

การศึกษาการแบ่งกลุ่มผู้โดยสารสายการบินตามแบบประเมินความพึงพอใจด้วยเทคนิคทางวิทยาการข้อมูล

ปัทมณ เกริกชัยวัน¹, โสภณ มงคลลักษณ์²

บทคัดย่อ

การให้ความสำคัญกับคุณภาพการบริการของสายการบินเป็นปัจจัยสำคัญที่มีอิทธิพลต่อความพึงพอใจของลูกค้า งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาวิธีการแบ่งกลุ่มลูกค้า โดยใช้เทคนิคการเรียนรู้ของเครื่อง เพื่อจัดกลุ่มของข้อมูลร่วมกับการวิเคราะห์ความสำคัญของการให้บริการของสายการบิน โดยประกอบไปด้วย 3 วิธีหลักดังนี้ 1. การแบ่งกลุ่มลูกค้าด้วยเทคนิคเคมีนส์ (K-means) 2. การเข้าใจการตัดสินใจของแบบจำลองในการเลือกใช้แต่ละฟีเจอร์ในการแบ่งกลุ่มด้วยเทคนิคต้นไม้ตัดสินใจ (Decision tree) 3. การหาปัจจัยที่สำคัญของการแบ่งกลุ่มจากฟีเจอร์อิมพอร์ตแดนซ์ (Feature importance)

โดยใช้ชุดข้อมูลแบบประเมินความพึงพอใจของผู้โดยสารในสายการบินในสหรัฐฯ ซึ่งดำเนินการสำรวจในปี 2015 และมีข้อมูลที่รวบรวมจากตัวอย่างผู้โดยสารจำนวน 129,879 คนที่ใช้บริการการบินแบบฟูลเซอร์วิส (Full Service) โดยผู้วิจัยเลือกเฉพาะกลุ่มผู้โดยสารที่ไม่พึงพอใจจำนวน 73,452 ราย ผู้วิจัยสนใจที่จะแบ่งกลุ่มลูกค้าที่มาใช้บริการตามการประเมินที่คล้ายคลึงกันของกลุ่มลูกค้า ซึ่งสามารถแบ่งได้เป็น 5 กลุ่มประเภทลูกค้าที่มีการประเมินคล้ายคลึงกันจากการแบ่งกลุ่มของ K-means ซึ่งให้ค่า Silhouette Score ที่ 0.149 และ Davies-Bouldin Index ที่ 1.830 และผู้วิจัยต้องการหาปัจจัยที่มีความสำคัญต่อความไม่พึงพอใจของผู้โดยสารแต่ละกลุ่ม โดยพบว่าแต่ละกลุ่มมีปัจจัยที่ไม่พึงพอใจที่แตกต่างกันออกไป แต่ทุกกลุ่มผู้โดยสารให้ความสำคัญกับการบริการอินเทอร์เน็ตบนเครื่องบิน

คำสำคัญ : การแบ่งกลุ่มลูกค้า, การหาปัจจัยที่สำคัญ, การเรียนรู้ของเครื่อง, เทคนิคเคมีน, เทคนิคต้นไม้ตัดสินใจ

¹ หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิทยาการข้อมูล คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ กรุงเทพฯ 10110

² คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ กรุงเทพฯ 10110

* Corresponding author: Tel.: 081-6379200 E-mail address: patipan.krirk@g.swu.ac.th

A STUDY OF AIRLINE PASSENGER SEGMENTATION BASED ON SATISFACTION EVALUATION USING DATA SCIENCE TECHNIQUES

Patipan Krirkchaiwan^{1*}, Sophon Mongkolluksamee²

Abstract

Focusing on the quality of airline services is a crucial factor influencing customer satisfaction. This research aims to study customer segmentation using machine learning techniques to cluster data alongside analyzing the importance of airline services. The research consists of three main methods: 1. Customer segmentation using the K-means technique 2. Understanding the decision-making of the model in selecting each feature for clustering using the Decision tree technique 3. Identifying the important factors for clustering using Feature importance

The study uses a dataset of passenger satisfaction surveys from airlines in the United States, conducted in 2015, with data collected from 129,879 passengers who used full-service airlines. The researcher selected only the dissatisfied passengers, totaling 73,452 individuals. The researcher is interested in segmenting customers based on similar assessments, which can be divided into five customer groups with similar evaluations using K-means clustering.

