Deposit Detection Case Study

AIMANEHI EHICHIOYA

Content

- Problem Definition
- EDA
- Data Preprocessing
- Model selection
- Model Evaluation Matrix
- Conclusion

Problem Statement

Nexus Bank is a financial institution that provide banking solutions such as term deposit, personal loan and mortgage financing.
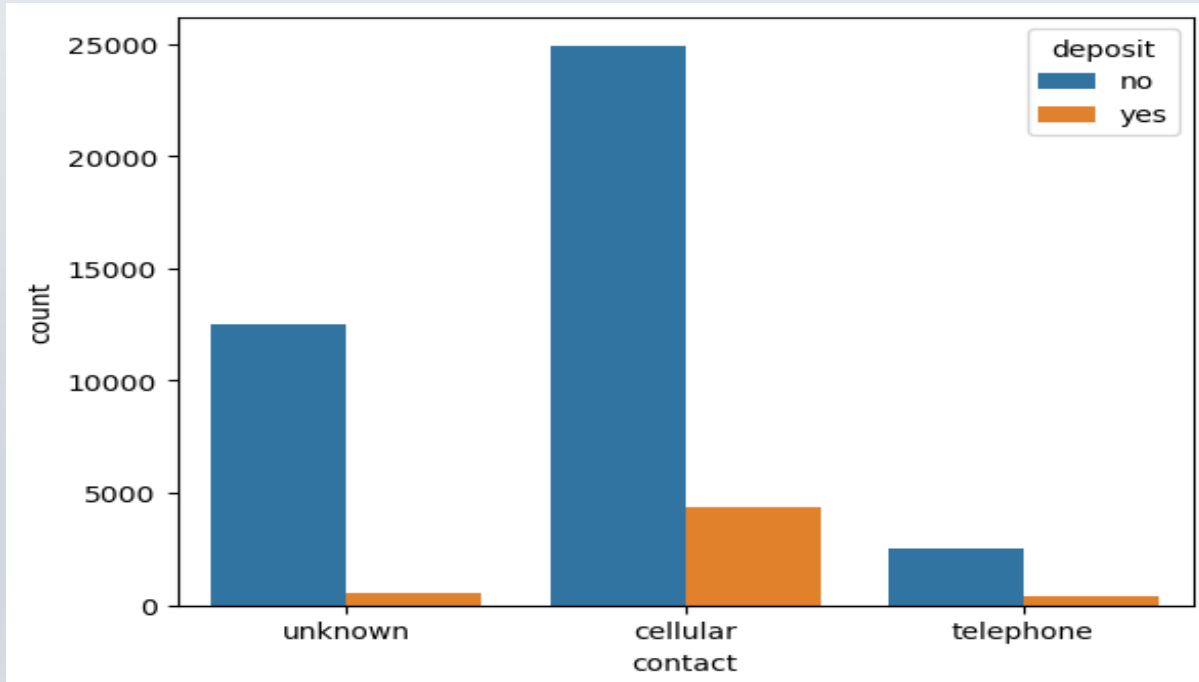
The board wants to:

- Anticipate future customer behavior and know the likelihood of deposit.
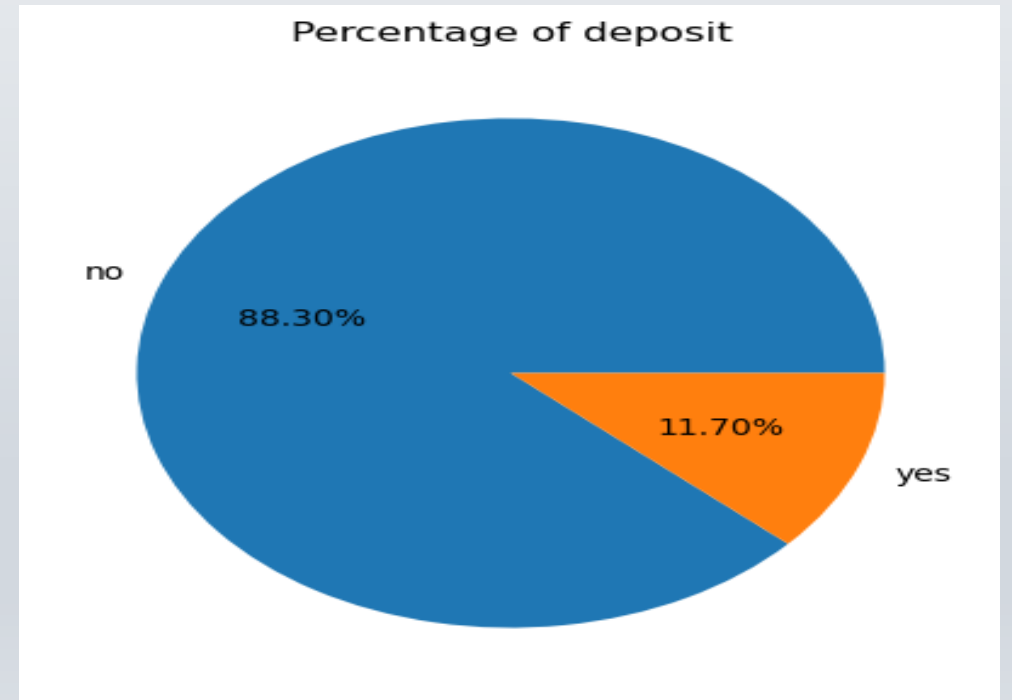- Optimize its operations, and improve customer deposits.

