Adam Murdock

Clark Ingram

<p align="center">Predicting Win rate from Game Data</p>

Our project's purpose is to utilize different types of data and statistics gathered from the game League of Legends to predict a professional player's win rate. We can then use this project to take data from any player's stats and use this tool to predict the expected performance of such players. Due to League of Legends' incredible popularity, there are several years of esport results we can pull our data from. This will allow us to gather information from previous years providing us with a vast quantity of data points to use as predictions. There is a need for this type of prediction tool, as teams are always scouting for new talent and these teams compete for large sums of money. This is a lucrative field, and one that has serious demand for any type of successful scouting tool.

League of Legends first came out on October 27, 2009. The game has several different "roles" in which said "role" defines the player's chosen playstyle. The roles, Top Lane, Middle Lane, Bottom Lane, Support, and Jungle, all have an important role within the game, which means that the performance of a single player could quite possibly change a game's outcome. The game's popularity over the years has skyrocketed, resulting in massive e-sport events. Professionals compete for incredibly large prize pools while signed for million dollar contracts.

Due to the massive scale that League of Legends operates at, there is increasing competition by not only the professional players, but also on betting sites. This has created a desire to have some sort of method or system to accurately predict player and team performances in matches. But this tool can also be used as a possible means of scouting for potential talent in upcoming players.

Traditionally, coaches and teams in North America will pay players from other parts of the world enormous amounts of money to come play for them. While this is great, there are

rules that limit the amount of talent scouted from overseas. This means that a majority of the team has to be from that team's home country. So this tool can be beneficial in finding talent to fill out the rest of the slots, because North America teams struggle to find good players from North America.

For our process of execution, first we went to a website that hosts data and stats from all the League of Legends esport events, ranging from the summer of 2020, back to the spring of 2013 in some cases. We attempted to write a scraper which would gather the information for us, and dump it into a spreadsheet, however we failed to get it to run efficiently due to the weird formatting of the website, and we found it much more efficient at this point to copy and paste the data into a spreadsheet ourselves. After spending a few hours copying about 5000 rows of data into spreadsheets spread out by year, we then realized it would be more efficient to compress the data into fewer spreadsheets. After eventually condensing all the information into a singular spreadsheet, we then cleaned up some of the data that was incomplete or unrecorded, such as if a player's amount of gold collected was not gathered, it wouldn't be conclusive data. We then took this data turning it into a csv file, then created a script to translate them into a npy format so as to allow the computer to quickly read the data we provided it. After having the data prepared, we went to work creating a neural net that would attempt to predict outcomes using the formatted data mentioned beforehand. We started the predictions by giving weights we thought would work and then we ran it for a large amount of cycles. We even attempted to modify the data by giving a player the winrate of their next season, so as to predict their next season's results, however we found the process time consuming and inconclusive. We continued to update the weights to what our code output, until we got weights that created errors and predictions that were reliable and acceptable.

Our results seem like they can be pretty promising. After a couple of iterations of fine tuning and changes, we got to a point where the predictions were pretty accurate. However as it ran, the prediction was skewing further away over time. While we do not know the reason behind this, we have our suspicions on the matter. Firstly, we believe that some of the data we chose can skew our results due to player health. Sometimes when a player cannot perform or for other reasons, one member of the team will switch out with another. This causes the switched in player, as they are like a reserve, to have less games and a more drastic win rate. Such as one player we found, as opposed to the normal 18 to 20 games, only played about 2 games, and had a 100% win rate. While that is great statistics for those two games, that doesn't accurately reflect this player's true stats due to the low amount of data provided. We believe that these examples, as this tends to be a common practice, subbing in players, are skewing our predictions and causing issues over time.

Currently our results could use some work, but we have made a lot of progress. To better fine tune our project, we would first go about finding a way to prevent subbing in players from skewing our data, whether it be by removing them from the data pool or by adding a minimum amount of games played to be allowed into our data pool. Another point of data that we do not interact much with is "Champions Played". Champions are characters the player can use, and this stat is the number of different champions they played during the season. While our code currently does not realize this, the more champions played often means the player is more flexible and often adapts to fit the needs of their team. We could also look deeper into the "Assists" statistic. Assists are when two players work together to defeat an opponent. We could use this to better analyze team statistics and find a way to calculate the win rate of a team based on their collaboration stats, such as assists and champions played, which highly suggest the level of teamwork. Overall it seems our project would end up being directed into a possible analysis of team data, and maybe even find ways to optimize and create better teams.

References

Aarushi Singh. 2019. How AI is Taking Predictive Analytics to the Next Level. (August 2019). Retrieved October 4, 2020 from
https://www.martechvibe.com/insights/staff-articles/how-ai-is-improving-predictive-analytics/

We plan to use this source as a "How to guide". Essentially we are using this as a reference guide for what our project should do, and what to expect while working on it. Also, if we ever get stuck or do not know where our next step should be, we intend to reference this article.

Kristian Medina. The history of League of Legends esports. Retrieved October 30, 2020 from
https://www.pinnacle.com/en/esports-hub/betting-articles/league-of-legends/history-of-league-of-legends/8sz2h2z4rrk83vs4

Used for general knowledge of the mechanics within league of legends and their esports history. Note, Clark is a nerd and kinda understands the game, but Adam is the real big brain here.

Lolesports Staff and Lolesports Staff The collect. 2019 LCS Scouting Grounds. Retrieved October 31, 2020 from
https://nexus.leagueoflegends.com/en-us/2019/06/2019-lcs-scouting-grounds/

General knowledge of the scouting scene, especially for the focus on future work. We also used this to attempt to place values on our weights based on what they deemed as important.

Seouk Jun Kim. 2020. Match Prediction in League of Legends Using Vanilla Deep Neural Network. (April 2020). Retrieved November 14, 2020 from
https://towardsdatascience.com/match-prediction-in-league-of-legends-using-vanilla-deep-neural-network-7cadc6fce7dd

Talks about different appeals to studying games within the targeted genre. While Dota is easier to access, it is not free. While League, its harder to access but it is free data and there is a lot more of it. It also talks about approaches and different ways we can score the data. A fantastic example on ways we can approach our project in ways that can create meaningful data.

Victoria Hodge, Sam Devlin, Nick Sephton, Florian Block, Anders Drachen and Peter Cowling. 2017. Win Prediction in Esports: Mixed-Rank Match Prediction in Multi-player Online Battle Arena Games. (November 2017). Retrieved October 4, 2020 from

https://www.researchgate.net/publication/321160741_Win_Prediction_in_Esports_Mixed-Rank_ Match_Prediction_in_Multi-player_Online_Battle_Arena_Games

This article was written about another MOBA game which shares a lot of similarities to League of Legends. We plan to use this as a parallel to our project and examine the steps and processes they used while we create ours. There will be times where we compare our results and see if the methods and coding would be beneficial to us as well.

2020. Predicting the Winning Side in Esports Using Tensorflow (2). (April 2020). Retrieved October 4, 2020 from https://mc.ai/predicting-the-winning-side-in-esports-using-tensorflow-2-2/

This source talks about a very similar idea to ours, however the stats focus on predicting a single game, whereas our scope is somewhat larger, as it focuses on an entire season. This source will help us as we will take this data and information to better our project. Taking in this information surely will help us to better refine our searches and data usage.