

การพยากรณ์การเกิดพายุฝนฟ้าคะนองระยะปัจจุบัน ด้วยเทคนิคการเรียนรู้ของเครื่อง กรณีศึกษาสนามบินภาคเหนือของประเทศไทย

Thunderstorm Nowcasting Using Machine Learning Techniques: A Case Study of the Airports in Northern Thailand

ชญาณิษฐ์ เมืองสง¹ ขจิตพรณ กฤตพลวิมาน^{1*} และรัชกฤช ธนพัฒน์ดล²

Chayanit Muangsong¹, Khajitpan Kritpolviman^{1*} and Ratchakrit Tanapattanadol²

¹แขนงวิชาเทคโนโลยีสารสนเทศและการสื่อสาร สาขาวิชาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช

¹Information and Communication Technology, School of Science and Technology,

Sukhothai Thammathirat Open University

²สำนักทะเบียนและวัดผล มหาวิทยาลัยสุโขทัยธรรมาธิราช

²Office of Registration Records and Evaluation, Sukhothai Thammathirat Open University

*Corresponding author: Khajitpan.mak@stou.ac.th

Received: December 7, 2023

Revised: February 13, 2024

Accepted: February 19, 2024

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อ (1) สร้างแบบจำลองในการพยากรณ์การเกิดพายุฝนฟ้าคะนองระยะปัจจุบัน และ (2) ประเมินประสิทธิภาพของแบบจำลองในการพยากรณ์การเกิดพายุฝนฟ้าคะนองระยะปัจจุบัน ที่มีระยะเวลาพยากรณ์ 1 และ 2 ชั่วโมงข้างหน้า สำหรับสนามบินภาคเหนือของประเทศไทยด้วยเทคนิคการเรียนรู้ของเครื่อง โดยชุดข้อมูลนำเข้า คือ ข้อมูลอุตุนิยมวิทยา ระหว่างเดือนมกราคม พ.ศ. 2558 ถึงเดือนธันวาคม พ.ศ. 2565 รวมทั้งสิ้น 364,382 ชุดข้อมูล และทำการแบ่งข้อมูลทั้งหมดออกเป็น 5 กลุ่มข้อมูล คือ กลุ่มข้อมูลรวมสนามบินภาคเหนือ กลุ่มข้อมูลสนามบินภาคเหนือตอนบน กลุ่มข้อมูลสนามบินภาคเหนือตอนล่าง กลุ่มข้อมูลสนามบินที่มีภูเขาล้อมรอบ 1-2 ด้าน และกลุ่มข้อมูลสนามบินที่มีภูเขาล้อมรอบ 3-4 ด้าน โดยพยากรณ์การเกิดพายุฝนฟ้าคะนองระยะปัจจุบันด้วยเทคนิคการเรียนรู้ของเครื่อง และสร้างแบบจำลองการจำแนกด้วยอัลกอริทึมมาตรฐาน 3 อัลกอริทึม ได้แก่ Naive Bayes Decision Tree และ Neural Networks และอัลกอริทึมมาตรฐานร่วมกับอัลกอริทึมการเรียนรู้แบบกลุ่ม ซึ่งได้แก่ Bagging AdaBoost และ Random Forest และทำการปรับความสมดุลของข้อมูลด้วยเทคนิคการสุ่มตัวอย่างเกิน SMOTE และเปรียบเทียบประสิทธิภาพของแบบจำลองด้วยวิธี 10-Fold Cross Validation ผลการวิจัยพบว่า แบบจำลองที่ดีที่สุดสำหรับกลุ่มข้อมูลรวมสนามบินภาคเหนือ ระยะเวลาการพยากรณ์ 1 ชั่วโมงข้างหน้า คือ Random Forest โดยมีค่า F-measure เท่ากับร้อยละ 76.45 และค่าพื้นที่ใต้กราฟเท่ากับ 0.888 และสำหรับระยะเวลาการพยากรณ์ 2 ชั่วโมงข้างหน้า คือ Neural Networks ร่วมกับ Bagging โดยมีค่า F-measure เท่ากับร้อยละ 30.62 และค่า AUC เท่ากับ 0.746

คำสำคัญ: พายุฝนฟ้าคะนอง พยากรณ์อากาศระยะปัจจุบัน การเรียนรู้ของเครื่อง แบบจำลองการจำแนก อัลกอริทึม การเรียนรู้แบบกลุ่ม

Abstract

The objectives of this research were to (1) create thunderstorm nowcasting models and (2) evaluate the efficiency of the nowcasting models for thunderstorm forecasting at northern Thailand's airports with forecasting times in one hour and two hours using machine learning techniques. The input datasets were obtained from the Meteorological Terminal Air Report (METAR) of nine airports located in northern Thailand. The hourly METAR reports generated by the Thai Meteorological Department during January 2015 and December 2022, covering a total of 364,382 datasets, were analyzed. All input data were divided into 5 groups: the group of all airports, airports in the upper northern, airports in the lower northern, airports surrounded by mountains, 1-2 sides, and 3-4 sides. By using machine learning algorithms, three classification standard algorithms were applied, including Naïve Bayes, Decision Tree, and Neural Networks. Three algorithmic methods with ensemble learning algorithms, including Bagging, AdaBoost, and Random Forest, were also used to create classification models. The Synthetic Minority Oversampling Technique (SMOTE) was used for balancing datasets, and the 10-fold cross-validation method was employed to evaluate predictive models. According to the results of the data group that included all airports for one-hour forecasting time, the Random Forest was the most effective model, with the F-measure value of 76.45% and the Area Under the Curve (AUC) of 0.888. For a two-hour forecasting time, the combination of neural networks and bagging was the most effective model, with an F-measure of 30.62% and an AUC of 0.746.

Keywords: thunderstorm, nowcasting, machine learning, classification model, ensemble learning algorithms



บทนำ

การพยากรณ์อากาศระยะปัจจุบัน หรือ Nowcasting เป็นการพยากรณ์สภาวะอากาศที่เกิดขึ้นในปัจจุบันสำหรับช่วงเวลาไม่เกิน 2 ชั่วโมง ซึ่งข้อมูลการพยากรณ์อากาศระยะปัจจุบันนี้มีความสำคัญอย่างยิ่งต่อการรายงานข่าวอากาศการบิน (Meteorological Terminal Air Report--METAR) เนื่องจากนำมาใช้สำหรับการวางแผนการบินของสายการบินภายในประเทศและการจัดการจราจรทางอากาศของหน่วยงานประจำท่าอากาศยาน (Thai Meteorological Department, 2023) โดยเฉพาะการเกิดปรากฏการณ์พายุฝนฟ้าคะนอง (thunderstorm) ที่มีสภาวะอากาศที่เลวร้ายที่สุด มีลมกระโชกและกระแสอากาศพัดขึ้นลงรุนแรง ส่งผลให้ในระหว่างการขึ้นลงของเครื่องบินนักบินไม่สามารถควบคุมหรือรักษาระดับบินไว้ได้เป็นเหตุให้เครื่องบิน

ประสบอุบัติเหตุได้ ดังนั้นการแจ้งเตือนเหตุการณ์ล่วงหน้าของการเกิดพายุฝนฟ้าคะนองระยะปัจจุบัน จึงช่วยเพิ่มประสิทธิภาพการตัดสินใจของนักบินหรือหน่วยงานต่าง ๆ ที่เกี่ยวข้องกับการจราจรทางอากาศโดยรอบสนามบิน สำหรับการพยากรณ์อากาศของกรมอุตุนิยมวิทยาในระยะปัจจุบันเพื่อออกข่าวอากาศการบินนั้นดำเนินการโดยผู้ปฏิบัติงานเป็นหลัก นั่นคือ นักอุตุนิยมวิทยาประจำสนามบินวิเคราะห์ข้อมูลสภาพอากาศที่ได้จากการตรวจวัดสารประกอบทางอุตุนิยมวิทยา โดยระบบตรวจอากาศอัตโนมัติที่ติดตั้งอยู่บริเวณสนามบิน แล้วพิจารณาร่วมกับข้อมูลอื่น ๆ เช่น แบบจำลองพยากรณ์เชิงตัวเลข เรดาร์ตรวจอากาศ และดาวเทียมตรวจอากาศ เป็นต้น จากนั้นรายงานข่าวอากาศการบินที่ได้จากการตรวจวัดในช่วงเวลานั้น ๆ และออกคำเตือนพยากรณ์แนวโน้มสภาพอากาศแปรปรวนหรือการเกิด

พายุฝนฟ้าคะนองที่จะเกิดขึ้นภายในระยะเวลา 2 ชั่วโมงข้างหน้า (trend forecast) ต่อท้ายข่าวอากาศการบินดังกล่าว โดยทำการประเมินผลความถูกต้องของการพยากรณ์จากข่าวอากาศการบินในชั่วโมงถัดไปว่ามีการแจ้งเตือนการเกิดพายุฝนฟ้าคะนองทันหรือไม่ ซึ่งความถูกต้องของการพยากรณ์ในแต่ละสถานีตรวจอากาศการบินขึ้นอยู่กับความรู้ความสามารถและประสบการณ์ในการทำงานของผู้ปฏิบัติงานแต่ละคน โดยการปฏิบัติงานของนักอุตุนิยมวิทยาประจำสนามบินของภาคเหนือมีความยากลำบากในการพยากรณ์อากาศประกอบข่าวอากาศการบินเนื่องจากภาคเหนือของประเทศไทยมีลักษณะภูมิประเทศโดยทั่วไปแบ่งออกเป็น 3 ลักษณะ คือ เทือกเขา ที่ราบระหว่างหุบเขา และที่ราบลุ่ม รวมถึงมีภูมิอากาศแบบร้อนชื้นสลับแล้ง สภาพภูมิอากาศทั่วไปอยู่ภายใต้อิทธิพลของลมมรสุมตะวันตกเฉียงใต้ และลมมรสุมตะวันออกเฉียงเหนือ นอกจากนี้ยังได้รับอิทธิพลจากพายุจรต่าง ๆ ที่พัดผ่านเข้ามาในพื้นที่ (Environment and Pollution Control Office 2, 2021) ด้วยลักษณะภูมิประเทศและภูมิอากาศดังกล่าวทำให้ลักษณะอากาศเกิดการเปลี่ยนแปลงอยู่ตลอดเวลาและเป็นเรื่องยากสำหรับการพยากรณ์อากาศระยะปัจจุบันได้อย่างถูกต้อง

ด้วยข้อจำกัดดังกล่าวจึงควรมีการใช้เทคนิคการเรียนรู้ของเครื่องช่วยในการพยากรณ์ ในปัจจุบันมีการวิจัยใช้เทคนิคการเรียนรู้ของเครื่องในการวิเคราะห์ข้อมูลเพื่อการพยากรณ์อากาศตามช่วงระยะเวลาต่าง ๆ ที่สนใจ และพบว่าพยากรณ์อากาศระยะสั้นลงทำให้ความแม่นยำเพิ่มขึ้น ตัวอย่างเช่น การวิจัยเกี่ยวกับการพยากรณ์ฝนระยะปัจจุบันด้วยเทคนิคการเรียนรู้ของเครื่องโดยอิงข้อมูลน้ำฝนตรวจวัดในกรุงเทพมหานครที่ทำการศึกษาระยะเวลาพยากรณ์ 30-180 นาทีข้างหน้า พบว่า ระยะเวลาพยากรณ์ 30 นาที มีความแม่นยำสูงสุดและเมื่อระยะเวลาพยากรณ์มากขึ้นทำให้ความแม่นยำลดลง โดยระยะเวลาพยากรณ์ 90 นาทีเป็นระยะเวลาพยากรณ์นานที่สุดที่สามารถยอมรับค่าความแม่นยำได้ (Srithagon, 2021) รวมถึงงานวิจัยการพยากรณ์ปริมาณน้ำฝนของประเทศไทยด้วยวิธีการเรียนรู้เชิงลึกสองแบบจำลองต่อกันสำหรับพยากรณ์ปริมาณฝนล่วงหน้า 1 ถึง 6 ชั่วโมงข้างหน้า พบว่า ระยะเวลาพยากรณ์ 1 ชั่วโมงมีค่าความคลาดเคลื่อนต่ำที่สุด และเมื่อระยะเวลาพยากรณ์นานขึ้นทำให้ค่าความคลาด

