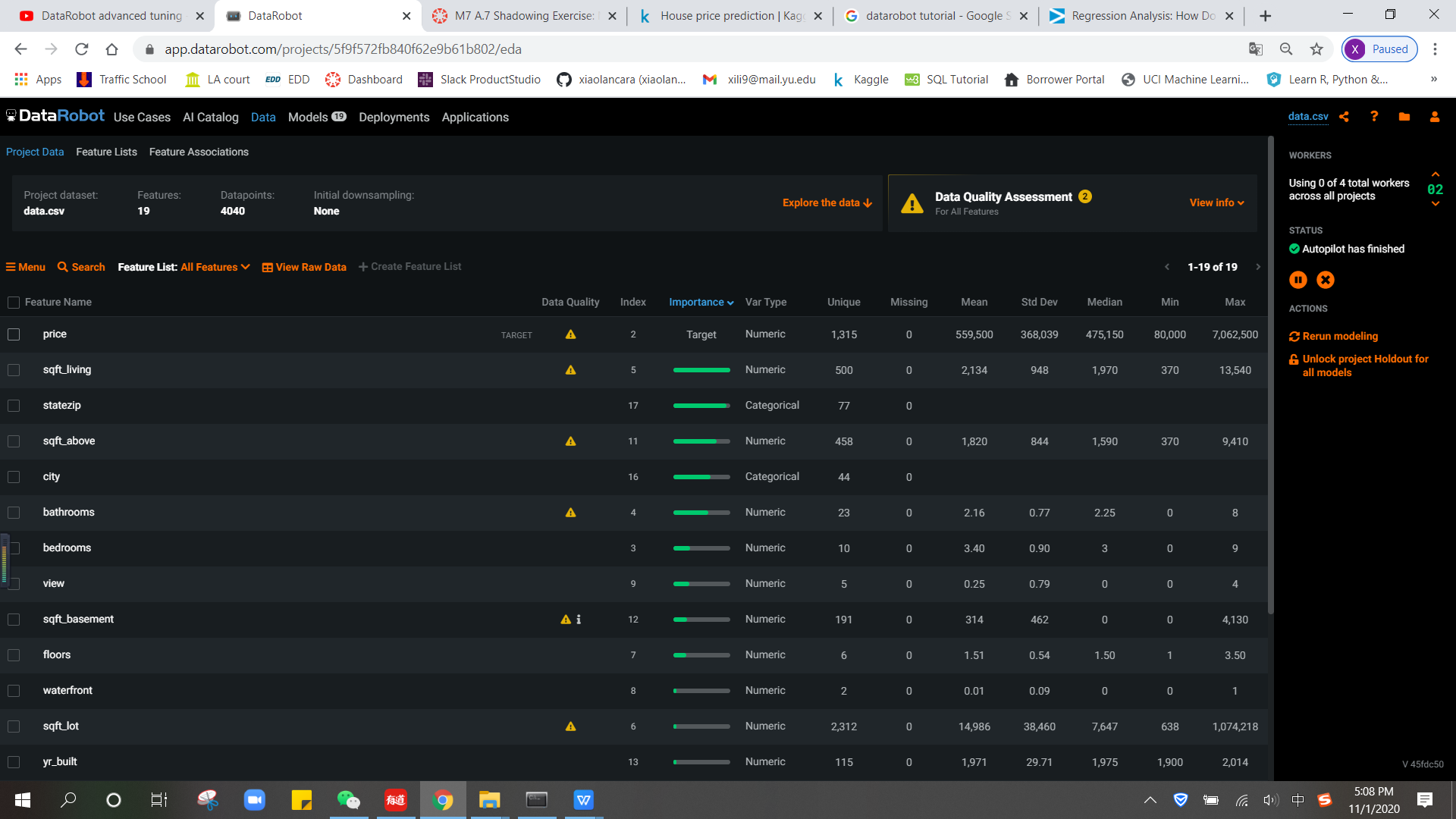
**Introduction:**

The dataset is from kaggle. <https://www.kaggle.com/shree1992/housedata>

The database contains information on real estate in 2014, including sqft\_living, statezip, bedrooms, bathrooms, and prices. In this assignment, I will use the training data to predict the target ‘prices’ and choose the best model by using datarobot.

