Logo, company name

Description automatically generatedLogo

Description automatically generated

Indonesian Journal of Science & Technology

Journal homepage: http://ejournal.upi.edu/index.php/ijost/

*Indonesian Journal of Science & Technology x(*x) (20xx) x-xx

Customers Lifetime Value-Based Segmentation using Hybrid K-means Clustering and Analytic Hierarchy Process: a Case Study of an Indonesian National Electricity Company

*Radit Rahmadhan1\*, Meditya Wasesa2*

1,2School of Business and Management, Institut Teknologi Bandung, Indonesia

Correspondence: E-mail: [radit\_rahmadhan@sbm-itb.ac.id](mailto:radit_rahmadhan@sbm-itb.ac.id), [meditya.wasesa@sbm-itb.ac.id](mailto:meditya.wasesa@sbm-itb.ac.id)

|  |  |  |
| --- | --- | --- |
| **A B S T R A C T S** |  | **A R T I C L E I N F O** |
| To effectively manage the increasing electricity demand, developing predictive analytics based on understanding the customers' electricity consumption patterns is essential. This study presents a hybrid customer segmentation analytics by combining the K-Means clustering, customer lifetime value concept, and analytic hierarchy process. The analytics is useful for decision-making in defining service strategies integrated with customer relationship management. This study uses more than 16 million records of customer electricity consumption data from January 2019 to December 2020. We use K-Means clustering to identify the initial market segments. Next, we evaluate and validate the customer segmentation results using the customer lifetime value concept and analytical hierarchy process. Three customer segments were identified. We propose a continuous replenishment program for the first customer group, less-profitable customers. This type of customer will implement partnership programs to encourage increased electricity consumption and retail account marketing, such as must carry out further customer profiling by providing service product information following customer profiles using CRM in line with the customer ID. While for the second and third customer groups, which are profitable customers, we propose business to business this type of customer will implement increase their energy consumption by offering premium service products without going out during peak usage and customer business development strategy such as by providing special executive accounts to customers to provide the best solutions and consultation on electrical problems.  © 2021 Tim Pengembang Jurnal UPI |  | ***Article History:***  *Received 00 Jun 20xx*  *Revised 00 Jul 20xx*  *Accepted 00 Jul 20xx*  *Available online 00 Sep 20xx*  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  ***Keyword:***  *Analytics,*  *Customer Analytics,*  *K-Means Clustering,*  *Electricity,*  *Customer Lifetime Value,*  *Customer Relationship Management,*  *Analytical Hierarchy Process.* |

**1. INTRODUCTION**

The electricity consumption in Indonesia continues to increase from 2015 to 2020 by 98.89%, with business customers dominating electricity consumption (Katadata, 2020). PT. PLN Persero is the only electricity provider in Indonesia providing higher electricity power for the entire region, including the West Sumatra region. While the electricity demand of business customers is increasing, electricity blackouts often occur up to a high frequency of four times a month. Based on the data analysis results that have been carried out, power outages cause the average electricity usage time for business customers to be under 50 hours per month. Based on information from the Commercial Manager of PLN for the West Sumatra Region, the incident was due to customers using power above 200 thousand using a higher peak load electricity usage time than electricity outside peak hours. During off-peak hours, customers rarely use it. Based on these problems, PT. PLN Persero West Sumatra must understand the characteristics of the customer's electricity use so that the use of electricity at times outside the peak load can be allocated resources that are appropriate and on target-to-target customer segmentation.

Customer segmentation is one way to understand and map customer preferences. According to previous research, customer segmentation refers to grouping customers based on similar characteristics (McLoughlin, Duffy, & Conlon, 2015). Thus, customer segmentation can predict future actions in consuming the services. That customers use and build relationships and enhance customer commitment to building a solid business (McLoughlin et al., 2015) ;(Ye, 2021) several previous studies discussed customer segmentation on customers' electricity consumption (Camero, Luque, Bravo, & Alba, 2018) (Gajowniczek & Zabkowski, 2018); (Lee, Kim, & Jang, 2020), and electricity demand (Jang, Jeong, Kim, & Joo, 2021) . The research context is more about finding new customer behavior patterns in consuming electricity, and more methods use a combination of K-Means and Self Organizing Maps (SOM) and other clustering methods (Camero et al., 2018) (Gajowniczek & Zabkowski, 2018), (Lee et al., 2020). Other studies use the regression method for customer segmentation (Jang et al., 2021). They want to predict future electricity consumption to meet electricity demand from customers. The results of several previous studies provide recommendations for optimization of the use of electricity to the electricity that has been provided consumption (Camero et al., 2018); (Gajowniczek & Zabkowski, 2018); (Lee et al., 2020). There are also other studies analyzing customer characteristics by applying the K-Means Clustering model by analyzing tariffs, power, the number of bills paid and then from the model results. The concept is used in Customer Relationship Management (CRM) to gain insight or make company business decisions (Afthoni et al., n.d.).

Previous research on customer segmentation commonly was based on total electricity consumption per day consumption (Camero et al., 2018);(Gajowniczek & Zabkowski, 2018), ;(Lee et al., 2020),. Another study only analyzed rates, electricity, and total bills by combining K-Means and CRM (Afthoni et al., n.d.). Therefore, this study fills the gap by analyzes based power, peak load electricity consumption, and peak external load electricity consumption by applying a combination of the K-Means clustering method (Bapna, Goes, Gupta, & Jin, 2004), customer lifetime value concept, and analytic hierarchy process. The method can handle large-sized data such as the one we use, i.e., data from PT. PLN Persero West Sumatra Region from 2019 to 2020. Data features are installed power at the customer, peak load electricity usage time, peak load electricity usage time. The analysis results are useful to improve future marketing strategy decisions. This improvement can help the company to optimize electrical power services.

