การใช้คลัสเตอร์แบบเคมีนและเคเนียร์เรสเนเบอร์ในการจัดกลุ่มลูกค้า และเพิ่มประสิทธิภาพ การกำหนดราคา: กรณีศึกษา บริษัทขนส่ง

พีรวิชญ์ ภิบาล

งานนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาการจัดการโลจิสติกส์และโซ่อุปทาน
คณะโลจิสติกส์ มหาวิทยาลัยบูรพา
2567

ลิขสิทธิ์เป็นของมหาวิทยาลัยบูรพา

การใช้คลัสเตอร์แบบเคมีนและเคเนียร์เรสเนเบอร์ในการจัดกลุ่มลูกค้า และเพิ่มประสิทธิภาพ การกำหนดราคา: กรณีศึกษา บริษัทขนส่ง

พีรวิชญ์ ภิบาล

งานนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาการจัดการ โลจิสติกส์และโซ่อุปทาน
คณะโลจิสติกส์ มหาวิทยาลัยบูรพา
2567

ลิขสิทธิ์เป็นของมหาวิทยาลัยบูรพา

USING K-MEANS CLUSTERING AND K-NEAREST NEIGHBOR TO SEGMENT CUSTOMERS AND INCREASE THE EFFICIENCY OF PRICE SETTING: A CASE STUDY OF TRANSPORT COMPANY

PEERAWIT PHIBAN

AN INDEPENDENT STUDY SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR MASTER DEGREE OF SCIENCE
IN LOGISTICS AND SUPPLY CHAIN MANAGEMENT
FACULTY OF LOGISTICS
BURAPHA UNIVERSITY
2024

COPYRIGHT OF BURAPHA UNIVERSITY

คณะกรรมการควบคุมงานนิพนธ์และคณะกรรมการสอบงานนิพนธ์ได้พิจารณางานนิพนธ์ของ พีรวิชญ์ ภิบาล ฉบับนี้แล้ว เห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาการจัดการโลจิสติกส์และโซ่อุปทาน ของมหาวิทยาลัยบูรพาได้

คณะกรรมการควบคุมงานนิพนธ์
อาจารย์ที่ปรึกษาหลัก

.....
(ผู้ช่วยศาสตราจารย์ ดร.ไพโรจน์ เร้าธนชลกุล)

คณะกรรมการสอบงานนิพนธ์

..... ประธาน
(รองศาสตราจารย์ ดร.ณกร อินทร์พยุง)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.มัธยะ ยุวมิตร)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.ไพโรจน์ เร้าธนชลกุล)

..... คณบดีคณะโลจิสติกส์
(รองศาสตราจารย์ ดร. ณกร อินทร์พยุง)
วันที่.....เดือน.....พ.ศ.....

บัณฑิตวิทยาลัย มหาวิทยาลัยบูรพา อนุมัติให้รับงานนิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาการจัดการโลจิสติกส์และโซ่อุปทาน ของมหาวิทยาลัยบูรพา

..... คณบดีบัณฑิตวิทยาลัย
(รองศาสตราจารย์ ดร.วิทวัส แจ้งเอี่ยม)
วันที่.....เดือน.....พ.ศ.....

65920239: สาขาวิชา: การจัดการโลจิสติกส์และโซ่อุปทาน; วท.ม. (การจัดการโลจิสติกส์และโซ่อุปทาน)
คำสำคัญ: คลัสเตอร์แบบเคมีน/ เคเนียร์เรสเนเบอร์/ การจัดกลุ่มลูกค้า

พีรวิชญ์ ภิบาล : การใช้คลัสเตอร์แบบเคมีนและเคเนียร์เรสเนเบอร์ในการจัดกลุ่มลูกค้า และเพิ่มประสิทธิภาพ การกำหนดราคา: กรณีศึกษา บริษัทขนส่ง. (USING K-MEANS CLUSTERING AND K-NEAREST NEIGHBOR TO SEGMENT CUSTOMERS AND INCREASE THE EFFICIENCY OF PRICE SETTING: A CASE STUDY OF TRANSPORT COMPANY) คณะกรรมการควบคุมงานนิพนธ์: ไพโรจน์ เร้าธนชลกุล ปี พ.ศ. 2567.

การจัดกลุ่มลูกค้า เป็นกลยุทธ์ที่ธุรกิจเกี่ยวกับโลจิสติกส์ต้องให้ความสำคัญ เนื่องจากลูกค้าที่ใช้บริการเพิ่มมากขึ้น ทำให้การจัดกลุ่มลูกค้าเป็นวิธีที่กำหนดกลุ่มเป้าหมายและการให้บริการ งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาการแบ่งกลุ่มลูกค้า เริ่มจากการเลือก และแปลงตัวแปรที่มีความสำคัญต่อการแบ่งกลุ่ม ได้แก่ จำนวนการใช้งานของลูกค้า ราคาที่จ่าย กำไรสุทธิ จากนั้น ใช้แบบจำลองคลัสเตอร์แบบเคมีน (K-Means clustering) เพื่อเพิ่มประสิทธิภาพในการจัดกลุ่มลูกค้า และการใช้แบบจำลองเคเนียร์เรสเนเบอร์ (K-Nearest Neighbors) เพื่อนำอัลกอริทึมไปใช้จำแนกกลุ่มข้อมูลใหม่ในอนาคต โดยที่ไม่ทำให้กลุ่มเดิมคลาดเคลื่อน ช่วยให้บริษัทสามารถลดต้นทุนในการกำหนดราคาโปรโมชั่นอย่างเป็นระบบ โดยใช้ Elbow method และ Silhouette score ในการวัดประสิทธิภาพแบบจำลองคลัสเตอร์แบบเคมีน (K-Means clustering) ผลการวิเคราะห์ พบว่า ค่า K = 3 เป็นค่าที่เหมาะสมที่สุด เมื่อทดลองวัดตั้งแต่ค่า K = 1 ถึง 10 ส่วนแบบจำลองเคเนียร์เรสเนเบอร์ (K-Nearest neighbors) โดยใช้ F1-Score และ Accuracy score ในการวัดประสิทธิภาพ ผลการทดลอง พบว่า Neighbor = 9 มีค่า F1-Score และ Accuracy score สูงสุด ซึ่งแสดงถึงประสิทธิภาพที่ดีที่สุดในการทำนายกลุ่ม แล้วจึงใช้แบบจำลองอัลกอริทึมเคเนียร์เรสเนเบอร์ มาจัดกลุ่มลูกค้าออกเป็น 3 กลุ่ม และให้โปรโมชั่นการกำหนดราคา ช่วยให้บริษัทสามารถเพิ่มกำไรเฉลี่ยร้อยละ 38.25 หรือเทียบเท่ากับ 32,838,644.16 บาท ซึ่งเป็นผลลัพธ์ที่ดีขึ้นมา กนอกจากนี้ ยังสามารถลดต้นทุนที่ใช้ในการให้โปรโมชั่นลงได้อีก 5,577,435.16 บาท

65920239: MAJOR: LOGISTICS AND SUPPLY CHAIN MANAGEMENT; M.Sc. (LOGISTICS AND SUPPLY CHAIN MANAGEMENT)
KEYWORDS: K-MEANS CLUSTERING/ K-NEAREST NEIGHBORS/ CUSTOMER SEGMENTATION

PEERAWIT PHIBAN : USING K-MEANS CLUSTERING AND K-NEAREST NEIGHBOR TO SEGMENT CUSTOMERS AND INCREASE THE EFFICIENCY OF PRICE SETTING: A CASE STUDY OF TRANSPORT COMPANY. ADVISORY COMMITTEE: PAIROJ RAOTHANACHONKUN, Ph.D. 2024.

Customer segmentation is a crucial strategy for businesses in the logistics sector. As the number of service users increases, segmentation helps in targeting and providing services efficiently. This research aims to study customer segmentation, starting with the selection and transformation of significant variables for segmentation, including customer usage frequency, payment amounts, and net profits. A K-Means clustering model is employed to enhance the efficiency of customer segmentation. Additionally, a K-Nearest neighbors (KNN) model is used to classify new data into existing clusters accurately. This helps the company systematically reduce promotional pricing costs. The elbow method and silhouette score are used to evaluate the K-Means clustering model' s performance. The analysis revealed that K = 3 is the optimal value when testing K values from 1 to 10. For the KNN model, the F1-score and accuracy score are used to measure performance. The experiment showed that neighbor = 9 yielded the highest F1-score and accuracy score, indicating the best prediction efficiency. The KNN algorithm is then used to segment customers into three groups and to allocate promotional pricing, helping the company increase profits by an average of 22.45%, equivalent to 32,838,644.16 baht. Additionally, it reduced the cost of promotions by 5,577,435.16 baht.

กิตติกรรมประกาศ

งานวิจัยนี้ สำเร็จลุล่วงไปด้วยดี ผู้วิจัยขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.ไพโรจน์ เร้าธนชลกุล อาจารที่ปรึกษาวิทยานิพนธ์ ผู้ดูแล ให้คำปรึกษา และควบคุมวิทยานิพนธ์อย่างใกล้ชิด ขอขอบพระคุณ รองศาสตราจารย์ ดร.ณกร อินทร์พยุง และผู้ช่วยศาสตราจารย์ ดร.มัธยะ ยุวมิตร ที่กรุณาให้เกียรติเป็นกรรมการงานนิพนธ์ ให้คำแนะนำและปรึกษาขั้นตอนการทำวิจัย การศึกษาและวิเคราะห์ผลการศึกษา การใช้คลัสเตอร์แบบเคมีนและเคเนียร์เรสเนเบอร์ในการจัดกลุ่มลูกค้าเพื่อเพิ่มประสิทธิภาพในการกำหนดราคา: กรณีศึกษา บริษัทขนส่ง ขอขอบพระคุณคณาจารย์ทุกท่านในคณะโลจิสติกส์ ท่านที่ให้คำแนะนำ และให้ความรู้อันเป็นประโยชน์ต่องานวิจัย การดำเนินงานและการทำงาน ขอขอบพระคุณบิดา มารดา และครอบครับที่ให้การสนับสนุนในด้านการศึกษาและกำลังใจในการทำงาน ขอบคุณเพื่อน พี่ ที่ให้คำแนะนำ ให้คำปรึกษา และกำลังใจในการทำงานวิจัย และขอบคุณผู้ให้คำสัมภาษณ์ทุกท่าน ที่ให้ความรู้ ข้อเสนอแนะ และการพัฒนาต่อในงานวิจัยนี้ ขอขอบพระคุณทุกหน่วยงานที่มีพระคุณทุกท่านอย่างสูงในการช่วยงานวิจัยนี้

สุดท้ายนี้ หากผลการศึกษามีข้อบกพร่องประการใด ผู้วิจัยขอน้อมรับไว้เพื่อการปรับปรุงแก้ไขในการศึกษาครั้งถัดไป

พีรวิชญ์ ภิบาล

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญตาราง ..... ฌ

สารบัญภาพ ..... ญ

บทที่ 1 บทนำ 1

ความเป็นมาและความสำคัญของปัญหา 1

วัตถุประสงค์ของการวิจัย 2

ขอบเขตของการวิจัย 2

ประโยชน์ที่ได้รับจากการวิจัย 2

นิยามศัพท์เฉพาะ 2

บทที่ 2 แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง 4

แนวคิดเกี่ยวกับการจัดการโลจิสติกส์ 4

แนวคิดทฤษฎีเกี่ยวกับการตั้งราคา 7

แนวคิดเกี่ยวกับการเก็บข้อมูลในรูปแบบอิเล็กทรอนิกส์ 9

แนวคิดทฤษฎีเกี่ยวกับการจัดกลุ่มลูกค้า 10

แนวคิดทฤษฎี Feature engineering 19

แนวคิดทฤษฎี Data mining 20

งานวิจัยที่เกี่ยวข้อง 21

บทที่ 3 วิธีดำเนินงานวิจัย 26

การเก็บรวบรวมข้อมูล 26
ขั้นตอนการดำเนินงานวิจัย 28
เครื่องมือที่ใช้ในการดำเนินงานวิจัย.....29
การวิเคราะห์ข้อมูล.....31
สรุปผลการดำเนินงาน 34
บทที่ 4 ผลการดำเนินงานวิจัย.....36
ศึกษาตัวแปรที่มีผลต่อการแบ่งกลุ่มข้อมูล 36
การแบ่งกลุ่มลูกค้าโดยใช้คลัสเตอร์แบบเคมีน.....39
การแบ่งกลุ่มโดยใช้เคเนียร์เรสเนเบอร์ 44
การกำหนดราคา.....46
บทที่ 5 สรุป อภิปรายผล และข้อเสนอแนะ 49
สรุปผลการวิจัย 49
อภิปรายผล.....50
ข้อเสนอแนะ 51
บรรณานุกรม 52
ภาคผนวก 56
ภาคผนวก ก 57
ภาคผนวก ข 60
ประวัติย่อของผู้วิจัย 65

สารบัญตาราง

หน้า

ตารางที่ 1 ตัวแปรทั้งหมดในงานวิจัย 31

ตารางที่ 2 การแปลงข้อมูลชนิดตัวเลขให้เป็นค่ามาตรฐาน 33

ตารางที่ 3 การแปลงข้อมูลชนิดข้อความให้เป็นตัวเลขและทำให้กลายเป็นค่ามาตรฐาน 34

ตารางที่ 4 จำนวนข้อมูลที่ถูกแบ่งกลุ่มโดย คลัสเตอร์แบบเคมีน 39

ตารางที่ 5 ค่าการวัดประสิทธิภาพ F1-score และ Accuracy score ของค่า Number of neighbor แต่ละค่า 45

ตารางที่ 6 ผลลัพธ์การจัดกลุ่มข้อมูลทั้งหมดและวัดค่าร้อยละกำไรเฉลี่ยของแต่ละกลุ่ม 47

ตารางที่ 7 ผลการกำหนดราคาโปรโมชันแบบจัดกลุ่ม.....47

ตารางที่ 8 ผลลัพธ์ร้อยละของรูปแบบการกำหนดราคา 48

สารบัญภาพ

หน้า

ภาพที่ 1 ห่วงโซ่คุณค่าทั่วไป....5
ภาพที่ 2 ความสามารถในการแข่งขันและ “3C” 6
ภาพที่ 3 แนวคิดของระบบโดยรวม 7
ภาพที่ 4 ต้นทุนการขนส่ง 9
ภาพที่ 5 การแบ่งกลุ่มโมเดลของ Machine learning 11
ภาพที่ 6 แผนภาพประเภทของ Machine learning โดยละเอียด 12
ภาพที่ 7 วิธีคิดการวัดระยะห่างโดยใช้สมการยุคลิด โมเดลเคเนียร์เรสเนเบอร์....13
ภาพที่ 8 K1, K3, K5 14
ภาพที่ 9 Confusion Matrix 15
ภาพที่ 10 วิธีคิดการวัดระยะห่างโดยใช้สมการยุคลิด โมเดลคลัสเตอร์แบบเคมีน 17
ภาพที่ 11 ระดับความแม่นยำมีแนวโน้มจะขึ้นไปสูงสุดเมื่อ K = 3 18
ภาพที่ 12 การหาจำนวนกลุ่มจากค่า Silhouette score 19
ภาพที่ 13 กระบวนการ Features engineering....20
ภาพที่ 14 กระบวนการทำงาน Data Mining 21
ภาพที่ 15 ตัวอย่างข้อมูลที่ใช้สำหรับวิเคราะห์การแบ่งกลุ่ม....27
ภาพที่ 16 ขั้นตอนการศึกษาดำเนินงาน....29
ภาพที่ 17 การแบ่งชุดข้อมูลฝึกสอนและทดสอบ....30
ภาพที่ 18 การแบ่งกลุ่มลูกค้าจากจำนวนการใช้งาน....38
ภาพที่ 19 การแบ่งข้อมูลโดยใช้ คลัสเตอร์แบบเคมีน ค่า K = 10....40
ภาพที่ 20 ค่า Silhouette score เมื่อค่า K = 10....41
ภาพที่ 21 ค่า Elbow method ของแต่ล่ะ คลัสเตอร์ เมื่อค่า K = 10....42

ภาพที่ 22 Silhouette score เมื่อ ค่า K = 3 42
ภาพที่ 23 การแบ่งข้อมูลโดยใช้ คลัสเตอร์แบบเคมีน ค่า K = 3 43
ภาพที่ 24 กระบวนการจัดกลุ่มด้วย เคเนียร์เรสเนเบอร์ 44
ภาพที่ 25 การวัดประสิทธิภาพ F1-score และ Accuracy score ของค่า Number of neighbor แต่ละค่า 46
ภาพที่ 26 ตัวอย่างการบันทึกข้อมูลบริษัทกรณีศึกษา ก่อนแก้ไขข้อมูล 58
ภาพที่ 27 ตัวอย่างการบันทึกข้อมูลบริษัทกรณีศึกษา หลังแก้ไขข้อมูล 59
ภาพที่ 28 ตัวอย่างชุดข้อมูลที่ใช้ในการวิเคราะห์ คลัสเตอร์แบบเคมีน 61
ภาพที่ 29 ตัวอย่างชุดข้อมูลที่ได้จากการวิเคราะห์ คลัสเตอร์แบบเคมีนและเคเนียร์เรสเนเบอร์.....62
ภาพที่ 30 การเลือกฟีเจอร์และทำข้อมูลให้เป็นมาตรฐานและจัดกลุ่มด้วยเคมีนคลัสเตอร์และเคเนียร์เบอร์ 63

บทที่ 1
บทนำ

ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันหลายบริษัทในหลากหลายอุตสาหกรรมยังคงใช้วิธีการให้โปรโมชันแก่ลูกค้าในรูปแบบที่ขาดการวิเคราะห์และไม่ได้อาศัยเทคนิคการจัดโปรโมชันที่มีประสิทธิภาพ ผลที่ตามมาคือข้อเสียทางการค้าที่สำคัญหลายประการ อาทิ การใช้ทรัพยากรที่ไม่เหมาะสม การเสียโอกาสในการสร้างรายได้เพิ่มเติม และการบริหารจัดการที่ขาดประสิทธิภาพ หนึ่งในปัญหาหลักของการให้โปรโมชันแบบสุ่มหรือไร้การวิเคราะห์คือ การที่บริษัทไม่สามารถระบุได้ว่าลูกค้ากลุ่มใดมีมูลค่าสูงและควรได้รับการส่งเสริมมากที่สุด ทำให้โปรโมชันที่ให้ไปอาจไม่ได้รับผลตอบแทนที่คุ้มค่า และส่งผลให้เกิดการสิ้นเปลืองทรัพยากรโดยไม่จำเป็น การใช้โปรโมชันอย่างไม่มีแบบแผนยังส่งผลให้บริษัทไม่สามารถใช้ประโยชน์จากข้อมูลที่มีอยู่เพื่อสร้างกลยุทธ์ทางการตลาดที่มีประสิทธิภาพได้อย่างเต็มที่ (Kotler & Keller, 2016)

การนำเทคนิคการเรียนรู้ของเครื่อง Machine learning และการวิเคราะห์ข้อมูล (Data analysis) มาประยุกต์ใช้ในการแบ่งกลุ่มลูกค้าและการกำหนดโปรโมชันที่เหมาะสม สามารถช่วยให้บริษัทสามารถจัดสรรทรัพยากรได้อย่างมีประสิทธิภาพยิ่งขึ้น และเพิ่มโอกาสในการสร้างรายได้จากลูกค้าแต่ละกลุ่ม การวิเคราะห์ข้อมูลเชิงลึกเกี่ยวกับพฤติกรรมการใช้บริการและข้อมูลทางการเงินของลูกค้า จะช่วยให้บริษัทสามารถระบุได้ว่าลูกค้ากลุ่มใดมีแนวโน้มที่จะสร้างกำไรสูงสุดและควรได้รับการส่งเสริมผ่านโปรโมชันที่ออกแบบมาเฉพาะสำหรับกลุ่มนั้น นอกจากนี้การแบ่งกลุ่มลูกค้าโดยใช้อัลกอริทึมการเรียนรู้ของเครื่องยังช่วยให้บริษัทสามารถปรับกลยุทธ์การตลาดให้สอดคล้องกับการเปลี่ยนแปลงของตลาดและพฤติกรรมของลูกค้าได้อย่างรวดเร็วและแม่นยำ ตัวอย่างเช่น การใช้วิธีการคลัสเตอร์แบบเคมีน (K-means clustering) ในการจัดกลุ่มลูกค้าและเคเนียร์เรสเนเบอร์ (K-nearest neighbors) ในการทำนายกลุ่มในอนาคตเพื่อนำไปใช้งานจริงโดยไม่ทำให้กลุ่มข้อมูลเดิมคลาดเคลื่อน สามารถช่วยให้บริษัทเข้าใจโครงสร้างของกลุ่มลูกค้าและสร้างโปรโมชันที่ตรงกับความต้องการและพฤติกรรมการใช้งานของลูกค้าในแต่ละกลุ่มได้ดียิ่งขึ้น การศึกษาและการประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่องในกระบวนการจัดการโปรโมชันยังสามารถช่วยให้บริษัทลดต้นทุนในการให้โปรโมชันโดยไม่จำเป็น และเพิ่มความสามารถในการแข่งขันในตลาด การกำหนดโปรโมชันที่เหมาะสมกับกลุ่มลูกค้าที่มีศักยภาพสูงยังช่วยให้บริษัทสามารถสร้างรายได้เพิ่มเติมจากลูกค้าเหล่านี้ได้อย่างมีประสิทธิภาพ (Pranshu & Tommy, 2019)

บริษัทกรณีศึกษาที่ดำเนินกิจการรับขนย้ายเครื่องจักรทั่วไปและรถขนส่งประสบปัญหาที่ลูกค้ามีความถี่ในการใช้บริการที่แตกต่างกัน และมักมีการต่อรองราคาก่อนจะตกลงวิ่งงาน ปัจจุบันผู้มีอำนาจในการตั้งราคาในบริษัทมักกำหนดราคาตามระเบียบของบริษัทโดยไม่อิงถึงข้อมูลที่มีการบันทึกไว้ ทำให้การตั้งราคาไม่เป็นมาตรฐาน บริษัทกรณีศึกษามีความต้องการระบบที่เป็นมาตรฐานในการตั้งราคาให้กับกลุ่มลูกค้าที่มีลักษณะการใช้บริการที่แตกต่างกัน โดยมีวัตถุประสงค์เพื่อสร้างความโปร่งใสในการตั้งราคา การนำระบบมาตรฐานในการตั้งราคาเข้ามาใช้จะช่วยให้บริษัทกรณีศึกษาสามารถตั้งราคาได้อย่างเป็นธรรม อีกทั้งยังช่วยเพิ่มประสิทธิภาพในการดำเนินงาน ลดความซับซ้อนในการเจรจาต่อรอง

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาลักษณะการกำหนดราคาค่าขนส่ง
2. เพื่อจัดกลุ่มลูกค้าโดยพิจารณาตัวแปรที่สำคัญ
3. เพื่อเพิ่มประสิทธิภาพการตั้งราคาสำหรับกลุ่มลูกค้า

ขอบเขตของการวิจัย

งานวิจัยนี้ใช้ข้อมูลบันทึกการวิ่งงานในการขนส่งของลูกค้า โดยนักศึกษาได้เลือกใช้ข้อมูลเฉพาะกลุ่มลูกค้าที่อยู่ในสมุดบันทึกงานเท่านั้น ซึ่งเป็นข้อมูลที่บันทึกตั้งแต่วันที่ 4 สิงหาคม พ.ศ. 2564 ถึงวันที่ 31 สิงหาคม พ.ศ. 2566 รวมทั้งสิ้น 4,270 รายการ จากจำนวนลูกค้าทั้งหมด 250 ราย โดยจัดกลุ่มลูกค้าด้วยอัลกอริทึม คือ คลัสเตอร์แบบเคมีน กับ เคเนียร์เรสเนเบอร์ และนำมาจัดกลุ่มลูกค้า

ประโยชน์ที่ได้รับจากการวิจัย

1. ทราบลักษณะการตั้งราคาค่าขนส่ง
2. ทราบกลุ่มลูกค้าโดยพิจารณาตัวแปรที่สำคัญ
3. ทราบแนวทางการตั้งราคาสำหรับกลุ่มลูกค้าที่มีประสิทธิภาพ

นิยามศัพท์เฉพาะ

คลัสเตอร์แบบเคมีน (K-means clustering) คือ เป็นเทคนิคที่ใช้ในการแบ่งกลุ่มข้อมูล โดยที่ข้อมูลที่มีอยู่จะถูกแบ่งออกเป็นกลุ่มย่อย โดยอาศัยคุณสมบัติหรือลักษณะที่คล้ายคลึงกันจับกลุ่มของข้อมูลลูกค้าบริษัทที่มีลักษณะใกล้เคียงกันเป็นกลุ่มเดียวกัน

เคเนียร์เรสเนเบอร์ (K-nearest neighbors) คือ ซึ่งเป็นอัลกอริทึมในการจำแนกประเภททำการสอนโดยใช้ข้อมูลของลูกค้าที่มีการแบ่งกลุ่มมาแล้วในการสอนอัลกอริทึมแล้วจึงนำอัลกอริทึมมาใช้กับข้อมูลบริษัทที่เป็นข้อมูลทดสอบเพื่อแยกว่าชุดข้อมูลเป็นข้อมูลกลุ่มไหน

การจัดกลุ่มลูกค้า (Customer) คือ กลุ่มบุคคลหรือองค์กรที่ใช้บริการจากบริษัท กรณีศึกษา ในบริบททางธุรกิจ ลูกค้าเป็นกลุ่มเป้าหมายสำคัญที่บริษัทมุ่งเน้นเพื่อให้เกิดการใช้บริการ ซึ่งลูกค้าสามารถแบ่งออกเป็นหลายประเภทตามลักษณะการใช้บริการ

บทที่ 2
แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

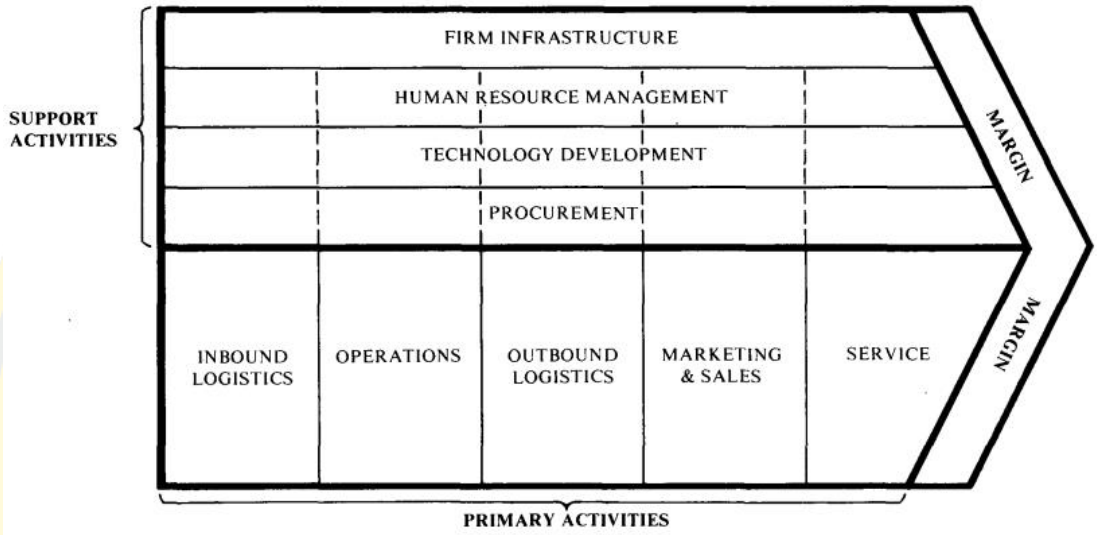
การศึกษาวิจัยเรื่องการใช้ Machine learning ในการจัดกลุ่มลูกค้าเพื่อเพิ่มประสิทธิภาพในการตั้งราคาผู้วิจัยได้สืบค้นศึกษาข้อมูล แนวคิดทฤษฎี ตำรา และงานวิจัยที่เกี่ยวข้อง เพื่อนำมาใช้เป็นแนวทางแก้ปัญหาและส่งเสริมงานวิจัยมีดังนี้

