Course Project

**Instructions**

The purpose of the course project is to apply the principles of machine learning covered in the course to a real application of your choice.

For the project, find an interesting dataset that allows you to create an interesting prediction model. Think of something you want to model using the dataset that could serve some useful purpose in an application area. **I suggest to** choose a dataset having at least 10 variables and consider possibly predicting multiple variables in your prediction output. However, there are also interesting problems using fewer variables which are acceptable; please choose a problem with a non-trivial data modelling aspect.

Once you’ve chosen a desired learning model target, set about creating an optimal trained model to perform your desired prediction task. Consider at least (3) separate model types in your evaluation and explore some hyper-parameter space of each model additionally. After evaluating these models and parameters, explain your choice for the final model that you think should serve as the best predictor available.

Based on project work, create a 10-15 min presentation to be given in class between November 14 and November 28, 2024 (exact schedule TBD through random selection process).

Also, create a final written report to document all the details of your project work. The report is relatively free-form however I suggest covering these points:

-Explain the problem you are solving and why it has some value to someone. Why is it useful?

-Explain the dataset used and any processing or steps required.

-Explain which models you use to try to solve your prediction problem.

-Evaluate models on the data and explain methodology. Present the results in a readable and understandable way using appropriate plots, figures and explanations.

-Evaluate the end results and try to determine if your predictor will achieve its goals.

Please cite any references or sources in your final report that you use during the course of the project.

**Marking Scheme**:

**Proposal**: 2%

-Identify data source(s)

-Explain what you want to model and briefly why it would be useful.

**Overall Content/Presentation**: 15%

-Overall Content of the project.

-Clear Overall Presentation

-Explanation of the problem you are trying to solve

-Methodology and Results

**Final Report**: 8%

-Explain the problem you are solving and why it has some value to someone. Why is it useful?

-Explain the dataset used and any processing or steps required.

-Explain which models you use to try to solve your prediction problem.

-Evaluate models on the data and explain methodology. Present the results in a readable and understandable way using appropriate plots, figures and explanations.

-Evaluate the end results and try to determine if your predictor will achieve its goals.

The report will also be judged on clarity in presentation of results.