

---

# Player Base Analysis on League of Legends

---

**Jason Ma**  
Mathematics of Big Data I  
Harvey Mudd College  
jyma@g.hmc.edu

## Abstract

This paper presents a few approaches to predict the win/loss ratio (WRR) for players in League of Legends given statistics accumulated over many ranked games. We then try to classify players based on their statistics to glean information about the different play styles present in the data. The methods used include Linear Regression, recursive feature elimination, and K Means clustering.

## 1 Introduction

League of Legends (LoL), created by Riot Games Inc., is the most popular Multiplayer Online Battle Arena (MOBA) video game in the world. It boasts over 100 million monthly players that compared to the 13 million of DOTA 2 - which is considered to be LoL's main competitor - show that LoL has excelled at player acquisition[1]. Since LoL is a free-to-play game, it is imperative for Riot Games to ensure that LoL is an enjoyable experience for all players to maintain their player base. To this end, it is crucial they have a firm grasp of the tendencies players have, and which types of playstyles their players most. Thus, the goal of this paper is to apply statistical data analysis to uncover the trends that will be most useful for Riot Games to balance their game around, and discover what classes of players exist.

It is Riot Games' policy to not provide a public dataset of player statistics, but they do allow individuals to use their API[2] to pull data for personal use. To focus the data analysis on the more dedicated players and reduce the number of API calls required, we chose to solely use statistics from LoL's ranked game mode, which is a more competitive environment. After 2 weeks of data gathering, about 1.5 million players were selected for analysis, of which only 110,000 had ranked statistics.

## 2 Data Processing

The data gathered from the API was first cut down into a csv file with each line representing a player and including the id of the champion - character in the game - they play, followed by 37 features which were aggregated statistics from the players' games on the champion. For analysis, this was further cut down into aggregating the player's statistics per game across all champions played, which left 10 features for every player. These features included win/loss ratio (WRR), kills per game, deaths per game, assists per game, gold earned per game, minions slain per game, and various types of damage dealt per game. In this paper, playstyle will refer to trends in the player features.

For the purpose of this paper, we will focus mostly on WRR. Since people are more likely to desire winning when playing a competitive game mode, we believe that we can approximate all players in ranked to be playing their best playstyle for winning. As such, the analysis of this data will most likely Additionally, this allows for the possibility of using these results to determine the more optimal playstyle in LoL.

### 3 Algorithms Applied

First to analyze the statistics that correlated most with a good WRR, we applied recursive feature elimination. The algorithm works by assigning features to weights - in this case with logistic regression - and then removing the features with the the smallest absolute magnitude. Then it recurses with the smaller feature set until the number of features is reduced to the desired amount. Initially, we chose to use LASSO regression for the feature selection, but we were unable to produce meaningful results.

Next, we applied linear regression to try and fit player features to their WRR. Since winning is especially important for players in ranked games, analyzing which features correlate with a higher WRR. Understanding which playstyle is most likely to win is crucial for Riot Games, so that they can balance LoL such that they can reward fun playstyles with higher chances of winning.

Finally, we applied K Means clustering in an attempt to classify players based on their playstyle. For this algorithm specifically, we chose to normalize the statistics across all players by their means, so that the resulting centroids will easier to analyze.

The algorithm for k-means is as follows:

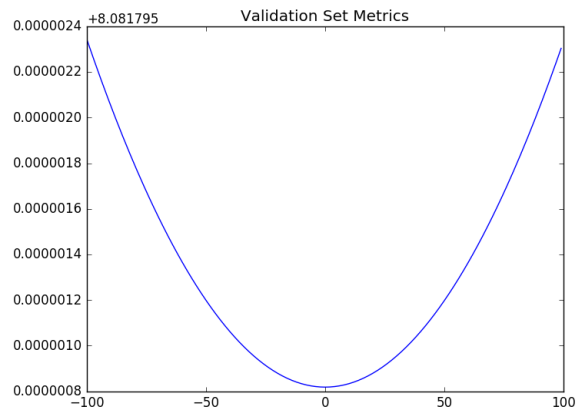
**Data:** Let  $lst$  be an array of vectors of  $N$  dimenions,  $Cent$  be randomly defined vectors of  $N$  dimensions, and  $C$  be the number of desired centroids  
**Result:** A list of centroids  
**while**  $Cent \neq oldCent$  **do**  
     $oldCent = cent$   
     $labels = getClosestCentroids(lst, Cent)$   $\triangleright$  Clusters data points according to the closest centroids  
     $centroids = getNewCentroids(lst, labels, C)$   $\triangleright$  Uses clusters to generate new centroids at their center  
**end**  
**return**  $Cent$

**Algorithm 1:** K-Means Clustering

### 4 Results

The Recursive Feature Elimination outputed an array of rankings of each of the features. After running various tests, on average the statistics that most heavily correlated with a higher WRR were assistsPerSession, killsPerSession, goldPerSession, and minionKillsPerSession.

For the linear regression, after testing various values, we found that the optimal  $\lambda^* = 0.0$ ,



Using the given regularization term, the produced RMSE values for the test and validation set were 8.425 and 7.818 respectively. These show that the features we choose to predict WRR are probably not the best features, and that there is probably a hidden combination of features that will perform better. Still, it's not completely far off from the correct answer, and

The results centroids generated from K-Means is given below.

Table 1: K-Means Centroids

WRR	Kills	Assists	Deaths	Gold	Minions	Heals	Damage	Physical Damage	Magic Damage
47.88	5.499	8.964	6	11693	124	2551	161937	44073	63747
45.26	5.042	8.987	5.798	12194	186	1505	340556	59659	72406
40.72	4.287	8.755	6.414	10603	103	2101	162021	30965	54054
44.83	4.812	9.119	6.153	11234	119	2517	451061	39711	57483
27.7	2.694	6.75	6.111	9119	106	823	596479	33247	44953

## 5 Conclusions

From the results, very clear that Kills, Minions Slain, and Gold earned are the strongest features that affect WRR. The recursive feature elimination method most often scored these features to be in the top 3-4 features for predicting WRR. Additionally, the higher WRR in the Table 1 are highly correlated with more Kills, more Gold Earned, and more minions slain.

With these results, we now have a better understanding of what exactly determines how well a player performs. We hope that in the future, we will have the opportunity to build upon these results and build better models to predict even more. In the future, we hope to incorporate a player's ranking in ranked, so that we can also start predicting which features best affect that trait as well. Another possible side project would be to analyze match data, and see which types of advantages best correlate with victory.

## 6 References

- [1] Tassi, Paul. "Riot Games Reveals 'League of Legends' Has 100 Million Monthly Players". In: *Forbes* (2016)
- [2] <https://developer.riotgames.com/general-policies>
- [3] Dipaolo, Conner. Solution Manual: Math 189R Homework 1, Problem 1. (2016)
- [4] Dipaolo, Conner. Solution Manual: Math 189R Homework 2, Problem 4. (2016)
- [5] Murphy, Kevin P. *Machine learning: a probabilistic perspective*. MIT press, 2012

### 6.1 Links

Project Github: <https://github.com/jasonma97/PlayerBaseAnalysis>

Riot API: <https://developer.riotgames.com>