

บทที่ 4

ผลการดำเนินงาน

จากการวิเคราะห์ข้อมูลเที่ยวบินที่นำมาใช้ในการศึกษาครั้งนี้ โดยใช้ชุดข้อมูล Flight Delays and Cancellations จาก Kaggle ซึ่งครอบคลุมเที่ยวบินภายในประเทศของสหรัฐอเมริกา ในปี พ.ศ. 2558 เนื่องจากข้อมูลเที่ยวบินมีปริมาณจำนวนมากและหลากหลาย จึงได้มีการสุ่มเลือกข้อมูลบางส่วนมาใช้ในการวิเคราะห์ เพื่อให้ได้ชุดข้อมูลที่มีความเหมาะสมต่อการสร้างโมเดล ทั้งนี้มีวัตถุประสงค์เพื่อวิเคราะห์ปัจจัยที่ส่งผลกระทบต่อความล่าช้าในการมาถึงของเที่ยวบิน (Arrival Delay) รวมถึงสร้างแบบจำลองเพื่อทำนายแนวโน้มของการเกิดความล่าช้า

ผลการวิเคราะห์ข้อมูลที่ได้ถูกนำเสนอในรูปแบบของตาราง กราฟ และค่าตัวชี้วัดทางสถิติ โดยผู้จัดทำโครงการได้นำเทคโนโลยีการทำเหมืองข้อมูล (Data Mining) และอัลกอริทึมการเรียนรู้ของเครื่อง (Machine Learning) มาประยุกต์ใช้ ทำให้สามารถอธิบายความสัมพันธ์ของปัจจัยที่เกี่ยวข้อง ตลอดจนสร้างโมเดลพยากรณ์ที่มีประสิทธิภาพและสามารถนำไปประยุกต์ใช้ได้จริง ผลการดำเนินงานทั้งหมดแสดงไว้ในบทนี้

4.1 ผลการดำเนินงาน

ผู้จัดทำได้ทำการจัดการและประมวลผลข้อมูลเที่ยวบินด้วยกระบวนการ CRISP-DM โดยเริ่มจากการทำความเข้าใจข้อมูล (Data Understanding) การเตรียมข้อมูล (Data Preparation) และการสร้างแบบจำลอง (Modeling) ในการวิเคราะห์เชิงจำแนกประเภท (Classification) ได้เลือกใช้เทคนิค Decision Tree และ Random Forest ซึ่งเป็นเทคนิคการทำเหมืองข้อมูลที่ช่วยแสดงกฎการตัดสินใจในรูปแบบโครงสร้างต้นไม้ (Decision Rules) เพื่อให้เข้าใจความสัมพันธ์ระหว่างปัจจัยต่าง ๆ ที่ส่งผลกระทบต่อความล่าช้า โดยกำหนดตัวแปรเป้าหมายคือ ARRIVAL_DELAY_CLASS แบ่งออกเป็น 3 กลุ่ม ได้แก่

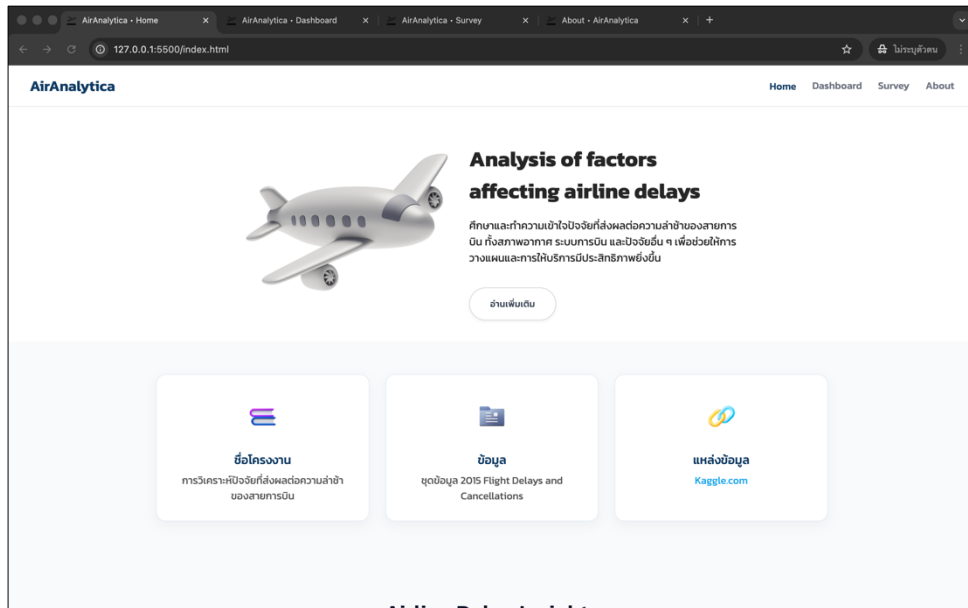
- ON TIME หมายถึง เที่ยวบินที่มาถึงตรงเวลา (ค่าความล่าช้าน้อยกว่า 5 นาที)
- SHORT DELAY หมายถึง เที่ยวบินที่ล่าช้าเล็กน้อย (5-14 นาที)
- DELAY หมายถึง เที่ยวบินที่ล่าช้าสูง (ตั้งแต่ 15 นาทีขึ้นไป)

