



การจัดกลุ่มลูกค้าด้วยเทคนิค RFM และการวิเคราะห์ตะกร้าสินค้าเพื่อทำการตลาดเฉพาะกลุ่ม
CUSTOMER SEGMENTATION USING RFM TECHNIQUES AND MARKET BASKET
ANALYSIS FOR TARGETED MARKETING

ธนกร กอวรพันธุ์

การจัดกลุ่มลูกค้าด้วยเทคนิค RFM และการวิเคราะห์ตะกร้าสินค้าเพื่อทำการตลาดเฉพาะกลุ่ม



สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล
คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
ปีการศึกษา 2567
ลิขสิทธิ์ของมหาวิทยาลัยศรีนครินทรวิโรฒ

CUSTOMER SEGMENTATION USING RFM TECHNIQUES AND MARKET BASKET
ANALYSIS FOR TARGETED MARKETING



An Master's Project Submitted in Partial Fulfillment of the Requirements
for the Degree of MASTER OF SCIENCE
(Data Science)

Faculty of Science, Srinakharinwirot University

2024

Copyright of Srinakharinwirot University

สารนิพนธ์
เรื่อง
การจัดกลุ่มลูกค้าด้วยเทคนิค RFM และการวิเคราะห์ตะกร้าสินค้าเพื่อทำการตลาดเฉพาะกลุ่ม
ของ
ธนกร กอวรพันธุ์

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล
ของมหาวิทยาลัยศรีนครินทรวิโรฒ

(รองศาสตราจารย์ นายแพทย์ฉัตรชัย เอกปัญญาสกุล)
คณบดีบัณฑิตวิทยาลัย

คณะกรรมการสอบปากเปล่าสารนิพนธ์

..... ที่ปรึกษาหลัก
(อาจารย์ ดร.ศุภร คนธภาคี)

..... ประธาน
(ผู้ช่วยศาสตราจารย์ ดร.จันตรี ผลประเสริฐ)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.ศิริสรรพ เหล่าหะเกียรติ)

ชื่อเรื่อง	การจัดกลุ่มลูกค้าด้วยเทคนิค RFM และการวิเคราะห์ตะกร้าสินค้าเพื่อทำ การตลาดเฉพาะกลุ่ม
ผู้วิจัย	ธนกร กอวรพันธุ์
ปริญญา	วิทยาศาสตร์มหาบัณฑิต
ปีการศึกษา	2567
อาจารย์ที่ปรึกษา	อาจารย์ ดร. ศุภร คนธภักดี

ในยุคดิจิทัลที่ข้อมูลมีบทบาทสำคัญในการขับเคลื่อนธุรกิจ การเข้าใจความต้องการและพฤติกรรมของลูกค้าเป็นปัจจัยสำคัญสู่ความสำเร็จ การวิจัยนี้มีวัตถุประสงค์เพื่อแบ่งกลุ่มลูกค้าตามพฤติกรรมการซื้อ โดยใช้เทคนิคการวิเคราะห์ RFM (Recency, Frequency, Monetary) ร่วมกับการวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis) เพื่อพัฒนากลยุทธ์การตลาดเฉพาะกลุ่มอย่างมีประสิทธิภาพ ข้อมูลที่ใช้เป็นข้อมูลธุรกรรมจากร้านค้าปลีกออนไลน์ในประเทศสหราชอาณาจักร ระหว่างวันที่ 1 ธันวาคม 2009 ถึง 9 ธันวาคม 2011 รวมทั้งสิ้น 1,067,371 รายการ การวิจัยประกอบด้วยขั้นตอนการสร้างฟีเจอร์จากค่า RFM และการจัดกลุ่มลูกค้าด้วยเทคนิค Machine Learning ได้แก่ Gaussian Mixture Model (GMM), K-Means Clustering และ Hierarchical Clustering โดยประเมินประสิทธิภาพด้วย Silhouette Score และ Davies-Bouldin Index (DBI) ผลลัพธ์จากการจัดกลุ่มถูกนำไปวิเคราะห์สินค้าด้วยหลักการ ABC Analysis และใช้เทคนิค FP-Growth เพื่อค้นหาความสัมพันธ์ของสินค้าที่มักถูกซื้อร่วมกัน ผลการศึกษาชี้ให้เห็นว่าการผสมผสานเทคนิค RFM และการวิเคราะห์ตะกร้าสินค้า สามารถสนับสนุนการกำหนดกลยุทธ์ทางการตลาดที่ตรงเป้าหมาย เพิ่มความพึงพอใจและความภักดีของลูกค้าได้อย่างมีประสิทธิภาพ

คำสำคัญ : การวิเคราะห์ RFM, การวิเคราะห์ตะกร้าสินค้า, การจัดกลุ่มลูกค้า, การเรียนรู้ของเครื่อง, กลยุทธ์การตลาดเฉพาะกลุ่ม

Title	CUSTOMER SEGMENTATION USING RFM TECHNIQUES AND MARKET BASKET ANALYSIS FOR TARGETED MARKETING
Author	THANAKORN KOWARAPUNTH
Degree	MASTER OF SCIENCE
Academic Year	2024
Thesis Advisor	Dr. Subhorn Khonthapagdee

In the digital era where data plays a vital role in driving business success, understanding customer needs and behaviors is a key factor. This study aims to segment customers based on purchasing behavior using the RFM (Recency, Frequency, Monetary) technique in combination with Market Basket Analysis to develop effective targeted marketing strategies. The dataset used consists of 1,067,371 transactions from a UK-based online retail store collected between December 1, 2009, and December 9, 2011. The research involves feature engineering based on RFM values and customer segmentation using machine learning techniques including Gaussian Mixture Model (GMM), K-Means Clustering, and Hierarchical Clustering. The performance of clustering models is evaluated using the Silhouette Score and Davies-Bouldin Index (DBI). The segmented data is then analyzed through ABC Analysis to classify product importance, and the FP-Growth algorithm is applied to discover frequent item associations. The results indicate that combining RFM analysis with Market Basket Analysis effectively supports the development of personalized marketing strategies that enhance customer satisfaction and foster loyalty.

Keyword : RFM Analysis, Market Basket Analysis, Customer Segmentation, Machine Learning, Targeted Marketing Strategy

กิตติกรรมประกาศ

ข้าพเจ้าขอกราบขอบพระคุณ อาจารย์ ดร. ศุภร คนธภักดี อาจารย์ที่ปรึกษาเป็นอย่างสูง ที่ได้กรุณาให้คำแนะนำอันทรงคุณค่า พร้อมทั้งชี้แนะแนวทางการดำเนินงานวิจัยอย่างใกล้ชิด ด้วยความเอาใจใส่ตลอดระยะเวลาของการจัดทำสารนิพนธ์ฉบับนี้ ทั้งยังช่วยชี้ให้เห็นถึงข้อบกพร่องและแนวทางแก้ไขอย่างสร้างสรรค์ ซึ่งมีส่วนสำคัญต่อความสำเร็จของงานวิจัยในครั้งนี้เป็นอย่างยิ่ง

ข้าพเจ้าขอขอบคุณ คณาจารย์ทุกท่านในสาขาวิชาวิทยาการข้อมูล คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ที่ได้ถ่ายทอดความรู้และให้ข้อเสนอแนะอันเป็นประโยชน์อย่างยิ่งต่อการเรียนรู้และการพัฒนางานวิจัยของข้าพเจ้า นอกจากนี้ข้าพเจ้าขอขอบพระคุณ คณะกรรมการสอบสารนิพนธ์ ที่ได้กรุณาสละเวลาให้ข้อคิดเห็นเชิงวิชาการอันทรงคุณค่า ซึ่งช่วยส่งเสริมให้งานวิจัยฉบับนี้มีความสมบูรณ์มากยิ่งขึ้น

ท้ายที่สุดนี้ ข้าพเจ้าขอขอบคุณ ครอบครัวและเพื่อน ๆ ทุกคน ที่คอยเป็นกำลังใจและให้การสนับสนุนด้วยความรักและความเข้าใจอย่างไม่มีเงื่อนไขตลอดระยะเวลาของการศึกษา โดยเฉพาะอย่างยิ่ง ข้าพเจ้าขอขอบคุณ ภรรยา “ไพลิน” และลูกชาย “น้องธีร์” ที่เป็นแรงใจสำคัญและอยู่เคียงข้างเสมอ

ข้าพเจ้าหวังเป็นอย่างยิ่งว่างานวิจัยฉบับนี้จะเป็นประโยชน์ต่อผู้ที่สนใจในสาขาวิชานี้ และหากมีข้อผิดพลาดประการใด ข้าพเจ้าขอน้อมรับไว้ด้วยความเคารพ และขออภัยมา ณ โอกาสนี้

ธนกร กอวรพันธุ์

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง	ญ
สารบัญรูปภาพ.....	ฎ
บทที่ 1 บทนำ	1
1.1 ภูมิหลัง.....	1
1.2 ความมุ่งหมายของงานวิจัย.....	2
1.3 ความสำคัญของงานวิจัย	2
1.4 ขอบเขตของงานวิจัย.....	2
1.5 กรอบแนวคิดในงานวิจัย	3
1.6 สมมติฐานในการวิจัย.....	4
บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง	5
2.1 การแบ่งกลุ่มลูกค้า (Customer Segmentation)	5
2.1.1 การแบ่งกลุ่มลูกค้าตามหลักการ RFM	6
2.1.2 การจัดกลุ่มลูกค้าด้วยแบบจำลอง (Customer Clustering)	6
2.2 กระบวนการในการสร้างหรือปรับปรุงคุณลักษณะ (Feature Engineering).....	7
2.3 การวิเคราะห์กลุ่มสินค้าด้วยหลักการ ABC (ABC Analysis).....	8
2.4 การวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis)	9
2.5 แบบจำลองที่ใช้ในงานวิจัย	11

2.5.1 Gaussian Mixture Model (GMM)	11
2.5.2 K-means Clustering.....	12
2.5.3 Hierarchical Clustering	13
2.6 การประเมินผลแบบจำลอง	15
2.6.1 Silhouette Score.....	15
2.6.2 Davies-Bouldin Index (DBI)	15
2.7 งานวิจัยที่เกี่ยวข้อง	15
2.7.1 Customer Segmentation by Using RFM Model and Clustering Methods : A Case Study in Retail Industry (Doğan และคณะ, 2018)	15
2.7.2 Analysis Of Increasing Student Service Satisfaction Using K-Means Clustering Algorithm and Gaussian Mixture Models (GMM) (Maulidya และคณะ, 2024)..	17
2.7.3 Using RFM Model and Market Basket Analysis for Segmenting Customers and Assigning Marketing Strategies to Resulted Segments (Maraghi และคณะ, 2020).....	18
2.7.4 A Comparative Study of RFM - Based Clustering Methods in Customer Segmentation (Ahammed และคณะ, 2023)	18
2.7.5 An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market (John และคณะ, 2023)	19
บทที่ 3 วิธีดำเนินการวิจัย.....	20
3.1 กระบวนการทำงานวิจัย.....	20
3.2 การทำความเข้าใจและตรวจสอบข้อมูล	21
3.2.1 การทำความเข้าใจข้อมูล.....	21
3.2.2 การตรวจสอบข้อมูล	22
3.3 วิเคราะห์ข้อมูลเชิงสำรวจ (Exploratory Data Analysis)	23

3.4 การเตรียมข้อมูล (Data Preprocessing)	27
3.5 การสร้างแบบจำลองจัดกลุ่มลูกค้า (Clustering Model)	29
3.6 การประเมินประสิทธิภาพแบบจำลอง (Model Evaluation)	32
3.7 การวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis).....	32
บทที่ 4 ผลการดำเนินการวิจัย.....	33
4.1 ผลลัพธ์การจัดกลุ่มด้วยแบบจำลอง Gaussian Mixture Model	33
4.2 ผลลัพธ์การจัดกลุ่มด้วยแบบจำลอง K-Means Clustering	35
4.3 ผลลัพธ์การจัดกลุ่มด้วยแบบจำลอง Hierarchical Clustering.....	37
4.4 เปรียบเทียบผลลัพธ์การจัดกลุ่ม	38
4.5 ผลการวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis)	41
บทที่ 5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ.....	44
5.1 สรุปผลการวิจัย	44
5.2 อภิปรายผลการวิจัย	47
5.3 ข้อเสนอแนะ.....	52
บรรณานุกรม.....	54
ประวัติผู้เขียน	56

สารบัญตาราง

	หน้า
ตาราง 1 แสดงคุณลักษณะของข้อมูล	21
ตาราง 2 แสดงการแบ่งช่วงค่า RFM ให้อยู่ในรูปแบบ Score	28
ตาราง 3 แสดงค่าเฉลี่ยของแต่ละตัวแปร RFM Score	34
ตาราง 4 ผลลัพธ์การวิเคราะห์กลุ่มสินค้าตามหลักการ RFM Analysis.....	42
ตาราง 5 จำนวนกฎความสัมพันธ์สำหรับลูกค้าแต่ละกลุ่ม	42
ตาราง 6 ตัวอย่างกฎความสัมพันธ์ที่ได้ของลูกค้าแต่ละกลุ่ม	43
ตาราง 7 เปรียบเทียบผลลัพธ์การจัดกลุ่ม	45
ตาราง 8 ตัวอย่างโปรโมชันสำหรับลูกค้าแต่ละกลุ่ม	45
ตาราง 9 แสดงผลลัพธ์การจัดกลุ่มตามหลัก RFM Analysis ของ GMM.....	47
ตาราง 10 แสดงผลลัพธ์การจัดกลุ่มตามหลัก RFM Analysis ของ K-Means	48
ตาราง 11 แสดงผลลัพธ์การจัดกลุ่มตามหลัก RFM Analysis ของ Hierarchical.....	49
ตาราง 12 แสดงตัวอย่างโปรโมชันที่ได้สำหรับลูกค้าแต่ละกลุ่ม โดยเลือกจากระดับความเชื่อมั่น (Confidence) สูงสุด.....	51
ตาราง 13 แสดงตัวอย่างโปรโมชันที่ได้สำหรับลูกค้าแต่ละกลุ่ม โดยเลือกจากจำนวนสินค้าที่ขายได้น้อยที่สุด	51
ตาราง 14 แสดงตัวอย่างโปรโมชันที่ได้สำหรับลูกค้าแต่ละกลุ่ม โดยเลือกจากยอดขายได้ที่ขายได้น้อยที่สุด	52

สารบัญรูปภาพ

	หน้า
ภาพประกอบ 1 แสดงรูปแบบการจัดกลุ่มของ Gaussian Mixture Model.....	11
ภาพประกอบ 2 แสดงรูปแบบการจัดกลุ่มของ K-means Clustering	12
ภาพประกอบ 3 แสดงรูปแบบการจัดกลุ่มของ Hierarchical Clustering.....	14
ภาพประกอบ 4 แสดงกระบวนการทำงานของงานวิจัย.....	20
ภาพประกอบ 5 แสดงตัวอย่างข้อมูล 5 แถวแรก.....	22
ภาพประกอบ 6 แสดงชนิดของข้อมูล.....	22
ภาพประกอบ 7 แสดงการจัดการค่าผิดปกติ (Outliers).....	22
ภาพประกอบ 8 แสดงการจัดการข้อมูลไม่ถูกต้อง (Incorrectly data) และข้อมูลที่สูญหายไป (Missing Values) และสร้างตัวแปรใหม่.....	23
ภาพประกอบ 9 แสดงผลหลังจากการตรวจสอบและแก้ไขข้อมูล.....	23
ภาพประกอบ 10 แสดงจำนวนธุรกรรมต่อเดือน.....	24
ภาพประกอบ 11 แสดงจำนวนขึ้นงานต่อเดือน	24
ภาพประกอบ 12 แสดงจำนวนรายได้ต่อเดือน	25
ภาพประกอบ 13 แสดงประเทศที่มีจำนวนธุรกรรมสูงสุด 10 อันดับ.....	25
ภาพประกอบ 14 แสดงการกระจายตัวของจำนวนสินค้าต่อธุรกรรม	26
ภาพประกอบ 15 แสดงการกระจายตัวของยอดรายได้ต่อธุรกรรม	26
ภาพประกอบ 16 แสดงการกระจายตัวของยอดรายได้ต่อลูกค้า	27
ภาพประกอบ 17 แสดงการวิเคราะห์ความสัมพันธ์ระหว่างคุณลักษณะ.....	27
ภาพประกอบ 18 แสดงการสร้างคุณลักษณะ recency, frequency, monetary	28
ภาพประกอบ 19 แสดงการสร้างคุณลักษณะ RecencyScore, FrequencyScore, MonetaryScore	28
ภาพประกอบ 20 แสดงการกระจายตัวของตัวแปร RFM	29

ภาพประกอบ 21 แสดงการหาจำนวนกลุ่มที่เหมาะสมที่สุดของแบบจำลอง GMM.....	30
ภาพประกอบ 22 แสดงการหาจำนวนกลุ่มที่เหมาะสมที่สุดของแบบจำลอง K-Means.....	31
ภาพประกอบ 23 แสดงการหาจำนวนกลุ่มที่เหมาะสมที่สุดของแบบจำลอง Hierarchical.....	31
ภาพประกอบ 24 แสดงการค้นหารายการสินค้าที่มีการซื้อบ่อย (Frequent Itemsets).....	32
ภาพประกอบ 25 แสดงการสร้างกฎความสัมพันธ์ (Association Rules) ระหว่างสินค้า.....	33
ภาพประกอบ 26 แสดงผลลัพธ์การจัดกลุ่มด้วย Gaussian Mixture Model ในรูปแบบ 3 มิติ.....	33
ภาพประกอบ 27 แสดงผลลัพธ์การจัดกลุ่มด้วย Gaussian Mixture Model ในรูปแบบ 2 มิติ.....	34
ภาพประกอบ 28 แสดงผลลัพธ์การวิเคราะห์พฤติกรรมของกลุ่มลูกค้า GMM ด้วย RFM Analysis.	35
ภาพประกอบ 29 แสดงค่าเฉลี่ย RFM ของกลุ่มลูกค้า GMM.....	35
ภาพประกอบ 30 แสดงผลลัพธ์การจัดกลุ่มด้วย K-Means Clustering ในรูปแบบ 3 มิติ.....	35
ภาพประกอบ 31 แสดงผลลัพธ์การจัดกลุ่มด้วย K-Means Clustering ในรูปแบบ 2 มิติ.....	36
ภาพประกอบ 32 แสดงผลลัพธ์การวิเคราะห์พฤติกรรมของกลุ่มลูกค้า K-Means ด้วย RFM Analysis	36
ภาพประกอบ 33 แสดงค่าเฉลี่ย RFM ของกลุ่มลูกค้า K-Means.....	36
ภาพประกอบ 34 แสดงผลลัพธ์การจัดกลุ่ม Hierarchical Clustering ในรูปแบบ 3 มิติ.....	37
ภาพประกอบ 35 แสดงผลลัพธ์การจัดกลุ่ม Hierarchical Clustering ในรูปแบบ 2 มิติ.....	37
ภาพประกอบ 36 แสดงผลลัพธ์การวิเคราะห์พฤติกรรมของกลุ่มลูกค้า Hierarchical ด้วย RFM Analysis	38
ภาพประกอบ 37 แสดงค่าเฉลี่ย RFM ของกลุ่มลูกค้า Hierarchical.....	38
ภาพประกอบ 38 ลักษณะของลูกค้าแต่ละกลุ่มของ GMM ตามหลัก RFM Analysis.....	39
ภาพประกอบ 39 ลักษณะของลูกค้าแต่ละกลุ่มของ K-Means ตามหลัก RFM Analysis	39
ภาพประกอบ 40 ลักษณะของลูกค้าแต่ละกลุ่มของ Hierarchical ตามหลัก RFM Analysis	40

บทที่ 1

บทนำ

1.1 ภูมิหลัง

ในยุคดิจิทัลที่ข้อมูลกลายเป็นหัวใจสำคัญในการขับเคลื่อนธุรกิจ การทำความเข้าใจถึงความต้องการและรูปแบบพฤติกรรมของลูกค้าเป็นกุญแจสำคัญสู่ความสำเร็จ การแบ่งกลุ่มลูกค้าอย่างแม่นยำทำให้ธุรกิจสามารถสร้างกลยุทธ์การตลาดเฉพาะกลุ่ม (Targeted Marketing) ที่มีประสิทธิภาพและตรงเป้าหมายมากขึ้น โดยเทคนิค RFM (Recency, Frequency, Monetary) เป็นหนึ่งในวิธีการวิเคราะห์ลูกค้าที่ยอดนิยม เนื่องจากเป็นเครื่องมือที่สามารถประเมินความคุ้มค่าและศักยภาพของลูกค้าได้ง่าย โดยการพิจารณาปัจจัย 3 ด้าน คือ ความถี่ในการซื้อ (Frequency) มูลค่ารวมของการใช้จ่าย (Monetary) และความถี่ของการซื้อครั้งล่าสุด (Recency) ข้อมูลนี้สามารถช่วยให้ธุรกิจแยกแยะลูกค้าตามมูลค่าที่สร้างให้กับธุรกิจ และจัดกลุ่มเพื่อพัฒนากลยุทธ์การตลาดที่แตกต่างกันตามกลุ่มลูกค้า

นอกจากการวิเคราะห์ RFM แล้ว การวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis) เป็นอีกหนึ่งเทคนิคที่สามารถค้นหาความสัมพันธ์ของสินค้าที่มักถูกซื้อพร้อมกัน ข้อมูลนี้ช่วยให้ธุรกิจสามารถพัฒนาการเสนอขายสินค้าแบบ Cross-selling หรือ Upselling ได้อย่างตรงจุด อีกทั้งยังสามารถช่วยปรับปรุงประสบการณ์ของลูกค้าโดยการนำเสนอสินค้าที่ตรงกับความต้องการเฉพาะบุคคล ด้วยการใช้เทคนิค RFM สามารถสร้างพีเจอร์ที่ครอบคลุมมิติต่าง ๆ ของพฤติกรรมซื้อสินค้า ช่วยให้การจัดกลุ่มลูกค้ามีความแม่นยำมากขึ้น และวิเคราะห์ตะกร้าสินค้าตามกลุ่มลูกค้าการทำเช่นนี้จะช่วยให้ธุรกิจสามารถพัฒนากลยุทธ์การตลาดเฉพาะกลุ่มที่ตอบโจทย์กลุ่มเป้าหมายได้อย่างมีประสิทธิภาพ อีกทั้งยังเสริมสร้างความภักดีให้แก่ลูกค้า

ดังนั้น การศึกษานี้จึงเห็นความสำคัญในการนำเทคนิค RFM และการวิเคราะห์ตะกร้าสินค้ามาใช้ร่วมกัน เพื่อพัฒนาการจัดกลุ่มลูกค้าที่มีความลึกซึ้งและนำไปสู่การพัฒนาพฤติกรรมการตลาดที่เจาะจงกลุ่มเป้าหมายได้มากขึ้น ซึ่งจะช่วยให้ธุรกิจสามารถเข้าถึงและตอบสนองต่อความต้องการของลูกค้าได้อย่างเหมาะสม และเพิ่มศักยภาพในการสร้างรายได้จากการทำการตลาดที่ตรงกลุ่มเป้าหมายมากขึ้น

1.2 ความมุ่งหมายของงานวิจัย

ในงานวิจัยนี้ ผู้วิจัยได้ตั้งความมุ่งหมายไว้ ดังนี้

1.2.1 เพื่อศึกษาและวิเคราะห์พฤติกรรมลูกค้าโดยใช้เทคนิค RFM (Recency, Frequency, Monetary) และสร้างพีเจอรจากผลการวิเคราะห์ เพื่อใช้ในการจัดกลุ่มลูกค้า (Clustering)

1.2.2 เพื่อศึกษาการจัดกลุ่มลูกค้าโดยใช้เทคนิคการเรียนรู้ของเครื่องแบบไม่มีผู้สอน (Unsupervised Machine Learning)

