การศึกษาวิธีการแบ่งกลุ่มลูกค้าตามพฤติกรรมการซื้อสินค้าโดยใช้เทคนิคเคมีน

กมลทิพย์ มนตรีสา[1*], เรืองศักดิ์ ตระกูลพุทธิรักษ์[2]

บทคัดย่อ

การทำธุรกิจและการตลาดในยุคปัจจุบัน ควรคำนึงถึงความหลากหลายและความแตกต่างในพฤติกรรมของลูกค้า การเข้าใจลักษณะและพฤติกรรมการซื้อสินค้าของลูกค้าช่วยสร้างความเข้าใจในกลุ่มลูกค้า และนำข้อมูลเหล่านี้มาใช้ในการตัดสินใจทางธุรกิจและกำหนดกลยุทธ์การตลาดที่มีประสิทธิภาพ ตอบโจทย์ต่อความต้องการของลูกค้าอย่างเหมาะสม งานวิจัยนี้มีความมุ่งหมายในการศึกษาวิธีการแบ่งกลุ่มลูกค้า (Clustering) โดยพิจารณาจากพฤติกรรมการซื้อของลูกค้า ด้วยเทคนิคการเรียนรู้ของเครื่องแบบไม่มีผู้สอนโดยใช้เทคนิค K-Means และแบบจำลอง RFM (การซื้อครั้งล่าสุด, ความถี่ในการซื้อ, ยอดค่าใช้จ่ายรวม) และเพิ่มตัวแปร Basket Size (ยอดการซื้อแต่ละครั้ง) และตัวแปร Day Type (วันที่มาใช้บริการ) ผู้วิจัยเลือกศึกษาธุรกิจประเภทร้านค้าปลีก ใช้ชุดข้อมูลสาธารณะจากเว็บไซต์ www.kaggle.com จัดเก็บข้อมูลในปี ค.ศ. 2019 ถึงต้นปี ค.ศ. 2023 มีจำนวนข้อมูลทั้งหมด 29,103 รายการ มี 7 คุณลักษณะ ใช้เทคนิค K-Means ในการแบ่งกลุ่มลูกค้า กำหนดค่า K ที่เหมาะสมด้วย Elbow Method และวัดประสิทธิภาพการแบ่งกลุ่มด้วย Silhouette Score และ Davies-Bouldin Index ผลการวิเคราะห์แบ่งกลุ่มลูกค้าสามารถแบ่งกลุ่มได้ 4 กลุ่ม และผลจากการศึกษาลักษณะเฉพาะของกลุ่มนิยามได้ดังนี้ กลุ่มที่ 1 มีจำนวน 259 ราย ให้คำนิยามเป็น “กลุ่มลูกค้ามาน้อย จ่ายน้อย ช้อปวันธรรมดา” กลุ่มที่ 2 มีจำนวน 140 ราย ให้คำนิยามเป็น “กลุ่มลูกค้ามาบ่อย จ่ายหนัก ช้อปวันธรรมดา” กลุ่มที่ 3 มีจำนวน 66 ราย ให้คำนิยามเป็น “กลุ่มลูกค้ามาบ่อย จ่ายหนัก ช้อปวันหยุด” กลุ่มที่ 4 มีจำนวน 42 ราย ให้คำนิยามเป็น “กลุ่มลูกค้ามาน้อย จ่ายน้อย ช้อปวันหยุด” การวิเคราะห์คุณลักษณะนี้สามารถช่วยให้ธุรกิจเข้าใจและจัดการ กับกลุ่มลูกค้าได้อย่างมีประสิทธิภาพและเหมาะสม เช่น การพัฒนาและการนำเสนอโปรโมชั่นหรือแนวทางการตลาดที่เหมาะสมสำหรับกลุ่มลูกค้าที่มีลักษณะและพฤติกรรมทางการซื้อเหล่านี้ และสามารถใช้ข้อมูลเพิ่มเติมเพื่อวิเคราะห์พฤติกรรมการซื้อของลูกค้าได้ละเอียดมากขึ้น

คำสำคัญ : การแบ่งกลุ่มลูกค้า, พฤติกรรมการซื้อ, การจัดกลุ่มแบบเคมีน, แบบจำลองอาร์เอฟเอ็ม

---

[1] หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิทยาการข้อมูล คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ กรุงเทพฯ 10110

[2] คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ กรุงเทพฯ 10110

* Corresponding author:Tel.: 084-7966797 E-mail address: kamontip.montrisa@g.swu.ac.th

A Study of Customer Segmentation Based on Purchasing Behaviour Using the K-Means Technique

Kamontip Montrisa[1*], Ruangsak Trakunphutthirak[2]

Abstract

In contemporary business and marketing, understanding the diversity and differences in customer behaviors is crucial. Analyzing and comprehending customers' purchasing patterns helps in grasping customer groups and utilizing this data for business decision-making and effective marketing strategies. This research aims to study customer segmentation using clustering techniques, considering customers' purchasing behaviors, employing K-Means and RFM models (Recency, Frequency, Monetary), along with additional variables like Basket Size and Day Type. The study focuses on retail businesses, utilizing a dataset from www.kaggle.com spanning from 2019 to early 2023, comprising 29,103 entries with 7 features. K-Means clustering is employed, determining the appropriate K value using the Elbow Method and evaluating the clustering performance using the Silhouette Score and Davies-Bouldin Index. The analysis results in 4 customer groups defined as follows: Group 1: Includes 259 customers. Defined as "Low-Volume, Low-Spending, Regular Shopping on Weekdays." Group 2: Includes 140 customers. Defined as "Frequent Shoppers, Heavy Spenders, Shopping on Weekdays." Group 3: Includes 66 customers. Defined as "Frequent Shoppers, Heavy Spenders, Shopping on Weekends." Group 4: Includes 42 customers. Defined as "Low-Volume, Low-Spending, Shopping on Weekends." Analyzing these characteristics can aid businesses in effectively managing and catering to customer groups. For instance, developing and presenting promotions or marketing strategies tailored to the specific characteristics and purchasing behaviors of these customer groups. Further data analysis can provide more detailed insights into customer purchasing behaviors, enabling businesses to better understand and serve their customers.