ผลการวิเคราะห์จากโมเดล Decision Tree พบว่ามีค่าความแม่นยำ (Accuracy) เท่ากับ 90.01% $\pm$ 0.27% ซึ่งถือว่าอยู่ในระดับสูง โดยสามารถจำแนกเที่ยวบินตรงเวลาและเที่ยวบินล่าช้าได้อย่างถูกต้อง ขณะที่โมเดล Random Forest มีค่าความแม่นยำเท่ากับ 89.30% $\pm$ 0.03% แสดงถึงประสิทธิภาพใกล้เคียงกัน ทั้งนี้จากการวิเคราะห์กฎจำแนกพบว่า DEPARTURE_DELAY, AIRLINE_DELAY และ LATE_AIRCRAFT_DELAY เป็นตัวแปรสำคัญที่ส่งผลโดยตรงต่อความล่าช้า

ส่วนการวิเคราะห์เชิงพยากรณ์ (Regression) ใช้เทคนิค Multiple Linear Regression (MLR) เพื่อศึกษาความสัมพันธ์ระหว่างตัวแปรตาม ARRIVAL_DELAY กับตัวแปรอิสระที่ผ่านการคัดเลือกด้วยเกณฑ์ค่าสหสัมพันธ์ (Correlation $\geq$ 0.4) รวมทั้งหมด 14 ตัว เช่น SCHEDULED_TIME, AIRTIME, DISTANCE, DEPARTURE_TIME, ARRIVAL_TIME, AIRLINE_DELAY, และ LATE_AIRCRAFT_DELAY เป็นต้น ผลการประเมินแบบจำลองพบว่า

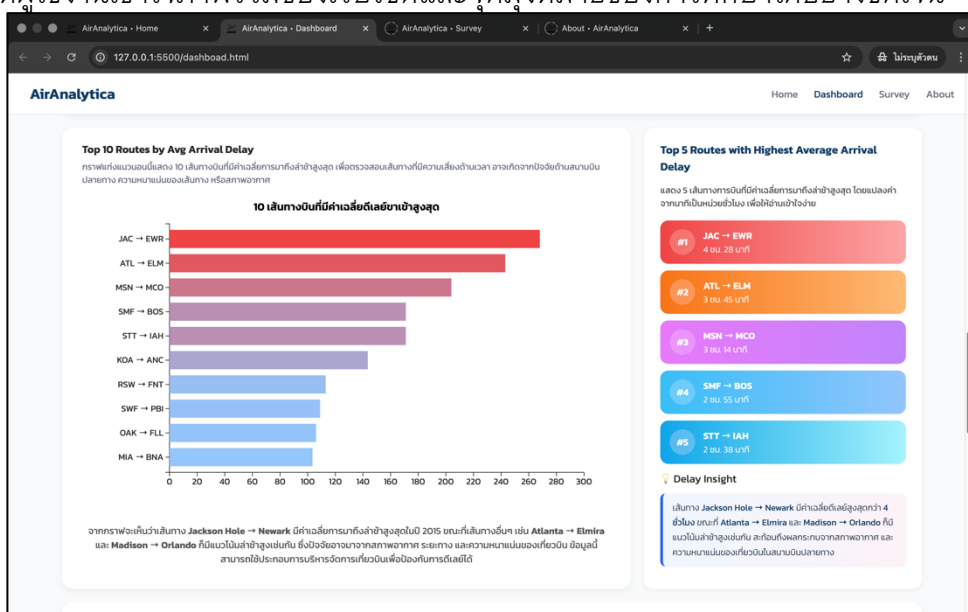
โมเดลมีค่า $R^2 = 0.873$, $MAE = 7.2$ นาที และ $RMSE = 9.875$ นาที แสดงให้เห็นว่าแบบจำลองสามารถพยากรณ์ค่าความล่าช้าได้อย่างแม่นยำ

ข้อมูลผลลัพธ์ทั้งหมดถูกนำเสนอในรูปแบบของตาราง กราฟ แดชบอร์ดเชิงโต้ตอบ (Interactive Dashboard) และแสดงผลลัพธ์แบบสอบถามจากโมเดล Classification และ Regression ผ่านเว็บไซต์ “AirAnalytica” ซึ่งพัฒนาโดยใช้ภาษา HTML, CSS, และ JavaScript เพื่อให้ผู้ใช้งานสามารถเข้าถึงข้อมูล วิเคราะห์แนวโน้ม และเข้าใจปัจจัยที่ส่งผลต่อความล่าช้าได้อย่างชัดเจนและเป็นระบบ



ภาพที่ 4.1 หน้า Home ของเว็บไซต์ AirAnalytica

จากภาพที่ 4.1 หน้านี้เป็นหน้าแรกของเว็บไซต์ AirAnalytica ซึ่งแสดงชื่อโครงการ และแหล่งข้อมูลที่ใช้ในการวิเคราะห์ (Flight Delays and Cancellations Dataset จาก Kaggle) เพื่อให้ผู้ใช้งานเข้าใจภาพรวมของเว็บไซต์และจุดมุ่งหมายของการศึกษาได้อย่างชัดเจน



