) ความกดอากาศต่ำสุด

2) ความชื้นสัมพัทธ์ต่ำสุด

3) ความชื้นสัมพัทธ์เฉลี่ย

4) อุณหภูมิสูงสุด

5) ความกดอากาศเฉลี่ย 8 เวลา

อย่างไรก็ตาม เพื่อให้การเปรียบเทียบกับแบบจำลอง LSTM มีความยุติธรรม จึงตัดสินใจใช้ตัวแปรนำเข้าทั้ง 10 ตัวเหมือนกันสำหรับทั้งสองแบบจำลอง

**3.3 การทดสอบโมเดล**

การทดสอบและประเมินประสิทธิภาพของแบบจำลองเป็นขั้นตอนสำคัญในการระบุว่าแบบจำลองใดเหมาะสมที่สุดสำหรับการพยากรณ์ปริมาณน้ำฝนในจังหวัดนครพนม งานวิจัยนี้ใช้วิธีการประเมินประสิทธิภาพหลายวิธีเพื่อให้ได้ผลการประเมินที่ครอบคลุมและเชื่อถือได้

### 3.3.1 วิธี K-Fold Cross Validation

ในการประเมินประสิทธิภาพของแบบจำลอง งานวิจัยนี้ใช้เทคนิค K-Fold Cross Validation โดยแบ่งข้อมูลทั้งหมดออกเป็น 5 ส่วนที่มีจำนวนตัวอย่างใกล้เคียงกัน โดยลำดับของข้อมูลจะเป็นการสุ่ม เพื่อป้องกันการลำเอียงจากลำดับข้อมูลจริง (randomized split) ในแต่ละรอบจะใช้ข้อมูล K-1 ส่วนสำหรับการฝึกโมเดล และอีก 1 ส่วนสำหรับการทดสอบโมเดล ดำเนินการซ้ำทั้งหมด K รอบ เพื่อให้ข้อมูลทุกส่วนถูกใช้เป็นชุดทดสอบ ผลการพยากรณ์ในแต่ละรอบจะถูกประเมินด้วยตัวชี้วัด Accuracy, Precision, Recall และ F1-Score พร้อมทั้งสรุปค่าประสิทธิภาพเฉลี่ย

### 3.3.2 วิธี Confusion Matrix

เพื่อประเมินผลของโมเดลในกรณีที่เป็นปัญหาการจำแนกประเภท (Classification) งานวิจัยนี้ใช้ Confusion Matrix เป็นเกณฑ์วัดผล เฉพาะข้อมูลของชุดทดสอบ (Test data set 20% ตั้งแต่วันที่ 24 กุมภาพันธ์ 2567 ถึง 7 กันยายน 2567) เปรียบเทียบค่าจริงกับค่าผลลัพธ์ที่ได้จากการพยากรณ์ของโมเดล โดยเป็นตารางที่แสดงจำนวนค่าที่คาดการณ์ถูกต้องและผิดพลาดในแต่ละประเภทของข้อมูล ใช้ตัวชี้วัด Precision, Recall, F1-Score และ Accuracy ในงานวิจัยนี้ ใช้ Confusion Matrix เพื่อประเมินความสามารถในการจำแนกระดับปริมาณน้ำฝนตามเกณฑ์ของกรมอุตุนิยมวิทยา 5 ระดับ มีขั้นตอนดังนี้:

1) แปลงค่าพยากรณ์ปริมาณน้ำฝนเป็นระดับตามเกณฑ์ของกรมอุตุนิยมวิทยา

ดังแสดงในรูปที่ 10

รูปภาพประกอบด้วย ข้อความ, ภาพหน้าจอ, ตัวอักษร

เนื้อหาที่สร้างโดย AI อาจไม่ถูกต้อง

*รูปที่ 10 เกณฑ์การตรวจวัดฝน ปริมาณฝนรายวัน ของกรมอุตุนิยมวิทยา*

จากรูปที่ 10 เกณฑ์การวัดฝน ปริมาณฝนรายวัน ที่ได้มาจากกรมอุตุนิยมวิทยา*(http://www.arcims.tmd.go.th/dailydata/DetailDailyRain.html)*

2) สร้าง Confusion Matrix ขนาด 5×5 เพื่อแสดงจำนวนตัวอย่างที่ถูกจำแนกถูกต้องและผิดพลาดในแต่ละระดับ

3) คำนวณตัวชี้วัดประสิทธิภาพสำหรับแต่ละระดับ ได้แก่ Precision, Recall และ

F1-Score

4) คำนวณค่า Accuracy โดยรวมของการจำแนกระดับปริมาณน้ำฝน

### 3.3.3 การวิเคราะห์ค่าความคลาดเคลื่อนของแบบจำลอง

นอกจากการประเมินความสามารถในการจำแนกระดับปริมาณน้ำฝนแล้ว ยังได้ทำการวิเคราะห์ค่าความคลาดเคลื่อนของการพยากรณ์ค่าปริมาณน้ำฝนเชิงปริมาณ โดยใช้ตัวชี้วัดดังนี้:

1) Mean Absolute Error (MAE): วัดค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ระหว่างค่าจริงและค่าที่ทำนายได้ คำนวณจากสมการ

MAE= (3.2)

โดยที่ ​ คือค่าจริง และ คือค่าที่ทำนายได้

2) Root Mean Square Error (RMSE): วัดรากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง ให้น้ำหนักกับความคลาดเคลื่อนขนาดใหญ่มากกว่า MAE คำนวณจากสมการ

RMSE= (3.3)

3) Coefficient of Determination (R²): วัดสัดส่วนของความแปรปรวนในตัวแปรตามที่อธิบายได้ด้วยตัวแปรอิสระ คำนวณจากสมการ

R2= (3.4)

โดยที่ คือค่าเฉลี่ยของค่าจริง

ค่า MAE และ RMSE ที่ต่ำ และค่า R² ที่สูง (เข้าใกล้ 1) แสดงถึงความแม่นยำที่สูงของแบบจำลอง

**3.3.4 การเปรียบเทียบประสิทธิภาพของแบบจำลอง**

หลังจากประเมินประสิทธิภาพของแบบจำลองทั้งสอง ได้ทำการเปรียบเทียบผลลัพธ์โดยพิจารณาปัจจัยต่างๆ ดังนี้:

1) ความแม่นยำในการพยากรณ์: เปรียบเทียบค่า Accuracy, Precision, Recall และ F1-Score จากการประเมินด้วย K-Fold Cross Validation และ Confusion Matrix

2) ความคลาดเคลื่อนของการพยากรณ์: เปรียบเทียบค่า MAE, RMSE และ R² ทั้งในชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบ

3) ความสามารถในการจำแนกระดับปริมาณน้ำฝน: วิเคราะห์ประสิทธิภาพในการจำแนกระดับปริมาณน้ำฝนแต่ละระดับ โดยเฉพาะความสามารถในการตรวจจับวันที่มีฝนตกหนักซึ่งมีความสำคัญในการเตือนภัยน้ำท่วม

4) ความสามารถในการทำนายวันที่มีฝนตกและไม่มีฝนตก: เปรียบเทียบค่า Precision และ Recall สำหรับการทำนายวันที่มีฝนตก (ปริมาณน้ำฝน > 0.1 มิลลิเมตร) และวันที่ไม่มีฝนตก

ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองทั้งสองจะถูกนำเสนอในบทที่ 4 และนำไปสู่การสรุปว่าแบบจำลองใดเหมาะสมที่สุดสำหรับการพยากรณ์ปริมาณน้ำฝนในจังหวัดนครพนม

# **บทที่ 4**

# **ผลการวิจัย**

บทนี้นำเสนอผลการทดลองของแบบจำลอง Long Short-Term Memory (LSTM) และ Support Vector Regression (SVR) สำหรับการพยากรณ์ปริมาณน้ำฝน โดยใช้กระบวนการประเมินผลผ่าน K-Fold Cross Validation, Confusion Matrix และ ตัวชี้วัดประสิทธิภาพของโมเดล เพื่อเปรียบเทียบประสิทธิภาพของแต่ละแบบจำลอง

นอกจากนี้ บทนี้ยังนำเสนอการวิเคราะห์เชิงลึกเกี่ยวกับ ข้อดี และ ข้อจำกัด ของแบบจำลองแต่ละประเภท ตลอดจนแนวทางในการปรับปรุงแบบจำลองให้สามารถพยากรณ์ปริมาณน้ำฝนได้อย่างแม่นยำมากยิ่งขึ้น เพื่อเพิ่มประสิทธิภาพของการคาดการณ์และความน่าเชื่อถือของผลลัพธ์

## **4.1 ผลลัพธ์จาก K-Fold Cross Validation**

### 4.1.1 ผลลัพธ์ 5-Fold Cross Validation

**ตารางที่ 2** ผลลัพธ์ของการทดสอบ 5 Fold Cross Validation

|  |  |  |
| --- | --- | --- |
| Fold | LSTM | SVR |
| 1 | 0.7546 | 0.7497 |
| 2 | 0.7477 | 0.7163 |
| 3 | 0.7559 | 0.7714 |
| 4 | 0.7723 | 0.7694 |
| 5 | 0.7835 | 0.7592 |
| Accuracy | 0.7628 | 0.7532 |

จากตารางที่ 2 แสดงผลการทดสอบประสิทธิภาพของแบบจำลองโดยใช้ 5-Fold Cross Validation พบว่าแบบจำลอง LSTM มีค่าความถูกต้องเฉลี่ย (Average Accuracy) เท่ากับ 76.28% ซึ่งสูงกว่าแบบจำลอง SVR ที่มีค่าความถูกต้องเฉลี่ยเท่ากับ 75.32% แม้ว่าความแตกต่างจะไม่มากนัก (0.96%)

นอกจากนี้ เมื่อพิจารณาค่าความถูกต้องในแต่ละ Fold พบว่าแบบจำลอง LSTM มีความสม่ำเสมอมากกว่า โดยมีค่าส่วนเบี่ยงเบนมาตรฐานของความถูกต้องเท่ากับ 0.014 เทียบกับแบบจำลอง SVR ที่มีค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.022 ซึ่งแสดงให้เห็นว่าแบบจำลอง LSTM มีความเสถียรในการพยากรณ์มากกว่า ไม่ว่าจะใช้ชุดข้อมูลใดในการฝึกและทดสอบ

**ตารางที่ 3** ตารางเปรียบเทียบประสิทธิภาพของโมเดล LSTM และ SVR

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| แบบจำลอง | Accuracy | Precision | Recall | F1-Score |
| LSTM | 76.28% | 27.47% | 26.42% | 26.92% |
| SVR | 75.32% | 22.03% | 23.05% | 22.51% |

จากตารางที่ 3 แสดงการเปรียบเทียบประสิทธิภาพโดยรวมของแบบจำลอง LSTM และ SVR พบว่าแบบจำลอง LSTM มีประสิทธิภาพสูงกว่าในทุกมิติของการวัดผล ทั้งในด้านความถูกต้อง (Accuracy) ความแม่นยำ (Precision) ความครบถ้วน (Recall) และค่า F1-Score

แม้ว่าทั้งสองแบบจำลองจะมีค่าความถูกต้องที่ค่อนข้างสูง (มากกว่า 75%) แต่มีค่า Precision, Recall และ F1-Score ที่ค่อนข้างต่ำ (ต่ำกว่า 30%) ซึ่งสะท้อนให้เห็นถึงความท้าทายในการพยากรณ์ปริมาณน้ำฝนที่มีความผันผวนสูงและมีข้อมูลไม่สมดุล โดยเฉพาะการพยากรณ์วันที่มีฝนตก ซึ่งมีสัดส่วนน้อยกว่าวันที่ไม่มีฝนตกในชุดข้อมูลฝึกสอน (32.91% เทียบกับ 67.09%)

อย่างไรก็ตาม แบบจำลอง LSTM มีค่า Precision สูงกว่า SVR ถึง 5.44% (27.47% เทียบกับ 22.03%) ซึ่งแสดงให้เห็นว่า LSTM มีความสามารถในการทำนายวันที่มีฝนตกได้แม่นยำกว่า SVR อย่างมีนัยสำคัญ

## **4.2 ผลลัพธ์จาก Confusion Matrix**

### 4.2.1 ผลลัพธ์ Confusion Matrix ของ LSTM

**ตารางที่ 4** ผลลัพธ์ LSTM Confusion Matrix

|  |  |  |  |
| --- | --- | --- | --- |
| เกณฑ์การวัดระดับ | Precision | Recall | F1-Score |
| ฝนวัดจำนวนไม่ได้ (น้อยกว่า 0 มิลลิเมตร) | 0.6667 | 0.3457 | 0.4553 |
| ฝนตกน้อย (ตั้งแต่ 0.1 มิลลิเมตร ถึง 10มิลลิเมตร) | 0.3256 | 0.5185 | 0.4 |
| ฝนตกปานกลาง (ตั้งแต่ 10.1 มิลลิเมตร ถึง 35มิลลิเมตร) | 0.2453 | 0.3514 | 0.2889 |
| ฝนตกหนัก (ตั้งแต่ 35.1 มิลลิเมตร ถึง 90 มิลลิเมตร) | 0.2857 | 0.1818 | 0.2222 |
| ฝนตกหนักมาก (ตั้งแต่ 90.1 มิลลิเมตร ขึ้นไป) | 0 | 0 | 0 |
| Accuracy: 37.44% | | | |

