

การปรับปรุงประสิทธิภาพสำหรับการจำแนกประเภทชนิดของกลุ่มเมมโดยใช้วิธีการ
เรียนรู้แบบรวมกลุ่ม



วิทยานิพนธ์เสนอบัณฑิตวิทยาลัย มหาวิทยาลัยนเรศวร เพื่อเป็นส่วนหนึ่งของการศึกษา
หลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์
กุมภาพันธ์ 2560
ลิขสิทธิ์เป็นของมหาวิทยาลัยนเรศวร

วิทยานิพนธ์ เรื่อง "การปรับปรุงประสิทธิภาพสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆโดยใช้
วิธีการเรียนรู้แบบรวมกลุ่ม"

ของ นายปัทม์ อุปการ

ได้รับการพิจารณาให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการสอบวิทยานิพนธ์
(ดร.วินัย วงษ์ไทย)

..... ประธานที่ปรึกษาวิทยานิพนธ์
(ผู้ช่วยศาสตราจารย์ ดร.เกรียงศักดิ์ เตมีย์)

..... กรรมการผู้ทรงคุณวุฒิภายใน
(ผู้ช่วยศาสตราจารย์ ดร.ไกรศักดิ์ เกษร)

..... กรรมการผู้ทรงคุณวุฒิภายนอก
(ดร.พรเทพ ไรจนวสุ)

อนุมัติ

(ดร.ภาณุ พุทธวงศ์)

รองคณบดีฝ่ายบริหารและวางแผน ปฏิบัติราชการแทน

คณบดีบัณฑิตวิทยาลัย

16 ก.พ. 2560

ประกาศคุณูปการ

วิทยานิพนธ์เล่มนี้สำเร็จลงได้ด้วยความกรุณาของ ผู้ช่วยศาสตราจารย์ ดร. เกรียงศักดิ์ เติมีย ที่ได้สละเวลาอันมีค่ามาเป็นทีปรึกษาในการทำวิทยานิพนธ์ครั้งนี้ และคอยให้ความรู้พร้อมทั้งให้คำแนะนำปรึกษาในด้านเทคนิคต่าง ๆ ในศาสตร์ที่เกี่ยวข้องกับการทำวิทยานิพนธ์ ตลอดจนการตรวจทานแก้ไขข้อบกพร่องของการวิทยานิพนธ์ฉบับนี้จนเสร็จสมบูรณ์เป็นอย่างดี

ขอกราบขอบพระคุณคณะท่านอาจารย์กรรมการสอบป้องกันวิทยานิพนธ์อันประกอบไปด้วย ดร.วินัย วงษ์ไทย ประธานกรรมการ ผู้ช่วยศาสตราจารย์ ดร.ไกรศักดิ์ เกษร กรรมการ ผู้ทรงคุณวุฒิภายใน และ ดร.พรเทพ วิจารณ์สุ กรรมาการผู้ทรงคุณวุฒิภายนอก ที่ได้กรุณาให้คำแนะนำตลอดจนแก้ไขข้อบกพร่องของวิทยานิพนธ์ด้วยความเอาใจใส่ จนทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลงได้อย่างสมบูรณ์และทรงคุณค่า

ขอกราบขอบพระคุณ ดร.ฉัตรนันทน์ สอนแก้ว อาจารย์คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏลำปาง ที่ให้ความอนุเคราะห์ข้อมูลสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆจากการเก็บรวบรวมโดยใช้ข้อมูลลักษณะของการแผ่รังสีดวงอาทิตย์ จากหอดูดาวแห่งชาติ ดอยอินทนนท์ จังหวัดเชียงใหม่ และขอขอบคุณ ดร. เอกสิทธิ์ พัทธวงศ์ศักดิ์ดา วิทยากรอบรมการใช้งานซอฟต์แวร์ (Open Source) ทางด้านการทำเหมืองข้อมูล (Data Mining) ผู้ร่วมก่อตั้งห้างหุ้นส่วนสามัญ ดาต้าคิวบ์ (Data Cube) ที่คอยให้คำปรึกษาและให้ความรู้เกี่ยวกับหลักการและทฤษฎีต่าง ๆ ทางศาสตร์วิชาที่เกี่ยวข้องกับการทำเหมืองข้อมูล ตลอดจนขอขอบคุณ พี่ ๆ เพื่อน ๆ ในสาขาวิชาเทคโนโลยีสารสนเทศ วิทยาการคอมพิวเตอร์ วิศวกรรมไฟฟ้า วิศวกรรมการจัดการ และวิทยาศาสตร์สิ่งแวดล้อม รวมถึงทุก ๆ คนที่คอยเป็นกำลังใจในการทำวิทยานิพนธ์ฉบับนี้ให้สำเร็จลงได้ด้วยดี

ผู้วิจัยหวังว่าวิทยานิพนธ์ฉบับนี้จะมีประโยชน์อยู่ไม่น้อย จึงขอมอบส่วนดีทั้งหมดนี้ให้แก่เหล่าคณาจารย์ที่ได้ประสิทธิ์ประสาทวิชาจนทำให้ผลงานวิจัยเป็นประโยชน์ต่อผู้ที่เกี่ยวข้อง และขอแสดงความกตัญญูตเวทีตาคุณ แม่บิดา มารดา และผู้มีพระคุณทุกท่านที่คอยให้ความห่วงใยและให้กำลังใจ ตลอดจนการสนับสนุนในทุก ๆ ด้านจนได้บรรลุผลการเรียนระดับบัณฑิตศึกษาในครั้งนี้ สำหรับข้อบกพร่องต่าง ๆ ที่อาจจะเกิดขึ้นในงานวิจัยนั้นผู้วิจัยขออภัยเพียงผู้เดียว และยินดีที่จะรับฟังคำแนะนำจากทุกท่านที่ได้เข้ามาศึกษา เพื่อเป็นประโยชน์ในการพัฒนางานวิจัยต่อไป

ปัทม์ อูปการ

ชื่อเรื่อง การปรับปรุงประสิทธิภาพสำหรับการจำแนกประเภทชนิดของ
กลุ่มเมฆโดยใช้วิธีการเรียนรู้แบบรวมกลุ่ม

ผู้วิจัย ปัทม์ อูปการ

สถานที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร. เกรียงศักดิ์ เตมีย

ประเภทสารนิพนธ์ วิทยานิพนธ์ วท.ม. สาขาวิชาวิทยาการคอมพิวเตอร์,
มหาวิทยาลัยนเรศวร, 2559

คำสำคัญ การจำแนกประเภทของเมฆ การจำแนกประเภทแบบเดี่ยว
การจำแนกประเภทแบบรวมกลุ่ม การวัดประสิทธิภาพ
แบบจำลอง ค่าที่ใช้วัดความถูกต้องในการจำแนกประเภท

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์ที่จะใช้วิธีการจำแนกประเภทแบบเดี่ยวและการจำแนกประเภทแบบรวมกลุ่ม เพื่อเปรียบเทียบความแตกต่างและการเพิ่มขึ้นของประสิทธิภาพการจำแนกประเภทชนิดของกลุ่มเมฆ ตามงานวิจัยของ ถิรนนท์ สอนแก้ว (รายงานการวิจัย, 2558) โดยที่ การจำแนกประเภทชนิดของกลุ่มเมฆนั้น ได้อาศัยข้อมูลพื้นฐานของค่าการแผ่รังสีจากดวงอาทิตย์ และภาพถ่ายของท้องฟ้าที่หอดูดาวแห่งชาติดอยอินทนนท์ จังหวัดเชียงใหม่ สำหรับวิธีที่ใช้ในการจำแนกประเภทประกอบด้วยวิธีการสร้างต้นไม้ช่วยตัดสินใจ (J48) และโครงข่ายประสาทเทียมหลายชั้น (Artificial Neural Networks : ANN) ซึ่งจัดเป็นการจำแนกประเภทแบบเดี่ยว และใช้วิธีการ Bagging, Boosting และ Random Forest สำหรับการจำแนกประเภทแบบรวมกลุ่ม ผลจากการทดลองและเปรียบเทียบประสิทธิภาพของการจำแนกประเภทแสดงให้เห็นว่า วิธีการจำแนกประเภทแบบรวมกลุ่มสามารถให้ประสิทธิภาพในการจำแนกประเภทชนิดของกลุ่มเมฆดีกว่าการจำแนกประเภทแบบเดี่ยว เนื่องจากมีความหลากหลายของแบบจำลองที่ช่วยในการวิเคราะห์ข้อมูลและร่วมกันตัดสินใจในการจำแนกประเภทชนิดของกลุ่มเมฆ

Title	PERFORMANCE IMPROVEMENT FOR CLOUD CLASSIFICATION USING ENSEMBLE METHOD
Author	Paphat Auppakar
Advisor	Assistant Professor Kreangsak Tamee, Ph.D.
Academic Paper	Thesis Master of Science Degree in Computer Science, Naresuan University, 2016
Keywords	Cloud Classification, Single Classification, Ensemble Classification, Cross-Validation, Accuracy

ABSTRACT

This research aimed to utilize single and ensemble numerical classification methods applied to the work done by Sonkaew et al. (research report, 2015) in classifying cloud types based on solar radiation measurements and all-sky camera images obtained from the Thai National Observatory (TNO) operated by the National Astronomical Research Institute of Thailand (NARIT) situated at Doi Inthanon national park in Chiang Mai. The Decision Tree (J48) and Artificial Neural Network (ANN) were used for the single classification methods, while the Bagging, Boosting, Voting and Random Forest were used for the ensemble classification methods. The experiment result shows that the ensemble classification can provide better performance than a single classification. Since there are a variety of models to help data analyze and decide on the classifying cloud types.

สารบัญ

บทที่

หน้า

1 บทนำ	1
ความเป็นมาของปัญหา	1
จุดมุ่งหมายของการศึกษา	3
ขอบเขตของการวิจัย	3
ประโยชน์ที่คาดว่าจะได้รับ	3
นิยามศัพท์เฉพาะ	4
สมมติฐานของการวิจัย	4
แผนการดำเนินงานวิจัย	5
2 เอกสารและงานวิจัยที่เกี่ยวข้อง	6
ทฤษฎีที่เกี่ยวข้องกับเมฆ	6
ความหมายของเมฆและการเกิดเมฆ	7
การเรียกชื่อเมฆ	9
การแบ่งชนิดของกลุ่มเมฆและลักษณะของเมฆในงานวิจัย	11
เทคนิคการจำแนกประเภทข้อมูล (Data Classification Techniques)	13
ต้นไม้ช่วยตัดสินใจ (Decision Tree)	14
โครงข่ายประสาทเทียม (Artificial Neural Networks)	23
การเรียนรู้แบบรวมกลุ่ม (Ensemble Learning)	36
Bagging Method	37
Boosting Method	38
Random Forest Method	41
มาตรวัดและการทดสอบประสิทธิภาพแบบจำลอง	42
ค่าที่ใช้สำหรับวัดประสิทธิภาพแบบจำลอง	42
การแบ่งข้อมูลเพื่อวัดประสิทธิภาพแบบจำลอง	42
งานวิจัยที่เกี่ยวข้อง	44

สารบัญ (ต่อ)

บทที่	หน้า
3 วิธีดำเนินการวิจัย.....	47
เครื่องมือที่ใช้ในการวิจัย.....	47
กรอบวิธีการดำเนินงานวิจัย.....	48
การเตรียมข้อมูลสำหรับการทดลอง.....	49
การสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยว.....	53
การสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม.....	57
4 ผลการวิจัย.....	67
วิเคราะห์โดยใช้วิธีการจำแนกประเภทแบบเดี่ยว.....	67
ผลลัพธ์จากการวิเคราะห์ข้อมูลโดยใช้อัลกอริทึม J48.....	68
ผลลัพธ์จากการวิเคราะห์ข้อมูลโดยใช้อัลกอริทึม ANN.....	70
วิเคราะห์โดยใช้วิธีการจำแนกประเภทแบบรวมกลุ่ม.....	73
ผลลัพธ์จากการวิเคราะห์ข้อมูลด้วยวิธี Bagging.....	73
ผลลัพธ์จากการวิเคราะห์ข้อมูลด้วยวิธี Boosting.....	79
ผลลัพธ์จากการวิเคราะห์ข้อมูลด้วยวิธี Random Forest.....	86
5 บทสรุป.....	90
สรุปผลการวิจัย.....	90
อภิปรายผล.....	90
ข้อเสนอแนะ.....	94

สารบัญ (ต่อ)

บทที่	หน้า
บรรณานุกรม	96
ภาคผนวก	101
ประวัติผู้วิจัย	129



สารบัญตาราง

ตาราง

หน้า

1 ขั้นตอนการดำเนินงานวิจัย และระยะเวลาการดำเนินงานวิจัย.....	5
2 ตัวอย่างข้อมูลสำหรับการจำแนกประเภทชนิดของเมฆ 2 กลุ่ม.....	17
3 ตัวอย่างข้อมูลสำหรับการจำแนกประเภทชนิดของเมฆ 2 กลุ่ม โดยแปลงค่า คุณลักษณะข้อมูลของเมฆจากตาราง 2 ให้อยู่ในรูปแบบตัวเลข.....	28
4 ตัวอย่างข้อมูลคุณลักษณะของเมฆที่ได้ทำการ Normalize.....	31
5 ข้อมูลกลุ่มตัวอย่างการแบ่งกลุ่มของเมฆและคุณลักษณะของเมฆ.....	49
6 จำนวนข้อมูลตัวอย่างของกลุ่มเมฆ 8 กลุ่ม สำหรับการวิเคราะห์ 3 คุณลักษณะ และ 7 คุณลักษณะ.....	50
7 จำนวนข้อมูลตัวอย่างของกลุ่มเมฆ 3 กลุ่ม สำหรับการวิเคราะห์ 3 คุณลักษณะ และ 7 คุณลักษณะ.....	50
8 ข้อมูลตัวอย่างคุณลักษณะของเมฆ โดยวิเคราะห์ 3 คุณลักษณะ 8 กลุ่ม.....	51
9 ข้อมูลตัวอย่างคุณลักษณะของเมฆ โดยวิเคราะห์ 7 คุณลักษณะ 8 กลุ่ม.....	51
10 ข้อมูลตัวอย่างคุณลักษณะของเมฆ โดยวิเคราะห์ 3 คุณลักษณะ 3 กลุ่ม.....	52
11 ข้อมูลตัวอย่างคุณลักษณะของเมฆ โดยวิเคราะห์ 7 คุณลักษณะ 3 กลุ่ม.....	52
12 พารามิเตอร์ที่กำหนดของอัลกอริทึม J48.....	54
13 พารามิเตอร์ที่กำหนดของอัลกอริทึม ANN.....	56
14 พารามิเตอร์ที่กำหนดของวิธี Bagging โดยใช้ J48 เป็นอัลกอริทึมพื้นฐาน.....	58
15 พารามิเตอร์ที่กำหนดของวิธี Bagging โดยใช้ ANN เป็นอัลกอริทึมพื้นฐาน.....	60
16 พารามิเตอร์ที่กำหนดของวิธี Boosting โดยใช้ J48 เป็นอัลกอริทึมพื้นฐาน.....	62
17 พารามิเตอร์ที่กำหนดของวิธี Boosting โดยใช้ ANN เป็นอัลกอริทึมพื้นฐาน.....	64
18 พารามิเตอร์ที่กำหนดของวิธี Random Forest.....	66
19 การกำหนดค่าพารามิเตอร์ที่ใช้ในการทดลองสำหรับ J48.....	68
20 ค่า Confusion Matrix โดยใช้ J48 วิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 1.....	68
21 ค่า Confusion Matrix โดยใช้ J48 วิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 2.....	69
22 ค่า Confusion Matrix โดยใช้ J48 วิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 3.....	69

สารบัญตาราง (ต่อ)

ตาราง

หน้า

23	ค่า Confusion Matrix โดยใช้ J48 วิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 4.....	69
24	การกำหนดค่าพารามิเตอร์ที่ใช้ในการทดลองสำหรับ ANN.....	71
25	ค่า Confusion Matrix โดยใช้ ANN วิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 1.....	71
26	ค่า Confusion Matrix โดยใช้ ANN วิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 2.....	72
27	ค่า Confusion Matrix โดยใช้ ANN วิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 3.....	72
28	ค่า Confusion Matrix โดยใช้ ANN วิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 4.....	72
29	การกำหนดค่าพารามิเตอร์ที่ใช้ในการทดลองด้วยวิธี Bagging โดยใช้ J48 เป็นพื้นฐานการเรียนรู้.....	74
30	ค่า Confusion Matrix วิธีการแบบ Bagging โดยใช้ J48 เป็นพื้นฐานการเรียนรู้ ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 1.....	74
31	ค่า Confusion Matrix วิธีการแบบ Bagging โดยใช้ J48 เป็นพื้นฐานการเรียนรู้ ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 2.....	75
32	ค่า Confusion Matrix วิธีการแบบ Bagging โดยใช้ J48 เป็นพื้นฐานการเรียนรู้ ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 3.....	75
33	ค่า Confusion Matrix วิธีการแบบ Bagging โดยใช้ J48 เป็นพื้นฐานการเรียนรู้ ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 4.....	76
34	การกำหนดค่าพารามิเตอร์ที่ใช้ในการทดลองด้วยวิธี Bagging โดยใช้ ANN เป็นพื้นฐานการเรียนรู้.....	76
35	ค่า Confusion Matrix วิธีการแบบ Bagging โดยใช้ ANN เป็นพื้นฐานการเรียนรู้ ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 1.....	77
36	ค่า Confusion Matrix วิธีการแบบ Bagging โดยใช้ ANN เป็นพื้นฐานการเรียนรู้ ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 2.....	77
37	ค่า Confusion Matrix วิธีการแบบ Bagging โดยใช้ ANN เป็นพื้นฐานการเรียนรู้ ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 3.....	78

สารบัญตาราง (ต่อ)

ตาราง

หน้า

38	ค่า Confusion Matrix วิธีการแบบ Bagging โดยใช้ ANN เป็นพื้นฐานการเรียนรู้ ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 4.....	78
39	การกำหนดค่าพารามิเตอร์ที่ใช้ในการทดลองด้วยวิธี Boosting โดยใช้ J48 เป็นพื้นฐานการเรียนรู้.....	80
40	ค่า Confusion Matrix วิธีการแบบ Boosting โดยใช้ J48 เป็นพื้นฐานการเรียนรู้ ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 1.....	80
41	ค่า Confusion Matrix วิธีการแบบ Boosting โดยใช้ J48 เป็นพื้นฐานการเรียนรู้ ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 2.....	81
42	ค่า Confusion Matrix วิธีการแบบ Boosting โดยใช้ J48 เป็นพื้นฐานการเรียนรู้ ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 3.....	81
43	ค่า Confusion Matrix วิธีการแบบ Boosting โดยใช้ J48 เป็นพื้นฐานการเรียนรู้ ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 4.....	82
44	การกำหนดค่าพารามิเตอร์ที่ใช้ในการทดลองด้วยวิธี Boosting โดยใช้ ANN เป็นพื้นฐานการเรียนรู้.....	82
45	ค่า Confusion Matrix วิธีการแบบ Boosting โดยใช้ ANN เป็นพื้นฐานการเรียนรู้ ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 1.....	83
46	ค่า Confusion Matrix วิธีการแบบ Boosting โดยใช้ ANN เป็นพื้นฐานการเรียนรู้ ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 2.....	84
47	ค่า Confusion Matrix วิธีการแบบ Boosting โดยใช้ ANN เป็นพื้นฐานการเรียนรู้ ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 3.....	84
48	ค่า Confusion Matrix วิธีการแบบ Boosting โดยใช้ ANN เป็นพื้นฐานการเรียนรู้ ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 4.....	85
49	การกำหนดค่าพารามิเตอร์ที่ใช้ในการทดลองด้วยวิธี Random Forest.....	86
50	ค่า Confusion Matrix วิธีการแบบ Random Forest สำหรับการวิเคราะห์ ข้อมูลกลุ่มตัวอย่างที่ 1.....	87

สารบัญตาราง (ต่อ)

ตาราง

หน้า

51	ค่า Confusion Matrix วิธีการแบบ Random Forest สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 2.....	87
52	ค่า Confusion Matrix วิธีการแบบ Random Forest สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 3.....	87
53	ค่า Confusion Matrix วิธีการแบบ Random Forest สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 4.....	88
54	ผลการทดลองสำหรับการวัดประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆทั้งหมด.....	91
55	การทดสอบค่า T-test โดยใช้อัลกอริทึม J48 เพื่อเปรียบเทียบประสิทธิภาพกับการจำแนกประเภทแบบรวมกลุ่ม.....	94
56	การทดสอบค่า T-test โดยใช้อัลกอริทึม ANN เพื่อเปรียบเทียบประสิทธิภาพกับการจำแนกประเภทแบบรวมกลุ่ม.....	94
57	ตัวดำเนินการที่ใช้สำหรับการทดลองในงานวิจัย.....	108

สารบัญภาพ

ภาพ

หน้า

1	แสดงการแบ่งชนิดของกลุ่มเมฆโดยวัดจากความสูงของพื้นโลก	7
2	แสดงการเกิดเมฆจากการยกตัวของอากาศ	8
3	แสดงการแบ่งชนิดของเมฆพื้นฐาน 10 ชนิด ตามกรมอุตุนิยมวิทยาโลก (WMO)	11
4	แสดงโครงสร้างของต้นไม้ช่วยตัดคลื่นใจ	14
5	แสดงค่า Entropy ของเหตุการณ์ที่จะเกิดขึ้น	15
6	แสดงเหตุการณ์ที่จะเกิดเมฆฝนและไม่ใช่เมฆฝนของแอตทริบิวต์สีของเมฆ	18
7	แสดงเหตุการณ์ที่จะเกิดเมฆฝนและไม่ใช่เมฆฝนของแอตทริบิวต์ความหนาแน่น	18
8	แสดงการแบ่งช่วงของข้อมูลที่เหมาะสมสำหรับแอตทริบิวต์ความสูงจากพื้นโลก	20
9	แสดงการเลือกแอตทริบิวต์เริ่มต้นของต้นไม้ช่วยตัดคลื่นใจ	22
10	แสดงโครงสร้างต้นไม้ช่วยตัดคลื่นใจที่เสร็จสมบูรณ์	23
11	แสดงโครงสร้างของเซลล์ประสาทในสมองของมนุษย์	24
12	แสดงโครงสร้างของสถาปัตยกรรมโครงข่ายประสาทเทียม	24
13	แสดงโครงสร้างสถาปัตยกรรมโครงข่ายประสาทเทียมแบบ MLP	25
14	แสดงเซลล์ประสาทในสถาปัตยกรรมโครงข่ายประสาทเทียม	26
15	แสดงการสุ่มค่าน้ำหนักเริ่มต้นและค่าเบี่ยงเบนในโครงข่ายประสาทเทียม	29
16	แสดงการนำข้อมูลเข้าในชั้น Input Layer ที่ผ่านกระบวนการ Normalization	30
17	แสดงผลลัพธ์ที่ได้จากกระบวนการ Feed-Forward	32
18	แสดงผลลัพธ์ของการปรับปรุงค่าน้ำหนักและค่าเบี่ยงเบนใหม่	33
19	แสดงการนำโครงข่ายประสาทเทียมมาประยุกต์ใช้ในการจำแนกประเภท	34
20	แสดงการเกิดปัญหา Local Trap	36
21	แสดงโครงสร้างพื้นฐานของการเรียนรู้แบบรวมกลุ่ม	37
22	แสดงโครงสร้างวิธีการทำงานแบบ Bagging	38
23	แสดงโครงสร้างวิธีการทำงานแบบ Boosting	39
24	แสดงโครงสร้างวิธีการทำงานแบบ Random Forest	41

สารบัญภาพ (ต่อ)

ภาพ	หน้า
25 แสดงการแบ่งข้อมูลเพื่อวัดประสิทธิภาพแบบจำลอง	43
26 แสดงกรอบวิธีการดำเนินงานวิจัยสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆ	48
27 แสดงขั้นตอนการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยวโดยใช้ J48	53
28 แสดงขั้นตอนการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยวโดยใช้ ANN.....	55
29 แสดงขั้นตอนการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่มด้วยวิธี Bagging โดยใช้ J48 เป็นอัลกอริทึมพื้นฐานการเรียนรู้	57
30 แสดงขั้นตอนการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่มด้วยวิธี Bagging โดยใช้ ANN เป็นอัลกอริทึมพื้นฐานการเรียนรู้	59
31 แสดงขั้นตอนการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่มด้วยวิธี Boosting โดยใช้ J48 เป็นอัลกอริทึมพื้นฐานการเรียนรู้.....	61
32 แสดงขั้นตอนการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่มด้วยวิธี Boosting โดยใช้ ANN เป็นอัลกอริทึมพื้นฐานการเรียนรู้.....	63
33 แสดงขั้นตอนการสร้างแบบจำลองแบบรวมกลุ่มด้วยวิธี Random Forest.....	65
34 แสดงประสิทธิภาพความถูกต้องสำหรับการจำแนกประเภทแบบเดี่ยวโดยใช้ J48	70
35 แสดงประสิทธิภาพความถูกต้องสำหรับการจำแนกประเภทแบบเดี่ยวโดยใช้ ANN	73
36 แสดงประสิทธิภาพความถูกต้องสำหรับการจำแนกประเภทแบบรวมกลุ่มวิธี Bagging	79
37 แสดงประสิทธิภาพความถูกต้องสำหรับการจำแนกประเภทแบบรวมกลุ่มวิธี Boosting....	85
38 แสดงประสิทธิภาพความถูกต้องสำหรับการจำแนกประเภทแบบรวมกลุ่ม วิธี Random Forest	89
39 แสดงผลการเปรียบเทียบประสิทธิภาพความถูกต้องสำหรับการจำแนกประเภท ชนิดของกลุ่มเมฆ.....	92
40 แสดงเว็บไซต์สำหรับดาวน์โหลดโปรแกรม Rapidminer Studio	102
41 แสดงขั้นตอนระหว่างรอการติดตั้งโปรแกรม Rapidminer Studio	102
42 แสดงขั้นตอนเมื่อทำการติดตั้งโปรแกรม Rapidminer Studio เสร็จสมบูรณ์	103

สารบัญภาพ (ต่อ)

ภาพ	หน้า
43 แสดงหน้าเริ่มต้นของโปรแกรม.....	103
44 แสดงหน้าเริ่มต้นการทำงานของโปรแกรม.....	104
45 แสดงขั้นตอนการติดตั้ง Extensions Packet สำหรับ Rapidminer	105
46 แสดงขั้นตอนการติดตั้ง Weka Extension Packages	105
47 แสดงขั้นตอนการรระหว่างติดตั้งโปรแกรม Weka Extension Packages.....	106
48 แสดงการติดตั้งโปรแกรม Weka Extension Packages ที่เสร็จสมบูรณ์.....	106
49 แสดงหน้าเริ่มต้นของโปรแกรมที่ได้ทำการติดตั้ง Weka Extension.....	107
50 แสดงการใช้งานตัวดำเนินการสำหรับข้อมูลนำเข้า	109
51 แสดงการเลือกไฟล์ข้อมูลสำหรับข้อมูลนำเข้า	110
52 แสดงการตั้งค่าไฟล์ข้อมูลสำหรับข้อมูลนำเข้า	111
53 แสดงตัวอย่างข้อมูลนำเข้า.....	111
54 แสดงการกำหนดชนิดของข้อมูลตัวอย่างสำหรับข้อมูลนำเข้า	112
55 แสดงการใช้งานตัวดำเนินการ X-Validation	112
56 แสดงหน้าจอโปรเซสย่อยของตัวดำเนินการ X- Validation	113
57 แสดงการเรียนรู้ข้อมูลด้วยตัวดำเนินการ J48	114
58 แสดงการเรียนรู้ข้อมูลด้วยตัวดำเนินการ ANN	114
59 แสดงโปรเซสย่อยของตัวดำเนินการ Bagging โดยใช้ J48 เป็นอัลกอริทึมพื้นฐาน สำหรับการเรียนรู้	114
60 แสดงโปรเซสย่อยของตัวดำเนินการ Bagging โดยใช้ ANN เป็นอัลกอริทึมพื้นฐาน สำหรับการเรียนรู้	115
61 แสดงโปรเซสย่อยของตัวดำเนินการ Boosting โดยใช้ J48 เป็นอัลกอริทึมพื้นฐาน สำหรับการเรียนรู้	115
62 แสดงโปรเซสย่อยของตัวดำเนินการ Boosting โดยใช้ ANN เป็นอัลกอริทึมพื้นฐาน สำหรับการเรียนรู้	116

สารบัญภาพ (ต่อ)

ภาพ	หน้า
63 แสดงการเรียนรู้ข้อมูลด้วยตัวดำเนินการ Random Forest.....	116
64 แสดงตัวดำเนินการสำหรับทดสอบแบบจำลองในส่วนของ Testing	117
65 แสดงตัวอย่างโปรเซสที่สมบูรณ์สำหรับการจำแนกประเภทแบบเดี่ยว	117
66 แสดงตัวอย่างโปรเซสที่สมบูรณ์สำหรับการจำแนกประเภทแบบรวมกลุ่ม	118
67 แสดงผลลัพธ์การทดสอบค่า T-test โดยเปรียบเทียบระหว่าง J48 กับ Bagging (J48)	119
68 แสดงผลลัพธ์การทดสอบค่า T-test โดยเปรียบเทียบระหว่าง J48 กับ Bagging (ANN)	120
69 แสดงผลลัพธ์การทดสอบค่า T-test โดยเปรียบเทียบระหว่าง J48 กับ Boosting (J48)	121
70 แสดงผลลัพธ์การทดสอบค่า T-test โดยเปรียบเทียบระหว่าง J48 กับ Boosting (ANN)	122
71 แสดงผลลัพธ์การทดสอบค่า T-test โดยเปรียบเทียบระหว่าง J48 กับ Random Forest.....	123
72 แสดงผลลัพธ์การทดสอบค่า T-test โดยเปรียบเทียบระหว่าง ANN กับ Bagging (J48)	124
73 แสดงผลลัพธ์การทดสอบค่า T-test โดยเปรียบเทียบระหว่าง ANN กับ Bagging (ANN)	125
74 แสดงผลลัพธ์การทดสอบค่า T-test โดยเปรียบเทียบระหว่าง ANN กับ Boosting (J48)	126
75 แสดงผลลัพธ์การทดสอบค่า T-test โดยเปรียบเทียบระหว่าง ANN กับ Boosting (ANN)	127
76 แสดงผลลัพธ์การทดสอบค่า T-test โดยเปรียบเทียบระหว่าง ANN กับ Random Forest.....	128

บทที่ 1

บทนำ

ความเป็นมาของปัญหา

การจำแนกประเภทชนิดของเมฆ (Cloud Classification) กำลังได้รับความนิยมเพิ่มมากขึ้นเรื่อย ๆ ในช่วงเวลาที่ผ่านมา ทั้งนี้เนื่องจากข้อมูลและสารสนเทศที่ได้รับจากเมฆแต่ละชนิดนั้นมีอิทธิพลเป็นอย่างมากต่อการพยากรณ์อากาศ ดังนั้นหากมีการพัฒนาและปรับปรุงขั้นตอนวิธีการจำแนกประเภทชนิดของเมฆให้มีประสิทธิภาพทั้งทางด้านความถูกต้องและแม่นยำแล้ว ย่อมเกิดประโยชน์ต่อนักวิจัยที่จะนำไปประยุกต์ใช้ในการพยากรณ์สภาพอากาศ ภูมิอากาศ อุทกวิทยา การบริหารจัดการทรัพยากรน้ำ และภาคการเกษตร (Mahrooghy, M., et al., 2011) โดยเฉพาะอย่างยิ่งในหน่วยงานทางด้านอุตุนิยมวิทยา สำหรับวิธีการที่ใช้ในการจำแนกประเภทชนิดของเมฆส่วนใหญ่อาศัยหลักเกณฑ์การจำแนกตามองค์การอุตุนิยมวิทยาโลก ซึ่งได้แบ่งชนิดของเมฆออกเป็น 10 ชนิด (World Meteorological Organization., 1975) ได้แก่ เมฆเซอร์รัส (Ci) เมฆเซอโรคิวมูลัส (Cc) เมฆเซอโรสเตรตัส (Cs) เมฆอัลโตคิวมูลัส (Ac) เมฆอัลโตสเตรตัส (As) เมฆนิมโบสเตรตัส (Ns) เมฆสเตรโตคิวมูลัส (Sc) เมฆสเตรตัส (St) เมฆคิวมูลัส (Cu) และ เมฆคิวโมโลนิมบัส (Cb) จากการศึกษาและทบทวนวรรณกรรมในช่วง 5 ปีที่ผ่านมาพบว่าในต่างประเทศมีนักวิจัยหลายคนได้พยายามค้นหาขั้นตอนวิธีการต่าง ๆ สำหรับการจำแนกประเภทชนิดของเมฆ ดังเช่นการจำแนกเมฆโดยการวิเคราะห์คุณลักษณะของเมฆด้วย Texture และใช้อัลกอริทึม kNN สำหรับการเรียนรู้ในการจำแนกชนิดของเมฆ (Heinle, A., et al., 2010) ต่อมาได้มีงานวิจัยของ Pallavi, V. P. and Vaithyanathan, V., (2013) ทำการวิเคราะห์คุณลักษณะของเมฆจากการแบ่งสีของภาพถ่าย โดยใช้อัลกอริทึมโครงข่ายประสาทเทียม (Artificial Neural Networks : ANN) เปรียบเทียบกับ Genetic Algorithm สำหรับขั้นตอนการจำแนกชนิดของเมฆ และในปี 2015 ได้มีการใช้ภาพถ่ายบนท้องฟ้า (Whole-Sky Ground-Based Images) โดยใช้ Cloud Coverage (CC) เป็นค่าที่ใช้แบ่งสัดส่วนของเมฆ สำหรับขั้นตอนในการจำแนกชนิดของเมฆใช้วิธี ANN และ Support Vector Machine (Taravat, A., et al., 2015) ในปีเดียวกันนี้มีงานวิจัยที่ได้คิดค้นวิธีการใหม่ ๆ เพื่อปรับปรุงการตรวจจับชนิดของเมฆจากระบบภาพถ่ายทั่วท้องฟ้าให้มีประสิทธิภาพเพิ่มมากขึ้น (Chauvin, R., et al., 2015)

สำหรับในประเทศไทยได้มีการศึกษาค้นคว้าด้านนี้ในหลายประเด็นบนพื้นฐานของความรู้ทางด้านอุตุนิยมนิยามวิทยา โดยล่าสุดตามรายงานการวิจัยของ ถิรนนท์ สอนแก้ว (2558) ได้ทำการศึกษาและวิจัยเกี่ยวกับการจำแนกประเภทชนิดของกลุ่มเมฆที่เก็บรวบรวมโดยใช้ข้อมูลลักษณะของการแผ่รังสีดวงอาทิตย์ จากหอดูดาวแห่งชาติดอยอินทนนท์ จังหวัดเชียงใหม่ซึ่งนำมาใช้เป็นตัวแปรหนึ่งที่บ่งบอกถึงคุณลักษณะของเมฆแต่ละชนิด โดยในงานวิจัยได้ทำการแบ่งชนิดของเมฆออกเป็น 2 ลักษณะคือ ลักษณะที่ 1 แบ่งชนิดของเมฆออกเป็น 8 กลุ่ม และลักษณะที่ 2 แบ่งชนิดของเมฆออกเป็น 3 กลุ่ม สำหรับขั้นตอนในการจำแนกประเภทชนิดของกลุ่มเมฆใช้หลักวิธีการวิเคราะห์ทางสถิติซึ่งเป็นวิธีแบบเชิงเส้น (Linear) เปรียบเทียบกับการวิเคราะห์โดยใช้โครงข่ายประสาทเทียมหลายชั้น (Multi-Layer Perceptron : MLP) ซึ่งเป็นวิธีแบบไม่เชิงเส้น (Non-linear) จากผลการทดลองพบว่าการใช้โครงข่ายประสาทเทียมหลายชั้นสามารถจำแนกประเภทชนิดของกลุ่มเมฆได้ดีกว่าการวิเคราะห์ทางหลักสถิติ โดยภาพรวมประสิทธิภาพของความถูกต้องในการจำแนกสูงที่สุดอยู่ที่ 85% สำหรับการวิเคราะห์ชนิดของกลุ่มเมฆ 3 กลุ่มโดยใช้ 7 คุณลักษณะแต่อย่างไรก็ตามประสิทธิภาพการจำแนกประเภทชนิดของเมฆยังไม่ดีพอเท่าที่ควร เนื่องจากวิธีการของโครงข่ายประสาทเทียมหลายชั้นจัดเป็นกระบวนการในการจำแนกประเภทแบบเดี่ยว (Single Classification) จากการศึกษาค้นคว้าทฤษฎีต่าง ๆ เพิ่มเติมพบว่า มีหลายวิธีการสำหรับการเพิ่มประสิทธิภาพให้กับการจำแนกประเภทของข้อมูล โดยเฉพาะอย่างยิ่งวิธีการจำแนกประเภทแบบผสมผสาน (Multiple Classification) หรือเรียกว่าวิธีการเรียนรู้แบบรวมกลุ่ม (Ensemble Learning) ซึ่งมีนักวิจัยหลายคนได้นำวิธีการนี้ไปประยุกต์ใช้กับข้อมูลประเภทอื่น ๆ แล้วพบว่าสามารถทำให้ประสิทธิภาพในการจำแนกประเภทของข้อมูลเพิ่มมากขึ้น เช่น การจำแนกชนิดของโรคมะเร็งเต้านมจากเครื่องเอกซเรย์เต้านมระบบดิจิทัลโดยใช้วิธีการของโครงข่ายประสาทเทียมพบว่าให้ประสิทธิภาพความถูกต้องอยู่ที่ 90% และเมื่อนำเอาวิธีการเรียนรู้แบบรวมกลุ่มมาใช้สามารถทำให้ประสิทธิภาพเพิ่มขึ้นถึง 99% (Pour, S.G., et al., 2012) ในปีเดียวกันมีการทดลองการจำแนกประเภทของ Spam Email โดยใช้อัลกอริทึม C4.5, Naive Bayes, SVM และ kNN ในการสร้างแบบจำลองแบบเดี่ยวเปรียบเทียบกับการสร้างแบบจำลองการเรียนรู้แบบรวมกลุ่มโดยใช้อัลกอริทึม C4.5 เป็นแบบจำลองพื้นฐานการเรียนรู้พบว่า การเรียนรู้แบบรวมกลุ่มให้ประสิทธิภาพความถูกต้องดีกว่าการใช้แบบจำลองการจำแนกประเภทแบบเดี่ยว (SHI, L., et al., 2012) ในปีต่อมามีงานวิจัยของ Tiwari A. and Prakash A. (2014) ได้ทำการทดลองโดยใช้ชุดข้อมูลของ SONAR Dataset และได้เปรียบเทียบระหว่างการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยวกับการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม 3 วิธีโดยใช้อัลกอริทึม J48 เป็น

แบบจำลองพื้นฐานการเรียนรู้ พบว่าแบบจำลองที่ใช้วิธีการเรียนรู้แบบรวมกลุ่มให้ประสิทธิภาพการจำแนกได้สูงสุดอยู่ที่ 86.07% และในปี 2015 ได้มีการนำวิธีการเรียนรู้แบบรวมกลุ่มไปประยุกต์ใช้กับข้อมูลทางด้านเวชชีวศาสตร์ (Meshram, S.B. and Shinde, S.M., 2015)

ดังนั้นในงานวิจัยนี้จึงมีแนวคิดที่จะทำการปรับปรุงประสิทธิภาพในการจำแนกประเภทชนิดของกลุ่มเมฆเพื่อแก้ไขปัญหาประสิทธิภาพในด้านของความถูกต้องและแม่นยำสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆที่ต่ำ โดยได้นำเอาหลักทฤษฎีวิธีการเรียนรู้แบบรวมกลุ่มมาใช้เพื่อช่วยเพิ่มประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆให้สูงขึ้น ซึ่งได้ทำการศึกษาหลักเกณฑ์การแบ่งกลุ่มชนิดของเมฆและการเลือกคุณลักษณะที่นำมาวิเคราะห์ในลักษณะเดียวกันกับรายงานการวิจัยของ ธีรนนท์ สอนแก้ว (2558)

จุดมุ่งหมายของการศึกษา

จากแนวคิดในงานวิจัย ผู้วิจัยได้ตั้งจุดมุ่งหมายของงานวิจัยไว้ดังนี้

1. ปรับปรุงประสิทธิภาพสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆโดยใช้เทคนิควิธีการเรียนรู้แบบรวมกลุ่ม

ขอบเขตของการวิจัย

ในการศึกษาค้นคว้าข้อมูล ผู้วิจัยได้กำหนดขอบเขตของการวิจัยไว้ดังนี้

1. ชุดข้อมูล (Dataset) จำนวน 4 ชุด สำหรับใช้ในการทดลอง ซึ่งเป็นของมูลจริงที่ได้จากสถาบันวิจัยดาราศาสตร์แห่งชาติ โดยเก็บรวบรวมข้อมูลภาพถ่ายบนท้องฟ้าจากหอดูดาวแห่งชาติ ดอยอินทนนท์ จังหวัดเชียงใหม่
2. ใช้อัลกอริทึม J48 และ ANN สำหรับขั้นตอนการจำแนกประเภทแบบเดี่ยว รวมถึงใช้เป็นอัลกอริทึมพื้นฐานการเรียนรู้ร่วมกับวิธีการเรียนรู้แบบรวมกลุ่มแบบ Bagging และ Boosting
3. ใช้วิธีการเรียนรู้แบบรวมกลุ่ม 3 วิธี ได้แก่ Bagging, Boosting, และ Random Forest

ประโยชน์ที่คาดว่าจะได้รับ

1. ได้แบบจำลองที่มีประสิทธิภาพความถูกต้องและแม่นยำสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆเพิ่มสูงขึ้น

2. ใช้เป็นทฤษฎีกรณีศึกษาสำหรับพัฒนาตัวแบบในการจำแนกประเภท (Classifier) เพื่อเพิ่มประสิทธิภาพความถูกต้องและแม่นยำให้การจำแนกประเภทชนิดของกลุ่มเมฆ และนำข้อมูลสารสนเทศที่ได้ไปประยุกต์ใช้ ก่อให้เกิดประโยชน์ต่อหน่วยงานด้านอุตุนิยมวิทยาในสาขาต่าง ๆ

3. สามารถนำข้อมูลสารสนเทศที่ได้จากการจำแนกประเภทชนิดของกลุ่มเมฆไปประยุกต์ใช้ในด้านศาสตร์ที่เกี่ยวข้องกับการพยากรณ์อากาศได้

นิยามศัพท์เฉพาะ

Cloud Classification = กระบวนการในการจำแนกประเภทของเมฆ

Ensemble Learning = การเรียนรู้แบบรวมกลุ่ม

Data Classification Techniques = เทคนิคการจำแนกประเภทข้อมูล

Cross Validation = วิธีการแบ่งข้อมูลเพื่อใช้วัดประสิทธิภาพแบบจำลอง

Accuracy = ค่าที่ใช้สำหรับวัดประสิทธิภาพความถูกต้องของแบบจำลอง

สมมติฐานของการวิจัย

ในการศึกษางานวิจัย ผู้วิจัยได้ทำการตั้งสมมติฐานไว้ดังนี้ เมื่อนำเทคนิควิธีการเรียนรู้แบบรวมกลุ่มมาใช้ในการจำแนกประเภทชนิดของกลุ่มเมฆ จะสามารถทำให้ประสิทธิภาพความถูกต้องและแม่นยำในการจำแนกประเภทชนิดของกลุ่มเมฆเพิ่มสูงขึ้น

แผนการดำเนินงานวิจัย

ตาราง 1 ขั้นตอนการดำเนินงานวิจัย และระยะเวลาการดำเนินงานวิจัย

	พ.ศ. 2557	พ.ศ. 2558	พ.ศ. 2559
ขั้นตอนการดำเนินงาน	พหุศาสตร์	พหุศาสตร์	พหุศาสตร์
	พหุสาขา	พหุสาขา	พหุสาขา
	พหุศาสตร์	พหุศาสตร์	พหุศาสตร์
	พหุสาขา	พหุสาขา	พหุสาขา
	พหุศาสตร์	พหุศาสตร์	พหุศาสตร์
	พหุสาขา	พหุสาขา	พหุสาขา
	พหุศาสตร์	พหุศาสตร์	พหุศาสตร์
	พหุสาขา	พหุสาขา	พหุสาขา
	พหุศาสตร์	พหุศาสตร์	พหุศาสตร์
	พหุสาขา	พหุสาขา	พหุสาขา
1. ศึกษาเอกสารงานวิจัยที่เกี่ยวข้อง			
2. ศึกษาเทคโนโลยี เครื่องมือสนับสนุน			
3. ศึกษากลุ่มตัวอย่างข้อมูล			
4. วิเคราะห์และออกแบบแบบจำลอง			
5. ทดสอบประสิทธิภาพของแบบจำลอง			
6. นำแบบจำลองใช้งานกับข้อมูลใหม่			
7. สรุปและอภิปรายผลการทดลอง			
8. เขียนบทความวิจัย			
9. จัดทำเอกสารวิทยานิพนธ์			
10. สรุปผลการวิจัย			

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การศึกษางานวิจัยในครั้งนี้ ผู้วิจัยได้นำเสนอแนวความคิด ทฤษฎี และผลงานวิจัยที่เกี่ยวข้องซึ่งมีรายละเอียดดังนี้

1. ทฤษฎีที่เกี่ยวข้องกับเมฆ
 - 1.1 ความหมายของเมฆและการเกิดเมฆ
 - 1.2 การเรียกชื่อเมฆ
 - 1.3 การแบ่งชนิดของกลุ่มเมฆและลักษณะของเมฆในงานวิจัย
2. เทคนิคการจำแนกประเภทข้อมูล (Data Classification Techniques)
 - 2.1 ต้นไม้ช่วยตัดสินใจ (Decision Tree)
 - 2.2 โครงข่ายประสาทเทียม (Artificial Neural Networks)
3. การเรียนรู้แบบรวมกลุ่ม (Ensemble Learning)
 - 3.1 Bagging Method
 - 3.2 Boosting Method
 - 3.3 Random forest Method
4. มาตรฐานวัดและการทดสอบประสิทธิภาพแบบจำลอง
5. งานวิจัยที่เกี่ยวข้อง

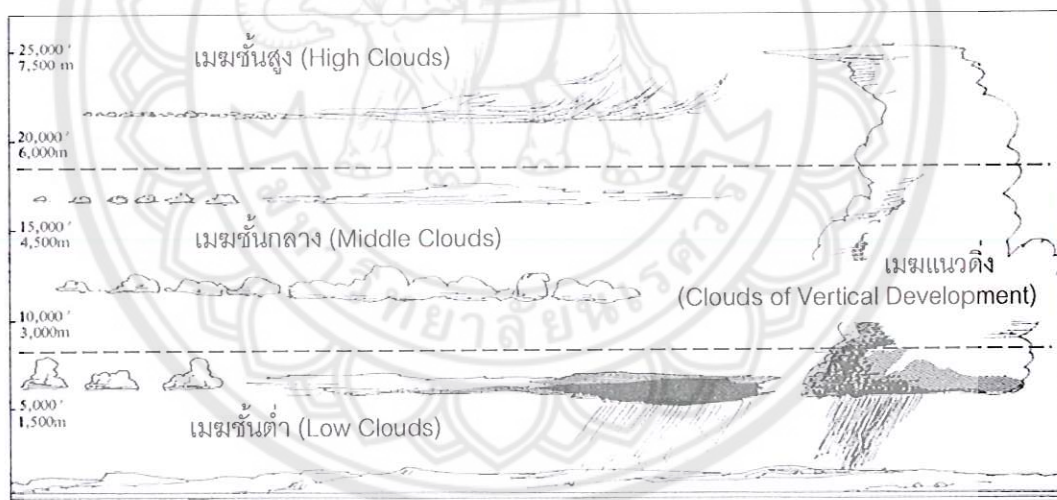
ทฤษฎีที่เกี่ยวข้องกับเมฆ

นักอุตุนิยมวิทยาตั้งแต่สมัยอดีตจนถึงปัจจุบันใช้การตรวจสอบสภาพอากาศจากการตรวจวัดด้วยเครื่องมือชนิดต่าง ๆ เพื่อการวัดความกดอากาศ อุณหภูมิ และปริมาณน้ำฝน นอกจากนี้แล้วการตรวจสอบสภาพอากาศยังรวมถึงการสังเกตและการบันทึกเมฆชนิดต่าง ๆ เนื่องจากเมฆบางชนิดสามารถบ่งบอกถึงสภาพอากาศได้เป็นอย่างดี เช่น เมฆอัลโตคิวมูลัส (Alto cumulus), เมฆอัลโตสเตรัส (Altostratus) และเมฆนิมโบสเตรตัส (Nimbostratus) เมฆเหล่านี้สามารถบ่งบอกถึงการเกิดพายุฝน หรืออาจมีฝนตกเกิดขึ้นได้ เป็นต้น โดยเมฆสามารถจำแนกได้ด้วยสายตาของคนเราได้เป็น 4 รูปแบบหลักดังนี้ เมฆสเตรตัส (Stratus), เมฆคิวมูลัส (Cumulus),

เมฆซีร์รัส (Cirrus) และเมฆนิมบัส (Nimbus) ดังนั้นการจำแนกประเภทของเมฆชนิดต่าง ๆ จึงมีบทบาทสำคัญมากสำหรับการพยากรณ์อากาศในปัจจุบัน

