

MAZSUSA / SYRIATEL-CUSTOMER-CHURN-ANALYSIS

<> Code

Issues

Pull requests

Actions

Projects

Wiki

Security

Insights

Settings

Analysis of the customer churn rate at Syriatel

☆ 0 stars

🔗 0 forks

👁 1 watching

🌿 1 Branch

🏷 0 Tags

📈 Activity

🌐 Public repository

🔗 1 Branch

🏷 0 Tags

🔗

🏷

🔍 Go to file

t













Go to file

+



Add file

Code

⋮

 MAZSUSA	final notebook ammend	b00621c · yesterday	
 IMAGES	adding into gitignore	yesterday	
 .gitignore	modifying gitignore	yesterday	
 CM.png	images	yesterday	
 CUSTOMER CHURN RATE.png	images	yesterday	
 PRESENTATION_SYRIATEL.pdf	presentation pdf	1 minute ago	
 README.md	Added images	yesterday	
 SyriaTel.csv	Adding data understanding	last week	
 image-1.png	images	yesterday	
 image-2.png	images	yesterday	
 index.ipynb	final notebook ammend	yesterday	

📖 README



SYRIATEL CUSTOMER CHURN RATE ANALYSIS



Business Understanding

Introduction: The project aims to address a critical issue faced by SyriaTel, a telecommunications company, by developing a classifier to predict which customers are likely to stop using their services in the near future. This binary classification problem focuses on identifying patterns that signal customer churn.

Stakeholders: The primary stakeholders for this project are the management and customer retention teams at SyriaTel. These stakeholders can utilize the insights gained from the classifier to pinpoint at-risk customers more accurately. By understanding the predictive patterns of customer behavior, they can implement targeted interventions to improve retention rates and reduce revenue loss.

Conclusion: By uncovering and analyzing predictable patterns in customer churn, this project provides valuable information that can help SyriaTel enhance their customer retention strategies. The implications for the company include reduced financial losses associated with customer attrition and improved overall profitability through more effective customer engagement and retention efforts.

Data Understanding

Data Sources and Suitability:

The dataset for this project is sourced from Kaggle and pertains to customer churn rates within a telecommunications context. It provides detailed information about customer behaviors and attributes, which are essential for predicting churn. This dataset includes various features related to customer demographics, service usage, and account details, making it a comprehensive resource for analyzing patterns that lead to customer attrition.

Dataset Size and Descriptive Statistics:

The dataset comprises 3333 records and includes several features such as:

- account length

- area code
- phone number
- the different plans
- Number of messages and calls
- The call charges
- Customer service calls
- Churn

Feature Justification:

Features included in the dataset are justified based on their relevance to understanding customer behavior. For instance, customer tenure is a critical indicator of potential churn, as longer tenures often correlate with higher retention rates. Service plans and billing information provide insights into customer satisfaction and usage patterns, which are pivotal for identifying churn risks. Each feature contributes valuable information for developing a predictive model.

Data Limitations:

Despite its utility, the dataset has some limitations. It may not capture all factors influencing churn, such as external market conditions or changes in service quality. Additionally, if the dataset is not representative of the entire customer base, predictions may be biased. These limitations should be considered when interpreting the results and developing strategies based on the model's predictions.

Data Preparation

Data preparation steps included converting column names to lowercase, replacing 'False'/'True' with 0/1 for the churn column, handling missing values and duplicates, and encoding categorical variables.

Data Modelling

Three models were compared: Logistic Regression, Decision Trees and Random Forest

Logistic Regression:

- Achieved an accuracy of 85.57%, precision of 58.33%, recall of 20.79%, F1-score of 30.65%, and AUC of 0.59.
- The confusion matrix revealed 21 true positives, 551 true negatives, 15 false positives, and 80 false negatives.

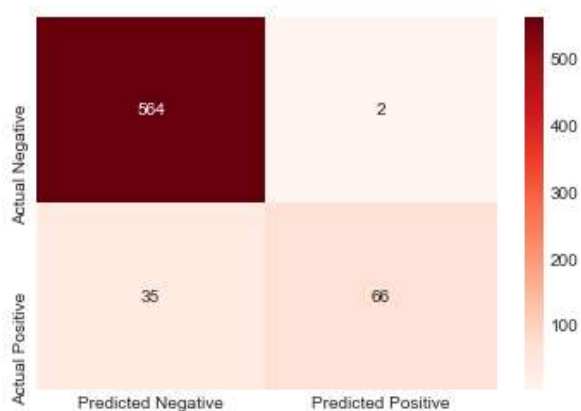
Decision Trees

- The model achieved an accuracy of 92.5%.
- The confusion matrix revealed 74 true positives, 543 true negatives, 23 false positives, and 27 false negatives.

