Bank Term Deposit Subscription Predictor

Title: Predicting Bank Deposit Subscriptions Using Machine Learning

Author: *Rolanda Azeem*
Date: *08/06/2025*

➔ Project Overview

This project focuses on leveraging operational data to support business goals through predictive analytics.The ultimate aim is to enhance the effectiveness of client outreach campaigns by enabling more targeted and informed decision-making.

Objective:

- To develop a machine learning model that predicts whether a client will subscribe to a term deposit, based on historical marketing campaign data.
- Deliver actionable insights to inform marketing strategy.

Scope:

- Clean and process pertinent call campaign data to train and evaluate a model
- Deploy model via a working ML app
- Identify key insights to inform marketing strategy

➔ Data Preparation

Dataset:

- Source: [https://limewire.com/d/x7Hsa#Ml2ucdiZhL]
- Number of observations: *[e.g., 41,188]*
- Target variable: **y** (Subscription: Yes/No)

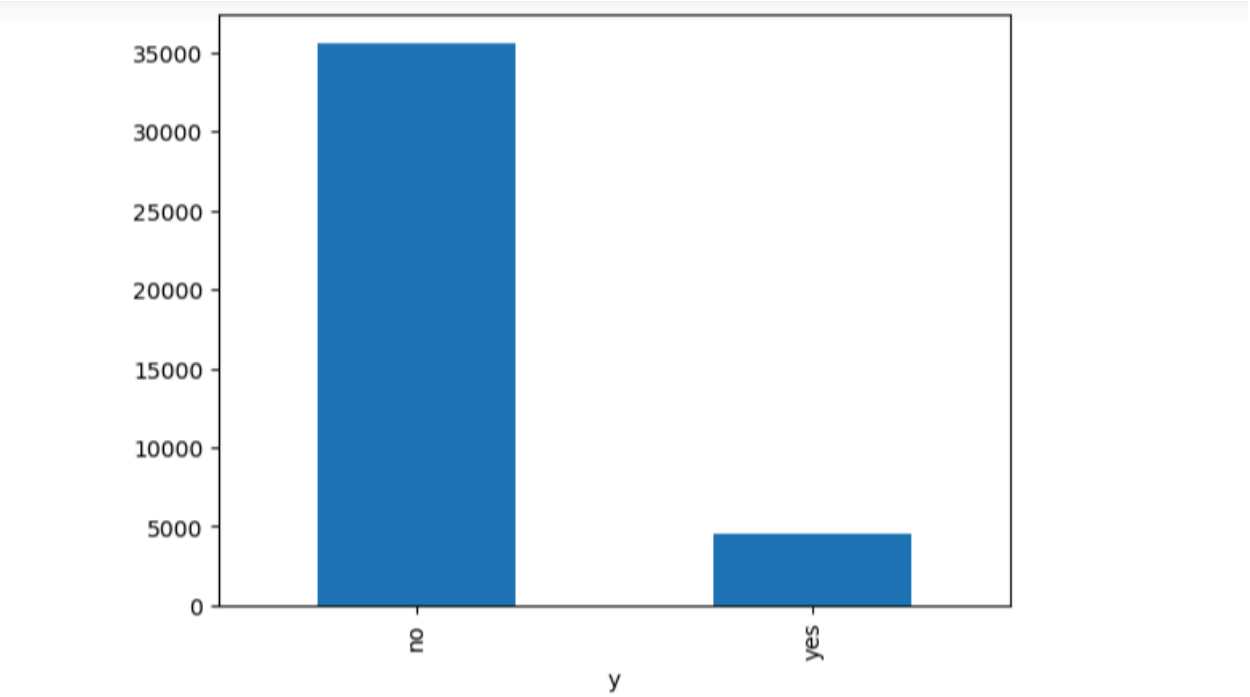
Data Cleaning:

- Removed Unknown entries where plausible (2.6%)
- Grouped redundant categories where possible (education)
- Changed and cleaned up strings where needed (“admin.”, “non-existent ~ unknown”)

