# LOAN APPROVAL MODEL

- NAME : KELVIN MWANGI KIBUNG'A(MODEL DEVELOPER)

# INTRODUCTION

- The financial field has to be handled with a lot of sensitivity especially when lending money to customers . The primary goal of this loan approval model project is to automate and improve the decision-making process for approving or rejecting loan applications. This will be done by building a machine learning model that predicts the likelihood of a loan applicant defaulting on a loan based on historical data.

# PROBLEM STATEMENT/BUSINESS QUESTION

- Loan approval decisions are traditionally made using manual processes, which are time-consuming and subject to human biases. Furthermore, lenders may struggle with accurately predicting which applicants will repay their loans. The problem is to design a model that uses historical data to identify patterns and predict whether a borrower is likely to default on a loan.

# DATA OVERVIEW

- Source :The data is from Kaggle datasets
- Features : Some of the features include past loan defaulting history , loan intent and income level .
- Data preparation for modelling : Dealing with outliers, dealing with class imbalance problem.

# METHODOLOGY

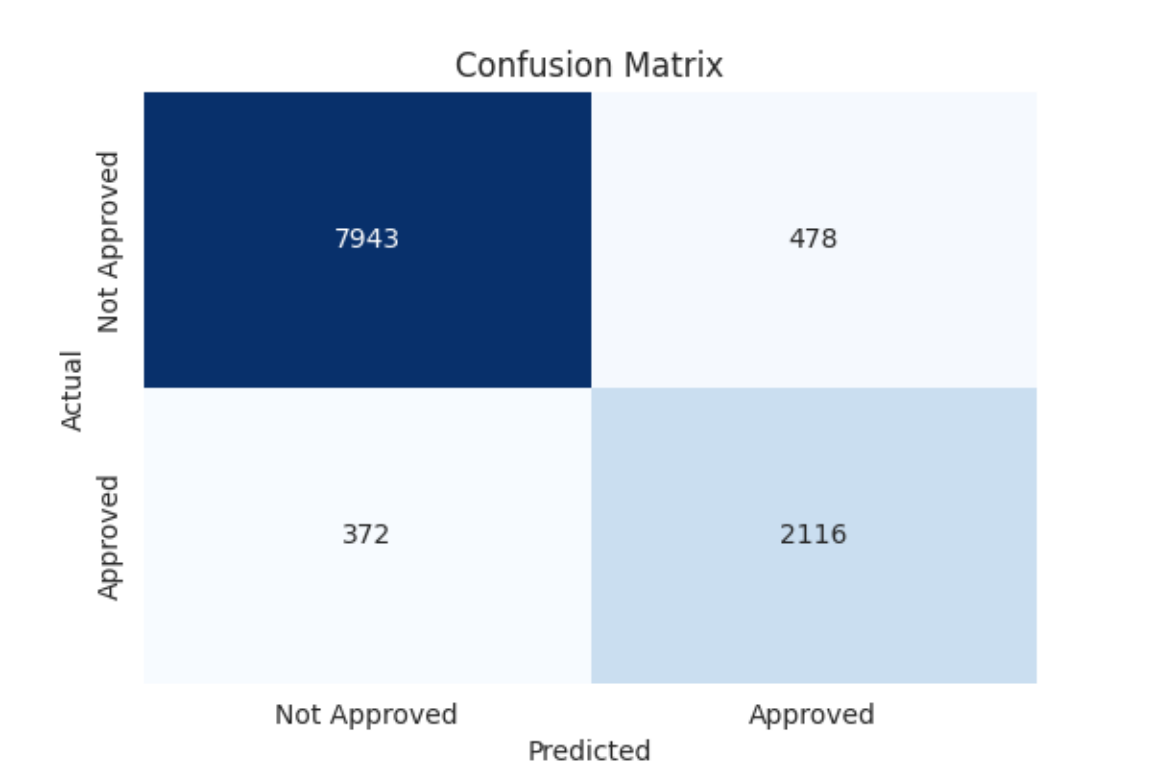
- **Model Selection**: We used Sequential model from keras as the baseline model , later implemented RandomForest and XGBoost model.

- **Training & Testing**: We used 75% of the data for training and 25% for testing the model.

- **Evaluation metrics : The metrics use for evaluation are accuracy, F1SCore , recall and precision.**

# MODEL PERFORMANCE

The models performed well and below is a confusion matrix from the xgboost model.


Confusion Matrix
Actual
Not Approved
Approved
7943
478
372
2116
Not Approved
Approved
Predicted

# INSIGHTS

- All the models have a high accuracy(89) and above. However our dataset being very imbalanced, we cannot assume our models have a good performance from the accuracy alone.
- Conveniently the precision ,recall and Fbetascore which are the guiding metrics of a model from imbalanced data,are very good too.
- The XGBClassifier and randomforest models have very good performances of the three metrics.

# RECOMMENDATIONS

- **1. Proactive Default Prevention Strategies**
- **2. Approve loans based on risk profiles**
- **3. Adjustment of interest rates and securities for high risk individuals**
- **4. Personalized Loan Products**