**Prediction Model of Customer Segmentation in Customer Relationship Management Electric Service Users Using Machine Learning**

Radit Rahmadhan1 and Meditya Wasesa2

1Institut Teknologi Bandung

**Abstract.**

The increasing number of electricity users in Indonesia does not necessarily mean positive growth for the only electricity provider in Indonesia. Therefore, understanding customer segmentation and customer preferences is very important to increase customer satisfaction (ie PT. PLN Persero customers per customer). In response, we present new insights into power user customer segmentation and preferences using customer relationship management (CRM) with the help of Key Account Marketing (KAM). We use PT. PLN Persero's consumer data, from 2019 to 2020, operate three machine learning for classification (Decision Tree, Random Forest Tree, Logistic Regression) and compare it with AUC ROC model to determine the segmentation model and customer preferences We propose four dominant customer segments and characterizing customer preferences for a single electricity provider in Indonesia. Finally, we offer a new framework using Customer relationship management (CRM) with the help of Key Account Marketing (KAM) to predict customer segmentation.

**Keyword**: Customer Relationship Management, Machine Learning, Key Account Marketing,

Prediction

# Introduction

The increasing use of electricity in Indonesia is increasing every year [1]. Based on data taken from katadata.co.id [1], the increase in the number of electricity users in Indonesia from 2010 to 2020 is 33.25 percent [1]. PT. PLN Persero as the only provider of electricity in Indonesia must provide electricity supply to the people of Indonesia in large quantities. However, PLN is having difficulties because electricity supply to remote areas is still limited. Based on information obtained from NewsSetup in the next 10 years the government plans to open the door for the private sector to enter the electricity transmission business. The government does numbers that aim to ease the financial burden of the PLN company [2]. Based on these problems, PLN is threatened because its customers can switch to using private electricity. According to Chiang, 2018 customer segmentation refers to the process of grouping customers into more specific ones to predict future customer actions or behavior. Based on research conducted by Andaleeb, the 2015 Customer Relationship Management (CRM) model is one way to better understand Customer Segmentation, from the customer segmentation model. It groups customers based on predetermined variables that aim to predict the amount of power that will be used, the meter used, the time of electricity consumption used by the customer [3].

Based on what PLN explained earlier as the sole electricity provider company in Indonesia, it is necessary to understand how to apply the customer segmentation model in customer relationship management. (CRM) is used to predict future customer actions or behavior based on the facilities provided by PLN that are used by customers. Customer Relationship Management (CRM) is used to predict customer satisfaction by understanding customer behavior, customer loyalty and customer feedback to the company. which aims to improve performance, attract customer interested and increase company profitability. However, the Customer Relationship Management model is not enough to understand customer segmentation, there is a new model that has not been widely used by previous researchers, namely the Key Account Marketing (KAM) model. Key Account Marketing (KAM) is a model that can also understand customer segmentation with a top-to-bottom approach, while Customer Relationship (CRM) itself is a bottom-up approach [5].

Based on previous research, Key Account Marketing (KAM) is used to increase wider sales and build relationships and increase customer commitment to build a strong business. This study creates a new model by combining the Customer Relationship Management model with Key Account Marketing to understand customer segmentation which aims to predict actions or actions. Behavior that will be carried out by customers in the future so that companies such as PLN can make companies more familiar with the characteristics of their customers. The dataset used in this study is PLN customer transaction data from 2019 to 2020. The research will be carried out by making a customer classification which will be divided into three customer classifications grouped based on the rental area, the power used by the customer. customers, service units available in the area, payments made by customers such as manual or electronic.

The data will be processed with several machine learning models, the models we use to group customers or customer segmentation such as K-Means Clustering, K-Nearest Neighbors, we will compare which model is faster in predicting customer segmentation in the incorporation of the customer framework. relationship management with key account marketing, using machine learning. Can help predictions in the model to further clarify customer segmentation more quickly and accurately, then can help companies, namely PLN, by innovating what actions will be taken in the future to retain customers such as adding electricity supply to remote areas because in the future there must be housing to be built, then customers who want to increase their business power such as MSMEs that require large amounts of electricity. We wanted to develop a predictive model by combining Customer Relationship Management (CRM) and Key Account Marketing to make it more effective.

