ATLANTIS PRESS
Advances in Economics, Business and Management Research, volume 65
1st International Conference on Economics, Business, Entrepreneurship, and Finance (ICEBEF 2018)

Telecom Customer's Segmentation Using Decision Tree to Increase Active Electronic Money Subscribers

I Gede Wiyana Ananta Noor, Maya Ariyanti, Andry Alamsyah
School of Economics and Business
Telkom University
Bandung, Indonesia
w.ananta@gmail.com

***Abstract*—The ABC telecommunication company as one of electronic money providers has more than 100 million customers. If it is compared to the number of electronic money customers which have growing potential. In December 2017, the number of electronic customers owned by ABC was on the third rank of electronic money ownership. Some efforts to increase the number of electronic money customers have been conducted but it has not achieved the expected target. Based on the aforementioned problem so that identifying future customers from potential telecommunication customers to be active electronic customers thus campaign activity can be done effectively and controllably. Therefore, a customer predicting model is needed to predict potential customers to be active electronic customers. This research creates model which can be used to predict future customers using telecommunication transaction act at ABC Company. The analysis used was telecommunication transaction data for all electronic money customers with 32 variables. Those variables were formed from variables such as voice, SMS and internet usage including other forming transaction such as customer's dominant location, operating system from device and device type used by customers. Forming method model used decision tree with accuracy (ACC) measuring evaluation, positive prediction value (PPV), negative prediction value (NPV), true positive rate (TPR) and true negative rate (TNR). Based on the evaluation result, this model can predict the future customers who will be the active electronic customers for 54,09%.**

Keywords—electronic money; prediction model; decision tree

I. INTRODUCTION

Based on Bank of Indonesia regulation number: 11/12/PBI/2009 about electronic money, electronic money is the payment tool that meets these elements:

- Issued based on the money value paid in advance by holder to the issuer;
- Money value which is saved electronically in a media such as server or chip;
- Used as the payment tool to the holder who is not the electronic money issuer; and
- Electronic money value which is deposited by the holder and managed by the issuer is not saving as what has been mentioned in the law regulating about banking.

Thus the electronic money owned by the communication company is server-based electronic money which is different from credit. The electronic money owned functions as saving media of electronic money and can be used for any transaction types such as: payment, purchasing and money transfer.

Indonesian government plans transaction without cash money. The ABC Company has introduced electronic money service earlier in 2007. Somehow based on the survey result conducted by JAKPAT – Mobile Survey Platform Indonesia, in December 2017 electronic money owned by ABC was the third electronic money ownership.

By using data gained from Bank Indonesia, the authors found that ownership percentage of ABC from 2010 to 2017 has experienced declining trend. It was influenced by the growing number of electronic money issuers who holds license from Bank Indonesia.

Some efforts to increase the customer number as well as active customers, such as: opening booth in malls until giving gifts to ABC customers. But those efforts have not increased the expected target amount of customers. This happened because the campaigned done so far has not had the right market target in which the campaigns were still random.

Telecommunication industry has market penetration higher than banking, so that the electronic money products which are expanded by telecommunication industry has bigger opportunity to dominate market. In another side, telecommunication revenue nowadays supported by voice and SMS transaction, keep shifting to revenue digital service based.

Banking account ownership ratio in Indonesia is only 36%, one rank below Cambodia for 39% (based on researches in Southeast Asia countries in 2014). The bank account ownership condition is contrast to mobile phone penetration in Indonesia for 125% [1], this becomes the opportunity for electronic money to get into society segment who does not have banking access.

Concerning the achievement of electronic money of ABC, that is the active customers in February 2018 were only 4.5 millions from all customers which are 19 millions, this achievement is far from target which was set in 2017 for 33

million. Those customer numbers are also very small compared to main telecommunication customers. It is caused by the absence of model that can be used to predict potential customers to be active electronic money customers.

