**Clustering model of electricity load profile using K-means clustering: A case study of electricity company in Indonesia**

Radit Rahmadhan1 and Meditya Wasesa2

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

**Abstract.**

**Background**: The increase in the burden of electricity consumption in Indonesia, especially the West Sumatra region, is significantly focused on electricity supply. Developing a predictive model related to the pattern of electricity consumption used by customers is very important in managing the electrical power provided by the company.

**Objective**: This study aims to develop a customer segmentation model that is used to analyze customer characteristics based on the use of profile loads. The model used as the basis for decision making in making service strategies that are integrated with Customer Relationship Management.

**Method**: The data used are customer transaction data of PT. PLN Persero, from January 2019 to December 2020. We used an unsupervised machine learning model, namely K-Means Clustering, with validation of the number of clusters using the Elbow method. We assess the customer grouping using customer lifetime value by determining the weight value using the Analytical Hierarchy Process.

**Results**: We obtain three customer groups based on electricity consumption profile. For the first customer group, which categorized into less-profitable customers, we propose continuous replenishment program and retail account marketing. While for the second and third customer groups, which are profitable customers, we propose business to business and customer business development strategy. Clustering using the K-Means Clustering model with customer lifetime value can help companies find the right marketing strategy according to the type of customer.

**Keyword**: *Clustering, K-Means Clustering, Electricity, Elbow Method, Customer Lifetime Value, Customer, Relationship Management, Load Profile, Unsupervised machine learning, Analytical Hierarchy Process*

# Introduction

The electricity consumption in Indonesia continues to increase from 2015 to 2020 by 98.89% with business customers dominate the biggest electricity consumption [1]. PT. PLN Persero is the only electricity provider in Indonesia providing higher electricity power for the entire region, including the West Sumatra region. While electricity demand of business customers is increasing, electricity blackouts often occur up to a high frequency of four times a month [2]. 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 their similar characteristics [3]. Thus, customer segmentation can be utilized to predict prospective actions in consuming the services. That customers use and build relationships and enhance customer commitment to building a solid business[3][4]. Several previous studies discussed customer segmentation on customers' electricity consumption [4], [7], [8], [10],[12]. and electricity demand [7], [9]– [11]. 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 [4], [7], [8], [10],[12]. Other studies use the regression method for customer segmentation [7], [9]– [11], 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 [4], [7], [8], [10],[12]. There are also other studies about 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 [11].

Previous research on customer segmentation commonly was based on total electricity consumption per day [4], [7], [8], [10],[12]. Another study only analyzed tariffs, electricity, and total bills by combining K-Means and CRM [11]. Therefore, this study fills the gap by analyzes based on power, peak load electricity consumption, and peak external load electricity consumption by applying a combination of the K-Means clustering method [12]. This study aims to develop a segmentation profile of electricity consumers by examining energy consumption patterns using the data described previously. Then validation method is needed to determine the best number of clusters to determine the clustering measurement. The Elbow method is used for the correct number of groups by looking at the Sum Square Error (SSE) value by looking at the slope of the specified curve [12][13].

Finally, the results from clustering will become the main direction of the Customer Lifetime Value (CLV) [14] method to determine the right grouping. The main factor in the selection of this model is the handling of several large data sets, such as data owned by PT. PLN Persero. The dataset used is the customer transaction data of 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. We want to break down business customers with potential high peak load electricity consumption into several dimensions. This finding can help the company in developing strategies for improving and optimizing electrical power services in the future.

The paper will be organized as follows. The first part describes the background of the problem, gaps in the research, and the purpose. The second part describes a 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.

# Literature Review

Table 1 presents an overview of previous studies that focuses on customer segmentation using transaction/ customer credentials data. As shown, we categorize related articles based on its business context, dataset, segmentation features, and the segmentation method.