ภาพที่ 4.2 หน้า Dashboard แสดงผลสถิติเที่ยวบินและกราฟ

จากภาพที่ 4.2 แสดงหน้าแดชบอร์ดหลัก (Main Dashboard) สำหรับการนำเสนอผลการวิเคราะห์ข้อมูลเที่ยวบินในรูปแบบเชิงสถิติและภาพกราฟิก โดยได้ออกแบบให้เป็นแดชบอร์ดแบบโต้ตอบ (Interactive Dashboard) ที่รวบรวมผลการวิเคราะห์ในหลายมิติ แดชบอร์ดดังกล่าวช่วยให้ผู้ใช้งานสามารถมองเห็นภาพรวมของข้อมูลได้อย่างเป็นระบบ เข้าใจแนวโน้ม และปัจจัยที่อาจส่งผลกระทบต่อความล่าช้าของเที่ยวบินในแต่ละช่วงเวลา

The screenshot shows the 'AirAnalytics' dashboard with a title 'แบบฟอร์มทดสอบโมเดลการประเมินความล่าช้าเที่ยวบิน' (Flight Delay Prediction Model Test Form). It contains two main sections:

- ส่วนที่ 1 : การจำแนกสถานะเที่ยวบิน (Decision Tree)**
 - ปัจจัยของสายการบิน (Airline Factors):**
 - สายการบินที่คุณใช้ประจำหรือไม่? (Do you use the airline regularly?) with buttons: ไม่ล่าช้า (No delay), ล่าช้าบางครั้ง (Sometimes delay), ล่าช้าบ่อย (Often delay).
 - สภาพอากาศในวันเดินทางเป็นอย่างไร? (What is the weather on the travel day?) with buttons: อากาศปกติ (Normal weather), ฝนตกหรือพายุ (Rain or storm).
 - ปัจจัยภายนอก (External Factors):**
 - เที่ยวบินของคุณโดยทั่วไปออกตรงเวลาหรือไม่? (Does your flight generally depart on time?) with buttons: ตรงเวลา (On time), ล่าช้าเล็กน้อย (เล็กน้อยกว่า 30 นาที) (Slightly late (less than 30 minutes)), ล่าช้ามาก (มากกว่า 30 นาที) (More than 30 minutes).
 - เที่ยวบินก่อนหน้าถึงตรงเวลาหรือไม่? (Did the previous flight arrive on time?) with buttons: ตรงเวลา (On time), ล่าช้า (Late).
- ส่วนที่ 2 : การพยากรณ์เวลาความล่าช้า (Multiple Linear Regression)**
 - เวลาออกตามกำหนด (Scheduled departure time)
 - Airtime (นาที) (Airtime (minutes))

ภาพที่ 4.3 หน้าแบบสอบถาม (Survey) สำหรับการเก็บข้อมูลประกอบการวิเคราะห์ จากภาพที่ 4.3 หน้านี้เป็นหน้าแบบสอบถามจำลองที่จัดทำขึ้นเพื่อเก็บข้อมูลจากผู้ใช้งานเกี่ยวกับปัจจัยที่อาจมีผลต่อความล่าช้า เช่น สายการบิน ช่วงเวลาเดินทาง หรือสภาพอากาศ ทั้งนี้ข้อมูลดังกล่าวถูกใช้เพื่อสาธิตการทำงานของระบบในเชิงโครงงานเท่านั้น ไม่ได้ใช้เพื่อการเก็บข้อมูลจริง

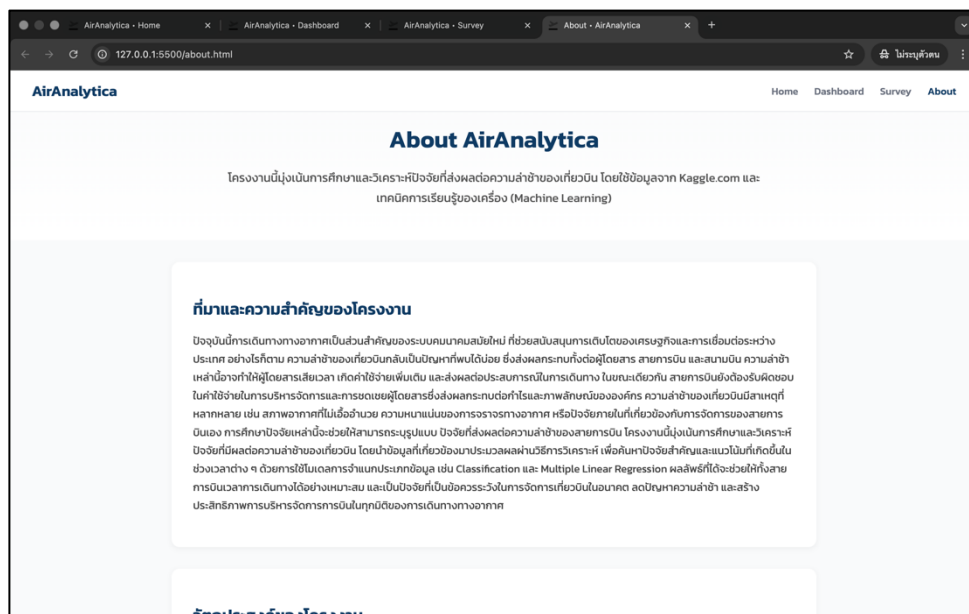
The screenshot shows the 'AirAnalytics' survey form with a modal window titled 'ผลการประเมินเที่ยวบิน' (Flight Delay Prediction Results). The modal displays two results:

- Decision Tree DELAY:** เที่ยวบินของคุณมีแนวโน้มล่าช้ามาก (Your flight has a high tendency to be late).
- Regression:** ล่าช้าเล็กน้อย = 16.9 นาที (Slightly late = 16.9 minutes). คาดว่าล่าช้าเล็กน้อย (Expected to be slightly late).

