Huarui Lu, Zhiwei Li, NUID:001050438

Ranking System

Spring 2020 INFO6205 Project

**Team Member**

The prediction of English Premier League in 2019-2020 season.

# Introduction

## What is Ranking System?

A ranking is a relationship between a set of items such that, for any two items, the first is either 'ranked higher than', 'ranked lower than' or 'ranked equal to' the second.[1] A ranking system is a system that analyzes the input provided by users to provide the ranking for each element.

Our task is to develop a ranking system which is able to evaluate the following expression where are elements from a set of competing elements :   
where is the probability that would beat if they met in a head to head matchup at neutral territory.

In this project, we created a ranking system for English Premier League (EPL). The input to our system is a set of prior encounters with a result. These results are scores of each game. The output is the probability of win-draw-lose in coming games in season 2019-2020 according to the schedule.

To present our prediction result, probability density function (PDF) is used to show the probability .

## ELO Rating System

The Elo rating system is a method for calculating the relative skill levels of players in zero-sum games such as chess. It is named after its creator Arpad Elo, a Hungarian-American physics professor.

The Elo system was originally invented as an improved chess rating system over the previously used Harkness system, but is also used as a rating system for multiplayer competition in a number of video games, association football, American football, basketball, Major League Baseball, table tennis, board games such as Scrabble and Diplomacy, and other games.[6]

## Performance Rating

The phrase "Elo rating" is often used to mean a player's chess rating as calculated by FIDE. However, this usage is confusing and misleading because Elo's general ideas have been adopted by many organizations. Each organization has a unique implementation, and none of them follows Elo's original suggestions precisely. It would be more accurate to refer to all of the above ratings as Elo ratings and none of them as the Elo rating.

Performance Rating is one of the unique implementations, which is learned and improved by us in this project.

Performance rating is a hypothetical rating that would result from the games of a single event only. Some chess organizations use the "algorithm of 400" to calculate performance rating. According to this algorithm, performance rating for an event is calculated in the following way: For each win, add your opponent's rating plus 400; For each loss, add your opponent's rating minus 400. And divide this sum by the number of played games.

This can be expressed by the following formula:

To improve the precise of our Ranking system, ELO Rating Principle and EPL regular rating principle are both used in our algorithm.

## Probability Density Function

In probability theory, a probability density function (PDF), or density of a continuous random variable, is a function whose value at any given sample (or point) in the sample space (the set of possible values taken by the random variable) can be interpreted as providing a relative likelihood that the value of the random variable would equal that sample.[2]

For our project, we will use two distribution to predict the goal difference in 92 games left, one is Skellam Distribution, the other is Poisson Distribution. Especially, the Skellam Distribution can be seen as two Poisson Distribution combined.

