

การกำหนดทางวิ่งของเครื่องบินขาเข้าจากข้อมูลการจราจรทางอากาศ  
โดยใช้แมชชีนเลิร์นนิง

LANDING RUNWAY ASSIGNMENT BY AIRPORT TRAFFIC  
USING MACHINE LEARNING



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต  
สาขาวิชาวิศวกรรมไฟฟ้าและคอมพิวเตอร์  
คณะวิศวกรรมศาสตร์  
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง  
พ.ศ. 2566

KMITL-2023-EN-M-027-055

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

LANDING RUNWAY ASSIGNMENT BY AIRPORT TRAFFIC  
USING MACHINE LEARNING



A THESIS SUBMITTED IN PARTIAL FULFILLMENT  
OF THE REQUIREMENT FOR THE DEGREE OF  
MASTER OF ENGINEERING IN ELECTRICAL AND COMPUTER ENGINEERING  
SCHOOL OF ENGINEERING  
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG  
2023

KMITL-2023-EN-M-027-055

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



COPYRIGHT 2023

SCHOOL OF ENGINEERING

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารลิขสิทธิ์สงวนไว้สําหรับการเชิงในเพื่อการศึกษาเท่านั้น เมื่อผู้ยู่ได้เห็นไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อวิทยานิพนธ์	การกำหนดทางวิ่งของเครื่องบินขาเข้าจากข้อมูลการจราจรทางอากาศโดยใช้แมชชีนเลิร์นนิง
นักศึกษา	นายวศรุต จิงสุวดี
รหัสประจำตัว	64601125
ปริญญา	วิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชา	วิศวกรรมไฟฟ้าและคอมพิวเตอร์
พ.ศ.	2566
อาจารย์ที่ปรึกษาวิทยานิพนธ์	รศ.ดร.ชวลิต เบญจางคประเสริฐ

## บทคัดย่อ

วิทยานิพนธ์ฉบับนี้นำเสนอแนวทางแก้ไขปัญหาระนาบที่ของผู้ควบคุมการจราจรทางอากาศของสนามบินสุวรรณภูมิ โดยลดปริมาณงาน และเพิ่มขีดความสามารถทางวิ่งด้วยการใช้แมชชีนเลิร์นนิงเป็นเครื่องมือในการช่วยตัดสินใจกำหนดทางวิ่งสำหรับเครื่องบินขาเข้าตามข้อมูลที่สำคัญ เช่น ข้อมูลการจราจรในสนามบินของเครื่องบินที่ขึ้น และลงบนทางวิ่งที่สนามบินสุวรรณภูมิ เนื่องจากปริมาณการจราจรทางอากาศที่เพิ่มขึ้น ส่งผลต่อความเครียดของผู้ควบคุมการจราจรทางอากาศ อาจส่งผลให้ประสิทธิภาพการจัดการจราจร และการตัดสินใจลดน้อยลง โดยงานวิจัยนี้ต้องการศึกษา และวิเคราะห์หาปัจจัยที่ส่งผลกับการกำหนดทางวิ่งให้กับเครื่องบินขาเข้า รวมทั้งการจัดการกับชุดข้อมูล เพื่อคัดเลือกชุดข้อมูลที่เหมาะสมกับการทำนายผลลัพธ์ของแบบจำลอง และนำไปทดสอบกับแบบจำลองประเภทจำแนกข้อมูล (Classification) ที่มีอัลกอริทึมแตกต่างกัน ในการศึกษาครั้งนี้ใช้แบบจำลองทั้งหมด 5 แบบ คือ Logistic Regression, Random Forest, Support Vector Machine, K-Nearest Neighbor และ Artificial Neural Networks เพื่อเปรียบเทียบหาแบบจำลองที่เหมาะสมกับการทำนาย โดยผลจากการเปรียบเทียบประสิทธิภาพการทำนายของแบบจำลองพบว่า แบบจำลอง Logistic Regression มีความแม่นยำ และถูกต้องมากที่สุด โดยมีค่าความถูกต้องรวมของการทำนายทางวิ่งร้อยละ 80.18 และได้ค่าประสิทธิภาพที่วัดได้จากพื้นที่ใต้เส้นโค้ง ROC (Area under curve, AUC) ร้อยละ 87.05 นอกจากนี้แบบจำลองใช้เวลาประมวลผลข้อมูล 0.1 วินาที

Thesis	Landing Runway Assignment by Airport Traffic using Machine Learning
Student	Mr.Wasarut Jungsuwadee
Student ID.	64601125
Degree	Master of Engineering
Program	Electrical and Computer Engineering
Year	2023
Thesis Advisor	Assoc.Prof.Dr.Chawalit Benjangkprasert

## ABSTRACT

This thesis proposes the solutions to the overwhelming burden of air traffic controllers by reducing workload and optimizing runway capacity using machine learning tools. Machine learning tools are applied to assign runways for incoming aircraft. The selected data to predict assigned landing runway based on critical information such as aerodrome traffic information of aircraft taking off and landing on the runway at Suvarnabhumi Airport, THAILAND. Due to the increasing air traffic that affects the ability to air traffic controller then the management of runway capacity cannot be managed efficiently. This thesis intends to study and analyze the radar data and air traffic data to select the influence feature that is proper for the classifier models. To get the appropriate model in terms of accuracy and the time to process of the model. Classifier models in this thesis are Logistic Regression, Random Forest, Support Vector Machine, K-Nearest Neighbor and Artificial Neural Networks, each model was compared with the same dataset and evaluated the model by using the performance measurement method. The result of model comparison shows that the Logistic Regression model has the most overall accuracy at 80.18 percent and 87.05 percent of Area under Receiver Operating Characteristic Curve (AUC). Furthermore, the Logistic Regression model took time to process the dataset at 0.1 seconds.

## กิตติกรรมประกาศ

วิทยานิพนธ์เล่มนี้สำเร็จได้ด้วยความกรุณาจากอาจารย์ที่ปรึกษา รศ.ดร.ชวลิต เบญจางคประเสริฐ ที่ให้ความช่วยเหลือ และชี้แนะแนวทางแก้ปัญหาตลอดจนให้ความรู้และประสบการณ์แก่ข้าพเจ้า

ขอขอบคุณ รศ.บุญยชนะ ภูระหงษ์ และ ดร.ไอศูรย์ กาญจนสุรัตน์ ที่ให้คำปรึกษา ให้คำแนะนำตลอดจนชี้แนะแนวทางการศึกษาในระดับชั้นของวิศวกรรมศาสตรมหาบัณฑิต จนนำมาสู่เป้าหมายการเรียนรู้ของตัวข้าพเจ้า และสำเร็จเป็นวิทยานิพนธ์ฉบับนี้ได้

ขอขอบคุณ บริษัท วิทย์การบินแห่งประเทศไทย จำกัด ที่ให้การสนับสนุนข้อมูลของสนามบินสุวรรณภูมิสำหรับการวิจัยนี้

สำหรับคุณงามความดีอันใดที่เกิดจากวิทยานิพนธ์ฉบับนี้ ข้าพเจ้าขอมอบให้กับบิดามารดา ซึ่งเป็นที่รักและเคารพยิ่ง ตลอดจนครูอาจารย์ที่เคารพทุกท่านที่ได้ประสิทธิ์ประสาทวิชาความรู้และถ่ายทอดประสบการณ์ที่ดีให้แก่ข้าพเจ้า

วศรุต จິงสูวดี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และห้ามอ้างถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ.....	II
กิตติกรรมประกาศ.....	III
สารบัญ.....	IV
สารบัญตาราง.....	VI
สารบัญรูป.....	VII
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา.....	2
1.3 สมมุติฐานของการศึกษา.....	2
1.4 ขอบเขตการวิจัย.....	2
1.5 ขั้นตอนของการศึกษา.....	2
บทที่ 2 หลักการและทฤษฎีที่เกี่ยวข้อง.....	3
2.1 การจัดการจราจรทางอากาศ.....	3
2.1.1 สนามบินนานาชาติสุวรรณภูมิ (VTBS).....	4
2.2 ปัญญาประดิษฐ์ (AI : Artificial Intelligence).....	6
2.3 การเรียนรู้ของเครื่องจักร (ML : Machine learning).....	7
2.3.1 Logistic Regression Analysis.....	9
2.3.2 Random Forest.....	11
2.3.3 K-Nearest Neighbors (KNN).....	12
2.3.4 Support Vector Machine (SVM).....	12
2.3.5 Artificial Neuron Network (ANN).....	16
2.4 การวัดและประเมินผลประสิทธิภาพของแบบจำลอง Classification.....	18
2.4.1 Confusion Matrix.....	18
2.4.2 Receiver Operating Characteristic (ROC) Curve.....	19
2.5 งานวิจัยที่เกี่ยวข้อง.....	21

## สารบัญ (ต่อ)

	หน้า
บทที่ 3 วิธีการดำเนินการวิจัย.....	29
3.1 การออกแบบการทดลอง.....	29
3.2 ข้อมูล.....	30
3.2.1 ข้อมูลเรดาร์ (Radar Data).....	31
3.2.2 ข้อมูลแผนการบิน (Flight Plan).....	32
3.2.3 การคัดเลือกข้อมูล Dataset ของแบบจำลอง.....	32
3.2.4 การจัดรูปแบบข้อมูล Dataset ของแบบจำลอง.....	34
3.3 การนำเข้าชุดข้อมูลเพื่อทดสอบในแบบจำลอง.....	37
3.4 การเปรียบเทียบประสิทธิภาพของแบบจำลอง.....	38
บทที่ 4 ผลการดำเนินการ.....	40
4.1 ประสิทธิภาพของแบบจำลอง.....	40
4.1.1 Confusion Matrix ของแบบจำลอง Logistic Regression.....	41
4.1.2 Confusion Matrix ของแบบจำลอง Artificial Neural Networks.....	41
4.1.3 Confusion Matrix ของแบบจำลอง Support Vector Machine - Radial Basis Function.....	42
4.1.4 Confusion Matrix ของแบบจำลอง Random Forest.....	43
4.1.5 Confusion Matrix ของแบบจำลอง K-Nearest Neighbor.....	43
4.1.6 ค่าความถูกต้อง ค่าความระลึก ค่าเฉลี่ยความถูกต้อง ค่าความเที่ยงตรง.....	44
4.1.7 พื้นที่ใต้เส้นโค้ง ROC (Area under curve, AUC).....	47
4.1.8 เวลาที่ใช้ประมวลผล.....	51
4.2 เปรียบเทียบประสิทธิภาพแบบจำลองกับงานวิจัยที่เกี่ยวข้อง.....	51
บทที่ 5 ปัญหาและแนวทางการแก้ไข.....	53
5.1 สรุปและวิจารณ์ผลการทดลอง.....	53
5.2 แนวทางการแก้ไข.....	54
บรรณานุกรม.....	55
ภาคผนวก ผลงานวิจัยที่ได้รับการตีพิมพ์.....	58
ประวัติผู้เขียน.....	60

# สารบัญตาราง

ตารางที่.....	หน้า
2.1 ความแม่นยำเปรียบเทียบระหว่างสองแบบจำลอง.....	21
2.2 ประสิทธิภาพการคาดการณ์และเวลาในการคำนวณสำหรับเทคนิค GB.....	24
2.3 ผลลัพธ์การเปรียบเทียบความแม่นยำระหว่าง 2 แบบจำลอง.....	26
2.4 ผลลัพธ์การเปรียบเทียบความแม่นยำระหว่าง Random Forest, XGBoost และ SVM.....	27
2.5 ค่าความแม่นยำของ Novel Linear SVM กับ Convolutional SVM.....	27
2.6 ผลลัพธ์เชิงสถิติของ KNN เทียบกับ ANN.....	28
2.7 Confusion matrices for the predicting runway conditions using XGBoost.....	28
3.1 ตัวอย่างข้อมูลเรดาร์.....	31
3.2 ตัวอย่างข้อมูลแผนการบิน.....	32
3.3 หัวข้อชุดข้อมูลอินพุตของแบบจำลอง.....	34
3.4 ตัวอย่างชุดข้อมูลการจราจรทางอากาศที่มี 20 NM รอบสนามบิน.....	36
3.5 ตัวอย่างชุดข้อมูล Dataset หลังจากการทำ Data Preprocessing.....	36
3.6 จำนวนชุดข้อมูลการฝึกฝนและการทดสอบ.....	38
4.1 สรุปพารามิเตอร์ที่ใช้สำหรับแบบจำลอง Logistic Regression.....	41
4.2 สรุปพารามิเตอร์ที่ใช้สำหรับแบบจำลอง Artificial Neural Networks.....	41
4.3 สรุปพารามิเตอร์ที่ใช้สำหรับแบบจำลอง SVM – RBF Kernel.....	42
4.4 สรุปพารามิเตอร์ที่ใช้สำหรับแบบจำลอง Random Forest.....	43
4.5 สรุปพารามิเตอร์ที่ใช้สำหรับแบบจำลอง K-Nearest Neighbor.....	43
4.6 ค่าความถูกต้อง ค่าความระลอก ค่าเฉลี่ยความถูกต้อง ค่าความเที่ยงตรง ของแบบจำลอง.....	44
4.7 ผลลัพธ์ของแบบจำลองในงานวิจัยอื่นที่เกี่ยวข้องกับการจัดการจราจรทางอากาศ.....	52

# สารบัญรูปภาพ

รูปที่.....	หน้า
1.1 ตัวอย่างหน้าจอแสดงผลข้อมูลสำหรับเจ้าหน้าที่ควบคุมการจราจรทางอากาศ.....	1
2.1 พื้นที่การควบคุมจราจรทางอากาศบริเวณสนามบิน.....	3
2.2 แผนผังทางวิ่งสนามบินนานาชาติสุวรรณภูมิ.....	4
2.3 ทางวิ่งของสนามบินนานาชาติสุวรรณภูมิ ที่ให้บริการในเงื่อนไขของลมทางทิศเหนือ.....	5
2.4 การจราจรทางอากาศที่ 25 NM รอบสนามบินนานาชาติสุวรรณภูมิ.....	6
2.5 แผนภาพการจำแนกรูปแบบของ Machine Learning.....	8
2.6 ฟังก์ชันโลจิสติก (Logistic Function).....	9
2.7 การทำงานของ Random Forest Classifier.....	12
2.8 ตัวอย่าง SVM ใน 2 มิติ.....	13
2.9 ตัวอย่างไฮเปอร์เพลนที่ใช้ในปัญหาที่ข้อมูลไม่เป็นเชิงเส้น.....	14
2.10 ส่วนประกอบของโครงข่ายประสาทเทียม.....	16
2.11 แผนผังโครงข่ายประสาทเทียม.....	17
2.12 รูปแบบของ Confusion Matrix.....	19
2.13 ROC Curve.....	20
2.14 ความแม่นยำของการจราจรทั้ง 4 รูปแบบ.....	22
2.15 ความแม่นยำของ KNN ในการทำนายหาความล่าช้าของเที่ยวบิน.....	23
2.16 Relationship between regression coefficient and iteration.....	24
2.17 Binary: Scatter plot. a) Decision Tree b) Random Forest.....	25
2.18 Multiclass: Scatter plot. a) Decision Tree b) Random Forest.....	26
3.1 ภาพรวมขั้นตอนการทดลอง.....	30
3.2 การจราจรของเครื่องบินบริเวณสนามบินบนหน้าจอ Air Situation Display (ASD).....	33
3.3 แผนภาพการคัดแยกข้อมูลเครื่องบินขาเข้าและขาออกสนามบินสุวรรณภูมิ.....	34
3.4 แผนภาพการควมรวมข้อมูลระหว่างแผนการบินและข้อมูลเรดาร์.....	35
3.5 ความสัมพันธ์ระหว่างชุดข้อมูลที่ใช้ในการจำแนกกับการกำหนดทางวิ่งของเครื่องบินขาเข้า.....	37
3.6 แผนภาพการป้อนข้อมูลเพื่อเรียนรู้และทดสอบของแบบจำลอง.....	38
3.7 แผนภาพการเปรียบเทียบประสิทธิภาพของแบบจำลอง.....	39

## สารบัญรูปภาพ (ต่อ)

รูปที่.....	หน้า
4.1 แผนภาพการนำเข้าข้อมูลและวัดประสิทธิภาพของแบบจำลอง.....	40
4.2 Confusion Matrix – Logistic Regression.....	41
4.3 Confusion Matrix – ANN.....	42
4.4 Confusion Matrix – SVM (RBF).....	42
4.5 Confusion Matrix – Random Forest.....	43
4.6 Confusion Matrix – KNN.....	44
4.7 ประสิทธิภาพของแบบจำลอง Logistic Regression.....	45
4.8 ประสิทธิภาพของแบบจำลอง Random Forest.....	45
4.9 ประสิทธิภาพของแบบจำลอง SVM with RBF.....	46
4.10 ประสิทธิภาพของแบบจำลอง KNN.....	46
4.11 ประสิทธิภาพของแบบจำลอง ANN.....	47
4.12 เปรียบเทียบประสิทธิภาพการทำนายของแบบจำลอง.....	47
4.13 Area under curve – Logistic Regression.....	48
4.14 Area under curve – Random Forest.....	48
4.15 Area under curve – SVM (RBF).....	49
4.16 Area under curve – KNN.....	49
4.17 Area under curve – ANN.....	50
4.18 Area under curve – All Models comparison.....	50
4.19 เวลาที่ใช้ไปในการทำนายของแต่ละแบบจำลอง.....	51

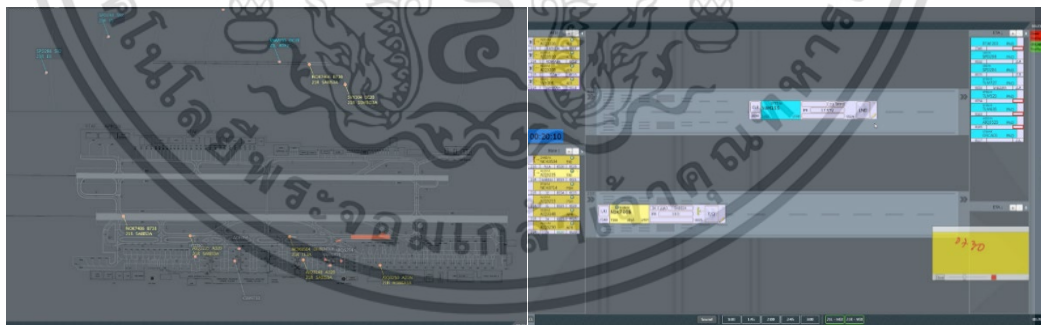
# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันการคมนาคมทางอากาศเป็นที่นิยมอย่างแพร่หลาย ทั้งการขนส่งผู้โดยสารและการขนส่งสินค้า จากปริมาณของเที่ยวบินที่เพิ่มสูงขึ้น ส่งผลให้ความหนาแน่นของอากาศยานที่ทำการขึ้นและลงในสนามบินมีเพิ่มมากขึ้นไปตามกัน ดังนั้นการจัดการจราจรทางอากาศภายในสนามบินจึงเป็นปัจจัยที่สำคัญ ที่จะช่วยให้การคมนาคมทางอากาศเป็นไปอย่างปลอดภัย ต่อเนื่อง และตรงเวลา ภายใต้ภาวะที่ความหนาแน่นของอากาศยานมีเพิ่มจำนวนมากขึ้นอย่างต่อเนื่อง

ด้วยระบบการบริหารการจราจรทางอากาศบริเวณสนามบินสุวรรณภูมิ ณ ปัจจุบัน อาศัยเทคโนโลยีการติดตามอากาศยานด้วยระบบเรดาร์ทั้งภาคอากาศและภาคพื้นดิน รวมทั้งระบบประมวลผลข้อมูลเที่ยวบิน (Flight Plan) เพื่อนำมาแสดงผลให้กับเจ้าหน้าที่ควบคุมการจราจรทางอากาศ สำหรับประกอบการตัดสินใจในการจัดลำดับเครื่องขึ้นและลง รวมถึงการกำหนดเส้นทางวิ่งของสนามบิน (Runway) ให้กับนักบิน ถึงแม้ว่าเจ้าหน้าที่ควบคุมการจราจรทางอากาศจะมีเทคโนโลยีที่ช่วยให้เห็นอากาศยานได้จากระยะไกล รวมทั้งยังมีการแจ้งเตือนต่าง ๆ ภายในสนามบินแล้วนั้น ก็ยังคงต้องอาศัยการตัดสินใจด้วยตนเองจากข้อมูลต่าง ๆ ที่แสดงบนหน้าจอ ดังรูปที่ 1.1 ที่จะต้องมีความถูกต้อง รวดเร็ว และแม่นยำ เพื่อให้ได้ประสิทธิภาพการรองรับเที่ยวบินบนทางวิ่งสูงสุด จากเหตุผลที่กล่าวมาจึงส่งผลต่อความเครียดของเจ้าหน้าที่ควบคุมการจราจรทางอากาศ และอาจเป็นผลให้การตัดสินใจของเจ้าหน้าที่ควบคุมการจราจรทางอากาศมีความล่าช้าหรือผิดพลาดได้



รูปที่ 1.1 ตัวอย่างหน้าจอแสดงผลข้อมูลสำหรับเจ้าหน้าที่ควบคุมการจราจรทางอากาศ

ด้วยความเครียดจากความกดดันของภาระงานที่เพิ่มขึ้น จึงก่อให้เกิดแนวความคิดการนำเทคโนโลยี Machine Learning เข้ามาช่วยแบ่งเบาภาระของผู้ควบคุมการจราจรทางอากาศสำหรับการกำหนดทางวิ่งของเครื่องบินขาเข้าจากข้อมูลการจราจรทางอากาศของสนามบินนานาชาติสุวรรณภูมิ โดยลดความเครียดของภาระงานที่ต้องอาศัยการพิจารณาข้อมูลจากหลากหลายด้านด้วยตนเอง ซึ่งจากการศึกษาเทคโนโลยี Machine Learning พบว่าเป็นเทคโนโลยีที่น่าไปใช้ในการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แก้ปัญหาต่าง ๆ ดังเช่น ในการศึกษาการประมวลผลล่วงหน้าโดยใช้อัลกอริทึมพื้นฐานในการสกัด  
หลอดเลือดในภาวะเบาหวานขึ้นจอตาโดย Kasi และคณะ [1,2] การใช้ Machine Learning เพื่อ  
ปรับปรุงการสกัดหลอดเลือด [3] การนำไปใช้เพื่อตรวจจับผู้บุกรุกด้วยผลการตรวจจับที่ยอดเยี่ยมเมื่อ  
รวมกับ R-CNN ที่เร็วกว่า [4] การปรับปรุงประสิทธิภาพการเรียนรู้ของเครื่องอีกอย่างหนึ่งดังงานวิจัย  
Simulated Annealing [5] การวิจัยการพัฒนาอัลกอริทึมที่มีประสิทธิภาพดีขึ้นและพฤติกรรมคอน  
เวอร์เจนซ์ [6] จะเห็นได้ว่า Machine Learning ใช้เพื่อวัตถุประสงค์ที่หลากหลายนอกเหนือจากการ  
ใช้งานทางการแพทย์และการระบุภัยคุกคามแล้วนั้น ยังสามารถใช้ในการจัดการการบินได้อีกด้วย

## 1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

1. เพื่อศึกษาและวิเคราะห์ปัจจัยที่มีผลกับการกำหนดทางวิ่งให้กับเครื่องบินขาเข้า
2. เพื่อสร้างแบบจำลอง Machine Learning ที่เหมาะสมสำหรับการกำหนดทางวิ่งให้กับ  
เครื่องบินขาเข้าของสนามบินสุวรรณภูมิ

## 1.3 สมมติฐานของการศึกษา

1. ระบุปัจจัยสำคัญที่มีผลต่อการตัดสินใจการกำหนดทางวิ่งให้กับเครื่องบินขาเข้าได้
2. ได้แบบจำลอง Machine Learning ที่สามารถกำหนดทางวิ่งให้กับเครื่องบินได้อย่าง  
แม่นยำและรวดเร็ว

## 1.4 ขอบเขตการวิจัย

1. ปัจจัยสำหรับการกำหนดทางวิ่งให้กับเครื่องบินขาเข้าของสนามบินสุวรรณภูมิ
2. แบบจำลอง Machine Learning ใช้ข้อมูลของเที่ยวบิน และตำแหน่งอากาศยานของ  
เที่ยวบินที่ขึ้นและลงในสนามบินสุวรรณภูมิสำหรับการเรียนรู้และทดสอบ

## 1.5 ขั้นตอนของการศึกษา

1. ศึกษางานวิจัยที่เกี่ยวข้องกับการใช้เทคโนโลยี Machine Learning
2. คัดเลือกประเภท Machine Learning ที่จะนำมาทดลอง
3. วิเคราะห์ข้อมูลแผนการบินและข้อมูลการจราจรทางอากาศของเครื่องบินที่ทำการขึ้นและ  
ลงสนามบินนานาชาติสุวรรณภูมิ เพื่อจัดทำชุดข้อมูล (Dataset) สำหรับทดสอบ
4. สร้างแบบจำลองที่เลือกมาทดลอง และทดสอบด้วย Dataset
5. วัดผลประสิทธิภาพ และเปรียบเทียบผลที่ได้ของแต่ละแบบจำลอง
6. สรุปผลการทดลอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และแจ้งอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

