



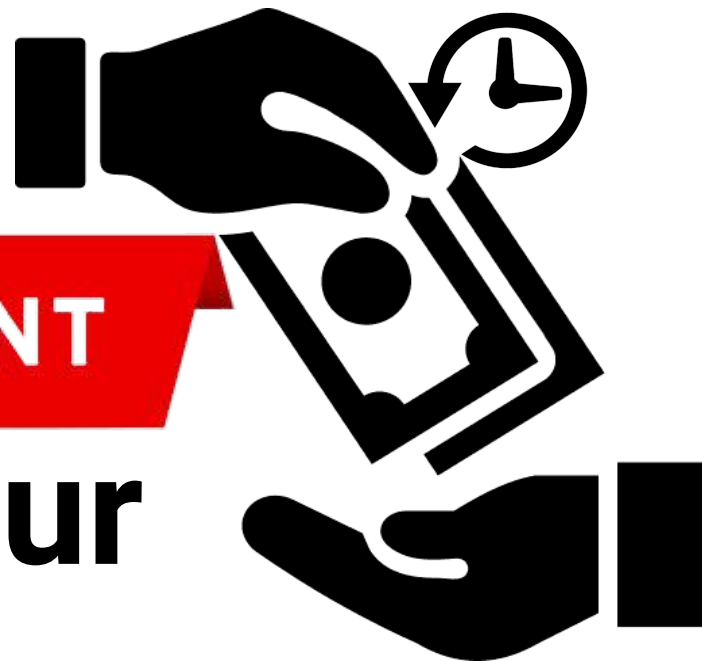
kaggle


KaggleX - Project Showcase

Predicting LOAN

REPAYMENT

Behaviour





Predicting loan repayment behavior of Nigerian banks' borrowers using Machine Learning Algorithms

Victoria Okesipe

BIO

Victoria Okesipe is a climate and data science enthusiast who believes in working hard towards achieving her goals. She is eager to gain industry-based work experience and challenge herself internationally. Her mantra is “Begin with the end in mind”.



Project Definition

- Project Summary
- Data Science Topic(s) Applied
- Lessons learnt

Project Description

Project Description

The goal of this project is to develop a machine learning model that accurately predicts whether a borrower will repay a loan on time or not. **The project will involve collecting data on borrowers' credit history, income, debt-to-income ratio, loan amount, loan term, and other relevant variables.** The data can be obtained from publicly available sources, kindly click [here](#).

Project Summary

Project Summary

The project involves preparing loan data for machine learning by cleaning, preprocessing, and feature engineering. Multiple machine learning algorithms will be evaluated, including logistic regression, decision trees, random forests, and gradient boosting, with the best performing model selected for hyperparameter tuning. The final model's performance will be evaluated on a test set using metrics such as accuracy, precision, recall, and F1 score, and a confusion matrix will be created to visualize its performance. The model will be deployed locally using Streamlit for easy use and accessibility.

Data Science Topic(s) Applied

Specific Data Science Topics Applied

- **Data Cleaning:** handling missing values, dealing with outliers, correcting data types, etc.
- **Data Preprocessing:** feature scaling, normalization, encoding categorical variables, handling imbalanced data, etc.
- **Feature Engineering:** creating new features from existing variables, selecting relevant features, etc.
- **Supervised Learning:** using labeled data to train a model for classification or regression tasks.

Specific Data Science Topics Applied

- **Model Selection:** evaluating different models and selecting the best performing one based on certain criteria.
- **Hyperparameter Tuning:** optimizing the model's hyperparameters to improve its performance.
- **Model Evaluation:** assessing the model's performance on an unseen test dataset using metrics such as accuracy, precision, recall, and F1 score.
- **Model Deployment:** deploying the final model locally using Streamlit, so that users can interact with it and get predictions.

Lessons Learnt

Working on this project provided several valuable lessons, including:

- The importance of data cleaning, feature engineering, model selection, hyperparameter tuning, model evaluation, and deployment.
- It highlighted the significance of good collaboration and communication skills and emphasized the value of documentation for reproducibility and sharing knowledge.
- .
- EDA was shown to be a valuable tool in providing insights into the data.
- Finally, working on this project demonstrated that careful consideration of various factors, such as data quality, model complexity, and scalability, is essential for successfully deploying a machine learning model.

Project Outcome

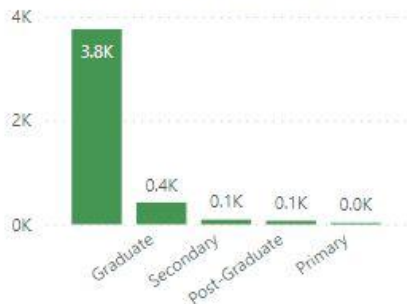


Loan
Amount by
Year

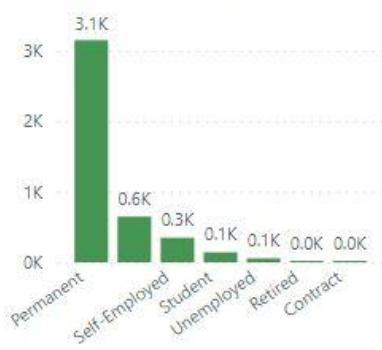


LOAN Dashboard Full Insight

Level of Education



Employment Status



Total Loan Applicants

4,368

Bank Name



Good Loan



Bad Loan



15 Days Term

919
21.0%

30 Days Term

3,129
71.6%

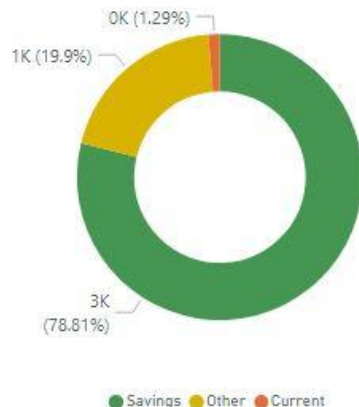
60 Days Term

288
6.6%

90 Days Term

32
0.7%

Bank Account Type



Project Outcome

The outcome of this project is a machine learning model that can be used by lending institutions to predict whether a borrower will **repay** a **loan** on **time** or **not**. This can help lenders make informed lending decisions and reduce the risk of default. Kindly find the link to the Python Notebook [here](#).



THANK YOU