

-:Task 03 Report:-

Decision Tree Classification using Bank Marketing Dataset

Submitted By:

Naima Shoukat

Submitted To:

Prodigy InfoTech

Course:

Data Science

Date:

07 August, 2025

■ Table of Contents:

1. Introduction
2. Problem Statement
3. Tools and Technologies
4. Methodology
 - Data Collection
 - Data Exploration
 - Data Cleaning
 - Model Building
 - Evaluation
5. Results and Interpretation
6. Conclusion

1■■ Introduction

In this task, a decision tree classifier is developed using the Bank Marketing dataset. The goal is to predict whether a customer will subscribe to a term deposit based on various demographic and behavioral features.

2■■ Problem Statement

To build a machine learning model using a decision tree to classify customer responses (yes/no) based on marketing data from a Portuguese banking institution.

3■■ Tools and Technologies

- Programming Language: Python
- Libraries: pandas, matplotlib, seaborn, scikit-learn
- Platform: Jupyter Notebook
- ML Algorithm: Decision Tree Classifier
- Dataset Source: UCI Machine Learning Repository (Bank Marketing Dataset)

■ 4.1 Data Collection

- Dataset: Bank Marketing
- Source: UCI Repository
- Target Column: y (yes/no - term deposit)
- Features include: age, job, marital status, education, default, balance, housing, loan, contact, duration, campaign, pdays, previous, poutcome.

■ 4.2 Data Exploration

- Inspected first few rows using `.head()`
- Used `.info()` and `.describe()` for structure and summary
- Checked for missing values and class distribution

■ 4.3 Data Cleaning

- Converted categorical variables to numeric using `LabelEncoder`
- Removed or imputed missing values
- Ensured all features were in proper format for model

■ 4.4 Model Building

- Features (X) and Target (y) defined
- Dataset split into train and test sets
- `DecisionTreeClassifier` initialized and trained

■ 4.5 Evaluation

- Predictions made on test data
- Evaluation metrics: accuracy, confusion matrix, classification report
- Model performance interpreted and visualized

5■■ Results and Interpretation

- Model achieved acceptable accuracy with clear classification metrics
- Confusion matrix showed class prediction performance
- Feature importance can be visualized to interpret model decisions

6■■ Conclusion

■ Observations:

- Decision tree performed effectively on the classification task
- Proper encoding and cleaning improved results
- Important features: duration, contact, poutcome, etc.

■■ Limitations:

- Dataset imbalance could bias model
- Decision trees prone to overfitting without pruning or ensemble methods
- Further improvement using Random Forest or boosting models possible