League of Legends Data Analysis: How to Win

M.T.*

Abstract

This project aims to analyze datasets from League of Legends (LoL), a highly competitive multiplayer online battle arena (MOBA) game, to uncover patterns and strategies that correlate with winning. Leveraging extensive match data, the study seeks to identify which combinations of champions yield the highest chances of victory, alongside deeper analyses to reveal nuanced insights into effective gameplay strategies. Through quantitative analysis and data-driven methodologies, this research endeavors to contribute to the understanding of game dynamics at various levels of play, potentially offering actionable guidance for players seeking to enhance their performance.

1 Introduction

League of Legends, developed by Riot Games, has established itself as a cornerstone of the esports industry, captivating millions of players and spectators worldwide. Its complexity, stemming from the vast array of champion combinations and strategic depth, presents a fertile ground for analytical exploration. This project is motivated by the hypothesis that certain champion combinations and strategies significantly influence the likelihood of winning matches. By dissecting match datasets, the study aims to uncover patterns that could inform player decisions, team compositions, and game strategies.

The research begins with a foundational analysis of champion win rates and progresses towards more complex inquiries, such as the impact of specific champion synergies, the role of early-game advantages, and the effectiveness of various strategic approaches across different levels of play. This project is not just an academic exercise but a practical guide aiming to bridge the gap between statistical analysis and actionable gameplay insights.

In doing so, the project will engage with existing literature and studies on esports analytics, game theory, and statistical modeling, situating its contributions within the broader context of gaming research. The proposed work outlines a methodology for data collection, analysis, and interpretation, aiming to provide a comprehensive evaluation of gameplay effectiveness in LoL. Through this research, we seek not only to enrich the academic discourse on esports analytics but also to offer tangible strategies for players and teams looking to improve their competitive edge.

2 Related Work

In the realm of League of Legends (LoL) analytics, numerous studies have sought to dissect and predict game outcomes, champion effectiveness, and the impact of game updates on player strategies. A seminal piece of work[1], delves into the performance of individual champions within professional tournaments, employing KDA (Kills, Deaths, Assists) metrics to gauge champion effectiveness across different positions. This analysis highlights the variability in champion performance and its implications on meta-game strategies, emphasizing the importance of champion selection and positional play in professional play. However, the study acknowledges the limitations of using KDA as a sole performance indicator, given the nuances of player skill and game dynamics.

Building on the quantitative analysis of champion performance, the same study explores the creation of a classification model to predict match outcomes in higher-tier ranked games. By leveraging a Decision Tree Model, the research demonstrates a high predictive accuracy, spotlighting the critical role of tower kills in determining match victors. This model's refinement further underscores the strategic importance of objective control, particularly tower and inhibitor destruction, in securing game victories.

Contrastingly, another study[2], focuses on the dynamic aspect of LoL, particularly the influence of balance updates issued by Riot Games. This research employs a classification system to analyze the relationship between balance updates and their subsequent impact on game statistics and champion performance. It highlights the complexities of game balancing and its ramifications on player behavior and champion viability, employing a compre-

*University of Colorado, Boulder

hensive methodology that spans data parsing, automated categorization, and advanced statistical analysis. Despite the intricate nature of balance changes, the study offers novel insights into optimizing game strategies in response to evolving game mechanics.

Lastly, a preparatory study[3], explores the potential of Artificial Intelligence (AI) techniques to predict and enhance game strategies. This research underscores the preliminary nature of applying AI in LoL analytics, emphasizing the necessity of exploratory data analysis to understand the critical variables that influence game outcomes. The future directions proposed include leveraging AI for decision support, champion optimization, and strategic planning against specific opponents. This approach signifies a burgeoning field of study that aims to meld advanced AI techniques with the rich dataset provided by LoL gameplay to unlock new dimensions of game strategy and player performance optimization.

Collectively, these studies underscore the multifaceted approach to analyzing LoL, from individual champion performance and predictive modeling to the broader implications of game updates and the prospective application of AI. Each work contributes to a deeper understanding of the game's dynamics, offering valuable insights for players, coaches, and game analysts alike.

