

## An Analytical Approach to League of Legends' Professional Games

**Sijun Hwang and Abe Payne**

Department of Systems Engineering  
United States Military Academy  
West Point, NY 10996

Corresponding author's Email: [Sijun.Hwang@westpoint.edu](mailto:Sijun.Hwang@westpoint.edu)

**Author Note:** CDT Sijun Hwang, from Orange County, California, is currently a student at the United States Military Academy. Majoring in Engineering Management, CDT Hwang would like to extend his gratitude to Major Payne as well as Mr. Ian Kloo for all assistance in this independent research. He would also like to extend his thanks to the Department of Systems Engineering at West Point.

**Abstract:** This paper uses both logistic regression as well as random forest to analyze which variables are important and significant to winning in League of Legends, a popular massively multiplayer online battle arena game. It looks into different variables that are seen as the most important from casual players and applies them to professional games. Additionally, this paper describes the results from these models and compares them to how professionals or casuals play League of Legends.

**Keywords:** League of Legends, Professional Play, LCS, Data Analysis, Logistic Regression, Random Forest

### 1. Introduction

Esports is continuing to grow, with games like League of Legends boasting some of the largest viewership. In the League of Legends 2019 World Finals, there were over 100 million viewers bringing in fans from different streaming services including Twitch and YouTube. In comparison, the Super Bowl had about 100.7 million viewers (Roundhill Team, 2020). Additionally, the esports industry has generated revenue that exceeds 1 billion dollars and audiences of more than 443 million around the world (Ayles, 2019)

League of Legends, also referred to as League or LoL, is a multiplayer online battle arena (MOBA). In League, there is a nexus, which acts as the final objective killed to win or lose a game. The goal for each summoner, or player, is to destroy the nexus to see the victory screen at the end of the game. However, there are eleven turrets that defend three main lanes that champions, or in-game characters, travel through. Three of these towers defend inhibitors at the entrance of each base. Figure 1 below has a top-down view of what the map looks like, showing each side. When inhibitors are destroyed, it gives only the allied lane stronger minions. Inhibitors respawn five minutes after being destroyed. In between these lanes is what is known as the jungle. In the jungle, there are different types of jungle camps or monsters that champions can kill for extra gold and experience. Additionally, the nexus spawns minions in intervals of thirty seconds in which they travel down each lane towards the enemy base. These minions also grant gold and experience. Gold is used to increase the combat power of a champion by purchasing different items, while experience will level up your champion. A higher-level means that one will be able to use more powerful abilities, increasing combat power.



Figure 1. A Top-Down View of the Map in LoL

From a spectator's perspective, League of Legends is starting to follow suit of what ESPN does with a game's box scores. Lolesports.com contains Videos-on-Demand, or VODs, that will give statistics of the different kills, deaths, and assists that each player received. Additionally, it gives numbers on different objectives that a team has accomplished in addition to other data.

## **2. Background Research**

To further understand League of Legends and what research has been done, Alex Leavitt, Brian C. Keegan, and Joshua Clark from USC, Harvard, and USC, respectively, investigated communication methods within League of Legends. This trio established three hypotheses that pings (quick pre-set messages) affect team performance. They discovered that pings differed based on team role and skill, pings varied based on kills and assists, and pings had a concave relationship with performance. (Leavitt, Keegan, & Clark, 2016) Additional research has been conducted regarding a data-driven approach that created different combat rules trained from a decision tree. The result from this research identified methods that new players can use to better be successful in these types of games. (Yang, Harrison, & Roberts, 2014). This research has dived into the finer intricacies of the game, but there has not yet been research on which factors contribute to winning and which factors within a game players should focus. In this paper, data from thousands of League of Legends esports games will be analyzed to determine which variables in the game help teams win at the highest level.

## **3. Modeling**

### **4. Methodology**

#### **3.1.1 Data Collection**

Data was first collected through oracleselixir.com. CSV files were extracted from their website that contained match data from every League of Legend league including regions such as China, Korea, North America, and Europe. Each CSV file contained a game ID for each game, the URL in which the match results are from, and the league in which the game was played in. Other columns include the different picks and bans (champions that one team can forbid the other from playing for just that match) that each game had, the game length, and the result of the game. More specific data is shown in these files, which included kills, deaths, assists, team kills per minute, dragons killed, wards placed, vision score, total gold, earned gold, as well as individual stats at 10 and 15 minutes into the game. In total, there are about 119 different columns for each player.

