|  |  |
| --- | --- |
| **Criteria** | **Points (0-10)** |
| **Content** | 10 |
| **Subject Knowledge** | 9 |
|  | **19/20 Excellent job!** |

**General Feedback**

**Content**

* Very good repo organization!
* Great scripts, nice use of relative path when reading in the data and nice comments throughout
* Maybe put the “preliminary” scripts in a separate folder? Just to make it more clear these are not the final scripts for the analysis reported in your paper
* Excellent document formatting, all tables and figures look great!

**Subject Knowledge**

* Great introduction! Establishes the public health problem and utility of prediction and some nice details on the literature (especially the PubMed search)
* Great detail on dataset and data cleaning
* Great detail on training, tuning, and testing, though I would like more detail on 1) tuning of the KNN (assuming you did the default bootstrap within your training folds, but should stated explicitly) and 2) tuning of RF (it seems a 5-fold CV was used within the training fold which is good! Though the term “test error” is unclear.)
* Nice interpretation of tables and discussion of collinearity
* With the Youden’s index, it seems like it was chosen based on the **test set** performance, and not the training folds. Since this is technically a tuning parameter, it would useful to select the threshold from the training set. This should also be explicit in your document
* Good detail in the discussion/conclusion. Would like to see this research question analyzed with more detail info as stated in the last paragraph
* Note that you could have done the ROC analysis with the RF as well, may have been interesting. Also variable importance measures from the RF would be interesting to see.