This method resulted in a Silhouette Score of 0.149 and a Davies-Bouldin Index of 1.830. The researcher aims to identify the factors important to the dissatisfaction of each passenger group, finding that each group has different dissatisfaction factors, but all passenger groups emphasize the importance of in-flight internet service.

Keywords : Customer segmentation, Feature important , Machine learning, K-means technique, Decision tree technique

¹ Data Science, Faculty of Science, Srinakharinwirot University, Bangkok, 10110, Thailand

² Faculty of Science, Srinakharinwirot University, Bangkok, 10110, Thailand

* Corresponding author: Tel.: 080-5488355 E-mail address: aekpreya.baisani@g.swu.ac.th

บทนำ

การเดินทางโดยเครื่องบินมีความสำคัญมากในชีวิตประจำวันของคนในยุคปัจจุบันเนื่องจากการเดินทางที่มีความรวดเร็วสะดวกสบายและปลอดภัยเป็นอย่างมาก นอกจากนี้ในปัจจุบันราคาตั๋วเครื่องบินที่ถูกต่ำกว่าสมัยก่อนเนื่องจากการแข่งขันทางด้านราคาจากสายการบินที่หลากหลายทำให้ผู้โดยสารมีทางเลือกมากขึ้น สร้างการแข่งขันอย่างเข้มข้นในธุรกิจการบินซึ่งปัจจุบันการเติบโตของอุตสาหกรรมการบินกลับมาสูงขึ้นอีกครั้งหลังจากผ่านพ้นสถานการณ์โควิด ตามความต้องการเดินทางของผู้โดยสารทั่วโลกทำให้เกิดการแข่งขันที่เพิ่มขึ้นระหว่างสายการบินต่างๆ ประกอบกับมีสายการบินที่เกิดขึ้นใหม่เป็นจำนวนมาก

การแข่งขันที่เพิ่มขึ้นนี้ทำให้สายการบินต้องมุ่งเน้นไปที่การบริการเพื่อสร้างความแตกต่างและดึงดูดผู้โดยสารให้เลือกใช้บริการของตน ปัจจัยเช่น ความสะดวกสบาย ความตรงต่อเวลา คุณภาพของบริการภาคพื้นและบนฟ้า และประสบการณ์ของผู้โดยสารล้วนมีส่วนสำคัญในการสร้างความพึงพอใจและความภักดีของผู้โดยสาร ดังนั้น การให้บริการที่เหนือกว่าคู่แข่งสามารถสร้างความได้เปรียบทางธุรกิจที่สำคัญ

การเข้าใจความสำคัญของการบริการและความต้องการของแต่ละกลุ่มผู้โดยสารจึง เป็นกุญแจสำคัญในการปรับปรุงและพัฒนาการบริการให้ตรงกับความต้องการของผู้โดยสาร ซึ่งช่วยให้สายการบินสามารถวางแผนและบริหารการบริการอย่างรอบคอบเพื่อสร้างความพึงพอใจและความภักดีในกลุ่มผู้โดยสาร ซึ่งจะส่งผลดีต่อความสำเร็จในระยะยาวของสายการบินทั้งในเชิงธุรกิจและความเชื่อมั่นของสายการบิน

วิธีดำเนินการ

ขั้นตอนที่ 1 : แนะนำชุดข้อมูลที่ใช้ในการศึกษา

ผู้วิจัยนำข้อมูลเกี่ยวกับการประเมินความพึงพอใจของลูกค้าที่มาใช้บริการสายการบินมาวิเคราะห์ในการทำวิจัยครั้งนี้ โดยใช้ชุดข้อมูลงานวิจัยนี้ได้รับมาจาก Kaggle Dataset ของ "The U.S. Airline Passenger Satisfaction Dataset" ซึ่งอธิบายเกี่ยวกับความพึงพอใจของผู้โดยสารในสายการบินในสหรัฐอเมริกา โดยการดำเนินการสำรวจที่สนามบินในปี 2015 โดยมีข้อมูลที่รวบรวมจากตัวอย่างผู้โดยสารจำนวน 129,879 คนที่ใช้บริการการบินแบบเต็มรูปแบบ(Full Service) โดยผู้วิจัยเลือกเฉพาะกลุ่มผู้โดยสารที่ไม่พึงพอใจจำนวน 73,452 ราย

