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

Radit Rahmadhan1 and Meditya Wasesa1

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 Master 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, Xgboost) and compare it with Logistic regression 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 Master Account Marketing (KAM) to predict customer segmentation.

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

Prediction

# Introduction

Customer Satisfaction refers to the company's process of providing services to customers. Based on data taken from katadata.co.id, the increase in the number of electricity users in Indonesia from 2010 to 2020 was 33.25 percent. With such a large number of users, of course it requires a very large electricity supply, but PLN is one of the sole providers in Indonesia. experiencing difficulties because the supply of electricity to remote areas is still limited. Based on this information, in the next 10 years the Indonesian government plans to open the door for the private sector to enter the electricity transmission business. This step aims to ease the financial burden of the stun company, namely PLN. This makes PLN threatened because its customers can switch to using private electricity. In this case, PLN needs to understand customer segmentation in CRM which is useful for predicting or predicting customer characteristics in using PLN services, recognizing customer characteristics as PLN's future innovations so that customers are satisfied with PLN services. Customer Relationship Management (CRM) is a method of understanding customer behavior through intense communication with customers to improve performance, attract customers, retain customers, and increase loyalty and profitability [1]. Most of the previous CRM studies predict in terms of customer loyalty and customer satisfaction rarely. Beginning in 2020, Key Account Marketing is a systematic approach to managing and developing customers to achieve maximum value and mutually beneficial results and increase revenue. The Key Account Marketing function accelerates the delivery of information and accelerates service to Customers/Prospective Customers [2]. Key Account Marketing (KAM) is the ultimate goal to increase sales (sales) and more broadly, build relationships and partnerships with customers to establish strong business partnerships. Key Account Marketing can also be said to increase the effectiveness of Customer Relationship Management with the existence of Key Account Marketing which can create a new framework that can predict the extent to which customer satisfaction with the services or products offered by the company can increase company revenues. [3]. The two methods previously described are expected to create a new model that predicts customer satisfaction more quickly and accurately [4]. Machine learning is a tool used to predict customer satisfaction. We wanted to develop a predictive model by combining CRM and KAM to make it more effective. Using this new prediction model is expected to make machine learning predict faster than the model.

## Research Question

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

1. How 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?

## Research Objective

The research study aims to understand how effective Customer Relationship Management is against Key Account Marketing and PT. PLN Persero advantage of applying the combination methods. The following will act as the objectives of the study:

1. To create a new framework by adding Key Account Marketing in predicting customer segmentation.
2. To increase the effectiveness or speed in predicting Customer Relationship Management by adding Key Account Marketing.

# Literature Review

## 2.1 Customer Relationship Management

According to (Payne 2012), it is stated that Customer Relationship Management is a business strategy that implements the management of relationships between companies and customers to maintain those aimed at the prosperity of the company or organization by optimizing the company's ability to find connections between companies or organizations and customers to obtain special meaning. Relational marketing improvises by presenting innovative strategies for marketing concepts, encouraging a move from marketing orientation to customer acquisition (transactional) to focus on customer retention or loyalty (VAVRA, 1993). There are important factors, namely quality, customer service, and aftermarket customer loyalty.

## 2.2 Key Account Marketing

Key Account Marketing, according to (Hult 2011), is a marketing company or group that has reached limits and an idiosyncratic management approach in managing specific customers to its customers, namely loyal customers. These customers are essential to a company's future development, for example, because they represent tremendous growth opportunities (Davies & Ryals, 2014; Homburg, Workman, & Jensen, 2002) or because working closely with customers allows the supplier company to be able to produce more products. (Hakanen, 2014). According to (Ahmmed & Noor, 2012), Key Account Marketing is an approach taken by supplier companies that target customer loyalty for various needs. Complex with special treatment aimed at the benefit of both parties. There are four keys contained in a marketing key account, namely

1. Earn

2. Save

3. Grow

4. Win Back

The four keys that have been mentioned are very closely related to the concept of customer relationship management. Literature on Key Account Marketing is still limited. Researchers develop Key Account Marketing by applying a conceptual framework to conceptualize and develop and test hypotheses. (Hunt, 1983, p., 10).

