การวิเคราะห์ข้อมูลจากเครื่องรับชำระอิเล็กทรอนิกส์ (EDC) เพื่อการใช้ประโยชน์สูงสุด

DATA ANALYTICS FROM THE ELECTRONIC DATA CAPTURE (EDC) MACHINE FOR MAXIMUM UTILIZATION

ศุภวิชญ์ รุจิเมธาภาส

SUPHAWIT RUJIMETHAPAS

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูลและการวิเคราะห์

ศูนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง คณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2568

KMITL-2025-SC-M-017-009

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

DATA ANALYTICS FROM THE ELECTRONIC DATA CAPTURE (EDC) MACHINE FOR MAXIMUM UTILIZATION

SUPHAWIT RUJIMETHAPAS

AN INDEPENDENT STUDY SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR
THE DEGREE OF MASTER OF SCIENCE IN DATA SCIENCE AND ANALYTICS
KMITL DIGITAL ANALYTICS AND INTELLIGENCE CENTER SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
2025

KMITL-2025-SC-M-017-009

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPYRIGHT 2025

SCHOOL OF SCIENCE

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อการค้นคว้าอิสระ	การวิเคราะห์ข้อมูลจากเครื่องรับชำระอิเล็กทรอนิกส์ เพื่อการใช้ประโยชน์สูงสุด
ชื่อนักศึกษา	นายศุภวิชญ์ รุจิเมธาภาส
รหัสประจำตัว	65056086
ปริญญา	วิทยาศาสตรมหาบัณฑิต (วิทยาการข้อมูลและการวิเคราะห์) ศูนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง
พ.ศ.	2568
อาจารย์ที่ปรึกษาการค้นคว้าอิสระ	รองศาสตราจารย์ ดร.ณัฐไชย์ ลีนาวงศ์

บทคัดย่อ

ปัจจุบันที่เทคโนโลยีมีบทบาทสำคัญในชีวิตประจำวัน การเปลี่ยนแปลงจากการใช้จ่ายด้วยธนบัตรเป็นสังคมไร้เงินสดทำให้การปรับตัวของธุรกิจเข้ากับเทคโนโลยีกลายเป็นสิ่งจำเป็น เครื่องรับชำระเงินอิเล็กทรอนิกส์ (Electronic Data Capture: EDC) เป็นหนึ่งในเครื่องมือที่ตอบโจทย์นี้ โดยสามารถรับชำระเงินผ่านบัตรเครดิตและช่วยให้ธนาคารวิเคราะห์พฤติกรรมการใช้จ่ายของลูกค้าได้อย่างมีประสิทธิภาพ งานวิจัยนี้จึงมุ่งเน้นการวิเคราะห์ข้อมูลจากเครื่องรับชำระเงินอิเล็กทรอนิกส์ เพื่อจัดกลุ่มร้านค้าตามยอดการชำระเงิน โดยใช้วิธีการแบ่งกลุ่มแบบค่าเฉลี่ยเค (K-Means Clustering) การวิเคราะห์แบบเป็นขั้นตอน (Hierarchical Clustering) และการแบ่งกลุ่มข้อมูลตามความหนาแน่น (Density-Based Spatial Clustering of Applications with Noise: DBSCAN) เพื่อระบุปัญหาที่บางร้านค้ามียอดชำระน้อยหรือไม่เป็นไปตามที่คาดหวัง การวิเคราะห์นี้มีเป้าหมายในการปรับปรุงประสิทธิภาพการใช้งานเครื่องรับชำระเงินให้เกิดประโยชน์สูงสุดแก่ธนาคารและธุรกิจ การประเมินประสิทธิภาพของแบบจำลองจัดกลุ่มทำด้วยดัชนีเดวีส์-โบลดิน (Davies-Bouldin Index) ซิลูเอต โคอิฟฟิเชียนท์ (Silhouette Coefficient) และดัชนีคาลินสกี้-ฮาราบาซ (Calinski-Harabasz Index) ผลการวิจัยสรุปได้ว่าวิธีการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเคเป็นแบบจำลองที่เหมาะสมที่สุด เนื่องจากให้ผลลัพธ์ที่ดีที่สุดโดยรวมทั้งในแง่ของการแบ่งกลุ่มที่ชัดเจน ความหนาแน่นภายในกลุ่ม และการแยกระหว่างกลุ่มที่ดี

คำสำคัญ : การจัดกลุ่มแบบค่าเฉลี่ยเค การจัดกลุ่มข้อมูลตามความหนาแน่น การจัดกลุ่มแบบเป็นขั้นตอน เครื่องรับชำระเงินอิเล็กทรอนิกส์ ซิลูเอต โคอิฟฟิเชียนท์ ดัชนีเดวีส์-โบลดิน ดัชนีคาลินสกี้-ฮาราบาซ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Independent Study Title	Data Analytics From The Electronic Data Capture (EDC) Machine for Maximum Utilization
Student Name	Suphawit Rujimethapas
Student ID	65056086
Degree	Master of Science (Data Science and Analytics) KMITL Digital Analytics and Intelligence Center
Year	2025
Independent Study Advisor	Assoc. Prof. Dr. Chartchai Leenawong

Abstract

In today's era, where technology plays a crucial role in daily life, the transition from cash-based transactions to a cashless society has made it essential for businesses to adapt to technological advancements. Electronic Data Capture (EDC) machines have become a key tool in this transformation, enabling credit card payments and allowing banks to analyze customer spending behavior effectively. This research analyzes data from EDC machines to categorize merchants based on their transaction volumes using K-Means Clustering, Hierarchical Clustering, and Density-Based Spatial Clustering of Applications with Noise (DBSCAN). The objective is to identify low or unexpected transaction volumes among certain merchants. This analysis aims to optimize the utilization of EDC machines to maximize benefits for both banks and businesses. The performance of the clustering models is evaluated using the Davies-Bouldin Index, Silhouette Coefficient, and Calinski-Harabasz Index. The results of the study indicate that the K-Means Clustering method is the most suitable model, as it provides the best overall results in terms of clear group separation, intra-group density, and inter-group distinction.

Keywords: K-Means Clustering, DBSCAN (Density-Based Clustering), Hierarchical Clustering, Electronic Data Capture, Silhouette Coefficient, Davies-Bouldin Index, Calinski-Harabasz Inde

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

งานวิจัยเรื่อง การวิเคราะห์ข้อมูลจากเครื่องรับชำระอิเล็กทรอนิกส์ (EDC) เพื่อการใช้ประโยชน์สูงสุดด้วยวิธีการเรียนรู้ของเครื่องสามารถดำเนินการจนประสบความสำเร็จลุล่วงไปด้วยดี เนื่องจากได้รับความกรุณาจาก รศ.ดร.ณัฐไชย์ ลีนาวงศ์ อาจารย์ที่ปรึกษาการค้นคว้าอิสระที่ให้คำปรึกษา ข้อเสนอแนะ เอื้อเฟื้อเอกสารต่างๆที่ใช้เป็นแนวทางในการวิเคราะห์ข้อมูล รวมถึงตรวจทานแก้ไขความถูกต้องตลอดจนติดตามงานทุกๆขั้นตอนของการดำเนินงาน จนกระทั่งงานวิจัยครั้งนี้สำเร็จเรียบร้อยด้วยดี ผู้วิจัยขอขอบพระคุณด้วยความเคารพอย่างสูงไว้ ณ โอกาสนี้

ขอขอบพระคุณ ผศ.ดร.พรรณทิพา วาณิชย์จิรัฐติกาล และดร.กิตติมศักดิ์ ในจิต คณะกรรมการการค้นคว้าอิสระที่ให้คำปรึกษา และคำแนะนำเพื่อความสมบูรณ์ยิ่งขึ้น

ขอขอบคุณคณาจารย์สาขาวิชาวิทยาการข้อมูลและการวิเคราะห์ทุกท่านที่ได้ประสิทธิ์ประสาทวิชาความรู้และช่วยเหลือให้คำแนะนำในเรื่องต่างๆมาโดยตลอด

ขอขอบพระคุณบิดามารดา ที่คอยให้กำลังใจและสนับสนุนผู้จัดทำการค้นคว้าอิสระมาโดยตลอด และขอขอบคุณเพื่อนคณะวิทยาศาสตร์สาขาวิทยาการข้อมูลและการวิเคราะห์ทุกท่านที่ให้ความรู้และช่วยเหลือในการวิจัย

ขอขอบคุณข้อมูลจากธนาคารที่อนุญาติให้นำข้อมูลออกมาใช้จนทำให้งานวิจัยนี้สำเร็จ

สุดท้ายนี้ผู้วิจัยหวังว่างานวิจัยฉบับนี้จะเป็นประโยชน์สำหรับหน่วยงานที่เกี่ยวข้องและผู้สนใจศึกษาต่อไป

นายศุภวิชญ์ รุจิเมธาภาส

พระจอมเกล้าเจ้าคุณทหาร ... เทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อ	**ก**
Abstract	**ข**
กิตติกรรมประกาศ	**ค**
สารบัญ	**ง**
สารบัญรูป	**ซ**
บทที่ 1 บทนำ	**1**
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของงานวิจัย	1
1.3 ขอบเขตของงานวิจัย	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ	2
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	**3**
2.1 เครื่องรูดบัตรแบบเคลื่อนที่ (Electronic Data Capture: EDC)	3
2.2 การจัดเตรียมข้อมูล (Data Preprocessing)	3
2.1.1 การปกปิดข้อมูล (Data Anonymization)	3
2.1.2 การวิเคราะห์อาร์เอฟเอ็ม (RFM Analysis)	3
2.3 การเรียนรู้ของเครื่อง (Machine Learning)	4
2.3.1 การแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค (K-Mean Clustering)	4
2.3.2 การวิเคราะห์แบ่งกลุ่มแบบเป็นขั้นตอน (Hierarchical Cluster Analysis)	5
2.3.3 การแบ่งกลุ่มข้อมูลตามความหนาแน่น (Density-Based Spatial Clustering of Applications with Noise: DBSCAN)	7
2.4 การประเมินประสิทธิภาพของแบบจำลอง (Model Evaluation)	8
2.4.1 ดัชนีเดวีส์-โบลดิน (Davies-Bouldin Index)	8
2.4.2 ซิลูเอต โคอิฟฟิเชียนท์ (Slhouette Coefficient)	8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

2.4.3 ดัชนีคาลินสกี้-ฮาราบาซ (Calinski-Harabasz Index) 9

2.5 งานวิจัยที่เกี่ยวข้อง 9

บทที่ 3 วิธีการดำเนินงานวิจัย 12

3.1 แผนผังขั้นตอนการทำงาน 12

3.2 การจัดเตรียมข้อมูล (Data Preprocessing) 13

3.2.1 การเก็บรวบรวมข้อมูล (Data Collection) 13

3.1.2 การวิเคราะห์อาร์เอฟเอ็ม (RFM Analysis) 14

3.3 การสร้างแบบจำลอง 15

3.4 การเปรียบเทียบประสิทธิภาพแบบจำลอง 15

บทที่ 4 ผลการวิจัยและการอภิปรายผล 16

4.1 ผลการวิเคราะห์ของการเรียนรู้ของเครื่อง 16

4.2 การเปรียบเทียบแบบจำลอง 22

4.2.1 ดัชนีเดวีส์-โบลดิน 22

4.2.2 ค่าซิลูเอต โคอิฟฟิเชียนท์ 23

4.2.3 ดัชนีคาลินสกี้-ฮาราบาซ 24

4.3 การวิเคราะห์เชิงกลยุทธ์ 25

4.3.1 กลุ่มร้านค้าที่ไม่ค่อยใช้บริการ หรือไม่ได้ใช้บริการมานาน 25

4.3.2 ร้านค้าที่ใช้บริการเป็นปกติ 26

4.3.3 ร้านค้าชั้นยอดที่ใช้บริการสูงสุด 26

บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ 27

5.1 สรุปผลการวิจัย 27

5.2 ข้อเสนอแนะ 28

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

เอกสารอ้างอิง	29
ภาคผนวก	31
ภาคผนวก ก	32
ภาคผนวก ข	37
ประวัติผู้เขียน	44

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

หน้า

ตารางที่ 3.1 ตัวอย่างข้อมูล 13
ตารางที่ 3.2 ตัวแปร ชื่อหัวข้อ และคำอธิบาย 14
ตารางที่ 3.3 แสดงผลการวิเคราะห์อาร์เอ็มเอฟของแต่ละร้านค้า 14
ตารางที่ 4.1 ผลการเปรียบเทียบประสิทธิภาพของแบบจำลอง 25

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

		หน้า
รูปที่	2.1 แสดงการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค	4
รูปที่	2.2 แสดงการวิเคราะห์แบ่งกลุ่มแบบเป็นขั้นตอน	6
รูปที่	2.3 แสดงการแบ่งกลุ่มข้อมูลตามความหนาแน่น	7
รูปที่	3.1 แผนผังขั้นตอนการทำงาน	12
รูปที่	4.1 กราฟแสดงผลการใช้วิธีเอลโบว์ในการหาจำนวนคลัสเตอร์	16
รูปที่	4.2 กราฟแสดงการวิเคราะห์ซิลูเอต	17
รูปที่	4.3 กราฟแสดงผลการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค	17
รูปที่	4.4 กราฟแสดงผลการแบ่งกลุ่มข้อมูลตามความหนาแน่น	18
รูปที่	4.5 กราฟแสดงผลการแบ่งกลุ่มข้อมูลตามความหนาแน่นในรูปแบบเดนโดแกรม	19
รูปที่	4.6 กราฟแสดงผลการแบ่งกลุ่มแบบเป็นขั้นตอน	20
รูปที่	4.7 กราฟแสดงผลการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเคหลังตัดข้อมูลที่มีมูลค่าสูง	21
รูปที่	4.8 กราฟแสดงผลการแบ่งกลุ่มข้อมูลตามความหนาแน่นหลังตัดข้อมูลที่มีมูลค่าสูง	21
รูปที่	4.9 กราฟแสดงผลการแบ่งกลุ่มแบบเป็นขั้นตอนหลังตัดข้อมูลที่มีมูลค่าสูง	21
รูปที่	4.10 กราฟแสดงผลการเปรียบเทียบประสิทธิภาพด้วยค่าเดวีส์โบดิน	23
รูปที่	4.11 กราฟแสดงผลการเปรียบเทียบประสิทธิภาพด้วยค่าซิลูเอต โคอิฟฟิเชียนท์	23
รูปที่	4.12 กราฟแสดงผลการเปรียบเทียบประสิทธิภาพด้วยค่าดัชนีคาลินสกี้-ฮาราบาซ	24

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันการดำเนินชีวิตของมนุษย์ส่วนใหญ่จะต้องมีเทคโนโลยีเข้ามาเกี่ยวข้อง จากเดิมที่การใช้จ่ายเป็นการจ่ายด้วยธนบัตรเป็นส่วนใหญ่แต่ในปัจจุบันได้เปลี่ยนมาเป็นสังคมไร้เงินสด ส่งผลให้การปรับธุรกิจให้มีความเข้ากับเทคโนโลยีเป็นสิ่งจำเป็นที่ขาดไม่ได้ เครื่องรับชำระเงินอิเล็กทรอนิกส์ (ELECTRONIC DATA CAPTURE) จึงเป็นสิ่งที่ตอบโจทย์ในสังคมยุคปัจจุบัน เนื่องจากเป็นเครื่องที่รับชำระเงินผ่านบัตรเครดิตของลูกค้า อีกทั้งยังเป็นเครื่องมือสำคัญที่ช่วยให้ธนาคารสามารถติดตามและวิเคราะห์พฤติกรรมของลูกค้าได้อย่างมีประสิทธิภาพ

การวิเคราะห์ข้อมูลจากเครื่องรับชำระเงินอิเล็กทรอนิกส์ (ELECTRONIC DATA CAPTURE) เป็นกระบวนการที่มีความสำคัญอย่างมากในการพัฒนาและปรับปรุงธุรกิจของธนาคารในยุคดิจิทัลและยุคการแข่งขันทางธุรกิจที่รุนแรง โดยเฉพาะเมื่อพบว่ามีปัญหาเกี่ยวกับยอดการชำระเงินที่ไม่เป็นไปตามที่คาดหวัง ปัญหาที่พบบ่อยในการใช้งานเครื่องรับชำระเงินอิเล็กทรอนิกส์ คือร้านค้าบางร้านมีเครื่องรับชำระเงินอิเล็กทรอนิกส์ แต่ไม่มียอดเข้ามาหรือมียอดน้อยเมื่อเปรียบเทียบกับคาดการณ์ ดังนั้นจึงต้องการแก้ไขปัญหานี้ เพื่อให้ธนาคารได้ประโยชน์สูงสุดจากการใช้เครื่องรับชำระเงินอิเล็กทรอนิกส์ของลูกค้า

จากปัญหาที่กล่าวมาจึงได้มีการนำข้อมูลจากเครื่องรับชำระเงินอิเล็กทรอนิกส์มาทำการวิเคราะห์การจัดกลุ่ม โดยใช้การแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค การวิเคราะห์แบ่งกลุ่มแบบเป็นขั้นตอน และการแบ่งกลุ่มข้อมูลตามความหนาแน่นในการจัดกลุ่มร้านค้าที่มียอดการชำระเงินมากและยอดการชำระเงินน้อย เพื่อนำมาวิเคราะห์และแก้ไขปัญหาให้การใช้เครื่องรับชำระเงินอิเล็กทรอนิกส์มีประโยชน์และประสิทธิภาพสูงสุดพร้อมทั้งช่วยให้ธุรกิจสามารถเติบโตและรักษาความเป็นอยู่ได้อย่างยั่งยืนในตลาดที่เปลี่ยนแปลงอย่างรวดเร็วในปัจจุบันและอนาคต

1.2 วัตถุประสงค์ของงานวิจัย

1) เพื่อแก้ไขปัญหาการใช้เครื่องรับชำระเงินอิเล็กทรอนิกส์อย่างไม่เกิดประโยชน์สูงสุด
2) เพื่อนำข้อมูลมาวิเคราะห์ในการคัดสรรกลุ่มลูกค้าเก่าและลูกค้าใหม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.3 ขอบเขตของงานวิจัย

ข้อมูลการใช้เครื่องรับชำระเงินอิเล็กทรอนิกส์จากธนาคารแห่งหนึ่ง โดยมีข้อมูลทั้งหมด 67,391 รายการ เป็นข้อมูลร้านค้า 214 ร้านค้า ซึ่งข้อมูลที่นำมาใช้เป็นข้อมูลการใช้บริการตั้งแต่วันที่ 1 มกราคม 2562 – 31 ธันวาคม 2562

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1) ได้แก้ไขปัญหาการใช้เครื่องรับชำระเงินอิเล็กทรอนิกส์เพื่อให้เกิดประโยชน์สูงสุด
2) ได้นำข้อมูลมาวิเคราะห์ในการคัดสรรกลุ่มลูกค้าเก่าและลูกค้าใหม่ เพื่อเชื่อมความสัมพันธ์กับกลุ่มลูกค้าได้มีประสิทธิภาพมากขึ้น
3) บริษัทได้กำไรจากการให้บริการเครื่องรับชำระเงินอิเล็กทรอนิกส์มากขึ้น