id	Gender	Customer Age	Type of Tra Class	Flight Dist.	Inflight wif	Departure.	Ease of On Gate	locat	Food and c	Online box	Seat comf	Inflight ent	On-board	Leg room	ε	Baggage h.	Checkin s	Inflight ser	Cleanlines	Departure	Arrival Del	satisfaction		
0	70172	Male	Loyal Cust	13	Personal T	Eco Plus	460	3	4	3	1	5	3	5	5	4	3	4	4	5	25	18	neutral or dissatisfied	
1	5047	Male	disloyal Cu	25	Business t	Business	235	3	2	3	3	1	3	1	1	5	3	1	4	1	1	6	neutral or dissatisfied	
2	110028	Female	Loyal Cust	26	Business t	Business	1142	2	2	2	2	5	5	5	4	3	4	4	4	5	0	0	satisfied	
3	24026	Female	Loyal Cust	25	Business t	Business	562	2	5	5	5	2	2	2	2	5	3	1	4	2	11	9	neutral or dissatisfied	
4	119299	Male	Loyal Cust	61	Business t	Business	214	3	3	3	3	4	5	5	3	3	4	4	3	3	3	0	0	satisfied
5	111157	Female	Loyal Cust	26	Personal T	Eco	1180	3	4	2	1	1	2	1	3	4	4	4	4	1	0	0	neutral or dissatisfied	
6	82113	Male	Loyal Cust	47	Personal T	Eco	1276	2	4	2	3	2	2	2	3	3	4	3	5	2	9	23	neutral or dissatisfied	
7	96462	Female	Loyal Cust	52	Business t	Business	2035	4	3	4	4	5	5	5	5	5	5	4	5	4	4	0	0	satisfied

โดยชุดข้อมูลประกอบไปด้วยข้อมูลส่วนตัวของผู้โดยสาร ดังตารางประกอบที่ 1 ,ข้อมูลของไฟล์ทโดยสารดังตารางประกอบที่ 2 และข้อมูลการบริการต่างๆของสายการบินดังตารางที่ 3

Variable	Description
Satisfaction	ความพึงพอใจของผู้โดยสาร
id	รหัสผู้โดยสาร
Age	อายุ
Gender	เพศ
Customer Type	ประเภทของลูกค้า

ตารางประกอบที่ 1 แสดงตัวแปรของข้อมูลส่วนตัวของผู้โดยสาร

Variable	Description
Type of Travel	ประเภทของการท่องเที่ยว
Class	ระดับชั้นที่นั่ง
Flight distance	ระยะทางของเที่ยวบิน
Departure Delay in Minutes	ความล่าช้าในการออกเดินทาง(ในหน่วยนาที)
Arrival Delay in Minutes	ความล่าช้าในการไปถึงปลายทาง(ในหน่วยนาที)

