Journal of
Information Systems Engineering
and Business Intelligence

Vol.8, No.2, October 2022
Available online at: http://e-journal.unair.ac.id/index.php/JISEBI

Segmentation using Customers Lifetime Value: Hybrid K-means Clustering and Analytic Hierarchy Process

Radit Rahmadhan[1)*], Meditya Wasesa[2)]

[1)2)]*School of Business and Management, Institut Teknologi Bandung, Indonesia*
Jl. Ganesha 10, Bandung
[1)]radit_rahmadhan@sbm-itb.ac.id, [2)]meditya.wasesa@sbm-itb.ac.id

Abstract

Background: Understanding customers' electricity consumption patterns is essential for developing predictive analytics, which is needed for effective supply and demand management.
Objective: This study aims to understand customers' segmentation and consumption behaviour using a hybrid approach combining the K-Means clustering, customer lifetime value concept, and analytic hierarchy process.
Methods: This study uses more than 16 million records of customers' electricity consumption data from January 2019 to December 2020. The K-Means clustering identifies the initial market segments. The results were then evaluated and validated using the customer lifetime value concept and analytical hierarchy process.
Results: Three customer segments were identified. Segment 1 has 282 business customers with a total capacity of 938,837 kWh, peak load usage of 27,827 kWh, and non-peak load usage of 115,194 kWh. Segment 2 has 508,615 business customers with a total capacity of 4,260 kWh, a peak load of 35 kWh, and a non-peak load of 544 kWh. Segment 3 has 37 business customers with a total capacity of 2,226,351 kWh, a peak load of 123.297 kWh, and a non-peak load of 390,803.
Conclusion: A business strategy that could be taken is to base customer relationship management (CRM) on the three-customer segmentation. For the least profitable segment, aside from retail account marketing, a continuous partnership program is needed to increase electricity consumption during the non-peak period. For the highly and moderately profitable segments, a premium business-to-business approach can be applied to accommodate their increasing energy consumption without excessive electricity use in the peak period. Special account executives need to be deployed to handle these customers.

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

Article history: Received 16 June 2022, first decision 23 August 2022, accepted 8 September 2022, available online 28 October 2022

I. INTRODUCTION

Electricity consumption in Indonesia increased by 98.89% from 2015 to 2020, with a major proportion taken by business customers [1]. However, the increasing demand is not always supported by adequate services. Electricity blackouts occur up to four times a month. This happens because customers use power above 200 thousand during the peak load than electricity outside peak hours. During the non-peak load hours, the usage is low. Based on these problems, electricity companies must understand customers' electricity use characteristics to maximise electricity distribution. For example, the low consumption by business customers due to the power cuts (under 50 hours per month) can be improved.

Segmentation groups customers based on similar characteristics [2] to improve understanding customer preferences that direct future actions [3]. Previous studies have segmented customers based on their daily electricity consumption [2]-[4]. The methods used include a hybrid between K-Means clustering and alternative clustering [3], [4] that aims to find patterns of electricity customer behaviour to predict future electricity consumption [2], [5] and optimise the ongoing electricity usage [5], [6]. Other studies analyse segmentation to gain insight or make company business decisions by combining K-Means clustering and a marketing strategy in customer relationship management (CRM) [5], [7].

To the best of our knowledge, articles combining classic K-Means clustering techniques with customer lifetime value (CLV) with a marketing strategy in CRM are lacking. Therefore, this study aims to develop a segmentation

[*] Corresponding author

ISSN 2443-2555 (online) 2598-6333 (print) © 2022 The Authors. Published by Universitas Airlangga.

doi: http://dx.doi.org/10.20473/jisebi.8.2.130-141

model that can reflect electricity consumption behaviour. These findings can help power companies improve their strategy in targeting customers according to their characteristics. We incorporate three variables in the segmentation model development: power capacity, peak load consumption, and non-peak load consumption. We utilise K-means clustering, analytic hierarchy process (AHP) approaches, and customer lifetime value aspects.

