



SYRIATEL TELECOMMUNICATIONS

CUSTOMER CHURN PREDICTION

BY KEVIN OGUDA

BUSINESS PROBLEM

Problem statement:

SyriaTel lacks a predictive mechanism to identify which customers are likely to leave the service. As a result, retention efforts are reactive rather than proactive, leading to higher operational costs and loss of revenue. The problem at hand is to:

Build a classification model that can predict whether a customer is likely to churn, based on their behavioral and account-related data.

Key Features:

- Customer Service Interactions
- Account Information
- Usage Metrics
- Billing and Payment Methods
- Demographics



Business Challenges & Solutions

Challenges

- Data quality and imbalance
- Feature relevance
- Interpretability
- Cost sensitivity

Solutions

- Exploratory data analysis
- Predictive modelling
- Feature engineering
- Model evaluation
- Actionable insights

Objectives

- To determine distribution of the churn.
- To establish correlation matrix for numerical features.
- To determine relationship of churn in relation to important features.
- To predict whether a customer will "soon" stop doing business with SyriaTel telecommunications company.
- To find a way to prevent customer churn.
- To find a way of reducing revenue loss incase of customer churn.

Data Collection & Preparation

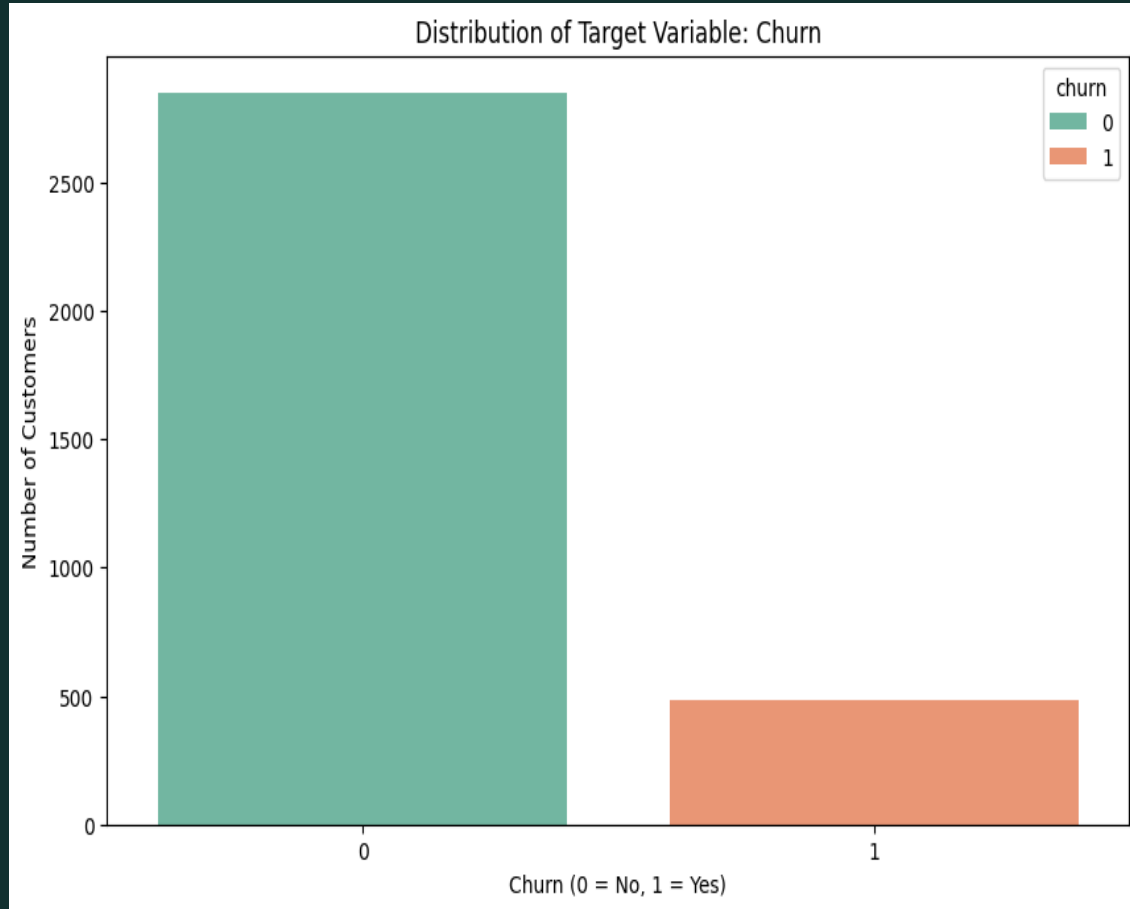
Data source

The dataset is sourced from: <https://www.kaggle.com/datasets/becksddf/churn-in-telecoms-dataset/data>

