

# A hybrid machine learning model for predicting customer churn under presence of uncertainty,

## Abstract

Nowadays, with regard to competitiveness and positioning in an unpredictable environment, the industry seeks to do its utmost to provide the customer with the services they need to compete in the long term and maintain their customers. Also, the cost of new customer attraction is much higher than the cost of maintaining their current customer. Companies may be able to handle a number of customers in a variety of ways, but they will not be well preserved, which is why marketers who have been thinking of attracting their customers so far have tended to keep them up to date. There are different approaches to customer retention, but the most acceptable way is to find people who want to leave the organization, which is referred to as a churned customer.

Some of the customers, with or without known reasons, leave the organizations. Correct prediction of churn customers, or in other words, the position of the customers in relation to the organization in any time period, can have a significant role in the management of costs. The purpose of this research is to reduce the organization's loss by correctly predicting churn customers. The Sina Bank is the case study for this research. To achieve the purpose, after preprocessing the data over six-month periods, three Classes of Loyal, tendentious to churn, and churning were defined, and monthly customers' transactions were allocated to one of the three classes. The ranking of customers in each period of time in three Classes is the innovation of this research. In other words, customers at any time period in this study were assigned to three groups Loyal, tendentious to churn, and churning. This allocation of customers' positions in each period in three modes is the innovation of this research. Then, by using classification techniques such as Decision Tree, Random Forest, Bayesian, Neural Networks, Logistic Regression, Linear Regression, and Support Vector Machine, the accuracy and efficiency of the proposed model in two classes and three classes model were analyzed. For determining the model's performance, the three classes model was compared with two classes approach by a set of criteria such as Accuracy, Recall, Precision, and Cost index. The results showed the effectiveness of the proposed model compared to the two classes approach, and the costs were less than the two classes approach. In both methods, the Support Vector Machine technique has the lowest cost.