The first part of this article describes the background of the problem, gaps in the research, the purpose, literature review on customer segmentation carried out in previous studies. Section 3 narrates the research method. Section 4 explains the results and discussion. Section 5 presents the conclusions, implications, current limitations, and future research.

**Table 1.** Reviewed Studies on Customer Segmentation in Electricity Consumption

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Article | Business Context | Dataset | Segmentation Features | Segmentation Method |
| (McLoughlin et al., 2015) | Electricity Load Profile in Ireland | Experimental data period January 1, 2009, to December 31, 2010, | Dwelling type, No. of bedrooms, Age, Social Class, Electronic Type | K-means, k-medoid and Self Organizing Maps (SOM) |
| (Toussaint & Moodley, 2020) | Electricity Consumption in South Africa | South Africa Electric Load Profile Data from 1994 to 2014 | X=Hour (load profile multiple one days)  Y= X multiple All household | K-Means  And Self Organizing Maps (SOM) |
| (Camero et al., 2018) | Electricity  Demand Signature in Andalusian | The load data of 64 buildings located in Andalusia, Spain | Identity, Industrial Division, Industrial Categories, Mean Power Consumption, Power Consumption | Variable selection (Feature Selection), Model (K-Means, Hierarchical Clustering, K-Medoid Clustering), Validation (Connectivity, Dunn and  Silhouette indexes) |
| (Jang et al., 2021) | Electricity Load Profile | Smart Metering Data in 2009 | Identity, Social Status, age, gender, Demand kWh, Income | Regression Ordinary Least Square (OLS), Evaluation (Root Mean Square Error (RMSE)) |
| (Lee et al., 2020) | Electricity Load Profile | Residential Demand Data from November 2017 until February 2018 | Identity, Daily Consumption, Load Profile, Peak Hour, Demand | K-means, Fuzzy C-Means (FCM) and Self Organizing Maps (SOM) |
| (Gajowniczek & Zabkowski, 2018) | Electricity Consumption Forecasting | Electricity Consumption Data from 46 homes in Texas | Identity, Time, Total kWh | Model (Artificial neural networks, regression  trees, random forest regression, 𝑘-nearest neighbors' regression,  and support vector regression), Evaluation (Naive forecast, random  forecast, the ARIMA model, and stepwise regression) |
| (Bañales, Dormido, & Duro, 2021) | Electricity Demand with Renewable Technologies | Half-hourly energy use for one-year data | Average energy use,  energy–temperature correlation, the entropy of the load-shape representative vector, and distance to  wind generation patterns. | Model (K-Medoids), Validities (average silhouette) |
| (Afthoni et al., n.d.) | Electricity Consumption in Indonesia | Customer Transaction in September 2021 | Rate, Power, Total kWh, Total Cost, Flash Time | Variable selection with correlation  Model (K-Means)  Validity (Silhouette Method)  Explores (Customer Relationship Management (CRM)) |

Previous studies in customer segmentation in electricity consumption have explored various dimensions of the customer clustering problem (McLoughlin et al., 2015); (Camero et al., 2018);(Gajowniczek & Zabkowski, 2018); (Lee et al., 2020). They use the context of electricity consumption as a case study to find out patterns of electricity use in predicting future electricity consumption. Several clustering models, one of which is often used, namely K-Means Clustering, have been explored customer grouping by considering patterns of electricity use and electricity demand to meet electricity consumption based on what has been prepared by the company (Ye, 2021);(Jang et al., 2021).

A context study of load profile electricity (McLoughlin et al., 2015) using experimental data by installing 4000 intelligent meters in several homes in Ireland with existing methods used to classify household electricity use, in general, can be divided into four categories, statistics, manipulation, time series, and clustering. Statistical methods have been widely used in the unregulated power market to form a standard load of Personal Classes (PC). A typical load PC is used for settlement purposes and estimates the amount and time of electricity used.

Research on electricity consumption in South Africa (Toussaint & Moodley, 2020) focuses on household customers, aiming to classify customers based on patterns and types of using electricity using the K-Means clustering model and Self Organizing Maps (SOM). They used internal and external validation to evaluate the clustering structure based on South African households expected daily electricity consumption behavior. Another study used electrical load data also in Andalusia, Spain (Camero et al., 2018), but the research context was about electricity demand. Using a combination model between K-Means clustering and K-medoid clustering, they determine interrelated variables to predict customer segmentation. This study aims to provide an alternative customer segmentation that can manage several types of customers. It then presents the segmentation results based on the characteristics of the load curve. Finally, they compare the two marks and provide solutions to the effects of classification and segmentation.

Research on the context of electricity load data (E. Lee et al., 2020) uses electricity demand data to predict electricity loads per day based on customers' heterogeneity of electricity demand behavior, then processed using a combination of K-Means clustering models and Self Organizing Maps (SOM) and Fuzzy C-Means. The segmentation results provide the proper group identification for electricity demand per day. The result shows a tremendous impact because it can save on utility costs based on electricity reduction by customers. Another study with the same context as (E. Lee et al., 2020), but this study uses data from smart meters in 2009 (Jang et al., 2021), they use a regression model with an evaluation of the root mean square error for customer segmentation based on electricity demand used, age, and Income from the customer. The aim is to find new customer electricity usage behavior patterns based on predetermined variables. Another study uses six regression models to predict daily electricity consumption based on the total electricity consumption used by customers (Gajowniczek & Zabkowski, 2018). They compared the models to find new patterns of customers' daily electricity usage.

