# Predicting League of Legends Match Results Based on Machine

**Authors:** Donghua Wang

**Date:** 2024-01-03T00:00:00+00:00

## Abstract

League of Legends (LoL) is a highly popular multiplayer online competitive game, featuring intricate game mechanics and team cooperation, making the prediction of match outcomes a challenging task. This study utilizes a dataset from Kaggle, comprising 9,879 ranked matches ranging from Diamond I to Master tier, to build a machine learning model predicting the ultimate winner, either the blue or red team, based on the features of the first 10 minutes of gameplay. Through steps such as data loading, preprocessing, and feature engineering, we provided effective inputs for the model. For model selection, we opted for the Logistic Regression algorithm, achieving a model accuracy of 0.7277 through data splitting and training. This accuracy robustly supports predictions of the winning side, whether blue or red. However, to further enhance model performance, we recommend exploring additional feature engineering methods, investigating alternative machine learning algorithms, and fine-tuning hyperparameters. The introduction of deep learning models is also a promising avenue to better capture the complex relationships within the game. Through these improvements, we anticipate increasing the model's predictive accuracy for future matches, offering valuable insights for game development and enhancement.

## Full Text

## Preamble

League of Legends (LoL) is a highly popular multiplayer online competitive game featuring intricate mechanics and team-based gameplay, making match outcome prediction a challenging task. This study utilizes a Kaggle dataset comprising 9,879 ranked matches from Diamond I to Master tier to build a machine learning model that predicts the winner (blue or red team) based on features from the first 10 minutes of gameplay. Through data loading, preprocessing, and feature engineering, we prepared effective inputs for the model. For model selection, we employed logistic regression, achieving an accuracy of

0.7277 through data splitting and training. This performance provides robust support for predicting the winning side. However, to further enhance model performance, we recommend exploring additional feature engineering methods, investigating alternative machine learning algorithms, and fine-tuning hyperparameters. The introduction of deep learning models also represents a promising avenue for better capturing the game's complex relationships. Through these improvements, we anticipate increasing predictive accuracy for future matches and offering valuable insights for game development and enhancement.

**Keywords:** League of Legends, Machine Learning, Match Outcome Prediction, Feature Engineering, Logistic Regression, Accuracy

## 1 Introduction

League of Legends (LoL), a globally popular multiplayer online competitive game, has attracted millions of players worldwide. Its unique mechanics, diverse champion selection, and strategic team cooperation make each match an unpredictable and challenging contest. For the esports community, accurate match outcome predictions not only enhance the viewing experience but also provide valuable insights for professional team adjustments and game balance optimization.

This study focuses on the first 10 minutes of LoL matches—a crucial period that often determines the direction of victory. Through detailed analysis of 9,879 high-tier matches (Diamond I to Master), we aim to build a machine learning model capable of predicting the final result based on early-game features. This predictive model holds potential as a beneficial tool in esports, offering players deeper game insights and providing opportunities for professional players and teams to enhance their tactics and strategies.

In our exploration, we conducted data loading and preprocessing, selecting logistic regression as our initial modeling tool. Through validation on training and testing sets, we achieved a model accuracy of 0.7272. However, we acknowledge this is merely a starting point. Future efforts should include more sophisticated feature engineering, in-depth exploration of different algorithms, and comprehensive model performance evaluation.

This study not only offers a new perspective on predicting LoL match results but also opens possibilities for applying machine learning to esports. Through this work, we aspire to make meaningful contributions to the LoL community, esports research, and the broader application of machine learning in the gaming industry. The source code is available in a GitHub repository: https://github.com/wdh7022/Predicting-League-of-Legends/tree/main

## 2.1 Domestic Research Status

In China, the rapid development of data science and machine learning has gradually drawn attention to esports research. League of Legends, as a widely popular

multiplayer online competitive game, has become a focal point for investigation. Researchers have delved into match data analysis, exploring key factors such as player behavior, team strategies, and match outcomes. Additionally, some studies have attempted to use machine learning techniques to predict esports match results. By modeling and analyzing pre-match data, researchers seek to identify crucial features influencing outcomes to provide more accurate predictions. There is also research focused on in-game element balance, offering guidance to developers for improving the gaming experience.

