ISYE 6740 – Spring 2021
Project Proposal

Team Member Names: Richard Moss

Project Title: Developing machine learning algorithms to identify exceptional League of Legends players using Korea's LCK Data

Problem Statement

League of Legends has been one of the largest video games in the world for more than 10 years. Built for competitive gameplay, League of Legends is a five-versus-five Multiplayer Online Battle Arena (MOBA) style game, in which teammates must pilot a selected character to accomplish a goal in opposition to an enemy team. In League of Legends, this goal is to destroy the enemies base, called a Nexus. To gain advantages, teams can complete objectives, farm minions, and kill enemy champions to achieve additional experience and gold, allowing them to buy items and increase their strength.

From a competitive perspective, the League of Legends ecosystem is made up of 4 major leagues in China, Korea, North America and Europe, 9 minor regions in Brazil, the Commonwealth of Independent States, Vietnam, Latin America, the Asia-Pacific Region, Brazil, Oceania, Turkey and Japan and over 100 amateur, regional, collegiate and high school leagues worldwide. With 180 million players worldwide, the League of Legends competitive scene is capped off each year with the World Championship. In 2019, the World Championship drew approximately 100 million unique viewers, as compared to the Super Bowl in that same year drawing 98 million (Pei, 2019). From an industry perspective, the top 10 most valuable teams are worth a combined $3.5 billion as seen in Figure 1, and generate around $1 billion in revenue, with ownership groups comprised of professional sports teams and players, venture capital groups, and millionaire investors (Knight, 2022).

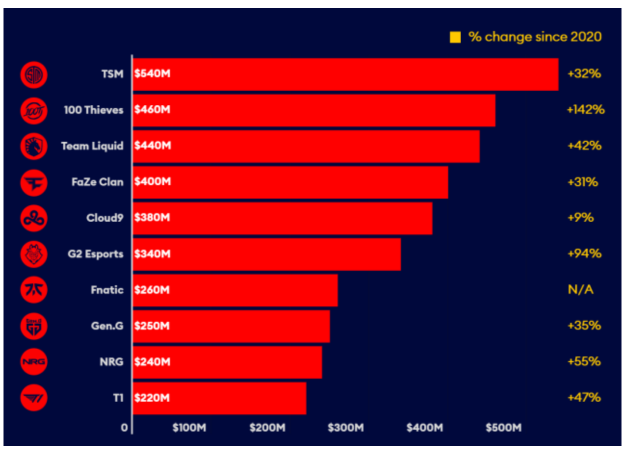

Figure 1. 2022 Forbes Most Valuable Esports Organizations

One of the most determining factors in popularity of organizations is winning. All teams on the top 10 list shown above have shown a repeated propensity to win their respective regions. As an example, the top 3 teams all compete in League of Legends' North American Championship (LCS). TSM is the "winningest" team of all time with 7 championships, Team Liquid has 4 and is the only team to win 4 straight titles, and despite being a newcomer to the League, 100 Thieves has 1 title and has attended the last 4 consecutive finals and 2 world championships. As such, it is clear that finding winning players is key to organizational success and profit. To find these players, this paper will develop machine learning applications based on players in Korea's League of Legends' Champions Korea (LCK). The LCK as a league is the dominant force winning the overwhelming majority of World Championship winning 7/11, and every Korean team except ONE has failed to progress to at least the top 8 of every team to ever attend.

Methodology

This application will be developed using two main datasets. The first dataset comes from Leaguepedia (Leaguepedia, 2022). Each split, (1/2 of a year) players play 18 best-of-three series. After each game in the series, a player of the game is awarded, and this running total is tracked to identify the best players. An example of this list from the 2022 LCK Spring Split is shown below.

"Player of the Game" Standings

Player of the Game Standings		
Rank	**Points**	**Player**
1	1600	Canyon
2	1000	Chovy
2	1000	Deft
4	900	Faker
4	900	Lava
Click for Full Standings		**[hide]**
6	800	Keria
6	800	Rascal
8	700	Doran
8	700	Dread
8	700	FATE
8	700	Gumayusi
8	700	Peanut
13	600	Oner

Figure 2. LCK Spring Split 2022 Player of the Game Standings

This data is tracked for every split going back to 2012. It allows us to identify the players who have best impacted winning for their team and will be used as the response for our algorithms. The second dataset will contain our predictor variables and comes from Oracle's Elixir (Oracle's Elixir, 2022). While this data goes back to 2012, detailed predictors are available after 2015 therefore Spring 2015 will act as the starting point for this machine learning application. Each year between 2015 and 2022 we have two splits of player data, for a total of 16 splits of data. The player's statistics track their kills, deaths, assists and advanced statistics such as gold difference compared to their enemy and percentage of their team's damage amongst many others. These statistics are averaged over each game the player plays that split, as shown below.

