**Final Report: Bank Marketing Term Deposit Prediction**

**Business Understanding**

A Portuguese bank conducted several telemarketing campaigns to promote term deposits. The business goal is to improve the effectiveness of these campaigns by identifying clients most likely to subscribe, allowing the bank to prioritize resources and reduce marketing costs.

**Business Objective:** Build a predictive model to determine whether a client will subscribe to a term deposit based on their profile and past interactions.

**Data Cleaning Summary**

• Loaded data from the UCI repository (bank-additional-full.csv)

• Identified and handled “unknown” values in job, marital, education, default, housing, and loan columns.

• Removed the duration column, which leaks target information.

• One-hot encoded all categorical features for model compatibility.

• Converted the target variable y from 'yes'/'no' to 1/0.

**Descriptive & Inferential Statistics**

• **Class Imbalance:** ~88.7% of clients did not subscribe (y=0), while only ~11.3% did (y=1).

• **Age:** Median age was slightly higher among subscribers.

• **Job & Education:** Certain jobs (e.g., management, retired) and higher education levels correlated with a higher likelihood of subscription.

• **Campaign History:** Clients with fewer prior contacts had higher subscription rates.

**Key Findings**

**Model Performance:**

• **Baseline (majority class):** ~88.7% accuracy.

• **Logistic Regression:** Matched baseline in accuracy but had **poor recall** for class 1.

• **Decision Trees and KNN:** Easy to interpret but struggled similarly due to class imbalance.

• **SVM:** Performed well on accuracy but slow to train and not significantly better.

**Insights for Business (non-technical):**

• The current model struggles to identify clients who say “yes”.

• Relying solely on accuracy is misleading; precision and recall are more meaningful.

• Some job roles, education levels, and previous interactions strongly influence outcomes.

**Next Steps & Recommendations**

1. **Handle Imbalance More Aggressively:**

• Adjust classification **threshold** or optimize for **recall/F1-score**.

2. **Try other Models:**

• Train **Random Forest** and **Gradient Boosting** (e.g., XGBoost or LightGBM).

3. **Feature Engineering:**

• Create more meaningful variables like “previously contacted”, “contact success rate”, or combine job + education.

4. **Hyperparameter Tuning:**

• Apply GridSearchCV to find best model configurations.

5. **Visual Reporting for Business Users:**

• Build a dashboard or visual report with high-confidence leads for marketing to act on.