SyriaTel Customer Churn Prediction

Using Machine Learning to Improve Customer Retention

Prepared by: Priscilla Otieno

Audience: Business & Technical Stakeholders

Outline

1. Beginning
2. Overview
3. Business understanding
4. Modelling
5. Evaluation
6. Recommendations
7. Next Steps

Why Customer Churn Matters

SyriaTel is experiencing customer churn, leading to revenue loss. Acquiring new customers is significantly more expensive than retaining existing ones, and churn often becomes visible only after the customer has already left.

> Key Business Question: Can we identify customers likely to churn early enough to intervene?

Project Goals

1

Predict Churn Likelihood

Determine which customers are most likely to churn in the near future.

2

Identify Key Drivers

Uncover the primary factors influencing customer churn behavior.

3

Support Proactive Strategies

Enable the implementation of targeted retention strategies before churn occurs.

Success Criteria

1

Prioritize Recall & F1-Score

Focus on accurately identifying churners rather than overall accuracy.

2

Actionable Predictions

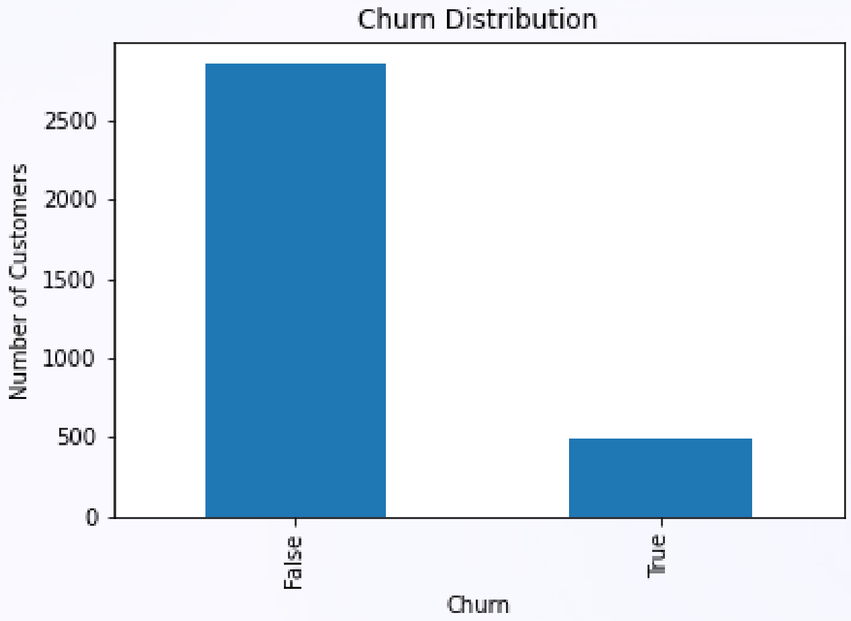
Translate model predictions into concrete, implementable business decisions.

Data Overview: Dataset Summary

- Customer-level telecommunications data, including:
  - Call usage patterns
  - Account and service plan details
  - Customer support interactions
- Target variable: Churn (Yes / No)

Key Observation

Churned customers represent a minority class, which informed our model evaluation strategy due to class imbalance.



This space will be used for the churn distribution plot from the notebook.

Modeling Approach: Why This Strategy?

1

Baseline Logistic Regression

Provided an interpretable and reliable benchmark for initial assessment.

2

Decision Tree Classifier

Capable of capturing non-linear relationships and highlighting key churn drivers.