The following will be a research question to guide the research process:

1. What is the impact of using a combination of Customer Relationship Management and Key Account Marketing methods in measuring customer segmentation?

2. How effective is the implementation of Key Account Marketing in Customer Relationship Management?

This research aims to find out how effective Customer Relationship Management is to Key Account Marketing and PT. PLN Persero advantages of applying combination methods. The following will act as a research objective:

1. Create a new framework by adding Key Account Marketing in predicting customer segmentation.

2. Increase effectiveness or speed in predicting Customer Relationship Management by adding Key Account Marketing.

This paper is structured as follows. Part 2 presents a literature review of previous research on Customer Relationship Management (CRM) using machine learning, Previous research on Key Account Marketing (KAM), Part 3 describes the research framework including data collection and research design. Part 4 presents the results of the study. The final section provides conclusions, offering future research potential and current limitations.

# Literature Review

There are a large number of studies related to Customer Relationship Management (CRM) that overcome the limitations of their data sets and propose various algorithms and techniques to identify customer segmentation of a particular data set. The following subsections discuss the related work of the most important challenges that researchers face with their respective proposed solutions.

## 2.1. Customer Segmentation in Customer Relationship Management using Machine Learning

Segmentation can be seen as a simplification of the messy complexity of dealing with many individual customers, each with different needs and potential values. Traditional customer segmentation methods are generally based on experience classification methods or simple statistical methods. Traditional statistical methods classify customers according to simple behavioral characteristics or attribute characteristics such as the category of product purchased or the region where they live.[6] This segmentation method cannot perform a more complex analysis that what kind of customers have high potential value and what kind of customers have credit. tall. With extensive implementation of EC and CRM, companies have been collecting more and more customer data. Traditional techniques such as multiple regression cannot cope with this level of complexity. As a result, the reliability and validity of the statistical functions used to generate segmentation or to build predictive models are factors that might contribute to CRM user dissatisfaction [7]

Data mining can be considered as a recently developed methodology and technology, becoming famous in 1994. The Sas Institute defines data mining as the process of selecting, exploring, and modeling large amounts of data to uncover previously unknown data patterns. Thus, data mining can be thought of as a process and technology for detecting previously unknown things to gain a competitive advantage. Data mining uses neural networks, decision trees, link analysis, and association analysis to find useful trends and patterns from the extracted data [6]. Data mining can yield important insights including predictive models and associations that can help companies understand their customers better. Many large companies today have terabytes of data, where they may be able to find more information about customers, markets, and competition than they need. Data mining enables marketers to better extract valuable business information from the 'data mountains' in enterprise systems. It's a potential solution to a big problem many companies face: data abundance and the relative scarcity of staff, technology, and time to change. Numbers and notes become meaningful. information about existing and potential customers. Data mining allows companies to measure consumer behavior based on 100 or more attributes instead of the three or four associated with traditional statistical modeling. The more attributes a company uses, the greater the complexity of the data and the greater the need for data mining tools. As practitioners enthusiastically searched for a profitable group of customers whose loyalties were stable, some academics began to question whether the segment was really an entity. A stable and much more fundamental whether they really exist [7].

The segmentation method based on data mining created by this paper can solve the above problems because the model can learn new information that is entered later. and get new rules. It provides full support for dynamic management processes in acquiring customers, retaining customer, and increasing customer value, customer satisfaction and Promoting customer loyalty. Establishing a mapping relationship between the attributes of the conception and the customer is the key step of the segmentation method based on data mining [18]. Customer data contains dispersive and advanced attributes. Assigning each customer attribute as a dimension and assigning each customer as a particle, all customers in the company can form a multidimensional space, which has been defined as a customer attribute space. Mapping relationships between customer attributes and conceptual categories can be built using analytical methods, or with sample learning methods. The analysis method analyzes the attribute characters of each conception category that must be possessed, then forms a mapping of the relationship between the attribute space and the conception space [12]. However, many mapping relationships between attribute space and conception space are not clear, it is necessary to use sample learning methods to build mapping relationships [11]. The sample learning method automatically generalizes the mapping relationship between attribute space and conception space by applying data mining technology to the same conceptual category. known in the company database. The data mining process is called sample learning [10].