Keywords : Customer Segmentation, Purchasing Behavior, K-Means Clustering, RFM Model

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

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

* Corresponding author:Tel.: 084-7966797 E-mail address: kamontip.montrisa@g.swu.ac.th

บทนำ

การทำธุรกิจและการตลาดในยุคปัจจุบันนี้ การคำนึงถึงความหลากหลายและความแตกต่างในพฤติกรรมของลูกค้าเป็นสิ่งที่สำคัญ การรับฟังและตอบสนองต่อความต้องการของลูกค้านั้นเป็นปัจจัยสำคัญอย่างมากในการแข่งขันในทางธุรกิจ หากเจ้าของธุรกิจต่าง ๆ ไม่สามารถทำความเข้าใจถึงพฤติกรรมของลูกค้าตนเองแล้ว อัตราการสูญเสียลูกค้าไปให้แก่คู่แข่งทางธุรกิจ ก็จะยิ่งเพิ่มสูงขึ้น ในสถานการณ์ปัจจุบันที่ลูกค้ามีบทบาทสำคัญในการกำหนดทิศทางการใช้สินค้าและบริการ ลูกค้าจะเลือกซื้อสินค้าหรือบริการที่ทำให้ตนพึงพอใจ ถ้าหากว่าตัวสินค้าหรือบริการที่ใช้อยู่ไม่สามารถตอบสนองความต้องการได้ ก็อาจเปลี่ยนไปใช้สินค้าหรือบริการจากผู้ให้บริการรายอื่น การแบ่งกลุ่มลูกค้าตามพฤติกรรมการซื้อสินค้า ถือเป็นกระบวนการสำคัญในการวิเคราะห์กลุ่มลูกค้าที่มีพฤติกรรมการซื้อสินค้าที่คล้ายคลึงกัน ข้อมูลที่เกี่ยวข้องหรือเป็นตัวแทนของลักษณะการซื้อสินค้าของลูกค้ามีความสำคัญในกระบวนการนี้ ทั้งนี้การเข้าใจลักษณะและพฤติกรรมการซื้อสินค้าของลูกค้ายังช่วยสร้างความเข้าใจในกลุ่มลูกค้า และนำข้อมูลนี้มาใช้ในการตัดสินใจทางธุรกิจและกำหนดกลยุทธ์การตลาดที่มีประสิทธิภาพ ตอบโจทย์ต่อความต้องการของลูกค้าอย่างเหมาะสม ช่วยในการวางแผนกลยุทธ์ที่สร้างความพึงพอใจและความภักดีให้กับลูกค้า ทั้งยังสร้างความเชื่อมั่นให้กับลูกค้าอย่างยั่งยืน นอกจากนี้การแบ่งกลุ่มลูกค้ายังช่วยในการสื่อสารกับกลุ่มลูกค้าผ่านช่องทางหรือแพลตฟอร์มที่เหมาะสมและช่วยในการค้นพบโอกาสใหม่ ๆ

จากที่กล่าวมาข้างต้น ผู้วิจัยจึงสนใจศึกษาวิธีการแบ่งกลุ่มลูกค้า เพื่อทำความเข้าใจถึงพฤติกรรมของลูกค้า ได้เลือกทำการศึกษาวิธีการแบ่งกลุ่มลูกค้าตามพฤติกรรมการซื้อสินค้า โดยใช้ชุดข้อมูลสาธารณะจากเว็บไซต์ www.kaggle.com เกี่ยวกับรายการขายปลีก จัดเก็บข้อมูลในปี 2019 ถึงต้นปี 2023 มีจำนวนข้อมูลทั้งหมด 29,103 รายการ มี 7 คุณลักษณะ เพื่อที่จะสามารถนำวิธีการนี้ไปประยุกต์ใช้ประโยชน์ทางด้านธุรกิจและวางแผนทำการตลาดได้ในอนาคต

งานวิจัยที่เกี่ยวข้อง

บทความวิจัยที่เกี่ยวข้อง มีตัวอย่างดังต่อไปนี้ ณรรฐคุณ วิรุฬห์ศรี และคณะ [1] ได้ทำการศึกษาวิเคราะห์แบ่งกลุ่มลูกค้าของบริษัทผลิตอาหารสัตว์เลี้ยงโดยใช้ข้อมูลการซื้อขายอาหารสัตว์เลี้ยงและศึกษาพฤติกรรมและลักษณะของลูกค้าในแต่ละกลุ่ม โดยใช้เทคนิคการวิเคราะห์แบ่งกลุ่มแบบ K-means clustering เก็บรวบรวมฐานข้อมูลการซื้อขายผลิตภัณฑ์อาหารสัตว์ของบริษัทกรณีศึกษาระหว่าง ปี ค.ศ. 2018-2020 และทำการรวมตารางข้อมูลในแต่ละปี ซึ่งมีลูกค้าทั้งหมด 1,431 ราย ผลการวิเคราะห์แบ่งกลุ่มพบว่า จากวิธี Elbow สามารถกำหนดจำนวนกลุ่มได้ 8 กลุ่ม และผลจากการศึกษาลักษณะเฉพาะของกลุ่มทำให้นิยามกลุ่มได้ดังนี้ “กลุ่มลูกค้าทั่วไป คืนสินค้าน้อย” (56.4%) “กลุ่มลูกค้าทั่วไปใจไม่นิ่ง” (15.7%) “กลุ่มลูกค้าห่างไกล” (13.2%) “กลุ่มลูกค้าขยันคืน” (6.0%) “กลุ่มลูกค้าซื้อง่ายขายคล่อง” (5.9%) “กลุ่มลูกค้ากระเป๋าหนัก” (1.3%) “กลุ่มลูกค้าที่มีแนวโน้มควรรักษาไว้” (0.8%) และ “กลุ่มลูกค้าขาจร” (0.79%) ผลที่ได้สามารถนำไปใช้ในการวางกลยุทธ์การขายให้เหมาะสมกับพฤติกรรมของลูกค้าในแต่ละกลุ่มเพื่อพัฒนาสินค้าและบริการให้ตอบสนองความต้องการของลูกค้าได้ดียิ่งขึ้น บทความวิจัยถัดไป เอกปรียา ใบสนิ [2] ได้ทำการศึกษาวิธีการแบ่งกลุ่มลูกค้าโดยใช้เทคนิคการเรียนรู้ของเครื่อง เพื่อจัดกลุ่มของข้อมูลร่วมกับการวิเคราะห์ข้อความ ด้วย 2 วิธี ดังนี้ 1) การแบ่งกลุ่มลูกค้าด้วยเทคนิคอาร์เอฟเอ็ม 2) การแบ่งกลุ่มด้วยเทคนิคอาร์เอฟเอ็มและการวิเคราะห์ข้อความด้วยเครื่องมือประมวลภาษาธรรมชาติโมดูลเอ็นแอลทีเค (Natural Language Toolkit) เปรียบเทียบผลของการจัดกลุ่มโดยใช้ฟีเจอร์ที่ได้จากเทคนิคอาร์เอฟเอ็มและเทคนิควิเคราะห์ข้อความมารวมกัน เปรียบเทียบกับการจัดกลุ่มโดยใช้

