DS 301: Mid-Term Project Details

Find a dataset and project:

- Your team needs to find a research paper around Classification which uses any of the classification algorithms you have studied in class till now. The research paper should have a dataset that can be found in Github, Kaggle or UCI.
- After finding the research paper, you need to provide your team details, research paper and the dataset link to me through slack. **Per group there can be 3-4 members.**

Project implementation:

- Perform all the necessary data preprocessing and feature engineering steps.
- Reproduce the steps mentioned in the research paper. For example, the paper worked on classification with Decision Trees, you need to perform these steps in your project. **If Naive Bayes or any other algorithm is there in the paper which is not yet taught in the class, you can ignore that part of implementation and work on the rest of the models which have been taught to you.**
    - Example: Research paper and Dataset (The dataset details has been mentioned in the research paper)
- Once you have successfully reproduced the steps, you need to create a 10-15 minute presentation to explain about your project. The presentations should cover the following:
    - Introduction to the research paper and problem being solved.
    - Introduction to the dataset including the dataset features and target and descriptive features.
    - Explanation of the methodology (algorithm/approach) used to solve the problem.
    - Show your reproduced code.
    - Suggest improvements to the model. (You don't need to implement improvements, but just indicate the suggestions in the presentation).

Project Code and Presentation Submission:

- To submit your project, you must create a public GitHub repository and upload all of your reproducibility code and presentation.
- It is important to include a README file with instructions on how to run your code.
- Once you have uploaded your work to GitHub, submit the GitHub link along with your presentation on Google Classroom in the Midterm Project submission.
- Please ensure that your GitHub repository is public and accessible to anyone with the link.

GitHub Repository Structure and Components:

- **Readme.md** - Provide the detailed summary of your project, the hierarchy of your project folder.
- **data** - a folder inside which you can place your dataset file. (.csv/.xlsx)
- **models** - a folder inside which you can place your ML modeling work.
- **Project Presentation File** - a file which is a project presentation. If you have a link to your presentation, then you can put it in Readme.md

Some of the useful resources for research papers:

- ChatGPT
- arXiv
- HAL
- Papers with code
- IEEE Xplore
- Google Scholar
- NeurIPS

Grading Rubric:

- Quality of the Research paper (10%)
- Reproduction of the paper implementation (30%)
- Quality of the work (10%)
- Presentation and engagement with the audience (50%)