THE DESIGN OF A CHURN PREDICTION MODEL USING DATA MINING FOR CUSTOMER RETENTION: A CASE OF POSTPAID CUSTOMERS OF PT. XYZ CELLULAR

Putu Eka Putra.ST[1], Gadang Ramantoko[2]

School of Economics and Business, Telkom University, INDONESIA.

[1]eputra17@gmail.com, [2]gadangramantoko@gmail.com

ABSTRACT

Based on data of sales report, churn, customer base postpaid retail segment (Regular) in the last five years at PT. XYZ Cellular, there has been an increase in the number of customer churn. The large number of customer churn indicates that the implemented retention programs have not been on target. Therefore, it is necessary to conduct a research to design a churn prediction model. The churn prediction will be useful for designing and deciding the retention program to be carried out and which customers will be targeted for retention programs. The data used in this study was derived from PT. XYZ Cellular. It was related to customer behavior including: (1) total billing, (2) conversation, SMS, usage, and (4) churn information data from January to March 2018. Data processing was done using data mining with logistic regression modeling with SPSS data modeler. Based on the results of data processing, Model Logistic Regression has sensitivity or ability to precisely predict customer churn from the group of churners of 97.9%. Five best predicting variables were: (1) average transaction of the use of voice among PT. XYZ Cellular users, (2) whether or not customers have changed their devices, (3) the average transaction of the use of voice calls to users other than PT. XYZ Cellular users, (4) average customer bills, and (5) whether or not the customers have overdue bills.

Keywords: Customer Retention, Big Data Analysis, Logistic Regression, Churn prediction score

INTRODUCTION

Background

Based on data from We Are Social website in January 2017, the number of cellular subscribers in Indonesia compared to the population in Indonesia has reached 142%. This shows that the penetration of cellular customers in Indonesia is very high. One of the challenges faced by the cellular telecommunication operators in Indonesia in the future is finding a way to reduce the number of customers who stop using their company services and move to competitor companies. The behavior of customers who discontinue the use of services provided by the cellular telecommunication operator companies is called churn. Churn has a direct impact on the reduction of profit and market share of the company. Therefore, churn management is a crucial strategy in the competition.

As a telecommunication company that provides cellular service, PT. XYZ Cellular has realized the importance of overcoming churn problem. This company is the leader of the entire telephone industry market. With such large market share and large number of customers, efforts need to be made to keep the market share value from falling. A more efficient way to maintain market share is needed because the cost of retaining customers is

cheaper than the cost of getting new customers. Hence, retaining customers and reducing churn is the most appropriate way for PT. XYZ Cellular to maintain its market share.

PT. XYZ Cellular has a database to store the data needed in running its business. Large amount of data which is stored electronically in the company's database can be used to find patterns of consumer behavior and characteristics. Proper processing of the data can produce useful knowledge to understand patterns of churn behavior and predict which customers will discontinue their subscriptions. The process of finding meaningful relationships, patterns and trends by examining a large set of data stored in storage by using pattern recognition techniques such as statistical techniques and mathematics is called data mining (Larose, 2016). Data mining as one of the data analysis tools has been proven in predicting the customers who will discontinue their subscriptions, when they will churn, and also accuracy of the prediction. Therefore, data mining can be used as an approach to analyze and predict churn of postpaid customers of PT. XYZ Cellular.

Formulation of Problem

Based on the description in the background of the study and due to the importance of churn prediction, it can be concluded that PT. XYZ Cellular needs a churn prediction to support the selection strategy of the target that will be given retention programs to reduce the number of customer churn which will impact on decreasing revenue. Data mining can be used as an analysis tool and customer churn prediction. It is one of the analytical features of CRM. From the available data, a prediction model is designed to help the implementation of effective and efficient retention programs.

Research Purpose

Based on the formulation of the problem, the purpose of this research is to create a churn prediction of PT. XYZ Cellular customers in regular segment using data mining.

LITERATURE REVIEW

Definition of Consumer Behavior

Many experts define consumer behavior. Hawkins and Motherbough (2010) in their book entitled Consumer Behavior Building Marketing Strategy define consumer behavior as a study of individuals, groups or organizations and the processes by which they select, use and dispose products, services, experiences or ideas to satisfy their needs. While Michael Solomon et al. (2006) in the book entitled Consumer Behavior: A European Perspective state that consumer behavior is a process when individuals or groups select, buy, use or dispose products, services, ideas and experiences to satisfy their needs. Consumer behavior is also defined as a study of how individuals, groups and organizations choose, buy, use, and dispose goods, services, ideas, or experiences to meet their needs and desires (Philip Kotler and Kevine Line Keller, 2012).

