DS 740 Final Project

Glenn Niles

League of Legends (LoL) is an online multiplayer online battle arena (MOBA) game published by Riot Games in 2009. The game pits two teams of five players controlling characters picked from a large roster attempting to destroy one another’s bases. As esports continue to grow there will be a larger market for professional teams to be begin to understand how the game works and what factors come into play that enhance each team’s chances of winning. The growth of esports will also create a need for Riot Games to carefully look at game balance to ensure certain characters or team compositions don’t create situations that are unwinnable against teams that don’t choose a setup that is optimal to all other setups.

For this project, a dataset of 9,879 from high-ranked games was collected, using 40 variables as predictors. The data is largely split between “blue team” and “red team” where 19 variables were tracked for each team. The 39th variable tracks whether the blue team or red team won the match. The 40th variable was the unique game identifier, which was removed from the dataset. The focus of this project was to determine which actions taken by the players had the largest affect on winning a game of LoL. Because of this, an attempt was made to focus on what made each team different during each match. This had an added upside of reducing the number of predictor variables there were to work with.

Seven of the variables were eliminated by finding the difference between blue team’s performance and red team’s performance. The different number of wards placed, the amount of kills, the number of towers destroyed, the difference in average experience level, the difference in total minions killed, and the difference in jungle minions killed were all used to create new variables that only looked at the difference between both teams.

Several variables were removed due to having high correlations with other variables in the dataset, some of these were removed before analysis. Since deaths are inversely related to kills, there would be one kill for every one death. Kills grant experience points and gold while deaths put that player on a respawn timer. Since kills are more rewarding, I removed the deaths variable to make analysis easier. Similarly, assists were highly correlated with kills but don’t offer as much reward as the person who gets the kill, these were also removed. The heralds variable kept causing trouble and seemed like it didn’t add much to the analysis of the dataset in the times that it did work so it was removed.

Experience points are highly correlated with character level. A character gains a level after crossing an experience threshold. Gaining a level provides the character with stat boosts to their attack strength, health, skill points, and at certain levels, new abilities a player can use. Levelling up creates a greater effect than individual experience points and is based upon experience gained. Also, a character is capped off at level 18 but still earns experience points, these extra experience points aren’t used, because of these reasons, experience gained was dropped from the dataset.

CS per Minutes and gold per minute were dropped from the dataset due to them being highly correlated with minions killed and total gold respectively. It would be interesting if experience per minute were tracked as well and to run an analysis as a time series to see how much each of these values would go towards predicting a winner. Gold difference and experience difference are highly correlated with total gold and total experience so these were also removed. This left us with a dataset with 15 variables, much less than the 40 that were in the original dataset.

Analysis began be creating histograms for each variable. Since the difference was taken on many these variables, a nice bell curve was achieved for most variables. Stepwise regression was applied to the data first, using both forwards and backwards steps. Both directions gave very similar results and found that blue dragons killed, total gold from both teams, towers destroyed, average level and total minions and jungle minions killed seemed to be the most important variables. The best model for backwards stepwise regression had an AIC of 10481.74.

Next, logistic regression was run on the dataset and the VIF of each variable was found. After running this, it was found that the VIFs for each variable was reasonably low, with kills having the highest VIF at 7.97. Most variables were around 1. Going from there, linear discriminant analysis was run on each of the variables found using stepwise logistic regression based on increasing p-values using double cross-validation. The cross-validation seemed to favor the models that had six and seven variables. The model predicted 0.731 correct values. This is not a great value but it’s not terrible and could be improved in the future.

A support vector machine was run on the dataset using various cost values and gamma values. This was left to run overnight as it was very resource intensive. The best parameters were found to be a cost value of 0.1 and a gamma value of 0.5. This gave us an error rate of 0.29 which is similar to the values from running the linear discriminant analysis (0.27). This once again, is not ideal but also is better than randomly guessing at who would be the winner.

Overall, I don’t think this analysis uncovered a great deal about what makes a team win in a game of League of Legends based on the variables we have to work with. While average level and the number of dragons killed seems to play a big role in giving players an advantage, I don’t think any of these variables are sufficient for to predict a winner accurately. This analysis would be better run as a time series, I believe, where gold earned, experience earned and minions killed per minute could paint a better picture of how successful teams play in the early game to gain an advantage on the other team. I suspect that team compositions and character usage play a higher role in winning teams than the individual actions of the player. Player skill more than likely plays a big role in individual games and tracking each individual player throughout a game would probably be more enlightening than focusing on the teams.

A screenshot of a cell phone

Description automatically generated

Figure 1Number of predictors LDA

Link to dataset: [https://www.kaggle.com/bobbyscience/league-of-legends-diamond-ranked-games-10-min/data#](https://www.kaggle.com/bobbyscience/league-of-legends-diamond-ranked-games-10-min/data)