Financial inclusion in East Africa

Introduction

Access to bank accounts enable households to build savings, make payments, improve their access to loans, insurances and related services. It also helps small businesses to build up their creditworthiness.

Therefore, access to bank accounts is an essential contributor to financial inclusion and long-term economic growth.

Introduction

The JAU bank - a **social bank from East Africa** - has developed a sustainable ETF savings plan (SES) to support people to build modest reserves and savings while warranting a certain ratio of withdrawal of capital per time period.

Now JAU wants us to:

- build a **model** that allows them to **identify bank account holders** as well as potential new customers to inform them about the SES.
- make **brief analysis** of the current status of financial inclusion in their region and typical features of their current and potential customers

Since production and distribution of information material are cheaper than the fictive margin the JAU needs to generate, we agreed on the evaluation metric **RECALL**:

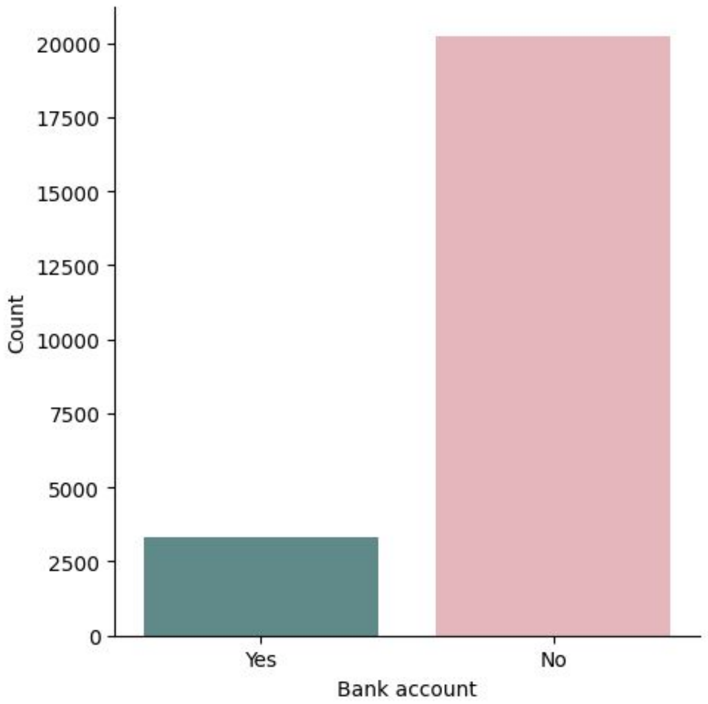
We want to identify as many of the current bank account holders as possible - even if this means accepting a few more false positives.

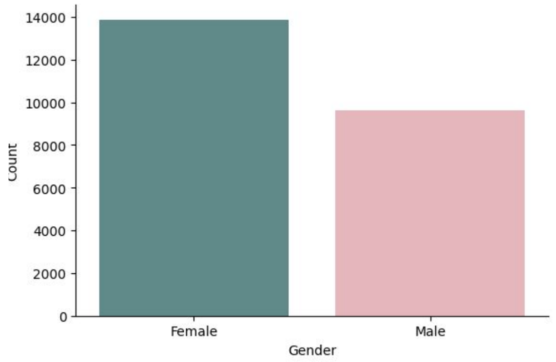
Perhaps these "false positives" are even potential new customers for a bank account?

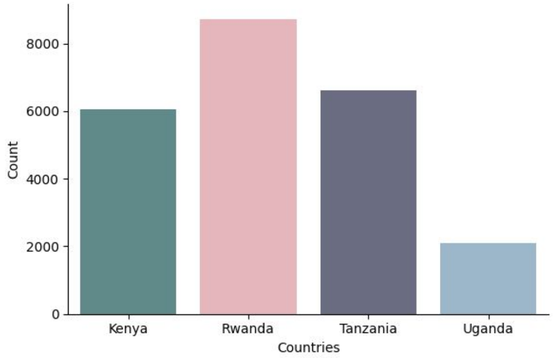
The dataset

- extracted from various surveys **from 2016 to 2018**
- contains the answers of slightly more than **23000 respondents from Kenya, Tanzania, Rwanda and Uganda**, to questions such as:
    - Do they live in a **rural or urban** area?
    - Do they have **access to a cell phone**?
    - What is the **size of their household**?
    - **How old** are they and what **gender** are they?
    - What is their **position in the household**?
    - Are they **married**?
    - What level of **education and job level** do they have?
- **relatively clean** (no missing data, no obvious outliers, no duplicates, etc.)
- main challenge: **imbalanced data**

