

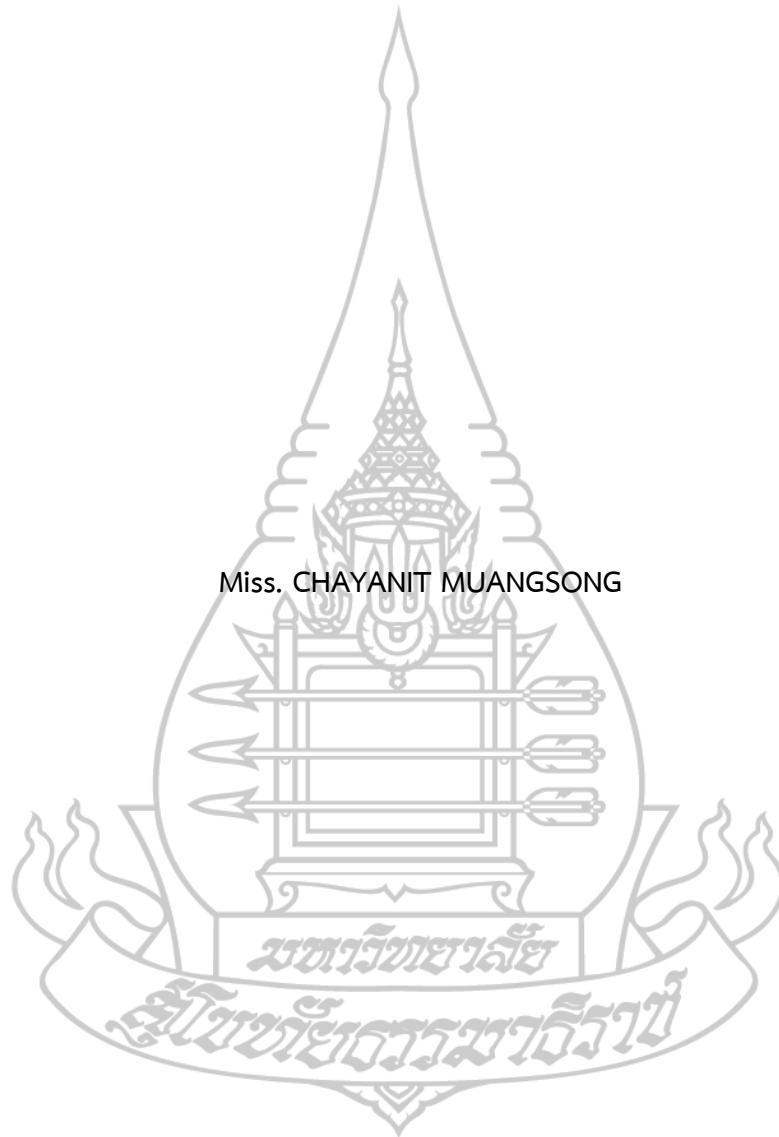
การพยากรณ์การเกิดพายุฝนฟ้าคะนองระยะปัจจุบันจากข่าวอากาศการบิน
ด้วยเทคนิคการเรียนรู้ของเครื่อง กรณีศึกษาสนามบินภาคเหนือของประเทศไทย



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
แขนงวิชาเทคโนโลยีสารสนเทศและการสื่อสาร สาขาวิชาวิทยาศาสตร์และเทคโนโลยี
มหาวิทยาลัยสุโขทัยธรรมมาธิราช

พ.ศ. 2566

Thunderstorm Nowcasting from Meteorological Terminal Air Report
Using Machine Learning Techniques:
A Case Study of the Airports in Northern Thailand



Miss. CHAYANIT MUANGSONG

A Thesis Submitted in Partial Fulfillment of the Requirements for
the Degree of Master of Science in Information and Communication Technology

School of Science and Technology
Sukhothai Thammathirath Open University

2023

หัวข้อวิทยานิพนธ์	การพยากรณ์การเกิดพายุฝนฟ้าคะนองระยะปัจจุบันจากข่าวอากาศการบิน ด้วยเทคนิคการเรียนรู้ของเครื่อง กรณีศึกษาสนามบินภาคเหนือของประเทศไทย
ชื่อและนามสกุล	นางสาวชญานิษฐ์ เมืองสง
แขนงวิชา / วิชาเอก	เทคโนโลยีสารสนเทศและการสื่อสาร
สาขาวิชา	วิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช
อาจารย์ที่ปรึกษา	1. ผู้ช่วยศาสตราจารย์ ดร.ขจิตพรธณ กฤตพลวิมาน
อาจารย์ที่ปรึกษาร่วม	2. อาจารย์ ดร.รัชกฤษ ชนพัฒน์ดล

วิทยานิพนธ์นี้ได้รับความเห็นชอบให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรระดับปริญญาโท เมื่อวันที่ 21 กุมภาพันธ์ พ.ศ. 2567

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.ดวงเพ็ญ เจตน์พิพัฒน์พงษ์)

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.ขจิตพรธณ กฤตพลวิมาน)

..... กรรมการ

(อาจารย์ ดร.รัชกฤษ ชนพัฒน์ดล)

..... ประธานกรรมการบัณฑิตศึกษา

(รองศาสตราจารย์ ดร.นราธิป ศรีราม)

ชื่อวิทยานิพนธ์ การพยากรณ์การเกิดพายุฝนฟ้าคะนองระยะปัจจุบันจากข่าวอากาศการบิน ด้วยเทคนิคการเรียนรู้ของเครื่อง กรณีศึกษาสนามบินภาคเหนือของประเทศไทย

ผู้วิจัย นางสาวชญานิษฐ์ เมืองสง รหัสนักศึกษา 2649600059

ปริญญา วิทยาศาสตรมหาบัณฑิต (เทคโนโลยีสารสนเทศและการสื่อสาร)

อาจารย์ที่ปรึกษา (1) ผู้ช่วยศาสตราจารย์ ดร.ชจิตพรธณ กฤตพลวิมาน (2) อาจารย์ ดร.รัชกฤษ ธนพัฒน์

ปีการศึกษา 2566

บทคัดย่อ

วิทยานิพนธ์นี้มีวัตถุประสงค์เพื่อ (1) สร้างแบบจำลองในการพยากรณ์การเกิดพายุฝนฟ้าคะนองระยะปัจจุบัน และ (2) ประเมินประสิทธิภาพของแบบจำลองในการพยากรณ์การเกิดพายุฝนฟ้าคะนองระยะปัจจุบัน ที่มีระยะเวลาพยากรณ์ 1 และ 2 ชั่วโมงข้างหน้า สำหรับสนามบินภาคเหนือของประเทศไทยด้วยเทคนิคการเรียนรู้ของเครื่อง โดยชุดข้อมูลนำเข้า คือ ข่าวอากาศการบินของสนามบินในพื้นที่ภาคเหนือจำนวน 9 สนามบินที่มีการรายงานข่าวอากาศการบินรายชั่วโมงโดยกรมอุตุนิยมวิทยา ระหว่างเดือนมกราคม พ.ศ. 2558 ถึงเดือนธันวาคม พ.ศ. 2565 รวมทั้งสิ้น 364,382 ชุดข้อมูล

ระเบียบวิธีวิจัยดำเนินการโดยแบ่งข้อมูลทั้งหมดออกเป็น 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 Random Forest Bagging และ AdaBoost และทำการปรับความสมดุลของข้อมูลด้วยเทคนิคการสุ่มตัวอย่างเกิน SMOTE และเปรียบเทียบประสิทธิภาพของแบบจำลองด้วยวิธี 10-Fold Cross Validation

ผลการวิจัยพบว่า แบบจำลองที่ดีที่สุดสำหรับกลุ่มข้อมูลรวมสนามบินภาคเหนือ ระยะเวลาการพยากรณ์ 1 ชั่วโมงข้างหน้า คือ Random Forest โดยมีค่า F-measure เท่ากับร้อยละ 76.62 ค่า Balanced Accuracy เท่ากับ 81.60 และค่า AUC เท่ากับ 0.918 และสำหรับระยะเวลาการพยากรณ์ 2 ชั่วโมงข้างหน้า คือ Neural Networks ร่วมกับ Bagging โดยมีค่า F-measure เท่ากับร้อยละ 27.72 ค่า Balanced Accuracy เท่ากับ 57.61 และค่า AUC เท่ากับ 0.681

คำสำคัญ พายุฝนฟ้าคะนอง, พยากรณ์อากาศระยะปัจจุบัน, การเรียนรู้ของเครื่อง, แบบจำลองการจำแนก, อัลกอริทึมการเรียนรู้แบบกลุ่ม

Thesis title: Thunderstorm Nowcasting from Meteorological Terminal Air Report Using Machine Learning Techniques: A Case Study of the Airports in Northern Thailand

Researcher: Miss. CHAYANIT MUANGSONG; **ID:** 2649600059;

Degree: Master of Science (Information and Communication Technology);

Thesis advisors: (1) Asst. Prof. Khajitpan Kritpolviman, Ph.D.; (2) Ratchakrit Tanapattanadol, Ph.D.; **Academic year:** 2023

Abstract

The objectives of this thesis 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.

In the research methodology, all 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 the Naive Bayes, Decision Tree, and Neural Networks. Three algorithm methods with ensemble learning algorithms including Random Forest, Bagging, and AdaBoost were also used to create Classification models. The oversampling technique, known as 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 76.62%, Balanced Accuracy of 81.60%, and the AUC of 0.918. For two-hour forecasting time, the combination of Neural Networks and Bagging was the most effective model, with the F-measure of 27.72%, Balanced Accuracy of 57.61%, and the AUC of 0.681.

Keywords: Thunderstorm, Nowcasting, Machine Learning, Classification model, Ensemble learning algorithms

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สามารถสำเร็จลุล่วงไปได้ด้วยดี เนื่องด้วยความช่วยเหลืออย่างดียิ่งจาก ผู้ช่วยศาสตราจารย์ ดร.ขจิตพรรณ กฤตพลวิมาน อาจารย์ที่ปรึกษาหลัก และอาจารย์ ดร.รัชกฤษ ธนพัฒน์ดล อาจารย์ที่ปรึกษาร่วม ผู้เป็นแรงผลักดันสำคัญตลอดจนให้ความรู้ คำปรึกษา คำแนะนำในการค้นคว้าหาข้อมูลเพื่อศึกษาเพิ่มเติมนำมาสู่การดำเนินการจนเสร็จสิ้นขั้นตอนการวิจัยได้ และสามารถปรับปรุงแก้ไขวิทยานิพนธ์ฉบับนี้กระทั่งเสร็จสมบูรณ์ ผู้วิจัยรู้สึกซาบซึ้งในความกรุณาของท่านและขอขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

ผู้วิจัยขอขอบพระคุณคณาจารย์สาขาวิชาวิทยาศาสตร์และเทคโนโลยี แขนงเทคโนโลยี สารสนเทศและการสื่อสารทุกท่านที่ได้ถ่ายทอดความรู้ซึ่งเป็นประโยชน์แก่ผู้วิจัย รวมถึงเพื่อน นักศึกษาปริญญาโททุกคนที่ให้กำลังใจ และคำแนะนำในการทำวิทยานิพนธ์ด้วยดีเสมอมา ตลอดจนผู้มีพระคุณทุกท่านที่ไม่ได้เอ่ยนามอันทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี

สุดท้ายนี้ ผู้วิจัยขอขอบพระคุณครอบครัว ได้แก่ คุณพ่อ คุณแม่ และผู้ช่วยศาสตราจารย์ ดร.โชติกา เมืองสง ผู้เป็นพี่สาว ที่ได้สนับสนุนและให้กำลังใจในการศึกษาและการทำงานวิจัยครั้งนี้จนสำเร็จ ประโยชน์อันใดอันจะพึงเกิดขึ้นจากวิทยานิพนธ์ฉบับนี้ผู้วิจัยขอมอบแก่ผู้มีพระคุณทุกท่านที่ได้กล่าวมา



นางสาวชฎานิชช์ เมืองสง

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ฎ
สารบัญรูปภาพ.....	ฏ
บทที่ 1 บทนำ.....	1
1. ความเป็นมาและความสำคัญของปัญหา.....	1
2. วัตถุประสงค์การวิจัย.....	3
3. กรอบแนวคิดการวิจัย.....	4
4. ขอบเขตของการวิจัย.....	5
5. นิยามศัพท์เฉพาะ.....	8
6. ประโยชน์ที่คาดว่าจะได้รับ.....	9
บทที่ 2 วรรณกรรมที่เกี่ยวข้อง.....	10
1. การพยากรณ์อากาศ.....	10
1.1. ความหมายของการพยากรณ์อากาศ.....	10
1.2. ระยะเวลาของการพยากรณ์อากาศ.....	11
1.3. ข้อมูลอุตุนิยมวิทยาจากการตรวจวัด.....	11
2. ปραกฏการณ์พายุฝนฟ้าคะนอง.....	12
2.1. สาเหตุของการเกิดพายุฝนฟ้าคะนอง.....	13
2.2. ลักษณะอากาศร้ายเนื่องจากพายุฝนฟ้าคะนอง.....	13
2.3. พายุฝนฟ้าคะนองกับผลกระทบต่อการบิน.....	14
3. สนามบินบริเวณภาคเหนือที่ทำการศึกษา.....	14
3.1. ทำอากาศยานแม่ฮ่องสอน.....	15
3.2. ทำอากาศยานนานาชาติเชียงใหม่.....	16

สารบัญ (ต่อ)

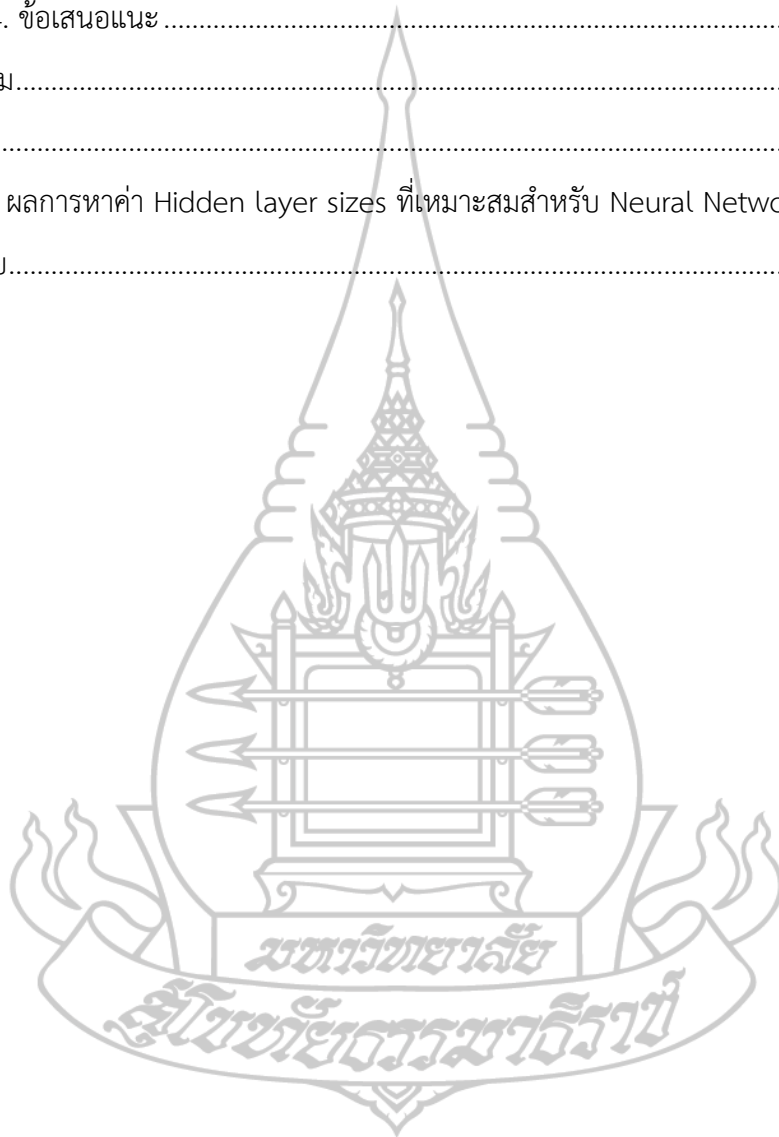
	หน้า
3.3. ทำอากาศยานแม่ฟ้าหลวงเชียงราย	17
3.4 ทำอากาศยานลำปาง	18
3.5 ทำอากาศยานแพร่.....	19
3.6 ทำอากาศยานน่านนคร.....	19
3.7 ทำอากาศยานนานาชาติแม่สอด	20
3.8 ทำอากาศยานสุโขทัย.....	21
3.9 ทำอากาศยานพิษณุโลก.....	22
4. ข่าวกาศการบิน	23
5. เทคนิคการเรียนรู้ของเครื่อง	32
5.1 การเตรียมข้อมูลก่อนสร้างแบบจำลอง (Data Preprocessing).....	32
5.2 การปรับข้อมูลไม่สมดุล (Imbalanced Data)	33
5.3 การคัดเลือกคุณลักษณะ (Feature Selection).....	35
5.4 การปรับค่าพารามิเตอร์ที่เหมาะสมของอัลกอริทึม	37
5.5 แบบจำลองการจำแนก (Classification Model)	38
5.6 อัลกอริทึมที่ใช้สำหรับการสร้างแบบจำลองการจำแนก.....	39
5.7 การประเมินประสิทธิภาพแบบจำลอง (Performance Evaluation).....	44
6. งานวิจัยที่เกี่ยวข้อง	48
6.1 งานวิจัยด้านการพยากรณ์อากาศ ด้วยเทคนิคการเรียนรู้ของเครื่อง	48
6.2 งานวิจัยด้านอื่นๆที่เกี่ยวข้องกับเทคนิคการเรียนรู้ของเครื่อง.....	51
บทที่ 3 วิธีดำเนินการวิจัย	56
1. เครื่องมือที่ใช้ในการวิจัย	56
2. การรวบรวมและเตรียมข้อมูลนำเข้า (Input)	56
2.1 การรวบรวมข้อมูล	56
2.2 การเตรียมข้อมูล	60
2.3 การแบ่งกลุ่มข้อมูล	66

สารบัญ (ต่อ)

	หน้า
2.4 การแบ่งชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบ	68
3. การสร้างแบบจำลอง (Process).....	69
3.1 การคัดเลือกอัลกอริทึม	69
3.2 โอเพอร์เรเตอร์ที่ใช้ในงานวิจัย.....	72
3.3 การปรับข้อมูลไม่สมดุล.....	75
3.4 การคัดเลือกคุณลักษณะ	82
3.5 การปรับค่าพารามิเตอร์ที่เหมาะสมในงานวิจัย	84
3.6 การสร้างแบบจำลองจากอัลกอริทึมมาตรฐาน.....	97
3.7 การสร้างแบบจำลองจากอัลกอริทึมการเรียนรู้แบบกลุ่ม	101
4. การประเมินประสิทธิภาพของแบบจำลอง (Evaluation).....	106
4.1 การวัดประสิทธิภาพของแบบจำลอง	106
4.2 การคัดเลือกแบบจำลองที่ดีที่สุด	107
4.3 การนำแบบจำลองไปใช้งานจริง	107
บทที่ 4 ผลการวิเคราะห์ข้อมูล	109
1. ผลการวิเคราะห์สำหรับการคัดเลือกคุณลักษณะ	109
2. ผลการปรับค่าพารามิเตอร์ที่เหมาะสมในงานวิจัย.....	118
2.1 สรุปผลการปรับค่าพารามิเตอร์สำหรับอัลกอริทึมมาตรฐาน	118
2.2 สรุปผลการปรับค่าสำหรับอัลกอริทึมการเรียนรู้แบบกลุ่ม	120
3. ผลการประเมินประสิทธิภาพแบบจำลอง	126
4. สรุปแบบจำลองที่เหมาะสมสำหรับแต่ละพื้นที่สนามบิน	133
5. การทดสอบนำแบบจำลองไปใช้งานจริง	135
บทที่ 5 สรุปการวิจัย อภิปรายผล และข้อเสนอแนะ	138
1. สรุปการวิจัย	138
2. อภิปรายผล.....	139
3. ข้อจำกัดของงานวิจัย	140

สารบัญ (ต่อ)

	หน้า
4. ข้อเสนอแนะ	140
บรรณานุกรม.....	143
ภาคผนวก.....	147
ก ผลการหาค่า Hidden layer sizes ที่เหมาะสมสำหรับ Neural Networks	147
ประวัติผู้วิจัย.....	158



สารบัญตาราง

	หน้า
ตารางที่ 1.1 จำนวนข้อมูลการบินร่วมกับ AWS ของ 5 สนามบินภาคเหนือ	6
ตารางที่ 2.1 ตัวอย่างการแปลความข่าวอากาศการบินของข่าว	31
ตารางที่ 2.2 Confusion Matrix สำหรับ Binary Classification	45
ตารางที่ 3.1 เวลาทำการของสถานีตรวจอากาศการบิน	58
ตารางที่ 3.2 จำนวนชุดข้อมูลข่าวอากาศการบินทั้งหมดแยกตามสถานีตรวจอากาศการบิน	59
ตารางที่ 3.3 จำนวนชุดข้อมูลนำเข้าทั้งหมดแบ่งตามกลุ่มข้อมูล	61
ตารางที่ 3.4 การแปลงข้อมูลที่เป็นตัวเลขให้อยู่ในรูปแบบของช่วงข้อมูล	63
ตารางที่ 3.5 รายละเอียดแอตทริบิวต์ของข้อมูลที่ใช้สำหรับทำการวิจัย	65
ตารางที่ 3.6 จำนวนข้อมูลที่นำเข้าสำหรับระยะเวลาพยากรณ์ 1 ชั่วโมง	68
ตารางที่ 3.7 จำนวนข้อมูลที่นำเข้าสำหรับระยะเวลาพยากรณ์ 2 ชั่วโมง	69
ตารางที่ 3.8 การทดสอบเบื้องต้นเพื่อคัดเลือกอัลกอริทึมที่เหมาะสมในงานวิจัย	70
ตารางที่ 3.9 การพิจารณาข้อจำกัดของแต่ละอัลกอริทึม	70
ตารางที่ 3.10 แสดงแบบจำลองทั้งหมดในงานวิจัย	72
ตารางที่ 3.11 โอเพอร์เรเตอร์สำหรับการสร้างแบบจำลองในงานวิจัย	73
ตารางที่ 3.12 ค่าที่เหมาะสมสำหรับ K (Number of Nearest Neighbors)	80
ตารางที่ 3.13 จำนวนข้อมูลข่าวอากาศการบินในปีพ.ศ. 2566 เพื่อทดสอบแบบจำลอง	108
ตารางที่ 4.1 ค่าน้ำหนักการคัดเลือกคุณลักษณะ สำหรับกลุ่มข้อมูลรวมสนามบินภาคเหนือ	110
ตารางที่ 4.2 ค่าน้ำหนักการคัดเลือกคุณลักษณะ สำหรับสนามบินภาคเหนือตอนบน	111
ตารางที่ 4.3 ค่าน้ำหนักการคัดเลือกคุณลักษณะ สำหรับสนามบินภาคเหนือตอนล่าง	112
ตารางที่ 4.4 ค่าน้ำหนักการคัดเลือกคุณลักษณะ สำหรับกลุ่มข้อมูลภูเขาล้อมรอบ 1-2 ด้าน	113
ตารางที่ 4.5 ค่าน้ำหนักการคัดเลือกคุณลักษณะ สำหรับกลุ่มข้อมูลภูเขาล้อมรอบ 3-4 ด้าน	114
ตารางที่ 4.6 สรุปผลการคัดเลือกคุณลักษณะของทุกแบบจำลอง	117
ตารางที่ 4.7 ผลการปรับค่าพารามิเตอร์ที่เหมาะสม สำหรับอัลกอริทึมมาตรฐาน	118
ตารางที่ 4.8 ผลการปรับค่าพารามิเตอร์ที่เหมาะสม สำหรับอัลกอริทึมการเรียนรู้แบบกลุ่ม	121
ตารางที่ 4.9 ผลการทดสอบประสิทธิภาพของแบบจำลอง สำหรับ 1 ชั่วโมง	127

สารบัญตาราง (ต่อ)

	หน้า
ตารางที่ 4.10 ผลการทดสอบประสิทธิภาพของแบบจำลอง สำหรับ 2 ชั่วโมง.....	131
ตารางที่ 4.11 สรุปผลการคัดเลือกแบบจำลองที่ดีที่สุดของแต่ละพื้นที่สนามบิน.....	134
ตารางที่ 4.12 ผลการทดสอบประสิทธิภาพของแบบจำลองจากการนำไปใช้จริง	135
ตารางที่ 4.13 เปรียบเทียบผลการทำนายจากการนำไปใช้จริงระยะเวลาพยากรณ์ 1 ชั่วโมง.....	136
ตารางที่ 4.14 เปรียบเทียบผลการทำนายจากการนำไปใช้จริงระยะเวลาพยากรณ์ 2 ชั่วโมง.....	137



สารบัญรูปภาพ

	หน้า
ภาพที่ 1.1 กรอบแนวคิดการวิจัย.....	4
ภาพที่ 2.1 ด้านหน้าของพายุฝนฟ้าคะนองจะมีอากาศไหลลงอย่างรุนแรง	13
ภาพที่ 2.2 สภาพพื้นที่บริเวณท่าอากาศยานแม่ฮ่องสอน	15
ภาพที่ 2.3 สภาพพื้นที่บริเวณท่าอากาศยานนานาชาติเชียงใหม่.....	16
ภาพที่ 2.4 สภาพพื้นที่บริเวณท่าอากาศยานแม่ฟ้าหลวงเชียงราย.....	17
ภาพที่ 2.5 สภาพพื้นที่บริเวณท่าอากาศยานลำปาง	18
ภาพที่ 2.6 สภาพพื้นที่บริเวณท่าอากาศยานแพร่.....	19
ภาพที่ 2.7 สภาพพื้นที่บริเวณท่าอากาศยานน่านนคร.....	20
ภาพที่ 2.8 สภาพพื้นที่บริเวณท่าอากาศยานนานาชาติแม่สอด	21
ภาพที่ 2.9 สภาพพื้นที่บริเวณท่าอากาศยานสุโขทัย.....	22
ภาพที่ 2.10 สภาพพื้นที่บริเวณท่าอากาศยานพิษณุโลก	23
ภาพที่ 2.11 Synthetic Minority Over-sampling Technique (SMOTE)	34
ภาพที่ 2.12 ตัวอย่างการตั้งค่าค้นหาโดย Optimize Parameter (Grid).....	38
ภาพที่ 2.13 การจำแนกด้วยอัลกอริทึม Decision Tree.....	40
ภาพที่ 2.14 ส่วนประกอบของ Neural Networks.....	41
ภาพที่ 2.15 หลักการทำงานของเทคนิค Bagging.....	42
ภาพที่ 2.16 หลักการทำงานของเทคนิค AdaBoost.....	43
ภาพที่ 2.17 หลักการทำงานของเทคนิค Random Forest.....	44
ภาพที่ 2.18 หลักการทำงานของ K-fold Cross Validation	45
ภาพที่ 2.19 กราฟ Receiver Operating Characteristic (ROC)	47
ภาพที่ 3.1 การสืบค้นข้อมูลข่าว METAR.....	57
ภาพที่ 3.2 นำเข้าข้อมูลใน Microsoft Excel.....	58
ภาพที่ 3.3 ตัวอย่างตารางข้อมูลสำหรับชุดข้อมูลของสนามบินนานาชาติเชียงใหม่	60
ภาพที่ 3.4 ลักษณะภูมิประเทศของสนามบินที่มีภูเขาล้อมรอบ 1-2 ด้าน.....	67

สารบัญรูปภาพ (ต่อ)

	หน้า
ภาพที่ 3.5 ลักษณะภูมิประเทศของสนามบินที่มีภูเขาล้อมรอบ 3-4 ด้าน.....	68
ภาพที่ 3.6 การกำหนดช่วงในการค้นหา K (Number of Nearest Neighbors)	76
ภาพที่ 3.7 การหาค่า K (Number of Neighbours) ที่เหมาะสมของเทคนิค SMOTE	76
ภาพที่ 3.8 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของ Naive Bayes.....	77
ภาพที่ 3.9 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของ Decision Tree.....	77
ภาพที่ 3.10 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของ Neural Networks	77
ภาพที่ 3.11 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของ Random Forest.....	78
ภาพที่ 3.12 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของ Naive Bayes ร่วมกับ Bagging.....	78
ภาพที่ 3.13 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของ Decision Tree ร่วมกับ Bagging.....	78
ภาพที่ 3.14 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของ Neural Networks ร่วมกับ Bagging	79
ภาพที่ 3.15 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของ Naive Bayes ร่วมกับ AdaBoost.....	79
ภาพที่ 3.16 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของ Decision Tree ร่วมกับ AdaBoost.....	79
ภาพที่ 3.17 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของ Neural Networks ร่วมกับ AdaBoost	80
ภาพที่ 3.18 การหาค่าน้ำหนักแอตทริบิวต์ด้วยวิธี Information Gain และ Chi-Square.....	83

สารบัญภาพ (ต่อ)

	หน้า
ภาพที่ 3.19 ค่าพารามิเตอร์ที่กำหนดได้ของอัลกอริทึม Naïve Bayes	84
ภาพที่ 3.20 ค่าพารามิเตอร์ที่กำหนดได้ของอัลกอริทึม Decision Tree	85
ภาพที่ 3.21 ค่าพารามิเตอร์ที่กำหนดได้ของอัลกอริทึม Neural Networks	86
ภาพที่ 3.22 การกำหนดค่าพารามิเตอร์ของอัลกอริทึม Random Forest	87
ภาพที่ 3.23 การกำหนดพารามิเตอร์ของเทคนิค Bagging	88
ภาพที่ 3.24 การกำหนดพารามิเตอร์ของเทคนิค AdaBoost	88
ภาพที่ 3.25 การปรับค่าพารามิเตอร์ที่เหมาะสมในงานวิจัย	89
ภาพที่ 3.26 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของ Decision Tree	90
ภาพที่ 3.27 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของ Neural Networks	90
ภาพที่ 3.28 การปรับการค้นหา Optimize Parameter (Grid) ของ Decision Tree	90
ภาพที่ 3.29 การปรับการค้นหา Optimize Parameter (Grid) ของ Neural Networks	91
ภาพที่ 3.30 การปรับค่าพารามิเตอร์ที่เหมาะสมสำหรับอัลกอริทึมการเรียนรู้แบบกลุ่ม	91
ภาพที่ 3.31 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของ Random Forest	92
ภาพที่ 3.32 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของ Decision Tree ร่วมกับ Bagging	92
ภาพที่ 3.33 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของ Neural Networks ร่วมกับ Bagging	93
ภาพที่ 3.34 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของ Decision Tree ร่วมกับ AdaBoost	93
ภาพที่ 3.35 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของ Neural Networks ร่วมกับ AdaBoost	93
ภาพที่ 3.36 การปรับการค้นหา Optimize Parameter (Grid) ของ Random Forest	94

สารบัญภาพ (ต่อ)

หน้า

ภาพที่ 3.37 การปรับการค้นหา Optimize Parameter (Grid) ของ Decision Tree ร่วมกับอัลกอริทึมการเรียนรู้แบบกลุ่ม	94
ภาพที่ 3.38 การปรับการค้นหา Optimize Parameter (Grid) ของ Neural Networks ร่วมกับอัลกอริทึมการเรียนรู้แบบกลุ่ม	95
ภาพที่ 3.39 การสร้างแบบจำลองจากอัลกอริทึม Neural Networks เพื่อค้นหาค่า Hidden layer size ด้วยวิธี Random Search.....	96
ภาพที่ 3.40 การสร้างแบบจำลองจากอัลกอริทึม Neural Networks ร่วมกับ Bagging เพื่อค้นหาค่า Hidden layer size ด้วยวิธี Random Search.....	96
ภาพที่ 3.41 การสร้างแบบจำลองจากอัลกอริทึม Neural Networks ร่วมกับ AdaBoost เพื่อค้นหาค่า Hidden layer size ด้วยวิธี Random Search.....	97
ภาพที่ 3.42 ขั้นตอนการสร้างแบบจำลองจาก Naive Bayes.....	98
ภาพที่ 3.43 Sub-Process ภายในโอเพอร์เรเตอร์ Naive-Cross Validation.....	98
ภาพที่ 3.44 วิธีการสร้างแบบจำลองจาก Decision Tree.....	99
ภาพที่ 3.45 Sub-Process ภายในของแบบจำลองจาก Decision Tree	99
ภาพที่ 3.46 วิธีการสร้างแบบจำลองจาก Neural Networks	100
ภาพที่ 3.47 Sub-Process ภายในของแบบจำลองจาก Neural Networks.....	100
ภาพที่ 3.48 การสร้างแบบจำลองการเรียนรู้แบบรวมกลุ่มเทคนิค Random Forest.....	101
ภาพที่ 3.49 Sub-Process ภายในของเทคนิค Random Forest.....	101
ภาพที่ 3.50 การสร้างแบบจำลองการเรียนรู้แบบรวมกลุ่มเทคนิค Bagging	102
ภาพที่ 3.51 Sub-Process ภายในเทคนิค Bagging ร่วมกับ Naive Bayes.....	103
ภาพที่ 3.52 Sub-Process ภายในเทคนิค Bagging ร่วมกับ Decision Tree	103
ภาพที่ 3.53 Sub-Process ภายในเทคนิค Bagging ร่วมกับ Neural Networks.....	103
ภาพที่ 3.54 การสร้างแบบจำลองการเรียนรู้แบบรวมกลุ่มเทคนิค AdaBoost	104
ภาพที่ 3.55 Sub-Process ภายในของเทคนิค AdaBoost ร่วมกับ Naive Bayes.....	104
ภาพที่ 3.56 Sub-Process ภายในเทคนิค AdaBoost ร่วมกับ Decision Tree	105

สารบัญภาพ (ต่อ)

	หน้า
ภาพที่ 3.57 Sub-Process ภายในของเทคนิค AdaBoost ร่วมกับ Neural Networks	105
ภาพที่ 3.58 ตัวอย่างการวัดประสิทธิภาพของแบบจำลองของอัลกอริทึม Random Forest.....	106
ภาพที่ 4.1 แผนภูมิแสดงค่าน้ำหนักการคัดเลือกคุณลักษณะด้วยวิธี Information Gain สำหรับระยะเวลาพยากรณ์ 1 ชั่วโมงข้างหน้า	115
ภาพที่ 4.2 แผนภูมิแสดงค่าน้ำหนักการคัดเลือกคุณลักษณะด้วยวิธี Information Gain สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมงข้างหน้า	116
ภาพที่ 4.3 แผนภูมิแสดงค่าน้ำหนักการคัดเลือกคุณลักษณะด้วยวิธี Chi-Square สำหรับระยะเวลาพยากรณ์ 1 ชั่วโมงข้างหน้า	116
ภาพที่ 4.4 แผนภูมิแสดงค่าน้ำหนักการคัดเลือกคุณลักษณะด้วยวิธี Chi-Square สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมงข้างหน้า	117
ภาพที่ 4.5 ค่า F-measure แต่ละอัลกอริทึมของกลุ่มข้อมูล ระยะเวลาพยากรณ์ 1 ชั่วโมงข้างหน้า.....	129
ภาพที่ 4.6 การเปรียบเทียบค่า AUC ของอัลกอริทึมแต่ละกลุ่มข้อมูล สำหรับระยะเวลาพยากรณ์ 1 ชั่วโมงข้างหน้า	129
ภาพที่ 4.7 ค่า F-measure แต่ละอัลกอริทึมของกลุ่มข้อมูล ระยะเวลาพยากรณ์ 2 ชั่วโมงข้างหน้า.....	132
ภาพที่ 4.8 การเปรียบเทียบค่า AUC ของอัลกอริทึมแต่ละกลุ่มข้อมูล สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมงข้างหน้า	133

บทที่ 1

บทนำ

1. ความเป็นมาและความสำคัญของปัญหา

การพยากรณ์อากาศระยะปัจจุบัน หรือ Nowcasting เป็นการพยากรณ์สภาวะอากาศที่เกิดขึ้นในปัจจุบันสำหรับช่วงเวลาไม่เกิน 2 ชั่วโมง ซึ่งข้อมูลการพยากรณ์อากาศระยะปัจจุบันนี้มีความสำคัญอย่างยิ่งต่อการรายงานข่าวอากาศการบิน (Meteorological Terminal Air Report: METAR) เนื่องจากนำมาใช้สำหรับการวางแผนการบินของสายการบินภายในประเทศและการจัดการจราจรทางอากาศของหน่วยงานประจำท่าอากาศยาน (กรมอุตุนิยมวิทยา, 2554) โดยเฉพาะการเกิดปรากฏการณ์พายุฝนฟ้าคะนอง (Thunderstorm) ที่มีสภาวะอากาศที่เลวร้ายที่สุด มีลมกระโชกและกระแสอากาศพัดขึ้นลงรุนแรง ส่งผลให้ในระหว่างการขึ้นลงของเครื่องบินนักบินไม่สามารถควบคุมหรือรักษาระดับบินไว้ได้เป็นเหตุให้เครื่องบินประสบอุบัติเหตุได้ ดังนั้นการแจ้งเตือนเหตุการณ์ล่วงหน้าของการเกิดพายุฝนฟ้าคะนองระยะปัจจุบัน จึงช่วยเพิ่มประสิทธิภาพการตัดสินใจของนักบินหรือหน่วยงานต่างๆ ที่เกี่ยวข้องกับการจราจรทางอากาศโดยรอบสนามบิน สำหรับการพยากรณ์อากาศของกรมอุตุนิยมวิทยาในระยะปัจจุบันเพื่อออกข่าวอากาศการบินนั้นดำเนินการโดยผู้ปฏิบัติงานเป็นหลัก นั่นคือ นักอุตุนิยมวิทยาประจำสนามบินวิเคราะห์ข้อมูลสภาพอากาศที่ได้จากการตรวจวัดสารประกอบทางอุตุนิยมวิทยาโดยระบบตรวจอากาศอัตโนมัติที่ติดตั้งอยู่บริเวณสนามบินแล้วพิจารณาพร้อมกับข้อมูลอื่นๆ เช่น แบบจำลองพยากรณ์เชิงตัวเลข เรดาร์ตรวจอากาศ และดาวเทียมตรวจอากาศ เป็นต้น จากนั้นรายงานข่าวอากาศการบินที่ได้จากการตรวจวัดในชั่วโมงนั้นๆ และออกคำเตือนพยากรณ์แนวโน้มสภาพอากาศแปรปรวนหรือการเกิดพายุฝนฟ้าคะนองที่จะเกิดขึ้นภายในระยะเวลา 2 ชั่วโมงข้างหน้า (Trend forecast) ต่อท้ายข่าวอากาศการบินดังกล่าว โดยทำการประเมินผลความถูกต้องของการพยากรณ์จากข่าวอากาศการบินในชั่วโมงถัดไปว่ามีการแจ้งเตือนการเกิดพายุฝนฟ้าคะนองทันหรือไม่ ซึ่งความถูกต้องของการพยากรณ์ในแต่ละสถานีตรวจอากาศการบินขึ้นอยู่กับความรู้ความสามารถและประสบการณ์ในการทำงานของผู้ปฏิบัติงานแต่ละคน นอกจากนี้เนื่องจากภาคเหนือของประเทศไทยมีลักษณะภูมิประเทศโดยทั่วไปแบ่งออกเป็น 3 ลักษณะ คือ

เทือกเขา ที่ราบระหว่างหุบเขา และที่ราบลุ่ม รวมถึงมีภูมิอากาศแบบร้อนชื้นสลับแล้ง สภาพภูมิอากาศทั่วไปอยู่ภายใต้อิทธิพลของลมมรสุมตะวันตกเฉียงใต้ และลมมรสุมตะวันออกเฉียงเหนือ นอกจากนี้ยังได้รับอิทธิพลจากพายุจรต่างๆ ที่พัดผ่านเข้ามาในพื้นที่ (สำนักงานสิ่งแวดล้อมภาคเหนือ, 2563) ด้วยลักษณะภูมิประเทศและภูมิอากาศดังกล่าวทำให้ลักษณะอากาศเกิดการเปลี่ยนแปลงอยู่ตลอดเวลาและเป็นเรื่องยากสำหรับการพยากรณ์อากาศระยะปัจจุบันได้อย่างถูกต้อง

ด้วยข้อจำกัดดังกล่าวจึงควรมีการใช้เทคนิคการเรียนรู้ของเครื่องช่วยในการพยากรณ์ ในปัจจุบันมีการวิจัยใช้เทคนิคการเรียนรู้ของเครื่องในการวิเคราะห์ข้อมูลเพื่อการพยากรณ์อากาศตามช่วงระยะเวลาต่างๆที่สนใจ เช่น การวิจัยเกี่ยวกับการพยากรณ์ฝนระยะปัจจุบันด้วยเทคนิคการเรียนรู้ของเครื่องโดยอิงข้อมูลน้ำฝนตรวจวัดในกรุงเทพมหานครที่ทำการศึกษาระยะเวลาพยากรณ์ 30-180 นาทีข้างหน้า พบว่า ระยะเวลาพยากรณ์ 30 นาที มีความแม่นยำสูงสุดและเมื่อระยะเวลาพยากรณ์มากขึ้นทำให้ความแม่นยำลดลง โดยระยะเวลาพยากรณ์ 90 นาทีเป็นระยะเวลาพยากรณ์นานที่สุดที่สามารถยอมรับค่าความแม่นยำได้ (สมปราชญ์ ศรีถกล, 2564) รวมถึงงานวิจัยการพยากรณ์ปริมาณน้ำฝนของประเทศไทยด้วยวิธีการเรียนรู้เชิงลึกสองแบบจำลองต่อกันสำหรับพยากรณ์ปริมาณฝนล่วงหน้า 1 ถึง 6 ชั่วโมงข้างหน้า พบว่า ระยะเวลาพยากรณ์ 1 ชั่วโมงมีค่าความคลาดเคลื่อนต่ำที่สุดและเมื่อระยะเวลาพยากรณ์นานขึ้นทำให้ค่าความคลาดเคลื่อนสูงขึ้น (เพ็ญลดา มะโนกิจ, 2562) อีกทั้งยังมีงานวิจัยเกี่ยวกับการประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่องเพื่อสร้างระบบเตือนภัยล่วงหน้าสำหรับการแจ้งเตือนฝนตกหนักในระยะเวลาพยากรณ์ระยะสั้นมาก โดยทำการศึกษาระยะเวลาพยากรณ์ 3 ชั่วโมงข้างหน้า ซึ่งพบว่าแบบจำลองที่นำเสนอสามารถคาดการณ์ฝนตกหนักในระยะเวลาพยากรณ์ 3 ชั่วโมงได้ดี (Moon et al., 2018) เป็นต้น

งานวิจัยนี้เลือกทำการศึกษาที่ระยะเวลาพยากรณ์ 1 และ 2 ชั่วโมงข้างหน้า โดยพิจารณาระยะเวลาตามข้อกำหนดขององค์การการบินพลเรือนระหว่างประเทศ (International Civil Aviation Organization: ICAO) ที่กำหนดให้สถานีตรวจอากาศการบินรายงานข่าวอากาศการบินทุกชั่วโมงและออกข่าวคำเตือนพยากรณ์แนวโน้มสภาพอากาศในระยะเวลาพยากรณ์ 2 ชั่วโมง (กองอุตุนิยมวิทยาการบิน, 2558) ดังนั้นผู้วิจัยจึงเลือกพิจารณาระยะเวลาพยากรณ์ 1 และ 2 ชั่วโมงข้างหน้า เพื่อให้ครอบคลุมระยะเวลาการออกข่าวคำเตือนพยากรณ์แนวโน้มการเกิดพายุฝนฟ้าคะนองและเพื่อให้การพยากรณ์การเกิดพายุฝนฟ้าคะนองมีความแม่นยำยิ่งขึ้น ผู้วิจัยได้ทำการศึกษานามบินในพื้นที่ภาคเหนือจำนวน 9 สนามบินจากทั้งหมด 11 สนามบิน (ยกเว้นสนามบินตากและสนามบินเพชรบูรณ์ที่ใช้สำหรับทางราชการเท่านั้น) โดยทั้ง 9 สนามบินนี้มีการรายงานข่าวอากาศการบินราย

ข้อมูลโดยกรมอุตุนิยมวิทยา และเนื่องจากลักษณะพื้นที่ภาคเหนือของประเทศไทยมีความหลากหลาย ผู้วิจัยจึงได้ทำการแบ่งพื้นที่ศึกษาออกเป็น 5 กลุ่มข้อมูล คือ (1) กลุ่มข้อมูลรวมสนามบินภาคเหนือ (2) กลุ่มข้อมูลสนามบินภาคเหนือตอนบน (3) กลุ่มข้อมูลสนามบินภาคเหนือตอนล่าง (4) กลุ่มข้อมูลสนามบินที่มีภูเขาล้อมรอบ 1-2 ด้าน และ (5) กลุ่มข้อมูลสนามบินที่มีภูเขาล้อมรอบ 3-4 ด้าน จากนั้นใช้เทคนิคการเรียนรู้ของเครื่องในการสร้างแบบจำลองเพื่อพยากรณ์การเกิดพายุฝนฟ้าคะนอง ซึ่งในงานวิจัยนี้ทำการสร้างแบบจำลองการจำแนก (Classification Model) โดยมีการแบ่งคลาสเพื่อจำแนกชนิด Binary ที่มีเพียงสองสถานะเท่านั้น เพื่อบ่งชี้ปรากฏการณ์การเกิดพายุฝนฟ้าคะนอง นั่นคือ ไม่เกิดพายุฝนฟ้าคะนอง (NO TSRA) และเกิดพายุฝนฟ้าคะนอง (TSRA) สร้างแบบจำลองด้วยอัลกอริทึมมาตรฐาน ได้แก่ Naive Bayes Decision Tree และ Neural Networks และสร้างแบบจำลองอัลกอริทึมมาตรฐานร่วมกับอัลกอริทึมการเรียนรู้แบบกลุ่ม (Ensemble Learning Algorithms) ได้แก่ Bagging AdaBoost และ Random Forest เพื่อเพิ่มประสิทธิภาพการพยากรณ์การเกิดพายุฝนฟ้าคะนองระยะปัจจุบันให้มีความถูกต้องมากยิ่งขึ้น จากนั้นเปรียบเทียบประสิทธิภาพของแบบจำลองด้วยวิธี 10-Fold Cross Validation โดยคัดเลือกแบบจำลองที่ดีที่สุดของแต่ละกลุ่มข้อมูลด้วยค่า F-measure สูงสุด โดยเมื่อได้แบบจำลองที่ดีที่สุดของแต่ละพื้นที่แล้วจะทำการเปรียบเทียบกับข้อมูลจริงในปีพ.ศ. 2566 เพื่อประเมินการใช้งานโดยเปรียบเทียบความถูกต้องระหว่างคำพยากรณ์โดยนักอุตุนิยมวิทยากับผลการพยากรณ์จากแบบจำลอง ซึ่งสามารถสร้างความมั่นใจให้แก่ผู้ปฏิบัติงานเมื่อนำแบบจำลองจากงานวิจัยนี้ไปใช้งานจริงในการพยากรณ์การเกิดพายุฝนฟ้าคะนองบริเวณสนามบิน

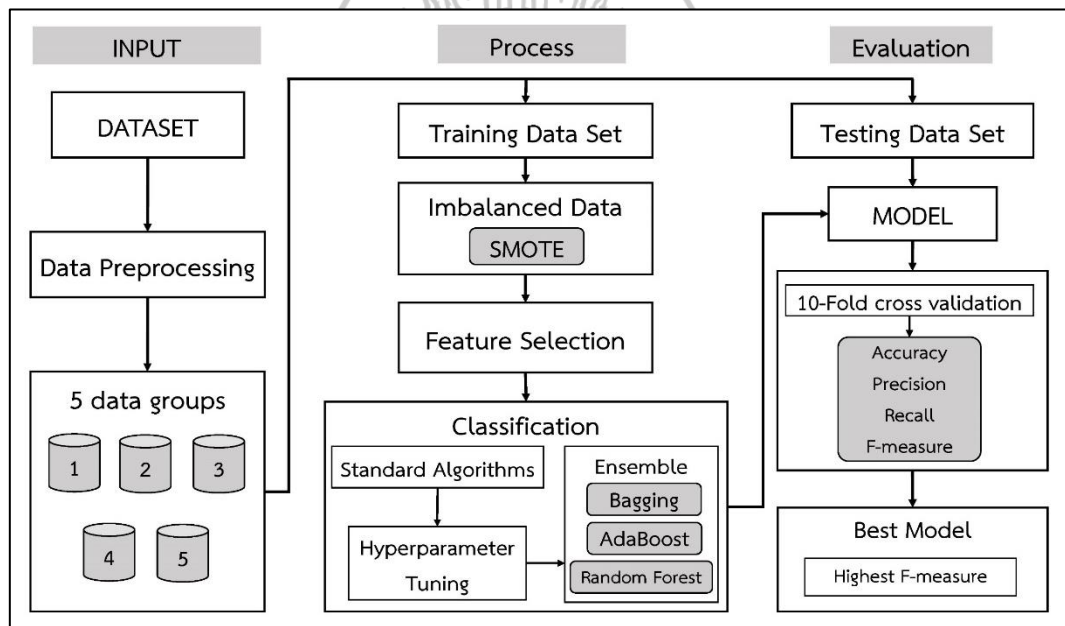
2. วัตถุประสงค์การวิจัย

2.1 เพื่อสร้างแบบจำลองในการพยากรณ์การเกิดพายุฝนฟ้าคะนองใน 1 และ 2 ชั่วโมงข้างหน้า สำหรับการพยากรณ์อากาศระยะปัจจุบัน บริเวณสนามบินภาคเหนือของประเทศไทย

2.2 เพื่อประเมินประสิทธิภาพของแบบจำลองในการพยากรณ์การเกิดพายุฝนฟ้าคะนองใน 1 และ 2 ชั่วโมงข้างหน้า สำหรับการพยากรณ์อากาศระยะปัจจุบัน บริเวณสนามบินภาคเหนือของประเทศไทย

3. กรอบแนวคิดการวิจัย

จากกรอบแนวคิดวิจัยดังแสดงในภาพที่ 1.1 มีการนำเข้าสู่ชุดข้อมูลข่าวอากาศการบิน จากนั้นชุดข้อมูลทั้งหมดจะเข้าสู่ขั้นตอนการเตรียมข้อมูลก่อนการสร้างแบบจำลอง 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 สูงสุดของแต่ละกลุ่มข้อมูล



ภาพที่ 1.1 กรอบแนวคิดการวิจัย

4. ขอบเขตของการวิจัย

4.1 รูปแบบการวิจัย การวิจัยเชิงพัฒนา

4.2 ขอบเขตข้อมูล ข้อมูลทำการวิจัยจากข้อมูลข่าวอากาศการบินจำนวน 9 แห่ง

ดังนี้

- 1) ท่าอากาศยานนานาชาติเชียงใหม่
- 2) ท่าอากาศยานแม่ฮ่องสอน
- 3) ท่าอากาศยานแม่ฟ้าหลวง เชียงราย
- 4) ท่าอากาศยานลำปาง
- 5) ท่าอากาศยานน่านนคร
- 6) ท่าอากาศยานแพร่
- 7) ท่าอากาศยานพิษณุโลก
- 8) ท่าอากาศยานสุโขทัย
- 9) ท่าอากาศยานนานาชาติแม่สอด

4.3 ขอบเขตด้านระยะเวลา ข้อมูลตั้งแต่วันที่ 1 มกราคม พ.ศ. 2558 ถึง วันที่ 31 ธันวาคม พ.ศ. 2565 รวมระยะเวลา 8 ปี

4.4 ขอบเขตด้านเนื้อหา ในตอนเริ่มต้นการศึกษาผู้วิจัยตั้งขอบเขตของข้อมูลไว้ว่า ใช้ข้อมูลการตรวจอากาศโดยกรมอุตุนิยมวิทยาจำนวน 5 ปี คือ พ.ศ.2561-2565 ของข่าวอากาศการบิน และข้อมูลจากสถานีตรวจวัดอากาศอัตโนมัติ (AWS) มาผนวกรวมกันเพื่อสร้างแบบจำลองที่เหมาะสมสำหรับสนามบินของภาคเหนือ แต่หลังจากการรวบรวมข้อมูลพบว่า ไม่สามารถใช้ข้อมูลเพื่อทำการศึกษาได้ทุกๆสนามบินของภาคเหนือ เนื่องจากต้องทำการคัดเลือกสนามบินที่มีการติดตั้งเครื่องมือตรวจวัดอากาศอัตโนมัติบริเวณเดียวกันหรือใกล้กับสนามบินแต่ละแห่ง โดยเครื่องมือดังกล่าวจะต้องติดตั้งในรัศมี 8 กม.จากจุดอ้างอิงสนามบิน (Aerodrome Reference Point : ARP) ตามกฎขององค์การการบินพลเรือนระหว่างประเทศ (ICAO) และบางแห่งเพิ่งได้รับการติดตั้งเครื่องมือใหม่ทำให้มีข้อมูลไม่เพียงพอ ดังนั้นจึงเหลือเพียง 5 สนามบินของภาคเหนือเท่านั้นที่สามารถนำข้อมูลมาทำการศึกษาได้ ซึ่งได้แก่ ท่าอากาศยานนานาชาติเชียงใหม่ ท่าอากาศยานแม่ฮ่องสอน ท่าอากาศยานลำปาง ท่าอากาศยานพิษณุโลก และท่าอากาศยานนานาชาติแม่สอด

จากตารางที่ 1.1 เมื่อเริ่มทำการรวบรวมและเตรียมข้อมูลของทั้ง 5 สนามบิน พบว่าระยะเวลาของการเกิดพายุฝนฟ้าคะนองจากการตรวจวัดของเครื่องมือตรวจอากาศการบิน และจากเครื่องมือ AWS มีความคลาดเคลื่อนกัน ทำให้แม้ว่าจะเกิดฝนตกในชั่วโมงนั้น แต่ถ้าเครื่องมืออย่างใดอย่างหนึ่งตรวจไม่พบฝนตก ข้อมูลชุดนั้นก็ต้องถูกตัดทิ้งไปส่งผลให้ชุดข้อมูลที่จะนำมาสร้างแบบจำลองมีจำนวนน้อยกว่าความเป็นจริง เนื่องจากความแตกต่างของจุดติดตั้งอุปกรณ์ที่อยู่คนละจุดกันนั่นเอง ดังนั้นจึงเกิดปัญหาของจำนวนข้อมูลเมื่อนำชุดข้อมูลของ 2 แหล่งข้อมูลมารวมกันแล้วมีข้อมูลของคลาสเกิดฝนฟ้าคะนอง (TSRA Class) น้อยเกินไป ผู้วิจัยจึงได้เริ่มขยายขอบเขตข้อมูลให้มากขึ้นจาก 5 ปี เป็น 8 ปี (พ.ศ.2558-2565) เพื่อให้มีข้อมูลที่เพียงพอ แต่พบว่า มีเพียงท่าอากาศยานนานาชาติเชียงใหม่เท่านั้นที่มีข้อมูลเครื่องมือ AWS ย้อนหลังนาน 8 ปี แต่สนามบินที่เหลือได้รับการติดตั้งเครื่องมือที่สามารถรวบรวมข้อมูลได้เพียง 5-6 ปีเท่านั้น ดังนั้นด้วยข้อมูลที่มีอยู่ทำให้ไม่สามารถนำชุดข้อมูลร่วมกันของ 2 แหล่งตรวจอากาศ คือ ข่าวกาศการบินและเครื่องมือ AWS ไปสร้างแบบจำลองได้ เพราะอาจทำให้แบบจำลองเกิดการเรียนรู้ที่ผิดพลาดจากชุดข้อมูลที่มีน้อยเกินไป

ตารางที่ 1.1 จำนวนข้อมูลการบินร่วมกับ AWS ของ 5 สนามบินภาคเหนือ (พ.ศ.2561-2565)

ชื่อสนามบิน	METAR	AWS	1 ชั่วโมง	2 ชั่วโมง
			TSRA	TSRA
1. ท่าอากาศยานนานาชาติเชียงใหม่	38,080	38,243	175	140
2. ท่าอากาศยานแม่ฮ่องสอน	23,829	37,300	50	45
3. ท่าอากาศยานพิษณุโลก	27,965	33,233	90	88
4. ท่าอากาศยานลำปาง	26,134	32,134	87	72
5. ท่าอากาศยานนานาชาติแม่สอด	23,611	29,568	45	32

ผู้วิจัยจึงกำหนดขอบเขตข้อมูลการวิจัยใหม่ โดยใช้เพียงข้อมูลข่าวกาศการบินเพียงแหล่งเดียวรวบรวมข้อมูลตั้งแต่วันที่ 1 ม.ค. 2558 ถึง 31 ธ.ค. 2565 รวมระยะเวลา 8 ปี ซึ่งเป็นข้อมูลรายชั่วโมงของข่าวกาศการบินของสนามบินภาคเหนือ 9 สนามบินที่รายงานโดยกรมอุตุนิยมวิทยา มีจำนวนข้อมูลทั้งหมด 364,382 ชุดข้อมูล ทำการศึกษาที่ระยะเวลาพยากรณ์ 1 และ 2

ชั่วโมงข้างหน้า และทำการแบ่งออกเป็น 5 กลุ่มข้อมูลแล้ววิเคราะห์ตามขอบเขตและลักษณะพื้นที่ของสนามบินภาคเหนือของประเทศไทย รวมถึงตามเกณฑ์การแบ่งพื้นที่ของกรมอุตุนิยมวิทยา ดังนี้

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

4.4.2 กลุ่มข้อมูลที่ 2 สนามบินภาคเหนือตอนบน ได้แก่ ท่าอากาศยานแม่ฮ่องสอน ท่าอากาศยานนานาชาติเชียงใหม่ ท่าอากาศยานแม่ฟ้าหลวงเชียงราย ท่าอากาศยานลำปาง ท่าอากาศยานแพร่ และท่าอากาศยานน่านนคร

4.4.3 กลุ่มข้อมูลที่ 3 สนามบินภาคเหนือตอนล่าง ได้แก่ ท่าอากาศยานนานาชาติแม่สอด ท่าอากาศยานสุโขทัย และท่าอากาศยานพิษณุโลก

4.4.4 กลุ่มข้อมูลที่ 4 สนามบินที่มีภูเขาล้อมรอบ 1-2 ด้าน ได้แก่ ท่าอากาศยานสุโขทัย และท่าอากาศยานพิษณุโลก

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

ในงานวิจัยนี้ได้สร้างแบบจำลองการจำแนกด้วยโปรแกรม RapidMiner Studio โดยทั้ง 5 กลุ่มข้อมูลจะถูกแบ่งออกเป็น 2 ส่วนเพื่อนำเข้าสู่แบบจำลอง คือ (1) ชุดข้อมูลฝึกฝน ใช้ข้อมูลในปี พ.ศ. 2558-2563 รวม 6 ปี นำเข้าเพื่อใช้สอนและปรับค่าพารามิเตอร์ที่เหมาะสม และ (2) ชุดข้อมูลทดสอบ ใช้ข้อมูลในปี พ.ศ. 2564-2565 รวม 2 ปี นำเข้าเพื่อทดสอบว่าสามารถทำงานได้ดีกับชุดข้อมูลที่ไม่เคยพบมาก่อนหรือไม่ ซึ่งในงานวิจัยนี้มีชุดข้อมูลไม่สมดุล คือ คลาสไม่เกิดพายุฝนฟ้าคะนอง (NO TSRA) มีจำนวนมากกว่าคลาสเกิดพายุฝนฟ้าคะนอง (TSRA) และผู้วิจัยได้เลือกทำกระบวนการเพื่อจัดการปรับความไม่สมดุลของข้อมูลก่อนการคัดเลือกคุณลักษณะ เนื่องจากข้อมูลที่ไม่สมดุลกันอาจทำให้แบบจำลองมีความลำเอียงในการเรียนรู้ที่ไม่เหมาะสม คุณลักษณะที่มีจำนวนน้อยอาจถูกมองข้ามและไม่ถูกพิจารณาว่าเป็นสิ่งสำคัญ ส่งผลต่อความแม่นยำของแบบจำลองได้ โดยปรับความสมดุลของข้อมูลด้วยเทคนิคการสุ่มตัวอย่างเกินหรือ เทคนิค SMOTE จากนั้นสร้างแบบจำลองด้วยอัลกอริทึมมาตรฐาน ได้แก่ Naive Bayes Decision Tree และ Neural Networks และสร้างแบบจำลองอัลกอริทึมมาตรฐานร่วมกับอัลกอริทึมการเรียนรู้แบบกลุ่ม ได้แก่ Random Forest Bagging และ AdaBoost เพื่อเพิ่มประสิทธิภาพการพยากรณ์การเกิดพายุฝนฟ้าคะนอง

ระยะปัจจุบันให้มีความถูกต้องมากยิ่งขึ้น และนำเข้าสู่ชุดข้อมูลทดสอบเพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองวิธี 10-Fold Cross Validation ด้วยค่า Accuracy Precision Recall และ F-measure โดยคัดเลือกแบบจำลองที่ดีที่สุดของแต่ละกลุ่มข้อมูลด้วยค่า F-measure สูงสุด ซึ่งจะสามารถวัดผลความถ่วงดุลของแบบจำลองที่หามาได้จากการคำนวณค่า Precision และ Recall ทำให้บังชี้การทำนายคลาสเกิดพายุฝนฟ้าคะนอง (TSRA Class) ได้ ประกอบกับการพิจารณาวัดประสิทธิภาพแบบจำลองทั้งหมดด้วยค่า Accuracy Precision Recall AUC และ Balanced Accuracy เพื่อพิจารณาประสิทธิภาพของแบบจำลองโดยภาพรวม

5. นิยามศัพท์เฉพาะ

5.1 การพยากรณ์อากาศระยะปัจจุบัน (Nowcasting) หมายถึง การรายงานสถานะอากาศที่เกิดขึ้นในปัจจุบัน และการคาดการณ์สภาพลมฟ้าอากาศสำหรับช่วงเวลาไม่เกิน 2 ชั่วโมง

5.2 ปรากฏการณ์พายุฝนฟ้าคะนอง (Thunderstorm) หมายถึง สภาพอากาศที่เกิดจากเมฆที่ก่อตัวขึ้นในทางตั้งขนาดใหญ่ที่เรียกว่า เมฆคิวมูโลนิมบัส ทำให้เกิดลักษณะอากาศร้าย เช่น ลมกระโชก ฟ้าผ่า ลูกเห็บตก และฝนตกหนัก เป็นต้น

5.3 ข่าวอากาศการบิน (Meteorological Terminal Air Report: METAR) หมายถึง การรายงานอากาศสนามบินแบบประจำ (Routine Meteorological Report) ซึ่งต้องตรวจวัดและรายงานทุก ๆ 30 นาที สำหรับสนามบินนานาชาติ หรือทุก ๆ ชั่วโมงสำหรับสนามบินภายในประเทศไทย เมื่อทำการตรวจวัดสภาพอากาศแล้วจะต้องเข้ารหัสเพื่อกระจายข่าวต่อไป

5.4 การเรียนรู้ของเครื่อง (Machine Learning) หมายถึง การเรียนรู้โดยอัตโนมัติจากการเรียนรู้ในอดีตและทำการปรับปรุงการทำงานได้เองโดยไม่ต้องมีมนุษย์คอยกำกับหรือเขียนโปรแกรมกำกับสั่งการเครื่อง ซึ่งมีหลักการทำงานคล้ายกับมนุษย์ที่จำเป็นต้องเรียนรู้ประสบการณ์เพื่อให้สามารถจำแนกได้อย่างถูกต้อง

5.5 แบบจำลองการจำแนก (Classification Model) หมายถึง วิธีการวิเคราะห์ข้อมูลที่จัดอยู่ในประเภทการเรียนรู้แบบมีผู้สอน ซึ่งจำเป็นต้องมีคำตอบ (Label) ให้แบบจำลองเรียนรู้จากข้อมูลนำเข้าที่ถูกป้อนเข้ามาเพื่อให้สามารถหาคำตอบตามกำหนดไว้ ซึ่งผลลัพธ์ของการจำแนกนี้จะอยู่ในรูปแบบของคำตอบที่เป็นตัวเลือก หรือกลุ่มข้อมูล

5.6 อัลกอริทึมการเรียนรู้แบบกลุ่ม (Ensemble Learning Algorithms) หมายถึง วิธีการอาศัยแบบจำลองการจำแนกข้อมูลมากกว่าหนึ่งตัว แต่ละแบบจำลองการจำแนกข้อมูลจะมีกระบวนการทำงานของตัวเอง และทุกแบบจำลองการจำแนกจะสร้างจากกลุ่มข้อมูลเดียวกัน เมื่อได้

ผลลัพธ์ของแต่ละแบบจำลอง ก็จะนำผลลัพธ์ที่ได้เหล่านั้นผ่านวิธีการรวบรวม (Combination Integration หรือ Vote) และตัดสินผลลัพธ์สุดท้าย (Final Decision) เพื่อให้ได้ผลลัพธ์การจำแนกข้อมูลเดี่ยวเท่านั้น

5.7 ข้อมูลไม่สมดุล (Imbalanced data) หมายถึง การที่ชุดข้อมูลประกอบด้วยข้อมูลกลุ่มส่วนมาก (Majority) และข้อมูลกลุ่มส่วนน้อย (Minority) ถ้าเข้าสู่กระบวนการสร้างแบบจำลอง โดยที่ไม่มีขั้นตอนการปรับสมดุลของข้อมูลก่อน ส่งผลให้แบบจำลองทำงานได้ผิดพลาด