II. LITERATURE REVIEW

A. Previous Segmentation Studies Based on Electricity Consumption Data

Table 1 presents previous studies on customer segmentation using electricity consumption data, including the exploration of variables in the segmentation analyses [2], [3], [6]. It can also be seen that K-Means Clustering has been a popular technique [8], [9].

TABLE 1
PREVIOUS STUDIES ON CUSTOMER SEGMENTATION BASED ON ELECTRICITY CONSUMPTION

Article	Business Context	Dataset	Segmentation Features	Segmentation Method
[2]	Electricity load profile in Ireland	Experimental data from 1 January 2009 to 31 December 2010	Housing typology, number of bedrooms, age, socio-economic class, electronic type	K-means, k-medoid, and self-organising maps (SOM)
[3]	Electricity consumption in South Africa	South African electric load profile data from 1994 to 2014	X= hour (load profile multiple one days) Y= X multiple all household	K-Means and self-organising maps (SOM)
[4]	Electricity demand signature in Andalusia	The load data of 64 buildings located in Andalusia, Spain	Profile, 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)
[6]	Electricity load profile	Smart metering data in 2009	Profile, social status, age, gender, demand kWh, income	Regression ordinary least square (OLS), evaluation (root mean square error (RMSE))
[5]	Electricity load profile	Residential demand data from November 2017 until February 2018	Profile, daily consumption, load profile, peak hour, demand	K-Means, Fuzzy C-Means (FCM), and self-organising maps (SOM)
[8]	Electricity consumption forecasting	Electricity consumption data from 46 homes in Texas	Profile, time, total kWh	Model (artificial neural networks, regression trees, random forest regression, k-nearest neighbours' regression, and support vector regression), evaluation (Naive forecast, random forecast, the ARIMA model, and stepwise regression)
[9]	Electricity demand with renewable technologies	Half-hourly energy use for one-year data	Average energy use, energy–temperature correlation, the entropy of the load-shape representative vector, and distance to wind generation patterns.	Model (K-Medoids), validities (average silhouette)

McLoughlin et al. 2015 [2] profiled electricity load using experimental data by installing 4,000 intelligent meters in several homes in Ireland. A study in South Africa [3] examined household customers, aiming to classify them based on the patterns and types using the K-Means clustering model and self-organising maps (SOM). Another study used electrical load data in Andalusia, Spain [4], but the research context was electricity demand, not consumption. The aim was to provide an alternative customer segmentation to manage different customer types based on the load curve's characteristics.

Jang et al. 2021 [6] used electricity demand data to predict daily electricity load using a combination of K-Means clustering models and self-organising maps (SOM) and Fuzzy C-Means. The result shows a tremendous impact on utility cost reduction. Lee et al. 2021 [5] use a regression model encompassing electricity demand, age, and income variables. Another study compared six regression models to predict daily electricity consumption [8]. They compared the models to discover the patterns of customers' daily electricity usage. Banales et al. 2021 [9] estimated energy reserves using customer electricity requests. The study applied the K-medoid model and silhouette method to customers' half-day electricity usage considering wind energy as an alternative energy source.

B. Previous Studies on Segmentation Based on Customer Lifetime Value

Previous studies have incorporated customer lifetime value (CLV) in defining customer segments [7], [10], [11]. Lee et al. 2021 [7] use the K-Means clustering model and CLV considering product preferences to predict behaviour in buying products. A previous study has also combined the CLV and K-Means models for segmentation [7]. The grouping uses the K-Means Clustering method based on the LRFM Model (length, recency, frequency, monetary). The CLV value is generated from the multiplication of the LRFM normalisation results. Then, the weight values were analysed using the analytical hierarchy process (AHP) and turned into a matrix. The results show that some segments had a high customer loyalty value, from which companies can create strategies for retention.

Li et al. 2018 [10] explored supermarket marketing using LRFM models and K-Means clustering mapped customers based on the same characteristics and classified them into a potential repurchase category (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. Research in pharmaceutical marketing [11] also has the same objective [12]. However, they use eight validation methods to determine the correct groupings. Another study examined the transport industry using the K-Means clustering and CLV models to group customers [10] with the same research objective [13]. It also has similar goals and models [13] to marketing research in telecommunication companies [14]. However, they do not use the CLV model but the neural network to classify priority customers after clustering results.