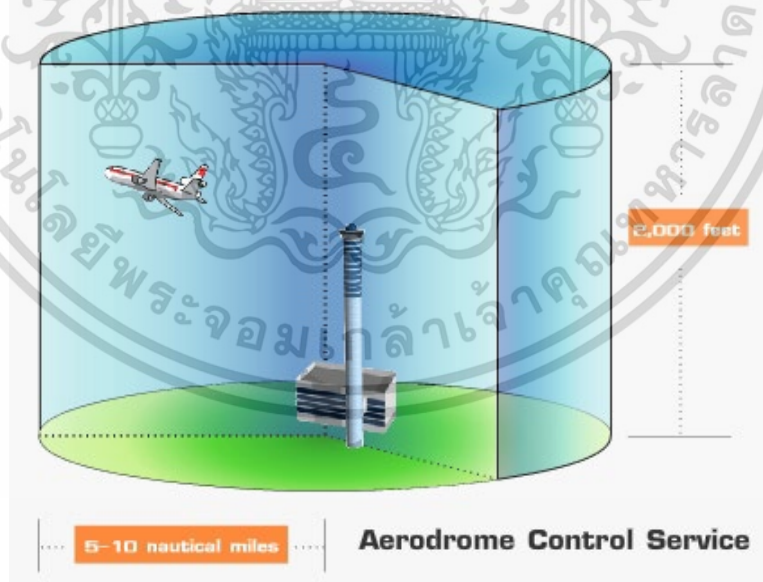
## บทที่ 2

### หลักการและทฤษฎีที่เกี่ยวข้อง

บทนี้จะนำเสนอหลักการและทฤษฎีที่เกี่ยวข้องกับงานวิจัย ซึ่งประกอบด้วยการจัดการจราจรทางอากาศของสนามบิน ความหมายของปัญญาประดิษฐ์ (Artificial Intelligence, AI) ทฤษฎีการเรียนรู้ของเครื่อง (Machine Learning, ML) ของแบบจำลองที่นำมาศึกษาในงานวิจัย และการวัดหาประสิทธิภาพของแบบจำลอง

#### 2.1 การจัดการจราจรทางอากาศ

การจัดการหรือการควบคุมจราจรทางอากาศบริเวณสนามบิน จะให้บริการควบคุมจราจรทางอากาศ ณ บริเวณสนามบิน โดยแต่ละสนามบินจะทำการควบคุมจราจรทางอากาศครอบคลุมรัศมี 5 - 10 ไมล์ทะเล (NM) โดยรอบสนามบิน ที่ความสูงตั้งแต่พื้นดินถึง 2,000 ฟุต (ft) โดยเฉลี่ย [7] ดังแสดงในรูปที่ 2.1 นอกจากนี้ยังมีหน้าที่ และความรับผิดชอบในการควบคุมจราจรทางอากาศให้กับอากาศยานบนทางวิ่ง ทางขับ ลานจอด ในพื้นที่รับผิดชอบบริเวณโดยรอบสนามบิน โดยผู้ควบคุมจราจรทางอากาศจะปฏิบัติงานอยู่บนหอบังคับการบิน และมีความรับผิดชอบในการการควบคุมจราจรทางอากาศบนทางวิ่ง และอากาศยานโดยรอบสนามบิน ให้ขับเคลื่อนด้วยความปลอดภัย เป็นระเบียบ และรวดเร็ว รวมทั้งกำหนดทางวิ่งขึ้น-ลง ให้เหมาะสมกับสภาพอากาศ และสถานการณ์ต่าง ๆ



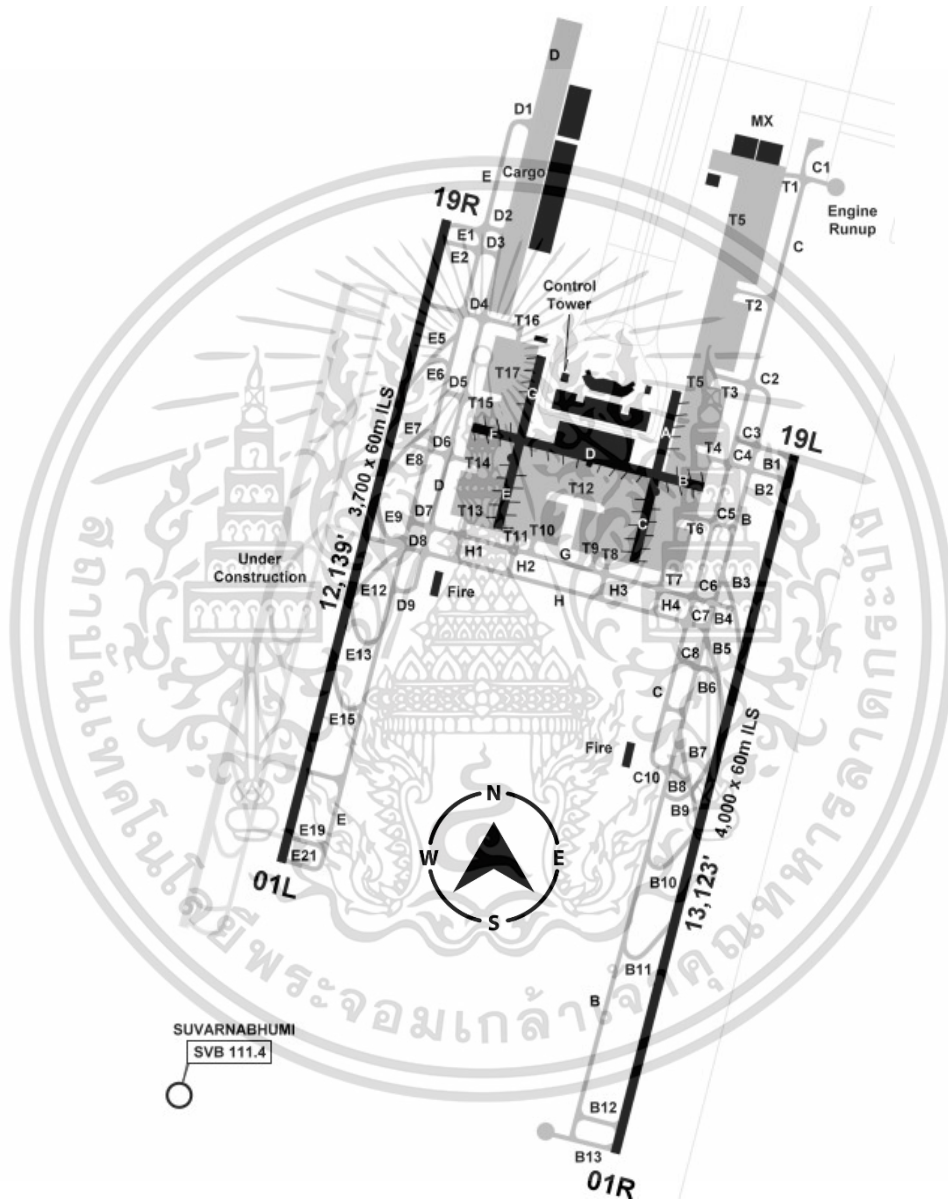
รูปที่ 2.1 พื้นที่การควบคุมจราจรทางอากาศบริเวณสนามบิน

<https://www.aerothai.co.th/sites/default/files/images/air-traffic-control/aerodome.jpg>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 2.1.1 สนามบินนานาชาติสุวรรณภูมิ (VTBS)

สนามบินนานาชาติสุวรรณภูมิมีทางวิ่งคู่ขนานกัน 2 ทางวิ่ง ดังแสดงในรูปที่ 2.2 ห่างกัน 2,200 เมตร โดยทางวิ่ง 01L/19R มีความยาว 4,000 เมตร และทางวิ่ง 01R/19L มีความยาว 3,700 เมตร ทางวิ่งทั้งสองทางสามารถใช้งานพร้อมกันเพื่ออำนวยความสะดวกให้เครื่องบินขาออก และขาเข้าทำการบินขึ้น และลงได้พร้อมกัน [7]

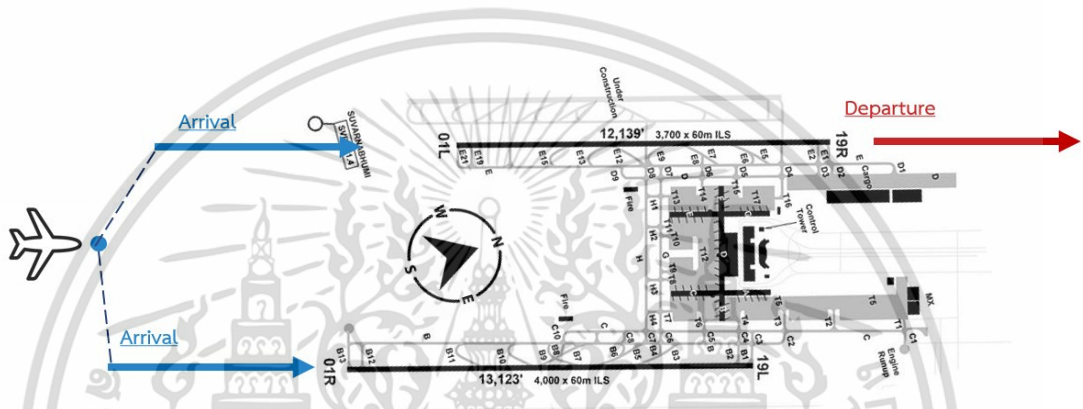


รูปที่ 2.2 แผนผังทางวิ่งสนามบินนานาชาติสุวรรณภูมิ

[https://geofs.fandom.com/wiki/Bangkok\\_Suvarnabhumi\\_Airport?file=BKK\\_charts.jpg](https://geofs.fandom.com/wiki/Bangkok_Suvarnabhumi_Airport?file=BKK_charts.jpg)

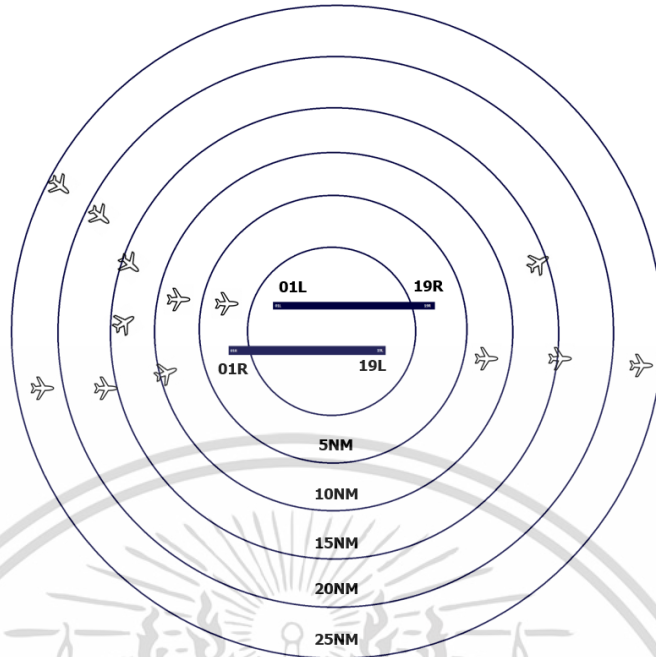
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การจัดการลำดับของเครื่องบินขาเข้า และขาออกที่เข้าคิวเพื่อลงจอด และขึ้นบินบนทางวิ่ง ทั้งสองทางภายในสนามบินนานาชาติสุวรรณภูมินั้น ผู้ควบคุมการจราจรทางอากาศที่ปฏิบัติงานอยู่บน หอบังคับการบินพิจารณา และตัดสินใจทิศทางของทางวิ่งสำหรับใช้ขึ้นหรือลง ขึ้นอยู่กับทิศทางลม โดยต้องกำหนดให้ใช้ทางวิ่งทิศตรงข้ามทิศทางลม [7] ซึ่งสนามบินนานาชาติสุวรรณภูมิ แบ่งทางวิ่ง ออกได้เป็น 4 ทิศทาง คือ 19R, 19L, 01R และ 01L โดยปกติ เมื่อมีลมมาจากทางทิศเหนือของ สนามบิน ทางวิ่ง 01R และ 01L จะถูกกำหนดให้เครื่องบินใช้ลงจอด และบินขึ้น ดังแสดงในรูปที่ 2.3 หากกล่าวอีกนัยหนึ่งทางวิ่ง 19R และ 19L จะถูกกำหนดให้ใช้ลงจอด และบินขึ้นในทิศทางของลมทิศ ใต้ เนื่องจากการนำเครื่องบินขึ้นหรือลงในทิศตรงกันข้ามกับลม จะเพิ่มประสิทธิภาพการลงจอด และ บินขึ้นได้อย่างปลอดภัย



รูปที่ 2.3 ทางวิ่งของสนามบินนานาชาติสุวรรณภูมิ ที่ให้บริการในเงื่อนไขของลมทางทิศเหนือ

โดยทั่วไปแล้ว ผู้ควบคุมการจราจรทางอากาศสนามบินนานาชาติสุวรรณภูมิจะใช้ข้อมูล การจราจรที่ 20 Nautical Mile, NM ในการประเมินระยะทางจากสนามบิน และกำหนดทางวิ่งลง จอดให้กับเครื่องบินที่เดินทางมาถึง ซึ่งผู้ควบคุมการจราจรทางอากาศสามารถมองเห็นข้อมูล การจราจรเข้าได้จากหน้าจอ Air Situation Display (ASD) ในกรณีที่การจราจรออกเบาบางไม่ หนาแน่น ผู้ควบคุมการจราจรทางอากาศจะกำหนดทางวิ่งทั้งสองทางวิ่งให้กับเครื่องบินขาเข้าสามารถ ลงได้พร้อมกันทั้งสองทางวิ่ง และจะจัดการจราจรให้มีเครื่องบินเข้ามาลงจอดบนทางวิ่ง 01L ที่บ่อย กว่าบนทางวิ่ง 01R ในกรณีที่การจราจรขาออกบนทางวิ่ง 01L น้อยกว่าการจราจรบนทางวิ่ง 01R ดัง รูปที่ 2.4 การจราจรทางอากาศที่ 25 NM รอบสนามบินนานาชาติสุวรรณภูมิ



รูปที่ 2.4 การจราจรทางอากาศที่ 25NM รอบสนามบินนานาชาติสุวรรณภูมิ

ทั้งนี้การกำหนดทางวิ่งให้กับเครื่องบินเข้ากรณีที่ทำให้ทำการลงพร้อมกันทั้งสองทางวิ่ง ผู้ควบคุมการจราจรทางอากาศจำเป็นต้องอาศัยการพิจารณาปริมาณการจราจรทางอากาศที่เห็นบน ASD ในขณะที่การพิจารณาใช้เวลาจำกัด เนื่องจากผู้ควบคุมการจราจรทางอากาศยังต้องพิจารณาข้อมูลอื่น ๆ อีกปริมาณมาก สำหรับที่ต้องใช้ในควบคุมการจราจร เช่น ความเร็วของเครื่องบินเพื่อกำหนดระยะห่างระหว่างเครื่องบิน [8] ดังนั้นในการศึกษานี้จึงใช้ข้อมูลแผนการบิน และข้อมูลการติดตามจากระบบเรดาร์ทุติยภูมิสนามบินนานาชาติสุวรรณภูมิที่ระยะทาง 20 NM จากสนามบินนำมาเป็นข้อมูลสำหรับแบบจำลอง

## 2.2 ปัญญาประดิษฐ์ (Artificial Intelligence, AI)

AI คือ เครื่องจักร (Machine) มีความสามารถในการทำความเข้าใจ เรียนรู้องค์ความรู้ต่าง ๆ เช่น การรับรู้ การเรียนรู้ การให้เหตุผล และการแก้ปัญหาต่าง ๆ เครื่องจักรที่มีความสามารถเหล่านี้ก็ถือว่าเป็น AI นั่นเอง [9] ในปัจจุบัน AI มีความสามารถหลายหลายรูปแบบ หลัก ๆ ดังนี้ [10]

- **การเรียนรู้ของเครื่องจักร (Machine Learning)** เป็นความสามารถในการเรียนรู้และวิเคราะห์ข้อมูลของเครื่องคอมพิวเตอร์ เพื่อใช้สำหรับการพยากรณ์ผลลัพธ์จากข้อมูลต่าง ๆ ที่ได้ทำการเรียนรู้

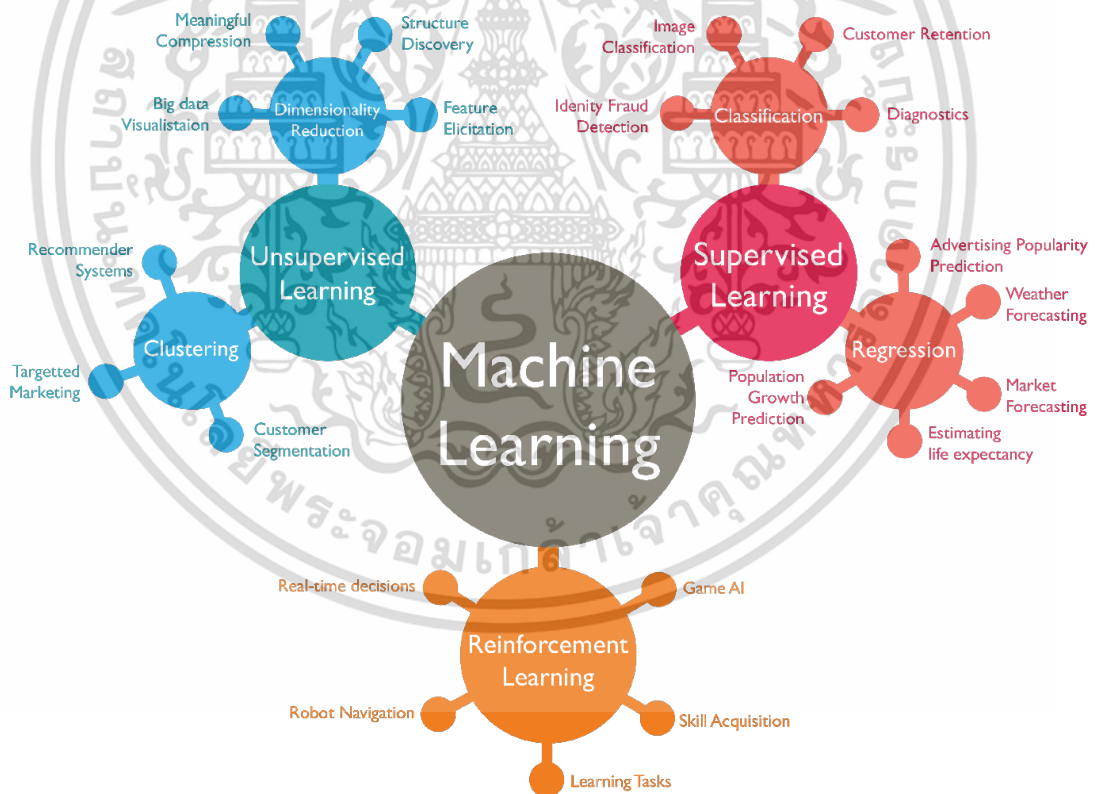
- **การเรียนรู้เชิงลึก (Deep Learning)** ใช้โครงข่ายประสาทเทียมขนาดใหญ่ที่มีหน่วยประมวลผลหลายชั้น โดยอาศัยประโยชน์จากความก้าวหน้าในศักยภาพของคอมพิวเตอร์ และเทคนิคในการเรียนรู้รูปแบบของข้อมูลปริมาณมหาศาลที่มีความซับซ้อนที่ได้รับการพัฒนาให้ดียิ่งขึ้นแล้ว ส่วนใหญ่ใช้งานในประเภทการจดจำภาพและคำพูด
- **ความสามารถในการวางแผน (Automated Planning, Scheduling & Optimization)** เป็นการทำให้เครื่องสามารถตัดสินใจ และเลือกการดำเนินงานที่บรรลุเป้าหมายอย่างมีประสิทธิภาพ
- **ความสามารถในการรู้จำคำพูด (Speech)** เป็นเทคนิคการทำให้เครื่องรู้จำเสียงพูด และโครงสร้างภาษาของมนุษย์และวิเคราะห์ข้อมูลจากเสียงนั้น
- **การประมวลผลภาพ (Computer Vision)** ใช้การจดจำรูปแบบ และการเรียนรู้เชิงลึกในการจดจำสิ่งที่อยู่ในภาพหรือวิดีโอ เมื่อเครื่องสามารถประมวลผลวิเคราะห์ และเข้าใจรูปภาพ มันจะสามารถจับภาพหรือวิดีโอได้แบบเรียลไทม์และตีความสภาพแวดล้อมได้
- **การประมวลผลภาษาธรรมชาติ (Natural Language Processing, NLP)** คือความสามารถของคอมพิวเตอร์ในการวิเคราะห์ ทำความเข้าใจและสร้างภาษามนุษย์ ซึ่งรวมถึงคำพูดด้วย ขั้นตอนถัดไปของ NLP คือ การโต้ตอบด้วยภาษาธรรมชาติ ซึ่งช่วยให้มนุษย์สามารถสื่อสารกับคอมพิวเตอร์ได้โดยใช้ภาษาเพื่อดำเนินการงานต่าง ๆ

## 2.3 การเรียนรู้ของเครื่องจักร (Machine learning, ML) [11]

Machine Learning คือ สมอ่งที่การเรียนรู้ทั้งหมดเกิดขึ้น วิธีที่เครื่องเรียนรู้เหมือนกับมนุษย์มนุษย์เรียนรู้จากประสบการณ์ ยังมีประสบการณ์มาก ยิ่งง่ายต่อการพยากรณ์ว่าสิ่งต่อไปอะไรจะเกิดขึ้น โดยเปรียบเทียบเมื่อประสบกับเหตุการณ์ที่ไม่เคยเจอมาก่อน มีความเป็นไปได้ที่ความสำเร็จจะลดลงกว่าเหตุการณ์ที่เคยเจอมาแล้ว เครื่องจักรสามารถถูกฝึกฝน (Train) ได้ในรูปแบบเดียวกันเพื่อที่จะเพิ่มความแม่นยำในการพยากรณ์ เครื่องจักรสามารถมองเห็นตัวอย่างที่ต้องการให้เห็นได้ เมื่อนำตัวอย่างที่เหมือนกันให้เครื่องจักรดู ซึ่งเครื่องจักรจะสามารถค้นหาผลลัพธ์นั้นจนพบได้ อย่างไรก็ตามเหมือนกับมนุษย์ ถ้าเครื่องจักรถูกสั่งให้ค้นหาในสิ่งที่ไม่เคยถูกฝึกฝนมาก่อน ก็ไม่สามารถที่จะพยากรณ์ผลลัพธ์ออกมาได้ถูกต้อง

โดย Machine Learning นั้นจำแนกออกได้เป็น 3 รูปแบบ คือ

1. การเรียนรู้โดยมีผู้สอน (Supervised Learning) เป็นการเรียนรู้ที่เครื่องจักรจะต้องอาศัยข้อมูลใน การฝึกฝน เปรียบเสมือนกับการเรียนการสอนของเด็กเล็ก โดยจำเป็นจะต้องอาศัยชุดข้อมูลต่าง ๆ ซึ่งประกอบด้วยชุดของข้อมูล และชุดผลลัพธ์ของข้อมูลที่ต้องการ โดยผลที่ได้จากการเรียนรู้ คือ Machine Learning สามารถคาดคะเนผลลัพธ์ที่เกิดขึ้นจากการได้รับข้อมูล
2. การเรียนรู้โดยไม่มีผู้สอน (Unsupervised Learning) เป็นการเรียนรู้ที่เครื่องจักรนั้นสามารถเรียนรู้ได้ด้วยตนเอง โดยไม่ต้องมีค่าเป้าหมายของแต่ละข้อมูล ซึ่งวิธีการคือมนุษย์จะเป็นผู้ใส่ข้อมูลต่าง ๆ และกำหนดสิ่งที่ต้องการจากข้อมูลเหล่านั้น โดยให้เครื่องจักรวิเคราะห์จากการจำแนก และสร้างแบบแผนจากข้อมูลที่ได้รับมา
3. การเรียนแบบเสริมกำลัง (Reinforcement Learning) เป็นการเรียนรู้สิ่งต่าง ๆ จากการลองผิดลองถูก ภายใต้แนวคิดที่ว่าจะเลือกกระทำสิ่งๆ ทำให้ได้ผลลัพธ์มากที่สุด โดยทำการเรียนรู้จากการลองผิดลองถูกในสถานการณ์ในอดีตหรือระบบจำลอง และพยายามที่จะพัฒนาระบบการตัดสินใจของตัวเองให้ดีขึ้นเรื่อย ๆ โดยที่อาจจะพัฒนาด้วยการพยายามสร้างแบบจำลองสถานการณ์ต่าง ๆ



รูปที่ 2.5 แผนภาพการจำแนกรูปแบบของ Machine Learning

<https://www.towardsdatascience.com/machine-learning-types-2-c1291d4f04b1>

### 2.3.1 Logistic Regression Analysis

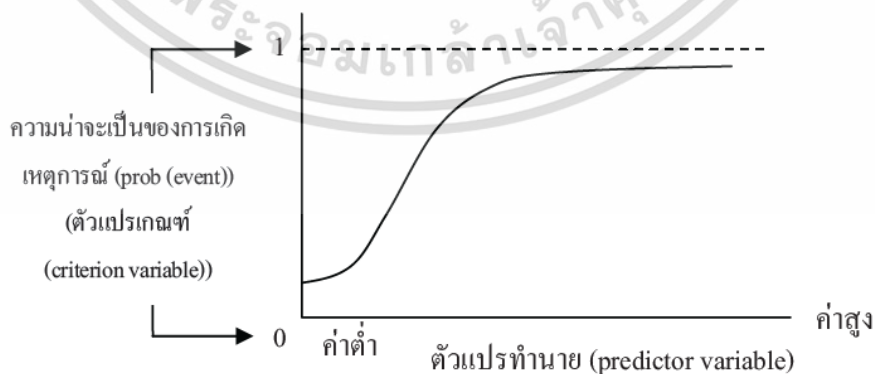
การวิเคราะห์การถดถอยโลจิสติก (Logistic Regression Analysis) เป็นเทคนิคการวิเคราะห์สถิติเชิงคุณภาพ (Qualitative Statistical Techniques) ที่มีวัตถุประสงค์เพื่อประมาณค่าหรือทำนายเหตุการณ์ที่สนใจว่าจะเกิดหรือไม่เกิดเหตุการณ์นั้น ภายใต้อิทธิพลของปัจจัยที่เกี่ยวข้องกับเหตุการณ์ [12] การวิเคราะห์การถดถอยโลจิสติกแบ่งเป็น 2 ประเภท คือ