5.8 เทคนิคการสุ่มเพิ่มตัวอย่างกลุ่มน้อย (Synthetic Minority Oversampling Technique : SMOTE) หมายถึง เป็นเทคนิคที่ใช้ในการแก้ปัญหาข้อมูลไม่สมดุล โดยเพิ่มจำนวนข้อมูลกลุ่มน้อยจากการสังเคราะห์ด้วยข้อมูลจริงในการสร้างข้อมูลขึ้นมาให้มีปริมาณข้อมูลใกล้เคียงกับข้อมูลของกลุ่มส่วนมาก โดยจะสุ่มเลือกข้อมูลกลุ่มส่วนน้อยขึ้นมา 1 และคำนวณหาค่าระยะห่างระหว่างค่าที่เลือกกับทุกๆ ค่า จากนั้นเลือกค่าที่ใกล้เคียงที่สุด

5.9 การคัดเลือกคุณลักษณะ (Feature Selection) หมายถึง การลดขนาดหรือมิติของข้อมูลและยังคงลักษณะสำคัญของข้อมูลไว้ อาจทำให้เหลือคุณลักษณะที่มีความสำคัญและดีที่สุดเพียงหนึ่งตัวหรือกลุ่มของตัวแปรเพียงหนึ่งกลุ่มที่มีความสำคัญต่อการพยากรณ์ ซึ่งหากได้คุณลักษณะที่ดีจะช่วยให้แบบจำลองมีประสิทธิภาพและทำงานได้เร็วขึ้น

6. ประโยชน์ที่คาดว่าจะได้รับ

6.1 นักอุตุนิยมวิทยาผู้ปฏิบัติงานได้ใช้แบบจำลองการพยากรณ์การเกิดพายุฝนฟ้าคะนองระยะปัจจุบันที่เหมาะสมสำหรับสนามบินบริเวณภาคเหนือของประเทศไทย ซึ่งสามารถปรับปรุงประสิทธิภาพของการพยากรณ์อากาศระยะปัจจุบันให้มีความแม่นยำมากยิ่งขึ้น

6.2 เจ้าหน้าที่หอบังคับการบินซึ่งทำหน้าที่ในการจัดการจราจรทางอากาศ มีแหล่งข้อมูลเพิ่มเติมที่ได้รับจากนักอุตุนิยมวิทยาผู้ปฏิบัติงานเพื่อประกอบการวางแผนการจราจรทางอากาศให้ปลอดภัยต่อสภาพอากาศร้ายของปรากฏการณ์พายุฝนฟ้าคะนอง

6.3 สายการบินและประชาชนทั่วไปที่เป็นผู้ใช้งานสนามบินบริเวณภาคเหนือของประเทศไทย ได้รับการแจ้งเตือนภัยการเกิดฝนฟ้าคะนองบริเวณสนามบินได้ทันต่อการเกิดสภาพอากาศร้าย

บทที่ 2

วรรณกรรมที่เกี่ยวข้อง

การศึกษาวิจัย เรื่อง การพยากรณ์การเกิดพายุฝนฟ้าคะนองระยะปัจจุบันจากข่าวอากาศ การบิน ด้วยเทคนิคการเรียนรู้ของเครื่อง กรณีศึกษาสนามบินภาคเหนือของประเทศไทย ผู้วิจัยได้ ทำการค้นคว้าแนวคิดทฤษฎีและงานวิจัยที่เกี่ยวข้องเพื่อใช้เป็นแนวทางในการศึกษา ดังนี้

1. การพยากรณ์อากาศ
2. ปราณุกรณ์พายุฝนฟ้าคะนอง
3. สนามบินบริเวณภาคเหนือที่ทำการศึกษา
4. ข่าวอากาศการบิน
5. เทคนิคการเรียนรู้ของเครื่อง
6. งานวิจัยที่เกี่ยวข้อง

1. การพยากรณ์อากาศ

1.1. ความหมายของการพยากรณ์อากาศ

กรมอุตุนิยมวิทยา (2554) ได้ให้ความหมายการพยากรณ์อากาศ หมายถึง การ คาดหมายสภาพลมฟ้าอากาศในอนาคต การที่จะพยากรณ์อากาศได้ต้องมียอดประกอบ 3 ประการ ประการแรก คือ ความรู้ความเข้าใจในปรากฏการณ์และกระบวนการต่าง ๆ ที่เกิดขึ้นในบรรยากาศ ซึ่งได้มาจากเฝ้าสังเกตและบันทึกไว้ โดยการสังเกตทำให้สามารถอธิบายถึงสาเหตุของการเกิดลักษณะ อากาศแบบต่างๆ ได้ ประการที่สอง คือ สภาพอากาศปัจจุบัน ซึ่งได้มาจากการตรวจอากาศ ที่มีทั้ง การตรวจอากาศผิวพื้น และการตรวจอากาศชั้นบนในระดับความสูงต่างๆ และประการสุดท้าย คือ ความสามารถที่จะผสมผสานองค์ประกอบทั้งสองข้างต้นเข้าด้วยกันเพื่อคาดการณ์การเปลี่ยนแปลง ของบรรยากาศที่จะเกิดขึ้นในอนาคต เมื่อมีความรู้ความเข้าใจในเรื่องราวของลมฟ้าอากาศ และมี ข้อมูลผลการตรวจอากาศแล้ว สิ่งที่ต้องทำเพื่อให้สามารถพยากรณ์อากาศได้ คือ การวิเคราะห์ข้อมูล ผลการตรวจอากาศเพื่อให้ทราบลักษณะอากาศปัจจุบัน และการคาดการณ์การเปลี่ยนแปลงของ ลักษณะอากาศที่กำลังเกิดขึ้นนั้นว่าจะมีทิศทางและความเร็วในการเคลื่อนที่อย่างไร และความรุนแรง

จะเพิ่มขึ้นหรือลดลงเพียงใด นั่นคือคาดการณ์ว่าบริเวณที่จะพยากรณ์นั้นจะอยู่ภายใต้อิทธิพลของปรากฏการณ์แบบใด แล้วจึงจัดทำคำพยากรณ์อากาศโดยพิจารณาจากลักษณะลมฟ้าอากาศที่สัมพันธ์กับปรากฏการณ์นั้น ๆ ต่อไป

1.2. ระยะเวลาของการพยากรณ์อากาศ

การพยากรณ์อากาศอาจเป็นการคาดการณ์สำหรับช่วงเวลาไม่กี่ชั่วโมงข้างหน้าจนถึงการคาดการณ์สิ่งที่จะเกิดขึ้นในอีกหลายปีจากปัจจุบัน โดยสามารถแบ่งชนิดของการพยากรณ์อากาศตามระยะเวลาที่คาดการณ์ได้ดังนี้

1.2.1 การพยากรณ์ระยะปัจจุบัน คือ การรายงานสภาวะอากาศที่เกิดขึ้นในปัจจุบัน และการคาดการณ์สภาพลมฟ้าอากาศสำหรับช่วงเวลาไม่เกิน 2 ชั่วโมง

1.2.2 การพยากรณ์ระยะสั้นมาก คือ การพยากรณ์สำหรับช่วงเวลาไม่เกิน 12 ชั่วโมง

1.2.3 การพยากรณ์ระยะสั้น คือ การพยากรณ์สำหรับระยะเวลาเกินกว่า 12 ชั่วโมงขึ้นไปจนถึง 3 วัน

1.2.4 การพยากรณ์อากาศระยะปานกลาง คือ การพยากรณ์สำหรับช่วงเวลาไม่เกินกว่า 3 วันขึ้นไปจนถึง 10 วัน

1.2.5 การพยากรณ์ระยะยาว คือ การพยากรณ์สำหรับช่วงเวลาระหว่าง 10 ถึง 30 วัน โดยปกติมักเป็นการพยากรณ์ว่าค่าเฉลี่ยของตัวแปรทางอุตุนิยมวิทยาในช่วงเวลานั้นจะแตกต่างไปจากค่าเฉลี่ยทางภูมิอากาศอย่างไร

1.2.6 การพยากรณ์ระยะนาน คือ การพยากรณ์ตั้งแต่ 30 วัน จนถึง 2 ปี

1.2.7 การพยากรณ์ภูมิอากาศ คือ การพยากรณ์ช่วงเวลามากกว่า 2 ปีขึ้นไป

1.3. ข้อมูลอุตุนิยมวิทยาจากการตรวจวัด

ข้อมูลอุตุนิยมวิทยาจากการตรวจวัดเป็นปัจจัยสำคัญสำหรับการพยากรณ์อากาศในทุกระยะการพยากรณ์ โดยจะใช้ข้อมูลการตรวจวัดจากสถานีอุตุนิยมวิทยาที่กระจายอยู่ทั่วประเทศ ได้แก่ ข้อมูลจากสถานีตรวจอากาศผิวพื้นที่มีการรายงานผลการตรวจวัดทุก 3 ชั่วโมงตลอด 24 ชั่วโมง ข้อมูลจากสถานีตรวจอากาศการบินที่มีการรายงานผลการตรวจวัดทุก 30 นาทีหรือ 1 ชั่วโมง และ ข้อมูลจากสถานีตรวจอากาศชั้นบนที่มีการรายงานผลการตรวจวัดวันละ 1 ครั้ง ข้อมูลจากสถานีตรวจอากาศจะประกอบด้วย สารประกอบทางอุตุนิยมวิทยาที่เป็นลักษณะอากาศที่เกิดขึ้นในแต่ละท้องถิ่น

เช่น ความกดอากาศ ลม อุณหภูมิ ความชื้นสัมพัทธ์ เมฆ และหยาดน้ำฟ้า เป็นต้น โดยเครื่องมือที่ใช้ในระบบการพยากรณ์อากาศสามารถแบ่งออกได้เป็นประเภทใหญ่ๆ ได้ดังนี้

1.3.1 เครื่องมือตรวจอากาศผิวพื้น ในสถานีตรวจอากาศผิวพื้นแต่ละสถานีจะมีสนามอุตุนิยมวิทยา ซึ่งเป็นพื้นที่ติดตั้งอุปกรณ์เพื่อตรวจวัดสารประกอบอุตุนิยมวิทยาต่างๆ ตามเวลาที่กำหนดไว้ในแต่ละวันตลอด 24 ชั่วโมง

1.3.2 เครื่องมือตรวจอากาศการบิน เป็นเครื่องมือระบบตรวจอากาศอัตโนมัติและระบบตรวจวัดวินเชียร์ (Automatic Weather Observation System: AWOS และ Low Level Wind Shear Alert System: LLWAS) สำหรับตรวจวัดสารประกอบทางอุตุนิยมวิทยาต่างๆ บริเวณสนามบิน ได้แก่ การตรวจวัดอุณหภูมิของอากาศ อุณหภูมิบนทางวิ่ง ความชื้นสัมพัทธ์ ความกดอากาศ ทิศทางและความเร็วลม ทิศนวิสัยบนทางวิ่ง ความสูงฐานเมฆ และการตรวจวัดวินเชียร์

1.3.3 เครื่องมือตรวจอากาศชั้นบน เนื่องจากการเปลี่ยนแปลงของลักษณะอากาศมีความเกี่ยวข้องกับการเปลี่ยนแปลงของสภาวะสารประกอบอุตุนิยมวิทยาในระดับชั้นบน ซึ่งอยู่นอกเหนือจากผิวพื้นของโลกด้วย ดังนั้นจึงมีความจำเป็นต้องตรวจวัดสารประกอบอุตุนิยมวิทยาในบรรยากาศด้วยโดยเฉพาะในชั้นโทรโปสเฟียร์ สารประกอบทางอุตุนิยมวิทยาที่ทำการตรวจวัด เช่น อุณหภูมิ ความกด ลมและความชื้นอากาศในระดับต่าง ๆ เป็นต้น การตรวจวัดข้อมูลเหล่านี้ใช้เครื่องมือหลักที่เรียกว่า Radiosonde

1.3.4 เครื่องมือตรวจอากาศพิเศษ เป็นเครื่องมือที่ใช้สำหรับตรวจปรากฏการณ์หรือลักษณะอากาศที่เกิดขึ้นเพื่อช่วยสนับสนุนในการวิเคราะห์พยากรณ์อากาศ โดยเครื่องมือตรวจอากาศพิเศษเหล่านี้มีหลายอย่าง เช่น เรดาร์ตรวจอากาศ ดาวเทียมอุตุนิยมวิทยา และเครื่องมือสำหรับตรวจวัดความสูงของคลื่น เป็นต้น

2. ปรากฏการณ์พายุฝนฟ้าคะนอง

พายุฝนฟ้าคะนอง (Thunderstorm) เป็นปรากฏการณ์ทางธรรมชาติที่เกิดขึ้นเป็นประจำทุกวันเหนือพื้นผิวโลก โดยการก่อตัวที่เกิดขึ้นในแต่ละพื้นที่จะเป็นไปตามฤดูกาล ในบริเวณใกล้เส้นศูนย์สูตรนั้นมีโอกาสที่จะเกิดพายุฝนฟ้าคะนองได้ตลอดปี เนื่องจากมีสภาพอากาศในเขตร้อนจึงมีอากาศร้อนอบอ้าว ซึ่งเอื้อต่อการก่อตัวของพายุฝนฟ้าคะนอง ส่วนบริเวณขั้วโลกเหนือ และขั้วโลกใต้ที่อยู่ในละติจูดที่สูงขึ้นไป มักจะเกิดขึ้นในฤดูร้อน สำหรับประเทศไทยพายุฝนฟ้าคะนองสามารถก่อตัว

ได้เกือบตลอดเวลาและในทุกพื้นที่ เนื่องจากมีภูมิอากาศในเขตร้อน (Tropic) โดยเฉพาะในเดือนมีนาคมถึงเดือนพฤษภาคม พายุฝนฟ้าคะนองที่เกิดขึ้นจะมีความรุนแรงกว่าปกติ จนเกิดเป็นลักษณะที่เรียกว่า “พายุฤดูร้อน” (กรมอุตุนิยมวิทยา, 2550)

2.1. สาเหตุของการเกิดพายุฝนฟ้าคะนอง

พายุฝนฟ้าคะนองเกิดจากเมฆที่ก่อตัวขึ้นในทางตั้งขนาดใหญ่ที่เรียกว่า เมฆคิวมูโลนิมบัส (Cumulonimbus: CB) ทำให้เกิดลักษณะอากาศร้าย เช่น ลมกระโชก ฟ้าผ่า ลูกเห็บตก และฝนตกหนัก เป็นต้น นอกจากนี้เมฆจะมีฐานเมฆต่ำและมีกระแสอากาศไหลลงรุนแรงจนทำให้เกิดเมฆเป็นลำคล้ายวงช้างยื่นจากใต้ฐานเมฆหนาที่ปลงมายังพื้นดิน ภายในของกลุ่มเมฆนี้将有ความกดอากาศต่ำมาก จนเกือบเป็นสุญญากาศดูสิ่งต่าง ๆ เข้าไปในกระแสอากาศ



ภาพที่ 2.1 ด้านหน้าของพายุฝนฟ้าคะนองจะมีอากาศไหลลงอย่างรุนแรง

ที่มา: <http://www2.aeromet.tmd.go.th/KnowledgeOTS.php>

2.2. ลักษณะอากาศร้ายเนื่องจากพายุฝนฟ้าคะนอง

2.2.1 พายุทอร์นาโด เป็นลักษณะอากาศร้ายรุนแรงที่สุดที่เกิดจากพายุฝนฟ้าคะนอง มีลักษณะเป็นลำเหมือนวงช้างยื่นออกมาจากฐานเมฆ มีลักษณะการหมุนวนบิดเป็นเกลียวมีเส้นผ่าศูนย์กลางประมาณ 1,000 ฟุต มักจะเกิดในที่ราบกว้างใหญ่ ซึ่งเกิดจากเมฆพายุฝนฟ้าคะนองรุนแรงที่มีฐานเมฆต่ำและมีกระแสอากาศไหลลงรุนแรงใกล้พื้นดินดูดเอาอากาศ และวัตถุต่างๆ หมุนวนขึ้นไปในกระแสอากาศนั้น ความรุนแรงนี้สามารถทำให้ต้นไม้ขนาดใหญ่หักโค่นได้ อีกทั้งยังทำให้บ้านเรือนและสิ่งก่อสร้างต่างๆ ได้รับความเสียหายตามแนวที่พายุพัดผ่าน

2.2.2 พายุลูกเห็บ เป็นลักษณะอากาศที่มักจะเกิดขึ้นจากพายุฝนฟ้าคะนองที่มี ยอดเมฆสูงมากซึ่งมีอากาศที่ปั่นป่วนรุนแรง โดยกระแสอากาศที่เคลื่อนที่ขึ้นไปในระดับสูงทำให้หยด น้ำเริ่มแข็งตัวเป็นหยดน้ำแข็งสะสมจนมีขนาดใหญ่ขึ้น และในที่สุดเมื่อกระแสอากาศพยุกรับไม่ได้ ก็จะ ตกลงมาเป็นลูกเห็บ

2.2.3 ฟ้าแลบและฟ้าผ่า เป็นลักษณะอากาศที่เกิดควบคู่กัน โดยเกิดจากการ ปลดปล่อยประจุอิเล็กตรอนระหว่างก้อนเมฆ หรือเกิดขึ้นระหว่างก้อนเมฆกับพื้นดิน เมื่อเกิดความต่าง ศักย์ไฟฟ้าระหว่างตำแหน่งทั้งสองค่าระดับหนึ่งก่อให้เกิดสนามไฟฟ้าขนาดใหญ่ เมื่อกำลังแรงพอก็จะ ทำให้เกิดเป็นกระแสไฟฟ้าไหลผ่านอากาศทำให้เกิดฟ้าแลบในก้อนเมฆ และกระแสไฟฟ้าที่ไหลผ่าน อากาศอย่างเฉียบพลันจากก้อนเมฆทำให้เกิดฟ้าผ่า

2.2.4 ฝนตกหนัก พายุฝนฟ้าคะนองสามารถก่อให้เกิดฝนตกหนัก และน้ำท่วม ฉับพลันได้ในพื้นที่ซึ่งเป็นที่ราบลุ่มหรือที่ต่ำ และพื้นที่ตามบริเวณเชิงเขา

2.3. พายุฝนฟ้าคะนองกับผลกระทบต่อการบิน

พายุฝนฟ้าคะนองเป็นปรากฏการณ์ทางธรรมชาติที่มีอันตรายต่อเครื่องบิน ทั้งขณะ ทำการบินอยู่บนอากาศ ขณะบินอยู่ในเส้นทางบิน หรือขณะกำลังขึ้นหรือลงที่สนามบิน เนื่องจากพายุ ฝนฟ้าคะนองนั้นทำให้เกิดลมกระโชกและมีกระแสอากาศพัดขึ้นลงอย่างรุนแรง ทำให้เครื่องบินที่ท่า การบินผ่านเข้าไปหรือบินอยู่บริเวณใกล้เคียงได้รับผลกระทบ บ่อยครั้งที่ทำให้ความสูงของเครื่องปี เปลี่ยนแปลงกระทันหันและนักบินไม่สามารถรักษาระดับบินไว้ได้ เป็นเหตุให้เครื่องบินประสบ อุบัติเหตุตกมาแล้วมากมาย บางครั้งเครื่องบินจะยังได้รับอันตรายจากลูกเห็บและน้ำแข็งเกาะจับ เครื่องบินอย่างรุนแรงอีกด้วย เครื่องบินจึงควรหลีกเลี่ยงการบินผ่านเข้าไปในพายุฝนฟ้าคะนอง และ ศึกษาระเบียบปฏิบัติต่างๆในการบินเมื่อพบพายุฟ้าคะนองจะช่วยให้นักบินสามารถผ่านพ้นพายุฟ้า คะนองได้อย่างปลอดภัย (กองอุตุนิยมวิทยาการบิน, 2558)

3. สนามบินบริเวณภาคเหนือที่ทำการศึกษา

งานวิจัยนี้ได้ทำการศึกษาสนามบินในพื้นที่ภาคเหนือจำนวน 9 สนามบินจากทั้งหมด 11 สนามบิน ยกเว้นสนามบินตากและสนามบินเพชรบูรณ์ที่ไม่ได้ใช้งานเพื่อบริการสายการบินพาณิชย์ ทั่วไปแต่ใช้สำหรับทางราชการเท่านั้น โดยทั้ง 9 สนามบินนี้มีการรายงานข่าวอากาศการบินรายชั่วโมง โดยกรมอุตุนิยมวิทยา ข้อมูลแต่ละท่าอากาศยานจากเอกสารแถลงข่าวการบิน (Aeronautical

Information Publication: AIP) โดยสำนักงานการบินพลเรือนแห่งประเทศไทย (2566) และข้อมูลลักษณะภูมิประเทศจากศูนย์ภูมิอากาศ กรมอุตุนิยมวิทยา (2566) รายละเอียดของแต่ละสนามบิน ดังนี้

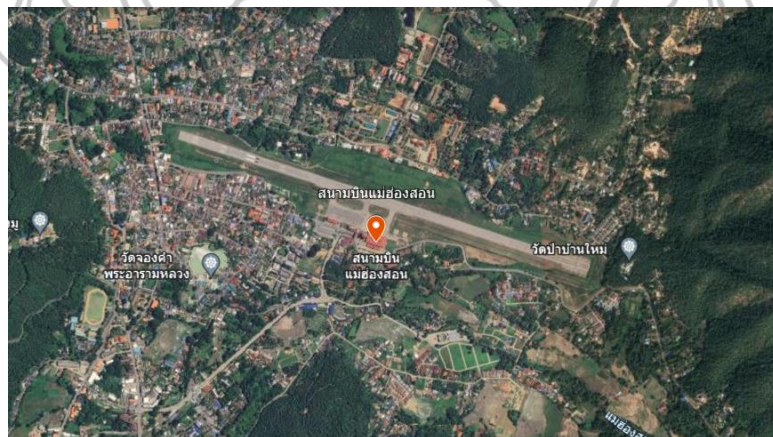
3.1. ท่าอากาศยานแม่ฮ่องสอน

3.1.1 ข้อมูลท่าอากาศยาน

- 1) รหัสสนามบิน ICAO: VTCH
- 2) ระดับความสูงอ้างอิง (Elevation): 929 ฟุต
- 3) พิกัดที่ตั้ง: 19° 18' 06" N 97° 58' 30" E
- 4) เวลารายงานข่าวการบิน: ทุก 1 ชั่วโมง เวลา 05.00-18.00 น.
- 5) ทางวิ่ง: วางตัวในแนว 110 และ 290 องศา ดังภาพที่ 2.2

3.1.2 ข้อมูลลักษณะภูมิประเทศ

สภาพภูมิประเทศส่วนใหญ่ของจังหวัด เป็นภูเขาสูงและป่าไม้ที่เหลือน้อยเป็นที่ราบสองฝั่งลำน้ำและที่ราบในหุบเขามีเทือกเขาที่สูงและสลับซับซ้อนมากมาย เช่น ดอยถ้ำหมากแงง ดอยหนองเขียว ดอยกิวลม ดอยแม่เหาะ และดอยลาน มีแม่น้ำสำคัญ 2 สาย คือ แม่น้ำปาย ไหลจากตอนเหนือของอำเภอปาย ผ่านท้องที่อำเภอเมือง ลงทางใต้เข้าสู่เขตสหภาพเมียนมา ไปบรรจบกับแม่น้ำสาละวิน (หรือคง) ที่สบปายและแม่น้ำยม ไหลจากทิศตะวันออกของอำเภอขุนยวม ผ่านอำเภอแม่สะเรียงลงไปทางใต้บรรจบกับแม่น้ำเมยในเขตอำเภอแม่ระมาด จังหวัดตาก



ภาพที่ 2.2 สภาพพื้นที่บริเวณท่าอากาศยานแม่ฮ่องสอน

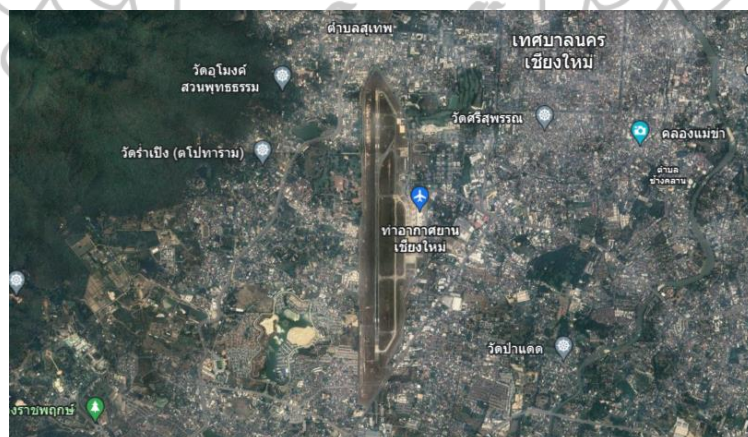
3.2. ทำอากาศยานนานาชาติเชียงใหม่

3.2.1 ข้อมูลทำอากาศยาน

- 1) รหัสสนามบิน ICAO: VTCC
- 2) ระดับความสูงอ้างอิง (Elevation): 1,036 ฟุต
- 3) พิกัดที่ตั้ง: 18° 46' 17" N 98° 57' 46" E
- 4) เวลารายงานข่าวการบิน: ทุก 30 นาที ตลอด 24 ชั่วโมง
- 5) ทางวิ่ง: วางตัวในแนว 180 และ 360 องศา ดังภาพที่ 2.3

3.2.2 ข้อมูลลักษณะภูมิประเทศ

สภาพภูมิประเทศสภาพภูมิประเทศส่วนใหญ่ของจังหวัดเป็นภูเขาและที่ราบเชิงเขา มีเทือกเขาแดนลาวอยู่ทางด้านเหนือของจังหวัดกั้นพรมแดนไทย-พม่า มียอดเขาที่สำคัญคือ ดอยผ้าห่มปกและดอยอ่างขาง ซึ่งอยู่ในอำเภอฝาง เทือกเขาถนนธงชัยอยู่ทางตะวันตกของแม่น้ำปิงมี ยอดเขาที่สำคัญคือ ดอยอินทนนท์ซึ่งอยู่ในอำเภอจอมทอง ซึ่งเป็นยอดเขาที่สูงที่สุดของประเทศไทย และดอยสุเทพซึ่งอยู่ในอำเภอเมือง เทือกเขาผีปันน้ำอยู่ทางตะวันออกของแม่น้ำปิง มียอดดอยขุนตาล อยู่ในอำเภอพร้าว ดอยสะเก็ด และสันกำแพง นอกจากนี้จังหวัดเชียงใหม่ยังเป็นแหล่งต้นน้ำลำธารหลายสาย เช่น แม่น้ำปิง (มีต้นน้ำจากดอยถั่วในอำเภอเชียงดาว) แม่น้ำแม่แจ่ม แม่น้ำฝาง เป็นต้น ส่วนพื้นที่ราบลุ่มของจังหวัดส่วนใหญ่เป็นที่ราบลุ่มน้ำแม่ปิง ครอบคลุมพื้นที่บางส่วนของอำเภอเชียงดาว แม่แตง เมือง สารภีและดอยเต่า อีกส่วนเป็นที่ราบลุ่มแม่น้ำกก ครอบคลุมพื้นที่อำเภอฝาง และแม่เอย



ภาพที่ 2.3 สภาพพื้นที่บริเวณทำอากาศยานนานาชาติเชียงใหม่

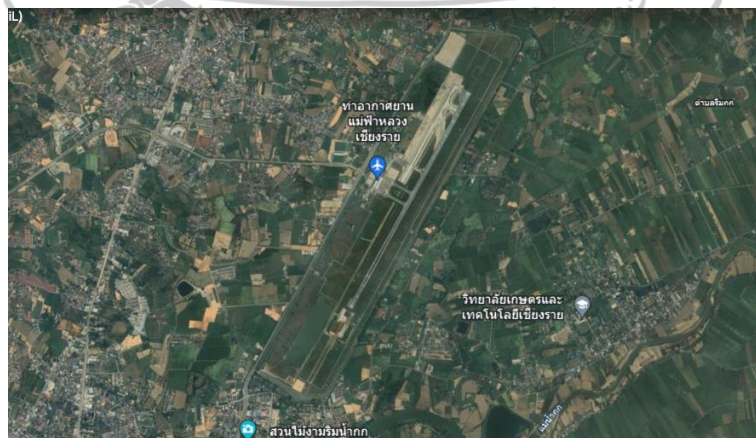
3.3. ท่าอากาศยานแม่ฟ้าหลวงเชียงราย

3.3.1 ข้อมูลท่าอากาศยาน

- 1) รหัสสนามบิน ICAO: VTCT
- 2) ระดับความสูงอ้างอิง (Elevation): 1,280 ฟุต
- 3) พิกัดที่ตั้ง: 19° 57' 08" N 99° 52' 59" E
- 4) เวลารายงานข่าวการบิน: ทุก 1 ชั่วโมง ตลอด 24 ชั่วโมง
- 5) ทางวิ่ง: วางตัวในแนว 30 และ 210 องศา ดังภาพที่ 2.4

3.3.2 ข้อมูลลักษณะภูมิประเทศ

สภาพภูมิประเทศสภาพภูมิประเทศเป็นลักษณะภูเขาล้อมรอบและมีที่ว่างระหว่างทิวเขา ด้านตะวันตกมีทิวเขาด้านตรงซัดทอดจากประเทศสหภาพพม่าลงมาทางใต้ ทางตะวันออกมีทิวเขาดอยยาวทอดจากอำเภอเชียงของลงมาถึงอำเภอเทิงและทิวเขาดอยผาหม่นซึ่งแบ่งกั้นพรมแดนไทย-ลาว ทอดยาวจากอำเภอเชียงของลงมาถึงอำเภอเชียงคำเลยไปถึงอำเภอทุ่งช้างและอำเภอปัว จังหวัดน่าน ดังนั้นที่อันกว้างใหญ่ของจังหวัดเชียงรายจึงเต็มไปด้วยภูเขายอดเขาสูงที่สำคัญ มีดอยตุง สูง 1,928 เมตร ดอยสามเส้นหรือดอยแม่สะลองสูง 1,673 เมตร ดอยยาว สูง 1,400 เมตร ดอยกู่สูง 1,608 เมตร และดอยภูลังกา สูง 1,693 เมตร ส่วนแม่น้ำที่สำคัญมีแม่น้ำโขง แม่น้ำกก และแม่น้ำลาว พื้นที่ราบของจังหวัดนี้ส่วนมากอยู่ระหว่างหุบเขา จึงเป็นที่ราบที่มีระดับสูงและมีขนาดไม่กว้างนักที่นับว่าสำคัญ ได้แก่ ที่ราบสูงเชียงราย ซึ่งอยู่ในระดับความสูง 378 เมตร มีทิวเขาล้อมรอบเกือบทุกด้านทางเหนือเป็นทิวเขาแดนลาว ทางตะวันตกมีทิวเขาดอยสามเส้นหรือดอยแม่สะลอง และดอยช้างในทิวเขาแดนลาวทางใต้ และทางตะวันออกเป็นแนวทิวเขาฝิ่ปันน้ำ



ภาพที่ 2.4 สภาพพื้นที่บริเวณท่าอากาศยานแม่ฟ้าหลวงเชียงราย

3.4 ท่าอากาศยานลำปาง

3.4.1 ข้อมูลท่าอากาศยาน

- 1) รหัสสนามบิน ICAO: VTCL
- 2) ระดับความสูงอ้างอิง (Elevation): 811 ฟุต
- 3) พิกัดที่ตั้ง: 18° 16' 20" N 99° 30' 15" E
- 4) เวลารายงานข่าวการบิน: ทุก 1 ชั่วโมง เวลา 06.00-20.00 น.
- 5) ทางวิ่ง: วางตัวในแนว 175 และ 355 องศา ดังภาพที่ 2.5

3.4.2 ข้อมูลลักษณะภูมิประเทศ

สภาพภูมิประเทศโดยทั่วไปเป็นรูปยารี่พื้นที่ส่วนใหญ่เป็นภูเขาที่มีราบสูงและที่ราบลุ่มในบริเวณแม่น้ำวัง อยู่สูงจากระดับน้ำทะเลประมาณ 268 เมตร พื้นที่ป่าประกอบด้วยไม้สักและไม้เบญจพรรณ จังหวัดลำปางมีภูเขาที่สำคัญ ได้แก่ ดอยขุนตาล กั้นเขตแดนจังหวัดลำปางกับจังหวัดลำพูน มีอุโมงค์เจาะลอดภูเขาที่ยาว 1,362 เมตร เป็นอุโมงค์รถไฟที่ยาวที่สุดในประเทศไทย ดอยผีปันน้ำหรือดอยนางแก้ว อยู่ในเขตอำเภอวังเหนือ กั้นเขตแดนระหว่างจังหวัดลำปางกับเชียงราย สูงประมาณ 1,500 เมตร ดอยผาจ้อ ดอยกึ่งมณฑาและดอยขุนแม่เหล็ก แม่น้ำที่สำคัญ ได้แก่ แม่น้ำวัง แม่น้ำจาว และแม่น้ำจาง



ภาพที่ 2.5 สภาพพื้นที่บริเวณท่าอากาศยานลำปาง

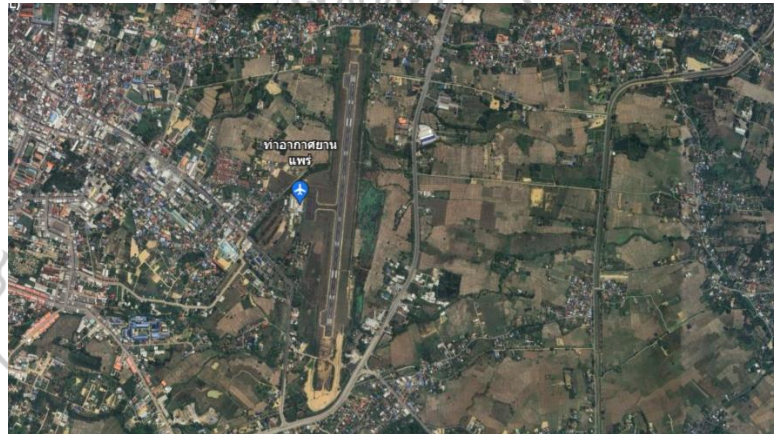
3.5 ท่าอากาศยานแพร่

3.5.1 ข้อมูลท่าอากาศยาน

- 1) รหัสสนามบิน ICAO: VTCP
- 2) ระดับความสูงอ้างอิง (Elevation): 538 ฟุต
- 3) พิกัดที่ตั้ง: $18^{\circ} 07' 54''$ N $100^{\circ} 09' 52''$ E
- 4) เวลาขานงานข่าวการบิน: ทุก 1 ชั่วโมง เวลา 05.00-18.00 น.
- 5) ทางวิ่ง: วางตัวในแนว 5 และ 185 องศา ดังภาพที่ 2.6

3.5.2 ข้อมูลลักษณะภูมิประเทศ

สภาพภูมิประเทศมีภูเขาล้อมรอบทั้ง 4 ทิศ พื้นที่ภายในวงรอบเป็นพื้นที่ราบ ซึ่งลาดเอียงไปทางใต้ลงแม่น้ำยม ลักษณะคล้ายอ่างหรือท้องกระทะ พื้นที่ราบส่วนใหญ่เป็นที่อยู่อาศัย และประกอบอาชีพในทางการเกษตรกรรม ภูเขาที่ล้อมรอบตัวจังหวัดแพร่ทั้ง 4 ทิศนี้มีลักษณะคล้ายภูเขาติดต่อกันเป็นพืด มีความสูงพอสมควรตามแนวตะเข็บติดต่อกับจังหวัดใกล้เคียง



ภาพที่ 2.6 สภาพพื้นที่บริเวณท่าอากาศยานแพร่

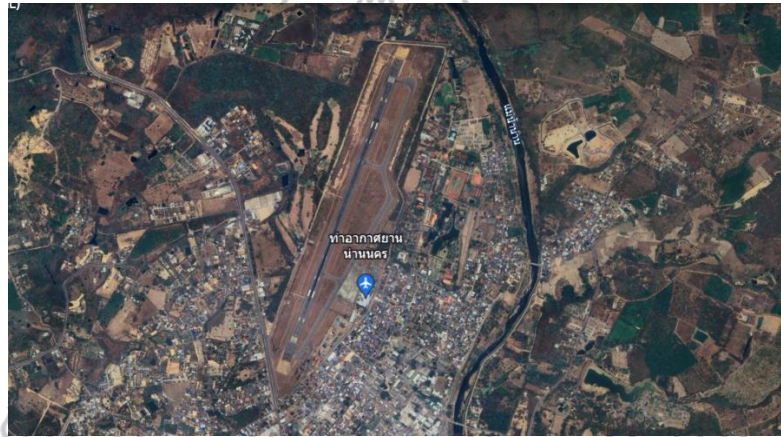
3.6 ท่าอากาศยานน่านนคร

3.6.1 ข้อมูลท่าอากาศยาน

- 1) รหัสสนามบิน ICAO: VTCN
- 2) ระดับความสูงอ้างอิง (Elevation): 685 ฟุต
- 3) พิกัดที่ตั้ง: $18^{\circ} 48' 28''$ N $100^{\circ} 47' 00''$ E
- 4) เวลาขานงานข่าวการบิน: ทุก 1 ชั่วโมง เวลา 05.00-21.00 น.
- 5) ทางวิ่ง: วางตัวในแนว 20 และ 300 องศา ดังภาพที่ 2.7

3.6.2 ข้อมูลลักษณะภูมิประเทศ

สภาพภูมิประเทศโดยทั่วไปเป็นป่าและภูเขา 3 ส่วน มีพื้นที่ราบ 1 ส่วน มีความหนาแน่นของป่าและภูเขาทางอยู่ด้านเหนือ ตะวันออกเฉียงเหนือและใต้ ส่วนพื้นที่ราบอยู่ทางด้านตะวันออกตามแนวฝั่งแม่น้ำ พื้นที่มีลักษณะลาดเอียงจากเหนือไปได้ ภูมิประเทศส่วนใหญ่เป็นทิวเขาสลับซับซ้อนทั่วไป พื้นที่ราบเหมาะแก่การเพาะปลูกมีน้อย ภูเขาที่สำคัญ ได้แก่ ภูเขาเขต อำเภอบัว ซึ่งอยู่ติดชายแดนลาวภูเขาเป็นดอยที่สูงที่สุดอยู่ในอำเภอบัว สูงถึง 1,980 เมตร ดอยผากิ อยู่ในเขตอำเภอบ้านหลวง ติดกับอำเภอยางชุมน้อย จังหวัดศรีสะเกษ แม่น้ำที่สำคัญ ได้แก่ แม่น้ำน่าน ซึ่งมีต้นน้ำอยู่ที่บริเวณดอยภูแวไหลผ่านจังหวัดอุดรธานี พิษณุโลก ไปบรรจบกันแม่น้ำปิงที่จังหวัดนครสวรรค์ ยาวประมาณ 300 กิโลเมตร



ภาพที่ 2.7 สภาพพื้นที่บริเวณท่าอากาศยานน่านนคร

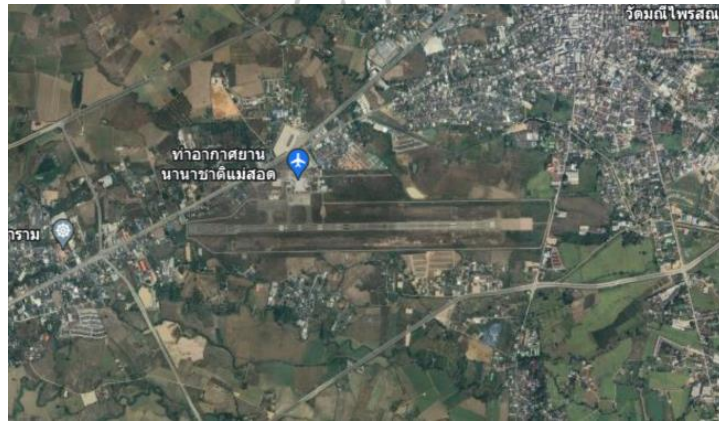
3.7 ท่าอากาศยานนานาชาติแม่สอด

3.7.1 ข้อมูลท่าอากาศยาน

- 1) รหัสสนามบิน ICAO: VTPM
- 2) ระดับความสูงอ้างอิง (Elevation): 690 ฟุต
- 3) พิกัดที่ตั้ง: 16° 41' 59" N 98° 32' 36" E
- 4) เวลารายงานข่าวการบิน: ทุก 1 ชั่วโมง เวลา 05.00-18.00 น.
- 5) ทางวิ่ง: วางตัวในแนว 90 และ 270 องศา ดังภาพที่ 2.8

3.7.2 ข้อมูลลักษณะภูมิประเทศ

สภาพภูมิประเทศส่วนใหญ่ของจังหวัดตากเป็นป่าและภูเขาสลับซับซ้อน มีเทือกเขาถนนธงชัยเป็นแนวยาวตั้งแต่เหนือจรดใต้แบ่งพื้นที่จังหวัดออกเป็น 2 ด้าน คือ ด้านตะวันออกและด้านตะวันตก โดยอำเภอแม่สอดอยู่ทางด้านตะวันตกของจังหวัด เต็มไปด้วยป่าไม้หนาแน่น และเทือกเขาสลับซับซ้อน สภาพดินอุดมสมบูรณ์กว่าด้านตะวันออกและมีที่ราบเพื่อการเกษตรอยู่มาก โดยเฉพาะท้องที่อำเภอแม่สอด และอำเภอแม่ระมาด มีทุ่งหญ้าอยู่บ้าง ป่าไม้ส่วนใหญ่เป็นป่าดงดิบ ป่าสน ป่าเบญจพรรณ และป่าแดง



ภาพที่ 2.8 สภาพพื้นที่บริเวณท่าอากาศยานนานาชาติแม่สอด

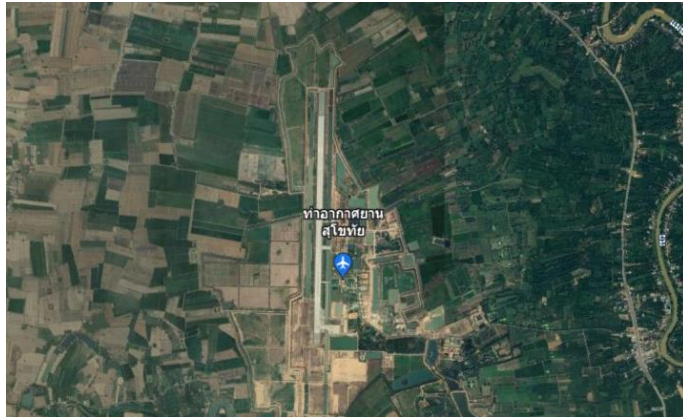
3.8 ท่าอากาศยานสุโขทัย

3.8.1 ข้อมูลท่าอากาศยาน

- 1) รหัสสนามบิน ICAO: VTPO
- 2) ระดับความสูงอ้างอิง (Elevation): 179 ฟุต
- 3) พิกัดที่ตั้ง: 17° 14' 16" N 99° 49' 06" E
- 4) เวลารายงานข่าวการบิน: ทุก 1 ชั่วโมง เวลา 05.00-18.00 น.
- 5) ทางวิ่ง: วางตัวในแนว 180 และ 360 องศา ดังภาพที่ 2.9

3.8.2 ข้อมูลลักษณะภูมิประเทศ

สภาพภูมิประเทศส่วนใหญ่ของจังหวัดสุโขทัยเป็นที่ราบลุ่ม มีแม่น้ำยมไหลผ่านเป็นระยะทางประมาณ 170 กิโลเมตร พื้นที่ทางด้านเหนือเป็นที่ราบสูงมีภูเขาขนาบเป็นพืดมาทางทิศตะวันตก พื้นที่ตอนกลางเป็นที่ราบลุ่มและตอนใต้เป็นที่ราบสูง จังหวัดสุโขทัยมีภูเขาขนาดย่อมอยู่หลายลูก ภูเขาที่สูงที่สุด คือ เขาหลวง ซึ่งสูงกว่าระดับน้ำทะเลปานกลางประมาณ 1,185 เมตร



ภาพที่ 2.9 สภาพพื้นที่บริเวณท่าอากาศยานสุโขทัย

3.9 ท่าอากาศยานพิษณุโลก

3.9.1 ข้อมูลท่าอากาศยาน

- 1) รหัสสนามบิน ICAO: VTPP
- 2) ระดับความสูงอ้างอิง (Elevation): 145 ฟุต
- 3) พิกัดที่ตั้ง: $16^{\circ} 46' 58'' N$ $100^{\circ} 16' 44'' E$
- 4) เวลาทำงานข่าวการบิน: ทุก 1 ชั่วโมง เวลา 05.00-21.00 น.
- 5) ทางวิ่ง: วางตัวในแนว 140 และ 320 องศา ดังภาพที่ 2.10

3.9.2 ข้อมูลลักษณะภูมิประเทศ

สภาพภูมิประเทศของจังหวัดพิษณุโลก แยกออกได้เป็น 2 ลักษณะ คือด้านตะวันออกและตะวันออกเฉียงเหนือมีลักษณะเป็นภูเขา และมียอดเขาสูงที่สุด ได้แก่ เขาข้อโปสูง 1,028 เมตร ส่วนทางด้านตะวันตกของจังหวัดเป็นที่ราบลุ่ม มีแม่น้ำสำคัญที่อยู่ในเขตอำเภอเมืองพิษณุโลก ได้แก่ แม่น้ำน่าน เป็นแม่น้ำที่สำคัญที่สุดของจังหวัด มีต้นกำเนิดอยู่ในภาคเหนือตอนบน บริเวณทิวเขาผีปันน้ำ ทิวเขาหลวงพระบางและทิวเขาเพชรบูรณ์ไหลลงมาทางใต้ผ่านจังหวัดน่าน อุตรดิตถ์และเข้าเขตจังหวัดพิษณุโลกทางตอนเหนือบริเวณบ้านโคกเทียม อำเภอพรหมพิราม ไหลแบ่งครึ่งเขตอำเภอเมืองพิษณุโลก ผ่านอำเภอบางกระพุ่มแล้วออกพ้นเขตจังหวัดที่บ้านนามคลีอำเภอบางกระพุ่ม ความยาวของแม่น้ำในเขตจังหวัดพิษณุโลก 127.5 กิโลเมตร



ภาพที่ 2.10 สภาพพื้นที่บริเวณท่าอากาศยานพนมเปญ

4. ข่าวอากาศการบิน

กรมอุตุนิยมวิทยา (2563) อธิบายความหมายของการรายงานข่าวอากาศการบิน (Meteorological Terminal Air Report: METAR) หมายถึง รูปแบบการรายงานผลการตรวจวัดค่า สสารประกอบอุตุนิยมวิทยาและสภาพอากาศที่เกิดขึ้นภายในบริเวณท่าอากาศยาน ซึ่งทำการรายงาน เป็นประจำทุกครึ่งชั่วโมงสำหรับสนามบินนานาชาติที่มีเที่ยวบินจำนวนมาก เช่น ท่าอากาศยานดอนเมือง ท่าอากาศยานสุวรรณภูมิ ท่าอากาศยานนานาชาติเชียงใหม่ เป็นต้น หรือทำการรายงานเป็น ประจำทุกหนึ่งชั่วโมงสำหรับสนามบินส่วนใหญ่ในประเทศ โดยมีวัตถุประสงค์เพื่อนำไปใช้สำหรับ ประกอบการวางแผนการบิน (Flight plan) และใช้สำหรับเผยแพร่กระจายข่าวออกไปให้แก่ หน่วยงานทางการบินที่เกี่ยวข้อง เช่น หอควบคุมสนามบิน หน่วยงานควบคุมการจราจรทาง อากาศใกล้เขตสนามบิน หน่วยงานควบคุมจราจรทางอากาศระดับภูมิภาค ศูนย์ข่าวสารการบิน สถานี สื่อสาร เป็นต้น โดยเนื้อหาและรูปแบบของรายงานข่าวอากาศการบินเป็นไปตามข้อกำหนดของ องค์การอุตุนิยมวิทยาโลก (World Meteorological Organization: WMO) และองค์การการบินพล เรือนระหว่างประเทศ (International Civil Aviation Organization: ICAO) รูปแบบของข่าว METAR ประกอบไปด้วยส่วนต่างๆ 13 ส่วน ดังนี้

4.1 ส่วนที่ 1 ประเภทของรายงาน (Identification of the type of report) โดย ระบุประเภทของรายงาน สำหรับรายงานอากาศการบินแบบประจำ คือ METAR

4.2 ส่วนที่ 2 รหัสท่าอากาศยาน (Location indicator) โดยระบุรหัสท่าอากาศยาน ด้วยตัวอักษรภาษาอังกฤษ 4 ตัวอักษรตามข้อกำหนดของ ICAO ซึ่งกำหนดไว้ในเอกสาร Doc 7910

ตัวอย่างเช่น รหัสท่าอากาศยานนานาชาติเชียงใหม่ คือ VTCC และรหัสท่าอากาศยานพิบูลย์โลก คือ VTPT เป็นต้น

4.3 ส่วนที่ 3 เวลาตรวจอากาศการบิน (Time of the Observation) โดยระบุเวลาที่ทำการตรวจอากาศการบิน ประกอบด้วยวันที่และเวลาเป็นชั่วโมงและนาที ตามเวลามาตรฐานสากล (Coordinate Universal Time: UTC (Z)) ตัวอย่างเช่น ตรวจอากาศการบินวันที่ 28 เวลา 08:00 น. ดังนั้นเวลาตรวจอากาศการบินในรายงาน คือ 280100Z

4.4 ส่วนที่ 4 การตรวจอากาศการบินด้วยระบบอัตโนมัติ (Identification of an automated) ถ้าสถานีตรวจอากาศการบินใช้ระบบตรวจอากาศแบบอัตโนมัติจะต้องระบุในรายงานด้วยคำว่า AUTO

4.5 ส่วนที่ 5 ลมผิวพื้น (Surface wind) โดยข้อมูลลมผิวพื้นสำหรับรายงานข่าว METAR นำมาจากเครื่องวัดลม (Anemometer) ที่ติดตั้งอยู่ในท่าอากาศยาน เครื่องวัดลมควรติดตั้งอยู่ในตำแหน่งที่เหมาะสมเพื่อให้เป็นค่าตัวแทนในการวัดสภาพของลมทั่วทั้งทางวิ่ง ซึ่งมีทางวิ่งเดียวหรือทางวิ่งทั้งหมดที่มีอยู่

4.5.1 รูปแบบรหัสลมผิวพื้น คือ dddffG_{f_m}KT d_nd_nd_nVd_xd_xd_x โดยที่ ddd คือ ทิศทางของลมผิวพื้นเฉลี่ยในช่วง 10 นาทีที่ผ่านมา, ff คือ ความเร็วของลมผิวพื้นเฉลี่ย ในช่วง 10 นาทีที่ผ่านมา, f_mf_m คือ ลมกระโชก (Gust) แรงสูงสุดในช่วง 10 นาทีที่ผ่านมา, d_nd_nd_nVd_xd_xd_x คือ ความแปรผันของทิศทางลมผิวพื้นตามเข็มนาฬิกาในช่วง 10 นาทีที่ผ่านมา

4.5.2 ลมกระโชก (Gust) คือ ลมที่มีความเร็วแปรปรวนไปจากความเร็วลมเฉลี่ย ตั้งแต่ 10 นอต ขึ้นไป แรงสูงสุดใน 10 นาทีก่อนเวลาตรวจอากาศ (หรือเกิดการเปลี่ยนแปลงที่มีความไม่ต่อเนื่องอย่างเห็นได้ชัด) จะรายงานเมื่อลมกระโชกมีค่ามากกว่าความเร็วลมเฉลี่ยตั้งแต่ 10 นอตขึ้นไป

4.5.3 กลุ่ม d_nd_nd_nVd_xd_xd_x คือ ความแปรผันของทิศทางลมผิวพื้นตามเข็มนาฬิกาในช่วง 10 นาทีที่ผ่านมา ตั้งแต่ 60 องศาหรือมากกว่า แต่น้อยกว่า 180 องศา และความเร็วลมเฉลี่ยมากกว่า 3 นอต ต้องรายงานทิศทางลมที่มีการแปรผันไปตามลำดับ โดยมีทิศทางตามเข็มนาฬิกา เช่น 290V090 หรือ 170V250

4.5.4 หลักเกณฑ์อื่นๆในการเข้ารหัสลมผิวพื้น ได้แก่

1) เมื่อมีการเปลี่ยนแปลงของทิศทางลมในช่วง 10 นาทีก่อนเวลาตรวจ (หรือเกิดการเปลี่ยนแปลงที่มีความไม่ต่อเนื่องอย่างเห็นได้ชัด) ตั้งแต่ 60 องศาหรือมากกว่า แต่น้อยกว่า 180 องศา และความเร็วลมเฉลี่ยน้อยกว่า 3 นอต ค่าลมในกรณีนี้จะต้องรายงานเป็นลมแปรปรวน (Variable: VRB) เช่น VRB01KT หรือ VRB02KT

2) เมื่อมีการเปลี่ยนแปลงทิศทางลมในช่วง 10 นาทีก่อนเวลาตรวจอากาศ (หรือเกิดการเปลี่ยนแปลงที่มีความไม่ต่อเนื่องอย่างเห็นได้ชัด) ตั้งแต่ 180 องศาหรือมากกว่า หรือไม่สามารถรายงานทิศทางลมได้เช่น เมื่อ พายุฝนฟ้าคะนองผ่านท่าอากาศยาน ค่าลมในกรณีนี้จะต้องรายงานเป็นลมแปรปรวน (Variable: VRB) เช่น VRB28KT

3) ทิศทางและความแปรปรวนของทิศทางลมผิวพื้นจะถูกรายงานให้เป็นจำนวนเต็ม 10 องศาที่ใกล้ที่สุด

4) เมื่อความเร็วลมต่ำกว่า 1 นอต ให้รายงานเป็นลมสงบ (Calm) โดยที่ ลมสงบ เข้ารหัสเป็น 00000KT

5) ความเร็วลมผิวพื้นเฉลี่ยและความเร็วสูงสุดจะถูกปิดเศษเป็นจำนวนเต็ม การรายงานความเร็วลม ผิวพื้นตั้งแต่ 01 ถึง 99 นอต (Knots) ถ้าความเร็วตั้งแต่ 100 นอตขึ้นไป เข้ารหัสเป็น “P99”

4.6 ส่วนที่ 6 ทิศนวิสัย (Visibility) ในรายงานข่าว METAR การรายงานค่าทัศนวิสัย เป็นการรายงานค่า Prevailing Visibility ซึ่งหมายถึง ค่าทัศนวิสัยในการมองเห็นที่ใกล้ที่สุดอย่างน้อยครึ่งวงกลมหรือมากกว่า 50% รอบๆท่าอากาศยาน โดยพื้นที่เหล่านี้อาจเป็นส่วนที่ต่อเนื่องกันหรือไม่ต่อเนื่องกันก็ได้ และในบางปรากฏการณ์ต้องรายงานค่าทัศนวิสัยต่ำสุดด้วย โดยที่การรายงานควรได้จากการตรวจวัดที่ระดับความสูงประมาณ 1.5 เมตรเหนือพื้นดินที่บริเวณจุดตรวจ ผู้ปฏิบัติงานควรตระหนักถึงข้อผิดพลาดที่อาจเป็นไปได้ ที่เกิดขึ้นจากการตรวจวัดในมุมเฉียง (slant visibility) เมื่อทำการตรวจวัดค่าทัศนวิสัยที่ระดับความสูงมากกว่า 1.5 เมตรเหนือพื้นดิน

4.6.1 รูปแบบรหัสทัศนวิสัย คือ VVVV VNVNVNDv โดยที่ VVVV คือ ค่า Prevailing Visibility, VNVNVNVN คือ ค่าทัศนวิสัยต่ำสุด, Dv คือ ทิศของทัศนวิสัยที่ต่ำสุดตามนियามการมองเห็นทางอุตุนิยมวิทยา (Meteorological Optical Range: MOR) รายงานด้วยตัวอักษรหนึ่งหรือสองตัว โดยระบุด้วยทิศทั้ง 8 ทิศ (ทิศทั้ง 8 ทิศ ได้แก่ N NE E SE S SW W NW) จากจุดอ้างอิงสนามบิน (Aerodrome Reference Point: ARP)

4.6.2 หลักเกณฑ์อื่นๆในการเข้ารหัสทัศนวิสัย

1) รหัส VVVV ในการรายงานข่าว METAR จะรายงานเป็นตัวเลขจำนวนสี่ตัว เมื่อทัศนวิสัยตั้งแต่ 10 กิโลเมตร ขึ้นไปจะใช้รหัส “9999”

2) ถ้ามีทัศนวิสัยในทิศทางอื่นซึ่งไม่ใช่ทัศนวิสัยที่เป็น Prevailing Visibility มีค่าต่ำกว่า 1500 เมตร หรือต่ำกว่า 50% ของ Prevailing Visibility ควรรายงานต่อจากค่า Prevailing Visibility และระบุทิศทางที่เกี่ยวข้องกับท่าอากาศยาน โดยอ้างอิงจากทิศทั้ง 8 ทิศ ถ้าตรวจพบทัศนวิสัยต่ำสุดมากกว่าหนึ่งทิศทางควรรายงานทิศที่มีความสำคัญที่สุด

3) เมื่อสภาพอากาศเข้าสู่เงื่อนไขที่ต้องรายงาน “CAVOK” จะใช้ CAVOK ในการรายงาน โดยไม่ต้องรายงานกลุ่มทัศนวิสัย

4.7 ส่วนที่ 7 พิสัยการมองเห็นบนทางวิ่ง (Runway Visual Range: RVR) เป็นการวัดค่าการมองเห็นในแนวนอนตามแนวทางวิ่ง การวัดค่า RVR ทำได้โดยใช้เครื่องมือที่เรียกว่า Transmissometer โดยระบบเครื่องมือตรวจวัดค่า RVR อาจมีเซ็นเซอร์อยู่ที่ ทักษดาวนโซน (Touchdown zone) จุดกึ่งกลาง (Mid-point) และจุดสิ้นสุด (Stop end) ของทางวิ่ง แต่ในการรายงานข่าว METAR จะมีการวัดค่าบริเวณที่ซาดาวนโซนเท่านั้น ถ้าหากค่า RVR ที่บริเวณที่ซาดาวนโซนไม่สามารถใช้ได้ การรายงานกลุ่ม RVR สำหรับทางวิ่งนั้นจะถูกละเว้นไม่ต้องรายงาน และจะรายงานเฉพาะเมื่อมีค่าทัศนวิสัยต่ำสุด หรือค่าการมองเห็นบนทางวิ่งมีค่าน้อยกว่า 1500 เมตรรูปแบบรหัสพิสัยการมองเห็นบนทางวิ่ง คือ RDRDR/VRVRVRVR โดยที่ R คือ ตัวระบุกลุ่ม, DRDR คือ ตัวกำหนดตำแหน่งของทางวิ่งที่อยู่ใกล้จุดที่ตรวจวัดค่า RVR ที่สุด ตามด้วยอักษร L, C หรือ R (ถ้ามีมากกว่าหนึ่งทางวิ่ง) เพื่อแยกแยะระหว่างทางวิ่งซ้าย กลาง ขวาตามลำดับ, VRVRVRVR คือ ค่า RVR ที่รายงานระยะเป็นเมตร

4.8 ส่วนที่ 8 สภาพอากาศปัจจุบัน (Present weather) ในการตรวจสภาพอากาศปัจจุบัน จะเกี่ยวข้องกับสภาพอากาศที่เกิดขึ้นในขณะที่ทำการตรวจบริเวณท่าอากาศยานเท่านั้น ในกรณีที่มีการรายงานปรากฏการณ์ในบริเวณใกล้เคียงกับท่าอากาศยาน (in the vicinity: VC) ภายในระยะประมาณ 8-16 กิโลเมตรจากจุดอ้างอิงสนามบิน อาจใช้จุดอ้างอิงจากการมองเห็นข้อมูลจากเรดาร์ตรวจสภาพอากาศ รายงานจากนักบิน และ เซ็นเซอร์อัตโนมัติเพื่อช่วยในการตัดสินใจ โดยถ้าไม่มีสภาพอากาศที่มีนัยสำคัญต่อการปฏิบัติงานด้านการบิน ในขณะที่ทำการตรวจอากาศทั้งในบริเวณท่าอากาศยานหรือบริเวณใกล้เคียงกับท่าอากาศยาน รหัสกลุ่มนี้จะถูกละเว้นไม่ต้องรายงาน ตัวอย่างของปรากฏการณ์สภาพอากาศที่สำคัญ ดังนี้

4.8.1 ฝนละออง (Drizzle: DZ) คือ หยดน้ำที่ละเอียดมากที่มีเส้นผ่าศูนย์กลางน้อยกว่า 0.5 มิลลิเมตร ตกลงมาจากเมฆ Stratiform บางๆ เมื่อละอองฝน ตกกระทบบนผิวน้ำจะไม่สามารถมองเห็นได้ แต่ถ้าตกปรอย ๆ อย่างต่อเนื่องอาจทำให้เกิดการไหลจากหลังคาและพื้นผิวทางวิ่ง ทัศนวิสัยมีความสัมพันธ์กับความความรุนแรงของฝนที่ตกลงมาและ จำนวนหยดน้ำ นอกจากนี้โดยทั่วไปฝนละอองที่ตกหนักจะตกจากฐานล่างของเมฆที่อยู่ต่ำ ฝนละอองที่ตกเบาๆ มีความสอดคล้องกับการไหลลงจากหลังคาเพียงเล็กน้อย ในขณะที่ฝนละอองที่ตกปานกลางมักเกี่ยวข้องกับการมองเห็นหรือทัศนวิสัยน้อยกว่า 3000 เมตร มีน้ำฝนไหลลงมาจากหลังคา ฝนละอองที่ตกหนักสอดคล้องกับอัตราการสะสมของปริมาณน้ำฝนมากกว่า 1 มิลลิเมตรต่อชั่วโมง

4.8.2 หมอก (Fog: FG) คือ ละอองน้ำแขวนลอยในอากาศที่มีขนาดเล็กมาก ทิศนวิสัยภายในละอองน้ำแขวนลอยในอากาศนั้น น้อยกว่า 1000 เมตร ความชื้นสัมพัทธ์มีค่าใกล้เคียงหรือเท่ากับ 100% (อุณหภูมิกอากาศและจุดน้ำค้างจะมีค่าเท่ากันหรือใกล้เคียงกันมาก)

4.8.3 ลูกเห็บ (Hail: GR) คือ ลูกบอลโปร่งแสงหรือทึบแสง หรือชิ้นน้ำแข็ง (Hailstone) ที่มีเส้นผ่านศูนย์กลาง 5 มิลลิเมตรขึ้นไป ตกจากก้อนเมฆที่ก่อตัวในทางตั้ง (TCU, CB) โดยมีกระแสอากาศไหลเวียนภายในก้อนเมฆที่รุนแรง เมื่อตกลง บนพื้นแข็งจะกระเด็นกระดอนมีเสียงดัง

4.8.4 ฟ้าหลัวแห้ง (Haze: HZ) คือ สิ่งเกิดจากฝุ่นละอองขนาดเล็กที่มีขนาดเล็กมากในอากาศ มองไม่เห็นด้วยตาเปล่า แต่มีจำนวนมากพอที่จะทำให้อากาศมีลักษณะเป็นสีซีดขาว หรือมีสีซีดจาง รายงานเมื่อตรวจพบทัศนวิสัยลดลงมาถึง 5000 เมตร มีความชื้นสัมพัทธ์มีค่าน้อยกว่า 80%

4.8.5 ฟ้าหลัวชื้น (Mist: BR) คือ ละอองน้ำขนาดเล็กแขวนลอยอยู่ในอากาศ หรืออนุภาคที่ดูดความเปียกชื้น ในอากาศ รายงานเมื่อทัศนวิสัยทั่วไปลดลงมาอยู่ในช่วงระหว่าง 1000 ถึง 5000 เมตร ความชื้นสัมพัทธ์มีค่า ตั้งแต่ 80% ขึ้นไป

4.8.6 ฝน (Rain: RA) คือ หยาดน้ำฟ้า (precipitation) ที่ตกลงมาจากเมฆ เป็นของเหลวมีเส้นผ่าศูนย์กลาง ของขนาดหยดน้ำเกิน 0.5 มิลลิเมตรขึ้นไป โดยความเข้มข้นของหยดน้ำ อาจแตกต่างกันไปตามความรุนแรง ของหยาดน้ำฟ้า ลักษณะทางธรรมชาติและแหล่งกำเนิด เช่น ขนาดเบา (Light) ขนาดปานกลาง (Moderate) ขนาดหนัก (Heavy) ตกเป็นระยะๆ หรือตกต่อเนื่องกัน ถ้าฝนที่ตกลงมาจากเมฆก่อตัวทางตั้งสามารถอธิบาย ได้ว่าเป็นฝนโปรย

4.8.7 ควัน (Smoke: FU) คือ สิ่งแขวนลอยในอากาศของอนุภาคขนาดเล็กที่เกิดจากการเผาไหม้ทำให้เกิดสี เทาหรือสีฟ้าแก่บรรยากาศ ควรรายงานเมื่อทัศนวิสัยทั่วไป (Prevailing Visibility) เท่ากับ 5000 เมตรหรือน้อยกว่า อาจมีการรายงานเมื่อทัศนวิสัยต่ำกว่า 1000 เมตรก็ได้ หากไม่มีละอองน้ำแขวนลอยและความชื้นสัมพัทธ์ไม่เกิน 90%

