**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 increase in electricity use 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 was 33.25 percent [1]. PT. PLN Persero as the sole electricity provider in Indonesia must provide electricity supply to the people of Indonesia in large quantities. However, PLN has difficulty because the 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 get into the electricity transmission business. The government conducted the figures that aimed to ease the financial burden of 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 in order to predict future customer actions or behavior. Based on research conducted by Andaleeb, 2015 Customer Relationship Management (CRM) model is one way to understand more about Customer Segmentation, from the customer segmentation model. It grouped customers based on predefined variables that aimed to predict customer satisfaction as well.[3]

Based on what PLN explained earlier as a single electricity provider company in Indonesia is needed to understand how to implement the customer segmentation model in customer relationship manegement. (CRM) is used to predict or predict future customer actions or behavior based on facilities provided by PLN used by customers. Customer Relationship Management (CRM) is used to predict customer satisfaction by understanding customer behavior, customer loyalty and customer feedback to companies. which aims to improve performance, attract customer interest and increase the profitability of the company. But it is not enough with the Customer Relationship Management model to understand customer segmentation, there is a new model that has not been widely used by researchers before, 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 an approach from the bottom. up [5].

Based on previous research Key Account Marketing (KAM) is used to increase broader sales and build relationships and increase customer commitment to establish a strong business. This study create a new model by combining the Customer Relationship Management model with Key Account Marketing to understand customer segmentation that aims to predict actions or actions. The behavior that will be taken by customers in the future so that companies such as PLN can make the company more familiar with the characteristics of its customers. The dataset used in this study data PLN customer transactions from 2019 to 2020. The research will be conducted by making a customer classification that will be divided into three customer classifications that are grouped by the area where the rental, 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 by several models of machine learning, models that KAmi use to classify such as Logistic Regression, Decision Tree and Random Forest Tree, later we will compare which models are faster at predicting customer segmentation in the merger of customer relationship management frameworks with key account marketing, using machine learning. Can help predict in the model to better clarify customer segmentation more quickly and accurately, then can help the company that is PLN by innovating what actions will be taken in the future to keep its 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 add power to their business such as MSMEs that require considerable electricity. We want 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 is a large amount of research related to Customer Relationship Management (CRM) that overcomes the limitations of their data sets and proposes 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 researchers face with their respective proposed solutions. In addition, we will focus our literature review to discuss related work that uses classification techniques namely Logistic Regression, Decision Tree, Random Forest Tree.

## 2.1. Customer Segmentation in Customer Relationship Management

Segmentation can be seen as a simplification of the messy complexity of dealing with multiple individual customers, each with different needs and potential value. Traditional customer segmentation methods are generally based on experience classification methods or simple statistical methods. Traditional statistical methods group customers according to simple behavioral characters or attribute characters such as the category of the purchased product or the region in which it lives.[6] This method of segmentation cannot perform a more complex analysis that what kind of customers have high potential value and what kind of customers have credit. high. With the extensive implementation of EC and CRM, companies have accumulated a growing number of customer data. Traditional techniques such as multiple regressions cannot overcome this level of complexity. As a result, the reliability and validity of statistical functions used to generate segmentation or to build predictive models become possible contributing factors to dissatisfaction. CRM user [7].