1. ความหมายของเมฆและการเกิดเมฆ

เมฆ (Cloud) คือกลุ่มของละอองน้ำขนาดเล็กซึ่งเกิดจากการควบแน่นของหยดน้ำในอากาศ แต่ลักษณะของเมฆชั้นสูงซึ่งจะมีอุณหภูมิต่ำกว่าจุดเยือกแข็งและเป็นกลุ่มของผลึกน้ำแข็งขนาดเล็ก เมื่อสะสมไอน้ำจนอิ่มตัวจะก่อให้เกิดหยาดน้ำฟ้าและตกลงมาสู่พื้นโลก หยดน้ำและผลึกน้ำแข็งที่มีพื้นผิว (Surface) นั้นจะสะท้อนแสงทำให้เราสามารถมองเห็นเป็นก้อนสีขาว ในบางครั้งบางมุมตกกระทบของแสงและเงาจากเมฆชั้นบนหรือเมฆชั้นที่อยู่ข้างเคียง ความหนาแน่นของหยดน้ำที่อยู่ในก้อนเมฆนั้นอาจทำให้มองเห็นเป็นเมฆสีเทาได้ ซึ่งในปัจจุบันมีการจำแนกประเภทของเมฆที่พบได้ตามธรรมชาติโดยแบ่งออกเป็น 4 ระดับ ตามความสูงของฐานเมฆจากพื้นโลกและเมฆที่ก่อตัวขึ้นในแนวดิ่ง สำหรับการจำแนกประเภทของเมฆโดยดูจากความสูงของฐานเมฆนั้นมีพื้นฐานมาจากนักอุตุนิยมวิทยาชาวอังกฤษชื่อ Luke Howard ในปี ค.ศ. 1802 แสดงในภาพ 1

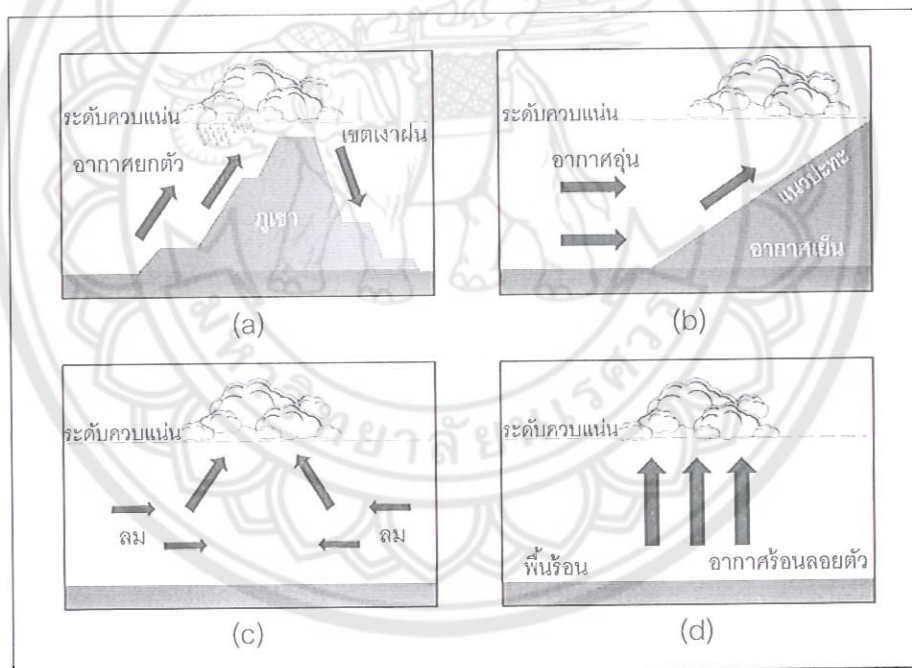


ภาพ 1 แสดงการแบ่งชนิดของกลุ่มเมฆโดยวัดจากความสูงของพื้นโลก

ในปี 1998 มีนักวิจัย Duchon and O' Malley, (1998) ทำการจำแนกประเภทของเมฆ โดยใช้ค่าการแผ่รังสีจากดวงอาทิตย์ที่วัดได้ด้วยเครื่องไพรานอมิเตอร์ มีวัตถุประสงค์เพื่อนำไปประยุกต์ใช้ในการจำแนกประเภทของเมฆโดยไม่จำเป็นต้องอาศัยการสังเกตด้วยตาเปล่า โดยปริมาณรังสีตกกระทบที่วัดได้จากเครื่องไพรานอมิเตอร์จะมีค่าที่แตกต่างกัน ขึ้นอยู่กับเมฆในแต่ละชนิดและมีค่าเปลี่ยนแปลงไปในแต่ละช่วงเวลาของวัน ค่ารังสีที่วัดได้ในทุก ๆ นาฬิกาจะถูกนำมาคำนวณหาค่า

ความคลาดเคลื่อนมาตรฐานและสัดส่วนที่วัดได้ต่อค่ารังสีของท้องฟ้าเพื่อทำการจำแนกประเภทชนิดของเมฆ โดย Nicola Chalmer, (2005) ได้ทำการพัฒนาต่อจากงานวิจัยเดิม เนื่องจากค่าความสอดคล้องของการจำแนกประเภทชนิดของเมฆระหว่างค่ารังสีดวงอาทิตย์ที่วัดได้จากเครื่องไพรานอมิเตอร์กับการจำแนกประเภทของเมฆด้วยตาเปล่ามีความสอดคล้องกันเพียง 45% จึงต้องทำให้ข้อมูลมีความสอดคล้องกันมากขึ้น โดยนำค่ารังสีดวงอาทิตย์ที่วัดได้มาเปรียบเทียบกับค่ารังสีรวม และพัฒนาชุดการทดลองสำหรับการวัดค่ารังสีดวงอาทิตย์ในแต่ละช่วงเวลา โดยได้ทำการแบ่งช่วงสูงสุด – ต่ำสุด ของรังสีดวงอาทิตย์ที่วัดได้ตามประเภทของเมฆ เพื่อให้สามารถจำแนกประเภทของเมฆได้แม่นยำยิ่งขึ้น

เมฆเกิดขึ้นได้จากการยกตัวของอากาศ สำหรับกลไกที่ทำให้เกิดการเคลื่อนตัวของอากาศในแนวดิ่งมี 4 กระบวนการ ดังภาพ 2



ภาพ 2 แสดงการเกิดเมฆจากการยกตัวของอากาศ

1.1 เกิดจากสภาพภูมิประเทศ

เมื่อกระแสลมปะทะภูเขา อากาศถูกบังคับให้ลอยสูงขึ้น (เนื่องจากไม่มีทางออกทางอื่น) จนถึงระดับควบแน่นจึงกลั่นตัวเป็นหยดน้ำ สังเกตจากบริเวณบนยอดเขาสูงมักมีเมฆปกคลุมอยู่ ทำให้บริเวณยอดเขาที่มีความชุ่มชื้นและอุดมไปด้วยป่าไม้ และเมื่อกระแสลมพัดผ่านยอด

เขา อากาศแห้งที่สูญเสียไอน้ำจะจมตัวลงจนมีอุณหภูมิสูงขึ้น ภูมิอากาศบริเวณหลังภูเขาจึงเป็นเขตที่แห้งแล้ง เรียกว่า "เขตเงาฝน" (Rain Shadow) แสดงในภาพ 2 (a)

1.2 เกิดจากแนวปะทะของอากาศ

อากาศร้อนมีความหนาแน่นต่ำกว่าอากาศเย็น เมื่ออากาศร้อนปะทะกับอากาศเย็นอากาศร้อนจะดันตัวขึ้น อุณหภูมิจะลดต่ำลงจนถึงระดับควบแน่นทำให้เกิดเมฆและกลายเป็นฝน แสดงในภาพ 2 (b)

1.3 เกิดจากอากาศบีบตัว

เมื่อกระแสลมพัดมาปะทะกันอากาศจะยกตัวขึ้น ทำให้อุณหภูมิลดต่ำลงจนเกิดอากาศอิ่มตัว ไอน้ำในอากาศควบแน่นเป็นหยดน้ำและกลายเป็นเมฆ แสดงในภาพ 2 (c)

1.4 เกิดจากการพาความร้อน

เมื่อพื้นผิวของโลกมีความแตกต่างกันจึงมีการดูดกลืนและคายความร้อนไม่เท่ากัน จึงมีผลทำให้กลุ่มอากาศที่ลอยอยู่มีอุณหภูมิแตกต่างกันไปด้วยโดยเฉพาะอย่างยิ่งในช่วงฤดูร้อน เช่น กลุ่มอากาศที่ลอยอยู่เหนือพื้นคอนกรีตจะมีอุณหภูมิสูงกว่ากลุ่มอากาศที่ลอยอยู่เหนือพื้นหญ้า กลุ่มอากาศที่มีอุณหภูมิสูงมีความหนาแน่นน้อยกว่าอากาศในบริเวณโดยรอบจะทำให้เกิดเป็นกลุ่มก้อนเมฆขึ้นได้ แสดงในภาพ 2 (d)

2. การเรียกชื่อเมฆ

การเรียกชื่อเมฆแบ่งออกเป็น 2 รูปร่างลักษณะ คือ เมฆก้อนและเมฆแผ่น นอกจากนี้จะเรียกชื่อเมฆตามลักษณะการก่อตัวของเมฆแล้ว เมฆยังถูกเรียกตามระดับชั้นความสูงของฐานเมฆจากพื้นโลกได้ดังนี้

2.1 เมฆชั้นต่ำ (Low Clouds)

เกิดขึ้นที่ระดับต่ำกว่า 2 กิโลเมตร สำหรับการเรียกชื่อจะเติมคำว่า "สเตรโต" ซึ่งแปลว่า "เป็นชั้น" เรียกชื่อเมฆที่มีลักษณะเป็นแผ่นว่า "เมฆสเตรตัส" (Stratus) และเมฆที่มีลักษณะเป็นก้อนว่า "เมฆคิวมูลัส" (Cumulus) ส่วนเมฆเป็นแผ่นที่ลอยชิดติดกัน เรียกว่า "เมฆสเตรโตคิวมูลัส" (Stratocumulus) ในกรณีที่เมฆฝนจะเพิ่มคำว่า "นิมโบ" หรือ "นิมบัส" ซึ่งแปลว่า "ฝน" เข้าไป เช่น เรียกเมฆก้อนที่มีฝนตกว่า "เมฆคิวมูโลนิมบัส" (Cumulonimbus) และเรียกเมฆแผ่นที่มีฝนตกว่า "เมฆนิมโบสเตรตัส" (Nimbostratus) ซึ่งชนิดของกลุ่มเมฆในชั้นต่ำจะประกอบไปด้วยเมฆดังต่อไปนี้

2.1.1 เมฆสเตรตัส (Stratus) เมฆแผ่นบางลอยสูงเหนือพื้นไม่มากนัก เช่น เมฆที่ลอยปกคลุมยอดเขามักเกิดขึ้นตอนเช้าหรือหลังฝนตก บางครั้งลอยต่ำปกคลุมพื้นดิน เรียกว่า "หมอก"

2.1.2 เมฆสเตรโตคิวมูลัส (Stratocumulus) เมฆก้อนลอยติดกันเป็นแพ ไม่มีรูปทรงที่ชัดเจน มีช่องว่างระหว่างก้อนเพียงเล็กน้อยมักเกิดขึ้นเวลาที่อากาศไม่ดี

2.1.3 เมฆนิมโบสเตรตัส (Nimbostratus) เมฆแผ่นสีเทาเกิดขึ้นเวลาที่อากาศมีเสถียรภาพทำให้เกิดฝนพรำ ๆ ฝนผ่านหรือฝนตกแดดออกไม่มีพายุฝนฟ้าคะนอง ฟ้าร้องฟ้าผ่า มักปรากฏให้เห็นสายฝนตกลงมาจากฐานเมฆ

2.2 เมฆชั้นกลาง (Middle Clouds)

เกิดขึ้นที่ระดับสูง 2 - 6 กิโลเมตร สำหรับการเรียกชื่อจะเดิมคำว่า "อัลโต" ซึ่งแปลว่า "ชั้นกลาง" ไว้ข้างหน้า เช่น เรียกเมฆก้อนชั้นกลางว่า "เมฆอัลโตคิวมูลัส" (Alto cumulus) และเรียกเมฆแผ่นชั้นกลางว่า "เมฆอัลโตสเตรตัส" (Alto stratus) ซึ่งชนิดของกลุ่มเมฆในชั้นกลางจะประกอบไปด้วยเมฆดังต่อไปนี้

2.2.1 เมฆอัลโตคิวมูลัส (Alto cumulus) เมฆก้อนสีขาวมีลักษณะคล้ายฝูงแกะลอยเป็นแพ มีช่องว่างระหว่างก้อนเล็กน้อย

2.2.2 เมฆอัลโตสเตรตัส (Alto stratus) เมฆแผ่นหนาส่วนมากมักมีสีเทาเนื่องจากบังแสงดวงอาทิตย์ไม่ให้ลอดผ่าน และเกิดขึ้นปกคลุมท้องฟ้าเป็นบริเวณกว้างมาก หรือปกคลุมไปทั่วท้องฟ้าทั้งหมด

2.3 เมฆชั้นสูง (High Clouds)

เกิดขึ้นที่ระดับสูงมากกว่า 6 กิโลเมตร สำหรับการเรียกชื่อจะเดิมคำว่า "เซโร" ซึ่งแปลว่า "ชั้นสูง" ไว้ข้างหน้า เช่น เรียกเมฆก้อนชั้นสูงว่า "เมฆเซโรคิวมูลัส" (Cirro cumulus) เรียกเมฆแผ่นชั้นสูงว่า "เมฆเซโรสเตรตัส" (Cirro stratus) และเรียกเมฆชั้นสูงที่มีรูปร่างเหมือนขนนกว่า "เมฆเซอรัส" (Cirrus) ซึ่งชนิดของกลุ่มเมฆในชั้นสูงจะประกอบไปด้วยเมฆดังต่อไปนี้

2.3.1 เมฆเซโรคิวมูลัส (Cirro cumulus) เมฆสีขาว เป็นผลึกน้ำแข็งมีลักษณะเป็นริ้วคลื่น มักเกิดขึ้นปกคลุมท้องฟ้าบริเวณกว้าง

2.3.2 เมฆเซโรสเตรตัส (Cirro stratus) เมฆแผ่นบางสีขาวเป็นผลึกน้ำแข็งปกคลุมท้องฟ้าเป็นบริเวณกว้างและโปร่งแสง บางครั้งเกิดการหักเหของแสงทำให้เกิดดวงอาทิตย์ทรงกลดและดวงจันทร์ทรงกลดเป็นรูปร่างกลมสี่คล้ายรุ้ง

2.3.3 เมฆเซอรัส (Cirrus) เมฆริ้วสีขาวคล้ายขนนกเป็นผลึกน้ำแข็ง ส่วนใหญ่จะเกิดขึ้นในวันที่มีอากาศดีและท้องฟ้าเป็นสีฟ้าเข้ม

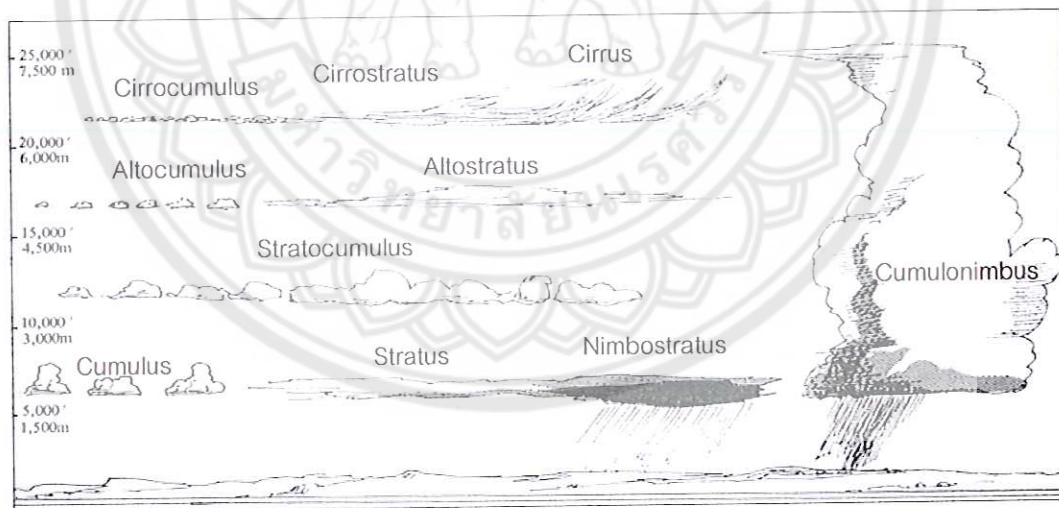
2.4 เมฆก่อตัวในแนวตั้ง (Clouds of Vertical Development)

2.4.1 เมฆคิวมูลัส (Cumulus) เมฆก้อนปุกปุย สีขาวเป็นรูปกะหล่ำ ก่อตัวในแนวตั้ง เกิดขึ้นจากอากาศไม่มีความเสถียร ฐานเมฆเป็นสีเทาเนื่องจากมีความหนามากพอที่จะบังแสงจากดวงอาทิตย์ ทำให้เกิดเงาและมักปรากฏให้เห็นเวลาอากาศดี

2.4.2 เมฆคิวมูโลนิมบัส (Cumulonimbus) เมฆก่อตัวในแนวตั้ง พัฒนามาจากเมฆคิวมูลัส มีขนาดใหญ่มากปกคลุมพื้นที่บริเวณกว้าง ทำให้เกิดพายุฝนฟ้าคะนอง หากมีกระแสลมชั้นบนพัดแรงจะทำให้ยอดเมฆรูปกะหล่ำกลายเป็นรูปทั่งตีเหล็กและต่อยอดออกมาเป็น เมฆเซอโรสเตรตัสหรือเมฆเซอรัส

3. การแบ่งชนิดของกลุ่มเมฆและลักษณะของเมฆในงานวิจัย

การแบ่งชนิดของเมฆสามารถทำได้หลายวิธีขึ้นอยู่กับเกณฑ์และจุดประสงค์ที่ใช้ในการแบ่งของผู้วิจัย ซึ่งตามหลักสากลนั้นทางองค์การกรมอุตุนิยมวิทยาโลก (World Meteorological Organization : WMO) ได้แบ่งชนิดของเมฆพื้นฐานออกเป็น 10 ชนิด แสดงในภาพ 3



ภาพ 3 แสดงการแบ่งชนิดของเมฆพื้นฐาน 10 ชนิด ตามกรมอุตุนิยมวิทยาโลก (WMO)

ดังนั้นวิทยานิพนธ์นี้จะใช้เกณฑ์การแบ่งตามรายงานการวิจัยของ ถิรนนท์ สอนแก้ว (2558) โดยทำการแบ่งชนิดของเมฆออกเป็น 8 กลุ่ม และ 3 กลุ่ม ซึ่งได้ทำการวิเคราะห์คุณลักษณะของเมฆ 2 แบบ คือ 3 คุณลักษณะ และ 7 คุณลักษณะ อธิบายรายละเอียดได้ดังต่อไปนี้

3.1 การแบ่งชนิดของกลุ่มเมฆ

3.1.1 แบ่งชนิดของเมฆออกเป็น 8 กลุ่ม

- 1) กลุ่มเมฆที่เป็นก้อนและอยู่ระดับต่ำ (Cu)
- 2) กลุ่มเมฆที่เป็นริ้ว ๆ และบาง (Ci, Cs)
- 3) กลุ่มเมฆที่เป็นก้อนและอยู่ระดับกลางขึ้นไป (Cc, Ac)
- 4) ท้องฟ้าโปร่งไม่มีเมฆ (Clear sky)
- 5) กลุ่มเมฆที่เป็นก้อนและแผ่อยู่รวมกัน (Sc)
- 6) กลุ่มเมฆที่เป็นแผ่น (St, As)
- 7) กลุ่มเมฆฝน (Cb, Ns)
- 8) กลุ่มที่ไม่ใช่เมฆ เช่น หมอก (Fog) หรือฝุ่นละออง

3.1.2 แบ่งชนิดของเมฆออกเป็น 3 กลุ่ม (แบ่งตามสภาพท้องฟ้า)

- 1) ท้องฟ้าประกอบไปด้วยเมฆที่มีลักษณะเป็นก้อน (Cu, Cc, Ac)
- 2) ท้องฟ้าที่เป็นท้องฟ้าโปร่ง (Clear sky, Ci and Cs)
- 3) ท้องฟ้าที่ประกอบไปด้วยเมฆที่มีลักษณะเป็นแผ่นและกระจาย

ค่อนข้างทั่วบนท้องฟ้า (Sc, St, As, Cb, Ns and Fog)

3.2 การแบ่งคุณลักษณะของเมฆ

การวิเคราะห์คุณลักษณะของเมฆโดยทั่วไปสามารถทำได้หลายวิธีไม่ว่าจะเป็น การสังเกตจากรูปทรง การสังเกตจากความสูงของฐานเมฆ การวัดความหนาแน่นของมวลอากาศ การวัดค่าจากการแผ่รังสีดวงอาทิตย์ การดึงคุณลักษณะของเมฆจากภาพถ่ายดาวเทียม รวมถึง การตรวจวัดค่าคุณลักษณะของเมฆด้วยเครื่องมือตรวจวัดสภาพอากาศในรูปแบบต่าง ๆ นอกเหนือจากนี้ยังมีวิธีการอื่น ๆ อีกมาก ดังนั้นวิทยานิพนธ์นี้เลือกใช้คุณลักษณะของเมฆตาม รายงานการวิจัยของ ถิรนนท์ สอนแก้ว (2558) ซึ่งคุณลักษณะเมฆที่นำมาใช้ในการวิเคราะห์ สามารถแบ่งออกเป็น 2 แบบ ได้แก่ แบบที่ 1 ทำการวิเคราะห์คุณลักษณะของเมฆ 3 คุณลักษณะ ได้แก่ ค่าอัตราส่วนรังสีจากดวงอาทิตย์ (AV) ค่าส่วนเบี่ยงเบนมาตรฐานของดวงอาทิตย์ (SD) และ ค่าอัตราส่วนของเมฆต่อท้องฟ้าทั้งหมด (CC) แบบที่ 2 ทำการวิเคราะห์คุณลักษณะของเมฆ 7 คุณลักษณะโดยเพิ่มตัวแปรที่เกี่ยวข้องกับทางอุตุนิยมวิทยาอีก 4 คุณลักษณะ ได้แก่ อุณหภูมิ (T) ความกดอากาศ (P) ความชื้นสัมพัทธ์ (RH) และความเร็วลม (WS)

สำหรับการหาค่า AV และ SD เป็นการวัดค่ารังสีจากดวงอาทิตย์ที่แผ่มายัง พื้นผิวโลกผ่านก้อนเมฆ เนื่องจากก้อนเมฆแต่ละชนิดมีระดับความสูงที่วัดได้จากพื้นผิวโลกและการ

กระจายตัวที่แตกต่างกันของก้อนเมฆรวมถึงความหนาแน่นที่แตกต่างกัน ดังนั้นรังสีที่ส่งผ่านก้อนเมฆแต่ละชนิดมายังพื้นผิวโลกจึงมีค่าไม่เท่ากัน สามารถแสดงในสมการ 1 และ 2 ตามลำดับ สำหรับในส่วนของคุณค่า CC เป็นการหาค่าอัตราส่วนของเมฆจากภาพถ่ายทั่วท้องฟ้าซึ่งเป็นการดึงคุณลักษณะของเมฆจากภาพถ่าย (Image Extraction) แสดงในสมการ 3

$$\text{อัตราส่วนรังสีจากดวงอาทิตย์ (AV)} = \frac{\text{ค่าเฉลี่ยรังสีจากดวงอาทิตย์ (A)}}{\text{ค่าเฉลี่ยรังสีจากดวงอาทิตย์ที่คำนวณได้ (M)}} \quad (1)$$

$$\text{ส่วนเบี่ยงเบนมาตรฐานของดวงอาทิตย์ (SD)} = \frac{\text{ค่าความคลาดเคลื่อนมาตรฐาน (S)}}{\text{ค่าเฉลี่ยรังสีจากดวงอาทิตย์ที่คำนวณได้ (M)}} \quad (2)$$

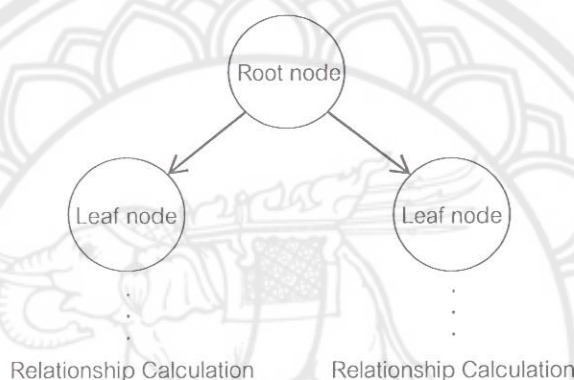
$$\text{อัตราส่วนของเมฆต่อท้องฟ้าทั้งหมด (CC)} = \frac{\text{ภาพในส่วนที่เป็นเมฆ (N_{\text{Cloud}})}}{\text{ภาพในส่วนที่เป็นท้องฟ้า (N_{\text{all_sky}})}} \quad (3)$$

เทคนิคการจำแนกประเภทข้อมูล (Data Classification Techniques)

การจำแนกประเภทข้อมูลจัดเป็นการเรียนรู้แบบมีผู้สอน (Supervised Learning) ซึ่งเป็นเทคนิคหนึ่งของการเรียนรู้ของเครื่อง (Machine Learning) และถูกจัดให้อยู่ในศาสตร์ของการทำเหมืองข้อมูล (Data Mining) การจำแนกประเภทข้อมูลเป็นวิธีที่นำเอาข้อมูลตัวอย่าง (Instance) ที่มีอยู่ในอดีตมาสอน (Training Data) เพื่อเรียนรู้รูปแบบที่จะเกิดขึ้นจากข้อมูลด้วยสมการหรือการวิเคราะห์ทางอัลกอริทึมที่ใช้ในการจำแนกประเภท และสร้างเป็นแบบจำลองสำหรับหาคำตอบให้กับข้อมูลใหม่ (Unseen Data) ตัวอย่างของอัลกอริทึมที่ใช้ในการจำแนกประเภทข้อมูลมีอยู่หลายวิธีการ เช่น Decision Tree, Naive Bayes, K-Nearest Neighbors, Artificial Neural Networks และ Support Vector Machine ฯลฯ เป็นต้น สำหรับวิทยานิพนธ์นี้ใช้อัลกอริทึม J48 ซึ่งเป็นหนึ่งในอัลกอริทึมของต้นไม้ช่วยตัดสินใจ และ โครงข่ายประสาทเทียมหลายชั้นในการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยว รวมไปถึงการใช้ร่วมกับวิธีการเรียนรู้แบบผสมผสานหรือเรียกอีกอย่างหนึ่งว่าวิธีการเรียนรู้แบบรวมกลุ่ม

1. ต้นไม้ช่วยตัดสินใจ (Decision Tree)

การสร้างต้นไม้ช่วยตัดสินใจถูกพัฒนาขึ้นโดย Quinlan, J. R., (1986) ซึ่งเป็นอัลกอริทึมที่นิยมนำมาใช้กันอย่างแพร่หลายในงานทางด้านการจำแนกประเภทข้อมูล สำหรับการสร้างต้นไม้ช่วยตัดสินใจนั้นจะขึ้นอยู่กับ การคำนวณค่าความสัมพันธ์ระหว่างแต่ละแอตทริบิวต์กับแอตทริบิวต์คำตอบ (Class Label) โดยแอตทริบิวต์ที่มีความสัมพันธ์กับแอตทริบิวต์คำตอบมากที่สุดจะถูกเลือกให้เป็นโหนดราก (Kaur, D., et al., 2015) ซึ่งชนิดของต้นไม้ช่วยตัดสินใจมีหลายชนิด สำหรับชนิดที่นิยมนำมาใช้ในการจำแนกประเภทข้อมูลได้แก่ ID3 และ C4.5 หรือ J48



ภาพ 4 แสดงโครงสร้างของต้นไม้ช่วยตัดสินใจ

จากภาพ 4 ต้นไม้ช่วยตัดสินใจประกอบไปด้วยโหนดเริ่มต้นของเหตุการณ์ (Root Node) และโหนดใบ (Leaf Node) โดยทั้ง 2 โหนดเชื่อมต่อกันด้วยกิ่ง (Branch) วิธีการของต้นไม้ช่วยตัดสินใจแบบ ID3 จะใช้ค่า Entropy เป็นตัววัดความแตกต่างของข้อมูล ดังนั้นถ้าข้อมูลมีความแตกต่างกันมาก Entropy จะมีค่าสูง แต่ถ้าข้อมูลมีความแตกต่างกันน้อย Entropy จะมีค่าต่ำ สำหรับการหาค่า Entropy นั้นคำนวณได้จากสมการ 4 (Han, J., Kamber, M. and Pei, J., 2012)

$$\text{Info}(D) = -\sum_{i=1}^m P_i \log_2 P_i \quad (4)$$

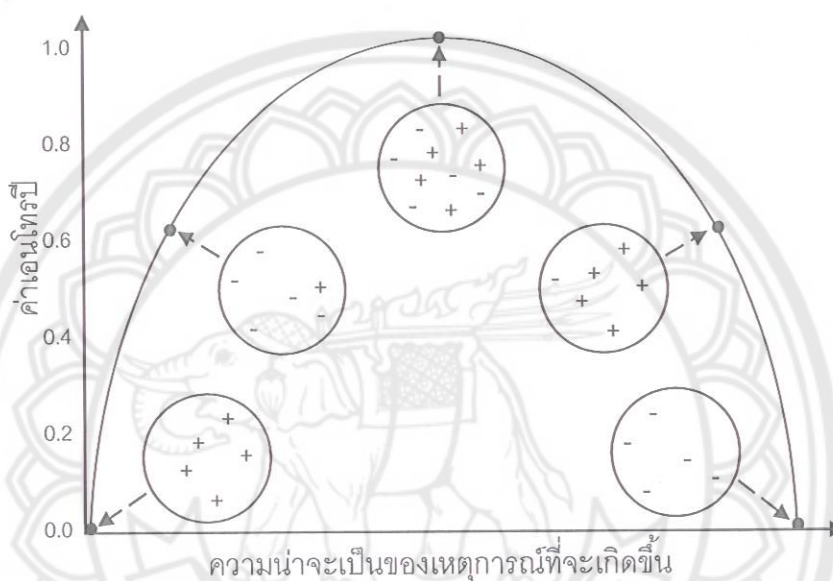
โดยที่

Info(D) คือ ค่า Entropy ก่อนการแบ่งข้อมูลตัวอย่างในแอตทริบิวต์คำตอบ

P คือ ความน่าจะเป็นที่เหตุการณ์ i จะเกิดขึ้นจากเหตุการณ์ทั้งหมด m เหตุการณ์

i คือ เหตุการณ์ที่จะเกิดขึ้น

m คือ จำนวนของเหตุการณ์



ภาพ 5 แสดงค่า Entropy ของเหตุการณ์ที่จะเกิดขึ้น

จากภาพ 5 แสดงตัวอย่างความน่าจะเป็นที่เหตุการณ์บวกและเหตุการณ์ลบจะเกิดขึ้น ถ้าเหตุการณ์ทั้งสองมีความน่าจะเป็นของเหตุการณ์บวกและเหตุการณ์ลบเท่ากันจะทำให้ค่า Entropy เท่ากับ 1 แต่หากมีเหตุการณ์ใดเพียงเหตุการณ์เดียวเกิดขึ้นจากข้อมูลตัวอย่างทั้งหมดจะทำให้ค่า Entropy ที่ได้มีค่าเท่ากับ 0 (ถ้าแอตทริบิวต์ไหนมีค่า Entropy สูงจะทำให้เป็นโหนดที่มีความเหมาะสมต่อการแบ่งข้อมูล) สำหรับอัลกอริทึม ID3 จะใช้ค่า Information Gain (IG) เป็นตัววัดความสัมพันธ์ระหว่างแต่ละแอตทริบิวต์กับแอตทริบิวต์คำตอบหรือเป็นค่าคาดหวังผลต่างของค่า Entropy ก่อนที่จะมีการแบ่งข้อมูลและหลังการแบ่งข้อมูลในแอตทริบิวต์นั้น ถ้าค่า IG ของแอตทริบิวต์นั้นมีความสัมพันธ์กับแอตทริบิวต์คำตอบมากที่สุดเมื่อนำมาเปรียบเทียบกับในแต่ละแอตทริบิวต์ในการเรียนรู้ชุดของข้อมูล แอตทริบิวต์นั้นจะถูกเลือกให้เป็นโหนดเริ่มต้นของเหตุการณ์ ซึ่งการหาค่า IG สามารถคำนวณได้จากสมการ 5 โดยที่ ค่า $\text{Info}_A(D)$ หาได้จากสมการ 6

$$\text{Gain}(A) = \text{Info}(D) - \text{Info}_A(D) \quad (5)$$

$$\text{Info}_A(D) = \sum_{j=1}^V \frac{|D_j|}{|D|} \times \text{Info}(D_j) \quad (6)$$

โดยที่

$\text{Info}_A(D)$ คือ ค่า Entropy หลังการแบ่งข้อมูลตัวอย่างในแอตทริบิวต์ A

$|D_j|$ คือ ข้อมูลตัวอย่างในแอตทริบิวต์ A ที่มีค่าเป็น j

$|D|$ คือ ข้อมูลตัวอย่างทั้งหมด

$\text{Info}(D_j)$ คือ ค่า Entropy ของแอตทริบิวต์ A ที่มีค่าเป็น j

A คือ แอตทริบิวต์

j คือ ค่าในแอตทริบิวต์

V คือ จำนวนของค่าในแอตทริบิวต์

อัลกอริทึม ID3 สามารถใช้งานได้เฉพาะกับข้อมูลที่มีลักษณะเป็นค่าแบบไม่ต่อเนื่อง (Discrete Data) ทำให้มีการพัฒนาอัลกอริทึม J48 เพื่อให้สามารถนำมาใช้กับการประมวลผลในชุดของข้อมูลที่มีลักษณะเป็นค่าแบบต่อเนื่อง (Continuous Data) และยังช่วยเพิ่มประสิทธิภาพการเรียนรู้ได้ดีขึ้น เนื่องจาก J48 มีการพัฒนาเทคนิคการตัดกิ่งของต้นไม้ (Pruning) ทำให้โครงสร้างของต้นไม้มีขนาดเล็กลงและประมวลผลได้เร็วขึ้นโดยไม่ทำให้ประสิทธิภาพลดลง จึงสามารถช่วยหลีกเลี่ยงปัญหา Overfitting ลงได้ ซึ่งเป็นปัญหาที่พบได้บ่อยกับอัลกอริทึมทางด้านการเรียนรู้ของเครื่อง

อัลกอริทึม J48 ถูกพัฒนาต่อจากอัลกอริทึม ID3 โดย Quinlan, J. R., (1993) ซึ่งโครงสร้างของอัลกอริทึม J48 จะใช้ค่าอัตราส่วน Gain Ratio (GR) เป็นตัวพิจารณาความสัมพันธ์ระหว่างแต่ละแอตทริบิวต์กับแอตทริบิวต์คำตอบแทนค่า IG ที่ถูกใช้ในอัลกอริทึม ID3 (Han, J., Kamber, M. and Pei, J., 2012) สำหรับการหาค่า GR สามารถคำนวณได้จากสมการ 7 โดยที่ค่า Split Information (SI) เป็นค่าที่แสดงถึงผลรวมของอัตราส่วนของข้อมูลตัวอย่างในแอตทริบิวต์ A ที่มีค่าเป็น j ต่อจำนวนข้อมูลตัวอย่างทั้งหมดหรือเป็นค่าที่แสดงถึงอำนาจในการจำแนกของกลุ่มตัวอย่างข้อมูลตามค่าที่เป็นไปได้ในแอตทริบิวต์ A สามารถคำนวณได้จากสมการ 8

$$\text{Gain Ratio}(A) = \frac{\text{Gain}(A)}{\text{SplitInfo}_A(D)} \quad (7)$$

$$\text{SplitInfo}_A(D) = - \sum_{j=1}^V \frac{|D_j|}{|D|} \times \log_2 \frac{|D_j|}{|D|} \quad (8)$$

โดยที่

$|D_j|$ คือ ข้อมูลตัวอย่างในแอตทริบิวต์ A ที่มีค่าเป็น j

$|D|$ คือ ข้อมูลตัวอย่างทั้งหมด

A คือ แอตทริบิวต์

j คือ ค่าในแอตทริบิวต์

V คือ จำนวนของค่าในแอตทริบิวต์

ตัวอย่างการสร้างแบบจำลองต้นไม้ช่วยตัดสินใจ J48

ตาราง 2 ตัวอย่างข้อมูลสำหรับการจำแนกประเภทชนิดของเมฆ 2 กลุ่ม

สีของเมฆ	ความหนาแน่นของเมฆ	ความสูงจากพื้นโลก (เมตร)	กลุ่มของเมฆ
ขาวโปร่งแสง	ต่ำ	3,000	เมฆฝน
ขาวโปร่งแสง	ต่ำ	7,000	ไม่ใช่เมฆฝน
เทา	สูง	2,300	เมฆฝน
เทา	สูง	1,000	เมฆฝน
ขาวโปร่งแสง	ต่ำ	5,500	ไม่ใช่เมฆฝน
ขาวโปร่งแสง	สูง	1,500	เมฆฝน
ขาวขุ่น	สูง	2,500	ไม่ใช่เมฆฝน
ขาวโปร่งแสง	สูง	1,500	เมฆฝน
เทา	ต่ำ	3,500	เมฆฝน
ขาวขุ่น	สูง	2,500	ไม่ใช่เมฆฝน

ตาราง 2 แสดงตัวอย่างข้อมูลสำหรับการจำแนกประเภทชนิดของเมฆ 2 กลุ่ม ซึ่งมีข้อมูลทั้งหมด 10 ตัวอย่าง 4 แอตทริบิวต์ ได้แก่ แอตทริบิวต์ที่แสดงถึงระดับสีของเมฆ, ความหนาแน่น, ความสูงของเมฆจากพื้นโลก และกลุ่มของเมฆซึ่งเป็นแอตทริบิวต์คำตอบ สามารถแสดงขั้นตอนการสร้างแบบจำลองต้นไม้ช่วยตัดสินใจ J48 ได้ดังตัวอย่างต่อไปนี้

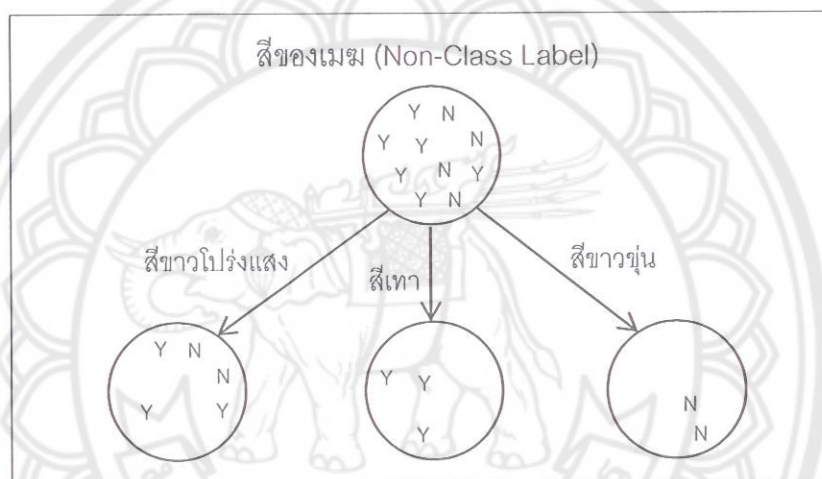
ขั้นตอนที่ 1 คำนวณหาค่า Entropy โดยพิจารณาจากแอตทริบิวต์คำตอบ และค่า Entropy ที่ไม่ใช่แอตทริบิวต์คำตอบ จากสมการ 4 โดยให้ Y แทนเมฆฝน และ N แทนไม่ใช่เมฆฝน

$$P_{\text{เมฆฝน}} = 6/10 = 0.6, P_{\text{ไม่ใช่เมฆฝน}} = 4/10 = 0.4$$

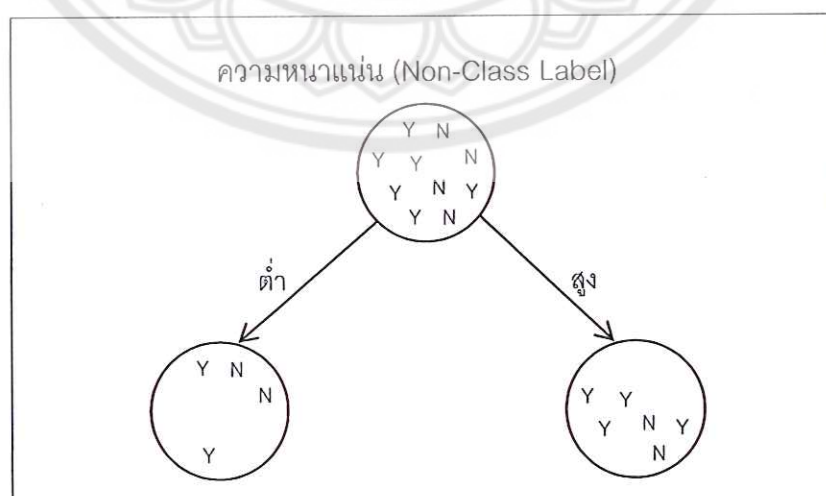
$$\text{Info(กลุ่มของเมฆ)} = - [(0.6)\log_2(0.6) + (0.4)\log_2(0.4)] = 0.97$$

ขั้นตอนที่ 2 คำนวณหาค่า IG ของทุก ๆ แอตทริบิวต์ จากสมการ 5 และ 6

ขั้นตอนที่ 3 คำนวณหาค่า GR ของทุก ๆ แอตทริบิวต์ จากสมการ 7 โดยที่ค่า SI คำนวณได้จากสมการ 8 ซึ่งแอตทริบิวต์ที่มีค่า GR สูงที่สุดจะถูกเลือกให้เป็นโหนดเริ่มต้นของเหตุการณ์



ภาพ 6 แสดงเหตุการณ์ที่จะเกิดเมฆฝนและไม่ใช่เมฆฝนของแอตทริบิวต์สีของเมฆ



ภาพ 7 แสดงเหตุการณ์ที่จะเกิดเมฆฝนและไม่ใช่เมฆฝนของแอตทริบิวต์ความหนาแน่น

จากภาพ 6 และ 7 สามารถคำนวณหาค่า IG, SI และ GR ของแต่ละแอตทริบิวต์ได้ดังต่อไปนี้

พิจารณาแอตทริบิวต์สีของเมฆ

สีขาวยิ่งแสง : $Y=3/5$, $N=2/5$

$$|D_{\text{สีขาวยิ่งแสง}}| / |D| = 5/10$$

$$\text{Info}(D_{\text{สีขาวยิ่งแสง}}) = - [(3/5 \log_2 3/5) + (2/5 \log_2 2/5)] = 0.97$$

สีเทา : $Y=3/3$, $N=0/3$

$$|D_{\text{สีเทา}}| / |D| = 3/10$$

$$\text{Info}(D_{\text{สีเทา}}) = - [(3/3 \log_2 3/3) + (0/3 \log_2 0/3)] = 0$$

สีขาวขุ่น : $Y=0/2$, $N=2/2$

$$|D_{\text{สีขาวขุ่น}}| / |D| = 2/10$$

$$\text{Info}(D_{\text{สีขาวขุ่น}}) = - [(0/2 \log_2 0/2) + (2/2 \log_2 2/2)] = 0$$

$$IG_{\text{สีของเมฆ}} = 0.97 - [5/10(0.97) + 3/10(0) + 2/10(0)] = 0.49$$

$$SI_{\text{สีของเมฆ}} = - [(5/10 \log_2 5/10) + (3/10 \log_2 3/10) + (2/10 \log_2 2/10)] = 1.49$$

$$GR_{\text{สีของเมฆ}} = 0.49/1.49 = 0.33$$

พิจารณาแอตทริบิวต์ความหนาแน่น

ความหนาแน่นต่ำ : $Y=2/4$, $N=2/4$

$$|D_{\text{ความหนาแน่นต่ำ}}| / |D| = 4/10$$

$$\text{Info}(D_{\text{ความหนาแน่นต่ำ}}) = - [(2/4 \log_2 2/4) + (2/4 \log_2 2/4)] = 1$$

ความหนาแน่นสูง : $Y=4/6$, $N=2/6$

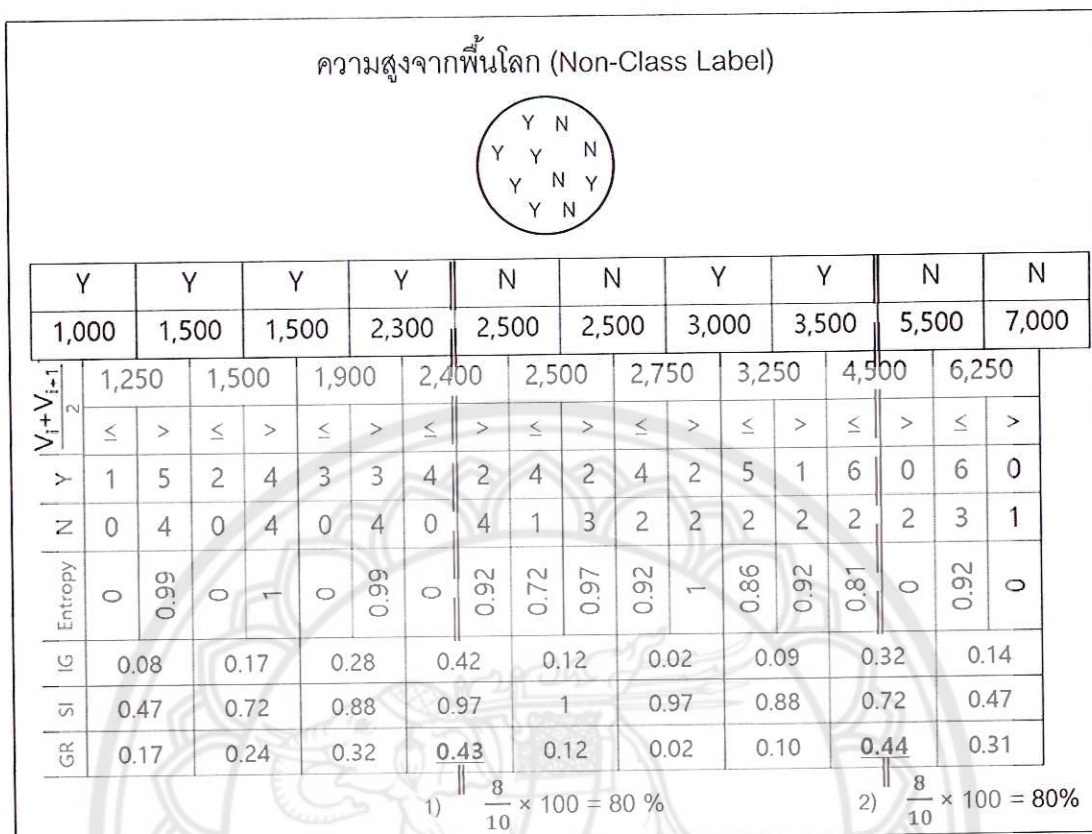
$$|D_{\text{ความหนาแน่นสูง}}| / |D| = 6/10$$

$$\text{Info}(D_{\text{ความหนาแน่นสูง}}) = - [(4/6 \log_2 4/6) + (2/6 \log_2 2/6)] = 0.92$$

$$IG_{\text{ความหนาแน่นของเมฆ}} = 0.97 - [4/10(1) + 6/10(0.92)] = 0.02$$

$$SI_{\text{ความหนาแน่นของเมฆ}} = - [(4/10 \log_2 4/10) + (6/10 \log_2 6/10)] = 0.97$$

$$GR_{\text{ความหนาแน่นของเมฆ}} = 0.02/0.97 = 0.02$$



ภาพ 8 แสดงการแบ่งช่วงของข้อมูลที่เหมาะสมสำหรับแอตทริบิวต์ความสูงจากพื้นโลก

จากภาพ 8 จะเห็นว่าการคำนวณค่า IG, SI และ GR มีความแตกต่างจากแอตทริบิวต์สีของเมฆ และความหนาแน่นของเมฆ เนื่องจากว่าแอตทริบิวต์ความสูงจากพื้นโลกลักษณะข้อมูลเป็นแบบค่าต่อเนื่อง ซึ่งการหาค่า GR ที่ข้อมูลมีลักษณะเป็นแบบค่าต่อเนื่องนั้นต้องทำการเรียงลำดับข้อมูลใหม่ (Sorting) จากนั้นน้อยไปหามาก $\{V_1, V_2, V_3, \dots, V_m\}$ โดยที่ V คือ ลำดับของการเรียงข้อมูล m ตัวอย่างที่ใช้ในการเรียนรู้ และตำแหน่งของการแบ่งช่วงข้อมูลแต่ละตำแหน่งหาได้จาก $(V_i + V_{i+1})/2$ (Kantardzic, M., 2011) สำหรับการหาช่วงที่มีความเหมาะสมที่สุดนั้นจะหาได้จากตำแหน่งของการแบ่งข้อมูลที่มีค่าความถูกต้องสูงสุดในการจำแนกข้อมูล (Roiger, R. J. and Geatz, M. W., 2003) ดังภาพจะพบว่ามีค่าความถูกต้องในการจำแนกประเภทชนิดของเมฆสูงสุดอยู่ 2 ตำแหน่งเท่ากันที่ 80% ได้แก่ ตำแหน่งที่หนึ่ง $(2,300+2,500)/2$ มีช่วงอยู่ระหว่าง $\leq 2,400$ และ $> 2,400$ ตำแหน่งที่สอง $(3,500+5,500)/2$ มีช่วงอยู่ระหว่าง $\leq 4,500$ และ $> 4,500$ ซึ่งทั้งสองตำแหน่งสามารถคำนวณค่า GR ได้ดังต่อไปนี้

