Optimizing Marketing Strategy for PT PLN Through Customer-Based Clustering Using Machine Learning

Anton Sujarwo[1,2], Fahdzi Muttaqien[1], Yandika Restu W[1,2] & Zuhri Arieffasa S[1,2]

[1] Institut Teknologi Bandung
[2] PT. PLN (Persero)
Email: mail.anton.sujarwo@gmail.com

Abstract. PT PLN (Persero) operates across Indonesia, with branch units managing diverse customer categories and volumes, including industrial, business, and residential customers. This variation in customer profiles presents challenges in optimizing marketing strategies and determining appropriate marketer allocations for each unit. To address this, this study applies machine learning, specifically the K-Means clustering algorithm, to group PLN units based on the number and type of customers managed. The Elbow Method identifies the optimal number of clusters, ensuring that the grouping reflects meaningful distinctions among units. Through clustering, PLN aims to enhance marketing strategies by tailoring resource allocation and optimizing marketer distribution. This study demonstrates that customer-based clustering can provide actionable insights for targeted marketing and resource optimization within PLN's operational framework.

Keywords: *customer segmentation; machine learning; clustering; unsupervised learning; k-means.*

1 Introduction

PT PLN (Persero) is Indonesia's primary electricity provider, with branch units that manage diverse customer categories across the nation's vast geographical landscape. Each unit serves a unique distribution of customer types, which generally fall into three main categories: industrial, business, and residential. The variation in customer profiles and volumes across different regions creates distinct challenges for PLN, particularly in formulating effective marketing strategies and allocating resources, such as the appropriate number of marketers, to each unit. With the evolving demand for tailored services and regional operational optimization, the need for a data-driven approach to customer segmentation and resource allocation becomes increasingly critical.

To address these challenges, this study employs machine learning, specifically the K-Means clustering algorithm, to group PLN units based on their customer data. Machine learning is a computational method that uses data to "learn" patterns and make predictions. This process involves data collection, model

Received _____, Revised _____, Accepted for publication _____

training, validation, and the application of the model to new data to produce the desired output. The applications of machine learning span various fields such as facial recognition, natural language processing, anomaly detection, and stock market prediction [1]. Clustering is a technique that segments data points into groups, or clusters, based on their similarities, providing an organized structure that enables better strategic planning and targeted resource deployment. By clustering PLN units according to the number and type of customers managed, the company can gain insights into the distinct characteristics of each unit's customer base, which, in turn, can guide the design of localized marketing strategies and the allocation of marketers. K-means is an effective method for uncovering hidden structures in data without explicit supervision [1], while the speed and simplicity of this algorithm are among its main advantages [2]. This is the rationale behind choosing the K-Means method for use in this study.

Determining the optimal number of clusters is essential to ensure that the segmentation captures meaningful distinctions between customer groups. For this purpose, the Elbow Method is applied, which helps identify the point at which adding more clusters does not significantly reduce within-cluster variance, indicating an optimal balance between cluster detail and interpretability. By identifying clusters based on customer types and volumes, PLN can improve marketing efficiency by aligning resources and strategies with the needs and potentials of specific customer segments.

Through this approach, PLN seeks to achieve a more efficient allocation of marketing resources and improve the effectiveness of its strategies across Indonesia's diverse regions. This study demonstrates how machine learning techniques can support public utility companies like PLN in addressing operational challenges, providing a framework for data-driven decision-making that aligns with the organization's broader goals of resource optimization and customer satisfaction.

2 Machine Learning

Machine learning is a branch of artificial intelligence (AI) that focuses on developing algorithms and techniques that allow computers to learn from data and make predictions or decisions without being explicitly programmed to perform specific tasks. Machine learning is a computational method that uses data to "learn" patterns and make predictions. This process involves data collection, model training, validation, and the application of the model to new data to produce the desired output. The applications of machine learning span various fields such as facial recognition, natural language processing, anomaly detection, and stock market prediction [1]. Machine learning can be classified into several main categories based on the type of task to be solved and the type of data

available. The main categories in machine learning include supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning [2].

2.1 Clustering

Clustering is a technique used in machine learning and pattern recognition to group a set of data points into clusters, such that points within the same cluster are more similar to each other than to points in other clusters. Clustering is an example of unsupervised learning, where the algorithm finds structure in data without predefined labels or categories.
The goal of clustering is to uncover the natural grouping of data points based on a measure of similarity or distance. Clustering can be used for a variety of applications, including data compression, anomaly detection, and preprocessing for supervised learning. There are several methods for clustering, including popular approaches like K-Means, Gaussian Mixture Models (GMM), and Hierarchical Clustering, each of which has unique applications based on the data structure and intended outcome [1].

