

Introduction

Customer churn, or customers dropping a company's product or services, is a top issue for companies across all industries. Not only does it impact revenues, but customer loss also increases the cost of acquiring new customers. Companies having knowledge about which customers will churn today allow them to practice proactive retention. This project will produce a machine learning model that will forecast customer churn using data gathered from Kaggle. From customer behavior and demographics, the model will identify if customers are likely to churn or not, and that would be useful information for companies to make improvements on customer retention.

Project Objectives

The primary goal of this project is to produce an accurate and effective machine learning model for customer churn prediction. Specific activities involve preprocessing and exploring the dataset to handle missing values and outliers, feature engineering for improved predictability, and training and comparing several classification models to compare performances. The models will be evaluated on accuracy, precision, recall, and ROC-AUC as measures of validity. The project also aims to visually represent significant findings to enable companies to view what causes churn and create targeted retention strategies.

Methodology

The project will be carried out sequentially from preprocessing the data to clean up and organize the dataset for analysis. It involves using missing value processing, encoding categorical variables, and numerical feature scaling. Feature engineering will be done following this in order to create new variables, such as "Support Calls per Tenure" and "Average Monthly Spend," which can provide more information regarding the behavior of customers. Different models of machine learning like Logistic Regression and Random Forest will be trained and cross validated to determine the best model. Cross validation techniques will be utilized to ensure that the model will perform well for unseen data. Finally, analytical illustrations, including unique graphical plots will be created to provide a clear presentation of the results in a visual format, which will emphasize the important trends and patterns.

Ethical Considerations

This project may pose some ethical concerns. One such worry is data bias, particularly in instances where the data set contains minimal observation of churned customers. To avoid this problem, techniques like class weight rescaling will be used to adjust the importance of different categories when classifying a class. Companies should also not rely solely on model estimates since it can lead to overaggressive or irrelevant retention. Probability scores instead of binary estimates can be able to include risk by giving space for human judgment in making decisions.

Projected Impacts

The impacts from this project, aim to have a reliable churn forecast model that will be able to determine possible risk customers. The model will be elevated by extensive tracking of drivers most related to churn, such as multiple support calls or overdue payments. The data will enable companies to develop tailored retention programs, including rebates or enhanced customer service. This project also provides results presented in a readable format, so that the stakeholders can easily interpret and act upon the results.

Conclusion

Customer churn prediction is a very important asset to a company's toolbox with the goal of achieving high customer retention and low revenue loss. The project uses the application of machine learning to create a very precise but also moral forecasting model. By taking a strong stance and not being open to possible bias, this project is aimed to garner answers that organizations can rely on. The final product of the project will be a pretrained model, detailed report, and visualization model to assist in future churn predictions.