**Fundamentals of Data Analytics - ITC-5203**

**Final Course Project**

Report Deadline: Friday April 08, 2024

Bank Marketing

The problem is related to direct marketing campaigns (phone calls) of a Portuguese banking institution. The **classification goal** is to predict whether the client subscribes a term deposit or not. The target class is the last attribute (*subscribed*) and has two values (*yes* and *no*).

The training set (trainset.csv) contains 3,196 subscribed and 26,076 unsubscribed records. The test set (testSet.csv) contains 1,444 subscribed and 1,047 unsubscribed records.

**Attribute Information:**

1. **age** (numeric)
2. **job**: type of job (categorical: 'admin.','blue-collar', 'entrepreneur', 'housemaid', 'management', 'retired','self-employed', 'services', 'student', 'technician', 'unemployed', 'unknown')
3. **marital**: marital status (categorical: 'divorced','married','single','unknown'; note: 'divorced' means divorced or widowed)
4. **education**: (categorical: 'basic.4y', 'basic.6y', 'basic.9y', 'high.school', 'illiterate', 'professional.course', 'university.degree', 'unknown')
5. **housing**: has housing loan? (categorical: 'no','yes','unknown')
6. **loan**: has personal loan? (categorical: 'no','yes','unknown')
7. **contact**: contact communication type (categorical: 'cellular','telephone')
8. **month**: last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
9. **day\_of\_week**: last contact day of the week (categorical: 'mon','tue','wed','thu','fri')
10. **duration**: last contact duration, in seconds (numeric).
11. **campaign**: number of contacts performed during this campaign and for this client (numeric, includes last contact)
12. **pdays**: number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)
13. **poutcome**: outcome of the previous marketing campaign (categorical: 'failure','nonexistent','success')
14. **nr.employed**: number of employees - quarterly indicator (numeric)
15. **Target Attribute**: Subscribed - has the client subscribed a term deposit? (binary: 'yes','no')

# Steps

The project involves the following steps:

1. **Data exploration**: try to know data and represent statistics for the important features among the features and the target attribute.
2. **Preprocessing the data**. The goal of this step is to extract features from records in the training set and use these features to test data sets. Note that the data have “**unknown**” values that need to be cleaned.
3. **Use a classification-learning method** provided by **python** to learn a model from the set of training examples. You can use any of the classification methods for this purpose.
4. **Test the learned model** on the test set and report the testing results.

# What to hand in:

1. Your programs (**Python**) for this project.
2. A **report** that contains:
   1. The objective of the project (you may use an introduction section to describe it).
   2. What learning method(s) you use to learn the model(s), and the testing results of the learned model(s) on the test data (such as classification accuracy or ROC curve).
   3. Any discussion and conclusion that you find during from the project.
3. Slides for **10 minutes presentation**
   1. There will be a presentation at the last lecture and each group must present their project.

# How to hand in:

* **Presentation:** You must submit your presentation by Monday April 8th at 11:59pm. You will present your project in class. The presentation schedule will be announced later.
* **Report and Python program:** You must upload your Python program and your report (as a PDF file) to Blackboard by Friday April 08.

# Marking Scheme (25 points)

Your project mark will consist of the following components:

* Your presentation (7 points)
* Clearness and organization of your report (10 points)
* Soundness and correctness of your solution (as described in your report and implemented in your programs if any) (8 points).

**No late submission is allowed.**