Exploratory Data Analysis

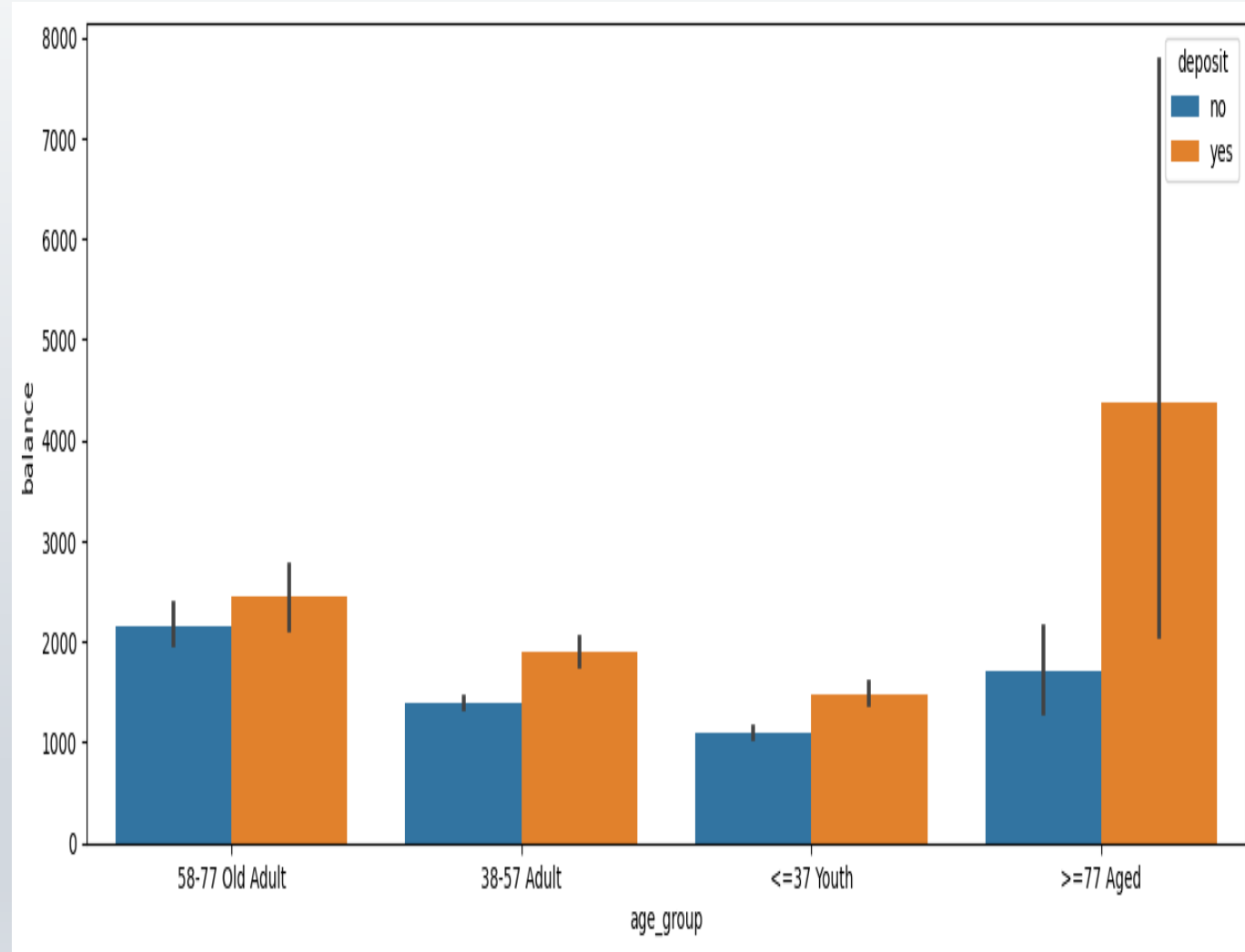➤ A lot of customers were contacted through cellular but few deposited

