

Kabutie / SyriaTel-

🔍

📧

👤

<> Code

⦿ Issues

🔗 Pull requests

▶ Actions

📁 Projects

📖 Wiki

🛡 Security

📊 Insights

⚙ Settings

👁

🔗

★

☆ 0 stars

🍴 1 fork

👁 0 watching

🔗 Branches

📈 Activity

🏷 Tags

🌐 Public repository

🔗

🔗 5 Branches

🏷 0 Tags

🔗

🏷

🔍 Go to file

t

Go to file

Add file ➕

Code

⋮

👤 Kabutie

Add files via upload

206bb8e · 8 minutes ago

🕒

📄 .gitignore	Resolve merge conflicts in .gitignor...	3 days ago
📄 Data Report.pdf	Add files via upload	10 minutes ago
📄 README.md	Replace README.md in main with v...	3 days ago
📄 bigml_59c28831336c6604c8...	Add files via upload	5 days ago
📄 presentation.pdf	Add files via upload	8 minutes ago
📄 student.ipynb	final	2 days ago

📖 README

✎ ⋮

# Customer Churn Prediction – SyriaTel Dataset

## Business Problem

Customer churn (when customers leave) is a major challenge for the telecom companies. Retaining customers is considered more cost-effective than acquiring new ones. By analyzing the historical customer data, we aim to identify patterns and predict which customers are more likely to churn. This allows the company to proactively take measures to improve customer retention.

**Goal:** Predict which customers are likely to churn and uncover what leads or affects churn. This enables the company to take proactive retention measures.

## Business Overview

The dataset comes from **Kaggle**, containing **3,333 customer records** with demographics, usage, and service-related outcomes.

The main task for the data set is to build a predictive model to classify whether a customer will:

- Stay (0)

- Churn (1)

## Business Understanding

### Stakeholders

- **Executives** – Ensure revenue growth and reduce churn.
- **Customer Service Teams** – Identify and retain at-risk customers(churners).
- **Data Science & Analytics Teams** – Build churn prediction models.
- **Marketing Teams** – Design targeted retention campaigns.

### Key Business Questions

1. What are the top churn factors driving customer churn?
2. How much revenue is lost due to customer churn over a given period?
3. Can a baseline machine learning model accurately predict whether a customer will churn?
4. Which machine learning model performs best in predicting churn based on classification metrics?
5. How does pricing impact customer retention, and what pricing strategies can be implemented to reduce churn without significantly impacting revenue?

## Data Understanding & Analysis

### Dataset Overview

Metric	Value
Total Records	3,333
Features	21
Target Variable	Churn (0 = Stayed, 1 = Churned)
Stayed Customers	2,850 (85.5%)
Churned Customers	483 (14.5%)

### Data Cleaning Performed

- Checked for missing values (none found).
- Removed irrelevant identifiers (e.g. phone number).
- Converted categorical variables into numerical (dummy encoding).
- Standardized column naming for consistency.
- Verified class balance and applied `class_weight="balanced"` in modeling.

### Data Analysis Highlights

- Customers making **frequent customer service calls** show higher churn likelihood.

- **International plan subscription** strongly correlates with churn.
- **Total day minutes/charges** contribute significantly to churn risk.
- Dataset is moderately imbalanced (**85/15 split**), requiring class-weighting techniques.

## Model & Results

Metric	Score
Accuracy	94%
Recall (Churned)	83%
Precision (Churned)	79%
F1-Score (Churned)	81%

## Confusion Matrix

	Predicted Stayed	Predicted Churned
Actual Stayed	564 (TN)	2 (FP)
Actual Churned	13 (FN)	88 (TP)

## Summary of Findings

- Around 15% of customers churned.
- Churn is strongly influenced by:
  - **International plan usage**
  - **High day-time calls/charges**
  - **Frequent customer service calls**
- Dataset imbalance exists, but **class weighting improved recall** for churners.
- Random Forest performed strongly with **high accuracy and balanced precision/recall**.

## Final Insights & Recommendations

1. **Active Retention:** Focus on customers with international plans and frequent service calls.
2. **Customer Support Quality:** Investigate service issues that trigger dissatisfaction.
3. **Targeted Offers:** Provide loyalty incentives to high-usage customers at risk.
4. **On-going Monitoring:** Initiate churn prediction in real-time dashboards.
5. **Data Expansion:** Include billing history, complaints, and competitor data for much more stronger insights.

## Releases

No releases published

[Create a new release](#)

## Packages

No packages published

[Publish your first package](#)

## Contributors 5



## Languages

● Jupyter Notebook 100.0%