

Customer Churn Prediction

SyriaTel

Business Impact: Reducing churn improves revenue and customer retention.

Business Understanding

- 1. Customer Churn Definition:** Churn occurs when a customer stops using a company's services.
- 2. Industry Impact:** In the telecom sector, retaining customers is more cost-effective than acquiring new ones.
- 3. Why It Matters:** High churn rates lead to significant revenue losses and increased marketing expenses.
- 4. Goal:** Understanding churn drivers to improve retention strategies and enhance customer satisfaction.

Problem Statement

Problem: SyriaTel is experiencing high customer attrition, leading to revenue decline and reduced market share.

Business Impact: Without intervention, churn could continue increasing, making customer acquisition costs unsustainable.

Objectives

- 1. Predict Churn:** Develop a machine learning model to classify customers into churn and non-churn groups.
- 2. Identify Key Drivers:** Determine the most influential factors leading to churn.
- 3. Provide Insights:** Offer data-driven recommendations for customer retention strategies.
- 4. Improve Customer Retention Efforts:** Enable SyriaTel to take proactive actions in reducing churn rates.

Exploratory Data Analysis (EDA)

Key Observations from Univariate & Bivariate Analysis

- **Customer Service Calls & Churn:** Customers with higher service calls tend to churn more.
- **International Plan Influence:** Customers on international plans have a higher churn rate.
- **Account Length:** No strong correlation between account tenure and churn.
- **Total Day Minutes & Charges:** Usage levels do not significantly impact churn likelihood.
- **International Calls & Night Minutes:** No clear relationship with churn was found.

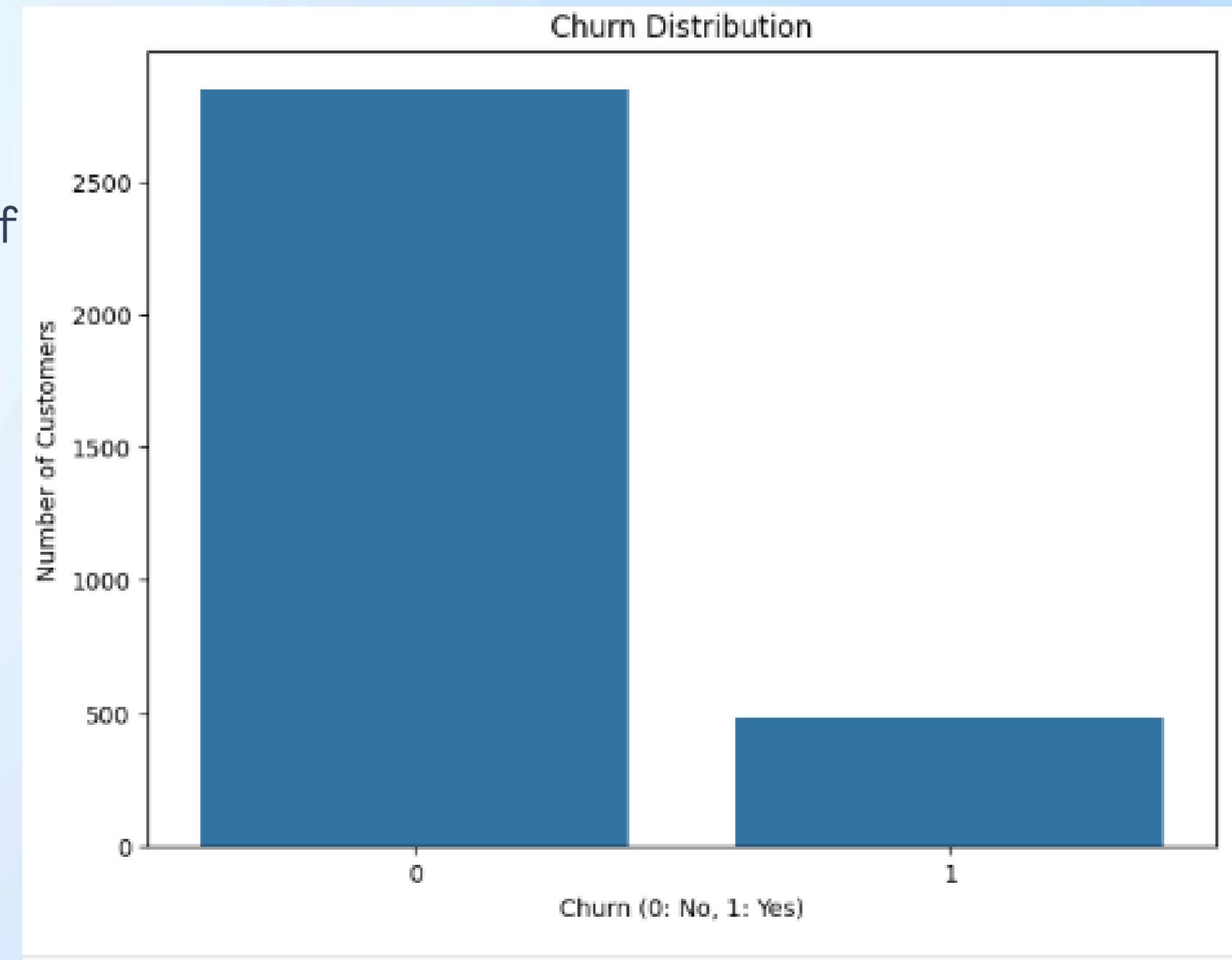
Exploratory Data Analysis (EDA)

