**Problem Statement: Loan Default Prediction in Kenya**

Loan default occurs when a borrower fails to meet repayment obligations, either by missing multiple payments or stopping entirely. It can happen with various loan types, including personal, mortgage, auto, student, or business loans. The stages of loan default include missed payments, delinquency (leading to late fees and credit bureau reports), default (typically after 30–90+ days of non-payment), and legal consequences such as asset seizure or lawsuits. The effects of loan default are severe, including credit score damage, higher interest rates on future loans, legal action, loss of collateral, and aggressive debt collection efforts.

In Kenya, loan defaults have been a growing concern, particularly in government-backed lending initiatives like the Hustler Fund. The State Department for Micro, Small, and Medium Enterprises (MSMEs) reported that default rates for Hustler Fund loans now surpass 50%, with outstanding repayments totaling KSh11 billion ($85 million) (Clynch, 2024). Additionally, the number of loan defaulters in Kenya has nearly doubled over the past year, with 3.85 million more borrowers being blacklisted by lenders, primarily digital lending platforms and mobile loan apps. Data from TransUnion Africa shows that the total number of negatively listed Kenyans reached 7.74 million in Q1 2024, a 99.2% increase from Q1 2023 (Kamau, 2024). Among those affected by high default rates are borrowers from Ruto’s flagship initiative, the Hustler Fund, which provides loans to individuals and small businesses. More than 13 million borrowers have defaulted on a total of Kes. 7 billion. (Kamau, 2024).

Several studies have attempted to model loan default prediction in Kenya, yet challenges persist. A 2016 study applied the Generalized Extreme Value Regression Model to address rare event scenarios in Kenyan banks (Wanjohi, 2016). A 2018 study used logistic regression to analyze factors influencing student loan defaults (Pauline Nyathira Kamau, 2018). More recently, a 2022 study developed a credit risk assessment model using machine learning techniques, achieving 93% accuracy by incorporating borrower details such as age, loan amount, marital status, and sex (Mutembete, 2022). Despite these advancements, accurately predicting loan defaults remains a challenge, necessitating further research and model refinement.

Our loan default prediction model aims to analyze a large dataset to estimate the probability of a borrower defaulting based on credit history and financial behavior. By examining past loan performances, lenders can make informed credit decisions and reduce default risks. This model is crucial for both banks and customers. For banks, it aids in risk management, credit decisioning, optimized interest rates, regulatory compliance, fraud detection, operational efficiency, and portfolio diversification. For customers, it ensures fair loan approvals, better loan terms, financial awareness, faster loan processing, and personalized loan offers. In summary, this model will help financial institutions optimize lending policies and mitigate credit risk while ensuring responsible lending and borrowing practices.