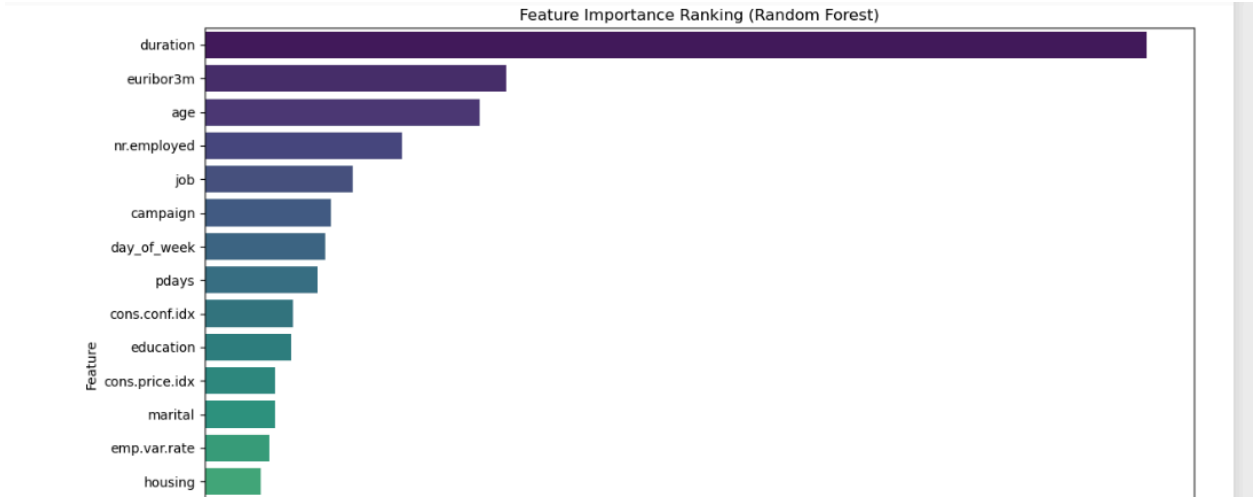
➔ Exploratory Data Analysis (EDA)

Key Findings:

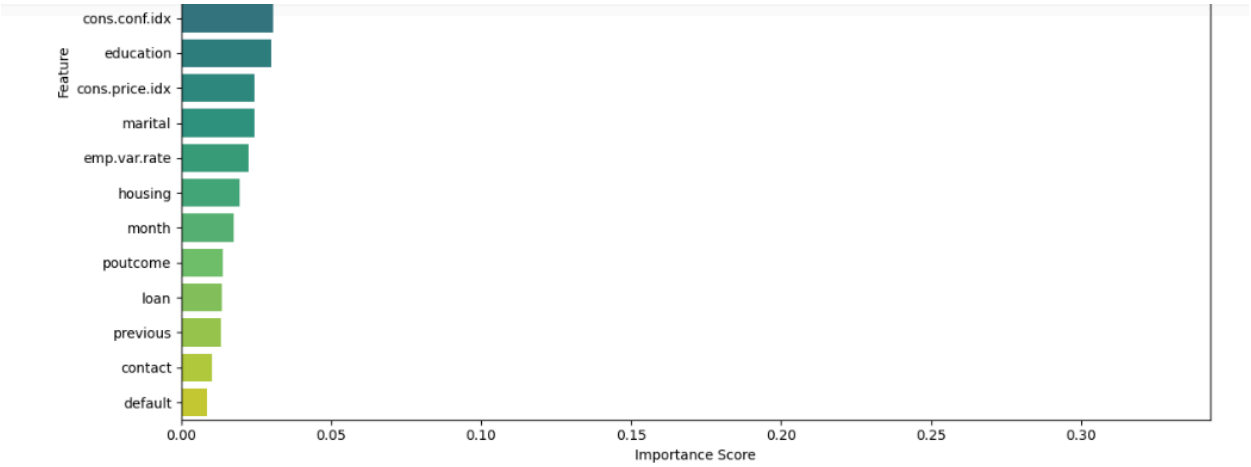
- **Y**: *Target variable shows an imbalance with over 60% of records being no*