### 4.2.2 ผลลัพธ์ Confusion Matrix ของ SVR

**ตารางที่ 5** ผลลัพธ์ SVR Confusion Matrix

|  |  |  |  |
| --- | --- | --- | --- |
| เกณฑ์การวัดระดับ | Precision | Recall | F1-Score |
| ฝนวัดจำนวนไม่ได้ (น้อยกว่า 0 มิลลิเมตร) | 0.5536 | 0.378 | 0.4493 |
| ฝนตกน้อย (ตั้งแต่ 0.1 มิลลิเมตร ถึง 10มิลลิเมตร) | 0.32 | 0.2963 | 0.3077 |
| ฝนตกปานกลาง (ตั้งแต่ 10.1 มิลลิเมตร ถึง 35มิลลิเมตร) | 0.1042 | 0.1351 | 0.1176 |
| ฝนตกหนัก (ตั้งแต่ 35.1 มิลลิเมตร ถึง 90 มิลลิเมตร) | 0.2424 | 0.3636 | 0.2909 |
| ฝนตกหนักมาก (ตั้งแต่ 90.1 มิลลิเมตร ขึ้นไป) | 0 | 0 | 0 |
| Accuracy: 30.61% | | | |

จากตารางที่ 4 และ 5 แสดงผลลัพธ์ Confusion Matrix ของแบบจำลอง LSTM และ SVR ในการจำแนกระดับปริมาณน้ำฝนตามเกณฑ์ของกรมอุตุนิยมวิทยา 5 ระดับ พบว่า

**1) ความถูกต้องโดยรวม**: แบบจำลอง LSTM มีค่าความถูกต้อง (Accuracy) ในการจำแนกระดับปริมาณน้ำฝน 37.44% ซึ่งสูงกว่าแบบจำลอง SVR ที่มีค่าความถูกต้อง 30.61% อย่างมีนัยสำคัญ (เพิ่มขึ้น 6.83%)

**2) การพยากรณ์วันที่ไม่มีฝนตก**: ทั้งสองแบบจำลองมีประสิทธิภาพสูงสุดในการพยากรณ์วันที่ไม่มีฝนตก (น้อยกว่า 0.1 มิลลิเมตร) โดย LSTM มีค่า Precision สูงกว่า (0.6667 เทียบกับ 0.5536) แต่ SVR มีค่า Recall สูงกว่าเล็กน้อย (0.3780 เทียบกับ 0.3457)

**3) การพยากรณ์ฝนตกน้อย**: LSTM มีประสิทธิภาพสูงกว่าในการพยากรณ์วันที่มีฝนตกน้อย (0.1-10 มิลลิเมตร) ทั้งในด้าน Precision (0.3256 เทียบกับ 0.3200) และ Recall (0.5185 เทียบกับ 0.2963) โดยมีค่า F1-Score สูงกว่าถึง 0.0923 (0.4000 เทียบกับ 0.3077)

**4) การพยากรณ์ฝนตกปานกลาง**: LSTM มีประสิทธิภาพสูงกว่าอย่างชัดเจนในการพยากรณ์วันที่มีฝนตกปานกลาง (10.1-35 มิลลิเมตร) โดยมีค่า Precision (0.2453 เทียบกับ 0.1042) และ Recall (0.3514 เทียบกับ 0.1351) สูงกว่ามาก

**5) การพยากรณ์ฝนตกหนัก**: SVR มีประสิทธิภาพสูงกว่าในการพยากรณ์วันที่มีฝนตกหนัก (35.1-90 มิลลิเมตร) ในด้าน Recall (0.3636 เทียบกับ 0.1818) แต่ LSTM มีค่า Precision สูงกว่า (0.2857 เทียบกับ 0.2424)

**6) การพยากรณ์ฝนตกหนักมาก**: ทั้งสองแบบจำลองไม่สามารถพยากรณ์วันที่มีฝนตกหนักมาก (มากกว่า 90.1 มิลลิเมตร) ได้เลย ซึ่งอาจเนื่องมาจากมีตัวอย่างน้อยเกินไปในชุดข้อมูลฝึกสอน

4.2.3 การวิเคราะห์ Confusion Matrix

เมื่อวิเคราะห์ Confusion Matrix ในรายละเอียด (ไม่ได้แสดงในตาราง) พบว่า:

**1) ความสับสนในการจำแนก**: ทั้งสองแบบจำลองมักสับสนระหว่างระดับปริมาณฝนที่ใกล้เคียงกัน เช่น ระหว่างวันที่ไม่มีฝนตกกับวันที่มีฝนตกน้อย หรือระหว่างวันที่มีฝนตกปานกลางกับวันที่มีฝนตกหนัก

**2) อัตราการทำนายผิดพลาด**: แบบจำลอง SVR มีแนวโน้มที่จะทำนายว่าไม่มีฝนตกในวันที่มีฝนตกจริง (False Negative) มากกว่า LSTM ซึ่งเป็นข้อผิดพลาดที่มีผลกระทบสูงในการเตือนภัยน้ำท่วม

**3) ความไม่สมดุลของข้อมูล**: ประสิทธิภาพที่ต่ำในการพยากรณ์วันที่มีฝนตกหนักและฝนตกหนักมากสะท้อนให้เห็นถึงปัญหาความไม่สมดุลของข้อมูล เนื่องจากมีตัวอย่างน้อยเกินไปในชุดข้อมูลฝึกสอน

ผลการวิเคราะห์ Confusion Matrix แสดงให้เห็นว่า แบบจำลอง LSTM มีความสามารถในการจำแนกระดับปริมาณน้ำฝนได้ดีกว่า SVR โดยรวม โดยเฉพาะในระดับฝนตกน้อยและฝนตกปานกลาง ซึ่งเป็นระดับที่พบบ่อยที่สุดในชุดข้อมูล อย่างไรก็ตาม ทั้งสองแบบจำลองยังมีข้อจำกัดในการพยากรณ์วันที่มีฝนตกหนักและฝนตกหนักมาก ซึ่งเป็นเหตุการณ์ที่สำคัญในการเตือนภัยน้ำท่วม

## **4.3 ผลการวิเคราะห์ค่าความคลาดเคลื่อนของโมเดล**

นอกจากการประเมินความสามารถในการจำแนกระดับปริมาณน้ำฝนแล้ว ยังได้ทำการวิเคราะห์ค่าความคลาดเคลื่อนของการพยากรณ์ค่าปริมาณน้ำฝนเชิงปริมาณ เพื่อประเมินความแม่นยำในการทำนายค่าจริงของปริมาณน้ำฝน

