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

**THESIS**

**In partial fulfilment of the requirements**

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

**From Institut Teknologi Bandung**

**By**

**Radit Rahmadhan**

**Student ID: 29020003**

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

****

**INSTITUT TEKNOLOGI BANDUNG**

**2022**

**ABSTRACT**

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

By

**Radit Rahmadhan**

**Student ID: 29020003**

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

Write abstract in English here. The font size in the abstract is 12, times new roman characters and not written in bold. Abstract consists of 500-800 words. The first sentence of the abstract is three spaces from the last line of the name of study program. The first word or the beginning of a new paragraph should be separated by a single space from the last sentence of the preceding paragraph. Abstracts are written in English and consist of 500-800 words. The end of abstract, which is still in the sheet of this abstract but in separate line from abstract, ended with keyword (keywords). Keywords (keywords) written at the bottom of the abstract are words, which are related and derived from the abstract. IT SHOULD NOT BE derived / taken from the body/ text of the thesis. The minimum numbers of keywords are 4 single words or a single meaningful two-word combination and maximum number of keywords are 7 single words or a single meaningful two-word combination.

The proposed research contained in the abstract should comprehensively carried out the research problems, which explaining reasons why the scientific issues need to be studied; the background; the research stages; the assumptions and hypotheses that need to be proved; as well as the research objectives to be achieved; the method used; the brief review; and the explanation of the results that will be obtained in the study. The contribution of the research results to the science should be clearly stated. Abstract pages and the preparation section of the thesis should be separately numbered from the main body page number of the thesis. The preparatory pages are numbered by using Roman numerals i, ii, iii, iv, …, x, xi, ... to distinguish it from the main body page numbers of the of thesis in the form of Arabic numerals.

Keywords: keyword 1, keyword 2, etc.

***ABSTRAK***

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

*Oleh*

**Radit Rahmadhan**

**29020003**

**(Program Studi Magister Sains Manajemen)**

Write abstract in Bahasa Indonesia here. The font size is 12 and not bold. The first sentence of the abstract should be 3 spaces away from the last line of the program name. The abstract content, which is an Indonesian translation of the abstract in English language, is printed in size 1; not bold; and in italic form. The first word or the beginning of a new paragraph is separated by a single space from the last sentence of the preceding paragraph. Abstract are written in Indonesian and consist of 500-800 words. The end of abstract, which is still in the sheet of this abstract but in separate line from abstract, ended with keyword. Keywords written at the bottom the abstract are words which are related and derived from the abstract, and SHOULD NOT BE derived / taken from the body / text thesis. The minimum numbers of keywords are 4 single words or a single meaningful two-word combination and maximum number of keywords are 7 single words or a single meaningful two-word combination.

Keywords: kata kunci 1, kata kunci 2, dst.

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

By

**Radit Rahmadhan**

**Student ID: 29020003**

**Master Program of Science in Management**

Institut Teknologi Bandung

Approved

December

Supervisor

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

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

**THE GUIDANCE FOR USING THE THESIS**

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

This thesis has to be cited as:

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

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

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

**STATEMENT OF AUTHORSHIP**

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

Signature:

Bandung, December 20,2022

**ACKNOWLEDGMENTS**

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

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

DASDASDASDASDASDASDASDSADASD

# Introduction

## Background

Electricity is a vital energy for the sustainability of human activities both for individuals, community groups and the industrial world (a). 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 (b). Community activities tend to increase over time. The increase in activities encourages an increase in the operation of equipment with electric power. (sdasdasd). 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 (1). 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 [3]. 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 [3][4].

Some previous research discussed customer segmentation on customer electricity consumption [4], [7], [8], [10], [12]. and electricity demand [7], [9]-[11]. The research context is more about finding new customer behavior patterns in consuming electricity and more methods that use a combination of K-Means and Self Organizing Maps (SOM) and other clustering methods [4], [7], [8], [10], [12].

Other studies use regression methods for customer segmentation [7], [9]-[11], 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 [4], [7], [8], [10], [12]. 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 [11].Previous research on customer segmentation is generally based on total electricity consumption per day [4], [7], [8], [10], [12]. Other studies only analyzed tariffs, electricity, and total bills by combining K-Means and CRM [11].

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

## Research Objectives

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

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

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

## Research Question

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

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

## Research approach and methods

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

## Research Scope and Limitations

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

1. This study focuses on the electricity consumption of business customers specific to the Padang region. The selection of variables used power capacity, peak load consumption, and non-peak load consumption. Further research could investigate a wider selection of regions and a deeper selection of variables.