Data Preparation

- Data Cleaning
- Data Transformation
- Data Structuring

Distribution of the churn variable



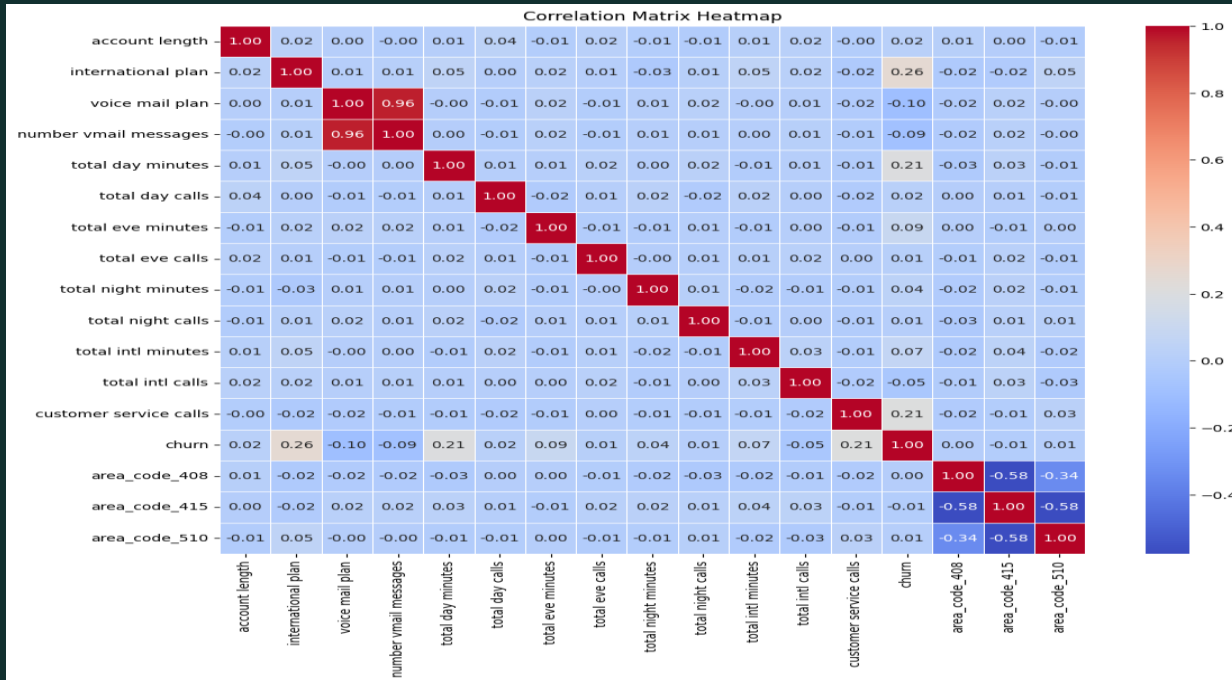
Interpretation:

The plot typically shows a much higher bar for non-churned customers (0) compared to churned customers (1), indicating that most customers stayed with the company while a smaller proportion left.

