**Snehashish Bommali - 1002127072**

**Data Mining Project 2**

**Classifier used: Linear SVC**

Hyper Parameters used – (  
 penalty l1 – regularisation,

C = 1.0 – cost ,

Loss – squared hinge ,

Fit intercept = true – to calculate intercept

Tol = 1e-1 – stopping criteria a solver will stop when there is a change in cost function is less

than the desired value.

Multi\_class = ‘ovr’ – multiclass strategy to one-vs-rest to choose better decision boundary

Max\_iter = 10000 – converges the solution at ten thousand iterations)

**Feature Selection methods:**

I have selected features which gave us the better accuracy – ['FG%', 'FT%', '3P%', ‘2P%’, 'TRB', 'AST', 'BLK', 'PTS', 'TOV', 'FGA']

I have standardised the test and train features with mean as 0 and unit variance with 1.

Feature Importance using the **Random Forest** model – which ensures the features which give us the most purity based on Gini impurity (purity of a node throughout the hierarchy of the tree).

Age: 0.028848979615054365 FT: 0.025312625262631408

G: 0.02667931401335561 FTA: 0.026707624057221723

GS: 0.018280878489898886 FT%: 0.033014243915052435

MP: 0.036068641647579745 ORB: 0.06349640544596484

FG: 0.023304072215996268 DRB: 0.05335732914752016

FGA: 0.03202045074385338 TRB: 0.049938717678566474

FG%: 0.06377285603217878 AST: 0.07314793275256307

3P: 0.0307598205342641 STL: 0.028457637900384643

3PA: 0.054018624065398105 BLK: 0.05614049879823214

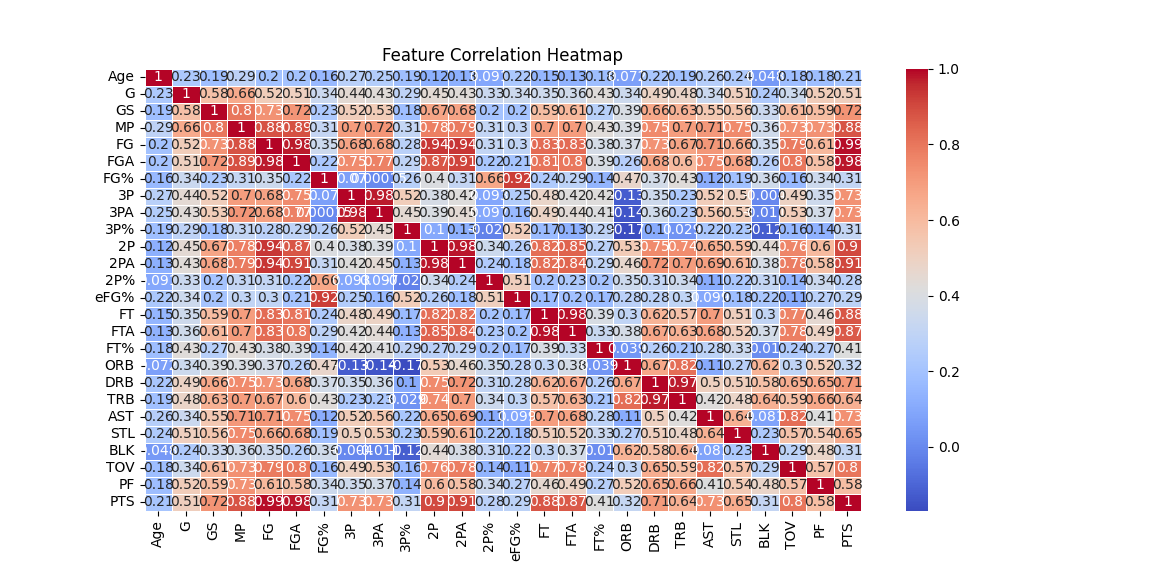
3P%: 0.03459500469239219 TOV: 0.0349569185421585

