Each team will consist of three students. They should use at least 6 different classification methods to perform experiments. They should compare these different methods on the training data and write a summary report that include how the models are trained and their performance on the training data. Based on the performance of the algorithms on training data, they should pick the best algorithm to run on the testing data. And the team needs to provide prediction results of the best algorithm on testing dataset.

Submission (zip all files into one and upload it on Canvas):

1. report (7pts): 1pt for overall structure, organization, and language of the report, 1pt for each approach, need to specify which library/function is used, how the model is trained (e.g., any hyper parameters and how their values are determined, any cross-validations), results on the training data (using one figure, that includes accuracy results from all algorithms, generated from within your code). Based on the results, you designate one Best Model, which will be used in step 4.   At the end of the report, please specify the effort from each team member as a percentage and this sentence "all team members agree with the specified effort." If all members have the same percentage, they will get the same score. Otherwise, individual members will get different scores based on their effort.

2. code for 6 models (6pts), 1pt for each model. Your code should should include a readme file to help TA run your code on Training data. By directly running your code on the same training data, TA should be able to reproduce the figure generated in your report.

3. code to run test data (3pts), 0.5pt for each model. Your trained model will be obtained from step 2. This section will allow TA to run the trained models on the test data and your program will report final prediction of each method.   A readme file is needed to help TA to run it.

4. prediction result of the Best Model on the test data (4pts). TA will take the final prediction result from your designated best model and calculate classification error rate.   If the error rate <=2.5%, 4pts, otherwise <= 5%, 3 pts; otherwise if <=10%, 2pts; otherwise if <=15%, 1pts; otherwise 0pt.

Timeline: 3 weeks.