1. แนวคิดเกี่ยวกับการจัดการโลจิสติกส์
2. แนวคิดทฤษฎีเกี่ยวกับการตั้งราคา
3. แนวคิดเกี่ยวกับการเก็บข้อมูลในรูปแบบอิเล็กทรอนิกส์
4. แนวคิดทฤษฎีเกี่ยวกับการจัดกลุ่มลูกค้า
5. แนวคิดทฤษฎี Feature engineering
6. แนวคิดทฤษฎี Data mining
7. งานวิจัยที่เกี่ยวข้อง

แนวคิดเกี่ยวกับการจัดการโลจิสติกส์

การจัดการโลจิสติกส์เริ่มต้นขึ้นจากการดำเนินงานทางทหาร โดยทัพอังกฤษก่อนสงครามโลกครั้งที่ 1 เป็นผู้นำการจัดการส่งเสบียงให้กองทัพ ซึ่งรวมถึงการสร้างและบริหารจัดการโครงสร้างพื้นฐาน เช่น ถนน รถไฟ ท่าเรือ สนามบิน สถานที่จัดเก็บสินค้า และยานพาหนะที่ใช้ในการลำเลียงอาวุธยุทโธปกรณ์ โลจิสติกส์ในยุคแรกเน้นไปที่การประกันความพร้อมใช้งานและความสามารถในการเคลื่อนย้ายทรัพยากรอย่างมีประสิทธิภาพ ก่อนปลายศตวรรษที่ 19 สหรัฐอเมริกาเริ่มมีการพัฒนาการกระจายสินค้าพืชทางการเกษตร ซึ่งมีการแลกเปลี่ยนสินค้าโภคภัณฑ์อย่างแพร่หลายทั่วทวีปอเมริกา การพัฒนานี้ได้ขยายตัวไปสู่การกระจายสินค้าด้านอุตสาหกรรมในช่วงปี ค.ศ. 1870 ทำให้เกิดการจัดการโลจิสติกส์ที่มีรูปแบบเป็นระบบมากขึ้น (JWD GROUP, 2020) ห่วงโซ่คุณค่า หรือ Value chain เป็นแนวคิดที่ได้รับการพัฒนาโดย Michael Porter และถูกนำมาใช้ทั่วโลกมานานกว่า 30 ปี เป็นเครื่องมือที่แบ่งกิจกรรมในบริษัทที่เกี่ยวข้องกับกลยุทธ์ โดยกิจกรรมเหล่านี้สามารถส่งผลให้ราคาสูงขึ้นหรือต้นทุนต่ำ ห่วงโซ่คุณค่าของบริษัทประกอบด้วยกิจกรรมหลักและกิจกรรมสนับสนุน ซึ่งช่วยเพิ่มประสิทธิภาพให้มากกว่าคู่แข่งหรือสร้างความแตกต่างเพื่อความได้เปรียบเชิงการแข่งขัน สามารถแบ่งออกได้ 2 ประเภท คือ กิจกรรมหลัก และกิจกรรมสนับสนุน โดยในกิจกรรมหลัก มีตำแหน่ง ได้แก่ ตำแหน่งขนส่งวัตถุดิบขาเข้า

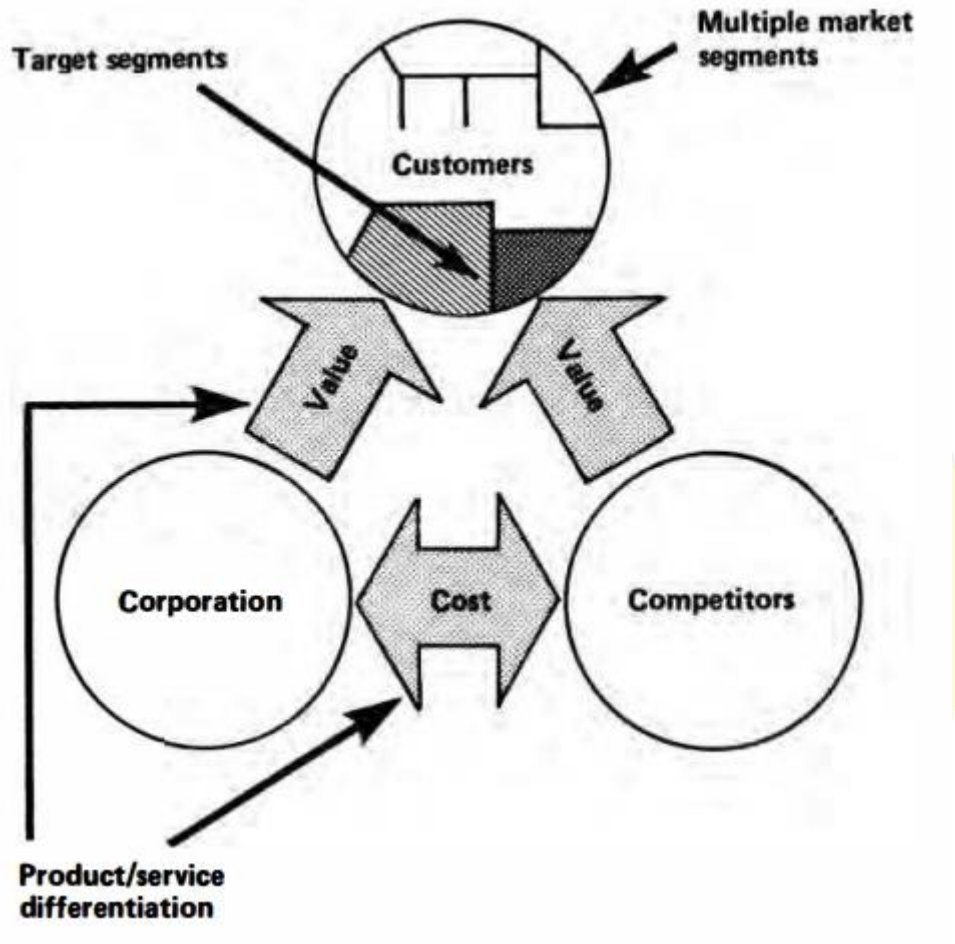
ตำแหน่งปฏิบัติการ ตำแหน่งขนส่งสินค้าขาออก ตำแหน่งขายและการตลาด ตำแหน่งบริการ ส่วนกิจกรรมสนับสนุน นั้นมีกิจกรรม ได้แก่ ตำแหน่งทรัพยากรมนุษย์ ตำแหน่งพัฒนาเทคโนโลยี ตำแหน่งจัดซื้อ โดยเพิ่มประสิทธิภาพให้มากกว่าคู่แข่งหรือสร้างความแตกต่างจากคู่แข่งเพื่อความได้เปรียบเชิงการแข่งขัน (Porter, 1985) ดังภาพที่ 3



ภาพที่ 1 ห่วงโซ่คุณค่าทั่วไป (Porter, 1985)

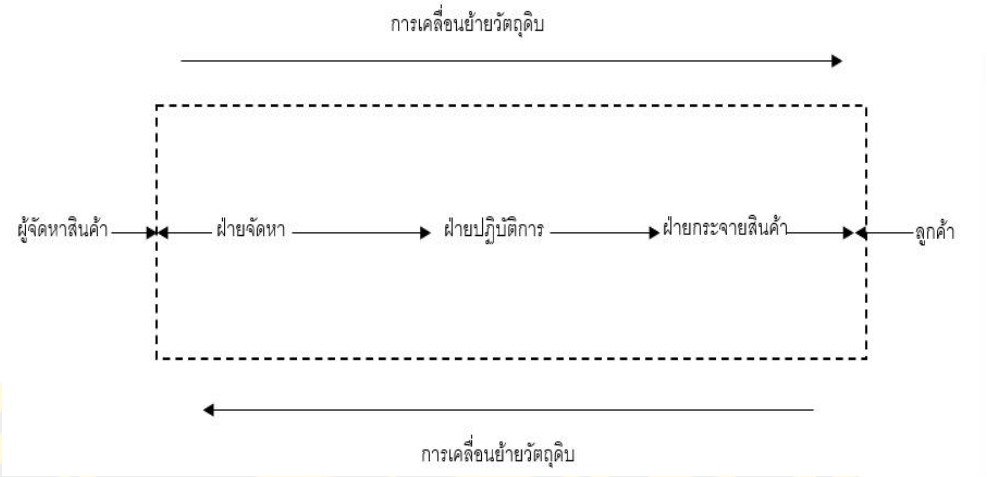
การจัดการโลจิสติกส์และห่วงโซ่อุปทานเป็นกระบวนการที่ซับซ้อนและสำคัญในการดำเนินธุรกิจสมัยใหม่ ทั้งสองแนวคิดนี้มีบทบาทสำคัญในการเพิ่มประสิทธิภาพในการดำเนินงานและสร้างความได้เปรียบเชิงการแข่งขัน การจัดการห่วงโซ่อุปทาน เป็นการวางแผนและการจัดการข้อมูลธุรกิจและผลิตภัณฑ์ในทุกขั้นตอนของการผลิตและการส่งมอบสินค้า การประสานงานในห่วงโซ่อุปทานมีความสำคัญอย่างยิ่งต่อการทำให้ทุกหน่วยงานภายในบริษัททำงานร่วมกันอย่างมีประสิทธิภาพ รวมถึงการเชื่อมโยงกับหน่วยงานภายนอก เช่น ลูกค้าและคู่แข่ง หลักการเชื่อมโยงสามเหลี่ยม หลักการเชื่อมโยงสามเหลี่ยมที่ประกอบด้วย บริษัท ลูกค้า และคู่แข่ง เป็นแนวคิดที่ช่วยสร้างความแตกต่างจากการแข่งขัน ความสัมพันธ์ที่ดีระหว่างสามกลุ่มนี้ช่วยให้การดำเนินงานมีประสิทธิภาพมากขึ้น ลดต้นทุนและเพิ่มกำไร ตลอดจนทำให้ธุรกิจมีความยั่งยืนในระยะยาว (Christopher, 1992) จากภาพตัวอย่างที่ 4 โดยเป็นระบบขององค์กรกิจกรรมข้อมูล ทรัพยากรที่เกี่ยวข้องกับ การวางแผน การย้าย การจัดเก็บ หรือบริการจากซัพพลายเออร์ไปยังลูกค้า ครอบคลุมการวางแผนและการจัดการกิจกรรมทั้งหมดที่เกี่ยวข้องกับการจัดหา การจัดซื้อ

การจัดการโลจิสติกส์รวมถึงส่วนที่สำคัญของการประสานงานและความร่วมมือกับพันธมิตร (Myerson, 2015)



ภาพที่ 2 ความสามารถในการแข่งขันและ “3C” (Ohmae, 1982)

ทั้งนี้กิจกรรมทั้งหมดที่ทำให้บรรลุเป้าหมายตั้งแต่การจัดการ วางแผน การขนส่งบริการไปยังลูกค้า ด้วยต้นทุนที่ต่ำที่สุด โดยประสานการดำเนินการของธุรกิจ ตลาดซัพพลายเออร์ และทั่วทั้งองค์กร ซึ่งสามารถมองเห็นได้ทั้งระบบ มีการรวมบริษัทให้เป็นหนึ่งเดียวทำให้บรรลุเป้าหมายได้สำเร็จ (ทวีศักดิ์ เทพพิทักษ์, 2548)



ภาพที่ 3 แนวคิดของระบบโดยรวม (ทวีศักดิ์ เทพพิทักษ์, 2548)

การจัดการโลจิสติกส์และโซ่อุปทาน มีความคล้ายกันแต่ไม่เหมือนกัน การเข้าใจความแตกต่างการจัดการโซ่อุปทานและโลจิสติกส์เป็นสิ่งสำคัญกับธุรกิจในการเพิ่มผลกำไรสูงสุด การจัดการห่วงโซ่อุปทานมุ่งเน้นไปที่การจัดการความสัมพันธ์กับซัพพลายเออร์และลูกค้าเพื่อเพิ่มประสิทธิภาพการไหลของวัสดุและบริการ โลจิสติกส์มุ่งเน้นไปที่การเคลื่อนย้ายทางกายภาพที่แท้จริงของสินค้าจากที่หนึ่งไปยังอีกที่หนึ่ง (YILDIZ, 2023)

แนวคิดทฤษฎีเกี่ยวกับการตั้งราคา

การตั้งราคาเป็นหนึ่งในปัจจัยสำคัญที่มีผลกระทบต่อความสำเร็จของธุรกิจ การตั้งราคาไม่ได้เป็นเพียงการกำหนดตัวเลขเท่านั้น แต่ยังเป็นกลยุทธ์ที่เกี่ยวข้องกับการแบ่งตลาด การกำหนดราคาสามารถส่งผลต่อการเพิ่มหรือลดการผลิต และยังสะท้อนถึงวัตถุประสงค์และกลยุทธ์ขององค์กรอีกด้วย แนวคิดทฤษฎีเกี่ยวกับการตั้งราคามีหลากหลายรูปแบบ ซึ่งแต่ละแบบมีความสำคัญและมีการประยุกต์ใช้แตกต่างกันไปในแต่ละสถานการณ์ การตั้งราคามีบทบาทสำคัญต่อกลยุทธ์ทางธุรกิจ การเลือกวิธีการตั้งราคาที่เหมาะสมกับสถานการณ์และเป้าหมายของธุรกิจจะช่วยให้ธุรกิจสามารถแข่งขันได้และเพิ่มความยั่งยืนในระยะยาว

1. การตั้งค่าราคา

การตั้งราคา Markup หรือ Markup on cost เป็นวิธีการตั้งราคาสินค้าหรือบริการโดยการเพิ่มกำไรที่ต้องการเข้าไปในต้นทุน วิธีนี้เป็นที่นิยมในธุรกิจค้าปลีกและการผลิต เนื่องจากง่ายต่อการคำนวณและการใช้งาน การตั้งราคาแบบ Markup ที่เหมาะสมจะช่วยให้ธุรกิจสามารถทำกำไร

สูงสุดได้โดยพิจารณาจากค่าความยืดหยุ่นของราคาสินค้า ซึ่งสามารถคำนวณได้ดังนี้ (สุจินดา เจียมศรีพงษ์, 2558)

$$Mark\ up = Price \times Precentmarkup$$

โดยให้ Price = ราคา

Percent markup = จำนวนเงินหรือเปอร์เซ็นต์ที่ใช้ในการเพิ่มราคาจากราคาเริ่มต้น

2. การตั้งค่าราคา Markdown

การตั้งราคา Markdown คือการลดราคาสินค้าหรือบริการจากราคาเดิมที่ตั้งไว้เพื่อกระตุ้นการขายหรือจัดการกับสินค้าคงคลัง การลดราคานี้อาจเกิดขึ้นได้จากหลายสาเหตุ เช่น การขายล้างสต็อก การลดราคาตามฤดูกาล หรือการส่งเสริมการขายในระยะสั้น การตั้งราคา Markdown เป็นกลยุทธ์ที่สำคัญสำหรับธุรกิจในการดึงดูดลูกค้าและเพิ่มยอดขายสมการคือ (สุจินดา เจียมศรีพงษ์, 2558)

$$Markdown = Price \times Percentmarkdown$$

โดยให้ Price = ราคาเริ่มต้น

Percentmarkdown = จำนวนเงินหรือเปอร์เซ็นต์ที่ใช้ในการลดราคาจากราคาเริ่มต้น

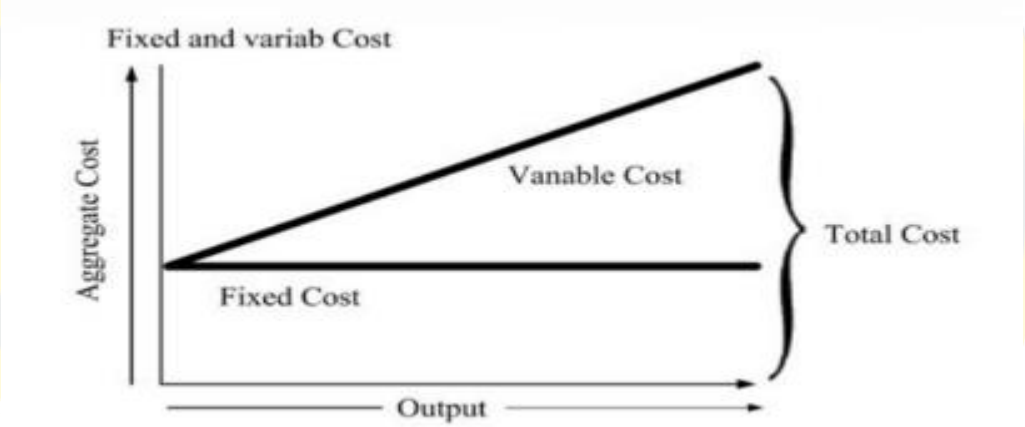
ต้นทุนและการตั้งราคาค่าขนส่ง โครงสร้างต้นทุนขนส่งนั้นมีอยู่หลายอย่างขึ้นอยู่กับลักษณะ สินค้าผู้ประกอบการ ตลาด ระยะทาง ลักษณะสินค้า ปริมาณสินค้า เงื่อนไข เป็นต้น ต้นทุนจะมีดังนี้

3. ต้นทุนคงที่ หรือ Fixed cost เป็นค่าใช้จ่ายที่ไม่มีการเปลี่ยนแปลงจำนวนคงที่ไม่ว่าจะมีการผลิตมากหรือน้อยก็ต้องเสียค่าใช้จ่ายเท่าเดิมตลอดเวลา เช่น เช่าอาคาร เช่าที่ดิน ประกันภัย ทะเบียนยานพาหนะ เป็นต้น

4. ต้นทุนผันแปร หรือ Variable cost เป็นค่าใช้จ่ายที่มีการผันแปรไปตามจำนวนการผลิต ผลิตมากจ่ายมากผลิตน้อยจ่ายน้อย ถ้ามีการขนส่งมากค่าใช้จ่ายจะเพิ่มขึ้นถ้ามีการขนส่งน้อยค่าใช้จ่ายจะลดลง เช่น ค่าน้ำมัน ค่าบำรุงรักษา ค่าใช้จ่ายในการขนส่ง เป็นต้น

5. ต้นทุนรวม หรือ Total cost โดยการเอาต้นทุนคงที่และต้นทุนผันแปรมาบวกเข้าด้วยกันซึ่งคือต้นทุนของการบริการและการขนส่งสินค้า ไม่สามารถต้นทุนของการขนส่งสินค้าหรือบริการได้ เช่น การขนส่งทางรางหรือรถไฟ ที่มีการขนสินค้ากับผู้ใช้บริการ อยู่ในขบวนเดียวกัน จึงกลายเป็นต้นทุนรวมเพราะแยกไม่ได้ว่าเป็นต้นทุนขนส่งผู้โดยสารเท่าใดหรือเป็นต้นทุนขนส่งสินค้าและบริการจำนวนเท่าใด เป็นต้น

6. ต้นทุนเที่ยวกลับ หรือ Back haul cost คือ ค่าใช้จ่ายการเสียโอกาสด้วยการที่ต้องขนส่งสินค้าบริการถึงปลายทางแล้ว เที่ยวกลับไม่มีสินค้ากลับมาด้วยจึงวิ่งรถเปล่ากลับ ดังนั้นต้องคิดต้นทุนเที่ยวกลับด้วย ซึ่งเป็นการทำให้เกิดการสูญเปล่าบริษัทที่รถขนส่งสินค้าต้องคำนึงถึงต้นทุนตรงนี้ด้วย



ภาพที่ 4 ต้นทุนการขนส่ง (คำนาย อภิปรัชญาสกุล, 2559)

แนวคิดเกี่ยวกับการเก็บข้อมูลในรูปแบบอิเล็กทรอนิกส์

การเก็บข้อมูลในรูปแบบอิเล็กทรอนิกส์ คือ การเปลี่ยนกระบวนการจากกระดาษมาเป็นระบบอัตโนมัติแบบไร้กระดาษ ซึ่งจะขึ้นอยู่กับ การวางแผน ขอบเขต ข้อกำหนดขององค์กร ความต้องการของผู้บริหารกับพนักงานมีความสอดคล้องกันจะมีทิศทางที่ประสบความสำเร็จได้อย่างมีประสิทธิภาพ (Xiong, 2020) เอกสารที่เป็นรูปแบบชนิดไฟล์ ช่วยลดการใช้ทรัพยากรกระดาษ ประหยัดค่าอุปกรณ์ในสำนักงาน เช่น หมึกพิมพ์ อุปกรณ์การเก็บเอกสาร เป็นต้น ยังช่วยลดการทำงานผิดพลาด ล่าช้า ป้องกันเอกสารสูญหาย ไฟล์ที่นิยมใช้ คือ PDF เพราะเป็นรูปแบบที่ทั่วโลกใช้งานกัน ใช้งานง่าย มีความรวดเร็วในการแสดงผล จะช่วยให้องค์กรจัดการการทำงานได้อย่างสมบูรณ์แบบและมีมาตราฐาน (Ditto, 2021) ซึ่งจะใช้คอมพิวเตอร์พร้อมโปรแกรม เช่น

เครื่องมือ Microsoft office โดยมี Excel, Power point, Word หรือเครื่องมือจาก Google workspace อย่าง Docs, Sheets, และ Slides เป็นต้น (Business & Technology, 2023)

แนวคิดทฤษฎีเกี่ยวกับการจัดกลุ่มลูกค้า

การจัดกลุ่ม หรือ Clustering อยู่ในประเภทของการเรียนรู้แบบไม่มีผู้สอนหรือ Unsupervised learning เป็นเทคนิคการแบ่งข้อมูลโดยจะเปิดเผยกลุ่มย่อยในชุดข้อมูลที่แตกต่างกันที่มีอยู่ หรือตัวแปรที่มี 2 กลุ่มขึ้นไป เพื่อให้แต่ละคลัสเตอร์อยู่ในกลุ่มที่เหมาะสม มากกว่าข้อมูลทั้งหมด โดยมีความต้องการให้แบ่งกลุ่มของวัตถุประสงค์ที่คล้ายกันอยู่ในกลุ่มเดียวกัน ถ้าแตกต่างจะอยู่ในกลุ่มอื่นที่มีการกำหนดค่าวัตถุประสงค์ไว้ โดยประเภทของการจัดกลุ่มมีดังนี้ (Joshi, 2022)

1. Partitioning methods คือ การทำคลัสเตอร์แบบ Centroid เป็นวิธีที่ง่ายที่สุดในประเภทการทำคลัสเตอร์ทั้งหมดมีประสิทธิภาพแต่ขึ้นอยู่กับวัตถุประสงค์เงื่อนไขการกำหนด เป็นการแบ่งกลุ่มข้อมูลจากจุดข้อมูล ที่กระจุกตัวใกล้กับจุด Centroid ที่สุด

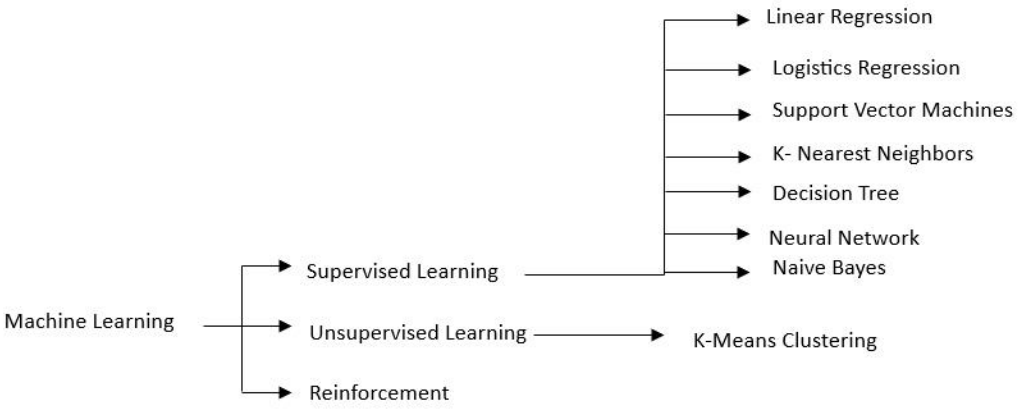
2. Hierarchical clustering คือ การจัดกลุ่มแบบลำดับชั้นหรือที่เรียกว่าการจัดกลุ่มตามการเชื่อมต่อนั้นขึ้นอยู่กับหลักการที่ว่าทุกวัตถุเชื่อมต่อกับระดับของความสัมพันธ์ เหมาะสำหรับข้อมูลที่มีลำดับชั้นหรือระยะห่างที่ใกล้เคียงกัน เช่น อนุกรมวิธาน การแบ่งกลุ่มลักษณะนี้มี 2 ประเภท คือ ล่างขึ้นบน Agglomerative และบนลงล่าง Divisive)

3. Density-based clustering คือ วิธีการจัดกลุ่มตามความหนาแน่นจะพิจารณาความหนาแน่นก่อนระยะทาง ของข้อมูลเกิดจากการกระจุกตัวของ Data point ที่เกาะกันอย่างหนาแน่น การกระจุกตัวที่เกิดจะมีรูปร่างและขนาดแตกต่างกัน และมีระดับความความหนาแน่นใกล้เคียงกัน วิธีการจัดกลุ่มนี้รวมถึงสัญญาณรบกวนและค่าผิดปกติในชุดข้อมูล ข้อมูลจะถูกแบ่งออกจากกันโดยพื้นที่ว่างของการกระจายตัว

4. Distribution-based clustering คือ เทคนิคการจัดกลุ่มที่เรารู้จักนั้นขึ้นอยู่กับความใกล้ชิด ความเหมือน ระยะทาง หรือองค์ประกอบ ความหนาแน่น การจัดกลุ่มตามการแจกแจงจะสร้างและจัดกลุ่มจุดข้อมูลตามความเป็นไปได้ของการแจกแจงความน่าจะเป็นแบบเดียวกัน