Table 1 Reviewed Studies on Customer Segmentation in Electricity Consumption

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Article | Business Context | Dataset | Segmentation  Features | Segmentation  Method |
| [4] | 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) |
| [6] | Electricity Consumption in South Africa | South Africa Electric Load Profile Data from 1994 to 2014 | X=Hour (load profile multiple 1 day)  Y= X multiple All household | K-Means  And Self Organizing Maps (SOM) |
| [5] | 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) |
| [10] | 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)) |
| [9] | Electricity Load Profile | Residential Demand Data during November,2017 until February, 2018 | Identity, Daily Consumption, Load Profile, Peak Hour, Demand | K-means, Fuzzy C-Means (FCM) and Self Organizing Maps (SOM) |
| [8] | 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) |
| [15] | Electricity Demand with Renewable Technologies | Half -hourly energy use for 1 year data | Average energy use,  energy–temperature correlation, entropy of the load-shape representative vector, and distance to  wind generation patterns. | Model (K-Medoids), Validities (average silhouette) |
| [11] | 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[4], [7], [8], [10],[12]. 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, has explored customer grouping by considering patterns of electricity use and electricity demand to meet electricity consumption based on what has been prepared by company[7], [9]– [11].

A context study of load profile electricity [4] 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 Personal Classes (PC). A typical load PC is used for settlement purposes and estimates the amount and Time of Use of electricity used. A series of PCs are manufactured for different market segments (e.g., residential, commercial, industrial) and derived on an average for all customers within a customer class.

Research on electricity consumption in South Africa [6] focus on household customers, which aims to classify customers based on patterns and types 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 the expected behavior of South African households' daily electricity consumption. Another study used electrical load data also in Andalusia, Spain [5], but the research context was about electricity demand. They determine interrelated variables to predict customer segmentation models using a combination model between K-Means clustering and K-medoid clustering. 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 [9] uses electricity demand data to predict electricity loads per day based on the heterogeneity of electricity demand behavior by customers, 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 [9], but this study uses data from smart meters in 2009 [10], 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 [8]. 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 [15] 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 [11]. They used data on customer electricity bills in September 2021 with predictors of power, rate, total kwh, flash sale, total cost, which were 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 to act to customers in the future according to the wisdom that has been carried out.

To the best of our knowledge, most previous studies in customer segmentation on electricity consumption more 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 [6], [8], [11], electricity load profile [9], [10] and daily electricity demand [5], [7], [15]. Then, only one study combined the concept of clustering with CRM[11]; the other research only compared the clustering model to find patterns of electricity use. However, in the concept of clustering electricity consumption for customer segmentation, no one has analyzed based on power, peak-load electricity consumption and off-peak-load electricity consumption and then combines them with the idea of ​​CLV [16] to determine the correct customer group. In this study, clustering was carried out using the K-Means method, with the number of clusters validated using the Elbow method. Then, the clustering results would be classified using CLV to determine the correct group to develop future company services.

# Method

Figure 1 presents the research framework in this study. The framework is adapted from standard methods for building predictive analytical models[17]. There are five stages: data collection, data preparation, choice variables, clustering model, exploration of cluster result.

Diagram

Description automatically generated with low confidence

Figure 1 The Research Framework

**3.1 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 Results of data collection

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

* 1. **Data Preparation**

This section presents the data that will be used for processing in the prediction model, which will be divided into 2, namely as follows:

1. 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 present 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 [18], 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 category

1. 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 results of the analysis of data focus and data cleaning obtained 13 variables with 508,934 data records used for model development.

Table 3 The Result of Data Preparation

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 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 analogue meter and E means 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 of 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 | Discounts given by the company 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 |

* 1. **Choice of 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 [18]. 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 The Result of Variable

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Data Type | Function | Variable Description |
| 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 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 of peak load kwh usage and peak external load kwh used by customers |
| Discount | Double | Discounts given by the company 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 |

* 1. **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[12]. 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[21].

Validation in this study uses the elbow method. The Elbow method in previous studies [13], [20]– [22] was used to determine the number of data clusters to be processed. This method visualizes the number of k = 2 until the number of k is determined. The exact number of groups [xx] is selected when a drastic change is inversely 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 [14].

Step 1: Determine the number of clusters with 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.

