Tian Wang. Predictive Analysis on eSports Games – A Case Study on League of Legends (LoL) eSports Tournaments. A Master's Paper for the M.S. in I.S. degree. April, 2018. 46 pages. Advisor: Arcot Rajasekar

This paper discusses about how different features influence League of Legends eSports game results. Logistic regression and decision trees are used as the main predictive analysis methods for making a prediction. In this study, three types of features: champion selection, in-game factors, and player performance, are tested to see how closely they are related to the game results. This paper also compares the similarity and difference between predictive analytics on traditional sports and eSports games, and discusses about potential approaches to improve eSports prediction accuracy in the future.

Headings:

Predictive Analysis

Information Analytics

Machine Learning

PREDICTIVE ANALYSIS ON ESPORTS GAMES – A CASE STUDY ON LEAGUE OF LEGENDS (LOL) ESPORTS TOURNAMENTS

by
Tian Wang

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

April 2018

Approved by

____

Arcot Rajasekar

Table of Contents

1. Introduction

Today, eSports, which is defined as well-organized multiplayer video game competitions between professional players, is becoming more and more popular with an increasing availability of online streaming media platforms. The most successful eSports competitions include Dota 2, League of Legends, Overwatch, and Counter-Strike: Global Offensive. E-Sports has been drawing people's attention, and the established fan base has been rapidly increasing over the past few years. In 2013, there were about 71.5 million people in the world that watched competitive eSports games. Particularly, 32 million people tuned in to the League of Legends World Championship Season 3 that year (Warr, 2014). It is obvious that the popularity of eSports games has been as much as, if not more than, physical sports like basketball or soccer. The audience for eSports was larger than the one that watched the last game of the NBA finals in the same year (Segal, 2014). In 2017, the global audience is expected to reach 385 million, with 191 million of them being regular viewers and others being occasional viewers (Hattenstone, 2017). The growth of the eSports economy has also proved the future potential of this field. According to 2017 Global eSports Market Report, eSports economy is estimated to grow to $696 million with a year-on-year growth of 41.3% in 2018 (Warman, 2017). Investors treat eSports as a field with positive impact on game revenues and being a stand-alone business in the future.

Forecasting has always been an important role in sports. Every major professional sports team has an analytics department or expert. These analysts will use player's statistics to predict future games and try to increase win rate by draft selection or roster adjustment. Similar to game predictions in traditional competitions like basketball (NBA) and football (NFL), predictive analytics can also be used to predict eSports games. However, eSports games can be more complex. Although currently there are already many well-designed approaches for sports analysis, methodologies on most eSports predictions have yet to be discovered. Compared to physical sports, eSports prediction would involve more information because the selection of characters could also be the factor that influences game results besides players' statistics. To be more specific, in traditional physical competitions, only players' and/or teams' performances need to be considered to predict the game since there is no other factor that can have a huge impact. Also, while most common video game genres for eSports are real-time strategy, fighting, and multiplayer online battle arena (also known as action real-time strategy, is a genre of video game, in which each player controls a single character in one of two teams and the objective is to destroy the opponent's main structure), besides player's performance, each character in the video game will have its own abilities, which lead to different advantages and disadvantages against each other.

Although eSports prediction is more difficult than traditional competitions because of complicated information, it is still worth exploring predictive analytic methods for this particular field, considering its huge impact on different parties. By using predictive models, eSports teams can set up specific strategies against different opponents based on the forecasts. Players will have a strategic advantage by predicting what the opposing team will do next. As for people involved in broadcasting, such as analysts and commentators, predicting draft selections will be useful for conversations and discussion during the drafting phase of a match.

In this paper, previous work on eSports prediction will be reviewed and discussed. After defining the problem and explaining the background, multiple statistics tools and predictive analytic models will be used to predict eSports games based on data from player's previous performance, draft selection, and features from the game itself, including character information and real-time game process statistics. The goal of the designed methodology is to answer following questions:

1) Which predictive analytics model(s) can be applied to eSports game prediction? How does each model perform with the dataset?
2) How does predictive analytics on eSports games differ from other prediction models, especially comparing with "traditional" sports analysis?
3) What are the advantages and disadvantages for predictive analytics in eSports?

2. Literature Review

2.1 eSports

Synonyms of the term eSports include electronic sports, cybersports, gaming, competitive computer gaming, and virtual sports (Jenny, Manning, Keiper, & Olrich, 2017). Although eSports has been popular worldwide in these days and been more accepted as a sport, and gamers are identified as athletes within society, the debate still continues about whether eSports can be truly considered as a sport. The conceptual quandary turns to be a pertinent issue for defining eSports and drawing boundaries for people's general understanding of what constitutes a sport. Overall, sporting activity still focuses on the body and physical activities (Witkowski, 2012), but people may argue that capabilities of eSports players are not measured by their physical prowess or finesse since they only sitting on their chairs while playing (Hamari, Max Sjöblom, 2017). However, eSports could be physically taxing based on the way that players control the game states of the game's software or system. For example, players are physically interacting with the computer in dancing video games.

As the same way of "traditional" sports having sub-cultures, there are different genres of eSports, such as multiplayer online batter arenas, first-person shooters, real-time strategy, or collectible card games. eSports can be classified into two eras: early arcade era from the 1980’s to 1990’s, in which eSports genres like NBA Jam and Virtua Racing being the most popular ones; the Internet era, in which other genres eSports games became more popular with the evolution of the Internet (Lee, Schoenstedt, 2011). A significant change from arcade era to the Internet era was the mode of eSports consumption, which changed from human-versus-machine to human-versus-human (Griffiths, Davies, & Chappell, 2003). This change of mode allowed players to participate in the tournaments and to play against each other.

Given the characteristics mentioned above, eSports has a lot of similarities comparing to “traditional” sports. Both of them include competitive environments that provide displays of skill and prowess (Michaluk, 2012). Also, both sports and eSports games provide heated entertainment to the audience viewing the spot. Although it is slightly different since eSports are viewed from an electronic screen (Hewitt, 2014), they still have a similar general format for audience to view. Like traditional sports, eSports has also been recognized as good spectator sport because of increased amount of televising through both regional network channels and national broadcasting.

To distinguish eSports from “traditional” sports, a specific and clear definition of eSports is necessary. There are only a few definitions being proposed regarding eSports from past conceptual and qualitative literature on this field. To define eSports, Wagner (2006) extended the general definition of sports from “an area of sport activities in which people develop and train mental or physical abilities…” with the addition of “in the use of information and communication technologies.” However, this definition doesn’t distinguish difference between an electronic sport activity or a “traditional” one, since most sports are assisted or mediated by computer currently (Witkowski, 2012). The key factor comes down to where the outcomes of the sport are manifested. All outcome-defining activities happen in the real world in traditional sports, while they happen in a “virtual world” in eSports. Therefore, eSports could be more specifically defined as “a form of sports where the primary aspects of the sport are facilitated by electronic systems; the input of players and teams as well as the output of the eSports system are mediated by human-computer interfaces (Hamari, Max Sjöblom, 2017).”

2.2 Predictive Analytics on Traditional Sports

In traditional sport competitions, predicting events from past game results has been discussed for a long time, and there has been extensive research on this topic. Min, Kim, Choe, Eom, & McKay (2008) proposed a framework to predict sports matches results by using Bayesian inference and rule-based reasoning along with an in-game time-series approach. A Bayesian network model represents the conditional dependencies among uncertain variables that can be both objective and subjective (Constantinou, Fenton, & Neil, 2012). The motivation was that sports matches are highly stochastic, but teams’ strategies can be approximated by logic rules. FRES (Football Result Expert System), which was the prediction model implemented based on the framework, was proved to be able to provide reasonable and stable predictions after testing and examination. Meanwhile, in 2009, an empirical study for forecast accuracy was presented by Spann and Skiera. Data was collected from 678-837 games of three seasons of the German premier soccer league. The results showed that a weight-based combination of the forecasts methods leads to a slightly higher forecast accuracy, while a rule-based combination improves the forecast accuracy substantially.

Data mining techniques have also been applied to sports prediction. In a study by Cao (2012), a model was built by using machine learning algorithms, including Simple Logistics Classifier, Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Naïve Bayes to predict the NBA game results. After evaluating the results, the study concluded that Simple Logistics Classifier yielded the best results with the highest accuracy rate. Another study constructed by Haghighat, Rastegari, and Nourafza (2013) reviewed previous research on data mining systems to predict sports results and evaluated the advantages and disadvantages of each system. Similar classification techniques were discussed in this study, including ANN, SVM, Bayesian Method, Decision Trees, Fuzzy System, and Logistic Regression. The study proved that use of machine learning and data mining techniques could improve prediction accuracy, but it also pointed out the problem on lack of a general and comprehensive set of statistics and difficulties on data collection.