เคลื่อนสูงขึ้น (Manokij, 2019) อีกทั้งยังมีงานวิจัยเกี่ยวกับการประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่องเพื่อสร้างระบบเตือนภัยล่วงหน้าสำหรับการแจ้งเตือนฝนตกหนักในระยะเวลาพยากรณ์ระยะสั้นมาก โดยทำการศึกษาระยะเวลาพยากรณ์ 3 ชั่วโมงข้างหน้า ซึ่งพบว่า แบบจำลองที่นำเสนอสามารถคาดการณ์ฝนตกหนักในระยะเวลาพยากรณ์ 3 ชั่วโมงได้ดี (Moon et al., 2019) เป็นต้น

งานวิจัยนี้เลือกทำการศึกษาที่ระยะเวลาพยากรณ์ 1 และ 2 ชั่วโมงข้างหน้า โดยพิจารณาระยะเวลาตามข้อกำหนดขององค์การการบินพลเรือนระหว่างประเทศ (International Civil Aviation Organization--ICAO) ที่กำหนดให้สถานีตรวจอากาศการบินรายงานข่าวอากาศการบินทุกชั่วโมงและออกข่าวค่าเตือนพยากรณ์แนวโน้มสภาพอากาศในระยะเวลาพยากรณ์ 2 ชั่วโมง (Aeronautical Meteorology Division, 2015) ดังนั้นผู้วิจัยจึงเลือกพิจารณาระยะเวลาพยากรณ์ 1 และ 2 ชั่วโมงข้างหน้า เพื่อให้ครอบคลุมระยะเวลาการออกข่าวค่าเตือนพยากรณ์แนวโน้มการเกิดพายุฝนฟ้าคะนอง และเพื่อให้การพยากรณ์การเกิดพายุฝนฟ้าคะนองมีความแม่นยำยิ่งขึ้น ผู้วิจัยได้ทำการศึกษาสนามบินในพื้นที่ภาคเหนือจำนวน 9 สนามบินจากทั้งหมด 11 สนามบิน (ยกเว้นสนามบินตากและสนามบินเพชรบูรณ์ที่ใช้สำหรับทางราชการเท่านั้น) โดยทั้ง 9 สนามบินนี้มีการรายงานข่าวอากาศการบินรายชั่วโมงโดยกรมอุตุนิยมวิทยา และเนื่องจากลักษณะพื้นที่ภาคเหนือของประเทศไทยมีความหลากหลาย ผู้วิจัยจึงได้ทำการแบ่งพื้นที่ศึกษาออกเป็น 5 กลุ่มข้อมูล คือ (1) กลุ่มข้อมูลรวมสนามบินภาคเหนือ (2) กลุ่มข้อมูลสนามบินภาคเหนือตอนบน (3) กลุ่มข้อมูลสนามบินภาคเหนือตอนล่าง (4) กลุ่มข้อมูลสนามบินที่มีภูเขาล้อมรอบ 1-2 ด้าน และ (5) กลุ่มข้อมูลสนามบินที่มีภูเขาล้อมรอบ 3-4 ด้าน จากนั้นใช้เทคนิคการเรียนรู้ของเครื่องในการสร้างแบบจำลองเพื่อพยากรณ์การเกิดพายุฝนฟ้าคะนอง ซึ่งในงานวิจัยนี้มีการแบ่งคลาสเพื่อจำแนกชนิด Binary ที่มีเพียงสองสถานะเท่านั้น เพื่อป้องกันปรากฏการณ์การเกิดพายุฝนฟ้าคะนอง นั่นคือ ไม่เกิดพายุฝนฟ้าคะนอง (NO TSRA) และเกิดพายุฝนฟ้าคะนอง (TSRA) จากนั้นเปรียบเทียบประสิทธิภาพของแบบจำลองด้วยวิธี 10-Fold Cross Validation ทำการพิจารณาค่าความแม่นยำ (accuracy) ค่าความเที่ยงตรง (precision)

ค่าความระลึก (recall) ค่าความถ่วงดุล (f-measure) และค่าพื้นที่ใต้กราฟ (Area Under the Curve--AUC) โดยคัดเลือกแบบจำลองที่ดีที่สุดของแต่ละกลุ่มข้อมูลด้วยค่า F-measure สูงสุดและพิจารณาค่า AUC ของแบบจำลอง

วัตถุประสงค์การวิจัย

1. เพื่อสร้างแบบจำลองในการพยากรณ์การเกิดพายุฝนฟ้าคะนองใน 1 และ 2 ชั่วโมงข้างหน้า สำหรับการพยากรณ์อากาศระยะปัจจุบัน บริเวณสนามบินภาคเหนือของประเทศไทย

2. เพื่อประเมินประสิทธิภาพของแบบจำลองในการพยากรณ์การเกิดพายุฝนฟ้าคะนองใน 1 และ 2 ชั่วโมงข้างหน้า สำหรับการพยากรณ์อากาศระยะปัจจุบัน บริเวณสนามบินภาคเหนือของประเทศไทย

แนวคิดทฤษฎีที่เกี่ยวข้อง

ผู้วิจัยทำการศึกษาแนวคิดทฤษฎีและการทบทวนวรรณกรรมที่เกี่ยวข้องกับงานวิจัย มีรายละเอียด ดังนี้

1. การเรียนรู้ของเครื่อง (machine learning)

เป็นส่วนหนึ่งของปัญญาประดิษฐ์ (Artificial Intelligence) โดยอัลกอริทึมสำหรับการเรียนรู้ของเครื่อง จะทำการเรียนรู้โดยอัตโนมัติจากการเรียนรู้ในอดีตและทำการปรับปรุงการทำงานได้เองโดยไม่ต้องเขียนโปรแกรมกำกับไว้หรือสั่งการเครื่อง การเรียนรู้ของเครื่องแบ่งออกเป็น Supervised Unsupervised และ Reinforcement Learning โดยที่อัลกอริทึมของ Supervised Learning จะเป็นการเรียนรู้แบบมีผู้สอน แต่ Unsupervised Learning จะเป็นการเรียนรู้แบบไม่มีผู้สอน โดยทั้ง 2 แบบ สามารถให้ค่าความแม่นยำได้เป็นอย่างดี (Gupta et al., 2022) ซึ่งในงานวิจัยนี้เป็นการสร้างแบบจำลองแบบ Supervised learning ประเภทแบบจำลองการจำแนก (classification model) จากอัลกอริทึมมาตรฐาน (standard classification) และอัลกอริทึมมาตรฐานร่วมกับอัลกอริทึมการเรียนรู้แบบกลุ่ม (ensemble learning algorithms) ซึ่ง Tanwanon et al. (2017) ได้ให้ความหมายของอัลกอริทึมการเรียนรู้แบบกลุ่มว่า เป็นวิธีการอาศัยแบบจำลองการจำแนกข้อมูล

มากกว่าหนึ่งตัว แต่ละแบบจำลองการจำแนกข้อมูลจะมีกระบวนการทำงานของตนเอง และทุกแบบจำลองการจำแนกจะสร้างจากกลุ่มข้อมูลเดียวกัน เมื่อได้ผลลัพธ์ของแต่ละแบบจำลอง ก็จะนำผลลัพธ์ที่ได้เหล่านั้นผ่านวิธีการรวบรวม (combination integration หรือ vote) และตัดสินใจผลลัพธ์สุดท้าย (final decision) เพื่อให้ได้ผลลัพธ์การจำแนกข้อมูลเดียวกัน

ในงานวิจัยนี้มีขั้นตอนที่สำคัญของกระบวนการสร้างแบบจำลอง โดยทบทวนวรรณกรรมและอ้างอิงงานวิจัยต่างๆ ที่เกี่ยวข้อง ดังนี้

1.1 การปรับข้อมูลไม่สมดุล (imbalanced data)

ในกรณีที่มีข้อมูล 2 กลุ่มหรือมากกว่า ประกอบด้วยข้อมูลกลุ่มส่วนมาก (majority) และข้อมูลกลุ่มส่วนน้อย (minority) ถ้านำข้อมูลทั้งหมดเข้าสู่กระบวนการสร้างแบบจำลองจะส่งผลให้แบบจำลองทำงานผิดพลาดนำไปสู่ปัญหาการแบ่งข้อมูลผิดกลุ่ม (Jantarakongkul et al., 2014) ซึ่งงานวิจัยโดย Dey and Pratap (2023) ได้ศึกษาเปรียบเทียบการปรับข้อมูลไม่สมดุลโดยใช้เทคนิคการสุ่มตัวอย่างเกิน (oversampling) 3 วิธี ได้แก่ SMOTE Borderline-SMOTE และ ADASYN และใช้อัลกอริทึมการจำแนก ได้แก่ SVM KNN GaussianNB Decision Tree และ Random Forest ผลการวิจัยดังกล่าวพบว่า การใช้เทคนิค SMOTE ร่วมกับ Random Forest มีค่า Accuracy สูงสุด ขณะที่ Borderline-SMOTE และ ADASYN มีค่า Accuracy ที่ดีเมื่อใช้กับอัลกอริทึม SVM Decision tree และ Random Forest และงานวิจัยโดย Phaeobang and Sinsomboonthong (2019) ได้ศึกษาการปรับความไม่สมดุลของข้อมูลด้วย 4 วิธี คือ วิธีการสุ่มตัวอย่างเกิน วิธีการสุ่มตัวอย่างเกินโดยเทคนิค SMOTE วิธีการสุ่มตัวอย่างลด และวิธีการสุ่มตัวอย่างผสมผสาน พบว่า เทคนิค SMOTE เป็นวิธีปรับความไม่สมดุลที่ทำให้ได้ค่าประสิทธิภาพของแบบจำลองสูงสุด นอกจากนี้งานวิจัยโดย Sridee (2021) ได้ศึกษาการประยุกต์ใช้การเรียนรู้ของเครื่องเพื่อตรวจจับงานเสียที่เกิดขึ้นในกระบวนการทดสอบฮาร์ดดิสก์ไดรฟ์ลักษณะชุดข้อมูลที่นำเข้ามีจำนวนหลายแอตทริบิวต์และมีความไม่สมดุลสูง โดยใช้อัลกอริทึม SVM และ XGBoost ร่วมกับ 3 เทคนิคในการจัดการข้อมูลที่ไม่สมดุล ได้แก่ SMOTE Different Cost Learner และ SMOTE ร่วมกับ

Different Cost พบว่า การใช้เทคนิค SMOTE ร่วมกับ อัลกอริทึม XGBoost มีค่าประสิทธิภาพสูงสุด

1.2 การคัดเลือกคุณลักษณะ (feature selection)

ในวิธีดำเนินการวิจัยควรมีการคัดเลือกคุณลักษณะที่เกี่ยวข้องกับข้อมูลสำหรับการสร้างแบบจำลองการเรียนรู้ที่มีประสิทธิภาพ รวมถึงใช้หน่วยความจำและเวลาประมวลผลลดลง (Chen et al., 2019) งานวิจัยโดย Khunsuk and Thongkam (2020) ได้ศึกษากระบวนการคัดเลือกคุณลักษณะสำหรับเพิ่มประสิทธิภาพในการจำแนกความคิดเห็นของลูกค้า โดยใช้การคัดเลือกคุณลักษณะของข้อความจำนวน 3 เทคนิค คือ Chi-Square Information Gain และ Gain Ratio แล้วเปรียบเทียบประสิทธิภาพด้วยอัลกอริทึม Naive Bayes SVM KNN และ C4.5 พบว่า การคัดเลือกคุณลักษณะแบบ Information Gain ร่วมกับเทคนิค Naive Bayes ให้ผลที่ดีที่สุด