1.2.3 เพื่อศึกษาตัวชี้วัด (metrics) ในการประเมินผลลัพธ์การจัดกลุ่มลูกค้า

1.2.4 เพื่อวิเคราะห์กลุ่มสินค้าตามหลักการของ ABC Analysis

1.2.5 เพื่อทำการตลาดเฉพาะกลุ่มลูกค้าด้วยเทคนิคการวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis)

1.3 ความสำคัญของงานวิจัย

งานวิจัยนี้ทำการศึกษาการวิเคราะห์พฤติกรรมลูกค้าด้วยเทคนิค RFM และจัดกลุ่มลูกค้าด้วยเทคนิคการเรียนรู้ของเครื่องแบบไม่มีผู้สอน (Unsupervised Machine Learning) จากนั้นทำการประเมินผลลัพธ์การจัดกลุ่มลูกค้า และวิเคราะห์กลุ่มสินค้าตามหลักการของ ABC Analysis สุดท้ายใช้เทคนิคการวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis) เพื่อทำการตลาดเฉพาะกลุ่มลูกค้า โดยงานวิจัยนี้ใช้ชุดข้อมูลธุรกรรมของร้านค้าปลีกออนไลน์แห่งหนึ่งจากแหล่งข้อมูลสาธารณะ Kaggle.com ประกอบด้วย 8 คุณลักษณะ และมีจำนวนข้อมูลทั้งหมด 1,067,371 รายการ

1.4 ขอบเขตของงานวิจัย

1.4.1 ประชากรที่ใช้ในการวิจัย

ข้อมูลการทำธุรกรรมของร้านค้าปลีกออนไลน์แห่งหนึ่งจากแหล่งข้อมูลสาธารณะ Kaggle.com ที่จัดเก็บข้อมูลระหว่างวันที่ 01/12/2009 ถึง 09/12/2011

1.4.2 กลุ่มตัวอย่างที่ใช้ในงานวิจัย

ข้อมูลการทำธุรกรรมที่จัดเก็บข้อมูลระหว่างวันที่ 01/12/2009 ถึง 09/12/2011 ซึ่งมีจำนวนข้อมูลทั้งหมด 1,067,371 รายการ และประกอบด้วย 8 คุณลักษณะ

1.4.3 ตัวแปรที่ศึกษา

InvoiceNo : หมายเลขใบแจ้งหนี้ เป็นเลขจำนวนเต็ม 6 หลักที่ระบุแต่ละธุรกรรมอย่างเฉพาะเจาะจง ถ้าหมายเลขนี้เริ่มต้นด้วยตัวอักษร 'c' หมายถึงการยกเลิก

StockCode : รหัสสินค้า (ผลิตภัณฑ์) เป็นเลขจำนวนเต็ม 5 หลักที่ระบุสินค้าแต่ละชนิดอย่างเฉพาะเจาะจง

Description : ชื่อสินค้า (ผลิตภัณฑ์)

Quantity : จำนวนของสินค้าแต่ละชนิดในแต่ละธุรกรรม

InvoiceDate : วันที่และเวลาของใบแจ้งหนี้

UnitPrice : ราคาสินค้าต่อหน่วย

CustomerID : รหัสของลูกค้า เป็นเลขจำนวนเต็ม 5 หลัก ที่ระบุลูกค้าแต่ละรายอย่างเฉพาะเจาะจง

Country : ชื่อประเทศ ชื่อประเทศที่ลูกค้าอาศัยอยู่

1.5 กรอบแนวคิดในงานวิจัย

งานวิจัยนี้ใช้ข้อมูลธุรกรรมของร้านค้าปลีกออนไลน์ที่ดำเนินการในสหราชอาณาจักรระหว่างวันที่ 01/12/2009 ถึง 09/12/2011 ซึ่งมีจำนวนข้อมูลทั้งหมด 1,067,371 รายการ และประกอบด้วย 8 คุณลักษณะ โดยศึกษาการจัดกลุ่มลูกค้าตามพฤติกรรมการซื้อ เพื่อสร้างกลยุทธ์การตลาดเฉพาะกลุ่ม (Targeted Marketing) ด้วยเครื่องมือ Machine Learning ในการสร้างแบบจำลอง โดยทำการแปลงข้อมูลให้เหมาะสมต่อการนำมาใช้ในการจัดกลุ่มและวิเคราะห์กลุ่มลูกค้า โดยใช้เทคนิคการสร้างฟีเจอร์ (Feature Engineering) ด้วยเทคนิค RFM เพื่อหาค่าความถี่ในการซื้อครั้งล่าสุด (Recency) ค่าความถี่ในการซื้อสินค้า (Frequency) และมูลค่ารวมของการใช้จ่าย (Monetary) ซึ่งข้อมูลเหล่านี้จะถูกนำไปสร้างฟีเจอร์ใหม่เพื่อใช้ในการจัดกลุ่มลูกค้า (Clustering) โดยใช้แบบจำลอง 3 ประเภท ได้แก่ Gaussian Mixture Model (GMM), K-Means Clustering และ Hierarchical Clustering เพื่อแบ่งกลุ่มลูกค้าตามลักษณะพฤติกรรมการซื้อที่คล้ายคลึงกัน และประเมินผลลัพธ์การจัดกลุ่มลูกค้าด้วยวิธี Silhouette Score และวิธี Davies-Bouldin Index (DBI) จากนั้นข้อมูลที่ถูกแบ่งกลุ่มแล้วจะถูกนำไปวิเคราะห์กลุ่มสินค้าตามหลักการของ ABC Analysis และใช้เทคนิคการวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis) ด้วย FP-Growth Algorithm เพื่อตรวจสอบความสัมพันธ์ระหว่างสินค้าที่ลูกค้ามักจะซื้อร่วมกัน ซึ่งประกอบด้วยค่า Support (การสนับสนุน), Confidence

(ความเชื่อมั่น), Lift (การยก) และ Conviction (การพิสูจน์) เพื่อพัฒนากลยุทธ์การตลาดเฉพาะกลุ่มที่ตอบโจทย์กลุ่มเป้าหมายได้

1.6 สมมติฐานในการวิจัย

1.6.1 การใช้เทคนิค RFM จะสามารถวิเคราะห์พฤติกรรมของลูกค้าที่มีลักษณะคล้ายกันได้ได้อย่างมีประสิทธิภาพ

1.6.2 การจัดกลุ่มลูกค้าด้วยแบบจำลอง Gaussian Mixture Model (GMM), K-Means Clustering และ Hierarchical Clustering เป็นวิธีการจัดกลุ่มลูกค้าที่มีหลักการทำงานที่แตกต่างกันตามลักษณะของข้อมูล

1.6.3 การประเมินผลลัพธ์การจัดกลุ่มลูกค้าด้วย Silhouette Score และ Davies-Bouldin Index (DBI) สามารถประเมินประสิทธิภาพการจัดกลุ่มลูกค้าตามลักษณะที่ต้องการได้

1.6.4 การจัดกลุ่มลูกค้าตามพฤติกรรมการซื้อและการวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis) สามารถช่วยสร้างกลยุทธ์การตลาดเฉพาะกลุ่มเพื่อเพิ่มยอดขายและความพึงพอใจของลูกค้า

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการวิจัยครั้งนี้ ผู้วิจัยได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง และได้นำเสนอตามหัวข้อต่อไปนี้

1. การแบ่งกลุ่มลูกค้า (Customer Segmentation)
2. กระบวนการในการสร้างหรือปรับปรุงคุณลักษณะ (Feature Engineering)
3. การวิเคราะห์กลุ่มสินค้าด้วยหลักการ ABC Analysis
4. การวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis)
5. แบบจำลองที่ใช้ในงานวิจัย
6. การประเมินผลแบบจำลอง
7. งานวิจัยที่เกี่ยวข้อง

2.1 การแบ่งกลุ่มลูกค้า (Customer Segmentation)

การแบ่งกลุ่มลูกค้า (Customer Segmentation) เป็นกระบวนการในการจำแนกลูกค้าออกเป็นกลุ่มย่อยต่าง ๆ โดยพิจารณาจากลักษณะร่วมที่มีความคล้ายคลึงกันภายในกลุ่ม และมีความแตกต่างกันระหว่างกลุ่ม การแบ่งกลุ่มลูกค้าอย่างมีประสิทธิภาพจะช่วยให้ธุรกิจเข้าใจความต้องการลูกค้าเฉพาะกลุ่ม ซึ่งนำไปสู่การพัฒนาผลิตภัณฑ์และบริการที่ตรงตามความต้องการเหล่านั้น และสร้างความได้เปรียบในการแข่งขันทางธุรกิจได้อย่างยั่งยืน โดยหลักเกณฑ์ในการแบ่งกลุ่มลูกค้ามี 4 ประเภท ได้แก่

- เกณฑ์ด้านประชากรศาสตร์ (Demographic Segmentation) เช่น อายุ, เพศ, การศึกษา, อาชีพ, สถานภาพสมรส เป็นต้น
- เกณฑ์ด้านภูมิศาสตร์ (Geographic Segmentation) เช่น ประเทศ, ภูมิภาค, จังหวัด, เขตที่อยู่อาศัย, สภาพภูมิอากาศ เป็นต้น
- เกณฑ์ด้านจิตวิทยา (Psychographic Segmentation) เช่น รูปแบบการดำเนินชีวิต (Lifestyle), บุคลิกภาพ (Personality), ค่านิยม (Values),ทัศนคติ (Attitudes), ความสนใจ (Interests) เป็นต้น
- เกณฑ์ด้านพฤติกรรม (Behavioral Segmentation) เช่น ความถี่ในการซื้อ, มูลค่าการซื้อ, ความภักดีต่อแบรนด์, โอกาสในการซื้อ เป็นต้น

ในงานวิจัยนี้จะใช้เกณฑ์ด้านพฤติกรรม (Behavioral Segmentation) โดยใช้เทคนิคการจัดกลุ่มลูกค้า ดังนี้

2.1.1 การแบ่งกลุ่มลูกค้าตามหลักการ RFM

RFM Analysis เป็นเทคนิคที่ใช้ในการวิเคราะห์เพื่อทำความเข้าใจพฤติกรรมของลูกค้าโดยพิจารณาจากข้อมูลที่ผ่านมาของลูกค้าแต่ละราย เพื่อให้ธุรกิจสามารถกำหนดกลยุทธ์การตลาดที่เหมาะสมตามกลุ่มลูกค้าได้ โดยพิจารณาจาก 3 ปัจจัย ได้แก่

- Recency (R) – การวัดระยะเวลาตั้งแต่ครั้งล่าสุดที่ลูกค้าทำการซื้อ โดยคำนวณจากวันที่ปัจจุบันลบด้วยวันที่ลูกค้าซื้อครั้งล่าสุด ยิ่งลูกค้าทำการซื้อเมื่อไม่นานนี้มากเท่าไร ยิ่งมีโอกาสสูงที่จะซื้ออีกในอนาคต ลูกค้าที่มีค่า Recency ต่ำ (เพิ่งซื้อไปไม่นาน) มีแนวโน้มที่จะตอบสนองต่อแคมเปญการตลาดได้ดีกว่าลูกค้าที่มีค่า Recency สูง (ไม่ได้ซื้อเป็นเวลานานแล้ว)

- Frequency (F) – การวัดความถี่ในการซื้อสินค้า โดยคำนวณจากจำนวนครั้งที่ลูกค้าทำการซื้อในช่วงเวลาที่กำหนด ยิ่งลูกค้าทำการซื้อบ่อยครั้งมากเท่าไร ยิ่งแสดงถึงความสัมพันธ์ที่ดีระหว่างลูกค้ากับธุรกิจ ลูกค้าที่ซื้อซ้ำบ่อย ๆ อาจหมายถึงลูกค้าที่มีความภักดีต่อแบรนด์และมีแนวโน้มสูงที่จะทำการซื้อในอนาคต

- Monetary (M) - การวัดมูลค่ารวมของการซื้อ โดยคำนวณจากผลรวมของมูลค่าการซื้อทั้งหมดที่ลูกค้าทำไว้ในช่วงเวลาที่กำหนด ยิ่งลูกค้ามีมูลค่าการซื้อรวมสูงมากเท่าไร ยิ่งแสดงถึงความสำคัญของลูกค้าต่อธุรกิจ ลูกค้าที่มียอดการซื้อสูงมักจะเป็นกลุ่มที่ธุรกิจควรให้ความสำคัญและดูแลเป็นพิเศษ

2.1.2 การจัดกลุ่มลูกค้าด้วยแบบจำลอง (Customer Clustering)

การจัดกลุ่มลูกค้า (Customer Clustering) เป็นวิธีการแบ่งกลุ่มลูกค้าตามลักษณะพฤติกรรมหรือคุณลักษณะที่คล้ายคลึงกัน เพื่อช่วยให้ธุรกิจสามารถเข้าใจพฤติกรรมของลูกค้าที่คล้ายคลึงกันมากขึ้น ซึ่งสามารถช่วยในการกำหนดกลยุทธ์การตลาด การพัฒนาสินค้า และการบริการที่ตรงกับความต้องการของลูกค้าแต่ละกลุ่ม โดยการจัดกลุ่มลูกค้าด้วยแบบจำลอง (Clustering Models) สามารถแบ่งออกเป็น 2 ประเภทหลัก ดังนี้

2.1.2.1 การจัดกลุ่มลูกค้าด้วยการเรียนรู้ของเครื่องแบบมีผู้สอน (Supervised Machine Learning)

เป็นการใช้ข้อมูลที่มีการระบุประเภทของกลุ่มลูกค้าไว้แล้ว เพื่อฝึกแบบจำลองให้สามารถระบุว่าคุณค่าคนใดควรอยู่ในกลุ่มใด ตัวอย่างเช่น Logistic Regression Decision Trees ซึ่งเป็นเทคนิคที่ใช้ข้อมูลที่มีการระบุประเภท (Labeled Data) เพื่อตัดสินใจว่าลูกค้าแต่ละคนควรจะถูกจัดให้อยู่ในกลุ่มใดตามคุณลักษณะที่วัดได้

2.1.2.2 การจัดกลุ่มลูกค้าด้วยการเรียนรู้ของเครื่องแบบไม่มีผู้สอน (Unsupervised Machine Learning) เป็นการจัดกลุ่มลูกค้าที่ใช้ข้อมูลที่ไม่มีการติดป้ายประเภทล่วงหน้า (Unlabeled Data) โดยอัลกอริทึมจะค้นหารูปแบบหรือโครงสร้างของข้อมูลที่คล้ายคลึงกัน เทคนิคการเรียนรู้แบบไม่มีผู้สอนเหมาะกับการวิเคราะห์ข้อมูลขนาดใหญ่ที่ยังไม่ทราบถึงประเภทของกลุ่มล่วงหน้า ตัวอย่างเช่น Gaussian Mixture Models (GMM), K-Means Clustering, Hierarchical Clustering เป็นต้น

2.2 กระบวนการในการสร้างหรือปรับปรุงคุณลักษณะ (Feature Engineering)

กระบวนการในการสร้างหรือปรับปรุงคุณลักษณะ (Feature Engineering) จากข้อมูลดิบที่เก็บมาอาจจะไม่พร้อมสำหรับการใช้ในแบบจำลอง ดังนั้นการสร้างหรือปรับปรุงคุณลักษณะจึงเป็นขั้นตอนสำคัญในการเตรียมข้อมูลให้พร้อมและเหมาะสม เพื่อช่วยให้แบบจำลองสามารถเรียนรู้จากข้อมูลได้มีประสิทธิภาพมากขึ้นและช่วยให้ผลลัพธ์แม่นยำมากขึ้น โดยการทำ Feature Engineering สามารถแบ่งได้เป็น 4 ประเภท ดังนี้

2.2.1 การจัดการกับค่าที่ขาดหาย (Handling Missing Data)

ในการทำงานกับข้อมูลจริง มักพบว่ามีความขาดหายบางส่วนขาดหายไป ดังนั้นการจัดการกับค่าที่ขาดหายเป็นสิ่งสำคัญเพื่อป้องกันไม่ให้แบบจำลองทำงานผิดพลาด การจัดการสามารถทำได้หลายวิธี เช่น

- การลบค่าที่ขาดหายไป ลบแถวหรือคอลัมน์ที่มีค่าขาดหายออกไปทั้งหมด เหมาะสำหรับกรณีที่ข้อมูลขาดหายไม่มาก
- การแทนที่ด้วยค่าทางสถิติ ใช้ค่าเฉลี่ย (Mean), มัธยฐาน (Median) หรือฐานนิยม (Mode) แทนที่ค่าที่ขาดหาย
- การประมาณค่า (Imputation) ใช้แบบจำลองทางสถิติหรือ Machine Learning เพื่อประมาณค่าที่ขาดหายไปตามความสัมพันธ์ของข้อมูลอื่น ๆ

2.2.2 การแปลงประเภทข้อมูล (Data Transformation)

ข้อมูลที่เราอาจจะอยู่ในรูปแบบที่ไม่เหมาะสมกับการเรียนรู้ของเครื่อง เช่น ข้อมูลหมวดหมู่ (Categorical Data) หรือข้อมูลที่มีการกระจายตัวไม่เหมาะสม จึงจำเป็นต้องทำการแปลงข้อมูลให้เหมาะสมมากขึ้น ตัวอย่างเช่น

- การเข้ารหัสหมวดหมู่ (Categorical Encoding) สำหรับข้อมูลที่เป็นหมวดหมู่ เช่น เพศ (ชาย/หญิง) หรือประเทศ สามารถแปลงเป็นค่าตัวเลขได้ เช่น การเข้ารหัสแบบ One-Hot Encoding หรือการใช้ Label Encoding
- การแปลงทางคณิตศาสตร์ ใช้การแปลงข้อมูล เช่น การสเกลข้อมูล (Scaling), การใช้ Log Transformation เพื่อให้ข้อมูลมีการกระจายตัวที่สมดุลขึ้น

2.2.3 การสร้างคุณลักษณะใหม่ (Feature Creation)

การสร้างคุณลักษณะใหม่จากข้อมูลเดิมสามารถเพิ่มประสิทธิภาพในการทำนายได้ โดยการรวมข้อมูลหลายคอลัมน์เข้าด้วยกันหรือสร้างคุณลักษณะใหม่ที่มีความสำคัญต่อปัญหาที่กำลังวิเคราะห์

- การรวมฟีเจอร์ (Feature Combination) รวมฟีเจอร์ที่มีอยู่ เช่น การรวมราคากับจำนวนสินค้าที่ซื้อเพื่อสร้างฟีเจอร์ "ยอดรวมการซื้อ"
- การสร้างคุณลักษณะตามเวลาหรือวัน สร้างฟีเจอร์ใหม่ที่เกี่ยวข้องกับวันในสัปดาห์หรือช่วงเวลาทำการซื้อสินค้า เพื่อดูว่ามีพฤติกรรมการซื้อเปลี่ยนแปลงไปตามเวลา

2.2.4 การเลือกคุณลักษณะ (Feature Selection)

การเลือกฟีเจอร์ที่มีความสัมพันธ์กับตัวแปรเป้าหมายเพื่อปรับปรุงประสิทธิภาพของแบบจำลองและลดความซับซ้อนของข้อมูล การเลือกฟีเจอร์ทำได้หลายวิธี เช่น

- การใช้ตัววัดความสัมพันธ์ (Correlation) ดูความสัมพันธ์ระหว่างฟีเจอร์กับตัวแปรเป้าหมาย เช่น การใช้ Pearson Correlation หรือการทดสอบ Chi-Square
- การใช้เทคนิคอัตโนมัติ (Automated Feature Selection) เช่น การใช้ Recursive Feature Elimination (RFE) หรือการวัดความสำคัญของฟีเจอร์ผ่านการใช้แบบจำลอง

2.3 การวิเคราะห์กลุ่มสินค้าด้วยหลักการ ABC (ABC Analysis)

การวิเคราะห์กลุ่มสินค้าด้วยหลักการ ABC (ABC Analysis) เป็นเครื่องมือในการบริหารจัดการสินค้าคงคลังที่มีประสิทธิภาพ โดยประยุกต์ใช้หลักการของ Pareto (Pareto Principle) หรือ "กฎ 80/20" มาใช้ในการจำแนกประเภทสินค้าตามระดับความสำคัญ เพื่อให้การบริหารจัดการทรัพยากรเป็นไปอย่างมีประสิทธิภาพสูงสุด โดยการวิเคราะห์แบบ ABC พิจารณาจากมูลค่าและปริมาณของสินค้า

- มูลค่าการขาย (Annual Consumption Value) สำหรับแต่ละสินค้า

$$\text{Annual Consumption Value} = \text{จำนวนขายต่อปี} \times \text{ราคาต่อหน่วย}$$

- เปอร์เซ็นต์การสะสมของมูลค่าการขาย (Cumulative Percentage)

$$\text{Cumulative Percentage} = \frac{\text{มูลค่าการขายของสินค้า}}{\text{มูลค่าการขายรวม}} \times 100$$

- กลุ่ม A (สินค้าที่มีความสำคัญสูง) สินค้ากลุ่มนี้มักมีมูลค่าการขายสูง หรือมีผลกระทบสำคัญต่อรายได้ของธุรกิจ (10-20% ของจำนวนรายการสินค้าทั้งหมด) คิดเป็น 70-80% ของมูลค่ารวมทั้งหมด
- กลุ่ม B (สินค้าที่มีความสำคัญปานกลาง) สินค้ากลุ่มนี้มีจำนวนสินค้ามากกว่ากลุ่ม A (20-30% ของจำนวนรายการสินค้าทั้งหมด) และมีมูลค่าในระดับกลาง คิดเป็น 15-20% ของมูลค่ารวมทั้งหมด
- กลุ่ม C (สินค้าที่มีความสำคัญต่ำ) สินค้ากลุ่มนี้มีจำนวนมากที่สุด (50-70% ของจำนวนรายการสินค้าทั้งหมด) และมีมูลค่าการขายที่ต่ำ คิดเป็น 5-10% ของมูลค่ารวมทั้งหมด

2.4 การวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis)

การวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis) เป็นเทคนิคในการวิเคราะห์พฤติกรรม การซื้อของลูกค้า ซึ่งช่วยให้ธุรกิจเข้าใจว่าลูกค้ามักซื้อสินค้าใดร่วมกันบ่อยครั้ง โดยการวิเคราะห์นี้สามารถนำไปใช้ในการกำหนดกลยุทธ์การตลาด การจัดโปรโมชั่น การปรับปรุงสินค้าที่นำเสนอ และการจัดสต็อกสินค้าให้มีประสิทธิภาพมากยิ่งขึ้น โดยอิงตามแนวคิดของ Association Rule Learning ซึ่งใช้ข้อมูลจากการซื้อในอดีตเพื่อนำมาวิเคราะห์เพื่อค้นหากฎความสัมพันธ์ระหว่างสินค้าในตะกร้าสินค้าที่ลูกค้าซื้อ ในการวิเคราะห์ตะกร้าสินค้านี้ อัลกอริทึมหลักที่ใช้ในการวิเคราะห์ได้แก่

- Apriori Algorithm ใช้ในการค้นหากฎความสัมพันธ์ โดยมุ่งเน้นไปที่การค้นหาสินค้าที่มักถูกซื้อร่วมกันบ่อยที่สุดในตะกร้าสินค้า อัลกอริทึมนี้ใช้หลักการ "การตัดสินใจ" เริ่มต้นด้วยการหาสินค้าที่มีการขายดีที่สุด (frequent itemsets) และสร้างกฎที่เกี่ยวข้องจากข้อมูลที่ได้ ข้อดีของ Apriori คือสามารถจัดการกับข้อมูลขนาดใหญ่ได้มีประสิทธิภาพ แต่มีข้อจำกัดในเรื่องเวลาในการประมวลผลเมื่อข้อมูลมีปริมาณมาก

- FP-Growth Algorithm ถูกพัฒนาขึ้นมาเพื่อแก้ไขปัญหาที่เกิดจากการใช้ Apriori Algorithm โดยไม่ต้องสร้าง candidates หลายชุด ซึ่งทำให้ประสิทธิภาพลดลง แทนที่จะสร้าง itemsets FP-Growth จะสร้าง frequent pattern tree (FP-tree) ที่ทำให้การค้นหากฎความสัมพันธ์ระหว่างสินค้าทำได้อย่างรวดเร็วและมีประสิทธิภาพมากขึ้น อัลกอริทึมนี้เหมาะสำหรับการ