พิจารณาช่วง $\leq 2,400$ และ $> 2,400$

$\leq 2,400$: $Y=4/4$, $N=0/4$

$$|D_{\leq 2,400}| / |D| = 4/10$$

$$\text{Info}(D_{\leq 2,400}) = - [(4/4 \log_2 4/4) + (0/4 \log_2 0/4)] = 0$$

$> 2,400$: $Y=2/6$, $N=4/6$

$$|D_{> 2,400}| / |D| = 6/10$$

$$\text{Info}(D_{> 2,400}) = - [(2/6 \log_2 2/6) + (4/6 \log_2 4/6)] = 0.92$$

$$IG_{\leq 2,400 \text{ และ } > 2,400} = 0.97 - [4/10(0) + 6/10(0.92)] = 0.42$$

$$SI_{\leq 2,400 \text{ และ } > 2,400} = - [(4/10 \log_2 4/10) + (6/10 \log_2 6/10)] = 0.97$$

$$GR_{\leq 2,400 \text{ และ } > 2,400} = 0.42/0.97 = 0.43$$

พิจารณาช่วง $\leq 4,500$ และ $> 4,500$

$\leq 4,500$: $Y=6/8$, $N=2/8$

$$|D_{\leq 4,500}| / |D| = 8/10$$

$$\text{Info}(D_{\leq 4,500}) = - [(6/8 \log_2 6/4) + (2/8 \log_2 2/4)] = 0.81$$

$> 4,500$: $Y=0/2$, $N=2/2$

$$|D_{> 4,500}| / |D| = 2/10$$

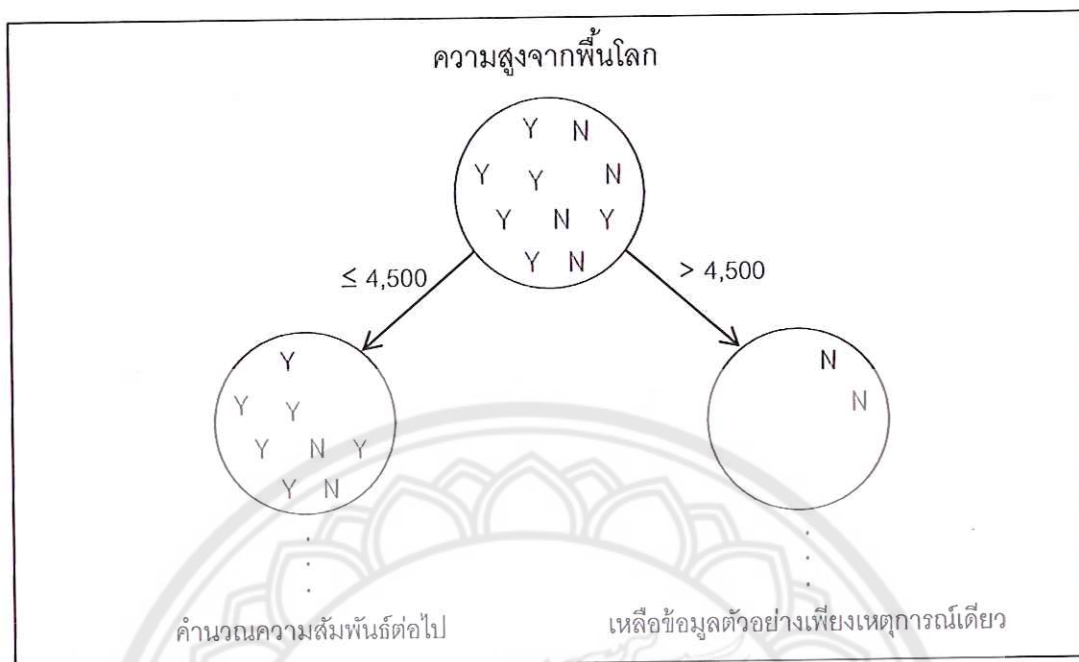
$$\text{Info}(D_{> 4,500}) = - [(0/2 \log_2 0/2) + (2/2 \log_2 2/2)] = 0$$

$$IG_{\leq 4,500 \text{ และ } > 4,500} = 0.97 - [8/10(0.81) + 2/10(0)] = 0.32$$

$$SI_{\leq 4,500 \text{ และ } > 4,500} = - [(8/10 \log_2 8/10) + (2/10 \log_2 2/10)] = 0.72$$

$$GR_{\leq 4,500 \text{ และ } > 4,500} = 0.32/0.72 = 0.44$$

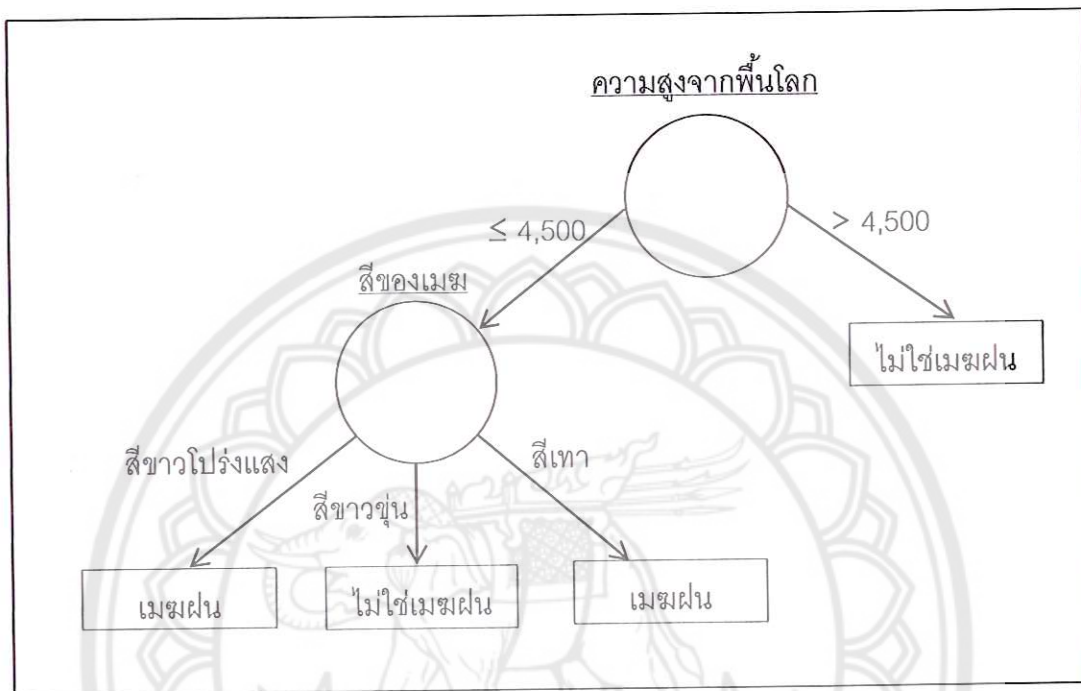
จากการแบ่งช่วงข้อมูลทั้งหมดในแอตทริบิวต์ความสูงจากพื้นโลก เมื่อพิจารณาค่า GR จะเห็นว่าช่วงตำแหน่งที่ $\leq 4,500$ และ $> 4,500$ ให้ค่า GR สูงสุดอยู่ที่ 0.44 และเมื่อทำการเปรียบเทียบกับแอตทริบิวต์ทั้งหมดในการคำนวณพบว่าค่า GR ของแอตทริบิวต์ความสูงจากพื้นโลก $>$ แอตทริบิวต์สีของเมฆ $>$ แอตทริบิวต์ความหนาแน่นของเมฆ โดยค่า GR อยู่ที่ 0.44, 0.33 และ 0.02 ตามลำดับ ดังนั้นแอตทริบิวต์ความสูงจากพื้นโลกจึงถูกเลือกให้เป็นโหนดเริ่มต้นของโครงสร้างแบบจำลองต้นไม้ช่วยตัดสินใจ J48 เนื่องจากมีค่า GR สูงที่สุด



ภาพ 9 แสดงการเลือกแอตทริบิวต์เริ่มต้นของต้นไม้ช่วยตัดสินใจ

จากภาพ 9 ทำการแตกกิ่งของต้นไม้ต่อไปจนกว่าแอตทริบิวต์นั้นจะเหลือข้อมูลตัวอย่างเพียงเหตุการณ์เดียว เมื่อพิจารณาช่วงที่ข้อมูลตัวอย่าง > 4500 จะพบว่ามีความตอบเหลือเพียงเหตุการณ์เดียวเท่านั้น คือ ไม่ใช่เมฆฝน ดังนั้นจึงไม่ต้องทำการแตกกิ่งของต้นไม้ต่อไปอีกเพราะเมื่อคำนวณค่า Entropy ในโหนดนี้จะมีค่าเท่ากับ 0 จึงไม่เหมาะสมกับการเป็นโหนดที่ใช้สำหรับการแบ่งข้อมูล ซึ่งจะแตกต่างจากช่วงที่ข้อมูลตัวอย่าง $\leq 4,500$ ที่มีคำตอบคงเหลืออยู่ 2 เหตุการณ์ คือ เมฆฝน และไม่ใช่เมฆฝน ดังนั้นในโหนดนี้จะต้องทำการแตกกิ่งของต้นไม้ต่อไปอีก หลังจากนั้นกลับไปทำขั้นตอนที่ 1 โดยให้แอตทริบิวต์ความสูงจากพื้นโลกเป็นโหนดเริ่มต้นของต้นไม้ช่วยตัดสินใจ และพิจารณาในส่วนของข้อมูลตัวอย่างที่เหลือ ได้แก่ ข้อมูลตัวอย่างที่ 1, 3, 4, 6, 7, 8, 9 และ 10 ดำเนินการเช่นนี้ไปเรื่อย ๆ จนไม่สามารถทำการแตกกิ่งของต้นไม้ต่อไปได้อีก จากภาพ 10 เป็นภาพแสดงตัวอย่างโครงสร้างของต้นไม้ช่วยตัดสินใจโดยใช้อัลกอริทึม J48 ในการจำแนกประเภทชนิดของเมฆ 2 กลุ่มที่เสร็จสมบูรณ์แล้วจากการเรียนรู้ชุดข้อมูล 10 ตัวอย่าง จะเห็นว่าแอตทริบิวต์ความสูงจากพื้นโลกและสีของเมฆก็เพียงพอที่จะใช้สำหรับการสร้างต้นไม้ช่วยตัดสินใจเพื่อทำการจำแนกชนิดของเมฆทั้ง 2 กลุ่มได้ สำหรับวิธีการนำแบบจำลองที่สร้างขึ้นไปใช้งานนั้นเมื่อมีข้อมูลตัวอย่างใหม่เข้ามา เช่น สีของเมฆเป็นสีขาว ความหนาแน่นของเมฆต่ำ ความสูงของเมฆจากพื้นโลกอยู่ที่ระดับ 4,670 เมตร แบบจำลองจะพิจารณาค่าคุณลักษณะต่าง ๆ ที่รับเข้ามา

จากนั้นแบบจำลองจะพยากรณ์ข้อมูลที่เข้ามาใหม่นี้ว่า 'ไม่ใช่เมฆฝน' เนื่องจากแบบจำลองจะพิจารณาค่าความสูงจากพื้นโลกเป็นอันดับแรก

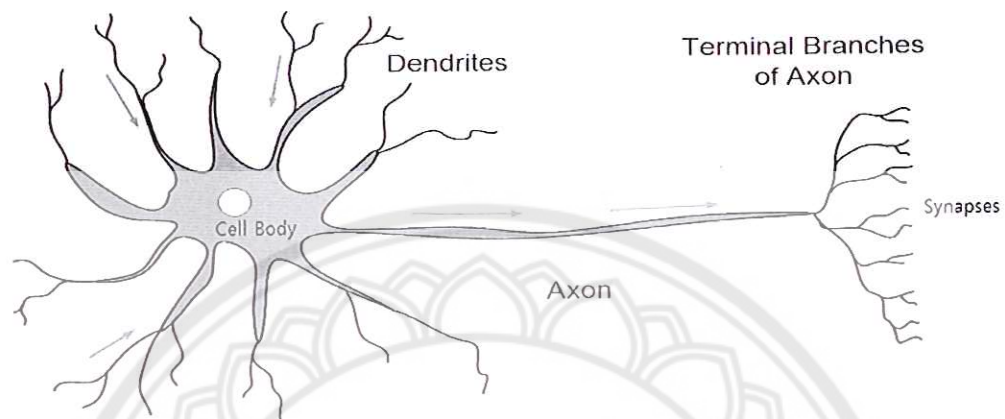


ภาพ 10 แสดงโครงสร้างต้นไม้ช่วยตัดสินใจที่เสร็จสมบูรณ์

2. โครงข่ายประสาทเทียม (Artificial Neural Networks : ANN)

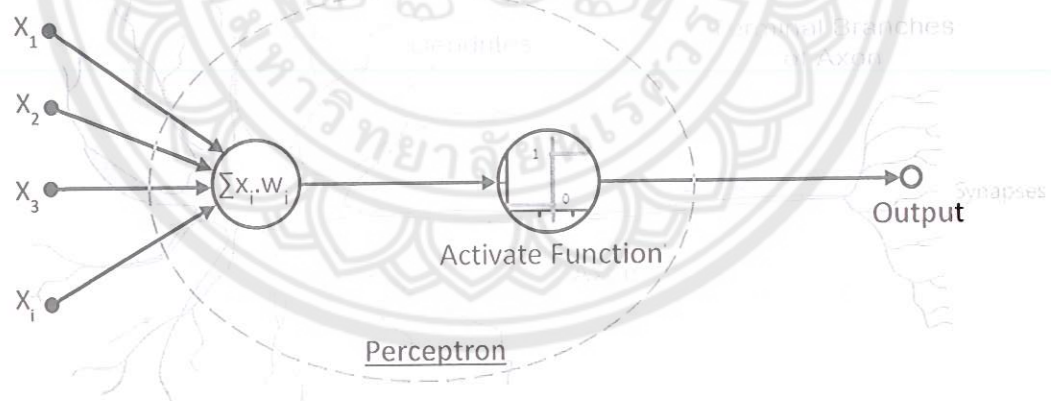
โครงข่ายประสาทเทียมจัดเป็นหนึ่งในอัลกอริทึมการเรียนรู้ของเครื่อง ซึ่งเป็นการสร้างแบบจำลองที่มีลักษณะการทำงานคล้ายกับเซลล์ประสาทในสมองของมนุษย์ เพื่อการเรียนรู้และจดจำรูปแบบของข้อมูล (Pattern Recognition) จากภาพ 11 เป็นลักษณะการทำงานของเซลล์ประสาทในสมองของมนุษย์ โดยในสมองจะประกอบด้วย เซลล์ประสาทหรือนิวรอน (Neurons) และจุดประสานประสาท (Synapses) ในแต่ละเซลล์ประสาทจะประกอบด้วยปลายในการรับกระแสประสาท (Dendrite) และปลายในการส่งกระแสประสาท (Axon) ซึ่งเซลล์เหล่านี้เมื่อถูกกระตุ้นด้วยสิ่งเร้าภายนอกหรือกระตุ้นด้วยเซลล์ด้วยกัน กระแสประสาทจะวิ่งผ่านเดนไดรต์เข้าสู่เซลล์ประสาท (Cell Body) ซึ่งในเซลล์ประสาทจะมีนิวเคลียส (Nucleus) เป็นตัวพิจารณาในการกระตุ้นเซลล์อื่น ๆ ถ้ากระแสประสาทที่รับเข้ามาแรงพอ นิวเคลียสก็จะกระตุ้นเซลล์อื่น ๆ ต่อไปผ่านทางท่อแอกซอนเพื่อส่งสัญญาณไปยังจุดประสานประสาท โดยทั่วไปนั้นสถาปัตยกรรมแบบจำลอง

โครงข่ายประสาทเทียมมีอยู่หลากหลายชนิดในแต่ละชนิดมีการนำไปประยุกต์ใช้กับงานที่มีลักษณะแตกต่างกันออกไปตามความเหมาะสม



ภาพ 11 แสดงโครงสร้างของเซลล์ประสาทในสมองของมนุษย์

ที่มา: Nelson, A. L., 2014

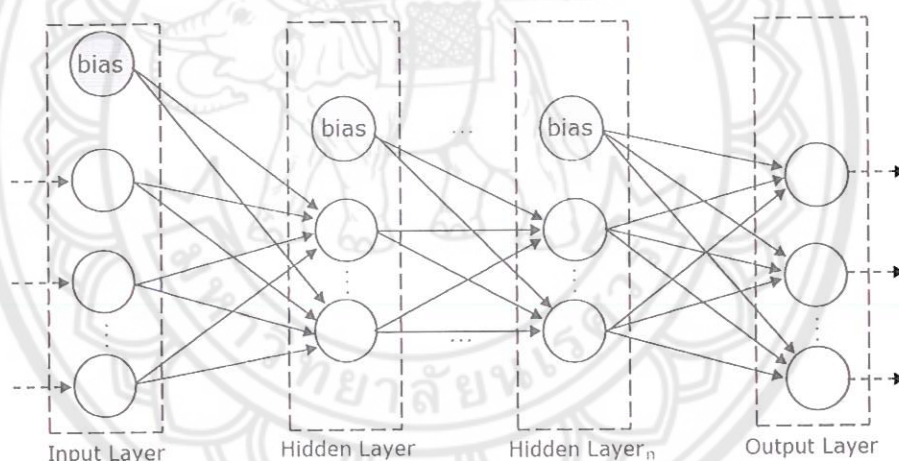


ภาพ 12 แสดงโครงสร้างของสถาปัตยกรรมโครงข่ายประสาทเทียม

ที่มา: Nelson, A. L., 2014

จากภาพ 12 แสดงโครงสร้างการทำงานของโครงข่ายประสาทเทียม โดยเริ่มจากการรับและส่งข้อมูลมายังส่วนที่เรียกว่า "Perceptron" ซึ่งสามารถเทียบได้กับเซลล์สมองของมนุษย์ โดยที่เพอร์เซปตรอนจะทำการรับข้อมูลที่อยู่ในรูปของตัวเลขเข้ามาคำนวณ จากนั้นทำการแปลงค่าโดยผ่านฟังก์ชันกระตุ้น (Activate Function) แล้วส่งค่าผลลัพธ์ไปยังนิวรอนหรือโหนดอื่น ๆ ต่อไป

สำหรับวิทยานิพนธ์นี้ใช้โครงข่ายประสาทเทียมแบบเพอร์เซปตรอนหลายชั้น เนื่องจากเป็นประเภทที่นิยมนำมาประยุกต์ใช้กับการจำแนกประเภทของข้อมูลในหลาย ๆ งานวิจัย และยังสามารถให้ประสิทธิภาพความถูกต้องสูงสำหรับกระบวนการเรียนรู้ข้อมูลเพื่อสร้างแบบจำลอง ซึ่งโครงข่ายประสาทเทียมประเภทนี้จะใช้วิธีการปรับหาค่าน้ำหนักของเส้นเชื่อมระหว่างโหนดแต่ละโหนดอย่างเหมาะสม เพื่อลดค่าความผิดพลาด (Error) ที่เกิดขึ้นในระหว่างกระบวนการเรียนรู้ข้อมูลให้น้อยที่สุด การทำงานในลักษณะนี้เรียกว่า การเรียนรู้แบบแพร่ย้อนกลับ (Back-propagation Learning)



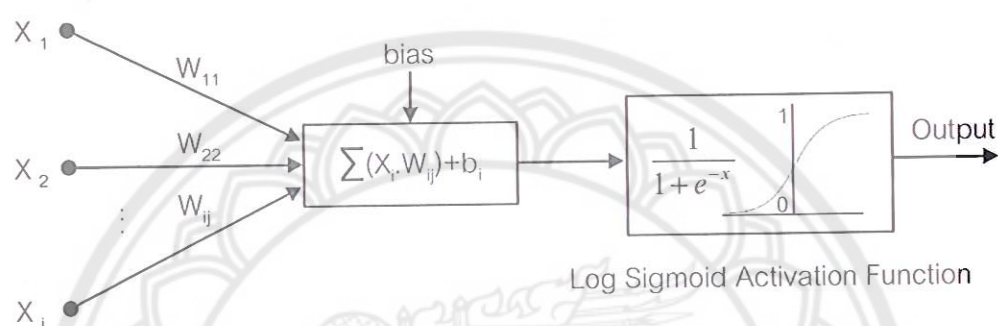
ภาพ 13 แสดงโครงสร้างสถาปัตยกรรมโครงข่ายประสาทเทียมแบบ MLP

จากภาพ 13 โครงข่ายประสาทเทียมแบบ MLP จะประกอบไปด้วย 3 ชั้น ได้แก่

1. Input Layer คือ ชั้นสำหรับรับข้อมูลนำเข้า ซึ่งจะแสดงถึงแอตทริบิวต์ต่าง ๆ ของข้อมูลเพื่อใช้ในการเรียนรู้ ในขั้นนี้จะทำการรวมผลคูณระหว่างค่าของข้อมูลที่ได้รับเข้ามากับค่าน้ำหนักของแต่ละเส้นเชื่อม (Weight) บวกกับค่าเบี่ยงเบน (Bias) และจะส่งผ่านฟังก์ชันกระตุ้นเพื่อทำการแปลงค่าข้อมูลโดยใช้ Log Sigmoid Function จากนั้นนำผลลัพธ์ส่งผ่านไปยังโหนดอื่น ๆ ในชั้นถัดไป

2. Hidden Layer คือ ชั้นที่รับค่าจากการคำนวณในชั้น Input Layer ส่งเข้ามา ซึ่งจำนวนของโหนดในชั้นนี้โดยทั่วไปค่าเริ่มต้นจะเท่ากับจำนวนแอดทริบิวต์ทั้งหมดหารสอง (สามารถปรับเพิ่มหรือลดจำนวนของโหนดในชั้นนี้ได้ และสามารถกำหนดให้มีชั้นซ่อนได้หลาย ๆ ชั้น)

3. Output Layer คือ ชั้นที่แสดงผลลัพธ์สุดท้ายที่ได้จากการคำนวณในชั้น Hidden Layer ส่งเข้ามา โดยจำนวนของโหนดในชั้นนี้จะขึ้นอยู่กับคลาสคำตอบ



ภาพ 14 แสดงเซลล์ประสาทในสถาปัตยกรรมโครงข่ายประสาทเทียม

จากภาพ 14 เป็นลักษณะการทำงานของเซลล์ประสาทในสถาปัตยกรรมโครงข่ายประสาทเทียม โดยเริ่มจากการรวมผลคูณระหว่างค่าของข้อมูลที่รับเข้า (X_i) กับค่าน้ำหนัก (W_{ij}) บวกกับค่าเบี่ยงเบนแสดงในสมการ 9

$$\text{sum}_j = \sum_{i=1}^n X_i \cdot W_{ij} + b_j \quad (9)$$

โดยที่

X_i คือ ค่าของข้อมูลนำเข้า

W_{ij} คือ เส้นเชื่อมระหว่างโหนด หรือค่าน้ำหนักระหว่างโหนด

i คือ จำนวนโหนดในชั้น Input Layer

j คือ จำนวนโหนดในชั้น Hidden Layer

n คือ จำนวนคุณลักษณะของข้อมูล

b คือ ค่าเบี่ยงเบน

จากนั้นนำผลลัพธ์ที่ได้จากการคำนวณมาแปลงค่าผ่านฟังก์ชันแบบ Log Sigmoid คำนวณได้จากสมการ 10 แล้วส่งค่าไปยังโหนดอื่น ๆ ต่อไปอีก จากชั้น Hidden Layer ไปยังชั้น Output Layer ลักษณะการทำงานแบบนี้เรียกว่า Feed-Forward (Krenker, A., Bester, J. and Kos, A., 2011)

$$\text{net}_j = \frac{1}{1 + e^{-\text{sum}_j}} \quad (10)$$

โดยที่

sum_j คือ ค่าผลรวมของผลคูณระหว่างค่าของข้อมูลนำเข้ากับค่าน้ำหนักของเส้นเชื่อมบวกกับค่าเบี่ยงเบน

เมื่อได้ค่าผลลัพธ์สุดท้ายในชั้น Output Layer โครงข่ายประสาทเทียมจะใช้วิธีการเรียนรู้แบบแพร่ย้อนกลับ โดยเริ่มจากการคำนวณหาค่าความผิดพลาดที่เกิดขึ้นจากกระบวนการเรียนรู้เพื่อใช้สำหรับปรับค่าน้ำหนักของเส้นเชื่อมระหว่างโหนดใหม่ แสดงในสมการ 11 (กรณีโหนดในชั้น Output Layer) และสมการ 12 (กรณีโหนดในชั้น Hidden Layer)

$$\delta_{j,k} = \begin{cases} \text{Output}_k (1 - \text{Output}_k) (\text{Actual}_i - \text{Output}_k) & (11) \\ \text{Output}_j (1 - \text{Output}_j) \sum_{\text{downstream}} W_{jk} \delta_j & (12) \end{cases}$$

โดยที่

Output_k คือ ค่าผลลัพธ์ที่ได้ของแต่ละโหนดในชั้น Output Layer

Output_j คือ ค่าผลลัพธ์ที่ได้ของแต่ละโหนดในชั้น Hidden Layer

Actual_i คือ ค่าข้อมูลจริง หรือค่าของคลาสคำตอบของข้อมูลตัวอย่าง

W_{jk} คือ ค่าน้ำหนักระหว่างโหนด

δ_j คือ ค่าความผิดพลาดที่เกิดจากกระบวนการเรียนรู้

i คือ จำนวนโหนดในชั้น Input Layer

j คือ จำนวนโหนดในชั้น Hidden Layer

k คือ จำนวนโหนดในชั้น Output Layer

จากนั้นทำการปรับค่าน้ำหนักที่ได้จากการคำนวณค่าความผิดพลาดที่เกิดขึ้นของเส้นเชื่อมใหม่ทั้งหมด คำนวณได้ตามสมการ 13 และ 14

$$bais_{(update)} = \eta\delta + W_{(current)} \quad (13)$$

$$W_{(update)} = \eta\delta O_{(previous)} + W_{(current)} \quad (14)$$

โดยที่

$O_{(previous)}$ คือ ผลลัพธ์ของโหนดในชั้นก่อนหน้า

$W_{(current)}$ คือ ค่าน้ำหนักของเส้นเชื่อมเดิม

η คือ ค่าอัตราการเรียนรู้

δ คือ ค่าความผิดพลาดที่เกิดขึ้นจากกระบวนการเรียนรู้

สำหรับฟังก์ชันกระตุ้นนั้นก็มีหลายประเภทสามารถอธิบายได้ในงานวิจัยของ Duch, W. และ Jankowski, N., (1999) การเลือกใช้ฟังก์ชันกระตุ้นที่เหมาะสมจะมีผลต่อการเพิ่มประสิทธิภาพให้กับการทำงานในโครงข่ายประสาทเทียม (Ozkan, C. and Erbek, F.S., 2003) โดยทั่วไปโครงข่ายประสาทเทียม MLP ส่วนใหญ่จะใช้ฟังก์ชันกระตุ้นแบบ Log Sigmoid Function สามารถอธิบายรายละเอียดการสร้างแบบจำลองด้วยโครงข่ายประสาทเทียม MLP ดังตัวอย่างต่อไปนี้

ตัวอย่างการสร้างแบบจำลองโครงข่ายประสาทเทียม MLP

ตาราง 3 ตัวอย่างข้อมูลสำหรับการจำแนกประเภทชนิดของเมฆ 2 กลุ่ม โดยแปลงค่าคุณลักษณะข้อมูลของเมฆจากตาราง 2 ให้อยู่ในรูปแบบตัวเลข

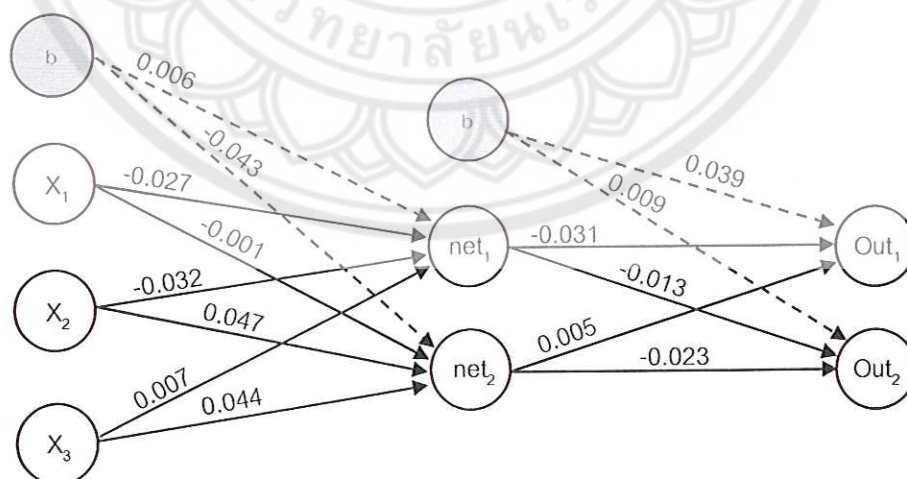
สีของเมฆ	ความหนาแน่นของเมฆ	ความสูงจากพื้นโลก (เมตร)	กลุ่มของเมฆ
1	1	3,000	1
1	1	7,000	0
2	2	2,300	1
2	2	1,000	1
1	1	5,500	0

ตาราง 3 (ต่อ)

สีของเมฆ	ความหนาแน่นของเมฆ	ความสูงจากพื้นโลก (เมตร)	กลุ่มของเมฆ
1	2	1,500	1
3	2	2,500	0
1	2	1,500	1
2	1	3,500	1
3	2	2,500	0

จากตาราง 3 แสดงตัวอย่างข้อมูลสำหรับการเรียนรู้เพื่อสร้างแบบจำลองโครงข่ายประสาทเทียม MLP ในการจำแนกประเภทชนิดของเมฆ 2 กลุ่ม ซึ่งประกอบด้วยแอตทริบิวต์สีของเมฆ, ความหนาแน่นของเมฆ, ความสูงจากพื้นโลก และกลุ่มของเมฆซึ่งเป็นแอตทริบิวต์คำตอบ โดย 1 แทนกลุ่มของเมฆที่เป็นเมฆฝน และ 0 แทนกลุ่มเมฆที่ไม่ใช่เมฆฝน

ขั้นตอนที่ 1 กำหนดจำนวนรอบสำหรับการเรียนรู้ และค่าความผิดพลาดที่ยอมรับได้ รวมถึงจำนวนโหนดในชั้น Hidden Layer กรณีไม่กำหนดจำนวนโหนดในชั้นนี้ จำนวนโหนดจะเท่ากับจำนวนแอตทริบิวต์ทั้งหมดหารสอง ($N_h = n/2$) ซึ่งได้มีนักวิจัยนำเสนอทฤษฎีเกี่ยวกับการกำหนดจำนวนโหนดในชั้น Hidden Layer ไว้หลากหลายวิธี

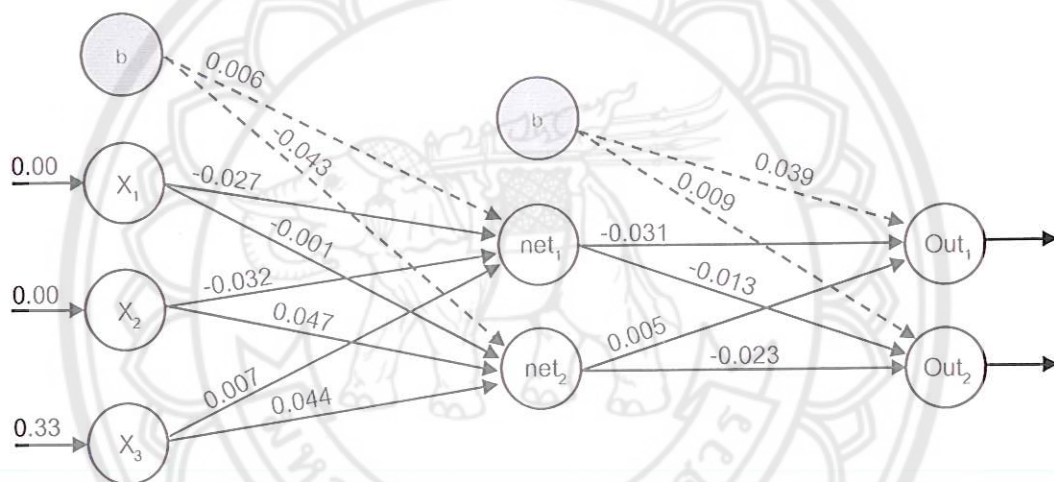


ภาพ 15 แสดงการสุ่มค่าน้ำหนักเริ่มต้นและค่าเบี่ยงเบนในโครงข่ายประสาทเทียม

โดยในงานวิจัยของ Sheela, K.G. และ Deepa, S.N., (2014) ได้นำเสนอการปรับปรุงวิธีการกำหนดจำนวนโหนดในชั้น Hidden Layer อย่างเหมาะสมอยู่ที่ $N_h = (4n^2 + 3) / (n^2 - 8)$ สำหรับในตัวอย่างนี้ได้กำหนดจำนวนโหนดในชั้น Hidden Layer เท่ากับ 2 โหนด

ขั้นตอนที่ 2 สุ่มค่าน้ำหนักของเส้นเชื่อมระหว่างโหนดทั้งหมด รวมถึงค่าเบี่ยงเบนของแต่ละโหนดในชั้น Hidden Layer และ Output Layer แสดงในภาพ 15

ขั้นตอนที่ 3 รับค่าของข้อมูลนำเข้าในแถวแรก แสดงได้ดังภาพ 16 สำหรับในขั้นตอนนี้จะต้องทำการแปลงค่าข้อมูล (Normalization) เพื่อให้ได้ค่าอยู่ระหว่าง 0 ถึง 1 จากสมการ 15 สามารถแสดงผลลัพธ์ของการทำนอมอลไลซ์ในตาราง 4



ภาพ 16 แสดงการนำข้อมูลเข้าในชั้น Input Layer ที่ผ่านกระบวนการ Normalization

$$\text{Normalization} = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \quad (15)$$

โดยที่

X_i คือ ค่าในตำแหน่งข้อมูลตัวที่ i

$X_{\max}$ คือ ค่ามากที่สุดสำหรับแอตทริบิวต์

$X_{\min}$ คือ ค่าน้อยที่สุดสำหรับแอตทริบิวต์

ตาราง 4 ตัวอย่างข้อมูลคุณลักษณะของเมฆที่ได้ทำการ Normalize

สีของเมฆ	ความหนาแน่นของเมฆ	ความสูงจากพื้นโลก (เมตร)	ชนิดของเมฆ
0.00	0.00	0.33	1
0.00	0.00	1.00	0
0.50	1.00	0.22	1
0.50	1.00	0.00	1
0.00	0.00	0.75	0
0.00	1.00	0.08	1
1.00	1.00	0.25	0
0.00	1.00	0.08	1
0.50	0.00	0.42	1
1.00	1.00	0.25	0

ขั้นตอนที่ 4 คำนวณผลรวมของผลคูณระหว่างค่าของข้อมูลนำเข้ากับค่าน้ำหนักของเส้นเชื่อมบวกกับค่าเบี่ยงเบน ซึ่งคำนวณได้จากสมการ 9 จากนั้นทำการแปลงค่าข้อมูลผ่านฟังก์ชันกระตุ้นแบบ Log Sigmoid ในสมการ 10

โหนดที่ 1 ชั้น Hidden Layer

$$net_1 = 0(-0.027) + 0(-0.032) + 0.33(0.007) + 0.006$$

$$net_1 = 0.008$$

$$f(net_1) = 1 / (1 + 2.72^{-0.008}) = 0.502$$

โหนดที่ 2 ชั้น Hidden Layer

$$net_2 = 0(-0.001) + 0(0.047) + 0.33(0.044) - 0.043$$

$$net_2 = -0.028$$

$$f(net_2) = 1 / (1 + 2.72^{0.028}) = 0.493$$

โหนดที่ 1 ชั้น Output Layer

$$Output_1 = 0.502(-0.031) + 0.493(0.005) + 0.039$$

$$Output_1 = 0.026$$

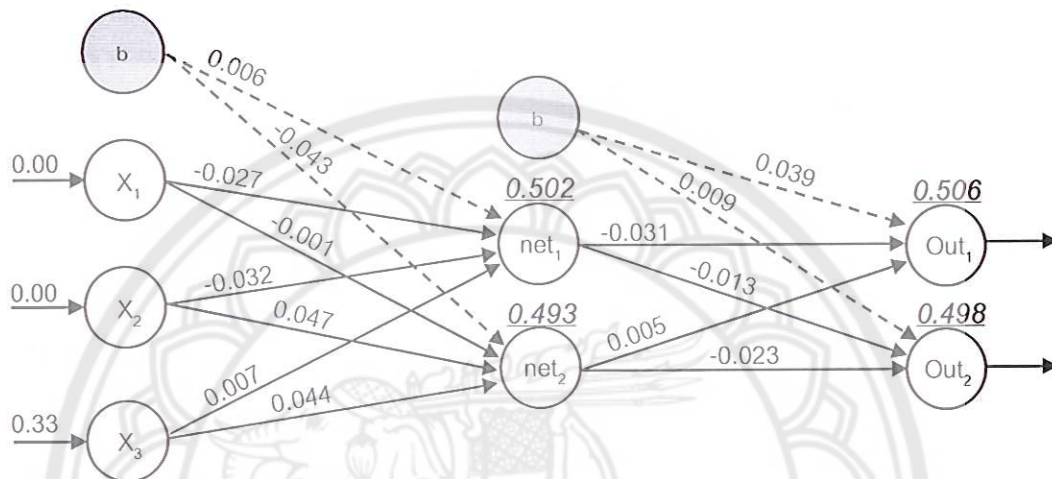
$$f(Output_1) = 1 / (1 + 2.72^{-0.026}) = 0.506$$

โหนดที่ 2 ชั้น Output Layer

$$\text{Output}_2 = 0.502(-0.013) + 0.493(-0.023) + 0.009$$

$$\text{Output}_2 = -0.009$$

$$f(\text{Output}_2) = 1/1 + 2.72^{0.009} = 0.498$$



ภาพ 17 แสดงผลลัพธ์ที่ได้จากกระบวนการ Feed-Forward

จากภาพ 17 เป็นการแสดงผลลัพธ์สุดท้ายที่ได้จากกระบวนการทำงานแบบ Feed-Forward เมื่อได้ค่าผลลัพธ์ในชั้น Output Layer โครงข่ายประสาทเทียมจะทำการหาข้อผิดพลาดจากวิธีการเรียนรู้แบบแพร่ย้อนกลับต่อไป

ขั้นตอนที่ 5 คำนวณค่าความผิดพลาดของโหนดในชั้น Output Layer จากสมการ 11 รวมทั้งปรับปรุงค่าเบี่ยงเบนจากสมการ 13 และค่าน้ำหนักของเส้นเชื่อมระหว่างโหนดทั้งหมดจากสมการ 14 สามารถแสดงรายละเอียดการคำนวณได้ ดังนี้

$$\delta_{(\text{Output1})} = 0.506(1-0.506)(1-0.506) = 0.123$$

$$\delta_{(\text{Output2})} = 0.498(1-0.498)(1-0.498) = 0.126$$

$$\text{Bias}_{\text{Output1}(\text{update})} = 0.1(0.123) + 0.039 = 0.051$$

$$\text{Bias}_{\text{Output2}(\text{update})} = 0.1(0.126) + 0.009 = 0.022$$

$$W_{11(\text{update})} = 0.1(0.123)(0.502) - 0.031 = -0.025$$

$$W_{12(\text{update})} = 0.1(0.126)(0.502) - 0.013 = -0.007$$

$$W_{21(\text{update})} = 0.1(0.123)(0.493) + 0.005 = 0.011$$

$$W_{22(\text{update})} = 0.1(0.126)(0.493) - 0.023 = -0.017$$

ขั้นตอนที่ 6 คำนวณค่าความผิดพลาดของโหนดในชั้น Hidden Layer จากสมการ 12 และปรับปรุงค่าเบี่ยงเบนกับค่าน้ำหนักของเส้นเชื่อมระหว่างโหนดทั้งหมด

$$\delta_{(\text{net1})} = 0.502(1-0.502)((-0.031(0.123)-0.013(0.126))) = -0.001$$

$$\delta_{(\text{net2})} = 0.493(1-0.493)((0.005(0.123)-0.023(0.126))) = -0.001$$

$$\text{Bias}_{\text{Output1}(\text{update})} = 0.1(-0.001) + 0.006 = 0.006$$

$$\text{Bias}_{\text{Output2}(\text{update})} = 0.1(-0.001) - 0.043 = -0.043$$

$$W_{11(\text{update})} = 0.1(-0.001)(0) - 0.027 = -0.027$$

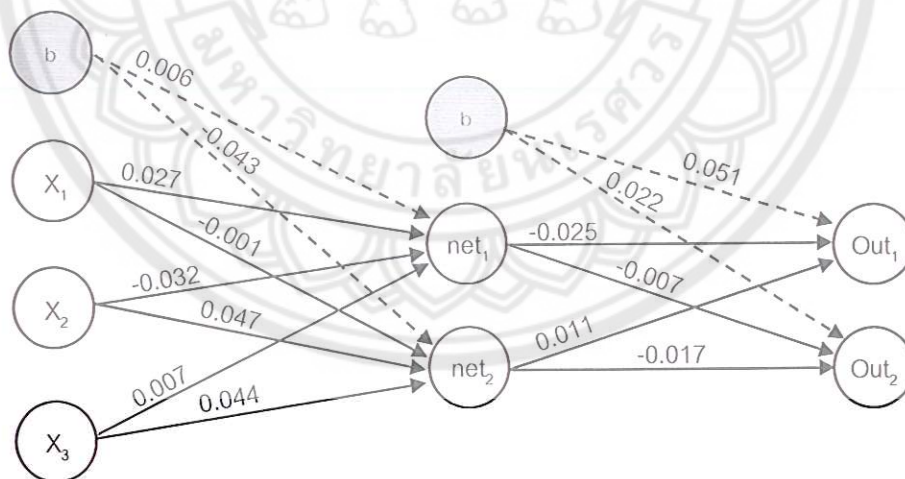
$$W_{12(\text{update})} = 0.1(-0.001)(0) - 0.001 = -0.001$$

$$W_{21(\text{update})} = 0.1(-0.001)(0) - 0.032 = -0.032$$

$$W_{22(\text{update})} = 0.1(-0.001)(0) + 0.047 = 0.047$$

$$W_{31(\text{update})} = 0.1(-0.001)(0.33) + 0.007 = 0.007$$

$$W_{32(\text{update})} = 0.1(-0.001)(0.33) + 0.044 = 0.044$$



ภาพ 18 แสดงผลลัพธ์ของการปรับปรุงค่าน้ำหนักและค่าเบี่ยงเบนใหม่

ขั้นตอนที่ 7 รับค่าข้อมูลนำเข้าใหม่ในแถวถัดไปแล้วกลับไปทำในขั้นตอนที่ 4 แต่ถ้าเป็นข้อมูลในแถวสุดท้ายหรืออ่านค่าข้อมูลตัวอย่างครบทุกข้อมูลแล้วให้ไปทำในขั้นตอนที่ 8

ขั้นตอนที่ 8 คำนวณค่าความผิดพลาดเฉลี่ย (Mean Squared Error : MSE) ของทุกแถวข้อมูลในสมการ 16 ถ้าค่าความผิดพลาดเฉลี่ยมีค่าน้อยกว่าค่าที่ยอมรับได้ให้จบกระบวนการเรียนรู้ แต่ถ้าค่าความผิดพลาดเฉลี่ยมีค่ามากกว่าค่าที่ยอมรับได้จะต้องทำการตรวจสอบว่าครบตามจำนวนรอบที่กำหนดไว้ในขั้นตอนที่ 1 หรือไม่ ถ้าครบให้จบกระบวนการเรียนรู้ ถ้าไม่ครบให้กลับไปทำในขั้นตอนที่ 3

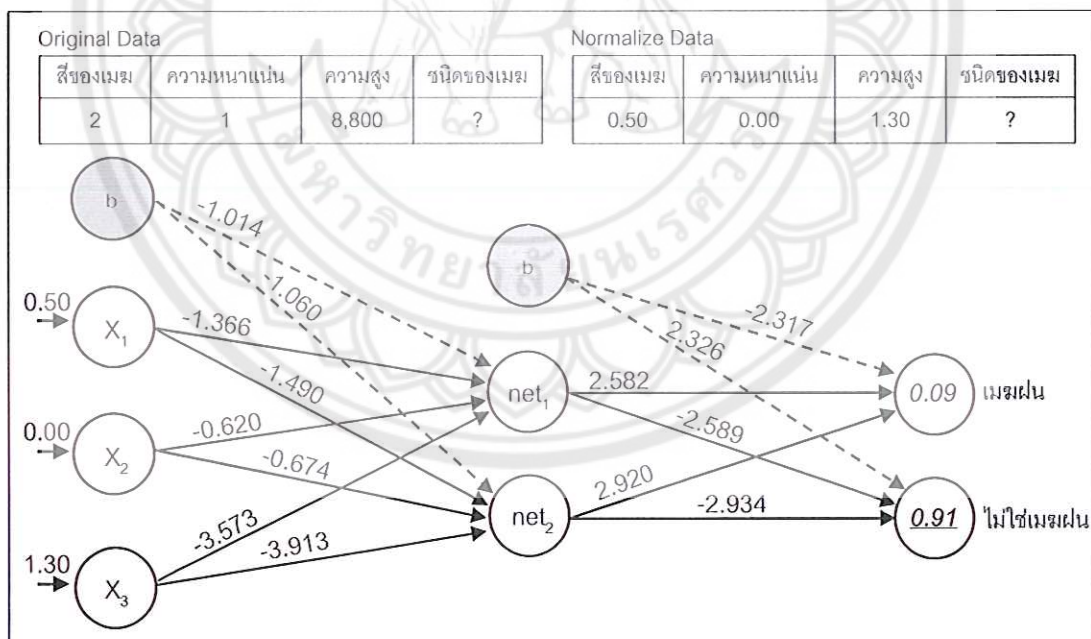
$$MSE = \frac{1}{n} \sum_{i=1}^n (X_{\text{Actual}} - X_k)^2 \quad (16)$$

โดยที่

X_{Actual} คือ ค่าของข้อมูลจริง

X_k คือ ค่าผลลัพธ์ของโหนดที่ k ในชั้น Output Layer

n คือ ตัวอย่างข้อมูลทั้งหมด



ภาพ 19 แสดงการนำโครงข่ายประสาทเทียมมาประยุกต์ใช้ในการจำแนกประเภท

ภาพ 19 เป็นการนำแบบจำลองโครงข่ายประสาทเทียมที่สร้างเสร็จสมบูรณ์แล้วไปประยุกต์ใช้จริงในการจำแนกประเภทชนิดของกลุ่มเมฆ ซึ่งค่าพารามิเตอร์ที่ได้กำหนดไว้สำหรับการออกแบบ

โครงข่ายประสาทเทียมตามที่ได้แสดงในตัวอย่างมีดังต่อไปนี้ จำนวนรอบที่ใช้ในการเรียนรู้ 500 รอบ อัตราการเรียนรู้อยู่ที่ 0.1 ค่าความผิดพลาดที่ยอมรับได้เท่ากับ 0.001 และค่าโมเมนตัมเท่ากับ 0 (ค่าโมเมนตัมจะขออธิบายในส่วนถัดไป) จากภาพ 19 จะเห็นว่าเมื่อมีค่าคุณลักษณะของข้อมูลใหม่เข้ามาในแบบจำลอง ได้แก่ สีของเมฆเท่ากับ 0.5, ความหนาแน่นของเมฆเท่ากับ 0.00 และความสูงของเมฆเท่ากับ 1.30 แบบจำลองโครงข่ายประสาทเทียมจะรับค่าเหล่านี้เข้ามาคำนวณเพื่อนำผลลัพธ์สุดท้ายที่ได้จากกระบวนการทำงานแบบ Feed-Forward มาพิจารณาในการจำแนกประเภทชนิดของกลุ่มเมฆ ดังตัวอย่างจะเห็นว่าผลลัพธ์สุดท้ายที่ได้จากโหนดในชั้น Output Layer คลาสคำตอบของกลุ่มที่เป็นเมฆฝนมีค่าเท่ากับ 0.09 ส่วนที่ไม่ใช่เมฆฝนมีค่าเท่ากับ 0.91 ดังนั้นแบบจำลองจะพยากรณ์ว่า ข้อมูลตัวอย่างที่เข้ามาใหม่นี้ไม่ใช่เมฆฝน เนื่องจากค่าผลลัพธ์ที่ได้มีค่าสูงกว่าคลาสคำตอบของกลุ่มที่เป็นเมฆฝน