Data mining can be thought of as a methodology and technology developed recently, becoming famous in the year l994. 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 extracted data[6]. Data mining can generate 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 to. Data mining allows marketers to better extract valuable business information from the 'mountain data' in a company's systems. This is a potential solution to the big problems many companies face: abundant data and relatively dearths of staff, technology,and time to change. Numbers and records become meaningful. information about existing customers 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 data and the greater the need for data mining tools. As practitioners enthusiastically seek out groups that benefit customers whose loyalties are stable, some academics are beginning to question whether the segment is actually an entity. A stable and many more fundamentals 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 customers and improving customer value, customer satisfaction and Promote customer loyalty. Establishing a mapping relationship between conception attributes and customers is a key step of the segmentation method based on data mining[18]. Customer data contains dispersive and continued attributes. Assigning each customer attribute as a dimension and designating each customer as a particle, the entire customer in the company can form a multidimensional space, which has been defined as the customer attribute space. The mapping relationship between customer attributes and conception categories can be constructed by analytical methods, or by sample learning methods. Analytical methods analyze the attribute character of each category of conception that must be possessed, further establishing a mapping of the relationship between attribute space and conceptionspace[12]. However many mapping relationships between attribute space and conception space are unclear, it is necessary to use sample learning methods to establish mapping relationships[11]. Sample learning methods automatically generalize the mapping relationship between attribute space and conception space by applying data mining technology to the same conception category. known in the company's database. The process of data mining is called sample learning. As for the rules in making customer classification [10].

1) Segmentation rule making

Sort customers by the Customer Segment Model and Segment Functions. After training the segment model, we get segment rules or network segment rules. We can effectively group new customers based on trained models.

2) Function Analysis

Functional analysis includes customer value analysis, credit analysis and promotion analysis. Based on basic mapping of relationships between customers and concepts. Furthermore, the need for new functions will be brought to CRM with the development of management practices. The new demand function will be added to the conceptual dimension and reconstruct the mapping relationship with customer characteristics.

Advantages of making a customer shake[10]

(1) Increase the promotional effect

Customer segmentation based on data mining can help companies to create appropriate promotional strategies, at the right time, with appropriate products and services, aimed at the appropriate customer.

(2) Analyze customer value and customer loyalty Customer value and customer loyalty are important to the company's management strategies and tactics. The company can confirm customer ratings according to expected value and loyalty analyzed with a segmentation model based on mining data.

(3) Analyze credit risk

Risk assessment is an effective way to evaluate a particular type of customer risk, usually the risk of default.

(4) Instruct the R&D of the new product

Companies can find out their customers' preferences by analyzing customers based on data mining, and ensuring that various requests will be realized in the new design.

(5) Confirming the target market

Customer segmentation based on data mining can make the target group of customers clear and find the market explicitly.

The key role of marketing is to identify customers or segments with the greatest potential for value creation and target them successfully with appropriate marketing strategies to reduce risk. These high lifetime value customers defect to competitors [10]. In this mode of construction, customer segmentation is the basic job of data mining according to known historical segmentation information[12]. The training data used to build segment forecast modes can be historical data or exogenous data obtained from experience or surveys. Because customer behavior is uncertain and inconsistent, researchers and managers must build dynamic customer segmentation models to objectively reflect characteristics. In the customer-centric era, customer segmentation results are related to the determination of company strategies and tactics. Best practices require marketers to develop their understanding of customer segmentation based on data mining techniques and use output to develop marketing strategies on a regular basis creative to maximize shareholder value[14].

## 2.2 Machine Learning in Customer Relationship Management Framework

Machine learning and data mining help companies create tools that can create and take action based on customer knowledge and information. Customer information is the basis for maintaining a long-term relationship with customers and is also known as relationship and customer management (CRM). Classification and segmentation of customer data sets is used to maintain efficient relationships with customers and further increase profitability and productivity. In this paper, the authors propose segmentation of customers based on demographic properties such as gender, age and spending scores and analyze data sets for interesting facts. Derivative attribute data sets are investigated for classification. Classification is used to categorize each customer into several classes, namely, 'gold', 'silver', 'elite' and 'occasional'. Comparison of different classification algorithms is simulated with the WEKA tool. Multi-layer perceptron (MLP) was found to be the best classification algorithm with 98.33% accuracy compared to Naïve Bayes, regression and J48.4.