C. Marketing Strategy in Customer Relationship Management

Two strategies that can increase profits and customer retention [12] are 1) sustainable marketing and 2) one-to-one marketing. Sustainable marketing aims to maintain and increase customer loyalty through long-term services [13]-[16], such as a continuous replenishment program. This can be directed to less profitable customers [15] because partnership programs can increase purchases [17]. Meanwhile, business-to-business programs can be applied to moderately profitable customers [16], [17]. Finally, high-end services can be provided to highly profitable customers to increase trust and loyalty[18]-[21].

One-to-one marketing focuses on individual approaches aligned with customers' unique needs [18], [19]. This program uses online news and databases of customer information, followed by personal interactions to understand their unique needs [20], [21]. It builds interactive marketing and post-marketing programs in developing customers using individual customer information [22]– [24]. This approach can be implemented in customer business development and retail account marketing. Customer business development aims for profitable customers and moderately profitable customers [22], [23] by assessing the benefits of marketing, finance, and management business processes [24], [25]. Retail account marketing is typically used for less profitable customers [26], [27]. The approach sees the customer as a partner in developing business opportunities. This program performs customer profiling using a CRM system [28], [29].

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

III. METHODS

Fig. 1 presents the research framework of this study that adopts standard methods for building predictive analytics [30]. The framework comprises five stages: data collection, data preparation, choice of variables, clustering model, and marketing strategy definition.

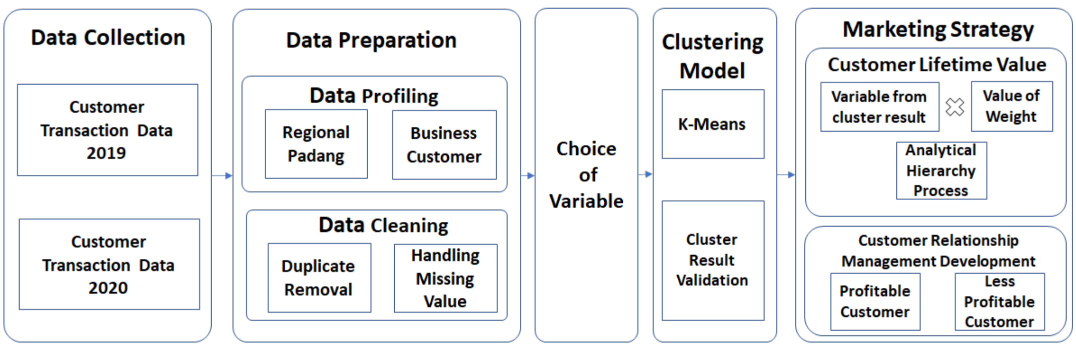

Fig. 1 The research framework

A. Data Collection

In this study, we obtained data from PT. PLN Persero, West Sumatra region. The data provided is customer transaction data from January 2019 to December 2020, which consists of 16,504,228 customers and 107 data variables. In this research, we used customer data in July 2020. Table 2 shows the descriptive statistics of the data set.

TABLE 2
DESCRIPTIVE OF DATA COLLECTION

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

B. Data Preparation

This section presents the data preparation processes for developing the prediction model, namely:

1) Data Profiling

The data selection starts by looking at the areas in West Sumatra that consumed the highest electricity. Fig. 2 presents the results of the plot analysis carried out in four areas of the service centre of PT. PLN Persero. The Padang area showed the highest electricity consumption compared to other sites.

