CMP 466 – Machine Learning and Data Mining- Spring 2021

Team Project- Assignment 1

Please submit a report including the following by the deadline:

1- Introduction about the project where you introduce the problem and your idea to solve it.

2- Literature review. At this stage, each team member is expected to summarize 3-10 recent research papers (depending on the literature availability, but not less than 3). Papers using the same datasets are highly recommended. Other papers using similar datasets/ideas are recommended. When summarizing any paper, try to focus on the purpose, methods used, dataset (public or private, its name), the performance metrics used and their values.

3- Description of the dataset to be used in your project: source, type, size (number of samples), dimensionality (number of features), class labels, the number of samples in each class, etc.

4- Familiarize yourselves with the data, by reading the data into your Python environment and visualizing some of its features vs. labels.

5- Preprocess the data using Scikit-learn library, such as encoding the categorical features:

https://scikit-learn.org/stable/modules/preprocessing.html#preprocessing

Submission: Please submit your assignment to iLearn by the deadline. Late submissions will be penalized according to the syllabus late policy. No submissions will be accepted after four days of the deadline.