Research on the context of looking for energy reserves based on the number of customer electricity requests (Bañales et al., 2021) uses data on customers' half-day electricity usage by selecting variables based on the average amount processed by adding wind variables as alternative electrical energy. This study uses the K-Medoid model and the Silhouette method to validate the number of clusters to apply an efficient time series clustering methodology that explicitly considers the pattern of renewable energy generation. Other research on the context of electricity consumption in Indonesia (Afthoni et al., n.d.). They used data on customer electricity bills in September 2021 with predictors of power, rate, total kWh, flash sale, total cost, tested for variable correlation. This research uses the K-Means Clustering model and the Silhouette Method as the number of clusters to get customer segmentation based on the characteristics of customers paying for electricity according to the power used. The clustering results will be explored using the CRM model to gain insight into customer action in the future according to the wisdom that has been carried out.

**1.2 Customer Lifetime Value in Customer Segmentation**

Previous studies in customer segmentation have explored various dimensions of customer clustering problems (Gustriansyah, Suhandi, & Antony, 2019; Janardhanan & Muthalagu, 2020; Marisa, Ahmad, Yusof, Fachrudin, & Aziz, 2019). Many of them use the marketing context as a case study. The K-Means clustering model and Customer Lifetime Value explores customer grouping by considering the specified product preferences and predicting customer behavior in buying products offered by the company (Z. J. Lee, Lee, Chang, & Sano, 2021)

A context study in marketing combines the Customer Lifetime Value (CLV) and K-Means models in each customer segment (Marisa et al., 2019). The grouping uses the K-Means Clustering method based on the LRFM (Length, Recency, Frequency, Monetary) model. The cluster formation process uses the Elbow method. The CLV value is generated from the multiplication of the LRFM normalization results, and then the LFRM weight value uses the Analytical Hierarchy Process (AHP). Based on the LRFM matrix, this cluster has a high loyalty value, with the symbol LRFM being a loyal customer (the best segment with a high customer loyalty value). Based on the LRFM symbol, companies can create strategies to retain customers and acquire loyal customers with high profitability.

Another study with a supermarket marketing context with the same objective and predictor variables used historical customer data processed with a combination of LRFM models to determine data selection on potential customer purchases (Z. J. Lee et al., 2021). The K-means clustering model to map customers based on the same characteristics is then classified to distinguish potential customers for repurchase and then validated using the elbow method. This study uses data from all AR-Pulsabiz pulse server operators in Malang, Indonesia, to predict the future of Small and Medium Enterprises. The number of potential customers who will become operators by using a combination of the K-Means Clustering model and the LRFM model to group customers to provide services according to priority.

Research in pharmaceutical marketing (Gustriansyah et al., 2019) also has the same objective (Marisa, Ahmad, Yusof, Fachrudin, & Aziz, 2019b), but they use eight validation methods in determining the correct number of groupings. Another transportation survey uses the K-Means Clustering model and the CLV model to group customers (Li et al., 2018) with the same research objective (Ye, 2021). It also has similar goals and models (Ye, 2021) to marketing research in Telecommunication Companies (Abdi & Abolmakarem, 2019). However, they do not use the CLV model but use the Neural Network to classify priority customers after getting the results from clustering.

**1.3 Marketing Strategy in Customer Relationship Management**

Two popular customer relationship strategies can lead to an increase in profits and customers retention (Hosseini, Maleki, & Gholamian, 2010), namely:

1. Sustainable Marketing

This program is a program to maintain and increase customer loyalty through special long-term services and increase value by studying the characteristics of customers (Cunha, Mendonça, Morais, Fernandes, & Letra, 2020; Foncubierta-Rodríguez, Galiana-Tonda, & del Mar Galiana Rubia, 2020; Malm, Pikkarainen, & Hyrkäs, 2020; Rao, Velidandla, Scott, & Drechsel, 2020). Implementing a sustainable marketing program from this concept will be explained as follows.

1. **Continuous Replenishment Program**

This program is used for less profitable customers (Tsao, Setiawati, Linh Vu, & Sudiarso, 2021). Approaches to programs such as partnership programs encourage increased use of the company's services to customers (Tsao, Setiawati, Linh Vu, & Sudiarso, 2021).

1. **Business to Business**

This program is used for profitable customers (Foncubierta-Rodríguez, Galiana-Tonda, & del Mar Galiana Rubia, 2020; Rao, Velidandla, Scott, & Drechsel, 2020). The approach to this program is like providing special executive services to customers to improve service, so that customer trust will increase and become more loyal (Baniasadi, Samari, Hosseini, & Najafabadi, 2021; Daat, Sanggenafa, & Larasati, 2021; Xie, Chen, Huang, & He, 2021; Yan, Qin, Nie, & Yang, 2018).

2. One to One Marketing

This program is an individual program that satisfies customers' unique needs (Gil-Quintana & Vida de León, 2021; Kafkas, Perdahçı, & Aydın, 2021). This program uses customer information from online news and databases, followed by personal interactions to meet customers' unique needs (Baniasadi, Samari, Hosseini, & Najafabadi, 2021; Xie, Chen, Huang, & He, 2021). Build interactive marketing and post-marketing programs in developing customers using individual customer information (Huynh, Trieu, van Nguyen, Tran, & Lam, 2021; Koponen, Julkunen, Gabrielsson, & Pullins, 2021; Yudhya, 2019). The application of the one-to-one marketing program from this concept will be explained as follows.

