**Data Analysis and Machine Learning on League of Legends**

Yiyi Zhou, Chang Liu

zhou.Yiy@husky.neu.edu

# liu.chang12@husky.neu.edu

INFO 7390 Advances Data Sci/Architecture, Fall 2017 Northeastern University

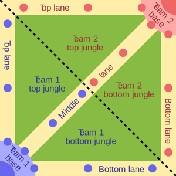
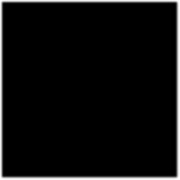
# Abstract

Based on selected world tournament match records dataset, this project aims to make analysis about region gaming styles, strategies, and the win rates behind it in the game of League of Legends. This project made a logic regression to analyze the effect of different parameter on the match result and used Log5 formula to find the win rate between each two champions. Furthermore, from these match records, with various champions selected by different players to content, a model was trained to predict match result based on champion selection. With the analysis in this project we answered how to pick champions according to the champion picked on the other side.

# Introduction

League of Legends (LOL) has been a typical and popular multiplayer online battle arena (MOBA) video game developed and published by Riot Games. Released in 2009, it has been a great popular MOBA game around the world.

During a classic match, players will be involved in a 5v5 team game to work with teammates and compete with the enemy team. The ultimate goal is to destroy the main base of the enemy’s team. For each match, players can select a different champion as the character to operate, and each champion has unique abilities and game roles. As shown in the figure, the entire map is divided into blue side and red side with two teams spawned in two corners. Right in the left bottom and right top are the blue side and



*Fig 1. A classic MOBA game map*

the red side main base respectively, and there are three lanes connecting two bases in total, called top lane, middle lane, and bottom lane.

The other diagonal vertical to the middle lane is a path called the river, and these two diagonals divide the rest part of the map into 4 parts, and they are all called jungles. Generally, a 5 players’ team would assign one to the top, one to the middle, one to the jungle, and two to the bottom, their position is called top, middle, jungle, bottom and support.

Usually the bottom is an archer (a champion’s role, usually called AD carry/ADC) who needs a long time of farming, so a support is assigned for him/her to protect this process, and they are assigned to the bottom lane to protect the dragon to be stolen from the enemy, which is an important strategic resource in the early game.

# Dataset

The datasets used in this project are from kaggle.com and other game statistic website.

**Kaggle.com:**  https://www.kaggle.com/chuckephron/ leagueoflegends (Professional Tournament) https://www.kaggle.com/lanls1/matchidv1

(Rank Matches) https://www.kaggle.com/xenogearcap/ league2016 (Rank Matches)

**Official Match History site:** http://matchhistory.na.leagueoflegends.com/ en/#page/landing-page

**Official stats API:** <https://developer.riotgames.com>

# Data Analysis

# The effect of different coefficients for the match result and logic regression.

# The match results of LOL depend on many different coefficients, like the gold gain of the team, the dragons it killed, the barons it killed etc. We made a logical regression to see the effect of different coefficients on our match result. We chose the value of game length, the dragon killed, the gold in 10 minutes, the gold in 20 minutes, the herald number, the baron value, the inhibits value, the tower knocked down and the kill times in a match as the independent variable.

The following is the regression result we got.

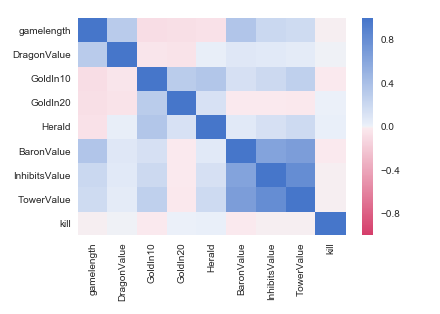
# 

*Fig 2. Logical regression*

To win a game, gold in 10, gold in 20 are important features, and so are the Baron value, the inhibits value, and the tower value. For every unit change in Baron, the log odds of being a winner will decrease by 0.2153 For every unit change in Tower I knocked down, the log odds of being a winner will decrease by 1.8170.

# 

We used the model we trained to predict the result of our test data. And we got the accuracy of 0.96. That’s very high.

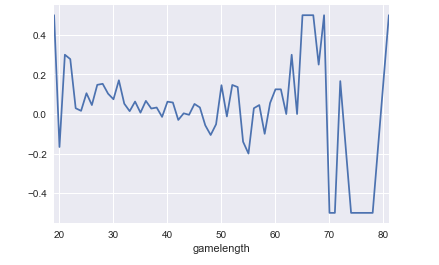
To see if there are some relationship between the independent values, we calculated for the correlation between different independent variables.

