Data Mining Project 2
Vikas Sable -6473

Preprocessing:

- Not considered % attributes as they are directly co-related to other attributes does not represent useful characteristics for classification.
- Also, here players who have played for less amount of time does not represent true classification. Based on that performed some experiments and found out threshold value for the 'MP' parameter.

Algorithm:

- Tried different algorithms such as Naïve Bayes, Decision Tree, SVM and their variants. Found out that Linear SVM was giving best accuracy for stratified cross-fold validation. Please, find the experiment results bellow.

Results:

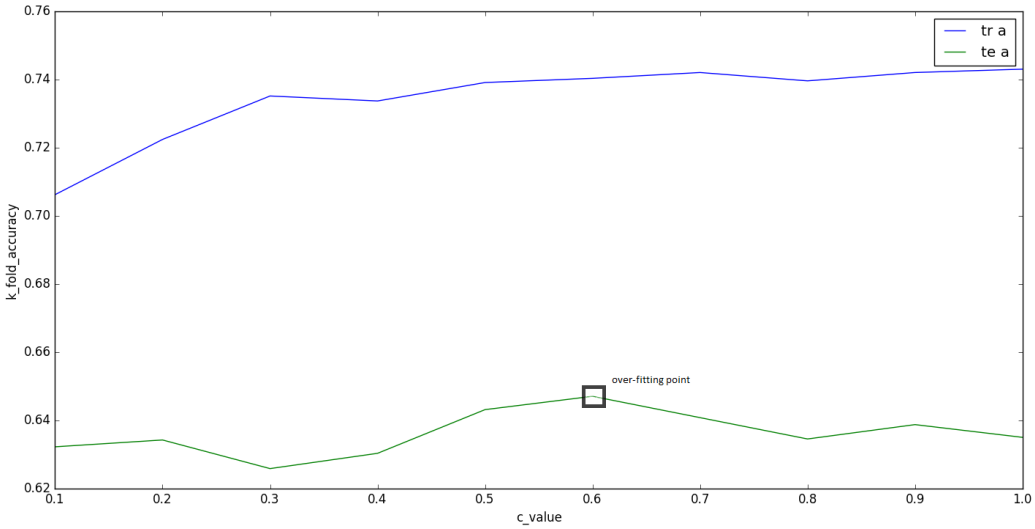
	Regularization parameter C	**Training Accuracy**	**Test Accuracy**
Minutes greater than 5 and all attributes	0.6	0.7402746293541255	0.6470007771841392
Minutes greater than 5 and removed attributes(Age, G, GS)	**0.7**	**0.7265140302636444**	**0.6535375020190006**
Minutes greater than 10 and all attributes	0.8	0.6237385666878529	0.6237385666878529
Minutes greater than 10 and removed attributes(Age, G, GS)	0.3	0.64093616601744	0.7337250153550098

Please, observe the graphs bellow. They are platted with regularization parameter C on X axis and cross-validation training and testing error on Y axis. In each graph there is one point marked which represents the over-fitting for C values.

