**ASSESSING PLAYERS VALUE AND DESCRIBING THE METRICS USED**

* Creating a function called ‘evaluate\_abilities’ which takes in player id as an input. This function calculates the average performance of a player before 2014 based on individual player abilities listed on epl\_players.csv which includes the following factors namely crossing , finishing ,header accuracy, short passing to name a few by taking the mean value of these parameters over the matches played during this timeline. To take into account the consistency of a player, all the years of data before 2014 are considered. The previous years i.e 2013,2012,2011 etc, This along with the goals and assists given by the player and the shots they have tried are also taken into account. The shots taken are calculated separately by creating an array called attributes which contains the shot misses as well as the goals. For each player, the goal types and misses are recorded and stored in a dataframe.
* Even though the valuing of players are heavily skewed in favour of the attacking position such as centre forward or attacking midfielders I’ve taken care of the bias by giving relative importance to the attributes that are not exclusive of attacking position such as assists, free kick and distance goals(long range shots). Age of the player is one of the crucial factor that determines his value and I’ve penalized the value of a player who’s on average older than the rest of the player in the year 2014. If the player misses big chance goal opportunity then he’s penalized accordingly. For rare goals scored such as bicycle kicks, free kicks etc the player is awarded points accordingly. The relative importance of each attributes are gauged by a weight vector for each of the attributes mentioned.

**Prediction of match results for season 2015-16**

* A function called gen\_result is defined to create a new column from the given epl\_matches\_train file which has the results of a match. Average team performance over the years has been constructed using a function gen from epl\_teams data keeping specific team attributes such as buildUpPlaySpeed, buildUpPlayDribbling, buildUpPlayPassingetc while leaving others like buildUpPlaySpeedClass, chanceCreationPassingClass etc. We use these team attributes as features and match results as label and create a classification model. Grid search is used to tune the hyperparameters of the given model (multiclass support vector machines,logistic (multinomial) regression, stochastic gradient descent classifier) with cross validation. Then the best model is taken by finding the best value for hyperparameters that increases the accuracy on validation set. Using this model the label for test data is predicted.