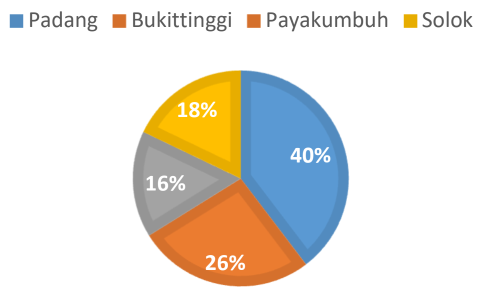

Fig. 2 Total electricity consumption based on region

The subsequent analysis looks at potential customers who use a higher total power consumption kWh. Fig. 3 presents the results of plot analysis based on total electricity consumption by customer category. Based on the regulations issued by the Indonesian government, customers are divided into five categories: household, social, government, business, and industrial. As shown in Fig. 3, business customers consume the highest kWh at around 37%, followed by industrial customers at 31% and other customers using electricity consumption below 15%. Therefore, this study focuses on business customers because they use higher electricity consumption than others and can increase company revenues.

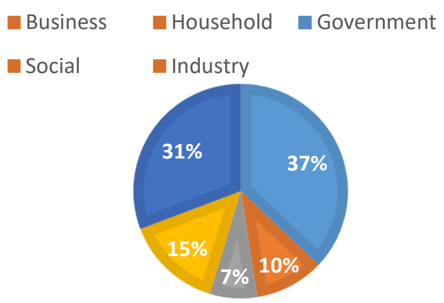

Fig. 3 Total electricity consumption based on customer energy

2) Data Cleaning

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

TABLE 3
DESCRIPTIVE OF DATA CLEANING

Variable	Data Type	Count	Max	Min	Variable Description
ID Customer	Integer	24,785	-	-	Identity of the customer
Customer Service Unit	String	12	-	-	Customer Service Units or service branches provided by the company which are in 4 customer service centres, namely Belanti, Painan, 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 the digital meter
Flash time	Double	27,904	4775.66	0	Electricity usage time by the customer
Total kWh	Integer	10,427	635,370	0	The total peak load kWh usage and peak external load kWh used by customers
Non-Peak Load	Integer	10,417	500,640	0	KWH used at peak external load by customers
Peak Load	Integer	1,515	146,580	0	KWH used at peak load by customers
Discount	Double	11	338,942	0	The company gives discounts based on the provisions of the company
Non-Peak Load Fee	Double	18,578	518,552,899	0	Payments made when using non-peak load
Peak Load Fee	Double	2,256	227,736,949	0	Payments made when using Peak Load
Total Cost	Double	21,621	732,079,768	0	The total cost paid by the customer

C. Choice of Variable

From the 13 variables of the dataset in Table 3, we select the variable with Integer or Double types in Table 4. Table 4 shows nine possible variables used in the clustering model. Because the study focuses on identifying customer segmentation on power based on peak and non-peak loads. This research will treat electricity consumption from 6.00 am to 4.59 pm as peak load and the rest as non-peak load. Thus, the peak load and non-peak load electricity consumption are the main features of the clustering model.

TABLE 4
DESCRIPTIVE OF POTENTIAL VARIABLE

Variable	Data Type	Function
Power	Integer	Predictor
Flash time	Double	
Non-Peak Load	Integer	
Peak Load	Integer	
Total kWh	Integer	
Discount	Double	
Non-Peak Load Fee	Double	
Peak Load Fee	Double	
Total Cost	Double	
Customer segmentation	Double	Predicted

D. Clustering Model

K-Means is one of the well-known unsupervised learning techniques for cluster analysis [31], aggregating or dividing datasets into several clusters according to the similarity value. The model is used because the algorithm is straightforward, and users can determine the number of clusters.

We determine the number of clusters (k) by the elbow method of validation [7], [10], [12]. The output is the number of data clusters to be processed. This method visualises the number of k = 2 until the k is determined. The number of groups is selected when a drastic change is proportional to the previous value. The value before the difference is the number of clusters. After the number of sets is determined, the processing begins 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 [9].

Step 1 : Determine the number of clusters with the elbow method
Step 2 : Each data point in the data set is assigned to the nearest centroid to generate a new centroid.
Step 3 : All data points are assigned to the nearest centroid, and then a new group is created.
Step 4 : The process is repeated until the stopping criteria are met.