*Fig 3. Heat map for different coefficients in this game*

Here we can see, the baron value, the inhibits value and the tower value have some effect on each other but not large. There is no noticeable relationship between the other independent values.

**2.The blue team’s advantage according to time**

A popular belief is that, the blue team have some advantages over the red team at the beginning of the match, so they are more likely to win at the beginning. We calculated the win rate of the blue team in different matches of different game length and compared it with 0.5. The result is shown as below.



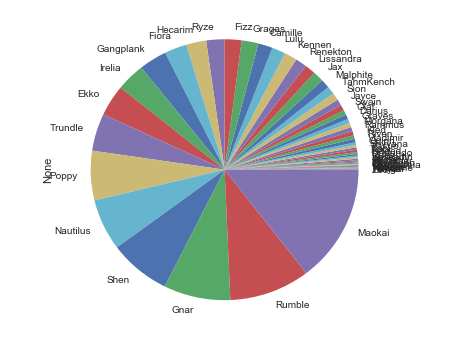
*Fig 4. Blue win rate according to time*

As we can see, during the beginning 50 minutes, the winning rate of blue team is slightly higher than 0.5, so it does have an advantage over the red team at the beginning. As time increases, this advantage disappears.

# 3.Pick Champions

We first made the statistic of the most popular champions in different location of the game(*Top,Jungle,Middle,ADC,Support*).

**Top:**



*Fig 5. Pie chart of popular champions in Top*

As we can see *Maokai, Rumble, Gnar, Shen, Nautilus* are the top five popular champions for Top.

We did the following steps to find the champion that has some advantage over the five champions above.

We first selected all the matches that *Maokai* takes part and grouped them by the Top champion name of the other team. And then we selected all the matches that *Maokai* took part when the team lose the match and again grouped them by the champion name of the other team. Later we divided them to calculate for the win rate of these champions over *Maokai.*

Considering the rigor of statistic, we filtered the champions that fight with *Maokai* for less than five times.

The result we got is shown below.



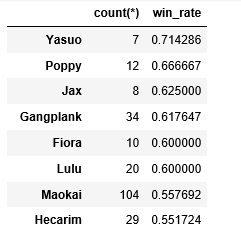
*Fig 6. Champions win rate over Maokai in Top*

According to the result, *Malphite* and *Sion* are powerful rivals of *Maokai*. In the step of picking champions, if one side choose to use *Maokai*, the other side will have an advantage if them pick *Malphite* or *Sion*.

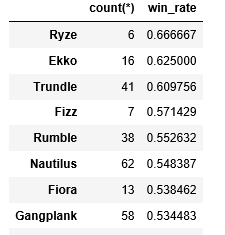
We did the similar work for all of the top five champions in Top. And the results are shown below.



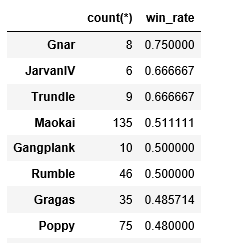
*Fig 7. Champions win rate over Rumble in Top*



*Fig 8. Champions win rate over Gnar in Top*



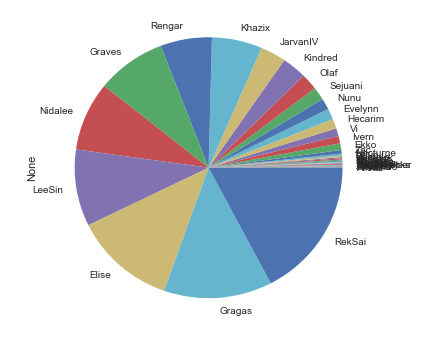
*Fig 9. Champions win rate over Shen in Top*



*Fig 10. Champions win rate over Nautilus in Top*

**Jungle:**

We made similar analysis for Jungle as we did for Top.



*Fig 11. Pie chart of popular champions in Jungle*

*RekSai, Gragas, Elist ,LeeSin* and *Nidalee* are popular champions for jungle.