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

เนื้อหาในบทนี้กล่าวถึงทฤษฎีและงานวิจัยที่เกี่ยวข้องกับการศึกษา ประกอบด้วย 5 ส่วนคือ นิยามของสินเชื่อ การจัดเตรียมข้อมูล การเรียนรู้ของเครื่อง มาตรวัดประสิทธิภาพ และงานวิจัยที่เกี่ยวข้อง

2.1 เครื่องรูดบัตรแบบเคลื่อนที่ (Electronic Data Capture: EDC)

อุปกรณ์ที่อยู่ในลักษณะของเครื่องรับจ่ายเงินที่เป็นได้ทั้งบัตรเครดิต และบัตรเดบิต สามารถรับเงินผ่านทางคิวอาร์โค้ด (QR Code) ได้ โดยจะเป็นการรับจ่ายผ่านรูปแบบกระเป๋าเงินดิจิทัล (e–Wallet) อาลีเพย์ (Alipay) เช่น การใช้แอปเป๋าตัง หลักการเบื้องต้นคือทำหน้าที่รับโอนเงินเข้ามาทางบัญชีธนาคารที่ได้ผูกกับเครื่องไว้ ส่งผลให้เพิ่มความสะดวกสบาย ปลอดภัยทั้งคนจ่ายเงิน และคนรับเงิน พร้อมเพิ่มยอดขายให้กับธุรกิจได้อย่างเต็มประสิทธิภาพ ทั้งนี้เครื่องรูดบัตรแบบเคลื่อนที่ในบางธนาคารยังสามารถรับ-จ่ายบัตรสวัสดิการแห่งรัฐได้

2.2 การจัดเตรียมข้อมูล (Data Preprocessing)

2.1.1 การปกปิดข้อมูล (Data Anonymization)

เป็นกระบวนการในการเปลี่ยนแปลงข้อมูลที่อ่อนไหวเพื่อปกป้องความเป็นส่วนตัว โดยยังคงรูปแบบและลักษณะทางสถิติของข้อมูลเอาไว้ มีหลายวิธีที่ใช้ในการปกปิดข้อมูลดังนี้

1. การใช้นามแฝง (Pseudonymization) เป็นการแทนที่ข้อมูลที่อ่อนไหวด้วยค่าที่ไม่เกี่ยวข้อง เช่น การใช้ตัวอักษรหรือตัวเลขแทนชื่อ
2. การทำให้ข้อมูลมีความเฉพาะเจาะจงน้อยลง (Generalization) เช่น การเปลี่ยนอายุเป็นช่วงอายุ
3. การลบหรือปกปิดข้อมูลที่อ่อนไหวออก (Suppression)

2.1.2 การวิเคราะห์อาร์เอฟเอ็ม (RFM Analysis)

เป็นการแบ่งกลุ่มลูกค้าตามพฤติกรรมที่แตกต่างกัน โดยนำข้อมูลหลัก 3 อย่างมาวิเคราะห์ ซึ่งประกอบไปด้วย ความถี่ล่าสุด (Recency) ความถี่ (Frequency) และมูลค่า (Monetary)

1. ความถี่ล่าสุด คือ ข้อมูลของลูกค้าที่มาซื้อสินค้าหรือใช้บริการครั้งล่าสุดเมื่อไหร่
2. ความถี่ คือ ความถี่ของลูกค้าในการมาซื้อสินค้าหรือใช้บริการ
3. มูลค่า คือ ปริมาณการซื้อหรือใช้บริการของลูกค้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3 การเรียนรู้ของเครื่อง (Machine Learning)

การเรียนรู้ของเครื่อง คือการให้เครื่องเรียนรู้งานจากตัวอย่างหรือประสบการณ์จำนวนหนึ่ง เพื่อให้ทำงานได้อย่างมีประสิทธิภาพ หลังจากเรียนรู้เสร็จสิ้นจากตัวอย่างจำนวนหนึ่งอย่างเพียงพอ เครื่องนี้จะสามารถนำไปใช้ในการประมวลผลของตัวอย่างใหม่ที่ไม่เคยพบมาก่อนได้อย่างมีประสิทธิภาพ ซึ่งการเรียนรู้ของเครื่องจะมีอยู่สองประเภทใหญ่ ๆ ได้แก่ การเรียนรู้แบบมีผู้สอน (Supervised Learning) และการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning)

1. การเรียนรู้แบบมีผู้สอน

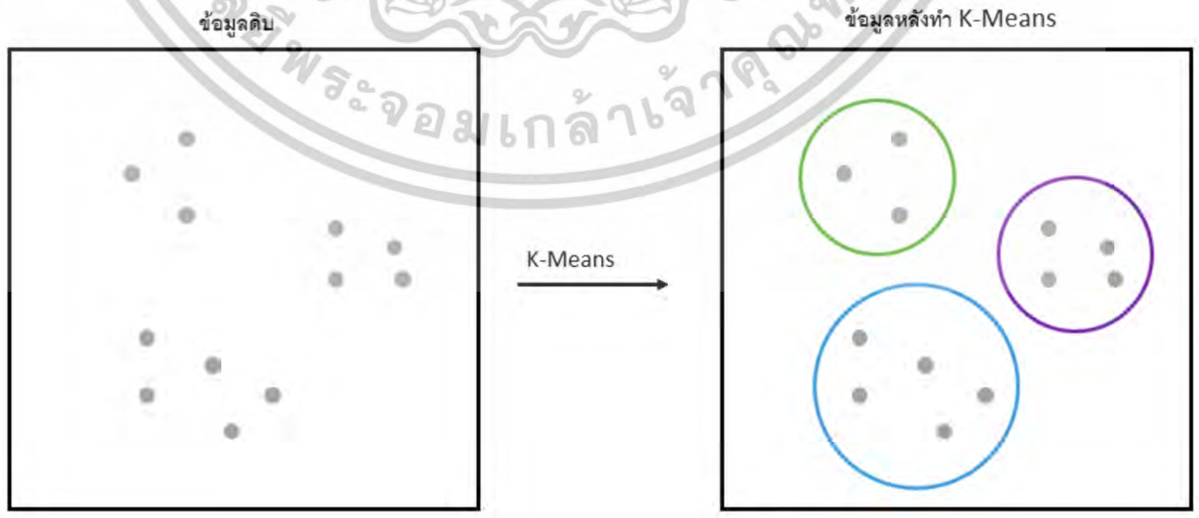
การเรียนรู้แบบมีผู้สอนเป็นการทำให้คอมพิวเตอร์สามารถหาคำตอบของปัญหาได้ด้วยตัวเอง หลังจากเรียนรู้จากชุดข้อมูลตัวอย่างไปแล้วระยะหนึ่ง

2. การเรียนรู้แบบไม่มีผู้สอน

การเรียนรู้แบบไม่มีผู้สอนเป็นการเรียนรู้ที่ให้เครื่องจักรสามารถเรียนรู้ได้ด้วยตนเอง โดยไม่ต้องมีค่าเป้าหมายของแต่ละข้อมูล ซึ่งวิธีการคือมนุษย์จะเป็นผู้ใส่ข้อมูลต่าง ๆ และกำหนดสิ่งที่ต้องการจากข้อมูลเหล่านั้น ทำให้เครื่องจักรวิเคราะห์จากการจำแนกและสร้างแบบแผนจากข้อมูลที่ได้รับมา

2.3.1 การแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค (K-Mean Clustering)

การแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค เป็นหนึ่งในการเรียนรู้แบบไม่มีผู้สอนของการเรียนรู้ของเครื่อง การทำการแบ่งกลุ่มข้อมูลหลักการคือการกำหนดจุดศูนย์กลางของกลุ่ม (Centroid) มา เค จุด และทำการคำนวณระยะห่างระหว่างชุดข้อมูล และจุดศูนย์กลางในหลักการของการวัดระยะทางแบบยุคลิด (Euclidean Distance) จากนั้นกำหนดให้ชุดข้อมูลที่มีระยะห่างน้อยที่สุดกับจุดศูนย์กลางของกลุ่มเป็นข้อมูลกลุ่มนั้น และคำนวณหากึ่งกลางใหม่ ในจุดศูนย์กลางของกลุ่มนั้น จากนั้นทำการวัดระยะทางแบบยุคลิด และวนทำแบบเดิม จนกระทั่งจุดศูนย์ถ่วงไม่มีการเปลี่ยนแปลง (สิรธีร์, 2565) แสดงดังรูปที่ 2.1



รูปที่ 2.1 แสดงการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 2.1 แสดงการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค ซึ่งการเลือกค่าเค จะใช้วิธีเอลโบว์เคิร์ (Elbow curve method) เป็นวิธีการใช้ในการเลือกค่าเคที่เหมาะสม โดยทำการหาค่าเฉลี่ยค่าความเฉื่อย (Inertia) หรือเป็นค่าของความคลาดเคลื่อน (Error) ระหว่างจุดศูนย์กลางของกลุ่ม และทุกจุดโดยเริ่มตั้งแต่ เค ที่ 1 ไปจนถึงจำนวนที่ต้องการ จากนั้นนำมาพิจารณาเลือก ค่า เค ที่เหมาะสมตามค่าความคลาดเคลื่อนที่อยู่ในจุดที่มีมุมหักเสมือนศอก (Elbow) โดยเป็นการเลือกค่าความแตกต่างของความคลาดเคลื่อนที่มากที่สุดเมื่อเทียบกับค่า เค ก่อนหน้า

- วิธีเอลโบว์ เป็นเทคนิคเพื่อกำหนดจำนวนกลุ่มที่เหมาะสม โดยเฉพาะในกระบวนการจัดกลุ่มข้อมูลแบบค่าเฉลี่ยเค หลักการของวิธีนี้คือการพิจารณาการลดลงของการสูญเสียภายในกลุ่ม(Within-Cluster Sum of Squares: WCSS) ซึ่งเป็นการวัดระยะห่างรวมระหว่างจุดข้อมูลภายในกลุ่มและจุดศูนย์กลางของกลุ่ม การคำนวณการสูญเสียภายในกลุ่มนี้จะดำเนินการซ้ำหลายครั้งโดยเพิ่มจำนวนกลุ่มขึ้นเรื่อยๆ โดยในแต่ละครั้งการสูญเสียภายในกลุ่มจะลดลง แต่จะเริ่มลดน้อยลงเมื่อถึงจุดหนึ่ง จุดดังกล่าวเรียกว่า "ข้อศอก" ซึ่งเป็นตำแหน่งที่เหมาะสมที่สุดในการเลือกจำนวนกลุ่มสำหรับกระบวนการจัดกลุ่ม การเลือกจำนวนกลุ่มในจุดข้อศอกช่วยลดความซับซ้อนของแบบจำลองโดยยังคงรักษาความแม่นยำไว้ หากเลือกจำนวนกลุ่มมากเกินไปอาจทำให้แบบจำลองซับซ้อนเกินความจำเป็น และเสี่ยงต่อการเกิดปัญหาการจำเพาะข้อมูลเกินไป (Overfitting) ในทางกลับกัน หากเลือกจำนวนน้อยเกินไป การแบ่งกลุ่มอาจไม่สามารถสะท้อนลักษณะเฉพาะของข้อมูลได้อย่างเหมาะสม

2.3.2 การวิเคราะห์แบ่งกลุ่มแบบเป็นขั้นตอน (Hierarchical Cluster Analysis)

ในการวิเคราะห์แบ่งกลุ่มแบบเป็นขั้นตอนนั้น มีวิธีการย่อยในการจัดกลุ่ม 2 วิธี คือ วิธีการรวมกลุ่ม (Agglomerative) และวิธีการแบ่งกลุ่ม (Divisive) การวิเคราะห์แบ่งกลุ่มแบบเป็นขั้นตอนด้วยวิธีการรวมกลุ่มเป็นวิธีการรวมหน่วยตัวอย่างเข้าเป็นกลุ่มเดียวกัน โดยเริ่มจากถือว่า 1 หน่วยตัวอย่างเป็น 1 กลุ่ม ถ้ามี เอ็น หน่วยตัวอย่าง แสดงว่ามี เอ็น กลุ่ม จากนั้นพิจารณาว่าจะรวมหน่วยตัวอย่างคู่ใดเข้าด้วยกัน ถ้าพิจารณาจากความคล้ายจะให้หน่วยตัวอย่างที่คล้ายกันมากที่สุดอยู่กลุ่มเดียวกัน แต่ถ้าพิจารณาจากความไม่คล้าย จะให้หน่วยตัวอย่างที่ต่างกันน้อยที่สุดอยู่ในกลุ่มเดียวกัน หลังจากนั้นทำการรวมกลุ่มต่างๆ เข้าด้วยกันทีละ 2 กลุ่ม จนกระทั่งสุดท้ายเหลือเพียง 1 กลุ่ม หรือทุกหน่วยตัวอย่างรวมอยู่ในกลุ่มเดียวกัน การวิเคราะห์แบ่งกลุ่มแบบเป็นขั้นตอนด้วยวิธีการแบ่งกลุ่ม จะตรงกันข้ามกับวิธีข้างต้น คือ จะทำการแบ่งกลุ่มหน่วยตัวอย่างทั้งหมดออกจากกันเป็น 2 กลุ่ม และทำไปเรื่อย ๆ จนกระทั่งได้จำนวนกลุ่มเท่ากับจำนวนหน่วยตัวอย่าง โดยที่วิธีการทั้ง 2 วิธีนี้ สามารถแสดงให้ดูได้ในแผนภาพ 2 มิติ ที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เรียกว่า เดนโดแกรม (Dendrogram) ซึ่งจะแสดงให้เห็นถึงการรวมกลุ่มหรือการแบ่งกลุ่มในแต่ละระดับ (ปราณี นิลกรณ์, 2548, น. 646) แสดงดังรูปที่ 2.2

รูปที่ 2.2 แสดงการวิเคราะห์แบ่งกลุ่มแบบเป็นขั้นตอน

(ที่มา Cluster Analysis (p. 72), โดย B. S. Everitt, S. Landau, M. Leese, และ D. Stahl, 2011, Chichester, UK: Wiley.)

จากรูปที่ 2.2 แสดงถึงกระบวนการวิเคราะห์แบ่งกลุ่มแบบเป็นขั้นตอน ซึ่งมีสองวิธีหลักคือ วิธีการรวมกลุ่ม และวิธีการแบ่งกลุ่ม ซึ่งสามารถอธิบายหลักการทำงานได้ดังนี้

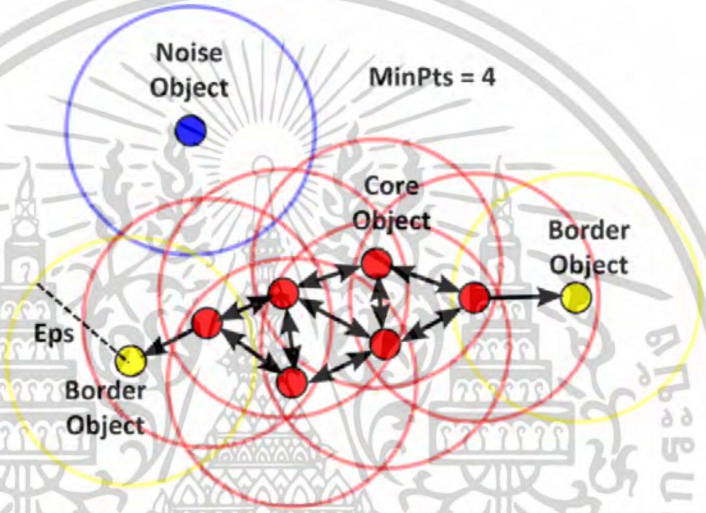
- วิธีการรวมกลุ่ม เริ่มจากการพิจารณาแต่ละจุดข้อมูลเป็นคลัสเตอร์เล็กๆ ที่ประกอบด้วยจุดเดียว จากนั้นจะค่อยๆ รวมคลัสเตอร์ที่ใกล้เคียงกันที่สุดเข้าด้วยกันไปเรื่อยๆ จนกระทั่งทุกจุดอยู่ในคลัสเตอร์เดียวกัน จากรูปจะเห็นการรวมจากซ้ายไปขวา เริ่มจากจุดข้อมูล a และ b รวมกันเป็นคลัสเตอร์ (a,b) จากนั้นรวม (a,b) เข้ากับคลัสเตอร์ที่เหลือ (c, d, และ e) จนกระทั่งสุดท้ายรวมเป็นคลัสเตอร์ใหญ่คลัสเตอร์เดียว (a,b,c,d,e)

- วิธีการแบ่งกลุ่ม เริ่มจากการพิจารณาข้อมูลทั้งหมดเป็นคลัสเตอร์เดียวขนาดใหญ่ จากนั้นจะแยกคลัสเตอร์ใหญ่นี้ออกเป็นคลัสเตอร์ย่อยๆ ที่มีความแตกต่างกันมากที่สุดเรื่อยๆ จนกระทั่งแต่ละจุดเป็นคลัสเตอร์ของตัวเอง จากรูปจะเห็นการแยกจากขวาไปซ้าย เริ่มจากคลัสเตอร์ใหญ่ (a,b,c,d,e) แยกออกเป็น (a,b) และ (c,d,e) จากนั้น (c,d,e) ถูกแยกออกเป็น c และ (d,e) สุดท้าย (d,e) แยกออกเป็น d และ e

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.3 การแบ่งกลุ่มข้อมูลตามความหนาแน่น (Density-Based Spatial Clustering of Applications with Noise: DBSCAN)

เป็นอัลกอริทึมที่ไม่ต้องกำหนดจำนวนกลุ่มข้อมูลที่ต้องการแบ่ง ใช้เพื่อค้นหาความสัมพันธ์และโครงสร้างของข้อมูลที่ยุ่งยากแต่ยังสามารถระบุรูปแบบและคาดการณ์พฤติกรรมของข้อมูลในอนาคตได้ เหมาะสำหรับข้อมูลที่มีการกระจายตัวกันแบบไม่เป็นกลุ่มก้อน มีรูปร่างเป็นรูปทรงต่างๆ ที่การแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเคไม่สามารถจัดกลุ่มได้ และจัดการกับข้อมูลผิดปกติ (Noise) ได้ดี ผลลัพธ์ที่ได้จากการจัดกลุ่มจะเป็นคลัสเตอร์ไม่จำกัดขนาดและรูปทรง ซึ่งสามารถอธิบายนิยามและหลักการทำงานได้ดังนี้



รูปที่ 2.3 แสดงการแบ่งกลุ่มข้อมูลตามความหนาแน่น

นิยามของอัลกอลิทึม จากรูปที่ 2.3 อธิบายนิยมของ DBSCAN ได้ดังนี้

1. ระยะเอปไซลอน (Eps) คือ รัศมีระหว่างจุดในกลุ่มข้อมูล ดังนั้นระยะห่างระหว่างจุดสองจุดต้องมีค่าน้อยกว่า หรือเท่ากับระยะเอปไซลอน