5. Fuzzy clustering คือ อัลกอริทึมทำงานโดยการกำหนดค่าสมาชิกให้กับจุดข้อมูลทั้งหมดที่เชื่อมโยงกับคลัสเตอร์ คำนวณจากระยะห่างระหว่างคลัสเตอร์และจุดข้อมูล หากค่าสมาชิกอยู่ใกล้กับคลัสเตอร์มากขึ้น ก็มีโอกาสสูงที่จะอยู่ในคลัสเตอร์นั้น

6. Constraint-based หรือ Supervised clustering คือ ข้อจำกัดเงื่อนไขถูกกำหนดให้เป็นคุณสมบัติที่ต้องการของผลลัพธ์ของการจัดกลุ่ม ซึ่งอาจอยู่ใน จำนวนคลัสเตอร์ ขนาดคลัสเตอร์ หรือ ตัวแปร ที่จำเป็นสำหรับการจัดคลัสเตอร์

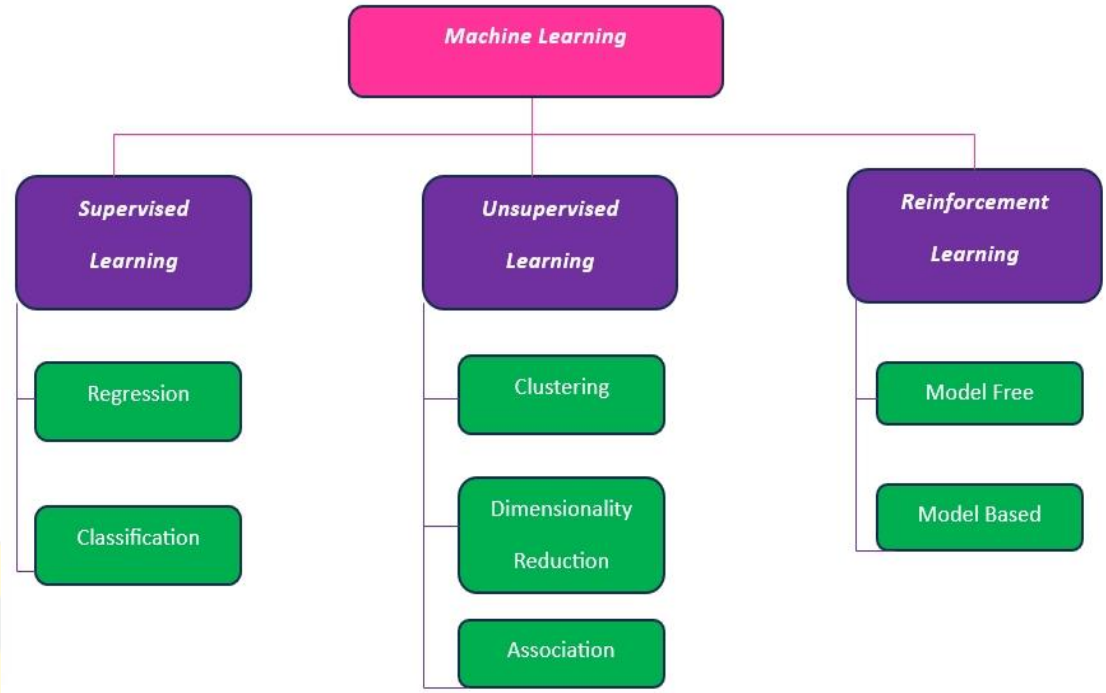


ภาพที่ 5 การแบ่งกลุ่มโมเดลของ Machine learning (บัญชา ปะสีละเตสัง, 2564)

7. Supervised learning เป็น Machine learning ที่ใช้ในการพยากรณ์ลักษณะเป้าหมาย โดยอาศัยข้อมูลที่เป็น Labeled ข้อมูลที่มีการระบุไว้ชัดเจนว่าข้อมูลนั้นมีลักษณะเป็นอย่างไรหรือเท่าไร เพื่อให้คอมพิวเตอร์ใช้ในการเรียนรู้และพยากรณ์ข้อมูลชุดอื่นว่าจะมีลักษณะเป้าหมายผลลัพธ์เป็นอย่างไร และยังสามารถแบ่งออกเป็นประเภทย่อยได้อีกมี Regression ใช้ในการพยากรณ์ลักษณะที่เป็นค่าตัวเลข เช่น การพยากรณ์ราคาบ้าน การพยากรณ์ราคาหุ้น เป็นต้น และ Classification ใช้ในการจำแนกประเภทลักษณะเป้าหมาย ตัวอย่างเช่น การพยากรณ์การป่วยหรือไม่ป่วย การจำแนกระหว่างส้มกับแตงโมง เป็นต้น

8. Unsupervised learning เป็น Machine learning ที่ใช้ข้อมูล Unlabeled data คือข้อมูลที่ไม่มีการระบุลักษณะเป้าหมาย ใช้ข้อมูลเหล่านั้นมาสร้างกฎหรือรูปแบบ เพื่อใช้ในการจัดกลุ่ม สามารถแบ่งออกได้ คือ Clustering ใช้ในการจัดกลุ่ม, Dimensionality reduction ใช้ในการลดลิมิตของข้อมูล และ Association ใช้ในการสร้างระบบแนะนำจากความเชื่อมโยงข้อมูล

9. Reinforcement learning เป็น Machine learning ที่ใช้ในการตัดสินใจที่ดีที่สุดเพื่อให้บรรลุเป้าหมายที่ตั้งไว้ ภายใต้เงื่อนไขที่กำหนด ผ่านการลองผิดลองถูก (นันทนัช ฟูสามป๊อก และคณะ, 2566)



ภาพที่ 6 แผนภาพประเภทของ Machine learning โดยละเอียด (นันทนัช ฟูสามป๊อก และคณะ, 2566)

เทคนิค Clustering มีเยอะพอสมควรที่สามารถนำมาแก้ไขปัญหาตามความเหมาะสม แต่ในงานวิจัยนี้มีปัญหาด้านการจัดกลุ่มลูกค้า จึงขออธิบายเฉพาะ Clustering ที่จะใช้ในงาน

10. K-nearest neighbors

เป็นวิธีการแบ่งคลาสสำหรับใช้จัดหมวดหมู่ข้อมูล ใช้หลักการเปรียบเทียบข้อมูลที่สนใจกับข้อมูลอื่นว่ามีความคล้ายคลึงมากน้อยเพียงใด หากข้อมูลที่กำลังสนใจนั้นอยู่ใกล้ข้อมูลใดมากที่สุด ระบบจะให้คำตอบเป็นเหมือนคำตอบของข้อมูลที่อยู่ใกล้ที่สุดนั้นลักษณะการทำงานแบบไม่ได้ใช้ข้อมูลชุดเรียนรู้ ในการสร้างแบบจำลองแต่จะใช้ข้อมูลนี้มาเป็นตัวแบบจำลองเลย (KongRuksiam, 2020) โดยจำเป็นจะต้องอาศัยข้อมูลที่เราแบ่งไว้ก่อนแล้ว ทำให้รู้ประเภทข้อมูลของสิ่งของที่เรายังไม่เคยแบ่งมาก่อน ในการใช้งานจริงยังมีเงื่อนไขบางอย่างที่เราต้องพิจารณาเป็นพิเศษ เช่น การกำจัดข้อมูลที่ไม่จำเป็นหรือข้อมูลที่มีแล้วจะทำให้ผลลัพธ์ผิดพลาด ใช้วิธีการคำนวณด้วยหลักการคณิตศาสตร์ระดับพื้นฐานได้ทำการเปรียบเทียบระยะห่างของตำแหน่งข้อมูล หลักการคำนวณระยะห่างและความเหมือนของ โมเดลเคเนียร์เรสเนเบอร์ โดยการวัดระยะห่าง 2 จุดบนแกน X และ Y สามารถใช้สมการยุคลิด Euclidean (บัญชา ปะสีละเตสัง, 2564)

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (yi - xi)^2}$$

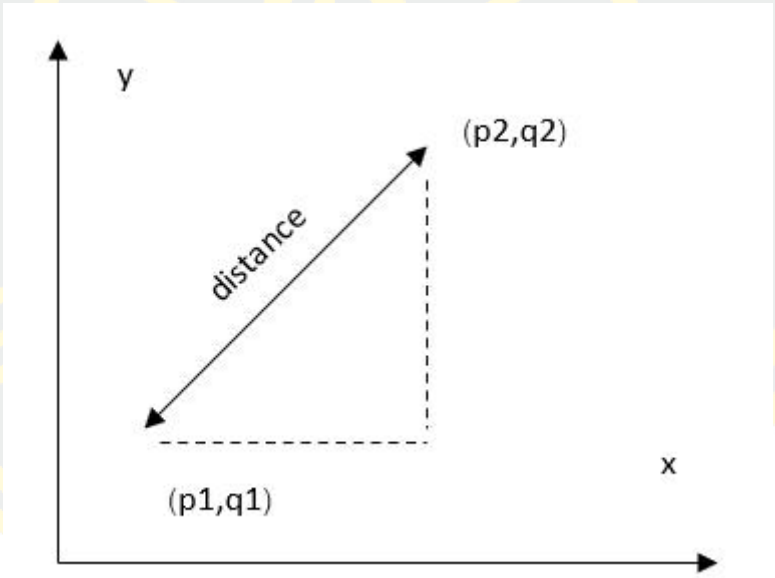
$$\text{distance} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

$$\text{distance} = \sqrt{(p^1 - p^2)^2 + (q_1 - q_2)^2}$$

โดยให้

x_1, y_1 คือ พิกัดจุด P

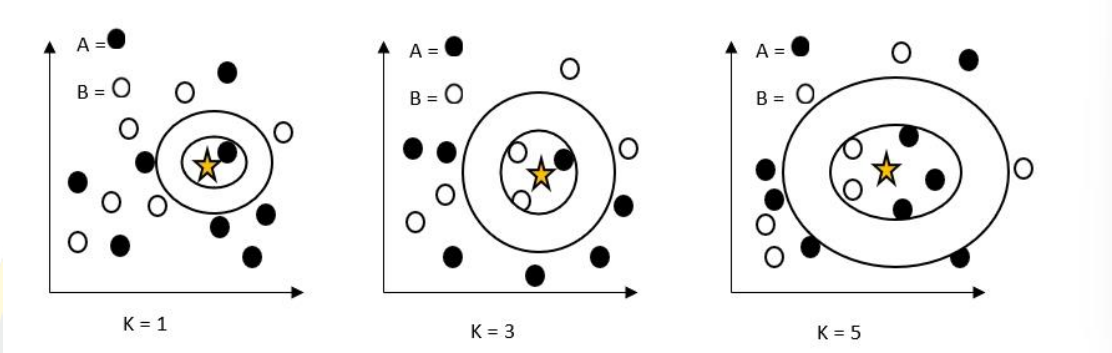
x_2, y_2 คือ พิกัดจุด q



ภาพที่ 7 วิธีคิดการวัดระยะห่างโดยใช้สมการยุคลิด โมเดลเคเนียร์เรสเนเบอร์
(บัญชา ปะสีละเตสัง, 2564)

ตามหลักการของโมเดลเคเนียร์เรสเนเบอร์ นั้น ตัว K หมายถึงจำนวนตำแหน่งข้างเคียงซึ่งอยู่ใกล้ตำแหน่งที่เราทำนายผลมากที่สุด โดยลักษณะตามที่ได้กล่าวมาทั้งหมด เราเลือกตำแหน่งข้อมูลที่อยู่ใกล้ที่สุดเพียงอันเดียวหรือเป็นกรณีที่ K=1 นั่นเอง แต่การกำหนดค่า K เป็น 1 อาจไม่ใช่ผลลัพธ์ที่เหมาะสมที่สุดของโมเดลนี้ ดังนั้น เราจำเป็นต้องเรียนรู้หลักการกำหนดค่า K เป็นตัวเลข

อื่น ๆ ซึ่งอาจให้ผลการทำนายที่ดีกว่า โดยหลักการก็คือ ให้พิจารณาจำนวนตำแหน่งข้างเคียงตามค่า K ที่ระบุ เช่น 3 ตำแหน่ง หรือ 5 ตำแหน่ง หรือจะมากกว่านี้ก็ได้ แล้วนับจำนวนเปรียบเทียบว่ามีคลาสใดมากที่สุด ก็ถือว่าผลการทำนายตรงกับคลาสนั้น เช่น ในภาพถัดไปเป็นกรณีที่ K = 1, 3 และ 5



ภาพที่ 8 K1, K3, K5 (บัญชา ปะสีละเตสัง, 2564)

ขั้นตอนการใช้งานโมเดลเคเนียร์เรสเนเบอร์มีดังนี้

การเตรียมข้อมูล ก่อนที่จะนำข้อมูลไปใช้ในโมเดล ต้องทำข้อมูลให้เหมาะกับการใช้งาน โดยการแยกข้อมูลออกเป็น Training set และ Test set โดย Training set จะถูกนำไปใช้ในการเทรนโมเดลเคเนียร์เรสเนเบอร์ และ Test set จะถูกนำไปใช้ในการทดสอบและประเมินประสิทธิภาพของโมเดลเคเนียร์เรสเนเบอร์

การกำหนดพารามิเตอร์ ซึ่งสามารถกำหนดได้ เช่น จำนวน K หรือจำนวนตัวอย่างที่ใช้ในการทำนาย รวมถึงการกำหนดวิธีการคำนวณระยะทางหรือความเหมือนของข้อมูลตัวอย่าง

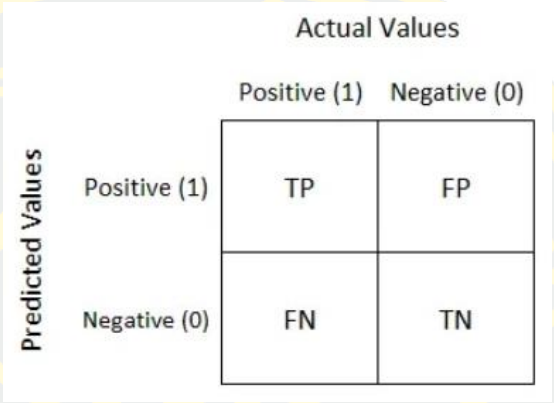
การเทรนโมเดล จะทำการค้นหาข้อมูลตัวอย่างที่ใกล้ที่สุดกับข้อมูลที่ต้องการจัดกลุ่ม โดยการค้นหานั้นจะใช้วิธีการคำนวณระยะทางและความคล้ายคลึงระหว่างตัวอย่างที่มีอยู่ใน Training set กับตัวอย่างที่ต้องการจัดกลุ่ม

การทำนาย เมื่อได้เทรนโมเดลและนำข้อมูล Test มาใช้ในการทดสอบโมเดลเคเนียร์เรสเนเบอร์ จะหา K ที่ใกล้ที่สุดกับข้อมูล test และจะให้ข้อมูล Test อยู่ในกลุ่มเดียวกับตัวอย่างที่มีจำนวนมากที่สุด

การประเมินผล เมื่อได้ทำการทดสอบโมเดลแล้วจะต้องทำ การประเมินผลโมเดลว่ามีประสิทธิภาพมากน้อยแค่ไหน (Kittimsakdi nicit, 2023)

การวัดประสิทธิภาพโมเดลเคเนียร์เรสเนเบอร์

สามารถทำได้หลายแบบ เช่น การใช้ Confusion matrix เพื่อวัดประสิทธิภาพของโมเดลในการทำนายกลุ่ม หรือการคำนวณค่า Accuracy, Precision, Recall, F1 score เป็นต้น Confusion matrix คือ ตารางสำคัญในการวัดความสามารถของ Machine learning ในการแก้ปัญหา



ภาพที่ 9 Confusion Matrix (Chengz, 2019)

โดยให้

TP คือ True positive การที่ทำนายว่า จริง แต่มีค่าเป็นจริง

TN คือ True negative การที่ทำนายว่า ไม่จริง แต่มีค่าไม่จริง

FP คือ False positive การที่ทำนายว่า จริง แต่มีค่าเป็นไม่จริง

FN คือ False negative การที่ทำนายว่า ไม่จริง แต่มีค่าเป็นจริง

โดยทั่วไปแล้วจะมีตัววัดที่นิยมใช้กันในงานวิจัยและการทำงาน และสมการ คือ (นันทนัช ฟูสามป๊อก และคณะ, 2566)

1) Precision เป็น สัดส่วนการวัดความแม่นยำของข้อมูลที่เป็น Positive หรือเชิงบวกโดยพิจารณาแยกทีละคลาส สามารถเขียนสมการได้ดังนี้

$$Precision = \frac{TP}{TP + FP}$$

2) Recall เป็นสัดส่วนการวัดความถูกต้องของการพยากรณ์ข้อมูลที่เป็น True positive และ False negative โดยพิจารณาแยกทีละคลาส สามารถเขียนสมการได้ดังนี้

$$Recall = \frac{TP}{TP + FN}$$

3) Accuracy เป็นสัดส่วนการวัดความถูกต้องของการพยากรณ์ข้อมูลที่เป็น True positive และ True negative โดยรวมทุกคลาส สามารถเขียนสมการได้ดังนี้

$$Accuracy = \frac{TP + TN}{n}$$

4) F1-Score คือ ค่าเฉลี่ยแบบ Harmonic mean หรืออัตราส่วนระหว่างจำนวนข้อมูลทั้งหมดกับผลรวมส่วนกลับของข้อมูลแต่ละตัว ระหว่าง Precision และ Recall สร้าง F1 สามารถเขียนสมการได้ดังนี้

$$F1 = 2 * \frac{Precision \ * \ Recall}{Precision \ + \ Recall}$$

11. K-means clustering

เป็นเทคนิคการจัดกลุ่มแบบ Partitioning methods เป็นการเรียนรู้โดยไม่มีผู้สอน การแก้ปัญหาการจัดกลุ่มจะใช้หลักการแบ่งข้อมูลออกเป็นกลุ่ม แล้วหาจุด Centroid ของแต่ละกลุ่มซึ่งจะพิจารณาก่อนว่าจะแบ่งข้อมูลออกเป็นกี่กลุ่ม หรือการกำหนดค่า K นั่นเองและทำการเลือกข้อมูลตำแหน่งใด ก็ได้มาเป็นจุด Centroid เริ่มแรกตามค่า ใช้วิธีคำนวณหาระยะห่างระหว่างจุดแล้วเปรียบเทียบตามแนวทางเดิม โดยคำนวณได้จากการคำนวณหาระยะทางระหว่างจุด เราจะใช้สมการยุคลิด Euclidean Equation (บัญชา ปะสีละเตสัง, 2564)

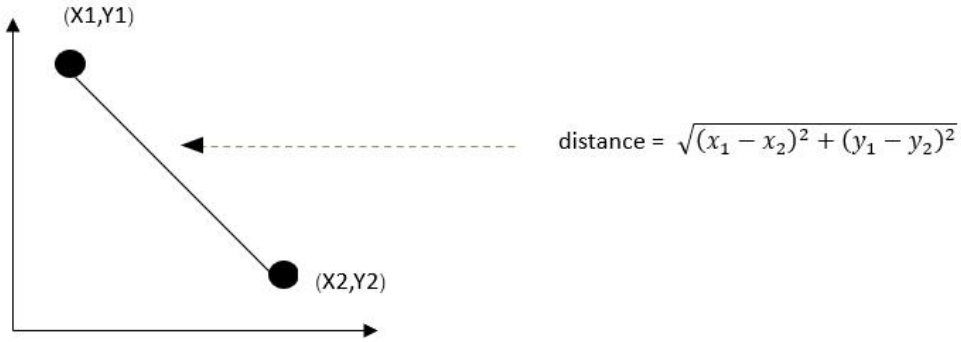
$$d(x, y) = \sqrt{\sum_{i=1}^{n} (yi - xi)^2}$$

$$distance = \sqrt{\left(x_1 - x_2\right)^2 + \left(y_1 - y_2\right)^2}$$

โดยให้

x_1, y_1 คือ พิกัดจุด y

x_2, y_2 คือ พิกัดจุด x



ภาพที่ 10 วิธีคิดการวัดระยะห่างโดยใช้สมการยุคลิด โมเดลคลัสเตอร์แบบเคมีน (บัญชา ปะสีละเตสัง, 2564)

โมเดลคลัสเตอร์แบบเคมีนจะใช้หลักการแบ่งข้อมูลออกเป็นกลุ่มแล้วหาจุด Centroid ของแต่ละกลุ่ม ซึ่งวิธีขั้นตอนโดยสังเขปคือ

พิจารณาก่อนว่าจะแบ่งข้อมูลออกเป็นกี่กลุ่ม หรือการกำหนดค่า K นั่นเอง

เลือกข้อมูลตำแหน่งใด ๆ ก็ได้มาเป็นจุด Centroid เริ่มแรกตามค่า K เช่น หากกำหนด K = 2 ก็ต้องเลือก 2 จุดมาเป็น Centroid

หาระยะห่างจาก Centroid แต่ละจุดไปยังตำแหน่งข้อมูล แล้วเทียบระยะห่างว่าจุดนั้นอยู่ใกล้ Centroid ใดมากที่สุด ก็จัดตำแหน่งข้อมูลนั้นให้อยู่ในกลุ่มเดียวกับ Centroid ดังกล่าว

ทำเช่นเดียวกับขั้นตอน 3 จนครบทุกตำแหน่ง ก็จะเป็นการจัดกลุ่มข้อมูลรอบแรก

คำนวณหาค่าเฉลี่ยของข้อมูลในแต่ละกลุ่ม ซึงจะได้ผลลัพธ์เป็นจุด Centroid ใหม่

เทียบตำแหน่ง Centroid ใหม่ กับตำแหน่ง Centroid เดิม หากตำแหน่งเปลี่ยนแปลงไปจากเดิมแสดงว่าเราต้องย้อนกลับไปจัดกลุ่มข้อมูลใหม่ โดยเทียบระยะห่างจากจุด Centroid ใหม่กับข้อมูลแต่ละตำแหน่ง หรือตามขั้นตอน 3 เป็นต้นมา

จากนั้นก็หาจุด Centroid ใหม่เทียบกับ Centroid เดิม ถ้าเปลี่ยนแปลงก็ย้อนกลับไปทำตั้งแต่ข้อ 3 ใหม่เป็นเช่นนี้ไปเรื่อยเรื่อย จนกว่า Centroid ใหม่จะเป็นตำแหน่งเดียวกันกับ Centroid เดิม จึงจะถือว่าการจัดกลุ่มล่าสุดคือการแบ่งกลุ่มที่ถูกต้อง และค่าเฉลี่ยล่าสุด Centroid ที่แท้จริงของแต่ละกลุ่ม

การวัดประสิทธิภาพของโมเดลคลัสเตอร์แบบเคมีนมีดังนี้

Elbow method เป็นวิธีสำหรับการหาค่า K ที่เหมาะสมที่สุดในการจัดกลุ่มแบบเคมีน กราฟจะแสดงค่าภายในคลัสเตอร์ผลรวมของกำลังสอง บนแกน y ที่สอดคล้องกับค่า ของ K บนแกน x ค่า คือเมื่อค่าที่ได้จากการคำนวณมีความผิดพลาดน้อยลงความชันของเส้นโค้งจะเริ่มเรียบ

และจะเกิดเป็นมุมที่คล้ายข้อศอก จะเป็นจุดที่แสดงค่า K ที่เหมาะสมที่สุด โดยใช้วิธีการคำนวณวัดค่า Sum of Square Error ของผลรวมระยะห่าง ระหว่างวัตถุประสงค์ กับ Centroid

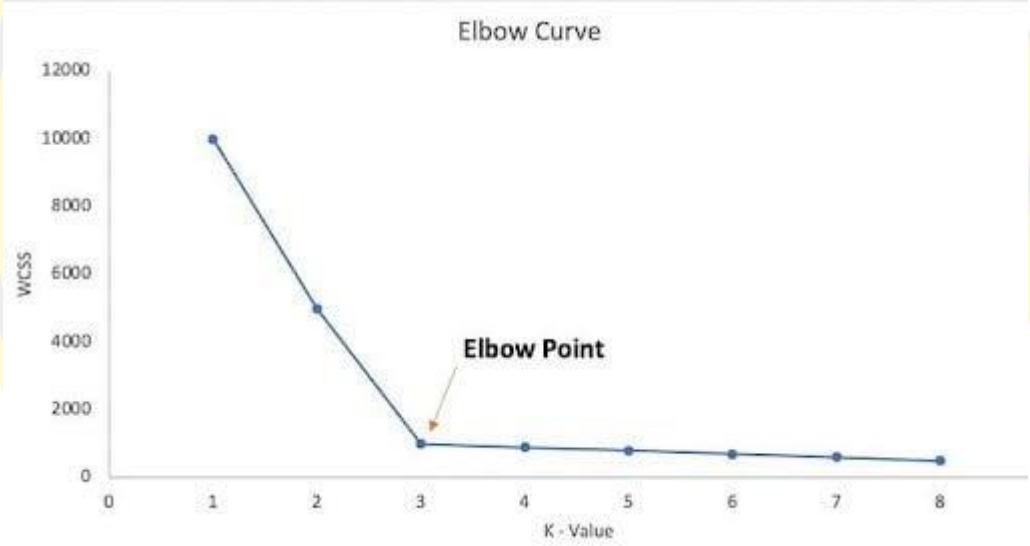
$$WCSS = \sum_{i=1}^{k} \sum_{j=1}^{n} \text{distance}\left(x_{ij} - \mu_i \right)^2$$

โดยให้

WCSS = คือผลรวมของระยะห่างระหว่างจุดข้อมูลกับจุดศูนย์กลางของกลุ่ม

k = จำนวนของกลุ่มที่เรากำลังทดลอง

n= จำนวนของข้อมูล



ภาพที่ 11 ระดับความแม่นยำมีแนวโน้มจะขึ้นไปสูงสุดเมื่อ K = 3 (Tomar, 2023)

Silhouette score เป็นตัววัดคุณภาพของการแบ่งกลุ่ม โดยวัดความคลาดเคลื่อนภายในกลุ่มและความคลาดเคลื่อนระหว่างกลุ่มผลลัพธ์ได้แสดงการวัดว่าแต่ละจุดในกลุ่มหนึ่งอยู่ใกล้กันมากเท่าใดกับจุดทั้งหมดในกลุ่มใกล้เคียง เมื่อเทียบกับกลุ่มอื่น ค่าของ Silhouette อยู่ในช่วง มีค่าใกล้กับ 1 แสดงว่าเป็นการหากลุ่มที่ดี Silhouette score อยู่ในช่วง -1, 1 โดยคะแนนสูงแสดงถึงการแบ่งกลุ่มที่ดี คะแนน 0 แสดงถึงการแบ่งกลุ่มที่ทับซ้อนกัน และคะแนนต่ำกว่า 0 แสดงถึงการแบ่งกลุ่มที่ไม่เหมาะสม สูตรคำนวณค่า Silhouette score (Boonsoong, 2020)

