Research Methodology (Dev’s Part so far , 1 algo more to go)

Collected the latest data set of thoracic surgery available on the Internet. After going through the data cleaning steps, which are changing alphabetic values given in the dataset to numeric (True = 1 , False = 0) , PRZ (level of affect on a person’s health due to lung cancer. Basically the stages of lung cancer)is a scale where 0 is the best and 5 is the worst in this dataset we just have till PRZ2 from PRZ0 which is non restricted and fully active , PRZ1which is able to carry out light work , PRZ2 resting 50% of the day. Next was the stages of lung cancer OC11 is the first stage which was given a value of -1 , then OC12 = -2 so on till OC14 = -4

Applied feature scaling and the data was divided into training and test with test size of 25%

Algorithm applied SVM , RANDOM FOREST

Accuracy of SVM :- 82.2 % (write it in statement )

SVM parameter tuning kernel = linear , random\_state = 2

Accuracy of Random Forest : - 83.05% (same ^)

Random forest tuning n\_estimators = 100 , criterion = ‘entropy , random\_state = 0

Search on net about these algo and write and elaborate in 200 words.