#### **3.1.2 Data Cleaning**

Creating the data frame started with binding each data frame into one large, total data frame. From there, the data frame was filtered into different leagues. This included LCS, LCK, LPL, and LEC, which are the North American, Korean, Chinese, and European leagues respectively. These leagues were chosen because they are known as the major regions that compete in the game. There are many minor regions that range from Brazil, Malaysia, Thailand, and even Latin America. From there, the total data frame was filtered by total game results for each side rather than the specific professional player. Additionally, if there were any rows that were incomplete, the entire row was removed. One large change that was made to this data frame was that matches were only pulled if they were on patch 10.01 or later. This is because elemental dragons were introduced to the game, largely changing both the style and strategy of the game.

With a new and large data frame, much of the data frame still needed to be cleaned. Columns were renamed for consistent syntax, and redundant or unnecessary columns were removed. Many of these observations and columns were filtered because they were either incomplete or would not assist in the overall research. This allowed for the data to be processed faster and eliminated noise that would have been dealt with otherwise.

#### **3.1.3 Model Building**

The first model involved making a generalized linear model. Logistic modeling was used because the dependent variable being analyzed was a categorical variable from 0 to 1 (with 0 being a loss and 1 being a victory). The model investigates which independent variables contribute to the result of the game. Logistic Model 1a assessed the variables most players consider the most important to winning the game. This first model included gold spent difference (gspd), kills, deaths, towers, dragons killed, barons, elders, and first herald. These variables are the most common and popular objectives on which teams typically focus. Many games in League will be dependent on who controls these objectives. Typically, control of these objectives leads to a win. This model was then modified multiple times to account for insignificant variables and potential confounding variables.

The second model looked at different aspects of the game that players consider but are not considered as important as the variables in the first model. These new variables included damage per minute, damage taken per minute, vision score, and creep score (total monsters killed).

The third and final model examined a random forest model, using the variables from Logistic Model 1a, Logistic Model 1b, and Logistic Model 1c. These three models were very similar, with just variables taken out each time. Logistic Model 1a had every variable listed before. Model 1b took out the *gspd* as it was an obvious confounding variable. Logistic Model 1c narrowed these variables even more only accounting for variables that the game labeled as “objectives” but were not monsters.

## 5. Results

Logistic Model 1a produced surprising results. For example, dragons, barons, elders, and first herald were all considered insignificant as shown below in Figure 2. Additionally, a variables’ significance was determined by its z-value or easier to understand p-value which is denoted in column “Pr(>|z|)” of Figure 2. Here, if the number is .05 or smaller, it would be considered significant to our model. The stars to the right of the number is a quick way to see which variables were significant.

Dragons typically force the game to progress, in which the Elder dragon spawns after one team has slain four dragons. This makes sense in the fact that many professional teams do not commit all of their resources in killing dragons and would often look for other objectives or plays to make if they do not have control of the area around the dragon.

The Barons parameter is much more surprising in the fact that Barons give the entire team gold, experience, and a temporary power-up to each individual player making them stronger. Additionally, Baron makes nearby allied minions stronger, allowing for teams to group and apply immense pressure, or to split and apply pressure in the three lanes.

The Elder parameter is the most interesting in its insignificance. Most players, coaches, and analysts would argue that killing the Elder dragon would mean that the end of the game is soon. Elders grant an individual power buff that allow players to ignore both armor and magic resist, as well as an execution that instantly kills a player if their health drops below 20 percent. Typically, even if a team is behind, killing the Elder dragon may signal a potential comeback.

Killing the first herald also is an interesting variable that was deemed insignificant. Typically, the first herald provides a strong minion that can be summoned for the team. This herald will crash head-first into turrets, assisting in killing turrets. Additionally, killing turrets grant gold and killing the first turret grants even more gold for the team.