Buying Decision Process

Philip Kotler and Kevine Line Keller (2012) in their book mentioned that there are 5 stages in the process of buying products, namely: (1) Problem Recognition, (2) Information Search, (3) Evaluation of Alternatives, (4) Purchase Decision, and (5) Post Purchase Behavior.

Customer Loyalty

Customers who repurchase or repair certain products or services show a commitment called consumer loyalty (Philip Kotler and Kevine Line Keller, 2012). Customer loyalty

Leena and Luna International, Chikusei, Japan.
(株) リナアンドルナインターナショナル, 筑西市,日本
ISSN: 2186-845X ISSN: 2186-8441 Print
www.ajmse. leena-luna.co.jp

is formed due to two main factors, namely customer trust and customer satisfaction (Ahmad,et al, 2015).

Data Mining

Data mining is defined as a process that uses mathematical techniques, statistics, artificial intelligence and machine learning to extract and identify useful information for gaining knowledge from databases (Bahari and Elayidom, 2015). Data mining can extract valuable potential from knowledge, models and rules from large amount of data (Guo and Qin, 2017). According to Rekha (2015), data mining is an analytical process designed to explore data to look for patterns or systematic relationships that are consistent between variables and then validated.

Logistic Regression

According to Kotu & Desphande (2015), logistic regression is used to predict the value of the target variable in the form of binary (0 or 1) using numeric variable input. Hosmer & Lemeshow (2000) explains that the regression method has become an integral component of each data analysis related to the description of the relationship between the response variable and one or more explanatory variables.

Variable

According to Baragoin, Corinne, Christian M. Andersen, Stephan Bayerl, Graham Bent, Jleun Lee, and Christoph Schommer, Mining Your Own Business in Telecoms Using DB2 Intelligent Miner for Data, International Business Machines (2001), data sources that can be used to predict customers churn on cellular telecommunication service operator companies are: (1) churn indicators, (2) customer information data, (3) call data, (4) customer indices obtained from transaction data.

RESEARCH METHODS

Data Collection Tools

The details of the initial variables used are described as follows:

Table 1 (Part-I). Variable Data Table

No.	Research variable	Data Type	Description	Data source
1	FLAG_BLOCK_H5	Non-metric	Customers who are blocked for 5 days are called churn	Churn indicator
2	Device_model_change	Non-metric	Whether or not customers have changed their devices	Customer Information - Psychograph
3	Count_Bad	Non-metric	Whether or not the customers have overdue bills	Customer Information - Psychograph
4	Trx_voice_onnet_avg	Nominal	The average transaction for the use of voice calls among PT. XYZ Cellular users	Customer Call Data
5	Trx_voice_offnet_avg	Nominal	The average transaction of the use of voice calls to users other than PT. XYZ Cellular users	Customer Call Data

Table 1 (Part-II). Variable Data Table

No.	Research variable	Data Type	Description	Data source
6	Trx_sms_avg	Nominal	Average transaction of SMS usage	Customer Call Data
7	Trx_broadband_avg	Nominal	Average transaction of data usage	Customer Call Data
8	Mou_voice_onnet_avg	Nominal	Average voice calls usage among PT. XYZ Cellular users in minutes	Customer Call Data
9	Mou_voice_offnet_avg	Nominal	Average voice calls usage to users other than PT. XYZ Cellular users in minutes	Customer Call Data
10	Vol_broadband_avg	Nominal	Average data usage in Megabyte	Customer Call Data
11	Tot_bill_amount_avg	Nominal	Average bill in rupiah	Billing and Payment Data
12	Rev_voice_onnet_avg	Nominal	The average total usage of voice calls among users of PT. XYZ Cellular in rupiah	Billing and Payment Data
13	Rev_voice_offnet_avg	Nominal	The average total usage of voice calls to users other than PT. XYZ Cellular users in rupiah	Billing and Payment Data
14	Rev_sms_avg	Nominal	Average total usage for SMS services in rupiah	Billing and Payment Data
15	Rev_ir_avg	Nominal	The average usage of International roaming in rupiah	Billing and Payment Data

Stages of Research

There is an unbroken Outer Ring line illustrating that iterations are carried out continuously to improve the results. This is done to answer business challenges by taking lessons from previous iterations. According to Chapman, et al. (2000), the process in data mining consists of 6 stages, as seen in Figure 1.

