PROJECT-2

6.Explain your method that led to better accuracy, what ideas or observations helped you achieve better accuracy on the dataset?

* Based on my observations for the position classification task, some columns that may not be necessary or could be dropped are <Player ,Tm, G, GS, MP,PTS>
* And the relevant Columns are <Pos, FG, FGA, FG%, 3P, 3PA, 3P%, 2P, 2PA, 2P%, eFG%, FT, FTA, FT%, B, AST, STL, BLK, TOV, PF>
* There are some redundant attributes and calculations such as eFG% = (FG + 0.5 \* 3P) / FGA. 2P% = 2P / 2PA. 3P% = 3P / 3PA. FT% = FT / FTA. AST/TOV = AST / TOV. STL/TOV = STL / TOV.
* So finally Dropping **<'Player','Tm','FT%','Pos','G','GS','PTS','MP','TOV','eFG%'>** attributes for better accuracy.
* Performed Scaling using Minmax scaler is used to scale the input data. scales each feature in the input data to a given range, typically between 0 and 1.
* One more observation is whenever the Minutes played (MP) is less than 15 minutes accuracy is very low and it is hard to predict the position. So, I have added a filter to improve the accuracy. This cutoff helps to filter out players who have only played a few games or for a limited minutes, which could introduce noise into the classification task.
* I have tried different classifiers with different sets of variables, I have noticed the Linear SVM generates the best accuracy for my model. On Regular test data it has **53.01%** and on cross-validation average accuracy **56.50%**