In Vucetic research, 2018 says that in customer segmentation the goal is to group customers in the feature space taking into account defined and potentially incomplete product preferences, so that the preferences of instances in one cluster are more similar than the preferences of customers in other clusters [13]. heuristic basis for this application that uses well-known algorithms such as K-means, and proposes a principled algorithm designed specifically for this type of clustering. It is based on the Plackett-Luce (PL) probabilistic ranking model. Each cluster is represented as a composite of Voronoi cells defined by a set of prototypes and assigned a set of PL label scores that rank the cluster-specific labels. Unknown PL cluster parameters and prototype positions were determined using supervised learning techniques. Cluster membership and rank for a new instance is determined by the membership of its closest prototype. The proposed algorithm is evaluated empirically on synthetic and real-life label ranking data. The OT-based method is superior to the heuristic-based supervised clustering approach. The proposed PL-based algorithm was also evaluated on the label rank prediction task. The results show that it is highly competitive with sophisticated label ranking algorithms, and highly accurate on partially ranked data [13]. Additional applications include: meta-learning [24], where, given a new data set, the task is to induce total rankings. available algorithms according to their suitability based on the properties of the data set; predict food preferences for new customers based on survey results, demographics, and other respondent characteristics [17]; determine the order of questions in a survey for a specific user based on the respondent's attributes. See [8] for an overview of label ranking applications in economics, operations research, and databases. Supervised clustering of label rank data is an open and non-trivial problem, which has not been discussed in the data mining literature. Traditionally, clustering techniques are unsupervised, and thus do not consider class membership (classification) or target value (regression). Supervised grouping [9, 10, 11], on the other hand, was not. It aims to generate the desired grouping with additional information (e.g., class labels). The first baseline approach uses the well-known K-means algorithm, either by treating the label rating as one of the features, or by first grouping based on the instant features only and then assigning a rating labels to clusters were obtained using the generalized Mallow model [18]. The first two basic approaches group label ratings [19], regardless of instance features, to obtain a predefined number of classes. It then trains multi-class classifiers using the newly formed classification data.

The proposed Plackett-Luce Mixed Model (MMPL) presents a general framework for label ranking that can be used for both supervised clustering and for prediction. It is based on a multi-prototype cluster representation, where the underlying cluster preferences are modeled using the cluster-specific Plackett-Luce score parameter. The model is fast, with linear training and predictive time, constant memory scaling with number of prototypes and able to work efficiently with incomplete ratings. K-means! Mallows algorithm first performs K-means clustering based on feature instances only, without taking into account the label rank. It then derives the k center rank for each cluster from the label rank to which this cluster belongs using the Mallows Model. This approach is expected to work well when the clusters in the feature space match the clusters in the label space. Naive K-means In this approach, label rank is treated as a feature. After this preprocessing stage, the newly formed data in the cluster uses the K-means benchmark algorithm. The k center rank for each cluster is derived from the cluster centroid obtained by sorting the last L attributes in ascending order. This ranking is used to predict the label rank of a new instance when its cluster membership is determined by finding the nearest cluster centroid in the original feature space.

In Aziz's research, 2019 customer segmentation in the Small Medium Enterprise model used the clustering method. Clustering is also referred to as data segmentation in some applications because the grouping of large data into groups has something in common with the group.[18] Clustering is the process of grouping physical and abstract datasets of objects into groups of objects that have similarities. [19], [20]. Clustering is a method that is widely used in various fields, including customer segmentation, customer behavior, customer profitability, forest fire data collection and so on. Various algorithms are used in clustering: K-Means, DBSCAN (Density Based Spatial Clustering of Application with Noise), SOM (Self Organizing Map) and others. The most commonly used method is the K-Means algorithm. [5].

However, K-Means clustering has a weakness in the accuracy of determining the number of clusters. [21], [5], [13], [12]. Many researchers have examined validation methods in determining the best number of clusters. One good method for determining the number of clusters is the Elbow Method. [22]. To determine the number of clusters using the Elbow method. This method is used in cluster analysis for interpretation and performance testing of the consistency level of the right number of clusters by looking at the SSE value. [23]. At some point the graph will drop drastically with a curve called the angled criterion. This value then becomes the best value of k or the number of clusters. [23].

