

Final project - 30 points

The final project is for you to experiment with all you have learnt using a dataset of your choice.

Here is the step by step instructions for doing the final project. I am not expecting you to do anything complicated. A simple prediction model would be enough.

Step 1: Find a dataset - here are some sources – www.data.gov or <https://archive.ics.uci.edu/ml/index.php>.

Select a dataset that you are most comfortable with – it need not be in healthcare and it need not be from the above two locations either – if you have access to some other data, pl. use that.

Step 2: Develop one or two simple hypotheses with predictors and target variable

Step 3: Use the most appropriate ML technique/s that you have learned to develop a prediction model. First, develop a few models – report the accuracy of each model and pick the most accurate one as your final model. Write a report on the process, methodology and about the model that is the most accurate.

What needs to be submitted:

Write the above in the format of a short journal paper (2-3 pages single space – not including references, tables, figures etc)

Introduction – Describe why you are developing this prediction model (e.g. how it will benefit the society). Please add any references if you can find any.

Research question and Hypotheses – Write about your predictors and target variable and a simple and clear hypothesis (1 or 2 would be good enough).

Method – Describe the ML method/s you are using. Describe all the steps you undertook before developing the model. Explain why you chose this ML technique over others.

Results: Report the accuracy of each model and describe the model that is most accurate.

Appendix: Add tables, figures, R screen shots etc.

Due date – check syllabus

