



# CS F415: DATA MINING

## SECOND SEMESTER 2018-19

### Assignment-4

INSTRUCTOR: DR. ARUNA MALAPATI

---

**Submission date and time: 26th April, 2019 23:59 hrs**

**Maximum marks: 25**

The goal of this assignment is to implement classification algorithms to correctly assign a class for the data point. You need to implement both **Naive Bayes** and **k-Nearest neighbours** and output the results separately.

You need to train your model either by splitting the data into training and test data (e.g, 80-20%) or check your error on a validation set.

**Dataset:**

- [Nursery Data Set](#)

You are free to use any other dataset provided that it is of comparable size as the datasets provided.

**Programming Languages:** C, C++, Java, Python

**Team Size:** 4

**Report:**

- Name and ID of team members.
  - Dataset used.
  - Pre-processing done on the data(if any).
  - Confusion Matrix
  - Accuracy, Precision and Recall values
-

---

**Submission Files:**

- Source code files
- Report in PDF format
- README

**Remarks:**

- All submission documents should be zipped together and submitted to CMS through one of the group member's account before deadline. Name of the file should be DM\_ASSN4\_201x0xxx\_201x0xxx\_201x0xxx\_201x0xxx.zip
- All source codes will be checked for plagiarism on Moss (for a Measure of Software Similarity). Any kind of plagiarism will lead to severe penalization.
- You are expected to demo your code and present your results as per the schedule that will be made available on CMS later.

**Evaluation:**

- Code & comments (15 marks)
- Report (5 marks)
- Viva (5 marks)

Please contact the following teaching assistants for any queries:

1. Harshith Thonupunoori(f20150071@hyderabad.bits-pilani.ac.in)