3 Proposed Work

Our project aims to extend the current understanding of strategic dynamics in League of Legends (LoL) by employing a multifaceted approach that integrates data analytics, machine learning (ML), and artificial intelligence (AI) techniques. Given the evolving nature of LoL's meta-game and the continuous introduction of balance changes, our work seeks to not only analyze historical data but also to predict future game outcomes and strategic shifts. The proposed work is structured around several key objectives(which are tentative to our project timeline depending on how things unfold):

Objective 1: Comprehensive Data Analysis Building on exploratory data analysis, we intend to conduct a comprehensive review of available LoL datasets, focusing on high-tier ranked matches and professional tournament play. This analysis will include advanced statistical methods to identify patterns and trends in champion selection, win rates, and game strategy evolution over time. By comparing these patterns across different patches and seasons, we aim to understand the impact of game changes on player strategies.

Objective 2: Predictive Modeling Leveraging the findings from our data analysis, we plan to develop and refine predictive models that can forecast match outcomes based on a variety of factors, including champion picks, team compositions, and in-game statistics such as tower kills and objective control. Our approach will explore various ML algorithms, from decision trees and logistic regression to more complex models like neural networks, to determine the most effective predictive framework.

Objective 3: Strategy Optimization Utilizing insights from our predictive models, our project will delve into strategy optimization for players and teams. This includes identifying key factors that contribute to winning matches and developing guidelines for champion selection, role assignment, and in-game decision-making. We will also explore the potential for real-time strategy adaptation based on early game performance indicators.

Objective 4: AI-Driven Insights As a forward-looking goal, we aim to incorporate AI techniques to not only predict outcomes but also to uncover underlying strategies that might not be immediately apparent through traditional analysis. This could involve the use of unsupervised learning to cluster similar play styles or reinforcement learning to simulate matches and derive optimal strategies under varying conditions.

The proposed work aims to bridge the gap between historical data analysis and the predictive modeling of game outcomes in League of Legends. By embracing a holistic approach that combines data analytics with advanced ML and AI techniques, we hope to offer actionable insights for players, teams, and the broader LoL community, ultimately contributing to a deeper understanding of the game's strategic landscape.

3.1 Data

For starting point, we decided to work with an already existing dataset of competitive League of Legends matches that happened from 2014 to 2018.[4] Later, if we feel a need to come back and gather more data, we will iterate back through our data acquisition and modeling. In the final report, we will also talk a bit about Riot Games API which is the primary official API for developers who wish to gather data directly through their API. It's also important to note that most of these public datasets are the result of different webpages scraping and crawling thus the veracity of the data should also be kept on mind if we note some nonsensical immediate results.

3.2 Explanatory Data Analysis

This dataset contains records of League of Legends competitive matches between 2015-2017. The matches include the NALCS, EULCS, LCK, LMS, and CBLoL leagues as well as the World Championship and Mid-Season Invitational tournaments.[4]

This dataset contains 57 columns regarding info such as the Year, the season, the League that the games took place alongside other info such as each team champions bans and picks, the names of the teams, the result and other in-game objective attributes such as teams golds, dragons, towers, inhibitors and etc.

We have a variety of data types in our dataset. There are some columns like team golds that are lists of integers, team towers that are lists of floats and strings and some other categorical variables such as names of champions which are strings.

We can start off by checking whether we have any duplicates ore missing values. There seems to be no duplicates and we have at max 38 records missing some columns. Since we have around 7k records, we can drop those records for now and proceed.

WE noticed that game length has a mean of 37 minutes with a std close to 8. Also we noticed that there are games as short as 17 minutes and as long 95 minutes. Those games can be later studied individually to gain further insights as the very short games in a competitive setting may be due to the dramatic skill gap between the teams or costly mistakes at the beginning of the match which deserve further investigation.

Additionally, we noted that Blue team wins mean is around 0.54 and Red team wins mean is 0.46. They are close enough to not raise immediate questions and may be due to some insignificant natural advantage due to the map structure. We may need a larger dataset to determine this hypothesis validity without any in-game knowledge.

