ABSTRACT

This research is based on condition in PT Telkomsel that we found some of internet data users changed to be data user to be non-data users. This is partly due to the desire of customers to benefits. In addition, competition between tele-communication providers is also one of the causes associated with offering more competitive prices and more favorable benefits.

With this research, the author wants to provide insight to Telkomsel in predicting data customers who will change to be non-data user accurately using predictive analytics. Then segmentation is made against the results of predictions based on customer behavior, in order to classify those customers that have the potential to change into non user data with appropriate service offerings.

Data resources derived from internal data Telkomsel related to customer behavior including total recharge, data usage, SMS, lapse information, data conversations, in the period January to March 2019. Data processing is done through data playing with modeling logistic regression using SPSS data modeler.

Logistic regression model is used to predict customers who will lapse from user data becomes non data users by using 14 input variables with numeric types. The based table analysis is taken from customer data that was reported in April 2019 amounting to 40,991 subscribers that will move become lapse and took customer samples which remained data customers for the month of 12,297 customers out of a total of 390,567 customers. The Analytics based table is divided into 2 data sets, namely 70% training data set and 30% testing data sets for both lapse and non-lapse customers from user data. So, from the total training & testing data set of 53,288 customers, it was predicted that 44,818 customers would lapse into non data users with model accuracy reaching 86% and resulting in a top 6 significant variable namely, Day of Broadband Average, Total Recharge Average, Length of Stay Average, Volume quota chat Average, Video Average Volume quota, and Payload Average.

The results of segmentation using k-means, obtained the number of clusters as much as 5 with silhouette index 0.43, which means the validity of segmentation is in the fair category.

The results of this study are expected to provide an overview of the model and prediction results of lapse and also the characteristics of prepaid customer segmentation that will lapse into non-user data, and can also be used by telecommunications operators in providing special offers and marketing programs that are appropriate to the characteristics of the segment.

Keyword: *Market Segmentation, Customer behavior, marketing strategy, regression, K-Means.*