### 4.3.1 พารามิเตอร์ของโมเดล

**ตารางที่ 6** ตารางแสดงพารามิเตอร์โมเดล LSTM

|  |  |  |  |
| --- | --- | --- | --- |
| แบบจำลอง | ชั้นของแบบจำลอง | Hyper-Parameter | Activation Function |
| LSTM | LSTM layer 1 | 50 | Tanh |
| FullyConnectedLayer | 1 |
| Regression Layer | - |

**ตารางที่ 7** ตารางแสดงพารามิเตอร์โมเดล SVR

|  |  |  |  |
| --- | --- | --- | --- |
| แบบจำลอง | Kernel | BoxConstraint | Epsilon |
| SVR | Radial basis function | 180000 | 0.05 |

4.3.2 ผลลัพธ์ค่าความคลาดเคลื่อน

**ตารางที่ 8** ผลลัพธ์ค่าความคลาดเคลื่อนของชุดข้อมูลทดสอบ

|  |  |  |  |
| --- | --- | --- | --- |
| ชุดข้อมูลฝึก | | | |
| แบบจำลอง | MAE | RMSE | R² |
| LSTM | 0.5518 | 1.9942 | 0.9839 |
| SVR | 1.9942 | 2.7251 | 0.9698 |
| ชุดข้อมูลทดสอบ | | | |
| แบบจำลอง | MAE | RMSE | R² |
| LSTM | 11.4488 | 19.4633 | -0.0832 |
| SVR | 19.0596 | 32.1085 | -1.9577 |

ตารางที่ 8 แสดงค่าความคลาดเคลื่อนของแบบจำลอง LSTM และ SVR ทั้งในชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบ จากผลการวิเคราะห์พบว่า:

**1) ประสิทธิภาพในชุดข้อมูลฝึกสอน**: ทั้งสองแบบจำลองแสดงประสิทธิภาพสูงในชุดข้อมูลฝึกสอน โดยแบบจำลอง LSTM มีค่าความคลาดเคลื่อนต่ำกว่า (MAE = 0.5518, RMSE = 1.9942) และมีค่า R² สูงกว่า (0.9839) เมื่อเทียบกับแบบจำลอง SVR (MAE = 1.9942, RMSE = 2.7251, R² = 0.9698)

**2) ประสิทธิภาพในชุดข้อมูลทดสอบ**: ทั้งสองแบบจำลองมีประสิทธิภาพลดลงอย่างมากในชุดข้อมูลทดสอบ โดยมีค่าความคลาดเคลื่อนสูงขึ้นและค่า R² เป็นลบ อย่างไรก็ตาม แบบจำลอง LSTM ยังคงมีประสิทธิภาพดีกว่า SVR อย่างชัดเจน โดยมีค่า MAE และ RMSE ต่ำกว่า (11.4488 และ 19.4633 เทียบกับ 19.0596 และ 32.1085) และมีค่า R² ที่ดีกว่า (-0.0832 เทียบกับ -1.9577)

**3) ความแตกต่างของชุดข้อมูล**: ค่า R² ที่เป็นลบในชุดข้อมูลทดสอบอาจเกิดจากความแตกต่างระหว่างลักษณะของชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบ โดยเฉพาะสัดส่วนของวันที่มีฝนตก (32.91% ในชุดฝึกสอน เทียบกับ 58.16% ในชุดทดสอบ) และความรุนแรงของฝนที่อาจแตกต่างกันตามฤดูกาล

### 4.3.3 การวิเคราะห์รูปแบบความคลาดเคลื่อน

เมื่อวิเคราะห์รูปแบบความคลาดเคลื่อนในรายละเอียด พบว่า

**1) ความคลาดเคลื่อนตามปริมาณน้ำฝน**: ทั้งสองแบบจำลองมีความคลาดเคลื่อนสูงในวันที่มีปริมาณน้ำฝนมาก โดยเฉพาะในวันที่มีฝนตกหนักและฝนตกหนักมาก ซึ่งสอดคล้องกับผลการวิเคราะห์ Confusion Matrix ที่แสดงประสิทธิภาพต่ำในการพยากรณ์ระดับปริมาณฝนที่สูง

**2) ความคลาดเคลื่อนตามฤดูกาล**: ความคลาดเคลื่อนมีแนวโน้มสูงขึ้นในช่วงฤดูฝน (พฤษภาคม-ตุลาคม) ซึ่งมีความผันผวนของปริมาณน้ำฝนสูง และต่ำลงในช่วงฤดูแล้ง (พฤศจิกายน-เมษายน)

**3) ความคลาดเคลื่อนในการพยากรณ์วันที่เริ่มมีฝนตก**: ทั้งสองแบบจำลองมีความคลาดเคลื่อนสูงในการพยากรณ์วันแรกที่เริ่มมีฝนตกหลังจากช่วงที่ไม่มีฝนตกติดต่อกันหลายวัน ซึ่งอาจเกิดจากการที่แบบจำลองให้น้ำหนักกับข้อมูลในช่วงก่อนหน้ามากเกินไป

**4.4 การเปรียบเทียบประสิทธิภาพระหว่างแบบจำลอง**

การเปรียบเทียบประสิทธิภาพระหว่างแบบจำลอง LSTM และ SVR ในการพยากรณ์ปริมาณน้ำฝนในจังหวัดนครพนม สามารถสรุปได้ตามประเด็นต่างๆ ดังนี้:

### 4.4.1 ด้านความแม่นยำในการพยากรณ์

**1) ความถูกต้องโดยรวม** แบบจำลอง LSTM มีค่าความถูกต้อง (Accuracy)

สูงกว่า SVR ทั้งในการประเมินด้วย 5-Fold Cross Validation (76.28% เทียบกับ 75.32%) และการประเมินด้วย Confusion Matrix (37.44% เทียบกับ 30.61%)

**2) แม่นยำในการจำแนกระดับปริมาณฝน** LSTM มีความสามารถในการจำแนกระดับปริมาณฝนได้ดีกว่า SVR โดยเฉพาะในระดับฝนตกน้อยและฝนตกปานกลาง ซึ่งพบบ่อยที่สุด

ในชุดข้อมูล

**3) ความคลาดเคลื่อนในการพยากรณ์เชิงปริมาณ** LSTM มีค่าความคลาดเคลื่อนต่ำกว่า SVR ทั้งในชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบ โดยในชุดข้อมูลทดสอบ LSTM มีค่า MAE และ RMSE เท่ากับ 11.4488 และ 19.4633 เทียบกับ 19.0596 และ 32.1085 ของ SVR

### 4.4.2 ด้านความเสถียรในการพยากรณ์

**1) ความสม่ำเสมอระหว่าง Folds** LSTM มีความสม่ำเสมอในผลลัพธ์ระหว่าง Folds มากกว่า SVR โดยมีค่าส่วนเบี่ยงเบนมาตรฐานของความถูกต้องเท่ากับ 0.014 เทียบกับ 0.022 ของ SVR

