Machine Learning – Programming Assignment

This is a group assignment.

You may apply any learning algorithm (under supervised, unsupervised or reinforcement learning) to solve any real-world problem based on any real-world dataset. You may use a publicly available dataset from a source like the following one (you can choose any other data source. Be sure to cite where you got your data from. **The dataset has to be real-world data that is published on the internet**). Following are two example datasources for ML datasets. You may use any additional data source as long as they are credible and publicly available.

http://archive.ics.uci.edu/ml/index.php

https://www.kaggle.com/datasets

Based on a dataset, you may apply machine learning algorithm and then get the test results. You may use Python as the programming language and Jupyter Notebook as the programming environment.

For supervised learning, **four** distinct algorithms should be applied on the dataset (team size 4). If the team size is three, then apply three algorithms.

For unsupervised learning/clustering, **three** distinct algorithms should be applied on the dataset (team size 3).

For reinforcement learning, **one or two** distinct algorithms should be applied (team size one or two, respectively).

Note: If you're performing supervised/unsupervised learning and if you do the assignment individually, make sure to compare at least two algorithms.

Submission

- The ids, emails of the members in a text file called 'members.txt'.
- A text file called 'submission.txt' containing the link to the dataset. If you're using your own dataset, make sure to upload to a shareable cloud directory and copy the link here.
- Link to a public Github repository containing the Jupyter Notebooks containing the code appended to the same 'submission.txt' file.
- Link to a youtube video containing the presentation/demo of the work done. Each member can use maximum 4 mins to explain their contribution to the project, appended to the same 'submission.txt' file.
- A report in **pdf format** with a description on the problem addressed, dataset used, methodology (algorithms used), results and discussion (including possible limitations/future work). **The report should compare and contrast the different algorithms used, in terms of their accuracy, using appropriate metrics.**

- The source code should be added as an appendix to the report.
- The entire submission bundle has to be uploaded to the courseweb link (which will be available close to the deadline) as a single zip file. The name of the zip file should be ML-assignment. **Only one person needs to upload the submission on behalf of the group.**

The report should have an appendix that contains all the source code (**added as text, not screen-shots**). **If the source code is not added as text in the appendix, it won't be accepted as a valid submission.**

The report is the main deliverable that will be marked. However, the code should be submitted to validate what is mentioned in the report. **Both the report and the source code should be there to be accepted as a valid submission.**

Please note if any of the above components are missing in the submission, it will not be accepted as a valid submission and will be given zero marks. Any resubmissions will be awarded a maximum mark of 45%.

All reports will be uploaded to Turnitin for plagiarism checking. **If the turnitin similarity is above 20%, marks will be penalized.**

Deadline: End of 10th week of the semester

Marks allocated: 20