1. การวิเคราะห์การถดถอยโลจิสติกทวิ (Binary Logistic Regression Analysis) ใช้กับตัวแปรตามที่แบ่งออกเป็น 2 กลุ่มย่อย (Dichotomous Variable) มี 2 ค่า คือ 0 กับ 1 เช่น กลุ่มที่มีเหตุการณ์กับกลุ่มที่ไม่มีเหตุการณ์
2. การวิเคราะห์การถดถอยโลจิสติกพหุกลุ่ม (Multinomial Logistic Regression Analysis) ใช้กับตัวแปรที่มีหลายค่ามากกว่า 2 กลุ่ม (Polytomous Variable) เช่น โรงพยาบาลมีมาตรฐานการให้บริการ สูง ปานกลาง และต่ำ

แบบจำลองการวิเคราะห์การถดถอยโลจิสติกทวิ (Binary Logistic Regression Analysis Model) กรณีตัวแปรทำนาย 1 ตัว ในการวิเคราะห์การถดถอยอย่างง่าย (Simple Regression Analysis) สมการที่แสดงความสัมพันธ์ระหว่าง  $x$  กับ  $y$  จะอยู่ในรูปเชิงเส้น ดังสมการที่ (2.1)

$$y = b_0 + b_1x + e \quad (2.1)$$

แต่สำหรับการวิเคราะห์โลจิสติกที่เป็นแบบทวิ ตัวแปรตามหรือตัวแปรเกณฑ์ ( $y$ ) มี 2 ค่า คือ ไม่เกิดเหตุการณ์ ( $y = 0$ ) หรือเกิดเหตุการณ์ ( $y = 1$ ) มีความสัมพันธ์กับตัวแปรทำนาย ( $x$ ) ไม่อยู่ในรูปเชิงเส้น ทั้งนี้เพราะตัวแปรตามมี 2 ค่า คือ 0 กับ 1 จึงเป็นไปได้ที่ความสัมพันธ์จะอยู่ในรูปเส้นตรง ซึ่งความสัมพันธ์ของตัวแปรของการวิเคราะห์โลจิสติกจะอยู่ในรูปคล้ายตัว S ดังรูปที่ 2.6



รูปที่ 2.6 ฟังก์ชันโลจิสติก (Logistic Function) [12]

โดยที่สมการของ Logistic Function แสดงได้ดังสมการที่ (2.2)

$$p(y) = \frac{1}{1+e^{-f(x)}} \quad (2.2)$$

เมื่อ  $p(y)$  คือ ความน่าจะเป็นของการเกิดเหตุการณ์  $y$ ,  $e$  คือ exponential function และ  $f(x)$  คือ ฟังก์ชันของตัวแปรทำนาย

กรณีตัวแปรทำนายมากกว่า 1 ตัว ในการวิเคราะห์เมื่อตัวแปรทำนายมีมากกว่า 1 ตัว จะได้ฟังก์ชันดังสมการที่ (2.3)

$$P_y = \frac{e^{b_0+b_1x_1+\dots+b_px_p}}{1+e^{b_0+b_1x_1+\dots+b_px_p}} \quad (2.3)$$

เมื่อ  $P_y$  เป็นความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจ จะได้  $Q_y$  หรือความน่าจะเป็นของการไม่เกิดเหตุการณ์ที่สนใจ ดังสมการที่ (2.4)

$$Q_y = 1 - P_y \quad (2.4)$$

หรือเขียนได้อีกแบบหนึ่ง คือ สมการที่ (2.5)

$$Q_y = 1 - \left( \frac{e^{b_0+b_1x_1+\dots+b_px_p}}{1+e^{b_0+b_1x_1+\dots+b_px_p}} \right) \quad (2.5)$$

จากความสัมพันธ์ระหว่างตัวแปรทำนายกับตัวแปรเกณฑ์ของการวิเคราะห์ถดถอยโลจิสติกไม่เป็นรูปเชิงเส้น จึงต้องมีการปรับให้ความสัมพันธ์ให้อยู่ในรูปเชิงเส้น ในรูปแบบของ odds หรือ odd ratio หมายถึง อัตราส่วนระหว่างโอกาสที่จะเกิดเหตุการณ์ที่สนใจ ( $y = 1$ ) กับโอกาสที่จะไม่เกิดเหตุการณ์ที่สนใจ ( $y = 0$ ) จะได้ดังสมการที่ (2.6)

$$odds = \frac{P_y}{Q_y} \quad (2.6)$$

ค่าของ odds แสดงถึงโอกาสที่จะเกิดเหตุการณ์ที่สนใจ เป็นที่เท่าของโอกาสที่จะไม่เกิดเหตุการณ์ที่สนใจ นั่น คือ ถ้า odds มีค่ามากกว่า 1 แสดงว่า โอกาสที่จะเกิดเหตุการณ์ที่สนใจนั้นมากกว่าโอกาสที่จะไม่เกิดเหตุการณ์ที่สนใจ

การเขียนโมเดลโลจิสติก จะอยู่ในรูป log ของ odds เรียกว่า logit หรือ logistic response function ซึ่ง logit เขียนในรูปสมการดังสมการที่ (2.7)

$$\log(\text{odds}) = \log\left(\frac{P_y}{Q_y}\right) \quad (2.7)$$

เมื่อ  $Q_y = 1 - P_y$  จะได้ดังสมการที่ (2.8)

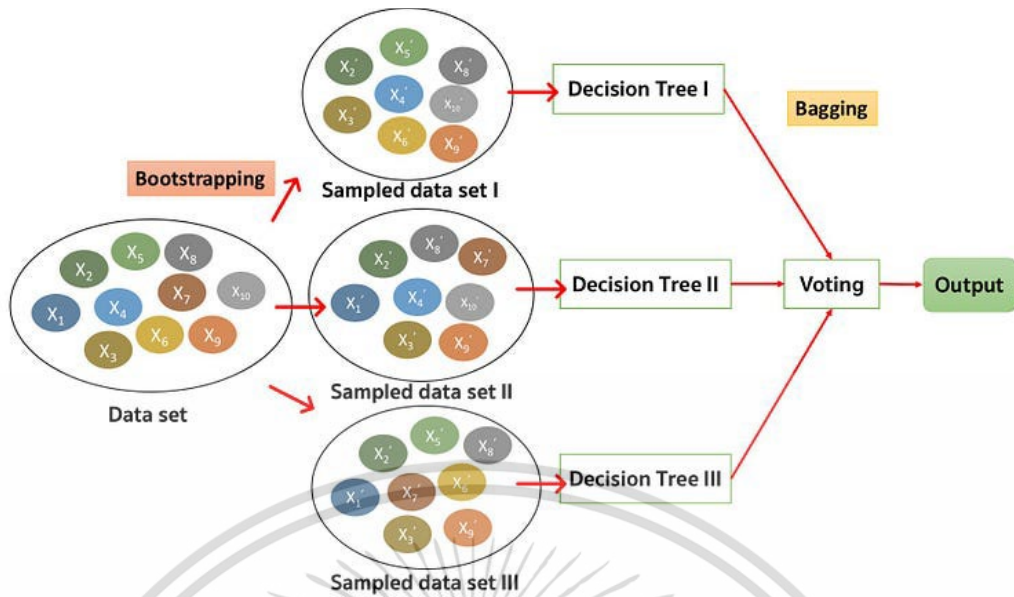
$$\text{logit} = \log(\text{odds}) = b_0 + b_1x_1 + \dots + b_px_p \quad (2.8)$$

เมื่อได้ log ของ odds ratio หรือ logit แล้ว รูปแบบของตัวแปรเกณฑ์ จึงสามารถทำนายได้ด้วยชุดของตัวแปรทำนายเชิงเส้นตรง

สำหรับการทำนายค่า  $y$  ที่เป็น  $P_y$  ในการวิเคราะห์ถดถอยโลจิสติกจะใช้สมการที่ (2.3) ตามวิธี Maximum Likelihood [12]

### 2.3.2 Random Forest [11]

สำหรับ Random Forest เป็นอัลกอริทึมที่นำมาใช้ได้ดีทั้งปัญหาแบบการถดถอย (Regression) และการจำแนก (Classification) ซึ่งจะนำเอาอัลกอริทึมแบบ Decision Trees ที่ได้หลาย ๆ ต้นมาฝึกฝนข้อมูลรวมกัน โดยมีหลักการนำเอาข้อมูลมาฝึกฝนกับแบบจำลองหลาย ๆ ครั้ง บนข้อมูลชุดเดียวกัน ในกรณีที่ปัญหา Classification จะใช้วิธีผลโหวตมากที่สุด (Majority vote) โดยค่าพยากรณ์ของ Decision Trees ที่ได้รับค่าผลโหวตมากที่สุด จะถูกเลือกให้เป็นค่าพยากรณ์ของปัญหา ดังรูปที่ 2.7 แต่ถ้าเป็นปัญหาวิเคราะห์แบบ Regression จะใช้วิธีคำนวณหาค่าเฉลี่ย (Mean) โดยนำเอา ค่าพยากรณ์ของทุก Decision Trees มาคำนวณหาค่าเฉลี่ยเพื่อแสดงเป็นค่าพยากรณ์ของปัญหา



รูปที่ 2.7 การทำงานของ Random Forest Classifier

[https://miro.medium.com/v2/resize:fit:4800/format:webp/1\\*iFgl9nTtiCupbck8D9mCXg.jpeg](https://miro.medium.com/v2/resize:fit:4800/format:webp/1*iFgl9nTtiCupbck8D9mCXg.jpeg)

### 2.3.3 K-Nearest Neighbors (KNN) [11,13]

อัลกอริทึม K-nearest neighbors (KNN) นับเป็นเทคนิคที่มีวิธีการไม่ซับซ้อน และเข้าใจได้ง่ายที่สุดที่ใช้ ในการจำแนกประเภทข้อมูล โดยหลักการทำงาน คือ จะใช้หลักการเปรียบเทียบความคล้ายคลึงกันของข้อมูลที่สนใจกับข้อมูลอื่นว่ามีค่าความคล้ายคลึงหรืออยู่ใกล้กับข้อมูลใดมากที่สุด k ตัว จากนั้นจะทำการตัดสินใจว่า ค่าตอบของข้อมูลที่สนใจนั้นควรเป็นคำตอบเดียวกับข้อมูลที่อยู่ใกล้ที่สุด k ตัวนั้น ทั้งนี้ k คือ ความถี่ของข้อมูลที่ อยู่ใกล้กับข้อมูลที่สนใจ โดย มาตรฐานระยะห่างที่นิยม ได้แก่ ระยะยูคลิด (Euclidean distance) ดังสมการที่ (2.9)

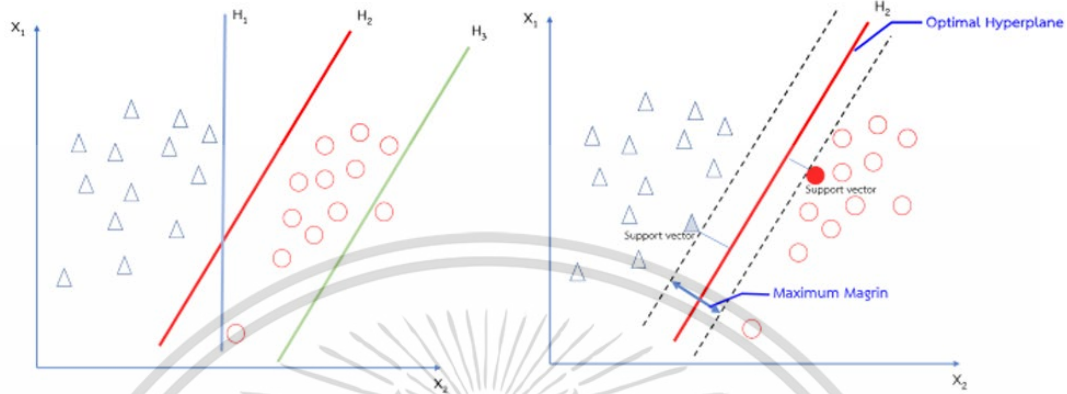
$$dist(p, q) = \sqrt{\sum_{k=1}^n (p_k - q_k)^2} \quad (2.9)$$

เมื่อ  $dist(p, q)$  คือ ระยะห่างของข้อมูลที่คล้ายคลึงกันระหว่างข้อมูลที่สนใจกับข้อมูลอื่น

### 2.3.4 Support Vector Machine (SVM) [11]

ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine, SVM) เป็นเทคนิคหนึ่งที่ได้รับค่านิยมอย่างแพร่หลายในงานที่เกี่ยวข้องกับการจัดจำรูปแบบตลอดจนการแก้ปัญหาการจัดกลุ่ม (Classification Problem) โดยเฉพาะกับปัญหาที่มีขนาดของข้อมูลไม่ใหญ่มาก แต่คุณลักษณะ (Features) ของข้อมูลมีเป็นจำนวนมาก

หลักการทํางานของ SVM จะอาศัยใช้การสร้างเส้นแบ่ง หรือไฮเปอร์เพลน (Hyperplane) ในการแบ่งแยกคลาสของข้อมูลออกจากกัน จากนั้นจะทำการหาว่าไฮเปอร์เพลนใดเป็นเส้นที่ใช้แยกคลาสของข้อมูลได้ดีที่สุด (Optimal Separating Hyperplane) ดังรูปที่ 2.8



รูปที่ 2.8 ตัวอย่าง SVM ใน 2 มิติ [13]

โดยอาศัยหลักการของการหาสัมประสิทธิ์ของสมการเพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ถูกต้องเข้าสู่กระบวนการสอนให้ระบบเรียนรู้ โดยเน้นไปยังเส้นแบ่งแยกกลุ่มข้อมูลได้ดีที่สุด เมื่อพิจารณาข้อมูลที่ประกอบด้วยข้อมูล 2 กลุ่ม ดังสมการที่ (2.10)

$$D = \{(x_i, y_i) ; i = 1, 2, \dots, n\} \quad (2.10)$$

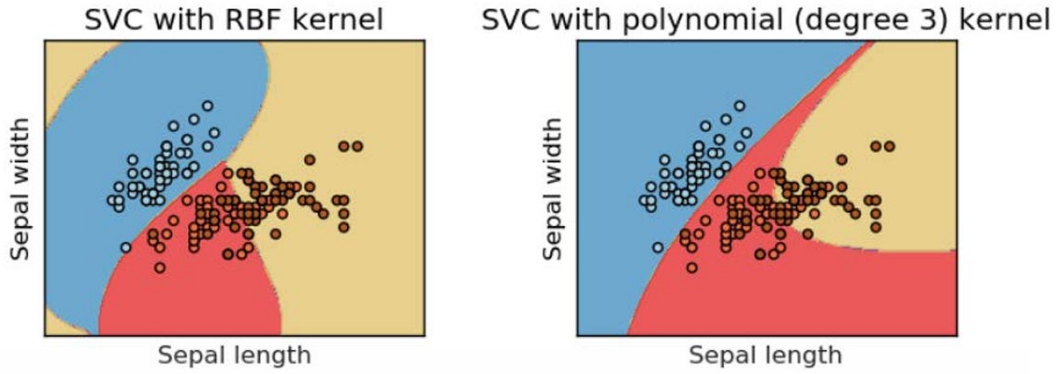
เมื่อ  $x_i = (x_{i1}, x_{i2}, \dots, x_{im}) \in R^m$   
 $y_i \in \{1, -1\}$  โดย 1 คือ ข้อมูลกลุ่ม 1 และ -1 คือ ข้อมูลกลุ่ม 2

ซึ่งเป็นการกำหนดกลุ่มเป้าหมายให้ SVM โดยที่ SVM นั้นมุ่งเป้าเพื่อหาฟังก์ชันการตัดสินใจที่สามารถแบ่งแยกค่าที่ไม่ทราบได้ ดังสมการที่ (2.11) และ (2.12)

$$f(x) = \text{sign}\{\sum_{k=1}^n w_k \varphi_k(x) \varphi_k(x_k) + b\} \quad (2.11)$$

$$\varphi(x) = [\varphi_1(x_1), \varphi_2(x_2), \dots, \varphi_n(x_n)]^T \quad (2.12)$$

อย่างไรก็ตาม สามารถใช้ SVM ในปัญหาที่มีลักษณะข้อมูลไม่เป็นเชิงเส้นได้ ดังรูปที่ 2.9



รูปที่ 2.9 ตัวอย่างไฮเปอร์เพลนที่ใช้ในปัญหาที่ข้อมูลไม่เป็นเชิงเส้น [13]

กลุ่มข้อมูล  $x$  จากสมการที่ (2.13) ไม่สามารถแบ่งแยกได้ด้วยสมการเส้นตรง แต่จะถูกแปลงให้อยู่ในรูปแบบที่สามารถใช้สมการเส้นตรงแบ่งแยกได้ โดยใช้เคอร์เนลฟังก์ชัน (Kernel Function) ดังสมการที่ (2.13)

$$K(x, x_k) = \varphi(x)\varphi(x_k) \quad (2.13)$$

เมื่อ	$\varphi(x)$	คือ	ฟังก์ชันสำหรับแปลงข้อมูลที่ไม่เป็นเชิงเส้นให้เป็นข้อมูลที่อยู่ในรูปเชิงเส้นสามารถแบ่งแยกได้
	$w_k$	คือ	ค่าน้ำหนักที่เชื่อมโยงจาก Feature Space ไปสู่ Output Space
	$b$	คือ	ค่าโน้มเอียง (Bias)
	$x_k$	คือ	ซัพพอร์ตเวกเตอร์ โดย $k = 1, 2, \dots, n_v$
	$n_v$	คือ	จำนวนซัพพอร์ตเวกเตอร์

วิธีการที่ใช้ในการหา Optimal Separating Hyperplane ที่ดีที่สุด คือ เพิ่มเส้นขอบ (Margin) ให้กับเส้นแบ่งทั้งสองข้าง และสร้างเส้นขอบที่สัมผัสกับค่าข้อมูลใน Feature Space ที่ใกล้ที่สุด ดังนั้นเส้นแบ่งที่มีเส้นขอบกว้างที่สุด จึงเป็นเส้นแบ่งที่ดีที่สุด และเรียกตำแหน่งการสัมผัสข้อมูลที่ใกล้ที่สุดจากการเพิ่มขอบนี้ว่า Support Vector เนื่องจากในบางกรณีการแบ่งแยกกลุ่มไม่สามารถทำได้ถูกต้องโดยสมบูรณ์

ดังนั้นจึงต้องมีการกำหนดตัวแปรสำหรับยอมรับค่าความผิดพลาดโดยการเพิ่มตัวแปร  $\delta$  (Slack Variable) ดังสมการที่ (2.14) และ (2.15)

$$w^T x + b \geq y - \delta_i ; y = 1 \quad (2.14)$$

$$w^T x + b \leq y + \delta_i ; y = -1 \quad (2.15)$$

จากการกำหนดค่า  $\delta_i > 0$  ทำให้โครงสร้างของซัพพอร์ตเวกเตอร์แมชชีนบรรลู่วัตถุประสงค์ใน 2 ส่วน คือ การเพิ่มระยะแบ่งแยกให้มากที่สุด และลดข้อผิดพลาดในการทำนายให้ต่ำที่สุด ดังสมการที่ (2.16)

$$\text{Minimize } \frac{1}{2} \|w\|^2 + c \sum_{i=1}^N \delta_i \quad (2.16)$$

$$\text{โดยที่ } y_i(w^T \varphi(x) + b) + \delta_i - 1 \geq 0$$

$$\delta_i \geq 0, i = 1, 2, \dots, N$$

เคอร์เนลฟังก์ชัน (Kernel Function) มีอยู่เป็นจำนวนมาก ซึ่งที่รู้จักกันดีมี 3 ชนิดด้วยกัน คือ

1. โพลีโนเมียล (Polynomial) แสดงดังสมการที่ (2.17)

$$K(x_i, x_j) = (x_i^T x_j + r)^\gamma ; \gamma > 0 \quad (2.17)$$

2. เรเดียลเบสฟังก์ชัน (Radial Basis Function-RBF) แสดงดังสมการที่ (2.18)

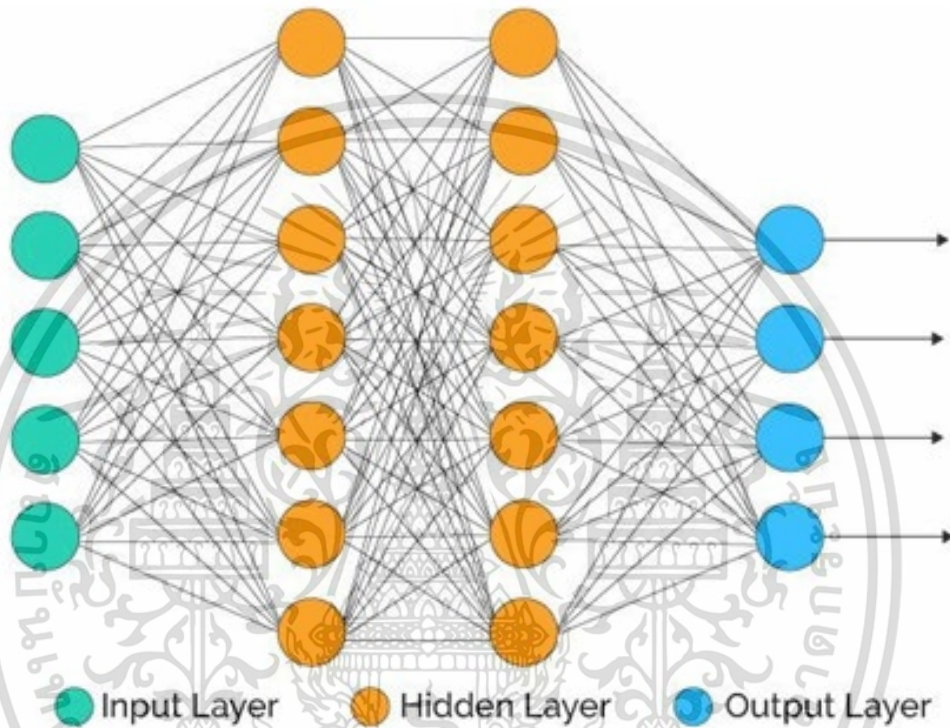
$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) ; \gamma > 0 \quad (2.18)$$

3. ซิกมอยด์ (sigmoid) แสดงดังสมการที่ (2.19)

$$K(x_i, x_j) = \tanh(\gamma x_i^T x_j - r) \quad (2.19)$$

### 2.3.5 Artificial Neuron Network (ANN) [13]

วิธีการสร้างแบบจำลองข้อมูลหรือโมเดลข้อมูล (Data Modeling) โดยวิธีโครงข่ายประสาทเทียมอัจฉริยะ (Artificial Neuron Network: ANN) เป็นวิธีการที่มีพื้นฐานมาจากการเลียนแบบการทำงานของสมองมนุษย์ซึ่งประกอบด้วยนิวรอน (Neurons) ที่ใช้ในการเชื่อมโยงเพื่อแก้ปัญหาต่าง ๆ โดยมีโครงสร้างการทำงาน ตามรูปที่ 2.10 ซึ่งประกอบด้วยข้อมูลเข้า (Input) 1 ชั้น และข้อมูลออก (Output) 1 ค่า จะเรียกว่า Perceptron

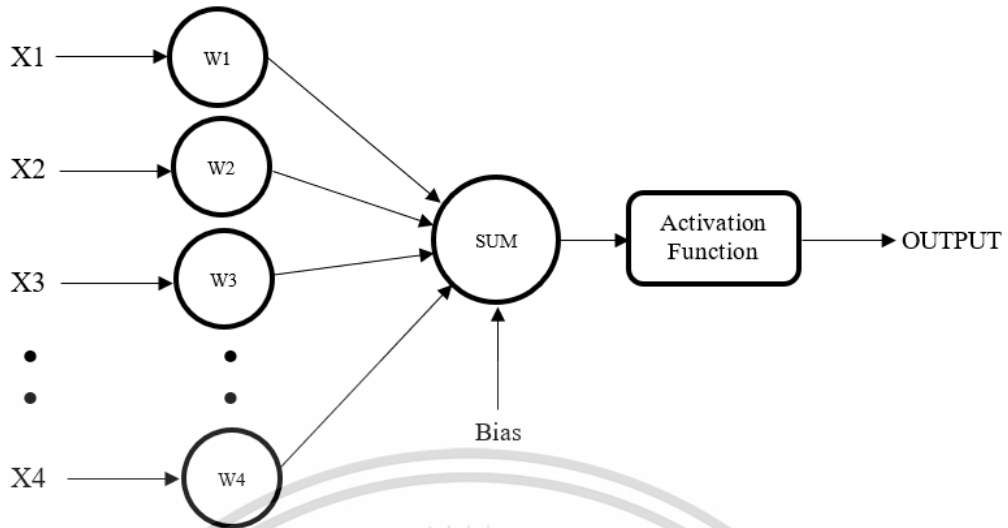


รูปที่ 2.10 ส่วนประกอบของโครงข่ายประสาทเทียม

โครงข่ายประสาทเทียมประกอบด้วย 3 ส่วน

1. ชั้นข้อมูลเข้า (Input Layer)
2. ชั้นซ่อน (Hidden Layer)
3. ชั้นข้อมูลออก (Output Layer)