Therefore, this study aims to generate Customer Lifetime Value (CLV) with the K-Means clustering algorithm based on the LRFM (Length, Recency, Frequency and Monetary) model. The dataset used as an experiment is SME data selling pulses of all operators for the period January 1, 2018 to June 30, 2018. Before the clustering process is carried out, normalization and transformation are carried out to produce a dataset that is ready for the clustering process. Determination of the best number of clusters is calculated by the elbow method. Then the CLV results from the weighting of the LFRM data and the analysis and ranking process.

The following is table 1 of the literature review that has been collected from previous research.

Table 1

Reviewed Studies on Customer Segmentation

|  |  |  |  |
| --- | --- | --- | --- |
| Paper Identifier | Features | Business Context | Techniques |
| Vucetic et al,2018 | Restaurant customer feedback data, The variables used to predict new customers are customer name, type of food ordered, food review, gender | Marketing | K-Means using PL based algorithm |
| Aziz et al, 2019 | SME customers are all pulse server operators AR-Pulsabiz Malang Indonesia. A total of 141 customers in the analysis period January 1, 2018 to June 30, 2018 The data will first be processed by query filtering in the form of preprocessing and transformation. The variables used are ID, Lenght, Recency, Frequency, Monetary | Small Medium Enterprise | K-Means Clustering and LRFM Model |
| Ye Jingyi, 2021 | This E-commerce data comes from the University of California, Irvine's Machine Learning Repository called Online Retail Data Set. This is a transnational dataset containing transactions that occurred between 12 January 2010 and 12 September 2011 from UK based and registered non-store online retailers and companies primarily selling gifts for all occasions There are a total of 541,909 cases and 8 attributes in the data set where each The row displays customer information and product information from one product from 2010 to 2011. Variables used are Invoice No., Stock Code, Description, Quantity, Invoice Date, Unit Price, Customer ID, Country, Total Price | E- Commerce | K-Means Clustering |
| Sano et al, 2021 | Supermarket data containing 4376 customer data, 105,170 transaction log data, 11,171 product data from datasets from January 1, 2017 to December 31, 2018. The variables used are ID,Product Name, Lenght, Recency, Frequency, Monetary | Marketing | AEF, RFM, k-means, Ward method, FCM, and the decision tree |
| Antony et al,2019 | The real-life dataset, which contained the sales data of a pharmacy in Palembang. It consist a transactional records between January and December 2015. The dataset contained 2.043 products, 399.738 sales transactions and 3.956.683 products sold. The variables used are ID,Product Name, Lenght, Recency, Frequency, Monetary | RFM Analysis | K-Means Method and eight indexes of validity to determine the optimal number of clusters namely Elbow Method, Silhouette Index, Calinski-Harabasz Index, Davies-Bouldin Index, Ratkowski Index, Hubert Index, Ball-Hall Index, and Krzanowski-Lai Index |
| Puh et al, 2020 | Questionnaire data consisting of 500 consumers in Croatia by applying a convenience sample through a highly structured questionnaire. The variables used are Demographic characteristics (Age, Gender, Education, Occupation, Monthly Income in HRK), Product ,Frequency,Percentage | Food Retailing | Latent Class Model |
| Abdi et al, 2018 | Customers of a telecom company. The dataset consists of 25 variables, with 24 predictor variables and 1 target variable. It consists of 1000 instances with 274 records labeled churned and 726 non-churned ones. The variables are divided into two groups: socio-demographic and Behavioral. The variables used are Socio-demographic attributes (Region, Age, Marital, Address, Income, Education, Employment, Retire, Gender), Behavioral Attributes (Hours of Usage (Longmon,Tollmon, Equipmon, Cardmon,Wiremon), Selected Service (Multiline, Voice, Pager, Internet, Callid, Callwait, Forward, Confer, Callcard, Wirelles, Churn)) | Telecom Company | K-Means Clustering, Neural network, |
| McLoughlin et al, 2014 | Experimental data by installing smart meters to more than 4000 residences in Ireland, from 2009 and 2010. The data used is information about the characteristics of the dwelling, occupants and equipment used in the house. The variable used are Dwelling type, No. of bedrooms,Age, Social Class, Electronic Type | Electricity | k-means, k-medoid and Self Organising Maps (SOM) |
| Li et al, 2012 | Historical data from the vehicle sharing platform database at the university detailed data of all customers from 30 November 2015 to 30 November 2017 was retrieved based on the Last Ride Time from the last driving date, for a total of 50715 records. Which contains the user ID, driving mileage, points, discounts and 29 other attributes. The Variable are used User id, current miles,cost, car id | Transportation | K-Means Method and KLRFMD model |
| Marisa et al, 2019 | The dataset is extracted from data from one of the SMEs engaged in the sale of electric pulses.The variable used are Name Customer, Average of transaction/ week, Payment System | Small Medium Enterprise | K-Means Clustering |
| Chindyana et al ,2021 | The dataset using google review rating dataset which contains 5456 tourist data. The Variable used are Id Customer, Gender, Place, Review | Tourism | K-Means Method and DBSCAN Method |
| Zhao et al,2021 | Using real data as an example is regarding the online customer segmentation of a company in the UK. This example was studied by Chen et al. [30]. The customer transaction dataset contains: 11 variables and a total of 22190 transaction records valid for approximately one year. 11 variables contain information  such as invoice number, quantity, price, address and zip code. | E-commerce | K-Means Clustering and RFM Model |

