Financial Inclusion in Africa

By: Ankamah Ellis Boateng

Executive Summary

Objective

To create a machine learning model to predict which individuals are most likely to have or use a bank account to access the financial inclusion in Kenya, Rwanda, Tanzania, and Uganda

Key Findings

Strategic investments into education system, mobile banking and gender equality will increase the financial index

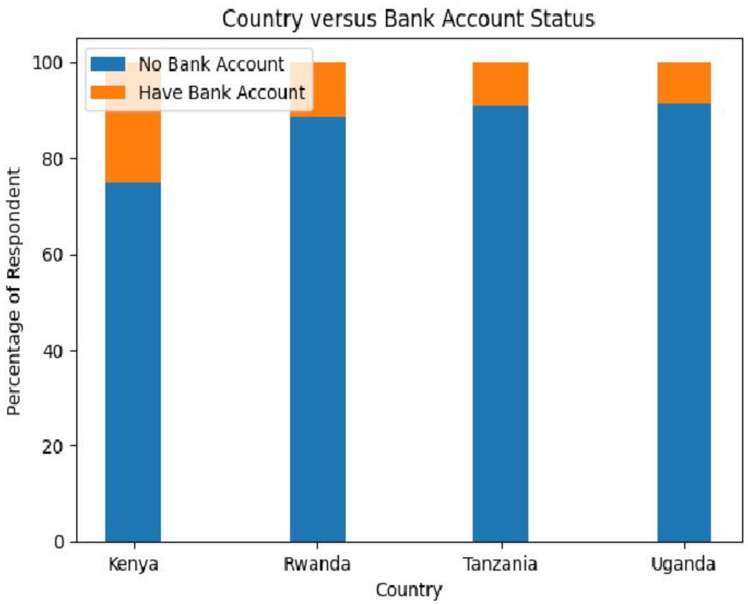
Data Overview

Source
Zindi: https://zindi.africa/competitions/financial-inclusion-in-africa/data

Description
The dataset contains response from 23524 interviewees about their personal information, demographics and whether they have a bank account or not

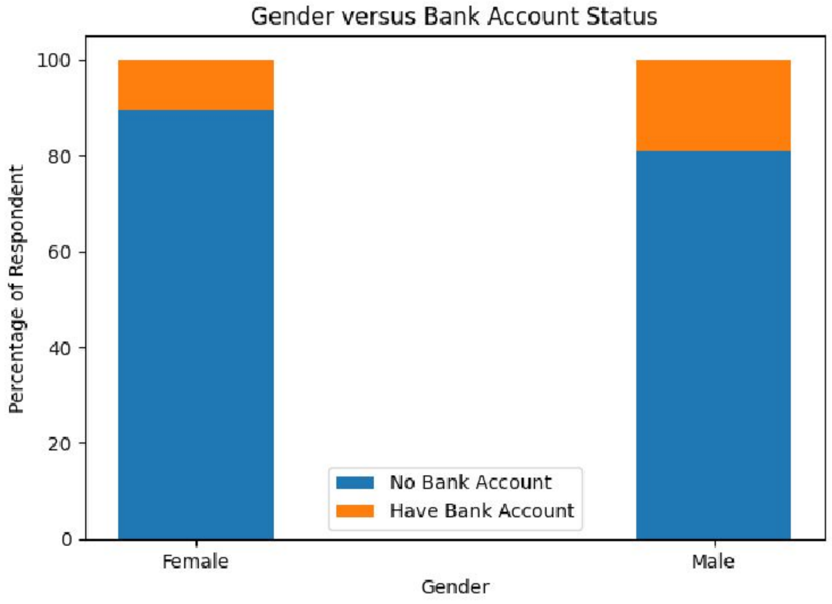
Processing
There were no missing values. Nonetheless, bank account column was encoded to binary (1-yes, 0-no). The unique id column was dropped because it was not relevant for model training

Exploratory Data Analysis



This chart provides an idea of the financial inclusion index in the observed countries with Kenya having the highest and Tanzania the least.

