

การวิเคราะห์การจัดกลุ่มเพื่อการจัดการมูลฝอยติดเชื้อที่เหมาะสมโดยใช้เทคนิคเหมืองข้อมูล

Cluster Analysis for Optimal Infectious Waste Management

Using Data Mining Techniques

รัชพล กิตติคุณ^{1*} และ จุฑามาศ จินตนา²
Ratchapon Kittikun^{1*} and Jutamat Jintana²

¹คณะนวัตกรรม เทคโนโลยี และการสร้างสรรค์ มหาวิทยาลัยฟาร์อีสเทอร์น

*Correspondence author E-mail: northratchapon@gmail.com

²ปร.ด. (วิศวกรรมอุตสาหการ) อาจารย์ ดร. คณะนวัตกรรม เทคโนโลยี และการสร้างสรรค์ มหาวิทยาลัยฟาร์อีสเทอร์น

E-mail: Jutamat@feu.edu

Received: November 18, 2024 / Revised: March 26, 2025 / Accepted: April 4, 2025

บทคัดย่อ

การจัดการมูลฝอยติดเชื้อเป็นประเด็นสำคัญที่ส่งผลกระทบต่อสิ่งแวดล้อม สุขอนามัย และต้นทุนการดำเนินงาน โดยเฉพาะในช่วงการระบาดของ COVID-19 ที่ทำให้ปริมาณมูลฝอยติดเชื้อเพิ่มขึ้นอย่างมีนัยสำคัญ การวิจัยเชิงวิเคราะห์นี้มีวัตถุประสงค์เพื่อ 1) เพื่อศึกษาคุณลักษณะ (Attribute) ของลูกค้าที่ใช้บริการจัดเก็บมูลฝอยติดเชื้อ และ 2) วิเคราะห์การจัดกลุ่มข้อมูลลูกค้าด้วยวิธีการ K-Means Algorithms โดยใช้ข้อมูลลูกค้า 540 ราย ประกอบด้วยสถานพยาบาลขนาดใหญ่ 37 แห่ง และสถานพยาบาลขนาดเล็ก 503 แห่ง ของบริษัท นิวโซค่านวย เชียงใหม่ จำกัด การวิเคราะห์ที่ใช้โปรแกรม Orange Data Mining พิจารณาจากตัวแปรปริมาณมูลฝอยติดเชื้อ (0.5-12 กก./เดือน) ระยะทาง (0.3-15 กม.) และความถี่ในการจัดเก็บ ผลการวิจัยพบว่าสามารถจัดกลุ่มลูกค้าได้ 3 กลุ่ม คือ 1) กลุ่มที่มีปริมาณมูลฝอยติดเชื้อต่ำ (85.66%) โดยมีปริมาณเฉลี่ย 1.2 กก./เดือน ระยะทางใกล้เคียง 1.8 กม. และความถี่การจัดเก็บปานกลาง 2) กลุ่มที่มีปริมาณมูลฝอยติดเชื้อสูง (11.43%) โดยมีปริมาณเฉลี่ย 8.5 กก./เดือน ระยะทางปานกลางเฉลี่ย 4.2 กม. และความถี่การจัดเก็บต่ำ และ 3) กลุ่มที่มีปริมาณมูลฝอยติดเชื้อปานกลาง (2.91%) โดยมีปริมาณเฉลี่ย 3.7 กก./เดือน ระยะทางไกลเฉลี่ย 8.5 กม. และความถี่การจัดเก็บสูง ผลการจัดกลุ่มมีความแม่นยำสูงด้วยค่า Silhouette Score 0.873 การเปรียบเทียบประสิทธิภาพก่อนและหลังการปรับปรุง พบการพัฒนาใน 4 ด้านหลัก ได้แก่ การจัดการเวลา การใช้ความจุรถ ความถี่การจัดเก็บ และการจัดการพื้นที่ ผลการวิจัยนี้สามารถนำไปประยุกต์ใช้ในการวางแผนเส้นทางการจัดเก็บที่เหมาะสม การออกแบบแพ็คเกจบริการเฉพาะกลุ่ม และการเพิ่มประสิทธิภาพการจัดการทรัพยากร เพื่อลดต้นทุนการดำเนินงานและยกระดับคุณภาพการให้บริการ อย่างไรก็ตาม การศึกษาในอนาคตควรพิจารณาปัจจัยเพิ่มเติม เช่น ประเภทของมูลฝอยติดเชื้อ และการวิเคราะห์ข้อมูลร่วมกับระบบ GIS เพื่อการพัฒนาที่ครอบคลุมและยั่งยืนมากยิ่งขึ้น

คำสำคัญ : การจัดกลุ่ม, มวลถ่วง, การวิเคราะห์ข้อมูล, K-Means, เหมืองข้อมูล

Abstract

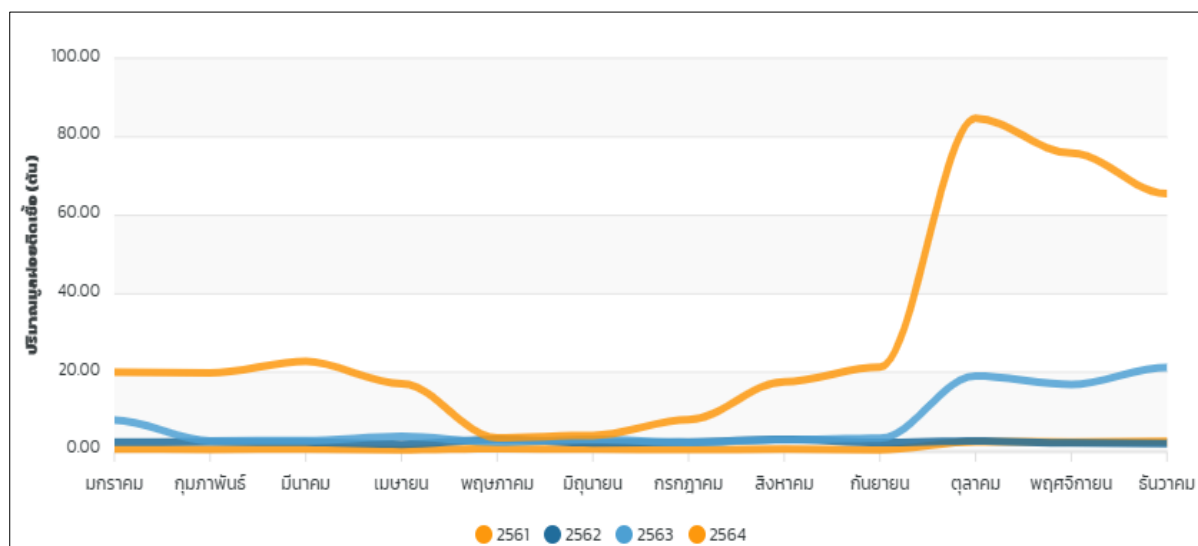
Infectious waste management is a critical issue affecting the environment, public health, and operational costs, particularly during the COVID-19 pandemic, which significantly increased infectious waste volumes. This analytical research aimed to 1) to study the attributes of customers using infectious waste collection services, and 2) analyze customer data clustering using K-Means Clustering method. The study utilized data from 540 customers, comprising 37 large healthcare facilities and 503 small healthcare facilities of New Chok Amnuay Chiang Mai Company Limited. The analysis was conducted through Orange Data Mining software, considering variables including infectious waste quantity (0.5-12 kg/month), distance (0.3-15 km), and collection frequency. The results revealed three distinct customer clusters: Cluster 1 (85.66%) characterized by low waste quantity (average 1.2 kg/month), short distance (average 1.8 km), and moderate collection frequency; Cluster 2 (11.43%) with high waste quantity (average 8.5 kg/month), moderate distance (average 4.2 km), and low collection frequency; and Cluster 3 (2.91%) with moderate waste quantity (average 3.7 kg/month), long distance (average 8.5 km), and high collection frequency. The clustering showed high efficiency with a Silhouette Score of 0.873. The comparison of operational efficiency before and after implementation showed improvements in four main areas: time management, vehicle capacity utilization, collection frequency, and area management. These findings can be applied to optimize collection routes, design cluster-specific service packages, and improve resource management, leading to reduced operational costs and enhanced service quality. However, future studies should consider additional factors such as infectious waste types and GIS integration for more comprehensive and sustainable development.

