**Machine Learning – Programming Assignment**

This is a group assignment. **The group size is 2. If you prefer, you can do it individually.**

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.

<https://archive.ics.uci.edu/datasets>

<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, **two** distinct algorithms should be applied on the dataset.

For unsupervised learning/clustering, **two** distinct algorithms should be applied on the dataset.

For reinforcement learning, **one** algorithm should be applied.

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 5 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 12th week of the semester**

**Marks allocated: 20**

**NOTES**

**Source:** https://www.kaggle.com/datasets/andrewmvd/heart-failure-clinical-data

Various supervised machine-learning techniques like **K-Nearest Neighbour, Decision Tree, Logistic Regression, Naïve Bayes, and Support Vector Machine (SVM)** model are used for predicting cardiac disease using a dataset that was collected from the repository of the University of California, Irvine (UCI).

[**https://www.youtube.com/watch?v=OO7Y5wQWnQs**](https://www.youtube.com/watch?v=OO7Y5wQWnQs) **– reference KNearest**