Multivariate Analysis:

- Combining multiple features (e.g., total day minutes and customer service calls) helped reveal more patterns.
- Clustering techniques showed that churned customers tend to group in certain regions of feature distributions.

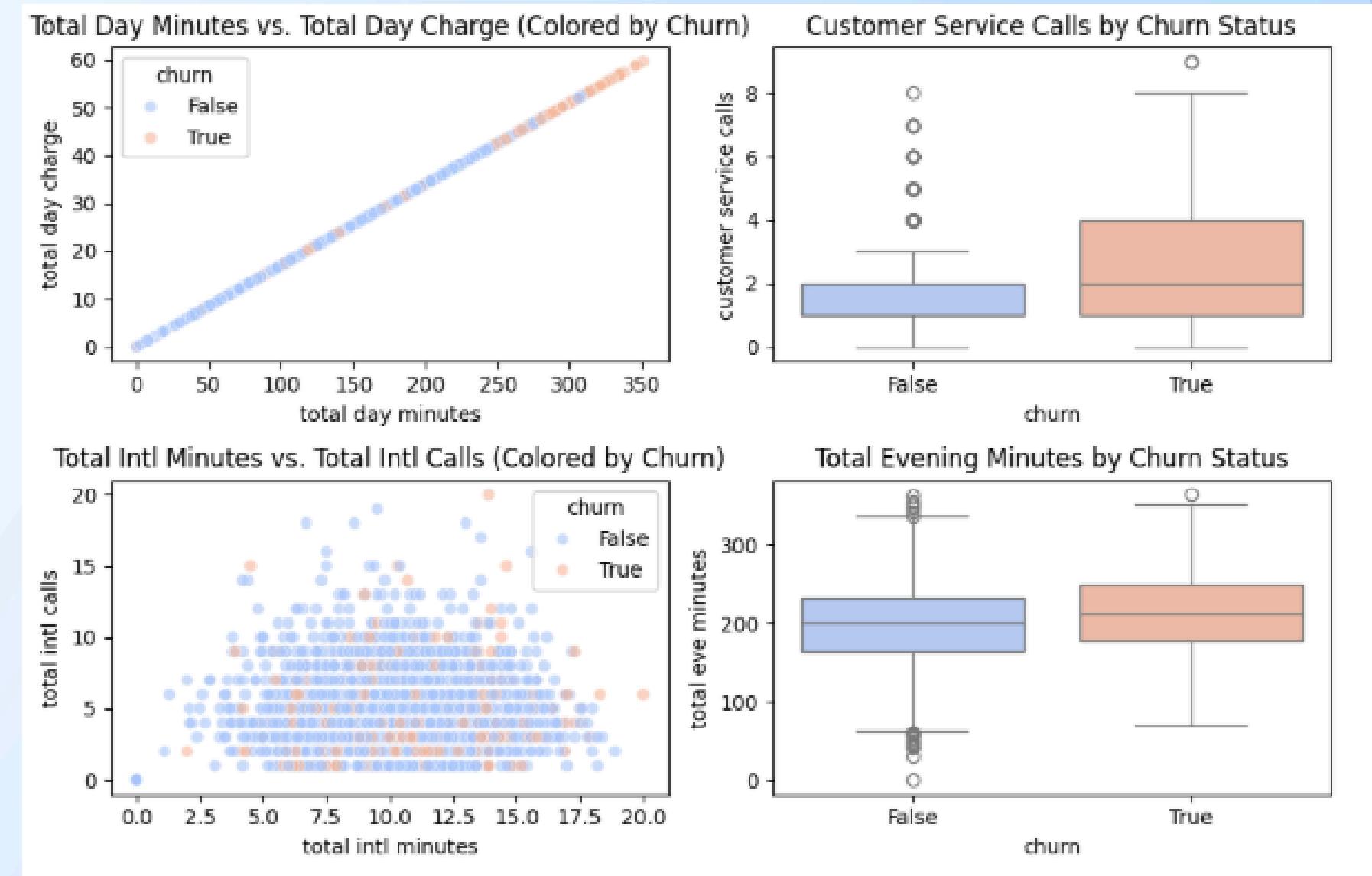
Churn Distribution

- The dataset is imbalanced, with a much larger proportion of customers who did not churn (0) compared to those who did (1).
- This imbalance suggests that resampling techniques (such as SMOTE) or weighted modeling approaches may be necessary to improve prediction performance.
- The low churn percentage emphasizes the need for precise targeting of at-risk customers to maximize retention efforts.



