**Advance Deep Learning**

**Final Course Project**

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. **Develop and train at least two Deep-learning methods** provided by **Python** to learn a model from the set of training examples. Compare both the models based on
   * Evaluate each model’s performance on the test set using the following metrics:

▪ Accuracy

▪ Precision

▪ Recall and F1 Score

* + Use a confusion matrix to visualize each model’s performance

1. **Test the learned model** on the test set and report the test results.

# What to hand in(Only one submission from each group):

1. Your programs (**Python) both pdf and ipynb** for this project.
2. Professional Report (PDF format): A structured report that includes:
   1. Introduction: Overview of the project, goals, and significance of analysis.
   2. Dataset Description and Preprocessing: Summary of the dataset, data preprocessing steps, and the rationale behind each choice.
   3. Model Descriptions: Explanation of models, including architecture and hyperparameter choices.
   4. Results and Findings: Detailed results from training, validation, and testing, presented in tabular and graphical formats.
   5. Analysis and Discussion: Interpretation of results, comparison of model strengths and weaknesses, and insights gained from analysis.
   6. Conclusion: Summary of key findings, final remarks on model effectiveness, and possible future improvements.
   7. References: Citations for any external resources, papers, or tools referenced in the project.
3. Slides for **10-15 minutes presentation**
   1. There will be a presentation at the last lecture and each group must present their project.
4. Group Responsibility Breakdown o A brief outline indicating each team member’s role and contributions, from data preprocessing to model analysis.

# How to hand in:

* **Presentation:** You must submit your presentation. 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 as well as ipynb) to Blackboard.

# Marking Scheme (35 points)

Your project mark will consist of the following components:

* Your presentation (10 points)
* Clearness and organization of your report (15 points)
* Soundness and correctness of your solution (as described in your report and implemented in your programs if any) (10points).

**No late submission is allowed.**