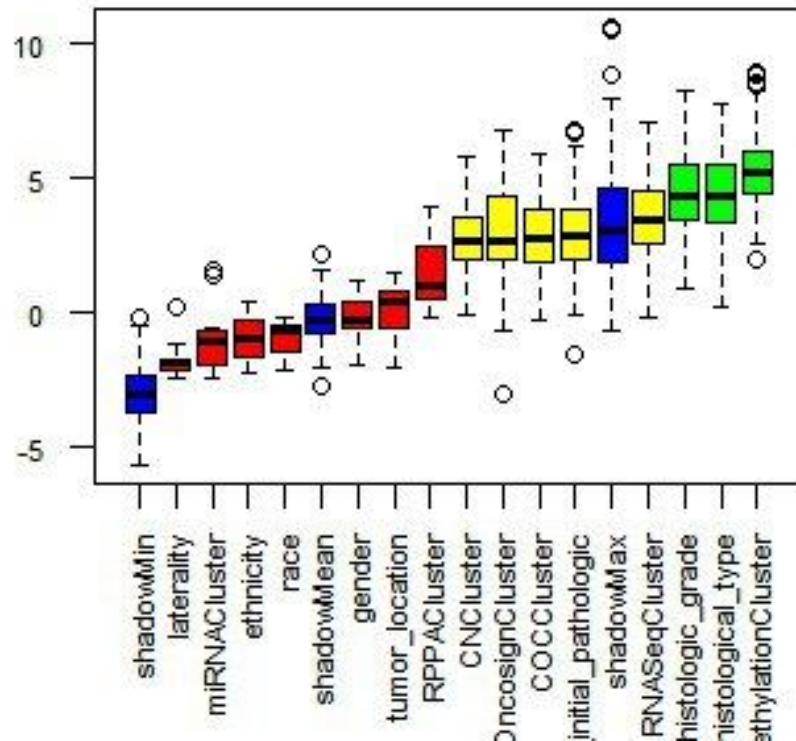


HOW BORUTA WORKS?

1. It creates randomness by creating shuffled copies of all the features (SHADOW FEATURES)
2. Then it will train random forest classifier on the extended data.
3. It applies a feature importance measure (default = Mean Decrease Accuracy)
4. For every iterations, it will check if the real feature is higher importance than the best of the shadow feature. They check this by evaluating the z score. So they check if the feature has a Z score higher than the maximum Z score of its shadow feature.
5. Throughout the iterations it will remove features that are not important
6. The algo stops when all the features are either confirmed or rejected or if it reaches maximum limit of random forest runs

Variable Importance



- Based on our Dataset, it has been confirmed that 3 attributes will be selected (GREEN)
- 7 are Rejected (RED)
- And there are 5 tentative attributes(YELLOW)
 - Run rough fix function
 - Algorithm decides that based on our dataset, 1 tentative attribute can be selected (age_at_initial_pathologic)

```

#plotting the rejected variables to see whi
meanImp  decision
MethylationCluster      5.285224 Confirmed
neoplasm_histologic_grade 4.446440 Confirmed
histological_type        4.390489 Confirmed
age_at_initial_pathologic 2.960639 Confirmed

```