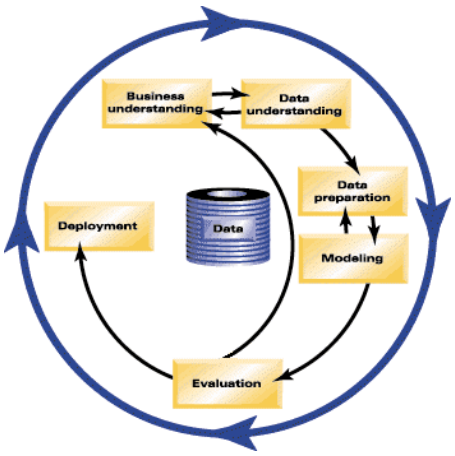

Figure 1. Stages of Designing a Churn Prediction Model

Source: Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0 Step-by-step data mining guide. SPSS Inc. Picked 2018

Leena and Luna International, Chikusei, Japan.
(株) リナアンドルナインターナショナル, 筑西市,日本
ISSN: 2186-845X ISSN: 2186-8441 Print
www.ajmse. leena-luna.co.jp

Population and Sample

The total population of the postpaid regular segment was 250,000 regular types on March 31, 2018 with a subscription period of more than 3 months. The amount of data was based on data received by researchers from PT. XYZ Cellular by paying attention to the confidentiality of data regulated in the policies of PT. XYZ Cellular in providing information to external parties. The data population consisted of churners and non-churners. The definition of churn in this study referred to customers who experienced late payments up to 5 days from the due date of payment set by PT. XYZ Cellular.

Data Collection and Data Sources

In this study the data used was derived from internal data of PT. XYZ Cellular. It was related to: (1) total bills, (2) conversation data, SMS, domestic and international data usage and (3) device change from January to March 2018. The data had gone through a process of aggregation according with the need to make predictive modeling. In the process of aggregation and preparation of analytics data mart, researchers used Teradata data warehouse tools. The data collection process is shown in Figure 2.

