BANK LOAN APPROVAL CLASSIFICATION PROJECT

NAME: LILIAN KABURO

GITHUB LINK: HTTPS://GITHUB.COM/LILIANKABURO/LOAN-APPROVAL-PROJECT/

INTRODUCTION

The Bank Loan Approval Classification Project is designed to develop a machine learning model that predicts whether a loan application should be approved or rejected based on a set of features related to the applicant's financial history, personal information, and other relevant criteria. The objective is to build a classification model that can assist bank decision-makers by providing quick, objective, and accurate recommendations for loan approval.

BUSINESS PROBLEM

- What is the main challenge of this project?---How can we create a predictive model to classify loan applications as either approved or denied?
- What key factors should the model take into account when deciding on loan approvals?---How can we include various factors such as income, credit score, loan amount, and employment status into the model to mirror the bank's decision-making process?
- What data do we have available for training the model?---What features of the loan applicants (e.g., age, education level, credit history) are present in the dataset, and how do they correlate with loan approval status (approved or denied)?
- How can machine learning techniques be applied to address this issue?---How can we utilize machine learning methods to develop a model that accurately predicts whether a new loan application will be approved or denied?
- What is the main objective of this project?---How can we automate the loan approval decision-making process by developing a predictive model, and what level of accuracy is necessary for practical implementation?

DATA SOURCE AND UNDERSTANDING

Data Source

The Dataset we are using is obtained from https://www.kaggle.com/datasets/.

Data Understanding>>>>> data understanding and data exploration using python pandas and other libraries

DATA PRE-PROCESSING

- Handling categorical Variables
- Handling high correlation
- Feature scaling
- Data splitting

MODELS BUILT

- Logistic Regression model
- decision Tree
- RandomForestClassifier
- KNearestNeighbours
- GradientBoostingClassifier Model

CONCLUSION

By leveraging machine learning models for loan approval classification, financial institutions can improve the speed, accuracy, and fairness of their loan decision-making process. The chosen model will provide a reliable and scalable way to assess loan applications, ultimately benefiting both the bank and its customers by enhancing decision-making efficiency and reducing risks associated with loan defaults.

RECOMMENDATIONS

1. Objective: Prioritize Loan Approvals

Random Forest strikes -- Ideal for businesses aiming to prioritize loan approvals without overly compromising on quality.

2. Objective: Minimize Financial Risk (Reduce False Approvals)

Gradient BoostingClassifier---Ideal for businesses that want to ensure a higher level of certainty in loan approvals, even if it means missing some approvals.

3. Objective: Simplicity and Ease of Use

KNN is easy to understand and doesn't require complex tuning or heavy computational resources.. Ideal for businesses with resource constraints and a need for simplicity, but less ideal for businesses where accuracy and minimizing missed opportunities are critical.

REFERENCES

- W3schools https://www.w3schools.com/
- Datacamp https://www.datacamp.com/
- Moringa School Instructure Canvas(you have to be a student at Moringa to access this)
- YouTube tutorials https://www.youtube.com/watch?v=XmSlFPDjKdc&t=905s
- Github Repositories https://github.com/
- kaggle datasets and kaggle code https://www.kaggle.com/
- Python Documentations https://docs.python.org/3/
- Chatgpt 4.0 https://chatgpt.com/ (when all else fail get this tool to debug)