

Weekly Update 04/27/2017 - 05/03/2017

Jeevith Bopaiah

University of Kentucky

May 3, 2017

- Improved the feature selection process using Random Forest Classifier
 - Identified minimal set of features that can predict the final result
- Explored different hyper parameters for AdaBoost Classifier
 - Tested for individual protein clusters.
 - Computed AUC for each cluster
 - Used 5 fold cross validation set to test the accuracy of the model.

- The classifier performed the best on 8 features.
 - nPyrimidines - 0.3291
 - nImidazoles - 0.1093
 - nArCONHR - 0.1084
 - PCR - 0.0717
 - Wi_D - 0.0624
 - TPSA - 0.0388
 - nCbH - 0.026
 - nCb- - 0.0231

AdaBoost Classifier Result

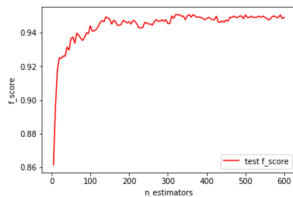


Figure: Protein Cluster 1

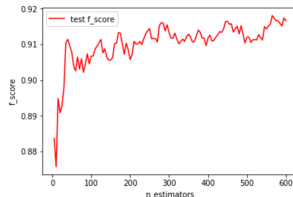


Figure: Protein Cluster 2

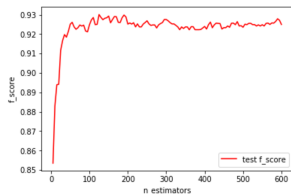


Figure: Protein Cluster 3

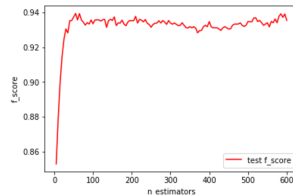


Figure: Protein Cluster 4

AdaBoost Classifier Result

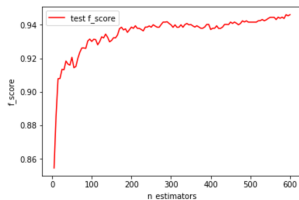


Figure: Protein Cluster 5

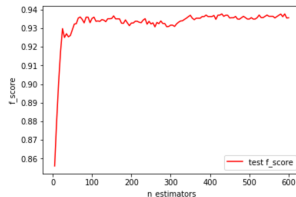


Figure: Protein Cluster 6

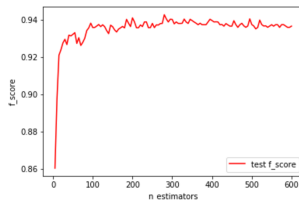


Figure: Protein Cluster 7

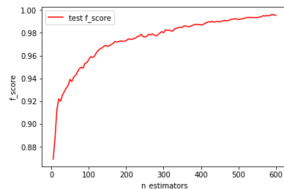


Figure: All Protein Clusters

AdaBoost Classifier Result

- Cluster 1

- Precision: 0.912
- Recall: 0.912
- F1 Score: 0.912
- AUC: 0.97

- Cluster 3

- Precision: 0.941
- Recall: 0.943
- F1 Score: 0.941
- AUC: 0.98

- Cluster 5

- Precision: 0.922
- Recall: 0.941
- F1 Score: 0.931
- AUC: 0.97

- Cluster 2

- Precision: 0.936
- Recall: 0.928
- F1 Score: 0.932
- AUC: 0.97

- Cluster 4

- Precision: 0.90
- Recall: 0.923
- F1 Score: 0.911
- AUC: 0.96

- Cluster 6

- Precision: 0.932
- Recall: 0.929
- F1 Score: 0.930
- AUC: 0.97

AdaBoost Classifier Result

- Cluster 7
 - Precision: 0.90
 - Recall: 0.925
 - F1 Score: 0.912
 - AUC: 0.96
- All Cluster
 - Precision: 0.989
 - Recall: 0.990
 - F1 Score: 0.990
 - AUC: 0.99
- Next Steps:
 - Generate the features for the FDA Approved Drugs
 - Test it using the predictive model.