4.8.8 พายุฟ้าคะนอง (Thunderstorm: TS) คือ การปลดปล่อยกระแสไฟฟ้าอย่างรวดเร็ว เกิดฟ้าแลบ (Lightning) เกิดเสียงฟ้าคะนอง(Thunder) พายุฝนฟ้าคะนองมีความสัมพันธ์กับเมฆ Cumulonimbus ในข่าว METAR เมื่อได้ยินเสียงฟ้าคะนองให้รายงาน TS หรือเมื่อได้ยินเสียงฟ้าคะนองบริเวณท่าอากาศยาน ให้รายงาน VCTS ประมาณ 8 - 16 กิโลเมตรจากจุดอ้างอิงสนามบิน การรายงานพายุฝนฟ้าคะนองที่มีฝนตกลงมาด้วยที่ท่าอากาศยาน “TS” ใช้เป็นคำอธิบาย และรวมกับคำย่อ ของฝน (RA) ที่เหมาะสม โดยใช้สัญลักษณ์ความรุนแรง (+, -) นำหน้า ขึ้นอยู่กับอัตราการตกของฝน

4.9 ส่วนที่ 9 เมฆ (Cloud) ใช้การตรวจด้วยสายตาของผู้ปฏิบัติงานด้านการตรวจอากาศการบินเป็นหลัก โดยอาจใช้ค่าที่ได้จากเครื่องมือตรวจอากาศอัตโนมัติเป็นตัวประกอบการตัดสินใจในการเข้ารหัส เช่น ความสูงของเมฆ โดยมองเป็นลำดับชั้นของแต่ละความสูงจากต่ำสุดไปหาสูงสุด และจำนวนของเมฆในแต่ละชั้นว่ามีปริมาณเท่าไรเมื่อเทียบกับทั้งท้องฟ้า ซึ่งต้องอนุมานแบ่งเมฆทั้งท้องฟ้าออกเป็น 8 ส่วน แล้วจึงนำจำนวนและความสูงมาเข้ารหัสตามวิธีเข้ารหัสกรุปเมฆและจะเข้ารหัสแบบพิเศษเมื่อมีกลุ่มเมฆ CB เกิดขึ้น

4.9.1 รูปแบบรหัสเมฆ คือ $N_5N_5N_5h_5h_5(CC)$ โดยที่ $N_5N_5N_5$ คือ จำนวนเมฆที่แสดงเป็น FEW SCT BKN OVC และ h_5h_5 คือ ความสูงของฐานเมฆเหนือท่าอากาศยานรายงานได้ทุก ๆ 100-10,000 ฟุต, (CC) คือ ชนิดของเมฆที่สามารถใช้รายงานได้เฉพาะ Cumulonimbus (CB) และหรือ Towering Cumulus (TCU) ในข่าว METAR

4.9.2 จำนวนเมฆ ($N_5N_5N_5$) อนุมานแบ่งเมฆทั้งท้องฟ้าออกเป็น 8 ส่วน คือ FEW เล็กน้อยหรือปกคลุมไม่เกิน 1-2 ส่วน SCT ปกคลุม 3-4 ส่วน BKN ปกคลุม 5-7 ส่วน และ OVC ปกคลุมเต็มท้องฟ้าคือ 8 ส่วน

4.10 ส่วนที่ 10 อุณหภูมิอากาศและอุณหภูมิจุดน้ำค้าง (Air temperature and Dewpoint temperature) ซึ่งอุณหภูมิทั้งสองที่รายงานในข่าว METAR จะเป็นอุณหภูมิอากาศและอุณหภูมิจุดน้ำค้าง ซึ่งอุณหภูมิ จุดน้ำค้างไม่สามารถวัดได้โดยตรงจากเครื่องมือวัดอุณหภูมิ อย่างไรก็ตามสามารถคำนวณได้จากอุณหภูมิกระเปาะแห้งและอุณหภูมิกระเปาะเปียก หรือความชื้นสัมพัทธ์ ซึ่งมีรูปแบบรหัส คือ $T'T' / T_dT_d$ โดยที่ $T'T'$ คือ อุณหภูมิอากาศที่รายงานเป็นจำนวนเต็มไม่มีจุดทศนิยม หน่วยวัดเป็นองศาเซลเซียส (Degree Celsius: °C) และถ้ามีจุดทศนิยมตั้งแต่ 0.5 ขึ้นไปให้ปัดขึ้น ต่ำกว่า 0.5 ให้ปัดทิ้ง, T_dT_d คือ อุณหภูมิจุดน้ำค้างที่รายงานเป็นจำนวนเต็มไม่มีจุดทศนิยม หน่วยวัดเป็นองศาเซลเซียส (Degree Celsius: °C) และถ้ามีจุดทศนิยมตั้งแต่ 0.5 ขึ้นไปให้ปัดขึ้น ต่ำกว่า 0.5 ให้ปัดทิ้ง

4.11 ส่วนที่ 11 ความกดอากาศ (Atmospheric pressure) โดยความกดอากาศใช้ในการบินเพื่อวัดความสูง (Altimeter setting) ของเครื่องบิน และด้วยเหตุนี้จึงเป็นสิ่งสำคัญที่จะต้องมีการวัดความกดอากาศที่ถูกต้องอยู่เสมอ โดยในการรายงานข่าว METAR จะรายงานความกดอากาศ QNH ซึ่งก็คือ ความกดอากาศที่ถูกต้องอยู่ที่ระดับน้ำทะเลปานกลางที่ขึ้นอยู่กับสภาวะบรรยากาศมาตรฐานสากล (International Standard Atmosphere: ISA) ตลอดจนความแตกต่างของความสูงและมีการปรับเศษทศนิยมทิ้งเพื่อรายงานเป็นจำนวนเต็มในหน่วย hectopascal (hPa) หรือ มิลลิบาร์ (Millibars) มีรูปแบบรหัส คือ $Q_P P_P P_P P_P$ โดยที่ Q คือ ตัวระบุกลุ่ม, $P_P P_P P_P P_P$ คือ ความกดอากาศที่ได้รับการหักแก้ที่ระดับน้ำทะเลปานกลาง (QNH)

4.12 ส่วนที่ 12 ข่าวสารเพิ่มเติม (Supplementary information) ได้แก่

4.12.1 สภาพอากาศที่ผ่านมา (Recent weather) คือ สภาพอากาศที่หยุดไปแล้ว หรือลดความรุนแรงลงจากรายงานครั้งล่าสุดหรือภายในชั่วโมงที่ผ่านมา แล้วแต่ว่าระยะเวลาใดจะสั้นกว่า ซึ่งอาจมีการรายงานกลุ่มสภาพอากาศที่ผ่านมาล่าสุดได้ถึงสามกลุ่ม แต่ละกลุ่มจะถูกสร้างขึ้นด้วยตัวย่อของสภาพอากาศปัจจุบันที่เหมาะสม แล้วนำหน้าด้วยตัวบ่งชี้ “RE” ตัวอย่างการรายงานสภาพอากาศที่ผ่านมา เช่น

- 1) RERA (Recent rain) สภาพอากาศที่ผ่านมามีฝนตกปานกลางหรือหนัก
- 2) RETSRA (Recent thunderstorm with rain) สภาพอากาศที่ผ่านมา มีพายุฝนฟ้าคะนองที่มีฝนตกเกิดขึ้น
- 3) RETS (Recent thunderstorm without precipitation) สภาพอากาศที่ผ่านมา มีพายุฟ้าคะนองที่ไม่มีฝนตกเกิดขึ้น

4.12.2 การตรวจและรายงานลมเฉือน (Wind Shear) โดยการรายงานลมเฉือนในประเทศไทยจะรายงานต่อเมื่อมีเครื่องมือตรวจวัดหรือระบบ LLWAS (Low Level Wind Shear Alert System) เท่านั้น ซึ่งจะถูกเพิ่มลงข่าว METAR ต่อท้ายกลุ่มความกดอากาศ มีรูปแบบรหัส คือ WS R_DR_DR หรือ WS ALL RWY

4.13 ส่วนที่ 13 การพยากรณ์แนวโน้มสภาวะอากาศบริเวณสนามบิน (Trend forecast) หมายถึง การคาดการณ์แนวโน้มการเปลี่ยนแปลงที่มีนัยสำคัญของสารประกอบอุตุนิยมวิทยาชนิดใดชนิดหนึ่ง หรือมากกว่าที่จะเกิดขึ้นบริเวณสนามบินภายใน 2 ชั่วโมงข้างหน้า ซึ่งสารประกอบอุตุนิยมวิทยานั้น ได้แก่ ลมผิวพื้น ทิศนวิสัย สภาพอากาศ และเมฆ โดยเป็นการพยากรณ์อากาศสำหรับการนำเครื่องบินลง ตามที่กำหนดโดยข้อตกลงการเดินทางอากาศในภูมิภาค แบ่งเป็น 2 กรณี ดังนี้

4.13.1 ไม่มีการเปลี่ยนแปลงสารประกอบอุตุนิยมวิทยาที่มีนัยสำคัญ หมายถึง การคาดการณ์ว่าไม่มีการเปลี่ยนแปลงสารประกอบอุตุนิยมวิทยาที่มีนัยสำคัญในเวลา 2 ชั่วโมงข้างหน้า จากรายงานข่าว METAR ฉบับนั้น ให้ต่อท้ายด้วย NOSIG (No Significant changes)

4.13.2 มีการเปลี่ยนแปลงสารประกอบอุตุนิยมวิทยาที่มีนัยสำคัญ หมายถึง การคาดการณ์ว่ามีการเปลี่ยนแปลงสารประกอบอุตุนิยมวิทยาที่มีนัยสำคัญในเวลา 2 ชั่วโมงข้างหน้า จากรายงานข่าว METAR ฉบับนั้น ให้ต่อท้ายด้วยรหัสรูปแบบเดียวกัน โดยใช้กลุ่มคำบอกการเปลี่ยนแปลงนำหน้าตามด้วยสารประกอบอุตุนิยมวิทยาที่คาดว่าจะมีการเปลี่ยนแปลงที่มีนัยสำคัญ มีรายละเอียดดังนี้

- 1) กลุ่มคำบอกการเปลี่ยนแปลง (change groups) ประกอบด้วยคำบ่งชี้การเปลี่ยนแปลงที่เรียกว่า “Change Indicator” (BECMG หรือTEMPO) และคาดการณ์เวลาเริ่มต้นและสิ้นสุดการเปลี่ยนแปลง

2) ลมผิวพื้น (Surface wind) ตามเกณฑ์การเปลี่ยนแปลงที่มีนัยสำคัญของลมผิวพื้น เช่น เมื่อคาดหมายว่าทิศทางลมเฉลี่ยเปลี่ยนแปลงไป 60° หรือมากกว่า โดยที่ความเร็วเฉลี่ยก่อน และ/หรือหลังการเปลี่ยนแปลงเป็น 10 นอตหรือมากกว่า เป็นต้น

3) ทิศนวิสัย (visibility) ตามเกณฑ์การเปลี่ยนแปลงที่มีนัยสำคัญของทัศนวิสัย เช่น เมื่อคาดหมายว่าทัศนวิสัยเลวลงและผ่านค่าหนึ่งค่าหรือมากกว่า ของค่า 150 350 600 800 1500 หรือ 3000 เมตร เป็นต้น

4) สภาพอากาศ (weather) ตามเกณฑ์การเปลี่ยนแปลงที่มีนัยสำคัญของสภาพอากาศ เช่น เมื่อคาดหมายสภาพอากาศชนิดเดียวหรือหลายชนิดหรือรวมกันของสภาพอากาศมีปรากฏการณ์พายุฝนฟ้าคะนองเกิดขึ้นหรือสิ้นสุด หรือ เปลี่ยนแปลงความรุนแรง เป็นต้น

5) เมฆในท้องฟ้า (cloud) หรือ ทัศนวิสัยในแนวตั้ง (vertical visibility) ตามเกณฑ์การเปลี่ยนแปลงที่มีนัยสำคัญของเมฆในท้องฟ้าหรือทัศนวิสัยในแนวตั้ง เช่น เมื่อคาดหมายว่าจำนวนเมฆเปลี่ยนแปลง ถ้าจำนวนของเมฆที่มีฐานอยู่ต่ำกว่า หรือคาดว่าลระดับต่ำกว่าหรือยกตัวขึ้นสูงกว่า 1500 ฟุต คาดหมายว่ามีการเปลี่ยนแปลงจำนวนเมฆจาก FEW-SCT เป็น BKN- OVC หรือจาก BKN-OVC เป็น FEW-SCT

ตัวอย่างข่าวอากาศการบินอยู่ในรูปแบบข้อความรหัส ดังนี้

ข่าว 1: METAR VTCC 011900Z 35010G25KT 1200 R36/P2000N +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=

และตัวอย่างการแปลความข่าวอากาศการบินของข่าว ดังตารางที่ 2.1

ตารางที่ 2.1 ตัวอย่างการแปลความข่าวอากาศการบินของข่าว

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 และมีลมกระโชก ความเร็ว 25 นอต
6	1200	ทัศนวิสัย 1200 เมตร
7	R36/1200	พิสัยการมองเห็นสูงสุดบนทางวิ่ง 36 เท่ากับ 1,200 เมตร
8	+TSRA	เกิดพายุฝนฟ้าคะนองขนาดหนัก
9	FEW025CB BKN035	มีเมฆ Cumulonimbus ฐาน 2500 ฟุต จำนวน 1-2 ส่วนจากทั้งหมด 8 ส่วน และมีเมฆฐาน 3500 ฟุต จำนวน 5-7 ส่วน จากทั้งหมด 8 ส่วน
10	25/22	อุณหภูมิอากาศ เท่ากับ 25 °C และอุณหภูมิจุดน้ำค้าง เท่ากับ 22 °C
11	Q1013	ความกดอากาศหักแก้สู่ระดับน้ำทะเลปานกลาง เท่ากับ 1013 มิลลิบาร์
12	RERA (ข่าวที่ 2)	ความรุนแรงของฝนที่ตกได้ลดกำลังลงจากข่าวก่อนหน้า
13	BECMG 3000 TSRA	คาดว่าใน 2 ชั่วโมงข้างหน้าจะเกิดพายุฝนฟ้าคะนองปานกลาง ทัศนวิสัยเท่ากับ 3,000 เมตร

จากตารางที่ 2.1 เนื้อหาข่าวอากาศการบินจะประกอบไปด้วย 13 ส่วน คือ (1) ประเภทของรายงาน (2) รหัสท่าอากาศยาน (3) เวลาตรวจอากาศการบิน (4) การตรวจอากาศการบินด้วยระบบอัตโนมัติ (5) ลมผิวพื้น (6) ทัศนวิสัย (7) พิสัยการมองเห็นบนทางวิ่ง (8) สภาพอากาศปัจจุบัน (9) เมฆ (10) อุณหภูมิอากาศและอุณหภูมิจุดน้ำค้าง (11) ความกดอากาศ (12) ข่าวสารเพิ่มเติม และ (13) การพยากรณ์แนวโน้มสภาพอากาศบริเวณสนามบิน ซึ่งส่วนที่ (1)-(12) ได้จากการตรวจวัดค่าสารประกอบทางอุตุนิยมวิทยาจริงด้วยคนและเครื่องมือตรวจอากาศการบิน และส่วนที่ (13) เป็นส่วนต่อท้ายข่าวที่เป็นการพยากรณ์แนวโน้มสภาพอากาศบริเวณสนามบินที่จะเกิดขึ้นภายใน 2 ชั่วโมง

ข้างหน้า หรือที่เรียกว่า Trend Forecasts โดยข่าวแต่ละชั่วโมงจะถูกรายงานนาฬิกาที่ 0 และสามารถส่งข่าวล่าช้าได้ไม่เกินนาฬิกาที่ 3

5. เทคนิคการเรียนรู้ของเครื่อง

การเรียนรู้ของเครื่อง (Machine learning) เป็นส่วนหนึ่งของปัญญาประดิษฐ์ (Artificial intelligence) โดยอัลกอริทึมสำหรับการเรียนรู้ของเครื่องจะทำการเรียนรู้โดยอัตโนมัติจากการเรียนรู้ในอดีตและทำการปรับปรุงการทำงานได้เองโดยไม่ต้องมีมนุษย์คอยกำกับหรือเขียนโปรแกรมกำกับสั่งการเครื่อง ซึ่งมีหลักการทำงานคล้ายกับมนุษย์ที่จำเป็นต้องเรียนรู้ประสบการณ์เพื่อให้สามารถจำแนกได้อย่างถูกต้อง โดยการเรียนรู้ของเครื่องจะเริ่มต้นด้วยการป้อนชุดข้อมูลพื้นฐานและคำสั่งต่างๆ เพื่อให้คอมพิวเตอร์เกิดกระบวนการฝึกฝน (Training) และสามารถจำแนกแยกแยะสิ่งต่างๆได้ และเพื่อการปรับปรุงประสิทธิภาพของการจำแนกให้ดีขึ้นจะต้องมีการป้อนชุดข้อมูลใหม่ๆ เพื่อให้เกิดการพัฒนาของระบบที่ประสิทธิภาพสูงขึ้น โดยการเรียนรู้ของเครื่องแบ่งเป็น 3 ประเภท คือ (1) การเรียนรู้แบบมีผู้สอน (Supervised Learning) (2) การเรียนรู้โดยไม่มีผู้สอน (Unsupervised Learning) และ (3) การเรียนรู้แบบเสริมกำลัง (Reinforcement Learning)

สำหรับในงานวิจัยนี้เป็นการสร้างแบบจำลองแบบการเรียนรู้แบบมีผู้สอน ประเภทแบบจำลองการจำแนก (Classification Model) จากอัลกอริทึมมาตรฐาน (Standard Classification) และอัลกอริทึมมาตรฐานร่วมกับอัลกอริทึมการเรียนรู้แบบกลุ่ม (Ensemble Learning Algorithms) การแบ่งคลาส (Class) เพื่อจำแนกชนิด Binary ที่มีเพียงสองสถานะเท่านั้น เพื่อป้องกันปรากฏการณ์การเกิดพายุฝนฟ้าคะนอง นั่นคือ ไม่เกิดพายุฝนฟ้าคะนอง (NO TSRA) และเกิดพายุฝนฟ้าคะนอง (TSRA) โดยมีขั้นตอนที่สำคัญของกระบวนการสร้างแบบจำลอง ดังนี้

5.1 การเตรียมข้อมูลก่อนสร้างแบบจำลอง (Data Preprocessing)

การเตรียมข้อมูลเป็นกระบวนการเริ่มต้นที่มีความสำคัญในการสร้างแบบจำลอง เนื่องจากข้อมูลที่นำเข้าสู่ผลโดยตรงต่อประสิทธิภาพของแบบจำลอง ซึ่งก่อนที่จะนำเข้าสู่ข้อมูลต้องมีการทำความเข้าใจและตรวจสอบคุณภาพของข้อมูลก่อน โดยทำให้ข้อมูลอยู่ในรูปแบบที่ถูกต้องหรือปรับเปลี่ยนให้อยู่ในรูปแบบที่เหมาะสม เพื่อให้แบบจำลองเกิดการเรียนรู้ด้วยเงื่อนไขที่ถูกต้อง โดยทั่วไปการเตรียมข้อมูลแบ่งการทำงานออกเป็น 4 ขั้นตอน คือ

5.1.1 การทำความสะอาดข้อมูล (Data Cleaning) เป็นขั้นตอนแรกของการเตรียมข้อมูลที่ต้องตรวจสอบว่ามีข้อผิดพลาดใดเกิดขึ้นกับข้อมูลบ้าง เช่น ความซ้ำซ้อนของข้อมูล (Duplicate data) ข้อมูลไม่ถูกต้อง (Incorrectly data) ข้อมูลเก่าล้าสมัย (Expired data) เกิดการสูญหายของข้อมูลบางส่วน (Missing Value) มีค่าข้อมูลที่มีความผิดปกติหรือแตกต่างไปจากข้อมูลในกลุ่ม (Outliers) เป็นต้น เมื่อพบข้อผิดพลาดของข้อมูลควรมีการปรับปรุงแก้ไขข้อผิดพลาดเหล่านั้นก่อน ซึ่งสามารถทำได้หลายวิธี เช่น การลบข้อมูลออกไปจากชุดข้อมูล (Removal) การปรับปรุงข้อมูลให้เป็นค่าที่ถูกต้อง (Correction) และการแทนค่า (Replacement) เป็นต้น

5.1.2 การผสานข้อมูล (Data Integration) เป็นขั้นตอนรวบรวมข้อมูลที่มีหลายแหล่งให้เป็นข้อมูลชุดเดียวกันเพื่อไม่ทำให้เกิดความซ้ำซ้อนของข้อมูล ซึ่งนำไปสู่ปัญหาความไม่สอดคล้องกันของข้อมูล (Data inconsistencies) อีกทั้งทำให้ข้อมูลมีคุณภาพและเพิ่มความเร็วในการประมวลผล

5.1.3 การแปลงข้อมูล (Data transformation) เป็นขั้นตอนการวิเคราะห์ข้อมูลที่อาจมาจากหลายแหล่งที่มีความแตกต่างกัน โดยสามารถทำได้หลายวิธี ดังนี้

1) Discretization เป็นกระบวนการในการแปลงข้อมูลที่เป็นตัวเลขให้อยู่ในช่วงข้อมูล เนื่องจากบางอัลกอริทึมการเรียนรู้ของเครื่องไม่สามารถรองรับข้อมูลที่เป็นตัวเลขได้ อีกทั้งการแปลงข้อมูลให้อยู่ในรูปของช่วงของข้อมูลทำให้แบบจำลองเกิดการเรียนรู้ที่ไม่ซับซ้อน และอาจทำให้ประสิทธิภาพของโมเดลสูงขึ้นได้.

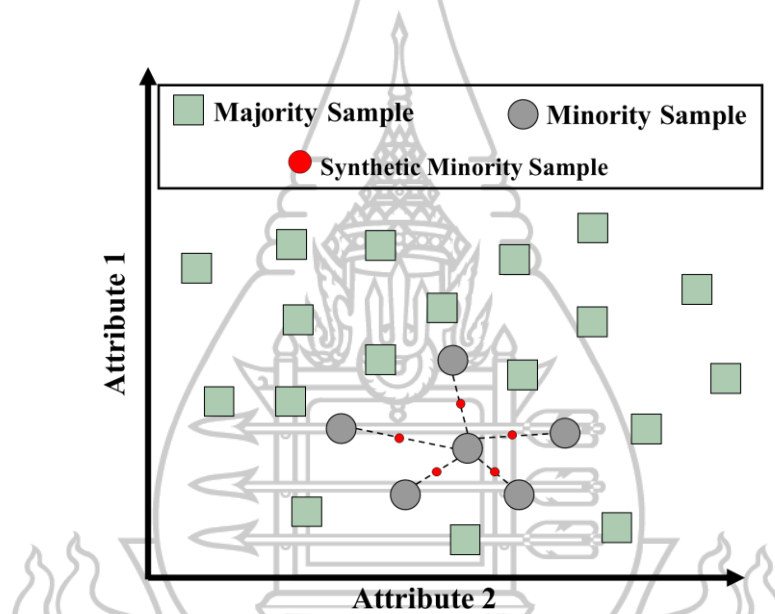
2) Aggregation เป็นกระบวนการแปลงข้อมูลให้อยู่ในรูปแบบสรุป ตัวอย่างของตัวดำเนินการ ได้แก่ Sum Min Max และ Count เป็นต้น โดยวิธีการนี้ทำให้แบบจำลองสามารถค้นหารูปแบบหรือเงื่อนไขที่ถูกต้องของข้อมูลง่ายขึ้น ส่งผลให้ประสิทธิภาพของแบบจำลองสูงขึ้น

3) Normalization เป็นกระบวนการแปลงข้อมูลให้อยู่ในช่วงที่กำหนด ซึ่งช่วยลดความซ้ำซ้อนของข้อมูล จึงทำให้เนื้อที่ในการจัดเก็บข้อมูลลดลงด้วย

5.2 การปรับข้อมูลไม่สมดุล (Imbalanced Data)

ข้อมูลไม่สมดุล คือ การที่ชุดข้อมูลมีจำนวนในคลาสที่แตกต่างกันมาก โดยปกติแล้วข้อมูลที่ดีนั้นควรมีความสมดุลของคลาสที่ใช้เพื่อการจำแนก ในกรณีที่ข้อมูลประกอบด้วยข้อมูลกลุ่มส่วนมาก (Majority) และข้อมูลกลุ่มส่วนน้อย (Minority) ถ้านำข้อมูลทั้งหมดเข้าสู่กระบวนการสร้างแบบจำลองโดยที่ไม่มีขั้นตอนการปรับสมดุลของข้อมูลก่อน จะส่งผลให้แบบจำลองทำงานได้ผิดพลาด ในงานวิจัยนี้ปรับข้อมูลไม่สมดุลด้วยวิธีการสังเคราะห์ข้อมูล (Data Resampling) เทคนิค SMOTE

5.2.1 เทคนิคการสุ่มตัวอย่างเกินของข้อมูลกลุ่มน้อยแบบสังเคราะห์ (Synthetic Minority Over-sampling Technique: SMOTE) เป็นเทคนิคที่ใช้ในการแก้ปัญหาข้อมูลไม่สมดุล โดยเพิ่มจำนวนข้อมูลกลุ่มน้อยจากการสังเคราะห์ด้วยข้อมูลจริงในการสร้างข้อมูลขึ้นมา ซึ่งนำตัวอย่างข้อมูลจากกลุ่มส่วนน้อยมาพิจารณาทีละตัวจนครบทุกตัวโดยสุ่มเลือกข้อมูลขึ้นมา 1 ค่า และหาค่าใกล้เคียงจำนวน K ตัว (Number of Nearest Neighbors) จากนั้นคำนวณหาระยะทาง (Euclidean distance) ระหว่างข้อมูลที่สุ่มเลือกกับข้อมูลใกล้เคียง K ตัว เพื่อหาค่าระยะทางที่น้อยที่สุดระหว่างค่าที่สุ่มกับค่าข้อมูลใกล้เคียงและทำการสร้างข้อมูลสังเคราะห์ขึ้นมา ทำให้การกระจายของกลุ่มข้อมูลมีความสมดุลมากขึ้น ดังภาพที่ 2.11



ภาพที่ 2.11 Synthetic Minority Over-sampling Technique (SMOTE)

5.2.2 การปรับความสมดุลของข้อมูลด้วยวิธีอื่นๆ ยังมีอีกหลายวิธีที่งานวิจัยอื่นๆ ใช้เพื่อปรับสมดุลของข้อมูลก่อนนำไปสร้างแบบจำลอง เช่น

1) วิธีสุ่มลด (Random Under-sampling) เป็นการปรับลดจำนวนข้อมูลกลุ่มส่วนมากด้วยการสุ่มเลือกจากข้อมูลเดิมให้มีจำนวนใกล้เคียงกับกลุ่มส่วนน้อย วิธีการนี้เหมาะสำหรับงานวิจัยที่มีข้อมูลนำเข้าจำนวนมากและมีจำนวนข้อมูลกลุ่มส่วนน้อยมากพอที่เมื่อทำการสุ่มลดแล้วจะยังสามารถสร้างแบบจำลองการจำแนกให้มีประสิทธิภาพ เนื่องจากวิธีการนี้ทำให้เสียข้อมูลบางส่วนและมีโอกาสเกิด Underfitting ได้

2) วิธีสุ่มเกิน (Random Over-Sampling) เป็นการปรับเพิ่มจำนวนข้อมูลกลุ่มส่วนน้อยด้วยวิธีการสุ่มเลือกจากข้อมูลเดิมให้มีจำนวนใกล้เคียงกับกลุ่มส่วนมาก โดยวิธีการนี้มีโอกาสเกิดการ Overfitting ขึ้นได้

3) วิธี Borderline-SMOTE เป็นวิธีการที่ปรับปรุงการทำงานของ SMOTE ให้ดีขึ้น โดยการสุ่มตัวอย่างสังเคราะห์ของข้อมูลกลุ่มน้อยด้วยเส้นแนวเขต ซึ่งจะทำการเลือกตัวอย่างที่อยู่ใกล้แนวเขตของข้อมูลกลุ่มน้อยที่จำแนกผิด และสร้างเฉพาะตัวอย่างสังเคราะห์ที่ยากในการจำแนก แต่ยังมีข้อจำกัดเนื่องจากอาจสร้างตัวอย่างบางส่วนที่ทับซ้อนกับตัวอย่างเชิงลบและวิธีการสุ่มตัวอย่างมากเกินไปจะทำให้ซ้ำทุกตัวอย่างที่เป็นบวกโดยไม่ต้องโดยคำนึงถึงปัจจัยอื่นๆ

4) วิธี ADASYN (Adaptive Synthetic Sampling Approach) เป็นวิธีการที่ปรับปรุงการทำงานของ SMOTE ให้ดีขึ้น ซึ่งในขั้นตอนการสังเคราะห์ข้อมูลนั้นไม่จำเป็นต้องพิจารณาข้อมูลทุกตัวของข้อมูลกลุ่มส่วนน้อย โดย ADASYN จะใช้วิธีการแจกแจงแบบถ่วงน้ำหนัก (Weight distribution) ของข้อมูลในกลุ่มส่วนน้อย โดยการสังเคราะห์ข้อมูลซึ่งขึ้นอยู่กับความสำคัญของข้อมูลนั้นๆ ถ้าข้อมูลโดยยากต่อการจำแนกก็จะให้ค่าของน้ำหนักข้อมูลนั้นมากและสังเคราะห์ข้อมูลชุดข้อมูลขึ้นมาในบริเวณนั้นๆ

5.3 การคัดเลือกคุณลักษณะ (Feature Selection)

การคัดเลือกคุณลักษณะเป็นขั้นตอนของการเลือกข้อมูลที่มีความสอดคล้องหรือสัมพันธ์กับคลาสที่ทำการศึกษา โดยข้อมูลที่นำเข้าไปเพื่อสร้างแบบจำลองนั้นจะประกอบไปด้วยคุณลักษณะหรือแอตทริบิวต์ (Attribute) จำนวนมาก ซึ่งอาจจะมีความเกี่ยวข้องหรือไม่เกี่ยวข้องกันก็ได้ ดังนั้นขั้นตอนนี้จึงมุ่งหมายเพื่อคัดเลือกแอตทริบิวต์ที่เกี่ยวข้องกับคลาสทำการศึกษา ซึ่งมีความสำคัญต่อการทำนายเพื่อช่วยลดมิติของข้อมูลและอาจช่วยให้การเรียนรู้ดำเนินการได้รวดเร็วและมีประสิทธิภาพมากขึ้น รวมถึงใช้หน่วยความจำและเวลาประมวลผลลดลง โดยมีวิธีการคัดเลือกที่เป็นที่นิยมใช้เพื่อสร้างแบบจำลองทั้งหมด 3 วิธี ดังนี้

5.3.1 Filter Method เป็นการคัดเลือกแอตทริบิวต์ด้วยการคำนวณค่าน้ำหนักที่เป็นค่าความสัมพันธ์ระหว่างแต่ละแอตทริบิวต์และคลาสที่ทำการศึกษา คัดเลือกโดยเรียงลำดับตามค่าน้ำหนักที่คำนวณได้ จากนั้นเลือกแอตทริบิวต์ที่มีค่าน้ำหนักมากกว่าเพื่อนำไปใช้สร้างแบบจำลองต่อไป เทคนิคในการคำนวณค่าน้ำหนักของแอตทริบิวต์ต่างๆ มีหลายวิธี เช่น Information Gain Chi-Square ANOVA และ ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) เป็นต้น โดยในงานวิจัยนี้ใช้ 2 เทคนิค คือ

1) Information Gain วิธีการนี้สามารถใช้ได้กับแอตทริบิวต์ที่เป็นนามินอล (nominal) เท่านั้น ซึ่งจะหาค่าน้ำหนักเพื่อใช้ในการแบ่งข้อมูลด้วยการคำนวณค่า Gain สำหรับแต่ละ

แอตทริบิวต์ ถ้าแอตทริบิวต์ใดมีค่าสูงสุดจะถูกเลือกให้มีความสำคัญต่อการจำแนก สามารถคำนวณได้ดังสมการที่ 2.1

$$\text{infoGain}(S, A) = \text{Entropy}(S) - \left| \frac{S_v}{S} \right| E(S_v) \quad (2.1)$$

โดยที่	A	คือ แอตทริบิวต์ A
	$ S_v $	คือ สมาชิกของแอตทริบิวต์ A ที่มีค่า v
	$ S $	คือ จำนวนสมาชิกของกลุ่มตัวอย่าง

2) Chi-Square วิธีการนี้คำนวณนี้ใช้คำนวณความถี่ที่เกิดขึ้นระหว่างแอตทริบิวต์ต่างๆเปรียบเทียบกับคลาสที่ทำการศึกษา ซึ่งแอตทริบิวต์ที่มีความเกี่ยวข้องมากจะมีค่าน้ำหนักมาก และแอตทริบิวต์ที่มีความเกี่ยวข้องน้อยจะมีค่าน้ำหนักน้อย ทำให้สามารถพิจารณากำจัดแอตทริบิวต์ที่ไม่เกี่ยวข้องออกได้ สามารถคำนวณค่า Chi-Square ทางสถิติได้ดังสมการที่ 2.2

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \quad (2.2)$$

โดยที่	χ^2	คือ ค่าสถิติ Chi-Square
	O_i	คือ ความถี่ที่ได้จากการสังเกต (Observed Frequency)
	E_i	คือ ความถี่ที่คาดหวัง (Expected Frequency)

5.3.2 Wrapper Method เป็นการคัดเลือกแอตทริบิวต์ด้วยการสร้างแบบจำลองการจำแนกขึ้นมาจากแอตทริบิวต์ที่กำหนดไว้ จากนั้นทดสอบประสิทธิภาพการทำงานของแบบจำลองและคัดเลือกชุดของแอตทริบิวต์ที่ทำให้แบบจำลองมีประสิทธิภาพมากที่สุดมาใช้เพื่อสร้างแบบจำลองต่อไป แบ่งออกเป็น 2 วิธี คือ

1) Forward Selection เป็นการสร้างแบบจำลองโดยเพิ่มทีละ แอตทริบิวต์ ถ้าประสิทธิภาพดีก็จะเก็บไว้และเลือกแอตทริบิวต์อื่นๆมาเพิ่มต่อไปจนประสิทธิภาพของแบบจำลองไม่ดีขึ้นจึงหยุดกระบวนการ

2) Backward Elimination เป็นการสร้างแบบจำลองที่เริ่มจากการใช้แอตทริบิวต์ทั้งหมดที่มีและตัดแอตทริบิวต์ทิ้งทีละแอตทริบิวต์ ถ้าประสิทธิภาพของแบบจำลองดีขึ้นก็จะตัดแอตทริบิวต์ทิ้งเรื่อยๆจนกว่าประสิทธิภาพจะลดลงจึงหยุดกระบวนการ

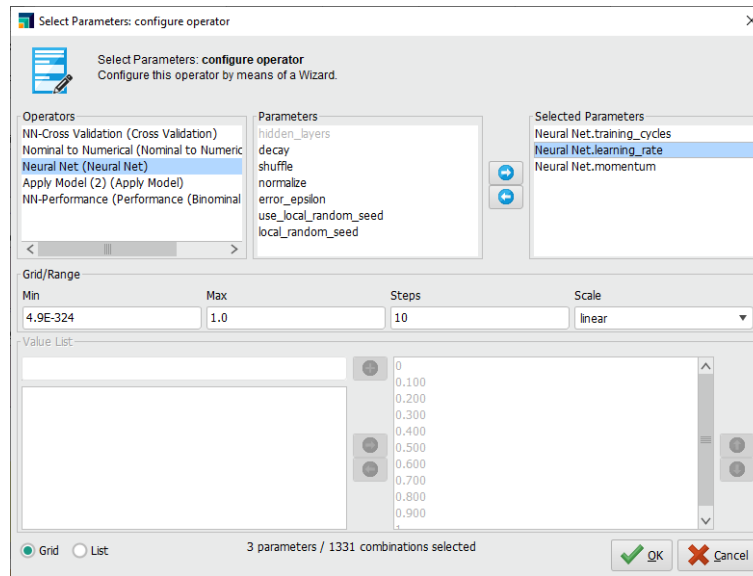
5.3.3 Embedded Method เป็นวิธีการคัดเลือกคุณลักษณะที่รวมข้อดีของ Filter และ Wrapper Method เข้าด้วยกัน ซึ่งมีวิธีการเลือกข้อมูลโดยจัดลำดับความสำคัญของแอตทริบิวต์ในระหว่างการฝึกฝนแบบจำลอง จากนั้นจึงเลือกแอตทริบิวต์ที่มีลำดับความสำคัญสูงสุด k ลำดับเข้าสู่กระบวนการสร้างแบบจำลอง

5.4 การปรับค่าพารามิเตอร์ที่เหมาะสมของอัลกอริทึม (Hyperparameter Tuning)

Hyperparameters คือ พารามิเตอร์ต่างๆ ที่ผู้ใช้สามารถกำหนดเองได้ก่อนที่แบบจำลองจะทำการเรียนรู้โดยการกำหนดโครงสร้างของแบบจำลองตั้งแต่ต้น ซึ่งค่าพารามิเตอร์ที่เหมาะสมของแต่ละอัลกอริทึมเป็นกระบวนการปรับค่าที่ไม่ได้ถูกเรียนรู้จากข้อมูล ตัวอย่างพารามิเตอร์ที่มีการปรับค่า เช่น ค่า Learning Rate ที่ใช้ในการควบคุมแต่ละขั้นของการเรียนรู้ อัลกอริทึม Neural Network หรือค่า n_estimators ซึ่งกำหนดจำนวนต้นไม้สำหรับการสร้างแบบจำลองด้วยอัลกอริทึม Random Forest เป็นต้น โดยการปรับค่าพารามิเตอร์ที่เหมาะสมมีผลต่อประสิทธิภาพของแบบจำลองในการเรียนรู้ของเครื่อง สามารถแบ่งการค้นหา Hyperparameters เป็น 3 วิธี ดังนี้

5.4.1 Manual Search วิธีนี้ทำได้โดยการเลือกค่า Hyperparameter ของแบบจำลองจากประสบการณ์และความคิดเห็นส่วนบุคคล โดยจะสร้างแบบจำลองขึ้นมาจากค่าที่เลือกและทำการวัดค่าความแม่นยำไปเรื่อยๆ จนกว่าจะได้ค่าความแม่นยำที่พึงพอใจ

5.4.2 Grid Search หรือ การค้นหาแบบกริด วิธีนี้ใช้ในการหาค่า Hyperparameter ด้วยการลองใช้พารามิเตอร์ที่กำหนดไว้ล่วงหน้าทุกชุด และประเมินประสิทธิภาพหรือความแม่นยำของแบบจำลองแต่ละชุด ซึ่งเป็นการลองสร้างโมเดลจากค่าของ Hyperparameter ทุกชุด รูปแบบของการทำงานจะคล้ายกริด โดยค่าทั้งหมดจะอยู่ในรูปของเมทริกซ์ (Matrix) พารามิเตอร์แต่ละชุดจะถูกนำมาพิจารณาและสังเกตความถูกต้อง เมื่อชุดของ Hyperparameter ทั้งหมดได้รับการประเมินแล้ว โมเดลที่มีชุดพารามิเตอร์ที่ให้ความแม่นยำสูงสุดจะถือว่าดีที่สุด โดยในวิจัยนี้ใช้โอเพอร์เรเตอร์ Optimize Parameter (Grid) ของโปรแกรม RapidMiner มาช่วยในการหาค่ากำหนดสำหรับพารามิเตอร์ต่างๆ ที่เหมาะสม โดยโอเพอร์เรเตอร์นี้ผู้ใช้สามารถกำหนดค่าของชุดพารามิเตอร์ของแต่ละอัลกอริทึมที่ทำการศึกษา โดยจำเป็นต้องกำหนดช่วงที่ต้องการให้โอเพอร์เรเตอร์ทำการค้นหาแบบกริด นั่นคือ ค้นหาทุกชุดพารามิเตอร์และคัดเลือกเพียงชุดเดียวที่ทำให้แบบจำลองมีประสิทธิภาพสูงสุด โดยในขั้นตอนนี้ใช้ระยะเวลาหากกำหนดให้มีการค้นหาในหลายพารามิเตอร์ ตัวอย่างการตั้งค่าค้นหาของโอเพอร์เรเตอร์ Optimize Parameter (Grid) แสดงดังภาพที่ 2.12



ภาพที่ 2.12 ตัวอย่างการตั้งค่าค้นหาโดย Optimize Parameter (Grid)

5.4.3 Random Search วิธีนี้มีความคล้ายคลึงกับการทำ Grid Search ที่จะลองใช้พารามิเตอร์ที่กำหนดไว้ล่วงหน้าในกริดทุกชุด แต่ Random Search จะทำการสุ่มเลือกค่าพารามิเตอร์จากกริดที่สร้างขึ้น ดังนั้นบางครั้งอาจจะได้แบบจำลองที่มีประสิทธิภาพที่สุดเหมือนกับ Grid Search แต่วิธีนี้มีประสิทธิภาพสูงในการใช้งานจริงเนื่องจากใช้เวลาในการสร้างแบบจำลองน้อย โดยในในวิจัยนี้ใช้วิธีการค้นหาค่า Hidden layer sizes ของอัลกอริทึม Neural Networks ที่เหมาะสม โดยการกำหนดค่าเป็น 1-8 ตามจำนวนแอตทริบิวต์สูงสุดที่นำเข้าไปสร้างแบบจำลอง ซึ่งจะสร้างแบบจำลองของแต่ละชุดข้อมูลและตั้งค่า Hidden layer sizes และเปรียบเทียบประสิทธิภาพเพื่อคัดเลือกค่าที่เหมาะสมที่สุดของแต่ละชุดข้อมูล

5.5 แบบจำลองการจำแนก (Classification Model)

Classification Model เป็นหนึ่งในวิธีการวิเคราะห์ข้อมูลที่ตั้งอยู่ในประเภทการเรียนรู้แบบมีผู้สอน ซึ่งจำเป็นต้องมีคำตอบ (Label) ให้แบบจำลองเรียนรู้จากข้อมูลนำเข้าที่ถูกป้อนเข้ามาเพื่อให้สามารถหาคำตอบตามกำหนดไว้ ซึ่งผลลัพธ์ของการจำแนกนี้จะอยู่ในรูปแบบของคำตอบที่เป็นตัวเลือก หรือกลุ่มข้อมูล เช่น Yes หรือ No เป็นหรือไม่เป็น A B หรือ C เป็นต้น ซึ่งสามารถวัดประสิทธิภาพด้วยการใช้ Confusion Matrix ถ้าเปรียบเทียบกับแบบจำลองการวิเคราะห์การถดถอยโลจิสติก (Regression Model) ซึ่งเป็นหนึ่งในแบบจำลองประเภทการเรียนรู้แบบมีผู้สอนเช่นเดียวกัน แต่มีความแตกต่างกันตรงที่ Regression Model มีคำตอบของแบบจำลองเป็นค่าตัวเลขที่จะเกิดขึ้นในอนาคตจากการนำเข้าไปสร้างแบบจำลอง ไม่ใช่เป็นคำตอบที่เป็นตัวเลือกเหมือน Classification Model โดยสามารถแบ่งออกได้เป็น 3 ประเภทหลัก ได้แก่

5.5.1 Binary Classification เป็นรูปแบบการจำแนกที่มีกระบวนการวิเคราะห์ข้อมูลที่น่าเข้าเพื่อให้ได้ผลลัพธ์ออกมาเป็นคำตอบเพียง 2 กลุ่ม เช่น ฝนตกหรือไม่ตก ไข้หรือไม่ ไข้ ไข้หรือไม่ ไข้ เป็นต้น

5.5.2 Multi-Class Classification เป็นการจำแนกประเภทแบบหลายคลาส โดยมีความคล้ายคลึงกับ Binary Classification แต่ได้ผลลัพธ์ออกมาเป็นคำตอบมากกว่า 2 กลุ่มขึ้นไป เช่น การจำแนกรูปภาพว่าเป็น คน สัตว์ สิ่งของ เป็นต้น

5.5.3 Multi-Label Classification เป็นการจำแนกประเภทแบบหลายคำตอบ โดยมีความคล้ายคลึงกับ Multi-Class Classification ตรงที่ข้อมูลชุดหนึ่งอาจจะมีข้อมูลที่มีคุณสมบัติเหมือนกันแต่สามารถให้ผลลัพธ์ที่มีความแตกต่างกัน ซึ่งการเรียนรู้ของเครื่องเพื่อจำแนกแบบหลายคำตอบจะมีความยากและซับซ้อนกว่าแบบหลายคลาส

5.6 อัลกอริทึมที่ใช้สำหรับการสร้างแบบจำลองการจำแนก

ในงานวิจัยนี้แบ่งอัลกอริทึมของการสร้างแบบจำลองเป็น 2 ขั้นตอนคือ อัลกอริทึมมาตรฐานและอัลกอริทึมการเรียนรู้แบบกลุ่ม โดยมีอัลกอริทึมที่ใช้ดังนี้

5.6.1 อัลกอริทึมมาตรฐาน (Standard Algorithms)

1) Naive Bayes เป็นอัลกอริทึมใน Classification Model ใช้ในการแบ่งกลุ่มหรือหาเหตุการณ์ที่จะเกิดขึ้นโดยใช้หลักความน่าจะเป็นของข้อมูลบนพื้นฐานของทฤษฎีเบย์ (Bayes Theorem) และสมมติฐานที่กำหนดในการเกิดของเหตุการณ์ต่างๆที่ใช้ในการจำแนกนั้นเป็นอิสระต่อกัน หลักการคำนวณความน่าจะเป็นแบบมีเงื่อนไขที่เรียกว่า Conditional Probability (Dietrich et al., 2015) สามารถคำนวณได้จากสมการที่ 2.3

$$P(h|D) = \frac{P(D|h) \times P(h)}{P(D)} \quad (2.3)$$

โดยที่

$P(h)$ คือ ความน่าจะเป็นของสมมติฐานที่คลาสเป็น h

$P(D)$ คือ ความน่าจะเป็นของสมมติฐานที่ข้อมูลเป็น D

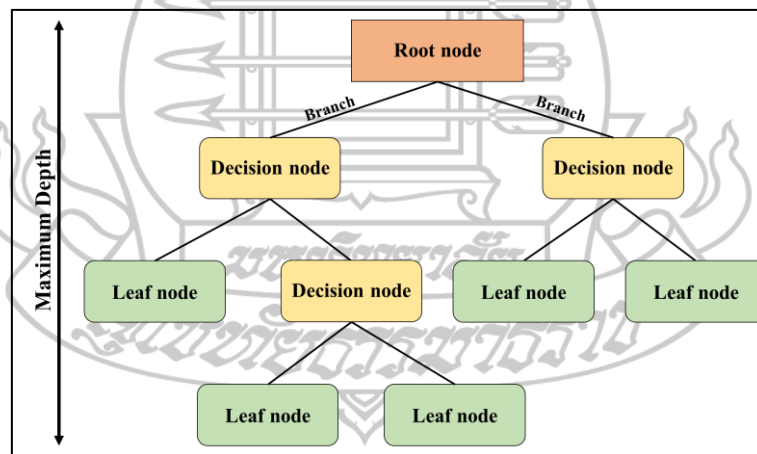
$P(h|D)$ คือ ค่าความน่าจะเป็นของสมมติฐานที่ข้อมูลเป็น D จะมีคลาสเป็น h

$P(D|h)$ คือ ค่าความน่าจะเป็นของสมมติฐานคลาส h จะมีข้อมูลเป็น D

ในปัญหาที่มีตัวแปรต้นหรือข้อมูลที่ต้องพิจารณามากกว่า 1 ค่าความน่าจะเป็นของเหตุการณ์สามารถคำนวณได้จากผลคูณของความน่าจะเป็นแต่ละข้อมูลที่มีคลาส h ดังสมการที่ 2.4

$$P(h|D) = P(D_1|h) \times P(D_2|h) \times \dots \times P(D_n|h) \times P(h) \quad (2.4)$$

2) Decision Tree หรือ ต้นไม้ตัดสินใจ ในการเรียนรู้ของเครื่องนั้น ต้นไม้ตัดสินใจเป็นแบบจำลองทางคณิตศาสตร์ที่ใช้ทำนายประเภทของวัตถุโดยพิจารณาจากลักษณะของวัตถุ ซึ่งมีลักษณะการตัดสินใจเป็นแผนผังต้นไม้ ประกอบด้วย (1) Node หรือที่เรียกว่า โหนดของต้นไม้ ใช้แทนตัวแปรต่างๆในการตัดสินใจ (2) Branch หรือส่วนที่เป็นกิ่ง ใช้แทนค่าที่เป็นได้ของตัวแปร (3) Leaf node หรือที่เรียกว่าโหนดใบ ใช้แทนคำตอบของการตัดสินใจหรือการจำแนก ซึ่งต้นไม้ตัดสินใจเป็นอัลกอริทึมที่ใช้ได้กับข้อมูลที่มีค่าต่อเนื่อง (Continuous values) และไม่ต่อเนื่อง (Discrete values) โดยที่หากใช้กับข้อมูลไม่ต่อเนื่องจะเรียกว่า ต้นไม้ตัดสินใจแบบจำแนก (Classification trees) ขณะที่หากใช้กับข้อมูลต่อเนื่องจะเรียกว่า ต้นไม้ตัดสินใจแบบถดถอย (Regression trees) ตัวอย่างการจำแนกด้วยอัลกอริทึม Decision Tree ดังภาพที่ 2.13



ภาพที่ 2.13 การจำแนกด้วยอัลกอริทึม Decision Tree

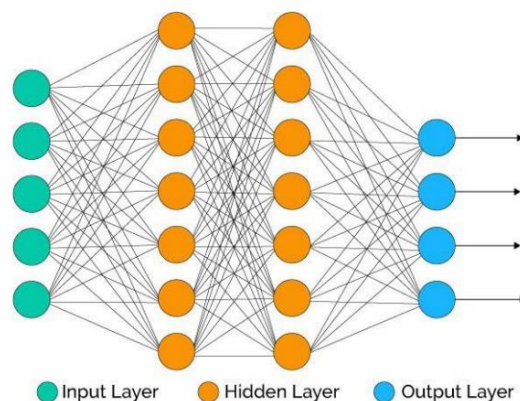
3) Neural Networks หรือ โครงข่ายประสาทเทียม เป็นอัลกอริทึมที่เลียนแบบการทำงานของสมองของมนุษย์ ประกอบด้วย นิวรอน คือ เซลล์ประสาท เดนไดร์ คือ เซลล์ประสาทในการรับกระแสประสาท และแอกซอน คือ การส่งกระแสประสาทจากแนวคิดดังกล่าวจึงได้กลายมากระบวนการทำงานของโครงข่ายประสาทเทียม ซึ่งหมายถึงเครือข่ายของโหนดที่เชื่อมต่อถึงกัน (Alghoul et al., 2018) โดยการเรียนรู้ของอัลกอริทึม Neural Networks ใช้

กระบวนการเรียนรู้ข้อมูลโดยการปรับค่าน้ำหนักให้เป็นค่าที่เหมาะสมที่สุด สามารถแบ่งการเรียนรู้ของอัลกอริทึม Neural Networks เป็น 2 ประเภท คือ การเรียนรู้แบบมีผู้สอน ที่จะมีการตรวจคำตอบเพื่อให้อัลกอริทึมปรับตัว โดยจะตรวจสอบว่าให้คำตอบที่ถูกหรือไม่ ถ้าตอบไม่ถูกต้องก็จะปรับตัวเองเพื่อให้ได้คำตอบที่ดีขึ้น และการเรียนรู้แบบไม่มีผู้สอน ที่เป็นการเรียนแบบไม่มีผู้แนะนำไม่มีการตรวจคำตอบว่าถูกหรือผิด แต่อัลกอริทึมจะจัดเรียงโครงสร้างด้วยตัวเองตามลักษณะของข้อมูลผลลัพธ์ที่ได้และจะสามารถจัดหมวดหมู่ของข้อมูลได้เอง โดยอัลกอริทึม Neural Networks ประกอบด้วย 3 ส่วน ดังภาพที่ 2.14 ได้แก่

(1) ชั้นข้อมูลเข้า (Input Layer) ชั้นนี้จะเป็นข้อมูลนำเข้า โดยจำนวนของโหนดในชั้นนี้ขึ้นอยู่กับจำนวนของข้อมูลนำเข้าที่จะนำมาสร้างแบบจำลอง เช่น ถ้าข้อมูลนำเข้าประกอบด้วย ทิศทางลม อุณหภูมิ ความกดอากาศ ทิศวิสัย รวม 4 ตัวแปร ดังนั้นชั้นข้อมูลก็จะมี 4 โหนด ซึ่งจะเรียกตัวแปรที่นำมาวิเคราะห์เหล่านี้ว่าคุณลักษณะหรือแอตทริบิวต์

(2) ชั้นซ่อน (Hidden Layer) เป็นชั้นที่อยู่ตรงกลาง ซึ่งมีผลอย่างมากต่อประสิทธิภาพในการเรียนรู้ของแบบจำลอง โดยชั้นซ่อนนี้จะมีกี่ชั้นก็ได้และแต่ละชั้นจะมีจำนวนนิวรอนเท่าใดก็ได้เช่นกันแต่ยิ่งกำหนดจำนวนมากโครงข่ายก็จะยิ่งมีความซับซ้อนมากยิ่งขึ้น แต่ในทางปฏิบัตินั้นการกำหนดชั้นซ่อน 1 ชั้นก็เพียงพอต่อการวิเคราะห์ข้อมูล (สายชล สีนสมบูรณ์ทอง,2563) ซึ่งการเพิ่มชั้นและจำนวนนิวรอนส่งผลต่อการทำงานของแบบจำลองทำให้แบบจำลองใช้เวลานานยิ่งขึ้นในการประมวลผล

(3) ชั้นข้อมูลออก (Output Layer) เป็นชั้นที่จะนำเอาข้อมูลจากการคำนวณไปใช้และจำนวนโหนดในชั้นนี้ขึ้นอยู่กับรูปแบบของข้อมูลออกที่จะนำไปใช้ต่อไป



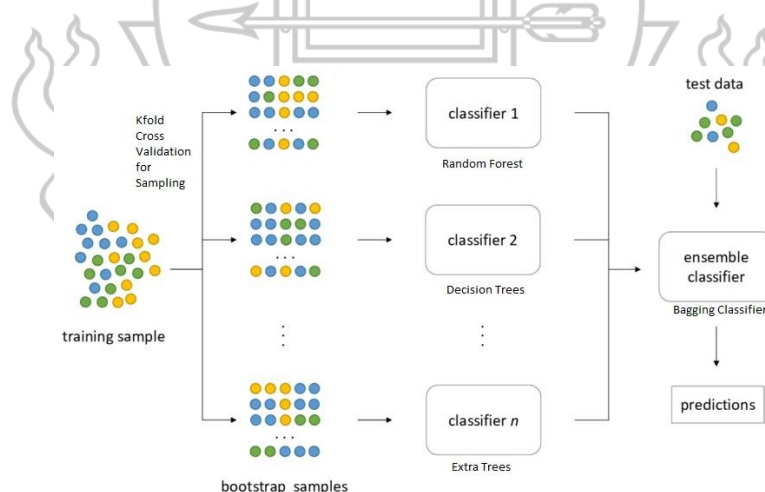
ภาพที่ 2.14 ส่วนประกอบของ Neural Networks

ที่มา: <https://pancharoen-c.medium.com/back-propagation-algorithm-from-scratch>

5.6.2 อัลกอริทึมการเรียนรู้แบบกลุ่ม (Ensemble Learning Algorithms)

ปรเมษฐ์ ธีรวานนท์ และคณะ (2560) ได้ให้ความหมายของอัลกอริทึมการเรียนรู้แบบกลุ่มว่า เป็นวิธีการอาศัยแบบจำลองการจำแนกข้อมูลมากกว่าหนึ่งตัว แต่แต่ละแบบจำลองการจำแนกข้อมูลจะมีกระบวนการทำงานของตนเอง และทุกแบบจำลองการจำแนกจะสร้างจากกลุ่มข้อมูลเดียวกัน เมื่อได้ผลลัพธ์ของแต่ละแบบจำลอง ก็จะนำผลลัพธ์ที่ได้เหล่านั้นผ่านวิธีการรวบรวม (Combination Integration หรือ Vote) และตัดสินผลลัพธ์สุดท้าย (Final Decision) เพื่อให้ได้ผลลัพธ์การจำแนกข้อมูลเดียวกันเท่านั้น โดยในงานวิจัยนี้ใช้ 3 เทคนิค ดังนี้

1) Bagging เป็นเทคนิคช่วยเพิ่มประสิทธิภาพในการสร้างแบบจำลองโดยทำงานร่วมกับเทคนิคการเรียนรู้ต่าง ๆ เช่น Decision Tree และ Random Forest เป็นต้น โดยจะถูกสร้างจากชุดข้อมูลเรียนรู้ที่กำหนดให้แบบใส่คืน (Sampling with replacement) กล่าวคือ เมื่อสุ่มข้อมูลขึ้นมาได้หนึ่งตัวจะคืนข้อมูลนั้นกลับเข้าไปยังชุดข้อมูลเดิม ทำให้ข้อมูลนั้นถูกสุ่มซ้ำอีกในอนาคต จากนั้นใช้หลักการรวมผลลัพธ์ด้วย Majority Vote เพื่อหาแบบจำลองที่สามารถตอบข้อผิดพลาดได้ดีที่สุดมาเป็นแบบจำลองสำหรับการจำแนกข้อมูลชุดนั้น ซึ่ง Bagging เป็นเทคนิคของการเรียนรู้แบบกลุ่มที่ช่วยลดความผิดพลาดของการจำแนกประเภทข้อมูล โดยมีเป้าหมายที่จะลดความผิดพลาดของแบบจำลองการจำแนกที่ถูกสร้างให้มีความจำเพาะและซับซ้อนมากเกินไป (Complexity) ทำให้เกิดปัญหา Overfitting และหลักการทำงานของเทคนิค Bagging ดังภาพที่ 2.15



ภาพที่ 2.15 หลักการทำงานของเทคนิค Bagging

ที่มา: <https://medium.com/ml-research-lab/bagging-ensemble-meta-algorithm-for-reducing-variance-c98fffa5489f>

2) AdaBoost หรือชื่อเต็มของเทคนิคนี้คือ Adaptive Boosting เป็นเทคนิคสำหรับช่วยเพิ่มประสิทธิภาพในการสร้างแบบจำลองการเรียนรู้แบบกลุ่มประเภท Boosting โดยทำงานร่วมกับเทคนิคการเรียนรู้ต่าง ๆ เช่น เทคนิค AdaBoost ร่วมกับ Random Forest หลักการทำงานเริ่มด้วยการสร้างชุดข้อมูลฝึกฝน (Training data) จากการกำหนดค่าน้ำหนัก (Weight) ให้กับข้อมูลแต่ละตัว จากนั้นนำข้อมูลเรียนรู้ไปสร้างแบบจำลองกับเทคนิคการเรียนรู้ ซึ่งในแต่ละรอบที่มีการสร้างแบบจำลองใหม่ค่าน้ำหนักจะถูกเปลี่ยนไป ถ้าแบบจำลองตอบผิดพลาด ข้อมูลตัวนั้นจะถูกเพิ่มค่าน้ำหนัก และถ้าแบบจำลองตอบถูกข้อมูลตัวนั้นก็จะถูกลดค่าน้ำหนักลง หลักการทำงานโดยสรุปคือ จะใช้เทคนิค AdaBoost รวบรวมแต่ละขั้นตอนของการเรียนรู้ที่จำแนกประเภทแล้วได้ผลลัพธ์ที่แย่ (Weak Learners) จะถูกปรับปรุงให้มีผลลัพธ์ที่ดีขึ้น (Strong Learners) และหลักการทำงานของเทคนิค AdaBoost ดังภาพที่ 2.16

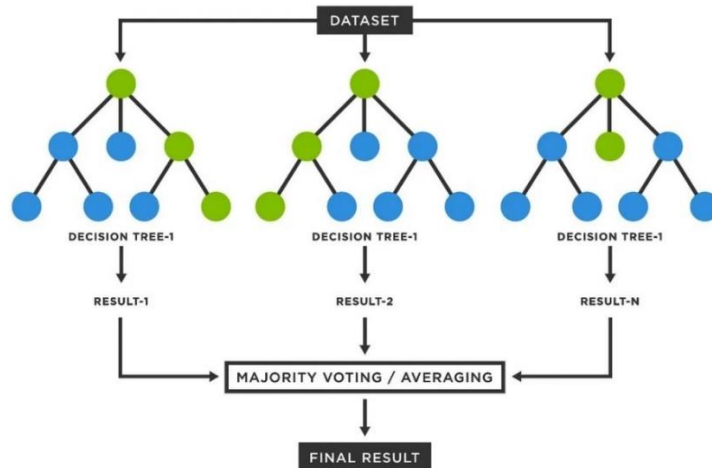


ภาพที่ 2.16 หลักการทำงานของเทคนิค AdaBoost

ที่มา: <https://www.geeksforgeeks.org/boosting-in-machine-learning-boosting-and-adaboost/>

3) Random Forest เป็นหนึ่งในเทคนิคของการเรียนรู้แบบกลุ่ม โดยหลักการทำงานของ Random Forest คือ สร้างแบบจำลองจากอัลกอริทึม Decision Tree หลายๆ ต้น ซึ่งแต่ละแบบจำลองจะได้รับชุดข้อมูลไม่เหมือนกันแต่ข้อมูลนั้นเป็นส่วนหนึ่งในชุดข้อมูลทั้งหมด และผลลัพธ์ที่ได้มาจาก Decision Tree ทำนายผลของตัวเอง แล้วสรุปผลการจำแนกด้วยวิธีการ Majority Vote ที่เป็นคำตอบซึ่งถูกเลือกมากที่สุด โดย Decision Tree เพียงต้นเดียวอาจยังไม่ได้

ผลลัพธ์ที่ดี แต่การนำมาเรียนรู้แบบกลุ่มจึงทำให้แบบจำลองมีความแม่นยำมากยิ่งขึ้น และหลักการ
ทำงานของเทคนิค Random Forest ดังภาพที่ 2.17



ภาพที่ 2.17 หลักการทำงานของเทคนิค Random Forest

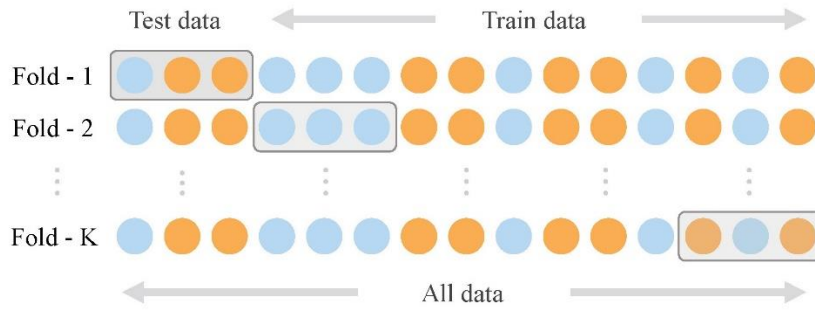
ที่มา: <https://divakarkumarp.medium.com/how-did-the-random-forest-algorithm-work-in-machine-learning-9e044573898b>

5.7 การประเมินประสิทธิภาพแบบจำลอง (Performance Evaluation)

การประเมินประสิทธิภาพเป็นขั้นตอนที่สำคัญในการสร้างแบบจำลองด้วยเทคนิคการเรียนรู้ของเครื่อง เพื่อให้ทราบถึงประสิทธิภาพของการจำแนก รวมถึงสามารถเปรียบเทียบประสิทธิภาพระหว่างอัลกอริทึมเพื่อให้สามารถคัดเลือกอัลกอริทึมที่ดีที่สุดสำหรับชุดข้อมูลนั้น โดยวิธีประเมินประสิทธิภาพสำหรับการพยากรณ์การเกิดพายุฝนฟ้าคะนองระยะปัจจุบันในงานวิจัยนี้ มีรายละเอียดดังต่อไปนี้

5.7.1 K-fold Cross Validation เป็นวิธีทางสถิติเพื่อประเมินประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่อง โดยกระบวนการนี้สามารถปรับเปลี่ยนค่า K ที่หมายถึงกลุ่มจำนวนของการแบ่งตัวอย่างออกเป็นกลุ่มย่อยขนาดเท่ากัน (folds) ก่อนการประเมินประสิทธิภาพ เพื่อให้ข้อมูลทุกตัวมีโอกาสเป็นชุดทดสอบและชุดฝึกฝน ป้องกันปัญหาการเลือกข้อมูลที่ดีและง่ายมาเป็นชุดข้อมูลทดสอบ โดยการฝึกฝนและทดสอบของแบบจำลองจะทำทั้งหมด K รอบ โดยรอบแรกใช้ข้อมูลชุดที่ 1 เป็นข้อมูลทดสอบ ส่วนข้อมูลชุดที่เหลือจะกลายเป็นชุดข้อมูลฝึกฝน จากนั้นสลับข้อมูลชุดที่ 2 มาเป็นชุดข้อมูลทดสอบและข้อมูลกลุ่มอื่นๆที่เหลือเป็นชุดข้อมูลฝึกฝน สลับเช่นนี้ไปเรื่อย ๆ จนครบ