Thus, the question in this research is how to form decision tree model from transaction history of telecommunication customers in ABC to be able to predict potential customers to be active electronic money customers. This research is aimed at analyzing data by running big data processing in ABC by using classification method which aims at making decision tree model from ABC transaction history and training set which have been defined before. Training set used supervised data set that data which has class attribute.

The result of this research is expected to contribute in several areas such as:

- For academic usage, this research can give opportunity for the writer to contribute the knowledge about applying one of data mining theory in marketing field.
- For practical usage, this research is expected to give contribution of modelling that can be used to predict ABC customers who have not become active electronic money customers and support Indonesia government in increasing inclusive monetary.

II. Literature Review

A. Segmentation and Marketing Target

Marketer is not connected to all customers in big, vast or varied market [2]. Identification on market segment is required to be done to serve the customers effectively. This decision needs comprehension and thorough strategic thinking what makes each segment unique and different. Identify and uniquely satiate the correct market segment often becomes the key success of market.

Facing the competition effectively, today there are companies focus on marketing target. They focus on possessing better potential market. The effective marketing target required the marketer to:

- Make dentification and differentiate buyer's group profiles in the need and wish (market segmentation).
- Select one or several segments that will be chosen (market target).
- For each target segment, make, communicate and give precise benefit to offer company's market (market posisioning).

B. Churn

Increasing company's profitability is not enough by attracting new customers, the company should their customers and develop their business. Increasing the number of customers but experiencing the high churn of customers is the same as pouring water into leaking bucket [2]. Some churn definition that may different for each industry [3].

TABLE I. Churn Definition

Data set	Definition of churn
Banking	The customer with below certain limit in certain period.
Retailing	The customer who changes the buying pattern for one period or customers without login for certain period.
Online gambling	The customer who never plays for certain period.
Retailing (newspaper)	The customer who has not renewed their product for certain period.
Mobile telecommunication	The customer who does not enjoy all the offered services or customers who move to other operators.
Online games	The player who permanently leaves the game.
Online retailing	The customer who moves to competitor.

C. Marketing Intelligence

Marketing intelligence is continual comprehending process, analyze and estimate internal and external company's environment related to customers, competitors and markets and then use the information and knowledge gained to support the company's market related with decision [4]. Nowadays in the comprehending process and analyze several companies uses big data to gain individual customer's information is the most suitable to meet their needs. By using a methodology, mining the transaction using technology and demographic information about customers (product or service which they purchase, contract requirement and expired date, complaining note or average waiting duration) and call center agent (average of handling duration and selling efficiency) to identify the optimal real time matching.

By using market analytic, the marketing manager expects to be able to avoid personalization which does not describe the customer's character. It is realized when we receive email, SMS and even chat which has been personalized but does not talk to use when we browse in a website and there is a banner of advertisement which is not suitable with our need. It may show us that profile, modelling and segmentation have not been done [3].

D. Decision Tree

Several classification techniques can be used to predict are logistic regression, artificial neural network, decision tree and support vector machine [3]. Decision tree is widely chosen for several reasons [5]. First, decision tree method is suitable to for analyzing complex relation between variables and does not need specific assumption about functional form. Decision tree is non-parametric method which does not depend on the probability knowledge about the phenomena being observed. Second, decision tree is more suitable to describe complicated relation pattern such as non linier between variables. Third, the structure of decision tree allows the decision rule extraction which can be used to find the pattern between variables. Decision tree provides effective method to categorize a set of cases on the database to different class by using resulted decision rule. Fourth, this research introduces predictive model based on decision tree which gives significant predictive power and allows to predict with high accuracy. Implementation of decision tree on data mining is classification model uses tree

structure or hierarchy structure. Classification activity is done by inserting variables which have been identified before. Those variables is gained from transformation of data transaction which is predicted to have influence on classes (designated variables).

Decision tree is mapping the problem solving alternatives which can be achieved from that problem. The concept of decision tree is changing data to be decision tree and decision rules. The data in decision tree is generally stated in form of table with variables and information inside of it will be made as criteria in making trees.

