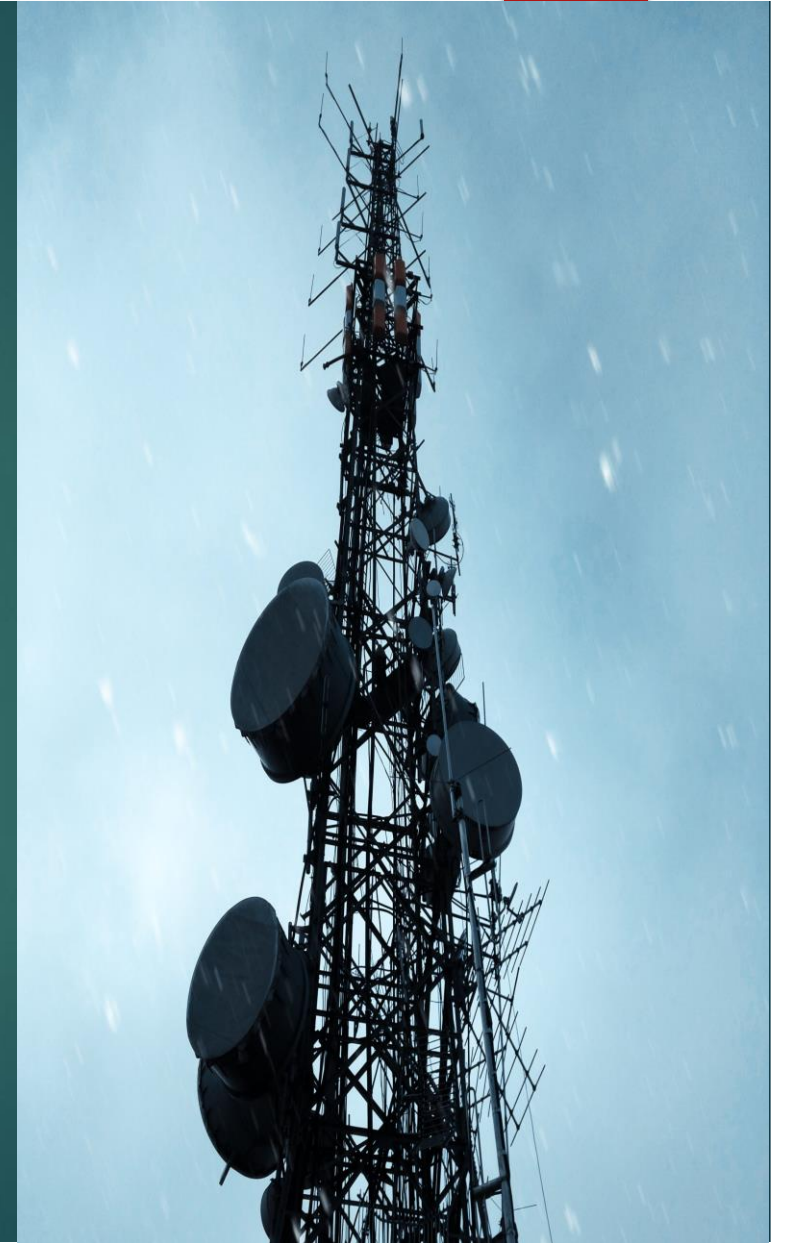# SyriaTel Customer Churn Analysis

Overview

▶ In business, "churn" refers to the rate at which customers stop doing business with a company over a specific period. It's often expressed as a percentage of the total customer base and indicates how well a business is retaining its customers.

▶ High churn rates can signal potential problems with customer satisfaction, product quality, or customer service.

Business Understanding

▶ The churn rate is the percentage of customers who stop using a service or product within a specific timeframe.

▶ This model aims to answer a few objectives such as:

1. What is the current churning rate acoording to the available dataset?

2. Can we be able to see the factors that affect the churning rates most?

3. Can we be able to predict the churning rate of the company in the time to come?

4. What is the accuracy of the prediction model that is created?

Data

▶ This project utilizes data from the Churn in Telecom dataset from Kaggle.

▶ The target variable in this dataset that we aimed to predict was the churn column.