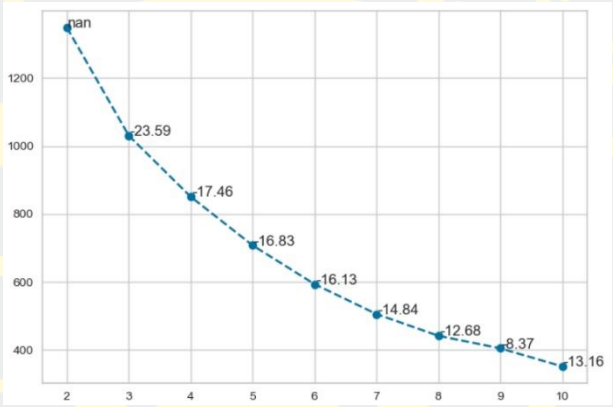
$$S_i = \frac{b_i - a_i}{max(a_i, b_i)}$$

โดยให้

S_i = Silhouette score ของกลุ่ม i

a_i = ค่าเฉลี่ยของระยะห่างระหว่างจุด i กับจุดในกลุ่มเดียวกัน

b_i = ค่าเฉลี่ยของระยะห่างระหว่างจุด i กับจุดในกลุ่มเดียวกัน



ภาพที่ 12 การหาจำนวนกลุ่มจากค่า Silhouette score (Kumar, 2021)

แนวคิดทฤษฎี Feature engineering

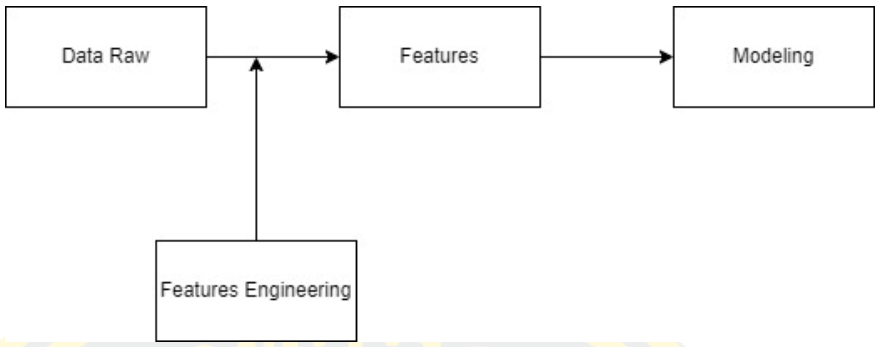
Feature engineering คือ เป็นกระบวนการแปลงคุณสมบัติของข้อมูลเก่าที่ยังไม่มีประสิทธิภาพ ไปเป็นข้อมูลใหม่ที่มีประสิทธิภาพในการพัฒนาโมเดลให้มีประสิทธิภาพมากกว่าการใช้ข้อมูลเก่า ตัวอย่างการทำ Feature engineering มีดังนี้

Scaling คือ เป็นการตรวจสอบ Features ให้มีมาตราส่วนที่เปรียบเทียบกันได้

Encoding คือ การแปลงข้อมูลที่ไม่ใช่ตัวเลขให้กลายเป็นรูปแบบของตัวเลข

Dimensionality reduction คือ ลดจำนวน Features โดยไม่สูญเสียข้อมูลสำคัญ ปรับปรุงประสิทธิภาพของโมเดล

การสร้าง Feature ใหม่ คือ การรวมหรือใช้การดำเนินการทางคณิตศาสตร์ กับ Features เดิมที่มีอยู่ (Narut Soontranon, 2024)



ภาพที่ 13 กระบวนการ Features engineering (Zhu, 2018)

แนวคิดทฤษฎี Data mining

Data mining หรือเหมืองข้อมูลเป็นวิธีการที่ใช้ในการค้นหาความรู้ (Knowledge) จากข้อมูลขนาดใหญ่ โดยกระบวนการนี้ทำการวิเคราะห์ข้อมูลจำนวนมากเพื่อค้นหารูปแบบ (Pattern) และความสัมพันธ์ (Association) ที่ซ่อนอยู่ในชุดข้อมูล กระบวนการทำงานของเหมืองข้อมูลได้รับการพัฒนาขึ้นในปี ค.ศ. 1996 โดยแบ่งออกเป็น 6 ขั้นตอนที่เรียกว่า Cross-Industry Standard Process for Data Mining (CRISP-DM) ดังนี้

1. การทำความเข้าใจธุรกิจและวิเคราะห์ปัญหา (Business understanding) ขั้นตอนนี้เกี่ยวกับการทำความเข้าใจปัญหาและเปลี่ยนแปลงปัญหานั้นให้อยู่ในรูปแบบโจทย์ของการวิเคราะห์ข้อมูล พร้อมทั้งวางแผนการดำเนินงาน

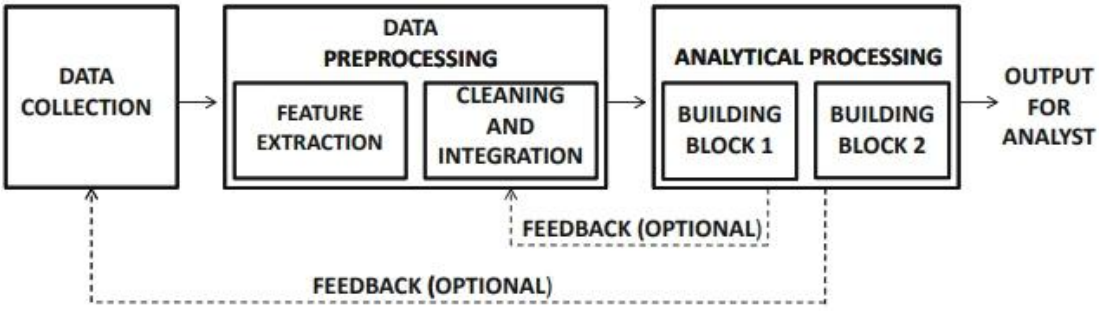
2. การทำความเข้าใจข้อมูล (Data understanding) เริ่มต้นจากการรวบรวมข้อมูลและตรวจสอบความถูกต้องของข้อมูลที่รวบรวมมา

3. การเตรียมข้อมูล (Data preparation) เป็นการแปลงข้อมูลดิบให้เป็นข้อมูลที่พร้อมสำหรับการวิเคราะห์ โดยทำการทำความสะอาดข้อมูล (Data cleaning)

4. การสร้างโมเดล (Modeling) วิเคราะห์ข้อมูลโดยใช้เทคนิคทางเหมืองข้อมูล เช่น การจำแนกประเภทข้อมูลหรือการแบ่งกลุ่มข้อมูล ขั้นตอนนี้อาจต้องย้อนกลับไปที่ขั้นตอนการเตรียมข้อมูลเพื่อปรับเปลี่ยนข้อมูลให้เหมาะสมกับแต่ละเทคนิค

5. การประเมินโมเดล (Model evaluation) ตรวจสอบผลการวิเคราะห์ข้อมูลเพื่อให้แน่ใจว่าตรงตามวัตถุประสงค์ที่ตั้งไว้ในขั้นตอนแรก และประเมินความน่าเชื่อถือของผลลัพธ์

6. การนำไปใช้ (Deployment) นำความรู้ที่ได้จากการวิเคราะห์ไปใช้ในทางปฏิบัติ (Aggarwal, 2014)



ภาพที่ 14 กระบวนการทำงาน Data Mining (Aggarwal, 2014)

งานวิจัยที่เกี่ยวข้อง

Novianti, Setyorini, and Rafflesia (2017) ได้ทำการวิเคราะห์ K-Means cluster ในการจัดกลุ่มจุดศูนย์กลางแผ่นดินไหว จังหวัดเบงกูลู ประเทศอินโดนีเซีย เป็นพื้นที่ที่มีแผ่นดินไหวสูงจึงต้องศึกษาลักษณะแผ่นดินไหวแต่ละครั้งในจังหวัดเบงกูลู ซึ่งจะใช้ข้อมูลจังหวัดเบงกูลูและพื้นที่โดยรอบตั้งแต่เดือนมกราคม พ.ศ. 2513 ถึงเดือนธันวาคม พ.ศ. 2558 โดยได้นำข้อมูลมาจากกรมอุตุนิยมวิทยาภูมิอากาศและธรณีฟิสิกส์ การจัดกลุ่มโดยใช้วิธีระยะทางแบบยุคลิดใช้ในการวิเคราะห์นี้ ตัวแปรคือละติจูดลองจิจูดและขนาด จำนวนคลัสเตอร์ที่เหมาะสมซึ่งเป็น 7 การวิเคราะห์ จะดำเนินการในการจัดกลุ่มศูนย์กลางแผ่นดินไหว ขั้นตอนที่ 1 กำหนดข้อมูลที่จะจัดกลุ่ม ขั้นตอนที่ 2 ใช้การวิเคราะห์ K-Means cluster กับข้อมูลแผ่นดินไหว ขั้นตอนที่ 3 คำนวณเกณฑ์สำหรับจำนวน K ที่เหมาะสมของ K-Means cluster ขั้นตอนที่ 4 ศึกษาแต่ละคลัสเตอร์ที่ส่งผลให้เกิดเป้าหมายนี้ สรุปข้อมูลที่ใช้ในการศึกษานี้มี 968 เหตุการณ์ มีการนำเสนอการแพร่กระจายของศูนย์กลางดังนี้

K-mean cluster 1 เป็นคลัสเตอร์ที่เกิดแผ่นดินไหวรอบ เซลูมา, เบงกูลูใต้ และคอร์ ทั้งในนอกชายฝั่งและในแผ่นดินใหญ่ มีแผ่นดินไหว 85.72% ได้แก่ แผ่นดินไหวตื้น และ 14.28% ของแผ่นดินไหวเป็นแผ่นดินไหวขนาดกลางที่มักเกิดขึ้นบนแผ่นดินใหญ่

K-mean cluster 2 ตั้งอยู่รอบ มูโกมูโกะ, เบงกูลูเหนือ และเรจัง เลบง ขนาดแผ่นดินไหวสูงสุดที่เกิดขึ้น คือแผ่นดินไหวปี 2544 ขนาด 7.4 ริคเตอร์

K-mean cluster 3 เกิดแผ่นดินไหวนอกชายฝั่งของเมืองเบงกูลู เป็นแผ่นดินไหวตื้น เกิดแผ่นดินไหวขนาด 7 ริคเตอร์, 7.3 ริคเตอร์ และ 7.9 ริคเตอร์

K-mean cluster 4 เกิดแผ่นดินไหวในทะเลทางตะวันตกของจังหวัดลัมปุง เหตุการณ์แผ่นดินไหว ส่วนใหญ่เกิดขึ้นในทะเล แผ่นดินไหวที่ใหญ่ที่สุดที่มีขนาด 7 ริคเตอร์ ของปี 2537

K-mean cluster 5 ประกอบด้วยเหตุการณ์แผ่นดินไหวที่เกิดขึ้นรอบหมู่เกาะเมนตาไว ของเกาะสุมาตราตะวันตกโดยเฉพาะบนเกาะซิโปรา เป็นการรวมกันของแผ่นดินไหวในทะเล และบนบก

K-mean cluster 6 เป็นกระจุกที่มีแผ่นดินไหวเกิดขึ้นที่ทะเลของมหาสมุทรอินเดีย ร้อยละ 92.67 เป็นแผ่นดินไหว 5 ริคเตอร์, 6 ริคเตอร์ ส่วนที่เหลือร้อยละ 7.3 คือ 6 ริคเตอร์, 7 ริคเตอร์ มี ร้อยละ 99.33 ของแผ่นดินไหวตื้น และแผ่นดินไหวสูงสุดที่ 6.5 ริคเตอร์

K-mean cluster 7 เป็นแผ่นดินไหวขนาดใหญ่ที่วัดได้ 7.7 ริคเตอร์ รวมถึงขนาด 7.2 ริคเตอร์ เป็นแผ่นดินไหวตื้น

สุชาวดี ปลั่งศรี (2562) ได้ทำการวิจัยเรื่องการจัดกลุ่มขนาดผลิตภัณฑ์โดยการใช้ K-means clustering เพื่อลดต้นทุนบรรจุภัณฑ์ มีโรงงานกรณีศึกษาเป็นโรงงานผลิตผ้าเบรก ซึ่งมีกล่องหลายขนาดที่ใช้ในหีบห่อ จึงทำให้เกิดผลเสียกับโรงงานกรณีศึกษามี 3 อย่าง คือ ต้นทุนหีบห่อ พื้นที่ในการสต็อกสินค้า รอบจำนวนการขนส่งที่ต่ำ เป็นต้น เนื่องจากไม่มีกระบวนการทำงานที่เป็นมาตราฐานโดยที่ไม่รู้ถึงน้ำหนักของหีบห่อและการวางซ้อนที่ชัดเจนเป็นมาตราฐาน ซึ่งได้มีการเน้น 2 ขั้นตอนในการดำเนินงานวิจัย โดยการใช้ K-means clustering ในการจัดลุ่ม เพื่อแบ่งขนาดการบรรจุหีบห่อ มีสถิติข้อมูลจำนวน 3 แบบ มีจำนวน 437 ข้อมูลโดยหาค่าน้ำหนักของหีบห่อ โดยทดสอบการวางบน Pallet ซึ่งมี 5 ขั้นตอนที่ดีและมีประสิทธิภาพ โดยผลลัพธ์ซึ่ง ลดต้นทุนได้ 2.9 ล้านบาท คิดเป็น 13.43% ลดพื้นที่จัดเก็บคงคลังได้ 2.19 ล้านลูกบาศก์นิ้ว คิดเป็น 48.65% เพิ่มปริมาณการ ขนส่งได้ 5,200 กล่อง/ รอบ คิดเป็น 86.7%

วณิชา แผลงรักษา และนิเวศ จิระวิชิตชัย (2562) ได้ทำการศึกษาเรื่องการแบ่งกลุ่มลูกค้าโดยใช้เทคนิคการจัดกลุ่ม K-means สำหรับการจัดการความสัมพันธ์กับลูกค้า ได้ใช้สถิติข้อมูลของกลุ่มผู้ซื้อสินค้าขายออนไลน์แห่งหนึ่ง ด้วยวิธีการดูออเดอร์การสั่งของผู้ซื้อมีจำนวน 1,000 คน โดยมีการดูจาก สถานะ เพศ รายได้ต่อเดือน บุตร การศึกษา อาชีพ การเดินทาง ที่อยู่ รถยนต์ส่วนบุคคล วิเคราะห์แบบลักษณะประชากรศาสตร์ และนำมาวิเคราะห์แบบ K-means clustering ต่อ จากการทดลองผลลัพธ์คือ มีการจำแนกได้ 7 K-means

K-mean 1 คือ รายได้น้อยกว่า 39,050 เหรียญ อาศัยอยู่ที่ยุโรป

K-mean 2 คือ ผู้ไม่มีรถยนต์ส่วนบุคคล มีการศึกษา รายได้ ระหว่าง 39,050 ไม่เกิน 71,062 เหรียญ

K-mean 3 คือ อายุน้อยกว่า 37 ไม่มีบุตร มีศึกษา

K-mean 4 คือ มีรายได้ระหว่าง 39,050 ไม่เกิน 71,062 เหรียญ อาศัยอยู่อเมริกาเหนือ การเดินทาง 16 กิโลเมตร มีรถอายุ 55-65 ปี

K-mean 5 คือ อาชีพผู้จัดการ ผู้ชำนาญการ มีรถ มีรายได้ระหว่าง 97,111 ไม่เกิน 127,371 เหรียญ การเดินทาง 16 กิโลเมตร ที่อยู่อาศัยแปซิฟิก

K-mean 6 คือ อาชีพช่างเทคนิค อายุ 46-55 ปี มีรถ มีบุตร ที่อยู่อาศัยอเมริกาเหนือ

K-mean 7 คือ อาชีพผู้จัดการ มีรายได้ระหว่าง 97,111 ไม่เกิน 127,371 เหรียญ มีรถ

การวิเคราะห์ K-means clustering เป็นการจัดกลุ่มผู้ที่มีประเภทลักษณะที่ใกล้เคียงกัน นำมาอยู่ในกลุ่มร่วมกัน ทำให้เห็นภาพที่ชัดเจนขึ้น

พสธร (2563) ทำการศึกษาการประเมินราคาเสนอขายห้องชุดด้วย Deep Neural network และ K-Means clustering algorithm โดยทำการนำข้อมูลมาจาก ZmyHome ซึ่งมีข้อมูลจำนวน 11,062 ชุด ซึ่งหลังจากที่ได้เก็บรวบรวมและทำการนำข้อมูลที่ไม่สมบูรณ์จะเหลือข้อมูลจำนวน 6,106 ชุด วิจัยนี้ต้องการที่จะเปรียบเทียบผลลัพธ์ จากวิธี K-Means กับวิธีการประเมินด้วยระยะทาง จึงได้แบ่งจำนวนคลัสเตอร์เป็น 39 คลัสเตอร์ และ Deep neural network ปัจจัยที่นำมาศึกษาแบ่งเป็น ปัจจัยด้านตัวโครงการ ปัจจัยด้านทำเล และปัจจัยด้านสภาพแวดล้อม และได้ทำการเอาข้อมูลบาง ปัจจัยออก เช่น จำนวนห้องน้ำ ราคาเสนอขายรวม อายุของประกาศ ผลวิเคราะห์ผลลัพธ์ที่ได้จาก การศึกษาด้วยอัลกอริทึม Deep neural network โดยเปรียบเทียบระหว่างข้อมูลที่ได้จากการแบ่ง คลัสเตอร์ด้วย K-Means กับข้อมูลที่ใช้ข้อมูลของคอนโดในรัศมี 5 กิโลเมตรรอบโครงการ การทำนายราคาได้มีการเปรียบเทียบระหว่างการคัดเลือก คอนโดมิเนียมด้วย K-Means และ Deep neural network พบว่าทั้งสองวิธีไม่ได้มีความแม่นยำในการทำนายราคาเสนอขายแตกต่างกัน แต่ การทำนายราคาโดยใช้ K-Means ในการคัดเลือก คอนโดมิเนียมมีอัตราการทำนายราคาสำเร็จที่ มากกว่า จำนวนประกาศที่สามารถประเมิน ราคาเสนอขายได้สำเร็จ 5,717 ประกาศหรือร้อยละ 93.63 มากกว่า Deep Neural Network สามารถประเมินราคาเสนอขายได้ 5,165 ประกาศหรือ ร้อยละ 84.59

Wiyono, Wibowo, Hidayatullah, and Dairoh (2020) ได้เปรียบเทียบ K-nearest neighbors, Support vector machines และการตัดสินใจแบบ Decision tree สำหรับประสิทธิภาพของ นักเรียนแบบการทำนายผล การที่นำข้อมูลนักเรียนที่ไม่ได้เรียนต่อหรือยังเรียนไม่จบมารวมกับ นักศึกษาที่สำเร็จการศึกษาจะส่งผลต่อข้อมูลจำนวนของนักเรียนสำเร็จการศึกษา ซึ่งการป้องกัน ไม่ให้นำข้อมูลนักเรียนมารวมกัน สามารถทำได้โดยการนำข้อมูลของนักเรียนมาทำนายผลการเรียน การศึกษา ดำเนินการโดยใช้อัลกอริทึม K-nearest neighbors, Support vector machines และ Decision tree เพื่อให้ได้แบบจำลองการทำนายที่ดีที่สุด ขั้นตอนการดำเนินงานมีดังนี้ การเก็บ รวบรวมข้อมูล, การประมวลผลล่วงหน้า, การสร้างแบบจำลอง, การเปรียบเทียบแบบจำลอง, การประเมินผล ผลลัพธ์คือ Support ใช้ค่า C = 1 มีความแม่นยำที่ดีที่สุดในการทำนายด้วย

ค่าความแม่นยำร้อยละ 95 สามารถทำนายนักศึกษาที่สำเร็จการศึกษาอยู่ 311 คน และนักเรียนที่ไม่ได้เรียนต่อหรือยังเรียนไม่จบ 53

Decision ใช้ค่า cp = 0.6689113 มีความแม่นยำในการทำนายร้อยละ 93 นักศึกษาที่สำเร็จการศึกษาอยู่ 308 คน และนักเรียนที่ไม่ได้เรียนต่อหรือยังเรียนไม่จบ 48 คน

K-nearest สามารถทำนายประสิทธิภาพของนักเรียน ได้ดีด้วย k = 3 มีค่าความแม่นยำในการทำนาย 92% นักศึกษาที่สำเร็จการศึกษา 309 คน และนักเรียนที่ไม่ได้เรียนต่อหรือยังเรียนไม่จบ 52 คน Support vector machines มีความแตกต่างร้อยละ 4 จาก Decision tree และ K-nearest neighbors

กาญจนมาส เปลี่ยนสกุล (2564) ได้ทำการศึกษาการจัดกลุ่มลูกค้าบริษัทยานยนต์ด้วยข้อมูลประชากรโดยใช้เทคนิคการเรียนรู้ของ Machine learning งานวิจัยนี้ได้เลือกข้อมูลประชากรสาธารณะของลูกค้าภายในบริษัทยานยนต์แห่งหนึ่ง และนำมาจัดกลุ่มทั้งหมด 4 กลุ่ม โดยใช้อัลกอริทึม Logistic regression, Naïve bayes, Support Vector Machine (SVM), Random forest และ Extreme Gradient Boosting (XGBoost) และแก้ปัญหาข้อมูลไม่สมดุลด้วยวิธีการ SMOTE วัดประสิทธิภาพด้วยค่า Accuracy, Precision, Recall, F1-Score ดูค่าความถูกต้อง ด้วย Confusion Matrix ผลลัพธ์ที่ได้คือส่วน Logistics regression ร่วมกับการใช้ SMOTE ให้ Accuracy ร้อยละ 47.21 Precision ร้อยละ 46.66 Recall ร้อยละ 47.21 F1-Score ที่ร้อยละ 46.54 45.87 ส่วน Naïve Bayes ร่วมกับการใช้ SMOTE ให้ Accuracy ร้อยละ 46.10 Precision ร้อยละ Recall ร้อยละ 45.11 F1-Score ที่ร้อยละ 46.10 ส่วน SVM ร่วมกับการใช้ SMOTE ให้ Accuracy ร้อยละ 47.88 Precision ร้อยละ 47.20 Recall ร้อยละ 47.88 F1-Score ร้อยละ 46.91 และ Random forest ร่วมกับการใช้ SMOTE ให้ประสิทธิภาพที่ดีที่สุด ที่ค่า Accuracy ร้อยละ 48.75, Precision ร้อยละ 48.10, Recall ร้อยละ 48.75 และ F1-Score ที่ร้อยละ 48.31

ภัครพล อาจอาษา (2564) ได้ทำการศึกษาการวิเคราะห์คุณภาพน้ำด้วยเทคนิคการจัดกลุ่มข้อมูล ของเขื่อนลำปะทาว จังหวัดชัยภูมิ ได้เก็บข้อมูลด้วยทุ่นลอยน้ำที่ติดตั้งระบบเซนเซอร์ตรวจคุณภาพน้ำจำนวน 5 พารามิเตอร์ประกอบด้วยเซนเซอร์วัด ค่าออกซิเจนที่อยู่ในน้ำ ค่าอุณภูมิน้ำ ค่ากรดเบสของน้ำ ค่าความขุ่นของน้ำ และค่าการนำไฟฟ้าของน้ำ เก็บรวบรวมข้อมูลคุณภาพน้ำในช่วงเดือน มกราคม-มีนาคม พ.ศ. 2564 มีจำนวน 13,608 ชุดข้อมูล รอบบริเวณเขื่อนจำนวน 5 จุด ผลการหาจำนวนกลุ่มที่ดีที่สุดได้ 5 กลุ่ม โดยวิธีการ PCA+K-Means ซึ่งพิจารณาจากวิธีการ Elbow method และ Silhouette score ที่ได้ค่าอยู่ที่ 0.7015

นิภาพร ชนะมาร (2566) ได้จัดกลุ่มข้อมูลเศรษฐกิจครัวเรือนด้วยการประยุกต์ใช้ K-Means clustering โดยใช้ข้อมูลเศรษฐกิจครัวเรือนจากฐานข้อมูลตำบลและฐานข้อมูลขนาดใหญ่สำหรับการพัฒนาท้องถิ่นอย่างยั่งยืน ทำการเก็บรวบรวมข้อมูล ตั้งแต่ปี พ.ศ. 2563 ถึง 2564 จำนวน 12 หมู่บ้าน 2,909 ครัวเรือน ได้เก็บข้อมูลทั้งหมด 10 ส่วน มี 17,933 ระเบียน และ 178 ปัจจัย ผลการวิจัยการจัดกลุ่มเศรษฐกิจครัวเรือน จัดกลุ่มได้ 3 กลุ่ม โดยพิจารณาจากค่า Centroid ที่เหมาะสม ดังนี้ ครัวเรือนที่อยู่ในระดับเศรษฐกิจน้อย มีจำนวน 304 ครัวเรือน ร้อยละ 3.375 ครัวเรือนที่อยู่ในระดับเศรษฐกิจปานกลาง มีจำนวน 544 ครัวเรือน ร้อยละ 2.597 ครัวเรือนที่อยู่ในระดับเศรษฐกิจสูงมีจำนวน 903 ครัวเรือน ร้อยละ 2.355

วาทิตย์ คำพรมา เพชร อิ่มทองคำ และจักรชัย โสอินทร์ (2562) วิเคราะห์ความรู้สึกเกี่ยวกับความคิดเห็นต่อโรงแรมในประเทศไทย โดยได้ใช้เทคนิคการวิเคราะห์ความรู้สึกในรูปแบบเหมืองความคิดเห็น ต่อการรับบริการ ได้มีการรวบรวมความคิดเห็นจากเว็บไซต์ APT TUBE จำนวน 10,000 ประโยคแล้วจึงทำการวิเคราะห์ความคิดเห็น โดยปรากฏว่าเทคนิค K-means ร่วมกับ K-Nearest neighbors ให้ความถูกต้องสูงสุดที่ร้อยละ 94.8 รองลงมาคือเทคนิค Decision tree ร้อยละ 89.7, K-Nearest neighbors ร้อยละ 88.2 และเทคนิค Support vector machine ร้อยละ 87.5 โดยเทคนิคที่ให้ค่าความถูกต้องมากที่สุด คือ เทคนิค K-means ร่วมกับ K-Nearest neighbors ให้ค่าความถูกต้องสูงที่สุดที่ร้อยละ 94.8

สุริยัน เขตบรรจง (2564) ศึกษาผลการจัดกลุ่มข้อมูลโดยใช้เทคโนโลยีปัญญาประดิษฐ์ AI แบ่งกลุ่มผลสัมฤทธิ์ทางการเรียนของนักเรียนกลุ่มตัวอย่างเป็น กลุ่มเก่ง กลุ่มปานกลาง กลุ่มอ่อน กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ คือ นักเรียนชั้นประถมศึกษาปีที่ 5 ปีการศึกษา 2561 โรงเรียนสามเสนนอก สำนักงานเขตดินแดง กรุงเทพมหานคร จำนวน 33 คน ได้มาจากการสุ่มตัวอย่างด้วยวิธีจับสลาก เครื่องมือที่ใช้ในการวิจัยครั้งนี้ โดยใช้วิธีการเรียนรู้การแบ่งกลุ่มข้อมูล K-mean ผลการวิจัย พบว่า การแบ่งกลุ่มของนักเรียนได้ กลุ่มเรียนเก่ง 11 คน กลุ่มเรียนปานกลาง 11 คน กลุ่มเรียนอ่อน 11 คน และการแบ่งกลุ่มของรายวิชาเป็น กลุ่มเด่น 4 รายวิชา กลุ่มกลาง 3 รายวิชา กลุ่มด้อย 3 รายวิชา โดยมีรายวิชาที่เด่นสุดคือ วิชาประวัติศาสตร์ รองมาคือวิชาการงานอาชีพ และรายวิชาที่สุดท้ายคือ วิชาภาษาอังกฤษ