วิเคราะห์ข้อมูลขนาดใหญ่และมีความซับซ้อน ทำให้สามารถจัดการกับข้อมูลได้มีประสิทธิภาพมากยิ่งขึ้น

- Eclat Algorithm หรือ Equivalence Class Transformation ใช้ในการค้นหาความสัมพันธ์ระหว่างสินค้าคล้ายกับ Apriori Algorithm แต่แตกต่างกันตรงที่ Eclat ใช้แนวทางการคำนวณแบบ intersection ของ transaction list ซึ่งช่วยให้สามารถค้นพบความสัมพันธ์ได้รวดเร็วและมีประสิทธิภาพ โดยเฉพาะเมื่อต้องจัดการกับข้อมูลจำนวนมาก อัลกอริทึมนี้เหมาะสำหรับกรณีที่ต้องการผลลัพธ์ที่รวดเร็วและแม่นยำในการวิเคราะห์ข้อมูล

การวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis) มักจะใช้ตัวแปรหลัก ๆ ในการคำนวณและวิเคราะห์ความสัมพันธ์ระหว่างสินค้าซึ่งสามารถนำไปใช้ในการสร้างกฎความสัมพันธ์ได้ โดยตัวแปรพร้อมสูตรการคำนวณมีดังนี้

- Support แสดงถึงความถี่ที่กลุ่มสินค้าถูกซื้อร่วมกันในข้อมูลทั้งหมด

$$Support(A) = \frac{\text{จำนวนการซื้อที่มี } A}{\text{จำนวนการซื้อทั้งหมด}}$$

- Confidence แสดงถึงความน่าจะเป็นที่ลูกค้าจะซื้อสินค้าหนึ่งเมื่อซื้ออีกสินค้าหนึ่ง

$$Confidence(A \rightarrow B) = \frac{Support(A, B)}{Support(A)}$$

- Lift แสดงถึงความสัมพันธ์ระหว่างสองสินค้าว่ามีความสัมพันธ์มากกว่าการสุ่มซื้อหรือไม่

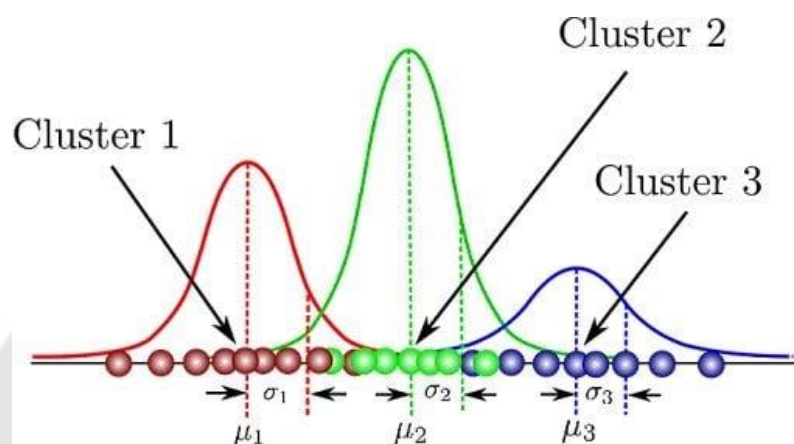
$$Lift(A, B) = \frac{Confidence(A \rightarrow B)}{Support(B)}$$

- Association Rules (กฎความสัมพันธ์) เป็นแนวทางในการวิเคราะห์ข้อมูลที่ใช้เพื่อค้นหาความสัมพันธ์ระหว่างสินค้าหรือเหตุการณ์ที่เกิดขึ้นร่วมกันในชุดข้อมูล โดยกฎนี้ประกอบไปด้วย 2 ส่วน คือ Antecedents คือสินค้าหรือกลุ่มสินค้าที่มีการซื้อหรือเกิดขึ้นก่อนในกฎความสัมพันธ์ และ Consequents คือสินค้าหรือกลุ่มสินค้าที่มีการซื้อหรือเกิดขึ้นหลังจากส่วนหน้าของกฎความสัมพันธ์

2.5 แบบจำลองที่ใช้งานวิจัย

2.5.1 Gaussian Mixture Model (GMM)

Gaussian Mixture Model (GMM) เป็นแบบจำลองทางสถิติที่ใช้สำหรับการวิเคราะห์ข้อมูลเชิงหลายมิติ โดยสมมติว่าข้อมูลใน dataset มาจากการผสมของ Gaussian distributions หลายตัว (Mixture of Gaussians) ซึ่งแต่ละตัวแทนกลุ่ม (cluster) ที่มีความน่าจะเป็นเฉพาะตัว เทคนิคนี้เหมาะสำหรับการจัดกลุ่มข้อมูลที่มีลักษณะซับซ้อน เช่น การกระจายตัวแบบไม่สมมาตร หรือมีหลายโหนด (peaks) ตามภาพประกอบ 1 โดยมีกระบวนการทำงาน ดังนี้



ภาพประกอบ 1 แสดงรูปแบบการจัดกลุ่มของ Gaussian Mixture Model

ที่ ม ๑ Gaussian Mixture Model Explained, Retrieved March 15, 2025, from <https://builtin.com/articles/gaussian-mixture-model>

ขั้นตอนที่ 1 สมมติว่าข้อมูลใน dataset ถูกสร้างขึ้นจาก Gaussian distributions หลายตัว ($k=1,2,\dots,K$) โดยแต่ละ Gaussian distribution ถูกกำหนดด้วยค่าเฉลี่ย (μ) และค่าโควาเรียนซ์ (Σ) และข้อมูลแต่ละจุด (x) มีความน่าจะเป็นที่จะมาจากแต่ละ cluster (p) และมีมิติข้อมูล (d)

$$N(x|\mu_k, \Sigma_k) = \frac{1}{(2\pi)^{d/2} |\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k)\right)$$

ขั้นตอนที่ 2 การผสมผสานของ Gaussians กำหนดให้ K คือจำนวน Gaussian distributions (หรือ clusters) และ π คือน้ำหนัก (weights) ของแต่ละ Gaussian distribution

$$P(x) = \sum_{k=1}^K \pi_k N(x|\mu_k, \Sigma_k)$$

ขั้นตอนที่ 3 ประมาณค่าพารามิเตอร์ โดยใช้ Expectation-Maximization (EM) algorithm

- E-step คำนวณความน่าจะเป็นที่ข้อมูล x จะอยู่ใน cluster k

$$r_{ik} = \frac{\pi_k N(x_i | \mu_k, \Sigma_k)}{\sum_{j=1}^K \pi_j N(x_i | \mu_j, \Sigma_j)}$$

- M-step: ปรับปรุงค่าพารามิเตอร์ใหม่

$$\mu_k = \frac{\sum_{i=1}^n r_{ik} x_i}{\sum_{i=1}^n r_{ik}}, \Sigma_k = \frac{\sum_{i=1}^n r_{ik} (x_i - \mu_k)(x_i - \mu_k)^T}{\sum_{i=1}^n r_{ik}}, \pi_k = \frac{\sum_{i=1}^n r_{ik}}{n}$$

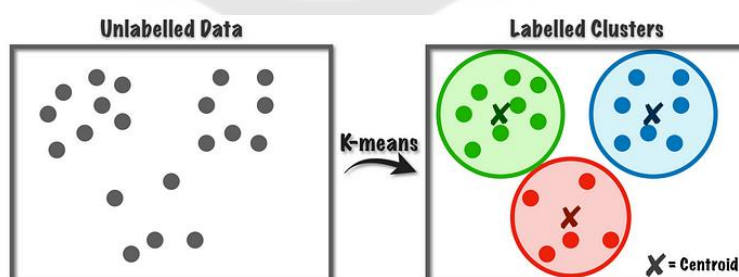
ขั้นตอนที่ 4 เลือกจำนวน cluster ที่เหมาะสม (K) ด้วย Bayesian Information Criterion (BIC) หรือ Akaike Information Criterion (AIC) โดยกำหนดให้ k คือ จำนวนพารามิเตอร์ n คือ จำนวนข้อมูล และ L คือ Likelihood

$$BIC = k \log(n) - 2 \log(L)$$

$$AIC = 2k - 2 \log(L)$$

2.5.2 K-means Clustering

K-means Clustering เป็นอัลกอริทึมที่นิยมใช้ในการจัดกลุ่มข้อมูลที่ต้องการแบ่งข้อมูลออกเป็น K กลุ่ม โดยการทำงานของแบบจำลองนี้จะเริ่มต้นด้วยการสุ่มเลือก K centroid และจัดกลุ่มข้อมูลโดยการคำนวณระยะห่างจากจุดข้อมูลไปยัง centroid ที่ใกล้ที่สุด และอัปเดต centroid โดยการคำนวณค่าเฉลี่ยของจุดข้อมูลในแต่ละกลุ่ม โดยจะทำซ้ำขั้นตอนการจัดกลุ่มและการอัปเดต centroid จนกว่าจะไม่มีเปลี่ยนแปลงในกลุ่ม ตามภาพประกอบ 2 โดยมีกระบวนการทำงาน ดังนี้



ภาพประกอบ 2 แสดงรูปแบบการจัดกลุ่มของ K-means Clustering

ที่ ม ๑ Understanding how K-Means Clustering Works (A detailed guide), Retrieved March 15, 2025, from <https://levelup.gitconnected.com/understanding-how-k-means-clustering-works-a-detailed-guide-9a2f8009a279>

ขั้นตอนที่ 1 Initialization : สุ่ม K จุดศูนย์กลาง (centroids) เริ่มต้น

ขั้นตอนที่ 2 Assignment : จัดกลุ่มข้อมูล โดยคำนวณระยะห่างระหว่างจุดข้อมูลกับ centroid และกำหนดจุดข้อมูลให้อยู่ในกลุ่มที่ใกล้ที่สุด

$$d(x, c_k) = \sqrt{\sum_{i=1}^n (x_i - c_{ki})^2}$$

ขั้นตอนที่ 3 Update : คำนวณค่า centroid ใหม่สำหรับแต่ละกลุ่ม โดยใช้ค่าเฉลี่ยของจุดข้อมูลในกลุ่มนั้น

$$c_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_i$$

ขั้นตอนที่ 4 Repeat : ทำซ้ำขั้นตอน Assignment และ Update จนกว่า centroid จะไม่เปลี่ยนแปลง

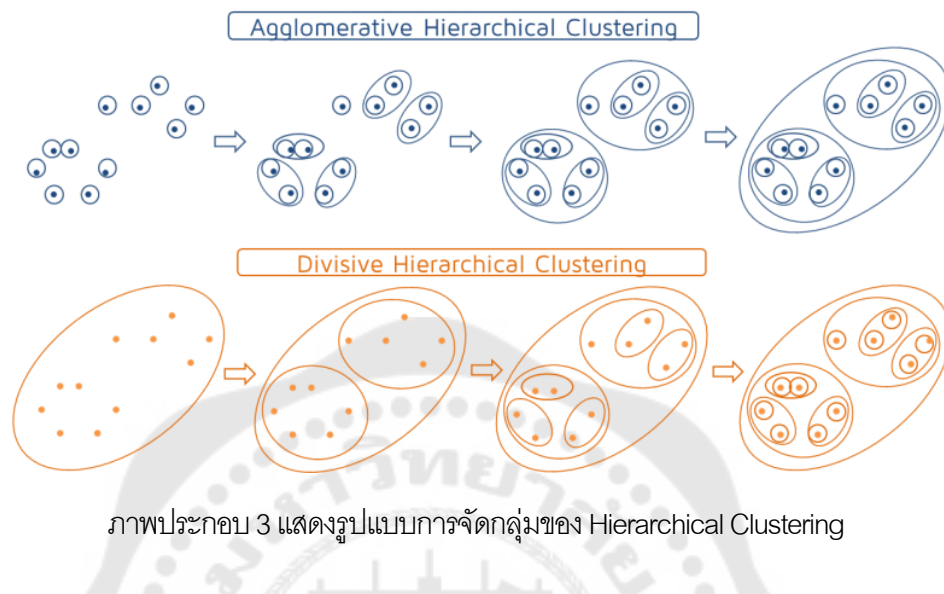
ขั้นตอนที่ 5 Elbow Method : เลือกจำนวน K ที่เหมาะสมโดยดูจากจุดที่ค่า Within-cluster Sum of Squares (WCSS) โดย WCSS จะลดลงเมื่อ K เพิ่มขึ้น และจุดที่กราฟเริ่มชะลอการลดลง (elbow point) บ่งบอกถึงจำนวน K ที่เหมาะสม

$$WCSS = \sum_{k=1}^K \sum_{i \in C_k} \|x_i - c_k\|^2$$

2.5.3 Hierarchical Clustering

Hierarchical Clustering เป็นเทคนิคการเรียนรู้ของเครื่องแบบไม่มีผู้สอนที่ใช้สำหรับการจัดกลุ่ม (clustering) ข้อมูล โดยจะสร้างโครงสร้างของกลุ่มในรูปแบบของต้นไม้ (dendrogram) ซึ่งแสดงถึงความสัมพันธ์ระหว่างกลุ่มต่าง ๆ ของข้อมูล เมื่อสร้าง dendrogram แล้ว สามารถกำหนดจำนวนกลุ่ม (clusters) ที่ต้องการได้ โดยการตัด dendrogram ที่ระดับความสูงที่เหมาะสม ซึ่งจะทำให้สามารถแยกข้อมูลออกเป็นกลุ่มที่มีความสัมพันธ์กันได้ วิธีการนี้สามารถแบ่งออกเป็นสองประเภทหลัก คือ Agglomerative (รวมกลุ่มจากจุดข้อมูลที่ละคู่) และ Divisive (แบ่งกลุ่มใหญ่เป็นกลุ่มย่อยลงไปเรื่อย ๆ) การทำงานของ Hierarchical Clustering จะพิจารณาความคล้ายคลึงกัน (similarity) หรือความแตกต่าง (dissimilarity) ระหว่างจุดข้อมูล โดยใช้มาตรฐานระยะทาง (distance metric) เช่น Euclidean Distance หรือ Manhattan Distance

ในการคำนวณความใกล้เคียงของข้อมูล ทำให้สามารถสร้างกลุ่มที่มีลักษณะคล้ายกัน ตามภาพประกอบ 3 โดยมีกระบวนการทำงาน ดังนี้



ภาพประกอบ 3 แสดงรูปแบบการจัดกลุ่มของ Hierarchical Clustering

ที่มา Hierarchical clustering, using it to invest, Retrieved March 15, 2025, from <https://quantdare.com/hierarchical-clustering/>

ขั้นตอนที่ 1 Initialization : ให้ข้อมูลแต่ละจุดเป็นกลุ่มของตัวเอง

ขั้นตอนที่ 2 Calculate Distance : คำนวณระยะห่างระหว่างกลุ่ม

- Single Linkage : ระยะทางที่ใกล้ที่สุดระหว่างจุดข้อมูลในสองกลุ่ม

$$D(A, B) = \min_{a \in A, b \in B} d(a, b)$$

- Complete Linkage : ระยะทางที่ใกล้ที่สุดระหว่างจุดข้อมูลในสองกลุ่ม

$$D(A, B) = \max_{a \in A, b \in B} d(a, b)$$

- Average Linkage : ค่าเฉลี่ยของระยะทางทั้งหมดระหว่างจุดข้อมูลในสองกลุ่ม

$$D(A, B) = \frac{1}{|A| \cdot |B|} \sum_{a \in A, b \in B} d(a, b)$$

ขั้นตอนที่ 3 Merge : รวมกลุ่มที่มีระยะห่างน้อยที่สุด

ขั้นตอนที่ 4 Repeat : ทำซ้ำจนเหลือกลุ่มเดียว

ขั้นตอนที่ 5 Cut the Dendrogram : กำหนดจำนวนกลุ่มโดยการตัด dendrogram

2.6 การประเมินผลแบบจำลอง

2.6.1 Silhouette Score

Silhouette Score เป็นตัวชี้วัดที่ช่วยในการประเมินความสามารถในการจัดกลุ่ม โดยคำนวณความคล้ายคลึงกันระหว่างจุดข้อมูลภายในกลุ่มและจุดข้อมูลในกลุ่มอื่น ๆ โดยมีค่าอยู่ระหว่าง -1 ถึง 1

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

- ค่าใกล้ 1 ข้อมูลในกลุ่มมีความใกล้เคียงกันสูง และอยู่ห่างจากกลุ่มอื่น ๆ
- ค่าใกล้ 0 ข้อมูลอยู่ในบริเวณขอบเขตระหว่างกลุ่ม
- ค่าใกล้ -1 ข้อมูลอาจจะถูกจัดกลุ่มในกลุ่มที่ไม่ถูกต้อง

2.6.2 Davies-Bouldin Index (DBI)

Davies-Bouldin Index (DBI) เป็นตัวชี้วัดที่คุณภาพของการจัดกลุ่มในเชิงสัมพัทธ์ ระหว่างกลุ่ม โดยการวัดความแยก (separation) และความแน่นหนา (compactness) ของกลุ่ม

$$DBI = \frac{1}{K} \sum_{i=1}^K \max_{j \neq i} (R_{ij})$$

- ค่าต่ำหมายความว่ากลุ่มมีความแน่นหนาสูงและแยกจากกันได้ดี
- ค่าที่สูงหมายความว่ามีการมีกลุ่มที่ไม่ชัดเจนหรือมีความทับซ้อนกัน

2.7 งานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้ศึกษาบทความวิจัยที่เกี่ยวข้องกับการจัดกลุ่มลูกค้าด้วยเทคนิคและแบบจำลองต่าง ๆ เพื่อศึกษาวิธีการและแนวทางเพื่อดำเนินการวิจัยดังกล่าว โดยได้ศึกษาจากบทความวิจัยดังนี้

2.7.1 Customer Segmentation by Using RFM Model and Clustering Methods : A Case Study in Retail Industry (Doğan และคณะ, 2018)

งานวิจัยนี้นำเสนอการวิเคราะห์การแบ่งกลุ่มลูกค้าในอุตสาหกรรมค้าปลีกกีฬาโดยใช้เทคนิค RFM (Recency, Frequency, Monetary) ร่วมกับแบบจำลองการจัดกลุ่ม เพื่อปรับปรุงระบบบัตรสมาชิกที่มีอยู่เดิมซึ่งแบ่งกลุ่มลูกค้าเพียงตามยอดการใช้จ่าย (Monetary) เท่านั้น โดยนำปัจจัยด้านความถี่ในการซื้อ (Frequency) และระยะเวลาการซื้อครั้งล่าสุด (Recency)

มาพิจารณาร่วมกับยอดขายง่าย เพื่อให้เข้าใจพฤติกรรมลูกค้าได้ดียิ่งขึ้นและสามารถพัฒนากลยุทธ์การตลาดที่เหมาะสมกับแต่ละกลุ่ม โดยใช้เทคนิค RFM Analysis ร่วมกับแบบจำลอง Two-step Clustering Analysis และ K-means Clustering Analysis

การดำเนินงานวิจัยนี้ได้เก็บรวบรวมข้อมูลลูกค้าจากฐานข้อมูลบัตรสมาชิกของบริษัทค้าปลีกกีฬาในประเทศตุรกี ระยะเวลาข้อมูลตั้งแต่ 1 มกราคม 2016 ถึง 31 ธันวาคม 2016 จำนวน 700,032 ราย โดยเริ่มต้นจากการวิเคราะห์ RFM เพื่อกำหนดค่า Recency, Frequency และ Monetary และใช้แบบจำลอง Two-step Clustering กับ K-means Clustering เพื่อทำการจัดกลุ่มลูกค้า และเปรียบเทียบผลการจัดกลุ่มลูกค้า

ผลการวิจัยนี้จะทำการเปรียบเทียบกับข้อมูลสมาชิกเดิมที่มี Bronze 694,647 ราย Gold 4,469 ราย และ Premium 916 ราย โดยแบบจำลอง Two-step Clustering สามารถแบ่งลูกค้าเป็น 3 กลุ่ม ได้แก่

- Bronze (40% หรือ 279,717 ราย): มีค่า $R=193.23$, $F=1.49$, $M=261.18$
- Gold (53.9% หรือ 377,379 ราย): มีค่า $R=72.31$, $F=1.57$, $M=282.07$
- Premium (6.1% หรือ 42,936 ราย): มีค่า $R=53.25$, $F=7.49$, $M=1,308.18$

โดยเมื่อเปรียบเทียบข้อมูลสมาชิกเดิมกับผลการแบ่งกลุ่มจากแบบจำลอง Two-step Clustering พบว่า

- สมาชิก Bronze จำนวน 377,379 ราย ควรได้รับการปรับเป็น Gold
- สมาชิก Bronze จำนวน 37,551 ราย ควรได้รับการปรับเป็น Premium
- สมาชิกทุกรายใน Gold จำนวน 4,469 ราย ควรได้รับการปรับเป็น Premium
- สมาชิก Premium เดิมยังคงอยู่ในระดับ Premium

ต่อมา แบบจำลอง K-means Clustering สามารถแบ่งลูกค้าเป็น 4 กลุ่ม ดังนี้

- Regular (92% หรือ 644,081 ราย): มีค่า $R=120.16$, $F=1.12$, $M=327.2$
- Loyal (514 ราย): มีค่า $R=88.5$, $F=2.63$, $M=719.2$
- Star (97 ราย): มีค่า $R=54.1$, $F=6.03$, $M=2,823.2$
- Advanced (55,340 ราย): มีค่า $R=111.7$, $F=2.01$, $M=439.1$

โดยเมื่อเปรียบเทียบข้อมูลสมาชิกเดิมกับผลการแบ่งกลุ่มจากแบบจำลอง K-means Clustering พบว่า

- กลุ่ม Regular ประกอบด้วยสมาชิก Bronze ทั้งหมด

- กลุ่ม Advanced ประกอบด้วยสมาชิก Bronze จำนวน 50,566 ราย, Gold จำนวน 4,469 ราย และ Premium จำนวน 305 ราย

- กลุ่ม Loyal และ Star ส่วนใหญ่มาจากสมาชิก Premium เดิม

เมื่อเปรียบเทียบกับข้อมูลสมาชิกเดิม พบว่ามีความแตกต่างถึง 60% เนื่องจากพิจารณาปัจจัยด้านความถี่และความเร็วในการกลับมาซื้อพร้อมด้วย และเห็นว่าการแบ่งกลุ่มโดยใช้เพียงยอดใช้จ่าย (M) อย่างเดียวไม่เพียงพอ การพิจารณาความถี่ (F) และความเร็วในการกลับมาซื้อ (R) ทำให้เข้าใจพฤติกรรมลูกค้าได้ดีขึ้น และสามารถจัดกลุ่มลูกค้าได้เหมาะสมกว่า

2.7.2 Analysis Of Increasing Student Service Satisfaction Using K-Means Clustering Algorithm and Gaussian Mixture Models (GMM) (Maulidya และคณะ, 2024)

งานวิจัยนี้นำเสนอการวิเคราะห์เปรียบเทียบระหว่างแบบจำลอง K-Means Clustering และ Gaussian Mixture Model (GMM) เพื่อวิเคราะห์ความพึงพอใจในการให้บริการของนักศึกษา โดยมุ่งเน้นการทำความเข้าใจโครงสร้างข้อมูลและความเหมาะสมของแบบจำลอง เพื่อช่วยให้สถาบันการศึกษาปรับปรุงบริการให้ตรงกับความต้องการของนักศึกษาได้อย่างมีประสิทธิภาพมากขึ้น

