

**CUSTOMER CHURN PREDICTION IN THE BANK
SECTOR: A STUDY BASED ON CREDIT CARD USERS**

**Kaushali Thripathy Mohotti Bandara Rathnayake
Mudiyanselage**

A dissertation submitted in partial fulfilment of the requirement for
Master of Science degree in Business Analytics

**Department of Computing
Informatics Institute of Technology, Sri Lanka
in collaboration with
Robert Gordon University, Aberdeen**

2022

Abstract

In the banking industry, technological advancement has been accelerating at a rapid rate as of late. To better manage their financial and property assets, people all around the world use banking services. At this point, all the technological improvements are being implemented in the banking sector in order to provide the customers with proper operational excellence. According to this point of view, the bank must provide its customers with cutting-edge applications that will help them save both time and money.

Therefore, the bank needs to conduct an analysis of the value of its customers in order to boost its marketing growth and increase its revenue. However, the banking industry continues to struggle with the prediction of customer churn, which is a hard issue for the sector when it comes to measuring the rise of earnings. With this point of view, our main objective is to make predictions regarding the number of customers that cancel their credit card accounts within the banking sector. This research paper uses the public available churn modeling data set extracted from Kaggle.

Hence, this research was initiated in order to identify what factors have impacted credit card customer churn and how those factors behave in this data set while having the main objective of the study is to build a predictive model based on several machine learning algorithms and to create an optimum model which using ensemble classification voting methods. At the end of this research, the algorithm CAT Boost (categorical boost) has seen to have performed (98% accuracy) overall the best without feature selected and Extra tree machine learning models have been performed (95% accuracy) with feature selected the best model while having certain tradeoffs as well, however. Furthermore, combined several models to create ensemble models that make use of different voting techniques. However, it was observed that these models did not outperform some of the individual machine learning models.