บทที่ 3
วิธีดำเนินงานวิจัย

การหาแนวทางการใช้ Machine learning ในการจัดกลุ่มลูกค้าเพื่อเพิ่มประสิทธิภาพในการตั้งราคาตอบสนองได้อย่างแม่นยำ กรณีศึกษาบริษัทขนส่ง จะมีวิธีการศึกษาโดยรวม 5 ส่วน มีการเก็บข้อมูลที่เกี่ยวข้องนำมาวิเคราะห์หาแนวทาง วิธีการ แก้ไขปรับปรุง ได้ดีขึ้นและเหมาะสม โดยจะมีขั้นตอนการศึกษาดังนี้

1. การเก็บรวบรวมข้อมูล
2. ขั้นตอนการดำเนินงานวิจัย
3. เครื่องมือที่ใช้ในการดำเนินงานวิจัย
4. การวิเคราะห์ข้อมูล
5. สรุปผลการดำเนินงาน

การเก็บรวบรวมข้อมูล

ในงานวิจัยนี้การเก็บรวบรวมข้อมูลได้ดำเนินการโดยใช้ข้อมูลทุติยภูมิ ซึ่งหมายถึงข้อมูลที่มีการบันทึกและรวบรวมไว้แล้วจากแหล่งข้อมูลที่มีอยู่เดิม โดยในการศึกษาครั้งนี้ ข้อมูลที่ใช้มาจากสถิติการใช้บริการของลูกค้าบริษัทกรณีศึกษาที่ถูกบันทึกอยู่ในสมุดบันทึกงานของบริษัท ข้อมูลที่รวบรวมมาใช้ในการวิจัยครอบคลุมช่วงเวลาตั้งแต่วันที่ 4 สิงหาคม พ.ศ. 2564 ถึง 31 สิงหาคม พ.ศ. 2566 โดยมีจำนวนข้อมูลลูกค้าทั้งหมด 4,270 รายการ จากจำนวนลูกค้าทั้งสิ้น 250 ราย ซึ่งข้อมูลเหล่านี้ประกอบไปด้วยรายละเอียดการใช้บริการของลูกค้าในช่วงเวลาที่ระบุไว้ นอกจากนี้การวิจัยยังได้ศึกษาขั้นตอนการทำงานของบริษัท รวมถึงการตั้งราคาสำหรับลูกค้าปัจจุบัน เพื่อวิเคราะห์ว่ามีการลดราคาหรือปรับเปลี่ยนราคาอย่างไร ราคาในปัจจุบันอยู่ที่เท่าไร และมีปัญหาอะไรที่เกิดขึ้นจากการตั้งราคา การวิเคราะห์นี้ทำโดยอ้างอิงจากสมุดบันทึกงานสถิติการใช้บริการของลูกค้าของบริษัทกรณีศึกษา ข้อมูลที่ได้จากบริษัทกรณีศึกษานี้ ถูกนำมาเปรียบเทียบเพื่อหาจุดแข็งและจุดอ่อน รวมถึงการวิเคราะห์หาวิธีการปรับปรุงเพื่อให้เกิดประสิทธิภาพสูงสุดในการดำเนินงานของบริษัท

Date	Pick-up location	Where to drop off	Vehicle ty	employing comp	drive	pricel	Cost	Sector to drop	distan	distd	um_toi	RESIDU	PercentResid	Kids	pms ca	equipm	laung	Toll fe	allowan	Jet ski	Ate the c	Expens
3/1/2023	Mahachai	Sathon	6 Hiab	P.168	Anucha	8400	1500	middle	39.4	394	3254.65	5145.35	158.092268	0	0	0	0	5	200	110	15.76	550.65	810
3/1/2023	Praksa	Tum ru	6 Hiab	Green Warehouse	Teerawat	4816	1500	middle	7	70	2477.83	2338.17	94.36361655	0	0	0	0	5	200	110	2.8	97.83	810
3/1/2023	Praksa	Tum ru	6 Hiab	Green Warehouse	korn	4816	1500	middle	7	70	1977.83	2838.17	143.4991885	0	0	0	0	5	200	110	2.8	97.83	310
4/1/2023	Rayong	Prawet	6 Hiab	Jay Tuk	Boonpeng	8960	1500	middle	172	1720	6933.87	2026.13	29.22076705	0	0	0	0	5	200	110	68.8	2403.87	1310
4/1/2023	Velkor	Rayong	6 Hiab	TOZEN	Khanchai	6160	1500	middle	131	1310	5750.86	409.14	7.114414192	0	0	0	0	5	200	110	52.4	1830.86	1110
5/1/2023	Min buri	Phetkasem	10 Hiab	Atthaphon	Swat	5936	2500	middle	48.1	481	4631.31	1304.69	28.17107903	0	0	0	4	0	200	110	24.05	840.31	810
5/1/2023	Bang Phli	Worakulchai	4 wheels	Worakulchai	Fen	4816	1000	middle	63.9	639	2321.11	2494.89	107.4869351	0	0	0	12	5	200	110	10.65	372.11	310
6/1/2023	Rayong	Oon Nut	10 Hiab	Pinnacle	Khanchai	7280	2500	middle	171	1710	8507.37	-1227.37	-14.42713788	0	0	0	4	5	200	110	85.5	2987.37	1310
6/1/2023	Rayong	Sai ha	10 Hiab	Pinnacle	korn	7840	2500	middle	219	2190	8825.93	-985.93	-11.17083412	0	0	0	4	5	200	110	109.5	3825.93	310
6/1/2023	Mahachai	Phetkasem	6 Hiab	H.Pattanarungruan	korn	6720	1500	middle	36	360	4673.14	5546.86	118.6966365	4	2000	1500	5	5	200	110	14.4	503.14	2310
6/1/2023	Bang Kruai	Nakhon Sawan	6 Hiab	Worakulchai	Swat	16800	1500	middle	280	2800	9823.28	6976.72	71.0223062	0	0	0	5	5	200	110	112	3913.28	1610
6/1/2023	Bang Kruai	Nakhon Sawan	6 Hiab	Worakulchai	Anucha	16800	1500	middle	280	2800	9823.28	6976.72	71.0223062	0	0	0	5	5	200	110	112	3913.28	1610
6/1/2023	Ratchaphruek	Min buri	10 Hiab	Abhichart	Khanchai	4480	2500	middle	55	550	4820.85	-340.85	-7.070329921	0	0	0	4	5	200	110	27.5	960.85	810
6/1/2023	chon buri	Chaengwattana	6 Hiab	LSB	Swat	7280	1500	middle	112	1120	5295.31	1984.69	37.48014753	0	0	0	5	5	200	110	44.8	1565.31	1110
7/1/2023	King Kaew	Rama song	6 Hiab	UPE	Teerawat	10080	1500	middle	55.7	557	5645.46	7934.54	140.5472716	4	2000	1500	5	5	200	110	22.28	778.46	2810
7/1/2023	Samut Prakan	Krmchlprathan pakkerrd	6 Hiab	M5M	sakda	5600	1500	middle	55.1	551	3631.08	1968.92	54.2240876	0	0	0	5	5	200	110	22.04	770.08	810
7/1/2023	Salaya	Nakhon Sawan	10 Hiab	Hydroengineering	korn	16800	2500	middle	298	2980	12996.06	7303.94	56.20118713	4	2000	1500	4	5	200	110	149	5206.06	2310
7/1/2023	Top Steel Enterprise Comp	Bang Prakong	6 Hiab	P.Rungruang	Prasit	5040	1500	middle	51.5	515	3544.76	1495.24	42.18169918	0	0	0	5	5	200	110	20.6	719.76	810
7/1/2023	Khlong Toei Pier	Asoke	6 Hiab	Air Thaiger	Anucha	3920	1500	middle	4.2	42	2410.7	1509.3	62.60837101	0	0	0	5	5	200	110	1.68	58.7	810
7/1/2023	Mahachai	Praksa	6 Hiab	H.Pattanamahachai	Boonpeng	11200	1500	middle	60.1	601	5750.96	8949.04	155.6094982	4	2000	1500	5	5	200	110	24.04	839.96	2810
7/1/2023	Padriw	MegaBangna	10 Hiab	TCH	Swat	5040	2500	middle	66.1	661	5125.77	-85.77	-1.673309571	0	0	0	4	5	200	110	33.05	1154.77	810
7/1/2023	Phuthth buchaa sam sib ho	Lat Krabang	10 Hiab	CCU	Khanchai	5600	2500	middle	45.9	459	4570.87	1029.13	22.5149698	0	0	0	4	5	200	110	22.95	801.87	810
7/1/2023	Mahachai	Sa Kaeo	6 Hiab	H.Pattanarungruan	Teerawat	13440	1500	middle	251	2510	11127.98	5812.02	52.22888611	4	2000	1500	5	5	200	110	100.4	3507.98	3610
7/1/2023	Top Steel Enterprise Comp	Tum ru	10 Hiab	Pornstl	sakda	5040	2500	middle	9.6	96	3573.71	1466.29	41.02991009	0	0	0	4	5	200	110	4.8	167.71	810
7/1/2023	Bang Phli	Khong song Nam	10 Hiab	Fueltech	Swat	5936	2500	middle	25.1	251	3999.5	1936.5	48.41855232	0	0	0	4	5	200	110	12.55	438.5	810
9/1/2023	Ske Ngam	Bang Yai	6 wheels	Ning	Pong	3360	1500	middle	38	380	2721.09	638.91	23.479929	0	0	0	5	5	200	110	15.2	531.09	310
9/1/2023	Mahachai	Kamphaeng Phet	6 Hiab	H.Pattanamahachai	Boonpeng	28000	1500	middle	393	3930	14532.57	16967.43	116.7545038	4	2000	1500	5	5	200	110	157.2	5492.57	3610
9/1/2023	M-SENKO	Ramintra	6 Hiab	CYS	Boonpeng	5600	1500	middle	38.1	381	3723.49	3876.51	104.1095854	1	500	1500	5	5	200	110	15.24	532.49	1310
9/1/2023	Chaengwattana	chon buri	6 Hiab	LSB	somjai	3920	1500	middle	108	1080	4399.41	-479.41	-10.89714303	0	0	0	5	5	200	110	43.2	1509.41	310
9/1/2023	Khlong Toei Pier	Tiwanon	10 Hiab	WPIL	Khanchai	5936	2500	middle	25.2	252	4002.24	1933.76	48.31694251	0	0	0	4	5	200	110	12.6	440.24	810
9/1/2023	Khlong Toei Pier	Tiwanon	6 Hiab	WPIL	Anucha	4816	1500	middle	25.2	252	2914.2	1901.8	65.25976254	0	0	0	0	5	200	110	10.08	352.2	810

ภาพที่ 15 ตัวอย่างข้อมูลที่ใช้สำหรับวิเคราะห์การแบ่งกลุ่ม

ขั้นตอนการดำเนินงานวิจัย

การศึกษาวิจัยนี้ประกอบด้วย 8 ขั้นตอนหลัก ซึ่งแต่ละขั้นตอนถูกออกแบบมาอย่างเป็นระบบเพื่อให้ได้ข้อมูลและผลการวิจัยที่มีประสิทธิภาพ โดยจะประยุกต์ใช้ทฤษฎีการทำเหมืองข้อมูล ดังนี้

1. เก็บรวบรวมข้อมูลบริษัทกรณีศึกษา เริ่มต้นด้วยการเก็บข้อมูลจากบริษัทที่เป็นกรณีศึกษา ข้อมูลนี้ประกอบด้วยสถิติการใช้บริการของลูกค้าในช่วงเวลาที่กำหนด จัดเก็บข้อมูลที่รวบรวมมาไว้ในโปรแกรม Excel เพื่อความสะดวกในการจัดการและวิเคราะห์ข้อมูลต่อไป

2. วิเคราะห์ปัญหาที่พบ วิเคราะห์ข้อมูลที่ได้เพื่อระบุปัญหาหลักที่เกิดขึ้นในการดำเนินงานของบริษัท รวมถึงปัญหาที่เกี่ยวข้องกับการตั้งราคาการบริการลูกค้า

3. ขั้นจัดเตรียมข้อมูลก่อนการประมวลผล ทำการจัดการข้อมูล เช่น การทำความสะอาดข้อมูล การจัดรูปแบบข้อมูลให้เป็นมาตรฐาน เพื่อให้ข้อมูลพร้อมสำหรับการวิเคราะห์ด้วยอัลกอริทึม Machine learning

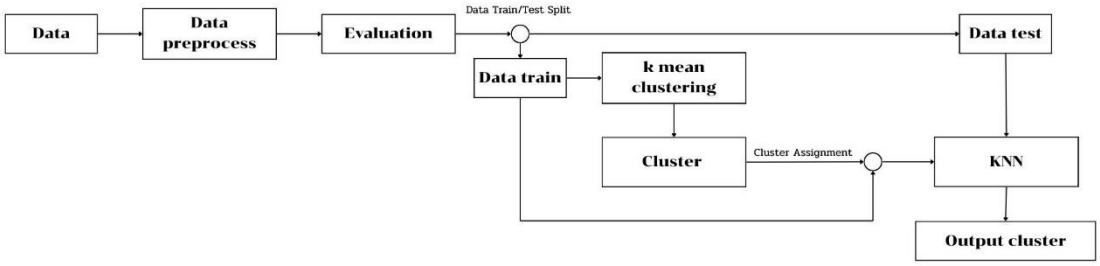
4. การสร้างแบบจำลองคลัสเตอร์แบบเคมีน นำข้อมูลที่ได้ไปใช้ในการสร้างแบบจำลองคลัสเตอร์แบบเคมีน เพื่อใช้ในการจำแนกกลุ่มลูกค้า โดยใช้ Elbow method และ Silhouette score ในการวัดประสิทธิภาพแบบจำลองคลัสเตอร์แบบเคมีน

5. การสร้างแบบจำลองเคเนียร์เรสเนเบอร์ นำข้อมูลกลุ่มที่ได้จากคลัสเตอร์แบบเคมีน มาทำการสร้างแบบจำลองเคเนียร์เรสเนเบอร์ โดยใช้ F1-Score และ Accuracy score ในการวัดประสิทธิภาพเพื่อให้แบบจำลองมีประสิทธิภาพในการจำแนกกลุ่มเทียบเท่าแบบจำลองคลัสเตอร์แบบเคมีนเพื่อนำแบบจำลองไปใช้จำแนกกลุ่มข้อมูลใหม่ในอนาคต โดยที่ไม่ทำให้กลุ่มข้อมูลเดิมที่เคยจำแนกคลาดเคลื่อน

6. นำผลลัพธ์การจัดกลุ่มลูกค้าที่ได้จากอัลกอริทึมเคเนียร์เรสเนเบอร์ มาวิเคราะห์ผลลัพธ์ที่ได้จากการจัดกลุ่มลูกค้าเพื่อวัดประสิทธิภาพในการกำหนดราคาของแต่ละกลุ่มลูกค้า

7. เปรียบเทียบประสิทธิภาพก่อนและหลังปรับปรุง เปรียบเทียบผลลัพธ์การดำเนินงานก่อนและหลังการปรับปรุงการตั้งราคาและการจัดกลุ่มลูกค้า เพื่อประเมินความสำเร็จของการปรับปรุงที่ทำขึ้น

8. สรุปผลการดำเนินงานวิจัย สรุปผลการวิจัยทั้งหมด รวมถึงข้อค้นพบ ข้อดี ข้อเสีย ข้อเสนอแนะในการดำเนินงานในอนาคต เพื่อให้สามารถนำไปใช้ประโยชน์ต่อไปได้



ภาพที่ 16 ขั้นตอนการศึกษาดำเนินงาน

เครื่องมือที่ใช้ในการดำเนินงานวิจัย

งานวิจัยนี้ได้ใช้การวิเคราะห์การแบ่งกลุ่มด้วย คลัสเตอร์แบบเคมีน และเคเนียร์เรสเนเบอร์ ดังต่อไปนี้

1. การวิเคราะห์ด้วยคลัสเตอร์แบบเคมีน

คลัสเตอร์แบบเคมีน ถูกใช้ในการแบ่งกลุ่มลูกค้าจากชุดข้อมูลที่เก็บรวบรวมและจัดการในโปรแกรม Microsoft excel ขั้นตอนมี

1.1 แบ่งชุดข้อมูล

ชุดข้อมูลถูกแบ่งออกเป็นสองส่วน: ชุดฝึกสอน Training set ร้อยละ 90 และชุดทดสอบ Testing set ร้อยละ 10 เพื่อให้สามารถประเมินประสิทธิภาพของโมเดลได้อย่างถูกต้อง

1.2 คำนวณหาค่า K

การหาค่า K ที่เหมาะสมทำได้โดยการเปรียบเทียบค่าผลรวมของความคลาดเคลื่อนภายในแต่ละคลัสเตอร์ Sum of squared errors สำหรับค่า K หลายค่า จากนั้นจึงเลือกค่า K ที่ลดความคลาดเคลื่อนได้มากที่สุด โดยวิธีนี้จะช่วยในการหาตำแหน่งของค่า K ที่ดีที่สุด

1.3 สุ่มค่าเริ่มต้นของศูนย์กลางคลัสเตอร์

เริ่มต้นด้วยการสุ่มตำแหน่งเริ่มต้นของศูนย์กลางคลัสเตอร์ หรือ Centroids จากนั้นใช้โมเดล คลัสเตอร์แบบเคมีนในการวิเคราะห์และแบ่งกลุ่มลูกค้าตามลักษณะการใช้บริการ

1.4 การแบ่งกลุ่มลูกค้า

โมเดลจะทำการปรับปรุงตำแหน่งศูนย์กลางคลัสเตอร์ซ้ำจนกว่าศูนย์กลางจะไม่เปลี่ยนแปลงมาก ซึ่งจะได้กลุ่มลูกค้าที่มีลักษณะคล้ายกันอยู่ในกลุ่มเดียวกัน

2. การวิเคราะห์ด้วย เคเนียร์เรสเนเบอร์

เคเนียร์เรสเนเบอร์ถูกใช้เพื่อเพิ่มความแม่นยำในการจัดกลุ่มลูกค้าโดยใช้ข้อมูลที่ได้จากคลัสเตอร์แบบเคมีน ขั้นตอนมี

2.1 เตรียมชุดข้อมูล

แยกชุดข้อมูลเป็นชุดฝึกสอนร้อยละ 90 และชุดทดสอบร้อยละ 10 เช่นเดียวกับในขั้นตอนของ คลัสเตอร์แบบเคมีน

2.2 ใช้ข้อมูลคลัสเตอร์ที่ได้จาก คลัสเตอร์แบบเคมีน

ข้อมูลที่ได้จากการแบ่งกลุ่มลูกค้าด้วย K คลัสเตอร์แบบเคมีน ถูกนำมาใช้เป็นข้อมูลเบื้องต้นในการฝึกสอนโมเดล เคเนียร์เรสเนเบอร์เพื่อเพิ่มความแม่นยำในการจัดกลุ่ม

2.3 การคำนวณระยะทาง

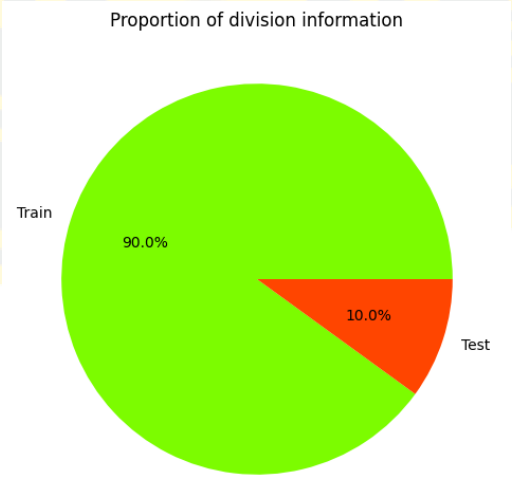
โมเดล เคเนียร์เรสเนเบอร์จะทำการหากลุ่มของลูกค้าที่ใกล้ที่สุดโดยคำนวณระยะทางระหว่างข้อมูลทดสอบกับข้อมูลในชุดฝึกสอน จากนั้นจึงจัดกลุ่มลูกค้าตามกลุ่มที่ใกล้ที่สุด

2.4 การปรับค่า N

ทดลองปรับค่า N หรือถูกเรียกว่า จำนวนเพื่อนบ้านที่ใกล้ที่สุด เพื่อหาค่าที่ให้ผลลัพธ์ดีที่สุด โมเดลจะทำการทำนายกลุ่มของข้อมูลทดสอบโดยพิจารณาจากกลุ่มของข้อมูลในชุดฝึกสอนที่ใกล้ที่สุด

2.5 การทำนายและประเมินผล

ผลลัพธ์ที่ได้จาก เคเนียร์เรสเนเบอร์จะถูกนำมาเปรียบเทียบกับผลลัพธ์จริงในชุดทดสอบเพื่อวัดประสิทธิภาพและความแม่นยำของการจัดกลุ่มลูกค้า



ภาพที่ 17 การแบ่งชุดข้อมูลฝึกสอนและทดสอบ

การวิเคราะห์ข้อมูล

1. ตัวแปรที่ใช้ในการวิจัย

ตารางที่ 1 ตัวแปรทั้งหมดในงานวิจัย

ชื่อตัวแปร	ชนิดข้อมูล	ความหมาย
Date	วันที่	วันที่ที่เกี่ยวข้องกับการจองหรือทำรายการนัดหมาย สามารถเป็นข้อมูลประเภทวันที่หรือวันที่และเวลาได้
Pick-up location	สถานที่รับสินค้า	สถานที่เริ่มต้นที่ที่ผู้ใช้ต้องการให้รถรับส่งไปรับ
Where to drop off	สถานที่ส่งสินค้า	สถานที่ปลายทางที่ผู้ใช้ต้องการให้รถรับส่งไปส่ง
Vehicle types	ประเภทรถ	ประเภทของรถพาหนะที่ต้องการใช้บริการ
Employing company	ชื่อบริษัท	บริษัทหรือผู้ให้บริการที่จัดหาบริการรถรับส่ง
Driver	ชื่อคนขับ	ข้อมูลเกี่ยวกับคนขับรถ
Pricer	บาท	ราคาที่ต้องจ่ายสำหรับการใช้บริการรถรับส่ง
Sector to drop off	ภูมิภาค	ส่วนที่ต้องการให้รถรับส่งไปส่ง อาจเป็นที่อยู่หรือพื้นที่
Distance	กิโลเมตร	ระยะทางระหว่างสถานที่เริ่มต้นและสถานที่ปลายทาง
Driver's trip fee	บาท	ค่าบริการของคนขับรถสำหรับการเดินทาง
Kids แar	จำนวนคน	ข้อมูลเกี่ยวกับรถเด็ก อาจมีค่าใช้จ่ายเพิ่มเติม
Childrens car trip fee	บาท	ค่าบริการของรถเด็กสำหรับการเดินทาง
Cost of equipment per set	บาท	ราคาของอุปกรณ์
Oil consumption rate	กิโลเมตรต่อบาท	อัตราการใช้น้ำมันของรถต่อหน่วยระยะทาง

ตารางที่ 1 (ต่อ)

ชื่อตัวแปร	ชนิดข้อมูล	ความหมาย
Toll fees	บาท	ค่าผ่านทางพิเศษ
Allowance	บาท	ค่าเบี้ยเลี้ยงหรือค่าใช้จ่ายเพิ่มเติม
Total cost of oil consumption	ลิตร	ค่าใช้จ่ายทั้งหมดที่เกี่ยวข้องกับการใช้น้ำมัน
Calculate the cost of oil	บาท	ค่าใช้จ่ายในการใช้น้ำมัน
Expenses	บาท	ค่าใช้จ่ายเบ็ดเตล็ด
Total cost	บาท	ค่าใช้จ่ายทั้งหมด
Residual	บาท	กำไรคงเหลือ
Percent residual	ร้อยละ	กำไรคงเหลือร้อยละ
Cost car	บาท	ต้นทุนค่ารถ
Cost distance car	บาท	ต้นทุนจากระยะทาง

2. สำหรับตัวแปรที่ใช้ในการแบ่งกลุ่ม

2.1 Employing company

2.2 Residual

2.3 Price

2.4 Percent residual

3. การแปลงชนิดข้อมูล

3.1 ข้อมูลที่เป็นจำนวนตัวเลข

Residual, Price, Distance, Driver's trip fee, Kids Car, Cost of equipment per set, Total cost of oil consumption, Calculate the cost of gas, Residual, Percent residual, Cost car, Cost distance car จะนำข้อที่เป็นตัวเลขในแต่ล่ะตัวแปรไปทำการหารกับจำนวนที่มากที่สุดในแต่ล่ะตัวแปรนั้น เพื่อให้ข้อมูลมีค่าอยู่ที่ 0 ถึง 1 แล้วจึงนำไปใช้กับโมเดล

$$x_{normalized} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

โดยให้

X คือ ค่าข้อมูลที่ต้องการ Normalize

X_{min} คือ ค่าข้อมูลที่น้อยที่สุดในชุดข้อมูล

X_{max} คือ ค่าข้อมูลที่มากที่สุดในชุดข้อมูล

ตารางที่ 2 การแปลงข้อมูลชนิดตัวเลขให้เป็นค่ามาตรฐาน

Total cost	Total cost Normalize
1134.94	0.056004386
4121.14	0.220693645
1620.61	0.08278914
7620.88	0.413704692
511.07	0.021597886
1930.53	0.099881262
1542.88	0.078502322
5278.7	0.284533207
2870.98	0.151747183
933.29	0.044883366
3037.31	0.160920301
10624.02	0.579328194
795.85	0.037303535
1136.85	0.056109722
1765.61	0.090785906

3.2 ข้อมูลที่เป็นข้อความ

Employing company, Sector to drop off จะทำการนำข้อมูลประเภทข้อความแปลงเป็นข้อมูลตัวเลข โดยใช้วิธีการ Label encoding โดยการกำหนดตัวเลขให้กับแต่ละค่าข้อความในคอลัมน์ที่ต้องการแปลง ตัวอย่างเช่น หากมีข้อมูลประเภทข้อความเช่น “ภาคเหนือ”, “ภาคกลาง”, “ภาคใต้”, “ภาคตะวันออกเฉียงเหนือ” การใช้ Label encoding จะแปลงเป็นเลข 0, 1, 2, 3 ตามลำดับ แล้วจึงนำมาหารค่าที่มากที่สุดเพื่อให้มีค่าอยู่ในช่วง 0 ถึง 1 เหมือนข้อมูลจำนวนตัวเลข