ฟีเจอร์ที่ได้จากเทคนิคอาร์เอฟเอ็ม ในการเปรียบเทียบค่า Adjusted Rand Index และค่า Normalized Mutual Information ระหว่างวิธีที่ 1 และวิธีที่ 2 ซึ่งให้ค่า ARI เท่ากับ 0.5116 และ NMI เท่ากับ 0.3646 ค่า Confusion Matrix ที่ได้จากการเปรียบเทียบทั้ง 2 วิธี ผลที่ได้จากการแบ่งกลุ่มไม่สอดคล้องกันมากนัก จากการเปรียบเทียบการแบ่งกลุ่มด้วยเทคนิคอาร์เอฟเอ็มร่วมกับการวิเคราะห์ข้อความ มีประสิทธิภาพดีกว่าวิธีที่ใช้ฟีเจอร์ที่ได้จากเทคนิคอาร์เอฟเอ็มเพียงเทคนิคเดียวในแง่ที่ผลการจัดกลุ่มที่ได้จะให้ข้อมูลเชิงลึกของการสั่งซื้อร่วมกับข้อมูลในส่วนของศักยภาพของลูกค้าจากข้อมูลอาร์เอฟเอ็ม บทความวิจัยถัดไป Kansal และคณะ [3] ได้ศึกษาอัลกอริทึมที่แตกต่างกัน 3 แบบ คือ K-Means, Agglomerative และ Mean shift ถูกนำมาใช้เพื่อแบ่งกลุ่มลูกค้าและเปรียบเทียบผลลัพธ์ของกลุ่มที่ได้รับจากอัลกอริทึมของชุดข้อมูลที่มีคุณสมบัติ 2 ประการ จำนวน 200 ตัวอย่าง คือ ค่าเฉลี่ยของปริมาณการซื้อและค่าเฉลี่ยการเข้าเยี่ยมชมของลูกค้าในร้านค้าเป็นประจำทุกปี ข้อมูลได้รับการปรับขนาดโดยใช้ Standard Scaler จำนวนกลุ่มของ K-means และ Agglomerative มีค่าเท่ากัน โดยให้รูปแบบเดียวกัน โดยประยุกต์การจัดกลุ่มลูกค้า 5 กลุ่ม แบ่งออกเป็น Careless, Careful, Standard, Target และ Sensible customers อย่างไรก็ตาม มีกลุ่ม 2 กลุ่มจากการใช้การแบ่งกลุ่มด้วย mean shift clustering ซึ่งถูกตั้งชื่อว่า High buyers and frequent visitors และ High buyers and occasional visitors ผลลัพธ์ของกลุ่มที่ได้รับจากทั้ง 3 อัลกอริทึมไม่มีความแตกต่างที่มีนัยสำคัญมากนักใน K-means และ Agglomerative clustering ดังนั้นอัลกอริทึมทั้งสองนี้สามารถจัดกลุ่มของข้อมูลได้ดีกว่า อัลกอริทึม Mean shift วัดผลการประเมินด้วย silhouette score บทความวิจัยถัดไป Dedi และคณะ [4] ได้ศึกษาการแบ่งกลุ่มลูกค้าเพื่อให้เจ้าของธุรกิจสามารถกำหนดว่าลูกค้ากลุ่มไหนที่ให้ประโยชน์มากที่สุดและลูกค้ากลุ่มไหนที่ไม่ให้ประโยชน์ การระบุเกณฑ์ของลูกค้าในการแบ่งกลุ่มตามค่า RFM (Recency, Frequency, Monetary) วิธีการจัดกลุ่มนี้ใช้อัลกอริทึม K-Means Clustering ผลลัพธ์จาก Elbow Method แบ่งกลุ่มได้ทั้งหมด 3 กลุ่ม พร้อมลักษณะของแต่ละกลุ่ม กลุ่มที่ 1 ประกอบด้วยลูกค้า 69 คน จัดให้เป็น “ลูกค้าที่มาช้อปทุกวัน” กลุ่มที่ 2 ประกอบด้วยลูกค้า 95 คน จัดให้เป็น “ลูกค้าที่ไม่มีความเคลื่อนไหว” กลุ่มที่ 3 มีลูกค้า 11 คน จัดให้เป็น “ลูกค้าชั้นดี” ค่าของความแตกต่างของ SSE คือ 2.7630 และค่าเฉลี่ยของ Silhouette Index คือ 0.7210 สามารถให้ผลการแบ่งกลุ่มลูกค้าตามค่า RFM ที่แม่นยำในการแบ่งกลุ่มลูกค้าได้ บริษัทสามารถใช้เป็นสื่อส่งเสริมการขาย จัดโปรโมชั่นให้กับลูกค้าชั้นดีได้

