**Bank Data Prediction**

**Advance Project 7**

**Report**

**Title: Bank Term Deposit Subscription Prediction**

**1. Objective**

To develop a machine learning model that predicts whether a bank client will subscribe to a term deposit based on their profile and interaction history.

**2. Solution**

* **Data Source:**  
  Public dataset from a Portuguese bank containing details of marketing campaigns.
* **Tech Stack:**  
  Python, Jupyter Notebook, Scikit-learn, Seaborn, Matplotlib, Pandas
* **Key Activities:**
  + Data loading and exploration
  + Detailed EDA with graphs and insights
  + Preprocessing (encoding, scaling)
  + Model building (Random Forest)
  + Evaluation with metrics and ROC-AUC
  + Inference on sample inputs
    1. **Solution Architecture**
* plaintext
* CopyEdit
* Raw Dataset (.csv)
* ↓
* Data Cleaning & EDA (Seaborn, Pandas)
* ↓
* Feature Engineering & Preprocessing
* ↓
* Model Pipeline (Sklearn: Preprocessor + Classifier)
* ↓
* Model Evaluation (Accuracy, F1, AUC)
* ↓
* Deployment-Ready Pipeline (.pkl)

**4. Methodology**

* **Step 1:** Understand business problem and data
* **Step 2:** Exploratory Data Analysis (EDA)
* **Step 3:** Data Preprocessing
* **Step 4:** Model Training and Validation
* **Step 5:** Performance Evaluation
* **Step 6:** Save model pipeline
* **Step 7:** Inference testing

**5. Time Taken**

| **Task** | **Time Estimate** |
| --- | --- |
| Requirement Understanding | 2 hours |
| Data Understanding & EDA | 4 hours |
| Preprocessing & Feature Setup | 2 hours |
| Model Training & Tuning | 2 hours |
| Evaluation & Reporting | 2 hours |
| Documentation & Packaging | 2 hours |
| **Total** | **14 hours** |

**6. Challenges Faced**

* **Imbalanced Target Class**:  
  Needed to handle a large skew toward "no" class in subscription
* **Categorical Explosion**:  
  Many one-hot encoded features from columns like job, education, etc.
* **Model Interpretability vs Accuracy**:  
  Balancing simple models (Logistic Regression) vs powerful black-box ones (Random Forest)

**7. Complexity**

| **Aspect** | **Level** |
| --- | --- |
| Dataset Size | Medium |
| Feature Engineering | Medium |
| Model Complexity | Low to Medium |
| Business Integration | High Impact |

**9. Business Impact**

* Reduced marketing costs by predicting high-conversion leads
* Improved campaign success rate through better targeting
* Enabled data-driven decision-making in marketing operations

**10. Reusability Tips for Future Projects**

* Modularize all training code using Pipeline
* Use ColumnTransformer for preprocessing consistency
* Save model and evaluation reports
* Follow ML lifecycle structure for documentation