Pluralsight Modeling Test

William Rinauto

**Comments**

I wanted to do a quick writeup to explain precisely what I am turning in for this assignment. I will email a version of this document to Scott when I’m ready to turn everything in, and I will also include it in the repo that gets deployed to github.

All relevant content for evaluating this project should be contained in ./PROJECT DELIVERABLES.

You’ll notice that my project was written mostly in R, while the instructions indicated that it should be written in Python 3. I know that I have more expertise when it comes to R, so I thought that I would be more able to fulfill/demonstrate the stated non python-related requirements of this assignment by building the model in R (critical thinking, data cleaning, and applied analysis).

My way of reconciling this was to create a jupyter notebook that “productionizes” the model originally developed in R. My hope is that this can be seen as a simple way of prototyping in R, but still productionizing in Pluralsight’s preferred language for deploying ML models.

**Project Summary**

-**rmd-model.html** is the rendered RMarkdown document that shows the cleaning and exploration of data, as well as the model building process. I tried to be very thorough with my comments here to explain my thought process and decisions.

In rmd-model.html, I test 3 types of predictive algorithms and selected the best model based on each model’s cross validation performance.

-**productionize.html** is the productionization of the model in python. It reads in the h2o model that was created in the R environment, prepares the data, and then generates and saves predictions. The main part of the python code are two functions, **cleanup\_data** and **create\_predictions**, which taken together, prepare data to be passed to the model, and call the model.

-**results.xlsx** is the predictions from the final model applied to the testing dataset

Note: I’m also including two additional files here as alternatives for the two html above. I ran into an obscure Latex error in both R and python when trying to knit/download my work as pdf. The pdf files included are the html counterparts, “ctrl-p” saved from chrome. (this does not do a great job preserving formatting so I recommend reading the html files if possible).