Another interesting thing to note is that the gold spent difference was a negative number. This -16.23 means that if the percentage of *gspd* were to increase by one percent, the log odds of winning decreases by 16.23. This result raises concern, as gold is the most direct measurement used in professional play to see the difference in how two teams played. Gold directly translates to combat power, and combat power translates to being the stronger team in order to win the game. This raised questions as to if any of the variables were collinear or confounding with each other. Upon a variance inflation factor test or VIF in R, none of the variables were found to be collinear. To solve for confounding variables, *gspd* was removed from the logistic model to see if it made a large difference in the coefficients.

Inhibitors was also a negative number, which rose some questions. Inhibitors are utilized by teams in order to bring pressure to a certain lane, forcing enemy champions to defend their base. This would allow teams to fight with a numbers advantage, giving teams a five-minute window necessary to win. However, recent analysis points out that depending on the state of the game, killing inhibitors may assist the enemy team (Sutton, 2021). These empowered minions would give the enemy extra gold and experience that would allow for them to shorten the gold gap between two teams. As long as a team can manage to avoid fighting while their inhibitor is down, destroying an inhibitor may not contribute towards winning.

```

Coefficients:
      Estimate Std. Error z value Pr(>|z|)
(Intercept) -12.46629   1.56774  -7.952 1.84e-15 ***
gspd        -16.42426   3.23742  -5.073 3.91e-07 ***
kills         0.55285   0.05979   9.246 < 2e-16 ***
deaths       -0.61304   0.05991 -10.233 < 2e-16 ***
inhibitors   -0.83759   0.28976  -2.891 0.00384 **
towers        1.95521   0.21966   8.901 < 2e-16 ***
dragons       0.22169   0.15268   1.452 0.14651
barons        0.19747   0.26734   0.739 0.46012
elders        0.24940   0.54780   0.455 0.64891
firstherald  -0.35760   0.35820  -0.998 0.31812
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 4469.38 on 3223 degrees of freedom
Residual deviance: 240.96 on 3214 degrees of freedom
AIC: 260.96

Number of Fisher Scoring iterations: 10
    
```

Figure 2. Coefficient Results for Logistic Model 1a

Logistic Model 1b is the same as Logistic Model 1a except the fact that gspd was taken out. There is no difference from this model compared to the other in terms of significance. Every variable that was significant before continued to be significant, while every variable that was insignificant continued to be insignificant.

Logistic Model 1c as shown in Figure 4 is the final product of the first logistic model accounting for only significant variables as well as confounding variables. This gives more visibility on the true effect that these variables may have, without any influence from insignificant variables. Likewise, it recognizes that these variables are still significant. All variables here are significant and show that, of the objectives that are laid out in game, towers and kills are the ones that contribute the most for winning.

```

Coefficients:
      Estimate Std. Error z value Pr(>|z|)
(Intercept) -11.14040   1.44504  -7.709 1.26e-14 ***
kills         0.45976   0.05022   9.155 < 2e-16 ***
deaths       -0.48174   0.04565 -10.554 < 2e-16 ***
inhibitors   -0.69445   0.27335  -2.541 0.0111 *
towers        1.70529   0.19814   8.607 < 2e-16 ***
dragons       0.25061   0.14542   1.723 0.0848 .
barons       -0.18581   0.24533  -0.757 0.4488
elders        0.51114   0.50778   1.007 0.3141
firstherald  -0.55158   0.33171  -1.663 0.0963 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 4469.38 on 3223 degrees of freedom
Residual deviance: 269.41 on 3215 degrees of freedom
AIC: 287.41

Number of Fisher Scoring iterations: 10
    
```

Figure 3. Coefficient Results for Logistic Model 1b

```

Coefficients:
      Estimate Std. Error z value Pr(>|z|)
(Intercept) -10.1754   1.2930  -7.870 3.56e-15 ***
kills         0.4666   0.0496   9.407 < 2e-16 ***
deaths       -0.4852   0.0450 -10.783 < 2e-16 ***
inhibitors   -0.5624   0.2695  -2.087 0.0369 *
towers        1.5855   0.1811   8.756 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 4469.38 on 3223 degrees of freedom
Residual deviance: 276.93 on 3219 degrees of freedom
AIC: 286.93

Number of Fisher Scoring iterations: 10
    
```

Figure 4. Coefficient Results for Logistic Model 1c

In Figure 5, the coefficients of the log odds from Figure 4 were taken out. These are shown in a small list below, with log odds accounting for a row and each variable has its own column. Kills is read as follows: for each increase in one kill for a team, their log odd of winning the game increased by approximately .467. This can be applied to every coefficient after that, with the log odd of winning the game either increasing or decreasing based on the coefficient shown.