2. จุดข้อมูลขั้นต่ำ (MinPts) คือ จำนวนจุดขั้นต่ำสุดสำหรับการสร้างจุดศูนย์กลางของกลุ่มข้อมูล

3. จุดศูนย์กลาง (Core Point) คือ จุดที่เป็นแกนหลักของกลุ่มข้อมูล มีจำนวนจุดที่ใกล้เคียงอยู่ภายในระยะเอปไซลอน มากกว่าหรือเท่ากับจุดข้อมูลขั้นต่ำ

4. จุดขอบเขต (Border Point) คือ จุดบนเส้นขอบของกลุ่มข้อมูลที่มีจำนวนจุดที่อยู่ใกล้เคียงภายในระยะเอปไซลอน น้อยกว่าจุดข้อมูลขั้นต่ำ แต่ยังอยู่ในบริเวณใกล้เคียงจุดศูนย์กลาง

5. จุดข้อมูลที่ไม่ได้อยู่ในกลุ่ม (Noise Point) คือ จุดใดๆ ที่ไม่ได้อยู่ในกลุ่มข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การทำงานของอัลกอลิทึม การประมวลผลเริ่มจากการกำหนดจุดศูนย์กลางทุกจุดในชุดข้อมูลจะถูกตั้งค่าเป็นไม่ถูกตรวจสอบ เมื่อจุดใดที่อยู่ภายในระยะเอปไซลอนถูกเข้าถึงจากจุดแกนกลาง จุดนั้นจะถูกทำเครื่องหมายว่าถูกตรวจสอบ จากนั้นดำเนินการสำรวจทุกจุดที่อยู่ในระยะเอปไซลอน กลุ่มข้อมูลคลัสเตอร์จะถูกสร้างเมื่อจำนวนจุดมากกว่าหรือเท่ากับจุดข้อมูลขั้นต่ำ จุดที่เหลือจะถูกสร้างเป็นคอลเลกชัน N ทำซ้ำเพื่อค้นหาจุดแกนกลาง และสร้างกลุ่มข้อมูลจนกว่าจะถึงจุดใดๆ ที่ไม่สามารถจัดสรรให้กับกลุ่มข้อมูลใดได้ ซึ่งจุดเหล่านั้นจะถูกกำหนดให้เป็นจุดที่อยู่นอกกลุ่มข้อมูล

2.4 การประเมินประสิทธิภาพของแบบจำลอง (Model Evaluation)

2.4.1 ดัชนีเดวีส์-โบลดิน (Davies-Bouldin Index)

เป็นเกณฑ์ที่ใช้ในการวัดคุณภาพการจัดกลุ่มที่ใช้ในการวิเคราะห์ เพื่อการแบ่งกลุ่มข้อมูล โดยดัชนีเดวีส์-โบลดิน คือ อัตราส่วนระหว่างผลรวมของการกระจายตัวของข้อมูลในกลุ่ม และ ระยะห่างระหว่างกลุ่ม จะเห็นว่าการแบ่งกลุ่มที่ดีนั้น การกระจายตัวในกลุ่มจะต้องน้อย และ ระยะห่างระหว่างแต่ละกลุ่มจะต้องมาก หากค่าของดัชนีเดวีส์-โบลดินมีค่าเล็กที่สุดจะทำให้ได้การแบ่งแยกของกลุ่มดีที่สุด ซึ่งสำหรับการคำนวณดัชนีเดวีส์-โบลดินเป็นดังสมการต่อไปนี้

$$\frac{1}{K}\sum_{k=1}^{K} \underset{k \neq j}{max}\left\{\frac{S(U_k)+S(U_i)}{d(U_k,U_i)}\right\}$$

โดย ค่า $S(U_k)+S(U_i)$ เป็นระยะห่างของข้อมูลภายในกลุ่ม k และกลุ่ม i

ค่า (U_k, U_i) เป็นระยะห่างระหว่างจุดกึ่งกลางกลุ่ม k กับกลุ่ม i

2.4.2 ซิลูเอต โคอิฟฟิเชียนท์ (Slhouette Coefficient)

ดัชนีขึ้นฐานที่นิยมนำมาวัดคุณภาพการจัดกลุ่ม เพื่อวัดว่าข้อมูลแต่ละจุดมีความเหมือนกับกลุ่มที่ตัวเองอยู่มากหรือน้อย โดยประเมินจากระยะห่างระหว่างข้อมูลเป็นหลักและระยะห่างภายในกลุ่มเดียวกัน ซิลูเอต โคอิฟฟิเชียนท์ จะมีค่าระหว่าง -1 และ 1 โดยกลุ่มที่มีคุณภาพจะมีค่าดัชนีเข้าใกล้ 1 กล่าวคือ การจัดกลุ่มจะมีระยะห่างระหว่างกลุ่มมากและ ระยะห่างภายในกลุ่มน้อย แต่หากดัชนีมีค่าเข้าใกล้ 0 แสดงว่า การจัดกลุ่มมีการทับซ้อนกัน อย่างไรก็ตามข้อจำกัดของ ซิลูเอต โคอิฟฟิเชียนท์ คือ มีความอ่อนไหวต่อข้อมูลรบกวน อาจทำให้ดัชนีวัดประสิทธิภาพธิภาพการจัดกลุ่มผิดจากความเป็นจริง (วิทวัส, 2565) ค่าของซิลูเอต โคอิฟฟิเชียนท์ เป็นดังสมการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

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

โดย $b(i)$ หมายถึง ระยะห่างระหว่างกลุ่ม

$a(i)$ หมายถึง ระยะห่างภายในกลุ่มเดียวกัน

2.4.3 ดัชนีคาลินสกี้-ฮาราบาซ (Calinski-Harabasz Index)

อัตราส่วนของความแปรปรวนระหว่างกลุ่มกับความแปรปรวนภายในกลุ่ม โดยระยะห่างระหว่างกลุ่ม คำนวณมาจากจุดศูนย์กลางของกลุ่มถึงจุดศูนย์กลางของ ส่วนระยะห่างระหว่างกลุ่ม คำนวณจากระยะห่างของข้อมูลกับจุดศูนย์กลางของกลุ่ม คาลินสกี้-ฮาราบาซ จะมีค่าอยู่ระหว่าง 0 จนถึงอนันต์ หากมีค่าดัชนีมากแสดงว่าการจัดกลุ่มข้อมูลนั้นมีคุณภาพหรือไม่มีการซ้อนทับกันระหว่างกลุ่ม ดังนั้นกลุ่มที่แบ่งแยกกันชัดเจน ควรจะมีค่าความแปรปรวนระหว่างกลุ่มมาก (วิทวัส, 2565) ดัชนีคาลินสกี้-ฮาราบาซ เป็นดังสมการ

$$s = \frac{tr(B_k)}{tr(W_k)} \times \frac{n_E - k}{k - 1}$$

โดย $tr(W_k)$ หมายถึง ความแปรปรวนภายในกลุ่ม

$tr(B_k)$ หมายถึง ความแปรปรวนระหว่างกลุ่ม

2.5 งานวิจัยที่เกี่ยวข้อง

Iran Finance Association (2022) ได้ศึกษาการเปรียบเทียบวิธีการจัดกลุ่มทางการเงินเพื่อการกระจายความเสี่ยงและเพิ่มประสิทธิภาพพอร์ตการลงทุน งานวิจัยนี้ได้ทำการเปรียบเทียบวิธีการจัดกลุ่มหุ้นสามแบบ ได้แก่ การแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค, การจัดกลุ่มแบบลำดับชั้นและการแพร่กระจายความสัมพันธ์ (Affinity Propagation) เพื่อวิเคราะห์ว่าวิธีใดที่สามารถจัดกลุ่มหุ้นตามรูปแบบการเคลื่อนไหวได้ดีที่สุด ผลการวิจัยพบว่าวิธีการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค มีประสิทธิภาพที่สุดในการจัดกลุ่มข้อมูลทางการเงินในงานวิจัยนี้ โดยให้ค่าคุณภาพการจัดกลุ่มสูงสุดที่ 0.422 เมื่อเทียบการจัดกลุ่มแบบลำดับชั้นและการแพร่กระจายความสัมพันธ์ซึ่งมีค่าสัมประสิทธิ์ซิลูเอตต์ต่ำกว่า

Baby and Sasirekha (2013) ได้สำรวจและวิเคราะห์อัลกอริทึมการจัดกลุ่มเชิงชั้นแบบเกาะกลุ่ม (Agglomerative Hierarchical Clustering Algorithm) ซึ่งเป็นวิธีการจัดกลุ่มข้อมูลที่เริ่มต้นด้วยการพิจารณาข้อมูลแต่ละหน่วยเป็นกลุ่มแยกกัน จากนั้นจึงทำการรวมกลุ่มที่มีความใกล้เคียงกันมากที่สุดเข้าด้วยกัน จนกว่าจะเหลือเพียงกลุ่มเดียวหรือจำนวนกลุ่มตามที่กำหนดไว้ เพื่อให้สามารถนำข้อมูลที่จัด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กลุ่มแล้วไปใช้ในการวิเคราะห์ต่อไปได้อย่างมีประสิทธิภาพ ผลของวิจัยนี้ได้สรุปว่าอัลกอริทึมการจัดกลุ่มเชิงชั้นแบบเกาะกลุ่มมีข้อดีหลายประการ เช่น สามารถแสดงการเรียงลำดับของข้อมูลที่เป็นประโยชน์ในการแสดงผลข้อมูล และสามารถสร้างกลุ่มขนาดเล็กที่อาจเป็นประโยชน์ในการค้นพบข้อมูลใหม่ๆ อย่างไรก็ตาม อัลกอริทึมนี้มีข้อจำกัดเช่น ต้องใช้เวลาในการประมวลผลมากเมื่อใช้กับชุดข้อมูลขนาดใหญ่ และมีความไวต่อการกำหนดค่าพารามิเตอร์เริ่มต้น

Liu and Zhu (2021) ได้ศึกษาเกี่ยวกับการพัฒนาระบบเตือนภัยล่วงหน้าทางการเงินโดยใช้การแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเคเพื่อระบุและวิเคราะห์ปัจจัยเสี่ยงทางการเงิน กำหนดความเป็นไปได้ของความเสี่ยงที่อาจเกิดขึ้น และเพื่อให้มีการป้องกันและจัดการความเสี่ยง ผลของการวิจัยพบว่าวิธีการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค สามารถหลีกเลี่ยงผลกระทบเชิงลบจากการกำหนดเกณฑ์โดยมนุษย์ได้อย่างมีประสิทธิภาพ ปรับกระบวนการพยากรณ์ความเสี่ยงทางการเงินอย่างต่อเนื่อง และแจกจ่ายชุดข้อมูลเป้าหมายไปยังแต่ละศูนย์กลางกลุ่มเพื่อหาวิธีการที่เหมาะสมที่สุด ทำให้สามารถจำแนกสถานะของความเสี่ยงทางการเงินได้อย่างแม่นยำและเป็นกลาง

Abdulhafedh (2021) ได้ศึกษาการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเคและการจัดกลุ่มแบบลำดับชั้นในการวิเคราะห์และจำแนกกลุ่มลูกค้าบัตรเครดิต, ปรับปรุงกลยุทธ์การตลาดและเปรียบเทียบประสิทธิภาพของเทคนิคต่างๆ เพื่อแนะนำการปรับปรุงการให้บริการลูกค้า ผลการวิจัยพบว่าการวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis: PCA) ร่วมกับการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเคและการจัดกลุ่มแบบลำดับชั้นเพิ่มประสิทธิภาพในการจัดกลุ่มจาก 3 กลุ่มเป็น 4 กลุ่ม ช่วยให้บริษัทสามารถปรับกลยุทธ์การตลาดที่เหมาะสมกับกลุ่มลูกค้าแต่ละกลุ่มได้ดียิ่งขึ้น

Dong and Jin (2007) ได้ศึกษาการแบ่งกลุ่มแบบลำดับขั้น การแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค และ แผนที่ลักษณะการจัดระเบียบตัวเอง (Self-Organizing Feature Map: SOFM) เพื่อเปรียบเทียบประสิทธิภาพและความเหมาะสมในการวิเคราะห์ข้อมูลการแสดงออกของยีนแต่ละประเภท ผลการวิจัยพบว่าวิธีการแต่ละแบบมีข้อดีและข้อเสียแตกต่างกัน และไม่มีวิธีใดที่เหมาะสมที่สุดสำหรับข้อมูลทุกประเภท การใช้วิธีการหลายแบบในการวิเคราะห์ข้อมูลชุดเดียวกันอาจช่วยให้ได้ผลลัพธ์ที่แม่นยำและครอบคลุมมากขึ้น

Alfian, Althof, Mufidah, Syarif and Wulandari (2024) ได้การเปรียบเทียบประสิทธิภาพของอัลกอริธึมการแบ่งกลุ่มข้อมูลตามความหนาแน่น, การแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค และการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเอ็กซ์ในการจัดกลุ่มข้อมูลการชอปปิง โดยหาว่าอัลกอริธึมใดให้ผลลัพธ์ที่ดีที่สุดในด้านการสร้างคลัสเตอร์ และใช้ดัชนีเดวีส์-โบลดินในการวัดความถูกต้อง ซึ่งผลงานวิจัยพบว่าการแบ่งกลุ่มข้อมูลตามความหนาแน่นให้ผลลัพธ์ที่ดีที่สุด โดยมีดัชนีเดวีส์-โบลดินอยู่ที่ 0.1973

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

วิทวัส แสงสว่าง (2565) งานวิจัยนี้มุ่งแก้ปัญหาการกำหนดราคาขายเฉลี่ยต่อตารางเมตรของโครงการที่อยู่อาศัย ซึ่งเป็นความท้าทายสำหรับผู้ประกอบการอสังหาริมทรัพย์ เนื่องจากวิธีแบ่งตามขอบเขตความเป็นเมืองไม่สะท้อนสภาพตลาดที่แท้จริงอีกต่อไป จากการทดสอบด้วยอัลกอริทึมการแบ่งกลุ่มข้อมูลตามความหนาแน่น และ การจัดกลุ่มข้อมูลตามความหนาแน่นแบบลำดับขั้น (Hierarchical Density-Based Spatial Clustering of Applications with Noise: HDBSCAN) กับข้อมูล 2,010 โครงการในกรุงเทพฯ พบว่า HDBSCAN ให้ผลการจัดกลุ่มที่มีคุณภาพดีที่สุด หลังจากนั้น ได้นำกลุ่มที่มีคุณภาพดีไปสร้างแบบจำลองเพื่อทำนายราคาเฉลี่ยต่อตารางเมตร โดยใช้ปัจจัยโครงการและสภาพตลาด พร้อมทดสอบผลการทำนายด้วย RMSE และ MAPE เพื่อทำให้การกำหนดราคามีความสมเหตุสมผลมากขึ้นในอนาคต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

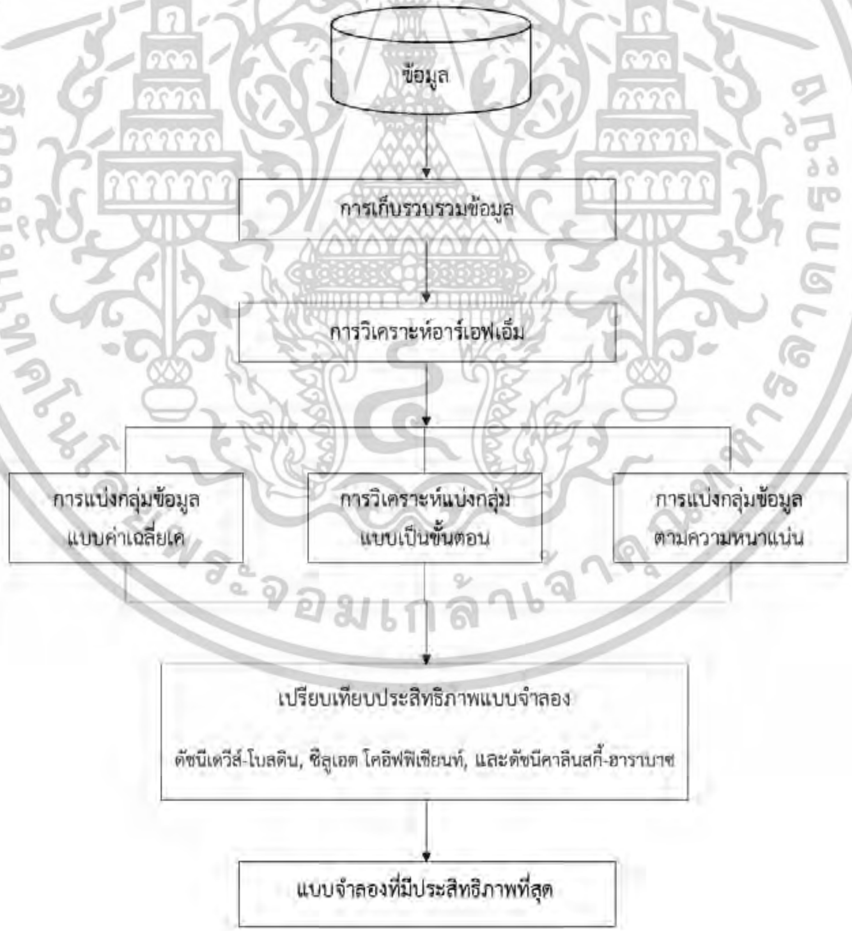
บทที่ 3

วิธีการดำเนินงานวิจัย

งานวิจัยนี้ได้มีการนำข้อมูลจากเครื่องรับชำระเงินอิเล็กทรอนิกส์มาทำการวิเคราะห์ โดยใช้การแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค การวิเคราะห์แบ่งกลุ่มแบบเป็นขั้นตอน และการแบ่งกลุ่มข้อมูลตามความหนาแน่นเพื่อนำมาวิเคราะห์และแก้ไขปัญหาการใช้เครื่องรับชำระเงินอิเล็กทรอนิกส์อย่างไม่เกิดประโยชน์สูงสุด ซึ่งใช้ดัชนีเดวีส์-โบลดิน, ค่าซิลูเอต โคอิฟฟิเชียนท์ และดัชนีคาลินสกี้-ฮาราบาซในการเปรียบเทียบประสิทธิภาพ

3.1 แผนผังขั้นตอนการทำงาน

แผนขั้นตอนการทำงานสามารถเขียนสรุปเป็นแผนผังได้ ดังรูปที่ 3.1



รูปที่ 3.1 แผนผังขั้นตอนการทำงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 การจัดเตรียมข้อมูล (Data Preprocessing)

3.2.1 การเก็บรวบรวมข้อมูล (Data Collection)

ผู้วิจัยได้ข้อมูลมาจากข้อมูลจริงของธนาคารแห่งหนึ่ง ซึ่งเป็นข้อมูลการใช้บัตรเครดิต และบัตรเดบิตของเครื่องรับชำระเงินอิเล็กทรอนิกส์ โดยชุดข้อมูลมีจำนวนทั้งสิ้น 67,391 รายการ เป็นข้อมูลร้านค้า 214 ร้านค้า แต่ข้อมูลที่นำมาใช้มีข้อมูลบางส่วนเป็นข้อมูลที่มีความอ่อนไหว ทางผู้วิจัยจึงได้มีการเปลี่ยนแปลงข้อมูลในส่วนนี้เพื่อปกป้องความเป็นส่วนตัว ตัวอย่างของข้อมูลและรายละเอียดของข้อมูลต่างๆ แสดงดังตาราง 3.1 และ 3.2