1.3 อัลกอริทึมสำหรับการสร้างแบบจำลอง

ที่ผ่านมาในงานวิจัยจำนวนหนึ่งที่ค้นหาอัลกอริทึมที่สามารถทำนายผลได้แม่นยำที่สุด อย่างไรก็ตามความเหมาะสมของอัลกอริทึมขึ้นอยู่กับลักษณะของชุดข้อมูลที่จะทำการศึกษาของแต่ละงานวิจัยและแนวทางการนำไปปรับใช้ เช่น งานวิจัยโดย Tanwanon et al. (2017) ได้ศึกษาการประยุกต์ใช้อัลกอริทึมการเรียนรู้แบบกลุ่มเพื่อพยากรณ์แนวโน้มของราคาหลักทรัพย์ในตลาดหลักทรัพย์แห่งประเทศไทย โดยใช้อัลกอริทึม Random Forest SVM KNN Naive Bayes และอัลกอริทึมการเรียนรู้แบบกลุ่มด้วยเทคนิค Boosting ผลการวิจัยพบว่า อัลกอริทึมการเรียนรู้แบบกลุ่มสามารถเพิ่มประสิทธิภาพในการพยากรณ์ได้ถึง 5%-14% และงานวิจัยโดย Nasritha et al. (2018) ได้ศึกษาการเปรียบเทียบเทคนิคการสุ่มตัวอย่างเพื่อการจำแนกข้อมูลที่ไม่สมดุล โดยใช้อัลกอริทึม CART Decision Tree Random Forest SVM และ Multilayer Perceptron ร่วมกับอัลกอริทึมการเรียนรู้แบบกลุ่ม ได้แก่ AdaBoost และ Bagging เพื่อสร้างแบบจำลองการจำแนก จากการวิจัยพบว่า เทคนิคการสุ่มตัวอย่างเกินสามารถปรับปรุงข้อมูลที่ไม่สมดุลได้ดีที่สุด และแบบจำลอง Random Forest

แบบจำลอง Random Forest ร่วมกับ AdaBoost และแบบจำลอง Random Forest ร่วมกับ Bagging มีค่าประสิทธิภาพการจำแนกที่ติดกับข้อมูลทั่วไคราะห์ในงานวิจัย

2. การทดสอบประสิทธิภาพของแบบจำลอง

การทดสอบประสิทธิภาพเป็นขั้นตอนที่บ่งชี้ว่าแบบจำลองที่สร้างขึ้นมีประสิทธิภาพการจำแนกข้อมูลที่ดีหรือไม่ ซึ่งมีเทคนิควิธีที่หลากหลาย เช่น งานวิจัยโดย Dey and Pratap (2023) ได้สร้างแบบจำลองเพื่อศึกษาเปรียบเทียบการใช้อัลกอริทึมที่หลากหลายร่วมกับการปรับข้อมูลไม่สมดุลด้วยเทคนิคต่าง ๆ โดยวัดประสิทธิภาพด้วยค่า Accuracy Precision Recall F-measure และ กราฟ Receiver Operating Characteristic--ROC นอกจากนี้ งานวิจัยของ Manokij (2019) ได้สร้างแบบจำลองการทำนายปริมาณน้ำฝนของประเทศไทยด้วยวิธีการเรียนรู้เชิงลึกสองแบบจำลองต่อกัน เพื่อทำนายปริมาณฝนล่วงหน้า 1 ถึง 6 ชั่วโมงข้างหน้า โดยข้อมูลที่ใช้ในงานวิจัยมีความไม่สมดุลระหว่างคลาสที่ฝนตกและฝนไม่ตก และเปรียบเทียบประสิทธิภาพด้วยค่า F-measure เพื่ออธิบายความสามารถในการทำนายคลาสฝนตก รวมถึงงานวิจัยโดย Moon et al. (2019) ที่สร้างแบบจำลองสำหรับการแจ้งเตือนฝนตกหนักในการพยากรณ์ระยะสั้นมาก โดยเปรียบเทียบประสิทธิภาพของแบบจำลองด้วยค่า F-measure เนื่องจากเป็นชุดข้อมูลไม่สมดุลและค่าอัตราส่วนของจำนวนเหตุการณ์ที่ทำนายได้ถูกต้องว่าฝนตก และงานวิจัยของ Senawong et al. (2022) ทำการจำแนกข้อมูลขนาดใหญ่โดยใช้การจัดกลุ่มด้วยวิธีเคมีนและวิธีการเรียนรู้เชิงลึก แล้ววัดประสิทธิภาพของแบบจำลองด้วยค่า Precision และค่า AUC

ในงานวิจัยนี้จะทำการวิเคราะห์ค่าต่าง ๆ จาก Confusion Matrix การเกิดพยากรณ์ผิดพลาดที่สามารถนำไปใช้กับการจำแนกคลาสชนิด Binary ของงานวิจัยนี้ได้ ดังตาราง 1 ที่แสดงค่า True Positive--TP False Positive--FP False Negative--FN และ True Negative--TN

ตาราง 1

Confusion Matrix

	แท้จริง “เกิดพายุฝนฟ้าคะนอง”	แท้จริง “ไม่เกิดพายุฝนฟ้าคะนอง”
ทำนายว่า “เกิดพายุฝนฟ้าคะนอง”	True Positive--TP	False Positive--FP
ทำนายว่า “ไม่เกิดพายุฝนฟ้าคะนอง”	False Negative--FN	True Negative--TN

จากตาราง 1 TP คือ จำนวนข้อมูลที่ทำนายว่าเกิดพายุฝนฟ้าคะนองแล้วเกิดขึ้นจริง
 FP คือ จำนวนข้อมูลที่ทำนายว่าเกิดพายุฝนฟ้าคะนองแต่แท้จริงไม่เกิด
 FN คือ จำนวนข้อมูลที่ทำนายว่าไม่เกิดพายุฝนฟ้าคะนองแต่แท้จริงเกิด
 TN คือ จำนวนข้อมูลที่ทำนายว่าไม่เกิดพายุฝนฟ้าคะนองแล้วไม่เกิดขึ้นจริง

และสามารถคำนวณได้ดังสมการที่ 1-4 Promla and Sanrach (2020)

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F\text{-measure} = \frac{2 \times (Precision \times Recall)}{(Precision+Recall)} \quad (4)$$

ค่า AUC คำนวณจากพื้นที่ใต้กราฟ ROC ซึ่งเป็นกราฟที่พล็อตระหว่างค่า Specificity (แกน x) และค่า Sensitivity (แกน y) เพื่อนำเสนอประสิทธิภาพของการจำแนกคลาสชนิด Binary โดยแบ่งผลลัพธ์ของการพยากรณ์ออกเป็น 2 กลุ่ม คือ กลุ่มที่เกิดเหตุการณ์และกลุ่มที่ไม่เกิดเหตุการณ์ (Lamprakhon, 2017) โดย AUC จะมีค่าอยู่ระหว่าง 0-1 เมื่อค่า AUC เข้าใกล้ 1 หมายความว่าแบบจำลองในภาพรวมสามารถจำแนกข้อมูลได้ดีมาก ซึ่งสามารถคำนวณได้ดังสมการที่ 5-7 (Senawong et al., 2022) ทั้งนี้เป็นที่น่าสังเกตว่าสมการที่ 3 และสมการที่ 6 ค่าตัวแปรที่ใช้คำนวณเป็นค่าเดียวกันแต่ทั้งสองสมการนี้ใช้แตกต่างกัน คือ สมการที่ 3 ใช้แสดงค่าความระลึกลำดับสำหรับหาประสิทธิภาพของแบบจำลอง ส่วนสมการที่ 6 นำเสนอ

ค่า True Positive Rate สำหรับพล็อตกราฟ ROC

$$AUC = \frac{Sensitivity + Specificity}{2} \quad (5)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (6)$$

$$Specificity = \frac{TN}{TN+FP} \quad (7)$$

และเกณฑ์การวัดประสิทธิภาพจากค่า AUC เป็น ดังนี้

“0.5 ≤ AUC < 0.7 คือ แบบจำลองมีประสิทธิภาพต่ำ”

“0.7 ≤ AUC < 0.8 คือ เกณฑ์มาตรฐานสำหรับแบบจำลองส่วนใหญ่”

“0.8 ≤ AUC < 0.9 คือ แบบจำลองทำงานได้ดี”

“AUC > 0.9 คือ แบบจำลองทำงานได้ดีมาก”

จากงานวิจัยต่าง ๆ ที่เกี่ยวข้อง ทำให้ผู้วิจัยมีแนวคิดพัฒนาและประเมินประสิทธิภาพของแบบจำลองในการพยากรณ์การเกิดพายุฝนฟ้าคะนองระยะปัจจุบันที่มีระยะเวลาพยากรณ์ 1 และ 2 ชั่วโมงข้างหน้า โดยมีการปรับสมดุลข้อมูลด้วยเทคนิค SMOTE เพื่อเพิ่มจำนวนข้อมูลกลุ่มส่วนน้อยให้มีจำนวนข้อมูลใกล้เคียงกับกลุ่มส่วนมาก โดยสุ่มตัวอย่างข้อมูลขึ้นมาหนึ่งตัวและหาค่าระยะห่างระหว่างข้อมูลที่เลือกกับข้อมูลที่ใกล้เคียงที่สุด (Chawla et al., 2002) และตามผลการวิจัยที่เสนอโดย Phaeobang and

Sinsomboonthong (2019) ซึ่งใช้เทคนิค SMOTE เพื่อให้แบบจำลองมีค่า Accuracy สูงสุด และผู้วิจัยดำเนินการคัดเลือกคุณลักษณะโดยใช้เทคนิค Information Gain และ Chi-Square สำหรับหาค่าน้ำหนักในการคัดเลือกคุณลักษณะที่เกี่ยวข้องและกำจัดคุณลักษณะที่ไม่เกี่ยวข้องออก หลังจากนั้นทำการสร้างแบบจำลองการจำแนกข้อมูลด้วยอัลกอริทึมมาตรฐาน ได้แก่ Naive Bayes Decision Tree และ Neural Networks และจากงานวิจัยโดย Tanwanon et al. (2017) พบว่า ใช้อัลกอริทึมการเรียนรู้แบบกลุ่มแล้วมีค่าประสิทธิภาพการจำแนกที่ดีขึ้น ดังนั้นผู้วิจัยจึงสร้างแบบจำลองอัลกอริทึมมาตรฐานร่วมกับอัลกอริทึมการเรียนรู้แบบกลุ่ม ได้แก่ Bagging AdaBoost และ Random Forest เพื่อศึกษาการเพิ่มประสิทธิภาพของแบบจำลองการพยากรณ์การเกิดพายุฝนฟ้าคะนองระยะปัจจุบันให้มีความถูกต้องและแม่นยำมากยิ่งขึ้น

