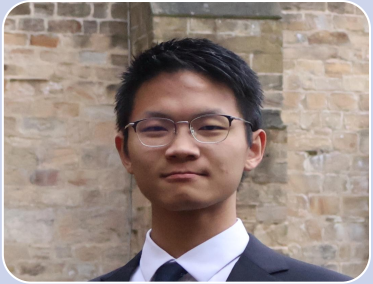
BankBytes

Final report

BANKBYTES



Mark Kimutai Kitur

kiturmark@gmail.com

Kenya

Dedan Kimathi university

Of technology

Data Science

Yue hu

hy1550278246@gmail.com

The United Kingdom

Durham University

Data Science

Rakshith Nagaraju

rakshith.nagaraj6@gmail.com

The United Kingdom

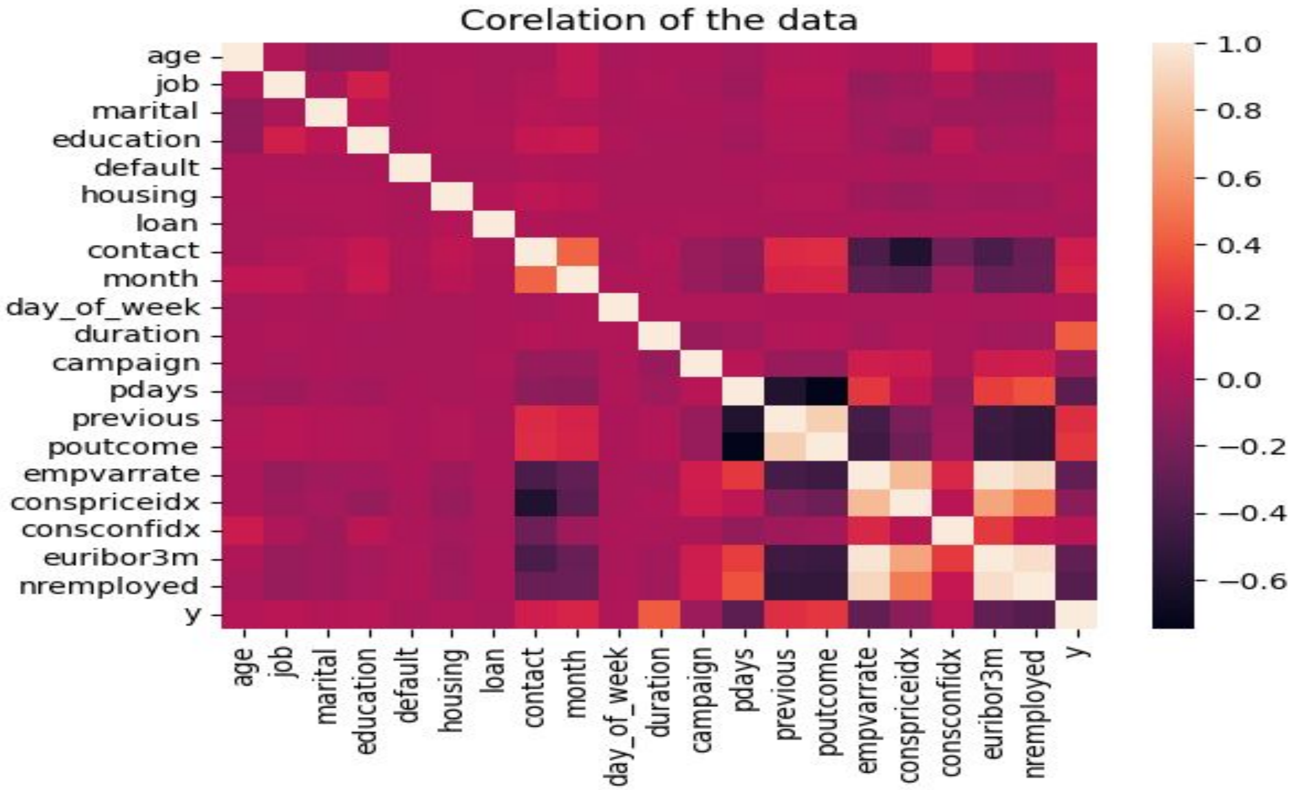
University of Liverpool

Data Science

Preparation of data for model building

For a model to clearly learn the data and make good prediction on unseen data, columns which have less correlation should be removed. In this case we removed columns: nremployed, empvarate and conspriceidx. (The correlation is shown in the next page).

A graph showing relation among the columns



Evaluation Metrics

- Evaluation Metrics are criteria used to assess the performance of Machine Learning models.