➤ The percentage of customer that didn't deposit is 88% which will lead to loss in the company.
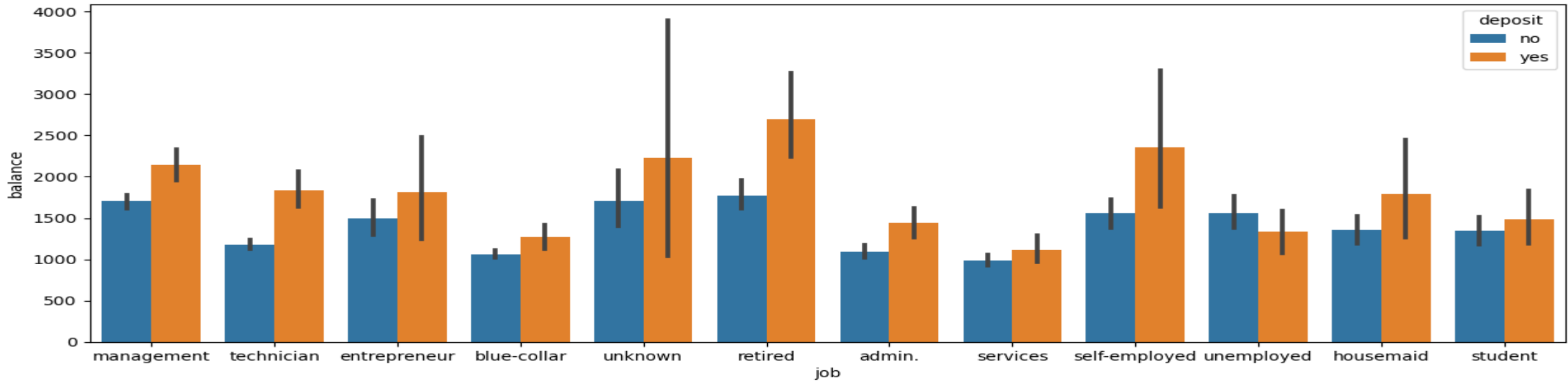
Exploratory Data Analysis

➢ The age is divided into 4groups which includes Youth(18-37), Adult(38-57), Old Adult(58-77) and Aged(above 77).

➢ From the chart the aged with high balance deposited

Exploratory Data Analysis

❖This chart shows that the occupation with the highest number of deposit are management, technicians, retired and students comparing it to the number of customer in the different occupation and those with high balance made deposits.