2.2 K-Means

K-Means is described as a widely-used clustering algorithm that groups data points into k clusters, where k is specified by the user. The algorithm aims to partition the data in such a way that the variance within each cluster is minimized.
Key Steps of K-Means :

1. Initialize Centroids: The algorithm begins by choosing k initial centroids (the centers of each cluster).
2. Assignment Step: Each data point is assigned to the nearest centroid, forming k clusters based on the minimum Euclidean distance.
3. Update Step: The centroids are then recalculated by taking the mean of all data points in each cluster.
4. Iteration: The assignment and update steps repeat until the centroids stabilize (i.e., no longer move), or until a pre-set number of iterations is reached.

Overall, K-Means is an effective algorithm for data grouping, especially when clusters are well-separated, and provides a simple and intuitive approach to unsupervised learning.

Several studies on customer segmentation using K-Means clustering have been conducted, including :

1. Mall Customer Segmentation: This study aimed to optimize marketing strategies by grouping customers based on their annual income and spending scores [3].
2. Customer Segmentation in Shopping Centers in Indonesia: This research identified customer characteristics, product categories, and target segmentation based on demographic and behavioral data [4].
3. Digital Native Market Segmentation Using Various Clustering Techniques: This study clustered digital consumers based on their behavior on social networking platforms, focusing on teenagers who are part of the digital native generation [5].

3 Methodology

The methodology employed in this research consists of four main stages: data collection, data preparation, clustering, interpretation, and evaluation.

Data Collection:
In this stage, data from various PLN units across Indonesia is collected, focusing on customer categories, including industrial, business, and residential customers. This data is primary data from 2023, sourced from PLN's internal division.

Data Preparation:
In this stage, data from various PLN units across Indonesia is collected, focusing on customer categories, including industrial, business, and residential customers. Data preparation involves cleaning the dataset by handling any missing or inconsistent values, standardizing data to ensure consistency, and performing feature scaling as necessary. This process ensures that the data used for clustering accurately represents customer volumes and categories managed by each unit.

Clustering:
After preparing the data, the K-Means clustering algorithm is applied to group PLN units based on their customer characteristics. The Elbow Method is utilized to determine the optimal number of clusters by identifying the point where adding more clusters does not significantly decrease the within-cluster variance. This stage results in clusters of PLN units with similar customer profiles, allowing for meaningful segmentation. The clustering method in this study is implemented using the Python programming language with the KMeans library from sklearn.

Interpretation:
Once clustering is complete, each cluster is analyzed to interpret the underlying characteristics of the grouped units. This involves examining the distribution of customer types (industrial, business, and residential) within each cluster to identify patterns and trends. This interpretation provides insights into the unique

customer management needs of each group, guiding resource allocation and marketing strategies.

Evaluation:
In the final stage, the clustering results are evaluated to assess their practicality and relevance in optimizing PLN's marketing strategies. This involves reviewing the clusters to ensure they align with PLN's operational objectives and provide actionable guidance for resource distribution. The evaluation process confirms that the clustering model provides a useful segmentation framework for supporting strategic decisions.
This methodology provides a structured approach to segmenting PLN units, ensuring data integrity, meaningful grouping, and practical insights to support optimized marketing and resource allocation strategies.

3.1 Data Preparation

Data preprocessing is a crucial step in the data analysis process to ensure that the data used is of high quality, clean, and ready for further analysis. Collected data often contains missing values, outliers, or input errors. The goal of data preprocessing is to ensure that the data used for K-Means clustering is valid and reliable. In this study, preprocessing included outlier detection and handling, using Z-score methods.

Z-score Method:
The Z-score method identifies outliers by measuring how far a data point deviates from the mean in terms of standard deviations. Data points with a Z-score beyond a certain threshold (commonly set at ±3) are flagged as potential outliers. This method is particularly useful when the data follows a normal distribution.