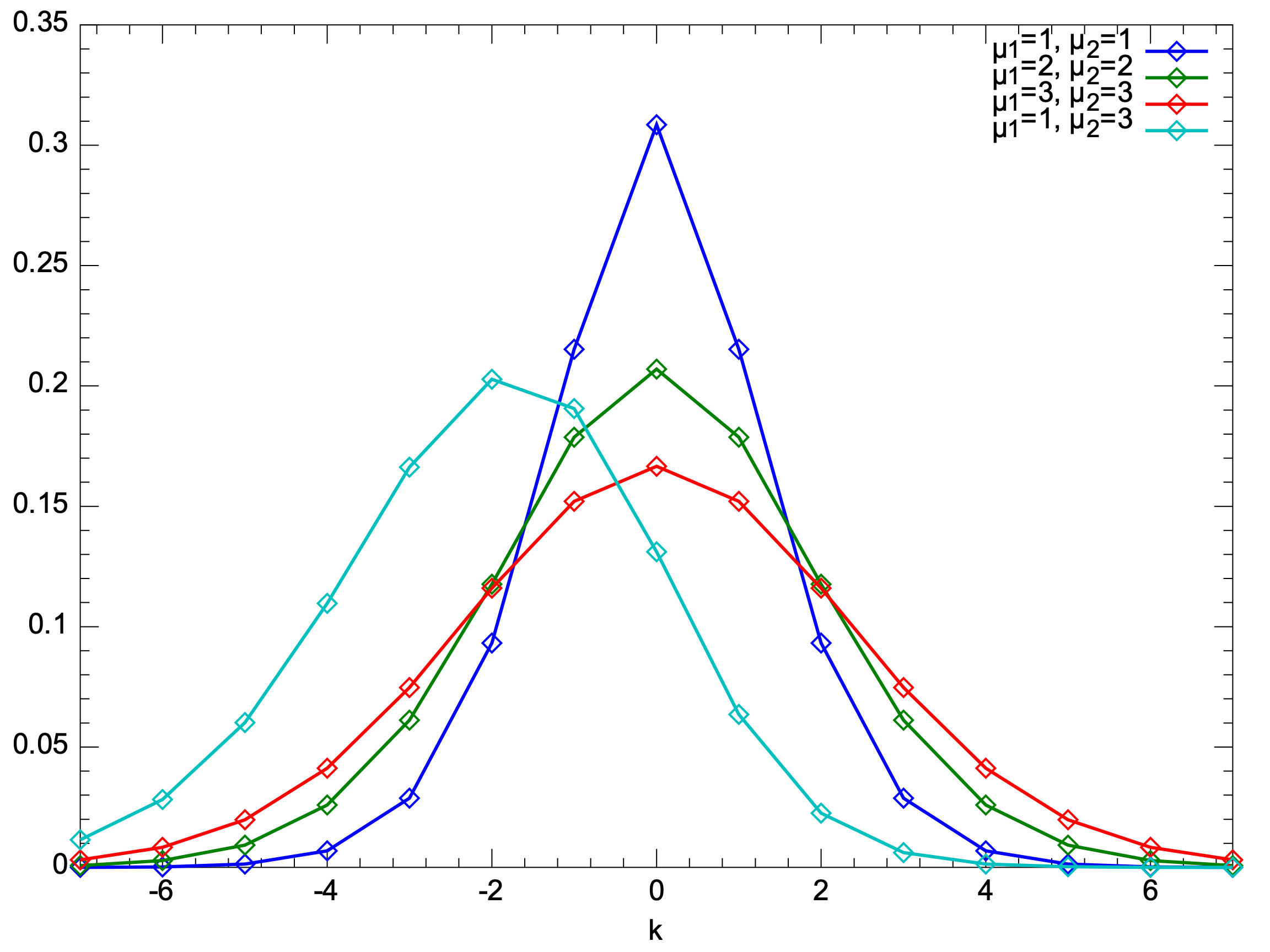


Fig.1. Probability density function of Skellam Distribution.