Keywords: Clustering, Infectious Waste, Data Analysis, K-Means, Data Mining

บทนำ

มูลฝอยติดเชื้อเป็นปัญหาสำคัญที่ส่งผลกระทบต่อสุขอนามัยและสิ่งแวดล้อม โดยเฉพาะในช่วงการระบาดของ COVID-19 ที่ทำให้ปริมาณมูลฝอยติดเชื้อเพิ่มขึ้นอย่างมาก (Department of Health, 2024a) การจัดการมูลฝอยติดเชื้อมีความท้าทายทั้งด้านปริมาณที่เพิ่มขึ้น ค่าใช้จ่ายสูง และผลกระทบต่อสิ่งแวดล้อม กรมควบคุมมลพิษได้เสนอแนวทางการจัดการ เช่น การคัดแยก การขนส่งที่ปลอดภัย และการกำจัด

อย่างถูกวิธี (Pollution Control Department, 2022) โดยภาพที่ 1 แสดงให้เห็นถึงแนวโน้มที่เพิ่มขึ้นของปริมาณมลพิษที่สะสมทั่วประเทศ โดยเฉพาะในปี 2564 ที่มีการระบาดของ COVID-19 อย่างรุนแรง ซึ่งเป็นปัจจัยสำคัญที่ตอกย้ำถึงความจำเป็นในการวางแผนระบบจัดการมลพิษที่ให้มีประสิทธิภาพมากยิ่งขึ้น



ภาพที่ 1 กราฟแสดงภาพรวมปริมาณมลพิษที่สะสมที่เกิดขึ้น รายเดือน (ตัน)

ที่มา: Department of Health (2024b)

อย่างไรก็ตาม หากไม่สามารถจำแนกลูกค้าออกเป็นกลุ่มที่มีลักษณะใกล้เคียงกันได้ ก็จะทำให้การวางแผนเส้นทางหรือการกำหนดรอบการจัดเก็บไม่สอดคล้องกับความต้องการจริง อันนำไปสู่ต้นทุนที่ไม่จำเป็น เช่น การใช้รถขนาดใหญ่ไปในพื้นที่ที่มีปริมาณขยะน้อย หรือการเก็บขยะในพื้นที่ห่างไกลบ่อยเกินความจำเป็น เป็นต้น ดังนั้น งานวิจัยนี้จึงเลือกใช้เทคนิค K-Means Algorithms ซึ่งเป็นเทคนิคการจัดกลุ่ม (clustering) ที่มีประสิทธิภาพในการแยกกลุ่มข้อมูลที่มีลักษณะใกล้เคียงกัน โดยไม่ต้องพึ่งพาตัวแปรอธิบายล่วงหน้า ช่วยให้สามารถจำแนกกลุ่มลูกค้าที่มีลักษณะใกล้เคียงกัน เช่น ปริมาณมลพิษที่สะสม ระยะทาง และความถี่ในการจัดเก็บ เพื่อใช้ในการออกแบบบริการและวางแผนเส้นทางได้อย่างมีประสิทธิภาพ

บริษัท นิวโซคอานวย เชียงใหม่ จำกัด เป็นผู้ให้บริการจัดการมลพิษที่สะสมในภาคเหนือของประเทศไทย ซึ่งประสบปัญหาสำคัญหลายประการที่ส่งผลกระทบต่อประสิทธิภาพการดำเนินงานและต้นทุน ปัญหาแรกคือการวางแผนเส้นทางที่ไม่มีประสิทธิภาพ เนื่องจากขาดข้อมูลเชิงลึกเกี่ยวกับลักษณะการใช้บริการของลูกค้าแต่ละราย ส่งผลให้เกิดการสิ้นเปลืองเชื้อเพลิงและเวลาในการให้บริการ ประการที่สอง การคาดการณ์ปริมาณมลพิษที่สะสมที่ไม่แม่นยำ ทำให้บางครั้งรถเก็บขนมีความจุไม่เพียงพอหรือมากเกินไป ส่งผลให้เสียโอกาสในการให้บริการหรือสิ้นเปลืองทรัพยากรโดยไม่จำเป็น นอกจากนี้ การขาดการจัดกลุ่มข้อมูลที่มีประสิทธิภาพ ทำให้ไม่สามารถให้บริการที่เหมาะสมกับความต้องการเฉพาะของลูกค้าแต่ละกลุ่มได้ ปัญหาสุดท้ายคือความไม่สอดคล้องระหว่างความถี่ในการเก็บขนกับปริมาณมลพิษของลูกค้าบางราย

ทำให้เกิดการสะสมของมูลฝอยติดเชื้อเกินกำหนด หรือการเก็บขนที่ไม่จำเป็น ซึ่งส่งผลกระทบต่อต้นทุนการดำเนินงานและคุณภาพการให้บริการ

เทคนิคการทำเหมืองข้อมูล (Data Mining) ในยุคดิจิทัลได้แสดงศักยภาพอย่างชัดเจนในการยกระดับประสิทธิภาพการดำเนินธุรกิจ โดยเฉพาะอย่างยิ่งในบริบทของการจัดการมูลฝอยติดเชื้อ (Chen, Chiang, & Storey, 2012) การประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลในธุรกิจนี้สามารถช่วยในการวิเคราะห์รูปแบบการเกิดมูลฝอยติดเชื้อ

การวิจัยนี้มุ่งเน้นการแก้ไขปัญหาการจัดการมูลฝอยติดเชื้อของบริษัท นิวโซคอานวย เชียงใหม่ จำกัด โดยใช้กระบวนการมาตรฐาน CRISP-DM เป็นแนวทางในการวิเคราะห์ข้อมูล วัตถุประสงค์หลักของการศึกษานี้คือ 1) เพื่อศึกษาโครงสร้างข้อมูลคุณลักษณะของลูกค้าที่ใช้บริการจัดเก็บมูลฝอยติดเชื้อ 2) เพื่อวิเคราะห์การจัดกลุ่มข้อมูลคุณลักษณะ ของลูกค้า ด้วยวิธีการ K-Means Clustering ผลการศึกษานี้จะช่วยให้บริษัทสามารถปรับปรุงการดำเนินงาน วิเคราะห์ข้อมูลลูกค้าปัจจุบัน และยกระดับคุณภาพการให้บริการได้อย่างมีประสิทธิภาพ

วัตถุประสงค์การวิจัย

1. เพื่อศึกษาคุณลักษณะ (Attribute) ของลูกค้าที่ใช้บริการจัดเก็บมูลฝอยติดเชื้อ
2. เพื่อวิเคราะห์การจัดกลุ่มข้อมูลคุณลักษณะของลูกค้า ด้วยวิธีการ K-Means Algorithms

วิธีการวิจัย

การวิจัยนี้แบ่งออกเป็น 2 ระยะ ตามวัตถุประสงค์ของการศึกษา ดังนี้

การวิจัยระยะที่ 1: เพื่อศึกษาคุณลักษณะ (attribute) ของลูกค้าที่ใช้บริการจัดเก็บมูลฝอยติดเชื้อ ข้อมูลที่ใช้ในการวิจัยครั้งนี้ มี 2 กลุ่ม คือ

- 1) ข้อมูลจำนวนลูกค้าจากสถานพยาบาลขนาดใหญ่ที่ตั้งอยู่ในจังหวัดเชียงใหม่ จำนวน 37 ราย
- 2) ข้อมูลจำนวนลูกค้าจากสถานพยาบาลขนาดเล็กในจังหวัดเชียงใหม่ จำนวน 503 ราย

โดยขั้นตอนในการวิจัยมีดังนี้

1. รวบรวมข้อมูลลูกค้าทั้งหมดจำนวน 540 ราย จากฐานข้อมูลของบริษัท นิวโซคอานวย เชียงใหม่ จำกัด โดยแบ่งเป็นลูกค้าจากสถานพยาบาลขนาดใหญ่ จำนวน 37 ราย และลูกค้าจากสถานพยาบาลขนาดเล็กจำนวน 503 ราย โดยข้อมูลที่เก็บรวบรวมจากลูกค้า ประกอบด้วย

- 1.1 รหัสลูกค้า เป็นรหัสเฉพาะที่ใช้ระบุตัวตนของลูกค้าแต่ละราย โดยไม่เปิดเผยข้อมูลส่วนบุคคล
- 1.2 ระยะทาง ระยะทางจากจุดจัดเก็บมูลฝอยติดเชื้อของบริษัทไปยังสถานที่ตั้งของลูกค้า

(กิโลเมตร)

- 1.3 ปริมาณน้ำหนักมูลฝอยติดเชื้อเฉลี่ย ค่าเฉลี่ยของน้ำหนักมูลฝอยติดเชื้อที่ลูกค้าแต่ละราย

(กิโลกรัม)

1.4 ความถี่ในการจัดเก็บ จำนวนครั้งที่มีการจัดเก็บมูลฝอยติดเชื้อในแต่ละเดือน ซึ่งบริษัทจะกำหนดไว้ตามแผนการให้บริการ

