CUSTOMER CHURN ANALYSIS

SYRIATEL CUSTOMER CHURN

By Susan Nduta Kanyora.

Introduction

Churning refers to the phenomenon where customers discontinue using a service or product. Customer churn can have a significant impact on business performance and profitability. Identifying customers who are likely to churn allows businesses to take proactive measures to retain those customers and minimize revenue loss.

Overview

The Syriatel project is aimed at predicting customer churn in the telecom industry to improve business performance and profitability. By analyzing customer behavior, usage patterns, and service interactions, we can identify customers who are likely to discontinue using the service. Through exploratory data analysis and the application of logistic regression and ensemble methods, we aim to develop accurate churn prediction models. This will enable us to proactively take measures to retain valuable customers and minimize revenue loss. The project focuses on leveraging data science techniques to address the business problem of customer churn and enhance decision-making processes in the telecom industry.

Business and Data Understanding

The objective is to predict customer churn and provide insights into customer behavior and factors influencing churn. The dataset used contains customer information, usage patterns, and service interactions. Exploratory data analysis is conducted to understand correlations and patterns, followed by the development of classification models using logistic regression and ensemble methods. Model performance is evaluated using metrics such as accuracy, precision, recall, and F1-score. The expected outcomes include accurate churn prediction models, insights into key factors influencing churn, and recommendations for targeted retention strategies to minimize customer churn and improve business performance.

Exploratory Data Analysis

Exploratory Data Analysis

During the exploratory data analysis (EDA) of the Syriatel dataset, we conducted various tasks to gain insights into customer behavior and churn. We addressed missing values, performed data cleaning, and examined descriptive statistics to understand the dataset's characteristics. Through correlation analysis and visualizations, we identified relationships between variables and gained valuable insights into factors influencing customer churn. These findings will guide our modeling efforts and help develop strategies for customer retention. The EDA sets the foundation for our predictive analysis and provides a solid understanding of the dataset.

Exploratory Data Analysis

Churn Distribution
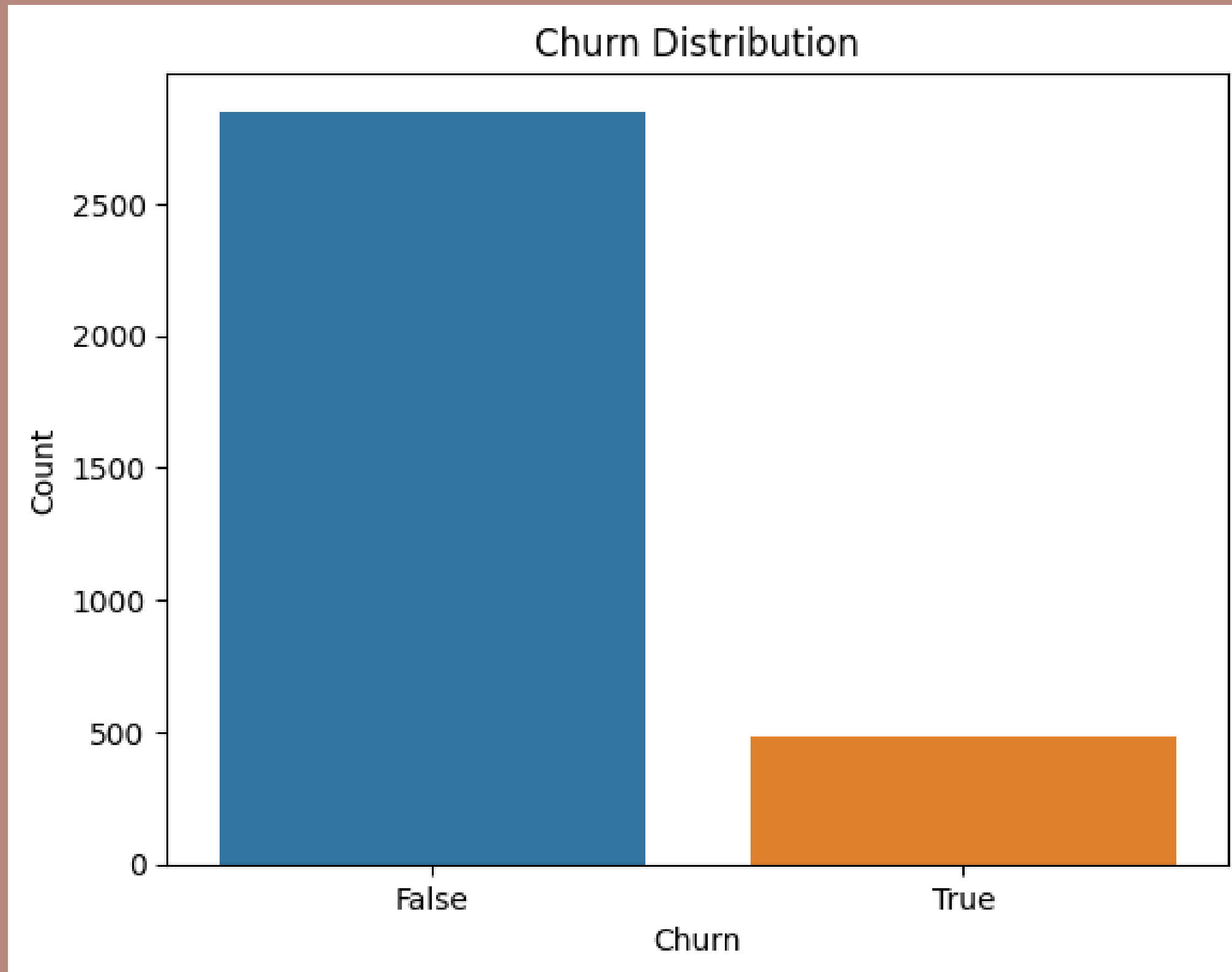
The graph suggests that the majority of customers did not churn, while a smaller proportion of customers churned.