This problem was further discussed in a study on near-term predictions for professional basketball (Yue, Lucey, Bialkowski, Matthews, 2014). The paper pointed out that it was challenging to develop predictive models with spatiotemporal data given the lack of readily available semantically rich representations of the game. In order to get a compact data representation that enabled efficient prediction, the study employed a latent factor modeling approach with combination of discriminative learning techniques and techniques for spatial regularization and non-negative matrix factorization. The approach was successfully validated by spatiotemporal tracking data from the 2012-2013 NBA season and the results showed accurate in-game predictions. The study constructed by Pretorius and Parry (2016) discussed another problem of sports prediction, that predicting the outcome of sports games could be difficult tasks because of the complex relationship between variables of interest. A random forest classification algorithm was employed in the study to predict match results in the 2015 Rugby World Cup. The prediction model performed well with an accuracy rate of 89.58%, which proved the effectiveness and efficiency of the machine learning based system.

Models of predictive analytics have been used not only for prediction of game results and player performances, but also for traditional sports gambling. Prediction of sports games results began to enhance the anticipation for the game and increase the odds value of people's most favorite teams (Wen, Hung, Hwang, Lin, 2016). Buchdahl (2003) investigated the quality of bookmakers' odds in all kinds of sports, especially football. His conclusion was that a small feature set will suffice to beat the bookmakers' odds in football when enough data is used for the training set in a prediction system. Once a larger and more intelligent feature set is collected and analyzed, the possibility of prediction football matches could be more accurate.

There are more researches that focus on similarities of predictive methods between different sports. Buursma (2010) believed that predicting models on football games could also be applied to similar sports like basketball, baseball, and ice hockey. Researchers have been exploring different features that influence baseball game results in past few years. The build of predictive models on major league baseball (MLB) has evolved in several stages. A simple method of prediction was provided by Kaigh (1995), which only used the home and away records of MLB teams. Barry and Hartigan (1993) built a more complex prediction model that considered team strengths over time and required huge number of parameters. However, these models built in early years were not accurate since many factors such as individual pitchers' efforts were not considered during the predictive analytics. In 2004, Yang and Swartz used a two-stage Bayesian model and a Markov chain Monte Carlo algorithm to test various factors in order to evaluate the probability of winning a game in major league baseball (MLB). The model was proved to be effective by the results and it can be extended to other sports.

2.3 eSports Prediction Methods

Predictive analytics on eSports share similarities with traditional sports prediction. The two most popular prediction methods for baseball, PECOTA and MARCEL, can also be applied to player performance prediction in electronic sports (Shim, Sharan, Srivastava, 2010). The study showed that players' past performance plays a significant role on predicting their future performance as traditional sports. Prediction models were built on buckets using discretization based on binning and histograms tended to have higher prediction accuracy. This systematic studies also provided an idea that analysis of player performance in different dimensions such as player demographics and archetypes could help game developers so that they would know if the game and characters are played as intended.

Unlike traditional sports such as tennis or football where a match is separated by the structure of the game, online role-playing or real-time strategy games don't have a natural structure (Schubert, Drachen, & Mahlmann, 2016). One example is DOTA 2, which is a team-based competitive online game. Study constructed in this paper analyzed movements of characters in order to detect engagement between groups of players.

Another difference between traditional sports and eSports is that features of the game itself should also be considered in eSports prediction. Yang and Roberts (2013) predicted the success of an eSports team based on the hero selection. Besides draft selection, another feature of eSports game is that it includes real-time gameplay data. In a study of DOTA 2 prediction, Yang, Qin, and Lei (2016) developed two stages to predict the winning team of DOTA 2. Prediction in the first stage was similar to the one in previous work, which used different features to predict the game result before a match begins. In the second stage, they introduced real-time gameplay data to predict during a match besides only considering pre-match information.

Several methodologies were discussed when predicting eSports match results, one of them was the analytical hierarchy process (AHP) data mining method used by Aryanata, Rahadi, and Sudarmojo (2017) in their study of DOTA 2 matches prediction. Four parameters (experience per minute and gold per minute, matchmaking ratio point, 3 head to head matches result, and the last 10 results matches) were used to predict the results of international DOTA 2 match. Results of model testing showed that AHP method performed well on game prediction based on information obtained before the match. The Hidden Markov Models were used by Tamassia, Raffe, Sifa, Drachen, Zambetta, & Hitchens (2016) to develop a churn prediction model for an online multi-player game called Destiny. The paper also evaluated the area under curve (ROC) of behavioral features in order to predict churn in Destiny.

Previous work also included outcome prediction on League of Legends (LoL). Ong, Deolalikar, & Peng (2015) built a framework using unsupervised learning to find LoL eSports players' behavior clusters, along with classification algorithms, to predict the outcome of matches. 21 features of player statistics such as average damage dealt and money earned were selected and evaluated by k-means and DP-means clustering models. The paper pointed out that time-dependent player statistics features (for example, how players perform early and later in the game) would improve the accuracy of predictive models.

Chen and Joachims (2016) studied eSports prediction from another perspective. The motivation of their study was that matchups and comparisons typically happen in different contexts that can alter the outcome. Besides the players' themselves, characteristics of the game, such as the time the game is played and the importance of the game may also have influence on the player's performance. Rote learning, Bradley-Terry model, and pairwise logistic regression model were applied to the framework described in the study.

Summerville, Cook, and Steenhuisen (2016) took further steps on eSports predictive analytics. In their paper, machine learning techniques were used to analyze the drafting phase of DOTA 2 matches. The paper discussed about the future of eSports analysis: besides the goal of helping game developers to evaluate the game and game characters that mentioned in previous works, with a wealth of available data, eSports games could be an area with huge potential, and predictive analytics on eSports games will have significant impact on broadcasting, organization, and competition around eSports.

2.4 League of Legends

League of Legends (LoL), a multi-player online battle arena video game developed and published by Riot Games, is a fast-paced, competitive game that blends the speed and intensity of a real-time strategy with role-playing game (RPG) elements. Details of LoL game play can be found on Riot's official website ("Game Info", League of Legends).

LoL features a symmetric map with three paths, or lanes. Lanes are known as "top", "mid", and "bot". At the beginning of each game, both of the two teams will assign one player to top lane, one to mid lane, and two players (ad carry and support) to bot lane. There are also a "jungle" area between each lane, and one player is responsible for this area. Typically, players will be assigned a certain role (“top”, “mid”, “adc”, “support”, and “jungler”) and choose the corresponding lane to go to. Each player in both teams will select a unique “champion”, which is the in-game characters created by Riot. Every week there will be a rotation of 15 champions offered to all players, but champions can also be purchased using points earned through games or real-world money. Champions are categorized by roles: assassin, fighter, mage, marksman, support, and tank. Usually, particular role of champion will be selected for each lane. For example, mage champions will be selected by mid laners, while tank or fighter will be selected by top laners.

The goal of LoL game is to penetrate and destroy the enemy team’s central base, which is called “Nexus” in LoL. In order to accomplish this goal, players need to destroy enemy towers (which are called “turret” in LoL) in at least one lane first. Ally turrets will help to protect players and the team base, while enemy turrets will attack the player under certain circumstances. Gold can be earned once an enemy turret is destroyed. The gold earned will be used to purchase items, which increase a champion’s stats (such as attack damage or attack speed) in order to make them stronger.

There are three “inhibitors” located in each team’s base, and every inhibitor is corresponded to one of the three lanes. Inhibitors can help to prevent enemy Super Minions from spawning in that specific lane, so an enemy inhibitor being destroyed will give the team a huge advantage over opponents. Minions (and Super Minions) are small warriors that automatically march along a set path towards the enemy base in order to assist the team destroy the enemy Nexus. Gold can also be earned by killing enemy minions. Two things are placed in the jungle area: jungle monsters and plants. Jungler, as well as any other players, can earn gold and experience by killing the jungle monsters. Plants grow in specific area in the Rift, and they can offer bonuses to players during the game.

Dragon, Rift Herald, and Baron Nashor are other key factors that potentially have significant impact on the game result besides turrets and inhibitors. The legendary Elder Dragon, along with four Element Dragons, will provide bonus movement speed, increased overall damage, increased damage to structures and epic monsters, or additional health and mana every few seconds if not in combat. If a team kill and pick up the Rift Herald, it will relentlessly push its way to the enemy's Nexus unless it is taken down by the enemy team. Baron Nashor has always been considered as the most powerful resource on the map. The Baron buff includes boosting the team's stats, granting a faster recall to team's base, doing more damage to the enemy, and receiving less damage from opponents.

2.5 LoL eSports

League of Legends (LoL) has a widespread professional tournaments began in 2011 published by Riot Games. Professional gamers will form into a team and compete against each other as a team of 5 people. Until June 2016, LoL has had about 30 million USD in prize money, 4083 players, and 1718 tournaments, compared to DOTA 2's 64 million USD in prize money, 1495 players, and 613 tournaments (Fossett, 2016). Riot Games organized the League Championship Series (LCS) in both North America and Europe, which are located in Los Angeles and Berlin respectively. Each LCS tournament contains 10 professional teams. There are also similar regional competitions in China (LPL), South Korea (LCK), Taiwan (LMS), and other regions. World Championship tournament, which involves top seeds from each regional competition, will be held annually.

