MIT license

0 stars 0 forks 0 watching Branches Activity
Tags

Public repository

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

Kanyenje-Brian Update README.md 605a589 · now

.ipynb_checkpoints	final	4 minutes ago
Data	models	13 hours ago
LICENSE	Initial commit	13 hours ago
Presentation.pdf	final	39 minutes ago
README.md	Update README.md	now
notebook.ipynb	final	4 minutes ago
requirements.txt	final	39 minutes ago

README MIT license

Loan Approval Classification

This project builds a machine learning model to predict whether a loan applicant is likely to be approved or rejected based on demographic, financial, and behavioral attributes. It supports lenders in minimizing loan defaults and improving credit risk assessments.

Business Problem

Lending institutions face increasing pressure to make accurate, data-driven decisions in evaluating loan applications. Misclassifying high-risk applicants can lead to defaults, while rejecting creditworthy applicants impacts customer satisfaction and business growth.

Goal: Use data science to distinguish between high-risk and creditworthy applicants and support better lending decisions.

🎯 Objectives

1. Analyze demographic factors associated with high loan default rates.
2. Test the relationship between credit score and loan status.
3. Build and validate machine learning models for accurate loan prediction.
4. Recommend the best-performing model based on precision, recall, F1-score, and AUC.

📊 Dataset Overview

- **Source**: Kaggle: Loan Approval Classification Data
- **Records**: 45,000
- **Target**: `loan_status` (0 = Rejected, 1 = Approved)

Features include:

- Age, gender, education, income
- Employment experience
- Loan amount, interest rate
- Credit score and credit history
- Previous loan default indicator

🔍 Exploratory Data Analysis (EDA)

- KDE and count plots for distributions and class comparisons
- Multicollinearity analysis using correlation heatmaps and t-tests
- Chi-square tests and statistical analysis for relationships between:
  - Gender and default
  - Age and default
  - Education level and approval

🧹 Data Preprocessing

- No missing or duplicate values
- Scaled numeric features using MinMaxScaler
- Ordinal encoding for education
- One-hot encoding for gender, home ownership, loan intent, previous defaults
- Feature selection with **RFECV**
- Class imbalance handled using **SMOTE** and **class weighting**

Models Developed

Model	Key Notes
Logistic Regression	Baseline model, fine-tuned with L1 penalty

Model	Key Notes
Decision Tree	Captures non-linear patterns and feature interactions
XGBoost	(Planned) Advanced boosting model for performance

📈 Model Evaluation (Best Version)

Metric	Logistic Regression	Decision Tree
Accuracy	89%	92%
F1 (Approved)	76%	79%
AUC Score	0.95	0.96

- Evaluation metrics: Accuracy, F1-score, Precision, Recall, AUC
- Performance validated on **unseen test data**

📁 Folder Structure

├── Data/ # Compressed dataset (ZIP) ├── Presentation.pdf # Slide presentation with findings ├── notebook.ipynb # Complete workflow in Jupyter ├── requirements.txt # Python dependencies ├── LICENSE # MIT License └── README.md # Project overview and documentation

🛠️ Setup Instructions

Follow these steps to set up the project on your local machine:

```
 1. Clone the repository
git clone https://github.com/your-username/loan-approval-classification.git

# 2. Navigate to the project directory
cd loan-approval-classification

# 3. (Optional) Create and activate a virtual environment
python -m venv venv
source venv/bin/activate        # On Windows: venv\Scripts\activate

# 4. Install the required packages
pip install -r requirements.txt
```

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

- **Jupyter Notebook** 100.0%