กรอบแนวคิดการวิจัย

จากกรอบแนวคิดวิจัยดังแสดงในภาพ 1 มีการนำเข้าสู่ชุดข้อมูลข่าวอากาศการบิน จากนั้นชุดข้อมูลทั้งหมดจะถูกเตรียมข้อมูลก่อนการสร้างแบบจำลอง โดยแบ่งชุดข้อมูลข่าวอากาศการบินทั้งหมดออกเป็น 5 กลุ่มข้อมูลตามขอบเขตและลักษณะพื้นที่ของสนามบินภาคเหนือของประเทศไทย และแบ่งแต่ละกลุ่มข้อมูลเป็น 2 ส่วน คือ (1) ชุดข้อมูลฝึกฝน (training data set) และ (2) ชุดข้อมูลทดสอบ (testing data set) เพื่อเข้าสู่ขั้นตอนการสร้างแบบจำลอง โดยใช้อัลกอริทึมมาตรฐานสำหรับสร้างแบบจำลองเดียวกัน ได้แก่ Naive Bayes Decision Tree และ Neural Networks และอัลกอริทึมมาตรฐานร่วมกับอัลกอริทึมการเรียนรู้แบบกลุ่ม ได้แก่ Bagging AdaBoost และ Random Forest จากนั้นทดสอบประสิทธิภาพโดยเลือกแบบจำลองที่มีค่า F-measure สูงสุดของแต่ละกลุ่มข้อมูลและพิจารณาค่า AUC ของแบบจำลองดังกล่าว

วิธีดำเนินการวิจัย

วิธีดำเนินการวิจัยแบ่งเป็น 4 ขั้นตอน ดังนี้

1. การเตรียมข้อมูลก่อนสร้างแบบจำลอง

งานวิจัยนี้รวบรวมข้อมูลตั้งแต่วันที่ 1 ม.ค. 2558 ถึง 31 ธ.ค. 2565 รวมระยะเวลา 8 ปี ซึ่งเป็นข้อมูลราย

ชั่วโมงของข่าวอากาศการบินของสนามบินภาคเหนือ 9 สนามบินซึ่งรายงานโดยกรมอุตุนิยมวิทยา มีจำนวนข้อมูลทั้งหมด 364,382 ชุดข้อมูลแล้วทำการแบ่งออกเป็น 5 กลุ่มข้อมูล เนื่องจากปัจจัยทางด้านลักษณะภูมิประเทศที่มีภูเขาล้อมรอบส่งผลทำให้สารประกอบทางอุตุนิยมวิทยาเกิดการเปลี่ยนแปลงตลอดเวลา ได้แก่ ทิศทางลม ความเร็วลม การเกิดลมกระโชก สภาพอากาศปัจจุบัน ความชื้นสัมพัทธ์ และความกดอากาศ แล้ววิเคราะห์ตามขอบเขตและลักษณะพื้นที่ของสนามบินภาคเหนือของประเทศไทย ตามเกณฑ์การแบ่งพื้นที่ของกรมอุตุนิยมวิทยา ดังนี้ (Northern Met., 2023)

กลุ่มข้อมูลที่ 1 รวมสนามบินภาคเหนือ ได้แก่ ท่าอากาศยานแม่ฮ่องสอน ท่าอากาศยานนานาชาติเชียงใหม่ ท่าอากาศยานแม่ฟ้าหลวงเชียงราย ท่าอากาศยานลำปาง ท่าอากาศยานแพร่ ท่าอากาศยานน่านนคร ท่าอากาศยานนานาชาติแม่สอด ท่าอากาศยานสุโขทัย และท่าอากาศยานพิษณุโลก

กลุ่มข้อมูลที่ 2 สนามบินภาคเหนือตอนบน ได้แก่ ท่าอากาศยานแม่ฮ่องสอน ท่าอากาศยานนานาชาติเชียงใหม่ ท่าอากาศยานแม่ฟ้าหลวงเชียงราย ท่าอากาศยานลำปาง ท่าอากาศยานแพร่ และท่าอากาศยานน่านนคร

กลุ่มข้อมูลที่ 3 สนามบินภาคเหนือตอนล่าง ได้แก่ ท่าอากาศยานนานาชาติแม่สอด ท่าอากาศยานสุโขทัย และท่าอากาศยานพิษณุโลก

กลุ่มข้อมูลที่ 4 สนามบินที่มีภูเขาล้อมรอบ 1-2 ด้าน ได้แก่ ท่าอากาศยานสุโขทัย และท่าอากาศยานพิษณุโลก

กลุ่มข้อมูลที่ 5 สนามบินที่มีภูเขาล้อมรอบ 3-4 ด้าน ได้แก่ ท่าอากาศยานแม่ฮ่องสอน ท่าอากาศยานนานาชาติเชียงใหม่ ท่าอากาศยานแม่ฟ้าหลวงเชียงราย ท่าอากาศยานลำปาง ท่าอากาศยานแพร่ ท่าอากาศยานน่านนคร และท่าอากาศยานนานาชาติแม่สอด

ข่าว 1 และ ข่าว 2 เป็นตัวอย่างข่าวอากาศการบินในรูปแบบข้อความรหัส ชุดข้อมูลทั้งหมดประกอบด้วยข้อมูลดิบที่เป็นเนื้อข่าว 13 ส่วน ดังแสดงในตาราง 2 ดังนี้

ข่าว 1: METAR VTCC 011900Z 35010G25KT 1200 R36/1200 +TSRA FEW025CB BKN035 25/22 Q1013 BECMG 3000 TSRA=

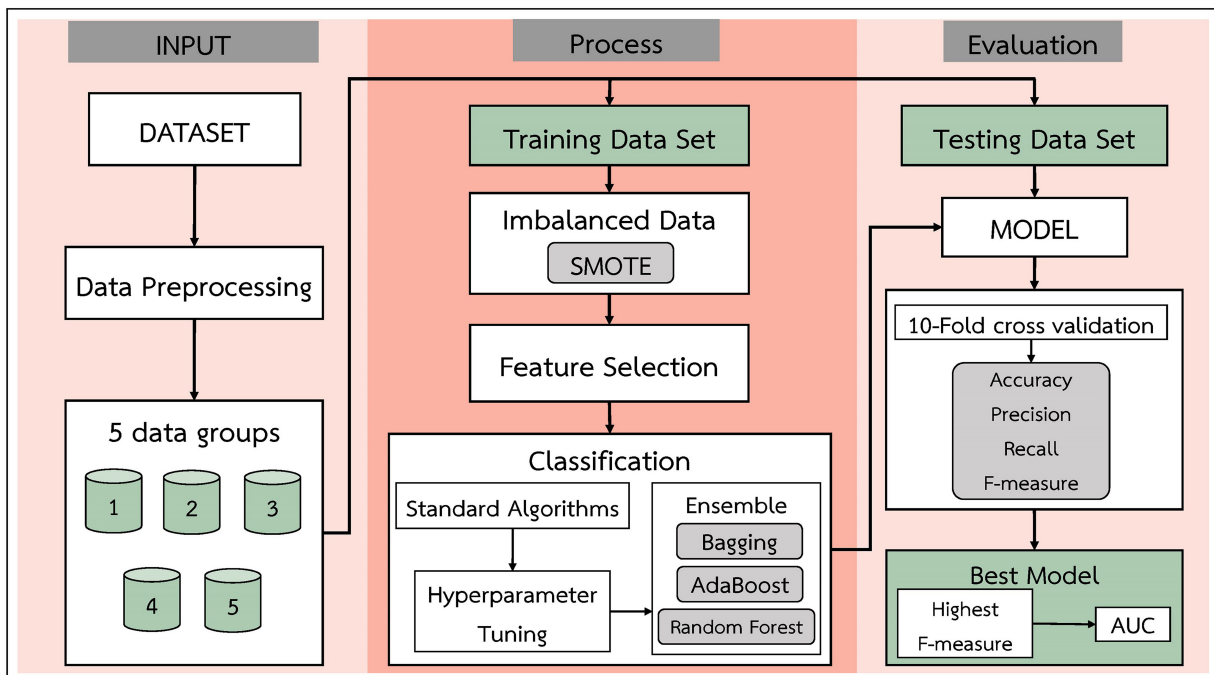
ข่าว 2: METAR VTCC 012000Z 07009KT
5000 TSRA FEW025CB BKN035 24/22 Q1013 RERA
TEMPO 2500 +TSRA=

จากตาราง 2 เนื้อหาข่าวอากาศการบินจะประกอบไปด้วย 13 ส่วน คือ (1) ประเภทของรายงาน (2) รหัสท่าอากาศยาน (3) เวลาตรวจอากาศการบิน (4) การตรวจอากาศการบินด้วยระบบอัตโนมัติ (5) ลมผิวพื้น (6) ทิศนวิสัย (7) พิสัยการมองเห็นบนทางวิ่ง (8) สภาพอากาศปัจจุบัน (9) เมฆ (10) อุณหภูมิอากาศและอุณหภูมิจุดน้ำค้าง (11) ความกดอากาศ (12) ข่าวสารเพิ่มเติม และ (13) การพยากรณ์แนวโน้มสภาวะอากาศบริเวณสนามบิน ซึ่งส่วนที่ (1)-(12) ได้จากการตรวจวัดค่าสารประกอบทางอุตุนิยมวิทยาจริงด้วยคนและเครื่องมือตรวจอากาศการบิน และส่วนที่ (13) เป็นส่วนต่อท้ายข่าวที่เป็นการพยากรณ์แนวโน้มสภาพอากาศบริเวณสนามบินที่จะเกิดขึ้นภายใน 2 ชั่วโมงข้างหน้า หรือที่เรียกว่าTrend Forecasts โดยข่าวแต่ละชั่วโมงจะถูกรายงานนาที่ที่ 0 และสามารถส่งข่าวล่าช้าได้ไม่เกินนาที่ที่ 3

เมื่อทำการรวบรวมข่าวอากาศการบินทั้งหมดแล้ว ข้อมูลทั้งหมดจะถูกปรับให้สมบูรณ์ด้วยโปรแกรม Microsoft Excel โดยทำความสะอาดข้อมูลเพื่อให้อยู่

ในรูปแบบที่ถูกต้องและสมบูรณ์ จากการวิเคราะห์พบว่าข้อมูลบางส่วนที่สูญหาย (missing value) เกิดจาก (1) ความผิดพลาดในขั้นตอนการรวบรวมข้อมูล จึงแทนที่ข้อมูลที่สูญหายด้วยค่าที่ถูกต้อง และ (2) ความบกพร่องของอุปกรณ์ตรวจวัดที่บันทึกค่าการตรวจวัดไม่ครบถ้วนสมบูรณ์ จึงได้ตัดชุดข้อมูลนั้นทิ้งไป จากนั้นทำการแปลงข้อมูลให้อยู่ในรูปแบบช่วงข้อมูลเพื่อให้มีการกระจายอย่างเป็นรูปแบบและลดเวลาในการประมวลผลของแบบจำลอง โดยรายละเอียดคุณลักษณะของข้อมูลที่ใช้ในการวิจัยแสดงดังตาราง 3

จากนั้นทั้ง 5 กลุ่มข้อมูลจะถูกแบ่งออกเป็น 2 ส่วนเพื่อนำเข้าสู่แบบจำลอง คือ (1) ชุดข้อมูลฝึกฝน ใช้ข้อมูลในปีพ.ศ. 2558-2563 รวม 6 ปี นำเข้าเพื่อใช้สอนและปรับค่าพารามิเตอร์ที่เหมาะสม และ (2) ชุดข้อมูลทดสอบ ใช้ข้อมูลในปีพ.ศ.2564-2565 รวม 2 ปี นำเข้าเพื่อทดสอบว่าสามารถทำงานได้ดีกับชุดข้อมูลที่ไม่เคยพบมาก่อนหรือไม่ ซึ่งในงานวิจัยนี้มีการแบ่งเป็น 2 คลาสเพื่อการจำแนกการเกิดพายุฝนฟ้าคะนอง นั่นคือ ไม่เกิดพายุฝนฟ้าคะนอง (NO TSRA) และเกิดพายุฝนฟ้าคะนอง (TSRA) โดยจำนวนชุดข้อมูลที่นำเข้าในการสร้างแบบจำลองแสดงดังตาราง 4