Below the results, there is a note: 'หมายเหตุ โมเดลทั้งสองอาจให้เกณฑ์ต่างกันในการประเมิน' (Note: Both models may give different criteria for evaluation). At the bottom of the modal, there is a button 'กดลบในแดชบอร์ดอีกครั้ง' (Click delete on dashboard again).

ภาพที่ 4.4 หน้าแสดงผลการประเมินจากแบบสอบถาม โดยสรุปผลตามข้อมูลที่กรอก

จากภาพที่ 4.4 หน้านี้แสดงผลการประเมินข้อมูลจากแบบสอบถาม โดยนำข้อมูลที่ผู้ใช้กรอกมาประมวลผลและแสดงในรูปแบบของกราฟ ตาราง หรือสรุปผล เพื่อให้เห็นแนวโน้มของความล่าช้าในปัจจัยต่าง ๆ เช่น ช่วงเวลาที่มีโอกาสล่าช้าสูง หรือประเภทของความล่าช้าที่พบได้บ่อยที่สุด



ภาพที่ 4.5 หน้าเกี่ยวกับโครงการ (About) ที่อธิบายที่มา ความสำคัญ และวัตถุประสงค์ของโครงการ AirAnalytica

จากภาพที่ 4.5 หน้านี้เป็นหน้าที่รวบรวมข้อมูลของผู้จัดทำ ที่มา วัตถุประสงค์ และเครื่องมือที่ใช้ในการพัฒนาเว็บไซต์ รวมถึงเทคโนโลยีที่ใช้ในการวิเคราะห์ข้อมูล เพื่อให้ผู้ใช้งานทราบข้อมูลพื้นฐานของโครงการและแนวทางการดำเนินงานโดยรวมของ AirAnalytica

4.2 การอภิปรายผล

4.2.1 การศึกษาทำความเข้าใจธุรกิจ (Business Understanding) ผลการศึกษาพบว่า ปัญหาความล่าช้าของเที่ยวบินเป็นปัญหาสำคัญที่ส่งผลกระทบต่อสายการบิน สนามบิน และผู้โดยสาร ซึ่งปัจจัยที่ทำให้เกิดความล่าช้าสามารถมาจากหลายสาเหตุ เช่น การจัดการของสายการบิน สภาพอากาศ หรือระบบการควบคุมการบิน โดยเฉพาะในเที่ยวบินภายในประเทศของสหรัฐอเมริกา ที่มีปริมาณเที่ยวบินจำนวนมากในแต่ละวัน การทำความเข้าใจเชิงลึกเกี่ยวกับปัจจัยที่ส่งผลต่อการล่าช้าจึงมีความสำคัญต่อการวางแผนและบริหารจัดการตารางเที่ยวบินให้มีประสิทธิภาพสูงสุด

4.2.2 การทำความเข้าใจข้อมูล (Data Understanding) ผู้วิเคราะห์ข้อมูลได้รวบรวมข้อมูลจากชุดข้อมูล Flight Delays and Cancellations (2015) ซึ่งประกอบด้วยข้อมูลจำนวนกว่า 5,819,079 ระเบียน และ 31 แอตทริบิวต์ เช่น เวลาการออกเดินทาง (DEPARTURE_TIME), เวลามาถึง (ARRIVAL_TIME), ระยะทาง (DISTANCE), ความล่าช้าในขั้นตอนต่าง ๆ (AIR_SYSTEM_DELAY, AIRLINE_DELAY, WEATHER_DELAY, LATE_AIRCRAFT_DELAY) เป็นต้น จากนั้นทำการสุ่มข้อมูลบางส่วน (Stratified Sampling 5%) เพื่อให้ขนาดข้อมูลเหมาะสมต่อการ

ประมวลผลและลดความซ้ำซ้อน โดยยังคงรักษาสัดส่วนของคลาสข้อมูลในแต่ละกลุ่มให้เท่าเดิม

4.2.3 การเตรียมข้อมูล (Data Preparation) ผู้วิเคราะห์ข้อมูลได้ดำเนินการทำความสะอาดข้อมูล (Data Cleaning) เพื่อลบประเด็นที่ซ้ำกันและแทนค่าที่ขาดหาย (Missing Values) ด้วยค่าเฉลี่ยหรือศูนย์ตามลักษณะของตัวแปร พร้อมทั้งสร้างแอตทริบิวต์ใหม่ (Generate Attribute) สำหรับการจำแนกกลุ่มเที่ยวบิน เช่น ตัวแปร ARRIVAL_DELAY_CLASS ซึ่งแบ่งสถานะเที่ยวบินออกเป็น 3 กลุ่ม ได้แก่ ON TIME, SHORT DELAY และ DELAY การเตรียมข้อมูลดังกล่าวช่วยเพิ่มความถูกต้องและความพร้อมในการสร้างแบบจำลอง