E. Marketing Strategy Definition

Customer lifetime value (CLV) is one way of defining customer value [7]. The model calculates the distance between the data point and the central cluster [32]. The higher the value, the more loyal the customer is. CLV is calculated based on the CLV rating determined for each segment [33] as in (1).

$$CLV = X_i * W_j + \cdots + X_n * W_n \tag{1}$$

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 AHP [34], which solves complex multi-criteria problems in a hierarchy [33]. It is helpful for integrated and fuzzy issues based on human brain assessment. The step from AHP is described below[3]:

1. Comparing variables based on cluster results.
2. Making a set of pairwise comparison matrices for each lower level with one matrix for each element.
3. The matrix results are required for assessment in each pairwise comparison.
4. Hierarchical synthesis determines the criterion weights taken from all eigenvectors.
5. After making all pairwise comparisons, consistency is determined using the eigenvalues in (2).

$$CI = \frac{\lambda - n}{n-1} \tag{2}$$

where:
CI: Consistency index
λ 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.

Based on the CLV results, we can then determine the appropriate service improvement strategies based on the concept of CRM [12] in Table 5.

TABLE 5
CUSTOMER RELATION STRATEGY

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

IV. RESULTS

The first step of analysis is choosing a combination of potential variables to be processed into a cluster model, in determining the combination of selected variables with a high total variance value. The total variance value is obtained from the sum between dimension 1 and dimension 2. Dimension is the data variance value obtained from the K-Means process with the combination of variables shown in Table 6. The blue colour indicates the variance between power (P), non-peak load (NPL), and peak load (PL) features, leading to the high data variance value of 97.7%. In complement, the dissimilarity between each cluster has an error value of around 2.3%.

TABLE 6
THE COMBINATION OF CLUSTERING VARIABLES

P	FT	TK	NPL	PL	NPLF	PLF	TC	D	DIM1	DIM2	TV
√	√	×	√	√	×	×	×	×	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

Next, we determine the number of clusters (k) using the elbow method. Fig. 4 shows the visualisation of the results. The magnitude of the total within-clusters sums of squared decline radically when we alter the number of clusters (k) from 3 to 4. Therefore, based on the method, the best grouping of the K-Means clustering model in the electricity consumption sector is at cluster number (k) of 3.

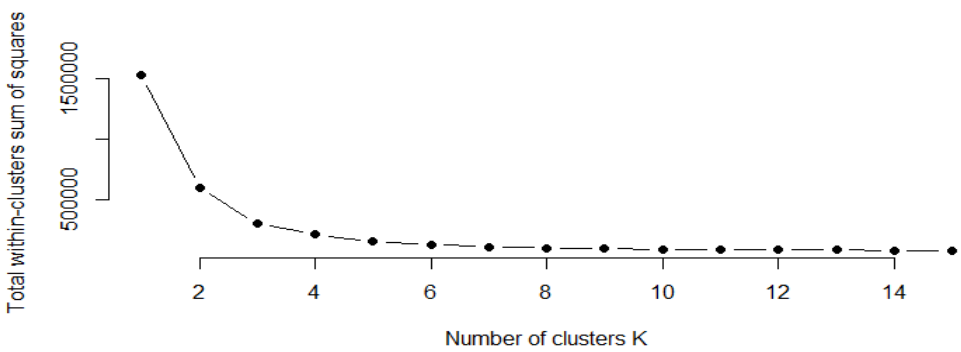

Fig. 4 The number of clusters determination

As shown in Fig. 5, the cluster number (k) of 3 groups customers into distinctive clusters. Using k=4 in Fig. 6, we see that there are outliers (the group with dark purple points) that have an indistinctive boundary with the other cluster (the group with light green points). Thus, we deem that three clusters are indeed in the correct grouping.

