**Spring 2020 INFO6205 Project**

Longtao Ling 001054615

Shuhang Li 001053704

Tong Yang 001302930

(alphabetical order)

****

**Ranking system of prediction of EPL unfinished season**

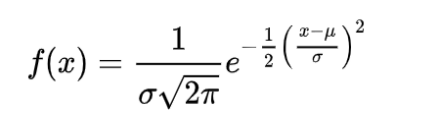
**Introduction**

We all know that the EPL is one of the most famous football leagues in the world. However, due to the big outbreak of the COVID-19, the world stopped its path, as well as the football leagues like EPL. Although Liverpool must end the season at the top of the table (it is mathematically impossible for any other team to pass them), the next five positions are important for the summer, as are the last three positions, which teams will be relegated. We created a system that can basically predict the ranking table of EPL as that you can see at the end of the season base on Normal distribution. Also we make the prediction that not only conclude the ranking score of teams, but the win rate of each 2 teams which is not easy. We have also considered the difference effect of HomeTeam and AwayTeam game. Our system can output tables of predict of win rate, final ranking board of win draw lose and scores.

**Normal Distribution**

In probability theory, a normal (or Gaussian or Gauss or Laplace–Gauss) distribution is a type of continuous probability distribution for a real-valued random variable.

The general form of its probability density function is



The normal distribution is the distribution of continuous random variables with two parameters μ and σ2. The first parameter μ is the mean value of the random variable following the normal distribution, and the second parameter σ2 is the variance of this random variable, so the normal distribution record as N (μ, σ2). Normal distribution is very common in nature. Normally, If an index is jointly influenced by several independent factors, and each factor cannot produce a dominant influence (Lindbergh condition), this index will obey the central limit theorem and converge to a normal distribution. In our project, the number of goals by each team is influenced by multiple factors, such as weather, home and away, injuries, etc. So using normal distribution is very suitable in our project.

**Aim of the project**

Design and implement a simulation for English Premium League (EPL)

● Calculate the win probability that team A will beat team B if they meet each other head to head match-up, and divide into home and away circumstance.

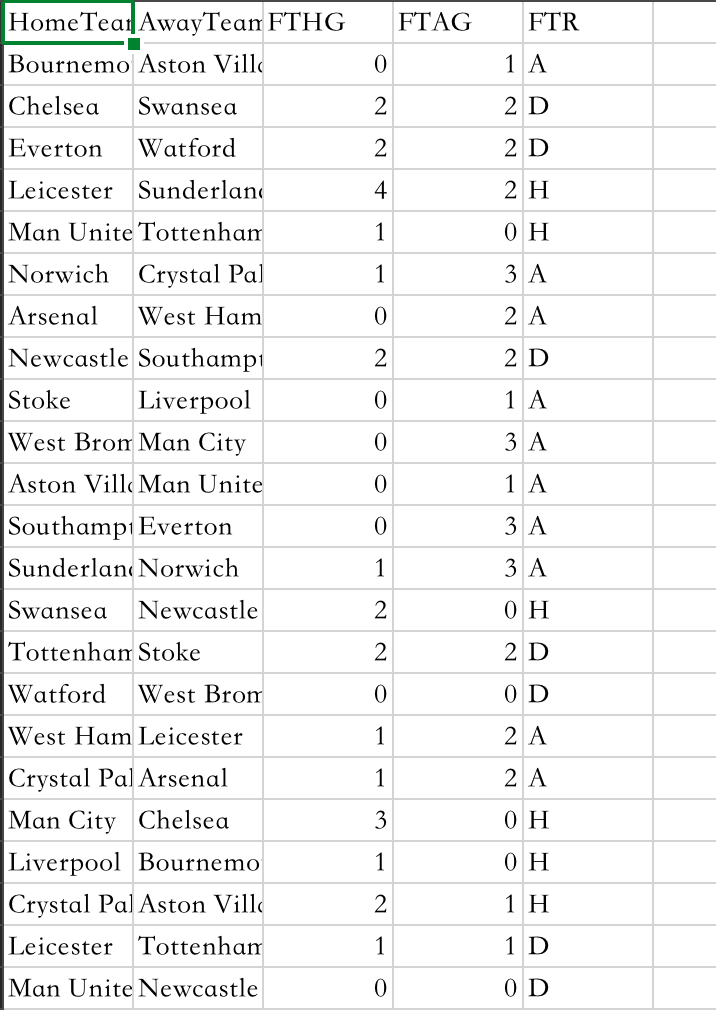
● Predict the remaining game of 19-20 season based on last five year data.

● Export a ranking table by their final score.