4.2.4 การสร้างแบบจำลอง (Modeling) ผู้วิเคราะห์ข้อมูลได้ทำการวิเคราะห์ข้อมูลเที่ยวบินเพื่อศึกษาปัจจัยที่ส่งผลต่อความล่าช้าในการมาถึงของเที่ยวบิน โดยใช้เทคนิคการทำเหมืองข้อมูล (Data Mining) เพื่อสร้างแบบจำลองการจำแนกและแบบจำลองการพยากรณ์ค่าความล่าช้า โดยเลือกใช้เทคนิค Decision Tree, Random Forest และ Multiple Linear Regression ซึ่งเป็นเทคนิคที่วิเคราะห์ข้อมูลเชิงคาดการณ์ (Predictive Analytics) โดยชุดข้อมูลที่นำมาใช้ในการสร้างแบบจำลองได้รับการสุ่มแบบชั้นเชิง (Stratified Sampling) เพื่อให้กลุ่มข้อมูลแต่ละประเภทมีสัดส่วนสมดุลกัน และได้ทำการคัดเลือกตัวแปรอิสระ (Independent Variables) ที่มีค่าสัมประสิทธิ์สหสัมพันธ์กับตัวแปรตาม (ARRIVAL_DELAY) ตั้งแต่ 0.4 ขึ้นไป

4.2.5 การวัดประสิทธิภาพของโมเดล (Evaluation) การประเมินประสิทธิภาพของแบบจำลองได้ดำเนินการด้วยวิธี 10-fold Cross Validation เพื่อให้ได้ค่าประเมินที่มีความแม่นยำและเสถียร โดยใช้ตัวชี้วัดหลัก ได้แก่ ค่า Accuracy, Precision, Recall และ F1-score ซึ่งผลลัพธ์จากการทดสอบพบว่า

4.2.5.1 โมเดล Decision Tree ให้ความแม่นยำสูงกว่าเล็กน้อย และมีความสามารถในการตีความได้ง่ายกว่าด้วยการแสดงผลในรูปแบบกฎการตัดสินใจ (Decision Rules) ซึ่งช่วยให้เข้าใจปัจจัยที่มีผลต่อความล่าช้าได้อย่างชัดเจน

โมเดลการจำแนกประเภท (Classification) การสร้างโมเดลการจำแนกประเภท ได้เลือกใช้เทคนิค Decision Tree และ Random Forest โดยตัวแปรเป้าหมายคือ ARRIVAL_DELAY_CLASS ซึ่งแบ่งออกเป็น 3 กลุ่ม ได้แก่

- ON TIME: เที่ยวบินที่มีค่า ARRIVAL_DELAY < 5 นาที
- SHORT DELAY: เที่ยวบินที่มีค่า ARRIVAL_DELAY ระหว่าง 5-14 นาที
- DELAY: เที่ยวบินที่มีค่า ARRIVAL_DELAY ≥ 15 นาที

โดยใช้ตัวแปรที่มีค่าสหสัมพันธ์กับ ARRIVAL_DELAY ตั้งแต่ 0.4 ขึ้นไป ได้แก่ DEPARTURE_DELAY, LATE_AIRCRAFT_DELAY, AIRLINE_DELAY และ AIR_SYSTEM_DELAY ผลการประเมินประสิทธิภาพของโมเดล Decision Tree

- Accuracy = 90.01% ± 0.27%
- Precision: ON TIME = 88.37%, DELAY = 99.45%, SHORT DELAY = 66.36%
- Recall: ON TIME = 99.13%, DELAY = 99.47%, SHORT DELAY = 11.44%

ผลการประเมินประสิทธิภาพของโมเดล Random Forest

- Accuracy = 89.30% ± 0.03%
- Precision: ON TIME = 86.93%, DELAY = 100.00%, SHORT DELAY = 0%
- Recall: ON TIME = 100.00%, DELAY = 98.78%, SHORT DELAY = 0%

ตารางที่ 4.1 ผลการวัดประสิทธิภาพของโมเดล Classification

โมเดล (Models)	Accuracy	Precision	Recall	F1-score
Decision Tree	90.01%	88.37%	99.13%	88.04%
Random Forest	89.30%	86.93%	100.00%	92.95%

4.2.5.2 ผลการวิเคราะห์ด้วยโมเดลการถดถอย (Regression) ในการวิเคราะห์ด้วยโมเดลการถดถอย ผู้จัดทำได้เลือกใช้เทคนิค Multiple Linear Regression โดยใช้ ARRIVAL_DELAY เป็นตัวแปรตาม และเลือกตัวแปรอิสระที่มีค่าสหสัมพันธ์ตั้งแต่ 0.4 ขึ้นไป จำนวน 14 ตัว

ผลการประเมินโมเดล Multiple Linear Regression พบว่า

- Root Mean Squared Error (RMSE) = 9.875 ± 0.592
- Mean Absolute Error (MAE) = 7.2
- $R^2 = 0.873$