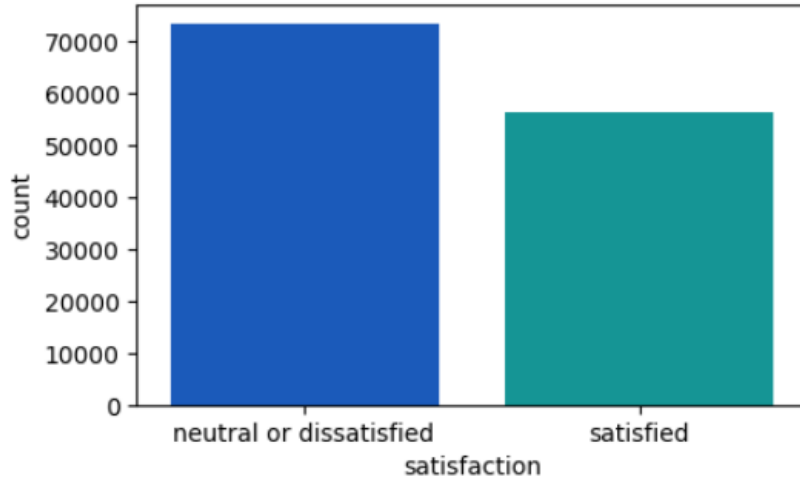
ตารางประกอบที่ 2 แสดงตัวแปรของไฟล์ท

Variable	Description
Inflight wifi service	การบริการอินเทอร์เน็ตบนเครื่อง
Ease of Online booking	การจองแบบออนไลน์
Inflight service	การให้บริการบนเครื่องของพนักงาน
Online boarding	การเช็คอินออนไลน์
Inflight entertainment	ความบันเทิงบนเครื่องบิน
Food and drink	อาหารและเครื่องดื่ม
Seat comfort	ความสบายของที่นั่ง
On-board service	บริการที่มีบนเครื่อง
Leg room service	พื้นที่วางขา
Departure/Arrival time convenient	ความเหมาะสมของเวลาเที่ยวบิน
Baggage handling	การจัดการเรื่องกระเป๋า
Gate location	ตำแหน่งGateโดยสาร
Cleanliness	ความสะอาด
Check-in service	การบริการเช็คอิน

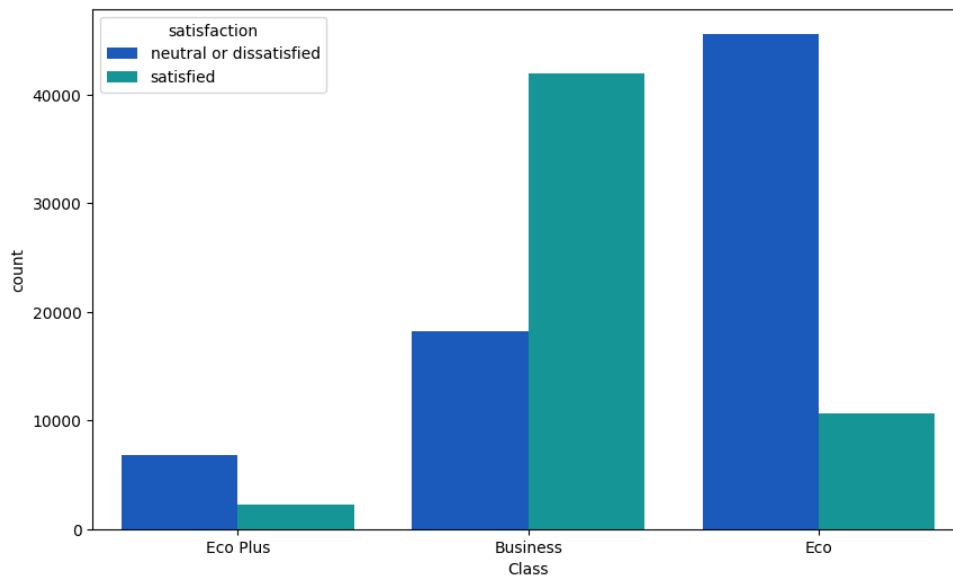
ตารางประกอบที่ 3 แสดงตัวแปรของการบริการของสายการบิน

ขั้นตอนที่ 2 : Exploratory Data Analysis : EDA

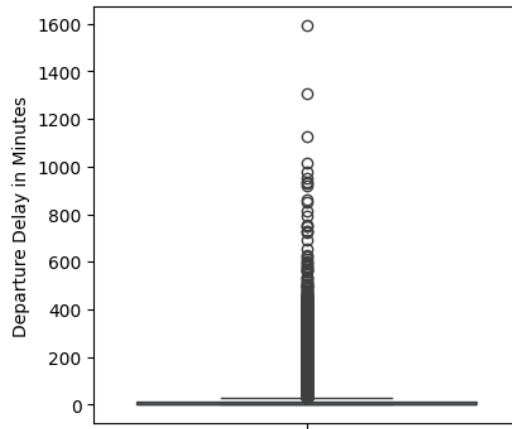
ใช้ภาษาไพทอนในการวิเคราะห์ข้อมูลและการเรียนรู้ของเครื่อง เริ่มต้นด้วยการนำเข้าโมดูลสำคัญสำหรับการสร้างแบบจำลอง ต่อมานำเข้าไฟล์ข้อมูลและข้อมูลที่ใช้สำหรับสร้างแบบจำลอง เริ่มกระบวนการตรวจสอบและสำรวจข้อมูลเบื้องต้น เพื่อหาข้อมูลเชิงลึกจากความพึงพอใจของลูกค้าโดยใช้ไลบรารี Pandas, Numpy, Seabone และ Matplotlib สำรวจข้อมูลกับข้อมูลคอลัมน์ชื่อความพึงพอใจดังรูปที่ 1