วิธีดำเนินการ

ภาพประกอบ 1 แสดงขั้นตอนการวิจัย

1. การทำความเข้าใจปัญหา (Business Understanding)

ในงานวิจัยนี้ผู้วิจัยได้ศึกษาเกี่ยวกับธุรกิจประเภทร้านค้าปลีก ซึ่งใช้เป็นชุดข้อมูลสาธารณะจากเว็บไซต์ www.kaggle.com โดยมีความต้องการแบ่งกลุ่มลูกค้า (Customer Segmentation) ตามพฤติกรรมการซื้อสินค้าอย่างเหมาะสม เพื่อให้สามารถจัดกลุ่มลูกค้าที่คล้ายคลึงกัน นำไปสู่การพัฒนากลยุทธ์และวางแผนการตลาด หรือการเพิ่มประสิทธิภาพในการบริหารจัดการร้านค้าปลีกให้ตอบโจทย์ตรงกับความต้องการของลูกค้าในทุก ๆ กลุ่ม

2. การทำความเข้าใจข้อมูล (Data Understanding)

รวบรวมข้อมูลที่เกี่ยวข้องและเชื่อถือได้ ตรวจสอบคุณภาพของข้อมูลเพื่อให้เข้าใจลักษณะของข้อมูลและสามารถจัดการกับข้อมูลให้เหมาะสมพอที่จะนำมาใช้วิเคราะห์ ในขั้นตอนนี้ผู้วิจัยได้รวบรวมข้อมูลการขายของร้านค้าปลีก จัดเก็บข้อมูลในปี ค.ศ. 2019 ถึงต้นปี 2023 มีจำนวนข้อมูลทั้งหมด 29,103 รายการ มี 7 คุณลักษณะ แสดงถึงข้อมูลของแต่ละบุคคลที่ไม่ซ้ำกัน ดังตารางที่ 1

ตารางที่ 1 ชื่อคอลัมน์ (Column Name) และคำอธิบาย (Description)

Column Name	Description
InvoiceID	รหัสใบเสร็จ
Date	วันที่ทำรายการ
ProductID	รหัสสินค้า (รหัสเฉพาะสำหรับสินค้าแต่ละรายการ)
TotalSales	ยอดขายรวมต่อใบเสร็จ
Discount	จำนวนส่วนลด
CustomerID	รหัสลูกค้า ที่ไม่ซ้ำกันสำหรับลูกค้าแต่ละราย
Quantity	จำนวนสินค้าที่ขายต่อใบเสร็จ

3. การเตรียมข้อมูล (Data Preparation)

1. การทำความสะอาดข้อมูล (Data Cleaning)

การทำความสะอาดข้อมูลนั้นตรวจสอบว่ามีข้อผิดพลาดอะไรเกิดขึ้นกับข้อมูลบ้าง เช่น มีการจัดเก็บข้อมูลที่ซ้ำซ้อน (Duplicate data) ข้อมูลไม่ถูกต้อง (Incorrectly data) ข้อมูลเก่าล้าสมัย (Expired data) เกิดการสูญหายของข้อมูลในบางส่วน (Missing Value) มีค่าข้อมูลที่มีความผิดปกติหรือแตกต่างไปจากข้อมูลในกลุ่ม (Outliers) เป็นต้น จากนั้นทำการปรับปรุงข้อมูล โดยวิธีการอาจใช้การแทนที่ด้วยการแก้ไขเป็นค่าที่ถูกต้องหรือแทนด้วยค่าใหม่หรืออาจทำการลบข้อมูลที่ไม่ถูกต้องเหล่านั้นออกไป เป็นต้น

2. การสร้างตัวแปรใหม่จากข้อมูลที่มีอยู่ (Feature Engineering)

การสร้างตัวแปรใหม่จากข้อมูลการซื้อสินค้า (Transaction) ที่เก็บมา โดยปกติแล้วข้อมูลที่มีไม่สามารถนำมาใช้ได้โดยตรง เนื่องจากการซื้อสินค้าหนึ่งครั้งมากกว่า 1 แถว และลูกค้า 1 คน มีการซื้อมากกว่า 1 ครั้ง ทำให้มีหมายเลขการสั่งซื้อที่แตกต่างกันไป และวัตถุประสงค์ของการวิเคราะห์ข้อมูลในกรณีนี้ต้องการแบ่งกลุ่มลูกค้าแต่ในข้อมูลที่มีจะเป็นรายละเอียดของการซื้อสินค้าแต่ละครั้ง ดังนั้น จึงต้องการสร้างเป็น profile ของลูกค้าแต่ละรายเสียก่อน โดยการสร้าง Profile นี้จะพิจารณาตัวแปรต่างๆ ดังนี้

1) Recency (R) คือ จำนวนวันที่ซื้อครั้งล่าสุดห่างจากวันคำนวณกี่วัน
2) Frequency (F) คือ จำนวนครั้งในการซื้อสินค้า
3) Monetary (M) คือ ยอดค่าใช้จ่ายรวมจากทุกตาราง
4) Basket Size (B) คือ ยอดค่าใช้จ่ายแต่ละครั้ง
5) Day_Type (D) คือ วันที่ลูกค้ามาใช้บริการ

	CustomerID	Recency	Frequency	Monetary	Basket Size	Day_Type
0	0.0	7.0	50.0	303317.932	6066.359	0
1	1.0	735.0	13.0	20476.729	1575.133	0
2	2.0	52.0	36.0	51963.712	1443.436	0
3	3.0	625.0	1.0	300.847	300.847	0
4	4.0	92.0	38.0	60977.720	1604.677	0
...	...	...	...	...	...	...
502	502.0	7.0	130.0	275282.339	2117.556	0
503	503.0	229.0	3.0	19930.508	6643.503	0
504	504.0	131.0	5.0	16088.983	3217.797	0
505	505.0	12.0	67.0	286124.865	4270.520	0
506	506.0	814.0	298.0	356847.178	1197.474	0

