



# Bank Customers Attrition Prediction Presentation

Ndanu Mwatu; PTO4; 24 October 2023

## Outline



1. Overview
2. Business Problem
3. Data Understanding
4. Preprocessing/ Preparation of Data
5. Data Modeling and Evaluation
6. Results
7. Conclusions / Recommendations
8. Next Steps

# Overview

This project focuses on solving a business problem using classification models through an iterative approach.



## Business Problem



- ❖ ABC Multinational Bank is concerned over high customer attrition rates.
- ❖ They seek a classification model to predict customer turnover - in order to apply mitigation measures beforehand.

## Data Understanding



- ❖ This project uses account holders' data for ABC Multinational Bank customers, downloaded from Kaggle.
- ❖ It comprises of 10,000 rows and 12 feature columns including; credit score, country, age, balance, estimated salary and churn (the target variable).

# Preprocessing/ Preparation of Data



## These include;

- Define X (predictors) and y (target)
- Split the data into training and test sets
  - Check for missing values
  - One hot encode categorical data
- Normalize/standardize numeric features

## Data Leakage

Preprocessing done after splitting the data into train and test sets -  
to prevent data leakage.

# Data Modeling and Evaluation

## Logistic regression, Decision Trees and Random Forest model types built

### Iterative Approach;

#### i. Logistic Regression Modelling

- Establish Baseline model
- Create Iterative models & determine the best

#### ii. Decision Trees Modelling

- Establish Baseline model
- Create Iterative models & determine the best

### Iterative Approach;

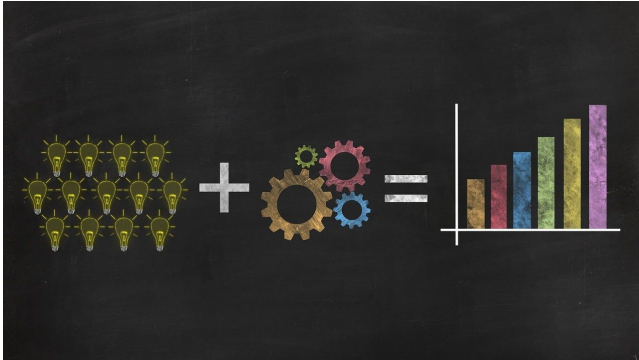
#### iii. Random Forest Modelling

- Establish Baseline model
- Create Optimal model - using **GridSearchCV**
- Feature Importances of the Random Forest Models

#### iv. Evaluate the Overall Best Model

**Evaluation Metrics:** Log loss, ROC curves, AUC, Precision, Recall, Accuracy, f1 Score

# Results



## Logistic Regression Modelling:

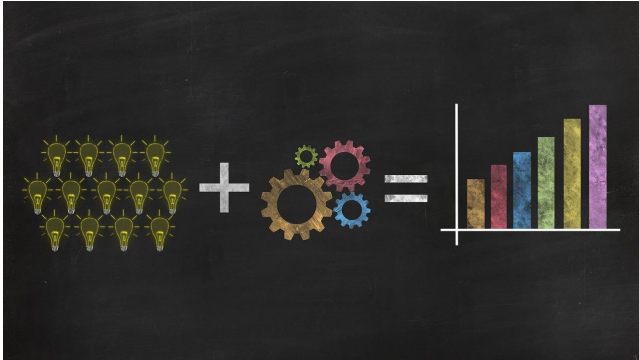
**Best** Performing model has;

- ❖ Log loss: 0.433
- ❖ Parameters:
  - Scaled data
  - Increased regularization
  - 'Saga' solver.

*Baseline and 6 iterations built.*



# Results



## Decision Trees Modelling:

**Best** Performing model has;

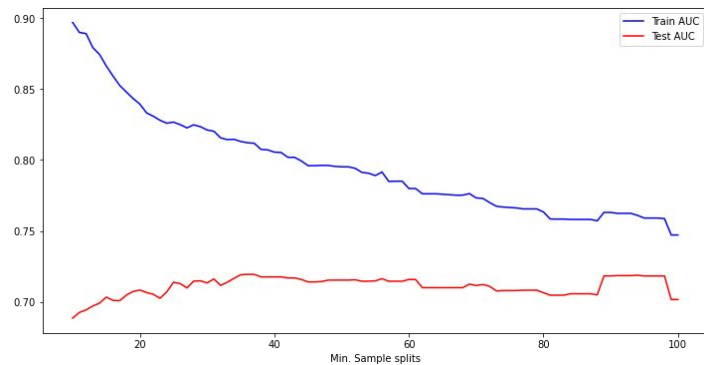
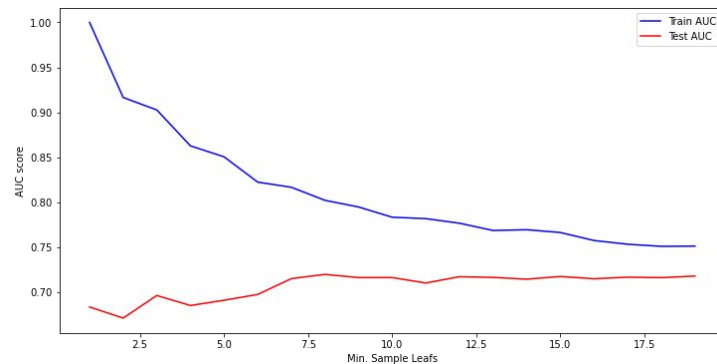
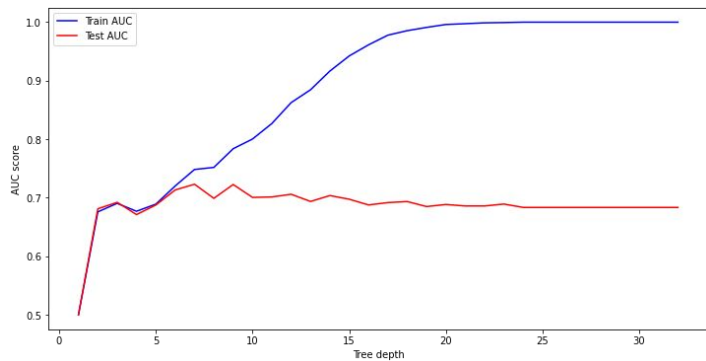
- ❖ Log loss: 0.406
- ❖ Parameters:
  - Max\_depth=6
  - Min\_samples\_split=91
  - Min\_samples\_leaf=16