Random Forest:

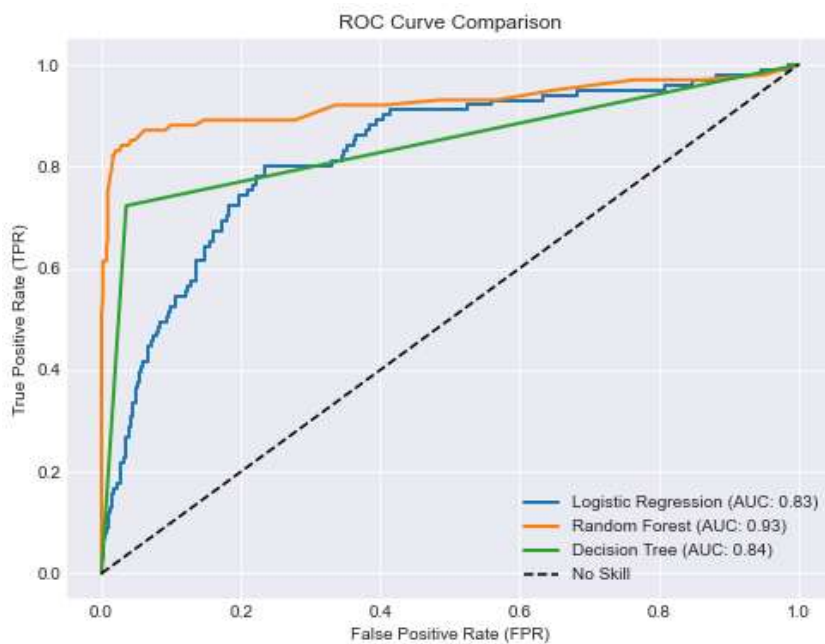
- Achieved an accuracy of 94.45%, F1-score of 78.1%, and AUC of 0.82.

- The confusion matrix revealed 66 true positives, 564 true negatives, 2 false positives, and 35 false negatives.
 - Hyperparameter tuning was performed using GridSearchCV to optimize the Random Forest model. Random Forest was the best performing model

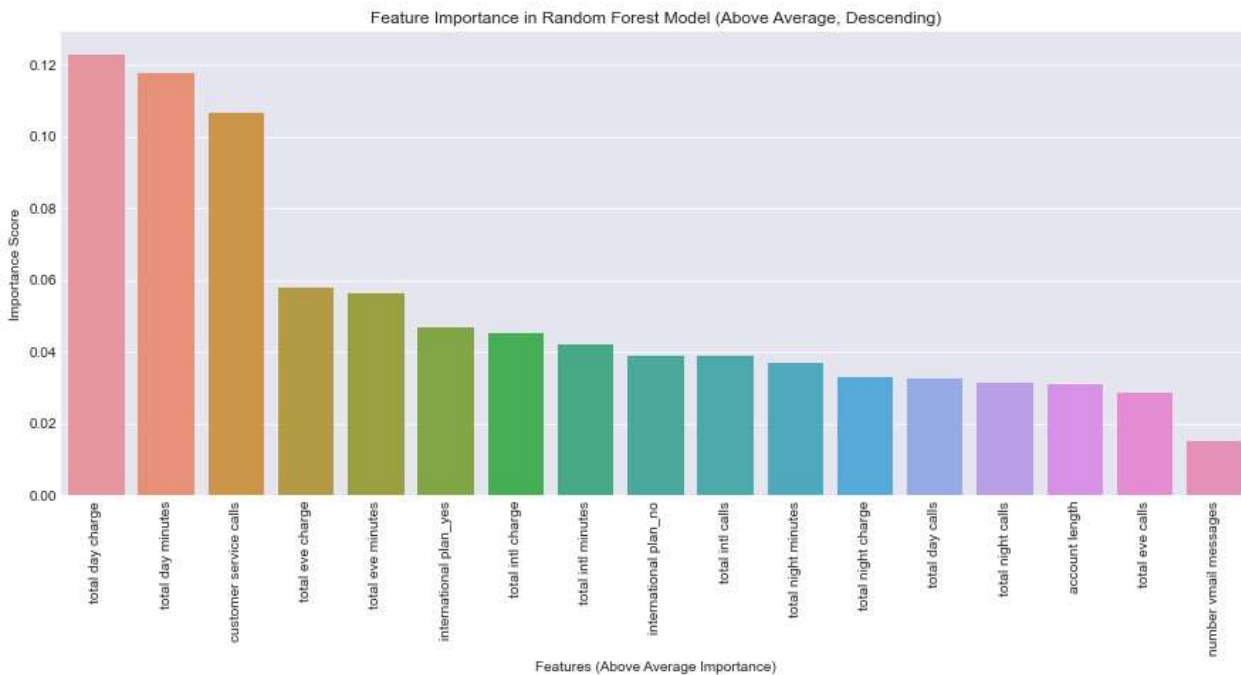


Model Evaluation

ROC curves were used to evaluate and compare the performance of the Logistic Regression and Random Forest models. The Random Forest model demonstrated superior performance by achieving a higher AUC compared to Logistic Regression.



Important features in the data set



- The top features are:
 - Total day charges and minutes
 - Customer service calls
 - Total evening charge and minutes
 - International plan

CONCLUSION

Random Forest model appears to be the best model to predict the customers likely to churn.

- The churn prediction analysis conducted for SyriaTel aimed to develop a classifier to identify customers likely to terminate their services. Through comprehensive data exploration, preparation, and modeling, several key findings emerged:
- Model Performance: Random Forest emerged as the most effective model for churn prediction, outperforming Logistic Regression and Decision Trees. It exhibited superior accuracy and predictive power, making it the preferred choice for SyriaTel's churn prediction system.
- Key Predictive Features: Total day charges and minutes, customer service calls, total evening charges and minutes and subscription to the international plan were identified as crucial indicators of churn. These insights provide valuable guidance for SyriaTel in devising proactive retention strategies targeted at high-risk customers.

RECOMMENDATIONS

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Languages

● Jupyter Notebook 100.0%