City University of Hong Kong 2022

# Transfer Strategy Optimization for Premier League Clubs

WANG Zhi Xuan[1], LI Xiao Yang[1], LIU Wei[1], HUANG Yu Qin[1], and SHA Xing Chen[1]

[1]City University of Hong Kong

April 13, 2022

## Abstract

Data analysis and algorithms have come to play an extremely important role in the football industry today. Big clubs look to gain a competitive edge on and off the pitch, and big data is allowing them to extract insights to improve the club management, recruitment strategy and their commercial efficiency. We study the performance of teams and players in English Premier League from 2007 to 2017. In addition, We take some English clubs as examples, for which arbitrary transfer optimization strategies will be presented based on data analysis.

**Keywords**
Premier league, Data analysis, Transfer strategy, Decision tree, Clustering algorithm, Data visualization

## 1 Introduction

The study of utilizing data analysis to optimize transfer strategy is inspired by successes of FC Midtjylland and FC Brentford. As the owner of both clubs, Matthew Benham extremely emphasizes the role of data analysis in transferring activity. Now, FC Brentford has successfully promoted, whose valuation and revenue has significantly rised over the years.

It is worth noting that the transfer profit is exactly the core of the club's revenue. The club well utilized the sports data model and cutting-edge data statistical research from Smartods and Matchbook to assist judgment in player recruitment, training and transfer.

For instance, the data model used by FC Brentford includes various recruitment indicators, one of which is expected goals. When selecting a striker, the club will not only emphasize the number of goals one has scored, but also examine the quantity and quality of opportunities one has created and their bonus to the team's performance, which will be reflected in the value of expected goals.

| Brentford's Transfer Market Activity | | | | | | | |
|---|---|---|---|---|---|---|---|
| Player | Purchased | | | Sold | | | Profit |
| | Year | Club | Fee | Year | Club | Fee | |
| Ollie Watkins | 2017 | Exeter City | £1.8m | 2020 | Aston Villa | £33m | **£31.2m** |
| Said Benrahma | 2018 | OGC Nice | £2.7m | 2020 | West Ham | £30m | **£27.3m** |
| Neal Maupay | 2017 | Saint-Ettiene | £1.8m | 2019 | Brighton | £20m | **£18.2m** |
| Andre Gray | 2014 | Luton Town | £0.5m | 2015 | Burnley | £12m | **£11.5m** |
| Scott Hogan | 2014 | Rochdale | £0.8m | 2017 | Aston Villa | £12m | **£11.2m** |
| Chris Mepham | 2016 | Brentford B | £0 | 2019 | Bournemouth | £11m | **£11m** |
| Ezri Konsa | 2018 | Charlton | £2.5m | 2019 | Aston Villa | £12m | **£9.5m** |
| Ryan Woods | 2015 | Shrewsbury | £1m | 2018 | Stoke | £6.5m | **£5.5m** |
| Nico Yennaris | 2014 | Arsenal | £0.2m | 2019 | Beijing Guoan | £5m | **£4.8m** |
| Jota | 2014 | Celta Vigo | £1.5m | 2017 | Birmingham | £6m | **£4.5m** |
| James Tarkowski | 2014 | Oldham | £0.3m | 2016 | Burnley | £4.5m | **£4.2m** |
| John Egan | 2016 | Gillingham | £0.4m | 2018 | Sheffield Utd | £4m | **£3.6m** |
| | | | £13.5m | | | £156m | **£142.5m** |

Figure 1: Brentford's Transfer Market Activity

Football economy is an economic form based on football and its related industries, which is realized through the exchange of players and the circulation of football-related products. In the modern football economy, the most important commodity is indeed the football player.

Clubs consider the potential commercial value of players in addition to their performance. For example, how well he has a reputation among fans, how many jerseys he can sell, whether he can attract more advertising sponsorship through this player, and so on.

In response, a British company called Orb Finance has designed a set of analytical tools to help club managers assess the commercial value of players. For example, after Bale's move from Tottenham to Real Madrid, the software predicts that he will bring £41 million in shirt sales to the club over the next six years, which is close to half of Bale's sky-high transfer fee. In addition, Orb predicts that if Bale's contribution to Real Madrid can reach the level of Tottenham Hotspur which accounts for 11.6% of total goals

of the team, his market value will reach £94 million. Therefore, the commercial value of players is also an important factor that we need to consider during the transfer period.