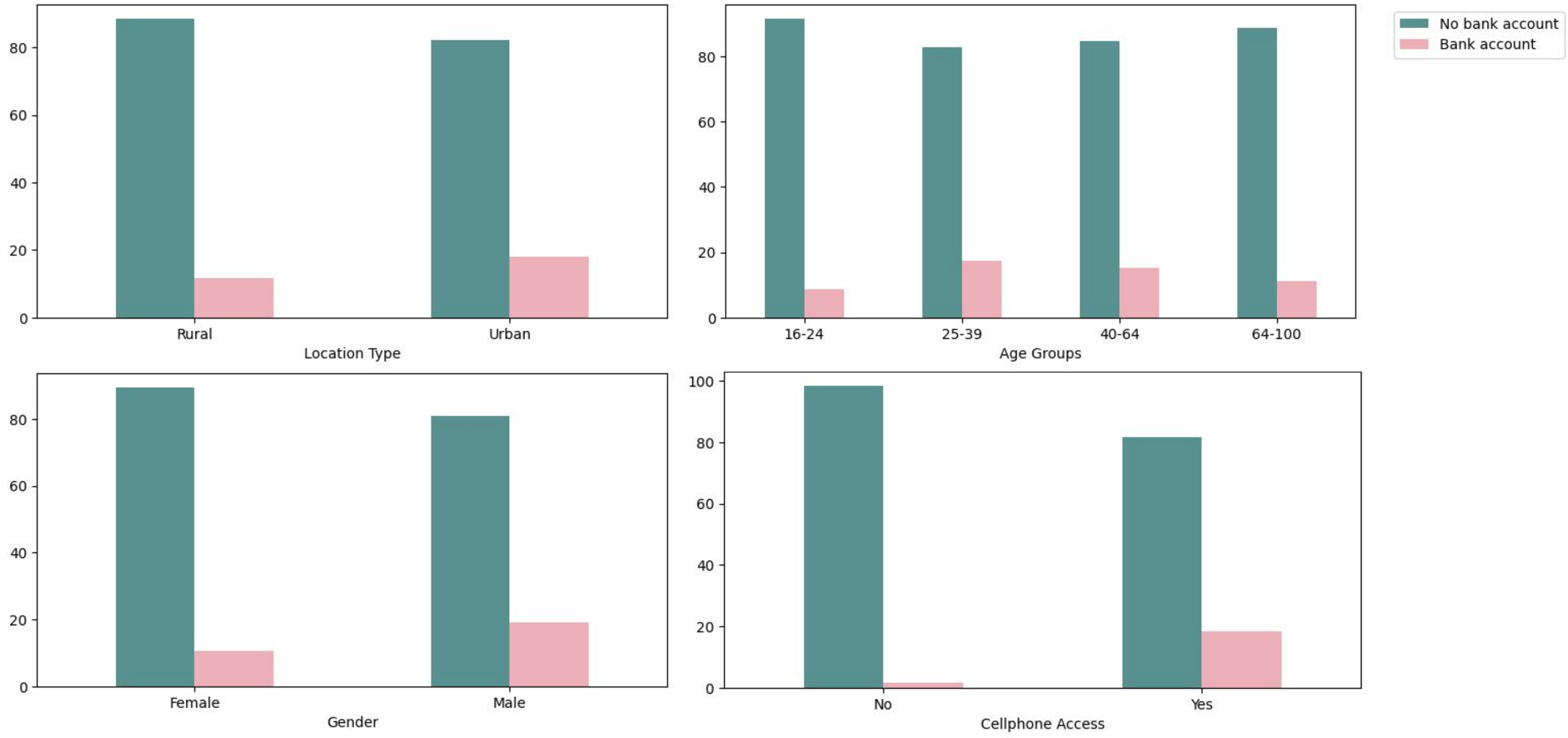
Analysis distributions in the entire dataset



