



SYRIATEL TELECOMMUNICATION COMPANY CUSTOMER CHURN PREDICTION

Business Overview

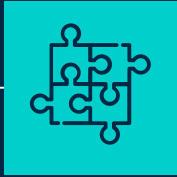
- SyriaTel, a telecommunication company is facing a challenge with increasing customer churn. They wish to get data-driven insights to understand the factors leading to the high churn rate and find ways to mitigate them.
- The analysis aims to use customer usage patterns, to identify customers that are at-risk of churning and factors that influence their decision to leave.
- By being able to predict churn and address customer's concerns promptly we will be able to foster customer loyalty, satisfaction and reduce churn.

PROBLEM STATEMENT

- SyriaTel, a telecommunications company, is experiencing a significant increase in customer churn. This phenomenon is adversely affecting the company's revenue and necessitates high acquisition costs for new customers



OBJECTIVES



01

Develop a predictive model to forecast customer churn.



02

Uncover patterns influencing churn decisions.



03

Enhance customer retention strategies based on data-driven insights

DATA UNDERSTANDING

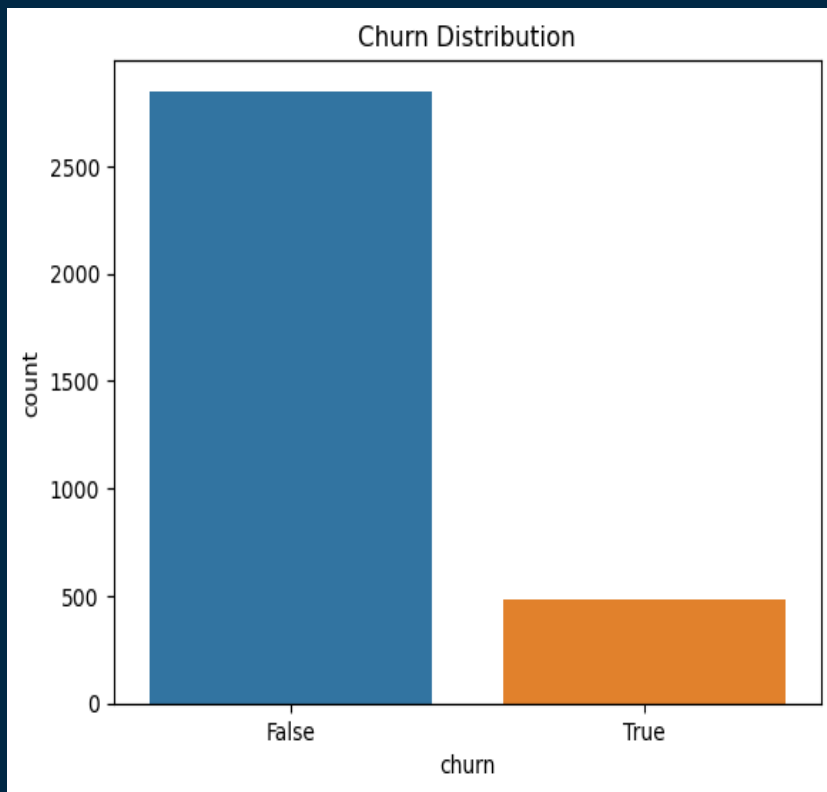
The data is obtained from Kaggle and is from SyriaTel Telecommunications

The target column is churn is a bool with True and False as values making this a binary classification problem.

