Customer Churn Prediction for SyriaTel

USING DECISION TREES AND REGRESSION MODELS TO IDENTIFY AT-RISK CUSTOMERS

Presented by: **Joram Mugesa**

Introduction

- SyriaTel, a leading telecommunications provider, faces the challenge of customer churn—when users stop using their services. High churn negatively impacts revenue and increases customer acquisition costs. To remain competitive, it's essential for SyriaTel to identify customers who are at risk of leaving and take proactive steps to retain them.

Stakeholders & Business Value

Stakeholders:

- 1. Customer Retention Team
- **2. Marketing Department**
- **3. Customer Service Team**
- **4.Product & Network Teams**
- **5.Executive Leadership**

Business Value:

- Business goal: Predict which customers are likely to churnOptimize investment & reduce risk
- Value: Enable targeted retention strategies and reduce revenue loss
- Approach: Built classification models using telecom usage data

Business Understanding

- SyriaTel loses revenue from customers who stop using services (churn)
- Early identification of churn allows personalized interventions
- Key question: Can we **predict churn** based on customer behavior?

Data Source

Source: bigml_data.csv

- Features include:
  - Demographics (e.g., state, account length)
  - Usage patterns (day/night/eve minutes & charges)
  - Plan subscriptions (international plan, voice mail)
  - Customer service interactions
  - Target: Churn(yes/No

Data Exploration & Cleaning

- Applied **One-Hot Encoding** to convert categorical data
- Scaled numerical features using **StandardScaler**
- Class distribution was highly imbalanced (majority = No Churn)
- Extracted temporal features (release month, year, quarter)
- Used **SMOTE** to balance the training dataset

Model 1 – Logistic Regression

High accuracy of 86% — overall good at classifying the dominant class (No Churn)

Very high recall for class 0 of 97% — almost never mislabels loyal customers

Interpretable coefficients — you can analyze which features increase or reduce churn risk linearly.

Very low recall for churners (0.25) — misses 75% of customers who actually churned.

Poor F1-score (0.34) for class 1 — weak balance between false positives and false negatives

The features and churn pattern might not follow a linear relationship, which logistic regression depends on

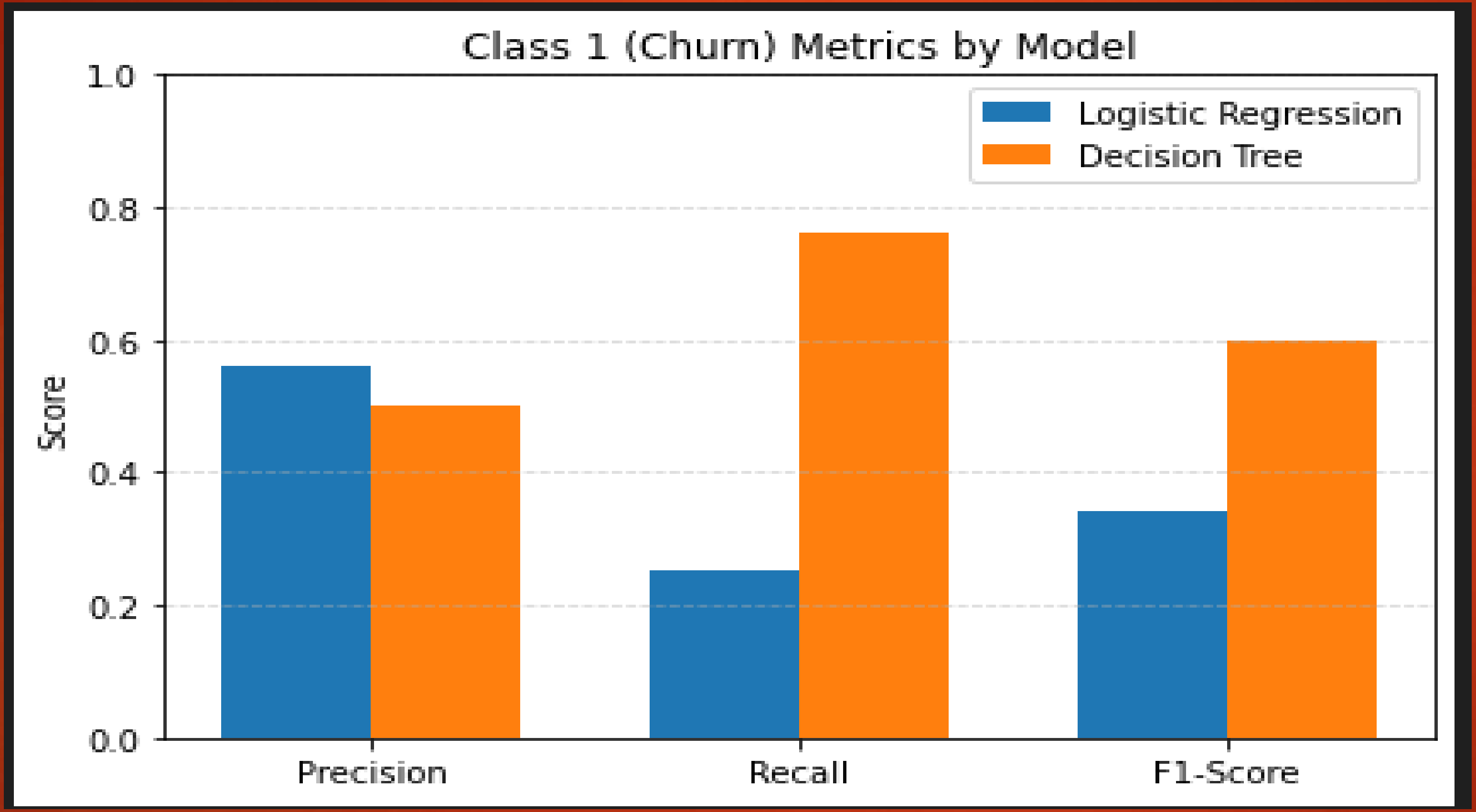
Logistic regression is too conservative — it prefers to avoid false positives, even if it means missing actual churners

Useful for understanding general trends but not reliable for churn prediction by itself.

Model 2 – Decision Tree (Recommended)

- Much higher recall for churners (76%) — captures the majority of customers who churn
- Balanced performance: F1-score for class 1 is 0.60, much better than logistic regression
- Captures nonlinear interactions like if customers have both an international plan and high service calls, they churn
- Better weighted and macro F1-scores — shows more balanced performance across classes
- Slightly lower accuracy than Logistic Regression (85% vs 86%)
- More prone to overfitting without proper pruning or depth control
- Less interpretable than logistic regression (but still explainable with tree diagrams and feature importance)

Model Comparison Chart



Business Insights

High churn risk linked to:

- Customers with international plans
- Frequent calls to customer service
- High daytime charges

Early intervention on these indicators can reduce churn

Recommendation

- Deploy the **Decision Tree model** to:

- Score existing customers weekly/monthly
- Flag high-risk accounts for retention follow-up

Use insights to drive **personalized offers and support**

Questions & Discussions