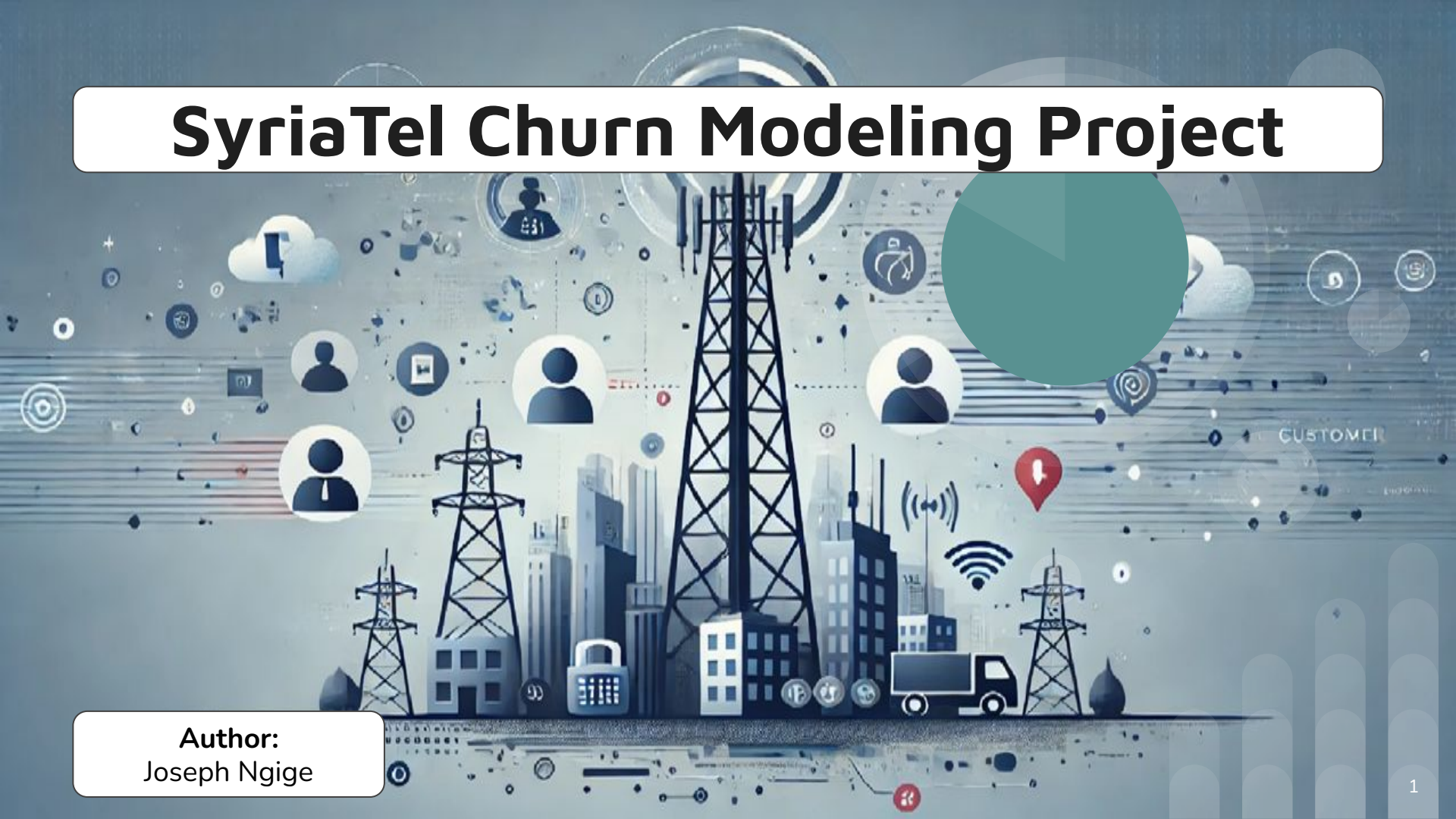
SyriaTel Churn Modeling Project



Author:
Joseph Ngige

Project Overview

Objective: Develop a classification model to predict customer churn at SyriaTel, helping the company reduce churn by identifying at-risk customers and implementing targeted retention strategies.