The data has 21 columns and 3333 rows of identifiers of customers



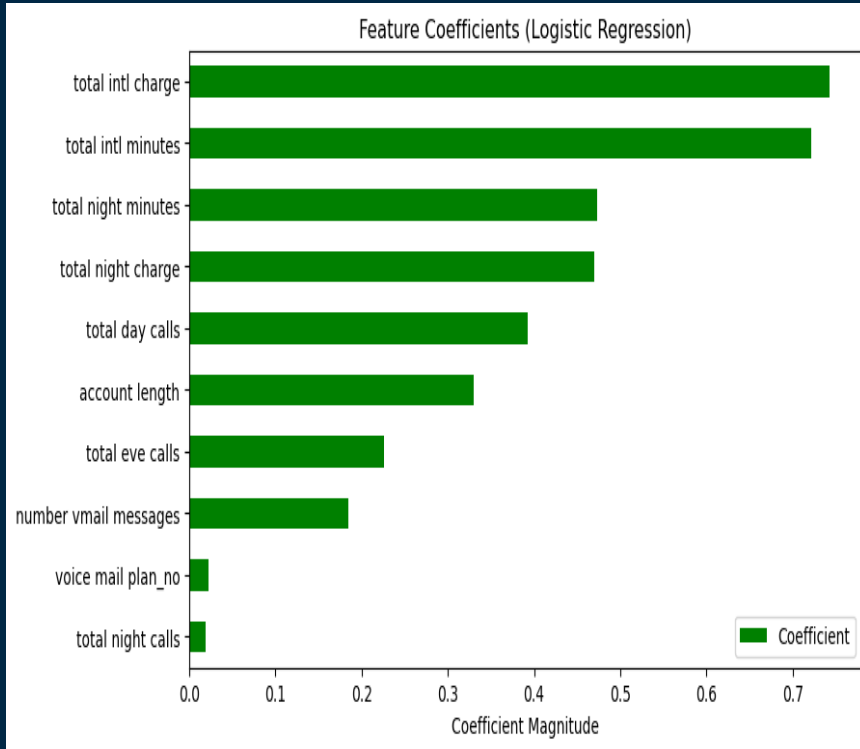
DATA ANALYSIS



- Our target variable is churn. 2850 represents customers that did not churn while 483 represents customers who churned.
- 14.5% of customers churned. We are dealing with an imbalanced classification problem.

