

การพยากรณ์ปริมาณน้ำฝนระยะสั้นในบริเวณพื้นที่สนามบินสุวรรณภูมิด้วยโครงข่าย
ระบบประสาทแบบย้อนกลับ

Short Term Precipitation Forecasting using Recurrent Neural Networks, a Case Study of Suvarnabhumi Airport

รักศณา ภูสีเขียว* และ สุรณพีร์ ภูมิตูฬสาร

Ragkana Phooseekhiwe and Suronapee Phoomvuthisarn*

ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

Department of Statistics, Chulalongkorn Business School, Chulalongkorn University

Received: November 16, 2021; Revised: March 31, 2022; Accepted: April 08, 2022; Published: June 29, 2022

ABSTRACT – Rainfall is one of the key important factors affecting human life. Accurate precipitation forecasting allows humans to better prepare for various activities that will happen in the future. However, in some situations, the availability of weather data is limited, making it difficult to make accurate forecasting. Currently, much research has chosen deep neural networks as an algorithm to train forecasting models. The main idea is to come up with relevant features at the architecture level. Based on this paradigm, it has been shown that the appropriate deep neural network architecture can flexibly mix and match features that are relevant in predictions. Consequently, most of the existing research then focuses on some of the techniques to improve the performance of the models without paying much attention on the issues of adding relevant features. However, when the training data is limited, deep neural networks might not scale very well, thus making it difficult to mix and match features for predictions. This imposes a research question of how effective the proposed forecasting models might be when not adding relevant features to the models. The aims of this research are to develop and compare the efficiency of various models for short-term precipitation forecasting with and without adding relevant features. The experiment consists of 2 parts: (1) The experiment by exploring whether the models with relevant featured provided can achieve higher accuracy compared to original models in equivalent environments, and (2) The experiment by comparing accuracy between 5 models (ARIMA, ARIMAX, RNN, LSTM and GRU). The weather dataset, which is very limited in quantity, used in this research was collected from Suvarnabhumi Airport. The results show that adding relevant features can enhance the forecasting performance of the model when the weather data is limited. In addition, the GRU model with relevant featured provided is the most effective prediction.

KEYWORDS: Long Short-Term Memory, Predictive analytics, Rainfall prediction, Recurrent Neural Network, Neural Network, Gated Recurrent Unit, Imbalanced data

บทคัดย่อ -- ปริมาณน้ำฝนนับเป็นปัจจัยสำคัญอย่างหนึ่งที่มีผลต่อการดำเนินชีวิตของมนุษย์ การพยากรณ์ปริมาณน้ำฝนที่มีความแม่นยำช่วยให้มนุษย์เตรียมพร้อมสำหรับกิจกรรมต่างๆ ที่จะเกิดขึ้นในอนาคตได้ดี อย่างไรก็ตามในบางสถานการณ์ ความพร้อมใช้งานของข้อมูลสภาพอากาศมีจำกัด ทำให้การพยากรณ์ปริมาณน้ำฝนอย่างแม่นยำนั้นเป็นเรื่องที่ยาก ปัจจุบันหลายๆ งานวิจัยที่เกี่ยวข้องได้เลือกโครงข่ายประสาทเทียมเชิงลึกเป็นอัลกอริทึมในการฝึกแบบจำลองเพื่อใช้ในการพยากรณ์

*Corresponding Author: ragkana.ph@gmail.com

แนวคิดหลักคือการสร้างตัวแปรคุณลักษณะ (Feature) ที่เกี่ยวข้องในระดับสถาปัตยกรรม จากหลักการนี้สถาปัตยกรรมโครงข่ายประสาทเทียมเชิงลึกที่เหมาะสมสามารถผสมผสานและจับคู่คุณลักษณะที่เกี่ยวข้องในการพยากรณ์ได้อย่างเหมาะสม ผลที่ตามมาจากงานวิจัยที่มีอยู่ส่วนใหญ่จึงมุ่งเน้นไปที่เทคนิคบางอย่างเพื่อปรับปรุงประสิทธิภาพของแบบจำลอง โดยไม่ได้ให้ความสำคัญกับการเพิ่มคุณลักษณะให้กับตัวแบบมากนัก อย่างไรก็ตามเมื่อข้อมูลการฝึกฝนมีจำนวนจำกัด โครงข่ายประสาทเทียมเชิงลึกอาจจะทำงานได้ไม่เต็มประสิทธิภาพมากนัก ทำให้การผสมผสานและจับคู่คุณลักษณะที่เกี่ยวข้องในการพยากรณ์ทำได้ไม่ดีตามไปด้วย สิ่งนี้ทำให้เกิดคำถามงานวิจัยว่าแบบจำลองการพยากรณ์ปริมาณน้ำฝนที่ได้ถูกนำเสนอมีประสิทธิภาพที่ดีเพียงพอรึหรือไม่ เมื่อไม่ได้มีการเพิ่มคุณสมบัติที่เกี่ยวข้องให้กับแบบจำลอง งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อพัฒนาและเปรียบเทียบประสิทธิภาพของแบบจำลองต่างๆ ในการพยากรณ์ปริมาณน้ำฝนสะสมในระยะสั้นที่มีและไม่มี การเพิ่มตัวแปรคุณสมบัติที่เกี่ยวข้อง โดยได้แบ่งการทดลองออกเป็น 2 ส่วนเพื่อวัดประสิทธิภาพ คือ 1) การเปรียบเทียบประสิทธิภาพของตัวแบบที่มีการเพิ่มตัวแปรคุณลักษณะที่เกี่ยวข้องว่ามีความถูกต้องแม่นยำขึ้นหรือไม่เมื่อเทียบกับแบบจำลองที่ไม่ได้มีการเพิ่มตัวแปรคุณลักษณะในสภาพแวดล้อมที่เทียบเท่ากัน และ 2) การเปรียบเทียบประสิทธิภาพในการพยากรณ์ปริมาณน้ำฝนสะสมของแบบจำลองที่สนใจศึกษา ได้แก่ ARIMA ARIMAX RNN LSTM และ GRU ข้อมูลที่นำมาใช้ในงานวิจัยนี้เป็นข้อมูลสภาพอากาศและปริมาณน้ำฝนสะสมที่รวบรวมมาจากพื้นที่สนามบินสุวรรณภูมิ จากผลการศึกษาทั้ง 2 ส่วนพบว่า การเพิ่มตัวแปรคุณลักษณะสามารถเพิ่มประสิทธิภาพการพยากรณ์ให้กับตัวแบบได้ในกรณีที่ข้อมูลนำมาฝึกฝนตัวแบบมีจำนวนจำกัด โดย แบบจำลอง GRU ให้ประสิทธิภาพในการพยากรณ์มากที่สุด

คำสำคัญ: แบบจำลองหน่วยความจำระยะสั้นแบบยาว, การวิเคราะห์เชิงทำนาย, การพยากรณ์ปริมาณน้ำฝน, โครงข่ายระบบประสาทแบบย้อนกลับ, เครือข่ายระบบประสาทเทียม, โครงข่ายประตูกลับ, ชุดข้อมูลไม่สมดุล

1. บทนำ

1.1 ที่มาและความสำคัญ

สภาพอากาศถือเป็นอีกหนึ่งสิ่งที่มีความเกี่ยวข้องกับมนุษย์มาตั้งแต่สมัยอดีต เพราะไม่ว่าจะเป็นการใช้ชีวิตประจำวัน การเดินทาง การประกอบอาชีพ (เช่น การทำการเกษตร ประมง เป็นต้น) หรือแม้แต่ใช้ในการตัดสินใจในการบริหารจัดการทรัพยากรระดับประเทศ ต่างก็มีความเกี่ยวข้องกับสภาพอากาศทั้งสิ้น แต่การจะพยากรณ์ปริมาณน้ำฝนอย่างแม่นยำ และมีประสิทธิภาพยังคงเป็นปัญหาสำคัญที่รัฐบาลหรือแม้แต่ภาคธุรกิจยังคงให้ความสนใจ เพราะนอกจากปัจจัยที่จะทำให้เกิดปรากฏการณ์ฝนตกนั้นจะมีความซับซ้อนแล้ว ตัวปรากฏการณ์ฝนตกเองก็เหมือนเป็นปรากฏการณ์แบบสุ่มที่ไม่อาจคาดเดาได้เช่นกัน เพื่อที่จะสามารถคาดการณ์หรือพยากรณ์ปริมาณน้ำฝนที่จะตกลงมาในช่วงต่างๆ ได้ จึงได้มีการนำเอาศาสตร์ต่างๆ เข้ามาช่วย

ในเริ่มแรก แบบจำลองที่มักถูกนำมาใช้ในการพยากรณ์ข้อมูลปริมาณน้ำฝนก็คือแบบจำลองอาร์มา(ARIMA) ดังเช่นในงานวิจัยของ (Wang et al., 2013) ที่มีการนำตัวแบบ SARIMA มาใช้ในการวิเคราะห์และพยากรณ์ กับงานวิจัยของ (Narayanan et al., 2013) ก็ได้ใช้ตัวแบบจำลอง ARIMA ในการพยากรณ์ปริมาณน้ำฝนในแต่ละสถานี พบว่าตัวแบบ ARIMA นั้นไม่ได้ตอบโจทย์สำหรับทุกสถานี เนื่องจากตัวแบบจำลอง ARIMA นั้นเป็นตัวแบบจำลองที่พิจารณาโดยใช้ตัวแปรเพียงตัวแปรเดียว แต่ในความเป็นจริง ปัจจัยหรือตัวแปรที่ส่งผลกับปริมาณน้ำฝนที่เกิดขึ้นในแต่ละช่วงเวลานั้น ไม่ได้มีเพียงปัจจัยเดียว เพื่อให้สามารถนำปัจจัยร่วมอื่นๆ มาพิจารณาพร้อมกับปริมาณน้ำฝนสะสมได้ ในงานวิจัยของ (Jalalkamali et al., 2015) จึงได้เลือกใช้แบบจำลอง ARIMAX ในการพยากรณ์ปริมาณน้ำฝนเพื่อคาดการณ์ปัญหาภัยแล้ง

ปัจจุบัน หลายๆ งานวิจัยที่เกี่ยวข้องกับการพยากรณ์อากาศหรือการพยากรณ์ปริมาณน้ำฝน ได้เลือกใช้โครงข่ายประสาท

เทียมเข้ามาช่วยในการพยากรณ์ อย่างเช่นในงานวิจัยของ (Hung et al., 2009) ที่ได้มีการนำโครงข่ายประสาทเทียม (ANN) เข้ามาช่วยในการวิเคราะห์นั้น ทำให้สามารถเพิ่มตัวแปรหรือคุณลักษณะด้านสภาพอากาศอื่นๆ เข้าไปเพื่อเพิ่มประสิทธิภาพในการพยากรณ์ ทั้งนี้ การสร้างตัวแบบโดยอาศัยโครงข่ายประสาทเทียมนั้นก็มีเทคนิคหรือวิธีการแตกแขนงแยกย่อยลงไปอีกมากมาย แต่ในการจัดการกับข้อมูลภูมิอากาศนั้น การใช้โครงข่ายระบบประสาทแบบย้อนกลับ (RNNs) จะเป็นตัวเลือกที่เหมาะสมที่สุดเนื่องจากโครงข่ายดังกล่าวนั้นถูกพัฒนามาเพื่อจัดการกับข้อมูลที่มีลักษณะเป็นลำดับ หรืออนุกรมเวลา

