**SEGMENTATION USING CUSTOMER LIFETIME VALUE  
HYBRID K-MEANS AND ANALYTIC HIERRARCHY PROCESS**

**THESIS**

**In partial fulfilment of the requirements**

**For the Degree of Master of Science in Management**

**From Institut Teknologi Bandung**

**By**

**Radit Rahmadhan**

**Student ID: 29020003**

**(Master Program of Science in Management)**

****

**INSTITUT TEKNOLOGI BANDUNG**

**2022**

# ABSTRACT

**SEGMENTATION USING CUSTOMER LIFETIME VALUE  
HYBRID K-MEANS AND ANALYTIC HIERRARCHY PROCESS**

By

**Radit Rahmadhan**

**Student ID: 29020003**

**(Master Program of Science in Management)**

Developing predictive analytics based on understanding customers' electricity consumption patterns is essential to effectively manage the increasing electricity demand. This study presents a hybrid approach to customer segmentation by combining K-Means clustering, the concept of customer lifetime value, and an analytic hierarchy process to better understand customers' electricity consumption behaviour. We use K-Means clustering to identify initial market segments. Next, we evaluate and validate the customer segmentation results using the customer lifetime value concept and the analytical hierarchy process. Segment 1 has 282 business customers with a total capacity of 938,837 kWh, peak load usage of 27,827 kWh, and non-peak load of 115,194. In segment 2, there are 508 customers with a total capacity of 938,837 kWh, a peak load usage of 27,827 kWh, and a non-peak load of 115,194. In segment 2, there are 508,615 business customers with a total power of 4,260 kWh, then a peak load of 35 kWh and a non-peak load of 544. In segment 3, there are 37 business customers with a total power of 2,226,351 kWh, then a peak load of 123,297 kWh and a non-peak load of 390,803. Strategies to be taken based on the segmentation of these three customers will be integrated with CRM. For the least profitable segment, we propose an ongoing partnership program to encourage increased electricity consumption during non-peak periods and retail account marketing. For profitable and medium profitable customers, we propose a premium business to business approach that can accommodate their increased energy consumption without excessive electricity usage during peak periods. This approach will be supported by dedicated executive accounts for these customers.

Keywords: Customer Analytics, Electricity, Customer Lifetime Value, Customer   
 Relationship Management, K-Means Clustering, Analytical Hierarchy   
 Process.

# ABSTRAK

***SEGMENTASI MENGGUNAKAN NILAI SEUMUR HIDUP PELANGGAN HIBRIDA K-MEANS DAN PROSES HIERARKI ANALITIK***

*Oleh*

**Radit Rahmadhan**

**29020003**

**(Program Studi Magister Sains Manajemen)**

*Mengembangkan analisis prediktif berdasarkan pemahaman pola konsumsi listrik pelanggan sangat penting untuk mengelola permintaan listrik yang meningkat secara efektif. Studi ini menyajikan pendekatan hibrida untuk segmentasi pelanggan dengan menggabungkan pengelompokan K-Means, konsep nilai masa pakai pelanggan, dan proses hierarki analitik untuk lebih memahami perilaku konsumsi listrik pelanggan. Kami menggunakan K-Means clustering untuk mengidentifikasi segmen pasar awal. Selanjutnya, kami mengevaluasi dan memvalidasi hasil segmentasi pelanggan dengan menggunakan konsep nilai seumur hidup pelanggan dan proses hirarki analitis. Segmen 1 memiliki 282 pelanggan bisnis dengan total kapasitas 938.837 kWh, penggunaan beban puncak 27.827 kWh, dan beban non puncak 115.194. Pada segmen 2, terdapat 508 pelanggan dengan total kapasitas 938.837 kWh, penggunaan beban puncak 27.827 kWh, dan beban non puncak 115.194. Pada segmen 2 terdapat 508.615 pelanggan bisnis dengan total daya sebesar 4.260 kWh, kemudian beban puncak sebanyak 35 kWh dan beban non puncak sebanyak 544. Pada segmen 3, terdapat 37 pelanggan bisnis dengan total daya sebesar 2.226.351 kWh, kemudian beban puncak sebanyak 123,297 kWh dan beban non puncak sebanyak 390. 803.Strategi yang akan diambil berdasarkan segmentasi tiga pelanggan ini akan diintegrasikan dengan CRM. Untuk segmen yang paling tidak menguntungkan, kami mengusulkan program kemitraan yang berkelanjutan untuk mendorong peningkatan konsumsi listrik selama periode non puncak dan pemasaran akun ritel. Untuk pelanggan yang menguntungkan dan sedang menguntungkan, kami mengusulkan pendekatan premium business to business yang dapat mengakomodasi peningkatan konsumsi energi mereka tanpa penggunaan listrik yang berlebihan pada periode puncak. Pendekatan ini akan didukung oleh rekening eksekutif khusus untuk pelanggan-pelanggan tersebut.*

Keywords: *Analisis Pelanggan, Listrik, Nilai Seumur Hidup Pelanggan, Pelanggan*

*Manajemen Hubungan Pelanggan, K-Means Clustering, Analytical Hierarchy*

*Proses.*

**SEGMENTATION USING CUSTOMER LIFETIME VALUE  
HYBRID K-MEANS AND ANALYTIC HIERRARCHY PROCESS**

By

**Radit Rahmadhan**

**Student ID: 29020003**

**Master Program of Science in Management**

Institut Teknologi Bandung

Approved

December

Supervisor

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Meditya Wasesa, S.T., M.Sc., Ph.D.)

**THE GUIDANCE FOR USING THE THESIS**

Unpublished masters’ thesis is registered and available in the Library of Bandung Institut of Technology, and is open for public, provided that the author owns the copyright in accordance with the intellectual property rights that are applicable in the Library of Institut Teknologi Bandung. Bibliographical references are allowed to be use in a limited manner, however the citation and summarization can only be proceed upon the author’s permission and must include the scientific norm of stating this thesis as the source.

This thesis has to be cited as:

Rahmadhan, Radit. (2022): *Segmentation using Customers Lifetime Value: Hybrid  
 K-Means Clustering and Analytic Hierarchy Process, Masters’ Thesis,* Institut Teknologi Bandung.

Reproduction or publication of parts or whole of the thesis must be under the consent of the Head of Graduate School, Bandung Institute of Technology.

*In dedication to my beloved parents, brothers, and family, supervisor and advisor, and friends who always support me.*

**STATEMENT OF AUTHORSHIP**

I hereby declare that I am the sole author of this thesis and to the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made. I further declare that this thesis has not been previously submitted to obtain a degree at this or any other higher education institution.

Signature:

Bandung, December 09,2022

**ACKNOWLEDGMENTS**

The introduction page is printed on a new page. On this page, the Masters’ students may have the opportunity to express their gratitude in writing to other mentors and or individuals who have provided guidance; advise and critics; as well as to those who have assisted in conducting the research; whether individuals or bodies that have provided financial assistance, and so forth.