ข้อมูลส่วนตัวของลูกค้าเช่นเพศ,อายุ และข้อมูลประเภทการเดินทางของลูกค้าเช่นคลาส,ลูกค้าประจำไม่ประจำ เพื่อพยายามแบ่งกลุ่มลูกค้าสายการบินเป็นกลุ่ม



จากนั้นหาค่าที่เป็นmissing value ที่ปรากฏในชุดข้อมูลโดยใช้ไลบรารี pandas,outlier ที่ปรากฏจากการ box plot ในชุดข้อมูลโดยใช้ไลบรารี Seaborn

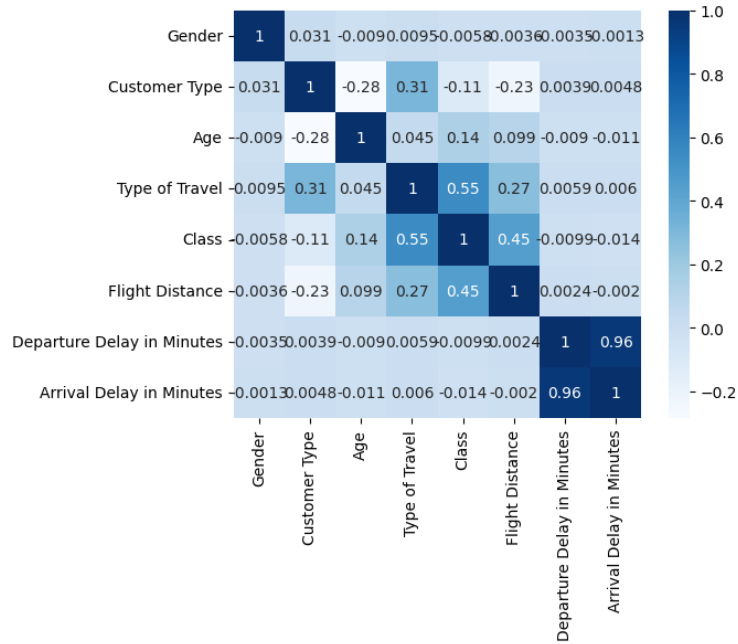


ขั้นตอนที่ 3 : Data Preprocessing and Cleaning

การจัดเตรียมข้อมูล ผู้วิจัยได้ตัดข้อมูลบางแถวออกไม่นำข้อมูลต่อไปนี้มีรวมวิเคราะห์ ได้แก่ การเดินทางที่ไกลกว่า 4,000 ไมล์ ไฟลท์ที่เลยในขาออกที่มากกว่า 1 ชั่วโมง 40 นาที และไฟลท์ที่ตีเลยในขาเข้า และเพื่อให้ข้อมูลเป็นประเภทเดียวกันผู้วิจัยใช้การทำ endcoding เพื่อให้ข้อมูลอยู่ในรูปแบบเดียวกัน

	Gender	Customer Type	Age	Type of Class	Class
0	0	0	13	0	1
1	0	1	25	1	2
2	1	0	26	1	2
3	1	0	25	1	2
4	0	0	61	1	2