ในงานวิจัยของ (Poornima & Pushpalatha, 2019) ได้ทำการพัฒนาแบบจำลอง Intensified LSTM และทำการเปรียบเทียบประสิทธิภาพของแบบจำลองดังกล่าวกับแบบจำลองอาร์มาและโครงข่ายระบบประสาทแบบย้อนกลับ ซึ่งพบว่าแบบจำลองที่เสนอนั้นมีความแม่นยำในการพยากรณ์มากกว่าอีก 2 แบบจำลอง ต่อมา (Yang et al., 2020) ได้ทดลองทำการเปรียบเทียบประสิทธิภาพของแบบจำลองหน่วยความจำระยะสั้นแบบยาวกับโครงข่ายประตูกลับ พบว่า นอกจากประสิทธิภาพของแบบจำลองทั้ง 2 จะดีหรือไม่ขึ้นอยู่กับจำนวนข้อมูลแล้วยังขึ้นกับและประเภทของข้อมูลที่น่าวิเคราะห์ด้วย เช่นเดียวกับในงานวิจัยของ (Aurnhammer & Frank, 2019) และ (Shewalkar, 2019) ได้ทำการศึกษาเปรียบเทียบประสิทธิภาพของโครงข่ายระบบประสาทแบบย้อนกลับกับชุดข้อมูลที่มีความหลากหลาย ซึ่งพบว่า แบบจำลอง GRU และ LSTM นั้น มีความเหมาะสมในการคาดการณ์หรือพยากรณ์กับชุดข้อมูลหลายๆประเภท แต่ทั้งนี้ ขึ้นอยู่กับตัวข้อมูลที่จะนำเข้ามาด้วยเช่นกัน อย่างไรก็ตาม งานวิจัยที่กล่าวมาข้างต้นมุ่งเน้นในเรื่องการปรับปรุงอัลกอริทึมของตัวแบบจำลอง จึงไม่ได้ทำการทดลองเกี่ยวกับการปรับปรุง แก๊ซข้อมูล หรือเพิ่มคุณลักษณะใดๆ เพื่อเพิ่มประสิทธิภาพในการพยากรณ์

งานวิจัยนี้จึงมุ่งเน้นการพัฒนาและวิเคราะห์เปรียบเทียบประสิทธิภาพในการพยากรณ์ปริมาณน้ำฝนสะสมล่วงหน้าในระยะสั้นร่วมกับการใช้ข้อมูลหลายตัวแปร โดยตัวแบบที่เหมาะสมจะมาจากผลการศึกษาและเปรียบเทียบใน 2 ประเด็นหลักคือ 1) การเปรียบเทียบประสิทธิภาพในการพยากรณ์ของตัวแบบที่ใช้ชุดข้อมูลที่แตกต่างกัน ระหว่างชุดข้อมูลผ่านการเพิ่มคุณลักษณะพิเศษกับชุดข้อมูลดั้งเดิม 2) เปรียบเทียบ

ประสิทธิภาพในการพยากรณ์ในช่วงเวลาถัดไปของโครงข่ายระบบประสาทแบบย้อนกลับที่สนใจศึกษา กับแบบจำลอง ARIMA และแบบจำลอง ARIMAX ซึ่งเป็นแบบจำลองที่เป็นที่นิยมทางสถิติในการพยากรณ์ข้อมูลประเภทอนุกรมเวลา และสำหรับโครงข่ายระบบประสาทแบบย้อนกลับที่สนใจนำมาศึกษาเปรียบเทียบ จะมีทั้งหมด 3 ตัว ได้แก่ โครงข่ายระบบประสาทแบบย้อนกลับ (RNN) แบบจำลองหน่วยความจำระยะสั้นแบบยาว (Long short-term memory) และโครงข่ายประตูกลับ (Gated Recurrent Unit)

งานวิจัยนี้ได้ใช้ข้อมูลสภาพอากาศรายชั่วโมงจากสถานีตรวจวัดสภาพอากาศสนามบินนานาชาติสุวรรณภูมิ จังหวัดสมุทรปราการ (ข้อมูลดังกล่าวได้รับการสนับสนุนจากกรมอุตุนิยมวิทยา) ชุดข้อมูลดังกล่าวมีการเก็บสะสมตั้งแต่ปี ค.ศ. 2016 ถึงปี ค.ศ. 2020 โดยเราจะนำชุดข้อมูลนี้มาสร้างโครงข่ายระบบประสาทแบบย้อนกลับ เพื่อพยากรณ์ปริมาณน้ำฝนในอีก 3 ชั่วโมงข้างหน้า และเพื่อเพิ่มประสิทธิภาพในการพยากรณ์ปริมาณน้ำฝน จึงจะทำการเพิ่มคุณลักษณะทางสถิติบางอย่างเข้าไปก่อนที่จะนำข้อมูลเข้าไปวิเคราะห์ในโครงข่ายระบบประสาทแบบย้อนกลับ พร้อมทั้งเปรียบเทียบค่าความแม่นยำในการพยากรณ์ที่ได้จากการพัฒนาตัวแบบต่างๆ

ทั้งนี้ จากผลการทดลองเปรียบเทียบประสิทธิภาพในการพยากรณ์ของตัวแบบที่ใช้ชุดข้อมูลที่แตกต่างกัน ระหว่างชุดข้อมูลผ่านการเพิ่มคุณลักษณะพิเศษทางสถิติกับชุดข้อมูลดั้งเดิม พบว่า การเพิ่มคุณลักษณะทางสถิติให้กับข้อมูลช่วยให้ประสิทธิภาพในการพยากรณ์ของแบบจำลองเพิ่มขึ้นกว่าการใช้เพียงคุณลักษณะดั้งเดิม และในบรรดาแบบจำลองทั้งหมดที่นำมาศึกษาเปรียบเทียบประสิทธิภาพในการพยากรณ์ชุดข้อมูลสภาพอากาศชุดนี้ พบว่า แบบจำลองโครงข่ายประตูกลับ (GRU) เป็นแบบจำลองที่มีประสิทธิภาพในการพยากรณ์ชุดข้อมูลนี้มากที่สุด

ซึ่งจากผลการพัฒนาและวิเคราะห์เปรียบเทียบประสิทธิภาพในการพยากรณ์ปริมาณน้ำฝนระยะสั้น องค์ความรู้ใหม่ที่ได้จากงานวิจัยนี้ก็คือการเพิ่มตัวแปรคุณลักษณะสามารถเพิ่มประสิทธิภาพการพยากรณ์ให้กับตัวแบบโครงข่ายระบบประสาทแบบย้อนกลับที่ถูกนำมาใช้อย่างแพร่หลายในปัจจุบันได้ ในกรณีที่ข้อมูลที่น่ามาฝึกฝนตัวแบบมีจำนวนจำกัด ทางผู้วิจัยหวังว่างานวิจัยฉบับนี้จะเป็นประโยชน์ และแนวทางให้กับผู้ที่ต้องการพยากรณ์ ศึกษา หรือต่อยอดเกี่ยวกับการเพิ่มประสิทธิภาพให้กับ

แบบจำลองในการพยากรณ์ปริมาณน้ำฝน หรือสภาพอากาศอื่นๆ แต่มีข้อมูลสภาพอากาศที่ค่อนข้างจำกัดในอนาคตต่อไป

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

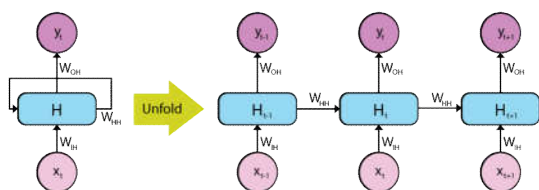
2.1 แบบจำลองอนุกรมเวลา (Time series)

แบบจำลอง ARIMA เป็นหนึ่งในแบบจำลองที่ใช้ในการวิเคราะห์ข้อมูลอนุกรมเวลาแบบตัวแปรเดี่ยว(Univariate variable) ซึ่งถูกต้องและมีความแม่นยำสูงโดยพิจารณาจากการวัดค่าความคลาดเคลื่อนของการพยากรณ์ เป็นการสร้างตัวแบบจากข้อมูลที่เก็บรวบรวมในอดีตตามลำดับเวลาอย่างต่อเนื่องเพื่อกำหนดรูปแบบของข้อมูลและพยากรณ์ข้อมูลในอนาคต โดยข้อมูลอนุกรมเวลาที่จะนำมาวิเคราะห์ได้นั้น ต้องมีความคงที่ (Stationary) และไม่มีแนวโน้ม(Trend) องค์ประกอบของแบบจำลอง ARIMA จะประกอบไปด้วยพารามิเตอร์สำคัญ 3 ตัว คือ p , d และ q

2.2 โครงข่ายระบบประสาทแบบย้อนกลับ (Recurrent

Neural Networks: RNN)

โครงข่ายระบบประสาทแบบย้อนกลับ (RNN) คือ โครงข่ายประสาทเทียม (Artificial Neuron Network : ANN) รูปแบบหนึ่ง ที่ออกแบบมาเพื่อแก้ปัญหาเกี่ยวกับข้อมูลประเภทอนุกรมเวลา (Time Series data) หลักการของ RNN คือการปรับรูปแบบของโครงข่ายประสาทเทียมแบบเดิม เพื่อให้มีชั้นซ่อน (Hidden state) ในการจดจำความรู้หรือข้อมูลก่อนหน้า มารวมเข้ากับข้อมูลตัวใหม่ที่เข้ามา (Input data) เพื่อใช้ในการทำนายหรือพยากรณ์

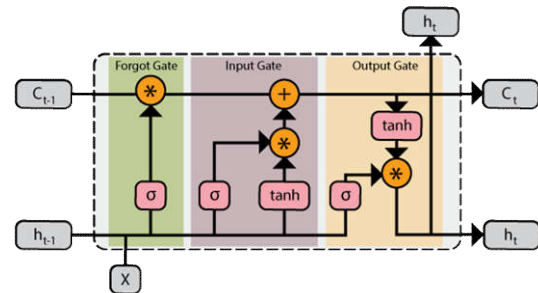


รูปที่ 1. โครงสร้างของโครงข่ายระบบประสาทแบบย้อนกลับ

2.3 แบบจำลองหน่วยความจำระยะสั้นแบบยาว (Long Short-Term Memory: LSTM)

แบบจำลองหน่วยความจำระยะสั้นแบบยาว(LSTM) เป็นรูปแบบหนึ่งของโครงข่ายระบบประสาทแบบย้อนกลับ (RNN)

