OPIM 5604 - Predictive Modeling

Telecom Customer Churn Analysis

(Team 4 - Vinay Reddy, Yoonhye, Reza Rafa, Tarun Kumar)

Telecom companies are facing increasing competition in the market, and customer retention is critical to their success. The objective of this project is to analyze customer churn in a telecommunications company using various customer demographics and behavioral variables. The goal is to predict the likelihood of a customer leaving the company, or churning," based on their usage patterns and other characteristics.

The analysis includes a dataset of 7,043 customers with 33 variables, including dependents, contracts, tenure months, partners, paperless billing, total charges, tech support, payment method, multiple lines, online security, streaming TV, internet service, streaming movies, online backup, senior citizens, phone service, gender, CLTV, monthly charges, and device protection.

The data was preprocessed, including handling missing values, encoding categorical variables, and scaling numerical features. The data set was partitioned with a split of 60%, 20%, and 20% in training, validation, and testing, respectively. The data set was also stratified with a split balance of 50% in the churn label. For exploratory data analysis, multivariate and distributional analyses were conducted to understand the relationships between variables and the target variable (churn label). In addition to that, the Y by X feature mosaic plot was explored to identify the relationship between the categorical and the target variable, Churn Label. The analysis was done based on the two statistical measures, which are likelihood and P value.

Logistic regression, boosted forest, neural models, decision tree, bootstrap, KNN, discriminant, naive bayes, and ensemble models were trained and evaluated on the dataset. The ensemble model outperformed the other models with an accuracy of 83.22%, a lift of 2.746, an accuracy of 73%, and an AUC of 0.8733 in predicting the churn status of a customer. The most important variables in the model were dependents, monthly charges, tenure months, contracts, and internet service.

The analysis concludes that customers who have high monthly charges, shorter tenure, and subscribe to fiber-optic Internet service are more likely to churn. On the other hand, customers with long tenure, dependents, year-on-year contracts, and low monthly charges are less likely to churn. Recommendations for reducing churn include offering incentives for longer-term contracts, offering referral discount and coupons to retain the customer and developing targeted retention strategies for customers who are at high risk of churn.

Overall, by leveraging customer data, advanced analytics, and predictive modeling, we can identify key churn drivers, forecast, and predict customer churn. The project provides insights into the factors that drive customer churn in the telecommunications industry and provides a framework for predicting churn and developing targeted retention strategies.