

Data Science Intern at Data Glacier

Project: Bank Marketing (Campaign)

Week 8: Deliverables

- **Group Name:** LimitLess Team
- **Team member's details:**
 1. **Name:** Chitra S Chaudhari
Email: chitraksonawane@gmail.com
Country: USA
Company: Create & Learn LLC, GrandCircus Detroit
Specialization: Data Science
 2. **Name:** Nuri Öztürk
Email: ozturknuri8@gmail.com
Country: Turkey
Specialization: Data Science
- **Batch Code:** LISUM 17
- **Date:** 17 February 2023
- **Submitted to** Data Glacier
- **Github Repo link:**
<https://github.com/ChitraChaudhari/Bank-Marketing-campaign->

Problem description

ABC Bank wants to be able to predict which clients are most likely to subscribe to a term deposit. In this way, the bank wants to save time and money by running the marketing campaign more effectively and successfully.

Data understanding

In this project, we have been given 4 datasets:

- Bank
- Bank-full
- Bank-additional
- Bank-additional-full

We started data exploration with the bank dataset, which is the smallest one. And as we examined the data, we decided to drop two features that did not contain specific information for the target column. You can find more information about the features below. The features "default" and "poutcome" were removed. There are no lower outliers in the dataset and the upper outliers in the "balance" and "campaign" features were also removed. For "age" and "duration" features, we applied a logarithmic transformation to eliminate outliers. Our focus now is to understand the information contained in other datasets with additional features and prepare the bank dataset for model building. After creating our first model with the bank dataset, we will try to improve our model with other datasets.

Attribute information:

- Input variables:
 - Bank client data:
 - age (numeric)
 - job: type of job (categorical: "admin.", "blue-collar", "entrepreneur", "housemaid", "management", "retired", "self-employed", "services", "student", "technician", "unemployed", "unknown")
 - marital: marital status (categorical: "divorced", "married", "single", "unknown"; note: "divorced" means divorced or widowed)

- education (categorical: "basic.4y", "basic.6y", "basic.9y", "high.school", "illiterate", "professional.course", "university.degree", "unknown")
- default: has credit in default? (categorical):
 - 98.2% No
 - 1.8% Yes
 - 0% Unknown
- housing: has a housing loan? (categorical: "no", "yes", "unknown")
- loan: has a personal loan? (categorical: "no", "yes", "unknown")
- Related to the last contact of the current campaign:
 - contact: contact communication type (categorical: "cellular", "telephone")
 - month: last contact month of the year (categorical: "jan", "feb", "mar", ..., "nov", "dec")
 - day_of_week: last contact day of the week (categorical: "mon", "tue", "wed", "thu", "fri")
 - 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 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.
- Other attributes:
 - campaign: number of contacts performed during this campaign and for this client (numeric, includes the last contact)
 - pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means the client was not previously contacted)
 - previous: number of contacts performed before this campaign and for this client (numeric)
 - poutcome: outcome of the previous marketing campaign (categorical)
 - 81.7% Unknown
 - 10.8% Failure
 - 4.1% Other
 - 3.3% Success
- Social and economic context attributes:
 - emp.var.rate: employment variation rate - quarterly indicator (numeric)
 - cons.price.idx: consumer price index - monthly indicator (numeric)

- cons.conf.idx: consumer confidence index - monthly indicator (numeric)
- euribor3m: euribor 3-month rate - daily indicator (numeric)
- nr.employed: number of employees - quarterly indicator (numeric)
- Output variable (desired target):
 - y - has the client subscribed to a term deposit? (binary: "yes", "no")