เพื่อแก้ปัญหาการลดลงของเกรเดียนต์ (Vanishing Gradient) เมื่อลำดับของข้อมูลที่ได้รับเข้ามามีจำนวนมากเกินไป โดยการทำงานของ LSTM นั้นจะมีความคล้ายคลึงกับ RNN แต่ในส่วน of ชั้นซ่อน (Hidden state) ที่ใช้สำหรับจดจำลำดับของข้อมูล ก่อนหน้านั้น LSTM จะมีการเรียนรู้ว่า เมื่อใดควรลืม (Forget) เขียน (Write) หรืออนุญาตให้อ่าน (Read) ซึ่งทำให้สามารถรองรับข้อมูลที่มีลำดับปริมาณมากที่เข้ามาได้

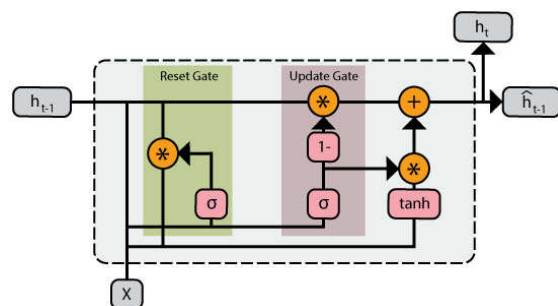


รูปที่ 2. แบบจำลองหน่วยความจำระยะสั้นแบบยาว

2.4 โครงข่ายประตูกักเก็บ (Gated Recurrent Unit:

GRU)

เป็นรูปแบบหนึ่งของโครงข่ายระบบประสาทแบบย้อนกลับ (RNN) มีจุดประสงค์เพื่อแก้ปัญหาการลดลงของเกรเดียนต์ (Vanishing Gradient) หรือเพื่อแก้ปัญหาการมีหน่วยความจำระยะสั้นของ RNN เช่นเดียวกับ LSTM ซึ่งการทำงานภายในของ GRU จะมีความคล้ายคลึงกับ LSTM เช่นกัน แต่จะลดความซับซ้อนของประตู(Gate) ลงเหลือเพียงแค่ 2 ประตูเพื่อให้เกิดความรวดเร็วในการคำนวณ ได้แก่ Update Gate เป็นประตูนำเข้าข้อมูลเพื่อนำไปคำนวณในการปรับสถานะของค่าเซลล์ และ Reset Gate เป็นประตูที่ใช้ในการตัดสินใจว่าจะปรับสถานะเซลล์เมื่อมีข้อมูลใหม่เข้ามาหรือไม่



รูปที่ 3. โครงข่ายประตูกักเก็บ

2.5 ปัญหาข้อมูลไม่สมดุล (Imbalanced data)

ปัญหาข้อมูลไม่สมดุล (Imbalanced data) เป็นปัญหาที่เกิดขึ้นในชุดข้อมูลที่มีการแบ่งประเภทของข้อมูลในชุดข้อมูล ออกเป็น 2 กลุ่ม (หรือมากกว่า) แล้วข้อมูลในกลุ่มใดกลุ่มหนึ่งมีจำนวนมากกว่าข้อมูลอีกกลุ่มหนึ่งมากการที่ข้อมูลในชุดข้อมูลหนึ่งๆ มีความไม่สมดุลเกิดขึ้น จะส่งผลกระทบต่อการทำงานของข้อมูลในกลุ่มส่วนน้อย เพราะโดยทั่วไปนั้น วิธีการจำแนกประเภท (Classification) ต่างๆ จะมีประสิทธิภาพสูงก็ต่อเมื่อข้อมูลในแต่ละกลุ่มในชุดข้อมูลนั้นมีจำนวนที่ใกล้เคียงกัน ซึ่งหากชุดข้อมูลมีความไม่สมดุล ก็จะส่งผลให้วิธีการจำแนกประเภทที่ใช้โดยทั่วไปนั้นมีโอกาสที่จะทำนายกลุ่มของชุดข้อมูลออกมาเป็นกลุ่มส่วนมาก อันนำไปสู่ปัญหาการแบ่งกลุ่มข้อมูลผิดกลุ่ม (Misclassification) เนื่องจากตัวแบบจะมองว่าการทำนายออกมาเป็นกลุ่มส่วนมากจะมีความแม่นยำมากกว่า ซึ่งก็จะทำให้ตัวแบบไม่สามารถทำนาย หรือแยกประเภทข้อมูลที่เป็นกลุ่มส่วนน้อยออกมาได้เลย

วิธีการแก้ปัญหาข้อมูลไม่สมดุลนั้น มีด้วยกันหลายวิธี แต่ในที่นี้จะขอยกมาถึงแค่วิธีเดียว ซึ่งเป็นวิธีที่ผู้วิจัยได้เลือกใช้ในการนำมาแก้ปัญหาความไม่สมดุลที่เกิดขึ้นในชุดข้อมูล ซึ่งก็คือวิธีการสุ่มตัวอย่างเพิ่มกลุ่มส่วนน้อยด้วยการสังเคราะห์ (Synthetic Minority Over-sampling Technique: SMOTE) เป็นวิธีที่ถูกเสนอขึ้นในปี 2002 โดย (Chawla et al., 2002) เป็นการสร้างกลุ่มตัวอย่างส่วนน้อยตัวอย่างใหม่ขึ้นมาอย่างสุ่ม โดยใช้หลักการกำหนดจำนวนเพื่อนบ้านที่อยู่ใกล้ที่สุดจำนวน k ตัว (k -nearest neighbors: KNN) แล้วทำการสุ่มสร้างข้อมูลขึ้นมาบนแนวเส้นที่เชื่อมต่อระหว่างตัวอย่างกลุ่มส่วนน้อยใดๆ

2.6 งานวิจัยที่เกี่ยวข้อง

เนื่องจากตัวข้อมูลสภาพอากาศนั้นเป็นข้อมูลเชิงอนุกรมเวลา ตัวแบบอนุกรมเวลา (Time series) จึงเป็นตัวเลือกลำดับต้นๆ ถูกนำมาใช้ในการพยากรณ์

(Wang et al., 2013) ได้ทำการพยากรณ์ปริมาณน้ำฝนสะสมในเมืองโจวกวัง มณฑลซานตง ประเทศจีน ซึ่งได้พบว่าปริมาณน้ำฝนนั้นมีสหสัมพันธ์อัตโนมัติที่โดดเด่น (Strong autocorrelation) จึงได้เลือกใช้ตัวแบบ SARIMA (Seasonal Autoregressive and Moving Average) ในการวิเคราะห์และพยากรณ์ ซึ่งหลังจากการเตรียมข้อมูล การกำหนดพารามิเตอร์ และเลือกตัวแบบที่ดีที่สุดตามวิธีการของการทำตัวแบบอนุกรม

เวลาแล้ว จากการทดสอบประสิทธิภาพพบว่า ค่าที่พยากรณ์ได้จากตัวแบบมีค่าใกล้เคียงกับปริมาณน้ำฝนสะสมจริง โดยมีค่าความคลาดเคลื่อนเพียงร้อยละ 27.42 เท่านั้น แต่ในงานวิจัยของ Wang และคณะ ไม่ได้กล่าวถึงการเปรียบเทียบแบบจำลอง SARIMA กับแบบจำลองอื่นๆ

(Narayanan et al., 2013) ได้ทำการพยากรณ์ปริมาณน้ำฝนสะสมในช่วงฤดูก่อนมรสุมตะวันตกเฉียงใต้ (ระหว่างเดือนมีนาคมถึงพฤษภาคม) ในภาคตะวันตกของประเทศอินเดีย โดยใช้ข้อมูลสภาพอากาศรายเดือน เฉพาะช่วงเดือนมีนาคมถึงพฤษภาคม จากสถานีวัดสภาพอากาศในประเทศอินเดีย 6 สถานี ในช่วงเวลา 60 ปี โดยมีการใช้ Mann Kendall test ในการทดสอบสหสัมพันธ์ระหว่างตัวแปร และใช้ตัวแบบ ARIMA ในการพยากรณ์ พบว่า ผลการพยากรณ์ทั้ง 6 สถานีมีความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ที่มากน้อยแตกต่างกันไป หมายความว่า การใช้ตัวแบบ ARIMA ในการพยากรณ์ปริมาณน้ำฝนสะสมอาจจะไม่ได้ตอบโจทย์สำหรับทุกสถานี หรือบางสถานีอาจจะต้องการการจัดการกับตัวแปรรับเข้าก่อนที่จะนำไปพยากรณ์ ทั้งนี้ งานวิจัยของ Narayanan และคณะ ก็ไม่ได้กล่าวถึงการเปรียบเทียบแบบจำลอง ARIMA กับแบบจำลองอื่นเช่นกัน

อย่างไรก็ตาม แบบจำลอง ARIMA ก็ยังมีข้อเสีย นั่นคือเป็นแบบจำลองที่พิจารณาโดยใช้ตัวแปรเพียงตัวเดียว ซึ่งในความเป็นจริงนั้น ปัจจัยที่ทำให้เกิดฝนหรือส่งผลต่อปริมาณน้ำฝนสะสมนั้น มีได้หลายปัจจัย และเพื่อเพิ่มประสิทธิภาพในการพยากรณ์จึงควรที่จะเอาปัจจัยสภาพอากาศด้านอื่นๆ มาพิจารณาด้วย เพื่อให้ตอบโจทย์ดังกล่าว แบบจำลอง ARIMAX จึงถูกยกขึ้นมาเพื่อทำการเปรียบเทียบกับแบบจำลอง ARIMA

(Wangdi et al., 2010) ได้เปรียบเทียบผลการพยากรณ์การเกิดไข้มาลาเรีย ระหว่างแบบจำลอง ARIMA ที่รับข้อมูลสถิติผู้ป่วยไข้มาลาเรียเพียงตัวแปรเดียว กับตัวแบบ ARIMAX ที่นอกจากจะรับข้อมูลของจำนวนผู้ป่วยแล้ว ยังรับข้อมูลด้านสภาพภูมิอากาศเพิ่มเข้าไปด้วย พบว่า ตัวแบบ ARIMAX ให้ผลการพยากรณ์ที่ดีกว่าตัวแบบ ARIMA

(Jalalkamali et al., 2015) ได้พัฒนาตัวแบบพยากรณ์ปริมาณน้ำฝนเพื่อคาดการณ์การเกิดปัญหายุ้งแล้ง โดยในงานวิจัยได้ทำการเปรียบเทียบผลการพยากรณ์ระหว่างตัวแบบ ARIMAX ตัวแบบโครงข่ายประสาทเทียม และตัวแบบ support-

vector machines (SVM) พบว่าในการพยากรณ์ระยะสั้น ตัวแบบ ARIMAX เป็นตัวแบบที่มีประสิทธิภาพมากที่สุด

นอกจากการใช้ตัวแบบ ARIMA และ ARIMAX การพยากรณ์โดยใช้แบบจำลองการเรียนรู้เชิงลึก โครงข่ายระบบประสาทแบบย้อนกลับ รวมไปถึงการทำ Feature Engineering เพื่อเพิ่มประสิทธิภาพให้กับตัวแบบก็ได้เข้ามามีบทบาทมากขึ้น