Exploratory Data Analysis

➢ The boxplot in the chat shows the datasets contain many outliers
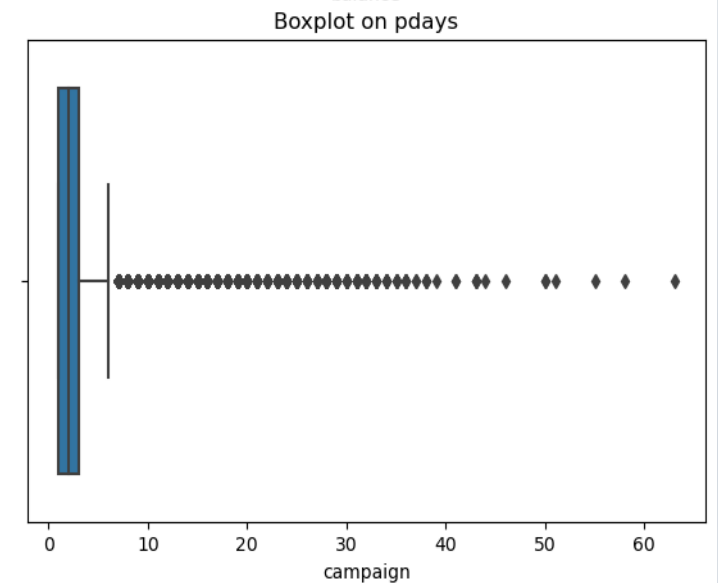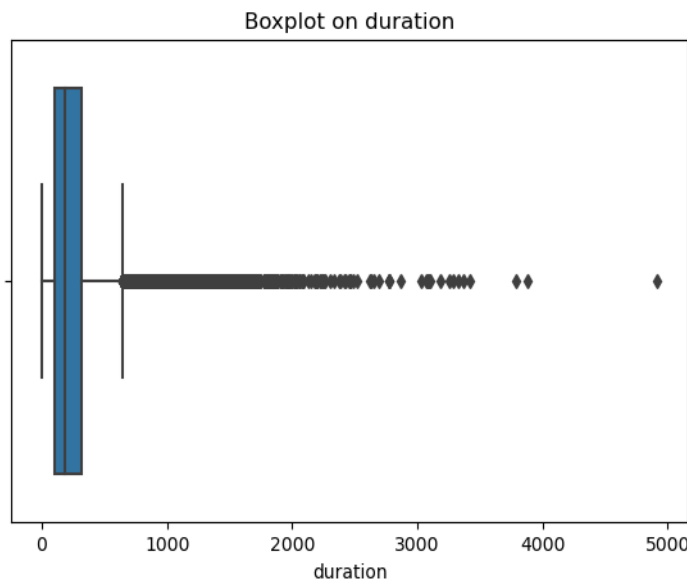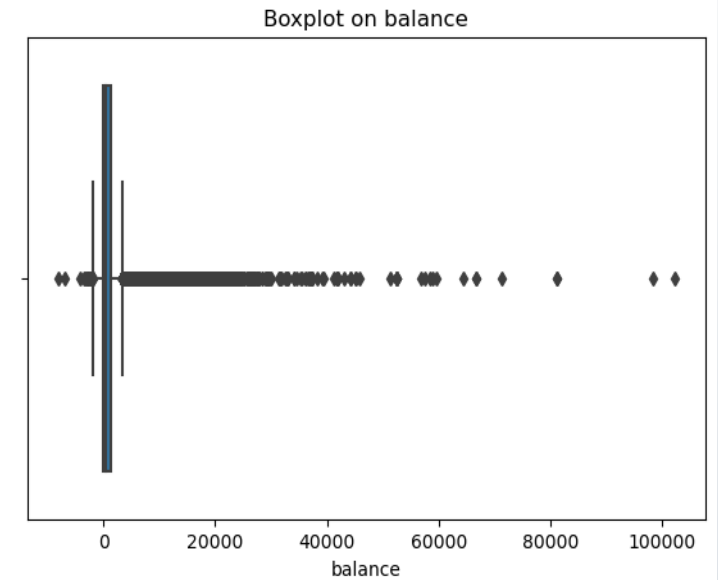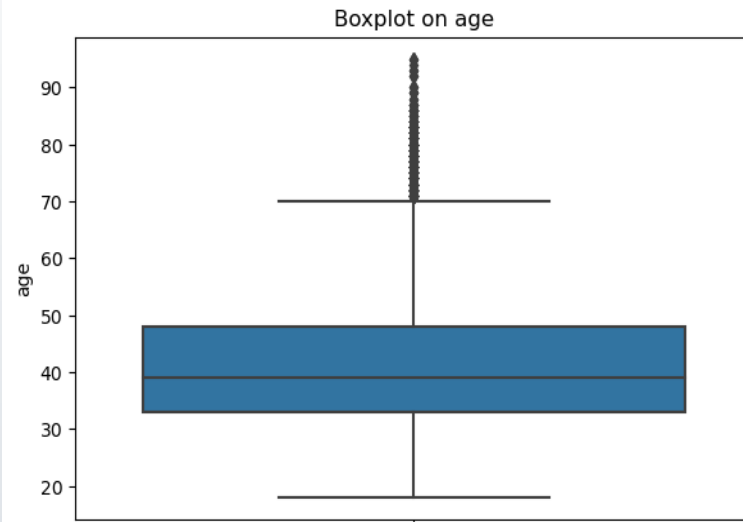
Exploratory Data Analysis

➢ There is no correlation within the dataset except the number of days that passed by after the customer was last contacted and the number of contacts made to the customer before this campaign.

