

**CHURN PREDICTION OF PREPAID
TELECOMMUNICATION CUSTOMERS IN SRI LANKA**

Hasitha Pathirage

A dissertation submitted in partial fulfilment of the requirement for
Master of Science degree in Advanced Software Engineering

School of Computing

Informatics Institute of Technology, Sri Lanka

in collaboration with

University Westminster, UK

2023

ABSTRACT

Customer is the most important asset to a company, so it is important to retain existing customers. Acquiring new customers is known to be more expensive than retaining existing customers. Because of this it is important to know beforehand whether a customer will leave the service or churn to take necessary precautions. Due to the current economic crisis in Sri Lanka customer behaviors are changing and all of the sectors including telecommunication is getting affected. So, it is important for telecommunication industry to identify churning customers. Furthermore, there is no research conducted on the prepaid customer churning in Sri Lankan domain and it is hard to identify prepaid customer churning due to the non-contractual behavior of the prepaid domain.

Considering above factors and filling the gaps found in churn prediction domain a churn prediction model for prepaid customers in Sri Lanka is built and tested. This model uses both neural networks, boosting classifiers with hybrid resampling methods to answer the class imbalance problem that occur in churn prediction dataset. The model is trained with a true feature rich dataset collected from a leading telecommunication service provider in Sri Lanka. The research investigates data preprocessing, three feature selection methods, 6 resampling methods which includes under sampling, over sampling and hybrid sampling. Also 30 classifiers were compared and tested. Furthermore, the research looks into snapshot ensemble method to improve neural network results.

The final model trained with SMOTE-ENN hybrid resampling method was able to achieve accuracy of 98.60%, precision of 98.64%, recall of 98.58%, F1 score of 98.61%, AUC score of 99.60% showing high and outstanding scores for the prediction model. Furthermore 1.1% performance improvements were measured with the use of snapshot ensemble.

Keywords: Machine learning, Boosting classifiers, neural networks, resampling.