Player Stats > LCK 2022 Summer Teams Champions Download This Table

Filter Positions | Filter Teams | Filter Players | Min. GP | Both Sides | All Results | 12.10 to 12.14

Jun 14, 2022 → Aug 14, 2022

Player	Team	Pos	GP	W%	CTR%	K	D	A	KDA	KP	KS%	DTH%	FB%	GD10	XPD10	CSD10	CSPM	CS
Aiming	KT Rolster	ADC	45	56%	44%	200	66	197	6.0	71.5%	36.0%	14.3%	18%	164	-7	5.7	9.8	3
Aria	KT Rolster	Middle	9	33%	56%	24	16	42	4.1	74.2%	27.0%	17.2%	33%	-190	-93	-5.9	9.0	2
Bdd	Nongshim RedForce	Middle	42	33%	60%	111	98	209	3.3	69.4%	24.1%	18.4%	12%	-7	-16	0.8	8.6	2
BeryL	DRX	Support	45	49%	49%	35	150	354	2.6	73.1%	6.6%	26.8%	24%	147	151	7.4	1.9	[illegible]
Bible	DWG KIA	Support	5	20%	40%	2	23	35	1.6	67.3%	3.6%	30.7%	40%	55	54	2.2	0.9	[illegible]
Canna	Nongshim RedForce	Top	42	33%	48%	89	104	185	2.6	59.4%	19.3%	19.5%	19%	123	80	2.1	8.3	2
Canyon	DWG KIA	Jungle	41	59%	59%	93	93	269	3.9	72.3%	18.6%	22.7%	29%	-93	31	1.1	6.3	1
Cheoni	Hanwha Life Esports	ADC	14	14%	57%	30	32	34	2.0	72.7%	34.1%	17.7%	21%	-447	-184	-9.6	9.0	3
Chovy	Gen.G	Middle	40	88%	53%	133	52	231	7.0	64.5%	23.6%	16.5%	20%	196	115	5.0	10.0	2
Clozer	Liiv SANDBOX	Middle	41	66%	39%	130	77	234	4.7	65.7%	23.5%	18.1%	20%	3	68	1.0	9.0	2
Croco	Liiv SANDBOX	Jungle	41	66%	54%	103	104	288	3.8	70.6%	18.6%	24.4%	39%	-2	-16	-1.5	5.6	1
Cuzz	KT Rolster	Jungle	45	56%	51%	108	98	276	3.9	69.2%	19.5%	21.2%	40%	61	5	1.2	5.6	1
Deft	DRX	ADC	45	49%	56%	149	104	231	3.7	71.4%	28.0%	18.6%	24%	-27	-64	-5.8	8.4	2
Delight	Fredit BRION	Support	42	26%	52%	19	111	234	2.3	70.3%	5.3%	21.7%	36%	-100	-47	-5.5	1.2	[illegible]
deokdam	DWG KIA	ADC	41	59%	51%	174	75	193	4.9	73.3%	34.7%	18.3%	34%	-29	-50	1.2	9.1	3
Doran	Gen.G	Top	40	88%	65%	90	75	227	4.2	56.2%	16.0%	23.8%	25%	62	72	1.2	8.6	2
Dove	Liiv SANDBOX	Top	41	66%	54%	79	86	270	4.1	63.0%	14.3%	20.2%	20%	-281	-162	-8.5	7.1	1

Figure 3. LCK Player Data from Oracle's Elixir

Some preprocessing will be done to prepare the data for analysis. First, the two datasets will be combined on team, player name and split so that each player's player of the game points are stored with their split averaged statistics. Any players with less than 10 games played will also be filtered out of the dataset to avoid substitute players with a small sample size of games from skewing the data. Finally, we can see there are different ranges for each of the categories, and the values of the categories tend to improve year-over-year as the game changes and players improve. As such, each predictor will be standardized for each year, such that players who are exceptional in their year over a number of predictors are identified and not compared to players in other years where the game may have changed.

Feature selection will be done in order to remove non-predictive features such as the players name, team name and position. We can also find multicollinearity within the feature set by finding the variance inflation factor of the predictors and identify related features which add linear dependency in the predictor set. The dataset will be split such that 14 random splits are used for training, and 2 for testing.

For analyzing the data, a few models will be built in order to find the best method of identifying outstanding players. First, multiple linear regression, ridge regression, stepwise regression using Bayesian information criterion, Lasso regression, elastic net and random forest will be used. From these models we can find the explained variance in the model, the error in the test and training set, and for some methods we can also perform feature selection. This will allow us to determine the model which best explains the variance in the player's player of the game rating. Next, we can take the coefficients from this model with the largest absolute values and identify them as features with a large impact on player performance. The reason absolute value is used rather than large positive value is since some variable with low values may impact player performance in a positive fashion, such as having less deaths in a game. Finally, using these coefficients we can describe players which best impact team success, and use this to create a statement which describes an ideal player, and can be applied to scouting the general high-ranking player base to identify prospective talents.

Works Cited

Knight, Brett. (2022). *The Most Valuable Esports Companies 2022.* Published in Forbes. Web.

Leaguepedia. (2022). *League of Legends Champions Korea.* Published in Lol Fandom. Web.

Pei, Annie. (2019). *This esports giant draws in more viewers than the Super Bowl, and it's expected to get even bigger.* Published on CNBC. Web.

Sevenhuysen, Tim. (2022). *Player Statistics by Split*. Published on Oracle's Elixir. Web.