**2/28/2021 Meeting:**

* **Test, Train Split**
  + Use same random seed for splits or save data to two csv files so train and test splits are the same for each model
* **Ensemble:**
  + Final ensemble in scikit learn with classifiers scikitlearn.ensemble.votingclassifier
  + Ask Dave about ensemble - can a heterogeneous model use different features for each base model?
    - Initial assumption: Maximize accuracy of each model and then use an ensemble at the end. Each base model has 1 vote for output, therefore the input features of each model can be different?
* **Interface for our program**
  + In terminal what numbers do you have and then we can build a dataframe for them and our program will run and display an output? Or fxn outside of ensemble that reads in a dataframe and put into our program.
* How to unit test this?
* **From Prof. Valleau on classifying / determining efficacy of models**
  + [Feature selection using Python for classification problems | by Richard Liang | Towards Data Science](https://towardsdatascience.com/feature-selection-using-python-for-classification-problem-b5f00a1c7028)
  + F1, etc. better for understanding if model is doing well or not as opposed to accuracy of the model
  + Need to look at histograms
    - Could look at residuals for booleans 0, 1, -1
    - Also ask about histogram
* **Action Items:**
  + Go through classifiers
  + MSE test for already made models
  + Decision tree stuff
  + Clarify questions for Monday
* ***Anaconda navigator for confused beginners like me.***