In this study, the North American League of Legends Championship Series (NA LCS) will be chosen as the evaluated target. NA LCS is the preeminent League of Legends eSports league in North American that stretches across two split, Spring and Summer. Every year there will be 10 teams participating in NA LCS, and each team will face every other team twice over the course of the season, for a total of 18 matches each. Until Summer 2017, matches are played in a Best of 3 format (the rule changes for upcoming season, according to Riot Games), and each team will only play once per day. Each split (Spring and Summer) will be divided into regular matches and playoffs. NA LCS playoffs are comprised of 6 teams in a seeded, single elimination tournament. All matches are in a Best of 5 format. There will also be a third/fourth place match ("About NA LCS". LoL eSports).

3. Methodology

LoL eSports will be chosen as a representative sample for the eSports predictive analytics in this paper. Riot, which is the developer for League of Legends (LoL), has open source data for LoL champions and matches. These data resources are credible and accurate since they are retrieved from Riot's official website. The Static Data API provided by Riot Developer draws its information directly from Data Dragon, which is a web service that centralizes LoL game data. By using the Static Data API, information can be pulled from Data Dragon directly. Data source from Amazon Web Services (AWS) will also be retrieved and used in this paper. Riot makes data on LoL match games available for the public to discover.

Data analysis will be divided into two parts. The first stage is to analyze statistics from the game process, which include champion selection and game properties. Two datasets including 200 LoL matched games will be analyzed using a cross-validation method. Examples of the attributes include but not limit to: first kill, first turret destroyed, and number of dragon killed. The second stage is to evaluate LoL eSports statistics, which include previous performance for each team and individual player. To evaluate the teams' previous performance, data for LoL eSports game results from 2015 to 2017 will be collected. Three datasets are included: team rosters, gold values and death information.

Player's performance will be analyzed based on individual statistics such as Kill/Death/Assist (KDA) ratio, damage per minute, gold earned per minute, and kill participation rate.

Given a dataset with huge amount of data, it is important to figure out which variables and criteria are relevant before finalizing appropriate analytics models. Different statistic tools will be used in this paper. Each individual category in the match game dataset will be evaluated in the first stage using regression models on R and WEKA. The results will provide a general understanding on which categories have more impact on the game results than others. After the first stage, data analytics will be constructed around categories correlated to the dependent variable, which is the game result. The data analytics modeling should be able to answer specific questions on LoL eSports game prediction, with examples as follows:

- Which criteria from the game process (first tower kill, total minions killed, total damage deal, etc.) is mostly related to the game results?
- Will Champion A always have advantages over Champion B? If so, which factor (health, attach speed, attach damage, ability, etc.) is the most important one?
- Is there any combination of champions (especially for bot lane) that always have a higher win rate than others in eSports games? How does it contribute to prediction of the game?

The data analysis models should be able to help with predicting LoL eSports results. By using the predictive models, teams could set up specific strategies against different opponents based on players' previous performance and champion selection. It is also easier for broadcasts and analysts to do analysis before the game starts given the champion selection.

3.1 Dataset

Static Data API provided by Riot Developer was used to collection information from League of Legends official website. Data of over 1000 matches were generated. After basic data cleanse and pre-processing, all the data were divided into two separate datasets with different categories.

The first dataset includes 200 matches from NA LCS, each row in the dataset represents a single game identified by a unique game ID with result and multiple properties during the game process. Unlike number of kills or gold earned, these in-game factors, such as first tower or first dragon, are objective, which means they are not supposed to be highly depended on the player's performance or champion selection. The dataset will be used to test and analyze how in-game factors influence the game results.

The second dataset includes 855 previous NA LCS matches. Each row in the dataset represents a single game with result, players in each team, and champion selected for each position. Statistics on each player's performance, for example, KDA ratio, kill participation, and percentage of damage, were also added into this dataset. The dataset will be used to analyze how champion selection and player's performance affect the game results.

3.2 Features

Three types of features will be analyzed in this study, which are described as follows:

In-Game Factors: objective properties during the game process.

- First blood: if a team has the first kill. Boolean value.
- First tower/dragon/Baron/inhibitor/Rift Herald: if a team gets the first tower, dragon, Baron, inhibitor, or Rift Herald. Boolean value.
- Tower/Dragon/Baron kills: total number of towers, dragons, or Barons a team kills during the game. Numeric value.

Champion Selection: a total of 134 champions are categorized into five positions, which are top, jungle, mid, bot, and support. For each position, champions may have advantages over each other based on its attack range, attack speed, etc.

Player Performance: statistics related to the player's performance. Usually represented as a total number or average number during a certain time period (one split or one season).

- KDA: kill/death/assist ratio.
- GD10: average gold earned difference at 10 minute.
- CSD10: average creep score difference at 10 minute.
- XPD10: average experience difference at 10 minute.

4. Models

4.1 Logistic Regression

Logistic regression is a statistical class probability estimation model that analyzes a dataset in which one or more independent variables are used to determine a categorical outcome by a cumulative logistic distribution. The outcome is usually coded as “0” or “1”, which represent “win” vs “lose” in this case study. Regression coefficients, which represent the change in the logit for each unit change in the predictor, will be examined in order to understand the contribution of individual predictors.

Likelihood ratio test and the Wald statistic are most common designed tests that are used to assess the significance of an individual predictor. A logistic regression is expected to provide a better fit to the data if it demonstrates an improvement over a model with fewer predictors. Likelihood ratio test compares the likelihood of data under the full model to the one with fewer predictors, and then analyzes if the observed difference is statistically significant. Wald test is used to evaluate the statistical significance of each coefficient in the model. It is calculated by the ratio of the square of the regression coefficient to the square of the standard error of the coefficient.

To enhance the prediction accuracy and interpretability, ridge and lasso regression analysis can be applied to shrink the coefficients that contribute most to the error. Both ridge and lasso regularization work by adding a penalty term to the log likelihood function, while the penalty term is β_1^2 in ridge regression and $|\beta_1|$ in case of lasso. Thus, the quantity to be minimized in the two cases are $L + \lambda \sum \beta_1^2 \ldots.$ for ridge regression and $L + \lambda \sum |\beta_1| \ldots.$ for lasso regression, where λ is a free parameter that is selected to minimize the out of sample error for the model.

Logistic regression models are implemented as glm() function in R. It lists the coefficient estimates, standard error of the estimates, z-statistics, and p-value as the output. Based on p-value, predictors can be treated as significant, have little predictive power, or even contribute to overfitting. Another method in R to test how variables affect the outcome is variable importance. Variable importance is an indication of which predictors are most useful for predicting the target variable, which is represented by the function varImp() in R. The method takes an object and calculate variable importance for regression models.

4.2 Decision Tree

Decision tree is a non-parametric supervised learning method for classification and regression. The model is used to predict the value of a target variable and represent the decision making visually and explicitly. Performance of a decision tree can be improved by pruning, which means to remove the branches that make use of features having low importance. Pruning helps to reduce the complexity of the tree, so that it increases the predictive power by reducing overfitting.

Two important concepts related to decision tree are entropy and information gain. Entropy is the degree of randomness of elements that measures the unpredictability and diversity of information content. Entropy can be calculated as $\sum_i - p_i \log_2 p_i$ with p_i being the proportion of class i in the data. The information gain is based on the decrease in entropy after the dataset is split on an attribute. The decision tree is constructed by finding the attribute that returns the highest information gain. In general terms, information gain is defined as follows:

$$IG(T, a) = H(T) - H(T|a)$$

Advantages of decision tree models include transparency and specificity. The decision tree explicates all the possible alternatives and traces each alternative to the corresponding conclusion, so that it allows easy comparison among the various alternatives. Another ability of decision tree is to assign specific values to problem, decisions, and outcomes of each decision. It reduces ambiguity in decision-making because every possible scenario from a decision is represented by a clear node, with all possible solutions viewed clearly in a single view.

5. Experiments

5.1 In-Game Factors

Considering most in-game factors are numerical or categorical dependent variables, logistic regression model is used to generate and analyze the result in this situation. Since the dependent variable, which is the game result, only has two outcomes: win or lose, the study will construct a binomial logistic regression model to analyze how different independent variables during the game process, such as first tower or first dragon, have impact on the game results.

The dataset used to test in-game factors gathers information from 200 previous NA LCS matches. Given each game has two teams participated, 400 rows of data in total were analyzed in this study. The results of games were recorded along with nine different game process variables: first blood (which team has the first kill), first tower, first inhibitor, first Baron, first Rift Herald (the neutral monster that helps team to push towers), total number of tower kills, dragon kills, and baron kills. The following binomial logistic regression model was generated by R, and results were shown below.

