

## 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:

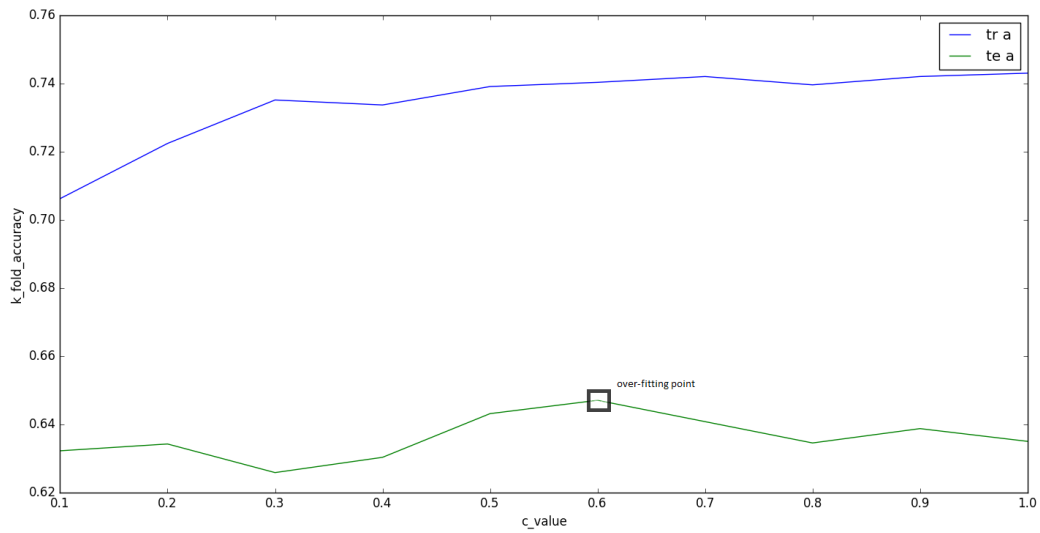
- Tried different algorithms such as Naïve Bayes, Decision Tree, SVM and their variants. Found out that Linear SVM was giving best accuracy for stratified cross-fold validation. Please, find the experiment results bellow.

#### Results:

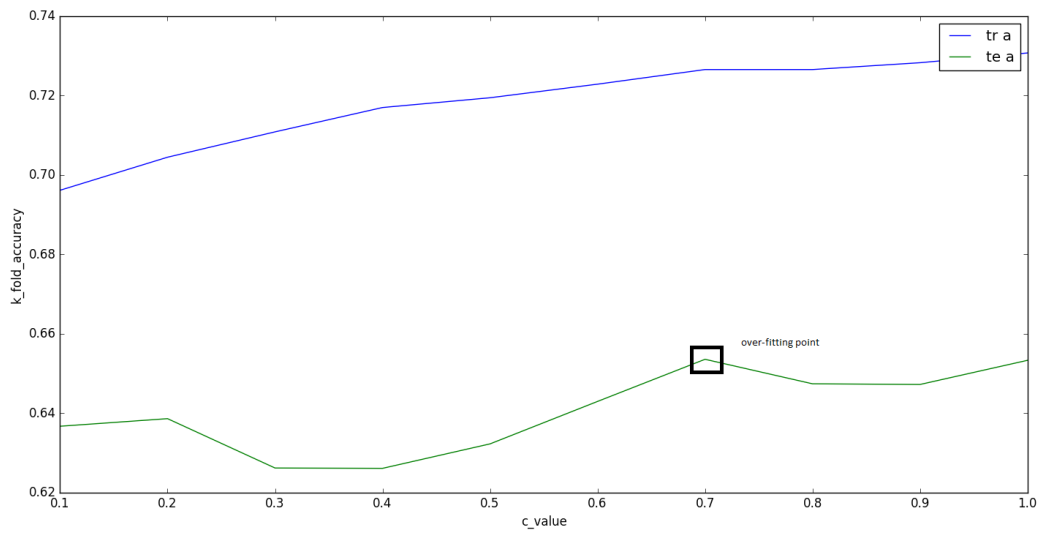
	Regularization parameter C	Training Accuracy	Test Accuracy
Minutes greater than 5 and all attributes	0.6	0.7402746293541255	0.6470007771841392
<b>Minutes greater than 5 and removed attributes(Age, G, GS)</b>	<b>0.7</b>	<b>0.7265140302636444</b>	<b>0.6535375020190006</b>
Minutes greater than 10 and all 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.

