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 National Electricity Company Case

*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** |
| This study presents a hybrid approach for customer segmentation by combining the K-Means clustering, customer lifetime value concept, and analytic hierarchy process to better understand the customers' electricity consumption behavior. This study uses more than 16 million records of customers' 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. For the less profitable costumers, we propose a continuous partnership program to encourage increased electricity consumption during non peak period and retail account marketing For the third group of customers are profitable customers and the second customers are middle profitable customers, we propose premium business to business approach which can accommodate their increasing energy consumption without excessive electricity use in the peak period. The approach will be supported by special executive accounts to these customers.  © 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,*  *Electricity,*  *Customer Lifetime Value,*  *Customer Relationship Management,*  *K-Means Clustering,*  *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). While the electricity demand of business customers is increasing, electricity blackouts often occur up to a high frequency of four times a month. Power outages cause the average electricity usage time for business customers to be under 50 hours per month. The incidents happened due to customers using power above 200 thousand using a higher peak load electricity usage time than electricity outside peak hours. During non peak load hours, customers rarely use it. Based on these problems, the company must understand the customer's electricity use characteristics to improve electricity usage allocation.

Customer segmentation refers to grouping customers based on similar characteristics (McLoughlin, Duffy, & Conlon, 2015). Customer segmentation facilitates a better understanding of customer preferences and future actions. Previous studies on electricity customers segmentation commonly consider daily electricity consumption (Camero, Luque, Bravo, & Alba, 2018; Gajowniczek & Zabkowski, 2018; E. Lee, Kim, & Jang, 2020).This study corresponds to the needs by incorporating power, peak load electricity consumption, and non peak load electricity consumption variables in the segmentation analysis. Moreover, we combine the classical K-means clustering technique with the customer lifetime value concept and analytic hierarchy process.

**1.1 Previous Studies on Customer Electricity Consumption Segmentation**

Table 1 presents previous studies on customer segmentation using electricity consumption data. Previous studies have explored various variables in the segmentation analyses (Camero et al., 2018; Z. J. Lee, Lee, Chang, & Sano, 2021; McLoughlin et al., 2015). K-Means Clustering has been a popular technique (Jang, Jeong, Kim, & Joo, 2021; Ye, 2021).

**Table 1.** Previous Studies on Customer Segmentation Based on 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)) |
| (E. 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) |

One study profiled electricity load (McLoughlin et al., 2015) using experimental data by installing 4,000 intelligent meters in several homes in Ireland. Another research 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). Another study used electrical load data in Andalusia, Spain (Camero et al., 2018), but the research context was electricity demand not consumption. The study aims to provide an alternative customer segmentation to manage several different types of customers based on the characteristics of the load curve.

One study (E. Lee et al., 2020) uses electricity demand data to predict daily electricity load using a combination of K-Means clustering models and Self Organizing Maps (SOM) and Fuzzy C-Means. The result shows a tremendous impact on utility costs reduction. Another study utilized smart meter data in 2009 (Jang et al., 2021), they using a regression model considering electricity demand used, age, and income variables. Another study compared six regression models to predict daily electricity consumption based (Gajowniczek & Zabkowski, 2018). They compared the models to find new patterns of customers' daily electricity usage.

Another study estimates energy reserves using customer electricity requests (Bañales et al., 2021). The study applied the K-medoid model and silhouette method on customers' half-day electricity usage considering wind energy as alternative energy source.

**1.2 Previous Studies on Customer Lifetime Value Based Segmentation**

Previous studies have incorporated customer lifetime consideration in defining customer segments (Gustriansyah, Suhandi, & Antony, 2019; Janardhanan & Muthalagu, 2020; Marisa, Ahmad, Yusof, Fachrudin, & Aziz, 2019) One latest study used K-Means clustering model and customer lifetime value considering the costumers' product preferences to predicting customer's behavior in buying products (Z. J. Lee, Lee, Chang, & Sano, 2021)

A study combines the Customer Lifetime Value (CLV) concept and K-Means models for the segmentation purpose (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 exploring supermarket marketing used 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, 2019). However, they use eight validation methods to determine the correct 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 relationship strategies can lead to an increase in profits and customers retention (Hosseini, Maleki, & Gholamian, 2010), namely (1) sustainable marketing and (2) one-to-one marketing. Sustainable marketing aims to maintain and increase customer loyalty through providing long-term services (Cunha, Mendonça, Morais, Fernandes, & Letra, 2020; Foncubierta-Rodríguez, Galiana-Tonda, & del Mar Galiana Rubia, 2020; Malm, Pikkarainen, & Hyrkäs, 2020; Rao, Velidandla, Scott, & Drechsel, 2020).

