Fantasy Football Team Prediction

https://github.com/277roshan/MachineLearningFinalProject

Team Members
Roshan Thapaliya
Sudheesh Bhattarai
Gauri Shankar Gautam

Project Proposal
Introduction
This project aims to use machine learning technique to predict the top 11 players in Fantasy Premier League, a website that lets you participate in a word-wide players who try to guess the most optimal combination of players who will do well in that gameweek. This is a very interesting topic because we are trying to find from data which player will do well even before the managers and the players know what is going to happen.

Objectives
In this project, we will propose techniques in machine learning to find out which players will perform best. We will measure our success based on the difference between the average score and our predicted score. Similarly, we will also see the difference between the highest score and the score we proposed. We can also see how consistent the prediction is by considering consequent games over a certain period.

Progress:
- We have been able to implement a Naive Bayes algorithm to predict the players for the next week fixtures.
- The implementation includes a prediction weather each player would be able to get more than 8 points in a game week. Getting more than 8 points for a player is considered pretty good for a fantasy player.
- The implementation of the Naive Bayes algorithm gave us a result with a 86.0% accuracy.

Calculation method

For a certain player, say X we have the historical data of how he has performed in past gameweeks.

We implement Gaussian Naive Bayes algorithm based on certain features. For now we experimented with goals scored, total points and clean sheet.

So for each player we have respective decision boundary and we decide if they will get points greater than or equal to 8 in next gameweek.

All the players that the algorithm suggested would get greater than or equal to 8 points in next gameweek were

Set 1

[u'Koscielny', u'Chambers', u'Chamberlain', u'Iwobi', u'Boruc', u'Cook', u'Daniels', u'Ak\xe9', u'Arter', u'Stanislas', u'Gosling', u'King', u'Heaton', u'Keane', u'F\xe0bregas', u'Willian', u'Moses', u'Dann', u'Ward', u'McArthur', u'Baines', u'Coleman', u'Dawson', u'Snodgrass', u'Maloney', u'Schmeichel', u'Mahrez', u'Sakho', u'Lovren', u'Clyne', u'Henderson', u'Man\xe9', u'Kolarov', u'Zabaleta', u'Clichy', u'Silva', u'G\xfcndogan', u'Iheanacho', u'de Gea', u'Jones', u'Shaw', u'Blind', u'Martial', u'Vald\xe9s', u'Gibson', u'Bertrand', u'van Dijk', u'Redmond', u'Ward-Prowse', u'Given', u'Shawcross', u'Adam', u'Whelan', u'Arnautovic', u'Bojan', u'Kaboul', u'Jones', u'Kon\xe9', u'Fern\xe1ndez', u'Sigurdsson', u'Fer', u'Lloris', u'Rose', u'Trippier', u'Chadli', u'Eriksen', u'Alli', u'Wanyama', u'Son', u'Janssen', u'Nyom', u'Capoue', u'Evans', u'Brunt', u'Fletcher', u'Phillips', u'Rond\xf3n', u'Randolph', u'Obiang', u'Payet', u'Feghouli', u'Negredo', u'Zu\xf1iga', u'Gueye', u'San\xe9', u'Guzan', u'Ram\xedrez', u'Llorente', u'Pogba', u'Defour', u'Fabio', u'Hendrick', u'Alonso', u'Slimani', u'Anichebe', u'Niang', u'Brady', u'Gabbiadini']

Players that actually scored

Players who actually scored >=8 in next gameweek were

Set 2

[u'King', u'Heaton', u'Willian', u'Mahrez', u'Sakho', u'Ward-Prowse', u'Eriksen']

For people who have to transfer in players every gameweek Set 1 was a good point to start. Backed by our ML algorithm, they had confidence that these players would score more than 8 points. Set 1 still contains players who are injured or who won't play in next gameweek. This injury update is shown in the website and managers can easily not choose to transfer them in. Likewise after getting set 1 they can also use their own decision to transfer players suited for their team.