ตารางที่ 4.2 ผลการวัดประสิทธิภาพของโมเดล Regression

โมเดล (Models)	RMSE	MAE	R^2
Multiple Linear Regression	9.875	7.2	0.873

4.2.5.3 การเปรียบเทียบประสิทธิภาพของทั้งสองโมเดล พบว่า Decision Tree มีความเหมาะสมกว่าสำหรับชุดข้อมูลนี้ เนื่องจากให้ค่าความแม่นยำสูงกว่า (90.01% เทียบกับ 89.30%) และสามารถตีความได้ง่ายผ่านโครงสร้างกฎที่ชัดเจน ขณะที่ Random Forest แม้จะมีความยืดหยุ่นสูงและทนทานต่อ Noise ในข้อมูล แต่ไม่สามารถจำแนกกลุ่ม SHORT DELAY ได้เลย ซึ่งอาจเกิดจากจำนวนข้อมูลในกลุ่มนี้มีสัดส่วนต่ำกว่าร้อยละ 2 ของทั้งหมด

ผลการจำแนกดังกล่าวสอดคล้องกับวัตถุประสงค์ของโครงการที่ต้องการระบุปัจจัยหลักที่ส่งผลต่อความล่าช้าของเที่ยวบิน โดยพบว่าตัวแปรที่มีอิทธิพลมากที่สุดคือ DEPARTURE_DELAY, รองลงมาคือ AIRLINE_DELAY และ LATE_AIRCRAFT_DELAY ซึ่งเป็นปัจจัยที่เกิดจากการดำเนินงานของสายการบินโดยตรง

จากการประเมินเชิงสถิติ พบว่าค่าความคลาดเคลื่อน (Error Rate) ของทั้งสองโมเดลอยู่ในระดับต่ำกว่า 10% แสดงให้เห็นว่าผลลัพธ์ที่ได้มีความน่าเชื่อถือสูง และโมเดลสามารถนำไปประยุกต์ใช้ในการคาดการณ์สถานะเที่ยวบินจริงในอนาคตได้อย่างมีประสิทธิภาพ โดยเฉพาะการนำผลการจำแนกจาก Decision Tree ไปต่อยอดในระบบแดชบอร์ดวิเคราะห์เที่ยวบิน เพื่อช่วยให้สายการบินหรือหน่วยงานควบคุมการบินสามารถตัดสินใจได้อย่างทันท่วงทีเมื่อตรวจพบปัจจัยเสี่ยงของความล่าช้า

4.2.5.4 กฎการจำแนกที่ได้จากโมเดล Decision Treeจากโครงสร้างต้นไม้ตัดสินใจ (Decision Tree Structure) ที่ได้จากการวิเคราะห์ พบว่าโมเดลสามารถสกัด กฎการจำแนก (Decision Rules) ออกมาได้ทั้งหมด 7 กฎหลัก ดังนี้

- ก. IF AIRLINE_DELAY > 0.5 THEN Result = DELAY
หมายความว่า หากเกิดความล่าช้าที่เกี่ยวข้องกับสายการบินมากกว่า 0.5 นาที เที่ยวบินจะถูกจัดอยู่ในกลุ่ม “ล่าช้า (DELAY)”
- ข. IF AIRLINE_DELAY ≤ 0.5 AND AIR_SYSTEM_DELAY > 0.5 THEN Result = DELAY หมายความว่า หากสายการบินไม่ล่าช้ามาก แต่ระบบควบคุมการบินมีความล่าช้ามากกว่า 0.5 นาที เที่ยวบินจะถูกจัดอยู่ในกลุ่ม “ล่าช้า”
- ค. IF AIRLINE_DELAY ≤ 0.5 AND AIR_SYSTEM_DELAY ≤ 0.5 AND LATE_AIRCRAFT_DELAY > 0.5 THEN Result = DELAY หมายความว่า หากไม่มีความล่าช้าจากสายการบินและระบบการบิน แต่เครื่องบินเที่ยวก่อนหน้ามาถึงล่าช้า เที่ยวบินนั้นจะมีแนวโน้มล่าช้าเช่นกัน
- ง. IF LATE_AIRCRAFT_DELAY ≤ 0.5 AND DEPARTURE_DELAY > 54.5 THEN Result = DELAY
หมายความว่า หากไม่มีความล่าช้าจากเครื่องบินเที่ยวก่อนหน้า แต่เที่ยวบินออกช้ากว่า 54.5 นาที จะถูกจัดอยู่ในกลุ่ม “ล่าช้า”
- จ. IF LATE_AIRCRAFT_DELAY ≤ 0.5 AND DEPARTURE_DELAY ≤ 19.5 THEN Result = ON TIME
หมายความว่า หากเที่ยวบินออกช้าหรือเร็วไม่เกิน 19.5 นาที จะถือว่า “ตรงเวลา (ON TIME)”
- ฉ. IF LATE_AIRCRAFT_DELAY ≤ 0.5 AND DEPARTURE_DELAY BETWEEN 19.5–49.5 THEN Result = SHORT DELAY หมายความว่า หากเที่ยวบินออกช้าในช่วง 20–50 นาที จะถูกจัดอยู่ในกลุ่ม “ล่าช้าระดับสั้น (SHORT DELAY)”
- ช. IF LATE_AIRCRAFT_DELAY ≤ 0.5 AND DEPARTURE_DELAY > 49.5 THEN Result = DELAY
หมายความว่า หากเที่ยวบินออกช้ามากกว่า 50 นาทีขึ้นไป จะถูกจัดให้อยู่ในกลุ่ม “ล่าช้า (DELAY)”