1.5 ประเภทของสถานพยาบาล การจำแนกลักษณะของสถานพยาบาลตามขนาดหรือประเภทการให้บริการ

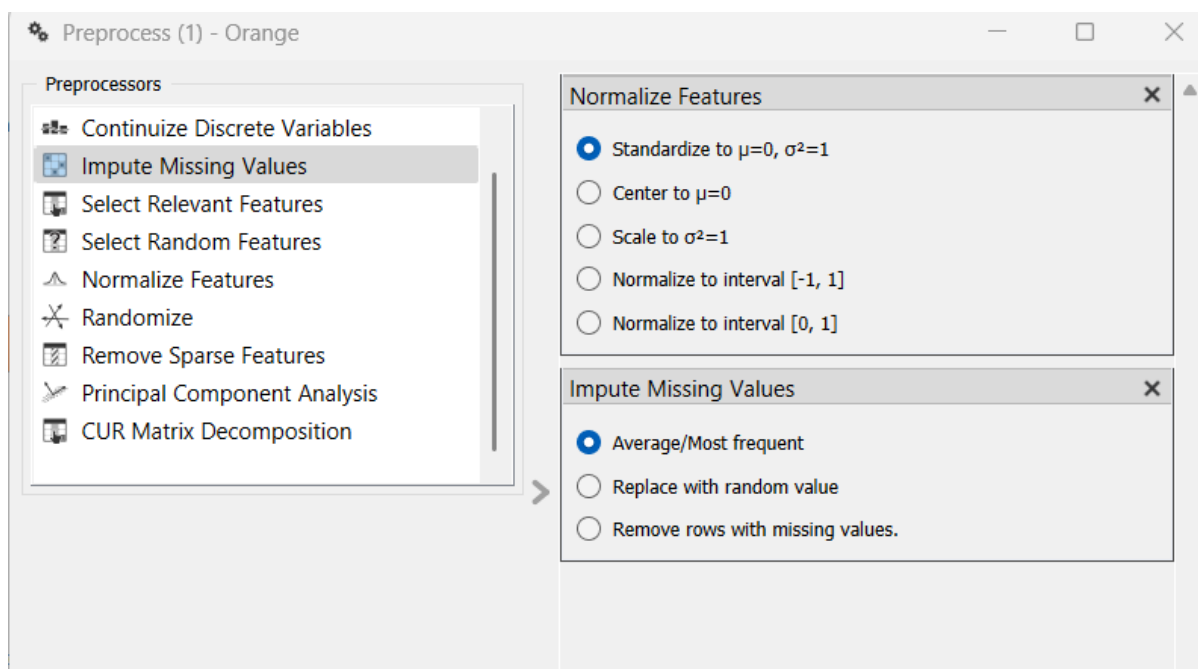
จากการรวบรวมข้อมูลของลูกค้าทั้งหมด 540 ราย พบว่ามีตัวแปรคุณลักษณะที่สำคัญในการนำมาวิเคราะห์การจัดกลุ่ม รายละเอียดของตัวแปรทั้งหมด แสดงไว้ในตารางที่ 1

ตารางที่ 1 ตัวแปรคุณลักษณะ Attribute พร้อมรายละเอียด ชนิด และค่าของตัวแปร

Attribute	Description	Type	Values
ID	รหัสลูกค้า	Nominal	รหัสที่ไม่ซ้ำกันสำหรับลูกค้าแต่ละราย
Waste_Avg	ปริมาณมูลฝอยติดเชื้อเฉลี่ยต่อเดือน	Integer	ค่าจำนวนกิโลกรัม (เช่น 0.5, 2.8, 12)
Distance	ระยะทางจากบริษัทถึงสถานที่ลูกค้า	Float	ระยะทางเป็นกิโลเมตร (เช่น 0.3-100 กิโลเมตร)
Collection_Freq	ความถี่ในการจัดเก็บ	Nominal	1 = ต่ำ (1 ครั้ง/เดือน), 2 = ปานกลาง (2-3 ครั้ง/เดือน), 3 = สูง (4 ครั้ง/เดือนขึ้นไป)
Customer_Type	ประเภทของสถานพยาบาล	Nominal	1 = สถานพยาบาลขนาดเล็ก, 2 = สถานพยาบาลขนาดกลาง, 3 = สถานพยาบาลขนาดใหญ่

2. ตรวจสอบและทำความสะอาดข้อมูล มีการดำเนินการจัดเตรียมข้อมูลอย่างเป็นระบบ เพื่อให้ข้อมูลมีความสมบูรณ์และเหมาะสมต่อการวิเคราะห์ ในขั้นตอนการเตรียมข้อมูล ได้ดำเนินการทำความสะอาดข้อมูลโดยใช้โมดูล Preprocess ของโปรแกรม Orange โดยในส่วนของการจัดการค่าขาดหาย (missing values) เลือกใช้วิธีแทนค่าด้วยค่าเฉลี่ยหรือค่าที่พบมากที่สุด (average/most frequent) เพื่อเติมเต็มข้อมูลที่หายไป และป้องกันไม่ให้ข้อมูลชุดใดต้องถูกตัดทิ้ง จากนั้นได้ทำการแปลงข้อมูลด้วยการ Normalize ซึ่งทำให้ข้อมูลมีค่าเฉลี่ยเป็นศูนย์ และส่วนเบี่ยงเบนมาตรฐานเป็นหนึ่ง ส่งผลให้ตัวแปรทั้งหมดมีมาตราส่วนเท่ากัน และเหมาะสมสำหรับการวิเคราะห์ด้วยเทคนิค K-Means

โดยภาพที่ 2 แสดงการเตรียมข้อมูล (data preprocessing) ในโปรแกรม Orange Data Mining



ภาพที่ 2 การตั้งค่าขั้นตอนการเตรียมข้อมูลในโปรแกรม Orange Data Mining

ตัวอย่าง: ลูกค้าจำนวน 540 ราย ที่บริษัทให้บริการจริง

เครื่องมือวิจัย: Google Sheets และโปรแกรม Orange Data Mining สำหรับรวบรวมและจัดรูปแบบข้อมูล

การวิเคราะห์: วิเคราะห์โครงสร้างข้อมูลเชิงพรรณนา เพื่อเตรียมข้อมูลให้พร้อมสำหรับการจัดกลุ่มในระยะที่ 2

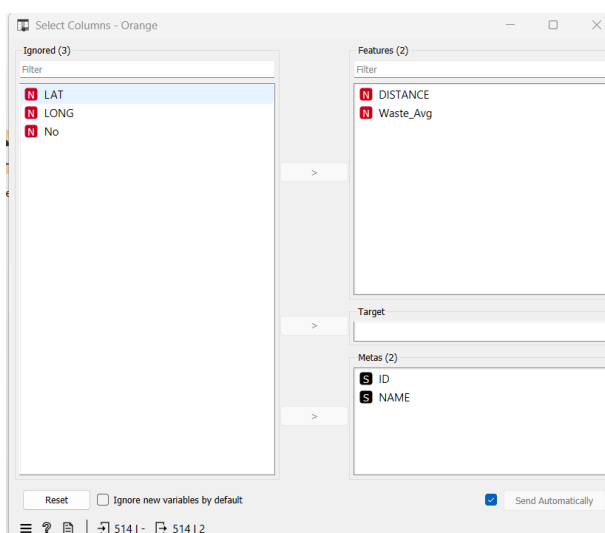
ก่อนเข้าสู่กระบวนการจัดกลุ่มด้วย K-Means ข้อมูลที่ผ่านการจัดเตรียมอย่างสมบูรณ์แล้ว ได้แก่ การเติมค่าขาดหาย การจัดการค่าผิดปกติ และการ Normalize ด้วย Z-Score ได้ถูกจัดเก็บอยู่ในรูปแบบตารางที่ประกอบด้วยตัวแปรหลัก ได้แก่ รหัสลูกค้า (Customer_ID), ปริมาณมูลฝอยติดเชื้อเฉลี่ย (Waste_Avg), ระยะทาง (Distance), ความถี่ในการจัดเก็บ (Frequency), ประเภทสถานพยาบาล (Type) และค่าที่ผ่านการ Normalize แล้ว (Waste_Std, Distance_Std) ดังแสดงใน ภาพที่ 3

NO	ID	Waste_Avg	Collection_Freq	DISTANCE	PACKAGE	Customer_Type
1	195	20	1	31.2	A1	1
2	452	31	1	73.7	A1	1
3	729	32	1	73.5	A1	1
4	848	14	1	73.7	A1	1
5	1821	7	1	12.1	A1	1
6	29	16	1	16.5	A1	1
7	138	13	1	47.8	A1	2
8	175	10	1	21.1	A1	2
9	354	10	1	15.1	A1	1
10	386	18	1	22.2	A1	1
11	434	5	1	20.1	A1	1
12	520	5	1	17.6	A1	1
13	910	91	1	20.2	A1	1
14	1230	1	2	17.7	A1	1
15	1752	6	1	12.4	A1	1
16	1758	16	1	12.3	A1	1
17	0305	5	1	9.6	A2	1
18	0493	11	1	26.2	A2	1
19	0675	39	1	59.7	A1	1
20	1067	9	1	142	A1	1

ภาพที่ 3 แสดงตารางโครงสร้างข้อมูลของลูกค้า ที่ผ่านการจัดเตรียมเบื้องต้นก่อนนำเข้าสู่กระบวนการจัดกลุ่มด้วย K-Means

การวิจัยระยะที่ 2: การวิเคราะห์การจัดกลุ่มข้อมูลด้วย K-Means

1. นำข้อมูลที่ผ่านการเตรียมไว้มาทำการจัดกลุ่มด้วยเทคนิค K-Means โดยใช้โปรแกรม Orange Data Mining โดยเลือกเฉพาะตัวแปร “Waste_Avg” และ “Distance” มาใช้ในการจัดกลุ่มตามวัตถุประสงค์ของการวิจัย ดังแสดงในภาพที่ 4



