BANK MARKETING ANALYSIS

ANALISI BIVARIATA

Rappresentazione barplot delle variabili categoriche, analisi bivariata tra le variabili categoriche e la Y.

Rappresentazione frequenza delle variabili numeriche, analisi bivariata tra le variabili numeriche e la Y.

FIND OUTLIERS

Trovare gli outliers nelle variabili numeriche.

CORRELATION

Esplora la correlazione tra le variabili numeriche

FEATURE ENGINEERING

Trasforma le variabili con YES/NO in variabili booleane;

Trasforma le variabili categoriche in variabili numeriche (ENCODING).

MODEL SELECTION

regressione logistica

albero decisionale

random forest

APPROFONDISCI TEMATICHE

Classificazione binaria di machine learning supervisionato sugli algoritmi

TRASFORMAZIONE E SCALING DATI (eventuale)

MACHINE LEARNING

Training test – set test

Algoritmi:

regressione logistica

albero decisionale

random forest

studiare accuracy, recall

identificazione miglior algoritmo

DASHBOARD POWER BI

Analisi descrittiva

VARIABILI

1 - **age**: (numeric)  
2 - **job**: type of job (categorical: 'admin.','blue-collar','entrepreneur','housemaid','management','retired','self-employed','services','student','technician','unemployed','unknown')  
3 - **marital**: marital status (categorical: 'divorced','married','single','unknown'; note: 'divorced' means divorced or widowed)  
4 - **education**: (categorical: primary, secondary, tertiary and unknown)  
5 - **default**: has credit in default? (categorical: 'no','yes','unknown')  
6 - **housing**: has housing loan? (categorical: 'no','yes','unknown')  
7 - **loan**: has personal loan? (categorical: 'no','yes','unknown')  
8 - **balance**: Balance of the individual.

Aii. Related with the last contact of the current campaign:

8 - contact: contact communication type (categorical: 'cellular','telephone')  
9 - month: last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')  
10 - day: last contact day of the week (categorical: 'mon','tue','wed','thu','fri')  
11 - duration: last contact duration, in seconds (numeric). Important note: this attribute highly affects the output target (e.g., if duration=0 then y='no'). Yet, the duration is not known before a call is performed. Also, after the end of the call y is obviously known. Thus, this input should only be included for benchmark purposes and should be discarded if the intention is to have a realistic predictive model.

Aiii. other attributes:

12 - campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)  
13 - pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)  
14 - previous: number of contacts performed before this campaign and for this client (numeric)  
15 - poutcome: outcome of the previous marketing campaign (categorical: 'failure','nonexistent','success')

Output variable (desired target):  
21 - y - has the client subscribed a term deposit? (binary: 'yes','no')

Pdays lo tolgo