*Fig 12. Champions win rate over Gragas in Jungle*



*Fig 13. Champions win rate over Elise in Jungle*

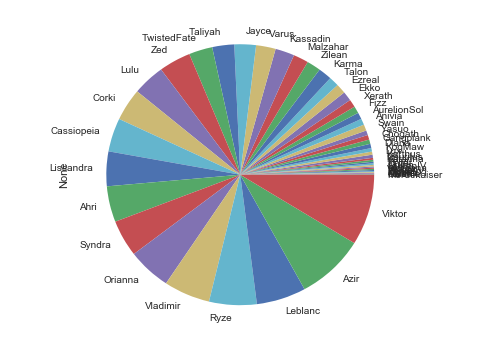


*Fig 14. Champions win rate over LeeSin in Jungle*

**

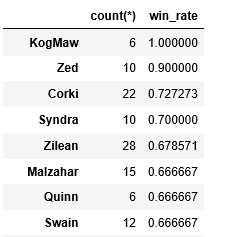
*Fig 15. Champions win rate over Nidalee in Jungle*

**Middle:**

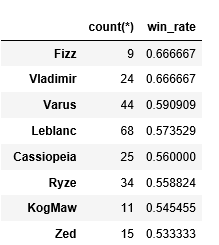


*Fig 16. Pie chart of popular champions in Middle*

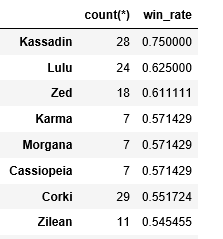
As we can see, *Viktor, Azir, Leblanc, Ryze* and *Vladmir* are the popular champions in this match.



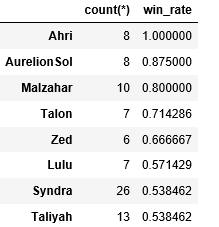
*Fig 17. Champions win rate over Victor in Middle*

**

*Fig 18. Champions win rate over Azir in Middle*

**

*Fig 19. Champions win rate over Leblanc in Middle*

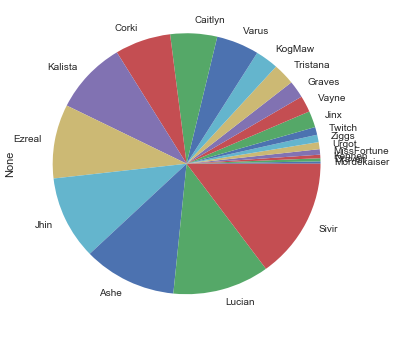


*Fig 20. Champions win rate over Ryze in Middle*



*Fig 21. Champions win rate over Vladimir in Middle*

**ADC:**

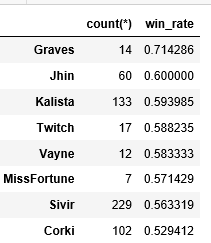


*Fig 22. Pie chart of popular champions in ADC*

As we can see, *Sirvir, Lucian, Ashe, Jhin*, and *Ezreal* are the popular champions at ADC in this match.



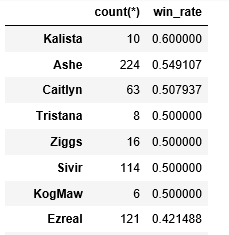
*Fig 23. Champions win rate over Sivir in ADC*

**

*Fig 24. Champions win rate over Lucian in ADC*

**

*Fig 25. Champions win rate over Ashe in ADC*

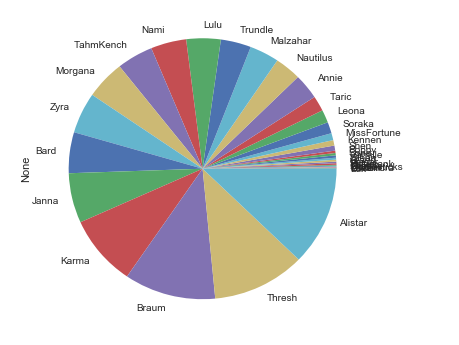
**

*Fig 26. Champions win rate over Jhin in ADC*

**

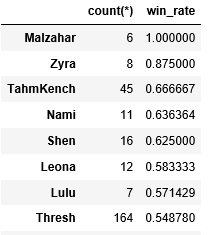
*Fig 27. Champions win rate over Ezreal in ADC*

**Support:**

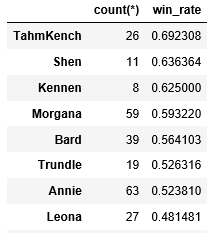


*Fig 28. Pie chart of popular champions in Support*

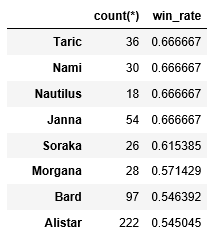
As we can see, *Alistar, Thresh, Brarum, Karma, Janna* are the popular champions in this match.