III. Modeling and Evaluation

The researcher will use variables which have been defined by concerning the value of each variable. The variable is defined toward the object which has been known for its class (predicted variable) that is active and inactive electronic money.

A. Variable and Data Set

Based on the previous, in analyzing data need, the authors define the data source needed such as:

- Key variable, is MSISDN (customer's number),
- Predicted variable, is the status of electronic money customer yaitu status pelanggan uang elektronik (active and inactive),
- Predictor variable, is:
  - The device used by the customers [6],
  - Billing and customer charging [7],
  - Conversation data, is voice transaction and SMS [7], and
  - Internet service transaction [7].

In gathering data, the authors used two main data sources those are telecommunication transaction and electronic money transaction. From the electronic money transaction, the active customer's list will be collected from the customer who has transaction for six months and inactive customer is not categorized as active customers. The customer list will be the main source to form the data mart. Variable description that variable description is formed by collecting data transaction for one month and customer's basic profile in ABC.

B. Model Analysis

In analyzing the research result, the authors will make confusion matrix like in table 2. Each value from confusion matrix is gained from model testing by using data testing.

TABLE II. Confusion Matrix

		Actual Condition	
		Active	***Inactive***
Prediction	***Active***	TP (True Positive)	FP (False Positive)
	Inactive	FN (False Negative)	TN (True Negative)

In this research analysis technique and data evaluation are using cross validation method K-Fold Cross Validation. In this technique data set is divided into some K-partition randomly with the same size. Some K-times experiment, in which each experiment used data partition for the-K number as the data testing and used the rest of other partitions as data training. The error level of all K partition is cross-validate error rate. This procedure has been tested extensively by several partition numbers, and 10 Fold Cross Validation has been considered enough and accurate, especially for big sample, although computation capacity enables for more partition [8].

C. Analysis Technique

This research used quantitative method by using data gained from ABC customers' transactions and listed as electronic money customers. The data is formed into data mart which has been known for the class those are active and active electronic money customers.

Several parameters used to predict and evaluate are:

1) Accuracy: Accuracy is the measurement to find the probability of precise predicted future customers both active and inactive to all data used as testing.

$$\text{Accuracy (ACC)} = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

2) Positive Predictive Value (PPV) or precision: Positive Predictive Value or precision is measurement to know probability future electronic money customer to be claimed as active.

$$\text{Positive prediction value (PPV) or precision} = \frac{TP}{TP + FP} \quad (2)$$

3) Negative Predictive Value (NPV): Negative Predictive Value is a measurement to find probability future electronic money to be claimed as inactive.

$$\text{Negative predictive value (NPV)} = \frac{TN}{FN + TN} \quad (3)$$

4) True Positive Rate (TPR) or recall or sensitivity: True Positive Rate or recall or sensitivity is measurement to find probability of predicted future electronic money customers to be claimed as active on all true active condition.

$$\text{True positive rate (TPR) or recall or sensitivity} = \frac{TP}{TP + FN} \quad (4)$$

5) True Negative Rate (TNR) or spesificity: True Negative Rate or spesificity is measurement to find probability of future electronic money customers who are predicted to be inactive to all true inactive condition.

$$\text{True negative rate (TNR) or spesificity} = \frac{TN}{FP + TN} \quad (5)$$

D. Data Presentation

Data set used in this research is all electronic money customers in February 2018 for 19,252,907 customers. By

using electronic money transaction data during September 2017 until February 2018 the active customers are known for 4,521,522 or for 23.48 % from population. The active customers have definition as owning transaction that caused the customers' credit reduced but does not include transaction for dormant account. Meanwhile for inactive customers for 14,731,385 or for 76.52 %. From all active and inactive electronic money customers, the authors make data mart by completing that data with transaction history of the customers in February 2018.

Before modeling, the data normality testing was taken by using IBM SPSS Modeler ver. 18.0. The testing result shows that invalid data or can be noise in the model.