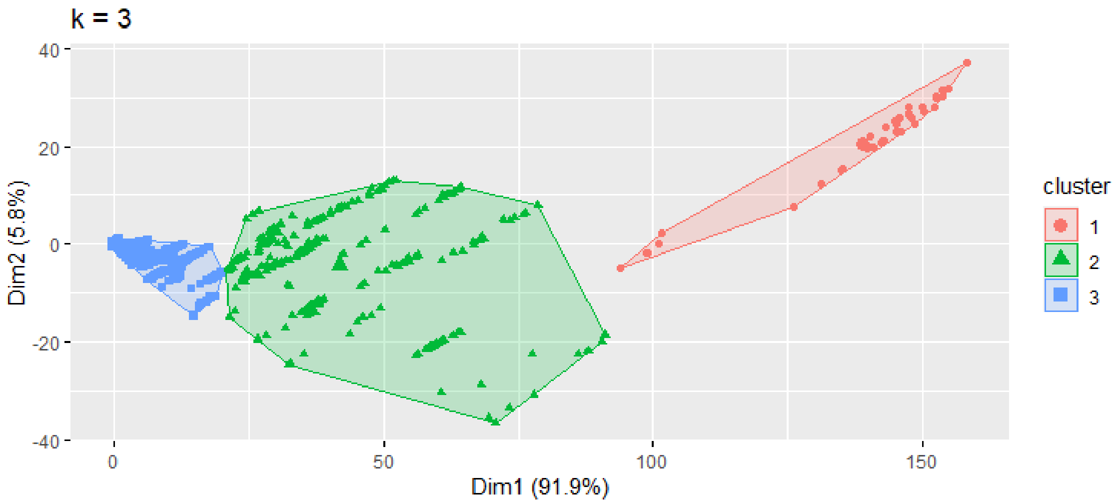

Fig. 5 Cluster visualization (k=3)

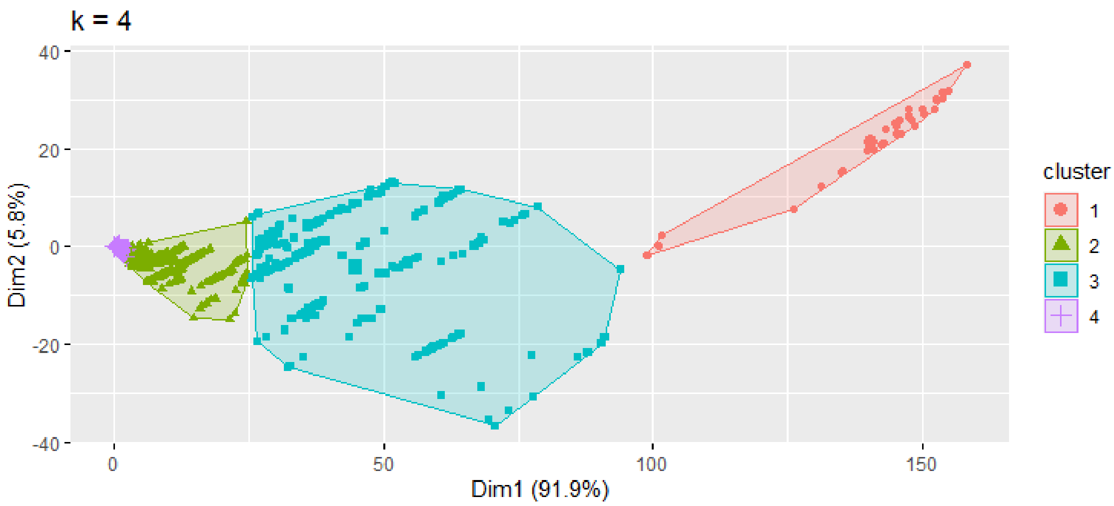

Fig. 6 Cluster visualization (k=4)

Table 7 depicts the details of the three-cluster using K=3. The first cluster represents 282 customers. The centroid of the first cluster is located at the total power use of 937,837 kWh, the peak load of 27,827 kWh, and the total non-peak load of 115,194 kWh, with customers using an installed capacity of above 10,600 kWh. The second cluster represents 508,615 customers. The centroid of the second cluster is located at the total power use of 4,260 kWh, the peak load of 35 kWh, and the non-peak load of 544 kWh, with customers using an installed capacity of below 10,600 kWh. The third cluster represents 37 customers. The centroid of the third cluster is located at the total power use of 2,226,351 kWh, the peak load of 123,297 kWh, and the non-peak load of 390,803 kWh, with customers using an installed capacity above 200,000 kWh.

