Cracking the Churn Code: A Data-Driven Strategy for SyriaTel

Transforming Customer Behavior into Actionable Retention Insights

Author: Laban Leploote

Business Understanding

Business Problem:

SyriaTel faces a "revolving door" of customers. High churn rates lead to massive revenue loss and inflated costs, as acquiring new customers is **5 to 10 times more expensive** than retaining existing ones.

Key Points:

Service Fatigue: Frequent support calls signal deep dissatisfaction.

Value Gap: High-usage and international users feel they aren't getting enough value for their spending.

Competitive Pressure: Loyal customers are susceptible to better offers if they feel neglected.

Goal: Shift from a reactive "wait and see" stance to a proactive strategy by predicting churn before it happens.

Objectives

Develop a Predictive Model

Build a high-performance machine learning model to classify at-risk customers accurately.

Identify Churn Drivers

Pinpoint the exact behaviors (e.g., usage hours, service calls) that lead to cancellations.

Optimize for Recall

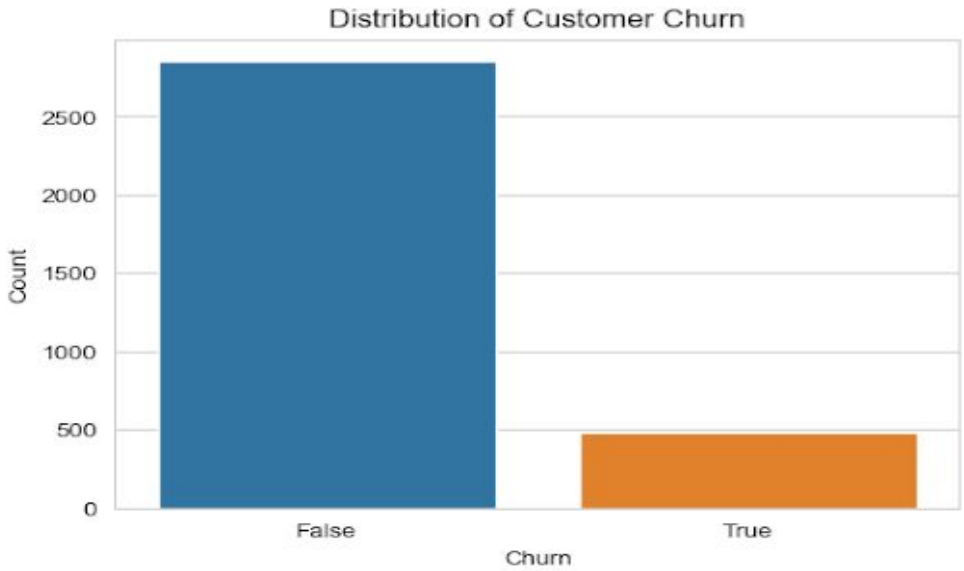
Prioritize identifying as many potential churners as possible to minimize lost revenue.

Deliver Actionable Insights

Provide the Marketing and Retention teams with data-driven strategies for personalized interventions.

Data Overview

- **Dataset:** SyriaTel Customer Churn data (3,333 records, 21 features).
- **Features:** Covers customer demographics, usage (Day, Eve, Night, Intl), and service metrics.
- **Churn Imbalance:** Only **14.5%** of customers churned. This imbalance required specialized techniques (like stratification and class weighting) to ensure the model could "see" the churners clearly.



Methodology – Preprocessing

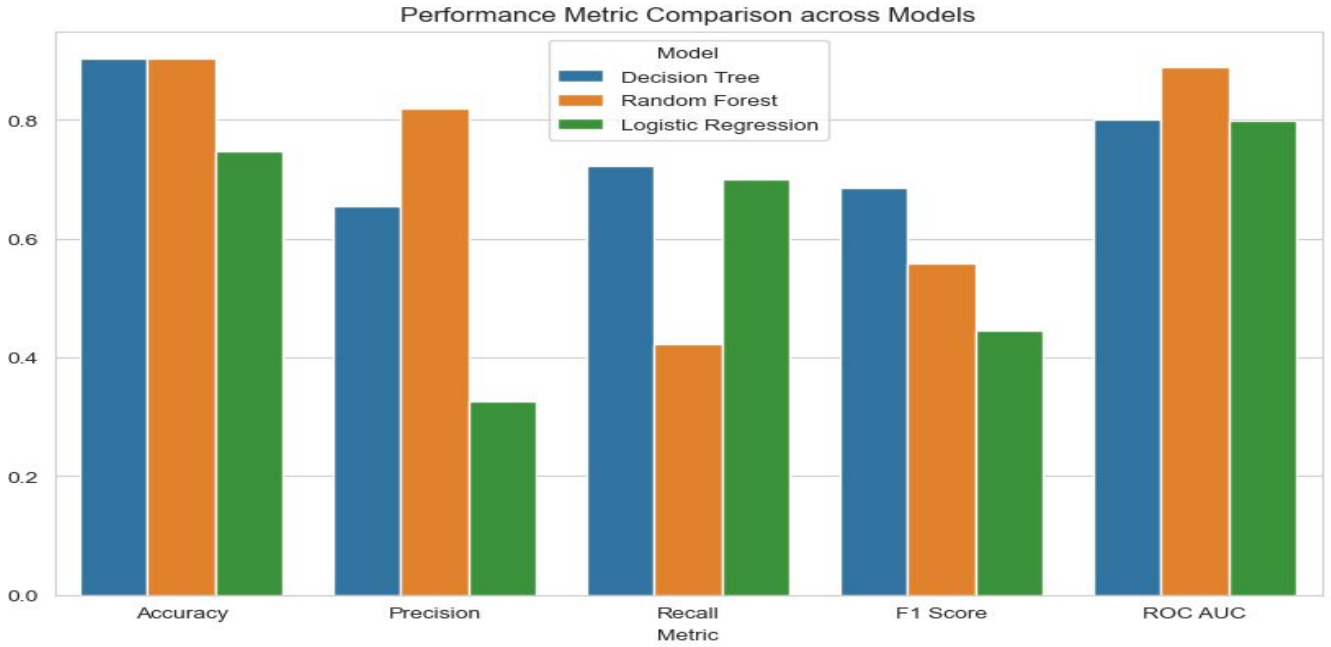
To ensure the models were accurate and unbiased, we performed the following:

- **Data Cleaning:** Standardized column names and confirmed zero missing values.
- **Feature Selection:** Dropped `phone_number` (unique ID) and `charge` columns (perfectly correlated with `minutes`) to prevent model redundancy.
- **Encoding:** Converted categorical data (e.g., `international_plan`, `state`) into numerical formats using Binary and One-Hot Encoding.
- **Splitting:** Used a **80/20 Stratified Split** to maintain the churn ratio in both training and test sets.
- **Scaling:** Applied **StandardScaler** to ensure high-value features (like total minutes) didn't overshadow smaller-scale metrics.

Model Comparison

We evaluated three core algorithms to find the best balance between catching churners and avoiding false alarms:

1. **Logistic Regression:** High recall (70%) but too many false alarms (low precision).
2. **Decision Tree:** A balanced performer but prone to overfitting without tuning.
3. **Random Forest (Winner):** Provided the best overall discrimination (ROC-AUC ~0.91) and was selected for further optimization.



Performance Results

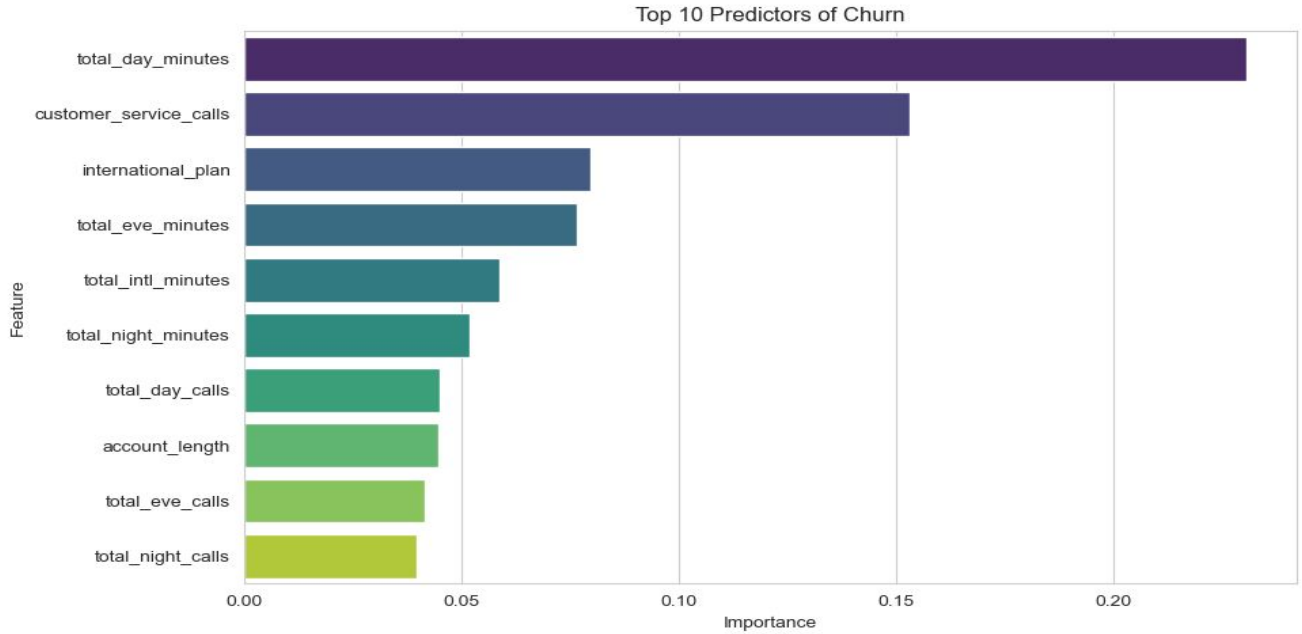
After hyperparameter tuning using `GridSearchCV`, the **Tuned Random Forest** became our champion model:

METRIC	SCORE	BUSINESS INTERPRETATION
Accuracy	92%	High overall reliability
Recall	66%	Identifies 2 out of 3 churners before they leave
Precision	80%	4 out of 5 are correct ensuring budget efficiency
F1-score	0.72	Excellent balance between catching churn and saving cost

Top 10 Drivers of Churn

The model revealed that churn isn't random—it's driven by three main factors:

1. **Total Day Minutes:** Heavy daytime users are highly price-sensitive and likely to switch for better rates.
2. **Customer Service Calls:** The most actionable signal; churn spikes dramatically after the **3rd call**.
3. **International Plan:** A "Value Gap" exists where international users don't feel the plan matches their needs.



Recommendations for Retention

The Proactive "3-Strike" Rule: Automatically flag any customer reaching their 3rd service call for a "Service Recovery" outreach by a senior agent.

High-Usage Loyalty Programs: Offer "Power User" discounts to the top 10% of daytime users to prevent them from being poached by competitors.

International Plan Audit: Review international rates against competitors and offer free trials to high-volume international callers who haven't yet joined the plan.

Probability-Based Intervention:

- **High Risk (>60%):** Direct calls and substantial renewal discounts.
- **Medium Risk (30-60%):** Automated "check-in" emails and feature highlights.