International Journal of Engineering & Technology, 8 (1.9) (2019) 578-580

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Big Data Analytic as the Foundation of Customer Retention

Anung Asmoro[1]*, Lukito Edi Nugroho[2], Selo[3]

[1,2,3]Department of Industrial Engineering, Electrical Engineering and Information Technology Department Gadjah Mada University Yogyakarta, Indonesia
**Corresponding author E-mail: anung.s3te14@mail.ugm.ac.id*

Abstract

Most companies spend energy and resources to build new business, but for companies have the customer base should its focus is trying in defending and improve relations with customers. This study aims to maintain telecommunications fixed broadband services customers in Indonesia based on big data analytic. This research result indicates the big data analytic can help predict customers that will terminate a relationship subscription. The implementation of the big data analytic produce a list of customers that is predicted will terminate a relationship with the company. Knowledge was used as a base to know the main cause of subscription and stop continued with business needs to be done to prevent it.

Keywords: *Customer retention, big data analytic, resources, customer base, prediction*

1. Introduction

In general, in the various industries, the costs required to maintain one existing customer is one-fifth of the cost it takes to get one new customer[1]. Though increasingly tight competition efforts, rapid changes in the political, economic, social, and technology put the effort in retaining customers (customer retention) into a competitive position. Especially in the telecommunications business confronts between incumbent operator with other operators that offer the service and new package make customers have the option to decide subscription to then switch to other providers. According to SAS Institute Report in [2] rates of customers who decided to unsubscribe in telecommunication industry around 30% per year with an upward trend in correlation with the OD the growth market.

Many studies already conducted related research like this, [1] which presents a unified analytic framework. This framework is used to detect the early warning signs of the impending termination of the subscription. The method used is using an algorithm learning that gives value to every customer associated potential they will break in a certain time duration. The value of each customer is dynamic as a result of the learning algorithm execution is dynamic.

Research [3] using the company's database as a data source to run the scenario of customer retention. The goal is to predict whether customers will subscribe to cut ties in the near future or not. The billing data telecommunications company used as the basis of analysis of prediction. In research [4] describes the modeling predictions for churner based on data mining method. The study also discusses the many uses of the model tree analysis in detail. While the focus is the churn of business theorist associated with the process flow and modeling techniques.

On research [5] explains in detail the reduced customer case study using logistic regression. Logistic regression is used to predict the results based on the value of a discrete continuous variable or the category of logistics. In research [6] describes a case study using the Oracle database which contains 50 thousand customers wireless telecommunication industry to predict the termination of subscription. This research uses the Naive Bayes algorithm of dynamical approach. Research [7] using Mining Mart to do an analysis of unsubscribing. Focus research on data preprocessing using Mining Mart.

In contrast to previous studies that use various methods of prediction, this research uses the approach of big data analytic and it doesn't just stop at prediction, but also know the main cause and continued with the effort required to prevent it.

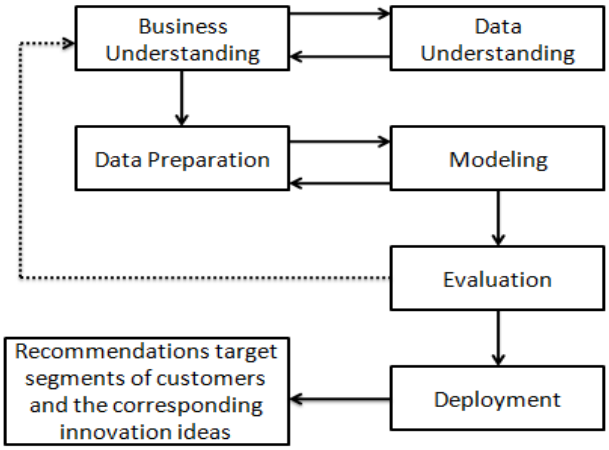

Fig.1: Core process of big data analytics.