Insight:

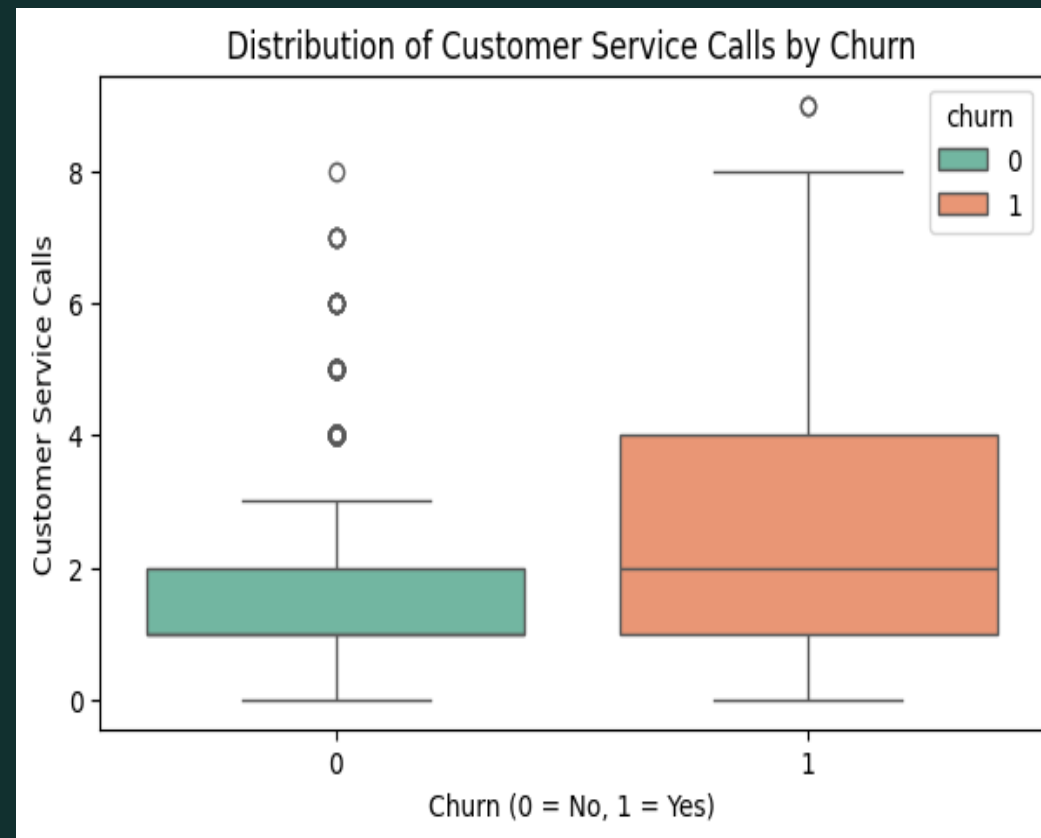
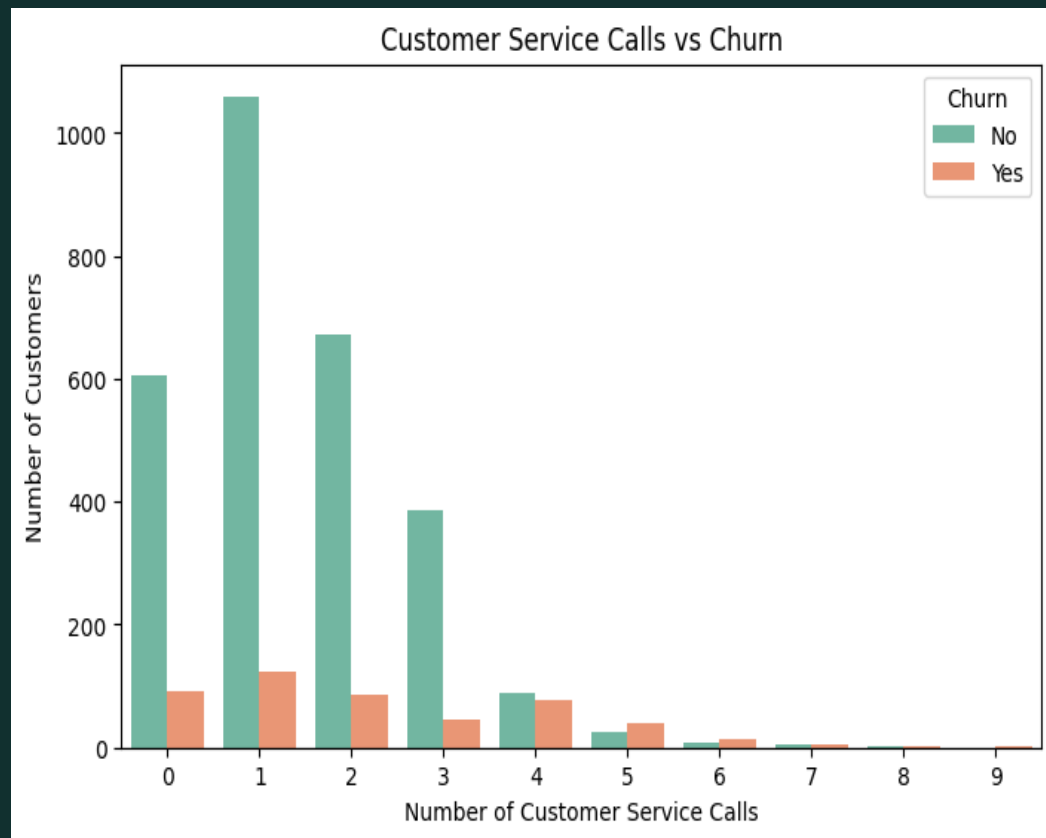
This highlights class imbalance in the dataset, which is common in churn problems. It means special care is needed during modeling to ensure the minority class (churners) is not overlooked.