สำหรับโมเมนตัม (Momentum) เป็นค่าที่กำหนดเพื่อช่วยเพิ่มอัตราการหวนให้กับค่าอัตราการเรียนรู้ของข้อมูลในการปรับค่าน้ำหนักของเส้นเชื่อมรวมถึงค่าเบี่ยงเบี่ยงในระหว่างกระบวนการเรียนรู้ข้อมูลของขั้นตอนการเรียนรู้แบบแพร่ย้อนกลับ และยังสามารถช่วยให้ประสิทธิภาพการทำงานของโครงข่ายประสาทเทียมดีขึ้น โดยการนำค่าโมเมนตัมที่กำหนดไปคูณกับค่าน้ำหนักเดิมตามสมการ 17 การกำหนดค่าโมเมนตัมให้เหมาะสมยังสามารถช่วยแก้ไขปัญห Local Trap ได้ ซึ่งเป็นปัญหาที่เกิดขึ้นจากการหาค่าความผิดพลาดไม่ลู่เข้าสู่จุดที่มีค่าต่ำสุดอย่างแท้จริงที่เกิดขึ้นภายในขั้นตอนการปรับค่าน้ำหนักของเส้นเชื่อมและค่าเบี่ยงเบี่ยง

$$W_{(update)} = \eta \delta O_{(Previous)} + \alpha W_{(current)} \quad (17)$$

โดยที่

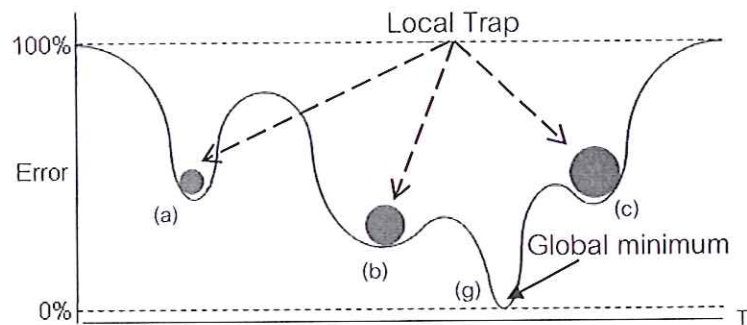
$O_{(Previous)}$ คือ ผลลัพธ์ของโหนดในชั้นก่อนหน้า

$W_{(current)}$ คือ ค่าน้ำหนักของเส้นเชื่อมเดิม

η คือ ค่าอัตราการเรียนรู้

δ คือ ค่าความผิดพลาดที่เกิดขึ้นจากกระบวนการเรียนรู้

α คือ ค่าโมเมนตัม



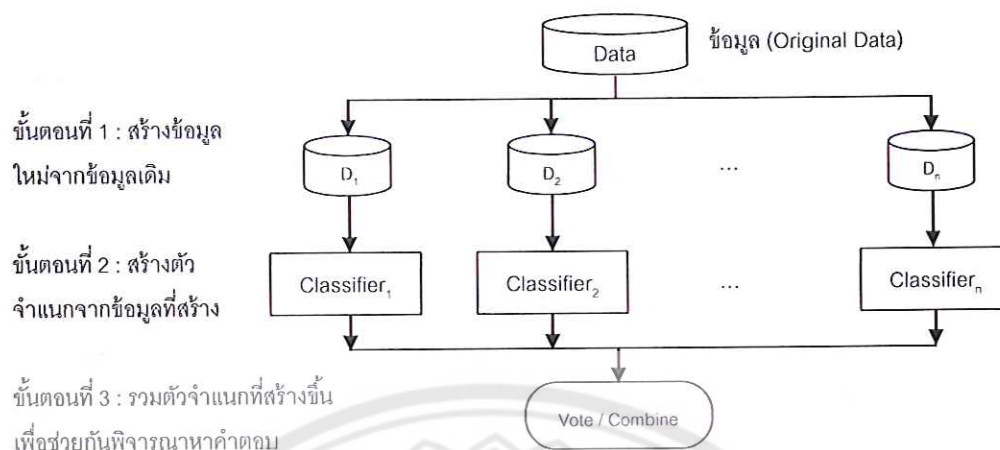
ภาพ 20 แสดงการเกิดปัญหา Local Trap

จากภาพ 20 แสดงการเกิดปัญหา Local Trap ที่ตำแหน่ง (a), (b) และ (c) ซึ่งเป็นปัญหาที่ทำให้ไม่สามารถพบค่าความผิดพลาดในจุดที่ต่ำสุดได้อย่างสมบูรณ์ ดังนั้นการกำหนดค่าโมเมนตัมที่เหมาะสมจะช่วยเพิ่มประสิทธิภาพในการเรียนรู้ให้กับโครงข่ายประสาทเทียมได้ เนื่องจากสามารถทำให้พบตำแหน่งค่าความผิดพลาดในจุดที่ต่ำสุด (Global Minimum) อย่างแท้จริงในตำแหน่ง (g)

การเรียนรู้แบบรวมกลุ่ม (Ensemble Learning)

การเรียนรู้แบบรวมกลุ่มเป็นวิธีการรวมเอากลุ่มของตัวจำแนกประเภทข้อมูลที่สร้างขึ้นหลาย ๆ ตัวจำแนกและมีความเป็นอิสระต่อกันมาพิจารณาร่วมกันเพื่อช่วยในการตัดสินใจสำหรับการหาคำตอบโดยใช้วิธีการรวม (Combine) หรือด้วยวิธีการโหวต (Voting) เพื่อให้ได้ผลลัพธ์ในการจำแนกข้อมูลที่มีประสิทธิภาพสูง ซึ่งเทคนิคการเรียนรู้แบบรวมกลุ่มนั้นมีอยู่หลากหลายวิธีการ แต่สำหรับวิธีการที่มีประสิทธิภาพที่ได้รับความนิยม ได้แก่ วิธี Bagging, Boosting และ Random Forest (Yang, P., Yang, Y.H., Zhou, B.B. and Zomaya, A.Y., 2010)

สำหรับวิทยานิพนธ์นี้ได้ทำการศึกษาและเปรียบเทียบเทคนิคการเรียนรู้แบบรวมกลุ่ม 3 วิธี ได้แก่ Bagging, Boosting และ Random Forest เพื่อเปรียบเทียบประสิทธิภาพสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆ

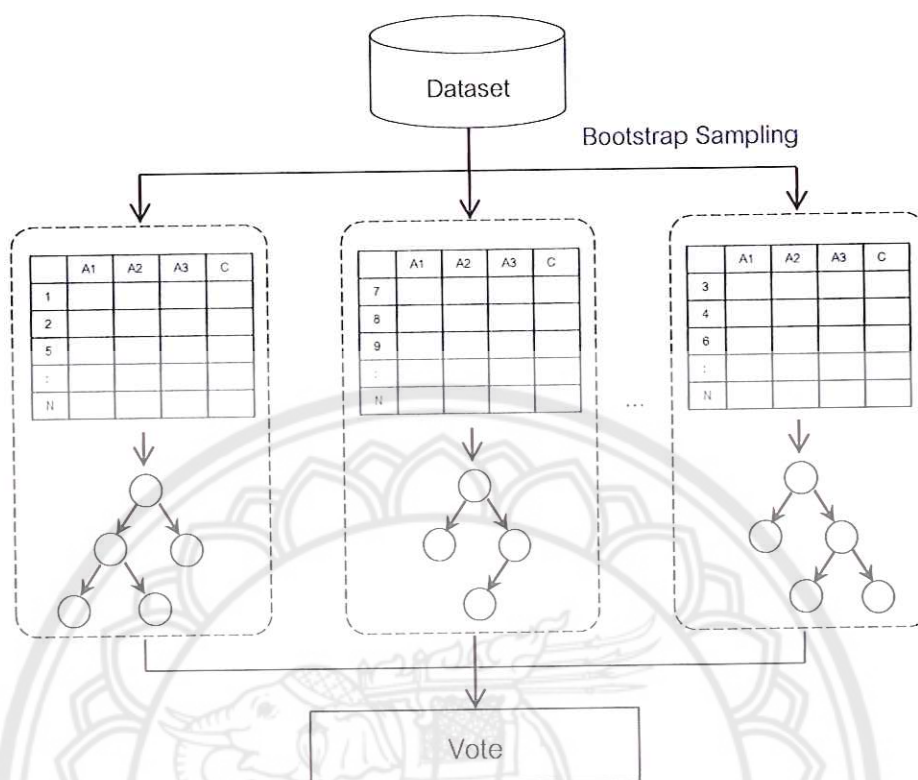


ภาพ 21 แสดงโครงสร้างพื้นฐานของการเรียนรู้แบบรวมกลุ่ม

จากภาพ 21 แสดงโครงสร้างพื้นฐานของแบบจำลองการเรียนรู้แบบรวมกลุ่ม ซึ่งจะประกอบไปด้วย 3 ขั้นตอนหลัก ได้แก่ ขั้นตอนที่ 1 การสร้างชุดข้อมูลตัวอย่างขึ้นมาใหม่หลาย ๆ ชุดที่มีลักษณะแตกต่างกันในแต่ละชุด ขั้นตอนที่ 2 สร้างตัวจำแนกประเภทข้อมูลหลาย ๆ ตัวจำแนกเพื่อเรียนรู้ชุดข้อมูลที่สร้างขึ้นในแต่ละชุด และขั้นตอนสุดท้ายเป็นการรวมตัวจำแนกประเภทหลาย ๆ ตัวจำแนกที่สร้างขึ้นจากขั้นตอนที่ 2 เพื่อร่วมกันตัดสินใจในการพิจารณาคำตอบโดยใช้วิธีการรวมแบบจำลองหรือการโหวตจากเสียงข้าง (Majority Vote) มากเพื่อให้ได้คำตอบที่ดีที่สุด

1. Bagging Method

วิธีการเรียนรู้แบบรวมกลุ่ม Bagging ถูกนำเสนอขึ้นโดย Breiman, L., (1996) Bagging หรือเรียกอีกอย่างว่า Bootstrap Aggregating เป็นวิธีหนึ่งของการเรียนรู้แบบรวมกลุ่มที่มีประสิทธิภาพ โดยส่วนใหญ่จะใช้งานร่วมกับการจำแนกประเภทในรูปแบบต้นไม้ช่วยตัดสินใจ แต่ก็สามารถใช้งานร่วมกับอัลกอริทึมการจำแนกประเภทในรูปแบบอื่น ๆ ได้ ซึ่งกระบวนการทำงานจะสุ่มชุดของข้อมูลตัวอย่างขึ้นมาใหม่โดยใช้วิธีที่เรียกว่า "Bootstrap Sampling" ซึ่งข้อมูลในแต่ละชุดที่ถูกสุ่มขึ้นมาจะถูกเรียนรู้ที่มีลักษณะต่างกันในการสร้างแบบจำลอง เนื่องจากข้อมูลที่ถูกเรียนรู้มีความแตกต่างกันทำให้เกิดความหลากหลายของแบบจำลอง ผลลัพธ์สุดท้ายจะทำการโหวตเสียงข้างมากจากแบบจำลองที่สร้างขึ้นเหล่านั้นเพื่อให้ได้ผลลัพธ์ที่ดีที่สุดเพียงหนึ่งคำตอบ วิธีการทำงานแบบ Bagging จะมีผลกระทบต่อการจำแนกประเภทข้อมูลที่ไม่ใช่เชิงเส้น (Non-linear) เมื่อข้อมูลที่ใช้ในการเรียนรู้ถูกเปลี่ยนแปลงเพียงเล็กน้อย (Sewell, M., 2007) สามารถแสดงวิธีการทำงานแบบ Bagging ได้ดังภาพ 22



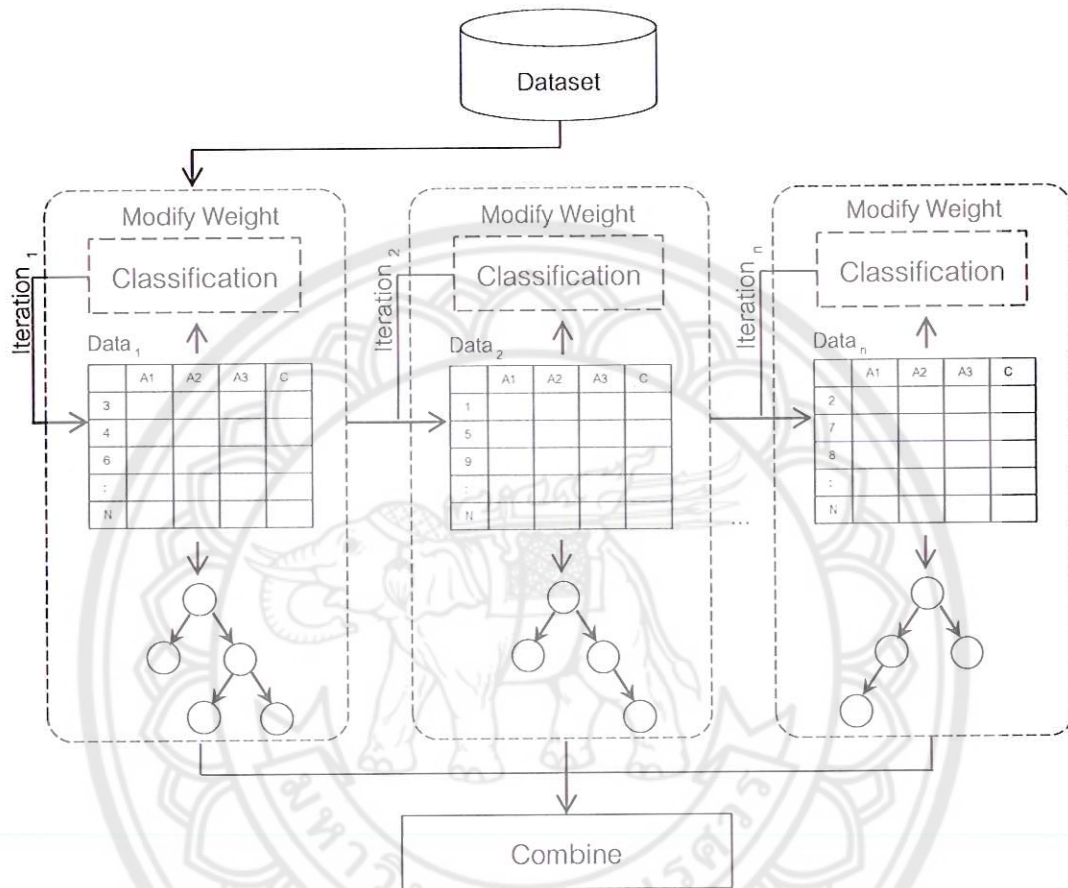
ภาพ 22 แสดงโครงสร้างวิธีการทำงานแบบ Bagging

ที่มา: Yang, P., Yang, Y.H., Zhou, B.B. and Zomaya, A.Y., 2010

2. Boosting Method

Boosting เป็นวิธีการเรียนรู้แบบรวมกลุ่มอีกหนึ่งวิธีที่นิยมนำมาใช้ในการสร้างแบบจำลองการเรียนรู้แบบรวมกลุ่ม ลักษณะจะแตกต่างจากวิธีการของ Bagging ในส่วนของ การถ่วงค่าน้ำหนักให้กับข้อมูลตัวอย่างที่ได้ทำการเรียนรู้โดยเน้นไปที่การหาความผิดพลาดที่เกิดขึ้น จากกระบวนการเรียนรู้ข้อมูล เรียกลักษณะวิธีการนี้ว่า "Weak Learning" และในขั้นตอนสุดท้าย จะใช้วิธีการรวมตัวจำแนกประเภทที่สร้างขึ้นหลาย ๆ ตัวจำแนกโดยพิจารณาจากค่าเฉลี่ยในการ ถ่วงน้ำหนัก (Mean Weight) และทำการโหวตเพื่อให้ได้ผลลัพธ์เพียงคำตอบเดียว สำหรับวิธีการ ของ Boosting ที่นิยมนำมาใช้ คือ Adaptive Boosting (AdaBoost) ซึ่งถูกพัฒนาขึ้น โดย Freund, Y. และ Schapire, R.E., (1996) วิธีการของ Adaboost จะทำการถ่วงค่าน้ำหนักให้กับข้อมูล ตัวอย่างที่ถูกเรียนรู้ของแต่ละรอบในการสร้างแบบจำลอง โดยข้อมูลตัวอย่างที่ทำการจำแนก

ประเภทได้ถูกต้องจะถูกลดค่าน้ำหนักลง ส่วนข้อมูลตัวอย่างที่จำแนกประเภทผิดพลาดจะถูกเพิ่มค่าน้ำหนักให้มีความสำคัญมากขึ้นเพื่อให้ข้อมูลนั้นมีโอกาสถูกเลือกในการเรียนรู้ของรอบต่อไป



ภาพ 23 แสดงโครงสร้างวิธีการทำงานแบบ Boosting (Adaboost)

ที่มา: Yang, P., Yang, Y.H., Zhou, B.B. and Zomaya, A.Y., 2010

สามารถแสดงวิธี Boosting แบบ Adaboost ได้ดังภาพ 23 เนื่องจาก AdaBoost เป็นอัลกอริทึมที่มีการปรับปรุงค่าน้ำหนักให้กับข้อมูล ดังนั้นการหาค่าน้ำหนักจะอาศัยค่าความผิดพลาดที่เกิดขึ้นจากกระบวนการเรียนรู้ของข้อมูลเป็นหลัก (Han, J., Kamber, M. and Pei, J., 2012)

$$\varepsilon_i = \sum_{k: c_i(x_k) \neq y_k} D_i(k) \quad (18)$$

เริ่มจากการคำนวณหาค่าความผิดพลาดที่เกิดขึ้นในการเรียนรู้ข้อมูลตามสมการ 18 โดยที่ D_i เป็นการกำหนดน้ำหนักให้กับข้อมูล $D_i = 1/m$ เมื่อ m คือ จำนวนข้อมูลตัวอย่างทั้งหมด และ C_i เป็นตัวจำแนกประเภทที่จำแนกข้อมูลตัวที่ k ผิดพลาด ซึ่ง ϵ_i คือ ผลรวมของค่าความผิดพลาดที่ได้จากตัวจำแนกประเภทที่จำแนกข้อมูลตัวที่ k ผิดพลาด หลังจากนั้นทำการคำนวณค่าน้ำหนักของข้อมูลในสมการ 19

$$\alpha_i = \frac{1}{2} \ln\left(\frac{1-\epsilon_i}{\epsilon_i}\right) \quad (19)$$

ทำการปรับปรุงค่าน้ำหนักข้อมูลตัวอย่างแต่ละตัวจากสมการ 20 โดยที่ Z_i คือ ปัจจัยความเป็นปกติ (Normalization Factor) หาได้จากสมการ 21

$$D_{i+1}(k) = \frac{D_i(k)}{Z_i} \times \begin{cases} e^{-\alpha_i} & \text{if } C_i(x_k) = y_k \\ e^{\alpha_i} & \text{if } C_i(x_k) \neq y_k \end{cases} \quad (20)$$

$$Z_i = 2\sqrt{\epsilon_i(1-\epsilon_i)} \quad (21)$$

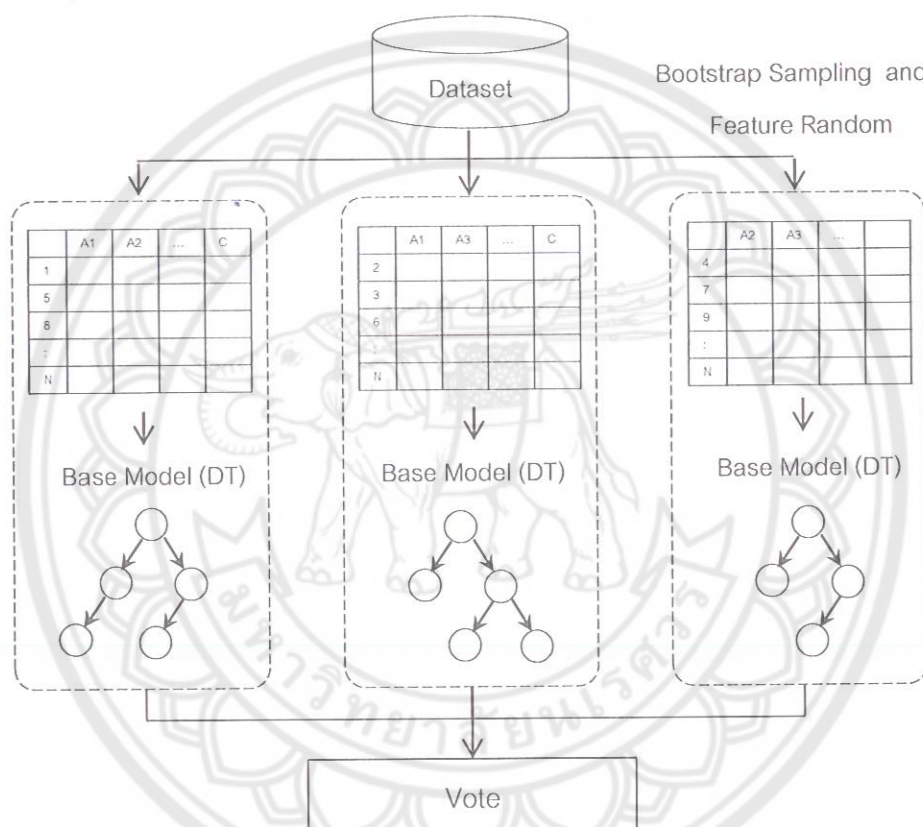
ผลลัพธ์สุดท้ายที่ได้จากกระบวนการจำแนกประเภทข้อมูลด้วยวิธี Boosting แบบ Adaboost สามารถแสดงได้ดังสมการ 22

$$C(x) = \text{Sign}\left(\sum_{i=1}^J \alpha_i c_i(x)\right) \quad (22)$$

โดยที่ α_i คือ ค่าน้ำหนักของข้อมูลทั้งหมดที่คำนวณได้ในแต่ละรอบจากการคำนวณหา ค่าความผิดพลาดในการจำแนกข้อมูล และ C_i คือ ตัวจำแนกประเภทข้อมูล

3. Random Forest Method

Random Forest ถูกพัฒนาขึ้นโดย Breiman, L., (2001) เป็นอัลกอริทึมการเรียนรู้แบบรวมกลุ่มโดยใช้วิธีการสร้างต้นไม้ช่วยตัดสินใจร่วมกันหลาย ๆ แบบจำลองเพื่อช่วยกันตัดสินใจจากการโหวตเสียงข้างมากจากแบบจำลองที่ถูกสร้างขึ้นทั้งหมดเพื่อให้ได้ผลลัพธ์ที่ดีที่สุดเพียงคำตอบเดียว



ภาพ 24 แสดงโครงสร้างวิธีการทำงานแบบ Random Forest

ที่มา: Yang, P., Yang, Y.H., Zhou, B.B. and Zomaya, A.Y., 2010

Random Forest จะมีลักษณะการทำงานคล้ายกับ Bagging โดยใช้วิธีการสุ่มชุดของข้อมูลตัวอย่างด้วยวิธี Bootstrap Sampling แต่จะมีส่วนขยายเพิ่มเติมตรงที่ Random Forest จะทำการสุ่มเลือกคุณลักษณะ (Random Feature) จากชุดตัวอย่างข้อมูลด้วย ทำให้ลดความสัมพันธ์ระหว่างคุณลักษณะลงได้ (Alam, M.S. and Vuong, S.T., 2013)

สำหรับอัลกอริทึมที่ใช้งานร่วมกับ Random Forest นั้นจะใช้พื้นฐานการเรียนรู้ของต้นไม้ช่วยตัดสินใจ (Tree-Base Model) ในการสร้างแบบจำลองหลาย ๆ แบบจำลอง ซึ่งจะช่วยในการปรับปรุงประสิทธิภาพให้สูงขึ้นถ้าเปรียบเทียบกับการใช้ต้นไม้ช่วยตัดสินใจเพียงหนึ่งแบบจำลอง (Ali, J., Khan, R., Ahmad, N. and Maqsood, I., 2012) สามารถแสดงวิธีการของ Random Forest ดังภาพ 24

มาตรวัดและการทดสอบประสิทธิภาพแบบจำลอง

1. ค่าที่ใช้สำหรับวัดประสิทธิภาพแบบจำลอง

งานวิจัยนี้ใช้ค่าความถูกต้อง (Accuracy) ในการพิจารณาประสิทธิภาพของแบบจำลอง เพื่อใช้ในการเปรียบเทียบแบบจำลองที่สร้างขึ้นทั้งหมด โดยค่าความถูกต้องดังกล่าวสามารถคำนวณได้จากสมการ 23 (Mittal, P. and Gill, N.S., 2014)

$$\text{Accuracy} = \frac{\text{Number of Correctly}}{\text{Total Number of Objects}} \times 100 \quad (23)$$

โดยที่ Number of Correctly คือ จำนวนข้อมูลตัวอย่างที่ทำการจำแนกประเภทชนิดของกลุ่มเมฆได้ถูกต้องทั้งหมด ส่วน Total Number of Objects คือ จำนวนข้อมูลตัวอย่างของกลุ่มเมฆทั้งหมด

2. การแบ่งข้อมูลเพื่อวัดประสิทธิภาพแบบจำลอง

การแบ่งข้อมูลเพื่อใช้สำหรับวัดประสิทธิภาพแบบจำลองนั้นมีหลายวิธีการ สำหรับในวิทยานิพนธ์นี้ ใช้วิธีการแบบ Cross-Validation Test ซึ่งเป็นวิธีที่ได้รับความนิยมสำหรับการแบ่งข้อมูลเพื่อวัดประสิทธิภาพแบบจำลอง เนื่องจากผลลัพธ์ที่ได้มีความน่าเชื่อถือ โดยหลักในการแบ่งข้อมูลด้วยวิธีนี้จะเริ่มจากการกำหนดค่า K หรือการแบ่งข้อมูลออกเป็น K ส่วนเท่า ๆ กัน ตัวอย่างเช่น กำหนดให้ K = 5 (5 Fold Cross-Validation) ข้อมูลจะถูกแบ่งออกเป็น 5 ส่วน โดยในแต่ละส่วนจะมีจำนวนข้อมูลเท่า ๆ กัน จากนั้นจะใช้ข้อมูล 4 ส่วนทำการเรียนรู้ และข้อมูลอีก 1 ส่วนจะถูกใช้ในการทดสอบประสิทธิภาพสลับกันไปจนครบทุกชุดข้อมูล ดังภาพ 25

รอบที่ 1	ข้อมูลชุดที่ 1	ข้อมูลชุดที่ 2	ข้อมูลชุดที่ 3	ข้อมูลชุดที่ 4	ข้อมูลชุดที่ 5
รอบที่ 2	ข้อมูลชุดที่ 1	ข้อมูลชุดที่ 2	ข้อมูลชุดที่ 3	ข้อมูลชุดที่ 4	ข้อมูลชุดที่ 5
รอบที่ 3	ข้อมูลชุดที่ 1	ข้อมูลชุดที่ 2	ข้อมูลชุดที่ 3	ข้อมูลชุดที่ 4	ข้อมูลชุดที่ 5
รอบที่ 4	ข้อมูลชุดที่ 1	ข้อมูลชุดที่ 2	ข้อมูลชุดที่ 3	ข้อมูลชุดที่ 4	ข้อมูลชุดที่ 5
รอบที่ 5	ข้อมูลชุดที่ 1	ข้อมูลชุดที่ 2	ข้อมูลชุดที่ 3	ข้อมูลชุดที่ 4	ข้อมูลชุดที่ 5

☐ Training Data ☐ Testing Data

ภาพ 25 แสดงการแบ่งข้อมูลเพื่อวัดประสิทธิภาพแบบจำลอง

สามารถอธิบายลักษณะการแบ่งข้อมูลเพื่อทดสอบประสิทธิภาพของแบบจำลองได้ดังตัวอย่างเช่น มีข้อมูลตัวอย่างทั้งหมดจำนวน 100 ตัวอย่าง ข้อมูลจะถูกแบ่งออกเป็น 5 ชุด โดยในแต่ละชุดจะมีข้อมูลทั้งหมด 20 ตัวอย่าง หลังจากนั้นทำการทดสอบประสิทธิภาพดังนี้

รอบที่ 1 นำข้อมูลชุดที่ 2,3,4,5 เรียนรู้เพื่อสร้างแบบจำลอง และใช้ข้อมูลชุดที่ 1 ทดสอบ

รอบที่ 2 นำข้อมูลชุดที่ 1,3,4,5 เรียนรู้เพื่อสร้างแบบจำลอง และใช้ข้อมูลชุดที่ 2 ทดสอบ

รอบที่ 3 นำข้อมูลชุดที่ 1,2,4,5 เรียนรู้เพื่อสร้างแบบจำลอง และใช้ข้อมูลชุดที่ 3 ทดสอบ

รอบที่ 4 นำข้อมูลชุดที่ 1,2,3,5 เรียนรู้เพื่อสร้างแบบจำลอง และใช้ข้อมูลชุดที่ 4 ทดสอบ

รอบที่ 5 นำข้อมูลชุดที่ 1,2,3,4 เรียนรู้เพื่อสร้างแบบจำลอง และใช้ข้อมูลชุดที่ 5 ทดสอบ

จากตัวอย่างจะได้ประสิทธิภาพของแบบจำลองทั้งหมด 5 แบบจำลอง จากนั้นทำการหาค่าเฉลี่ยความถูกต้องของแบบจำลองทั้ง 5 ดังนั้นวิธีการนี้จึงมีความน่าเชื่อถือสูงสำหรับการวัดประสิทธิภาพแบบจำลองเนื่องจากข้อมูลทุกตัวจะถูกใช้ในการเรียนรู้และการทดสอบประสิทธิภาพทำให้เกิดความเอนเอียงของข้อมูล และยังสามารถช่วยหลีกเลี่ยงปัญหาของการ Overfitting ให้น้อยลงได้

งานวิจัยที่เกี่ยวข้อง

เนื่องจากสภาพอากาศมีส่วนสำคัญมากต่อการดำรงชีวิตในปัจจุบันไม่ว่าจะเป็นการเกิดลักษณะของสภาพอากาศแปรปรวน การเกิดพายุร้อนพายุผ่า รวมถึงการเกิดฝนตก ซึ่งเมฆเป็นปัจจัยหนึ่งที่มีบทบาทสำคัญต่อการพิจารณาของสาเหตุเหล่านั้นได้เป็นอย่างดี จากอดีตจนถึงปัจจุบันมีนักวิจัยจำนวนมากให้ความสนใจเกี่ยวกับวิธีการจำแนกประเภทของเมฆ เนื่องจากสามารถนำเอาข้อมูลสารสนเทศที่ได้จากการจำแนกประเภทของเมฆไปใช้ประโยชน์ในการวิเคราะห์ลักษณะของสภาพภูมิอากาศได้ และทำให้เกิดประโยชน์ต่อหลาย ๆ หน่วยงาน โดยเฉพาะในหน่วยงานที่เกี่ยวข้องกับทางด้านอุตุนิยมวิทยา ซึ่งวิธีการจำแนกประเภทของเมฆนั้นมีผู้วิจัยจำนวนมากได้นำเสนอเทคนิค แนวคิด และทฤษฎี ที่ใช้การจำแนกประเภทของเมฆดังนี้

1. การจำแนกประเภทของเมฆจากภาพถ่ายดาวเทียม GOES-8 โดยทำการแบ่งเมฆออกเป็น 6 ชนิดร่วมกับลักษณะพื้นดินและพื้นน้ำอีก 4 ชนิด และใช้ค่า Singular Value Decomposition (SVD), Wavelet Packets (WP) ซึ่งเป็นการดึงเอาคุณลักษณะเด่นจากภาพถ่าย รวมถึงคุณลักษณะของ Gray-Level Co-occurrence Matrix (GLCM) ที่นำมาใช้ในการวิเคราะห์ สำหรับขั้นตอนการจำแนกประเภทของเมฆได้ทำการเปรียบเทียบระหว่างวิธีการ Artificial Neural Networks ทั้งหมด 2 รูปแบบคือ Probability Neural Network (PNN) และ Self-Organized feature Map (SOM) ผลการทดลองพบว่า การใช้อัลกอริทึม PNN ให้ประสิทธิภาพดีที่สุดในการจำแนกประเภทของเมฆ (Tian, B., et al., 1999)

2. การศึกษาการจำแนกเมฆจากภาพถ่ายดาวเทียม METEOSAT 7 โดยแบ่งประเภทของเมฆออกเป็น 6 ชนิดและใช้วิธีการดึงคุณลักษณะของเมฆจากภาพถ่ายแบบ Texture Feature ได้แก่ Statistical Features (SF), Spatial Gray-Level Dependence Matrices (SGLDM), Gray-Level Difference Statistics (GLDS), Neighborhood Gray Tone Difference Matrix (NGTDM), Statistical Feature Matrix (SFM), Laws Texture Energy Measures (TEM), Fractal Dimension Texture Analysis (FDTA) และ Fourier Power Spectrum (FPS) รวมทั้งหมด 55 คุณลักษณะ สำหรับขั้นตอนการจำแนกประเภทของเมฆใช้ Self-Organizing Feature Map (SOFM) และ K-nearest neighbor (kNN) ผลการทดลองพบว่า วิธีการ kNN ให้ประสิทธิภาพการจำแนกดีกว่า SOFM และพบว่าคุณลักษณะของ SGLDM, NGTDM, GLDS, SF เป็นคุณลักษณะที่ดีต่อการจำแนกประเภทของเมฆ (Christodoulos, I., et al., 2003)

3. การจำแนกประเภทของเมฆจากภาพถ่ายดาวเทียม Kalpana - 1 โดยใช้วิธีการดึงคุณลักษณะของ Texture Feature ในรูปภาพแบบ Gabor Transform และใช้ Support Vector

Machine (SVM) เรียนรู้ข้อมูลในการจำแนกประเภทของเมฆ ผลการทดลองพบว่า ความแม่นยำในการจำแนกประเภทของเมฆสูงขึ้นจากเดิมเฉลี่ยอยู่ที่ 92.30% (Chethan, H.K., Raghavendra, R. and Kumar, G.H., 2009)

4. การปรับปรุงวิธีการจำแนกประเภทของเมฆ โดยใช้เทคนิค Current window-based Clustering จากดาวเทียมอุตุนิยมวิทยา FengYun-2C (FY-2C) ของจีน โดยแบ่งเมฆเป็น 6 ประเภท พื้นดินและพื้นน้ำอีก 2 ประเภท สำหรับขั้นตอนการจำแนก ได้ทำการเปรียบเทียบวิธีการแบบ Artificial Neural Networks ทั้งหมด 6 แบบ กับ PCA และ SVM ซึ่งพบว่า Artificial Neural Networks ทั้งหมด 6 แบบและ PCA ให้ประสิทธิภาพสูงกว่า SVM เฉลี่ย 10% และ SOM กับ PNN ให้ประสิทธิภาพดีที่สุดจากผลการทดลอง (Liu, Y., Xia, J., Shi, C.X. and Hong, Y., 2009)

5. การจำแนกประเภทของเมฆจากสภาพอากาศโดยใช้ภาพถ่ายดาวเทียม ได้นำเสนอวิธีการใหม่ คือ Discrete Wavelet Transform ในการสกัดคุณลักษณะของเมฆเพื่อปรับปรุงประสิทธิภาพในการจำแนกประเภทของเมฆ ซึ่งได้ทำการแบ่งเมฆออกเป็น 3 ชนิด ได้แก่ เมฆชั้นต่ำ (Low), เมฆชั้นกลาง (Medium) และเมฆชั้นสูง (High) โดยในงานวิจัยได้ทำการเปรียบเทียบอัลกอริทึมที่ใช้ในการจำแนกประเภทอีก 3 แบบ ผลการทดลองพบว่า วิธีการที่นำเสนอใหม่นี้ให้ประสิทธิภาพดีที่สุดเมื่อเปรียบเทียบกับอัลกอริทึมอื่น ๆ เฉลี่ยอยู่ที่ 94.37% (Kuril, S., Saini, I. and Saini, B. S., 2013)

6. ในปี 2014 ได้มีการนำเสนอเทคนิคใหม่ ๆ คือ CCSI-ODSR เพื่อแก้ไขปัญหาในการจำแนกประเภทของเมฆจากภาพถ่ายดาวเทียม โดยแบ่งประเภทของเมฆออกเป็น 4 ประเภท ร่วมกับพื้นดินและน้ำอีก 2 ประเภท สำหรับในขั้นตอนการเรียนรู้ได้ทำการเปรียบเทียบ 3 อัลกอริทึม ได้แก่ SVM, SRC และ CCSI-ODSR จากผลการทดลองแสดงให้เห็นว่าอัลกอริทึม CCSI-ODSR ที่ได้นำเสนอขึ้นมาใหม่ ให้ประสิทธิภาพในการจำแนกประเภทของเมฆโดยเฉลี่ยสูงกว่าอัลกอริทึม SVM และ SRC (Jin W., Wang, L., Zeng, X., Liu, Z. and Fu, R., 2014)

7. การจำแนกประเภทของเมฆจากภาพถ่ายท้องฟ้าที่มีอิทธิพลต่อพลังงานแสงอาทิตย์ Photovoltaic (PV) โดยแบ่งประเภทของเมฆออกเป็น 4 ประเภท ซึ่งได้ทำการดึงคุณลักษณะของเมฆด้วยวิธีการแบบ Spectral Features กับ Textural Features และใช้ SVM สำหรับการจำแนกประเภทของเมฆ ผลการทดลองพบว่าสามารถทำการจำแนกประเภทของเมฆโดยเฉลี่ยค่าความถูกต้องสูงสุดอยู่ที่ 96.44% (Zhen, Z., et al., 2015)

ในปัจจุบันได้มีงานวิจัยของ ถิรนนท์ สอนแก้ว (2558) ทำการจำแนกประเภทของเมฆโดยใช้ค่าอัตราส่วนที่วัดได้จากการแผ่รังสีจากดวงอาทิตย์ (AV) ค่าส่วนเบี่ยงเบนมาตรฐานของดวง

อาทิตย (SD) และค่าอัตราส่วนของเมฆต่อท้องฟ้าทั้งหมด (CC) รวมถึงการนำค่าของข้อมูลทางอุตุนิยมวิทยาที่เกี่ยวข้องกับคุณลักษณะของเมฆ ได้แก่ อุณหภูมิ (T) ความกดอากาศ (P) ความชื้นสัมพัทธ์ (RH) และความเร็วลม (WS) มาทำการวิเคราะห์เพิ่มเติม โดยในงานวิจัยได้แบ่งชนิดของเมฆออกเป็น 8 กลุ่ม และ 3 กลุ่ม สำหรับใช้ในการทดลองตามลำดับ ส่วนขั้นตอนในการจำแนกประเภทชนิดของกลุ่มเมฆให้หลักวิธีการวิเคราะห์ทางสถิติเปรียบเทียบกับโครงข่ายประสาทเทียมหลายชั้น ผลการทดลองแสดงให้เห็นว่า โครงข่ายประสาทเทียมให้ประสิทธิภาพในการจำแนกประเภทชนิดของเมฆดีกว่าวิธีการวิเคราะห์ทางสถิติ จากงานวิจัยนี้พบว่าประสิทธิภาพความถูกต้องและแม่นยำในการจำแนกประเภทชนิดของกลุ่มเมฆยังมีประสิทธิภาพอยู่ในระดับเกณฑ์ที่ต่ำ ดังนั้นในวิทยานิพนธ์นี้จึงได้นำเสนอเทคนิคทฤษฎีการเรียนรู้แบบรวมกลุ่มมาประยุกต์ใช้เพื่อปรับปรุงประสิทธิภาพสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆให้เพิ่มสูงขึ้น



บทที่ 3

วิธีดำเนินการวิจัย

งานวิจัยนี้ได้ทำการเปรียบเทียบระหว่างวิธีการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยว และแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม โดยมุ่งเน้นในเรื่องของการหาประสิทธิภาพ ความถูกต้องและแม่นยำสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆที่ดีที่สุด ซึ่งผู้วิจัยได้ทำการศึกษาแนวคิด ทฤษฎี ตลอดจนเอกสารและงานวิจัยที่เกี่ยวข้องเพื่อเป็นพื้นฐานความรู้สำหรับการทำวิจัยในครั้งนี้ โดยได้กำหนดวิธีการดำเนินงานวิจัยซึ่งประกอบด้วยเนื้อหา ดังต่อไปนี้

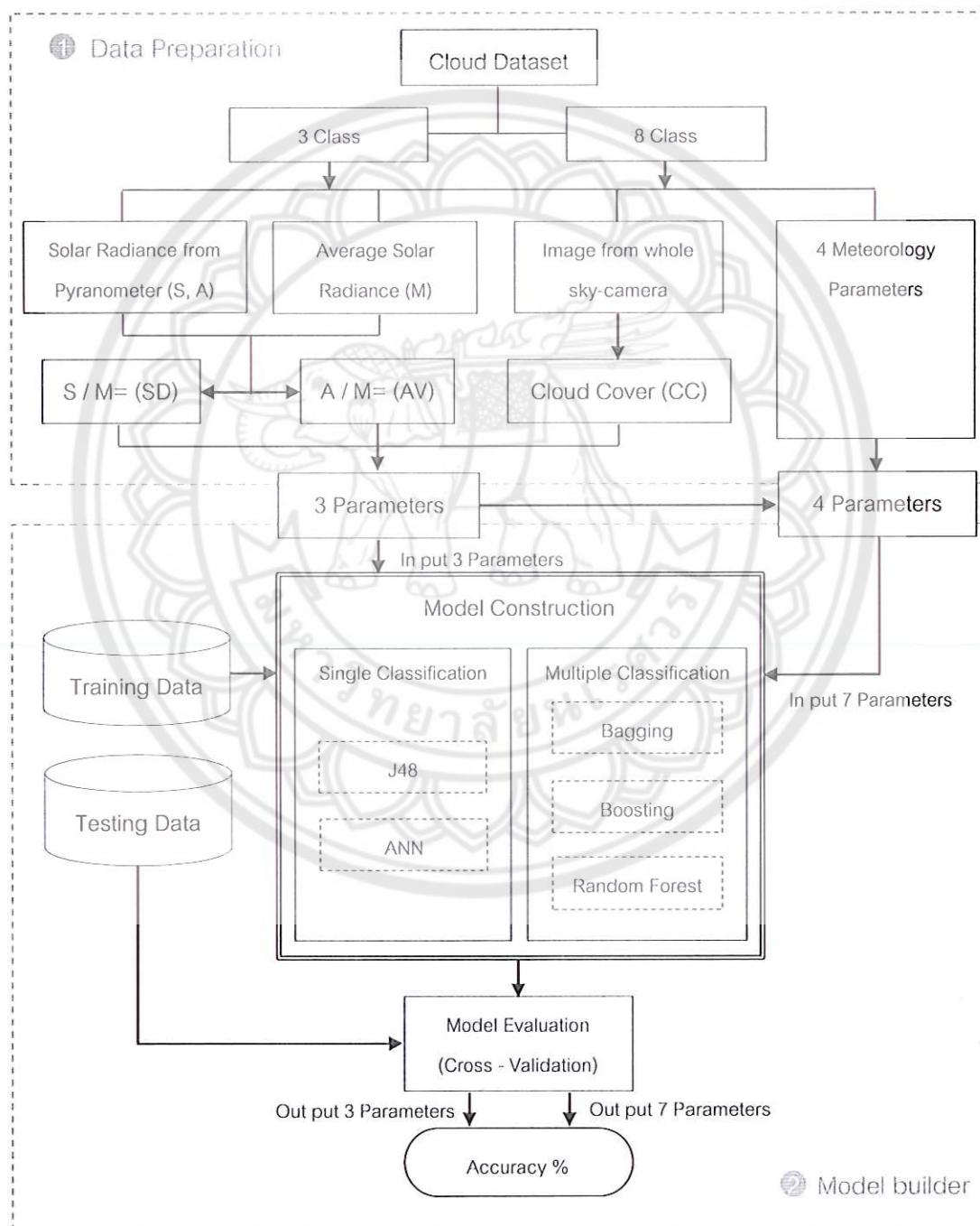
1. เครื่องมือที่ใช้ในการวิจัย
2. กรอบวิธีการดำเนินงานวิจัย
 - 2.1 การเตรียมข้อมูลสำหรับการทดลอง
 - 2.2 การสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยว
 - 2.3 การสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม

เครื่องมือที่ใช้ในการวิจัย

1. ฮาร์ดแวร์ (Hardware)
 - 1.1 คอมพิวเตอร์ Notebook 1 เครื่อง
 - 1.2 CPU Intel(R) Core(TM) i7-4700HQ @ 2.40 GHz
 - 1.3 RAM DDR 3 ขนาด 4 GB, Hard Disk 700 GB
 - 1.4 Graphics Card Intel(R) HD Graphics 4600
2. ซอฟต์แวร์ (Software)
 - 2.1 Windows 8.1 Professional (64-bit)
 - 2.2 RapidMiner Studio Academia 7.1.001
 - 2.3 Weka 7.0.0 (Extension)
 - 2.4 Microsoft Word 2013
 - 2.5 Microsoft Excel 2013

กรอบวิธีการดำเนินงานวิจัย

ในหัวข้อนี้ได้นำเสนอกรอบวิธีการดำเนินงานวิจัย หลักการ รวมไปถึงการออกแบบการทดลอง โดยได้แบ่งการทดลองออกเป็น 2 ส่วนหลัก ๆ คือ ส่วนที่ 1 การเตรียมข้อมูลเพื่อใช้สำหรับการทดลอง ส่วนที่ 2 การสร้างแบบจำลองในการจำแนกประเภท แสดงรายละเอียดในภาพ 26



ภาพ 26 แสดงกรอบวิธีการดำเนินงานวิจัยสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆ

จากขั้นตอนการเตรียมข้อมูลในภาพ 26 ทำการแบ่งกลุ่มของเมฆออกเป็น 2 แบบสำหรับการทดลอง คือ แบ่งกลุ่มของเมฆออกเป็น 8 กลุ่ม และ 3 กลุ่ม ตามที่ได้กล่าวไว้ในบท 2 โดยการทำงานจะเริ่มจาก นำข้อมูลที่ได้เตรียมไว้เข้าสู่กระบวนการเรียนรู้ในการจำแนกประเภทชนิดของกลุ่มเมฆ ซึ่งในขั้นตอนนี้ ได้ใช้อัลกอริทึม J48 และ ANN เรียนรู้ข้อมูลตัวอย่างในแต่ละชุดเพื่อใช้สำหรับการจำแนกประเภทชนิดของกลุ่มเมฆในการสร้างแบบจำลองการจำแนกประเภทแบบเดียว และทำการเปรียบเทียบกับการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม 3 วิธี ได้แก่ Bagging, Boosting และ Random Forest โดยจะใช้อัลกอริทึม J48 และ ANN เป็นอัลกอริทึมพื้นฐานที่ใช้งานร่วมกับวิธีการเรียนรู้แบบรวมกลุ่ม 2 วิธี คือ Bagging และ Boosting (สามารถนำอัลกอริทึมอื่น ๆ มาประยุกต์ใช้งานร่วมกับวิธีการของ Bagging และ Boosting ได้) ส่วนวิธีการของ Random Forest จะใช้ต้นไม้ช่วยตัดสินใจซึ่งเป็นอัลกอริทึมพื้นฐานในการเรียนรู้ของตัวอัลกอริทึมเอง (ไม่สามารถนำอัลกอริทึมอื่น ๆ มาประยุกต์ใช้งานร่วมกับวิธีการของ Random Forest ได้) เมื่อได้แบบจำลองในแต่ละวิธีจากนั้นจะใช้วิธีการ Cross- Validation สำหรับวัดประสิทธิภาพแบบจำลองที่สร้างขึ้นในแต่ละแบบจำลอง เพื่อให้ได้แบบจำลองที่มีความน่าเชื่อถือ และในลำดับสุดท้ายจะดำเนินการเปรียบเทียบผลลัพธ์ของประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆที่ได้จากการทดลองทั้งหมด

1. การเตรียมข้อมูลสำหรับการใช้ในการทดลอง

สำหรับในขั้นตอนนี้ ข้อมูลตัวอย่างเป็นข้อมูลที่ได้รับคำแนะนำมาจากการเก็บรวบรวมข้อมูลของสถาบันวิจัยดาราศาสตร์แห่งชาติ (องค์การมหาชน) โดยได้ทำการแบ่งชุดข้อมูลที่ใช้สำหรับการดำเนินการทดลองออกเป็น 4 กลุ่มตัวอย่าง แสดงในตาราง 5

ตาราง 5 ข้อมูลกลุ่มตัวอย่างการแบ่งกลุ่มของเมฆและคุณลักษณะของเมฆ

กลุ่ม ตัวอย่าง (Dataset)	จำนวน ตัวอย่าง (Instants)	ค่าพารามิเตอร์ (Parameters)		ลักษณะการแบ่งกลุ่มเมฆ (Classes)	
		3 คุณลักษณะ	7 คุณลักษณะ	3 กลุ่ม	8 กลุ่ม
1	7,367	✓			✓
2	7,367	✓		✓	
3	7,367		✓		✓
4	7,367		✓	✓	

จากตาราง 5 ประกอบด้วยกลุ่มตัวอย่างของข้อมูล 4 ชุด โดยแต่ละกลุ่มตัวอย่างจะประกอบด้วยจำนวนข้อมูล 7,367 ตัวอย่าง และแบ่งประเภทของเมฆออกเป็น 8 กลุ่ม และ 3 กลุ่ม โดยใช้ 3 คุณลักษณะ และ 7 คุณลักษณะสำหรับการวิเคราะห์ข้อมูลในการจำแนกประเภทชนิดของกลุ่มเมฆ ซึ่งจำนวนตัวอย่างชนิดของเมฆแต่ละกลุ่มนั้นแสดงได้ในตาราง 6 และ 7

ตาราง 6 จำนวนข้อมูลตัวอย่างของกลุ่มเมฆ 8 กลุ่ม สำหรับการวิเคราะห์ 3 คุณลักษณะ และ 7 คุณลักษณะ

