Statistical Machine Learning **Bonus Assignment**

Instructions: You are allowed to discuss but the final answer should be your own. Any instance of cheating will be considered as academic dishonesty and penalty will be applied.

1. Restrict to using only Python for coding assignments.
2. You are free to use math libraries like *Numpy, Pandas*; and use *Matplotlib, Seaborn* library for plotting.
3. Use of machine learning library scikit-learn is allowed.
4. **No Deep learning library/technique is allowed**.

Link to Kaggle Challenge: https://www.kaggle.com/t/0306555c319e423ba32410b71ab2d3cc

You are given a dataset (uploaded on Kaggle). You are required to use and train any classifier, and then make a submission.

- You would need to preprocess the data to filter out outliers, irrelevant or redundant indices, etc
- You may need to balance the data during training.
- You can use any technique for feature extraction.
- There are marks for scoring (on the basis of Kaggle rankings).
- Make end-to-end codes, i.e., given input data, output csv should be generated.
- Save your final models/weights and submit them along with the codes.

Evaluation criteria will be correctness, viva, and marks/rank scored on Kaggle.

Submission format: Please submit a report describing your methodology, observations and results only in the PDF format. All the python code files need to be submitted with the following naming format: "**<Roll_Number>_file.py**" in a zipped folder with source folder named as "**bonus_<Roll_Number>.zip**"

Note: Submit all codes, models/weights, on Classroom in a zip file in specified format.