League of Legends Artificial Intelligence Coaching

Project Proposal

DS450-01

Data Science Senior Capstone

Kadin McWilliams

kadinmcwilliams@gmail.com

1/12/2025

Executive Summary

This project aims to identify the most critical gameplay variables in League of Legends through quantitative analysis. Using data from 500 personal games and Riot Games' API, the project will explore over 100 gameplay variables, including objectives, damage, vision, gold income, and player behaviors. By analyzing this data set and applying machine learning models, the project seeks to predict game outcomes, validate predictions using random accounts and future games, and highlight key gameplay factors that influence success.

The primary model for this analysis will be an artificial neural network, chosen for its ability to handle complex, non-linear data and its scalability for continuous updates. Additional models include gradient boosting machines for high prediction accuracy and random forest classifiers for robustness with large datasets. These models will help identify strengths and weaknesses in individual player performances, offering tailored coaching recommendations to increase the chances of winning.

Python will be used for all coding involved in this project. The main libraries that will be used are pandas, numpy, and sklearn for data manipulation, data analysis and machine learning. Riot Games' developer tools will be used for the data collection. The project's ultimate goal is to provide players with a clear understanding of which gameplay variables matter most and provide personal recommendations so they can improve their competitive rank in League of Legends.

Project Idea

This project aims to address what the most important variables of gameplay in league of legends are. There are many theories and ideas on what the most important gameplay mechanics to focus on are, but limited quantitative evidence to back up those theories. To accomplish this goal, I will take game data from 500 games that I have played and perform an exploratory data analysis as well as train multiple machine learning models. After completing training I will use the model to predict the outcome of future games that I play as well as predicting random accounts game outcomes to validate the model. Assuming the model is valid, I will use the important values from the machine learning models to explain what the most important gameplay variables are. This model can then be used to show what gameplay variables need to be improved for any individual player, offering individual coaching by highlighting specific player weaknesses.

Background

League of Legends is one of the most popular games in the world, currently being the fifth most played game internationally (Demirkol, 2024). One main draw for many players is the competitive ranked mode in which players can rank up by winning games. However, League of Legends is a complex game and how to win games can be difficult to understand. Seeing which gameplay variables to focus on to win games are not always clear to even the best players.

Currently there are a few solutions to helping players identify the mistakes they are making to help improve their gameplay. The first option that players have is they can review their own games using a replay tool provided by the game. The replay tool is free to use but due to the complexity of the tool, many players who may not be as knowledgeable will have trouble and not be able to utilize it properly. These specific pros and cons means that a player can get a more tailored and specific information about performance, but only if they have the time and knowledge necessary. The second and more common way to identify gameplay performance is to use third party websites like op.gg, u.gg and league of graphs. These websites provide individual statistics such as game win rates, champion performance statistics, individual trends and a game score metric. Not all websites have all of the previously mentioned services and the analysis is often surface level. The lack of consistency between websites means that players would have to look at multiple different websites to get all of the information they need. The overall amount of effort and time to gather these statistics is inefficient. The final approach that people use to improve is hiring coaches to help them improve their gameplay. This method is one of the most effective methods of improvement but does cost money to use this method.

The data set is obtained from the Riot Games' API, this API is available to anyone who wants to access the data. The individual matches contain a large quantity of variables reaching over 100, most of which are integers. There are expected variables such as the number of main objectives taken, with who and how many of those objectives each player took. The APItracks the number and type of pings that each player makes. There are also specific variables under these main categories: damage, vision, gold income,experience, teammate behaviors, items bought, map information, wins, team bans and champions picked. The final main category of variables is under a different APIthat has the champion stats, participation information, damage and position on the map for every 100th of a millisecond in the game. There is a large number of variables that are all seemingly important, the majority of the variables are integers with a few being floating point and boolean variables.

Modeling

The main model that I would like to use for completing this project would be using an artificial neural network. I believe that this will be the best model due to the complex and non linear nature of the data set that is being presented. I also think that the robustness of the model will help with being able to apply the trained model to future games played and other accounts. Finally with new data constantly flowing in the scalability of the model is another attractive quality. The second model I would like to use is gradient boosting machines for making winning predictions. This model type tends to have a high accuracy which will be needed for more practical results, however there are some drawbacks. Gradient boosting machines tend to be prone to overfitting data. Overfitting the data could overshadow important gameplay variables, meaning that players will lose out on potential improvements to win more games. The final machine learning model that I plan to use is random forest classification. The pros to using random forest classification is that it is robust like the artificial neural networks and is able handle large data sets really well. The main drawback with the model is that the data set is less interpretable, meaning that drawing conclusions from the results will be more difficult.

Tools

I plan to use the Riot Games developer tool to gather all of the data and run tests on my machine learning models. The developer tool will help with understanding what the variables are, navigating the data set and gathering new data easily from the internet. I also plan to use python for all of the coding in this project. I find python to be easy to use and implement the gathering of data from the API and the machine learning models. I plan currently to use the following python libraries: pandas, numpy, seaborn, matplotlib, sklearn, time and request. To help with the coding of this project I will be using various online tools such as youtube, chat gpt and various discussion forms.

Conclusion

This project aims to identify the most critical gameplay variables in League of Legends using quantitative analysis. By analyzing data from 500 personal games and employing machine learning models, the project seeks to predict game outcomes and validate these predictions with random accounts and future games. The specific models used will be: an artificial neural network, due to its ability to handle non-linear relationships and scalability, Gradient boosting machines, for high prediction accuracy and random forest classifiers for robustness with large datasets. The project will utilize data from Riot Games' API, encompassing over 100 variables such as objectives

taken, player statistics, and in-game events. The findings will highlight gameplay factors that influence success, offering tailored coaching recommendations to improve individual player performance.

Reference

Demirkol, O. (2024, September 27). *League of Legends Player Count*. The Spike.

https://www.thespike.gg/league-of-legends/beginner-guides/league-of-legends-player-count