K ชุด และคำนวณหาประสิทธิภาพเฉลี่ย หลักการทำงานของ K-fold Cross Validation ดังที่กล่าวมาแล้วนั้นแสดงไว้ดังภาพที่ 2.18



ภาพที่ 2.18 หลักการทำงานของ K-fold Cross Validation

ที่มา: <https://www.philschmid.de/k-fold-as-cross-validation-with-a-bert-text-classification-example>

5.7.2 Confusion Matrix เป็นตารางสรุปจำนวนข้อมูลที่แบบจำลองมีการทำนาย การเกิดพายุฝนฟ้าคะนองได้ถูกต้องและไม่ถูกต้อง

ตารางที่ 2.2 Confusion Matrix สำหรับ Binary Classification

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

จากตารางที่ 2.2 TP คือ จำนวนข้อมูลที่ทำนายว่าเกิดพายุฝนฟ้าคะนองแล้วเกิดขึ้นจริง
 FP คือ จำนวนข้อมูลที่ทำนายว่าเกิดพายุฝนฟ้าคะนองแต่แท้จริงไม่เกิด
 FN คือ จำนวนข้อมูลที่ทำนายว่าไม่เกิดพายุฝนฟ้าคะนองแต่แท้จริงเกิด
 TN คือ จำนวนข้อมูลที่ทำนายว่าไม่เกิดพายุฝนฟ้าคะนองแล้วไม่เกิดขึ้นจริง

โดยจำนวนข้อมูลที่ทำนายได้จากตาราง Confusion Matrix นี้ สามารถคำนวณค่าต่างๆเพื่อประเมินประสิทธิภาพของแบบจำลองเพื่อการพยากรณ์การเกิดพายุฝนฟ้าคะนองระยะปัจจุบันได้ ดังนี้

1) Accuracy คือ ค่าความถูกต้องของแบบจำลอง หาได้จากอัตราส่วนของค่าที่แบบจำลองทำนายได้ถูกต้องเทียบกับค่าผลลัพธ์การทำนายทั้งหมด โดยพิจารณาการทำนายทุกคลาส สามารถคำนวณได้จากสมการที่ 2.5

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.5)$$

2) Precision คือ ค่าความแม่นยำของการจำแนกคลาส Positive หาได้จากอัตราส่วนระหว่างจำนวนที่ทำนายถูกต้องของ Positive เทียบกับจำนวนทั้งหมดที่ถูกจำแนกว่าเป็น Positive สามารถคำนวณได้จากสมการที่ 2.6

$$Precision = \frac{TP}{TP + FP} \quad (2.6)$$

3) Recall คือ ค่าความระลึก เป็นการวัดความถูกต้องของแบบจำลองโดยพิจารณาแยกทีละคลาส หาได้จากอัตราส่วนระหว่างจำนวนที่ทำนายถูกต้องของคลาสที่กำลังพิจารณา/จำนวนที่ทำนายทั้งถูกและผิดของคลาสที่กำลังพิจารณา สามารถคำนวณได้จากสมการที่ 2.7

$$Recall = \frac{TP}{TP + FN} \quad (2.7)$$

4) F-measure คือ ค่าความถ่วงดุล โดยเป็นค่าประสิทธิภาพโดยรวมซึ่งเป็นวิธีการประเมินประสิทธิภาพของแบบจำลองที่ดีกว่าค่า Accuracy ในกรณีเป็นข้อมูลแบบไม่สมดุล เพราะค่า F-measure จะให้ความสำคัญกับ False Positive และ False Negative ร่วมด้วย แบบจำลองที่สามารถจำแนกได้ดีจะมีค่าใกล้เคียงหนึ่ง สามารถคำนวณได้จากค่า Precision และ Recall ดังสมการที่ 2.8

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (2.8)$$

5) Balanced Accuracy คือ ค่าความถูกต้องสมดุล โดยเป็นค่าที่แสดงถึงการทำนายความถูกต้องแม่นยำในการทำงานเฉลี่ยของทั้งสองคลาส ซึ่งใช้งานกับการจัดการข้อมูลของแต่ละคลาสไม่สมดุลกัน ดังสมการที่ 2.9

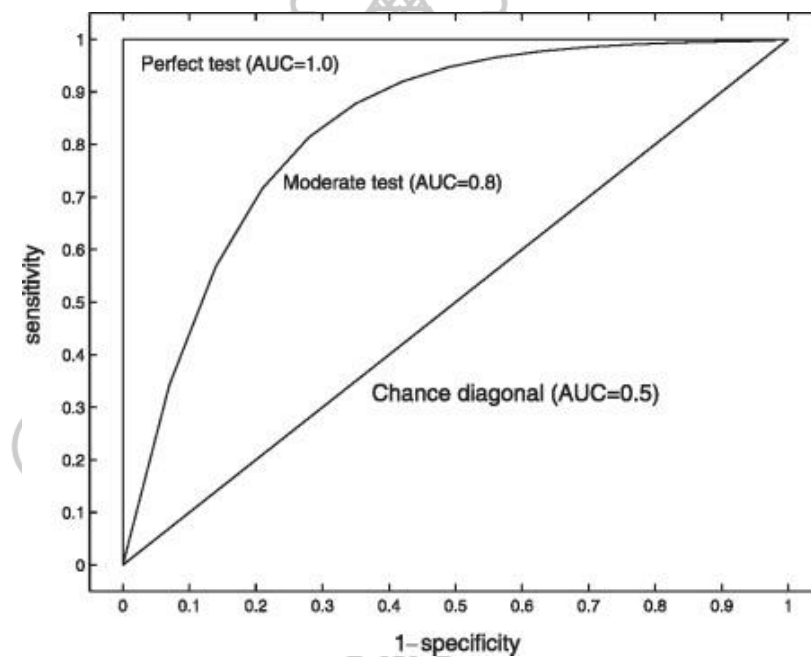
$$Balanced Accuracy = \frac{Sensitivity + Specificity}{2} \quad (2.9)$$

โดยที่

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (2.10)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (2.11)$$

5.7.3 Area Under the Curve (AUC) คือ ค่าพื้นที่ใต้กราฟ Receiver Operating Characteristic (ROC) ซึ่งเป็นกราฟที่พล็อตระหว่างค่า 1- Specificity (แกน x) และค่า Sensitivity (แกน y) ดังภาพที่ 2.19 โดยกราฟนี้แสดงให้เห็นถึงประสิทธิภาพของการจำแนกคลาสชนิด Binary โดยแบ่งผลลัพธ์ของการพยากรณ์ออกเป็น 2 กลุ่ม คือ กลุ่มที่เกิดเหตุการณ์และไม่เกิดเหตุการณ์



ภาพที่ 2.19 กราฟ Receiver Operating Characteristic (ROC)

ที่มา: Devos et al. (2007, pp.302)

ค่า AUC เป็นพื้นที่ใต้กราฟของ ROC curve ระหว่าง (0,0) กับ (1,1) โดยการแปลผลเพื่อแสดงประสิทธิภาพของแบบจำลองนั้น โดย AUC จะมีค่าอยู่ระหว่าง 0-1 เมื่อค่าเข้าใกล้ 1 หมายความว่า แบบจำลองในภาพรวมสามารถจำแนกข้อมูลได้ดีมาก และเกณฑ์การวัดประสิทธิภาพจากค่า AUC เป็นดังนี้ (นันทพัชร เสนาวงศ์ และคณะ, 2565)

$0.5 \leq AUC < 0.7$	คือ	แบบจำลองมีประสิทธิภาพต่ำ
$0.7 \leq AUC < 0.8$	คือ	เกณฑ์มาตรฐานสำหรับแบบจำลองส่วนใหญ่
$0.8 \leq AUC < 0.9$	คือ	แบบจำลองทำงานได้ดี
$AUC > 0.9$	คือ	แบบจำลองทำงานได้ดีมาก

6. งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับการศึกษาครั้งนี้ ดังนี้

6.1 งานวิจัยด้านการพยากรณ์อากาศ ด้วยเทคนิคการเรียนรู้ของเครื่อง

สมปราษฎ์ ศรีถกล (2564) ได้ศึกษาการพยากรณ์ฝนระยะปัจจุบันด้วยเทคนิคการเรียนรู้ของเครื่อง โดยใช้ข้อมูลปริมาณน้ำฝนจากการตรวจวัดสะสม 5 นาทีจาก 130 สถานีตรวจวัดที่กระจายในพื้นที่กรุงเทพมหานคร ในปี ค.ศ.2017-2020 ศึกษาจำนวน 6 ระยะเวลาพยากรณ์ คือ 30, 60, 90, 120, 150 และ 180 นาทีข้างหน้า มีการจำแนกประเภทของลักษณะเหตุการณ์การเกิดฝนออกเป็น 2 กลุ่มคือ ฝนตก (Rain Class) และฝนไม่ตก (Dry Class) ด้วยอัลกอริทึม CART MLP Random Forest และ SVM จากการวิจัยพบว่า การพยากรณ์ฝนระยะปัจจุบันด้วยข้อมูลน้ำเข้า คือ ปริมาณน้ำฝนเพียงอย่างเดียวเป็นไปได้ โดยความแม่นยำสูงสุดที่ระยะเวลาพยากรณ์ 30 นาที ซึ่งมีความแม่นยำสูงขึ้นเมื่อระยะเวลาพยากรณ์สั้นลง และวัดประสิทธิภาพแบบจำลองด้วยค่า F-measure โดยเฉลี่ยที่เวลาพยากรณ์ 90 นาทีเป็นระยะเวลาที่นานที่สุดที่มีค่าความแม่นยำที่ยอมรับได้ อันเป็นช่วงเวลาที่เพียงพอในการปรับตัวก่อนฝนตก โดย Random Forest ให้ผลลัพธ์ที่ดีที่สุด โดยมีค่า F-measure ของ Random Forest CART SVM และ MLP เท่ากับ 0.73 0.69 0.63 และ 0.59 ตามลำดับ

เฟื่องลดา มะโนกิจ (2562) ได้ศึกษาการพยากรณ์ปริมาณน้ำฝนของประเทศไทย ด้วยวิธีการเรียนรู้เชิงลึกสองแบบจำลองต่อกัน ได้แก่ แบบจำลองการจำแนกประเภทเพื่อนำมาจำแนกว่าในช่วงชั่วโมงข้างหน้าฝนจะตกหรือไม่ และแบบจำลองวิเคราะห์การถดถอยเพื่อทำนายปริมาณน้ำฝนที่เกิดขึ้นจริงเฉพาะช่วงที่ฝนตกเท่านั้นโดยใช้เทคนิคการใส่ข้อมูลที่ถูกทำนายแล้วเป็นข้อมูลรับเข้าเพื่อทำนายในขั้นตอนที่สอง ซึ่งข้อมูลน้ำเข้าแบ่งตามภูมิภาคของประเทศไทย ประกอบไปด้วย 5 คุณลักษณะ ได้แก่ ปริมาณน้ำฝน อุณหภูมิ ความชื้นสัมพัทธ์ และความดันอากาศ ระยะเวลาพยากรณ์ 1 ถึง 6 ชั่วโมงข้างหน้า ซึ่งข้อมูลที่ใช้ในงานวิจัยมีความไม่สมดุลระหว่างคลาสที่ฝนตกและฝนไม่ตก จึงเปรียบเทียบประสิทธิภาพด้วยค่า F-measure มาเป็นตัวชี้วัดในข้อมูลชุดนี้เพื่ออธิบายความสามารถ

ในการทำนายคลาสฝนตก จากการวิจัยพบว่า ระยะเวลาพยากรณ์ 1 ชั่วโมงมีค่าความคลาดเคลื่อนต่ำที่สุดและเมื่อระยะเวลาพยากรณ์นานขึ้นทำให้ค่าความคลาดเคลื่อนสูงขึ้น

Moon, Kim, Lee, and Moon (2018) ได้ศึกษาการประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่องเพื่อสร้างระบบเตือนภัยล่วงหน้าสำหรับการแจ้งเตือนฝนตกหนักในการพยากรณ์ระยะสั้นมาก โดยจะสร้างสัญญาณการเตือนภัยเมื่อคาดการณ์ว่าจะถึงเกณฑ์ที่แนะนำของฝนตกหนักที่จะเกิดขึ้นภายใน 3 ชั่วโมงข้างหน้า ข้อมูลนำเข้า ได้แก่ ทิศทางและความเร็วลมม ความเร็วลมในแนวตั้ง และแนวราบ อุณหภูมิ ความชื้นสัมพัทธ์ ความดันบรรยากาศ ความกดอากาศที่ระดับน้ำทะเลปานกลาง การตรวจพบฝน และปริมาณน้ำฝนรายชั่วโมงตั้งแต่ 1-12 ชั่วโมง ใช้อัลกอริทึม Logistic regression เพื่อทำนายว่าต้องมีการแจ้งเตือนล่วงหน้าหรือไม่ จากการวิจัยพบว่า แบบจำลองที่นำเสนอสามารถทำนายฝนตกหนักระยะเวลาพยากรณ์ 3 ชั่วโมงข้างหน้าได้ดี โดยเปรียบเทียบประสิทธิภาพด้วยค่า F-measure และค่าอัตราส่วนของจำนวนเหตุการณ์ที่ทำนายได้ถูกต้องว่าฝนตก

Pirone, Cimorelli, Giudice, and Pianese (2022) ได้ศึกษาการพยากรณ์ปริมาณฝนระยะปัจจุบัน โดยใช้เทคนิคการเรียนรู้ของเครื่อง กรณีศึกษาทางตอนใต้ของอิตาลี โดยได้นำเสนอแบบจำลองการทำนายปริมาณน้ำฝนระยะปัจจุบันโดยใช้โครงข่ายประสาทเทียมแบบป้อนข้อมูลไปข้างหน้า (feed-forward neural network) ซึ่งได้นำเข้าข้อมูลปริมาณน้ำฝนของแต่ละพื้นที่ มีระยะเวลาพยากรณ์ คือ 30 นาที 1 ชั่วโมง และ 2 ชั่วโมงข้างหน้า จากการวิจัยพบว่า ระยะเวลาพยากรณ์ 30 นาทีได้แบบจำลองที่มีประสิทธิภาพสูงสุด ขณะที่ระยะเวลาพยากรณ์ 2 ชั่วโมงมีประสิทธิภาพต่ำที่สุด อีกทั้งแบบจำลองที่สร้างขึ้นทำให้เกิดการทำนายและเรียนรู้ความสัมพันธ์ที่ช่วยอธิบายช่วงเวลาวิวัฒนาการของการเกิดฝนได้ คุณลักษณะที่สำคัญของแบบจำลองคือ ใช้เวลาในการประมวลผลสั้น จึงเหมาะสมสำหรับระบบเตือนภัยล่วงหน้าแบบเรียลไทม์

Ojo and Ogunjo (2022) ได้ศึกษาการสร้างแบบจำลองการเรียนรู้ด้วยเครื่องสำหรับการทำนายปริมาณน้ำฝนรายเดือนและรายปีสำหรับประเทศไนจีเรีย โดยใช้การวิเคราะห์การถดถอยเชิงพหุหลายตัวแปร (multivariate polynomial regressions: MPR) และอัลกอริทึมการเรียนรู้ด้วยเครื่อง 12 อัลกอริทึม ได้แก่ โครงข่ายประสาทเทียม (Artificial Neural Networks: ANN) 3 ชนิด ระบบอนุมานฟัซซี่แบบปรับตัวได้ (Adaptive Neuro- Fuzzy Inference system: ANFIS) 4 ชนิด และซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine : SVM) 5 ชนิด เพื่อประมาณค่าปริมาณน้ำฝนรายเดือนและรายปีในบริเวณพื้นที่เขตร้อน มีข้อมูลนำเข้า ได้แก่ ข้อมูลปริมาณน้ำฝนภาคพื้นดิน ลองจิจูด ละติจูด และระดับความสูง จากการวิจัยพบว่า ในแต่ละเดือนมีแบบจำลองที่

เหมาะสมสำหรับการพยากรณ์แตกต่างกัน และสำหรับการพยากรณ์รายปีการใช้แบบจำลองของอัลกอริทึม ANFIS มีผลลัพธ์ที่ดีที่สุด

Sunori et al. (2021) ได้ศึกษาการจำแนกระดับความรุนแรงของฝนโดยใช้เทคนิคการเรียนรู้ของเครื่อง ซึ่งต้องการจำแนกระดับความรุนแรงของฝน 2 ชนิด คือ ฝนตกหนัก (Huge Rainfall) และฝนตกกำลังปานกลาง (Moderate Rainfall) มีข้อมูลนำเข้า ได้แก่ อุณหภูมิสูงสุด และความชื้นสัมพัทธ์ สร้างแบบจำลองด้วยอัลกอริทึมต่างๆ ได้แก่ Gaussian Naive Bayes Kernel Naive Bayes Fine KNN Cubic KNN Fine tree และ Coarse tree จากการวิจัยพบว่า ทุกแบบจำลองในงานวิจัยนี้มีประสิทธิภาพการจำแนกที่ดี โดยแบบจำลองที่มีความแม่นยำสูงสุด คือแบบจำลองจากอัลกอริทึม Cubic KNN เท่ากับ 90.6% รองลงมาคือ Fine KNN เท่ากับ 89.4%

Shah, Shukla, and Priamvada (2022) ได้ศึกษาการสร้างแบบจำลองการเรียนรู้แบบกลุ่มเพื่อพยากรณ์ปริมาณน้ำฝนระยะปัจจุบัน โดยใช้ข้อมูลจากสถานีตรวจอากาศอัตโนมัติ ซึ่งมีข้อมูลนำเข้า ได้แก่ ความชื้น อุณหภูมิ ความกดอากาศ ข้อมูลแสงแดด และปริมาณน้ำฝน สร้างแบบจำลองจากอัลกอริทึมการเรียนรู้แบบกลุ่ม ได้แก่ Bagging Boosting Random Forest AdaBoost และ XgBoost มีระยะเวลาพยากรณ์ 4 ชั่วโมงข้างหน้า จากการวิจัยพบว่า แบบจำลองจากอัลกอริทึม Random Forest และ XgBoost มีค่าความแม่นยำสูงสุดใกล้เคียงกัน คือ เท่ากับ 80% และดีกว่า AdaBoost ที่มีความแม่นยำเท่ากับ 65%

Alnawas, Khafaji, and Azeez (2022) ได้ศึกษาการพยากรณ์ปริมาณน้ำฝนรายวันสำหรับจังหวัด Thi-Qar ประเทศอิรัก ด้วยเทคนิคการเรียนรู้ของเครื่อง โดยการใช้ข้อมูลสภาพอากาศในอดีตที่มีข้อมูลนำเข้า ได้แก่ ปริมาณน้ำฝน อุณหภูมิ ความชื้นสัมพัทธ์ ความเร็วลม ค่าการระเหย และค่ารังสีแสงแดด โดยได้สร้างแบบจำลองการคาดการณ์ปริมาณน้ำฝนโดยใช้อัลกอริทึม Deep learning Decision Tree และ Naive Bayes จากการวิจัยพบว่า แบบจำลองอัลกอริทึม deep learning ได้ค่าความแม่นยำสูงสุด เท่ากับ 91.59%

Ghada, Estrella, and Menzel (2022) ได้ศึกษาการสร้างแบบจำลองการเรียนรู้ของเครื่องเพื่อจำแนกประเภทของฝนตามขนาดการกระจายตัวของเม็ดฝนประกอบการตรวจวัดชนิดของกลุ่มเมฆ ข้อมูลนำเข้า ได้แก่ ขนาดของเม็ดฝนจากเครื่องมือตรวจวัด และการตรวจวัดชนิดของกลุ่มเมฆด้วยคน โดยจะพยากรณ์เพื่อแบ่งแยกชนิดของฝนที่ตกว่า เป็นฝนเกิดจากการพาความร้อนหรือเป็นฝนที่เกิดจากเมฆแผ่น สร้างแบบจำลองด้วยอัลกอริทึม ได้แก่ KNN Naive Bayes

Conditional Trees และ Random Forest จากการวิจัยพบว่า แบบจำลองจากอัลกอริทึม Random Forest ได้ค่าความถูกต้องสูงสุด

Geetha and Nasira (2014) ได้ศึกษาการทำเหมืองข้อมูลสำหรับการใช้งานทางด้านอุตุนิยมวิทยา โดยสร้างแบบจำลองการทำนายปริมาณน้ำฝนรายปี มีข้อมูลนำเข้า ได้แก่ อุณหภูมิ อุณหภูมิจุดน้ำค้าง ความกดอากาศ ทิศนวิสัย ความเร็วลม ลมกระโชก อุณหภูมิ และปริมาณน้ำฝนสะสม ด้วยอัลกอริทึม Decision Tree จากการวิจัยพบว่า แบบจำลองในงานวิจัยนี้สามารถทำงานได้ดีโดยมีค่าความถูกต้องเท่ากับ 80.67%

6.2 งานวิจัยด้านอื่นๆที่เกี่ยวข้องกับเทคนิคการเรียนรู้ของเครื่อง

Dey and Pratap (2023) ได้ศึกษาเปรียบเทียบการใช้เทคนิคการสุ่มเกิน 3 วิธี คือ SMOTE Borderline-SMOTE และ ADASYN มีข้อมูลนำเข้า 3 ชุดข้อมูล คือ (1) ชุดข้อมูลโรคเบาหวานประเภทที่ 2 (2) ชุดข้อมูลโรคมะเร็งเต้านม และ (3) ชุดข้อมูลการเรียกร้องสินไหม ประกัยภัย ด้วยอัลกอริทึมการจำแนก ได้แก่ SVM KNN Gaussian NB Decision Tree และ Random Forest ทดสอบประสิทธิภาพด้วยค่า Accuracy Precision Recall F-measure และ ROC curve จากการวิจัยพบว่า อัลกอริทึม Random Forest สามารถทำงานได้ดีที่สุดกับทั้ง 3 เทคนิคของการสุ่มเกิน และการใช้ SMOTE ร่วมกับ Random Forest มีค่า Accuracy สูงสุด ขณะที่ Borderline-SMOTE และ ADASYN มีค่า Accuracy ที่ดีเมื่อใช้กับอัลกอริทึม SVM Decision tree และ Random Forest

อัจฉรา แผ้วบาง และสายชล สีนสมบูรณ์ทอง (2562) ได้ศึกษาการปรับความไม่สมดุลของข้อมูลด้วย 4 วิธี คือ วิธีการสุ่มเกิน วิธีการสุ่มเกินโดยเทคนิค SMOTE วิธีการสุ่มลด และวิธีการสุ่มผสมผสาน มีข้อมูลนำเข้า 3 ชุดข้อมูล คือ (1) ชุดข้อมูลเคมีบำบัดมะเร็งลำไส้ใหญ่ระยะ B/C (2) ชุดข้อมูลโรคที่มีความผิดปกติของโปรตีน และ (3) ชุดข้อมูลการรักษาอาการปวดศีรษะขั้นรุนแรง โดยใช้อัลกอริทึมการจำแนก 5 อัลกอริทึม ได้แก่ KNN Neural Networks SVM Rule-based และ Stochastic Gradient Descent (SGD) ทดสอบประสิทธิภาพด้วยค่าความถูกต้อง ค่าความไว ค่าความจำเพาะ ค่าคลาดเคลื่อนกำลังสองเฉลี่ย และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย จากการวิจัยพบว่า วิธีที่มีประสิทธิภาพสูงสุดคือ อัลกอริทึม Rule-based ปรับความไม่สมดุลด้วยเทคนิค SMOTE

พิชญะ พรหมลา และจรัญ แสนราช (2563) ได้ศึกษาการเปรียบเทียบประสิทธิภาพการวิเคราะห์ความพึงพอใจเกี่ยวกับการจัดการเรียนการสอนด้วยกระบวนการวิเคราะห์ความรู้สึก ข้อมูลนำเข้า คือ ชุดข้อมูลวิเคราะห์ความคิดเห็นที่ได้จากแบบสอบถาม ใช้เทคนิคการเรียนรู้แบบ

รวมกลุ่มเพื่อจำแนกข้อมูล ได้แก่ Vote Bagging และ Random Forest ร่วมกับเทคนิควิธีมาตรฐาน ได้แก่ Decision Tree Naive Bayes และ K-NN ทดสอบประสิทธิภาพด้วยค่า Accuracy Precision Recall F-measure และ AUC จากการวิจัยพบว่า เทคนิคการเรียนรู้แบบรวมกลุ่มเพื่อจำแนกข้อมูลแบบ Vote มีประสิทธิภาพมากที่สุด

ธีรยุทธ คุณสุข และจारी ทองคำ (2563) ได้ศึกษากระบวนการคัดเลือกคุณลักษณะสำหรับเพิ่มประสิทธิภาพในการจำแนกความคิดเห็นของลูกค้า ชุดข้อมูลนำเข้าจาก เว็บไซต์ wongnai.com ใช้ 3 เทคนิคในการคัดเลือกคุณลักษณะของข้อความ คือ Chi-Square Information Gain และ Gain Ratio ด้วยอัลกอริทึม Naive Bayes SVM KNN และ C4.5 ทดสอบประสิทธิภาพด้วยค่า Accuracy Precision และ Recall จากการวิจัยพบว่า การคัดเลือกคุณลักษณะแบบ Information Gain ร่วมกับเทคนิค Naive Bayes ให้ผลที่ดีที่สุดในการจำแนกความคิดเห็น โดยมีค่า Accuracy เท่ากับ 89.08% ค่า Precision เท่ากับ 89.12 % และค่า Recall เท่ากับ 89.10 %

ปรเมษฐ์ ธีรวานนท์ และคณะ (2560) ได้ศึกษาการประยุกต์ใช้แบบจำลองการเรียนรู้แบบกลุ่ม เพื่อพยากรณ์แนวโน้มของราคาหลักทรัพย์ในตลาดหลักทรัพย์แห่งประเทศไทย ชุดข้อมูลนำเข้า คือ ข้อมูลหลักทรัพย์จากตลาดหลักทรัพย์แห่งประเทศไทย ใช้อัลกอริทึม Random Forest SVM KNN Naive Bayes และอัลกอริทึมการเรียนรู้แบบกลุ่มด้วยเทคนิค Boosting โดยทดสอบประสิทธิภาพด้วยค่า Accuracy จากการวิจัยพบว่า อัลกอริทึมการเรียนรู้แบบกลุ่มด้วยวิธีการถ่วงน้ำหนักสามารถเพิ่มประสิทธิภาพในการพยากรณ์ได้ถึง 5%-14%

กาญจน์ ณ ศรีระ และคณะ (2560) ได้ศึกษาการเปรียบเทียบเทคนิคการสุ่มตัวอย่างเพื่อการจำแนกข้อมูลที่ไม่สมดุล ด้วย 3 วิธี คือ เทคนิคการสุ่มเพิ่มตัวอย่างส่วนน้อย เทคนิคการสุ่มลดตัวอย่างส่วนมาก และเทคนิคการสุ่มตัวอย่างซ้ำ ชุดข้อมูลนำเข้า 3 ชุดข้อมูล คือ (1) ชุดข้อมูลโรคมะเร็งเต้านม (2) ชุดข้อมูลโรคเบาหวาน และ (3) ชุดข้อมูลภาวะแทรกซ้อนที่เท้าในผู้ป่วยเบาหวาน ใช้ อัลกอริทึม CART Decision Tree Random Forest SVM และ Multilayer Perceptron ร่วมกับอัลกอริทึมการเรียนรู้แบบกลุ่ม ได้แก่ AdaBoost และ Bagging เพื่อสร้างแบบจำลองสำหรับจำแนกข้อมูลการเกิดและไม่เกิดโรค ทดสอบประสิทธิภาพด้วยค่า Precision Recall และ F-measure จากการวิจัยพบว่า วิธีการสุ่มตัวอย่างซ้ำสามารถปรับปรุงข้อมูลที่ไม่สมดุลได้ดีที่สุด และแบบจำลอง Random Forest แบบจำลอง AdaBoost ร่วมกับ Random Forest และแบบจำลอง Bagging ร่วมกับ Random Forest มีค่าประสิทธิภาพการจำแนกที่ดีกับข้อมูลในงานวิจัย

ธรรมบุญ ปัญญาทิพย์ และคณะ (2565) ได้ศึกษาการเพิ่มประสิทธิภาพจำแนกข้อมูลผลกระทบโควิด-19 ต่อผู้ป่วยมะเร็งระดับ ข้อมูลนำเข้า คือ ชุดข้อมูล COVID-19 effect on Liver Cancer จาก Kaggle ใช้อัลกอริทึมเพื่อจำแนก ได้แก่ SVM Naive Bayes KNN Decision Tree และ Neural Networks และศึกษาการเพิ่มประสิทธิภาพการจำแนกข้อมูลด้วยการใช้อัลกอริทึมการเรียนรู้แบบกลุ่มเทคนิค Vote และ Random Forest จากการวิจัยพบว่า (1) การจำแนกข้อมูลผลกระทบโควิด-19 ต่อผู้ป่วยมะเร็งระดับ ด้วยการแบ่งชุดข้อมูล 80:20 อัลกอริทึมที่จำแนกดีที่สุด ได้แก่ Decision Tree และ Neural Networks มีความถูกต้องเท่ากับ 100% (2) การแบ่งชุดข้อมูลทดสอบ ประสิทธิภาพ 5-Fold Cross Validation อัลกอริทึมที่จำแนกดีที่สุด ได้แก่ Decision Tree มีความถูกต้องเท่ากับ 99.6% (3) ผลการเพิ่มประสิทธิภาพการจำแนกข้อมูลด้วยอัลกอริทึมการเรียนรู้แบบรวมกลุ่ม ทดสอบประสิทธิภาพ 5-Fold Cross Validation นั้น อัลกอริทึม Random Forest มีความถูกต้องสูงสุด เท่ากับ 99.8%

อรุณี ศรีดี (2564) ได้ศึกษาการประยุกต์ใช้การเรียนรู้ของเครื่องเพื่อตรวจจับงานเสียที่เกิดขึ้นในกระบวนการทดสอบฮาร์ดดิสก์ไดรฟ์ ลักษณะชุดข้อมูลที่นำเข้ามีจำนวนหลายแอตทริบิวต์และมีความไม่สมดุลสูง ใช้วิธีการเลือกคุณลักษณะด้วย Filter Method และ Embedded Method เพื่อมิติของข้อมูล ใช้อัลกอริทึม SVM และ XGBoost ร่วมกับ 3 เทคนิคในการจัดการข้อมูลที่ไม่สมดุล ได้แก่ SMOTE Different Cost Learner และ SMOTE กับ Different Cost ทดสอบประสิทธิภาพด้วยค่า Accuracy Precision Recall F-measure ROC AUC และ PRC AUC จากการวิจัยพบว่า แบบจำลองอัลกอริทึม XGBoost ร่วมกับ SMOTE ให้ประสิทธิภาพที่ดีที่สุด มีค่า Accuracy Precision Recall F-measure เท่ากับ 97% โดยที่ค่า ROC AUC เท่ากับ 91% และ PRC AUC เท่ากับ 73%

หนึ่งฤทัย ประเสริฐ (2564) ได้ศึกษาการออกแบบและพัฒนาแบบจำลองการจำแนกความผิดปกติของเสียงหัวใจ โดยนำเข้าข้อมูลเสียงหัวใจเพื่อสร้างแบบจำลองการจำแนกระหว่างเสียงหัวใจปกติ (Normal) และเสียงหัวใจผิดปกติ (Abnormal) ซึ่งใช้เทคนิค SMOTE ในการปรับความไม่สมดุลของข้อมูล ก่อนสร้างแบบจำลองด้วยอัลกอริทึม Neural Network Deep Learning และ SVM ประเมินประสิทธิภาพของแบบจำลองด้วยค่า Accuracy Precision Recall F-measure เป็นตัวตัดสินหลักของงานวิจัย นอกจากนั้นยังพิจารณาค่า Balance accuracy ROC-curve และ AUC ร่วมด้วย ผลวิจัยพบว่า อัลกอริทึม Neural Network มีประสิทธิภาพในการจำแนกสูงที่สุด

จากการทบทวนวรรณที่เกี่ยวข้องผู้วิจัยพบว่า งานวิจัยทางการพยากรณ์อากาศสามารถสร้างแบบจำลองได้ทั้งการวิเคราะห์การถดถอยโลจิสติก (Regression Model) และการจำแนก (Classification Model) ซึ่งเป็นแบบจำลองประเภทการเรียนรู้แบบมีผู้สอนเช่นเดียวกัน แต่มีความแตกต่างกันตรงที่ Regression Model มีคำตอบของแบบจำลองเป็นค่าตัวเลขที่จะเกิดขึ้นในอนาคตจากการนำเข้าข้อมูลเพื่อสร้างแบบจำลอง ไม่ใช่เป็นคำตอบที่เป็นตัวเลือกเหมือน Classification Model โดยมีหลายงานวิจัยที่สร้างแบบจำลองแบบ Regression Model ดังนำเสนอโดย Pirone, Cimorelli, Giudice, and Pianese (2022) ที่ต้องการทำนายปริมาณน้ำฝนสำหรับการพยากรณ์ระยะปัจจุบัน แต่ในงานวิจัยครั้งนี้ผู้วิจัยศึกษาการสร้าง Classification Model โดยพบว่าการวิจัยส่วนใหญ่ได้สร้างแบบจำลองการจำแนกโดยแบ่งคลาสของการทำนายออกเป็น 2 คลาส ซึ่งจัดเป็นชนิด Binary ที่มีเพียง 2 สถานะเท่านั้น ดังนำเสนอโดย สมปราชญ์ ศรีถกล (2564) และเฟื่องลดา มะโนกิจ (2562) แบ่งคลาสออกเป็นฝนตกและฝนไม่ตก รวมถึง Sunori et al. (2021) แบ่งคลาสออกเป็นฝนตกหนัก (Huge Rainfall) และฝนตกกำลังปานกลาง (Moderate Rainfall) และ Ghada, Estrella, and Menzel (2022) ที่แบ่งคลาสการทำนายออกเป็นฝนเกิดจากการพาความร้อนและฝนที่เกิดจากเมฆแผ่น ซึ่งการสร้างแบบจำลองการจำแนกชนิด Binary มีความรวดเร็วในการฝึกฝนมากกว่าแบบจำลองการจำแนกชนิด Multi-Class หรือ Multi-Label ทำให้มีความสะดวกในการเตรียมข้อมูลและเลือกวิธีการฝึกฝนที่เหมาะสมได้ง่ายยิ่งขึ้น อีกทั้งในกรณีที่ข้อมูลไม่สมดุลสามารถจัดการกับปัญหาความไม่สมดุลได้ง่ายกว่า เนื่องจากต้องมีการตัดสินใจเพียงแค่สองกลุ่มเท่านั้น

สำหรับการเลือกศึกษาในแต่ละระยะเวลาการพยากรณ์นั้น ผู้วิจัยพบว่า งานวิจัยโดย สมปราชญ์ ศรีถกล (2564) ได้ศึกษาการพยากรณ์อากาศระยะปัจจุบันถึงปานกลาง (ไม่เกิน 12 ชั่วโมง) โดยแบ่งเป็น 6 ระยะเวลาพยากรณ์ คือ 30, 60, 90, 120, 150 และ 180 นาทีข้างหน้า รวมถึง Pirone, Cimorelli, Giudice, and Pianese (2022) ที่ศึกษาการพยากรณ์อากาศระยะปัจจุบัน โดยแบ่งเป็น 3 ระยะเวลาพยากรณ์ คือ 30 นาที 1 ชั่วโมง และ 2 ชั่วโมงข้างหน้า ซึ่งทั้ง 2 งานวิจัยนี้มีจุดมุ่งหมายเพื่อให้เกิดการแจ้งเตือนฝนตกให้แก่ประชาชนได้ทันเวลาจึงจำเป็นต้องมีการแบ่งเวลาการพยากรณ์ให้สั้นลง นอกจากนั้นยังมีการพยากรณ์อากาศระยะที่ยาวนานกว่านี้ ดังนำเสนอโดย เฟื่องลดา มะโนกิจ (2562) ที่กำหนดระยะเวลาการพยากรณ์เป็น 1 ถึง 6 ชั่วโมงข้างหน้า สำหรับการแจ้งเตือนภัยระดับน้ำฝนที่สะสมซึ่งอาจก่อให้เกิดน้ำท่วมได้ทันต่อเหตุการณ์ และในงานวิจัยครั้งนี้ผู้วิจัยเลือกศึกษาระยะเวลาพยากรณ์ 1 และ 2 ชั่วโมงข้างหน้า เพื่อให้ครอบคลุมระยะเวลา 2 ชั่วโมงของการออกคำพยากรณ์ Trend Forecast สำหรับข่าวอากาศการบินที่ต้อง

รายงานทุก 1 ชั่วโมงนั่นเอง ในส่วนของอัลกอริทึมที่ใช้ศึกษามีหลากหลายทั้งที่เป็นอัลกอริทึมสำหรับ Classification Model และ Regression Model ดังนำเสนอโดย Ojo and Ogunjo (2022) สร้าง Regression Model โดยใช้โครงข่ายประสาทเทียม (Artificial Neural Networks: ANN) 3 ชั้น ระบบอนุมานฟัซซีแบบปรับตัวได้ (Adaptive Neuro- Fuzzy Inference system: ANFIS) 4 ชั้น และซัพพอร์ตเวกเตอร์แมทชีน (Support Vector Machine : SVM) 5 ชั้น เพื่อประมาณค่าปริมาณน้ำฝนรายเดือนและรายปี นอกจากนี้ Sunori et al. (2021) สร้าง Classification Model โดยใช้ Gaussian Naive Bayes, Kernel Naive Bayes, Fine KNN, Cubic KNN, Fine tree และ Coarse tree รวมถึงมีการใช้อัลกอริทึมการเรียนรู้แบบกลุ่มที่พบว่าทำให้ประสิทธิภาพของแบบจำลองเพิ่มขึ้นกว่าการใช้อัลกอริทึมมาตรฐานเพียงอย่างเดียว ดังนำเสนอโดยพิชญะ พรหมลา และจรัญ แสนราช (2563) ที่เลือกใช้ Vote Bagging และ Random Forest ร่วมกับอัลกอริทึมมาตรฐาน ได้แก่ Decision Tree Naive Bayes และ K-NN ซึ่งจากการทบทวนวรรณกรรมนั้นอัลกอริทึมที่นำมาสร้างแบบจำลองมีความหลากหลาย การเลือกนำไปใช้ขึ้นอยู่กับความเหมาะสมกับข้อมูลที่ผู้วิจัยต้องมีส่วนดำเนินการเพื่อพิจารณาในเบื้องต้น

สำหรับการวัดประสิทธิภาพของแบบจำลองนั้น ผู้วิจัยพบว่า งานวิจัยที่มีชุดข้อมูลแบบไม่สมดุลและจะใช้ค่า F-measure เป็นตัวตัดสินในการเลือกแบบจำลองที่ดีที่สุด ดังนำเสนอโดย สมปราชญ์ ศรีถกล (2564) เพ็องลดา มะโนกิจ (2562) รวมถึง Moon, Kim, Lee, and Moon (2018) เป็นต้น อีกทั้งใช้ค่า Accuracy Precision Recall Balance accuracy และ AUC เพื่อพิจารณาประสิทธิภาพของแบบจำลองด้านอื่น ๆ ร่วมกัน ดังนั้นผู้วิจัยมีแนวคิดพัฒนาและประเมินประสิทธิภาพของแบบจำลองในการพยากรณ์การเกิดพายุฝนฟ้าคะนองระยะปัจจุบัน ที่มีระยะเวลาพยากรณ์ 1 และ 2 ชั่วโมงข้างหน้า และทำการสร้างแบบจำลองการจำแนกข้อมูลด้วยอัลกอริทึมมาตรฐาน ได้แก่ Naive Bayes Decision Tree และ Neural Networks และสร้างแบบจำลองอัลกอริทึมมาตรฐานร่วมกับอัลกอริทึมการเรียนรู้แบบกลุ่ม ได้แก่ Bagging AdaBoost และ Random Forest เพื่อเพิ่มประสิทธิภาพการพยากรณ์การเกิดพายุฝนฟ้าคะนองระยะปัจจุบันให้มีความถูกต้องมากยิ่งขึ้น

บทที่ 3

วิธีดำเนินการวิจัย

การศึกษาวิจัย เรื่อง การพยากรณ์การเกิดพายุฝนฟ้าคะนองระยะปัจจุบันจากข่าวอากาศการบิน ด้วยเทคนิคการเรียนรู้ของเครื่อง กรณีศึกษาสนามบินภาคเหนือของประเทศไทย ผู้วิจัยได้ทำการค้นคว้าแนวคิดทฤษฎีและงานวิจัยที่เกี่ยวข้องเพื่อใช้เป็นแนวทางในการศึกษาตามกรอบแนวคิดวิจัย ดังนี้

1. เครื่องมือที่ใช้ในการวิจัย
2. การรวบรวมและเตรียมข้อมูลนำเข้า (Input)
3. การสร้างแบบจำลอง (Process)
4. การประเมินประสิทธิภาพของแบบจำลอง (Evaluation)

1. เครื่องมือที่ใช้ในการวิจัย

- | | |
|------------------------|--|
| 1.1 RapidMiner Studio | สำหรับสร้างแบบจำลองเพื่อวิเคราะห์เชิงทำนาย |
| 1.2 Microsoft Excel | สำหรับจัดการข้อมูลและสร้างกราฟแผนภูมิ |
| 1.3 Microsoft Power Bi | สำหรับแสดงผลเป็นแผนภูมิภาพต่าง ๆ |

2. การรวบรวมและเตรียมข้อมูลนำเข้า (Input)

2.1 การรวบรวมข้อมูล

2.1.1 ข้อมูลที่ใช้ในงานวิจัย

รวบรวมข้อมูลตั้งแต่วันที่ 1 มกราคม 2558 ถึง 31 ธันวาคม 2565 รวมระยะเวลา 8 ปี ซึ่งเป็นข้อมูลรายชั่วโมงของข่าวอากาศการบิน (METAR) ของสนามบินภาคเหนือจำนวน 9 สนามบิน รายงานโดยกรมอุตุนิยมวิทยา มีจำนวนข้อมูลทั้งหมด 364,382 ชุดข้อมูล โดยใช้ข้อมูลจากสนามบิน ดังต่อไปนี้

- 1) ทำอากาศยานนานาชาติเชียงใหม่ (VTCC)
- 2) ทำอากาศยานแม่ฮ่องสอน (VTCH)
- 3) ทำอากาศยานแม่ฟ้าหลวง เชียงราย (VTCT)
- 4) ทำอากาศยานลำปาง (VTCL)
- 5) ทำอากาศยานน่านนคร (VTCN)
- 6) ทำอากาศยานแพร่ (VTCP)
- 7) ทำอากาศยานพิษณุโลก (VTPP)
- 8) ทำอากาศยานสุโขทัย (VTPO)
- 9) ทำอากาศยานนานาชาติแม่สอด (VTPM)

2.1.2 แหล่งที่มาของข้อมูล

1) การสืบค้นข้อมูล รวบรวมข้อมูลตั้งแต่วันที่ 1 ม.ค. 2558 ถึง 31 ธ.ค. 2565 ระยะเวลารวม 8 ปี โดยใช้ข้อมูลตรวจวัดรายชั่วโมง สืบค้นเว็บไซต์กระจายข่าวทางอุตุนิยมวิทยา <http://www.ogimet.com> โดยข้อมูลที่ได้จากเว็บไซต์ถูกจัดเก็บในรูปแบบของไฟล์ .txt รายละเอียดการสืบค้นข้อมูลดังภาพที่ 3.1 นำเข้าข้อมูลแยกตามสนามบินในรูปแบบรหัสข้อความจากการสืบค้น เพื่อจัดการข้อมูลในขั้นตอนต่อไปด้วย Microsoft Excel ดังภาพที่ 3.2

Professional information about meteorological conditions in the world

OGIMET

Metar/Speci/Taf reports selection query

ICAO INDEXES	TYPE	SORT ORDER	NIL REPORTS	FORMAT
VTCC	SA	Oldest first	NIL report excluded	TXT

TIME INTERVAL

Year	Month	Day	Hour
BEGIN: 2015	January	01	00
END: 2022	December	31	23

send Reset

You have to set:

- The ICAO indexes from desired stations, with a comma or space separating indexes. If you don't know the index, you can visit [this page](#)
- The type of report you want to get
 - ALL It will show METAR, SPECI and TAF reports
 - SA METAR and SPECI.
 - SP only SPECI.
 - FC Only short TAF reports (validity 9 Hours).
 - FT Only large TAF reports (18 or 24 Hours).
- Order of displayed results. You can select chronological or reverse order
- You also can decide whether to get "NIL" void reports
- Output format
 - HTML Push HTML output format

Professional information about meteorological conditions in the world

OGIMET

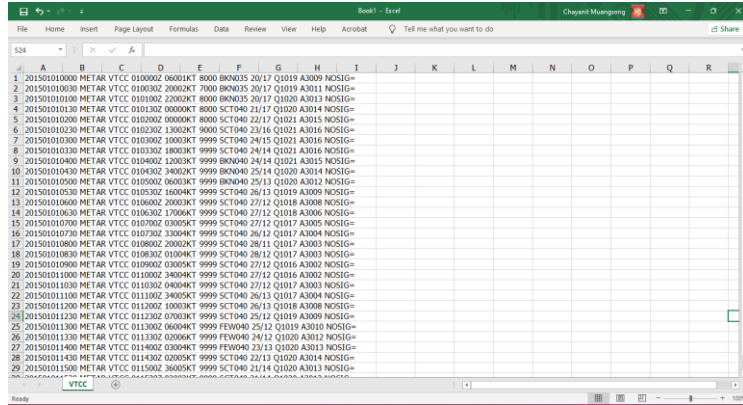
Version española

METEOSAT
[Latest meteosat / metar](#)
[Meteosat-9 metar](#)
[Meteosat-9 VIS-IR loop](#)

WEATHER MODEL FORECAST
[GFS world weather maps](#)

METEOGRAMS
[Grimet asco](#)
[Grimet meteo](#)

ภาพที่ 3.1 การสืบค้นข้อมูลข่าว METAR จากเว็บไซต์กระจายข่าวทางอุตุนิยมวิทยา OGIMET



ภาพที่ 3.2 นำเข้าข้อมูลใน Microsoft Excel

2) จำนวนชุดข้อมูลข่าวอากาศการบิน ขึ้นอยู่กับเวลาทำการและช่วงเวลาตรวจอากาศแต่ละสถานี เวลาทำการของแต่ละสถานีตรวจอากาศการบินแสดงดังตารางที่ 3.1 โดยทำอากาศยานนานาชาติเชียงใหม่และทำอากาศยานแม่ฟ้าหลวงเชียงรายมีจำนวนข่าวอากาศการบินมากกว่าสถานีอื่นๆ เนื่องจากเปิดทำการและส่งข่าวตรวจอากาศการบินตลอด 24 ชั่วโมง

ตารางที่ 3.1 เวลาทำการของสถานีตรวจอากาศการบิน

สถานีตรวจอากาศการบิน	เวลาทำการ	ช่วงเวลาการตรวจ
1. ทำอากาศยานนานาชาติเชียงใหม่	ตลอด 24 ชั่วโมง	30 นาที
2. ทำอากาศยานแม่ฮ่องสอน	05.00 – 18.00 น.	1 ชั่วโมง
3. ทำอากาศยานแม่ฟ้าหลวงเชียงราย	ตลอด 24 ชั่วโมง	1 ชั่วโมง
4. ทำอากาศยานลำปาง	06.00 – 20.00 น.	1 ชั่วโมง
5. ทำอากาศยานน่านนคร	05.00 – 20.00 น.	1 ชั่วโมง
6. ทำอากาศยานแพร่	05.00 – 18.00 น.	1 ชั่วโมง
7. ทำอากาศยานพิษณุโลก	06.00 – 22.00 น.	1 ชั่วโมง
8. ทำอากาศยานสุโขทัย	05.00 – 18.00 น.	1 ชั่วโมง
9. ทำอากาศยานนานาชาติแม่สอด	05.00 – 18.00 น.	1 ชั่วโมง

ในงานวิจัยครั้งนี้เลือกพิจารณาข้อมูลนำเข้า (Input) เฉพาะผลการตรวจอากาศ ทุก 1 ชั่วโมงเท่านั้น ดังนั้นข่าวอากาศการบินของท่าอากาศยานนานาชาติเชียงใหม่จะมีข้อมูลนำเข้าที่ลดลงจากข้อมูลข่าวทั้งหมดที่ได้ทำการตรวจวัด และพิจารณาตัดชุดข้อมูลที่เกิดความผิดพลาดในการรายงานข่าวออกไป ซึ่งในเนื้อหาข่าวปรากฏสัญลักษณ์พิเศษอื่นๆนอกเหนือจากการรายงานข่าวปกติ เช่น #, ///, \$, ---, AAA เป็นต้น ซึ่งสัญลักษณ์ดังกล่าวทำให้ไม่สามารถระบุค่าพารามิเตอร์ต่างๆทางอุตุนิยมวิทยาที่ตรวจวัดได้ ดังนั้นจำนวนชุดข้อมูลข่าวอากาศการบินทั้งหมดแยกตามสถานีตรวจอากาศการบินแสดงดังตารางที่ 3.2

ตารางที่ 3.2 จำนวนชุดข้อมูลข่าวอากาศการบินทั้งหมดแยกตามสถานีตรวจอากาศการบิน

สถานีตรวจอากาศการบิน	ชุดข้อมูล METAR	ชุดข้อมูลนำเข้าจริง
1. ท่าอากาศยานนานาชาติเชียงใหม่	138,242	69,071
2. ท่าอากาศยานแม่ฮ่องสอน	30,930	28,222
3. ท่าอากาศยานแม่ฟ้าหลวงเชียงราย	64,830	63,293
4. ท่าอากาศยานลำปาง	35,670	35,528
5. ท่าอากาศยานน่านนคร	36,557	36,528
6. ท่าอากาศยานแพร่	39,325	28,035
7. ท่าอากาศยานพิษณุโลก	43,395	43,350
8. ท่าอากาศยานสุโขทัย	32,968	32,947
9. ท่าอากาศยานนานาชาติแม่สอด	27,414	27,409

2.1.3 การจัดเก็บข้อมูลในรูปแบบตาราง

ข้อมูลที่ได้มาจะรวบรวมอยู่ในรูปแบบรหัสข้อความ ซึ่งในขั้นตอนการเตรียมข้อมูลก่อนสร้างแบบจำลองได้จัดการข้อมูลใน Microsoft Excel ดังนี้

1) แปลงข้อมูล เป็นรหัสข้อความเป็นตารางข้อมูลโดยแยกออกมาเป็นแต่ละคอลัมน์ของค่าพารามิเตอร์ทางอุตุนิยมวิทยาที่ตรวจวัดได้ทุก 1 ชั่วโมง จากนั้นทำการพิจารณาพารามิเตอร์ที่จะนำไปศึกษาต่อไป ซึ่งจากส่วนประกอบของข่าวอากาศการบินทั้ง 13 ส่วน ผู้วิจัยได้ตัดข้อมูลส่วนประกอบที่ไม่สำคัญ (ส่วนที่ 1 และ 3) และส่วนประกอบที่ต้องใช้เครื่องมือพิเศษสำหรับการ

ตรวจวัดเท่านั้น (ส่วนที่ 7 และ 12) ซึ่งสถานีตรวจอากาศการบินของภาคเหนือส่วนใหญ่ไม่มีเครื่องมือ นั้น ดังนั้นจึงมีส่วนประกอบของข่าวอากาศการบินจำนวน 7 ส่วนประกอบที่นำไปศึกษาต่อไป ได้แก่ ส่วนที่ 2 รหัสท่าอากาศยาน ส่วนที่ 5 ลมผิวพื้น ส่วนที่ 6 ทิศนวิสัย ส่วนที่ 8 สภาพอากาศปัจจุบัน ส่วนที่ 9 เมฆ ส่วนที่ 10 อุณหภูมิอากาศและอุณหภูมิจุดน้ำค้าง และส่วนที่ 11 ความกดอากาศ

2) ตัวอย่างตารางข้อมูล ชุดข้อมูลที่ได้อีกก่อนเข้าสู่ขั้นตอนการเตรียมข้อมูลก่อนสร้างแบบจำลอง ตัวอย่างตารางข้อมูลสำหรับชุดข้อมูลของสนามบินนานาชาติเชียงใหม่ แสดงดังภาพที่ 3.3

ส่วนที่ 2	ส่วนที่ 5	ส่วนที่ 6	ส่วนที่ 8	ส่วนที่ 9	ส่วนที่ 10	ส่วนที่ 11	
VTCC	21004KT	8000	-RA	BKN035	26	18	Q1009
VTCC	27006KT	4000	RA	BKN035	23	20	Q1010
VTCC	24005KT	6000	-RA	BKN035	23	20	Q1011
VTCC	23004KT	7000	BR	BKN035	23	20	Q1010
VTCC	20005KT	3000	RA	BKN035	23	20	Q1011
VTCC	21006KT	3000	RA	BKN035	22	20	Q1011
VTCC	22003KT	6000	BR	BKN035	22	20	Q1012
VTCC	15002KT	2000	RA	BKN030	22	20	Q1013
VTCC	16004KT	3000	RA	BKN030	22	20	Q1014
VTCC	16004KT	3000	RA	BKN030	22	20	Q1014
VTCC	16004KT	3000	RA	BKN030	22	20	Q1015
VTCC	18004KT	3000	RA	BKN030	22	20	Q1014
VTCC	20005KT	7000	RA	BKN030	21	19	Q1013
VTCC	20005KT	7000	RA	BKN030	20	19	Q1012
VTCC	20005KT	2000	RA	BKN030	21	19	Q1011
VTCC	22005KT	6000	RA	BKN030	20	19	Q1012

ภาพที่ 3.3 ตัวอย่างตารางข้อมูลสำหรับชุดข้อมูลของสนามบินนานาชาติเชียงใหม่

2.2 การเตรียมข้อมูล

ก่อนนำชุดข้อมูลทั้งหมดไปสร้างแบบจำลองต้องมีการเตรียมข้อมูลก่อน เพื่อให้ชุดข้อมูลของแต่ละสนามบินมีรูปแบบของข้อมูลที่เหมือนกัน ขั้นตอนดังนี้

2.2.1 การทำความสะอาดข้อมูล เป็นกระบวนการที่เกี่ยวข้องกับการตรวจสอบและแก้ไขข้อมูล เพื่อให้ชุดข้อมูลอยู่ในรูปแบบที่ถูกต้องและสมบูรณ์ โดยเมื่อจัดการรวมข้อมูลเรียบร้อยแล้ว ตารางรวบรวมข้อมูลจะมีบางส่วนที่เกิดการสูญหายของข้อมูล (Missing Value) หรือมีค่าของข้อมูลที่มีความผิดปกติหรือแตกต่างไปจากข้อมูลในกลุ่ม จึงตรวจสอบข้อมูลอีกครั้งเพราะอาจเกิดจากความผิดพลาดในขั้นตอนการรวบรวมข้อมูล หรือความบกพร่องของอุปกรณ์การตรวจที่ทำให้

แสดงผลการตรวจไม่ครบถ้วนสมบูรณ์ ซึ่งพบว่า ค่าที่สูญหายเกิดจากความผิดพลาดในการเปลี่ยนข้อความรหัสของ METAR จึงแทนที่ข้อมูลที่สูญหายหรือข้อมูลที่มีค่าที่ผิดปกติด้วยค่าที่ถูกต้อง แต่ยังมีชุดข้อมูลบางส่วนที่มีพารามิเตอร์ทางอุตุนิยมวิทยาไม่ครบถ้วน พบว่าในวันที่และเวลานั้น ๆ อุปกรณ์เกิดความบกพร่องจึงไม่สามารถวัดค่าได้ตามปกติ จึงได้ตัดชุดข้อมูลนั้นทิ้งไปและเหลือจำนวน ข้อมูล นำเข้าเพื่อศึกษาต่อดังแสดงในตารางที่ 3.3

ตารางที่ 3.3 จำนวนชุดข้อมูลนำเข้าทั้งหมดแบ่งตามกลุ่มข้อมูล

กลุ่มข้อมูล	จำนวนชุดข้อมูล
รวมสนามบินภาคเหนือ	364,382
ภาคเหนือตอนบน	260,676
ภาคเหนือตอนล่าง	103,706
ภูเขาล้อมรอบ 1-2 ด้าน	76,297
ภูเขาล้อมรอบ 3-4 ด้าน	288,085

2.2.2 การผสานข้อมูล เป็นขั้นตอนรวมชุดข้อมูลที่มาจากแต่ละสถานีตรวจอากาศ การบินให้เป็นข้อมูลชุดเดียวกัน

2.2.3 การแปลงข้อมูล เป็นอีกขั้นตอนหนึ่งที่สำคัญในกระบวนการเตรียมข้อมูลเพื่อนำไปสร้างแบบจำลอง โดยข้อมูลข่าวอากาศการบินประกอบไปด้วยเนื้อหา 13 ส่วน ในงานวิจัยนี้มีรายละเอียดการจัดการข้อมูลแต่ละส่วน ดังนี้

1) ส่วนที่ 1 ประเภทของรายงาน ไม่ได้นำเข้าเป็นแอตทริบิวต์เพื่อสร้างแบบจำลอง เนื่องจากมีประเภทของรายงานเหมือนกันในทุกชุดข้อมูล

2) ส่วนที่ 2 รหัสท่าอากาศยาน ใช้สำหรับแบ่งกลุ่มข้อมูลเป็น 5 กลุ่ม ได้แก่ กลุ่มสนามบินรวมภาคเหนือ กลุ่มสนามบินภาคเหนือตอนบน กลุ่มสนามบินภาคเหนือตอนล่าง กลุ่มสนามบินที่มีภูเขาล้อมรอบ 1-2 ด้าน และกลุ่มสนามบินที่มีภูเขาล้อมรอบ 3-4 ด้าน

3) ส่วนที่ 3 เวลาตรวจอากาศการบิน แบ่งออกเป็น 2 ส่วนคือ เดือน และ ช่วงเวลาของวัน

4) ส่วนที่ 4 การตรวจอากาศการบินด้วยระบบอัตโนมัติ ไม่ได้นำเข้าเป็นแอตทริบิวต์เพื่อสร้างแบบจำลอง เนื่องจากมีประเภทของรายงานเหมือนกันในทุกชุดข้อมูล

- 5) ส่วนที่ 5 ลมผิวพื้น แบ่งพารามิเตอร์ของการตรวจวัดออกเป็น 3 ส่วน คือ ทิศทางลม ความเร็วลม และการเกิดลมกระโชก
 - 6) ส่วนที่ 6 ทิศนวิสัย นำเข้าเป็นแอตทริบิวต์เพื่อสร้างแบบจำลอง
 - 7) ส่วนที่ 7 พิสัยการมองเห็นบนทางวิ่ง ไม่ได้นำเข้าเป็นแอตทริบิวต์เพื่อสร้างแบบจำลอง เนื่องจากสนามบินบริเวณภาคเหนือส่วนใหญ่ไม่ได้มีการรายงานค่านี้เนื่องจากต้องใช้เครื่องมือเฉพาะ และจะรายงานก็ต่อเมื่อค่าทัศนวิสัยบนทางวิ่งต่ำกว่า 1,500 เมตรเท่านั้น
 - 8) ส่วน 8 สภาพอากาศปัจจุบัน นำเข้าเป็นแอตทริบิวต์เพื่อสร้างแบบจำลอง อีกทั้งใช้สภาพอากาศปัจจุบันที่เกิดขึ้นในช่วงถัดไป เป็นตัวกำหนด Label ของแบบจำลอง
 - 9) ส่วนที่ 9 เมฆ พิจารณาเฉพาะชนิดของเมฆที่ก่อตัวทางตั้งที่ทำให้เกิดพายุฝนฟ้าคะนองได้ ได้แก่ เมฆชนิด Cumulonimbus (CB) และ Towering Cumulus (TCU) ซึ่งเป็นเมฆระยะเริ่มต้นก่อนพัฒนาเป็นเมฆชนิด CB
 - 10) ส่วนที่ 10 อุณหภูมิอากาศและอุณหภูมิจุดน้ำค้าง นำไปคำนวณให้เป็นค่าความชื้นสัมพัทธ์ (Humidity)
 - 11) ส่วนที่ 11 ความกดอากาศ นำเข้าเป็นแอตทริบิวต์เพื่อสร้างแบบจำลอง
 - 12) ส่วนที่ 12 ขาวสารเพิ่มเติม ไม่ได้นำเข้าเป็นแอตทริบิวต์เพื่อสร้างแบบจำลอง
 - 13) ส่วนที่ 13 การพยากรณ์แนวโน้มสภาวะอากาศบริเวณสนามบิน ไม่ได้นำเข้าเป็นแอตทริบิวต์เพื่อสร้างแบบจำลอง แต่จะใช้เป็นข้อมูลเพื่อทดสอบแบบจำลองในขั้นตอนสุดท้าย โดยจะทำการเปรียบเทียบการพยากรณ์ของแบบจำลองกับการพยากรณ์ด้วยผู้ปฏิบัติงานจริง
- ผู้วิจัยได้ทำการแปลงชุดข้อมูล จากตัวเลขให้อยู่ในรูปแบบของช่วงข้อมูลเพื่อลดความซับซ้อนของข้อมูล มีรายละเอียดการแปลงข้อมูลดังตารางที่ 3.4

ตารางที่ 3.4 การแปลงข้อมูลที่เป็นตัวเลขให้อยู่ในรูปแบบของช่วงข้อมูล

แอตทริบิวต์	ความหมายของค่าข้อมูล		
1. Month	Jan ถึง Dec	เดือนมกราคม ถึง ธันวาคม	
2. Time_of_day	Morning	เวลา 06.00-12.00 น.	
	Afternoon	เวลา 13.00-18.00 น.	
	Evening	เวลา 19.00-24.00 น.	
	Night	เวลา 01.00-05.00 น.	
3. W_direct	CLAM	ลมสงบ	ความเร็วลม 0 KT
	N	ทิศเหนือ	337.5 ถึง 22.4 องศา
	NE	ทิศตะวันออกเฉียงเหนือ	22.5 ถึง 67.4 องศา
	E	ทิศตะวันออก	67.5 ถึง 112.4 องศา
	SE	ทิศตะวันออกเฉียงใต้	112.5 ถึง 157.4 องศา
	S	ทิศใต้	157.5 ถึง 202.4 องศา
	SW	ทิศตะวันตกเฉียงใต้	202.5 ถึง 247.4 องศา
	W	ทิศตะวันตก	247.5 ถึง 292.4 องศา
	NW	ทิศตะวันตกเฉียงเหนือ	292.5 ถึง 337.4 องศา
4. W_speed	CLAM	ลมสงบ	ความเร็วลม 0 KT
	LIGHT	ลมอ่อน	ความเร็วลม 1-10 KT
	MODERATE	ลมปานกลาง	ความเร็วลม 11-16 KT
	STRONG	ลมแรง	ความเร็วลม 17 KT ขึ้นไป
5. GUST	YES	มีทิศทางและความเร็วของลมกระโชก	
	NO	ไม่มีทิศทางและความเร็วของลมกระโชก	
6. Visibility	A	ทัศนวิสัยมากกว่า 5000 เมตร	
	B	ทัศนวิสัย 3000-5000 เมตร	
	C	ทัศนวิสัยต่ำกว่า 3000 เมตร	

ตารางที่ 3.4 (ต่อ)

แอตทริบิวต์		ความหมายของค่าข้อมูล
7. WW	NO	ไม่มีปรากฏการณ์ฝนตก
	Light Rain	เกิดฝนขนาดเล็กเบา (-DZ, DZ, +DZ, -RA, -SHRA)
	Moderate Rain	เกิดฝนขนาดปานกลาง (RA, SHRA)
	Heavy Rain	เกิดฝนหนัก (+RA, +SHRA)
	TSRA	เกิดพายุฝนฟ้าคะนอง (TS, -TSRA, TSRA, +TSRA)
8. M_Cloud	NO	ไม่มีเมฆที่ทำให้เกิดฝนฟ้าคะนอง
	TCU	มีเมฆฝนฟ้าคะนองที่ก่อตัวทางตั้งในชั้นเริ่มก่อตัว
	CB	มีเมฆที่ก่อตัวทางตั้งในชั้นทำให้เกิดฝนฟ้าคะนอง
9. Humidity	Dry	ความชื้นสัมพัทธ์น้อยกว่า 80%
	High Moisture	ความชื้นสัมพัทธ์มากกว่าหรือเท่ากับ 80%
10. QNH	Low	ความกดอากาศน้อยกว่า 1013 มิลลิบาร์
	MSL	ความกดอากาศเท่ากับ 1013 มิลลิบาร์
	High	ความกดอากาศมากกว่า 1013 มิลลิบาร์

2.2.4 แอตทริบิวต์ที่นำเข้าสู่ขั้นตอนการสร้างแบบจำลอง จากการทำความสะอาดข้อมูลในขั้นตอนที่ผ่านมา ผู้วิจัยสรุปผลการคัดเลือกได้จำนวนคุณลักษณะหรือแอตทริบิวต์ (Attribute) จำนวน 8 แอตทริบิวต์เพื่อนำเข้าสู่ขั้นตอนของการสร้างแบบจำลอง ดังตารางที่ 3.5