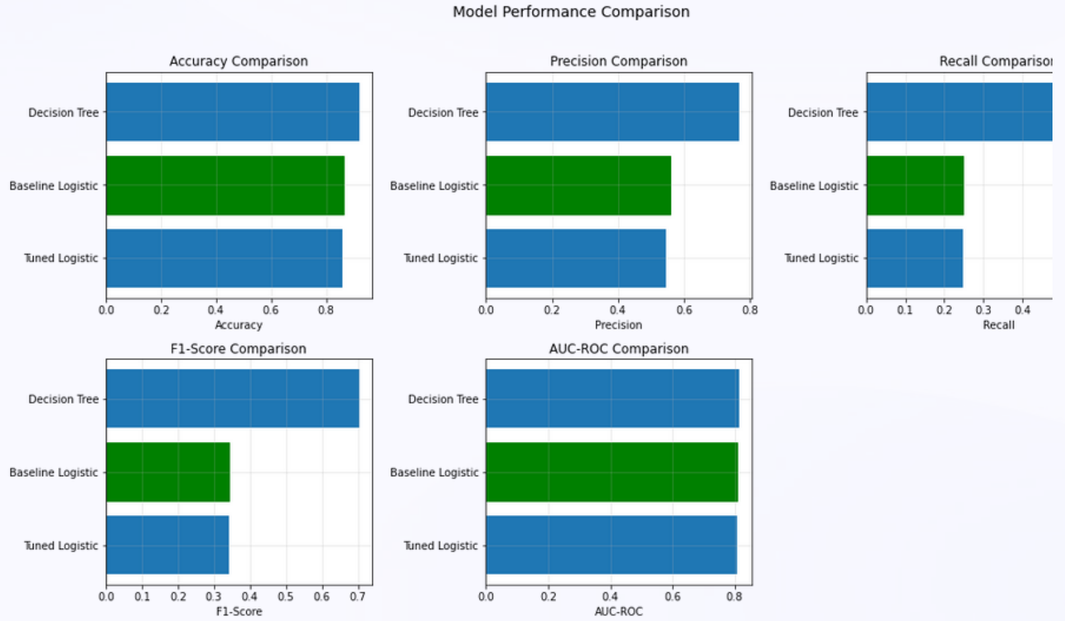
3

Tuned Logistic Regression

Optimized for enhanced performance while maintaining interpretability for business use.

This structured, incremental approach ensures transparency and measurable improvement in our predictive capabilities.

This space will be used for a decision tree or model overview visual from the notebook.



Evaluation Metrics: How Model Performance Was Measured

Accuracy

Overall correctness of the model's predictions.

Precision

Focuses on controlling costs associated with retention actions.

Recall

Measures the model's ability to identify all actual churners.

F1-Score

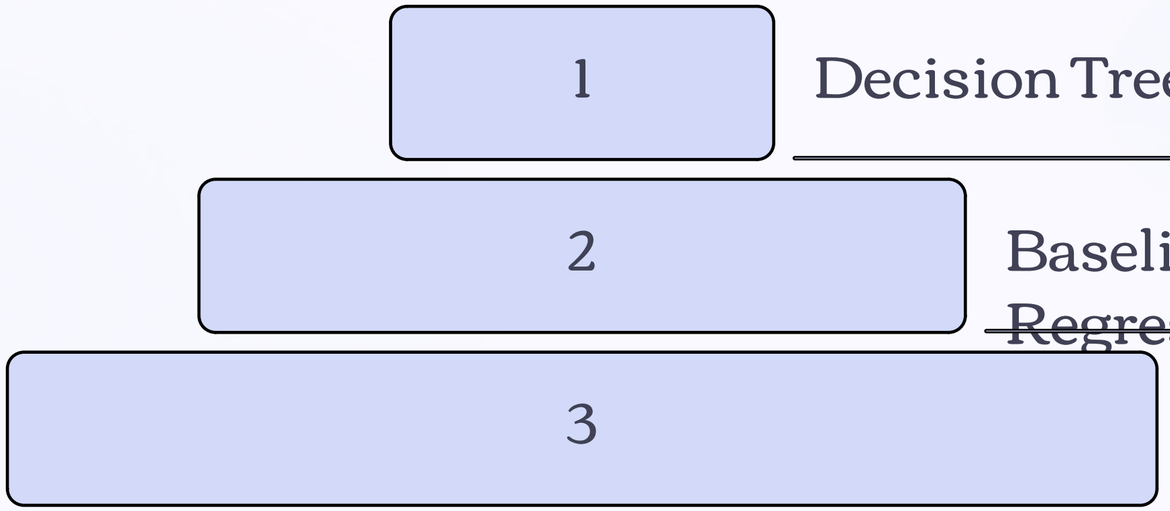
Provides a balanced measure between precision and recall.

AUC-ROC

Evaluates the model's capacity to distinguish churners from non-churners.

> Primary Focus: Missing a churner is more costly than contacting a non-churner.