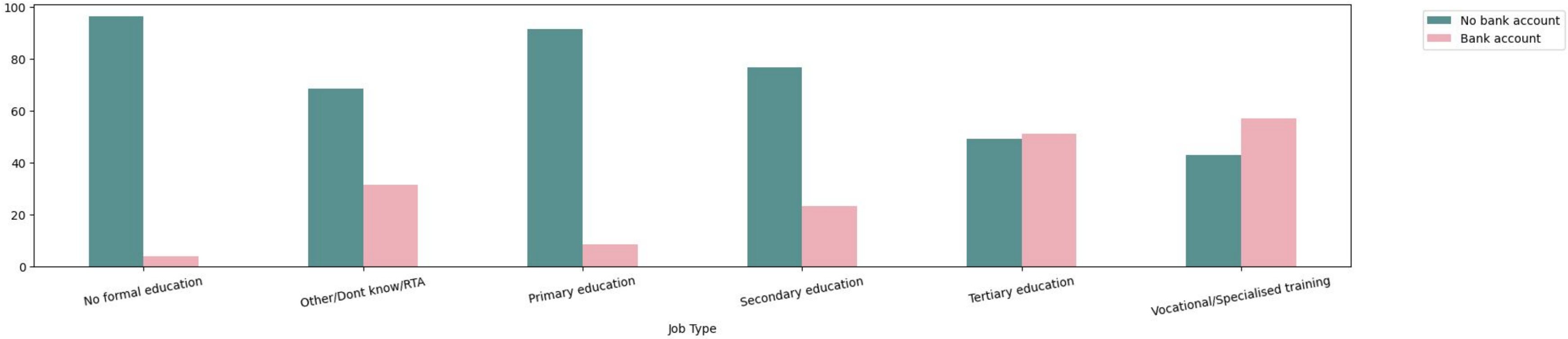


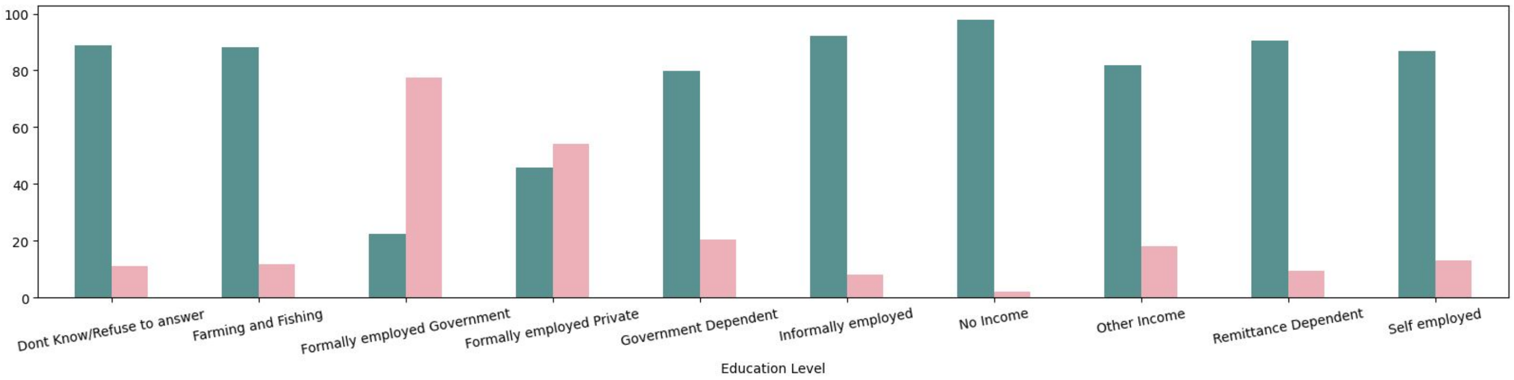


Analysis distributions of the target variable



Analysis distributions of the target variable





Analysis profiles of typical bank account holders

Zahir, 42, lives together with his wife and two of their children in a rural region. With his primary education he has his own business as a bike mechanic.