ตารางที่ 3.5 รายละเอียดแอตทริบิวต์ของข้อมูลที่ใช้สำหรับทำการวิจัย

แอตทริบิวต์	คำอธิบาย	ประเภทข้อมูล	ค่าข้อมูล
1. Month	เดือน	Polynomial	Jan/Feb/Mar/Apr/May/June/ Jul/Aug/Sep/Oct/Nov/Dec
2. Time_of_day	ช่วงเวลาของวัน	Polynomial	Morning Afternoon Evening Night
3. W_direct	ทิศทางลม	Polynomial	CLAM VRB N NE E SE S SW W NW
4. W_speed	ความเร็วลม	Polynomial	CLAM LIGHT BREEZE MODERATE BREEZE STRONG BREEZE
5. GUST	ลมกระโชก	Binomial	YES/NO
6. Visibility	ทัศนวิสัย	Polynomial	A/B/C
7. WW	สภาพอากาศ ปัจจุบัน	Polynomial	NO Light rain Moderate Rain Heavy Rain TSRA VC Rain Obscurations

ตารางที่ 3.5 (ต่อ)

แอตทริบิวต์	คำอธิบาย	ประเภทข้อมูล	ค่าข้อมูล
8. M_Cloud	เมฆฝนฟ้าคะนอง	Polynomial	NO CB TCU
9. Humidity	ความชื้นสัมพัทธ์	Binomial	Dry High Moisture
10. QNH	ความกดอากาศ	Polynomial	Low/MSL/High

2.2.5 การแบ่งคลาสเพื่อจำแนก ในงานวิจัยนี้ จำแนกประเภทของลักษณะเหตุการณ์เพื่อสร้างแบบจำลองออกเป็น 2 คลาส โดยมีนิยามดังนี้

1) ไม่เกิดพายุฝนฟ้าคะนอง (NO TSRA Class) หมายถึง การไม่มีสภาพอากาศสำคัญ ฝนละออง (Drizzle) ฝนธรรมดา (Rain) ฝนชุก (Shower) การเกิดฝนระยะไกลในระยะ 8-16 กิโลเมตร (VCSH) ฟ้าหลัวชื้น (Damp haze) ฟ้าหลัวแห้ง (Haze) หมอก (Fog) และควันไฟ (Smoke)

2) เกิดพายุฝนฟ้าคะนอง (TSRA Class) หมายถึง เหตุการณ์ทางสภาพอากาศ ได้แก่ การเกิดฟ้าร้อง (TS) และการเกิดพายุฝนฟ้าคะนองขนาดเบา ปานกลาง และหนัก

2.3 การแบ่งกลุ่มข้อมูล

ทำการแบ่งออกเป็น 5 กลุ่มข้อมูลโดยวิเคราะห์ตามขอบเขตและลักษณะพื้นที่ของสนามบินภาคเหนือของประเทศไทย รวมถึงตามเกณฑ์การแบ่งพื้นที่ของกรมอุตุนิยมวิทยา ดังนี้

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

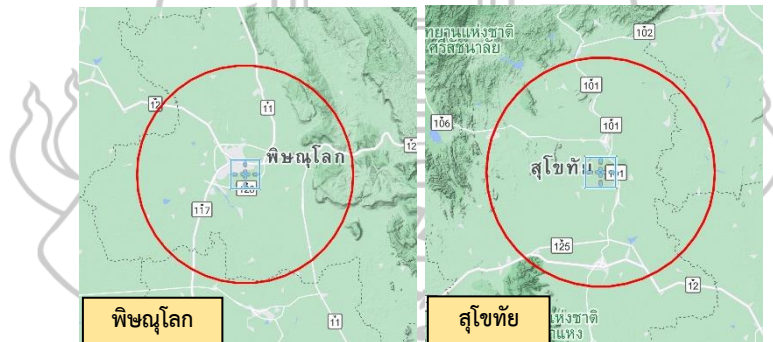
2.3.2 กลุ่มข้อมูลที่ 2 สนามบินภาคเหนือตอนบน ได้แก่ ทำอากาศยานแม่ฮ่องสอน ทำอากาศยานนานาชาติเชียงใหม่ ทำอากาศยานแม่ฟ้าหลวงเชียงราย ทำอากาศยานลำปาง ทำอากาศยานแพร่ และทำอากาศยานน่านนคร

2.3.3 กลุ่มข้อมูลที่ 3 สนามบินภาคเหนือตอนล่าง ได้แก่ ทำอากาศยานนานาชาติแม่สอด ทำอากาศยานสุโขทัย และทำอากาศยานพิษณุโลก

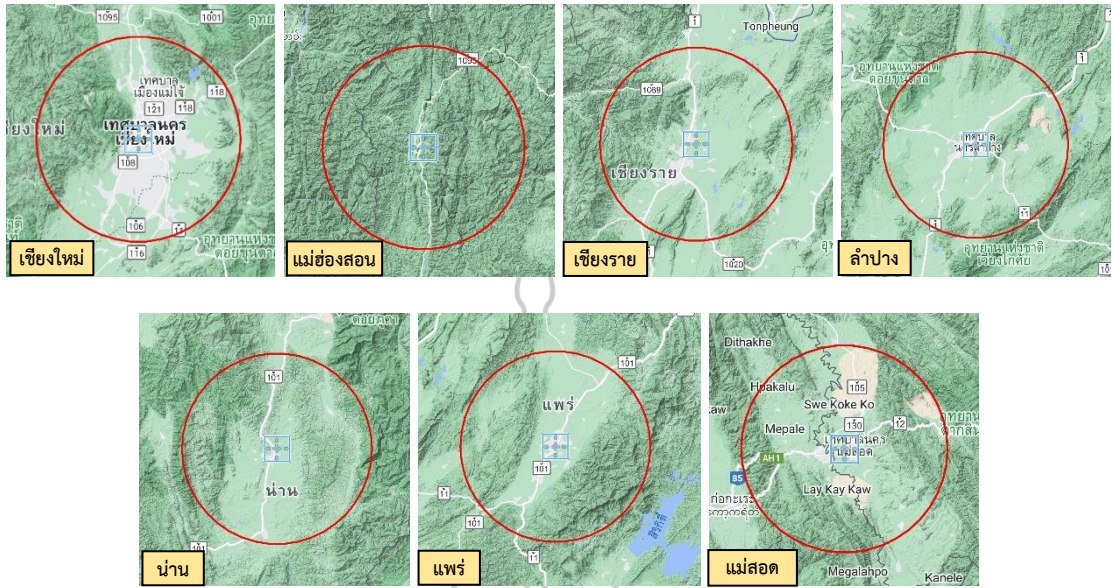
2.3.4 กลุ่มข้อมูลที่ 4 สนามบินที่มีภูเขาล้อมรอบ 1-2 ด้าน ได้แก่ ทำอากาศยานสุโขทัย และทำอากาศยานพิษณุโลก

2.3.5 กลุ่มข้อมูลที่ 5 สนามบินที่มีภูเขาล้อมรอบ 3-4 ด้าน ได้แก่ ทำอากาศยานแม่ฮ่องสอน ทำอากาศยานนานาชาติเชียงใหม่ ทำอากาศยานแม่ฟ้าหลวงเชียงราย ทำอากาศยานลำปาง ทำอากาศยานแพร่ ทำอากาศยานน่านนคร และทำอากาศยานนานาชาติแม่สอด

โดยข้อมูลกลุ่มที่ 4 และ 5 พิจารณาตามลักษณะภูมิประเทศตำแหน่งที่ตั้งของภูเขาล้อมรอบจำนวน 4 ด้าน จากภาพที่ 3.4 และภาพที่ 3.5 เส้นวงกลมสีแดงแสดงเขตรัศมี 30 กิโลเมตร จากจุดอ้างอิงสนามบิน ซึ่งในการพิจารณาเพื่อแบ่งกลุ่มข้อมูลนั้นทำได้โดยการแบ่งวงกลมออกเป็น 4 ส่วนและคัดเลือกตามแผนที่ลักษณะภูมิประเทศของแต่ละสนามบินว่าควรจัดอยู่ในกลุ่มข้อมูลสนามบินที่มีภูเขาล้อมรอบ 1-2 ด้าน หรือควรจัดอยู่ในกลุ่มข้อมูลสนามบินที่มีภูเขาล้อมรอบ 3-4 ด้าน



ภาพที่ 3.4 ลักษณะภูมิประเทศของสนามบินที่มีภูเขาล้อมรอบ 1-2 ด้าน



ภาพที่ 3.5 ลักษณะภูมิประเทศของสนามบินที่มีภูเขาล้อมรอบ 3-4 ด้าน

2.4 การแบ่งชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบ

ในขั้นตอนนี้ชุดข้อมูลในงานวิจัยทั้ง 5 กลุ่มข้อมูลจะถูกแบ่งออกเป็น 2 ส่วนเพื่อนำเข้าสู่แบบจำลอง คือ (1) ชุดข้อมูลฝึกฝน (Training Data Set) ใช้ข้อมูลในปีพ.ศ. 2558-2563 รวม 6 ปี เพื่อใช้เรียนรู้และปรับค่าพารามิเตอร์ที่เหมาะสม และ (2) ชุดข้อมูลทดสอบ (Testing Data Set) ใช้ข้อมูลในปีพ.ศ.2564-2565 รวม 2 ปี เพื่อทดสอบว่าสามารถทำงานได้ดีกับชุดข้อมูลที่ไม่เคยพบมาก่อนหรือไม่ จำนวนข้อมูลที่น่าใช้ในการสร้างแบบจำลองแบ่งตามระยะเวลาพยากรณ์ 1 และ 2 ชั่วโมงข้างหน้า และแบ่งแยกตามคลาสของการจำแนก ดังตารางที่ 3.6 และ ตารางที่ 3.7

ตารางที่ 3.6 จำนวนข้อมูลที่น่าใช้ในการสร้างแบบจำลองสำหรับระยะเวลาพยากรณ์ 1 ชั่วโมง

กลุ่มข้อมูล	จำนวนชุดข้อมูล	Training Data Set		Testing Data Set	
		NO TSRA	TSRA	NO TSRA	TSRA
รวมสนามบินภาคเหนือ	364,382	258,773	2,023	102,547	1,039
ภาคเหนือตอนบน	260,676	185,667	1,531	72,653	825
ภาคเหนือตอนล่าง	103,706	73,106	492	29,894	214
ภูเขาล้อมรอบ 1-2 ด้าน	76,297	55,348	437	20,334	178
ภูเขาล้อมรอบ 3-4 ด้าน	288,085	203,425	1,586	82,213	861

ตารางที่ 3.7 จำนวนข้อมูลที่น่าใช้ในการสร้างแบบจำลองสำหรับระยะเวลาพยากรณ์ 2 ชั่วโมง

กลุ่มข้อมูล	จำนวนชุดข้อมูล	Training Data Set		Testing Data Set	
		NO TSRA	TSRA	NO TSRA	TSRA
รวมสนามบินภาคเหนือ	364,382	258,782	2,014	102,556	1,030
ภาคเหนือตอนบน	260,676	185,659	1,539	72,650	828
ภาคเหนือตอนล่าง	103,706	73,123	475	29,906	202
ภูเขาล้อมรอบ 1-2 ด้าน	76,297	55,363	422	20,346	166
ภูเขาล้อมรอบ 3-4 ด้าน	288,085	203,419	1,592	82,210	864

3. การสร้างแบบจำลอง (Process)

3.1 การคัดเลือกอัลกอริทึม

สำหรับการสร้างแบบจำลองในงานวิจัยนี้ เลือกใช้อัลกอริทึมประเภทจำแนกข้อมูล (Classification) โดยชุดข้อมูลมีคลาสชนิด Binary ที่มีประเภทหรือกลุ่มของคลาสเพียงสองสถานะเท่านั้น นั่นคือ ไม่เกิดฝนฟ้าคะนอง (NO TSRA) และเกิดฝนฟ้าคะนอง (TSRA) ส่วนแอตทริบิวต์ของชุดข้อมูลเป็นประเภท Polynomial และ Binomial ซึ่งผู้วิจัยได้ทดสอบเบื้องต้นเพื่อคัดเลือกอัลกอริทึมมาตรฐานที่จะนำมาใช้ ได้แก่ Naive Bayes Decision Tree Neural Networks KNN และ SVM โดยการสร้างแบบจำลองจากอัลกอริทึมแต่ละชนิดที่สามารถนำมาใช้กับประเภทของข้อมูลในงานวิจัยนี้ได้ นำเข้าชุดข้อมูลรวมสนามบินภาคเหนือ พ.ศ.2558 ระยะเวลา 1 ปี สำหรับระยะเวลาพยากรณ์ 1 ชั่วโมงข้างหน้า และประเมินประสิทธิภาพด้วยค่า Accuracy Precision Recall และ F-measure ซึ่งผู้วิจัยคัดเลือกอัลกอริทึมที่เหมาะสมกับลักษณะข้อมูลของการวิจัยนี้ โดยพิจารณาจากค่า F-measure สูงสุดของการสร้างแบบจำลองเบื้องต้น ซึ่งสามารถบ่งชี้ถึงความสามารถในการจำแนกคลาสของการเกิดฝนฟ้าคะนองได้ดี ผลการทดสอบเบื้องต้นเพื่อคัดเลือกอัลกอริทึมที่เหมาะสมในงานวิจัย ดังตารางที่ 3.8 และประกอบกับพิจารณาข้อจำกัดของแต่ละอัลกอริทึมเช่นกัน ดังตารางที่ 3.9

ตารางที่ 3.8 การทดสอบเบื้องต้นเพื่อคัดเลือกอัลกอริทึมที่เหมาะสมในงานวิจัย

อัลกอริทึม	Accuracy	Precision	Recall	F-measure
Neural Networks	99.61	99.19	64.27	78.00
Decision Tree	99.57	98.04	61.47	75.57
Naive Bayes	98.79	45.58	64.10	53.28
KNN	99.18	100.00	24.34	39.15
SVM	98.92	100.00	18.00	35.00

ตารางที่ 3.9 การพิจารณาข้อจำกัดของแต่ละอัลกอริทึม

อัลกอริทึม	ข้อจำกัด
Naive Bayes	<ul style="list-style-type: none"> - ตั้งสมมติฐานว่าแอตทริบิวต์ข้อมูลเป็นอิสระไม่ขึ้นต่อกัน ซึ่งอาจไม่ตรงกับความเป็นจริงในบางกรณี - ไม่มีการจัดการกับข้อมูลที่ขาดหาย ต้องปรับปรุงข้อมูลให้เสร็จก่อนเท่านั้น
Decision Tree	<ul style="list-style-type: none"> - มีความยุ่งยากในการกำหนดค่าพารามิเตอร์ที่เหมาะสม - ประสิทธิภาพลดลงเมื่อใช้กับข้อมูลที่มีความซับซ้อนมาก - ไม่มีการจัดการกับข้อมูลที่ขาดหาย ต้องปรับปรุงข้อมูลให้เสร็จก่อนนำมาเข้าแบบจำลอง - มีโอกาสเกิด Overfitting กับชุดข้อมูลใหม่
Neural Networks	<ul style="list-style-type: none"> - มีจำนวนพารามิเตอร์ที่ต้องฝึกฝนจำนวนมาก ทำให้ใช้เวลาในการประมวลผลนาน - การกำหนดค่าพารามิเตอร์ไม่มีสูตรตายตัวในการเลือกวิธีเชื่อมต่อ layer ใช้เวลานานในการหาค่าพารามิเตอร์ที่เหมาะสม

ตารางที่ 3.9 (ต่อ)

อัลกอริทึม	ข้อจำกัด
K- Nearest Neighbor (KNN)	<ul style="list-style-type: none"> - ใช้ระยะเวลาในการประมวลผลนาน เนื่องจากคำนวณระยะห่างระหว่างข้อมูล - มีความยุ่งยากในการกำหนดค่าพารามิเตอร์ K ที่เหมาะสม - อาจเกิดความผิดพลาดเมื่อมีลักษณะของข้อมูลที่ไม่สมดุล
Support Vector Machines (SVM)	<ul style="list-style-type: none"> - มีความยุ่งยากในการกำหนดค่าพารามิเตอร์ที่เหมาะสม - เหมาะกับข้อมูลที่มีแอตทริบิวต์จำนวนมากแต่มีปริมาณข้อมูลน้อยถึงปานกลาง หากใช้งานกับชุดข้อมูลที่มีขนาดใหญ่ เวลาที่ใช้ในการฝึกจะเพิ่มขึ้น - มีความสามารถในการจัดกลุ่มข้อมูลที่เป็นเชิงเส้นได้ดี แต่อาจมีปัญหาในการจัดกลุ่มข้อมูลที่มีความซับซ้อน - มีความแม่นยำน้อยลงเมื่อชุดข้อมูลเป็นแบบหลายคลาส

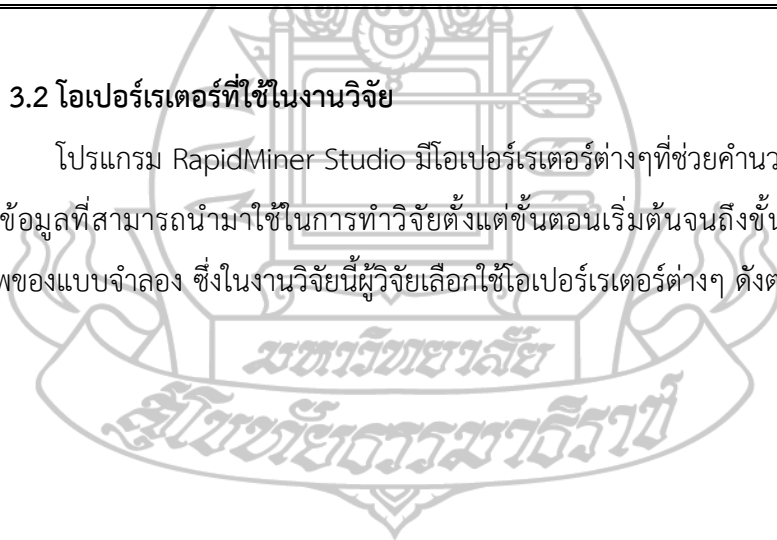
โดยผลจากการคัดเลือกอัลกอริทึม พบว่า อัลกอริทึม Naive Bayes Decision Tree และ Neural Networks มีความเหมาะสมกับข้อมูลในงานวิจัยนี้ โดยแบ่งการศึกษาเป็น 2 ส่วนคือ ส่วนที่ 1 การสร้างแบบจำลองโดยอัลกอริทึมมาตรฐาน ได้แก่ Naive Bayes Decision Tree และ Neural Networks และส่วนที่ 2 การสร้างแบบจำลองโดยอัลกอริทึมมาตรฐานร่วมกับอัลกอริทึมการเรียนรู้แบบกลุ่ม ได้แก่ Bagging AdaBoost และ Random Forest ซึ่งแบบจำลองทั้งหมดในงานวิจัย ดังตารางที่ 3.10

ตารางที่ 3.10 แสดงแบบจำลองทั้งหมดในงานวิจัย

แบบจำลอง	ตัวย่อในงานวิจัย	ประเภท
Naive Bayes	NB	อัลกอริทึมมาตรฐาน
Decision Tree	DT	อัลกอริทึมมาตรฐาน
Neural Networks	NN	อัลกอริทึมมาตรฐาน
Random Forest	RF	อัลกอริทึมการเรียนรู้แบบกลุ่ม
Naive Bayes ร่วมกับ Bagging	NA+Bag	อัลกอริทึมการเรียนรู้แบบกลุ่ม
Decision Tree ร่วมกับ Bagging	DT+Bag	อัลกอริทึมการเรียนรู้แบบกลุ่ม
Neural Networks ร่วมกับ Bagging	NN+Bag	อัลกอริทึมการเรียนรู้แบบกลุ่ม
Naive Bayes ร่วมกับ AdaBoost	Na+Ada	อัลกอริทึมการเรียนรู้แบบกลุ่ม
Decision Tree ร่วมกับ AdaBoost	DT+Ada	อัลกอริทึมการเรียนรู้แบบกลุ่ม
Neural Networks ร่วมกับ AdaBoost	NN + Ada	อัลกอริทึมการเรียนรู้แบบกลุ่ม

3.2 โอเพอร์เรเตอร์ที่ใช้ในงานวิจัย

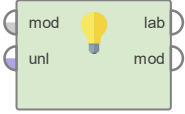
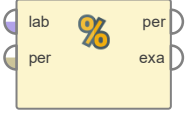
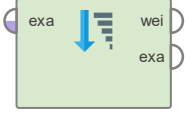
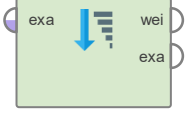
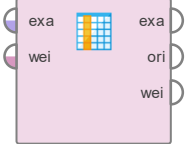
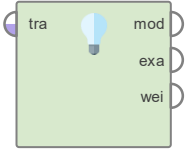
โปรแกรม RapidMiner Studio มีโอเพอร์เรเตอร์ต่างๆที่ช่วยคำนวณ วิเคราะห์ และประมวลผลข้อมูลที่สามารถนำมาใช้ในการทำวิจัยตั้งแต่ขั้นตอนเริ่มต้นจนถึงขั้นตอนประเมินผล ประสิทธิภาพของแบบจำลอง ซึ่งในงานวิจัยนี้ผู้วิจัยเลือกใช้โอเพอร์เรเตอร์ต่างๆ ดังตารางที่ 3.11



ตารางที่ 3.11 โอเปอร์เรเตอร์สำหรับการสร้างแบบจำลองในงานวิจัย

โอเปอร์เรเตอร์	คำอธิบาย
	<p>Write Excel</p> <p>นำเข้าไฟล์ Excel สู่ RapidMiner</p>
	<p>Select Attributes</p> <p>ใช้สำหรับเลือกนำเข้าแอตทริบิวต์ที่ต้องการ</p>
	<p>Set Role</p> <p>ใช้กำหนดบทบาทของแต่ละแอตทริบิวต์</p>
	<p>Retrieve</p> <p>ใช้ดึงข้อมูลที่ต้องการมาใช้ในกระบวนการ</p>
	<p>SMOTE Upsampling</p> <p>เทคนิค SMOTE เพื่อปรับสมดุลของข้อมูล</p>
	<p>Multiply</p> <p>ใช้สำหรับการสร้างข้อมูลนำเข้าหลายๆชุด</p>
	<p>Optimize Parameters (Grid)</p> <p>ใช้สำหรับการค้นหาค่าพารามิเตอร์ที่เหมาะสมที่สุดของแบบจำลองเพื่อให้ได้ผลลัพธ์ที่ดีที่สุดในการทำนายผล</p>
	<p>Cross Validation</p> <p>ใช้สำหรับการประเมินประสิทธิภาพของแบบจำลองโดยการแบ่งชุดข้อมูลออกเป็นส่วนย่อย (k-fold)</p>

ตารางที่ 3.11 (ต่อ)

โอเปอเรเตอร์	คำอธิบาย
Nominal to Nume ... 	Nominal to Numerical ใช้สำหรับการแปลงข้อมูลที่เป็นข้อความหรือหมวดหมู่ให้เป็นข้อมูลตัวเลข
Apply Model 	Apply Model ใช้สำหรับนำแบบจำลองที่สร้างแล้วไปใช้ทำนายข้อมูลชุดใหม่
Performance 	Performance ใช้สำหรับวัดประสิทธิภาพของแบบจำลอง
Weight by Inform ... 	Weight by Information Gain ใช้สำหรับการคัดเลือกแอตทริบิวต์ ด้วยค่า Information Gain
Weight by Chi Sq ... 	Weight by Chi Squared Statistic ใช้สำหรับการคัดเลือกแอตทริบิวต์ ด้วยค่า Chi Squared
Select by Weights 	Select by Weights ใช้สำหรับการเลือกแอตทริบิวต์ที่มีความสำคัญสูงในการทำนาย
Naive Bayes 	Naive Bayes ใช้อัลกอริทึม Naive Bayes สำหรับทำนายผลข้อมูล
Decision Tree 	Decision Tree ใช้อัลกอริทึม Decision Tree สำหรับทำนายผลข้อมูล

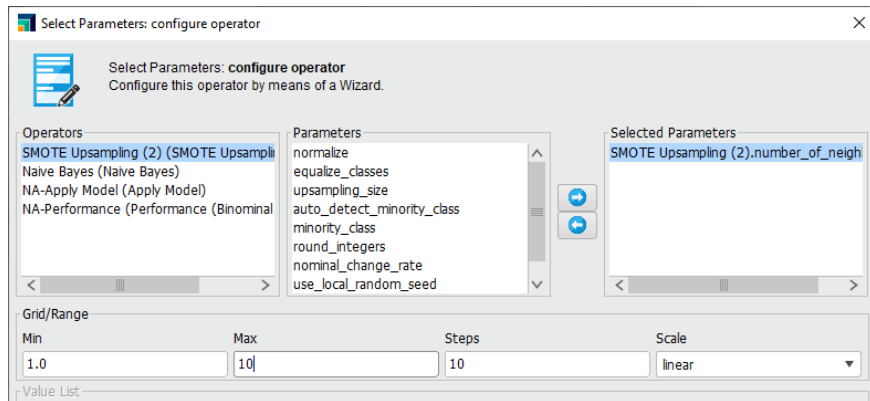
ตารางที่ 3.11 (ต่อ)

โอเปอเรเตอร์		คำอธิบาย
	Neural Net	ใช้อัลกอริทึม Neural Networks สำหรับทำนายผลข้อมูล
	Bagging	เทคนิค Bagging ของอัลกอริทึมการเรียนรู้แบบกลุ่ม ใช้สำหรับทำนายด้วยแบบจำลองหลายตัวที่สุ่มเพื่อสร้างโมเดลหลายๆโมเดล แล้วนำผลลัพธ์ที่ได้มารวมกันเพื่อทำนาย
	AdaBoost	เทคนิค Boosting ของอัลกอริทึมการเรียนรู้แบบกลุ่ม ใช้สำหรับทำนายด้วยการรวมกันของหลายๆ โมเดลย่อย (weak learners) เพื่อสร้างโมเดลที่แข็งแกร่งขึ้น (strong learner) ที่มีความสามารถในการทำนายได้ดีกว่า
	Random Forest	แบบจำลองอัลกอริทึมการเรียนรู้แบบกลุ่มที่ใช้ อัลกอริทึม Random Forest สำหรับทำนายผลข้อมูล

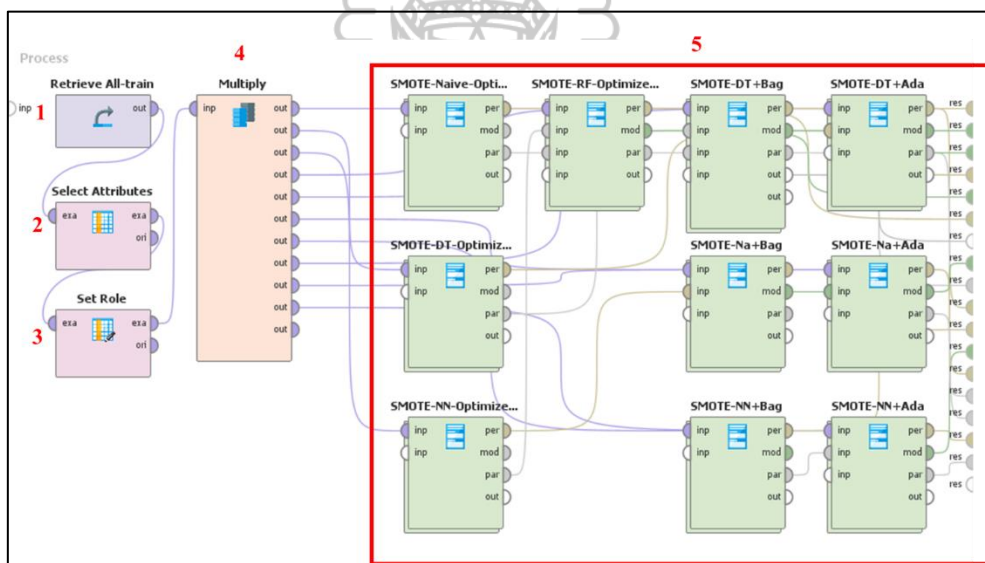
3.3 การปรับข้อมูลไม่สมดุล

ในงานวิจัยนี้มีชุดข้อมูลไม่สมดุล คือ คลาสไม่เกิดพายุฝนฟ้าคะนอง (NO TSRA) มีจำนวนมากกว่าคลาสเกิดพายุฝนฟ้าคะนอง (TSRA) ซึ่งผู้วิจัยได้เลือกทำกระบวนการเพื่อจัดการปรับความไม่สมดุลของข้อมูลก่อนการคัดเลือกคุณลักษณะ เนื่องจากข้อมูลที่ไม่สมดุลกันอาจทำให้แบบจำลองมีความลำเอียงในการเรียนรู้ที่ไม่เหมาะสม ซึ่งคุณลักษณะที่มีจำนวนน้อยอาจถูกมองข้ามและไม่ถูกพิจารณาว่าเป็นสิ่งสำคัญ ส่งผลต่อความแม่นยำของแบบจำลองได้ โดยปรับความสมดุลของข้อมูลด้วยเทคนิคการสุ่มตัวอย่างเกิน หรือ เทคนิค SMOTE ในงานวิจัยนี้ใช้โอเปอเรเตอร์ Optimize

Parameter (Grid) ช่วยในการหาค่าที่เหมาะสมสำหรับ K (Number of Nearest Neighbors) ของแต่ละอัลกอริทึม โดยกำหนดช่วงในการค้นหาเท่ากับ 1-10 ดังภาพที่ 3.6



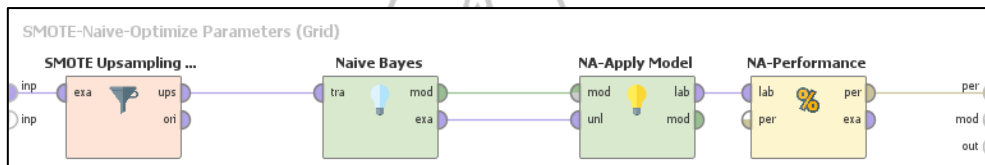
ภาพที่ 3.6 การกำหนดช่วงในการค้นหา K (Number of Nearest Neighbors) เท่ากับ 1-10



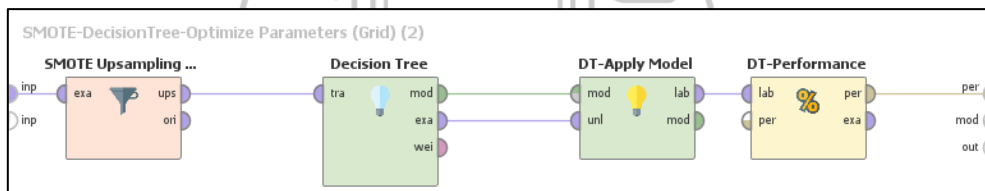
ภาพที่ 3.7 การหาค่า K (Number of Neighbors) ที่เหมาะสมของเทคนิค SMOTE

จากภาพที่ 3.7 แสดงวิธีการหาค่า K จากเทคนิค SMOTE ของอัลกอริทึมมาตรฐาน ได้แก่ Naive Bayes Decision Tree และ Neural Networks รวมถึงของอัลกอริทึมการเรียนรู้แบบกลุ่ม ได้แก่ Random Forest เทคนิค Bagging ร่วมกับอัลกอริทึมมาตรฐาน และเทคนิค AdaBoost ร่วมกับอัลกอริทึมมาตรฐาน สามารถอธิบายขั้นตอนของโอเปอเรเตอร์ตามหมายเลข 1-5 ได้ดังนี้

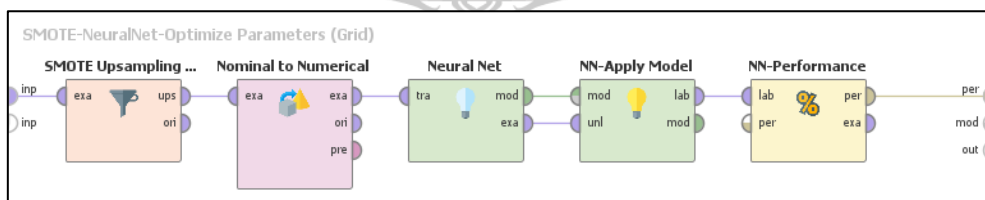
- 1) Retrieve นำเข้าข้อมูลของแต่ละกลุ่มข้อมูลเพื่อสร้างแบบจำลอง
- 2) Select Attributes เลือกแอตทริบิวต์ที่นำมาใช้ทำนาย โดยเลือกแอตทริบิวต์ทั้งหมดเนื่องจากยังไม่ได้ทำขั้นตอนการคัดเลือกคุณลักษณะ
- 3) Set Role เลือกคลาสค่าตอบที่เป็นระยะเวลาพยากรณ์ คือ 1 หรือ 2 ชั่วโมง
- 4) Multiply ต้องการนำ Output ที่มีการ Set Role แล้วไปใช้แยกส่วนกัน
- 5) Optimize Parameters (Grid) ต้องการหาค่า K ของเทคนิค SMOTE ที่เหมาะสม โดย Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของแต่ละอัลกอริทึมดังภาพที่ 3.8 ถึง ภาพที่ 3.17



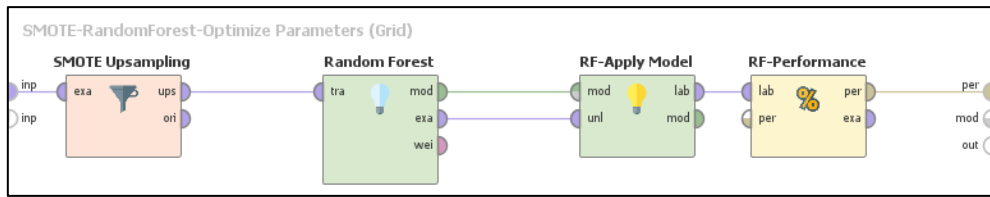
ภาพที่ 3.8 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของ Naive Bayes



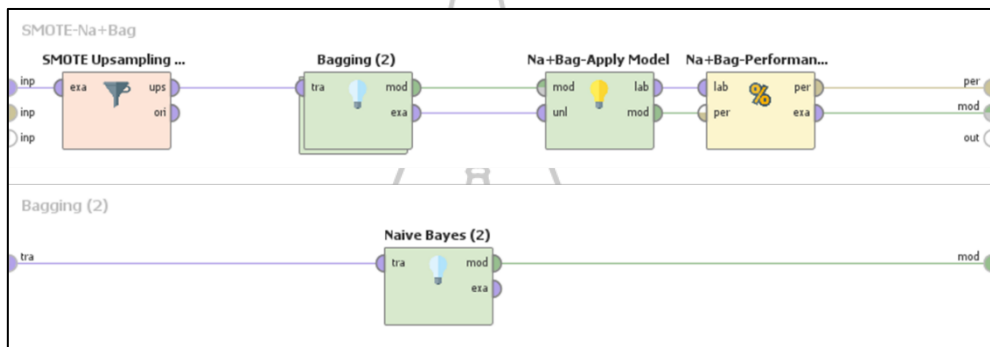
ภาพที่ 3.9 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของ Decision Tree



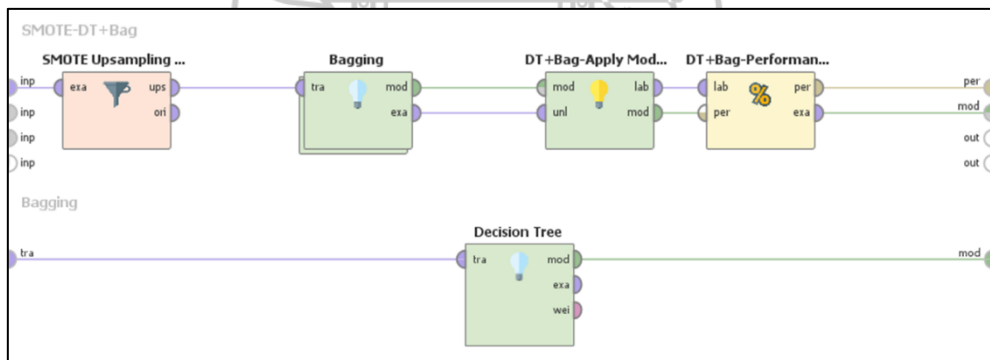
ภาพที่ 3.10 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของ Neural Networks



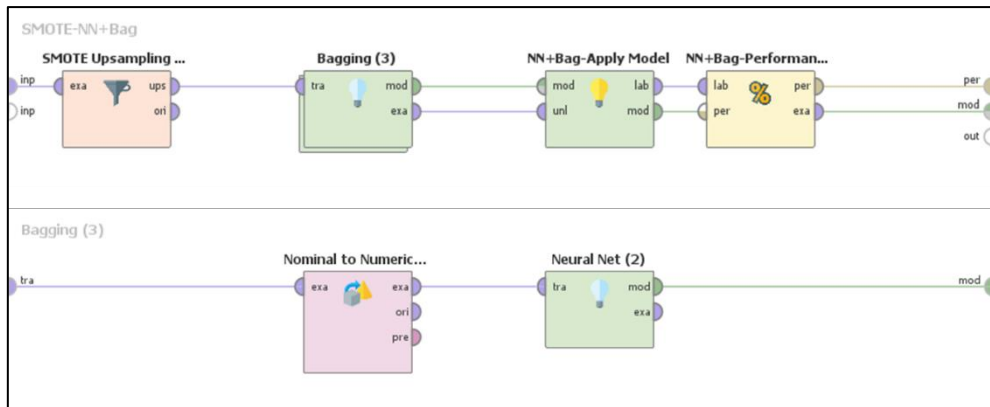
ภาพที่ 3.11 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid)
ของ Random Forest



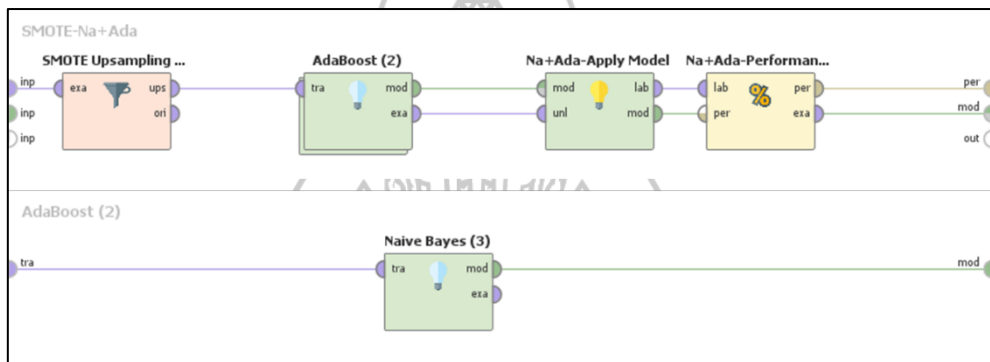
ภาพที่ 3.12 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid)
ของ Naive Bayes ร่วมกับ Bagging



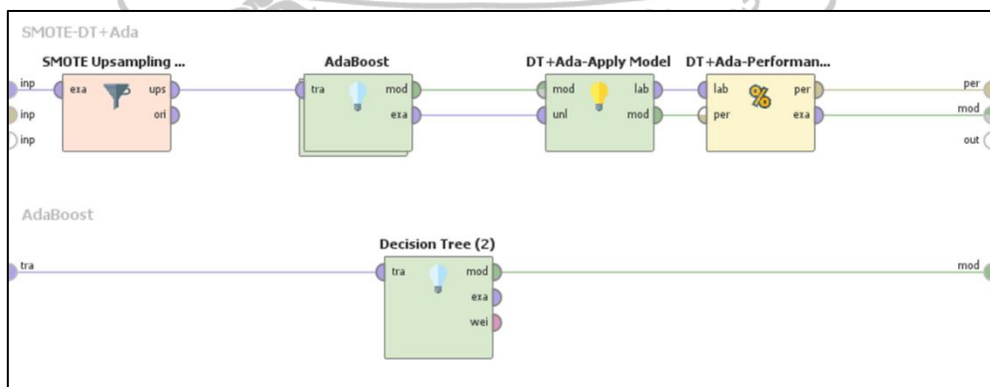
ภาพที่ 3.13 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid)
ของ Decision Tree ร่วมกับ Bagging



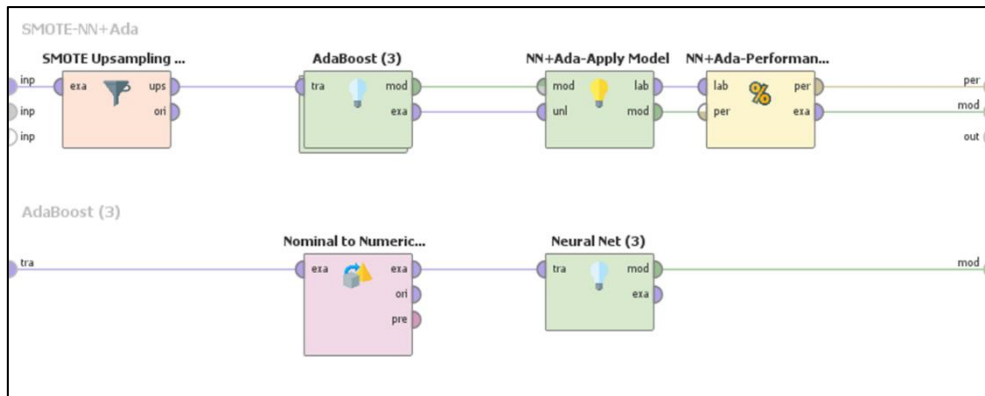
ภาพที่ 3.14 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid)
ของ Neural Networks ร่วมกับ Bagging



ภาพที่ 3.15 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid)
ของ Naive Bayes ร่วมกับ AdaBoost



ภาพที่ 3.16 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid)
ของ Decision Tree ร่วมกับ AdaBoost



ภาพที่ 3.17 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid)
ของ Neural Networks ร่วมกับ AdaBoost

ผู้วิจัยปรับสมดุลของข้อมูลโดยปรับความสมดุลของข้อมูลด้วยเทคนิคการสุ่มตัวอย่างเกิน หรือ เทคนิค SMOTE จากการหาค่าที่เหมาะสมสำหรับ K (Number of Nearest Neighbors) ของแต่ละอัลกอริทึม พบว่า ได้ค่า K เท่ากับ 1 ในทุกแบบจำลอง ในทั้ง 2 ระยะเวลาพยากรณ์ ผลการทำ Optimize Parameter (Grid) ดังตารางที่ 3.12

ตารางที่ 3.12 ค่าที่เหมาะสมสำหรับ K (Number of Nearest Neighbors) ของแต่ละอัลกอริทึม

กลุ่มข้อมูล	อัลกอริทึม	ค่า K (เทคนิค SMOTE)	
		1 ชั่วโมง	2 ชั่วโมง
รวมสนามบินภาคเหนือ	Naive Baye	1	1
	Decision Tree	1	1
	Neural Networks	1	1
	Random Forest	1	1
	Naive Baye ร่วมกับ Bagging	1	1
	Decision Tree ร่วมกับ Bagging	1	1
	Neural Networks ร่วมกับ Bagging	1	1
	Naive Baye ร่วมกับ AdaBoost	1	1
	Decision Tree ร่วมกับ AdaBoost	1	1
Neural Networks ร่วมกับ AdaBoost	1	1	

ตารางที่ 3.12 (ต่อ)

กลุ่มข้อมูล	อัลกอริทึม	ค่า K (เทคนิค SMOTE)	
		1 ชั่วโมง	2 ชั่วโมง
สนามบินภาคเหนือตอนบน	Naive Baye	1	1
	Decision Tree	1	1
	Neural Networks	1	1
	Random Forest	1	1
	Naive Baye ร่วมกับ Bagging	1	1
	Decision Tree ร่วมกับ Bagging	1	1
	Neural Networks ร่วมกับ Bagging	1	1
	Naive Baye ร่วมกับ AdaBoost	1	1
	Decision Tree ร่วมกับ AdaBoost	1	1
Neural Networks ร่วมกับ AdaBoost	1	1	
สนามบินภาคเหนือตอนล่าง	Naive Baye	1	1
	Decision Tree	1	1
	Neural Networks	1	1
	Random Forest	1	1
	Naive Baye ร่วมกับ Bagging	1	1
	Decision Tree ร่วมกับ Bagging	1	1
	Neural Networks ร่วมกับ Bagging	1	1
	Naive Baye ร่วมกับ AdaBoost	1	1
	Decision Tree ร่วมกับ AdaBoost	1	1
Neural Networks ร่วมกับ AdaBoost	1	1	
สนามบินที่มีภูเขาล้อมรอบ 1-2 ด้าน	Naive Baye	1	1
	Decision Tree	1	1
	Neural Networks	1	1
	Random Forest	1	1
	Naive Baye ร่วมกับ Bagging	1	1
	Decision Tree ร่วมกับ Bagging	1	1
	Neural Networks ร่วมกับ Bagging	1	1
	Naive Baye ร่วมกับ AdaBoost	1	1
	Decision Tree ร่วมกับ AdaBoost	1	1
Neural Networks ร่วมกับ AdaBoost	1	1	

ตารางที่ 3.12 (ต่อ)

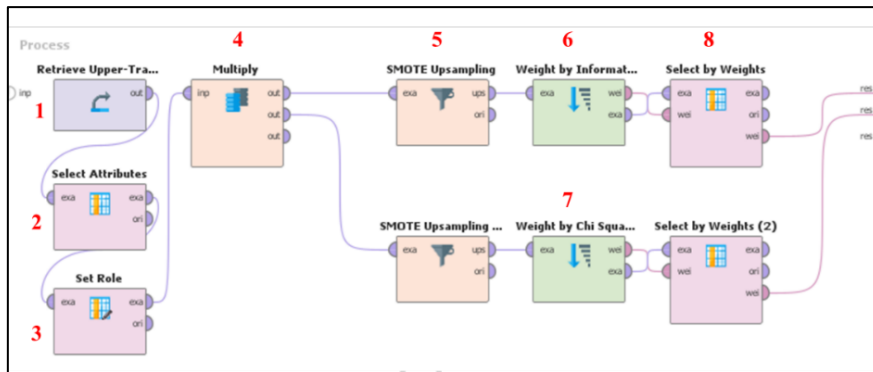
กลุ่มข้อมูล	อัลกอริทึม	ค่า K (เทคนิค SMOTE)	
		1 ชั่วโมง	2 ชั่วโมง
สนามบินที่มีภูเขาล้อมรอบ 3-4 ด้าน	Naive Baye	1	1
	Decision Tree	1	1
	Neural Networks	1	1
	Random Forest	1	1
	Naive Baye ร่วมกับ Bagging	1	1
	Decision Tree ร่วมกับ Bagging	1	1
	Neural Networks ร่วมกับ Bagging	1	1
	Naive Baye ร่วมกับ AdaBoost	1	1
	Decision Tree ร่วมกับ AdaBoost	1	1
	Neural Networks ร่วมกับ AdaBoost	1	1

3.4 การคัดเลือกคุณลักษณะ

ผู้วิจัยจะทำขั้นตอนของการคัดเลือกคุณลักษณะต่อจากการปรับชุดข้อมูลไม่สมดุล ซึ่งในขั้นตอนนี้มีความสำคัญอย่างยิ่งในการคัดเลือกแอตทริบิวต์ที่มีความเกี่ยวข้องกับตัวแปรที่สนใจ และกำจัดแอตทริบิวต์ที่ไม่เกี่ยวข้องออกก่อนนำข้อมูลไปสร้างแบบจำลอง โดยจะช่วยลดมิติของข้อมูล และอาจทำให้การเรียนรู้ดำเนินการได้เร็วขึ้นและมีประสิทธิภาพมากขึ้น ผู้วิจัยใช้วิธีหาค่าน้ำหนัก ความเกี่ยวข้องกับตัวแปรที่สนใจด้วย Information Gain และ Chi-Square ของทั้งหมด 10 แอตทริบิวต์ที่เป็นข้อมูลนำเข้าเพื่อสร้างแบบจำลอง ซึ่งในขั้นตอนการคัดเลือกคุณลักษณะนี้จะกำหนดค่า K ที่เหมาะสมของการปรับความสมดุลข้อมูลเทคนิค SMOTE ตามผลการศึกษาที่ได้จากขั้นตอนที่ผ่านมา รายละเอียดของการคัดเลือกคุณลักษณะ ดังนี้

3.4.1 Information Gain ใช้สำหรับหาค่าน้ำหนักในการคัดเลือกแอตทริบิวต์ที่เกี่ยวข้อง ในงานวิจัยครั้งนี้จะทำการจัดเรียงแอตทริบิวต์ที่มีความเกี่ยวข้องที่สุดจากมากไปน้อยและจะพิจารณาค่าน้ำหนักของแต่ละแอตทริบิวต์ การทำเช่นนี้ทำให้สามารถพิจารณากำจัดแอตทริบิวต์ที่ไม่เกี่ยวข้องออกได้ โดยการหาค่า Information Gain

3.4.2 Chi-Square ใช้สำหรับหาค่าน้ำหนัก ซึ่งแอตทริบิวต์ที่มีความเกี่ยวข้องมากจะมีค่าน้ำหนักมาก และแอตทริบิวต์ที่มีความเกี่ยวข้องน้อยจะมีค่าน้ำหนักน้อย ทำให้สามารถพิจารณากำจัดแอตทริบิวต์ที่ไม่เกี่ยวข้องออกได้โดยการหาค่า Chi-Square



ภาพที่ 3.18 การหาค่าน้ำหนักแอตทริบิวต์ด้วยวิธี Information Gain และ Chi-Square

จากภาพที่ 3.18 สามารถอธิบายขั้นตอนการคัดเลือกคุณลักษณะที่มีโอเปอร์เรเตอร์ตามหมายเลข 1-8 ได้ดังนี้

1. Retrieve นำเข้าข้อมูลของแต่ละกลุ่มข้อมูลเพื่อสร้างแบบจำลอง
2. Select Attributes เลือกแอตทริบิวต์ที่นำมาใช้ทำนาย โดยเลือกแอตทริบิวต์ทั้งหมดเนื่องจากยังไม่ได้ทำขั้นตอนการคัดเลือกคุณลักษณะ
3. Set Role เลือกคลาสคำตอบที่เป็นระยะเวลาพยากรณ์ คือ 1 หรือ 2 ชั่วโมง
4. Multiply ต้องการนำ Output ที่มีการ Set Role แล้วไปใช้แยกส่วนกัน
5. SMOTE Upsampling การปรับสมดุลข้อมูล โดยกำหนดค่าตามผลลัพธ์ของขั้นตอนการปรับค่าที่เหมาะสมสำหรับ K
6. Weight by Information Gain ใช้สำหรับหาค่าน้ำหนักของแอตทริบิวต์ด้วยค่า Information Gain
7. Weight by Chi Squared Statistic ใช้สำหรับหาค่าน้ำหนักของแอตทริบิวต์ ด้วยค่า Chi Squared
8. Select by Weights เรียงลำดับค่าน้ำหนักของแอตทริบิวต์จากมากไปน้อย

โดยเกณฑ์ในการตัดแอตทริบิวต์ก่อนนำไปสร้างแบบจำลอง จะพิจารณาจากค่าน้ำหนักความเกี่ยวข้องกับตัวแปรด้วยวิธี Information Gain และ Chi-Square ที่เรียงลำดับความสำคัญมากไปหาน้อย หรือลำดับที่ 1-10 ของงานวิจัยนี้ มีเกณฑ์ดังนี้

1. วิธี Information Gain จะพิจารณาตัดแอตทริบิวต์ที่มีค่าน้ำหนักความเกี่ยวข้องกับตัวแปรวิธี Information Gain เท่ากับ 0.00 ออก เนื่องจากไม่มีความสัมพันธ์กับตัวแปร (ธงไชย พ้องเสียง และจารี ทองคำ, 2566)

2. พิจารณาตัดแอตทริบิวต์ไม่เกิน 2 แอตทริบิวต์ หากลำดับที่ 8 ไม่มีค่าน้ำหนักความเกี่ยวข้องกับตัวแปรวิธี Information Gain และ Chi-Square เท่ากับ 0 เนื่องจากต้องการให้แบบจำลองมีข้อมูลนำเข้าที่หลากหลายตัวแปรอย่างน้อย 8 แอตทริบิวต์

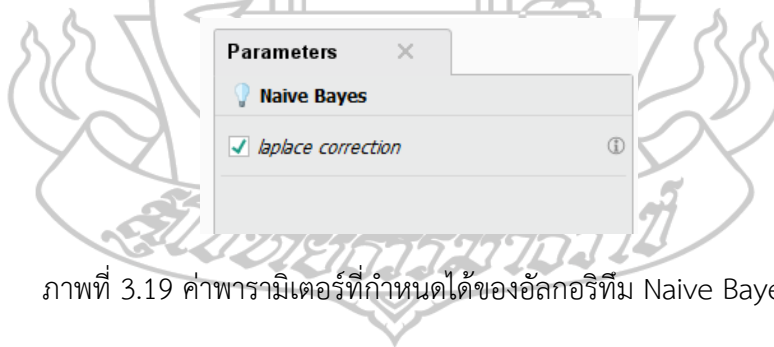
3. พิจารณาแอตทริบิวต์ในลำดับที่ 9 ซึ่งมีความเกี่ยวข้องกับตัวแปรน้อย หากค่าน้ำหนักความเกี่ยวข้องกับตัวแปรวิธี Information Gain และ Chi-Square มีเปอร์เซ็นต์ส่วนต่างกับลำดับก่อนหน้ามากกว่าหรือเท่ากับ 100% ให้ตัดแอตทริบิวต์ในลำดับที่ 9 รวมถึงลำดับที่ 10 ที่มีความเกี่ยวข้องน้อยออก

3.5 การปรับค่าพารามิเตอร์ที่เหมาะสมในงานวิจัย

ผู้วิจัยต้องการกำหนดค่า Hyperparameter ซึ่งเป็นพารามิเตอร์ต่างๆ ที่มีการกำหนดโครงสร้างของแบบจำลองตั้งแต่เริ่มต้น เนื่องจากมีผลต่อประสิทธิภาพของแบบจำลองงานวิจัยนี้จึงมีการหาค่าพารามิเตอร์ที่เหมาะสมของแต่ละอัลกอริทึม รายละเอียดดังนี้

3.5.1 ค่าพารามิเตอร์ที่สามารถปรับค่าได้สำหรับอัลกอริทึมมาตรฐาน

1) Naive Bayes ไม่ต้องมีการปรับค่าพารามิเตอร์ใดๆ ดังภาพที่ 3.19



ภาพที่ 3.19 ค่าพารามิเตอร์ที่กำหนดได้ของอัลกอริทึม Naive Bayes

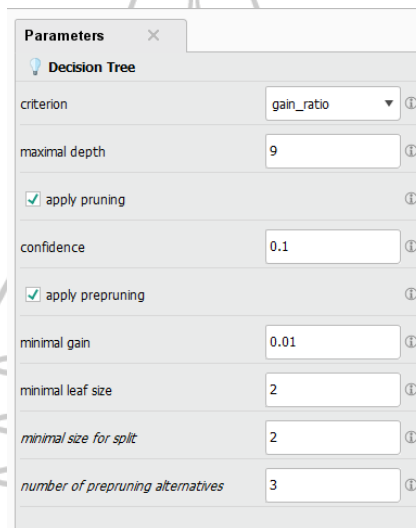
2) Decision Tree หาค่าที่เหมาะสมของพารามิเตอร์ ดังภาพที่ 3.20 ได้แก่

(1) Criterion คือ เป็นค่าที่ใช้กำหนดเกณฑ์ในการเลือกแอตทริบิวต์ที่ต้นไม้มักจะทำการแตกกิ่ง โดยจะมีการปรับค่าของการแตกกิ่งให้เหมาะสม ซึ่งสามารถเลือกได้หนึ่งเกณฑ์จากทั้งหมด 5 เกณฑ์ ได้แก่ Gain_Ratio Information_Gain Gini_Index Accuracy และ Least_Square

(2) Maximum Depth คือ การกำหนดค่าความลึกของต้นไม้ โดยหากตั้งค่าเป็น “-1” จะไม่มีการกำหนดค่าความลึกของต้นไม้ และต้นไม้จะถูกสร้างต่อไปจนกว่าจะถูกให้หยุดการสร้างด้วยเกณฑ์อื่นๆ และหากตั้งค่าเป็น “1” ต้นไม้ที่ถูกสร้างขึ้นจะมีเพียงโหนดเดียวเท่านั้น

(3) minimal leaf size คือ ขนาดของใบไม้ ซึ่งเป็นจำนวนข้อมูลในแต่ละกลุ่มข้อมูลที่มีกาแบ่งย่อยออกมาแล้ว ต้นไม้จึงจะถูกสร้างขึ้นในลักษณะที่ใบไม้ทุกใบมีจำนวนขั้นต่ำใน Leaf Node ตามที่กำหนดไว้

(4) minimal size for split คือ ขนาดของโหนดหรือจำนวนของข้อมูลซึ่งจะมีเฉพาะโหนดที่มีขนาดมากกว่าหรือเท่ากับค่าที่ตั้งค่าไว้เท่านั้นจึงจะมีการแตกกิ่งออกไป



Parameter	Value
criterion	gain_ratio
maximal depth	9
apply pruning	<input checked="" type="checkbox"/>
confidence	0.1
apply prepruning	<input checked="" type="checkbox"/>
minimal gain	0.01
minimal leaf size	2
minimal size for split	2
number of prepruning alternatives	3

ภาพที่ 3.20 ค่าพารามิเตอร์ที่กำหนดได้ของอัลกอริทึม Decision Tree

3) Neural Networks หาค่าที่เหมาะสมของพารามิเตอร์ ดังภาพที่ 3.21 ได้แก่

(1) Training Cycle คือ รอบการฝึกฝน เป็นจำนวนรอบที่ Neural Networks จะถูกฝึกฝนในแต่ละครั้ง ในแต่ละรอบการฝึกฝนแบบจำลองจะใช้ชุดข้อมูลการฝึกฝนที่กำหนดมาในการปรับค่าน้ำหนักเพื่อให้มีการเรียนรู้เกิดขึ้น การกำหนดรอบการฝึกฝนให้เหมาะสมเป็นสิ่งสำคัญเช่นกัน เนื่องจากถ้าฝึกฝนนานเกินไปอาจทำให้เกิดการเรียนรู้เกินไป (overfitting) และถ้าฝึกฝนน้อยเกินไปอาจทำให้แบบจำลองไม่มีประสิทธิภาพในการทำนายข้อมูลใหม่ (underfitting)

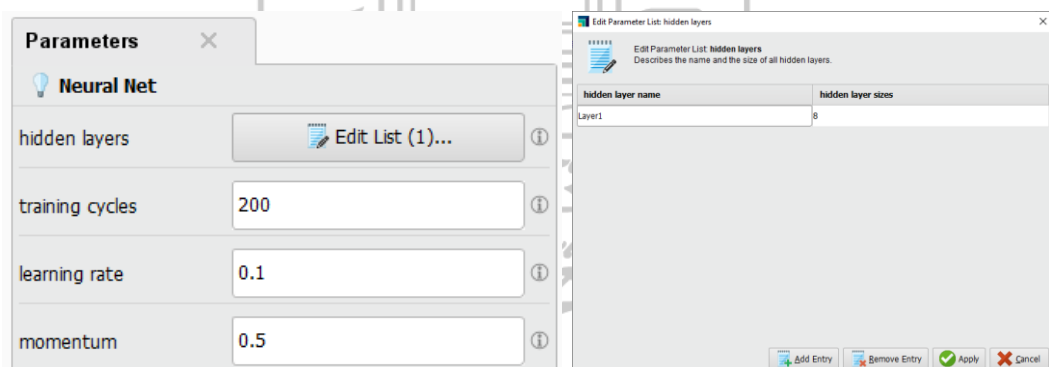
(2) Learning Rate คือ อัตราการเรียนรู้ เป็นพารามิเตอร์ที่กำหนดอัตราการปรับค่าน้ำหนักในแต่ละรอบการฝึกฝน (training iteration) การเลือกค่า Learning Rate ที่

เหมาะสมเป็นสิ่งสำคัญ เนื่องจากค่าที่สูงกว่าอาจทำให้การฝึกอบรมไม่เสถียร ส่วนค่าที่ต่ำกว่าอาจทำให้การฝึกอบรมช้าลง

(3) Momentum คือ ความเร็วในการอัปเดต เป็นพารามิเตอร์ที่กำหนดความเร็วในการอัปเดตค่าน้ำหนักในแต่ละรอบการฝึก ค่าที่สูงกว่าอาจทำให้การฝึกอบรมเกิดการเดินเร็วข้ามจุดต่อเนื่อง ส่วนค่าที่ต่ำกว่าอาจทำให้การฝึกอบรมช้าลง

(4) Hidden layers คือ พารามิเตอร์นี้ใช้สำหรับกำหนดชื่อและขนาดของ Hidden Layers สามารถกำหนดโครงสร้างของ Neural Network ได้ด้วยพารามิเตอร์นี้ ในงานวิจัยนี้ระบุจำนวน 1 เลเยอร์ ซึ่งเพียงพอสำหรับการวิเคราะห์ข้อมูล การกำหนดมากกว่า 1 ชั้นจะทำให้แบบจำลองใช้เวลานานในการประมวลผล

(5) Hidden layer size คือ จำนวน Node ภายในชั้น Hidden layers ซึ่งทุกครั้งจำนวนของ Node จะมากกว่าค่าที่ตั้งไว้ค่าหนึ่งเสมอเพราะจะมี Node ที่เพิ่มเติมเข้ามาในแต่ละชั้น แต่ Node ที่เพิ่มขึ้นนี้ไม่ได้เชื่อมกับชั้นก่อนหน้า ถ้าหากตั้งค่าเป็น -1 ขนาดของชั้นนั้นจะถูกคำนวณจากจำนวนของ attributes ของชุดข้อมูลที่ใส่เข้าไป จากสมการ (จำนวน Attributes + จำนวนของ Class)/2+1 แต่ถ้าหากไม่ตั้งค่าใด ๆ ก็จะเป็นค่าของประเภทข้อมูลและขนาดที่เท่ากับสมการข้างต้น ซึ่งถ้ามีแค่ชั้นเดียวโดยไม่มี Node ใด ๆ Input node ก็จะเชื่อมกับ Output node โดยที่ไม่มี Hidden layer



ภาพที่ 3.21 ค่าพารามิเตอร์ที่กำหนดได้ของอัลกอริทึม Neural Networks

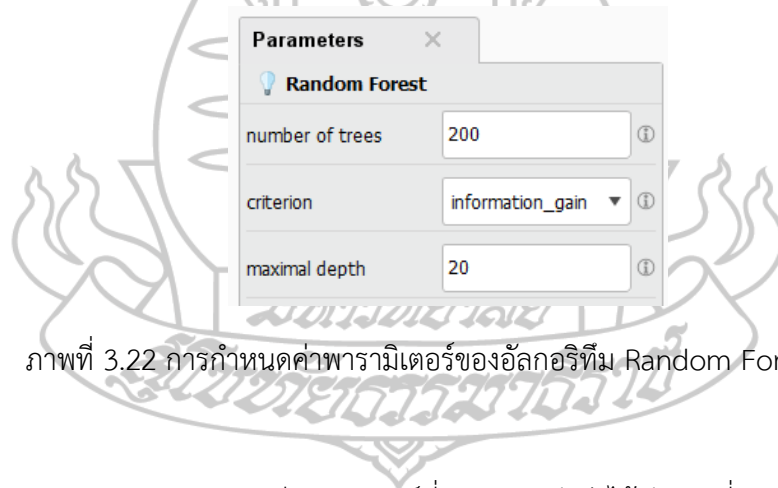
3.5.2 ค่าพารามิเตอร์ที่สามารถปรับค่าได้สำหรับอัลกอริทึมการเรียนรู้แบบกลุ่ม

1) Random Forest สามารถสร้างแบบจำลองได้เองโดยไม่ต้องทำงานร่วมกับอัลกอริทึมมาตรฐาน โดยหาค่าที่เหมาะสมของพารามิเตอร์สำหรับ Random Forest ดังภาพที่ 3.22 ได้แก่

(1) Number of Trees (numTrees) คือ ค่ากำหนดจำนวนต้นไม้ที่จะสร้าง และต้นไม้แต่ละต้นจะมีการเลือกชุดข้อมูลย่อย ซึ่งจำนวนต้นไม้ที่มากขึ้นจะทำให้ประสิทธิภาพของแบบจำลองดีขึ้น จนถึงจุดหนึ่งที่ค่าประสิทธิภาพคงที่จนจำนวนต้นไม้ไม่ส่งผลต่อการทำนายอีกต่อไป ค่าเริ่มต้นที่ใช้ใน RapidMiner คือ 100

(2) Criterion คือ เป็นค่าที่ใช้กำหนดเกณฑ์ในการเลือกแอตทริบิวต์ที่ต้นไม้จะทำการแตกกิ่ง โดยจะมีการปรับค่าของการแตกกิ่งให้เหมาะสม ซึ่งสามารถเลือกได้หนึ่งเกณฑ์จากทั้งหมด 5 เกณฑ์ ได้แก่ Gain_Ratio Information_Gain Gini_Index Accuracy และ Least_Square

(3) Maximum Depth (maxDepth) คือ ค่ากำหนดความลึกสูงสุดของต้นไม้ใน Random Forest ซึ่งมีผลต่อความซับซ้อนและการจัดกลุ่มของข้อมูล ค่าเริ่มต้นที่ใช้ใน RapidMiner คือ 10