The digital revolution and the increasing amount of data generated by companies/organizations in recent decades have led to great interest in the field of machine learning and deep learning. Technology organizations and companies use Machine L-based predictive analytics to gain an edge over their competitors. The purpose of ML techniques is to find 'hidden' information in data, which is almost impossible to do in traditional ways based on human analytical skills.5 According to (Cioca et al., 2013; Rahman and Khan, 2017). Machine Learning techniques are used to mine data for business intelligence and sellable strategies for customers such as classifying them in different categories, creating promotional scheme strategies, and to improve customer relationship management (CRM) In the current scenario, business processes are becoming increasingly customer-oriented and placed as top management priorities [6] .Due to technological advances in e-commerce, M-commerce, virtual marketing and digital marketing, each product is just a click away from customers. Based on the statement from (Singh and Agrawal, 2019a; Adebiyi et al., 2016) This leads to very intense competition, which is necessary to ensure that consumers receive the highest possible quality standards to maintain them and reduce stirring rates[7]. Crm's focus is to expand customer service and support in customer retention. Customers are very important to every company and organization. For the identification and retention of their target customers, it is very important to have data analysis, which is used to explore valuable insights and trends to know the metrics and nature of the customer. According to (Singh et al., 2018; Yadav et al., 2018it is important to observe the most important key factors influencing a customer's purchasing decision to purchase any product and service[8]. Machine Learnings is one of the popular data analyses that governs the structure of analytical models, which is valuable for growth in purchasing behavior. ML techniques are widely used in customer segmentation prediction, customer life value (CLTV), churning, sales, etc. Based on a statement from (Sgaier et al., 2017) Customer segmentation is useful in understanding what demographic and psychographic subpopulation is inside your customers in business cases and leveraging this information to improve profits, image, value, and inventory management.[9] Companies in any business already realize that gaining new customers is not enough for lasting success and efforts need to be made to identify customer segmentation towards retention. In this paper, segmentation techniques and customer classification for business intelligence analysis purposesare proposed[10]. The demographic property of the customer is taken as a parameter of customer segmentation to know the analytics about the customer. That will help in CRM, efficiency and productivity of shopping malls. Furthermore, machine learning techniques are used to predict customer judgment [9].

## 2.2.1 Customer Relationship Management using Logistic Regression

Recurrence examination is a type of prescient display method that explores the relationship between the needy (target) and autonomous variables [11]. This procedure is used to anticipate, show timing and find causal relationships between these factors [12]. For example, the link between rush driving and the number of road accidents by drivers is best concentrated through relapse. Strategic regression is a machine learning grouping calculation used to forecast the likelihood of clear variables. In strategic relapsing, bound variables are paired variables that contain coded information 1 (really, achievement, and so on.) or 0 (no, disappointed, and so on.). The relapsing model strategically predicts P(Y=1) as component X. Calculated regression is one of the most well-known approaches to adjusting models for clear information, especially for double reaction information in data modeling[12].

## 2.2.2 Customer Relationship Management using Decision Tree

The decision tree is said to be one of the common methodologies, used in predicting as well as estimating customer churn problems. Based on the method of dividing and conquering, decisions were developed.[13] But the decision tree does have some limitations such as not being usable for non-linear and complex relationships between attributes. However, it has been observed that the decision tree method does try to improve classification accuracy [13]. In the JST paper along with the decision tree it is used to perform customer churn predictions and it is known that decision trees outperform neural networks in terms of accuracy[14]. Presented a DT application classification methodology to analyze churn rates in the telecommunications industry. Here, the ID3 decision tree is used and observed that the customer area is one of the main classification features, the other gives two results to the customer for churn using several methods such as K-means clustering, QUEST, CART, Logistic Regression, neural network, exhaustive CHAId[13]. Here, it is observed that CHAID performs much better than the other methodologies mentioned. Note that its accuracy is about 60%, which is much better than other methodologies. In addition, other decision trees do not stand in front for exhaustive CHAID.[14]