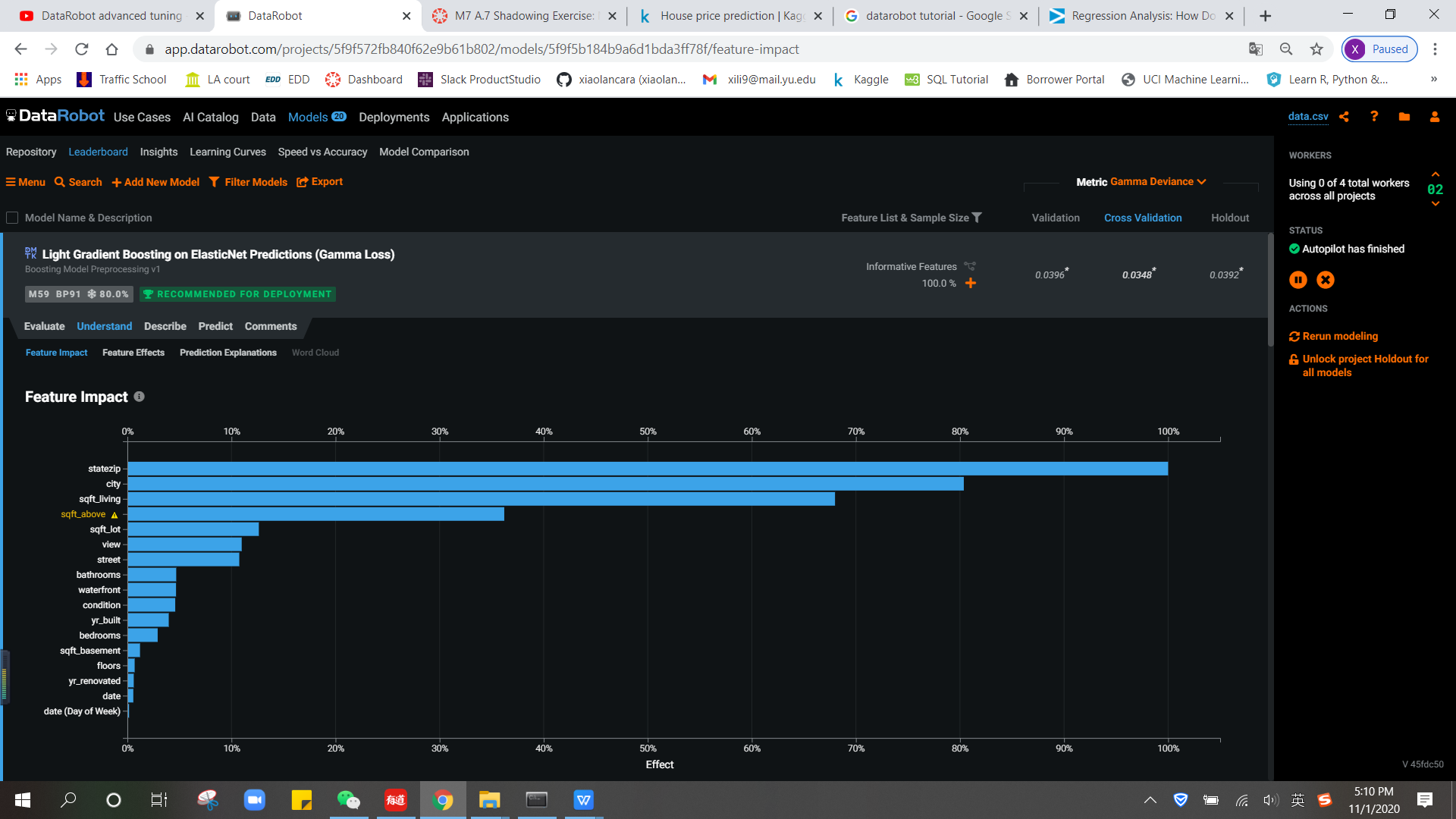
#### ****Model Development:****

1. First, I split the data to data(training data) and data\_test. Then import the raw data into the DataRobot. With the target ‘price’ I choose , the correlation importance of features were shown as below.

