Used Python, Jupyter Notebook, numpy, pandas, sklearn

We used random forest to generate the model

We compared the model’s prediction with the testing database’s labels to find the average difference between the predicted points and the students’ actual points. This is our model’s error margin. We subtracted this number from 100 to get a percentage for accuracy.

Our dataset consisted of 1000 students. We used 250 to train the model, and the other 750 were used for testing.

**Math predictor model:** Our initial model had an accuracy of 87 percent but using a double for loop to iterate initial random states 0 to 100, we were able to increase the accuracy of the model to 89.43 percent. What this means is that our predictions of the students’ grades are off by +- 10 points on average. While this is not the most accurate model, we believe that with more data, the model could improve significantly. We could also continue to tweak the current model to fit the current dataframe by further adjusting our hyperparameters.

**Reading and writing predictor model:** We applied the same technique to the model for predicting the reading and writing scores. We achieved an accuracy of 89.52 and 89.76 respectively.

Made up six random students and predicted their scores.