DS 301: Final Project Details

Find a dataset and project:

- Your team needs to find a research paper around Classification. 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 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 make following significant contributions to the project as described below:
  - Testing the methodology of the paper on similar datasets and evaluating the model performance on different datasets.
  - Experimenting different model parameters and creating an upgraded version of the existing methodology. This will let you explore the model performance by changing different parameters.
  - Try adding different classification models to the experimentation process. For example, if the paper talks about Random Forest, Decision Tree, you can try to perform Logistic Regression or SVM and evaluate the model performance. After doing this, you need to provide arguments on whether your contributions have improved the model performance or not.

Project Report:

- You need to prepare a report about your project, your contributions, challenges and conclusion. The template of the report is provided along with.
- Submit your GitHub repo link and project report to the classrooms before the due date.

Project Presentation:

- Your team needs to present the project report walking through the research paper details, individual contributions towards the project, challenges and learning.

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 Report.pdf** - a pdf file which is a project report
- **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:

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

Project Submission:

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

Grading Rubric:

- Quality of the Research paper (10%)
- Contributions made to the paper implementation (40%)
- Quality of the work (20%)
- Quality of the report (10%)
- Presentation and engagement with the audience (20%)