1. **Customer Business Development**

This program is used for profitable customers (Borisavljević & Radosavljević, 2021; Daat, Sanggenafa, & Larasati, 2021) by assessing the benefits of marketing, finance, and management business processes (Koponen, Julkunen, Gabrielsson, & Pullins, 2021; Kulej-Dudek, 2021). This program aims to explore the customer's business development by providing the best solutions and consulting regarding customers' services (Borisavljević & Radosavljević, 2021; Kulej-Dudek, 2021)

1. **Retail Account Marketing**

This program is used for less profitable customers (Yan et al., 2018; Yudhya, 2019). The approach to this program sees the customer as a partner to develop business opportunities. This program performs customer profiling further by using CRM, which is more integrated into the application (Dias, de Oliveira, Filho, & Rodrigues, 2021; Sekizaki, Nishizaki, & Hayashida, 2016).

To the best of our knowledge, most previous studies on customer segmentation on electricity consumption focus on predicting electricity consumption and electricity demand per day used by customers because it affects electricity supply or looking for other electricity alternatives. Previous research focused on household customers by identifying daily electricity consumption (Afthoni et al., n.d.; Gajowniczek & Zabkowski, 2018; Toussaint & Moodley, 2020), electricity load profile (Jang et al., 2021; E. Lee et al., 2020) and daily electricity demand (Bañales et al., 2021; Camero et al., 2018; Hyland et al., 2013). Then, only one study combined the concept of clustering with CRM (Afthoni et al., n.d.); the other research only compared the clustering model to find patterns of electricity use. However, in the idea of clustering electricity consumption for customer segmentation, no one has analyzed based power, peak-load electricity consumption and off-peak-load electricity consumption and then combined them with the concept of CLV (Marisa et al., 2019) to determine the correct customer group. In this study, clustering was carried out using the K-Means method, with the number of clusters being validated using the Elbow method. Then, the clustering results will be classified using CLV. Calculation of CLV will involve the value of the clustering variable and the weight value of the clustering variable value. The weight value will be calculated using the Analytical Hierarchy Process. The results from the CLV will be used to determine marketing strategies based on the concept of Customer Relationship Management on the right customer segmentation results to develop the company's services in the future.

Graphical user interface, diagram, application, Word

Description automatically generated

**Figure 1.** The Research Framework

**2. METHODS**

Figure 1 presents the research framework of this study. The framework is adapted from standard methods for building predictive analytical models (Schoenherr & Speier-Pero, 2015). There are five stages: data collection, data preparation, choice variables, clustering model, marketing strategy.

**2.1. Data Collection**

In this study, we used data from PT. PLN Persero. of the West Sumatra zone. Our research uses customer transaction data from January 2019 to December 2020, consisting of 16,504,228 and 107 data variables. Table 2 shows the data that has been taken from 2 years.

**Table 2.** Result of data collection

|  |  |  |  |
| --- | --- | --- | --- |
| Data | Year | Row | Variable |
| Customer Transactions history | 2019 | 7,945,689 | 107 |
| Customer Transactions history | 2020 | 8,558,539 | 107 |

**2.2. Data Preparation**

This section presents the data preparation processes for developing the prediction model, namely:

**A. Data Profiling**

This section presents the focus of the data, which will be selected based on the data analysis to be carried out. The study starts by looking at the areas in West Sumatra that use the highest electricity. Figure 2 presents based on the results of the plot analysis that has been carried out in 4 areas of the service center of PT. PLN Persero, the Padang area, has the highest electricity consumption compared to other sites.

**Figure 2.** Total electricity consumption based on region

The subsequent analysis looks at potential customers who use higher total kWh. Figure 3 presents the results of plot analysis based on total electricity consumption by customer category. Based on the regulations issued by the Indonesian government [49], customers are divided into five categories, namely household, social, government, business, and industrial. Based on the results of the analysis plot that business customers have carried out, the highest use of electricity is around 37%, followed by industrial customers as much as 31% and other customers using electricity consumption below 15%. Therefore, in this study, the focus is on business customers because they use higher electricity consumption than others and can increase company revenues.

**Figure 3.** Total electricity consumption based on customer energy

**B. Data Cleaning**

This section presents a further analysis of the data focuses carried out previously. This analysis is used to clean or remove data rows if there are duplicate data rows or missing data rows. The results of data cleaning will find potential predictor variables based on the number of data variants contained in the variable. Finally, Table 3 shows the analysis results of data focus and data cleaning obtained 13 variables with 508,934 data records used for model development.

**Table 3.** Result of data cleaning

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Data Type | Count | Max | Min | Variable Description |
| ID Customer | Integer | 24,785 | - | - | Identity of the customer |
| Customer Service Unit | String | 12 | - | - | Customer Service Units or service branches provided by the company which are in 4 customer service centers namely Belanti, Painan, Indarung, Pariaman, Lubuk Basung, Lubuk Sikaping, Koto tuo, Baso, Sijunjung, Sungai Rumbai, Kayu Aro, Sawah Lunto, Batusangkar, Lintau, Lima Puluh Kota and others |
| Data Entry Date | Date | 24 | 2020/12 | 2019/01 | Admin enters data per 1 month |
| Rates | Categorical | 3 | - | - | B1 means a business that uses electricity from 450 kWh to 5500 kWh, B2 means a business that uses electricity from 6600 to 200 thousand kWh, B3 means a business that uses 200 thousand kWh of electrical power and above |
| Power | Integer | 43 | 2,425,000 | 450 | Power used by customers such as 450 kwh,900 kwh,1,300 kwh, 2,200 kwh,3,300 kwh, 7,700 kwh,15,400 kwh,132,000 kwh, 200,000 kwh and others |
| Meter Code | Categorical | 5 | - | - | M means analog meter, and E means the digital meter |
| Flash time | Double | 2,7904 | 4775.66 | 0 | Electricity usage time by customer |
| Total KWH | Integer | 1,0427 | 635,370 | 0 | The total peak load kWh usage and peak external load kWh used by customers |
| KWH Off – Load | Integer | 10,417 | 500,640 | 0 | KWH used at peak external load by customers |
| KWH Peak Load | Integer | 1,515 | 146,580 | 0 | KWH used at peak load by customers |
| Discount | Double | 11 | 338,942 | 0 | The company gives discounts based on the provisions of the company, such as using unused kWh by the company or because of a natural disaster |
| Peak Offload Fee | Double | 18,578 | 518,552,899 | 0 | Payments made when using Peak Offload |
| Peak Load Fee | Double | 2,256 | 227,736,949 | 0 | Payments made when using Peak Load |
| Total Cost | Double | 21,621 | 732,079,768 | 0 | The total cost paid by the customer |