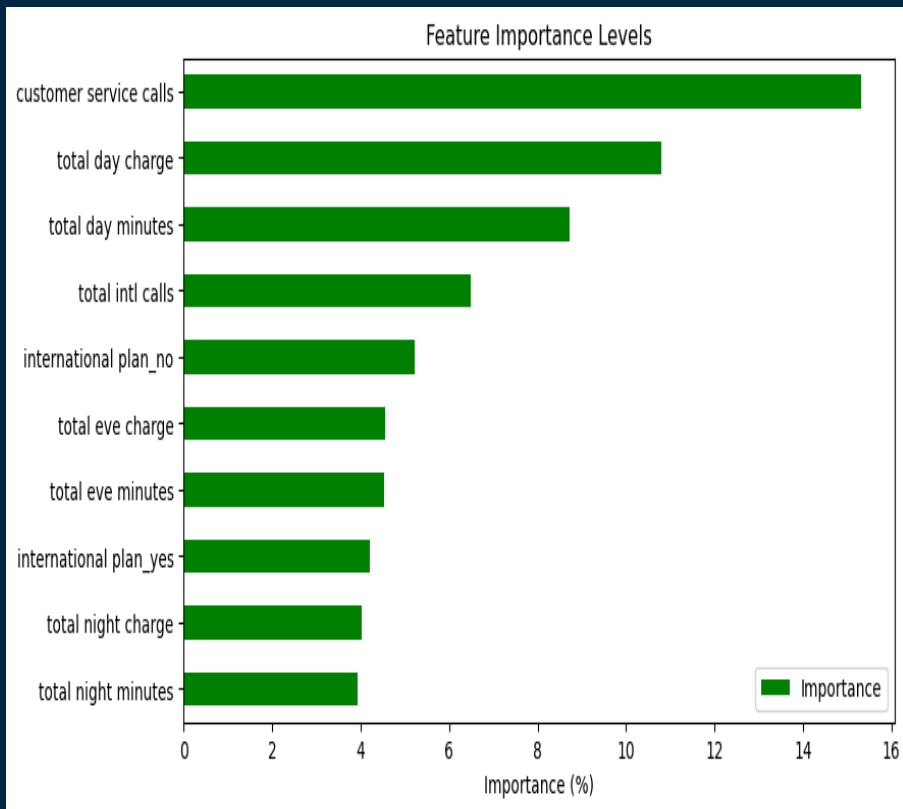
MODELLING

LOGISTIC REGRESSION MODEL



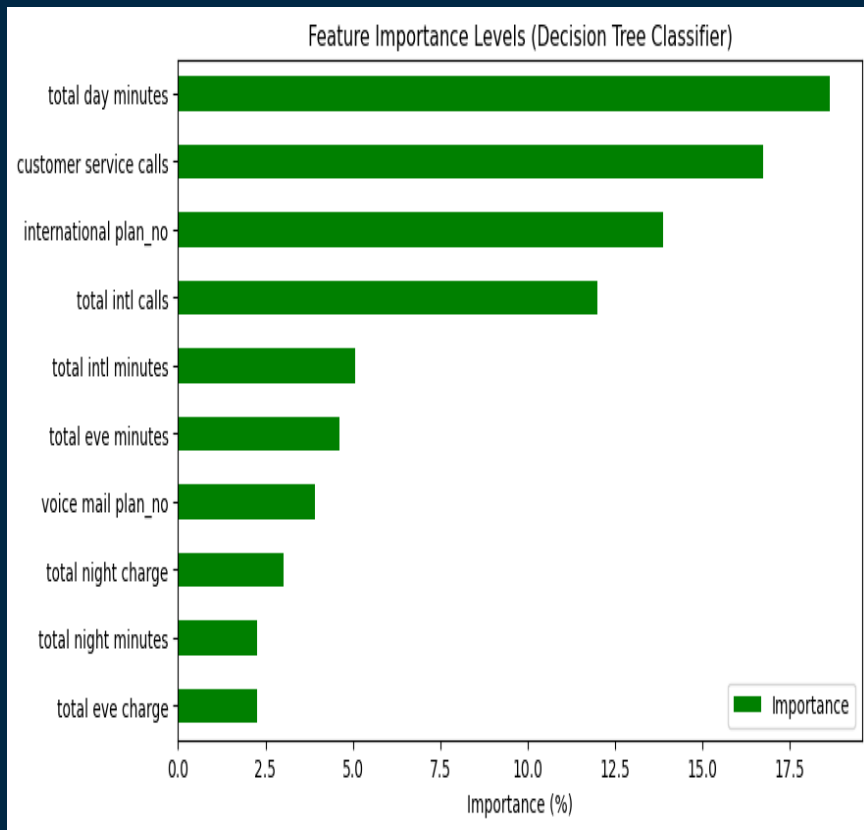
- The logistic regression model is our baseline model
- Accuracy: Achieved an accuracy of 84% on the test data.
- Precision: Precision for class 0 (non-churn) was 86%. Precision for class 1 (churn) was 73%.
- Recall: Recall for class 0 was 97%. However, recall for class 1 was 50%, suggesting some difficulty in identifying all churning customers.
- F1-score: The F1-score, which balances precision and recall, was 82%.
- Confusion Matrix: The confusion matrix shows 709 true negatives), 26 false positives, 99 false and 125 true positives
- Our most important features contributing to churn were: total international charge, total international minutes, total night minutes, and total night charge.

