

# CS5101 Machine Learning Lab (Aug - Dec 2021)

## Assignment 9

---

### Rules:

1. You must submit your code in a single python .ipynb notebook with the naming format as follows: Firstname Lastname assignment9.ipynb
  2. Submit the .csv file mentioned in the following naming format: Firstname\_Lastname.csv
  3. For each question, create a separate text block containing the question followed by a code block containing the solution.
  4. Your code must be properly commented explaining each step clearly.
  5. If any of the above instructions are not followed, the penalty will be there for the same.
  6. Your code and answers will be checked for plagiarism and if found plagiarised, zero marks will be provided for assignment. So make sure you actually code and solve the questions rather than noting down the answers.
  7. NOTE: The total mark for this assignment is 5 marks.
  8. Deadline for submission is 8-Nov 23:00 IST
- 

### Questions:

1. You are provided with the following data files:
  - a. train.csv
  - b. test.csv

[3 marks] Learn a model which can classify given data correctly. This is a multiclass data imbalance problem with high dimensions data. You can see the classes are named as n, p1, p2, p3,...So the classes as n are negative ones and with p in it as positive.

You may need to include **more than one type of classifier** in your model and do **some preprocessing** to remove un important features.

- Submit your model diagram along with hyper parameters in a pdf.
  - Submit your rational for such model.
  - Submit your predictions for the test.csv file as mentioned in rules.
2. [2 marks]Perform spectral clustering for a given dataset.
-