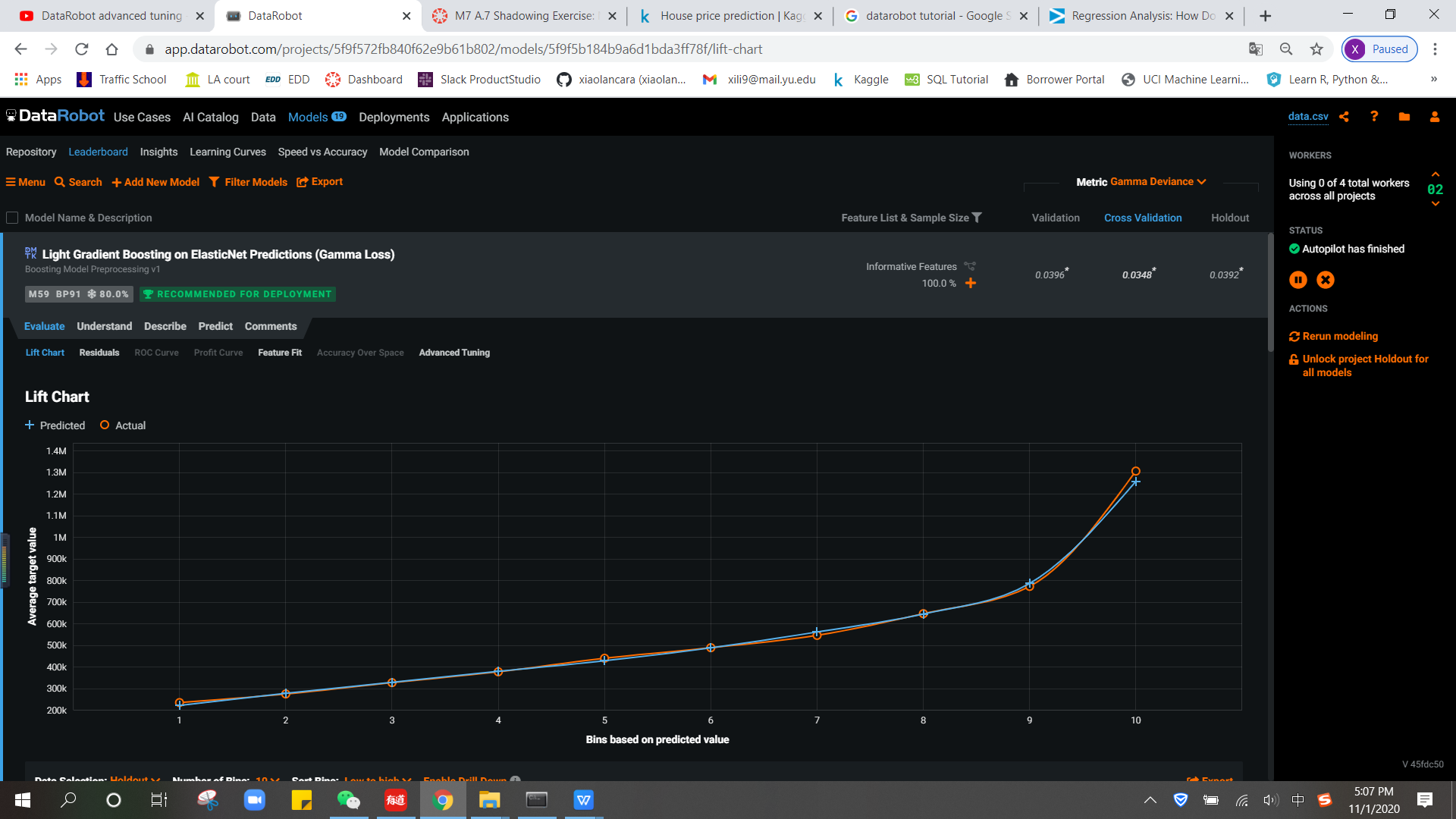


1. DataRobot provides 20 models for predict. I choose the recommended model Gradient Boosting to explore.

