

Predicting Customer Churn For SyriaTel Company



Author :Ann Mwangi



Business understanding

- Customers are important in any business or organization. The more the customers, the more the organization makes in profit. From customers we can get two groups of customers, New customers and existing customers. Existing customers are crucial to any business as getting a new customer is not easy. Studies have shown that acquiring new customers is 6 to 7 times more expensive than retaining an old one.



Data understanding

- In this project we will make use of "bigml_59c28831336c6604c800002a.csv" data from SyriaTel company which was obtained from Kaggle. The dataset includes how many minutes they spend talking, how many calls they make and how much they are charged during day, evening and night periods. To get the monthly customer churn rate = $(\text{Customers lost during the month} / \text{Total customers at the beginning of the month}) * 100$



Objectives

- To find out the **features** that are most important to our target variable
- To come up with a predictive model that predicts whether a customer will churn soon
- Come up with recommendations for customers predicted to churn.

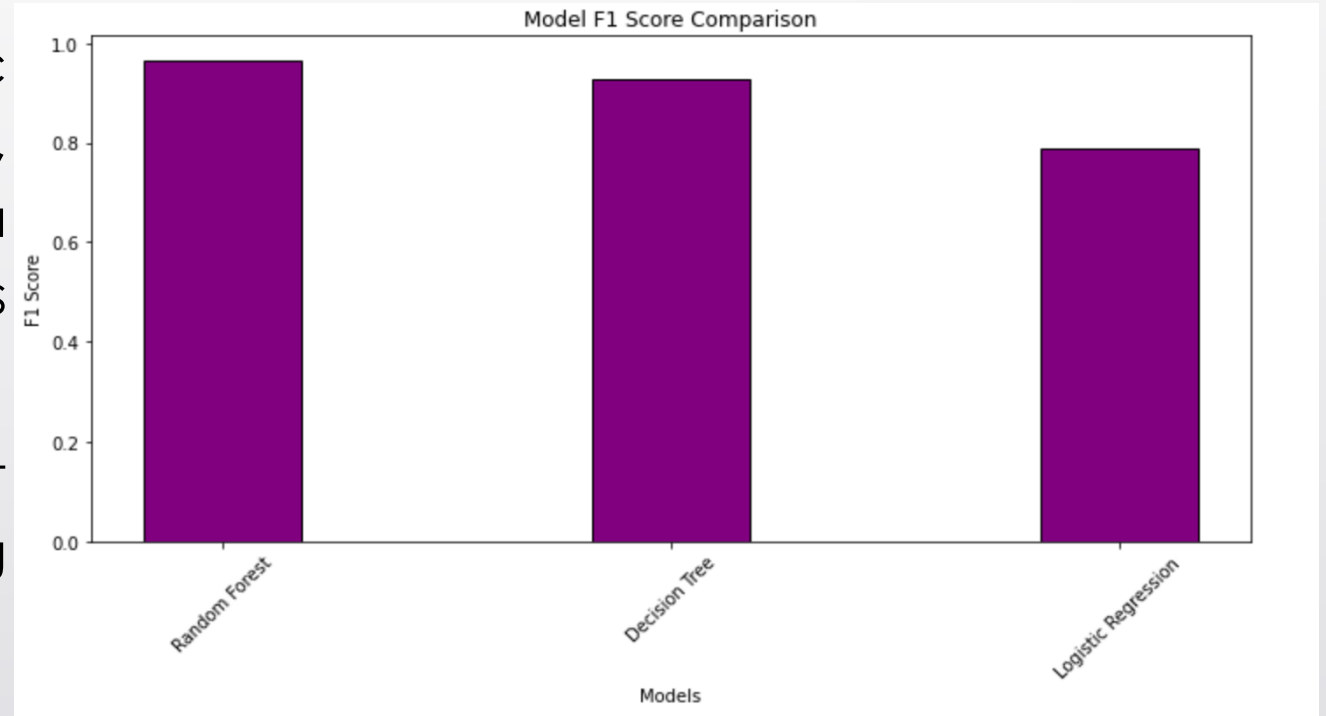


Models Implemented

- Random Forest
 - High performance and robustness
 - Good balance between accuracy and interpretability
 - Used for feature importance analysis
- Decision Tree
 - Simple and interpretable model
 - Sensitive to hyperparameters and dataset characteristics
 - Useful for understanding decision boundaries
-
- Logistic Regression
 - Linear model suitable for binary classification
 - Provides probability estimates
 - Less computationally intensive than ensemble methods

