SEGMENTATION USING CUSTOMER LIFETIME VALUE HYBRID K-MEANS AND ANALYTIC HIERARCHY 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 HIERARCHY 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 160,205 business customers with a total capacity of 911,000 kWh, peak load usage of 365,000 kWh, and non-peak load of 546,000. In segment 2, there are 200,123 customers with a total capacity of 435,400 kWh, a peak load usage of 184,864 kWh, and a non-peak load of 250,536. In segment 3, there are 148,287 business customers with a total power of 1,250,000 kWh, then a peak load of 489,655 kWh and a non-peak load of 760,345. 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 BERBASIS CUSTOMER LIFETIME VALUE DENGAN TEKNIK HYBRID K-MEANS DAN ANALYTIC HIERARCHY PROCESS

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 terus meningkat secara efektif. Studi ini menyajikan pendekatan hibrida untuk segmentasi pelanggan dengan menggabungkan pengelompokan K-Means, konsep nilai masa pakai pelanggan, dan proses hirarki 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 analitik. Segmen 1 memiliki 160.205 pelanggan bisnis dengan total kapasitas 911.000 kWh, penggunaan beban puncak 365.000 kWh, dan beban non-puncak 546.000. Pada segmen 2, terdapat 200.123 pelanggan dengan total kapasitas 435.400 kWh, penggunaan beban puncak 184.864 kWh, dan beban non-puncak 250.536. Pada segmen 3, terdapat 148.000 pelanggan bisnis dengan total kapasitas 435.400 kWh, penggunaan beban puncak 184.864 kWh, dan beban non-puncak 250.536. Pada segmen 3, terdapat 148.287 pelanggan bisnis dengan total daya sebesar 1.250.000 kWh, kemudian beban puncak sebesar 489.655 kWh dan beban non puncak sebesar 760.345. Strategi yang akan diambil berdasarkan segmentasi ketiga 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 menengah, kami mengusulkan pendekatan bisnis premium ke bisnis yang dapat mengakomodasi peningkatan konsumsi energi mereka tanpa penggunaan listrik yang berlebihan selama periode puncak. Pendekatan ini akan didukung oleh akun eksekutif khusus untuk pelanggan ini.

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

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

by
Radit Rahmadhan
Student ID: 29020003
Master Program of Science in Management

Institut Teknologi Bandung

Approved

12 December 2022

Advisor,

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.

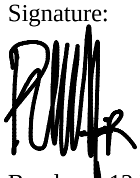
Reproduction or publication of parts or whole of the thesis must be under the consent of the Head of Graduate School, Bandung Institute of Technology.

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, 12 December 2022

Radit Rahmadhan

ACKNOWLEDGMENTS

First of all, there are no words that deserve to be said other than sincere gratitude to ALLAH SWT who has given me health and strength so that I can finally complete this master's thesis on time.

Then, I would like to express my deepest gratitude and respect to my thesis and research advisor, Meditya Wasesa. He inspired me to be the best and independent researcher and made me realize to optimize my capacity in research. It is one of my honors throughout my study journey at Institut Teknologi Bandung to have Mr. Meditya as my advisor.

Thanks to Mrs. Yuliani Dwi Lestari, Ph.D, director of MSM/DSM program, Mrs. Anti Novarianti, and Mrs. Anjar Listyorini, who have helped my study administration and many things to support my research and thesis with good and excellent service, special thanks to all MSM lecturers generally SBM-ITB lecturers for their patience and sincerity to help us (MSM students) improve our research and critical thinking skills.

My deepest and sincere gratitude to the people I love the most in the world and for the rest of my life, my parents and cousins. The main source of inspiration who always provides prayers, motivation, energy, and material support. Thank you for putting education in such a high position for me to achieve. I believe this thesis would never have been completed without the presence of my significant others.

I cannot forget my friends in MSM who went through difficult times together during this study. Especially, Dinda Mas Agus (MSM 2019), and Abdillah who have helped my thesis project and always discuss and solve problems together,

Mr. Mas Hendra aka BOS MSM 2020, Igun and David supervision teammates Mr. Medit and the PCM team (Liga, Kibil, Afif, Rafi Ndari, Rafi Fathur), Bunda Intan, Bunda Nadya, Taffy, Imas, Irsa, Tiwi and other MSM- General 2020 friends who have been kind and fun during online and offline discussions.

Finally, I would like to thank all those who supported my study journey and thesis completion that I cannot mention one by one. May Allah SWT repay your kindness.

Bandung, 9 December,2022

Radit Rahmadhan

TABLE OF CONTENT

LIST OF APPENDICES

LIST OF FIGURES

LIST OF TABLES

Chapter I Introduction

I.1 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).

Customer segmentation is a way for a company to group its customers based on common characteristics or needs. This can be very useful for a monopoly like PLN (Perusahaan Listrik Negara) because it can help the company improve its efficiency and increase customer satisfaction (Asikin & Suhartana, n.d.). By understanding the specific needs of different customer groups, PLN can tailor its services to better meet those needs. For example, the company can differentiate between customers who use peak and non-peak electricity and provide electricity accordingly(Alpay et al., n.d.; Pardede et al., 2019). According to the Manager of PLN West Sumatra, 80% of the electricity used by customers is being utilized, while the remaining 20% is not being used due to a high number of non-peak loads.

Commented [RRS1]: Added the Statement Monopoly Company action from "Give monopoly some people said that is the monopoly. A company did not consider, not quite, not, consider about how they can a segmented their custom model getting up "input from Pak Manahan

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.

I.2 Research Questions

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 in monopoly company 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?

I.3 Research Objectives

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.

I.4 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 modified 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.

I.5 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.

I.6 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.

Chapter II 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.

II.1 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.

II.2 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 behaviour 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.

II.3 Marketing Strategy of Monopoly Company 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 and Middle Profitable Customer (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 and Middle Profitable Customer (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).

II.3.1 Marketing Strategy in Monopoly Company

The marketing strategy of a monopoly company in customer relationship management (CRM) may involve several different approaches(Asikin & Suhartana, n.d.). Some potential strategies could include:

1. Personalized marketing: This involves tailoring marketing efforts to the specific needs and preferences of individual customers. This could include

things like targeted email campaigns, personalized offers, or customized product recommendations.

2. Customer loyalty programs: These programs are designed to encourage customers to continue doing business with the company by offering rewards or incentives for their loyalty. This could include discounts, free products or services, or exclusive access to certain events or promotions.

3. Customer service: Providing excellent customer service is a key part of any CRM strategy. This could involve things like easy-to-use self-service options, responsive social media support, or a dedicated customer service team.