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-10.65654	1.52654	-6.981	2.93e-12	***
teams.0.firstBloodTRUE	1.24506	0.52463	2.373	0.0176	*
teams.0.firstTowerTRUE	-1.11152	0.61438	-1.809	0.0704	.
teams.0.firstInhibitorTRUE	0.89857	0.56401	1.593	0.1111	
teams.0.firstBaronTRUE	1.58307	0.85803	1.845	0.0650	.
teams.0.firstDragonTRUE	0.33219	0.61322	0.542	0.5880	
teams.0.firstRiftHeraldTRUE	-0.18412	0.54350	-0.339	0.7348	
teams.0.towerKills	1.36617	0.22209	6.151	7.68e-10	***
teams.0.dragonKills	0.09436	0.27334	0.345	0.7299	
teams.0.baronKills	-0.46039	0.74256	-0.620	0.5353	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figure 1. Logistic Regression on multiple in-game factors

	Overall
teams.0.firstBloodTRUE	2.3732324
teams.0.towerKills	6.1514149
teams.0.firstBaronTRUE	1.8450086
teams.0.firstTowerTRUE	1.8091820
teams.0.firstInhibitorTRUE	1.5931825
teams.0.firstDragonTRUE	0.5417218
teams.0.firstRiftHeraldTRUE	0.3387713
teams.0.dragonKills	0.3452117
teams.0.baronKills	0.6200054

Figure 2. Variable Importance

The results showed that four independent variables, total number of tower kills, first blood, first Baron, and first tower, had small p-values. As both firstBloodTRUE and towerKills are less than 0.05, these two variables are significant in the logistic regression model. In particular, tower kills had an extremely small p-value, which indicates that this independent variable does have significant influence on the outcome, or does help to predict the result. Variables like firstTowerTRUE and firstBaronTRUE also have little predictive power. In other words, if a team has more tower kills than the enemy team, gets the first blood, first tower, and first Baron, this team will have a higher chance to win the game. The variable importance tested by R also indicates how tower kills and first blood have more influence on the game results over other variables.

```
J48 pruned tree
------------------

teams/0/towerKills <= 6: Fail (182.0/3.0)
teams/0/towerKills > 6
|   teams/0/towerKills <= 8
|   |   teams/0/firstBaron = FALSE
|   |   |   teams/0/firstBlood = TRUE: Win (17.0/6.0)
|   |   |   teams/0/firstBlood = FALSE
|   |   |   |   teams/0/firstInhibitor = FALSE: Fail (5.0)
|   |   |   |   teams/0/firstInhibitor = TRUE: Win (3.0/1.0)
|   |   teams/0/firstBaron = TRUE: Win (42.0/6.0)
|   teams/0/towerKills > 8: Win (151.0/3.0)
```

Figure 3. J48 Decision Tree

A J48 decision tree constructed using WEKA further proved this hypothesis. From results of the J48 tree shown below, it is obvious that number of tower kills has a significant influence on determining whether a team wins or not, while getting the first Baron and first blood are also considered as categories that have some level of influence on the game results.

5.2 Champion Selection

Two steps of analysis were constructed for champion selection. First step was to figure out if a champion selected would lead to a specific champion selected for the same position from another team. Five bar plots were generated by R graphics, as the results shown below.

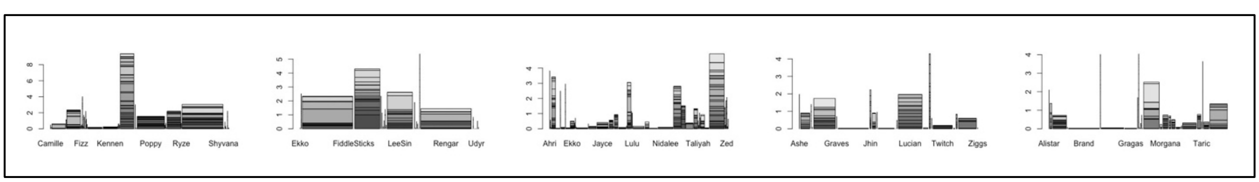

Figure 4. Bar plots on champion selection for five positions

The x-axis in above features indicates the pick by blue-side teams who pick the champion first in order, and the y-axis represents the pick by red-side teams that pick the champion after blue-side teams. From the results, it is clear and obvious that certain champions have been teams' favorite picks while some champions barely being picked. Also, once the blue-side team picks one champion, the red-side team will most likely narrow down their choice for the same position, which proves that hypothesis that champions are not randomly selected by team. In other words, teams may have specific strategy on using a combination of champions, or select certain champion against the opponent team.

The next step was to test the dataset by decision tree model to see if any combination of champion selection, or selection one champion against another at the same position could lead to a team win. Every combination of champions from different positions on the team or from the same position but different team were evaluated separately by constructing a decision tree. Given the size of the decision tree, only one example from each situation is shown below.

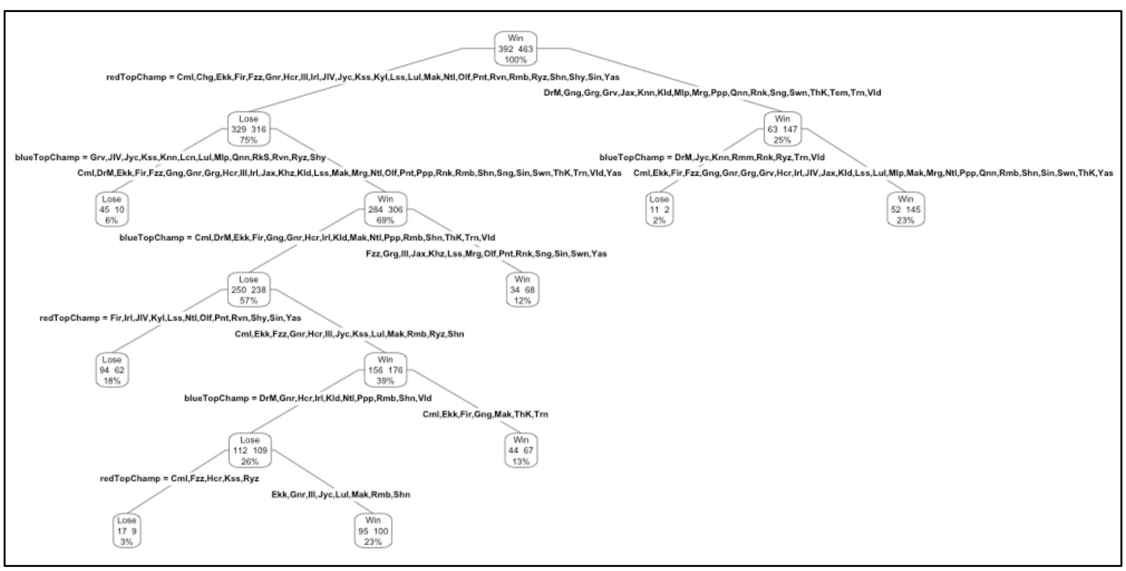

Figure 5. Decision Tree: Selection on Top for Team 1 and Team 2

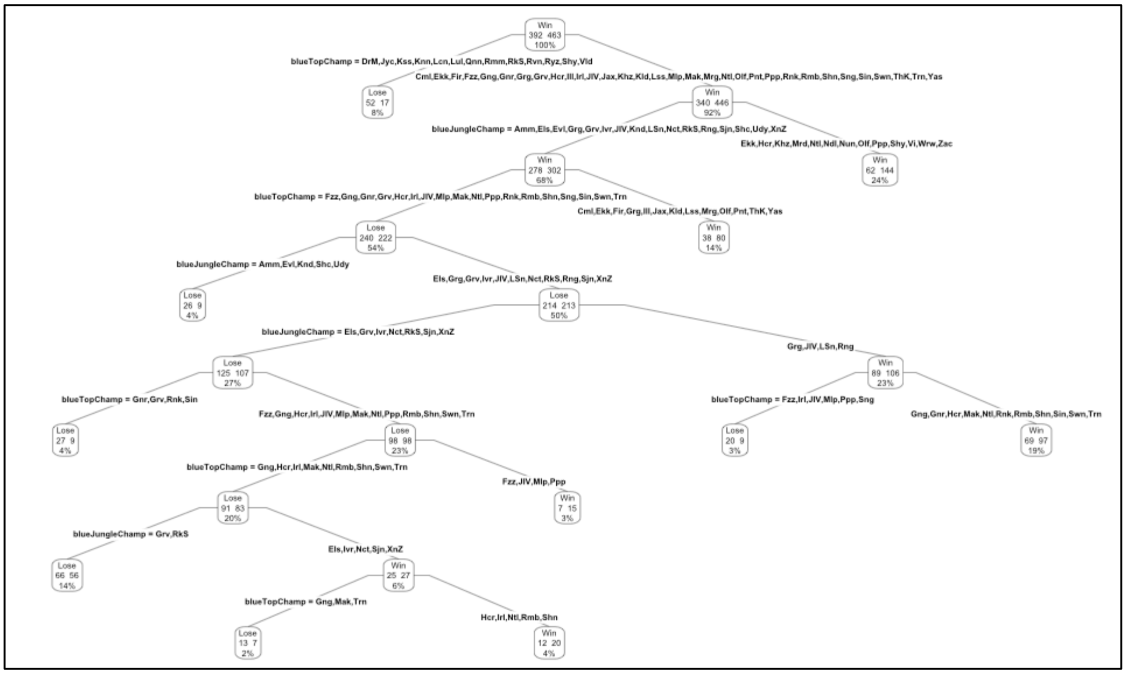

Figure 6. Decision Tree: Selection on Top and Jungle for Team 1

The first example proves the hypothesis that champion A may have advantages over B at the same position. The second example of the decision tree models proves that certain combination of champion selection will lead to the team's victory. That says, if a team selects champion A on the top, and then select a corresponding champion B on the jungle, the probability of winning the game will be increased. However, the accuracy rates for all the decision tree models built are around 55%. Even though the prediction model is not completely random and is able to provide insights on predicting the game results at some level, it is not supposed to be a strongly proof in this situation.

