# MACHINE LEARNING APPROACH FOR CUSTOMER CATEGORIZATION USING DATA MINING TO INCREASE PROFITABILITY

Sujah A.M.A.[1*] and Rathnayake R.M.K.T.[2]
[1]Department of Computing & Information Systems, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Sri Lanka
[2]Department of Physical Sciences & Technology, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Sri Lanka
*ameersujah@gmail.com

Customer plays a crucial role for the success of any businesses. It is difficult to survive in the competitive business environment without attracting the customers. Since, most successful businesses continuously do the research and development part for their customers in the fields of customer identification, customer attraction, customer retention, and customer development to achieve a high level of customer relationship. All the competitors are trying to make more customer profit to survive in their business. To achieve this, businesses required to increase their capabilities on understanding customer behavioural patterns and preferences. The major aim of this study is to categorize the customers based on behaviours and develop a prediction model to predict future customer categories. The current study is carried under the two phases. Initial phase is focusing on clustering the customers based on their behaviours by using K-means++ algorithm. Recency, Frequency and Monetary (RFM) attributes are used to cluster the customers and the second phase is to develop the Artificial Neural Network (ANN) model to predict future customer's category based on their usage behaviours. Dataset consists of 5,000 customer details in a particular business with RFM attributes to cluster, train and test the model. K-means++ algorithm used to cluster and the final weighted cluster centroids are calculated based on the weights of RFM attributes. Target attributes are generated by analysing the final weighted cluster centroids for the ANN model. Confusion matrix used to evaluate the performance of ANN model. Existing researches applied unsupervised or supervised learning algorithms separately. But this research study integrates both. Therefore, this model well fit for any business industries without any errors.

***Keywords:*** *Customer relationship, K-means++, Profitability, RFM attributes*