In the forewords, authors may use all kinds of writing varieties. However it is advisable to keep it in a standard written sentences. Acknowledgments should be made in excessive and limited only to the "scientifically related".

**TABLE OF CONTENT**

[ABSTRACT i](#_Toc121490514)

[ABSTRAK ii](#_Toc121490515)

[Chapter I Introduction 1](#_Toc121490516)

[I.1 Background 1](#_Toc121490517)

[I.2 Research Objectives 3](#_Toc121490518)

[I.3 Research Question 4](#_Toc121490519)

[I.4 Research approach and methods 4](#_Toc121490520)

[I.5 Research Scope and Limitations 4](#_Toc121490521)

[I.6 Writing Structure 5](#_Toc121490522)

[Chapter II Literature Review 6](#_Toc121490523)

[II.1 Previous Segmentation Studies Based on Electricity Consumption Data 6](#_Toc121490524)

[II.2 Previous Studies on Segmentation Based on Customer Lifetime Value 9](#_Toc121490525)

[II.3 Marketing Strategy in Customer Relationship Management 11](#_Toc121490526)

[II.4 Research Position 13](#_Toc121490527)

[Chapter III Research Methodology 16](#_Toc121490528)

[III.1 Research Philosophical Position 16](#_Toc121490529)

[III.2 Research Framework 19](#_Toc121490530)

[III.3 Data Collection 21](#_Toc121490531)

[III.4 Data Preparation 21](#_Toc121490532)

[III.5 Choice of Variable 24](#_Toc121490533)

[III.6 Clustering Model 25](#_Toc121490534)

[III.7 Marketing Strategy Definition 26](#_Toc121490535)

[Chapter IV Results and Analysis 29](#_Toc121490536)

[IV.1 Result of Clustering Model 29](#_Toc121490537)

[IV.2 Result Customer Lifetime Value (CLV) 32](#_Toc121490538)

[IV.3 Implement Customer Relationship Management Strategies 34](#_Toc121490539)

[Chapter V Summary And Conclusion 36](#_Toc121490540)

[V.1 Conclusion 36](#_Toc121490541)

[V.2 Research and Practical Implications 38](#_Toc121490542)

[V.3 Limitation and Further Research 38](#_Toc121490543)

[REFERENCES 39](#_Toc121490544)

[APPENDICES 46](#_Toc121490545)

**LIST OF APPENDICES**

**LIST OF FIGURES**

[Figure III.1 Research onion (Saunders et al., 2016) 18](#_Toc121490661)

[Figure III.2 The research framework 20](#_Toc121490662)

[Figure III.3 Total Electricity Consumption Bases on Region 22](#_Toc121490663)

[Figure III.4 Total electricity consumption based on customer energy 22](#_Toc121490664)

[Figure IV.1 The Number of clusters of K 30](file:///C:\Users\radit\Downloads\Writing\Draft%20Tesis\fulldraft.docx#_Toc121490665)

[Figure IV.2 Cluster Visualization (k=3) 31](#_Toc121490666)

[Figure IV.3 Cluster Visualization (k=4) 31](#_Toc121490667)

**LIST OF TABLES**

[Table II.1 CUSTOMER SEGMENTATION BASED ON ELECTRICITY CONSUMPTION 7](#_Toc121490725)

[Table II.2 Criteria of literature on customer segmentation 13](#_Toc121490726)

[Table III.1 Pragmatism philosophy assumptions (Saunders et al., 2016) 17](#_Toc121490727)

[Table III.2 Descriptive of Data Collection 21](#_Toc121490728)

[Table III.3 Descriptive of Data Cleaning 23](#_Toc121490729)

[Table III.4 Descriptive of Potential Variable 25](#_Toc121490730)

[Table III.5 Customer Relation Strategy 28](#_Toc121490731)

[Table IV.1 The Combination of Clustering Variables 29](#_Toc121490732)

[Table IV.2 The Detail of The Clustering Results 32](#_Toc121490733)

[Table IV.3 Weight of AHP Results 32](#_Toc121490734)

[Table IV.4 Result of Customer Lifetime Value in Each Cluster 33](#_Toc121490735)

[Table IV.5 Result of Customer Ranking 33](#_Toc121490736)

[Table IV.6 Insight from CRM Decision Development 34](#_Toc121490737)

# Introduction

## Background

Electricity is a vital energy for the sustainability of human activities both for individuals, community groups and the industrial world (Yan et al., 2018). As the development of electrical energy is more widely used to carry out activities with enormous benefits where various equipment to meet the needs of life are operated using electrical energy (Azadeh & Faiz, 2011). Community activities tend to increase over time. The increase in activities encourages an increase in the operation of equipment with electric power (Hyland et al., 2013). During the electricity consumption period from 2015 to 2020, Indonesia experienced an increase in electricity consumption of around 98.99% with business customers dominating the largest electricity consumption (Katadata, 2020). PT PLN Persero is the only electricity provider in Indonesia that provides higher power for all regions, including West Sumatra. While the electricity demand of business customers is increasing, power outages often occur up to a high frequency of four times a month. Based on the results of the data analysis that has been carried out, power outages cause the average electricity usage time of business customers to be below 50 hours per month. This is due to customers who use power above 200 thousand during peak load rather than electricity outside peak hours. During non-peak load hours, the usage is low.

Based on this, power companies should understand the characteristics of customers' electricity usage to maximize electricity distribution. For example, the low consumption of business customers due to power outages (under 50 hours per month) can be improved. Customer segmentation is one way to understand and map customer preferences. According to previous research, customer segmentation refers to the grouping of customers based on similar characteristics (McLoughlin et al., 2015). Thus, customer segmentation can be utilized to predict prospective actions in consuming services. That customers use and build relationships and increase customer commitment to build a solid business (Camero et al., 2018; Park et al., 2018a).

Some previous research discussed customer segmentation on customer electricity consumption (Camero et al., 2018; Gajowniczek & Zabkowski, 2018; Z. J. Lee et al., 2021; Li et al., 2018) and electricity demand (Bañales et al., 2021; Gustriansyah et al., 2019; Z. J. Lee et al., 2021; Yan et al., 2018). The research context is more about finding new customer behavior patterns in consuming electricity and more methods that use a combination of K-Means and Self Organizing Maps (SOM) and other clustering methods (Bañales et al., 2021; Gustriansyah et al., 2019; Z. J. Lee et al., 2021; Marisa et al., 2019; Yan et al., 2018) .

Other studies use regression methods for customer segmentation (Bañales et al., 2021; Gustriansyah et al., 2019; Z. J. Lee et al., 2021; Yan et al., 2018), they want to predict future electricity consumption to meet electricity demand from customers. The results of some previous studies provide recommendations for optimizing the use of electricity against the electricity that has been provided (Bañales et al., 2021; Gustriansyah et al., 2019; Z. J. Lee et al., 2021; Marisa et al., 2019; Yan et al., 2018). There are also other studies on analyzing customer characteristics by applying the K-Means Clustering model by analyzing tariffs, power, the number of bills paid and then from the results of the model. The concept is used in Customer Relationship Management (CRM) to gain insight or make company business decisions (Gustriansyah et al., 2019). Previous research on customer segmentation is generally based on total electricity consumption per day (Bañales et al., 2021; Gustriansyah et al., 2019; Z. J. Lee et al., 2021; Marisa et al., 2019; Yan et al., 2018).Other studies only analyzed tariffs, electricity, and total bills by combining K-Means and CRM (Gustriansyah et al., 2019).