He owns a cellphone.

Asante, 62, lives in a rural region together with two of her sisters. Her husband died five years ago. With her primary education she is farming her entire life.

She owns a cellphone that some of her grandchildren are allowed to use.

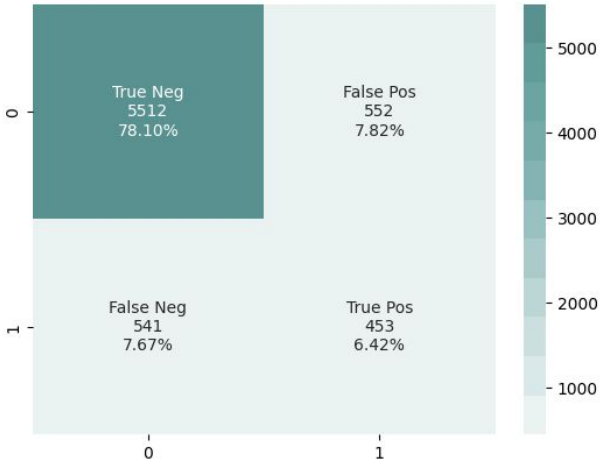
Nala, 33, lives in a medium sized city and uses her cellphone every day. Her college education helped her being successful in her own business as agent for female writers and other artists. She lives together with her fiancé and his two younger sisters.

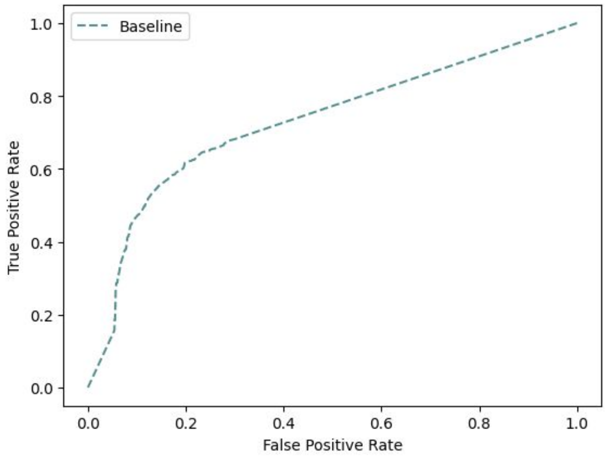
Kovu, 32, lives together with his brother and sister-in-law in the capital of his country. He works as an employed lawyer in a global lawfirm.

He owns two cellphones, one private and one business.

Baseline Model Decision Tree

- Simple **Decision Tree** with all features
- F1-Score: **0.44**
- Recall: **0.42**





Final Model XGBoost Classifier

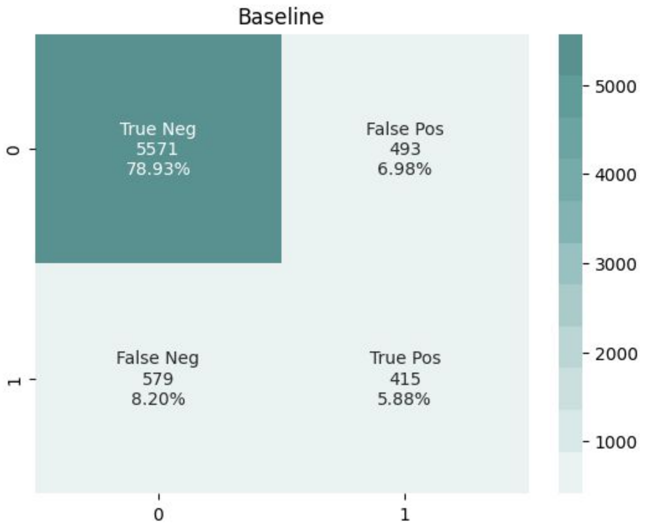
What is the XGBoost Classifier?

- implementation of gradient boosted decision trees
- combines the estimates of a set of models
- models are added sequentially to correct the errors made by existing models, until no further improvements can be made
- uses a gradient descent algorithm to minimize the loss when adding new models
- allows us to specify class weights: by increasing the weight of the minority class, we can give it more influence

How did we use it?