ตารางที่ 3.1 ตัวอย่างข้อมูล

Merchant ID	Merchant Name	Trx.Date	Card No.	Trx Amount	Credit Merchant
16500070001	1010c1d0c3a8e6ead13b2569c4229ff5	05/10/2019 10:54:23	de94b4c511155c15aebb40dbc3fe1a4e	1,190.00	1,164.53
16500200002	c8ac7c7bfd6ff4b790ea0dcfa3b69a49	26/09/2019 23:23:17	b21353ce72b3a29200d475b31cac1883	2,000.00	1,965.76
16500200002	c8ac7c7bfd6ff4b790ea0dcfa3b69a49	29/10/2019 16:26:38	38b1c9d795645c299c1f784b56db5336	2,000.00	1,965.76
16500070001	1010c1d0c3a8e6ead13b2569c4229ff5	07/10/2019 11:42:06	3ccd79eb7b3abf23bbe53852910a889f	798.00	780.92
16500470003	ad66cda3338706cd97a695619e277eec	09/11/2019 17:27:15	c0f3187a8b3e024e2498bb406a4efa49	3,590.00	3,532.38
16500680009	0cb9ad6cc273385f1500c33a020838bf	18/06/2019 16:01:57	eb510f331b6299c156b76aece769c50a	8,653.87	8,514.97
16502150002	d2b0d83088133bb5385fa4d94391f0bf	28/07/2019 09:06:04	28effeaf71341646064eb67c73288f9b	1,350.00	1,332.67
16500800007	7a6754e2a3ef95bdfd70bbad1a3e305b	28/01/2019 20:25:26	78c4f32afaf1ac43c59bda04cab6a54b	7,800.00	7,649.77
16500470003	ad66cda3338706cd97a695619e277eec	31/08/2019 18:20:16	57a56df2c2cd3349ccbb786feb3e3373	9,990.00	9,829.66
16500070001	1010c1d0c3a8e6ead13b2569c4229ff5	27/07/2019 14:12:03	32e518a230a0ba9842aa12f3f890650d	540.00	528.44
16500470003	ad66cda3338706cd97a695619e277eec	21/07/2019 13:47:00	174dd0ece3bf9a2cbb8f4fa556e74680	1,490.00	1,466.09
16500070001	1010c1d0c3a8e6ead13b2569c4229ff5	23/07/2019 17:19:39	f1cc93f1bf19daeeda44ce6b5c2644e9	1,010.00	988.39
16500200002	c8ac7c7bfd6ff4b790ea0dcfa3b69a49	03/09/2019 01:14:27	6de671a0c3d4b9d99f7f3aa0074c34dd	3,500.00	3,440.08
16500470003	ad66cda3338706cd97a695619e277eec	16/04/2019 09:42:20	e973adfa9ee9d58533a99ee754bfd3cb	7,490.00	7,369.79
16500680010	1a7400cccb85d374748674f33912325e	27/08/2019 15:58:17	b0062bd239bb33ef13b70eb301ff7c03	2,620.43	2,581.17

ตารางที่ 3.2 ตัวแปร ชื่อหัวข้อ และคำอธิบาย

ลำดับ	ชื่อหัวข้อ	คำอธิบาย
1	merchant_id	รหัสร้านค้า
2	trx_amount	จำนวนเงินของการใช้เครื่องรับชำระเงินอิเล็กทรอนิกส์
3	credit_merchant	จำนวนเงินที่ร้านค้าได้รับหลังจากหักค่าธรรมเนียมและภาษี
4	datetime	วันที่ใช้บริการเครื่องรับชำระเงินอิเล็กทรอนิกส์

3.1.2 การวิเคราะห์อาร์เอฟเอ็ม (RFM Analysis)

การวิเคราะห์อาร์เอฟเอ็ม ผู้วิจัยทำการรวบรวมข้อมูลการใช้บริการเครื่องรับชำระเงินอิเล็กทรอนิกส์ของแต่ละร้านค้าโดยจัดกลุ่มตามรหัสร้านค้า จากนั้นสร้างตัวแปรใหม่ขึ้นโดยแสดงตัวอย่างดังตารางที่ 3.2 ซึ่งมีรายละเอียดดังนี้

1. จำนวนวันที่ลูกค้าซื้อล่าสุด เป็นการดูค่าความถี่ข้อมูลของลูกค้าที่มาซื้อสินค้าครั้งล่าสุด ลบด้วยวันที่กำหนดให้เป็นวันที่ปัจจุบันนั่นก็คือวันที่ 1 มกราคม 2563 หากค่าจำนวนวันที่ลูกค้าซื้อล่าสุดมีค่ามากจะแสดงถึงการไม่ได้ใช้บริการของเครื่องรับชำระเงินอิเล็กทรอนิกส์
2. จำนวนการใช้บริการของแต่ละร้านค้า เป็นการดูค่าความถี่ของการใช้บริการเครื่องรับชำระเงินอิเล็กทรอนิกส์
3. ยอดรวมของแต่ละร้านค้า เป็นการดูจำนวนเงินทั้งหมดที่ใช้เครื่องรับชำระเงินอิเล็กทรอนิกส์ของแต่ละร้านค้า

ตารางที่ 3.3 แสดงผลการวิเคราะห์อาร์เอ็มเอฟของแต่ละร้านค้า

Merchant_Id	Recency	Frequency	Monetary
16500070001	20	82	95688.88
16500070002	27	14	32880.0
16500070004	145	10	26221.0
16500170002	61	22	366199.0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3 การสร้างแบบจำลอง

นำชุดข้อมูลมาสร้างแบบจำลอง โดยแบบจำลองที่เลือกนำมาจากการศึกษาวิจัยที่เกี่ยวข้องรวมถึงการศึกษาข้อมูลของแต่ละแบบจำลอง แบบจำลองที่เลือกประกอบด้วย 3 แบบจำลองดังนี้

1. การแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค
2. การแบ่งกลุ่มข้อมูลตามความหนาแน่น
3. การวิเคราะห์แบ่งกลุ่มแบบเป็นขั้นตอน

3.4 การเปรียบเทียบประสิทธิภาพแบบจำลอง

การเปรียบเทียบประสิทธิภาพแบบจำลองด้วยมาตรวัดประสิทธิภาพจำนวน 3 ค่า ดังนี้

1. ดัชนีเดวีส์-โบลดิน
2. ซิลูเอต โคอิฟฟิเชียนท์
3. ดัชนีคาลินสกี้-ฮาราบาซ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

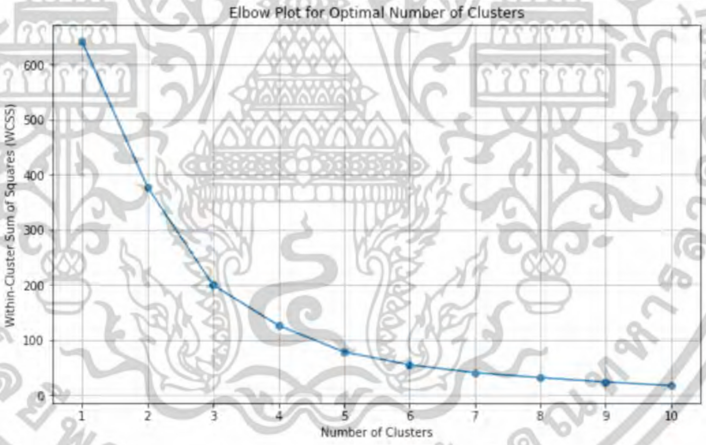
บทที่ 4

ผลการวิจัยและการอภิปรายผล

งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์และแก้ไขปัญหาการใช้เครื่องรับชำระเงินอิเล็กทรอนิกส์อย่างไม่เกิดประโยชน์สูงสุด โดยวิธีการเรียนรู้ของเครื่อง ซึ่งนำข้อมูลมาชุดข้อมูลจริงของธนาคารแห่งหนึ่ง จากนั้นทำการการเตรียมข้อมูล เพื่อจะไปสู่ขั้นตอนของการสร้างแบบจำลอง โดยแบบจำลองเป็นแบบจำลองการเรียนรู้ของเครื่องแบบไม่มีผู้สอน 3 วิธี คือ การแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค การแบ่งกลุ่มข้อมูลตามความหนาแน่น และการวิเคราะห์แบ่งกลุ่มแบบเป็นขั้นตอน ขั้นตอนสุดท้ายทำการเปรียบเทียบประสิทธิภาพแบบจำลอง โดยผู้วิจัยกำหนดใช้ดัชนีเดวีส์-โบลดิน, ค่าซิลูเอต โคอิฟฟิเชียนท์ และดัชนีคาลินสกี้-ฮาราบาซในการเปรียบเทียบประสิทธิภาพ

4.1 ผลการวิเคราะห์ของการเรียนรู้ของเครื่อง

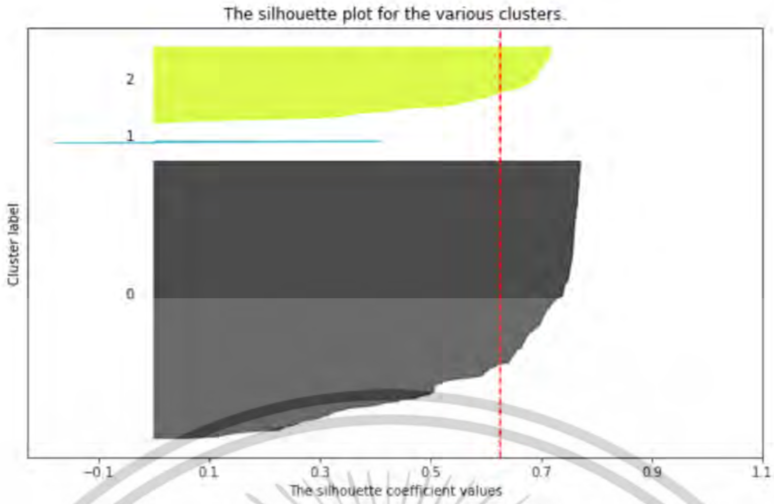
จากการสร้างแบบจำลองการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเคและแบบจำลองการแบ่งกลุ่มข้อมูลตามความหนาแน่นซึ่งใช้วิธีเอลโบว์ในการหาจำนวนคลัสเตอร์ที่เหมาะสม แสดงผลดังรูปที่ 4.1



รูปที่ 4.1 กราฟแสดงผลการใช้วิธีเอลโบว์ในการหาจำนวนคลัสเตอร์

จากรูปที่ 4.1 เป็นกราฟที่แสดงผลการใช้วิธีเอลโบว์ในการหาจำนวนคลัสเตอร์ ซึ่งจากกราฟจะเห็นว่าจุดหักศอกของเส้นกราฟจะอยู่ที่จุด 3 ดังนั้นผู้วิจัยจึงเลือกจำนวนลัสเตอร์ที่ 3 กลุ่ม ในการสร้างแบบจำลองการแบ่งกลุ่มแบบค่าเฉลี่ยเคและแบบจำลองการแบ่งกลุ่มข้อมูลตามความหนาแน่น หลังจากการใช้วิธีเอลโบว์เพื่อหาจำนวนกลุ่มที่เหมาะสมสำหรับการแบ่งกลุ่มแบบค่าเฉลี่ยเค ผู้วิจัยได้ใช้การวิเคราะห์ซิลูเอตเพื่อตรวจสอบความถูกต้องและประสิทธิภาพของการจัดกลุ่มที่ได้ โดยการวิเคราะห์ซิลูเอตจะช่วยวัดระดับการกระจายตัวของข้อมูลภายในกลุ่มที่จัดไว้และความแตกต่างจากกลุ่มอื่นๆ แสดงผลดังรูปที่ 4.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

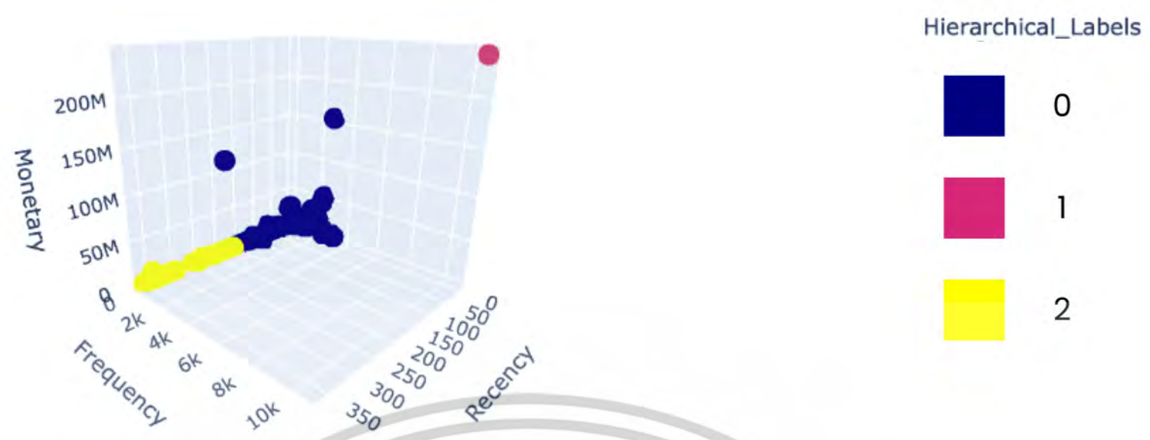


รูปที่ 4.2 กราฟแสดงการวิเคราะห์ซิลูเอต

จากรูปที่ 4.2 เป็นกราฟแสดงการวิเคราะห์ซิลูเอต ซึ่งจากกราฟจะเห็นว่าค่าซิลูเอตโดยเฉลี่ยสำหรับการแบ่งกลุ่ม 3 กลุ่มนั้นมีค่าเฉลี่ยอยู่ในระดับที่เหมาะสม เนื่องจากมีค่าเข้าใกล้ 1 ซึ่งบ่งชี้ว่าการแบ่งกลุ่มนี้สามารถแบ่งแยกข้อมูลได้อย่างชัดเจนและมีความแม่นยำในการจัดกลุ่ม ต่อไปจะเป็นการแสดงผลของการวิเคราะห์ข้อมูลด้วยวิธีการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค แสดงผลดังรูปที่ 4.3 การวิเคราะห์ข้อมูลด้วยการแบ่งกลุ่มข้อมูลตามความหนาแน่น แสดงผลดังรูปที่ 4.4

รูปที่ 4.3 กราฟแสดงผลการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



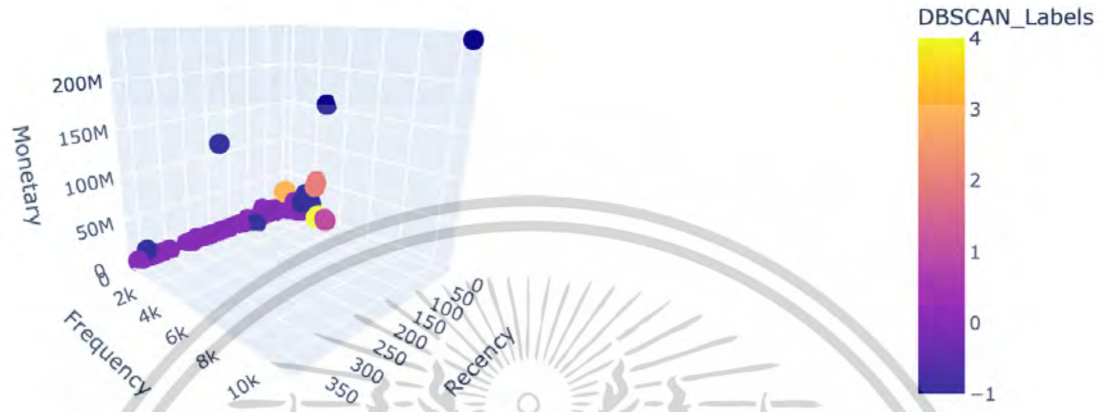
รูปที่ 4.4 กราฟแสดงผลการแบ่งกลุ่มข้อมูลตามความหนาแน่น

จากรูปที่ 4.3 เป็นกราฟแสดงผลของการวิเคราะห์การจัดกลุ่มข้อมูลแบบค่าเฉลี่ยเค และรูปที่ 4.4 เป็นกราฟแสดงผลการวิเคราะห์การจัดกลุ่มข้อมูลตามความหนาแน่นในรูปแบบ 3 มิติโดยใช้ตัวแปร จำนวนวันที่ลูกค้าซื้อล่าสุด, จำนวนการใช้บริการของแต่ละร้านค้า, และยอดรวมของแต่ละร้านค้า ข้อมูลถูกแบ่งออกเป็นสามคลัสเตอร์ (แสดงด้วยสีที่แตกต่างกัน) ซึ่งแต่ละคลัสเตอร์แสดงถึงกลุ่มข้อมูลที่มีความคล้ายคลึงกัน จากกราฟข้อมูลสีเหลืองเป็นกลุ่มที่ร้านค้าไม่ได้ใช้บริการเครื่องรับบัตรเครดิตนานที่สุด จึงเป็นกลุ่มที่ทางธนาคารจำเป็นต้องกระตุ้นร้านค้าหรือปิดบริการ กลุ่มต่อมาคือกลุ่มสีน้ำเงิน เป็นกลุ่มที่ใช้บริการเป็นปกติอาจจะมีการนำโปรโมชั่นเข้าไปนำเสนอเพื่อกระตุ้นร้านค้าให้ใช้บริการเครื่องรับบัตรเครดิตของธนาคารมากขึ้น และกลุ่มสุดท้ายคือกลุ่มสีชมพู กลุ่มนี้เป็นกลุ่มร้านค้าชั้นยอดที่ใช้บริการเครื่องรูดบัตรของธนาคารได้อย่างต่อเนื่องและใช้เป็นจำนวนมากร้านค้ากลุ่มนี้จึงเป็นกลุ่มที่ไม่ควรให้เกิดการยกเลิกการใช้บริการ ดังนั้นจึงอาจจะมีการลดค่าธรรมเนียม หรือของสัมมนาคุณเพื่อเป็นการขอบคุณที่ใช้บริการของธนาคารมาตลอด นอกจากนี้กราฟแสดงผลการวิเคราะห์การจัดกลุ่มข้อมูลตามความหนาแน่นสามารถแสดงในรูปแบบเดนโดแกรม (Dendrogram) ด้วยวิธีการแบ่งกลุ่ม (Divisive) แสดงดังรูปที่ 4.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 4.5 กราฟแสดงผลการแบ่งกลุ่มข้อมูลตามความหนาแน่นในรูปแบบเดนโดแกรม