4. Data-driven marketing: Monopoly companies may have access to a large amount of data about their customers. This data can be used to create more targeted and effective marketing efforts. For example, companies may use data analytics to identify trends or patterns in customer behaviour, and then use this information to create more personalized marketing campaigns.

For a monopoly like PLN (Perusahaan Listrik Negara), customer segmentation is very important because it can help the company improve operating efficiency and increase customer satisfaction(Alpay et al., n.d.; Asikin & Suhartana, n.d.; Pardede et al., 2019). Here are some narratives about the importance of customer segmentation in the monopoly company PLN:

1. Knowing customer needs: By categorizing customers based on similar characteristics, PLN can find out more specific customer needs. For example, PLN can differentiate between household customers and industrial customers, so that it can provide tariffs that suit the electricity consumption needs of each customer.

2. Improve operating efficiency: By knowing the specific needs of customers, PLN can manage resources more efficiently. For example, PLN can allocate resources to meet the higher needs of industrial customers, thereby improving the efficiency of the company's operations.

3. Improving customer satisfaction: By providing products and services that match customer needs, PLN can increase customer satisfaction. For example, PLN can provide cheaper electricity tariffs for household customers who have low electricity consumption, thereby increasing the satisfaction of these customers.
4. Increase customer loyalty: By providing products and services that meet customer needs, PLN can increase customer loyalty to the company. For example, industrial customers who are satisfied with the services they receive from PLN are more likely to continue using the company's services rather than switching to another company.

Commented [RRS2]: Added the Statement Monopoly Company action from "Give monopoly some people said that is the monopoly. A company did not consider, not quite, not, consider about how they can a segmented their custom model getting up" input from Pak Manahan

II.4 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	1	Electricity Load Profile

	themes		
		2	Electricity Consumption
		3	Electricity Demand
		4	Marketing Context
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
		5	Customer (behaviour, satisfaction)
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.

Table II.3 Research Positioning

	Business Context				Data				Features					Method						
Study	1	2	3	4	1	2	3	4	1	2	3	4	5	1	2	3	4	5	6	7
(McLoughlin et al., 2015)	√				√				√	√				√	√					
(Toussaint & Moodley, 2020)		√				√			√	√	√			√		√				
(Camero et al., 2018)			√				√		√	√	√			√	√					
(Jang et al., 2021)	√				√				√	√	√							√		
(E. Lee et al., 2020)	√						√		√			√		√		√	√			
(Gajowniczek & Zabkowski, 2018)		√					√		√	√	√							√		
(Bañales et al., 2021)			√				√				√	√			√					
(Afthoni et al., n.d.)		√					√		√	√	√			√						√
(Foncubierta-Rodríguez et al., 2020)				√				√					√	√					√	
(Ye, 2021)				√									√	√					√	
(Gil-Quintana & Vida de León, 2021)				√									√	√					√	
(Borisavljević & Radosavljević, 2021; Xie et al., 2021)				√									√	√					√	√

Chapter III Research Method

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.

III.1 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.

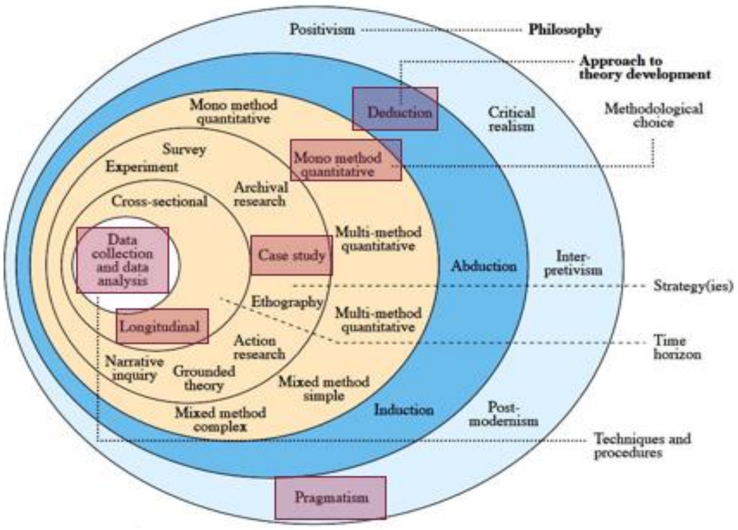
The business phenomenon of this study is Understanding electricity consumption patterns is essential to effectively manage the increasing demand for electricity.

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 (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.



Note(s) : Red Boxes indicate the research philosophy, approach to the theory development, methodological choice, research strategy, and time horizon used this study.

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.

III.2 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 modelling 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.

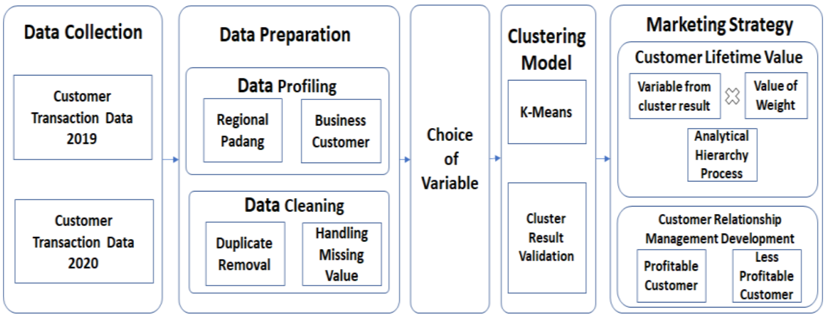


Figure III.2 The research framework

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.).

III.3 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 are higher in that month.

Table III.2 Data Description

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

III.4 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.

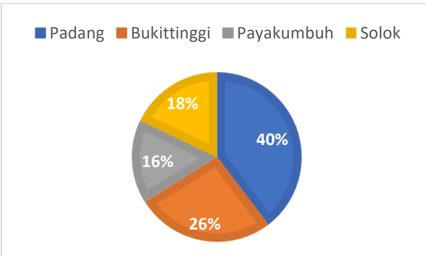


Figure III.3 Total Electricity Consumption Bases on Region

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.

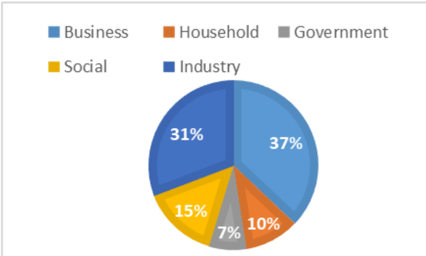


Figure III.4 Total electricity consumption based on customer type

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.

Table III.3 Data Cleaning Description

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

III.5 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 Predictive model specification

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

III.6 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 (Bapna et al., 2004).

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.