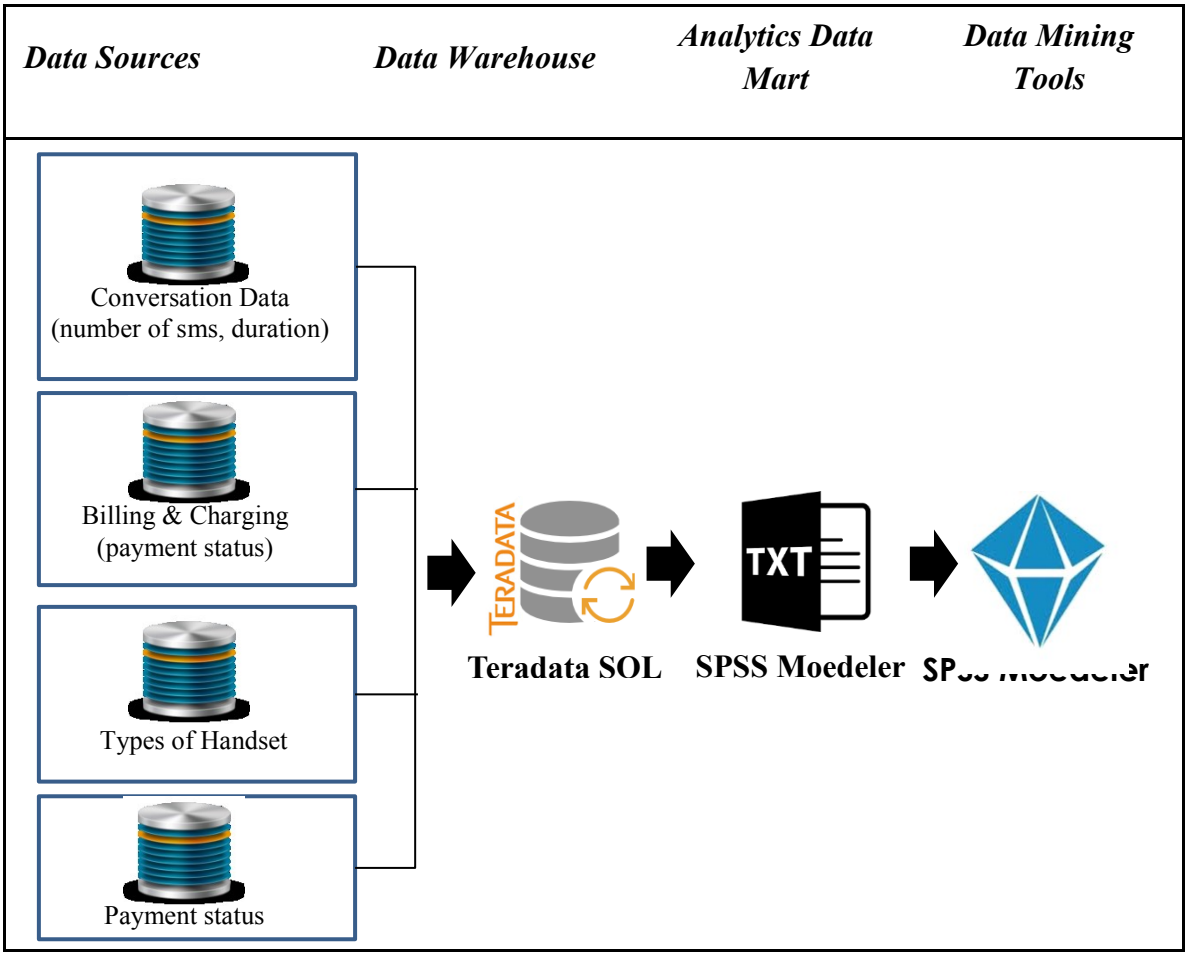

Figure 2. Stages of Data Collection

Model Validity

To test the validity of the model that has been made, the researchers used Confusion matrix. Confusion matrix is used to measure the validity of the model with binary target values (0 or 1). Each value of confusion matrix is obtained from the results of model testing.

Table 2. Confusion Matrix

		Predicted Condition	
		Not Churn	Churn
Actual Condition	Not Churn	TN (*True Negative*)	FN (*False Negative*)
	Churn	FP (*False Positive*)	TP (*True Positive*)

True positive rate (TPR) or sensitivity = $\frac{TP}{TP+FP}$

True negative rate (TNR) or specificity = $\frac{TN}{TN+FN}$

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

The higher the accuracy and sensitivity of a model is the better the model to be chosen.

RESEARCH RESULTS AND DISCUSSION

Research Results

After data processing using SPSS Modeler version 18.0, the following results were obtained:

Table 3. Iteration of Prediction Models Formation with Forwards Variable Selection Method Stepwise

		B	S.E.	Wald	df	Sig.	Exp (B)
Step 1a	trx_voice_offnet_avg	-0.063	0.001	3537.978	1	0	0.939
	Constant	1.739	0.009	35916.81	1	0	5.691
Step 2b	trx_voice_offnet_avg	-0.06	0.001	3136.664	1	0	0.942
	device_model_change	0.922	0.017	2815.893	1	0	2.514
	Constant	1.369	0.011	15874.47	1	0	3.932
Step 3c	trx_voice_offnet_avg	-0.059	0.001	2899.649	1	0	0.942
	device_model_change	1.069	0.018	3658.47	1	0	2.912
	COUNT_BAD(1)	-1.672	0.033	2630.657	1	0	0.188
	Constant	2.781	0.032	7636.251	1	0	16.133
Step 4d	trx_voice_onnet_avg	-0.007	0	1525.224	1	0	0.993
	trx_voice_offnet_avg	-0.05	0.001	1967.181	1	0	0.951
	device_model_change	1.037	0.018	3390.143	1	0	2.82
	COUNT_BAD(1)	-1.68	0.033	2628.34	1	0	0.186
	Constant	3.028	0.033	8500.054	1	0	20.662

Leena and Luna International, Chikusei, Japan.
(株) リナアンドルナインターナショナル, 筑西市,日本
ISSN: 2186-845X ISSN: 2186-8441 Print
www.ajmse. leena-luna.co.jp