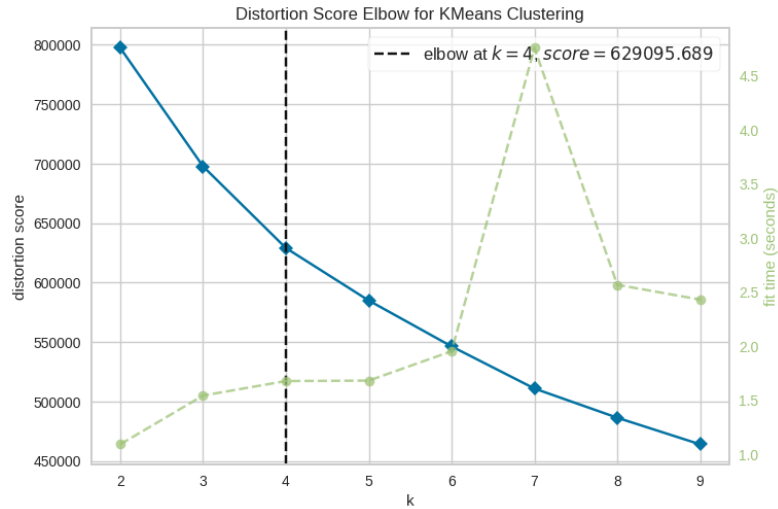
เมื่อทำแปลงข้อมูลให้อยู่ในรูปแบบเดียวกันแล้วผู้วิจัยได้วิเคราะห์ความสัมพันธ์ของข้อมูลจากค่า correlation เนื่องจากการตีเลย์ขาเข้าและขาออกมีค่าความสัมพันธ์ที่สูงมาก เพราะเมื่อเครื่องบินออกเดินทางช้าก็แปรผันตามว่าจะถึงปลายทางช้าด้วยจึงลบปัจจัยการตีเลย์ขาเข้า หลังจากนั้นผู้วิจัยได้ทำการสเกลข้อมูลด้วยการทำ StandardScaler เพื่อให้ข้อมูลอยู่ในสเกลเดียวกัน



และนำข้อมูลไปผ่านการลดมิติของข้อมูลด้วยการทำ PCA (Principal Component Analysis) เพื่อลดความสำคัญของปัจจัยที่ไม่ได้ส่งผลต่อความพึงพอใจ และนำข้อมูลที่ผ่านการ PCA แล้ว โดยผู้วิจัยคำนึงถึงค่า Eigence Value และ Variance ratio ในการเลือก PC(Principle Component)

ขั้นตอนที่ 4 : การแบ่งกลุ่มข้อมูล(K-means)

ผู้วิจัยวิเคราะห์หาค่าจำนวนกลุ่มที่เหมาะสมที่สุดต่อข้อมูลชุดนี้ โดยผู้วิจัยเลือกมา 2 วิธี ได้แก่ Elbow Method และ Silhouette Score โดยใช้ Silhouette Score และ Davies-Bouldin Index เป็นเทคนิคที่ใช้วัดตัวอย่างนั้นๆ มีความเหมือนกับกลุ่มที่อยู่มากเพียงใด เมื่อเทียบกับกลุ่มอื่นๆ และทำการเลือกจำนวนกลุ่มที่ดีที่สุดของข้อมูลชุดนี้



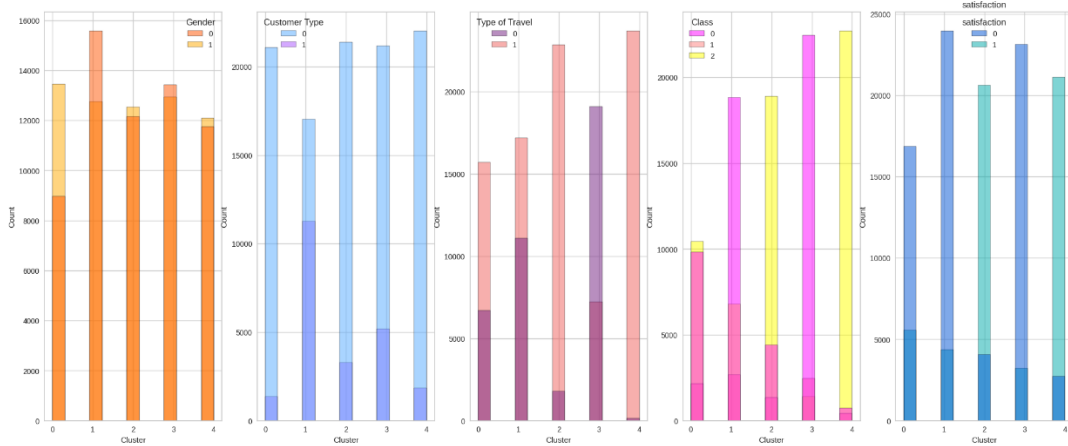
ขั้นตอนที่ 5 : การหาปัจจัยที่มีความสำคัญต่อลูกค้าแต่ละกลุ่ม