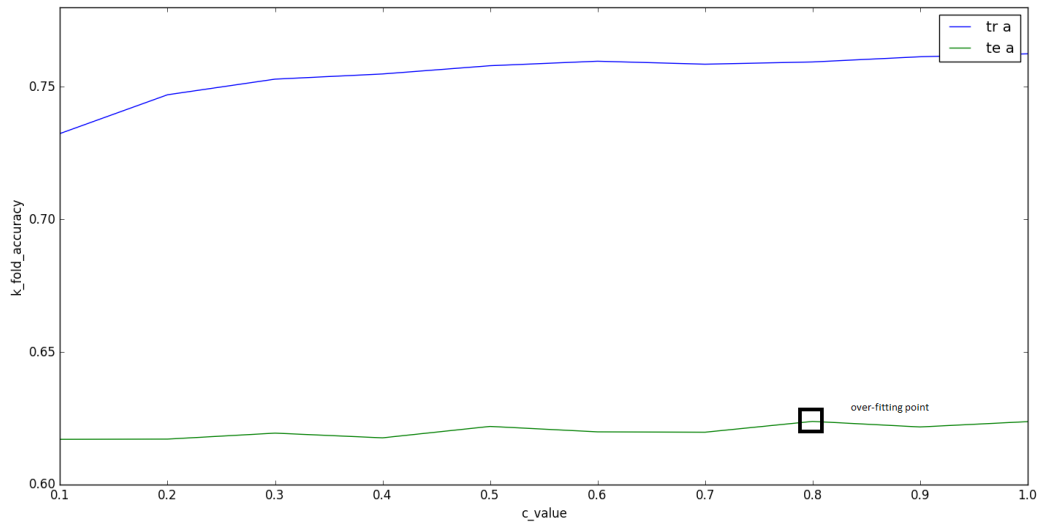
Data Mining Project 2  
Vikas Sable -6473  
**Minutes greater than 5 and all attributes**



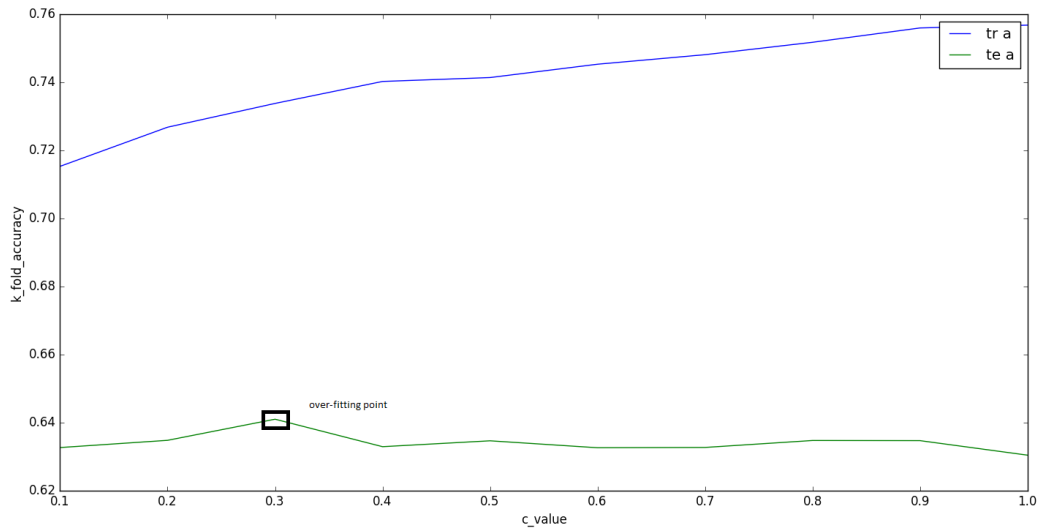
**Minutes greater than 5 and removed attributes (Age, G, GS)**



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:

```
[py35] C:\Users\Vikas\Google Drive\UTA sem 3\DM\DM_project_2\python final classifier.py
k fold Training set score: [0.72413793103448276, 0.73086419733086418, 0.7272727272727279, 0.71568627450980393, 0.7272727272727279, 0.71882640586707064, 0.73414634146341462, 0.73774509803921573, 0.73105134474327627, 0.71813725490196079]
k fold Test set score: [0.66000000000000003, 0.57999999999999996, 0.5714285714285714, 0.6666666666666663, 0.70833333333333337, 0.65957446080510634, 0.73913043478260865, 0.68069505117391306, 0.63043478260869568, 0.71111111111111114]
k fold Mean training accuracy: 0.7265140302636444
k fold Mean testing accuracy :0.6535375020190006
-----75% 25% model-----
Training set score: 0.716
Test set score: 0.647
Confusion matrix:
Predicted C PF PG SF SG All
True
C      16  5  1  0  0  23
PF     11 16  0  1  0  28
PG      0  0 21  0  3  24
SF      0  7  0 10  0  23
SG      0  1  0  3 16  24
All     27 27 26 14 25 110
```

### References:

- [http://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.StratifiedKFold.html](http://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedKFold.html)
- [http://scikit-learn.org/stable/modules/cross\\_validation.html](http://scikit-learn.org/stable/modules/cross_validation.html)
- <http://www.svms.org/parameters/>
- [https://en.wikipedia.org/wiki/Basketball\\_positions](https://en.wikipedia.org/wiki/Basketball_positions)
- [http://www.basketball-reference.com/leagues/NBA\\_2016\\_per\\_game.html](http://www.basketball-reference.com/leagues/NBA_2016_per_game.html)
- [https://en.wikipedia.org/wiki/Basketball\\_statistics](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.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>