About Dataset

Data Description:
The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed.

Domain: Banking

Context:
Leveraging customer information is paramount for most businesses. In the case of a bank, attributes of customers like the ones mentioned below can be crucial in strategizing a marketing campaign when launching a new product.

Learning Outcomes:
● Exploratory Data Analysis
● Preparing the data to train a model
● Training and making predictions using an Ensemble Model
● Comparing model performances

Objective:
The classification goal is to predict if the client will subscribe (yes/no) a term deposit (variable y).

Steps and tasks:

1. Import the necessary libraries
2. Read the data as a data frame
3. Perform basic EDA which should include the following and print out your insights at every step.
   a. Shape of the data
   b. Data type of each attribute
   c. Checking the presence of missing values
   d. 5 point summary of numerical attributes
   e. Checking the presence of outliers
4. Prepare the data to train a model – check if data types are appropriate, get rid of the missing values etc
5. Train a few standard classification algorithms, note and comment on their performances along different metrics.
6. Build the ensemble models and compare the results with the base models.
   Note: Random forest can be used only with Decision trees.
7. Compare performances of all the models

References:
● Data analytics use cases in Banking
● Machine Learning for Financial Marketing

1. age

- **Type**: Numerical

- **Description**: Represents the age of the client.
- **Significance**: Age can influence the likelihood of a client subscribing to a term deposit. For example, younger clients may be more willing to save for the future, while older clients may have more financial stability.

---

2. job

- **Type**: Categorical
- **Description**: The type of job the client has (e.g., "admin.", "technician", "blue-collar", "management", etc.).
- **Significance**: Different professions might have varying levels of disposable income and saving habits, affecting their likelihood of subscribing to a term deposit.

---

3. marital

- **Type**: Categorical
- **Description**: The marital status of the client (e.g., "single", "married", "divorced").
- **Significance**: Marital status can influence financial priorities. For instance, married clients might prioritize long-term savings like term deposits compared to single clients.

---

4. education

- **Type**: Categorical
- **Description**: The highest level of education attained by the client (e.g., "primary", "secondary", "tertiary", "unknown").
- **Significance**: Education level often correlates with financial literacy and income, both of which can impact the likelihood of investing in a term deposit.

---

5. default

- **Type**: Categorical
- **Description**: Whether the client has credit in default (e.g., "yes" or "no").
- **Significance**: Clients with a history of default are less likely to invest in term deposits due to financial instability.

---

6. balance

- **Type**: Numerical
- **Description**: The average yearly balance in the client's bank account (in euros).
- **Significance**: Clients with higher account balances are generally more likely to invest in term deposits due to greater disposable income.

---

7. housing

- **Type**: Categorical
- **Description**: Whether the client has a housing loan (e.g., "yes" or "no").
- **Significance**: Having a housing loan might reduce the likelihood of subscribing to a term deposit due to existing financial commitments.

---

8. loan

- **Type**: Categorical
- **Description**: Whether the client has a personal loan (e.g., "yes" or "no").
- **Significance**: Similar to housing loans, personal loans can impact a client's ability to subscribe to a term deposit.

---

9. contact

- **Type**: Categorical
- **Description**: The type of communication used to contact the client (e.g., "cellular", "telephone", "unknown").
- **Significance**: Some communication methods might be more effective than others in influencing a client's decision to subscribe.

---

10. day

- **Type**: Numerical
- **Description**: The last day of the month on which the client was contacted (e.g., 1–31).
- **Significance**: May help in identifying trends based on specific days (e.g., weekends or weekdays).

---

11. month

- **Type**: Categorical
- **Description**: The last month of the year when the client was contacted (e.g., "jan", "feb", "mar", etc.).
- **Significance**: Certain months might have higher subscription rates due to seasonal trends or marketing strategies.

---

12. duration

- **Type**: Numerical
- **Description**: The duration of the last call with the client (in seconds).

- **Significance**: Longer calls might indicate a higher likelihood of subscription, as more time was spent engaging the client.

---

13. campaign

- **Type**: Numerical
- **Description**: The number of contacts performed during the campaign for this client.
- **Significance**: Clients contacted multiple times may either have a higher chance of subscribing or show signs of fatigue, which can be analyzed.

---

14. pdays

- **Type**: Numerical
- **Description**: Number of days since the client was last contacted from a previous campaign (-1 if not contacted previously).
- **Significance**: Helps to determine the effectiveness of re-engaging clients from previous campaigns.

---

15. previous

- **Type**: Numerical
- **Description**: Number of contacts performed before the current campaign for this client.
- **Significance**: Indicates whether frequent engagement leads to better results.

---

16. poutcome

- **Type**: Categorical
- **Description**: The outcome of the previous marketing campaign for this client (e.g., "success", "failure", "unknown").
- **Significance**: A successful outcome in a previous campaign may increase the likelihood of subscribing to the current term deposit.

---

17. Target (or y)

- **Type**: Categorical
- **Description**: The target variable indicating whether the client subscribed to a term deposit or not (e.g., "yes" or "no").
- **Significance**: This is the dependent variable the predictive model aims to classify.

Summary

Each column provides critical insights into client characteristics, campaign details, or the subscription outcome. These attributes will help in:

- Understanding client behavior.
- Analyzing campaign effectiveness.
- Building predictive models for better marketing strategies.