ภาพ 1 กรอบแนวคิดวิจัย

ตาราง 2

ตัวอย่างการแปลความข่าวอากาศการบิน

ส่วนที่	เนื้อหาข่าว	การแปลความข่าว
1	METAR	ประเภทข่าว METAR
2	VTCC	รหัสท่าอากาศยานนานาชาติเชียงใหม่
3	011900Z	วันที่ 1 เวลา 19.00 UTC หรือเวลา 02.00 น.
4	-	หากส่งผ่านระบบอัตโนมัติจะแสดง AUTO
5	35010G25KT	ทิศทางลม 350 องศา ความเร็วลม 10 นอต และมีลมกระโชกความเร็ว 25 นอต
6	1200	ทัศนวิสัย 1200 เมตร
7	R36/1200	พิธีกรรมการมองเห็นสูงสุดบนทางวิ่งหมายเลข 36 เท่ากับ 1,200 เมตร
8	+TSRA	เกิดพายุฝนฟ้าคะนองขนาดหนัก
9	FEW025CB BKN035	มีเมฆ Cumulonimbus ฐาน 2,500 ฟุต จำนวน 1-2 ส่วนจากทั้งหมด 8 ส่วน และมีเมฆฐาน 3,500 ฟุต จำนวน 5-7 ส่วน จากทั้งหมด 8 ส่วน
10	25/22	อุณหภูมิอากาศ เท่ากับ 25 °C และอุณหภูมิจุดน้ำค้าง เท่ากับ 22 °C
11	Q1013	ความกดอากาศหักแก้มสูงระดับน้ำทะเลปานกลาง เท่ากับ 1,013 มิลลิบาร์
12	RERA (ข่าวที่ 2)	ความรุนแรงของฝนที่ตกได้ลดลงจากข่าวก่อนหน้า
13	BECMG 3000 TSRA	คาดว่าใน 2 ชั่วโมงข้างหน้าจะเกิดพายุฝนฟ้าคะนองปานกลาง ทัศนวิสัยเท่ากับ 3,000 เมตร

ตาราง 3

รายละเอียดคุณลักษณะ (attribute) ของข้อมูลที่ใช้ในการวิจัย

คุณลักษณะ	คำอธิบาย	ประเภทข้อมูล	ค่าข้อมูล
W_direct	ทิศทางลม	Polynomial	CLAM, VRB, N, NE, E, SE, S, SW, W, NW
W_speed	ความเร็วลม	Polynomial	CLAM, LIGHT BREEZE, MODERATE BREEZE, STRONG BREEZE
GUST	การเกิดลมกระโชก	Binomial	YES, NO
WW	สภาพอากาศปัจจุบัน	Polynomial	NO, Light rain, Moderate Rain, Heavy Rain, TSRA, VC Rain, Obscurations
M_Cloud	เมฆพายุฝนฟ้าคะนอง	Polynomial	NO, CB, TCU
QNH	ความกดอากาศ	Polynomial	Low, MSL, High
Humidity	ความชื้นสัมพัทธ์	Binomial	Dry, High Moisture
Visibility	ทัศนวิสัย	Polynomial	A, B, C

ตาราง 4

จำนวนข้อมูลที้นำเข้าในการสร้างแบบจำลอง

กลุ่มข้อมูล	จำนวนชุดข้อมูล	Training Data Set				Testing Data Set			
		พยากรณ์ 1 ชั่วโมง		พยากรณ์ 2 ชั่วโมง		พยากรณ์ 1 ชั่วโมง		พยากรณ์ 2 ชั่วโมง	
		NO TSRA	TSRA	NO TSRA	TSRA	NO TSRA	TSRA	NO TSRA	TSRA
รวมสนามบินภาคเหนือ	364,382	258,773	2,023	258,782	2,014	102,547	1,039	102,556	1,030
ภาคเหนือตอนบน	260,676	185,667	1,531	185,659	1,539	72,653	825	72,650	828
ภาคเหนือตอนล่าง	103,706	73,106	492	73,123	475	29,894	214	29,906	202
ภูเขาล้อมรอบ 1-2 ด้าน	76,297	55,348	437	55,363	422	20,334	178	20,346	166
ภูเขาล้อมรอบ 3-4 ด้าน	288,085	203,425	1,586	203,419	1,592	82,213	861	82,210	864

2. การสร้างแบบจำลอง

2.1 การปรับชุดข้อมูลไม่สมดุล

ในงานวิจัยนี้มีชุดข้อมูลไม่สมดุล คือ คลาสไม่เกิดพายุฝนฟ้าคะนอง (NO TSRA) มีจำนวนมากกว่าคลาสเกิดพายุฝนฟ้าคะนอง (TSRA) และผู้วิจัยได้เลือกทำกระบวนการเพื่อจัดการปรับความไม่สมดุลของข้อมูลก่อนการคัดเลือกคุณลักษณะ เนื่องจากข้อมูลที่ไม่สมดุลกันทำให้แบบจำลองมีความคลาดเคลื่อนในการเรียนรู้ ซึ่งคุณลักษณะที่มีจำนวนน้อยอาจถูกมองข้ามและไม่ถูกพิจารณาว่าเป็นสิ่งสำคัญส่งผลต่อความแม่นยำของแบบจำลองได้ โดยปรับความสมดุลของข้อมูลด้วยเทคนิคการสุ่มตัวอย่างเกิน หรือเทคนิค SMOTE ในงานวิจัยนี้ใช้โอเพอเรเตอร์ Optimize Parameter (Grid) ช่วยในการหาค่าที่เหมาะสมสำหรับ K (number of nearest neighbors) ของแต่ละอัลกอริทึม

2.2 การคัดเลือกคุณลักษณะ

ในงานวิจัยนี้ผู้วิจัยเลือกวิธีการคัดเลือกคุณลักษณะที่มีความเกี่ยวข้องกับตัวแปรที่สนใจและกำจัดคุณลักษณะที่ไม่เกี่ยวข้องออกก่อนนำข้อมูลไปสร้างแบบจำลอง เพื่อหาค่าน้ำหนักความเกี่ยวข้องกับตัวแปรที่สนใจโดยใช้วิธี Information Gain และ Chi-Square ข้อมูลนำเข้าเพื่อสร้างแบบจำลองมี 8 คุณลักษณะ ได้แก่ ทิศทางลม ความเร็วลม การเกิดลมกระโชก ทิศวนวีสัย สภาพอากาศปัจจุบัน เมฆพายุฝนฟ้าคะนอง ความชื้นสัมพัทธ์ และความกดอากาศ

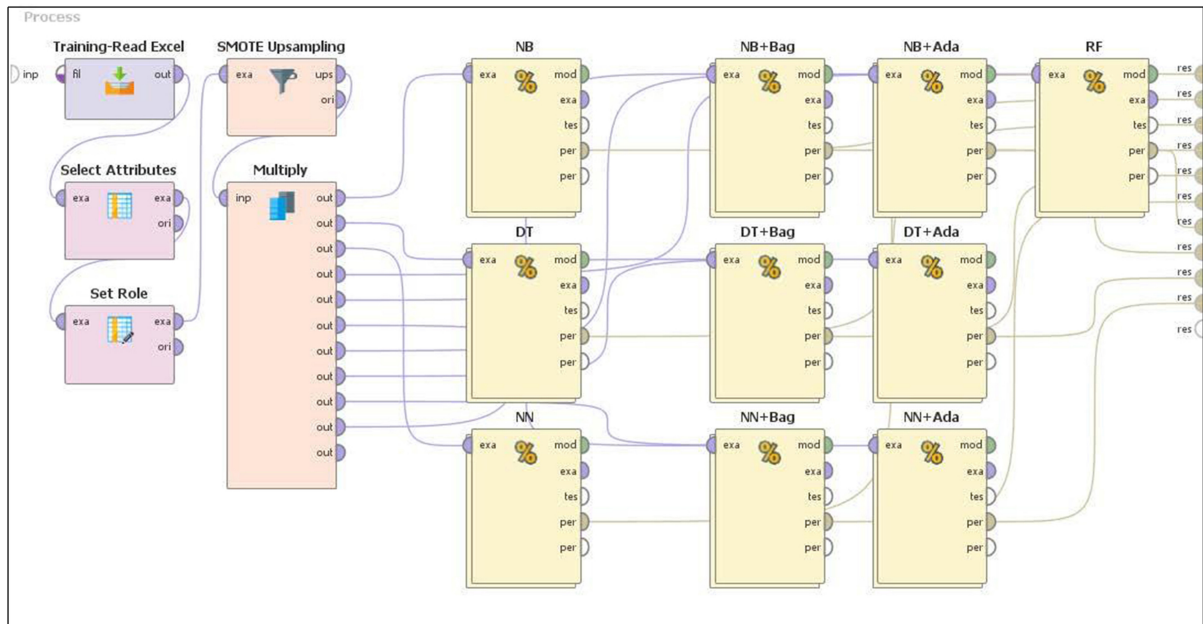
2.3 การปรับค่าพารามิเตอร์ที่เหมาะสมในงานวิจัย

ขั้นตอนการปรับค่าพารามิเตอร์มีผลต่อประสิทธิภาพของแบบจำลองในการเรียนรู้ของเครื่อง โดยจากงานวิจัยของ Thummikarat (2020) เกี่ยวกับการวินิจฉัยโรคพาร์กินสัน

โดยใช้เทคนิคการเรียนรู้ของเครื่อง ได้ใช้วิธีการค้นหาแบบกริด (Grid Search) เพื่อปรับค่าพารามิเตอร์ที่เหมาะสมที่สุดและให้ได้ผลลัพธ์ของแบบจำลองที่เหมาะสมโดยไม่เกิดการ Over Fitting หรือ Under Fitting ในงานวิจัยนี้ผู้วิจัยได้ใช้โอเพอเรเตอร์ Optimize Parameter (Grid) มาช่วยในการหาพารามิเตอร์ที่เหมาะสมของอัลกอริทึม ได้แก่ Decision Tree Neural Networks และ Random Forest ส่วน Naive Bayes ไม่มีการปรับค่าพารามิเตอร์

2.4 การสร้างแบบจำลองอัลกอริทึมมาตรฐานและอัลกอริทึมการเรียนรู้แบบกลุ่ม

จากภาพ 2 แบ่งการสร้างแบบจำลองเป็น 2 ส่วน คือ ส่วนที่ 1 การสร้างแบบจำลองโดยอัลกอริทึมมาตรฐาน ได้แก่ Naive Bayes Decision Tree และ Neural Networks และส่วนที่ 2 การสร้างแบบจำลองโดยอัลกอริทึมมาตรฐานร่วมกับอัลกอริทึมการเรียนรู้แบบกลุ่ม ได้แก่ Bagging AdaBoost และ Random Forest และในขั้นตอนนี้จะใช้ชุดข้อมูลฝึกฝนในการเรียนรู้ของแบบจำลอง และปรับค่าพารามิเตอร์ของอัลกอริทึมตามขั้นตอนก่อนหน้า ดังนั้นทั้ง 5 กลุ่ม ข้อมูลจึงถูกนำไปวิเคราะห์เพื่อหาแบบจำลองที่เหมาะสมโดยแบบจำลองที่ถูกสร้างได้แก่ (1) Naive Bayes--NB (2) Decision Tree--DT (3) Neural Networks--NN (4) Naive Bayes+Bagging--NB+Bag (5) Decision Tree+Bagging--DT+Bag (6) Neural Networks+Bagging--NN+Bag (7) Naive Bayes+AdaBoost--NB+Ada (8) Decision Tree+AdaBoost--DT+Ada (9) Neural Networks+AdaBoost--NN+Ada และ (10) Random Forest--RF