Implementation of sustainable marketing can be in the form of a continuous replenishment program. This can be applied to 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). Another implementation is business to business program applicable to profitable customers and middle profitable customers (Foncubierta-Rodríguez, Galiana-Tonda, & del Mar Galiana Rubia, 2020; Rao, Velidandla, Scott, & Drechsel, 2020). The approach provides special executives to customers to improve service. This way, customer trust and loyalty will increase (Baniasadi, Samari, Hosseini, & Najafabadi, 2021; Daat, Sanggenafa, & Larasati, 2021; Xie, Chen, Huang, & He, 2021; Yan, Qin, Nie, & Yang, 2018).

One to one marketing focuses on individual approach aligned with customers' unique needs (Gil-Quintana & Vida de León, 2021; Kafkas, Perdahçı, & Aydın, 2021). This program uses online news and databases of customer information, 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).

One on one marketing program can be implemented in the form of customer business development and retail account marketing. Costumer business development aims for profitable customers and middle 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). Retail account marketing normally used for less profitable customers (Yan et al., 2018; Yudhya, 2019). The approach sees the customer as a partner to develop business opportunities. This program performs customer profiling further by using CRM system (Dias, de Oliveira, Filho, & Rodrigues, 2021; Sekizaki, Nishizaki, & Hayashida, 2016).

After revieweing the literature, we offer a new approach in costumer electricity segmentation by incorporating power, peak load electricity consumption, and peak external load electricity consumption variables in the segmentation analysis. Moreover, we combine the classical K-means clustering technique with the customer lifetime value concept and analytic hierarchy process.

Graphical user interface, diagram, application, Word

Description automatically generated

**Figure 1.** The Research Framework

**2. METHOD**

Figure 1 presents the research framework of this study that adopts standard methods for building predictive analytics (Shmueli & Koppius 2011). The framework consists of five stages: data collection, data preparation, choice of variables, clustering model, marketing strategy definition.

**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 (see Table 2).

**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**

The data selection is started 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 of power consumption 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 , customers are divided into five categories, namely household, social, government, business, and industrial. As shown in Figure 3 business customers consume the highest use of electricity at around 37%, followed by industrial customers as much as 31% and other customers using electricity consumption below 15%. Therefore, this study focuses 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 analysis is used to handles duplicate data rows or missing data rows. The results of data cleaning will find potential predictors in the dataset. 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 |
| Non Peak Load | Integer | 10,417 | 500,640 | 0 | KWH used at peak external load by customers |
| 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 |
| Non Peak Load Fee | Double | 18,578 | 518,552,899 | 0 | Payments made when using non peak load |
| 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 of Variable**

From the 13 variables of the dataset in Table 3, we select the variable with Integer or Double types (Table 4). Table 4 shows nine possible variables used in the clustering model. Because the study focuses on identifying customer segmentation on power based on peak load and the non peak load. This research will treat electricity consumption from 6.00 am to 4.59 pm as peak load, and the rest as non peak load electricity consumption. Thus the peak load and non-peak load electricity consumption are the main features for 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 |
| Non Peak Load | Integer | KWH used at peak external load by customers |
| 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 |
| Non Peak Load Fee | Double | Payments made when using non peak load |
| 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**

Commonly, K-means is one of the well-known unsupervised learning techniques for cluster analysis (Bapna et al., 2004) 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.

We determint the number of clusters (k) by elbow method of validation(Hosseini et al., 2010; Janardhanan & Muthalagu, 2020; Marisa et al., 2019) 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 Definition**

CLV is one way of defining customer value (Marisa et al., 2019). The model calculates the distance between the data point and the central cluster (Hosseini et al., 2010). The higher the value, the more loyal the customer is. CLV is calculated based on the CLV rating determined for each segment (Khajvand, Zolfaghar, Ashoori, & Alizadeh, 2011) 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 CLV results, then we can determine the appropriate service improvement strategies based on the concept of customer relationship management (Hosseini et al., 2010) (see Table 5).

**Table 5.** Customer Relations Strategy

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

**3. RESULTS AND DISCUSSION**

In the first step of analysis we aim to define the combination of features which will cover the largest data variance. As shown in Table 6, the combination of power (P), non peak load (NPL), and peak load (PL) features lead to the high data variance value of 97.7%. In complement, the dissimilarity between each cluster has an error value of around 2.3%.

**Table 6.** The Combination of Clustering Variables

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| P | FT | TK | NPL | PL | NPLF | 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 % |

Chart, line chart, scatter chart

Description automatically generated*Desc: P: Power, FT: Flash Time, TC: Total KWH, NPL: Non Peak Load, PL: Peak Load, NPLF: Non Peak Load Fee, PLF: Peak Load Fee, TC: Total Cost, D: Discount, DIM1: Dimension1, DIM2:Dimension2, TV: Total Variant*

**Figure 4.** The Number of Clusters Determination

Next, we determine the number of cluster (k) using the elbow method. Figure 4 shows the visualization of the results. As shown the magnitude of the total within-clusters sum of squared descent radically when we alter the noumber of clusters (k) from 3 to 4. Thus, based on the method the best grouping of the K-Means clustering model in the electricity consumption sector is at cluster number (k) of 3.

