**BANK SUBSCRIPTION PREDICTION REPORT**

By

Peniel Tenkoramah Twum

**Dataset Overview**

The data contains 41,188 records with customer information and campaign contact results.

The target variable is y (whether a client subscribed to a term deposit: yes/no), which was mapped to a binary variable y\_binary (1 for yes, 0 for no).

**Key Insights from Exploratory Data Analysis (EDA)**

Demographic & Behavior Patterns:

Age Group: Most clients are between 30–45 years, with decreasing subscription rates in older age brackets.

Marital + Education: Combining these shows nuanced insights:

Single clients with tertiary or professional education show higher subscription rates.

Married individuals with only basic education tend to not subscribe.

Job Type: White-collar jobs like "management" and "technician" have better subscription conversion.

Month & Contact Timing: Most contacts were in May, but subscription rates were better in March and December.

**Financial & Campaign Attributes:**

Duration of the last contact was highly predictive but dropped for modeling (data leakage risk).

Pdays = 999 often indicates no previous contact; most clients had this value.

Campaign Count: Too many calls reduced success, excessive contact likely annoys clients.

Economic indicators like euribor3m and emp.var.rate slightly influenced outcomes but required normalization.

**Feature Engineering**

Created new variables like:

age cluster (grouped age ranges)

marital\_edu (combined marital status and education level)

Outliers were capped for variables like age and campaign duration using the IQR method.

Unknown entries in categorical variables were removed to improve model reliability.

**Predictive Modeling Results**

**Models Used:**

Logistic Regression

Random Forest

XGBoost (tuned)

**Performance Summary:**

Model Accuracy F1-Score (Yes) Recall (Yes) Precision (Yes)

Logistic Regression 0.89 0.35 0.25 0.62

Random Forest 0.87 0.37 0.30 0.49

XGBoost (Tuned) 0.88 0.38 0.30 0.53

**Insight**: All models perform well on 'No' predictions, but struggle with detecting actual 'Yes' (subscribers) due to class imbalance (majority did not subscribe).

**Recommendations and Next Steps**

Use class balancing techniques (e.g., SMOTE, class weights).

Investigate high-impact variables like job, month, poutcome, and euribor3m.

Consider stacking or ensemble models.

* A/B test future marketing campaigns using top predictors (timing, contact method, job segment).

**Final Insight:**

The model can reliably identify non-subscribers, helping the bank avoid wasting resources, but needs improvement to better target potential subscribers especially those in underrepresented positive cases.