This is a collaborative project classifying loan applications of small businesses as either approved or not approved.

Credit Alchemy: Small Business Creditworthiness Demystified

Predicting Credit-Worthiness of Small Businesses

This task is aimed at building a loan approval classifier that predicts whether or not a loan application for a small business will be approved, based on the business' financial and credit history. After evaluating our different models, both stand-alones and ensemble, we found that XGBoost predicts the target variable

Case study of businesses supported by the U.S. Small Business Administration(SBA) found here.

Project Overview

This project focuses on utilizing the dataset from the U.S. Small Business Administration to develop a predictive model for loan application approval. By analyzing relevant factors and historical loan data, the project aims to create a reliable system that assists the SBA in making informed decisions while minimizing the risk of defaults. The ultimate goal is to provide the SBA with a robust loan approval model that enhances their ability to make accurate and efficient lending decisions.

Stakeholders: Small Businesses, Government, and Lending institutions such as the SBA

Project Objectives

1. Conduct a comprehensive analysis of the dataset from the U.S. Small Business Administration to identify patterns and trends for accurate loan approval predictions.
2. Develop a robust machine learning model that utilizes the identified patterns and trends to predict loan approval outcomes effectively.
3. Deploy and evaluate the developed machine learning model in the loan approval process of the U.S. Small Business Administration, continuously optimizing its predictive capabilities for informed lending decisions.

Success Metrics

We settled on Accuracy as our success metric and concluded the following that where:

The Accuracy score is at least 85 % or higher the model would classify/filter the loan applications as either Approved or Rejected.

The accuracy score is preferred in a loan classification/filtering model as it provides a straightforward measure of how well the model correctly predicts loan approvals or rejections, allowing for a clear assessment of its overall performance and effectiveness in minimizing false positives and false negatives, which is crucial in this case.

Data Understanding

The data science process that was utilised in this project was the CRISP-DM process.

The data was sourced from the official SBA Open Data source. It had 899,164 rows and 27 columns.

Data Preparation included the following:

- Handling of missing values.

- Changing of data types to appropriate format.
- Appropriately dropping missing values.
- Feature Engineering.
- Pre-processing of the data

Modeling and Evaluation

The models that were used in this project included the following:

1. Logistic Regression - This was the baseline model
2. Decision Trees
3. Random Forest
4. Support Vector Machine
5. XGBoost
6. Neural Networks

As the performance of the Baseline model did not meet our metrics of success, we proceeded to attempt other models to see if it will improve. The best performing model was XGBoost, as it has a test accuracy of upto **91%** . We therefore used it as our model for deploying the app.

Deployment

The next step in the project was to create a web application that would help us to meet our project objectives effectively. This loan filtering app was deployed on Streamlit.

To deploy the app, you will need to have Streamlit installed.

You can install Streamlit by running the following commands in your terminal:

Install Streamlit:

```
pip install streamlit
```

The Streamlit documentation is provided for your convenience, in case of any errors.

To run the app locally run the following on the terminal:

```
streamlit run sba.py
```

The input features were: `Gross Approval` , `Industry` , `Loan Term` , `Number of Employees` , `NewExist` , and `UrbanRural` to predict loan approval.

Upon doing a feature importance on the pre-trained model, we found out that the **Loan Term** and the **Gross Approval amount** are the two key features in determining whether a loan would pass the initial filtering stage.

Recommendations

Based on the analysis that we have done, we came up with five main areas into which the SBA can dive deeper, these areas included:

1. Loan Term
2. Size of the Business
3. Location
4. Business Existence Period
5. Industry

Conclusion

- With the power of data, SBA can now make accurate predictions and conduct an initial filtration without human intervention. Clients can also get a feel of the major factors that would deter them from getting a loan and adjust accordingly by utilizing some of the existing services that are provided by the SBA.
- This would be helpful also in addressing the issue of unemployment as it would further help encourage more entrepreneurs and small businesses to seek support in order to leverage their businesses; which will also increase the employment rate as a domino effect.

Navigating the Repository

- To access this repository locally, you will need to clone it by following these steps:

```
  git clone
https://github.com/LynneMutwiri/Credit_Worthiness_of_Small_Businesses.git
  cd Credit_Worthiness_of_Small_Businesses
```

- Under the Deliverables folder, you will find the following files stored in pdf format:

```
Presentation
Write-up
Jupyter Notebook
GitHub
```

- To review the full analysis kindly refer to the Jupyter Notebook as well as the Presentation

Project Collaborators

1. Brian Njau
2. Grace Nyongesa
3. Josiah Okumu
4. Leo Kariuki
5. Lynne Mutwiri
6. Stella Kitur

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Contributors 4


LynneMutwiri


stellacherotich Stella Cherotich


Brian-Njau


leokariuki leokariuki

Languages

- **HTML** 75.8%
- **Jupyter Notebook** 24.1%
- **Python** 0.1%

Suggested Workflows

Based on your tech stack


Actions Importer

Automatically convert CI/CD files to YAML for GitHub Actions.

Set up


SLSA Generic generator

Generate SLSA3 provenance for your existing release workflows

Configure

Jekyll using Docker image

Configure

Package a Jekyll site using the jekyll/builder Docker image.

More workflows

Dismiss suggestions