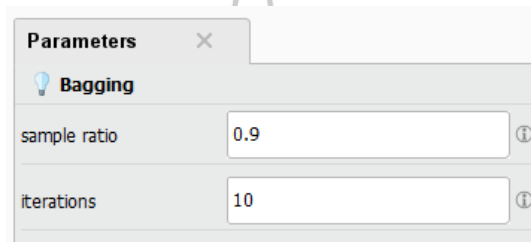
ภาพที่ 3.22 การกำหนดค่าพารามิเตอร์ของอัลกอริทึม Random Forest

2) Bagging มีพารามิเตอร์ที่สามารถปรับค่าได้ดังภาพที่ 3.23 คือ

(1) Sample Ratio คือ ค่าสัดส่วนของข้อมูลตัวอย่างที่จะถูกใช้ในแต่ละรอบในกระบวนการ Bagging โดยค่าเริ่มต้นของพารามิเตอร์นี้คือ 0.9 ซึ่งหมายถึงใช้ข้อมูลตัวอย่าง 90% ในแต่ละรอบ หากต้องการใช้ข้อมูลตัวอย่างเพียงบางส่วน สามารถตั้งค่าสัดส่วนที่เหมาะสมในการสร้างโมเดลส่วนย่อย งานวิจัยนี้จะตั้งค่าเป็น 0.9

(2) Iteration คือ จำนวนรอบในกระบวนการ Bagging นั่นคือจำนวนโมเดลส่วนย่อยที่จะถูกสร้างขึ้นในกระบวนการ Bagging ค่าเริ่มต้นของพารามิเตอร์นี้คือ 10 ซึ่งหมายถึงการสร้างโมเดลส่วนย่อย 10 ตัว ในงานวิจัยนี้จะตั้งค่าเป็น 10

(3) พารามิเตอร์ตามอัลกอริทึมมาตรฐานที่ทำงานร่วมกัน ได้แก่ Naive Bayes Decision Tree และ Neural Networks

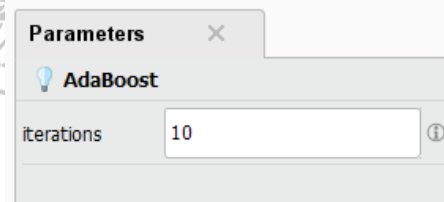


ภาพที่ 3.23 การกำหนดพารามิเตอร์ของเทคนิค Bagging

3) AdaBoost มีพารามิเตอร์ที่สามารถปรับค่าได้ดังภาพที่ 3.24 คือ

(1) Iteration ที่เป็นจำนวนรอบในกระบวนการ AdaBoost แสดงถึงจำนวนโมเดลส่วนย่อยที่จะถูกสร้างขึ้นในกระบวนการ AdaBoost ค่าเริ่มต้นของพารามิเตอร์นี้คือ 10 ซึ่งหมายถึงการสร้างโมเดลส่วนย่อย 10 ตัว และในงานวิจัยนี้จะตั้งค่าเป็น 10

(2) พารามิเตอร์ตามอัลกอริทึมมาตรฐานที่ทำงานร่วมกัน ได้แก่ Naive Bayes Decision Tree และ Neural Networks

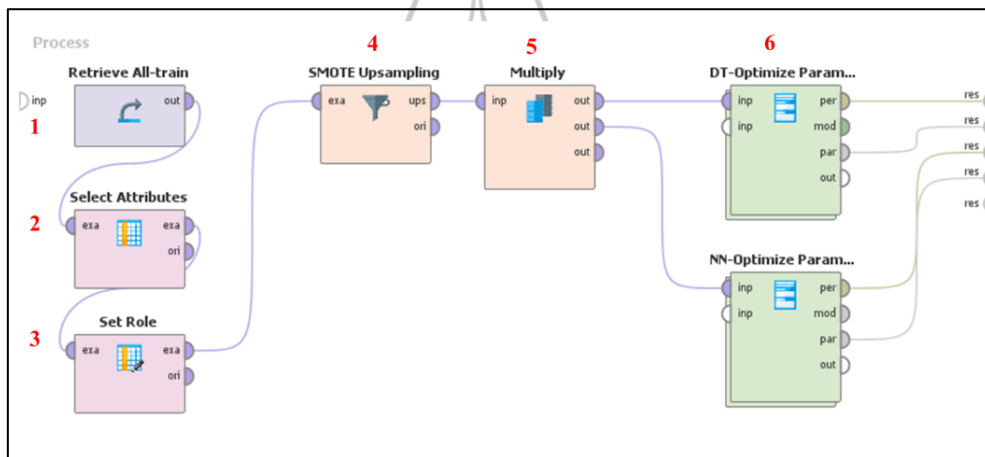


ภาพที่ 3.24 การกำหนดพารามิเตอร์ของเทคนิค AdaBoost

3.5.3 ขั้นตอนการ Hyperparameter วิธี Grid Search ด้วย โอเปอเรเตอร์

Optimize Parameter (Grid)

1) ขั้นตอนการ Hyperparameter สำหรับอัลกอริทึมมาตรฐาน ได้แก่ Decision Tree Neural Networks ส่วน Naive Bayes ไม่ต้องการปรับค่าพารามิเตอร์ใดๆ โดยผู้วิจัยได้ใช้โอเปอเรเตอร์ Optimize Parameter (Grid) มาช่วยในการหาค่ากำหนดสำหรับพารามิเตอร์ต่าง ๆ ที่เหมาะสม ซึ่งก่อนทำขั้นตอนนี้จะต้องทำการปรับสมดุลของข้อมูลและคัดเลือกคุณลักษณะมาแล้วเพื่อให้ได้ค่า K (Number of Neighbours) ของเทคนิค SMOTE และจำนวนแอตทริบิวต์ที่เหมาะสม และเข้าสู่การสร้างแบบจำลองจริง ดังภาพที่ 3.25

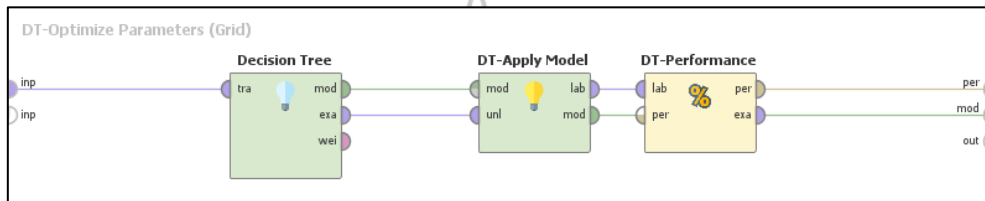


ภาพที่ 3.25 การปรับค่าพารามิเตอร์ที่เหมาะสมสำหรับอัลกอริทึมมาตรฐาน

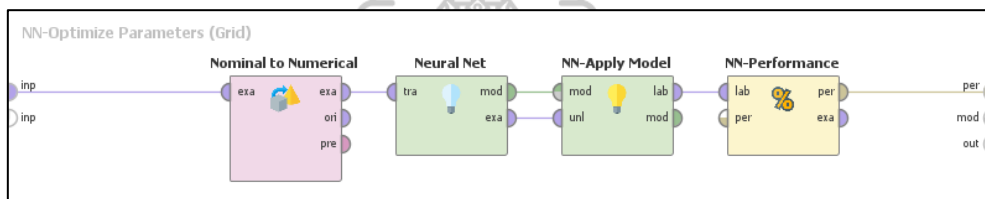
จากภาพที่ 3.25 สามารถอธิบายขั้นตอนตามหมายเลข 1-6 ได้ดังนี้

1. Retrieve นำเข้าข้อมูลของแต่ละกลุ่มข้อมูลเพื่อสร้างแบบจำลอง
2. Select Attributes เลือกแอตทริบิวต์ที่นำมาใช้ทำนาย โดยเลือกจากผลลัพธ์ของขั้นตอนในการคัดเลือกคุณลักษณะ
3. Set Role เลือกคลาสคำตอบที่เป็นระยะเวลาพยากรณ์ คือ 1 หรือ 2 ชั่วโมง
4. SMOTE Upsampling การปรับสมดุลข้อมูล โดยกำหนดค่าตามผลลัพธ์ของขั้นตอนการปรับค่าที่เหมาะสมสำหรับ K
5. Multiply ต้องการนำ Output ที่มีการ Set Role แล้วไปใช้แยกส่วนกัน

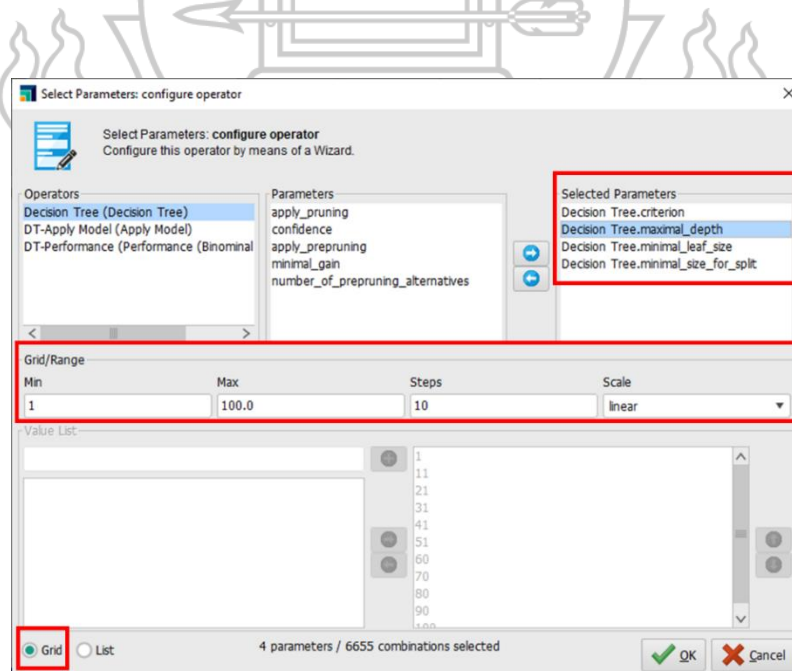
6. Optimize Parameters (Grid) ต้องการหาค่าพารามิเตอร์ที่เหมาะสมโดย Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของแต่ละอัลกอริทึมดังภาพที่ 3.26-3.27 และการปรับการค้นหาของแต่ละอัลกอริทึมดังภาพที่ 3.28-3.29



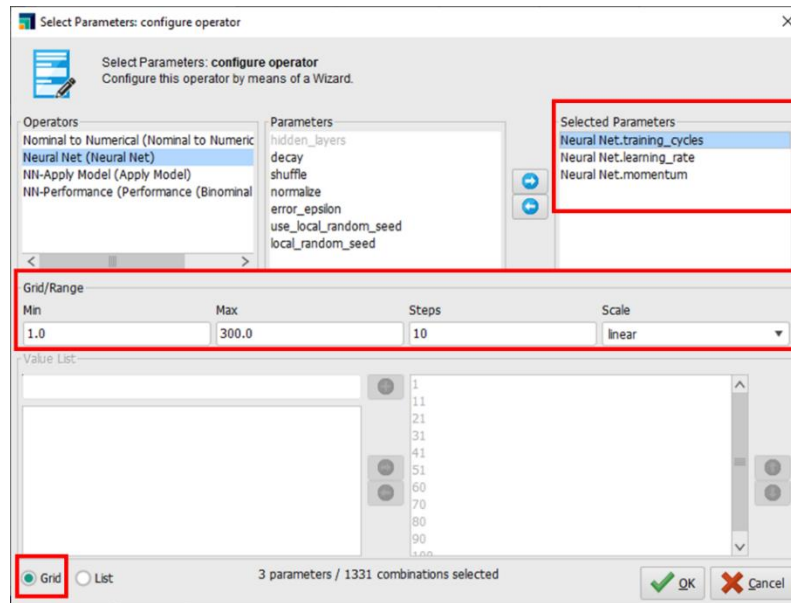
ภาพที่ 3.26 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของ Decision Tree



ภาพที่ 3.27 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของ Neural Networks

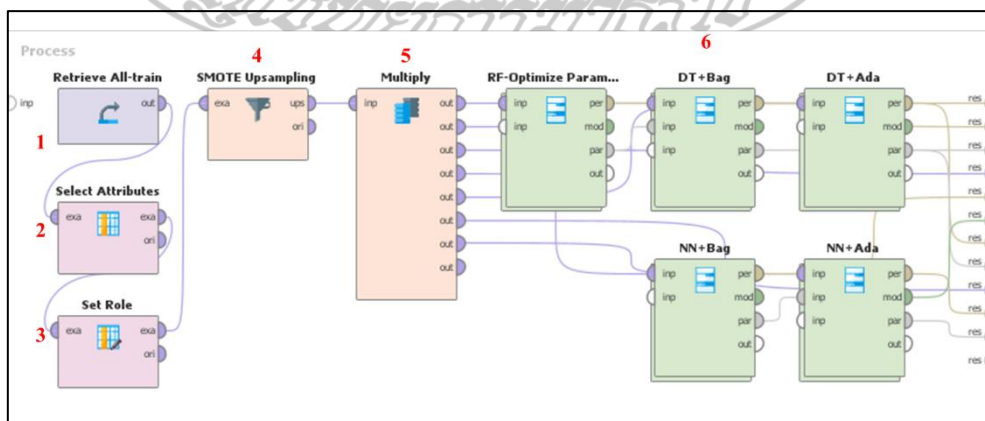


ภาพที่ 3.28 การปรับการค้นหา Optimize Parameter (Grid) ของ Decision Tree



ภาพที่ 3.29 การปรับการค้นหา Optimize Parameter (Grid) ของ Neural Networks

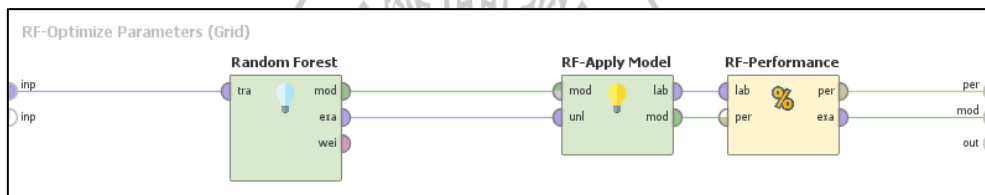
2) ขั้นตอนการ Hyperparameter สำหรับอัลกอริทึมการเรียนรู้แบบกลุ่ม ได้แก่ Random Forest เทคนิค Bagging และเทคนิค AdaBoost แต่สำหรับการใช้งานร่วมกับอัลกอริทึม Naive Bayes ไม่ต้องการปรับตั้งค่าเพิ่มเติม โดยผู้วิจัยได้ใช้โอเพอเรเตอร์ Optimize Parameter (Grid) มาช่วยในการหาค่ากำหนดสำหรับพารามิเตอร์ต่าง ๆ ที่เหมาะสม ซึ่งก่อนทำขั้นตอนนี้จะต้องทำการปรับสมดุลของข้อมูลและคัดเลือกคุณลักษณะมาแล้วเพื่อให้ได้ค่า K (Number of Neighbours) ของเทคนิค SMOTE และจำนวนแอตทริบิวต์ที่เหมาะสม และเข้าสู่การสร้างแบบจำลองจริง ดังภาพที่ 3.30



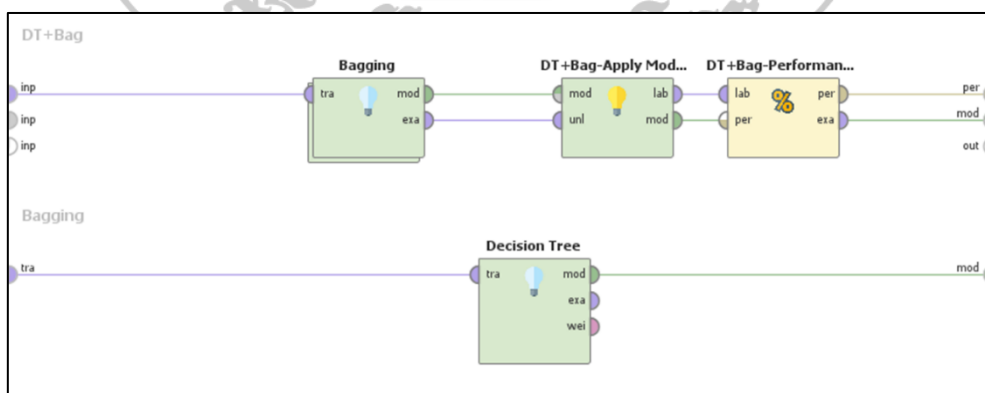
ภาพที่ 3.30 การปรับค่าพารามิเตอร์ที่เหมาะสมสำหรับอัลกอริทึมการเรียนรู้แบบกลุ่ม

จากภาพที่ 3.30 สามารถอธิบายขั้นตอนตามหมายเลข 1-6 ได้ดังนี้

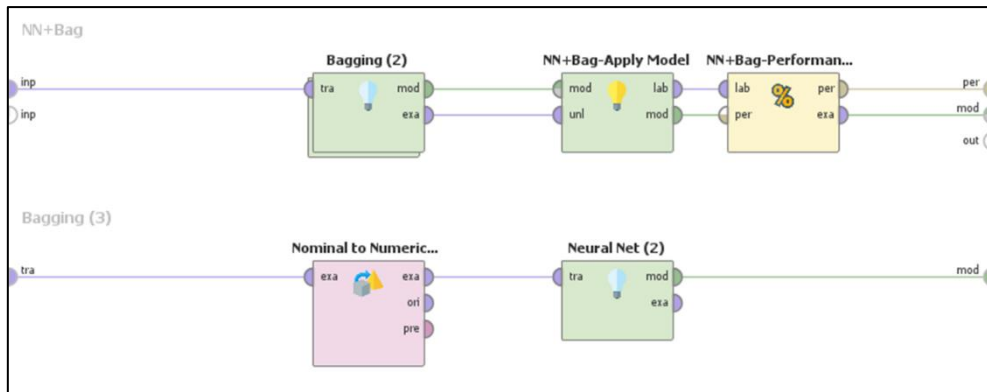
- 1.Retrieve นำเข้าข้อมูลของแต่ละกลุ่มข้อมูลเพื่อสร้างแบบจำลอง
- 2.Select Attributes เลือกแอตทริบิวต์ที่นำมาใช้ทำนาย โดยเลือกจากผลลัพธ์ของขั้นตอนในการคัดเลือกคุณลักษณะ
- 3.Set Role เลือกคลาสค่าตอบที่เป็นระยะเวลาพยากรณ์ คือ 1 หรือ 2 ชั่วโมง
- 4.SMOTE Upsampling การปรับสมดุลข้อมูล โดยกำหนดค่าตามผลลัพธ์ของขั้นตอนการปรับค่าที่เหมาะสมสำหรับ K
- 5.Multiply ต้องการนำ Output ที่มีการ Set Role แล้วไปใช้แยกส่วนกัน
- 6.Optimize Parameters (Grid) ต้องการหาค่าพารามิเตอร์ที่เหมาะสมโดย Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของแต่ละอัลกอริทึมดังภาพที่ 3.31-3.35 และการปรับการค้นหาของแต่ละอัลกอริทึมดังภาพที่ 3.36-3.38



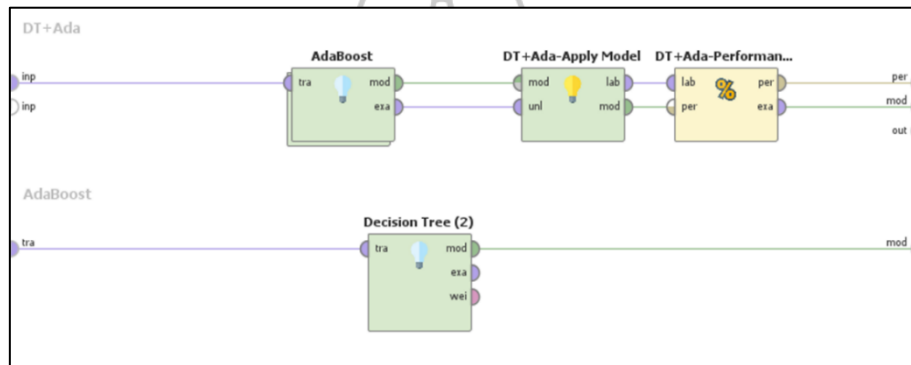
ภาพที่ 3.31 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของ Random Forest



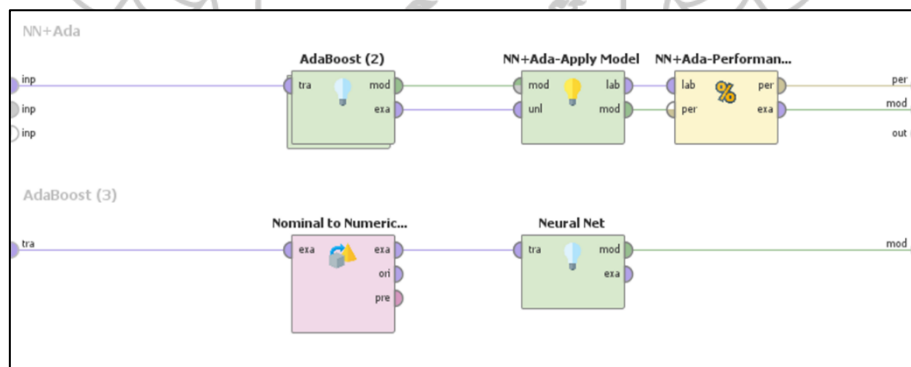
ภาพที่ 3.32 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของ Decision Tree ร่วมกับ Bagging



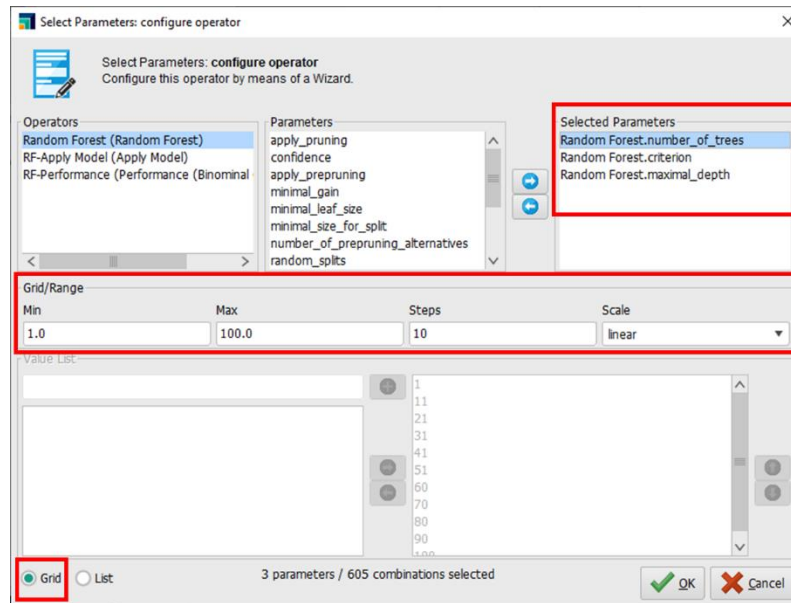
ภาพที่ 3.33 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของ Neural Networks ร่วมกับ Bagging



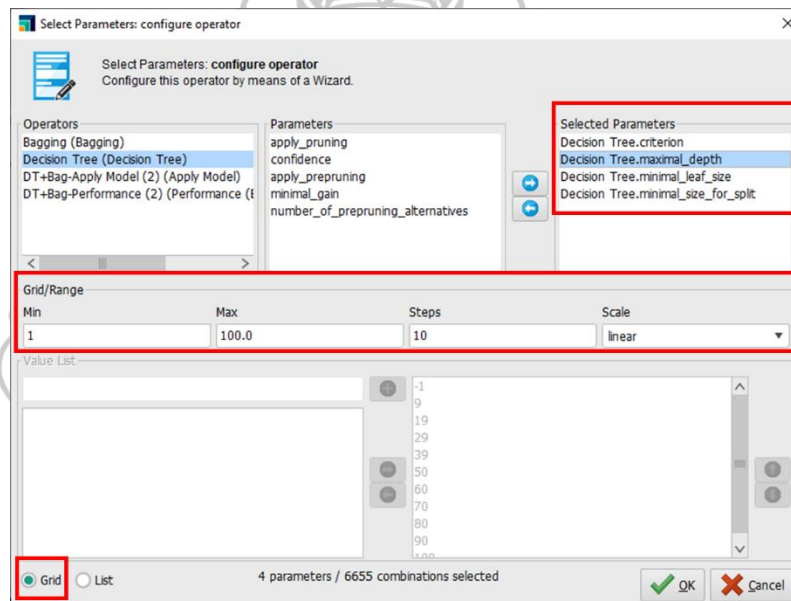
ภาพที่ 3.34 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของ Decision Tree ร่วมกับ AdaBoost



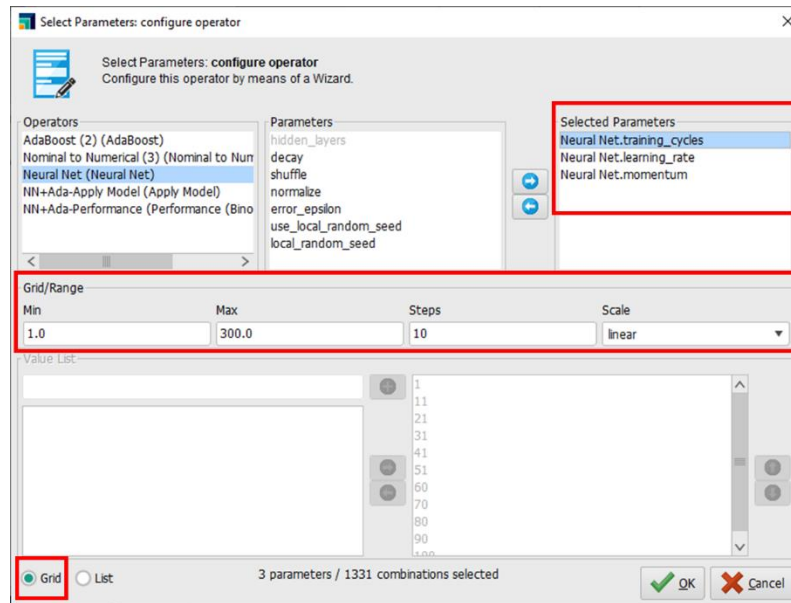
ภาพที่ 3.35 Sub-Process ภายในโอเพอร์เรเตอร์ Optimize Parameter (Grid) ของ Neural Networks ร่วมกับ AdaBoost



ภาพที่ 3.36 การปรับการค้นหา Optimize Parameter (Grid) ของ Random Forest



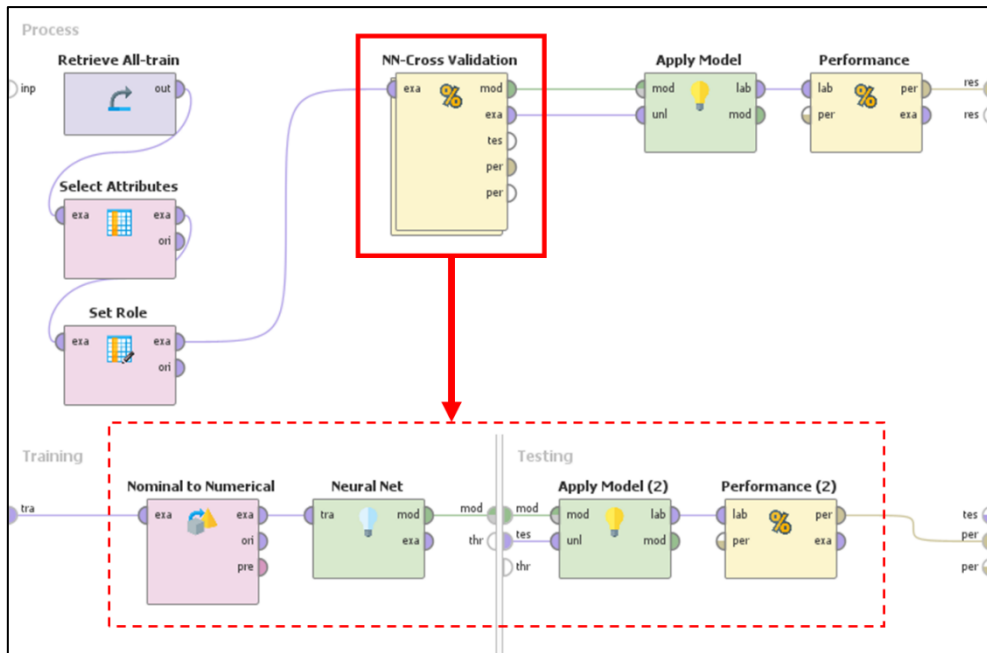
ภาพที่ 3.37 การปรับการค้นหา Optimize Parameter (Grid) ของ Decision Tree ร่วมกับอัลกอริทึมการเรียนรู้แบบกลุ่ม



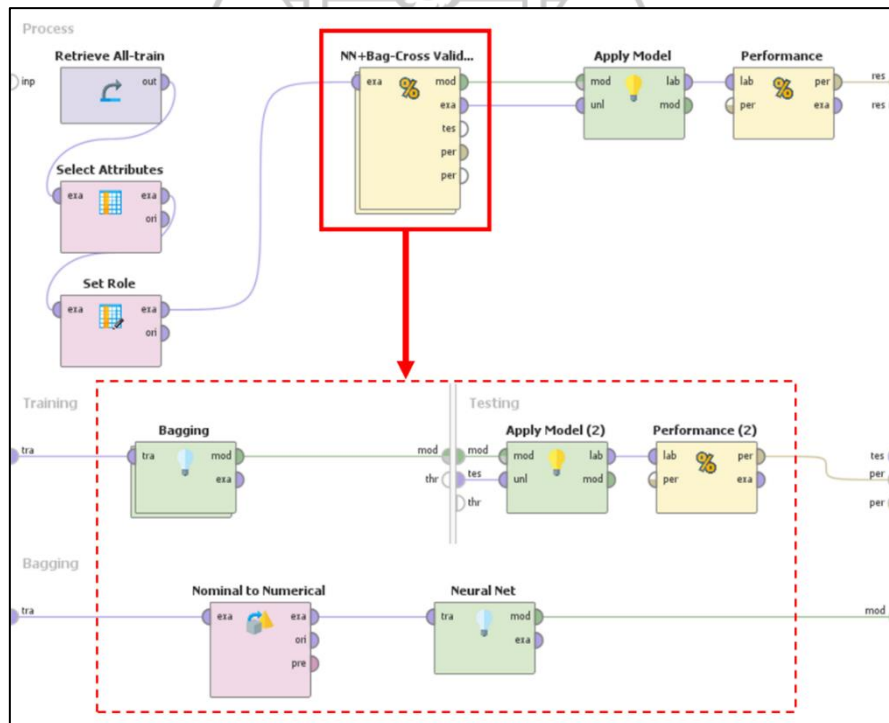
ภาพที่ 3.38 การปรับการค้นหา Optimize Parameter (Grid) ของ Neural Networks ร่วมกับอัลกอริทึมการเรียนรู้แบบกลุ่ม

3.5.4 ขั้นตอนการ Hyperparameter วิธี Random Search

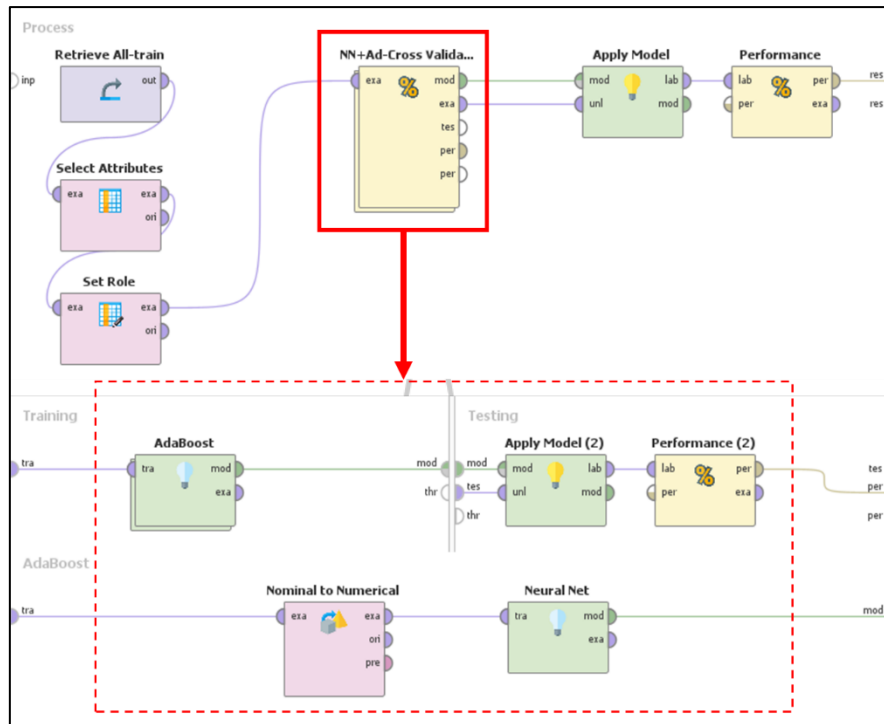
จากภาพที่ 3.38 การปรับการค้นหา Optimize Parameter (Grid) ของ Neural Networks ไม่สามารถกำหนดการค้นหาค่าที่เหมาะสมสำหรับ Hidden Layers ได้ ซึ่งวิธี Random Search จะทำการสุ่มเลือกค่าพารามิเตอร์จากกริดที่สร้างขึ้น ดังนั้นบางครั้งอาจจะไม่ได้แบบจำลองที่มีประสิทธิภาพที่สุดเหมือนกับ Grid Search แต่วิธีนี้มีประสิทธิภาพสูงในการใช้งานจริง เนื่องจากใช้เวลาในการสร้างแบบจำลองน้อย โดยในงานวิจัยนี้กำหนดค่า Hidden layers ซึ่งเป็นโครงสร้างของ Neural Network จำนวน 1 เลเยอร์ ซึ่งเพียงพอสำหรับการวิเคราะห์ข้อมูล การกำหนดมากกว่า 1 ชั้นจะทำให้แบบจำลองใช้เวลานานในการประมวลผล และทำการค้นหาค่า Hidden layer size หรือจำนวน Node ภายในชั้น Hidden layers ที่เหมาะสม เริ่มค้นหาโดยการสร้างแบบจำลองจากอัลกอริทึม Neural Networks โดยกำหนดค่าพารามิเตอร์ต่างๆที่เหมาะสมตามขั้นตอนก่อนหน้านี้ ได้แก่ Training Cycle Learning Rate และ Momentum จากนั้นกำหนดค่า Hidden layer size เท่ากับ 1-8 ตามจำนวนแอตทริบิวต์สูงสุดที่นำเข้าไปเพื่อสร้างแบบจำลอง โดยเปรียบเทียบประสิทธิภาพเพื่อคัดเลือกค่าที่เหมาะสมที่สุดของแต่ละชุดข้อมูลจากค่า F-measure สูงสุด การสร้างแบบจำลองเพื่อค้นหาค่า Hidden layer size แสดงดังภาพที่ 3.39-3.41 และผลการทดสอบทั้งหมดในภาคผนวก ก



ภาพที่ 3.39 การสร้างแบบจำลองจากอัลกอริทึม Neural Networks เพื่อค้นหาค่า Hidden layer size ด้วยวิธี Random Search



ภาพที่ 3.40 การสร้างแบบจำลองจากอัลกอริทึม Neural Networks ร่วมกับ Bagging เพื่อค้นหาค่า Hidden layer size ด้วยวิธี Random Search

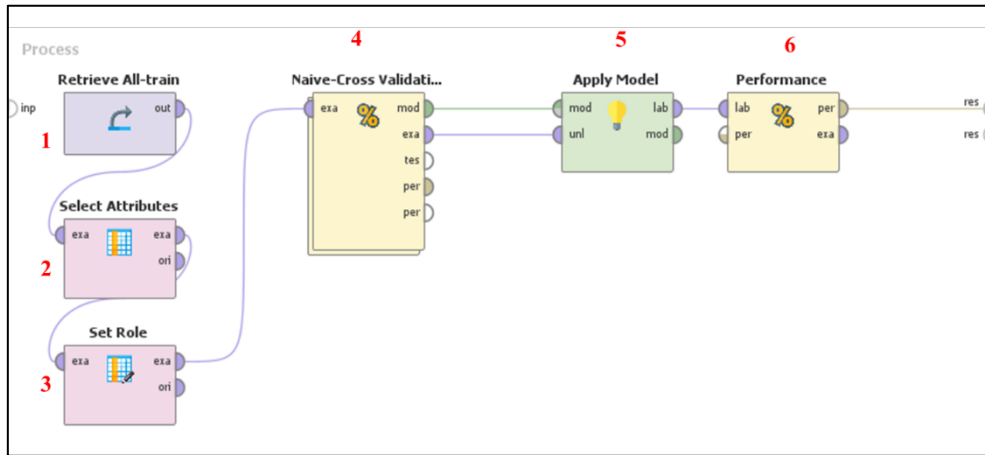


ภาพที่ 3.41 การสร้างแบบจำลองจากอัลกอริทึม Neural Networks ร่วมกับ AdaBoost เพื่อค้นหาค่า Hidden layer size ด้วยวิธี Random Search

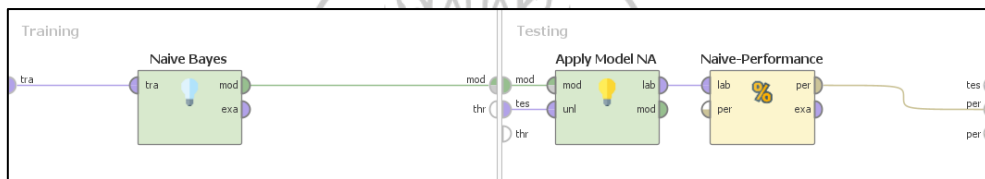
3.6 การสร้างแบบจำลองจากอัลกอริทึมมาตรฐาน

ในขั้นตอนนี้ผู้วิจัยจะทำการสร้างแบบจำลองจากอัลกอริทึมมาตรฐาน โดยปรับค่าต่างๆของแต่ละแบบจำลองตามผลลัพธ์ที่ได้มาจากขั้นตอนก่อนหน้า รวมถึงนำเข้าสู่ข้อมูลของงานวิจัย ทั้ง 5 กลุ่มข้อมูล คือ กลุ่มข้อมูลที่ 1 รวมสนามบินภาคเหนือ กลุ่มข้อมูลที่ 2 สนามบินภาคเหนือตอนบน กลุ่มข้อมูลที่ 3 สนามบินภาคเหนือตอนล่าง กลุ่มข้อมูลที่ 4 สนามบินที่มีภูเขาล้อมรอบ 1-2 ด้าน และกลุ่มข้อมูลที่ 5 สนามบินที่มีภูเขาล้อมรอบ 3-4 ด้าน ซึ่งมีวิธีการสร้างแบบจำลองดังต่อไปนี้

3.6.1 การสร้างแบบจำลองจากอัลกอริทึม Naive Bayes



ภาพที่ 3.42 ขั้นตอนการสร้างแบบจำลองจาก Naive Bayes

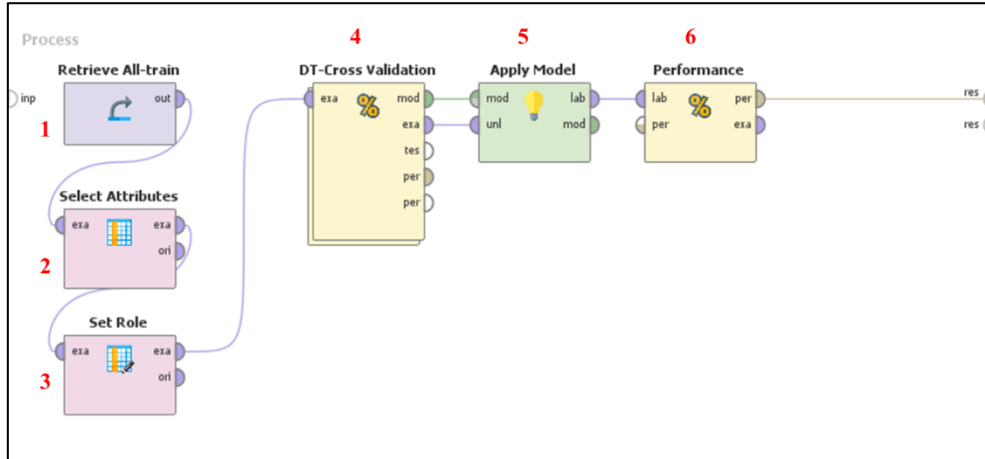


ภาพที่ 3.43 Sub-Process ภายในโอเปอร์เรเตอร์ Naive-Cross Validation

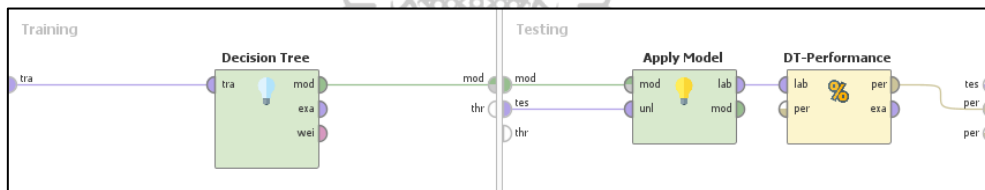
จากภาพที่ 3.42 สามารถแสดงได้ดังสามารถอธิบายขั้นตอนของโอเปอร์เรเตอร์ตามหมายเลข 1-6 ได้ดังนี้

1. Retrieve นำเข้าข้อมูลของแต่ละกลุ่มข้อมูลเพื่อสร้างแบบจำลอง
2. Select Attributes เลือกแอตทริบิวต์ที่นำมาใช้ทำนาย โดยเลือกจากผลลัพธ์ของขั้นตอนในการคัดเลือกคุณลักษณะ
3. Set Role เลือกคลาสค่าตอบที่เป็นระยะเวลาพยากรณ์ คือ 1 หรือ 2 ชั่วโมง
4. Cross Validation ตรวจสอบความแม่นยำของแบบจำลองจากอัลกอริทึม Naive Bayes โดยตั้งค่า folds เท่ากับ 10 และ Sub-Process ภายในโอเปอร์เรเตอร์ ดังภาพที่ 3.43
5. Apply Model นำแบบจำลองที่สร้างขึ้นมาใช้ทำนายข้อมูลทดสอบ
6. Performance ทดสอบประสิทธิภาพของแบบจำลองที่สร้างขึ้น

3.6.2 การสร้างแบบจำลองจากอัลกอริทึม Decision Tree



ภาพที่ 3.44 วิธีการสร้างแบบจำลองจาก Decision Tree

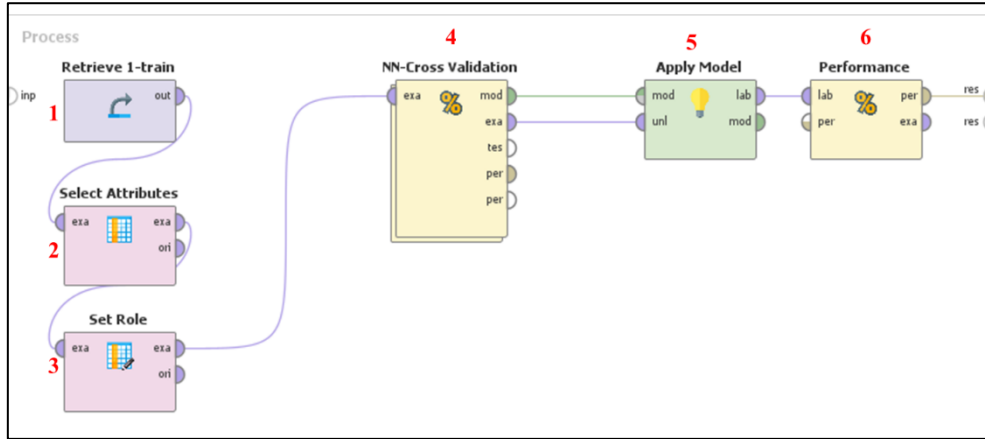


ภาพที่ 3.45 Sub-Process ภายในของแบบจำลองจาก Decision Tree

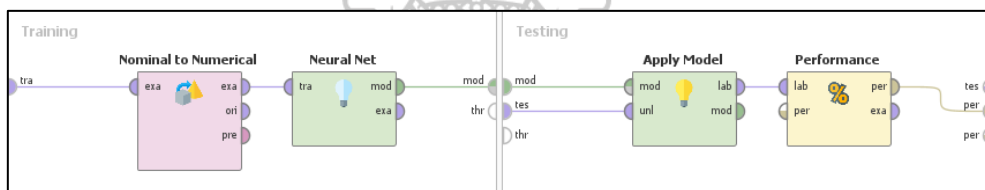
จากภาพที่ 3.44 สามารถแสดงได้ดังสามารถอธิบายขั้นตอนของโอเปอร์เรเตอร์ตามหมายเลข 1-6 ได้ดังนี้

- 1.Retrieve นำเข้าข้อมูลของแต่ละกลุ่มข้อมูลเพื่อสร้างแบบจำลอง
- 2.Select Attributes เลือกแอตทริบิวต์ที่นำมาใช้ทำนาย โดยเลือกจากผลลัพธ์ของขั้นตอนในการคัดเลือกคุณลักษณะ
- 3.Set Role เลือกคลาสค่าตอบระยะเวลาพยากรณ์ คือ 1 หรือ 2 ชั่วโมง
- 4.Cross Validation ตรวจสอบความแม่นยำของแบบจำลองจากอัลกอริทึม Decision Tree โดยตั้งค่า folds เท่ากับ 10 และ Sub-Process ภายในโอเปอร์เรเตอร์ ดังภาพที่ 3.45
- 5.Apply Model นำแบบจำลองที่สร้างขึ้นมาใช้ทำนายข้อมูลทดสอบ
- 6.Performance ทดสอบประสิทธิภาพของแบบจำลองที่สร้างขึ้น

3.6.3 การสร้างแบบจำลองจากอัลกอริทึม Neural Networks



ภาพที่ 3.46 วิธีการสร้างแบบจำลองจาก Neural Networks



ภาพที่ 3.47 Sub-Process ภายในของแบบจำลองจาก Neural Networks

จากภาพที่ 3.46 สามารถแสดงได้ดังสามารถอธิบายขั้นตอนของโอเปอร์เรเตอร์ตามหมายเลข 1-6 ได้ดังนี้

1.Retrieve นำเข้าข้อมูลของแต่ละกลุ่มข้อมูลเพื่อสร้างแบบจำลอง
 2.Select Attributes เลือกแอตทริบิวต์ที่นำมาใช้ทำนาย โดยเลือกจากผลลัพธ์ของขั้นตอนในการคัดเลือกคุณลักษณะ

3.Set Role เลือกคลาสค่าตอบที่เป็นระยะเวลาพยากรณ์ คือ 1 หรือ 2 ชั่วโมง

4.Cross Validation ตรวจสอบความแม่นยำของแบบจำลองจากอัลกอริทึม Neural Networks โดยตั้งค่า folds เท่ากับ 10 และ Sub-Process ภายในโอเปอร์เรเตอร์ดังภาพที่ 3.47 โดยมีโอเปอร์เรเตอร์ Nominal to Numerical เพื่อแปลงข้อมูลที่เป็นข้อความให้เป็นตัวเลข เนื่องจากอัลกอริทึม Neural Networks ใช้ค่าตัวเลขในการคำนวณและประมวลผลข้อมูล

5. Apply Model นำแบบจำลองที่สร้างขึ้นมาใช้ทำนายข้อมูลทดสอบ

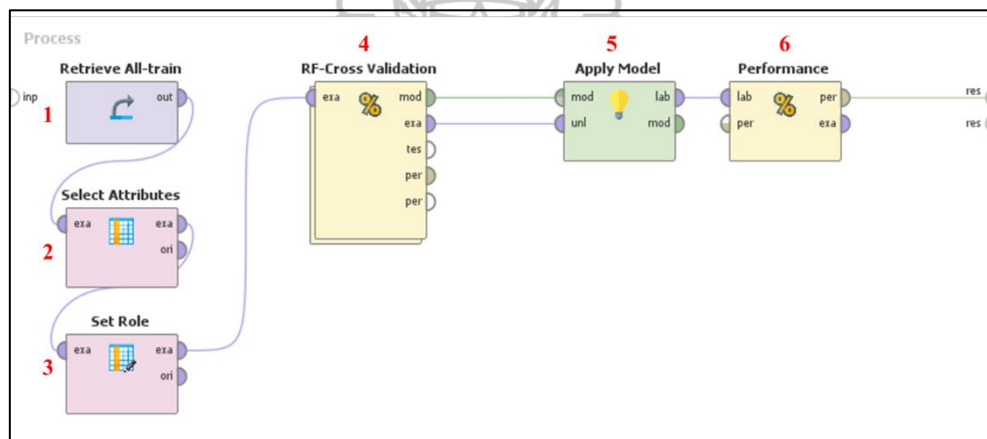
6. Performance ทดสอบประสิทธิภาพของแบบจำลองที่สร้างขึ้น

3.7 การสร้างแบบจำลองจากอัลกอริทึมการเรียนรู้แบบกลุ่ม

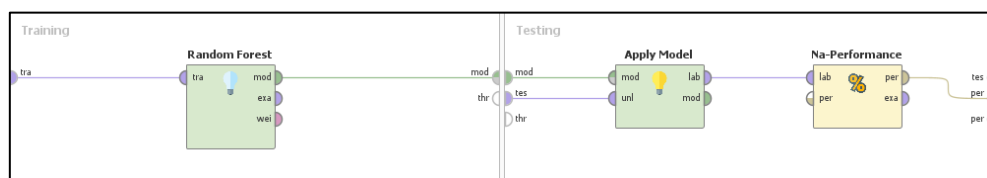
ผู้วิจัยพบว่า การสร้างแบบจำลองโดยใช้อัลกอริทึมมาตรฐานเพียงอย่างเดียวยังไม่สามารถทำนายคลาสเกิดฝนฟ้าคะนอง (TSRA Class) ได้อย่างแม่นยำเท่าที่ควร จึงต้องการศึกษาการสร้างแบบจำลองจากอัลกอริทึมมาตรฐานร่วมกับอัลกอริทึมการเรียนรู้แบบกลุ่มให้ทำนายผลว่าสามารถทำให้ประสิทธิภาพในการทำนายคลาส positive เพิ่มขึ้นได้หรือไม่ โดยการประเมินประสิทธิภาพจะดำเนินการเช่นเดียวกับการทดสอบอัลกอริทึมมาตรฐาน ใช้ 3 เทคนิคดังนี้

3.7.1 Random Forest

สร้างแบบจำลองจากอัลกอริทึม Random Forest โดยปรับค่าต่างๆของแต่ละแบบจำลองตามผลลัพธ์ที่ได้มาจากการปรับค่าพารามิเตอร์ที่เหมาะสม



ภาพที่ 3.48 การสร้างแบบจำลองการเรียนรู้แบบรวมกลุ่มเทคนิค Random Forest



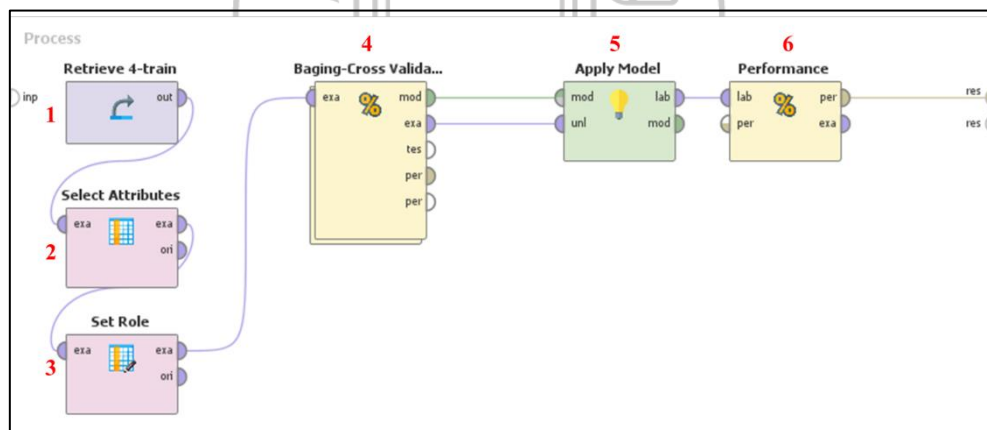
ภาพที่ 3.49 Sub-Process ภายในของเทคนิค Random Forest

จากภาพที่ 3.48 สามารถแสดงได้ดังสามารถอธิบายขั้นตอนของโอเปอเรเตอร์ตามหมายเลข 1-6 ได้ดังนี้

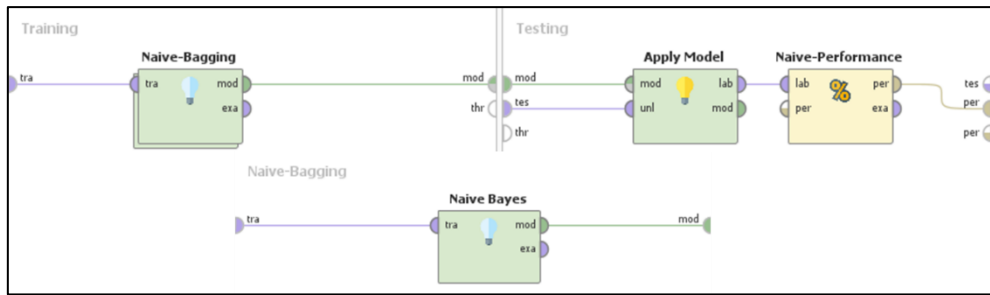
1. Retrieve นำเข้าข้อมูลของแต่ละกลุ่มข้อมูลเพื่อสร้างแบบจำลอง
2. Select Attributes เลือกแอตทริบิวต์ที่นำมาใช้ทำนาย โดยเลือกจากผลลัพธ์ของขั้นตอนในการคัดเลือกคุณลักษณะ
3. Set Role เลือกคลาสค่าต่อระยะเวลาพยากรณ์ คือ 1 หรือ 2 ชั่วโมง
4. Cross Validation ตรวจสอบความแม่นยำของแบบจำลองจากอัลกอริทึม Random Forest โดยตั้งค่า folds เท่ากับ 10 และ Sub-Process ภายในโอเปอเรเตอร์ดังภาพที่ 3.49
5. Apply Model นำแบบจำลองที่สร้างขึ้นมาใช้ทำนายข้อมูลทดสอบ
6. Performance ทดสอบประสิทธิภาพของแบบจำลองที่สร้างขึ้น

3.7.2 Bagging

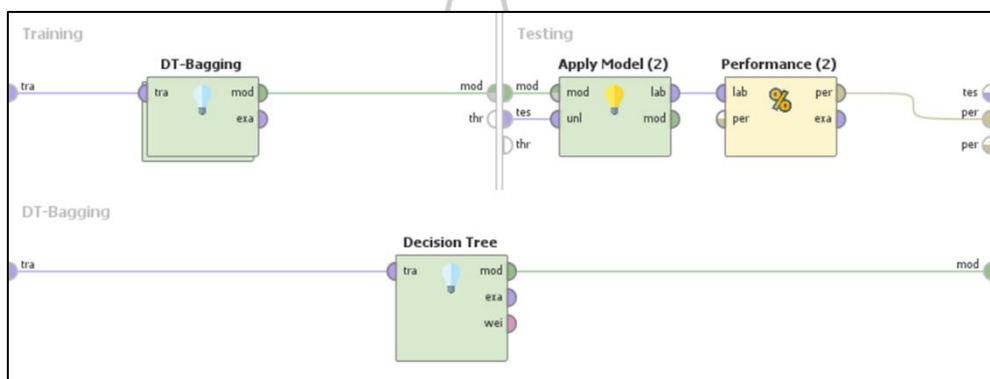
นำค่าพารามิเตอร์ที่เหมาะสมสำหรับแต่ละอัลกอริทึมที่ได้ดำเนินการในขั้นตอนที่ผ่านมาเพื่อสร้างแบบจำลองจากเทคนิค Bagging



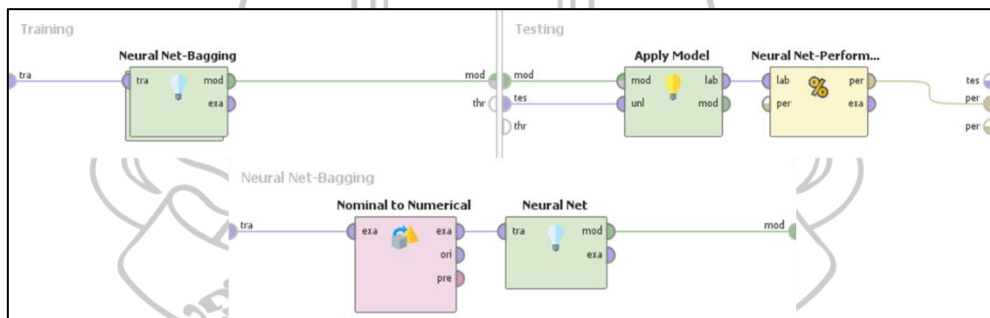
ภาพที่ 3.50 การสร้างแบบจำลองการเรียนรู้แบบรวมกลุ่มเทคนิค Bagging



ภาพที่ 3.51 Sub-Process ภายในเทคนิค Bagging ร่วมกับ Naive Bayes



ภาพที่ 3.52 Sub-Process ภายในเทคนิค Bagging ร่วมกับ Decision Tree



ภาพที่ 3.53 Sub-Process ภายในเทคนิค Bagging ร่วมกับ Neural Networks

จากภาพที่ 3.50 สามารถแสดงได้ดังสามารถอธิบายขั้นตอนของโอเปอร์เรเตอร์ตามหมายเลข 1-6 ได้ดังนี้

1. Retrieve นำเข้าข้อมูลของแต่ละกลุ่มข้อมูลเพื่อสร้างแบบจำลอง
2. Select Attributes เลือกแอตทริบิวต์ที่นำมาใช้ทำนาย โดยเลือกจากผลลัพธ์ของขั้นตอนในการคัดเลือกคุณลักษณะ

3. Set Role เลือกคลาสคำตอบที่เป็นระยะเวลาพยากรณ์ คือ 1 หรือ 2 ชั่วโมง

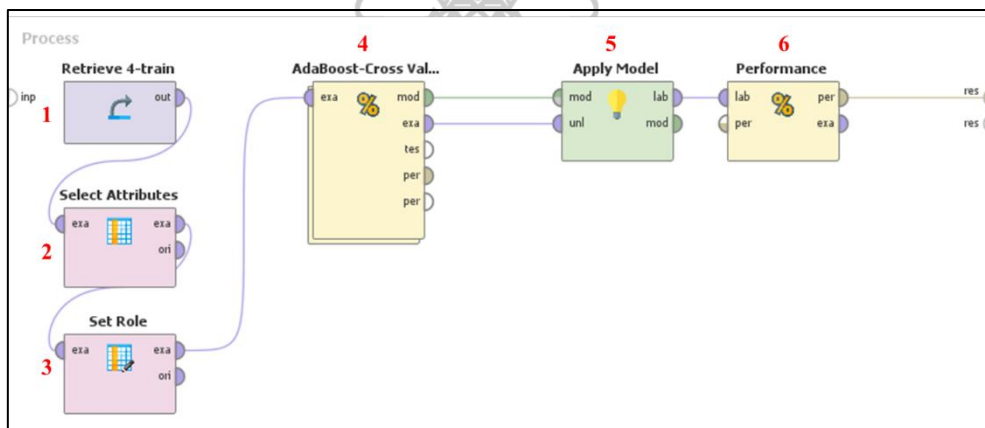
4. Cross Validation ตรวจสอบความแม่นยำของแบบจำลองจากแต่ละอัลกอริทึม โดยตั้งค่า folds เท่ากับ 10 และ Sub-Process ภายในโอเพอร์เรเตอร์ดังภาพที่ 3.51-3.53 โดยภายในโอเพอร์เรเตอร์ Bagging มีอัลกอริทึมที่ทำงานร่วมกัน

5. Apply Model นำแบบจำลองที่สร้างขึ้นมาใช้นำข้อมูลทดสอบ

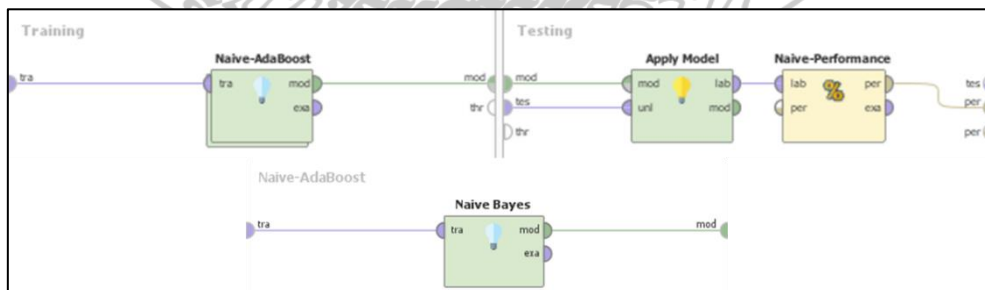
6. Performance ทดสอบประสิทธิภาพของแบบจำลองที่สร้างขึ้น

3.7.3 AdaBoost

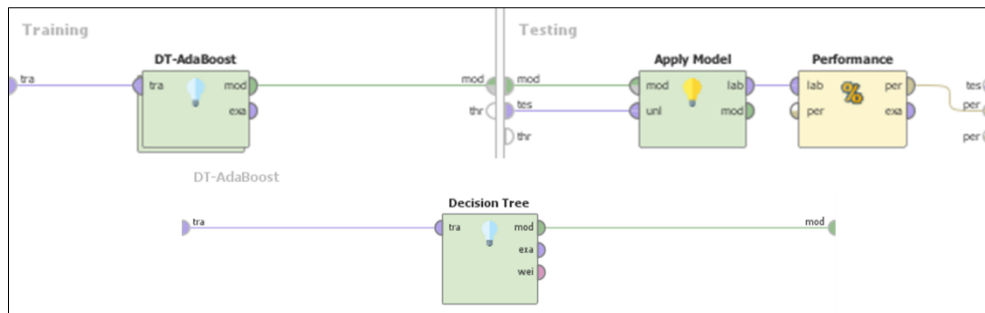
นำค่าพารามิเตอร์ที่เหมาะสมสำหรับแต่ละอัลกอริทึมที่ได้ดำเนินการในขั้นตอนที่ผ่านมาเพื่อสร้างแบบจำลองจากเทคนิค AdaBoost



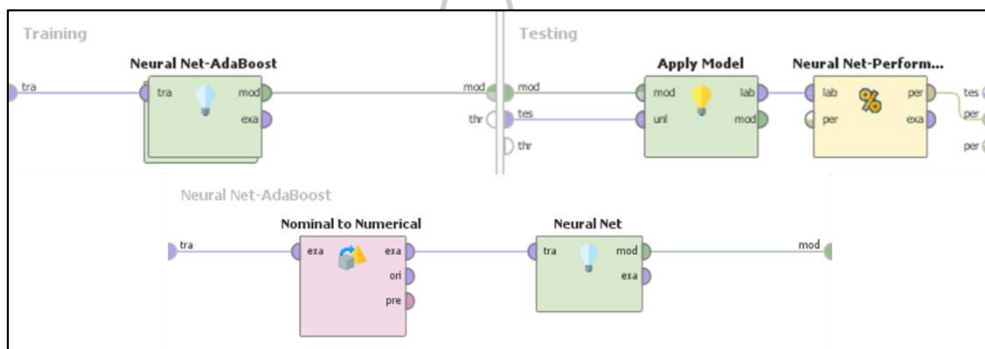
ภาพที่ 3.54 การสร้างแบบจำลองการเรียนรู้แบบรวมกลุ่มเทคนิค AdaBoost



ภาพที่ 3.55 Sub-Process ภายในของเทคนิค AdaBoost ร่วมกับ Naive Bayes



ภาพที่ 3.56 Sub-Process ภายในเทคนิค AdaBoost ร่วมกับ Decision Tree



ภาพที่ 3.57 Sub-Process ภายในของเทคนิค AdaBoost ร่วมกับ Neural Networks

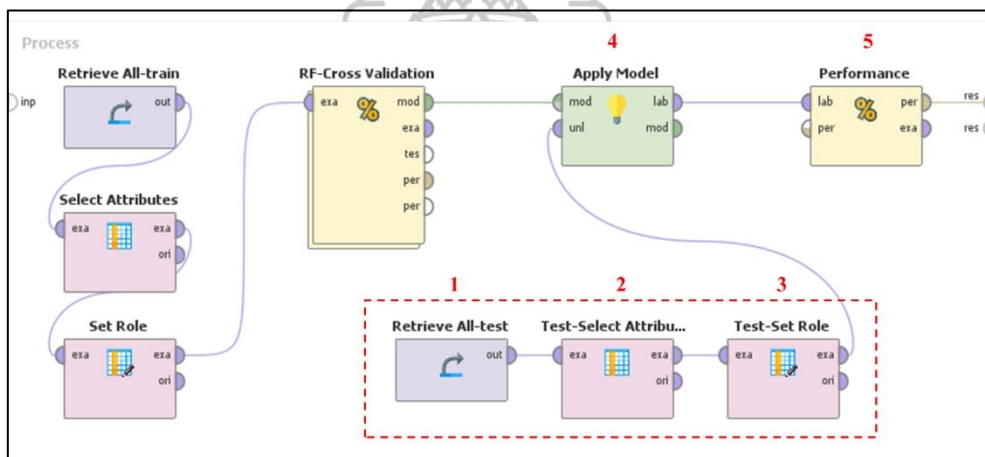
จากภาพที่ 3.54 สามารถแสดงได้ดังสามารถอธิบายขั้นตอนของโอเปอร์เรเตอร์ตามหมายเลข 1-6 ได้ดังนี้