จากกฎข้างต้นจะเห็นได้ว่า ตัวแปร DEPARTURE_DELAY ปรากฏอยู่ในทุกกฎสำคัญของโมเดล ซึ่งสะท้อนให้เห็นถึงความสัมพันธ์โดยตรงระหว่างเวลาการออกเดินทางและความล่าช้าในการถึงปลายทางของเที่ยวบิน โดยสามารถสรุปได้ว่า “ความล่าช้าในการออกเดินทาง” เป็นปัจจัยสำคัญที่สุดที่ส่งผลต่อการเกิดความล่าช้าในภาพรวมของเที่ยวบิน

แบบจำลอง Multiple Linear Regression (MLR) ไม่ได้ถูกนำมาใช้เพื่อเปรียบเทียบประสิทธิภาพพร้อมกับโมเดลการจำแนกประเภท เนื่องจากมีจุดประสงค์ในการวิเคราะห์ที่

แตกต่างกัน โดยโมเดล MLR ถูกออกแบบมาเพื่อ พยากรณ์ค่าความล่าช้าเชิงตัวเลข (เช่น ค่าจำนวนนาฬิกาของการมาถึงล่าช้า) และถูกนำไปประยุกต์ใช้ในส่วนของ เว็บไซต์ AirAnalytica เพื่อแสดงผลการคำนวณในหน้าทำนายความล่าช้า (Prediction Page) โดยตรง ไม่ได้อยู่ในขอบเขตของการประเมินเปรียบเทียบความแม่นยำของโมเดลจำแนก แต่ใช้เพื่อเสริมความสมบูรณ์ในการนำเสนอข้อมูลเชิงปริมาณแก่ผู้ใช้งานเว็บไซต์

4.2.6 การนำโมเดลไปใช้งานจริง (Deployment) ขั้นตอนสุดท้ายของกระบวนการวิเคราะห์ข้อมูลตามกรอบแนวคิด CRISP-DM ซึ่งมีวัตถุประสงค์เพื่อนำผลลัพธ์จากแบบจำลองที่พัฒนาแล้วมาใช้งานในรูปแบบที่สามารถเข้าถึงและสื่อสารกับผู้ใช้งานได้อย่างมีประสิทธิภาพในการศึกษาครั้งนี้ ผู้วิเคราะห์ข้อมูลได้นำผลลัพธ์จากการสร้างแบบจำลองทั้ง Decision Tree, Random Forest และ Multiple Linear Regression มาประยุกต์ใช้ผ่านระบบเว็บไซต์ต้นแบบชื่อว่า “AirAnalytica” ซึ่งออกแบบขึ้นเพื่อแสดงผลการวิเคราะห์และการพยากรณ์ความล่าช้าของเที่ยวบินในรูปแบบที่เข้าใจง่ายและเป็นเชิงโต้ตอบ (Interactive Dashboard) เว็บไซต์ AirAnalytica ถูกพัฒนาโดยใช้ภาษา HTML, CSS, และ JavaScript (D3.js) สำหรับการสร้างกราฟและแผนภูมิแบบโต้ตอบ เพื่อสื่อสารผลลัพธ์ของการวิเคราะห์ข้อมูลในรูปแบบภาพ (Data Visualization) โครงสร้างเว็บไซต์แบ่งออกเป็น 4 หน้าหลัก ดังนี้

4.2.6.1 หน้า Home แสดงชื่อโครงการ ที่มา วัตถุประสงค์ และแหล่งข้อมูลที่ใช้ในการวิเคราะห์

4.2.6.2 หน้า Dashboard แสดงข้อมูลเชิงสถิติและผลการวิเคราะห์ เช่น กราฟเปรียบเทียบสาเหตุของความล่าช้า

4.2.6.3 หน้า Survey แบบสอบถาม เป็นการประยุกต์ใช้ กฎการตัดสินใจจากโมเดล Decision Tree เพื่อสร้างแบบสอบถามเชิงโต้ตอบ ผู้ใช้งานสามารถเลือกปัจจัยที่เกี่ยวข้องกับเที่ยวบิน

หน้า Result ผลลัพธ์การพยากรณ์ หน้านี้เป็นส่วนแสดงผลของการประมวลผลข้อมูลจากแบบสอบถามที่ผู้ใช้งานกรอก โดยระบบจะนำข้อมูลที่ได้รับมาคำนวณและประมวลผลร่วมกันจาก ได้แก่ Decision Tree และ Multiple Linear Regression