5.3 Player Performance

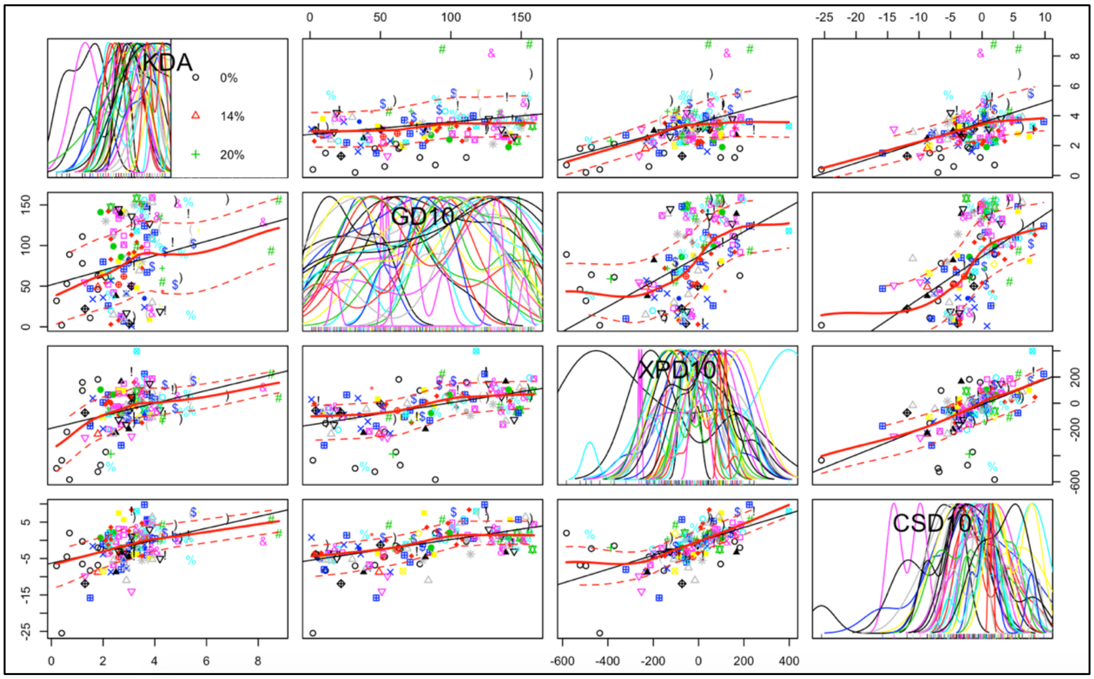

Figure 7. Scatterplot Matrices: player's statistics and win rate

Scatterplot matrices were generated to determine the correlation between player's performance, which is represented by multiple variables, and the win rate. Four typical categories of players' statistics were tested in this case study: KDA ratio, GD10, XPD10, and CSD10, with each of them recorded as a numerical value. The win rate for each player is displayed as a percentage, which is calculated as the number of winning games divided by the total number of games played. Results from the scatterplot matrices constructed by R are shown above. The result pinpoints that these four variables have similar correlations to the target outcome – win rate. The higher a player's KDA ratio is, the larger difference on gold earned, experience, and creep score a player has over the opponents, the higher change the team will win the game.

6. Discussion

Similar to traditional sports prediction, predictive analytics on eSports closely related to player's performance and in-game factors. Statistical models like logistic regression and decision tree are also perform well on eSports prediction. Most features that are statistically proved to have a positive impact on the game results can be explained with circumstances in a real LoL game. The accuracy rate for each tested model was over 50 percentages, which means the results of built model are not completely random and it can be used to predict eSports games.

However, even though the prediction model in this study has decent performance, LoL eSports game results are still difficult to predict given all the statistical features. In this study, the three types of features being analyzed are assumed to have no correlation with each other, and variables under each feature are assumed to be independent. In reality, all the variables from three types of features have some level of correlation, and they are mutable as time passes by. For example, a player's performance is not only based on his own play, but also affected by his teammates' performance and the team's overall play style. He may have better performance if his teammates are outstanding, or his synergy with a particular teammate is extremely great. As for champion selections, Riot updates the game frequently, with each new version of the game including plenty of changes on champion's properties. The player's choice on champion is possibly determined by multiple factors that are difficult to analyze as statistical data: the player's proficiency on the champion, whether the champion is playable under a specific version or not, how well the champion coordinates with other ally champions, etc. Comparing with traditional sports prediction, eSports prediction is more complicated given the complexity of a huge amount of dependent variables. Thus, a more detailed analysis with different models should be constructed in order to understand the eSports prediction.

7. Conclusion

In this study, both logistic regression and decision tree models are considered efficient in predicting LoL eSports game results based on information on in-game factors and player's performance. However, a more complex statistical model needs to be implemented in order to predict the results using the champion selection information. Similarity and difference between traditional sports prediction and eSports prediction can be discussed after this study. In general, eSports prediction is similar to the prediction on traditional sports, since both of them have a binary dependent target value (game result can only be whether win or lose) and are rely on features related to players involved and the game itself, with the player's performance has a huge influence on the game result. The key factor that distinguishes eSports prediction from traditional sports and makes it more complex is another independent variable besides players and games: champions, which is a unique property of eSports games. Unlike traditional sports, the format of eSports is not people playing against each other directly. Instead, players will manipulate a game character (Champion in League of Legends) to go against the opponent. Champions with different properties play significant roles in eSports prediction. They are separated from players and game processes, and will not be affected by any of these two features. To construct the predictive analytics on eSports appropriately, the combination of players, champions, and in-game factors should be considered, which makes eSports prediction more complicated than traditional sports.

Several problems have raised during the model building and analyzing process. Data temporality was one of the potential issues with the databases used in this study. All the data are retrieved from previous matches in NA LCS, which means the value of target variable (result of each game) is already known in this case. Since it is unable to test the built models with a reasonable amount of new data being the “unknown data”, a set of previous matches was randomly selected to test the models. Furthermore, cross-validation method was used to estimate the accuracy rate of a predictive model’s performance, which would be effective in limiting problems like overfitting. Another difficulty faced during the experiments was the selection among a large amount of features. Given the situation that not all the features in the databases have impact on the eSports game results and it could potentially have negative influence on the model’s performance, only part of the features is expected to be used in this study. It is important to avoid the situation that only the “most” contributed features are selected for the model after comparing the correlation, which may lead to bias on performance estimates. In this case, all the selected and then tested features in this study were based on previous discussions on eSports prediction and general knowledge on League of Legends.

From the experiment process and results from the model, advantages and disadvantage on eSports predictive analytics can be discussed. Comparing with traditional sports, a huge advantage of eSports prediction is that all the features can be generated automatically and accurately during the game. For example, the League of Legend case studied in this paper, is completely electronic and based on computer. That says, all the data generated and retrieved using procedures built by programming languages. However, many data related to traditional sports are recorded by people. For example, referees are the people that decide if a player touch the ball in a basketball game. In this way, data used in eSports prediction is expected to be more accurate, which leads to a potentially better performance on the prediction model. In the other hand, eSports has its own disadvantage because sometimes statistics don't correctly represent the fact. For example, in LoL eSports, players in the rank 1 team always tend to have better statistics than the ones on the team with a lower rank. Even though statistics will tell these players have better performance, it doesn't truly reflect the player's actual ability. It is important to set up the standards on feature selection (for example, if one category from player's statistics highly depends on team's performance, it should not be included) when building eSports prediction models.

Studying in SILS has helped a lot in case of the understanding of information science. Knowledge gained from SILS courses allows student to discover principles and impacts of information, and advances information access, use, and management to solve real-world problems and improve the quality of life for people. Inspiration and motivation behind this study was especially from the course of Information Analytics, which introduces analytical techniques to deal with large data sets. Several predictive models, such as logistic regression, linear regression, support vector machines, and maximum likelihood estimation, were learned from this class. The class also compared the similarity and difference between these models, and defined their advantages and disadvantages. Based on what has been learned from this class, logistic regression and decision tree models were chosen for this study because of their properties.

