# MINDD TP1

# Business Understanding

## Objective

The objective of the project is to build data-driven models to predict the success of telemarketing calls aimed at selling long-term deposits.

In this context, by applying data mining, it is expected to identify influencing factors related to customers and campaigns that can significantly improve business decisions.

Marketing is targeted at a specific segment of customers of a certain type. The purpose of this project is to develop a model, using the CRISP-DM methodology, that can identify and prioritize these customers

## The data

The data is related to direct marketing campaigns of a banking institution.

The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required

to access if the product (bank term deposit) is (or not) subscribed.

It consists of:

* Attribute information bank client data
* Social and economic context attributes
* Other attributes (Related to the contacts made with the client prior and during campaigns)

## What is a term deposit

"A term deposit is a savings' tool where money is deposited into an account at a financial institution. Term deposit investments usually have short-term maturities ranging from one month to a few years, have varying levels of required minimum deposits, and pay a fixed interest rate to the investor."

Quoted from “Term Deposit: Definition, How It's Used, Rates, and How to Invest” (https://www.investopedia.com/terms/t/termdeposit.asp), James Chen.

## Determining factors for individual customers to opt for term deposits

"... there are 5 groups of factors that greatly affect the decisions of individual customers to choose a savings' bank: Service quality, Safety, Related effects, Benefits financial benefits, convenience"

Tuan, L. A., Nhu, M. T. Q., & Nhan, N. le. (2021). Factors Affecting the Decision of Selecting Banking to Save Money of Individual Customers – Experimental in Da Nang City. Advances in Science, Technology and Engineering Systems Journal, 6(3), 409–417. https://doi.org/10.25046/aj060345

In other words, we can expect higher conversion rates if these factors are present. It would therefore be useful to analyze the data to determine whether these elements can be identified and quantified.

## Determining factors for telemarketing success

"High-quality, accurate customer data and well-targeted segments are key drivers of telemarketing success."

ICTSD. (2021). Factors affecting telemarketing productivity. International Centre for Trade and Sustainable Development. https://www.ictsd.org/unraveling-productivity-challenges-in-the-telemarketing-department

"Timing calls appropriately and implementing consistent follow-ups significantly improve customer engagement and conversion rates."

Tuan, L. A., Nhu, M. T. Q., & Nhan, N. L. (2018). Factors influencing customer purchasing behavior in telemarketing. So09.tci-thaijo.org. https://so09.tci-thaijo.org/index.php/PMR/article/view/5570

We can also expect higher conversion rates if these factors are present. It would be useful to analyze the data and determine whether these elements can be identified and quantified.

# Plan

## I. Collect data

### II. Clean Data

As mentioned in bank-information.txt, there is missing client categorical information that is defined as "unknown".

If these values are not present in a great part of the data, the clients that have missing data should be excluded because they would either give unreliable information about the "unknown" group, or if they are absorbed into other groups, it would also not provide reliable information.

It must be also investigated if there is abnormal data and also the conversion of data to proper numeric values and units.

### III. Explorative Data Analysis

* Understand the relation between properties. Try to unravel groups of customers and how they behave.
* Investigate the relation between conversion rates and the various factors
* Determine the social economic factors and how they sway the conversion rates.
* Determine the quality of the telemarketing campaign and how it affects conversion rates (converstion into long term deposits).

### IIII. Create a classification machine learning model

Create a model using various algorithms to fit the data into the best possible solution

## II. Data Understanding

**Collect initial data:** Acquire the necessary data and (if necessary) load it into your analysis tool.

**Describe data:** Examine the data and document its surface properties like data format, number of records, or field identities.

**Explore data:** Dig deeper into the data. Query it, visualize it, and identify relationships among the data.

**Verify data quality:** How clean/dirty is the data? Document any quality issues.

### Columns Identified

|  |  |  |  |
| --- | --- | --- | --- |
| # | Attribute | Description | Type |
| 1 | **age** | Client’s age | Numeric |
| 2 | **job** | Type of job (admin., blue-collar, entrepreneur, housemaid, management, retired, self-employed, services, student, technician, unemployed, unknown) | Categorical |
| 3 | **marital** | Marital status (divorced, married, single, unknown) — note: “divorced” includes widowed | Categorical |
| 4 | **education** | Education level (basic.4y, basic.6y, basic.9y, high.school, illiterate, professional.course, university.degree, unknown) | Categorical |
| 5 | **default** | Has credit in default? (no, yes, unknown) | Categorical |
| 6 | **housing** | Has housing loan? (no, yes, unknown) | Categorical |
| 7 | **loan** | Has personal loan? (no, yes, unknown) | Categorical |
| 8 | **contact** | Type of contact communication (cellular, telephone) | Categorical |
| 9 | **month** | Last contact month of year (jan–dec) | Categorical |
| 10 | **day\_of\_week** | Last contact day of the week (mon–fri) | Categorical |
| 11 | **duration** | Duration of last contact (in seconds).  Strongly correlated with target; should **only be used for benchmark models**, not for real predictive deployment. | Numeric |
| 12 | **campaign** | Number of contacts performed during this campaign for this client (includes last contact) | Numeric |
| 13 | **pdays** | Days passed since last contact in a previous campaign (999 = client not previously contacted) | Numeric |
| 14 | **previous** | Number of contacts performed before this campaign | Numeric |
| 15 | **poutcome** | Outcome of the previous marketing campaign (failure, nonexistent, success) | Categorical |
| 16 | **emp.var.rate** | Employment variation rate (quarterly indicator) | Numeric |
| 17 | **cons.price.idx** | Consumer price index (monthly indicator) | Numeric |
| 18 | **cons.conf.idx** | Consumer confidence index (monthly indicator) | Numeric |
| 19 | **euribor3m** | Euribor 3-month rate (daily indicator) | Numeric |
| 20 | **nr.employed** | Number of employees (quarterly indicator) | Numeric |
| 21 | **y** | Target variable: Has the client subscribed to a term deposit? (yes, no) | Binary (Target) |