**2.3. Choice Variable**

This section presents predictor variables that will later be used in the clustering model. From the 13 variables in Table 3, the variable to be selected is of type Integer or Double because the process in the clustering model, the focus is on predicting customer segmentation on power based on peak load and peak external load used by customers in the future. Still, the ID\_Customer variable is not included in the predictor because this variable is not needed in the clustering model. This research will expect the peak load, which the usage time is from 6.00 am to 4.59 pm and the peak external load, which is from 5.00 pm to 5.59 am [49]. Based on this explanation, the kWh off-loads, and kWh Peak Load variables are used as predicted in the clustering model. Table 4 shows 9 possible variables used in the clustering model.

**Table 4.** Result of variable

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Data Type | Function | Variable |
| Power | Integer | Predictor | Power used by customers such as 450 kwh,900 kwh,1,300 kwh, 2,200 kwh,3,300 kwh, 7,700 kwh,154,00 kwh,132,000 kwh,200,000 kwh and others |
| Flash time | Double | Electricity usage time by the customer |
| KWH Off-Load | Integer | KWH used at peak external load by customers |
| KWH Peak Load | Integer | KWH used at peak load by customers |
| Total kWh | Integer | The total peak load kWh usage and peak external load kWh used by customers |
| Discount | Double | The company gives discounts based on the provisions of the company, such as using unused kWh by the company or because of a natural disaster |
| Peak Offload Fee | Double | Payments made when using Peak Offload |
| Peak Load Fee | Double | Payments made when using Peak Load |
| Total Cost | Double | The total cost paid by the customer |
| Customer segmentation | Double | Predicted | The results of the cluster based on the model |

**2.4. Clustering Model**

This study aims to develop a prediction model with customer segmentation or clustering that can provide accurate predictions of customers who have the potential to use peak load and peak external load electricity consumption. However, this research still examines the clustering model and its ease of implementation. We use the K-Means Clustering model to group customers.

Commonly, K-means is one of the well-known unsupervised learning techniques for cluster analysis (Bapna et al., 2004). Cluster analysis is used to aggregate or divide the data set into several clusters according to the similarity value. The situation in this model is used because this algorithm has simplicity and ease of use, and users can determine the number of clusters themselves. This number of clusters (k) needs to be determined by validation (Hosseini et al., 2010).

Validation in this study uses the elbow method. The Elbow method in previous studies (Janardhanan & Muthalagu, 2020; Marisa et al., 2019) was used to determine the number of data clusters to be processed. This method visualizes the number of k = 2 until the k is determined. The exact number of groups is selected when a drastic change is proportional to the previous value. The value before the difference is the number of clusters. After the number of sets is determined, the processing will continue by starting with randomly generated centroids and iteratively calculating new centroids to gather to the last group. The steps in the k-means model are described as follows (Marisa et al., 2019)

**Step 1**: Determine the number of clusters with the elbow method

**Step 2**: Each data point in the data set will be assigned to the nearest centroid, and then a new centroid is generated.

**Step 3**: To recalculate a new cluster by assigning all data points to the nearest centroid, and then a new group is created.

**Step 4**: The process will be repeated between step 2 and step 3 until the stopping criteria are met.

**2.5. Marketing Strategy**

In this section, we will present a process that aims to gain insight from the results of the clustering model. This insight can be developed to improve CRM using CLV. CLV is one way of defining customer value (Marisa et al., 2019). The model calculates the distance between zero and the central cluster as high refers to most customer loyalty (Hosseini et al., 2010)CLV is usually used in calculating customer profitability. CLV is done after segmenting customers. CLV is calculated based on the CLV rating determined for each segment (Khajvand, Zolfaghar, Ashoori, & Alizadeh, 2011). CLV equation calculation is as follow:

*Where*:

*X = variables values from cluster results*

*N = end of the variable and weight based on the number of clustered variables*

*W = weight of each value of cluster result*

*I = start of the variable*

*J = start of the weight*

The weight value is obtained using calculations from the Analytical Hierarchy Process (AHP) (A. Hosseini & Hosseini, 2020). AHP solves complex multi-criteria problems into a hierarchy (Khajvand et al., 2011). It is helpful for integrated and fuzzy issues based on human brain assessment. The step from AHP is described below (Park, Kim, Lee, & Whang, 2018):

1. Comparing variables based on cluster results.
2. Make a set of pairwise comparison matrices for each lower level with one matrix for each element
3. The results of the matrix are required for assessment in each pairwise comparison
4. Hierarchical synthesis is now used to determine the criterion weights taken from all eigenvectors.
5. After making all pairwise comparisons, consistency is determined using the eigenvalues with the formula:

*Where:*

*CI = Consistency index*

*λ maximum = the eigenvalue of the predetermined variable value*

*n=number of criteria*

1. Steps 3 to 5 are performed for all levels in the hierarchy.

Based on the results of CLV, then we can determine the targeting that aims to

develop customer service improvement strategies based on the concept of customer relationship CRM (Hosseini et al., 2010), which is described in table 5.