*Baseline and 4 iterations built.*

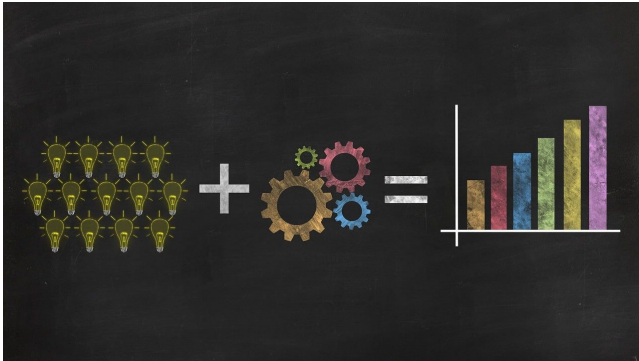
# Results



## Decision Trees Parameters



# Results



## Random Forests Modelling:

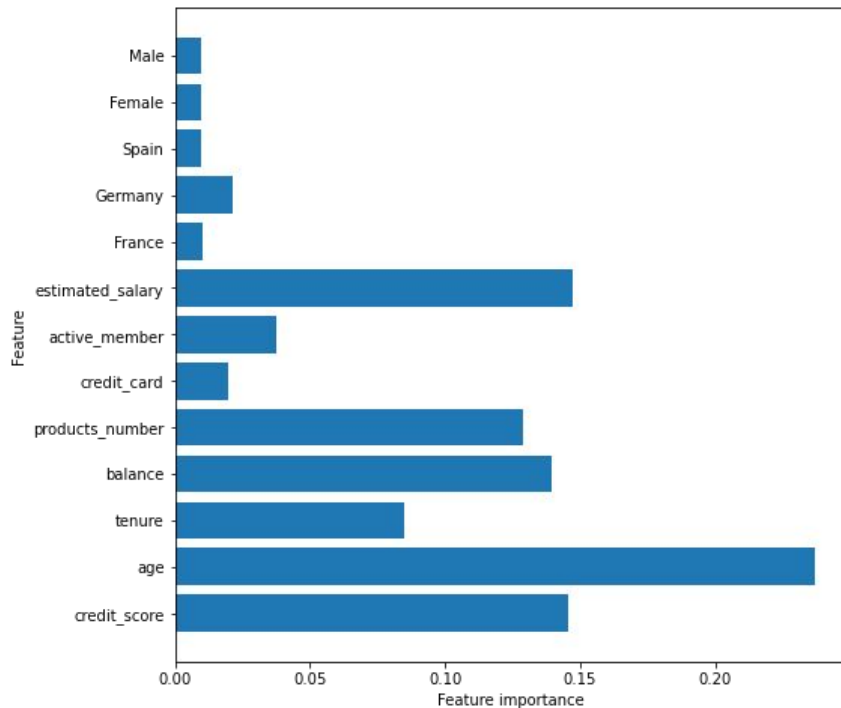
**Best** Performing model has;

- ❖ Log loss: 0.36
- ❖ Parameters:
  - Criterion='entropy'
  - Max\_depth=9
  - Min\_samples\_leaf=16
  - Min\_samples\_split=91
  - n\_estimators=100

*Baseline and iteration using  
GridSearchCV parameters*

# Results

## Feature Importance for Random Forest Model



**Age** is the feature with highest importance.

## Results: Overall Best Model Evaluation - Optimal Random Forest Model



- ★ Training Precision: 0.866
- ★ Testing Precision: 0.8591
  
- ★ Training Recall: 0.394
- ★ Testing Recall: 0.380
  
- ★ Training Accuracy: 0.863
- ★ Testing Accuracy: 0.864
  
- ★ Training F1-Score: 0.541
- ★ Testing F1-Score: 0.527

# Conclusions/ Recommendations



1. The client advised to make use of the overall best model to predict customers that are likely to leave the Bank and target intervention strategies.
2. This model may not be sufficient to decide on best candidates for provision of loans, thus the client should be cautious in utilising it as such.
3. The client advised to pay attention to age of customers, credit\_score, estimated salary and balance as features of importance when designing intervention strategies to retain customers.

## Next Steps



For additional insights, further analysis is proposed in the following areas;

- ❖ Further tuning is proposed of the hyperparameters of the best performing model in order to lead to better performance metrics particularly the f1 Score. This further analysis could include use of XGBoost.
- ❖ Adjustment of the model's Recall and Precision could be done in line with the focus of the Bank. A further discussion with the Bank to understand their needs and focus is needed e.g. are the intervention measures likely to be too costly, in which case, the client would want a model that is even more precise?



## References

All images (*that are not visualizations generated from data*) are downloaded from [www.pixabay.com](https://www.pixabay.com)





**THANK YOU**