Project Notes

Sunday, May 17, 2020

7:56 PM

* 1. Data cleaning/wrangling
     1. Use missForest to impute missing values
     2. Organize columns in data
     3. Convert data to correct data types
  2. Visualize the data
     1. ggpairs()
     2. Figure out which geoms to use with different types of data
     3. Iterate through all variables against one or two (attrition/jobSatisfaction)
        + Investigate plots some more?
  3. Feature Engineering
  4. Split data into train and test
  5. Run a random forest FULL model and look at varImp()
  6. Benchmark models
  7. Kmeans clustering or association rules, which one makes sense to use?
  8. Conclusion
  9. Create report

Meeting Schedule

~~\*Mon 5/18 - Wed 5/20 (afternoon)~~

~~Sun 5/24 (evening)~~

\*Tues 5/26 – Fri 5/29 (afternoon/evening)

Report

Sat 5/30 (afternoon)

Sun 5/31 (evening)

\*optional