**Table 5.** Customer Relations Strategy

|  |  |  |
| --- | --- | --- |
| Customer Type | Sustainable Marketing | One to one Marketing |
| Profitable Customer | Business To Business | Customer Business Development |
| Less Profitable Customer | Continuous Replenishment Program | Retail Account Marketing |

**3. RESULTS AND DISCUSSION**

The first step is to find the correct variables in the clustering model by combining the predetermined variables with the K-Means clustering model. Based on the results from table 6, the selected variables are based on the high data variance value of 97.7% because the high variance represents a high dissimilarity between each cluster with an error value of around 2.3%. Based on these results, the selected variables are Power, Peak Off-Load and Peak Load because they have the highest data variance. The appropriate variables are marked in yellow.

**Table 6.** The Combination of Variables

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| P | FT | TK | POL | PL | POLF | PLF | TC | D | DIM1 | DIM2 | TV |
| v | v | x | v | v | x | x | x | x | 69.2% | 25.1 % | 94.3 % |
| v | v | x | v | v | v | v | v | x | 79.7 % | 14.3 % | 94.0 % |
| v | v | x | v | v | v | v | x | v | 65.7 % | 14.4 % | 80.1 % |
| v | v | x | v | v | v | v | v | v | 69.7 % | 12.6 % | 82.3 % |
| v | v | v | x | x | x | x | x | v | 47.3 % | 25.1 % | 72.4 % |
| v | v | v | x | x | x | x | v | v | 57.1 % | 20.1 % | 77.2 % |
| v | v | v | x | x | x | x | v | x | 71.4 % | 25.1 % | 96.5 % |
| v | x | x | v | v | v | v | x | x | 92.5 % | 5.1 % | 97.5 % |
| v | x | x | v | v | x | x | x | x | 91.9 % | 5.8 % | 97.7 % |
| v | x | x | v | v | x | x | v | x | 93.2% | 4.4 % | 97.6 % |

*Desc: P: Power, FT: Flash Time, TC: Total KWH, POL: Peak Off Load, PL: Peak Load, POLF: Peak Off Load Fee, PLF: Peak Load Fee, TC: Total Cost, D: Discount, DIM1: Dimension1, DIM2:Dimension2, TV: Total Variant*

The second step, after getting the variables that were processed previously then continued with determining the number of clustering using the elbow method to get the best number of clusters (k). Figure 4 shows the number of groups based on the results of the predictor variables previously described using the Elbow method. When skewed, the correct number of clusters is determined by looking at the line graph. From Figure 4, the chart starts to descend at points 3 and 4.

Chart, line chart, scatter chart

Description automatically generatedThe third step, after getting the best grouping from the elbow method between 3 and 4, visualization at points 3 and point 4 uses the K-Means clustering model. Based on the visualization results, the best grouping of the K-Means clustering model in the electricity consumption sector is at point 3. However, the analysis results show that at point 4, there are outliers (groups at the dark purple point) in the distribution. The study of the k-means effect in Figure 5 and Figure 6 can be seen below.

**Figure 4.** The number of clusters of k

Chart

Description automatically generated

**Figure 5.** Cluster result of k=3

Chart

Description automatically generated

**Figure 6.** Cluster result of k=4

Based on the results of clustering using K-Means clustering, Table 7 presents three different customer groups are finding. The first group represents 937,837 total powers used total electricity consumption at peak load of 27,827 kWh and total electricity consumption when peak off-load is 115,194 kWh with customers using installed capacity above 10,600 kWh. The second group describes as many as 4,260 full powers used total electricity consumption at peak load of 35 kWh and total electricity consumption at peak load of 544 kWh, with customers using installed capacity between 450 kWh to 10,600 kWh. The third group describes 2,226,351 full powers used total electricity consumption at peak load of 123,297 kWh and total electricity consumption at peak load time of 390,803 kWh with customers using installed capacity above 200,000 kWh.

**Table 7.** The result of clustering

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Cluster** | **Number of Customer** | **Total Power (kWh)** | **KWH Peak Off Load (kWh)** | **KWH Peak Load (kWh)** | **Installed Power (kWh)** |
| 1 | 282 | 937,837 | 115,194 | 27,827 | 11,000 -200,000 |
| 2 | 508,615 | 4,260 | 544 | 35 | 450- 10600 |
| 3 | 37 | 2,226,351 | 390,803 | 123,297 | >200,000 |

The fourth step is to determine the customer's lifetime value. But previously defined the variables used for CLV; these variables were adopted from the Range, Frequency, and Monetary (RFM) variable model from the grouping results carried out in table 7. This study adopted the RFM variable model (S. M. S. Hosseini et al., 2010b; Khajvand et al., 2011) according to the variables we got from the grouping. Previously, the variables we used were power, kWh Peak Off Load, kWh Peak Load. These variables will be used to calculate CLV. Table 8 shows the weight value of each variable from the AHP calculation.