III.7 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:

$$CLV = X_i * W_j + \cdots + X_n * W_n$$

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

   $$CI = \frac{\lambda \max - n}{n - 1}$$

   Where:
   CI = Consistency index
   $\lambda\ maximum$ = the eigenvalue of the predetermined variable value
   $n = number\ of\ criteria$

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

In this study only applies 4 AHP steps, namely comparing variables, creating a set of pair-wise comparison matrices, and scoring each pair-wise comparison to determine the objectives (CLV weight value) and the last one making all pair-wise comparisons.

Commented [RRS3]: Revised the statement of using AHP in this thesis is based on Pak Manahan's "Using a combination of K-Means and AHP clearly and where AHP is used".

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

<table>
<tr><th>Customer Type</th><th>Sustainable Marketing</th><th>One To One Marketing</th></tr>
<tr><td>Profitable Customer</td><td rowspan="2">Business To Business</td><td rowspan="2">Customer Business Development</td></tr>
<tr><td>Middle Profitable Customer</td></tr>
<tr><td>Less Profitable Customer</td><td>Continuous Replenishment Program</td><td>Retail Account Marketing</td></tr>
</table>

Chapter IV 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.

IV.1 Results of K-Means Clustering

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 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.

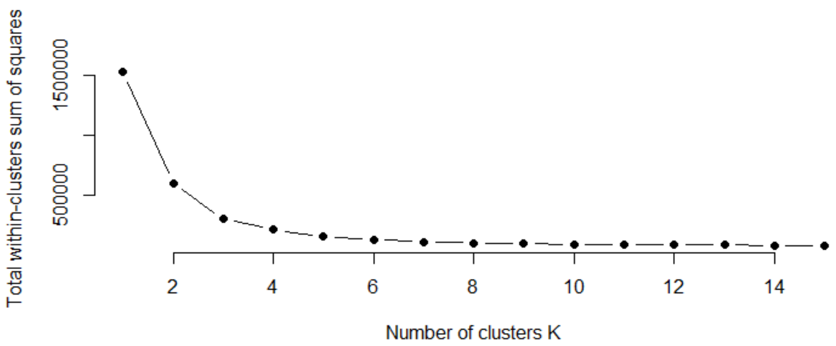


Figure IV.1 The Number of clusters

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.

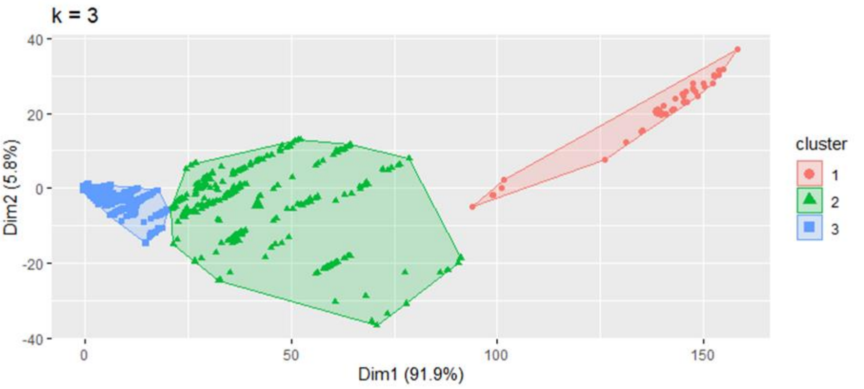


Figure IV.2 Cluster Visualization (k=3)

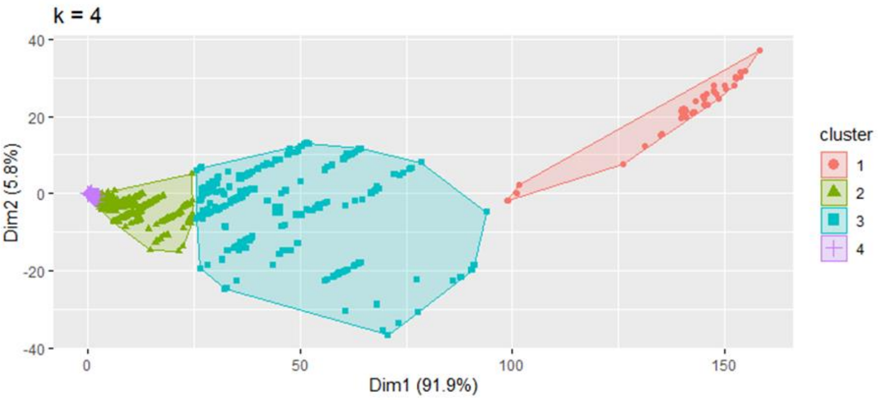


Figure IV.3 Cluster Visualization (k=4)

Based on the results of clustering using K-Means clustering, Table IV.2 present three different customer groups are finding. The first group represents 911,000 total powers used total electricity consumption at peak load of 365,000 kWh and total electricity consumption when peak off-load is 546,000 kWh with customers using installed capacity above 10,600 kWh.

The second group describes as much as 435,400 full powers used total electricity consumption at peak load of 184,864 kWh and total electricity consumption at peak load of 250,536 kWh with customers using installed capacity between 450 kWh to 10,600 kWh. The third group describes as much as 1,250,000 full powers used total electricity consumption at peak load of 489,655 kWh and total electricity consumption at peak load time of 760,345 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	160,205	911,000	546,000	365,000	11,000 -200,000
2	200,123	435,400	250,536	184,864	450- 10600
3	148,287	1,250,000	760,345	489,655	>200,000

Commented [RRS4]: Revised of corrections from Pak Fajar and Pak Manahan " First, the rules of segmentation, one of which is substantial. The segmentation result of 37 is not substantial and Second, the reason for segmentation is to make PLN more profitable. But PLN has 80% capacity, when the result is only 37 out of 500,000, it is less than 1% of the customers. So where is the other 79% that PLN has to sell?"

IV.2 Result of Analytical Hierarchy Process (AHP) and Customer Lifetime Value (CLV)

In determined 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. In this study, we used AHP to determine the value of the weight value of CLV then after the value was found we calculated the value of the cluster result variable in table IV.2.

IV.2.1 Result of Analytical Hierarchy Process (AHP)

In determined the CLV weight value, variables from the previous cluster results are needed, namely power, kWh Peak Off Load, kWh Peak Load. We will use these variables to calculate CLV. There are three layers in AHP: the goal, criteria, and alternative. Before entering the AHP steps, we will determine the layer. Table IV.3 shows the results of determining the layer of AHP.

Commented [RRS5]: Revised of using AHP in this thesis is based on Pak Manahan's "Using a combination of K-Means and AHP clearly and where AHP is used".

Table IV.3 Three Layer of AHP

Goal	Weight Value for CLV
Criteria	Variable (Power, Peak Off-Load, Peak Load)
Alternative	Made three different comparisons to find the value of the goal, such as (Power1, Peakload2, Peakoffload3).