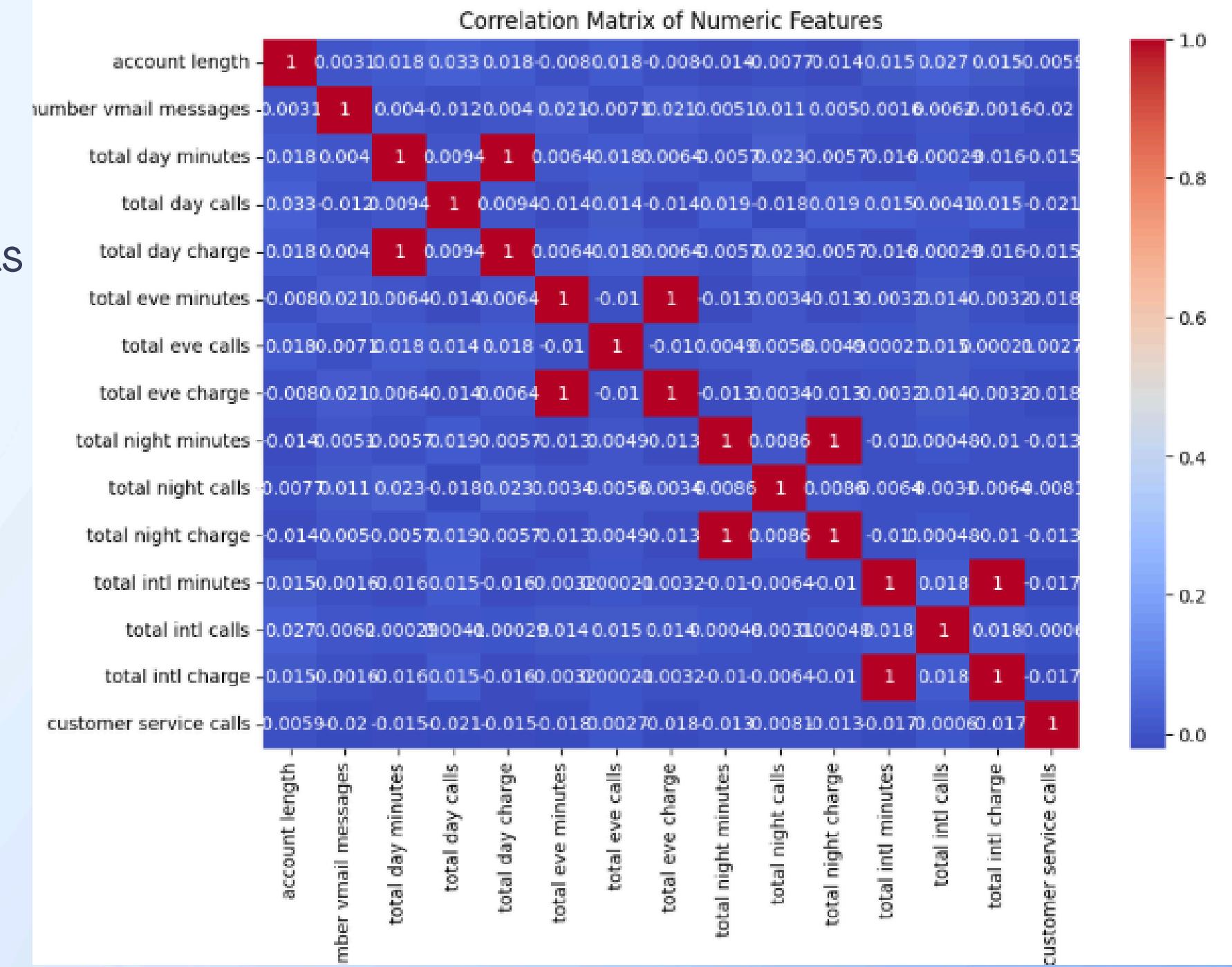
Feature Analysis & Churn Impact

- Total Day Minutes vs. Total Day Charge (Top Left Chart):** A perfect linear relationship, confirming that charges are a direct function of minutes used.
- Total International Minutes vs. Total International Calls (Bottom Left Chart):** No clear trend, meaning international usage patterns vary significantly among customers.
- Customer Service Calls & Churn (Top Right Chart):** Churned customers tend to make more service calls, indicating that customer dissatisfaction is a key driver of churn.
- Total Evening Minutes & Churn (Bottom Right Chart):** No significant difference in evening minutes between churned and non-churned customers, implying that evening usage is not a strong predictor of churn.



Correlation Matrix of Numeric Features

- The correlation matrix reveals that total day minutes and total day charge are highly correlated, which is expected as charges are directly proportional to minutes used.
- Similar high correlations exist between total evening minutes and total evening charge, as well as total night minutes and total night charge.
- Customer service calls show little correlation with other numeric features but play a significant role in churn prediction.



Conclusion

Best Model: Tuned Random Forest provided the highest accuracy and balanced performance.

- **Customer Service Calls:** A strong churn predictor, suggesting that dissatisfied customers contact support frequently before leaving.
- **International Plan:** Customers on international plans show higher churn rates, possibly due to pricing issues or alternative providers.