ภาพ 2 การสร้างแบบจำลองด้วย RapidMiner Studio

3. การทดสอบประสิทธิภาพแบบจำลอง

ในงานวิจัยนี้ใช้การทดสอบประสิทธิภาพแบบจำลองด้วยวิธี 10-Fold Cross Validation โดยนำเข้าสู่ชุดข้อมูลทดสอบ ซึ่งเป็นข้อมูลในปี พ.ศ.2564-2565 รวม 2 ปี เพื่อทดสอบแต่ละแบบจำลองว่าสามารถทำงานได้ดีกับชุดข้อมูลที่ไม่เคยพบมาก่อนหรือไม่ และทำการวัดประสิทธิภาพแบบจำลองเหล่านี้ด้วยค่า Accuracy Precision Recall และค่า F-measure

4. การคัดเลือกแบบจำลองที่ดีที่สุด

ในงานวิจัยนี้มีการคัดเลือกแบบจำลองที่ดีที่สุดสำหรับแต่ละกลุ่มข้อมูล โดยพิจารณาจากค่า F-measure สูงสุด ซึ่งจะสามารถวัดผลจากคลาส Positive ของแบบจำลองที่หามาได้จากการคำนวณค่า Precision และ Recall ทำให้บ่งชี้การทำนายคลาสเกิดพายุฝนฟ้าคะนอง (TSRA Class) ได้ จากนั้นพิจารณาค่า AUC ของแบบจำลองแต่ละกลุ่มข้อมูล

เครื่องมือที่ใช้ในการวิจัย

1. Microsoft Excel เพื่อจัดเก็บข้อมูล เตรียมข้อมูล และวิเคราะห์ข้อมูลด้วยแผนภูมิต่าง ๆ

2. RapidMiner Studio เป็นเครื่องมือที่ใช้ในการวิเคราะห์ข้อมูลและสร้างแบบจำลองเพื่อวิเคราะห์เชิงทำนาย ซึ่งมีโอเพอร์เรเตอร์ที่หลากหลายช่วยให้สามารถประมวลผลข้อมูลขนาดใหญ่ได้อย่างรวดเร็ว อีกทั้งยังมีความยืดหยุ่นในการปรับค่าพารามิเตอร์ของแบบจำลองตามความต้องการของผู้ใช้ และมีการแสดงผลข้อมูลในรูปแบบแผนภูมิทำให้ช่วยในการวิเคราะห์ข้อมูลได้ง่ายยิ่งขึ้น

ผลการวิจัย

1. ผลการสร้างแบบจำลอง

1.1 ผลการปรับชุดข้อมูลไม่สมดุล

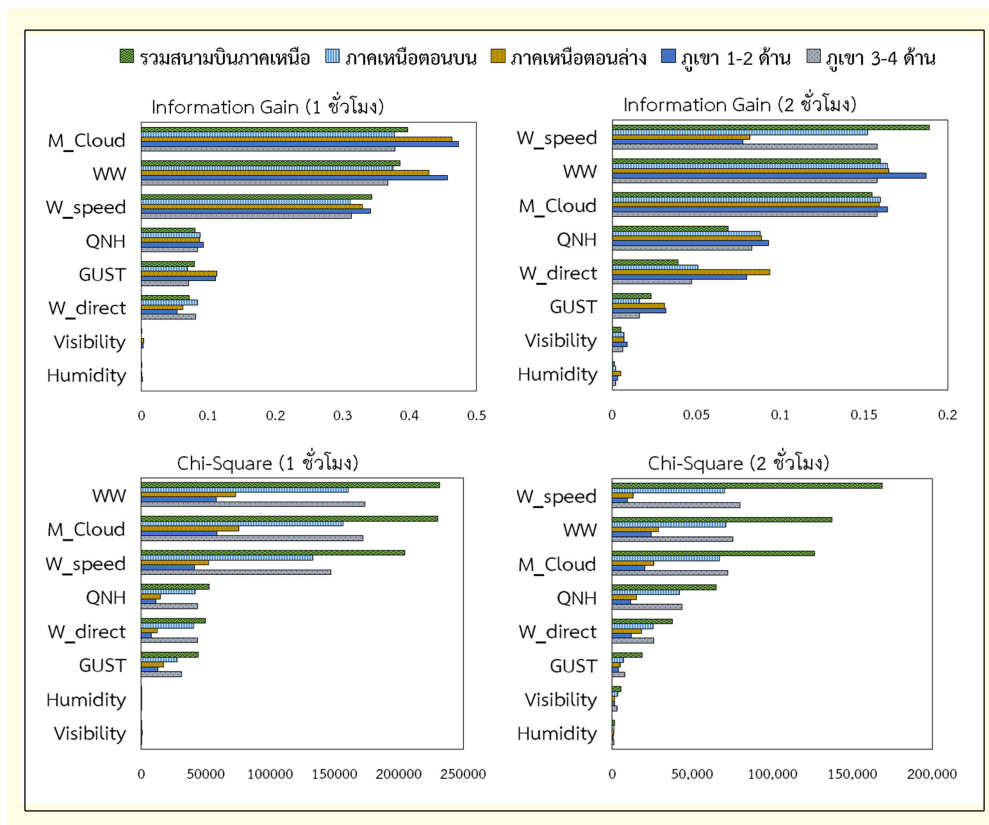
ผลการหาค่า K ที่เหมาะสมสำหรับ K-Nearest Neighbors โดยการปรับชุดข้อมูลไม่สมดุลด้วยเทคนิค SMOTE ด้วยโอเพอร์เรเตอร์ Optimize Parameter (Grid) พบว่า ได้ค่า K ของแต่ละอัลกอริทึมเท่ากับ 1 ในทุกแบบจำลอง

1.2 ผลการคัดเลือกคุณลักษณะ

จากภาพ 3 แผนภูมิแสดงค่าน้ำหนักการคัดเลือกคุณลักษณะด้วยวิธี Information Gain และ Chi-Square

พบว่า สำหรับระยะเวลาพยากรณ์ 1 และ 2 ชั่วโมงข้างหน้า นั้น ลำดับค่าน้ำหนักของคุณลักษณะมีความใกล้เคียงกันในทุก ๆ แบบจำลองทั้ง 2 ระยะเวลาพยากรณ์ อย่างไรก็ตามพบว่า คุณลักษณะความชื้นสัมพัทธ์ (humidity) และทัศนวิสัย (visibility) มีค่าน้ำหนักความเกี่ยวข้องกับตัวแปรน้อยที่สุดแตกต่างจากคุณลักษณะอื่น ๆ อย่างชัดเจน จึงได้ตัดทั้งสองคุณลักษณะดังกล่าวออก ทำให้เหลือเพียง 6 คุณลักษณะที่เป็นข้อมูลสำคัญนำเข้าสู่แบบจำลองการพยากรณ์การเกิดพายุฝนฟ้าคะนองระยะปัจจุบัน ได้แก่ ทิศทางลม (W_direct) ความเร็วลม (W_speed) การเกิดลมกระโชก (GUST) สภาพอากาศปัจจุบัน (WW) เมฆพายุฝนฟ้าคะนอง (M_Cloud) และความกดอากาศ (QNH)

1.3 ผลการปรับค่าพารามิเตอร์ที่เหมาะสมในงานวิจัย
ในงานวิจัยนี้ใช้โอเพอเรเตอร์ Optimize Parameter (Grid) ซึ่งเป็นวิธีการค้นหาแบบกริดสำหรับหาค่าพารามิเตอร์ที่เหมาะสม โดยใช้ค่าพารามิเตอร์ในตาราง 5 ในการสร้างแบบจำลองขั้นตอนต่อไปและท้ายที่สุดพบว่า สำหรับระยะเวลาพยากรณ์ 1 ชั่วโมงข้างหน้า ผลการปรับค่าพารามิเตอร์ไม่ทำให้แบบจำลองที่สร้างขึ้นเกิด Over Fitting หรือ Under Fitting เนื่องจากเมื่อนำเข้าข้อมูลทดสอบแล้วแบบจำลองยังสามารถทำนายการเกิดพายุฝนฟ้าคะนองได้ดี แต่สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมงข้างหน้า ถึงแม้จะปรับค่าพารามิเตอร์ต่าง ๆ แล้วก็ตามแต่แบบจำลองที่ทำการศึกษายังทำนายการเกิดพายุฝนฟ้าคะนองได้ไม่ดี



ภาพ 3 แผนภูมิแสดงค่าน้ำหนักการคัดเลือกคุณลักษณะด้วยวิธี Information Gain และ Chi-Square

ตาราง 5

ผลการปรับค่าพารามิเตอร์ที่เหมาะสมในงานวิจัย

อัลกอริทึม	พารามิเตอร์	รวมสนามบิน		ภาคเหนือ		ภาคเหนือ		ภูเขาล้อมรอบ		ภูเขาล้อมรอบ	
		ภาคเหนือ		ตอนบน		ตอนล่าง		1-2 ด้าน		3-4 ด้าน	
		1 ชม.	2 ชม.	1 ชม.	2 ชม.	1 ชม.	2 ชม.	1 ชม.	2 ชม.	1 ชม.	2 ชม.
Decision Tree	Criterion	gini_index	gini_index	gini_index	gini_index	gain_ratio	inform_gain	gini_index	inform_gain	gain_ratio	gini_index
	Maximum Depth	9	9	8	8	8	8	8	8	9	9
	minimal leaf size	2	2	2	2	2	2	2	4	2	2
	minimal size for split	9	2	2	2	2	2	2	2	2	2
Neural Networks	Training Cycle	282	138	45	166	80	291	294	255	262	152
	Learning Rate	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
	Momentum	0.7	0.9	0.5	0.8	0.6	0.6	0.8	0.9	0.7	0.7
Random Forest	Number of Trees	19	81	22	51	14	12	13	71	13	81
	Random Forest	gain_ratio	gain_ratio	gain_ratio	gain_ratio	gain_ratio	gain_ratio	inform_gain	gain_ratio	gain_ratio	gini_index
	Maximum Depth	7	7	12	12	10	10	12	12	8	8

2. ผลการทดสอบประสิทธิภาพของแบบจำลอง

ผลการทดสอบประสิทธิภาพในการพยากรณ์การเกิดพายุฝนฟ้าคะนอง สำหรับระยะเวลาพยากรณ์ 1 ชั่วโมง ข้างหน้า แสดงดังตาราง 6 เมื่อพิจารณาค่า Accuracy ของทุกชุดข้อมูล พบว่า มีค่า Accuracy มากกว่า 98% ในทุกอัลกอริทึม โดยเกิดจากการทำนายคลาสไม่เกิดพายุฝนฟ้าคะนองได้ถูกต้องซึ่งเป็นคลาสที่มีมากกว่านั่นเอง แต่หากพิจารณาจากค่า F-measure ที่สามารถบ่งชี้เฉพาะการทำนายของคลาสการเกิดพายุฝนฟ้าคะนอง พบว่า ค่า F-measure สูงสุด สำหรับ (1) กลุ่มข้อมูลรวมสนามบินภาคเหนือ คือ Random Forest เท่ากับ 76.45% (2) กลุ่มข้อมูลภาคเหนือตอนบน คือ Neural Networks เท่ากับ 77.64% และ Neural Networks ร่วมกับ AdaBoost เท่ากับ 77.64% (3) กลุ่มข้อมูลภาคเหนือตอนล่าง คือ

Neural Networks เท่ากับ 70.69% Neural Networks ร่วมกับ Bagging เท่ากับ 70.69% และ Neural Networks ร่วมกับ AdaBoost เท่ากับ 70.69% (4) กลุ่มข้อมูลสนามบินที่มีภูเขาล้อมรอบ 1-2 ด้าน คือ Neural Networks ร่วมกับ Bagging เท่ากับ 69.12% และ (5) กลุ่มข้อมูลสนามบินที่มีภูเขาล้อมรอบ 3-4 ด้าน คือ Random Forest เท่ากับ 77.53%