2P: 0.025986979076683788 PF: 0.029538040033731297

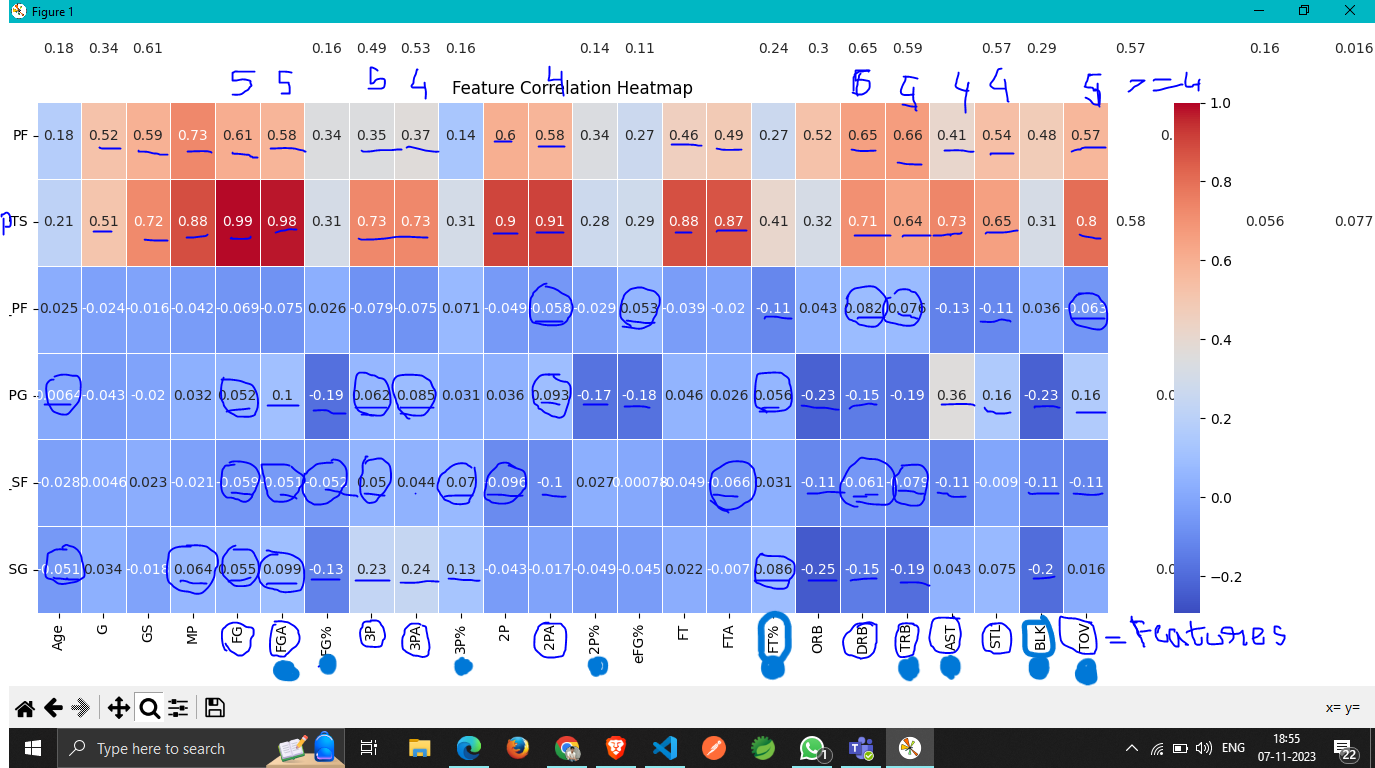
2PA: 0.0330440830240845 PTS: 0.030817105283961258

2P%: 0.045264202164569375 eFG%: 0.04247101486670302

Feature selection using the C**orelation Heatmap** between the features inputs and Pos.



From the above heat map we can identify the features which are corelated with pos.



After keen observation and further research including brute force feature selections, I have taken these features for my model training which eventually gave about

Test Accuracy: 52.9570%

Training Accuracy: 58.4000%

Average Test Accuracy Across All Folds: 52.33%

**Confusion Matrix:**

Rows represent actual values, columns represents predicted values.

Common mapping of positions   
0: Point Guard (PG)

1: Shooting Guard (SG)

2: Small Forward (SF)

3: Power Forward (PF)

4: Center (C)

**Results:**

Task 1:

Training Accuracy: 52.9570

Test Accuracy: 58.4000

----------------------------------------

Task 2:

Confusion Matrix:

0 1 2 3 4 All

0 14 2 1 0 0 17

1 10 7 3 0 11 31

2 0 2 10 0 9 21

3 5 4 2 0 15 26

4 2 2 9 0 17 30

All 31 17 25 0 52 125

----------------------------------------

Task 3:

Accuracy for Each Fold:

Fold 1 Test\_Accuracy: 60.00%

Fold 2 Test\_Accuracy: 42.00%

Fold 3 Test\_Accuracy: 52.00%

Fold 4 Test\_Accuracy: 50.00%

Fold 5 Test\_Accuracy: 54.00%

Fold 6 Test\_Accuracy: 50.00%

Fold 7 Test\_Accuracy: 48.00%

Fold 8 Test\_Accuracy: 53.06%

Fold 9 Test\_Accuracy: 57.14%

Fold 10 Test\_Accuracy: 57.14%

Average Test Accuracy Across All Folds: 52.33%

----------------------------------------