**Table 8.** Weight of AHP results

|  |  |
| --- | --- |
| **Variable** | **Weight** |
| Power | 0.237 |
| kWh Peak Off-Load | 0.391 |
| kWh Peak Load | 0.712 |

After getting the variables based on the cluster results that have been done and the correct weight value, the next step calculates the CLV value per group. The calculation is taken from the multiplication between the variable and the weight. NP refers to the standard cluster of the amount of power used by the customer as Weighted Power, NKPOL refers to the usual group of the amount of electricity at the time of peak off-load used by the customer is Weighted kWh Peak Off-Load, NKPL refers to the standard cluster of the amount of electricity at load time The height used by the customer is the Weighted kWh Peak Load. Table 9 presents the average CLV estimated for each.

Finally, after finding CLV in each customer segmentation, we can rank it based on that value. The ranking is based on the highest CLV value, so segment 3 gets the first rank because the value equals 768,236.6, segment 1 receives the second rank

because the value is equal to 287,121, and segment 2 gets the third rank because the value is equal 338.6. Table 10 presents device assignments in customer segmentation.

**Table 9.** Result of Customer Lifetime Value in each cluster

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Centroid** | **Number of Customer** | **NP** | **NKPOL** | **NKPL** | **CLV Value** |
| Segment 1 | 282 | 222,267.4 | 45,040.85 | 19,812.82 | 287,121 |
| Segment 2 | 508,615 | 100.962 | 212.704 | 24.9 | 338.586 |
| Segment 3 | 37 | 527,645.2 | 152,804 | 877,787.46 | 768,236.6 |

**Table 10.** Result of Customer Ranking

|  |  |  |  |
| --- | --- | --- | --- |
| **Segment** | **Number of Customers** | **CLV Value** | **Ranking** |
| 1 | 282 | 287,121 | 2 |
| 2 | 508,615 | 338.586 | 3 |
| 3 | 37 | 768,236.6 | 1 |

The last step is insight from customer segmentation development in each cluster which assesses the purpose of developing a customer service improvement strategy proposed with this model more efficiently. Therefore, targeting will be carried out from the ranking results, which are used to determine the target market based on profitable or less-profitable customers, as shown in Table 11.

**Table 11.** Insight from CRM decision development

|  |  |  |  |
| --- | --- | --- | --- |
| **Segment** | **Number of Customers** | **Ranking** | **Strategy Targeting** |
| 1 | 282 | 2 | Profitable Customer |
| 2 | 508,615 | 3 | Less-Profitable Customer |
| 3 | 37 | 1 | Profitable Customer |

Based on Table 11, there are two targeting strategies. The third group of 37 customers and the second group of 282 customers are profitable customers with installed power between 11,000 kWh to 200,000 are profitable customers with an installed capacity of more than 200,000 kWh and above; therefore, the right strategy is for the long term, namely business to business, This type of customer will increase their energy consumption by offering premium service products without going out during peak usage, while for one-to-one marketing, namely customer business development, by providing special executive accounts to customers to provide the best solutions and consultation on electrical problems.

The first group of 508,615 customers are less-profitable customers with installed power between 450 kWh to 10,600 kWh; therefore, the right strategy is to do that for the long term, namely the Continuous Replenishment Program; this type of customer will implement partnership programs to encourage increased electricity consumption such as providing bonuses in the form of vouchers for purchasing electrical equipment, Umrah tickets, car or motorcycle giveaways then partnering with electronic equipment manufacturers to substitute non-electrical equipment into electricity-based ones such as electric stoves, electric sewing machines, electric vehicles, etc.). One to one marketing strategy, namely the concept of Retail Account Marketing, is PT. PLN Persero must carry out further customer profiling by providing service product information following customer profiles using CRM integrated into PLN Mobile according to customer ID.

**4. CONCLUSION**

Developing predictive analytics based on understanding the customers' electricity consumption patterns is essential to manage the increasing electricity demand effectively. This study presents a hybrid customer segmentation model by combining the K-Means clustering, customer lifetime value concept, and analytic hierarchy process. This study uses more than 16 million records of customer electricity consumption data from January 2019 to December 2020. We use K-Means clustering to identify the initial market segments. Next, we evaluate and validate the customer segmentation results using the customer lifetime value concept and analytical hierarchy process.

Based on the analysis, it was found that there are three different customer segments from the combination of K-Means and CLV models based on power, peak load and peak off-load. Segment 1 has 282 business customers with a total capacity of 938,837 kWh, peak load usage of 27,827 kWh, and peak off-loads of 115,194. In part two, there are 508,615 business customers with a total power of 4,260 kWh, then peak load as much as 35 kWh and peak off-load as much as 544. In segment 3, there are 37 business customers with a total power of 2,226,351 kWh, then peak load of as much as 123.297 kWh and peak off-load 390,803. The strategy that will be taken based on this three-customer segmentation will be integrated with CRM. The second and third segmentation strategy is a profitable customer, so the right strategy is business to business for the long term. In contrast, the short-term strategy used is customer business development. Meanwhile, for segmentation, one strategy used for a long time is the Continuous Replenishment Program, and for the short term, Retail Account Marketing is used.

In terms of the contribution to the literature, this study presents a predictive model using segmentation or customer grouping based on electricity consumption used by business customers in electricity companies. This model can reflect customer behavior towards consuming the consumed electricity load. In most cases, individual customer characteristics show a positive or negative relationship, with each class exhibiting different patterns of electrical load consumption.

In terms of managerial implications, this finding can inform companies to provide more optimal power based on the characteristics of their customers. In addition, this research help companies improve their targeting strategy for their customer and the corresponding revenue. However, this study only focuses on business customers and only uses a combination of k-means clustering with the concept of CRM, namely CLV. Future studies can explore other clustering methods and CRM ideas in the further business context.

**5. ACKNOWLEDGMENT**

**6. AUTHORS’ NOTE**

The authors declare that there is no conflict of interest regarding the publication of this article. Authors confirmed that the paper was free of plagiarism.