**2) ความทนทานต่อ Overfitting** แม้ว่าทั้งสองแบบจำลองจะมีปัญหา Overfitting แต่ LSTM มีความทนทานมากกว่า โดยมีค่า R² ในชุดข้อมูลทดสอบเท่ากับ -0.0832 เทียบกับ -1.9577 ของ SVR

**3) ความสามารถในการปรับตัวกับข้อมูลที่แตกต่าง** LSTM มีความสามารถ

ในการปรับตัวกับข้อมูลที่มีลักษณะแตกต่างจากชุดข้อมูลฝึกสอนได้ดีกว่า SVR โดยเฉพาะในช่วงที่มีความผันผวนของปริมาณฝนสูง

### 4.4.3 ด้านความสามารถในการจับความสัมพันธ์ในข้อมูล

**1) ความสามารถในการเรียนรู้ข้อมูลลำดับเวลา** LSTM มีความสามารถโดดเด่นในการเรียนรู้รูปแบบระยะยาวในข้อมูลลำดับเวลา ทำให้สามารถพยากรณ์ปริมาณฝนได้แม่นยำกว่า SVR ซึ่งพิจารณาเฉพาะข้อมูล ณ เวลาปัจจุบัน

**2) ความสามารถในการจับความสัมพันธ์ไม่เชิงเส้น** แม้ว่าทั้ง LSTM และ SVR จะมีความสามารถในการจับความสัมพันธ์ไม่เชิงเส้น แต่ LSTM สามารถเรียนรู้ความสัมพันธ์ที่ซับซ้อนกว่าได้ดีกว่า โดยเฉพาะในบริบทของข้อมูลอนุกรมเวลา

4.4.4 ด้านข้อจำกัดของแบบจำลอง

**1) ความสามารถในการพยากรณ์ฝนตกหนักและฝนตกหนักมาก** ทั้งสองแบบจำลองมีข้อจำกัดในการพยากรณ์วันที่มีฝนตกหนักและฝนตกหนักมาก ซึ่งเป็นเหตุการณ์ที่สำคัญในการเตือนภัยน้ำท่วม อย่างไรก็ตาม LSTM ยังคงมีประสิทธิภาพดีกว่า SVR ในการพยากรณ์ฝนตกหนัก

**2) ปัญหาข้อมูลไม่สมดุล** ทั้งสองแบบจำลองได้รับผลกระทบจากปัญหาข้อมูลไม่สมดุล โดยเฉพาะการมีตัวอย่างน้อยเกินไปสำหรับวันที่มีฝนตกหนักและฝนตกหนักมาก แม้จะมีการใช้เทคนิคต่างๆ เพื่อแก้ไขปัญหานี้

**3) ความแตกต่างระหว่างชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบ** ทั้งสองแบบจำลองมีประสิทธิภาพลดลงอย่างมากในชุดข้อมูลทดสอบ ซึ่งอาจเกิดจากความแตกต่างของลักษณะข้อมูลระหว่างชุดข้อมูลฝึกสอน (มีวันที่ไม่มีฝนตก 67.09%) และชุดข้อมูลทดสอบ (มีวันที่มีฝนตก 58.16%)

# 

# **บทที่ 5**

# **สรุปผล อภิปรายผล และข้อเสนอแนะ**

ในงานวิจัยเรื่อง แบบจำลองการพยากรณ์ปริมาณน้ำฝนด้วยเทคนิคเรียนรู้แบบเครื่อง กรณีศึกษา : จังหวัดนครพนม เป็นการนำเสนอสรุปผลการวิจัย การอภิปรายผลการวิจัยในภาพรวม และข้อเสนอแนะสำหรับการพัฒนาและประยุกต์ใช้แบบจำลองการพยากรณ์ปริมาณน้ำฝนในอนาคต โดยเชื่อมโยงกับวัตถุประสงค์ของงานวิจัยและองค์ความรู้ที่ได้จากการศึกษา รวมถึงประโยชน์ในการนำไปใช้งานจริงและแนวทางการวิจัยต่อยอดในอนาคต

**5.1 สรุปผลการวิจัย**

การวิจัยนี้มุ่งเน้นการพัฒนาและเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์ปริมาณน้ำฝนโดยใช้เทคนิค Machine Learning กรณีศึกษาจังหวัดนครพนม โดยเปรียบเทียบแบบจำลอง Long Short-Term Memory (LSTM) และ Support Vector Regression (SVR) สามารถสรุปผลการวิจัยตามวัตถุประสงค์ได้ดังนี้

### 5.1.1 การพัฒนาแบบจำลองการพยากรณ์ปริมาณน้ำฝนที่มีความแม่นยำ

ในการพัฒนาแบบจำลองการพยากรณ์ปริมาณน้ำฝนรายวันสำหรับจังหวัดนครพนม ได้ดำเนินการตามขั้นตอนอย่างเป็นระบบ เริ่มจากการรวบรวมและเตรียมข้อมูลอุตุนิยมวิทยาจากกรมอุตุนิยมวิทยาจังหวัดนครพนมย้อนหลัง 2 ปี 8 เดือน (ตั้งแต่ 1 มกราคม 2565 จนถึง 7 กันยายน 2567) ทำการวิเคราะห์ข้อมูล ทำความสะอาดข้อมูล และแปลงข้อมูลให้เหมาะสมก่อนนำไปใช้พัฒนาแบบจำลอง

การพัฒนาแบบจำลอง LSTM ได้ออกแบบโครงสร้างที่ประกอบด้วยชั้น LSTM ที่มี 70 หน่วย ตามด้วยชั้นเชื่อมต่อเต็มรูปแบบและชั้นถดถอย ใช้ข้อมูลตัวแปรทางอุตุนิยมวิทยา 10 ตัวแปร ในรูปแบบลำดับเวลา (Time Sequence) ย้อนหลัง 7 วัน และปรับแต่งพารามิเตอร์ต่างๆ อย่างเหมาะสม ขณะที่แบบจำลอง SVR ใช้ Radial Basis Function Kernel ด้วยค่า BoxConstraint เท่ากับ 180,000 และค่า Epsilon เท่ากับ 0.05

ผลการพัฒนาแบบจำลองพบว่า แบบจำลอง LSTM ให้ค่าความถูกต้อง (Accuracy) ในการเปรียบเทียบกับค่าจริงมากที่สุดที่ 76.28% จากการประเมินด้วย 5-Fold Cross Validation ขณะที่แบบจำลอง SVR ให้ค่าความถูกต้องสูงสุดที่ 75.32% ซึ่งบรรลุวัตถุประสงค์ที่ตั้งไว้คือการพัฒนาแบบจำลองที่มีความแม่นยำไม่ต่ำกว่า 75%

