

WaitingaJames /
dsc-phase-3-project-v2-3

<> Code

Pull requests

Actions

Projects

Wiki

Security

Insights

Settings

View license

0 stars

61 forks

0 watching

1 Branch

0 Tags

Activity

Public repository · Forked from [learn-co-curriculum/dsc-phase-3-project-v2-3](#)

main

1 Branch

0 Tags

Go to file

t

Go to file

+

Add file

Code

...

This branch is 4 commits ahead of [learn-co-curriculum/dsc-phase-3-project-v2-3:main](#) .

Contribute

Sync fork

WaitingaJames

Final Commit4

2 minutes ago

images	move image to folder	4 years ago
.canvas	copy edits	3 years ago
.gitignore	initial commit	4 years ago
CONTRIBUTING.md	Final Commit2	15 minutes ago
LICENSE.md	Final Commit2	15 minutes ago
Presentation.pdf	Final Commit	16 minutes ago
README.md	Final Commit	16 minutes ago
README.pdf	Final Commit	16 minutes ago
Syriatel Data Analysis Project.pptx	Final Commit	16 minutes ago
phase3project-2new.ipynb	Final Commit	16 minutes ago
phase3project-2new.pdf	Final Commit	16 minutes ago
syriatel.csv	Final Commit	16 minutes ago

README

License

Customer churn prediction model

Project Overview

In response to SyriaTel's pressing need to address customer churn, we propose developing a robust classifier to predict customer attrition. Leveraging predictive analytics, our goal is to uncover discernible patterns within SyriaTel's telecommunications data. By employing advanced machine learning techniques, specifically binary classification, we aim to equip SyriaTel with a predictive model capable of identifying customers likely to churn in the near future. Our primary focus is on assisting SyriaTel in minimizing revenue loss by proactively addressing customer retention. Through this project, we aim to unlock actionable insights that will empower SyriaTel to implement targeted strategies to mitigate customer churn, ultimately fostering long-term business sustainability

Research Questions

1. Which specific features in the SyriaTel dataset are the most significant predictors of customer churn, and how do these features influence the likelihood of churn?
2. How do various predictive models (Logistic Regression, Random Forest, Decision Trees) differ in their accuracy and ability to predict customer churn in the telecommunications sector, and which model provides the best balance between sensitivity (recall) and specificity?
3. To what extent does the preprocessing of the dataset, such as handling of categorical variables, normalization, and addressing class imbalance through techniques like SMOTE, affect the predictive performance of churn models?

Goals

1. Develop machine learning models capable of forecasting customer churn through the analysis of customer characteristics.
2. Evaluate the constructed machine learning models and ascertain the most precise model for prediction.
3. The objective of the analysis is to pinpoint the particular features that exert a substantial influence on the customer churn rate at SyriaTel. By leveraging these insights, valuable recommendations can be formulated to mitigate churn rates within the company, thus enhancing customer retention Data

Methodology

The project followed a standard machine learning workflow:

1. Data Loading & Cleaning
2. Exploratory Data Analysis
3. Preprocessing & Feature engineering
4. Model Development & Evaluation-Fitting various models
5. Model Optimization

Key libraries used include Numpy, Pandas, Matplotlib, Seaborn, Scikit-learn, and imblearn.

Key Findings

Some highlights from the analysis which showed a greater impact on the customer churn rate include: -• Total Day Minutes • Total International Minutes • Total Day Charge • Total Evening Charge • Customer Service calls • International Calls • Total Charge ()

Random forest model performed best with a higher level of accuracy with highest AUC score

-

Releases

No releases published
[Create a new release](#)

Packages

No packages published
[Publish your first package](#)

Languages

- Jupyter Notebook 100.0%