▶ The features of this dataset include locational information as well as plan details such as call minutes, charges, customer services calls and whether the customer had an international plan and/or voice mail plan.
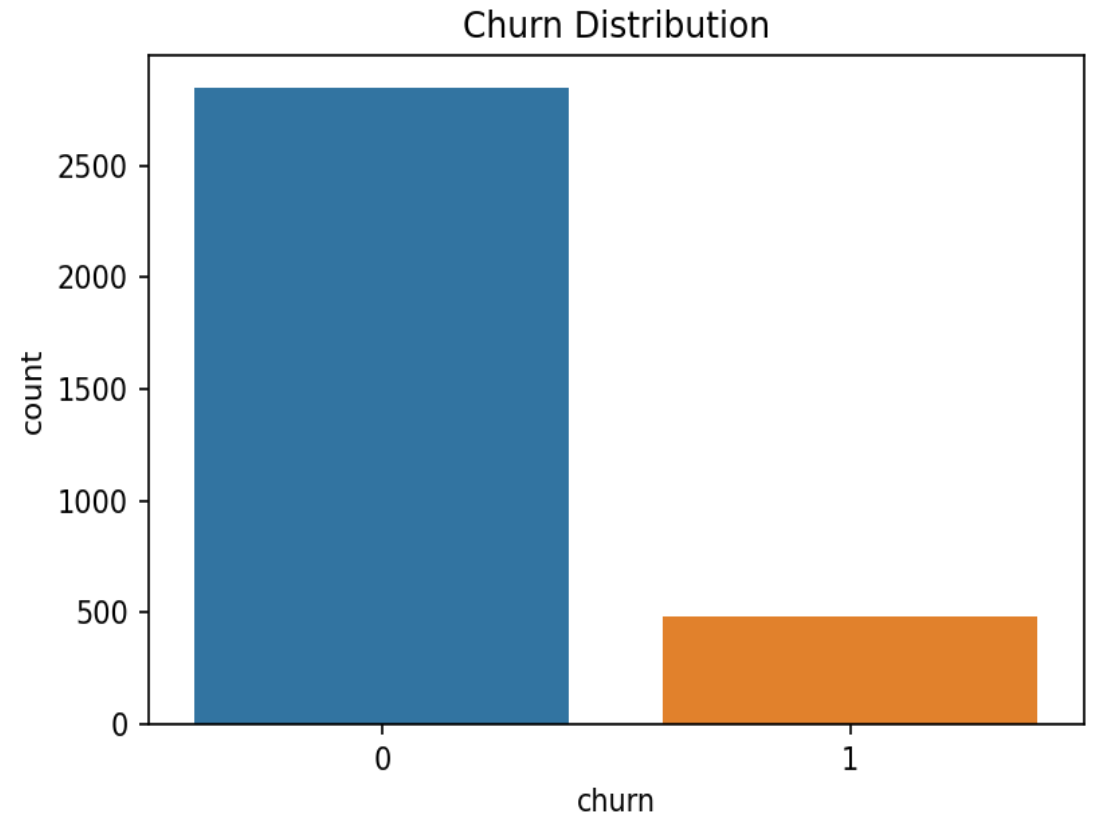
Methods

This project utilizes:

- Logistic regression
- Decision trees
- Identifying class imbalance
- SMOTE method
- Feature importance and optimization
- Hyperparameter tuning

Results

▶ The class imbalance shown alongside can be used to determine the total churning rate which is about 14%.

Churn Distribution

Results

From this graph we get to see that the factors that affect churning most are:

- ▶ Total day minutes
- ▶ Customer service calls
- ▶ International plan
- ▶ Total international calls
- ▶ Total day charge