Outcome: Iterated on predictive models, with Model 5 (Tuned Decision Tree) identified as the best performer. Key recommendations for SyriaTel include:

1. **Deploy Model 5** to target at-risk customers with:
    - High usage plans, Improved international service; Enhanced customer service and Region-specific retention strategies.
2. **Monitor and refine the model** to ensure continued effectiveness.
3. **Explore ensemble methods** for potential performance improvements.

Outline

- Business Understanding
- Data Understanding
- Data Analysis & Preparation
- Modeling
- Evaluation
- Conclusion, Recommendations & Next Steps

Business Understanding

Challenge:

Customer churn is a critical issue for SyriaTel, leading to revenue loss and higher customer acquisition costs.

Objective:

Identify behavioral patterns that predict customer churn, enabling proactive engagement with at-risk customers.

Impact:

By targeting likely churners with focused retention strategies, SyriaTel can optimize marketing efforts, improve profitability, and sustain customer loyalty.

Data Understanding

Curated dataset containing information about SyriaTel customers and includes features related to customer behavior, demographics, and service usage.

Dataset Overview:

- **Rows:** 3,333
- **Columns:** 21
- **Key Features:** State, Account length, International plan, Voicemail plan, Total day/evening/night/international minutes, Customer service calls.
- **Target Variable:** Binary indicator of customer churn (True/False).

Provides a comprehensive foundation for building a predictive model to identify at-risk customers at SyriaTel.

Data Analysis & Preparation

Key Insights from EDA:

- Identified an imbalance in the churn class (only 14.49% churned).
- Strong correlation found between high daytime usage, frequent customer service calls, and increased churn.
- Significant variation in churn rates across different states.

Data Preparation Steps:

- Removed irrelevant/redundant columns (e.g., phone number, area code) to prevent multicollinearity.
- Split the dataset into training and test sets (80:20) before pre-processing to avoid data leakage.
- Transformed categorical variables to numerical format to enhance model accuracy and predictive power.

Data Analysis & Preparation

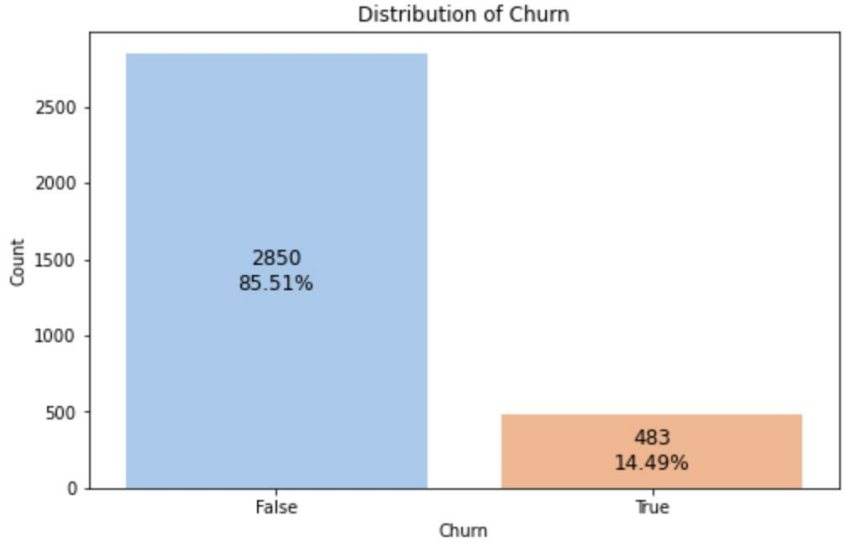


Figure 1: Churn Count Plot - *Indicates that the churn class is imbalanced, with fewer customers labeled as churned, 483 (14.49%) compared to those who did not churn, 2,850 (85.51%).*