## 2.2 Machine Learning Methods in Esports Research

In the realm of competitive gaming, particularly League of Legends, researchers have extensively explored machine learning applications for forecasting match outcomes. This endeavor involves employing various methods to enhance prediction accuracy and gain deeper insights into underlying match patterns. The following six approaches represent prominent techniques in this domain:

### 2.2.1 Logistic Regression

Logistic regression serves as a valuable linear model for addressing binary classification problems, a common scenario in esports. In this context, scientists harness its power to establish predictive models that correlate match outcomes with numerous features. League of Legends presents researchers with the challenge of predicting results based on various in-game factors, and logistic regression's ability to handle binary outcomes effectively makes it a prominent tool. By leveraging this framework, researchers aim to uncover relationships between diverse features—such as player statistics, team composition, and early-game events—and match outcomes. The logistic regression framework quantifies each feature's impact on the likelihood of victory for either team, providing not only a predictive model but also insights into the relative importance of different factors. Furthermore, logistic regression serves as a foundation for exploring more advanced techniques, as researchers often extend investigations to incorporate additional algorithms, feature engineering methods, and hyperparameter tuning for continuous refinement and improved accuracy.

### 2.2.2 Decision Tree Algorithm

Characterized by a tree-like structure representing decision processes, decision trees prove instrumental for modeling decision-making in multiplayer online games like League of Legends. Their inherent structure offers intuitiveness and interpretability, making them valuable for analyzing and predicting match outcomes. In a decision tree, each node corresponds to an attribute, and each branch signifies a decision based on that attribute. This hierarchical representation allows for a clear breakdown of decision-making processes. In LoL research, decision trees analyze key match features such as player statistics, team composition, and early-game events. Traversing the tree reveals decisions at each

node, providing insights into the factors contributing most significantly to results. The appeal of decision trees lies in their transparent framework, enabling players, analysts, and researchers to easily comprehend decision paths and identify critical features influencing outcomes.

### 2.2.3 Support Vector Machine (SVM)

Support Vector Machines, known for robustness in classification and regression, prove particularly effective for non-linear relationships. In League of Legends research, SVM emerges as a powerful algorithm for modeling match data and providing accurate classification of results. SVM operates by finding the optimal hyperplane that maximally separates data points from different classes. This capability makes it well-suited for capturing complex interactions in multiplayer online games. For match outcome prediction, SVM analyzes features such as player performance metrics, team composition, and early-game events. By mapping these features into high-dimensional space, SVM creates a hyperplane that distinctly classifies matches into different outcomes. Its strength lies in navigating intricate relationships, enabling accurate predictions even with non-linear and complex patterns.

### 2.2.4 Deep Learning and Neural Networks

Deep learning techniques, characterized by deep neural networks, play a pivotal role in comprehending intricate non-linear relationships. In the dynamic context of League of Legends, researchers harness neural networks to predict match outcomes and capture complex game features more effectively. Deep neural networks consist of multiple interconnected layers that process and transform input data, automatically learning intricate patterns and representations. This architecture makes them well-suited for analyzing diverse in-game features such as player statistics, team composition, and early-game events. By training on historical match data, the network learns to recognize complex patterns and dependencies, enhancing its predictive capability. The adaptability of deep learning models allows them to uncover hidden structures, providing nuanced understanding of factors influencing results. As researchers continue exploring deep learning in gaming, these techniques hold potential for unraveling deeper insights into LoL's competitive landscape.

### 2.2.5 Ensemble Learning

Ensemble learning methods, exemplified by random forests and gradient boosting trees, amalgamate multiple models to enhance overall performance. In League of Legends research, these methods integrate various models to adapt to the complexity inherent in match data. Random forests construct multiple decision trees during training, outputting the mode of classes or mean prediction of individual trees. This approach effectively handles diverse features, providing comprehensive analysis of outcomes. Gradient boosting trees build decision trees sequentially, with each tree correcting errors of the previous one, allowing

the model to adapt and improve over time. In LoL, gradient boosting trees capture intricate patterns and refine predictive accuracy. By combining strengths of multiple models, ensemble learning excels where individual models fall short, offering a holistic approach to predicting match outcomes in the rich and multifaceted domain of esports.

#### 2.2.6 Clustering Algorithms

Clustering algorithms, exemplified by K-means, play a crucial role in exploring match data by grouping similar instances and revealing patterns. In League of Legends research, clustering unveils similarities and differences among match types, providing nuanced understanding of dynamics. K-means divides datasets into predetermined clusters representing instances with shared characteristics, such as matches with similar player statistics, team compositions, or early-game events. This reveals distinct patterns and trends not immediately apparent through other methods. Through clustering, researchers gain valuable insights into match data diversity, identifying distinct archetypes based on various features. This understanding benefits strategy tailoring, player behavior analysis, and comprehension of LoL's multifaceted nature. Clustering serves as a valuable tool for organizing match data, facilitating deeper exploration and contributing to comprehensive understanding of the diverse LoL esports landscape.