If we try to get a sense of team win rates and their respective number of games, we'll see that there are teams with a high win rate such as 100% but a few games played(3) and also teams with low win rate such as 25% and still a few number of games(4). This may need to be investigated further to get a sense of elite teams among the competitors. One interesting finding based on that is we found that the average number of games played by different blue teams are about 31 games. We also noticed that the average win rate and standard deviation for blue teams that played more than 30 games(roughly the average) is around 0.55 and 0.12. We can perform another query based on the general 2-std rule of thumb to allocate some outliers(elite teams). We set the threshold for win rate to 0.79(0.55 + 2 * 0.12) and we get only 3 teams. "DFM", "LYN" and "ahq" with respective (39, 0.82), (39,0.92) and (96,0.79) which the first number in the tuple is the number of games and the second one is the win rate.

3.3 Model Training

Since we just started our EDA, we are not yet sure which features are the most important in deciding a victory. For a starter approach, we decided to go with some basic features such as all the champions picks (10 of them for both teams), team names and year, league and season the matches took place in and we chose the blue teams result as target variable. (Note that we have both red team and blue team result in every record which are the negation of each other) Later we may decide to drop some of these features, add some other columns from the original dataset or even engineer some features ourselves.

For starters, we create a random forest with these features as input and the result of the match(for blue team) as the target variable. We have used the sklearn One Hot Encoder for transforming our categorical features(all features in this case) After splitting dataset to train and test using sklearns train_test_split function (20% test size), we use the sklearns Pipline fit method to train our random forest model with 10000 trees as estimators. Finally, we use the Piplines predict method on our test size and we get the evaluation metrics which will be discussed in the next section.

4 Evaluation

The success of our project hinges on the rigorous evaluation of the data analysis techniques, predictive models, and strategic recommendations we propose. To ensure our findings and models are both accurate and applicable to real-world gameplay, we will employ a multifaceted evaluation approach. The list is not exhaustive but encompasses statistical validation, model performance metrics, and qualitative feedback from the LoL community which are tentaive at the moment depending on how much we will get to employ given our project timeline. We will not necessarily do the following but one may wish to consider different aspects of project evaluation:

- **Data Analysis Validation:**
  - **Statistical Significance** Makes assessments for statistical significance of findings

from exploratory data analysis (EDA) to ensure that observed patterns and trends are not due to random chance. Techniques such as t-tests, ANOVA, or chi-squared tests will be used where applicable.

  - **Comparative Analysis** By comparing our data-driven insights with established meta-game strategies and professional gameplay trends, we can validate the relevance and accuracy of our analysis.

- **Predictive Model Assessment:**
  - **Accuracy Metrics** For each predictive model developed (e.g., Decision Tree, Neural Networks), we will evaluate performance using appropriate metrics such as accuracy, precision, recall, F1 score, and AUC-ROC curve. This will help us identify the most effective model for predicting match outcomes and champion viability.
  - **Crossvalidation** To ensure one's model generalizability, they can employ k-fold cross-validation, allowing them to test the model's performance on different subsets of the data.
  - **Feature Importance** Analyzing the importance of different features in predictive models will provide insight into game dynamics and help refine strategy recommendations.

- **Strategy Effectiveness Evaluation:**
  - **Simulation and Back_testing** Where possible, one can simulate matches using their strategic recommendations to test their effectiveness in controlled environments. Additionally, backtesting their strategies against historical match data will provide an empirical basis for their validity.
  - **Community Engagement** Soliciting feedback from the LoL community, including players, coaches, and analysts, will be crucial. One can plan to share their strategic recommendations through forums and social media platforms to gather qualitative feedback on their practicality, usability, and success rate in actual gameplay.

- **Implementation and Iterative Improvement:**
  - **Live Testing** Implementing our strategies in live gameplay situations, either by ourselves or by collaborating with players of various skill levels, will offer direct insights into their effectiveness in the current game meta.
  - **Iterative Refinement** Based on the outcomes of live testing and community feedback, one can iteratively refine their strategies and models, continuously improving their accuracy and applicability.

4.1 Model Evaluation

As mentioned in the last section, we used the classification_report from sklearns metrics library to calculate our models different metrics which are accuracy, precision,recall, f1-score and support in this case.