507 rows × 6 columns

ภาพประกอบ 2 แสดงถึงการสร้างคอลัมน์ RFM+B+D

4. การสร้างแบบจำลอง (Modeling)

ในการศึกษานี้ ผู้วิจัยทำการทดลองสร้างแบบจำลอง 2 แบบ โดยแบบจำลองที่ 1 จะเป็นการทดลองแบ่งกลุ่มด้วยข้อมูล RFM แบบจำลองที่ 2 จะเป็นการทดลองแบ่งกลุ่มด้วยข้อมูล RFMBD เพิ่มตัวแปรจากยอดใช้จ่ายต่อครั้ง ร่วมกับการจัดกลุ่มตามวันที่มาใช้บริการ โดยข้อมูลที่นำมาใช้ในการทดลองแบบจำลองแต่ละแบบ มีรายละเอียดดังนี้

แบบจำลองที่ 1 การจัดกลุ่มด้วยการใช้ข้อมูล RFM

	Recency	Frequency	Monetary
0	7	50	303317.93
1	735	13	20476.73
2	52	36	51963.71
3	625	1	300.85
4	92	38	60977.72

แบบจำลองที่ 2 การจัดกลุ่มด้วยการใช้ข้อมูล RFMBD

	Recency	Frequency	Monetary	Basket Size	Day_Type
0	7.00	50.00	303317.93	6066.36	0
1	735.00	13.00	20476.73	1575.13	0
2	52.00	36.00	51963.71	1443.44	0
3	625.00	1.00	300.85	300.85	0
4	92.00	38.00	60977.72	1604.68	0

5. การวัดประสิทธิภาพ (Evaluation)

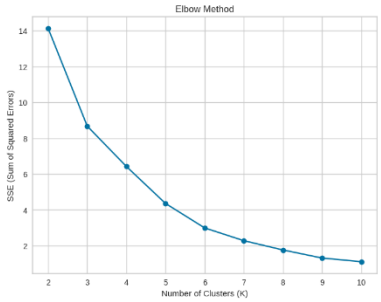
ในงานวิจัยนี้ ผู้วิจัยได้ใช้วิธีวิเคราะห์และวัดประสิทธิภาพของผลลัพธ์กลุ่มข้อมูลจากทั้ง 2 แบบจำลองด้วยค่า Silhouette Score และ Davies-Bouldin Index

ผลการวิจัยและอภิปรายผลการวิจัย

1. ผลการแบ่งกลุ่ม ทำ K-Means Clustering ด้วย RFM

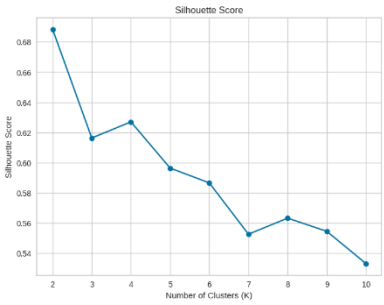
จากการสร้างแบบจำลองแบ่งกลุ่มด้วย RFM โดยใช้ข้อมูลที่ได้จากขั้นตอนการวิเคราะห์ข้อมูลลูกค้า การทำ Clustering การแบ่งกลุ่มข้อมูลที่เลือกค่า K ที่ดีที่สุดในการหาจำนวนกลุ่มที่เหมาะสม โดยใช้ Elbow Method และพิจารณาจากค่า Silhouette Score และค่า Davies-Bouldin Index ผลลัพธ์จากการจัดกลุ่มแสดงได้ดังต่อไปนี้

เมื่อพิจารณาจากการหักศอก ได้จำนวนกลุ่มเท่ากับ 4 กลุ่ม ดังภาพประกอบ 3



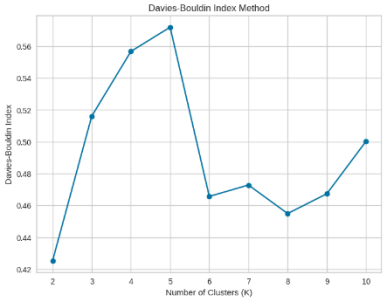
ภาพประกอบ 3 แสดงถึงผลการหาจำนวนกลุ่ม RFM ที่เหมาะสมด้วยวิธี Elbow

เมื่อพิจารณาจากค่าเฉลี่ยของ Silhouette Score อยู่ที่ 0.6884 ซึ่งมีค่าเข้าใกล้ 1 มากที่สุด ได้จำนวนกลุ่มเท่ากับ 2 กลุ่ม ดังภาพประกอบ 4



ภาพประกอบ 4 แสดงถึงผลการหาจำนวนกลุ่ม RFM ที่เหมาะสมด้วยวิธี Silhouette Score

เมื่อพิจารณาจากค่าเฉลี่ยของ Davies-Bouldin Index อยู่ที่ 0.4253 ซึ่งมีค่าน้อยที่สุด ได้จำนวนกลุ่มเท่ากับ 2 กลุ่ม ดังภาพประกอบ 5

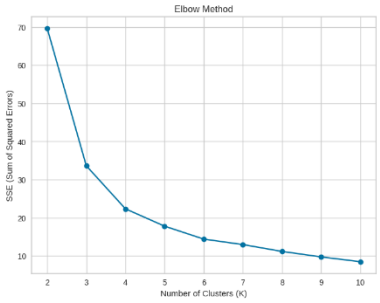


ภาพประกอบ 5 แสดงถึงผลการหาจำนวนกลุ่ม RFM ที่เหมาะสมด้วยวิธี Davies-Bouldin Index

2. ผลการแบ่งกลุ่ม ทำ K-Means Clustering RFMBD

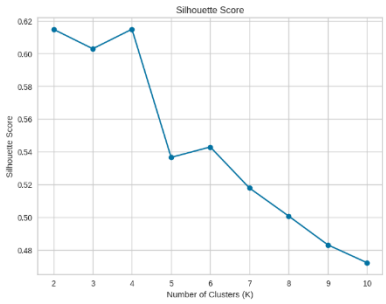
จากการสร้างแบบจำลองแบ่งกลุ่มด้วย RFM และเพิ่มตัวแปร B กับ D โดยใช้ข้อมูลที่ได้จากขั้นตอนการวิเคราะห์ข้อมูลลูกค้า การทำ Clustering การแบ่งกลุ่มข้อมูลที่เลือกค่า K ที่ดีที่สุดในการหาจำนวนกลุ่มที่เหมาะสม โดยใช้ Elbow Method และพิจารณาจากค่า Silhouette Score และค่า Davies-Bouldin Index ผลลัพธ์จากการจัดกลุ่มแสดงได้ดังต่อไปนี้