Here are the AHP steps we used in this study: Step One, we will use 3 variables based on the cluster results. Table IV.4 shows variables used to compare variables.

Table IV.4 Compared Variables

Total Power (kWh)	KWH Peak Off Load (kWh)	KWH Peak Load (kWh)
911,000	546,000	365,000
435,400	250,536	184,864
1,250,000	760,345	489,655

Second step, create a set of pairwise comparison matrices for each lower level with one matrix for each element. In this research, 3 matrices were created as alternatives to determine the criteria. Green colour is the first criterion, blue colour is the second criterion, yellow colour is the third criterion.

Third step, the results of the matrix are needed for assessment in each pairwise comparison. The right value for comparison will be marked in red.Table IV.5 shows the results of the second step and the third step.

Table IV.5 Matrix from Pairwaise Comparasion

	C (P1, P2, P3)	C (POL1, POL2, POL3)	C (PL1, PL2, PL3)	C (P1, P2, P3)	C (POL1, POL2, POL3)	C (PL1, PL2, PL3)	C (P1, P2, P3)	C (POL1, POL2, POL3)	C (PL1, PL2, PL3)
C (P1, P2, P3)	1	-0,00001098	-0,00001094	1	-0,00007676	-0,00036749	1	-0,0003383	0,001931089
C (POL1, POL2, POL3)	911000	100000000	-0,00000478	2250000	100000000000	-0,000008789	4345	10000000000	0,0234546589
C (PL1, PL2, PL3)	91375	20892600	10000000000	214872	662500000000	-0,000088965	517865	34536354700	1

Description: P: Power, POL: Peak Off-Load, PL: Peak Load

To determine the weight of the criteria in this study, we will use the hierarchy synthesis and consider all the red-marked values in Table IV.5 from all the eigenvectors. The results of these calculations will be compared with the values in

Table IV.3. The final weight values for each variable from the 4-step AHP can be found in Table IV.6, which will be used to calculate the CLV.

Table IV.6 Weight of AHP Results

Variable	Weight
Power	0.235
kWh Peak Off-Load	0.680
kWh Peak Load	0.343

IV.2.2 Result of Customer Lifetime Value (CLV)

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.7 presents the average CLV estimated for each.

Table IV.7 Result of Customer Lifetime Value in Each Cluster

Centroid	Number of Customer	NP	NKPOL	NKPL	CLV Value
Segment 1	160,205	214,085	371,280	125,195	710,560
Segment 2	200,123	102,319	170,365	63,408	336,092
Segment 3	148,287	293,750	517,036	167,952	978,736

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 978,736, segment 1 receives the second rank because the value is equal 710,560 and segment 2 gets the third rank because the value is equal 336,092. Table IV.8 presents device assignments in customer segmentation.

Commented [RRS6]: Revised of corrections from Pak Fajar and Pak Manahan " First, the rules of segmentation, one of which is substantial. The segmentation result of 37 is not substantial and Second, the reason for segmentation is to make PLN more profitable. But PLN has 80% capacity, when the result is only 37 out of 500,000, it is less than 1% of the customers. So where is the other 79% that PLN has to sell?"

Table IV.8 Result of Customer Ranking

Segment	Number of Customers	CLV Value	Ranking
1	160,205	710,560	2
2	200,123	336,092	3
3	148,287	978,736	1

IV.3 The Implementation of 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.9.

Table IV.9 Insight from CRM Decision Development

Segment	Number of Customers	Ranking	Strategy Targeting
1	160,205	2	Middle Profitable Customer
2	200,123	3	Less-Profitable Customer
3	148,287	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 160,205 business customers with a total capacity of 911,000 kWh, peak load usage of 365,000 kWh, and non-peak load of 546,000. In part two, there are 200,123 business customers with a total power of 435,400 kWh, then a peak load of 184,864 kWh and a non-peak load of 250,536. In segment 3, there are 148,287 business customers with a total power of 1,250,000 kWh, a peak load of 489,655 kWh, and non-peak load of 760,345.

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 435,000 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).

Commented [RRS7]: Revision from pak fajar based on the comment "The segmentation model based on CLV is good, but the implications are not substantial enough for PLN to get greater profits".

Chapter V Conclusion

V.1 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 modified 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 is a combination of power, peak load, and non-peak load. When determining the number of clusters, both 3 and 4 clusters had the same number, but cluster 3 had a more precise distribution. The combination of K-Means and CLV results identified three different customer segments. Segment 1 consists of 160,205 business customers with a total capacity of 911,000 kWh, peak load usage of 365,000 kWh, and non-peak load of 546,000. Segment 2 consists of 200,123 business customers with a total power of 435,400 kWh, peak load of 184,864 kWh, and non-peak load of 250,536. Segment 3 consists of 148,287 business customers with a total power of 1,250,000 kWh, peak load of 489,655 kWh, and non-peak load of 760,345.

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 435,000 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.

V.2 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 such as PLN can allocate resources to meet the higher needs of industrial customers, to improve the efficiency of the company's operations.
3. In addition, this research help companies improve their targeting strategy for their customer and the corresponding revenue as PLN can differentiate based on the power used, so as to provide tariffs that match the electricity consumption needs of each customer. .

V.3 Limitations 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

Alpay, H., Makale, K., & Tarihi, B. (n.d.). *Türk Yönetim ve Ekonomi Araştırmaları Dergisi Örgüt Kültürü ve Hizmet Kalitesi İlişkisi: Sistematik Bir Tarama Organization Culture and Service Quality Relationship: A systematic search*. https://orcid.org/0000-0002-3256-982X

Asikin, Z., & Suhartana, Lw. (n.d.). *Comparative Study of Post-Marriage Nationality Of Women in Legal Systems of Different Countries International Journal of Multicultural and Multireligious Understanding Procurement of Goods and Services by PT PLN (Persero) Mataram Branch Anti-Monopoly Perspective and Unfair Business Competition*. https://doi.org/10.18415/ijmmu.v7i3.1526

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

Pardede, A. M. H., Andrayani, F. W., Nurdin, Abdullah, D., Erliana, C. I., Dana, R. D., Dikananda, A. R., Fasya, F., Rahaningsih, N., Andriani, M., Subekti, P., Elisa, E., Sohibun, Noor Hasan Siregar, M., Ponidi, Winarno, E., & Ketut Sudarsana, I. (2019). Application of Data Mining Prediction of Electricity Deviation Flow Using Metode Backpropogation at PLN Binjai Area. *Journal of Physics: Conference Series*, *1363*(1). https://doi.org/10.1088/1742-6596/1363/1/012067

