INFOCHURN- CUSTOMER RETENTION ENGINE WITH CHURN PREDICTION FOR MOTOR INSURANCE IN SRI LANKA A CASE STUDY TO CEYLINCO GENERAL PLC

Alpitiya Ranasingha Arachchilage Thilini Prarthana

A dissertation submitted in partial fullfilment of the requirement for
Bachelor of Science (Honours) degree in Business Information Systems

Department of Computing

Informatics Institute of Technology, Srilanka

in collaboration with

University of Westminster, UK

2020

Abstract

Companies are striving to explore new ways to retain their clientele to sustain their business growth due to economic volatility and heated competition in the insurance industry around the world. Retaining customers is more cost-effective than acquiring new customers thus the customer churn, a critically addressed topic in both industry and academia.

Despite being a globally addressed problem, the Sri Lankan insurance industry has lagged in leveraging technology-based solutions to outsmart churn and strategically retain its clients. Instead, on the verge of trusting their gut feeling of "the customer will renew", the backdoor of the companies are losing millions of money and customers. This is majorly problematic in the motor insurance segment where the renewal period is only a year. Following the case study methodology, this study was aimed at analyzing the churn behaviour of motor insurance customers of Ceylinco General Insurance, a leading insurance company in Sri Lanka and to implement a data-driven solution to retain their most revenue generated customer segment.

The broader discipline of data mining using machine learning techniques was critically screened in deriving a churn prediction solution. The literature survey was used to analyse features, algorithms and methodologies of existing systems under a common evaluation criterion. Findings from the literature were reviewed with the industry experts to understand that feature engineering needs to be revised since there is a gap of motor insurance customer details captured in Sri Lanka compared to that of other countries. With using a real-life data set containing 478,157 records of motor insurance customers with 55 features, an extensive data preparation process was carried out. Pre-processing tasks included scaling data, handling class imbalance with SMOTENN sampling technique while the most important features were selected using Pearson's correlation. Rather than having only a binary classification approach to the churn prediction like the majority of the existing solutions, the author attempted regression models such that the churn result is depicted also as a percentage value between 0 to 100%. From the classification algorithms that were selected, the Gradient Boosting classifier derived the highest F1 score of 91.59%. From the regression algorithms that were selected, Random Forest Regressor performed the best. The results are aimed at maximizing profits by