Correlation matrix for numerical features

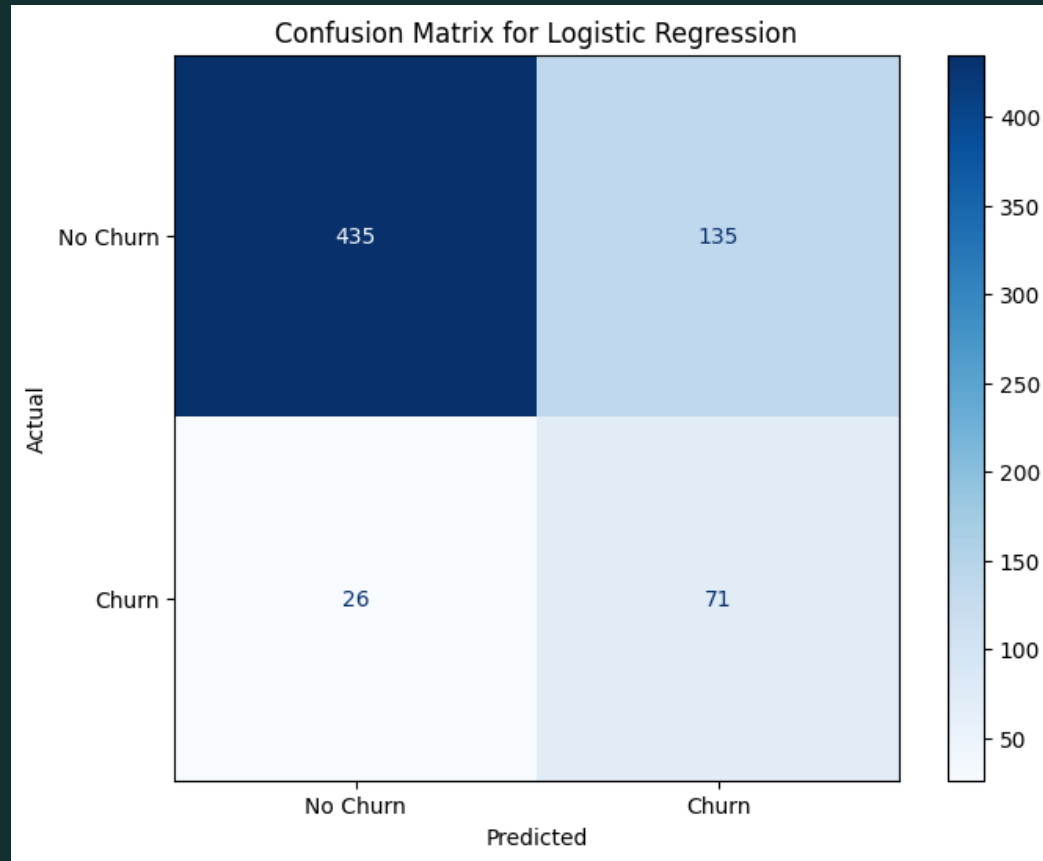


1. **High positive values (close to 1)** indicate a strong direct relationship between two features (as one increases, so does the other).
2. **High negative values (close to -1)** indicate a strong inverse relationship (as one increases, the other decreases).
3. **Values near 0** mean little or no linear relationship