(Srachoom, 2007) ได้พัฒนาโครงข่ายประสาทเทียมเพื่อให้สามารถพยากรณ์ปริมาณน้ำฝนสะสมระยะสั้น จากข้อมูลสภาพอากาศย้อนหลังก่อนวันที่ต้องการพยากรณ์ได้ โดยข้อมูลที่ใช้เป็นข้อมูลพื้นที่อำเภอเมือง จังหวัดเชียงใหม่ ได้แก่ ความกดอากาศสูงสุด-ต่ำสุด อุณหภูมิสูงสุด-ต่ำสุด ความชื้นสัมพัทธ์สูงสุด-ต่ำสุด และปริมาณน้ำฝน ย้อนหลัง 9 ปี (พ.ศ. 2541-2549) และเพื่อให้โครงข่ายประสาทได้เรียนรู้แนวโน้มของสภาพอากาศก่อนวันฝนตกได้อย่างชัดเจน จึงได้เพิ่มชุดข้อมูลของสภาพอากาศย้อนหลัง 3 วัน ก่อนวันพยากรณ์ และเพิ่มดัชนีแสดงฤดูกาลเข้าไปด้วย ส่งผลให้มีตัวแปรต้นทั้งหมด 24 ตัว พบว่าโครงข่ายประสาทเทียมที่เหมาะสมมี 1 ชั้นซ่อน มีจำนวน 30 นิวรอน มีความถูกต้องเฉลี่ยร้อยละ 73.36

(Hung et al., 2009) ได้ทำการพยากรณ์ปริมาณน้ำฝนในเขตกรุงเทพมหานครแบบรายงานผลได้ทันที (real time) ด้วยโครงข่ายประสาทเทียม (ANN) โดยทำการเปรียบเทียบระหว่างแบบ Simple multilayer perceptron และแบบ Generalized feedforward ร่วมกับการปรับจูนพารามิเตอร์ ข้อมูลที่นำมาใช้เป็นข้อมูลที่ทำกรเก็บข้อมูลแบบรายชั่วโมงตลอดระยะเวลา 4 ปี จาก 75 สถานีตรวจวัดสภาพอากาศในกรุงเทพมหานคร พบว่า ในการพยากรณ์ปริมาณน้ำฝนในช่วงเวลา 1-6 ชั่วโมงนั้น ยิ่งชั่วโมงเพิ่มขึ้นเท่าไร ค่าความคลาดเคลื่อนของการพยากรณ์ก็จะยิ่งเพิ่มขึ้นเช่นกัน นอกจากนี้ในงานวิจัยนี้ยังได้ทำการหาความสัมพันธ์ระหว่างตัวแปรต้นกับตัวแปรตามเพื่อพิจารณาว่าตัวแปรใดบ้างที่ส่งผลกับปริมาณน้ำฝน พบว่า เมื่อนำข้อมูลสภาพอากาศจากสถานีตรวจวัดใกล้เคียง (ในรัศมี 21 ตารางกิโลเมตร) มาพิจารณาพร้อมกับสถานีตรวจวัดที่สนใจจะพยากรณ์แล้วนั้น ส่งผลให้ค่าความคลาดเคลื่อนลดลงถึงร้อยละ 71.19 เมื่อเทียบกับการพิจารณาข้อมูลจากสถานีตรวจวัดที่สนใจเพียงสถานีเดียว

(Hernández et al., 2016) ได้ศึกษาเรื่องการพัฒนาระบบแบบจำลองการเรียนรู้เชิงลึกเพื่อทำนายปริมาณน้ำฝนสะสมรายวัน โดยใช้ข้อมูลสภาพอากาศรายวันจากสถานีตรวจวัดใน

เมือง Manizales ประเทศ Colombia สิ่งที่น่าสนใจของงานวิจัยนี้คือการสร้างตัวแปรต้นจำนวน 47 ตัว โดยมาจากข้อมูลพื้นฐานที่เก็บรวบรวมมา ได้แก่ อุณหภูมิ ความชื้นสัมพัทธ์ ความกดอากาศ ความเข้มของแสงอาทิตย์ ความเร็วลม และทิศทางลม ในส่วนของข้อมูลที่สร้างขึ้นจะมาจากการข้อมูลพื้นฐานแต่ละตัวอีกที โดยการนำมา 1) วัดย้อนหลังไป 3 วัน 2) ค่าเฉลี่ยในช่วง 5 วันย้อนหลัง 3) ความแตกต่างของอุณหภูมิที่วัดได้ในช่วงเวลา 4:00 น. และ 24:00 น. เป็นต้น เนื่องจากตัวแปรต้นที่มีจำนวนมาก จึงได้มีการนำตัวเข้ารหัสอัตโนมัติ (Autoencoder) มาช่วยในการสกัดคุณลักษณะสำคัญ และทำการหาความสัมพันธ์ของแต่ละตัวแปร ก่อนที่จะส่งต่อไปยัง Multilayer Perceptron (MLP) เพื่อทำหน้าที่ในการพยากรณ์ปริมาณน้ำฝนสะสม

นอกจากการทำ Feature Engineering เพื่อเพิ่มประสิทธิภาพให้กับตัวแบบแล้ว งานวิจัยข้างต้นยังกล่าวถึงการจัดการกับข้อมูลเบื้องต้น เพื่อไม่ให้เกิดแนวโน้มการพยากรณ์ผิดกลุ่ม (Misclassification) เนื่องจากข้อมูลที่นำมาใช้นั้นมีทั้งข้อมูลช่วงที่ฝนตก และข้อมูลในช่วงที่ฝนไม่ตก หากนำข้อมูลในช่วงที่ฝนไม่ตกมาทำนายรวมด้วยก็จะทำให้ไม่สามารถทำนายปริมาณน้ำฝนได้อย่างถูกต้อง โดยงานวิจัย (Srachoom, 2007) ได้แก้ไขปัญหาดังกล่าวด้วยการเพิ่มดัชนีฤดูกาลด้วยเลขไบนารีขนาด 3 บิต เพื่อให้แบบจำลองสามารถรับรู้กลุ่มประเภทของข้อมูลตามฤดูกาลได้ และงานวิจัย (Hung et al., 2009) ได้ทำการศึกษาเปรียบเทียบโดยการนำเข้าสู่ชุดข้อมูลหลายๆแบบ พบว่า นอกจากการใช้ข้อมูลสภาพอากาศจากสถานีตรวจวัดใกล้เคียงมาพิจารณาพร้อมกับสถานีตรวจวัดที่สนใจจะพยากรณ์แล้ว การใส่ข้อมูลนำเข้าที่จะทำการพยากรณ์มีความแม่นยำมากที่สุดก็คือ ใส่ข้อมูลเฉพาะวันที่ฝนตก งานวิจัย (Hernández et al., 2016) จะมีความคล้ายงานวิจัย (Srachoom, 2007) แต่จะแบ่งข้อมูลด้วยการเพิ่มดัชนีรายเดือนด้วยเลขไบนารีขนาด 12 บิต

ในส่วนของการพยากรณ์ปริมาณน้ำฝนสะสมที่เกิดขึ้น นอกจากแบบจำลอง ARIMA และโครงข่ายประสาทเทียมที่ถูกกล่าวถึงในงานวิจัยข้างต้น ในปัจจุบันได้มีการนำความรู้เรื่องแบบจำลองการเรียนรู้เชิงลึกเข้ามาเพื่อช่วยทำนาย โดยแบบจำลองการเรียนรู้เชิงลึกที่นิยมนำมาใช้ในการทำนายปริมาณน้ำฝนคือ แบบจำลองที่มีความเกี่ยวข้องกับอนุกรมเวลา ได้แก่ RNN LSTM และ GRU โดยงานวิจัย (Salman et al., 2018) กับ (Fan et

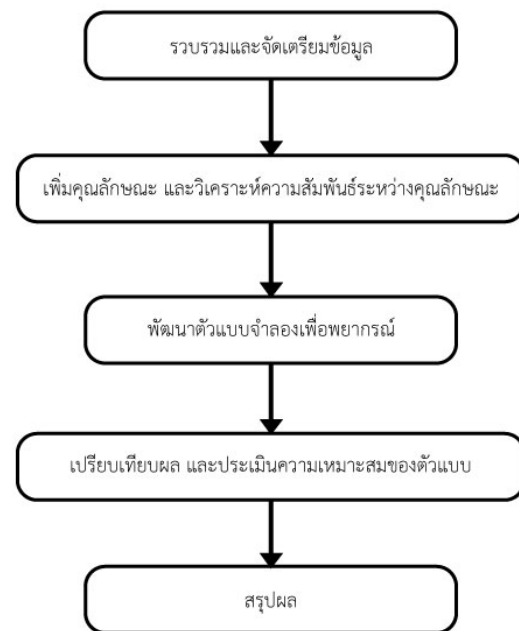
al., 2020) ได้ทำการทดลองพัฒนาแบบจำลอง LSTM เพื่อพยากรณ์ปริมาณน้ำฝน และนำแบบจำลองที่ได้ไปเปรียบเทียบกับแบบจำลองอื่นๆ และ (Aurnhammer & Frank, 2019) กับ (Yang et al., 2020) ได้ทำการทดลองเพื่อเปรียบเทียบความแม่นยำในการพยากรณ์ระหว่างแบบจำลองโครงข่ายระบบประสาทย้อนกลับ และพบว่า แบบจำลอง LSTM และ GRU มีความโดดเด่น และสามารถให้ผลการพยากรณ์ที่มีความถูกต้องและรวดเร็ว อย่างไรก็ตาม ในด้านความเหมาะสมในการเลือกใช้ตัวแบบจำลองโครงข่ายระบบประสาทแบบย้อนกลับนั้นขึ้นอยู่กับตัว Short-term memory ที่ได้เข้าไปด้วยเช่นกัน โดยทั้ง 2 งานวิจัยพบว่า ในชุดข้อมูลที่มีขนาดเล็ก แบบจำลอง GRU จะมีประสิทธิภาพในการเรียนรู้ที่ดีกว่า แต่สำหรับชุดข้อมูลขนาดใหญ่ แบบจำลอง LSTM จะมีประสิทธิภาพการเรียนรู้ที่ดีที่สุด

จากการศึกษางานวิจัยทั้งงานทางด้านการทำ Feature Engineering และด้านการปรับปรุงตัวแบบโครงข่ายระบบประสาทแบบย้อนกลับพบว่า งานวิจัยทั้ง 2 แบบจะมุ่งเน้นในส่วนของการปรับปรุงไฮเปอร์พารามิเตอร์ให้สามารถพยากรณ์ชุดข้อมูลให้มีความแม่นยำมากที่สุดเป็นหลักอยู่ ไม่ได้กล่าวถึงการทำ Feature Engineering มากนัก และไม่ได้มีการเปรียบเทียบให้เห็นว่าการเพิ่มคุณลักษณะทางสถิติต่างๆ ให้กับชุดข้อมูลนั้น ทำให้ประสิทธิภาพของการพยากรณ์เพิ่มขึ้นหรือไม่ จึงนำมาสู่จุดมุ่งหมายในการศึกษาข้อแรก คือการเปรียบเทียบประสิทธิภาพในการพยากรณ์ของตัวแบบที่ใช้ชุดข้อมูลที่แตกต่างกัน ระหว่างชุดข้อมูลผ่านการเพิ่มคุณลักษณะพิเศษกับชุดข้อมูลดั้งเดิม เพื่อดูว่าการเพิ่มคุณลักษณะนั้นส่งผลที่ดีกับประสิทธิภาพในการพยากรณ์จริงหรือไม่ และสำหรับจุดมุ่งหมายข้อที่สอง คือ เปรียบเทียบประสิทธิภาพในการพยากรณ์ในช่วงเวลาถัดไปของโครงข่ายระบบประสาทแบบย้อนกลับที่สนใจศึกษากับแบบจำลองอาร์มา จะเป็นการทดลองเพื่อหาแบบจำลองที่มีความเหมาะสมกับประเภทชุดข้อมูลที่นำมาศึกษามากที่สุด