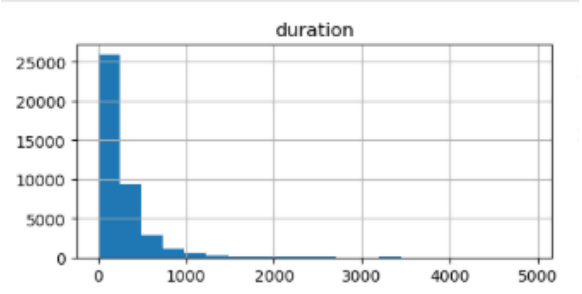
- **Most Important Features**: *Duration, euribor3m , and age*

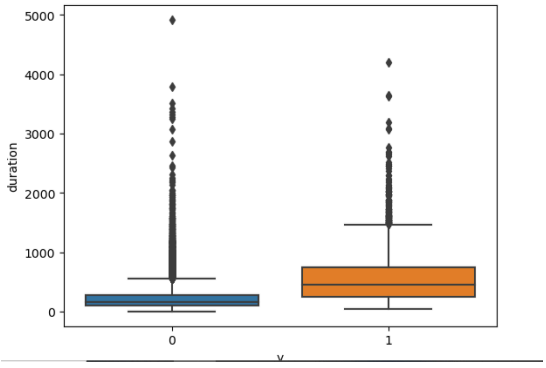


- **Least Important Features**: previous, *contact, default*

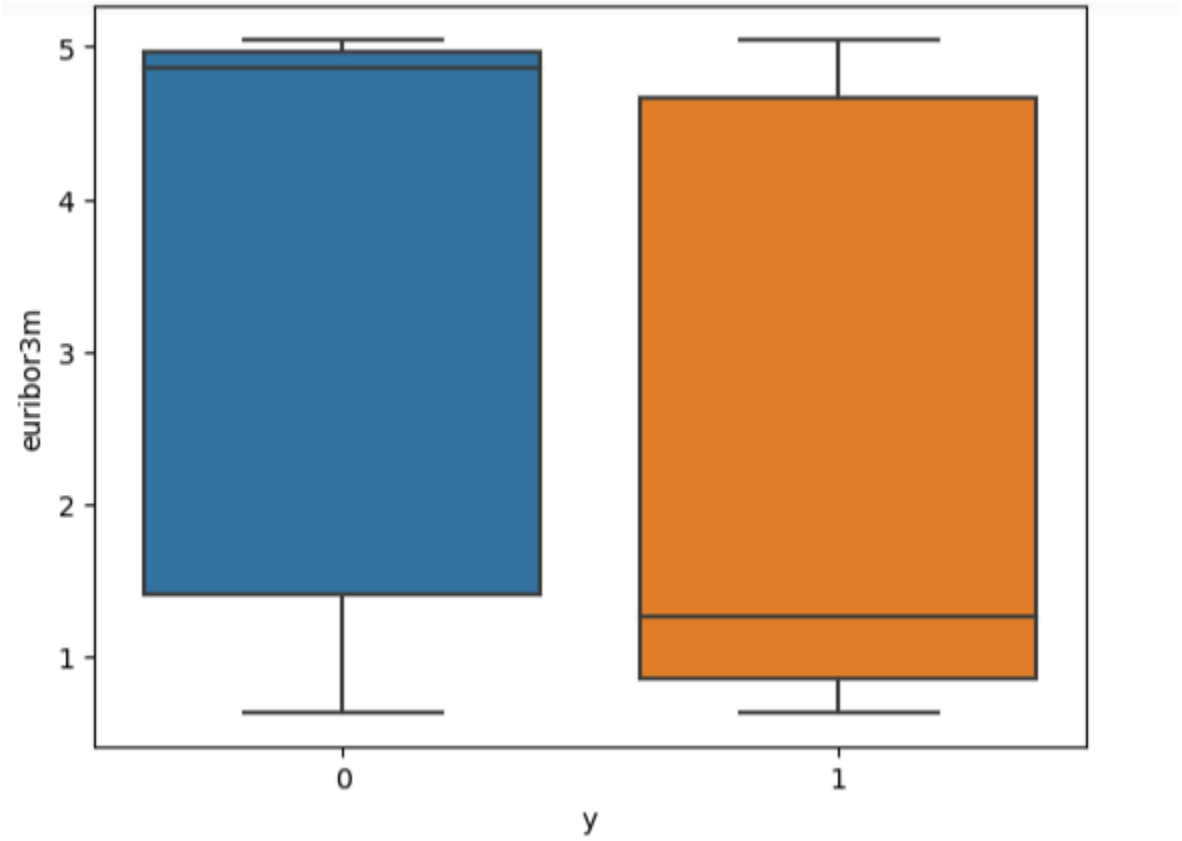


- **Duration**: Has a *positively skewed distribution ; most calls are short ; longer calls positively correlate with subscriptions(yes).*