For future study and research, the correlation between selected champions for each team and each position is expected to be reviewed and take into account the use of parameters in the predictive analytics. That says, if two features have a strongly correlation with each other, they should be tested as a whole when analyzing the prediction model. Also, it is important to consider the level of influence on game results from different types of features when building prediction model on eSports games. One potential approach is to use weight variables. Given different features may have different levels of impact on the game results, each feature can be assigned with a particular weight. Features with relatively large weights are supposed to have more influence in the analysis than the ones have smaller weight.

8. Bibliography

"About NA LCS". LoL eSports. http://www.loleSports.com/en_US/na-lcs/na_2017_summer/about

Aryanata, G. A., Rahadi, P. S. A. D., & Sudarmojo, Y. S. (2017). Prediction of DOTA 2 Match Result by Using Analytical Hierarchy Process Method. *International Journal of Engineering and Emerging Technology, Vol. 2, No. 1,* 22-25.

Barry, D. & Hartigan, J. A. (1993). Choice models for predicting divisional winners in major league baseball. *Journal of the American Statistical Association, 88*. 766-774.

Buchdal, J. (2003). Rating Systems For Fixed Odds Football Match Prediction. *Fixed Odds Sports Betting: The Essential Guide.*

Buursma, D. (2010). Predicting sports events from past results. In *14th Twente Student Conference on IT*.

Cao, C. (2012). Sports data mining technology used in basketball outcome prediction.

Chen, S., & Joachims, T. (2016). Predicting Matchups and Preferences in Context. In *KDD*. 775-784.

Constantinou, A. C., Fenton, N. E., & Neil, M. (2012). pi-football: A Bayesian network model for forecasting Association Football match outcomes. *Knowledge-Based Systems*, *36*, 322-339.

Fossett, Wyatt. (2016). "How Videogames Became a Sport, and Why They're Here to Stay (Hint: Money!)". *Techvibes.*

"Game Info". League of Legends. http://gameinfo.na.leagueoflegends.com/en/game-info/

Griffiths, M. D., Davies, M. N. O., & Chappell, D. (2003). Breaking the stereotype: the case of online gaming. *Cyberpsychology & Behavior, Vol. 6, Iss. 1.* 81-91.

Haghighat, M., Rastegari, H., & Nourafza, N. (2013). A review of data mining techniques for result prediction in sports. *Advances in Computer Science: an International Journal*, *2*(5), 7-12.

Hamari, J. j., & Sjöblom, M. m. (2017). What is eSports and why do people watch it?. *Internet Research*, *27*(2), 211-232.

Hattenstone, Simon. (2017). "The rise of eSports: are addiction and corruption the price of its success?". *The Guardian.*

Hewitt, E. (2014). Will eSports ever become widely accepted as official sports and how will they affect the way we entertain ourselves if they do?. *Computers for everyone*, 81-83.

Jenny, S. E., Manning, R. D., Keiper, M. C., & Olrich, T. W. (2017). Virtual (ly) Athletes: Where eSports Fit Within the Definition of "Sport". *Quest*, *69*(1), 1-18.

Kaigh, W. D. (1995). Forecasting baseball games. *Chance 8(2).* 33-37.

Lee, D., & Scho

enstedt, L. J. (2011). Comparison of eSports and traditional sports consumption motives. *The ICHPER-SD Journal of Research in Health, Physical Education, Recreation, Sport & Dance, 6*(2), 39-44.

Michaluk, T. (2012). Changes in the meaning of physicality in modern sport - from disabled sport to e-sport. *Physiotherapy / Fizjoterapia. 20(1*), 64-70.

Min, B., Kim, J., Choe, C., Eom, H., & McKay, R. B. (2008). A compound framework for sports results prediction: A football case study. *Knowledge-Based Systems*, *21*(7), 551-562.

Ong, H. Y., Deolalikar, S., & Peng, M. (2015). Player Behavior and Optimal Team Composition for Online Multiplayer Games. *arXiv preprint arXiv:1503.02230*.

Pretorius, A., & Parry, D. A. (2016). Human Decision Making and Artificial Intelligence: A Comparison in the Domain of Sports Prediction. In *Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists* (p. 32). ACM.

Schubert, M., Drachen, A., & Mahlmann, T. (2016). ESports analytics through encounter detection. In Sloan, M., ed., *Proceedings of the MIT Sloan Sports Analytics Conference*.

Segal, David. (2014). “Behind League of Legends, E-Sports’s Main Attraction”. *NY Times.*

Shim, K. J., Sharan, R., & Srivastava, J. (2010). Player performance prediction in massively multiplayer online role-playing games (MMORPGs). *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. 71-80.

Spann, M., & Skiera, B. (2009). Sports forecasting: a comparison of the forecast accuracy of prediction markets, betting odds and tipsters. Journal of Forecasting, 28(1), 55-72.

Summerville, A., Cook, M., & Steenhuisen, B. (2016). Draft-Analysis of the Ancients: Predicting Draft Picks in DotA 2 Using Machine Learning. In *Twelfth Artificial Intelligence and Interactive Digital Entertainment Conference*.

Tamassia, M., Raffe, W., Sifa, R., Drachen, A., Zambetta, F., & Hitchens, M. (2016). Predicting player churn in destiny: A Hidden Markov models approach to predicting player departure in a major online game. In *Computational Intelligence and Games (CIG), 2016 IEEE Conference.* 1-8.

Wagner, M. (2006). On the scientific relevance of eSport. *Proceedings of the 2006 International Conference on Internet Computing and Conference on Computer Game Development, CSREA Press, Las Vegas, NV.* 437-440.

Warman, Peter. (2017). “ESPORTS REVENUES WILL REACH $696 MILLION THIS YEAR AND GROW TO $1.5 BILLION BY 2020 AS BRAND INVESTMENT DOUBLES”. *NEWZOO.*

Warr, Philippa. (2014). "eSports in numbers: Five mind-blowing stats". *Red Bull.*

Wen, Y., Hung, K., Hwang, Y., & Lin, Y. F. (2016). Sports lottery game prediction system development and evaluation on social networks. *Internet Research, 26*(3), 758-788.

Witkowski, E. (2012). On the digital playing field: how we ‘do sport’ with networked computer games. *Games and Culture*, Vol. 7 No. 5. 349-374.

Yang, P., & Roberts, D. L. (2013). Knowledge discovery for characterizing team success or failure in ARTS games. In *Computational Intelligence in Games (CIG), IEEE Conference on*.

Yang, T. Y., & Swartz, T. (2004). A two-stage Bayesian model for predicting winners in major league baseball. *Journal of Data Science*, *2*(1), 61-73.

Yang, Y., Qin, T., & Lei, Y. H. (2016). Real-time eSports Match Result Prediction. *arXiv preprint arXiv:1701.03162*.

Yue, Y., Lucey, P., Carr, P., Bialkowski, A., & Matthews, I. (2014). Learning fine-grained spatial models for dynamic sports play prediction. *In Data Mining (ICDM), 2014 IEEE International Conference on IEEE.* 670-679.

Appendix A: Data Dictionary

Field Name	Data Type	Example	Description
gameId	Number	2585564750	a unique ID for each game
platformId	Text	NA1	a unique ID for the platform (in which region the game is played)
gameDuration	Number	2038	the time duration for a game
seasonId	Number	7	in which season the game is played
teams/0/teamId	Number	100	a unique ID for the team
teams/0/win	Text	Win	indicate if team 0 wins or loses
teams/0/firstBlood	Boolean	TRUE	if team 0 gets first blood
teams/0/firstTower	Boolean	TRUE	if team 0 gets first tower
teams/0/firstInhibitor	Boolean	TRUE	if team 0 gets first inhibitor
teams/0/firstBaron	Boolean	TRUE	if team 0 gets first baron
teams/0/firstDragon	Boolean	TRUE	if team 0 gets first dragon
teams/0/firstRiftHerald	Boolean	TRUE	if team 0 gets the rift herald
teams/0/towerKills	Number	4	the total number of towers team 0 kills
teams/0/inhibitorKills	Number	2	the total number of inhibitors team 0 kills
teams/0/baronKills	Number	1	the total number of barons team 0 kills
teams/0/dragonKills	Number	4	the total number of dragons team 0 kills
teams/0/riftHeraldKills	Number	0	the total number of rift herald team 0 kills
teams/0/dominionVictoryScore	Number	0	the score of dominion victory for team 0
teams/1/teamId	Number	100	a unique ID for team 1
teams/1/win	Text	Fail	indicate if team 1 wins or loses
teams/1/firstBlood	Boolean	FALSE	if team 1 gets first blood
teams/1/firstTower	Boolean	TRUE	if team 1 gets first tower
teams/1/firstInhibitor	Boolean	TRUE	if team 1 gets first inhibitor
teams/1/firstBaron	Boolean	TRUE	if team 1 gets first baron
teams/1/firstDragon	Boolean	TRUE	if team 1 gets first dragon
teams/1/firstRiftHerald	Boolean	TRUE	if team 1 gets the rift herald
teams/1/towerKills	Number	4	the total number of towers team 1 kills
teams/1/inhibitorKills	Number	2	the total number of inhibitors team 1 kills
teams/1/baronKills	Number	1	the total number of barons team 1 kills
teams/1/dragonKills	Number	4	the total number of dragons team 1 kills
teams/1/riftHeraldKills	Number	0	the total number of rift herald team 1 kills
teams/1/dominionVictoryScore	Number	0	the score of dominion victory for team 1
bResult	Text	Win	the result for team on the blue side
rResult	Text	Lose	the result for team on the red side
blueTopChamp	Text	Irelia	Champion selection for blue team's top
blueJungleChamp	Text	Rengar	Champion selection for blue team's jungle
blueMiddleChamp	Text	Ahri	Champion selection for blue team's mid
blueADCChamp	Text	Jinx	Champion selection for blue team's adc