## 2.2 Key Account Marketing

Based on previous research conducted by (Hult 2019) Key Account Marketing is used for marketing companies or groups that have reached the limit and an idiosyncratic management approach in managing certain customers to their customers, namely loyal customers. These customers are critical to a company's future development, for example, because they represent a tremendous growth opportunity (Davies & Ryals, 2019; Homburg, Workman, & Jensen, 2002) or because working with customers allows the supplying company to produce more. products (Hakanen, 2019). According to (Ahmmed & Noor, 2019), Key Account Marketing is an approach taken by supplier companies that targets customer loyalty for various needs. Complex with special treatment aimed at the interests of both parties. There are four keys contained in the marketing key account are Earn, Save, Grow, Win Back. The four keys that have been mentioned are closely related to the concept of customer relationship management. The approach of Customer Relationship Management from top to bottom, while Key Account Marketing is the opposite.The literature on Key Account Marketing is still limited. The researchers developed Key Account Marketing by applying a conceptual framework to conceptualize and develop and test hypotheses. (Hunt, 1983, p., 10). Based on previous research, no one has combined the concept of Customer Relationship Management with Key Account Marketing and in previous studies no one has used data analytic methods or machine learning in Key Account Marketing. The following is table 1 of the literature review that has been collected from previous research.

Table 2

Reviewed Studies on Key Account Marketing

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Article | Model | Forecasted Value | Methods | Business Context |
| Health, 2011 | KAM | Company Revenue | Logistic Regression | Hotel |
| Wang & Brennan, 2014 | KAM | Employee Peformance | Linear Regression | Company |
| Ivens et al., 2018 | KAM | Employee Peformance | Linear Regression | Company |
| Ahmmed & Noor, 2018 | KAM | B2B | Logistic Regression | Company |

# Method

Our study focuses on segmenting customers using PLN data for the West Sumatra zone by conducting a comprehensive comparison between Logistic Regression, Decision Tree and Random Forest Tree. All three algorithms have the same input data with the same ratio of training and testing. The input data for the three algorithms is processed through filters to reduce noise and remove unwanted data. After that, the cleaned data was divided into training and test sets. The training set is then modeled using an algorithm to give the desired output. The following subsections will provide more details on this process.

Figure 1 shows the framework in this study. This was adapted from the standard method for constructing predictive analytical models [26]. There are five stages: data collection; selecting relevant predictor variables, determining appropriate customer segmentation, determining potential prediction methods, evaluating, validating, and selecting the best

predictive model, and finally reporting the research results.



Figure 1 Research Framework

**3.1 Data Collection**

In this study, we used data from PT. PLN Persero. The data taken is only 1 area because the fields for each region are the same, the data taken by PLN is the West Sumatra zone. Our research uses customer transaction data from January 2019 to December 2020 which consists of 16,811,662 and 90 data variables in 2 years. Table 2 shows descriptive statistics for the data. Several records were removed from the data set because they showed illogical conclusions i.e. duplicate records or missing values.