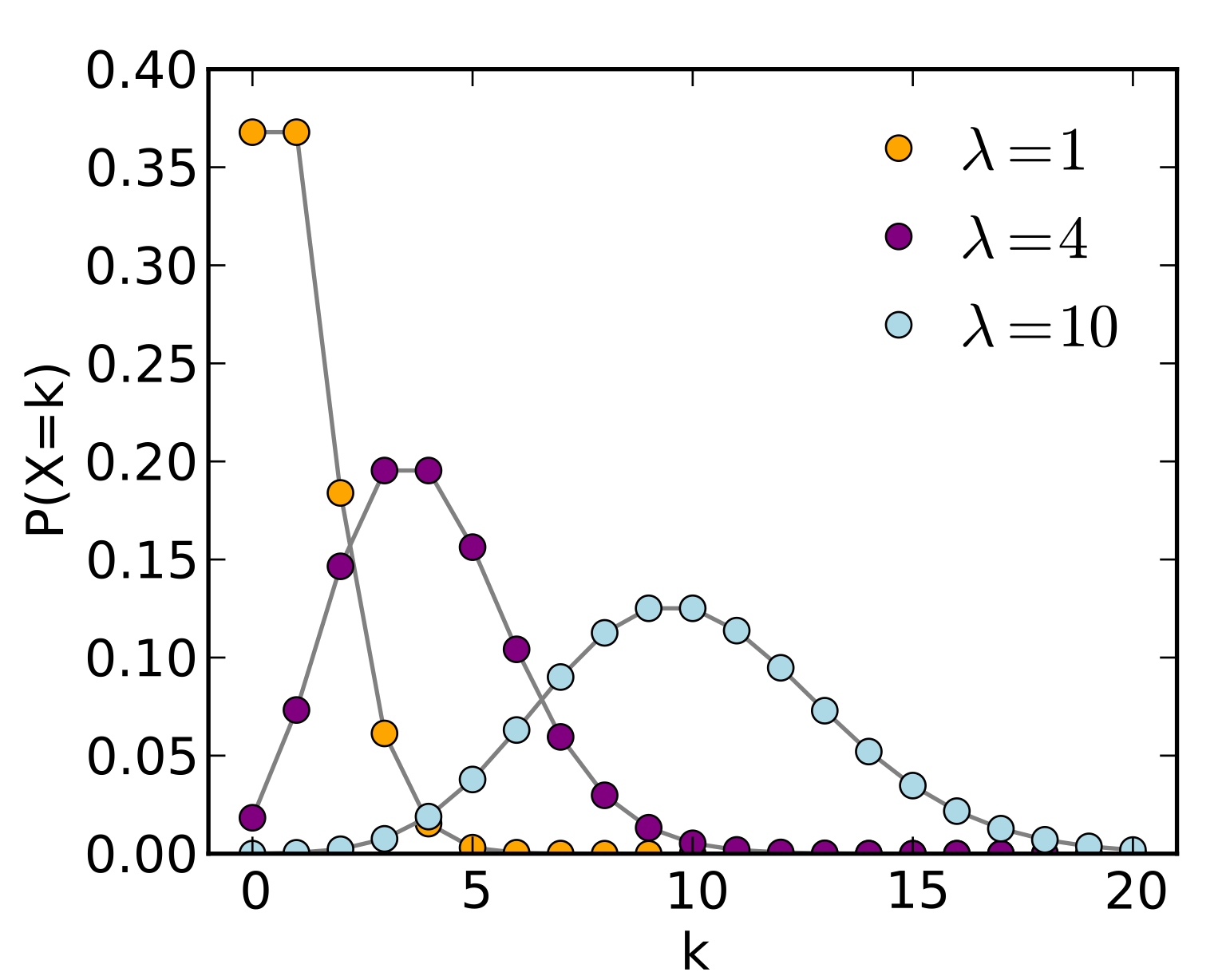


Fig.2. Probability density function of Poisson Distribution.

The probability mass function of Skellam Distribution is:

[3]

And is the modified Bessel function of the first kind.

[4]

The probability mass function of Poisson Distribution is:

[5]

# Aim of the Project

Designed the EPL Ranking System.

* Calculate the win, draw and loss probability that team A will beat team B if they meet with each other in a head to head matchup.
* Draw the probability density function for goal difference in each game.
* Predict those coming games. “Complete the season” and give a final table with 38 matches “played” by each team.
* Predict the final rank of those twenty teams with the prediction of 92 coming games.

# Dataset used in this project

In this project, dataset used by us is only EPL 2019-2020 season result (those games finished). Data is saved in a CSV file.

图片包含 游戏机, 电脑

描述已自动生成

Fig.3. An overview of data in 2019-2020 season

For each match, we will extract the Home Team, Away Team, Home Team Goal, Away Team Goal, Result, Home Fouls Commit, Away Fouls Commit, Home Shot on Target, and Away Shot on Target as parameters to calculate the probability of the result in future.

# Project Description

## Import Dataset

We import dataset of the match history in 2019-2020 season by CVSReader.jar and create nine classes which will be useful for our team to store data or realize the function my need.

## Rank Teams by current matches

By ELO Rating Principle and EPL regular rating principle, we primarily calculate the rank of those twenty teams by analyzing past 288 matches.

## Predict the Future Matches

In light of the whole season each team will play 19 games in their home with others, 92 games need to play in the future. With team directory and match history, we can obtain the future games stored into an Array List.