In this report, we mainly focus on indicators such as goals per match, shoot accuracy, passes per match, cross accuracy, tackle success and clearance per match, etc. In terms of the market value of players, page views, age, fpl points and fpl value will be put more emphasis on instead.

### 1.1 Research Target

We studied the change trend of the ranking of clubs, which have been premier league clubs over the decade, from season 07/08 to season 16/17. In the line chart below, it can be clearly seen that Manchester City has changed from middle class clubs to strong contingents over the decade. In the following discussion, transfer history of Manchester City Football Club will be focused on to figure out how the transformation happens. Apart from that, Sunderland is the only relatively weak club that secured their premier league survival over the decade. Thus, Sunderland FC will not be listed in our research targets.

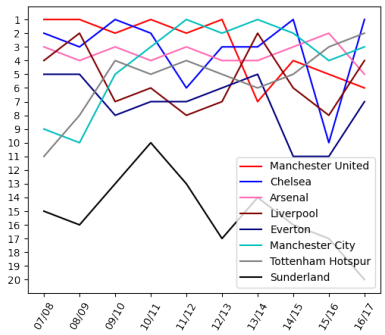



Figure 2: Premier League Ranking Trends

## 2 Methodologies

In this sections, methods used will be presented. Supervised learning methods (e.g., decision tree) are adopted to label the overall performance of premier league players. Besides, we utilize k means clustering algorithm to classify players into typical groups.

### 2.1 Decision Tree

Decision tree is a method of machine learning. Each non-leaf node represents a test on a feature attribute, each branch represents the output of that feature attribute on a range, and each leaf node holds a category. The process of using a decision tree for decision making is to start from the root node, test the corresponding feature attributes in the item to be classified, and select the output branch according to its value until it reaches the leaf node, and the category stored by the leaf node is used as the decision result.

Before annual ranks for players are published every year, it is difficult to tell which players has more potential, especially considering different people may focus on different parts. For examples, some people prefer better performance in the aspect of goals per match, while others pay more attention to shooting accuracy. Then, how to give a fair classification of player's performance level can be a considerable problem if we want to get a head start for our transferring strategy. Normally it is reasonable that different estimations, including clubs, appearance, etc., can all contribute to players' final ranks or performance levels in different degrees, and that is why we need an comprehensive and nature way to use measured data collected from games this year to provide a reasonable prediction about how well the player is. Decision tree is quite a reasonable and ideal model to assist us complete this task.

After data collection and discussion, six attributes are finally chosen as the attributes of training data to construct the decision tree, which are, goals per match, shoot accuracy, passes per match, cross accuracy, tackle success, and the ratio of clearance and appearance.

Moreover, since decision tree is a kind of model which try to learn the structure of training data, find the most interest features with splitting threshold, and assign each record to a purer child model in order to label the node with corresponding decision path, it is obvious that different positions of players can be affected by different features or thresholds, which means it is not satisfactory to construct a general decision tree same for all positions. Therefore, we performed some preprocessing for those training data, including integration, denoising, subsample, and finally splatted them to different training data sets for midfielder, forward, defender, respectively. Here, different from decision tree for market value prediction, we use gini index as the impurity measure in order to simplify the training calculation while still maintain the satisfactory data split performance.

Complexity of decision tree is also an important issue for ideal performance of classification. We design different decision trees to test the performance of our predictive model, and finally choose $maxdepth = 2, 3, 3$ for defender, forwarder, midfielder respectively to avoid over-

fitting and underfitting effects. An example is as shown above.

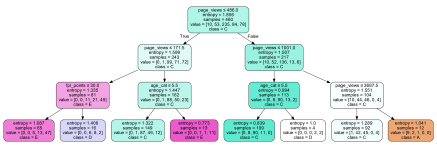


Figure 3: Examples of Decision Tree

In terms of the decision tree for performance analysis, we set up level 1 to 8. In terms of the decision tree for market value evaluation, we set up level A, B, C and D. The classification criterion can be referred in the following diagram.