TABLE 7
THE DETAIL OF THE CLUSTERING RESULTS

Cluster	Number of Customers	Total power (kWh)	Non-Peak Load (kWh)	Peak Load (kWh)	Installed power (kWh)
1	282	937,837	115,194	27,827	11,000 -200,000
2	508,615	4,260	544	35	450- 10600
3	37	2,226,351	390,803	123,297	>200,000

The fourth step is to determine CLV, which is commonly determined by using range, frequency, and monetary (RFM) [32], [34]. In this study, we use the Power, Non-Peak Load, and Peak Load to calculate CLV. Before that, it is necessary to calculate the weight value using the AHP formula (see Equation 2). Table 8 shows the weight value of each variable from the AHP calculation.

TABLE 8
WEIGHT OF AHP RESULTS

Variable	Weight
Power	0.237
Non-Peak Load	0.391
Peak Load	0.712

Next, we calculate the CLV value per group by multiplying the clustering features variable in Table 7 and the AHP weights in Table 8 as seen in (1). Table 9 presents the average CLV estimation for each cluster. NP refers to the standard amount of power. NNPL refers to the usual amount of electricity during non-peak time. NPL refers to the amount during peak time. The height used by the customer is the weighted peak load.

TABLE 9
RESULT OF CUSTOMER LIFETIME VALUE IN EACH CLUSTER

Centroid	Number of Customers	NP	NNPL	NPL	CLV Value
Segment 1	282	222,267.4	45,040.85	19,812.82	287,121
Segment 2	508,615	100.962	212.704	24.9	338.586
Segment 3	37	527,645.2	152,804	877,787.46	768,236.6

Finally, the values can be ranked after finding CLV in each customer segmentation. Segment 3 came first with a value of 768,236.6. Segment 1 is the second with the value of 287,121, and segment 2 is the third with the value of 338.6. Table 10 presents the customer segmentation rank.

TABLE 10
RESULT OF CUSTOMER RANKING

Segment	Number of Customers	CLV Value	Ranking
1	282	287,121	2
2	508,615	338.586	3
3	37	768,236.6	1

The last step is to define the corresponding marketing strategy based on the clustering results. As depicted in Table 11, we characterise each cluster based on its profitability: marginally, moderately, and highly profitable.

TABLE 11
INSIGHT FROM CRM DECISION DEVELOPMENT

Segment	Number of Customers	Ranking	Strategy Targeting	Programs
1	282	2	Moderately Profitable Customer	Premium Service and Campaign to Customer
2	508,615	3	Marginally Profitable Customer	Partnership Program and Substitute Equipment from non-electrical to electrical
3	37	1	Highly Profitable Customer	Special Premium Service Product and Special Executive Accounts

IV. DISCUSSION

The analysis has shown 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 generate the same output. However, in the visualization of cluster 3, the distribution is more precise. The combination of K-Means and CLV results found three different customer segments. Segment 1 has 282 business customers with a total capacity of 938,837 kWh, peak load usage of 27,827 kWh, and non-peak load usage of 115,194 kWh. Segment 2 has 508,615 business customers with a total capacity of 4,260 kWh, a peak load of 35 kWh, and a non-peak load of 544 kWh. Segment 3 has 37 business customers with a total capacity of 2,226,351 kWh, a peak load of 123.297 kWh, and a non-peak load of 390,803.

Two marketing strategies can be used, i.e., sustainable marketing and one-on-one marketing. The profitable customer group can be maintained using sustainable marketing, namely business-to-business, by offering premium service products to use more electricity during non-peak periods. One-to-one marketing can be provided via special account executives to provide the best solutions and consultations for electrical problems experienced by customers. For the moderately profitable group, sustainable business-to-business marketing can be done by offering premium services without compromising consumption habits during peak seasons. One-to-one marketing also applies to increase electricity use during the non-peak load period.