**Data used in this project**

We use English Premier League 2015-2020 result in our project.(In each row, we have home team, away team, home team goals, away team goals and wins or losses,please check it in our data file match.csv)

**A part of data in 5 year’s match:**



For the prediction of the remaining games,we get the schedule of it (In file unplayed.csv):

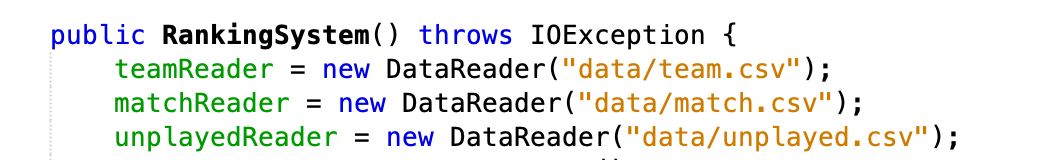


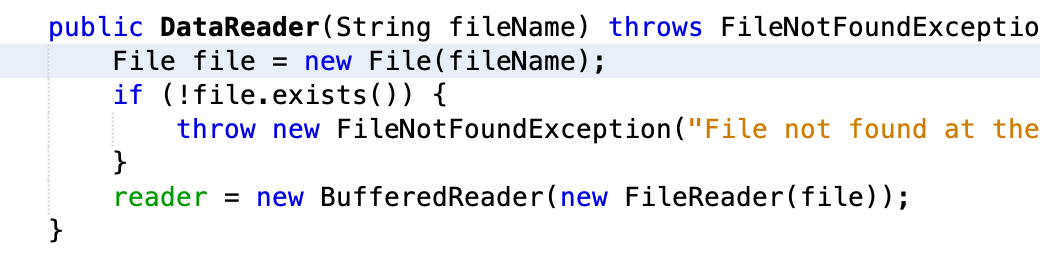
**Implementation:**

We totally have six class in our project.

|  |  |
| --- | --- |
| Class name | Class description |
| Analysis | Class use normal distribution to calculate and predict the result |
| DataReader | Read information from csv |
| DataWriter | Generate the result of win rate and rank |
| RankingSystem | The main class of the project |
| Match | Consist of all match information |
| Team | Consist of all team information |

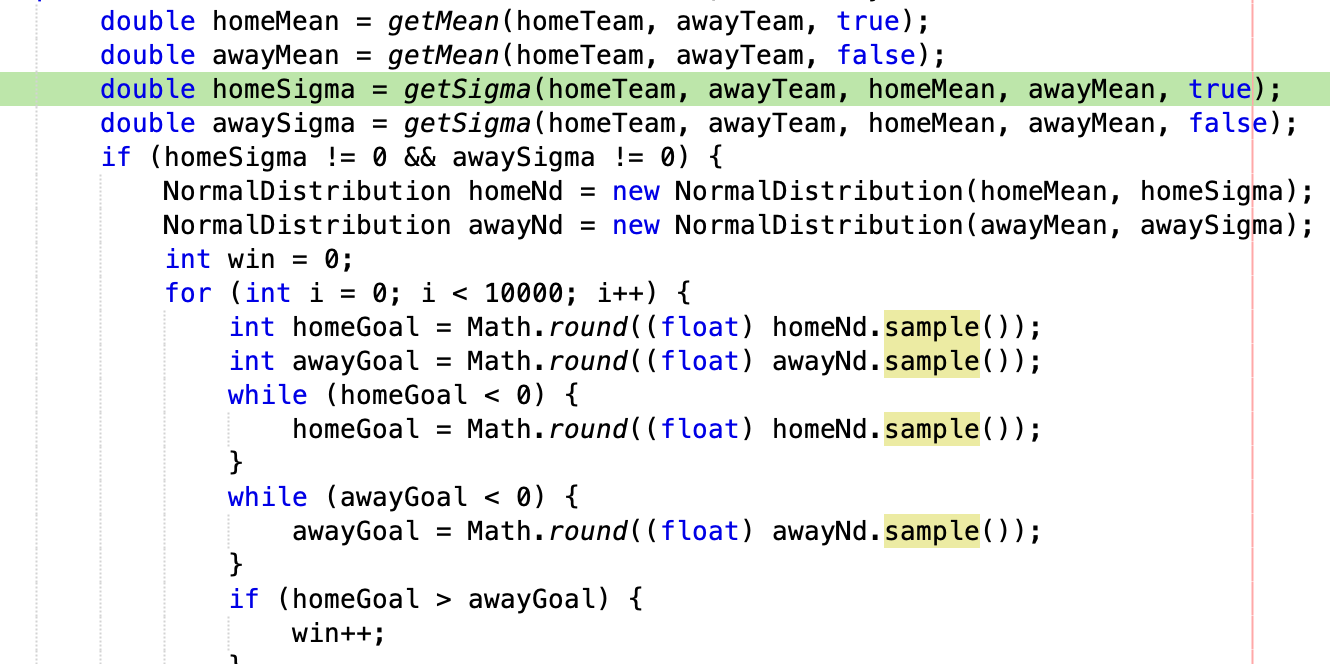
First,we use BufferedReader to import team, match, unplayed match data.