โดยโครงข่ายประสาทเทียมจะนำข้อมูลเข้าและค่าน้ำหนักแต่ละปลายประสาทมาคูณกัน และนำค่าทั้งหมดมารวมกัน หากค่าที่รวมกันนี้เกินกว่าค่าแบ่งที่กำหนดไว้ (Threshold,  $T$ ) จะทำการส่งต่อข้อมูลไปยังฟังก์ชันกระตุ้น (Activation Function) ดังรูปที่ 2.11 แสดงแผนผังแสดงเซลล์ประสาทของโครงข่ายประสาทเทียม



รูปที่ 2.11 แผนผังโครงข่ายประสาทเทียม

จำนวนอินพุตระบุจำนวนคุณสมบัติที่เลือกสำหรับแบบจำลอง (แสดงเป็น  $m$ ) ผลลัพธ์ของ ANNs ถูกกำหนดโดยผลลัพธ์ของฟังก์ชันการกระตุ้น ดังแสดงในสมการที่ (2.20) และ (2.21)

$$k = \sum_{i=1}^m w_i x_i + b \quad (2.20)$$

$$Output = f(k) \quad (2.21)$$

เมื่อ  $k$  คือ ผลรวมของน้ำหนัก (แสดงด้วย  $w$ )

$x$  คือ อินพุต

$b$  คือ ไบแอส

ฟังก์ชันกระตุ้นจะถูกนำไปใช้ทั้ง Hidden node และ Output node ซึ่งทั้งสองโหนดอาจจะใช้ฟังก์ชันกระตุ้นที่เหมือนหรือต่างกันได้ แต่ส่วนมากจะใช้ฟังก์ชันแบบไม่เป็นเชิงเส้น ฟังก์ชันกระตุ้นจะมีหลากหลายรูปแบบ ดังต่อไปนี้

- ฟังก์ชันกระตุ้นซิกมอยด์ (Sigmoid Activation Function) เป็นฟังก์ชันทางคณิตศาสตร์ที่มีลักษณะเป็น “S-curve” หรือเรียกว่า “Sigmoid curve” จะมีค่าระหว่าง 0 และ 1 ฟังก์ชันนี้จะใช้เมื่อต้องการทำนายความน่าจะเป็น (Probability) ของข้อมูลออก ดังสมการ (2.22)

$$f(k) = \frac{1}{1+e^{-k}} \quad (2.22)$$

- ฟังก์ชันเรกติไฟด์ลินีเยยูนิต (Rectified Linear Units, ReLu) เป็นฟังก์ชันที่นิยมใช้งานมากที่สุดในโครงข่ายประสาทเทียมแบบบิด (Convolutional Neural Networks: CNN) และโครงข่ายประสาทเทียมอัจฉริยะ (ANN) ฟังก์ชันนี้ ถ้าข้อมูลเข้ามีค่ามากกว่าศูนย์ ข้อมูลออกจะเป็นค่าบวก และถ้าข้อมูลเข้ามีค่าศูนย์หรือติดลบ ข้อมูลออกจะมีค่าเป็นศูนย์

## 2.4 การวัดและประเมินผลประสิทธิภาพของแบบจำลอง Classification

วิธีการมาตรฐานที่ใช้ในการประเมินหรือวัดประสิทธิภาพของแบบจำลอง นั้นมีหลากหลายวิธี วิธีการหนึ่งที่เป็นพื้นฐาน คือ การแบ่งหมวดหมู่ของผลลัพธ์ โดยจะต้องมีข้อมูลผลลัพธ์จริง และผลลัพธ์ที่ได้จากแบบจำลองแล้วนำมาเทียบกันดู โดยในเรื่องการประเมินนี้ มีคำสำคัญที่จะต้องทราบความหมายอยู่หลายคำ หลัก ๆ คือ

- True Positive (TP) คือ ทำนายว่า “จริง” และมีค่าเป็น “จริง”
- True Negative (TN) คือ ทำนายว่า “ไม่จริง” และมีค่าเป็น “ไม่จริง”
- False Positive (FP) คือ ทำนายว่า “จริง” แต่มีค่าเป็น “ไม่จริง”
- False Negative (FN) คือ ทำนายว่า “ไม่จริง” แต่มีค่าเป็น “จริง”
- Accuracy ค่าความถูกต้อง
- Precision ค่าความเที่ยงตรง
- Sensitivity ค่าความไว
- Recall ค่าความระลึก

### 2.4.1 Confusion Matrix [14]

ในการเปรียบเทียบข้อมูลที่ได้จากการเรียนรู้ และการทำนายของแบบจำลอง สามารถแสดงเป็นตารางหรือเมทริกซ์ ที่เรียกว่า Contingency Table หรือ Confusion Matrix ได้ดังรูปที่ 2.12

		Predicted	
		Negative (N) -	Positive (P) +
Actual	Negative -	True Negatives (TN)	False Positives (FP) Type I error
	Positive +	False Negatives (FN) Type II error	True Positives (TP)

รูปที่ 2.12 รูปแบบของ Confusion Matrix

จากรูปที่ 2.12 ตัวแปรทั้ง 4 ตัวนี้ (TP, FP, TN, และ FN) เป็นตัวที่ใช้บ่งบอกประสิทธิภาพของแบบจำลอง ผ่านค่าสำคัญที่มีนิยามดังนี้

- ค่าความถูกต้อง (Accuracy) สามารถคำนวณได้ดังสมการที่ (2.23)

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.23)$$

- ค่าความเที่ยงตรง (Precision) สามารถคำนวณได้ดังสมการที่ (2.24)

$$Precision = \frac{TP}{TP+FP} \quad (2.24)$$

- ค่าความระลึก (Recall) สามารถคำนวณได้ดังสมการที่ (2.25)

$$Recall = \frac{TP}{TP+FN} \quad (2.25)$$

- ค่าความจำเพาะ (Specificity) สามารถคำนวณได้ดังสมการที่ (2.26)

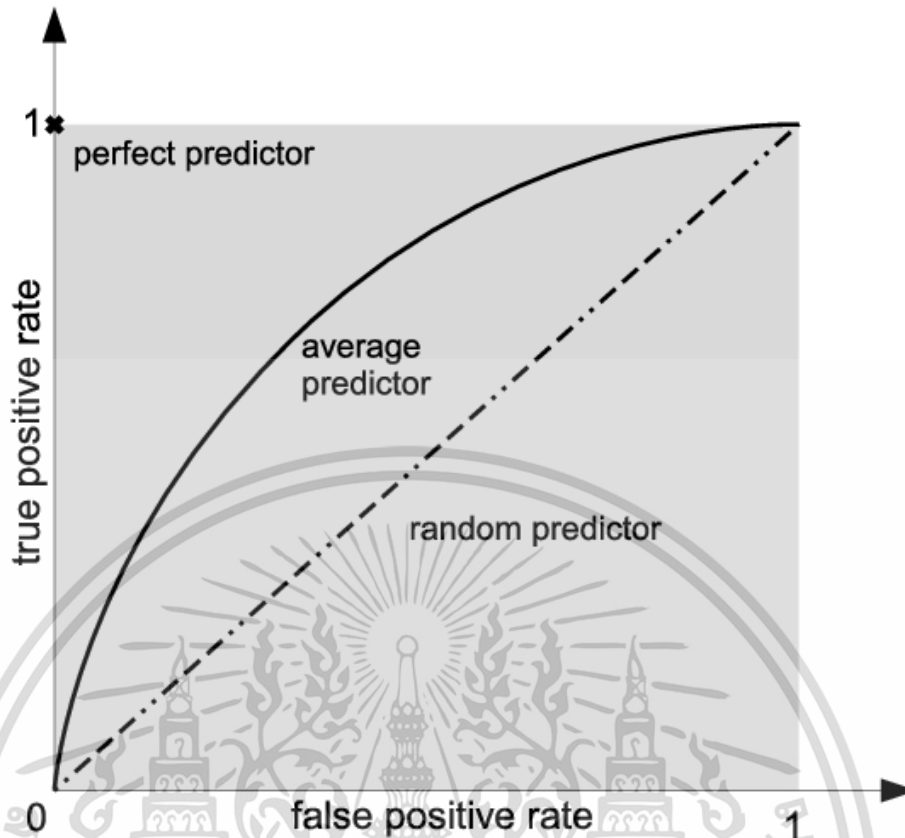
$$Specificity = \frac{TN}{TN+FP} \quad (2.26)$$

- ค่าเฉลี่ยความถูกต้อง (F1-Score) สามารถคำนวณได้ดังสมการที่ (2.27)

$$F_1 = 2 \frac{Precision \cdot Recall}{Precision+Recall} \quad (2.27)$$

#### 2.4.2 Receiver Operating Characteristic (ROC) Curve [15]

ROC Curve เป็นเส้นที่ใช้วัดถึงประสิทธิภาพของแบบจำลองแบบ Classification ว่าสามารถทำนายประเด็นที่สนใจได้อย่างแม่นยำขนาดไหน โดยทั่วไปนิยมวัดประสิทธิภาพของแบบจำลองแบบ Binary เส้น ROC นั้นวัดสัดส่วนของ True Positive เทียบกับสัดส่วนของ False Positive Sensitivity และ (1-Specificity) จึงถูกเรียกอีกชื่อหนึ่งว่า True Positive Rate และ False Positive Rate ดังรูปที่ 2.13



รูปที่ 2.13 ROC Curve [16]

โดยตัวแปรที่ใช้บอกความถูกต้องของแบบจำลองได้แก่ ความถูกต้อง และพื้นที่ใต้เส้นโค้ง ROC (Area Under Curve, AUC) เกิดจากจุดคู่อันดับระหว่าง True Positive Rate ในแกน Y และ False Positive Rate ในแกน X สร้าง เป็นเส้นโค้ง และคำนวณพื้นที่ใต้เส้นโค้งนั้น ซึ่งแบบจำลองที่มีความถูกต้องในระดับที่ยอมรับได้ จุดคู่อันดับนี้ต้องอยู่ในตำแหน่งเหนือเส้น Random Predictor หากต่ำกว่าเส้น Random Predictor แบบจำลองนั้นนับว่าเป็นแบบจำลองที่ใช้งานไม่ได้

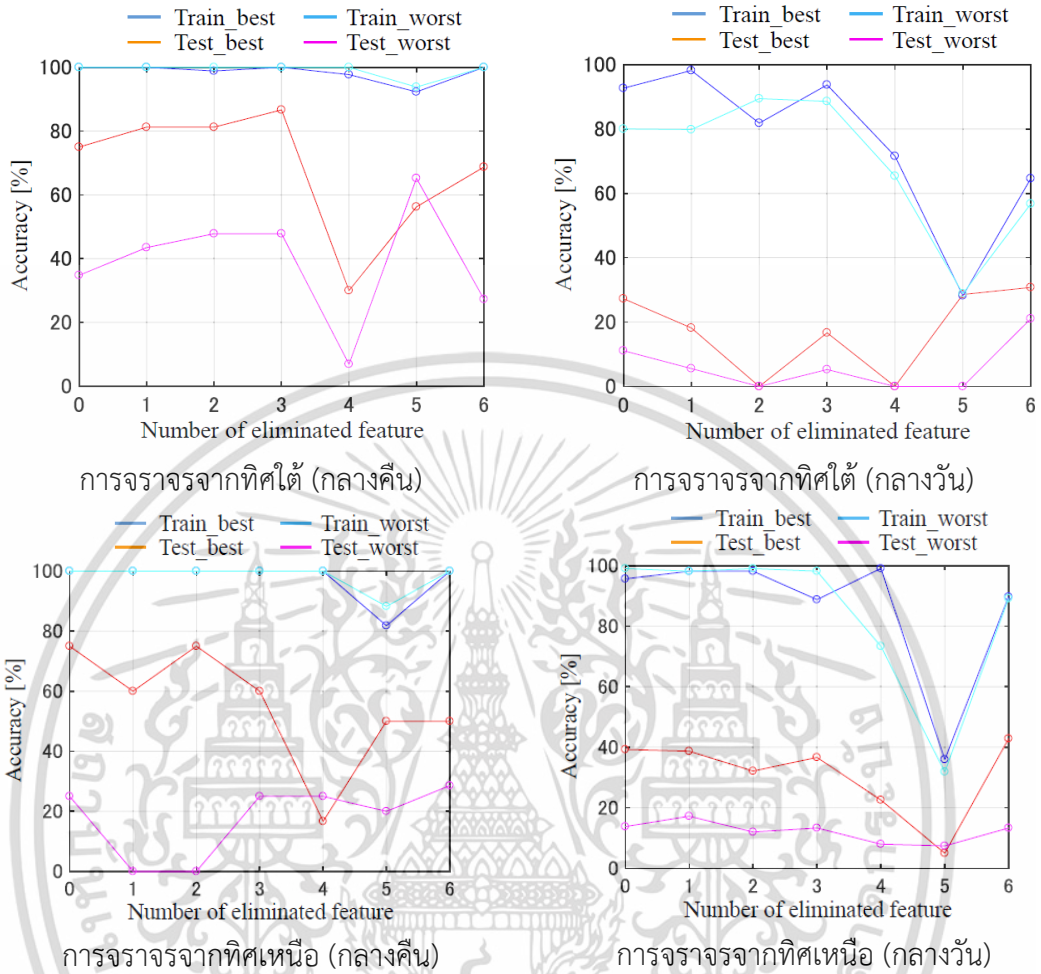
## 2.5 งานวิจัยที่เกี่ยวข้อง

ปี ค.ศ. 2016 งานวิจัยเรื่อง Predicting Runway Allocation with Support Vector Machine and Logistic Regression [17] ได้ศึกษาการทำนายการจัดสรรทางวิ่งให้กับเครื่องบินด้วย Machine Learning โดยใช้เทคนิคของ SVM และ Logistic Regression มุ่งเน้นไปที่การใช้ Machine Learning ช่วยในการตัดสินใจเพื่อปรับปรุงการดำเนินงานของผู้ควบคุมการจราจรทางอากาศ ช่วยในการคาดการณ์การจัดสรรทางวิ่งที่เหมาะสมที่สุดสำหรับเที่ยวบินขาเข้าและขาออกแต่ละเที่ยวบิน ด้วยชุดข้อมูลที่ประกอบไปด้วยสภาพอากาศ การซ่อมบำรุง น้ำหนักเครื่องบิน และการจราจร ณ ขณะนั้นๆ โดยให้ความแม่นยำของการทำนายแบบจำลองทั้งสองแบบ 55% สำหรับแบบจำลอง SVM with linear kernel และ 56% สำหรับแบบจำลอง Logistic Regression ดังรายละเอียดดังตารางที่ 2.1

ตารางที่ 2.1 ความแม่นยำเปรียบเทียบระหว่างสองแบบจำลอง

Algorithm	Classification Accuracy	Precision	Recall	F1-score
SVM with linear kernel	55%	49%	55%	47%
Logistic Regression	56%	51%	56%	49%

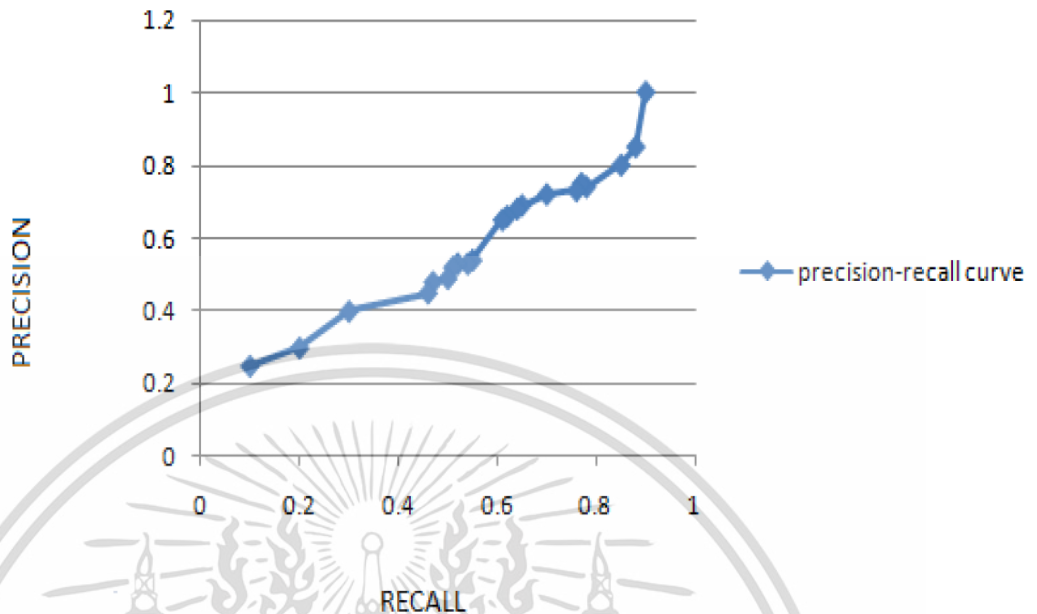
ปี ค.ศ. 2017 งานวิจัยเรื่อง Modeling of Runway Assignment Strategy by Human Controllers using Machine Learning [18] ได้ศึกษาการออกแบบจำลองโครงข่ายประสาทเทียมแบบ Multi-Layer Perceptron เพื่อมาทำนายทางวิ่งลงจอดกำหนดให้กับเครื่องบินขาเข้าสำหรับท่าอากาศยานนานาชาติโตเกียว ภายใต้ขั้นตอนการสลับทางวิ่งเพื่อลงจอดเพื่อเพิ่มความสามารถในการรองรับปริมาณเที่ยวบินเพิ่มขึ้น โดยผู้ควบคุมจราจรทางอากาศจะสั่งให้อากาศยานขาเข้าเปลี่ยนทางวิ่งที่จะใช้ลง ทั้งนี้เพื่อลดภาระงานได้มีการพัฒนาแบบจำลองโครงข่ายประสาทเทียมแยกอิสระจำนวน 4 แบบ ซึ่งพบว่าอาคารผู้โดยสารที่กำหนดหรือ Feature ที่ 4 เป็นหนึ่งในคุณสมบัติหลัก เนื่องจากความแม่นยำมีแนวโน้มลดลง เมื่อไม่มีข้อมูลอาคารผู้โดยสารที่กำหนด ทั้งนี้ ยังพบว่าความแม่นยำของข้อมูลการฝึกนั้นลดลงเมื่อไม่มีข้อมูลสถานการณ์การจราจรหรือ Feature ที่ 5 เช่นกัน แสดงให้เห็นว่าข้อมูลอาคารผู้โดยสารที่กำหนด และข้อมูลสถานการณ์การจราจร เป็นคุณสมบัติหลักที่ถูกใช้ในการทำนายของแบบจำลอง ดังรูปที่ 2.14



รูปที่ 2.14 ความแม่นยำของการจรรยาจรรทั้ง 4 รูปแบบ

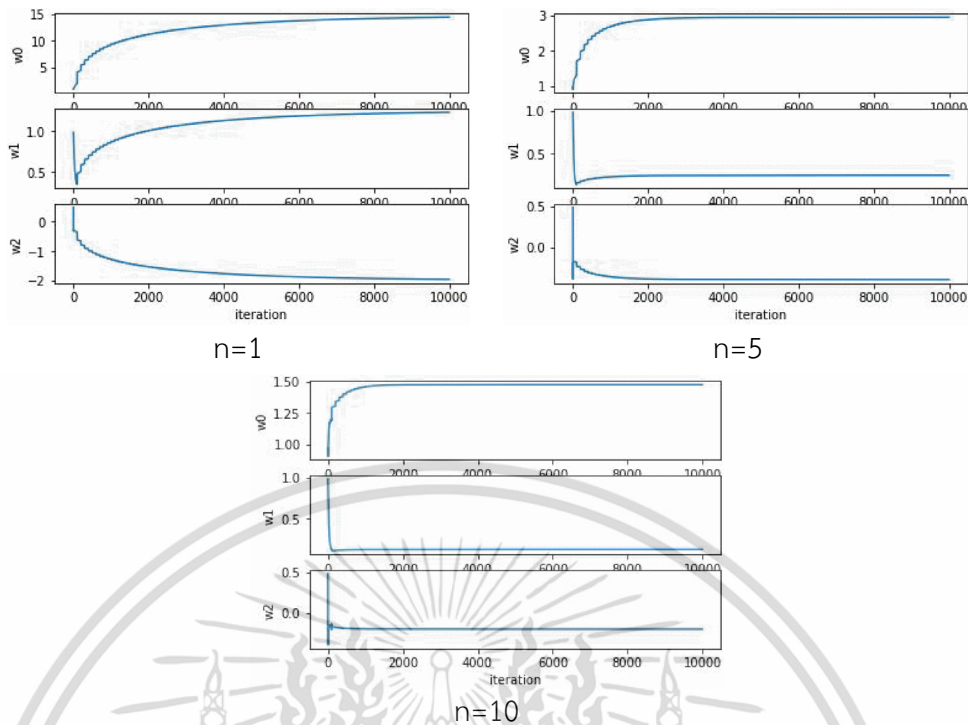
ปี ค.ศ. 2018 งานวิจัยเรื่อง Prediction of Airline delays using K Nearest Neighbor algorithm [19] ได้ศึกษาการใช้อัลกอริทึม KNN เพื่อทำนายหาความล่าช้าของเที่ยวบิน ที่เกิดจากสภาพอากาศแปรปรวน โดยเฉพาะอย่างยิ่ง แบบจำลองในการศึกษานี้สร้างขึ้นจากข้อมูลสภาพอากาศและเที่ยวบินในอดีตโดยใช้อัลกอริทึมการเรียนรู้ของเครื่อง K Nearest Neighbor ซึ่งให้ค่าความแม่นยำในการทำนาย 90% ค่าความไว 95% และค่าความเที่ยงได้ 90% ดังรูปที่ 2.15

## precision-recall curve



รูปที่ 2.15 ความแม่นยำของ KNN ในการทำนายหาความล่าช้าของเที่ยวบิน

ปี ค.ศ. 2019 งานวิจัยเรื่อง Logistic Regression Model Optimization and Case Analysis [20] ได้ศึกษาการเพิ่มประสิทธิภาพแบบจำลองการถดถอยโลจิสติก โดยการปรับตัวแปร  $n$  ของ Sigmoid function แตกต่างกันว่า  $n=1,5$  และ  $10$  เพื่อดูความสัมพันธ์ระหว่างตัวแปร  $n$  และค่าประสิทธิภาพการทำงานของอัลกอริทึม ซึ่งจากผลการทดสอบจะเห็นได้ว่าเมื่อ  $n=1$  ค่าสัมประสิทธิ์การถดถอยแบบวนซ้ำ 8000 รอบ เมื่อ  $n=5$  จะค่าสัมประสิทธิ์การถดถอยแบบวนซ้ำ 2500 รอบ เมื่อ  $n=10$  ค่าสัมประสิทธิ์การถดถอยแบบวนซ้ำ 1,000 รอบ สรุปได้ว่ายิ่ง  $n$  มาก ยิ่งต้องการจำนวนการวนซ้ำที่น้อยลง ดังรูปที่ 2.16 คือ ประสิทธิภาพการทำงานของอัลกอริทึมจะสูงขึ้น



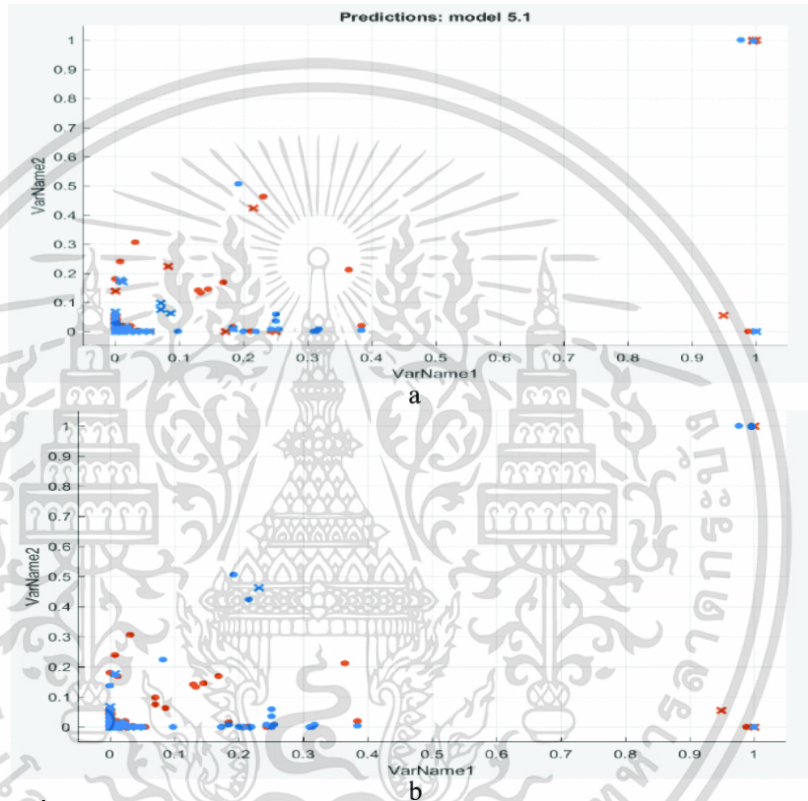
รูปที่ 2.16 Relationship between regression coefficient and iteration