RANDOM FOREST CLASSIFIER



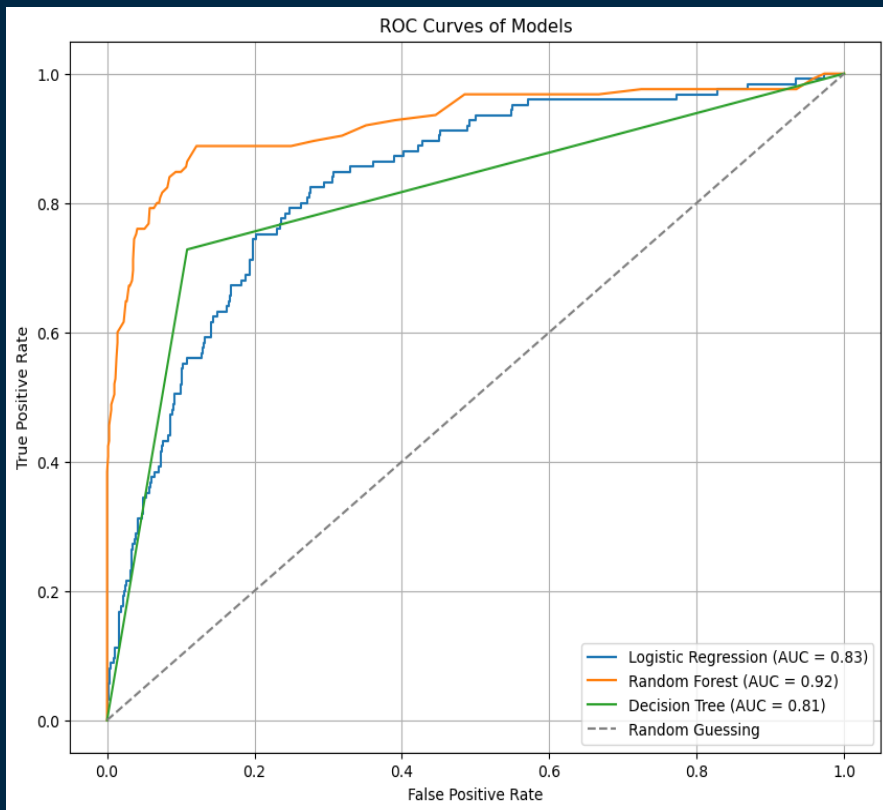
- Accuracy: Achieved an accuracy of 94% on the test data, indicating strong predictive power.
- Precision: Precision for class 0 (non-churn) was 96%. The precision for class 1 (churn) was 81%.
- Recall: Recall for class 0 was 97%. Recall for class 1 was 74%, F1-score: The F1-score was 87%, indicating overall excellent performance.
- Confusion Matrix: The confusion matrix indicates 709 true negatives, 20 false positives, 32 false negatives, and 93 true positives.
- The most important features in our random forest classifier are customer service calls, total day charge, total day minutes, and total international calls.