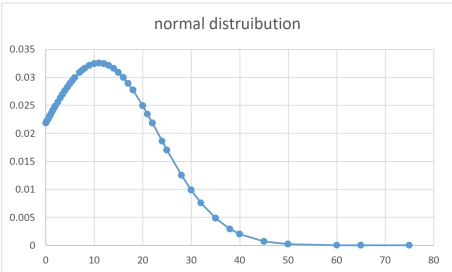



Figure 4: Normal Distribution

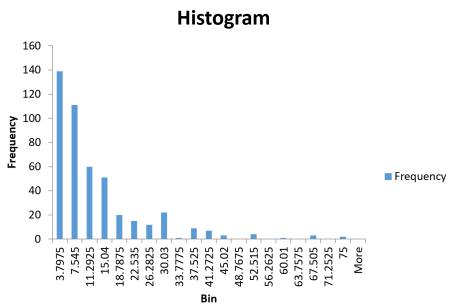



Figure 5: Histogram

## Feature selection

Feature selection determines which features are used to make judgments. In the training dataset, there may be many properties for each sample, and the role of different properties may be large or small. Therefore, the role of feature selection is to screen out features that are more relevant to the classification results, that is, features with strong classification ability. The criterion commonly used in feature selection is information gain.

## Decision tree generation

After selecting the feature, it is triggered from the root node, calculates the information gain of all features for the node, selects the feature with the largest information gain as the node feature, and establishes child nodes according to the different values of the feature; use the same method for each child node to generate new child nodes until the information gain is small or there are no features to choose from.

## Decision tree pruning

The main purpose of pruning is to combat overfitting and reduce the risk of overfitting by actively removing some branches.

## 2.2 K-means clustering

Cluster analysis is a typical unsupervised learning method. In our report, we use K-means clustering algorithm to classify different kinds of players into different clusters.

The motivation is that putting young elements in the teams is a common phenomenon while it is challenging to keep the stability of teams based on such changes. Therefore, building reasonable player classification model is necessary to support extension and recruitment strategy optimization. For instance, amateur premier league clubs are tend to sign young players who have similar characteristics with local franchise players and key players to avoid the tactic inconsistency. In our implementations, we select key players from our research targets as centroids in order to find their potential substitution. For instance, we select Kevin De Bruyne, Sergio Agüero, João Cancelo, Fernandinho and Gabriel Jesus from Manchester City to demonstrate the intuitional idea. The example is illustrated as follows.

## Complexity Analysis

Time complexity of K-means clustering algorithm is $O(inkm)$ and Space complexity of K-means algorithm is $O(nm)$, where $k, i, m, n$ denotes the number of clusters, iterations, dimensions and sample nodes. It can be easily seen that K-means clustering has linear time complexity $O(n)$ over $k$ clusters and number of iterations $i$, which targets at minimizing

$$\sum_{x_i \in S} \sum_{k} z_{i,k} |x_i - \lambda|^2$$

where $z_{i,j} \in \{0, 1\}$, $z_{i,k} = 1$ and $z_{i,j} = 0$ for all $j \neq k$. Euclidean distance is adopted in our implementation.

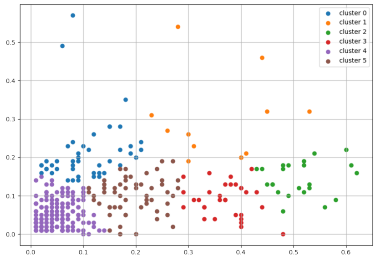



Figure 6: K-means clustering based on score contribution(x axis) and assist contribution(y axis)

| Player Name | Player ID | Cluster ID | xG | xA |
|---|---|---|---|---|
| João Cancelo | 184 | 0 | 0.08 | 0.19 |
| Kevin De Bruyne | 337 | 1 | 0.44 | 0.46 |
| Sergio Agüero | 356 | 2 | 0.54 | 0.21 |
| Gabriel Jesus | 319 | 3 | 0.38 | 0.15 |
| Fernandinho | 144 | 4 | 0.06 | 0.07 |
| Ferrán Torres | 317 | 5 | 0.37 | 0.13 |