การวิเคราะห์การแบ่งกลุ่มแบบเป็นขั้นตอน เป็นวิธีที่ช่วยในการแบ่งกลุ่มข้อมูลโดยไม่ต้องกำหนดจำนวนกลุ่มล่วงหน้า ผลการศึกษาของการใช้วิธีนี้ได้ทดสอบการแบ่งกลุ่มข้อมูลตามลำดับขั้นตอนโดยใช้เกณฑ์วัดระยะทางระหว่างข้อมูลแต่ละจุด โดยผลการวิเคราะห์แสดงในรูปที่ 4.6

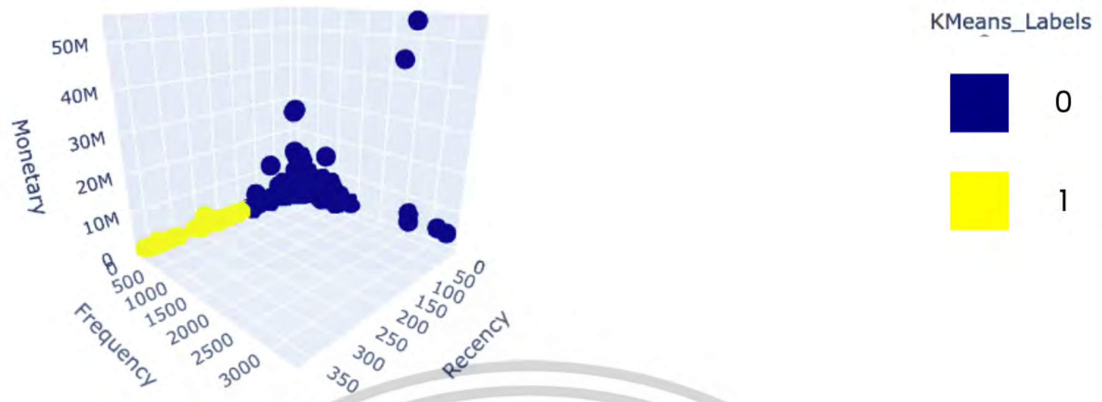


รูปที่ 4.6 กราฟแสดงผลการแบ่งกลุ่มแบบเป็นขั้นตอน

จากรูปที่ 4.6 เป็นกราฟแสดงผลการวิเคราะห์การแบ่งกลุ่มแบบเป็นขั้นตอน สีที่ปรากฏบนกราฟจะแสดงถึงกลุ่มที่แบ่งออกจากกัน ซึ่งการแบ่งกลุ่มแบบเป็นขั้นตอนสามารถจัดกลุ่มข้อมูลที่มีความหนาแน่นสูงเข้าไว้ด้วยกัน และสามารถแยกจุดข้อมูลที่อยู่โดดเดี่ยวหรืออยู่นอกกลุ่มออกมาเป็นข้อมูลผิดปกติได้อย่างชัดเจน จากกราฟแสดงให้เห็นกลุ่มสีน้ำเงิน (ค่า -1) กลุ่มนี้ประกอบด้วยจุดข้อมูลที่ไม่ได้ถูกจัดอยู่ในกลุ่มใดเลย ซึ่งหมายถึงร้านค้าที่มีพฤติกรรมแตกต่างจากกลุ่มอื่นอย่างมาก เช่น ร้านค้าที่เลิกใช้บริการไปแล้วหรือเป็นร้านค้าที่มียอดการใช้บริการสูงกว่าร้านอื่นเป็นอย่างมาก กลุ่มสีอื่นเป็นกลุ่มที่ร้านค้าที่มีพฤติกรรมการใช้งานต่อเนื่อง แต่แตกต่างกันในด้านยอดใช้จ่ายและความถี่การใช้บริการ

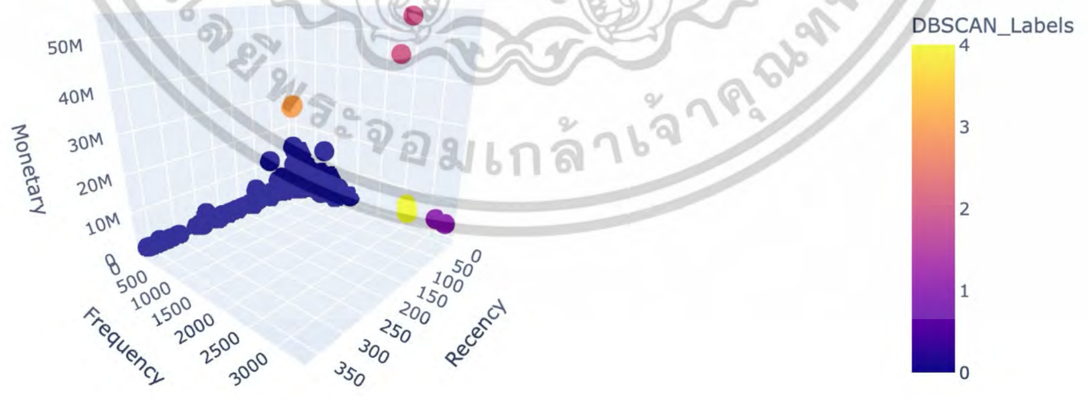
หลังจากนำแบบจำลองทั้ง 3 วิธีมาใช้ในการแบ่งกลุ่มข้อมูล ผลการวิเคราะห์จากรูปที่ 4.3, 4.4, 4.5 และ 4.6 แสดงให้เห็นว่ามีร้านค้าบางแห่งที่มีความโดดเด่นอย่างชัดเจนในทุกด้าน ทั้งในแง่ความถี่ในการใช้บริการ ความถี่การใช้บริการของแต่ละร้านค้า และยอดรวมการใช้บริการที่สูงกว่าร้านค้าอื่นๆ อย่างมาก ด้วยเหตุนี้ผู้วิจัยจึงพิจารณาจัดข้อมูลร้านค้าเหล่านี้ให้อยู่ในหมวดข้อมูลที่มีมูลค่าสูง (Extremely Valuable Data) และได้ตัดข้อมูลส่วนนี้ออกจากการวิเคราะห์เพื่อให้ผลลัพธ์ที่ได้มีความชัดเจนมากยิ่งขึ้น หลังจากการตัดข้อมูลที่มีมูลค่าสูงออกไปแล้ว การแบ่งกลุ่มข้อมูลใหม่ด้วยแบบจำลองทั้ง 3 วิธีให้ผลลัพธ์ดังนี้ รูปที่ 4.7 แสดงผลการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค, รูปที่ 4.8 แสดงผลการแบ่งกลุ่มตามความหนาแน่น และรูปที่ 4.9 แสดงผลการแบ่งกลุ่มแบบเป็นขั้นตอน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.7 กราฟแสดงผลการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเคหลังตัดข้อมูลที่มีมูลค่าสูง

รูปที่ 4.8 กราฟแสดงผลการแบ่งกลุ่มข้อมูลตามความหนาแน่นหลังตัดข้อมูลที่มีมูลค่าสูง



รูปที่ 4.9 กราฟแสดงผลการแบ่งกลุ่มแบบเป็นขั้นตอนหลังตัดข้อมูลที่มีมูลค่าสูง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากกราฟในรูปที่ 4.7, 4.8 และ 4.9 แสดงผลลัพธ์ของการแบ่งกลุ่มข้อมูลด้วยแบบจำลองทั้ง 3 วิธี หลังจากทำการตัดข้อมูลที่มีมูลค่าสูงออกไปแล้ว พบว่าโครงสร้างของการจัดกลุ่มมีความชัดเจนมากขึ้น โดยการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเคและการแบ่งกลุ่มตามความหนาแน่นสามารถแบ่งข้อมูลออกเป็นสองกลุ่มข้อมูลที่แสดงด้วยสีที่แตกต่างกัน ซึ่งกลุ่มแรกมีจำนวนการใช้บริการค่อนข้างสูง แต่มีความถี่และจำนวนเงินรวมที่ต่ำกว่าแสดงด้วยจุดสีเหลือง ขณะที่กลุ่มที่สองมีการกระจายตัวกว้างขึ้นในด้านจำนวนเงินรวม แสดงถึงร้านค้าที่มีความสม่ำเสมอในการใช้บริการและยอดรวมสูงกว่ากลุ่มแรกแสดงด้วยจุดสีน้ำเงิน ในกรณีของการแบ่งกลุ่มแบบเป็นขั้นตอนซึ่งเป็นวิธีการแบ่งกลุ่มที่อาศัยความหนาแน่นของข้อมูล พบว่ามีการแบ่งข้อมูลออกเป็นหลายกลุ่มมากกว่าวิธีอื่น ซึ่งอาจเกิดจากโครงสร้างของข้อมูลที่ยังคงมีการกระจายตัวของจุดข้อมูลที่แตกต่างกันในมิติของอาร์เอฟเอ็ม ถึงแม้จะมีการตัดข้อมูลที่มีมูลค่าสูงออกไปแล้ว นอกจากนี้ การแบ่งกลุ่มแบบเป็นขั้นตอนยังสามารถจำแนกจุดข้อมูลที่อยู่นอกกลุ่มหลักให้เป็นกลุ่มขนาดเล็กหรือข้อมูลนอกกลุ่ม (Outlier) ได้ ส่งผลให้ในกราฟแสดงจุดที่มีสีแตกต่างกันมากขึ้นเมื่อเทียบกับการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเคและการแบ่งกลุ่มตามความหนาแน่น

จากการวิเคราะห์นี้ การตัดข้อมูลที่มีมูลค่าสูงช่วยให้สามารถมองเห็นลักษณะของกลุ่มข้อมูลที่อยู่ในเกณฑ์เฉลี่ยได้อย่างชัดเจนยิ่งขึ้น ทำให้การแบ่งกลุ่มสามารถระบุความแตกต่างระหว่างกลุ่มร้านค้าในลักษณะที่แม่นยำมากขึ้น การแบ่งกลุ่มนี้สะท้อนให้เห็นถึงพฤติกรรมการใช้งานร้านค้าตามตัวแปรจำนวนวันที่ลูกค้าซื้อล่าสุด, จำนวนการใช้บริการของแต่ละร้านค้าและยอดรวมของแต่ละร้านค้า ที่ถูกลดความซับซ้อนลงหลังจากตัดข้อมูลที่มีความโดดเด่นสูงสุดออกไป โดยกลุ่มสีเหลืองเป็นกลุ่มที่มีค่าจำนวนการใช้บริการเครื่องรับชำระเงินอิเล็กทรอนิกส์สูง แต่มีค่าจำนวนวันที่ลูกค้าซื้อล่าสุดและยอดรวมของแต่ละร้านค้าอยู่ในระดับต่ำกว่ากลุ่มสีอื่นๆ ขณะที่กลุ่มสีน้ำเงินมีการกระจายตัวที่กว้างขึ้นในด้านยอดรวมของแต่ละร้านค้าแสดงถึงร้านค้าที่มีความสม่ำเสมอในการใช้บริการและยอดรวมสูงกว่ากลุ่มสีเหลือง

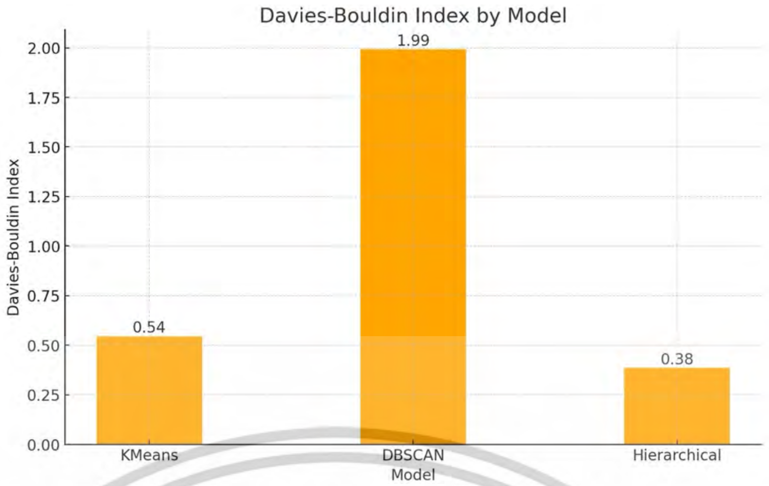
4.2 การเปรียบเทียบแบบจำลอง

จากการนำแบบจำลองการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค การวิเคราะห์แบ่งกลุ่มแบบเป็นขั้นตอนและการแบ่งกลุ่มข้อมูลตามความหนาแน่นไปทำการแบ่งกลุ่มข้อมูลการใช้เครื่องรับชำระเงินอิเล็กทรอนิกส์ของแต่ละร้านค้า และใช้ดัชนีเดวีส์-โบลดิน, ค่าซิลูเอต โคอิฟฟิเชียนท์ และดัชนีคาลินสกี้-ฮาราบาซในการเปรียบเทียบประสิทธิภาพของแบบจำลองและประเมินความเหมาะสมของการจัดกลุ่มในแต่ละแบบจำลอง

4.2.1 ดัชนีเดวีส์-โบลดิน

ดัชนีเดวีส์-โบดิน เป็นตัวชี้วัดที่ใช้ประเมินความกะทัดรัดและการกระจายตัวของกลุ่มที่ได้ โดยค่าเดวีส์-โบดิน ที่ต่ำสะท้อนถึงการแบ่งกลุ่มที่มีประสิทธิภาพมากกว่า แสดงผลดังรูปที่ 4.10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

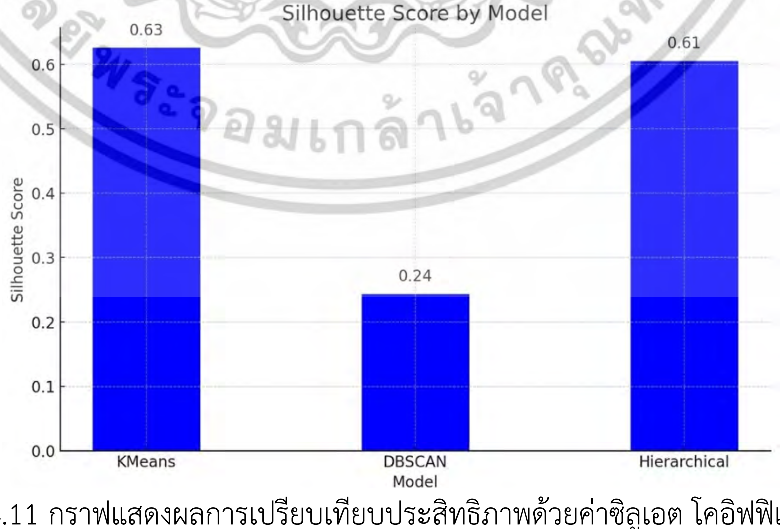


รูปที่ 4.10 กราฟแสดงผลการเปรียบเทียบประสิทธิภาพด้วยค่าเดวีส์โบดิน

จากกราฟในรูป 4.10 แสดงผลการเปรียบเทียบประสิทธิภาพของ 3 แบบจำลองด้วยค่าเดวีส์โบดิน ซึ่งจากกราฟพบว่าการแบ่งกลุ่มข้อมูลตามความหนาแน่น ให้ค่าเดวีส์-โบดินสูงสุดที่ 1.992 ซึ่งหมายถึงว่าการจัดกลุ่มแบบนี้อาจมีการทับซ้อนของข้อมูลในกลุ่มต่างๆมากที่สุด และมีประสิทธิภาพน้อยที่สุดเมื่อเทียบกับแบบอื่น ขณะที่การแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค ให้ค่าเดวีส์-โบดินที่ 0.544 ซึ่งต่ำกว่าการแบ่งกลุ่มตามความหนาแน่น แสดงถึงการกระจายข้อมูลที่ดีกว่าและมีความเป็นกลุ่มที่ชัดเจนมากขึ้น การวิเคราะห์กลุ่มแบบเป็นขั้นตอนให้ค่าเดวีส์-โบดินที่ 0.385 ซึ่งเป็นค่าที่ต่ำสุดแสดงถึงประสิทธิภาพสูงสุดในการจัดกลุ่มข้อมูลในลักษณะนี้ เนื่องจากกลุ่มมีความแตกต่างกันชัดเจนมากกว่าแบบอื่น

4.2.2 ค่าซิลูเอต โคอิฟฟิเชียนท์

ค่าซิลูเอต โคอิฟฟิเชียนท์เป็นตัวชี้วัดที่ใช้ประเมินคุณภาพของการจัดกลุ่มข้อมูล โดยค่าซิลูเอตโคอิฟฟิเชียนท์ที่สูงกว่าบ่งบอกถึงกลุ่มที่มีความชัดเจนและการกระจายตัวของข้อมูลภายในกลุ่มที่ดี แสดงผลดังรูปที่ 4.11



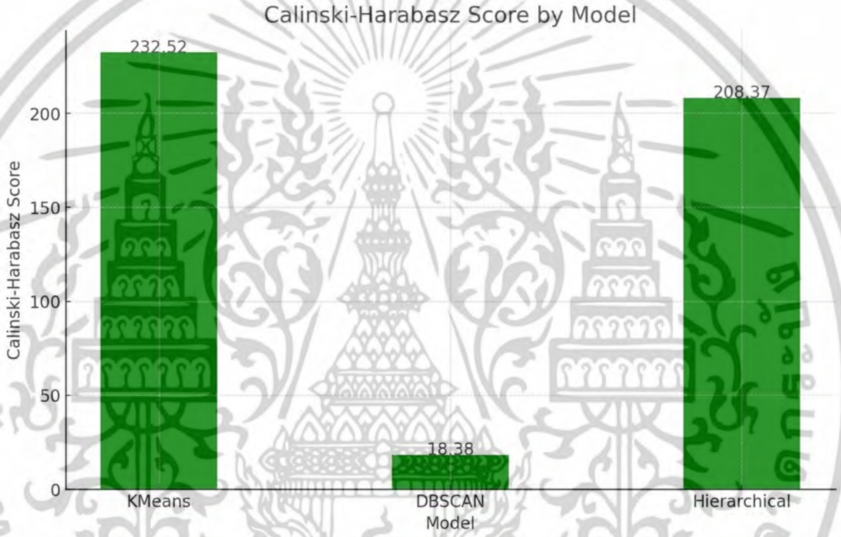
รูปที่ 4.11 กราฟแสดงผลการเปรียบเทียบประสิทธิภาพด้วยค่าซิลูเอต โคอิฟฟิเชียนท์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากกราฟในรูป 4.11 แสดงผลการเปรียบเทียบประสิทธิภาพของ 3 แบบจำลองด้วยค่าซิลูเอต โคอิฟฟิเชียนท์ ซึ่งจากกราฟพบว่าการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค ให้ค่าซิลูเอต โคอิฟฟิเชียนท์สูงสุดที่ 0.626 ซึ่งบ่งบอกถึงการจัดกลุ่มที่ดี การวิเคราะห์แบ่งกลุ่มแบบเป็นขั้นตอน มีค่าซิลูเอต โคอิฟฟิเชียนท์ 0.606 ใกล้เคียงกัน แสดงว่าประสิทธิภาพยังดีการแบ่งกลุ่มตามความหนาแน่น มีค่าซิลูเอต โคอิฟฟิเชียนท์ต่ำสุดที่ 0.244 ซึ่งอาจบ่งชี้ถึงการกระจายตัวที่ไม่ชัดเจน