ภาพที่ 4 การเลือกตัวแปรสำหรับการจัดกลุ่มข้อมูลในโปรแกรม Orange Data Mining

ในการจัดกลุ่มข้อมูลด้วย K-Means ผู้วิจัยเลือกใช้ตัวแปรจำนวน 2 ตัว ได้แก่ “Waste_Avg” และ “Distance” ซึ่งเป็นตัวแปรเชิงปริมาณที่มีผลโดยตรงต่อการให้บริการและต้นทุนของบริษัท

การเลือกตัวแปรเพียง 2 ตัวนี้ ช่วยให้สามารถแสดงผลในลักษณะ 2 มิติที่เข้าใจง่าย ขณะเดียวกัน ยังคงสะท้อนความแตกต่างที่สำคัญของกลุ่มลูกค้าได้อย่างชัดเจน จึงเหมาะสมกับวัตถุประสงค์ของการวิจัยนี้

2. ทำการกำหนดจำนวนกลุ่ม $K=3$ โดยอิงจากค่าความเหมาะสมของการจัดกลุ่ม (Silhouette Score) สำหรับงานวิจัยนี้ เลือกใช้ K-means clustering เพราะมีจุดเด่นในเรื่องความสามารถในการจัดกลุ่มข้อมูลเชิงตัวเลข เช่น ปริมาณมูลฝอยติดเชื้อ ระยะทาง และความถี่ในการเก็บขยะ อีกทั้งยังสามารถกำหนดจำนวนกลุ่มได้ล่วงหน้า ($K=3$) ซึ่งตรงกับวัตถุประสงค์ของการวิจัย นอกจากนี้ K-means ยังมีการคำนวณที่รวดเร็วและประสิทธิภาพสูงเมื่อวิเคราะห์ข้อมูลขนาดใหญ่ เช่น ชุดข้อมูลลูกค้า 540 รายในงานวิจัยนี้ ผลลัพธ์ที่ได้จาก K-means ยังง่ายต่อการตีความ เช่น การแสดงจุดศูนย์กลางของแต่ละกลุ่ม (Centroids) ที่ช่วยให้เข้าใจลักษณะเฉพาะของลูกค้าแต่ละกลุ่มและสนับสนุนการวางแผนกลยุทธ์ที่ชัดเจน การเลือก K-means clustering แทน DBSCAN หรือ Hierarchical Clustering สอดคล้องกับลักษณะข้อมูลของงานวิจัยที่กระจายตัวอย่างสม่ำเสมอ ไม่มี Noise ที่ชัดเจน และต้องการจำนวนกลุ่มที่กำหนดล่วงหน้า การใช้ K-means ใน Orange Data Mining จึงเป็นทางเลือกที่เหมาะสมและตอบโจทยความต้องการของการวิเคราะห์ข้อมูลในบริบทนี้ ด้วยลักษณะของข้อมูล หากข้อมูลมีการกระจายตัวของจำนวนกลุ่มที่ชัดเจน ยกตัวอย่างเช่นจำนวน $K=3$ ซึ่งเป็นผลที่สอดคล้องกับค่า K ที่เลือก และผลการวิเคราะห์ด้วย K-means clustering บนซอฟต์แวร์ Orange Data Mining แสดงให้เห็นค่า Silhouette Score ซึ่งใช้เป็นตัวชี้วัดคุณภาพของการจัดกลุ่ม โดยค่าที่ใกล้ 1 แสดงถึงการจัดกลุ่มที่ดีมาก เพราะข้อมูลในแต่ละกลุ่มมีความคล้ายคลึงกันสูง (Cohesion) และมีความแตกต่างชัดเจนจากกลุ่มอื่น จากตารางที่ 2 Silhouette Scores ที่ได้ ค่า $K=3$ มีค่า Silhouette Score สูงที่สุดที่ 0.873 ซึ่งบ่งชี้ว่าการจัดกลุ่มเมื่อแบ่งข้อมูลออกเป็น 3 กลุ่มนั้นมีความเหมาะสมที่สุดเมื่อเทียบกับจำนวนกลุ่มอื่น ๆ เช่น $K=4$ (0.809) และ $K=5$ (0.658) ที่มีค่า Silhouette Score ลดลง การลดลงนี้แสดงว่าการเพิ่มจำนวนกลุ่มเกินกว่า 3 ทำให้กลุ่มที่ได้เริ่มทับซ้อนกันมากขึ้นหรือมีการแยกกลุ่มที่ไม่จำเป็น ส่งผลให้ประสิทธิภาพในการจัดกลุ่มลดลง

ตารางที่ 2 ค่า Silhouette Score สำหรับการกำหนดจำนวนคลัสเตอร์

จำนวนคลัสเตอร์ (K)	Silhouette Score	การตีความผลลัพธ์
3	0.873	การจัดกลุ่มมีความเหมาะสมสูงที่สุดในจำนวน 3 กลุ่ม เนื่องจากมีค่า Silhouette Score สูงสุด
4	0.809	การจัดกลุ่มยังเหมาะสม แต่ค่า Silhouette Score ต่ำกว่าค่า $K=3$
5	0.658	ประสิทธิภาพในการจัดกลุ่มเริ่มลดลงเมื่อเพิ่มจำนวนคลัสเตอร์
6	0.653	ความเหมาะสมของการจัดกลุ่มลดลงอย่างชัดเจนเมื่อเทียบกับ K ที่น้อยกว่า
7	0.688	ค่า Silhouette Score เพิ่มขึ้นเล็กน้อย แต่ยังต่ำกว่า $K=3$
8	0.664	ประสิทธิภาพในการจัดกลุ่มยังคงต่ำเมื่อเทียบกับค่า $K=3$

การเลือก $K=3$ สอดคล้องกับบริบทของงานวิจัย เนื่องจากสามารถสะท้อนลักษณะลูกค้า 3 กลุ่มหลักที่แตกต่างกัน ได้แก่ ปริมาณมูลฝอยติดเชื้อ และระยะทาง การเลือกจำนวนกลุ่มนี้ยังช่วยลดความซับซ้อนในการวิเคราะห์และตีความผลลัพธ์ อีกทั้งยังสนับสนุนการวางแผนกลยุทธ์ที่ตรงเป้าหมายสำหรับลูกค้าแต่ละกลุ่ม ดังนั้น การกำหนด $K=3$ ไม่เพียงแต่เหมาะสมในเชิงเทคนิค แต่ยังตอบโจทย์การนำผลไปใช้ในเชิงปฏิบัติสำหรับการพัฒนาการให้บริการของบริษัทอีกด้วย

3. วิเคราะห์ผลการจัดกลุ่มด้วย K-Means โดยพิจารณาความแตกต่างของลักษณะข้อมูลในแต่ละกลุ่ม โดยการประเมินผล (Evaluation) ใช้ตัวชี้วัด เช่น Silhouette Score เพื่อวัดคุณภาพของการจัดกลุ่ม โดยเน้นความแตกต่างระหว่างกลุ่มและความสอดคล้องภายในกลุ่ม ผลลัพธ์การจัดกลุ่มพบว่า ลูกค้าแบ่งออกเป็น 3 กลุ่ม ได้แก่ กลุ่มที่ 1 ลูกค้าที่มีปริมาณมูลฝอยติดเชื้อต่ำ ระยะทางไกล และความถี่ปานกลาง กลุ่มที่ 2 ลูกค้าที่มีปริมาณมูลฝอยติดเชื้อสูง ระยะทางปานกลาง และความถี่ต่ำ และกลุ่มที่ 3 ลูกค้าที่มีปริมาณมูลฝอยติดเชื้อปานกลาง ระยะทางไกล และความถี่สูง ผลการวิเคราะห์นี้ช่วยให้บริษัทสามารถออกแบบ กลยุทธ์การให้บริการ เช่น การปรับเส้นทางหรือรอบการเก็บขยะ และเพิ่มประสิทธิภาพการจัดการมูลฝอย ติดเชื้อในภาพรวม

การวิเคราะห์ข้อมูลจะใช้กระบวนการมาตรฐาน CRISP-DM (Cross-Industry Standard Process for Data Mining) ซึ่งประกอบด้วย 6 ขั้นตอน (Bemthuis, Govers, & Asadi, 2024) โดยใช้โปรแกรม Orange Data Mining ในการวิเคราะห์ข้อมูล ในขั้นตอนการสร้างแบบจำลอง ผู้วิจัยใช้เทคนิค K-means clustering โดยกำหนดค่า K เป็น 3 กลุ่ม (Han, Kamber, & Pei, 2011) เพื่อทดสอบความแตกต่างระหว่างกลุ่มที่ได้จากการวิเคราะห์ (Fisher, 1927) การดำเนินการวิจัยได้รับการอนุมัติจากผู้บริหารบริษัท นิวัชกร อำนวย เชียงใหม่ จำกัด โดยคำนึงถึงจริยธรรมการวิจัยและการรักษาความลับของข้อมูลลูกค้า ผู้วิจัยได้ดำเนินการเก็บรวบรวมข้อมูลตามขั้นตอนที่กำหนด รวมถึงการตรวจสอบความถูกต้องและความสมบูรณ์ของข้อมูลก่อนนำไปวิเคราะห์

