
 **ndutakanyora** commit 4 hours ago  


53 lines (29 loc) · 5.02 KB


Preview


Code


Blame


Raw











# SyriaTel Customer Churn

---

By: Susan Nduta Kanyora



## Project Overview

---

The project aims to leverage machine learning techniques for analyzing telecommunications data and customer behavior patterns in order to gain insights and improve services.

## Business Problem

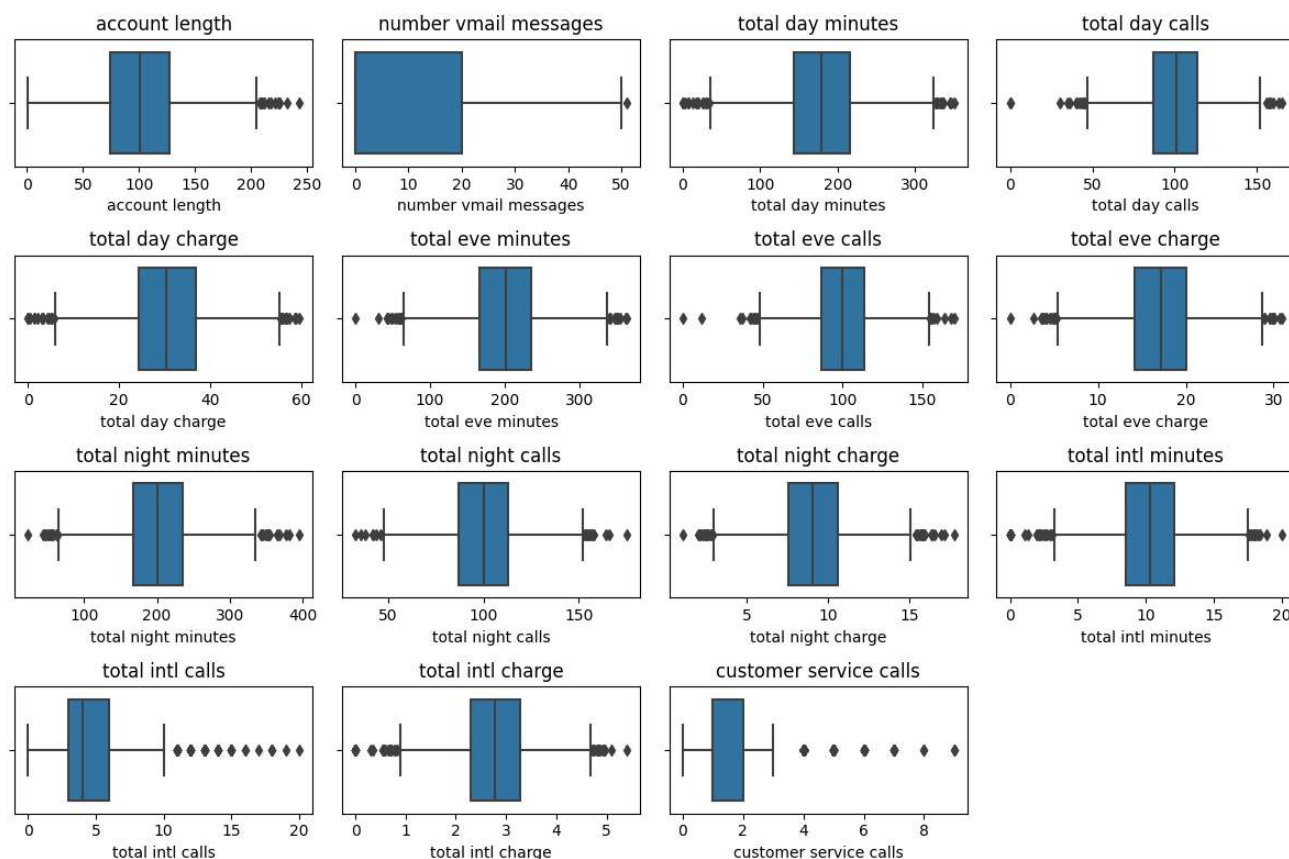
The project focuses on accurately identifying potential churners among SyriaTel's customer base. By employing machine learning techniques, the project aims to predict customers who are at risk of churning, enabling SyriaTel to implement targeted strategies and interventions to retain those customers, mitigate revenue loss, and enhance overall business performance and profitability.

## The Data

The dataset to be used is the SyriaTel Customer Churn dataset.

## Data Cleaning and EDA

Before delving into the analysis, some data cleaning procedures were performed to ensure data quality. These included managing missing values, handling duplicates, and addressing outliers. Once the data was cleaned, the EDA focused on investigating various aspects of the industry, such as customer churn. Customer churn refers to customers discontinuing the use of services, and understanding its impact on business performance and profitability was a key objective. The EDA aimed to identify factors and patterns associated with customer churn, enabling businesses to take proactive measures to retain customers and minimize revenue loss. Furthermore, the analysis explored correlations between different variables in the dataset, providing insights into the relationships between customer characteristics, service usage, and churn rates.



## Modelling training and evaluation

---

The analysis involved building and evaluating multiple models to predict customer churn in the telecom industry using the given dataset. The initial data exploration revealed that minimal data cleaning was required, with a focus on handling missing values, duplicates, and outliers.

The exploratory data analysis (EDA) aimed to answer important questions related to churn prediction, such as the relationship between price and the month of sale, the impact of view on price, and identifying the factors most correlated with price. This helped gain insights into the dataset and understand the patterns and relationships within the data.

Several models were trained and evaluated, including logistic regression, random forest, support vector machines (SVM), and boosting classifiers. These models were assessed using various evaluation metrics such as accuracy, precision, recall, and F1-score to determine their performance in predicting churn.

The logistic regression model achieved an accuracy of 86% and a precision of 65%, indicating its ability to correctly classify churned customers. The random forest model outperformed others with an accuracy of 95%, suggesting its strong predictive power. The SVM model achieved an accuracy of 84.8%, while the boosting classifier showed promising results with an accuracy of 90.7% and a precision, recall, and F1-score of 69.3%.

The models' evaluation provided valuable insights into customer churn prediction and identified the important features contributing to churn. The feature importance analysis highlighted the variables that had a significant impact on churn, allowing for targeted retention strategies.

Overall, the models demonstrated their effectiveness in predicting churn using the provided dataset. However, further improvements can be made by fine-tuning the models, incorporating additional features, and gathering more data. Regular monitoring and refinement of churn prediction strategies are crucial to effectively address customer churn in the telecom industry.

## Conclusion

---

A logistic regression model was trained to predict customer churn, achieving an accuracy of 86%. To enhance prediction accuracy, ensemble methods such as Random Forest, SVM, and Boosting classifiers were employed. Among these, the Random Forest classifier outperformed the others with an impressive accuracy of 95.5%, surpassing the SVM classifier's accuracy of 84.9%. Model evaluation metrics including precision, recall, and F1-score provided valuable insights into the models' performance. Visualizations such as the confusion matrix, ROC curve, and precision-recall curve aided in comprehending the models' classification results. Additionally, feature importance analysis identified the significant variables in predicting customer churn. The findings highlight the Random Forest classifier as a promising approach for identifying potential churned customers.

## Recommendations

---

Here are some of the recommendations as: 1.Customer Service Improvement. 2.Offer Incentives for International Plan. 3.Targeted Marketing Campaigns. 4.Proactive Customer Retention Strategies. 5.Continuous Monitoring and Analysis.