### 5.1.2 การเปรียบเทียบประสิทธิภาพของแบบจำลอง LSTM และ SVR

ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองทั้งสองพบว่า แบบจำลอง LSTM มีประสิทธิภาพสูงกว่า SVR ในทุกมิติของการวัดผล โดยมีค่าความถูกต้องสูงกว่าทั้งในการประเมินด้วย 5-Fold Cross Validation (76.28% เทียบกับ 75.32%) และการประเมินด้วย Confusion Matrix (37.44% เทียบกับ 30.61%) นอกจากนี้ LSTM ยังมีค่าความแม่นยำ (Precision) ความครบถ้วน (Recall) และ F1-Score สูงกว่า SVR

เมื่อพิจารณาความสามารถในการจำแนกระดับปริมาณฝนตามเกณฑ์ของกรมอุตุนิยมวิทยา พบว่า LSTM มีประสิทธิภาพสูงกว่า SVR ในการจำแนกระดับฝนตกน้อยและฝนตกปานกลาง ซึ่งพบบ่อยที่สุดในชุดข้อมูล แต่ทั้งสองแบบจำลองยังมีข้อจำกัดในการพยากรณ์วันที่มีฝนตกหนักและฝนตกหนักมาก

ในด้านความคลาดเคลื่อนของการพยากรณ์เชิงปริมาณ LSTM มีค่าความคลาดเคลื่อนต่ำกว่า SVR ทั้งในชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบ โดยในชุดข้อมูลทดสอบ LSTM มีค่า MAE และ RMSE เท่ากับ 11.4488 และ 19.4633 เทียบกับ 19.0596 และ 32.1085 ของ SVR และมีค่า R² ที่ดีกว่า (-0.0832 เทียบกับ -1.9577) แม้ว่าจะยังเป็นค่าลบในชุดข้อมูลทดสอบ ซึ่งสะท้อนถึงความท้าทายในการพยากรณ์ปริมาณน้ำฝนที่มีความผันผวนสูง

### 5.1.3 การวิเคราะห์ปัจจัยที่มีอิทธิพลต่อปริมาณน้ำฝนและข้อจำกัดของแบบจำลอง

การวิเคราะห์ความสำคัญของตัวแปรนำเข้าพบว่า ปัจจัยทางอุตุนิยมวิทยาที่มีอิทธิพลสูงสุดต่อปริมาณน้ำฝนในจังหวัดนครพนม

ข้อจำกัดสำคัญของแบบจำลองที่พบในงานวิจัยนี้ ได้แก่

**1) ข้อมูลไม่สมดุล** การมีตัวอย่างน้อยเกินไปสำหรับวันที่มีฝนตกหนักและฝนตกหนักมาก ทำให้แบบจำลองมีความแม่นยำต่ำในการพยากรณ์ระดับปริมาณฝนที่สูง

**2) ความผันผวนของปริมาณฝน** ความผันผวนสูงของปริมาณฝนในจังหวัดนครพนม โดยเฉพาะในช่วงฤดูฝน ทำให้การพยากรณ์มีความท้าทายมากขึ้น

**3) ความแตกต่างระหว่างชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบ** ความแตกต่างของลักษณะข้อมูลระหว่างชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบ ทั้งในด้านสัดส่วนของวันที่มีฝนตกและความรุนแรงของฝน ส่งผลให้ประสิทธิภาพของแบบจำลองลดลงอย่างมากในชุดข้อมูลทดสอบ

**4) ข้อจำกัดด้านปริมาณข้อมูล** การใช้ข้อมูลย้อนหลังเพียง 2 ปี 8 เดือน อาจไม่เพียงพอสำหรับการเรียนรู้รูปแบบที่ซับซ้อนของปริมาณฝนในระยะยาว โดยเฉพาะรูปแบบตามฤดูกาลที่อาจมีความแปรปรวนระหว่างปี

**5.2 อภิปรายผล**

ผลการวิจัยพบว่า แบบจำลอง LSTM มีประสิทธิภาพสูงกว่า SVR ในการพยากรณ์ปริมาณน้ำฝนในจังหวัดนครพนม ซึ่งสอดคล้องกับงานวิจัยของ Fan et al. (2020) และ รักษ์คณา ภูสีเขียว (2560) ที่พบว่า LSTM มีความแม่นยำสูงกว่าแบบจำลองอื่นๆ ในการพยากรณ์ปริมาณฝน การที่ LSTM มีประสิทธิภาพสูงกว่า SVR สามารถอธิบายได้ด้วยคุณสมบัติเฉพาะของ LSTM ในการเรียนรู้ความสัมพันธ์ระยะยาวในข้อมูลอนุกรมเวลา และความสามารถในการจดจำรูปแบบที่ซับซ้อนของปริมาณฝนที่เปลี่ยนแปลงตามฤดูกาล

อย่างไรก็ตาม ความแม่นยำในการจำแนกระดับปริมาณฝนของทั้งสองแบบจำลองยังไม่สูงมากนัก (37.44% สำหรับ LSTM และ 30.61% สำหรับ SVR) เมื่อเทียบกับงานวิจัยอื่นๆ เช่น Aderyani et al. (2022) ที่รายงานความแม่นยำสูงถึง 70-80% ในการพยากรณ์ปริมาณฝนระยะสั้น ความแตกต่างนี้อาจเกิดจากหลายปัจจัย ได้แก่

**1) ลักษณะภูมิอากาศของพื้นที่ศึกษา** จังหวัดนครพนมมีลักษณะภูมิอากาศแบบมรสุมเขตร้อน ซึ่งมีความผันผวนของปริมาณฝนสูง โดยเฉพาะในช่วงฤดูฝน ทำให้การพยากรณ์มีความท้าทายมากกว่าพื้นที่ที่มีปริมาณฝนสม่ำเสมอ

**2) ปริมาณและคุณภาพของข้อมูล** งานวิจัยนี้ใช้ข้อมูลย้อนหลังเพียง 2 ปี 8 เดือน ซึ่งอาจไม่เพียงพอสำหรับการเรียนรู้รูปแบบที่ซับซ้อนของปริมาณฝนในระยะยาว เมื่อเทียบกับงานวิจัยอื่นที่ใช้ข้อมูลย้อนหลัง 5-10 ปี หรือใช้ข้อมูลที่มีความละเอียดสูงกว่า เช่น ข้อมูลรายชั่วโมง