3. วิธีดำเนินงานวิจัย

งานวิจัยนี้มีจุดประสงค์ในการพัฒนาแบบจำลองที่เหมาะสมต่อการพยากรณ์ปริมาณน้ำฝนในพื้นที่สนามบินสุวรรณภูมิ ดังนั้นในงานวิจัยนี้จะเน้นไปที่ชั่วโมงที่มีฝนตกเพื่อทำนายปริมาณ

น้ำฝนที่จะเกิดขึ้น ทั้งนี้ ขั้นตอนในการดำเนินการวิจัยจะแบ่งออกได้ดังรูปที่แสดงต่อไปนี้



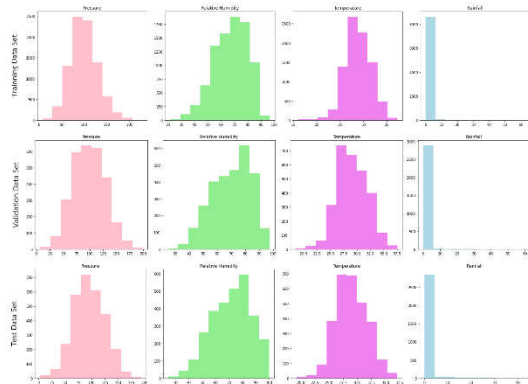
รูปที่ 4. ขั้นตอนการดำเนินงาน

3.1 รวบรวมและจัดเตรียมข้อมูล

งานวิจัยนี้ได้รับการสนับสนุนข้อมูลสภาพอากาศบริเวณสนามบินสุวรรณภูมีย้อนหลัง 5 ปี (ตั้งแต่ปี 2016 จนถึงปี 2020) จากกรมอุตุนิยมวิทยาประเทศไทย โดยชุดข้อมูลที่ได้จะมีทั้งหมด 4 คุณลักษณะ ได้แก่ อุณหภูมิรายชั่วโมง (Celsius) ความชื้นสัมพัทธ์รายชั่วโมง(Percent) ความกดอากาศรายชั่วโมง(hPa) และปริมาณน้ำฝนสะสม 3 ชั่วโมง (Millimeter)

ทั้งนี้ เนื่องจากข้อมูลทั้ง 4 คุณลักษณะมีความถี่ในการเก็บข้อมูลที่แตกต่างกัน และข้อมูลที่เราสนใจที่จะพยากรณ์ก็คือข้อมูลปริมาณน้ำฝน ดังนั้น ในเบื้องต้นจึงจำเป็นต้องทำการเปลี่ยนข้อมูลอุณหภูมิ ความชื้นสัมพัทธ์ และความกดอากาศที่เป็นข้อมูลรายชั่วโมงให้กลายเป็นข้อมูลราย 3 ชั่วโมงเหมือนกับข้อมูลของปริมาณน้ำฝนสะสมก่อน ซึ่งสามารถทำได้โดยการหาค่าเฉลี่ยของข้อมูลในทุกช่วง 3 ชั่วโมง โดยหลังจากที่ปรับความถี่ของช่วงเวลาของข้อมูลทั้ง 4 คุณลักษณะให้ตรงกันแล้วจะทำให้เรามีข้อมูลทั้งหมด 14,616 จุด แต่ก่อนที่จะนำเข้าแบบจำลองเพื่อพยากรณ์ เราจำเป็นต้องแบ่งข้อมูลออกเป็น 3 ส่วน ดังนี้ 1) ข้อมูลตั้งแต่ปี 2016 ถึงปี 2018 เป็นชุดข้อมูลฝึกสอน (Training data set) 2) ข้อมูลปี

2019 เป็นชุดข้อมูลตรวจสอบ (Validation data set) 3) ข้อมูลปี 2020 เป็นชุดข้อมูลทดสอบ (Test data set)



รูปที่ 5. การกระจายตัวในแต่ละคุณลักษณะของทั้ง 3 ชุดข้อมูล

จากรูปที่ 5 ซึ่งแสดงการกระจายตัวของข้อมูลในแต่ละคุณลักษณะจากแต่ละชุดข้อมูลที่เราได้ทำการแบ่งไว้ เราจะพบว่าข้อมูลอุณหภูมิและความกดอากาศมีลักษณะใกล้เคียงกับการกระจายตัวแบบปกติ (Normal Distribution) ในขณะที่ข้อมูลความชื้นสัมพัทธ์มีลักษณะเบ้ซ้าย (Left skewed) แต่สำหรับข้อมูลปริมาณน้ำฝนนั้น เราจะเห็นได้อย่างชัดเจนเลยว่าข้อมูลนั้นมีความไม่สมดุลเลย เนื่องจากปริมาณชั่วโมงที่ฝนไม่ตกนั้นมีมากกว่าปริมาณชั่วโมงที่ฝนตก และเมื่อมาพิจารณาอย่างละเอียดก็พบว่า ปริมาณชั่วโมงที่ฝนตกนั้น (ปริมาณน้ำฝนสะสมมากกว่า 0.1 มิลลิเมตร) มีเพียงร้อยละ 10 จากข้อมูลทั้งหมด ซึ่งปัญหาข้อมูลไม่สมดุลที่เกิดขึ้นดังกล่าว สามารถส่งผลกับการพยากรณ์ได้ จึงได้เลือกใช้วิธีการสุ่มตัวอย่างเพิ่มกลุ่มส่วนน้อยด้วยการสังเคราะห์ (Synthetic Minority Over-sampling Technique: SMOTE) สุ่มสร้างตัวอย่างชั่วโมงที่มีฝนตกขึ้นมาเพื่อให้แบบจำลองสามารถทำนายปริมาณน้ำฝนสะสมในชั่วโมงที่มีฝนตกได้อย่างแม่นยำมากขึ้น

3.2 เพิ่มคุณลักษณะ และวิเคราะห์ความสัมพันธ์ระหว่างคุณลักษณะ

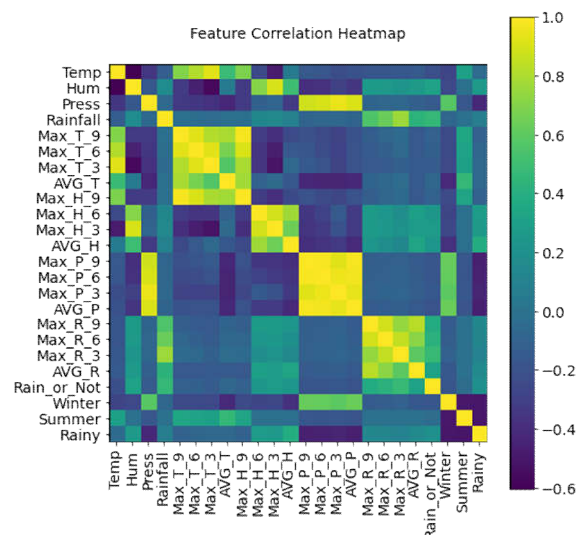
3.2.1 การเพิ่มข้อมูลคุณลักษณะทางสถิติ

เพื่อให้แบบจำลองสามารถเรียนรู้แนวโน้มของสภาพอากาศก่อนฝนจะตกได้ดียิ่งขึ้น จึงได้ทำการเพิ่มค่าคุณลักษณะทางสถิติให้กับชุดข้อมูลก่อนที่จะนำเข้าแบบจำลอง โดยเป็นการเพิ่มความสัมพันธ์ที่เกี่ยวข้องกันในเชิงเวลาให้กับแต่ละคุณลักษณะ

ได้แก่ ค่าสูงสุด เปรียบเทียบช่วงเวลาปัจจุบัน กับ 9 6 และ 3 ชั่วโมงก่อนหน้า ค่าเฉลี่ยในช่วง 15 ชั่วโมง ของทุกคุณลักษณะ อีกทั้งยังเพิ่มคุณลักษณะที่บ่งบอกฤดูกาล เช่น ช่วงเดือน พฤศจิกายนถึงเดือนกุมภาพันธ์จะเป็นฤดูหนาว ช่วงเดือนมีนาคมถึงเดือนมิถุนายนเป็นฤดูร้อน และช่วงเดือนกรกฎาคมถึงเดือนตุลาคมเป็นฤดูฝน ซึ่งทำให้คุณลักษณะที่เราจะนำเข้าแบบจำลองมีทั้งหมด 23 คุณลักษณะ

3.2.2 วิเคราะห์ความสัมพันธ์ระหว่างคุณลักษณะ

ในการเพิ่มประสิทธิภาพในการพยากรณ์ปริมาณน้ำฝนในช่วงเวลาถัดไปให้มีความแม่นยำมากยิ่งขึ้น งานวิจัยฉบับนี้จึงได้ทำการเพิ่มคุณลักษณะบางอย่างตามที่ไดกล่าวถึงไปในหัวข้อที่แล้วเข้ามาด้วย หลังจากนั้นจึงได้ทำการวิเคราะห์หาความสัมพันธ์ (Correlation) ระหว่างคุณลักษณะ ซึ่งได้ผลลัพธ์ดังรูปต่อไปนี้



รูปที่ 6. แผนภาพแสดงค่าสหสัมพันธ์ระหว่างแต่ละคุณลักษณะ

อย่างไรก็ตาม การนำคุณลักษณะทั้งหมด 23 คุณลักษณะเข้าสู่แบบจำลองอาจจะไม่ส่งผลให้ประสิทธิภาพของแบบจำลองเพิ่มมากขึ้นกว่าเดิม เนื่องจากบางคุณลักษณะก็ไม่ได้มีความสัมพันธ์กับตัวคุณลักษณะปริมาณน้ำฝนเลย ดังนั้น เพื่อเป็นการเปรียบเทียบให้เห็นผลว่าการเพิ่มคุณลักษณะนั้นสามารถช่วยเพิ่มประสิทธิภาพให้กับแบบจำลองได้จริง จึงได้ทำการแบ่งชุดข้อมูลก่อนจะนำเข้าแบบจำลองออกเป็น 3 ชุด ดังนี้

- 1) ชุดข้อมูลที่มีเพียง 4 คุณลักษณะดั้งเดิม ได้แก่ ปริมาณน้ำฝนสะสม อุณหภูมิ ความกดอากาศ และความชื้นสัมพัทธ์

