

An Unsupervised Machine Learning Approach for the Identification of the Churn Behavior of Customers

06 Nov.
ET23

S. Prasanth^{1(*)}, R. M. K. T. Rathnayaka²

¹*Department of Computing and Information Systems, Faculty of Applied Sciences, Sabaragamuwa 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*

(*)Email :senthnanprasanth007@gmail.com

It is true and practical that customers are the most precious asset for any successful organization in the present business world. Therefore, it is inevitable for an organization to retain its customer base in a stable position. In order to do this, a very effective churn prediction strategy should be followed by the organization concerned. As per certain churn predictions carried out in the past, researchers in this field had been motivated in establishing and administering much more successful businesses. Most of the instances in the business fields, the supervised machine learning techniques had been taken into consideration in the process of churn prediction. On the contrary to the above, the unsupervised machine learning techniques have been utilized in this study. In this process, 10, 000 post-paid subscriber details from a local "Telecommunication Company" which consists of 20 attributes have been obtained and analyzed. Further, the Principal Component Analysis (PCA) and K-Means clustering algorithm have been employed aiming at minimizing the dimensionality between the said attributes and to make clusters to find the churners and non-churners respectively. The results obtained from the above mentioned PCA have revealed that the specific 16 principal components represent the whole 20 features that are considered as important to cover the entire data. Further, totally 6 clusters have been generated and it was noted that some particular features tend to show higher contribution. This characteristic was identified during PCA. This was further analyzed through each cluster. In conclusion, the proposed approach revealed that out of the 6 clusters three (3) representing 4888 are churners and the other three (3) representing 5112 are non-churners.

Keywords: Churn, Principal Component Analysis (PCA), K-Means clustering, Unsupervised Machine Learning