Step 5e	tot_bill_amount_avg	0	0	486.898	1	0	1
	trx_voice_onnet_avg	-0.006	0	933.61	1	0	0.994
	trx_voice_offnet_avg	-0.043	0.001	1311.54	1	0	0.958
	device_model_change	1.037	0.018	3373.308	1	0	2.82
	COUNT_BAD(1)	-1.651	0.033	2532.882	1	0	0.192
	Constant	3.188	0.034	8913.846	1	0	24.245
Step 6f	tot_bill_amount_avg	0	0	463.677	1	0	1
	trx_voice_onnet_avg	-0.006	0	914.446	1	0	0.994
	trx_voice_offnet_avg	-0.043	0.001	1316.628	1	0	0.958
	trx_broadband_avg	-0.002	0	420.475	1	0	0.998
	device_model_change	1.022	0.018	3262.025	1	0	2.778
	COUNT_BAD(1)	-1.653	0.033	2532.858	1	0	0.192
	Constant	3.244	0.034	9104.613	1	0	25.627
Step 7g	tot_bill_amount_avg	0	0	508.412	1	0	1
	trx_voice_onnet_avg	-0.007	0	1110.153	1	0	0.994
	trx_voice_offnet_avg	-0.044	0.001	1344.429	1	0	0.957
	rev_voice_onnet_avg	0	0	209.423	1	0	1
	trx_broadband_avg	-0.002	0	408.886	1	0	0.998
	device_model_change	1.015	0.018	3209.851	1	0	2.76
	COUNT_BAD(1)	-1.656	0.033	2536.341	1	0	0.191
	Constant	3.234	0.034	9022.353	1	0	25.379
Step 8h	tot_bill_amount_avg	0	0	349.828	1	0	1
	trx_voice_onnet_avg	-0.007	0	1155.732	1	0	0.993
	trx_voice_offnet_avg	-0.044	0.001	1359.158	1	0	0.957
	rev_voice_onnet_avg	0	0	199.385	1	0	1
	trx_broadband_avg	-0.002	0	398.206	1	0	0.998
	rev_ir_avg	0	0	174.072	1	0	1
	device_model_change	1.012	0.018	3182.5	1	0	2.751
	COUNT_BAD(1)	-1.658	0.033	2538.356	1	0	0.191

	Constant	3.22	0.034	8924.647	1	0	25.028
Step 9i	tot_bill_amount_avg	0	0	351.204	1	0	1
	trx_voice_onnet_avg	-0.007	0	1171.342	1	0	0.993
	trx_voice_offnet_avg	-0.042	0.001	1170.059	1	0	0.959
	rev_voice_onnet_avg	0	0	226.209	1	0	1
	rev_sms_avg	0	0	43.948	1	0	1
	trx_broadband_avg	-0.002	0	396.903	1	0	0.998
	rev_ir_avg	0	0	175.229	1	0	1
	device_model_change	1.012	0.018	3179.536	1	0	2.751
	COUNT_BAD(1)	-1.657	0.033	2537.049	1	0	0.191
	Constant	3.233	0.034	8964.593	1	0	25.363
Step 10j	tot_bill_amount_avg	0	0	326.849	1	0	1
	trx_voice_onnet_avg	-0.007	0	1193.555	1	0	0.993
	trx_voice_offnet_avg	-0.036	0.002	515.688	1	0	0.965
	rev_voice_onnet_avg	0	0	237.533	1	0	1
	rev_voice_offnet_avg	0	0	33.83	1	0	1
	rev_sms_avg	0	0	39.096	1	0	1
	trx_broadband_avg	-0.002	0	394.797	1	0	0.998
	rev_ir_avg	0	0	176.748	1	0	1
	device_model_change	1.011	0.018	3170.833	1	0	2.747
	COUNT_BAD(1)	-1.657	0.033	2537.216	1	0	0.191
	Constant	3.237	0.034	8975.481	1	0	25.448
Step 11k	tot_bill_amount_avg	0	0	334.061	1	0	1
	trx_voice_onnet_avg	-0.007	0	1198.631	1	0	0.993
	trx_voice_offnet_avg	-0.036	0.002	508.755	1	0	0.965
	vol_broadband_avg	0	0	33.63	1	0	1
	rev_voice_onnet_avg	0	0	238.678	1	0	1
	rev_voice_offnet_avg	0	0	33.742	1	0	1
	rev_sms_avg	0	0	39.385	1	0	1
	trx_broadband_avg	-0.002	0	375.241	1	0	0.998
	rev_ir_avg	0	0	174.38	1	0	1
	device_model_change	1.012	0.018	3177.684	1	0	2.751
	COUNT_BAD(1)	-1.654	0.033	2526.895	1	0	0.191
	Constant	3.224	0.034	8873.468	1	0	25.119
Step 12l	tot_bill_amount_avg	0	0	346.005	1	0	1
	trx_voice_onnet_avg	-0.007	0	1128.906	1	0	0.993
	trx_voice_offnet_avg	-0.035	0.002	490.272	1	0	0.965
	trx_sms_avg	0.001	0	24.351	1	0	1.001