- **Account Length & Usage:** These factors had minimal impact, indicating that tenure alone does not determine loyalty.
- **Decision Tree vs. Random Forest:** Random Forest showed better generalization and predictive power compared to other models.
- **Model Improvement:** Tuning hyperparameters further enhanced model accuracy and precision.

Recommendations

1. **Deploy the Tuned Random Forest Model** for real-time churn prediction within SyriaTel's operational system.
2. **Enhance Customer Support** by reducing service complaints and improving issue resolution times.
3. **Focus on High-Risk Customers** through personalized retention campaigns and incentives.
4. **Regularly Update the Model** to adapt to changing customer behavior and market trends.
5. **Monitor Model Performance** continuously to ensure accuracy and effectiveness.

Findings

- The Random Forest model achieved 93.2% accuracy, making it the best-performing model for predicting customer churn.
- Customer service calls and international plans were the strongest churn indicators.
- Usage features (total day minutes, charges) showed minimal impact on churn.
- Account length had little correlation with customer loyalty.

Next Steps

- Implement the Random Forest Model to proactively identify customers at risk of churning.
- Enhance Customer Support by addressing service complaints efficiently.
- Develop Targeted Retention Strategies with personalized offers and engagement tactics.
- Monitor and Evaluate Model Performance regularly to ensure sustained accuracy.
- Continuously Update the Model with new data to adapt to evolving customer behaviors.

Thank You