Data Analysis & Preparation

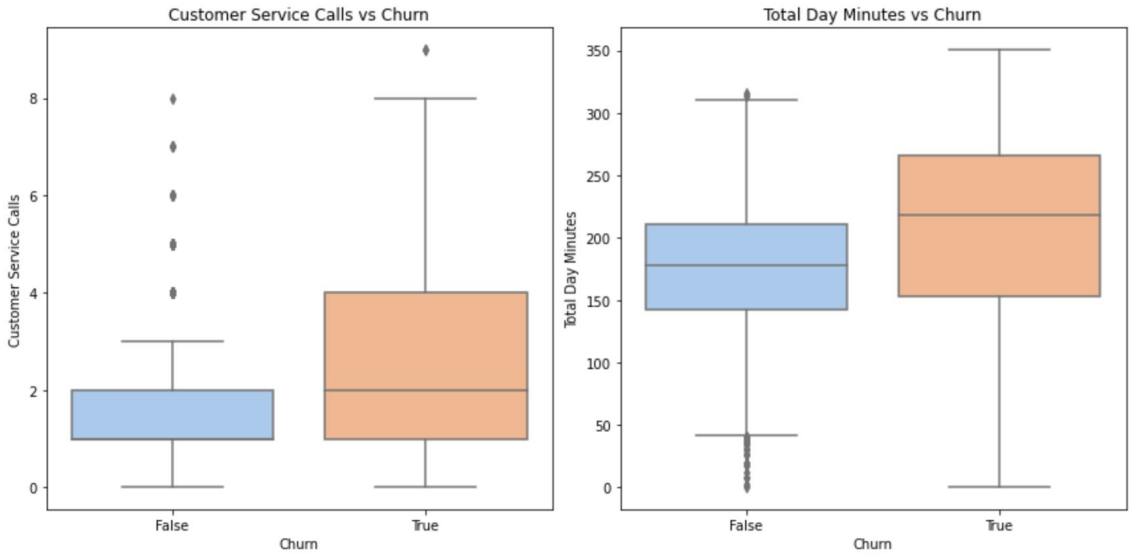


Figure 2: Churn Box plots - *Indicates that customers who churn tend to have a higher median customer service calls and total day minutes usage.*

Modeling

Approach: Employed an iterative modeling process, starting with basic models and increasing complexity.

Key Models:

- **Model 1:** Simple Logistic Regression struggled to identify churners.
- **Model 2:** Applied SMOTE to address class imbalance, improving churn detection but increasing false positives.
- **Model 3:** Tuned Logistic Regression, optimizing regularization (C) and penalization (l1/l2) to enhance performance.
- **Model 5:** Tuned Decision Tree, adjusted hyperparameters (max_depth, min_samples_split, min_samples_leaf) to balance accuracy and robustness.

Outcome: Model 5 emerged as the best-performing model, effectively balancing accuracy and robustness in predicting customer churn.

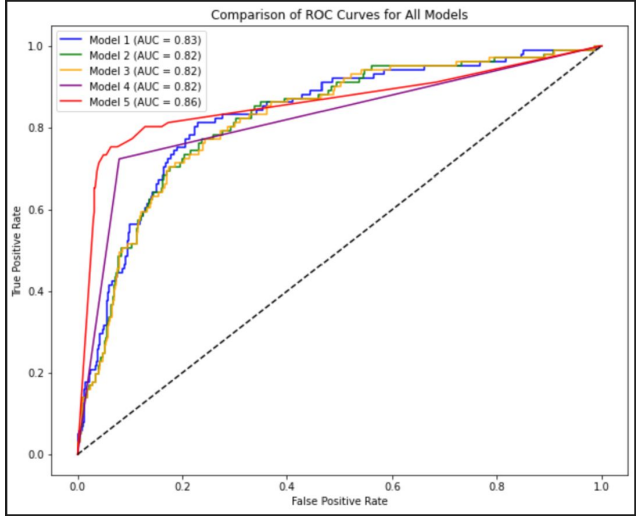
Evaluation

Key Metrics: Evaluation centered on finding the best model for predicting customer churn.