เมื่อพิจารณาจากการหักศอกที่ได้ ได้จำนวนกลุ่มเท่ากับ 4 กลุ่ม ดังภาพประกอบ 6



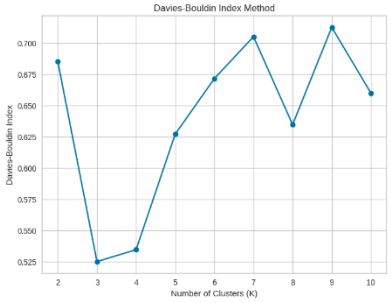
ภาพประกอบ 6 แสดงถึงผลการหาจำนวนกลุ่ม RFMBD ที่เหมาะสมด้วยวิธี Elbow

เมื่อพิจารณาจากค่าเฉลี่ยของ Silhouette Score อยู่ที่ 0.6148 ซึ่งมีค่าเข้าใกล้ 1 มากที่สุด ได้จำนวนกลุ่มเท่ากับ 4 กลุ่ม ดังภาพประกอบ 7



ภาพประกอบ 7 แสดงถึงผลการหาจำนวนกลุ่ม RFMBD ที่เหมาะสมด้วยวิธี Silhouette Score

เมื่อพิจารณาจากค่าเฉลี่ยของ Davies-Bouldin Index อยู่ที่ 0.5249 ซึ่งมีค่าน้อยที่สุด ได้จำนวนกลุ่มเท่ากับ 3 กลุ่ม ดังภาพประกอบ 8



ภาพประกอบ 8 แสดงถึงผลการหาจำนวนกลุ่ม RFMBD ที่เหมาะสมด้วยวิธี Davies-Bouldin Index

3. ผลการประเมินของการแบ่งกลุ่มด้วย RFM และการจัดกลุ่มด้วย RFMBD

ตารางที่ 2 ผลการประเมินของการแบ่งกลุ่มของข้อมูล RFM

Cluster RFM	Silhouette Score	Davies-Bouldin Index
2	0.6884	0.4253
3	0.6163	0.5161
4	0.6270	0.5567
5	0.5964	0.5712
6	0.5866	0.4657
7	0.5526	0.4728
8	0.5632	0.4549
9	0.5545	0.4673
10	0.5331	0.5002

จากตารางที่ 2 จะเห็นได้ว่า Cluster ที่มี Silhouette Score สูงและ Davies-Bouldin Index ต่ำสุด จะถือว่าเป็นการแบ่งกลุ่มที่ดีที่สุด โดย Cluster 2 มีผลการประเมินที่ดีที่สุดในทั้งสองดัชนีนี้ ควรพิจารณาการเลือกจำนวน Cluster ที่เหมาะสมตามวัตถุประสงค์ของการวิเคราะห์ข้อมูลและการนำไปใช้งาน

จากผลลัพธ์ของ Elbow Method และ Silhouette Score และ Davies-Bouldin Index ผู้วิจัยเลือกจัดกลุ่มข้อมูลเป็น 2 กลุ่ม โดยได้ผลลัพธ์ดังนี้ กลุ่มที่ 1 มีจำนวน 333 ราย กลุ่มที่ 2 มีจำนวน 174 ราย

ตารางที่ 4 ค่าเฉลี่ยของ Cluster RFM

Cluster RFM	Recency	Frequency	Monetary
0	157.5706	82.7237	219,442.8836
1	1,130.4195	7.1667	8,698.3188

ตารางที่ 3 ผลการประเมินของการแบ่งกลุ่มของข้อมูล RFMBD

Cluster RFMBD	Silhouette Score	Davies-Bouldin Index
2	0.6147	0.6850
3	0.6029	0.5249
4	0.6148	0.5344
5	0.5366	0.6270
6	0.5430	0.6715
7	0.5179	0.7051
8	0.5008	0.6346
9	0.4832	0.7126
10	0.4723	0.6599

จากตารางที่ 3 จะเห็นได้ว่า Cluster ที่มี Silhouette Score สูงสุด คือ Cluster 4 และ Davies-Bouldin Index ต่ำสุด คือ Cluster 3 การเลือกค่าที่เหมาะสมจึงต้องพิจารณาความเหมาะสมตามวัตถุประสงค์และลักษณะของข้อมูล

Silhouette Score จะวัดว่าข้อมูลมีความคล้ายคลึงกับคลัสเตอร์ของตัวเอง เมื่อเปรียบเทียบกับคลัสเตอร์อื่นๆ

Davies-Bouldin Index จะประเมินความคล้ายคลึงโดยเฉลี่ยระหว่างแต่ละคลัสเตอร์และคลัสเตอร์ที่คล้ายกันมากที่สุด

จากผลลัพธ์ของ Elbow Method และ Silhouette Score และ Davies-Bouldin Index ผู้วิจัยเลือกจัดกลุ่มข้อมูลเป็น 4 กลุ่ม ได้ผลลัพธ์ดังนี้ กลุ่มที่ 1 มีจำนวน 259 ราย กลุ่มที่ 2 มีจำนวน 140 ราย กลุ่มที่ 3 มีจำนวน 66 ราย กลุ่มที่ 4 มีจำนวน 42 ราย