Table 3

Descriptive Statistics of The Datasets

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Data** | **Attribute** | **Statistics** | **Raw Data** | **Filtered Data** |
| Customer transactions 2019 | Number of Records  Period | Count  Min  Max | 8,005,831  1/1/2019  31/12/2019 | -  1/3/2019  20/12/2019 |
| customer transactions 2020 | Number of Records  Period | Count  Min  Max | 8,705,831  1/1/2020  31/12/2020 | -  5/7/2020  28/12/2020 |

* 1. **Choice of Variable**

Predictor variables are based on data obtained from PLN Zone of West Sumatra. There are about 90 variables but not all of them are used because the variables will be determined by two models, namely CRM and KAM. Therefore, the selected variable has a high potential from the 2 models. Tables 4 and 5 show detailed information about the predictor variables.

Table 4

Choice of Predictor Variable CRM

|  |  |
| --- | --- |
| **Variable / Data Types** | **Variable Description** |
| Rate | The price of electricity per month |
| Month | Date of being a customer |
| Payment Code | Paid payment method |
| Early Reading Rate | Initial reading time |
| End Reading Rate | Final reading time |

Table 5

Choice of Predictor Variable KAM

|  |  |
| --- | --- |
| **Variable / Data Types** | **Variable Description** |
| Peak Load Electricity Consumption | Power consumed outside of peak load times |
| Beginning of peak load time | First Crucial time of use of electricity |
| End of Peak Load Time | Last Crucial time of use of electricity |
| Peak Load Lighting Payment | The payment that must be paid for using electricity |
| Off peak Time Payment | Payments that must be paid for using electricity outside a predetermined load |
| Payments Of Street Lamp Lighting Expenses | Payment to be paid while the street lighting is being operated |
| Transformer Rental Fee | The payment that must be paid is stated in the bill according to the agreement that has been signed with a stamp duty |
| Electricity Payment Tax | The agreed price for using the electricity |

* 1. **Choice of Customer Segmentation**

This study aims to determine customer segmentation from pre-selected variables. Customer segmentation is divided into 3, namely Home, Medium Industry, and Large Industry. The classification is based on the pre-selected CRM and KAM variables, namely based on the power used, payment methods, hours used by customers and others. Table 6 shows the customer segmentation.

Table 6

Customer Segmentation

|  |  |
| --- | --- |
| **Segmentation** | **Classification** |
| House | Rate, Month, Payment Code, Early Payment Code, End Reading Rate, Payments Of Street Lamp Lighting Expenses, Electricity Payment Tax |
| Small Industry | Rate, Month, Payment Code, Early Payment Code, End Reading Rate, Peak Load Electricity Consumption, Beginning of peak load time, End of Peak Load Time, Peak Load Lighting Payment ,Payments Of Street Lamp Lighting Expenses, Electricity Payment Tax |
| Big Industry | Rate, Month, Payment Code, Early Payment Code, End Reading Rate method, Payments Of Street Lamp Lighting Expenses, Transformer Rental Fee, Peak Load Electricity Consumption, Beginning of peak load time, End of Peak Load Time, Peak Load Lighting Payment, Electricity Payment Tax |

* 1. **Choice of Potential Method**

The focus of the research is to develop a customer segmentation prediction model from a combination of Customer Relationship Management (CRM) and Key Account Marketing (KAM) methods with coefficients and standard errors that can accurately predict whether customer satisfaction affects the company. Table 4 shows the prediction model that will be used in this study. Seeing the many built-in models in machine learning models, the researchers examined the ensemble model and the single model. In general, the ensemble model is more accurate in predicting than the single model [27]. However, the single model still outperformed the ensemble model. Researchers use K-Means and K-Nearest Neighbors.

**3.4.1 K-Means**

Generally, K-means is one of the well-known unsupervised learning techniques for cluster analysis. Cluster analysis is used to aggregate or divide a data set into several clusters according to similarity values. For k-means, it is necessary to determine the number of clusters (k). It starts with randomly generated centroids and iteratively computes new centroids to converge to the last cluster. There are four steps in k-means [26].

Step 1: The position of k centroids is generated randomly.

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 the new cluster by assigning all data points to the nearest centroid, and then a new cluster is created.

Step 4: The process will be repeated between step 2 and step 3 until the stopping criteria have been met.