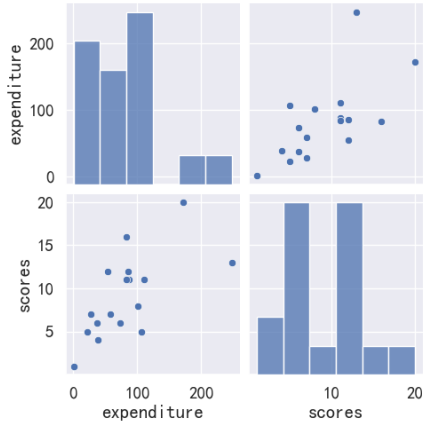
Figure 7: clustering table

## 2.3 Pearson Correlation Analysis

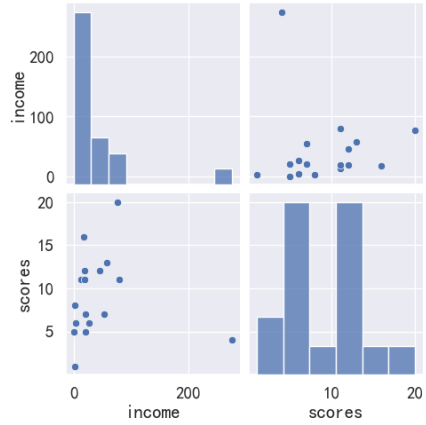
In terms of transferring the strategy of football clubs, the most intuitive indicator to be thought of is correlation. That's because data correlation is to discover the associative relations between data entries. Therefore, if the correlation between the transferring and the competence of clubs is known, the reasons to produce such results could be exploited to form a strategy. For example, suppose a positive linear relationship exists between transferring times and scores in a particular season. In that case, it is possible that the higher the player liquidity is for a club, the higher the winning possibility it has.

To start with, the first and foremost thing to do for correlation checking is to find proper indicators. From diverse sources, we collected many datasets and decided to use the score as the indicator of club capability. For the indicator of transferring, we chose transferring frequency, balance, expenditure and revenue of each club in different seasons. By doing this, we can obtain a general clue about transferring, which could be helpful when we further look deep down into it with more elements considered.

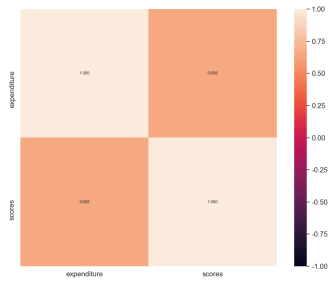
Then, with python libraries pandas and seaborn, we are able to do the raw data processing, correlation calculation and visualization. The results show that the correlation coefficients of the time of member transferring in, the time of member transferring out, balance, expenditure, and income regarding cost are $-0.0148, 0.0105, -0.4366, 0.6582, -0.0389$. The result shows no relationship between frequency and performance, nor income and performance. However, balance and performance have a weak linear relationship, and expenditure and performance have a moderate linear relationship. It can be seen that a club spending more on recruiting players has more chance to win as well as earn more.




(a) correlation coefficient between points and income




(b) correlation coefficient between points and expenditure




(c) Correlation Heatmap

# 3 Results

In this section, we first evaluate signings of Manchester City in the past 3 years. Apart from that, two transfer lists will be presented based on methods introduced previously.

## 3.1 Transfer History

Several significant signings of Manchester City are shown in the figure below.

| From | Player Name | To | Fee |
|---|---|---|---|
| Manchester City | Jack Grealish | Aston Villa | 105.75 |
| Manchester City | Rúben Dias | Benfica | 61.2 |
| Manchester City | Nathan Aké | Bournemouth | 40.77 |
| Manchester City | Ferran Torres | Valencia | 30.15 |
| Manchester City | João Cancelo | Juventus | 58.5 |
| Manchester City | Rodri | Atlético Madrid | 56.43 |
| Manchester City | Riyad Mahrez | Leicester | 61.02 |

Figure 8: Manchester City: Transfer History and Fee(Million)

### Midfielder

Midfielders in the list will be focused on first in this section.

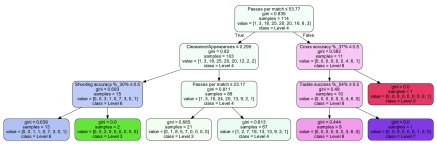


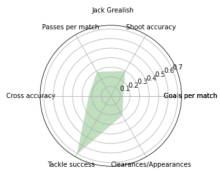
Figure 9: Midfielder, depth = 3

On 5 August 2021, Manchester City announced that they had signed Grealish on a six-year contract that would run until 2027.It was reported by numerous outlets that the transfer fee paid to Aston Villa was for the amount of £100 million, which constituted the most expensive transfer of an English player ever, as well as the highest fee ever paid by a British club.Grealish was given the number 10 shirt by the team which had previously been worn by club legend Sergio Agüero, who departed from City the month prior after ten seasons with the team.

