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?
        + Print out all ggplots in HTML file
        + Give some simple metrics on attrition
3. Feature Engineering
4. Split data into train and test
5. Run a random forest FULL model and look at varImp()
   1. Add simple decision tree
   2. ~~Run models again on a few variables and compare models~~
6. Benchmark models
7. Run Association rule mining
8. Conclusion
9. Create report

Meeting Schedule

Coding

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

~~Sun 5/24 @ 7PM – 9:30PM~~

~~\*Fri 5/29 @ 2PM – 6PM~~

Report

Sat 5/30 @ 3PM

Sun 5/31 @ 7PM

Mon 6/1 - Wed 6/3, Evening

 \*optional