DECISION TREE CLASSIFIER



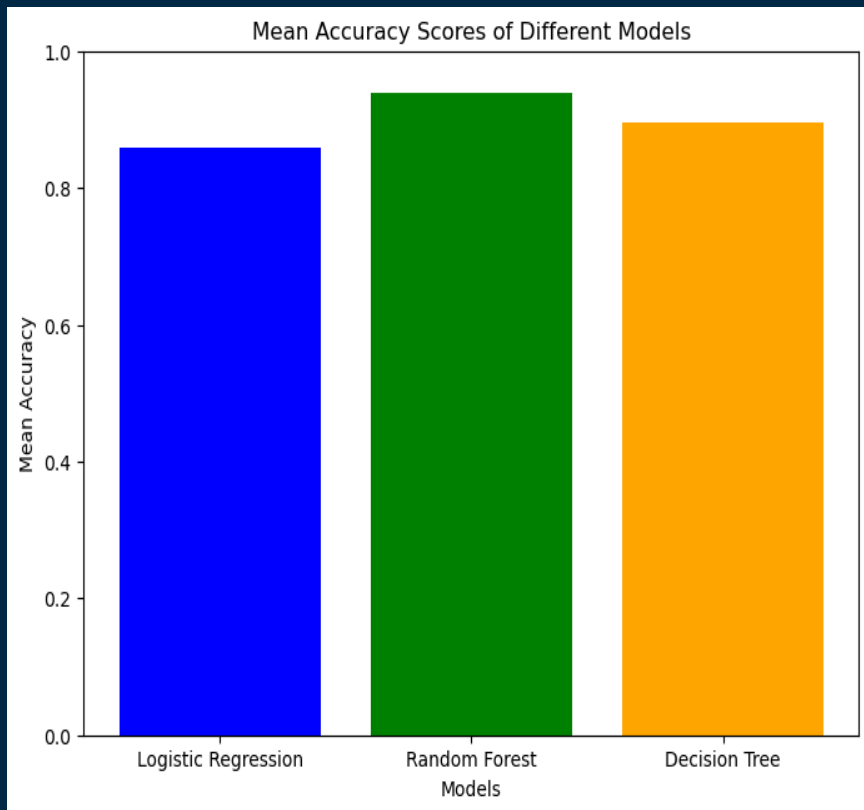
- Accuracy: Achieved an accuracy of approximately 89% on the test data.
- Precision: Precision for class 0 was around 96%. Precision for class 1 (churn) was approximately 81%.
- Recall: Recall for class 0 was approximately 86%,. Recall for class 1 was around 83%.
- F1-score: The F1-score, balancing precision and recall, was approximately 83%.
- Confusion Matrix: The confusion matrix showed 709 true negatives, 20 false positives, 21 false negatives, and 104 true positives.
- Our most important features contributing to churn were total day minutes, customer service calls, international plan_no, total international calls, and total international minutes.

ROC-AUC CURVES



- Logistic Regression (AUC = 0.83): The curve demonstrated decent performance, with an AUC score of 0.83, indicating a satisfactory ability to distinguish between positive and negative classes.
- Random Forest (AUC = 0.92): The ROC-AUC curve exhibited excellent performance, with an AUC score of 0.92, indicating a high ability to distinguish between positive and negative classes.
- Decision Tree (AUC = 0.81): The curve showed good performance, with an AUC score of 0.81, indicating a satisfactory ability to distinguish between positive and negative classes.
- Overall, the ROC-AUC curves confirmed the effectiveness of the models, with random forest outperforming logistic regression and decision tree models in terms of AUC score.