Correspondingly, this research study aims to fill the gap by developing a segmentation model that can reflect electricity consumption behavior. The findings can help electricity companies improve their strategies in targeting customers according to their characteristics. We include three variables in the development of the segmentation model: power capacity, peak load consumption, and non-peak load consumption. We used K-means, Analytic Hierarchy Process (AHP) approach, and customer lifetime value aspects. The dataset used is the customer transaction data of PT PLN Persero West Sumatra Region from 2019 to 2020.

## Research Objectives

The previous section highlighted the need for an accurate electricity consumption customer segmentation prediction model that can divide customers based on the right segmentation. It also discusses the appropriate marketing strategy according to the characteristics of their customers. Therefore, this thesis focuses on developing a hybrid model of electricity consumption customer segmentation in West Sumatra Zone using electricity customer transaction data from January 2019 to December 2020. The prediction model is developed based on a hybrid model that is a combination of machine learning, namely K-Means Clustering, Analytic Hierarchy Process (AHP) approach, Customer Lifetime Value (CLV) Aspect and Customer Relationship Management (CRM).

For this purpose, the research questions are formulated as follows:

1. How to develop an accurate customer segmentation model according to the characteristics of electricity customers using West Sumatra Zone business customer transaction data?
2. How to implement marketing strategies according to customer criteria based on the results of the customer segmentation model?

## Research Question

Considering the research question formulated above, the objectives of this study are defined as follows:

1. To develop a hybrid model of customer segmentation to find the right grouping of electricity customers according to their consumption patterns.
2. To apply the concept of Customer Relationship Management (CRM) Strategy according to the characteristics of its customers in order to meet the demand for electricity effectively.

## Research approach and methods

This thesis is a design science study that focuses on developing a hybrid model of customer segmentation models. This research uses data on business customer transactions of PT PLN Persero in West Sumatra region from January 2019 to December 2020. This research adopted and modify the predictive analytics framework by (Shmueli & Koppius, 2011) to develop a hybrid segmentation model, which consists of data collection, data preparation, model development, model evaluation, model usage and reporting.

## Research Scope and Limitations

To emphasize the focus of this study, the research scope and limitations are defined as follows:

1. This study focuses on the electricity consumption of business customers specific to the Padang region. The selection of variables used power capacity, peak load consumption, and non-peak load consumption. Further research could investigate a wider selection of regions and a deeper selection of variables.
2. This study used one month of business customer transaction data. Future studies can use one year or more to be further examined as input for segmentation models.
3. This study examines proposals using a combination of machine learning models namely K-Means Clustering, Customer Lifetime Value and Customer Relationship Management (CRM) strategies. Other advance methods can be investigated in further research.

## Writing Structure

This thesis is organized as follows: Chapter I presents the overview of the research background, research questions and correspondence objectives, research approach and method, research scope and limitations, and the writing structure of this thesis. Chapter II reviews related literature, identify the knowledge gap and presents the position of this study. Chapter III discusses the research philosophy, paradigm and methodology used in this study, which consist of data collection, data preparation, choice of variables, clustering model, and marketing strategy. Chapter IV presents the empirical results and analysis of the proposed hybrid segmentation model. Finally, Chapter V concludes the findings of this study, contributions, and present limitations alongside suggestions for future research.

# Literature Review

This chapter present a review of the related literature of this study. The discussion of relevant concepts in this study is presented, including Customer Segmentation Based on Electricity Consumption Data, Customer Segmentation Based on Customer Lifetime Value. The related literature is classified as to clarify the knowledge gap and this research’s position. Finally, the proposed research is presented at the end of this chapter.

## Previous Segmentation Studies Based on Electricity Consumption Data

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.

Previous studies in customer segmentation in electricity consumption have explored various dimensions of the customer clustering problem (Bañales et al., 2021; Gustriansyah et al., 2019; Z. J. Lee et al., 2021; Marisa et al., 2019; Yan et al., 2018). 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 (Bañales et al., 2021; Gustriansyah et al., 2019; Z. J. Lee et al., 2021; Yan et al., 2018).

Table II.1 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 1 day)  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 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) |
| (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 et al., 2021) | 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) |
| (Afthoni et al., n.d.) | Electricity Consumption in Indonesia | Customer Transaction in September 2021 | Rate, Power, Total kWh, Total Cost, Flash Time | Variable selection with correlation  Model (K-Means)  Validity (Silhouette Method)  Explores (Customer Relationship Management (CRM)) |

A context study of load profile electricity by (Camero et al., 2018) 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 by (E. Lee et al., 2020)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 by (Jang et al., 2021) used electrical load data also in Andalusia, Spain, 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 by (Bañales et al., 2021) 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 (Bañales et al., 2021), but this study uses data from smart meters in 2009 (Li et al., 2018), they use a regression model with an evaluation of the root mean square error for customer segmentation based on electricity demand used, age, and income from the customer. The aim is to find new customer electricity usage behavior patterns based on predetermined variables. Another study uses six regression models to predict daily electricity consumption based on the total electricity consumption used by customers (Gajowniczek & Zabkowski, 2018). They compared the models to find new patterns of customers' daily electricity usage.

Research on the context of looking for energy reserves based on the number of customer electricity requests by (Tsao et al., 2021) uses data on customers' half-day electricity usage by selecting variables based on the average amount processed by adding wind variables as alternative electrical energy. This study uses the K-Medoid model and the Silhouette method to validate the number of clusters to apply an efficient time series clustering methodology that explicitly considers the pattern of renewable energy generation. Other research on the context of electricity consumption in Indonesia by (Gustriansyah et al., 2019). 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.

## Previous Studies on Segmentation Based on Customer Lifetime Value

Previous studies in customer segmentation have explored various dimensions of customer clustering problems (Foncubierta-Rodríguez et al., 2020; Gil-Quintana & Vida de León, 2021; Rao et al., 2020). Many of them use the marketing context as a case study. The K-Means clustering model and Customer Lifetime Value explores customer grouping by considering the specified product preferences and predicting customer behavior in buying products offered by the company (Ye, 2021).

A context study in marketing combines the Customer Lifetime Value (CLV) and K-Means models in each customer segment (Foncubierta-Rodríguez et al., 2020). The grouping uses the K-Means Clustering method based on the LRFM (Length, Recency, Frequency, Monetary) model. The cluster formation process uses the Elbow method. The CLV value is generated from the multiplication of the LRFM normalization results, and then the LFRM weight value uses the Analytical Hierarchy Process (AHP). Based on the LRFM matrix, this cluster has a high loyalty value, with the symbol LRFM being a loyal customer (the best segment with a high customer loyalty value). Based on the LRFM symbol, companies can create strategies to retain customers and acquire loyal customers with high profitability.