การดำเนินงานวิจัยนี้ได้เก็บรวบรวมข้อมูลเก็บรวบรวมข้อมูลผ่านแบบสอบถามจากนักศึกษาจำนวน 1,000 คน ซึ่งรวบรวมในช่วงสิ้นภาคการศึกษา ปีการศึกษา 2022/2023 ประกอบด้วย 17 คุณลักษณะที่เกี่ยวข้องกับบริการต่าง ๆ ของมหาวิทยาลัย โดยดำเนินการจัดกลุ่มนักศึกษาด้วยแบบจำลอง K-Means Clustering และ Gaussian Mixture Model (GMM) และประเมินผลการจัดกลุ่มแต่ละแบบจำลองด้วยค่า Silhouette Score

ผลการวิจัยพบว่าแบบจำลอง K-Means Clustering แบ่งกลุ่มนักศึกษาออกเป็น 3 กลุ่ม มีค่า Silhouette Score = 0.44528 และแบบจำลอง Gaussian Mixture Model (GMM) แบ่งกลุ่มนักศึกษาออกเป็น 7 กลุ่ม มีค่า Silhouette Score = -0.500119 เมื่อเปรียบเทียบกันพบว่า K-Means Clustering ให้ผลลัพธ์ที่ดีกว่า GMM เนื่องจากมีค่า Silhouette Score เป็นบวก และสูงกว่า แสดงถึงการจัดกลุ่มที่มีประสิทธิภาพมากกว่า

2.7.3 Using RFM Model and Market Basket Analysis for Segmenting Customers and Assigning Marketing Strategies to Resulted Segments (Maraghi และคณะ, 2020)

งานวิจัยนี้นำเสนอแบบจำลองการจัดการลูกค้าสัมพันธ์ (CRM) แบบบูรณาการ โดยใช้การวิเคราะห์ RFM ร่วมกับการแบ่งกลุ่มลูกค้าด้วยแบบจำลอง K-means Clustering และการวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis) เพื่อทำความเข้าใจพฤติกรรมลูกค้าและกำหนดกลยุทธ์ทางการตลาดที่เหมาะสมสำหรับแต่ละกลุ่ม

การดำเนินงานวิจัยนี้ได้เก็บรวบรวมข้อมูลธุรกรรมการซื้อขายจากร้านค้าแห่งหนึ่งในประเทศจีน จำนวน 32,266 ราย โดยเริ่มต้นจากวิเคราะห์ RFM เพื่อกำหนดค่า Recency, Frequency และ Monetary ใช้แบบจำลอง K-means Clustering เพื่อทำการจัดกลุ่มลูกค้า และวิเคราะห์ ABC เพื่อคัดกรองสินค้าสำคัญ และใช้การวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis) ด้วย Apriori Algorithm ในการวิเคราะห์กฎความสัมพันธ์ของสินค้าในแต่ละกลุ่มลูกค้า เพื่อกำหนดกลยุทธ์การตลาด

ผลการวิจัยพบว่าแบบจำลอง K-Means Clustering แบ่งกลุ่มลูกค้าออกเป็น 6 กลุ่ม ผลการวิเคราะห์ ABC สามารถลดจำนวนสินค้าจาก 2,012 รายการ เหลือ 805 รายการ (กลุ่ม A และ B) โดยจำนวนธุรกรรมลดลงเพียง 1% (จาก 119,578 เหลือ 118,574) และผลการวิเคราะห์ Association Rules ได้กฎความสัมพันธ์รวม 2,656 กฎ และประสิทธิภาพของ Apriori Algorithm หลังตัดสินค้าจากการวิเคราะห์ ABC พบว่าค่าเฉลี่ย Confidence เพิ่มขึ้นจาก 63.4% เป็น 64.11% เวลาในการประมวลผลลดลงจาก 2:56 นาที เหลือ 1:10 นาที

2.7.4 A Comparative Study of RFM - Based Clustering Methods in Customer Segmentation (Ahammed และคณะ, 2023)

งานวิจัยนี้นำเสนอการศึกษาเปรียบเทียบวิธีการแบ่งกลุ่มลูกค้าโดยใช้เทคนิค RFM (Recency, Frequency, Monetary) ร่วมกับแบบจำลองการจัดกลุ่ม (Clustering) 3 รูปแบบ ได้แก่ K-Means Clustering, Fuzzy C-Means Clustering และ Graph-based Clustering เพื่อวิเคราะห์พฤติกรรมและแบ่งกลุ่มลูกค้าในธุรกิจค้าปลีก

การดำเนินงานวิจัยนี้ได้เก็บรวบรวมข้อมูลการทำธุรกรรมของลูกค้าจำนวน 23,053 รายการ จากลูกค้า 5,506 ราย โดยมีการบันทึกข้อมูลการซื้อขายในช่วงปี 2011 ถึง 2014 โดยมีขั้นตอนการวิจัย ดังนี้

- การประมวลผลข้อมูลเบื้องต้น (Data Pre-processing) เพื่อกำจัดค่าลบในข้อมูลปริมาณ ราคา และยอดขายรวม และสร้างตัวแปรยอดขายสุทธิโดยคำนวณจากปริมาณคูณราคา

- ทำการวิเคราะห์ RFM เพื่อกำหนดค่า Recency, Frequency และ Monetary
- ใช้เทคนิค Elbow Method เพื่อหาค่า k ที่เหมาะสมสำหรับการจัดกลุ่ม (k = 5)
- จัดกลุ่มลูกค้าด้วยแบบจำลอง 3 แบบ ได้แก่ K-Means, Fuzzy C-Means และ Graph-based clustering

- ประเมินผลการจัดกลุ่มด้วยเมตริกต่าง ๆ เช่น Silhouette Score, Davies-Bouldin Index และ Calinski-Harabasz Index

ผลการวิจัยแสดงให้เห็นว่า K-Means Clustering มีประสิทธิภาพดีที่สุดในการแบ่งกลุ่มลูกค้า โดยมีผลการประเมินดังนี้:

- K-Means Clustering ได้คะแนน Silhouette Score สูงสุดที่ 0.4230 แสดงถึงคุณภาพการจัดกลุ่มที่ดี Davies-Bouldin Score ที่ 0.8482 บ่งชี้ว่ากลุ่มมีการแบ่งแยกที่ชัดเจน และ Calinski-Harabasz Score ที่ 7,735.96 แสดงถึงความแตกต่างระหว่างกลุ่มที่ดี
- Fuzzy C-Means ได้ผลลัพธ์รองลงมา โดยมี Silhouette Score ที่ 0.3917 Davies-Bouldin Score ที่ 0.9477 และ Calinski-Harabasz Score ที่ 7,486.61
- Graph-Based Clustering มีประสิทธิภาพต่ำที่สุด โดยมี Silhouette Score ที่ 0.3040 Davies-Bouldin Score ที่ 1.2034 และ Calinski-Harabasz Score ที่ 3,208.95

K-Means Clustering ยังมีความรวดเร็วในการประมวลผล โดยใช้เวลาเพียง 0.05 วินาที ซึ่งเหมาะสมสำหรับการนำไปใช้งานจริง การแบ่งกลุ่มที่ได้สามารถนำไปใช้ในการพัฒนากลยุทธ์ทางการตลาดที่เฉพาะเจาะจงสำหรับลูกค้าแต่ละกลุ่มได้อย่างมีประสิทธิภาพ

2.7.5 An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market (John และคณะ, 2023)

งานวิจัยนี้นำเสนอการศึกษาการแบ่งกลุ่มลูกค้าโดยใช้เทคนิค RFM (Recency, Frequency, Monetary) และเปรียบเทียบประสิทธิภาพแบบจำลองการจัดกลุ่ม (Clustering) แบบต่าง ๆ ได้แก่ K-means Clustering, Gaussian Mixture Model (GMM), DBSCAN (Density-Based Spatial Clustering of Applications with Noise), BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) และ Agglomerative Clustering เพื่อพัฒนา

แบบจำลองการแบ่งกลุ่มลูกค้าที่มีประสิทธิภาพ สำหรับการแบ่งกลุ่มลูกค้าในตลาดค้าปลีกของสหราชอาณาจักร สำหรับปรับปรุงกระบวนการตัดสินใจทางธุรกิจ

การดำเนินงานวิจัยนี้ใช้ชุดข้อมูลค้าปลีกออนไลน์จาก UCI Machine Learning Repository ซึ่งประกอบด้วยข้อมูลลูกค้าจำนวน 541,909 รายการและมี 8 คุณลักษณะ โดยมีขั้นตอนการวิจัย ดังนี้

- การประมวลผลข้อมูลเบื้องต้น (Data Pre-processing)
- ทำการวิเคราะห์ RFM เพื่อกำหนดค่า Recency, Frequency และ Monetary
- ใช้ Principal Component Analysis (PCA) เพื่อลดมิติข้อมูล
- จัดกลุ่มลูกค้าด้วยแบบจำลองต่าง ๆ
- ประเมินผลการจัดกลุ่มด้วย Silhouette Score

ผลการวิจัยแสดงให้เห็นว่าการใช้ PCA ร่วมกับ Gaussian Mixture Model (GMM) ให้ประสิทธิภาพการจัดกลุ่มที่ดีที่สุด โดยมี Silhouette Score 0.80 ซึ่งสูงกว่าแบบจำลองอื่น ๆ อย่างมีนัยสำคัญ โดยผลการทดสอบแบบจำลองอื่น ๆ มีดังนี้ K-means ได้ 0.64, DBSCAN ได้ 0.626, BIRCH ได้ 0.64 และ Agglomerative Clustering ได้ 0.64

นอกจากนี้ การวิเคราะห์ด้วย RFM ยังช่วยแบ่งกลุ่มลูกค้าได้เป็น 5 กลุ่มคือ: ลูกค้าที่สูญเสีย (Lost customers) 31%, ลูกค้ามูลค่าต่ำ (Low-value customers) 30%, ลูกค้ามูลค่าปานกลาง (Medium-value customers) 21%, ลูกค้ามูลค่าสูง (High-value customers) 10% และลูกค้าระดับสูงสุด (Top customers) 8%

ความสำเร็จของ GMM เกิดจากความสามารถในการจับรูปแบบการกระจายตัวของข้อมูล และการใช้ PCA ช่วยลดปัญหาเรื่องมิติข้อมูลสูง ทำให้สามารถระบุกลุ่มลูกค้าได้แม่นยำกว่า ในขณะที่แบบจำลองอื่น ๆ เช่น K-means มีข้อจำกัดเรื่องการเลือกจุดศูนย์กลางเริ่มต้นแบบสุ่ม ทำให้ผลลัพธ์มีความแปรปรวน ± 0.06 ในการทดสอบหลายครั้ง

จากการศึกษางานวิจัยที่เกี่ยวข้อง พบว่างานวิจัยส่วนใหญ่มุ่งเน้นเพื่อเปรียบเทียบการจัดกลุ่มลูกค้าด้วยแบบจำลองต่าง ๆ เท่านั้น โดยยังไม่ได้นำผลการจัดกลุ่มไปใช้ในเชิงกลยุทธ์ต่อ ยอดอย่างครบถ้วน ผู้วิจัยจึงได้ต่อยอดแนวคิดของงานวิจัย 2.7.3 Using RFM Model and Market Basket Analysis for Segmenting Customers and Assigning Marketing Strategies to Resulted Segments (Maraghi และคณะ, 2020) โดยการนำแนวคิดการวิเคราะห์ RFM การจัดกลุ่มลูกค้า และการวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis) โดยปรับใช้เครื่องมือและเทคนิคที่เหมาะสม โดยใช้ Gaussian Mixture Models (GMM) และ Hierarchical

Clustering เพื่อเปรียบเทียบผลลัพธ์การจัดกลุ่มกับ K-means Clustering เพื่อเพิ่มความแม่นยำ หลังจากนั้นวิเคราะห์กลุ่มสินค้าตามหลักการของ ABC Analysis และใช้เทคนิคการวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis) เพื่อออกแบบกลยุทธ์การตลาดที่เหมาะสมกับกลุ่มลูกค้าในงานวิจัยนี้ได้



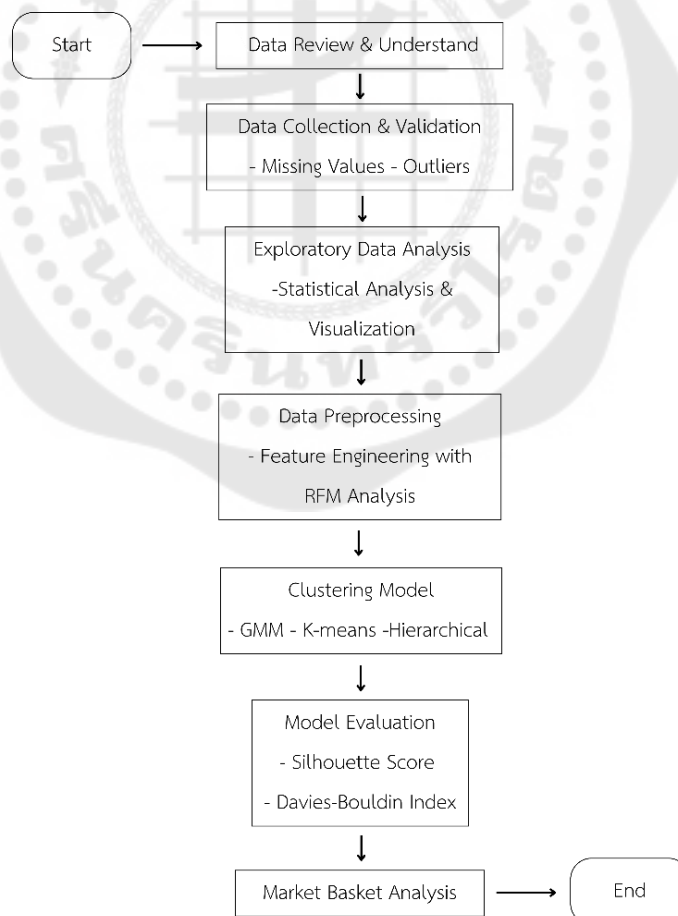
บทที่ 3

วิธีดำเนินการวิจัย

ในการวิจัยครั้งนี้ ผู้วิจัยได้ดำเนินการตามขั้นตอนดังนี้

1. กระบวนการทำงานวิจัย
2. การทำความเข้าใจและตรวจสอบข้อมูล
3. วิเคราะห์ข้อมูลเชิงสำรวจ (Exploratory Data Analysis)
4. การเตรียมข้อมูล (Data Preprocessing)
5. การสร้างแบบจำลองจัดกลุ่มลูกค้า (Clustering Model)
6. การประเมินประสิทธิภาพแบบจำลอง (Model Evaluation)
7. การวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis)

3.1 กระบวนการทำงานวิจัย



ภาพประกอบ 4 แสดงกระบวนการทำงานของงานวิจัย

จากภาพประกอบ 4 แสดงให้เห็นถึงกระบวนการทำงานวิจัย เริ่มต้นจากการนำเข้าข้อมูลธุรกรรมของร้านค้าปลีกออนไลน์แห่งหนึ่ง จากนั้นทำความเข้าใจข้อมูล ตรวจสอบความสมบูรณ์ของข้อมูล และวิเคราะห์ข้อมูลเชิงตรวจสอบ (Exploratory Data Analysis) เพื่อเข้าใจข้อมูลเชิงลึก จากนั้นทำการเตรียมข้อมูล (Data Preprocessing) โดยการสร้างคุณลักษณะข้อมูล (Feature Engineering) เพื่อให้ข้อมูลอยู่ในรูปแบบที่พร้อมต่อการนำมาจัดกลุ่มด้วยแบบจำลอง (Clustering Model) ได้แก่ Gaussian Mixture Model, K-means Clustering, และ Hierarchical Clustering และประเมินประสิทธิภาพแบบจำลอง (Model Evaluation) ด้วยตัวชี้วัด Silhouette Score และ Davies-Bouldin Index เพื่อเลือกโมเดลที่เหมาะสมที่สุดในการทำการวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis) เพื่อสร้างกลยุทธ์การตลาดตามพฤติกรรมผู้บริโภคเฉพาะกลุ่ม

3.2 การทำความเข้าใจและตรวจสอบข้อมูล

3.2.1 การทำความเข้าใจข้อมูล

งานวิจัยนี้ใช้ชุดข้อมูลธุรกรรมของร้านค้าปลีกออนไลน์แห่งหนึ่งจากแหล่งข้อมูลสาธารณะ Kaggle.com มีข้อมูลทั้งหมด 1,067,371 รายการ มี 8 คุณลักษณะ ดังตาราง 1 แสดงตัวอย่างข้อมูลตามภาพประกอบ 5 และแสดงชนิดข้อมูลตามภาพประกอบ 6

ตาราง 1 แสดงคุณลักษณะของข้อมูล

ลำดับ	ชื่อคุณลักษณะ	คำอธิบาย
1	Invoice	หมายเลขใบแจ้งหนี้
2	StockCode	รหัสสินค้า
3	Description	ชื่อสินค้า
4	Quantity	จำนวนสินค้าที่ขายต่อใบแจ้งหนี้
5	InvoiceDate	วันที่ของใบแจ้งหนี้
6	Price	ราคาสินค้า
7	Customer ID	รหัสลูกค้า
8	Country	ประเทศ

	Invoice	StockCode	Description	Quantity	InvoiceDate	Price	Customer ID	Country
0	489434	85048	15CM CHRISTMAS GLASS BALL 20 LIGHTS	12	2009-12-01 07:45:00	6.95	13085.0	United Kingdom
1	489434	79323P	PINK CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
2	489434	79323W	WHITE CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
3	489434	22041	RECORD FRAME 7" SINGLE SIZE	48	2009-12-01 07:45:00	2.10	13085.0	United Kingdom
4	489434	21232	STRAWBERRY CERAMIC TRINKET BOX	24	2009-12-01 07:45:00	1.25	13085.0	United Kingdom

ภาพประกอบ 5 แสดงตัวอย่างข้อมูล 5 แถวแรก

```
##### Shape #####
(1067371, 8)
##### Types #####
Invoice      object
StockCode    object
Description   object
Quantity      int64
InvoiceDate   object
Price         float64
Customer ID   float64
Country       object
dtype: object
```

ภาพประกอบ 6 แสดงชนิดของข้อมูล

3.2.2 การตรวจสอบข้อมูล

ดำเนินการตรวจสอบข้อมูลและปรับปรุงข้อมูลเพื่อให้อยู่ในรูปแบบที่ถูกต้อง สมบูรณ์ เช่น การจัดการค่าผิดปกติ (Outliers) การจัดการข้อมูลไม่ถูกต้อง (Incorrectly data) การจัดการข้อมูลที่สูญหายไป (Missing Values) และสร้างตัวแปรใหม่ เป็นต้น ตามภาพประกอบ 7 ถึง 9

```
[ ] def outlier_thresholds(dataframe, variable):
    quartile1 = dataframe[variable].quantile(0.01)
    quartile3 = dataframe[variable].quantile(0.99)
    interquartile_range = quartile3 - quartile1
    up_limit = quartile3 + 1.5 * interquartile_range
    low_limit = quartile1 - 1.5 * interquartile_range
    return low_limit, up_limit

def replace_with_thresholds(dataframe, variable):
    low_limit, up_limit = outlier_thresholds(dataframe, variable)
    dataframe.loc[(dataframe[variable] > up_limit), variable] = up_limit

[ ] replace_with_thresholds(df, "Quantity")
    replace_with_thresholds(df, "Price")
```

ภาพประกอบ 7 แสดงการจัดการค่าผิดปกติ (Outliers)

```
# Missing values are deleted. Canceled Invoices are not received and a new variable was created.

def data_prep(dataframe):
    dataframe.dropna(axis=0, inplace=True)
    dataframe = dataframe[~dataframe["Invoice"].str.contains("C", na=False)]
    dataframe = dataframe[dataframe["Quantity"] > 0]
    dataframe["TotalPrice"] = dataframe["Quantity"] * dataframe["Price"]
    return dataframe
```

ภาพประกอบ 8 แสดงการจัดการข้อมูลไม่ถูกต้อง (Incorrectly data) และข้อมูลที่สูญหายไป (Missing Values) และสร้างตัวแปรใหม่

```
##### Shape #####
(805620, 9)
##### Types #####
Invoice      object
StockCode    object
Description  object
Quantity     float64
InvoiceDate  object
Price        float64
Customer ID  float64
Country      object
TotalPrice   float64
dtype: object

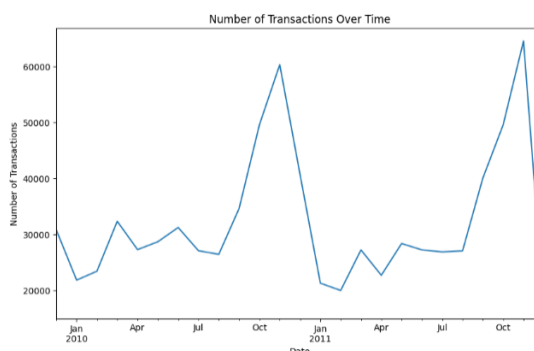
##### NA #####
Invoice      0
StockCode    0
Description  0
Quantity     0
InvoiceDate  0
Price        0
Customer ID  0
Country      0
TotalPrice   0
dtype: int64
```

ภาพประกอบ 9 แสดงผลหลังจากการตรวจสอบและแก้ไขข้อมูล

3.3 วิเคราะห์ข้อมูลเชิงสำรวจ (Exploratory Data Analysis)

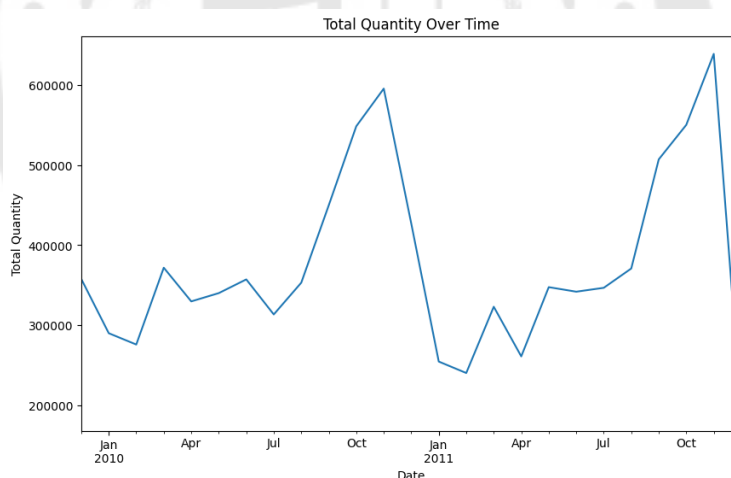
เป็นขั้นตอนสำคัญในกระบวนการทำงานกับข้อมูล โดยมีเป้าหมายเพื่อทำความเข้าใจข้อมูลเบื้องต้น ทั้งในด้านโครงสร้าง ความสัมพันธ์ และลักษณะเฉพาะ เพื่อระบุปัญหาที่อาจเกิดขึ้นและกำหนดแนวทางการวิเคราะห์หรือการสร้างแบบจำลองในขั้นต่อไป

เริ่มจากแนวโน้มจำนวนธุรกรรมที่เกิดขึ้นตามช่วงเวลา ตั้งแต่เดือนธันวาคม 2009 ถึง ธันวาคม 2011 โดยพบว่าจำนวนธุรกรรมมีความผันผวนเป็นระยะ และมีจุดสูงสุดในช่วงไตรมาสสุดท้ายของแต่ละปี (ประมาณเดือนตุลาคม) ซึ่งอาจสะท้อนถึงฤดูกาลหรือปัจจัยส่งเสริมการขายที่เกิดขึ้นในช่วงเวลาดังกล่าว อย่างไรก็ตาม หลังจากถึงจุดสูงสุด จำนวนธุรกรรมมีแนวโน้มลดลงอย่างรวดเร็ว ก่อนที่จะค่อย ๆ พ้นตัวขึ้นใหม่ในช่วงปีถัดไป ตามภาพประกอบ 10