Data Preprocessing

Encoding

➢ The Label Encoder was used to convert all categorical variable to numerical variable.

Feature Engineering

➢ Selecting the most important features for the machine learning model.

MinMax Scaler

➢ To normalize the datasets and removed outliers the minmaxscaler was used.

Model Selection

Splitting the data into training and testing sets is a common technique used to evaluate the performance of a machine.

The datasets was split into **80/20** ratio for training and testing respectively.

Model used to train and test the datasets are:

➢ XGBoost Classifier

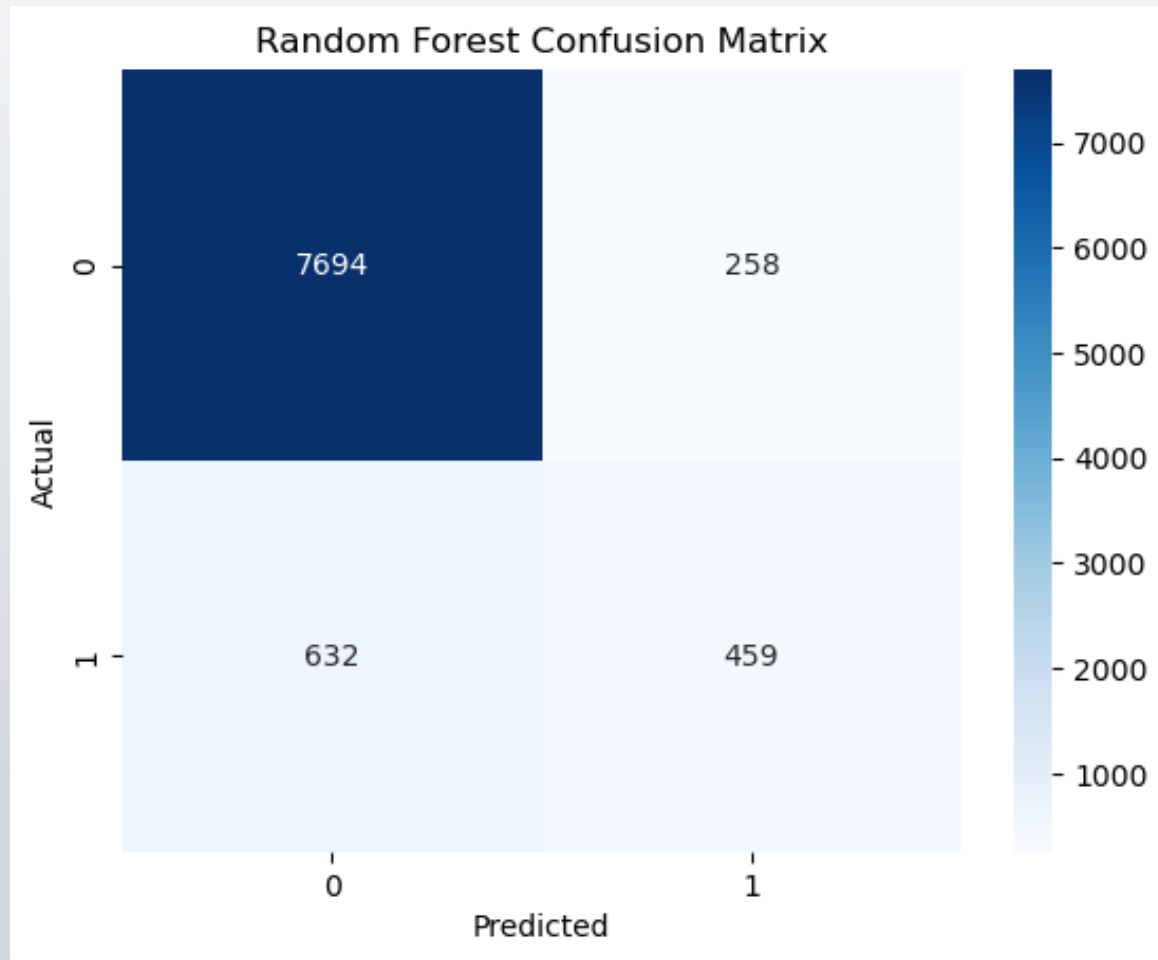➢ Random Forest Classifier

➢ Logistic Regression

Evaluation Matrix

XGBoost Classifier

➢The Accuracy is 0.9077739688

➢The Precision is 0.6576687116

➢The Recall is 0.491292392

➢ The F1 score 0.56243441

➢AUC Score is 0.72810343

The confusion matrix indicates that the number of true positives is 7673

False positives is 279, false negative is 555 and true negative is 536.