2) ชุดข้อมูลผ่านการเพิ่มคุณลักษณะทั้งสิ้น 19 คุณลักษณะ และจะรวมกับคุณลักษณะดั้งเดิม ทำให้มีคุณลักษณะรวมทั้งสิ้น 23 คุณลักษณะ

3) ชุดข้อมูลที่เพิ่มเฉพาะคุณลักษณะที่มีความสัมพันธ์กับปริมาณน้ำฝนสะสมเท่านั้น ซึ่งจะมีเพียง 7 คุณลักษณะ ซึ่งทำให้ข้อมูลชุดที่ 3 จะมีคุณลักษณะทั้งหมด 11 คุณลักษณะ โดยคุณลักษณะที่เพิ่มเข้ามาคือ ปริมาณน้ำฝนสูงสุด เปรียบเทียบช่วงเวลาปัจจุบัน กับ 9 6 และ 3 ชั่วโมงก่อนหน้า ค่าเฉลี่ยในช่วง 15 ชั่วโมง และฤดูกาล

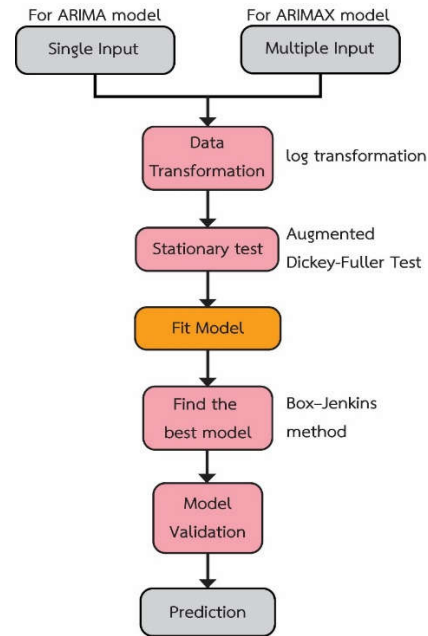
3.3 พัฒนาตัวแบบจำลองเพื่อพยากรณ์

3.3.1 แบบจำลองกลุ่มอาร์มา

ในกลุ่มนี้จะประกอบไปด้วย 2 แบบจำลอง ได้แก่ แบบจำลอง ARIMA และแบบจำลอง ARIMAX ทั้ง 2 แบบจำลองจะมีขั้นตอนการพัฒนาตัวแบบและการเขียนโปรแกรมที่ค่อนข้างคล้ายกัน ต่างกันแค่เพียงตัวแบบ ARIMA จะมีข้อมูลนำเข้าเพียงข้อมูลเดียว นั่นคือข้อมูลปริมาณน้ำฝน แต่สำหรับตัวแบบ ARIMAX นั้น จะสามารถนำเข้าข้อมูลคุณลักษณะอื่นๆ ที่นอกเหนือจากปริมาณน้ำฝนเข้าแบบจำลองได้ โดยทำให้ข้อมูลคุณลักษณะอื่นๆ อยู่ในรูปแบบของเมทริกซ์ ซึ่งนั่นทำให้เราสามารถทำการทดลองชุดข้อมูลทั้ง 3 ชุดที่เรามีกับแบบจำลอง ARIMAX ได้

จากผลการทดสอบความเหมาะสม พบว่าตัวแบบ ARIMA(1,1,3) เป็นตัวแบบที่มีความเหมาะสมมากที่สุด และสำหรับตัวแบบอนุกรมเวลา ARIMAX พบว่า ตัวแบบที่เหมาะสมที่สุดสำหรับชุดข้อมูลชุดที่ 1 คือ ARIMAX (1,1,4) ตัวแบบที่เหมาะสมที่สุดสำหรับชุดข้อมูลชุดที่ 2 คือ ARIMAX (1,1,2) และตัวแบบที่เหมาะสมที่สุดสำหรับชุดข้อมูลชุดที่ 3 คือ ARIMAX (3,1,2)

สำหรับขั้นตอนการทำงานของโปรแกรมโดยคร่าว สามารถแสดงได้ดังรูปต่อไปนี้



รูปที่ 7. แผนภาพแสดงขั้นตอนการดำเนินการในการพัฒนาตัวแบบอนุกรมเวลา ARIMA และ ARIMAX

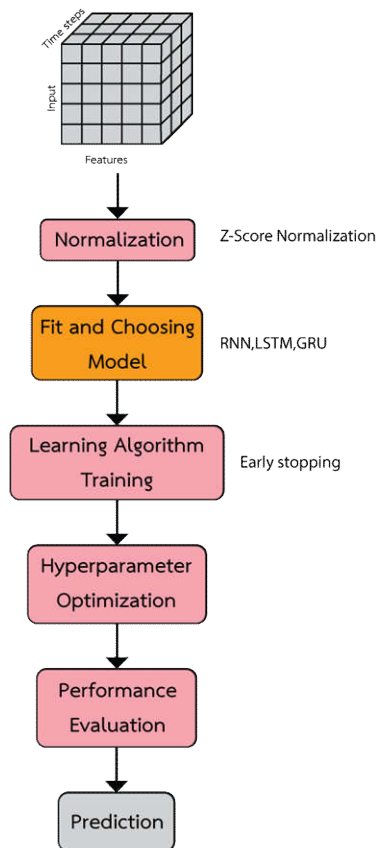
3.3.2 แบบจำลองกลุ่มโครงข่ายระบบประสาทแบบ

ย้อนกลับ

ในกลุ่มนี้จะประกอบไปด้วย 3 แบบจำลอง ได้แก่ แบบจำลองหน่วยความจำระยะสั้นแบบยาว (LSTM) โครงข่ายระบบประสาทแบบย้อนกลับ (RNN) และโครงข่ายประสาทวนกลับ (GRU) ซึ่งหลังจากที่เราจัดการกับข้อมูลเบื้องต้นเสร็จแล้ว (ในหัวข้อ 3.1 และ 3.2) เราจะทำการการปรับปรุงโครงสร้างข้อมูล (Normalization) ซึ่งจากรูปที่ 3.3 เราจะเห็นได้ว่าชุดข้อมูลในแต่ละคุณลักษณะมีการกระจายตัวที่แตกต่างกัน ดังนั้น จึงจำเป็นต้องปรับปรุงโครงสร้างของข้อมูลเพื่อลดความซับซ้อนและหลีกเลี่ยงความผิดพลาดของข้อมูลก่อนที่จะนำไปในแบบจำลอง โดยจะทำการปรับปรุงโครงสร้างข้อมูลด้วยวิธี Z-Score Normalization หรือ Standardization หลังจากนั้นจึงนำข้อมูลทั้งหมดเข้าแบบจำลองเพื่อทำการฝึกสอนโดยจะทำการ

แปลงข้อมูลนำเข้าให้อยู่ในรูปแบบอาร์เรย์ 3 มิติ คือ [Input, Timestep, Features] โดยที่ Input คือจำนวนข้อมูลที่เราจะนำเข้าแบบจำลอง Timestep คือ จำนวนช่วงเวลาก่อนหน้า ($t+n$) ซึ่งในที่นี้เราได้ใช้ทั้งหมด 80 ช่วงเวลา ($t+80$) หรือ 240 ชั่วโมงก่อนหน้า (10 วัน) เพื่อทำนายปริมาณน้ำฝนสะสมในอีกช่วงเวลาถัดไป

สำหรับขั้นตอนการทำงานของโปรแกรมโดยคร่าวสามารถแสดงได้ดังรูปต่อไปนี้



รูปที่ 8. แผนภาพแสดงกระบวนการทำงานของแบบจำลอง
โครงข่ายระบบประสาทแบบย้อนกลับ

พารามิเตอร์ที่ใช้ในการฝึกสอน ได้ทำการทดสอบทั้งหมด 500 Epochs และมีการใช้ Adaptive Moment Estimation (ADAM) เป็นตัว Optimizer เพื่อเพิ่มประสิทธิภาพในการทำนายผลของแบบจำลอง ในการวัดประสิทธิภาพของตัวแบบได้ทำการเลือก ฟังก์ชันสูญเสีย (Loss function) คือ Mean Absolute Error (MAE) ซึ่งจะมีความอ่อนไหวกับชุดข้อมูลที่มีค่าผิดปกติ (Outlier) น้อยกว่า Mean Square Error (MSE) และ

Root Mean Square Error (RMSE) เนื่องจากการนำค่าความคลาดเคลื่อน (Error) มาใส่ สัมบูรณ์ (Absolute) เหตุผลที่เลือก MAE เนื่องจากข้อมูลปริมาณน้ำฝนสะสมนั้นจะมีความเอนเอียงเนื่องจากช่วงเวลาที่ฝนตกมีน้อยกว่าช่วงเวลาที่ฝนไม่ตก (ดังที่แสดงในรูปที่ 3.4) อีกทั้ง ปริมาณน้ำฝนสะสมในบางช่วงก็มีปริมาณที่สูงมากกว่าปกติ ซึ่งอาจจะเกิดจากอิทธิพลของพายุหรือปัจจัยอื่นๆ และข้อมูลเหล่านี้เองที่กลายเป็นค่าผิดปกติ (Error) ในชุดข้อมูล การเลือกใช้ MAE ในการประเมินประสิทธิภาพของตัวแบบจึงมีความเหมาะสมกับชุดข้อมูลนี้มากกว่าค่าอื่นๆ

สำหรับเทคนิคการฝึกสอน ได้เลือกใช้เทคนิค Early stopping ซึ่งเป็นเทคนิคในการสร้างความสมดุลระหว่างค่าความเอนเอียง (Bias) และค่าความแปรปรวน (Variation) เพื่อไม่ให้แบบจำลองประสบปัญหา Overfitting (แบบจำลองอิงกับข้อมูลในชุดข้อมูลฝึกสอนมากเกินไป คือมีค่าความถูกต้องในการทำนายข้อมูลฝึกสอนสูง แต่ค่าความถูกต้องในการทำนายข้อมูลชุดทดสอบต่ำ) โดยการสังเกตค่า Validation loss ในทุกๆ Epochs ด้วยเงื่อนไขว่า หากค่าดังกล่าวไม่ลดลงติดต่อกัน 10 Epochs ให้หยุดการฝึกสอน

หลังจากนั้นจึงได้นำแบบจำลองแต่ละตัวมาทำการปรับจูนไฮเปอร์พารามิเตอร์ เพื่อให้ได้ไฮเปอร์พารามิเตอร์ที่มีความเหมาะสมกับแต่ละตัวแบบ และช่วยให้ได้ประสิทธิภาพในการพยากรณ์ที่สูงสุด โดยในการปรับจูนค่าไฮเปอร์พารามิเตอร์ (Hyper-Parameter Tuning) ในแต่ละตัวแบบที่เราเลือกมาใช้นั้น มีการปรับจูนค่าต่างๆ ดังนี้

แบบจำลอง	ชั้นของแบบจำลอง	ไฮเปอร์พารามิเตอร์	ฟังก์ชันกระตุ้น
RNN	SimpleRNN layer 1	32	Tanh
	SimpleRNN layer 2	64	
	Dense	32	
	Reshape	[output, number of features]	