ภาพประกอบ 10 แสดงจำนวนธุรกรรมต่อเดือน

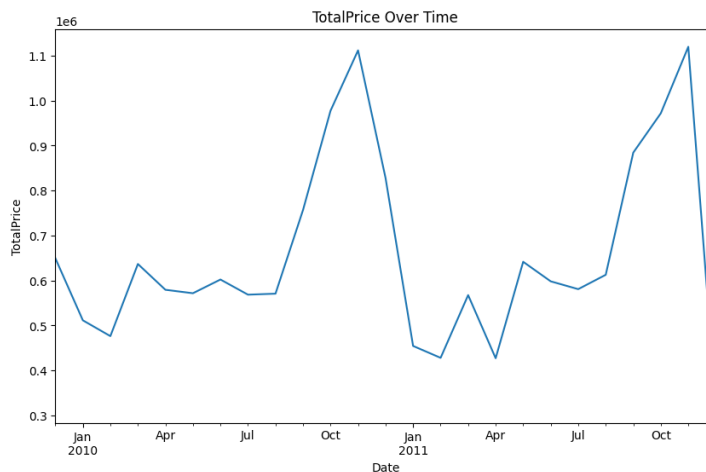
ต่อมา แนวโน้มปริมาณสินค้าที่ขายได้ตามช่วงเวลา ตั้งแต่เดือนธันวาคม 2009 ถึง ธันวาคม 2011 โดยพบว่าปริมาณสินค้าที่ขายมีความผันผวนตามฤดูกาล ลักษณะการเปลี่ยนแปลงของข้อมูลมีแนวโน้มคล้ายกับจำนวนธุรกรรมที่เกิดขึ้น โดยมีจุดสูงสุดในช่วงไตรมาสสุดท้ายของแต่ละปี (ประมาณเดือนตุลาคม) ซึ่งอาจเกิดจากช่วงโปรโมชั่นหรือเทศกาลสำคัญ อย่างไรก็ตาม หลังจากแต่ละจุดสูงสุด ปริมาณสินค้าที่ขายได้ลดลงอย่างรวดเร็วก่อนที่จะเริ่มฟื้นตัวขึ้นในช่วงปีถัดไป ตามภาพประกอบ 11



ภาพประกอบ 11 แสดงจำนวนชิ้นงานต่อเดือน

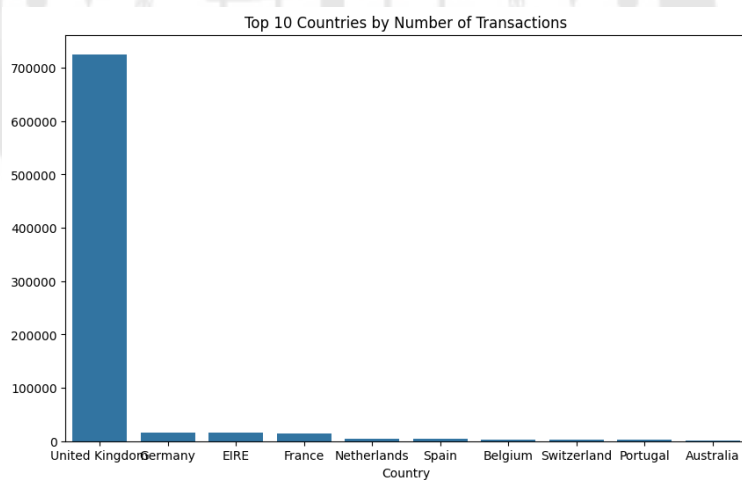
สุดท้าย แสดงแนวโน้มรายได้รวมจากการขายสินค้า (Total Price) ตามช่วงเวลา ตั้งแต่เดือนธันวาคม 2009 ถึงธันวาคม 2011 โดยพบว่ารายได้จากการขายมีความผันผวนอย่างมีรูปแบบชัดเจน โดยมีจุดสูงสุดในช่วงไตรมาสสุดท้ายของแต่ละปี (ประมาณเดือนตุลาคม) ซึ่งอาจเกิดจากช่วงเทศกาลหรือกิจกรรมส่งเสริมการขายที่กระตุ้นยอดขาย อย่างไรก็ตาม หลังจากแต่ละจุดสูงสุด รายได้ลดลงอย่างรวดเร็วก่อนที่จะเริ่มฟื้นตัวในช่วงถัดไป สะท้อนให้เห็นถึงพฤติกรรม

ซื้อของลูกค้าที่อาจได้รับอิทธิพลจากปัจจัยทางฤดูกาลและโปรโมชั่นทางการตลาด ตามภาพประกอบ 12



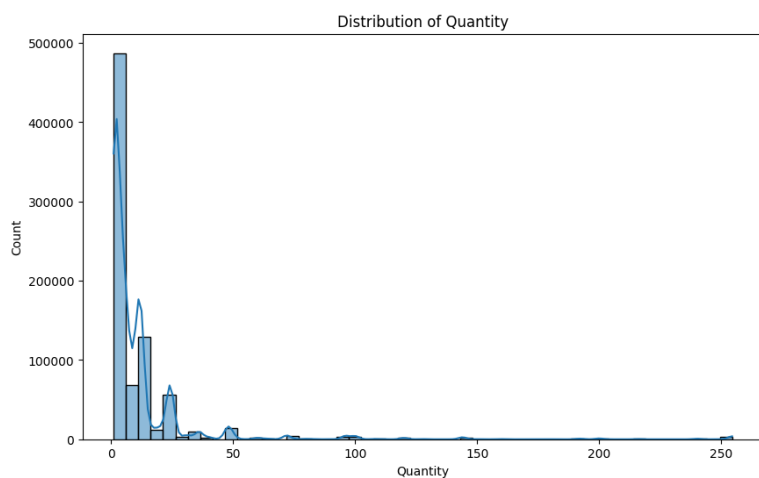
ภาพประกอบ 12 แสดงจำนวนรายได้ต่อเดือน

แสดงข้อมูลจำนวนธุรกรรม (Transactions) มากที่สุด 10 อันดับ จะเห็นได้ว่า United Kingdom มีจำนวนธุรกรรมสูงที่สุด ตามภาพประกอบ 13



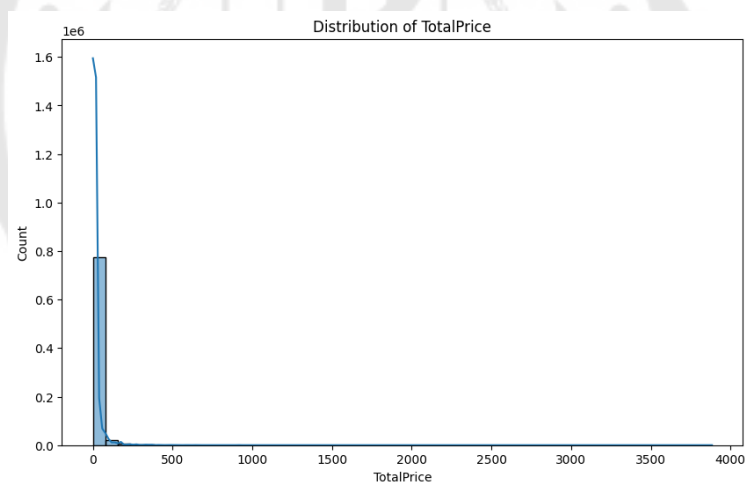
ภาพประกอบ 13 แสดงประเทศที่มีจำนวนธุรกรรมสูงที่สุด 10 อันดับ

แสดงการกระจายตัวของจำนวนสินค้าต่อธุรกรรม พบว่าจำนวนสินค้าที่ซื้อในแต่ละครั้งมีปริมาณต่ำกว่า 50 ซึ่งแสดงให้เห็นว่าธุรกิจนี้มีลูกค้าส่วนใหญ่ที่ซื้อสินค้าจำนวนน้อยต่อธุรกรรม ตามภาพประกอบ 14



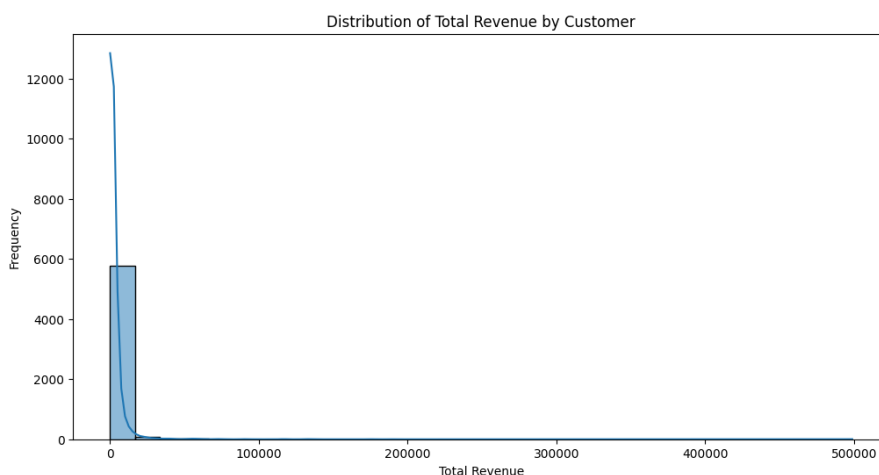
ภาพประกอบ 14 แสดงการกระจายตัวของจำนวนสินค้าต่อธุรกรรม

แสดงการกระจายตัวของยอดรายได้ต่อธุรกรรม พบว่ายอดรายได้ในการซื้อแต่ละครั้งมีปริมาณต่ำกว่า 500 ซึ่งแสดงให้เห็นว่าธุรกิจนี้มีลูกค้าส่วนใหญ่ที่มียอดใช้จ่ายจำนวนน้อยต่อธุรกรรม ตามภาพประกอบ 15



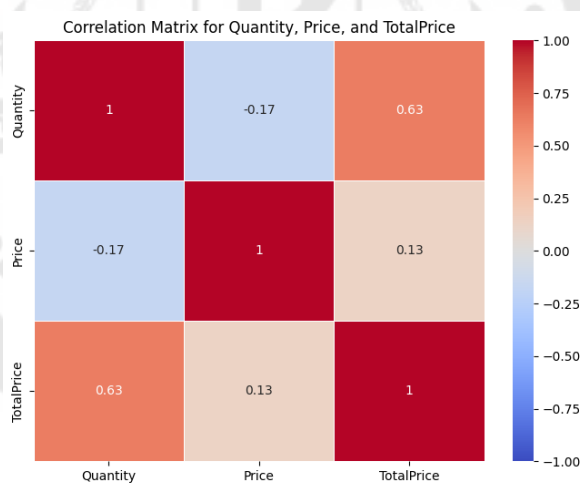
ภาพประกอบ 15 แสดงการกระจายตัวของยอดรายได้ต่อธุรกรรม

แสดงการกระจายตัวของยอดรายได้ต่อลูกค้า พบว่ายอดรายได้ต่อลูกค้ามีปริมาณต่ำกว่า 100,000 ซึ่งแสดงให้เห็นว่าธุรกิจนี้มีลูกค้าส่วนใหญ่ที่มียอดใช้จ่ายต่อรายไม่มากนัก ตามภาพประกอบ 16



ภาพประกอบ 16 แสดงการกระจายตัวของยอดรายได้ต่อลูกค้า

แสดงการวิเคราะห์ความสัมพันธ์ระหว่างคุณลักษณะ พบว่ามีความสัมพันธ์ในทางบวกระหว่าง Quantity และ TotalPrice 63%, ระหว่าง Price และ TotalPrice 13% ตามภาพประกอบ 17



ภาพประกอบ 17 แสดงการวิเคราะห์ความสัมพันธ์ระหว่างคุณลักษณะ

3.4 การเตรียมข้อมูล (Data Preprocessing)

งานวิจัยนี้ใช้การสร้างคุณลักษณะ (Feature Engineering) จาก RFM Analysis เป็นการวิเคราะห์ข้อมูลลูกค้ารายบุคคล โดยอาศัยข้อมูลพฤติกรรมการซื้อสินค้าตาม Customer ID ได้แก่ recency, frequency และ monetary ตามภาพประกอบ 18



	recency	frequency	monetary
Customer ID			
12346.0	326	12	637.540
12347.0	3	8	5633.320
12348.0	76	5	2019.400
12349.0	19	4	3812.745
12350.0	311	1	334.400

ภาพประกอบ 18 แสดงการสร้างคุณลักษณะ recency, frequency, monetary

เนื่องจากตัวแปร recency, frequency และ monetary เป็นข้อมูลตัวเลขที่มีช่วงค่าแตกต่างกันมาก จำเป็นต้องมีการแปลงค่าตัวแปรให้อยู่ในรูปของคะแนนหรือ Score โดยการแบ่งข้อมูลตามการแบ่งควอไทล์(quantiles) ให้แต่ละกลุ่มมีจำนวนข้อมูลใกล้เคียงกัน โดยกำหนด RecencyScore, FrequencyScore และ MonetaryScore เป็นค่าตั้งแต่ 1 - 5 เพื่อให้สามารถนำไปใช้ในการจัดกลุ่มลูกค้าด้วยแบบจำลองได้อย่างมีประสิทธิภาพ ตามภาพประกอบ 19 และตาราง 2



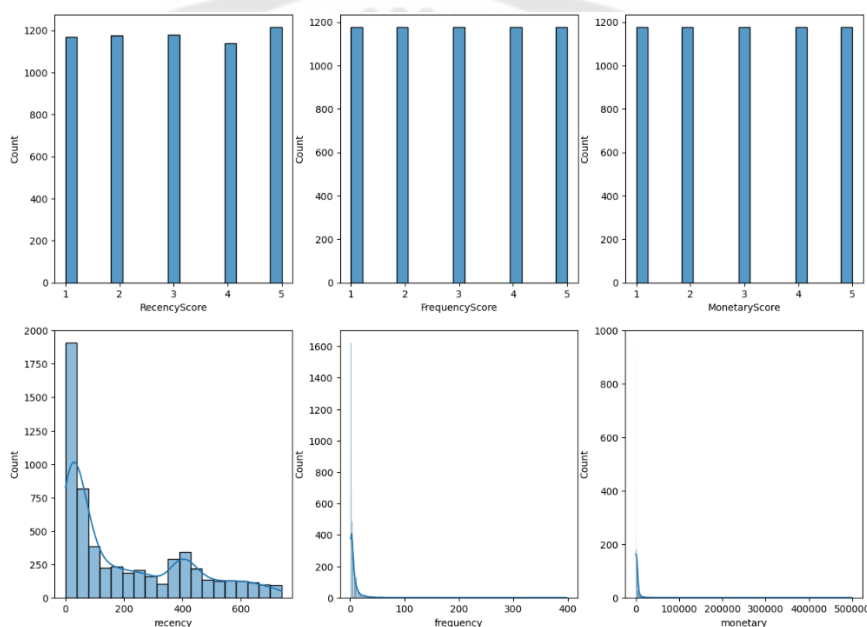
	recency	frequency	monetary	RecencyScore	FrequencyScore	MonetaryScore
Customer ID						
12346.0	326	12	637.540	2	5	3
12347.0	3	8	5633.320	5	4	5
12348.0	76	5	2019.400	3	4	4
12349.0	19	4	3812.745	5	3	5
12350.0	311	1	334.400	2	1	2

ภาพประกอบ 19 แสดงการสร้างคุณลักษณะ RecencyScore, FrequencyScore, MonetaryScore

ตาราง 2 แสดงการแบ่งช่วงค่า RFM ให้อยู่ในรูปแบบ Score

Recency Score	Recency Range	Frequency Score	Frequency Range	Monetary Score	Monetary Range
5	(1.0, 20.0)	1	(1.0, 1176.4)	1	(2.95, 281.93)
4	(20.0, 59.0)	2	(1176.4, 2351.8)	2	(281.93, 609.52)
3	(59.0, 190.0)	3	(2351.8, 3527.2)	3	(609.52, 1221.13)
2	(190.0, 411.0)	4	(3527.2, 4702.6)	4	(1221.13, 2898.66)
1	(411.0, 739.0)	5	(4702.6, 5878.0)	5	(2898.66, 498798.15)

เมื่อทำการวิเคราะห์การกระจายตัวของ recency, frequency และ monetary พบว่าตัวแปรแต่ละตัวมีการกระจายตัวที่ไม่สมมาตร (skewed distribution) โดย recency มีค่าต่ำสุดกระจุกตัวอยู่ที่ศูนย์และค่อย ๆ ลดลงเมื่อระยะเวลาห่างจากการซื้อครั้งล่าสุดเพิ่มขึ้น Frequency มีค่าส่วนใหญ่กระจุกตัวที่จำนวนครั้งซื้อที่ต่ำ และค่อย ๆ ลดลงเมื่อจำนวนครั้งซื้อเพิ่มขึ้น และ Monetary แสดงให้เห็นว่าลูกค้าส่วนใหญ่มียอดใช้จ่ายต่ำ และมียอดใช้จ่ายสูงสุดกระจุกตัวในกลุ่มลูกค้าบางราย โดยหลังจากทำการจัดกลุ่ม RFM Score ด้วยการแบ่งข้อมูลตามควอไทล์ (Quantile-based Binning) พบว่าการแบ่งกลุ่มเป็น 5 ระดับทำให้ข้อมูลมีการกระจายตัวที่สมดุลมากขึ้น โดยแต่ละ Score มีจำนวนตัวอย่างใกล้เคียงกัน ซึ่งช่วยการจัดกลุ่มลูกค้ามีความแม่นยำมากขึ้น ตามภาพประกอบ 20



ภาพประกอบ 20 แสดงการกระจายตัวของตัวแปร RFM

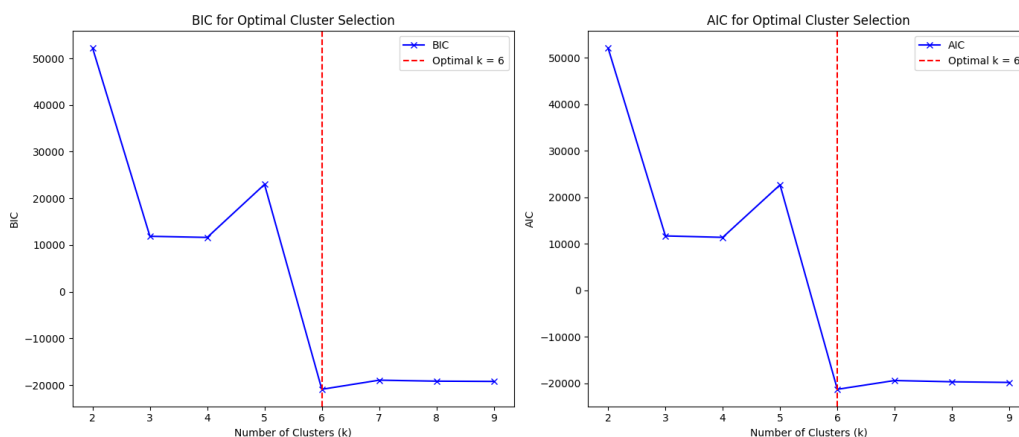
3.5 การสร้างแบบจำลองจัดกลุ่มลูกค้า (Clustering Model)

นำข้อมูลที่ผ่านการเตรียมข้อมูลเรียบร้อยแล้ว มาดำเนินการสร้างแบบจำลองจัดกลุ่มลูกค้า โดยงานวิจัยนี้จะใช้แบบจำลอง 3 ประเภท ได้แก่ Gaussian Mixture Model (GMM), K-Means Clustering และ Hierarchical Clustering

ผู้วิจัยเลือกใช้วิธีการหาจำนวนกลุ่มที่เหมาะสม (K) เฉพาะสำหรับแต่ละแบบจำลอง เนื่องจากแต่ละแบบจำลองมีหลักการจัดกลุ่มที่แตกต่างกัน โดย Gaussian Mixture Model (GMM) ใช้การประมาณ Gaussian Distribution หลายมิติ K-Means Clustering ใช้ centroid ส่วน Hierarchical Clustering ใช้โครงสร้างแบบลำดับขั้น ดังนั้น ผู้วิจัยจึงคิดว่าการกำหนดค่า K

แยกตามแบบจำลอง จะทำให้แบบจำลองมีความเหมาะสมสูงสุด และสะท้อนศักยภาพของแบบจำลองได้อย่างแท้จริง อย่างไรก็ตาม ในการพัฒนาเพิ่มเติมในอนาคต อาจมีการทดสอบการเปรียบเทียบผลลัพธ์โดยใช้ค่า K เดียวกัน เพื่อประเมินในมุมมองที่แตกต่างด้วย

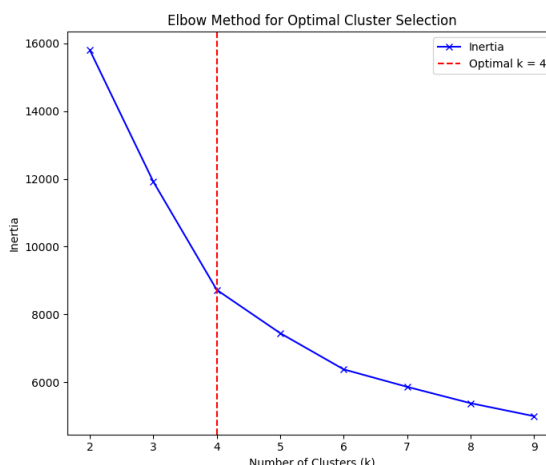
เริ่มต้นด้วยแบบจำลอง Gaussian Mixture Model (GMM) โดยทำการหาจำนวนกลุ่มที่เหมาะสมที่สุด (K) สำหรับ Gaussian Mixture Model (GMM) ด้วยเกณฑ์ BIC (Bayesian Information Criterion) และ AIC (Akaike Information Criterion) โดยทำการดูจุดที่ค่าของ BIC และ AIC ต่ำที่สุด ทำให้ได้ค่า K เท่ากับ 6 ตามภาพประกอบ 21



ภาพประกอบ 21 แสดงการหาจำนวนกลุ่มที่เหมาะสมที่สุดของแบบจำลอง GMM

จากนั้นสร้างแบบจำลองสำหรับจัดกลุ่มลูกค้าโดยกำหนดพารามิเตอร์ให้ `n_components` เท่ากับจำนวนกลุ่มจากค่า BIC (Bayesian Information Criterion) ซึ่งช่วยให้เลือกจำนวนกลุ่มที่เหมาะสมที่สุด

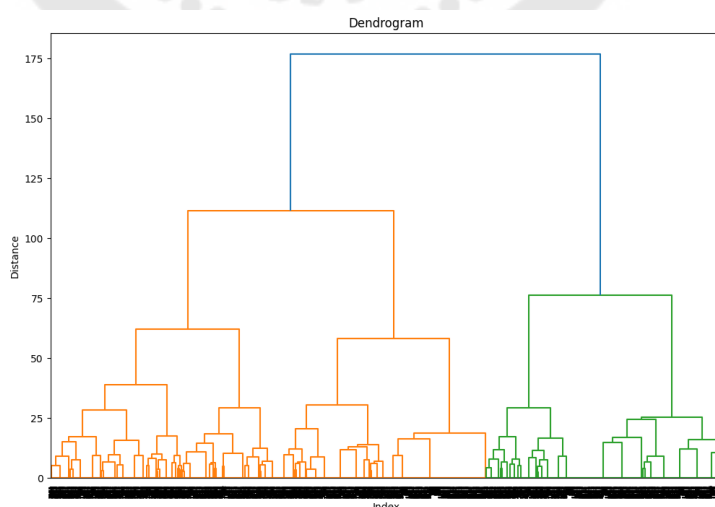
ต่อมา ทำการหาจำนวนกลุ่มที่เหมาะสมที่สุด (K) สำหรับ K-Means Clustering (K-Means) ด้วยวิธี Elbow Method เพื่อแสดงค่า Inertia เทียบกับจำนวนกลุ่ม โดยมองหาจุดที่การลดลงของ Inertia เริ่มช้าลง (จุดหักศอก) ทำให้ได้ค่า K เท่ากับ 4 ตามภาพประกอบ 22



ภาพประกอบ 22 แสดงการหาจำนวนกลุ่มที่เหมาะสมที่สุดของแบบจำลอง K-Means

จากนั้นสร้างแบบจำลองสำหรับจัดกลุ่มลูกค้าโดยกำหนดพารามิเตอร์ให้ $n_clusters = k_inertia$ ถูกกำหนดให้เป็นจำนวนคลัสเตอร์ที่เหมาะสมที่สุด โดยพิจารณาจากจุดหักศอก (Elbow Point) ของค่า Inertia ซึ่งช่วยให้เลือกจำนวนกลุ่มที่เหมาะสมที่สุด