* 1. **Exploration of Cluster Result**

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 [14]. The model calculates the distance between zero and the central cluster as a high value and refers to most of the customer loyalty in it [23]. 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[24]. 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) [25]. AHP solves complex multi-criteria problems into a hierarchy [26]. It is helpful for integrated and fuzzy issues based on human brain assessment [27]. The step from AHP is described below [22]– [25]:

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
6. 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 [29] 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 |

There are two programs from the customer relationship strategy. If the company chooses the right approach, it will increase profits and retain customers [29], [30] as follows.

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 [28]– [31]. Implementing a sustainable marketing program from this concept will be explained as follows.

A. Continuous Replenishment Program

This program is used for less profitable customers[35]. Approaches to programs such as partnership programs to encourage increased use of the company's services to customers [35], [36].

B. Business to Business

This program is used for profitable customers[37], [38]. 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 [36]– [39].

2. One to One Marketing

This program is an individual program aimed at satisfying customers' unique needs [43], [44]. This program uses customer information from online news and databases, followed by personal interactions to meet customers' unique needs [45], [46]. Build interactive marketing and post-marketing programs in developing customers using individual customer information [42]– [44]. The application of the one-to-one marketing program from this concept will be explained as follows.

A. Customer Business Development

This program is used for profitable customers [48], [49]. The approach to this program is to assess the benefits of marketing, finance, and management business processes [50], [51]. This program aims to explore the customer's business development by providing the best solutions and consulting regarding customers' services [40]– [42], [44].

B. Retail Account Marketing

This program is used for less profitable customers [47], [53]. 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 [54], [55].

# Result and Discussion

This section, what is done based on the previously made design is as follows: The first step is to find the correct variables to be applied 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 Combine Variable**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| P | FT | TK | POL | PL | POLF | PLF | TC | D | DIM 1 | 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*

Chart, line chart, scatter chart

Description automatically generatedThe 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. The correct number of clusters is determined by looking at the line graph when skewed. From Figure 4, the chart starts to descend at points 3 and 4

Figure 4 The Number of clusters of K

In the 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 green point) in the distribution. The study of the k-means effect in Figure 5 and Figure 6 can be seen below.