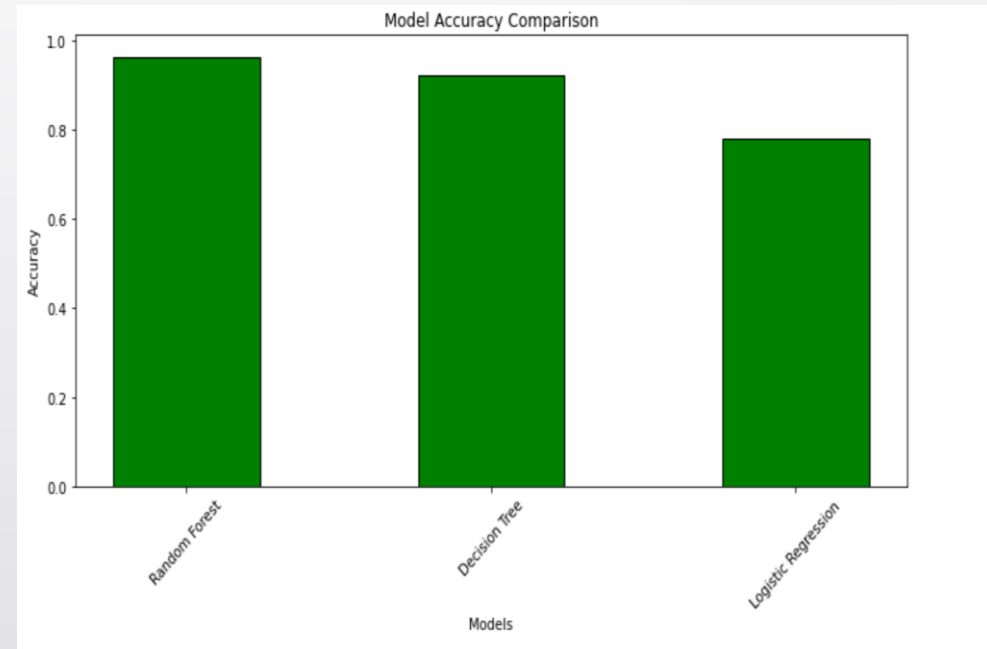
Performance comparison

- The F1 score is the harmonic mean of precision and recall, providing a single measure of a classifier's performance across both classes. It ranges from 0 to 1.
- we can see that random forest has the highest at 0.96 indicating a good performance



Model Accuracy comparison

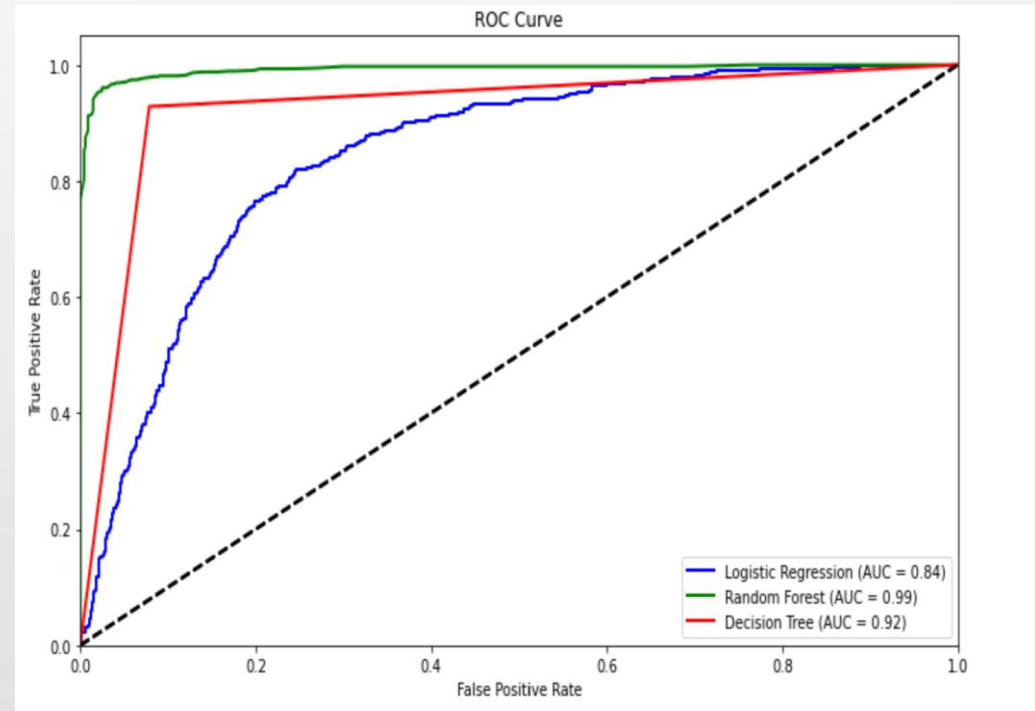
- Random forest was at 97.06% being the highest.
- Decision tree was at 92% and lastly logistic regression at 78 %



ROC curves comparison

Observations from the three Models:

- Logistic Regression showed a moderate performance with an AUC of 0.75.
- Random Forest demonstrated superior performance with an AUC of 0.85.
- Decision Tree had the lowest AUC among the three models at 0.72.
- These results suggest that Random Forest performs best in distinguishing between positive and negative cases, followed by Logistic Regression, and then Decision Tree.





Recommendations

- Customer care agents could be trained on giving a better customer experience to reduce the churn that increases with customer service calls
- The company can invest more in terms of network and retails outlets on states that have a high churn to mitigate that
- Reward or give discounts to customers with higher total costs to prevent churn



Conclusion

- Customers are important to any business and investing on ways to mitigate churn is crucial. Syrial Tel can employ the recommended steps to mitigate churn which will inturn increase its sales. Customer service is important and can make or break a business. The customers at Syrial Tel could be calling because they already have an issue, whether they churn or not depends mostly on how they were handled. Good customer service means a better resolution and satisfaction and that will prevent a customer from churning. Customers who spend more will feel more appreciated and recognized when they are rewarded.



Next steps

- Optimize model performance through hyperparameter tuning, focusing on Random Forest's key parameters.
- Analyze feature importance to refine the feature set and gain deeper insights into the problem domain.
- Implement cross-validation to assess model generalization and compare performance across different folds.
-