ผลการวิจัยนี้จะช่วยให้บริษัทสามารถออกแบบกลยุทธ์การบริการที่เหมาะสมกับลูกค้าแต่ละกลุ่ม ปรับปรุงเส้นทางการเก็บมูลฝอยติดเชื้อ และเพิ่มประสิทธิภาพการจัดการทรัพยากร ซึ่งจะส่งผลต่อการลดต้นทุนการดำเนินงานและส่งเสริมความปลอดภัยด้านสุขอนามัยและสิ่งแวดล้อมในระยะยาว

เครื่องมือวิจัย

1. ฐานข้อมูลลูกค้า

ฐานข้อมูลนี้เป็นข้อมูลที่รวบรวมจากลูกค้าของบริษัทในกระบวนการจัดเก็บมูลฝอยติดเชื้อในพื้นที่จริง ซึ่งถือเป็นข้อมูลสำคัญที่ช่วยให้การวิเคราะห์เป็นไปอย่างถูกต้อง ฐานข้อมูลนี้ประกอบด้วยตัวแปรหลัก เช่น ปริมาณมูลฝอยติดเชื้อที่ลูกค้าแต่ละรายสร้างขึ้น ระยะทางจากสถานพยาบาลถึงบริษัท และความถี่ในการจัดเก็บขยะที่กำหนดตามแผนการดำเนินงานของบริษัท โดยข้อมูลความถี่จะบันทึกจำนวนครั้งที่พนักงานเข้าไปรับมูลฝอยติดเชื้อในแต่ละเดือนให้กับลูกค้าแต่ละราย โดยข้อมูลดังกล่าวสะท้อนพฤติกรรมการให้บริการจริง

และสอดคล้องกับแพ็กเกจที่ลูกค้าเลือกใช้งาน เช่น แพ็กเกจรายเดือนแบบ 1 ครั้ง/เดือน หรือ 4 ครั้ง/เดือน เป็นต้น ข้อมูลเหล่านี้ไม่เพียงแต่สะท้อนถึงพฤติกรรมการใช้บริการของลูกค้าเท่านั้น แต่ยังเป็นตัวชี้วัดที่สำคัญในการจำแนกลักษณะของลูกค้าแต่ละกลุ่ม เพื่อนำไปสู่การวิเคราะห์เชิงลึก

2. ซอฟต์แวร์วิเคราะห์ข้อมูล

ซอฟต์แวร์ที่เลือกใช้ เช่น Orange เป็นเครื่องมือที่มีประสิทธิภาพสูงสำหรับการจัดการข้อมูลขนาดใหญ่และการวิเคราะห์ข้อมูลเชิงลึก ในซอฟต์แวร์ Orange Data Mining มีเครื่องมือสำหรับการจัดกลุ่ม (Clustering) ให้เลือกหลากหลาย เช่น K-means, Hierarchical Clustering, และ DBSCAN แต่ละเทคนิคมีจุดเด่นและการประยุกต์ใช้ที่แตกต่างกัน K-means clustering เหมาะกับข้อมูลที่มีโครงสร้างชัดเจน และระบุกลุ่มเอง ในขณะที่ Hierarchical Clustering ช่วยให้เห็นความสัมพันธ์ระหว่างกลุ่มในลักษณะลำดับขั้น ส่วน DBSCAN เหมาะกับข้อมูลที่มีการกระจุกตัวเป็นรูปทรงหลายเหลี่ยม (Polygon) เช่น มีการกระจายตัวแบบไม่สมมาตรหรือมีจุด Noise จำนวนมาก

3. Google Sheet สำหรับการเก็บรวบรวมและจัดการข้อมูล

Google Sheets เป็นเครื่องมือสำคัญสำหรับการเก็บรวบรวมและจัดการข้อมูลในงานวิจัยเกี่ยวกับการจัดการมูลฝอยติดเชื้อ ด้วยความสามารถในการสร้างและจัดเก็บฐานข้อมูลที่ยืดหยุ่น Google Sheets ช่วยให้สามารถรวบรวมข้อมูลจากลูกค้าได้อย่างเป็นระบบ เช่น ชื่อและที่อยู่ของลูกค้า ปริมาณมูลฝอยติดเชื้อ ระยะทางจากสถานพยาบาลถึงบริษัท ความถี่ในการจัดเก็บ และค่าใช้จ่ายรายเดือน ทีมงานสามารถบันทึกและอัปเดตข้อมูลเหล่านี้ได้แบบเรียลไทม์จากพื้นที่ปฏิบัติงาน ลดข้อผิดพลาดและเพิ่มความสะดวกในการจัดการข้อมูล

นอกจากนี้ Google Sheets ยังมีฟังก์ชันที่ช่วยในกระบวนการจัดการข้อมูล เช่น การใช้สูตรคำนวณสำหรับสรุปค่าเฉลี่ยหรือรวมข้อมูล การกรองและเรียงลำดับข้อมูลเพื่อตรวจสอบประเด็นสำคัญ และการใช้ Conditional Formatting เพื่อเน้นข้อมูลที่ต้องการติดตาม เช่น ลูกค้าที่มีปริมาณมูลฝอยติดเชื้อสูงผิดปกติ หรือความถี่ในการจัดเก็บที่ต้องปรับปรุง การสร้าง Pivot Table และแผนภูมิใน Google Sheets ยังช่วยในการสรุปข้อมูลและนำเสนอผลลัพธ์เบื้องต้นในรูปแบบที่เข้าใจง่าย เช่น กราฟแสดงปริมาณมูลฝอยติดเชื้อตามกลุ่มลูกค้า หรือสัดส่วนประเภทลูกค้าในแต่ละพื้นที่

4. เอกสารการดำเนินงานของบริษัท

เอกสาร เช่น ตารางการจัดเก็บขยะและบันทึกค่าใช้จ่าย สามารถแปลงเป็นข้อมูลดิจิทัลและจัดเก็บใน Google Sheets เพื่อเชื่อมโยงข้อมูลเหล่านี้กับฐานข้อมูลลูกค้าได้อย่างสะดวก นอกจากนี้ Google Sheets ยังช่วยในการวางแผนรอบการให้บริการ หรือการติดตามต้นทุนอย่างเป็นระบบ โดยสามารถแชร์ข้อมูลให้กับทีมงานที่เกี่ยวข้องเพื่อการปรับปรุงกระบวนการดำเนินงานได้แบบเรียลไทม์

การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลในงานวิจัยนี้ใช้กระบวนการมาตรฐาน CRISP-DM (Cross-Industry Standard Process for Data Mining) ซึ่งเป็นแนวทางที่ประกอบด้วยหกขั้นตอน เริ่มจากการทำความเข้าใจข้อมูล (Data Understanding) โดยข้อมูลที่ใช้มาจากฐานข้อมูลลูกค้าของบริษัท นิวิโชคอำนวยการ เชียงใหม่ จำกัด ซึ่งมีรายละเอียดสำคัญ เช่น รหัสลูกค้า ระยะทางจากสถานที่ตั้งของลูกค้าไปยังบริษัท ปริมาณน้ำหนักมูลฝอยติดเชื้อเฉลี่ย และความถี่ในการจัดเก็บ จากนั้นมีการตรวจสอบและสำรวจข้อมูลเบื้องต้น เช่น การกระจายตัวของข้อมูล และการคำนวณสถิติพื้นฐาน (Mean, Median, Standard Deviation) เพื่อประเมินคุณภาพและความครบถ้วนของข้อมูล

ผลการวิจัย

การศึกษานี้มีวัตถุประสงค์เพื่อศึกษาคุณลักษณะ (Attribute) ของลูกค้าที่ใช้บริการจัดเก็บมูลฝอยติดเชื้อและนำผลการวิเคราะห์ไปจัดกลุ่มเพื่อเพิ่มประสิทธิภาพการจัดการ ด้วยการใช้เทคนิคเหมืองข้อมูล K-Means ผลการวิเคราะห์ครอบคลุม 2 ส่วนหลัก ได้แก่ ผลการศึกษาโครงสร้างข้อมูลคุณลักษณะของลูกค้า และผลการวิเคราะห์การจัดกลุ่มข้อมูลโดยใช้วิธีการ K-Means ดังนี้

1. ผลการศึกษาโครงสร้างข้อมูลคุณลักษณะของลูกค้าที่ใช้บริการจัดเก็บมูลฝอยติดเชื้อ

ผลการศึกษาโครงสร้างข้อมูลของลูกค้า จำนวน 540 ราย ซึ่งรวบรวมจากฐานข้อมูลการดำเนินงานจริงของบริษัท นิวิโชคอำนวยการ เชียงใหม่ จำกัด พบว่าข้อมูลมีตัวแปรสำคัญ 3 ตัว ได้แก่