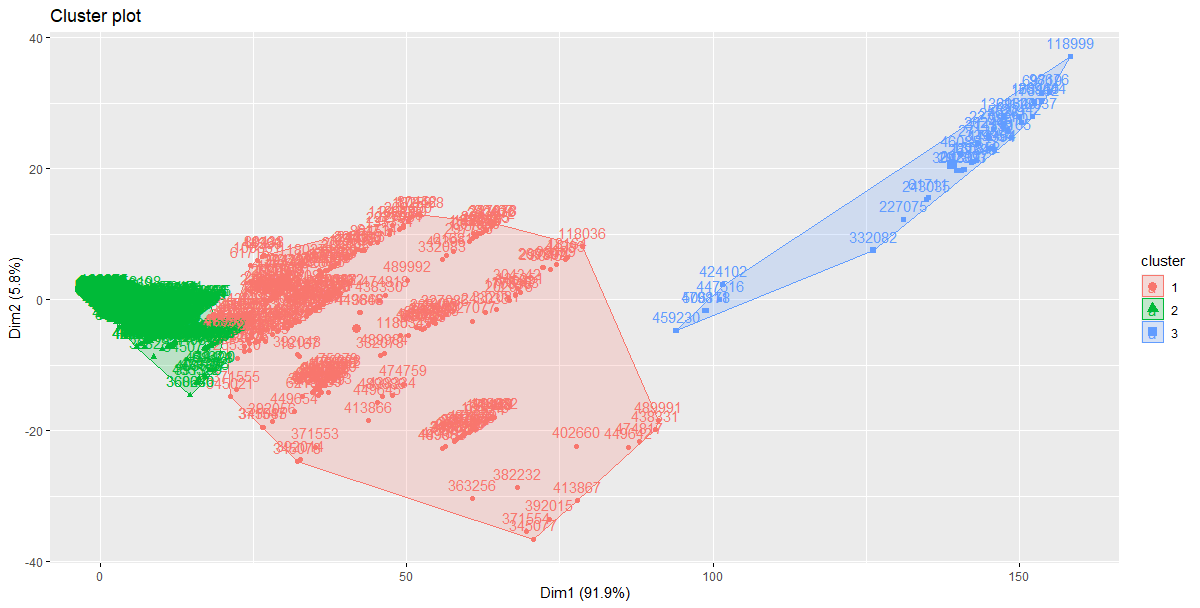


Figure 5 Cluster result of k = 3

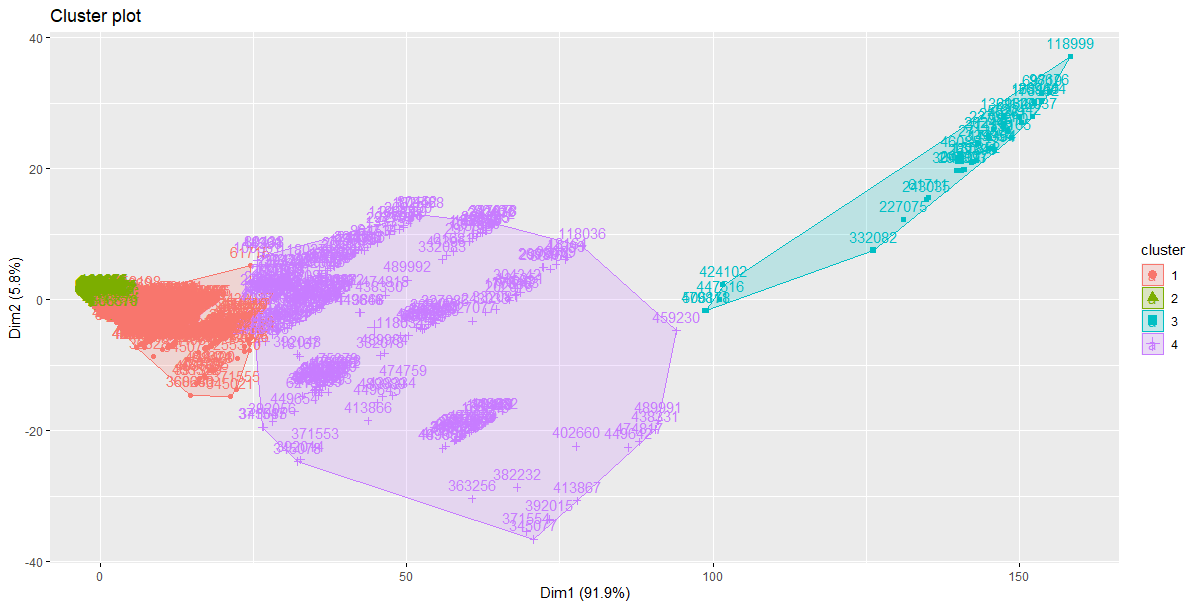


Figure 6 Cluster result of k = 4

Based on the results of clustering using K-Means clustering, Table 7 present 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 much 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 as much as 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 lifetime value. But previously defined the variables used to 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 [23], [24]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 |

The next step, after getting the variables based on the cluster results that have been done and the correct weight value, 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.

Table 9 The 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 |

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 that segment 3 gets the first rank because the value is equal 768,236.6, segment 1 receives the second rank because the value is equal 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 The 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, from the ranking results, targeting will be carried out, which is used to determine the target market based on profitable or less-profitable customers as show in Table 11.

Table 11 Insights 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.

# Conclusion and Future Work

This paper presents a clustering method that aims to find patterns of behavior of electricity users by predicting customers who use electricity at peak loads and customers who use electricity when not at peak loads in West Sumatra, Indonesia, using an unsupervised machine learning, namely the clustering model. The clustering method is applied to the K-Means Clustering model with validation of the number of groupings using the elbow method. The Elbow method is the most suitable for verification and is used to define customer grouping before acting in the grouping model. Then, the clustering results will be applied to CLV by determining the weight value of each variable using the AHP. From the results of CLV, each segment will be ranked based on the highest CLV value and then targeted based on profitable customers or less-profitable customers to get this insight so that it can be developed for company decision making.

