group memebers

1. Furtuna G/Slassie mit/ur/046/12
2. Betelhem Haftu mit/ur/273/12
3. Meles Mengesha mit/ur/159/12
4. Ayda Angesom mit/ur/298/12
5. Yorkabel Ngatu mit/ur/158/12
6. Meles H/Slassie mit/ur/190/12
7. Makda Haile mit/ur/101/12
8. Yemane G/Michael mit/ur/231/12
9. Michiale Adhanom mit/ur/166/12

Classification Report – Bank Marketing Dataset

This presentation covers building a Random Forest model to predict bank term deposit subscriptions.

M by Meles Haileselassie

Project Introduction

Objective

Predict client subscription to term deposits using supervised learning.

Significance

Enhances bank marketing, reduces contact costs, improves campaign efficiency.

Dataset

From UCI, sourced from Portuguese bank marketing campaigns.

Exploratory Data Analysis (EDA)

Dataset Overview

- 11,162 records
- 16 predictors + 1 target
- Mixed categorical and numerical features

Target distribution slightly imbalanced (·53% no, 47% yes).

Key Observations

- Age range: 18–95, mean ·41
- Skewed balance; some negative
- Contact duration varies widely
- No missing values detected

Most clients contacted by cellular, majority married with secondary education.

Methodology

Data Preprocessing

Label encoding for categoricals; standard scaling of numerical features.

No missing data imputation.