สัญลักษณ์ (Symbol)	ชนิดของกลุ่มเมฆ (Cloud Type)	จำนวนตัวอย่าง (Instants)
C1	Cu	985
C2	Ci, Cs	663
C3	Cc, Ac	201
C4	Clear sky	1,291
C5	Sc	441
C6	St, As	1,625
C7	Cb, Ns	1,898
C8	Fog	263

ตาราง 7 จำนวนข้อมูลตัวอย่างของกลุ่มเมฆ 3 กลุ่ม สำหรับการวิเคราะห์ 3 คุณลักษณะ และ 7 คุณลักษณะ

สัญลักษณ์ (Symbol)	ชนิดของกลุ่มเมฆ (Cloud Type)	จำนวนตัวอย่าง (Instants)
C1	Cu, Cc , Ac	1,186
C2	Clear sky, Ci and Cs	1,954
C3	Sc, St, As, Cb, Ns and Fog	4,227

จากตาราง 6 และ 7 แสดงถึงจำนวนข้อมูลตัวอย่างชนิดของเมฆแต่ละกลุ่ม จากข้อมูลทั้งหมด 4 กลุ่มตัวอย่างที่ใช้ในการทดลอง สำหรับค่าของประเภทข้อมูลที่นำมาวิเคราะห์ทั้ง 4 กลุ่มตัวอย่างนี้ จะมีค่าของข้อมูลเป็นตัวเลขต่อเนื่องทั้งหมด ตามที่ได้แสดงรายละเอียดในตาราง 8-11

ตาราง 8 ข้อมูลตัวอย่างคุณลักษณะของเมฆ โดยวิเคราะห์ 3 คุณลักษณะ 8 กลุ่ม

Ex	AV	SD	CC	Class
1	0.18	0.039	58.117	C6
2	0.77	0.349	21.390	C4
3	0.73	0.307	63.131	C8
4	0.73	0.307	34.020	C4
5	0.85	0.135	19.675	C4
6	0.84	0.138	34.678	C8
7	0.34	0.137	58.581	C1
8	0.85	0.143	56.667	C8
9	0.76	0.195	85.572	C5
:	:	:	:	:
7,367	0.39	0.365	93.202	C6

ตาราง 9 ข้อมูลตัวอย่างคุณลักษณะของเมฆ โดยวิเคราะห์ 7 คุณลักษณะ 8 กลุ่ม

Ex	AV	SD	CC	T	P	RH	WS	Class
1	0.04	0.007	85.947	13.6	100.0	998.6	3.6	C7
2	0.04	0.010	79.115	13.7	100.0	1000.4	4.0	C7
3	1.02	0.227	70.746	17.1	83.0	1003.5	0.0	C1
4	0.78	0.391	70.496	16.9	86.0	1003.4	0.0	C6
5	0.46	0.074	90.989	15.6	91.0	1003.4	0.0	C6
6	0.48	0.059	79.378	15.3	91.0	1003.6	0.0	C1
7	1.04	0.049	21.842	14.6	93.0	1007.3	0.0	C2
8	0.90	0.154	78.017	16.7	90.0	1009.6	0.9	C3
9	0.83	0.228	78.570	17.0	88.0	1009.4	0.9	C5
:	:	:	:	:	:	:	:	:
7,367	0.46	0.156	93.837	12.7	98.0	1010.1	0.9	C8

ตาราง 10 ข้อมูลตัวอย่างคุณลักษณะของเมฆ โดยวิเคราะห์ 3 คุณลักษณะ 3 กลุ่ม

Ex	AV	SD	CC	Class
1	0.27	0.050	88.848	C3
2	0.85	0.135	19.675	C2
3	0.29	0.060	90.954	C1
4	0.36	0.175	81.554	C1
5	0.77	0.349	21.390	C2
6	0.27	0.050	88.848	C3
7	0.29	0.067	89.208	C3
8	0.29	0.060	90.954	C1
9	0.36	0.170	81.554	C1
:	:	:	:	:
7,367	1.14	0.018	44.140	C2

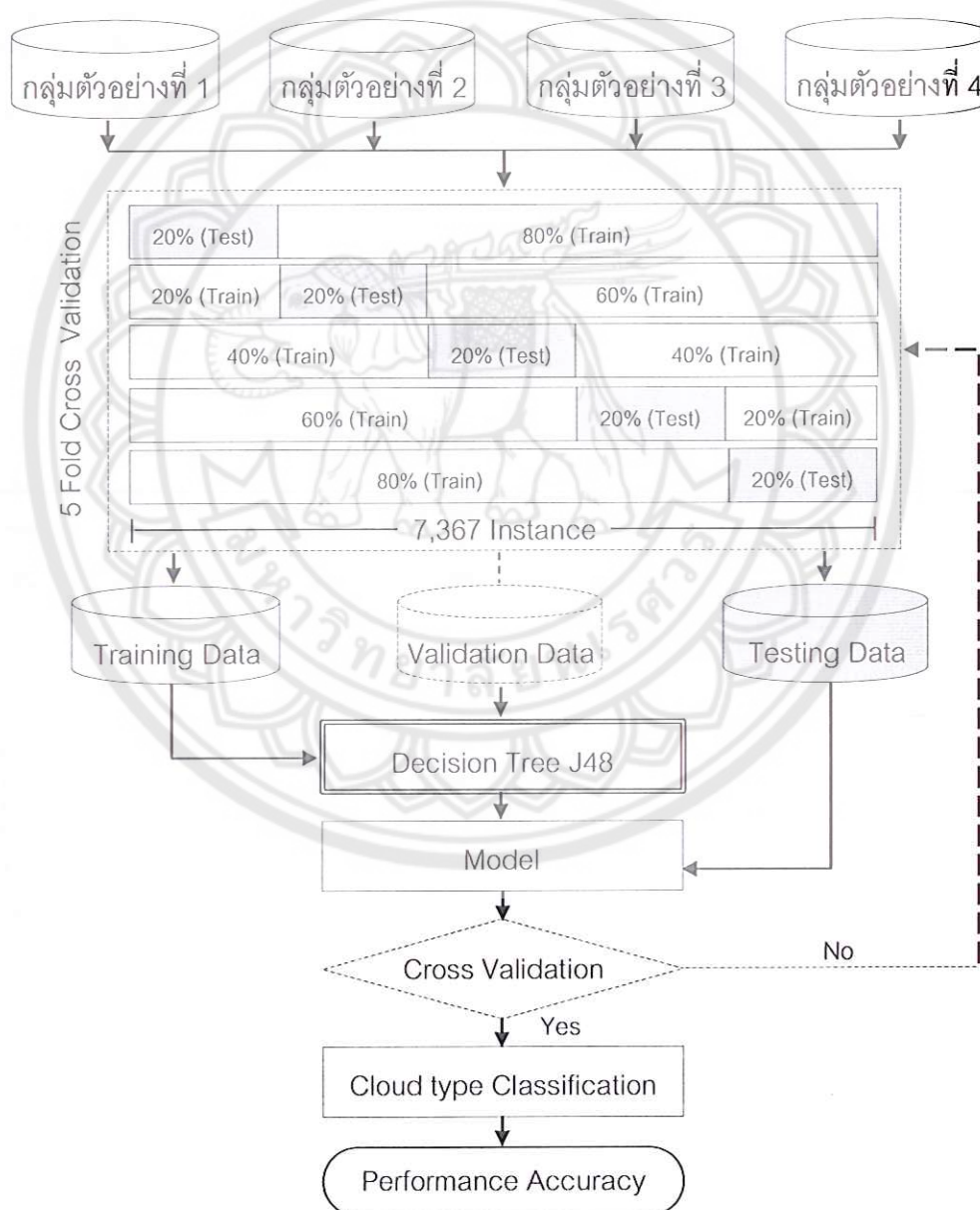
ตาราง 11 ข้อมูลตัวอย่างคุณลักษณะของเมฆ โดยวิเคราะห์ 7 คุณลักษณะ 3 กลุ่ม

Ex	AV	SD	CC	T	P	RH	WS	Class
1	0.21	0.031	88.848	12.7	96.0	1016.8	0.0	C3
2	1.19	0.007	45.676	16.4	44.0	1016.8	0.4	C1
3	0.19	0.053	88.678	13.0	96.0	1016.8	0.0	C3
4	1.25	0.026	38.602	16.0	39.0	1016.8	0.4	C2
5	1.18	0.009	47.013	16.9	40.0	1016.7	0.0	C1
6	0.98	0.001	43.927	15.2	30.0	1011.1	0.0	C2
7	0.98	0.001	43.930	15.4	31.0	1010.9	0.0	C2
8	0.96	0.242	72.730	14.0	71.0	1010.7	0.0	C1
9	1.04	0.164	71.050	14.7	73.0	1010.5	0.0	C1
:	:	:	:	:	:	:	:	:
7,367	0.20	0.079	61.233	13.8	96.0	1007.6	1.3	C3

2. การสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยว

ในหัวข้อนี้จะแสดงถึงขั้นตอนการใช้อัลกอริทึม J48 และ ANN ในการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยว ในการเรียนรู้กลุ่มตัวอย่างในแต่ละชุดที่ได้เตรียมไว้เพื่อหาประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆ สามารถแสดงรายละเอียดดังต่อไปนี้

2.1 การสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยวโดยใช้อัลกอริทึม J48



ภาพ 27 แสดงขั้นตอนการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยวโดยใช้ J48

ภาพ 27 เป็นขั้นตอนการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยวโดยใช้อัลกอริทึม J48 ซึ่งสามารถอธิบายเป็นลำดับขั้นตอนดังต่อไปนี้

2.1.1 เริ่มต้นจากการนำชุดข้อมูลแต่ละชุดที่เตรียมไว้ คือ กลุ่มตัวอย่างที่ 1, 2, 3 และ 4 ตามลำดับเข้าสู่กระบวนการเรียนรู้ที่ละชุด เพื่อหาประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆ

2.1.2 ทำการแบ่งข้อมูลสำหรับทดสอบประสิทธิภาพแบบจำลองที่สร้างขึ้น ในขั้นตอนนี้จะทำการแบ่งข้อมูลเพื่อใช้สำหรับการทดสอบด้วยวิธีการแบบ 5 Fold - Cross Validation โดยที่ชุดข้อมูลของแต่ละกลุ่มตัวอย่างจะประกอบไปด้วยข้อมูลตัวอย่างที่ใช้ในการทดลองจำนวน 7,376 ตัวอย่าง ซึ่งจำนวนตัวอย่างทั้งหมดนี้จะถูกแบ่งออกเป็นส่วน ๆ เท่ากัน ข้อมูลทุก ๆ ส่วนจะถูกเรียนรู้และทดสอบในการวัดประสิทธิภาพของแบบจำลองที่สร้างขึ้น

2.1.3 ทำการเรียนรู้และทดสอบชุดข้อมูลกลุ่มตัวอย่างในแต่ละกลุ่มด้วยอัลกอริทึม J48 สำหรับการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยว ขั้นตอนนี้ได้ทำการกำหนดค่าพารามิเตอร์ให้กับอัลกอริทึม J48 และปรับค่าพารามิเตอร์ให้เหมาะสมที่สุดสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆ โดยพารามิเตอร์ดังกล่าวได้อธิบายไว้ในตาราง 12

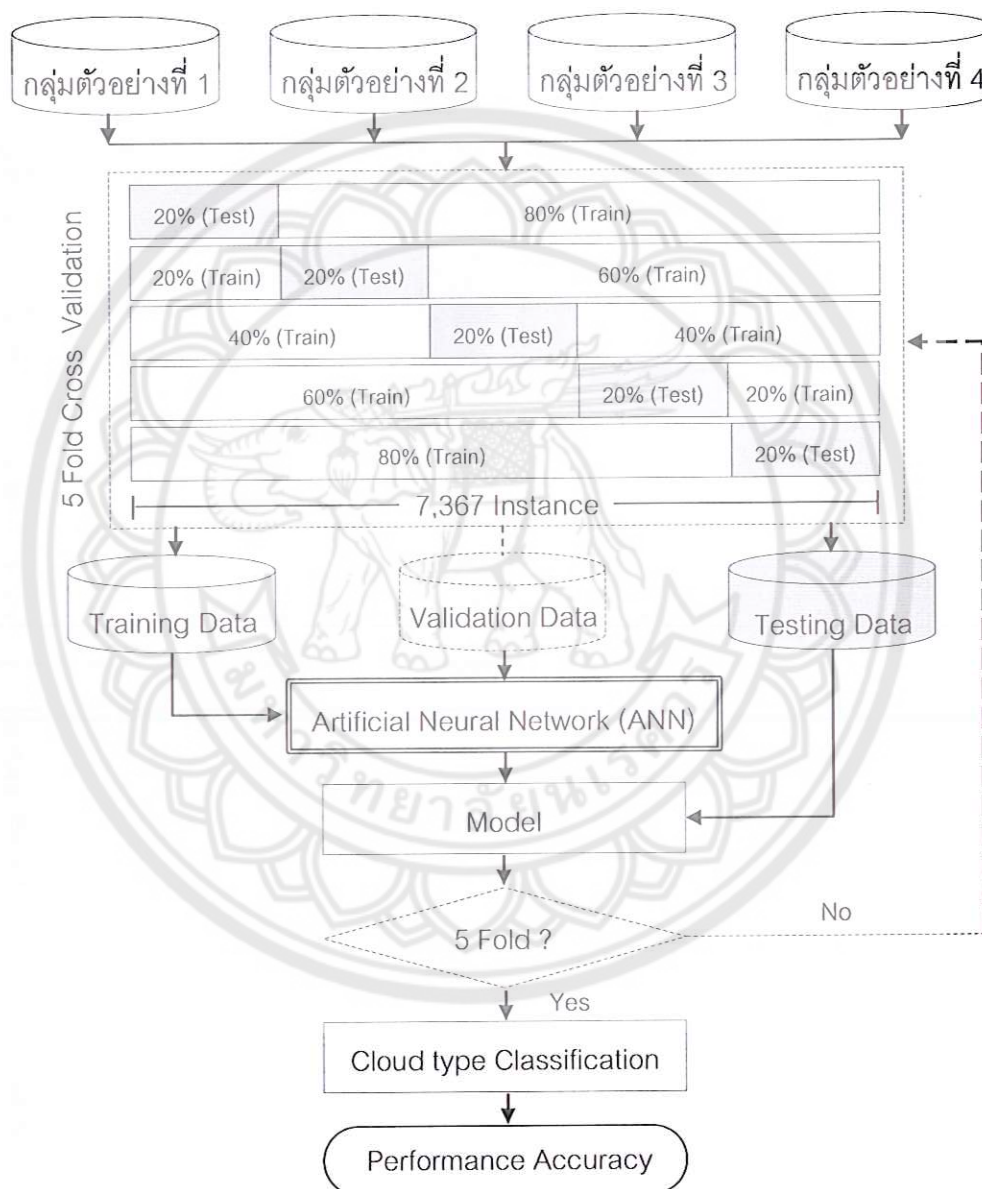
ตาราง 12 พารามิเตอร์ที่กำหนดของอัลกอริทึม J48

พารามิเตอร์	ความหมาย
Confidence Factor	ค่าความเชื่อมั่นหรือค่าที่ใช้สำหรับพิจารณาการตัดกิ่งของต้นไม้ช่วยตัดสินใจ J48
Minimum Number	จำนวนข้อมูลตัวอย่างขั้นต่ำที่เกิดขึ้นได้ในโหนดใบของต้นไม้ช่วยตัดสินใจ J48

2.1.4 ตรวจสอบเงื่อนไขการทำ Cross Validation ผลลัพธ์ของการทำ 5 Fold - Cross Validation จะได้แบบจำลองมาทั้งหมด 5 แบบจำลอง จากนั้นจะทำการหาค่าเฉลี่ยของแบบจำลองที่สร้างขึ้นทั้งหมดเพื่อให้ได้ประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆ

2.1.5 ผลลัพธ์สุดท้ายของกระบวนการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยวโดยใช้อัลกอริทึม J48 จะได้ประสิทธิภาพความถูกต้องสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆในแต่ละกลุ่มตัวอย่าง

2.2 การสร้างแบบจำลองโดยใช้อัลกอริทึม ANN



ภาพ 28 แสดงขั้นตอนการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยวโดยใช้ ANN

ภาพ 28 เป็นขั้นตอนการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยวด้วยอัลกอริทึม ANN ซึ่งสามารถอธิบายเป็นลำดับขั้นตอนดังต่อไปนี้

2.2.1 เริ่มต้นจากการนำชุดข้อมูลแต่ละชุดที่เตรียมไว้ คือ กลุ่มตัวอย่างที่ 1, 2, 3 และ 4 ตามลำดับเข้าสู่กระบวนการเรียนรู้ที่ละชุด เพื่อหาประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆ

2.2.2 ทำการแบ่งข้อมูลสำหรับทดสอบประสิทธิภาพแบบจำลองที่สร้างขึ้น ในขั้นตอนนี้จะทำการแบ่งข้อมูลเพื่อใช้สำหรับการทดสอบด้วยวิธีการแบบ 5 Fold - Cross Validation โดยที่ชุดข้อมูลของแต่ละกลุ่มตัวอย่างจะประกอบไปด้วยข้อมูลตัวอย่างที่ใช้ในการทดลองจำนวน 7,376 ตัวอย่าง ซึ่งจำนวนตัวอย่างทั้งหมดนี้จะถูกแบ่งออกเป็นส่วน ๆ เท่ากัน ข้อมูลทุก ๆ ส่วนจะถูกเรียนรู้และทดสอบในการวัดประสิทธิภาพของแบบจำลองที่สร้างขึ้น

2.2.3 ทำการเรียนรู้และทดสอบชุดข้อมูลกลุ่มตัวอย่างในแต่ละกลุ่มด้วยอัลกอริทึม ANN สำหรับการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยว ขั้นตอนนี้ได้ทำการกำหนดค่าพารามิเตอร์ให้กับอัลกอริทึม ANN และปรับค่าพารามิเตอร์ให้เหมาะสมที่สุดสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆ โดยพารามิเตอร์ดังกล่าวได้อธิบายไว้ในตาราง 13

ตาราง 13 พารามิเตอร์ที่กำหนดของอัลกอริทึม ANN

พารามิเตอร์	ความหมาย
Training cycles	จำนวนรอบที่ใช้ในการเรียนรู้ชุดข้อมูลตัวอย่าง
Hidden Node	จำนวนโหนดในชั้นซ่อน
Learning rate	อัตราการเรียนรู้
Momentum	ค่าโมเมนตัม

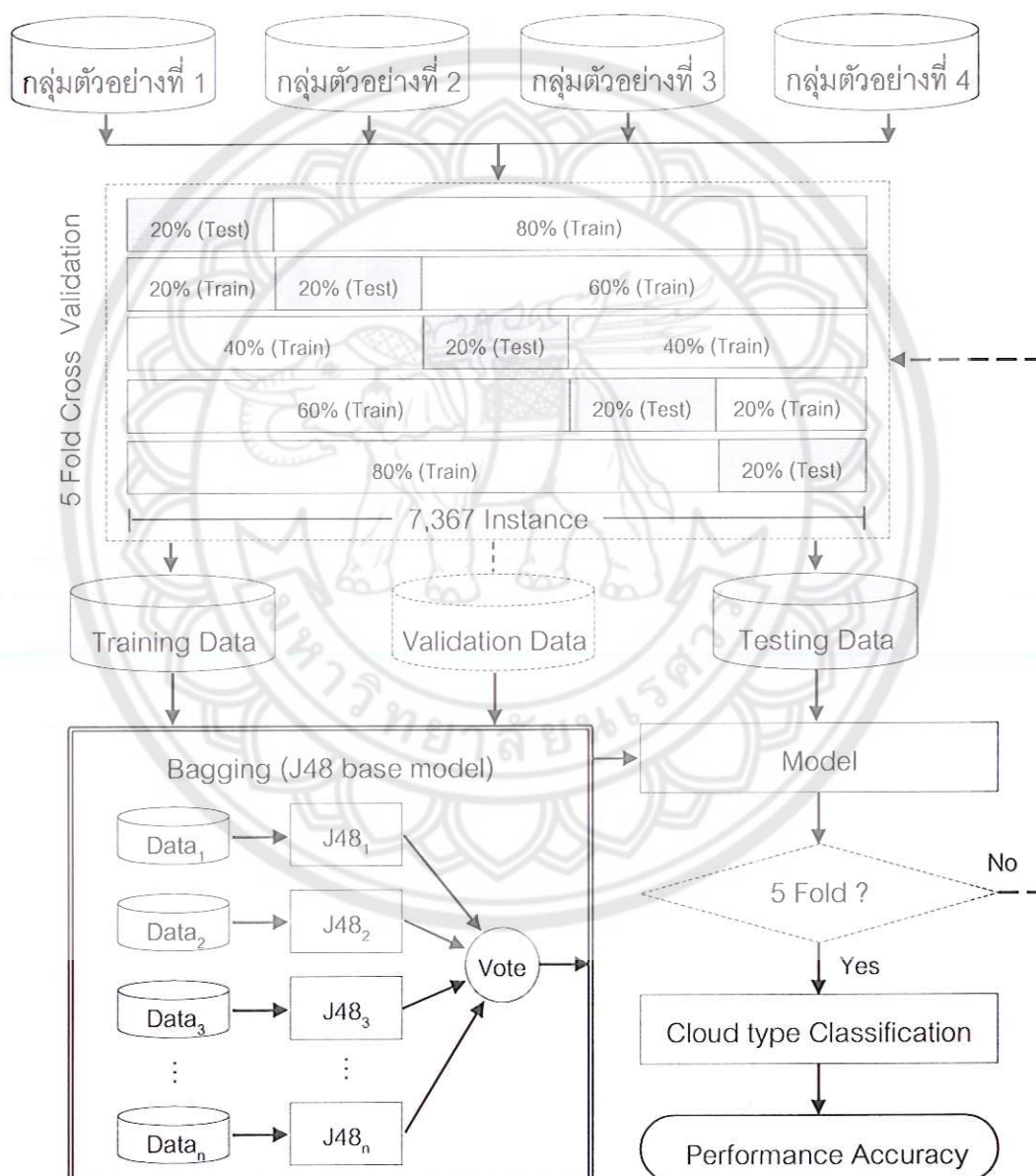
2.2.4 ตรวจสอบเงื่อนไขการทำ Cross Validation ผลลัพธ์ของการทำ 5 Fold - Cross Validation จะได้แบบจำลองมาทั้งหมด 5 แบบจำลอง จากนั้นจะทำการหาค่าเฉลี่ยของแบบจำลองที่สร้างขึ้นทั้งหมดเพื่อให้ได้ประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆ

2.2.5 ผลลัพธ์สุดท้ายจะได้ประสิทธิภาพความถูกต้องสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆโดยใช้อัลกอริทึม ANN ของแต่ละกลุ่มตัวอย่าง

3. การสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม

ในหัวข้อนี้จะกล่าวถึงขั้นตอนวิธีการสร้างแบบจำลองการเรียนรู้แบบรวมกลุ่ม 3 วิธี ได้แก่ Bagging, Boosting และ Random Forest รวมไปถึงการกำหนดค่าพารามิเตอร์ของแต่ละวิธีการในการสร้างแบบจำลองที่ใช้ในการทดลอง สามารถแสดงรายละเอียดได้ดังต่อไปนี้

3.1 การสร้างแบบจำลอง Bagging โดยใช้ J48 เป็นอัลกอริทึมพื้นฐาน



ภาพ 29 แสดงขั้นตอนการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่มด้วยวิธี Bagging โดยใช้ J48 เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้

ภาพ 29 เป็นขั้นตอนการสร้างแบบจำลองการจำแนกประเภทการเรียนรู้แบบรวมกลุ่มด้วยวิธี Bagging โดยใช้ J48 เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ สามารถอธิบายเป็นลำดับขั้นตอนได้ดังต่อไปนี้

3.1.1 เริ่มต้นจากการนำชุดข้อมูลของแต่ละชุดที่เตรียมไว้ คือ กลุ่มตัวอย่างที่ 1, 2, 3 และ 4 ตามลำดับเข้าสู่กระบวนการเรียนรู้ทีละชุด ในการสร้างแบบจำลองการจำแนกประเภทชนิดของกลุ่มเมฆ และหาประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆทั้ง 4 กลุ่มตัวอย่าง

3.1.2 ทำการแบ่งข้อมูลสำหรับทดสอบประสิทธิภาพแบบจำลองที่สร้างขึ้น ในขั้นตอนนี้จะทำการแบ่งข้อมูลเพื่อใช้สำหรับการทดสอบด้วยวิธีการแบบ 5 Fold - Cross Validation โดยที่ชุดข้อมูลของแต่ละกลุ่มตัวอย่างจะประกอบไปด้วยข้อมูลตัวอย่างที่ใช้ในการทดลองจำนวน 7,376 ตัวอย่าง ซึ่งจำนวนตัวอย่างทั้งหมดนี้จะถูกแบ่งออกเป็น ส่วน ๆ เท่ากัน ข้อมูลทุก ๆ ส่วนจะถูกเรียนรู้และทดสอบในการวัดประสิทธิภาพของแบบจำลองที่สร้างขึ้น

3.1.3 ทำการเรียนรู้และทดสอบชุดข้อมูลกลุ่มตัวอย่างในแต่ละกลุ่มด้วยวิธี Bagging โดยใช้ J48 เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ในการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม ขั้นตอนนี้ได้ทำการกำหนดค่าพารามิเตอร์ให้กับวิธี Bagging และปรับค่าพารามิเตอร์ให้เหมาะสมที่สุดสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆ โดยพารามิเตอร์ดังกล่าวได้อธิบายไว้ในตาราง 14

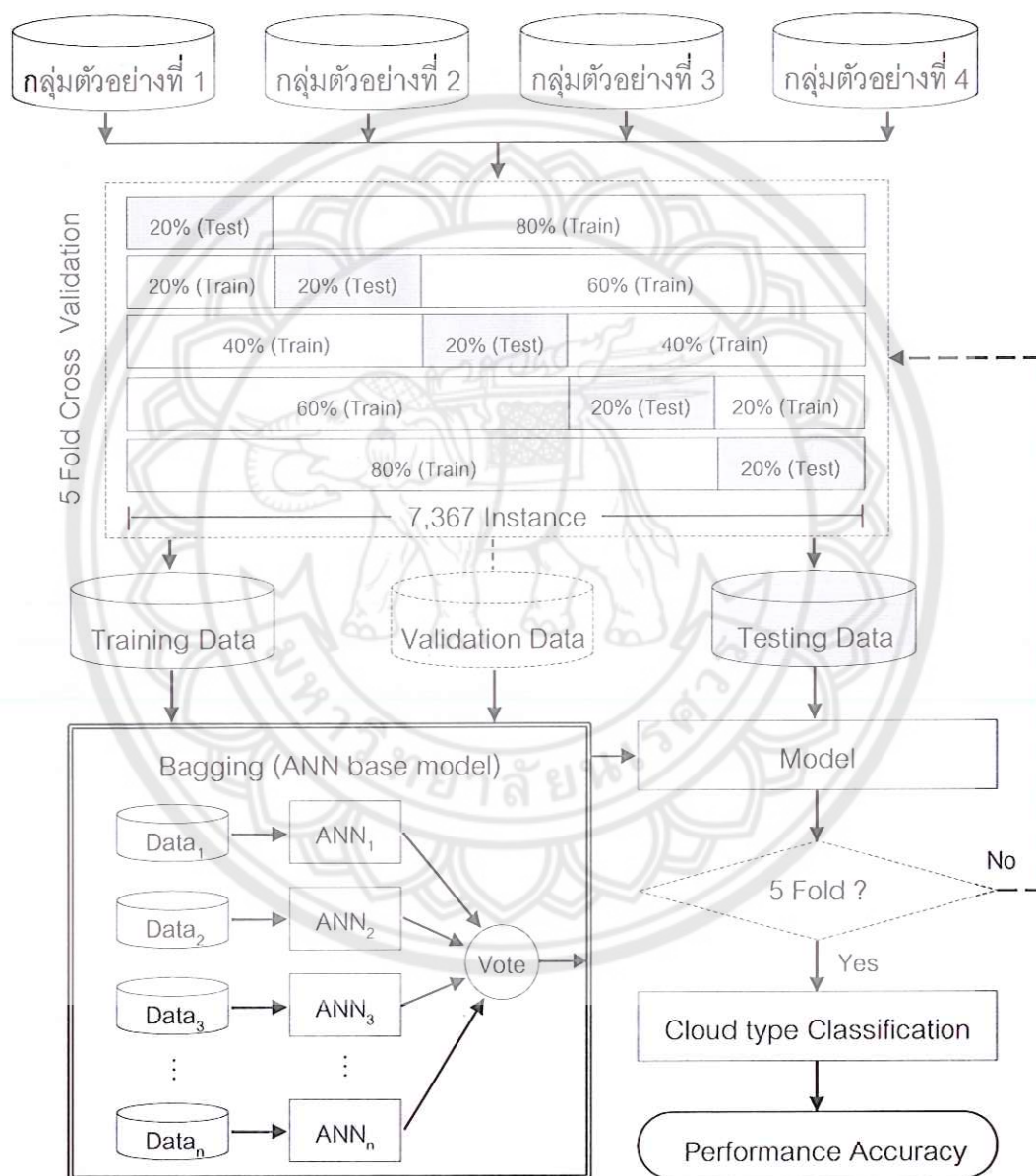
ตาราง 14 พารามิเตอร์ที่กำหนดของวิธี Bagging โดยใช้ J48 เป็นอัลกอริทึมพื้นฐาน

พารามิเตอร์	ความหมาย
Number of Model	จำนวนของแบบจำลองที่สร้างขึ้น
Sampling Data	ค่าที่ใช้สำหรับพิจารณาการสุ่มเลือกข้อมูลตัวอย่างเพื่อสร้างข้อมูลชุดใหม่

3.1.4 ตรวจสอบเงื่อนไขการทำ Cross Validation ผลลัพธ์ของการทำ 5 Fold - Cross Validation จะได้แบบจำลองมาทั้งหมด 5 แบบจำลอง จากนั้นจะทำการหาค่าเฉลี่ยของแบบจำลองที่สร้างขึ้นทั้งหมดเพื่อให้ได้ประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆ

3.1.5 ผลลัพธ์สุดท้ายของกระบวนการสร้างแบบจำลองแบบ Bagging โดยใช้ อัลกอริทึม J48 จะได้ประสิทธิภาพความถูกต้องสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆในแต่ ละกลุ่มตัวอย่าง

3.2 การสร้างแบบจำลอง Bagging โดยใช้ ANN เป็นอัลกอริทึมพื้นฐาน



ภาพ 30 แสดงขั้นตอนการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม ด้วยวิธี Bagging โดยใช้ ANN เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้

ภาพ 30 เป็นขั้นตอนการสร้างแบบจำลองการจำแนกประเภทการเรียนรู้แบบรวมกลุ่มด้วยวิธี Bagging โดยใช้อัลกอริทึม ANN สามารถอธิบายเป็นลำดับขั้นตอนได้ดังต่อไปนี้

3.2.1 เริ่มต้นจากการนำชุดข้อมูลของแต่ละชุดที่เตรียมไว้ คือ กลุ่มตัวอย่างที่ 1, 2, 3 และ 4 ตามลำดับเข้าสู่กระบวนการเรียนรู้ที่ละชุด ในการสร้างแบบจำลองการจำแนกประเภทชนิดของกลุ่มเมฆ และหาประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆทั้ง 4 กลุ่มตัวอย่าง

3.2.2 ทำการแบ่งข้อมูลสำหรับทดสอบประสิทธิภาพแบบจำลองที่สร้างขึ้น ในขั้นตอนนี้จะทำการแบ่งข้อมูลเพื่อใช้สำหรับการทดสอบด้วยวิธีการแบบ 5 Fold - Cross Validation โดยที่ชุดข้อมูลของแต่ละกลุ่มตัวอย่างจะประกอบไปด้วยข้อมูลตัวอย่างที่ใช้ในการทดลองจำนวน 7,376 ตัวอย่าง ซึ่งจำนวนตัวอย่างทั้งหมดนี้จะถูกแบ่งออกเป็นส่วน ๆ เท่ากัน ข้อมูลทุก ๆ ส่วนจะถูกเรียนรู้และทดสอบในการวัดประสิทธิภาพของแบบจำลองที่สร้างขึ้น

3.2.3 ทำการเรียนรู้และทดสอบชุดข้อมูลกลุ่มตัวอย่างในแต่ละกลุ่มด้วยวิธี Bagging โดยใช้ ANN เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ในการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม ขั้นตอนนี้ได้ทำการกำหนดค่าพารามิเตอร์ให้กับวิธี Bagging และปรับค่าพารามิเตอร์ให้เหมาะสมที่สุดสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆ โดยพารามิเตอร์ดังกล่าวได้อธิบายไว้ในตาราง 15

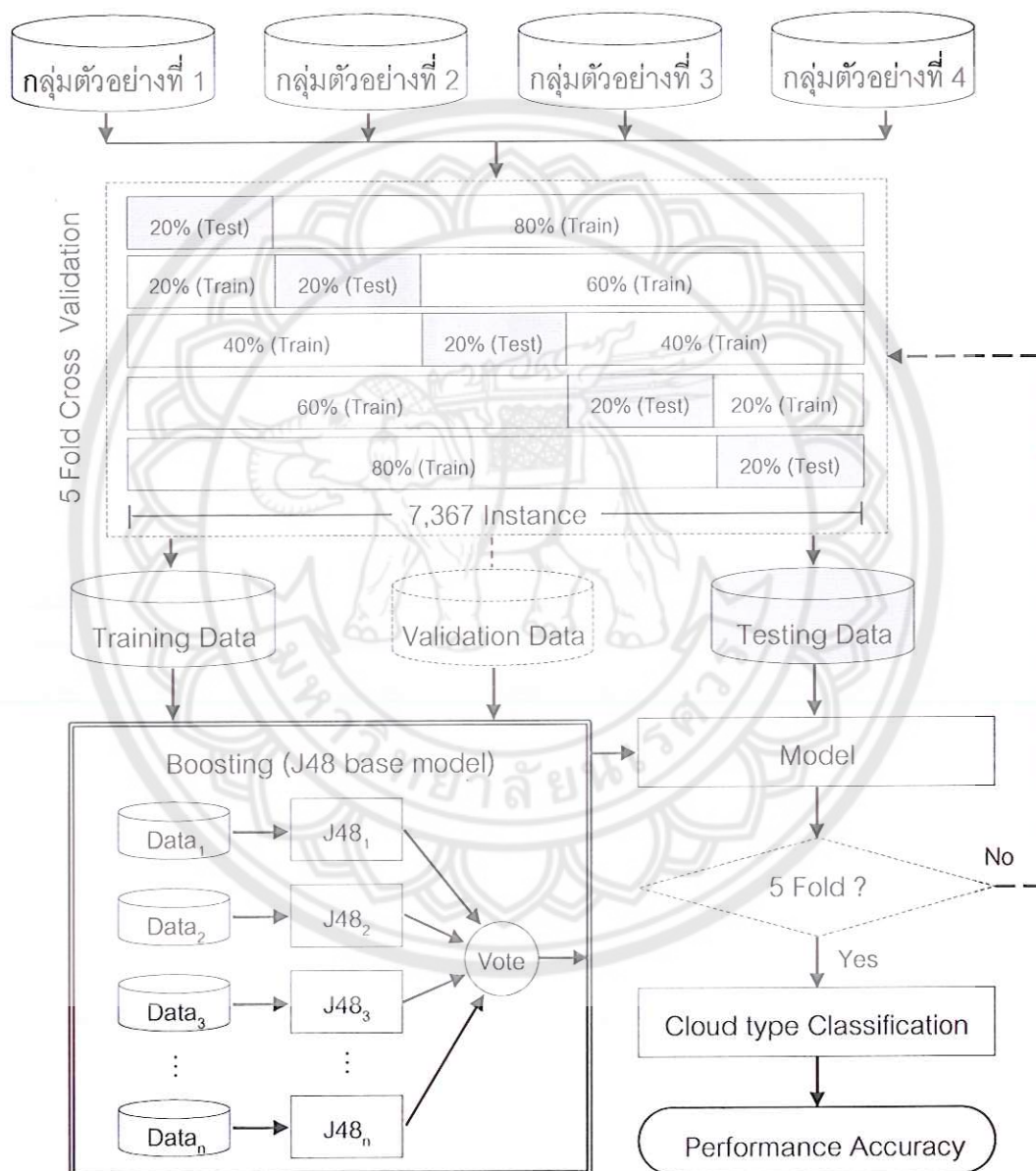
ตาราง 15 พารามิเตอร์ที่กำหนดของวิธี Bagging โดยใช้ ANN เป็นอัลกอริทึมพื้นฐาน

พารามิเตอร์	ความหมาย
Number of Model	จำนวนของแบบจำลองที่สร้างขึ้น
Sampling Data	ค่าที่ใช้สำหรับพิจารณาการสุ่มเลือกข้อมูลตัวอย่างเพื่อสร้างข้อมูลชุดใหม่

3.2.4 ตรวจสอบเงื่อนไขการทำ Cross Validation ผลลัพธ์ของการทำ 5 Fold - Cross Validation จะได้แบบจำลองมาทั้งหมด 5 แบบจำลอง จากนั้นจะทำการหาค่าเฉลี่ยของแบบจำลองที่สร้างขึ้นทั้งหมดเพื่อให้ได้ประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆ

3.2.5 ผลลัพธ์สุดท้ายของกระบวนการสร้างแบบจำลองแบบ Bagging โดยใช้ อัลกอริทึม ANN จะได้ประสิทธิภาพความถูกต้องสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆในแต่ละกลุ่มตัวอย่าง

3.3 การสร้างแบบจำลอง Boosting โดยใช้ J48 เป็นอัลกอริทึมพื้นฐาน



ภาพ 31 แสดงขั้นตอนการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่มด้วยวิธี Boosting โดยใช้ J48 เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้

ภาพ 31 เป็นขั้นตอนการสร้างแบบจำลองการจำแนกประเภทการเรียนรู้แบบรวมกลุ่มด้วยวิธี Boosting โดยใช้อัลกอริทึม J48 สามารถอธิบายเป็นลำดับขั้นตอนได้ดังต่อไปนี้

3.3.1 เริ่มต้นจากการนำชุดข้อมูลของแต่ละชุดที่เตรียมไว้ คือ กลุ่มตัวอย่างที่ 1, 2, 3 และ 4 ตามลำดับเข้าสู่กระบวนการเรียนรู้ทีละชุด ในการสร้างแบบจำลองการจำแนกประเภทชนิดของกลุ่มเมฆ และหาประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆทั้ง 4 กลุ่มตัวอย่าง

3.3.2 ทำการแบ่งข้อมูลสำหรับทดสอบประสิทธิภาพแบบจำลองที่สร้างขึ้น ในขั้นตอนนี้จะทำการแบ่งข้อมูลเพื่อใช้สำหรับการทดสอบด้วยวิธีการแบบ 5 Fold - Cross Validation โดยที่ชุดข้อมูลของแต่ละกลุ่มตัวอย่างจะประกอบไปด้วยข้อมูลตัวอย่างที่ใช้ในการทดลองจำนวน 7,376 ตัวอย่าง ซึ่งจำนวนตัวอย่างทั้งหมดนี้จะถูกแบ่งออกเป็น ส่วน ๆ เท่ากัน ข้อมูลทุก ๆ ส่วนจะถูกเรียนรู้และทดสอบในการวัดประสิทธิภาพของแบบจำลองที่สร้างขึ้น

3.3.3 ทำการเรียนรู้และทดสอบชุดข้อมูลกลุ่มตัวอย่างในแต่ละกลุ่มด้วยวิธี Boosting โดยใช้ J48 เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ในการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม ขั้นตอนนี้ได้ทำการกำหนดค่าพารามิเตอร์ให้กับวิธี Boosting และปรับค่าพารามิเตอร์ให้เหมาะสมที่สุดสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆ โดยพารามิเตอร์ดังกล่าวได้อธิบายไว้ในตาราง 16

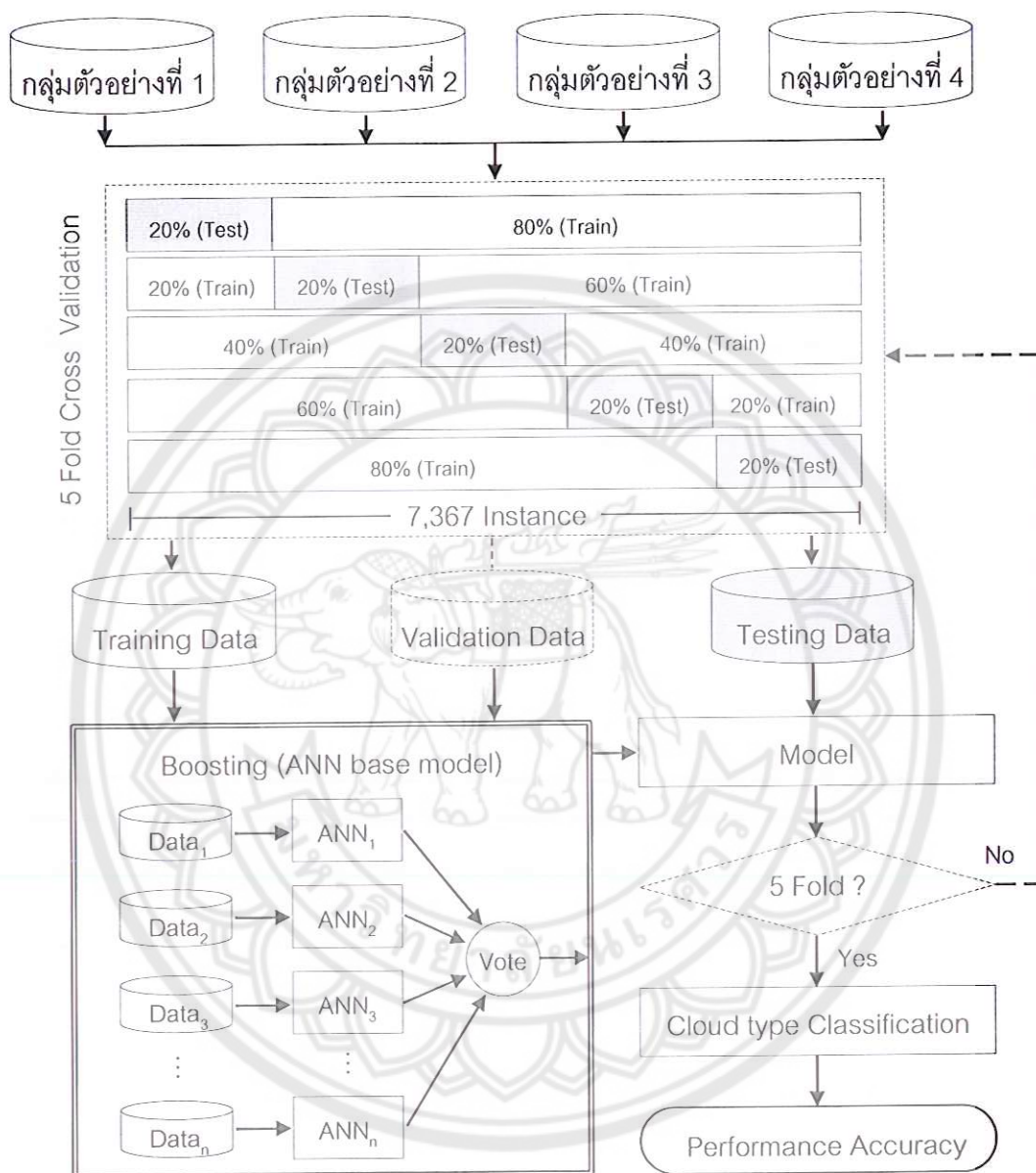
ตาราง 16 พารามิเตอร์ที่กำหนดของวิธี Boosting โดยใช้ J48 เป็นอัลกอริทึมพื้นฐาน

พารามิเตอร์	ความหมาย
Number of Model	จำนวนของแบบจำลองที่สร้างขึ้น

3.3.4 ตรวจสอบเงื่อนไขการทำ Cross Validation ผลลัพธ์ของการทำ 5 Fold - Cross Validation จะได้แบบจำลองมาทั้งหมด 5 แบบจำลอง จากนั้นจะทำการหาค่าเฉลี่ยของแบบจำลองที่สร้างขึ้นทั้งหมดเพื่อให้ได้ประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆ

3.3.5 ผลลัพธ์สุดท้ายของกระบวนการสร้างแบบจำลองแบบ Boosting โดยใช้ อัลกอริทึม J48 จะได้ประสิทธิภาพความถูกต้องสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆในแต่ละกลุ่มตัวอย่าง

3.4 การสร้างแบบจำลอง Boosting โดยใช้ ANN เป็นอัลกอริทึมพื้นฐาน



ภาพ 32 แสดงขั้นตอนการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่มด้วยวิธี Boosting โดยใช้ ANN เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้

ภาพ 32 เป็นขั้นตอนการสร้างแบบจำลองการจำแนกประเภทการเรียนรู้แบบรวมกลุ่มด้วยวิธี Boosting โดยใช้อัลกอริทึม ANN สามารถอธิบายเป็นลำดับขั้นตอนได้ดังต่อไปนี้

3.4.1 เริ่มต้นจากการนำชุดข้อมูลของแต่ละชุดที่เตรียมไว้ คือ กลุ่มตัวอย่างที่ 1, 2, 3 และ 4 ตามลำดับเข้าสู่กระบวนการเรียนรู้ทีละชุด ในการสร้างแบบจำลองการจำแนกประเภทชนิดของกลุ่มเมฆ และหาประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆทั้ง 4 กลุ่มตัวอย่าง

3.4.2 ทำการแบ่งข้อมูลสำหรับทดสอบประสิทธิภาพแบบจำลองที่สร้างขึ้น ในขั้นตอนนี้จะทำการแบ่งข้อมูลเพื่อใช้สำหรับการทดสอบด้วยวิธีการแบบ 5 Fold - Cross Validation โดยที่ชุดข้อมูลของแต่ละกลุ่มตัวอย่างจะประกอบไปด้วยข้อมูลตัวอย่างที่ใช้ในการทดลองจำนวน 7,376 ตัวอย่าง ซึ่งจำนวนตัวอย่างทั้งหมดนี้จะถูกแบ่งออกเป็นส่วน ๆ เท่ากัน ข้อมูลทุก ๆ ส่วนจะถูกเรียนรู้และทดสอบในการวัดประสิทธิภาพของแบบจำลองที่สร้างขึ้น

3.4.3 ทำการเรียนรู้และทดสอบชุดข้อมูลกลุ่มตัวอย่างในแต่ละกลุ่มด้วยวิธี Boosting โดยใช้ ANN เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ในการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม ขั้นตอนนี้ได้ทำการกำหนดค่าพารามิเตอร์ให้กับวิธี Boosting และปรับค่าพารามิเตอร์ให้เหมาะสมที่สุดสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆ โดยพารามิเตอร์ดังกล่าวได้อธิบายไว้ในตาราง 17

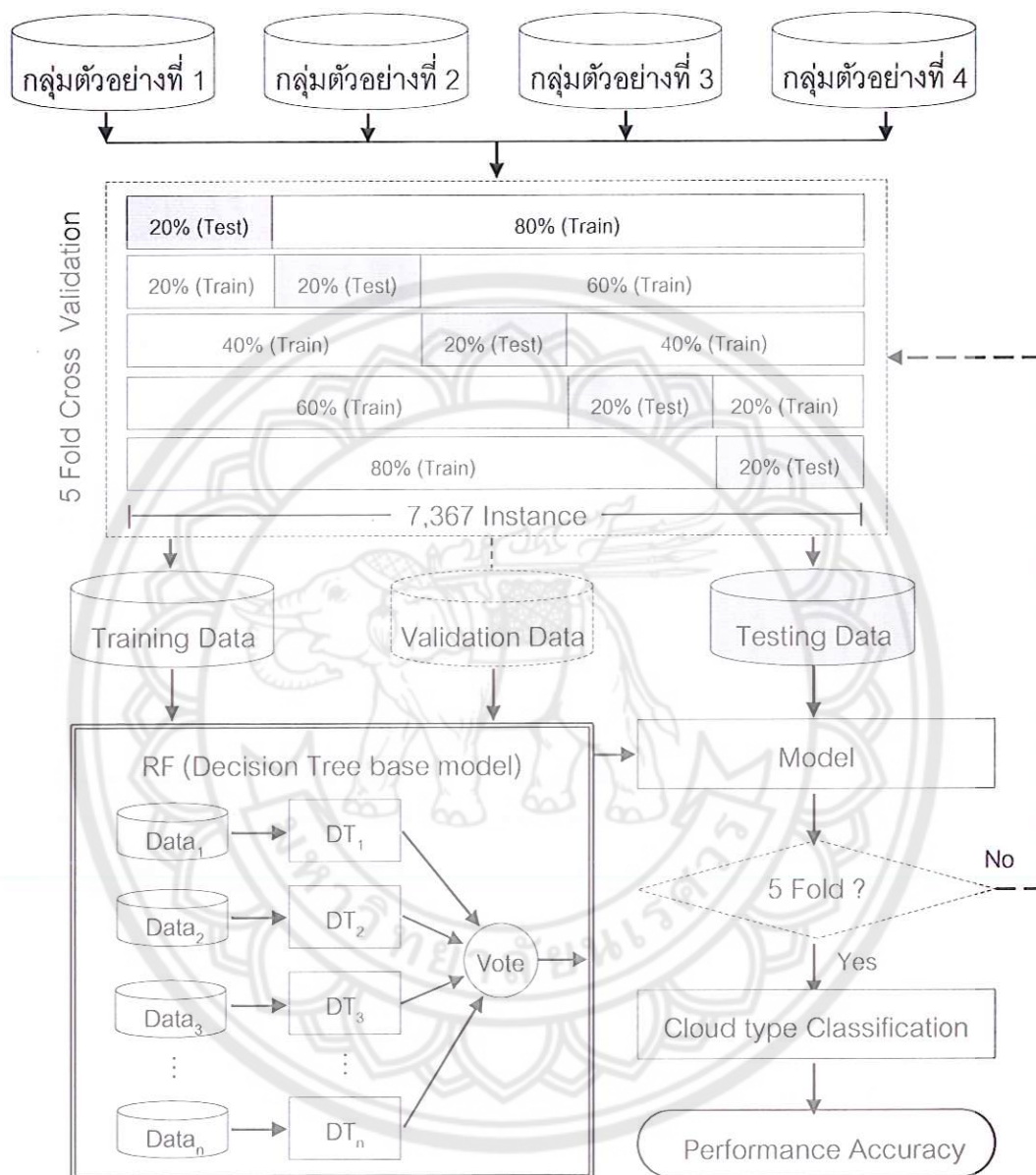
ตาราง 17 พารามิเตอร์ที่กำหนดของวิธี Boosting โดยใช้ ANN เป็นอัลกอริทึมพื้นฐาน