Exploratory Data Analysis Cont'd

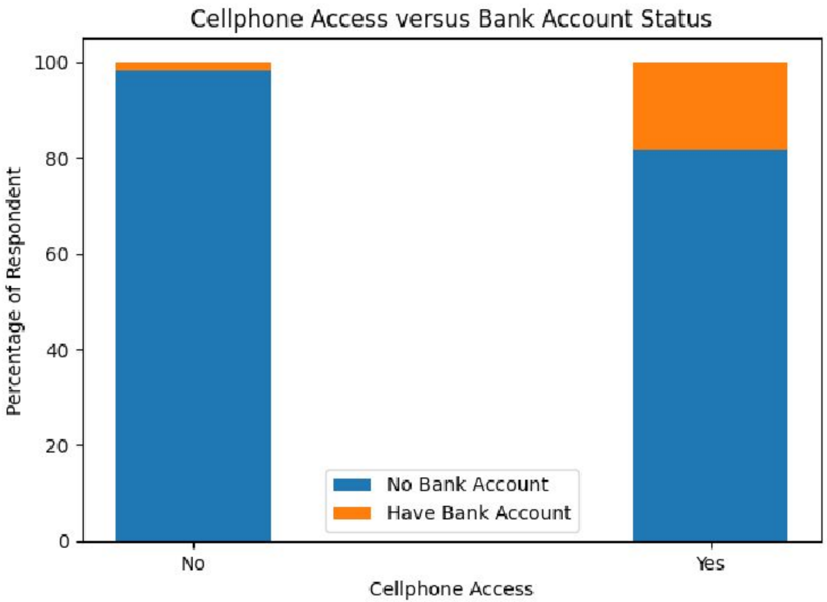


From our dataset, there are more females(59%) than males(41%)

Nonetheless, the percentage of females with bank accounts are lesser than males

Further suggesting some gender bias in the financial inclusion index

Exploratory Data Analysis Cont'd

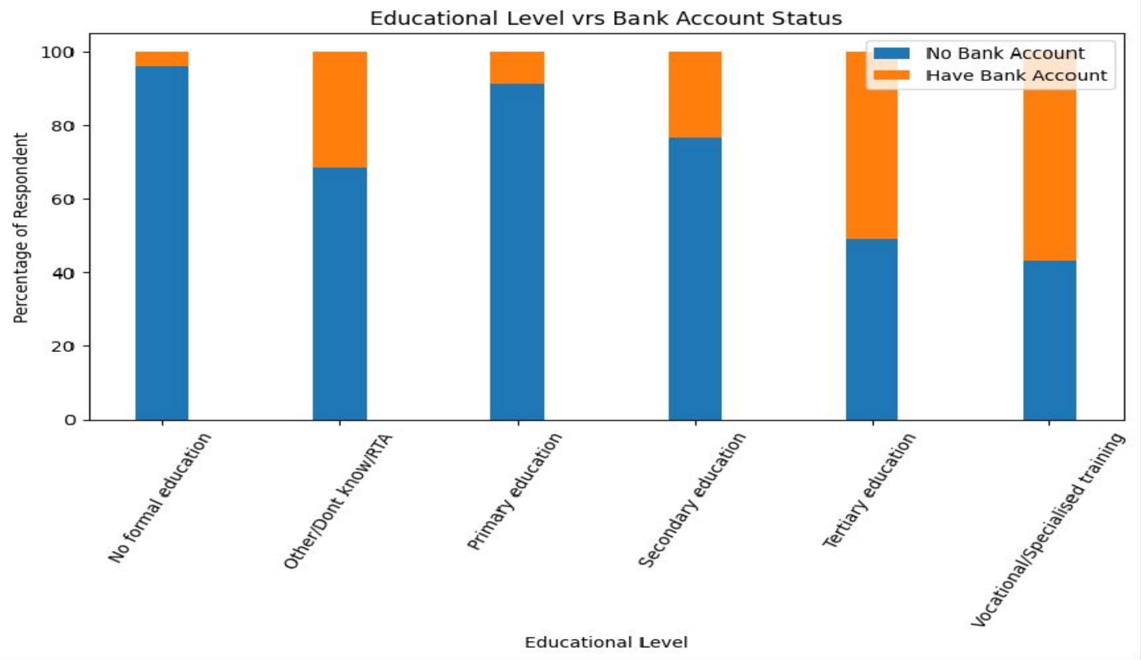


Almost all respondents without cellphones, about 98%, have no bank account

Just about 20% percent of those with cellphones have bank accounts

This raises questions about the mobile banking in our observed countries

Exploratory Data Analysis Cont'd



It is worth noting from the data that, as education level increases, the percentage of respondents with bank accounts increases

Methodology

- Perform machine learning analysis on the dataset using python
- Split the dataset into training set(70%) and validation set(30%)
- Use DictVectorizer to perform the one-hot encoding of categorical variables
- Classification algorithms used: Logistic Regression, Random Forest Classifier, Decision Tree Classifier, Adaboost Classifier, Gradient Boosting Classifier, and XGBoost Classifier
- Evaluation Metrics: accuracy score and F1 score

Results

Model	Accuracy Score	F1 Score	Recall Score	Precision Score
Logistic Regression	88.54	0.44	0.32	0.67
Random Forest Classifier	86.03	0.42	0.37	0.49
Decision Tree Classifier	82.57	0.36	0.36	0.36
Ada Boost Classifier	88.2	0.4	0.29	0.66
Gradient Boosting Classifier	88.14	0.46	0.37	0.61
XG Boost Classifier	88.1	0.46	0.37	0.61

Logistic Regression Model was selected because it has the highest accuracy score of 88.54 and comparatively good F1 score of 0.44

Conclusion

Model

Model accuracy of about 89% is good for making predictions

Recommendation

Promoting and improving the quality of the education system can increase financial inclusion

Investments into mobile banking and infrastructure

Encourage females to take advantage of the banking system to reduce the bias in the financial inclusion index

Thank you.