

main1 Branch0 Tags

Go to file

Add file

Code

libertron Notebook4819c86 · 32 minutes ago2 Commits		
Data	Notebook	32 minutes ago
Images	Notebook	32 minutes ago
PDFs	Notebook	32 minutes ago
.gitignore	Initial commit	yesterday
LICENSE	Initial commit	yesterday
Main.ipynb	Notebook	32 minutes ago
README.md	Notebook	32 minutes ago

README

GPL-3.0 license

Churn in Syria Telecommunication Company

Overview

We are going to perform a predictive analysis for a telecommunication company in order to provide solid insights for potential future churn.

Specifically, this will cover:

- Performing a train-test split to evaluate model performance on unseen data
- Applying appropriate preprocessing steps to training and test data
- Identifying overfitting and underfitting

Business Understanding

The primary goal of this project is to help the company to predict potential churn through the implemented model and perhaps take serious actions to prevent them.

Data Understanding

I will be using the Churn in Telecom's dataset, modeling the `churn` based on all other numeric features of the dataset. ([dataset here](#))

Data Preparation

Before analysis, the dataset underwent the following preprocessing steps: Cleaning: Removed duplicate entries, fixed inconsistent genre labeling, and corrected invalid values.

Data Analysis

For the current matter I will build **two models** :

- A Logistic regression as the baseline model.
- A Decision tree model as the second one, more complexe and finally tune it for more improvement.

For this reason I will first perform a **train-test split**, so that I am fitting the model using the training dataset and evaluating the model using the testing dataset.

Requirements

- Perform a Train-Test Split
- Fit a Logistic regression Model
- Fit a Decision tree Model
- Fit a Decision tree Tuned Model (Improve the previous model)
- Compare the models
- Determine feature importanceestablished universes.

Visualization

We utilized a variety of visual tools to present our findings clearly:

Confusion matrix for a better obeservation of TP FP FN TN Bar Chat to show best model Bar plot to display feature importance

Recommendations

After training and analysis we come out with some insights that could help predict and potentially reduce churn in the company business.

- Best Model Recommendation would be Tuned Decision tree since it has the highest ROC AUC score compared to Logistic Regression.
- Customers who spend more time per day tend to churn the most, that may due to the service charging fees. Therefore I would recommend to propose affordable service bundle.
- Another factor which is related to churn is customer service calls. Either the customer is not satisfy after calling, or he cannot find people to take his case in consideration. Then I would recommend to improve customer service experience.

Code Quality

All code was written in Python using industry-standard libraries such as Pandas, Matplotlib, and skeatlearn . Key characteristics of the codebase include:

- Modularity: Functions are separated for loading, cleaning, analyzing, and visualizing data to promote reusability.
- Documentation: Inline comments and docstrings are used to clarify purpose and logic.
- Efficiency: Vectorized operations and optimized queries reduce runtime for large datasets.
- Reproducibility: Code is organized in a Jupyter Notebook, allowing anyone to rerun the analysis with minimal setup.
- Version Control: Git was used to track changes and manage collaboration.

Repo Structure

Data/churn_dataset.csv

Partials/*.py [some modules]

PDFs/Github.pdf

PDFs/Notebook.pdf

PDFs/Presentation.pdf

.gitignore

LICENSE

Main.ipynb

README.md

About

Churn in Telecommunication Dataset

Readme

GPL-3.0 license

Activity

0 stars

0 watching

0 forks

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Languages

Jupyter Notebook 100.0%