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

Appendix A The Dataset

A.1. The Actual Data Set

NAMA_UPI	NAMA_AP	UNITUP	NAMA_UP	BLTH	IDPEL	PEMDA	NAMA	PNJ	NAMAPNJ	NOBANG	TARIP
UIW SUMATERA BARAT	UP3 PADANG	13103	ULP INDARUNG	201901	1,3103E+11	1	MUFFIDAH	JL	JL TNH SIRAH NAN XX	0	R1
UIW SUMATERA BARAT	UP3 PADANG	13103	ULP INDARUNG	201901	1,31031E+11	1	MANSUR	JL	Jl.AMPALU RT.001 RW.02 PANGGAM	0	R1
UIW SUMATERA BARAT	UP3 PADANG	13103	ULP INDARUNG	201901	1,31031E+11	1	ELMIDA	JL	Jl.RIMBO DATA RT.01 RW.01	0	R1
UIW SUMATERA BARAT	UP3 PADANG	13103	ULP INDARUNG	201901	1,31031E+11	1	NILVI RAHMI	KR	GRIYA PERMATA BIRU No.B/11	0	R1
UIW SUMATERA BARAT	UP3 PADANG	13103	ULP INDARUNG	201901	1,3103E+11	1	ZULMASRI	XX	KK.PENGAMBIRAN PERMAI II I/10	0	R1
UIW SUMATERA BARAT	UP3 PADANG	13103	ULP INDARUNG	201901	1,3103E+11	1	SUSI HELMI	JL	JL.TELUK NIBUNG RT.04/II	0	R1
UIW SUMATERA BARAT	UP3 PADANG	13103	ULP INDARUNG	201901	1,3103E+11	1	SUMARNO	XX	KT.LUAR LM.MANIS H17	0	R1
UIW SUMATERA BARAT	UP3 PADANG	13103	ULP INDARUNG	201901	1,3103E+11	1	RULMI	XX	KOTO LUAR RT2/RWII	0	R1
UIW SUMATERA BARAT	UP3 PADANG	13103	ULP INDARUNG	201901	1,31031E+11	1	RAMADANI.SH	JL	JL.KAPALO KOTO RT 05/II PAUH	0	R1
UIW SUMATERA BARAT	UP3 PADANG	13103	ULP INDARUNG	201901	1,3103E+11	1	ASMAN	JL	JL.IRIGASI PSR.BARU RT 02/I PA	0	R1
UIW SUMATERA BARAT	UP3 PADANG	13103	ULP INDARUNG	201901	1,3103E+11	1	M A R N E L L Y	XX	DURIAN TARUNG RT20/VIII BLK SD	0	R1
UIW SUMATERA BARAT	UP3 PADANG	13103	ULP INDARUNG	201901	1,3103E+11	1	ULIL ABSAR/MELAYU	XX	KOTO PRK.PISANG	0	R1
UIW SUMATERA BARAT	UP3 PADANG	13103	ULP INDARUNG	201901	1,31031E+11	1	ASNIDAR	JL	Jl.SAKO RT.03 RW.04	0	R1
UIW SUMATERA BARAT	UP3 PADANG	13103	ULP INDARUNG	201901	1,3103E+11	1	ELIDAWATI	KR	NUANSA INDAH No.A/02	0	R1

UIW SUMATERA BARAT	UP3 PADANG	13103	ULP INDARUNG	201901	1,31036E+11	1	AGUSMAN	JL	TANJUNG SABA		R1
UIW SUMATERA BARAT	UP3 PADANG	13103	ULP INDARUNG	201901	1,3103E+11	1	WISNIN	XX	KP BATU DEPAN 55	0	R1
UIW SUMATERA BARAT	UP3 PADANG	13103	ULP INDARUNG	201901	1,3103E+11	1	NURITA	XX	KK.HARBAINDO H/5-20 PAMPANGAN	0	R1
UIW SUMATERA BARAT	UP3 PADANG	13101	ULP BELANTI	201901	1,31011E+11	1	DARMI B	XX	PS RAYA BARAT II NO 50	0	B1
UIW SUMATERA BARAT	UP3 PADANG	13101	ULP BELANTI	201901	1,31011E+11	1	MARDIANIS	XX	GAJAH MADA NO.16 RT.001/002	0	B1
UIW SUMATERA BARAT	UP3 PADANG	13108	ULP BALAI SELASA	201901	1,3108E+11	4	POSKO DAMKAR	DS	BALAI SELASA	0	P1
UIW SUMATERA BARAT	UP3 PADANG	13108	ULP BALAI SELASA	201901	1,31083E+11	4	TOWER LLAJR	PS	PASAR SILAUT	0	P1

DAYA	KOGOL	NOPEL	KDDK	THBLMUT	JNSMUT	KDPEMBMETER	KET_KDPEMBMETER	TGLBACALALU	TGLBACAAKHIR	TGLRUBAH	SLALWBP	SAHLWBP	JAMNYALA	SAHLWBP_CABUT
1300	0	AH013448	ACATSQC15600	201301	EFGJ	M	MEKANIK/MANUAL	20181125	20181226	20121220	50912	51225	240,769	NULL
1300	0	AH051344	ACAPGWC08100	201203	CEFGL	M	MEKANIK/MANUAL	20181125	20181226	20120207	29032	29272	184,615	NULL
1300	0	AH057145	ACAPHIC13200	200805	A	M	MEKANIK/MANUAL	20181125	20181226	20080416	16204	16534	253,846	NULL
2200	0	AH061680	ACAKTKD01600	201707	J	M	MEKANIK/MANUAL	20181126	20181227	20170608	11726	12319	269,545	NULL
900	0	AH027166	ACAAPJD17700	199812	A	M	MEKANIK/MANUAL	20181126	20181227	19000101	24903	25038	150	NULL
900	0	AH032454	ACASSCC31100	200005	A	M	MEKANIK/MANUAL	20181125	20181226	19000101	31304	31426	135,556	NULL
1300	0	AH033707	ACACPFD07300	201302	EFGJ	M	MEKANIK/MANUAL	20181126	20181227	20130109	26960	27122	124,615	NULL
450	0	AH026105	ACACPFD20100	199111	A	M	MEKANIK/MANUAL	20181126	20181227	19000101	49672	49716	97,778	NULL
1300	0	AH050804	ACAKTKD09900	200403	A	M	MEKANIK/MANUAL	20181126	20181227	19000101	67519	67921	309,231	NULL
130	0	AH049	ACAPRXC	201309	EFJ	M	MEKANIK/M	20181125	20181226	20130829	43416	43656	184,615	NULL