As shown in Figure 5, the cluster number of 3 separates the customers into clear different clusters. Using k=4 (Figure 6), we see that there are outliers (group with dark purple points) that have indistinctive boundary with the other cluster (group with light green points). Thus, we deem that three clusters is indeed the right grouping.

Chart

Description automatically generated

**Figure 5.** Cluster Visualization (k=3)

Chart

Description automatically generated

**Figure 6.** Cluster Visualization (k=4)

Table 7 depicts the details of the three cluster. The first cluster represents 282 cosustomers. The centroid of the first cluster is located at the total powers used of 937,837 kWh, total electricity consumption at peak load of 27,827 kWh, total non peak load electricity consumption of 115,194 kWh, and customers using installed capacity above 10,600 kWh. The second cluster represents 508,615 customers. The centroid of the second cluster is located at the total powers used of 4,260 kWh, total electricity consumption at peak load of 35 kWh, total non peak load electricity consumption of 544 kWh, and customers using installed capacity below 10,600 kWh. The third cluster represents 17 customers. The centroid of the third cluster is located at the total powers used of 2,226,351 kWh, total electricity consumption at peak load of 123,297 kWh, total non peak load electricity consumption of 390,803 kWh, and customers using installed capacity above 200,000 kWh

**Table 7.** The Details of the Clustering Results.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Cluster** | **Number of Customer** | **Total Power (kWh)** | **Non Peak Load (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. According to the variables we got from the grouping, this study adopted the RFM variable model (S. M. S. Hosseini et al., 2010b; Khajvand et al., 2011). Previously, the variables we used were power, Non Peak Load, Peak Load. These variables will be used to calculate CLV. Table 8 shows the weight value of each variable from the AHP calculation.

Next, we calculate 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, NNPL refers to the usual group of the amount of electricity at the time of non peak load used by the customer is weighted non peak load, NPL refers to the standard cluster of the amount of electricity at load time The height used by the customer is the weighted peak load. Table 9 presents the average CLV estimated for each.

**Table 8.** Weight of AHP results

|  |  |
| --- | --- |
| **Variable** | **Weight** |
| Power | 0.237 |
| Non PeakLoad | 0.391 |
| Peak Load | 0.712 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Centroid** | **Number of Customer** | **NP** | **NNPL** | **NPL** | **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 9.** Result of customer lifetime value in each cluster

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 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 | Middle Profitable Customer |
| 2 | 508,615 | 3 | Less Profitable Customer |
| 3 | 37 | 1 | Profitable Customer |

Based on Table 11, there are two targeting strategies, namely (1) sustainable marketing and (2) one-on-one marketing. The third group is profitable customers, the right strategy in this group is for sustainable marketing, namely business to business, by offering premium service products to use more electricity during non-peak load, while one-to-one marketing is customer business development, by provide special executive accounts to customers to provide the best solutions and consultations for electrical problems. The second group is the middle profitable customer. Therefore, the right strategy for this group is for sustainable marketing, namely business to business, by increasing energy consumption by offering premium product services without leaving during peak load periods, while one-to-one marketing is by providing campaigns to customers to increase electricity usage at peak load.

The first group is customers who are less profitable because the total power used in one month is 4,260 kWh. Therefore, the right strategy is to do this for the long term, namely the Continuous Replenishment Program; This type of customer will carry out a partnership program to encourage an increase in electricity consumption, such as giving 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 non peak load. Segment 1 has 282 business customers with a total capacity of 938,837 kWh, peak load usage of 27,827 kWh, and non peak 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 non peak 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 non peak load as 390,803. The strategy that will be taken based on this three-customer segmentation will be integrated with CRM. The third segmentation strategy is profitable customers and the second segmentation is middle profitable customers, so the right strategy is business to business for the long term. In contrast, the short-term strategy used is customer business development. As for segmentation, one of the strategies used for the long term 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**

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*. https://doi.org/10.1186/s13731-021-00179-z

Camero, A., Luque, G., Bravo, Y., & Alba, E. (2018). Customer segmentation based on the electricity demand signature: The andalusian case. *Energies*, *11*. https://doi.org/10.3390/en11071788

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*. https://doi.org/10.5171/2020.585068

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*. https://doi.org/10.1155/2018/3683969

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*. https://doi.org/10.3390/en14196130

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

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*. https://doi.org/10.3390/en13061348

Lee, Z. J., Lee, C. Y., Chang, L. Y., & Sano, N. (2021). Clustering and classification based on distributed automatic feature engineering for customer segmentation. *Symmetry*, *13*. https://doi.org/10.3390/sym13091557

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*. https://doi.org/10.1155/2018/4671850

Ye, J. (2021). Analysis on E-commerce Order Cancellations Using Market Segmentation Approach. *ACM International Conference Proceeding Series*, 33–40. Association for Computing Machinery.

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.