Leena and Luna International, Chikusei, Japan.
(株) リナアンドルナインターナショナル, 筑西市,日本
ISSN: 2186-845X ISSN: 2186-8441 Print
www.ajmse. leena-luna.co.jp

	vol_broadband_avg	0	0	30.798	1	0	1
	rev_voice_onnet_avg	0	0	243.591	1	0	1
	rev_voice_offnet_avg	0	0	31.026	1	0	1
	rev_sms_avg	0	0	63.442	1	0	1
	trx_broadband_avg	-0.002	0	373.966	1	0	0.998
	rev_ir_avg	0	0	169.289	1	0	1
	device_model_change	1.011	0.018	3169.461	1	0	2.748
	COUNT_BAD(1)	-1.657	0.033	2534.517	1	0	0.191
	Constant	3.223	0.034	8866.542	1	0	25.097
	tot_bill_amount_avg	0	0	321.736	1	0	1
	trx_voice_onnet_avg	-0.007	0	1135.97	1	0	0.993
	trx_voice_offnet_avg	-0.032	0.002	251.739	1	0	0.969
	trx_sms_avg	0.001	0	25.132	1	0	1.001
	vol_broadband_avg	0	0	30.403	1	0	1
	mou_voice_offnet_avg	-0.005	0.002	7.375	1	0.007	0.995
Step	rev_voice_onnet_avg	0	0	245.339	1	0	1
13m	rev_voice_offnet_avg	0	0	10.56	1	0.001	1
	rev_sms_avg	0	0	66.316	1	0	1
	trx_broadband_avg	-0.002	0	373.748	1	0	0.998
	rev_ir_avg	0	0	171.011	1	0	1
	device_model_change	1.01	0.018	3162.858	1	0	2.746
	COUNT_BAD(1)	-1.658	0.033	2535.892	1	0	0.191
	Constant	3.22	0.034	8840.608	1	0	25.025

a Variable(s) entered on step 1: trx_voice_offnet_avg.

b Variable(s) entered on step 2: device_model_change.

c Variable(s) entered on step 3: COUNT_BAD.

d Variable(s) entered on step 4: trx_voice_onnet_avg.

e Variable(s) entered on step 5: tot_bill_amount_avg.

f Variable(s) entered on step 6: trx_broadband_avg.

g Variable(s) entered on step 7: rev_voice_onnet_avg.

h Variable(s) entered on step 8: rev_ir_avg.

i Variable(s) entered on step 9: rev_sms_avg.

j Variable(s) entered on step 10: rev_voice_offnet_avg.

k Variable(s) entered on step 11: vol_broadband_avg.

l Variable(s) entered on step 12: trx_sms_avg.

m Variable(s) entered on step 13: mou_voice_offnet_avg.

By using the 13th iteration from Table 3 and using a mathematical logistic regression linear formula, then the following results were obtained:

$$\pi = \frac{e^{3.22+(0*Total\_bill\_amount\_avg)+\cdots.(1.658*Count\_Bad)}}{1+e^{3.22+(0*Total\_bill\_amount\_avg)+\cdots.(1.652*Count\_Bad)}}$$

From the total data set of 156,054 customers, it was predicted that 127,687 customers would churn. The top 5 significant variables from the model which affect customers to churn are illustrated in Figure 3.

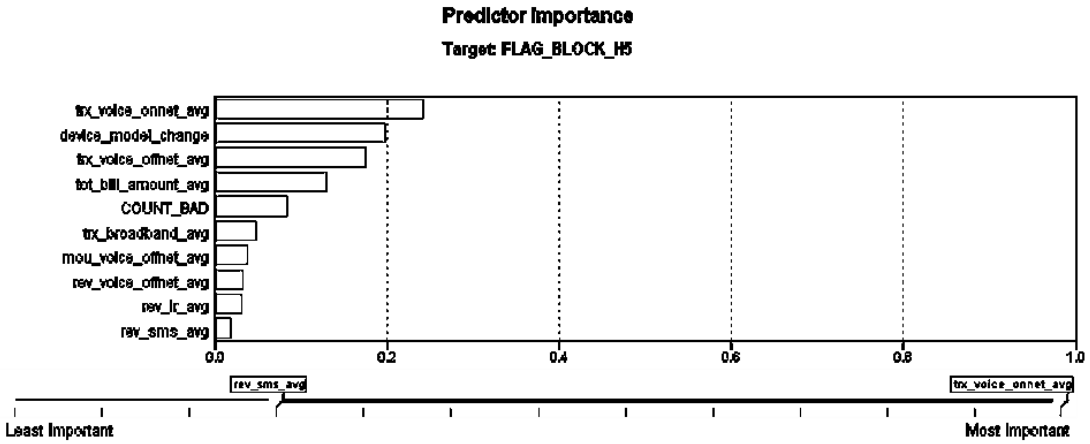

Figure 3. Visualization of Top 5 significant variables of the Logistic Regression Algorithm