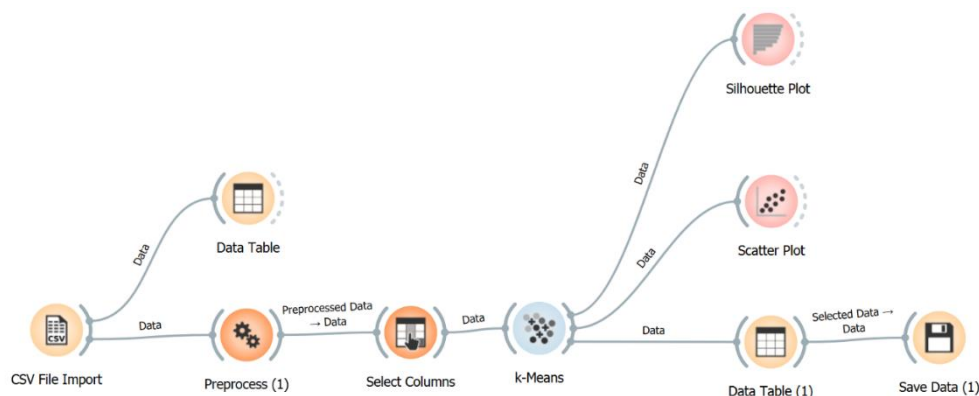
1.1 ปริมาณมูลฝอยติดเชื้อเฉลี่ยต่อเดือน มีค่าตั้งแต่ 0.5 กิโลกรัม ถึง 12 กิโลกรัม โดยมีค่าเฉลี่ยอยู่ที่ 2.8 กิโลกรัม (ส่วนเบี่ยงเบนมาตรฐาน = 1.9)

1.2 ระยะทางจากศูนย์จัดเก็บถึงลูกค้า มีช่วงระยะทางตั้งแต่ 0.3 กิโลเมตร ถึง 15 กิโลเมตร ค่าเฉลี่ยระยะทางอยู่ที่ 3.4 กิโลเมตร (ส่วนเบี่ยงเบนมาตรฐาน = 2.6) โดยผู้วิจัยได้แบ่งช่วงของระยะทางออกเป็น 3 ระดับ ได้แก่ ระยะทางใกล้ (0.3–5.0 กิโลเมตร), ระยะทางปานกลาง (มากกว่า 5.0–10.0 กิโลเมตร), และระยะทางไกล (มากกว่า 10.0–15.0 กิโลเมตร) เพื่อใช้ในการวิเคราะห์และจัดกลุ่มข้อมูลในขั้นตอนต่อไป

1.3 ความถี่ในการจัดเก็บขยะ ถูกกำหนดเป็น 3 ระดับ ได้แก่ ความถี่ต่ำ (1 ครั้ง/เดือน), ความถี่ ปานกลาง (2-3 ครั้ง/เดือน) และความถี่สูง (4 ครั้ง/เดือนขึ้นไป) โดยลูกค้าส่วนใหญ่ (ประมาณ 78%) อยู่ในระดับความถี่ปานกลาง

2. ผลการวิเคราะห์การจัดกลุ่มข้อมูลคุณลักษณะของลูกค้า ด้วยวิธีการ K-Means Clustering

จากภาพที่ 5 กระบวนการวิเคราะห์ข้อมูลด้วย K-Means Clustering ในโปรแกรม Orange Data Mining ตั้งแต่การนำเข้าข้อมูลจนถึงการแสดงผลลัพธ์



ภาพที่ 5 กระบวนการวิเคราะห์ข้อมูลด้วย K-Means ใน Orange Data Mining

โดยการวิเคราะห์ด้วย K-Means Clustering ใช้ค่าพารามิเตอร์ $K=3$ ซึ่งเหมาะสมที่สุดตามค่า Silhouette Score ที่สูงสุด (0.873) ผลการจัดกลุ่มลูกค้าแบ่งออกเป็น 3 กลุ่ม ดังนี้

กลุ่มที่ 1 (C1)

- จำนวน 463 ราย (85.66%)
- ลักษณะเด่น ปริมาณมูลฝอยต่ำ ระยะทางไกล ความถี่การจัดเก็บปานกลาง
- สถิติสำคัญ ปริมาณมูลฝอยเฉลี่ย 1.2 กิโลกรัม/เดือน ระยะทางเฉลี่ย 1.8 กิโลเมตร

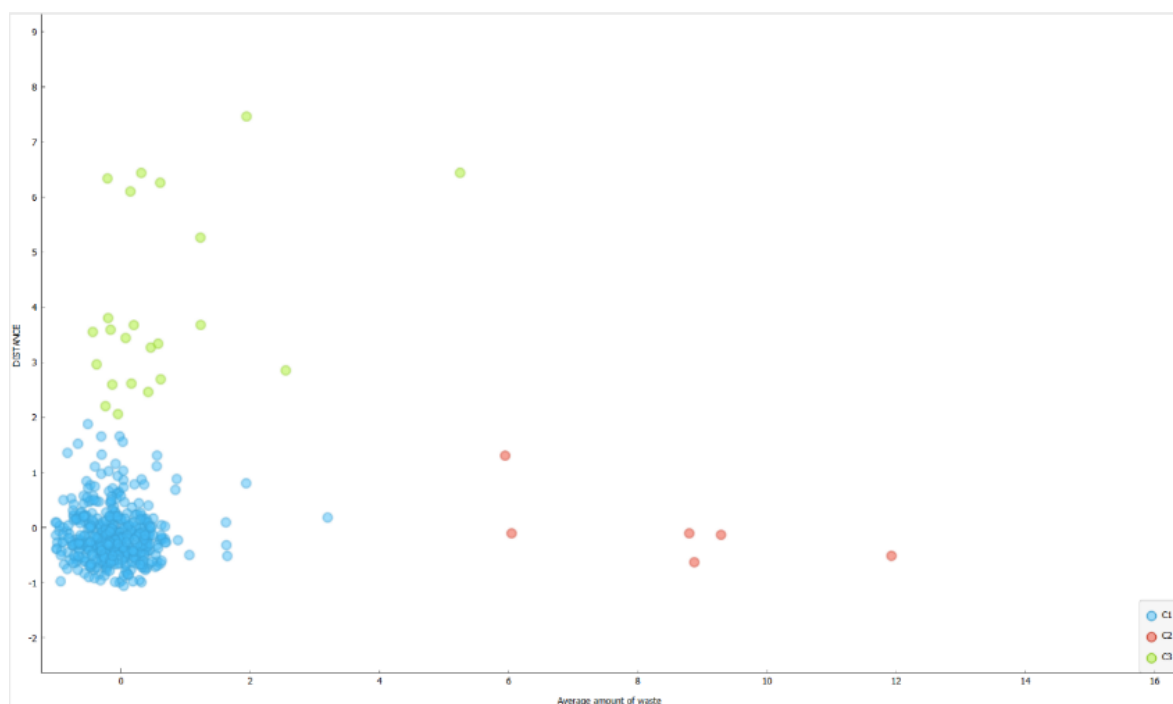
กลุ่มที่ 2 (C2)

- จำนวน 62 ราย (11.43%)
- ลักษณะเด่น ปริมาณมูลฝอยสูง ระยะทางปานกลาง ความถี่การจัดเก็บต่ำ
- สถิติสำคัญ ปริมาณมูลฝอยเฉลี่ย 8.5 กิโลกรัม/เดือน ระยะทางเฉลี่ย 4.2 กิโลเมตร

กลุ่มที่ 3 (C3)

- จำนวน 15 ราย (2.91%)
- ลักษณะเด่น ปริมาณมูลฝอยปานกลาง ระยะทางไกล ความถี่การจัดเก็บสูง
- สถิติสำคัญ ปริมาณมูลฝอยเฉลี่ย 3.7 กิโลกรัม/เดือน ระยะทางเฉลี่ย 8.5 กิโลเมตร

จากผลการจัดกลุ่มข้อมูลด้วยเทคนิค K-Means พบว่ากลุ่มลูกค้าสามารถจำแนกได้ตามลักษณะของ ปริมาณมูลฝอยเฉลี่ยและระยะทางจากบริษัท โดยการกระจายตัวของแต่ละกลุ่มแสดงในภาพที่ 6 ซึ่งแสดงให้เห็นถึงลักษณะเฉพาะของกลุ่มลูกค้าแต่ละกลุ่มได้อย่างชัดเจน



ภาพที่ 6 การกระจายตัวของกลุ่มลูกค้าตามความสัมพันธ์ระหว่างปริมาณมูลฝอยติดเชื้อเฉลี่ย (กิโลกรัม/เดือน) และระยะทาง (กิโลเมตร) จากการวิเคราะห์ด้วย K-Means Clustering

ผลการวิเคราะห์แสดงให้เห็นว่ากลุ่มที่ 1 ซึ่งเป็นกลุ่มลูกค้าส่วนใหญ่ มีระยะทางใกล้และปริมาณมูลฝอยต่ำเหมาะสมกับการให้บริการที่ใช้ทรัพยากรอย่างคุ้มค่า ในขณะที่กลุ่มที่ 2 มีปริมาณมูลฝอยสูงและระยะทางปานกลาง แสดงถึงลักษณะของสถานพยาบาลขนาดใหญ่ที่จัดเก็บขยะได้เบื้องต้น ขณะที่กลุ่มที่ 3 ซึ่งมีจำนวนน้อยที่สุด ต้องการการให้บริการที่มีความถี่สูงเนื่องจากระยะทางไกลและข้อจำกัดด้านการจัดเก็บ