Fig.1 shows a generic model of the big data analytic core process consisting of the ten stages. The ten stages of process can be summed up into five main stages, i.e., beginning with exploration and incorporation with an understanding of business and data, data preparation, model development, evaluation, and deployment. High rates of termination of subscription and move to other providers is a matter of business in General and fixed broadband, in particular, are the focus of this research. This problem is a result of the climate of business fixed broadband service competition happening in Indonesia. The competition is open in Indonesia begins with Act No. 36 of 1999 which mentions that the telecommunication's sector is already not a monopoly again. Thus, customers have many options and enable it to move between the provider and the services. This has to be one of the causes of increased levels of churn. This problem is translated into a big problem the analytic data, namely how to recognize signs of potential customers to decide to unsubscribe, to predict which customers are likely to do so.

Many databases that are owned by a telecommunication's company, one of the group's database identified related to the above problem is the Group's database of Customer behavior: the data profile, revenue, payment, usage, package, complaint, caring, and win back the customer. Understanding of data begins with a gather data, get in-depth data understanding, identify data quality problems, finding the initial idea of data usage, or to detect important data used to form a hypothesis of information. The condition of the customers with a breakup of the company is the condition the customer has submitted a request to unsubscribe or subscribers that are not retained after successfully conducted customer retention. In the billing database & collection are customers who already pay no bills for at least 2 consecutive months so as not to get the Bill/bill.

The next step after understanding the business, there is an understanding of the data and the preparation of the data. In general, the data needed is data: profile of customers, revenue, payment, usage, package, complaint, caring, and win back the customer. This data is sourced in a different database, including database: customers, Revenue, TREMS, MDF, ISISKA, and Caring. Then of all the related database, data pulled into one big table-big data, predictive models because the data is required in the form of a single table. After the data is prepared in a single big table of data in the data preparation process has been done, the next process related data is data cleansing (data cleansing), namely eliminating the outlawed data and handle missing and invalid value. After cleaning the data, then build data history for each variable. Proceed with making a variable derived from the historical data (average, etc.). This data is then merged in a big table. The big table is to be used in the modeling process.

Modeling beginning with an exploration of the data with the aim of knowing the data patterns that can be used to compile the modeling hypotheses. Then proceed with identifying and selecting the most important variable and explanatory variables. Identification and selection of the most important variables and identifiers are done in iterative. After it's done checking for multicollinearity.

Multicollinearity is the existence of a linear relationship between explanatory variables. The presence of multicollinearity is one of the causes of the occurrence of the modeling results of over fitting. The next step is to determine which data training and testing data that are later used in modeling, that data will be used to build the data. For the later models produced will be tested with data testing. The modeling was done in this study using the method of logistic regression, i.e., by matching data history customers with the functioning of the logistic curve with the help of logic software SPSS Modeler.

2. Discussion and Conclusion

The main objective of this work was to maintain customers telecommunications fixed broadband services in Indonesia based on big analytics data. This research uses the approach of big data analytics and it doesn't just stop at prediction, but also know the main cause and continued with the effort required to prevent it. The result of this is a big data analytic predicted customers who will likely perform the termination of subscription. And so easy to understand, the results are visualized in the form of a table. To be more easily visualized in the form of a report in the form of dashboards. One example of this is the deployment results as illustrated in Figure 2 below.