E. Preprocessing

Before learning process, data cleansing was done previously to prepare data which did not contain noise information. In this stage, the authors conducted several processes:

1) Reclassify: Reclassify is aimed to transform one value in the data set into other category. In the IBM SPSS Modeler, reclassify is often used as the node to recategorize a data. In this research was used to make value consistency for a value which has the same meaning such as: "null", "?", and "undefine".

2) Select: Node select is aimed to select or delete a group of data with certain condition. In this research is used to delete data for several conditions: first total usage and total voice offnet which have negative values and second device os, device type, dominan recharge channel, and dominan location has "null" value.

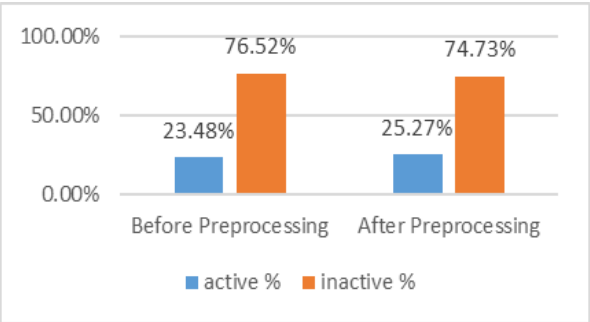

Fig. 1. The comparison before and after preprocessing.

In the figure 1 percentage before and after preprocessing have changed for less than 2%.

F. Model Forming

Model forming was made by using node C5.0 as well as prepare output type K-Fold Cross Validation. Implementing it, the authors used K=10 which were 10 partition for 10 times of iteration, in which each iteration used data for one of partitions in turn as data testing and use the other partition left as data training. 10 Fold Cross Validation has been considered as adequate and accurate, especially for big sample, although calculation capacity enables more partition.

G. Model

The decision tree model which is visualized in the form of rules representing the formed tree structure. Rule is the representation of root to leaf. Thus 1 leaf represents 1 rule. Although this research aims to predict the future customers that will be the future active electronic money customer, the modelling in this research produced 4,161 rules for active electronic money and 4,458 rules for inactive electronic money.

H. Evaluation and Discussion

The first evaluation is accuracy (ACC). ACC is the most used reference of a model performance. ACC shows capability of a model in predicting accurately for all predictions. In this research accuracy is gained from probability of accuracy predicting both predicting active and inactive electronic money for actual condition. From the modelling results, it gains ACC for 87.14%.

The second is Positive Predictive Value (PPV). Model evaluation used PPV which is called as precision is model evaluation to measure probability which is correctly predicted as positive from all positive predictions. From the modelling results, it gains PPV for 54.09%. It shows that modelling capacity this model's capacity to predict the future customers will be active electronic money accurately for 54.09%. This shows that there is risk for 45.91% for customers who are predicted as active electronic customers but actually does not become active electronic money customers.

The third evaluation is Negative Predictive Value (NPV). Evaluation model uses NPV is by measuring probability which is measured to be negative correctly from all negative predictions. From the modelling results, it gains NPV for 93.71%. This shows that this model's capacity to predict the future customers will not be active electronic money customers for 93.71%.

The fourth is True Positive Rate (TPR). TPR is model evaluation by measuring probability predicted as positive correctly for all samples whose conditions are positive. From the modelling, it gains TPR for 74.40%. This shows that from all active electronic money customers who are predicted correctly for 74.40%.

The fifth is the True Negative Rate (TNR). Model evaluation used TNR is model evaluation by calculating probability which is predicted as negative correctly for all samples whose conditions are negative. From the modelling, it gains TNR for 90.15%. It shows that from all inactive money electronic customers have been predicted correctly for 90.15%.

IV. Results and Discussion

This research yielded model evaluated with measurement ACC 83.70%, PPV 54.09%, NPV 93.71%, TPR 74.40% and TNR 90.15%.

Accuracy in this study shows the precision of predicting customers to be electronic money or not, making a higher value than research, as did Govindaraju et al. which resulted in accuracy of 70.94% in the prediction model of customer PT.

Telkomsel [7]. Another study of churn predictions by Lee et al. obtained an accuracy of 68.57% with decision tree technique [5].

