Predicting The Top 4 Teams in

League Of Legend Esports

Christian Juarbe

1.Business Understanding: League of Legend(LoL) is a multiplayer online battle arena(MOBA) video game developed and published by Riot Games. It was released in October 27,2009 and received well by the users. Even after years the game was released, the popularity of LoL was on the rise. It became so big that in 2011, The first major LoL Tournament was held in Jönköping, Sweden for a prize pool of $100,000. This marked a big step into making LoL the biggest esport game as of today. So now you are asking what is esports and why does it matter? Esport simply means Electronics sports where video games are used to compete with other players somewhat similar to regular sports. Esport are on the rise and it seems it won't be stopping anytime soon. One of the biggest esport prize pools for winning had $219 million dollars involved showing how much money is being invested in esports. Even the International Olympic Committee has acknowledged the potential of esport and are considering, in the near future, to consider it as a legitimate sport. So with the esport market making a rise, esport Teams are attempting all possible actions to achieve any kind of advantage over other teams.

The problem we end up running into is how do we determine what makes a team win, in particular, in LoL. LoL's main objective to win is to destroy the enemy bases nexus. You can think of it as destroying an enemy main headquarters which determines if you achieve victory or defeat. The only way to achieve this is by getting through towers of the enemy base to get access to the enemy nexus. What is important is that depending on the skill of the player, the champions they use(Playable character in game) and strategies determine how they will get and the difficulty to the enemy nexus. Since the game has so many options within it such as champion, items(in game tools to help strengthen your champ) and tactics, it can be difficult to determine your chances of winning. So what can help us determine a team victory are in game stats and out of game stats to see what stats do the victors have and how much impact it has on their victory. With this information, It can help esport and regular players increase their chance of victory in League of legend.

2. Data Understanding: So I used two sources of data to learn more about the way victory is achieved in League of Legend. The first one,which we will call Esports data(edata), is from a website called gol.gg which records all stats from all pro players from esports tournaments and playoffs. So I collect data regarding all teams that played in Season 4,Season 5, Season 6,Season 7, and Season 8 by using pd.read. I collected the data by using pd.read\_csv from the url. My second source of data comes from kaggle which is information about 10,000 league of legend ranked diamond games. More so data exactly 10 mins in each game.We will call this data Ranked Data. Thankfully the data was already cleaned up and prepared by the author Michel`s fanboi. He collected the data by using the league of legend api tools.

3.Data Preparation: With Ranked data, there wasn't much to prep it for. The data is clean and ready to be analyzed. All I did was added a column for red wins and gave it a 1 anytime the blue team lost. I did this to see how closely related the win conditions are for each side. I had to put in a lot of work for Edata to be clean and ready for analysis. First I had to drop season 5 because there was too much missing data to try to replace it with the mean of that column. Second I had combined each season dataframe into one big data frame.Next I had to drop two columns called Herpg% and Herpg. I drop these columns because it is a feature that did not exist in season 4 and has a lot of missing data in other seasons . Next I have replaced any leftover missing data with the mean of that column.Then I added a column called top 4 which had the top 4 teams that made it in the World championship. Finally I changed any column types that were objects to a float so I could use them in statistical analysis.With that being clean, I can finally analyze the data and get meaningful information from the data.

4.Modeling/Methodology: With both Edata and Ranked data, I used the correlation matrix to help determine what independent variables having the highest correlation with the y target variable. Looking at the ranked data, I picked out independent variables based on the correlation value of the columns with respect to the victory of the blue team. Any correlation with a positive value equal to or more than 0.25 were included to be the independent variables in the model. I believe that a positive correlation of .25 for more would make the model very accurate. I ended up with 8 variables. I will consider in the future to add high negative correlation because It can have a good impact on the model's accuracy. Then I set the target variable as the blue victory. I also made another model based on the red wins and see if the same opposite columns will give the exact same result as the blue model. For esports data, When I look at the correlation matrix, The correlation was not high for the top 4 in the data. So I picked columns that had a correlation of .15 or higher with top 4.

5.Evaluation/Results; After testing both models and looking at the two different data sets, I was able to come to some clear results on winning in league of legends for esports players, teams, and regular players. We will start with the esports data. I was able to tell that being able to predict the top 4 of the world championship based on the summer split with these columns would be impossible. The model without weights wasn't able to predict the team at all. The model with weights was only able to predict the right team 9% of the time. What it does show is that esport games from summer split are much harder to predict the top 4 of worlds with this set of data features. If we have more data in both the amount of rows and columns we might be able to predict the top 4 winners of worlds based on the summer split. What I want to do when going back to revisit the data is instead of doing a top 4, make a model to predict what team makes it to the league of legend world championship(Top16). I believe the model would have enough data to make a good prediction to what team would make it in the top 16 and could be used in a real world prediction. We would also be able to see the correlation with the data and what would be the most important features for a team to be in the top 16.

Now when we will look at the ranked data we were able to determine what features are most important. With the features we picked for the models, the logistic regression model has a 73% accuracy. That is not bad especially considering this is only 10 minute data into the games. Also what's even better is that now players have a better idea of how to achieve victory by looking at the feature in the model. I can even try to improve it by adding the high negative correlation to see if it will improve the model's accuracy. Another option would be adding even more data to improve the model`s accuracy. There are a lot of options for this data and model to have impact esports and players by improving their chance of winning. I believe the model with more data can be improved to a 85% accuracy. To further check how valid the data is, I also created a model for the opposite team victory and found it was exactly the same as the blue team model. So it's safe to say the game is mostly balanced. It won't be perfectly balanced because each side of the map is different but the winning conditions are the same for both sides.

**6.Conclusion:** I was able to learn a lot from this capstone project and the models. One thing I learned was the amount of data necessary to make a great machine learning model. In my Esport data, It has potential for growth and even becoming a well used model but it simply does not have enough data yet. So with enough time, It can be used in the future of league of legend esports. Also making a top 16 can also make the model more accurate and will still be useful for an esports team. Next would be the ranked data and the impacts of the results. We were able to determine the most impactful features for Victory in league of legend. We are also able to determine that the winning condition for each side is exactly the same. Finally we were able to build a model that can make a prediction of what team will win based on 10minute statistics. Overall I was able to learn the important aspects of machine learning and how to use the data to find out new information about it as well to present that information to other people.

g