## 2.3 Customer Satisfaction

According to Yeh et al. (2019), customer satisfaction is assessed from the service from the company, according to Fan, Chen, & Miao. (2018) Customers who are satisfied with the services provided by the company will repurchase the product, according to J. K.C. Chen, Batchuluun, & Limitation. Furthermore, (2015) customer satisfaction is related to customer perceptions of service offerings, which are compared to the standard performance expected by customers. For example, suppose a customer is satisfied with his service offering. In that case, it will develop an intention to buy again in the future, be willing to share their experiences with others, pay no attention to competing brands, and even reject service offers from other brands (Yeh et al., 2019). According to Taghizadeh et al. (2016), service providers must continue innovating in their service offerings to increase customer satisfaction. Based on previous research, this study describes consumer satisfaction as consumer behavior after experiencing a specific product or service, whether the offer meets their expectations or not, and ultimately affects the future behavior of the brand.

## 2.4 Machine Learning

Machine learning techniques consist of 2, namely supervised or unsupervised. Supervised machine learning techniques, algorithms used to study the relationship between input and output. The algorithm can extrapolate to calculate the output value for new input data. Unsupervised learning techniques are concluded to create a natural structure that exists in the data set. Unsupervised learning techniques do not require labeled data or training data sets, so they are helpful for data expansion. A common unsupervised learning technique is grouping a set of data grouped into several clusters. Objects in the same cluster are more similar to each other than objects in different collections.

Table 1

Reviewed Studies on CRM and KAM using Machine Learning

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Article | Model | Forecasted Value | Methods | Business Context |
| Chiang,2018 | CRM | Customer Loyal | Data Mining | Airplane |
| Health, 2011 | KAM | Company Revenue | Simple Regression | Hotel |
| Lee et al. l, 2011 | CRM | Customer Loyal | Simple Regression | Company |
| Rodriguez & Boyer, 2020 | CRM | Sales Peformance | Simple Regression | Company |
| Tworek, K., & Sałamacha, A. (2019 | CRM | Customer Peformance | Simple Regression | Company |
| Kim & Lee, 2015 | CRM | Consumer Segmentation | Hybrid Methodologi | Company |
| Coda & de Castro, 2019 | CRM | B2B | Simple Regression | Company |
| Madsen & Johanson, 2016 | CRM | Customer Loyal | Cluster Analisis | Company |
| Harbin et al., 2016) | CRM | Customer Loyal | Customer Segmentation | Company |
| Demo et al., 2018 | CRM | Customer Loyal | Data Mining | Airlines |
| Yuen & Chan, 2018 | CRM | Customer Loyal | E- CRM | Company |
| Wang & Brennan, 2014 | KAM | Employee Peformance | Interview | Company |
| Ivens et al., 2018 | KAM | Employee Peformance | Simple Regression | Company |
| Ahmmed & Noor, 2018 | KAM | B2B | Simple Regression | Company |

# Methodology

Diagram

Description automatically generated

Figure 1 Combining CRM Framework with KAM

Before going to the framework of methods, figure 1 shows the framework in this study. It was adapted from the CRM and KAM methods. Several steps must be done, namely preparing the data to be predicted, determining segmentation with KAM, determining the service with the highest prediction, determining which prediction model to use. Below are the methods are taken by the CRM and KAM framework.

## 3.1. Method

Figure 2 shows the framework in this study. It is adapted from the standard method for constructing predictions, analytic models. There are five stages: collecting data; selecting relevant predictor variables; determine potential prediction methods; evaluate, validate, and select the best prediction model; and finally reported research results.



