horizontal line

**Prakhar Gangwar Rishav Roy**

106116063 106116071

Fantasy Premier League Team of the Week

**CSPE21 - Machine Learning Project Methodology Report**

# OVERVIEW

To obtain the Team of the Week for world’s most watched football league - English Premier league. Players playing in the league get points every week, based on their performance that gameweek. Points are rewarded based on the goals scored, assists made, clean sheets kept and saves made. Points are deducted based on the number of red cards, yellow cards and penalties conceded.

The official website for fantasy premier league is <https://fantasy.premierleague.com/>.

Our aim is to get the team of XI with the highest points every gameweek.

# DATASET

Although the best data-set to use is the official Premier League API, <https://fantasy.premierleague.com/drf/bootstrap-static>, the problem with it is that the data gets refreshed every season. Hence only the in depth per-gameweek analysis of every player is erased. We thus have to use data from other sources like the following.

Season: 2013-14 <https://public.tableau.com/profile/turd1461#!/vizhome/FPL2013-14playerstats8_1/DetailedbyGameweek>

Season: 2014-15 <https://public.tableau.com/views/FPLplayerstats2014-15/DetailedbyGameweek?%3AshowVizHome=no>

# PAPERS REFERRED

Not a lot of work has been done in this field. Even if there is some work done, it is rarely open sourced, since these works would change the betting industry. Also most papers are done on American Football, which is a completely different ball game than “Soccer”. Still, there are a few papers, we will be referring to during the duration of our project.

1. American Football

<http://cs229.stanford.edu/proj2012/Kapania-FantasyFootballAndMachineLearning.pdf>

1. Another one on American Football

<http://cs229.stanford.edu/proj2015/105_report.pdf>

1. Small read on EPL.

<http://blog.minitab.com/blog/starting-out-with-statistical-software/predicting-the-barclays-premier-league-with-regression-analysis>

# RESULTS TO BE ACHIEVED

Since American Football scoring is based on the yards run, it is much easier to come up with a model with high accuracy. We will be using different parameters for different set of players - goalkeepers, defenders, midfielders, attackers. Since we have a wide range of points a player can get, we will use regression methods to obtain their gameweek points.

Unlike the papers we will also take into consideration opposition strength, home/away games, probability of him playing the next game and current form of the player- All of which are present in our dataset.

# METHODOLOGY

1. We will use data from 3 seasons. Out of which 2 will be for training and 1 for testing. On top of this we can apply k-fold cross validation to obtain a better accuracy.
2. We will take into consideration only the players who have more than 50 points in the entire season - removing the outliers.
3. There will be separate code for different player types. We will try out different regression techniques and use the one with best result.
4. In the results list, we will list out the players with score above 8 points.
5. Forming the best team among these players will be done manually.
6. Bagging with other models like SVMs, ANNs, etc to improve performance.

horizontal line