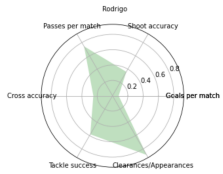
According to the statistics, taking Jack Grealish as examples, his performance can be quantified as follows. (Appearnce = 71, Goals per match = 0.13, Shoot accuracy = 0.30, Tackle success = 0.71, Clearance = 17, Passes per match = 29.55, Cross accuracy = 0.23) According to our decision tree analysis, Jack Grealish will be rated as level 6 (refer to figure 9).

On 3 July 2019, Manchester City met the terms of Rodri's £62.6 million release clause, enabling him to buy out the remainder of his contract with Atlético and leave the club. The transfer was a new record transfer fee paid by Manchester City. He signed a five-year contract.

The performance of Rodri can be quantified as follows. (Goals per match = 0.08, Shoot accuracy = 0.36, Tackle success = 0.58, Clearance per match = 0.888889, Passes per match = 74.25, Cross accuracy = 0.23) Similarly, Rodri will be rated as level 7.




(a) Jack Grealish




(b) Rodri

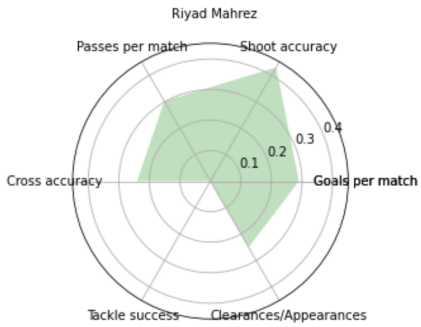
### Striker

In terms of strikers, the decision tree analysis model is illustrated as follows.

On 10 July 2018, Manchester City confirmed the signing of Mahrez on a five-year contract. The transfer fee of £60m made Mahrez the most expensive African footballer, and also Manchester City's most expensive signing and a record transfer fee received by Leicester City.

According to our analysis, Riyad Mahrez is rated as level 6. (Goals per match = 0.29, Shoot accuracy = 0.43, Tackle success = 0, Clearance per match = 0, Passes per match = 30.25, cross accuracy = 0.24)

(c) Striker, depth = 3




(d) Mahrez

### Defender

In terms of defenders, the decision tree analysis model is illustrated as follows.

On 7 August 2019, Cancelo joined Premier League champions Manchester City in a six-year deal worth £27.4 million plus Danilo being sent to Juventus in part-exchange, equalling to £60 million, making him the most expensive right-back ever.

According to our analysis, João Cancelo is rated as level 7. (Goals per match = 0, Shoot accuracy = 0, Tackle success = 0.60, Clearance per match = 0.941176, Passes per match = 59.41, Cross accuracy = 0.15)

Meanwhile, Nathan Aké is rated as level 6. (Goals per match = 0, Shoot accuracy = 0, Tackle success = 0.63, Clearance per match = 0.514966, Passes per match = 38.33, Cross accuracy = 0.2)

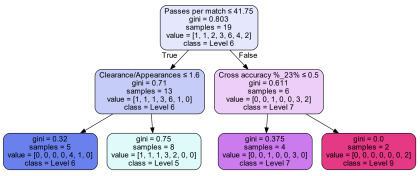



Figure 10: Defender, depth = 2

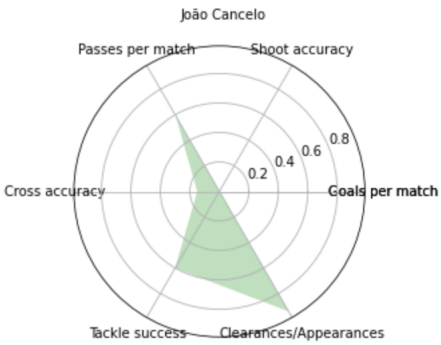



Figure 11: Cancelo

Apparently, defenders with high value of passes per match are more dominant in our machine learning rating criterion, which reflects that modern football tactics have increasingly higher requirement for the passing abilities of traditional center back players.