Another study with a supermarket marketing context with the same objective and predictor variables used historical customer data processed with a combination of LRFM models to determine data selection on potential customer purchases (Ye, 2021) . The K-means clustering model by (Koponen et al., 2021) 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 by (Gil-Quintana & Vida de León, 2021) also has the same objective(Abdi & Abolmakarem, 2019) , but they use eight validation methods in determining the correct number of groupings. Another transportation survey uses the K-Means Clustering model and the CLV model to group customers (Kafkas et al., 2021) with the same research objective(Park et al., 2018a). It also has similar goals and models by (Park et al., 2018b) to marketing research in Telecommunication Companies (Baniasadi et al., 2021). However, they do not use the CLV model but use the Neural Network to classify priority customers after getting the results from clustering.

## Marketing Strategy in Customer Relationship Management

Previous research, marketing strategy in customer segmentation was determined based on the CLV result, then we can evaluate target that aims to develop customer service improvement strategies based on the concept of customer relationship CRM (Xie et al., 2021).

There are two programs from the customer relationship strategy. If the company chooses the right approach, it will increase profits and retain customers by (Borisavljević & Radosavljević, 2021; Xie et al., 2021) 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 (Dias et al., 2021; Sekizaki et al., 2016; Yudhya, 2019). 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 (Yudhya, 2019). Approaches to programs such as partnership programs to encourage increased use of the company's services to customers (Dias et al., 2021; Yudhya, 2019).

B. Business to Business

This program is used for profitable customers (Kulej-Dudek, 2021; Sekizaki et al., 2016). The approach to this program is like providing special executive services to customers to improve service, so that customer trust will increase and become more loyal (Baniasadi et al., 2021; R. Hosseini et al., 2021; Kafkas et al., 2021; Xie et al., 2021).

2. One to One Marketing

This program is an individual program aimed at satisfying customers' unique needs (Gil-Quintana & Vida de León, 2021; Kafkas et al., 2021). This program uses customer information from online news and databases, followed by personal interactions to meet customers' unique needs (Baniasadi et al., 2021; Chen et al., 2015). Build interactive marketing and post-marketing programs in developing customers using individual customer information (Borisavljević & Radosavljević, 2021; Kulej-Dudek, 2021). 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 (Borisavljević & Radosavljević, 2021; Daat et al., 2021). The approach to this program is to assess the benefits of marketing, finance, and management business processes (Borisavljević & Radosavljević, 2021; Kulej-Dudek, 2021). This program aims to explore the customer's business development by providing the best solutions and consulting regarding customers' services (Borisavljević & Radosavljević, 2021; Daat et al., 2021; Koponen et al., 2021; Kulej-Dudek, 2021).

B. Retail Account Marketing

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

## Research Position

In order to identify the gap and position of this study, this study classified the literature based on four criteria, as shown in Table II.2 . The study business context are categorized into three : Electricity Load Profile,Electricity Consumption, Electricity Demand. The data used in the prediction into four : Experimental Data,Electric Load Profile data,Electric consumption data, Customer Transaction Data. The featured used into four : Load Profile, Total kWh, Daily consumption, Demand. Seven different method are K-Means, K-Medoid, Self-organising Maps, Fuzzy K-Means, Regression, CLV, CRM Strategy. Finally, this study also classified the data period used to develop the prediction segmentation model.

Table II.2 Criteria of literature on customer segmentation

|  |  |  |  |
| --- | --- | --- | --- |
| Criteria | Description | Code | Specification |
| Business Context | Investigated business themes | 1 | Electricity Load Profile |
|  | 2 | Electricity Consumption |
|  | 3 | Electricity Demand |
| Data | The category of Internet data used in the prediction model | 1 | Experimental Data |
|  | 2 | Electric Load Profile data |
|  | 3 | Electric consumption data |
|  | 4 | Customer Transaction Data |
| Features | The variables used in his research | 1 | Load Profile |
|  | 2 | Total kWh |
|  | 3 | Daily consumption |
|  | 4 | Demand |
| Method | The method to develop prediction segmentation model | 1 | K-Means |
|  | 2 | K-Medoid |
|  | 3 | Self-organising Maps |
|  | 4 | Fuzzy K-Means |
|  | 5 | Regression |
|  |  | 6 | CLV |
|  |  | 7 | CRM Strategy |

Table II.3 shows the research position of this study. 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 (Afthoni et al., n.d.; Gajowniczek & Zabkowski, 2018; Toussaint & Moodley, 2020), electricity load profile (E. Lee et al., 2020)and daily electricity demand (Camero et al., 2018; Z. J. Lee et al., 2021).

To the best our knowledge, it is difficult to find only one study that combines the concept of clustering with CRM. Other studies only compare clustering models to find electricity usage patterns. However, in the concept of electricity consumption clustering for customer segmentation, no one has analyzed based on power, peak load electricity consumption and off-peak electricity consumption and then combined with the idea of CLV (Marisa et al., 2019) to determine the right customer group. In this study, clustering is performed using the K-Means method, with the number of clusters validated using the Elbow method. Then, the clustering results will be classified using CLV. The CLV calculation will involve the value of the clustering variable and the weight value of the clustering variable value. The weight value will be calculated using the Analytical Hierarchy Process. The CLV results will be used to determine marketing strategies based on the concept of Customer Relationship Management on the right customer segmentation results to develop company services in the future.

# Research Methodology

This chapter presents the methodology used to conduct this study. First, researcher philosophy, assumptions, and research design are discussed. Second, we discuss the design science research approach used in this study. Lastly, we explain the research framework used to conduct this study, which consists of data collection, data preparation, choice of variables, clustering model, marketing strategy.

## Research Philosophical Position

Research philosophy is a system of beliefs and assumptions about the development of knowledge, which includes assumptions about human knowledge (epistemological assumptions), the nature of reality in the research (ontological assumptions) and the role of values and ethics within the research process (axiological assumptions) (Burrel &Morgan,1979), These assumptions constitute a credible research philosophy, which will influence the research topic’s comprehension, methodological choice, research strategy and data collection techniques, analysis procedures and findings interpretation (Saunders et al., n.d.)

There are five research philosophies in business and management research: positivism, critical realism, interpretivism, postmodernism and pragmatism (Saunders et al.,2016). Positivism research is value-free research working with an observable social reality that is typically deductive and uses quantitative analysis method to provide explanations. Unlike positivism, interpretivism is value-bond research that is typically inductive and uses qualitative methods to narrate and interpret new understanding. Critical realism is value-laden research that is more concerned research that is more concerned with investigations of anomalies, silences, and absences. Pragmatism research is value-driven research that more concerned with practical consequences and problem-solving.