ตารางที่ 5 ค่าเฉลี่ยของ Cluster RFMBD

Cluster RFMBD	Recency	Frequency	Monetary	Basket Size	Day_Type
0	1,110.2286	6.7714	8,560.0576	1,450.2134	0
1	155.6100	67.0270	191,683.1485	3,261.9892	0
2	109.3636	154.1061	354,585.7849	3,480.6156	1
3	1,100.2619	7.5000	8,118.6285	1,160.1186	1

สรุปผลการวิจัย

การศึกษาวิธีการแบ่งกลุ่มลูกค้าตามพฤติกรรมการซื้อสินค้าของร้านค้าปลีกแห่งหนึ่ง ด้วยแบบจำลอง RFM และมีการเพิ่มตัวแปรยอดการซื้อสินค้าแต่ละครั้ง (Basket Size) วันที่มาใช้บริการ (Day Type) โดยใช้เทคนิค K-Means ในการทำ Clustering การทดลองแบ่งเป็น 2 แบบจำลอง โดยแบบจำลองที่ 1 คือ การใช้ข้อมูล RFM และแบบจำลองที่ 2 คือ การใช้ข้อมูล RFMBD โดยพิจารณาจำนวนกลุ่มที่เหมาะสมด้วยวิธี Elbow ค่าเฉลี่ยของ Silhouette Score และค่าเฉลี่ยของ Davies-Bouldin index พบว่า แบบจำลองที่ 1 การใช้ข้อมูล RFM จำนวนกลุ่มที่เหมาะสมของลูกค้าจากผลลัพธ์ของ Elbow Method เมื่อพิจารณาจากการหักศอก ได้จำนวนกลุ่มเท่ากับ 4 กลุ่ม เมื่อพิจารณาจากค่าเฉลี่ยของ Silhouette Score อยู่ที่ 0.6884 ซึ่งมีค่าเข้าใกล้ 1 มากที่สุด ได้จำนวนกลุ่มเท่ากับ 2 กลุ่ม และพิจารณาจากค่าเฉลี่ยของ Davies-Bouldin Index อยู่ที่ 0.4253 ซึ่งมีค่าน้อยที่สุด ได้จำนวนกลุ่มเท่ากับ 2 กลุ่ม ดังนั้น ผู้วิจัยจึงเลือกแบ่งกลุ่มข้อมูลลูกค้า RFM ออกเป็น 2 กลุ่ม โดยได้ผลลัพธ์ดังนี้ กลุ่มที่ 1 มีจำนวน 333 ราย ให้คำนิยามเป็น “กลุ่มลูกค้าซื้อเยอะ จ่ายหนัก” เนื่องจากว่าลูกค้าในกลุ่มนี้มักจะมีความต้องการและความสนใจในการซื้อสินค้าหรือบริการในปริมาณมาก เข้ามาใช้บริการอย่างสม่ำเสมอและมักจะมีการใช้จ่ายสูงในแต่ละครั้งที่เข้ามาใช้บริการ ดังนั้นธุรกิจอาจมองเห็นกลุ่มลูกค้านี้เป็นกลุ่มเป้าหมายที่สำคัญ ในการพัฒนากลยุทธ์การตลาดและการบริการในอนาคตได้ กลุ่มที่ 2 มีจำนวน 174 ราย ให้คำนิยามเป็น “กลุ่มลูกค้าซื้อน้อย คิดนาน” เนื่องจากว่าลูกค้าในกลุ่มนี้มักจะมีลักษณะที่ใช้เวลาในการตัดสินใจในการซื้อสินค้าหรือบริการนานและมักเน้นการซื้อสินค้าในปริมาณน้อยในแต่ละครั้งที่เข้ามาใช้บริการ อาจเป็นเพราะไม่มีความต้องการในการซื้อสินค้าหรือบริการบ่อยครั้งหรือมีความต้องการน้อยลง ซึ่งอาจต้องการความเชื่อมั่นหรือข้อมูลเพิ่มเติมก่อนที่จะตัดสินใจในการซื้อสินค้าหรือบริการในอนาคต ในส่วนของแบบจำลองที่ 2 การใช้ข้อมูล RFMBD จำนวนกลุ่มที่เหมาะสมของลูกค้าจากผลลัพธ์ของ Elbow Method เมื่อพิจารณาจากการหักศอก ได้จำนวนกลุ่มเท่ากับ 4 กลุ่ม เมื่อพิจารณาจากค่าเฉลี่ยของ Silhouette Score อยู่ที่ 0.6148 ซึ่งมีค่าเข้าใกล้ 1 มากที่สุด ได้จำนวนกลุ่มเท่ากับ 4 กลุ่ม และพิจารณาจากค่าเฉลี่ยของ Davies-Bouldin Index อยู่ที่ 0.5249 ซึ่งมีค่าน้อยที่สุด ได้จำนวนกลุ่มเท่ากับ 3 กลุ่ม จะเห็นว่ามีการแบ่งกลุ่มได้แตกต่างกัน ค่า Elbow Method กับ Silhouette Score แบ่งออกเป็น 4 กลุ่ม ในขณะที่ Davies-Bouldin Index แบ่งออกเป็น 3 กลุ่ม ดังนั้น การเลือกจำนวนกลุ่มที่เหมาะสม จะขึ้นอยู่กับลักษณะของข้อมูลและวัตถุประสงค์ของการวิเคราะห์ ผู้วิจัยจึงเลือกแบ่งกลุ่มข้อมูลเป็น 4 กลุ่ม ได้ผลลัพธ์ดังนี้ กลุ่มที่ 1 มีจำนวน 259 ราย ให้คำนิยามเป็น “กลุ่มลูกค้ามาน้อย จ่ายน้อย ช้อปวันธรรมดา” เนื่องจากว่าลูกค้าในกลุ่มนี้มีการซื้อสินค้าน้อย อาจจะเป็นลูกค้าที่ไม่มีความต้องการหรือไม่ได้ใช้บริการบ่อยครั้ง อาจต้องพิจารณาในการวิเคราะห์เพิ่มเติมว่าเหตุผลที่ทำให้ลูกค้าในกลุ่มนี้ไม่ทำซื้อสินค้าเป็นประจำคืออะไร ต้องมีการให้บริการหรือส่วนลดพิเศษที่เหมาะสมเพื่อเพิ่มความน่าสนใจและสร้างความพึงพอใจในการซื้อสินค้ายิ่งขึ้น เช่น โปรโมชั่นพิเศษหรือส่วนลดสำหรับการซื้อครั้งต่อไป กลุ่มที่ 2 มีจำนวน 140 ราย ให้คำนิยามเป็น “กลุ่มลูกค้ามาบ่อย จ่ายหนัก ช้อปวันธรรมดา” เนื่องจากว่าลูกค้าในกลุ่มนี้มักจะมาใช้บริการบ่อย มักมีความสนใจและความต้องการในการใช้บริการของธุรกิจ และมีจำนวนเงินที่ใช้ในแต่ละครั้งสูง มีแนวโน้มมาใช้บริการในวันธรรมดา ซึ่งอาจเป็นเพราะมีความสะดวกและเหมาะสมในการมาใช้บริการในวันทำงาน ลูกค้าในกลุ่มนี้อาจมีความต้องการที่มากขึ้นเกี่ยวกับสินค้าหรือบริการที่ธุรกิจมี และมักมีความพึงพอใจในการบริการที่ได้รับ กลุ่มที่ 3 มีจำนวน 66 ราย ให้คำนิยามเป็น “กลุ่มลูกค้ามาบ่อย จ่ายหนัก ช้อปวันหยุด” เนื่องจากว่าลูกค้าในกลุ่มนี้มักจะมาใช้บริการบ่อย มีจำนวนเงินที่ใช้ในแต่ละครั้งและยอดรวมสูง มีความสนใจและความต้องการในการใช้บริการของธุรกิจ และมีแนวโน้มมาใช้บริการในวันหยุดเพื่อรับประสบการณ์การบริการที่ดีและเพลิดเพลินไปกับสินค้าหรือบริการที่ชื่นชอบในช่วงเวลาว่างๆ กลุ่มที่ 4 มีจำนวน 42 ราย ให้คำนิยามเป็น “กลุ่มลูกค้ามาน้อย จ่ายน้อย ช้อปวันหยุด” เนื่องจากว่าลูกค้าในกลุ่มนี้ไม่ค่อยซื้อสินค้า ซึ่งอาจเป็นผลมาจากการไม่พบความน่าสนใจในสินค้าหรือบริการ หรืออาจเกิดจากปัจจัยอื่นๆ เช่น ปัญหาในการสื่อสารหรือการตลาดที่ไม่เหมาะสม

