

Churn Analysis and Predictive Modeling using Machine Learning



Project Overview



This Project analyses customer's unsubscription from SyriaTel's Telecom and also does predictive modeling by examining various factors e.g. state, account length, number of voice mail messages and many others.

Business Understanding



Customer churn is a major challenge for SyriaTel's Telecom company as acquiring new customers is significantly more expensive than retaining existing ones. The company has observed a steady increase in customer attrition, which directly impacts revenue, customer lifetime value and market competitiveness.



Objectives

1. To Determine whether state is a factor affecting churning.
2. To find out which state has the highest churn.
3. To determine whether new customers churn faster than long-term customers.
4. To determine which feature strongly drives churning.
5. To determine the best-performing machine learning model for predicting SyriaTel telecom customer churn.

Data Understanding



- The dataset was clean i.e. it has no missing values and no duplicates.
- The dataset contains 20 features i.e. international plan, voice mail plan, state, total evening, day and night minutes on a call etc. and one target which is churn.

Modelling



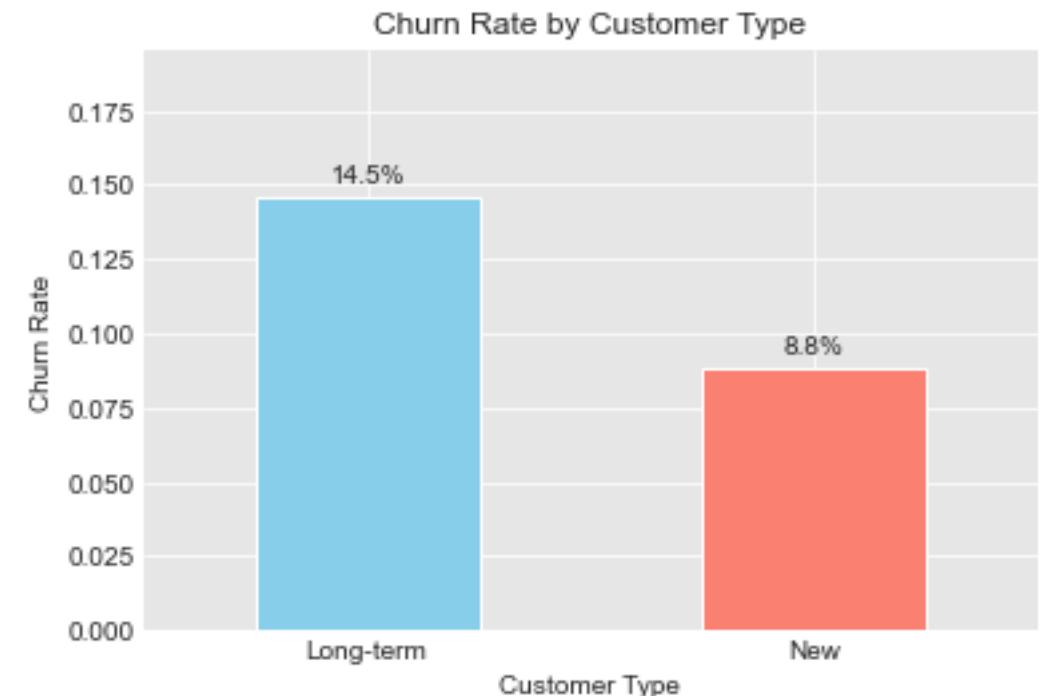
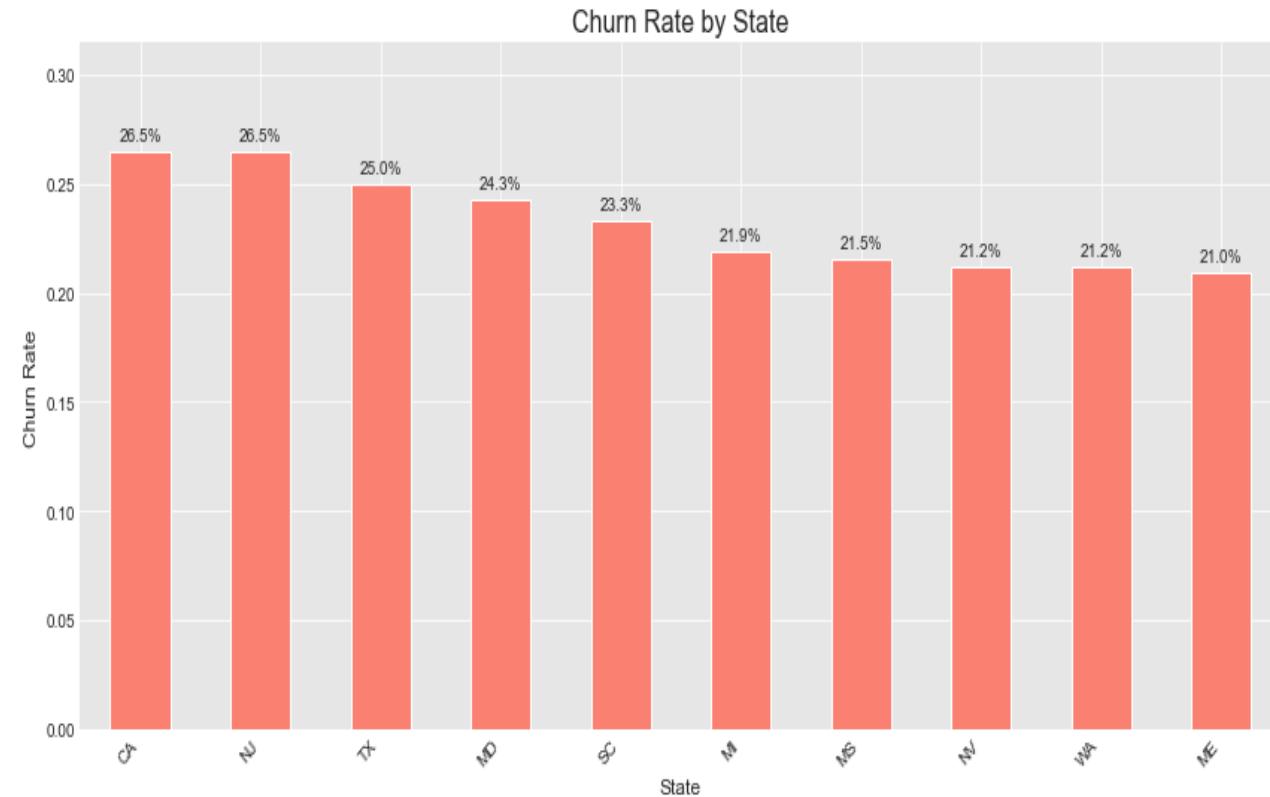
- I formulated the churn problem as a binary classification task. Several models were evaluated, including:
 1. Logistic Regression (baseline)
 2. Decision Tree Classifier
- The final model was selected based on validation AUC score and model interpretability