	(intercept)	kills	deaths	inhibitors	towers
Log odds	-10.1754	0.467	-0.485	-0.562	1.586

Figure 5. Coefficient for each Variable

The results from the second set of models are shown below. This first one was with the variables from Logistic Model 1c, with an added variation each time. Figure 6 below indicates that damage per minute was significant, which would seem self-explanatory. The more damage you deal to the enemy team, the enemy would either die or be forced to retreat. This would mean that you would be free to grab other objectives on the map as well as pressure the enemy base. Figure 7 displays the same model as Figure 6, but with damage taken per minute substituting damage per minute. The reason that the variables from Logistic Model 1c are kept the same is that while we hold those initial variables constant, both damage per minute and damage taken per minute are still significant.

```

Coefficients:
      Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.161e+01  1.463e+00 -7.938 2.05e-15 ***
kills        4.202e-01  5.241e-02  8.016 1.09e-15 ***
deaths      -5.163e-01  4.787e-02 -10.784 < 2e-16 ***
inhibitors  -5.877e-01  2.740e-01 -2.145  0.0319 *
towers       1.577e+00  1.821e-01  8.660 < 2e-16 ***
dpm          1.288e-03  5.119e-04  2.516  0.0119 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 4469.38 on 3223 degrees of freedom
Residual deviance: 270.33 on 3218 degrees of freedom
AIC: 282.33

Number of Fisher Scoring iterations: 10
    
```

Figure 6. Coefficient Results for Damage per Minute

```

Coefficients:
      Estimate Std. Error z value Pr(>|z|)
(Intercept)  -7.2768652  1.4416288  -5.048 4.47e-07 ***
kills         0.5092997  0.0513970   9.909 < 2e-16 ***
deaths       -0.4312323  0.0472251  -9.131 < 2e-16 ***
inhibitors   -0.5248505  0.2783727  -1.885  0.0594 .
towers        1.6532650  0.1918677   8.617 < 2e-16 ***
damagetakenperminute -0.0016509  0.0004143  -3.985 6.75e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 4469.38 on 3223 degrees of freedom
Residual deviance: 259.73 on 3218 degrees of freedom
AIC: 271.73

Number of Fisher Scoring iterations: 10
    
```

Figure 7. Coefficient Results for Damage Taken per Minute

These next two models investigate variables that are not associated with interacting with the enemy. Essentially, these variables are factors in the game that players can focus on their own. Both creep score, which is the amount of monsters you kill in a game, and vision score, a score given to you based on how often a player granted or denied vision for their team, are scores that theoretically can have zero interactions with the enemy. The results are shown below in Figures 8 and 9.

```

Coefficients:
      Estimate Std. Error z value Pr(>|z|)
(Intercept) -20.49376    3.28547  -6.238 4.44e-10 ***
kills         0.50032    0.05237   9.554 < 2e-16 ***
deaths       -0.46107    0.04676  -9.860 < 2e-16 ***
inhibitors   -0.43821    0.27640  -1.585 0.112868
towers        1.56107    0.18388   8.490 < 2e-16 ***
cspm          0.29471    0.08264   3.566 0.000362 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 4469.38 on 3223 degrees of freedom
Residual deviance: 263.55 on 3218 degrees of freedom
AIC: 275.55

Number of Fisher Scoring iterations: 10
    
```