XGBoost Confusion Matrix
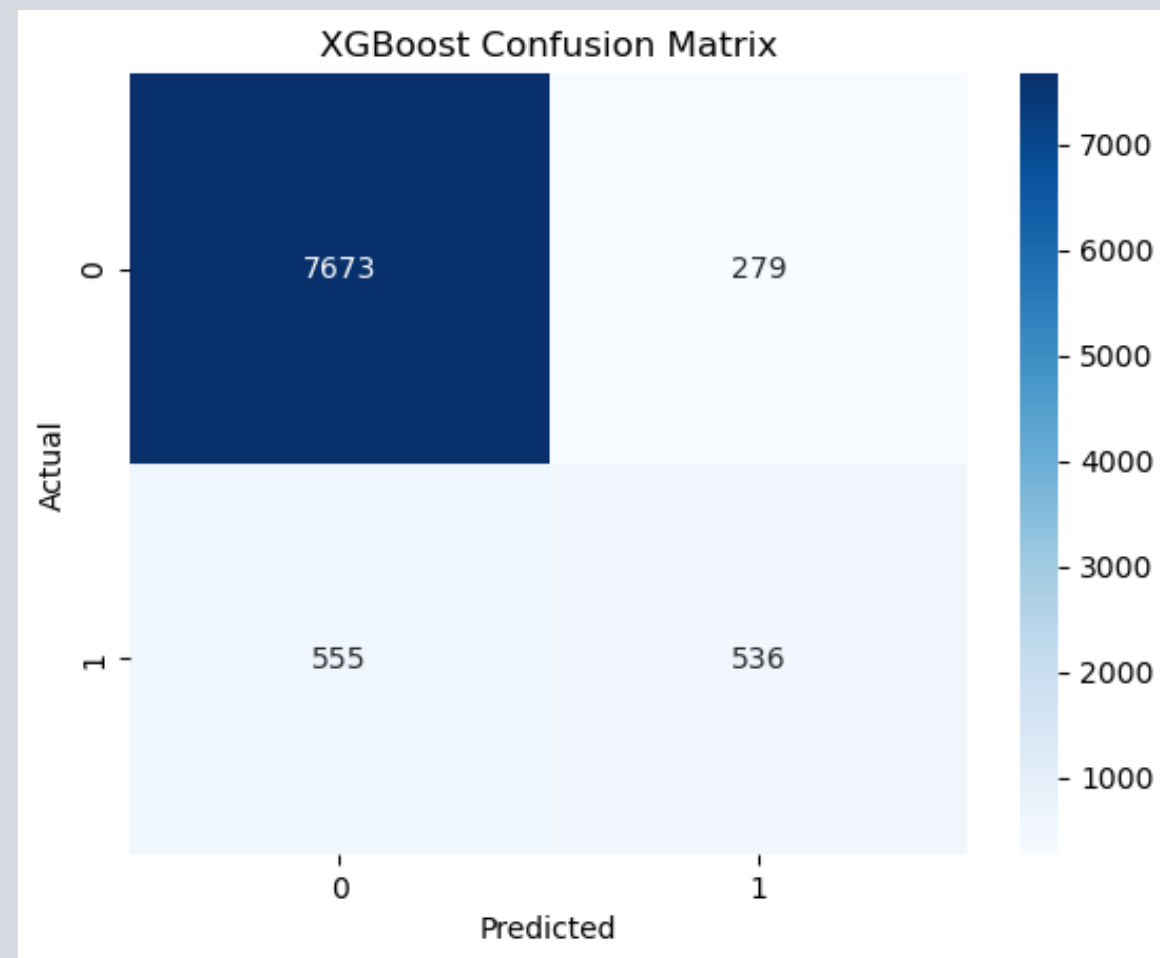
Evaluation Matrix

Random Forest Classifier

➤ The Accuracy is 0.90158133

➤ The Precision is 0.6401673

➤ The Recall is 0.4207149

➤ The F1 score 0.5077433

➤ AUC Score is 0.694135

The confusion matrix indicates that the number of true positives is 7694

False positives is 258, false negative is 632 and true negative is 459.