Relationship between customer service calls and churn

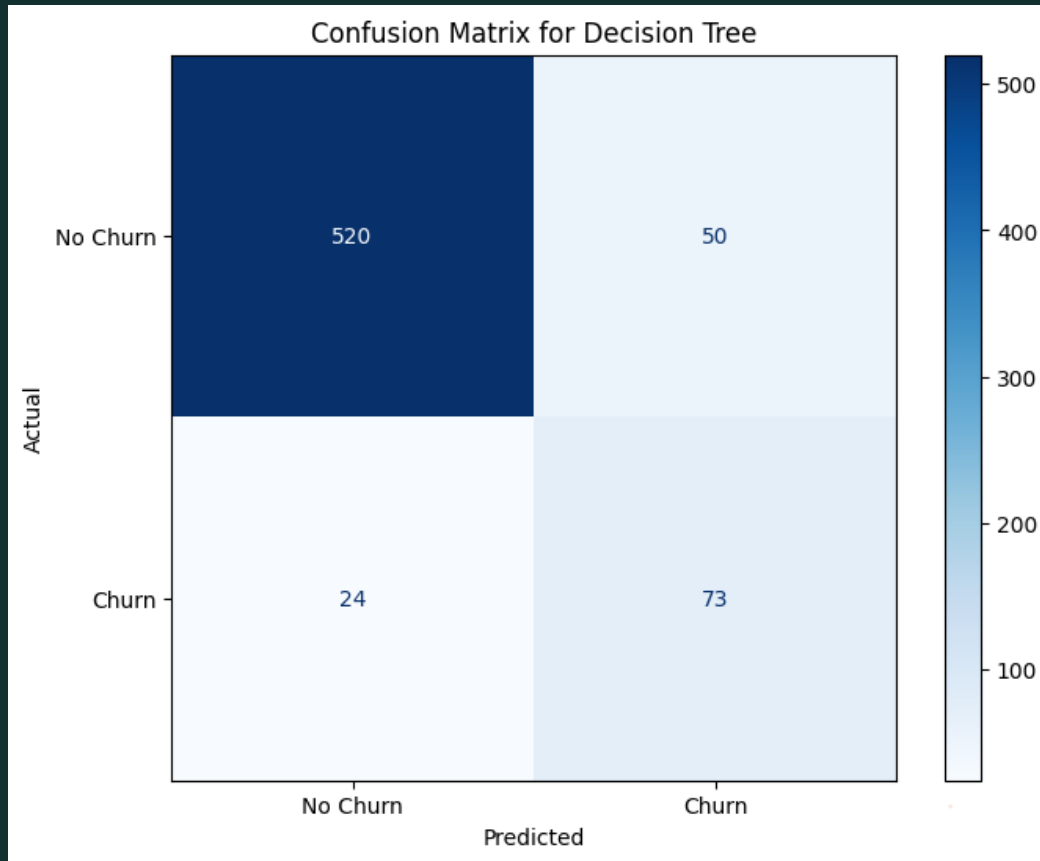


Logistic Regression Performance



1. **Negative (TN)= 435:** The number of customers who did not churn and were correctly predicted as non-churners.
2. **False Positive (FP)= 135:** The number of customers who did not churn but were incorrectly predicted as churners (Type I error).
3. **False Negative (FN)=26:** The number of customers who churned but were incorrectly predicted as non-churners (Type II error).
4. **True Positive (TP)=71:** The number of customers who churned and were correctly predicted as churners.

