Predictive Churn Modeling Optimizing Customer Retention for SyriaTel

Western Onzere

Business Understanding

Problem Statement

- SyriaTel, like all telecom companies, faces customer churn—where users discontinue services. Churn presents a major financial risk due to:
  - High Customer Acquisition Costs – Retaining existing customers is more cost-effective than acquiring new ones.
  - Revenue Loss – Lost customers reduce recurring revenue, impacting profitability.
  - Competitive Pressure – Customers can easily switch to competitors, making retention critical.

Business Objectives

Objective:

- Identify high-risk customers early so SyriaTel can intervene before they churn.

Key Goals:

- Predict churn risk to prioritize early intervention—retention is cheaper than acquisition.
- Reduce revenue loss by retaining high-risk customers with targeted incentives.
- Minimize false negatives to ensure actual churners are detected.

Key Trade-off:

  - SyriaTel prioritizes catching actual churners over avoiding false alarms. False negatives (missed churners) are costlier than false positives (offering discounts to loyal customers).

Key Metric

Business Impact of Missing Churners

- False negatives (missed churners) mean lost customers and lost revenue.
- Retaining a customer is cheaper than acquiring a new one.
- High recall ensures we identify most customers at risk before they leave.

Trade-off: Recall vs. Precision

- Increasing recall may lead to more false positives (loyal customers misclassified as churners).
- However, wrongly offering a discount to a few loyal users is less costly than missing real churners.

Adjusting the Probability Threshold

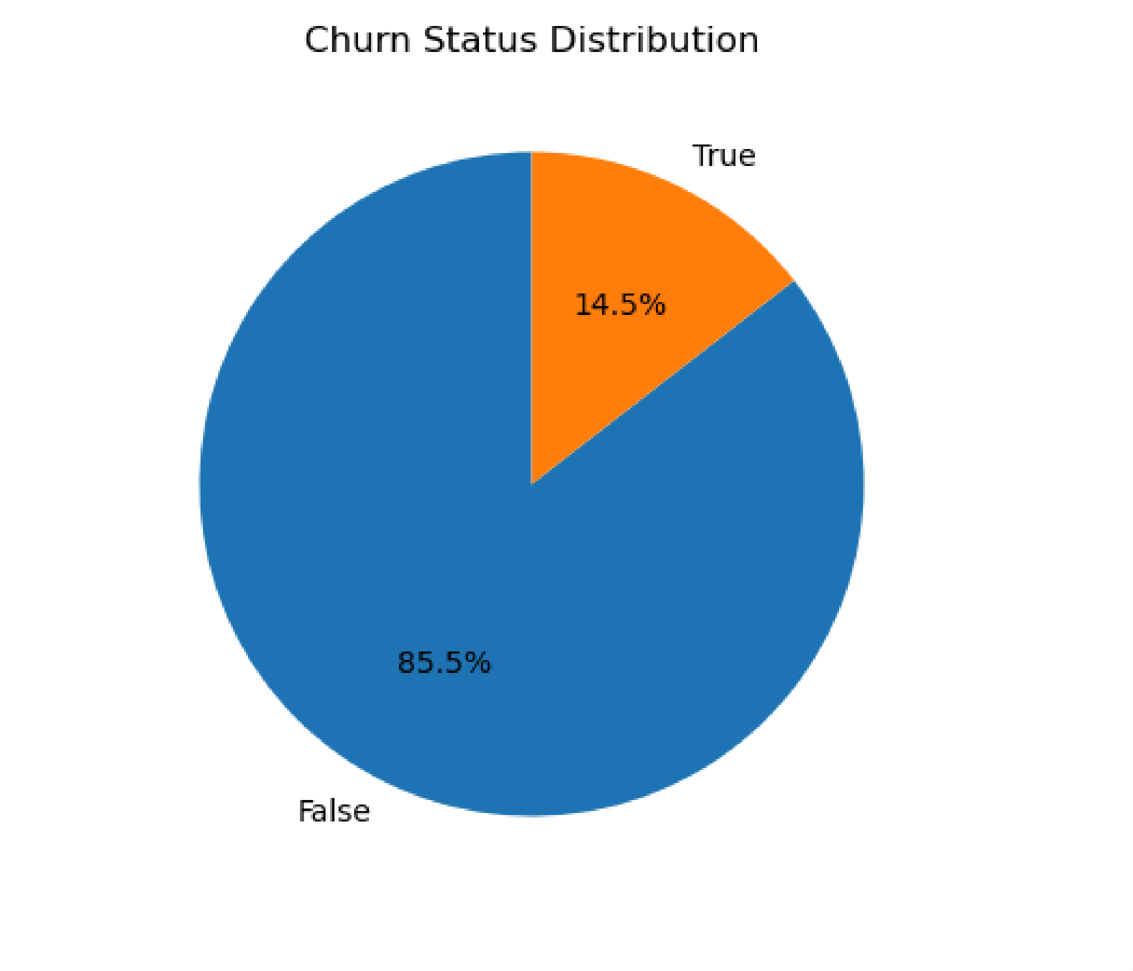
- Lowering the classification threshold can increase recall.

