

ABSTRACT

One of the challenges faced by PT Telkom Indonesia in the process of transition to FTTH services is how to retain current customers and attract customers to migrate toward Indihome product. Telkom currently has a home customer service around 8.3 million, and only about 26% are profiling well. These conditions make some difficulties for company to identify and recognize customer that potential to give future profit. Therefore, company needs a model to segment customers for determining priority of customer migration process to FTTH product so that can maximizing profit from the existing customer. Based on Pareto 80/20 rule, companies that can find 20% of profitable customers will be able to sustain its business revenue.

This research suggests to apply clustering process using principal component analysis (PCA) combined with K-Means, then continue it with classification process using customer lifetime value (CLV) combined with support vector machine and logistic regression. The primary data source is based on call data record and payment record of the existing customer. First, the primary data will be analyzed using PCA. Some attribute will be selected based on significant number from PCA then K-means algorithm will form cluster. Optimal cluster is determined by Calinski Harabasz and Davies Bouldin index. For each cluster, CLV score will be computed to define whether it is upper cluster that have potential future profit or lower cluster that needs more effort to migrate toward Indihome. Last, a classification model using SVM and logistic regression is build so that when there is new customers data, the process should not start from beginning.

As a result, the best model is the principal component analysis as the feature selection technique. The clustering measurement method to define optimal cluster is using Calinski Harabasz. While, for building a classification model a logistic regression gives the best accuracy result and improvement of performance algorithm with the F-Test score of 99.16%, the accuracy of about 99.93%, and the precision of 98.84%. The principal Component Analysis as the feature selection can improve the performance logistic regression up to 2% compared to non PCA implementation and RFM model. While, the top list of profitable customers that can be inferred using this model is 33.62% of total population.

Keyword: Principal component analysis, Customer lifetime value, K-Means, Support Vector Machine, Logistic Regression