The business phenomenon of this study is Understanding electricity consumption patterns is essential to effectively manage the increasing demand for electricity. This study starts with a problem and aims to contribute practical solutions that inform future practices for decision makers. The research method will be determined based on the problem and research questions that emphasize practical solutions and outcomes. Therefore, based on its philosophical position, this study is categorized under the research philosophy of pragmatism. This philosophy is achieved by analyzing theories, concepts, and research findings not in the abstract but regarding instruments of thinking and acting with their practical implications in specific contexts (Saunders et al., 2016). Therefore, the practical solutions and outcomes of this study provide interesting research contributions.

Table III.1 Pragmatism philosophy assumptions (Saunders et al., 2016)

|  |  |
| --- | --- |
| Assumption | Description |
| Ontology | * Complex, rich, external * Reality as the practical consequences of ideas * The flow of processes, experiences, and practices |
| Epistemology | * The practical meaning of knowledge * Theories and knowledge enable successful action * Focus on problems and practices * Problem-solving and future practices as the contributions |
| Axiology | * Value-driven research * Research initiated and sustained by the researcher’s doubts and beliefs * Researcher’s reflexive |
| Typical Method | * Following researcher problems and questions * Range of method mixed, multiple, qualitative, quantitative, action research * Emphasis on practical solutions and outcomes |

To better understand this research approach and design, Figure 3.1 illustrates the research onion of underlying research philosophy, theory development approach, methodological choices, research strategy, time horizon, and data collection. The two outer layers of research philosophy and theory development approach will influence how the research questions are answered (Saunders et al., 2016). As a pragmatism study, this research uses deductive inference to evaluate propositions considering the existing literature. This research designs a research strategy to test the proposed predictive analytics model (developed based on the read academic literature) in the specific segmentation of electricity consumption customers in West Sumatra, Indonesia.

Diagram, venn diagram

Description automatically generated

Figure III.1 Research onion (Saunders et al., 2016)

The subsequent three layers (i.e., methodological choice, research strategy, and time horizon) focus on the research design. In terms of methodological choices, this research is a quantitative study. We used quantitative data collection techniques followed by quantitative analysis procedures. In the analysis procedure, we also use quantitative methods in which the dataset used (transaction data of PLN West Sumatra Zone customers) is analysed.

This research is a single case study that focuses on a topic or phenomenon in a real-life setting (Understanding electricity consumption patterns is essential to effectively manage increasing electricity demand). The research was conducted over a longitudinal time horizon, involving time series data from January 2019 to December 2020. Finally, we used quantitative data collection to gather secondary data from PT PLN Persero West Sumatra Zone.

## Research Framework

This research adopted and modified a new method in the research paradigm of information systems, that is, design science research method. In contrast to the behavioural science paradigm, which seeks to establish and validate notions of human or organizational behavior, the design science paradigm aims to produce creative information technology artifacts that can solve organizational problems (Hevner et al., 2004). Fundamental to design science research is the process of problem-solving. Therefore, design science research must be tailored to local conditions in order to provide solutions and valuable results for solving problems. Knowledge of design challenges and their solutions is learned through the conception and implementation of artifacts (Hevner et al., 2004). Finally, this approach includes design evaluations to convincingly demonstrate that the research contributes to the applied context in a practical way (Gregor & Hevner, 2013).

This study is a design science study focused on machine learning-based predictive analytics. According to (Shmueli & Koppius, 2011), predictive analytics in quantitative empirical modeling refers to developing and testing models to make empirical predictions. Thus, it consists of two components: (1) an empirical forecasting model developed to predict new observations, and (2) an assessment of predictive power, measured in terms of the accuracy of out-of-sample forecasts. Compared to interpretive statistical modelling, which focuses on minimizing model bias, predictive analytics minimizes combined bias and variance (Shmueli & Koppius, 2011). This study adopted and modified the research framework of Shmueli & Koppius (2011) to develop machine learning-based predictive analytics.

The research framework of this study consists of five main steps, namely (1) data collection, (2) data preparation, (3) choice of variables, (4) Clustering model, (5) marketing strategy definition, as presented in Figure III.2. First, we collected the data from PT. PLN Persero West Sumatera Zone from January 2019 to December 2020. In the second step, we perform data profiling followed by data cleaning to eliminate data duplication and missing data. The third step is the variable selection phase to get the right variables in the clustering model phase followed by determining the marketing strategy.

Graphical user interface, diagram, application, Word

Description automatically generated

Figure III.2 The research framework

In the Clustering Model, we use 3 variables: power capacity, peak load consumption, and non-peak load consumption, we use these three predictor variables to predict customer segmentation (See Table III.5) using K-Means followed by cluster validation(S. M. S. Hosseini et al., 2010a).

The cluster results will be used to calculate customer lifetime value (CLV) followed by finding the weight values using Analytical Hierarchy Process. Finally, the CLV results are used as a proposal for implementing strategies on customer characteristics using the concept of customer relationship management strategies (Grbovic et al., n.d.).

## Data Collection

This study focuses on the segmentation of business customers of PT PLN Persero West Sumatra zone. Table III.2 summarizes the data sources used in this study. First, we collected PLN customer transaction data from January 2019 to December 2020. Customer transaction data in 2019 consists of 7,945,689 customers and has 107 variables then in 2020 it consists of 8,558,539 and has the same number and variable names as the previous year. Finally, all the data obtained when calculated consists of 16,504,228 customers and 107 data variables. In this study, we used data from July 2020 because customers and the amount of power used were higher in that month.

Table III.2 Descriptive of Data Collection

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

## Data Preparation

This phase consists of data profiling and data cleaning. In our first step of data profiling, we focus the data on regions and customer types. This section presents the data focus that will be selected based on the data analysis that will be conducted. The study starts by looking at the regions in West Sumatra that use the highest electricity. Figure III.3 presents based on the results of the plot analysis that has been carried out in 4 service center areas of PT PLN Persero, Padang area has the highest electricity consumption compared to other locations.

Figure III.3 Total Electricity Consumption Bases on Region

Further analysis looks at the potential customers who use a higher total kWh. Figure III.4 presents the results of the plot analysis based on total electricity consumption by customer category. Based on regulations issued by the Indonesian government [49], customers are divided into five categories, namely household, social, government, business, and industry. Based on the results of the plot analysis, business customers have the highest electricity usage of around 37%, followed by industrial customers as much as 31% and other customers use electricity consumption below 15%. Therefore, this study focuses on business customers because they use higher electricity consumption than others and can increase company revenue.

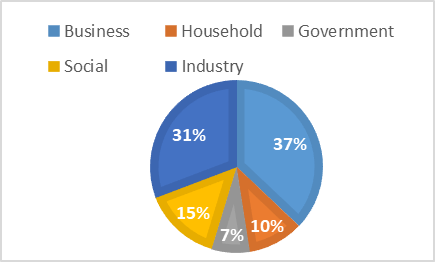


Figure III.4 Total electricity consumption based on customer energy

The second step is data cleaning, we perform duplicate removal and missing value handling. This analysis is used to handle duplicate data rows or missing data rows. Data cleaning aims to find potential predictors in the dataset. Finally, Table III.3 shows the results of the data focus analysis and data cleaning obtained 13 variables with 508,934 from the results of data profiling and data cleaning. The data will be used for model development.