Model Performance Comparison



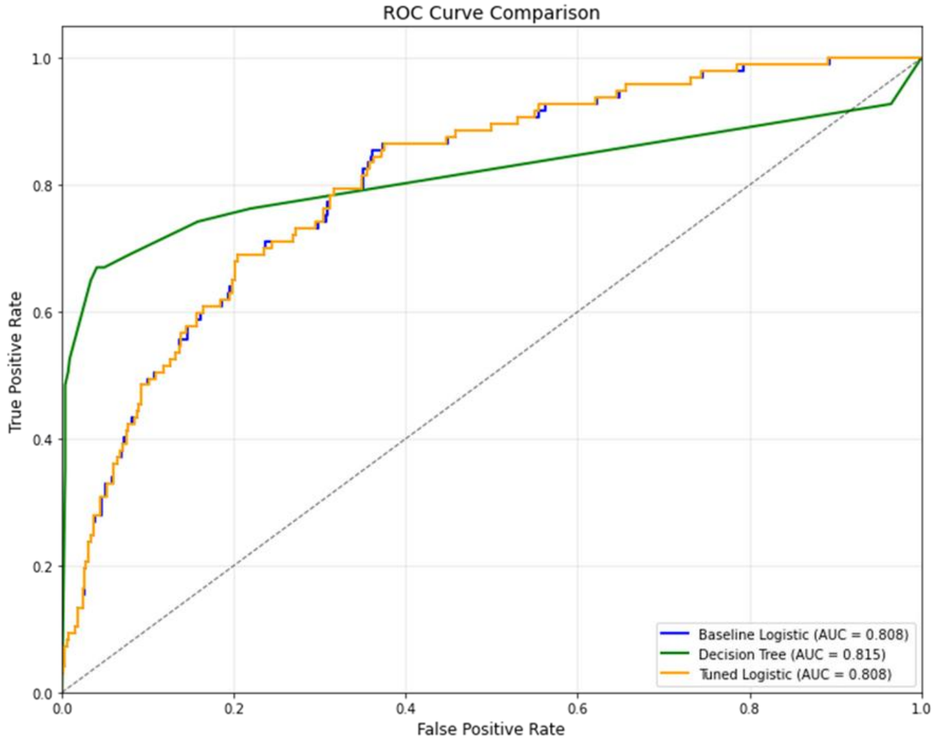
Decision Tree Classifier

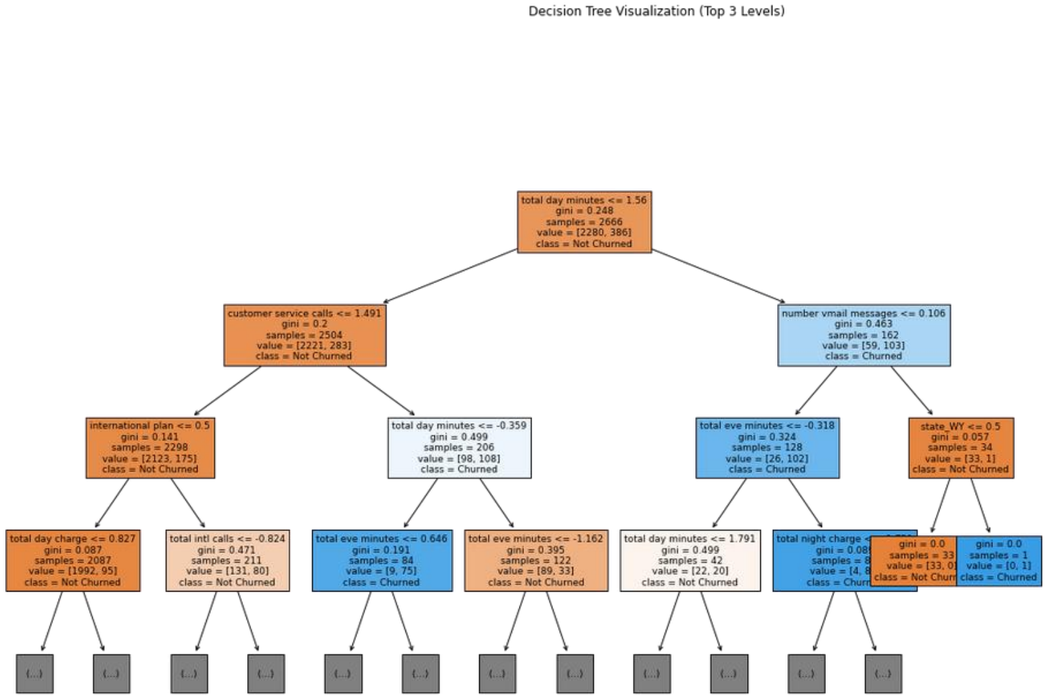
Baseline Logistic Regression

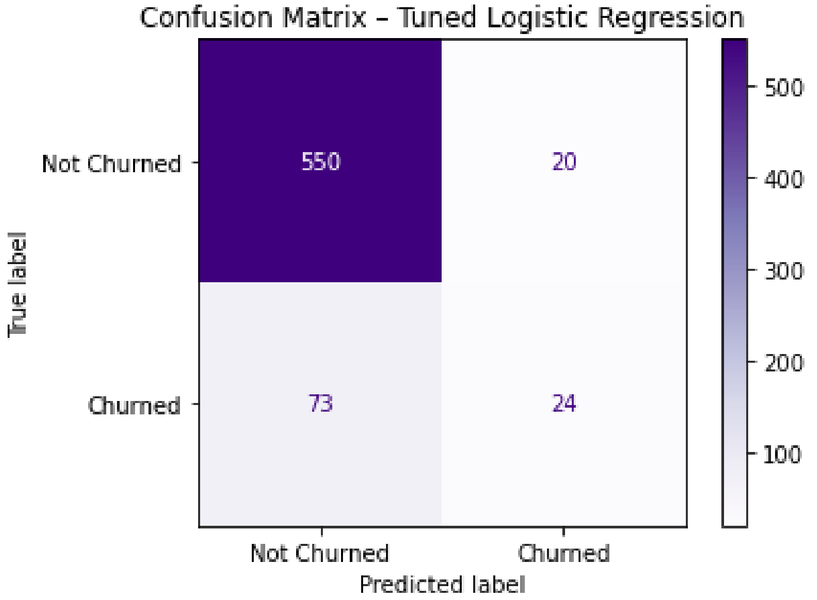
Tuned Logistic Regression (Best)

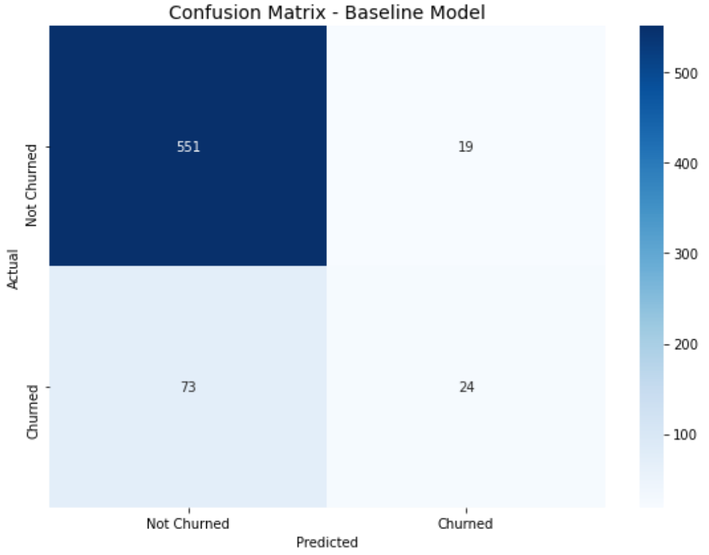
Why Tuned Logistic Regression Won

- Achieved the highest F1-Score among all models.
- Demonstrated strong recall on identifying actual churners.
- Offers a stable and interpretable solution for business application.


ROC Curve Comparison
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Baseline Logistic (AUC = 0.808)
Decision Tree (AUC = 0.815)
Tuned Logistic (AUC = 0.808)


Decision Tree Visualization (Top 3 Levels)
total day minutes <= 1.56
gini = 0.248
samples = 2666
value = [2280, 386]
class = Not Churned
customer service calls <= 1.491
gini = 0.2
samples = 2504
value = [2221, 283]
class = Not Churned
number vmail messages <= 0.106
gini = 0.463
samples = 162
value = [59, 103]
class = Churned
international plan <= 0.5
gini = 0.141
samples = 2298
value = [2123, 175]
class = Not Churned
total day minutes <= -0.359
gini = 0.499
samples = 206
value = [98, 108]
class = Churned
total eve minutes <= -0.318
gini = 0.324
samples = 128
value = [26, 102]
class = Churned
state_WY <= 0.5
gini = 0.057
samples = 34
value = [33, 1]
class = Not Churned
total day charge <= 0.827
gini = 0.087
samples = 2087
value = [1992, 95]
class = Not Churned
total intl calls <= -0.824
gini = 0.471
samples = 211
value = [131, 80]
class = Not Churned
total eve minutes <= 0.646
gini = 0.191
samples = 84
value = [9, 75]
class = Churned
total eve minutes <= -1.162
gini = 0.395
samples = 122
value = [89, 33]
class = Not Churned
total day minutes <= 1.791
gini = 0.499
samples = 42
value = [22, 20]
class = Not Churned
gini = 0.0
samples = 33
value = [33, 0]
class = Not Churned
gini = 0.0
samples = 1
value = [0, 1]
class = Churned
(...)