PPV in this study shows that predictions made in predictions will generate electronic money. This study resulted in a PPV of 54.09% and a lower ranking than previous studies, as did Govindaraju et al. that generated PPV of 58.02% in the prediction model of customer PT. Telkomsel said it was good enough [7]. This PPV measurement is formed from 4,161 rules for Electronic money, which is the existing rules, which can be used for Active money of 54.09%. It also indicates that 45.91% of the cluster of customers who will make electronic money, not in accordance with predicted results. It's still better that can cause a bunch of rules, to find customers who can get it out soon. Electronic money without customer fees available in the same place as LBA.

Another way with an NPV that states the right amount for prediction will be money, in this study, the model formed yields 4,458 rules for Electronic money inactive. With a NPV generating of 93.71%, use this model's capabilities in electronic money. It also explains the company's set of rules here to get the customer contact numbers and services needed for money, but can be used for electronic money only by 6.29%. Will be used for greater things for those associated with existing products of this rule.

This study uses predictor variables such as devices used by customers, billing and charging customers, both voice and SMS data and internet usage. This is in general also consistent with research conducted by Govindaraju et al. and Lee et al [6,7]. However, there are predictor variables that are not used in the Govindaraju et al. such as number of days of service and data in each month, customer USIM usage, operating system of device used by customer and customer location for the first time [7]. In addition there are predictor variables that cannot enter into Lee et al. such as the length of time the customer, the number of days of service and data in each month, the purchase of credit by the customer, the duration of voice usage and the amount of SMS usage [6].

V. CONCLUSIONS

This study provides evidence that the approach to data mining can be applied into the field of marketing. This research uses data of telecommunication transaction from all customer of electronic money either active or inactive electronic money customer. This indicates that from ABC subscriber transaction history data, the decision tree model can be used to predict the prospect will become active customer of electronic money.

The resulting model is a rule of 8,619 of which 4,161 rules can be used to predict that potential customers will become active electronics customers and 4,458 rules to predict potential customers who will become inactive electronic money customers. The capability of the resulting model can predict the prospect will be an active electronically active customer with 54.09% of all potential customers who are predicted to be active E-money. This model has the ability to predict that potential customers will become inactive electronic money customers better than active electronic money customers of 29.61%.

REFERENCES

[1] R. Hamdani, Brownbag Session, Telkomsel Sharing Session, 2016.

[2] P. Kotler and K. L. Keller, Marketing Management, vol. 15, Harlow, Essex: Pearson Education Limited, 2016.

[3] S. Khodabandehlou and M. Z. Rahman, "Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior," Journal of Systems and Information Technology, vol. 19, no. 1/2, pp. 65-93, 2017.

[4] A. Alamsyah, W. Rahmah and H. Irawan, "Sentiment Analysis Based On Appraisal Theory for Marketing Intelligence in Indonesia's Mobile Phone Market," Journal of Theoretical and Applied Information Technology, vol. 82, no. 2, pp. 335-340, 2015.

[5] S. Lee and J. Y. Choeh, "Exploring the determinants of and predicting the helpfulness of online user reviews using decision trees," Management Decision, vol. 55, no. 4, pp. 681-700, 2017.

[6] E.-B. Lee, J. Kim and S.-G. Lee, "Predicting customer churn in mobile industry using data mining technology," Industrial Management & Data Systems, vol. 117, no. 1, pp. 90-109, 2017.

[7] R. Govindaraju, T. Simatupang and T. A. Samadhi, "Perancangan Sistem Prediksi Churn Pelanggan PT. Telekomunikasi Seluler dengan Memanfaatkan Proses Data Mining," Jurnal Informatika, vol. 9, no. 1, pp. 33-42, Mei 2008.

[8] G. Cui, M. L. Wong, G. Zhang and L. Li, "Model selection for direct marketing: performance criteria and validation methods," Marketing Intelligence & Planning, vol. 26, no. 3, pp. 275-292, 2008.