Table III.3 Descriptive 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 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 |

## Choice of Variable

This section describes the predictor variables used in the clustering model. Among the 13 variables in Table III.3, the selected variable type is Integer or Double, because the focus of the process in the clustering model is to segment customers by power according to the peak load used by customers and the external power supply to predict the future peak load. However, The variable ID\_Customer is not included in the predictors because it is not required for the clustering model. This study will estimate the peak load from 6:00 am to 4:59 pm and the external peak load from 5:00 pm to 5:59 am (Katadata, 2020). Based on this interpretation, the variables kWh external load and kWh peak load are used as predictions in the clustering model. Table III.4 shows the 9 variables available for clustering models.

Table III.4 Descriptive of Potential 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 |

## Clustering Model

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

Commonly, K-means is one of the well-known unsupervised learning techniques for cluster analysis (Bapna et al., 2004). Cluster analysis is used to aggregate or divide the data set into several clusters according to the similarity value. The situation in this model is used because this algorithm has simplicity and ease of use, and users can determine the number of clusters themselves. This number of clusters (k) needs to be determined by validation (A. Hosseini & Hosseini, 2020).

Validation in this study uses the elbow method. The Elbow method in previous studies (Celik, 2009; Kim & Lee, 2015; Z. J. Lee et al., 2021) 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 (Gustriansyah et al., 2019) 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 (Marisa et al., 2019).

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.

## Marketing Strategy Definition

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

Where:

*X = variables values from cluster results*

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

*W = weight of each value of cluster result*

*I = start of the variable*

*J = start of the weight*

The weight value is obtained using calculations from the Analytical Hierarchy Process (AHP) (Al & Al-Harbi, n.d.). AHP solves complex multi-criteria problems into a hierarchy(Parvaneh et al., n.d.). It is helpful for integrated and fuzzy issues based on human brain assessment (Agustine et al., 2021). The step from AHP is described below (Al & Al-Harbi, n.d.; Parvaneh et al., n.d.):

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 formula

Where:

CI = Consistency index

= the eigenvalue of the predetermined variable value

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

Based on the results of CLV, then we can determine the targeting that aims to develop customer service improvement strategies based on the concept of customer relationship CRM (S. M. S. Hosseini et al., 2010b) which is described in table III.5.

Table III.5 Customer Relation 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 |

# Results and Analysis

In this chapter, the results and analysis of this study are presented. This study was conducted according to the research framework presented in the previous chapter (see Figure III.2). Therefore, Therefore, this chapter will present the data processing as described in sections III.6 and III.7. The last section presents the strategy that matches the company's criteria in each segment.

## Result of Clustering Model

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 IV.1, 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 dark blue.

Table IV.1 The Combination of Clustering Variables

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

*Description: 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*

The second step, after getting the variables that were processed previously then continued with determining the number of clustering using the elbow method to get the best number of clusters (k). Figure 4 shows the number of groups based on the results of the predictor variables previously described using the Elbow method. 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

Chart, line chart, scatter chart

Description automatically generated

Figure IV.1 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 IV.2 and Figure IV.3 can be seen below.

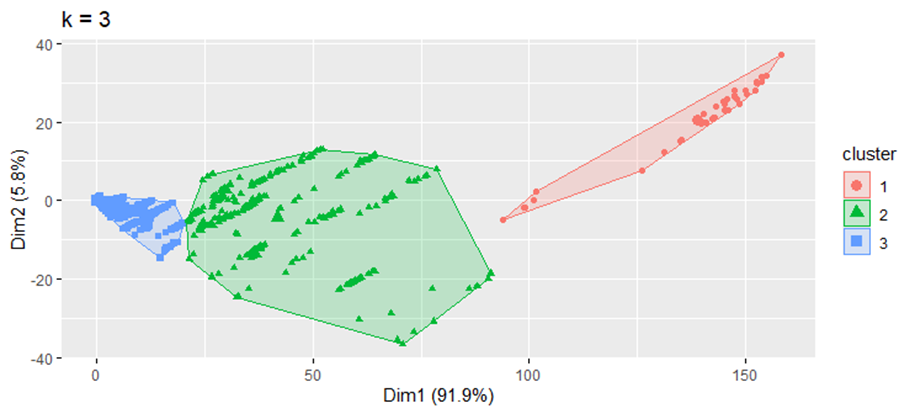


Figure IV.2 Cluster Visualization (k=3)

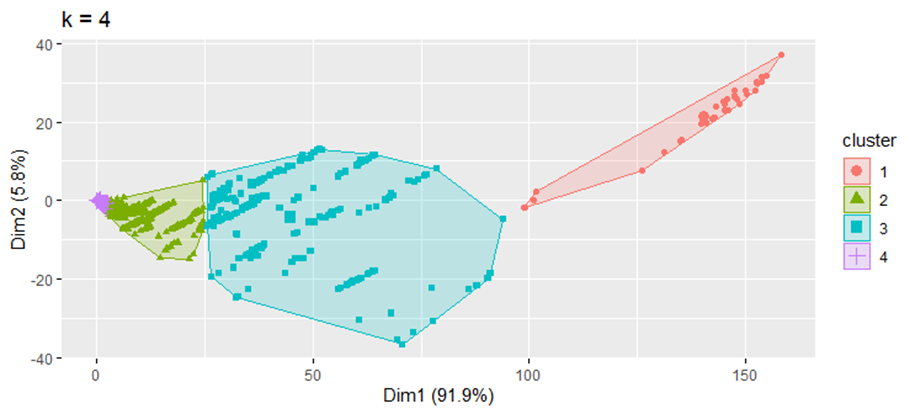


Figure IV.3 Cluster Visualization (k=4)

Based on the results of clustering using K-Means clustering, Table IV.2 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 IV.2 The Detail of The Clustering Results

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **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 |

## Result Customer Lifetime Value (CLV)

The first 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 [53], [54] 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 IV.3 shows the weight value of each variable from the AHP calculation.

Table IV.3 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 IV.4 presents the average CLV estimated for each.

Table IV.4 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 IV.5 presents device assignments in customer segmentation.

Table IV.5 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 |

## Implement Customer Relationship Management Strategies

In this section, we will explain the insights from the development of customer segmentation in each cluster that assesses the goal of developing the proposed customer service improvement strategy 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 shown in Table IV.6.

Table IV.6 Insight from CRM Decision Development

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

Based on the analysis, the combination of variables with the highest total value of variant by entering power, peak load, and peak non load. In determining the number of clusters, the number of clusters 3 and 4 have the same number, but in the visualization of cluster 3, the distribution is more precise. The combination of K-Means and CLV results found three different customer segments. Segment 1 has 282 business customers with a total capacity of 938,837 kWh, peak load usage of 27,827 kWh, and non-peak load of 115,194. In part two, there are 508,615 business customers with a total power of 4,260 kWh, then a peak load of 35 kWh and a non-peak load of 544. In segment 3, there are 37 business customers with a total power of 2,226,351 kWh, a peak load of 123,297 kWh, and non-peak load of 390.803.