## 3 Experimental Setup

To construct an effective model for predicting League of Legends match outcomes, we implemented rigorous experimental settings to ensure performance and reliability. Key aspects of our setup included:

**Data Loading and Cleaning:** We obtained a dataset of 9,879 high-tier matches (Diamond I to Master) from Kaggle. During loading, we performed necessary cleaning, including handling missing values, addressing outliers, and standardizing features to ensure data quality and consistency.

**Feature Selection and Engineering:** We carefully selected features relevant to match outcomes, including champion kills, deaths, gold, experience, and level. Through appropriate preprocessing and engineering, our goal was to extract crucial information to help the model effectively capture match dynamics.

**Model Selection:** Initially, we chose logistic regression as our modeling tool, considering its applicability and interpretability for binary classification. Simultaneously, we explored various models including decision trees, support vector machines, and deep learning to investigate performance differences among algorithms.

**Training and Validation:** We split the data into training and testing sets (80% training, 20% validation). To avoid overfitting, we employed cross-validation techniques for iterative model tuning and validation.

**Performance Evaluation:** Accuracy served as our primary metric, supplemented by precision, recall, and other relevant measures. Through comprehensive evaluation, we aimed to assess overall model performance.

Through this experimental setup, our objective was to establish a predictive model with strong generalization capabilities for real match scenarios, providing substantial contributions to the League of Legends community and esports research.

## 4 Experimental Results

In our study, we employed three distinct machine learning algorithms—Logistic Regression, Support Vector Machine (SVM), and Decision Tree—to predict League of Legends match outcomes. All three algorithms exhibited similar overall prediction accuracy, achieving approximately 72%. However, this performance does not necessarily represent exceptional predictive capability, as the game's complexity introduces numerous influencing factors including individual skill levels, team cooperation, and strategic decisions.

While the current predictive accuracy may not be ideal, we view this as an initial exploration. Future research directions should involve more sophisticated feature engineering, experimentation with alternative machine learning algorithms, and potentially the introduction of deep learning methods. Through continuous refinement of models and algorithms, we aim to enhance prediction accuracy for League of Legends match outcomes, providing deeper insights for esports research and practical applications.

## 5 Conclusion

In conclusion, our study employed machine learning models to predict League of Legends match outcomes based on the first 10 minutes of gameplay data. Using three algorithms—Logistic Regression, Support Vector Machine (SVM), and Decision Tree—we achieved similar overall prediction accuracy of approximately 72%. Our research provides initial insights into machine learning applications in esports while underscoring the need for further model refinement. Future efforts should concentrate on optimizing feature engineering, exploring advanced algorithms, and incorporating more complex techniques like deep learning. Through these endeavors, we aim to enhance the model's predictive accuracy for League of Legends matches, offering deeper insights for esports research and practical applications. As the esports industry continues to evolve, machine learning technologies hold promise as potent tools for understanding and optimizing match results.

## 6 Discussion

The final prediction accuracy of only 72% highlights the challenges machine learning models face when forecasting League of Legends match outcomes. This

limitation primarily stems from the complexity of human behavior and battlefield variability, making it difficult to fully capture and predict the gaming environment. Machine learning algorithms often encounter constraints when dealing with nonlinear, dynamic, and highly interactive scenarios, constrained by model simplifications and specific feature limitations.

In esports, factors such as individual player skills, team collaboration, and tactical decisions form a vast and intricate information network. The dynamic and highly subjective nature of these factors makes accurate capture and quantification challenging, thereby reducing prediction reliability. Additionally, the inherent randomness and uncertainty in games pose further challenges, causing models to struggle when responding to unforeseen circumstances and unexpected changes.

We acknowledge that machine learning models still have limitations in predicting human behavior, especially in highly dynamic and interactive gaming environments. While algorithms demonstrate excellence in handling large-scale data, they require more sophisticated modeling for complex human decision-making processes and tactical variations. Future research should explore more intricate model structures, comprehensive feature engineering, and deeper understanding of in-game mechanisms to enhance model adaptability and accuracy. Throughout this process, it is crucial to maintain humility regarding the diversity of human behavior and gaming dynamics, recognizing that achieving complete prediction of game results may always remain a challenge.

*Note: Figure translations are in progress. See original paper for figures.*