0		321	14700				ANUAL							
450	0	AH021381	ACADTGD17800	199411	A	M	MEKANIK/MANUAL	20181126	20181227	19000101	20181	20264	184,444	NULL
450	0	AH045021	ACAPHVD23800	199111	A	M	MEKANIK/MANUAL	20181126	20181227	19000101	11335	11348	28,889	NULL
900	0	AH061214	ACABGSC09300	201008	A	M	MEKANIK/MANUAL	20181125	20181226	20100714	14712	14864	168,889	NULL
2200	0	AH036049	ACAUGRD01600	201107	CEG	M	MEKANIK/MANUAL	20181126	20181227	20110615	45509	45678	76,818	NULL
900	0	XX5596327	ACAJULD12708	201812	A	E	ELEKTRONIK	20181127	20181227	20181115	35	126	101,111	NULL
450	0	AH011297	ACAKBUD08900	201610	J	M	MEKANIK/MANUAL	20181126	20181227	20160914	1555	1585	66,667	NULL
900	0	AH047299	ACAPPGC14500	200203	A	M	MEKANIK/MANUAL	20181125	20181226	19000101	23140	23227	96,667	NULL
900	0	AA120308	AAAPBAD09900	199608	A	M	MEKANIK/MANUAL	20181126	20181229	19000101	23422	23594	191,111	NULL
1300	0	AA121196	AAANGFD12100	199706	A	M	MEKANIK/MANUAL	20181126	20181227	19000101	53463	53657	149,231	NULL
900	3	AJ003749	AHABLID00300	201405	HIKL	M	MEKANIK/MANUAL	20181128	20181229	20100901	10491	10582	101,111	NULL
900	2	AJ301101	AHFLNVA11700	199609	A	M	MEKANIK/MANUAL	20181125	20181226	19000101	6568	6615	52,222	NULL
4400	3	AJ252926	AHETPUE12100	201602	EFJ	M	MEKANIK/MANUAL	20181129	20181230	20160120	27276	27825	124,773	NULL

A.2. PLN Business Customer Data July 2020

IDPEL	NAMA_UP	BLTH	TARIP	DAYA	KDPEMBMETER	JAMNYALA	PEMKWH	KWHLWBP	KWHWBP	RPDISKON	RPLWBP	RPWBP	RPPLN
131011203089	ULP BELANTI	201901	B1	900	M	191.111	172	108	64	0	45360	29760	98970
131011211969	ULP BELANTI	201901	B1	1300	M	149.231	194	194	0	0	187404	0	187404
131010230831	ULP BELANTI	201901	B1	1300	E	100.769	131	131	0	0	126546	0	126546
131010404303	ULP BELANTI	201901	B1	900	M	183.333	165	108	57	0	45360	26505	95715
131010595607	ULP BELANTI	201901	B1	2200	M	129.091	284	284	0	0	312400	0	315400
131094026980	ULP SUNGAI PENUH	201901	B1	900	M	264.444	238	108	130	0	45360	60450	129660
131090010054	ULP SUNGAI PENUH	201901	B1	2200	M	68.636	151	151	0	0	166100	0	166100
131090140889	ULP SUNGAI PENUH	201901	B1	1300	M	262.308	341	341	0	0	329406	0	332406
131090155417	ULP SUNGAI PENUH	201901	B1	900	M	204.444	184	108	76	0	45360	35340	104550
131092024746	ULP SUNGAI PENUH	201901	B1	900	M	257.778	232	108	124	0	45360	57660	126870
131090058544	ULP SUNGAI PENUH	201901	B1	1300	M	153.077	199	199	0	0	192234	0	192234
131090011186	ULP SUNGAI PENUH	201901	B1	1300	M	140.000	182	182	0	0	175812	0	175812
131093029967	ULP SUNGAI PENUH	201901	B1	900	M	100.000	90	90	0	0	37800	0	61650
131090086562	ULP SUNGAI PENUH	201901	B1	2200	M	212.727	468	468	0	0	514800	0	517800
131090144431	ULP SUNGAI PENUH	201901	B1	2200	M	40.000	88	88	0	0	96800	0	96800
131090055443	ULP SUNGAI PENUH	201901	B1	900	M	17.778	16	16	0	0	6720	0	30570
131090161234	ULP SUNGAI PENUH	201901	B1	900	M	27.778	25	25	0	0	10500	0	34350
131090023752	ULP SUNGAI PENUH	201901	B1	900	M	176.667	159	108	51	0	45360	23715	92925
131090021023	ULP SUNGAI PENUH	201901	B1	900	M	210.000	189	108	81	0	45360	37665	106875
131090027353	ULP SUNGAI PENUH	201901	B1	1300	M	131.538	171	171	0	0	165186	0	165186
131090015965	ULP SUNGAI PENUH	201901	B1	2200	M	123.182	271	271	0	0	298100	0	301100

131090027105	ULP SUNGAI PENUH	201901	B1	1300	E	263.846	343	343	0	0	331338	0	334338
131090012042	ULP SUNGAI PENUH	201901	B1	2200	M	40.000	88	88	0	0	96800	0	96800
131090139442	ULP SUNGAI PENUH	201901	B1	900	M	196.667	177	108	69	0	45360	32085	101295
131090025541	ULP SUNGAI PENUH	201901	B1	900	M	237.778	214	108	106	0	45360	49290	118500
131095010099	ULP SUNGAI PENUH	201901	B1	450	M	264.444	119	30	89	0	7620	37380	55575
131090026519	ULP SUNGAI PENUH	201901	B1	1300	M	532.308	692	692	0	0	668472	0	671472
131090110522	ULP SUNGAI PENUH	201901	B1	1300	M	289.231	376	376	0	0	363216	0	366216
131090025184	ULP SUNGAI PENUH	201901	B1	900	M	57.778	52	52	0	0	21840	0	45690
131090023902	ULP SUNGAI PENUH	201901	B1	900	M	268.889	242	108	134	0	45360	62310	131520
131090023951	ULP SUNGAI PENUH	201901	B1	1300	M	146.154	190	190	0	0	183540	0	183540
131095009233	ULP SUNGAI PENUH	201901	B1	450	M	175.556	79	30	49	0	7620	20580	38775
131090137007	ULP SUNGAI PENUH	201901	B1	900	M	0.000	0	0	0	0	0	0	23850
131090139491	ULP SUNGAI PENUH	201901	B1	1300	M	66.923	87	87	0	0	84042	0	84042
131090025802	ULP SUNGAI PENUH	201901	B1	900	M	0.000	0	0	0	0	0	0	23850
131090023458	ULP SUNGAI PENUH	201901	B1	3500	M	125.429	439	439	0	0	482900	0	485900
131090041922	ULP SUNGAI PENUH	201901	B1	1300	M	40.000	52	52	0	0	50232	0	50232
131090184358	ULP SUNGAI PENUH	201901	B1	4400	M	178.409	785	785	0	0	863500	0	866500
131099238216	ULP SUNGAI PENUH	201901	B1	2200	M	90.909	200	200	0	0	220000	0	220000
131090053351	ULP SUNGAI PENUH	201901	B1	1300	M	40.000	52	52	0	0	50232	0	50232
131090166284	ULP SUNGAI PENUH	201901	B1	900	M	215.556	194	108	86	0	45360	39990	109200
131091519379	ULP SUNGAI PENUH	201901	B1	2200	M	104.091	229	229	0	0	251900	0	254900