*Fig 29. Champions win rate over Alistar in Support*

**

*Fig 30. Champions win rate over Thresh in Support*



*Fig 31. Champions win rate over Brarum in Support*

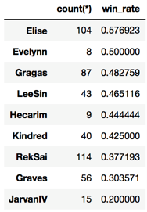


*Fig 32. Champions win rate over Karma in Support*



*Fig 33. Champions win rate over Janna in Support*

The aim of the above analysis is to find a proper choice for players to get a counter pick against the enemy’s choice, as the last enemy’s selected champion can be seen in the ban/pick phase before the game begins. To implement this, here follows the example.



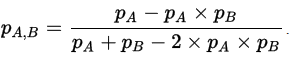
*Fig 34. Champions win rate over Nidalee in Jungle*

During the match in tournaments matches, *Elise* can be a good choice to counter *Nidalee* as a jungle position pick. *Elise* is a great popular Jungle champion with a high rate of selection as well. She is quite strong and aggressive in the early game phase, especially in scenarios like quick-gank and jungle encounter fight. She is able to kill other champions before level of 6 due to her special skill mechanisms of high outbreak of damage, which is obviously easy to take advantage to champions willing to farm and get rid of early encountering such as *Nidalee.*

**4.Using Log5 to find the possibility a character wins another character:**

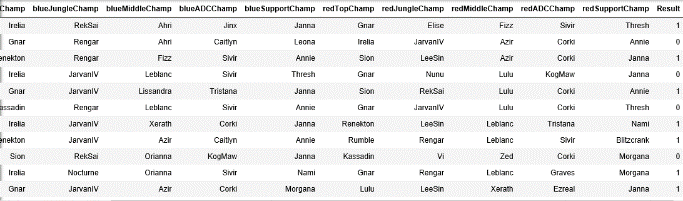
**Log 5** is a formula invented by Bill James to estimate the probability that team A will win a game, based on the true winning percentage of Team A and Team B.

The Log5 estimate for the probability of A defeating B is :



In our analysis we used Log5 to find the wining rate between two champions by first calculating for the winning rate of the specific champion among all ( we defined the champion to win by the team result, like if Rumble showed in two matches in blue side, and in these two matches the blue side wins then the winning rate of Rumble is 100%) and later calculate their win rate between each other.

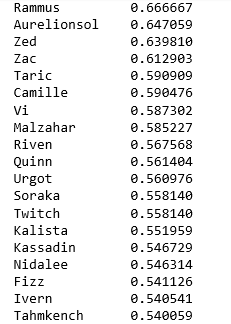
1. We first got the Champion participated in every match from both sides and the result of these matches. The graph is showing as below(Columns contain the champion in each location and the result).



*Fig 35. Match records with champion information*

1. We suppose that champions appear in matches for less than 30 times are champions with low using frequencies, and we remove the champions from our analysis.

Here is part of the win rate of every champion.



*Fig 36. Champions win rate of all the champions*

1. We chose the top five champions with the highest winning rate, and the last five champions with the lowest wining rate and calculated for the winning rate of them between each other using **Log5:**



*Fig 37. Winning rate between the strongest five champions and the weakest five champions.*

A screenshot of a cell phone

Description generated with high confidence

*Fig 38. Heat map of winning rate between each pair of the five strongest champions and the five weakest champions*

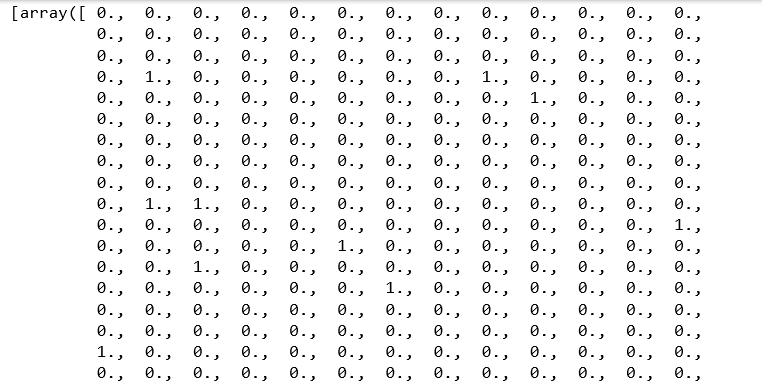
Here we can see that the possibility of *Rammus* to win *Drmundo* is higher than 0.85 which is large. The winning rates of *Renekton, Xerath , Evelynn* and *Lux* between each other are around 0.5 which means that the level difference among these champions are not significant. *Drmundo* is a weak champion, the power difference between *Drmundo* and the other four last powerful champions is not neglectable.