## 2.2.3 Customer Relationship Management using Random Forest Tree

Random forests are supervised learning algorithms used for both classification and regression.[15] But in any case, this is mainly used for classification issues. As we know that forests are made up of trees and more trees means stronger forests. Similarly, random forest algorithms make tree decisions on data samples and then get predictions from each and ultimately choose thebest solution throughvoting. This is a better ensemble method than a single decision tree because it reduces over-fitting by flattening the results. It addresses the problem of overfitting by flattening or combining the results of different decision trees. Random forests work well for a large number of data items rather than a single decision tree. Random forests have fewer variances than single decision trees. Random forests are very flexible and have very high accuracy[10]. Scaling data is not required in random forest algorithms. It maintains good accuracy even after providing data without scaling. The Random Forest algorithm maintains good accuracy even as most ofthe data islost.[15] Complexity is the main drawback of random forest algorithms. Random forest construction is much more difficult and time consuming than decision trees. More computing resources are needed to implement the Random Forest algorithm. It's less intuitive if we have a large collection of decision trees[17].

The prediction process using random forests is very time consuming compared to otheralgorithms. Some conventional algorithms such as Decision Tree, Genetics algorithm, neural network and tree classification have been proposed. The above algorithm is able to estimate the churn rate. However, they have some problems such as a decision tree that is lacking with the same class probability problem that can drastically reduce performance[18]. Similarly, in the case of genetic algorithms, it is highly unlikely to recognize the possibilities associated with estimates that result in low performance and in the case of techniques such as state-of-art resulting in some errors. Therefore, the above discussion can conclude that conventional RF techniques do not produce efficient results in terms of large data sets and also perform very poorly when they are unbalanced. Our proposed MRF (Modified Random Forest) method performs better across a variety of parameters such as accuracy and other durability[19]. Our scheme mainly has an additional layer of RV (Random Variables) which helps to perform the model much better along with our methods of assisting in minimizing Gaussian noise and also helps in reducing regression and classification problems. Our methods help in building a variety of different trees from specific training data sets. The proposed method helps in focusing the estimation of consumer churn rates in telecommunication services. In research conducted by Irfan et al the Random Forest Algorithm (RF) worked well with 88.63% of correctly classified instances[20]. Creating an effective retention policy is an important task of CRM to prevent churners. After classification,the proposed model segments the data of moving customers by categorizing customers who quit in groups using cosine similarities to provide retention offers. group-based. The paper also identifies churn factors that are important in determining the root cause of churn[21]. By knowing the significant churn factors of customer data, CRM can increase productivity, recommend relevant promotions to groups of customers who are likely to churn. based on similar patterns of behavior, and excessively enhance the company's marketing campaigns. The proposed churn prediction model is evaluated using metrics, such as accuracy, precision,recall, f-measure, and the receiver's operating characteristic area (ROC).22 The results revealed that our proposed churn prediction model resulted in better churn classification using RF algorithms and customer profiles using k-means clustering. In addition, italso provides the factors behind customer churn churn through rules generated using the selected classification algorithm attributes.[17]

## 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 1

Reviewed Studies on CRM and KAM using Machine Learning

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Article | Model | Forecasted Value | Methods | Business Context |
| Chiang,2018 | CRM | Customer Loyal | Linear Regression | Airplane |
| Health, 2011 | KAM | Company Revenue | Logistic Regression | Hotel |
| Lee et al. l, 2011 | CRM | Customer Loyal | Decision Tree, Logistic Regression | Company |
| Rodriguez & Boyer, 2020 | CRM | Sales Peformance | Linear Regression | Company |
| Tworek, K., & Sałamacha, A. (2019 | CRM | Customer Peformance | Linear Regression | Company |
| Kim & Lee, 2015 | CRM | Consumer Segmentation | Random Forest Tree | Company |
| Coda & de Castro, 2019 | CRM | B2B | Logistic Regression, Decision Tree | Company |
| Madsen & Johanson, 2016 | CRM | Customer Loyal | Cluster Using KKN | Company |
| Harbin et al., 2016) | CRM | Customer Loyal | Random Forest Tree | Company |
| Demo et al., 2018 | CRM | Customer Loyal | Cluster using KKN | Airlines |
| Yuen & Chan, 2018 | CRM | Customer Loyal | Logistic Regression | Company |
| 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 |
| This Study | CRM & KAM | Customer Segmentation | Logistic Regression, Decision Tree random forest Tree | 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 2

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 3 and 4 show detailed information about the predictor variables.