### Market Value

We studied the market value attributes of some of the sample nodes with decision tree analysis model introduced in previous sections, which is listed below.

(a) Market Value Prediction Model

| Player Name | Market Value(million) | Age Group | Page View | FPL Value | FPL Points | Level |
|---|---|---|---|---|---|---|
| Riyad Mahrez | 30 | 3 | 1753 | 8.5 | 120 | B |
| Nathan Ake | 8 | 2 | 1274 | 5 | 55 | B |
| Kevin De Bruyne | 65 | 3 | 2252 | 10 | 199 | B |
| Sergio Aguero | 65 | 4 | 4046 | 11.5 | 175 | A |

(b) Result

We set up 4 levels of player market value, which is A, B, C, D in descending order. It can be easily seen that the recruitment operations of Manchester City Football Club perform excellent under our evaluation mechanism.

## 3.2 Transfer Recommendation

In order to select the optimized recruitment targets, on one hand, we aimed at those who perform the best in our evaluation mechanism. On the other hand, we aimed at the best fit substitutions to the current squad of Manchester City by clustering analysis.

We sift the data set of forward player, midfielder player and defender player through the following conditions respectively, according to the decision tree(refer to figure 7, 8).

### Striker

Lacking of excellent center forward has been regarded as the most challenging problem for Manchester City over the seasons.

$$passes \quad per \quad match > 21.275(337 \quad results)$$

$$goals \quad per \quad match > 0.305(16 \quad results)$$

It is found that the filtering result among 16 players includes Sergio Agüero and Raheem Sterling. On 28 July 2011, Manchester City confirmed that Agüero had signed a five-year contract with the club. The fee was reported to be in the region of £35 million. Besides, on 12 July 2015, a deal was agreed for the transfer of Raheem Sterling to Manchester City for an initial £44 million, with a further potential £5 million in add-ons, subject to personal terms and a medical, which would make him the most expensive English player of all time. Both of the signings prove the correctness of our evaluation model.

Eventually, the recommended signings are listed below(rating in descending order), where players from Manchester City are excluded.

| Name | Jersey Number | Club | Nationality | Age |
|---|---|---|---|---|
| Mohamed Salah | 11 | Liverpool | Egypt | 28 |
| Sadio Mané | 10 | Liverpool | Senegal | 28 |
| Raúl Jiménez | 9 | Wolverhampton Wanderers | Mexico | 29 |
| Anthony Martial | 9 | Manchester United | France | 24 |
| Son Heung-Min | 7 | Tottenham-Hotspur | South Korea | 28 |
| Roberto Firmino | 9 | Liverpool | Brazil | 28 |
| Christian Benteke | 17 | Crystal Palace | Belgium | 29 |
| Gareth Bale | 9 | Tottenham Hotspur | Wales | 31 |

Figure 12: Striker Recommendation

### Midfielder and Defender

$$passes \quad per \quad match > 53.77(48 \quad results)$$

$$cross \quad accuracy > 0.5(1 \quad results)$$

After filtering, Eric García is the only recommended players for Manchester City. In reality, García moved from Barcelona at the age of 17 to Manchester City. He made his Premier League debut on 21 September 2019. In his final season at the club, he won the league and was runner-up in the UEFA Champions League, which proved the success of this transfer. It is worth noting that our transfer recommendation highly overlaps with the actual transfer of Manchester City.

### Commercially Valuable Players

$$page \quad view > 3697.5(13 \quad results)$$

$$age \quad category \leq 1.5(2 \quad results)$$

Except Gabriel Jesus, who is already in the squad of Manchester City, only Dele Alli from Tottenham hotspur satisfies our requirements. In the process of filtering, we found that players' social media views and age are the most significant factors other than their performance and previous market value, in terms of their current market value.

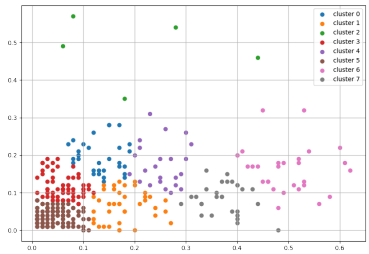
### Substitutions

In this section, we mainly focus on how to utilize clustering analysis to find the optimized substitution for the current roster of Manchester City.