ปี ค.ศ. 2019 งานวิจัยเรื่อง A machine learning model to predict runway exit at Vienna airport [21] ได้ศึกษาการทำนายทางออกของทางวิ่งสำหรับเครื่องบินด้วย Machine Learning ณ สนามบินนานาชาติเวียนนา ประเทศออสเตรีย โดยแบ่งประเภทชุดข้อมูลเป็น 3 แบบตามน้ำหนักของเครื่องบิน คือ หนัก กลาง และเบา สำหรับแบบจำลอง ซึ่งใช้แบบจำลอง Gradient Boosting (GB) ในการทำนาย และใช้ข้อมูลทาง ความเร็ว น้ำหนัก ของเครื่องบินเป็นชุดข้อมูลนำเข้าแบบจำลอง และได้ค่าความแม่นยำหลังทดสอบกับแบบจำลองในแต่ละประเภทน้ำหนัก ดังผลลัพธ์ในตารางที่ 2.2

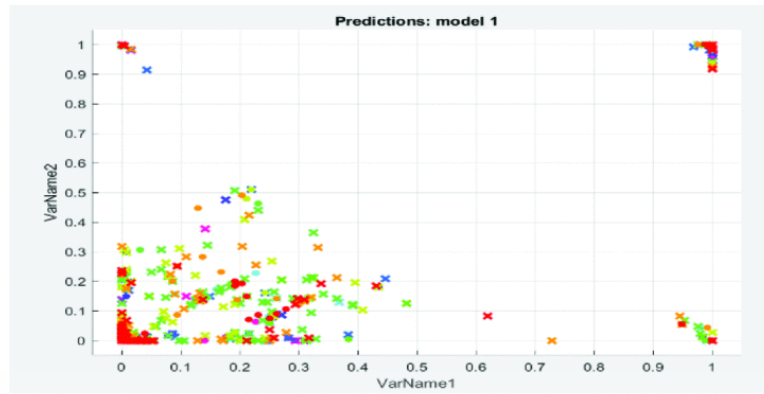
ตารางที่ 2.2 ประสิทธิภาพการคาดการณ์และเวลาในการคำนวณสำหรับเทคนิค GB

Aircraft Weight	Forecast performance (Accuracy in %)	F1-Score	AUC	Computation Time (s)
Heavy	98	Precision = 97.8% Recall = 98.3% F1 = 98%	81.00%	5.5
Medium	95	Precision = 95.8% Recall = 95.6% F1 = 95%	82.00%	5.5
Light	98	Precision = 94.3% Recall = 98.4% F1 = 96%	78.00%	3.5

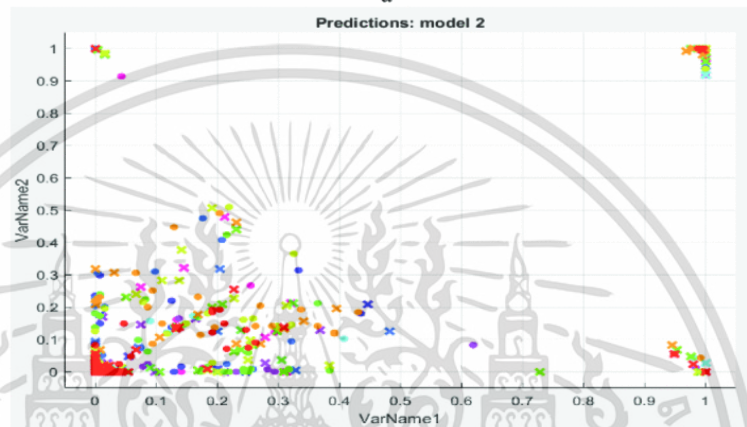
ปี ค.ศ. 2021 งานวิจัยเรื่อง Kannada Alphabets Recognition using Decision Tree and Random Forest Models [22] ได้ศึกษาเกี่ยวกับการใช้แบบจำลอง Decision Tree เปรียบเทียบกับการใช้แบบจำลอง Random Forest ในการระบุตัวอักษร Kannada โดยทดสอบการจำแนกทั้งแบบ Binary Classification และ Multiclass Classification ซึ่งได้ผลของความแม่นยำในการจำแนกในกรณี Binary Classification 82% - 84.6% สำหรับ Decision Tree และ 86.4% - 89.39% สำหรับ Random Forest โดยมีรูปแบบการกระจายตัวของข้อมูลดังรูปที่ 2.17 สำหรับ Binary Classification และรูปที่ 2.18 สำหรับ Multiclass Classification



รูปที่ 2.17 Binary: Scatter plot. a) Decision Tree b) Random Forest



a



b

รูปที่ 2.18 Multiclass: Scatter plot. a) Decision Tree b) Random Forest

ได้ผลของความแม่นยำในการจำแนกในกรณี Multiclass Classification 40.6% - 47% สำหรับ Decision Tree และ 72% - 76.7% สำหรับ Random Forest ดังตารางที่ 2.3

ตารางที่ 2.3 ผลลัพธ์การเปรียบเทียบความแม่นยำระหว่าง 2 แบบจำลอง

Type of Classification	Decision Tree		Random Forest	
	Accuracy	Error	Accuracy	Error
Binary Classification	82% - 84.6%	18% - 15.4%	86.4% - 89.39%	13.6% - 10.61%
Multiclass Classification	40.6% - 47%	59.4% - 53%	72% - 76.7%	28% - 23.3%

จากการเปรียบเทียบพบว่าแบบจำลอง Random Forest ให้ค่าความแม่นยำดีกว่า Decision Tree อย่างเห็นได้ชัดโดยเฉพาะในการจำแนกแบบ Multiclass Classification

ปี ค.ศ. 2021 งานวิจัยเรื่อง A machine learning framework for predicting ATC conflict resolution strategies for conformal automation [23] ได้ศึกษาการทำนายกลยุทธ์ในการแก้ไขปัญหาการจราจรชนกัน ด้วยการนำ Machine Learning 3 เทคนิค ประกอบไปด้วย Random Forest, XGBoost และ SVM มาเปรียบเทียบกับชุดข้อมูลชุดเดียวกัน จะได้ผลลัพธ์ของการทำนายจากแบบจำลองของ Random Forest มีประสิทธิภาพโดยรวมแม่นยำที่สุด ดังรายละเอียดในตารางที่ 2.4

ตารางที่ 2.4 ผลลัพธ์การเปรียบเทียบความแม่นยำระหว่าง Random Forest, XGBoost และ SVM

Metric	Model Performance (ATC A, ATC B)		
	Random Forest	XGBoost	SVM
Classification accuracy, % (Choice of aircraft)	95.1, 93.7	95.01, 95.07	92.45, 92.8
MAE, Nm (Maneuver initiation distance)	0.38, 0.52	0.47, 0.45	0.77, 1.39
MAE, Nm (Heading angle)	5.15, 3.66	5.36, 3.15	5.70, 5.55
MAE, Nm (Cross track deviation)	1.18, 1.63	1.29, 1.69	1.24, 1.59
Classification accuracy, % (Merging waypoint)	93.60, 99.20	92.09, 98.80	93.30, 98.80

ปี ค.ศ. 2022 งานวิจัยเรื่อง Power Theft Detection Using Novel Linear SVM Algorithm and Compared With Convolutional SVM Algorithm For Accuracy [24] ได้ศึกษาการใช้ Novel Linear SVM Algorithm และ Convolutional SVM Algorithm เพื่อทดสอบประสิทธิภาพ และเปรียบเทียบความแม่นยำในการตรวจจับโจรขโมยไฟฟ้า โดยแบ่งกลุ่มตัวอย่างให้กับแบบจำลองละ 10 ตัวอย่างจากการทดลองพบว่าผลลัพธ์ความแม่นยำเฉลี่ย (Mean Accuracy) ของ Novel Linear SVM Method ได้ร้อยละ 90.24 มากกว่า Convolutional SVM ที่ได้ความแม่นยำเฉลี่ยร้อยละ 81.84 การศึกษาแสดงให้เห็นว่าอัลกอริทึม Linear SVM ใหม่มีความแม่นยำในการตรวจจับการขโมยพลังงานอย่างมีนัยสำคัญ และเชื่อถือได้มากกว่าเมื่อเปรียบเทียบกับโมเดล Convolutional SVM

ตารางที่ 2.5 ค่าความแม่นยำเฉลี่ยของ Novel Linear SVM กับ Convolutional SVM

Group	Total Number of Testing	Mean Accuracy	Std. Deviation	Std. Error Mean
Linear SVM	20 times	90.24%	1.10	0.24
Convolutional SVM	20 times	81.84%	1.36	0.30

ปี ค.ศ. 2022 งานวิจัยเรื่อง Comparative Analysis using K - Nearest Neighbor with Artificial Neural Network to Improve Accuracy for Predicting Road Accidents [25] ได้ศึกษาการใช้ K - Nearest Neighbor (KNN) และ Artificial Neural Network (ANN) เพื่อทดสอบประสิทธิภาพ และเปรียบเทียบความแม่นยำในการทำนายการเกิดอุบัติเหตุบนท้องถนน ผลของการทดสอบแสดงให้เห็นว่า การทำนายด้วย KNN ให้ค่าความแม่นยำเฉลี่ยร้อยละ 81.22 ซึ่งมากกว่า ANN ที่มีความแม่นยำเฉลี่ยร้อยละ 69.22

ตารางที่ 2.6 ผลลัพธ์เชิงสถิติของ KNN เทียบกับ ANN

Algorithm	Total Number of Testing	Mean Accuracy	Std. Deviation	Std. Error Mean
K-Nearest Neighbors	10 times	81.22%	6.83	2.16
Artificial Neural Networks	10 times	69.22%	5.72	1.81

ปี ค.ศ. 2022 งานวิจัยเรื่อง A decision support system for safer airplane landings: Predicting runway conditions using XGBoost and explainable AI [26] ได้ศึกษาการทำนายเงื่อนไขของสภาพทางวิ่ง เพื่อเพิ่มความปลอดภัยในการนำเครื่องบินลงจอดบนทางวิ่ง โดยใช้เทคนิค Extreme Gradient Boosting นำไปทดสอบ และเปรียบเทียบกับแบบจำลองอื่นๆ ซึ่งได้ผลลัพธ์ในการทดสอบออกมาเป็นรูปแบบ Confusion Matrices ดังตารางที่ 2.7

ตารางที่ 2.7 Confusion matrices for the predicting runway conditions using XGBoost

		XGBoost		Runway		Scenario		Snowtam	
		Slippery	Non-Slippery	Slippery	Slippery	Non-Slippery	Slippery	Non-Slippery	Non-Slippery
Actual	Slippery	4,740	423	3,905	1,258	4,223	940	4,006	1,157
	Non-Slippery	28,863	166,482	46,967	148,378	78,894	116,451	20,679	174,666

## บทที่ 3

# วิธีการดำเนินการวิจัย

### 3.1 การออกแบบการทดลอง

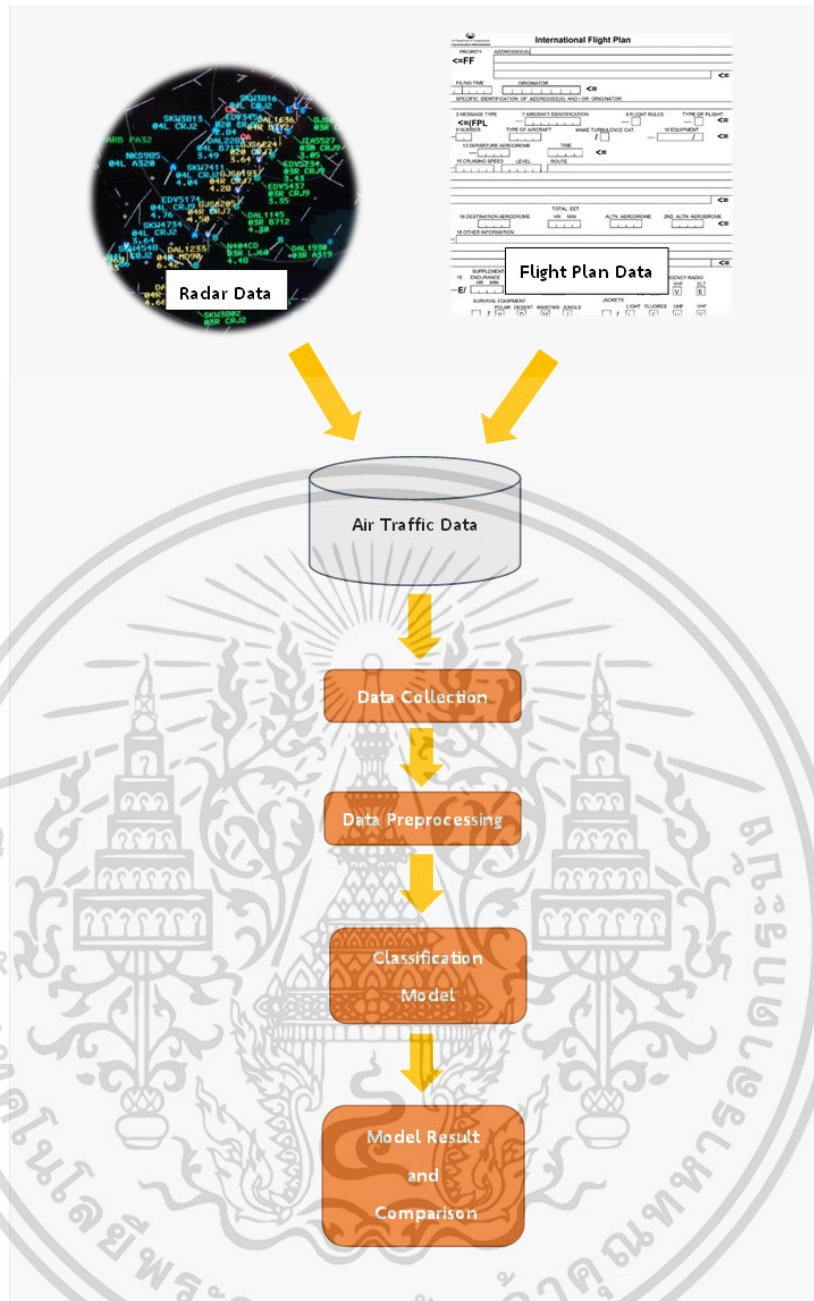
จากการศึกษาบทความหรืองานวิจัยที่มีความเกี่ยวข้องกับการศึกษา Machine Learning เพื่อค้นหาและทดสอบแบบจำลองที่เหมาะสมที่จะนำมาใช้ในการกำหนดทางวิ่งของเครื่องบินเข้า ณ สนามบินสุวรรณภูมิ โดยนำเอา Machine Learning ที่ใช้ในการจำแนก (Classification) หลากหลายอัลกอริทึมมาเปรียบเทียบกัน ได้แก่

- Logistic Regression
- Random Forest
- Support Vector Machine (SVM) - Radial Basis Function (RBF)
- K-Nearest Neighbor (KNN)
- Artificial Neural Networks (ANN)

โดยนำข้อมูลที่สนใจมาเข้าดำเนินการตามกระบวนการขั้นตอนดังต่อไปนี้

1. คัดเลือกชุดข้อมูล (Data Collection)
2. การประมวลผลข้อมูลล่วงหน้า (Data Preprocessing)
3. นำชุดข้อมูลที่ได้เข้าสู่แบบจำลอง Classification แต่ละอัลกอริทึม
4. วัดผลประสิทธิภาพของผลลัพธ์ที่ได้และเปรียบเทียบประสิทธิภาพระหว่างแบบจำลองแต่ละแบบ

ดังแสดงในรูปที่ 3.1 แสดงภาพรวมขั้นตอนการทดลอง



รูปที่ 3.1 ภาพรวมขั้นตอนการทดลอง

### 3.2 ข้อมูล

ในการศึกษานี้ใช้ข้อมูลการจราจรของสนามบินสุวรรณภูมิ โดยใช้ข้อมูลจำนวน 7 วัน รวมปริมาณเครื่องบินทั้งหมด 1,479 ลำ สำหรับการฝึกฝนและทดสอบทำนายเพื่อการกำหนดทางวิ่งลงจอดสำหรับการจราจรภายในรัศมี 20 NM รอบสนามบิน ซึ่งข้อมูลที่น่ามาใช้ในการศึกษานี้จะแบ่งออกได้เป็น 2 ประเภท คือ

1. ข้อมูลเรดาร์ (Radar Data)
2. ข้อมูลแผนการบิน (Flight Plan)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.2.1 ข้อมูลเรดาร์ (Radar Data)

เป็นข้อมูลที่ได้รับมาจากระบบเรดาร์สนามบินสุวรรณภูมิ ประกอบไปด้วยข้อมูลตำแหน่งของเครื่องบิน รวมถึงข้อมูลจำเพาะต่าง ๆ ของเครื่องบิน ตัวอย่างดังตารางที่ 3.1

ตารางที่ 3.1 ตัวอย่างข้อมูลเรดาร์

Time (s)	Aircraft Identification	X[NM]	Y[NM]	Aircraft Address	FL	SQUAWK	RHO [NM]
2635.7265	RTP186	-56.2734	174.390	0xa04d1f	260	4411	183.34 375
2635.843, 2635.81	CAL5555, THD102	-3.21093, -4.96093	20.9140, 25.843	0x89911c, 0x88530d	340, 197	265, 107, 217	21.898 4375, 26.585 9375
2636.125, 2636.1093	AIQ5500, FIN141	37.8281, 21.83593	193.765, 123.671	0x88045a, 0x461f4c	310, 400	45, 360, 206	197.58 59375, 125.85 9375
2636.4843, 2636.4843, 2636.4843	THD002, TVJ200	17.4531, 76.2031, 108.625	23.2890, 101.117, 144.398	0x885314, 0x885970	10, 329, 203	11, 100, 422, 504, 414	29.117 1875, 126.81 25, 180.70 3125
2636.8593	THD040	50.35156	25.25781	0x885305	243.5	3301	56.484 375

โดย Time คือ เวลาของข้อมูลเครื่องบิน ณ เวลานั้น ๆ เป็นหน่วย วินาที (s)

Aircraft Identification คือ ชื่อเรียกขานของเครื่องบินแต่ละลำ

X[NM] คือ ตำแหน่งของเครื่องบินในแกน X ห่างจากสถานีเรดาร์ เป็นหน่วย ไมล์ทะเล (NM)

Y[NM] คือ ตำแหน่งของเครื่องบินในแกน Y ห่างจากสถานีเรดาร์ เป็นหน่วย ไมล์ทะเล (NM)

Aircraft Address คือ เลขทะเบียนของเครื่องบินแต่ละลำ

FL คือ ความสูง หรือ แพนดามบิน (Flight Level)

SQUAWK คือ เลขระบุตัวตนของเครื่องบินแต่ละลำ

RHO [NM] คือ ระยะห่างระหว่างเครื่องบินกับเรดาร์ เป็นหน่วย ไมล์ทะเล (NM)

### 3.2.2 ข้อมูลแผนการบิน (Flight Plan)

เป็นข้อมูลรายละเอียดในการบินของของเครื่องบินแต่ละลำ ทั้งสนามบินต้นทาง สนามบินปลายทาง วันเวลาที่ทำการบิน รวมถึงสถานะการติดต่อกับหอบังคับการบินและทางวิ่งที่ถูกกำหนดให้กับเครื่องบินเพื่อนำเครื่องขึ้นหรือลงสนามบิน ตัวอย่างดังตารางที่ 3.2

ตารางที่ 3.2 ตัวอย่างข้อมูลแผนการบิน

Topic	Detail	Value
AFLIGHTDATE	วันที่ของแผนการบิน	24-11-2021 12:00:00 AM
ASSRCODE	เลขระบุตัวตนของเครื่องบินแต่ละลำ	136
AREGISTRATION	ทะเบียนของเครื่องบินแต่ละลำ	A6EFL
AMODES	เลขทะเบียนของเครื่องบินแต่ละลำ	9003378
ACALLSIGN	ชื่อเรียกขานของเครื่องบินแต่ละลำ	UAE9317
AAPPDEPFLAG	เครื่องขาเข้า (A) หรือ ขาออก (D)	A
ADEST	สนามบินปลายทาง	VTBS
AGATE	หลุมจอด	523
ARUNWAY	ทางวิ่งที่กำหนด	01L
AACTYPE	แบบของเครื่องบิน	B77L
AFIRSTCONTACT	เวลาที่ นักบินติดต่อกับหอบังคับการบินครั้งแรก	24-11-2021 12:35:53 AM
AISSUEDFROM WORKSTATION	ตำแหน่งของผู้ควบคุมการจราจรทางอากาศที่ควบคุมเครื่องบิน ณ เวลานั้น	DEFAULT;SBLTLCL04WS
AETD	เวลาเครื่องบินออก กรณีเป็นขาออก	-
AETA	เวลาเครื่องบินถึง กรณีเป็นขาเข้า	24-11-2021 12:27:23 AM
AFPLSTATE	สถานะของแผนการบิน	1ST

### 3.2.3 การคัดเลือกข้อมูล Dataset ของแบบจำลอง

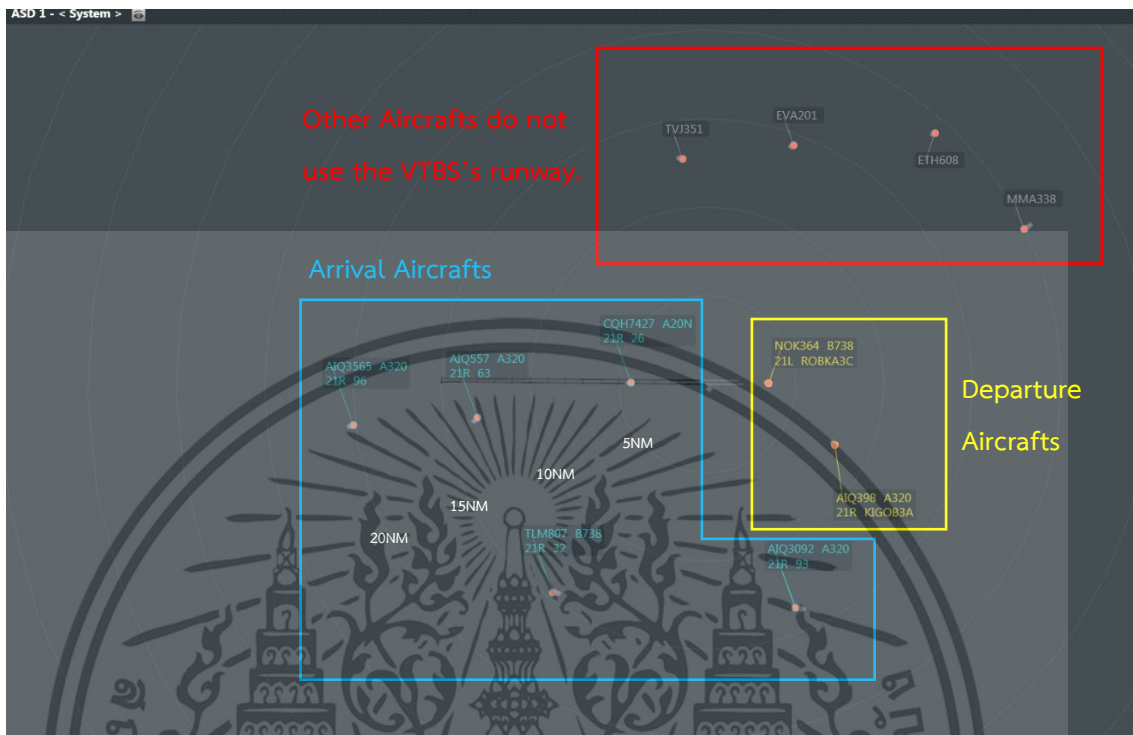
ข้อมูลเรดาร์และข้อมูลแผนการบินที่ได้รับมาเป็นข้อมูลที่ยังไม่ได้มีการจัดระเบียบและคัดเลือกชุดข้อมูลสำหรับนำมาเข้าเป็นอินพุตให้กับแบบจำลอง ดังนั้นจึงจำเป็นต้องทำการวิเคราะห์ข้อมูลและคัดเลือกชุดข้อมูล (Data Collection) ที่เป็นปัจจัยสำคัญและมีอิทธิพลกับกับผลลัพธ์ของการกำหนดทางวิ่ง เพื่อให้ได้ผลลัพธ์ของการทำนายมีความเที่ยงตรงและถูกต้องมากที่สุด

การคัดเลือกข้อมูลในการศึกษานี้ใช้หลักการของการบริหารการจราจรทางอากาศของหอบังคับการบินที่สนามบินสุวรรณภูมิ ในกรณีของทิศทางการบินที่พัฒนาจากทางทิศเหนือ ซึ่งจะใช้ทางวิ่งที่ 01R และ 01L ในการให้บริการขึ้น-ลงของเครื่องบิน ปกติแล้วผู้ควบคุมการจราจรทางอากาศจะเริ่มติดต่อกับนักบินเพื่อจัดลำดับคิวและกำหนดทางวิ่งลงจอดให้กับเครื่องบินขาเข้าเมื่อเครื่องบินอยู่ในระยะ 20 NM ห่างจากทางวิ่ง หรือประมาณ 6 นาทีก่อนนำเครื่องลงจอด โดยการจะพิจารณาว่าจะกำหนดทางวิ่ง 01L หรือ 01R ให้ในการลงจอด ผู้ควบคุมการจราจรทางอากาศจะใช้ข้อมูลการจราจร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ของเครื่องบินที่อยู่ในระยะทำการบินขึ้นและลงในช่วงนั้น หรือที่ 20 NM บนหน้าจอ Air Situation Display (ASD) บนจอควบคุมการบิน เป็นปัจจัยหลักในการตัดสินใจ ดังแสดงในรูปที่ 3.2