พารามิเตอร์	ความหมาย
Number of Model	จำนวนของแบบจำลองที่สร้างขึ้น

3.4.4 ตรวจสอบเงื่อนไขการทำ Cross Validation ผลลัพธ์ของการทำ 5 Fold - Cross Validation จะได้แบบจำลองมาทั้งหมด 5 แบบจำลอง จากนั้นจะทำการหาค่าเฉลี่ยของแบบจำลองที่สร้างขึ้นทั้งหมดเพื่อให้ได้ประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆ

3.4.5 ผลลัพธ์สุดท้ายของกระบวนการสร้างแบบจำลองแบบ Boosting โดยใช้ อัลกอริทึม ANN จะได้ประสิทธิภาพความถูกต้องสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆในแต่ละกลุ่มตัวอย่าง

3.5 การสร้างแบบจำลอง Random Forest



ภาพ 33 แสดงขั้นตอนการสร้างแบบจำลองแบบรวมกลุ่มด้วยวิธี Random Forest

ภาพ 33 เป็นขั้นตอนการสร้างแบบจำลองการจำแนกประเภทการเรียนรู้แบบรวมกลุ่มด้วยวิธี Random Forest สามารถอธิบายเป็นลำดับขั้นตอนได้ดังต่อไปนี้

3.5.1 เริ่มต้นจากการนำชุดข้อมูลของแต่ละชุดที่เตรียมไว้ คือ กลุ่มตัวอย่างที่ 1, 2, 3 และ 4 ตามลำดับเข้าสู่กระบวนการเรียนรู้ที่แต่ละชุด ในการสร้างแบบจำลองการจำแนกประเภทชนิดของกลุ่มเมฆ และหาประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆทั้ง 4 กลุ่มตัวอย่าง

3.5.2 ทำการแบ่งข้อมูลสำหรับทดสอบประสิทธิภาพแบบจำลองที่สร้างขึ้น ในขั้นตอนนี้จะทำการแบ่งข้อมูลเพื่อใช้สำหรับการทดสอบด้วยวิธีการแบบ 5 Fold - Cross Validation โดยที่ชุดข้อมูลของแต่ละกลุ่มตัวอย่างจะประกอบไปด้วยข้อมูลตัวอย่างที่ใช้ในการทดลองจำนวน 7,376 ตัวอย่าง ซึ่งจำนวนตัวอย่างทั้งหมดนี้จะถูกแบ่งออกเป็นส่วน ๆ เท่ากัน ข้อมูลทุก ๆ ส่วนจะถูกเรียนรู้และทดสอบในการวัดประสิทธิภาพของแบบจำลองที่สร้างขึ้น

3.5.3 ทำการเรียนรู้และทดสอบชุดข้อมูลกลุ่มตัวอย่างในแต่ละกลุ่มด้วยวิธี Random Forest ในการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม ขั้นตอนนี้ได้ทำการกำหนดค่าพารามิเตอร์ และปรับค่าพารามิเตอร์ให้เหมาะสมที่สุดสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆ โดยพารามิเตอร์ดังกล่าวได้อธิบายไว้ในตาราง 18

ตาราง 18 พารามิเตอร์ที่กำหนดของวิธี Random Forest

พารามิเตอร์	ความหมาย
Number of Model	จำนวนของแบบจำลองที่สร้างขึ้น
Number of Feature	ค่าที่ใช้สำหรับพิจารณาการสุ่มเลือก คุณลักษณะของชุดข้อมูลตัวอย่าง

3.5.4 ตรวจสอบเงื่อนไขการทำ Cross Validation ผลลัพธ์ของการทำ 5 Fold - Cross Validation จะได้แบบจำลองมาทั้งหมด 5 แบบจำลอง จากนั้นจะทำการหาค่าเฉลี่ยของแบบจำลองที่สร้างขึ้นทั้งหมดเพื่อให้ได้ประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆ

3.5.5 ผลลัพธ์สุดท้ายของกระบวนการสร้างแบบจำลองการเรียนรู้แบบรวมกลุ่มด้วยวิธี Random Forest จะได้ประสิทธิภาพความถูกต้องสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆในแต่ละกลุ่มตัวอย่าง

บทที่ 4

ผลการวิจัย

ในบทนี้ผู้วิจัยได้นำเสนอผลลัพธ์ที่ได้จากการทดลองสำหรับการวัดประสิทธิภาพในการจำแนกประเภทชนิดของกลุ่มเมฆ รวมไปถึงแสดงการปรับค่าพารามิเตอร์ให้เหมาะสมในแต่ละอัลกอริทึมรวมทั้งวิธีการเรียนรู้แบบรวมกลุ่มดังที่ได้กล่าวมาข้างต้น ซึ่งจะนำชุดข้อมูล 4 กลุ่มตัวอย่างในการทดลองตามรายละเอียดที่กล่าวมาในบทที่ 3 โดยในการทดลองได้ทำการเปรียบเทียบระหว่างการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยว และการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม เพื่อเปรียบเทียบหาประสิทธิภาพสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆที่ดีที่สุด และใช้ค่าจากตาราง Confusion Matrix ที่ได้จากระบวนการ Cross Validation ซึ่งเป็นค่าสำหรับใช้วัดประสิทธิภาพความถูกต้องสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆทั้งหมด สามารถคำนวณค่าความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆได้จากสมการ 23 ดังที่กล่าวไว้ในบท 2 สำหรับค่า Recall ที่แสดงในตาราง Confusion Matrix ในงานวิจัยนี้หมายถึง ค่าที่แสดงประสิทธิภาพความถูกต้องสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆในแต่ละชนิด โดยในงานวิจัยได้ทำการแบ่งผลลัพธ์ของการทดลองออกเป็น 2 ส่วน ดังต่อไปนี้

ส่วนที่ 1 วิเคราะห์โดยใช้อัลกอริทึมสำหรับการจำแนกประเภทแบบเดี่ยว

1. อัลกอริทึม J48
2. อัลกอริทึม ANN

ส่วนที่ 2 วิเคราะห์โดยใช้อัลกอริทึมสำหรับการจำแนกประเภทแบบรวมกลุ่ม

1. วิธีการแบบ Bagging
2. วิธีการแบบ Boosting
3. วิธีการแบบ Random Forest

วิเคราะห์โดยใช้วิธีการจำแนกประเภทแบบเดี่ยว

ผลการวิเคราะห์ข้อมูลสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆ ในการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยว ได้ทำการเปรียบเทียบ 2 อัลกอริทึม ได้แก่ J48 และ ANN

1. ผลลัพธ์จากการวิเคราะห์ข้อมูลโดยใช้อัลกอริทึม J48

จากผลการวิเคราะห์ข้อมูลในการจำแนกประเภทชนิดของกลุ่มเมฆโดยใช้อัลกอริทึม J48 ในการทดลองได้กำหนดค่าพารามิเตอร์ตามที่ได้แสดงในตาราง 19 เพื่อหาค่าพารามิเตอร์ที่เหมาะสมสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆ

ตาราง 19 การกำหนดค่าพารามิเตอร์ที่ใช้ในการทดลองสำหรับ J48

พารามิเตอร์	ค่าที่ใช้กำหนดในการทดลอง
Confidence factor	0.1, 0.2, 0.3, 0.4 และ 0.5
Minimum number	1, 2, 3, 4 และ 5

ตาราง 19 แสดงค่าพารามิเตอร์ที่ได้กำหนดในการทดลอง จากผลของการทดลอง พบว่า การกำหนดค่าพารามิเตอร์ที่เหมาะสมในการจำแนกประเภทชนิดของกลุ่มเมฆสำหรับ อัลกอริทึม J48 คือ ค่าความเชื่อมั่นอยู่ที่ 0.1 และจำนวนข้อมูลตัวอย่างขั้นต่ำที่เกิดขึ้นได้ในโหมด ไบเทากับ 2 จะให้ประสิทธิภาพความถูกต้องสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆเหมาะสม ที่สุด โดยคำนวณประสิทธิภาพความถูกต้องได้จากตาราง 20-23

ตาราง 20 ค่า Confusion Matrix โดยใช้ J48 วิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 1

Class		Prediction								
		C1	C2	C3	C4	C5	C6	C7	C8	Recall
Actual	C1	527	80	12	124	38	160	36	8	53.50%
	C2	107	292	13	190	8	40	11	2	44.04%
	C3	54	18	22	22	11	57	15	2	10.95%
	C4	58	88	4	1,126	1	6	8	0	87.22%
	C5	127	17	12	3	32	179	58	13	7.26%
	C6	147	27	17	15	66	617	721	15	37.97%
	C7	44	8	7	16	20	321	1,473	9	77.61%
	C8	9	4	3	2	14	78	118	35	13.31%

ตาราง 21 ค่า Confusion Matrix โดยใช้ J48 วิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 2

Class		Prediction			
		C1	C2	C3	Recall
Actual	C1	481	260	445	40.56%
	C2	145	1,734	75	88.74%
	C3	206	65	3,956	93.59%

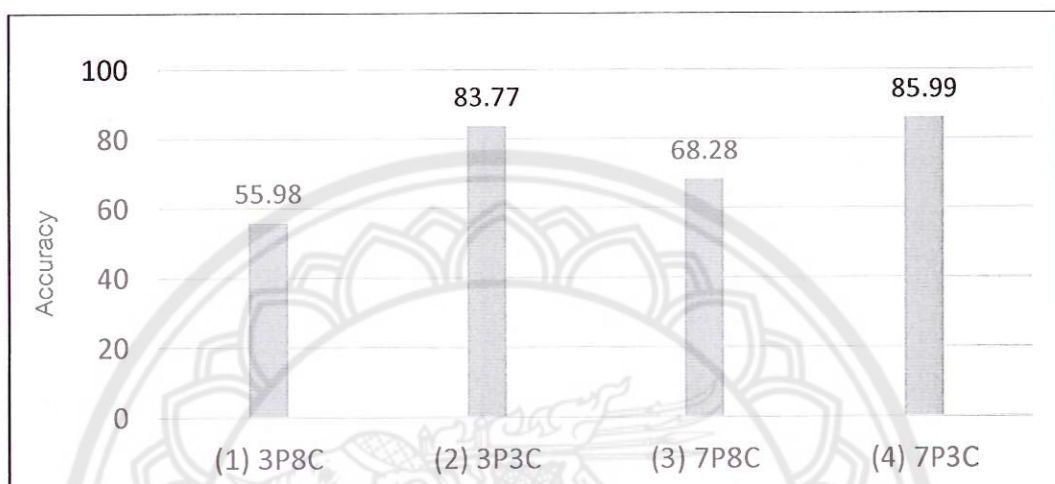
ตาราง 22 ค่า Confusion Matrix โดยใช้ J48 วิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 3

Class		Prediction								
		C1	C2	C3	C4	C5	C6	C7	C8	Recall
Actual	C1	587	70	28	72	69	106	44	9	59.59%
	C2	74	422	9	108	8	34	7	1	63.65%
	C3	40	16	68	9	4	42	19	3	33.83%
	C4	72	69	7	1,138	0	5	0	0	88.15%
	C5	95	20	7	3	139	111	54	12	31.52%
	C6	134	22	21	13	75	1,005	326	29	61.85%
	C7	40	4	8	4	33	267	1,512	30	79.66%
	C8	9	0	7	1	9	46	32	159	60.46%

ตาราง 23 ค่า Confusion Matrix โดยใช้ J48 วิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 4

		Prediction			
		C1	C2	C3	Recall
Actual	C1	612	196	378	51.60%
	C2	139	1,759	56	90.02%
	C3	201	62	3,964	93.78%

จากการคำนวณหาประสิทธิภาพความถูกต้องที่ได้จากตาราง Confusion Matrix ของข้อมูลกลุ่มตัวอย่างที่ 1-4 สำหรับการจำแนกประเภทชนิดของกลุ่มเมฆในการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยวโดยใช้อัลกอริทึม J48 สามารถสรุปผลการทดลองได้ดังภาพ 34



ภาพ 34 แสดงประสิทธิภาพความถูกต้องสำหรับการจำแนกประเภทแบบเดี่ยวโดยใช้ J48

จากภาพ 34 แสดงประสิทธิภาพความถูกต้องที่ได้จากผลการทดลองในการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยวโดยใช้อัลกอริทึม J48 ซึ่งคำนวณประสิทธิภาพความถูกต้องจากตาราง Confusion Matrix ที่ได้กำหนดค่าพารามิเตอร์อย่างเหมาะสมของกลุ่มตัวอย่างข้อมูลที่ 1-4 ตามตาราง 20-23 โดยประสิทธิภาพความถูกต้องสูงสุดในการจำแนกประเภทชนิดของกลุ่มเมฆ คือ กลุ่มตัวอย่างที่ 4 อยู่ที่ 85.99% รองลงมาคือ กลุ่มตัวอย่างที่ 2 กลุ่มตัวอย่างที่ 3 และกลุ่มตัวอย่างที่ 1 โดยประสิทธิภาพความถูกต้องอยู่ที่ 83.77%, 68.28% และ 55.98% ตามลำดับ

2. ผลลัพธ์จากการวิเคราะห์ข้อมูลโดยใช้อัลกอริทึม ANN

จากผลการวิเคราะห์ข้อมูลในการจำแนกประเภทชนิดของกลุ่มเมฆโดยใช้อัลกอริทึม ANN ในการทดลองได้กำหนดค่าพารามิเตอร์ตามที่ได้แสดงในตาราง 24 เพื่อหาค่าพารามิเตอร์ที่เหมาะสมสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆ

ตาราง 24 การกำหนดค่าพารามิเตอร์ที่ใช้ในการทดลองสำหรับ ANN

พารามิเตอร์	ค่าที่ใช้กำหนดในการทดลอง
Training Cycles	10,000 รอบ
Hidden Node	5, 10, 15 และ 20
Learning Rate	0.01
Momentum	0.3

ตาราง 24 แสดงการกำหนดค่าพารามิเตอร์ที่ใช้ในการทดลองของอัลกอริทึม ANN จากการทดลองได้กำหนด Training Cycles เท่ากับ 10,000 รอบ Hidden Node อยู่ที่ 5, 10, 15 และ 20 โหนด Learning Rate เท่ากับ 0.01 และค่า Momentum เท่ากับ 0.3 ผลของการทดลองพบว่า การกำหนด Hidden Node เท่ากับ 15 โหนด ให้ประสิทธิภาพความถูกต้องเหมาะสมที่สุด สามารถคำนวณประสิทธิภาพความถูกต้องได้จากตาราง 25-28

ตาราง 25 ค่า Confusion Matrix โดยใช้ ANN วิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 1

Class	Prediction								Recall
	C1	C2	C3	C4	C5	C6	C7	C8	
C1	548	81	4	125	3	188	30	6	55.63%
C2	130	239	0	244	3	37	10	0	36.05%
C3	68	9	9	22	1	80	11	1	4.48%
C4	41	95	1	1,145	0	7	2	0	88.69%
C5	132	18	0	3	5	228	48	7	1.13%
C6	153	15	12	21	1	741	679	3	45.60%
C7	35	0	1	18	1	377	1,466	0	77.24%
C8	12	3	1	2	1	92	120	32	12.17%

ตาราง 26 ค่า Confusion Matrix โดยใช้ ANN วิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 2

Class		Prediction			Recall
		C1	C2	C3	
Actual	C1	538	251	397	45.36%
	C2	148	1,744	62	89.25%
	C3	223	72	3,932	93.02%

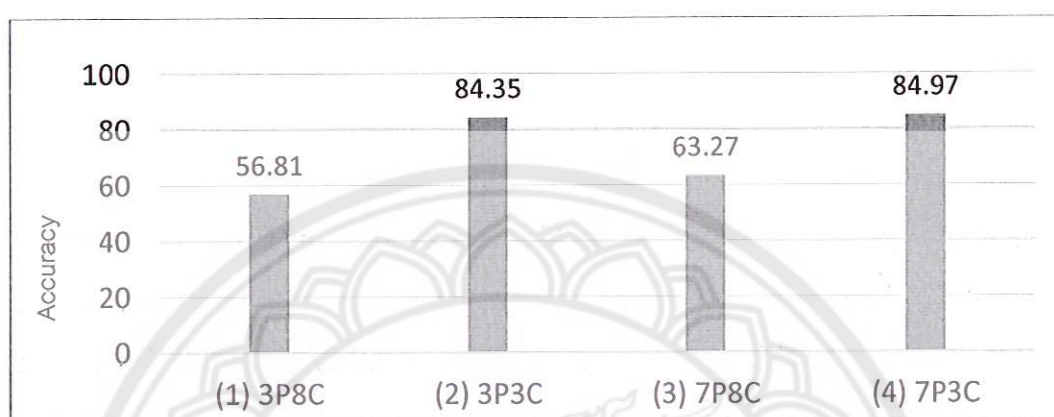
ตาราง 27 ค่า Confusion Matrix โดยใช้ ANN วิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 3

Class		Prediction								Recall
		C1	C2	C3	C4	C5	C6	C7	C8	
Actual	C1	609	79	7	83	20	130	41	16	61.83%
	C2	105	354	2	136	9	49	7	1	53.39%
	C3	53	11	6	14	5	90	19	3	2.99%
	C4	60	107	0	1,114	0	8	2	0	86.29%
	C5	131	15	4	1	37	196	38	19	8.39%
	C6	116	21	5	12	13	920	503	35	56.62%
	C7	36	5	2	3	1	337	1,507	7	79.40%
	C8	4	2	1	1	3	68	70	114	43.35%

ตาราง 28 ค่า Confusion Matrix โดยใช้ ANN วิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 4

Class		Prediction			Recall
		C1	C2	C3	
Actual	C1	594	218	374	50.08%
	C2	154	1,724	76	88.23%
	C3	215	70	3,942	93.26%

จากการคำนวณหาประสิทธิภาพความถูกต้องที่ได้จากตาราง Confusion Matrix ของข้อมูลกลุ่มตัวอย่างที่ 1-4 สำหรับการจำแนกประเภทชนิดของกลุ่มเมฆในการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยวโดยใช้อัลกอริทึม ANN สามารถสรุปผลการทดลองได้ดังภาพ 35



ภาพ 35 แสดงประสิทธิภาพความถูกต้องสำหรับการจำแนกประเภทแบบเดี่ยวโดยใช้ ANN

ภาพ 35 แสดงประสิทธิภาพความถูกต้องจากผลการทดลองในการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยวโดยใช้อัลกอริทึม ANN ซึ่งคำนวณประสิทธิภาพความถูกต้องจากตาราง Confusion Matrix ที่ได้กำหนดค่าพารามิเตอร์อย่างเหมาะสมของกลุ่มตัวอย่างข้อมูล 1-4 ตามตาราง 25-28 โดยประสิทธิภาพความถูกต้องสูงสุดในการจำแนกประเภทชนิดของกลุ่มเมฆคือ กลุ่มตัวอย่างที่ 4 อยู่ที่ 84.97% รองลงมาคือ กลุ่มตัวอย่างที่ 2 กลุ่มตัวอย่างที่ 3 และกลุ่มตัวอย่างที่ 1 โดยประสิทธิภาพความถูกต้องอยู่ที่ 84.35%, 63.27% และ 56.81% ตามลำดับ

วิเคราะห์โดยใช้วิธีการจำแนกประเภทแบบรวมกลุ่ม

ผลการวิเคราะห์ข้อมูลสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆ ในการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม ได้ทำการเปรียบเทียบ 3 วิธี ได้แก่ Bagging, Boosting และ Random Forest

1. ผลลัพธ์จากการวิเคราะห์ข้อมูลด้วยวิธี Bagging

จากผลการทดลองด้วยวิธี Bagging สำหรับการวิเคราะห์ประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆ สามารถแสดงผลการทดลองโดยแบ่งออกเป็น 2 รูปแบบดังต่อไปนี้

1.1 ผลลัพธ์จากการวิเคราะห์ข้อมูลด้วยวิธี Bagging โดยใช้อัลกอริทึม J48

ผลลัพธ์ของการวิเคราะห์ข้อมูลในการจำแนกประเภทชนิดของกลุ่มเมฆด้วยวิธี Bagging โดยใช้ J48 เป็นอัลกอริทึมพื้นฐานการเรียนรู้ สำหรับการทดลองนี้ได้กำหนดค่าพารามิเตอร์ตามที่แสดงในตาราง 29 เพื่อหาค่าพารามิเตอร์ที่เหมาะสมให้กับการจำแนกประเภทชนิดของกลุ่มเมฆ ดังนี้

ตาราง 29 การกำหนดค่าพารามิเตอร์ที่ใช้ในการทดลองด้วยวิธี Bagging โดยใช้ J48 เป็นพื้นฐานการเรียนรู้

พารามิเตอร์	ค่าที่ใช้กำหนดในการทดลอง
Number of Model	50, 100, 150, 200 และ 300
Sampling Data	0.5, 0.6, 0.7, 0.8 และ 0.9

ตาราง 29 แสดงค่าพารามิเตอร์ที่ได้กำหนดในการทดลองของวิธี Bagging โดยใช้ J48 เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ข้อมูล ในส่วนนี้ได้ทำการทดลองโดยสร้างแบบจำลอง J48 จำนวน 50, 100, 150, 200 และ 300 แบบจำลอง ค่าที่ใช้สำหรับสุ่มเพื่อสร้างข้อมูลชุดใหม่ กำหนดค่าอยู่ที่ 0.5, 0.6, 0.7, 0.8 และ 0.9 ผลการทดลองพบว่า การสร้าง J48 จำนวน 150 แบบจำลอง และใช้ค่าของการสุ่มเพื่อสร้างข้อมูลชุดใหม่อยู่ที่ 0.9 ให้ประสิทธิภาพเหมาะสมที่สุด ซึ่งสามารถคำนวณค่าความถูกต้องที่กำหนดค่าพารามิเตอร์อย่างเหมาะสมได้จากตาราง Confusion Matrix ในตาราง 30-33 ของกลุ่มตัวอย่างข้อมูลที่ 1-4

ตาราง 30 ค่า Confusion Matrix วิธีการแบบ Bagging โดยใช้ J48 เป็นพื้นฐานการเรียนรู้ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 1

Class	Prediction								Recall
	C1	C2	C3	C4	C5	C6	C7	C8	
Actual	C1	546	81	8	127	29	156	33	55.43%
	C2	103	321	7	182	7	37	6	48.42%
	C3	70	11	21	19	7	60	13	10.45%
	C4	49	86	2	1,144	1	1	8	88.61%

ตาราง 30 (ต่อ)

Class		Prediction								
		C1	C2	C3	C4	C5	C6	C7	C8	Recall
Actual	C5	130	17	4	4	36	181	60	9	8.16%
	C6	140	23	18	17	35	698	684	10	42.95%
	C7	40	1	5	19	13	376	1,440	4	75.87%
	C8	12	4	2	2	9	91	111	32	12.17%

ตาราง 31 ค่า Confusion Matrix วิธีการแบบ Bagging โดยใช้ J48 เป็นพื้นฐานการ
เรียนรู้ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 2

		Prediction			
		C1	C2	C3	
Actual	Class				Recall
	C1	521	235	430	43.93%
	C2	155	1,727	72	88.38%
	C3	214	70	3,943	93.28%

ตาราง 32 ค่า Confusion Matrix วิธีการแบบ Bagging โดยใช้ J48 เป็นพื้นฐานการ
เรียนรู้ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 3

Class		Prediction								
		C1	C2	C3	C4	C5	C6	C7	C8	Recall
Actual	C1	651	51	7	67	44	118	32	15	66.09%
	C2	70	448	4	96	2	37	6	0	67.57%
	C3	45	8	61	10	4	54	13	6	30.35%
	C4	48	44	5	1,186	2	6	0	0	91.87%

ตาราง 32 (ต่อ)

Class		Prediction								
		C1	C2	C3	C4	C5	C6	C7	C8	Recall
Actual	C5	109	15	2	0	122	139	40	14	27.66%
	C6	108	18	8	12	31	1,157	264	27	71.20%
	C7	39	3	2	2	9	241	1,587	15	83.61%
	C8	6	0	0	0	7	49	33	168	63.88%

ตาราง 33 ค่า Confusion Matrix วิธีการแบบ Bagging โดยใช้ J48 เป็นพื้นฐานการเรียนรู้ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 4

Class		Prediction			Recall
		C1	C2	C3	
Actual	C1	653	175	358	55.06%
	C2	117	1,776	61	90.89%
	C3	197	56	3,974	94.01%

1.2 ผลลัพธ์จากการวิเคราะห์ข้อมูลด้วยวิธี Bagging โดยใช้อัลกอริทึม ANN

ผลลัพธ์ของการวิเคราะห์ข้อมูลในการจำแนกประเภทชนิดของกลุ่มเมฆด้วยวิธี Bagging โดยใช้ ANN เป็นอัลกอริทึมพื้นฐานการเรียนรู้ สำหรับการทดลองนี้ได้กำหนดค่าพารามิเตอร์ตามที่แสดงในตาราง 34 เพื่อหาค่าพารามิเตอร์ที่เหมาะสมให้กับการจำแนกประเภทชนิดของกลุ่มเมฆ ดังนี้

ตาราง 34 การกำหนดค่าพารามิเตอร์ที่ใช้ในการทดลองด้วยวิธี Bagging โดยใช้ ANN เป็นพื้นฐานการเรียนรู้

พารามิเตอร์	ค่าที่ใช้กำหนดในการทดลอง
Number of Model	10, 15 และ 20
Sampling Data	0.5, 0.6, 0.7, 0.8 และ 0.9

ตาราง 34 แสดงค่าพารามิเตอร์ที่ได้กำหนดในการทดลองของวิธี Bagging โดยใช้ ANN เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ข้อมูล ในส่วนนี้ได้ทำการทดลองโดยสร้างแบบจำลอง ANN จำนวน 10, 15, และ 20 แบบจำลอง สำหรับค่าที่ใช้สำหรับสุ่มเพื่อสร้างข้อมูลชุดใหม่ กำหนดค่าอยู่ที่ 0.5, 0.6, 0.7, 0.8 และ 0.9 ผลการทดลองพบว่า การสร้าง ANN จำนวน 10 แบบจำลอง และใช้ค่าของการสุ่มเพื่อสร้างข้อมูลชุดใหม่อยู่ที่ 0.9 ให้ประสิทธิภาพเหมาะสมที่สุด ซึ่งสามารถคำนวณค่าความถูกต้องที่กำหนดค่าพารามิเตอร์อย่างเหมาะสม ได้จากตาราง Confusion Matrix ในตาราง 35-38 ของกลุ่มตัวอย่างข้อมูลที่ 1-4

ตาราง 35 ค่า Confusion Matrix วิธีการแบบ Bagging โดยใช้ ANN เป็นพื้นฐาน
การเรียนรู้ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 1

Class		Prediction								
		C1	C2	C3	C4	C5	C6	C7	C8	Recall
Actual	C1	542	84	1	123	4	196	29	6	55.03%
	C2	121	245	1	244	4	39	9	0	36.95%
	C3	68	8	5	22	0	85	13	0	2.49%
	C4	41	92	0	1,147	0	8	2	1	88.85%
	C5	131	17	0	3	4	236	41	9	0.91%
	C6	144	11	10	18	0	748	689	5	46.03%
	C7	36	0	0	20	0	374	1,468	0	77.34%
	C8	8	4	1	3	4	92	121	30	11.41%

ตาราง 36 ค่า Confusion Matrix วิธีการแบบ Bagging โดยใช้ ANN เป็นพื้นฐาน
การเรียนรู้ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 2

Class		Prediction			
		C1	C2	C3	Recall
Actual	C1	549	244	393	46.29%
	C2	156	1,739	59	89.00%
	C3	211	73	3,943	93.28%

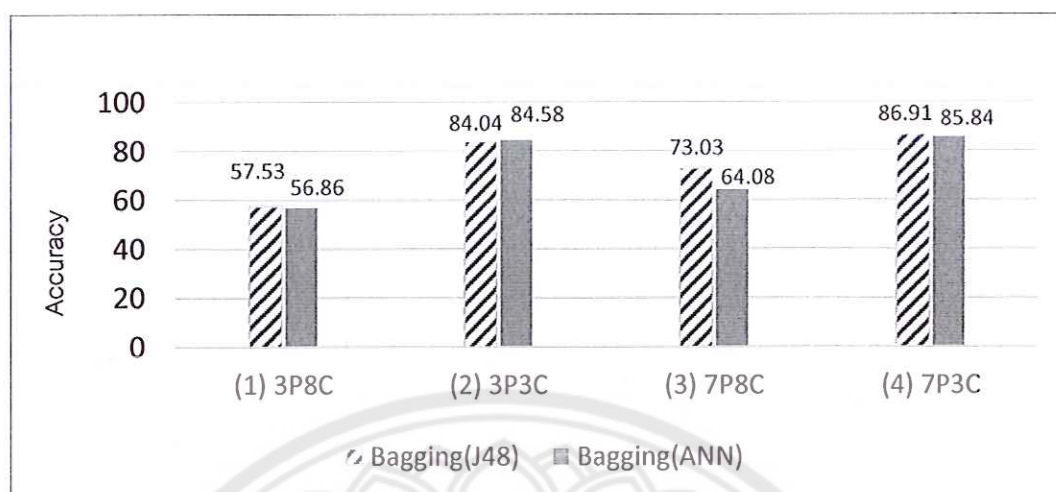
ตาราง 37 ค่า Confusion Matrix วิธีการแบบ Bagging โดยใช้ ANN เป็นพื้นฐาน
การเรียนรู้ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 3

Class		Prediction								
		C1	C2	C3	C4	C5	C6	C7	C8	Recall
Actual	C1	611	59	0	97	24	143	36	15	62.03%
	C2	105	343	1	144	9	54	7	0	51.73%
	C3	59	5	2	14	1	105	11	4	1.00%
	C4	57	81	0	1,146	0	5	2	0	88.77%
	C5	119	9	0	1	54	209	30	19	12.24%
	C6	97	18	4	10	13	954	497	32	58.71%
	C7	37	1	0	4	1	346	1,505	4	79.29%
	C8	4	0	1	1	2	81	68	106	40.30%

ตาราง 38 ค่า Confusion Matrix วิธีการแบบ Bagging โดยใช้ ANN เป็นพื้นฐาน
การเรียนรู้ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 4

		Prediction			
		C1	C2	C3	
Actual	Class				Recall
	C1	638	189	359	53.79%
	C2	150	1,721	83	88.08%
	C3	209	53	3,965	93.80%

จากการคำนวณหาประสิทธิภาพความถูกต้องที่ได้จากตาราง Confusion Matrix ของข้อมูลกลุ่มตัวอย่างที่ 1-4 สำหรับการจำแนกประเภทชนิดของกลุ่มเมฆในการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่มด้วยวิธี Bagging โดยใช้อัลกอริทึม J48 ในตาราง 30-33 และอัลกอริทึม ANN ในตาราง 35-38 สามารถสรุปผลการทดลองได้ดังภาพ 36



ภาพ 36 แสดงประสิทธิภาพความถูกต้องสำหรับการจำแนกประเภทแบบรวมกลุ่มวิธี Bagging

จากภาพ 36 แสดงประสิทธิภาพความถูกต้องที่ได้จากผลการทดลองในการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่มด้วยวิธี Bagging โดยใช้อัลกอริทึม J48 ซึ่งกลุ่มข้อมูลตัวอย่างที่ให้ประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆสูงสุด คือ กลุ่มตัวอย่างที่ 4 อยู่ที่ 86.91% รองลงมาคือ กลุ่มตัวอย่างที่ 2 กลุ่มตัวอย่างที่ 3 และกลุ่มตัวอย่างที่ 1 โดยประสิทธิภาพความถูกต้องอยู่ที่ 84.04%, 73.03% และ 57.53% ตามลำดับ สำหรับการจำแนกประเภทแบบรวมกลุ่มด้วยวิธี Bagging โดยใช้อัลกอริทึม ANN ให้ประสิทธิภาพของแต่ละกลุ่มตัวอย่างมีลักษณะไปในทางเดียวกันกับการใช้อัลกอริทึม J48 คือ กลุ่มตัวอย่างที่ 4 จะให้ประสิทธิภาพความถูกต้องการจำแนกประเภทชนิดของกลุ่มเมฆสูงที่สุด รองลงมาคือ กลุ่มตัวอย่างที่ 2 กลุ่มตัวอย่างที่ 3 และกลุ่มตัวอย่างที่ 1 โดยประสิทธิภาพความถูกต้องอยู่ที่ 85.84%, 84.58%, 64.08% และ 56.86% ตามลำดับ จากภาพรวมจะเห็นว่าการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่มด้วยวิธี Bagging โดยใช้อัลกอริทึม J48 ให้ประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆดีกว่าการใช้อัลกอริทึม ANN

2. ผลลัพธ์จากการวิเคราะห์ข้อมูลด้วยวิธี Boosting

จากผลการทดลองด้วยวิธี Boosting สำหรับการวิเคราะห์ประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆ สามารถแสดงผลการทดลองโดยแบ่งออกเป็น 2 รูปแบบดังต่อไปนี้

2.1 ผลลัพธ์จากการวิเคราะห์ข้อมูลด้วยวิธี Boosting โดยใช้อัลกอริทึม J48

ผลลัพธ์ของการวิเคราะห์ข้อมูลในการจำแนกประเภทชนิดของกลุ่มเมฆด้วยวิธี Boosting โดยใช้ J48 เป็นอัลกอริทึมพื้นฐานการเรียนรู้ สำหรับการทดลองนี้ได้กำหนดค่าพารามิเตอร์ตามที่แสดงในตาราง 39 เพื่อหาค่าพารามิเตอร์ที่เหมาะสมให้กับการจำแนกประเภทชนิดของกลุ่มเมฆ ดังนี้

ตาราง 39 การกำหนดค่าพารามิเตอร์ที่ใช้ในการทดลองด้วยวิธี Boosting โดยใช้ J48 เป็นพื้นฐานการเรียนรู้

พารามิเตอร์	ค่าที่ใช้กำหนดในการทดลอง
Number of Model	50, 100, 150, 200 และ 300

ตาราง 39 แสดงค่าพารามิเตอร์ที่ได้กำหนดในการทดลองของวิธี Boosting โดยใช้ J48 เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ข้อมูล ในส่วนนี้ได้ทำการทดลองโดยสร้างแบบจำลอง J48 จำนวน 50, 100, 150, 200 และ 300 แบบจำลอง ผลการทดลองแสดงให้เห็นว่าข้อมูลกลุ่มตัวอย่างที่ 1 สามารถสร้าง J48 ได้จำนวน 11 แบบจำลอง ข้อมูลกลุ่มตัวอย่างที่ 2 สามารถสร้าง J48 ได้จำนวน 12 แบบจำลอง ข้อมูลกลุ่มตัวอย่างที่ 3 สามารถสร้าง J48 ได้จำนวน 16 แบบจำลอง และข้อมูลกลุ่มตัวอย่างที่ 4 สามารถสร้าง J48 ได้จำนวน 11 แบบจำลอง ซึ่งสามารถคำนวณค่าความถูกต้องที่กำหนดค่าพารามิเตอร์อย่างเหมาะสม ได้จากตาราง Confusion Matrix ในตาราง 40-43 ของกลุ่มตัวอย่างข้อมูล 1-4

ตาราง 40 ค่า Confusion Matrix วิธีการแบบ Boosting โดยใช้ J48 เป็นพื้นฐานการเรียนรู้ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 1

Class		Prediction								
		C1	C2	C3	C4	C5	C6	C7	C8	Recall
Actual	C1	524	84	22	107	51	148	42	7	53.20%
	C2	109	335	11	165	4	34	3	2	50.53%
	C3	54	26	20	17	11	58	14	1	9.95%
	C4	56	96	2	1,123	1	5	6	2	86.99%

ตาราง 40 (ต่อ)

Class		Prediction								
		C1	C2	C3	C4	C5	C6	C7	C8	Recall
Actual	C5	125	13	8	2	62	150	63	18	14.06%
	C6	159	28	27	13	68	666	637	27	40.98%
	C7	30	3	4	11	25	480	1,322	23	69.65%
	C8	11	5	4	1	15	84	97	46	17.49%

ตาราง 41 ค่า Confusion Matrix วิธีการแบบ Boosting โดยใช้ J48 เป็นพื้นฐาน
การเรียนรู้ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 2

		Prediction			
		C1	C2	C3	
Actual	Class				Recall
	C1	552	234	400	46.54%
	C2	182	1,700	72	87.00%
	C3	274	61	3,892	92.07%

ตาราง 42 ค่า Confusion Matrix วิธีการแบบ Boosting โดยใช้ J48 เป็นพื้นฐาน
การเรียนรู้ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 3

Class		Prediction								
		C1	C2	C3	C4	C5	C6	C7	C8	Recall
Actual	C1	643	54	11	68	59	111	31	8	65.28%
	C2	71	444	8	100	7	28	5	0	66.97%
	C3	38	10	81	8	5	44	12	3	40.30%
	C4	53	48	3	1,182	0	5	0	0	91.56%

ตาราง 42 (ต่อ)

Class		Prediction								
		C1	C2	C3	C4	C5	C6	C7	C8	Recall
Actual	C5	84	10	3	1	170	109	49	15	38.55%
	C6	108	16	17	9	70	1,103	281	21	67.88%
	C7	34	5	5	2	21	264	1,557	10	82.03%
	C8	8	1	2	2	11	48	36	155	58.94%

ตาราง 43 ค่า Confusion Matrix วิธีการแบบ Boosting โดยใช้ J48 เป็นพื้นฐาน
การเรียนรู้ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 4

		Prediction			Recall
		C1	C2	C3	
Actual	C1	671	173	342	56.58%
	C2	119	1,778	57	90.99%
	C3	215	56	3,956	93.59%

2.2 ผลลัพธ์จากการวิเคราะห์ข้อมูลด้วยวิธี Boosting โดยใช้อัลกอริทึม ANN

ผลลัพธ์ของการวิเคราะห์ข้อมูลในการจำแนกประเภทชนิดของกลุ่มเมฆด้วยวิธี Boosting โดยใช้ ANN เป็นอัลกอริทึมพื้นฐานการเรียนรู้ สำหรับการทดลองนี้ได้กำหนดค่าพารามิเตอร์ตามที่แสดงในตาราง 44 เพื่อหาค่าพารามิเตอร์ที่เหมาะสมให้กับการจำแนกประเภทชนิดของกลุ่มเมฆ ดังนี้

ตาราง 44 การกำหนดค่าพารามิเตอร์ที่ใช้ในการทดลองด้วยวิธี Boosting โดยใช้ ANN เป็นพื้นฐานการเรียนรู้

พารามิเตอร์	ค่าที่ใช้กำหนดในการทดลอง
Number of Model	10, 20 และ 30

ตาราง 44 แสดงค่าพารามิเตอร์ที่ได้กำหนดในการทดลองของวิธี Boosting โดยใช้ ANN เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้ข้อมูล ในส่วนนี้ได้ทำการทดลองโดยสร้างแบบจำลอง ANN จำนวน 10, 20 และ 30 แบบจำลอง ผลการทดลองแสดงให้เห็นว่า ข้อมูลกลุ่มตัวอย่างที่ 1 สามารถสร้าง ANN ได้จำนวน 3 แบบจำลอง ข้อมูลกลุ่มตัวอย่างที่ 2 สามารถสร้าง ANN ได้จำนวน 5 แบบจำลอง ข้อมูลกลุ่มตัวอย่างที่ 3 สามารถสร้าง ANN ได้จำนวน 6 แบบจำลอง และข้อมูลกลุ่มตัวอย่างที่ 4 สามารถสร้าง ANN ได้จำนวน 6 แบบจำลอง โดยค่าของการสุ่มเพื่อสร้างข้อมูลชุดใหม่อยู่ที่ 0.9 ให้ความเหมาะสมที่สุด ซึ่งสามารถคำนวณค่าความถูกต้องที่กำหนดค่าพารามิเตอร์อย่างเหมาะสมได้จากตาราง Confusion Matrix ในตาราง 45-48 ของกลุ่มตัวอย่างข้อมูลที 1-4

ตาราง 45 ค่า Confusion Matrix วิธีการแบบ Boosting โดยใช้ ANN เป็นพื้นฐาน
การเรียนรู้ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 1

Class		Prediction								
		C1	C2	C3	C4	C5	C6	C7	C8	Recall
Actual	C1	545	80	2	116	4	194	39	5	55.33%
	C2	130	236	0	243	5	40	9	0	35.60%
	C3	66	9	5	22	0	85	14	0	2.49%
	C4	44	97	0	1,133	0	10	6	1	87.76%
	C5	135	17	0	2	6	231	42	8	1.36%
	C6	147	16	9	18	1	759	672	3	46.71%
	C7	41	0	0	16	2	381	1,458	0	76.82%
	C8	12	4	0	2	1	89	124	31	11.79%

ตาราง 46 ค่า Confusion Matrix วิธีการแบบ Boosting โดยใช้ ANN เป็นพื้นฐาน
การเรียนรู้ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 2

		Prediction			
		C1	C2	C3	Recall
Actual	C1	533	254	399	44.94%
	C2	144	1,747	63	89.41%
	C3	213	77	3,937	93.14%

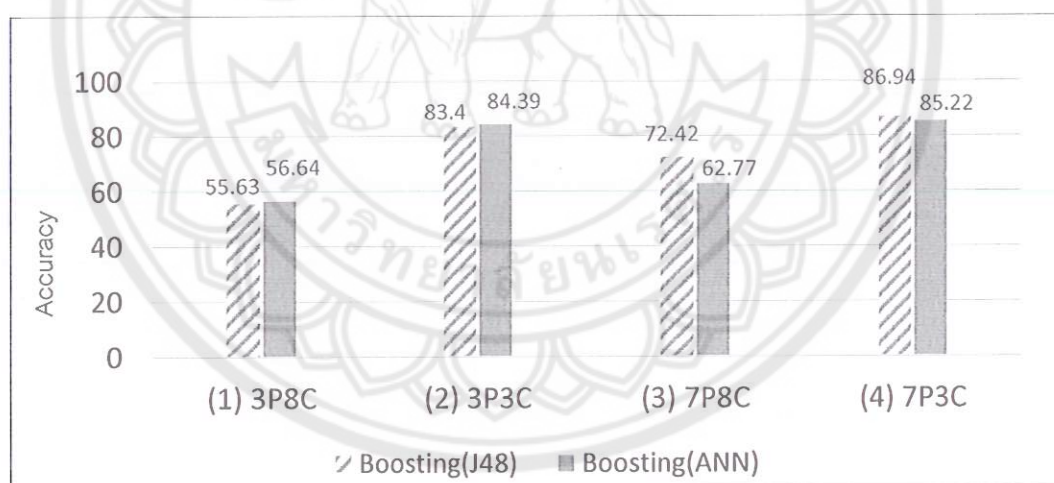
ตาราง 47 ค่า Confusion Matrix วิธีการแบบ Boosting โดยใช้ ANN เป็นพื้นฐาน
การเรียนรู้ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 3

Class		Prediction								
		C1	C2	C3	C4	C5	C6	C7	C8	Recall
Actual	C1	593	75	5	91	26	140	41	14	60.20%
	C2	113	338	1	142	11	47	11	0	50.98%
	C3	55	6	13	12	2	91	18	4	6.47%
	C4	72	95	1	1,113	1	6	3	0	86.21%
	C5	127	9	3	1	46	194	39	22	10.43%
	C6	103	24	8	12	14	906	513	45	55.75%
	C7	39	1	2	5	1	340	1,502	8	79.14%
	C8	4	0	0	1	0	68	77	113	42.97%

ตาราง 48 ค่า Confusion Matrix วิธีการแบบ Boosting โดยใช้ ANN เป็นพื้นฐาน
การเรียนรู้ในการสร้างแบบจำลอง สำหรับการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่ 4

Class		Prediction			
		C1	C2	C3	Recall
Actual	C1	596	210	380	50.25%
	C2	141	1,728	85	88.43%
	C3	215	58	3,954	93.54%

จากการคำนวณหาประสิทธิภาพความถูกต้องที่ได้จากตาราง Confusion Matrix ของข้อมูลกลุ่มตัวอย่างที่ 1-4 สำหรับการจำแนกประเภทชนิดของกลุ่มเมฆในการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่มด้วยวิธี Boosting โดยใช้อัลกอริทึม J48 ในตาราง 40-43 และอัลกอริทึม ANN ในตาราง 45-48 สามารถสรุปผลการทดลองได้ดังภาพ 37



ภาพ 37 แสดงประสิทธิภาพความถูกต้องสำหรับการจำแนกประเภทแบบรวมกลุ่ม
วิธี Boosting

จากภาพ 37 แสดงประสิทธิภาพความถูกต้องที่ได้จากผลการทดลองในการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่มด้วยวิธี Boosting โดยใช้อัลกอริทึม J48 ซึ่งข้อมูลกลุ่มตัวอย่างที่ให้ประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆสูงสุด คือ กลุ่มตัวอย่างที่ 4 อยู่ที่ 86.94% รองลงมาคือ กลุ่มตัวอย่างที่ 2 กลุ่มตัวอย่างที่ 3 และกลุ่มตัวอย่างที่ 1

โดยประสิทธิภาพความถูกต้องอยู่ที่ 83.40%, 72.42% และ 55.63% ตามลำดับ สำหรับการจำแนกประเภทแบบรวมกลุ่มด้วยวิธี Boosting โดยใช้อัลกอริทึม ANN ให้ประสิทธิภาพของแต่ละกลุ่มตัวอย่างมีลักษณะไปในทางเดียวกันกับการใช้อัลกอริทึม J48 คือ กลุ่มตัวอย่างที่ 4 จะให้ประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆสูงที่สุดลดลงมา คือ กลุ่มตัวอย่างที่ 2 กลุ่มตัวอย่างที่ 3 และกลุ่มตัวอย่างที่ 1 โดยประสิทธิภาพความถูกต้องอยู่ที่ 85.22%, 84.39%, 62.77% และ 56.64% ตามลำดับ จากภาพรวมจะเห็นว่าการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่มด้วยวิธี Boosting โดยใช้ J48 เป็นอัลกอริทึมพื้นฐานการเรียนรู้ให้ประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆดีกว่าการใช้อัลกอริทึม ANN

3. ผลลัพธ์จากการวิเคราะห์ข้อมูลด้วยวิธี Random Forest

จากผลการวิเคราะห์ข้อมูลในการจำแนกประเภทชนิดของกลุ่มเมฆด้วยวิธี Random Forest ในการทดลองได้กำหนดค่าพารามิเตอร์ตามที่ได้แสดงในตาราง 49 เพื่อหาค่าพารามิเตอร์ที่เหมาะสมสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆ ดังนี้

ตาราง 49 การกำหนดค่าพารามิเตอร์ที่ใช้ในการทดลองด้วยวิธี Random Forest

พารามิเตอร์	ค่าที่ใช้กำหนดในการทดลอง
Number of Model	50, 100, 150, 200 และ 300
Number of Feature	1, 2, 3, 4 และ 5

ตาราง 49 แสดงค่าพารามิเตอร์ที่ได้กำหนดในการทดลองของวิธี Random Forest โดยในการทดลองได้สร้างต้นไม้ช่วยตัดสินใจจำนวน 50, 100, 150, 200 และ 300 แบบจำลอง ค่าที่ใช้สำหรับพิจารณาการสุ่มเลือกคุณลักษณะอยู่ที่ 1, 2, 3, 4 และ 5 จากผลการทดลองพบว่า การสร้างต้นไม้ช่วยตัดสินใจจำนวน 200 แบบจำลอง โดยค่าสำหรับพิจารณาการสุ่มเลือกคุณลักษณะเท่ากับ 1 ให้ประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆเหมาะสมที่สุด ซึ่งสามารถคำนวณค่าความถูกต้องที่กำหนดค่าพารามิเตอร์อย่างเหมาะสม ได้จากตาราง Confusion Matrix ในตาราง 50-53 ของกลุ่มตัวอย่างข้อมูลที่ 1-4

ตาราง 50 ค่า Confusion Matrix วิธีการแบบ Random Forest สำหรับการวิเคราะห์
ข้อมูลกลุ่มตัวอย่างที่ 1

Class		Prediction								
		C1	C2	C3	C4	C5	C6	C7	C8	Recall
Actual	C1	572	76	9	98	49	149	30	2	58.07%
	C2	104	347	14	160	2	32	2	2	52.34%
	C3	66	13	21	20	10	57	11	3	10.45%
	C4	55	85	1	1,142	1	2	5	0	88.46%
	C5	124	9	8	3	61	162	59	15	13.83%
	C6	145	22	24	13	61	745	595	20	45.85%
	C7	31	2	9	17	17	468	1,336	18	70.39%
	C8	11	6	4	1	13	94	94	40	15.21%