Data Understanding

- Total Records: 3,333 customers
- Key Features
  - Customer Information: state, account length, area code, phone number
  - Service Plans: international plan, voice mail plan, number vmail messages
  - Usage Metrics:
    - Daytime, Evening, Nighttime, and International
- Customer Interaction: customer service calls
- No missing values
- No duplicate records found

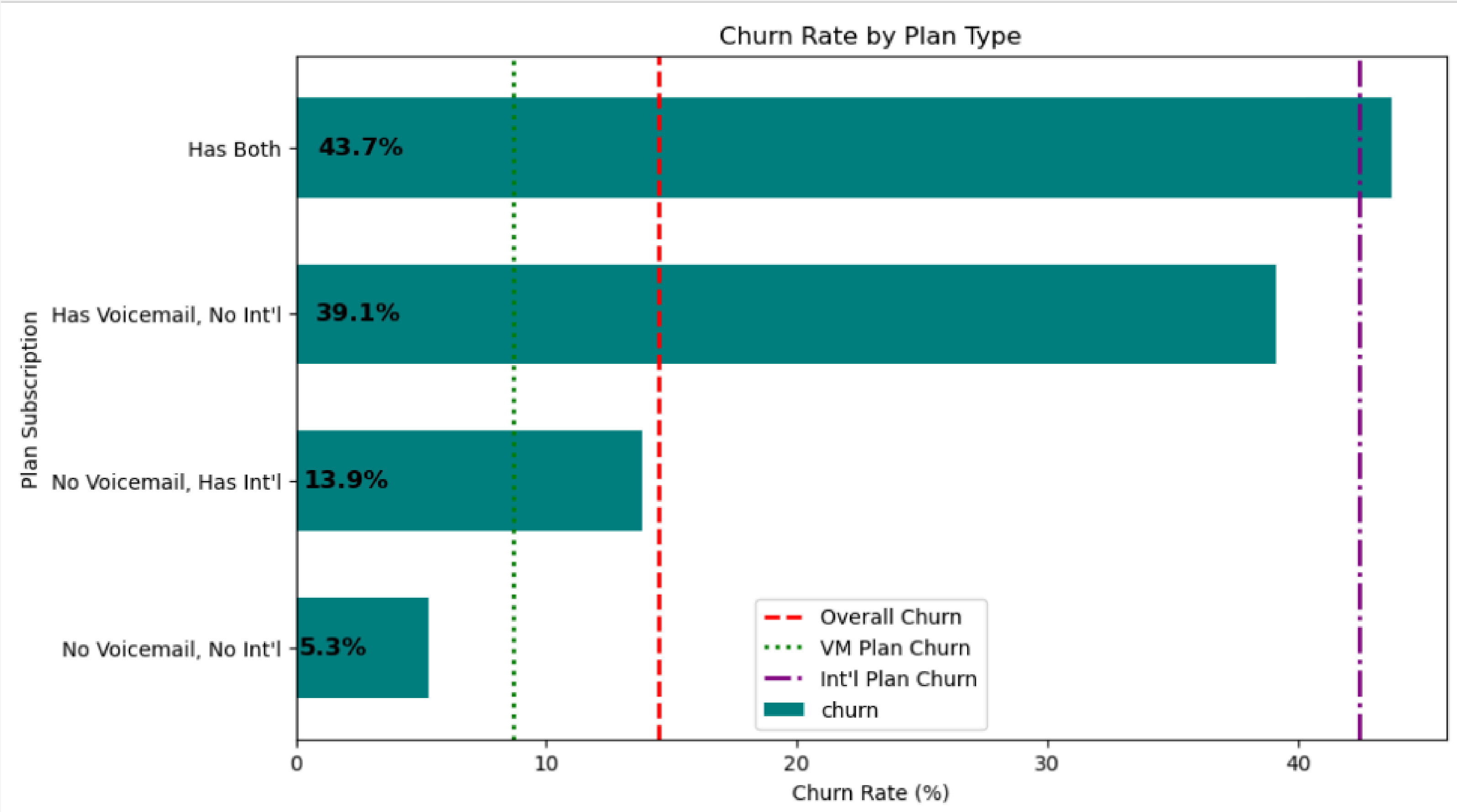
Data Understanding

- **Target Variable Distribution**
  - Churn Rate: 14% churners, 86% non-churners (requires balancing strategies).