- 1.Retrieve นำเข้าข้อมูลของแต่ละกลุ่มข้อมูลเพื่อสร้างแบบจำลอง
- 2.Select Attributes เลือกแอตทริบิวต์ที่นำมาใช้ทำนาย โดยเลือกจากผลลัพธ์ของขั้นตอนในการคัดเลือกคุณลักษณะ
- 3.Set Role เลือกคลาสคำตอบที่เป็นระยะเวลาพยากรณ์ คือ 1 หรือ 2 ชั่วโมง
- 4.Cross Validation ตรวจสอบความแม่นยำของแบบจำลองจากแต่ละอัลกอริทึม โดยตั้งค่า folds เท่ากับ 10 และ Sub-Process ภายในโอเปอร์เรเตอร์ภาพที่ 3.55-3.57 โดยภายในโอเปอร์เรเตอร์ AdaBoost มีอัลกอริทึมที่ทำงานร่วมกัน
5. Apply Model นำแบบจำลองที่สร้างขึ้นมาใช้ทำนายข้อมูลทดสอบ
- 6.Performance ทดสอบประสิทธิภาพของแบบจำลองที่สร้างขึ้น

4. การประเมินประสิทธิภาพของแบบจำลอง (Evaluation)

4.1 การวัดประสิทธิภาพของแบบจำลอง

ดำเนินการโดยนำชุดข้อมูลทดสอบ (Testing Data) ที่แบบจำลองไม่เคยเรียนรู้มาก่อน ซึ่งเป็นข้อมูลในปีพ.ศ.2564-2565 รวม 2 ปี นำเข้าสู่แบบจำลอง เพื่อทดสอบแต่ละอัลกอริทึมว่าสามารถทำงานได้ดีกับชุดข้อมูลที่ไม่เคยพบมาก่อนหรือไม่ และเปรียบเทียบประสิทธิภาพของแบบจำลองด้วยวิธี 10-Fold Cross Validation โดยวัดประสิทธิภาพแบบจำลองทั้งหมดด้วยค่า Accuracy Precision Recall F-measure AUC และ Balanced Accuracy เพื่อพิจารณาประสิทธิภาพของแบบจำลองโดยภาพรวม ซึ่งผู้วิจัยได้เพิ่มขั้นตอนประเมินประสิทธิภาพลงในขั้นตอนการสร้างแบบจำลองเริ่มต้น แสดงดังภาพที่ 3.58



ภาพที่ 3.58 ตัวอย่างการวัดประสิทธิภาพของแบบจำลองของอัลกอริทึม Random Forest

จากภาพที่ 3.58 สามารถอธิบายขั้นตอนการประเมินประสิทธิภาพของแบบจำลองตามหมายเลข 1-5 ได้ดังนี้

1. Retrieve นำเข้าสู่ชุดข้อมูลทดสอบแบบจำลอง
2. Select Attributes เลือกแอตทริบิวต์ที่นำมาใช้ทำนาย โดยเลือกจากผลลัพธ์ของขั้นตอนในการคัดเลือกคุณลักษณะ
3. Set Role เลือกคลาสคำตอบที่ระยะเวลาพยากรณ์ คือ 1 หรือ 2 ชั่วโมง
4. Apply Model นำแบบจำลองที่สร้างขึ้นมาใช้ทำนายข้อมูลทดสอบ
5. Performance ทดสอบประสิทธิภาพของแบบจำลองที่สร้างขึ้น

4.2 การคัดเลือกแบบจำลองที่ดีที่สุด

ในงานวิจัยนี้มีการคัดเลือกแบบจำลองที่ดีที่สุดสำหรับแต่ละกลุ่มข้อมูล โดยพิจารณา ค่า F-measure สูงสุด ซึ่งจะสามารถวัดผลความถ่วงดุลของแบบจำลองที่หามาได้จากการคำนวณค่า Precision และ Recall ทำให้บ่งชี้การทำนายคลาสเกิดพายุฝนฟ้าคะนอง (TSRA Class) ได้ ประกอบกับการพิจารณาวัดประสิทธิภาพแบบจำลองทั้งหมดด้วยค่า Accuracy Precision Recall AUC และ Balanced Accuracy เพื่อพิจารณาประสิทธิภาพของแบบจำลองโดยภาพรวม ในงานวิจัยนี้จะสรุปได้แบบจำลองที่เหมาะสมสำหรับแต่ละกลุ่มข้อมูล และ 2 ระยะเวลา ดังนี้

1. แบบจำลองรวมที่เหมาะสมสำหรับ 9 สนามบินของภาคเหนือ ระยะเวลาพยากรณ์ คือ 1 และ 2 ชั่วโมงข้างหน้า รวม 2 แบบจำลอง
2. แบบจำลองที่เหมาะสมสำหรับสนามบินภาคเหนือตอนบน 2 ระยะเวลาพยากรณ์ คือ 1 และ 2 ชั่วโมงข้างหน้า รวมเป็น 2 แบบจำลอง
3. แบบจำลองที่เหมาะสมสำหรับสนามบินภาคเหนือตอนล่าง 2 ระยะเวลาพยากรณ์ คือ 1 และ 2 ชั่วโมงข้างหน้า รวมเป็น 2 แบบจำลอง
4. แบบจำลองที่เหมาะสมสำหรับสนามบินที่มีภูเขาล้อมรอบ 1-2 ด้าน ระยะเวลาพยากรณ์ คือ 1 และ 2 ชั่วโมงข้างหน้า รวม 2 แบบจำลอง
5. แบบจำลองที่เหมาะสมสำหรับสนามบินที่มีภูเขาล้อมรอบ 3-4 ด้าน ระยะเวลาพยากรณ์ คือ 1 และ 2 ชั่วโมงข้างหน้า รวม 2 แบบจำลอง

4.3 การนำแบบจำลองไปใช้งานจริง

เมื่อได้แบบจำลองที่ดีที่สุดของทั้ง 5 กลุ่มข้อมูลทั้งระยะเวลาพยากรณ์ 1 และ 2 ชั่วโมงข้างหน้าแล้ว ขั้นตอนนี้จะเป็นการนำแบบจำลองที่ได้ไปใช้จริง โดยการทดสอบนำเข้าข้อมูลข่าวอากาศการบินในปีพ.ศ. 2566 ระยะเวลา 1 ปี และทำการวัดประสิทธิภาพด้วยค่า Accuracy Precision Recall F-measure AUC และ Balanced Accuracy โดยทำการเปรียบเทียบจำนวนครั้งของผลการทำนายคลาสิกการเกิดฝนฟ้าคะนองได้ถูกต้อง (True Positive) จากตาราง Confusion matrix เปรียบเทียบกับจำนวนครั้งที่ทำนายถูกต้องจากนักอุตุนิยมวิทยาผู้ปฏิบัติงานจริงจากการออกข่าวพยากรณ์แนวโน้มสภาพอากาศร้าย (Trend Forecast) ทำให้สามารถทราบได้ว่าการพยากรณ์ด้วยแบบจำลองหรือด้วยผู้ปฏิบัติงานจริงวิธีการใดมีความแม่นยำในการทำนายการเกิดพายุฝนฟ้าคะนองได้ดีกว่า โดยจำนวนข้อมูลข่าวอากาศการบินในปีพ.ศ. 2566 ที่ทำการทดสอบในขั้นตอนนี้ดังตารางที่ 3.13

ตารางที่ 3.13 จำนวนข้อมูลข่าวอากาศการบินในปีพ.ศ. 2566 ที่นำเข้าเพื่อทดสอบแบบจำลอง

กลุ่มข้อมูล	จำนวนชุด ข้อมูล	1 ชั่วโมง		2 ชั่วโมง	
		NO TSRA	TSRA	NO TSRA	TSRA
รวมสนามบินภาคเหนือ	55,441	54,847	594	54,901	540
ภาคเหนือตอนบน	38,773	38,293	480	38,293	480
ภาคเหนือตอนล่าง	16,668	16,554	114	16,608	60
ภูเขาล้อมรอบ 1-2 ด้าน	11,594	11,498	96	11,552	42
ภูเขาล้อมรอบ 3-4 ด้าน	43,847	43,349	498	43,349	498



บทที่ 4

ผลการวิเคราะห์ข้อมูล

การวิจัย เรื่อง การพยากรณ์การเกิดพายุฝนฟ้าคะนองระยะปัจจุบันจากข่าวอากาศการ บิน ด้วยเทคนิคการเรียนรู้ของเครื่อง กรณีศึกษาสนามบินภาคเหนือของประเทศไทย ผู้วิจัยได้ ดำเนินการตามกรอบการวิจัยเพื่อให้ได้ผลการวิเคราะห์ตามวัตถุประสงค์ โดยมีผลการวิเคราะห์ข้อมูล ดังนี้

1. ผลการวิเคราะห์สำหรับการคัดเลือกคุณลักษณะ
2. ผลการปรับค่าพารามิเตอร์ที่เหมาะสมในงานวิจัย
3. ผลการประเมินประสิทธิภาพแบบจำลอง
4. สรุปแบบจำลองที่เหมาะสมสำหรับแต่ละพื้นที่สนามบิน
5. การทดสอบนำแบบจำลองไปใช้งานจริง

1. ผลการวิเคราะห์สำหรับการคัดเลือกคุณลักษณะ

1.1 ผู้วิจัยใช้วิธีหาค่าน้ำหนักความเกี่ยวข้องของตัวแปรที่สนใจด้วยวิธี Information Gain และ Chi-Square ของทั้งหมด 10 แอตทริบิวต์ที่เป็นข้อมูลนำเข้าเพื่อสร้างแบบจำลอง พบว่า สำหรับ ระยะเวลาพยากรณ์ 1 ชั่วโมงข้างหน้า การเรียงลำดับแอตทริบิวต์ที่มีความสำคัญมากที่สุดเหมือนกัน ลำดับที่ 1-5 ได้แก่ เมฆฝนฟ้าคะนอง สภาพอากาศปัจจุบัน ความเร็วลม เดือน และช่วงเวลา โดยที่ทั้ง 2 วิธีการหาค่าน้ำหนักความเกี่ยวข้องของตัวแปรในแต่ละกลุ่มข้อมูลมีการเรียงลำดับความสำคัญจาก ลำดับ 1-10 เหมือนกัน ยกเว้นกลุ่มข้อมูลรวมสนามบินภาคเหนือที่มีการสลับลำดับความสำคัญใน ลำดับที่ 7 และ 8 ระหว่างแอตทริบิวต์การเกิดลมกระโชกและทิศทางลมแต่จะพบว่า ค่าน้ำหนักจาก การคำนวณของแอตทริบิวต์ทั้ง 2 ลำดับนั้นมีค่าใกล้เคียงกันมาก แต่สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมงข้างหน้า แอตทริบิวต์ที่มีความสำคัญมากที่สุดไม่แน่นอนเปลี่ยนแปลงไปตามแต่ละกลุ่มข้อมูล โดยสำหรับลำดับความสำคัญทั้ง 2 วิธีในบางกลุ่มข้อมูลมีการเรียงลำดับไม่ตรงกัน เนื่องจากในแต่ละ ลำดับของแอตทริบิวต์ที่สลับกันค่าน้ำหนักใกล้เคียงกันมาก โดยค่าน้ำหนักความเกี่ยวข้องของตัวแปรมี ค่าน้อยกว่าเมื่อเปรียบเทียบกับระยะเวลาพยากรณ์ 1 ชั่วโมงข้างหน้า ซึ่งลำดับค่าน้ำหนักของแต่ละแอต ทริบิวต์เรียงลำดับจากมากที่สุดไปน้อยที่สุด แสดงดังตารางที่ 4.1 ถึงตารางที่ 4.5

ตารางที่ 4.1 คำนวณน้ำหนักการคัดเลือกคุณลักษณะด้วยวิธี Information Gain และ Chi-Square
สำหรับกลุ่มข้อมูลรวมสนามบินภาคเหนือ

ระยะเวลา	ลำดับ	Information Gain		Chi-Square	
		แอตทริบิวต์	ค่าน้ำหนัก	แอตทริบิวต์	ค่าน้ำหนัก
1 ชั่วโมง	1	M_Cloud	0.40	M_Cloud	230,307
	2	WW	0.38	WW	227,159
	3	W_speed	0.30	W_speed	173,046
	4	Month	0.17	Month	107,991
	5	Time_of_day	0.13	Time_of_day	85,178
	6	QNH	0.08	QNH	55,525
	7	GUST	0.08	W_direct	46,719
	8	W_direct	0.07	GUST	44,131
	9	Humidity	0.00	Humidity	1,288
	10	Visibility	0.00	Visibility	433
2 ชั่วโมง	1	Month	0.17	Month	103,574
	2	WW	0.16	WW	97,811
	3	M_Cloud	0.16	M_Cloud	90,804
	4	W_speed	0.12	W_speed	75,700
	5	Time_of_day	0.10	Time_of_day	66,745
	6	QNH	0.09	QNH	58,986
	7	W_direct	0.04	W_direct	27,880
	8	GUST	0.02	GUST	11,724
	9	Visibility	0.01	Visibility	4,040
	10	Humidity	0.00	Humidity	1,574

ตารางที่ 4.2 ค่าน้ำหนักการคัดเลือกคุณลักษณะด้วยวิธี Information Gain และ Chi-Square สำหรับกลุ่มข้อมูลสนามบินภาคเหนือตอนบน

ระยะเวลา	ลำดับ	Information Gain		Chi-Square	
		แอตทริบิวต์	ค่าน้ำหนัก	แอตทริบิวต์	ค่าน้ำหนัก
1 ชั่วโมง	1	M_Cloud	0.38	M_Cloud	156,996
	2	WW	0.37	WW	156,982
	3	W_speed	0.28	W_speed	116,183
	4	Month	0.18	Month	81,400
	5	Time_of_day	0.12	Time_of_day	57,917
	6	QNH	0.09	QNH	43,325
	7	W_direct	0.08	W_direct	38,856
	8	GUST	0.07	GUST	27,794
	9	Humidity	0.00	Humidity	567
	10	Visibility	0.00	Visibility	205
2 ชั่วโมง	1	Month	0.17	Month	76,817
	2	WW	0.16	WW	69,906
	3	M_Cloud	0.16	M_Cloud	66,110
	4	W_speed	0.11	W_speed	50,525
	5	QNH	0.10	QNH	47,296
	6	Time_of_day	0.09	Time_of_day	46,134
	7	W_direct	0.05	W_direct	24,938
	8	GUST	0.02	GUST	6,939
	9	Visibility	0.01	Visibility	2,786
	10	Humidity	0.00	Humidity	1,358

ตารางที่ 4.3 คำนวณการคัดเลือกคุณลักษณะด้วยวิธี Information Gain และ Chi-Square
สำหรับกลุ่มข้อมูลสนามบินภาคเหนือตอนล่าง

ระยะเวลา	ลำดับ	Information Gain		Chi-Square	
		แอตทริบิวต์	ค่าน้ำหนัก	แอตทริบิวต์	ค่าน้ำหนัก
1 ชั่วโมง	1	M_Cloud	0.46	M_Cloud	75,400
	2	WW	0.43	WW	72,835
	3	W_speed	0.33	W_speed	52,561
	4	Month	0.19	Month	33,841
	5	Time_of_day	0.13	Time_of_day	24,962
	6	GUST	0.11	GUST	17,462
	7	QNH	0.09	QNH	14,606
	8	W_direct	0.05	W_direct	8,907
	9	Visibility	0.01	Visibility	1,018
	10	Humidity	0.00	Humidity	403
2 ชั่วโมง	1	Month	0.18	Month	32,642
	2	WW	0.16	WW	28,799
	3	M_Cloud	0.16	M_Cloud	25,592
	4	Time_of_day	0.12	Time_of_day	23,002
	5	QNH	0.09	QNH	14,746
	6	W_speed	0.08	W_speed	13,608
	7	W_direct	0.04	W_direct	7,753
	8	GUST	0.03	GUST	5,021
	9	Visibility	0.01	Visibility	1,262
	10	Humidity	0.00	Humidity	613

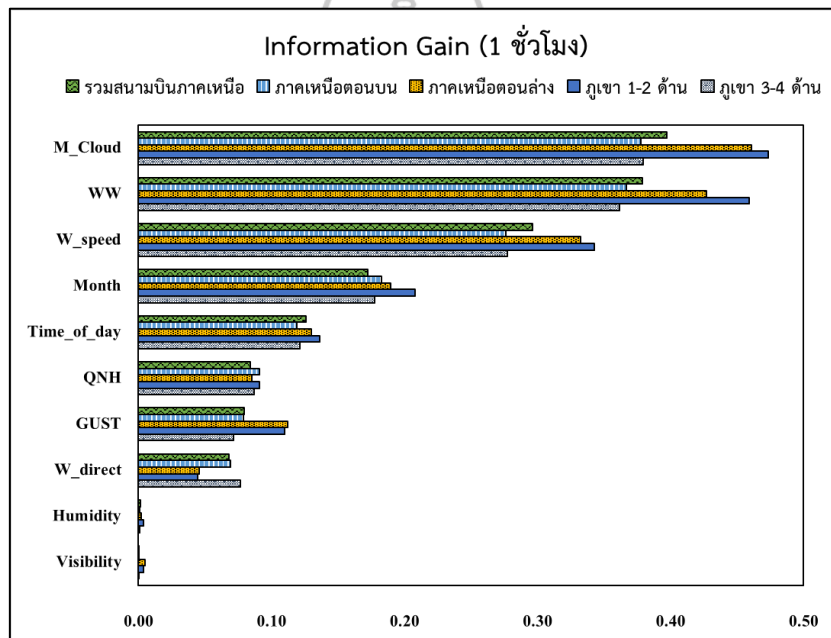
ตารางที่ 4.4 คำนวณการคัดเลือกคุณลักษณะด้วยวิธี Information Gain และ Chi-Square
สำหรับกลุ่มข้อมูลสนามบินที่มีภูเขาล้อมรอบ 1-2 ด้าน

ระยะเวลา	ลำดับ	Information Gain		Chi-Square	
		แอตทริบิวต์	ค่าน้ำหนัก	แอตทริบิวต์	ค่าน้ำหนัก
1 ชั่วโมง	1	M_Cloud	0.47	M_Cloud	58,821
	2	WW	0.46	WW	58,677
	3	W_speed	0.34	W_speed	41,327
	4	Month	0.21	Month	28,002
	5	Time_of_day	0.14	Time_of_day	19,734
	6	GUST	0.11	GUST	12,873
	7	QNH	0.09	QNH	11,204
	8	W_direct	0.04	W_direct	6,651
	9	Visibility	0.00	Visibility	594
	10	Humidity	0.00	Humidity	558
2 ชั่วโมง	1	Month	0.20	Month	26,543
	2	WW	0.19	WW	24,345
	3	M_Cloud	0.16	M_Cloud	19,978
	4	Time_of_day	0.13	Time_of_day	18,961
	5	QNH	0.09	QNH	11,077
	6	W_speed	0.08	W_speed	9,778
	7	W_direct	0.03	W_direct	5,214
	8	GUST	0.03	GUST	4,040
	9	Visibility	0.01	Visibility	1,471
	10	Humidity	0.00	Humidity	370

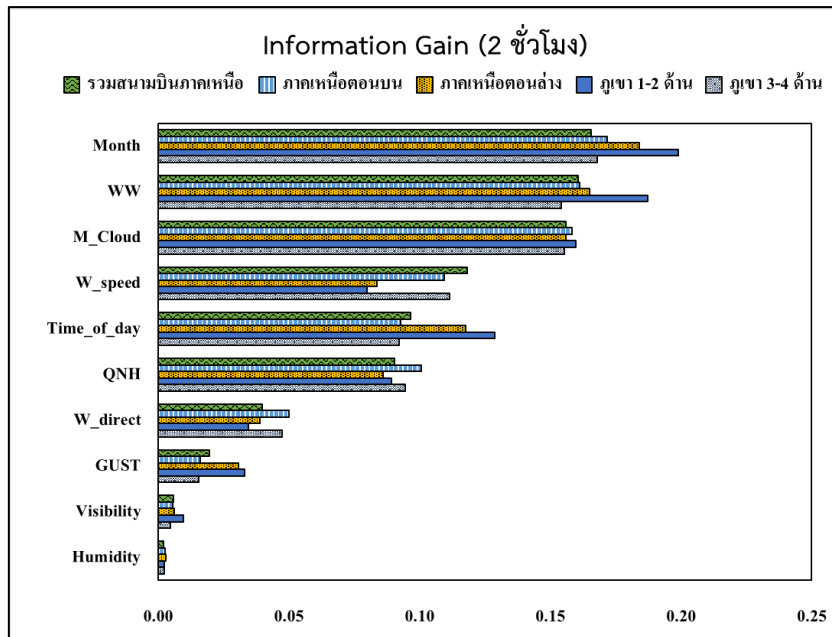
ตารางที่ 4.5 ค่าน้ำหนักการคัดเลือกคุณลักษณะด้วยวิธี Information Gain และ Chi-Square สำหรับกลุ่มข้อมูลสนามบินที่มีภูเขาล้อมรอบ 3-4 ด้าน

ระยะเวลา	ลำดับ	Information Gain		Chi-Square	
		แอตทริบิวต์	ค่าน้ำหนัก	แอตทริบิวต์	ค่าน้ำหนัก
1 ชั่วโมง	1	M_Cloud	0.38	M_Cloud	172,593
	2	WW	0.36	WW	170,402
	3	W_speed	0.28	W_speed	127,907
	4	Month	0.18	Month	87,180
	5	Time_of_day	0.12	Time_of_day	64,528
	6	QNH	0.09	QNH	45,388
	7	W_direct	0.08	W_direct	41,378
	8	GUST	0.07	GUST	31,492
	9	Humidity	0.00	Humidity	716
	10	Visibility	0.00	Visibility	249
2 ชั่วโมง	1	Month	0.17	Month	82,439
	2	M_Cloud	0.16	WW	73,816
	3	WW	0.15	M_Cloud	70,992
	4	W_speed	0.11	W_speed	56,531
	5	QNH	0.09	Time_of_day	50,211
	6	Time_of_day	0.09	QNH	49,034
	7	W_direct	0.05	W_direct	25,995
	8	GUST	0.02	GUST	7,441
	9	Visibility	0.00	Visibility	2,656
	10	Humidity	0.00	Humidity	1,376

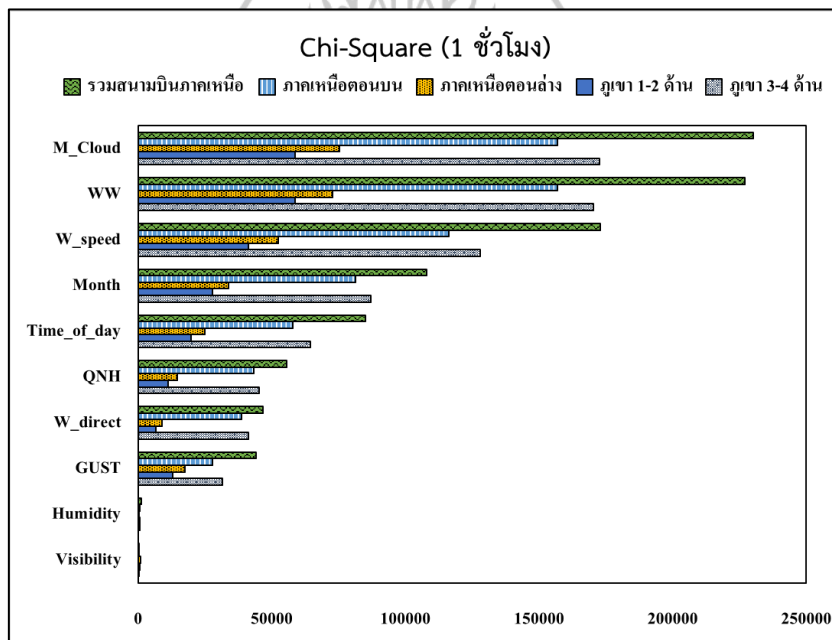
1.2 จากตารางที่ 4.1-4.5 การคัดเลือกคุณลักษณะสำหรับระยะเวลาพยากรณ์ 1 และ 2 ชั่วโมงข้างหน้า ผู้วิจัยพบว่า ลำดับค่าน้ำหนักความเกี่ยวข้องของตัวแปรที่สนใจในลำดับน้อยที่สุด คือ ลำดับที่ 9 และ 10 เหมือนกันในทุกๆแบบจำลองทั้ง 2 ระยะเวลาพยากรณ์ นั่นคือ ความชื้นสัมพัทธ์ (Humidity) และทัศนวิสัย (Visibility) และเมื่อพิจารณาตัดแอตทริบิวต์ที่มีค่าน้ำหนักความเกี่ยวข้องเท่ากับ 0.00 ออก รวมถึงได้พิจารณาแอตทริบิวต์ที่มีค่าในลำดับที่ 9 ซึ่งมีเปอร์เซ็นต์ส่วนต่างกับลำดับที่ 8 มากกว่าหรือเท่ากับ 100% ดังนั้นจึงได้ตัดแอตทริบิวต์ความชื้นสัมพัทธ์ (Humidity) และทัศนวิสัย (Visibility) ออกเนื่องจากได้ค่าน้ำหนักความเกี่ยวข้องกับตัวแปรน้อยที่สุดและมีค่าน้ำหนักต่างจากแอตทริบิวต์อื่นๆอย่างชัดเจนในทั้ง 2 วิธีคัดเลือกคุณลักษณะ ค่าน้ำหนักของแต่ละวิธีของการคัดเลือกคุณลักษณะแสดงดังภาพที่ 4.1-4.4



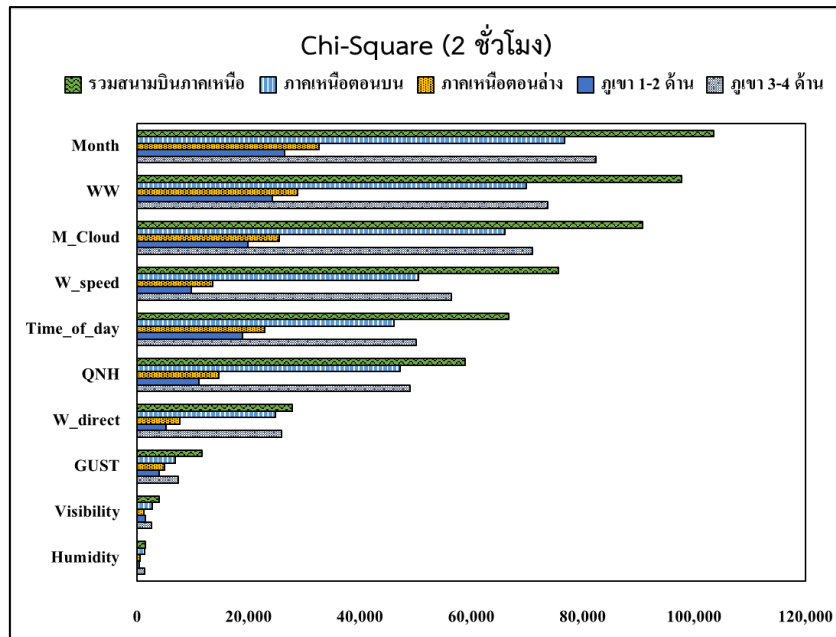
ภาพที่ 4.1 แผนภูมิแสดงค่าน้ำหนักด้วยวิธี Information Gain สำหรับระยะเวลาพยากรณ์ 1 ชั่วโมง



ภาพที่ 4.2 แผนภูมิแสดงค่าน้ำหนักด้วยวิธี Information Gain สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมง



ภาพที่ 4.3 แผนภูมิแสดงค่าน้ำหนักด้วยวิธี Chi-Square สำหรับระยะเวลาพยากรณ์ 1 ชั่วโมง



ภาพที่ 4.4 แผนภูมิแสดงค่าน้ำหนักด้วยวิธี Chi-Square สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมง

1.3 สรุปผลการคัดเลือกคุณลักษณะที่นำเข้าไปสร้างแบบจำลองสำหรับระยะเวลาพยากรณ์ 1 และ 2 ชั่วโมงข้างหน้า พบว่า มีแอตทริบิวต์ที่มีความเกี่ยวข้องกับตัวแปรที่สนใจและนำเข้าไปสร้างแบบจำลองจำนวน 8 แอตทริบิวต์ในทุกแบบจำลอง รายละเอียดดังตารางที่ 4.6

ตารางที่ 4.6 สรุปผลการคัดเลือกคุณลักษณะของทุกแบบจำลอง

คุณลักษณะ	คำอธิบาย	ประเภทข้อมูล	การนำเข้าสู่แบบจำลอง
M_Cloud	เมฆพายุฝนฟ้าคะนอง	Polynomial	นำเข้า
WW	สภาพอากาศปัจจุบัน	Polynomial	นำเข้า
W_speed	ความเร็วลม	Polynomial	นำเข้า
Month	เดือน	Polynomial	นำเข้า
Time_of_day	ช่วงเวลา	Polynomial	นำเข้า
QNH	ความกดอากาศ	Polynomial	นำเข้า
GUST	การเกิดลมกระโชก	Binomial	นำเข้า
W_direct	ทิศทางลม	Polynomial	นำเข้า
Humidity	ความชื้นสัมพัทธ์	Binomial	ไม่นำเข้า
Visibility	ทัศนวิสัย	Polynomial	ไม่นำเข้า

2. ผลการปรับค่าพารามิเตอร์ที่เหมาะสมในงานวิจัย

ในงานวิจัยนี้มีขั้นตอนของกระบวนการกำหนดค่า Hyperparameter ซึ่งมีการกำหนดโครงสร้างของแบบจำลองตั้งแต่เริ่มต้น เนื่องจากมีผลต่อประสิทธิภาพของแบบจำลอง ผลการหาค่าพารามิเตอร์ที่เหมาะสมของแต่ละอัลกอริทึม ดังนี้

2.1 สรุปผลการปรับค่าพารามิเตอร์สำหรับอัลกอริทึมมาตรฐาน

อัลกอริทึมมาตรฐาน ได้แก่ Decision Tree และ Neural Networks ทำการค้นหาค่าพารามิเตอร์ที่เหมาะสมในเบื้องต้นของแต่ละอัลกอริทึมด้วยวิธี Grid Search และใช้วิธี Random Search ในการหาค่า Hidden layer sizes ที่เหมาะสมสำหรับอัลกอริทึม Neural Networks เนื่องจากวิธี Grid Search ในโปรแกรม RapidMiner ไม่สามารถหาค่าที่เหมาะสมได้ โดยผู้วิจัยกำหนดค่าค้นหาตั้งแต่ 1-8 ผลการทดสอบสำหรับการหาค่า Hidden layer sizes แสดงดังภาคผนวก ก และผลการปรับค่าพารามิเตอร์ที่เหมาะสม สำหรับอัลกอริทึมมาตรฐานแสดงดังตารางที่ 4.7

ตารางที่ 4.7 ผลการปรับค่าพารามิเตอร์ที่เหมาะสม สำหรับอัลกอริทึมมาตรฐาน ระยะเวลาพยากรณ์ 1 และ 2 ชั่วโมง

อัลกอริทึม	พารามิเตอร์	ค่าที่เหมาะสม	
		1 ชม.	2 ชม.
กลุ่มข้อมูลที่ 1 รวมสนามบินภาคเหนือ			
Decision Tree	Criterion:	gain_ratio	gini_index
	Maximum Depth (maxDepth)	12	12
	minimal leaf size	2	4
	minimal size for split	2	2
Neural Networks	Training Cycle	282	140
Neural Networks	Learning Rate	0.1	0.1
	Momentum	0.7	0.9
	Hidden layer sizes	6	6

ตารางที่ 4.7 (ต่อ)

อัลกอริทึม	พารามิเตอร์	ค่าที่เหมาะสม	
		1 ชม.	2 ชม.
กลุ่มข้อมูลที่ 2 สนามบินภาคเหนือตอนบน			
Decision Tree	Criterion:	gini_index	gini_index
	Maximum Depth (maxDepth)	8	8
	minimal leaf size	4	4
	minimal size for split	2	4
Neural Networks	Training Cycle	80	170
	Learning Rate	0.1	0.1
	Momentum	0.5	0.8
	Hidden layer sizes	6	8
กลุ่มข้อมูลที่ 3 สนามบินภาคเหนือตอนล่าง			
Decision Tree	Criterion:	gain_ratio	gain_ratio
	Maximum Depth (maxDepth)	8	8
	minimal leaf size	2	2
	minimal size for split	2	2
Neural Networks	Training Cycle	100	290
	Learning Rate	0.1	0.1
	Momentum	0.6	0.6
	Hidden layer sizes	8	8
กลุ่มข้อมูลที่ 4 สนามบินที่มีภูเขาล้อมรอบ 1-2 ด้าน			
Decision Tree	Criterion:	gini_index	information_gain
	Maximum Depth (maxDepth)	12	8
	minimal leaf size	2	2
	minimal size for split	2	2

ตารางที่ 4.7 (ต่อ)

อัลกอริทึม	พารามิเตอร์	ค่าที่เหมาะสม	
		1 ชม.	2 ชม.
Neural Networks	Training Cycle	294	250
	Learning Rate	0.1	0.1
	Momentum	0.8	0.9
	Hidden layer sizes	6	6
กลุ่มข้อมูลที่ 5 สนามบินที่มีภูเขาล้อมรอบ 3-4 ด้าน			
Decision Tree	Criterion:	gini_index	gini_index
	Maximum Depth (maxDepth)	9	10
	minimal leaf size	2	4
	minimal size for split	2	2
Neural Networks	Training Cycle	260	180
	Learning Rate	0.1	0.1
	Momentum	0.7	0.9
	Hidden layer sizes	8	8

2.2 สรุปผลการปรับค่าสำหรับอัลกอริทึมการเรียนรู้แบบกลุ่ม

อัลกอริทึมการเรียนรู้แบบกลุ่ม ได้แก่ Random Forest อัลกอริทึมมาตรฐาน ร่วมกับเทคนิค Bagging และอัลกอริทึมมาตรฐานร่วมกับเทคนิค AdaBoost ได้ทำการค้นหาค่าพารามิเตอร์ที่เหมาะสมของแต่ละอัลกอริทึมด้วยวิธี Grid Search และใช้วิธี Random Search ในการหาค่า Hidden layer sizes ที่เหมาะสมสำหรับอัลกอริทึม Neural Networks ที่ทำงานร่วมกับอัลกอริทึมการเรียนรู้แบบกลุ่ม เนื่องจากวิธี Grid Search ในโปรแกรม RapidMiner ไม่สามารถหาค่าที่เหมาะสมได้ โดยผู้วิจัยกำหนดค่าค้นหาตั้งแต่ 1-8 ผลการทดสอบสำหรับการหาค่า Hidden layer sizes ดังภาคผนวก ก และผลการปรับค่าพารามิเตอร์ที่เหมาะสม สำหรับอัลกอริทึมการเรียนรู้แบบกลุ่มแสดงดังตารางที่ 4.8

ตารางที่ 4.8 ผลการปรับค่าพารามิเตอร์ที่เหมาะสม สำหรับอัลกอริทึมการเรียนรู้แบบกลุ่ม สำหรับ
ระยะเวลาพยากรณ์ 1 และ 2 ชั่วโมง

อัลกอริทึม	พารามิเตอร์	ค่าที่เหมาะสม	
		1 ชม.	2 ชม.
กลุ่มข้อมูลที่ 1 รวมสนามบินภาคเหนือ			
Random Forest	Number of Trees (numTrees)	19	80
	Criterion:	gain_ratio	gain_ratio
	Maximum Depth (maxDepth)	7	7
Decision Tree +Bagging	Criterion:	gain_ratio	gini_index
	Maximum Depth (maxDepth)	10	10
	minimal leaf size	2	4
	minimal size for split	2	2
Neural Networks +Bagging	Training Cycle	200	175
	Learning Rate	0.1	0.1
	Momentum	0.8	0.9
	Hidden layer sizes	8	6
Decision Tree + AdaBoost	Criterion:	gain_ratio	gini_index
	Maximum Depth (maxDepth)	12	12
	minimal leaf size	4	2
	minimal size for split	2	2
Neural Networks +AdaBoost	Training Cycle	150	120
	Learning Rate	0.1	0.1
	Momentum	0.8	0.8
	Hidden layer sizes	8	6

ตารางที่ 4.8 (ต่อ)

อัลกอริทึม	พารามิเตอร์	ค่าที่เหมาะสม	
		1 ชม.	2 ชม.
กลุ่มข้อมูลที่ 2 สนามบินภาคเหนือตอนบน			
Random Forest	Number of Trees (numTrees)	60	50
	Criterion:	gain_ratio	gain_ratio
	Maximum Depth (maxDepth)	40	15
Decision Tree + Bagging	Criterion:	gini_index	gini_index
	Maximum Depth (maxDepth)	8	8
	minimal leaf size	2	4
	minimal size for split	2	4
Neural Networks +Bagging	Training Cycle	45	120
	Learning Rate	0.1	0.1
	Momentum	0.5	0.9
	Hidden layer sizes	8	8
Decision Tree + AdaBoost	Criterion:	gini_index	gini_index
	Maximum Depth (maxDepth)	10	8
	minimal leaf size	4	4
	minimal size for split	2	4
Neural Networks +AdaBoost	Training Cycle	60	100
	Learning Rate	0.1	0.1
	Momentum	0.5	0.7
	Hidden layer sizes	8	8

ตารางที่ 4.8 (ต่อ)

อัลกอริทึม	พารามิเตอร์	ค่าที่เหมาะสม	
		1 ชม.	2 ชม.
กลุ่มข้อมูลที่ 3 สนามบินภาคเหนือตอนล่าง			
Random Forest	Number of Trees (numTrees)	15	12
	Criterion:	gain_ratio	gain_ratio
	Maximum Depth (maxDepth)	7	10
Decision Tree + Bagging	Criterion:	gain_ratio	gini_index
	Maximum Depth (maxDepth)	10	8
	minimal leaf size	2	2
	minimal size for split	2	2
Neural Networks +Bagging	Training Cycle	80	180
	Learning Rate	0.1	0.1
	Momentum	0.6	0.8
	Hidden layer sizes	6	8
Decision Tree + AdaBoost	Criterion:	gain_ratio	gain_ratio
	Maximum Depth (maxDepth)	10	8
	minimal leaf size	2	2
	minimal size for split	2	2
Neural Networks +AdaBoost	Training Cycle	122	185
	Learning Rate	0.1	0.1
	Momentum	0.7	0.6
	Hidden layer sizes	8	8

ตารางที่ 4.8 (ต่อ)

อัลกอริทึม	พารามิเตอร์	ค่าที่เหมาะสม	
		1 ชม.	2 ชม.
กลุ่มข้อมูลที่ 4 สนามบินที่มีภูเขาล้อมรอบ 1-2 ด้าน			
Random Forest	Number of Trees (numTrees)	20	70
	Criterion:	gain_ratio	gain_ratio
	Maximum Depth (maxDepth)	15	10
Decision Tree + Bagging	Criterion:	gini_index	gain_ratio
	Maximum Depth (maxDepth)	12	8
	minimal leaf size	2	2
	minimal size for split	2	2
Neural Networks +Bagging	Training Cycle	205	200
	Learning Rate	0.1	0.1
	Momentum	0.8	0.7
	Hidden layer sizes	8	6
Decision Tree + AdaBoost	Criterion:	gain_ratio	information_gain
	Maximum Depth (maxDepth)	8	10
	minimal leaf size	4	2
	minimal size for split	4	2
Neural Networks +AdaBoost	Training Cycle	180	200
	Learning Rate	0.1	0.1
	Momentum	0.7	0.6
	Hidden layer sizes	6	8

ตารางที่ 4.8 (ต่อ)

อัลกอริทึม	พารามิเตอร์	ค่าที่เหมาะสม	
		1 ชม.	2 ชม.
กลุ่มข้อมูลที่ 5 สนามบินที่มีภูเขาล้อมรอบ 3-4 ด้าน			
Random Forest	Number of Trees (numTrees)	15	80
	Criterion:	gain_ratio	gini_index
	Maximum Depth (maxDepth)	5	8
Decision Tree + Bagging	Criterion:	gini_index	gini_index
	Maximum Depth (maxDepth)	10	8
	minimal leaf size	2	2
	minimal size for split	2	2
Neural Networks +Bagging	Training Cycle	245	215
	Learning Rate	0.1	0.1
	Momentum	0.7	0.7
	Hidden layer sizes	6	8
Decision Tree + AdaBoost	Criterion:	gini_index	gini_index
	Maximum Depth (maxDepth)	9	10
	minimal leaf size	4	4
	minimal size for split	2	2
Neural Networks +AdaBoost	Training Cycle	225	175
	Learning Rate	0.1	0.1
	Momentum	0.8	0.7
	Hidden layer sizes	8	8

จากตารางที่ 4.7 และ 4.8 ผลการปรับค่าพารามิเตอร์ที่เหมาะสม ผู้วิจัยได้ทำการสร้างแบบจำลองของแต่ละอัลกอริทึมโดยกำหนดค่าเริ่มต้นของชุดพารามิเตอร์ตามตารางดังกล่าว ซึ่งพบว่าสำหรับระยะเวลาพยากรณ์ 1 ชั่วโมงข้างหน้า ผลการปรับค่าพารามิเตอร์ไม่ทำให้แบบจำลองที่สร้างขึ้นเกิด Overfitting หรือ Underfitting เนื่องจากเมื่อนำเข้าข้อมูลทดสอบแล้วแบบจำลองยังสามารถทำนายการเกิดพายุฝนฟ้าคะนองได้ดี แต่สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมงข้างหน้า ถึงแม้จะปรับค่าพารามิเตอร์ต่างๆแล้วก็ตาม แต่แบบจำลองยังทำนายการเกิดพายุฝนฟ้าคะนองไม่ได้

3. ผลการประเมินประสิทธิภาพแบบจำลอง

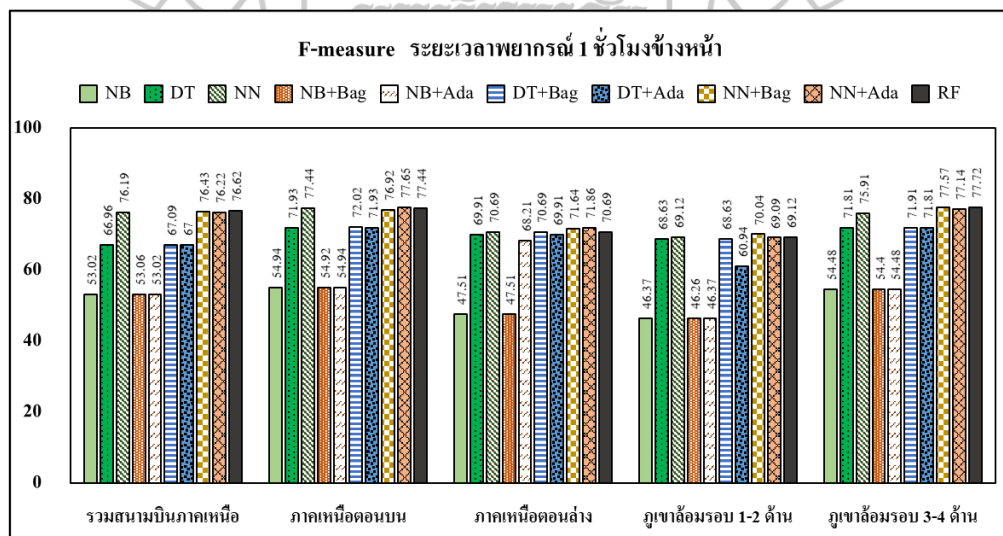
ผู้วิจัยได้สร้างแบบจำลองจากอัลกอริทึมมาตรฐาน ได้แก่ Naive Bayes Decision Tree และ Neural Networks รวมถึงสร้างแบบจำลองจากอัลกอริทึมมาตรฐานร่วมกับอัลกอริทึมการเรียนรู้แบบกลุ่ม ได้แก่ Random Forest Bagging และ AdaBoost โดยทำการปรับค่าพารามิเตอร์ต่างๆที่เหมาะสมกับแต่ละอัลกอริทึม จากนั้นทำการวัดประสิทธิภาพแบบจำลองทั้งหมดด้วยชุดข้อมูลทดสอบ (Testing Data) ที่เป็นข้อมูลในปี 2564-2565 รวม 2 ปี เพื่อทดสอบแบบจำลองของแต่ละอัลกอริทึมว่าสามารถทำงานได้ดีกับชุดข้อมูลที่ไม่เคยพบมาก่อนหรือไม่ และทำการวัดประสิทธิภาพแบบจำลองเหล่านี้ด้วย Accuracy Precision Recall F-measure AUC และ Balanced Accuracy รวมไปถึงเปรียบเทียบระยะเวลาในการดำเนินการของแบบจำลองตั้งแต่เริ่มต้นจนสิ้นสุดกระบวนการ ดังตารางที่ 4.9

ตารางที่ 4.9 ผลการทดสอบประสิทธิภาพของแบบจำลอง สำหรับระยะเวลาพยากรณ์ 1 ชั่วโมง

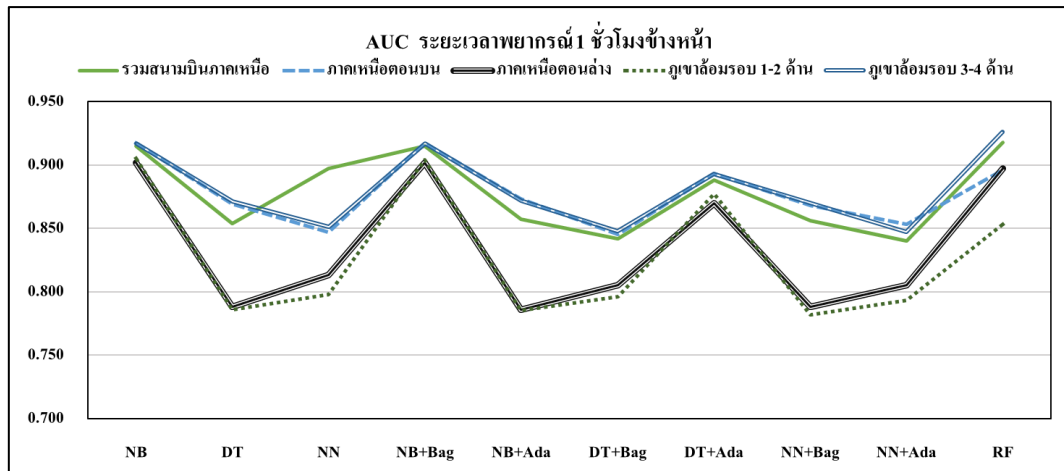
PERFORMANCE	แบบจำลอง									
	อัลกอริทึมมาตรฐาน			อัลกอริทึมการเรียนรู้แบบกลุ่ม						
	NB	DT	NN	NB+ Bag	DT+ Bag	NN+ Bag	NB+ Ada	DT+ Ada	NN+ Ada	RF
กลุ่มข้อมูลที่ 1 รวมสนามบินภาคเหนือ										
Accuracy	98.85	99.49	99.60	98.85	99.50	99.61	98.85	99.49	99.60	99.61
Precision	44.97	96.73	96.05	45.03	96.91	96.21	44.97	96.90	96.19	97.19
Recall	64.58	51.20	63.14	64.58	51.30	63.40	64.58	51.20	63.11	63.23
F-measure	53.02	66.96	76.19	53.06	67.09	76.43	53.02	67.00	76.22	76.62
AUC	0.915	0.854	0.897	0.915	0.857	0.842	0.888	0.856	0.840	0.918
Balanced Accuracy	81.89	75.59	81.56	82.28	75.64	81.30	82.28	75.59	81.16	81.60
Execution Time	0:00:01	0:00:04	1:59:43	0:00:14	0:00:27	9:50:12	0:00:34	0:00:24	6:45:30	0:00:36
กลุ่มข้อมูลที่ 2 สนามบินภาคเหนือตอนบน										
Accuracy	98.78	99.50	99.58	98.78	99.50	99.57	98.78	99.50	99.58	99.57
Precision	46.84	96.15	96.90	46.87	96.15	96.33	46.84	96.15	94.92	95.06
Recall	66.42	57.45	64.48	66.30	57.58	64.02	66.42	57.45	65.70	65.33
F-measure	54.94	71.93	77.44	54.92	72.02	76.92	54.94	71.93	77.65	77.44
AUC	0.917	0.869	0.847	0.917	0.873	0.845	0.893	0.868	0.853	0.896
Balanced Accuracy	82.78	78.71	82.22	83.14	78.78	81.59	83.20	78.71	82.42	82.65
Execution Time	0:00:01	0:00:05	0:23:46	0:00:08	0:00:13	5:25:12	0:00:20	0:00:13	4:50:22	0:01:27
กลุ่มข้อมูลที่ 3 สนามบินภาคเหนือตอนล่าง										
Accuracy	99.09	99.67	99.68	99.09	99.68	99.68	99.63	99.67	99.69	99.68
Precision	40.26	100.00	100.00	40.26	100.00	99.17	89.39	100.00	100.00	100.00
Recall	57.94	53.74	54.67	57.94	54.67	56.07	55.14	53.74	56.07	54.67
F-measure	47.51	69.91	70.69	47.51	70.69	71.64	68.21	69.91	71.86	70.69
AUC	0.902	0.788	0.813	0.902	0.786	0.805	0.870	0.788	0.805	0.897
Balanced Accuracy	78.66	76.87	77.34	78.97	77.34	77.73	77.57	76.87	78.01	77.34
Execution Time	0:00:01	0:00:05	0:13:05	0:00:04	0:00:04	2:20:15	0:00:08	0:00:04	3:10:18	0:00:12

ตารางที่ 4.9 (ต่อ)

PERFORMANCE	แบบจำลอง									
	อัลกอริทึมมาตรฐาน			อัลกอริทึมการเรียนรู้แบบกลุ่ม						
	NB	DT	NN	NB+ Bag	DT+ Bag	NN+ Bag	NB+ Ada	DT+ Ada	NN+ Ada	RF
กลุ่มข้อมูลที่ 4 สนามบินที่มีภูเขาล้อมรอบ 1-2 ด้าน										
Accuracy	98.88	99.59	99.59	98.88	99.59	99.60	98.88	99.51	99.59	99.59
Precision	39.76	100.00	100.00	39.60	100.00	98.98	39.76	100.00	96.94	100.00
Recall	55.62	52.25	52.81	55.62	52.25	54.19	55.62	43.82	53.67	52.81
F-measure	46.37	68.63	69.12	46.26	68.63	70.04	46.37	60.94	69.09	69.12
AUC	0.905	0.786	0.798	0.904	0.785	0.796	0.877	0.782	0.793	0.853
Balanced Accuracy	77.44	76.13	76.41	77.81	76.13	76.72	77.81	71.91	76.47	76.40
Execution Time	0:00:01	0:00:04	0:27:50	0:00:02	0:00:05	1:32:10	0:00:07	0:00:02	1:50:15	0:00:08
กลุ่มข้อมูลที่ 5 สนามบินที่มีภูเขาล้อมรอบ 3-4 ด้าน										
Accuracy	98.85	99.53	99.60	98.85	99.54	99.61	98.85	99.53	99.61	99.61
Precision	46.17	96.29	95.61	46.05	96.30	96.86	46.17	96.29	96.32	97.04
Recall	66.43	57.26	62.95	66.43	57.38	64.69	66.43	57.26	64.33	64.81
F-measure	54.48	71.81	75.91	54.40	71.91	77.57	54.48	71.81	77.14	77.72
AUC	0.917	0.871	0.851	0.917	0.872	0.848	0.893	0.870	0.847	0.926
Balanced Accuracy	82.81	78.62	81.46	83.20	78.68	81.94	83.20	78.62	81.76	82.39
Execution Time	0:00:01	0:00:05	1:51:29	0:00:03	0:00:28	2:40:32	0:00:35	0:00:12	3:10:25	0:00:14



ภาพที่ 4.5 ค่า F-measure แต่ละอัลกอริทึมของกลุ่มข้อมูล ระยะเวลาพยากรณ์ 1 ชั่วโมงข้างหน้า



ภาพที่ 4.6 การเปรียบเทียบค่า AUC ของอัลกอริทึมแต่ละกลุ่มข้อมูล
สำหรับระยะเวลาพยากรณ์ 1 ชั่วโมงข้างหน้า

จากตารางที่ 4.9 ผลการทดสอบประสิทธิภาพในการพยากรณ์การเกิดพายุฝนฟ้าคะนองสำหรับระยะเวลาพยากรณ์ 1 ชั่วโมงข้างหน้า เมื่อพิจารณาพบว่า มีค่า Accuracy มากกว่า 98% ในทุกอัลกอริทึมเกิดจากการทำนายคลาสไม่เกิดพายุฝนฟ้าคะนองได้ถูกต้องซึ่งเป็นคลาสที่มีมากกว่านั่นเอง โดยหากข้อมูลในงานวิจัยเป็นข้อมูลไม่สมดุลกันระหว่างสองคลาส โดยมีคลาสของการไม่เกิดมากกว่าคลาสเกิด การใช้ค่า Accuracy ในการตัดสินใจแบบจำลองจะทำให้เกิดการโอนเอียงจากข้อมูลเพราะทำนายคลาสไม่เกิดได้ถูกต้องมากกว่า ดังนั้นค่า F-measure จึงมีความเหมาะสมมากกว่าที่จะนำมาใช้เป็นตัวชี้วัดของข้อมูลชุดนี้ (เฟื่องลดา มะโนกิจ, 2562) และเมื่อเปรียบเทียบค่า F-measure ของแต่ละกลุ่มข้อมูล พบว่า ค่าสูงสุดของแต่ละกลุ่มข้อมูลมีค่า ดังนี้

1. กลุ่มข้อมูลรวมสนามบินภาคเหนือ คือ Random Forest เท่ากับ 76.62%
2. กลุ่มข้อมูลภาคเหนือตอนบน คือ Networks ร่วมกับ AdaBoost เท่ากับ 77.65%
3. กลุ่มข้อมูลภาคเหนือตอนล่าง คือ Neural Networks ร่วมกับ AdaBoost เท่ากับ 71.86%
4. กลุ่มข้อมูลสนามบินที่มีภูเขาล้อมรอบ 1-2 ด้าน คือ Neural Networks ร่วมกับ Bagging เท่ากับ 70.04%
5. กลุ่มข้อมูลสนามบินที่มีภูเขาล้อมรอบ 3-4 ด้าน คือ Random Forest เท่ากับ 77.72%

ผู้วิจัยได้ใช้ค่า F-measure ในการตัดสินใจแบบจำลองที่ดีที่สุดของแต่ละกลุ่มข้อมูล และได้พิจารณาค่าของการวัดประสิทธิภาพอื่นๆ ได้แก่ Accuracy Precision Recall AUC และ

Balanced Accuracy รวมไปถึงเปรียบเทียบระยะเวลาในการดำเนินการของแบบจำลองตั้งแต่เริ่มต้นจนถึงสิ้นสุดกระบวนการ เพื่อพิจารณาประสิทธิภาพของแบบจำลองโดยภาพรวม โดยค่า Balanced Accuracy จะมีค่าน้อยกว่า Accuracy ซึ่งทำให้สามารถทราบค่าความถูกต้องสมดุลของทั้งสองคลาสได้ อีกทั้งเมื่อพิจารณาค่า AUC พบว่า มีความสอดคล้องกับค่า Balanced Accuracy เนื่องจาก AUC เป็นพื้นที่ใต้กราฟที่พล็อตระหว่างค่า $1 - \text{Specificity}$ และค่า Sensitivity (Devos et al., 2007) เช่นเดียวกับ Balanced Accuracy ที่สามารถคำนวณได้จากค่า Specificity และ Sensitivity เช่นเดียวกัน แต่จะเห็นว่าค่า AUC ไม่สอดคล้องกับ F-measure ที่จะสามารถวัดผลความถ่วงดุลของแบบจำลองที่หามาได้จากการคำนวณค่า Precision และ Recall ดังเช่นเมื่อพิจารณาแบบจำลองของกลุ่มข้อมูลภาคเหนือตอนบนนั้น อัลกอริทึม Naive Bayes มีค่า F-measure เท่ากับ 54.94% และค่า AUC เท่ากับ 0.917 เปรียบเทียบกับอัลกอริทึม Neural Networks ร่วมกับ AdaBoost มีค่า F-measure เท่ากับ 77.65% และค่า AUC เท่ากับ 0.853 ซึ่งอัลกอริทึม Neural Networks ร่วมกับ AdaBoost ถูกเลือกเป็นแบบจำลองที่ดีที่สุดสำหรับกลุ่มข้อมูลนี้เนื่องจากมีค่า F-measure ที่มากกว่าสอดคล้องกับการศึกษาโดย Kung CK. (2020) ที่พบว่า หากแบบจำลองที่สร้างขึ้นมีค่า AUC ที่สูงนั้น ไม่ได้หมายความว่าค่า Precision ที่เป็นค่าหนึ่งที่ใช้คำนวณค่า F-measure จะสูงตามไปด้วย แม้ทั้งสองค่านี้จะเป็นค่าที่บอกถึงประสิทธิภาพของแบบจำลอง แต่ไม่ควรนำมาใช้คู่กันเพราะจะทำให้เกิดความสับสนในการเลือกแบบจำลองได้ โดยจะใช้ค่า AUC ในการวัดประสิทธิภาพของแบบจำลองโดยภาพรวม

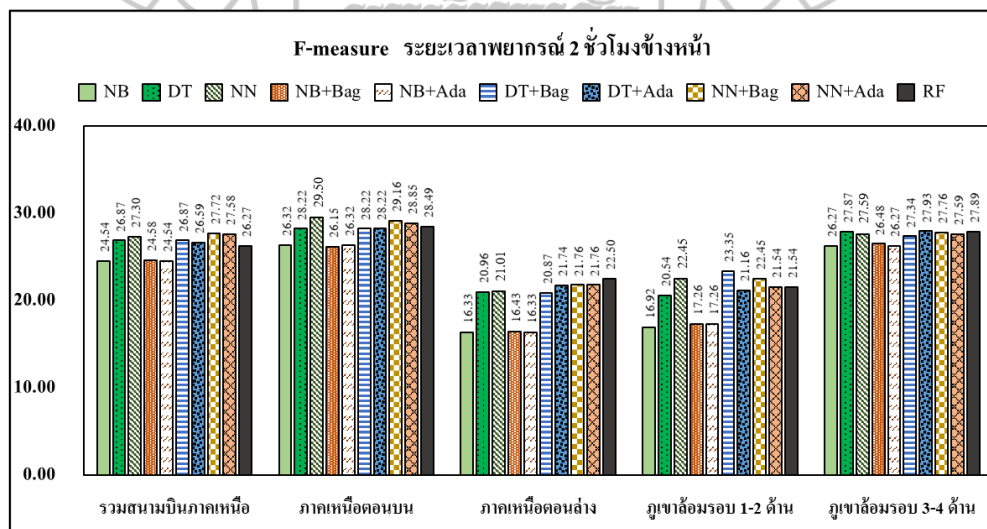


ตารางที่ 4.10 ผลการทดสอบประสิทธิภาพของแบบจำลอง สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมง

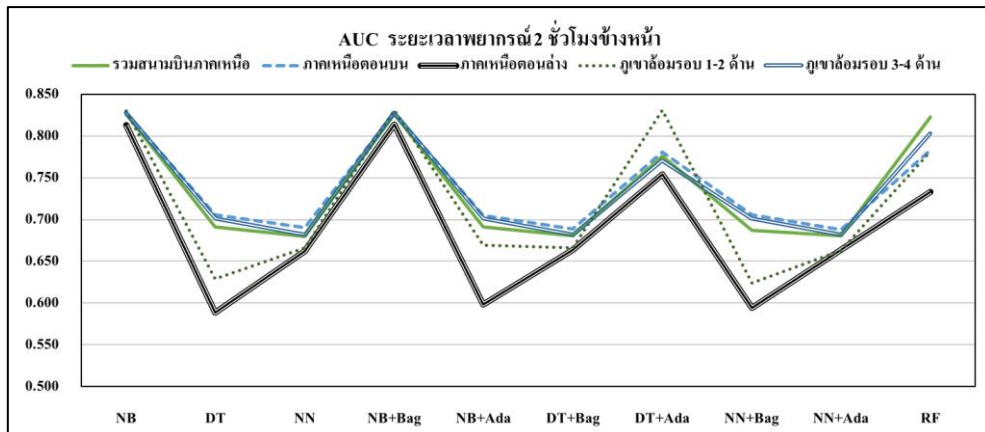
การทดสอบ ประสิทธิภาพ	แบบจำลอง									
	อัลกอริทึมมาตรฐาน			อัลกอริทึมการเรียนรู้แบบกลุ่ม						
	NB	DT	NN	NB+ Bag	DT+ Bag	NN+ Bag	NB+ Ada	DT+ Ada	NN+ Ada	RF
กลุ่มข้อมูลที่ 1 รวมสนามบินภาคเหนือ										
Accuracy	98.26	99.15	99.15	98.26	99.15	99.15	98.26	99.14	99.15	99.14
Precision	21.58	92.05	93.75	21.59	92.05	93.85	21.58	88.95	93.30	91.33
Recall	28.45	15.73	15.97	28.54	15.73	16.26	28.45	15.63	16.18	15.34
F-measure	24.54	26.87	27.30	24.58	26.87	27.72	24.54	26.59	27.58	26.27
AUC	0.826	0.691	0.680	0.826	0.691	0.681	0.776	0.687	0.681	0.823
Balanced Accuracy	63.71	57.86	57.98	64.26	57.86	57.61	64.22	57.81	57.57	57.66
Execution Time	0:00:01	0:00:05	2:10:45	0:00:12	0:00:19	2:45:15	0:00:28	0:00:16	3:01:18	0:02:14
กลุ่มข้อมูลที่ 2 สนามบินภาคเหนือตอนบน										
Accuracy	98.03	99.04	99.02	98.02	99.04	99.02	98.03	99.04	99.02	99.02
Precision	22.72	92.00	80.65	22.53	92.00	80.00	22.72	92.00	79.35	78.69
Recall	31.28	16.67	18.05	31.16	16.67	17.83	31.28	16.67	17.63	17.39
F-measure	26.32	28.22	29.50	26.15	28.22	29.16	26.32	28.22	28.85	28.49
AUC	0.828	0.706	0.690	0.828	0.705	0.689	0.781	0.706	0.688	0.783
Balanced Accuracy	65.03	58.33	59.00	65.56	58.33	58.31	65.61	58.33	58.21	58.67
Execution Time	0:00:01	0:00:04	1:05:47	0:00:10	0:00:11	1:55:30	0:00:14	0:00:13	2:35:15	0:01:02
กลุ่มข้อมูลที่ 3 สนามบินภาคเหนือตอนล่าง										
Accuracy	98.64	99.40	99.38	98.65	99.40	99.38	98.64	99.40	99.38	99.38
Precision	13.89	88.89	67.57	14.04	85.71	68.42	13.89	89.29	70.27	71.05
Recall	13.89	11.88	12.44	19.80	11.88	12.94	19.80	12.38	12.87	71.05
F-measure	16.33	20.96	21.01	16.43	20.87	21.76	16.33	21.74	21.76	22.50
AUC	0.813	0.589	0.662	0.814	0.598	0.664	0.754	0.594	0.664	0.733
Balanced Accuracy	56.53	55.93	56.20	59.88	55.93	56.06	59.88	56.18	56.02	57.03
Execution Time	0:00:01	0:00:04	0:18:25	0:00:03	0:00:04	2:45:20	0:00:08	0:00:04	3:05:17	0:00:10

ตารางที่ 4.10 (ต่อ)

การทดสอบ ประสิทธิภาพ	แบบจำลอง									
	อัลกอริทึมมาตรฐาน			อัลกอริทึมการเรียนรู้แบบกลุ่ม						
	NB	DT	NN	NB+ Bag	DT+ Bag	NN+ Bag	NB+ Ada	DT+ Ada	NN+ Ada	RF
กลุ่มข้อมูลที่ 4 สนามบินที่มีภูเขาล้อมรอบ 1-2 ด้าน										
Accuracy	98.37	99.28	99.26	98.41	99.26	99.26	98.41	99.27	99.25	99.25
Precision	14.41	100.00	73.33	14.91	76.67	70.97	14.91	86.96	70.00	72.41
Recall	20.48	11.45	13.25	20.48	13.77	13.33	20.48	12.05	12.73	12.65
F-measure	16.92	20.54	22.45	17.26	23.35	22.45	17.26	21.16	21.54	21.54
AUC	0.830	0.629	0.666	0.831	0.669	0.666	0.831	0.624	0.663	0.782
Balanced Accuracy	59.74	55.73	56.61	60.22	56.88	56.19	60.22	56.02	55.89	56.30
Execution Time	0:00:01	0:00:04	0:32:15	0:00:03	1:37:15	1:37:15	0:00:03	0:00:03	1:45:50	0:00:27
กลุ่มข้อมูลที่ 5 สนามบินที่มีภูเขาล้อมรอบ 3-4 ด้าน										
Accuracy	98.22	99.12	99.11	98.24	99.10	99.11	98.22	99.11	99.11	99.09
Precision	23.11	91.61	88.68	23.36	87.50	89.31	23.11	89.38	88.68	79.78
Recall	30.44	16.44	16.34	30.56	16.20	16.44	30.44	16.55	16.34	16.90
F-measure	26.27	27.87	27.59	26.48	27.34	27.76	26.27	27.93	27.59	27.89
AUC	0.828	0.701	0.682	0.828	0.701	0.682	0.771	0.701	0.682	0.803
Balanced Accuracy	64.69	58.21	58.15	65.27	58.09	57.69	65.21	58.26	57.64	58.44
Execution Time	0:00:01	0:00:03	1:45:33	0:00:03	0:00:15	2:30:15	0:00:29	0:00:11	3:00:20	0:02:09