ตารางที่ 3 การแปลงข้อมูลชนิดข้อความให้เป็นตัวเลขและทำให้กลายเป็นค่ามาตรฐาน

Employing company	Employing company encode	Employing company normalize
Atthaphon	0	0
PKC	1	0.004149378
Mechanic	2	0.008298755
Pinnacle	3	0.012448133
Sikko	4	0.01659751
U Plus	5	0.020746888
Chang Aam	6	0.024896266
TOZEN	7	0.029045643
LSB	8	0.033195021
WPIL	9	0.037344398
WPIL	9	0.037344398
U Plus	11	0.045643154
Charnvit	12	0.049792531
LSB	8	0.033195021
TOZEN	14	0.058091286

สรุปผลการดำเนินงาน

เพื่อสรุปผลการดำเนินงานและเปรียบเทียบประสิทธิภาพระหว่างกระบวนการทำงานเก่าและกระบวนการทำงานใหม่ที่ได้รับการปรับปรุง เราได้ดำเนินการดังนี้

1. การกำหนดราคา

การตั้งราคาในกระบวนการใหม่ที่ปรับปรุงด้วยการวิเคราะห์ข้อมูลและการจัดกลุ่มลูกค้าถูกนำมาเปรียบเทียบกับวิธีการตั้งราคาแบบเดิมที่เคยใช้ เพื่อดูว่ามีการเปลี่ยนแปลงในด้านรายได้ของบริษัทกรณีศึกษาอย่างไร โดยจะคำนวณการเปลี่ยนแปลงในรายได้เป็นเปอร์เซ็นต์ จะมีดังต่อไปนี้

เปรียบเทียบผลการตั้งราคาแบบ Markdown กับการตั้งราคาแบบเดิม

วิเคราะห์การเปลี่ยนแปลงในด้านรายได้ของบริษัทกรณีศึกษา

แสดงผลการเปลี่ยนแปลงในรูปแบบเปอร์เซ็นต์เพื่อให้เห็นภาพรวมของผลกระทบที่เกิดขึ้นจากการปรับปรุงการตั้งราคา

2. การจัดกลุ่มลูกค้า

ในการจัดกลุ่มลูกค้า กระบวนการใหม่ที่ใช้การวิเคราะห์ด้วย คลัสเตอร์แบบเคมีน และเคเนียร์เรสเนเบอร์ ถูกนำมาเปรียบเทียบกับวิธีการเดิมที่ยังไม่ได้แบ่งกลุ่มลูกค้า โดยมีขั้นตอนดังนี้

นำข้อมูลที่ผ่านการวิเคราะห์และแบ่งกลุ่มมาแสดงให้บริษัทกรณีศึกษาเห็นภาพรวมของกลุ่มลูกค้าอย่างชัดเจน

เปรียบเทียบการจัดกลุ่มลูกค้าระหว่างวิธีการเดิม ที่ยังไม่ได้แบ่งกลุ่ม กับวิธีการที่แบ่งกลุ่มด้วย คลัสเตอร์แบบเคมีน และเคเนียร์เรสเนเบอร์

ใช้ข้อมูลจากการแบ่งกลุ่มลูกค้าอย่างมีประสิทธิภาพเพื่อนำไปตั้งราคา

บทที่ 4
ผลการดำเนินงานวิจัย

บทนี้เป็นการแสดงผลทดลองใช้อัลกอลิทึมของ คลัสเตอร์แบบเคมีนในการจัดกลุ่มลูกค้าและเคเนียร์เรสเนเบอร์ในการนำผลการจัดกลุ่มไปใช้งานกับข้อมูลใหม่ในอนาคต เพื่อทำการจำแนกกลุ่มลูกค้าและทดสอบค่าความถูกต้องของการจำแนกกลุ่มแล้วจึงนำกลุ่มที่เหมาะสมที่สุดมาตั้งราคาส่วนลดของลูกค้าแต่ล่ะกลุ่ม โดยใช้ทฤษฎี Markdown มีวิธีดำเนินการดังนี้

ศึกษาตัวแปรที่มีผลต่อการแบ่งกลุ่มข้อมูล

เนื่องจากวัตถุประสงค์ที่ต้องการคือจัดกลุ่มลูกค้าเพื่อให้ทราบถึงการใช้บริการ โดยตัวแปรที่ใช้ใน Machine learning มีดังนี้

Employing company เป็นตัวแปรที่นำมาทำการแปลง โดยการใช้ทฤษฎี Features engineer เพื่อแปลงข้อมูลชื่อลูกค้าที่มาใช้บริการ ให้กลายเป็นจำนวนการใช้งานทั้งหมดของลูกค้าแต่ละราย เนื่องจาก ถ้าหากจะทำการแบ่งกลุ่มลูกค้าเพื่อทำการลดราคาให้กับกลุ่มลูกค้า ต้องคำนึงถึงความถี่และจำนวนรอบในการใช้งาน

Price เป็นราคาที่แต่ละบริษัทจ่ายในการใช้บริการ ตัวแปรนี้นำมาใช้ในการช่วยแบ่งกลุ่มเนื่องจากลูกค้ามียอดการใช้บริการที่ต่างกัน เลยเหมาะสำหรับการนำไปแบ่งกลุ่มลูกค้า

Residual ตัวแปรนี้บอกถึงผมกำไรรวมที่บริษัทได้ในแต่ละรอบในการขนส่ง

Percent residual ตัวแปรนี้บอกถึงเปอร์เซ็นต์กำไรรวมที่บริษัทได้ในแต่ละรอบในการขนส่ง

ดังนั้น ตัวแปรที่สำคัญที่สุด คือ จำนวนการใช้งานของลูกค้าโดยมาจากตัวแปร Employing company ที่ถูกแปลงให้อยู่ในรูปจำนวนของการใช้งาน จึงทดลองแบ่งกลุ่มจากจำนวนการใช้งานของลูกค้าเพื่อศึกษากลุ่มลูกค้าแต่ละกลุ่ม

ส่วนตัวแปรที่ไม่ได้นำมาใช้ใน Machine learning แต่นำมาใช้ในการวิเคราะห์ในส่วนการคำนวณตัวแปรร่วม จะมีดังนี้

Date ตัวแปรนี้เป็นตัวบอกวันที่ในการใช้บริการของลูกค้า ที่ไม่ได้นำตัวแปรมาใช้เพราะว่า ได้ใช้ตัวแปร Employing company มาแปลงให้อยู่ในรูปแบบจำนวนการใช้งานของลูกค้าแต่ละรายแทนการใช้ Date

Pick-up location และ Where to drop off 2 ตัวแปรนี้ได้บอกสถานที่รับส่ง ซึ่งไม่สามารถตีความหมายเพื่อนำไปใช้ได้และมีตัวแปร Distance หรือระยะทางการขนส่งมาใช้แทน

Vehicle types ตัวแปรนี้บอกถึงลักษณะของรถที่ใช้ในการขนส่ง โดยได้ใช้ ตัวแปร Oil consumption rate เพื่อแยกชนิดของรถขนส่งอยู่แล้ว

Driver ตัวแปรนี้ได้บอกชื่อของคนขับซึ่งไม่สามารถตีความหมายและนำใช้ในการวิเคราะห์แบ่งกลุ่มได้ เนื่องจากลูกค้าแต่ละรายไม่ได้ระบุคนขับเฉพาะเจาะจง หรือเกิดการสุ่มคนขับในแต่ละรอบการขนส่ง

Sector to drop off โดยตัวแปรนี้จะมีการช่วยบอกภูมิภาคในการใช้เส้นการขนส่งที่จะระบุสถานที่รับส่งสินค้าแต่ละจุดทั่วประเทศไทย

Distance ตัวแปรนี้บอกถึงระยะทางของการขนส่งสินค้า ซึ่งมีการส่งผลกับตัวแปรกับค่าใช้จ่ายได้

Driver's trip fee ตัวแปรนี้จะบอกถึงค่าเที่ยวที่ให้คนขับรถ ในแต่ละรอบการขนส่งสินค้าจำนวนค่าเที่ยวขึ้นอยู่กับระยะทาง

Kids car ตัวแปรนี้บอกถึงจำนวนพนักงานขนของที่บริษัทใช้แต่ละรอบในการขนส่งสินค้า

Childrens car trip fee ตัวแปรนี้บอกถึงค่าใช้จ่ายของบริษัทในการจ้างพนักงานขนของในแต่ละรอบในการขนส่งสินค้า

Cost of equipment per set ตัวแปรนี้บอกถึงค่าใช้จ่ายอุปกรณ์ในแต่ล่ะรอบการขนส่งของลูกค้าที่ใช้บริการ

Oil consumption rate ตัวแปรนี้บอกอัตราการกินน้ำมันของรถที่แต่ล่ะลูกค้าเลือกใช้บริการ

Toll fees ตัวแปรนี้บอกถึงการบริการค่าผ่านทางพิเศษซึ่งมีค่าใช้จ่ายเท่ากัน จะมีผลต่อค่าใช้จ่ายรวมของบริษัท

Allowance ตัวแปรนี้บอกค่าเบี้ยเลี้ยงคนขับรถ ซึ้งมีผลกับค่าใช้จ่ายรวมของบริษัท

Total cost of oil consumption ตัวแปรนี้บอกจำนวนการบริโภคน้ำมันหน่วยลิตรในแต่ละรอบการขนส่งของลูกค้าแต่ล่ะราย

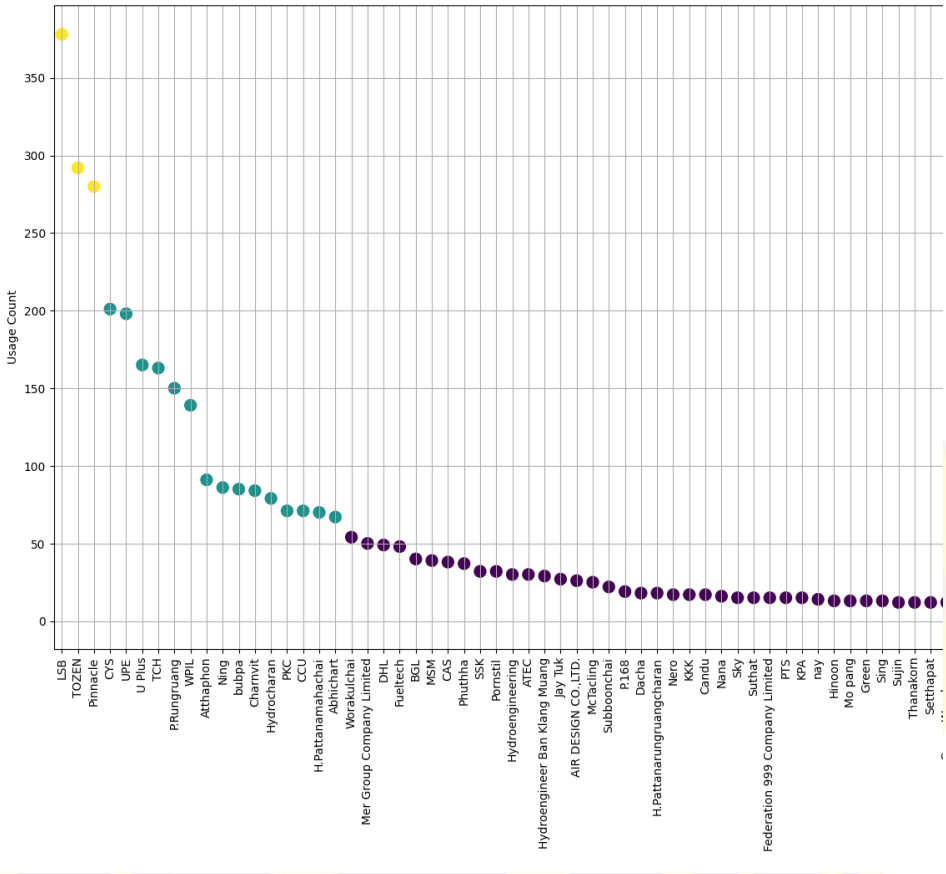
Calculate the cost of oil ตัวแปรนี้บอกจำนวนค่าใช้จ่ายการบริโภคน้ำมันหน่วยบาทของแต่ละรอบการขนส่งของแต่ล่ะบริษัท

Expenses ตัวแปรนี้บอกค่าใช้จ่ายเบ็ดเตล็ดของบริษัท

Total cost ตัวแปรนี้บอกถึงค่าใช้จ่ายสุทธิของบริษัทในแต่ล่ะรอบการใช้งาน

Cost car ตัวแปรนี้จะบอกถึงต้นทุนค่ารถแต่ละชนิด

Cost distance car ตัวแปรนี้จะบอกถึงค่าขนส่งที่คิดจากระยะทาง



ภาพที่ 18 การแบ่งกลุ่มลูกค้าจากจำนวนการใช้งาน

หากแบ่งกลุ่มลูกค้าตามจำนวนการใช้งานของลูกค้าแต่ล่ะราย จะแบ่งออกได้ดังนี้

ลูกค้ากลุ่มที่ 1 จะมีจำนวนการใช้งานมากที่สุดตั้งแต่ 201 ครั้งขึ้นไป

ลูกค้ากลุ่มที่ 2 จะมีจำนวนการใช้งานอยู่ที่ 54 ถึง 201 ครั้ง

ลูกค้ากลุ่มที่ 3 จะมีจำนวนการใช้งานต่ำกว่า 54 ครั้ง

จะพบว่า หากใช้ตัวแปรจำนวนการใช้งานอย่างเดียวก็สามารถแบ่งกลุ่มลูกค้า และมีความสมเหตุสมผล เพราะว่า การให้ส่วนลดควรจะกำหนดตามการใช้งานของลูกค้าเป็นหลัก

การแบ่งกลุ่มลูกค้าโดยใช้คลัสเตอร์แบบเคมีน

ในการแบ่งกลุ่มลูกค้า ในการแบ่งกลุ่มลูกค้า ข้อมูลทั้งหมดที่ใช้ในการวิเคราะห์มีจำนวน 4,270 รายการ เพื่อให้ได้ผลลัพธ์ที่น่าเชื่อถือและสามารถทดสอบประสิทธิภาพของโมเดลได้ ข้อมูลเหล่านี้ถูกแบ่งออกเป็นสองชุด ได้แก่

ข้อมูล Train จำนวนร้อยละ 90 ของข้อมูลทั้งหมด ซึ่งมีจำนวน 3,826 รายการ

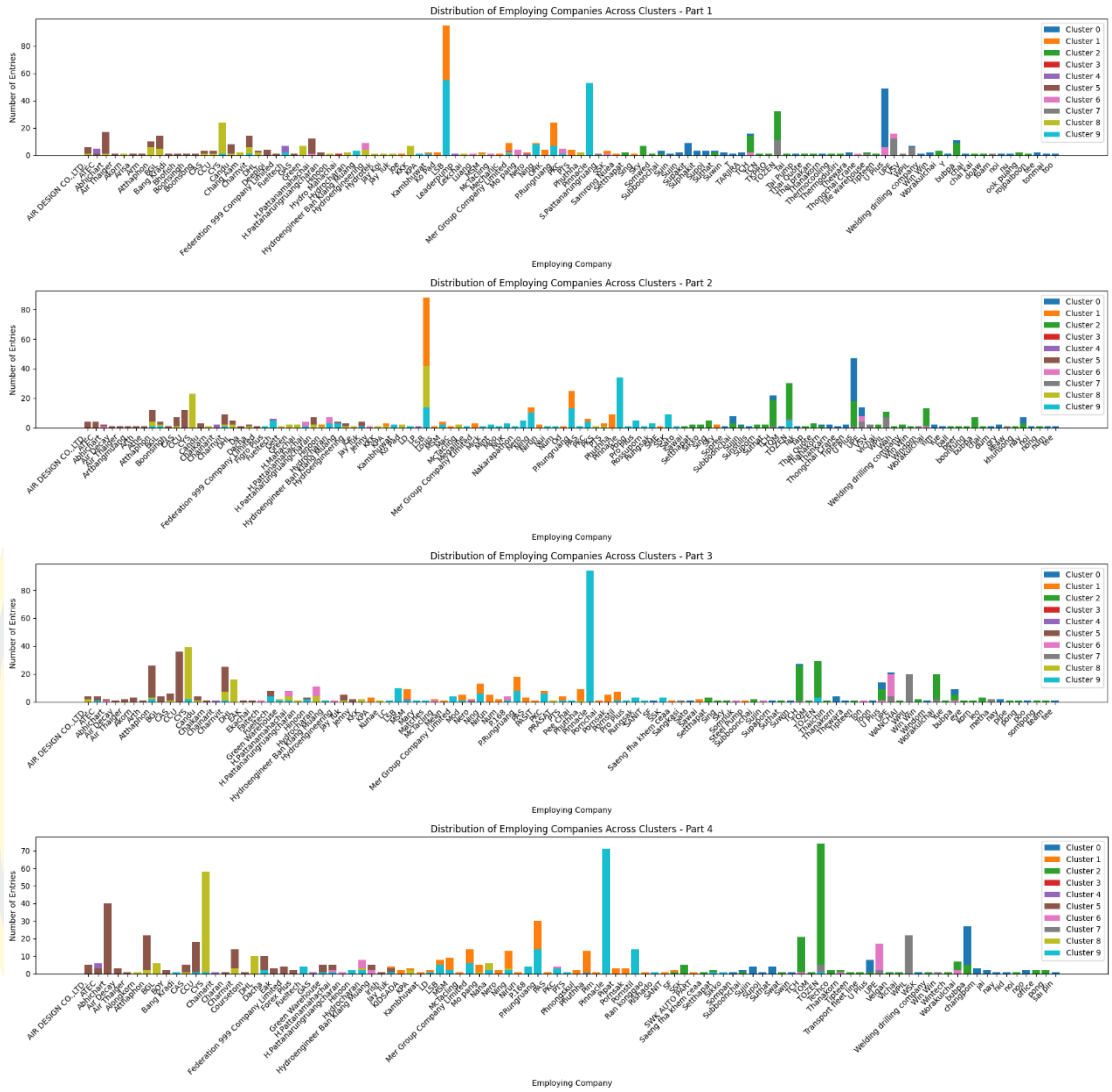
ข้อมูล Test จำนวนร้อยละ 10 ของข้อมูลทั้งหมด ซึ่งมีจำนวน 444 รายการ

ข้อมูล Train ถูกนำมาใช้ในการสร้างโมเดลการแบ่งกลุ่ม ในขณะที่ข้อมูล Test ถูกใช้เพื่อประเมินความสามารถของโมเดลในการจัดกลุ่มข้อมูลใหม่โดยได้นำข้อมูล Train มาแบ่งกลุ่มทั้งหมด 10 กลุ่ม คือ K = 10 แล้วมาวัดประสิทธิภาพโดยใช้ Elbow method ในการหาจำนวนกลุ่มที่เหมาะสมโดยการวัดค่าเฉลี่ยของความคลาดเคลื่อนภายในกลุ่ม Within cluster sum of squares กราฟที่ได้จะแสดงความสัมพันธ์ระหว่างจำนวนกลุ่ม K กับค่า WCSS จุดที่กราฟเปลี่ยนแปลงอย่างชัดเจนหรือมีรูปร่างคล้ายข้อศอก Elbow เป็นจุดที่แสดงจำนวนกลุ่มที่เหมาะสม และ Silhouette score วัดความถูกต้องของการแบ่งกลุ่มโดยพิจารณาจากระยะห่างระหว่างข้อมูลในกลุ่มเดียวกันและระยะห่างระหว่างกลุ่มต่าง ๆ ค่านี้อยู่ระหว่าง -1 ถึง 1 โดยค่าที่ใกล้ 1 แสดงว่าการจัดกลุ่มมีความแม่นยำสูง

โดยการทดลองแบ่งกลุ่ม K = 10 จะได้ผลดังนี้

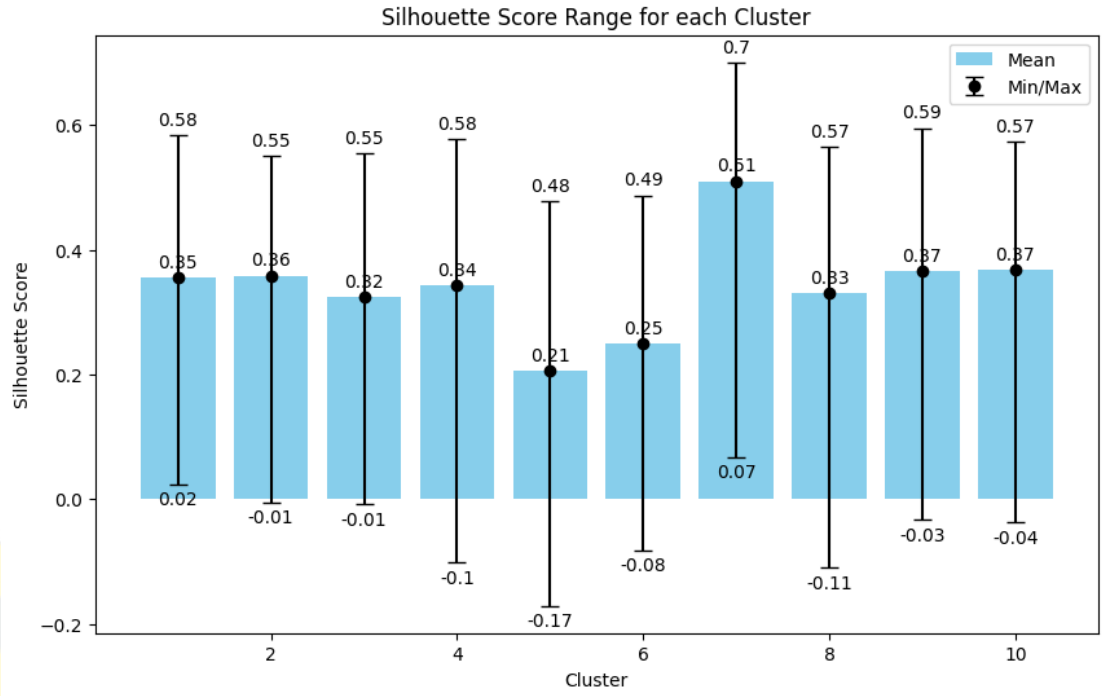
ตารางที่ 4 จำนวนข้อมูลที่ถูกแบ่งกลุ่มโดย คลัสเตอร์แบบเคมีน

Cluster	จำนวนข้อมูล
0	580
1	432
2	127
3	569
4	590
5	32
6	196
7	519
8	663
9	118



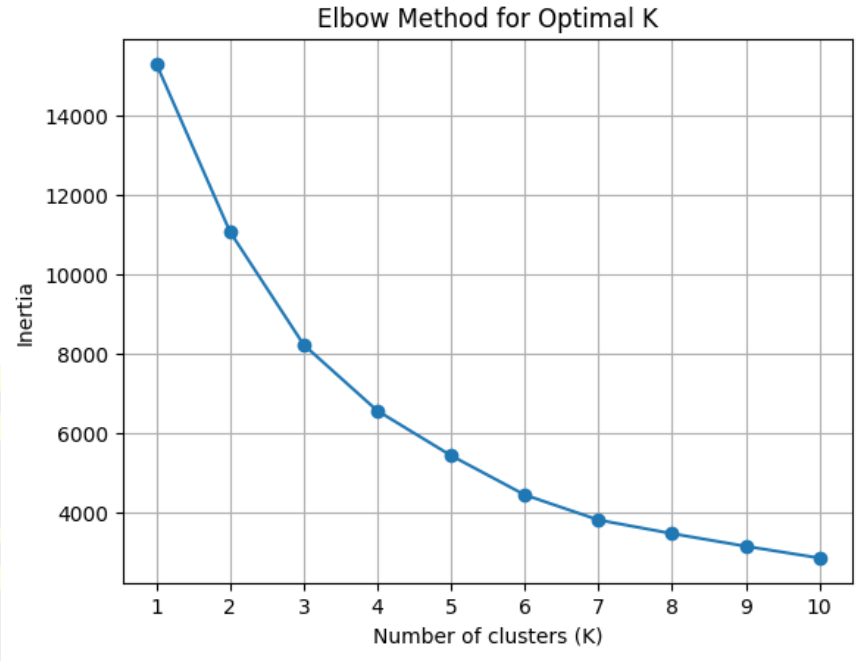
ภาพที่ 19 การแบ่งข้อมูลโดยใช้ คลัสเตอร์แบบเคมีน ค่า K = 10

จากการวิเคราะห์เพิ่มเติมในภาพที่ 19 พบว่าการใช้ K = 10 ทำให้เกิดความแปรปรวนในการแบ่งกลุ่มค่อนข้างมาก ตัวอย่างเช่น ลูกค้ารายหนึ่ง เช่น LSB ถูกจัดอยู่ในกลุ่มที่ 9, 8 และ 1 ซึ่งแสดงให้เห็นว่ามีความไม่สม่ำเสมอในการจัดกลุ่มและการแยกกลุ่มที่ไม่ชัดเจน จำนวนการแบ่งกลุ่มที่เกิดขึ้นนี้ยังมีจำนวนที่ใกล้เคียงกันในแต่ละกลุ่ม ซึ่งบ่งชี้ว่าการใช้ K = 10 ทำให้เกิดความแปรปรวนสูงเกินไปในการจัดกลุ่มลูกค้าสังเกตได้ว่าความหนาแน่นในกลุ่ม จะเกิดการแบ่งกลุ่มที่ไม่ชัดเจนในกลุ่มทั้งหมด มีความแปรปรวนที่ไม่มีความชัดเจนในการแบ่งกลุ่ม จากการใช้ทฤษฎี Silhouette score จะเห็นว่าเกิดค่า Silhouette score ในบางกลุ่มที่มีค่าต่ำ แสดงว่าเกิดการแบ่งกลุ่มที่ไม่ชัดเจน ในบาง Cluster



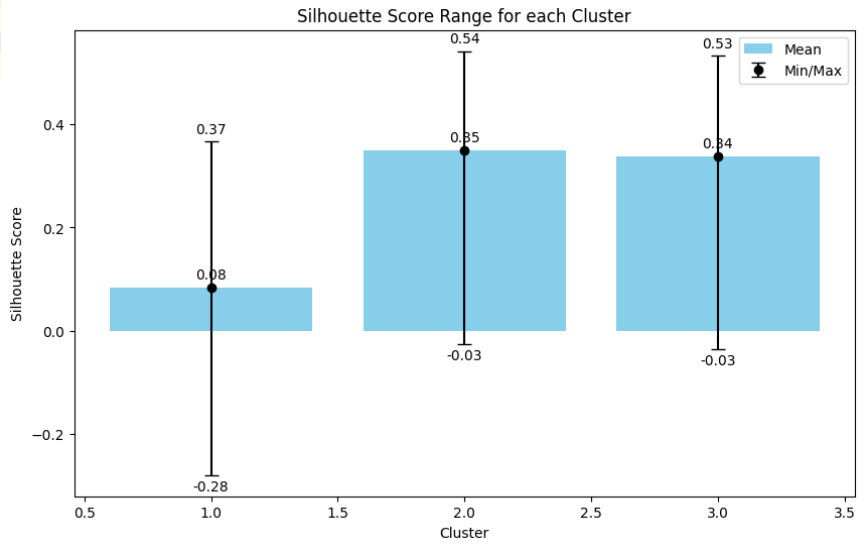
ภาพที่ 20 ค่า Silhouette score เมื่อค่า K = 10