Decision Tree Performance



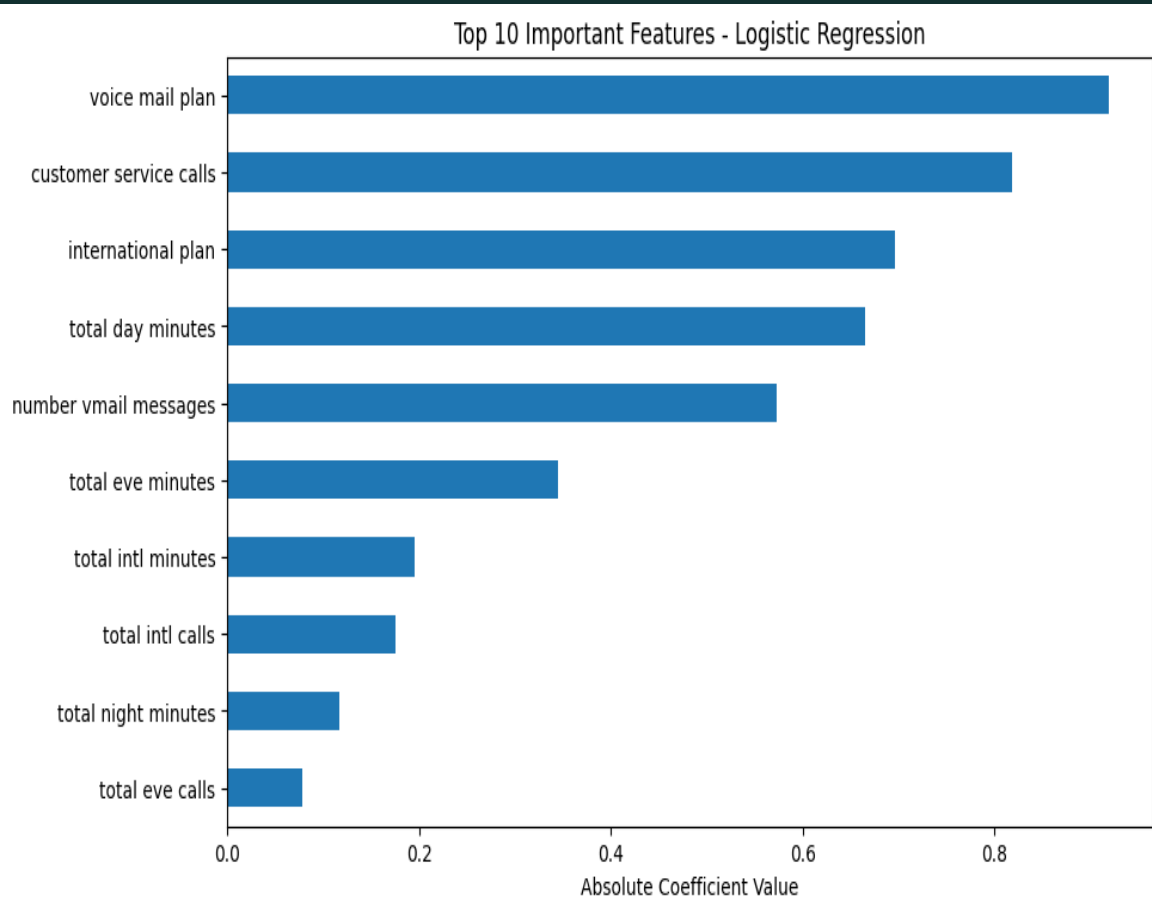
True Positives (TP):73 — Correctly predicted churners.

True Negatives (TN): 520 — Correctly predicted non-churners.

False Positives (FP):50 — Non-churners incorrectly flagged as churners.