We select 8 key players of Manchester City from the analysis results mentioned above as centroids. Attributes tuple $(xG, xA)$ and is defined, where xG, xA denote score contribution, assist contribution. We use k means clustering algorithm to classify players into several categories in terms of both attribute tuples respectively.

The results are shown below.

Some clusters can be clearly identified as certain types of players while some may not. For instance, cluster 5 can be identified as traditional defender players and cluster 6 can be defined as free-scoring strikers with high efficiency.




(a) clustering based on (xG, xA)

| Name | Player ID | Cluster ID | xG | xA |
|---|---|---|---|---|
| Sergio Agüero | 356 | 6 | 0.54 | 0.21 |
| Kevin De Bruyne | 337 | 2 | 0.44 | 0.46 |
| Raheem Sterling | 334 | 7 | 0.43 | 0.17 |
| Riyad Mahrez | 292 | 4 | 0.26 | 0.27 |
| Bernardo Silva | 223 | 0 | 0.12 | 0.16 |
| João Cancelo | 184 | 3 | 0.08 | 0.19 |
| Rodri | 176 | 1 | 0.08 | 0.06 |
| Kyle Walker | 8 | 5 | 0.01 | 0.02 |

(b) cluster table

For simplicity, we take Kevin De Bruyne ($playerID = 337$, $clusterID = 2$)in the figure above as an example. We identified cluster 2 as creative midfielder player. Other players in cluster 2 can be potential substitution for Kevin De Bruyne, playing similar role as creative midfielder in the match.

| Name | Club | Age | xG | xA |
|---|---|---|---|---|
| Cengiz Ünder | Leicester City | 23 | 0.08 | 0.57 |
| Alireza Jahanbakhsh | Brighton | 26 | 0.28 | 0.54 |
| Xherdan Shaqiri | Liverpool FC | 28 | 0.06 | 0.49 |
| Jack Grealish | Manchester City | 24 | 0.18 | 0.35 |

Figure 13: Potential substitution for Kevin De Bruyne

As expected, an interesting fact is that Jack Grealish, who was regarded as the most influential signing of Manchester City in 2021, is in our list of one of the best substitution for Kevin De Bruyne.

## 4 Discussion

To test our work, we conducted an experiment with Fantasy Premier League. Fantasy Premier League (FPL) is an officially held online competition in which players compete against other players around the world by selecting "fantasy" teams and scoring points based on the actual performance of their chosen players. With more

than 6 million competitors, it stimulates the formation of a Premier League team, with each player having to build a strong team on a limited budget based on their own analysis of player stats.

## 4.1 Random Sample Team

We started with a random team. It can be easily seen that the random team did not perform well, ranking near the bottom of nine million teams. We make some transferring strategies to improve this team.

| | |
|---|---|
| Overall Rank: | 9,034,289 |
| Total Players: | 9,088,492 |

(a) Ranking of the random team for the first week

(b) The roster of random team

## 4.2 Undervalued Player

We trained a neural network to estimate the market value of players based on the official statistics. Then we compare it with the current market price on Fantasy Premier League and select some of the high quality players that are undervalued by the market. In the picture below, the y axis is the loss function value between the real market prices and our predictions. The x axis is the number of training epochs. After training 1000 epochs, the loss becomes relatively steady and small. Therefore our model learns how to estimate the market price according to their performance in the game well.

## 4.3 Transferring strategy

We select the players through the previous decision tree and statistics to estimate their performance of the players. Then we transfer players from existing teams who are expensive but not

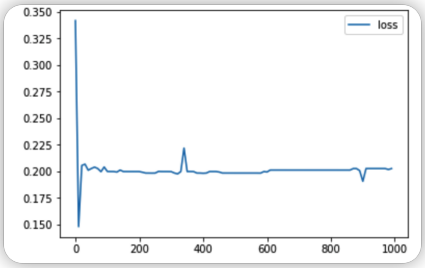
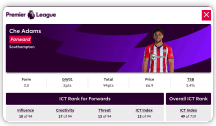



Figure 14: Training result

expected to perform well to high quality players. For example, we compare the two forwards, Che Adams and Christian Benteke. We found that Benteke is dramatically underestimated by the market and has better expected performance according to our decision tree than Adams who is currently on our team. The official statistics also prove our decision. Benteke only cost £6.0 while Adams will cost £6.9. Benteke also has a higher ICT index, which is an official index from the premier league where a higher ICT index implies better performance. Therefore we decide to replace Adams with Benteke as our center forward.



