2021AI511ML
Assignment 2/Hackathon 1
Rules and Regulations

Timings

Competition begins: ~~**3PM, Friday, 8 October**~~
5:40 PM, Friday, 8 October
Competition ends: **10PM, Sunday, 10 October**

These timings are absolute and non-negotiable -- down to the last second.

Groups

There will be six groups formed from the existing TA groups.

Group	**TAs**
Group 1	Aniruddh KB; Krishna Phalgun and Shashank Reddy; Sharmila Mani and Kalyan Verma
Group 2	Saksham Agrwal and Rohit Garg; Vemuru Srihari
Group 3	Sarthak Khoche; Ayush Yadav
Group 4	Rachit Uddyan Yagnik; Shourabh Payal
Group 5	Prateksha U; Mili Goyal
Group 6	Swasti Shreya Mishra; Anurag P

For each group, there will be separate competitions. You will only be competing with the students of your particular group.

Contest medium and description

The competition will be in the form of two Kaggle contests for each group -- one for regression and one for classification. The links for the contests will be shared by your TAs.

Kaggle contests are simple.

Prior to the contest you will have to make your Kaggle accounts, join the contest and **merge your teams to form your pairs**.

You are provided a train dataset and a test dataset. The train dataset has data with labels (y) and the test dataset has only features, no labels. Your objective is to predict the labels for the test dataset, by training on the train dataset. You submit your predictions on Kaggle as a .csv . Kaggle automatically scores your submission using the mentioned metric and returns a score. It ranks teams based on the top submission.

Usually there is a limit on the number of submissions per day, but since this is a short contest we will make it as permissive as possible. You will be allowed 20 submissions per day.

There are two leaderboards -- the public leaderboard and the private leaderboard. The public leaderboard score is available throughout the contest and is only scored on 60% (or so) of the test data. The private leaderboard is on the full test dataset and is only available after the contest ends. You have to choose 2-3 submissions for the private leaderboard. This is done to prevent overfitting to the public leaderboard.

Teams

You will be allowed and encouraged to form teams of 2 (though teams of 1 are still allowed). You must form teams only with students of your own group (though different TAs are allowed).

For example, a team consisting of one student from Aniruddh KB's TA cohort and one from Sharmila Mani's TA cohort is allowed -- because that is the same group for the contest. However, a team of one student from Saksham Agarwal's set and from Shourabh Payal's set is not allowed.

A Google form will be rolled out where you have to mention your team.

Evaluation

Some part of the grade component will be assigned to the Kaggle score (**score, not rank**) . Some part will be assigned to the approaches used.

After the end of the contest, please submit information on what all models you used in the following format:

EDIT: We are modifying the thing to be submitted a little bit. Apart from this, we expect a small document of 4-5 slides listing your approaches, your challenges etc.

Preprocessing	Model	Hyperparameters	Local validation score	Kaggle score (if submitted)

The purpose of this is to demonstrate that you have put your best foot forward and that you have actually tried this on your own.

Please be concise here. In case there is no change in a particular column, you can write "same as above".

You will have to submit this on LMS by **23:59, Wednesday, 13th October**. It is sufficient for one student from the team to submit -- both students need not submit.

Software and models allowed

All packages are allowed, including Scikit-Learn -- including those not discussed in class.
However, there is a restriction on the models allowed. You must restrict yourself to:

1. Linear/polynomial regression
2. Logistic/softmax regression
3. Bayes models (naive and non-naive)
4. SVM (no kernels)
5. K Means (if you wish to do clustering)
6. PCA/SVD

You are allowed to combine these models in whatever way you wish to (blending, stacking etc). This is within reasonable limits. For instance, even though neural networks can be thought of as an aggregation of many smaller models, **neural networks are not allowed.**

Datasets may be imbalanced in nature. You are allowed to use packages to deal with this.

You may choose to use your laptop, Kaggle, Colab, or even AWS. While AWS use is allowed, it is not compulsory. There will be no extra marks given for use of AWS, and no penalty for not using AWS.

Plagiarism note

You are not allowed to copy from other publicly available Kaggle notebooks.
You are not allowed to ask other teams for help or collaborate with them.
You are not allowed to ask other IIITB/IBAB students for help, or on other forums outside.