## Get the Necessary Parameters for Prediction

Significantly, in order to want to predict the goal difference in future games, we need the average goals of each team in the past. However, the mathematical average goal, which is calculated by the total goals divided by the total number of games, could not perfectly reflect the strengthen in future games, because different team has different defensive strengthen. In light of this, we use The Possible Goal Difference of A team in each future game. For here, the Possible Goal Difference of A team equals the average number of shot on target of A team minus the average number of successful defenses of B team. It is:

## Apply Proper Distribution for Prediction

After calculating the possible goal difference of A team in each future game, we find that there are some PGD (Possible Goal Difference) is negative. If we want to use the Skellam distribution, the PGD of two teams in a game must be positive. Meanwhile, if PGD of A is negative, we can definitely assume the possible of A’s goal is 0. Now, we just use the Poisson Distribution to predict how about the B’s goal, where is the PGD of A – PGD of B (PGD of B <0). So, we use those two distributions to predict the goal difference in the future. (All the distribution graphs are stored into main/resources, and their file type is png)

## Get Probability of Result and Final Rank in Future Matches

Finally, we need to predict the probability that home team win, draw or lose in each future game. For those predicted by Skellam Distribution, the probability of home team win is just the sum of the probability of goal difference larger than zero, the probability of draw is the probability of goal difference is zero, and the probability of lose is the probability of goal difference less than zero. For those predicted by Poisson Distribution, the probability of winner is the probability of goal difference larger than zero, the probability of draw is the probability of goal difference is zero, and the probability of lose is zero. After gaining those probability of each future match, we assume the result of future game is determined by the biggest probability of win, draw or lose. Then we rank them again to obtain the final rank result with all the 380 matches in this season by those two principles separately.

# Implementation

## RankingSystem

电脑屏幕截图

描述已自动生成

## DataReader

手机屏幕截图

描述已自动生成

## DrawMath

手机截图图社交软件的信息

描述已自动生成

## FutureMatchDictory

手机屏幕截图

描述已自动生成

## Match

电脑屏幕的截图

描述已自动生成

## MatchDirectory

手机屏幕截图

描述已自动生成

## sortHelper

手机屏幕截图

描述已自动生成

## Team

电脑屏幕的截图

描述已自动生成

## TeamDirectory

手机截图图社交软件的信息

描述已自动生成

# Output

## Current Rank

图片包含 游戏机

描述已自动生成Fig.4 Current rank of teams by EPL point.

图片包含 游戏机

描述已自动生成

Fig.5 Current rank of teams by ELO rating points.

Those rank tables are stored in main/resources/ Rank result of EPL Principle and main/resources/ Rank result of ELO Rating Principle, Separately.

## Win/Draw/Loss Probability

图片包含 游戏机

描述已自动生成

Fig.6 A screenshot of output in command line

All the results are output in command line.

## Probability Density Function Graph

地图的截图

描述已自动生成

图片包含 游戏机

描述已自动生成

地图的截图

描述已自动生成

图片包含 游戏机

描述已自动生成

Fig.7 Four examples of product density function graph.

All distributions are stored in main/resources/Distribution Graphs

## Final Table图片包含 游戏机, 电脑 描述已自动生成

Fig.8 The result of all games in 2019-2020 season.

## 图片包含 游戏机 描述已自动生成Final Rank

Fig.9 Final rank of teams by EPL point.

图片包含 游戏机

描述已自动生成

Fig.10 Final rank of teams by ELO rating point.

These final rank tables are stored in “main/resources/Rank result of EPL Principle” and “main/resources/ Rank result of ELO Rating Principle”, Separately.

# Conclusion and Future Implementations

1. According to both rating principle, the champion will be absolutely Liverpool. And Bournemouth, Norwich and Watford will be relegated. In light of the similarity of two rating principle, we have confidence to say our prediction is approximately rational and logical.
2. In the future, we would analysis the effect of home team advantage. And quantify it into a parameter in our prediction, which make our prediction and ranking system more precise. And we could also make a GUI for our ranking system and do some visualization work, which can improve our project into an interactive system.

# References

[1] <http://www.merriam-webster.com/dictionary/ranking>

[2] Grinstead, Charles M.; Snell, J. Laurie (2009). "Conditional Probability - Discrete Conditional" (PDF). Grinstead & Snell's Introduction to Probability. Orange Grove Texts. ISBN 161610046X. Retrieved 2019-07-25.

[3] <https://en.wikipedia.org/wiki/Skellam_distribution>

[4] <https://en.wikipedia.org/wiki/Bessel_function#Modified_Bessel_functions_:_I.CE.B1.2C_K.CE.B1>

[5] <https://en.wikipedia.org/wiki/Poisson_distribution>

[6] [https://en.wikipedia.org/wiki/Elo\_rating\_system](https://en.wikipedia.org/wiki/Elo_rating_system#cite_note-2)