Evaluation

- Model performance was evaluated using **ROC curve and Area Under the Curve**
The final model achieved:
 - **Train ROC AUC:** 0.78
 - **Test ROC AUC:** 0.81

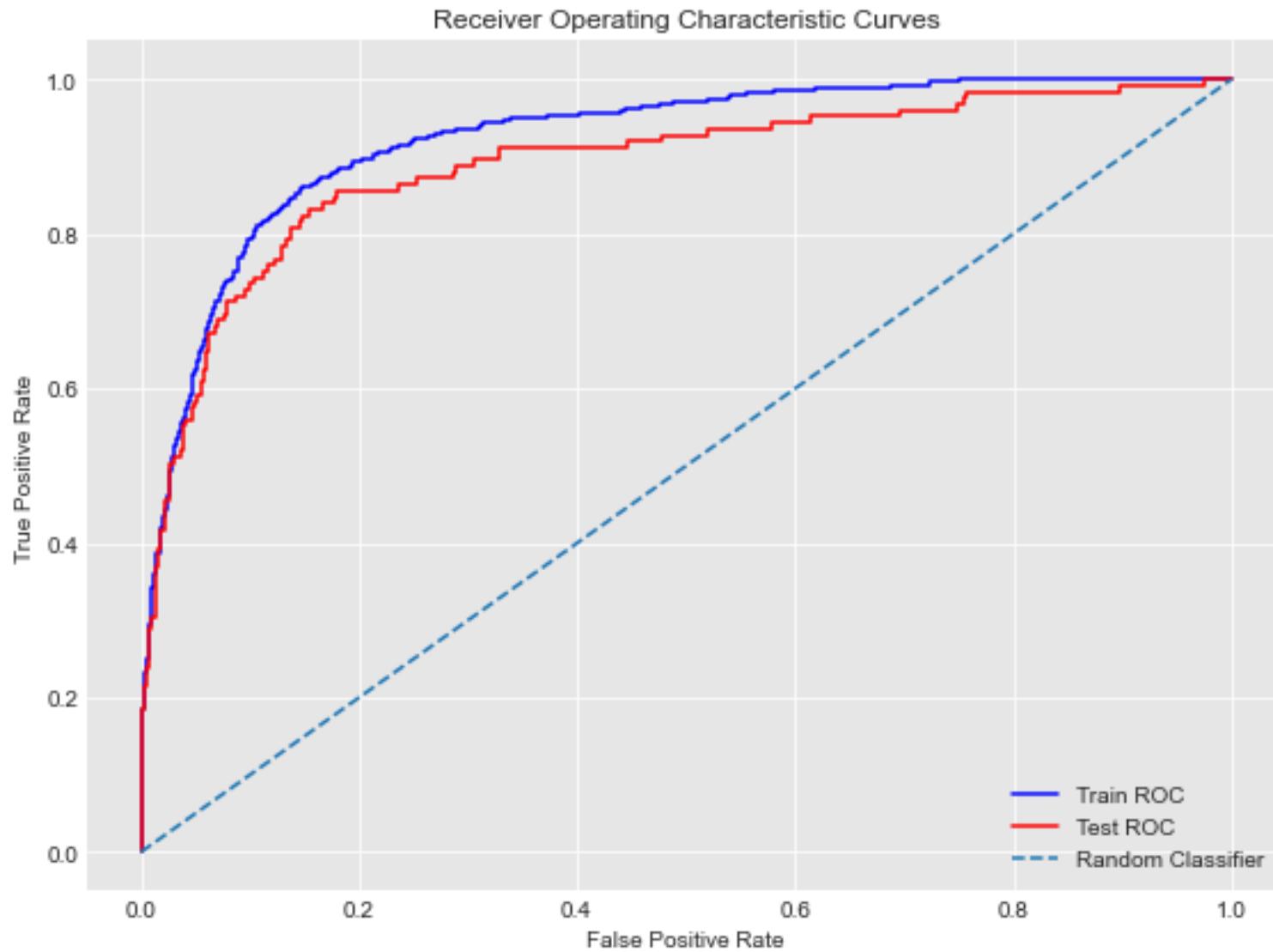
Visualizations



On average California and New Jersey lead in churn rate.
26 in 100 customers both in CA and NJ churn.