blueSupportChamp	Text	Janna	Champion selection for blue team's support
redTopChamp	Text	Gnar	Champion selection for red team's top
redJungleChamp	Text	Jax	Champion selection for red team's jungle
redMiddleChamp	Text	LeBlanc	Champion selection for red team's mid
redADCChamp	Text	Caitlyn	Champion selection for red team's adc
redSupportChamp	Text	Thresh	Champion selection for red team's support
PLAYER	Text	Hai	Player's name
TEAM	Text	Golden Guardians	Player's team
POS	Text	Middle	Player's position
GP	Number	18	Number of games played by the player
W%	Number	22%	Win rate (winning game/total number of games)
K	Number	34	Total number of kills by the player
D	Number	52	Total number of deaths by the player
A	Number	87	Total number of assists by the player
KDA	Number	2.3	Kill/Death/Assist ratio
KP	Number	77.60%	Kill participation
DTH%	Number	23.70%	Death percentage (player's death/team total deaths)
FB%	Number	17%	First blood rate (how frequently the player gets first blood)
GD10	Number	-320	Gold difference at 10 minute
XPD10	Number	-230	Experience difference at 10 minute
CSD10	Number	-8.7	Creep score difference at 10 minute
CSPM	Number	9	Creep score per minute
DPM	Number	539	Damage dealt per minute
DMG%	Number	28.60%	Damage percentage (player's damage/team's total damage)
GOLD%	Number	23.50%	Gold percentage (player's earned gold/team's total gold)
WPM	Number	0.58	Ward placed per minute
WCPM	Number	0.26	Ward cleared per minute

Appendix B: Databases

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	PLAYER	TEAM	POS	GP	W%	K	D	A	KDA	KP	DTH%	FB%	GD10	XPD10	CSD10	CSPM	DPM	DMG%	GOLD%
2	JayJ	FlyQuest	Support	2	0%	1	6	6	1.2	70.00%	26.10%	100%	-7	97	-6.5	1	222	13.00%	9.5
3	Shrimp	FlyQuest	Jungle	2	0%	1	10	5	0.6	60.00%	30.30%	0%	-307	-219	-4.5	4.5	150	8.10%	14.8
4	Contractz	Golden Guar	Jungle	18	22%	27	56	86	2	72.40%	25.60%	28%	-211	-195	-3.8	4.9	243	12.50%	16.1
5	Deftly	Golden Guar	ADC	18	22%	49	36	64	3.1	72.40%	16.40%	17%	-1	29	0.8	10	576	28.60%	26.7
6	Hai	Golden Guar	Middle	18	22%	34	52	87	2.3	77.60%	23.70%	17%	-320	-230	-8.7	9	539	28.60%	23.5
7	Lourlo	Golden Guar	Top	18	22%	34	29	69	3.6	66.00%	13.20%	6%	-157	-88	-1.2	9	407	22.60%	23.7
8	Matt	Golden Guar	Support	18	22%	12	46	110	2.7	78.20%	21.00%	22%	-13	-92	-3.7	1.6	135	7.80%	10.0
9	zig	OpTic Gamin	Top	12	25%	9	35	37	1.3	43.80%	26.30%	25%	-318	-263	-9.8	6.9	302	18.00%	18.3
10	Akaadian	OpTic Gamin	Jungle	18	28%	21	33	84	3.2	68.20%	16.70%	50%	75	81	0.7	4.9	179	10.50%	16.5
11	Arrow	OpTic Gamin	ADC	18	28%	44	31	66	3.5	71.40%	15.70%	33%	-79	15	-5.5	9.3	509	28.80%	25.9
12	LemonNatio	OpTic Gamin	Support	18	28%	8	49	93	2.1	65.60%	24.70%	28%	41	-7	-0.3	1.6	147	8.90%	10.7
13	PowerOfEvil	OpTic Gamin	Middle	18	28%	67	32	58	3.9	81.20%	16.20%	44%	21	-93	-3.8	10.1	571	33.20%	27.3
14	Damonte	Echo Fox	Middle	3	33%	8	6	14	3.7	71.00%	16.70%	33%	73	-74	5.7	9	435	20.00%	24.6
15	Dhokla	OpTic Gamin	Top	6	33%	5	18	22	1.5	55.10%	27.70%	17%	-270	-174	-15.8	8.3	310	19.90%	22.1
16	Flame	FlyQuest	Top	18	33%	33	31	81	3.7	73.10%	14.30%	22%	177	66	5.3	9	464	24.20%	23.8
17	Fly	FlyQuest	Middle	12	33%	22	29	60	2.8	75.90%	22.50%	33%	-107	-70	-4.8	8.4	537	26.90%	21.4
18	Keane	FlyQuest	Middle	6	33%	14	20	25	2	81.30%	22.70%	0%	-79	6	-1	9.3	495	27.20%	23.0
19	PapaChau	Echo Fox	Support	3	33%	2	9	22	2.7	77.40%	25.00%	33%	-115	-321	-8	1.1	216	9.30%	8.2
20	WILDTURTLE	FlyQuest	ADC	18	33%	58	37	64	3.3	78.20%	17.10%	17%	-84	-131	-3.3	10.4	595	30.80%	27.4
21	AnDa	FlyQuest	Jungle	16	38%	17	39	99	3	79.50%	21.20%	25%	-80	-197	-4.6	5.1	168	8.80%	16.0
22	Stunt	FlyQuest	Support	16	38%	10	45	110	2.7	82.20%	23.20%	13%	53	167	-3.3	1.8	157	8.80%	11.3
23	Biofrost	Counter Logi	Support	18	39%	9	45	138	3.3	78.20%	21.20%	22%	15	-28	-1.2	1.7	141	6.80%	10.7
24	Darshan	Counter Logi	Top	18	39%	37	42	88	3	66.50%	19.80%	44%	10	55	2.6	8.2	450	22.30%	22.8
25	huhi	Counter Logi	Middle	18	39%	51	42	98	3.5	79.30%	19.80%	39%	-52	-149	-4.3	9.3	634	29.60%	24.0
26	Reignover	Counter Logi	Jungle	18	39%	18	43	120	3.2	73.40%	20.30%	61%	105	159	1.7	4.9	193	9.40%	15.8
27	Stixxay	Counter Logi	ADC	18	39%	73	40	66	3.5	73.90%	18.90%	22%	119	19	4.7	10	701	31.80%	26.7
28	Apollo	Clutch Gamir	ADC	20	55%	45	31	84	4.2	75.40%	20.70%	0%	-76	-72	-0.5	9.8	484	26.70%	24.5