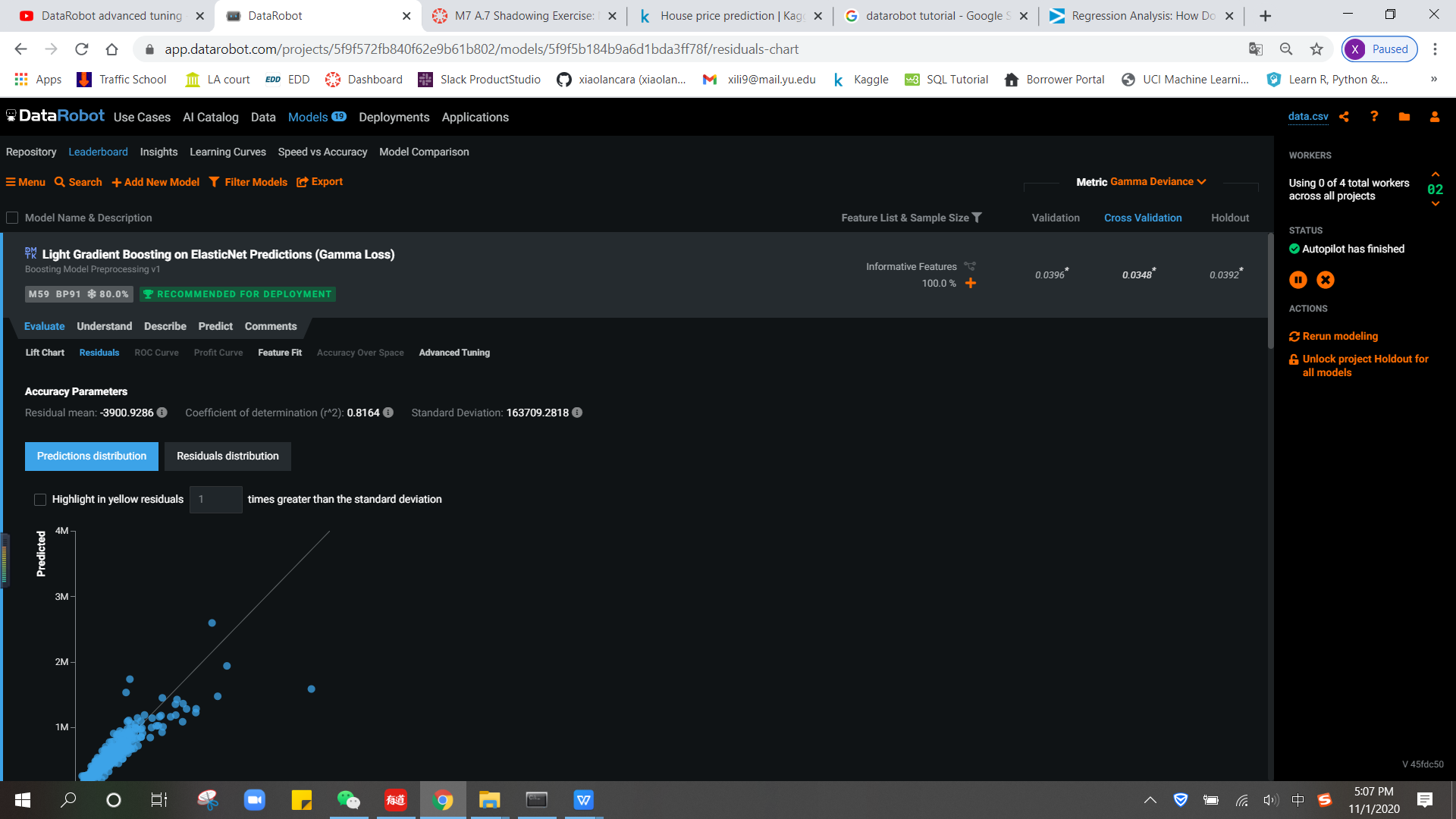
We can tell the highly correlated features from feature impact as below.



