

stellahmnene / Phase3\_project

🔍

🔒

📧

🖨

<> Code

🔴 Issues

🔗 Pull requests

🔄 Actions

📁 Projects

📖 Wiki

🛡 Security

📈 Insights

⚙ Setting

👁

🔗

☆

☆ 0 stars

🔗 0 forks

👁 1 watching

🔗 Branches

📈 Activity

🏷 Tags

🌐 Public repository

🔗 1 Branch

🏷 0 Tags

🔍 Go to file









Go to file

+



Add file

<> Code

⋮

 stellahmnene Delete Github in PDF.pdf	c892c6e · now	
 .gitignore	Initial commit	5 hours ago
 README.md	Update README.md	2 hours ago
 SyriaTel.pdf	Upload Powerpoint in PDF format	44 minutes ago
 SyriaTelChurn_Stella (Fin).pdf	Notebook to PDF	3 minutes ago
 SyriaTelChurn_Stella.ipynb	Created using Colab	5 hours ago
 SyriaTel_choladeck.pptx	Upload PPT	42 minutes ago

📖 README

# Telecom Churn Analysis and Customer Retention Strategies

## Overview

Customer churn is a critical issue for telecom companies, as losing customers leads to significant revenue loss and increased acquisition costs. This project aims to leverage machine learning to predict customer churn and provide actionable insights to improve customer retention strategies.

## Business Problem

Telecom companies face the challenge of customer attrition, where customers discontinue their services for various reasons, including pricing, service quality, and competition. Understanding the key drivers of churn and predicting which customers are likely to leave can help businesses take proactive measures to enhance customer satisfaction and retention.

Business Questions General Business Questions: -What are the key factors driving customer churn in the telecom industry?

W-hat customer segments are at the highest risk of churning?

-What strategic actions can telecom companies take to reduce churn based on model insights?

Modeling-Specific Business Questions: -What classification model performs best in predicting customer churn?

-What are the most important features influencing the model's predictions?

-How do different machine learning techniques (Logistic Regression, Decision Trees, etc.) compare in terms of accuracy, recall, and precision?

-How well does our model generalize to unseen data?

-What is the impact of hyperparameter tuning on model performance?

-What threshold should be used for classification to balance false positives and false negatives?

## Methods Applied

---

A structured data science approach was applied to predict customer churn. The dataset was first explored to understand its distribution, handle missing values, and encode categorical variables. Feature engineering was performed to enhance model performance. Multiple classification models were built, including Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). Hyperparameter tuning was applied to optimize model performance, and evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC were used to assess each model. The best-performing model was selected based on its ability to balance predictive power and interpretability, followed by business recommendations to help SyriaTel reduce customer churn.

## Results

---

The models were evaluated using key classification metrics, with Random Forest and Gradient Boosting emerging as the best-performing models based on accuracy, recall, and ROC-AUC scores. Random Forest provided a good balance between interpretability and performance, while Gradient Boosting offered a slightly higher predictive accuracy. Key features influencing churn included customer service calls, total day minutes, and total evening minutes, indicating that high service usage and frequent customer service interactions correlate with higher churn rates.

## Conclusion

---

The analysis revealed that customer dissatisfaction, high service usage, and frequent customer support interactions are strong predictors of churn. While advanced models like Gradient Boosting provided superior predictive power, simpler models like Logistic Regression offered easier interpretability. This suggests that businesses should balance predictive accuracy with actionable insights when deploying machine learning models for customer retention.

## Recommendations



### Releases

No releases published

[Create a new release](#)

### Packages

No packages published

[Publish your first package](#)

### Languages

- Jupyter Notebook 100.0%