



## **Bank Marketing Data Set**

**CPIT 440**

**Data Mining and Warehousing**

**Prepared by: Najat Ali**

**Prepared for: Hind Almisbahi**

# Phase 1:

## ❖ Key words:

Deposit: refers to money or assets held at a bank. When a customer makes a deposit, they place money in the bank. The bank holds the money for the customer for a set amount of time under certain conditions.

## ❖ Problem Goal:

The data is related with direct marketing campaigns (phone calls) of a Portuguese banking institution. The classification goal is to predict if the client will subscribe to a term deposit.

## ❖ Data Mining Function:

Classification

## ❖ Data type:

Multivariate (string and numeric)

- The attributes (age-balance-day-duration-campaign-pdays-previous) are numeric.
- The other attributes (job-marital-education-default-housing-loan-contact-month-poutcome-y) are string.

## ❖ Data Splits:

Data is split for testing, and the test dataset consists of 10% of the examples and 17 inputs randomly selected.

## ❖ Data Source:

UCI Machine Learning Repository:

<https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>

❖ Related Paper Summarization:

A Data-Driven Approach to Predict the  
Success of Bank Telemarketing

By

S. Moro, P. Cortez and P. Rita.

They propose a data mining (DM) approach to predict the success of telemarketing calls to sell bank long-term deposits. In this paper, a semi-automatic feature selection mechanism was explored in the modeling phase, performed with the data prior to July 2012 and that allowed to select a reduced set of 22 features.

They compared four DM models: logistic regression, decision trees (DT), neural network (NN) and support vector machine. Two metrics were used to compare these models, both at the modeling and rolling window evaluation phases: the area of the receiver operating characteristic curve (AUC) and the area of the LIFT cumulative curve (ALIFT). In both metrics and phases, the most effective results were obtained by the NN, which resulted in an AUC of 0.80 and ALIFT of 0.67 during the rolling windows evaluation.

Accordingly, by selecting only the most likely buyers, the proposed DSS (Decision Support System) improves the efficiency of the bank's telemarketing campaigns (i.e., reducing client intrusiveness and contact costs).

## Phase 2:

### ❖ Problem statement:

Direct marketing campaigns (phone calls) by a Portuguese banking institution are the subject of this data set. Using classification, the goal is to increase term deposit subscriptions by predicting if the client will subscribe or not.

### ❖ Data exploration and visualization that are submitted in the proposal:

To improve understanding of the data, we use a lot of visualization methods in the code part, specifically phase 1.

Data preprocessing: we check for duplication, split the data into train and test sets, handle missing values, outliers, normalization and handle categorical features.

### ❖ Models Training: We use four models to predict the target attribute.

Due to the use of some of them in the dataset paper, they are used here:

- Logistic Regression,
- Decision Tree
- Support vector machine
- Random Forest

### ❖ Models' evaluation:

We used accuracy, precision, and recall performance metrics for evaluating the models.

	Model	Accuracy	Precision	Recall
0	Logistic Regression	0.897822	0.618081	0.318441
1	Decision Tree	0.869402	0.443471	0.480989
2	Support Vector Machine	0.903129	0.443471	0.321293
3	Random forest	0.905783	0.670648	0.373574

- Random forest has the most accuracy, precision, and recall performance metrics compared to other models

❖ **Tools:**

Jupyter: notebook environment

The language: python

Libraries:

- pandas
- Numpy
- Matplotlib
- Sklearn
- Seaborn
- 

- ❖ **Difficulties you have faced and challenges:** The time of the project is coming with all other projects exams and events.

- ❖ **Future work:** enhance the modeling part by using NN\_model which is used by the paper as the best model.

❖ Reference:

- (2014) *A data-driven approach to predict the success of bank telemarketing*, *Decision Support Systems*. Available at:  
<https://www.sciencedirect.com/science/article/abs/pii/S016792361400061X>