จากทฤษฎี Elbow method ที่เลือกค่า K จากจุดที่กราฟมีลักษณะหักศอกอย่างชัดเจน พบว่าค่า K = 3 เป็นค่าที่เหมาะสมที่สุด เนื่องจากกราฟแสดงถึงการเปลี่ยนแปลงที่ชัดเจนขององศาการหักศอก ซึ่งหมายความว่าการแบ่งกลุ่มที่ K = 3 จะทำให้การจัดกลุ่มมีประสิทธิภาพและความเสถียรที่ดีที่สุด



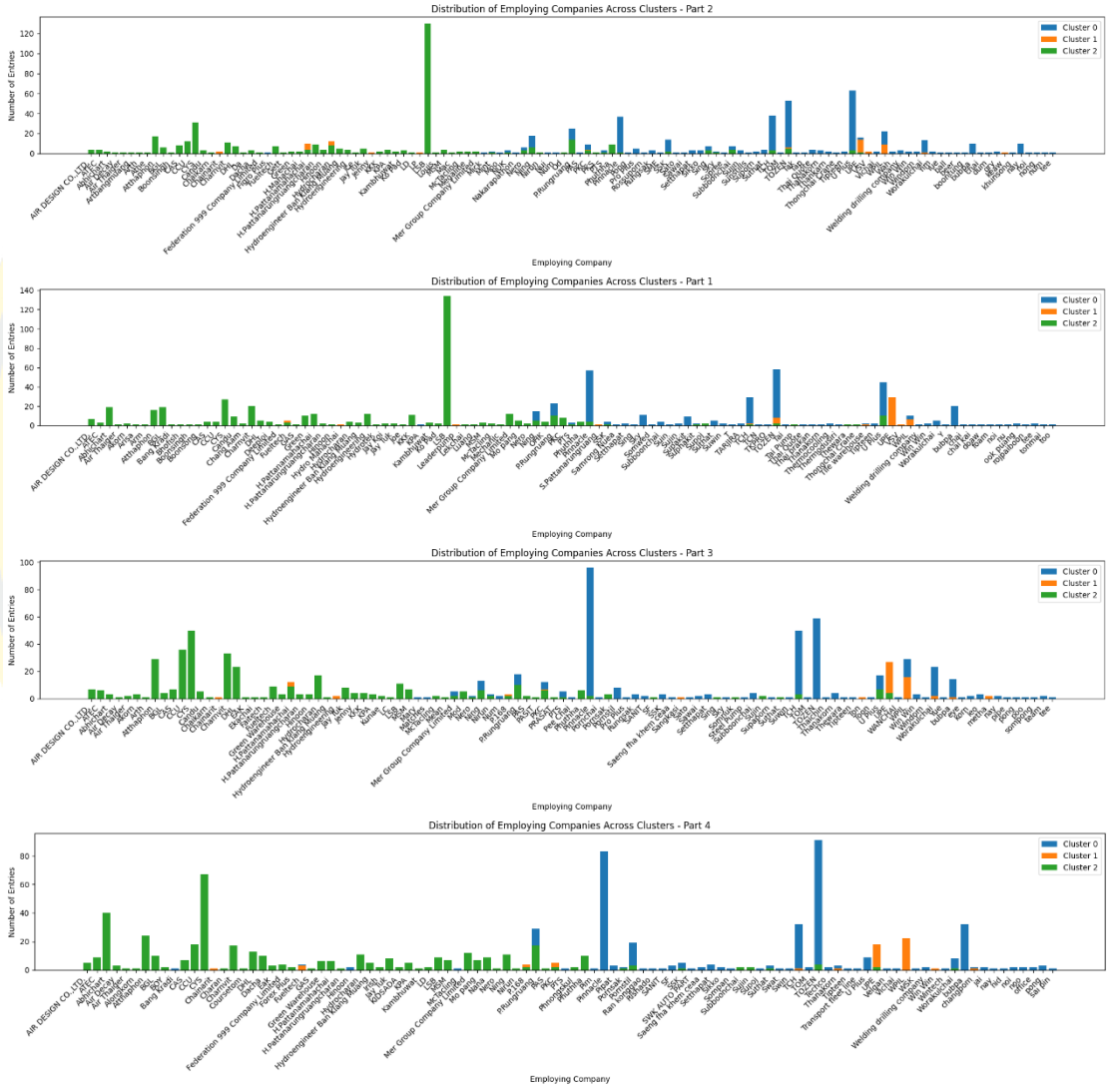
ภาพที่ 21 ค่า Elbow method ของแต่ละ คลัสเตอร์ เมื่อค่า K = 10

หลังจากแบ่งกลุ่มใหม่ด้วย K = 3 เราได้ทำการประเมินประสิทธิภาพการแบ่งกลุ่มอีกครั้ง โดยใช้ Silhouette score และพบว่า ค่า Silhouette score สูงขึ้นเมื่อเทียบกับการแบ่งกลุ่มด้วย K = 10 ซึ่งบ่งบอกว่าการแบ่งกลุ่มใหม่มีความแม่นยำและสม่ำเสมอมากขึ้น



ภาพที่ 22 Silhouette score เมื่อ ค่า K = 3

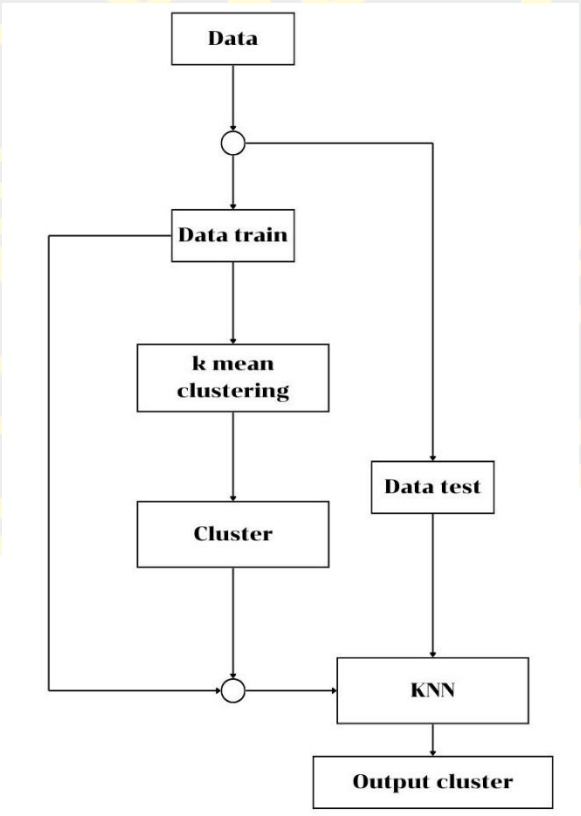
จากการทดลองนี้จะได้กลุ่มทั้งหมด 3 กลุ่ม และมีการวัดประสิทธิภาพการแบ่งกลุ่มโดยใช้ Silhouette score จะสังเกตได้ว่าค่าเฉลี่ยของ Silhouette score ในแต่ละคลัสเตอร์มีค่าดีกว่าการแบ่งกลุ่มด้วยคลัสเตอร์แบบเคมีน ที่ค่า K = 10 ซึ่งแสดงถึงความเหมาะสมและความแม่นยำที่สูงขึ้นของการแบ่งกลุ่มด้วย K = 3



ภาพที่ 23 การแบ่งข้อมูลโดยใช้ คลัสเตอร์แบบเคมีน ค่า K = 3

การแบ่งกลุ่มโดยใช้เคเนียร์เรสเนเบอร์

ในการใช้ เคเนียร์เรสเนเบอร์ สำหรับการจำแนกกลุ่มนั้น เราจำเป็นต้องมีกลุ่มที่ใช้สอนโมเดลในการเรียนรู้ก่อน เนื่องจาก เคเนียร์เรสเนเบอร์ เป็นอัลกอริทึมการเรียนรู้แบบมีผู้สอน หรือ Supervised learning ดังนั้น จึงต้องมีข้อมูลที่มีการระบุผลลัพธ์ของกลุ่ม ไว้ล่วงหน้า เพื่อให้โมเดลสามารถเรียนรู้และทำการทำนายได้อย่างถูกต้อง ในกระบวนการนี้ เราจะใช้ผลลัพธ์ของกลุ่มจากการทำคลัสเตอร์แบบเคมีน ที่ได้จากชุดข้อมูลฝึกสอน หรือ Training data มาเป็นผลเฉลย สำหรับการสอนโมเดลเคเนียร์เรสเนเบอร์ ซึ่งทำให้โมเดลสามารถเรียนรู้จากการจัดกลุ่มที่มีอยู่แล้ว เมื่อโมเดลเคเนียร์เรสเนเบอร์ ได้รับการฝึกสอนเสร็จสมบูรณ์แล้ว นำไปใช้กับชุดข้อมูลทดสอบ หรือ Test data เพื่อทำการทำนายกลุ่มใหม่ การใช้วิธีการนี้จะช่วยให้สามารถใช้ประโยชน์จากทั้ง 2 อัลกอริทึม คือ คลัสเตอร์แบบเคมีน และเคเนียร์เรสเนเบอร์ โดยใช้คลัสเตอร์แบบเคมีน ในการสร้างกลุ่มเบื้องต้น และใช้ เคเนียร์เรสเนเบอร์ ในการทำนายกลุ่มสำหรับข้อมูลใหม่อย่างมีประสิทธิภาพ

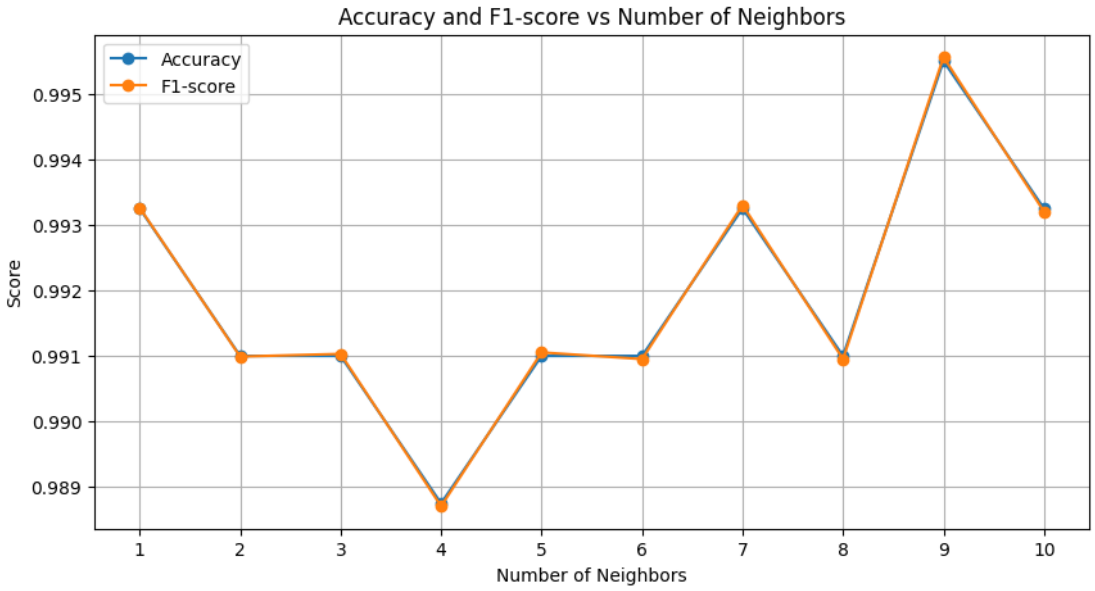


ภาพที่ 24 กระบวนการจัดกลุ่มด้วย เคเนียร์เรสเนเบอร์

หลังจากที่โมเดล เคเนียร์เรสเนเบอร์ ได้ทำการทำนายกลุ่มสำหรับข้อมูลชุดทดสอบแล้ว จะทำการวัดประสิทธิภาพของโมเดลโดยใช้ F1-Score และ Accuracy score เพื่อวัดประสิทธิภาพความถูกต้องของโมเดลนอกจากนี้ เพื่อหาจำนวน Neighbors ที่เหมาะสมที่สุดในการทำนาย เพื่อให้ได้ผลการทำนายกลุ่มให้ใกล้เคียงกับกลุ่มต้นฉบับมากที่สุด จะทำการทดลองโดยปรับค่าจำนวน Neighbors และวัดผลลัพธ์ที่ได้จากแต่ละค่า โดยเลือกค่าที่ทำให้ค่า Error หรือค่าความผิดพลาดน้อยที่สุดเพื่อใช้ในโมเดล กระบวนการนี้จะช่วยให้ได้โมเดลเคเนียร์เรสเนเบอร์ ที่มีประสิทธิภาพสูงสุดในการจำแนกกลุ่มข้อมูลใหม่ด้วยความแม่นยำ

ตารางที่ 5 ค่าการวัดประสิทธิภาพ F1-score และ Accuracy score ของค่า Number of neighbor แต่ละค่า

Number of neighbors	F1-score	Accuracy score
1	0.9932551901197879	0.9932432432432432
2	0.9910214189671811	0.990990990990991
3	0.9910214189671811	0.990990990990991
4	0.9888072078322195	0.9887387387387387
5	0.9910442524232997	0.990990990990991
6	0.9910442524232997	0.990990990990991
7	0.9932956159714518	0.9932432432432432
8	0.9915552475683007	0.9914954954954955
9	0.9955552475683007	0.9954954954954955
10	0.9933755042450696	0.9932432432432432



ภาพที่ 25 การวัดประสิทธิภาพ F1-score และ Accuracy score ของค่า Number of neighbor แต่ละค่า

จากการตรวจสอบข้อมูลในตารางที่ 5 เราพบว่า Neighbor ที่มีค่า F1-Score และ Accuracy score สูงสุดนั้นคือ Neighbor = 9 นั่นหมายความว่าในโมเดล เคเนียร์เรสเนเบอร์ ที่ใช้จำนวน Neighbor = 9 ส่งผลให้การทำนาย มีประสิทธิภาพใกล้เคียงกับกลุ่มต้นฉบับที่จัดด้วยเคมีนแบบคลัสเตอร์มากที่สุด เมื่อวัดด้วยค่า F1-Score และ Accuracy score ซึ่งเป็นตัวบ่งชี้ที่สำคัญในการประเมินประสิทธิภาพของโมเดล ว่ามีผลลัพธ์ใกล้เคียงกับผลเฉลยหรือกลุ่มต้นฉบับมากแค่ไหน

การกำหนดราคา

หลังจากการจัดกลุ่มด้วยคลัสเตอร์แบบเคมีน แล้วผลที่ได้จากการแบ่งกลุ่มทั้งหมด โดยบริษัทกรณีศึกษาจะทำการให้โปรโมชันกับกลุ่มลูกค้าสูงสุดได้ไม่เกินร้อยละ 12 เนื่องจากเครือข่ายบริษัทขนส่งพันธมิตร ได้กำหนดราคาโปรโมชันไว้สูงสุดที่ร้อยละ 12 โดยปกติการกำหนดราคาโปรโมชันจะขึ้นอยู่กับงบประมาณและนโยบายของแต่ละบริษัท จึงทำการแบ่งโปรโมชัน ของบริษัทกรณีศึกษาเป็น 3 กลุ่มดังนี้

ตารางที่ 6 ผลลัพธ์การจัดกลุ่มข้อมูลทั้งหมดและวัดค่าร้อยละกำไรเฉลี่ยของแต่ละกลุ่ม

Cluster	Quantity	Percentage
1	418	73.05
2	1733	63.37
3	2119	16.26

จากตารางที่ 6 จะเห็นว่าค่าร้อยละกำไรเฉลี่ยของแต่ละกลุ่มจะมี ค่าค่อนข้างห่างกัน จึงมีส่วนช่วยในการกำหนดราคาส่วนลดให้กับลูกค้าแต่ละกลุ่มได้ชัดเจน โดยจะให้โปรโมชันดังนี้

ตารางที่ 7 ผลการกำหนดราคาโปรโมชันแบบจัดกลุ่ม

Cluster	Percentage	Class	Promotion/ Percentage	Customers count
1	73.05	A	12	14
2	63.37	B	8	98
3	16.26	C	Without promotion	137

ผลการทดลองเพื่อวิเคราะห์ผลของการกำหนดราคาโปรโมชันในธุรกิจ ในการทดลองนี้มีการเปรียบเทียบผลกระทบของการกำหนดราคาโปรโมชันในสองรูปแบบดังนี้

1. การกำหนดราคาที่บริษัทใช้ดั้งเดิม

- บริษัทใช้ราคาโปรโมชันที่คงที่ทั้งหมด โดยไม่พิจารณาข้อมูลหรือกลุ่มลูกค้า
- ราคาโปรโมชันลดร้อยละ 12 จากราคาปกติ
- ผลลัพธ์กำไรเฉลี่ยร้อยละ 15.80

2. การกำหนดราคาแบบจัดกลุ่ม

- บริษัทแบ่งกลุ่มลูกค้าตามคลัสเตอร์เคมีนและเคเนียร์เรสเนเบอร์
- กลุ่ม A รับส่วนลดร้อยละ 12, กลุ่ม B รับส่วนลดร้อยละ 8, และกลุ่ม C ไม่ได้รับส่วนลด
- ผลลัพธ์กำไรเฉลี่ยร้อยละ 38.25

ตารางที่ 8 ผลลัพธ์ร้อยละของรูปแบบการกำหนดราคา

Pricing model	**Percentage**
Legacy discount	15.80
Group discount	38.25

จากตารางที่ 8 จะเห็นจากการเปรียบเทียบระหว่างการกำหนดราคาที่บริษัทใช้ดั้งเดิม และแบบจัดกลุ่ม มีผลต่อกำไรและการลดต้นทุนที่สำคัญของธุรกิจ ด้วยความแตกต่างในกำไรเฉลี่ยร้อยละ 15.80 และ 22.45 ตามลำดับ รวมถึงมูลค่าที่ต่างกันอย่างมากระหว่างการกำหนดราคาทั้ง 2 รูปแบบ การกำหนดราคาที่บริษัทใช้ดั้งเดิม ร้อยละ 15.80 หรือ 27,261,209 บาท ทำให้เราเห็นได้อย่างชัดเจนว่ากลยุทธ์การกำหนดราคาโปรโมชันแบบจัดกลุ่มมีผลสร้างค่าเพิ่มมากกว่าแบบเดิม ด้วยการใช้กลยุทธ์การกำหนดราคาแบบจัดกลุ่ม บริษัทสามารถเพิ่มกำไรเฉลี่ยร้อยละ 22.45 หรือเทียบเท่ากับ 32,838,644.16 บาท ซึ่งเป็นผลลัพธ์ที่ดีขึ้นมาก นอกจากนี้ยังสามารถลดต้นทุนที่ใช้ในการให้โปรโมชันลงได้อีก 5,577,435.16 บาท โดยคำนวณจากข้อมูลเก่า และข้อมูลชุดทดสอบที่เก็บเมื่อ 25 เมษายน 2567 ถึง 25 พฤษภาคม 2567

ดังนั้น การใช้กลยุทธ์การกำหนดราคาโปรโมชันแบบจัดกลุ่มนั้น ไม่เพียงเพิ่มกำไรและยอดขายเท่านั้น แต่ยังช่วยลดต้นทุนและเพิ่มมูลค่าให้กับธุรกิจได้อย่างมีประสิทธิภาพและยั่งยืน สามารถสร้างความแตกต่างในกำไรเฉลี่ยได้มากขึ้น เนื่องจากมีการปรับราคาโปรโมชันให้เข้ากับลูกค้าในแต่ละกลุ่ม ทำให้สามารถเพิ่มยอดขายและกำไรได้อย่างมีประสิทธิภาพมากยิ่งขึ้น ดังนั้น การใช้กลยุทธ์การกำหนดราคาโปรโมชันแบบจัดกลุ่มน่าสนใจและคุ้มค่าที่ควรพิจารณาในการพัฒนากลยุทธ์การตลาดของธุรกิจต่อไป

บทที่ 5
สรุป อภิปรายผล และข้อเสนอแนะ

การวิจัยนี้มีวัตถุประสงค์หลักเพื่อพัฒนาและปรับปรุงการแบ่งกลุ่มลูกค้าสำหรับการกำหนดราคาส่วนลดโดยใช้อัลกอริทึม คลัสเตอร์แบบเคมีน และเคเนียร์เรสเนเบอร์ โดยสามารถสรุปผลการวิจัยได้ดังนี้

สรุปผลการวิจัย

1. การแบ่งกลุ่มลูกค้าด้วย คลัสเตอร์แบบเคมีน

การทดลองใช้ค่า K = 10 ทำให้เกิดความแปรปรวนในกลุ่มลูกค้าสูง และไม่มีความชัดเจนในการแบ่งกลุ่ม ส่งผลให้การใช้ Silhouette score ในการประเมินประสิทธิภาพแสดงค่า Silhouette score ในบางกลุ่มที่มีค่าต่ำการเลือกค่า K ที่เหมาะสมด้วย Elbow method พบว่าค่า K = 3 เป็นค่าที่เหมาะสมที่สุด เนื่องจากมีลักษณะกราฟที่หักชัดเจนเมื่อใช้ค่า K = 3 ในการแบ่งกลุ่มข้อมูล พบว่ามีการแบ่งกลุ่มที่ชัดเจนมากขึ้น และค่าเฉลี่ยของ Silhouette score ในแต่ละกลุ่มมีค่าที่ดีกว่าการใช้ค่า K = 10

2. การแบ่งกลุ่มด้วย เคเนียร์เรสเนเบอร์

โมเดล เคเนียร์เรสเนเบอร์ ได้รับการสอนด้วยผลลัพธ์จาก คลัสเตอร์แบบเคมีน ด้วยค่า K = 3 การทดสอบด้วยข้อมูลทดสอบ พบว่า Neighbor ที่มีประสิทธิภาพสูงสุดคือ Neighbor = 9 โดยมีค่า F1-Score และ Accuracy score สูงที่สุดโดยการทดสอบจะทำการทดสอบกับผลเฉลยคือ กลุ่มที่ได้จากคลัสเตอร์แบบเคมีนเพื่อให้แบบจำลองเคเนียร์เรสเนเบอร์ มีประสิทธิภาพในการจำแนกกลุ่มข้อมูลใหม่ในอนาคต โดยที่ไม่ทำให้กลุ่มข้อมูลเดิมคลาดเคลื่อน

3. การกำหนดราคาส่วนลด

ผลการวิเคราะห์กำไรเฉลี่ยในแต่ละกลุ่มหลังการแบ่งกลุ่มพบว่า มีความแตกต่างกันอย่างชัดเจน โดยกลุ่มที่มีกำไรสูงที่สุด คลัสเตอร์ 1 ได้รับโปรโมชันลดร้อยละ 12 กลุ่มที่มีกำไรปานกลาง คลัสเตอร์ 2 ได้รับโปรโมชันลดร้อยละ 8 และกลุ่มที่มีกำไรต่ำที่สุด คลัสเตอร์ 3 ไม่ได้รับโปรโมชัน การกำหนดราคาส่วนลดแบบจัดกลุ่มมีผลให้กำไรรวมเฉลี่ยเพิ่มขึ้นเป็นร้อยละ 38.25 เมื่อเปรียบเทียบกับวิธีการกำหนดราคาแบบเดิมที่มีผลกำไรรวมเฉลี่ยร้อยละ 15.80

4. ตัวแปรที่มีผลต่อการกำหนดราคาค่าขนส่ง

ตัวแปรทั้งหมดนี้สามารถแบ่งออกได้เป็นสองกลุ่ม คือกลุ่มที่เกี่ยวกับต้นทุน และกลุ่มที่เกี่ยวกับผลกำไร

ตัวแปรเกี่ยวกับต้นทุนจะมีดังนี้ Distance, Driver's trip fee, Children car trip fee, Cost of equipment per set, Oil consumption rate, Toll fees, Allowance, Total oil consumption, Calculate the cost of oil, Expenses, Cost car, Cost distance car

ตัวแปรเกี่ยวกับผลกำไรจะมีดังนี้ Pricer, Residual, Percent residual ในการกำหนดราคาสินค้าหรือบริการรวมทั้งค่าขนส่ง จะสามารถคำนวณได้จากราคาต้นทุน และทำการราคาผลกำไรที่ต้องการเข้าไปเพื่อให้สามารถกำหนดราคาได้อย่างเหมาะสม

5. ตัวแปรที่สำคัญในการจัดกลุ่มลูกค้า

จะขึ้นอยู่กับหลายปัจจัยร่วมกันโดยตัวแปรที่มีผลต่อการจัดกลุ่มคือตัวแปร

Employing company ชื่อบริษัทลูกค้าที่ถูกนำมาแปลงเพื่อให้เห็นจำนวนรอบในการใช้บริการ

Price ราคาที่ลูกค้าต้องจ่าย การพิจารณาราคาจะช่วยให้สามารถแยกกลุ่มลูกค้าที่มีระดับการใช้จ่ายต่างกันออกไป

Residual การพิจารณากำไรคงเหลือจะช่วยให้สามารถแยกกลุ่มลูกค้าที่สร้างกำไรให้กับบริษัทได้มากน้อยต่างกันออกไป ซึ่งเป็นข้อมูลสำคัญในการวางกลยุทธ์การตลาดและการบริการ

Percent residual กำไรคงเหลือร้อยละแสดงถึงกำไรในรูปแบบเปอร์เซ็นต์ที่ได้รับจากลูกค้า ซึ่งจะช่วยให้เห็นภาพรวมของผลตอบแทนจากลูกค้าแต่ละกลุ่มได้ชัดเจนขึ้น

อภิปรายผล

งานวิจัยนี้และงานวิจัยที่เกี่ยวข้องมีการใช้เทคนิค Machine learning และตัวชี้วัดประสิทธิภาพในการประเมินผลการวิเคราะห์ข้อมูล อย่างไรก็ตาม มีความแตกต่างในเป้าหมายของการวิจัย วิธีการจัดการกับข้อมูลไม่สมดุล ข้อมูลและตัวแปรที่ใช้ในการวิจัย และวิธีการวิเคราะห์ข้อมูล ซึ่งทำให้งานวิจัยแต่ละชิ้นมีความเฉพาะเจาะจงในบริบทที่แตกต่างกัน

การแบ่งกลุ่มลูกค้าด้วย คลัสเตอร์แบบเคมีน และเคเนียร์เรสเนเบอร์ สามารถปรับปรุงประสิทธิภาพการกำหนดราคาส่วนลดได้อย่างมีนัยสำคัญ การเลือกค่า K ที่เหมาะสมโดยใช้ Elbow method และการตรวจสอบด้วย Silhouette score ช่วยให้ได้กลุ่มลูกค้าที่มีความชัดเจนและสามารถนำไปใช้งานได้จริง การใช้เคเนียร์เรสเนเบอร์เพื่อเสริมการแบ่งกลุ่มช่วยเพิ่มความแม่นยำในการ

จำแนกลูกค้า การวิเคราะห์กำไรเฉลี่ยในแต่ละกลุ่มยังชี้ให้เห็นถึงความสำคัญของการจัดกลุ่มลูกค้าในการกำหนดราคาส่วนลด

ข้อเสนอแนะ

1. การปรับปรุงและทดสอบอัลกอริทึม

ควรมีการทดสอบการแบ่งกลุ่มด้วยอัลกอริทึมอื่น เช่น Hierarchical clustering หรือ DBSCAN เพื่อเปรียบเทียบผลลัพธ์และประสิทธิภาพการปรับปรุงฟีเจอร์ที่ใช้ในการแบ่งกลุ่มด้วยการทดลองเพิ่มหรือลดฟีเจอร์บางตัวเพื่อหาชุดฟีเจอร์ที่ดีที่สุด

2. การวิเคราะห์ข้อมูลเพิ่มเติม

ควรมีการวิเคราะห์ข้อมูลเพิ่มเติมเพื่อหาตัวแปรที่มีผลต่อการใช้บริการของลูกค้าอย่างละเอียด เช่น พฤติกรรมการใช้งานของลูกค้าในช่วงเวลาทั้งหมด หรือปัจจัยภายนอกที่อาจมีผลต่อการใช้งาน

3. การใช้ผลการแบ่งกลุ่มในการปรับกลยุทธ์ทางการตลาด

ใช้ผลการแบ่งกลุ่มในการพัฒนากลยุทธ์ทางการตลาดที่เฉพาะเจาะจงต่อกลุ่มลูกค้าแต่ละกลุ่ม เช่น การจัดโปรโมชันเฉพาะสำหรับลูกค้ากลุ่มที่มีกำไรสูงเพื่อเพิ่มความประทับใจ

4. การปรับปรุงระบบการเก็บข้อมูล

ควรปรับปรุงระบบการเก็บข้อมูลให้มีความแม่นยำและครอบคลุมทุกตัวแปรที่สำคัญเพื่อให้ได้ข้อมูลที่สามารถนำมาใช้ในการวิเคราะห์ได้อย่างมีประสิทธิภาพ

การวิจัยนี้แสดงให้เห็นถึงความสำคัญของการใช้เทคโนโลยี Machine learning ในการปรับปรุงและพัฒนาการบริหารจัดการลูกค้า และการกำหนดราคาส่วนลดให้มีประสิทธิภาพมากยิ่งขึ้น

บรรณานุกรม

กาญจนมาส เปลี่ยนสกุล. (2564). *การจัดกลุ่มลูกค้าบริษัทยานยนต์ด้วยข้อมูลประชากร โดยใช้เทคนิคการเรียนรู้ของเครื่อง*. เข้าถึงได้จาก http://irithesis.swu.ac.th/dspace/bitstream/123456789/1702/1/gs631130341.pdf

คำนาย อภิปรัชญาสกุล. (2559). *การจัดการขนส่งและการกระจายสินค้าเชิงกลยุทธ์ Strategic transport and distribution management*. กรุงเทพฯ: โฟกัสมีเดีย แอนด์พับลิชชิ่ง.

