Telecommunication Customer Segmentation Based on the Profitability with a Novel Data Mining Approach

Sujah M.A. Ameer[1*], Kapila R.M.[1], Tharanga Rathnayaka[2]

[1]Department of Computing and Information Systems, Faculty of Applied Sciences, Sabaraga-muwa University of Sri Lanka, Belihuloya, Sri Lanka, [2]Department of Physical Sciences and Technology, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka

[*]E-mail: ameersujah@gmail.com

Technological advancement and globalization give high priority to telecommunication industry and it is highly competitive because of multiple service providers that provide different solutions to their customers. As a result, customers are rapidly moving from one service provider to another based on their requirements. Furthermore, human communications have been moving far from traditional to modern such as Skype, FaceTime and Social Media. Therefore, mobile operators are under a real threat of loosing revenue. To solve this issue, they need to increase their capabilities on understanding customer behavior patterns and preferences. The major aim of this study is to cluster the customers based on the profit which service provider earns and develop a model to predict customer profitability level in the future and clustering the customers to provide different promotions. This study was carried in three phases. First phase was the comparison of K-means and K-means++ approaches of clustering algorithm and choosing the best one by using Within Cluster Sum of Square (WCSS) and processing time. The result was K-means++ algorithm. Second phase was focusing on clustering the customers based on their behaviors by using K-means++ and developing an Artificial Neural Network (ANN) model to predict customer profitability level in the future. Finally, choose the highest profitable customer cluster and apply K-means++ algorithm to provide different promotions. Dataset consists of 12,000 prepaid customer details of service provider-x with 15 attributes such as monthly service charge, monthly outgoing call duration, duration of roaming calls, customer care calls, customer lifetime revenue, device's web capability and so. Confusion matrix was used to evaluate the performance of ANN model and the constructed ANN model gave the accuracy of 97.53%. Existing researches use unsupervised or supervised learning algorithms separately. But this study integrates both algorithms and getting high prediction accuracy of 97.53%. Therefore, this model fits well for telecommunication industries.

Keywords: Profitability, clustering, neural network, K-means