2. This study used one month of business customer transaction data. Future studies can use one year or more to be further examined as input for segmentation models.

3. This study examines proposals using a combination of machine learning models namely K-Means Clustering, Customer Lifetime Value and Customer Relationship Management (CRM) strategies. Other advance methods can be investigated in further research.

## Writing Structure

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

# Literature Review

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

## Previous Segmentation Studies Based on Electricity Consumption Data

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

Previous studies in customer segmentation in electricity consumption have explored various dimensions of the customer clustering problem[4], [7], [8], [10],[12]. They use the context of electricity consumption as a case study to find out patterns of electricity use in predicting future electricity consumption. Several clustering models, one of which is often used, namely K-Means Clustering, has explored customer grouping by considering patterns of electricity use and electricity demand to meet electricity consumption based on what has been prepared by company[7], [9]– [11].

Table II.1 CUSTOMER SEGMENTATION BASED ON ELECTRICITY CONSUMPTION

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Article | Business Context | Dataset | Segmentation  Features | Segmentation  Method |
| [4] | Electricity Load Profile in Ireland | Experimental data period January 1, 2009, to December 31, 2010, | Dwelling type, No. of bedrooms, Age, Social Class, Electronic Type | K-means, k-medoid and Self Organizing Maps (SOM) |
| [6] | Electricity Consumption in South Africa | South Africa Electric Load Profile Data from 1994 to 2014 | X=Hour (load profile multiple 1 day)  Y= X multiple All household | K-Means  And Self Organizing Maps (SOM) |
| [5] | Electricity  Demand Signature in Andalusian | The load data of 64 buildings located in Andalusia, Spain | Identity, Industrial Division, Industrial Categories, Mean Power Consumption, Power Consumption | Variable selection (Feature Selection), Model (K-Means, Hierarchical Clustering, K-Medoid Clustering), Validation (Connectivity, Dunn and  Silhouette indexes) |
| [10] | Electricity Load Profile | Smart Metering Data in 2009 | Identity, Social Status, age, gender, Demand kWh, Income | Regression Ordinary Least Square (OLS), Evaluation (Root Mean Square Error (RMSE)) |
| [9] | Electricity Load Profile | Residential Demand Data during November,2017 until February, 2018 | Identity, Daily Consumption, Load Profile, Peak Hour, Demand | K-means, Fuzzy C-Means (FCM) and Self Organizing Maps (SOM) |
| [8] | Electricity Consumption Forecasting | Electricity Consumption Data from 46 homes in Texas | Identity, Time, Total kWh | Model (Artificial neural networks, regression  trees, random forest regression, 𝑘-nearest neighbors’ regression,  and support vector regression), Evaluation (Naive forecast, random  forecast, the ARIMA model, and stepwise regression) |
| [15] | Electricity Demand with Renewable Technologies | Half -hourly energy use for 1 year data | Average energy use,  energy–temperature correlation, entropy of the load-shape representative vector, and distance to  wind generation patterns. | Model (K-Medoids), Validities (average silhouette) |
| [11] | Electricity Consumption in Indonesia | Customer Transaction in September 2021 | Rate, Power, Total kWh, Total Cost, Flash Time | Variable selection with correlation  Model (K-Means)  Validity (Silhouette Method)  Explores (Customer Relationship Management (CRM)) |

A context study of load profile electricity [4] using experimental data by installing 4000 intelligent meters in several homes in Ireland with existing methods used to classify household electricity use, in general, can be divided into four categories, statistics, manipulation, time series, and clustering. Statistical methods have been widely used in the unregulated power market to form a standard load Personal Classes (PC). A typical load PC is used for settlement purposes and estimates the amount and Time of Use of electricity used. A series of PCs are manufactured for different market segments (e.g., residential, commercial, industrial) and derived on an average for all customers within a customer class.

Research on electricity consumption in South Africa [6] focus on household customers, which aims to classify customers based on patterns and types using electricity using the K-Means clustering model and Self Organizing Maps (SOM). They used internal and external validation to evaluate the clustering structure based on the expected behavior of South African households' daily electricity consumption. Another study used electrical load data also in Andalusia, Spain [5], but the research context was about electricity demand. They determine interrelated variables to predict customer segmentation models using a combination model between K-Means clustering and K-medoid clustering. This study aims to provide an alternative customer segmentation that can manage several types of customers. It then presents the segmentation results based on the characteristics of the load curve. Finally, they compare the two marks and provide solutions to the effects of classification and segmentation.

