

MSC PROJECT REPORT

ECOMMERCE CUSTOMER CHURN TIME TO EVENT PREDICTION SYSTEM USING
XGBOOST AND SURVIVAL ANALYSIS

Yalage Chathura Buddhika Jayathilake

A report submitted as part of the requirements for the degree of MSc Big Data Analytics at Robert Gordon
University, Aberdeen, Scotland

Abstract

An important step for a success in any type of company is customer retention. Studies have shown that hiring new recruits is 27% costlier than keeping the existing customers. Therefore every company has addressed the issue of churning. This is truer for ecommerce companies as there are many competitors in the market due to several facts like the ease of entering, can sell irrespective of the geographical location, delivery methods etc. So in order to thrive in the ecommerce world, companies must put a lot of effort to identify the churning customers and make proactive actions to retain them. In order to take proactive actions, identifying the churning customers is not enough. Hence the time of churning has also considered in this study. Several approaches including different ML models, biased datasets and prediction types have been used in previous studies in order to achieve similar goals and they were discussed in the literature review. To conduct the study ecommerce purchasing data of 99456 customers of a Brazilian ecommerce site have been recorded which has 8 features over a period of 6 months. By using purchase frequency and the wait between the purchases, the churning customers have been identified for the dataset. In order to get the time of churning, this study has used survival analysis techniques together with XGBoost machine learning method to identify the highest probable period of churning in a customer's lifetime. There are many future works possible to carry forward this way of research and those are discussed in the future work section. Furthermore this type of study can be applied to any of the ecommerce companies which record the purchase data of customers and their customer features.