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 the use of peak load 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 the context of 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 to develop targeting strategy for their customer and increase the company 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 with the other context of business.

# References

[1] Katadata, “Konsumsi Listrik Nasional Terus Meningkat,” *www.databook.com*, Jan. 09, 2020. https://databoks.katadata.co.id/datapublish/2020/01/10/konsumsi-listrik-nasional-terus-meningkat (accessed Jan. 04, 2022).

[2] Yusrizal, “Sering Mati Lampu, Legislator Minta PLN Segera Bertindak Di Agam,” *Antara Sumbar*, Jun. 17, 2017. https://sumbar.antaranews.com/berita/206333/sering-mati-lampu-legislator-minta-pln-segera-bertindak-di-agam (accessed Jan. 15, 2022).

[3] J. Ye, “Analysis on E-commerce Order Cancellations Using Market Segmentation Approach,” in *ACM International Conference Proceeding Series*, Jan. 2021, pp. 33–40. doi: 10.1145/3450588.3450596.

[4] F. McLoughlin, A. Duffy, and M. Conlon, “A clustering approach to domestic electricity load profile characterisation using smart metering data,” *Applied Energy*, vol. 141, pp. 190–199, Mar. 2015, doi: 10.1016/j.apenergy.2014.12.039.

[5] A. Camero, G. Luque, Y. Bravo, and E. Alba, “Customer segmentation based on the electricity demand signature: The andalusian case,” *Energies*, vol. 11, no. 7, 2018, doi: 10.3390/en11071788.

[6] W. Toussaint and D. Moodley, “Clustering Residential Electricity Consumption Data to Create Archetypes that Capture Household Behaviour in South Africa,” *South African Computer Journal*, vol. 32, no. 2, pp. 1–34, 2020, doi: 10.18489/SACJ.V32I2.845.

[7] M. Hyland, E. Leahy, and R. S. J. Tol, “The potential for segmentation of the retail market for electricity in Ireland,” *Energy Policy*, vol. 61, pp. 349–359, Oct. 2013, doi: 10.1016/j.enpol.2013.05.052.

[8] K. Gajowniczek and T. Zabkowski, “Simulation Study on Clustering Approaches for Short-Term Electricity Forecasting,” *Complexity*, vol. 2018, Apr. 2018, doi: 10.1155/2018/3683969.

[9] E. Lee, J. Kim, and D. Jang, “Load profile segmentation for effective residential demand response program: Method and evidence from Korean pilot study,” *Energies*, vol. 16, no. 3, Mar. 2020, doi: 10.3390/en13061348.

[10] M. Jang, H. C. Jeong, T. Kim, and S. K. Joo, “Load profile-based residential customer segmentation for analyzing customer preferred time-of-use (Tou) tariffs,” *Energies*, vol. 14, no. 19, Oct. 2021, doi: 10.3390/en14196130.

[11] R. Afthoni, M. Hamdhani, A. Fitri Karimah, H. Patria, J. Analitika Bisnis, and F. Magister Manajemen Teknologi, “Seminar Nasional Teknik dan Manajemen Industri dan Call for Paper.”

[12] R. Bapna, P. Goes, A. Gupta, and Y. Jin, “USER HETEROGENEITY AND ITS IMPACT ON ELECTRONIC AUCTION MARKET DESIGN: AN EMPIRICAL EXPLORATION 1,” 2004.

[13] Z. J. Lee, C. Y. Lee, L. Y. Chang, and N. Sano, “Clustering and classification based on distributed automatic feature engineering for customer segmentation,” *Symmetry*, vol. 13, no. 9, Sep. 2021, doi: 10.3390/sym13091557.

[14] F. Marisa, S. S. S. Ahmad, Z. I. M. Yusof, Fachrudin, and T. M. A. Aziz, “Segmentation model of customer lifetime value in Small and Medium Enterprise (SMEs) using K-Means Clustering and LRFM model,” *International Journal of Integrated Engineering*, vol. 11, no. 3, pp. 169–180, 2019, doi: 10.30880/ijie.2019.11.03.018.