In determining the marketing strategy, recall that we discussed two marketing strategies in the literature review, namely (1) sustainable marketing and (2) one-on-one marketing. The third group is profitable customers, the right strategy in this group is sustainable marketing, namely business to business, by offering premium service products specifically to use more electricity during non-peak periods. Simultaneously, one can operationalize one-to-one marketing by providing special account executives to customers to provide the best solutions and consultations for electrical problems experienced by customers. The second group is the profitable intermediate customers. We propose sustainable business for business marketing. By offering premium services without the need to abandon consumer consumption habits during peak seasons. Another way, one-to-one marketing campaigns to customers to increase electricity use during the non-peak load period.

The first group is less profitable customers because the monthly electricity consumption is 4,260 kWh. Therefore, we propose a Continuous Replenishment Program. For this type of customer, the company is advised to conduct 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, and car or motorcycle prizes. Other forms of partnership with electronic equipment manufacturers to replace non-electrical equipment with electricity-based ones such as electric stoves, sewing machines, electric vehicles, etc.) can also be offered.

Based on this research, the focus is on customer segmentation so that the marketing strategy implemented can maximize the use of electricity provided by the company. Previous research only focused on classifying customers based on patterns and types of electricity use, electricity demand, and the use of the K-Means grouping model (Jang et al., 2021; E. Lee et al., 2020; McLoughlin et al., 2015; Toussaint & Moodley, 2020).

# Summary And Conclusion

## Conclusion

This research presents an application of hybrid segmentation of electricity consumption customers specifically in Padang region in Indonesia. Two main research questions were formulated. To answer the research questions, we conducted a predictive analytics study of customer segmentation to develop and implement an appropriate marketing strategy using hybrid customer segmentation.

***RQ1***: “*How to develop an accurate customer segmentation model according to the characteristics of electricity customers using West Sumatra Zone business customer transaction data*?”

This study adopted and modify the predictive analytics framework developed by Shmueli & Koppius (2011) to construct the predict segmentation models. The research framework consists of five stages: (1) data collection, (2) data preparation, (3) choice of variables, (4) Clustering model, (5) marketing strategy definition.

1. First, we collected data from PT PLN Persero West Sumatera Zone from January 2019 until December 2020.
2. Second, we performed data preparation (i.e., data profiling and data cleaning). Data profiling is used to focus on the field area and electricity consumption business customers. Data cleaning is used to remove duplicate data and missing data is removed.
3. Third, we select from the 13 variables obtained from the data preparation then the variables will be selected based on the predictor and predicted needed to be processed in the customer segmentation hybrid model. Finally, there are 9 variables that are used to be processed in the next process.
4. Fourth, 9 variables will be selected which are possible before being processed in the Clustering Model. There are 3 variables processed in the cluster model, the results are 3 segments from the processing
5. Finally, the segment results will be calculated into Customer Lifetime Value but previously the weight value will be sought from the AHP value of the 3 cluster result variables, then there is an appropriate segment ranking from the processing results.

***RQ2***: “*How to implement marketing strategies according to customer criteria based on the results of the customer segmentation model?*”

Based on the analysis, the combination of variables with the highest total value of variant by entering power, peak load, peak non load. In determining the number of clusters, the number of clusters 3 and 4 has the same number, but in the visualization of cluster 3 the distribution is more precise. The results of the combination of K-Means and CLV found three different customer segments. Segment 1 has 282 business customers with a total capacity of 938,837 kWh, peak load usage 27,827 kWh, and non-peak load 115,194. In part two, there are 508,615 business customers with a total power of 4,260 kWh, then a peak load of 35 kWh and a non-peak load of 544. In segment 3, there are 37 business customers with a total power of 2,226,351 kWh, then a peak load of 123,297 kWh and non-peak load of 390.803.

The third group is profitable customers, the right strategy in this group is sustainable marketing, namely business to business, by offering premium service products specifically to use more electricity during non-peak periods. Simultaneously, one can operationalize one-to-one marketing by providing special account executives to customers to provide the best solutions and consultations for electrical problems experienced by customers. The second group is the profitable intermediate customers. We propose sustainable business for business marketing. By offering premium services without the need to abandon consumer consumption habits during peak seasons. Another way, one-to-one marketing campaigns to customers to increase electricity use during the non-peak load period.

The first group is customers who are less profitable because the total electricity consumption per month is 4,260 kWh. Therefore, we propose a Continuous Replenishment Program. For this type of customer, the company is advised to conduct 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 prizes. Other forms of partnership with electronic equipment manufacturers to replace non-electrical equipment with electricity-based ones such as electric stoves, electric sewing machines, electric vehicles, etc.) can also be offered.

## Research and Practical Implications

Several contributions can be expected from this research. These contributions can be seen from three perspectives:

1. In term 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.
2. In terms of managerial implications, this finding can inform companies to provide more optimal power based on the characteristics of their customers.
3. In addition, this research help companies improve their targeting strategy for their customer and the corresponding revenue.

## Limitation and Further Research

Not without limitations, this study opens opportunities for further research. First, this research focuses on the field area and business customers only due to limitations in processing data, so it is not possible to use all the data obtained. In the future, it can investigate all regions and all types of customers to be more accurate in calculations.

Second, the research only uses 3 variables and combines the K-Means Clustering model, Customer Lifetime Value followed by the Customer Relationship Management Strategy Concept. In the future, it can investigate using the 13 variables that have been found and then can use other machine learning clustering models.

# REFERENCES

Abdi, F., & Abolmakarem, S. (2019). Customer Behavior Mining Framework (CBMF) using clustering and classification techniques. *Journal of Industrial Engineering International*, *15*, 1–18. https://doi.org/10.1007/s40092-018-0285-3

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

Agustine, P., Parung, H., Davey, P., & Frid, C. (2021). Management Strategies to Protect Coastal Areas from Oil-Polluted Seawater (A Case Study of Coastal Areas in Bekasi Regency). *IOP Conference Series: Earth and Environmental Science*, *921*(1). https://doi.org/10.1088/1755-1315/921/1/012049

Al, K. M., & Al-Harbi, -Subhi. (n.d.). *Application of the AHP in project management*. www.elsevier.com/locate/ijproman

Azadeh, A., & Faiz, Z. S. (2011). A meta-heuristic framework for forecasting household electricity consumption. *Applied Soft Computing Journal*, *11*(1), 614–620. https://doi.org/10.1016/j.asoc.2009.12.021

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*(12). 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*(1). https://doi.org/10.1186/s13731-021-00179-z

Bapna, R., Goes, P., Gupta, A., & Jin, Y. (2004). User Heterogeneity and Its Impact on Electronic Auction Market Design: An Empirical Exploration. In *Source: MIS Quarterly* (Vol. 28, Issue 1). http://www.jstor.orgStableURL:http://www.jstor.org/stable/25148623

Borisavljević, K., & Radosavljević, G. (2021). Application of logistics model in analysing relationship marketing in travel agencies. *Zbornik Radova Ekonomskog Fakultet Au Rijeci*, *39*(1), 87–112. https://doi.org/10.18045/zbefri.2021.1.87