4.2.3 ดัชนีคาลินสกี้-ฮาราบาซ

ดัชนีคาลินสกี้-ฮาราบาซเป็นตัวชี้วัดที่พิจารณาทั้งความกระจายตัวภายในกลุ่มและระยะห่างระหว่างกลุ่ม ค่าคาลินสกี้-ฮาราบาซที่สูงบ่งบอกถึงกลุ่มที่แยกตัวจากกันอย่างชัดเจน แสดงผลดังรูปที่ 4.12



รูปที่ 4.12 กราฟแสดงผลการเปรียบเทียบประสิทธิภาพด้วยค่าดัชนีคาลินสกี้-ฮาราบาซ

จากกราฟในรูป 4.12 แสดงผลการเปรียบเทียบประสิทธิภาพของ 3 แบบจำลองด้วยค่าคาลินสกี้-ฮาราบาซ ซึ่งจากกราฟพบว่าการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค ให้คะแนนสูงสุดที่ 232.522 ซึ่งแสดงถึงการแยกกลุ่มได้ดี การวิเคราะห์แบ่งกลุ่มแบบเป็นขั้นตอน ให้คะแนนรองลงมาที่ 208.374 ซึ่งยังแสดงถึงการแยกตัวของกลุ่มที่ดีเช่นกันแต่ไม่เท่ากับการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค ส่วนการแบ่งกลุ่มตามความหนาแน่น ให้คะแนนต่ำสุดที่ 18.384 อาจแสดงถึงการทับซ้อนระหว่างกลุ่ม

จากกราฟทั้ง 3 กราฟที่แสดงผลการเปรียบเทียบประสิทธิภาพของแบบจำลอง พบว่าวิธีการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเคเป็นแบบจำลองที่เหมาะสมที่สุดสำหรับชุดข้อมูลนี้ แสดงผลดังตารางที่ 4.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.1 ผลการเปรียบเทียบประสิทธิภาพของแบบจำลอง

	เดวีส์-โบลดิน	ซิลูเอต โคอิฟฟิเชียนท์	คาลินสกี้-ฮาราบาซ
การแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค	0.544	0.626	232.522
การวิเคราะห์แบ่งกลุ่มแบบเป็นขั้นตอน	0.385	0.606	208.374
การแบ่งกลุ่มข้อมูลตามความหนาแน่น	1.992	0.244	18.384

จากตารางที่ 4.1 แสดงผลการเปรียบเทียบประสิทธิภาพของแบบจำลองการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค การวิเคราะห์แบ่งกลุ่มแบบเป็นขั้นตอน และการแบ่งกลุ่มข้อมูลตามความหนาแน่น โดยการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเคมีประสิทธิภาพสูงที่สุด เนื่องจากมีค่าซิลูเอต โคอิฟฟิเชียนท์สูงที่สุด มีค่าอยู่ที่ 0.626 แสดงถึงแบบจำลองวิธีนี้มีการแบ่งกลุ่มที่ชัดเจน ดัชนีคาลินสกี้-ฮาราบาซสูงสุด มีค่าอยู่ที่ 232.52 แสดงถึงความหนาแน่นภายในกลุ่มที่ดีมีการแยกระหว่างกลุ่มที่ชัดเจน ส่วนการแบ่งกลุ่มแบบเป็นขั้นตอนให้ผลลัพธ์ใกล้เคียงกับการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเคโดยมีดัชนีเดวีส์-โบลดินต่ำที่สุด มีค่าอยู่ที่ 0.385 ซึ่งแสดงถึงความแตกต่างระหว่างกลุ่มที่ดี ในขณะที่การแบ่งกลุ่มข้อมูลตามความหนาแน่น มีประสิทธิภาพต่ำที่สุดอาจเนื่องจากไม่เหมาะกับลักษณะการกระจายตัวของข้อมูลชุดนี้ โดยสรุป KMeans เป็นโมเดลที่แนะนำให้ใช้สำหรับการจัดกลุ่มข้อมูลนี้ เนื่องจากให้ผลลัพธ์ที่ดีที่สุดโดยรวม ทั้งในแง่ของการแบ่งกลุ่มที่ชัดเจน ความหนาแน่นภายในกลุ่ม และการแยกระหว่างกลุ่มที่ดี ดังนั้นจึงสรุปได้ว่าการแบ่งกลุ่มแบบค่าเฉลี่ยเคมีความสม่ำเสมอของกลุ่มที่ดีที่สุด ขณะที่การแบ่งกลุ่มแบบเป็นขั้นตอนมีความสามารถในการแยกกลุ่มที่ชัดเจนที่สุด

4.3 การวิเคราะห์เชิงกลยุทธ์

จากการศึกษาพฤติกรรมการใช้เครื่องรับชำระเงินอิเล็กทรอนิกส์ของร้านค้า สามารถจำแนกกลุ่มผู้ใช้บริการออกเป็นสามกลุ่มหลัก ได้แก่ ร้านค้าที่ไม่ค่อยใช้บริการหรือไม่ได้ใช้มาเป็นเวลานาน, ร้านค้าที่ใช้บริการอย่างสม่ำเสมอและร้านค้าชั้นยอดที่มียอดการใช้บริการสูงสุด โดยการวิเคราะห์กลุ่มดังกล่าวมีความสำคัญต่อการกำหนดกลยุทธ์ทางธุรกิจของธนาคารและผู้ให้บริการระบบรับชำระเงิน เนื่องจากสามารถช่วยให้เข้าใจพฤติกรรมของร้านค้าและนำไปสู่การพัฒนานโยบายที่เหมาะสมเพื่อส่งเสริมการใช้บริการและรักษาความสัมพันธ์กับลูกค้า

4.3.1 กลุ่มร้านค้าที่ไม่ค่อยใช้บริการ หรือไม่ได้ใช้บริการมานาน

ร้านค้าในกลุ่มนี้เคยใช้เครื่องรับชำระเงินอิเล็กทรอนิกส์ของธนาคารแต่ไม่ได้ใช้งานมาเป็นระยะเวลานาน ซึ่งอาจมีสาเหตุมาจากปัจจัยหลายประการ เช่น ปัญหาด้านค่าธรรมเนียม ความขัดข้องของอุปกรณ์ หรือการเปลี่ยนไปใช้บริการของธนาคารคู่แข่ง ดังนั้นเพื่อกระตุ้นให้ร้านค้าเหล่านี้กลับมาใช้บริการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ควรมีมาตรการที่เหมาะสม เช่น การติดต่อสอบถามถึงสาเหตุที่เลิกใช้ พร้อมนำเสนอแนวทางแก้ไข การลดค่าธรรมเนียมสำหรับร้านค้าที่กลับมาใช้บริการ หรือการให้สิทธิพิเศษบางประการ เช่น การยกเว้นค่าธรรมเนียมในช่วงระยะเวลาหนึ่ง หรือการปรับปรุงบริการหลังการขายเพื่อเพิ่มความสะดวกในการใช้งาน

4.3.2 ร้านค้าที่ใช้บริการเป็นปกติ

ร้านค้าในกลุ่มนี้ยังคงใช้บริการเครื่องรับชำระเงินอิเล็กทรอนิกส์ของธนาคารอย่างสม่ำเสมอ อย่างไรก็ตามก็ควรมีการส่งเสริมให้ร้านค้าเหล่านี้เพิ่มปริมาณการใช้บริการให้สูงขึ้น โดยอาจใช้กลยุทธ์ทางการตลาด เช่น การให้ส่วนลดค่าธรรมเนียมหากมีปริมาณธุรกรรมเพิ่มขึ้น การจัดโปรแกรมสะสมแต้มสำหรับร้านค้า หรือการเสนอสิทธิพิเศษด้านสินเชื่อและบริการทางการเงินอื่นๆ เพื่อจูงใจให้ร้านค้ารักษาความสัมพันธ์กับธนาคารในระยะยาว

4.3.3 ร้านค้าชั้นยอดที่ใช้บริการสูงสุด

กลุ่มนี้ประกอบด้วยร้านค้าที่มีการใช้เครื่องรับชำระเงินอิเล็กทรอนิกส์ของธนาคารอย่างต่อเนื่องและมียอดการใช้บริการสูง ธนาคารควรให้ความสำคัญกับการรักษาลูกค้ากลุ่มนี้เป็นพิเศษ โดยอาจพิจารณามอบสิทธิประโยชน์เพิ่มเติม เช่น การลดค่าธรรมเนียมพิเศษ การให้เครดิตเงินคืน หรือการเชิญเข้าร่วมกิจกรรมพิเศษ เช่น สัมมนา หรือโปรแกรมลูกค้าระดับวีไอพี นอกจากนี้การเสนอผลิตภัณฑ์ทางการเงินที่เหมาะสม เช่น สินเชื่อธุรกิจ หรือบริการที่ปรึกษาทางการเงิน อาจช่วยให้ร้านค้ากลุ่มนี้มีความมั่นใจและเลือกใช้บริการของธนาคารอย่างต่อเนื่อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

จากการศึกษาการวิเคราะห์และแก้ไขปัญหาการใช้เครื่องรับชำระเงินอิเล็กทรอนิกส์อย่างไม่เกิดประโยชน์สูงสุด โดยวิธีการเรียนรู้ของเครื่อง เพื่อจะไปสู่ขั้นตอนของการสร้างแบบจำลอง โดยแบบจำลองเป็นแบบจำลองการเรียนรู้ของเครื่องแบบไม่มีผู้สอน 3 วิธี คือ การแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค การแบ่งกลุ่มข้อมูลตามความหนาแน่น และการวิเคราะห์แบ่งกลุ่มแบบเป็นขั้นตอน ขั้นตอนสุดท้ายทำการเปรียบเทียบประสิทธิภาพแบบจำลอง โดยผู้วิจัยกำหนดใช้ดัชนีเดวีส์-โบลดิน, ค่าซิลูเอต โคอิฟฟิเชียนท์ และดัชนีคาลินสกี้-ฮาราบาซในการเปรียบเทียบประสิทธิภาพ โดยศึกษาครั้งนี้ได้จัดทำขึ้นเพื่อวิเคราะห์ว่าร้านค้าไหนที่มีการใช้งานเครื่องรับชำระเงินอิเล็กทรอนิกส์ไม่เป็นไปตามเป้าหมายของธนาคาร จึงได้สรุปผลวิจัยและข้อเสนอแนะได้ดังนี้

5.1 สรุปผลการวิจัย

ผลการวิเคราะห์พบว่า การแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเคเป็นแบบจำลองที่เหมาะสมที่สุดสำหรับชุดข้อมูลนี้ โดยมีค่าซิลูเอต โคอิฟฟิเชียนท์สูงสุด 0.626 ซึ่งบ่งชี้ถึงการแบ่งกลุ่มที่ชัดเจน และมีดัชนีคาลินสกี้-ฮาราบาซสูงสุด 232.52 แสดงถึงความหนาแน่นภายในกลุ่มที่ดีและการแยกระหว่างกลุ่มที่ชัดเจน การวิเคราะห์แบ่งกลุ่มแบบเป็นขั้นตอนให้ผลลัพธ์ใกล้เคียงกับการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค โดยมีค่าดัชนีเดวีส์-โบลดินต่ำที่สุด 0.385 ซึ่งแสดงถึงความแตกต่างระหว่างกลุ่มที่ดี ในขณะที่การแบ่งกลุ่มข้อมูลตามความหนาแน่นมีประสิทธิภาพต่ำที่สุดในทุกเมตริก อาจเนื่องจากไม่เหมาะกับลักษณะการกระจายตัวของข้อมูลชุดนี้ โดยสรุปการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเคเป็นแบบจำลองที่แนะนำให้ใช้สำหรับการจัดกลุ่มข้อมูลนี้ เนื่องจากให้ผลลัพธ์ที่ดีที่สุดโดยรวมทั้งในแง่ของการแบ่งกลุ่มที่ชัดเจน ความหนาแน่นภายในกลุ่ม และการแยกระหว่างกลุ่มที่ดี

ผลของการจัดกลุ่มของทั้งสามแบบจำลองเป็นไปในทางทิศทางเดียวกัน เนื่องจากทั้งสามแบบจำลองให้ผลของการจัดกลุ่มใกล้เคียงกันทำให้สามารถนำไปปรับใช้ในการทำงานได้ทั้งสามแบบจำลอง ซึ่งจะทำให้ธนาคารสามารถนำผลวิจัยนี้ไปวางแผนการปรับการให้บริการการใช้เครื่องรับชำระเงินอิเล็กทรอนิกส์ได้อย่างมีประโยชน์สูงสุด โดยกลุ่มแรกคือกลุ่มที่ไม่ได้ใช้เครื่องรูดบัตรของธนาคารนานเกินกว่าที่ธนาคารกำหนด ทางธนาคารจะมีการเข้าไปตักเตือน หรือกระตุ้นร้านค้าเพื่อให้ร้านค้ากลับมาใช้เครื่องรูดของทางธนาคารอีกครั้ง กลุ่มต่อมาคือกลุ่มที่สองเป็นกลุ่มที่ใช้งานตามปกติ ไม่ได้มียอดหรือการใช้งานที่มีความถี่มากกว่าปกติ ในกลุ่มนี้ธนาคารยังไม่ต้องทำสิ่งใดกับร้านค้านี้มาก แต่อาจจะมีการเอาโปรโมชั่นไปนำเสนอ เพื่อให้ร้านค้ามีแนวโน้มใช้บริการมากขึ้น เช่น ลดค่าบริการเครื่องลง เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และกลุ่มสุดท้ายคือกลุ่มที่ร้านค้าที่ใช้บริการของธนาคารมีความถี่และมียอดมากเป็นพิเศษ ร้านค้ากลุ่มนี้ทางธนาคารก็จะดูแลและให้บริการเป็นพิเศษ หรืออาจเพิ่มสิทธิพิเศษเพิ่มเติมให้ เช่น การลดค่าธรรมเนียมหรือบริการค่าซ่อมเครื่องรูดบัตรของธนาคารให้โดยไม่คิดค่าใช้จ่ายเป็นต้น

5.2 ข้อเสนอแนะ

1. เพิ่มการวิเคราะห์พฤติกรรมของร้านค้าตามช่วงเวลา ศึกษาลักษณะการใช้เครื่องรับชำระเงินตามฤดูกาลหรือช่วงเวลาที่สำคัญ เช่น ช่วงเทศกาลหรือวันหยุด ซึ่งอาจช่วยให้สามารถออกแบบแคมเปญส่งเสริมการใช้บริการได้อย่างมีประสิทธิภาพ
2. พัฒนาระบบแนะนำแบบเรียลไทม์ หากสามารถวิเคราะห์ข้อมูลในแบบเรียลไทม์ได้ อาจสร้างประโยชน์อย่างมากในการให้คำแนะนำหรือโปรโมชั่นที่ตรงกับพฤติกรรมการใช้บริการของร้านค้า เช่น หากระบบตรวจพบว่าร้านค้าใดกำลังมีความถี่ในการใช้งานลดลง ระบบสามารถแนะนำโปรโมชั่นพิเศษทันที เพื่อกระตุ้นการใช้บริการอย่างต่อเนื่อง
3. ใช้ข้อมูลที่มาจากหลายธนาคาร เนื่องจากการวิจัยครั้งนี้ใช้ข้อมูลจากธนาคารแห่งหนึ่งเพียงชุดเดียว จึงอาจไม่สามารถสรุปแนวโน้มพฤติกรรมการใช้เครื่องรับชำระเงินอิเล็กทรอนิกส์ของร้านค้าทั้งหมดได้ ในอนาคต การใช้ข้อมูลจากหลายธนาคารหรือผู้ให้บริการที่หลากหลายจะช่วยให้การวิเคราะห์มีความแม่นยำและเชื่อถือได้มากยิ่งขึ้น นอกจากนี้ ควรพิจารณาเพิ่มตัวแปรอื่น ๆ เช่น ประเภทสินค้าและบริการของร้านค้า หรือช่วงเวลาการใช้งาน เพื่อให้ได้ข้อมูลที่ครบถ้วนและสะท้อนถึงพฤติกรรมที่แท้จริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

วิทวัส แสงสว่าง. 2565. “แบบจำลองทำนายราคาขายเฉลี่ยต่อพื้นที่ของโครงการที่อยู่อาศัยด้วยเทคนิคการวิเคราะห์ถดถอยร่วมกับการจัดกลุ่มตามความหนาแน่นพื้นที่กรุงเทพมหานคร” วิทยานิพนธ์มหาบัณฑิต สาขาวิชาวิศวกรรมข้อมูลขนาดใหญ่ คณะวิศวกรรมศาสตร์, มหาวิทยาลัยธุรกิจบัณฑิตย์.

อุไรวรรณ์ อังคะเวย์. 2560. “การค้าหาจุดสนใจโดยใช้ข้อมูลเชิงพื้นที่จากรถแท็กซี่ด้วยอัลกอริทึมการจัดกลุ่มตามความหนาแน่นที่กำหนดพารามิเตอร์บบอัตโนมัติ” วิทยานิพนธ์มหาบัณฑิต สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ คณะวิศวกรรมศาสตร์, มหาวิทยาลัยจุฬาลงกรณ์.

อัญญ์กาญจน์ พิชยยุทธ. 2556. “การแบ่งกลุ่มลูกค้ารายย่อยโดยใช้ปัจจัยด้านพฤติกรรมการซื้อน้ำแข็งซอง” วิทยานิพนธ์มหาบัณฑิต สาขาการจัดการ คณะบริหารธุรกิจ, มหาวิทยาลัยหอการค้าไทย.

Ang, G., Tian, R., & Wu, B. "An Overview of Clustering Methods in The Financial World," Proceedings of the 2022 7th International Conference on Financial Innovation and Economic Development (ICFIED 2022), Advances in Economics, Business and Management Research, vol. 211, 2022, pp. 524-528.

Abdulhafedh, “Incorporating K-means, Hierarchical Clustering and PCA in Customer Segmentation” Journal of City and Development, Vol. 3, No. 1 (2021): 12–30.

Diana L., Francisco A., Soumaya Y. & Ada A. (2017). A multi-start algorithm to design a multi-class classifier for a multi-criteria ABC inventory classification problem. Expert systems with application, 81(1), 12-21.

Do, J. H., & Choi, D.-K. "Clustering Approaches to Identifying Gene Expression Patterns from DNA Microarray Data," Molecules and Cells, vol. 25, no. 2, 2008, pp. 279-288.

Hulett, C., Hall, A., & Qu G. (2012) Dynamic Selection of k Nearest Neighbors in Instance-Based Learning. Information Reuse and Integration.: 85-92.

K. Sasirekha & P. Baby, "Agglomerative Hierarchical Clustering Algorithm - A Review," International Journal of Scientific and Research Publications, 3, no. 3 (2013): 1-3,

Nourahmadi, Rasti, & Sadeqi, "A Comparative Approach to Financial Clustering Models: A Study of the Companies Listed on Tehran Stock Exchange," Iranian Journal of Finance, 6, no. 4 (2022): 31-55,

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Jaroonchokanan, Termsaithong, & Suwanna, “Dynamics of Hierarchical Clustering in Stocks Market During Financial Crises,” Physica A, 2022.