LSTM	LSTM layer 1	64	Tanh
	LSTM layer 2	256	
	Dense	32	
	Reshape	[output, number of features]	
GRU	GRU layer 1	32	Tanh
	GRU layer 2	64	
	Dense	1	
	Reshape	[output, number of features]	

ตารางที่ 1. แสดงการปรับจูนไฮเปอร์พารามิเตอร์ที่ชั้น (Layer) ต่างๆ ของตัวแบบแต่ละตัว

3.4 เปรียบเทียบผลการพยากรณ์ และประเมินความเหมาะสมของตัวแบบ

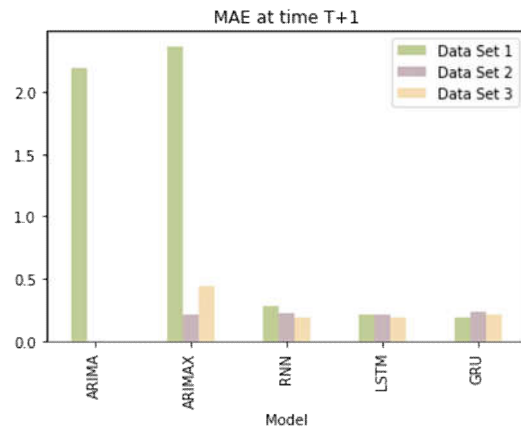
3.4.1 ผลการทดลองโดยการเปรียบเทียบประสิทธิภาพในการพยากรณ์ปริมาณน้ำฝนสะสมในอีก 1 ช่วงเวลาถัดไป

ในส่วนการเปรียบเทียบประสิทธิภาพในการพยากรณ์ของแต่ละแบบจำลอง กับแต่ละชุดข้อมูล จะทำการพิจารณาจากค่า Mean Absolute Error (MAE) และจะเป็นการเปรียบเทียบการทำนายปริมาณน้ำฝนสะสมในอีก 3 ชั่วโมงถัดไป ($t+3$)

	ARIMA	ARIMAX	RNN	LSTM	GRU
ชุดที่ 1	2.195	2.371	0.283	0.215	0.193
ชุดที่ 2		0.216	0.226	0.218	0.235
ชุดที่ 3		0.444	0.195	0.188	0.216

ตาราง 2. แสดงค่า MAE ใน 1 ช่วงเวลาถัดไป

ค่า MAE ที่ได้ สามารถสรุปได้เป็นแผนภาพดังนี้



รูปที่ 9. แผนภาพแสดงค่า MAE ใน 1 ช่วงเวลาถัดไป

จากการทดลองพบว่า แบบจำลองที่ให้ค่าเฉลี่ยความผิดพลาดกำลังสองน้อยที่สุดคือ แบบจำลองหน่วยความจำระยะสั้นแบบยาว

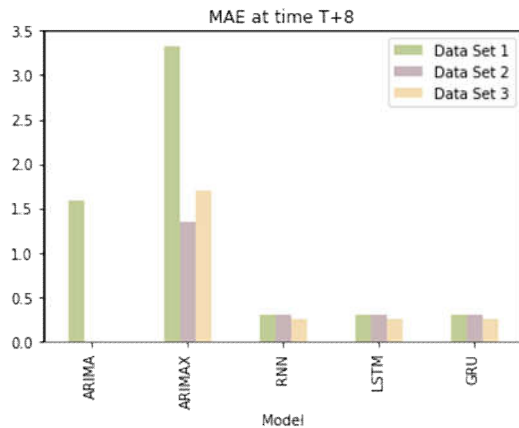
3.4.2 ผลการทดลองโดยการเปรียบเทียบประสิทธิภาพในการพยากรณ์ปริมาณน้ำฝนสะสมในอีกหลายช่วงเวลาถัดไป

การพิจารณาจากค่า Mean Absolute Error (MAE) และจะเป็นการเปรียบเทียบการทำนายปริมาณน้ำฝนสะสมในอีก 24 ชั่วโมงถัดไป ($t+8$)

	ARIMA	ARIMAX	RNN	LSTM	GRU
ชุดที่ 1	1.589	3.335	0.310	0.310	0.286
ชุดที่ 2		1.349	0.310	0.329	0.297
ชุดที่ 3		1.709	0.26	0.248	0.242

ตาราง 3. แสดงค่า MAE ใน 8 ช่วงเวลาถัดไป

ค่า MAE ที่ได้ สามารถสรุปได้เป็นแผนภาพดังนี้



รูปที่ 10. แผนภาพแสดงค่า MAE ใน 8 ช่วงเวลาถัดไป

จากการทดลองในการพยากรณ์ข้อมูลในอีกหลายช่วงเวลาถัดไปพบว่า การเพิ่มเฉพาะคุณลักษณะทางสถิติให้กับชุดข้อมูลก่อนที่จะนำเข้าแบบจำลองส่งผลให้ค่าเฉลี่ยความผิดพลาดกำลังสองลดลง แบบจำลองที่ให้ค่าเฉลี่ยความผิดพลาดกำลังสองน้อยที่สุดคือ โครงข่ายประตูกลับ

4. บทสรุปและการอภิปราย

งานวิจัยนี้มุ่งเน้นการพัฒนาและวิเคราะห์เปรียบเทียบประสิทธิภาพในการพยากรณ์ปริมาณน้ำฝนสะสมล่วงหน้าในระยะสั้นร่วมกับการใช้ข้อมูลหลายตัวแปร ข้อมูลที่นำมาใช้ในงานวิจัยนี้เป็นข้อมูลสภาพอากาศและปริมาณน้ำฝนสะสมในพื้นที่สนามบินสุวรรณภูมิที่ได้รับการสนับสนุนจากกรมอุตุนิยมวิทยา ประเทศไทย

ในการพัฒนาตัวแบบ ผู้วิจัยได้เตรียมชุดข้อมูลนำเข้าไว้ 3 ชุด คือ 1) ชุดข้อมูลดั้งเดิม (รวม 4 คุณลักษณะ) 2) ชุดข้อมูลที่เพิ่มคุณลักษณะทางสถิติ (รวม 23 คุณลักษณะ) 3) ชุดข้อมูลที่เพิ่มเฉพาะคุณลักษณะทางสถิติที่เกี่ยวข้องกับปริมาณน้ำฝนสะสม (รวม 11 คุณลักษณะ) เนื่องจากเมื่อลองพิจารณาข้อมูลที่อยู่ในชุดข้อมูลแล้วพบว่า จำนวนชั่วโมงที่มีฝนตกนั้น มีน้อยกว่าจำนวนชั่วโมงที่ไม่มีฝนตก โดยจำนวนชั่วโมงที่มีฝนตกนั้น คิดเป็นร้อยละ 10 จากจำนวนชั่วโมงที่ไม่มีฝนตก ซึ่งก่อให้เกิดปัญหาชุดข้อมูลไม่สมดุลตามมา จึงได้จัดการกับปัญหาดังกล่าวด้วยการเลือกใช้วิธีการสุ่มตัวอย่างเพิ่มกลุ่มส่วนน้อยด้วยการสังเคราะห์ (SMOTE) เพื่อให้มีจำนวนชั่วโมงที่มีฝนตกมากพอที่จะทำให้แบบจำลองสามารถทำนายปริมาณน้ำฝนสะสมได้

จากนั้นจึงได้ทำการพัฒนาตัวแบบจำลองที่เกี่ยวข้องกับการจัดการข้อมูลที่อยู่ในรูปของอนุกรมเวลาขึ้น ได้แก่

1. ตัวแบบ ARIMA เป็นตัวแบบที่รับข้อมูลนำเข้าเฉพาะข้อมูลปริมาณน้ำฝน จากการพัฒนาและคัดเลือกตัวแบบที่ดีที่สุดด้วยวิธีวิธีบอกซ์แอนด์เจนกินส์ พบว่าตัวแบบที่ให้ค่าความคลาดเคลื่อนน้อยที่สุด คือตัวแบบ ARIMA (1,1,3)
2. ตัวแบบ ARIMAX เป็นตัวแบบที่รับข้อมูลคุณลักษณะอื่นๆ มาพิจารณาพร้อมกับข้อมูลปริมาณน้ำฝนที่เราต้องการจะพยากรณ์ จากการพัฒนาและคัดเลือกตัวแบบที่ดีที่สุดด้วยวิธีบอกซ์แอนด์เจนกินส์ จะทำให้เราได้ตัวแบบ ARIMAX ที่ดีที่สุดสำหรับแต่ละชุดข้อมูล ได้แก่

- ข้อมูลชุดที่ 1 ตัวแบบที่ดีที่สุด คือ ตัวแบบ ARIMAX (1,1,4)

- ข้อมูลชุดที่ 2 ตัวแบบที่ดีที่สุด คือ ตัวแบบ ARIMAX (1,1,2)

- ข้อมูลชุดที่ 3 ตัวแบบที่ดีที่สุด คือ ตัวแบบ ARIMAX (3,1,2)

3. แบบจำลองการเรียนรู้เชิงลึก เป็นตัวแบบที่รับข้อมูลนำเข้าหลายตัวแปร โดยในงานวิจัยนี้ได้เลือกใช้ตัวแบบจำลองการเรียนรู้เชิงลึก 3 ตัว มาเพื่อใช้ในการเปรียบเทียบ ได้แก่ โครงข่ายระบบประสาทแบบย้อนกลับ (RNN) แบบจำลองหน่วยความจำระยะสั้นแบบยาว (LSTM) และโครงข่ายประตูกลับ (GRU)

ในส่วนการทดลอง งานวิจัยนี้ได้แบ่งการทดลองออกเป็น 2 ส่วน ดังนี้

1) การทดลองโดยการเปรียบเทียบประสิทธิภาพระหว่างการพัฒนาตัวแบบโดยการเพิ่มคุณลักษณะทางสถิติ

จากการทดลองในเบื้องต้นพบว่า การเพิ่มเฉพาะคุณลักษณะทางสถิติให้กับชุดข้อมูลก่อนที่จะนำไปทำการพัฒนาตัวแบบนั้น ส่งผลให้ประสิทธิภาพในการพยากรณ์ปริมาณน้ำฝนสะสมในช่วงเวลาใกล้เคียงเพิ่มขึ้น (ในอีก 8 ช่วงเวลาถัดไป) โดยเฉพาะในชุดข้อมูลชุดที่ 3 ที่มีการเพิ่มเฉพาะคุณลักษณะทางสถิติที่มีความสัมพันธ์กับปริมาณน้ำฝนสะสมให้กับชุดข้อมูลก่อนที่จะนำเข้าแบบจำลองส่งผลให้ Mean Absolute Error (MAE) ลดลงมากกว่าการใช้เพียงคุณลักษณะดั้งเดิมเพียง 4 คุณลักษณะ แต่สำหรับการพยากรณ์ในระยะสั้น (1 ช่วงเวลาถัดไป) การเพิ่มคุณลักษณะทางสถิติให้กับชุดข้อมูลในแบบจำลองโครงข่ายระบบประสาทแบบย้อนกลับยังไม่สามารถเพิ่มประสิทธิภาพในการพยากรณ์ให้ดีขึ้นได้มากนัก ยกเว้น