สรุปและอภิปรายผล

การวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์คุณลักษณะของลูกค้าที่ใช้บริการจัดเก็บมูลฝอยติดเชื้อและนำไปจัดกลุ่มด้วยเทคนิคเหมืองข้อมูล K-Means Clustering เพื่อพัฒนาประสิทธิภาพในการจัดการมูลฝอยติดเชื้อ โดยจากการวิเคราะห์ข้อมูลลูกค้าจำนวน 540 ราย พบว่าลูกค้าสามารถแบ่งออกเป็น 3 กลุ่มหลักตามตัวแปรปริมาณมูลฝอยติดเชื้อ ระยะทาง และความถี่ในการจัดเก็บ โดยกลุ่มที่ 1 เป็นลูกค้าส่วนใหญ่ที่มีปริมาณมูลฝอยติดเชื้อต่ำและอยู่ใกล้บริษัท กลุ่มที่ 2 เป็นลูกค้าที่มีปริมาณมูลฝอยติดเชื้อสูงที่สุดและอยู่ในระยะทางปานกลาง ขณะที่กลุ่มที่ 3 เป็นลูกค้าที่อยู่ในพื้นที่รอบนอก มีปริมาณมูลฝอยติดเชื้อปานกลาง แต่ต้องการความถี่ในการจัดเก็บสูง

ผลการจัดกลุ่มแสดงถึงศักยภาพในการนำผลลัพธ์ไปปรับปรุงการให้บริการ โดยสามารถออกแบบเส้นทางการจัดเก็บที่เหมาะสม ลดต้นทุนการดำเนินงาน และตอบสนองความต้องการเฉพาะของลูกค้าในแต่ละกลุ่ม เช่น การเพิ่มความถี่การจัดเก็บในพื้นที่ห่างไกลสำหรับกลุ่มที่ 3 หรือการจัดเส้นทางที่มีประสิทธิภาพใน

เขตเมืองสำหรับกลุ่มที่ 1 นอกจากนี้ ผลการวิเคราะห์ยังแสดงให้เห็นว่าการใช้ K-Means Clustering มีประสิทธิภาพสูง โดยค่า Silhouette Score ที่ 0.873 ซึ่งว่าการจัดกลุ่มนี้เหมาะสมที่สุด

อภิปรายผล

ผลการวิเคราะห์ข้อมูลด้วยเทคนิค K-Means Clustering ในงานวิจัยนี้พบว่า กลุ่มลูกค้าที่มีปริมาณมูลฝอยติดเชื้อมากกลับมีความถี่ในการจัดเก็บน้อย ซึ่งสอดคล้องกับผลการศึกษาของ Han, Kamber and Pei (2011) ที่ชี้ว่าการวางแผนจัดการขยะสำหรับสถานพยาบาลขนาดใหญ่สามารถใช้รอบการเก็บแบบห่างได้ หากมีระบบจัดการภายในที่ดี และแสดงถึงความแตกต่างที่ชัดเจนระหว่างกลุ่มลูกค้าทั้ง 3 กลุ่ม ซึ่งช่วยให้สามารถออกแบบบริการและวางแผนกลยุทธ์การจัดการมูลฝอยติดเชื้อได้อย่างเหมาะสม กลุ่มที่ 1 (C1) ซึ่งมีลูกค้าคิดเป็นสัดส่วน 85.66% ของทั้งหมด เป็นกลุ่มที่มีระยะทางใกล้ ปริมาณมูลฝอยติดเชื้อมาก และความถี่ในการจัดเก็บอยู่ในระดับปานกลาง สะท้อนถึงความต้องการบริการที่ไม่ซับซ้อน เหมาะสำหรับการวางแผนเส้นทางการให้บริการในพื้นที่ใกล้บริษัท ซึ่งจะช่วยลดต้นทุนการเดินทางและเพิ่มความคุ้มค่าในการดำเนินงาน กลุ่มที่ 2 (C2) ซึ่งมีลูกค้า 11.43% มีปริมาณมูลฝอยติดเชื้อสูงที่สุด ระยะทางปานกลาง และความถี่การจัดเก็บต่ำ กลุ่มนี้ส่วนใหญ่เป็นสถานพยาบาลขนาดใหญ่ เช่น โรงพยาบาลที่สามารถจัดการมูลฝอยติดเชื้อได้เบื้องต้น การลดความถี่ในการจัดเก็บจะช่วยเพิ่มประสิทธิภาพการใช้ทรัพยากร เช่น การลดจำนวนรอบการขนส่ง ในขณะที่กลุ่มที่ 3 (C3) มีลูกค้าเพียง 2.91% แต่เป็นกลุ่มที่อยู่ในพื้นที่รอบนอก มีปริมาณมูลฝอยติดเชื้อปานกลาง และต้องการความถี่การจัดเก็บสูง กลุ่มนี้ต้องการการให้บริการแบบเฉพาะกิจ เช่น การจัดเส้นทางพิเศษเพื่อเข้าถึงพื้นที่ที่ห่างไกล

นอกจากนี้ ลักษณะการกระจายตัวของกลุ่มลูกค้าที่แบ่งออกตามระยะทางและปริมาณมูลฝอยติดเชื้อยังสอดคล้องกับการวิเคราะห์ของ Chen, Chiang and Storey (2012) ซึ่งใช้การทำเหมืองข้อมูลในการจำแนกกลุ่มลูกค้าเพื่อปรับปรุงกระบวนการโลจิสติกส์ โดยพบว่าการใช้ข้อมูลระยะทางและปริมาณร่วมกันสามารถลดต้นทุนและเพิ่มประสิทธิภาพได้อย่างมีนัยสำคัญ

ผลการวิเคราะห์แสดงให้เห็นว่า K-Means Clustering ที่ใช้ค่า $K = 3$ มีค่า Silhouette Score สูงสุดที่ 0.873 ซึ่งบ่งชี้ถึงคุณภาพการจัดกลุ่มที่เหมาะสมที่สุด เมื่อเทียบกับการเพิ่มจำนวนคลัสเตอร์อื่น ๆ ผลลัพธ์นี้ช่วยให้เข้าใจถึงลักษณะพฤติกรรมของลูกค้าแต่ละกลุ่ม และสามารถนำไปวางแผนกลยุทธ์การให้บริการได้อย่างมีประสิทธิภาพ เช่น การเพิ่มความถี่การจัดเก็บในพื้นที่ที่ต้องการบริการสูง หรือการลดรอบการขนส่งในกลุ่มที่มีปริมาณขยะสะสมสูงแต่ต้องการความถี่การจัดเก็บต่ำ อย่างไรก็ตาม ข้อจำกัดของการวิเคราะห์ในครั้งนี้คือการใช้ตัวแปรเพียง 3 ตัว ได้แก่ ปริมาณมูลฝอยติดเชื้อ ระยะทาง และความถี่การจัดเก็บ ซึ่งอาจไม่ครอบคลุมถึงปัจจัยอื่น ๆ ที่อาจส่งผลต่อการจัดกลุ่ม เช่น ประเภทขยะหรือค่าใช้จ่าย นอกจากนี้ ความแม่นยำของข้อมูล เช่น พิกัดสถานที่ตั้ง หรือความถูกต้องของการบันทึกข้อมูลปริมาณมูลฝอยติดเชื้อ ก็อาจส่งผลต่อคุณภาพของผลลัพธ์

การประยุกต์ใช้ผลการวิจัย ผลการวิเคราะห์การจัดกลุ่มลูกค้าด้วย K-Means Clustering สามารถนำไปประยุกต์ใช้ในการพัฒนาการให้บริการที่เหมาะสมสำหรับลูกค้าแต่ละกลุ่ม สำหรับกลุ่มที่ 1 ซึ่งเป็นกลุ่มที่มีปริมาณมูลฝอยติดเชื้อต่ำและระยะทางใกล้ บริษัทฯ สามารถปรับปรุงประสิทธิภาพการให้บริการโดยใช้รถขนาดเล็กที่ประหยัดน้ำมัน วางแผนเส้นทาง ซึ่งจะแบ่งพื้นที่เป็นโซนและกำหนดจุดรวมขยะย่อย นอกจากนี้ยังสามารถพัฒนาระบบจองคิวออนไลน์ให้ลูกค้าเลือกช่วงเวลาที่จะสะดวก และเสนอแพ็คเกจราคาพิเศษสำหรับการจัดเก็บแบบรวมกลุ่มในพื้นที่เดียวกัน

สำหรับกลุ่มที่ 2 ซึ่งมีปริมาณมูลฝอยติดเชื้อสูงและระยะทางปานกลาง บริษัทฯ ควรจัดรถขนาดใหญ่พิเศษเฉพาะกลุ่มเพื่อรองรับปริมาณมูลฝอยติดเชื้อที่สูง พร้อมติดตั้งระบบติดตาม GPS และเซ็นเซอร์วัดปริมาณในถังขยะเพื่อการบริหารจัดการแบบ real-time การจัดทำแผนฉุกเฉินและการจัดเตรียมรถสำรองกรณีปริมาณเกินคาดการณ์ก็เป็นสิ่งจำเป็น รวมถึงการเสนอส่วนลดพิเศษสำหรับการจัดเก็บในช่วง off-peak เพื่อกระจายปริมาณงานให้สมดุล