Accuracy: 0.6143704680290046

	precision	recall	f1-score	support
0	0.63	0.41	0.50	701
1	0.61	0.79	0.69	816
accuracy			0.61	1517
macro avg	0.62	0.60	0.59	1517
weighted avg	0.62	0.61	0.60	1517

Figure 1: Baseline Model Scores

5 Conclusion

This checkpoint report outlines the preliminary framework for an ambitious project aimed at dissecting the complex dynamics of League of Legends through data-driven analysis. Our objective is to bridge the gap between raw data and actionable insights that can inform strategic decisions in-game. By leveraging historical match data, we anticipate uncovering patterns and strategies that transcend conventional wisdom, offering players and teams a competitive edge.

The exploration of related works has already highlighted the vast potential for analytics within the realm of LoL and esports more broadly. Our project seeks to contribute to this growing body of knowledge by applying a unique combination of exploratory data analysis, predictive modeling, and strategy optimization. While the task ahead is challenging, the potential rewards for the LoL community and the field of esports analytics are significant.

We are still iterating through different phases of our project from data gathering to our model training and deciding which questions can be answered given our dataset and models. In evaluating the performance of our initial random forest classifier designed to predict match outcomes in League of Legends, we observed an overall accuracy of 61.4%. This indicates a promising start, as the model performs significantly better than random guessing. However, the precision, recall, and F1-score metrics reveal areas for improvement, particularly in accurately identifying match losses.

The model demonstrated a higher capability in predicting victories (Class 1) with a recall of 79% and an F1-score of 0.69, compared to its performance in predicting losses (Class 0), where it showed a recall of only 41% and an F1-score of 0.50. This discrepancy suggests that while the model is relatively effective in identifying winning conditions, it struggles to accurately capture the nuances leading to a team's loss.

It can be shown that there are some in_game objectives that are more important than others, especially in different time windows during the match. There can be assymetric differences between the red and blue team in securing some in_game objectives due to the map balances for the two teams.Specifically,Baron and dragon kills are highly impactful objectives to secure before 30 minutes to increase odds of winning the game. These map monsters could be prioritized over towers or champion kills (i.e. thirsting) to more efficiently improve a team's odds of winning.

Additionally, Securing an objective as the red team and denying that objective to the enemy blue team have different impacts on the winning. This could be due to slight differences the position of objectives on the map, relative to the red and blue bases. For example, Baron spawns on the blue side of the river (the map's midline), making it more difficult for the red team to secure. This could explain why red team taking Baron leads to a greater increase in winning odds than blue team taking Baron. This is corroborated by the opposite position of Dragon and opposite relative impact on odds when slaying Dragon.[5]

Given these insights, future work should focus on enhancing the model's sensitivity towards losing conditions, possibly by incorporating more granular ingame data, exploring alternative feature engineering techniques, or experimenting with different model architectures. Enhancing the model's ability to discern the complexities of game outcomes will be crucial for improving its predictive accuracy and reliability.

References

[1] B. Zhao, Z. Jamal, R. Barney, K. Madadi, and Y. Jain, "League of legends: An exploratory data analysis," Medium, 2023, dataRes at UCLA. [Online]. Available: https://ucladatares.medium.com/league-of-legends-an-exploratory-data-analysis-11f6022f18be

[2] A. Kica, A. La Manna, L. O'Donnell, and T. Paolillo, "Analysis of data gathered from league of legends and the riot games api," Computer Science, WPI, Tech. Rep., 2016, advisor: Mark Claypool. Major Qualifying Project MQP-MLC-LL15. Terms ABC 2015-2016.

[3] J. Morales-García, A. Llanes-Castro, M. Curado, and F. Arcas-Túnez, "An exploratory data analysis for league of legends professional match data," in *Workshop Proceedings of the 19th International Conference on Intelligent Environments (IE2023)*, 2023.

[4] C. EPHRON, "League of legends," https://www.kaggle.com/datasets/chuckephron/leagueoflegends/data, 2018.

[5] R. Pili, "Predicting league of legends winner from 30 mins," https://www.kaggle.com/code/ryanpili/predicting-league-of-legends-winner-from-30-minsTrain-Classification-Models, 2022.