สุดท้ายนี้ ทำการหาจำนวนกลุ่มที่เหมาะสมที่สุด (K) สำหรับ Hierarchical Clustering ด้วยการสร้าง Dendrogram เพื่อช่วยสังเกตจุดตัดของลำดับขั้นการรวมกลุ่ม โดยใช้ Ward's method ซึ่งช่วยลดความแปรปรวนภายในกลุ่มเพื่อให้กลุ่มที่ได้มีความคล้ายคลึงกันมากที่สุด จากนั้นจะแสวงหา threshold distance ที่เหมาะสม โดยใช้ค่าจากช่วงที่มีการเปลี่ยนแปลงระยะห่างมากที่สุด โดยกำหนด $max_distance = distances[-10]$ เพื่อช่วยเลือกจุดตัดของ Dendrogram อย่างเป็นระบบ ทำให้ได้ค่า K เท่ากับ 10 ตามภาพประกอบ 23



ภาพประกอบ 23 แสดงการหาจำนวนกลุ่มที่เหมาะสมที่สุดของแบบจำลอง Hierarchical

จากนั้นสร้างแบบจำลองสำหรับจัดกลุ่มลูกค้าโดยกำหนดพารามิเตอร์ให้ใช้ค่า linked ซึ่งเป็นผลลัพธ์จากการสร้างโครงสร้างลำดับขั้นด้วย Ward's method และใช้ max_distance เป็นเกณฑ์ในการแบ่งกลุ่ม โดยเลือกค่าที่เหมาะสมจากช่วงที่มีการเปลี่ยนแปลงระยะห่างมากที่สุดใน linkage matrix เพื่อให้จำนวนกลุ่มอยู่ในช่วงที่ต้องการ นอกจากนี้ กำหนด criterion = 'distance' เพื่อใช้ระยะห่างเป็นเงื่อนไขหลักในการตัดสินใจว่าข้อมูลแต่ละจุดควรถูกจัดอยู่ในกลุ่มใด ซึ่งช่วยให้สามารถแบ่งกลุ่มลูกค้าได้อย่างเหมาะสมตามโครงสร้างของข้อมูล

3.6 การประเมินประสิทธิภาพแบบจำลอง (Model Evaluation)

ผู้วิจัยได้ทำการประเมินประสิทธิภาพของแบบจำลองเพื่อเปรียบเทียบประสิทธิภาพการจัดกลุ่มลูกค้าด้วยตัวชี้วัด Silhouette Score และตัวชี้วัด Davies-Bouldin Index (DBI)

3.7 การวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis)

นอกจากการจัดกลุ่มลูกค้า ผู้วิจัยได้ทำการวิเคราะห์กลุ่มสินค้าตามหลักการของ ABC Analysis และใช้เทคนิคการวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis) เพื่อตรวจสอบความสัมพันธ์ระหว่างสินค้าที่ลูกค้ามักจะซื้อพร้อมกัน ซึ่งประกอบด้วยค่า Support (การสนับสนุน), Confidence (ความมั่นใจ), Lift (การยก) และ Conviction (การพิสูจน์) โดยกำหนดให้ Minimum Support คือ 0.05 และ Minimum Confidence คือ 0.7 ด้วย FP-Growth Algorithm ตามภาพประกอบ 24 และ 25

```
frequent_itemsets = fpgrowth(basket, min_support=0.05, use_colnames=True)
```

```
[143] frequent_itemsets.head()
```

	support	itemsets
0	0.079598	(DOORMAT UNION FLAG)
1	0.066303	(DOORMAT NEW ENGLAND)
2	0.060508	(DOORMAT SPOTTY HOME SWEET HOME)
3	0.059144	(EDWARDIAN PARASOL NATURAL)
4	0.054713	(EDWARDIAN PARASOL BLACK)

ภาพประกอบ 24 แสดงการค้นหารายการสินค้าที่มีการซื้อบ่อย (Frequent Itemsets)

```

[144] rules = association_rules(frequent_itemsets, metric="confidence", min_threshold=0.7, num_itemsets=len(frequent_itemsets))
[145] rules.head()

```

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	representativity	leverage	conviction	zh
0	(ROSES REGENCY TEACUP AND SAUCER)	(REGENCY CAKESTAND 3 TIER)	0.088802	0.223965	0.070735	0.796545	3.556568	1.0	0.050846	3.814287	
1	(ROSES REGENCY TEACUP AND SAUCER)	(GREEN REGENCY TEACUP AND SAUCER)	0.088802	0.080961	0.067326	0.758157	9.364441	1.0	0.060136	3.800152	
2	(GREEN REGENCY TEACUP AND SAUCER)	(ROSES REGENCY TEACUP AND SAUCER)	0.080961	0.088802	0.067326	0.831579	9.364441	1.0	0.060136	5.410239	
3	(GREEN REGENCY TEACUP AND SAUCER)	(REGENCY CAKESTAND 3 TIER)	0.080961	0.223965	0.064428	0.795789	3.553194	1.0	0.046296	3.800174	

ภาพประกอบ 25 แสดงการสร้างกฎความสัมพันธ์ (Association Rules) ระหว่างสินค้า



บทที่ 4

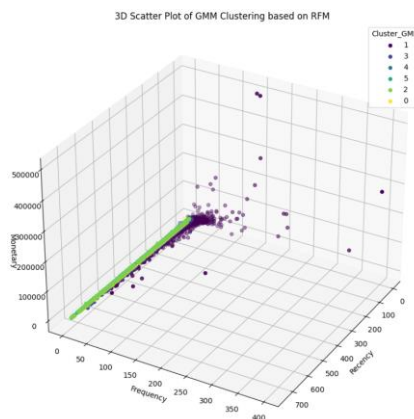
ผลการดำเนินการวิจัย

งานวิจัยนี้ได้ทำการศึกษาการวิเคราะห์พฤติกรรมลูกค้าของร้านค้าปลีกออนไลน์แห่งหนึ่งจากแหล่งข้อมูลสาธารณะ Kaggle.com ด้วยเทคนิค RFM และจัดกลุ่มลูกค้าด้วยเทคนิคการเรียนรู้ของเครื่องแบบไม่มีผู้สอน (Unsupervised Machine Learning) จากนั้นทำการประเมินผลลัพธ์การจัดกลุ่มลูกค้า วิเคราะห์กลุ่มสินค้าตามหลักการของ ABC Analysis และใช้เทคนิคการวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis) เพื่อทำการตลาดเฉพาะกลุ่มลูกค้า โดยผู้วิจัยได้ดำเนินการวิจัยตามกระบวนการเพื่อให้บรรลุวัตถุประสงค์ที่กำหนดไว้ ดังนี้

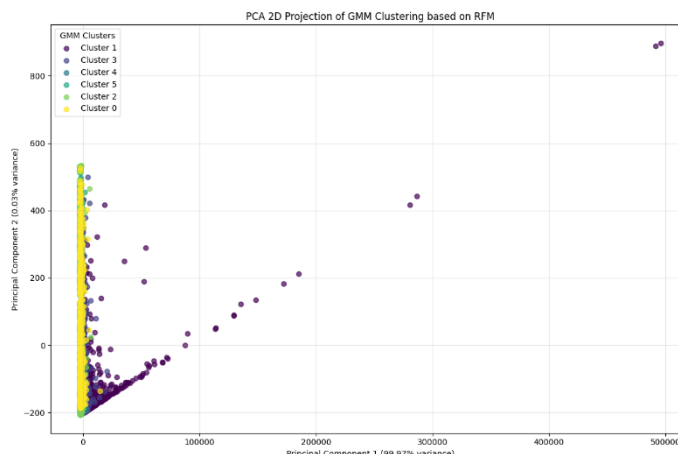
1. ผลลัพธ์การจัดกลุ่มด้วยแบบจำลอง Gaussian Mixture Model
2. ผลลัพธ์การจัดกลุ่มด้วยแบบจำลอง K-Means Clustering
3. ผลลัพธ์การจัดกลุ่มด้วยแบบจำลอง Hierarchical Clustering
4. เปรียบเทียบผลลัพธ์การจัดกลุ่ม
5. ผลการวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis)

4.1 ผลลัพธ์การจัดกลุ่มด้วยแบบจำลอง Gaussian Mixture Model

จากการสร้างแบบจำลอง Gaussian Mixture Model (GMM) พบว่าการจัดกลุ่มลูกค้าด้วยจำนวนกลุ่มที่เหมาะสมที่สุด (K) ด้วยเกณฑ์ BIC (Bayesian Information Criterion) และ AIC (Akaike Information Criterion) ซึ่งได้ค่า K เท่ากับ 6 ทำให้ได้ค่า Silhouette score เท่ากับ 0.165 หมายความว่ากลุ่มที่ได้ อาจมีความทับซ้อนกันมาก และ Davies-Bouldin Index เท่ากับ 1.826 หมายความว่ากลุ่มที่แบ่งยังมีความคล้ายกันสูง ทำให้การแยกกลุ่มไม่ดีนักตามภาพประกอบที่ 26 และ 27



ภาพประกอบ 26 แสดงผลลัพธ์การจัดกลุ่มด้วย Gaussian Mixture Model ในรูปแบบ 3 มิติ



ภาพประกอบ 27 แสดงผลลัพธ์การจัดกลุ่มด้วย Gaussian Mixture Model ในรูปแบบ 2 มิติ

หลังจากจัดกลุ่มแล้ว ผู้วิจัยได้นำผลลัพธ์ที่ได้จากการจัดกลุ่มมาทำการวิเคราะห์พฤติกรรมการซื้อของลูกค้าจากการใช้เทคนิค RFM โดยการแบ่งลูกค้าออกเป็นกลุ่มต่าง ๆ ตามคะแนนที่ได้รับในแต่ละด้าน โดยใช้ค่าเฉลี่ยของแต่ละตัวแปรเป็นเกณฑ์ในการจัดกลุ่ม ตามตาราง 3 หากคะแนนของลูกค้าในแต่ละด้านสูงกว่าค่าเฉลี่ยที่คำนวณจากกลุ่มลูกค้าทั้งหมด จะถูกจัดให้เป็น High (H) และหากต่ำกว่าหรือเท่ากับค่าเฉลี่ยจะถูกจัดให้เป็น Low (L) ซึ่งแสดงให้เห็นลักษณะของลูกค้าแต่ละกลุ่มตามพฤติกรรมการซื้อได้ โดยสามารถจัดกลุ่มตามหลัก RFM Analysis ได้ 3 กลุ่ม เช่น Cluster 0 มีค่า RecencyScore เฉลี่ย 2.586 ต่ำกว่าค่าเฉลี่ย 3.01 จัดให้เป็น L มีค่า FrequencyScore เฉลี่ย 3.0 เท่ากับค่าเฉลี่ย 3.0 จัดให้เป็น L และมีค่า MonetaryScore เฉลี่ย 2.879 น้อยกว่าค่าเฉลี่ย 3.0 จัดให้เป็น L สรุป Cluster 0 มีกลุ่มตาม RFM Analysis คือ LLL เป็นต้น ตามภาพประกอบที่ 28 และสรุปผลการจัดกลุ่มตาม RFM Analysis ตามภาพประกอบที่ 29

ตาราง 3 แสดงค่าเฉลี่ยของแต่ละตัวแปร RFM Score

RecencyScore_mean	FrequencyScore_mean	MonetaryScore_mean
3.01	3.0	3.0

	count	recency_mean	frequency_mean	monetary_mean	RecencyScore_mean	FrequencyScore_mean	MonetaryScore_mean	RFM Analysis gmm
cluster_GMM								
0	983	226.193286	3.048830	1021.680525	2.585961	3.0	2.878942	LLL
1	1176	51.114796	19.988946	10030.264149	4.187075	5.0	4.744048	HHH
2	1175	289.666383	1.619574	532.651815	2.401702	2.0	2.015319	LLL
3	1175	125.639149	5.768511	2056.301611	3.418723	4.0	3.774468	HHH
4	193	10.466321	3.165803	1061.707772	5.000000	3.0	2.974093	HLL
5	1176	352.134354	1.000000	337.459109	2.060374	1.0	1.571429	LLL

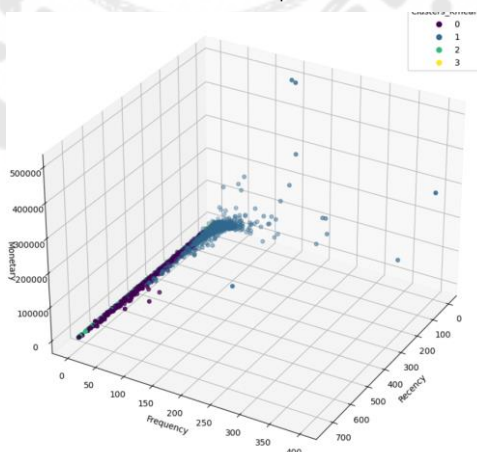
ภาพประกอบ 28 แสดงผลลัพธ์การวิเคราะห์พฤติกรรมของกลุ่มลูกค้า GMM ด้วย RFM Analysis

	count	recency_mean	frequency_mean	monetary_mean
RFM Analysis gmm				
HHH	2	88.376972	12.878728	6043.282880
HLL	1	10.466321	3.165803	1061.707772
LLL	3	289.331341	1.889468	630.597150

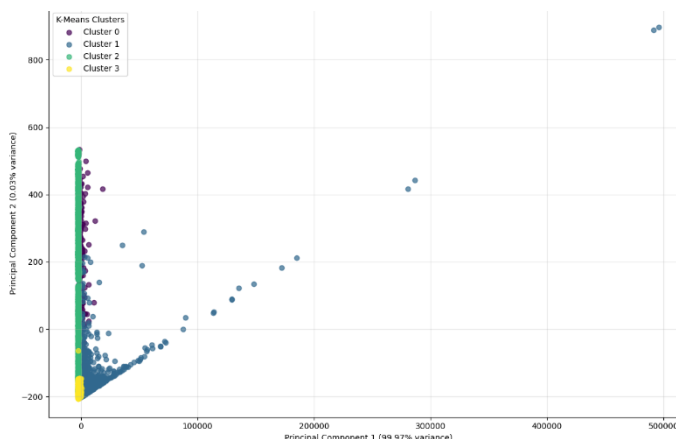
ภาพประกอบ 29 แสดงค่าเฉลี่ย RFM ของกลุ่มลูกค้า GMM

4.2 ผลลัพธ์การจัดกลุ่มด้วยแบบจำลอง K-Means Clustering

จากการสร้างแบบจำลอง K-Means Clustering พบว่าการจัดกลุ่มลูกค้าด้วยจำนวนกลุ่มที่เหมาะสมที่สุด (K) ด้วยวิธี Elbow Method ซึ่งได้ค่า K เท่ากับ 4 ทำให้ได้ค่า Silhouette score เท่ากับ 0.398 และ Davies-Bouldin Index เท่ากับ 0.956 หมายความว่ากลุ่มที่ได้มีความแยกกันอย่างชัดเจน เหมาะกับการใช้งานจริงมากที่สุด ตามภาพประกอบที่ 30 และ 31



ภาพประกอบ 30 แสดงผลลัพธ์การจัดกลุ่มด้วย K-Means Clustering ในรูปแบบ 3 มิติ



ภาพประกอบ 31 แสดงผลลัพธ์การจัดกลุ่มด้วย K-Means Clustering ในรูปแบบ 2 มิติ

หลังจากจัดกลุ่มแล้ว ผู้วิจัยได้นำผลลัพธ์ที่ได้จากการจัดกลุ่มมาทำการวิเคราะห์พฤติกรรม การซื้อของลูกค้าจากการใช้เทคนิค RFM โดยการแบ่งลูกค้าออกเป็นกลุ่มต่าง ๆ ตามคะแนนที่ได้รับในแต่ละด้าน โดยใช้ค่าเฉลี่ยของแต่ละตัวแปรเป็นเกณฑ์ในการจัดกลุ่ม หากคะแนนของลูกค้าในแต่ละด้านสูงกว่าค่าเฉลี่ยที่คำนวณจากกลุ่มลูกค้าทั้งหมด จะถูกจัดให้เป็น High (H) และหากต่ำกว่าค่าเฉลี่ยจะถูกจัดให้เป็น Low (L) ซึ่งแสดงให้เห็นลักษณะของลูกค้าแต่ละกลุ่มตามพฤติกรรม การซื้อได้ โดยสามารถจัดกลุ่มตามหลัก RFM Analysis ได้ 4 กลุ่ม ตามภาพประกอบที่ 32 และ 33

	count	recency_mean	frequency_mean	monetary_mean	RecencyScore_mean	FrequencyScore_mean	MonetaryScore_mean	Analysis Kmeans
cluster_Kmean								
0	1543	272.007129	4.024627	1366.478461	2.180817	3.229423	3.296824	LHH
1	1595	37.865831	16.159248	8212.790621	4.286520	4.693417	4.679624	HHH
2	1766	385.772367	1.216308	279.746436	1.784258	1.437146	1.459796	LLL
3	974	25.868583	2.915811	761.575843	4.455852	2.697125	2.571869	HLL

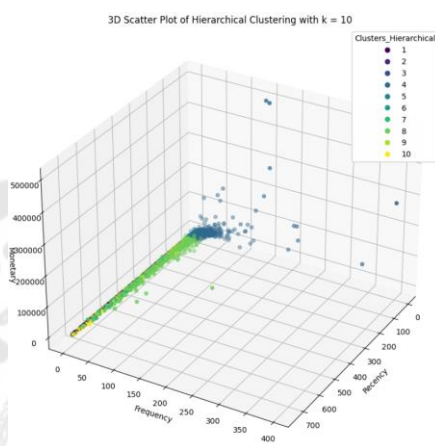
ภาพประกอบ 32 แสดงผลลัพธ์การวิเคราะห์พฤติกรรมของกลุ่มลูกค้า K-Means ด้วย RFM Analysis

	count	recency_mean	frequency_mean	monetary_mean
RFM Analysis Kmeans				
HHH	1	37.865831	16.159248	8212.790621
HLL	1	25.868583	2.915811	761.575843
LHH	1	272.007129	4.024627	1366.478461
LLL	1	385.772367	1.216308	279.746436

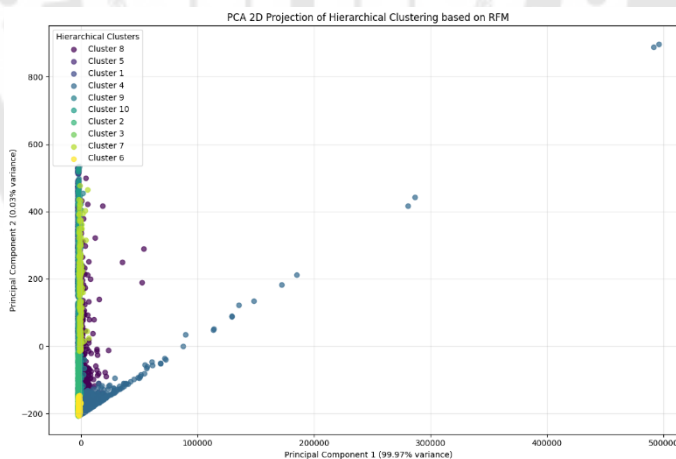
ภาพประกอบ 33 แสดงค่าเฉลี่ย RFM ของกลุ่มลูกค้า K-Means

4.3 ผลลัพธ์การจัดกลุ่มด้วยแบบจำลอง Hierarchical Clustering

จากการสร้างแบบจำลอง Hierarchical Clustering พบว่าการจัดกลุ่มลูกค้าด้วยจำนวนกลุ่มที่เหมาะสมที่สุด (K) ด้วยการสร้าง Dendrogram ซึ่งได้ค่า K เท่ากับ 10 ทำให้ได้ค่า Silhouette score เท่ากับ 0.352 และ Davies-Bouldin Index เท่ากับ 1.148 หมายความว่า การแบ่งกลุ่มมีความยังมีความคล้ายคลึงกันอยู่บ้าง และอาจมีบางจุดที่ถูกจัดอยู่ในกลุ่มที่ไม่เหมาะสมตามภาพประกอบที่ 34 และ 35



ภาพประกอบ 34 แสดงผลลัพธ์การจัดกลุ่ม Hierarchical Clustering ในรูปแบบ 3 มิติ



ภาพประกอบ 35 แสดงผลลัพธ์การจัดกลุ่ม Hierarchical Clustering ในรูปแบบ 2 มิติ

หลังจากจัดกลุ่มแล้ว ผู้วิจัยได้นำผลลัพธ์ที่ได้จากการจัดกลุ่มมาทำการวิเคราะห์พฤติกรรม การซื้อของลูกค้าจากการใช้เทคนิค RFM โดยการแบ่งลูกค้าออกเป็นกลุ่มต่าง ๆ ตามคะแนนที่ได้รับในแต่ละด้าน โดยใช้ค่าเฉลี่ยของแต่ละตัวแปรเป็นเกณฑ์ในการจัดกลุ่ม หากคะแนนของลูกค้าในแต่ละด้านสูงกว่าค่าเฉลี่ยที่คำนวณจากกลุ่มลูกค้าทั้งหมด จะถูกจัดให้เป็น High (H) และหากต่ำกว่าค่าเฉลี่ยจะถูกจัดให้เป็น Low (L) ซึ่งแสดงให้เห็นลักษณะของลูกค้าแต่ละกลุ่มตาม

พฤติกรรมที่ซื้อได้ โดยสามารถจัดกลุ่มตามหลัก RFM Analysis ได้ 6 กลุ่ม ตามภาพประกอบที่ 36 และ 37

cluster_hrc	count	recency_mean	frequency_mean	monetary_mean	RecencyScore_mean	FrequencyScore_mean	MonetaryScore_mean	Hierarchical
1	1362	449.910426	1.221733	246.613510	1.475771	1.485316	1.371512	LLL
2	400	78.972500	1.232500	329.857000	3.430000	1.310000	1.637500	HLL
3	309	21.779935	2.113269	336.881311	4.605178	2.339806	1.585761	HLL
4	898	18.340757	21.930958	11494.109943	4.640312	5.000000	4.807350	HHH
5	512	24.574219	5.576172	2610.089104	4.464844	3.769531	4.269531	HHH
6	456	25.311404	3.717105	927.541614	4.462719	3.129386	3.046053	HHH
7	257	375.245136	4.509728	1871.929455	1.610895	3.533074	3.980545	LHH
8	532	151.488722	9.857143	4290.945274	2.742481	4.458647	4.539474	LHH
9	341	316.853372	1.612903	1098.023003	1.982405	1.653959	3.266862	LLH
10	811	258.699137	3.655980	804.473158	2.282367	3.295931	2.684340	LHL

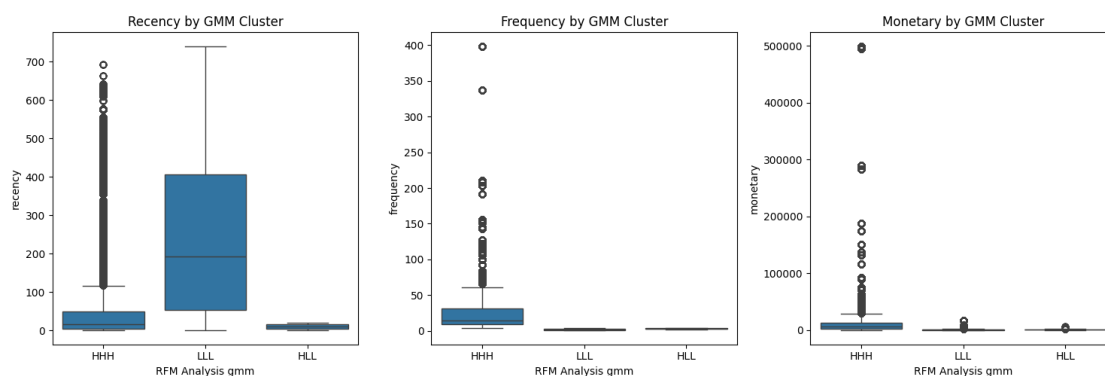
ภาพประกอบ 36 แสดงผลลัพธ์การวิเคราะห์พฤติกรรมของกลุ่มลูกค้า Hierarchical ด้วย RFM Analysis

	count	recency_mean	frequency_mean	monetary_mean
RFM Analysis Hierarchical				
HHH	3	22.742126	10.408078	5010.580220
HLL	2	50.376218	1.672884	333.369155
LHH	2	263.366929	7.183435	3081.437365
LHL	1	258.699137	3.655980	804.473158
LLH	1	316.853372	1.612903	1098.023003
LLL	1	449.910426	1.221733	246.613510

