CUSTOMER CLASSIFICATION - Bank Case Study

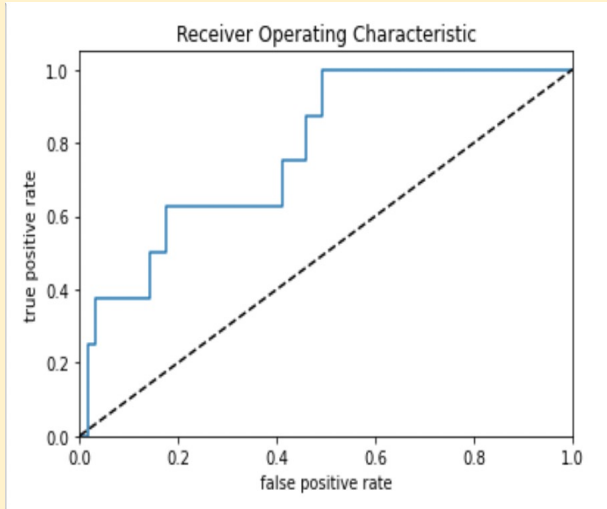
HYPOTHESIS:

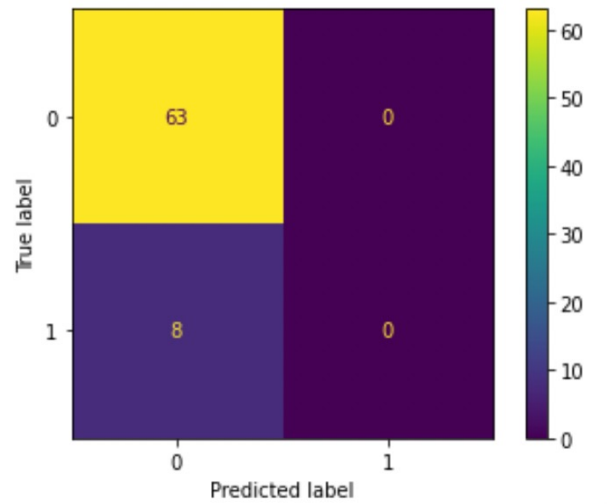
- **Ho:** the model **CAN NOT** predict accurately whether new customers will be "good" or "bad".
- **Ha**: the model **CAN** predict accurately whether new customers will be "good" or "bad".

STRATEGIC IMPACT:

An accurate logistic model would allow the bank to **increase efficiency** whilst **mitigating risk** and avoid bankrupcy. It other words, it would help in becoming **more profitable** as well as more **reliable.**

VISUALISING ACCURACY:





- **ROC curve plot** (left) shows the model is better at predicting than the threshold (dotted line), accuracy being 78%
- **Confusion Matrix** (right) is showing that the model was not capable of predicting "bad" customers (status B)

NEXT STEPS

The model is not ready to be deployed, however, the following measures could be applied to improve its accuracy:

- Get new data points:
  - External: such as income, household income
  - Internal: obtaining gender from birthdate
- Apply feature engineering to existing columns
- Drop highly correlated fields
- "Play" with the % of data assigned to training and testing
- Apply SMOTE and TOMEK LINKS methods to deal with unbalanced data
- Keep iterating until we either proof Ha or we confirm Ho.