Modelling

Our models were designed to be accurate and insightful while being easily understandable for non-technical stakeholders. We employed logistic regression as a baseline model, achieving an accuracy of 86%. To further improve our predictions, we explored ensemble methods such as Random Forest, SVM, and Boosting classifiers. Among them, Random Forest stood out with an impressive accuracy of 95.5%, surpassing the SVM classifier's accuracy of 84.9%. We evaluated our models using metrics like precision, recall, and F1-score, providing a comprehensive understanding of their performance. Additionally, we conducted feature importance analysis to identify the key factors influencing customer churn. Our findings suggest that the Random Forest classifier is a valuable tool for identifying potential churned customers.

Evaluation

The models were assessed using various evaluation metrics, including accuracy, precision, recall, and F1-score. The accuracy metric provided an overall measure of how well the models predicted customer churn. Precision measured the proportion of correctly identified churned customers among all predicted churn cases, while recall measured the proportion of correctly identified churned customers among all actual churn cases. The F1-score, which combines precision and recall, provided a balanced evaluation of the models' predictive power. Additionally, visualizations such as the confusion matrix, ROC curve, and precision-recall curve were employed to gain a deeper understanding of the models' classification results and their ability to correctly identify churned customers.

Conclusion

The exploratory data analysis conducted on the Syriatel dataset revealed correlations among various numerical variables, indicating relationships among customer behaviors. A logistic regression model achieved an accuracy of 86% in predicting customer churn. To enhance the prediction accuracy, ensemble methods including Random Forest, SVM, and Boosting classifiers were employed. Among these, the Random Forest classifier outperformed the others with an accuracy of 95.5%, surpassing the SVM classifier's accuracy of 84.9%. Model evaluation metrics such as precision, recall, and F1-score provided insights into the models' performance. Visualizations such as the confusion matrix, ROC curve, and precision-recall curve facilitated the interpretation of the models' classification results. Feature importance analysis identified the significant variables in predicting customer churn. Overall, the findings indicate that the Random Forest classifier is a promising approach for identifying potential churned customers.

Recommendations

-    **1**   **Customer Service Improvement.**

-    **2**   **Offer Incentives for International Plan.**

-    **3**   **Targeted Marketing Campaigns**

-    **4**   **Proactive Customer Retention Strategies**

-    **5**   **Continuous Monitoring and Analysis**