ภาพประกอบ 37 แสดงค่าเฉลี่ย RFM ของกลุ่มลูกค้า Hierarchical

4.4 เปรียบเทียบผลลัพธ์การจัดกลุ่ม

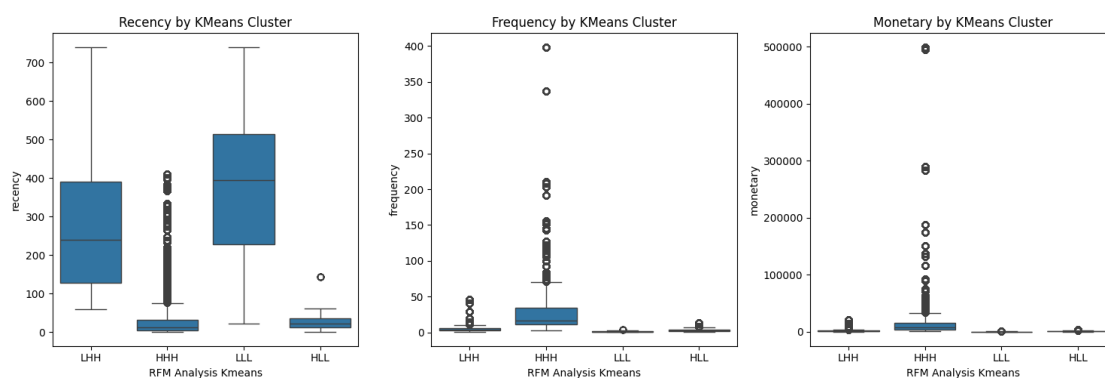
จากผลการประเมินประสิทธิภาพแบบจำลอง พบว่า K-Means Clustering เป็นวิธีที่มีประสิทธิภาพสูงสุด เนื่องจากสามารถแยกกลุ่มลูกค้าได้ชัดเจนที่สุดตามค่า Silhouette Score และ Davies-Bouldin Index รวมถึงจำนวนกลุ่มที่สอดคล้องกับ RFM Analysis มากที่สุด ในขณะที่ Gaussian Mixture Model (GMM) ให้ผลลัพธ์ที่ไม่ดีนักเนื่องจากการทับซ้อนของกลุ่มสูง ส่วน Hierarchical Clustering แม้จะช่วยให้เห็นโครงสร้างของข้อมูลได้ดี แต่การนำไปใช้งานจริงอาจไม่เหมาะสมเท่ากับ K-Means โดยสามารถเปรียบเทียบลักษณะของลูกค้าแต่ละกลุ่มตามหลัก RFM Analysis ตามภาพประกอบ 38 ถึง 40



ภาพประกอบ 38 ลักษณะของลูกค้าแต่ละกลุ่มของ GMM ตามหลัก RFM Analysis

Gaussian Mixture Model (GMM) ได้จำแนกกลุ่มที่เหมาะสม 6 กลุ่ม ตามการคำนวณเชิงสถิติ แต่เมื่อนำไปใช้กับ RFM Analysis เหลือเพียง 3 กลุ่ม โดยมีลักษณะของกลุ่มลูกค้า ดังนี้

- กลุ่ม HHH มีค่าเฉลี่ย Recency ต่ำ แสดงให้เห็นว่าลูกค้ามีการซื้อสินค้าล่าสุดไม่นาน มีค่าเฉลี่ย Frequency สูง และกระจายมากกว่ากลุ่มอื่น แสดงให้เห็นว่าลูกค้าในกลุ่มนี้ซื้อสินค้าบ่อย และมีค่าเฉลี่ย Monetary สูงมาก และกระจายตัวสูง แสดงให้เห็นว่าลูกค้ากลุ่มนี้ใช้จ่ายเงินเยอะกว่ากลุ่มอื่น
- กลุ่ม LLL มีค่าเฉลี่ย Recency สูง และช่วงค่ากระจายกว้าง แสดงให้เห็นว่าลูกค้าในกลุ่มนี้ส่วนใหญ่ไม่ได้ซื้อสินค้ามานาน มีค่าเฉลี่ย Frequency ต่ำมาก แสดงให้เห็นว่าลูกค้าในกลุ่มนี้มีแนวโน้มที่จะซื้อของไม่บ่อย และมีค่าเฉลี่ย Monetary ต่ำมาก แสดงให้เห็นว่าลูกค้าในกลุ่มนี้ไม่ได้ใช้จ่ายมาก
- กลุ่ม HLL มีค่าเฉลี่ย Recency ต่ำ แสดงให้เห็นว่าลูกค้ามีการซื้อสินค้าล่าสุดไม่นาน มีค่าเฉลี่ย Frequency ต่ำมาก แสดงให้เห็นว่าลูกค้าในกลุ่มนี้มีแนวโน้มที่จะซื้อของไม่บ่อย และมีค่าเฉลี่ย Monetary ต่ำมาก แสดงให้เห็นว่าลูกค้าในกลุ่มนี้ไม่ได้ใช้จ่ายมาก



ภาพประกอบ 39 ลักษณะของลูกค้าแต่ละกลุ่มของ K-Means ตามหลัก RFM Analysis

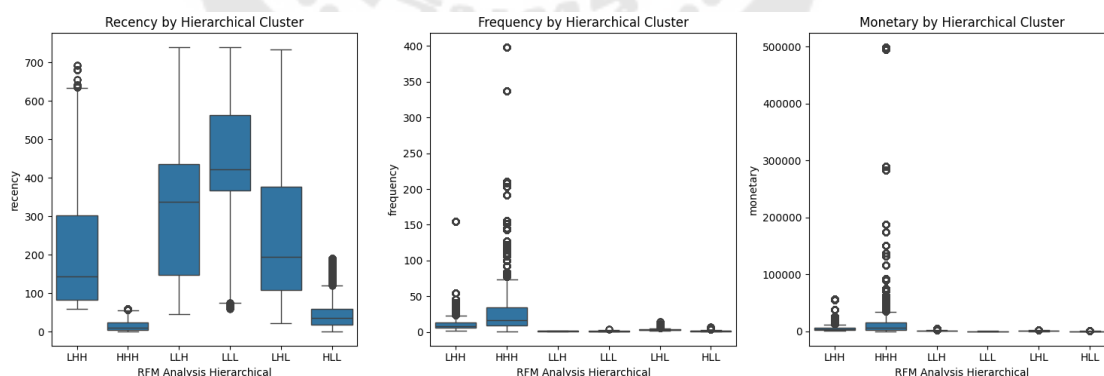
K-Means Clustering ได้จำนวนกลุ่มที่เหมาะสมคือ 4 กลุ่ม และสอดคล้องกับ RFM Analysis ที่ได้ 4 กลุ่ม เช่นกัน โดยมีลักษณะของกลุ่มลูกค้า ดังนี้

- กลุ่ม LHH มีค่าเฉลี่ย Recency สูง และกระจายตัวกว้าง แสดงให้เห็นว่าลูกค้าในกลุ่มนี้ส่วนใหญ่ไม่ได้ซื้อสินค้ามานาน มีค่าเฉลี่ย Frequency สูง และมีการกระจายตัว แสดงให้เห็นว่าลูกค้าในกลุ่มนี้ซื้อสินค้าบ่อย และมีค่าเฉลี่ย Monetary สูง และมีการกระจายตัว แสดงให้เห็นว่าลูกค้ากลุ่มนี้ใช้จ่ายเงินเยอะกว่ากลุ่มอื่น

- กลุ่ม HHH มีค่าเฉลี่ย Recency ต่ำ แสดงให้เห็นว่าลูกค้ามีการซื้อสินค้าล่าสุดไม่นาน มีค่าเฉลี่ย Frequency สูง และกระจายมากกว่ากลุ่มอื่น แสดงให้เห็นว่าลูกค้าในกลุ่มนี้ซื้อสินค้าบ่อย และมีค่าเฉลี่ย Monetary สูง และกระจายตัวสูง แสดงให้เห็นว่าลูกค้ากลุ่มนี้ใช้จ่ายเงินเยอะกว่ากลุ่มอื่น

- กลุ่ม LLL มีค่าเฉลี่ย Recency สูงมาก และช่วงค่ากระจายกว้าง แสดงให้เห็นว่าลูกค้าในกลุ่มนี้ส่วนใหญ่ไม่ได้ซื้อสินค้ามานาน มีค่าเฉลี่ย Frequency ต่ำมาก แสดงให้เห็นว่าลูกค้าในกลุ่มนี้มีแนวโน้มที่จะซื้อของไม่บ่อย และมีค่าเฉลี่ย Monetary ต่ำมาก แสดงให้เห็นว่าลูกค้าในกลุ่มนี้ไม่ได้ใช้จ่ายมาก

- กลุ่ม HLL มีค่าเฉลี่ย Recency ต่ำ แสดงให้เห็นว่าลูกค้ามีการซื้อสินค้าล่าสุดไม่นาน มีค่าเฉลี่ย Frequency ต่ำ แสดงให้เห็นว่าลูกค้าในกลุ่มนี้มีแนวโน้มที่จะซื้อของไม่บ่อย และมีค่าเฉลี่ย Monetary ต่ำ แสดงให้เห็นว่าลูกค้าในกลุ่มนี้ไม่ได้ใช้จ่ายมาก



ภาพประกอบ 40 ลักษณะของลูกค้าแต่ละกลุ่มของ Hierarchical ตามหลัก RFM Analysis

Hierarchical Clustering ได้จำนวนกลุ่มที่เหมาะสมได้ 10 กลุ่ม แต่เมื่อนำไปใช้จริงจาก RFM Analysis ได้ 6 กลุ่ม โดยมีลักษณะของกลุ่มลูกค้า ดังนี้

- กลุ่ม LHH มีค่าเฉลี่ย Recency สูง และกระจายตัวกว้าง แสดงให้เห็นว่าลูกค้าในกลุ่มนี้ส่วนใหญ่ไม่ได้ซื้อสินค้ามานาน มีค่าเฉลี่ย Frequency สูง และมีการกระจายตัว แสดงให้เห็นว่าลูกค้าในกลุ่มนี้ซื้อสินค้าบ่อย และมีค่าเฉลี่ย Monetary สูง และมีการกระจายตัว แสดงให้เห็นว่าลูกค้ากลุ่มนี้ใช้จ่ายเงินเยอะกว่ากลุ่มอื่น

- กลุ่ม HHH มีค่าเฉลี่ย Recency ต่ำ แสดงให้เห็นว่าลูกค้ามีการซื้อสินค้าล่าสุดไม่นาน มีค่าเฉลี่ย Frequency สูง และกระจายมากกว่ากลุ่มอื่น แสดงให้เห็นว่าลูกค้าในกลุ่มนี้ซื้อสินค้าบ่อย และมีค่าเฉลี่ย Monetary สูง และกระจายตัวสูง แสดงให้เห็นว่าลูกค้ากลุ่มนี้ใช้จ่ายเงินเยอะกว่ากลุ่มอื่น

- กลุ่ม LLH มีค่าเฉลี่ย Recency สูง และช่วงค่ากระจายกว้าง แสดงให้เห็นว่าลูกค้าในกลุ่มนี้ส่วนใหญ่ไม่ได้ซื้อสินค้ามานาน มีค่าเฉลี่ย Frequency ต่ำมาก แสดงให้เห็นว่าลูกค้าในกลุ่มนี้มีแนวโน้มที่จะซื้อของไม่บ่อย และมีค่าเฉลี่ย Monetary สูงเล็กน้อย แสดงให้เห็นว่าลูกค้ากลุ่มนี้ใช้จ่ายเงินสูงกว่าค่าเฉลี่ย

- กลุ่ม LLL มีค่าเฉลี่ย Recency สูงมาก และช่วงค่ากระจายกว้าง แสดงให้เห็นว่าลูกค้าในกลุ่มนี้ส่วนใหญ่ไม่ได้ซื้อสินค้ามานาน มีค่าเฉลี่ย Frequency ต่ำมาก แสดงให้เห็นว่าลูกค้าในกลุ่มนี้มีแนวโน้มที่จะซื้อของไม่บ่อย และมีค่าเฉลี่ย Monetary ต่ำมาก แสดงให้เห็นว่าลูกค้าในกลุ่มนี้ไม่ได้ใช้จ่ายมาก

- กลุ่ม LHL มีค่าเฉลี่ย Recency สูงมาก และช่วงค่ากระจายกว้าง แสดงให้เห็นว่าลูกค้าในกลุ่มนี้ส่วนใหญ่ไม่ได้ซื้อสินค้ามานาน มีค่าเฉลี่ย Frequency สูง และมีการกระจายตัว แสดงให้เห็นว่าลูกค้าในกลุ่มนี้ซื้อสินค้าบ่อย และมีค่าเฉลี่ย Monetary ต่ำ แสดงให้เห็นว่าลูกค้าในกลุ่มนี้ไม่ได้ใช้จ่ายมาก

- กลุ่ม HLL มีค่าเฉลี่ย Recency ค่อนข้างต่ำ แสดงให้เห็นว่าลูกค้ามีการซื้อสินค้าล่าสุดไม่นาน มีค่าเฉลี่ย Frequency ต่ำ แสดงให้เห็นว่าลูกค้าในกลุ่มนี้มีแนวโน้มที่จะซื้อของไม่บ่อย และมีค่าเฉลี่ย Monetary ต่ำ แสดงให้เห็นว่าลูกค้าในกลุ่มนี้ไม่ได้ใช้จ่ายมาก

4.5 ผลการวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis)

ก่อนทำการวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis) ผู้วิจัยได้ทำการวิเคราะห์กลุ่มสินค้าตามหลักการของ ABC Analysis โดยทำการแบ่งกลุ่มสินค้าออกเป็น 3 กลุ่ม ประกอบด้วยกลุ่ม A, B และ C ทำให้ได้สินค้าที่มียอดขายต่ำ (กลุ่ม C) คิดเป็นร้อยละ 5 จากยอดขายรวมทั้งหมด จำนวน 2,409 ชิ้น และสินค้าที่มียอดขายสูง (กลุ่ม A และ B) คิดเป็นร้อยละ

95 ของยอดขายรวมทั้งหมด จำนวน 2,222 ชิ้น โดยการนำหลัก ABC Analysis มาใช้ก่อนการวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis) ช่วยให้มุ่งเน้นสินค้าที่มีความสำคัญต่อธุรกิจ และเพิ่มประสิทธิภาพของการวิเคราะห์กฎความสัมพันธ์ ได้ผลลัพธ์ตามตาราง 4

ตาราง 4 ผลลัพธ์การวิเคราะห์กลุ่มสินค้าตามหลักการ RFM Analysis

กลุ่มสินค้า	จำนวนสินค้า	จำนวนที่ขายได้	ยอดขาย	คิดเป็นร้อยละ
A	1,014	6,553,490	13,153,860	80
B	1,208	1,990,020	2,466,827	15
C	2,409	838,078	822,997	5

ต่อมา ผู้วิจัยได้นำกลุ่มลูกค้าที่ได้จากการจัดกลุ่มด้วยแบบจำลอง K-Means มาทำการวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis) เพื่อตรวจสอบความสัมพันธ์ระหว่างสินค้าที่ลูกค้าแต่ละกลุ่มมักจะซื้อ โดยใช้ FP-Growth Algorithm เพื่อหาชุดสินค้าที่มักจะถูกซื้อพร้อมกัน กำหนดให้ Minimum Support เท่ากับ 0.05 หมายความว่าชุดสินค้ารวมกันที่เกิดขึ้นต้องมีความถี่เกิดขึ้นในข้อมูลอย่างน้อย 5% จากลูกค้าทั้งหมด หลังจากนั้นหากฎความสัมพันธ์ระหว่างสินค้าที่ซื้อพร้อมกันด้วย Association rules โดยกำหนดให้ Minimum Confidence เท่ากับ 0.7 หมายความว่าความเชื่อมั่นในกฎนั้นต้องมีค่าอย่างน้อย 70% เพื่อให้กฎนั้นเป็นที่ยอมรับได้ ทำให้ได้จำนวนกฎความสัมพันธ์แต่ละกลุ่มลูกค้าตามตาราง 5 และตัวอย่างกฎความสัมพันธ์ของลูกค้าแต่ละกลุ่มตามตาราง 6

ตาราง 5 จำนวนกฎความสัมพันธ์สำหรับลูกค้าแต่ละกลุ่ม

RFM Group	จำนวนกฎความสัมพันธ์ที่ได้
HHH	528,814
LHH	14
HLL	5
LLL	0

ตาราง 6 ตัวอย่างกฎความสัมพันธ์ที่ได้ของลูกค้าแต่ละกลุ่ม

RFM Group	Antecedents	Consequents	support	confidence	lift
HHH	(WICKER WREATH SMALL)	(HEART OF WICKER SMALL)	0.053	0.825	2.719
HHH	(72 SWEETHEART FAIRY CAKE CASES)	(60 TEATIME FAIRY CAKE CASES)	0.167	0.702	2.441
LHH	(GREEN REGENCY TEACUP AND SAUCER)	(ROSES REGENCY TEACUP AND SAUCER)	0.062	0.849	10.571
LHH	(RED HANGING HEART T-LIGHT HOLDER)	(WHITE HANGING HEART T-LIGHT HOLDER)	0.101	0.847	3.001
HLL	(BAKING SET SPACEBOY DESIGN)	(BAKING SET 9 PIECE RETROSPOT)	0.078	0.826	4.421
HLL	(WOODEN TREE CHRISTMAS SCANDINAVIAN)	(WOODEN STAR CHRISTMAS SCANDINAVIAN)	0.067	0.880	6.593

บทที่ 5

และข้อเสนอแนะ สรุปผลการวิจัย อภิปรายผล

งานวิจัยนี้ได้ทำการศึกษาศึกษาการวิเคราะห์พฤติกรรมลูกค้าของร้านค้าปลีกออนไลน์แห่งหนึ่งจากแหล่งข้อมูลสาธารณะ Kaggle.com ด้วยเทคนิค RFM และจัดกลุ่มลูกค้าด้วยเทคนิคการเรียนรู้ของเครื่องแบบไม่มีผู้สอน (Unsupervised Machine Learning) จากนั้นทำการประเมินผลลัพธ์การจัดกลุ่มลูกค้า วิเคราะห์กลุ่มสินค้าตามหลักการของ ABC Analysis และใช้เทคนิคการวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis) โดยสามารถสรุปผลได้ดังต่อไปนี้

1. สรุปผลการวิจัย
2. อภิปรายผลการวิจัย
3. ข้อเสนอแนะ

5.1 สรุปผลการวิจัย

ผู้วิจัยได้ทำการจัดกลุ่มลูกค้าด้วยเทคนิคการเรียนรู้ของเครื่องแบบไม่มีผู้สอน (Unsupervised Machine Learning) โดยใช้แบบจำลอง 3 แบบจำลอง ได้แก่ Gaussian Mixture Model (GMM) หาจำนวนกลุ่มที่เหมาะสมที่สุด (K) ด้วยเกณฑ์ BIC (Bayesian Information Criterion) และ AIC (Akaike Information Criterion) K-Means Clustering หาจำนวนกลุ่มที่เหมาะสมที่สุด (K) ด้วยวิธี Elbow Method และ Hierarchical Clustering หาจำนวนกลุ่มที่เหมาะสมที่สุด (K) ด้วยการสร้าง Dendrogram

หลังจากจัดกลุ่มด้วยแบบจำลองเรียบร้อยแล้ว ทำการประเมินประสิทธิภาพแบบจำลองด้วยค่า Silhouette score และ Davies-Bouldin Index หลังจากนั้นเพื่อที่จะทำความเข้าใจลักษณะของลูกค้าแต่ละกลุ่ม ผู้วิจัยได้ทำการวิเคราะห์พฤติกรรมการซื้อของลูกค้าแต่ละกลุ่มด้วยเทคนิค RFM

จากผลการประเมินประสิทธิภาพแบบจำลอง พบว่า K-Means Clustering เป็นวิธีที่มีประสิทธิภาพสูงสุด เนื่องจากสามารถแยกกลุ่มลูกค้าได้ชัดเจนที่สุดตามค่า Silhouette Score และ Davies-Bouldin Index รวมถึงจำนวนกลุ่มที่สอดคล้องกับ RFM Analysis มากที่สุด และสามารถนำผลการจัดกลุ่มไปวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis) เพื่อสร้างกลยุทธ์การตลาดเฉพาะกลุ่มได้ โดยสามารถเปรียบเทียบผลลัพธ์ได้ตามตาราง 6

ตาราง 7 เปรียบเทียบผลลัพธ์การจัดกลุ่ม

แบบจำลอง	Optimal Clusters	Clusters (RFM Analysis)	Silhouette Score	Davies-Bouldin Index
Gaussian Mixture Model	6	3	0.165	1.826
K-Means Clustering	4	4	0.398	0.956
Hierarchical Clustering	10	6	0.352	1.148

จากนั้น ทำการวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis) สำหรับกลุ่มลูกค้าที่ได้จากการจัดกลุ่มด้วยแบบจำลอง K-Means เพื่อหาความสัมพันธ์ระหว่างสินค้าที่ซื้อพร้อมกันด้วย ทำให้ได้ความสัมพันธ์ จำนวน 528,833 กฎ โดยสามารถนำกฎความสัมพันธ์ดังกล่าวมาสร้างโปรโมชั่นเฉพาะกลุ่มลูกค้าได้ ตัวอย่างตามตาราง 8

ตาราง 8 ตัวอย่างโปรโมชั่นสำหรับลูกค้าแต่ละกลุ่ม

RFM Group	Product	Promoted	support	confidence	lift
HHH	WATERING CAN GREEN DINOSAUR	WATERING CAN PINK BUNNY	0.056	0.824	6.882
HHH	WATERING CAN GREEN DINOSAUR	WATERING CAN BLUE ELEPHANT	0.068	0.750	7.037
HHH	WATERING CAN GREEN DINOSAUR	WATERING CAN PINK BUNNY	0.071	0.785	6.553
HHH	WATERING CAN GREEN DINOSAUR	WATERING CAN BLUE ELEPHANT	0.056	0.788	7.390
HLL	WOODEN TREE CHRISTMAS SCANDINAVIAN	WOODEN STAR CHRISTMAS SCANDINAVIAN	0.051	0.853	5.231
HLL	BAKING SET SPACEBOY DESIGN	PAPER CHAIN KIT 50'S CHRISTMAS	0.052	0.806	2.634

ตาราง 8 (ต่อ)

RFM Group	Product	Promoted	support	confidence	lift
HLL	BAKING SET SPACEBOY DESIGN	BAKING SET 9 PIECE RETROSPOT	0.053	0.832	2.556
HLL	WOODEN TREE CHRISTMAS SCANDINAVIAN	WOODEN STAR CHRISTMAS SCANDINAVIAN	0.068	0.880	6.593
HLL	WOODEN TREE CHRISTMAS SCANDINAVIAN	WOODEN HEART CHRISTMAS SCANDINAVIAN	0.051	0.863	4.699
LHH	PINK REGENCY TEACUP AND SAUCER	REGENCY TEA PLATE ROSES , GREEN REGENCY TEACUP AND SAUCER	0.063	0.927	12.016
LHH	PINK REGENCY TEACUP AND SAUCER	REGENCY TEA PLATE GREEN , GREEN REGENCY TEACUP AND SAUCER	0.060	0.922	11.864
LHH	PINK REGENCY TEACUP AND SAUCER	REGENCY TEA PLATE GREEN , REGENCY TEA PLATE ROSES	0.058	0.885	11.110
LHH	PINK REGENCY TEACUP AND SAUCER	REGENCY TEA PLATE GREEN , ROSES REGENCY TEACUP AND SAUCER	0.053	0.750	9.886
LHH	PINK REGENCY TEACUP AND SAUCER	ROSES REGENCY TEACUP AND SAUCER	0.117	0.949	5.064
HHH	WATERING CAN GREEN DINOSAUR	WATERING CAN PINK BUNNY	0.056	0.824	6.882