Research on the context of electricity load data [9] uses electricity demand data to predict electricity loads per day based on the heterogeneity of electricity demand behavior by customers, then processed using a combination of K-Means clustering models and Self Organizing Maps (SOM) and Fuzzy C-Means. The segmentation results provide the proper group identification for electricity demand per day. The result shows a tremendous impact because it can save on utility costs based on electricity reduction by customers. Another study with the same context as [9], but this study uses data from smart meters in 2009 [10], they use a regression model with an evaluation of the root mean square error for customer segmentation based on electricity demand used, age, and income from the customer. The aim is to find new customer electricity usage behavior patterns based on predetermined variables. Another study uses six regression models to predict daily electricity consumption based on the total electricity consumption used by customers [8]. They compared the models to find new patterns of customers' daily electricity usage.

Research on the context of looking for energy reserves based on the number of customer electricity requests [15] uses data on customers' half-day electricity usage by selecting variables based on the average amount processed by adding wind variables as alternative electrical energy. This study uses the K-Medoid model and the Silhouette method to validate the number of clusters to apply an efficient time series clustering methodology that explicitly considers the pattern of renewable energy generation. Other research on the context of electricity consumption in Indonesia [11]. They used data on customer electricity bills in September 2021 with predictors of power, rate, total kwh, flash sale, total cost, which were tested for variable correlation. This research uses the K-Means Clustering model and the Silhouette Method as the number of clusters to get customer segmentation based on the characteristics of customers paying for electricity according to the power used. The clustering results will be explored using the CRM model to gain insight to act to customers in the future according to the wisdom that has been carried out.

## Previous Studies on Segmentation Based on Customer Lifetime Value

Previous studies in customer segmentation have explored various dimensions of customer clustering problems [16]–[18]. Many of them use the marketing context as a case study. The K-Means clustering model and Customer Lifetime Value explores customer grouping by considering the specified product preferences and predicting customer behavior in buying products offered by the company [13]

A context study in marketing combines the Customer Lifetime Value (CLV) and K-Means models in each customer segment [16]. 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 [13]. The K-means clustering model to map customers based on the same characteristics is then classified to distinguish potential customers for repurchase and then validated using the elbow method. This study uses data from all AR-Pulsabiz pulse server operators in Malang, Indonesia, to predict the future of Small and Medium Enterprises. The number of potential customers who will become operators by using a combination of the K-Means Clustering model and the LRFM model to group customers to provide services according to priority.

Research in pharmaceutical marketing [18] also has the same objective [14], 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 [19] with the same research objective [3]. It also has similar goals and models [3] to marketing research in Telecommunication Companies [20]. However, they do not use the CLV model but use the Neural Network to classify priority customers after getting the results from clustering.

## Marketing Strategy in Customer Relationship Management

Previous research, marketing strategy in customer segmentation was determined based on the CLV result, then we can evaluate target that aims to develop customer service improvement strategies based on the concept of customer relationship CRM [21] .

There are two programs from the customer relationship strategy. If the company chooses the right approach, it will increase profits and retain customers [21], [22] as follows.

1. Sustainable Marketing

This program is a program to maintain and increase customer loyalty through special long-term services and increase value by studying the characteristics of customers [28]– [31]. Implementing a sustainable marketing program from this concept will be explained as follows.

A. Continuous Replenishment Program

This program is used for less profitable customers[27]. Approaches to programs such as partnership programs to encourage increased use of the company's services to customers [27], [28].

B. Business to Business

This program is used for profitable customers[29], [30]. The approach to this program is like providing special executive services to customers to improve service, so that customer trust will increase and become more loyal [36]– [39].

2. One to One Marketing

This program is an individual program aimed at satisfying customers' unique needs [35], [36]. This program uses customer information from online news and databases, followed by personal interactions to meet customers' unique needs [37], [38]. Build interactive marketing and post-marketing programs in developing customers using individual customer information [42]– [44]. The application of the one-to-one marketing program from this concept will be explained as follows.

A. Customer Business Development

This program is used for profitable customers [40], [41]. The approach to this program is to assess the benefits of marketing, finance, and management business processes [42], [43]. This program aims to explore the customer's business development by providing the best solutions and consulting regarding customers' services [40]– [42], [44].

B. Retail Account Marketing

This program is used for less profitable customers [39], [45]. 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 [46], [47].

## Research Position

In order to identify the gap and position of this study, this study classified the literature based on

# Research Methodology

## Research Philosophical Position

## Research Framework

## Data Collection

## Data Preparation

## Choice of Variable

## Clustering Model

## Marketing Strategy Definition

### Analytical Hierarchy Process

Dsdf

ssd

# Results and Analysis