จากการศึกษาการจำแนกกลุ่ม ผู้วิจัยเลือกใช้แบบจำลอง K-means ในการแบ่งกลุ่ม โดยกำหนดจำนวนกลุ่มของข้อมูลเป็น 4 กลุ่ม หลังจากนั้น กลุ่มของลูกค้าที่แบ่งได้จะถูกเพิ่มเป็นคอลัมน์ใหม่ในชุดข้อมูล และผู้วิจัยทำการตรวจสอบข้อมูลในแต่ละกลุ่มของลูกค้า เพื่อสามารถระบุประเภทต่างๆ ของลูกค้าในแต่ละกลุ่มที่มาใช้บริการได้

ในการศึกษาี้ ผู้วิจัยได้นำกลุ่มลูกค้าแต่ละกลุ่มมาหาปัจจัยที่สำคัญที่ส่งผลต่อความพึงพอใจของลูกค้าแต่ละกลุ่ม และระบุปัญหาการบริการของแต่ละกลุ่มและสามารถระบุประเภทของแต่ละกลุ่มได้ ด้วยแบบจำลองต้นไม้ตัดสินใจ (Decision Tree) ซึ่งผู้วิจัยได้ทำการแบ่งข้อมูลออกเป็น 2 ส่วน คือข้อมูลสำหรับฝึกและข้อมูลสำหรับทดสอบด้วยโมดูล train-test-split เพื่อทดสอบว่าโมเดลที่เลือกมาใช้นั้นมีประสิทธิภาพในการอธิบายการตัดสินใจของการแบ่งกลุ่ม หลังจากนั้นดูฟีเจอร์อิมพอร์ตเทนซ์ (Feature Importance) เพื่อดูปัจจัยที่สำคัญต่อการแบ่งกลุ่ม และใช้ bar plot จากไลบรารี Seaborn ทำให้ผู้วิจัยเห็นภาพรวมและสรุปได้ง่ายขึ้น

ผลการวิจัยและอภิปรายผลการวิจัย

การทดลองนี้ใช้แบบจำลองเคมีน(K-means) สำหรับการแบ่งกลุ่มลูกค้า มีจำนวนทั้งสิ้น 129,879เรคคอร์ด โดยใช้จำนวนกลุ่มของลูกค้า5กลุ่มให้ค่า Silhouette Score และ Davies-Bouldin Index ที่ 0.149 และ 1.83 ตามลำดับ



หลังจากการวิเคราะห์แต่ละกลุ่มแล้วพบว่า

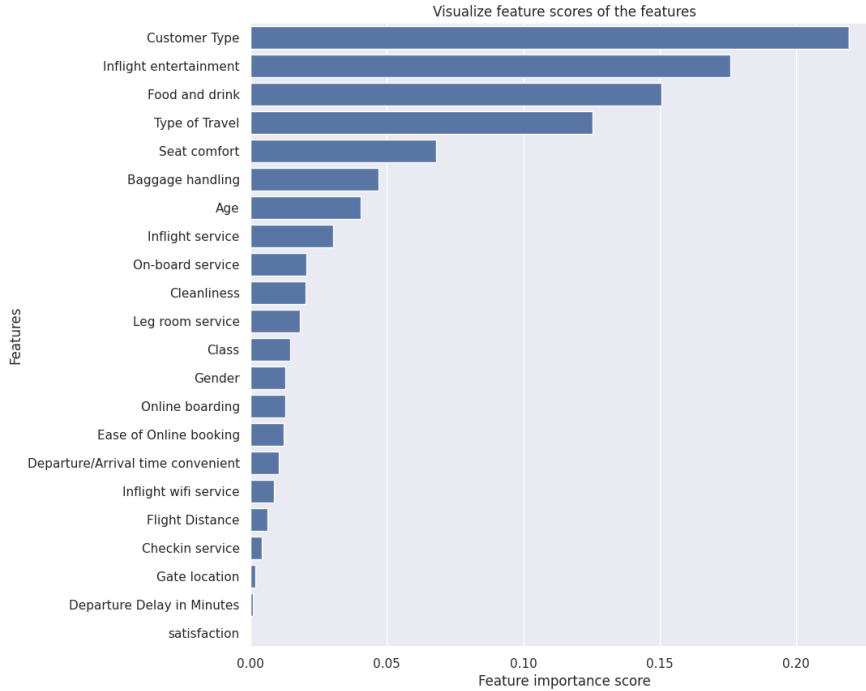
ลูกค้ากลุ่มที่1:ส่วนมากเป็นเพศหญิง ที่เดินทางเป็นประจำด้วยการเดินทางแบบส่วนตัว นั่งในส่วนชั้นโดยสารประหยัดและธุรกิจ ใกล้เคียงกัน โดยรวมแล้วไม่พึงพอใจกับการบริการ