The following is an outline of the outlier detection performed in this study:

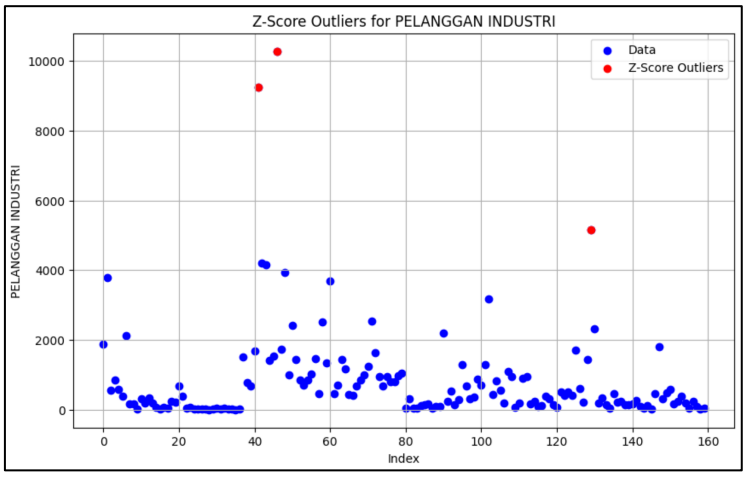

Figure 1 Z-Score

3.2 K-Means Clustering

The clustering of units is carried out using customer count data categorized into business, industrial, and residential customers. To determine the optimal number of clusters, the Elbow Method is applied, yielding the following results:

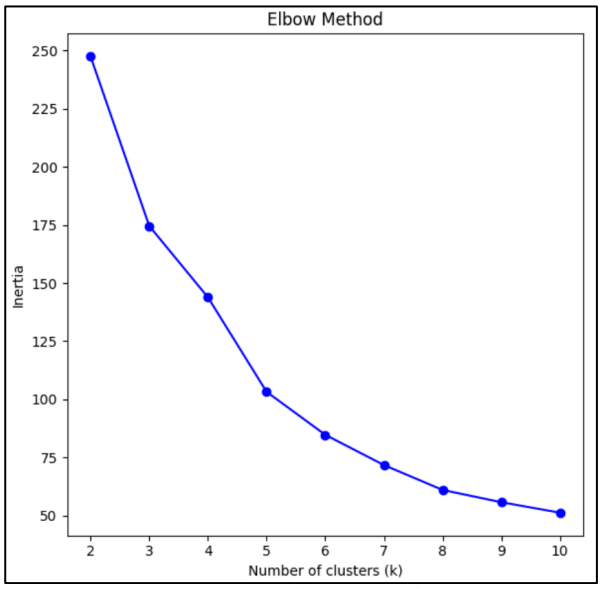

Figure 2 Elbow Method

Based on the above graph using the Elbow Method, the number of clusters is set to $k = 5$, resulting in the following data clustering outcomes :

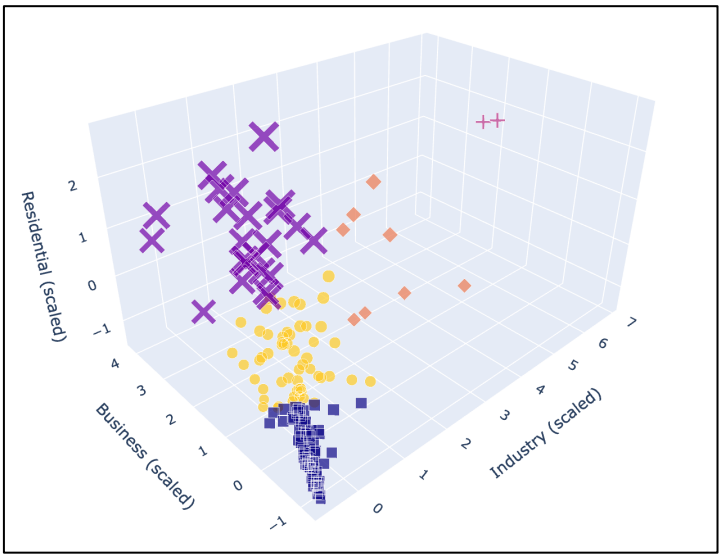

Figure 3 K-Means Clustering

4 Results and Discussion

From the experimental results, the number of unit for each cluster is obtained as in the table.

Table 1 Summary of cluster.