Wulandari, V., Syarif, Y., Alfian, Z., Althof, M. A., & Mufidah, M. (2024). Comparison of Density-Based Spatial Clustering of Applications with Noise (DBSCAN), K-Means and X-Means Algorithms on Shopping Trends Data. IJATIS: Indonesian Journal of Applied Technology and Innovation Science, 1(1), 1-8. doi: 10.57152/IJATIS.v1i1.1135

Zhangyao Zhu & Na Liu, "Early Warning of Financial Risk Based on K-Means Clustering Algorithm," Complexity, 2021

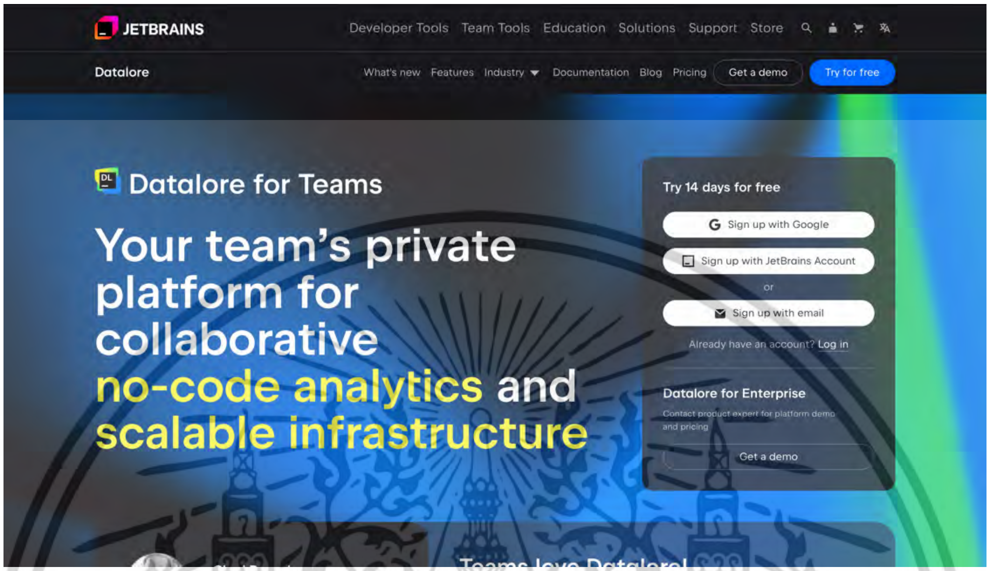
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก

ขั้นตอนการเตรียมข้อมูล



รูปที่ ก.1 โปรแกรมที่ใช้ในการวิเคราะห์ข้อมูลของแบบจำลอง

Merchant ID	Merchant Name	Terminal ID	Trx.Date	Trx.Time	Trace No.	Card No.	Card Type	Card Brand	Input By	Trx Type	Currency	Trx Amount	Fee	VAT	Credit Merchant	CC Card Type
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-10-05	10:54:23	37653400000100000000000	de94b4c511165c15aebb40dbc3fe1a4e	Credit Card	Local VISA	Chip/Circuit	Purchase	THB	1,190.00	23.80	1.67	1,164.53	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-07-27	14:12:03	425524000003	32e518a230a0ba9842aa12f3f890650d	Debit Card	Local VISA	Chip/Circuit	Purchase	THB	540.00	10.80	0.76	528.44	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-07-23	17:19:39	47737800000200000000000	f1cc93f1bf19daeeda44ce6b5c2644e9	Credit Card	Local VISA	Chip/Circuit	Purchase	THB	1,010.00	20.20	1.41	988.39	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-10-07	11:42:06	35959800000100000000000	3ccd79eb7b3abf23bbe53852910a889f	Credit Card	Local VISA	Chip/Circuit	Purchase	THB	798.00	15.96	1.12	780.92	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-08-24	18:53:35	52114000000300000000000	a757a67556bc2dc82907fd9d5ebf79ca	Credit Card	Local VISA	Chip/Circuit	Purchase	THB	864.00	17.28	1.21	845.51	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-10-14	13:52:10	429208000001	6b84ec32eff49dd7debc31f9daa5f149	Debit Card	Local VISA	Chip/Circuit	Purchase	THB	1,808.00	36.16	2.53	1,769.31	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-09-18	09:07:07	32281800000100000000000	3dcff2b79611b35778a917207c019474	Credit Card	Local VISA	Chip/Circuit	Purchase	THB	1,196.00	23.92	1.67	1,170.41	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-07-27	13:07:00	400014000002	7aefb3ba6c7de0c06d6a0428edc958a0	Debit Card	Local VISA	Chip/Circuit	Purchase	THB	637.00	12.74	0.89	623.37	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-09-12	10:09:43	34285800000100000000000	a46e9a1cf5dc5dd9f1f9c05a2b050437	Credit Card	Local VISA	Chip/Circuit	Purchase	THB	801.00	16.02	1.12	783.86	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-08-01	11:32:20	39271000000100000000000	539734404c61d4928da038fe0297921f	Credit Card	Local VISA	Chip/Circuit	Purchase	THB	1,084.00	21.68	1.52	1,060.80	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-08-24	12:53:18	387428000001	0d83c325fc2d3f1899db84ce843d6b49	Debit Card	Local VISA	Chip/Circuit	Purchase	THB	550.00	11.00	0.77	538.23	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-08-14	18:06:01	50873100000200000000000	afe33f947100c65854008a0c7c97ae7b	Credit Card	Local VISA	Chip/Circuit	Purchase	THB	2,580.00	51.60	3.61	2,524.79	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-08-21	14:50:44	438078000001	e0807764156ba957eecd4d5faa67e6be	Debit Card	Local VISA	Chip/Circuit	Purchase	THB	845.00	16.90	1.18	826.92	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-07-25	12:52:05	385782000001	599b89ebffea65c7365ce74a52e0b854	Debit Card	Local VISA	Chip/Circuit	Purchase	THB	513.00	10.26	0.72	502.02	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-07-17	12:17:54	390385000001	762c085055e35a43e65a1d577cc48ca4	Debit Card	Local VISA	Chip/Circuit	Purchase	THB	1.00	0.02	-	0.98	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-10-05	10:57:50	377876000002	b53e31a75b85766fe71a302c043db289	Debit Card	Local ITMX	Chip/Circuit	Purchase	THB	501.00	10.02	0.70	490.28	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-08-20	09:55:22	332268000001	ce442675dc44d9e0dae27938d63c59d4	Debit Card	Local VISA	Chip/Circuit	Purchase	THB	1,221.00	24.42	1.71	1,194.87	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-10-12	11:23:46	36401200000100000000000	f96d4864bbf8a1097421f493e307b2d7	Credit Card	Local VISA	Chip/Circuit	Purchase	THB	1,038.00	20.76	1.45	1,015.79	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-09-11	18:05:16	521816000002	9fae928dc8348a0d9120ac21b907b905	Debit Card	Local VISA	Chip/Circuit	Purchase	THB	509.00	10.18	0.71	498.11	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-09-16	12:34:22	37725800000100000000000	24ab10f05f271bc7c44ab7de8def2ba9	Credit Card	Local VISA	Chip/Circuit	Purchase	THB	1,540.00	30.80	2.16	1,507.04	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-12-11	15:04:15	46667400000100000000000	b34916e55c0c491b0f94aece9dfe4154	Credit Card	Local VISA	Chip/Circuit	Purchase	THB	700.00	14.00	0.98	685.02	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-08-26	08:56:58	307408000001	603bf5c9e518163064f67501962ec2ef	Debit Card	Local VISA	Chip/Circuit	Purchase	THB	116.00	2.32	0.16	113.52	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-09-12	12:31:54	38787800000200000000000	c09269e4433f4350ba421ea19b5a7c17	Credit Card	Local VISA	Chip/Circuit	Purchase	THB	1,503.13	30.06	2.10	1,470.97	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-08-15	10:27:52	33743600000100000000000	d9b5310056b9f2d227de07c878006205	Credit Card	Local VISA	Chip/Circuit	Purchase	THB	517.00	10.34	0.72	505.94	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-08-19	14:48:08	43740600000200000000000	215c5012a34e9aa7a4f2f9fa174ae52f	Credit Card	Local VISA	Chip/Circuit	Purchase	THB	734.00	14.68	1.03	718.29	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-10-03	12:24:13	42758800000100000000000	1b0ee8cb16da285f5b5e78ad8b5a4a6c	Credit Card	Local VISA	Chip/Circuit	Purchase	THB	978.00	19.56	1.37	957.07	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-09-08	10:47:14	36522400000100000000000	c3b4a31726cc8a0415051e3c476e8181	Credit Card	Local VISA	Chip/Circuit	Purchase	THB	1,143.00	22.86	1.60	1,118.54	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-08-18	12:41:19	38155200000100000000000	bd708c663397de79c19e2d7ed9185288	Credit Card	Local VISA	Chip/Circuit	Purchase	THB	2,349.50	46.99	3.29	2,299.22	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-08-14	10:15:23	33996500000100000000000	741315eb58aa6a9ad3b6bfa7bbd5b33d	Credit Card	Local VISA	Chip/Circuit	Purchase	THB	504.00	10.08	0.71	493.21	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-08-10	16:05:06	485516000001	3fb413708bdc057699d30123d41bcd08	Debit Card	Local VISA	Chip/Circuit	Purchase	THB	815.50	16.31	1.14	798.05	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-08-15	10:35:02	339682000002	c45d5c89109166b2cf35b42bdd04d2d6	Debit Card	Local VISA	Chip/Circuit	Purchase	THB	760.00	15.20	1.06	743.74	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-09-14	14:55:44	44186600000100000000000	9ab190a12807a8814036ef931db81a96	Credit Card	Local VISA	Chip/Circuit	Purchase	THB	886.00	17.72	1.24	867.04	
016500070001	1010c1d0c3a8e6ead13b2569c4229ff5	016500070001101	2019-07-13	13:35:59	424722000001	d69913cbdee49956d940c5a0627a81fa	Debit Card	Local VISA	Chip/Circuit	Purchase	THB	1,046.00	20.92	1.46	1,023.62	

รูปที่ ก.2 ข้อมูลจริงจากธนาคารแห่งหนึ่ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
import pandas as pd
import numpy as np
```

```
data_path = "/data/notebook_files/Real_data_2019.csv"
raw = pd.read_csv(data_path)
raw = raw[['Merchant ID', 'Trx.Date', 'Trx.Time', 'Trx Amount', 'Fee', 'VAT', 'Credit Merchant']]
raw = raw[raw['Merchant ID'] != 16500740001] # remove test ID
raw['datetime'] = pd.to_datetime(raw['Trx.Date'] + ' ' + raw['Trx.Time'])
raw = raw.drop(['Trx.Date', 'Trx.Time'], axis=1)
raw.columns = raw.columns.str.lower().str.replace(' ', '_')
raw['fee'] = raw['fee'].str.strip().str.replace(',', '')
raw['fee'] = raw['fee'].str.strip().str.replace('-', '')
raw['credit_merchant'] = raw['credit_merchant'].str.strip().str.replace(',', '')
raw['credit_merchant'] = raw['credit_merchant'].str.strip().str.replace('-', '')
raw['trx_amount'] = raw['trx_amount'].str.strip().str.replace(',', '')
raw['trx_amount'] = raw['trx_amount'].str.strip().str.replace('-', '')
raw['fee'] = pd.to_numeric(raw['fee'])
raw['trx_amount'] = pd.to_numeric(raw['trx_amount'])
raw['credit_merchant'] = pd.to_numeric(raw['credit_merchant'])
raw['percent_fee'] = raw['fee'] / raw['trx_amount'] * 100
raw['percent_fee'] = round(raw['percent_fee'], 1)
raw.head(10)
```

รูปที่ ก.3 การตัดข้อมูลออกและเลือกแค่ข้อมูลที่จะนำไปใช้

	merchant_id	trx_amount	fee	vat	credit_merchant	datetime	percent_fee
0	16500070001	1190.0	23.80	1.67	1164.53	2019-10-05 10:54:23	2.0
1	16500070001	540.0	10.80	0.76	528.44	2019-07-27 14:12:03	2.0
2	16500070001	1010.0	20.20	1.41	988.39	2019-07-23 17:19:39	2.0
3	16500070001	798.0	15.96	1.12	780.92	2019-10-07 11:42:06	2.0
4	16500070001	864.0	17.28	1.21	845.51	2019-08-24 18:53:35	2.0
5	16500070001	1808.0	36.16	2.53	1769.31	2019-10-14 13:52:10	2.0
6	16500070001	1196.0	23.92	1.67	1170.41	2019-09-18 09:07:07	2.0
7	16500070001	637.0	12.74	0.89	623.37	2019-07-27 13:07:00	2.0
8	16500070001	801.0	16.02	1.12	783.86	2019-09-12 10:09:43	2.0
9	16500070001	1084.0	21.68	1.52	1060.80	2019-08-01 11:32:20	2.0

รูปที่ ก.4 ตัวอย่างข้อมูลหลังจากเลือกแค่ข้อมูลที่จะนำไปใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```python
raw['datetime'] = pd.to_datetime(raw['datetime'])
latest_date = pd.to_datetime('2020-01-01')

# Calculate R, F, M for each merchant
rfm = raw.groupby('merchant_id').agg({
    'datetime': lambda x: (latest_date - x.max()).days,
    'credit_merchant': 'count',
    'trx_amount': 'sum'
}).rename(columns={'datetime': 'Recency', 'credit_merchant': 'Frequency', 'trx_amount': 'Monetary'})

# Calculate min, max, sum, avg, median of Monetary for each merchant
monetary_stats = raw.groupby('merchant_id')['trx_amount'].agg(['min', 'max', 'mean', 'median'])
result = rfm.join(monetary_stats)
result
```

	Recency	Frequency	Monetary	min	max	mean	median
merchant_id							
16500070001	20	82	9.568888e+04	1.0	6500.75	1166.937561	805.5
16500070002	27	14	3.288000e+04	505.0	5457.00	2348.571429	1747.5
16500070004	145	10	2.622100e+04	1020.0	4990.00	2622.100000	1870.5
16500170002	61	22	3.661990e+05	1.0	59400.00	16645.409091	12650.0
16500200001	4	389	8.336265e+06	235.0	241335.00	21429.987147	10220.0
...	...	...	...	...	...	...	...
16502140043	26	106	1.367580e+05	3.0	4400.00	1290.169811	1000.0
16502140044	5	163	1.988551e+06	506.0	230000.00	12199.699387	4611.0
16502150001	0	2905	1.566373e+08	180.0	433775.00	53919.903614	28000.0
16502150002	0	10589	2.406373e+08	50.0	350000.00	22725.216271	7500.0
16502150007	1	54	1.201369e+07	4500.0	2100000.00	222475.822222	150000.0

214 rows × 7 columns

รูปที่ ก.5 การวิเคราะห์อาร์เอฟเอ็ม

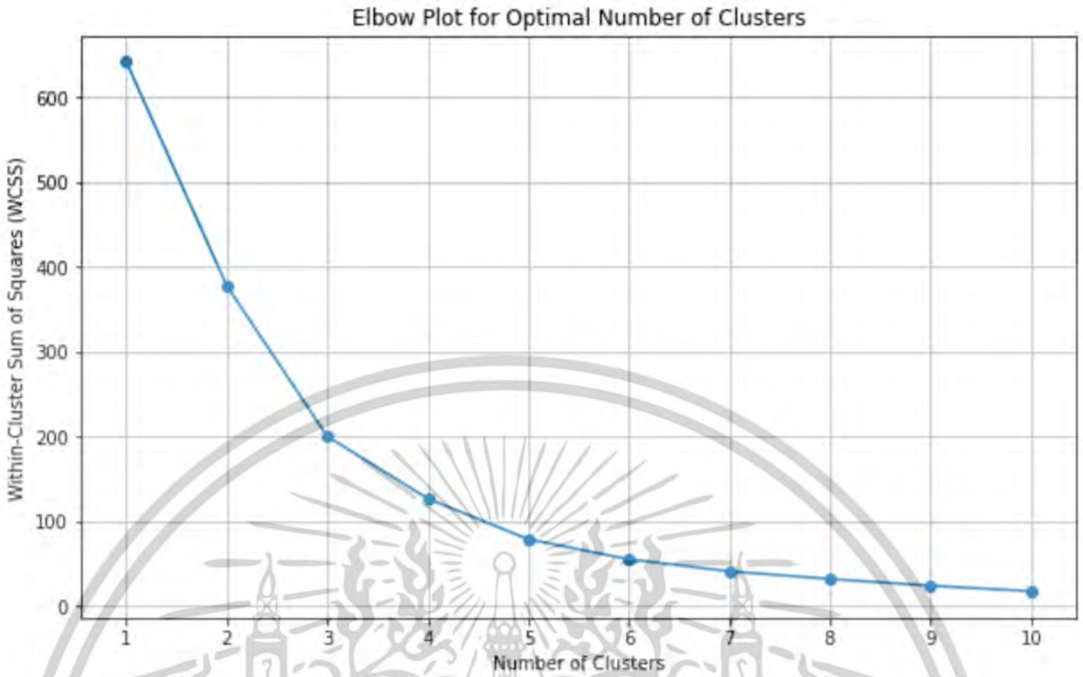
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
from sklearn.cluster import KMeans, DBSCAN, AgglomerativeClustering
from sklearn.metrics import silhouette_samples, silhouette_score
import matplotlib.pyplot as plt
import matplotlib.cm as cm
```

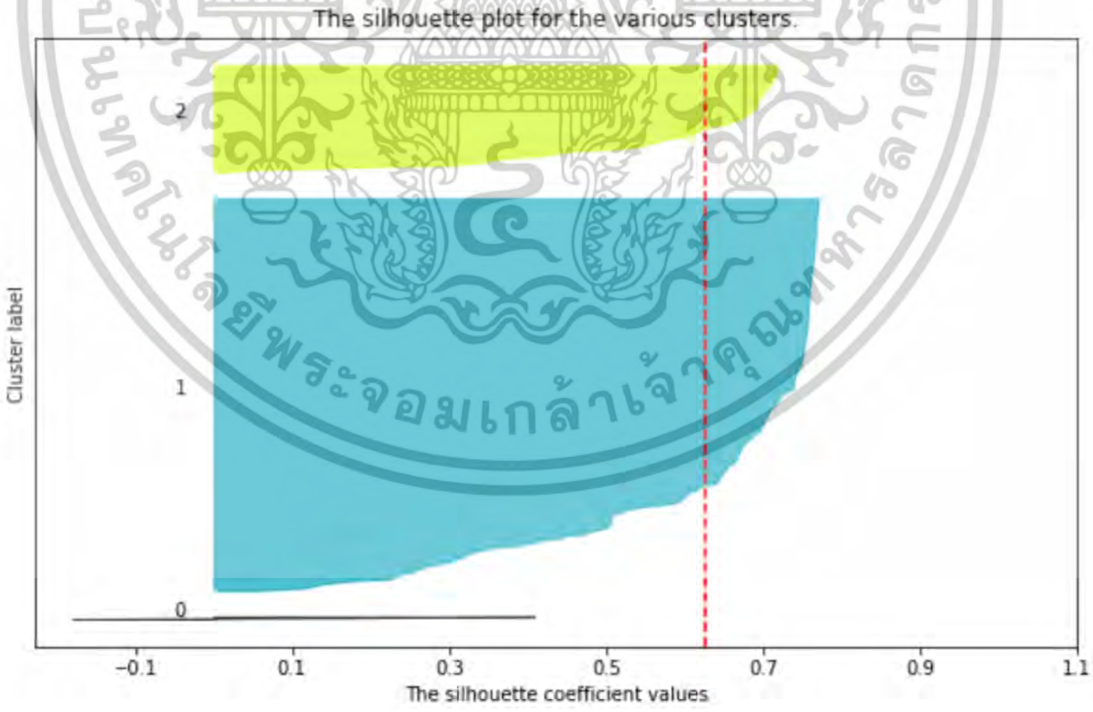
```
# Use only RFM for clustering
rfm_features = rfm[['Recency', 'Frequency', 'Monetary']]