- **Accuracy:** Overall measure of correct predictions but can be misleading with imbalanced data.
- **Recall:** Prioritized for identifying most customers likely to churn (Crucial).
- **Precision:** Ensures accuracy in predicting churn.
- **F1-Score:** Balances Precision and Recall.
- **AUC:** Measures model's ability to distinguish between churners and non-churners.

Figure 4: *Model Summary Scores*

Model	Accuracy	Precision (Churn)	Recall (Churn)	F1-Score (Churn)	AUC
Model 1 (Logistic Regression)	0.859070	0.606061	0.198020	0.298507	0.83
Model 2 (Logistic Regression with SMOTE)	0.766117	0.368421	0.762376	0.496774	0.82
Model 3 (Tuned Logistic Regression)	0.761619	0.360677	0.742574	0.485437	0.82
Model 4 (Decision Tree)	0.890955	0.618644	0.722772	0.666667	0.82
Model 5 (Tuned Decision Tree)	0.917541	0.871795	0.732673	0.796904	0.86

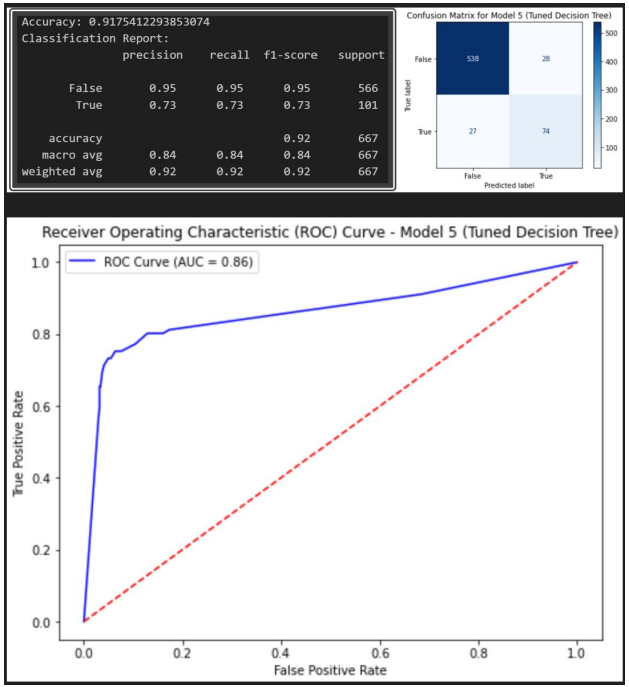


Evaluation

Final Model:

- **Tuned Decision Tree (Model 5):** Best performer after iterative testing and refinement.
- **Key Metrics:**
  - Accuracy: 91.75%
  - AUC: 86%
  - Recall: 72.28%
  - Precision: 87.15%
- **Outcome:** Model 5 effectively balances accuracy and robustness in predicting customer churn.

Figure 4: *Model 5 Scores*



Conclusion

Model 5 effectively predicts customer churn for SyriaTel, enabling targeted retention strategies to reduce churn, optimize resources, boost loyalty, and increase revenue.

Recommendations & Next Steps:

1. **Implement Targeted Retention Campaigns:** Personalize offers for high-risk customers (e.g., high usage plans, improved international service); Enhance customer service and adopt region-specific strategies.

2. **Monitor & Refine the Model:** Regularly retrain the model with new data to adapt to changing behaviors.

3. **Explore Ensemble Methods:** Consider methods like Random Forest or Gradient Boosting to enhance model performance.

Any Questions?

Thank You!

Email: elngige@gmail.com
GitHub: @engige
LinkedIn: linkedin.com/joseph ngige