**Final Project**

You have to write a paper (6 to 8 pages) about your project.

**Choose a dataset from**

UC Irvine Machine Learning Repository (<https://archive.ics.uci.edu/ml/index.php>)

with at least 5000 instances and 20 attributes for classification or regression.

**Compare** how the different approaches seen in class perform on this dataset to

**Predict** accurately the classes or the values of the unlabeled data.

**Determine** what are the best hyper-parameters for each approach you are using.

-we need to split our dataset 75% of the data for training and 25% of the data for testing

**Linear Regression Single Variable**

-We want to minimize the error between our points and a line of best fit.

**Cross Validation**

-Rotate the 25% testing data through the data set, record the results for how well each method does and compare the results to each other. Since, we divided the data into 4 blocks, this is called Four-Fold Cross Validation.

**Python libraries**

-pip

-pandas

-Matplotlib Plotting

-sklearn

**Presentation**

You have to include a presentation of the research questions and the chosen methods to tackle them, Relevant Literature Review, a presentation of the results and discussion and a conclusion/future work.