Figure 2 Prediction model framework

1. **PLN Data Collection**

In this study, we use PLN West Sumatra zone data. Our research used customer transaction data from January 2019 to December 2020, which consisted of 19,200,000 and 80 PLN customer transaction variables in 2 years. Table 2 shows the descriptive statistics for the dataset.

Table 2

Descriptive Statistics of The Datasets

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Data** | **Attribute** | **Statistics** | **Raw Data** | **Filtered Data** |
| Customer transactions 2019 | Number of Records  Period | Count  Min  Max | 9.600.000  1/1/2019  31/12/2019 | 2.000.000  1/3/2019  20/12/2019 |
| customer transactions 2020 | Number of Records  Period | Count  Min  Max | 9.600.000  1/1/2020  31/12/2020 | 1.000.000  5/7/2020  28/12/2020 |

1. **Choice of Variable**

The variable predictor is based on data obtained from the PLN West Sumatra Zone. There are about 80 variables, but not all of them are used then the variables are used only partly because it is not possible to do all the computations. Therefore, the selected variable has the highest potential. Table 3 shows detailed information about the variable predictors.

Table 3

Choice of Predictor Variable

|  |  |
| --- | --- |
| **Variable/Data Types** | **Variable Description** |
| Blok3 | Crucial time of use of electricity |
| BATH | Date of being a customer |
| TARIF | The price of electricity per month |
| KWHBP | The number of kWh used at peak electricity loads |
| SAHLWBP | Power consumed outside of peak load times |
| KWHLBP | Payments that must be paid outside of the peak load |
| RPLWBP | Payments that must be paid for using electricity outside a predetermined load |
| RPBEBAN | The payment that must be paid for using electricity within a specified time |
| RPPTL | The payment that must be paid for using electricity |
| RPBPJU | Payment to be paid while the street lighting is being operated |
| RPPLN | the agreed price for using the electricity |
| RPG | The payment that must be paid is stated in the bill |
| RPTAG\_MAT | The payment that must be paid is stated in the bill according to the agreement that has been signed with a stamp duty |
| KDPEMBAYARAN | Payment code that has been determined |
| RPBLOK3 | the payment used for using the block 3 |
| KOGOL | Group code that has been determined |
| RPBK1 | Price to be paid if you are late 1 |
| RPBK2 | The fee to be paid if you are late 2 |
| RPBK3 | The cost to be paid if you are late 3 |
| TGLBACA\_AWAL | Initial reading time |
| TGLBACA\_AKHIR | Final reading time |

1. **Choisce of Potensial Method**

This research aims to develop a combined prediction model between CRM and KAM 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 multiple models in one machine learning model, researchers investigated both the ensemble model and the single model. In general, ensemble models are more accurate in predicting than single models. However, the single model still outperforms the ensemble model. Researchers used Linear Regression, Naïve Bayes. In the ensemble group, researchers used a cluster approach using the K-Means algorithm.

Table 4

Prediction Model Choice

|  |  |  |
| --- | --- | --- |
| **Classifiers Category** | **Classifiers by Group** | **Model** |
| Ensemble | Parallel/Bagging | Cluster |
| Singular | Bayes | Naïve Bayes |
|  | Regression | Regresi Linear |

**C.1. Linear Regression**

Generally, Linear Regression is used to describe and test hypotheses. Therefore, choosing the correct variable and avoiding highly correlated variables must be considered when using Linear Regression. Furthermore, the variable predictors in linear regression can be categorical or numeric, and the target variable for linear regression is binary or dichotomous. Therefore, Linear Regression cannot predict the target variable for more than two classes. Although Linear 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. Linear regression formulation is stated as follows:

Graphical user interface, text, application, email

Description automatically generated

**C.2. Naïve Bayes**

Naïve Bayes is an algorithm contained in the classification technique. Naïve Bayes is a classification using probability and statistical methods introduced by Thomas Bayes. Bayes' theorem is predicting future opportunities based on past experiences. The theorem is combined with Naïve, where it is assumed that the conditions between the attributes are mutually independent. The Naïve Bayes classification thinks that the presence or absence of specific characteristics of a class has nothing to do with the elements of other courses. Here's the formulization from Naïve Bayes:

Graphical user interface, text

Description automatically generated

Information:

𝑃: Opportunity

𝑋𝑖: Attribute to i

𝑥𝑖: The value of the ith attribute

𝑌: The class you're looking for

𝑦𝑗: Y subclass you are looking for

𝜇: Mean, expresses the average of all attributes

𝜎: Standard deviation, representing the variants of all attributes

**C.3. Cluster K-Means**

K-Means is an algorithm in data mining that can be used to group/cluster data. There are many approaches to creating clusters, creating rules that dictate membership in the same group based on equality among its members. Another system is to develop a set of functions that measure some of the grouping properties as a function of several clustering parameters. The K-Means method is a method that belongs to a distance-based clustering algorithm that divides data into several clusters, and this algorithm only works on numerical attributes. The K-Means process is grouped with the following algorithms:

1. First, determine the number of groups.
2. Allocate data into groups randomly.
3. Calculate the center of the group from the data in each group.
4. Allocate each data to the nearest centroid/average. The following is the formalization of K Means.

A screenshot of a computer

Description automatically generated with medium confidence

1. Return to Step 3 if there is still data that has moved groups or if the change in the centroid value is above the specified threshold value, or if the value of the objective function used is still above the fixed threshold value.
2. **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 19,200,000 records, the researcher grouped the dataset to fold number one to number nine, which consisted of 192,000 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 (192,000 records), the training set consists of 1,544,000 (192,000 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 1,500,000 records (190,000 x 7-fold + 192,000 records from the previous ten-fold) and the validation set consists of 192,000 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 values were used to determine the prediction score better than accuracy. Mathematically, the researchers formulated the ABK score as follows:

A screenshot of a computer

Description automatically generated with medium confidence

1. **Model Use and Reporting**

They are comparing model development times and predictive scores based on the performance of each model. The best prediction model with the best predictive AUC-ROC score will help decision-making formulating the combination of CRM and KAM that is suitable in a better way.

# 4. Result

Table 5 and Table 6 show the results of the evaluation of the machine learning classification model that was built. Supervised machine learning models create models automatically from the training data set. The learning algorithm identifies and develops generalizable patterns that reflect the relationship between the dependent (target) and independent variables. Based on the built design, the model can then be built predictions of target variables based on the observed independent variables. To test the model's accuracy, the prediction results of the model that are created are then compared with the actual value of the target variable. If the target variable is categorical, the AUC-ROC score is usually used to evaluate how well the model can distinguish between different categorical variables (classes).

The AUC-ROC score is a simple evaluation method that is considered better than the predictive accuracy score as an evaluation method. Tenfold three-way and ten-fold cross-validation procedures evaluated linear Regression, Naïve Bayes, and Clustering K-Means methods. Table 5 shows the results of the three-way separation and comparison between data with and without standardization process data. Data standardization increases the mean and decreases the predicted standard deviation of the AUC-ROC score from linear regression.

Naïve Bayes is modeled in both validation and test conditions. However, the data standardization process cannot improve the AUC-ROC score prediction from clustering using K-Means. Naïve Bayes has the lowest AUC-ROC score but has the fastest time model development. Furthermore, the K-Means clustering classifier had an AUC-ROC score of 15.03% higher than Naïve Bayes 2.93%. Therefore, we conclude the best performing K-Means Clustering. The results of the tenfold cross-validation procedure are shown in Table 6. The fold column indicates the order of the fold, and the rest of the column shows the score for each Technique. Finally, the average score and standard deviation are at the bottom of the table. Linear regression and k-means clustering resulted in better AUC-ROC scores after undergoing the data standardization process. Naïve Bayes provides the fastest processing times but the lowest scores.