For the marginally profitable group, we propose a continuous replenishment program because the monthly electricity consumption is 4,260 kWh. 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 maximise 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 [2], [9], [10], [12].

V. CONCLUSIONS

Understanding electricity consumption patterns are critical to effectively managing increasing electricity demand. This study presents a hybrid customer segmentation model by combining K-Means clustering, customer lifetime value concept, analytic hierarchy process, and marketing strategy in CRM. We evaluate the number of clusters with the elbow method but choose K = 3 due to the visualisation of the correct cluster distribution. In the proper ranking, CLV values arto order to implement a marketing strategy according to customer criteria.

These findings can inform the company to provide more optimal strength based on the characteristics of its customers. In addition, this research helps companies improve their targeting strategies for their customers and revenue accordingly. In terms of the contribution of the literature, this study presents a predictive model using segmentation or customer grouping based on electricity consumption used by business customers. This model can reflect customer behaviour towards consuming electricity. In most cases, the characteristics of individual customers show a positive or negative relationship, with each class showing a different electricity consumption pattern. However, this study only focuses on business customers and combines K-Means clustering with the CRM concept and CLV. Future studies may explore other clustering methods in different contexts.

Author Contributions: *Radit Rahmadhan*: Conceptualization, Software, Investigation, Methodology, Writing - Original Draft, Data Collection. *Meditya Wasesa*: Supervision, Methodology, Writing-Original Draft, Writing - Review & Editing.

Funding: This research received no specific grant from any funding agency.

Acknowledgements: The authors would like to thank PT. PLN West Sumatra Area for the data access.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

[1] Katadata, "National Electricity Consumption Continues to Increase," *www.databook.com*, Jan. 09, 2020. https://databoks.katadata.co.id/datapublish/2020/01/10/konsumsi-listrik-nasional-terus-meningkat (accessed Jan. 04, 2022).

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

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

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

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

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

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

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

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

[10] H. Li, X. Yang, Y. Xia, L. Zheng, G. Yang, and P. Lv, "K-LRFMD: Method of Customer Value Segmentation in Shared Transportation Filed Based on Improved K-means Algorithm," in *Journal of Physics: Conference Series*, Jul. 2018, vol. 1060, no. 1. doi: 10.1088/1742-6596/1060/1/012012.

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

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

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

[14] F. Abdi and S. Abolmakarem, "Customer Behavior Mining Framework (CBMF) using clustering and classification techniques," *Journal of Industrial Engineering International*, vol. 15, pp. 1–18, Dec. 2019, doi: 10.1007/s40092-018-0285-3.

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

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

[17] K. C. Rao, S. Velidandla, C. L. Scott, and P. Drechsel, "Business Models for Fecal Sludge Management in India," *Resource Recovery & Reuse Series*, vol. 18, 2020.

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

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

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

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

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

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

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

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

[26] Q. Yan, C. Qin, M. Nie, and L. Yang, "Forecasting the Electricity Demand and Market Shares in Retail Electricity Market Based on System Dynamics and Markov Chain," *Math Probl Eng*, vol. 2018, 2018, doi: 10.1155/2018/4671850.

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

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

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

[30] G. Shmueli and O. R. Koppius, "Predictive Analytics in Information Systems," 2011.

[31] R. Bapna, P. Goes, A. Gupta, and Y. Jin, "User Heterogeneity and Its Impact on Electronic Auction Market Design: An Empirical Exploration," 2004. [Online]. Available: http://www.jstor.orgStableURL:http://www.jstor.org/stable/25148623

[32] A. Hosseini and R. Hosseini, "Model selection for count timeseries with applications in forecasting number of trips in bike-sharing systems and its volatility," Nov. 2020, [Online]. Available: http://arxiv.org/abs/2011.08389

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

[34] A. Hosseini, M. FallahNezhad, Y. ZareMehrjardi, and R. Hosseini, “Seasonal Autoregressive Models for Estimating the Probability of Frost in Rafsanjan,” *Journal of Nuts*, vol. 3, no. 2, pp. 45–52, 2012.

Publisher’s Note: Publisher stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.