**Loan Default Prediction**

This project aims to develop a machine learning model to predict the likelihood of a borrower defaulting on a loan, which is crucial for banks and financial institutions as it enables proactive risk management, reduces financial losses, and improves lending policies. The target audience includes banks and financial institutions, credit risk analysts, regulatory bodies, and fintech companies. The model can reduce bad debt and loan write-offs, improve lending policies by identifying high-risk applicants, and enhance customer profiling, leading to personalized lending solutions.

In Kenya's evolving financial sector, both lenders and borrowers have considerable obstacles. Challenges encountered by lenders include elevated default rates, political and economic volatility, and corruption.   
Borrowers face challenges include limited access to credit, high and variable interest rates, rigorous loan qualification criteria, and cultural and societal barriers, particularly impacting women.   
  
Loan default constitutes a substantial challenge within Kenya's financial sector, impacting the profitability and sustainability of lending institutions. The execution of governmental initiatives like the Hustler Fund underscores the necessity of addressing this issue. Launched in late 2022, the Hustler Fund aimed to provide accessible credit to Kenyan residents, offering loans at an annual interest rate of 8%. As of August 2023, more than 29% of the outstanding loan portfolio was classified as at-risk, with nearly 3 billion shillings in defaults, indicating a default rate nearly twice that of commercial banks.  
  
The dataset consists of historical loan application records, credit scores, and transaction history from multiple sources. Key features include demographic information, loan details, credit history, and transactional behavior. The model will use Logistic Regression, Decision Trees, Random Forest, XGBoost, and Deep Learning techniques. The target variable is Loan\_Approval\_Status (Default/Non-Default).  
  
The project will evaluate the model's accuracy, precision, recall, F1-Score, and ROC-AUC Score. The minimum viable product (MVP) involves a working Logistic Regression model with at least 70% accuracy. The project will also develop a web application for real-time loan approval assessment and use Python libraries for data handling and modeling. The analysis will be performed locally during development and cloud-based environments for scaling.  
  
In conclusion, this project aims to leverage machine learning to enhance loan approval and risk assessment processes, reducing financial losses and improving lending efficiency. The final product will be an end-to-end deployable application with real-world utility.