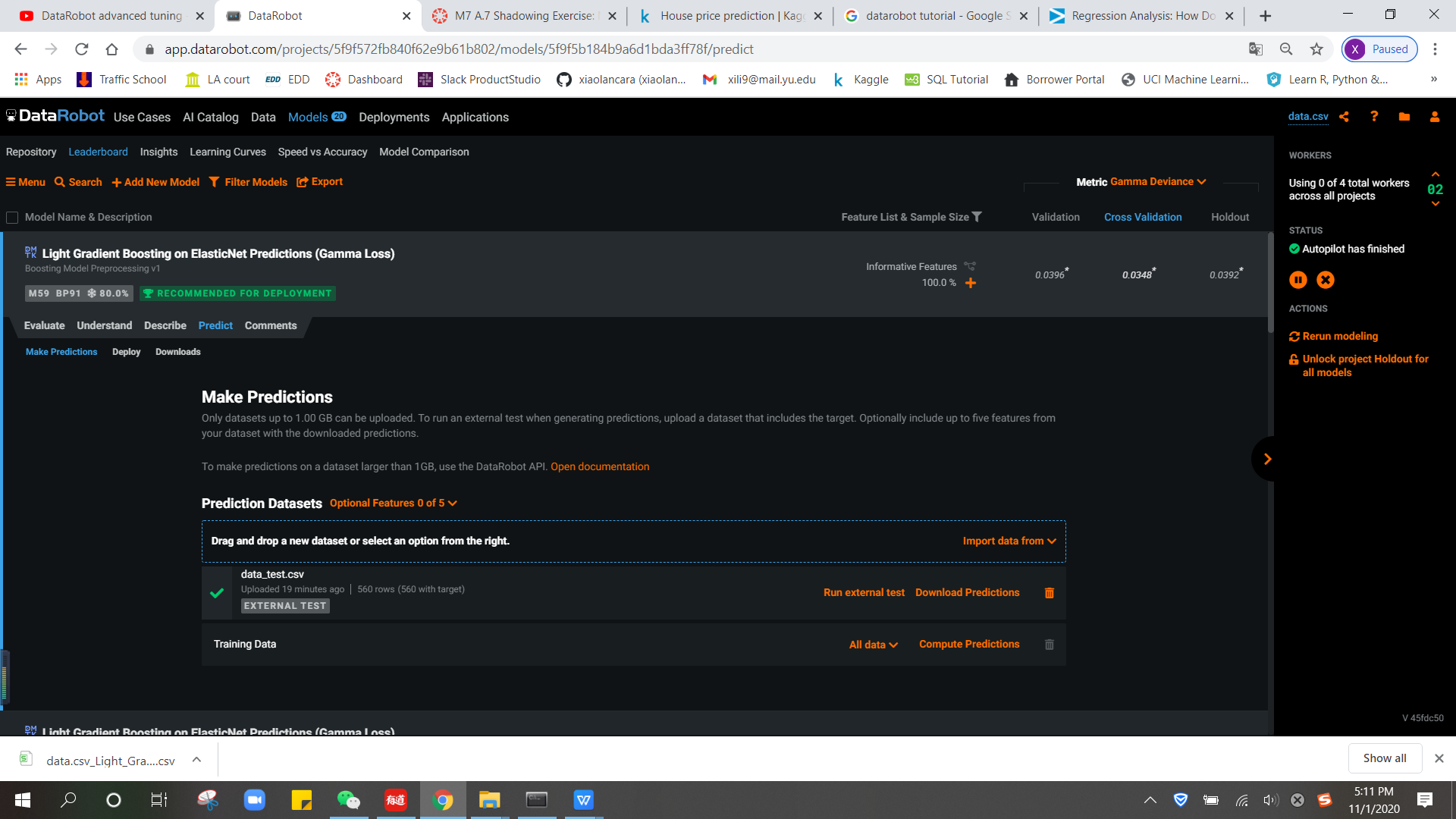
1. By using the Evaluate lift chart, we can see the predict line and Actual line is almost the same for now.