After the churn prediction model with logistic regression algorithm was formed, the researchers evaluated the model using the measurement of accuracy as follows:

Table 4. Iteration of Prediction Model based on Input Variables

Iteration	Observed		Predicted		
			Flag_Block_H5		Percentage Correct
			0	1	
Step 1	Flag_Block_H5	0	720	18985	3.7%
		1	1055	88380	98.8%
	Overall Percentage				81.5%
Step 2	Flag_Block_H5	0	935	18770	4.7%
		1	1082	88353	98.8%
	Overall Percentage				81.7%
Step 3	Flag_Block_H5	0	1294	18411	6.6%
		1	1215	88220	98.6%
	Overall Percentage				81.9%
Step 4	Flag_Block_H5	0	1722	17983	8.7%
		1	1625	87810	98.2%
	Overall Percentage				81.9%
Step 5	Flag_Block_H5	0	1945	17760	9.9%
		1	1654	87781	98.2%
	Overall Percentage				82.1%
Step 6	Flag_Block_H5	0	2116	17589	10.7%
		1	1751	87684	98.0%
	Overall Percentage				82.1%
Step 7	Flag_Block_H5	0	2265	17440	11.5%
		1	1775	87660	98.0%
	Overall Percentage				82.3%
Step 8	Flag_Block_H5	0	2352	17353	11.9%
		1	1829	87606	98.0%
	Overall Percentage				82.3%

Leena and Luna International, Chikusei, Japan.
(株) リナアンドルナインターナショナル, 筑西市,日本
ISSN: 2186-845X ISSN: 2186-8441 Print
www.ajmse. leena-luna.co.jp

Step 9	Flag_Block_H5	0	2348	17357	11.9%
		1	1828	87607	98.0%
	Overall Percentage				82.3%
Step 10	Flag_Block_H5	0	2377	17328	12.1%
		1	1846	87589	97.9%
	Overall Percentage				82.3%
Step 11	Flag_Block_H5	0	2412	17293	12.2%
		1	1868	87567	97.9%
	Overall Percentage				82.3%
Step 12	Flag_Block_H5	0	2430	17275	12.3%
		1	1884	87551	97.9%
	Overall Percentage				82.3%
Step 13	Flag_Block_H5	0	2439	17266	12.4%
		1	1887	87548	97.9%
	Overall Percentage				82.3%

Table 4 shows that the total percentage of logistic regression models until iteration 13 is 82.3%. Furthermore, to measure the accuracy and sensitivity of the model, the researchers used confusion matrix.

Table 5. Confusion Matrix for Training Data Set

		Predicted Condition		
	Total Population	**Not Churn**	**Churn**	**Total**
Actual Condition	Not Churn	2,439	17,266	19,705
	Churn	1,887	87,548	89,435
	Total	4,326	104,814	109,140
Accuracy	82.5%			
Sensitivity:	97.9%			

Table 6. Confusion Matrix for Testing Data Set

		Predicted Condition		
	Total Population	**Not Churn**	**Churn**	**Total**
Actual Condition	Not Churn	1.051	7,611	8,662
	Churn	795	37,457	38,252
	Total	1,846	45,068	46,914
Accuracy	82.1%			
Sensitivity	97.9%			

DISCUSSION

Based on the evaluation results of the prediction model using confusion matrix for testing the data set, it was known that:

1. Based on the test results using Confusion Matrix, it was found that out of 46,914 customers, 45,068 customers or 96.1% of customers were predicted to churn and 1,846 or around 4% of customers would not churn.
2. When compared with the actual condition of total customers who were predicted to churn, 83% or 37,457 customers could be correctly predicted and the remaining 17% or 7,611 could not be correctly predicted.
3. Of the total customers who were predicted not to churn, there were 57% or 1,051 customers could be correctly predicted not to churn and 43% of the remaining 795 were incorrectly predicted.
4. The sensitivity or the ability to predict churn customers appropriately from the group of churners was 97.9%
5. The five best variables that could be used to predict customer churn are shown in Table 7.