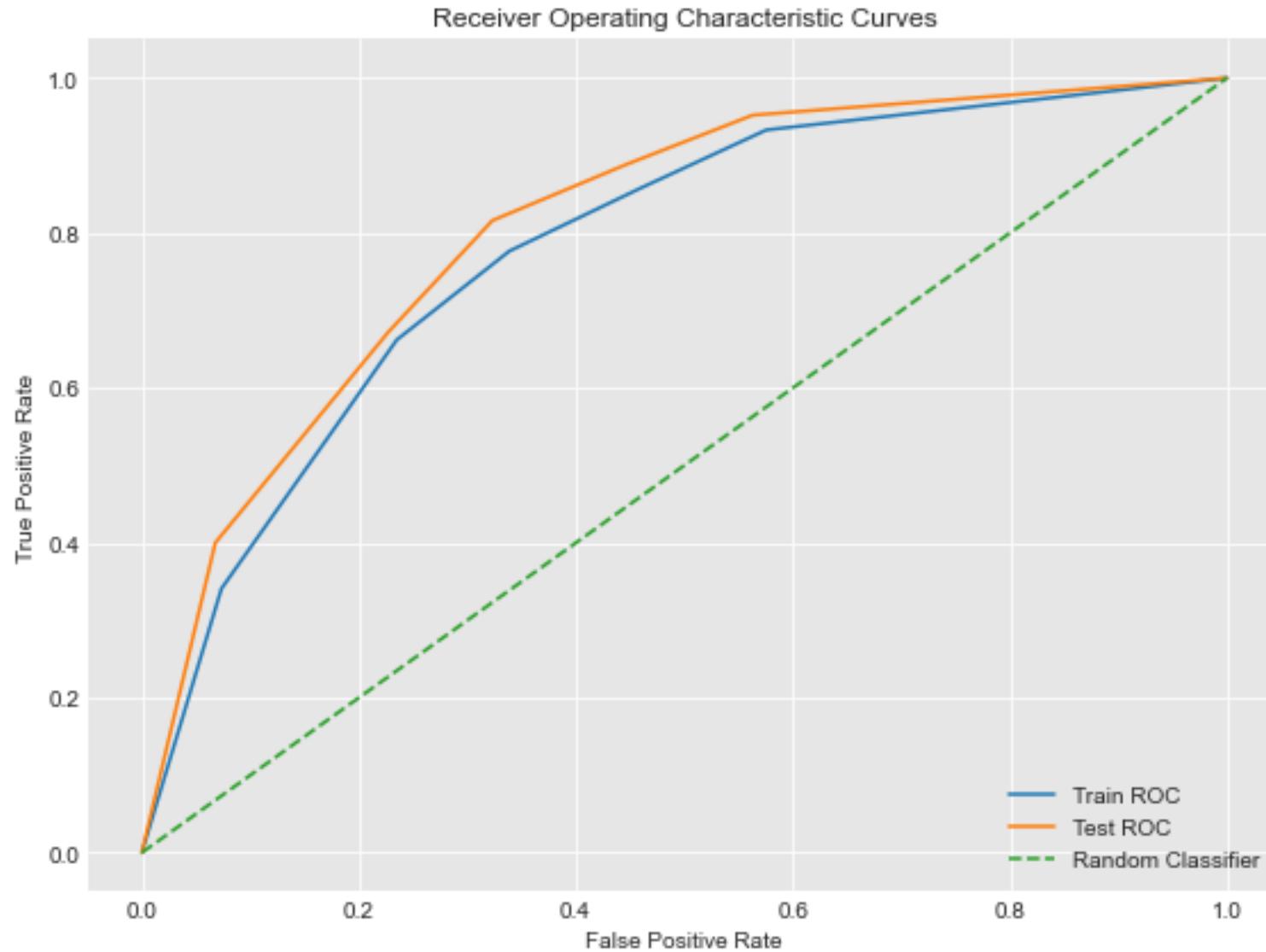
To the contrary, Long term customers seem to
churn faster than short term.

Evaluation



- Model performance was also evaluated using **ROC curve and Area Under the Curve**.
- The final Logistic Regression model achieved:
 - Train ROC AUC: 0.92**
 - Test ROC AUC: 0.89**

Evaluation Cont.



- Model performance was evaluated using **ROC curve and Area Under the Curve**.
- The final Decision Tree achieved:
Train ROC AUC: 0.78
Test ROC AUC: 0.81

Conclusions



The analysis shows that customer churn can be predicted with reasonable accuracy using state, account length and service features like total eve minutes on a call etc.

The model enables the business to:

- Identify the strongest driver of churning, international plan and voice mail plan.
- Target retention efforts more effectively.



Recommendations

1. Re-evaluate the international plan offering.
2. Target international plan users with retention incentives.
3. Promote and expand voice mail plan adoption.
4. Prioritize retention efforts in high-churn state.

Next Steps

- 
1. Future work will explore ensemble methods such as Random Forests.
 2. Deploy & Automation in systems.
 3. Techniques such as SHAP applied to better understand both global and individual-level drivers of churn.



Thank You

Name: Victor Kipkemboi Chepkonga.

LinkedIn:

<https://www.linkedin.com/in/victor-kipkemboi-34b302387/>

GitHub Repo: <https://github.com/Victor-Mufasa/Phase-3-Project/tree/main>