Feature Importance - Decision Tree
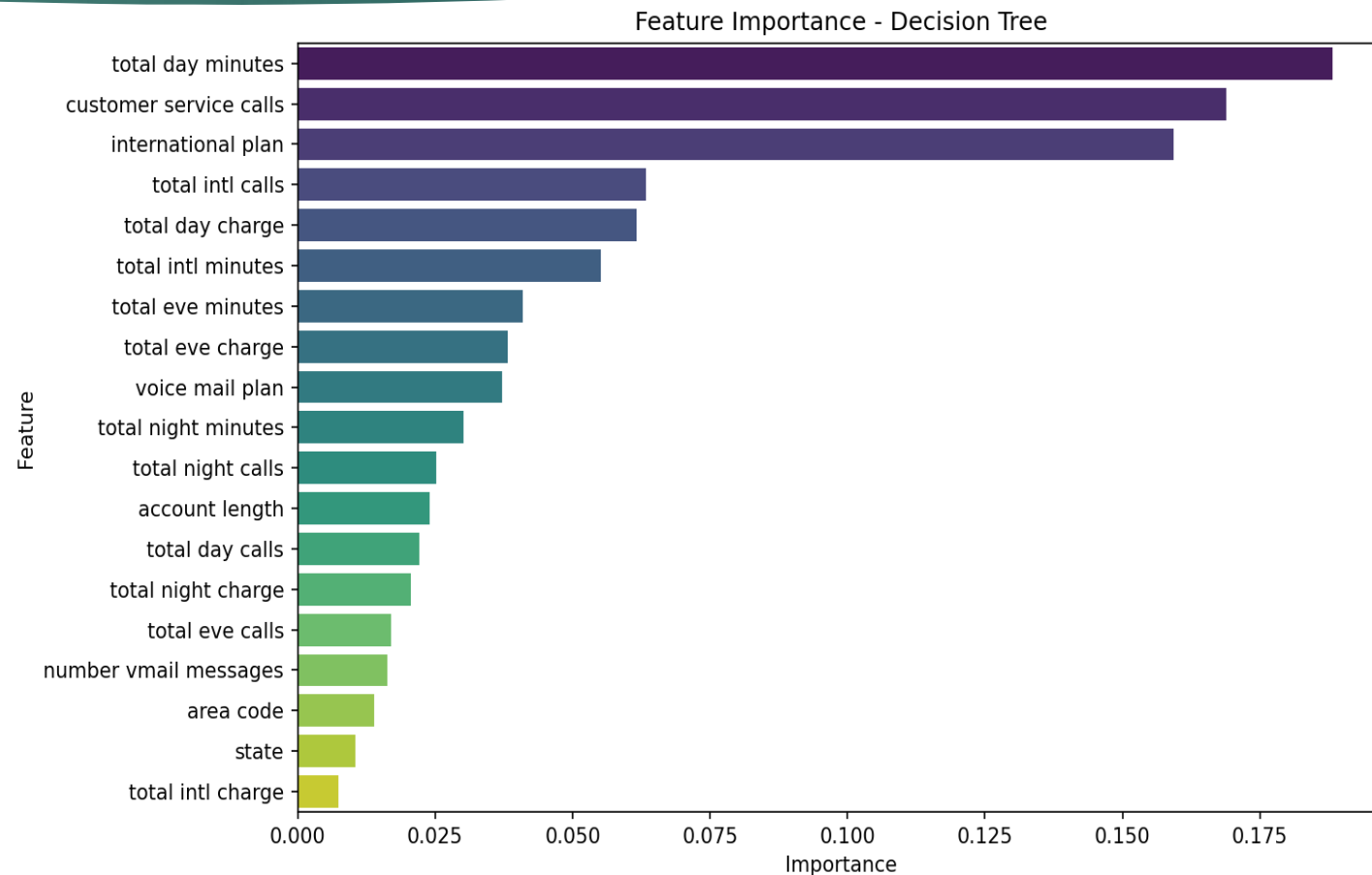
Results

▶ From the following tables we can conclude that the final prediction model had an accuracy of more than 93% and a ROC AUC of more than 0.88.

▶ This means the model is highly reliable as a predictor.

```
==== Decision Tree (Top 16 Features) Evaluation ====
>>> Train Metrics:
Accuracy: 0.9557
Precision: 0.9521
Recall: 0.7277
F1 Score: 0.8249
ROC AUC: 0.9182

>>> Test Metrics:
Accuracy: 0.9385
Precision: 0.8947
Recall: 0.6733
F1 Score: 0.7684
ROC AUC: 0.8888

Classification Report (Test):
              precision    recall  f1-score   support

           0       0.94      0.99      0.96       566
           1       0.89      0.67      0.77       101

    accuracy                           0.94       667
   macro avg       0.92      0.83      0.87       667
weighted avg       0.94      0.94      0.93       667
```

Conclusions

▶ High churn is linked to service dissatisfaction, particularly among international users and those who frequently contact customer service.

▶ Proactive engagement with high-usage customers and those with frequent support calls may help reduce churn.

▶ Predictive models can flag potential churners with >93% accuracy using only 16 key features.

Recommendations

- Operational Strategy

- Customer Support

- Marketing & Offers

- Model Deployment

Next Steps

▶ Apply ensemble models (Random Forest, XGBoost) for potentially better performance.

▶ Try SHAP or LIME for explainable AI insights.

▶ Consider customer lifetime value alongside churn prediction for better ROI targeting.