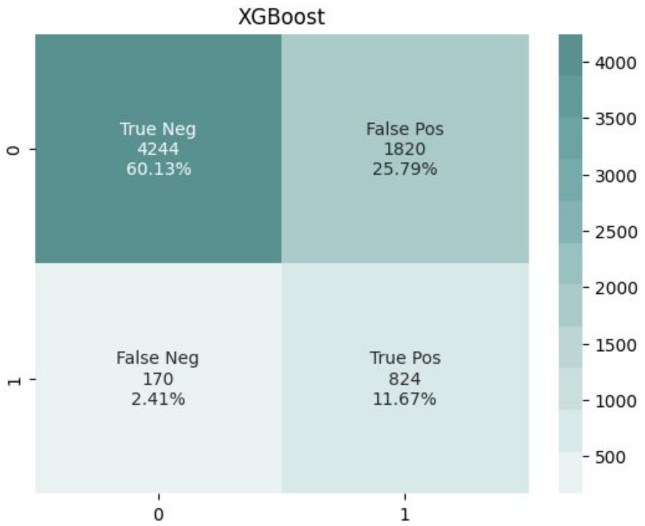
- trained the model on the entire, but simplified, encoded and scaled data
- optimized the model for a high recall
- then applied a relatively high threshold (0.738) to make the final predictions
- decision for this threshold is based on a break even calculation.