Evaluation Matrix

Logistic Regression

➢The Accuracy is 0.883556341

➢The Precision is 0.58796296

➢The Recall is 0.116406966

➢ The F1 score 0.19433817

➢AUC Score is 0.55260740

The confusion matrix indicates that the number of true positives is 7863

False positives is 89, false negative is 964 and true negative is 127.
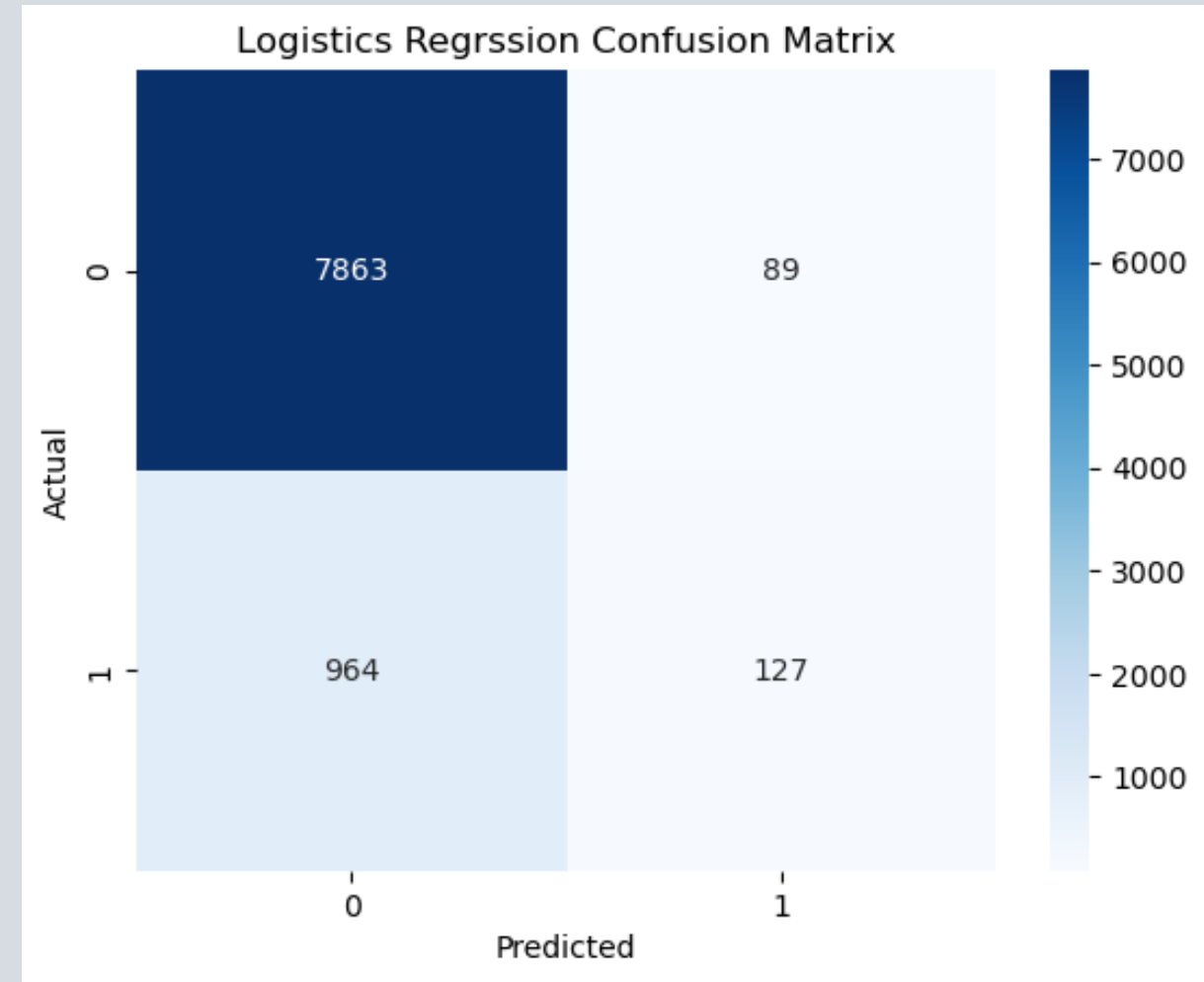
Conclusion

The model chosen based on the accuracy, interpretability, complexity and scalability of the model.

The primary metrics important to solve bank deposit accuracy, Which make XGBoost Classifier the best model for this datasets.

Thank you