Camero, A., Luque, G., Bravo, Y., & Alba, E. (2018). Customer segmentation based on the electricity demand signature: The andalusian case. *Energies*, *11*(7). https://doi.org/10.3390/en11071788

Celik, T. (2009). Unsupervised change detection in satellite images using principal component analysis and κ-means clustering. *IEEE Geoscience and Remote Sensing Letters*, *6*(4), 772–776. https://doi.org/10.1109/LGRS.2009.2025059

Chen, D. Q., Preston, D. S., & Swink, M. (2015). How the use of big data analytics affects value creation in supply chain management. *Journal of Management Information Systems*, *32*(4), 4–39. https://doi.org/10.1080/07421222.2015.1138364

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*(6), 1312–1321. https://doi.org/10.13189/ujaf.2021.090610

Dias, F. M., de Oliveira, M. P. V., Filho, H. Z., & Rodrigues, A. L. (2021). Analytical guidance or intuition? what guides management decisions on the most important customer value attributes in the supermarket retail? *Revista Brasileira de Marketing*, *20*(2), 385–414. https://doi.org/10.5585/REMARK.V20I2.16106

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*, *99*, 273–308. https://doi.org/10.7203/CIRIEC-E.99.14602

Gajowniczek, K., & Zabkowski, T. (2018). Simulation Study on Clustering Approaches for Short-Term Electricity Forecasting. *Complexity*, *2018*. https://doi.org/10.1155/2018/3683969

Gil-Quintana, J., & Vida de León, E. (2021). Educational influencers on instagram: Analysis of educational channels, audiences, and economic performance. *Publications*, *9*(4). https://doi.org/10.3390/publications9040043

Grbovic, M., Djuric, N., & Vucetic, S. (n.d.). *Supervised Clustering of Label Ranking Data*. https://epubs.siam.org/page/terms

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*(1), 470–477. https://doi.org/10.11591/ijeecs.v18.i1.pp470-477

Hosseini, A., & Hosseini, R. (2020). *Model selection for count timeseries with applications in forecasting number of trips in bike-sharing systems and its volatility*. http://arxiv.org/abs/2011.08389

Hosseini, R., Yang, K., Chen, A., & Patra, S. (2021). *A flexible forecasting model for production systems*. http://arxiv.org/abs/2105.01098

Hosseini, S. M. S., Maleki, A., & Gholamian, M. R. (2010a). Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. *Expert Systems with Applications*, *37*(7), 5259–5264. https://doi.org/10.1016/j.eswa.2009.12.070

Hosseini, S. M. S., Maleki, A., & Gholamian, M. R. (2010b). Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. *Expert Systems with Applications*, *37*(7), 5259–5264. https://doi.org/10.1016/j.eswa.2009.12.070

Hyland, M., Leahy, E., & Tol, R. S. J. (2013). The potential for segmentation of the retail market for electricity in Ireland. *Energy Policy*, *61*, 349–359. https://doi.org/10.1016/j.enpol.2013.05.052

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*(19). https://doi.org/10.3390/en14196130

Kafkas, K., Perdahçı, Z. N., & Aydın, M. N. (2021). 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*, *16*(7), 2965–2980. https://doi.org/10.3390/jtaer16070162

Katadata. (2020, January 9). *National Electricity Consumption Continues to Increase*. Www.Databook.Com. https://databoks.katadata.co.id/datapublish/2020/01/10/konsumsi-listrik-nasional-terus-meningkat

Khajvand, M., Zolfaghar, K., Ashoori, S., & Alizadeh, S. (2011). Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study. *Procedia Computer Science*, *3*, 57–63. https://doi.org/10.1016/j.procs.2010.12.011

Kim, K. Y., & Lee, B. G. (2015). Marketing insights for mobile advertising and consumer segmentation in the cloud era: A Q-R hybrid methodology and practices. *Technological Forecasting and Social Change*, *91*, 78–92. https://doi.org/10.1016/j.techfore.2014.01.011

Koponen, J., Julkunen, S., Gabrielsson, M., & Pullins, E. B. (2021). An intercultural, interpersonal relationship development framework. *International Marketing Review*, *38*(6), 1189–1216. https://doi.org/10.1108/IMR-11-2019-0267

Kulej-Dudek, E. (2021). Ecolabnet service packages as a response to the needs of manufacturing enterprises in the SME sector of the Baltic Sea Region. *Production Engineering Archives*, *27*(4), 265–271. https://doi.org/10.30657/pea.2021.27.35

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*(3). 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*(9). https://doi.org/10.3390/sym13091557

Li, H., Yang, X., Xia, Y., Zheng, L., Yang, G., & Lv, P. (2018). K-LRFMD: Method of Customer Value Segmentation in Shared Transportation Filed Based on Improved K-means Algorithm. *Journal of Physics: Conference Series*, *1060*(1). https://doi.org/10.1088/1742-6596/1060/1/012012

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*(3), 169–180. https://doi.org/10.30880/ijie.2019.11.03.018

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. https://doi.org/10.1016/j.apenergy.2014.12.039

Park, P., Kim, D., Lee, S., & Whang, J. (2018a). Toward an economically sustainable casino industry: A development of customer value indicators using an analytic hierarchy process. *Sustainability (Switzerland)*, *10*(11). https://doi.org/10.3390/su10114255

Park, P., Kim, D., Lee, S., & Whang, J. (2018b). Toward an economically sustainable casino industry: A development of customer value indicators using an analytic hierarchy process. *Sustainability (Switzerland)*, *10*(11). https://doi.org/10.3390/su10114255

Parvaneh, A., Abbasimehr, H., & Tarokh, M. J. (n.d.). *Integrating AHP and Data Mining for Effective Retailer Segmentation Based on Retailer Lifetime Value*.

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

Saunders, M., Lewis, P., Thornhill, A., Lewis, S. •, & Thornhill, •. (n.d.). *Research methods for business students fi fth edition*. www.pearsoned.co.uk

Sekizaki, S., Nishizaki, I., & Hayashida, T. (2016). 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)*, *194*(4), 27–41. https://doi.org/10.1002/eej.22743

Shmueli, G., & Koppius, O. R. (2011). Predictive Analytics in Information Systems. In *Source: MIS Quarterly* (Vol. 35, Issue 3).

Toussaint, W., & Moodley, D. (2020). Clustering Residential Electricity Consumption Data to Create Archetypes that Capture Household Behaviour in South Africa. *South African Computer Journal*, *32*(2), 1–34. https://doi.org/10.18489/SACJ.V32I2.845

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*(4), 2545–2565. https://doi.org/10.1051/ro/2021111

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*(6), 3393–3415. https://doi.org/10.3934/jimo.2020125

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. https://doi.org/10.1145/3450588.3450596

Yudhya, T. B. (2019). Retail store image: A study of the matahari department store (at Bandung Indonesia). *Humanities and Social Sciences Reviews*, *7*(5), 98–102. https://doi.org/10.18510/hssr.2019.7513

# APPENDICES