**3.4.2 K- Nearest Neighbors**

The K-Nearest Neighbor algorithm is a classification algorithm based on the nearest neighbor. The example above is just a very simple example of implementing this algorithm.

If in other cases there are more than 2 independent variables, to calculate the distance we can use the Euclidean Distance formula. Similar to Pythagoras, only Euclidean Distance has more than 2 dimensions. [31]:

A picture containing diagram

Description automatically generated

* 1. **Validity Index for Determining the Optimal Number of Clusters in the k-Means Method**

Elbow method (EM) [16] is a method used to determine the optimal number of clusters, by looking at the percentage comparison between the number of clusters that will form an angle on the curve. If the value of the first cluster with the value of the second cluster forms an angle (angled) on the curve or the largest decreasing value, the cluster value is the best cluster value. The best number of 'k' clusters will be selected at that point (turning point). This method is a visual method that looks at the total intra-cluster variation or the total Within-Clusters Sum of Squares (WSS) as a function of the number of clusters. The larger the number of clusters k, the smaller the AMPL value or vice versa.

The Calinski-Harabasz Index (CHI) [19] evaluates the validity of the cluster based on the calculation of Between-Clusters Sum of Square (BSS) and WSS. CHI measures the separation ratio based on the maximum distance between the centroids and measures compactness based on the sum of the distances between each data and the centroid. A compact and well-separated cluster configuration is expected to have high inter-cluster variance and relatively low intra-cluster variance [20, 21].

* 1. **Model Use and Reporting**

They compared the model development time and prediction scores based on the performance of each model. The best predictive model with the best predictive AUC-ROC score will help decision makers in formulating the right combination of CRM and KAM in a better way.

# References

Adams, P., Prakobtham, S., Limpattaracharoen, C., Suebtrakul, S., Vutikes, P., Khusmith, S., Wilairatana, P., Adams, P., & Kaewkungwal, J. (2017). Ethical issues of informed consent in malaria research proposals submitted to a research ethics committee in Thailand: A retrospective document review. *BMC Medical Ethics*, *18*(1). https://doi.org/10.1186/s12910-017-0210-0

Ahmmed, K., & Noor, N. A. M. (2018). Managing key accounts in the readymade garments industry: To what extent Bangladeshi companies perform. *Journal of Asian Finance, Economics, and Business*, *5*(4), 57–65. https://doi.org/10.13106/jafeb.2018.vol5.no4.57

Ahuja, V., & Medury, Y. (2010). Corporate blogs as e-CUSTOMER RELATIONSHIP MANAGEMENT tools - Building consumer engagement through content management. *Journal of Database Marketing and Customer Strategy Management*, *17*(2), 91–105. https://doi.org/10.1057/dbm.2010.8

AlHarbi, A., Heavin, C., & Carton, F. (2016). Improving customer-oriented decision-making through the customer interaction approach. *Journal of Decision Systems*, *25*, 50–63. https://doi.org/10.1080/12460125.2016.1187417

Ang, L. (2011). Community relationship management and social media. *Journal of Database Marketing and Customer Strategy Management*, *18*(1), 31–38. <https://doi.org/10.1057/dbm.2011.3>

Bell, E., Bryman, A., & Harley, B. (2018). “Business Research Methods” 2nd edition. Oxford university press.

Craig, C. A. & Allen, M.W. (2013), "Sustainability information sources: employee knowledge, perceptions, and learning," Journal of Communication Management, Vol. **17** No. 4, pp. 292-307. https://doi.org/10.1108/JCOM-05-2012-0035

Chiang, W. Y. (2018). Identifying high-value airline customers for strategies of online marketing systems: An empirical case in Taiwan. *Kybernetes*, *47*(3), 525–538. https://doi.org/10.1108/K-12-2016-0348

Coda, R. C., & de Castro, G. H. C. (2019). Business-to-business marketing: Brazilian scientific production from 2008 to 2018. *RAE Revista de Administracao de Empresas*, *59*(4), 258–270. https://doi.org/10.1590/S0034-759020190404

Demo, G., Rozzett, K., Fogaça, N., & Souza, T. (2018). Development and validation of a customer relationship scale for airline companies. *Brazilian Business Review*, *15*(2), 105–109. https://doi.org/10.15728/bbr.2018.15.2.1