IDPEL	NAMA_UP	BLTH	TARIP	DAYA	KDPEMBMETER	JAMNYALA	PEMKWH	KWHLWBP	KWHWBP	RPDISKON	RPLWBP	RPWBP	RPPLN
1.3101e+11	ULP KURANJI	202004	B1	2200	M	61.364	135	135	0	0	148500	0	148500
1.3101e+11	ULP KURANJI	202005	B1	2200	M	84.091	185	185	0	0	203500	0	203500
1.3101e+11	ULP KURANJI	202006	B1	2200	M	142.273	313	313	0	0	344300	0	347300
1.3101e+11	ULP KURANJI	202007	B1	2200	M	113.182	249	249	0	0	273900	0	276900
1.3101e+11	ULP KURANJI	202008	B1	2200	M	99.091	218	218	0	0	239800	0	239800
1.3101e+11	ULP KURANJI	202009	B1	2200	M	159.545	351	351	0	0	386100	0	389100
1.3101e+11	ULP KURANJI	202010	B1	2200	M	153.636	338	338	0	0	371800	0	374800
1.3101e+11	ULP KURANJI	202011	B1	2200	M	165.909	365	365	0	0	401500	0	404500
1.3101e+11	ULP KURANJI	202012	B1	2200	M	68.182	150	150	0	0	165000	0	165000
1.3101e+11	ULP KURANJI	201901	B1	1300	M	220.769	287	287	0	0	277242	0	280242
1.3101e+11	ULP KURANJI	201902	B1	1300	M	146.923	191	191	0	0	184506	0	184506
1.3101e+11	ULP KURANJI	201903	B1	1300	M	61.538	80	80	0	0	77280	0	77280
1.3101e+11	ULP KURANJI	201904	B1	1300	M	55.385	72	72	0	0	69552	0	69552
1.3101e+11	ULP KURANJI	201905	B1	1300	M	53.077	69	69	0	0	66654	0	66654
1.3101e+11	ULP KURANJI	201906	B1	1300	M	46.154	60	60	0	0	57960	0	57960
1.3101e+11	ULP KURANJI	201907	B1	1300	M	52.308	68	68	0	0	65688	0	65688
1.3101e+11	ULP KURANJI	201908	B1	1300	M	60.000	78	78	0	0	75348	0	75348
1.3101e+11	ULP KURANJI	201909	B1	1300	M	51.538	67	67	0	0	64722	0	64722
1.3101e+11	ULP KURANJI	201910	B1	1300	M	50.769	66	66	0	0	63756	0	63756
1.3101e+11	ULP KURANJI	201911	B1	1300	M	50.769	66	66	0	0	63756	0	63756
1.3101e+11	ULP KURANJI	201912	B1	1300	M	47.692	62	62	0	0	59892	0	59892

132221281761	ULP TABING	201901	B2	23000	A	40.000	920	920	0	0	1349898	0	1355898
132221281761	ULP TABING	201902	B2	23000	A	40.000	920	920	0	0	1349898	0	1355898
132221281761	ULP TABING	201903	B2	23000	A	50.652	1165	1165	0	0	1709381	0	1715381
132221281761	ULP TABING	201904	B2	23000	A	51.739	1190	1190	0	0	1746063	0	1752063
132221281761	ULP TABING	201905	B2	23000	A	53.696	1235	1235	0	0	1812091	0	1818091
132221281761	ULP TABING	201906	B2	23000	A	58.696	1350	1350	0	0	1980828	0	1986828
132221281761	ULP TABING	201907	B2	23000	A	50.565	1163	1163	0	0	1706447	0	1712447
132221281761	ULP TABING	201908	B2	23000	A	54.304	1249	1249	0	0	1832633	0	1838633
132221281761	ULP TABING	201909	B2	23000	A	52.652	1211	1211	0	0	1776876	0	1782876
132221281761	ULP TABING	201910	B2	23000	A	54.217	1247	1247	0	0	1829698	0	1835698
132221281761	ULP TABING	201911	B2	23000	A	54.870	1262	1262	0	0	1851707	0	1857707
132221281761	ULP TABING	201912	B2	23000	A	53.261	1225	1225	0	0	1797418	0	1803418
132221281761	ULP TABING	202001	B2	23000	A	52.478	1207	1207	0	0	1771007	0	1777007
132221281761	ULP TABING	202002	B2	23000	A	53.087	1221	1221	0	0	1791549	0	1797549
132221281761	ULP TABING	202003	B2	23000	A	51.565	1186	1186	0	0	1740194	0	1746194
132221281761	ULP TABING	202004	B2	23000	A	54.783	1260	1260	0	0	1848773	0	1854773
132221281761	ULP TABING	202005	B2	23000	A	46.696	1074	1074	0	0	1575859	0	1581859
132221281761	ULP TABING	202006	B2	23000	A	42.783	984	984	0	0	1443804	0	1449804
132221281761	ULP TABING	202007	B2	23000	A	46.739	1075	1075	0	0	1577326	0	1583326
132221281761	ULP TABING	202008	B2	23000	A	47.826	1100	1100	0	0	1614008	0	1620008
132221281761	ULP TABING	202009	B2	23000	A	47.957	1103	1103	0	0	1618410	0	1624410

A.3. Result Cluster

DAYA	Cluster	KWHLWBP	KWHWBP
1110000	1	59008	16848
690000	1	52600	14832
555000	1	59528	12216
1110000	1	174896	36608
555000	1	133938	30132
690000	1	107480	20816
1385000	1	199488	65712
865000	1	213300	50160
555000	1	193288	0
1730000	1	169400	33400
555000	1	190080	42400
345000	1	82104	25200
690000	1	125552	27320
690000	1	62232	12336
1110000	1	71280	20368
690000	1	113496	24080
1385000	1	198464	65072
865000	1	186600	44240
555000	1	159896	33096
345000	1	81380	24972
690000	1	95648	18528

283	**900**	**2**	**108**	**64**
284	1300	2	194	0
285	1300	2	131	0
286	900	2	108	57
287	2200	2	284	0
288	900	2	108	130

289	2200	2	151	0
290	1300	2	341	0
291	900	2	108	76
292	900	2	108	124
293	1300	2	199	0
294	1300	2	182	0
295	900	2	90	0
296	2200	2	468	0
297	2200	2	88	0
298	900	2	16	0
299	900	2	25	0
300	900	2	108	51
301	900	2	108	81
302	1300	2	171	0
303	2200	2	271	0

