**Machine Learning**

**Programming Assignment**

This is an individual 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.

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

<https://www.kaggle.com/datasets>

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

Submission

* The dataset (or the link to the dataset if it’s too large to be uploaded)
* The Jupyter Notebook containing the code.
* A **10-12** page pdf report with a description on the problem addressed, dataset used, methodology (algorithm used), results and discussion (including possible limitations/future work).
* 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 your id (for example, IT15234566.zip).

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 **30%**, **10%** of the marks will be reduced. For **50%** similarity, **50%** of the marks will be reduced. For reports with **80%** similarity, **no marks** will be given.

**Note:** If your submission size is larger than 10 MB, you may upload the submission to dropbox and share a dropbox link (please use dropbox only). Make sure that the download link is publicly accessible. If the download link is not accessible, I can give you another attempt to upload your submission but I will have to cap the mark at 45%, as it is effectively a resubmission and otherwise it would be unfair for those uploaded the assignment on time. If you are uploading the dropbox link, make sure to include that in a text file (readme file) and indicate your name, registration number, contact number and contact email in the same text file, so that I can contact you if there’s any issue in accessing the dropbox directory.

**Deadline: End of 11th week of the lectures**

**Late submission deadline: End of 12th week of lectures (20% marks will be capped)**

**Please note no submission will be accepted after the 12th week.**

**Marks allocated: 30**