

# An Unsupervised Machine Learning Approach for Churn Prediction

S. Prasanth<sup>1</sup>, R. M. K. T. Rathnayaka<sup>2</sup>, H. Arumawadu<sup>3</sup>

<sup>1a</sup>Department of Computing and Information Systems, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka

<sup>2</sup> Department of Physical Science and Technology, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka

<sup>3</sup>Virtusa (Pvt) Ltd, 752 Dr Danister De Silva Mawatha, Colombo

---

Customer churn is one of the critical issues faced by the firms nowadays. Telecommunication industry is no exception to this rule. In this industry, keeping the existing subscriber (customer) is more valuable than acquiring a new subscriber (attracting new customers costs approximately 5 times higher than retaining the existing customers). Therefore, predicting the attrition behavior of customers in advance is a significant task. This behavior has triggered most of the researchers to focus on developing the churn prediction model in several industries. Anyhow, in most of the time supervised machine learning techniques have been incorporated in this regard. But in here, an unsupervised machine learning approach has been proposed. A local telecommunication company can be approached for the purpose of conducting this research. Around 10,000 postpaid subscriber details with 20 attributes have been obtained and analyzed during this research. Further, Principal Component Analysis (PCA) and K-means clustering algorithm have been utilized with the intention of reducing the dimensionality between features and to find the churners and non-churners respectively. The results obtained from the PCA have revealed that, 16 principal components which represent all the 20 features are considered as most important aspects to cover the entire data. Moreover, totally 6 clusters have been generated and some particular features that tend to show high contributions were identified during the principal component analysis have been analyzed towards each cluster. The proposed approach has finally revealed that out of the 6 clusters three (3) representing 4888 are churners and the other three (3) representing 5112 are non-churners. It could be ensured that, this approach would assist the future researchers to have a promising start for combining the unsupervised technique with the supervised one.

*Keywords: Churn; Unsupervised Machine Learning; Principal Component Analysis; K-Means Clustering*

---

Corresponding author: Tel.: +94-77-072-1234

E-mail address: senthanprasanth007@gmail.com