ในส่วนของกลุ่มที่ 3 ซึ่งมีปริมาณมูลฝอยติดเชื้อปานกลางแต่อยู่ในระยะทางไกล บริษัทฯ ควรใช้รถขนาดกลางที่มีระบบควบคุมอุณหภูมิพิเศษสำหรับการขนส่งระยะไกล วางแผนเส้นทางแบบ Multi-stop routing เพื่อเพิ่มประสิทธิภาพในการเก็บหลายจุดในเส้นทางเดียวกัน การจัดตั้งสถานีพักขยะย่อยในพื้นที่ห่างไกลจะช่วยลดความเสี่ยงในการขนส่ง และการเสนอระบบสมาชิกพิเศษที่รับประกันการจัดเก็บตรงเวลาจะช่วยสร้างความมั่นใจให้กับลูกค้า

การเปรียบเทียบประสิทธิภาพก่อนและหลังการปรับปรุง

การเปรียบเทียบประสิทธิภาพการดำเนินงานหลังจากการนำผลการวิเคราะห์การจัดกลุ่มด้วย K-Means Clustering มาประยุกต์ใช้ แสดงให้เห็นการพัฒนาในหลายด้าน ดังนี้

1. ด้านการจัดการเวลา ก่อนการปรับปรุง บริษัทใช้รูปแบบการจัดเก็บแบบเดียวกันสำหรับลูกค้าทุกราย โดยไม่ได้คำนึงถึงความแตกต่างของปริมาณมูลฝอยติดเชื้อและระยะทาง หลังการปรับปรุง มีการแบ่งการจัดเก็บตามลักษณะของแต่ละกลุ่ม โดยกลุ่ม C1 ซึ่งเป็นกลุ่มใหญ่ที่สุด (85.66%) ใช้การจัดเก็บแบบรวมกลุ่มในพื้นที่เดียวกัน กลุ่ม C2 (11.43%) ปรับเป็นการจัดเก็บครั้งละมากแต่ลดความถี่ลง และกลุ่ม C3 (2.91%) ใช้การจัดเก็บแบบวางแผนเส้นทางรวม ที่สามารถจัดเก็บได้หลายจุดในเส้นทางเดียวกัน

2. ด้านการใช้ความจุรถ การปรับปรุงทำให้สามารถเลือกใช้รถขนส่งที่เหมาะสมกับปริมาณมูลฝอยติดเชื้อของแต่ละกลุ่ม จากเดิมที่ใช้รถขนาดเดียวกันสำหรับการจัดเก็บ เปลี่ยนเป็นการใช้รถขนาดเล็กสำหรับกลุ่ม C1 ที่มีปริมาณเฉลี่ย 1.2 กิโลกรัมต่อเดือน รถขนาดใหญ่สำหรับกลุ่ม C2 ที่มีปริมาณเฉลี่ย 8.5 กิโลกรัมต่อเดือน และรถขนาดกลางสำหรับกลุ่ม C3 ที่มีปริมาณเฉลี่ย 3.7 กิโลกรัมต่อเดือน

3. ด้านความถี่การจัดเก็บ มีการปรับความถี่การจัดเก็บให้สอดคล้องกับความต้องการของแต่ละกลุ่ม จากเดิมที่ใช้ความถี่แบบตายตัว เป็นการกำหนดความถี่ปานกลาง (2-3 ครั้งต่อเดือน) สำหรับกลุ่ม C1 ความถี่ต่ำ (1 ครั้งต่อเดือน) สำหรับกลุ่ม C2 และความถี่สูง (4 ครั้งต่อเดือนขึ้นไป) สำหรับกลุ่ม C3

4. ด้านการจัดการพื้นที่ การจัดการพื้นที่ที่ได้รับการปรับปรุงจากระบบกระจายตัวแบบไม่มีโซนเป็นการแบ่งพื้นที่ตามระยะทางที่ชัดเจน พื้นที่ใกล้ (ระยะทางเฉลี่ย 1.8 กิโลเมตร) พื้นที่ปานกลาง (ระยะทางเฉลี่ย 4.2 กิโลเมตร) และพื้นที่ไกล (ระยะทางเฉลี่ย 8.5 กิโลเมตร)

ประสิทธิภาพของการจัดกลุ่มได้รับการยืนยันด้วยค่า Silhouette Score ที่ 0.873 ซึ่งแสดงถึงความแม่นยำสูงในการจำแนกกลุ่ม โดยรวมการปรับปรุงนี้ส่งผลให้การดำเนินงานมีประสิทธิภาพมากขึ้น ทั้งในด้านการใช้ทรัพยากรและการจัดการเวลา อย่างไรก็ตาม ควรมีการติดตามและประเมินผลอย่างต่อเนื่องเพื่อการพัฒนาที่ยั่งยืนต่อไป

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลการวิจัยไปใช้ประโยชน์

1. ใช้ผลการจัดกลุ่มลูกค้าเพื่อวางแผนเส้นทางการจัดเก็บที่เหมาะสม เช่น เพิ่มรอบการจัดเก็บสำหรับลูกค้ากลุ่มที่มีความถี่สูง และจัดเส้นทางเฉพาะสำหรับพื้นที่ที่อยู่ห่างไกล
2. ออกแบบแพ็คเกจบริการเฉพาะกลุ่ม เช่น แพ็คเกจราคาประหยัดสำหรับลูกค้าที่มีปริมาณขยะน้อย และแพ็คเกจแบบครบวงจรสำหรับลูกค้าที่มีปริมาณขยะสูง
3. จัดสรรทรัพยากร เช่น รถเก็บขนและบุคลากร ให้เหมาะสมกับลักษณะของแต่ละกลุ่มลูกค้า เพื่อลดต้นทุนการดำเนินงานและเพิ่มประสิทธิภาพ
4. ใช้ผลการวิเคราะห์เพื่อสนับสนุนการตัดสินใจเชิงกลยุทธ์ เช่น การขยายพื้นที่ให้บริการ หรือการเพิ่มการลงทุนในเทคโนโลยีใหม่ ๆ เพื่อพัฒนาคุณภาพการให้บริการ
5. นำผลการจัดกลุ่มไปผสานกับระบบติดตามผลแบบเรียลไทม์ (real-time monitoring) เพื่อปรับปรุงการดำเนินงานและการให้บริการอย่างต่อเนื่อง

ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

1. เพิ่มตัวแปรที่เกี่ยวข้อง เช่น ประเภทของมูลฝอยติดเชื้อ ค่าใช้จ่ายในการจัดเก็บ และผลกระทบต่อสิ่งแวดล้อม เพื่อให้การวิเคราะห์ครอบคลุมและละเอียดมากยิ่งขึ้น
2. ผสานการวิเคราะห์ข้อมูลร่วมกับระบบ GIS (geographic information system) เพื่อสร้างแผนที่เส้นทางการจัดเก็บที่มีประสิทธิภาพ
3. ศึกษาลักษณะเชิงคุณภาพ เช่น การสัมภาษณ์ลูกค้า เพื่อเข้าใจความต้องการหรือข้อจำกัดที่ไม่สามารถวิเคราะห์ได้จากข้อมูลเชิงปริมาณ
4. ศึกษาผลกระทบในระยะยาว เช่น การลดต้นทุน การเพิ่มความพึงพอใจของลูกค้า และการลดผลกระทบต่อสิ่งแวดล้อม เพื่อประเมินประสิทธิภาพของกลยุทธ์การจัดการ
5. ขยายขอบเขตการวิจัยไปยังพื้นที่หรือภูมิภาคอื่น เพื่อเปรียบเทียบผลลัพธ์และสร้างแนวทางที่สามารถนำไปปรับใช้ในบริบทที่หลากหลายมากขึ้น

6. เพื่อเพิ่มความรอบด้านของการวิเคราะห์ ควรมีการเปรียบเทียบผลการจัดกลุ่มด้วยอัลกอริทึมอื่นในกลุ่มเดียวกัน เช่น K-Modes ซึ่งเหมาะกับข้อมูลเชิงกลุ่ม และ K-Medoids ซึ่งทนต่อค่าผิดปกติได้ดีกว่า K-Means การทดลองเปรียบเทียบจะช่วยให้ยืนยันความเหมาะสมของอัลกอริทึมที่เลือกใช้ และเพิ่มความน่าเชื่อถือให้ผลการวิจัย

รายการอ้างอิง

- Bemthuis, R. H., Govers, R. R., & Asadi, A. (2024). *A CRISP-DM-based methodology for assessing agent-based simulation models using process mining*. arXiv. <https://doi.org/10.48550/arXiv.2404.01114>
- Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36(4), 1165-1188.
- Department of Health. (2024a). *Infectious waste management during COVID-19 pandemic*. Department of Health.
- Department of Health. (2024b). *Overview of infectious waste management situation in Thailand*. Department of Health.
- Fisher, R. A. (1927). Statistical methods for research workers. *Protoplasma*, 23, 282.
- Han, J., Kamber, M., & Pei, J. (2011). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.
- Pollution Control Department. (2022). *Guidelines for infectious waste management*. Pollution Control Department.