

Predicting customer churn for a telecommunication company

Ayesha Ahmed¹, Aidan Slingsby¹, Daniel Hulme¹, Guy Lansley¹ and Dan le Marquand²
¹University College London, ²National Broadband Provider

Project Background

Customer retention is not an easy task for modern businesses in today's highly competitive world. Customer churn is a common issue in every industry because consumers are very keen on securing good customer service, offers and additional services, alongside high-quality products. It is difficult to retain customers who have decided to move to a competitor. Hence, predicting which customers are likely to churn out can be of great value to a company in terms of building and planning measures which would help them to retain their customers and to prevent losses.

Data and Methods

The data used for this study consisted of 127,000 company-customer interaction data, customer churn data, revenue data, customer geographic and demographic data, opt in/out history, and web usage data. The data were highly unbalanced with only 93.5% non-churned and 6.5% churned customers. To resolve this issue, we sampled the data using 3 sampling techniques: over-sampling, under-sampling, and Synthetic Minority Over-sampling technique (SMOTE). These techniques were combined with machine learning tools like Logistic Regression, Random Forests, Support Vector Machine, Decision Trees and Naive Bayes to predict the customer churn. The methodology is summarized in Figure 1.

Key Findings

The sampling techniques and machine learning tools revealed SMOTE to be the best sampling technique along with Random Forest which yielded 90.7% accuracy; Support Vector Machine yielded 89% accuracy; and Naive Bayes contributed the least accuracy with 78%. The other performance measures were F-1 Score, precision, recall and receiver operating characteristic (ROC) curves.

Figure 2, provides the comparison for the final test-3 models by the testing data predictions. After the exploratory analysis and data modelling, we located the top important variables which influence customer churn: customers with a tenure period of 36 months, monthly website visits, low revenue band customers, broadband download speed, faulty web pages viewed, and total emails received. Additionally, sending no emails, outbound calls, or direct mail to customers can increase customer churn by almost 50%.

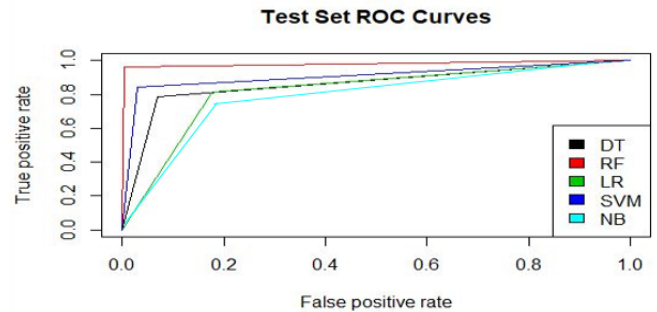


Figure 2: ROC curves

Value of the Research

The results of this study demonstrate the importance of data sampling techniques to overcome data imbalance and improve the performance of prediction. If implemented by telecommunication companies, the predictive models created in this work could represent an opportunity for businesses to take precautions to prevent the losses which occur due to customer churn. Also, by analysing the results, companies could better understand customer behaviour in general and create plans for wide range of customers with similar behaviours.

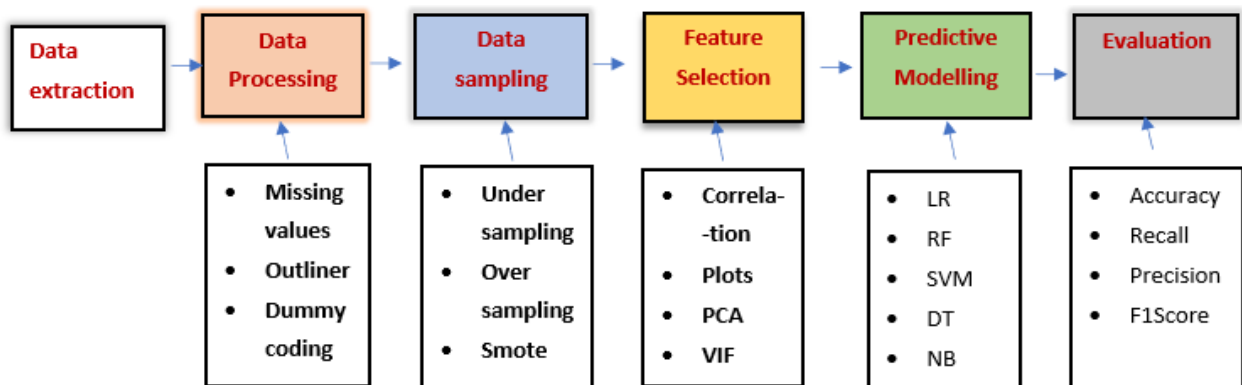


Figure 1: Flow diagram for the methodology