โมเดล Decision Tree ใช้สำหรับจำแนกสถานะของเที่ยวบินตามกฎการตัดสินใจ (Decision Rules) ที่ได้จากการวิเคราะห์ในบทก่อนหน้า เพื่อทำนายผลเชิงคุณภาพของเที่ยวบิน เช่น ตรงเวลา (ON TIME), ล่าช้าระดับสั้น (SHORT DELAY), หรือ ล่าช้า (DELAY) ขณะที่โมเดล Multiple Linear Regression (MLR) จะใช้ในการคำนวณค่าความล่าช้าเชิงเวลา (หน่วยเป็นนาที) เพื่อแสดงผลเชิงปริมาณของการพยากรณ์

4.2.6.4 หน้า About แสดงข้อมูลเกี่ยวกับผู้จัดทำ ที่มา วัตถุประสงค์ เครื่องมือ และเทคโนโลยีที่ใช้ในการพัฒนาเว็บไซต์

4.3 บทสรุป

จากผลการดำเนินงานทั้งหมดในโครงงานนี้ สามารถสรุปได้ว่าผู้วิเคราะห์ข้อมูลได้ดำเนินการวิเคราะห์ชุดข้อมูล Flight Delays and Cancellations Dataset (2015) จากเว็บไซต์ Kaggle โดยใช้กระบวนการวิเคราะห์ข้อมูลตามกรอบแนวคิด CRISP-DM อย่างเป็นระบบ

ประกอบด้วยการทำทำความเข้าใจธุรกิจ (Data Understanding) การเตรียมข้อมูล (Data Preparation) การสร้างแบบจำลอง (Modeling) และการประเมินผล (Evaluation) เพื่อศึกษาปัจจัยที่ส่งผลต่อความล่าช้าในการมาถึงของเที่ยวบิน (Arrival Delay) และสร้างแบบจำลองเพื่อพยากรณ์แนวโน้มความล่าช้าในอนาคต

ผลการวิเคราะห์แสดงให้เห็นว่า แบบจำลอง Decision Tree และ Random Forest สามารถจำแนกสถานะเที่ยวบินได้อย่างมีประสิทธิภาพ โดยแบบจำลอง Decision Tree ให้ค่าความแม่นยำสูงสุดที่ $90.01\% \pm 0.27\%$ และสามารถอธิบายความสัมพันธ์ระหว่างปัจจัยได้อย่างชัดเจนผ่านกฎการตัดสินใจ (Decision Rules) ขณะที่แบบจำลอง Random Forest ให้ค่าความแม่นยำ $89.30\% \pm 0.03\%$ ซึ่งมีประสิทธิภาพใกล้เคียงกัน แต่มีข้อจำกัดในการจำแนกกลุ่ม SHORT DELAY เนื่องจากสัดส่วนข้อมูลในกลุ่มนี้มีน้อย

ในส่วน of แบบจำลอง Multiple Linear Regression (MLR) ซึ่งใช้ในการพยากรณ์ค่าความล่าช้าเชิงเวลา (นาฬิกา) พบว่าโมเดลมีค่า $R^2 = 0.873$, $MAE = 7.2$ นาที, และ $RMSE = 9.875$ นาที แสดงให้เห็นว่าแบบจำลองสามารถอธิบายความแปรปรวนของข้อมูลได้ถึง 87.3% และมีความแม่นยำในระดับสูง ผลการวิเคราะห์สรุปได้ว่า ตัวแปรที่มีอิทธิพลมากที่สุดต่อความล่าช้าในการมาถึงของเที่ยวบินคือ DEPARTURE_DELAY, รองลงมาคือ AIRLINE_DELAY และ LATE_AIRCRAFT_DELAY ซึ่งเป็นปัจจัยที่เกี่ยวข้องกับการดำเนินงานของสายการบินโดยตรง

ผลลัพธ์จากแบบจำลองทั้งหมดถูกนำไปประยุกต์ใช้ในระบบเว็บไซต์ต้นแบบ AirAnalytica เพื่อแสดงผลการวิเคราะห์และการพยากรณ์ในรูปแบบเชิงโต้ตอบ (Interactive Visualization) โดยมีการผสานผลลัพธ์ของทั้งสองโมเดลเข้าด้วยกัน โมเดล Decision Tree ถูกนำมาใช้ในการสร้างแบบสอบถาม (Survey Page) เพื่อจำลองการทำนายสถานะเที่ยวบินจากกฎการตัดสินใจ ขณะที่โมเดล MLR ถูกนำมาใช้ในการคำนวณค่าความล่าช้าเชิงเวลา และแสดงผลในหน้า Result ในรูปแบบตัวเลขและกราฟแบบเรียลไทม์ การผสานโมเดลทั้งสองทำให้เว็บไซต์สามารถสื่อสารข้อมูลได้ทั้งเชิงคุณภาพและเชิงปริมาณอย่างครบถ้วน

โดยสรุป การดำเนินโครงการครั้งนี้ทำให้สามารถพัฒนาเครื่องมือวิเคราะห์ข้อมูลที่ช่วยให้เข้าใจปัจจัยสำคัญที่ส่งผลต่อความล่าช้าในการเดินทางทางอากาศ พร้อมทั้งนำเสนอในรูปแบบเว็บไซต์ที่ใช้งานง่าย มีความเป็นมิตรต่อผู้ใช้ และสามารถนำไปประยุกต์ใช้ได้จริงในการวางแผนและบริหารจัดการเที่ยวบินในอนาคต