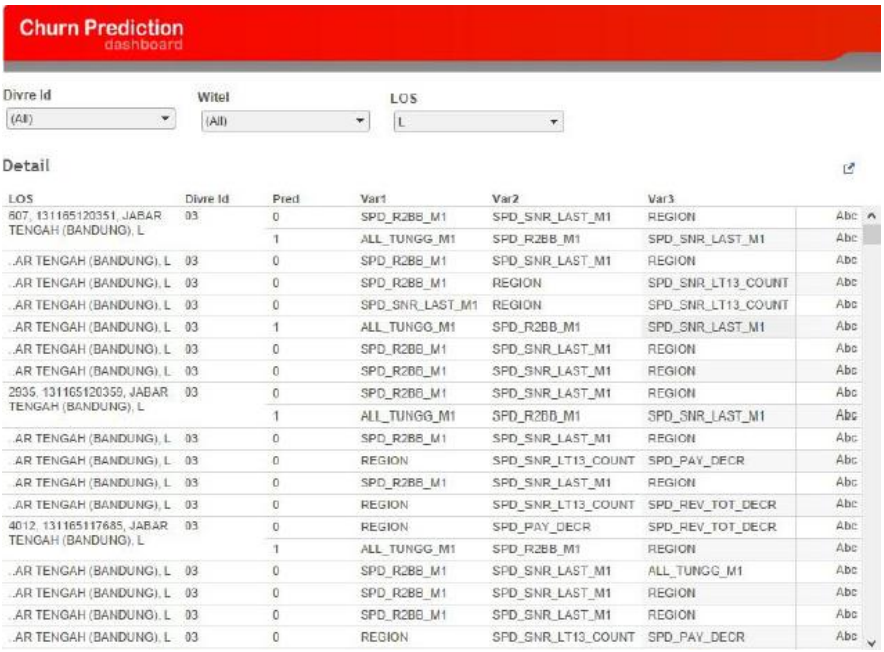

Fig.2: Visualization of results of deployment.

Whereas after 360-degree application integrates seamlessly with Customer care in the form of a dashboard report, illustrated in Figure 3 below.

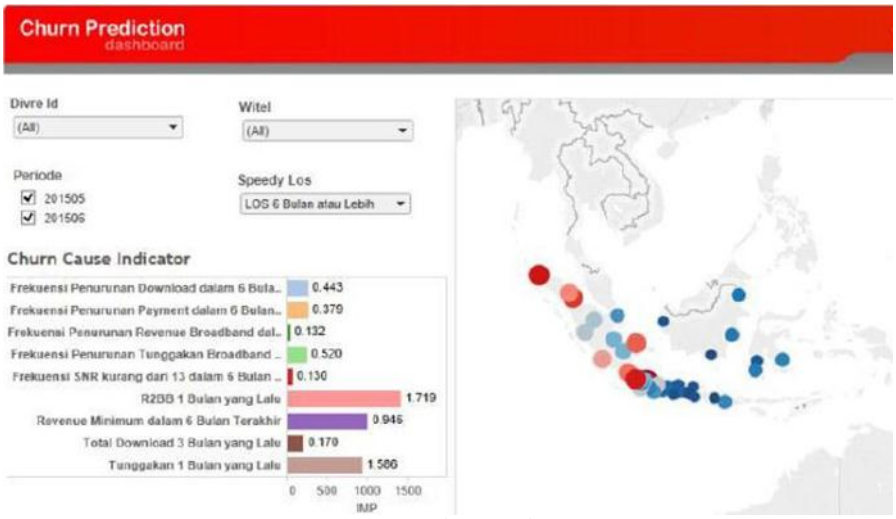

Fig.2: Visualization of results of deployment.

This research shows a big data analytic can help predict which customers will subscribe to end the relationship. Implementation of Big Data Analytic to generate a list of customers who have predicted will end the relationship with the company. Further knowledge is used as a basis to know the main causes of unsubscribing and continued with the efforts that need to be done to prevent it.

References

[1] E. Indicators, "Behavioral Modeling for Churn Prediction : Early Indicators and Accurate Predictors of Custom Defection and Loyalty," pp. 677–680, 2015.
[2] M. Richeldi and A. Perrucci, "Churn Analysis Case Study," 2002.
[3] N. A. Haris, O. Support, and S. Division, "Data Mining in Churn Analysis Model for Telecommunication Industry," vol. 1, no. 19, pp. 19–27, 2010.\
[4] N. Forhad, S. Hussain, R. M. Rahman, and A. Definition, "Churn Analysis : Predicting Churners," pp. 237–241, 2014.
[5] S. V Nath, "Customer Churn Analysis in the Wireless Industry : A Data Mining Approach Customer Churn Analysis in the Wireless Industry : A Data Mining Approach," no. 561, pp. 1–20.
[6] T. Mutanen, "Customer churn analysis – a case study," 2006.
[7] U. Yabas, H. C. Cankaya, and T. Ince, "Customer churn prediction for telecom services," Proc. - Int. Comput. Softw. Appl. Conf., pp. 358–359, 2012.