15.095 Machine Learning Project Proposal

Predicting League of Legends Ranked Match Wins

Team Members:

- *David Liszewski, dliszew1@mit.edu*
- *Tatdanai Asavamongkolkul, tatdanai@mit.edu*

Problem Summary

The eSports industry has been growing, and continues to grow rapidly. In 2016, viewership had reached 335 million people, and is expected to reach an audience of 557 million in 2021[1]. League of Legends (LoL) is widely considered the most popular eSports game, leading all eSports titles in total hours watched and peak viewership in 2019[2]. The game features 2 teams of 5 players each, who play as characters (champions) and work together to destroy their opponent's base (nexus). As LoL's popularity has grown, so has the monetary prizes for winning the League of Legends World Championship, the ultimate prize for the best LoL team in the world. The 2018 World Championship had a prize pool of $6.4 million, with the grand prize set at $2.4 million[3]. Through the popularity and high stakes of the game, it naturally leads to a desire to understand the reasons and factors behind their wins. A better understanding of what controls the outcomes of a match could give players an advantage on their opponents. Through this increased knowledge, the decisions a team makes before the match, e.g. champion selection strategy, could be impacted in a significant way. Better decisions may then lead to increased success and increased earnings for both teams and players.

Data

Past research[5] on the similar problem used limited data that focused on the regular players which have a wide range of skill sets. In this project, we will focus on the pro players level; open source data on the League of Legends' tournament in all regions are used[4]. The data includes around 15,000 matches from 2018 to 2020. The raw data will be cleaned and manipulated into an appropriate format with 123 features; including match-specific, champion-specific, and players-specific features. The list of all features is presented in the Appendix, Table A.1. More features may be added as needed. Note, the data manipulation and initial analysis will be done in R, while the machine learning and optimization tasks will be done in Julia.

Approach

In this project, both predictive and prescriptive methods will be applied.

<u>Predictive methods</u>: The data will be split into training, validating, and testing sets. As the effect of each factor might change over time, further analysis will be done to determine the best way to split the data, discussed in the "challenges" section below. Past research[5] has mainly considered logistic regression, random forest and neural network methods. Since interpretation is considered important for this project, we will build on past research by first applying interpretable cutting-edge methods such as Optimal Classification Tree (OCT) and OCT-H methods. These will be compared with the conservative methods such as logistic regression and CART. Then, if time allows, random forest and boosted trees will also be

studied and compared with the former models. These models will be used to understand the problem better, and make conclusions about which features are most important for winning matches.

Prescriptive methods: In this step, we will use a data-driven approach such as Optimal Prescriptive Trees (OPT) to enable the player-specific decision making. This will allow us to make direct recommendations to players on champion selection strategy. The results will be compared with the predictive approach. If time allows, other models such as prescriptive random forest will also be studied. The ideal method would be interpretable enough for players to understand the decision-making process.

Challenges

One of the main challenges of this project is determining where to split the data for training and testing. Over the years that our data is found, the game itself has received many updates, with new versions improving on previous versions. These updates can affect champion strength and team dynamics. Therefore, a model trained on 2018 data alone may not be able to predict well on 2020 data and therefore contain biases, as the game may have inherently changed over the two years. Further analysis and research will be done to determine the best method of data splitting.

References

[1] "The Incredible Growth of ESports [+ ESports Statistics]." Influencer Marketing Hub, 6 Jan. 2020, influencermarketinghub.com/growth-of-esports-stats/

[2] Fitch, Adam. "The Most Watched Esports Events of 2019." Esports Insider, 31 Mar. 2020, esportsinsider.com/2020/01/esports-viewership-2019/

[3] Heath, Jerome, and Sam Nordmark. "The 10 Largest Prize Pools in Esports." Dot Esports, 3 Sept. 2020, dotesports.com/general/news/biggest-prize-pools-esports-14605.

[4] https://oracleselixir.com/

[5] Yin, Jihan. "Predicting League of Legends Ranked Match Outcomes with Machine Learning," 2018, https://hackernoon.com/league-of-legends-predicting-wins-in-champion-select-with-machine-learning-6496523a7ea7.

Appendix

Table A.1: list of features

Feature No.	Match-specific Features
1	Starting side (blue, red)
2	League type (US, EU, etc.)
3	Playoff (yes, no)
	Champion-specific Features (5 per sample)
4-8	Average win rate of the champion
9-13	Average win rate of the champion for each corresponding role
14-18	Average win rate of the champion against its matchup
19-23	Average play rate of the champion

	Player-specific Features (5 per sample)
24-28	Win rate of the player
29-33	Average KDA ratio
34-38	Average relative KDA ratio
39-43	Average relative death
44-48	Average rate of being a victim of first blood
49-53	Average rate of getting a first blood
54-58	Average damage per minute
59-63	Average wards placed
64-68	Average wards killed
69-73	Average total closing-shot
74-78	Average jungle monster killed
79-83	Average enemy jungle monster killed
84-88	Average gold spent
89-93	Average closing-shot different against opponent at 10 minutes
94-98	Average gold different against opponent at 10 minutes
99-103	Average experience different against opponent at 10 minutes
104-108	Average closing-shot different against opponent at 15 minutes
109-113	Average gold different against opponent at 15 minutes
114-118	Average experience different against opponent at 15 minutes
119-123	Total games played with the chosen champion