DAYA	Cluster	KWHLWBP	KWHWBP
2075000	3	469500	126360
2425000	3	381300	140310
2075000	3	455190	120870
2425000	3	370500	137040
2075000	3	442050	118050
2425000	3	329370	123960
2075000	3	489930	134790
2425000	3	377250	145020
2425000	3	367410	142770
2075000	3	500640	134730
2425000	3	387120	146580

2075000	3	487860	145950
2425000	3	356100	134280
2075000	3	460440	128880
2075000	3	489930	132180
2425000	3	365940	137460
2425000	3	353610	134010
2075000	3	492420	131580
2425000	3	348510	131790
2075000	3	461400	123000
2425000	3	317700	117930

A.4. Calculation Cluster Result

Cluster	**DAYA**	**KWHLWBP**	**KWHWBP**
1	937837	115194	27827
2	4260	544	35
3	2226351	390803	123297

Appendix B R Code

B.1. Clustering K-Means Code

```
#Install Package
install.packages("tidyverse")
install.packages("cluster")
install.packages("factoextra")
install.packages("data.table")
install.packages("gridExtra")
install.packages("xlsx")

#Library
library(tidyverse)
library(cluster)
library(factoextra)
library(data.table)
library(gridExtra)
library(xlsx)

#Initial Data use tibble
df <- as_tibble(bisni4)

#filter Data use tibble
df1<- df %>% select(DAYA,KWHLWBP,KWHWBP)


#Visualisasi Data
View(df2)

#Filter Condition
df3 <- df %>% filter(BLTH > 202007)
df4<- df3 %>% select(DAYA,KWHLWBP,KWHWBP)

#Remove Missing Value
data1<-na.omit(df1)
data2<-na.omit(df2)
data3<-na.omit(df4)

#Scale
data4<- scale(data1)
```

```
data5<- scale(data2)
data6<- scale(data3)
head(data4)

#Clustering Distance Measure
distance <- get_dist(data1)
fviz_dist(distance, gradient = list(low = "#00AFBB",
                                    mid  =  "white",  high  =
"#FC4E07"))

#K-Means Clustering
k2 <- kmeans(data4, centers = 2, nstart = 25)
str(k2)
k2

#Cluster Plot
fviz_cluster(k2, data = data4)

#Alternative Cluster Plot
data7 %>%
  as_tibble() %>%
  mutate(cluster = k2$cluster,
         state = row.names(data7)) %>%
  ggplot(aes(DAYA, KWHLWBP, color = factor(cluster), label =
KWHWBP)) +
  geom_text()

#Alternative Cluster Plot1
k3 <- kmeans(data4, centers = 3, nstart = 25)
k4 <- kmeans(data4, centers = 4, nstart = 25)
k5 <- kmeans(data4, centers = 5, nstart = 25)

# plots to compare
p1  <-  fviz_cluster(k2,  geom  =  "point",  data  =  data4)  +
ggtitle("k = 2")
p2  <-  fviz_cluster(k3,  geom  =  "point",   data  =  data4)  +
ggtitle("k = 3")
p3  <-  fviz_cluster(k4,  geom  =  "point",   data  =  data4)  +
ggtitle("k = 4")
```

```
p4 <- fviz_cluster(k5, geom = "point",  data = data4) + ggtitle("k = 5")


grid.arrange(p1, p2, p3, p4, nrow = 2)


#Determining Optimal Clusters
#Elbow Method

set.seed(123)

# function to compute total within-cluster sum of square
wss <- function(k) {
  kmeans(data4, k, nstart = 10 )$tot.withinss
}

# Compute and plot wss for k = 1 to k = 15
k.values <- 1:15

# extract wss for 2-15 clusters
wss_values <- map_dbl(k.values, wss)

plot(k.values, wss_values,
     type="b", pch = 19, frame = FALSE,
     xlab="Number of clusters K",
     ylab="Total within-clusters sum of squares")

#Optimal Number Cluster
set.seed(123)
fviz_nbclust(data4, kmeans, method = "wss")

#Average Silhouette Method

# function to compute average silhouette for k clusters
avg_sil <- function(k) {
  km.res <- kmeans(data10, centers = k, nstart = 25)
  ss <- silhouette(km.res$cluster, dist(data10))
  mean(ss[, 3])
```

```
}

# Compute and plot wss for k = 2 to k = 15
k.values <- 2:15

# extract avg silhouette for 2-15 clusters
avg_sil_values <- map_dbl(k.values, avg_sil)

plot(k.values, avg_sil_values,
     type = "b", pch = 19, frame = FALSE,
     xlab = "Number of clusters K",
     ylab = "Average Silhouettes")

#Optimal Number Cluster
fviz_nbclust(data10, kmeans, method = "silhouette")

#Gap Statistic Method

set.seed(123)
gap_stat <- clusGap(data10, FUN = kmeans, nstart = 25,
                    K.max = 10, B = 50)
# Print the result
print(gap_stat, method = "firstmax")

fviz_gap_stat(gap_stat)

#Final
set.seed(123)
final <- kmeans(data4, 3, nstart = 25)
print(final)
head(final)
fviz_cluster(final, data = data4)

final2 <- data1 %>%
  mutate(Cluster = final$cluster) %>%
  group_by(Cluster) %>%
  summarise_all("mean")

final1 <-data1 %>%
```

```
  mutate(Cluster = final$cluster) %>%
  select(DAYA,Cluster,KWHLWBP,KWHWBP) %>%
  arrange(Cluster)

view(final2)

view(final2)

write.csv2(final1,"wwee.csv")
write.csv2(final2,"wwee1.csv")
```

B.2. AHP CODE

```
#Install Package
install.packages("tidyverse")
install.packages("gridExtra")
install.packages("data.table")
install.packages("ahpsurvey")


#Library
library(tidyverse)
library(gridExtra)
library(data.table)
library(ahpsurvey)

#Initial Data use tibble
df <- as_tibble(wwee1)
df1<- df %>% select(DAYA,KWHLWBP,KWHWBP)

df1 %>%
  ahp.mat(df1, negconvert = TRUE) %>%
  head(3)

df2 <- df1 %>%
  ahp.mat(df1, negconvert = T)

amean <- ahp.aggpref(df2,df1, method = "arithmetic")
amean
```

B.3. CLV Code

```
#Install Package
install.packages("tidyverse")
install.packages("gridExtra")
install.packages("data.table")

#Library
library(tidyverse)
library(gridExtra)
library(data.table)


CLV <- (Power *WP) + (POL *WPL) + (PL *WPL)
```