Statistics on Player's Performance

	A	B	C	D	E	F	G	H	I	J	K	L
1	bResult	rResult	blueTopChar	blueJungleCh	blueMiddleC	blueADCCha	blueSupport	redTopCham	redJungleCh	redMiddleCh	redADCCham	redSupportChamp
2	Win	Lose	Irelia	RekSai	Ahri	Jinx	Janna	Gnar	Elise	Fizz	Sivir	Thresh
3	Lose	Win	Gnar	Rengar	Ahri	Caitlyn	Leona	Irelia	JarvanIV	Azir	Corki	Annie
4	Win	Lose	Renekton	Rengar	Fizz	Sivir	Annie	Sion	LeeSin	Azir	Corki	Janna
5	Lose	Win	Irelia	JarvanIV	Leblanc	Sivir	Thresh	Gnar	Nunu	Lulu	KogMaw	Janna
6	Win	Lose	Gnar	JarvanIV	Lissandra	Tristana	Janna	Sion	RekSai	Lulu	Corki	Annie
7	Lose	Win	Kassadin	Rengar	Leblanc	Sivir	Annie	Gnar	JarvanIV	Lulu	Corki	Thresh
8	Win	Lose	Irelia	JarvanIV	Xerath	Corki	Janna	Renekton	LeeSin	Leblanc	Tristana	Nami
9	Win	Lose	Renekton	JarvanIV	Azir	Caitlyn	Annie	Rumble	Rengar	Leblanc	Sivir	Blitzcrank
10	Lose	Win	Sion	RekSai	Orianna	KogMaw	Janna	Kassadin	Vi	Zed	Corki	Morgana
11	Win	Lose	Irelia	Nocturne	Orianna	Sivir	Nami	Gnar	Rengar	Leblanc	Graves	Morgana
12	Win	Lose	Gnar	JarvanIV	Azir	Corki	Morgana	Lulu	LeeSin	Xerath	Ezreal	Janna
13	Win	Lose	Jax	JarvanIV	Kassadin	Graves	Morgana	Gnar	RekSai	Ahri	Corki	Thresh
14	Win	Lose	Irelia	Vi	Lissandra	Kalista	Annie	Gnar	JarvanIV	Ahri	Sivir	Leona
15	Lose	Win	Lulu	LeeSin	Xerath	Corki	Janna	Sion	JarvanIV	Cassiopeia	Kalista	Blitzcrank
16	Win	Lose	Maokai	JarvanIV	Leblanc	Kalista	Annie	Lulu	Rengar	Morgana	Tristana	Janna
17	Win	Lose	Rumble	Vi	Zed	Caitlyn	Morgana	Gnar	RekSai	Leblanc	Corki	Janna
18	Lose	Win	Gnar	JarvanIV	Azir	KogMaw	Nami	Maokai	LeeSin	Zed	Sivir	Janna
19	Win	Lose	Lissandra	JarvanIV	Xerath	Tristana	Morgana	Gnar	Nunu	Ahri	Caitlyn	Janna
20	Win	Lose	Gnar	LeeSin	Leblanc	KogMaw	Nami	Malphite	JarvanIV	Orianna	Graves	Janna
21	Lose	Win	Sion	Rengar	Ahri	Sivir	Janna	Lissandra	JarvanIV	Corki	Kalista	Thresh
22	Lose	Win	Lulu	RekSai	Corki	Kalista	Morgana	Rumble	LeeSin	Ahri	Graves	Annie
23	Win	Lose	Maokai	Vi	Kassadin	Graves	Annie	Irelia	JarvanIV	Corki	Sivir	Leona
24	Lose	Win	RekSai	Vi	Azir	Corki	Thresh	Maokai	LeeSin	Xerath	Sivir	Morgana
25	Win	Lose	Lissandra	RekSai	Xerath	Vayne	Nami	Sion	JarvanIV	Cassiopeia	Sivir	Morgana
26	Lose	Win	Gnar	JarvanIV	Kassadin	Sivir	Janna	Irelia	Vi	Nidalee	Jinx	Morgana
27	Win	Lose	Rumble	Vi	Leblanc	Graves	Morgana	Maokai	JarvanIV	Xerath	Lucian	Janna
28	Win	Lose	Lissandra	LeeSin	Ezreal	Sivir	Alistar	Irelia	Vi	Azir	Corki	Janna

Champion Selection

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	
1	gameId	platformId	gameDuratic	seasonId	gameVersior	teams/0/teamId	teams/0/win	teams/0/first	teams/0/first	teams/0/first	teams/0/first	teams/0/first	teams/0/first	teams/0/tow	teams/0/inhi	teams/0/bar	teams/0/dra	teams/0/vile	tea
2	2585564750	NA1	2038	9	7.17.200.395	100	Fail	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	4	0	0	0	0	
3	2585564753	NA1	2140	9	7.17.200.395	100	Win	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	11	4	1	3	0	
4	2585564758	NA1	1925	9	7.17.200.395	100	Win	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	11	2	0	3	0	
5	2585564759	NA1	2062	9	7.17.200.395	100	Fail	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	3	0	0	0	0	
6	2585564766	NA1	1979	9	7.17.200.395	100	Win	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	8	1	1	1	0	
7	2585564769	NA1	2181	9	7.17.200.395	100	Fail	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	7	1	0	3	0	
8	2585564771	NA1	1672	9	7.17.200.395	100	Fail	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	0	0	0	1	0	
9	2585564773	NA1	1536	9	7.17.200.395	100	Win	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	8	1	1	1	0	
10	2585564776	NA1	2010	9	7.17.200.395	100	Fail	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	5	0	0	1	0	
11	2585564780	NA1	2129	9	7.17.200.395	100	Fail	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	5	0	0	2	0	
12	2585564781	NA1	1533	9	7.17.200.395	100	Fail	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	0	0	0	0	0	
13	2585564783	NA1	1899	9	7.17.200.395	100	Win	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	10	1	1	4	0	
14	2585564784	NA1	2278	9	7.17.200.395	100	Win	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	11	4	1	4	0	
15	2585564785	NA1	1825	9	7.17.200.395	100	Fail	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	3	0	0	2	0	
16	2585564793	NA1	1850	9	7.17.200.395	100	Fail	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	5	0	0	1	0	
17	2585564794	NA1	1546	9	7.17.200.395	100	Win	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	10	2	1	3	0	
18	2585564800	NA1	1810	9	7.17.200.395	100	Win	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	10	3	1	2	0	
19	2585564801	NA1	2087	9	7.17.200.395	100	Fail	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	3	0	0	1	0	
20	2585564802	NA1	1698	9	7.17.200.395	100	Fail	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	5	0	0	0	0	
21	2585564806	NA1	1816	9	7.17.200.395	100	Fail	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	2	0	0	2	0	
22	2585564807	NA1	1933	9	7.17.200.395	100	Fail	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	2	0	0	1	0	
23	2585564808	NA1	1980	9	7.17.200.395	100	Fail	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	4	0	0	1	0	
24	2585564809	NA1	1723	9	7.17.200.395	100	Win	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	7	1	0	2	0	
25	2585564812	NA1	1943	9	7.17.200.395	100	Fail	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	2	0	0	1	0	
26	2585564814	NA1	1981	9	7.17.200.395	100	Win	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	9	2	1	2	0	
27	2585564817	NA1	2302	9	7.17.200.395	100	Fail	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	8	0	1	3	0	
28	2585564819	NA1	1574	9	7.17.200.395	100	Fail	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	0	0	0	0	0	

In-game Factors

Appendix C: Parameter Settings and Codes

Codes in R

```
> mydata <- read.csv("ingame.csv")
> head(mydata)
> summary(mydata)
> mylogit <- glm(teams.0.win ~ teams.0.firstTower + teams.0.firstBaron +
teams.0.firstBlood + teams.0.firstInhibitor + teams.0.firstDragon +
teams.0.firstRiftHerald + teams.0.towerKills + teams.0.inhibitorKills +
teams.0.baronKills + teams.0.dragonKills + teams.0.riftHeraldKills +
teams.0.dominionVictoryScore, data = mydata, family = 'binomial')
> summary(mylogit)
> confint(mylogit)
> library("caret")
> myImp <- varImp(mylogit, scale = FALSE)
> summary(myImp)

> mydata1 <- read.csv("champs.csv")
> summary(mydata1)
> plot(mydata1$blueTopChamp, mydata1$redTopChamp)
> plot(mydata1$blueJungleChamp, mydata1$redJungleChamp)
> plot(mydata1$blueMiddleChamp, mydata1$redMiddleChamp)
> plot(mydata1$blueADCChamp, mydata1$redADCChamp)
> plot(mydata1$blueSupportChamp, mydata1$redSupportChamp)

> library(rpart)
> library(rpart.plot)
> fit <- rpart(bResult ~ blueTopChamp + redTopChamp, method="class", data=mydata1)
> prp(fit)
> fit <- rpart(bResult ~ blueJungleChamp + redJungleChamp, method="class",
data=mydata1)
> prp(fit)
> fit <- rpart(bResult ~ blueMiddleChamp + redMiddleChamp, method="class",
data=mydata1)
> prp(fit)
> fit <- rpart(bResult ~ blueADCChamp + redADCChamp, method="class",
data=mydata1)
> prp(fit)
```

```
> fit <- rpart(bResult ~ blueSupportChamp + redSupportChamp, method="class",
data=mydata1)
> prp(fit)
> fit <- rpart(bResult ~ blueTopChamp + blueJungleChamp, method="class",
data=mydata1)
> prp(fit)
> fit <- rpart(bResult ~ blueJungleChamp + blueMiddleChamp, method="class",
data=mydata1)
> prp(fit)
> fit <- rpart(bResult ~ blueADCChamp + blueSupportChamp, method="class",
data=mydata1)
> prp(fit)

> library(car)
> mydata2 <- read.csv("players.csv")
> scatterplot.matrix(~KDA + GD10 + XPD10 + CSD10 | W., data = mydata2)
```

WEKA

Parameters for J48 Decision Tree:

batchSize = 100
binarySplits = False
collapseTree = True
confidenceFactor = 0.25
debug = False
doNotCheckCapabilities = False
doNotMakeSplitPointActualValue = False
minNumObj = 5
numDecimalPlaces = 2
numFolds = 3
reducedErrorPruning = False
saveInstanceData = False
seed = 1
subtreeRaising = True