**Kuiper Comments**

1. Best school as a variable has problems because it lacks evenness—to make it work, consider grouping as strongly agree vs all else.
   1. However, Kuiper does not like this question because it is very subjective.
2. Kuiper would prefer a variable like time spent on homework because it is more objective.
   1. We should group the data into relatively equal sized groups — for example, less than 2, 2-5, more than 5
3. We can try to figure out how well we can predict the time spent on school using a variety of models.
   1. We can tweak our models by using stumps or multiple decision trees
   2. This requires evaluating confusion matrices
4. We can use three response variables
5. We should include NAs in our analysis; we can also compare the model with NAs to the model without NAs.
6. We should use 20 variables in our model that we think will be meaningful. We should consider prior research and shouldn’t restrict variables unless there is prior research.
7. Our split for the data should be 70% training, 30% testing
   1. We should check to make sure the counts are roughly accurate
8. We should create a random forest and create an ROC.
9. We should NOT use RMSE because RMSE only applies to numerical variables.
10. We should check the efficacy of the model using accuracy? instead
11. We will treat our variables as categorical.
12. We should check the validity of our model using graphs ?
13. We should try our models with and without the variables that we think are suspicious.
14. The Machine Learning Summary sheet is our guide for this project.