IT496: Introduction to Data Mining

Autumn AY 2023-24

Instructor: Dr. Arpit Rana, arpit_rana@daiict.ac.in

Prerequisites: Programming in Python

Lab Session [LT-02]: Friday, 14:00 – 16:00 Hrs

TA contact info: Himanshu Beniwal (himanshubeniwal@iitgn.ac.in) – Head TA
Vipasha Vaghela (202211002@daiict.ac.in) – DA-IICT TA-1
Aksh Patel (202211037@daiict.ac.in) – DA-IICT TA-2

Preliminary Schedule of the Course Projects:

Week	Course Projects	Action	Due[1]
Week-4 [14 Aug 2023]	Regression or Classification Problems on Structured Data using classical ML techniques	**CP-1:** will be Released	Sunday, 10 Sept. 2023
Week-8 [11 Sept 2023]		Evaluation	
Week-9 [18 Sept 2023]	Classification Problems on Unstructured Data using Neural (shallow or deep) Networks	**CP-2:** will be Released	Sunday, 15 Oct 2023
Week-13 [16 Oct 2023]		Evaluation	
Week-14 [23 Oct 2023]	Unsupervised Learning Problems on Structured or Unstructured Data	**CP-3:** will be Released	Sunday, 19 Nov 2023
Week-18 [20 Nov 2023]		Evaluation	

1 – Course Projects (CPs) are due at 11:59 PM on the due date listed.

Academic Honesty

- The following constitute plagiarism on assignment submissions:
  - Copying any segment of code from any source.
  - Submitting code that you did not write yourself personally.
- **Students suspected of plagiarism on an assignment will be given a ZERO.**

Teams

- In total, **149 students** have registered for IT496 as Regular/Audit.
- The lab policy will remain the same for those who are auditing this course.
- There will be **30 teams** in total, each of size 5 (only one team of size 4).
- Members of the team will work together on course projects and the case study.
    - 50% marks will be based on your team efforts, so choose your members wisely!!
    - There will be no change in the team throughout the semester once it is registered. No request will be entertained!
- Every team member needs to know the concepts, code, and claims they submit. Anyone can be asked anything about their assignments.
- Team registration will open (we will share a Google form) on **Aug 8, 2023**. You must register your team by **Aug 9, 2023, at 11:59 PM.**
- Course Projects would be submitted and reviewed on *GitHub Classroom*. Details are provided in the section below.

Course Projects

- There are three-course projects in total; the themes are mentioned in the preliminary schedule above.
- For *each of the three* projects -
    - We will provide **six** publicly available datasets.
    - Each dataset will be assigned to **five** teams.
- During lab hours, there shall be two machine learning challenges without prior announcement. The three most proficient teams shall be acknowledged with a bonus of up to 30% on their respective scores.

GitHub Classroom Assignment Submission.

- We will provide you with a GitHub classroom invitation link, and you have to accept the assignment that will direct you to the course project repository.
- For each course project, you will receive an invite that you have to accept. In the particular assignment, teams will be formed.
- The team leader has to create a team in the classroom, adding other members to the team.
- In the assignment repository, we will share an assignment document that will include information related to dataset allocation, general instructions, evaluation criteria, and general protocols.
- You may clone the base file to your local machine and commit your changes to the repository.
- You will not be able to commit any changes after the deadline is over. Therefore, late submissions are not possible. No requests will be entertained!
- In the project evaluation, your code will be reviewed to check your responses and claims.

Case Study Presentation

Each group needs to follow the case study presentation policy as below.

- The case study presentation holds 10% weight out of 100.
- Each student group/ team will jointly present one case study on a specified day. However, each team member will be assessed individually based on his/ her contribution to this exercise.
- The theme of the case study is " AI for Nation's Priority and Global Interest ." Under this theme, student groups are supposed to present examples of AI-based startups (preferably Indian ones) that are addressing national issues or societal problems. For this, you can refer to white papers/ blogs available on their websites to get authentic information about their ideas and implementations. A few examples were briefly covered in the first lecture of this course.
- Your presentation should at least include (i) the problem definition, (ii) the motivation behind the startup, (iii) the solution they have proposed, and (iv) the impact they have registered on society. The overall motive is to cover how an AI-based solution to a problem has been converted into a business idea—entrepreneurial thinking.
- On the last slide of the presentation, you need to add a group photograph of your team and the names and student IDs of all members.
- Each team has to submit (i) the presentation slides, (ii) any reference material (e.g., white paper), and (iii) links to the startup website, their blogs, etc., just before the presentation.
- You also need to advertise your talk either through a poster or through a short video at least a day before your presentation.
- There will be 30 case-study sessions. The duration of each session will be, at most, 20 minutes. The exact schedule will be released in the month of October this year.