1. **Using K-Nearest Neighbors to predict Match Results**

A popular belief among LOL players is that the champions chosen in a match have effect on the match result. A team with a good composition is more likely to win the game. We used the K-nearest neighbors for classification and regression that predicts the match result. We chose to use the KNN in the consider that champions may have some talents and qualities complementary towards other champions so we need to take them as a group instead of merely calculating about the attendance of each champion.

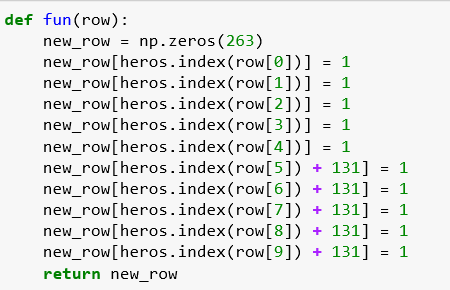
To achieve our goal, we built an array with 262 columns with each column representing a champion in the game as there were 131 champions occurred in our dataset and we used different columns to represent if this champion was in the red side or in the blue side (like if the first character participated in a match in the blue side, then the first column in that row was set to ‘1’, and if it participated in the red side, then the 1+131 = 132 column in this row was set to ‘1’. The array has 3802 rows representing the 3802 matches in our dataset. In total, each row had 10 cells set to 1 with five of them on the left side and five of them on the right side.

How our array looks like:



*Fig 39. Example of the array in KNN*

The following is the code we used to build our array.

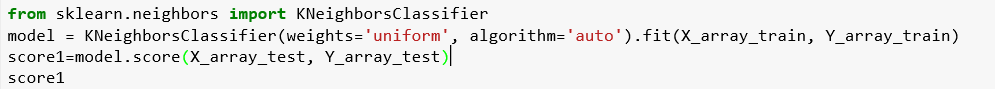


*Fig 40. Function we used in KNN*

We divided our data into training data and testing data.

We trained our model using K-Neighbors Classifier in sklearn.neibbors. We used the uniform weight and auto algorithm to train the model.

The following is the code we used to get our model.

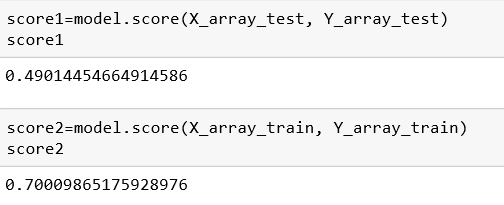


*Fig 41. The model get from KNN*

To test our model, we used our test data.

The following is our accuracy.

The accuracy of our model to predict our test model is about 0.5, and the accuracy to predict our training data itself is 0.7.



*Fig 42. Accuracy of our KNN model*

We came to the conclusion that LOL was a well-designed game. The fighting ability of different composition of champions are at a balanced level. The champions picked don’t have a noticeable effect on the match result.

# Conclusion

As the data analysis part shown, some obvious trend and results of analysis could be addressed by comparing specific attributes of dataset. Such as the gold proportion rate to each position and their win rate. The above analysis could be informative to players focusing on how to win a game more effectively, especially for those professional e-sport teams.

The other part of prediction/pick is to make a counter pick selection guide against a specific champion selected by the enemy. This could be quite practical and useful for players at ban/pick phase, especially those who are not that familiar with all the champions features and abilities.

We tried to train a model to predict the match result according to the composition of both side of the team and thus help our players to make decisions. However, this model turned out to be not very helpful with an accuracy of about 0.5. We think part of the reason is that our data comes from professional games, so the team composition are more likely to be well-organized in both side. Besides, the LOL is a well-designed game so the power levels of the champions are balanced at an acceptable level.

# Acknowledgment

We would like to show our gratitude to professor Nik Bear Brown for guiding us and encouraging us during the course of this project.

# References

[1]. [kaggle.com](http://kaggle.com/) League of Legends: https:// www.kaggle.com/chuckephron/ leagueoflegends (Professional Tournament) https://www.kaggle.com/lanls1/matchidv1 [2]. Riot developer API: [https:// developer.riotgames.com](https://developer.riotgames.com/) [3]. KNN algorithm-wiki: [https:// en.wikipedia.org/wiki/Knearest\_neighbors\_algorithm](https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm) [4]. KNN algorithm specification: [http://](http://blog.csdn.net/u011067360/article/details/23941577)

[blog.csdn.net/u011067360/article/details/ 23941577](http://blog.csdn.net/u011067360/article/details/23941577)

[5]. LOL stats: [http://www.lolking.net](http://www.lolking.net/) [6]. LOL champions ranking: http:// champion.gg/statistics/