No	Cluster	Sysmbol	Number of Branch
1	Cluster 1	Blue Squre	70
2	Cluster 2	Yellow Circle	56
3	Cluster 3	Purple "X"	24
4	Cluster 4	Orange Square	8
5	Cluster 5	Purple "+"	4

Cluster 1 (Blue Square):
This cluster represents units managing a relatively small number of customers across all categories (industrial, business, and residential). These units likely operate in areas with lower demand or in regions with limited industrial or commercial activity. Marketing strategies in this cluster could focus on gradual customer base expansion and enhancing service quality to attract new clients. Resource allocation for marketers may be minimal, as the customer base is smaller.

Cluster 2 (Yellow Circle):
Units in this cluster manage a medium volume of customers in each category, suggesting they operate in moderately populated or semi-industrialized regions. Marketing strategies could focus on retaining existing customers while gradually expanding services to medium-sized businesses and residential customers. These units might require a balanced approach to marketing resource allocation, targeting both customer retention and moderate growth.

Cluster 3 (Purple "X"):
This cluster includes units that manage a large number of business and residential customers but only a medium number of industrial clients. These units may be located in regions with developed commercial and residential sectors but limited industrial activity. Marketing efforts here could prioritize business and residential customer needs, emphasizing service reliability and targeted campaigns to expand the industrial customer base. Marketing resources should focus more on high-demand residential and business sectors.

Cluster 4 (Orange Square):
Units in this cluster manage a large number of industrial customers and a medium number of business and residential clients. This configuration suggests they are in regions with significant industrial activity, with moderate commercial and residential growth. Marketing strategies in this cluster could focus on supporting industrial clients with tailored services while expanding offerings for business and residential customers. Resource allocation should prioritize industrial marketing efforts but also support gradual expansion in the other categories.

Cluster 5 (Purple "+"):
This cluster encompasses units managing a large volume of customers across all categories, indicating operation in highly populated or industrialized regions. These units likely require substantial resources for customer service and marketing to maintain high satisfaction across customer types. Marketing strategies could emphasize advanced customer engagement, loyalty programs, and personalized services for industrial, business, and residential customers alike. This cluster may require the highest allocation of marketers to effectively support the broad and diverse customer base.

Each cluster's analysis provides insights into strategic marketing focus areas, with recommendations for resource allocation tailored to the customer management characteristics of each group.

5 Conclusions

The clustering analysis of PT PLN (Persero) units based on customer management characteristics reveals five distinct groups, each with unique patterns of customer composition across industrial, business, and residential categories. This segmentation provides strategic insights into how marketing resources and strategies can be optimized for each cluster:

Cluster 1, which manages a smaller customer base across all categories, may benefit from targeted growth strategies and minimal resource allocation, focusing on gradual expansion and customer acquisition in low-demand areas.

Cluster 2, managing a medium number of customers across categories, requires a balanced marketing approach, with resources aimed at retaining existing customers while promoting moderate growth in business and residential sectors.

Cluster 3 focuses on high-demand residential and business sectors with a medium number of industrial customers. Marketing efforts for these units should prioritize business and residential customer needs, with potential expansion into industrial services.

Cluster 4 serves a large number of industrial customers but medium levels of business and residential clients. For these units, the primary focus should be on maintaining strong industrial relationships while gradually expanding commercial and residential offerings.

Cluster 5, representing units with a large customer base in all categories, will need the highest level of marketing support to maintain engagement across diverse sectors. These units should implement advanced marketing strategies to cater to varied customer needs.

Overall, the clustering results enable a more efficient and targeted allocation of marketing resources, allowing PT PLN to tailor its strategies to the specific customer profiles managed by each unit. This data-driven approach provides a foundation for enhancing customer engagement, optimizing marketer distribution, and improving the overall effectiveness of PLN's marketing operations across Indonesia.

However, to achieve better clustering results, future research could compare this approach with other clustering methods such as the GMM or Hierarchical methods.

6 References

[1] Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
[2] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.)*. Springer.
[3] M. G. Pradana and H. T. Ha, "*K-Means Clustering Method For Customer Segmentation Based On Potential Purchases*," J. Appl. Data Sci., vol. 2, no. 1, pp. 19-25, Jan. 2021.
[4] I. A. Mulyo and J. Heikal, "*Customer Clustering Using The K-Means Clustering Algorithm in Shopping Mall in Indonesia*," Management Anal. J., vol. 12, no. 4, pp. 366-371, 2022
[5] M. A. Uddin et al., "*Data-driven Strategies for Digital Native Market Segmentation using Clusteri*ng," Int. J. Cogn. Comput. Eng., vol. 5, pp. 178-191, 2024.

Acknowledgement

This research was supported by the PT PLN (Persero).