Statistical Machine Learning
Kaggle Competition **Max marks:100**

Instructions:

1. You are allowed to discuss but the final answer/implementation should be your own. Any instance of cheating will be considered as academic dishonesty and penalty will be applied.
2. Restrict to using only Python/Matlab for coding assignments.
3. You are free to use math libraries like Numpy, Pandas; and use Matplotlib, Seaborn library for plotting.
4. The use of machine learning library scikit-learn is allowed.
5. **No Deep learning library/technique is allowed.**

Link to Kaggle Challenge:
https://www.kaggle.com/t/cf467eb15a5b4cb79ad6231d53c1b797

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.
- You may use techniques like SVM, bagging, boosting, PCA, LDA, Multilayer Perceptron(up to 5 hidden layers), etc. These are just example techniques that can be useful. However, You can use other models also from scikit-learn.
- There are marks for scoring (on the basis of Kaggle rankings).
- Look and review also all the instructions, data, rules carefully mentioned on the Competition Page before starting the challenge.

Dos:

- Make end-to-end codes, i.e., given input data, output CSV should be generated. Your code should able to generate predictions and accuracy for any given test file.
- Save your final models/weights and submit them along with the codes.

Dont's:

- Don't use someone's code/model, you have to code and build your model on your own.
- You are not allowed to use any kind of pre-trained or deep learning models.
- If found for any of these cases, the penalty will be applied strictly.

Evaluation criteria will be correctness, demo, and marks/rank scored on Kaggle. You will be evaluated based on how you come up with the proposed approach and what led to you here, why your technique worked, what are the experiments made by you, which techniques failed and why. Hence, It would be good if you cover all these things in your report itself.

Submission format: Please submit a detailed report describing all your methodologies, observations, experiments, results, analysis, and inferences only in PDF format. All the python code files, models, the report need to be submitted with the following:

- **Naming Format:**

  " <Roll_Number>_file.py"
  "<Roll_Number>_model"
  "<Roll_Number>_Report.pdf"

- All files in a folder with a source folder named "Kaggle Challenge".
- **Zip file name:** "Kaggle_<Roll_Number>.zip ".

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