Data Mining (IIT Guwahati, Winter 2021)

Instructor

Dr. Amit Awekar (awekar@)

Teaching Assistants

Nidhi Ahlawat (nidhi18@)
Rahul Vats (rvats@)
Manish Gupta (manishgupta@)
Kapil Kukreja (kkapil@)

Course Texts

Introduction to Data Mining, Tan, Steinbach, and Kumar, Pearson Publication

Data Mining and Machine Learning: Fundamental Concepts and Algorithms, Mohammed J. Zaki and Wagner Meira, Jr., Cambridge University Press
https://dataminingbook.info/book_html/

Course will also cover related research literature.

Course Prerequisites

Basic understanding of design and analysis of data structures and algorithms.

Course Purpose

Learn principles of designing effective and efficient data mining algorithms.
This course does not deal with how to use a particular data mining tool.
This course will **NOT** explore Deep Learning.

Grading

A weighted average grade will be calculated as follows:

- Course teaching
    - Course will be taught through video lectures uploaded to the Youtube by the instructor.
    - We will have interactive doubt clearing and discussion session each week.
    - The class teaching will account for about twenty classes.
    - Course topics:
        - Clustering (5 classes)
        - Frequent pattern mining (10 classes)
        - Classification (3 classes)
        - Kernel methods (2 classes)
- Oral exams: 20 points (2 exams, each exam will have 10 points)
    - Exam 1: Clustering and Frequent pattern mining
    - Exam 2: Frequent pattern mining, Classification, and Kernel methods
- Course project: 70 points
    - The course project will involve design and implementation of incremental clustering algorithms.
    - Each project group can have four members.
    - There will be a project milestone to complete per week.
    - You will have to typically meet the course instructor per week to discuss the course project.

- Peer review: 10 points
  - You expected to write critical review for your classmates about their course submissions.
  - The review is supposed to constructively point out the weaknesses and help improve your peers their work.
- Course schedule
  - Week 1: Group formation and problem definition (2 points)
  - Week 2: Review of the selected algorithm (8 points)
  - Week 3: Review of the related work (5 points)
  - Week 4: Coding of the selected algorithm in C++   (10 points)
  - Week 5: Proposed methodology, iteration 1   (5 points)
  - Week 6: Proposed methodology, iteration 2   (10 points)
  - Week 7: Oral exam 1 (10 points)
  - Week 8: Experimental results, iteration 1 (5 points)
  - Week 9: Experimental results, iteration 2 (10 points)
  - Week 10: Report, iteration 1 (5 points)
  - Week 11: Report, iteration 2 (10 points)
  - Week 12: Oral exam 2 (10 points)

The course will have only even grades (AA, BB, CC, and DD).
A final course score of 90 or above is guaranteed a course grade of AA, 80 or above = BB, 65 or above = CC, 40 or above = DD and F for less than 40. It is theoretically possible for everyone in the class to get an AA (or the opposite). Your grade depends only on how you do, and not on how everyone else in the class does. *No incomplete* will be given in this course.

Academic Integrity

The IITG policies against academic dishonesty will be strictly enforced. For the first instance of academic dishonesty, instructor will take appropriate action. However, second instance of academic dishonesty guarantees you the F grade.