รูปที่ 3.2 การจราจรของเครื่องบินบริเวณสนามบินบนหน้าจอ Air Situation Display (ASD)

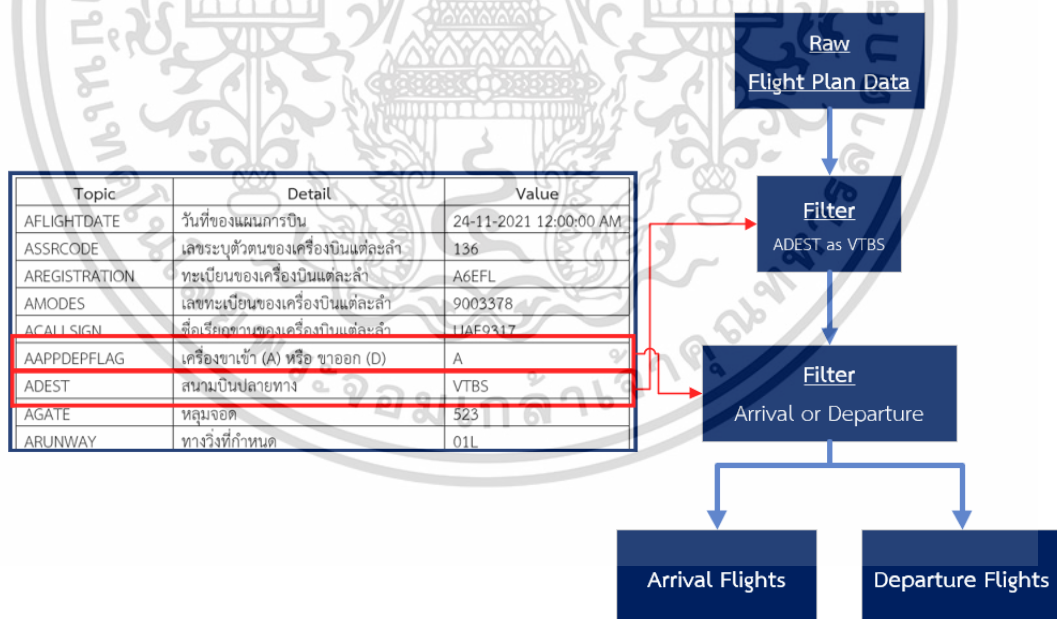
จากรูปที่ 3.2 จะพบว่าการแสดงตำแหน่งของเครื่องบินจะถูกแบ่งออกเป็น 3 สี คือ สีน้ำเงิน แสดงเครื่องบินที่จะเข้ามาลงจอดในสนามบิน สีเหลืองแสดงเครื่องบินที่บินขึ้นจากสนามบิน และสีเทา แสดงเครื่องบินที่ไม่ได้มีแผนการบินขึ้นหรือลงที่สนามบินที่สนใจ ซึ่งการแบ่งประเภทของเครื่องบินที่ขึ้นหรือลง ออกเป็นแต่ละสีบนหน้าจอ ASD นั้น ได้มาจากการควบคุมกันระหว่างข้อมูลของแผนการบินกับข้อมูลตำแหน่งของเครื่องบินในระบบประมวลผลและแสดงออกมาเป็นภาพบนหน้าจอ ASD ให้กับผู้ควบคุมการจราจรทางอากาศเห็น ซึ่งสามารถสรุปหัวข้อของชุดข้อมูลที่ถูกคัดเลือกมาเป็นปัจจัยหลักที่นำมาใช้ในแบบจำลองได้ ดังตารางที่ 3.3

### ตารางที่ 3.3 หัวข้อชุดข้อมูลอินพุตของแบบจำลอง

คุณลักษณะ	ความหมาย
Runway	Assigned landing runway
Arrival Traffic using runway 01R	Number of arrival aircrafts are around 20NM far from aircraft target toward to runway 01R.
Arrival Traffic runway using 01L	Number of arrival aircrafts are around 20NM far from aircraft target toward to runway 01L.
Departure Traffic runway using 01R	Number of departure aircrafts are around 20NM far from aircraft target ran out runway 01R.
Departure Traffic runway using 01L	Number of departure aircrafts are around 20NM far from aircraft target ran out runway 01L.

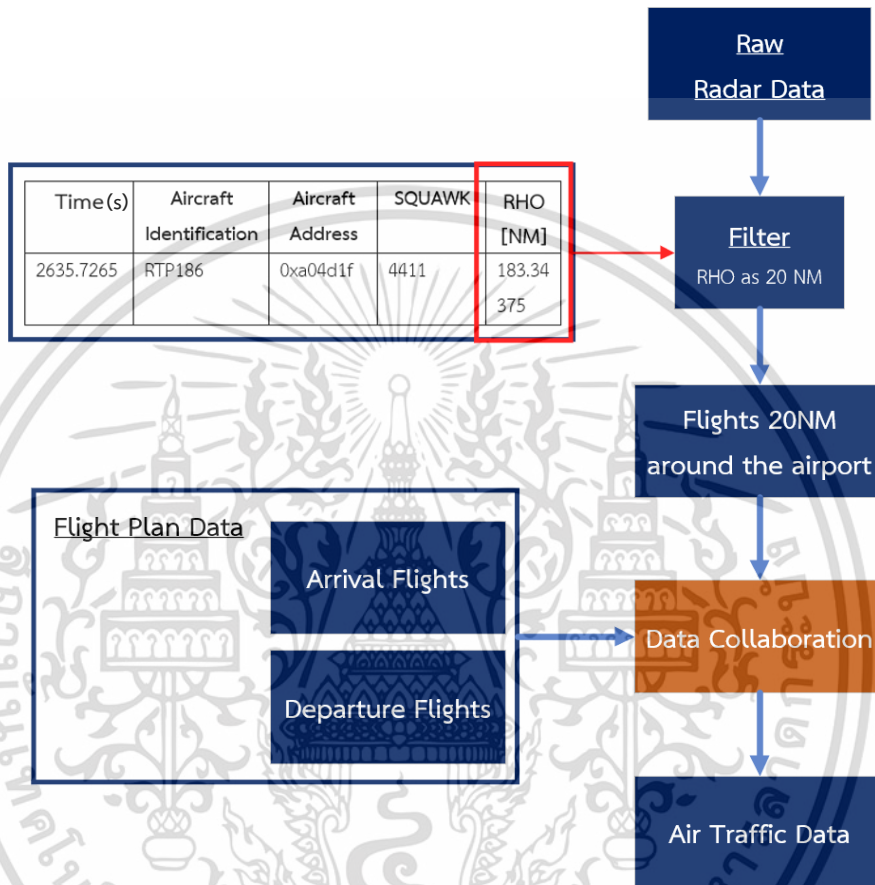
#### 3.2.4 การจัดรูปแบบข้อมูล Dataset ของแบบจำลอง

เมื่อกำหนดหัวข้อของชุดข้อมูลได้แล้วนั้น ในส่วนถัดไปจะต้องนำข้อมูลของเรดาร์และแผนการบินมาจัดรูปแบบ (Data Preprocessing) เพื่อให้ได้ชุดข้อมูล Dataset ตามที่ต้องการ โดยการกรองข้อมูลของเครื่องบินที่มีแผนการบินต้นทางและสนามบินปลายทางเป็นสนามบินสุวรรณภูมิ และแยกออกเป็น เครื่องบินขาเข้าและขาออก ดังรูปที่ 3.3



รูปที่ 3.3 แผนภาพการคัดแยกข้อมูลเครื่องบินขาเข้าและขาออกสนามบินสุวรรณภูมิ

จากนั้นนำข้อมูลแผนการบินมาควบรวมกับข้อมูลเรดาร์เพื่อกรองเอาเครื่องบินที่อยู่ในรัศมี 20 NM รอบสนามบิน โดยการใช้ข้อมูลชื่อเรียกขานของเครื่องบิน (Aircraft Identification) และเลขทะเบียนของเครื่องบินแต่ละลำ (Aircraft Address) ในการจับคู่ข้อมูลระหว่างแผนการบิน และ ตำแหน่งเครื่องบินของข้อมูลเรดาร์ ดังรูปที่ 3.4



รูปที่ 3.4 แผนภาพการควบรวมข้อมูลระหว่างแผนการบินและข้อมูลเรดาร์

เมื่อนำข้อมูลแผนการบินและข้อมูลเรดาร์มารวมเข้ากันแล้วจะได้เป็นชุดข้อมูลของการจราจรทางอากาศที่สนใจบริเวณ 20 NM ของสนามบิน ดังตัวอย่างในตารางที่ 3.4

ตารางที่ 3.4 ตัวอย่างชุดข้อมูลการจราจรทางอากาศรัศมี 20 NM รอบสนามบิน

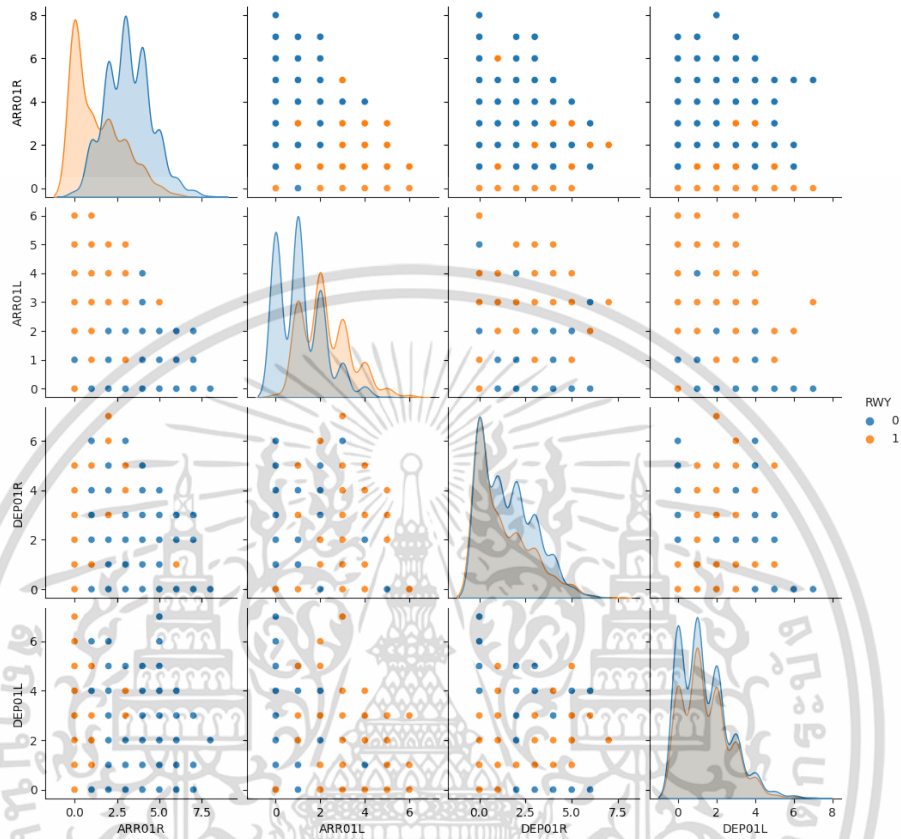
Date and Time	Callsign	Aircraft Address	Traffic Type	Assigned Runway	Distance (NM)
24/11/2021 21:23:58	BOX531	3949968	Arrival	01R	15.4453125
24/11/2021 01:10:55	FIN141	4595532	Arrival	01R	19.203125
24/11/2021 01:36:20	GFA152	8994958	Arrival	01R	12.109375
24/11/2021 00:35:53	UAE9317	9003378	Arrival	01L	17.3828125

หลังจากนั้น จึงทำการหาปริมาณของการจราจรของเครื่องบินขาเข้าและขาออกที่ใช้ทางวิ่งที่ 01R และ 01L เพื่อจัดทำเป็นชุดข้อมูล Dataset สำหรับนำเข้าทดสอบในแบบจำลอง โดยแยกออกเป็น ปริมาณการจราจรของเครื่องบินขาเข้าที่ใช้ทางวิ่ง 01R และ 01L และปริมาณการจราจรของเครื่องบินขาออกที่ใช้ทางวิ่ง 01R และ 01L ดังตารางที่ 3.5 แสดงตัวอย่างชุดข้อมูล Dataset ที่ได้จัดรูปแบบแล้ว ทั้งนี้จะกำหนดให้ทางวิ่ง 01L แทนด้วยค่า 1 และทางวิ่งที่ 01R ด้วยค่า 0 เนื่องจากแบบจำลองจะจำแนกออกเป็น Binary

ตารางที่ 3.5 ตัวอย่างชุดข้อมูล Dataset หลังจากการทำ Data Preprocessing

Callsign	Runway	Arrival traffic		Departure traffic	
		Runway 01R	Runway 01L	Runway 01R	Runway 01L
TVJ139	01R	4	3	2	2
ETH608	01L	3	2	2	2
QTR836	01L	2	2	2	2
AIQ4403	01R	2	2	2	3
OMA2815	01L	3	1	1	3
THD286	01R	3	1	1	3
THD155	01R	4	2	2	2
TVJ221	01R	4	3	2	3
KAL351	01L	4	2	2	3
TFT206	01R	5	2	2	2

หลังจากที่ได้จัดรูปแบบชุดข้อมูลที่จะใช้ในการทดสอบในแบบจำลองแล้ว เพื่อตรวจสอบว่าชุดข้อมูลที่คัดเลือกและจัดรูปแบบมาขึ้นมีความเหมาะสมกับการทำนายผลลัพธ์ของแบบจำลองหรือไม่ โดยทำการสร้างแผนภาพความสัมพันธ์ของชุดข้อมูลเพื่อตรวจสอบ ดังรูปที่ 3.5

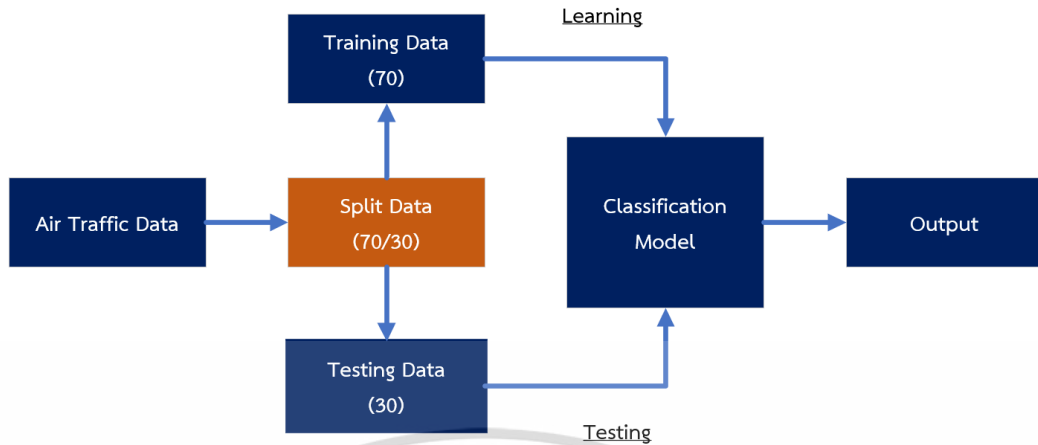


รูปที่ 3.5 ความสัมพันธ์ระหว่างชุดข้อมูลที่ใช้ในการจำแนกกับการกำหนดทางวิ่งของเครื่องบินขาเข้า

จากรูปที่ 3.5 จะพบว่าชุดข้อมูลที่ได้ทำการคัดเลือกมามีความเหมาะสมที่จะนำมาใช้จำแนกการกำหนดทางวิ่งของเครื่องบินขาเข้าได้อย่างเหมาะสม สืบเนื่องจากผลลัพธ์ของทางวิ่ง 0 และ 1 ที่กลุ่มข้อมูลค่อนข้างแยกออกจากกัน

### 3.3 การนำเข้าสู่ชุดข้อมูลเพื่อทดสอบในแบบจำลอง

ในการศึกษาครั้งนี้ เพื่อฝึกฝนและทดสอบแบบจำลองนั้น จะทำการแบ่งชุดข้อมูลออกเป็นข้อมูลสำหรับการฝึกฝน (Training) และข้อมูลสำหรับการทดสอบ (Test) เพื่อให้แบบจำลองแต่ละแบบทำการเรียนรู้จากชุดข้อมูล และทดสอบแบบจำลองด้วยชุดข้อมูลที่เหลือที่แบบจำลองยังไม่เคยเจอ ดังรูปที่ 3.6



รูปที่ 3.6 แผนภาพการป้อนข้อมูลเพื่อเรียนรู้และทดสอบของแบบจำลอง

โดยแบ่งออกเป็น 70 เปอร์เซ็นต์ สำหรับชุดข้อมูล Training และ 30 เปอร์เซ็นต์ สำหรับชุดข้อมูล Test ดังตารางที่ 3.6 สรุปจำนวนชุดข้อมูลการฝึกฝนและการทดสอบ

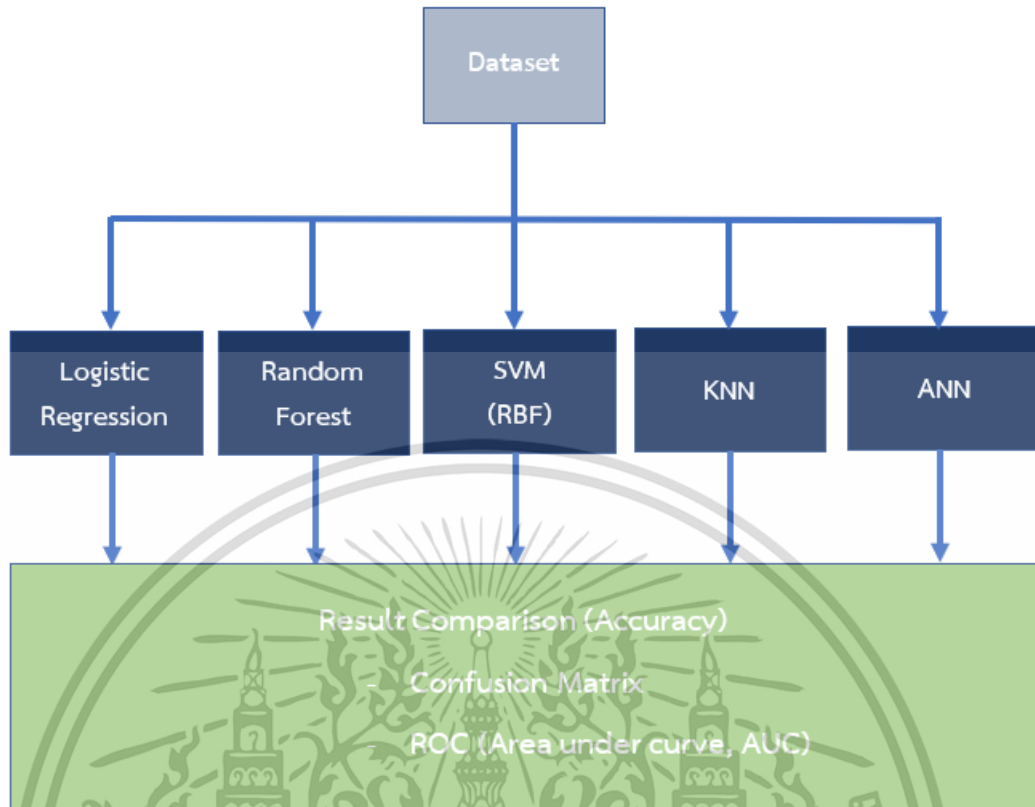
ตารางที่ 3.6 จำนวนชุดข้อมูลการฝึกฝนและการทดสอบ

	Training	Test
X	1035	444
Y	1035	444

กำหนดให้ X คือ ข้อมูลการจราจรเครื่องบินขาเข้าบนทางวิ่ง 01R, เครื่องบินขาเข้าบนทางวิ่ง 01L, เครื่องบินขาออกบนทางวิ่ง 01R และการจราจรเครื่องบินขาออกบนทางวิ่ง 01L ขณะที่ Y คือ ข้อมูลของทางวิ่งที่กำหนด

### 3.4 การเปรียบเทียบประสิทธิภาพของแบบจำลอง

เนื่องจากการจำแนกข้อมูลโดยใช้ Machine Learning มีวิธีการหลากหลายหลากหลาย ขึ้นอยู่กับความเหมาะสมในการนำมาใช้กับข้อมูลที่สนใจ ในการศึกษาครั้งนี้จึงได้เลือกอัลกอริทึมที่แตกต่างกันมาเปรียบเทียบกันเพื่อหาอัลกอริทึมที่เหมาะสมที่สุด ดังรูปที่ 3.7



รูปที่ 3.7 แผนภาพการเปรียบเทียบประสิทธิภาพของแบบจำลอง

โดยประสิทธิภาพของแบบจำลองจะถูกวัดด้วยค่าความแม่นยำของแบบจำลองจาก Confusion Matrix และพื้นที่ใต้เส้นโค้ง ROC (Area Under the Curve, AUC) รวมถึงเปรียบเทียบเวลาที่ใช้ในการประมวลผลเพื่อทำนายผลลัพธ์ของแต่ละแบบจำลอง

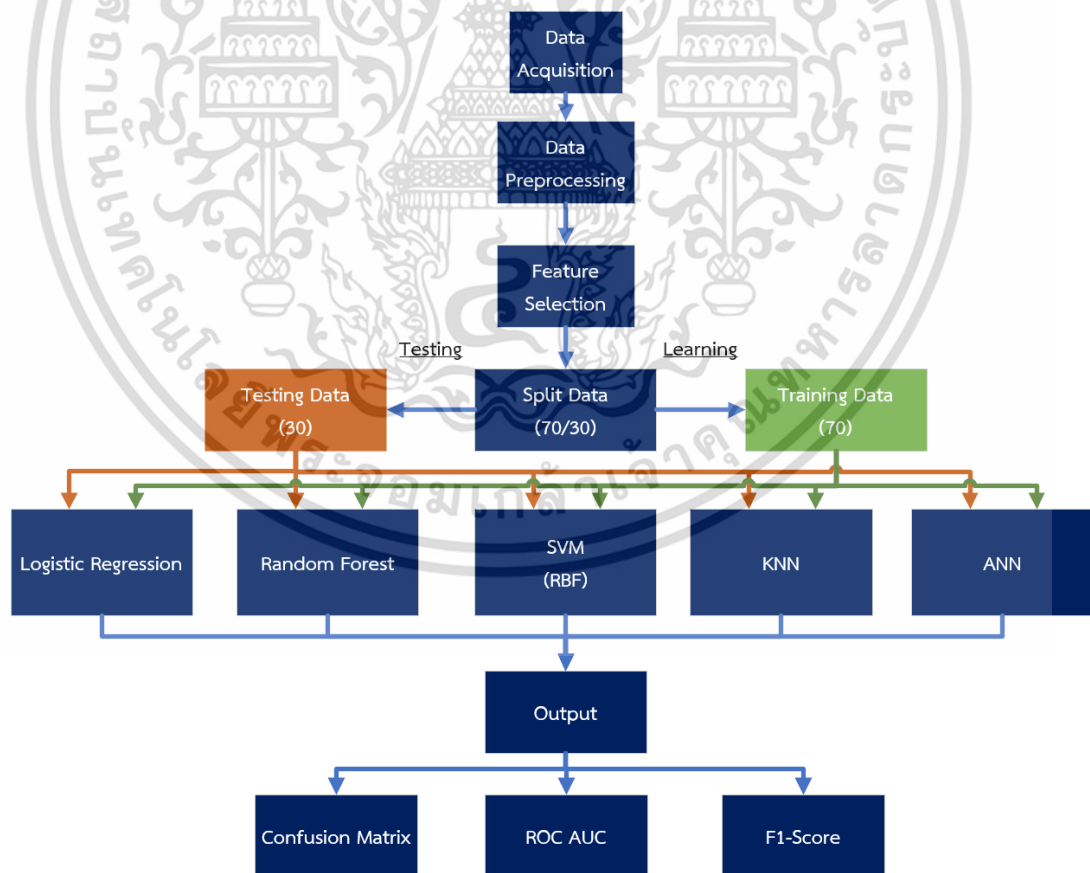
## บทที่ 4

# ผลการดำเนินการ

ในบทนี้จะกล่าวถึงผลลัพธ์ที่ได้จากการทดลอง ซึ่งประกอบไปด้วย ผลลัพธ์ของการทำนายด้วยแบบจำลองที่แตกต่างกัน ซึ่งใช้กลุ่มข้อมูลในการเรียนรู้และทดสอบชุดเดียวกัน พร้อมทั้งแสดงจำนวนผลลัพธ์ที่ทำนายถูกและผิด รวมทั้งเปรียบเทียบค่าความแม่นยำ และเวลาที่ใช้ในการประมวลผลในแต่ละแบบจำลอง

### 4.1 ประสิทธิภาพของแบบจำลอง

จากการนำเข้าสู่ข้อมูลเพื่อทำการเรียนรู้ให้กับแบบจำลองแต่ละแบบ สามารถวัดประสิทธิภาพของแบบจำลองแต่ละแบบของการศึกษานี้ แสดงด้วย Confusion Matrix ค่าความถูกต้อง (Accuracy) ค่าความระลึก (Recall) ค่าเฉลี่ยความถูกต้อง (F1-Score) ค่าความเที่ยงตรง (Precision) พื้นที่ใต้เส้นโค้ง ROC (Area Under the Curve, AUC) และเวลาที่ใช้ประมวลผล ดังรูปที่ 4.1