Final Model Baseline vs XGBoost Classifier



F1-Score: 0.**44**

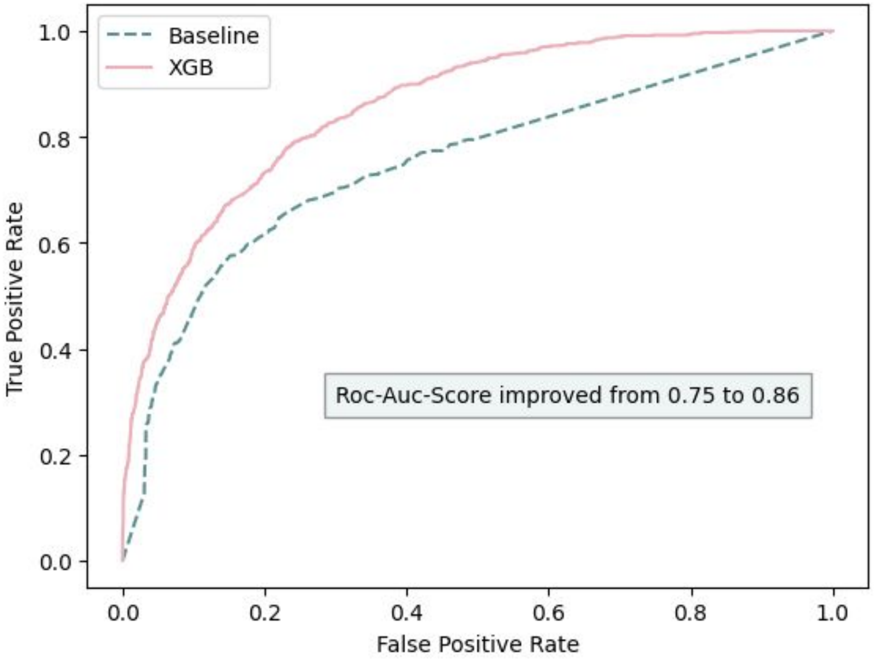
Recall: 0.**42**



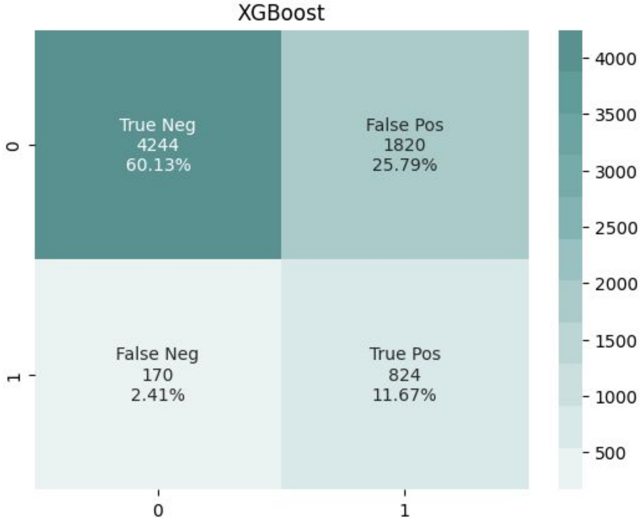
F1-Score: **0.45**

Recall: **0.83**

Final Model Baseline vs XGBoost Classifier



Final Model Cost Benefit Analysis of Final Model

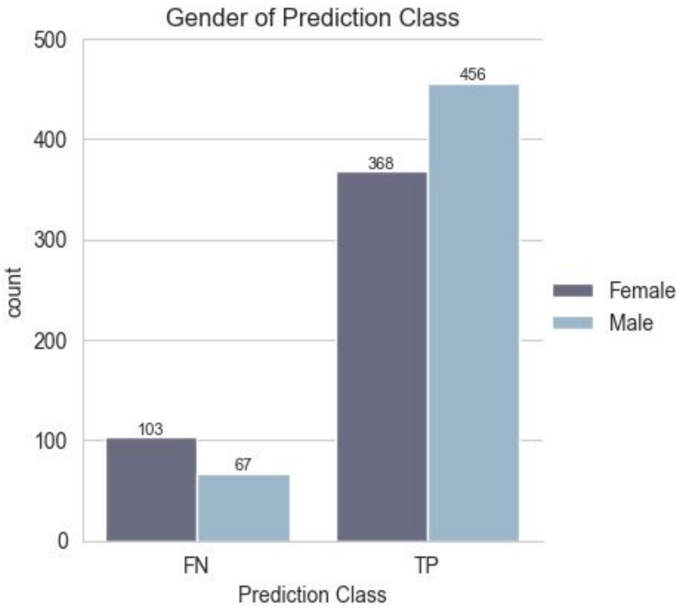


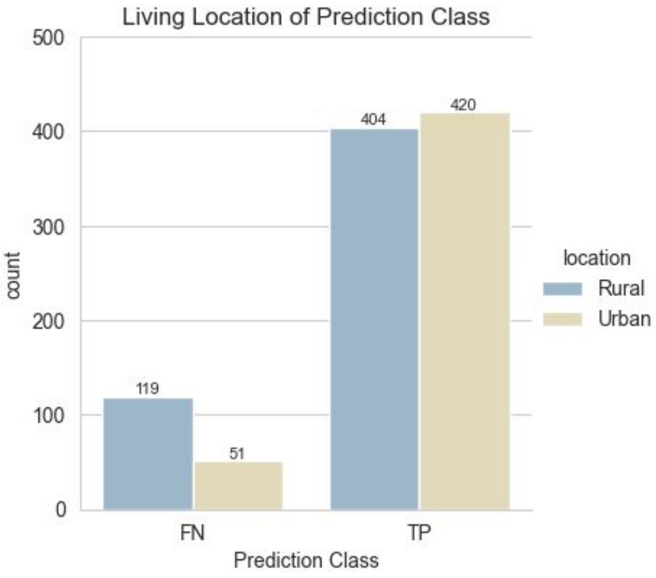
F1-Score: **0.45**

Recall: **0.83**

- Predictions always have downfalls:
  - too many unidentified bank account owners (FNs) result in lost sales
  - too many falsely predicted bank account owners in sunk costs (FPs)
- Cut off point based on following assumptions
  - 10% acquisition rate
  - lost sales due to unidentified bank account owners (FNs) of 100€ p.a. per saver over 10 years
  - sunk marketing costs due to falsely predicted bank account owners (FPs) of 8€
  - 0,5% acquisition rate of FPs

Error Analysis





Error Analysis

- Only 3 well educated bank account owners from Kenia were not identified and only 8 from Rwanda.
- In contrast, 80 of the non identified bank account owners are poorly educated and from Rwanda.