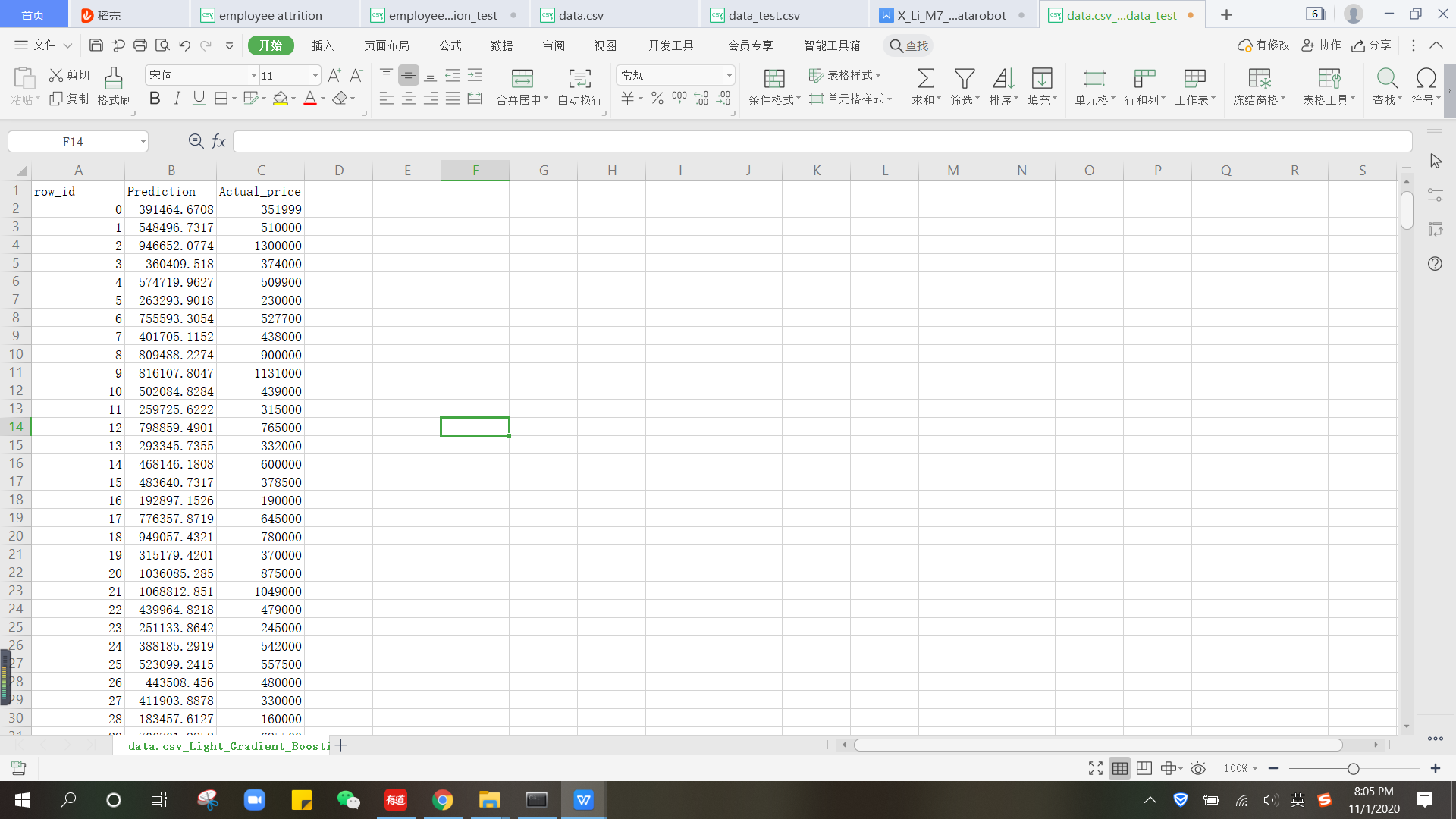
1. The r2 value of Gradient Boosting models is 0.8164, which demonstrate it is a highly predict model.



1. import the data\_test into the model and make prediction

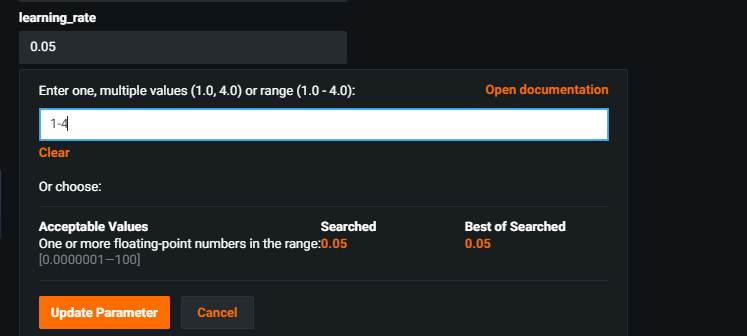


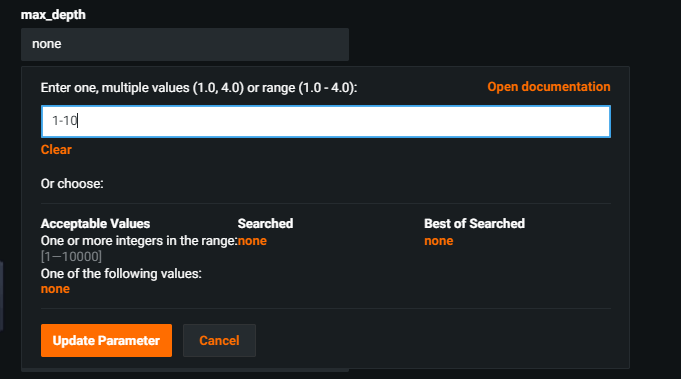
The predict price for data\_test is as below. We can add the Actual\_price to compare the result.