[15] S. Bañales, R. Dormido, and N. Duro, “Smart meters time series clustering for demand response applications in the context of high penetration of renewable energy resources,” *Energies*, vol. 14, no. 12, Jun. 2021, doi: 10.3390/en14123458.

[16] F. Marisa, S. S. S. Ahmad, Z. I. M. Yusof, Fachrudin, and T. M. A. Aziz, “Segmentation model of customer lifetime value in Small and Medium Enterprise (SMEs) using K-Means Clustering and LRFM model,” *International Journal of Integrated Engineering*, vol. 11, no. 3, pp. 169–180, 2019, doi: 10.30880/ijie.2019.11.03.018.

[17] G. Shmueli and O. R. Koppius, “PREDICTIVE ANALYTICS IN INFORMATION SYSTEMS RESEARCH 1.”

[18] “Permen ESDM 31 Tahun 2014\_TARIF TENAGA LISTRIK”.

[19] R. Gustriansyah, N. Suhandi, and F. Antony, “Clustering optimization in RFM analysis based on k-means,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 18, no. 1, pp. 470–477, 2019, doi: 10.11591/ijeecs.v18.i1.pp470-477.

[20] P. Bholowalia and A. Kumar, “EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN,” 2014.

[21] A. D. Savitri, F. Abdurrachman Bachtiar, and N. Y. Setiawan, “Segmentasi Pelanggan Menggunakan Metode K-Means Clustering Berdasarkan Model RFM Pada Klinik Kecantikan (Studi Kasus : Belle Crown Malang),” 2018. [Online]. Available: http://j-ptiik.ub.ac.id

[22] F. Marisa, S. Sakinah Syed Ahmad, Z. Izzah Mohd Yusof, T. Mohammad Akhriza, W. Purnomowati, and R. Kumar Pandey, “The Analyze of Relationship between Revenue and Customer Payment Methods in Small Medium Enterprise Based on Clustering K-Means,” in *Journal of Physics: Conference Series*, Jul. 2021, vol. 1908, no. 1. doi: 10.1088/1742-6596/1908/1/012021.

[23] S. M. S. Hosseini, A. Maleki, and M. R. Gholamian, “Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty,” *Expert Systems with Applications*, vol. 37, no. 7, pp. 5259–5264, Jul. 2010, doi: 10.1016/j.eswa.2009.12.070.

[24] M. Khajvand, K. Zolfaghar, S. Ashoori, and S. Alizadeh, “Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study,” in *Procedia Computer Science*, 2011, vol. 3, pp. 57–63. doi: 10.1016/j.procs.2010.12.011.

[25] K. M. Al and -Subhi Al-Harbi, “Application of the AHP in project management.” [Online]. Available: www.elsevier.com/locate/ijproman

[26] “AN-EVALUATION-MODEL-FOR-FOREIGN-DIRECT-INVESTMENT-PERFORMANCE-OF-FREE-TRADE-PORT-ZONESPromet--Traffic--Traffico”.

[27] P. Agustine, H. Parung, P. Davey, and C. Frid, “Management Strategies to Protect Coastal Areas from Oil-Polluted Seawater (A Case Study of Coastal Areas in Bekasi Regency),” in *IOP Conference Series: Earth and Environmental Science*, Dec. 2021, vol. 921, no. 1. doi: 10.1088/1755-1315/921/1/012049.

[28] P. Park, D. Kim, S. Lee, and J. Whang, “Toward an economically sustainable casino industry: A development of customer value indicators using an analytic hierarchy process,” *Sustainability (Switzerland)*, vol. 10, no. 11, Nov. 2018, doi: 10.3390/su10114255.

[29] S. M. S. Hosseini, A. Maleki, and M. R. Gholamian, “Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty,” *Expert Systems with Applications*, vol. 37, no. 7, pp. 5259–5264, Jul. 2010, doi: 10.1016/j.eswa.2009.12.070.