ลูกค้ากลุ่มที่2:ส่วนมากเป็นเพศชาย ที่เดินทางเป็นประจำและเป็นครั้งคราว ด้วยการเดินทางแบบส่วนตัวและแบบธุรกิจ นั่งในส่วนชั้นโดยสารประหยัด โดยรวมแล้วไม่พึงพอใจกับการบริการ

ลูกค้ากลุ่มที่3:เป็นเพศชายและหญิงพอกัน ที่เดินทางเป็นประจำด้วยการเดินทางแบบส่วนตัว นั่งในส่วนชั้นโดยสารชั้นธุรกิจ โดยรวมแล้วพึงพอใจกับการบริการ

ลูกค้ากลุ่มที่4:เป็นเพศชายและหญิงพอกัน ที่เดินทางเป็นประจำด้วยการเดินทางแบบธุรกิจ นั่งในส่วนชั้นโดยสารชั้นประหยัด โดยรวมแล้วไม่พึงพอใจกับการบริการ

ลูกค้ากลุ่มที่5:เป็นเพศชายและหญิงพอกัน ที่เดินทางเป็นประจำด้วยการเดินทางแบบส่วนตัว นั่งในส่วนชั้นโดยสารชั้นธุรกิจ โดยรวมแล้วพึงพอใจกับการบริการ



ปัจจัยที่สำคัญต่อการแบ่งกลุ่มผู้โดยสารที่ไม่พึงพอใจคือ ประเภทของผู้โดยสาร ความบันเทิงบนเครื่องบิน อาหารและเครื่องดื่ม

สรุปผลการวิจัย

งานวิจัยนี้เป็นการศึกษาวิธีการแบ่งกลุ่มลูกค้าด้วยเทคนิคเคมีน(K-means)และหาปัจจัยสำคัญจากเทคนิคต้นไม้ตัดสินใจ (Decision Tree) และเทคนิคพีเจอร์ที่สำคัญ(Feature Importance) ผลการทดลองพบว่าวิธีการจัดกลุ่มด้วยเทคนิคเคมีน(K-means) สามารถระบุกลุ่มลูกค้าที่แตกต่างกันได้ดีที่สุดที่ 5 กลุ่ม และจากการหาปัจจัยที่สำคัญที่มีผลต่อแบ่งกลุ่มความพึงพอใจด้วยวิธีการพีเจอร์ที่สำคัญ(Feature Importance) พบว่าทุกกลุ่มลูกค้าให้ความสำคัญกับประเภทของลูกค้า ความบันเทิงบนเครื่องบิน และอาหารและเครื่องดื่ม

กิตติกรรมประกาศ

การจัดทำวิจัยได้รับการสนับสนุนจากบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ในการนำเสนอผลงานวิจัย ผู้วิจัยจึงขอขอบคุณมา ณ ที่นี้

เอกสารอ้างอิง

- [1] Maryani I, Riana D, Astuti RD, Ishaq A, Pratama EA, editors. Customer Segmentation based on RFM model and Clustering Techniques With K-Means Algorithm. 2018Third International Conference on Informatics and Computing (ICIC); 2018: IEEE.
- [2] Chen D, Sain SL, Guo K. Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. *Journal of Database Marketing & Customer Strategy Management*. 2012;19(3):197-208.
- [3] Kansal T, Bahuguna S, Singh V, Choudhury T, editors. Customer segmentation using k-means clustering. 2018International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS); 2018: IEEE.
- [4] Hu X, Zhang H, Wu X, Chen J, Xiao Y, Xue Y, et al., editors. A Novel Approach for Customer Segmentation Based on Biclustering. *International Conference on Web Information Systems Engineering*; 2013: Springer.
- [5] Wang B, Miao Y, Zhao H, Jin J, Chen Y. A biclustering-based method for market segmentation using customer pain points. *Engineering Applications of Artificial Intelligence*. 2016;47:101-9.
- [6] Chen, D. (2015). UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Public Analytics group, London South Bank University, School of Engineering