Table 5

Evaluation Result of Ten-Fold Three-Way Split Procedure

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Fold | **Regresion Linear** | | | | | | **Naïve Bayes** | | | | | | **Clustering Using K-Means** | | | | | | | |
| **Not- Standardized Data** | | | **Standardized Data** | | | **Not- Standardized Data** | | | **Standardized Data** | | | **Not- Standardized Data** | | | | **Standardized Data** | | | |
| VF | MVS | TS | VF | MVS | TS | VF | MVS | TS | VF | MVS | TS | VF | MVS | TS | VF | | MVS | TS |
| 1 | 2 | 0,587 | 0,389 | 2 | 0,677 | 0,600 | 1 | 0,628 | 0,604 | 1 | 0,627 | 0,604 | 2 | 0,739 | 0,707 | 2 | | 0,793 | 0,759 |
| 2 | 2 | 0,568 | 0,545 | 2 | 0,656 | 0,610 | 2 | 0,636 | 0,639 | 2 | 0,636 | 0,638 | 2 | 0,736 | 0,722 | 2 | | 0,791 | 0,760 |
| 3 | 2 | 0,576 | 0,525 | 2 | 0,687 | 0,620 | 1 | 0,629 | 0,626 | 1 | 0,628 | 0,630 | 2 | 0,740 | 0,723 | 2 | | 0,790 | 0,770 |
| 4 | 2 | 0,765 | 0,512 | 2 | 0,788 | 0,780 | 1 | 0,639 | 0,636 | 1 | 0,638 | 0,636 | 2 | 0,723 | 0,740 | 2 | | 0,770 | 0,790 |
| 5 | 3 | 0,812 | 0,634 | 3 | 0,814 | 0,812 | 2 | 0,629 | 0,621 | 2 | 0,630 | 0,610 | 3 | 0,734 | 0,710 | 3 | | 0,790 | 0,754 |
| 6 | 3 | 0,934 | 0,867 | 3 | 0,935 | 0,932 | 3 | 0,631 | 0,607 | 3 | 0,632 | 0,608 | 3 | 0,741 | 0,715 | 3 | | 0,788 | 0,749 |
| 7 | 3 | 0,834 | 0,676 | 3 | 0,843 | 0,834 | 3 | 0,650 | 0,603 | 3 | 0,649 | 0,604 | 3 | 0,742 | 0,682 | 3 | | 0,793 | 0,738 |
| 8 | 3 | 0,546 | 0,555 | 3 | 0,555 | 0,500 | 3 | 0,643 | 0,602 | 3 | 0,643 | 0,601 | 3 | 0,743 | 0,685 | 3 | | 0,791 | 0,749 |
| 9 | 3 | 0,965 | 0,978 | 3 | 0,978 | 0,912 | 3 | 0,643 | 0,610 | 3 | 0,643 | 0,610 | 3 | 0,742 | 0,669 | 3 | | 0,794 | 0,741 |
| 10 | 3 | 0,897 | 0,917 | 3 | 0,912 | 0,900 | 3 | 0,651 | 0,585 | 3 | 0,651 | 0,584 | 3 | 0,747 | 0,609 | 3 | | 0,796 | 0,659 |
| AVG |  | 0,749 | 0,660 |  | 0,785 | 0,75 |  | 0,781 | 0,755 |  | 0,699 | 0.672 |  | 0,758 | 0,725 |  | | 0,779 | 0,751 |
| STD |  | 0,006 | 0.050 |  | 0,004 | 0,027 |  | 0,004 | 0,027 |  | 0,008 | 0,021 |  | 0,003 | 0,045 |  | | 0,003 | 0,032 |
| SPT |  | 2,9 |  |  | 263 |  |  | 2,63 |  |  | 1,33 |  |  | 13,08 |  |  | | 101,3 |  |

