**REPORT**

**OBJECTIVE**

Building a model that predicts who is likely to have a bank account in a population given certain feature.

**METHODOLOGY**

**Data collection**

The dataset used for this project was gotten from Zindi data science community

**Data Preprocessing**

The dataset used does not contain any missing values but I found out that some columns contain outliers which make up a little percent of the total dataset. These outliers were replaced with the mode of the distribution.

**Feature Engineering**

The data was encoded using LabelEncoder and OneHotEncoder. The former was used for EDA while the later was used for getting the patterns in the data.

**Model Used**

The model used for building the model is the RandomForestClassifier. This was done using cross validation and the best hyperparameters were used after tuning.

**MODEL EVALUATION**

Evaluation Metrics: Mean Absolute Error (MAE)

Result: 0.1156134882402947

**INSIGHT**

One of the key findings in this data is that the population with bank accounts is smaller than that without a bank account. This may be due to some reasons such as;

1. No orientation on the importance of having a bank account
2. From the distribution based on education, people without formal education at least a primary school are those without bank accounts
3. The kind of job also influences the likelihood of an individual owning a bank account
4. Also, other criteria posed by the bank might hinder people from getting a bank account