- **Feature Engineering**
  - Total Usage: Combined total minutes, total calls, and total charges across all time periods.
  - Average Call Duration: total minutes / total calls (helps understand customer engagement).
  - Activity Score: Composite score based on total usage & customer service interactions.
- **Outlier Detection & Handling**
  - Checked for extreme usage values that could skew model performance.



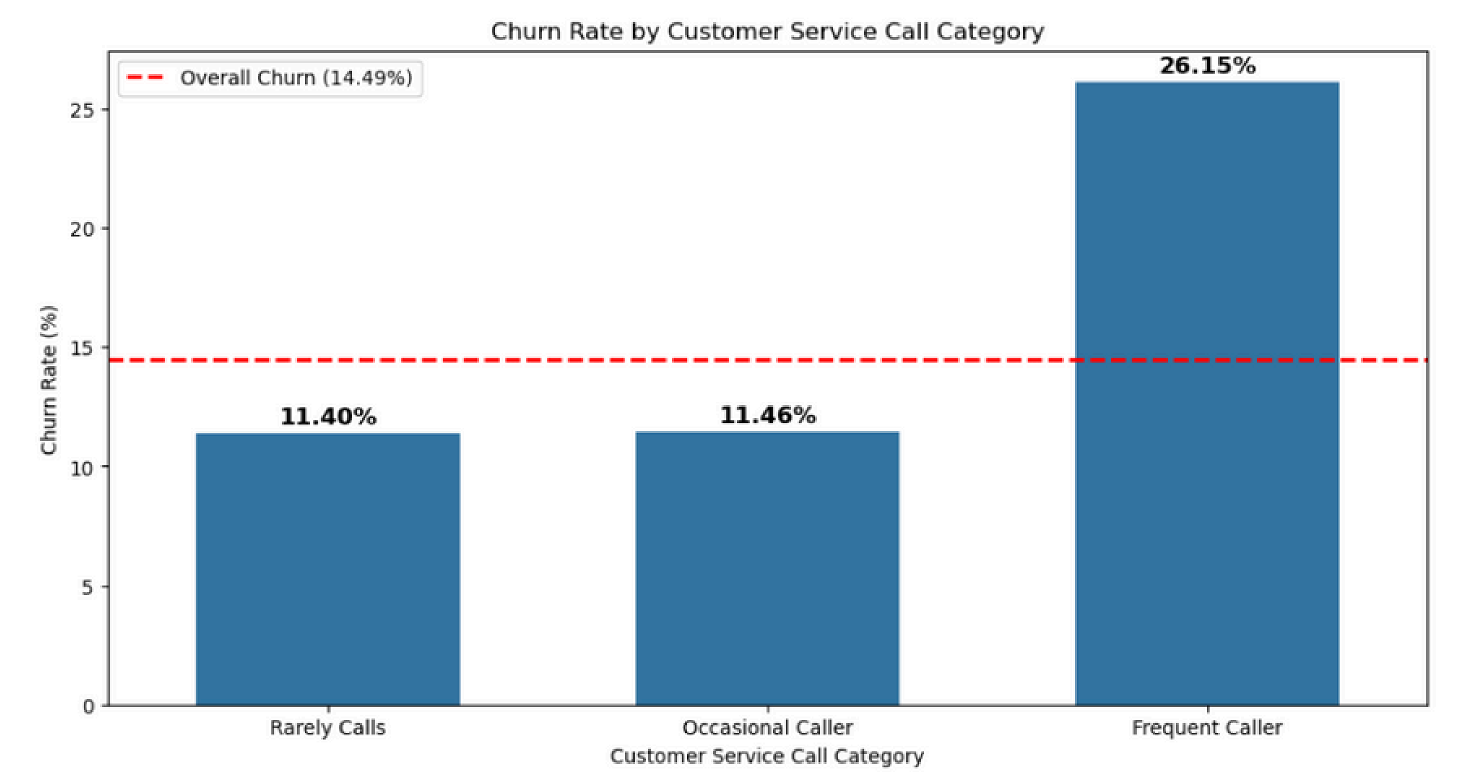
Exploratory Data Analysis: Customer Behavior