การเลือกช่วงเวลาในการใช้บริการที่ไม่เหมาะสม หรือความผิดหวังจากประสบการณ์ที่ผ่านมา ดังนั้น การสร้างกลยุทธ์ในการดึงดูดและครอบคลุมกลุ่มลูกค้านี้อาจจำเป็นต้องพัฒนาผ่านการปรับปรุงประสบการณ์การบริการหรือการสร้างความน่าสนใจในผลิตภัณฑ์หรือบริการอย่างเหมาะสม

จะเห็นได้ว่าการวิเคราะห์ข้อมูล RFMBD ที่มีการเพิ่มตัวแปรเพิ่มเติมเข้าไปเพื่อให้การวิเคราะห์และการจัดกลุ่มลูกค้ามีความครอบคลุมมากขึ้น และสามารถใช้ข้อมูลเพิ่มเติมเพื่อวิเคราะห์พฤติกรรมการซื้อของลูกค้าได้ละเอียดมากขึ้น อย่างไรก็ตาม การเลือกใช้รูปแบบของการจัดกลุ่มขึ้นอยู่กับวัตถุประสงค์และความต้องการของธุรกิจในการทำนายพฤติกรรมของลูกค้าและการจัดกลุ่มลูกค้าที่เหมาะสมกับกลยุทธ์การตลาดและการขายสินค้าของธุรกิจนั้นๆ

กิตติกรรมประกาศ

การจัดทำวิจัยได้รับการสนับสนุนจากบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ในการนำเสนอผลงานวิจัย ผู้วิจัยจึงขอขอบคุณมา ณ ที่นี้

เอกสารอ้างอิง

[1] ณรรฐคุณ วิรุฬห์ศรี และคณะ. (2565). Clustering Customers Using Their In-Depth Buying Behavior: A Pet Food Manufacturing Company Case Study. วารสารวิทยาศาสตร์ลาดกระบัง. https://li01.tci-thaijo.org/index.php/science_kmitl/article/download/250784/173839/918760

[2] เอกปรียา ใบสนิ. (2563). THE STUDY OF CUSTOMER SEGMENTATION BY USING RFM MODEL AND TEXT ANALYTICS. http://ir-ithesis.swu.ac.th/dspace/handle/123456789/1083

[3] Kansal, T., Bahuguna, S., Singh, V., และ Choudhury, T. (2018, 21-22 Dec. 2018). Customer Segmentation using K-means Clustering. 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS),

[4] Dedi, Dzulhaq, M. I., Sari, K. W., Ramdhan, S., Tullah, R., และ Sutarman. (2019, 16-17 Oct. 2019). Customer Segmentation Based on RFM Value Using K-Means Algorithm. 2019 Fourth International Conference on Informatics and Computing (ICIC),

[5] สุทธิพงศ์ ผ่องแผ้ว. (2555). STUDY OF THE SEED CLUSTERING PROCESS USING STRUCTURAL DATA. http://ir-ithesis.swu.ac.th/dspace/handle/123456789/49

[6] วิทยา พรพัชรพงศ์. (2549, 2012). การจัดกลุ่มลูกค้าและการทำเหมืองข้อมูล (Customer Segmentation and Data Mining). GotoKnow. https://www.gotoknow.org/posts/56616