Banking Data Analysis Project Report

1. Objective

The goal of this project is to analyze a marketing dataset from a bank to understand customer behavior and predict whether a client will subscribe to a term deposit.

2. Dataset Overview

Dataset: bankmarketing.csv

Source: UCI Machine Learning Repository

Total Records: ~45,000

Features include: age, job, marital, education, default, balance, housing, loan, contact, day, month, duration, campaign, pdays, previous, poutcome, y

3. Steps Performed

Step 1: Data Loading and Overview

Step 2: Data Cleaning

Step 3: Feature Engineering

Step 4: Exploratory Data Analysis (EDA)

Step 5: Handling Categorical Variables

Step 6: Feature Scaling

Step 7: Model Building

Step 8: Model Evaluation

Step 9: Hyperparameter Tuning

Step 10: Model Deployment

Step 11: GitHub Repository Creation

Step 12: Final Presentation & Report

4. Data Cleaning

- Removed duplicates

- Checked null values

- Processed inconsistent categorical entries

- Transformed 'y' to binary labels (Yes/No -> 1/0)

5. Feature Engineering & EDA

- Correlation matrix to find relevant features

- Distribution plots: Age, Target, Education, Job

- Insights: Duration is highly indicative of the target

6. Modeling

- Algorithm Used: Random Forest Classifier

- Evaluation Metrics: Accuracy, Precision, Recall, F1 Score

- Model Accuracy: ~88%

7. Deployment Plan

- Save model using joblib

- Create Streamlit/Flask app

- Deploy on Render for free hosting

- Include prediction form with input fields

8. GitHub Structure

- notebooks/

- src/

- bankmarketing.csv

- model.pkl, encoder.pkl, scaler.pkl

- README.md

- requirements.txt

- .gitignore

9. Visualizations

- Correlation Matrix

- Target Distribution

- Age Distribution

- Job/Education/Marital Status vs Target

10. Conclusion

The Banking Data Analysis project was successful in delivering a robust machine learning model with solid business insights. It showcases data cleaning, visualization, modeling, and deployment skills with a complete GitHub project repository.