- These metrics provide quantitative measures that help in understanding how well a model performs.

- These measures are used to measure and guide the selection process and tuning of the models which are eventually deployed in real world applications.

Metrics Used

- Accuracy
- Precision
- Recall
- F1-Score
- AUC-ROC

Accuracy

Accuracy is used to gauge the overall correctness of a model, especially when the dataset has balanced classes.

Measures the proportion of correctly classified instances out of the total instances.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Instances}}$$

Helps in quickly understanding how often the model is making correct predictions.

Precision

Precision focuses on the quality of positive predictions.

Proportion of true positive predictions among all positive predictions.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

Recall

Recall focuses on the coverage of actual positive instances.

Proportion of true positive predictions among all actual positives.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

Purpose of Precision vs Recall

These metrics help in understanding the types of errors a model is making, which is crucial in applications where the cost of false positives and false negatives differ significantly (e.g., medical diagnosis, fraud detection).

Precision is prefered in Medical diagnosis as having to check for something that may not be true is better than ignoring it thinking it was negative.

Recall is preferred in Fraud detection because you can check for fraud in something that is not actually a fraud but the other way around is harmful.

F1-Score

F1-Score combines precision and recall into a single metric.

Similar to F1 Score but gives more weight to Recall.

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Provides a balanced measure of a model's performance, particularly useful for imbalanced datasets where a single metric like accuracy can be misleading.

AUC-ROC

AUC-ROC is used to evaluate the performance of a binary classification model.

Usually used in Financial models.

ROC Curve: Plots True Positive Rate (Recall) against False Positive Rate.

AUC: Represents the probability that a randomly chosen positive instance is ranked higher than a randomly chosen negative instance.

Allows for easy comparison of multiple models' performance in terms of their discriminatory power.

By using these evaluation metrics, we can:

- Ensure that the models meet the required criteria for performance.
- Adjust the hyperparameters that can fine tune the performance to address specific business needs.
- Make informed decisions on which models to use during deployment.

Linear model

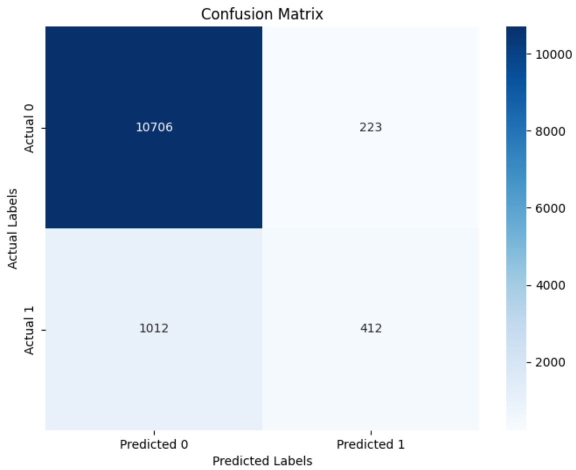
1. Classification Report

category	precision	recall	F1-score
0	0.91	0.98	0.95
1	0.65	0.29	0.40

	precision	recall	F1-score
Accuracy			0.90
Macro avg	0.78	0.63	0.67
Weighted avg	0.88	0.90	0.88

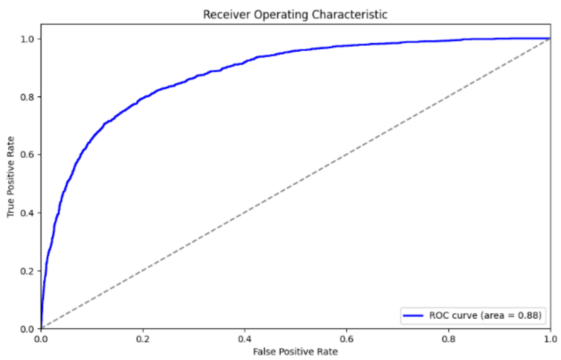
The results from the two tables demonstrate that this model is effective in classifying category 1, indicating that the majority of customers are unlikely to purchase the bank's product service. However, the performance in category 1 is not as strong as desired. One contributing factor could be the imbalanced sample and limited sample size.

2. Confusion matrix



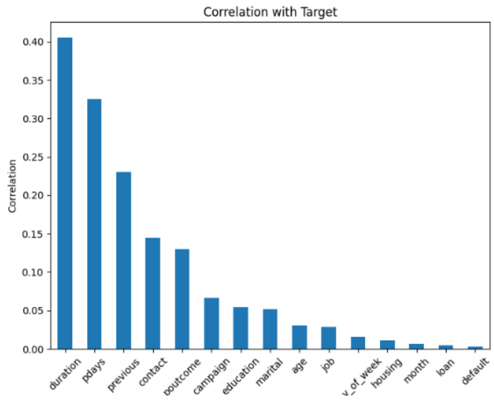
The overall accuracy of the model is approximately 90%. The majority of the samples are predicted to be category 0. However, the model's performance could be improved.