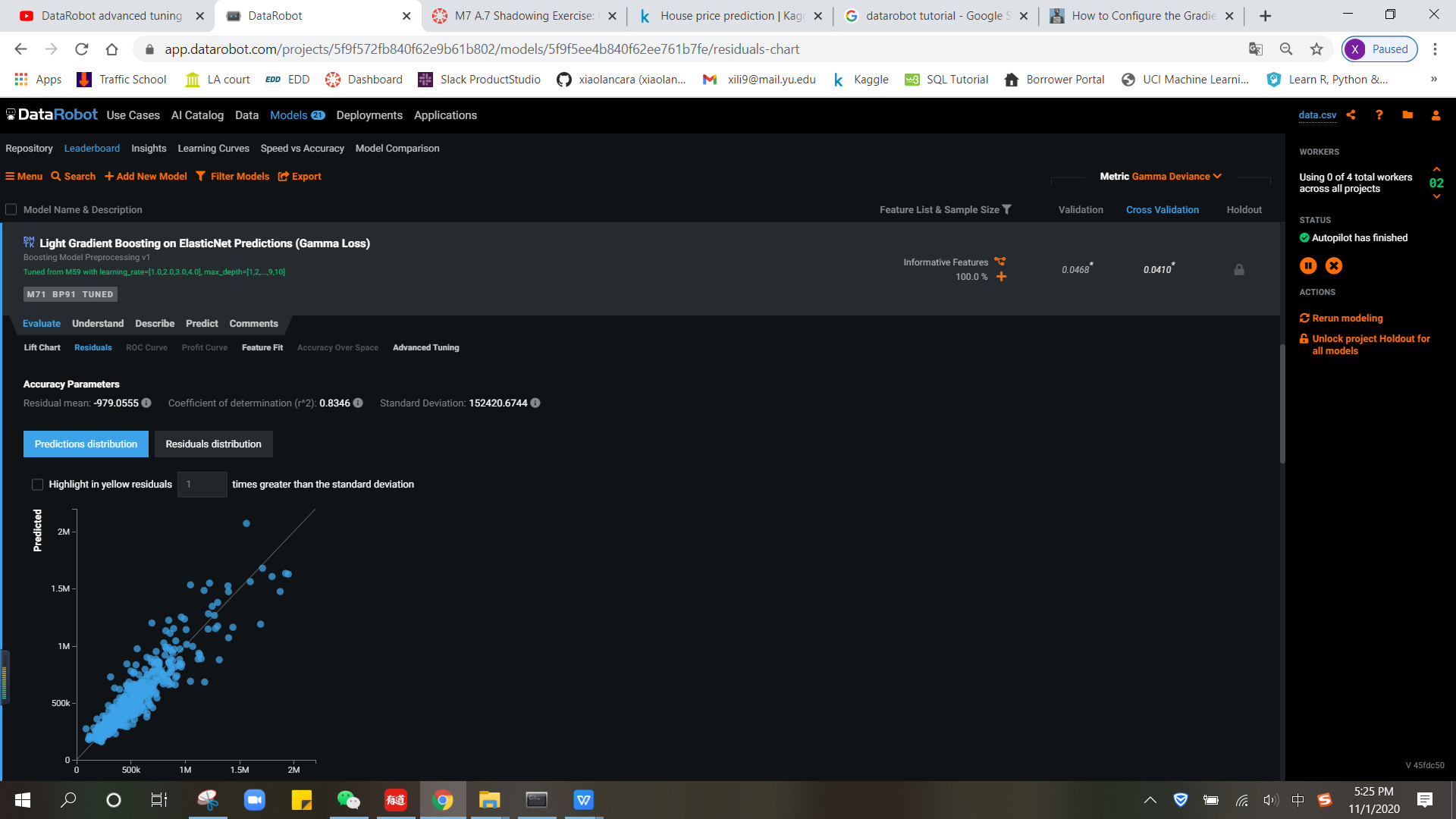
#### ****Hypertuning:****

1. On the Advanced Tuning tab, I change the parameter learning\_rate into a range 1-4, max\_depth into 1-10 range.

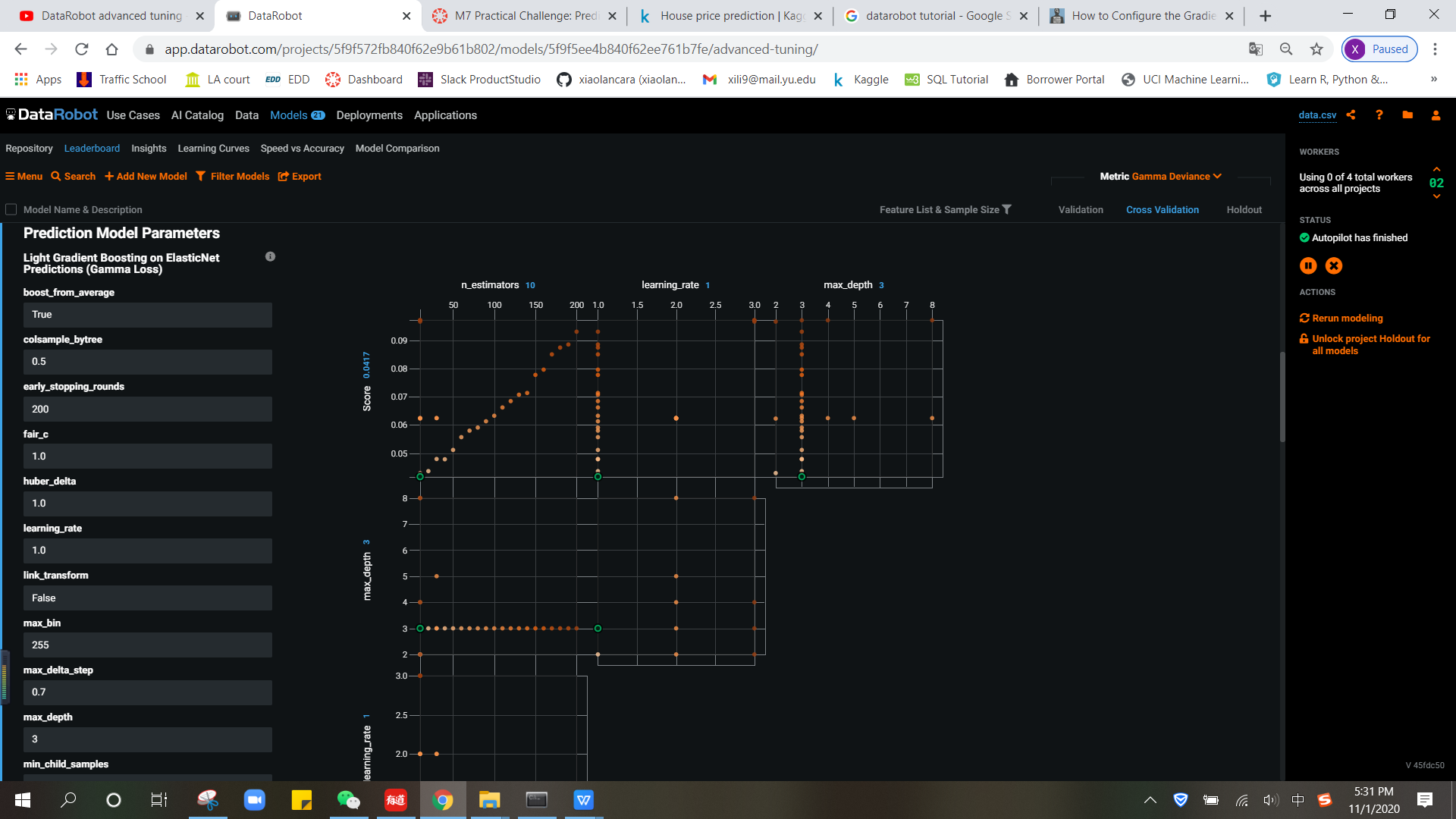




1. After Tuning, we can see the predict result r2 value has been improved to 0.8346



When we go to check which value is the best for parameter, we can tell from the current tuning which learning rate is 1, max\_depth is 3. Using the DataRobot is easily to help choose the best parameter with hyper tuning.



**Conclusion:**

**Datarobot finds the most important features in a quick way and provides a number of prediction models to users, ranking them according to their prediction rate, and providing the most recommended models. It also provides advanced tuning choice for users to avoid machine error.**