Figure 8. Coefficient Results for CSPM (Creep Score per Minute)

```

Coefficients:
      Estimate Std. Error z value Pr(>|z|)
(Intercept)  -7.527384    1.450582  -5.189 2.11e-07 ***
kills         0.436246    0.052637   8.288 < 2e-16 ***
deaths       -0.460888    0.049440  -9.322 < 2e-16 ***
inhibitors   -0.499950    0.273351  -1.829 0.06740 .
towers        1.535215    0.188154   8.159 3.37e-16 ***
visionscore  -0.007154    0.002463  -2.905 0.00367 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 3600.21 on 2596 degrees of freedom
Residual deviance: 234.16 on 2591 degrees of freedom
(627 observations deleted due to missingness)
AIC: 246.16

Number of Fisher Scoring iterations: 9
    
```

Figure 9. Coefficient Results for Vision Score

The results from these models indicate that both cspm and vision score are significant as indicators for winning games. In this, CSPM is the team's creep score per minute and vision score is the aggregate vision score for the entire team.

The final model utilized Random Forest to perform regression and classification tasks. Random Forests uses a technique that combines multiple decision trees in determining the final output. This means that multiple decision trees are run parallel to each other in which predictions are made. From there, all the predictions are averaged to make a Random Forest Prediction (Bakshi, 2020).

Below in Figure 10, are the results from the random forest model. It is important to note that the Mean Decrease Accuracy is what matters for importance. The Mean Decrease Accuracy is how much the accuracy of the model decreases if that variable is dropped from the model. Mean Decrease Gini is a measure of the variable importance based on the Gini impurity index used for the calculation of splits in trees (Bhalla, 2014).

	0	1	MeanDecreaseAccuracy	MeanDecreaseGini
kills	9.513793	34.74902	35.793668	163.23016
deaths	15.334839	37.94354	36.294333	153.24732
inhibitors	16.496734	40.51047	40.571848	337.34216
towers	17.236776	36.82901	37.682603	428.60535
dpm	2.631246	27.34192	26.966556	44.20918
damagetakenperminute	10.980015	15.38259	18.643848	21.34754
cspm	10.912450	4.49609	10.778307	30.34515
visionscore	15.531290	-11.34996	1.797289	18.29611

Figure 10. Mean Decrease Accuracy and Mean Decrease GINI for random forest Model

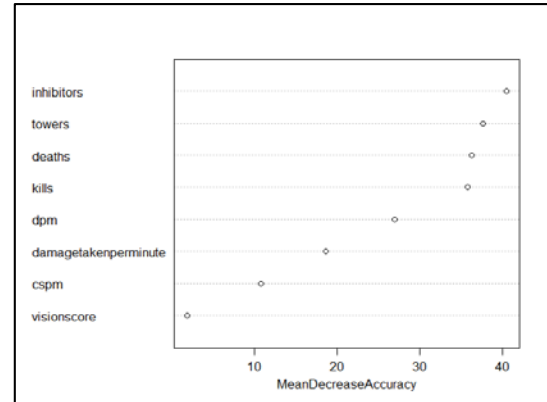


Figure 11. Graph of Variables by the Mean Decrease Accuracy

This random forest regression used all the variables that were deemed significant in Logistic Model 1c, as well as all the added variables in Logistic Model 2. It is important to note that random forest would not have run unless there was absolutely no NAs in the dataset, causing for these models to be ran on about three thousand observations rather than four thousand. Looking at both Figure 10 and 11, the results provide a new insight. This new model put inhibitors and towers at the forefront of what gave the model the most importance. Specifically, it means that if inhibitors were to be taken out of the random forest model, it would drop the model accuracy by 35.79. Figure 11 is an easy way to visualize which ones are the most important variables from top to bottom.

## 6. Discussion

This research has confirmed what most players of the popular video game, League of Legends, already know. However, this information is still valuable and still worth noting in order to confirm that these variables do matter when trying to win. New players can use this information to better understand the game and to lessen the learning curve when playing this game. For example, the objectives that are in game that players can most easily grasp (inhibitors, towers, deaths, kills), are confirmed to be the most important. This may allow newer players to focus on these four objectives when first playing the game before focusing on more minor parts of the game such as damage per minute, damage taken per minute, creep score per minute, and vision score. These later variables are much more difficult to master in-game, and many professionals tend to focus on these variables in order to give them the winning edge. The data that has come from these professional games give good insight into how the game should be played at the highest level.