รูปที่ 4.1 แผนภาพการนำเข้าสู่ข้อมูลและวัดประสิทธิภาพของแบบจำลอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

#### 4.1.1 Confusion Matrix ของแบบจำลอง Logistic Regression

เพื่อเปรียบเทียบข้อมูลที่ได้จากกระบวนการเรียนรู้และทำนายของแบบจำลอง Logistic Regression ซึ่งมีการตั้งค่าพารามิเตอร์ของแบบจำลองดังตารางที่ 4.1

ตารางที่ 4.1 สรุปพารามิเตอร์ที่ใช้สำหรับแบบจำลอง Logistic Regression

Parameters	Value
Penalty	l2
tol	0.0001
C	1
Solver	lbfgs

สามารถแสดงผลที่ได้ด้วย Confusion Matrix ดังรูปที่ 4.2

	01L	01R
01L Assigned	215	43
01R Assigned	45	141

รูปที่ 4.2 Confusion Matrix – Logistic Regression

#### 4.1.2 Confusion Matrix ของแบบจำลอง Artificial Neural Networks

เพื่อเปรียบเทียบข้อมูลที่ได้จากกระบวนการเรียนรู้และทำนายของแบบจำลอง Artificial Neural Networks ซึ่งมีการตั้งค่าพารามิเตอร์ของแบบจำลองดังตารางที่ 4.2

ตารางที่ 4.2 สรุปพารามิเตอร์ที่ใช้สำหรับแบบจำลอง Artificial Neural Networks

Parameters	Value
Layers of model	4
Units of hidden layer	3
Epoch	50
Activation function	ReLU
Optimization function	Adam
Error function	Binary cross-entropy

สามารถแสดงผลที่ได้ด้วย Confusion Matrix ดังรูปที่ 4.3

	01L	01R
01L Assigned	204	54
01R Assigned	45	141

รูปที่ 4.3 Confusion Matrix – ANN

#### 4.1.3 Confusion Matrix ของแบบจำลอง Support Vector Machine - Radial Basis Function

เพื่อเปรียบเทียบข้อมูลที่ได้จากกระบวนการเรียนรู้และทำนายของแบบจำลอง Support Vector Machine - Radial Basis Function ซึ่งมีการตั้งค่าพารามิเตอร์ของแบบจำลองดังตารางที่ 4.3

ตารางที่ 4.3 สรุปพารามิเตอร์ที่ใช้สำหรับแบบจำลอง SVM – RBF Kernal

Parameters	Value
Kernal	Rbf
Gamma	1
C	1

สามารถแสดงผลลัพธ์ที่ได้ด้วย Confusion Matrix ดังรูปที่ 4.4

	01L	01R
01L Assigned	203	55
01R Assigned	56	130

รูปที่ 4.4 Confusion Matrix – SVM (RBF)

#### 4.1.4 Confusion Matrix ของแบบจำลอง Random Forest

เพื่อเปรียบเทียบข้อมูลที่ได้จากกระบวนการเรียนรู้และทำนายของแบบจำลอง Random Forest ซึ่งมีการตั้งค่าพารามิเตอร์ของแบบจำลองดังตารางที่ 4.4

ตารางที่ 4.4 สรุปพารามิเตอร์ที่ใช้สำหรับแบบจำลอง Random Forest

Parameters	Value
N_estimators	100
Criterion	Gini
Min_samples_split	2
Min_samples_leaf	1
Min_weight_fraction_leaf	0.0
Ccp_alpha	0.0

สามารถแสดงผลลัพธ์ที่ได้ด้วย Confusion Matrix ดังรูปที่ 4.5

	01L	01R
01L Assigned	199	59
01R Assigned	57	129

รูปที่ 4.5 Confusion Matrix – Random Forest

#### 4.1.5 Confusion Matrix ของแบบจำลอง K-Nearest Neighbor

เพื่อเปรียบเทียบข้อมูลที่ได้จากกระบวนการเรียนรู้และทำนายของแบบจำลอง K-Nearest Neighbor ซึ่งมีการตั้งค่าพารามิเตอร์ของแบบจำลองดังตารางที่ 4.5

ตารางที่ 4.5 สรุปพารามิเตอร์ที่ใช้สำหรับแบบจำลอง K-Nearest Neighbor

Parameters	Value
N_neighbors	50
Weights	Uniform
Leaf_size	30
Metric	euclidean

สามารถแสดงผลลัพธ์ที่ได้ด้วย Confusion Matrix ดังรูปที่ 4.6

	01L	01R
01L Assigned	227	31
01R Assigned	58	128

รูปที่ 4.6 Confusion Matrix – KNN

จากผลลัพธ์ของ Confusion Matrix ทั้งหมด จะพบว่าจำนวนของผลการทำนายที่ถูกต้องในการกำหนดทางวิ่ง 01R และ 01L ของแบบจำลอง Logistic Regression มีปริมาณความถูกต้องของข้อมูลมากที่สุดเมื่อเทียบกับแบบจำลองอื่น ๆ

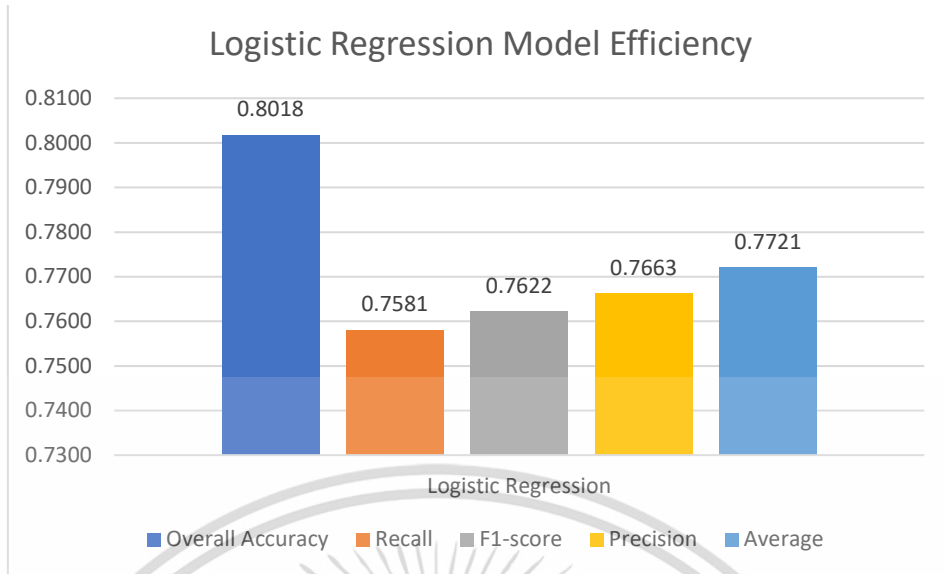
#### 4.1.6 ค่าความถูกต้อง ค่าความระลึก ค่าเฉลี่ยความถูกต้อง ค่าความเที่ยงตรง

ค่าวัดผลประสิทธิภาพของแบบจำลองแต่ละแบบแสดงได้ดังตารางที่ 4.6

ตารางที่ 4.6 ค่าความถูกต้อง ค่าความระลึก ค่าเฉลี่ยความถูกต้อง ค่าความเที่ยงตรง ของแบบจำลอง

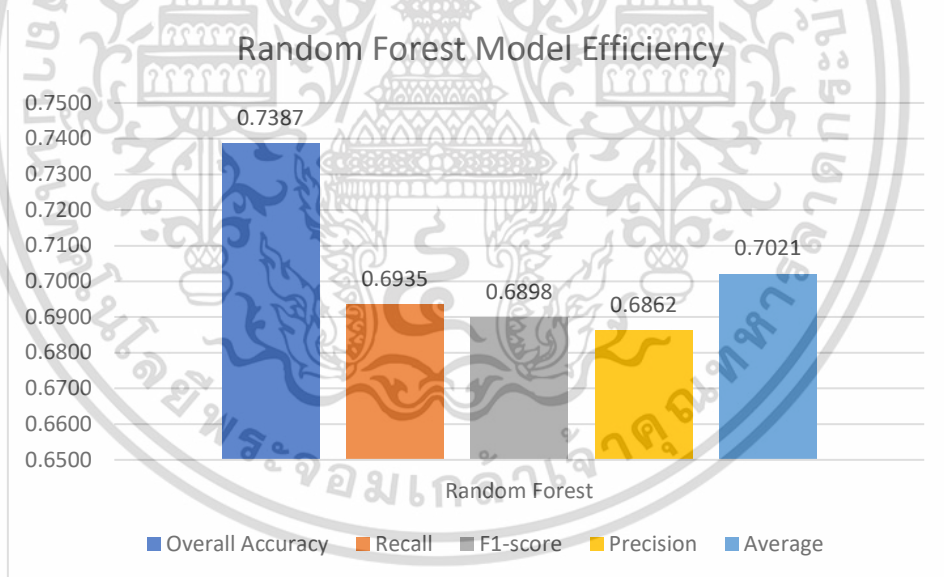
Model	Overall Accuracy	Recall	F1-score	Precision	Average
Logistic Regression	80.18%	75.81%	76.22%	76.63%	77.21%
Random Forest	73.87%	69.35%	68.98%	68.62%	70.21%
SVM (RBF)	75.00%	69.89%	70.08%	70.27%	71.31%
KNN	79.95%	68.82%	74.20%	80.50%	75.87%
ANN	77.70%	75.81%	74.02%	72.31%	74.96%

นำค่าประสิทธิภาพที่คำนวณได้ของแต่ละแบบจำลองมาแสดงเป็นกราฟ ดังรูปที่ 4.7 - 4.11



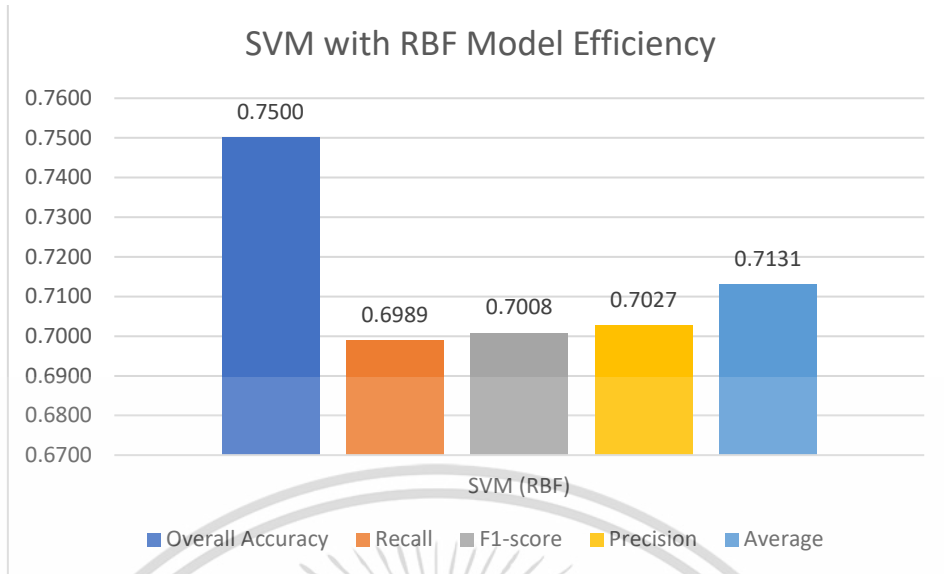
รูปที่ 4.7 ประสิทธิภาพของแบบจำลอง Logistic Regression

จากรูปที่ 4.7 พบว่าแบบจำลอง logistic Regression ให้ผลลัพธ์ค่าความถูกต้องรวมสูงที่สุด ร้อยละ 80.18 ได้ค่าความเที่ยงตรง ค่าเฉลี่ยความถูกต้อง และค่าความระลึกลดลงตามลำดับ โดยได้ค่าประสิทธิภาพโดยเฉลี่ยรวมร้อยละ 77.21



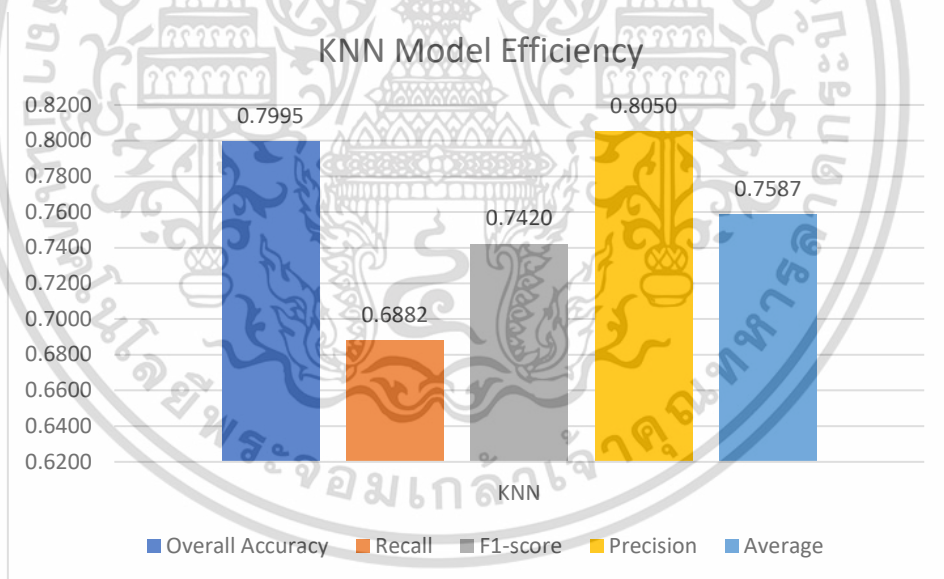
รูปที่ 4.8 ประสิทธิภาพของแบบจำลอง Random Forest

จากรูปที่ 4.8 พบว่าแบบจำลอง Random Forest ให้ผลลัพธ์ค่าความถูกต้องรวมสูงที่สุด ร้อยละ 73.87 ได้ค่าความระลึกล ค่าเฉลี่ยความถูกต้อง และค่าความเที่ยงตรง ลดลงตามลำดับ โดยได้ค่าประสิทธิภาพโดยเฉลี่ยรวมร้อยละ 70.21



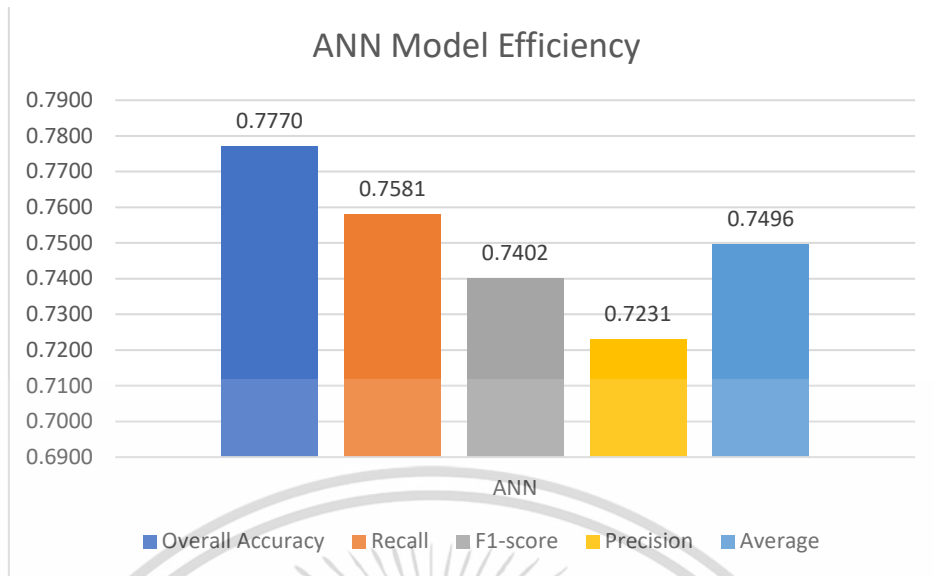
รูปที่ 4.9 ประสิทธิภาพของแบบจำลอง SVM with RBF

จากรูปที่ 4.9 พบว่าแบบจำลอง SVM with RBF ให้ผลลัพธ์ค่าความถูกต้องรวมสูงที่สุดร้อยละ 75.00 ได้ค่าความเที่ยงตรง ค่าเฉลี่ยความถูกต้อง และค่าความระลึกลดลงตามลำดับ โดยได้ค่าประสิทธิภาพโดยเฉลี่ยรวมร้อยละ 71.31



รูปที่ 4.10 ประสิทธิภาพของแบบจำลอง KNN

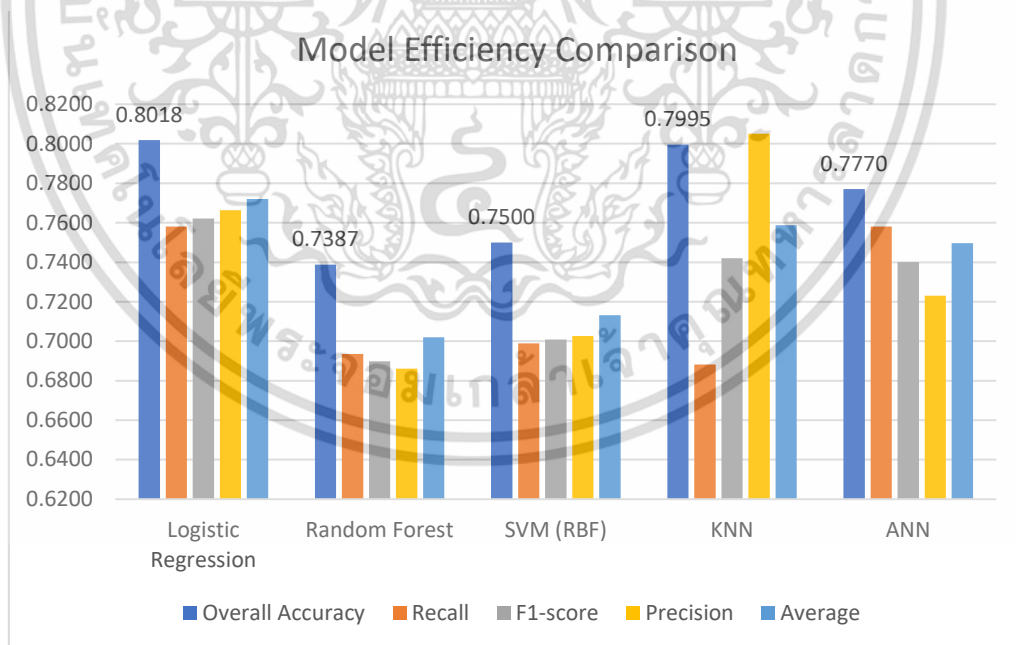
จากรูปที่ 4.10 พบว่าแบบจำลอง KNN ให้ผลลัพธ์ค่าความเที่ยงตรงสูงที่สุดร้อยละ 80.05 ได้ค่าความถูกต้องรวม ค่าเฉลี่ยความถูกต้อง และค่าความระลึกลดลงตามลำดับ โดยได้ค่าประสิทธิภาพโดยเฉลี่ยรวมร้อยละ 75.87



รูปที่ 4.11 ประสิทธิภาพของแบบจำลอง ANN

จากรูปที่ 4.11 พบว่าแบบจำลอง ANN ให้ผลลัพธ์ค่าความถูกต้องรวมสูงสุด ร้อยละ 77.70 ได้ค่าความระลึก ค่าเฉลี่ยความถูกต้อง และค่าความเที่ยงตรง ลดลงตามลำดับ โดยได้ค่าประสิทธิภาพโดยเฉลี่ยรวมร้อยละ 74.96

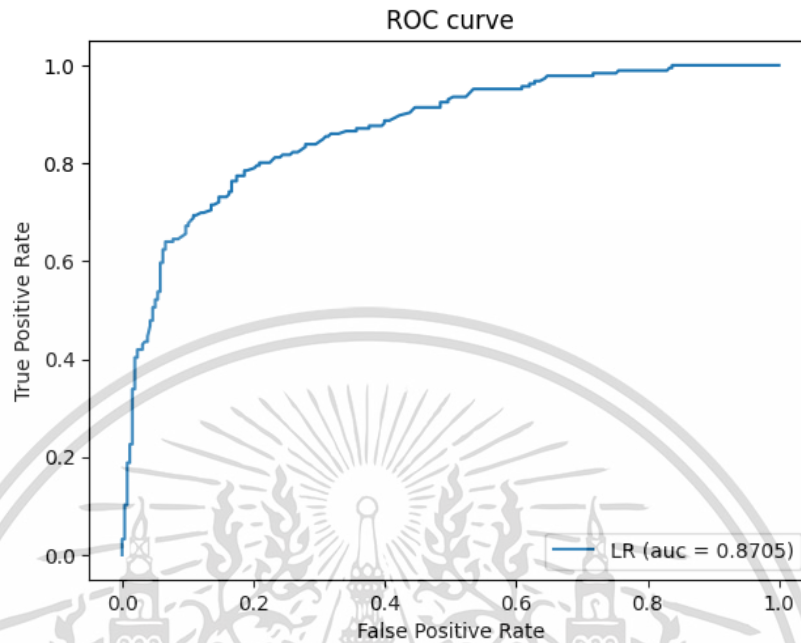
เมื่อเปรียบเทียบผลประสิทธิภาพค่าความถูกต้องรวมของแบบจำลองทั้งหมด จะพบว่าแบบจำลอง Logistic Regression มีค่าความถูกต้องรวมสูงสุดร้อยละ 80.18 ดังรูปที่ 4.12



รูปที่ 4.12 เปรียบเทียบประสิทธิภาพการทำนายของแบบจำลอง

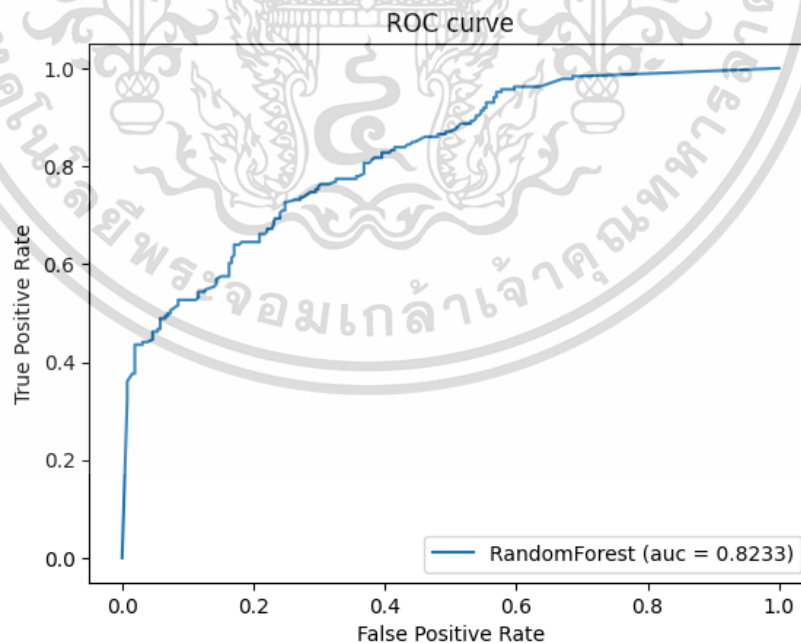
#### 4.1.7 พื้นที่ใต้เส้นโค้ง ROC (Area Under Curve, AUC)

ค่าประสิทธิภาพที่วัดได้จากพื้นที่ใต้เส้นโค้ง ROC ของแต่ละแบบจำลองแสดงดังรูปที่ 4.13 – รูปที่ 4.17



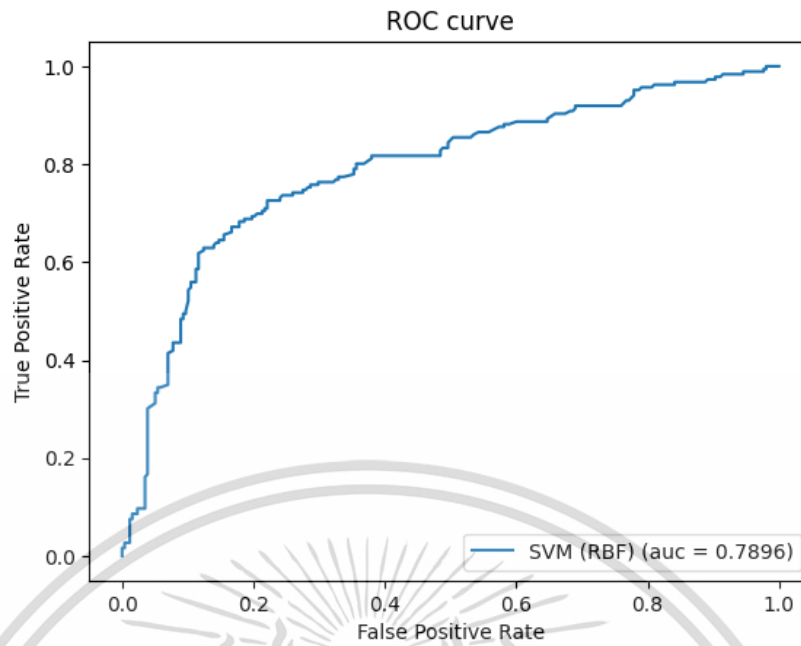
รูปที่ 4.13 Area under curve - Logistic Regression

จากรูปที่ 4.13 พบว่าแบบจำลอง Logistic Regression มีพื้นที่ใต้เส้นโค้ง ROC ร้อยละ 87.05