ผลการทดสอบประสิทธิภาพในการพยากรณ์การเกิดพายุฝนฟ้าคะนอง สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมง ข้างหน้า แสดงดังตาราง 7 เมื่อพิจารณาค่า Accuracy ของทุกชุดข้อมูล พบว่า มีค่า Accuracy มากกว่า 98% ในทุกอัลกอริทึม โดยเกิดจากการทำนายคลาสไม่เกิดพายุฝนฟ้าคะนองได้ถูกต้องซึ่งเป็นคลาสที่มีมากกว่านั่นเอง และเมื่อพิจารณาจากค่า F-measure ซึ่งสามารถบ่งชี้เฉพาะ

ถึงการทำนายของคลาสการเกิดพายุฝนฟ้าคะนอง พบว่าแบบจำลองที่ทำการศึกษาในครั้งนี้มีค่า F-measure ของแต่ละชุดข้อมูลต่ำไม่ถึง 50% นั่นคือ ยังไม่สามารถทำนายการเกิดพายุฝนฟ้าคะนอง สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมงข้างหน้าได้อย่างถูกต้อง

3. ผลการคัดเลือกแบบจำลองที่ดีที่สุด

การคัดเลือกแบบจำลองที่ดีที่สุดจะพิจารณาผลการทดสอบประสิทธิภาพของแบบจำลองจากตาราง 6 และ 7 เพื่อทำการเปรียบเทียบค่า F-measure สูงสุดของแต่ละกลุ่มข้อมูล โดยบางกลุ่มข้อมูลที่อัลกอริทึมมีค่า F-measure เท่ากันนั้น ผู้วิจัยจะพิจารณาค่า AUC ที่มาก

กว่าเพื่อเลือกเป็นแบบจำลองที่ดีที่สุดสำหรับกลุ่มข้อมูลนั้น โดยสรุปผลการคัดเลือกแบบจำลองที่ดีที่สุดของแต่ละกลุ่มข้อมูลที่ถูกนำมาวิเคราะห์ตามลักษณะพื้นที่สนามบินแสดงดังตาราง 8 ซึ่งสำหรับระยะเวลาพยากรณ์ 1 ชั่วโมง พบว่ามีค่า F-measure ระหว่าง 69.12-77.64% และค่า AUC ระหว่าง 0.8-0.9 ซึ่งจัดว่าเป็นแบบจำลองที่ทำงานได้ดี (Senawong et al., 2022) แต่สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมง ในทุก ๆ อัลกอริทึมยังมีประสิทธิภาพการทำนายการเกิดพายุฝนฟ้าคะนองได้ไม่ดีเท่าที่ควร โดยพิจารณาจากค่า F-measure ที่มีค่าระหว่าง 21.39-30.62% และเมื่อพิจารณาค่า AUC พบว่า เป็นแบบจำลองที่มีประสิทธิภาพต่ำใกล้เคียงกับแบบจำลองมาตรฐานทั่วไปเท่านั้น

ตาราง 6

ผลการทดสอบประสิทธิภาพของแบบจำลองสำหรับระยะเวลาพยากรณ์ 1 ชั่วโมง

กลุ่มข้อมูล	Performance	แบบจำลอง									
		อัลกอริทึมมาตรฐาน					อัลกอริทึมการเรียนรู้แบบกลุ่ม				
		NB	DT	NN	NB+Bag	DT+Bag	NN+Bag	NB+Ada	DT+Ada	NN+Ada	RF
รวมสนามบินภาคเหนือ	Accuracy	98.88	99.49	99.61	98.88	99.49	99.61	98.88	99.49	99.61	99.61
	Precision	45.99	96.90	97.31	45.99	96.90	97.32	45.99	96.90	97.31	97.32
	Recall	64.10	51.11	62.75	64.10	51.11	62.85	64.10	51.11	62.75	62.95
	F-measure	53.56	66.92	76.30	53.56	66.92	76.37	53.56	66.92	76.30	76.45
	AUC	0.884	0.856	0.888	0.884	0.856	0.851	0.859	0.758	0.886	0.888
ภาคเหนือตอนบน	Accuracy	98.77	99.50	99.57	98.78	99.50	99.58	98.77	99.50	99.57	99.58
	Precision	46.50	96.15	93.66	46.94	96.13	96.15	46.50	96.15	93.66	95.40
	Recall	66.06	57.45	66.30	66.06	57.21	57.58	66.06	57.45	66.30	65.33
	F-measure	54.58	71.93	77.64	54.88	71.73	72.02	54.58	71.93	77.64	77.55
	AUC	0.891	0.869	0.862	0.891	0.873	0.880	0.867	0.790	0.831	0.896
ภาคเหนือตอนล่าง	Accuracy	99.07	99.67	99.68	99.07	99.67	99.68	99.07	99.67	99.68	99.67
	Precision	39.42	100.00	100.00	39.42	100.00	100.00	39.42	100.00	100.00	98.32
	Recall	57.48	53.74	54.67	57.48	53.74	54.67	57.48	53.74	54.67	54.67
	F-measure	46.77	69.91	70.69	46.77	69.91	70.69	46.77	69.91	70.69	70.27
	AUC	0.863	0.788	0.833	0.863	0.788	0.852	0.829	0.769	0.859	0.840
ภูเขา 1-2 ด้าน	Accuracy	98.87	99.59	99.59	98.87	99.50	99.59	98.87	99.59	99.59	99.58
	1-2 ด้าน	39.20	100.00	98.95	39.20	100.00	100.00	39.20	100.00	98.95	97.92
	Recall	55.06	52.25	52.81	55.06	42.13	52.81	55.06	52.25	52.81	52.81
	F-measure	45.79	68.63	68.86	45.79	59.29	69.12	45.79	68.63	68.86	68.61
	AUC	0.864	0.786	0.814	0.864	0.736	0.867	0.832	0.761	0.764	0.797

ตาราง 6 (ต่อ)

กลุ่มข้อมูล	Performance	แบบจำลอง									
		อัลกอริทึมมาตรฐาน					อัลกอริทึมการเรียนรู้แบบกลุ่ม				
		NB	DT	NN	NB+Bag	DT+Bag	NN+Bag	NB+Ada	DT+Ada	NN+Ada	RF
ภูเขา 3-4 ด้าน	Accuracy	98.89	99.49	99.60	98.89	99.49	99.61	98.89	99.49	99.60	99.61
	Precision	47.53	96.40	93.71	47.58	96.40	96.20	47.53	96.40	93.71	96.21
	Recall	65.97	52.85	65.74	66.09	52.85	64.69	65.97	52.85	65.74	64.92
	F-measure	55.25	68.27	77.27	55.32	68.27	77.36	55.25	68.27	77.27	77.53
	AUC	0.890	0.864	0.867	0.889	0.864	0.898	0.851	0.767	0.893	0.895

ตาราง 7

ผลการทดสอบประสิทธิภาพของแบบจำลองสำหรับระยะเวลาพยากรณ์ 2 ชั่วโมง

กลุ่มข้อมูล	Performance	แบบจำลอง									
		อัลกอริทึมมาตรฐาน					อัลกอริทึมการเรียนรู้แบบกลุ่ม				
		NB	DT	NN	NB+Bag	DT+Bag	NN+Bag	NB+Ada	DT+Ada	NN+Ada	RF
รวมสนามบินภาคเหนือ	Accuracy	98.25	99.14	99.15	98.25	99.15	99.09	98.25	99.14	99.15	99.14
	Precision	21.15	91.33	91.95	21.19	93.49	62.39	21.15	91.33	91.95	91.81
	Recall	27.96	15.34	15.53	28.06	15.34	20.29	27.96	15.34	15.53	15.24
	F-measure	24.08	26.27	26.58	24.14	26.36	30.62	24.08	26.27	26.58	26.14
	AUC	0.762	0.688	0.746	0.762	0.687	0.746	0.711	0.600	0.578	0.746
ภาคเหนือตอนบน	Accuracy	98.09	99.04	99.04	98.09	99.04	99.05	98.09	99.04	99.04	99.04
	Precision	23.34	95.00	91.89	23.32	93.01	95.04	23.34	95.00	91.89	93.01
	Recall	30.56	16.06	16.43	30.56	16.06	16.18	30.56	16.06	16.43	16.06
	F-measure	26.46	27.48	27.87	26.45	27.39	27.66	26.46	27.48	27.87	27.39
	AUC	0.775	0.705	0.761	0.775	0.705	0.762	0.725	0.582	0.727	0.758
ภาคเหนือตอนล่าง	Accuracy	98.62	99.40	99.40	98.61	99.40	99.40	98.62	99.40	99.40	99.40
	Precision	13.95	88.89	92.00	13.80	100.00	88.46	13.95	88.89	92.00	88.46
	Recall	20.30	11.88	11.39	20.30	10.89	11.39	20.30	11.88	11.39	11.39
	F-measure	16.53	20.96	20.26	16.43	19.64	20.18	16.53	20.96	20.26	20.18
	AUC	0.788	0.589	0.703	0.704	0.598	0.701	0.656	0.562	0.703	0.678
ภูเขา 1-2 ด้าน	Accuracy	98.36	99.28	99.27	98.36	99.28	99.25	98.36	99.28	99.27	99.26
	1-2 ด้าน	14.52	100.00	86.96	14.52	95.24	76.00	14.52	100.00	86.96	80.00
	Recall	21.08	10.84	12.05	21.08	12.05	11.45	21.08	10.84	12.05	12.05
	F-measure	17.20	19.57	21.16	17.20	21.39	19.90	17.20	19.57	21.16	20.94
	AUC	0.721	0.635	0.704	0.717	0.624	0.710	0.703	0.554	0.560	0.700

ตาราง 7 (ต่อ)

กลุ่มข้อมูล	Performance	แบบจำลอง									
		อัลกอริทึมมาตรฐาน					อัลกอริทึมการเรียนรู้แบบกลุ่ม				
		NB	DT	NN	NB+Bag	DT+Bag	NN+Bag	NB+Ada	DT+Ada	NN+Ada	RF
ภูเขา 3-4 ด้าน	Accuracy	98.21	99.11	99.11	98.21	99.11	99.11	98.21	99.11	99.11	99.11
	Precision	22.43	93.24	90.73	22.45	93.24	93.84	22.43	93.24	90.73	90.73
	Recall	29.51	15.97	15.86	29.51	15.97	15.86	29.51	15.97	15.86	15.86
	F-measure	25.49	27.27	27.00	25.50	27.27	27.13	25.49	27.27	27.00	27.00
	AUC	0.773	0.701	0.760	0.773	0.701	0.764	0.714	0.580	0.671	0.759

ตาราง 8

สรุปผลการคัดเลือกแบบจำลองที่ดีที่สุดของแต่ละพื้นที่สนามบิน

พื้นที่สนามบิน	แบบจำลองที่ดีที่สุด	การทดสอบประสิทธิภาพ		
		Accuracy	F-measure	AUC
ระยะเวลาพยากรณ์ 1 ชั่วโมงข้างหน้า				
รวมสนามบินภาคเหนือ	Random Forest	99.61	76.45	0.888
ภาคเหนือตอนบน	Neural Networks	99.57	77.64	0.862
ภาคเหนือตอนล่าง	Neural Networks+AdaBoost	99.68	70.69	0.859
ภูเขาล้อมรอบ 1-2 ด้าน	Neural Networks+Bagging	99.59	69.12	0.867
ภูเขาล้อมรอบ 3-4 ด้าน	Random Forest	99.61	77.53	0.895
ระยะเวลาพยากรณ์ 2 ชั่วโมงข้างหน้า				
รวมสนามบินภาคเหนือ	Neural Networks+Bagging	99.09	30.62	0.746
ภาคเหนือตอนบน	Neural Networks	99.04	27.87	0.761
ภาคเหนือตอนล่าง	Decision Tree	99.40	20.96	0.589
ภูเขาล้อมรอบ 1-2 ด้าน	Decision Tree + Bagging	99.28	21.39	0.624
ภูเขาล้อมรอบ 3-4 ด้าน	Decision Tree	99.11	27.27	0.701

