makau99 / LoanDefaultAnalysis

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

LoanDefaultAnalysis Public

1 Branch 0 Tags Go to file t Go to file Add file Code

makau99 Rename clean.xls to data/clean.xls 534978d · 4 minutes ago

.ipynb_checkpoints	limitations	19 minutes ago
data	Rename clean.xls to data...	4 minutes ago
default+of+credit...	Developing outline & dat...	5 days ago
README.md	Update README.md	9 minutes ago
Untitled.ipynb	eda	5 days ago
creditdefault.ipynb	limitations	19 minutes ago

README

LoanDefaultAnalysis

About

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

Credit Risk Analysis

Overview

This project analyzes credit risk using machine learning models to predict the likelihood of default. We leverage a dataset containing customer credit history and financial behavior to build a predictive model.

Business and Data Understanding

Stakeholder Audience

Our primary stakeholders include:

- **Financial Institutions:** Banks and credit issuers looking to improve credit risk assessment.
- **Risk Analysts:** Professionals who analyze credit data to minimize defaults.
- **Regulatory Bodies:** Organizations that monitor fair lending practices.

Dataset Choice

The dataset contains information on credit balances, payment history, and demographic details. It is structured with features such as:

- **Credit Limit**
- **Payment Delays**
- **Total Bill Amount**
- **Default Status**
- **Demographic Features (e.g., Gender, Age)**

Modeling

We implemented and evaluated multiple machine learning models:

-Logistic Regression

-K-Nearest Neighbors (KNN)

-XGBoost Classifier

-LightGBM Classifier

-Random Forest Classifier

-Multi-Layer Perceptron (MLP) Classifier

The best-performing model was LightGBM, which effectively handled categorical data without encoding. The model had a 73% accuracy and 77% ROC AUC score. It could recall defaulters 67% of the time.

Evaluation

Performance Metrics

We evaluated the model on the test dataset using:

- **Confusion Matrix**
- **Classification Report (Precision, Recall, F1-score)**
- **Accuracy Score**

Conclusion

- Higher credit limits do not always correlate with higher default risk.
- Payment delays are key indicators of default.
- The model provides a reasonable prediction of credit risk, aiding financial institutions in decision-making.