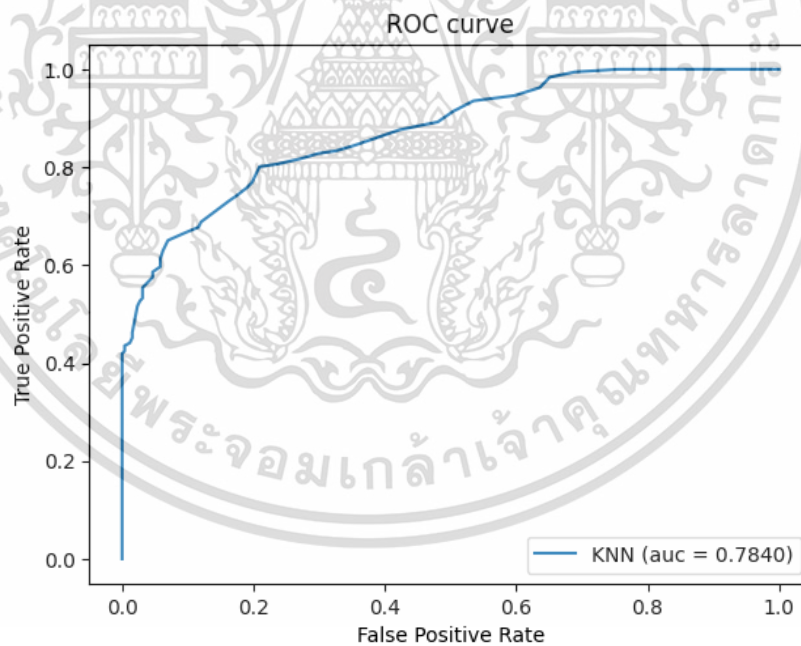
รูปที่ 4.14 Area under curve – Random Forest

จากรูปที่ 4.14 พบว่าแบบจำลอง Random Forest มีพื้นที่ใต้เส้นโค้ง ROC ร้อยละ 82.33



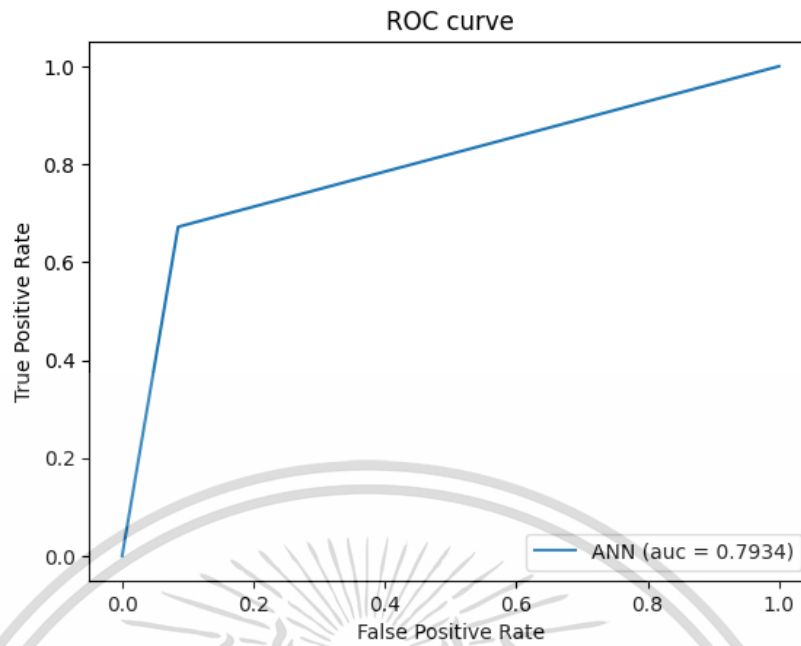
รูปที่ 4.15 Area under curve – SVM (RBF)

จากรูปที่ 4.15 พบว่าแบบจำลอง SVM (RBF) มีพื้นที่ใต้เส้นโค้ง ROC ร้อยละ 78.96



รูปที่ 4.16 Area under curve – KNN

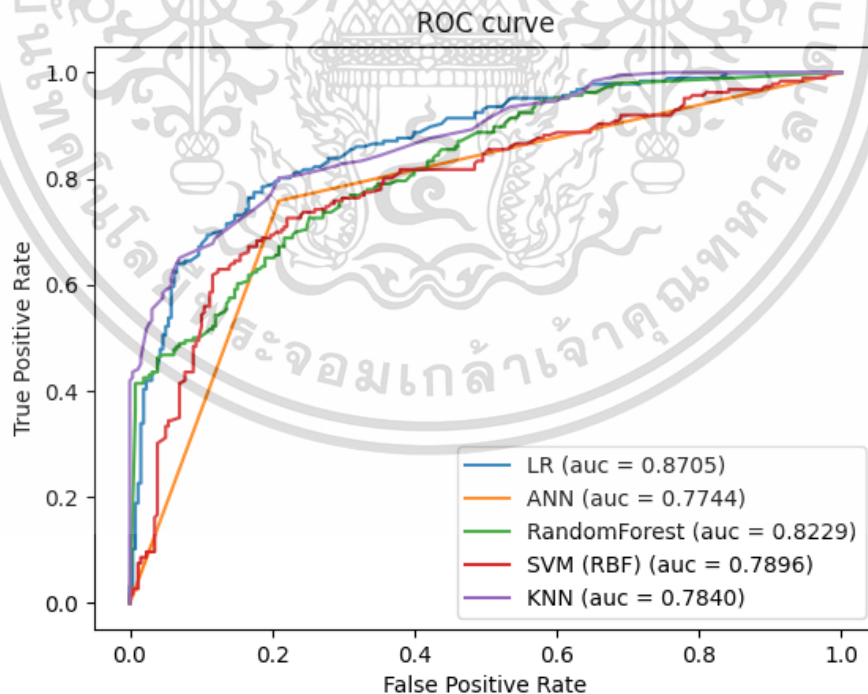
จากรูปที่ 4.16 พบว่าแบบจำลอง KNN มีพื้นที่ใต้เส้นโค้ง ROC ร้อยละ 78.40



รูปที่ 4.17 Area under curve – ANN

จากรูปที่ 4.17 พบว่าแบบจำลอง ANN มีพื้นที่ใต้เส้นโค้ง ROC ร้อยละ 79.34

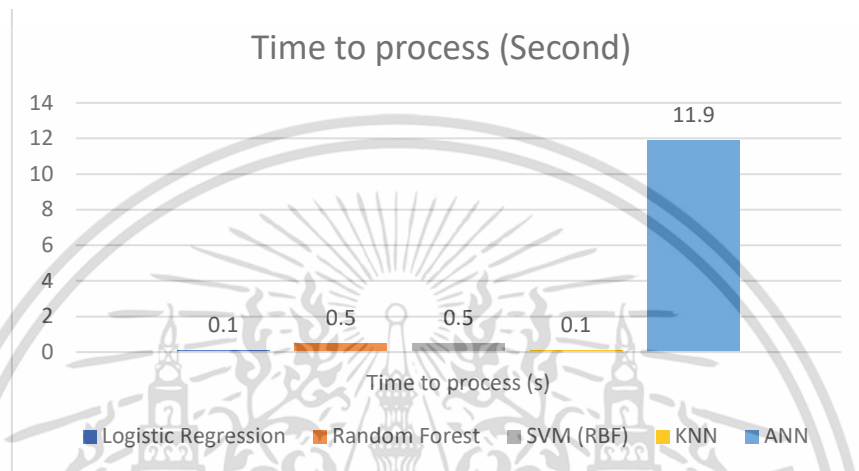
เมื่อนำผลพื้นที่ใต้เส้น ROC ของแต่ละแบบจำลองมาเปรียบเทียบกันจะพบว่าแบบจำลอง Logistic Regression มีค่ามากที่สุดร้อยละ 87.05 ดังรูปที่ 4.18



รูปที่ 4.18 Area under curve – All Models comparison

#### 4.1.8 เวลาที่ใช้ประมวลผล

นอกเหนือจากการวัดค่าประสิทธิภาพของแบบจำลองด้วยค่าความแม่นยำแล้ว ความรวดเร็วในการทำนายถือเป็นอีกปัจจัยหนึ่งที่สำคัญ และเป็นตัวบ่งชี้ถึงประสิทธิภาพของแบบจำลองอีกด้วย โดยเครื่องคอมพิวเตอร์แม่ข่ายที่ใช้ทดสอบแบบจำลองมีคุณลักษณะ คือ หน่วยประมวลผลกลาง (CPU) แบบ 8 แกนหลัก (8 core) และมีหน่วยความจำหลัก (RAM) ชนิด ECC DDR3 จำนวน 16 GB. ซึ่งเวลาที่ใช้ทำนายของแต่ละแบบจำลองในหน่วยวินาที แสดงดังรูปที่ 4.19



รูปที่ 4.19 เวลาที่ใช้ในการทำนายของแต่ละแบบจำลอง

จากรูปที่ 4.19 จะพบว่าแบบจำลอง Logistic Regression และ KNN ใช้เวลาในการประมวลผลและทำนายได้ภายในเวลา 0.1 วินาที ในขณะที่แบบจำลอง ANN ใช้เวลาในการประมวลผลและการทำนายมากที่สุด 11.9 วินาที

#### 4.2 เปรียบเทียบประสิทธิภาพแบบจำลองกับงานวิจัยที่เกี่ยวข้อง

ในการศึกษานี้ได้นำผลการทดสอบของแบบจำลอง Machine Learning ประเภทอื่น ๆ มาเปรียบเทียบกับการศึกษานี้ โดยแสดงให้เห็นถึงผลลัพธ์ของแบบจำลองที่มีรูปแบบของชุดข้อมูลที่แตกต่างกันไปในแต่ละหัวข้อ ที่เกี่ยวข้องกับการให้บริการจราจรทางอากาศ ดังตารางที่ 4.7

ตารางที่ 4.7 ผลลัพธ์ของแบบจำลองในงานวิจัยอื่นที่เกี่ยวข้องกับการจัดการจราจรทางอากาศ

Author	Dataset	Method	ACC	Recall	Specificity	Precision	F1-Score	AUC
F. Herrema et al. [21]	54,679 Arrival flights	Gradient Boosting	81.00%	98.30%	-	97.80%	98.00%	-
S. Mokhtarimousavi et al. [27]	21,891 Arrival flights	ABC-SVM	96.10%	96.40%	89.00%	-	97.40%	93.70%
A. Midtjord et al. [26]	13 Norwegian Airports. Weather data, Runway reports and Flight data	XGBoost	-	91.80%	85.20%	-	-	94.80%
G. Priyanka [19]	US domestic airline traffic data from 2013 to 2016	K-Nearest-Neighbor Classifier (KNN)	90.00%	95.00%	-	90.00%	-	-
Y. Guleria et al. [23]	612 Conflicting Flight Plans	Multiple Random Forest models	95.10%	-	-	-	-	-
K. Eisinga [17]	Similar flights per day	SVM with linear kernel	55.00%	55.00%	-	49.00%	47.00%	-
	(2008 and 2016) Weather conditions Maintenance planning. (2010 to 2016)	Logistic Regression	56.00%	56.00%	-	51.00%	49.00%	-
Proposed Method	1,479 Flight Data records	Logistic Regression	80.18%	75.81%	83.33%	76.63%	76.22%	87.05%

## บทที่ 5

### ปัญหาและแนวทางการแก้ไข

#### 5.1 สรุปและวิจารณ์ผลการทดลอง

วิทยานิพนธ์ฉบับนี้ มุ่งเน้นไปที่การศึกษาปัจจัยที่เกี่ยวข้องกับการตัดสินใจของผู้ควบคุมการจราจรทางอากาศ ในการกำหนดทางวิ่งให้กับเครื่องบินขาเข้าของสนามบินสุวรรณภูมิ และสร้างแบบจำลองสำหรับการกำหนดทางวิ่งให้กับเครื่องบินขาเข้า เพื่อลดภาระงาน ความเครียดที่เกิดขึ้น และเวลาในการตัดสินใจของผู้ควบคุมการจราจรทางอากาศ และเพิ่มความสามารถการรองรับปริมาณเที่ยวบินขาเข้า ณ สนามบินสุวรรณภูมิได้มากยิ่งขึ้น

ผู้วิจัยได้นำข้อมูลการจราจรของเครื่องบินบนทางวิ่ง และบริเวณน่านฟ้ารอบสนามบินสุวรรณภูมิเป็นปัจจัยหลักในการจัดทำชุดข้อมูลเพื่อทดสอบในแบบจำลอง ซึ่งในการวิจัยนี้นำเสนอแบบจำลอง Machine Learning ประเภทที่ใช้ในการจำแนก (Classification) จำนวน 5 แบบ ได้แก่ แบบจำลองการวิเคราะห์การถดถอยโลจิสติก (Logistic Regression Analysis) แบบจำลองการสุ่มป่าไม้ (Random Forest) แบบจำลอง K-Nearest Neighbors (KNN) แบบจำลองซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) และแบบจำลองโครงข่ายประสาทเทียม (Artificial Neuron Network) มาทดสอบหาแบบจำลองที่สามารถกำหนดทางวิ่งให้กับเครื่องบินได้อย่างแม่นยำและรวดเร็วที่สุด โดยใช้การเปรียบเทียบค่าประสิทธิภาพของแบบจำลอง ประกอบด้วย ค่าความถูกต้อง ค่าความระลึก ค่าเฉลี่ยความถูกต้อง และค่าความเที่ยงตรง

จากการทดสอบแบบจำลองกำหนดทางวิ่งของเครื่องบินขาเข้า พบว่าผลทำนายที่ได้จากแบบจำลอง Logistic Regression มีค่าความถูกต้องรวมสูงที่สุด เมื่อเทียบกับค่าความถูกต้องรวมของแบบจำลองอื่น โดยได้ค่าความถูกต้องรวมของการทำนายทางวิ่งร้อยละ 80.18 และได้ค่าประสิทธิภาพที่วัดได้จากพื้นที่ใต้เส้นโค้ง ROC (Area Under the Curve, AUC) สูงที่สุด ร้อยละ 87.05 โดยใช้เวลาในการประมวลผลข้อมูลเพียง 0.1 วินาที ในการทดสอบแบบจำลองทุกแบบในการศึกษานี้ พบว่าข้อมูลอินพุตของแบบจำลองที่ได้เลือกใช้เป็นตัวแปรที่สำคัญในการทำนายทางวิ่ง ซึ่งสอดคล้องกับกลยุทธ์ปัจจุบันที่ดำเนินการโดยผู้ควบคุมการจราจรทางอากาศของสนามบิน นอกจากนี้ยังเห็นได้ชัดว่ากระบวนการเตรียมข้อมูลเพื่อหาข้อมูลที่เกี่ยวข้องเป็นขั้นตอนที่สำคัญในกระบวนการจัดชุดข้อมูลสำหรับแบบจำลองเพื่อให้ได้ผลลัพธ์ที่แม่นยำ ทั้งนี้ผลการทำนายที่ได้จากแบบจำลองเป็นเพียงแค่การแนะนำให้ผู้ใช้งาน แต่การกำหนดทางวิ่งจะคงยังถูกตัดสินใจโดยผู้ควบคุมการจราจรทางอากาศโดยตรง การกำหนดทางวิ่งอาจเปลี่ยนไปจากสถานการณ์ปกติได้ ขึ้นอยู่กับสถานการณ์ปัจจุบันหรือสถานการณ์ที่ไม่สามารถควบคุมได้ ซึ่งปัจจัยดังกล่าวเป็นปัจจัยที่แบบจำลองไม่สามารถควบคุมได้ ส่งผลให้ผลการจำลองยังคงมีค่าความผิดพลาด แต่ยังอยู่ในช่วงที่ยอมรับได้ ดังจุดประสงค์ของการศึกษา

ปัญหาที่พบจากการวิเคราะห์ปัจจัยเพื่อนำมาเป็นข้อมูลเรียนรู้ให้กับแบบจำลองต่าง ๆ พบว่า ข้อมูลที่ใช้เรียนรู้จะเป็นข้อมูลที่ได้จากการเก็บข้อมูล ณ เวลานั้น ๆ เท่านั้น โดยการจราจรของเครื่องบินบริเวณรอบสนามบินจริง จะมีการเปลี่ยนแปลงรูปแบบได้ตลอดเวลา ขึ้นอยู่กับสถานการณ์สภาพอากาศในช่วงเวลาต่าง ๆ รวมถึงรูปแบบการตัดสินใจที่แตกต่างกันระหว่างผู้ควบคุมการจราจรทางอากาศที่มีการผลัดเปลี่ยนกันไปตามแต่ละเวร ซึ่งแบบจำลองที่ได้รับการเรียนรู้จากข้อมูลชุดเดิมนั้น จะส่งผลให้การทำนายในกรณีที่สถานการณ์มีการเปลี่ยนแปลงไปเกิดความผิดพลาดมากขึ้นไปกว่าเดิม

## 5.2 แนวทางการแก้ไข

ปัญหาที่เกิดขึ้นสามารถแก้ไขได้โดยการเก็บค่าการตัดสินใจจากเจ้าหน้าที่ควบคุมการจราจรทางอากาศในทุก ๆ ช่วงต้นของการเข้ากะที่มีการเปลี่ยนเวร รวมถึงในระหว่างการเข้าเวร เพื่อนำเข้ามาเป็นข้อมูลเรียนรู้ให้กับแบบจำลองอย่างต่อเนื่อง เช่น ทุก 30 นาที เพื่อให้แบบจำลองสามารถทำนายได้แม่นยำแม้ในกรณีที่มีรูปแบบของการตัดสินใจที่เปลี่ยนแปลงไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## เอกสารอ้างอิง

- [1] I. Kanjanasurat, B. Purahong, C. Pintavirooj, N. Satayarak and C. Benjangkprasert, "Blood vessel extraction and optic disk localization for diabetic retinopathy," 10th International Conference on Biomedical Engineering and Technology (ICBET), September 2020, pp. 112-116.
- [2] I. Kanjanasurat and N. Satayarak, "Optic disk and fovea localization by using the direction of blood vessels and morphology operation," 11th International Conference on Biomedical Engineering and Technology (ICBET), March 2021, pp. 14-19.
- [3] K. Tenghongsakul, I. Kanjanasurat, B. Purahong and A. Lasakul, "Retinal blood vessel extraction by using pre-processing and internet mode," 24<sup>th</sup> International Computer Science and Engineering Conference (ICSEC), December 2020, pp. 1-5.
- [4] K. Srijakkot, I. Kanjanasurat, N. Wiriyakrieng, M. Lartwatechakul and C. Benjangkprasert, "Intruder detection by using faster R-CNN in power substation," International Conference on Computing and Information Technology (IC2IT), March 2020, pp.159-167.
- [5] V. Chutchavong, T. Pumee, S. Thongkrait and T. Anuwongpinit, "An improved performance simulated annealing based on evolution strategies for single objective optimization problems," 6th International Conference on Engineering, Applied Sciences and Technology (ICEAST), July 2020, pp. 1-4.
- [6] P. Sarnmeta, W. Inthakon, D. Chumpungam and S. Suantai, "On convergence and complexity analysis of an accelerated forward-backward algorithm with linesearch technique for convex minimization problems and applications to data prediction and classification," J Inequal Appl, August 2021, pp. 141.
- [7] THE CIVIL AVIATION AUTHORITY OF THAILAND (CAAT). "AIRAC AIP - THAILAND." [Online]. Available : <https://aip.caat.or.th/2019-11-07-AIRAC/html/index-en-GB.html>. 2019.
- [8] THE CIVIL AVIATION AUTHORITY OF THAILAND (CAAT). Manual of standards, Air Traffic Management Services: Air Traffic Services (CAAT-ANS-MOSATS), Dec 2018.
- [9] ปรีชาพล ชูศรี, จีรณา น้อยมณี และเกษม พันธุ์สิน. เทคโนโลยีปัญญาประดิษฐ์ สำหรับการบริหารงานและการบริการภาครัฐ. กรุงเทพมหานคร : บริษัท ส.พิจิตรการพิมพ์ จำกัด. พ.ศ. 2562.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- [10] SAS Institute Inc. “What is AI? and how is it important?” [Online]. Available : [https://www.sas.com/th\\_th/insights/analytics/what-is-artificial-intelligence.html](https://www.sas.com/th_th/insights/analytics/what-is-artificial-intelligence.html). 2021.
- [11] S. Shalev-Shwartz, S. Ben-David, Understanding machine learning: from theory to algorithms. New York : Cambridge University Press. 2014.
- [12] ยุทธ ไกยวรรณ. “หลักการและการใช้การวิเคราะห์การถดถอยโลจิสติกสำหรับการวิจัย” วารสารวิจัยมหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย. ปีที่ 4 , ฉบับที่ 1, มกราคม-มิถุนายน พ.ศ. 2555. หน้า 1-12.
- [13] ไกรศักดิ์ เกษร. วิทยาศาสตร์ข้อมูล (Data Science). พิษณุโลก : ภาควิชาวิทยาการคอมพิวเตอร์ และเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยนเรศวร. พ.ศ. 2564.
- [14] A. Tharwat, “Classification assessment methods,” Applied computing and informatics, vol. 17, no. 1, August 2018, pp. 168-192.
- [15] J. Fan, S. Upadhye and A. Worster, “Understanding receiver operating characteristic (ROC) curves,” Canadian Journal of Emergency Medicine, vol. 8, no. 1, January 2006, pp. 19-20.
- [16] F. Salfner, M. Lenk and M. Malek, “A survey of online failure prediction methods,” [Online]. Available : [https://www.researchgate.net/figure/Sample-ROCplots-A-perfect-failure-predictor-shows-a-truepositive-rate-of-one-and-a\\_fig4\\_220565792](https://www.researchgate.net/figure/Sample-ROCplots-A-perfect-failure-predictor-shows-a-truepositive-rate-of-one-and-a_fig4_220565792), 2020.
- [17] K. Eisinga, “Predicting runway allocation with support vector machine and logistic regression,” Master of Science in Communication and Information Sciences, Master Track Data Science: Business and Governance, at the School of Humanities of Tilburg University, July 2016.
- [18] Y. Nakamura, R. Mori and H. Aoyama, “Modeling of runway assignment strategy by human controllers using machine learning,” IEEE/AIAA 36th Digital Avionics Systems Conference (DASC), September 2017, pp. 1-7.
- [19] G. Priyanka, “Prediction of airline delays using K-nearest neighbor algorithm,” International Journal of Emerging Technology and Innovative Engineering, vol. 4, no. 5, August 2018, pp. 87-90.
- [20] X. Zou, Y. Hu, Z. Tian and K. Shen, “Logistic regression model optimization and case analysis,” IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT), October 2019, pp. 135-139
- [21] F. Herremaa, R. Currana, S. Hartjesa, M. Ellejmib, S. Bancroftb and M. Schultzc, “A machine learning model to predict runway exit at Vienna airport,” Transportation Research Part E: Logistics and Transportation Review, vol. 131, November 2019, pp. 329-342.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- [22] K. Dutta, S. A., A. Victor, A. Nathu, M. Habib and D. Parashar, “Kannada alphabets recognition using decision tree and random forest models,” 3rd International Conference on Intelligent Sustainable Systems (ICISS), December 2020, pp. 534-541.
- [23] Y. Guleria, P. Tran, D. Pham, N. Durand and S. Alam, “A machine learning framework for predicting ATC conflict resolution strategies for conformal automation,” 11th SESAR Innovation Days (SIDs), 2021.
- [24] K. Bhupal Reddy, M. Vishnu Priya and Murugesan, “Power theft detection using novel linear SVM algorithm and compared with convolutional SVM algorithm for accuracy,” 14th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS), November 2022, pp. 1-9.
- [25] T. Deva Prakash and V. Nagaraju, “Comparative analysis using K-nearest neighbor with artificial neural network to improve accuracy for predicting road accidents,” International Conference on Edge Computing and Applications (ICECAA), October 2022, pp. 1092 - 1096.
- [26] A. Midtjord, R. Bin and A. Huseby, “A decision support system for safer airplane landings: Predicting runway conditions using XGBoost and explainable AI,” Cold Regions Science and Technology, vol. 199(103556), July 2022, pp. 1-15.
- [27] S. Mokhtarimousavi and A. Mehrabi, “Flight delay causality: Machine learning technique in conjunction with random parameter statistical analysis,” International Journal of Transportation Science and Technology, vol. 12, March 2023, pp. 230-244.



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## ผลงานวิจัยที่ได้รับการตีพิมพ์

- [1] I. Kanjanasurat, W. Jungsuwadee, A. Lasakul and C. Benjangkprasert, "Comparison of logistic regression and random forest algorithms for airport's runway assignment," *Journal of Physics: Conference series*, vol. 2497 (012016), 2023, pp.1-6.
- [2] I. Kanjanasurat, W. Jungsuwadee, B. Purahong and C. Benjangkprasert, "Landing runway assignment by airport traffic using machine learning," 3<sup>rd</sup> International Conference on Industrial Control Network and System Engineering Research (ICNSER 2022), May 2022, pp. 40-45.



## ประวัติผู้เขียน

ชื่อ-นามสกุล นายวศรุต จิ่งสุวดี  
วัน เดือน ปีเกิด 17 สิงหาคม 2537  
สถานที่เกิด กรุงเทพมหานคร  
ที่อยู่ 55/202 หมู่บ้านศุภาลัยไพรด์บางนา-ลาดกระบัง ถนน วัดศรีวารีน้อย  
ตำบล ศีรษะจรเข้ชั้น้อย อำเภอ บางเสาธง จังหวัด สมุทรปราการ 10540  
โทร. 087-469-7849  
ประวัติการศึกษา 2559 วิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมโทรคมนาคม  
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ประสบการณ์การทำงาน  
พ.ศ. 2561-ปัจจุบัน ตำแหน่งวิศวกรระบบจราจรทางอากาศ  
บริษัท วิทยุการบินแห่งประเทศไทย จำกัด



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และ 60 วินาทีถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้