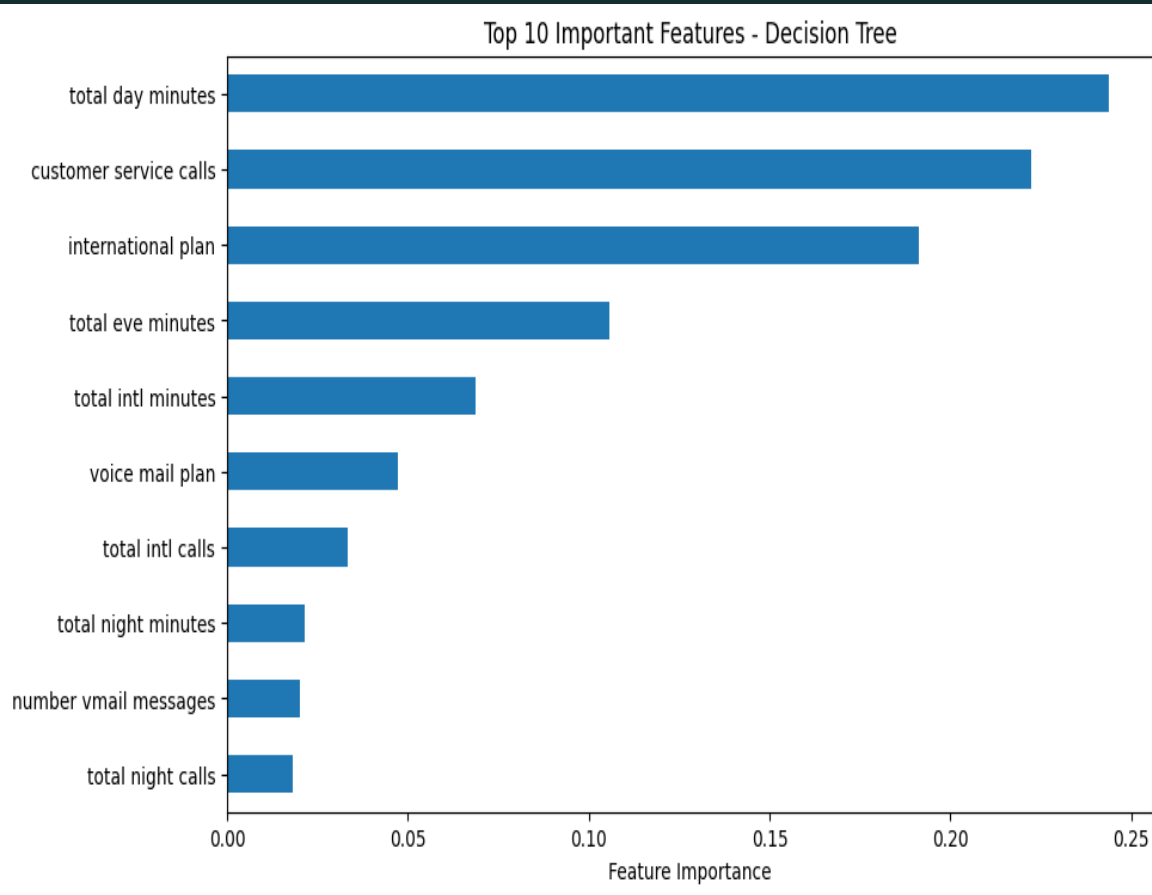
False Negatives (FN):24 — Churners missed by the model.

Important Features for Logistic Regression



- Voicemail plan doesn't work well with clients.
- More Customer Service Calls indicates dissatisfaction as it record high churn.
- International plan has high churn risk maybe due to cost.
- Total night minutes is relatively better than total day minutes as it record low churn.

Important Features for Decision Tree



- **Total day minutes records high churn maybe due to customer dissatisfaction whereas total night calls serves better as it records low churn.**
- **Customer Service calls are not satisfying as it records high churn.**
- **International plan also records high churn risk possibly because of cost**

Recommendations and next step

1. **Deploy the Decision Tree Model**-This model offers higher precision and accuracy, reducing wasted retention efforts.
2. **Target At-Risk Customers**-Focus retention campaigns such as special offers, discounts or improved service on the customers flagged as likely to churn.
3. **Monitor and Adjust**-Regularly monitor the model's performance and update it with new data to maintain accuracy.
4. **Address Key Churn Drivers**-Pay special attention to customers with frequent customer service calls, as this is a strong churn indicator.
5. **Optimize Resource Allocation**-By reducing false positives, the Decision Tree model helps SyriaTel allocate retention resources more efficiently, focusing on customers who are truly at risk.
6. **Continuous Improvement**-Investigate the reasons behind customer churn and use insights from the model to inform business decisions.



Thank You

KEVIN OGUDA

KEVINOGUDA@STUDENT.MORINGASCHOOL.COM