VF = Validation Fold, MVS= Maximum Validation AUC -ROC Score, TS= Test AUC- ROC Score, AVG= Average AUG-ROC Score, STD= Standard Deviation, SPT= Sum of Processing Time in Minutes

Table 6

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Regresion Linear** | | **Naïve Bayes** | | **Clustering Using K-Means** | |
| Fold | **Not- Standardized Data** | **Standardized Data** | **Not- Standardized Data** | **Standardized Data** | **Not- Standardized Data** | **Standardized Data** |
| 1 | 0,748 | 0,771 | 0,609 | 0,609 | 0,714 | 0,756 |
| 2 | 0,750 | 0,779 | 0,620 | 0,619 | 0,727 | 0,764 |
| 3 | 0,759 | 0,784 | 0,630 | 0,630 | 0,729 | 0,772 |
| 4 | 0,782 | 0,800 | 0,652 | 0,652 | 0,740 | 0,793 |
| 5 | 0,732 | 0,767 | 0,621 | 0,620 | 0,718 | 0,759 |
| 6 | 0,714 | 0,748 | 0,611 | 0,611 | 0,713 | 0,751 |
| 7 | 0,648 | 0,741 | 0,613 | 0,612 | 0,684 | 0,738 |
| 8 | 0,681 | 0,752 | 0,617 | 0,617 | 0,687 | 0,753 |
| 9 | 0,669 | 0,741 | 0,600 | 0,600 | 0,675 | 0,739 |
| 10 | 626 | 0,675 | 0,580 | 0,580 | 0,606 | 0,658 |
| Average AUC-ROC | 0,711 | 0,756 | 0,615 | 0,615 | 0,699 | 0,748 |
| Standard Deviation | 0,052 | 0,034 | 0,019 | 0,19 | 0,039 | 0,036 |
| Model Development Time (in seconds) | 19,00 | 16,00 | 10,00 | 10,00 | 53,00 | 376,00 |

# Conclusion

From the mean AUC-ROC score, the k-means cluster model appears to be superior in both evaluation procedures. The naïve Bayes model achieved a 0.773 mean AUC-ROC score and a ten-fold 0.781 three-way range under tenfold cross-validation conditions. The ensemble method outperforms the single strategy from the classifier category, which means that the ensemble method is better than the single method at dealing with bias, noise, and variance. Naïve Bayes produces the lowest scores due to inaccuracies but has the fastest processing times due to its simplicity. The process of standardizing the data increases the AUC-ROC score for Linear Regression and Cluster K-Means.

Interestingly the processing time after data standardization is reduced in Linear Regression. It is influenced by customer satisfaction, and data standardization can handle the negatives of the customer satisfaction case. That is why the score increases and the model development time decreases in the Linear Regression model. There are many advantages of the k-means cluster model, such as handling customer satisfaction and customer service. The CRM and KAM methods show different mean AUC-ROC values ​​in the same model. The Cluster K-Means and Linear Regression resulted in a tenfold higher mean score in the cross-validation method than in the tenfold three-way separation procedure. However, another model produced higher AUC-ROC scores in the three-event split procedure. In addition, the mean difference in AUC-ROC scores between models was higher in the cross-validation method than in the three-event split procedure. This means that the number of training sets affects the test score of each model.

# Discussion

Considering the importance of demand forecasts in the context of income management, this study analyzes three learning techniques to predict the likelihood of ordering a list of accommodations. We evaluated the AUC-ROC score of each model using two different evaluation methods, namely tenfold three-way split and ten-fold cross-validation procedures. In terms of the AUC-ROC score, the K-Means Cluster classifier outperformed other models, namely Linear Regression, Naïve Bayes. Naïve Bayes has the lowest AUC-ROC score. Linear Regression Performance in predicting the most superior likelihood of customer satisfaction. These findings can tell companies to improve their predictions and revenue management responses. In terms of its contribution to literacy, this study informs a method of predicting the likelihood of customer satisfaction.

# 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).