

Phase_3_Project

Public

🔗

🔗 1 Branch

🏷 0 Tags

🔗

📁

🔍 Go to file

t

Go to file

📄 About

Add file ▾

Code

⚙

marymusyoka

added pdf presentation

1e06e6b · 2 minutes ago

🕒

📁 Data	added gitignore file	last week
📄 .gitignore	added gitignore file	last week
📄 Presentation....	added pdf presenta...	2 minutes ago
📄 README.md	added files to read...	15 hours ago
📄 index.ipynb	additional changes	13 hours ago

No description, website, or topics provided.

📖

Readme

📈

Activity

★

0 stars

👁

1 watching

🔗

0 forks

Releases

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

Packages

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

Languages

● Jupyter Notebook 100.0%

📖 README

✎

☰

Predicting Customer Churn in Telecommunications: A Data-Driven Approach to Enhancing Retention Strategies at SyriaTel

Overview

This project aims to predict customer churn for SyriaTel, a telecommunications company, using a binary classification model. The goal is to identify patterns and indicators that can predict whether a customer will stop doing business with SyriaTel. This can help the company develop targeted retention strategies.

Dataset

The dataset used for this project consists of 3333 entries and 21 columns, including various features related to customer account information and call metrics. Key features include:

state: Customer's state.

account_length: Length of the account.

area_code: Area code of the customer.

phone_number: Customer's phone number.

international_plan: Whether the customer has an international plan.

voice_mail_plan: Whether the customer has a voice mail plan.

number_vmail_messages: Number of voice mail messages.

Various call metrics such as total_day_minutes, total_eve_minutes, and total_night_minutes.

churn: Target variable indicating whether the customer has churned (boolean).

Methodology

The project involves the following steps:

Data Preprocessing: Cleaning and preparing the dataset for analysis.

Exploratory Data Analysis (EDA): Analyzing the data to understand key patterns and relationships.

Model Training: Training a classification model to predict churn using features from the dataset.

Model Evaluation: Assessing model performance using metrics such as accuracy.

Hyperparameter Tuning: Using GridSearchCV to find the optimal hyperparameters for the Decision Tree classifier.

Features

Data Cleaning: Handling missing values and encoding categorical variables.

Exploratory Data Analysis: Visualizations and statistical summaries to understand the data.

Modeling: Decision Tree classifier with hyperparameter tuning using GridSearchCV.

Performance Metrics:

Accuracy, precision, recall, and F1-score.

Project Objectives

Predict Customer Churn: Develop a model to accurately predict whether a customer will churn based on their features and call metrics.

Feature Analysis: Identify key factors that influence customer churn.

Model Evaluation: Evaluate model performance using various metrics and refine the model for improved accuracy.

Tools and Libraries

Python: Programming language used for data analysis and modeling.

Pandas: Data manipulation and analysis.

Scikit-learn: Machine learning library for building and evaluating models.

Matplotlib/Seaborn: Data visualization.

Workflow

Data Exploration and Preprocessing

Load Data: Import and explore the dataset using Pandas.

Clean Data: Handle missing values, encode categorical variables, and normalize features.

Feature Engineering: Create or modify features to improve model performance.

Model Building

Train-Test Split: Divide the dataset into training and testing sets.

Model Selection: Experiment with various classification models:

Logistic Regression: A baseline model for classification.

Decision Tree: A decision tree classifier optimized using GridSearchCV.

RandomForest: A random forest classifier.

Hyperparameter Tuning: Optimize model parameters to enhance performance.

Model Evaluation

Evaluation Metrics:

Accuracy: Overall correctness of the model.

Precision: Ratio of true positives to the sum of true and false positives.

Recall: Ratio of true positives to the sum of true positives and false negatives.

F1 Score: Harmonic mean of precision and recall.

Feature Importance: Analyze which features most influence churn predictions.

Results and Insights

Model Performance: Summary of the best-performing model and its parameters.

Feature Importance: Ranking of features based on their impact on churn prediction.

Recommendations:

Strategies for improving customer retention based on insights from the model.