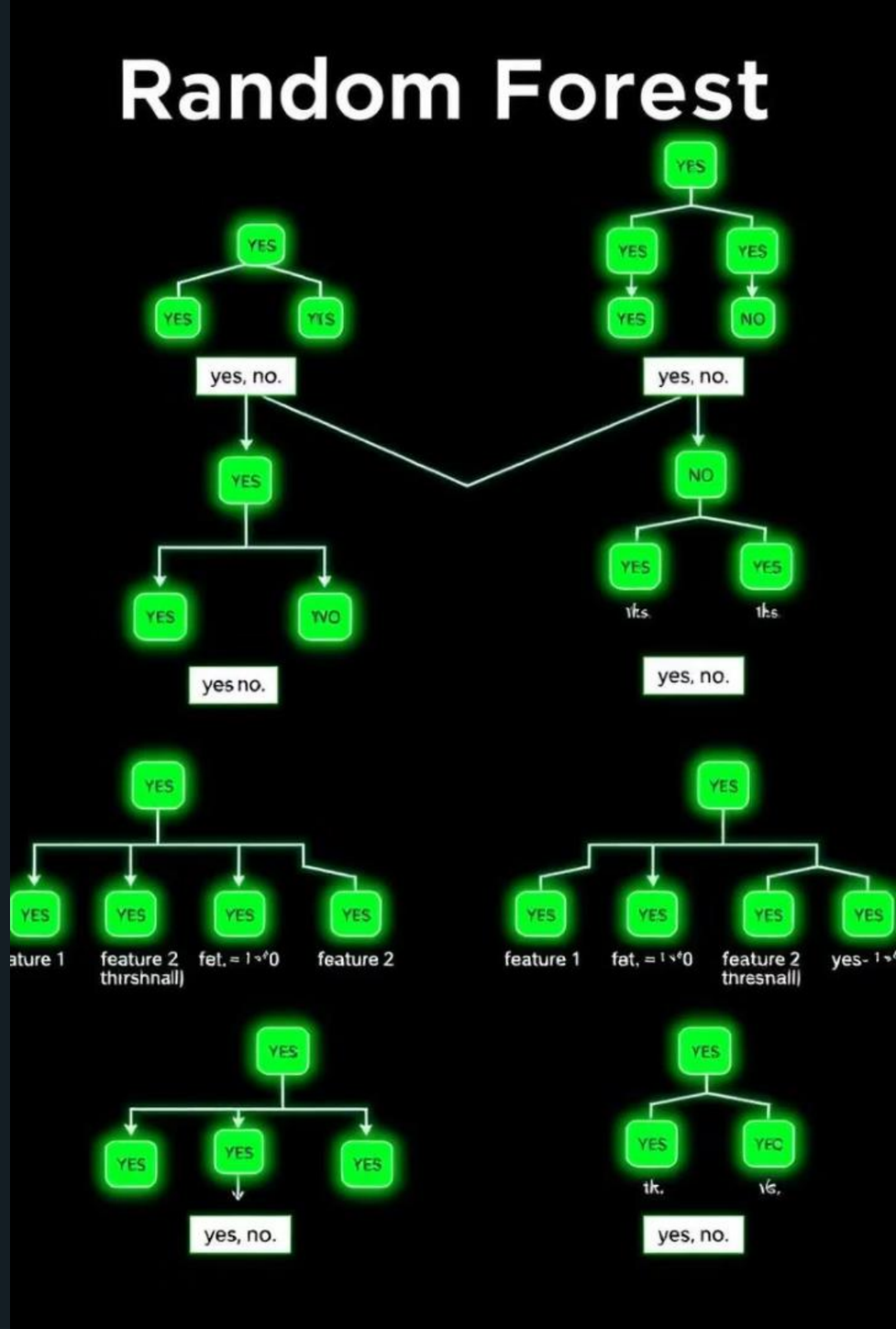
Feature Selection

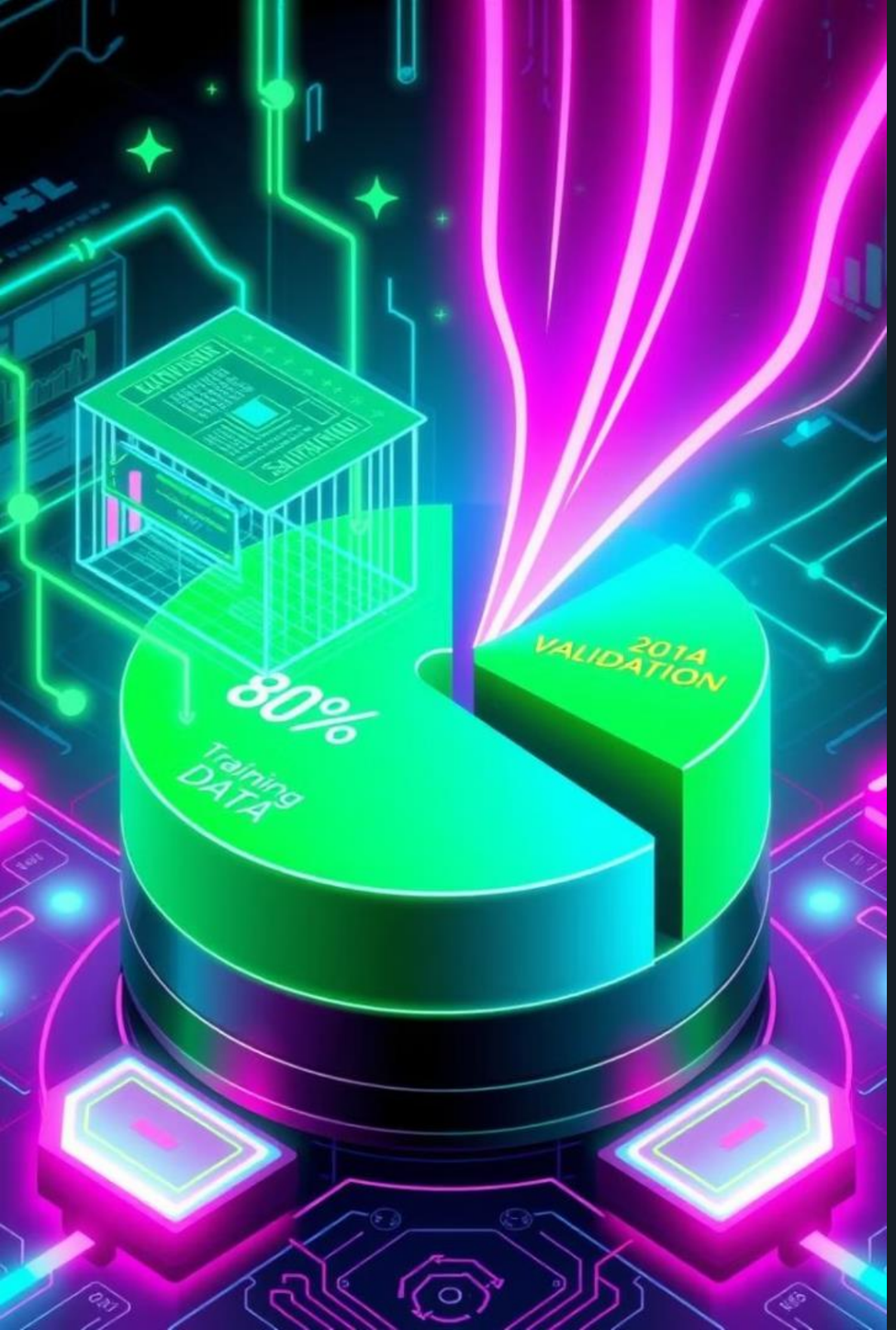
All original features retained except duration excluded for real-time deployment.

Model Choice

Random Forest classifier chosen for robust handling of mixed data and interpretability.

Default 100 trees, no complex tuning.





Model Training

Data Split

80% training, 20% testing with stratified sampling to preserve class distribution.

Reproducibility

Random seed fixed at 42.

Baseline Model

Trained without hyperparameter tuning, providing performance benchmark.

Evaluation and Results

Performance Metrics

- Accuracy: 85.31%
- Precision (yes): 83%
- Recall (yes): 87.%
- F1-score (yes): 85.5 %

Top Features

1. Duration (excluded for deployment)
2. Previous campaign outcome (poutcome)
3. Account balance
4. Days since last contact (pdays)
5. Age

Conclusion and Next Steps

1

Summary

Random Forest model achieved strong accuracy and balanced metrics.

2

Challenges

Feature leakage (duration), mild class imbalance present.

3

Future Improvements

- Hyperparameter tuning and alternative models
- Address imbalance with resampling (SMOTE)
- Exclude duration for deployment scenarios