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

**General Feedback**

**Content**

* Nicely formatted repo! Really like the separation of the supervised, unsupervised learning code and a separate output folder
* Would put the scripts in the main folder into a subfolder called “test” or “prelim” to store
* Great tables, figures, etc.!

**Subject Knowledge**

* Great introduction on the scientific background for the analysis. Maybe include the metrics on how well the machine learning algos in the literature did?
* Great detail on variables, outcome, and data characteristics. Which variables were removed due to correlations being high?
* Good detail on the training and tuning methods used, though would like more detail on the tuning. Was an additional CV process used for tuning, the bootstrap, etc.?
* Confused on the term “full dataset” for the penalized regression models
* Were the 10 predictors for the RF selected from the whole dataset or within the training folds only?
* I really like the use of supervised and unsupervised learning both together, as well as separately to ID unique traits
* Really nice interpretation of the results and cross-refencing of the figures and tables
* Would like to see some discussion of on the loading on the other eigenvectors (i.e. other components used, I believe the 2nd and 3rd were also used).
* Outstanding discussion at the end in all ways
* Really liked how you tied together all of the supervised learning results, specifically the LASSO-determined significant predictors and the RF variable importances