# Standardize the data
scaler = StandardScaler()
scaled_data = scaler.fit_transform(rfm_features)

# Determine the optimal number of clusters using the elbow method
wcss = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters=i, random_state=42, n_init=10)
    kmeans.fit(scaled_data)
    wcss.append(kmeans.inertia_)

# Plot the elbow plot
plt.figure(figsize=(10, 6))
plt.plot(range(1, 11), wcss, marker='o')
plt.title('Elbow Plot for Optimal Number of Clusters')
plt.xlabel('Number of Clusters')
plt.ylabel('Within-Cluster Sum of Squares (WCSS)')
plt.xticks(range(1, 11))
plt.grid(True)
plt.show()

# Apply KMeans clustering
optimal_clusters = 3  # Replace this with the optimal number of clusters from the elbow plot
kmeans = KMeans(n_clusters=optimal_clusters, random_state=42, n_init=10)
kmeans_labels = kmeans.fit_predict(scaled_data)
rfm['KMeans_Labels'] = kmeans_labels

# Silhouette plot
fig, ax = plt.subplots(1, 1, figsize=(10, 6))

silhouette_avg = silhouette_score(scaled_data, kmeans_labels)
print("For n_clusters =", optimal_clusters, "The average silhouette_score is :", silhouette_avg)

sample_silhouette_values = silhouette_samples(scaled_data, kmeans_labels)
y_lower = 10
for i in range(optimal_clusters):
    ith_cluster_silhouette_values = sample_silhouette_values[kmeans_labels == i]
    ith_cluster_silhouette_values.sort()
    size_cluster_i = ith_cluster_silhouette_values.shape[0]
    y_upper = y_lower + size_cluster_i
    color = cm.nipy_spectral(float(i) / optimal_clusters)
    ax.fill_betweenx(np.arange(y_lower, y_upper), 0, ith_cluster_silhouette_values, facecolor=color, edgecolor=color, alpha=0.7)
    ax.text(-0.05, y_lower + 0.5 * size_cluster_i, str(i))
    y_lower = y_upper + 10

ax.set_title("The silhouette plot for the various clusters.")
ax.set_xlabel("The silhouette coefficient values")
ax.set_ylabel("Cluster label")
ax.axvline(x=silhouette_avg, color="red", linestyle="--")
ax.set_yticks([])
ax.set_xticks(np.arange(-0.1, 1.1, 0.2))
plt.show()
```

รูปที่ ก.6 การหาจำนวนคลัสเตอร์ที่เหมาะสมด้วยวิธีเอลโบว์และการวิเคราะห์ซิลูเอต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ ก.7 แผนภาพจำนวนคลัสเตอร์ที่เหมาะสมด้วยวิธีเอลโบว์



รูปที่ ก.8 แผนภาพจำนวนคลัสเตอร์ที่เหมาะสมด้วยการวิเคราะห์ซิลูเอต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

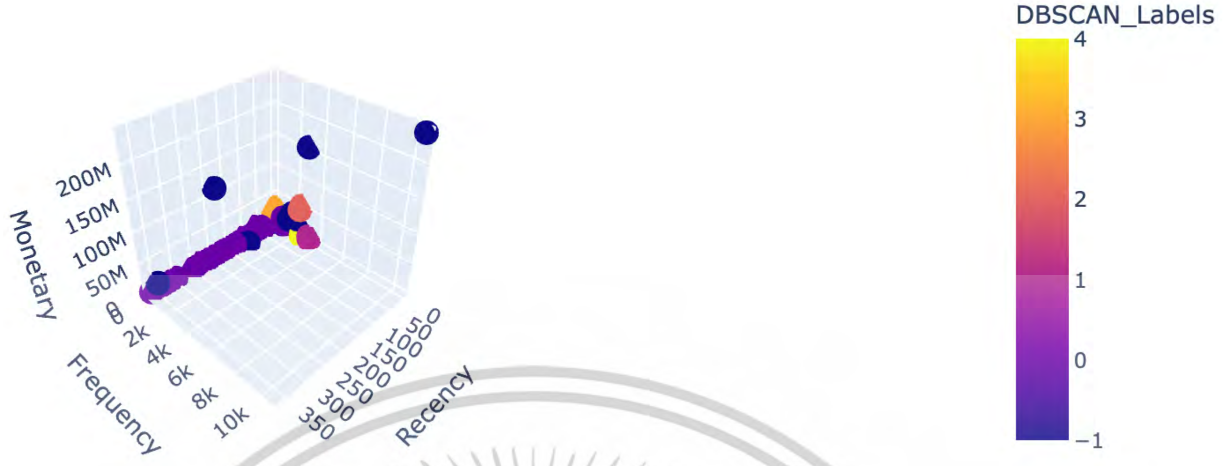
ภาคผนวก ข

```
# Reset index to include merchant_id as a column
rfm.reset_index(inplace=True)

# Use only RFM for clustering
rfm_features = rfm[['Recency', 'Frequency', 'Monetary']]

# Standardize the data
scaler = StandardScaler()
scaled_data = scaler.fit_transform(rfm_features)

# Apply KMeans clustering
kmeans = KMeans(n_clusters=3, random_state=42, n_init=10)
rfm['KMeans_Labels'] = kmeans.fit_predict(scaled_data)

# Apply DBSCAN clustering
dbscan = DBSCAN(eps=0.5, min_samples=2)
rfm['DBSCAN_Labels'] = dbscan.fit_predict(scaled_data)

# Apply Hierarchical clustering
hierarchical = AgglomerativeClustering(n_clusters=3)
rfm['Hierarchical_Labels'] = hierarchical.fit_predict(scaled_data)

# Plotting the results using Plotly
fig_kmeans = px.scatter_3d(rfm, x='Recency', y='Frequency', z='Monetary', color='KMeans_Labels', hover_data=['merchant_id'], title='KMeans Clustering')
fig_dbscan = px.scatter_3d(rfm, x='Recency', y='Frequency', z='Monetary', color='DBSCAN_Labels', hover_data=['merchant_id'], title='DBSCAN Clustering')
fig_hierarchical = px.scatter_3d(rfm, x='Recency', y='Frequency', z='Monetary', color='Hierarchical_Labels', hover_data=['merchant_id'], title='Hierarchica

fig_kmeans.show()
fig_dbscan.show()
fig_hierarchical.show()
```

รูปที่ ข.1 การสร้างแบบจำลองการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค, การวิเคราะห์แบ่งกลุ่มแบบเป็นขั้นตอน และ การแบ่งกลุ่มข้อมูลตามความหนาแน่น

รูปที่ ข.2 แผนภาพแสดงผลแบบจำลองการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ ข.3 แผนภาพแสดงผลแบบจำลองการวิเคราะห์แบ่งกลุ่มแบบเป็นขั้นตอน

รูปที่ ข.4 แผนภาพแสดงผลแบบจำลองการแบ่งกลุ่มข้อมูลตามความหนาแน่น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
rfm = rfm[rfm['DBSCAN_Labels'] != -1]

rfm
```

	merchant_id	Recency	Frequency	Monetary	KMeans_Labels	DBSCAN_Labels	Hierarchical_Labels
0	16500070001	20	82	95688.88	0	0	0
1	16500070002	27	14	32880.00	0	0	0
2	16500070004	145	10	26221.00	2	0	0
3	16500170002	61	22	366199.00	0	0	0
4	16500200001	4	389	8336265.00	0	0	0
...	...	...	...	...	...	...	...
206	16502140038	7	10	10954.00	0	0	0
208	16502140042	61	4	881840.00	0	0	0
209	16502140043	26	106	136758.00	0	0	0
210	16502140044	5	163	1988551.00	0	0	0
213	16502150007	1	54	12013694.40	0	0	0

206 rows × 7 columns

รูปที่ ข.5 การตัดกลุ่มข้อมูลที่มีมูลค่าสูง

```
# If necessary, apply KMeans clustering on 'result'
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler

# Standardizing the data
scaler = StandardScaler()
scaled_data_rfm = scaler.fit_transform(rfm[['Recency', 'Frequency', 'Monetary']])

# Plotting the results using Plotly
fig_kmeans = px.scatter_3d(rfm, x='Recency', y='Frequency', z='Monetary', color='KMeans_Labels', title='KMeans Clustering')
fig_dbscan = px.scatter_3d(rfm, x='Recency', y='Frequency', z='Monetary', color='DBSCAN_Labels', title='DBSCAN Clustering')
fig_hierarchical = px.scatter_3d(rfm, x='Recency', y='Frequency', z='Monetary', color='Hierarchical_Labels', title='Hierarchical Clustering')

# Display the plots
fig_kmeans.show()
fig_dbscan.show()
fig_hierarchical.show()
```

รูปที่ ข.6 การสร้างแบบจำลองการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค, การวิเคราะห์แบ่งกลุ่มแบบเป็นขั้นตอน และการแบ่งกลุ่มข้อมูลตามความหนาแน่น หลังตัดกลุ่มข้อมูลที่มีมูลค่าสูงออก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ ข.7 แผนภาพแบบจำลองแสดงผลการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค หลังตัดข้อมูลที่มีมูลค่าสูงออก

รูปที่ ข.8 แผนภาพแบบจำลองการวิเคราะห์แบ่งกลุ่มแบบเป็นขั้นตอน หลังตัดข้อมูลที่มีมูลค่าสูงออก

รูปที่ ข.9 แผนภาพแบบจำลองการแบ่งกลุ่มข้อมูลตามความหนาแน่น หลังตัดข้อมูลที่มีมูลค่าสูงออก

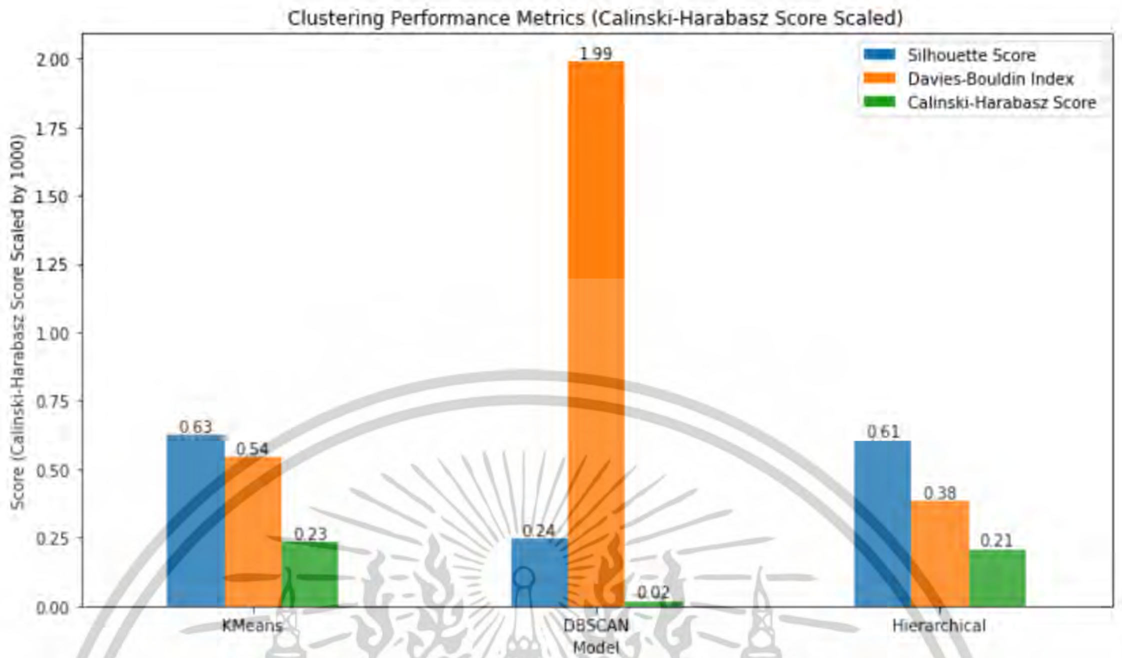
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
def calculate_scores(data, labels):
    silhouette = silhouette_score(data, labels)
    davies_bouldin = davies_bouldin_score(data, labels)
    calinski_harabasz = calinski_harabasz_score(data, labels)
    return silhouette, davies_bouldin, calinski_harabasz

kmeans_scores = calculate_scores(scaled_data, rfm['KMeans_Labels'])
dbscan_scores = calculate_scores(scaled_data, rfm['DBSCAN_Labels'])
hierarchical_scores = calculate_scores(scaled_data, rfm['Hierarchical_Labels'])

scores_df = pd.DataFrame({
    'Model': ['KMeans', 'DBSCAN', 'Hierarchical'],
    'Silhouette Score': [kmeans_scores[0], dbscan_scores[0], hierarchical_scores[0]],
    'Davies-Bouldin Index': [kmeans_scores[1], dbscan_scores[1], hierarchical_scores[1]],
    'Calinski-Harabasz Score': [kmeans_scores[2], dbscan_scores[2], hierarchical_scores[2]]
})

scores_df.set_index('Model').plot(kind='bar', figsize=(10, 6))
plt.title('Clustering Performance Metrics')
plt.ylabel('Score')
plt.xticks(rotation=0)
plt.tight_layout()
plt.show()

display(scores_df)
```

```
scores_df_clean = scores_df.replace('N/A', None).dropna()

scores_df_scaled = scores_df_clean.copy()

scores_df_scaled['Calinski-Harabasz Score'] = scores_df_scaled['Calinski-Harabasz Score'] / 1000

scores_df_scaled.set_index('Model').plot(kind='bar', figsize=(10, 6))
ax = scores_df_scaled.set_index('Model').plot(kind='bar', figsize=(10, 6))

for container in ax.containers:
    ax.bar_label(container, fmt='%.2f')

plt.title('Clustering Performance Metrics (Calinski-Harabasz Score Scaled)')
plt.ylabel('Score (Calinski-Harabasz Score Scaled by 1000)')
plt.xticks(rotation=0)
plt.tight_layout()
plt.show()
```

รูปที่ ข.10 การเปรียบเทียบแบบจำลอง ดัชนีเดวีส์-โบลดิน, ค่าซิลูเอต โคอิฟฟิเชียนท์ และดัชนีคาลินสกี้-ฮาราบาซ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ ข.11 แผนภาพแสดงผลการเปรียบเทียบประสิทธิภาพของแบบจำลอง

```
# Calculate average values and count of Recency, Frequency, Monetary for KMeans_Labels clusters 0 and 1
pivot_rfm_kmeans = rfm.pivot_table(values=['Recency', 'Frequency', 'Monetary'],
                                   index=['KMeans_Labels'],
                                   aggfunc={'Recency': [np.mean, np.size],
                                            'Frequency': [np.mean, np.size],
                                            'Monetary':[np.mean, np.size]})

# Display the pivot table
pivot_rfm_kmeans.round(0).reset_index()
```

	KMeans_Labels	Frequency		Monetary		Recency	
		mean	size	mean	size	mean	size
0	0	320.0	166	2918943.0	166	26.0	166
1	1	6747.0	2	198637318.0	2	0.0	2
2	2	17.0	46	2960119.0	46	261.0	46

รูปที่ ข.12 การวิเคราะห์อาร์เอ็มเอฟของแบบจำลองการแบ่งกลุ่มข้อมูลแบบค่าเฉลี่ยเค

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
# Calculate average values and count of Recency, Frequency, Monetary for KMeans_Labels clusters 0 and 1
pivot_rfm_kmeans = rfm.pivot_table(values=['Recency', 'Frequency', 'Monetary'],
                                   index=['DBSCAN_Labels'],
                                   aggfunc={'Recency': [np.mean, np.size],
                                            'Frequency': [np.mean, np.size],
                                            'Monetary':[np.mean, np.size]})

# Display the pivot table
pivot_rfm_kmeans.round(0).reset_index()
```

	DBSCAN_Labels	Frequency		Monetary		Recency	
		mean	size	mean	size	mean	size
0	-1	2422.0	8	74801555.0	8	80.0	8
1	0	158.0	198	1250489.0	198	79.0	198
2	1	3235.0	2	4592745.0	2	3.0	2
3	2	2374.0	2	49561033.0	2	0.0	2
4	3	62.0	2	26730280.0	2	8.0	2
5	4	2636.0	2	5103650.0	2	4.0	2

รูปที่ ข.13 การวิเคราะห์อาร์เอ็มเอฟของแบบจำลองการแบ่งกลุ่มข้อมูลตามความหนาแน่น

```
# Calculate average values and count of Recency, Frequency, Monetary for KMeans_Labels clusters 0 and 1
pivot_rfm_kmeans = rfm.pivot_table(values=['Recency', 'Frequency', 'Monetary'],
                                   index=['Hierarchical_Labels'],
                                   aggfunc={'Recency': [np.mean, np.size],
                                            'Frequency': [np.mean, np.size],
                                            'Monetary':[np.mean, np.size]})

# Display the pivot table
pivot_rfm_kmeans.round(0).reset_index()
```

	Hierarchical_Labels	Frequency		Monetary		Recency	
		mean	size	mean	size	mean	size
0	0	330.0	170	4453844.0	170	28.0	170
1	1	10589.0	1	240637315.0	1	0.0	1
2	2	15.0	43	469624.0	43	267.0	43

รูปที่ ข.14 การวิเคราะห์อาร์เอ็มเอฟของแบบจำลองการแบ่งกลุ่มแบบเป็นขั้นตอน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ	นายศุภวิชญ์ รุจิเมธาภาส
วัน เดือน ปีเกิด	01 ตุลาคม 2541
ที่อยู่ปัจจุบัน	99/104 หมู่บ้านอนันดา ถนนกิ่งแก้ว ซอยกิ่งแก้ว19 ตำบลบางพลีใหญ่ อำเภอบางพลี จังหวัดสมุทรปราการ 10540
ประวัติการศึกษา	(2564) เศรษฐศาสตรบัณฑิต สาขาเศรษฐศาสตร์ เกรดเฉลี่ย 2.93 (มหาวิทยาลัยศรีนครินทรวิโรฒ) (2567) วิทยาศาสตร์มหาบัณฑิต สาขาวิทยาการข้อมูลและการวิเคราะห์ (สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง) เกรดเฉลี่ย 3.70

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้