# DSA5101 Python Machine Learning Project

Li Xiaoli

# Data Set

- For our group project, you will use Bank Marketing data to predict whether a client will subscribe a term deposit. readme.txt file contains some useful information about the data
- You are required to use **trainiingdata.txt** as training data to build your machine learning models, and use **testdata.txt** to evaluate the model performance.
- You should implement at least 3 different machine learning (classification) algorithms (not necessarily taught as class), and then compare them to report their performance.
- Note if you want to perform oversampling, you should only do them on the training set. You should not do oversampling on the combined training and test data, as they will generate over-optimistic performance.
- You can perform feature engineering (include generate new features and select part of features to build models).

# Problem Statement and Performance Evaluation

- The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed.
- You should report precision, recall, F-measure for yes class and overall accuracy.

# Group & Submission

- **Group:** Each group has around ~4 students – forming by yourself.
- **Submission**
  - Each group selects one student as representative to submit your zipped files (including 1) presentation slides with your names and matriculation numbers), 2) well-documented codes, 3) recorded Audio/video presentations where anyone in the group need to present parts of slides, with total presentation duration around 15 mins) and upload to Luminus\Files\ Assignments/Projects/Python ML Projects_Submission by **Oct 1 (11.59 pm)**
  - **All students in the same group should work together and will get the same scores (unless other group members complain you do not make good contributions)**

# Suggested Topics for Project Slides

- Simple Data Set and Problem Statement Description.
- Dataset Pre-processing (including data exploration & visualization, feature engineering,  feature selection, etc.) .
- Experimental Study and Analysis.
- Summary of Project Achievements (including the insights from the project, e.g., feature importance analysis, how to use the prediction results for business).
- Future Directions for further Improvements.

# Thank You

Contact: xlli@i2r.a-star.edu.sg if you have questions