

kaigush / phase_3_project_moringa

🔍

📧

👤

<> Code

🔍 Issues

🔗 Pull requests

▶ Actions

📁 Projects

📖 Wiki

🛡 Security

📈 Insights

⚙ Settings

👁 Watch 0 ▾

🔗

★

▾

★ 0 stars

🔗 0 forks

👁 0 watching

🔗 Branches

📉 Activity

🏷 Tags

🌐 Public repository

🔗

🔗 1 Branch

🏷 0 Tags

🔗

🏷







🔍 Go to file t

Go to file



Add file ➕

Code

⋮

 kaigush first commit	d3c3473 · 9 minutes ago	
 README.md	first commit	9 minutes ago
 phase3_project.pdf	first commit	9 minutes ago
 presentation.pdf	first commit	9 minutes ago
 syriatelcurn.csv	first commit	9 minutes ago

📖 README



SyriaTel Customer Churn Prediction

Overview

This project focuses on predicting customer churn for SyriaTel, a telecommunications company. Churn occurs when customers stop using the service, directly impacting revenue. By identifying at-risk customers through behavioral patterns, SyriaTel can implement proactive retention strategies like personalized offers or improved service.

Business Understanding

The primary goal is to build a binary classification model to determine whether a customer is likely to leave (churn) or stay.

Target Variable: Churn (Yes / No).

Objectives:

- Identify high-risk customers.
- Provide probability scores for churn risk.
- Gain actionable insights into features driving churn.

Data Understanding

The dataset includes 3,333 customer records with 21 features.

Key Features:

Usage: Total day/evening/night/international minutes, calls, and charges.

Plans: International and voice mail plan status.

Service: Number of customer service calls.

Account: Account length and area code.

Data Integrity: The analysis confirmed zero duplicate records and zero missing values.

Methodology

The technical workflow included:

Data Preprocessing:

Encoding categorical variables (e.g., plan types).

Feature selection (dropping non-predictive identifiers).

Feature scaling using StandardScaler to normalize usage metrics.

Modeling:

Logistic Regression: Used as a baseline model to establish churn probabilities.

Decision Tree Classifier: Implemented to capture complex, non-linear relationships in customer behavior.

Evaluation Results

The Decision Tree model proved superior for business application:

Overall Accuracy: 90%.

Precision (Churners): 0.65 — 65% of predicted churners actually left, reducing "false alarm" marketing costs.

Recall (Churners): 0.73 — Caught 73% of actual churners.

F1-Score: 0.69 — Balanced performance significantly better than Logistic Regression (0.47).

Actionable Recommendations Service Call Thresholds: Frequent service calls are a high churn indicator.

Customers reaching 3+ calls should receive automatic loyalty check-ins.

High-Volume Targeting: High day charges correlate strongly with churn. Monitor these valuable users for retention offers before they depart.

Plan Personalization: Analyze users with international plans who have low usage to offer more cost-effective



Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)