

Machine Learning - Marketing Campaign Efficacy

Final Project
Columbia University Bootcamp

About the project

Reasons for Topic Selection

- Machine learning models can help improve direct marketing campaigns from the bank to target potential future customers
- This information can be used by other banks or businesses to determine the efficiency of direct marketing campaigns
- The team sourced information that resulted in a Y/N target to understand and learn how data can be used in machine learning to improve revenue outcomes

Analysis & Goals

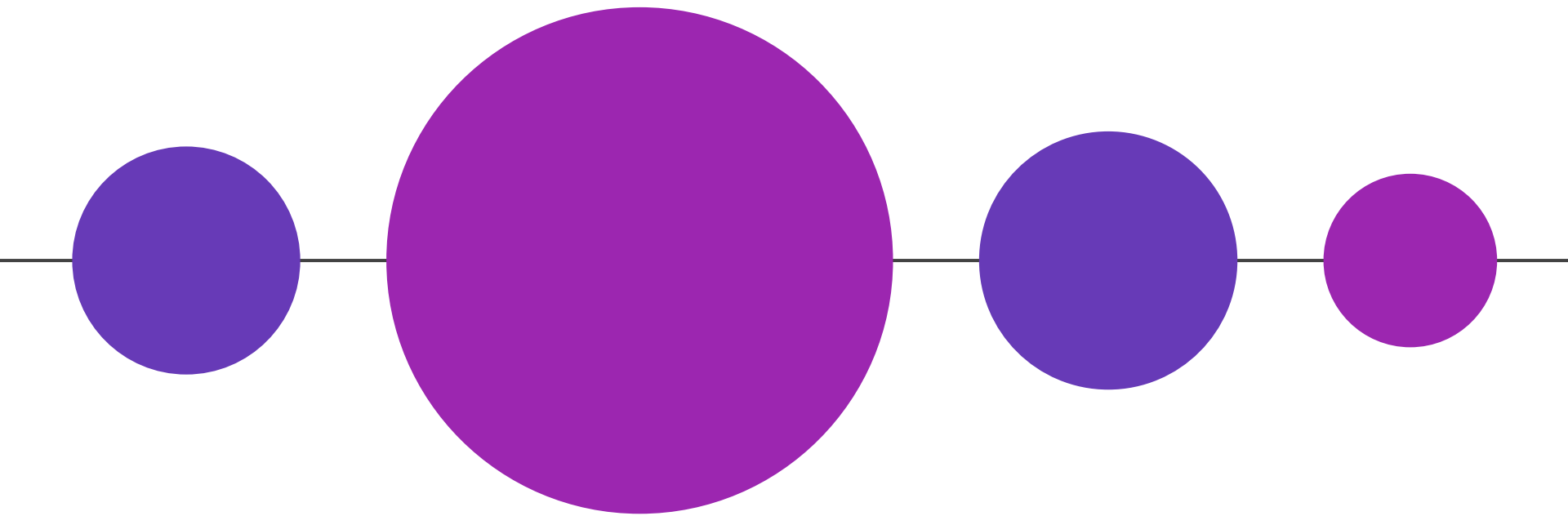
- The classification goal is to predict if the client will subscribe a term deposit.
- Which demographic variables have the most importance while determining whether the client subscribes or not?
- What variables would be most valuable in determining likelihood of subscriptions
- How to determine how to group character descriptions into numerical bins for analysis

Description of the Data Source

- The data is related with direct marketing campaigns of a Portuguese banking institution.
- The marketing campaigns were based on phone calls.
- The dataset we are using is [bank-full.csv](#) with all examples and 17 inputs, ordered by date (older version of this dataset with less inputs).
- Number of instances= 45211 and Number of Attributes=17

About the project : Machine Learning Modeling

- Jupyter Notebook
 - a. Scripts in Jupyter dependant on Pandas and SQL engine relational database
 - — — i. Panda used for preliminary data preprocessing
 - 1. Read in raw .csv data and preprocessed for model input
 - a. Clean the data to remove nulls, irrelevant data columns, convert string to numerical values, create bins of numerical data ranges to reduce noise
 - 2. Feature selection:
 - a. Dropped balance and contact columns
 - b. Converted strings to value
 - c. Grouping
 - i. age_group = ["<18", "18-28", "28-38", ">38"]
 - ii. duration_group = ["<103", "103-180", "180-319", ">319"]
 - 3. Training and Testing Classification
 - a. Training 75%
 - b. Testing 25%
 - ii. SQL engine
 - 1. Postgres : will contain static database for team project use
 - 2. SQL Lite / Alchemy
 - a. Table 1 : Features or dependent variables
 - b. Table 2 : Target or independent variables
- Logistic Regression
- Random Forest Decision Trees



Portfolio samples