ตาราง 51 ค่า Confusion Matrix วิธีการแบบ Random Forest สำหรับการวิเคราะห์
ข้อมูลกลุ่มตัวอย่างที่ 2

		Prediction			Recall
		C1	C2	C3	
Actual	C1	557	229	400	46.96%
	C2	146	1,748	60	89.46%
	C3	285	61	3,881	91.81%

ตาราง 52 ค่า Confusion Matrix วิธีการแบบ Random Forest สำหรับการวิเคราะห์
ข้อมูลกลุ่มตัวอย่างที่ 3

Class		Prediction								
		C1	C2	C3	C4	C5	C6	C7	C8	Recall
Actual	C1	724	30	4	64	31	105	20	7	73.50%
	C2	62	473	4	87	1	32	4	0	71.34%

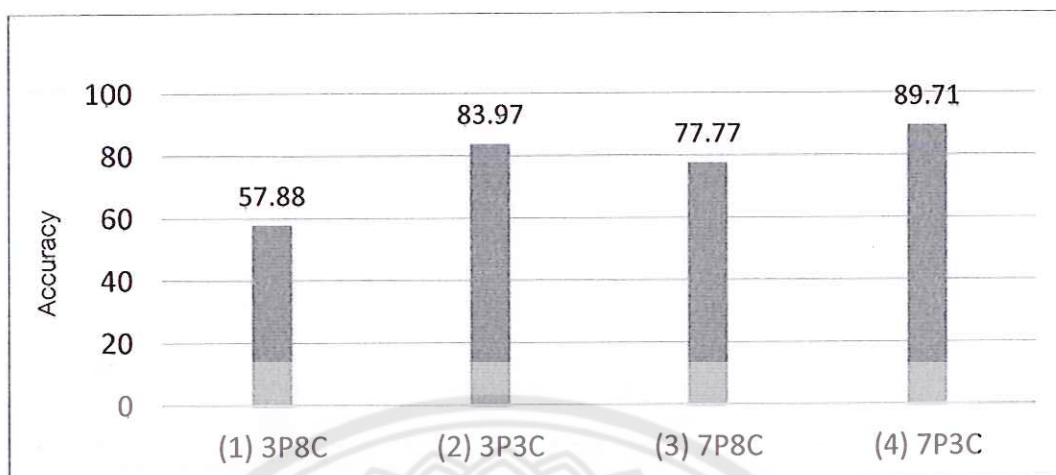
ตาราง 52 (ต่อ)

Class	Prediction								Recall
	C1	C2	C3	C4	C5	C6	C7	C8	
Actual	C3	32	7	92	11	3	42	13	45.77%
	C4	42	25	0	1,219	0	4	0	94.42%
	C5	101	11	1	0	151	127	35	34.24%
	C6	89	12	5	7	29	1,247	225	76.74%
	C7	33	2	1	0	6	202	1,650	86.93%
	C8	2	0	0	2	7	49	30	65.78%

ตาราง 53 ค่า Confusion Matrix วิธีการแบบ Random Forest สำหรับการวิเคราะห์
ข้อมูลกลุ่มตัวอย่างที่ 4

		Prediction			Recall
		C1	C2	C3	
Actual	C1	733	132	321	61.80%
	C2	83	1,814	57	92.84%
	C3	126	39	4,062	96.10%

จากการคำนวณหาประสิทธิภาพความถูกต้องที่ได้จากตาราง Confusion Matrix ของข้อมูลกลุ่มตัวอย่างที่ 1-4 สำหรับการจำแนกประเภทชนิดของกลุ่มเมฆในการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่มโดยใช้วิธี Random Forest สามารถสรุปผลการทดลองได้ดังภาพ 38



ภาพ 38 แสดงประสิทธิภาพความถูกต้องสำหรับการจำแนกประเภทแบบรวมกลุ่ม
วิธี Random Forest

จากภาพ 38 แสดงประสิทธิภาพความถูกต้องที่ได้จากผลการทดลองในการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่มโดยใช้วิธี Random Forest ซึ่งสามารถคำนวณหาประสิทธิภาพความถูกต้องจากตาราง Confusion Matrix ที่ได้กำหนดค่าพารามิเตอร์อย่างเหมาะสม ของกลุ่มตัวอย่างข้อมูล 1-4 ตามตาราง 50-53 โดยประสิทธิภาพความถูกต้องสูงสุดในการจำแนกประเภทชนิดของกลุ่มเมฆ คือ กลุ่มตัวอย่างที่ 4 อยู่ที่ 89.71% รองลงมาคือ กลุ่มตัวอย่างที่ 2 กลุ่มตัวอย่างที่ 3 และกลุ่มตัวอย่างที่ 1 โดยประสิทธิภาพความถูกต้องอยู่ที่ 83.97%, 77.77% และ 57.88% ตามลำดับ

บทที่ 5

บทสรุป

สรุปผลการวิจัย

งานวิจัยในครั้งนี้มีจุดประสงค์เพื่อศึกษาและทำการเปรียบเทียบประสิทธิภาพความถูกต้องของการจำแนกประเภทชนิดของกลุ่มเมฆในการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยว (Single Classification) กับการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม (Multiple Classification) หรือเรียกอีกอย่างหนึ่งว่าการเรียนรู้แบบรวมกลุ่ม (Ensemble Learning) เพื่อหาประสิทธิภาพความถูกต้องสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆที่ดีที่สุด จากผลการวิจัยพบว่า วิธีการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่มสามารถช่วยเพิ่มประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆได้ดีกว่าวิธีการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยว ซึ่งมีความสอดคล้องตรงตามสมมติฐานที่ได้ตั้งไว้ โดยประสิทธิภาพความถูกต้องสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆสูงที่สุดอยู่ที่ 89.71% สำหรับวิธีการของ Random Forest ที่ได้ทำการทดลองในกลุ่มตัวอย่างที่ 4 โดยแบ่งกลุ่มของเมฆออกเป็น 3 กลุ่ม และวิเคราะห์จากคุณลักษณะของเมฆ 7 คุณลักษณะ

อภิปรายผล

จากผลการทดลองข้อมูลกลุ่มตัวอย่างที่ 1-4 พบว่า ตัวอย่างข้อมูลกลุ่มที่ 1 และตัวอย่างข้อมูลกลุ่มที่ 2 ประสิทธิภาพความถูกต้องสำหรับการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยว อัลกอริทึม ANN ให้ประสิทธิภาพดีกว่าอัลกอริทึม J48 โดยประสิทธิภาพอยู่ที่ 56.81% และ 84.35% ตามลำดับ ซึ่งในทางตรงกันข้ามตัวอย่างข้อมูลกลุ่มที่ 3 และตัวอย่างข้อมูลกลุ่มที่ 4 อัลกอริทึม J48 จะให้ประสิทธิภาพดีกว่าอัลกอริทึม ANN โดยประสิทธิภาพอยู่ที่ 68.28% และ 85.99% ตามลำดับ และเมื่อพิจารณาประสิทธิภาพความถูกต้องสำหรับการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่มทั้ง 3 วิธี จะพบว่าโดยส่วนใหญ่แล้ววิธี Random Forest จะให้ประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆดีกว่าวิธีการจำแนกประเภทแบบรวมกลุ่มวิธีการอื่น ๆ โดยประสิทธิภาพความถูกต้องในการจำแนกสูงสุดของวิธี Random Forest ในกลุ่มตัวอย่างที่ 1-4 อยู่ที่ 57.88%, 83.97%, 77.77% และ 89.71% ตามลำดับ ดังตาราง 54 เนื่องจากวิธีการของ Random Forest คล้ายกับวิธีการ Bagging แต่เพิ่มเติมในส่วนของการฟังก์ชัน

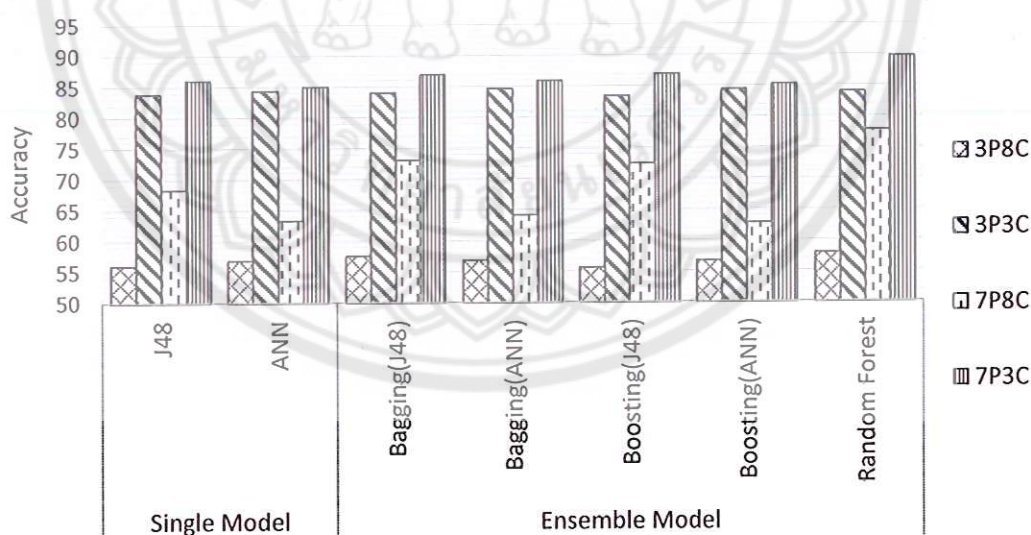
การทำงานแบบสุ่มเลือกคุณลักษณะของข้อมูลที่ใช้ในการวิเคราะห์เข้ามาด้วย ทำให้ลดค่าสหสัมพันธ์ (Correlation) ของคุณลักษณะลงในการสร้างต้นไม้แต่ละต้นที่มีความเป็นอิสระต่อกัน (ค่าสหสัมพันธ์ คือ ค่าความสัมพันธ์ระหว่างตัวแปรตั้งแต่ 2 ตัวขึ้นไป) ยิ่งกำหนดค่าการสุ่มคุณลักษณะของข้อมูลน้อย ก็จะทำให้ค่าสหสัมพันธ์น้อยลงตามเช่นกัน จึงทำให้ต้นไม้ในแต่ละต้นที่ถูกสร้างขึ้นเพื่อใช้สำหรับการจำแนกประเภทข้อมูลมีโครงสร้างต้นไม้ที่มีขนาดเล็ก ทำงานได้เร็ว และให้ประสิทธิภาพดีกว่าวิธีการแบบ Boosting (Han, J., Kamber, M., and Pei, J., 2012) อีกทั้งยังให้ประสิทธิภาพความถูกต้องสำหรับการจำแนกประเภทข้อมูลสูงขึ้นอย่างมีนัยสำคัญเมื่อเทียบกับการจำแนกประเภทแบบเดียวด้วยอัลกอริทึม J48 (Ali, J., Khan, R., Ahmad, N. and Maqsood, I., 2012)

สำหรับการทดลองในงานวิจัยนี้ได้กำหนดค่าที่ใช้สำหรับสุ่มเลือกคุณลักษณะข้อมูลอยู่ที่ 1, 2, 3, 4 และ 5 ตามลำดับ จากผลการทดลองทำให้ทราบว่า ค่าการสุ่มเลือกคุณลักษณะข้อมูลมีค่าต่ำสุดเท่ากับ 1 ให้ความเหมาะสมมากที่สุดสำหรับการสร้างแบบจำลองด้วยวิธี Random Forest ซึ่งมีความสอดคล้องและตรงตามหลักเกณฑ์ดังที่ได้กล่าวไว้ข้างต้น ทั้งนี้วิธีการที่จะทำให้ Random Forest จำแนกข้อมูลได้อย่างมีประสิทธิภาพสูงนั้นไม่ได้ขึ้นอยู่กับข้อกำหนดค่าการสุ่มเลือกคุณลักษณะของข้อมูลเพียงอย่างเดียว แต่อาจขึ้นอยู่กับปัจจัยอื่น ๆ อีกหลายประการ เช่น จำนวนของข้อมูลที่ใช้ในการวิเคราะห์ คลาสคำตอบของข้อมูล การกำหนดพารามิเตอร์สำหรับการสร้างต้นไม้ในแต่ละต้น จำนวนของต้นไม้ที่สร้างขึ้น เป็นต้น

ตาราง 54 ผลการทดลองสำหรับการวัดประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆทั้งหมด

Cloud Dataset	Single Model (%)		Ensemble Model (%)				
	J48	ANN	Bagging		Boosting		Random Forest
			J48	ANN	J48	ANN	
(1) 3P8C	55.98	56.81	57.53	56.86	55.63	56.64	<u>57.88</u>
(2) 3P3C	83.77	84.35	84.04	<u>84.58</u>	83.40	84.39	83.97
(3) 7P8C	68.28	63.27	73.03	64.08	72.42	62.77	<u>77.77</u>
(4) 7P3C	85.99	84.97	86.91	85.84	86.94	85.22	<u>89.71</u>

จากผลการทดลองในตาราง 54 ทำให้ทราบว่า การแบ่งประเภทของเมฆออกเป็น 3 กลุ่ม โดยใช้ 7 คุณลักษณะให้ความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆสูงที่สุด และเมื่อเปรียบเทียบระหว่างอัลกอริทึมและวิธีการที่ใช้สำหรับการจำแนกประเภทชนิดของกลุ่มเมฆโดยใช้แบบจำลองการจำแนกประเภทแบบเดียวกับแบบจำลองการจำแนกประเภทแบบรวมกลุ่มพบว่า แบบจำลองการจำแนกประเภทแบบรวมกลุ่ม ให้ประสิทธิภาพความถูกต้องดีกว่าแบบจำลองการจำแนกประเภทแบบเดี่ยว สามารถแสดงได้ดังกราฟในภาพ 39 ซึ่งเมื่อพิจารณาตัวอย่างข้อมูลกลุ่มที่ 1 (3P8C) และตัวอย่างข้อมูลกลุ่มที่ 2 (3P3C) จะเห็นว่าประสิทธิภาพความถูกต้องในการจำแนกประเภทชนิดของกลุ่มเมฆระหว่างการจำแนกประเภทแบบเดี่ยวกับการจำแนกประเภทแบบรวมกลุ่มให้ประสิทธิภาพความถูกต้องเพิ่มขึ้นไม่แตกต่างกันมาก แต่เมื่อพิจารณาตัวอย่างข้อมูลกลุ่มที่ 3 (7P8C) และตัวอย่างข้อมูลกลุ่มที่ 4 (7P3C) ประสิทธิภาพจะแตกต่างกันอย่างชัดเจน ทั้งนี้อาจขึ้นอยู่กับคุณลักษณะของข้อมูลที่ใช้ในการวิเคราะห์ที่มีจำนวนมากประกอบกับมีคลาสคำตอบจำนวนน้อยทำให้ส่งผลดีต่อประสิทธิภาพในการจำแนกชนิดของกลุ่มเมฆเพิ่มสูงขึ้น และในทางตรงกันข้ามหากจำนวนคุณลักษณะของข้อมูลที่ใช้ในการวิเคราะห์มีจำนวนน้อยและคลาสคำตอบมีจำนวนมาก อาจส่งผลกระทบต่อประสิทธิภาพในการจำแนกชนิดของกลุ่มเมฆให้ต่ำลงได้



ภาพ 39 แสดงผลการเปรียบเทียบประสิทธิภาพความถูกต้องสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆ

จากภาพ 39 แสดงให้เห็นว่าประสิทธิภาพความถูกต้องสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆโดยรวมของกลุ่มตัวอย่างที่ใช้ในการทดลองทั้งหมด 4 ชุด ซึ่งตัวอย่างข้อมูลกลุ่มที่ 4 แบ่งชนิดเมฆออกเป็น 3 กลุ่มและใช้คุณลักษณะของเมฆ 7 คุณลักษณะสำหรับการวิเคราะห์ มีประสิทธิภาพความถูกต้องสูงที่สุด เนื่องจากมีจำนวนคุณลักษณะที่ใช้ในการวิเคราะห์มากกว่าประกอบกับมีจำนวนของคลาสคำตอบ (ชนิดของกลุ่มเมฆ) มีน้อย และเมื่อพิจารณาจากกลุ่มตัวอย่างที่ 4 เพียงกลุ่มตัวอย่างเดียวสำหรับการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม จะเห็นว่าวิธี Random Forest ให้ประสิทธิภาพความถูกต้องสูงสุดอยู่ที่ 89.71% รองลงมา คือ Boosting (J48), Bagging (J48), Bagging (ANN) และ Boosting (ANN) โดยประสิทธิภาพอยู่ที่ 86.94%, 86.91%, 85.84% และ 85.22% ตามลำดับ

จากผลสรุปการทดลองดังกล่าวข้างต้นเมื่อใช้สถิติการทดสอบค่า T-test เพื่อหาความสัมพันธ์ของประสิทธิภาพในการจำแนกประเภทแบบเดี่ยวและการจำแนกประเภทแบบรวมกลุ่ม จากผลการทดสอบพบว่า การจำแนกประเภทแบบรวมกลุ่มวิธี Bagging(J48), Boosting(J48), Boosting(ANN) และ Random Forest สามารถเพิ่มประสิทธิภาพความถูกต้องให้กับการจำแนกประเภทชนิดของกลุ่มเมฆได้ดีกว่าการจำแนกประเภทแบบเดี่ยวโดยใช้อัลกอริทึม J48 อย่างมีนัยสำคัญทางสถิติที่ระดับความเชื่อมั่น 95% (ค่า Sig < 0.05) ดังตาราง 55 และเมื่อใช้วิธีการจำแนกประเภทแบบรวมกลุ่มเปรียบเทียบกับวิธีการจำแนกประเภทแบบเดี่ยวโดยใช้อัลกอริทึม ANN พบว่า การจำแนกประเภทแบบรวมกลุ่มวิธี Bagging(J48), Bagging(ANN), Boosting(J48) และ Random Forest สามารถเพิ่มประสิทธิภาพความถูกต้องให้กับการจำแนกประเภทชนิดของกลุ่มเมฆได้ดีกว่าการจำแนกประเภทแบบเดี่ยวโดยใช้อัลกอริทึม ANN อย่างมีนัยสำคัญทางสถิติที่ระดับความเชื่อมั่น 95% ดังตาราง 56 สามารถแสดงรายละเอียดการทดสอบค่า T-test ได้ในภาคผนวก ค ซึ่งโดยภาพรวมจากผลการทดสอบค่า T-test เพื่อหาประสิทธิภาพสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆมีความสอดคล้องเป็นไปตามสมมติฐานที่ตั้งไว้ คือ เมื่อนำเทคนิควิธีการเรียนรู้แบบรวมกลุ่มมาใช้สำหรับการจำแนกประเภทชนิดของกลุ่มเมฆ จะสามารถทำให้ประสิทธิภาพความถูกต้องและแม่นยำในการจำแนกประเภทชนิดของกลุ่มเมฆเพิ่มสูงขึ้น

ตาราง 55 การทดสอบค่า T-test โดยใช้อัลกอริทึม J48 เพื่อเปรียบเทียบประสิทธิภาพกับ
การจำแนกประเภทแบบรวมกลุ่ม

J48 compare	Paired Differences				t	Sig. (2-tailed)
	Mean	Std. Deviation	95% Confidence			
			Lower	Upper		
Bagging (J48)	-1.87	1.77	-2.7	-1.05	-4.74	*0
Bagging (ANN)	0.66	2.14	-0.33	1.66	1.39	0.18
Boosting (J48)	-1.09	1.89	-1.98	-0.21	-2.59	*0.018
Boosting (ANN)	1.25	2.59	0.04	2.46	2.16	*0.044
Random Forest	-3.83	3.59	-5.51	-2.15	-4.77	*0

ตาราง 56 การทดสอบค่า T-test โดยใช้อัลกอริทึม ANN เพื่อเปรียบเทียบประสิทธิภาพกับ
การจำแนกประเภทแบบรวมกลุ่ม

ANN compare	Paired Differences				t	Sig. (2-tailed)
	Mean	Std. Deviation	95% Confidence			
			Lower	Upper		
Bagging (J48)	-3.03	4.07	-4.93	-1.12	-3.33	*0.004
Bagging (ANN)	-0.49	0.37	-0.66	-0.32	-5.99	*0
Boosting (J48)	-2.25	4.28	-4.25	-0.24	-2.35	*0.03
Boosting (ANN)	0.09	0.28	-0.04	0.23	1.49	0.151
Random Forest	-4.98	5.95	-7.77	-2.2	-3.74	*0.001

ข้อเสนอแนะ

จากการศึกษา ทดลอง และการวิเคราะห์หาผลลัพธ์สำหรับกระบวนการจำแนกประเภท
ชนิดของกลุ่มเมฆ ผู้วิจัยพบปัญหาและได้มีข้อเสนอแนะเพิ่มเติมดังต่อไปนี้

1. จากปัญหาที่พบในขั้นตอนของการเตรียมข้อมูลเบื้องต้น เนื่องจากว่าข้อมูลที่นำมาใช้
สำหรับการทดลองในงานวิจัยนี้ยังคงมีปัญหาเกี่ยวกับความไม่สมดุลกันของจำนวนตัวอย่างข้อมูล

(Imbalance Data) ซึ่งเป็นปัญหาที่เกิดขึ้นกับข้อมูลทุกประเภท รวมถึงปัญหาข้อมูลที่มี Noise อาจทำให้กระบวนการสำหรับการจำแนกประเภทของข้อมูลไม่มีประสิทธิภาพเท่าที่ควร ดังนั้นหากผู้วิจัยศึกษาหาวิธีสำหรับการกำจัด Noise ของข้อมูลได้อย่างเหมาะสม อาจทำให้ประสิทธิภาพในการจำแนกประเภทชนิดของกลุ่มเมฆเพิ่มสูงขึ้นไม่มากนักน้อย ทั้งนี้ขึ้นอยู่กับองค์ประกอบหลาย ๆ อย่าง เช่น จำนวนตัวอย่างข้อมูล คุณลักษณะที่นำมาใช้สำหรับการวิเคราะห์ข้อมูล ความหลากหลายของคลาสคำตอบ การกำหนดค่าพารามิเตอร์ให้เหมาะสมกับอัลกอริทึมที่ใช้สำหรับการจำแนกประเภทข้อมูล เป็นต้น

2. การวิจัยนี้มุ่งเน้นที่การสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม โดยใช้วิธี Bagging, Boosting และ Random Forest ซึ่งนอกเหนือจากที่กล่าวมาแล้วยังมีวิธีการอื่น ๆ อีกมาก สำหรับการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม ตัวอย่าง เช่น Stacking, Voting, Random Subspace และ Hybrid Experts เป็นต้น

3. การนำอัลกอริทึมใหม่ ๆ มาประยุกต์ใช้สำหรับกระบวนการในการจำแนกประเภทของข้อมูล สามารถเพิ่มประสิทธิภาพความถูกต้องให้กับการจำแนกประเภทได้ เช่น อัลกอริทึมโครงข่ายประสาทเทียมแบบ Deep Learning การใช้งาน Genetic Algorithm (GA) มาประยุกต์ใช้ในการพิจารณาในการคัดเลือกคุณลักษณะของข้อมูล (Feature Selection) หรือการลดมิติของข้อมูลลง รวมไปถึงการนำเอาอัลกอริทึมแบบผสมผสาน (Hybrid) มาประยุกต์ใช้สำหรับกระบวนการเรียนรู้ข้อมูลเพื่อสร้างแบบจำลองในการเพิ่มประสิทธิภาพให้กับการจำแนกประเภทของข้อมูล



บรรณานุกรม

- ถิรนนท์ สอนแก้ว. (2558). การจำแนกชนิดของเมฆโดยใช้ข้อมูลการแผ่รังสีดวงอาทิตย์จากหอดูดาวแห่งชาติ ดอยอินทนนท์. เชียงใหม่: สถาบันดาราศาสตร์แห่งชาติ (องค์การมหาชน).
- Alam, M. S., & Vuong, S. T. (2013, August). Random forest classification for detecting android malware. In *Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCoM), IEEE International Conference on and IEEE Cyber, Physical and Social Computing* (pp. 663-669). Beijing: Beihang University.
- Ali, J., Khan, R., Ahmad, N., & Maqsood, I. (2012). Random Forests and Decision Trees. *IJCSI International Journal of Computer Science Issues*, 9(5), 272- 278.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123-140.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32.
- Chauvin, R., Nou, J., Thil, S., Traore, A., & Grieu, S. (2015). Cloud detection methodology based on a sky-imaging system. *Energy Procedia*, 69, 1970-1980.
- Chethan, H.K., Raghavendra, R., & Kumar, G.H. (2009). Texture based approach for cloud classification using SVM. In *Advances in Recent Technologies in Communication and Computing, 2009. ARTCom'09. International Conference on* (pp. 688-690). Kerala: Kottayam.
- Christodoulos, I., Christodoulou, Michaelides, S.C., & Pattichis, C.S. (2003). Multifeature Texture Analysis for the Classification of Clouds in Satellite Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 41(11), 2662-2668.
- Duch, W., & Jankowski, N. (1999). Survey of neural transfer functions. *Neural Computing Surveys*, 2(1), 163-212.
- Freund, Y., & Schapire, R. E. (1996, July). Experiments with a new boosting algorithm. In *icml*, 96, 148-156.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques* (3rd ed.). Edition. USA: 225 Wyman Street, Waltham, MA 02451.

- Heinle, A., Macke, A., & Srivastav, A. (2010). Automatic cloud classification of whole sky images. *Atmospheric Measurement Techniques*, 3(3), 557–567.
- Jin W., Wang, L., Zeng, X., Liu, Z., & Fu, R. (2014). Classification of clouds in satellite imagery using over-complete dictionary via sparse representation. *Pattern Recognition Letters*, 49, 193-200.
- Kantardzic, M. (2011). *Data mining: concepts, models, methods, and algorithms*. Canada: John Wiley & Sons, Inc., Hoboken, New Jersey.
- Kaur, D., Bedi, R., & Gupta, S.K. (2015). Review of Decision Tree Data Mining Algorithm: ID3 and C4.5. In *Proceedings of International Conference on Information Technology and Computer Science 2015*. Bali: Inna Grand Kuta Bali Hotel.
- Krenker, A., Bester, J., & Kos, A. (2011). Introduction to the Artificial Neural Networks. In Kenji Suzuki (Ed.), *Artificial Neural Networks-Methodological Advances and Biomedical Applications*. China: InTech China, Unit 405, Office Block, Hotel Equatorial Shanghai.
- Kuril, S., Saini, I., & Saini, B. S. (2013). Cloud classification for weather information by artificial neural network. *International Journal of Applied Physics and Mathematics*, 3(1), 28-30.
- Larose, D. T. (2014). *Discovering knowledge in data: an introduction to data mining*. Canada: John Wiley & Sons.
- Liu, Y., Xia, J., Shi, C.X., & Hong, Y. (2009). An Improved Cloud Classification Algorithm for China's FY-2C Multi-Channel Images Using Artificial Neural Network. *Journal Sensors* 2009, 9(7), 5558-5579.
- Mahrooghy, M., Younan, N. H., Anantharaj, V. G., & Aanstoos, J. (2011, July). High resolution satellite precipitation estimate using cluster ensemble cloud classification. In *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International* (pp. 2645-2648). British Columbia: Vancouver Convention Center.

- Meshram, S.B., & Shinde, S.M. (2015). A Survey on Ensemble Methods for High Dimensional Data Classification in biomedicine Field. *International Journal of Computer Applications*, 111(11), 5-7.
- Mittal, P., & Gill, N. S. (2014). A comparative analysis of classification techniques on medical data sets. *International Journal of Research in Engineering and Technology*, 3(06), 454-460.
- Nelson, A. L. (2014). *Introduction to Artificial Neural Networks*. [PowerPoint slides]. Retrieved January 23, 2016, from http://www.powershow.com/view/3c778f-N2YzM/Introduction_to_Artificial_Neural_Networks_powerpoint_ppt_presentation.
- Ozkan, C., & Erbek, F.S. (2003). The Comparison of Activation Functions or Multispectral Landsat TM Image Classification. *Photogrammetric Engineering & Remote Sensing*, 69(11), 1225-1234.
- Pallavi, V. P., & Vaithiyanathan, V. (2013). Combined Artificial Neural Network and Genetic Algorithm for Cloud Classification. *International Journal of Engineering and Technology*, 5(2), 787-794.
- Pour, S. G., McLeod, P., Verma, B., & Maeder, A. (2012). Comparing Data Mining with Ensemble Classification of Breast Cancer Masses in Digital Mammograms. In *The Second Australian Workshop on Artificial Intelligence in Health AIH 2012* (p. 55). Sydney: Sydney Harbour Marriott Hotel.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1), 81-106.
- Quinlan, J. R. (1993). *C4.5: programs for machine learning*. San Mateo, CA: Morgan Kaufmann.
- Richard, J. R., & Michael, W. G. (2003). *Data Mining: A tutorial-based primer*. NY: Pearson Education.
- Sewell, M. (2007). *Ensemble Learning*. London: Department of Computer Science, University College London.
- Sheela, K.G., & Deepa, S.N. (2014). Review on Methods to Fix Number of Hidden Neurons in Neural Networks. *Journal of Scientific & Industrial Research (JSIR)*, 73(10), 686-688.

- Shi, L., Ywang, Q., Ma, X., Weng, M., & Qiao, H. (2012). Spam Email Classification Using Decision Tree Ensemble. *Journal of Computational Information Systems*, 8(3), 949–956.
- Taravat, A., Del Frate, F., Cornaro, C., & Vergari, S. (2015). Neural Networks and Support Vector Machine Algorithms for Automatic Cloud Classification of Whole-Sky Ground-Based Images. *IEEE Geoscience and Remote Sensing Letters*, 12(3), 666–670.
- Tian, B., Shaikh, M. A., Azimi-Sadjadi, M. R., Haar, T. H. V., & Reinke, D. L. (1999). A study of cloud classification with neural networks using spectral and textural features. *IEEE Transactions on Neural Networks*, 10(1), 138–151.
- Tiwari A., & Prakash A. (2014). Improving classification of J48 algorithm using bagging, boosting and blending ensemble methods on SONAR dataset using WEKA. *International Journal of Engineering and Technical Research*, 2(9), 207–209.
- World Meteorological Organization. (1975). *International cloud atlas* (Vol. 1). Geneva: World Meteorological Organization.
- Yang, P., Hwa Yang, Y., B Zhou, B., & Y Zomaya, A. (2010). A review of ensemble methods in bioinformatics. *Current Bioinformatics*, 5(4), 296–308.
- Zhen, Z., Wang, F., Sun, Y., Mi, Z., Liu, C., Wang, B., & Lu, J. (2015, February). SVM based cloud classification model using total sky images for PV power forecasting. In *Innovative Smart Grid Technologies Conference (ISGT), 2015 IEEE Power & Energy Society* (pp. 1–5). Washington: Grand Hyatt Washington.

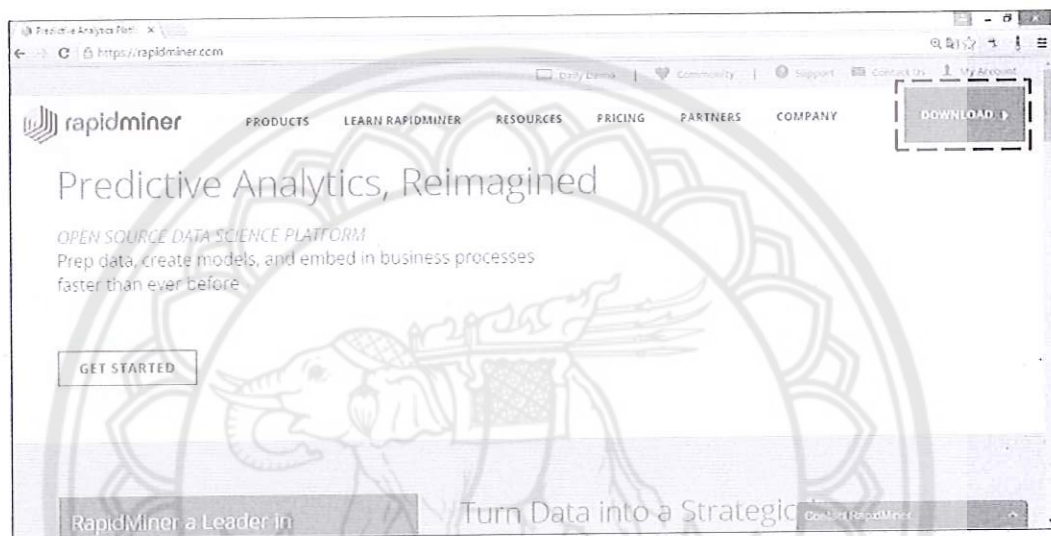


ภาคผนวก

มหาวิทยาลัยนเรศวร

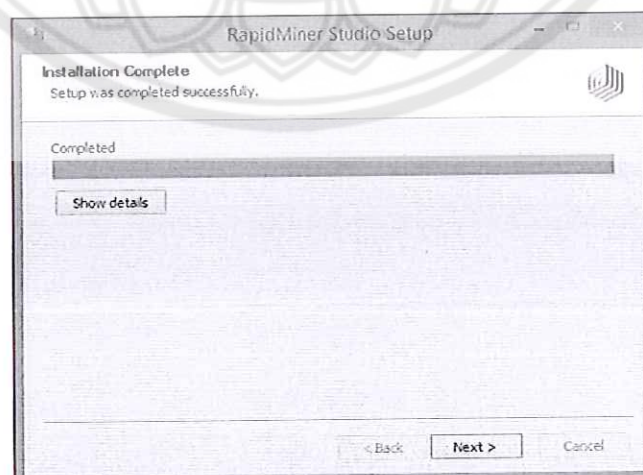
ภาคผนวก ก การดาวน์โหลด และติดตั้งโปรแกรม Rapidminer Studio

Rapidminer Studio เป็นซอฟต์แวร์ประเภทโอเพนซอร์สที่ใช้ในการวิเคราะห์ข้อมูลในด้านการทำเหมืองข้อมูลหรือการวิเคราะห์ทางสถิติ สำหรับการดาวน์โหลดและติดตั้งโปรแกรม Rapidminer Studio สามารถดาวน์โหลดได้ที่เว็บไซต์ www.rapidminer.com แสดงดังภาพ 40



ภาพ 40 แสดงเว็บไซต์สำหรับดาวน์โหลดโปรแกรม Rapidminer Studio

เข้าเว็บไซต์ www.rapidminer.com คลิก DOWNLOAD โปรแกรม และทำการติดตั้งโปรแกรม

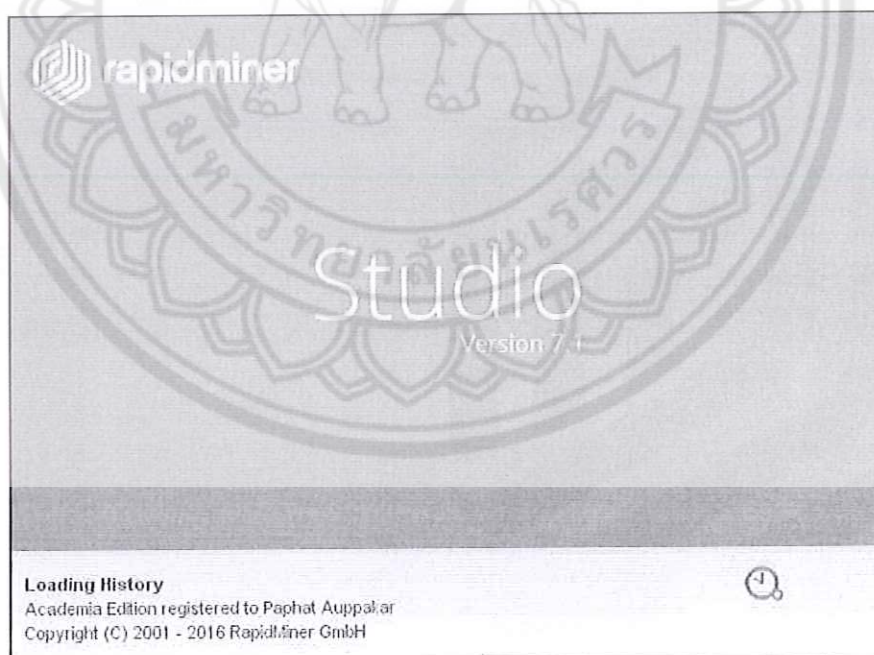


ภาพ 41 แสดงขั้นตอนระหว่างรอการติดตั้งโปรแกรม Rapidminer Studio

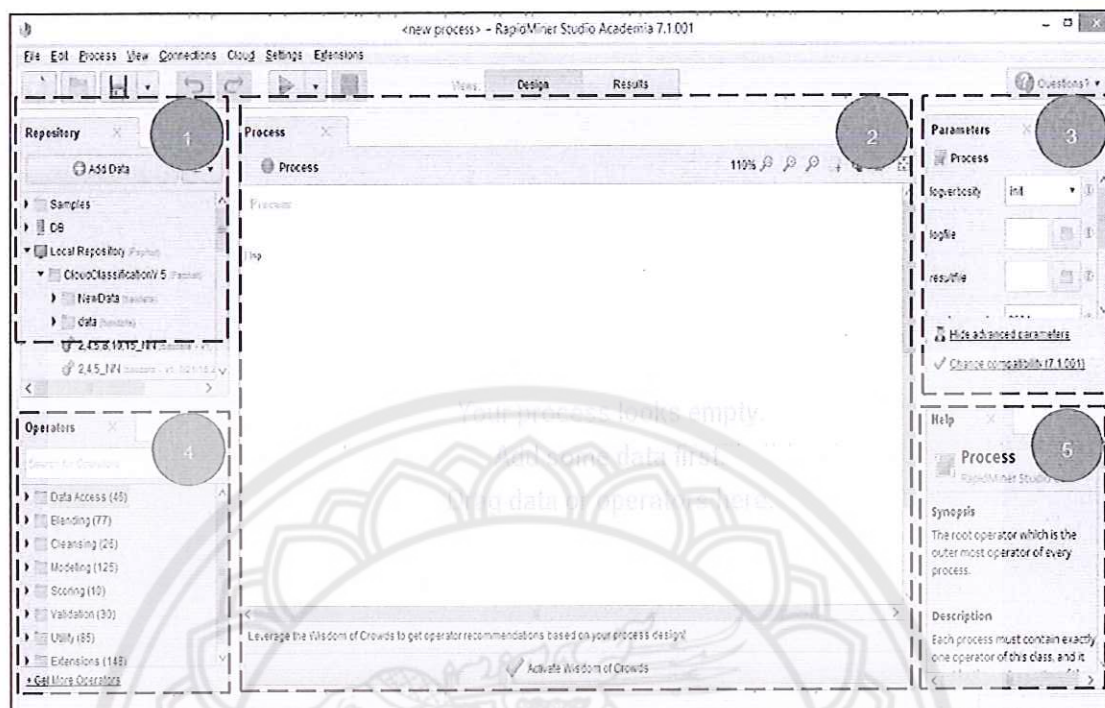


ภาพ 42 แสดงขั้นตอนเมื่อทำการติดตั้งโปรแกรม Rapidminer Studio เสร็จสมบูรณ์

เมื่อทำการติดตั้งเสร็จ โปรแกรมจะทำการโหลดหน้าต่างเริ่มต้นของโปรแกรมขึ้นมา ดังภาพ 43



ภาพ 43 แสดงหน้าต่างเริ่มต้นของโปรแกรม



ภาพ 44 แสดงหน้าเริ่มต้นการทำงานของโปรแกรม

จากภาพ 44 เป็นหน้าเริ่มต้นการทำงานของโปรแกรม โดยแบ่งการทำงานออกเป็น 5 ส่วนหลัก สามารถอธิบายความหมายของแต่ละส่วนได้ดังนี้

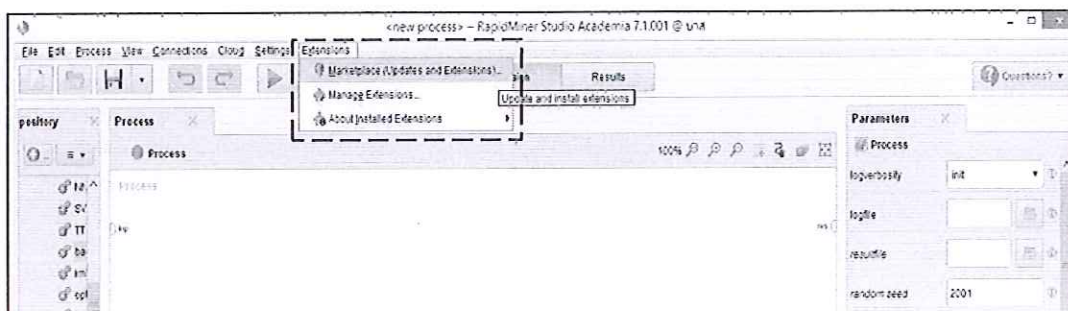
ส่วนที่ 1 Repository คือ ส่วนที่มีหน้าที่เก็บประวัติไฟล์การทำงาน หรือโปรเจกงานที่ได้ออกแบบและสร้างขึ้น

ส่วนที่ 2 Process คือ ส่วนที่มีหน้าที่เอาไว้สร้างงานหรือออกแบบงาน โดยนำเอาเครื่องมือในส่วนของ Operators มาใช้งาน

ส่วนที่ 3 Parameters คือ ส่วนที่มีไว้กำหนดค่าตัวแปรของตัวดำเนินการต่าง ๆ ที่มีอยู่ในส่วนของ Operators

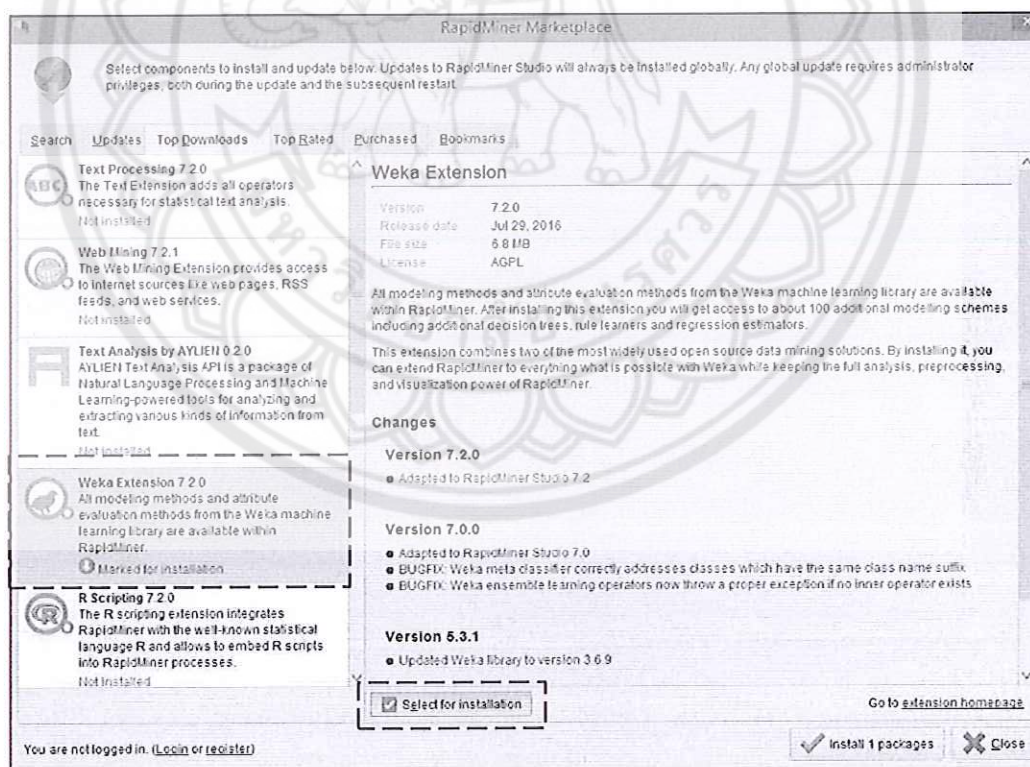
ส่วนที่ 4 Operators คือ ส่วนที่เก็บรวบรวมตัวดำเนินการต่าง ๆ หรืออัลกอริทึมต่าง ๆ เช่น ตัวดำเนินการสำหรับกรองข้อมูล ตัวดำเนินการสำหรับการสร้างแบบจำลองสำหรับการวิเคราะห์ข้อมูล ฯลฯ เป็นต้น

ส่วนที่ 5 Help คือ ตัวช่วยสำหรับอธิบายความหมายของตัวดำเนินการที่มีอยู่ในส่วนของ Operator รวมไปถึงการอธิบายลักษณะตัวแปรต่าง ๆ ที่ผู้ใช้กำหนด

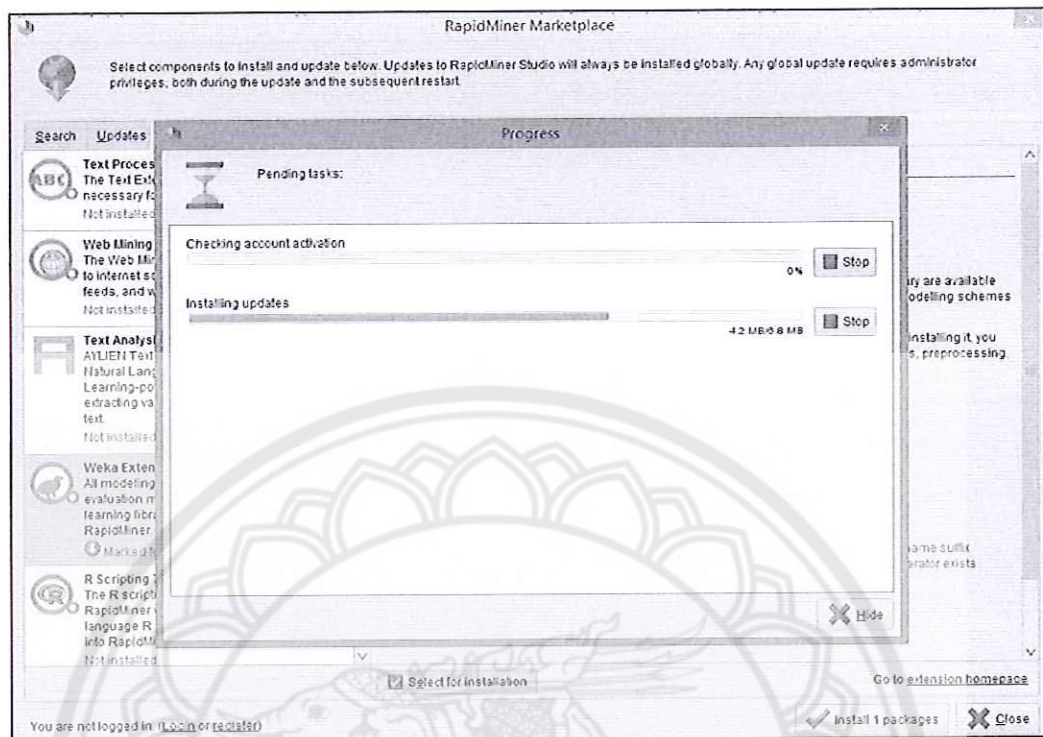


ภาพ 45 แสดงขั้นตอนการติดตั้ง Extensions Packet สำหรับ Rapidminer

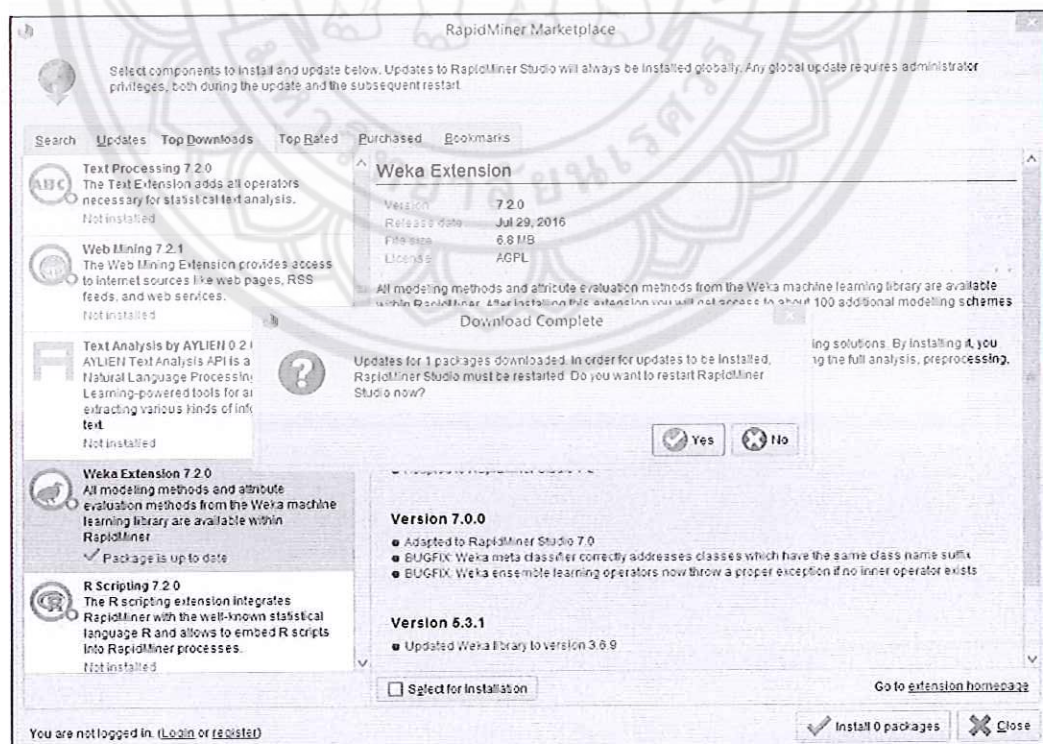
จากภาพ 45 เป็นขั้นตอนการติดตั้งหรือการนำเอา Packet อื่น ๆ มาใช้งานร่วมกับ Rapidminer Studio สำหรับในงานวิจัยนี้ได้นำ Weka Extension มาใช้งานร่วมกับโปรแกรมด้วย ดังภาพ 46-48 โดยอันดับแรกเข้าไปที่แท็บ Extensions แล้วเลือก Marketplace (Updates and Extensions) จากนั้นเข้าไปที่แท็บ Top Downloads แล้วเลือก Weka Extension เพื่อทำการติดตั้ง