Our successful picks

Joshua King
Bournemouth
Midfielder

Form	GW 29	Total	Price	TSB
11.7	8pts	124pts	£5.8	11.9%

Influence	Creativity	Threat	ICT Index
607.0	320.6	750.0	167.7

28	WHU (H) 3 - 2	18	90	3	0	0	2	0	0	1	0	0	0	3

Joshua King whom we correctly predicted would score over 8 got 18 points in gameweek 28.

Willian Borges Da Silva

Chelsea

Midfielder

Form	GW 29	Total	Price	TSB
5.5	10pts	87pts	£7.0	3.4%

Influence	Creativity	Threat	ICT Index
366.2	623.1	394.0	138.0

29	STK (A) 1 - 2	10	87	1	0	0	1	0	0	0	0	0	0	3

Willian scored 10 in his next gameweek.

Riyad Mahrez

Leicester

Midfielder

Form	GW 29	Total	Price	TSB
6.7	8pts	91pts	£8.9	8.5%

Influence	Creativity	Threat	ICT Index
410.2	521.4	629.0	156.4

29	WHU (A) 2 - 3	8	86	1	0	0	2	0	0	0	0	0	0	1

Mahrez scored exactly 8

Mamadou Sakho

Crystal Palace

Defender

Form	GW 29	Total	Price	TSB
9.0	9pts	24pts	£4.7	1.0%

Influence	Creativity	Threat	ICT Index
65.8	1.7	12.0	7.9

29	WAT (H) 1 - 0	9	90	0	0	1	0	0	0	0	0	0	0	3

Sakho scored 9

James Ward-Prowse

Southampton

Midfielder

Form	GW 29	Total	Price	TSB
6.0	10pts	68pts	£5.0	0.4%

Influence	Creativity	Threat	ICT Index
345.4	516.2	310.0	117.5

29	TOT (A) 2 - 1	10	90	1	0	0	2	0	0	0	0	0	0	3

Ward-Prowse score 10

Christian Eriksen

Spurs

Midfielder

Form	GW 29	Total	Price	TSB
7.3	9pts	153pts	£8.7	14.7%

Influence	Creativity	Threat	ICT Index
694.4	1186.6	858.0	273.7

29	SOU (H) 2 - 1	9	85	1	0	0	1	0	0	0	0	0	0	2

Eriksen scored 9

Tom Heaton

Burnley

Goalkeeper

Form	GW 29	Total	Price	TSB
3.7	10pts	119pts	£5.1	19.1%

Influence	Creativity	Threat	ICT Index
767.8	20.3	0.0	78.9

29	SUN (A) 0 - 0	10	90	0	0	1	0	0	0	0	0	0	5	3

Heaton scored 10.

Notable picks who scored nearly 8 from Set 1
1. Chadli scored 5 which is pretty good
2. Coleman scored 6
3. Alli scored 7

Challenges:
- Soccer is a highly unpredictable game with an underdog team beating the League leaders. Hence the prediction is very difficult even based on previous season datas.
- Fantasy Premier League started recently and hence most of the data are very current. Since the algorithm require more data for it to train itself, there is a lack of historical data.
- Data pre-processing is very clumsy.
- Very less data is available during the beginning of a season.
- English Premier League sees many transfers during the transfer window. Players coming newly to the League don't have any data associated with them. They only have a price range to give any estimate.

Lessons learned:
- There are many challenges when the data are not enough to train the algorithm properly.
- Soccer is a highly unpredictable platform with the added constraint that the players may come and go any time. This causes a dearth of data for the algorithm to predict about the player's performance.
- Naive Bayes algorithm provides a general Machine Learning platform. More algorithms are to be implement to get better prediction.

Further Improvements:
- Using K mean clustering algorithm to recognize who will score, get clean sheet, get assists etc in the next match.
- Try to implement and validate results using more algorithms like Logistic regression, Random forest etc

Member's Contribution:
- Roshan Thapaliya, Sudheesh Bhattarai: Implement the Naive Bayes algorithm.

- Gauri Shankar Gautam, Roshan Thapaliya: Data preprocessing algorithm analysis

- Gauri Shankar Gautam, Sudheesh Bhattarai: Testing and tutorial completion