ADA 442 FINAL PROJECT

Bank Marketing Data Insights And Model Deployment

Exploring strategies and methodologies for effective bank marketing data analysis and model deployment in the financial industry.

Donald Jordy
Deniz Caner Akdeniz
Serdar Topkaya
Azerhan BAĞCI
Presenters

Analysis Objectives

Aims in Predicting Client Subscription to term deposits

Data Source

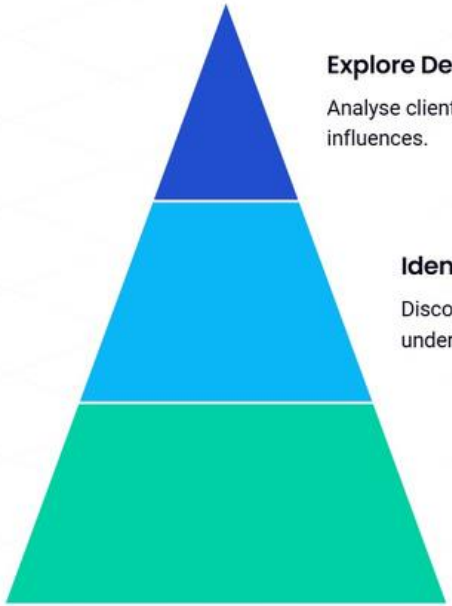
'bank-additional.csv' from UCI ML Repository (4119 instances with 20 features)

Key Features

Age, job type, marital status, education, and economic indicators such as Euribor rates , just to name a few

Target Variable

Whether a client subscribes to a term deposit.



Explore Demographics & Indicators

Analyse client demographics and economic factors to understand influences.

Identify Patterns & Correlations

Discover data patterns and correlations to enhance understanding.

Develop Predictive Model

Create a model to predict client subscription for better marketing strategies.

Exploratory Data Analysis

Key Insights from Our Data Exploration

Inspected variable distribution

Analysed how variables are spread across different values.

Identified target imbalances

Detected imbalances in the target variable distribution.

Utilised visualisation techniques

Implemented various visual tools to explore data.

Histograms

Displayed frequency distribution of variables.

Conducted correlation analysis

Used heatmaps to identify feature relationships .

Data Preprocessing Insights

Key Steps in Data Processing & Engineering

Data Cleaning

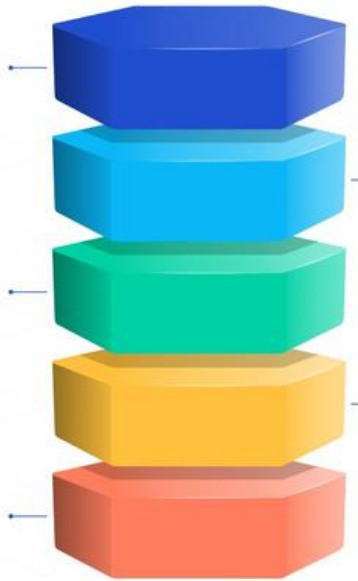
Confirmed no missing values in the dataset, ensuring data integrity.

Handling Imbalances

Used oversampling and undersampling to balance the target variable.

Feature Engineering

Created new features to enhance model insights and predictive power.



Label Encoding

Converted categorical variables into numerical format for ML models.

Feature Scaling

feature selection in order to select the best 15 features

Model Selection Process

Evaluating Performance Metrics

Perfomance Metrics

- Accuracy
- F1- score
- RMSE
- ROC AUC (to address the imbalanced dataset)

Random Forest

Selected for its superior accuracy and stability

Decision Tree

A fundamental model for both classification and regression tasks, offering simplicity, interpretability, and a foundation for more advanced ensemble methods.

XGBoost

a cutting-edge model for achieving high performance in structured data tasks, excelling in both classification and regression problems.

Logistic Regression

A baseline model for binary classification tasks.

Optimising Random Forest

Enhancements to the Random Forest Model

Hyperparameter Tuning

Optimised parameters to enhance model performance using GridsearchCV.

Pipeline Creation & Deployment

Structured pipeline improves training, ensuring workflow consistency. The model is deployed via a Streamlit app, offering ease of use for stakeholders.

Key Findings, Conclusion & Recommendations

Insights from Machine Learning on Marketing

1

Influential Features

Highly correlated features: `pdays` (days since last contact) and `nr.employed` (number of employees).

2

Targeting Strategy

Focus on clients aged 41 with low Euribor 3-month rates for efficient marketing.

3

Enhance Targeted Marketing

Use identified features to improve marketing campaigns.