Table 3

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 4

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 5 shows the customer segmentation.

Table 5

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. Looking at the number of built-in models in the machine learning model, the researcher investigated 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 Logistic Regression, Decision Tree, Random Forest Tree. In the ensemble group, the researcher used a random forest tree approach.

**3.4.1 Logistic Regression**

Generally, Logistic Regression is used to describe and test hypotheses [28]. Choosing the correct variable and avoiding highly correlated variables must be considered when using Logistic Regression [29]. The variable predictors in logistic regression can be categorical or numeric, and the target variable for linear regression is binary or dichotomous. Therefore, Logistic Regression cannot predict the target variable for more than two classes. Although Logistic Regression may have some disadvantages, it can often compete with other machine learning techniques, such as neural networks, machine support vectors, random forest, and gradient enhancement. Logistic regression formulation is stated as follows: follows:

Graphical user interface, text, application, email

Description automatically generated

**3.4.2 Decision Tree**

Decision Tree can solve classification problems. As the name suggests, the Decision Tree algorithm looks like a tree structure. It has a root node, a leaf node, and a branch; and some advantages, such as nonparametric, adaptive to any dataset, and can deal with non-linear relationships [30]. Decision Tree is an algorithm used in decision trees [32] and uses the Gini index to evaluate splits. The best score is 0, and the worst score is the same value for each class. The formalization of the Gini index is stated as follows [31]:

Text

Description automatically generated

**3.4.3 Random Forest Tree**

The ensemble method uses a Random Forest for the classifier, which consists of a Decision Tree that is formed randomly and independently of the sample dataset. It uses the law of large numbers, so it is not overfit and can be good for prediction [33]. Furthermore, it can be used for any dataset because it does not require distribution assumptions [34] but the disadvantage is that it can be biased because the sample consists of different compositions of predictive labels [35]. The formalization of the random forest classifier is stated as follows [36]:

Graphical user interface, text, application

Description automatically generated

* 1. **Evaluation, Validation, and Model Selection**

To measure the prediction performance of the predetermined model, the researcher used two different evaluation methods, namely three-way split ten folds and ten-fold cross procedures. In a tenfold three-way data separation procedure, we perform two groupings of data. In the first grouping of data, the researcher divides the data set into ten equal parts or folds.

The dataset is split into ten folds and is not evenly divided. From 16,811,662 records, the researcher grouped the dataset to fold number one to number nine, which consisted of 168,000 records. The number tenfold consists of 193,000 notes. The second grouping is more functional. First, the training set is used to match the data points to the proposed model. Second, the validation set is used to evaluate the most accurate model trained in the training set. The third set of tests is used to generate the final predictive score for each generated model. The number of data records used in training, validation, and test sets was adjusted according to the fold number category. If the test is set to number ten (168,000 records), the training set consists of 1,244,000 (168,000 x 8 times) and the validation set consists of 192,000 records.

If the test is not set to number ten (192,000 records), the training set consists of 1,200,000 records (160,000 x 7-fold + 168,000 records from the previous ten-fold) and the validation set consists of 168,000 records. In total, there are 900 test combinations. In the second procedure, tenfold cross-validation, we split the data into ninefold for training and one fold for testing.

In total, there are ten test combinations. Predictive scores of the evaluated models using ten-fold and tenfold cross-validation procedures were compared. The model with the highest predictive score is selected. In this study, the receiver operating features (ROC) or simplified AUC value were used to determine the prediction score better than accuracy. Graphical user interface, text, application

Description automatically generatedMathematically, the researchers formulated the ABK score as follows:

* 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).