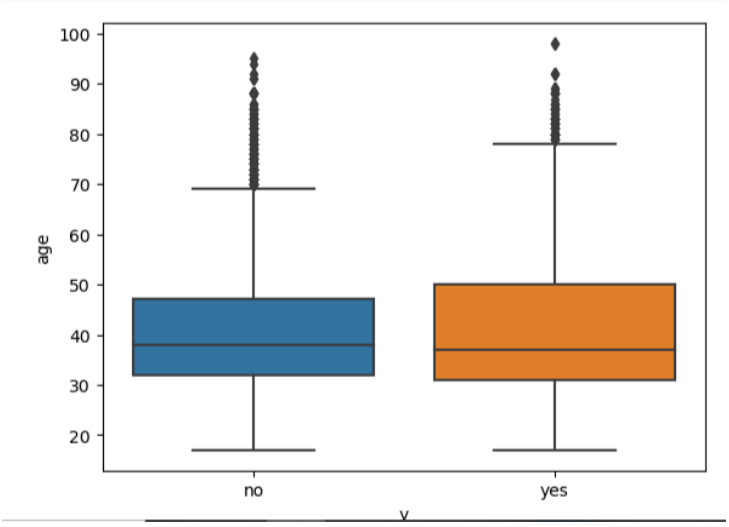




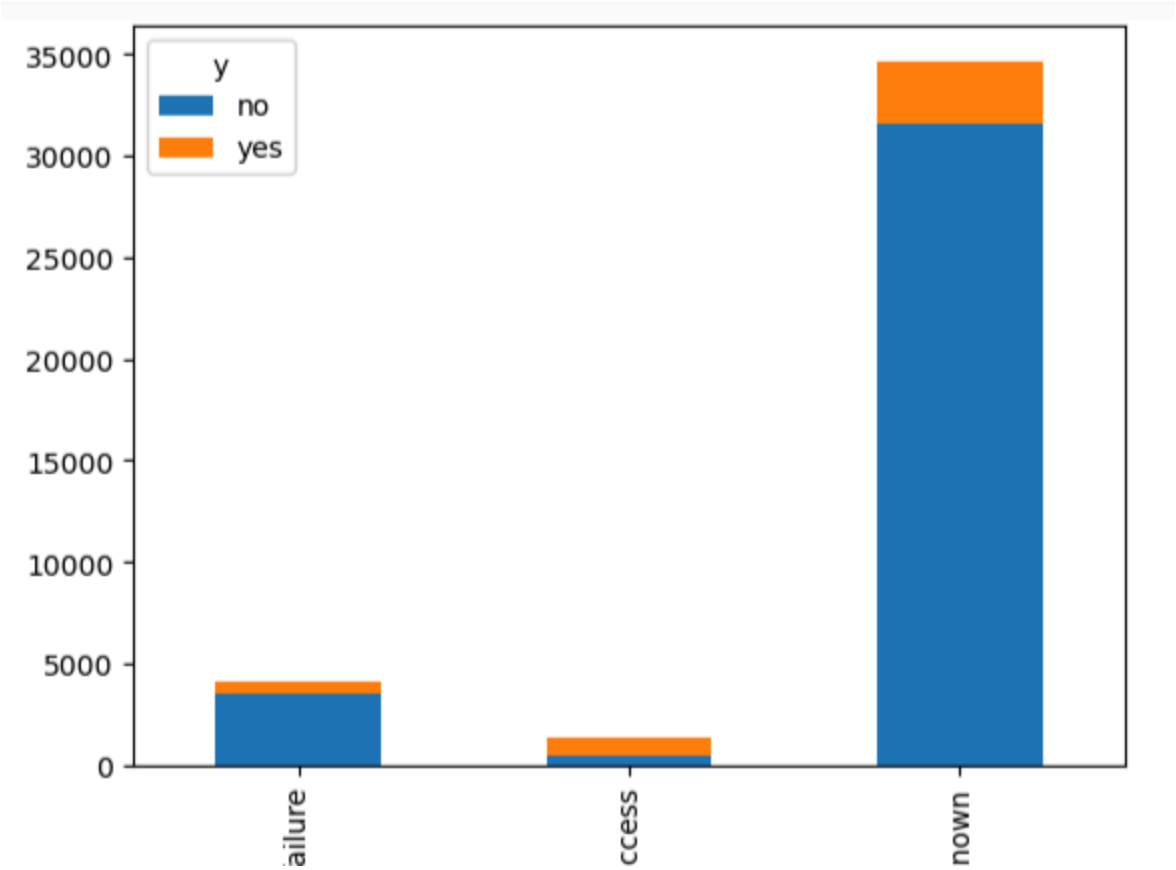
- **Euribor 3m**: *People are likely to subscribe when interest rates are lower possibly valuing stability over higher returns during such periods.*



- **Age**: *Older people (30 - 50) are a little more likely to subscribe to bank term deposits.*



- **Poutcome**: Previously positive *outcomes is a good indicator of future subscriptions(yes).*



➔ Methodology

- Most recent time-order full dataset was split (80% train / 20% test), assuming rows are time bound to prevent leakage
- One-hot encoding for categorical variables
- Standard scaling for numerical variables to prevent large numerals from effecting undue influence.
- Used scale_pos_weight to improve target variable imbalance
- Dropped low- ranked columns: **`default`**, **`contact`**, **`previous`** `to try and improve XGBoost performance`
- Adjusted prediction probability(0.3) to also improve target variable imbalance.
- Selected XGBoost over Logistic Regression though it performed better because XGBoost works better for :
    - imbalance data (scale_pos_weight and threshold tuning)
    - will handle non-linear and complex feature relationships better
    - Will handle the outliers in the database better

➔ Model Performance

Evaluation Metrics:

- Accuracy: *[73.2 %]*
- Precision : *[62 %] of subscriptions*
- *Recall: [33 %] of identifying subscriptions*
- F 1 score: *[73 %]*

➔ Insights & Recommendations

Insights:

- Clients whose calls lasted for longer durations are more likely to subscribe
- Older clients are slightly more likely to subscribe
- Economic indicators such as **`euribor3m`** correlate with decision-making

Recommendations for the Marketing Team:

- Develop highly engaging and possibly unconventional call scripts to keep clients on calls longer.
- Prioritize clients between the ages of 30 - 50 .
- Take advantage of time periods with low interest rates to market extensively.
- Do not neglect previously successful conversions.

➔ Deployment

- Model deployed via Streamlit
- Real-time input via sidebar form
- Outputs likelihood of subscription and probability of subscription

➔ Appendix

- Code repo link: [https://github.com/rolanda4/rolly-bankdeposit]
- Deployed App: [https://rolly-bankdeposit.streamlit.app/]