การอภิปรายผล

วัตถุประสงค์ของงานวิจัยนี้ คือ เพื่อสร้างและประเมินประสิทธิภาพของแบบจำลองเพื่อพยากรณ์การเกิดพายุฝนฟ้าคะนองระยะปัจจุบัน ซึ่งที่ผ่านมาการพยากรณ์อากาศบริเวณสนามบินพิจารณาออกคำพยากรณ์โดยนักอุตุนิยมวิทยาที่ปฏิบัติงานประจำสนามบินซึ่งใช้ข้อมูลจากการตรวจวัดสภาพอากาศ พิจารณาร่วมกับข้อมูลแบบจำลองพยากรณ์เชิงตัวเลขแบบภาพรวมของพื้นที่ประเทศไทย และวิเคราะห์ข้อมูลจากดาวเทียมตรวจอากาศร่วมกับเรดาร์ตรวจอากาศ

ที่ติดตั้งบางพื้นที่นอกเขตสนามบิน โดยงานวิจัยนี้นำเสนอการวิเคราะห์ข้อมูลข่าวอากาศการบินของสนามบินในพื้นที่ภาคเหนือจากเครื่องมือตรวจอากาศระบบอัตโนมัติที่ติดตั้งบริเวณสนามบินแต่ละแห่งโดยตรงโดยใช้เทคนิคการเรียนรู้ของเครื่องเพื่อให้การพยากรณ์มีความแม่นยำมากขึ้น จากผลการวิจัยพบว่า สำหรับระยะเวลาพยากรณ์ 1 ชั่วโมงข้างหน้าสำหรับพื้นที่รวมสนามบินภาคเหนือทั้งหมดคุณลักษณะที่มีความสำคัญมากที่สุด 3 ลำดับแรกของการพยากรณ์การเกิดพายุฝนฟ้าคะนองระยะปัจจุบันในพื้นที่

สนามบินภาคเหนือของประเทศไทย คือ เมฆฝนฟ้าคะนอง สภาพอากาศปัจจุบัน และความเร็วลม โดยแบบจำลอง Random Forest มีประสิทธิภาพการพยากรณ์การเกิดพายุฝนฟ้าคะนองได้ดีที่สุด เนื่องจาก แบบจำลอง Random Forest มีหลักการของการใช้ต้นไม้หลายต้นในการทำนาย และสรุปผลการทำนายด้วยการ Vote ผลลัพธ์ที่ถูกเลือกมากที่สุด ซึ่งเป็นคุณลักษณะที่ช่วยลดความผิดพลาดของคุณลักษณะของการเกิด Over Fitting ของแบบจำลอง และสำหรับระยะเวลาพยากรณ์ 2 ชั่วโมงข้างหน้าพบว่า คุณลักษณะที่มีความสำคัญมากที่สุดไม่แน่นอนโดยเปลี่ยนแปลงไปตามแต่ละกลุ่มข้อมูล ซึ่งทุกแบบจำลองมีประสิทธิภาพต่ำในการพยากรณ์การเกิดพายุฝนฟ้าคะนอง ซึ่งสอดคล้องกับผลการวิจัยของ Srithagon et al. (2021) ที่พบว่า แบบจำลองมีความแม่นยำต่ำลงเมื่อระยะเวลาพยากรณ์นานขึ้น สาเหตุเกิดจากชุดข้อมูลนำเข้าไม่สามารถสร้างรูปแบบการเรียนรู้ที่ถูกต้องให้แก่แบบจำลองได้ อาจเนื่องด้วยลักษณะภูมิอากาศประเทศไทยที่ตั้งอยู่ในเขตร้อนและสารประกอบทางอุตุนิยมวิทยาเปลี่ยนแปลงรวดเร็วแบบไม่มีรูปแบบแน่นอน ส่งผลให้การคาดการณ์สภาพอากาศล่วงหน้าในระยะเวลาพยากรณ์ 2 ชั่วโมงมีความแม่นยำน้อยลง และในงานวิจัยนี้แบ่งชุดข้อมูลทั้งหมดเป็น 5 กลุ่มข้อมูลตามขอบเขตและลักษณะพื้นที่ของสนามบิน พบว่า แบบจำลองที่ดีที่สุดของแต่ละกลุ่มข้อมูลให้ค่าความแม่นยำที่แตกต่างกันเพียงเล็กน้อยเท่านั้น โดยการเลือกแบบจำลองใดไปใช้ประโยชน์นั้นขึ้นอยู่กับวัตถุประสงค์ของผู้ใช้งาน เช่น การเลือกใช้อัลกอริทึม Random Forest ซึ่งมีประสิทธิภาพการจำแนกที่ดีเพียงพอสำหรับทุกกลุ่มพื้นที่สนามบิน ซึ่ง

ค่า AUC ก็ใกล้เคียงกันกับแบบจำลองที่ดีที่สุดในแต่ละกลุ่มข้อมูลในระดับที่ยอมรับได้ และมีความเร็วในการประมวลผลสูงกว่าการใช้อัลกอริทึม Neural Networks แต่หากต้องการประสิทธิภาพสูงสุดสำหรับพยากรณ์การเกิดพายุฝนฟ้าคะนองระยะปัจจุบันของแต่ละพื้นที่ ควรเลือกใช้แบบจำลองที่มีประสิทธิภาพสูงสุดของแต่ละกลุ่มข้อมูลของทั้ง 5 พื้นที่สนามบิน

ข้อเสนอแนะ

แบบจำลองที่ผู้วิจัยนำเสนอในบทความวิจัยนี้สามารถพัฒนาต่อยอดงานวิจัยในอนาคตได้ดังนี้ (1) ศึกษาระยะเวลาการพยากรณ์อื่น ๆ เพิ่มเติม เช่น การพยากรณ์ระยะสั้นมากสำหรับช่วงเวลาไม่เกิน 12 ชั่วโมง และการพยากรณ์ระยะสั้นสำหรับช่วงเวลา 1-3 วันเพื่อการแจ้งเตือนภัยที่ทันเวลา (2) เพิ่มพื้นที่การศึกษาในภูมิภาคอื่น ๆ ของประเทศไทย (3) พยากรณ์ระดับความรุนแรงของฝนที่ระดับต่าง ๆ ซึ่งการพัฒนางานวิจัยดังกล่าวจะเป็นประโยชน์อย่างยิ่งสำหรับการสนับสนุนหน่วยงานทางด้านการควบคุมการจราจรทางอากาศ และการวางแผนการบินของสายการบินในพื้นที่ขณะเกิดฝนตกหนักและพายุฝนฟ้าคะนอง นอกจากนี้ควรมีการทดสอบแบบจำลองหลายครั้งและวิเคราะห์หาค่าเฉลี่ยของค่าเบี่ยงเบนมาตรฐานของประสิทธิภาพแบบจำลอง สำหรับทดสอบความสามารถในการใช้งานทั่วไป (generalization) ของแบบจำลอง



References

- Aeronautical Meteorology Division. (2015). *Meteorological messages for aviation and dissemination*. Retrieved from <http://www2.aeromet.tmd.go.th/KnowledgeNews1.php>
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16(2002), 321-357. <https://doi.org/10.1613/jair.953>
- Chen, H., Li, T., Fan, X., & Luo, C. (2019). Feature selection for imbalanced data based on neighborhood rough sets. *Information Sciences*, 483(2019), 1-20. <https://doi.org/10.1016/j.ins.2019.01.041>

- Dey, I., & Pratap, V. (2023). A Comparative Study of SMOTE, Borderline-SMOTE, and ADASYN Oversampling Techniques using Different Classifiers. *Proceeding of 2023 3rd International Conference on Smart Data Intelligence* (pp. 294-302). Trichy, India: ICSMDI.
- Environment and Pollution Control Office 2. (2021). *Analysis of area potential to support water pollution risks in 17 provinces of the northern region*. Retrieved from <https://shorturl.asia/yEVaS>
- Gupta, V., Mishra, V. K., Singhal, P., & Kumar, A. (2022). An overview of supervised machine learning algorithm. *Proceeding of International Conference on System Modeling & Advancement in Research Trends* (pp. 87-92). Moradabad, India: IEEE.
- Jantarakongkul, B., Rasmeequan, S., Rimcharoen, S., Kulkasem, P., Chinnasarn, K., Rodtook, A., Vorraboot, P., & Onpans, J. (2014). *Optimal methods for classification of highly imbalanced datasets* (Research report). Bangkok: National Research Council of Thailand (NRCT). (in Thai)
- Khunsuk, T., & Thongkam, J. (2020). Feature selection method for improving customer reviews classification. *RMUTI Journal Science and Technology*, 13(1), 129-143. (in Thai)
- Lamprakhon, B. (2017). *A proposed method to compare areas under the roc curves for a single dataset: A case study of credit scoring model* (Master's thesis). Chulalongkorn University. Bangkok. (in Thai)
- Manokij, F. (2019). *Thailand's precipitation forecasting using deep learning approach model* (Master's thesis). Chulalongkorn University. Bangkok. (in Thai)
- Moon, S. H., Kim, Y. H., Lee, Y. H., & Moon, B. R. (2019). Application of machine learning to an early warning system for very short-term heavy rainfall. *Journal of Hydrology*, 568(2019), 1042-1054. <https://doi.org/10.1016/j.jhydrol.2018.11.060>
- Nasritha, K., Kerdprasop, K., & Kerdprasop, N. (2018). Comparison of sampling techniques for imbalanced data classification. *Journal of Applied Informatics and Technology*, 1(2018), 20-37. <https://doi.org/10.14456/jait.2018.2> (in Thai)
- Northern Met Center. (2023). *Today forecast*. Retrieved from <http://www.cmmet.tmd.go.th/forecast/todayforecast.jpg> (in Thai)
- Phaeobang, A., & Sinsomboonthong, S. (2019). Adjusting the imbalanced data with 5 classification methods. *Thai Journal of Science and Technology*, 9(4), 418-435. <https://doi.org/10.14456/tjst.2020.46> (in Thai)
- Promla, P., & Sanrach, C. (2020). The comparison of efficiency on the analysis of satisfaction on teaching performance using sentiment analysis by ensemble technique. *KKU Research Journal (Graduate Studies)*, 20(4), 140-149. (in Thai)
- Senawong, N., Wichitchan, S., & Kumphon, O. (2022). Large-scale data classification based on K-means clustering and deep learning. *The Journal of King Mongkut's University of Technology North Bangkok*, 32(4), 957-965. (in Thai)

- Sridee, A. (2021). *Applying machine learning techniques to detect failure in hard disk drive test process* (Master's thesis). Chulalongkorn University. Bangkok. (in Thai)
- Srithagon, S. (2021). *Rainfall nowcasting using learning machines based on rain gauge data in Bangkok Metropolis* (Master's thesis). Kasetsart University. Bangkok. (in Thai)
- Tanwanon, P., Yingsayree, C., Pongphet, W., & Kangkajit, T. (2017). Predict stock price trends in stock exchange of Thailand using ensemble model. *Journal of information Science and Technology*, 7(1), 12-21. <https://doi.org/10.14456/jist.2017.2> (in Thai)
- Thai Meteorological Department. (2023). *Meteorological knowledge*. Retrieved from <https://www.tmd.go.th/info/การพยากรณ์อากาศ-1>
- Thummikarat, H. (2020). *Parkinson's disease diagnosis using machine learning* (Master's thesis). Chulalongkorn University. Bangkok. (in Thai)