The data is generate and store in the static teamList, matchList and unplayedList.

Then we import Normal Distribution package to analysis the data. For each two team in the chart, we respectively calculate its mean and standard deviation at home and away.



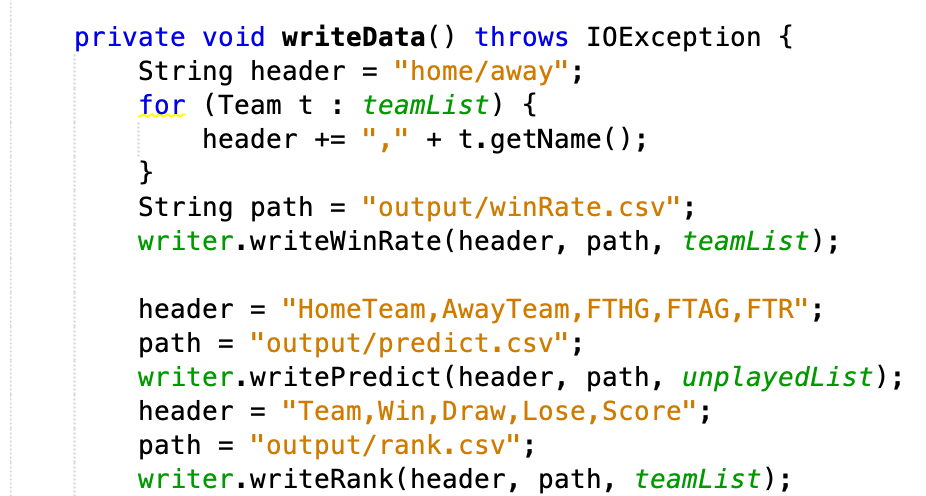
Due to the normal distribution, the goal may generate as negative , since the goal cant below 0, we have fault tolerance to regenerate the data. And In this season, since some team is the first time appear in the English Primer League in the past five year, and in the match history, it only have one record, so there not exist standard deviation, we can’t use the same algorithm to calculate it. So we divide into four circumstance.

|  |  |
| --- | --- |
| homeSigma != 0 && awaySigma != 0 | homeSigma == 0 && awaySigma != 0 |
| homeSigma != 0 && awaySigma == 0 | homeSigma == 0 && awaySigma == 0 |

For the prediction of the remaining game, we also use normal distribution to get the goals of each team, and compare the goals of each team to get which team win.For every win game, the team get three, every draw get one and get zero if they lose. And then use Math,round to get the goals as integer.

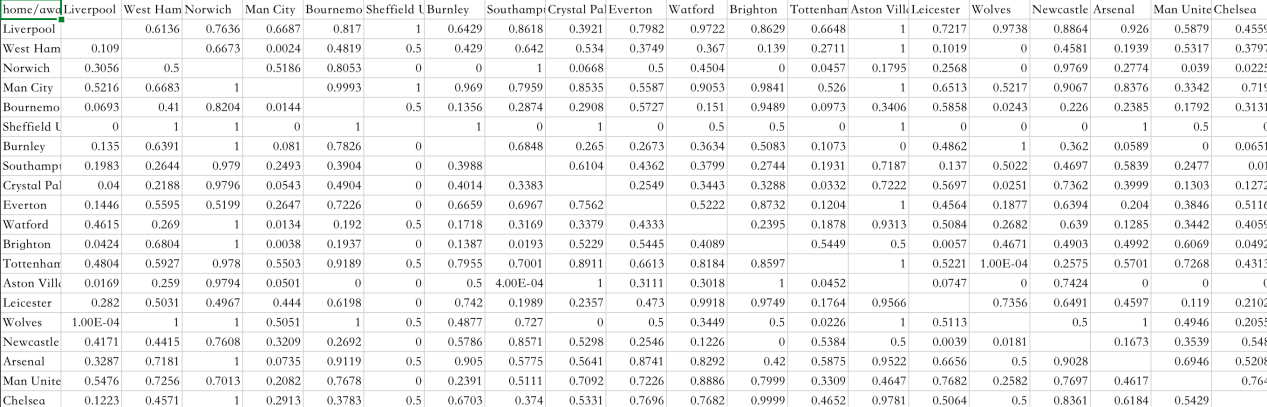


Finally, we generate csv to show output.



**Output:**

**Win Rate Chart:**



**Unplayed match prediction:**



**Final Rank:**



**Conclusion:**

● Since the number of goals is influenced by many factors, we find it appropriate to use the normal distribution model and then compare the number of each team goals to determine which team wins

● As everyone expected, Liverpool will win their first Premier League title

● However, football match cannot be easily predicted by number of goals, There are too many uncertain factors on the field, like key player’s performance, sudden injury, whether the team participates in gambling etc, so our predictions are insufficient to some extent.