สำหรับแบบจำลอง ARIMAX ซึ่งเห็นได้อย่างชัดเจนว่า ในการพยากรณ์ระยะสั้นนั้น การเพิ่มคุณลักษณะทางสถิติให้กับชุดข้อมูล ส่งผลให้ประสิทธิภาพในการพยากรณ์ดีขึ้นเกือบเทียบเท่ากับโครงข่ายระบบประสาทแบบย้อนกลับ แต่เมื่อช่วงเวลาไกลออกไป ประสิทธิภาพในการพยากรณ์กลับลดลงอย่างมาก

อย่างไรก็ตาม สำหรับข้อมูลชุดที่ 2 หากสามารถพัฒนาตัวแบบให้มีความซับซ้อนได้มากกว่านี้ อาจจะส่งผลให้การสกัดคุณลักษณะในแต่ละชั้นของแบบจำลองโครงข่ายระบบประสาทแบบย้อนกลับทำได้ดีขึ้น และอาจจะส่งผลให้ประสิทธิภาพของการพยากรณ์เพิ่มขึ้น อีกทั้ง ถึงแม้จะพัฒนาตัวแบบจำลองจนสามารถพยากรณ์ชั่วโมงที่มีฝนตกได้ แต่เมื่อพิจารณาจากผลของการพยากรณ์แล้วก็พบว่ายังคงมีปัญหาในเรื่องที่ไม่สามารถทำนายชั่วโมงที่มีปริมาณน้ำฝนสะสมมากกว่าปกติได้ ซึ่งเป็นส่วนที่ต้องพัฒนาต่อไป

2) การทดลองโดยการเปรียบเทียบประสิทธิภาพในการพยากรณ์ปริมาณน้ำฝนสะสมในช่วงเวลาถัดไปของแต่ละแบบจำลอง

การทดลองเปรียบเทียบประสิทธิภาพในการพยากรณ์ของตัวแบบแต่ละตัวในช่วงเวลาที่แตกต่างกัน งานวิจัยได้แบ่งการทดลองส่วนนี้ออกเป็น 2 ส่วนเพื่อวัดประสิทธิภาพ คือ เปรียบเทียบประสิทธิภาพในการพยากรณ์ปริมาณน้ำฝนสะสมใน 1 ช่วงเวลาถัดไป (3 ชั่วโมงถัดไป) และเปรียบเทียบประสิทธิภาพในการพยากรณ์ใน 8 ช่วงเวลาถัดไป (24 ชั่วโมง) พบว่า แบบจำลองในกลุ่มโครงข่ายระบบประสาทแบบย้อนกลับให้ประสิทธิภาพในการพยากรณ์ที่ดีกว่าแบบจำลองในกลุ่มอาร์มา โดยในการพยากรณ์ปริมาณน้ำฝนสะสมในอีก 1 ช่วงเวลาถัดไปนั้น แบบจำลอง LSTM ให้ประสิทธิภาพในการพยากรณ์ได้ดีที่สุด แต่ในการพยากรณ์ในอีก 8 ช่วงเวลาถัดไปนั้น พบว่า แบบจำลอง GRU ให้ประสิทธิภาพในการพยากรณ์ได้ดีที่สุด อีกทั้ง เมื่อเปรียบเทียบกับแบบจำลองโครงข่ายประสาทแบบย้อนกลับตัวอื่นๆ แบบจำลอง GRU ยังเป็นแบบจำลองที่มีการประมวลผลที่ไว เนื่องจากมีจำนวนประตุน้อย ดังนั้น สำหรับการเลือกแบบจำลองโครงข่ายระบบประสาทย้อนกลับเพื่อทำนายปริมาณน้ำฝนสะสม แบบจำลอง GRU จึงเป็นตัวเลือกที่น่าสนใจ

5. ข้อเสนอแนะ

5.1 ข้อจำกัดในงานวิจัย

ในการพัฒนาตัวแบบจำลองเพื่อทำนายปริมาณน้ำฝนสะสมในพื้นที่สนามบินนานาชาติสุวรรณภูมิมีข้อจำกัดดังต่อไปนี้

1. ข้อมูลสภาพอากาศในไทยส่วนมาเป็นข้อมูลที่ไม่ได้เผยแพร่แบบสาธารณะ ทำให้ข้อมูลสภาพอากาศที่นำมาใช้สร้างตัวแบบมีน้อย และจำกัด
2. จำนวนสถานีตรวจวัดสภาพอากาศในประเทศไทยมีจำนวนไม่มาก (เฉลี่ยจังหวัดละ 1 แห่ง) จึงไม่สามารถนำข้อมูลสภาพอากาศจากสถานีใกล้เคียงมาพิจารณาร่วมได้
3. ด้านสมรรถภาพของคอมพิวเตอร์ เนื่องจากทรัพยากรด้านคอมพิวเตอร์ที่จำกัดของผู้วิจัย จึงไม่สามารถพัฒนาตัวแบบที่มีความซับซ้อน หรือนำข้อมูลช่วงเวลาในอดีตได้เข้าไปมากกว่านี้ได้

5.2 แนวการวิจัยในอนาคต

ทั้งนี้ เพื่อให้แบบจำลองมีประสิทธิภาพ และความแม่นยำในการพยากรณ์มากยิ่งขึ้น ในอนาคตควรจะมีการเพิ่มเติมในส่วนต่างๆ ดังนี้

- 1) ด้านข้อมูล ควรนำข้อมูลพื้นที่ หรือข้อมูลเกี่ยวกับสภาพอากาศด้านอื่นๆ เช่น ปริมาณแสงแดด ทิศทางลม ความหนาแน่นของเมฆ ฯ มาพิจารณาด้วย เนื่องจากทั้งลักษณะทางภูมิศาสตร์ และปัจจัยด้านสภาพอากาศอื่นๆ ก็มีผลต่อปริมาณน้ำฝนสะสม
- 2) การปรับเปลี่ยนความยาวของช่วงเวลาในอดีต (Timestep) ที่จะนำเข้าแบบจำลองให้มีความยาวเพิ่มขึ้นมากกว่า 80 ช่วงเวลา เพื่อเปรียบเทียบประสิทธิภาพในการพยากรณ์ของแต่ละแบบจำลองกับแต่ละชุดข้อมูลเพื่อดูว่าการเพิ่มคุณลักษณะให้ชุดข้อมูลนั้นจะส่งผลต่อประสิทธิภาพการพยากรณ์จริงๆ หรือไม่ และแบบจำลองใดที่เหมาะสมกับข้อมูลชุดนี้มากที่สุด
- 3) การลองใช้ Optimization Algorithm ตัวอื่นๆ ในการเพิ่มประสิทธิภาพ และเปรียบเทียบผลการพยากรณ์
- 4) การใช้การเรียนรู้แบบกลุ่ม (Ensemble Learning Method) เข้ามาเป็นตัวช่วยในการพัฒนาตัวแบบให้สามารถพยากรณ์ได้อย่างแม่นยำ

เอกสารอ้างอิง

- [1] Aurnhammer, C., and Frank, S. L., “Comparing gated and simple recurrent neural network architectures as models of human sentence processing”, 2019.
- [2] Chawla, N. V., Bowyer K. W., Hall, L. O., and Kegelmeyer W. P., “SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research”, vol. 16, pp. 321-357, 2002.
- [3] Chung, J., Gulcehre, C., Cho, K., and Bengio, Y., “Empirical evaluation of gated recurrent neural networks on sequence modeling”, 2014.
- [4] Fan, H., Jiang, M., Xu, L., Zhu, H., Cheng, J., and Jiang, J., “Comparison of long short-term memory networks and the hydrological model in runoff simulation,” *Water*, vol. 12 no. 1, pp. 175, 2020.
- [5] Hernández, E., Sanchez-Anguix, V., Julian, V., Palanca, J., and Duque, N., “Rainfall prediction: A deep learning approach. International Conference on Hybrid Artificial Intelligence Systems,” 2016.
- [6] Hochreiter, S., and Schmidhuber, J., “Long short-term memory. *Neural Comput*,” vol. 9, no.8, pp. 1735-1780, 1997.
- [7] Hung, N. Q., Babel, M. S., Weesakul, S., and Tripathi, N., “An artificial neural network model for rainfall forecasting in Bangkok, Thailand,” *Hydrology and Earth System Sciences*, vol.13(8), pp. 1413-1425, 2009.
- [8] Jalalkamali, A., Moradi, M., and Moradi, N., “Application of several artificial intelligence models and ARIMAX model for forecasting drought using the Standardized Precipitation Index,” *International journal of environmental science and technology*, vol.12(4), pp. 1201-1210, 2015.
- [9] Manokij, F., “Thailand's Precipitation Forecasting Using Deep Learning Approach,” Chulalongkorn University, Bangkok, 2019.
- [10] Narayanan, P., Basistha, A., Sarkar, S., and Sachdeva, K., “Trend analysis and ARIMA modelling of pre-monsoon rainfall data for western India,” *Comptes Rendus Geoscience*, vol. 345(1), pp. 22-27, 2013.
- [11] Poornima, S., and Pushpalatha, M., “Prediction of rainfall using intensified LSTM based recurrent neural network with weighted linear units,” *Atmosphere*, vol.10(11), pp. 668, 2019.
- [12] Salman, A. G., Heryadi, Y., Abdurahman, E., and Suparta, W., “Weather Forecasting Using Merged Long Short-Term Memory Model (LSTM) and Autoregressive Integrated Moving Average (ARIMA) Model,” *J. Comput. Sci.*, vol. 14(7), pp. 930-938, 2018.
- [13] Sanguansat, P., “Artificial Intelligence with Machine Learning,” IDC Premier Limited, 2019.
- [14] Shewalkar, A., “Performance evaluation of deep neural networks applied to speech recognition: RNN, LSTM and GRU,” *Journal of Artificial Intelligence and Soft Computing Research*, vol. 9(4), pp. 235-245, 2019.
- [15] Srachoom, C., “Application of Artificial Neural Network for Weather Forecast,” Chiang Mai University. Chiang Mai, 2007.
- [16] Sukawat, D., “Weather forecast Knowledge,” [Online], Available: <https://www.tmd.go.th/info/info.php?FileID=1>.
- [17] Wang, S. W., Feng, J., and Liu, G., “Application of seasonal time series model in the precipitation forecast,” *Mathematical and Computer Modelling*, vol. 58(3-4), pp. 677-683, 2013.
- [18] Wangdi, K., Singhasivanon, P., Silawan, T., Lawpool Sri, S., White, N. J., and Kaewkungwal, J., “Development of temporal modelling for forecasting and prediction of malaria infections using time-series and ARIMAX analyses: a case study in endemic districts of Bhutan,” *Malaria Journal*, vol. 9(1), pp. 1-9, 2010.
- [19] Yang, S., Yu, X., and Zhou, Y., “LSTM and GRU neural network performance comparison study: Taking Yelp review dataset as an example,” 2020 International workshop on electronic communication and artificial intelligence (IWECAI), 2020.