ภาพที่ 4.7 ค่า F-measure แต่ละอัลกอริทึมของกลุ่มข้อมูล ระยะเวลาพยากรณ์ 2 ชั่วโมงข้างหน้า



ภาพที่ 4.8 การเปรียบเทียบค่า AUC ของอัลกอริทึมแต่ละกลุ่มข้อมูล สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมงข้างหน้า

สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมงข้างหน้า จากตารางที่ 4.10 ผลการทดสอบประสิทธิภาพในการพยากรณ์การเกิดพายุฝนฟ้าคะนอง สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมงข้างหน้า เมื่อพิจารณาค่า Accuracy ของทุกชุดข้อมูล พบว่า มีค่า Accuracy มากกว่า 98% ในทุกอัลกอริทึม โดยเกิดจากการทำนายคลาสไม่เกิดพายุฝนฟ้าคะนองได้ถูกต้องซึ่งเป็นคลาสที่มีมากกว่านั่นเอง และจากการเปรียบเทียบค่า F-measure ของแต่ละกลุ่มข้อมูล พบว่า แบบจำลองที่ทำการศึกษาในครั้งนี้มีค่า F-measure ของแต่ละชุดข้อมูลต่ำไม่ถึง 50% นั่นคือ ยังไม่สามารถทำนายการเกิดพายุฝนฟ้าคะนอง สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมงข้างหน้าได้อย่างถูกต้อง

4. สรุปแบบจำลองที่เหมาะสมสำหรับแต่ละพื้นที่สนามบิน

การคัดเลือกแบบจำลองที่ดีที่สุดจะพิจารณาผลการทดสอบประสิทธิภาพของแบบจำลอง จากตารางที่ 4.9 และตารางที่ 4.10 เพื่อทำการเปรียบเทียบค่า F-measure สูงสุดของแต่ละกลุ่มข้อมูล โดยสรุปผลการคัดเลือกแบบจำลองที่ดีที่สุดของแต่ละกลุ่มข้อมูล สำหรับระยะเวลาพยากรณ์ 1 ชั่วโมง พบว่า มีค่า F-measure ระหว่าง 70.04% - 77.72% แต่สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมง ในทุกๆอัลกอริทึมยังมีประสิทธิภาพการทำนายการเกิดพายุฝนฟ้าคะนองได้ไม่ดีเท่าที่ควร โดยพิจารณาจากค่า F-measure ที่มีค่าระหว่าง 22.50% - 29.50% และเมื่อพิจารณาค่า AUC พบว่าเป็นแบบจำลองที่มีประสิทธิภาพต่ำใกล้เคียงกับแบบจำลองมาตรฐานทั่วไปเท่านั้น สรุปผลการคัดเลือกแบบจำลองที่ดีที่สุดของแต่ละพื้นที่สนามบิน ดังตารางที่ 4.11

ตารางที่ 4.11 สรุปผลการคัดเลือกแบบจำลองที่ดีที่สุดของแต่ละพื้นที่สนามบิน

พื้นที่สนามบิน	แบบจำลองที่ดีที่สุด	การทดสอบประสิทธิภาพ		
		F-measure	Balanced Accuracy	AUC
ระยะเวลาพยากรณ์ 1 ชั่วโมงข้างหน้า				
รวมสนามบินภาคเหนือ	Random Forest	76.62	81.60	0.918
ภาคเหนือตอนบน	Neural Networks + AdaBoost	77.65	82.42	0.853
ภาคเหนือตอนล่าง	Neural Networks + AdaBoost	71.86	78.01	0.805
ภูเขาล้อมรอบ 1-2 ด้าน	Neural Networks + Bagging	70.04	76.72	0.796
ภูเขาล้อมรอบ 3-4 ด้าน	Random Forest	77.72	82.39	0.926
ระยะเวลาพยากรณ์ 2 ชั่วโมงข้างหน้า				
รวมสนามบินภาคเหนือ	Neural Networks + Bagging	27.72	57.61	0.681
ภาคเหนือตอนบน	Neural Networks	29.50	59.00	0.690
ภาคเหนือตอนล่าง	Random Forest	22.50	57.03	0.733
ภูเขาล้อมรอบ 1-2 ด้าน	Decision Tree + Bagging	23.35	56.88	0.669
ภูเขาล้อมรอบ 3-4 ด้าน	Decision Tree + AdaBoost	27.93	58.26	0.701

ตามเกณฑ์จากการศึกษาโดยนันทพัชร เสนาวงศ์ และคณะ (2565) ผลการทดสอบค่า AUC ของแบบจำลองที่ดีที่สุดแต่ละกลุ่มข้อมูล สำหรับระยะเวลาพยากรณ์ 1 ชั่วโมงข้างหน้าในงานวิจัยสรุปได้ดังนี้

1. กลุ่มข้อมูลรวมสนามบินภาคเหนือ คือ Random Forest เท่ากับ 0.918 จัดเป็นแบบจำลองที่ทำงานได้ดีมาก
2. กลุ่มข้อมูลภาคเหนือตอนบน คือ Networks ร่วมกับ AdaBoost เท่ากับ 0.853 จัดเป็นแบบจำลองที่ทำงานได้ดี
3. กลุ่มข้อมูลภาคเหนือตอนล่าง คือ Neural Networks ร่วมกับ AdaBoost เท่ากับ 0.805 จัดเป็นแบบจำลองที่ทำงานได้ดี
4. กลุ่มข้อมูลสนามบินที่มีภูเขาล้อมรอบ 1-2 ด้าน คือ Neural Networks ร่วมกับ Bagging เท่ากับ 0.796 จัดเป็นแบบจำลองที่ทำงานได้ตามเกณฑ์มาตรฐานสำหรับแบบจำลองส่วนใหญ่

5. กลุ่มข้อมูลสนามบินที่มีภูเขาล้อมรอบ 3-4 ด้าน คือ Random Forest เท่ากับ 0.926 จัดเป็นแบบจำลองที่ทำงานได้ดีมาก

5. การทดสอบนำแบบจำลองไปใช้งานจริง

ผลการทดสอบจากการนำเข้าข้อมูลข่าวอากาศการบินในปีพ.ศ. 2566 ระยะเวลา 1 ปี และทำการวัดประสิทธิภาพด้วยค่า Accuracy Precision Recall F-measure AUC และ Balanced Accuracy และจำนวนครั้งของผลการทำนายคลาสิกการเกิดฝนฟ้าคะนองได้ถูกต้อง (True Positive) จาก Confusion matrix เปรียบเทียบกับจำนวนครั้งที่ทำนายถูกต้องจากนักอุตุนิยมวิทยาผู้ปฏิบัติงานจริงจากการออกข่าวพยากรณ์แนวโน้มสภาพอากาศร้าย (Trend Forecast) ดังตารางที่ 4.12-4.14 ตารางที่ 4.12 ผลการทดสอบประสิทธิภาพของแบบจำลองจากการนำไปใช้จริง

พื้นที่ สนามบิน	แบบจำลอง ที่ดีที่สุด	การทดสอบประสิทธิภาพ						
		Accuracy	Precision	Recall	F-measure	AUC	Balanced Accuracy	Execution Time
สำหรับระยะเวลาพยากรณ์ 1 ชั่วโมงข้างหน้า								
รวม	RF	99.72	93.11	79.63	85.84	0.996	89.78	0:00:06
ตอนบน	NN+Ada	99.59	86.56	79.17	82.70	0.994	89.51	0:36:19
ตอนล่าง	NN+Ada	99.84	93.14	83.33	87.96	0.916	91.65	0:16:49
1-2 ด้าน	NN+Bag	99.83	96.34	82.29	88.76	0.997	91.13	0:37:23
3-4 ด้าน	RF	99.66	96.70	71.69	82.34	0.995	85.83	0:00:08
สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมงข้างหน้า								
รวม	NN+Bag	99.13	82.42	13.89	23.77	0.775	56.93	2:14:36
ตอนบน	NN	98.89	67.88	15.66	30.15	0.774	57.77	0:13:27
ตอนล่าง	RF	99.71	100.00	18.33	30.99	0.776	59.17	0:00:02
1-2 ด้าน	DT+Bag	99.64	unknown	0.00	unknown	0.500	50.00	0:00:02
3-4 ด้าน	DT+Ada	98.88	66.67	2.41	4.65	0.733	51.20	0:00:02

จากตารางที่ 4.12 ผลการทดสอบประสิทธิภาพของแบบจำลองจากการนำไปใช้จริง สำหรับระยะเวลาพยากรณ์ 1 ชั่วโมงข้างหน้า พบว่า แบบจำลองที่ดีที่สุดของแต่ละพื้นที่สนามบิน เมื่อทดสอบโดยการนำเข้าข้อมูลปีพ.ศ. 2566 แล้วทุกๆพื้นที่สนามบินมีประสิทธิภาพการพยากรณ์การเกิดฝนฟ้าคะนองได้ดี นั่นคือ F-measure เท่ากับ 82.34%-88.76% และเมื่อพิจารณาค่าอื่นๆประกอบเพื่อดูประสิทธิภาพโดยรวมของการจำแนกทั้งสองคลาสของแบบจำลอง ได้แก่ Balanced Accuracy เท่ากับ 85.83%-91.65% และ AUC เท่ากับ 0.916-0.997 ซึ่งตามเกณฑ์การวัดประสิทธิภาพจากค่า AUC จัดได้ว่าเป็นแบบจำลองที่ทำงานได้ดีมาก (นันทชพร เสนาวงศ์ และคณะ, 2565)

แต่สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมงข้างหน้า พบว่า แบบจำลองที่ดีที่สุดของแต่ละพื้นที่สนามบินยังไม่สามารถพยากรณ์การเกิดฝนฟ้าคะนองได้ดีเท่าที่ควร โดยเมื่อทดสอบนำเข้าข้อมูลพื้นที่สนามบินที่มีภูเขาล้อมรอบ 1-2 ด้านในแบบจำลอง Decision Tree ร่วมกับ Bagging นั้นผลลัพธ์แสดงให้เห็นว่าไม่สามารถทำนายการเกิดพายุฝนฟ้าคะนองได้เลย สำหรับพื้นที่สนามบินอื่นๆ สามารถทำนายการเกิดพายุฝนฟ้าคะนองได้บ้าง แต่ประสิทธิภาพการพยากรณ์การเกิดพายุฝนฟ้าคะนองยังมีค่าน้อย โดยมีค่า F-measure เท่ากับ 4.65%-30.99% เท่านั้น และเมื่อพิจารณาค่าอื่นๆประกอบเพื่อดูประสิทธิภาพโดยรวมของการจำแนกทั้งสองคลาสของแบบจำลอง ได้แก่ Balanced Accuracy เท่ากับ 50.00%-59.17% และค่า AUC เท่ากับ 0.500-0.776 ถึงแม้ว่าจะจัดเป็นเกณฑ์มาตรฐานสำหรับแบบจำลองส่วนใหญ่ แต่ที่แบบจำลองในการทดสอบครั้งนี้มีค่า AUC สูง เนื่องจากสามารถทำนายคลาสของการไม่เกิดพายุฝนฟ้าคะนองได้ถูกต้องซึ่งเป็นข้อมูลกลุ่มมากกว่านั่นเอง

ตารางที่ 4.13 เปรียบเทียบผลการทำนายคลาสการเกิดฝนฟ้าคะนองได้ถูกต้องจากการนำไปใช้จริง สำหรับระยะเวลาพยากรณ์ 1 ชั่วโมงข้างหน้า

พื้นที่สนามบิน	จำนวนครั้งที่เกิดพายุฝนฟ้าคะนองทั้งหมด	ทำนายโดยแบบจำลองที่ดีที่สุด		ทำนายโดยนักอุตุนิยมวิทยา	
		ทายถูก (ครั้ง)	คิดเป็น (%)	ทายถูก (ครั้ง)	คิดเป็น (%)
รวมสนามบิน	594	473	79.63	298	50.17
ดอนบน	480	380	79.17	244	50.83
ดอนล่าง	114	95	83.33	54	47.37
1-2 ด้าน	96	79	82.29	45	46.88
3-4 ด้าน	498	352	70.68	253	50.80

ตารางที่ 4.14 เปรียบเทียบผลการทำนายคลาสการเกิดฝนฟ้าคะนองได้ถูกต้องจากการนำไปใช้จริง สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมงข้างหน้า

พื้นที่สนามบิน	จำนวนครั้งที่เกิดพายุฝนฟ้าคะนองทั้งหมด	ทำนายโดยแบบจำลองที่ดีที่สุด		ทำนายโดยนักอุตุนิยมวิทยา	
		ทายถูก (ครั้ง)	คิดเป็น (%)	ทายถูก (ครั้ง)	คิดเป็น (%)
รวมสนามบิน	540	75	13.89	266	49.26
ดอนบน	480	93	19.38	234	48.75
ตอนล่าง	60	11	18.33	32	53.33
1-2 ด้าน	42	0	0.00	20	47.62
3-4 ด้าน	498	12	2.41	246	49.40

จากตารางที่ 4.13 เปรียบเทียบผลการทำนายคลาสการเกิดฝนฟ้าคะนองได้ถูกต้องจากการนำไปใช้จริง สำหรับระยะเวลาพยากรณ์ 1 ชั่วโมงข้างหน้า พบว่า แบบจำลองที่ดีที่สุดสำหรับแต่ละพื้นที่สนามบินในงานวิจัยนี้มีจำนวนครั้งการทำนายการเกิดพายุฝนฟ้าคะนองได้ถูกต้อง 70.68%-83.33% ซึ่งมากกว่าการทำนายด้วยนักอุตุนิยมวิทยาผู้ปฏิบัติงานที่สามารถทำนายการเกิดพายุฝนฟ้าคะนองได้ถูกต้องเท่ากับ 46.88%-50.83% เท่านั้น ดังนั้นแบบจำลองในงานวิจัยนี้สามารถใช้เป็นเครื่องมือสนับสนุนในการพยากรณ์การเกิดฝนฟ้าคะนอง สำหรับระยะเวลาพยากรณ์ 1 ชั่วโมงข้างหน้าได้ดี

จากตารางที่ 4.14 เปรียบเทียบผลการทำนายคลาสการเกิดฝนฟ้าคะนองได้ถูกต้องจากการนำไปใช้จริง สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมงข้างหน้า พบว่า แบบจำลองที่ดีที่สุดสำหรับแต่ละพื้นที่สนามบินในงานวิจัยนี้มีจำนวนครั้งการทำนายการเกิดพายุฝนฟ้าคะนองได้ถูกต้อง 0%-19.38% ซึ่งน้อยกว่าการทำนายด้วยนักอุตุนิยมวิทยาผู้ปฏิบัติงานที่สามารถทำนายการเกิดพายุฝนฟ้าคะนองได้ถูกต้องเท่ากับ 47.62%-53.33% โดยเมื่อทดสอบนำเข้าสู่ข้อมูลพื้นที่สนามบินที่มีภูเขาล้อมรอบ 1-2 ด้านในแบบจำลอง Decision Tree ร่วมกับ Bagging นั้นผลลัพธ์แสดงให้เห็นว่าไม่สามารถทำนายการเกิดพายุฝนฟ้าคะนองได้เลย ดังนั้นแบบจำลองในงานวิจัยนี้ไม่สามารถใช้เป็นเครื่องมือสนับสนุนในการพยากรณ์การเกิดฝนฟ้าคะนอง สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมงข้างหน้าได้

บทที่ 5

สรุปการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัย เรื่อง การพยากรณ์การเกิดพายุฝนฟ้าคะนองระยะปัจจุบันจากข่าวอากาศการ
บิน ด้วยเทคนิคการเรียนรู้ของเครื่อง กรณีศึกษาสนามบินภาคเหนือของประเทศไทย มีรายละเอียด
สรุปผลการวิจัย ดังนี้

1. สรุปการวิจัย

วิทยานิพนธ์นี้มีวัตถุประสงค์เพื่อสร้างและประเมินประสิทธิภาพของแบบจำลองในการ
พยากรณ์การเกิดพายุฝนฟ้าคะนองสำหรับระยะเวลาพยากรณ์ 1 และ 2 ชั่วโมงข้างหน้า บริเวณ
สนามบินภาคเหนือของประเทศไทย ซึ่งแบ่งคลาสของการจำแนกที่ต้องการทำนายผลลัพธ์ออกเป็น 2
คลาส คือ ไม่เกิดพายุฝนฟ้าคะนอง (NO TSRA Class) และเกิดพายุฝนฟ้าคะนอง (TSRA Class) โดย
ใช้โปรแกรม RapidMiner Studio ดำเนินการสร้างแบบจำลอง แบ่งการศึกษาเป็น 2 ส่วนคือ ส่วนที่
1 การสร้างแบบจำลองโดยอัลกอริทึมมาตรฐาน ได้แก่ Naive Bayes Decision Tree และ Neural
Networks และส่วนที่ 2 การสร้างแบบจำลองโดยอัลกอริทึมมาตรฐานร่วมกับอัลกอริทึมการเรียนรู้
แบบกลุ่ม ได้แก่ Random Forest Bagging และ AdaBoost จากผลการวิจัยพบว่า

1.1 สำหรับระยะเวลาพยากรณ์ 1 ชั่วโมงข้างหน้า พบว่า คุณลักษณะที่มีความสำคัญมาก
ที่สุด 5 ลำดับแรกคือ เมฆฝนฟ้าคะนอง สภาพอากาศปัจจุบัน ความเร็วลม เดือน และช่วงเวลา ส่วน
ความชื้นสัมพัทธ์ และทัศนวิสัย มีความเกี่ยวข้องกับตัวแปรที่ทำการศึกษาน้อยที่สุด เมื่อดำเนินการปรับ
ค่ากำหนดของพารามิเตอร์ที่เหมาะสมแล้วแบบจำลอง Random Forest มีประสิทธิภาพการ
พยากรณ์การเกิดพายุฝนฟ้าคะนองได้ดีที่สุดสำหรับพื้นที่รวมสนามบินภาคเหนือ

1.2 สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมงข้างหน้า พบว่า ลำดับคุณลักษณะที่มี
ความสำคัญมากที่สุดไม่แน่นอนโดยเปลี่ยนแปลงไปตามแต่ละกลุ่มข้อมูล แต่ความชื้นสัมพัทธ์ และทัศน
วิสัย มีความเกี่ยวข้องกับตัวแปรที่ทำการศึกษาน้อยที่สุด เมื่อดำเนินการปรับค่ากำหนดของ
พารามิเตอร์ที่เหมาะสมแล้ว ทุกแบบจำลองมีประสิทธิภาพต่ำในการพยากรณ์การเกิดพายุฝนฟ้า
คะนอง ซึ่งสอดคล้องกับผลการวิจัยของสมปราชญ์ (2564) ที่พบว่าแบบจำลองมีความแม่นยำลดลง

เมื่อระยะเวลาพยากรณ์นานขึ้น สาเหตุเกิดจากชุดข้อมูลนำเข้าไม่สามารถสร้างรูปแบบการเรียนรู้ที่ถูกต้องให้แก่แบบจำลองได้ อาจเนื่องด้วยลักษณะภูมิอากาศประเทศไทยที่ตั้งอยู่ในเขตร้อนและสารประกอบทางอุตุนิยมวิทยาเปลี่ยนแปลงรวดเร็วแบบไม่มีรูปแบบแน่นอน ส่งผลให้การคาดการณ์สภาพอากาศล่วงหน้าในระยะเวลาพยากรณ์ 2 ชั่วโมงมีความแม่นยำน้อยลง

1.3 การทดสอบโดยการนำแบบจำลองไปใช้จริงเปรียบเทียบกับพยากรณ์โดยนักอุตุนิยมวิทยาสำหรับระยะเวลาพยากรณ์ 1 ชั่วโมงข้างหน้า พบว่า แบบจำลองที่ดีที่สุดของแต่ละพื้นที่สนามบินมีประสิทธิภาพการทำนายการเกิดพายุฝนฟ้าคะนองได้ถูกต้องมากกว่าการพยากรณ์โดยคน ดังนั้นแบบจำลองในงานวิจัยนี้สามารถใช้เป็นเครื่องมือสนับสนุนในการพยากรณ์การเกิดพายุฝนฟ้าคะนองสำหรับระยะเวลาพยากรณ์ 1 ชั่วโมงข้างหน้าได้ดี

1.4 การทดสอบโดยการนำแบบจำลองไปใช้จริงเปรียบเทียบกับพยากรณ์โดยนักอุตุนิยมวิทยาสำหรับระยะเวลาพยากรณ์ 2 ชั่วโมงข้างหน้า พบว่า แบบจำลองที่ดีที่สุดของแต่ละพื้นที่สนามบินไม่สามารถทำนายผลการเกิดพายุฝนฟ้าคะนองได้ถูกต้อง ซึ่งการพยากรณ์โดยคนได้ผลได้ดีกว่า ดังนั้นสำหรับการพยากรณ์การเกิดพายุฝนฟ้าคะนอง ระยะเวลาพยากรณ์ 2 ชั่วโมงข้างหน้า ไม่สามารถใช้แบบจำลองในงานวิจัยนี้เป็นเครื่องมือสนับสนุนได้

2. อภิปรายผล

ในงานวิจัยนี้แบ่งชุดข้อมูลทั้งหมดเป็น 5 กลุ่มข้อมูลตามขอบเขตและลักษณะพื้นที่ของสนามบิน โดยจากการวิจัยพบว่า แบบจำลองที่ดีที่สุดของแต่ละกลุ่มข้อมูลให้ค่าความแม่นยำที่แตกต่างกันเพียงเล็กน้อยเท่านั้น โดยการเลือกแบบจำลองใดไปใช้ประโยชน์นั้นขึ้นอยู่กับวัตถุประสงค์ของผู้ใช้งาน เช่น การเลือกใช้อัลกอริทึม Random Forest ซึ่งมีประสิทธิภาพการจำแนกที่ดีเพียงพอสำหรับทุกกลุ่มพื้นที่สนามบิน ซึ่งค่า AUC ใกล้เคียงกันกับแบบจำลองที่ดีที่สุดในแต่ละกลุ่มข้อมูลในระดับที่ยอมรับได้ และมีความเร็วในการประมวลผลสูงกว่าการใช้อัลกอริทึม Neural Networks แต่หากต้องการประสิทธิภาพสูงสุดสำหรับพยากรณ์การเกิดพายุฝนฟ้าคะนองระยะปัจจุบันของแต่ละพื้นที่ ควรเลือกใช้แบบจำลองที่มีประสิทธิภาพสูงสุดของแต่ละกลุ่มข้อมูลของทั้ง 5 พื้นที่สนามบินสำหรับการนำแบบจำลองในงานวิจัยนี้ไปใช้กับพื้นที่อื่นๆของประเทศ หากต้องการประสิทธิภาพสูงสุดให้พิจารณาลักษณะภูมิประเทศของจังหวัดๆนั้น ได้แก่ หากมีลักษณะภูมิประเทศเป็นแบบมีภูเขาล้อมรอบ 1-2 ด้านและ 3-4 ด้าน สามารถเลือกใช้แบบจำลองตามงานวิจัยนี้ได้ แต่หากมีลักษณะพื้นที่เป็นที่ราบหรือลักษณะอื่นๆ ควรเลือกใช้แบบจำลองสำหรับพื้นที่รวมสนามบินภาคเหนือ นั่นคือ

อัลกอริทึม Random Forest ที่ให้ผลการพยากรณ์ที่ดี เนื่องจากในงานวิจัยนี้ค้นพบว่า คุณลักษณะ ทิศทางลมซึ่งเป็นตัวแปรที่มีความแตกต่างกันไปในแต่ละพื้นที่ของประเทศไทยมีลำดับความสำคัญ น้อยต่อการพยากรณ์การเกิดพายุฝนฟ้าคะนอง จึงสามารถพิจารณานำแบบจำลองในงานวิจัยนี้ไปใช้ สำหรับพื้นที่อื่นๆในประเทศไทยได้เช่นเดียวกัน

3. ข้อจำกัดของงานวิจัย

ในงานวิจัยนี้ยังมีข้อจำกัดในงานวิจัย รายละเอียดดังต่อไปนี้

3.1 ข้อมูลที่ใช้ในงานวิจัยรวบรวมจากเว็บไซต์กระจายข่าวทางอุตุนิยมวิทยา พบว่า มี ข้อมูลข่าวอากาศการบินบางส่วนที่มีการรายงานข่าวไม่สมบูรณ์ ซึ่งเกิดจากความผิดพลาดในการ รายงานข่าวและเครื่องมือการตรวจอากาศการบินขัดข้อง ทำให้ต้องพิจารณาตัดชุดข้อมูลเหล่านั้นทิ้ง ไป จึงมีข้อมูลนำเข้าลดลงจากข้อมูลข่าวทั้งหมดที่ได้ทำการตรวจวัดและรายงานจริง

3.2 ข้อมูลข่าวอากาศการบินที่นำเข้าเพื่อทำการวิจัยในครั้งนี้มีรูปแบบของข่าวที่ หลากหลายในแต่ละสนามบิน เนื่องจากแต่ละแห่งมีชนิดและจำนวนของอุปกรณ์ตรวจวัดที่แตกต่างกัน จึงต้องมีการจัดการข้อมูลในขั้นตอนการเตรียมข้อมูลก่อนสร้างแบบจำลองให้เหมาะสมและเป็น รูปแบบเดียวกัน

3.3 ข้อมูลข่าวอากาศการบินที่จะนำเข้าเพื่อใช้ในการสร้างแบบจำลองจะต้องมีการแปลง รูปแบบเป็นช่วงข้อมูลตามการศึกษาในงานวิจัยนี้ก่อน จึงจะทำให้มีประสิทธิภาพของการพยากรณ์ การเกิดพายุฝนฟ้าคะนองได้ดี

4. ข้อเสนอแนะ

แบบจำลองเพื่อการพยากรณ์การเกิดพายุฝนฟ้าคะนองระยะปัจจุบันที่ผู้วิจัยนำเสนอใน วิทยานี้สามารถพัฒนาต่อยอดงานวิจัยในอนาคตได้ซึ่งการพัฒนาวิจัยนี้จะเป็นประโยชน์อย่างยิ่ง สำหรับการสนับสนุนหน่วยงานทางด้านการควบคุมการจราจรทางอากาศ และการวางแผนการบิน ของสายการบินในพื้นที่ที่ขณะเกิดฝนตกหนักและพายุฝนฟ้าคะนอง ดังนี้

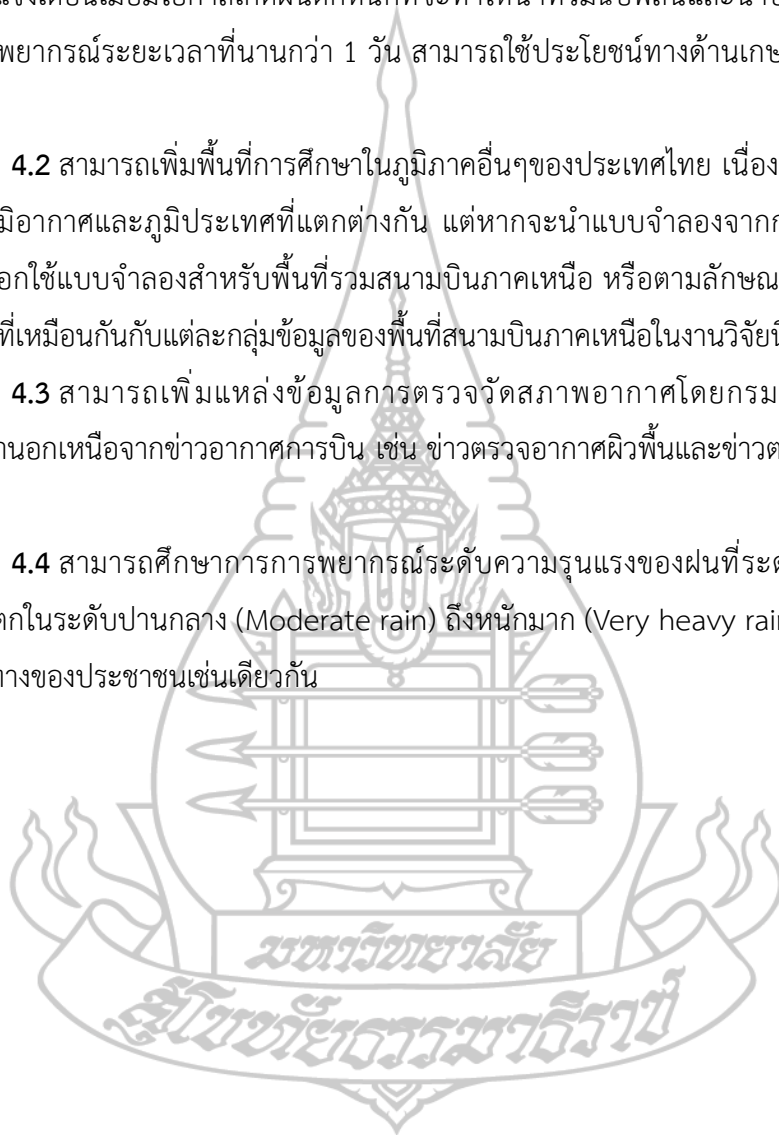
4.1 สามารถทำการศึกษาระยะเวลาการพยากรณ์อื่นๆเพิ่มเติมนอกจากในงานวิจัยนี้ อาจ มีการศึกษาระยะเวลาที่สั้นกว่านี้ หรือศึกษาการเพิ่มประสิทธิภาพของแบบจำลองที่ระยะเวลายาวนาน กว่านี้ เพื่อการแจ้งเตือนภัยการเกิดฝนตกหนักและพายุฝนฟ้าคะนองที่ทันเวลา โดยทางอุตุนิยมวิทยา

จะใช้ข้อมูลการพยากรณ์ที่สั้นกว่า 1 ชั่วโมง สำหรับเป็นการพยากรณ์อากาศที่จะแสดงผลผ่านแอปพลิเคชันแบบเรียลไทม์ให้แก่ประชาชน ส่วนการพยากรณ์อากาศ 1 และ 2 ชั่วโมงข้างหน้าใช้สำหรับการพยากรณ์อากาศการบิน สำหรับการพยากรณ์อากาศที่มีระยะเวลาเกินกว่า 2 ชั่วโมงถึง 1 วันจะใช้สำหรับการแจ้งเตือนเมื่อมีโอกาสเกิดฝนตกหนักที่จะทำให้ น้ำท่วมฉับพลันและน้ำป่าไหลหลาก และสำหรับการพยากรณ์ระยะเวลาที่นานกว่า 1 วัน สามารถใช้ประโยชน์ทางด้านเกษตรกรรมและการท่องเที่ยว

4.2 สามารถเพิ่มพื้นที่การศึกษาในภูมิภาคอื่นๆของประเทศไทย เนื่องจากแต่ละภูมิภาคมีลักษณะภูมิอากาศและภูมิประเทศที่แตกต่างกัน แต่หากจะนำแบบจำลองจากการวิจัยนี้ไปใช้ให้พิจารณาเลือกใช้แบบจำลองสำหรับพื้นที่รวมสนามบินภาคเหนือ หรือตามลักษณะภูมิประเทศของจังหวัดต่างๆที่เหมือนกันกับแต่ละกลุ่มข้อมูลของพื้นที่สนามบินภาคเหนือในงานวิจัยนี้

4.3 สามารถเพิ่มแหล่งข้อมูลการตรวจวัดสภาพอากาศโดยกรมอุตุนิยมวิทยาที่ทำการศึกษานอกเหนือจากข่าวอากาศการบิน เช่น ข่าวตรวจอากาศผิวพื้นและข่าวตรวจอากาศชั้นบน เป็นต้น

4.4 สามารถศึกษาการการพยากรณ์ระดับความรุนแรงของฝนที่ระดับต่างๆ เช่น ฝนธรรมดาแต่ตกในระดับปานกลาง (Moderate rain) ถึงหนักมาก (Very heavy rain) ที่ส่งผลกระทบต่อการเดินทางของประชาชนเช่นเดียวกัน





บรรณานุกรม

- กรมอุตุนิยมวิทยา. (2554). การพยากรณ์อากาศ. <https://www.tmd.go.th/info/การพยากรณ์อากาศ-1>
- กองอุตุนิยมวิทยาการบิน. (2558). การบริการข้อมูลอุตุนิยมวิทยาการบิน. <http://www2.aeromet.tmd.go.th/KnowledgeService.php>
- กองอุตุนิยมวิทยาการบิน. (2558). พายุฟ้าคะนอง. <http://www2.aeromet.tmd.go.th/KnowledgeOTS.php>
- กาญจน์ ณ ศรีระ, กิตติศักดิ์ เกิดประสพ, และนิตยา เกิดประสพ. (2018). การเปรียบเทียบเทคนิคการสุ่มตัวอย่างเพื่อการจำแนกข้อมูลที่ไม่สมดุล. *Journal of Applied Informatics and Technology*, 1(2018), 20-37. <https://doi.org/10.14456/jait.2018.2>
- ไกรศักดิ์ เกสร. (2560). โครงข่ายประสาทเทียมอัจฉริยะ. <https://csit.nu.ac.th/kraisak/ds/ds/chapter07/Chapter07.pdf>
- ธงไชย พ้องเสียง และจารี ทองคำ. (2566). การเพิ่มประสิทธิภาพแบบจำลองเพื่อการพยากรณ์การรักรักษาโรคความดันโลหิตสูงด้วยการคัดเลือกปัจจัย. *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยอุบลราชธานี*, 1(2566), 13-20.
- ธรรมนุญ ปัญญาทิพย์, ปณิตดา โปธินาม, และคมกริช อ่อนประสงค์. การเพิ่มประสิทธิภาพจำแนกข้อมูลผลกระทบโควิด-19 ต่อผู้ป่วยมะเร็งเรื้อรัง. *วารสารวิชาการการจัดการเทคโนโลยี มหาวิทยาลัยราชภัฏมหาสารคาม*, 2(2020), 66-78. <https://ph02.tci-thaijo.org/index.php/itm-journal/article/view/247698>
- ธีรยุทธ คุณสุข และจารี ทองคำ. (2563). กระบวนการคัดเลือกคุณลักษณะสำหรับเพิ่มประสิทธิภาพในการจำแนกความคิดเห็นของลูกค้า. *RMUTI JOURNAL Science and Technology*, 13(2020), 129-143.
- นันทชพร เสนาวงศ์, สุภาวดี วิชิตชาญ, และอรวิษญ์ กุมพล. (2565). การจำแนกข้อมูลขนาดใหญ่โดยใช้การจัดกลุ่มด้วยวิธีเคมีนและวิธีการเรียนรู้เชิงลึก. *The Journal of King Mongkut's University of Technology North Bangkok*, 32(4), 957-965.
- เบญจพร เอี่ยมประโคน. (2560). วิธีการเปรียบเทียบพื้นที่ใต้โค้ง ROC สำหรับข้อมูลชุดเดียวกัน: กรณีศึกษาแบบจำลองคะแนนเครดิต [วิทยานิพนธ์ปริญญาโท]. Chula Digital Collections. <https://digital.car.chula.ac.th/cgi/viewcontent.cgi?article=3019&context=chulaetd>

ปรเมษฐ์ ชันวานนท์, ชัยกร ยิ่งเสรี, วรพล พงษ์เพ็ชร, และธนภัทร ชังคะจิตร. (2560). การประยุกต์ใช้โมเดลการเรียนรู้แบบรวมกลุ่มเพื่อพยากรณ์แนวโน้มของราคาหลักทรัพย์ในตลาดหลักทรัพย์แห่งประเทศไทย. *Journal Of Information Science Technology*. 1(2017),12-21.

พิชญา พรหมลา และจรัญ แสนราช. (2563). การเปรียบเทียบประสิทธิภาพการวิเคราะห์ความพึงพอใจเกี่ยวกับการจัดการเรียนการสอนด้วยกระบวนการวิเคราะห์ความรู้สึกโดยใช้เทคนิคการรวมกลุ่มเพื่อจำแนกข้อมูล. *KKU Research Journal (Graduate Studies)*, 4(2563), 140-149.

เพ็ญลดา มะโนกิจ. (2562). การพยากรณ์ปริมาณน้ำฝนของประเทศไทยด้วยวิธีการเรียนรู้เชิงลึก [วิทยานิพนธ์ปริญญาโทมหาบัณฑิต]. Chula Digital Collections.

<https://digital.car.chula.ac.th/cgi/viewcontent.cgi?article=10499&context=chulaetd>

ศูนย์ภูมิอากาศ กรมอุตุนิยมวิทยา. (2537). ลักษณะอากาศรายจังหวัด.

<http://climate.tmd.go.th/map/thailand>

สมปราษฎ์ ศรีถกล. (2564). การพยากรณ์ฝนปัจจุบันด้วยเครื่องจักรการเรียนรู้ โดยอิงข้อมูลน้ำฝนตรวจวัดในกรุงเทพมหานคร [วิทยานิพนธ์ปริญญาโทมหาบัณฑิต]. ฐานข้อมูลดิจิทัล มก.

สายชล สิ้นสมบุรณ์ทอง. (2563). การเปรียบเทียบประสิทธิภาพในการทำนายผลความไม่สมดุลของข้อมูลในการจำแนกด้วยเทคนิคการทำเหมืองข้อมูล. *Thai Science and Technology Journal (TSTJ)*, 28(3), 383-393. <https://doi.org/10.14456/tstj.2020.31>

สำนักงานบินพลเรือนแห่งประเทศไทย. (2566, 19 ตุลาคม). เอกสารแถลงข่าวการบิน (Aeronautical Information Publication: AIP). <https://aip.caat.or.th/2023-11-30-AIRAC/html/index-en-GB.html>

สำนักงานสิ่งแวดล้อมภาคเหนือ. (2563, 19 กุมภาพันธ์). การวิเคราะห์ศักยภาพพื้นที่เพื่อรองรับความเสี่ยงด้านมลพิษทางน้ำ. <https://epo02.pcd.go.th/th/news/detail/54855>

หัสพล ธีมมิกรัตน์. (2020). การวินิจฉัยโรคพาร์กินสันโดยใช้การเรียนรู้ของเครื่อง [วิทยานิพนธ์ปริญญาโทมหาบัณฑิต]. Chula Digital Collections.

<https://doi.org/10.58837/CHULA.THE.2020.1134>

หนึ่งฤทัย ประเสริฐ. (2021). การออกแบบและพัฒนาแบบจำลองการจำแนกความผิดปกติของเสียงหัวใจ [วิทยานิพนธ์ปริญญาโทมหาบัณฑิต]. Suranaree University of Technology Intellectual Repository: SUTIR.

<http://sutir.sut.ac.th:8080/sutir/browse?type=author&value=หนึ่งฤทัย+ประเสริฐ>

อัจฉรา แพ้วบาง และสายชล สีนสมบูรณ์ทอง. (2562). การปรับความไม่สมดุลของข้อมูลด้วยการจำแนก 5 วิธี. *Thai Journal of Science and Technology*, 9(4), 418–435.

<https://doi.org/10.14456/tjst.2020.46>

อรุณี ศรีดี. (2564). การประยุกต์ใช้การเรียนรู้ของเครื่องเพื่อตรวจจับงานเสียที่เกิดขึ้นในกระบวนการทดสอบฮาร์ดดิสก์ไดรฟ์ [วิทยานิพนธ์ปริญญาโท]. Chula Digital Collections. https://www.cp.eng.chula.ac.th/~prabhas//thesis/Thesis_Arunee_2021.pdf

Alnawas, A., Khafaji, N. J. A., & Azeez, H. H. (2021). Precipitation Forecast for Thi-Qar Province of Iraq Utilizing Machine Learning Approaches. *Proceedings of 2nd International Multi-Disciplinary Conference Theme: Integrated Sciences and Technologies*. Turkey: EAI. <http://dx.doi.org/10.4108/eai.7-9-2021.2314897>

Ashish P. (2019). Bagging-Ensemble meta-Algorithm for Reducing variance. <https://medium.com/ml-research-lab/bagging-ensemble-meta-algorithm-for-reducing-variance-c98fffa5489f>

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. doi: 10.1016/j.ins.2019.01.041

Devos, A. K., Huffel, S. V., Simonetti, A. W., Graaf, M. V. D., Heerschap, A., Buydens, L. M.C. (2007). Chapter 11 - Classification of Brain Tumours by Pattern Recognition of Magnetic Resonance Imaging and Spectroscopic Data. In Azzam F.G. Taktak and Anthony C. Fisher (Eds.), *Outcome Prediction in Cancer* (pp. 285-318). Elsevier Science.

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.

Geetha, A., & Nasira, G. M. (2014). Data mining for meteorological applications: Decision trees for modeling rainfall prediction. *Proceedings of 2014 IEEE International Conference on Computational Intelligence and Computing Research*. India: IEEE. <https://doi.org/10.1109/ICCIC.2014.7238481>

Ghada, W., Estrella, N., & Menzel, A. (2019). Machine Learning Approach to Classify Rain Type Based on Thies Disdrometers and Cloud Observations.

Atmosphere, 10(2019). <https://doi.org/10.3390/atmos10050251>

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 (pp. 87-92). Moradabad, India: IEEE.

Kumar D. (2021). How did The Random Forest Algorithm work in Machine Learning?. <https://divakarkumarp.medium.com/how-did-the-random-forest-algorithm-work-in-machine-learning-9e044573898b>

Kung CK. (2020). ทำไมค่า ROC AUC Score เยอะ แต่ Precision Recall Score ถึงน้อยกันั้นน่ะ?. <https://medium.com/kbtg-life/what-will-happen-about-average-precision-score-and-roc-auc-score-when-data-is-highly-imbalanced-4e40b4d9e005>

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. ScienceDirect, 568(2019), 1042-1054.

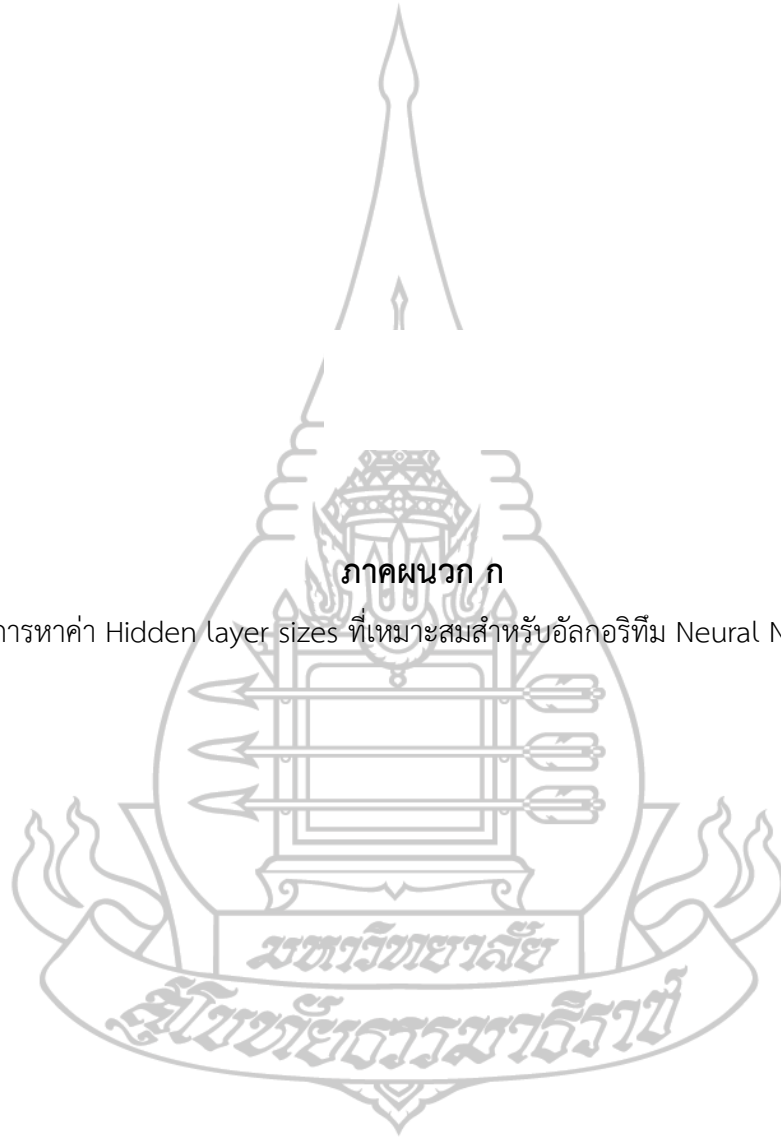
Ojo, O. S., & Ogunjo, S. T. (2022). Machine learning models for prediction of rainfall over Nigeria. Scientific African, 16(2022). doi.org/10.1016/j.sciaf.2022.e01246

Pirone, D. Cimorelli, L., Giudice, G. D., Pianese, D. (2022). Rainfall Nowcasting Exploiting Machine-Learning Techniques: A Case Study in Southern Italy. Environmental Sciences Proceedings. 21(2022). <https://doi.org/10.3390/environsciproc2022021049>

Raman R. (2020). Boosting in Machine Learning | Boosting and AdaBoost. <https://www.geeksforgeeks.org/boosting-in-machine-learning-boosting-and-adaboost>

Shah, N. H., Shukla, B. P., & Priamvada, A. (2022). Ensemble-based Model for Rainfall Nowcasting using Automatic Weather Station Data. JOURNAL OF Engineering Science and Technology Review, 15(2022), 111-116. <https://doi.org/10.25103/jestr.154.16>

Sunori, S.K., Babu, L. R. A., Shermina, J., Sam, D., Justin, S., Maurya, S., & Juneja, P. (2021). Classification of Rainfall Levels using Various Machine Learning Techniques. Proceeding of IEEE Mysore Sub Section International Conference. India: IEEE. <https://doi.org/10.1109/MysuruCon52639.2021.9641084>



ภาคผนวก ก

ผลการหาค่า Hidden layer sizes ที่เหมาะสมสำหรับอัลกอริทึม Neural Networks

ตารางภาคผนวกที่ 1 ผลการทดสอบเพื่อหาค่า Hidden layer sizes ที่เหมาะสมสำหรับ อัลกอริทึม Neural Networks สำหรับระยะเวลาพยากรณ์ 1 ชั่วโมง กลุ่ม ข้อมูลที่ 1 รวมสนามบินภาคเหนือ

อัลกอริทึม	Hidden layer sizes	การวัดประสิทธิภาพ				
		Accuracy	Precision	Recall	F-measure	AUC
Neural Networks	1	99.60	95.59	63.14	76.05	0.856
	2	99.60	95.58	62.95	75.91	0.855
	3	99.60	94.89	63.05	75.76	0.865
	4	99.60	95.18	63.18	75.95	0.876
	5	99.60	95.62	63.29	76.16	0.876
	6	99.60	96.05	63.14	76.19	0.897
	7	99.60	95.76	63.04	76.03	0.885
	8	99.60	96.05	63.14	76.19	0.893
Bagging+ Neural Networks	1	99.59	94.83	62.51	75.35	0.837
	2	99.60	95.30	62.83	75.73	0.839
	3	99.60	95.75	63.13	76.09	0.840
	4	99.60	95.61	63.09	76.02	0.840
	5	99.60	95.59	62.80	75.80	0.839
	6	99.60	95.89	62.97	76.02	0.840
	7	99.60	96.04	63.04	76.12	0.840
	8	99.61	96.21	63.40	76.43	0.842
AdaBoost + Neural Networks	1	99.60	95.13	62.80	75.66	0.839
	2	99.60	95.58	62.91	75.88	0.839
	3	99.60	95.87	62.98	76.02	0.840
	4	99.60	95.90	63.16	76.16	0.841
	5	99.61	96.17	63.09	76.20	0.840
	6	99.60	96.04	63.04	76.12	0.840
	7	99.60	95.89	62.97	76.02	0.840
	8	99.60	96.19	63.11	76.22	0.840

ตารางภาคผนวกที่ 2 ผลการทดสอบเพื่อหาค่า Hidden layer sizes ที่เหมาะสมสำหรับ อัลกอริทึม Neural Networks สำหรับระยะเวลาพยากรณ์ 1 ชั่วโมง กลุ่ม ข้อมูลที่ 2 สนามบินภาคเหนือตอนบน

อัลกอริทึม	Hidden layer sizes	การวัดประสิทธิภาพ				
		Accuracy	Precision	Recall	F-measure	AUC
Neural Networks	1	99.57	96.17	63.96	76.82	0.845
	2	99.57	96.15	63.87	76.75	0.844
	3	99.57	96.17	64.16	76.97	0.846
	4	99.57	96.55	64.36	77.24	0.847
	5	99.58	96.72	64.32	77.26	0.846
	6	99.58	96.90	64.48	77.44	0.847
	7	99.58	96.54	64.40	77.26	0.847
	8	99.57	96.36	64.20	77.06	0.846
Bagging+ Neural Networks	1	99.55	94.07	63.03	75.48	0.840
	2	99.55	94.44	63.12	75.67	0.840
	3	99.56	94.84	63.42	76.01	0.842
	4	99.56	95.02	63.35	76.01	0.842
	5	99.56	95.59	63.73	76.47	0.843
	6	99.57	95.97	64.06	76.83	0.845
	7	99.57	95.96	63.94	76.74	0.845
	8	99.57	96.33	64.02	76.92	0.845
AdaBoost + Neural Networks	1	99.56	93.97	64.79	76.70	0.849
	2	99.56	94.35	64.96	76.95	0.850
	3	99.57	94.53	65.05	77.07	0.850
	4	99.56	94.36	65.01	76.98	0.850
	5	99.57	94.55	65.29	77.24	0.851
	6	99.57	94.72	65.29	77.30	0.851
	7	99.57	94.75	65.66	77.56	0.853
	8	99.58	94.92	65.70	77.65	0.853

ตารางภาคผนวกที่ 3 ผลการทดสอบเพื่อหาค่า Hidden layer sizes ที่เหมาะสมสำหรับ อัลกอริทึม Neural Networks สำหรับระยะเวลาพยากรณ์ 1 ชั่วโมง กลุ่ม ข้อมูลที่ 3 สนามบินภาคเหนือตอนล่าง

อัลกอริทึม	Hidden layer sizes	การวัดประสิทธิภาพ				
		Accuracy	Precision	Recall	F-measure	AUC
Neural Networks	1	99.64	98.15	50.24	66.46	0.776
	2	99.65	98.18	51.18	67.29	0.781
	3	99.66	100.00	51.64	68.11	0.783
	4	99.66	100.00	52.80	69.11	0.789
	5	99.67	100.00	53.74	69.91	0.794
	6	99.67	99.15	54.67	70.48	0.794
	7	99.67	99.15	54.67	70.48	0.798
	8	99.68	100.00	54.67	70.69	0.813
Bagging+ Neural Networks	1	99.67	96.52	53.88	69.16	0.794
	2	99.68	98.26	54.33	69.97	0.797
	3	99.67	96.58	54.33	69.54	0.797
	4	99.68	98.29	54.76	70.34	0.799
	5	99.67	96.64	54.76	69.91	0.799
	6	99.68	99.17	56.07	71.64	0.805
	7	99.68	97.50	55.45	70.69	0.802
	8	99.68	99.17	55.87	71.47	0.804
AdaBoost + Neural Networks	1	99.66	97.35	52.88	68.54	0.789
	2	99.66	98.21	52.63	68.54	0.788
	3	99.66	98.25	53.08	68.92	0.790
	4	99.67	99.14	54.25	70.12	0.796
	5	99.68	100.00	54.67	70.69	0.798
	6	99.67	97.48	54.98	70.30	0.800
	7	99.68	98.32	55.19	70.69	0.801
	8	99.69	100.00	56.07	71.86	0.805

ตารางภาคผนวกที่ 4 ผลการทดสอบเพื่อหาค่า Hidden layer sizes ที่เหมาะสมสำหรับ อัลกอริทึม Neural Networks สำหรับระยะเวลาพยากรณ์ 1 ชั่วโมง กลุ่ม ข้อมูลที่ 4 สนามบินที่มีภูเขาล้อมรอบ 1-2 ด้าน

อัลกอริทึม	Hidden layer sizes	การวัดประสิทธิภาพ				
		Accuracy	Precision	Recall	F-measure	AUC
Neural Networks	1	99.56	98.84	48.85	65.38	0.769
	2	99.56	97.70	49.13	65.38	0.771
	3	99.57	96.67	50.29	66.16	0.776
	4	99.57	97.83	51.14	67.16	0.781
	5	99.57	96.77	51.43	67.16	0.782
	6	99.59	100.00	52.81	69.12	0.798
	7	99.59	98.94	52.54	68.63	0.788
	8	99.59	100.00	52.81	69.12	0.743
Bagging+ Neural Networks	1	99.56	96.67	50.00	65.91	0.775
	2	99.57	96.77	51.14	66.91	0.781
	3	99.57	96.81	51.70	67.41	0.783
	4	99.59	97.92	53.11	68.86	0.790
	5	99.59	98.96	53.07	69.09	0.790
	6	99.58	96.88	52.54	68.13	0.788
	7	99.59	96.94	53.67	69.09	0.793
	8	99.60	98.98	54.19	70.04	0.796
AdaBoost + Neural Networks	1	99.57	96.59	50.00	65.89	0.775
	2	99.58	97.75	50.88	66.92	0.779
	3	99.57	95.51	50.30	65.89	0.776
	4	99.58	96.70	51.16	66.92	0.781
	5	99.59	96.77	52.33	67.92	0.787
	6	99.59	96.94	53.67	69.09	0.793
	7	99.59	96.88	53.14	68.63	0.791
	8	99.58	96.88	52.54	68.13	0.788

ตารางภาคผนวกที่ 5 ผลการทดสอบเพื่อหาค่า Hidden layer sizes ที่เหมาะสมสำหรับ อัลกอริทึม Neural Networks สำหรับระยะเวลาพยากรณ์ 1 ชั่วโมง กลุ่ม ข้อมูลที่ 5 สนามบินที่มีภูเขาล้อมรอบ 3-4 ด้าน

อัลกอริทึม	Hidden layer sizes	การวัดประสิทธิภาพ				
		Accuracy	Precision	Recall	F-measure	AUC
Neural Networks	1	99.59	94.33	61.78	74.66	0.834
	2	99.59	94.63	61.95	74.88	0.835
	3	99.59	94.81	62.26	75.16	0.836
	4	99.59	94.67	62.22	75.09	0.836
	5	99.59	95.13	62.44	75.39	0.837
	6	99.60	95.17	62.74	75.63	0.839
	7	99.60	95.32	62.87	75.77	0.839
	8	99.60	95.61	62.95	75.91	0.851
Bagging+ Neural Networks	1	99.59	95.03	63.39	76.05	0.842
	2	99.60	95.40	63.71	76.40	0.843
	3	99.60	95.95	64.04	76.81	0.845
	4	99.60	95.77	63.88	76.64	0.844
	5	99.61	96.15	64.21	77.00	0.846
	6	99.61	96.86	64.69	77.57	0.848
	7	99.61	96.33	64.29	77.12	0.846
	8	99.61	96.49	64.25	77.14	0.846
AdaBoost + Neural Networks	1	99.60	96.09	63.72	76.63	0.843
	2	99.60	96.27	63.76	76.72	0.844
	3	99.60	95.91	63.68	76.54	0.843
	4	99.61	96.46	63.89	76.87	0.844
	5	99.60	96.46	63.82	76.81	0.844
	6	99.61	96.48	64.17	77.07	0.846
	7	99.61	96.31	64.17	77.02	0.846
	8	99.61	96.32	64.33	77.14	0.847

ตารางภาคผนวกที่ 6 ผลการปรับค่า Hidden layer sizes ที่เหมาะสมสำหรับอัลกอริทึม Neural Networks สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมง กลุ่มข้อมูลที่ 1 รวมสนามบินภาคเหนือ

อัลกอริทึม	Hidden layer sizes	การวัดประสิทธิภาพ				
		Accuracy	Precision	Recall	F-measure	AUC
Neural Networks	1	99.15	91.43	15.55	26.58	0.678
	2	99.15	90.75	15.30	26.19	0.676
	3	99.15	92.09	15.79	26.96	0.679
	4	99.15	92.13	15.88	27.09	0.679
	5	99.15	92.57	15.71	26.87	0.679
	6	99.15	93.75	15.97	27.30	0.680
	7	99.15	91.53	15.71	26.82	0.678
	8	99.15	90.91	15.55	26.56	0.678
Bagging+ Neural Networks	1	99.15	92.49	15.56	26.64	0.678
	2	99.15	92.57	15.74	26.91	0.679
	3	99.15	93.18	15.95	27.24	0.680
	4	99.15	92.18	16.02	27.30	0.680
	5	99.15	92.78	16.18	27.56	0.681
	6	99.15	93.85	16.26	27.72	0.681
	7	99.15	92.66	15.95	27.22	0.680
	8	99.15	92.57	15.74	26.91	0.679
AdaBoost + Neural Networks	1	99.15	91.33	15.41	26.38	0.677
	2	99.15	92.40	15.38	26.38	0.677
	3	99.15	93.02	15.56	26.67	0.678
	4	99.15	92.74	16.10	27.44	0.680
	5	99.15	92.27	16.18	27.54	0.681
	6	99.15	93.30	16.18	27.58	0.681
	7	99.15	93.22	15.99	27.30	0.680
	8	99.15	93.18	15.92	27.20	0.680

ตารางภาคผนวกที่ 7 ผลการปรับหาค่า Hidden layer sizes ที่เหมาะสมสำหรับอัลกอริทึม Neural Networks สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมง กลุ่มข้อมูลที่ 2 สนามบินภาคเหนือตอนบน

อัลกอริทึม	Hidden layer sizes	การวัดประสิทธิภาพ				
		Accuracy	Precision	Recall	F-measure	AUC
Neural Networks	1	99.01	77.78	16.95	27.83	0.684
	2	99.01	77.60	17.19	28.15	0.686
	3	99.02	78.14	17.31	28.34	0.686
	4	99.03	81.01	17.55	28.86	0.688
	5	99.02	78.57	17.29	28.34	0.686
	6	99.02	78.80	17.53	28.68	0.687
	7	99.02	79.78	17.61	28.85	0.688
	8	99.02	80.65	18.05	29.50	0.690
Bagging+ Neural Networks	1	99.02	78.21	16.99	27.92	0.685
	2	99.02	78.57	17.31	28.37	0.686
	3	99.02	79.56	17.41	28.57	0.687
	4	99.02	79.67	17.51	28.71	0.687
	5	99.02	79.35	17.63	28.85	0.688
	6	99.02	79.67	17.51	28.71	0.687
	7	99.02	79.89	17.73	29.02	0.688
	8	99.02	80.00	17.83	29.16	0.689
AdaBoost + Neural Networks	1	99.01	76.54	16.67	27.37	0.683
	2	99.01	76.67	16.79	27.54	0.684
	3	99.02	77.35	17.03	27.92	0.685
	4	99.02	77.47	17.13	28.06	0.685
	5	99.02	78.02	17.23	28.23	0.686
	6	99.02	78.57	17.29	28.34	0.686
	7	99.02	78.69	17.43	28.54	0.687
	8	99.02	79.35	17.63	28.85	0.688

ตารางภาคผนวกที่ 8 ผลการปรับค่า Hidden layer sizes ที่เหมาะสมสำหรับอัลกอริทึม Neural Networks สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมง กลุ่มข้อมูลที่ 3 สนามบินภาคเหนือตอนล่าง

อัลกอริทึม	Hidden layer sizes	การวัดประสิทธิภาพ				
		Accuracy	Precision	Recall	F-measure	AUC
Neural Networks	1	99.36	57.58	9.64	16.52	0.648
	2	99.37	58.82	10.15	17.32	0.651
	3	99.37	61.76	10.55	18.03	0.653
	4	99.36	57.14	10.15	17.24	0.651
	5	99.37	62.86	11.06	18.80	0.655
	6	99.38	63.89	11.62	19.66	0.658
	7	99.38	64.86	12.06	20.34	0.660
	8	99.38	67.57	12.44	21.01	0.662
Bagging+ Neural Networks	1	99.36	54.84	8.72	15.04	0.643
	2	99.36	54.55	9.23	15.79	0.646
	3	99.36	57.58	9.64	16.52	0.648
	4	99.37	57.14	10.20	17.32	0.651
	5	99.37	62.86	11.11	18.88	0.655
	6	99.38	63.89	11.62	19.66	0.658
	7	99.38	69.44	12.44	21.10	0.662
	8	99.38	68.42	12.94	21.76	0.664
AdaBoost + Neural Networks	1	99.36	54.84	8.72	15.04	0.643
	2	99.36	55.88	9.69	16.52	0.648
	3	99.36	54.29	9.74	16.52	0.648
	4	99.37	58.82	10.15	17.32	0.651
	5	99.37	60.00	10.66	18.10	0.653
	6	99.37	62.86	11.11	18.88	0.655
	7	99.37	64.86	12.00	20.25	0.660
	8	99.38	70.27	12.87	21.76	0.664

ตารางภาคผนวกที่ 9 ผลการปรับค่า Hidden layer sizes ที่เหมาะสมสำหรับอัลกอริทึม Neural Networks สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมง กลุ่มข้อมูลที่ 4 สนามบินที่มีภูเขาล้อมรอบ 1-2 ด้าน

อัลกอริทึม	Hidden layer sizes	การวัดประสิทธิภาพ				
		Accuracy	Precision	Recall	F-measure	AUC
Neural Networks	1	99.23	59.26	9.88	16.93	0.649
	2	99.24	64.29	11.04	18.85	0.655
	3	99.24	66.67	12.12	20.51	0.660
	4	99.24	65.52	11.59	19.69	0.658
	5	99.25	70.00	12.65	21.43	0.663
	6	99.26	73.33	13.25	22.45	0.666
	7	99.25	72.41	12.65	21.54	0.663
	8	99.25	68.97	12.12	20.62	0.660
Bagging+ Neural Networks	1	99.24	58.62	10.56	17.89	0.653
	2	99.24	62.07	11.11	18.85	0.655
	3	99.24	65.52	11.59	19.69	0.658
	4	99.24	63.33	11.59	19.59	0.658
	5	99.25	66.67	12.20	20.62	0.661
	6	99.26	70.97	13.33	22.45	0.666
	7	99.25	70.97	13.25	22.34	0.666
	8	99.24	66.67	12.12	20.51	0.660
AdaBoost + Neural Networks	1	99.24	60.71	10.56	17.99	0.653
	2	99.24	64.29	11.04	18.85	0.655
	3	99.24	62.07	11.11	18.85	0.655
	4	99.25	63.33	11.73	19.79	0.658
	5	99.25	66.67	12.27	20.73	0.661
	6	99.24	63.33	11.66	19.69	0.658
	7	99.25	64.52	12.35	20.73	0.661
	8	99.25	70.00	12.73	21.54	0.663

ตารางภาคผนวกที่ 10 ผลการปรับค่า Hidden layer sizes ที่เหมาะสมสำหรับอัลกอริทึม Neural Networks สำหรับระยะเวลาพยากรณ์ 2 ชั่วโมง กลุ่มข้อมูลที่ 5 สนามบินที่มีภูเขาล้อมรอบ 3-4 ด้าน

อัลกอริทึม	Hidden layer sizes	การวัดประสิทธิภาพ				
		Accuracy	Precision	Recall	F-measure	AUC
Neural Networks	1	99.10	86.54	15.70	26.57	0.678
	2	99.11	87.18	15.83	26.80	0.679
	3	99.10	87.82	15.89	26.92	0.679
	4	99.11	87.34	16.03	27.09	0.680
	5	99.11	87.97	16.13	27.25	0.681
	6	99.11	88.54	16.14	27.31	0.681
	7	99.11	88.05	16.24	27.42	0.681
	8	99.11	88.68	16.34	27.59	0.682
Bagging+ Neural Networks	1	99.10	86.54	15.72	26.60	0.678
	2	99.10	87.18	15.80	26.75	0.679
	3	99.10	87.82	15.89	26.92	0.679
	4	99.11	87.34	16.03	27.09	0.680
	5	99.11	87.97	16.13	27.25	0.681
	6	99.11	88.61	16.24	27.45	0.681
	7	99.11	89.24	16.32	27.59	0.681
	8	99.11	89.31	16.44	27.76	0.682
AdaBoost + Neural Networks	1	99.11	87.18	15.83	26.80	0.679
	2	99.11	87.82	15.93	26.97	0.680
	3	99.11	87.26	15.95	26.97	0.680
	4	99.11	88.54	16.13	27.28	0.681
	5	99.11	87.97	16.14	27.28	0.681
	6	99.11	88.61	16.22	27.42	0.681
	7	99.11	88.05	16.24	27.42	0.681
	8	99.11	88.68	16.34	27.59	0.682

ประวัติผู้วิจัย

ชื่อ	นางสาวชญาณิชฐ์ เมืองสง
วัน เดือน ปี เกิด	20 กันยายน 2531
สถานที่เกิด	อ.เมือง จ.ตาก
ประวัติการศึกษา	วิศวกรรมศาสตรบัณฑิต (วิศวกรรมสิ่งแวดล้อม) มหาวิทยาลัยเกษตรศาสตร์
สถานที่ทำงาน	ส่วนอากาศการบินตาก ศูนย์อุตุนิยมวิทยาภาคเหนือ กรมอุตุนิยมวิทยา
ตำแหน่ง	นักอุตุนิยมวิทยา