[30] A. Irawan, “Analisis Customer Relationship Management (CRM) terhadap Customer Relationship Quality (CRQ) dan Customer Lifetime Value (CLV) Rumah Zakat di Indonesia”, [Online]. Available: http://customerattuned.com/blog/the-value-of-trust/.

[31] “Real Bounce Forward Experimental Evidence on Destination Crisis Marketing, Destination Trust, e-WOM and Global Expat’s Willingness to Travel during and after COVID-19 Enhanced Reader”.

[32] Á. Leal, E. W. Mainardes, and L. M. Pascuci, “Environmental marketing: Acceptance of price premium in the Brazilian apparel industry,” *Revista Brasileira de Marketing*, vol. 20, no. 3, Jul. 2021, doi: 10.5585/REMARK.V20I3.19495.

[33] S. R. Moro, P. A. Cauchick-Miguel, and G. H. de S. Mendes, “Literature analysis on product-service systems business model: A promising research field,” *Brazilian Journal of Operations and Production Management*, vol. 19, no. 1. Associacao Brasileira de Engenharia de Producao, Jun. 04, 2022. doi: 10.14488/BJOPM.2021.043.

[34] T. Balanovska, O. Gogulya, K. Dramaretska, V. Voskolupov, and V. Holik, “USING MARKETING MANAGEMENT TO ENSURE COMPETITIVENESS OF AGRICULTURAL ENTERPRISES,” 2021. [Online]. Available: http://are

[35] Y. C. Tsao, M. Setiawati, T. Linh Vu, and A. Sudiarso, “Designing a supply chain network under a dynamic discounting-based credit payment program,” *RAIRO - Operations Research*, vol. 55, no. 4, pp. 2545–2565, Jul. 2021, doi: 10.1051/ro/2021111.

[36] D. C. de Freitas, L. G. de Oliveira, and R. L. C. Alcântara, “A theoretical framework to adopt collaborative initiatives in supply chains,” *Gestao e Producao*, vol. 26, no. 3, 2019, doi: 10.1590/0104-530X-4194-19.

[37] K. C. Rao, S. Velidandla, C. L. Scott, and P. Drechsel, “Business Models for Fecal Sludge Management in India,” *RESOURCE RECOVERY & REUSE SERIES*, vol. 18.

[38] M. J. Foncubierta-Rodríguez, F. Galiana-Tonda, and M. del Mar Galiana Rubia, “Chambers of Commerce: A new Management. The balanced scorecard approach for spanish chambers,” *CIRIEC-Espana Revista de Economia Publica, Social y Cooperativa*, no. 99, pp. 273–308, Jul. 2020, doi: 10.7203/CIRIEC-E.99.14602.

[39] C. R. Cunha, V. Mendonça, E. P. Morais, J. Fernandes, and I. Letra, “Using mobile devices and apps to assist the elder population in rural areas and generate business opportunities,” *IBIMA Business Review*, vol. 2020, Dec. 2020, doi: 10.5171/2020.585068.

[40] H. Malm, M. Pikkarainen, and E. Hyrkäs, “Impacts of open innovation on company business models: A case study of demand-driven co-creation,” *Journal of Innovation Management*, vol. 8, no. 3, pp. 75–108, Dec. 2020, doi: 10.24840/2183-0606\_008.003\_0006.

[41] B. Gavurova, M. Cepel, J. Belas, and J. Dvorsky, “Strategic Management in SMEs and Its Significance for Enhancing the Competitiveness in the V4 Countries-A Comparative Analysis,” *Management and Marketing*, vol. 15, no. 4, pp. 557–569, Dec. 2020, doi: 10.2478/mmcks-2020-0032.

[42] R. Schneider *et al.*, “Differential impact of government lockdown policies on reducing air pollution levels and related mortality in Europe,” *Scientific Reports*, vol. 12, no. 1, p. 726, Dec. 2022, doi: 10.1038/s41598-021-04277-6.