1. **REFERENCES**

Afthoni, R., Hamdhani, M., Fitri Karimah, A., Patria, H., Analitika Bisnis, J., & Magister Manajemen Teknologi, F. (n.d.). *Seminar Nasional Teknik dan Manajemen Industri dan Call for Paper* (Vol. 1).

Bañales, S., Dormido, R., & Duro, N. (2021). Smart meters time series clustering for demand response applications in the context of high penetration of renewable energy resources. *Energies*, *14*. https://doi.org/10.3390/en14123458

Baniasadi, N., Samari, D., Hosseini, S. J. F., & Najafabadi, M. O. (2021). Strategic study of total innovation management and its relationship with marketing capabilities in palm conversion and complementary industries. *Journal of Innovation and Entrepreneurship*, *10*.

Borisavljević, K., & Radosavljević, G. (2021). Application of logistics model in analysing relationship marketing in travel agencies. *Zbornik Radova Ekonomskog Fakultet Au Rijeci*, *39*, 87–112.

Camero, A., Luque, G., Bravo, Y., & Alba, E. (2018). Customer segmentation based on the electricity demand signature: The andalusian case. *Energies*, *11*.

Cunha, C. R., Mendonça, V., Morais, E. P., Fernandes, J., & Letra, I. (2020). Using mobile devices and apps to assist the elder population in rural areas and generate business opportunities. *IBIMA Business Review*, *2020*.

Daat, S. C., Sanggenafa, M. A., & Larasati, R. (2021). The role of intellectual capital on financial performance of smes. *Universal Journal of Accounting and Finance*, *9*, 1312–1321.

Foncubierta-Rodríguez, M. J., Galiana-Tonda, F., & del Mar Galiana Rubia, M. (2020). Chambers of Commerce: A new Management. The balanced scorecard approach for spanish chambers. *CIRIEC-Espana Revista de Economia Publica, Social y Cooperativa*, 273–308.

Gajowniczek, K., & Zabkowski, T. (2018). Simulation Study on Clustering Approaches for Short-Term Electricity Forecasting. *Complexity*, *2018*.

Gustriansyah, R., Suhandi, N., & Antony, F. (2019). Clustering optimization in RFM analysis based on k-means. *Indonesian Journal of Electrical Engineering and Computer Science*, *18*, 470–477.

Hosseini, S. M. S., Maleki, A., & Gholamian, M. R. (2010). Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. *Expert Systems with Applications*, *37*, 5259–5264.

Huynh, H. T. N., Trieu, H. D. X., van Nguyen, P., Tran, T. G., & Lam, L. N. H. (2021). Explicating brand equity in the information technology sector in Vietnam. *Administrative Sciences*, *11*. https://doi.org/10.3390/admsci11040128

Janardhanan, S., & Muthalagu, R. (2020). Market segmentation for profit maximization using machine learning algorithms. *Journal of Physics: Conference Series*, *1706*. IOP Publishing Ltd.

Jang, M., Jeong, H. C., Kim, T., & Joo, S. K. (2021). Load profile-based residential customer segmentation for analyzing customer preferred time-of-use (Tou) tariffs. *Energies*, *14*.

Koponen, J., Julkunen, S., Gabrielsson, M., & Pullins, E. B. (2021). An intercultural, interpersonal relationship development framework. *International Marketing Review*, *38*, 1189–1216.

Kulej-Dudek, E. (2021). Ecolabnet service packages as a response to the needs of manufacturing enterprises in the SME sector of the Baltic Sea Region. *Production Engineering Archives*, *27*, 265–271.

Lee, E., Kim, J., & Jang, D. (2020). Load profile segmentation for effective residential demand response program: Method and evidence from Korean pilot study. *Energies*, *16*.

Malm, H., Pikkarainen, M., & Hyrkäs, E. (2020). Impacts of open innovation on company business models: A case study of demand-driven co-creation. *Journal of Innovation Management*, *8*, 75–108.

Marisa, F., Ahmad, S. S. S., Yusof, Z. I. M., Fachrudin, & Aziz, T. M. A. (2019). Segmentation model of customer lifetime value in Small and Medium Enterprise (SMEs) using K-Means Clustering and LRFM model. *International Journal of Integrated Engineering*, *11*, 169–180.

McLoughlin, F., Duffy, A., & Conlon, M. (2015). A clustering approach to domestic electricity load profile characterisation using smart metering data. *Applied Energy*, *141*, 190–199.

Rao, K. C., Velidandla, S., Scott, C. L., & Drechsel, P. (2020). Business Models for Fecal Sludge Management in India. *Resource Recovery & Reuse Series*, *18*.

Tsao, Y. C., Setiawati, M., Linh Vu, T., & Sudiarso, A. (2021). Designing a supply chain network under a dynamic discounting-based credit payment program. *RAIRO - Operations Research*, *55*, 2545–2565.

Xie, W., Chen, B., Huang, F., & He, J. (2021). Coordination Of A Supply Chain With A Loss-Averse Retailer Under Supply Uncertainty And Marketing Effort. *Journal of Industrial and Management Optimization*, *17*, 3393–3415.

Yan, Q., Qin, C., Nie, M., & Yang, L. (2018). Forecasting the Electricity Demand and Market Shares in Retail Electricity Market Based on System Dynamics and Markov Chain. *Mathematical Problems in Engineering*, *2018*.

Yudhya, T. B. (2019). Retail store image: A study of the matahari department store (at Bandung Indonesia). *Humanities and Social Sciences Reviews*, *7*, 98–102.