ภาพ 46 แสดงขั้นตอนการติดตั้ง Weka Extension Packages



ภาพ 47 แสดงขั้นตอนการระหว่างติดตั้งโปรแกรม Weka Extension Packages



ภาพ 48 แสดงการติดตั้งโปรแกรม Weka Extension Packages ที่เสร็จสมบูรณ์

จากภาพ 48 เมื่อได้ทำการติดตั้งโปรแกรมอย่างสมบูรณ์แล้วให้ทำการคลิก Yes เพื่อเริ่มสแตร์ทโปรแกรมใหม่ จากนั้นจะสังเกตเห็นว่ามีสัญลักษณ์ของ Weka ปรากฏขึ้น ดังภาพ 49

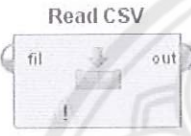
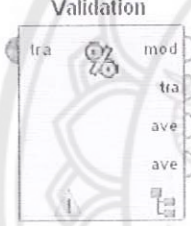
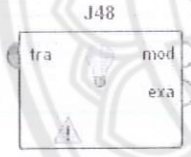
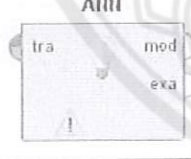
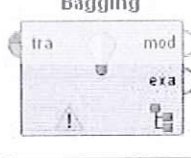
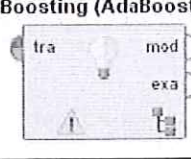


ภาพ 49 แสดงหน้าเริ่มต้นของโปรแกรมที่ได้ทำการติดตั้ง Weka Extension

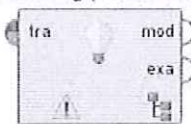
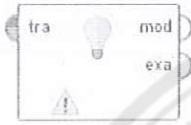
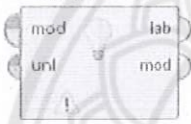
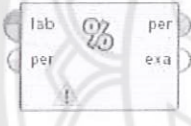
ภาคผนวก ข การออกแบบการทดลองงานวิจัยโดยใช้โปรแกรม Rapidminer Studio

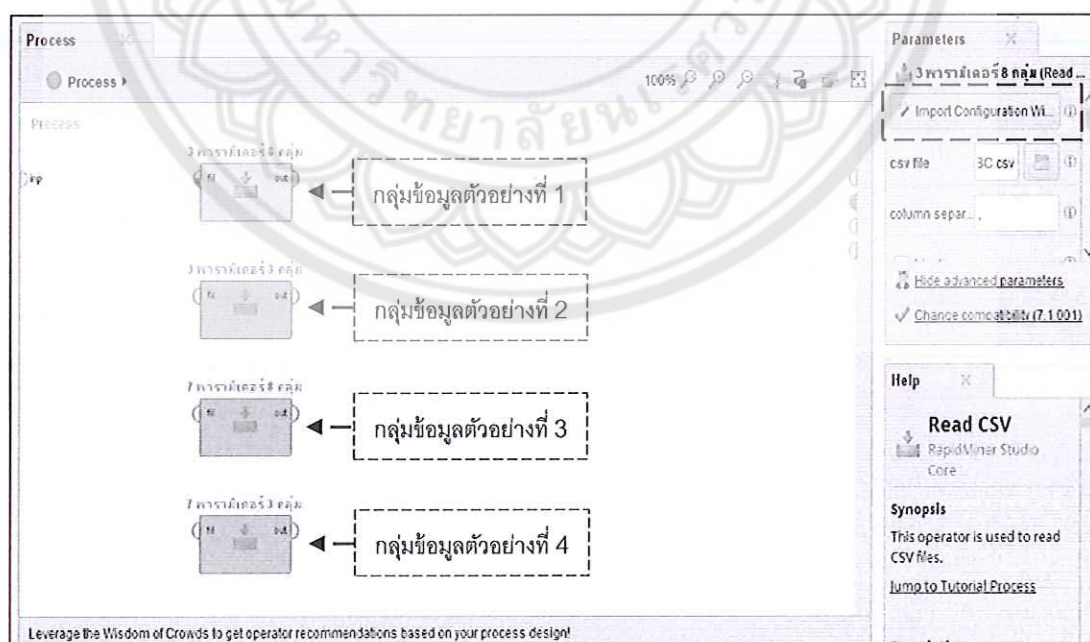
สำหรับการออกแบบและทดลองในงานวิจัย ผู้วิจัยได้ใช้ตัวดำเนินการ (Operator) ในการออกแบบ และวิเคราะห์ข้อมูลสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆ ดังตาราง 57

ตาราง 57 ตัวดำเนินการที่ใช้สำหรับการทดลองในงานวิจัย

Operator	ความหมาย
	ตัวดำเนินการที่ใช้สำหรับอ่านไฟล์เข้ามา เพื่อใช้ในการวิเคราะห์ข้อมูล
	ตัวดำเนินการที่ใช้สำหรับแบ่งข้อมูลเพื่อใช้ในการทดสอบประสิทธิภาพของข้อมูล หรือการทำ Cross Validation
	ตัวดำเนินการที่ใช้สำหรับสร้างแบบจำลองโดยใช้หลักการเรียนรู้แบบต้นไม้ช่วยตัดสินใจ J48 (สำหรับการสร้างแบบจำลองแบบเดียว)
	ตัวดำเนินการที่ใช้สำหรับสร้างแบบจำลองโดยใช้หลักการเรียนรู้แบบโครงข่ายประสาทเทียม (สำหรับการสร้างแบบจำลองแบบเดียว)
	ตัวดำเนินการที่ใช้สำหรับสร้างแบบจำลองโดยใช้หลักการเรียนรู้แบบรวมกลุ่มด้วยวิธี Bagging (สำหรับการสร้างแบบจำลองแบบรวมกลุ่ม)
	ตัวดำเนินการที่ใช้สำหรับสร้างแบบจำลองโดยใช้หลักการเรียนรู้แบบรวมกลุ่มด้วยวิธี Boosting (สำหรับการสร้างแบบจำลองแบบรวมกลุ่ม)

ตาราง 57 (ต่อ)

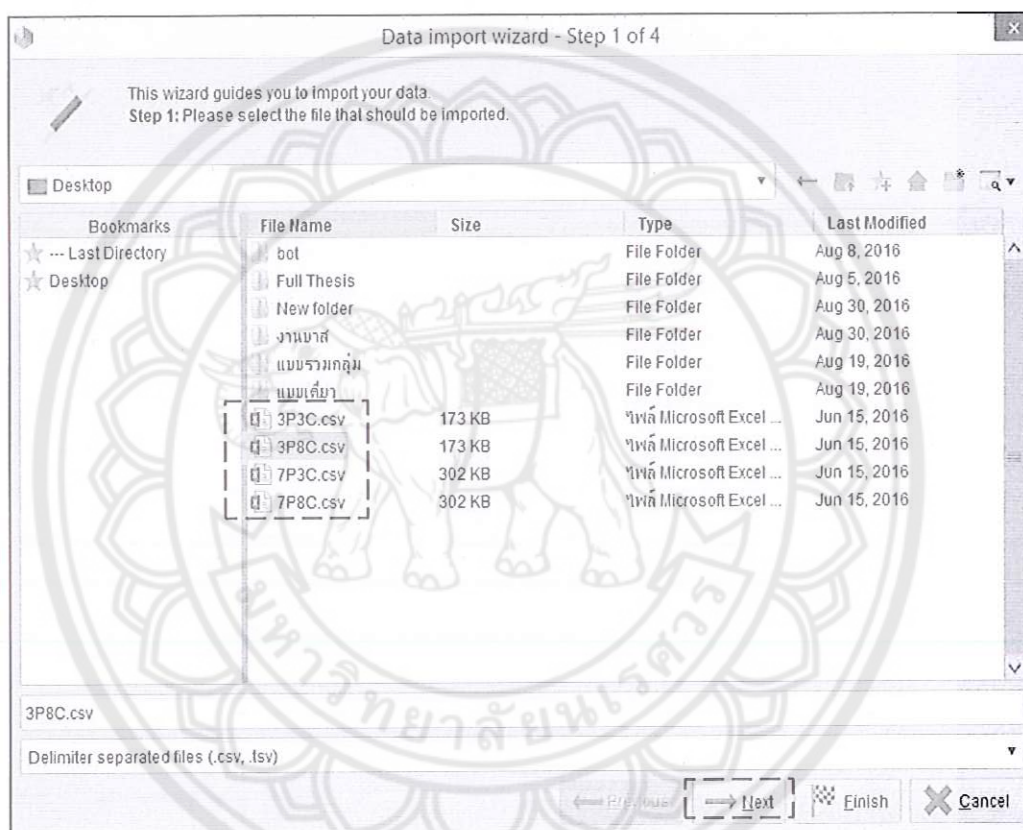
Operator	ความหมาย
Boosting (AdaBoost) 	ตัวดำเนินการที่ใช้สำหรับสร้างแบบจำลองโดยใช้หลักการเรียนรู้แบบรวมกลุ่มด้วยวิธี Boosting (สำหรับการสร้างแบบจำลองแบบรวมกลุ่ม)
Random Forest 	ตัวดำเนินการที่ใช้สำหรับสร้างแบบจำลองโดยใช้หลักการเรียนรู้แบบรวมกลุ่มด้วยวิธี Random Forest (สำหรับการสร้างแบบจำลองแบบรวมกลุ่ม)
Apply Model 	ตัวดำเนินการที่ใช้สำหรับการนำแบบจำลองที่สร้างขึ้นไปใช้งานจริง หรือ ใช้สำหรับขั้นตอนในการทดสอบแบบจำลองที่สร้างขึ้นในแต่ละรอบ ที่ได้จากกระบวนการเรียนรู้ข้อมูล
Performance 	ตัวดำเนินการสำหรับแสดงผลประสิทธิภาพของแบบจำลองที่สร้างขึ้น



ภาพ 50 แสดงการใช้งานตัวดำเนินการสำหรับข้อมูลนำเข้า

จากภาพ 50 แสดงการเริ่มต้นของการออกแบบการทดลอง โดยเริ่มจากการดึงข้อมูล (Import Data) ที่เราได้เตรียมไว้เข้ามาในโปรแกรม ซึ่งมีรายละเอียดดังต่อไปนี้

นำตัวดำเนินการ Read CSV เข้ามายังหน้าต่างโปรเซสในโปรแกรม Rapidminer แล้วคลิกที่ปุ่ม Import Configuration Wizard... จะปรากฏหน้าต่างดังภาพ 51 จากนั้นทำการเลือกไฟล์ข้อมูลที่เราได้เตรียมไว้และกดปุ่ม Next จะปรากฏหน้าต่างดังภาพ 52



ภาพ 51 แสดงการเลือกไฟล์ข้อมูลสำหรับข้อมูลนำเข้า

ทำการเลือก "Comma" ตามที่ได้แสดงในส่วนของ Column Space จากนั้นทำการกดปุ่ม Next จะแสดงดังภาพ 53 ซึ่งเป็นหน้าจอที่ใช้สำหรับแสดงผลของข้อมูลที่เรานำเข้า โดยจะแสดงจำนวนและชื่อแอตทริบิวต์ของข้อมูล

Data import wizard - Step 2 of 4

This wizard guides you to import your data.
Step 2: Please specify how the file should be parsed and how columns are separated.

File Reading

File Encoding: x-windows-874

☐ Trim Lines

☐ Skip Comments: #

Column Separation

☒ Comma "," ☐ Space

☐ Semicolon ";" ☐ Tab

☐ Regular Expression: /s|/t/

Escape Character: \

☒ Use Quotes: "

AV	SD	CC	Class
0.29	0.0606714	90.954	1
1.19	0.0790071	25.427	2
0.4	0.225714	55.803	3
1.02	0.0270143	50.519	4
0.62	0.291229	80.8635	5

Row, Column	Error	Original value	Message
-------------	-------	----------------	---------

Previous Next Finish Cancel

ภาพ 52 แสดงการตั้งค่าไฟล์ข้อมูลสำหรับข้อมูลนำเข้า

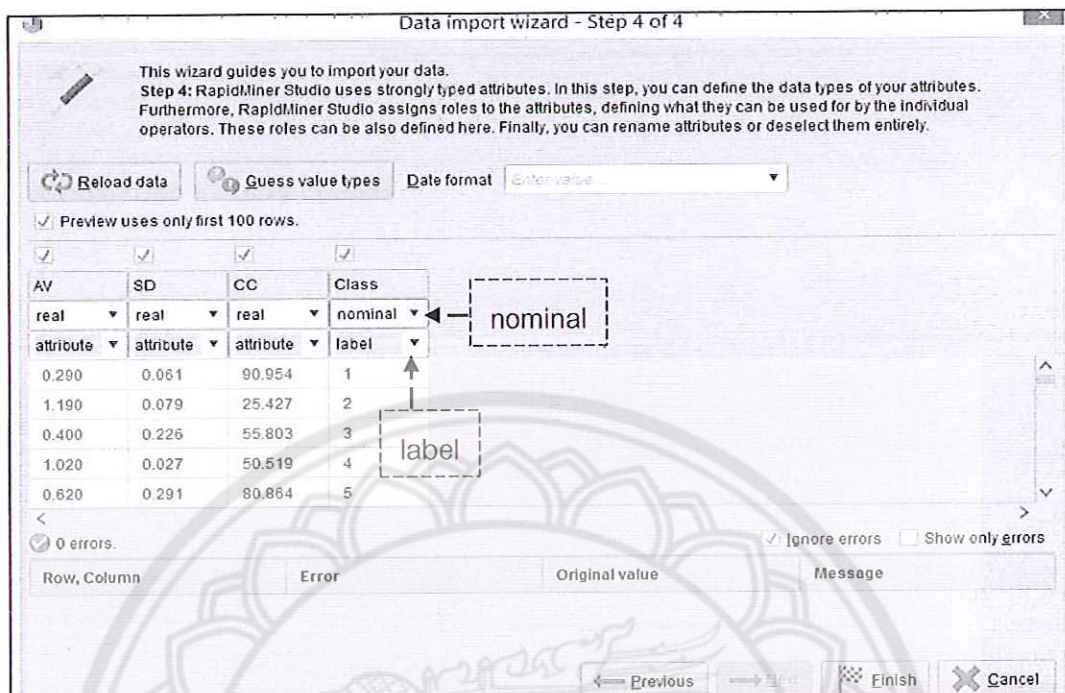
Data import wizard - Step 3 of 4

This wizard guides you to import your data.
Step 3: In RapidMiner Studio, each attribute can be annotated. The most important annotation of an attribute is its name - a row with this annotation defines the names of the attributes. If your data does not contain attribute names, do not set this property. If further annotations are contained in the rows of your data file, you can assign them here.

Annotat...	att1	att2	att3	att4
Name	AV	SD	CC	Class
-	0.29	0.0606714	90.954	1
-	1.19	0.0790071	25.427	2
-	0.4	0.225714	55.803	3
-	1.02	0.0270143	50.519	4
-	0.62	0.291229	80.8635	5
-	0.36	0.135979	75.89	6
-	0.69	0.338107	87.05	7
-	0.37	0.08225	86.107	8
-	0.36	0.0903357	86.5385	8
-	0.35	0.0945214	87.81	8
-	0.34	0.0979357	90.065	8
-	0.3	0.0720857	86.7595	8
-	0.25	0.0439357	85.5135	8
-	0.26	0.0448714	84.541	8

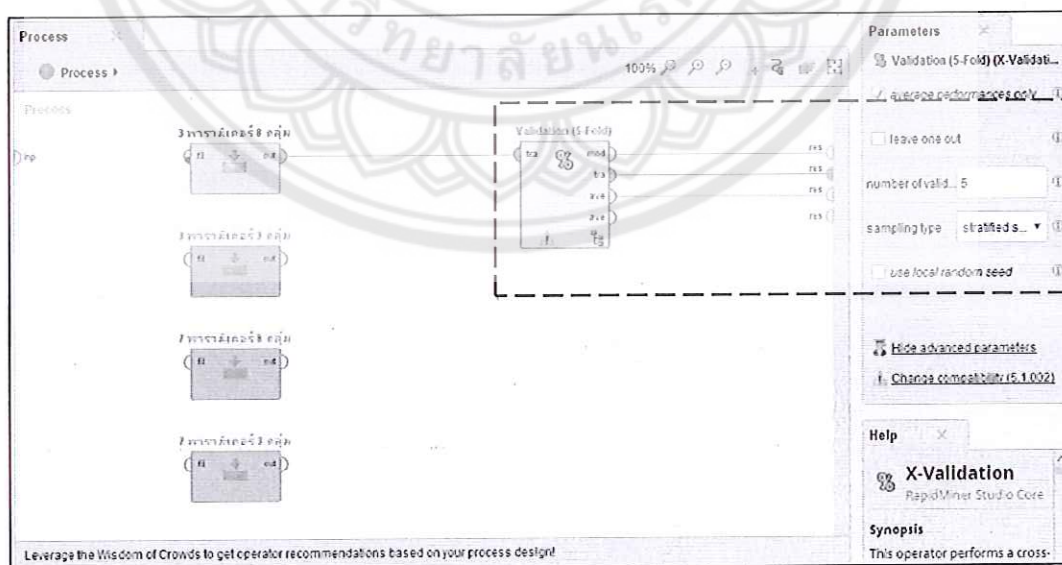
Previous Next Finish Cancel

ภาพ 53 แสดงตัวอย่างข้อมูลนำเข้า



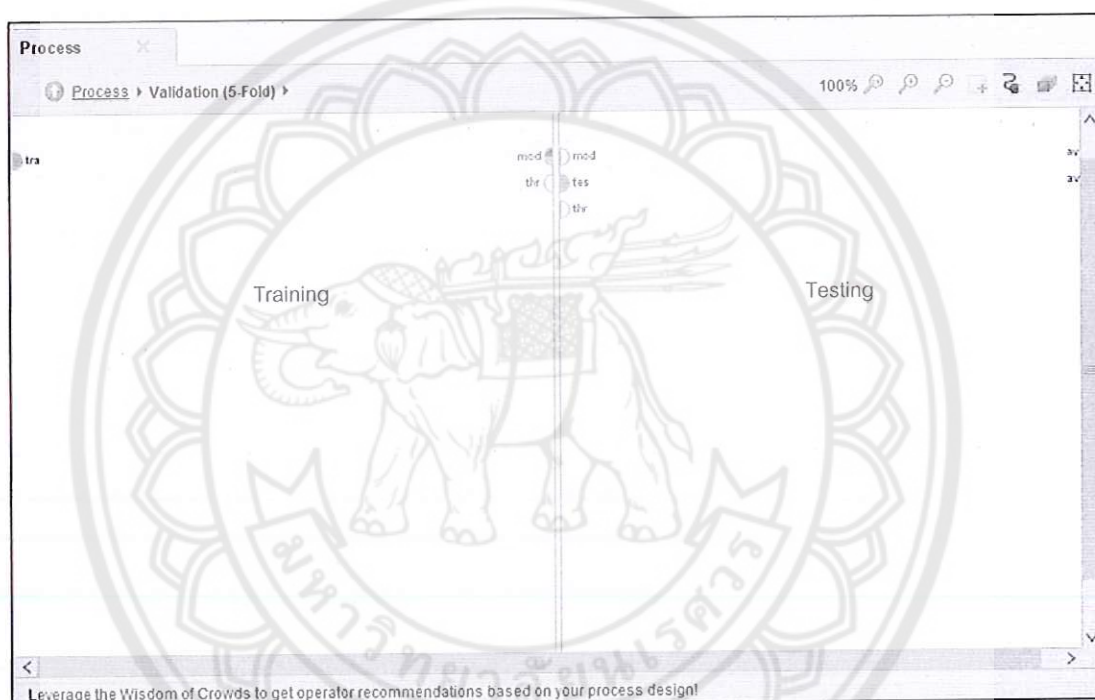
ภาพ 54 แสดงการกำหนดชนิดของข้อมูลตัวอย่างสำหรับข้อมูลนำเข้า

จากภาพ 54 เป็นขั้นตอนสุดท้ายของการเตรียมข้อมูลนำเข้าเพื่อใช้สำหรับการจำแนกประเภทชนิดของเมฆ โดยในขั้นตอนนี้จะต้องทำการกำหนดชนิดของข้อมูลที่ Attribute Class ดังภาพตัวอย่าง



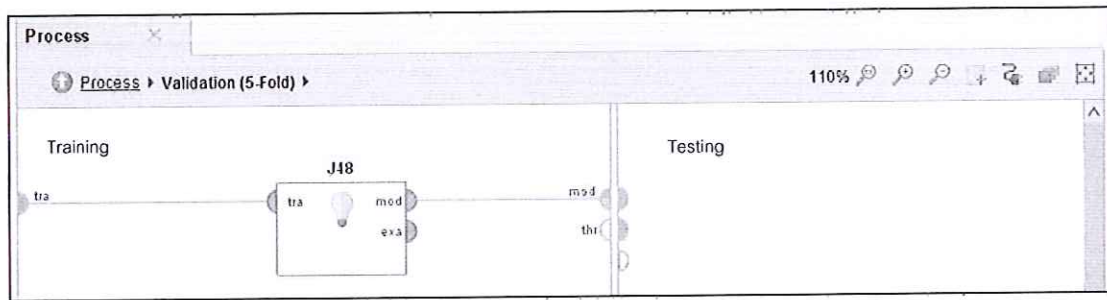
ภาพ 55 แสดงการใช้งานตัวดำเนินการ X-Validation

จากภาพ 55 แสดงการใช้งานตัวดำเนินการสำหรับการวิเคราะห์ข้อมูล โดยเริ่มจากนำตัวดำเนินการ X- Validation และปรับค่าตัวแปร number of validation เท่ากับ 5 หมายถึง ทำการแบ่งข้อมูลออกเป็นส่วน ๆ เท่ากัน 5 ส่วน ข้อมูลทุก ๆ ส่วนจะถูกเรียนรู้และทดสอบ จากนั้นทำการเชื่อมต่อการทำงานของตัวดำเนินการดังกล่าวตัวอย่าง ตัวดำเนินการ X- Validation จะมีลักษณะเป็นแบบตัวดำเนินการย่อย (Sub Process) โดยเมื่อทำการดับเบิลคลิกเข้าไปที่ตัวดำเนินการนี้ จะแสดงหน้าจอของโปรเซสดังภาพ 56

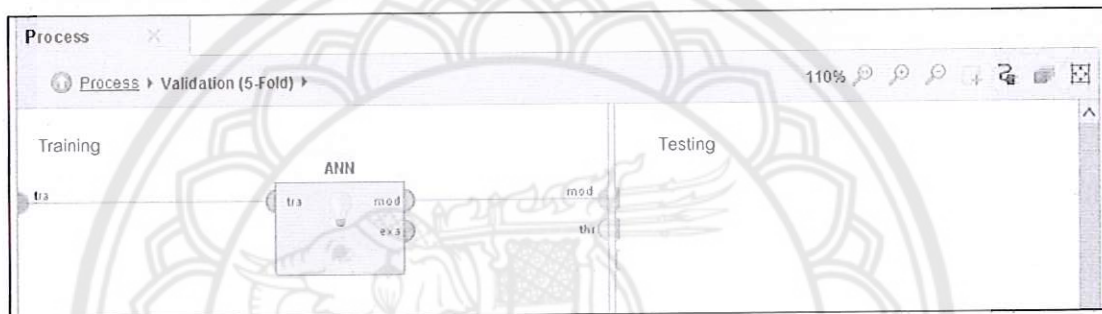


ภาพ 56 แสดงหน้าจอโปรเซสย่อยของตัวดำเนินการ X- Validation

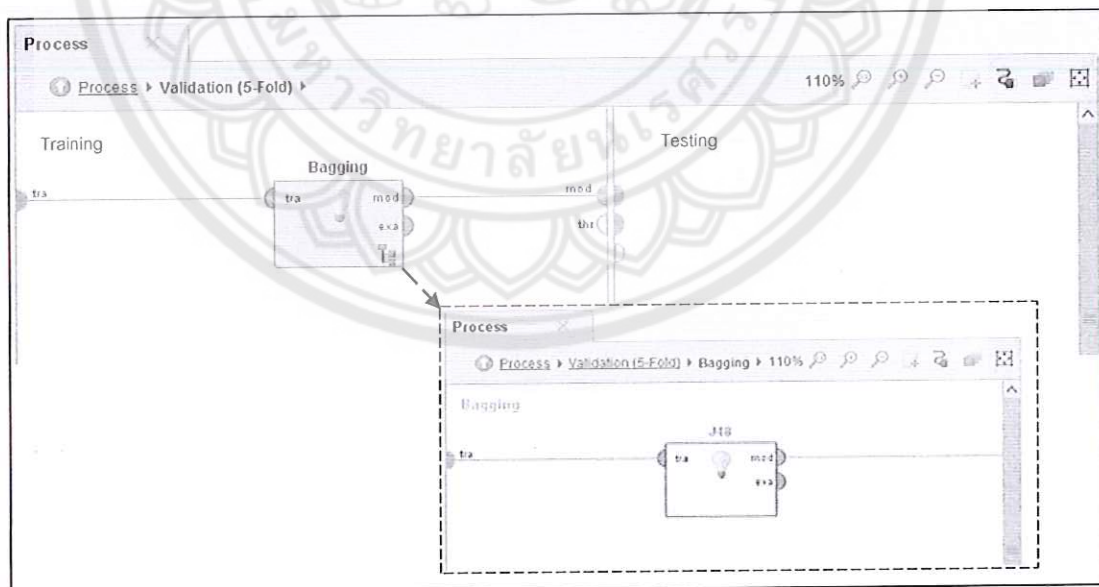
จากภาพ 56 จะเป็นหน้าจอโปรเซสย่อยของตัวดำเนินการ X- Validation โดยหน้าต่างการออกแบบในโปรเซสย่อยนี้จะแบ่งออกเป็น 2 ส่วน คือ ส่วนที่ 1 การ Training ใช้สำหรับออกแบบการเรียนรู้ข้อมูลเพื่อใช้สำหรับสร้างแบบจำลอง และส่วนที่ 2 การ Testing ใช้สำหรับการทดสอบแบบจำลองที่สร้างขึ้นจากส่วนที่ 1 เพื่อวัดประสิทธิภาพของแบบจำลอง ซึ่งในงานวิจัยนี้ได้ออกแบบการทดลองในส่วนที่ 1 โดยใช้ตัวดำเนินการในการเรียนรู้เพื่อสร้างแบบจำลอง 5 ตัวดำเนินการ คือ ANN, J48, Bagging, Boosting และ Random Forest แสดงดังภาพ 57 และ 58 (สำหรับการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยว) และภาพ 59-63 (สำหรับการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม)



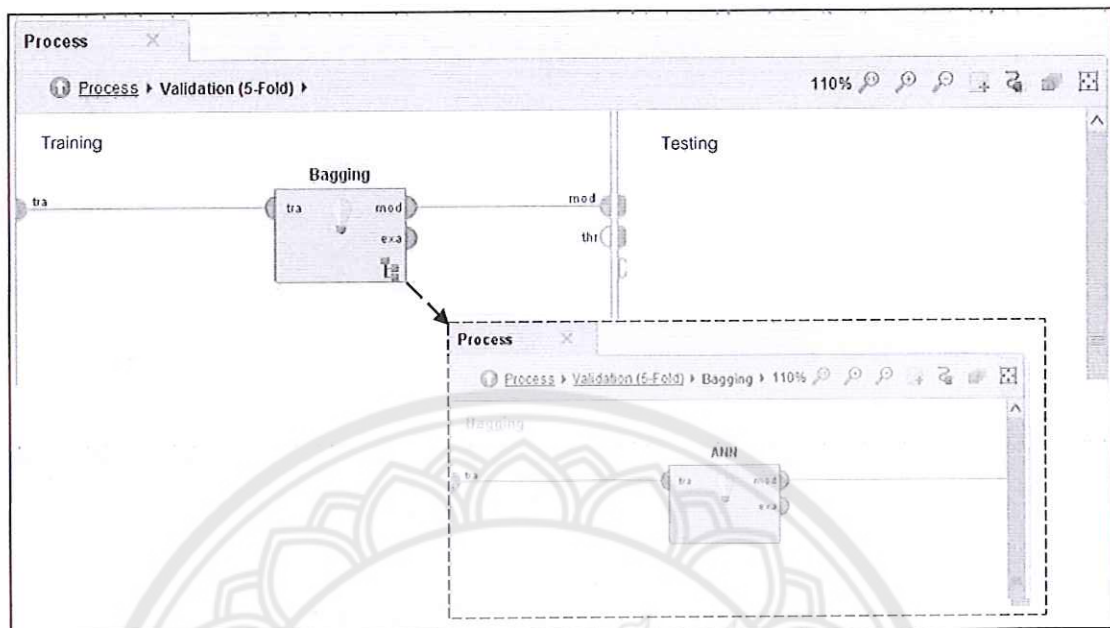
ภาพ 57 แสดงการเรียนรู้ข้อมูลด้วยตัวดำเนินการ J48



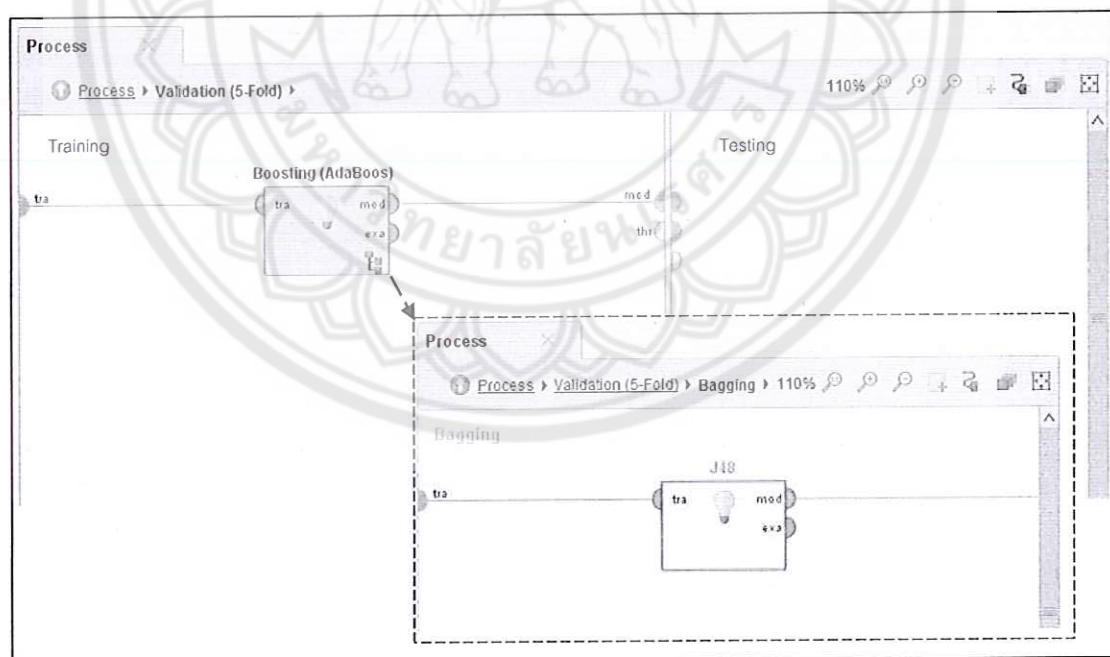
ภาพ 58 แสดงการเรียนรู้ข้อมูลด้วยตัวดำเนินการ ANN



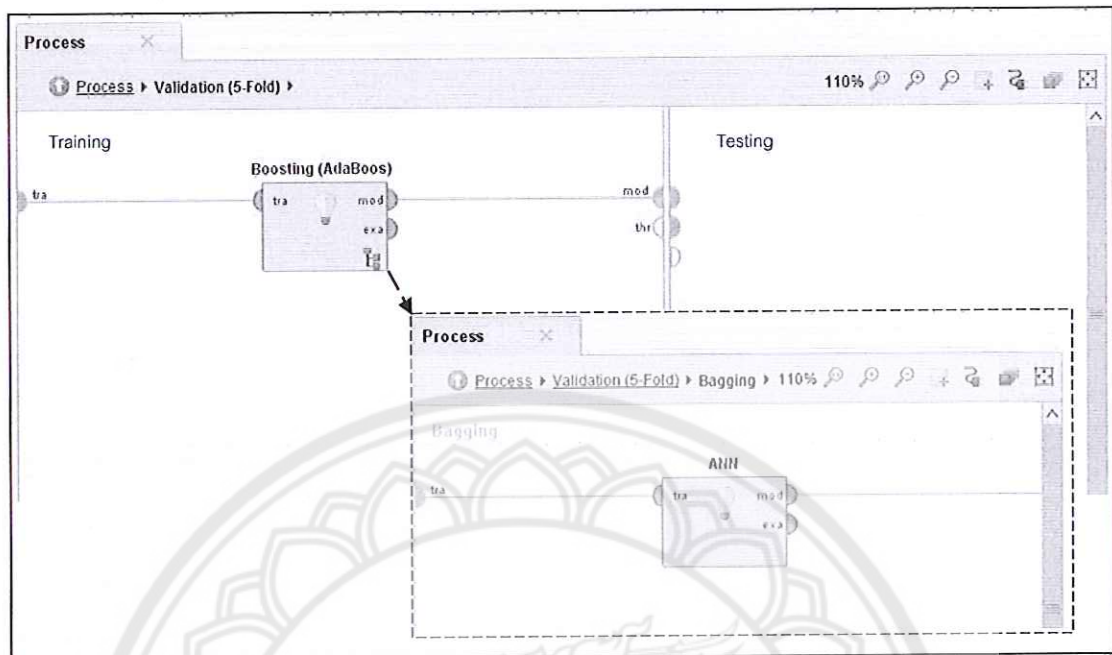
ภาพ 59 แสดงโปรเซสย่อยของตัวดำเนินการ Bagging โดยใช้ J48 เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้



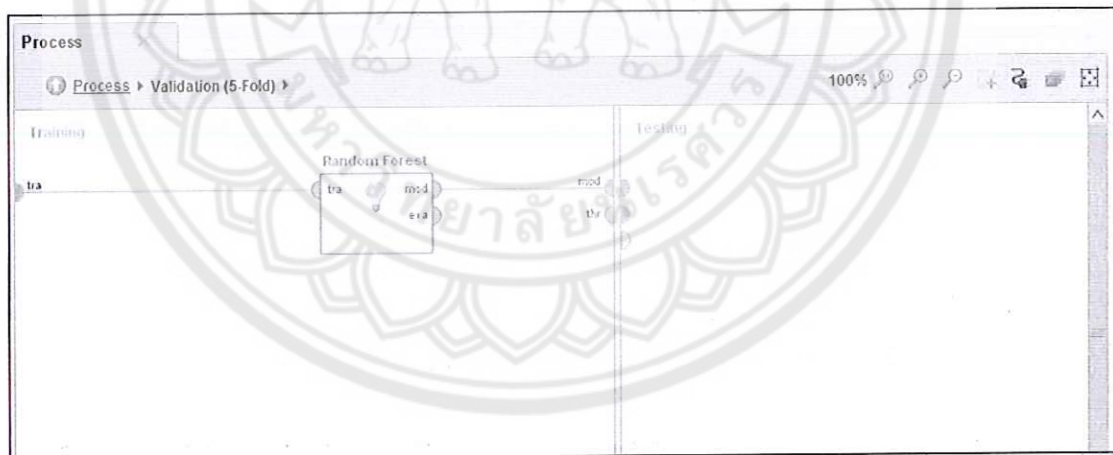
ภาพ 60 แสดงโปรเซสย่อยของตัวดำเนินการ Bagging โดยใช้ ANN เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้



ภาพ 61 แสดงโปรเซสย่อยของตัวดำเนินการ Boosting โดยใช้ J48 เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้

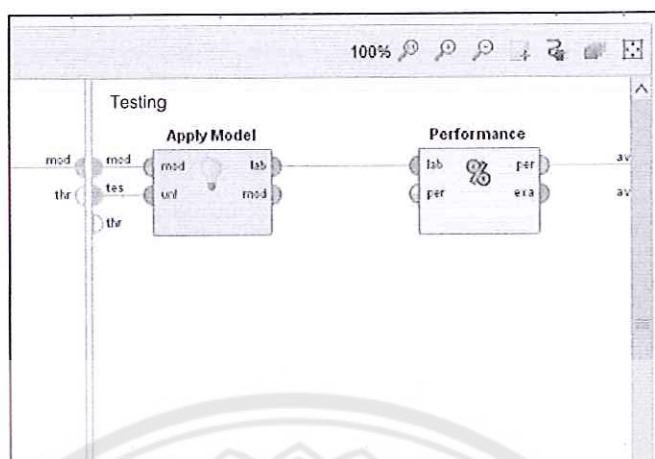


ภาพ 62 แสดงโปรเซสย่อยของตัวดำเนินการ Boosting โดยใช้ ANN เป็นอัลกอริทึมพื้นฐานสำหรับการเรียนรู้

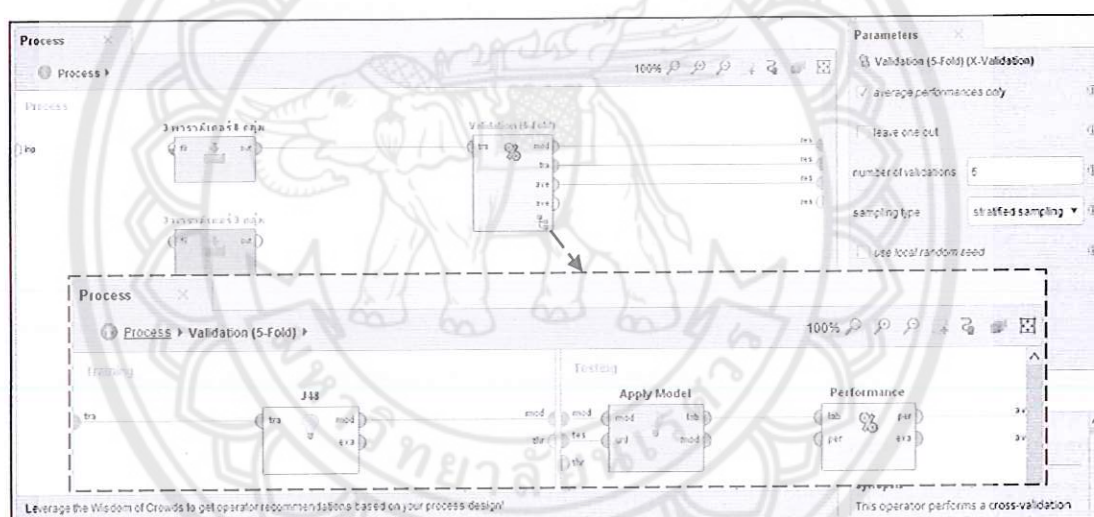


ภาพ 63 แสดงการเรียนรู้ข้อมูลด้วยตัวดำเนินการ Random Forest

สำหรับในส่วนของการ Testing ของทุก ๆ การทดลองในงานวิจัยจะประกอบไปด้วย 2 ตัวดำเนินการ ได้แก่ Apply Model และ Performance แสดงดังภาพ 64

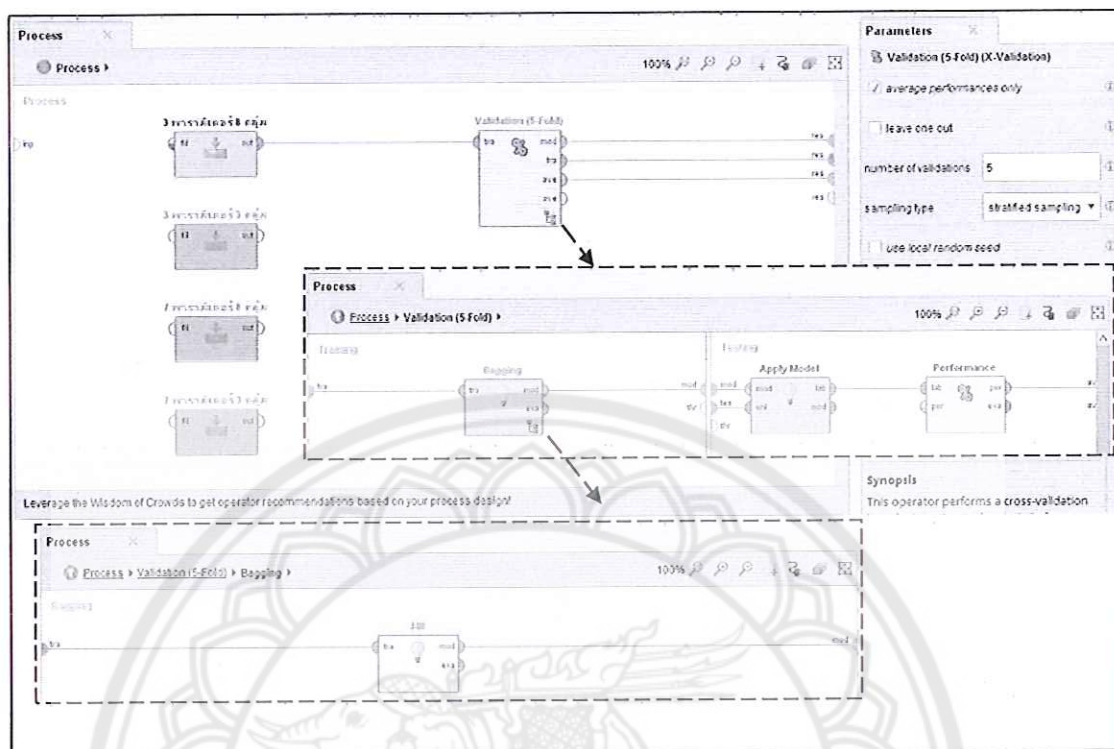


ภาพ 64 แสดงตัวดำเนินการสำหรับทดสอบแบบจำลองในส่วนของการ Testing



ภาพ 65 แสดงตัวอย่างโปรเซสที่สมบูรณ์สำหรับการจำแนกประเภทแบบเดี่ยว

จากภาพ 65 เป็นโปรเซสตัวอย่างที่ได้ทำการออกแบบการทดลองที่เสร็จสมบูรณ์แล้วสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆในการสร้างแบบจำลองการจำแนกประเภทแบบเดี่ยว และภาพ 66 เป็นโปรเซสตัวอย่างที่ได้ทำการออกแบบการทดลองที่เสร็จสมบูรณ์แล้วสำหรับการจำแนกประเภทชนิดของกลุ่มเมฆในการสร้างแบบจำลองการจำแนกประเภทแบบรวมกลุ่ม



ภาพ 66 แสดงตัวอย่างโปรเซสที่สมบูรณ์สำหรับการจำแนกประเภทแบบรวมกลุ่ม

ภาคผนวก ค ผลลัพธ์การทดสอบค่า T-test เพื่อหาความสัมพันธ์ของประสิทธิภาพการจำแนกประเภทชนิดของกลุ่มเมฆ

Table: Paired Sample Statistics

	Mean	N	Std. Deviation	S.E. Mean
Pair 1				
J48	73.51	20	12.52	2.8
Bagging_J48	75.38	20	11.83	2.65

Table: Paired Samples Correlations

	N	Correlation	Sig.
Pair 1			
J48 & Bagging_J48	20	0.99	0

Table: Paired Samples Test

Pair 1	J48 - Bagging_J48	Paired Differences					
		95% Confidence Interval of the Difference					
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t
		-1.87	1.77	0.4	-2.7	-1.05	-4.74
							19
							0

ภาพ 67 แสดงผลลัพธ์การทดสอบค่า T-test โดยเปรียบเทียบระหว่าง J48 กับ Bagging (J48)

Table: Paired Sample Statistics

	Mean	N	Std. Deviation	S.E. Mean
Pair 2				
J48	73.51	20	12.52	2.8
Bagging_ANN	72.84	20	12.97	2.9

Table: Paired Samples Correlations

	N	Correlation	Sig.
Pair 2			
J48 & Bagging_ANN	20	0.99	0

Table: Paired Samples Test

Pair 2	J48 - Bagging_ANN	Paired Differences					
		95% Confidence Interval of the Difference					
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t
		0.66	2.14	0.48	-0.33	1.66	1.39
							df
							19
							Sig. (2-tailed)
							0.18

ภาพ 68 แสดงผลลัพธ์การทดสอบค่า T-test โดยเปรียบเทียบระหว่าง J48 กับ Bagging (ANN)

Table: Paired Sample Statistics

	Mean	N	Std. Deviation	S.E. Mean
Pair 3				
J48	73.51	20	12.52	2.8
Boosting_J48	74.6	20	12.51	2.8

Table: Paired Samples Correlations

	N	Correlation	Sig.
Pair 3			
J48 & Boosting_J48	20	0.99	0

Table: Paired Samples Test

Pair 3	J48 - Boosting_J48	Paired Differences					
		95% Confidence Interval of the Difference					
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t
		-1.09	1.89	0.42	-1.98	-0.21	-2.59
							df
							19
							Sig. (2-tailed)
							0.018

ภาพ 69 แสดงผลลัพธ์การทดสอบค่า T-test โดยเปรียบเทียบระหว่าง J48 กับ Boosting (J48)

Table: Paired Sample Statistics

	Mean	N	Std. Deviation	S.E. Mean
Pair 4				
J48	73.51	20	12.52	2.8
Boosting_ANN	72.26	20	13.07	2.92

Table: Paired Samples Correlations

	N	Correlation	Sig.
Pair 4			
J48 & Boosting_ANN	20	0.98	0

Table: Paired Samples Test

Pair 4	J48 - Boosting_ANN	Paired Differences					
		95% Confidence Interval of the Difference					
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t
		1.25	2.59	0.58	0.04	2.46	2.16
							19
							0.044

ภาพ 70 แสดงผลลัพธ์การทดสอบค่า T-test โดยเปรียบเทียบระหว่าง J48 กับ Boosting (ANN)

Table: Paired Sample Statistics

	Mean	N	Std. Deviation	S.E. Mean
Pair 1				
ANN	72.35	20	12.85	2.87
Bagging_J48	75.38	20	11.83	2.65

Table: Paired Samples Correlations

	N	Correlation	Sig.
Pair 1			
ANN & Bagging_J48	20	0.95	0

Table: Paired Samples Test

Pair 1	ANN - Bagging_J48	Paired Differences					
		95% Confidence Interval of the Difference					
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t
		-3.03	4.07	0.91	-4.93	-1.12	-3.33
							df
							19
							Sig. (2-tailed)
							0.004

ภาพ 72 แสดงผลลัพธ์การทดสอบค่า T-test โดยเปรียบเทียบระหว่าง ANN กับ Bagging (J48)

Table: Paired Sample Statistics

	Mean	N	Std. Deviation	S.E. Mean
Pair 3				
ANN	72.35	20	12.85	2.87
Boosting_J48	74.6	20	12.51	2.8

Table: Paired Samples Correlations

	N	Correlation	Sig.
Pair 3			
ANN & Boosting_J48	20	0.94	0

Table: Paired Samples Test

Pair 3	ANN - Boosting_J48	Paired Differences					
		95% Confidence Interval of the Difference					
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t
		-2.25	4.28	0.96	-4.25	-0.24	-2.35
							df
							19
							Sig. (2-tailed)
							0.03

ภาพ 74 แสดงผลลัพธ์การทดสอบค่า T-test โดยเปรียบเทียบระหว่าง ANN กับ Boosting (J48)

Table: Paired Sample Statistics

	Mean	N	Std. Deviation	S.E. Mean
Pair 4				
ANN	72.35	20	12.85	2.87
Boosting_ANN	72.26	20	13.07	2.92

Table: Paired Samples Correlations

	N	Correlation	Sig.
Pair 4			
ANN & Boosting_ANN	20	1	0

Table: Paired Samples Test

Pair 4	ANN - Boosting_ANN	Paired Differences					
		95% Confidence Interval of the Difference					
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t
		0.09	0.28	0.06	-0.04	0.23	1.49
							df
							19
							Sig. (2-tailed)
							0.151

ภาพ 75 แสดงผลลัพธ์การทดสอบค่า T-test โดยเปรียบเทียบระหว่าง ANN กับ Boosting (ANN)

Table: Paired Sample Statistics

	Mean	N	Std. Deviation	S.E. Mean
Pair 5				
ANN	72.35	20	12.85	2.87
Random_Forest	77.33	20	12.31	2.75

Table: Paired Samples Correlations

	N	Correlation	Sig.
Pair 5			
ANN & Random_Forest	20	0.89	0

Table: Paired Samples Test

Pair 5	ANN - Random_Forest	Paired Differences					
		95% Confidence Interval of the Difference					
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t
		-4.98	5.95	1.33	-7.77	-2.2	-3.74
							19
							0.001

ภาพ 76 แสดงผลลัพธ์การทดสอบค่า T-test โดยเปรียบเทียบระหว่าง ANN กับ Random Forest