5.2 อภิปรายผลการวิจัย

งานวิจัยนี้ได้ทำการศึกษากลุ่มลูกค้าด้วยแบบจำลอง Gaussian Mixture Model (GMM), K-Means Clustering และ Hierarchical Clustering จากนั้นตรวจสอบลักษณะของกลุ่มลูกค้าตามหลัก RFM Analysis เพื่อทราบพฤติกรรมการซื้อของกลุ่มลูกค้า ทำให้ได้ผลการวิจัยดังนี้

แบบจำลอง Gaussian Mixture Model (GMM) ได้จำนวนกลุ่มที่เหมาะสมที่สุด (K) ด้วยเกณฑ์ BIC (Bayesian Information Criterion) และ AIC (Akaike Information Criterion) เท่ากับ 6 ทำให้ได้ค่า Silhouette score เท่ากับ 0.165 หมายความว่ากลุ่มที่ได้อาจมีความทับซ้อนกันมาก และ Davies-Bouldin Index เท่ากับ 1.826 หมายความว่ากลุ่มที่แบ่งยังมีความคล้ายกันสูง ทำให้การแยกกลุ่มไม่ดีนัก และเมื่อนำไปใช้กับ RFM Analysis เหลือเพียง 3 กลุ่ม ตามตาราง 9 ตาราง 9 แสดงผลลัพธ์การจัดกลุ่มตามหลัก RFM Analysis ของ GMM

กลุ่มตามหลัก RFM Analysis	จำนวนกลุ่ม ตาม GMM	จำนวน ลูกค้า	ค่าเฉลี่ย Recency	ค่าเฉลี่ย Frequency	ค่าเฉลี่ย Monetary
HHH	2	2,351	88.377	12.879	6,043.283
HLL	1	193	10.466	3.166	1,061.708
LLL	3	3,334	289.331	1.889	630.597

- กลุ่ม HHH มีลูกค้าจำนวน 2,351 ราย ค่า Recency เฉลี่ย 88.377 แสดงว่าลูกค้ามีการซื้อสินค้าล่าสุดห่างกับครั้งก่อนไม่นาน ค่า Frequency เฉลี่ย 12.879 แสดงว่าลูกค้ามีความถี่ในการซื้อสินค้าบ่อย และมีค่า Monetary เฉลี่ย 6,043.283 แสดงว่าลูกค้ามีการใช้จ่ายสูง ซึ่งถือว่าเป็นลูกค้าที่สำคัญต่อธุรกิจและต้องรักษาลูกค้ากลุ่มนี้ไว้

- กลุ่ม HLL มีลูกค้าจำนวน 193 ราย ค่า Recency เฉลี่ย 10.466 แสดงว่าลูกค้ามีการซื้อสินค้าล่าสุดห่างกับครั้งก่อนไม่นาน ค่า Frequency เฉลี่ย 3.166 แสดงว่าลูกค้ามีความถี่ในการซื้อสินค้าปานกลาง และมีค่า Monetary เฉลี่ย 1,061.708 แสดงว่าลูกค้ามีการใช้จ่ายปานกลาง ซึ่งถือว่าเป็นลูกค้าเป้าหมายที่อาจเป็นลูกค้าสำคัญในอนาคต

- กลุ่ม LLL มีลูกค้าจำนวน 3,334 ราย ค่า Recency เฉลี่ย 289.331 แสดงว่าลูกค้าไม่ได้มาซื้อสินค้าจากครั้งล่าสุดนานแล้ว ค่า Frequency เฉลี่ย 1.889 แสดงว่าลูกค้ามีความถี่ในการซื้อสินค้าไม่บ่อย และมีค่า Monetary เฉลี่ย 630.597 แสดงว่าลูกค้าไม่ได้ใช้จ่ายมาก ซึ่งถือว่าเป็นลูกค้าทั่วไปที่ไม่ค่อยสำคัญต่อธุรกิจ

แบบจำลอง K-Means Clustering ได้จำนวนกลุ่มที่เหมาะสมที่สุด (K) ด้วยวิธี Elbow Method เท่ากับ 4 ทำให้ได้ค่า Silhouette score เท่ากับ 0.398 และ Davies-Bouldin Index เท่ากับ 0.956 หมายความว่ากลุ่มที่ได้มีความแยกกันอย่างชัดเจน เหมาะกับการใช้งานจริงมากที่สุด และสอดคล้องกับ RFM Analysis ที่ได้ 4 กลุ่มเช่นกัน ตามตาราง 10

ตาราง 10 แสดงผลลัพธ์การจัดกลุ่มตามหลัก RFM Analysis ของ K-Means

กลุ่มตามหลัก RFM Analysis	จำนวนกลุ่ม ตาม K-Means	จำนวน ลูกค้า	ค่าเฉลี่ย Recency	ค่าเฉลี่ย Frequency	ค่าเฉลี่ย Monetary
HHH	1	1,595	37.866	16.159	8,212.791
HLL	1	974	25.869	2.916	761.576
LHH	1	1,543	272.007	4.025	1,366.478
LLL	1	1,766	385.772	1.216	279.746

- กลุ่ม HHH มีลูกค้าจำนวน 1,595 ราย ค่า Recency เฉลี่ย 37.866 แสดงว่าลูกค้ามีการซื้อสินค้าล่าสุดห่างกับครั้งก่อนไม่นาน ค่า Frequency เฉลี่ย 16.159 แสดงว่าลูกค้ามีความถี่ในการซื้อสินค้าบ่อย และมีค่า Monetary เฉลี่ย 8,212.791 แสดงว่าลูกค้ามีการใช้จ่ายสูง ซึ่งถือว่าเป็นลูกค้าที่สำคัญต่อธุรกิจและต้องรักษาลูกค้ากลุ่มนี้ไว้

- กลุ่ม HLL มีลูกค้าจำนวน 974 ราย ค่า Recency เฉลี่ย 25.868 แสดงว่าลูกค้ามีการซื้อสินค้าล่าสุดห่างกับครั้งก่อนไม่นาน ค่า Frequency เฉลี่ย 2.916 แสดงว่าลูกค้ามีความถี่ในการซื้อสินค้าไม่บ่อยนัก และมีค่า Monetary เฉลี่ย 761.576 แสดงว่าลูกค้ามีการใช้จ่ายปานกลาง ซึ่งถือว่าเป็นลูกค้าใหม่ที่เพิ่งมาเริ่มใช้บริการ

- กลุ่ม LHH มีลูกค้าจำนวน 1,543 ราย ค่า Recency เฉลี่ย 272.007 แสดงว่าลูกค้ามีการซื้อสินค้าล่าสุดห่างกับครั้งก่อนค่อนข้างนาน ค่า Frequency เฉลี่ย 4.025 แสดงว่าลูกค้ามีความถี่ในการซื้อสินค้าปานกลาง และมีค่า Monetary เฉลี่ย 1,366.478 แสดงว่าลูกค้ามีการใช้จ่ายค่อนข้างสูง ซึ่งถือว่าเป็นลูกค้าสำคัญที่เริ่มหายไปและควรติดตามกลับมา

- กลุ่ม LLL มีลูกค้าจำนวน 1,766 ราย ค่า Recency เฉลี่ย 385.772 แสดงว่าลูกค้าไม่ได้มาซื้อสินค้าจากครั้งล่าสุดนานแล้ว ค่า Frequency เฉลี่ย 1.216 แสดงว่าลูกค้ามีความถี่ในการซื้อสินค้าไม่บ่อย และมีค่า Monetary เฉลี่ย 279.746 แสดงว่าลูกค้าไม่ได้ใช้จ่ายมาก ซึ่งถือว่าเป็นลูกค้าทั่วไปที่ไม่ค่อยสำคัญต่อธุรกิจ

แบบจำลอง Hierarchical Clustering ได้จำนวนกลุ่มที่เหมาะสมที่สุด (K) ด้วยการสร้าง Dendrogram เท่ากับ 10 ทำให้ได้ค่า Silhouette score เท่ากับ 0.352 และ Davies-Bouldin Index เท่ากับ 1.148 หมายความว่า การแบ่งกลุ่มมีความยังมีความคล้ายคลึงกันอยู่บ้าง และอาจมีบางจุดที่ถูกจัดอยู่ในกลุ่มที่ไม่เหมาะสม และเมื่อนำไปใช้จริงจาก RFM Analysis ได้ 6 กลุ่มตามตาราง 9

ตาราง 11 แสดงผลลัพธ์การจัดกลุ่มตามหลัก RFM Analysis ของ Hierarchical

กลุ่มตามหลัก RFM Analysis	จำนวนกลุ่มตาม Hierarchical	จำนวน ลูกค้า	ค่าเฉลี่ย Recency	ค่าเฉลี่ย Frequency	ค่าเฉลี่ย Monetary
HHH	3	1,866	22.742	10.408	5,010.580
HLL	2	709	50.376	1.673	333.369
LHH	2	789	263.367	7.183	3,081.437
LHL	1	811	258.699	3.656	804.473
LLH	1	341	316.853	1.613	1,098.023
LLL	1	1,362	449.910	1.222	246.614

- กลุ่ม HHH มีลูกค้าจำนวน 1,866 ราย ค่า Recency เฉลี่ย 22.742 แสดงว่าลูกค้ามีการซื้อสินค้าล่าสุดห่างกับครั้งก่อนไม่นาน ค่า Frequency เฉลี่ย 10.408 แสดงว่าลูกค้ามีความถี่ในการซื้อสินค้าบ่อย และมีค่า Monetary เฉลี่ย 5,010.580 แสดงว่าลูกค้ามีการใช้จ่ายสูง ซึ่งถือว่าเป็นลูกค้าที่สำคัญต่อธุรกิจและต้องรักษาลูกค้ากลุ่มนี้ไว้

- กลุ่ม HLL มีลูกค้าจำนวน 709 ราย ค่า Recency เฉลี่ย 50.376 แสดงว่าลูกค้ามีการซื้อสินค้าล่าสุดในระยะปานกลาง ค่า Frequency เฉลี่ย 1.673 แสดงว่ามีความถี่ในการซื้อน้อย และมีค่า Monetary เฉลี่ย 333.369 แสดงว่าลูกค้ามีการใช้จ่ายค่อนข้างต่ำ ซึ่งถือว่าเป็นลูกค้าที่ซื้อสินค้าบางช่วงเวลาและไม่ได้เป็นลูกค้าหลัก

- กลุ่ม LHH มีลูกค้าจำนวน 789 ราย ค่า Recency เฉลี่ย 263.367 แสดงว่าลูกค้าไม่ได้ซื้อสินค้ามาระยะหนึ่งแล้ว ค่า Frequency เฉลี่ย 7.183 แสดงว่าลูกค้ามีความถี่ในการซื้อสินค้าปานกลาง และมีค่า Monetary เฉลี่ย 3,081.437 แสดงว่าลูกค้ามีการใช้จ่ายค่อนข้างสูง ซึ่งถือว่าเป็นลูกค้าที่กลับมาซื้อเป็นครั้งคราวแต่มีมูลค่าการซื้อสูง

- กลุ่ม LHL มีลูกค้าจำนวน 811 ราย ค่า Recency เฉลี่ย 258.699 แสดงว่าลูกค้าไม่ได้ซื้อสินค้ามานานแล้ว ค่า Frequency เฉลี่ย 3.656 แสดงว่าลูกค้ามีความถี่ในการซื้อสินค้าปานกลาง และมีค่า Monetary เฉลี่ย 804.473 แสดงว่าลูกค้ามีการใช้จ่ายปานกลาง ซึ่งถือว่าเป็นลูกค้าที่เคยซื้อแต่ปัจจุบันไม่ได้กลับมาซื้อบ่อย

- กลุ่ม LLH มีลูกค้าจำนวน 341 ราย ค่า Recency เฉลี่ย 316.853 แสดงว่าลูกค้าไม่ได้ซื้อสินค้ามานาน ค่า Frequency เฉลี่ย 1.613 แสดงว่าลูกค้ามีความถี่ในการซื้อสินค้าน้อยมาก และค่า Monetary เฉลี่ย 1,098.023 แสดงว่าลูกค้ามีการใช้จ่ายปานกลางถึงสูง ซึ่งถือว่าเป็นลูกค้าที่ซื้อสินค้าราคาแพงแต่ไม่บ่อย

- กลุ่ม LLL มีลูกค้าจำนวน 1,362 ราย ค่า Recency เฉลี่ย 449.910 แสดงว่าลูกค้าไม่ได้ซื้อสินค้ามานานมาก ค่า Frequency เฉลี่ย 1.222 แสดงว่าลูกค้ามีความถี่ในการซื้อสินค้าน้อยมาก และค่า Monetary เฉลี่ย 246.613 แสดงว่าลูกค้ามีการใช้จ่ายที่ต่ำ ซึ่งถือว่าเป็นลูกค้าที่หายไปหรือเลิกใช้บริการแล้ว

เมื่อเปรียบเทียบผลจัดกลุ่มลูกค้าด้วยแบบจำลอง Gaussian Mixture Model (GMM), K-Means Clustering และ Hierarchical Clustering และใช้ RFM Analysis ในการวิเคราะห์พฤติกรรมของกลุ่มลูกค้า พบว่า K-Means Clustering ให้ผลลัพธ์ที่เหมาะสมที่สุด เนื่องจากมีค่าตัวชี้วัดของการแบ่งกลุ่มที่ดีที่สุด โดยมีค่า Silhouette score สูงสุด และค่า Davies-Bouldin Index ต่ำสุด และสามารถแบ่งกลุ่มลูกค้าได้ชัดเจนและสอดคล้องกับ RFM Analysis มากที่สุด

ถึงแม้ว่า GMM จะช่วยให้เห็นความน่าจะเป็นของการเป็นสมาชิกในแต่ละกลุ่ม แต่ยังพบว่ากลุ่มที่ได้มีความทับซ้อนกันสูง ทำให้ความแตกต่างระหว่างกลุ่มไม่ชัดเจน ส่วน Hierarchical Clustering สามารถระบุจำนวนกลุ่มที่เหมาะสมผ่าน Dendrogram แต่ยังมีปัญหาเรื่องกลุ่มที่คล้ายกันมาก ส่งผลให้การนำไปใช้งานจริงอาจไม่สะดวกเท่ากับ K-Means

ต่อจากนี้ การนำไปประยุกต์ใช้ในเชิงธุรกิจ K-Means Clustering เป็นวิธีที่เหมาะสมที่สุดสำหรับการจัดกลุ่มลูกค้าเพื่อนำไปใช้วางแผนกลยุทธ์ทางการตลาด เพื่อรักษาลูกค้าให้ใช้บริการอย่างต่อเนื่อง (Customer Retention) โดยผู้วิจัยใช้เทคนิคการวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis) เพื่อหาความสัมพันธ์ระหว่างสินค้าที่ซื้อร่วมกัน สำหรับการออกโปรโมชั่นสำหรับลูกค้าแต่ละกลุ่มได้ ตัวอย่างตามตาราง 12 ถึง 14

ตาราง 12 แสดงตัวอย่างโปรโมชั่นที่ได้สำหรับลูกค้าแต่ละกลุ่ม โดยเลือกจากระดับความเชื่อมั่น (Confidence) สูงสุด

RFM Group	Product	Category	Promoted	support	confidence	lift
HHH	WATERING CAN GREEN DINOSAUR	A	WATERING CAN PINK BUNNY	0.056	0.824	6.882
HLL	WOODEN TREE CHRISTMAS SCANDINAVIAN	B	WHITE HANGING HEART T-LIGHT HOLDER	0.050	0.952	2.208
LHH	PINK REGENCY TEACUP AND SAUCER	A	GREEN REGENCY TEACUP AND SAUCER	0.053	1.000	6.088

ตาราง 13 แสดงตัวอย่างโปรโมชั่นที่ได้สำหรับลูกค้าแต่ละกลุ่ม โดยเลือกจากจำนวนสินค้าที่ขายได้น้อยที่สุด

RFM Group	Product	Category	Promoted	support	confidence	lift
HHH	WATERING CAN GREEN DINOSAUR	A	WATERING CAN BLUE ELEPHANT	0.068	0.750	7.037
HLL	BAKING SET SPACEBOY DESIGN	A	BAKING SET 9 PIECE RETROSPOT	0.050	0.851	2.616
LHH	BAKING SET SPACEBOY DESIGN	A	BAKING SET 9 PIECE RETROSPOT	0.050	0.851	2.616

ตาราง 14 แสดงตัวอย่างโปรโมชันที่ได้สำหรับลูกค้าแต่ละกลุ่ม โดยเลือกจากยอดขายได้ที่ขายได้น้อยที่สุด

RFM Group	Product	Category	Promoted	support	confidence	lift
HHH	WATERING CAN	A	WATERING CAN	0.068	0.750	7.037
	GREEN DINOSAUR		BLUE ELEPHANT			
HLL	WOODEN TREE	B	WOODEN STAR	0.053	0.894	5.482
	CHRISTMAS		CHRISTMAS			
LHH	SCANDINAVIAN	A	SCANDINAVIAN	0.137	0.932	5.672
	PINK REGENCY		GREEN REGENCY			
LHH	TEACUP AND	A	TEACUP AND	0.137	0.932	5.672
	SAUCER		SAUCER			

ดังนั้น การวิเคราะห์ตะกร้าสินค้า (Marker Basket Analysis) เพื่อหาความสัมพันธ์ระหว่างสินค้าที่มักถูกซื้อร่วมกัน เพื่อนำมาออกแบบกลยุทธ์ทางการตลาด เช่น การทำโปรโมชัน Cross-Selling และ Up-Selling เพื่อเพิ่มยอดขายและความพึงพอใจของลูกค้า ร่วมกับการจัดกลุ่มลูกค้า (Clustering) เพื่อให้ได้กลยุทธ์ที่ตรงกับพฤติกรรมของลูกค้าแต่ละกลุ่มอย่างมีประสิทธิภาพ

5.3 ข้อเสนอแนะ

1. ปรับปรุงกระบวนการ Data Preprocessing โดยใช้เทคนิคอื่นแทนการใช้ RFM Score เพื่อเพิ่มประสิทธิภาพในการจัดกลุ่มลูกค้า เช่น การสร้างคุณลักษณะใหม่ (Feature Engineering) ที่สะท้อนพฤติกรรมลูกค้าได้ลึกซึ้งยิ่งขึ้น เช่น Customer Lifetime Value (CLV), Churn Rate การลดมิติข้อมูลด้วยเทคนิคเช่น PCA, t-SNE และ UMAP เพื่อให้เห็นโครงสร้างของข้อมูลชัดเจนขึ้น การจัดการ Outliers ด้วยวิธีเช่น IQR หรือ Z-Score เพื่อลดผลกระทบจากข้อมูลที่ผิดปกติ

2. Gaussian Mixture Model (GMM) เหมาะสำหรับข้อมูลที่มีการกระจายตัวแบบไม่เป็นทรงกลมและมีความเป็นไปได้ที่ข้อมูลจะอยู่ในลักษณะของการแจกแจงแบบ Gaussian หลายชุด อย่างไรก็ตาม ในการทดลองพบว่า GMM ให้ค่า Silhouette Score และ Davies-Bouldin Index ที่ไม่ดีนักเมื่อเทียบกับแบบจำลองอื่น อาจเกิดจากการที่ข้อมูลไม่ได้กระจายตัวแบบ Gaussian อย่างแท้จริง ควรทดลองใช้ PCA หรือ t-SNE เพื่อลดมิติของข้อมูลก่อนทำ GMM และพิจารณาปรับเปลี่ยนพารามิเตอร์ของแบบจำลองเพื่อดูผลลัพธ์ที่เหมาะสมขึ้น

3. K-Means Clustering เหมาะสำหรับข้อมูลที่กระจายตัวแบบเป็นกลุ่ม (Clusters) ที่มีลักษณะเป็นทรงกลม (Convex Shape) อย่างไรก็ดีตาม แม้ว่าจะเป็นแบบจำลองที่ให้ผลลัพธ์ที่ดีที่สุดในงานวิจัยนี้ แต่ค่า Silhouette Score และ Davies-Bouldin Index ยังสามารถปรับปรุงได้ ควรพิจารณาการ ลดมิติข้อมูล (Dimensionality Reduction) การเปลี่ยนระยะทางที่ใช้คำนวณ (Distance Metrics) เพื่อให้ผลลัพธ์มีความแม่นยำและสะท้อนพฤติกรรมของลูกค้าได้ดียิ่งขึ้น

4. Hierarchical Clustering เหมาะกับการวิเคราะห์โครงสร้างความสัมพันธ์ของข้อมูล โดยไม่ต้องกำหนดจำนวนกลุ่มล่วงหน้า อย่างไรก็ดีตาม ข้อจำกัดคือใช้ทรัพยากรมากเมื่อข้อมูลมีขนาดใหญ่ และจากผลลัพธ์พบว่าค่า Silhouette Score และ Davies-Bouldin Index ไม่ดี ควรพิจารณาใช้ Agglomerative Clustering โดยลองเปลี่ยน linkage method (Single, Complete, Average, Ward) และใช้ PCA เพื่อลดมิติข้อมูล ก่อนทำการจัดกลุ่มเพื่อเพิ่มประสิทธิภาพของแบบจำลอง

5. การวิเคราะห์ตะกร้าสินค้า (Market Basket Analysis) ใช้สำหรับวิเคราะห์ความสัมพันธ์ของสินค้าที่มักถูกซื้อพร้อมกัน เพื่อใช้ในการสร้างโปรโมชั่นและกลยุทธ์การตลาด อย่างไรก็ดีตาม จากการวิจัยพบว่า กฎความสัมพันธ์ที่ได้นั้นมีมากเกินไป ทำให้นำไปใช้จริงในธุรกิจไม่ได้ ควรพิจารณาการ ปรับค่าขั้นต่ำของ Support และ Confidence ให้เหมาะสมกับพฤติกรรมการซื้อของลูกค้า และอาจใช้เทคนิค Segmentation-Based Association Rule Mining เพื่อวิเคราะห์กฎความสัมพันธ์แยกตามกลุ่มลูกค้า เพื่อให้ผลลัพธ์สามารถนำไปใช้กับแต่ละกลุ่มได้อย่างมีประสิทธิภาพมากขึ้น

บรรณานุกรม

- Ahammed, K., Azad, R. U., Al Jaber, M., Rony, M. A. T., และ Ahammad, M. (2023). A comparative study of RFM-based clustering methods in customer segmentation. 2023 26th International Conference on Computer and Information Technology (ICCIT),
- Doğan, O., Ayçin, E., และ Bulut, Z. (2018). Customer segmentation by using RFM model and clustering methods: a case study in retail industry. *International Journal of Contemporary Economics and Administrative Sciences*, 8.
- John, J. M., Shobayo, O., และ Ogunleye, B. (2023). An exploration of clustering algorithms for customer segmentation in the UK retail market. *Analytics*, 2(4), 809-823.
- Maraghi, M., Adibi, M. A., และ Mehdizadeh, E. (2020). Using RFM model and market basket analysis for segmenting customers and assigning marketing strategies to resulted segments. *Journal of Applied Intelligent Systems and Information Sciences*, 1(1), 35-43.
- Maulidya, A., Sitorus, Z., Siahaan, A. P. U., และ Iqbal, M. (2024). Analysis Of Increasing Student Service Satisfaction Using K-Means Clustering Algorithm and Gaussian Mixture Models (GMM). *International Journal Of Computer Sciences and Mathematics Engineering*, 3(1), 29-35.

ประวัติผู้เขียน