ฆฤณ ชินประสาทศักดิ์, ปรีชาพล อินทรสุข, นันทนัช ฟูสามป๊อก และณัฐพล สนพะเนาว์. (2566). *แก่นแท้ของ Deep Learning AI ฉบับอธิบายด้วยภาพ*. สมุทรปราการ: วัน โอ ไฟว์ ดิจิตอลพริ้นติ้ง.

ทวีศักดิ์ เทพพิทักษ์. (2548). *การจัดการลอจิสติกส์ Logistics management*. กรุงเทพฯ: ธรรมกมล การพิมพ์.

นิภาพร ชนะมาร. (2566). การจัดกลุ่มข้อมูลเศรษฐกิจ ครัวเรือน ด้วยการประยุกต์ใช้เทคนิคการจัดกลุ่มแบบ K-Means Clustering. *วารสารวิทยาศาสตร์วิศวกรรมศาสตร์และเทคโนโลยี, 3*(1), 18-31.

บัญชา ปะสีละเตสัง. (2564). *สร้างการเรียนรู้สำหรับ AI ด้วย Python machine learning*. กรุงเทพฯ: วี พริ้นท์.

พสธร วิทยาปรีชาพล. (2563). การประเมินราคาเสนอขายห้องชุดด้วย Deep neural network และ K-means clustering algorithm. เข้าถึงได้จาก https://digital.library.tu.ac.th/tu_dc/frontend/Info/z39?ctx_ver=Z39.88-2004&genre=Book&title.

ภัครพล อาจอาษา. (2564). *การวิเคราะห์คุณภาพน้ำด้วยเทคนิคการจัดกลุ่มข้อมูล*. เข้าถึงได้จาก http://olarik.it.msu.ac.th/wp-content/uploads/2021/10/complete_62011284504.pdf

วณิชา แผลงรักษา และนิเวศ จิระวิชิตชัย. (2562). การแบ่งกลุ่มลูกค้าโดยใช้เทคนิคการทำคลัสเตอร์แบบเคมีน สำหรับการบริหารลูกค้าสัมพันธ์. *PKRU SciTech Journal, 3*(2), 1-10.

วาทิตย์ คำพรมา เพชร อิ่มทองคำ และจักรชัย โสอินทร์. (2562). แบบจำลองการวิเคราะห์ความรู้สึกแบบผสมสำหรับความคิดเห็นต่อโรงแรมในประเทศไทยโดย ใช้ K-means และ K-NN. *Walailak Procedia, 2019*(7), NCIs34-NCIs34.

บรรณานุกรม (ต่อ)

สุริยัน เขตบรรจง. (2564). เทคโนโลยีปัญญาประดิษฐ์ AI กับการแบ่งกลุ่มข้อมูลผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้นประถมศึกษา ปีที่ 5 โดยใช้วิธีการเรียนรู้การแบ่งกลุ่มข้อมูลชนิด K-mean (Artificial Intelligence (AI) Technology and the Categorization of Fifth Grade Students' Academic Achievements Using K-means Clustering Learning Algorithm). *Journal of Education Studies, 49*(2), 1-11.

สุจินดา เจียมศรีพงษ์. (2558). *BUSINESS ECONOMICS เศรษฐศาสตร์ธุรกิจ*. พิษณุโลก: สำนักพิมพ์มหาวิทยาลัยนเรศวร.

สุชาวดี ปลั่งศรี. (2562). *การจัดกลุ่มขนาดผลิตภัณฑ์โดยการใช้ K-means Clustering เพื่อลดต้นทุนบรรจุภัณฑ์*. เข้าถึงได้จาก http://ithesis-ir.su.ac.th/dspace/bitstream/123456789/2926/1/61405317.pdf

Aggarwal, C. (2014). *Data Mining The Textbook*. New York USA.

Business & Technology. (2023). *Paperless Office คืออะไร ต้องมีตัวช่วยอะไรบ้างในการทำงาน*. Retrieved from https://aigencorp.com/tools-for-paperless-office/

Blanchard, T., Behera, B. & Bhatnagar, P. (2019). *Data science for marketing analytics*. Packt Publishing.

Boonsoong. (2020). *Machine learning classification*. Retrieved from https://medium.com/@patcharaboonsoong/season-4-machine-learning-classification-a8b66fa4ddb5.

Chengz. (2019). *วัดประสิทธิภาพ Model จาก Confusion Matrix*. Retrieved from https://medium.com/@cheng3374/%E0%B8%A7%E0%B8%B1%E0%B8%94%E0%B8%9B%E0%B8%A3%E0%B8%B0%E0%B8%AA%E0%B8%B4%E0%B8%97%E0%B8%98%E0%B8%B4%E0%B8%A0%E0%B8%B2%E0%B8%9E-model-%E0%B8%88%E0%B8%B2%E0%B8%81-confusion-matrix-69d391bcd48

Christopher. (1992). *Logistics & Supply Chain Management*. Great Britain pearson education limited.

Ditto. (2021). *ลดต้นทุนกระดาษ ปรับองค์กรเข้าสู่ดิจิทัลด้วย E-Document เอกสารอิเล็กทรอนิกส์*. Retrieved from https://www.dittothailand.com/dittonews/e-document/

JWD GROUP. (2020). *จุดกำเนิดและความเป็นมาของ โลจิสติกส์ ที่คุณไม่เคยรู้*. เข้าถึงได้จาก https://jwd-group.com/th/knowledge_bases/aboutlogistics/

บรรณานุกรม (ต่อ)

Joshi. (2022). *What is Clustering in Machine Learning: Types and Methods*. Retrieved from https://www.analytixlabs.co.in/blog/types-of-clustering-algorithms/

Kotler, Philip & Lane, K. K. (2016). *Marketing management.* เข้าถึงได้จาก https://so04.tci-thaijo.org/index.php/jkbu/article/view/104793

Kong Ruksiam. (2020). Machine learning (EP.5) *การคำนวณเพื่อนบ้านใกล้สุด (K-nearest Neighbors)*. Retrieved from https://kongruksiam.medium.com/%E0%B8%AA%E0%B8%A3%E0%B8%B8%E0%B8%9B-machine-learning-ep-4-%E0%B9%80%E0%B8%9E%E0%B8%B7%E0%B9%88%E0%B8%AD%E0%B8%E0%B8%94-k-nearest-neighbors-787665f7c09d

Kumar. (2021). *Elbow method vs Silhouette score - Which is Better*?. Retrieved from https://vitalflux.com/elbow-method-silhouette-score-which-better/

Kittimsakdi nicit. (2023). *K-Nearest Neighbors (KNN) จัดกลุ่มข้อมูลและทำนายค่าใน Machine Learning ด้วย Python*. Retrieved from https://kittimasak.com/k-nearest-neighbors-knn-machine-learning-python/

Myerson. (2015). *Supply Chain and Logistics Management Made Easy*. Pearson Education, Inc. : Paul Boge.

Novianti, Setyorini, & Rafflesia. (2017). K-Means cluster analysis in earthquake epicenter clustering. *International Journal of Advances in Intelligent Informatics, 3*(2), 81-89.

Narut Soontranon. (2024). *Feature Engineering อาวุธลับของ Data Scientist*. Retrieved from https://www.nerd-data.com/feature-engineering-data-prep/

Ohmae. (1982). *The Mind of the Strategist*. United States of America: McGraw-Hill.

Porter. (1985). *Competitive advantage creating and sustaining superior performance.* Avenue of the American New York, NY

Tomar. (2023). *Stop Using Elbow method in K-Means Clustering*. Retrieved from https://builtin.com/data-science/elbow-method

Wiyono, Wibowo, Hidayatullah, & Dairoh. (2020). Comparative study of KNN, SVM and Decision tree algorithm for student's performance prediction. *(IJCSAM) International Journal of Computing Science and Applied Mathematics, 6*(2), 50-53.

บรรณานุกรม (ต่อ)

Xiong. (2020). *Going Paperless: A Journey Worth Taking*. Retrieved from https://forbes.com /sites/forbestechcouncil/2020/04/02/going-paperless-a-journey-worth-taking/?sh= 409765895ca1

Yildiz. (2023). *Fundamentals and Strategies. Logistics and Supply Chain Management*. Retrieved from https://www.researchgate.net/publication/369302425

Zhu. (2018). *Comparison of Model Performance for Basic and Advanced Modeling Approaches to Crime Prediction*. Retrieved from https://www.researchgate.net/publication/ 329215185_Comparison_of_Model_Performance_for_Basic_and_Advanc ed_Modeling_Approaches_to_Crime_Prediction

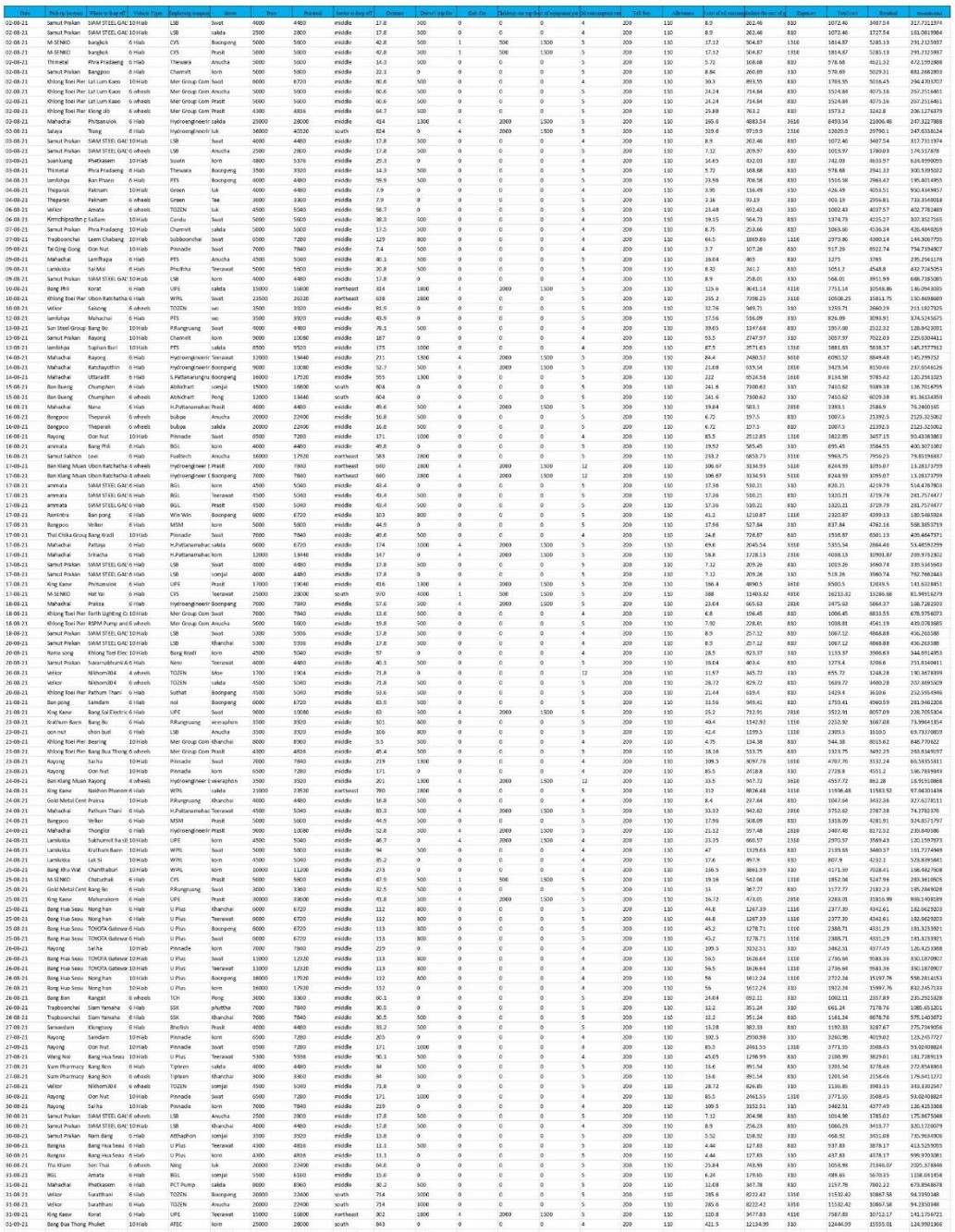
บรรณานุกรม

ภาคผนวก

ภาคผนวก ก

ตัวอย่างข้อมูลที่ใช้ในการศึกษา

ภาพที่ 26 ตัวอย่างการบันทึกข้อมูลบริษัทกรณีศึกษา ก่อนแก้ไขข้อมูล



ภาพที่ 27 ตัวอย่างการบันทึกข้อมูลบริษัทกรณีศึกษา หลังแก้ไขข้อมูล

ภาคผนวก ข

การวิเคราะห์ด้วยคลัสเตอร์แบบเคมีน และเคเนียร์เรสเนเบอร์

Employing company	Price	Ressidual	Percent
LSB	3942.4	191.94	5.117772
LSB	2464	-1286.46	-34.3014
CYS	4928	1185.13	31.66367
CYS	4928	1185.13	31.66367
Thewara	4928	2306.32	87.97107
Charnvit	4928	2636.31	115.0378
Mer Group Company Limited	5913.6	1104.05	22.95537
Mer Group Company Limited	4928	1297.16	35.72617
Mer Group Company Limited	4928	1297.16	35.72617
Mer Group Company Limited	4238.08	517.88	13.92076
Hydroengineering	24640	10506.46	74.33707
Hydroengineering	35481.6	13711.7	62.98467
LSB	3942.4	191.94	5.117772
LSB	2464	-233.97	-8.67208
Suwin	4730.88	1195.85	33.82857
Thewara	3449.6	827.92	31.57975
PTS	3942.4	326.82	9.039214
Green	3942.4	936.91	31.17329
Green	2956.8	974.61	49.16834
TOZEN	4435.2	1345.77	43.56046
Candu	4928	670.27	15.74243
Charnvit	4928	1189.34	31.81193
Subboonchai	6406.4	-363.46	-5.3688
Pinnacle	6899.2	3407.94	97.61347
PTS	4435.2	1259.2	39.64736
Phuthha	4928	2168.8	78.60249
LSB	3942.4	696.39	21.45372
UPE	14784	2392.86	19.31106
WPIL	23161.6	4773.35	25.9587
TOZEN	3449.6	-129.11	-3.60772
PTS	3449.6	684.51	24.75543
P.Rungruang	3942.4	-1296.28	-24.7444
Charnvit	8870.4	1442.43	19.4189
PTS	8377.6	245.97	3.024855
Hydroengineering	11827.2	2126.68	21.92336
Hydroengineering	8870.4	3413.86	62.56456
S.Pattanarungruang	15769.6	585.02	3.852724
Abhichart	14784	-166.62	-1.11447
Abhichart	11827.2	-3123.42	-20.8916
H.Pattanamahachai	3942.4	-1446.7	-26.8449
Pinnacle	6406.4	-1626.45	-20.2475

ภาพที่ 28 ตัวอย่างชุดข้อมูลที่ใช้ในการวิเคราะห์ คลัสเตอร์แบบเคมีน

Employing company	Price	Ressidual	Percent	Cluster	Predicted Cluster
LSB	3942.4	191.94	5.117772	2	2
LSB	2464	-1286.46	-34.3014	2	2
CYS	4928	1185.13	31.66367	1	1
CYS	4928	1185.13	31.66367	1	1
Thewara	4928	2306.32	87.97107	1	1
Charnvit	4928	2636.31	115.0378	1	1
Mer Group Company Limited	5913.6	1104.05	22.95537	1	1
Mer Group Company Limited	4928	1297.16	35.72617	1	1
Mer Group Company Limited	4928	1297.16	35.72617	1	1
Mer Group Company Limited	4238.08	517.88	13.92076	2	1
Hydroengineering	24640	10506.46	74.33707	0	0
Hydroengineering	35481.6	13711.7	62.98467	0	0
LSB	3942.4	191.94	5.117772	2	2
LSB	2464	-233.97	-8.67208	2	2
Suwin	4730.88	1195.85	33.82857	2	2
Thewara	3449.6	827.92	31.57975	2	2
PTS	3942.4	326.82	9.039214	2	2
Green	3942.4	936.91	31.17329	1	1
Green	2956.8	974.61	49.16834	1	1
TOZEN	4435.2	1345.77	43.56046	2	2
Candu	4928	670.27	15.74243	1	1
Charnvit	4928	1189.34	31.81193	1	1
Subboonchai	6406.4	-363.46	-5.3688	2	2
Pinnacle	6899.2	3407.94	97.61347	1	1
PTS	4435.2	1259.2	39.64736	1	1
Phuthha	4928	2168.8	78.60249	1	1
LSB	3942.4	696.39	21.45372	1	1
UPE	14784	2392.86	19.31106	0	0
WPIL	23161.6	4773.35	25.9587	0	0
TOZEN	3449.6	-129.11	-3.60772	2	2
PTS	3449.6	684.51	24.75543	2	2
P.Rungruang	3942.4	-1296.28	-24.7444	2	2
Charnvit	8870.4	1442.43	19.4189	1	1
PTS	8377.6	245.97	3.024855	2	2
Hydroengineering	11827.2	2126.68	21.92336	0	0
Hydroengineering	8870.4	3413.86	62.56456	0	0
S.Pattanarungruang	15769.6	585.02	3.852724	0	0
Abhichart	14784	-166.62	-1.11447	1	1
Abhichart	11827.2	-3123.42	-20.8916	1	1
H.Pattanamahachai	3942.4	-1446.7	-26.8449	1	1
Pinnacle	6406.4	-1626.45	-20.2475	2	2

ภาพที่ 29 ตัวอย่างชุดข้อมูลที่ได้จากการวิเคราะห์ คลัสเตอร์แบบเคมีนและเคเนียร์เรสเนเบอร์

```python
import pandas as pd
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
import matplotlib.pyplot as plt
from sklearn.metrics import f1_score

max_n = 5
max_n = max_n+1
# อ่านข้อมูลจากไฟล์
file_path = "data.xlsx"
data = pd.read_excel(file_path)
# ลบช่องว่างในชื่อบริษัท
data['Employing company'] = data['Employing company'].str.strip()
encoder_company = LabelEncoder()
data['Employing company'] = encoder_company.fit_transform(data['Employing company'])

train_data = data

# เลือกฟีเจอร์ที่ต้องการใช้
features = train_data[['Employing company','pricereal','RESIDUAL','PercentResiduan']]

# เข้ารหัสคอลัมน์ที่เป็นอักษร
encoder = LabelEncoder()
for col in features.columns:
    if features[col].dtype == 'object':
        features[col] = encoder.fit_transform(features[col].astype(str).apply(lambda x: x.strip()))

# ปรับขนาดฟีเจอร์ให้มีค่าเฉลี่ย 0 และความแปรปรวน 1
scaler = StandardScaler()
scaled_features = scaler.fit_transform(features)

# กำหนดจำนวนคลัสเตอร์
n_clusters = 10  # จำนวนคลัสเตอร์
kmeans = KMeans(n_clusters=n_clusters)
kmeans.fit(scaled_features)

# เพิ่มข้อมูลคลัสเตอร์ลงใน train_data
train_data['Cluster'] = kmeans.labels_

# อ่านข้อมูลสำหรับการทดสอบ
test_data = pd.read_excel(file_path)

# ลบช่องว่างในชื่อบริษัท
test_data['Employing company'] = test_data['Employing company'].str.strip()
# encoder_company = LabelEncoder()
test_data['Employing company'] = encoder_company.fit_transform(test_data['Employing company'])

# เลือกฟีเจอร์สำหรับข้อมูลทดสอบ
test_features = test_data[['Employing company','pricereal','RESIDUAL','PercentResiduan']]
for col in test_features.columns:
    if test_features[col].dtype == 'object':
        test_features[col] = encoder.transform(test_features[col].astype(str).apply(lambda x: x.strip()))

# ปรับขนาดฟีเจอร์ของข้อมูลทดสอบ
test_scaled_features = scaler.transform(test_features)
test_data['Cluster'] = kmeans.predict(test_scaled_features)
```

ภาพที่ 30 การเลือกฟีเจอร์และทำข้อมูลให้เป็นมาตรฐานและจัดกลุ่มด้วยเคมีนคลัสเตอร์และเคเนียร์เบอร์

```python
# ปรับขนาดฟีเจอร์ของข้อมูลทดสอบ
test_scaled_features = scaler.transform(test_features)
test_data['Cluster'] = kmeans.predict(test_scaled_features)

accuracy_scores = []
f1_scores = []
confusion_matrices = []

# วนลูปเพื่อทดสอบ KNN ด้วยจำนวนเพื่อนบ้านที่แตกต่างกัน
for n_neighbors in range(2, 5):
    knn = KNeighborsClassifier(n_neighbors=9)
    knn.fit(scaled_features, train_data['Cluster'])
    test_data['Predicted Cluster'] = knn.predict(test_scaled_features)

# แมพชื่อบริษัทกับรหัสที่เข้ารหัส
company_mapping = {i: label for i, label in enumerate(encoder_company.classes_)}
test_data['Employing company'] = test_data['Employing company'].apply(lambda x: company_mapping[x])

# บันทึกผลลัพธ์เป็นไฟล์ CSV
test_data.to_csv("result_alldata/output_alldata.csv", index=False)

# นับความถี่ของบริษัทในแต่ละคลัสเตอร์ที่ถูกทำนาย
company_cluster_freq = test_data.groupby(['Predicted Cluster', 'Employing company']).size().unstack(fill_value=0)

# หาคลัสเตอร์ที่มีความถี่สูงสุดสำหรับแต่ละบริษัท
company_cluster_max_freq = company_cluster_freq.idxmax(axis=0)

# แสดงผลคลัสเตอร์ที่แนะนำสำหรับแต่ละบริษัท
print("คลัสเตอร์ที่แนะนำสำหรับแต่ละบริษัทตามความถี่สูงสุด:")
print(company_cluster_max_freq)
company_cluster_max_freq = company_cluster_freq.idxmax(axis=0)
company_cluster_max_freq.to_csv("company_cluster_recommendation_alldata.csv", header=["Recommended Cluster"])

# คำนวณ Inertia สำหรับแต่ละจำนวนคลัสเตอร์ (K)
inertia = []
n_clusters = 10
for k in range(1, n_clusters+1):
    kmeans = KMeans(n_clusters=k)
    kmeans.fit(scaled_features)
    inertia.append(kmeans.inertia_)

# พล็อตกราฟ Elbow Method เพื่อหา K ที่เหมาะสม
plt.plot(range(1, n_clusters+1), inertia, marker='o')
plt.xlabel('จำนวนคลัสเตอร์ (K)')
plt.ylabel('Inertia')
plt.title('Elbow Method สำหรับหาค่า K ที่เหมาะสม')
plt.xticks(range(1, n_clusters+1))
plt.grid(True)
plt.show()
```

ภาพที่ 30 (ต่อ)

ประวัติย่อของผู้วิจัย

ชื่อ-สกุล	นายพีรวิชญ์ ภิบาล	
วัน เดือน ปี เกิด	3 ตุลาคม พ.ศ. 2539	
สถานที่เกิด	จังหวัดสมุทรปราการ	
สถานที่อยู่ปัจจุบัน	9/ 331 หมู่ 5 ถนนพุทธรักษา ซอยนาคดี ตำบลแพรกษาใหม่ อำเภอเมือง จังหวัดสมุทรปราการ	
ประวัติการศึกษา	พ.ศ. 2563	คณะโลจิสติกส์และเทคโนโลยีการบิน (การจัดการโลจิสติกส์) มหาวิทยาลัยเซาธ์อีสท์บางกอก
	พ.ศ. 2567	วิทยาศาสตรมหาบัณฑิต (การจัดการโลจิสติกส์และโซ่อุปทาน) มหาวิทยาลัยบูรพา

มหาวิทยาลัยบูรพา
BURAPHA UNIVERSITY