5-FOLD CROSS VALIDATION

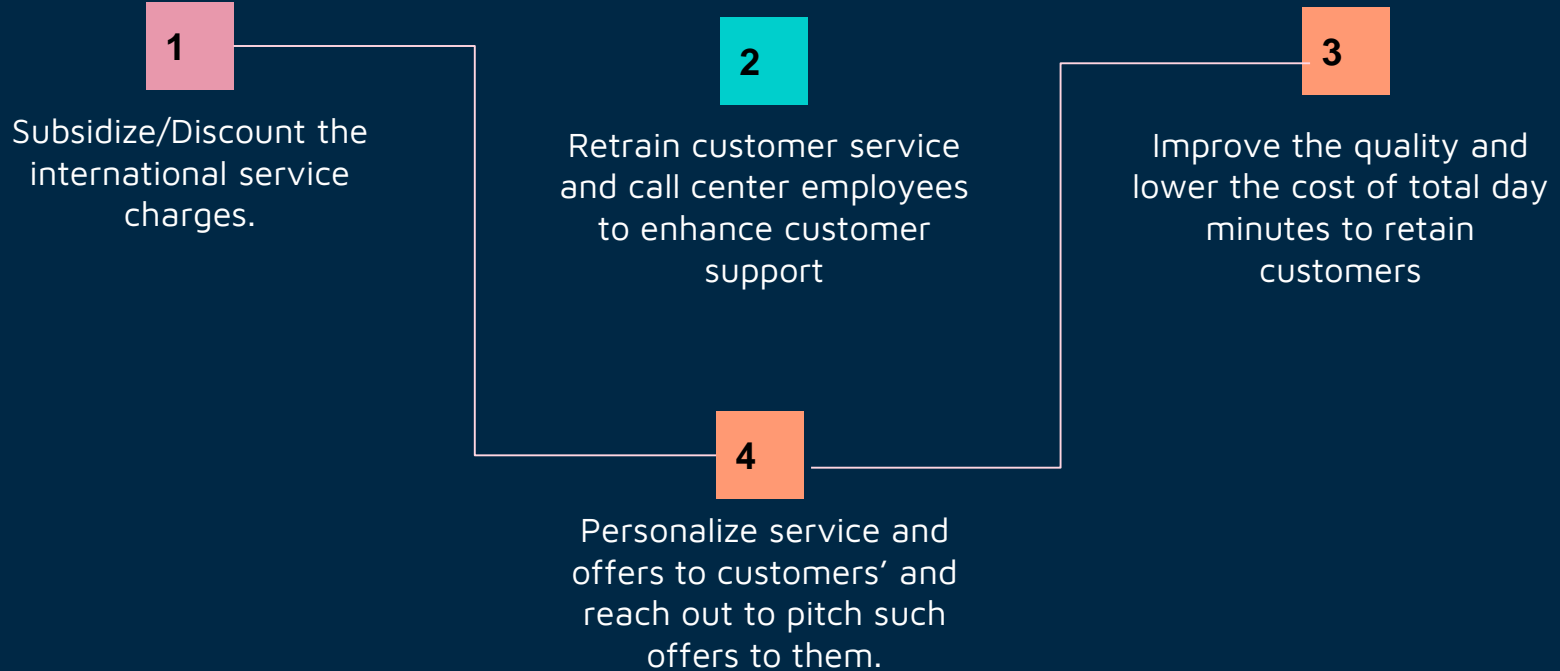


- **Logistic Regression:**
Mean Accuracy: 0.85897
CV Scores: [0.73046, 0.88681, 0.89136, 0.89369, 0.89252]
The model demonstrates consistent performance across folds, with an average accuracy of approximately 0.86.
- **Random Forest:**
Mean Accuracy: 0.93812
CV Scores: [0.93582, 0.93116, 0.93692, 0.94276, 0.94393]
The random forest model exhibits high accuracy across folds, with an average accuracy of approximately 0.94.
- **Decision Tree:**
Mean Accuracy: 0.89468
CV Scores: [0.86464, 0.88798, 0.89836, 0.92056, 0.90187]
The decision tree model shows consistent performance, with an average accuracy of approximately 0.89.
- Overall, the random forest model achieves the highest average accuracy, followed by logistic regression and decision tree models. This indicates that the random forest model is the most effective in predicting customer churn in this dataset.

CONCLUSION

- **Model Performance:** The Random Forest model outperformed both Logistic Regression and Decision Tree models in terms of accuracy, with an average accuracy score of approximately 94%. This indicates that Random Forest may be the most suitable model for predicting customer churn.
- **Feature Importance:** Important features such as total international charges, total international minutes, total night minutes, total night charge, and total day calls were identified across various models. These features can significantly influence the likelihood of customer churn and should be closely monitored by SyriaTel.
- **Business Recommendations:** Based on the identified influential features, SyriaTel should focus its efforts on addressing factors such as international call charges, nighttime usage patterns, and customer service calls to reduce churn rates. Implementing targeted retention strategies based on these insights can help enhance customer satisfaction and loyalty.
- **Further Analysis:** Continuous monitoring and analysis of customer churn patterns are essential for SyriaTel to stay proactive in retaining its customer base. Additionally, exploring more advanced modeling techniques and incorporating additional data sources could further improve churn prediction accuracy.

RECOMMENDATION



The background is a dark navy blue. It is decorated with various geometric elements: small squares in white, light blue, and orange, and thin vertical lines of the same colors. These elements are scattered across the frame, creating a modern, minimalist aesthetic.

THANK YOU!