(a) Adams

(b) Benteke

We continue to do several transfers like this. The results showed that in the first game week after the transfers, the team ranks significantly improved.

| | |
|---|---|
| GW Rank | 5,991,946 |

Figure 15: Ranking after transfer

Over the next few weeks, those underrated players we picked did well in the real world, making their power indicator go up and pushing the team up the rankings. Then we go on to get lower-priced but higher-quality players by transferring higher-priced players and based on previous results, the more expensive a player is, the better he tends to perform. So that's how we continue to increase the average value of the players on the team and improve the team rankings.

## Conclusions

Over the next few weeks, those underrated players we picked did well in the real world, making their power indicator go up and pushing the team up the rankings. Then we go on to get lower-priced but higher-quality players by transferring higher-priced players and based on previous results, the more expensive a player is, the better he tends to perform. So that's how we continue to increase the average value of the players on the team and improve the team rankings.

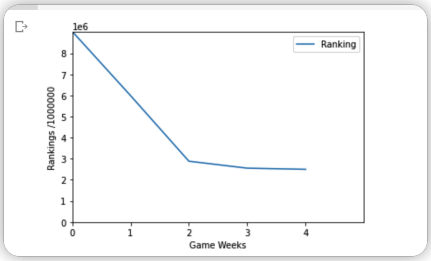



Figure 16: Conclusion

## Acknowledgements

The authors thank anonymous reviewers for the constructive comments. WANG Zhi Xuan, LIU Wei, LI Xiao Yang, HUANG Yu Qin and SHA Xing Chen are supported by City University of Hong Kong. Data sets are mainly from kaggle and shared by anonymous reviewers. The authors would like to thank the open source resource shared on the internet. Related transfer records are from the official website of Premier League.

| Name | EID | Contribution |
|---|---|---|
| WANG Zhi Xuan | zhixuwang4 | 20%(Decision tree, Visualization, Report) |
| LI Xiao Yang | xyli45 | 20%(K means clustering, Coordination, Report) |
| LIU Wei | wliu283 | 20%(Decision Tree, Report) |
| HUANG Yu Qin | yqhuang23 | 20%(Correlation Analysis, Report) |
| SHA Xing Chen | xingchsha2 | 20%(Neural Network, Report) |

Figure 17: Contribution

## References

[1] Baboota, R., Kaur, H. (2019). Predictive analysis and modelling football results using machine learning approach for English Premier League. International Journal of Forecasting, 35(2), 741-755.

[2] Bush, M., Barnes, C., Archer, D. T., Hogg, B., Bradley, P. S. (2015). Evolution of match performance parameters for various playing positions in the English Premier League. Human movement science, 39, 1-11.

[3] Freitas, P. (2021, May 4). All Premier League matches 2010-2021. Kaggle. Retrieved March 27, 2022, from https://www.kaggle.com/datasets/pablohfreitas/all-premier-league-matches-20102021

[4] Nalla, Z. (2018, July 12). Premier league. Kaggle. Retrieved March 27, 2022, from https://www.kaggle.com/datasets/zaeemnalla/premier-league

[5] ShubhamMaurya. (2017, August 3). English Premier League Players Dataset, 2017/18. Kaggle. Retrieved March 27, 2022, from https://www.kaggle.com/datasets/mauryashubham/english-premier-league-players-dataset

[6] Vaastav. (n.d.). Vaastav/fantasy-Premier-League: Creates A .CSV file of all players in the English player league with their respective team and total fantasy points. GitHub. Retrieved March 27, 2022, from https://github.com/vaastav/Fantasy-Premier-League

[7] "Joao Cancelo: Man City complete deal for Juventus full-back as Danilo moves other way". BBC Sport. 7 August 2019. Archived from the original on 7 August 2019. Retrieved 7 August 2019.

[8] "Riyad Mahrez: Overview". Premier League. Retrieved 28 May 2021.

[9] Wilson, Paul; Jackson, Jamie (4 July 2019). "City's new £62.8m signing Rodri says Manchester may be getting 'more blue'". The Guardian. Retrieved 19 February 2020.

[10] "2021/22 Premier League squads confirmed". Premier League. 10 September 2021. Retrieved 12 September 2021.