Many League of Legends professional teams have dedicated coaches and analysts that dive into these numbers, examining what the best possible strategy is. This research informs casual players that even if the professional scene focuses on certain objectives, it may not be what immediately should be focused on. For example, Logistic Model 1b consisted of every objective that are considered major objectives in the game. However, it indicates that dragons, barons, elders, and first herald are not significant to winning the game. This grants insight for casual players that focusing on maximizing kills and minimizing deaths is something more important to consider. If a player is contesting dragon control with the enemy and is risking a fight they are going to lose, it might be better for that team to back off and focus on other objectives such as creep score per minute or even vision score. This same logic can be applied to barons, elders, and first herald. Often, professionals completely give up dragon, baron, or herald control because it simply is not worth dying. This research is a reflection on pro-players' play style and confirms why the professionals do what they do. Instead, professionals opt to apply pressure on other sides of the map.

## 7. Limitation

There were many limitations when performing this research. For example, what started as a data frame with about 12,000 observations, the data frame got reduced to about 4,000 observations after filtering for incomplete data. Even if there were still about 4,000 observations to look at, much of the data that could have been useful in calculating the overall significance and importance of variables have been overlooked. Many of these variables had small typos, and the format in which these variables were recorded were inconsistent. Additionally, many of the columns that could have been important such as heralds in total had to be deleted due to improper calculations. Other data such as creep score per minute also had many observations filtered out from the model because of incomplete data. One other variable that would have been useful would have been team quality. The quality of the team going into many of these matches would have been useful to use, as it could have been another independent variable to consider. For example, if a team was on a win-streak or recently signed a very good player, the team quality of that team could have gone up while the team quality of another could have gone down.

## 8. Conclusion

This research is just the start of more research that can be done in the esports field. As esports continues to rise, many people are starting to see this industry as a career opportunity. The amount of analysis that can be done in just a couple of years of professional gaming leads to a wider array of possibilities in which analysts and enthusiasts alike can utilize their expertise to quantify how to win. As more research continues to come out, it will be much easier for players and coaches to justify why the game should be played differently and make strategies accordingly. League of Legends is a game that is developing rapidly and there already is a difference in play style from when the game first started.

## 9. Further Research

Further research may include diving into many of the other variables that were not considered in this research. There were variables such as damage mitigated, monsters killed in own jungle/enemy jungle, and even champion picks that were not used. One can even investigate the same data this paper focused on and compare it to the skill of the player. Would these same strategies work for casual players or are these strategies easier to accomplish after having a base knowledge of the game? It would be effective to see how investigating different levels of expertise influence one team's approach to winning. Other research may include looking at specific parts of the game including the time and noting the impact it has on winning and losing.

## 10. References

- Ayles, J. (2019, December 3). *Global Esports Revenue Reaches more than \$1 Billion as Audience Figures Exceed 433 Million*. Retrieved from Forbes: <https://www.forbes.com/sites/jamesayles/2019/12/03/global-esports-revenue-reaches-more-than-1-billion-as-audience-figures-exceed-433-million/?sh=14044f131329>
- Bakshi, C. (2020, June 8). *Random Forest Regression*. Retrieved from gitconnected: <https://levelup.gitconnected.com/random-forest-regression-209c0f354c84>
- Bhalla, D. (2014, November). *A Complete Guide to Random Forest in R*. Retrieved from Listen Data: <https://www.listendata.com/2014/11/random-forest-with-r.html>
- Leavitt, A., Keegan, B. C., & Clark, J. (2016). Ping to Win? Non-Verbal Communication and Team Performance in Competitive Online Multiplayer Games. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 4337-4350). San Jose: Association for Computing Machinery, New York.
- Roundhill Team. (2020, September 25). *Esports Viewership vs. Sports in 2020*. Retrieved from Roundhill Investments: <https://www.roundhillinvestments.com/research/esports/esports-viewership-vs-sports>
- Yang, P., Harrison, B., & Roberts, D. L. (2014). Identifying Patterns in Combat that are Predictive of Success in MOBA Games. *Foundations of Digital Games 2014* (p. 36). Ft. Lauderdale: Society for the Advancement of the Science of Digital Games.