Dolnicar, S., Grabler, K., Grün, B., & Kulnig, A. (2011). Key drivers of airline loyalty. *Tourism Management*, *32*(5), 1020–1026. <https://doi.org/10.1016/j.tourman.2010.08.014>

Dufour, I. F., Richard, M.-C., & Li, J. (2019). Theorizing from secondary

Qualitative data: A comparison of two data analysis methods. *Cogent*

*Education*, *6*(1). <https://doi.org/10.1080/2331186X.2019.1690265>

Etikan, I., Alkassim, R., & Abubakar, S. (2016). Comparison of Snowball Sampling and Sequential Sampling Technique. *Biometrics & Biostatistics Internationa Journal*, *3*(1). https://doi.org/10.15406/bbij.2016.03.00055

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>

Ivens, B. S., Leischnig, A., Pardo, C., & Niersbach, B. (2018). Key account management as a firm capability. *Industrial Marketing Management*, *74*, 39–49. https://doi.org/10.1016/j.indmarman.2017.09.026

Lewis, M. (2005). Research note: A dynamic programming approach to customer relationship pricing. *Management Science*, *51*(6), 986–994. https://doi.org/10.1287/mnsc.1050.0373

Madsen, D. Ø., & Johanson, D. (2016). Examining Customer Relationship Management from a management fashion perspective. *Cogent Business and Management*, *3*(1). https://doi.org/10.1080/23311975.2016.1161285

Nor Azila Mohd Noor Othman Yeop Abdullah, A., & Ahmmed Research Fellow Othman Yeop Abdullah, K. (n.d.). KEY ACCOUNT MANAGEMENT STRATEGY IN BUSINESS-TO-BUSINESS RELATIONSHIP: A PROPOSED RESEARCH FRAMEWORK. *International Journal of Business, Economics, and Law*, *2*, 1.

Ou, Y. C., Verhoef, P. C., & Wiesel, T. (2017). The effects of customer equity drivers on loyalty across services industries and firms. *Journal of the Academy of Marketing Science*, *45*(3), 336–356. https://doi.org/10.1007/s11747-016-0477-6

Rodriguez, M., & Boyer, S. (2020). The impact of mobile Customer Relationship Management (CUSTOMER RELATIONSHIP MANAGEMENT) on sales collaboration and sales performance. *Journal of Marketing Analytics*, *8*(3), 137–148. <https://doi.org/10.1057/s41270-020-00087-3>

Saunders, M., Lewis, P., & Thornhill, A. (2009). Research methods for business students. Pearson education.

Sugiyono.2013.Metode Penelitian Pendidikan.Bandung:Alfabeta.

Taherdoost, H. (2016). Sampling Methods in Research Methodology; How to Choose a Sampling Technique for Research. In*ternational Journal of Academic Research in Management (IJARM)* (Vol. 5, Issue 2). https://ssrn.com/abstract=3205035

Tolga Akçura, M., & Srinivasan, K. (2005). Research note: Customer intimacy and cross-selling strategy. *Management Science*, *51*(6), 1007–1012. https://doi.org/10.1287/mnsc.1050.0390

Tworek, K., & Sałamacha, A. (2019). CUSTOMER RELATIONSHIP MANAGEMENT influence on organizational performance - The moderating role of IT reliability. *Engineering Management in Production and Services*, *11*(3), 96–105. https://doi.org/10.2478/emj-2019-0024

Wang, X. L., & Brennan, R. (2014). A framework for key account management and revenue management integration. *Industrial Marketing Management*, *43*(7), 1172–1181. https://doi.org/10.1016/j.indmarman.2014.06.006

Woodcock, N., Broomfield, N., Downer, G., & Starkey, M. (2011). The evolving data architecture of social Customer Relationship Management. *Journal of Direct, Data, and Digital Marketing Practice*, *12*(3), 249–266. https://doi.org/10.1057/dddmp.2010.45

Yuen, F. T., & Chan, S. L. (2010). System Dynamics Modelling in CUSTOMER RELATIONSHIP MANAGEMENT: Window Fashions Gallery. *International Journal of Engineering Business Management* (Vol. 2, Issue 2).