1.1. Scope of the project.

When it comes to marketing, historical data should always drive strategy and planning. Predictive analytics is the next level of using that data for marketing success.

Predictive analytics is the use of data, statistical algorithms and AI techniques to identify possible future outcomes. This can help companies stay ahead of the curve and assess the future of their marketing efforts.

Scope of this project was building a predictive Model, using historical data and AI algorithms, able to forecast whether a customer will respond positively, or negatively, to a given Marketing Campaign. Predictive information such as this is, of course, very valuable for organizations, since it can help companies make more sound decisions when it comes to strategic planning, marketing spending, customer and target market segmentation, ROI prediction, and understand the main drivers/reasons that lead to the success (or unsuccess) of a marketing campaign.

1.2. Data Understanding.

The dataset contains information about a tele-marketing campaign from a Portuguese banking institution. Purpose of this campaign was to prompt their clients to subscribe for a specific financial product (Term Deposit). After each call, clients were asked their intentions of either subscribing to the product (indicating a successful campaign) or not (unsuccessful campaign).

The dataset had 41,188 instances of calls to clients (rows) and 21 variables (columns): 11 object-types (including the Target Variable) and 10 numeric-types, 5 floats and 5 integers. In some cases, same client was contacted multiple times, although each call was considered independent from another even if the client was the same.

- Variables can be divided into following groups:
  1. Customer-related variables:
  a. age: client's age (range: 17 - 98)
  b. job: client's type of job (12 categories)
  c. marital: marital status (4 categories: 'married', 'single', 'divorced', 'unknown')
  d. education: level of education (8 categories: 'illiterate', 'basic.4y', 'basic.6y', 'basic.9y', 'high.school', 'professional.course', 'university.degree', 'unknown')
  e. default: if the client has credit in default (3 categories: 'no', 'unknown', 'yes')
  f. housing: if the client has a mortgage on his/her house (3 categories: 'no', 'unknown', 'yes')
  g. loan: if the client has a personal loan (3 categories: 'no', 'unknown', 'yes')

2. Variables related to last contact (current campaign):

a. contact: type of communication (2 categories: 'telephone', 'cellular')

b. month: month of last contact (10 categories)

c. day_of_week: day of last contact (5 categories)

d. duration: call duration, in seconds (range: 0 - 4,918 (more than 1 hour and 20 minutes))

e. campaign: number of contacts performed during this campaign and for this client (range: 1 - 56)

3. Other marketing-related variables:
a. pdays: number of days from last contact (range: 0 - 999)
b. previous: number of contacts performed before this campaign and for this client (range: 0 - 7)
c. poutcome: outcome of previous marketing campaign (3 categories: 'nonexistent', 'failure', 'success')

4. Socioeconomic variables:

a. emp.var.rate: employement variation rate - quarterly indicator (range: (-3.4) - 1.4)

b. cons.price.idx: consumer price index - monthly indicator (range: 92.201 - 94.767)

c. cons.conf.idx: consumer confidence index - monthly indicator (range: (-50.8) - (-26.9))

d. euribor3m: euribor 3 month rate - daily indicator (range: 0.634 - 5.045)

e. nr.employed: number of employees - quarterly indicator (range: 4963.6 - 5228.1)

2. Methodology.

In order to build an efficient Classification model, with focus on prediction, I followed the below methodology:

- Defined the scope of the project.
- Selected an analytic, predictive, machine-learning driven approach.
- Data understanding: content and format analysis.
- Data Exploration:
- Data Cleaning
- Exploratory Data Analysis for Feature Selection.
- Data Preparation for modeling
- Model Development:
- Classifier 1: K-Nearest Neighbors (KNN)
- Classifier 2: Random Forest
- Classifier 3: Support Vector Machines
- Classifier 4: Ensemble Models - Stacking Classifier
- Model Evaluation: compare performance of different models, and select the best for the project purpose.