**3) เกณฑ์การประเมินที่แตกต่างกัน** งานวิจัยนี้ใช้เกณฑ์การจำแนกระดับปริมาณฝน 5 ระดับตามเกณฑ์ของกรมอุตุนิยมวิทยา ซึ่งมีความละเอียดสูงกว่าการแบ่งเพียง 2-3 ระดับในงานวิจัยอื่น ทำให้การประเมินความแม่นยำมีความท้าทายมากขึ้น

**4) ข้อมูลไม่สมดุล** การมีวันที่ไม่มีฝนตกมากถึง 67.09% ในชุดข้อมูลฝึกสอน ทำให้แบบจำลองมีแนวโน้มที่จะเรียนรู้การพยากรณ์วันที่ไม่มีฝนตกได้ดีกว่า แต่มีความแม่นยำต่ำในการพยากรณ์วันที่มีฝนตก โดยเฉพาะฝนตกหนักและฝนตกหนักมาก ซึ่งมีความสำคัญในการเตือนภัยน้ำท่วม

**5.3 ข้อเสนอแนะ**

จากผลการวิจัยและข้อจำกัดที่พบ สามารถเสนอแนะแนวทางการพัฒนาและประยุกต์ใช้แบบจำลองการพยากรณ์ปริมาณน้ำฝนในอนาคตได้ดังนี้

### 5.3.1 ข้อเสนอแนะสำหรับการนำผลการวิจัยไปใช้ประโยชน์

**1) การนำไปใช้ในการวางแผนการเกษตร** แบบจำลอง LSTM ที่พัฒนาขึ้นสามารถนำไปประยุกต์ใช้ในการพยากรณ์ปริมาณน้ำฝนระยะสั้น (1-7 วัน) เพื่อสนับสนุนการตัดสินใจในภาคการเกษตรของจังหวัดนครพนม เช่น การวางแผนการเพาะปลูก การให้น้ำ และการเก็บเกี่ยวผลผลิต

**2) การพัฒนาระบบเตือนภัยน้ำท่วมและภัยแล้ง** แม้ว่าแบบจำลองจะมีข้อจำกัดในการพยากรณ์วันที่มีฝนตกหนักและฝนตกหนักมาก แต่สามารถใช้เป็นส่วนหนึ่งของระบบเตือนภัยน้ำท่วมและภัยแล้งในพื้นที่ได้ โดยอาจต้องใช้ร่วมกับแบบจำลองอื่นๆ และข้อมูลจากแหล่งอื่น เช่น ข้อมูลระดับน้ำในแม่น้ำและข้อมูลจากดาวเทียมอุตุนิยมวิทยา

**3) การบูรณาการกับระบบบริหารจัดการน้ำ** แบบจำลองที่พัฒนาขึ้นสามารถนำไปบูรณาการกับระบบบริหารจัดการน้ำของจังหวัดนครพนม เพื่อช่วยในการคาดการณ์ปริมาณน้ำที่จะไหลเข้าอ่างเก็บน้ำ การวางแผนการระบายน้ำ และการจัดสรรน้ำเพื่อการอุปโภคบริโภคและการเกษตร

**4) พัฒนาแอปพลิเคชันสำหรับเกษตรกรและประชาชน** ควรพัฒนาแอปพลิเคชันที่ใช้งานง่ายสำหรับเกษตรกรและประชาชนทั่วไป ให้สามารถเข้าถึงข้อมูลการพยากรณ์ปริมาณน้ำฝนได้ โดยอาจใช้ระบบการแจ้งเตือนล่วงหน้าเมื่อคาดการณ์ว่าจะมีฝนตกหนักหรือภาวะแห้งแล้ง

### 5.3.2 ข้อเสนอแนะสำหรับการวิจัยในอนาคต

**1) การเพิ่มปริมาณและคุณภาพของข้อมูล** ควรเพิ่มปริมาณข้อมูลที่ใช้ในการฝึกแบบจำลอง โดยใช้ข้อมูลย้อนหลังที่มากขึ้น (เช่น 5-10 ปี) และเพิ่มความละเอียดของข้อมูล (เช่น ข้อมูลรายชั่วโมงแทนรายวัน) นอกจากนี้ ควรรวบรวมข้อมูลจากหลายแหล่ง เช่น สถานีตรวจวัดอากาศอัตโนมัติในพื้นที่ใกล้เคียง เพื่อเพิ่มความครอบคลุมและความถูกต้องของข้อมูล

**2) การทดลองใช้เทคนิค Machine Learning ขั้นสูง** ควรทดลองใช้เทคนิค Machine Learning ขั้นสูงที่มีประสิทธิภาพในการจัดการกับข้อมูลอนุกรมเวลาที่มีความซับซ้อน เช่น Transformer, Attention Mechanism, และ Neural Ordinary Differential Equations (NODE) ซึ่งอาจช่วยเพิ่มความแม่นยำในการพยากรณ์ปริมาณฝน

**3) การพัฒนาแบบจำลองแบบผสม (Hybrid Models)** ควรพัฒนาแบบจำลองแบบผสมที่รวมจุดแข็งของ LSTM และโมเดลอื่นๆ เช่น LSTM-CNN ที่ใช้ CNN ในการสกัดคุณลักษณะเชิงพื้นที่และ LSTM ในการจับความสัมพันธ์เชิงเวลา หรือ LSTM-SVR ที่ใช้ LSTM ในการพยากรณ์แนวโน้มระยะยาวและ SVR ในการปรับแต่งการพยากรณ์ระยะสั้น

**4) การใช้เทคนิคในการจัดการข้อมูลไม่สมดุล** ควรทดลองใช้เทคนิคขั้นสูงในการจัดการข้อมูลไม่สมดุล เช่น SMOTE, ADASYN, Class Weighting, และ Focal Loss เพื่อเพิ่มความแม่นยำในการพยากรณ์วันที่มีฝนตกหนักและฝนตกหนักมาก ซึ่งมีความสำคัญในการเตือนภัยน้ำท่วม

การใช้ข้อมูลเชิงพื้นที่และข้อมูลจากดาวเทียม: ควรเพิ่มการใช้ข้อมูลเชิงพื้นที่ (Spatial Data) และข้อมูลจากดาวเทียมอุตุนิยมวิทยา เช่น ข้อมูลเมฆจากดาวเทียม Himawari และข้อมูลฝนจากดาวเทียม GPM (Global Precipitation Measurement) ซึ่งอาจช่วยเพิ่มความแม่นยำในการพยากรณ์ปริมาณฝนในพื้นที่ที่ไม่มีสถานีตรวจวัด

**5) การพัฒนาแบบจำลองเฉพาะพื้นที่และฤดูกาล** ควรพัฒนาแบบจำลองเฉพาะสำหรับแต่ละพื้นที่ย่อยในจังหวัดนครพนม และแบบจำลองเฉพาะสำหรับแต่ละฤดูกาล เนื่องจากรูปแบบและปัจจัยที่มีผลต่อปริมาณฝนอาจแตกต่างกันตามลักษณะภูมิประเทศและช่วงเวลาของปี

การศึกษาผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศ: ควรศึกษาผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อรูปแบบของปริมาณฝนในจังหวัดนครพนม และพัฒนาแบบจำลองที่สามารถปรับตัวกับการเปลี่ยนแปลงเหล่านี้ได้ โดยอาจใช้เทคนิค Transfer Learning หรือ Adaptive Learning

**6) การพัฒนาระบบพยากรณ์แบบเรียลไทม์** ควรพัฒนาระบบพยากรณ์ปริมาณฝนแบบเรียลไทม์ที่สามารถปรับปรุงการพยากรณ์ตามข้อมูลที่ได้รับในเวลาจริง โดยใช้เทคนิค Online Learning หรือ Incremental Learning ซึ่งจะช่วยเพิ่มความแม่นยำในการพยากรณ์ระยะสั้น

การวิจัยเชิงสหวิทยาการ: ควรส่งเสริมการวิจัยเชิงสหวิทยาการระหว่างนักวิทยาการคอมพิวเตอร์ นักอุตุนิยม

# **บรรณานุกรม**

Aderyani, F. R., Mousavi, S. J., & Jafari, F. (2022). *Short-term rainfall forecasting using machine learning-based approaches of PSO-SVR, LSTM and CNN*. Journal of Hydrology. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0022169422010332>

Hassan, M. M., Rony, M. A. T., Khan, M. A. R., Hassan, M. M., Yasmin, F., Nag, A., Zarin, T. H., Bairagi, A. K., Alshathri, S., & El-Shafai, W. (2023). Machine learning-based rainfall prediction: Unveiling insights and forecasting for improved preparedness. *IEEE Access*, 11, 1-14. <https://doi.org/10.1109/ACCESS.2023.3333876>

Hernández, E., Sanchez-Anguix, V., Julian, V., Palanca, J., & Duque, N. (2016). Rainfall prediction: A deep learning approach. International Conference on Hybrid Artificial Intelligence Systems,

Hung, N. Q., Babel, M. S., Weesakul, S., & Tripathi, N. (2009). An artificial neural network model for rainfall forecasting in Bangkok, Thailand. *Hydrology and Earth System Sciences*, 13(8), 1413-1425.

Leys, C., Klein, O., Bernard, P., & Licata, L. (2019). Detecting Outliers: Do Not Use Standard Deviation Around the Mean, Use Absolute Deviation Around the Median. *Journal of Experimental Social Psychology*, 74, 139–156.

P Sai, et al., "Meteorological Feature Selection and Transformation for Rainfall Prediction: Enhancing Model Performance through Relevant Attributes," *Journal of Climate Science*, vol. 29, no. 2, pp. 135-150, 2024.

Pan, X., Wang, Y., Li, Z., Zhang, H., & Liu, J. (2020). Water level prediction model based on GRU and CNN. *Water*, 12(1), Article 12.

Qiu, J., Zhang, Q., Zhang, D., Jiang, S., Wu, A., & Yu, B. (2020). A hybrid artificial intelligence model for predicting the strength of foam-cemented paste backfill. *Materials*, 13(5), Article 1103.

Robertson, D.E., et al. (2023). *Bias Correction for Rainfall Prediction: Adjusting Zero and Non-Zero Rainfall for Improved Forecasting Accuracy.* Journal of Hydrology.

เจนจิรา มงคลเมือง, ศศิธร พิศพล, & เฉลิมวุฒิ คาเมือง. (2564). การพยากรณ์ปริมาณน้ำฝนรายเดือนของจังหวัดบุรีรัมย์ด้วยวิธีฟังก์ชันการถ่ายโอน. *วารสารวิทยาศาสตร์ประยุกต์*, 20(1), 154-166. <https://doi.org/10.14416/j.appsci.2021.01.011>

ทิพาภรณ์ หอมดี, วชิรกรณ์ เสนาวัง, พงฬ์นธี มณีกุล, ธิปธราดล แผงนาวิน, & ภากร ครุฑแสงอนันต์. (2565). การพยากรณ์ปริมาณน้ำฝนรายเดือนในจังหวัดนครพนม ณ สถานีโครงการน้ำก่ำตอนล่าง. ใน *เอกสารการประชุมวิชาการวิศวกรรมโยธาแห่งชาติ ครั้งที่ 27*.

มะโนกิจ, เฟื่องลดา. (2562). การพยากรณ์ปริมาณน้ำฝนของประเทศไทยด้วยวิธีการเรียนรู้เชิงลึก*[วิทยานิพนธ์ปริญญาโท, จุฬาลงกรณ์มหาวิทยาลัย]. Chulalongkorn University Theses and Dissertations (Chula ETD).* <https://digital.car.chula.ac.th/chulaetd/9500>

ยุพิน ไชยสมภาร, & ทวี ชัยพิมลผลิน. (2559). การพยากรณ์ระดับน้ำโดยแบบจำลองโครงข่ายประสาทเทียมด้วยข้อมูลน้ำฝนจากแบบจำลอง WRF-ECHAM5. *วารสารมหาวิทยาลัย*

*ศรีนครินทรวิโรฒ (สาขาวิทยาศาสตร์และเทคโนโลยี)*, 9(1).

รักษ์คณา ภูสีเขียว. (2564). การพยากรณ์ปริมาณน้ำฝนระยะสั้นในบริเวณพื้นที่สนามบินสุวรรณภูมิด้วยโครงข่ายระบบประสาทแบบย้อนกลับ *[วิทยานิพนธ์ปริญญาโท, จุฬาลงกรณ์มหาวิทยาลัย]. Chulalongkorn University Theses and Dissertations (Chula ETD).* <https://digital.car.chula.ac.th/chulaetd/5602>

**ประวัติผู้เขียน**

**ชื่อ-สกุล** กฤษณกร เทพพิทักษ์

**วัน เดือน ปี เกิด** 29 กรกฎาคม 2545

**สถานที่เกิด** นครพนม

**ที่อยู่ปัจจุบัน**  75 หมู่ 1 บ้านหนองฮี ตำบลหนองฮี อำเภอปลาปาก จังหวัดนครพนม 48160

**วุฒิการศึกษา**

**วุฒิ ชื่อสถาบัน ปีที่สำเร็จการศึกษา**

มัธยมศึกษาตอนปลาย โรงเรียนพัชรกิติยาภา ๑ นครพนม 2564

มัธยมศึกษาตอนปลาย4-5 โรงเรียนเรณูนครวิทยานุกูล 2562

มัธยมศึกษาตอนต้น โรงเรียนเรณูนครวิทยานุกูล 2560