Confusion Matrix – Tuned Logistic Regression
True label
Predicted label
Not Churned
Churned
550
20
73
24
500
400
300
200
100


Confusion Matrix - Baseline Model
Actual
Predicted
Not Churned
Churned
551
19
73
24
500
400
300
200
100

Key Insights: What Drives Customer Churn

High Customer Service Calls

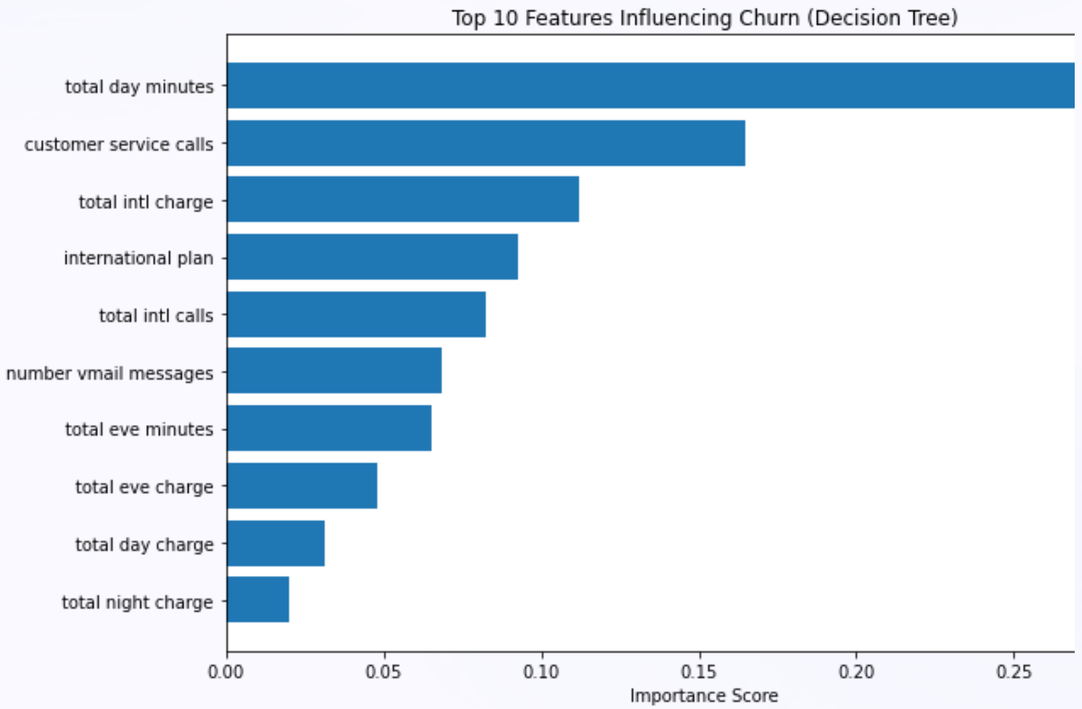
Frequent interactions with customer service often signal dissatisfaction.

Certain Service Plans

Specific service plans are correlated with higher churn rates.

Usage & Account Patterns

Distinct usage behaviors and account characteristics precede churn.



Business Value

These insights enable targeted interventions and support the efficient allocation of retention budgets.

This space will be used for top churn-driving features from the notebook.

Final Recommendation: Recommended Business Action

Deploy Tuned Logistic Regression Model

Implement the best-performing model to predict churn likelihood.

Focus Retention Efforts

Direct retention campaigns towards identified high-risk customers.

Design Targeted Offers

Leverage model insights to create personalized incentives and solutions.

Monitor & Retrain Periodically

Continuously assess model performance and update with fresh data.

Expected Impact

- Reduced customer churn rates.
- Improved Return on Investment (ROI) for retention initiatives.

Limitations & Next Steps

Limitations

- Model relies on historical data; customer behavior may evolve.
- Assumptions made regarding retention success rates.

Next Steps

- Incorporate additional customer feedback data for enriched insights.
- Explore cost-sensitive decision thresholds to optimize interventions.
- Integrate the model into live decision-making systems for real-time impact.