- **Churn Distribution & Service Plans**
  - Overall churn rate is ~14.5%.
  - Customers with an international plan tend to experience higher churn.
  - Voice mail plan patterns reveal that while fewer subscribe, plan type influences churn dynamics.
- **Usage & Billing Trends**
  - Elevated total minutes and charges are linked with a higher likelihood of churn.
  - Binned analyses indicate that medium-to-high usage segments show noticeable increases in churn.



Key Insights & Patterns

- **Service Plan Impact**
  - Analysis of voice mail plans shows that non-subscribers form a larger churn group
  - The combination of both plans (voice mail and international) is associated with the highest churn risk.
- **Customer Behavior & Interaction**
  - Higher volume of customer service calls correlates with increased churn.
  - Churn rates vary significantly by state, suggesting geographical factors affect customer retention.
- **Statistical Findings**
  - Significant predictors ($p < 0.05$) include number of voicemail messages, total international minutes/calls/charge, customer service calls, and other usage-based metrics.



Modeling

Baseline – Logistic Regression (High Accuracy, Poor Recall)

- Accuracy: ~87%, but Recall for churners: 28% → Missed most at-risk customers.
- Identified non-churners well, but performed poorly in flagging actual churners (High false negatives)

Decision Tree – Higher Recall but Overfitting Risk

- Recall jumped to ~81%, capturing far more churners.
- Better class separation, meaning the model distinguished churners from non-churners much better.
- Overfitting risk – model performed well on training data but lacked generalizability.

Model tuning

SMOTE – Addressing Class Imbalance

- Applied SMOTE to create a balanced dataset (upsampling minority class).
- Higher false positives, leading to unnecessary retention efforts for loyal customers.

Hyperparameter Tuning – Limited Improvement

- Random search tuning applied to Decision Trees
- Small improvements in precision but minimal recall gains.(Optimization didn’t enhance performance).

Threshold adjustments (0.33)

- Slightly Lower Recall (80%) from the default model (81%)
- Significantly Lower Precision. Precision drops from 84% to 71%, meaning more false positives
- Reduced AUC Scores – ROC (0.87) & PR (0.77) weaker ability to separate churners from non-churners

Best Model : Random Forest

Threshold Optimization

- Adjusting the classification threshold to 0.33 led to significant recall gains (83%).

Final Model Performance:

- Recall (Churners): 83% → A major improvement in detecting at-risk customers.
- Precision: Still strong (94%) ensuring reliable churn predictions.
- ROC-AUC: 0.91 → Excellent class separation.
- PR-AUC: 0.90 → Strong ability to identify churners in imbalanced data.

Impact:

- Early Intervention: Captures 83% of at-risk customers, allowing proactive action.
- Balanced Trade-off: Minimizes false negatives (missed churners) while keeping false positives under control.

Conclusion

Objective: Identify and Retain High-Risk Customers

- Successfully developed a Random Forest model with a threshold of 0.33 to predict churn.
- Ensures early intervention and minimizes customer loss.

Minimizing False Negatives (Missed Churners)

- Recall improved to 83%, capturing most at-risk customers.
- Directly aligns with SyriaTel’s priority of reducing false negatives.

Optimized Performance with Strong Model Metrics

- Precision: 94%, reducing unnecessary retention efforts.
- ROC-AUC: 0.91, showing strong separation between churners and non-churners.
- PR-AUC: 0.90, confirming the model’s effectiveness in handling imbalanced data.

Recommendations

- Deploy the Random Forest Model with threshold 0.33 to maximize recall while keeping false positives manageable.
- Implement a Model Monitoring System to track recall, precision, and business impact over time, ensuring continuous effectiveness.
- Retrain the Model Regularly with updated data to adapt to changing customer behaviors and improve predictive power.
- Leverage Explainable AI (SHAP, Feature Importance Analysis) to understand key drivers of churn and refine intervention strategies.

WESTERN ONZERE

+254 71-216-0286

onzerewestern@gmail.com

https://www.linkedin.com/in/western-onzere-ml/

https://www.youtube.com/@MinichAnalytics/featured