3. ROC curve



The AUC value obtained by this model is 0.8828, indicating that the model has high accuracy and stability in distinguishing positive and negative samples. This is an excellent performance, and the AUC value also shows that the linear model has high robustness and can perform well with a low false alarm rate.

4. Correlation between independent and dependent variables



The plot indicates that the duration of the last contact has the strongest correlation with the dependent variable, suggesting that this factor has a significant impact on the likelihood of a customer purchasing the bank's products. Additionally, the number of days that have elapsed since the last contact also has a notable influence on the dependent variable. However, the presence of a default credit status or loan does not appear to have a substantial effect on the dependent variable.

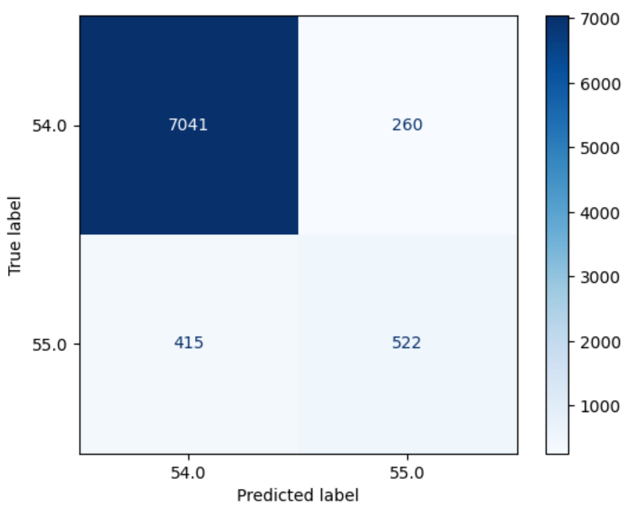
5. Compare with other models:

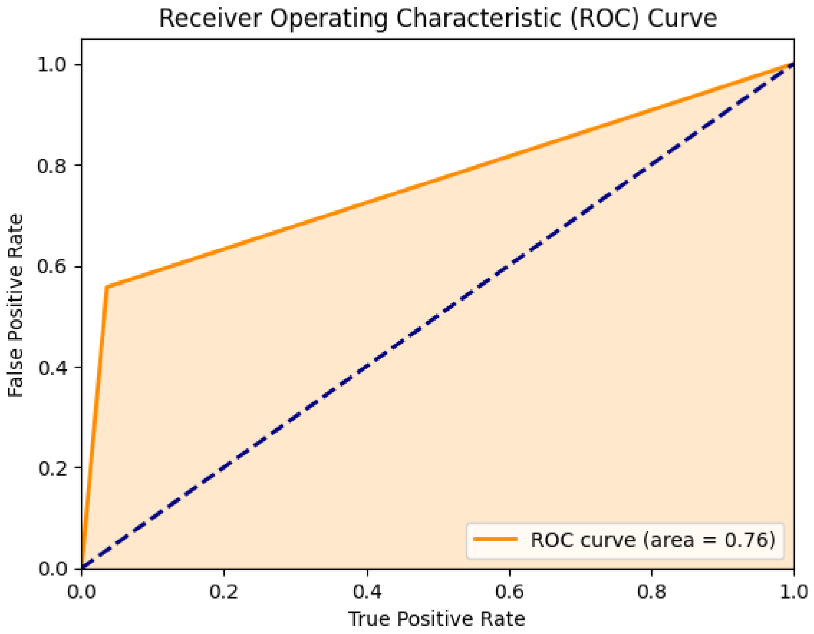
 a. Ensemble (AUC: around 0.67)

category	precision	recall	F1-score
0	0.92	0.97	0.95
1	0.65	0.37	0.47

	precision	recall	F1-score
Accuracy			0.90
Macro avg	0.78	0.67	0.71
Weighted avg	0.89	0.90	0.89

 b. Boosting (Acc.: 0.91%)





The results demonstrate that the boosting model exhibits slight superiority over the linear model in terms of overall performance. The ensemble model, on the other hand, demonstrates comparable performance to the linear model, with the notable distinction of more accurately predicting category 1. However, both models' AUC value is lower than that of the linear model, indicating that they may not be as robust as the latter. Furthermore, since the linear model is the simplest of the three, it can be calculated more rapidly than the other two, resulting in greater efficiency than the boosting and ensemble models.