

Informatics Institute of Technology
in Collaboration with University of
Westminster, UK



Delphi – A Telco Churn Prediction System based on Ensemble Learning

A dissertation by

Mr. Apiseg Rajendran

w1761737 / 2018086

Supervised by

Mr. Torin Wirasingha

May 2023

Submitted in partial fulfilment of the requirements for the BSc (Hons)
Computer Science degree at
the University of Westminster

Abstract

Predicting customer churn in the Telco industry is a major problem. Customer churn refers to the percentage of customers who terminate their services with a company in a given period, often due to dissatisfaction or the availability of better alternatives. Predicting churn is essential for companies to retain customers and improve their services. However, accurately predicting churn is a challenging task due to the complexity of the factors that influence customer behaviour. Traditional machine learning algorithms may not provide sufficient accuracy, and the computational cost of ensembling them may be prohibitively high. Therefore, this research proposes a solution that utilizes ensemble learning with synthetic data for efficient and accurate churn prediction.

To solve this problem, we propose a stacking ensemble model that combines six machine learning algorithms: RF, KNN, SVM, logistic regression, XGBoost and CNN. Each algorithm contributes to the final prediction, and their predictions are combined through a logistic regression meta model. Ensembling improves prediction accuracy by reducing the risk of overfitting and increasing model robustness. However, hyperparameter tuning is a crucial step in ensembling that involves selecting the optimal combination of hyperparameters for each algorithm. This process is computationally expensive and may not be feasible for large datasets. To address this challenge, we propose to use distributed processing with Optuna on AWS SageMaker to optimize hyperparameters efficiently.

The proposed ensemble model with distributed processing for hyperparameter tuning is expected to provide accurate predictions of customer churn in the Telco industry. The performance of the model will be evaluated using several metrics, including accuracy, auc, recall, and F1-score, on a large dataset. The results will be compared against traditional machine learning algorithms and other state-of-the-art churn prediction models to demonstrate the effectiveness of the proposed approach. The outcome of this research will provide insights into the effectiveness of ensemble learning and distributed processing for customer churn prediction and help Telco companies to make informed decisions to retain customers and improve services.

Keywords: customer churn, ensemble learning, distributed processing, hyperparameter tuning, machine learning algorithms.

Subject Descriptors: Telco industry, customer behaviour, predictive analytics.