[43] J. Gil-Quintana and E. Vida de León, “Educational influencers on instagram: Analysis of educational channels, audiences, and economic performance,” *Publications*, vol. 9, no. 4, Dec. 2021, doi: 10.3390/publications9040043.

[44] K. Kafkas, Z. N. Perdahçı, and M. N. Aydın, “Discovering customer purchase patterns in product communities: An empirical study on co-purchase behavior in an online marketplace,” *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 16, no. 7, pp. 2965–2980, Dec. 2021, doi: 10.3390/jtaer16070162.

[45] N. Baniasadi, D. Samari, S. J. F. Hosseini, and M. O. Najafabadi, “Strategic study of total innovation management and its relationship with marketing capabilities in palm conversion and complementary industries,” *Journal of Innovation and Entrepreneurship*, vol. 10, no. 1, Dec. 2021, doi: 10.1186/s13731-021-00179-z.

[46] W. Xie, B. Chen, F. Huang, and J. He, “Coordination Of A Supply Chain With A Loss-Averse Retailer Under Supply Uncertainty And Marketing Effort,” *Journal of Industrial and Management Optimization*, vol. 17, no. 6, pp. 3393–3415, Nov. 2021, doi: 10.3934/jimo.2020125.

[47] Q. Yan, C. Qin, M. Nie, and L. Yang, “Forecasting the Electricity Demand and Market Shares in Retail Electricity Market Based on System Dynamics and Markov Chain,” *Mathematical Problems in Engineering*, vol. 2018, 2018, doi: 10.1155/2018/4671850.

[48] S. C. Daat, M. A. Sanggenafa, and R. Larasati, “The role of intellectual capital on financial performance of smes,” *Universal Journal of Accounting and Finance*, vol. 9, no. 6, pp. 1312–1321, Dec. 2021, doi: 10.13189/ujaf.2021.090610.

[49] K. Borisavljević and G. Radosavljević, “Application of logistics model in analysing relationship marketing in travel agencies,” *Zbornik Radova Ekonomskog Fakultet au Rijeci*, vol. 39, no. 1, pp. 87–112, Jun. 2021, doi: 10.18045/zbefri.2021.1.87.

[50] E. Kulej-Dudek, “Ecolabnet service packages as a response to the needs of manufacturing enterprises in the SME sector of the Baltic Sea Region,” *Production Engineering Archives*, vol. 27, no. 4, pp. 265–271, Dec. 2021, doi: 10.30657/pea.2021.27.35.

[51] J. Koponen, S. Julkunen, M. Gabrielsson, and E. B. Pullins, “An intercultural, interpersonal relationship development framework,” *International Marketing Review*, vol. 38, no. 6, pp. 1189–1216, Oct. 2021, doi: 10.1108/IMR-11-2019-0267.

[52] H. T. N. Huynh, H. D. X. Trieu, P. van Nguyen, T. G. Tran, and L. N. H. Lam, “Explicating brand equity in the information technology sector in Vietnam,” *Administrative Sciences*, vol. 11, no. 4, Dec. 2021, doi: 10.3390/admsci11040128.

[53] T. B. Yudhya, “Retail store image: A study of the matahari department store (at Bandung Indonesia),” *Humanities and Social Sciences Reviews*, vol. 7, no. 5, pp. 98–102, Sep. 2019, doi: 10.18510/hssr.2019.7513.

[54] S. Sekizaki, I. Nishizaki, and T. Hayashida, “Impact of Retailer and Consumer Behavior on Voltage in Distribution Network under Liberalization of Electricity Retail Market,” *Electrical Engineering in Japan (English translation of Denki Gakkai Ronbunshi)*, vol. 194, no. 4, pp. 27–41, Mar. 2016, doi: 10.1002/eej.22743.

[55] F. M. Dias, M. P. V. de Oliveira, H. Z. Filho, and A. L. Rodrigues, “Analytical guidance or intuition? what guides management decisions on the most important customer value attributes in the supermarket retail?,” *Revista Brasileira de Marketing*, vol. 20, no. 2, pp. 385–414, 2021, doi: 10.5585/REMARK.V20I2.16106.