Table 7. Top 5 Significant Variables

No	Input variable	Description	Normalized Sensitivity Score
1	Trx_voice_onnet_avg	The average transaction for voice usage among PT. XYZ Cellular users	0.24
2	Device_model_change	Whether or not customers have changed their devices	0.2
3	Trx_voice_offnet_avg	The average transaction of the use of voice to users other than PT. XYZ Cellular users	0.18
4	Tot_bill_amount_avg	Average customer bills	0.13
5	Count_Bad	Whether or not the customers have overdue bills	0.08

6. Using this model, PT. XYZ Cellular can determine the customers who must be given a retention program, especially the 37,457 customers who were predicted to churn. Even though there were 795 customers who did not get retention programs because they were predicted not to churn but actually they churn, the number was still considered small. The business risk was not too significant compared to the revenue obtained from the customer that had been successfully retained.

CONCLUSIONS AND RECOMMENDATIONS

Conclusion

Based on the results of data collection and the purpose of the study, the researchers conclude that:

1. Based on the data presentation, costumer churn occurred in the group of customers who changed devices, made late payments, had average bills of less than 6,775,862, and had an average international roaming usage of 3,195,332 to 6,195,332. This was indicated by the high probability of churn when compared with the customer group who never changed devices, never made late payments, had an average bill of 3,195,332 to 6,195,332, and had average international roaming usage less from 3,195,332.

Leena and Luna International, Chikusei, Japan.
(株) リナアンドルナインターナショナル, 筑西市,日本
ISSN: 2186-845X ISSN: 2186-8441 Print
www.ajmse. leena-luna.co.jp

2. Based on the test results using the confusion matrix, a churn prediction model using Logistic Regression algorithm, the sensitivity or ability to accurately predict customer churn of customer churn group was 97.9%
3. The five best variables that could be used to predict costumer churn were:
   a. The average transaction for voice usage among PT. XYZ Cellular users
   b. Whether or not customers have changed their devices
   c. The average transaction of the use of voice to users other than PT. XYZ Cellular users
   d. Average customer bills
   e. Whether or not the customers have overdue bills
4. Churn prediction score could be used as a way to determine retention strategies through the selection of customers who were predicted to churn. They should be given a retention program. Therefore, the retention program would be effective and the business risk could be minimized.

SUGGESTION

Based on the results of the study, some suggestions for PT. XYZ Cellular are stated as follows:

1. PT. XYZ Cellular should immediately apply the prediction models with logistic regression algorithms to reduce the postpaid customer churn in the regular segment, focusing on the customers who are predicted to churn

PT. XYZ Cellular can utilize variables that significantly influence customers to churn as a reference in making and implementing retention products.

REFERENCES

[1] Ahmad, S. A. (2015). *Customer loyalty is formed due to two main factors, namely customer trust and customer satisfaction*. Retreievd from https://www.researchgate.net/publication/326742546_Determinants_of_Customer_Loyalty_A_Review_and_Future_Directions.

[2] Bahari, F., & Elaydion, S. (2014). An efficient CRM-data mining framework for the prediction of customer behavior. *International Conference on Information and Communication Technologies*.

[3] Baragoin, B. (2001). Mining your own business in telecoms using DB2 intelligent miner for data. *International Business Machines.*

[4] Chapman, S. (2000). *CRISP-DM 1.0 Step-by-step data mining guide.* SPSS Inc. Retrieved from https://www.the-modeling-agency.com/crisp-dm.pdf.

[5] Guo, F., & Qin, H. (2017). Data mining techniques for customer relationship management. *Journal of Physics Conference Series*.

[6] Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression*. Canada: John Wiley & Sons, Inc.

[7] Kotu, V., & Desphande, B. P. (2015). *Predictive analytics and data mining.* Waltham, USA: Elsevier Inc.

[8] Philip, K., & Kevin, L. K. (2012). *Marketing management.* Retrieved from https://trove.nla.gov.au/work/4070618.

[9] Rekha, S. (2015). A data mining model for customer relationship management. *IJECS, 4* (8).

[10] Saini, N. (2016). Churn prediction in telecommunication industry using decision tree. *Streamed Info-Ocean, 1* (1).

Leena and Luna International, Chikusei, Japan.
(株) リナアンドルナインターナショナル, 筑西市,日本
ISSN: 2186-845X ISSN: 2186-8441 Print
www.ajmse. leena-luna.co.jp