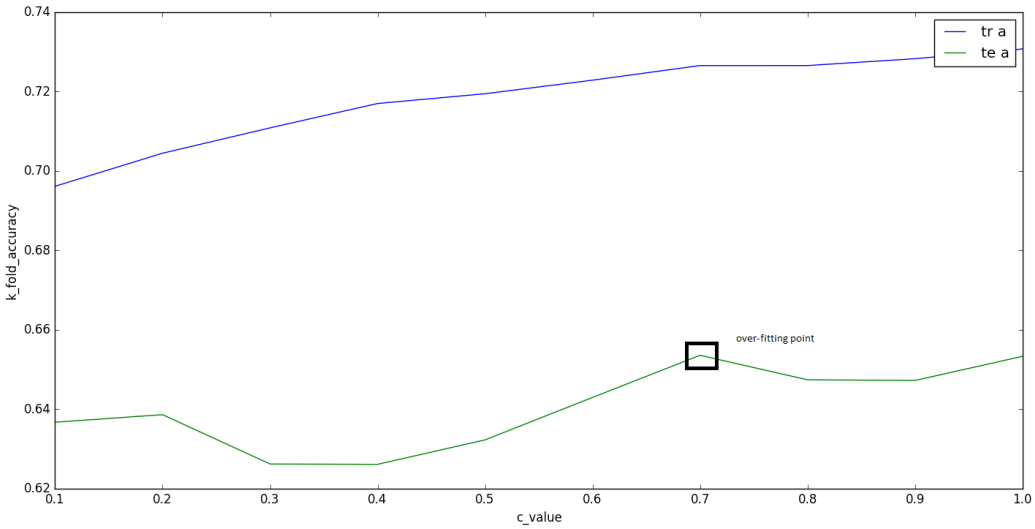
Data Mining Project 2

Vikas Sable -6473

Minutes greater than 5 and all attributes



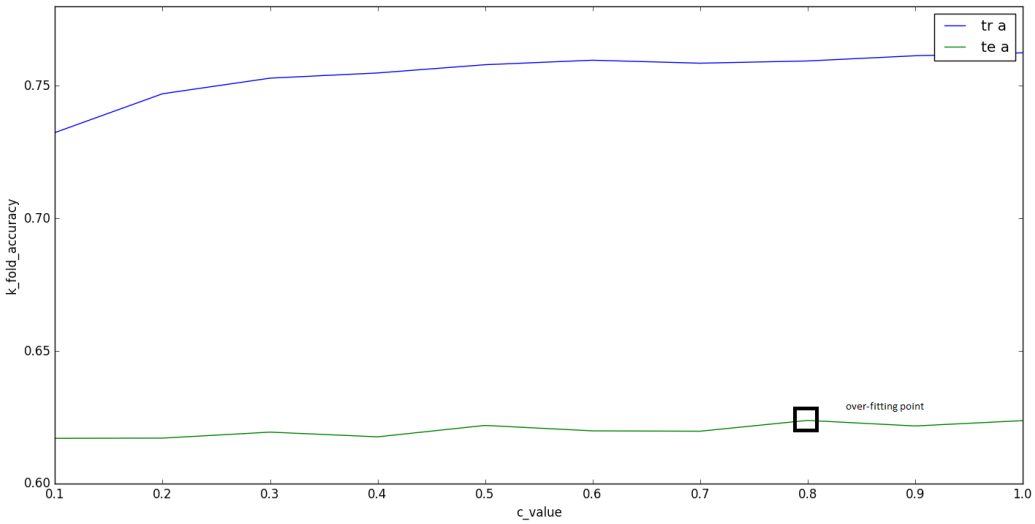
Minutes greater than 5 and removed attributes (Age, G, GS)



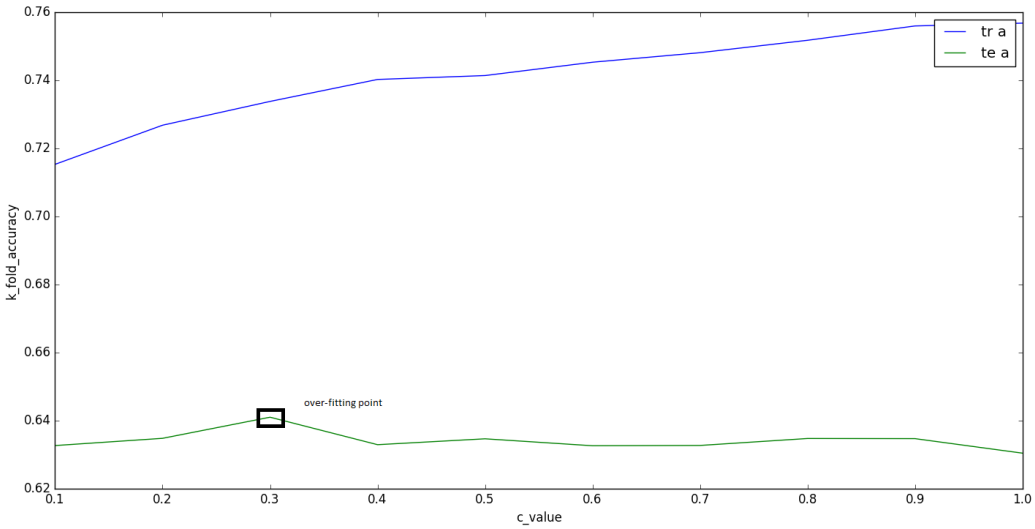
Data Mining Project 2

Vikas Sable -6473

Minutes greater than 10 and all attributes

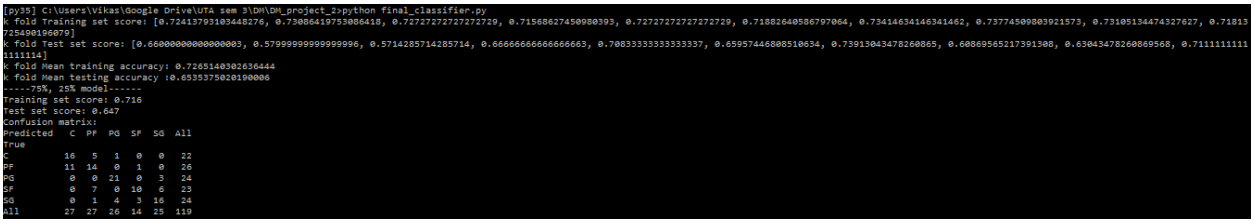


Minutes greater than 10 and removed attributes (Age, G, GS)



Data Mining Project 2
Vikas Sable -6473

Out-put Final Classifier:



References:

- http://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedKFold.html
- http://scikit-learn.org/stable/modules/cross_validation.html
- http://www.svms.org/parameters/
- https://en.wikipedia.org/wiki/Basketball_positions
- http://www.basketball-reference.com/leagues/NBA_2016_per_game.html.
- https://en.wikipedia.org/wiki/Basketball_statistics.
- http://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html#sklearn.neighbors.KNeighborsClassifier
- http://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html#sklearn.naive_bayes.GaussianNB
- http://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html#sklearn.svm.LinearSVC
- http://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html