LEAGUE OF LEGENDS: PREDICTING GAME RESULTS WITH EARLY GAME VARIABLES

Final Exam for Big Data Analytics

Henrique Campos Amorim – S150304

Department of Digitalisation

Dr. José Parra-Moyano

November 25, 2022

Table of Contents

List of Figures

Abstract

Esports generate public and detailed data for all games, are expected to grow as an industry by a 21,9% CAGR, and involve several stakeholders such as the game producer, professional teams, and third-party software companies. Specifically looking at League of Legends, this research focused on understanding how to predict the outcome of matches. It did so given a market potential of such a solution to a monthly active player base of 180 million. The research question is: How well can past data of up to 10 minutes of gameplay predict the match outcome? The analysis builds on past research regarding game prediction in MOBAs and draws on methods such as Decision Trees and Neural Networks for prediction models. The dataset utilized contained 9.598 games collected in 2020 from high-level players and the results show that gold difference is the most important variable. From a commercial perspective, the accuracy results of around 74% indicate that more data is needed to build a third-party software that players would be willing to buy. However, the lack of high predictive power also illustrates that the game developer, Riot Games, actively tweaks the game so that early performance does not automatically translate to wins. Finally, it is recommended for third-party software companies to continue exploring and including additional data to increase model accuracy, while on the other hand it is recommended that Riot Games doubles down on the continuous improvement strategy to guarantee teams have a chance of winning despite poor initial performance.

Key words: esports, neural networks, game prediction, data analytics, decision trees

Introduction

Revenues of the esports industry, comprised of all the competitive multiplayer video games or electronic sports, are on the track to achieve 21,9% CAGR until 2030 reaching USD 5.74 billion (Straits Research, 2022). Different modalities, or games, have achieved recognition as legitimate sports and both players and streamers alike have become celebrities. Looking specifically at League of Legends (LoL), a game produced by Riot Games (Riot) and first published in 2009, the numbers of monthly players skyrocketed to 180 million in 2022 since the launch of their Netflix series Arcane. Famous players such as @Bjergsen and @brttOficial have 1,3 million and 1 million followers on Twitter while LoL's official account (@lolesports) has over 2,3 million followers also on Twitter. Overseeing and investing in the scenario, Riot had revenues of USD 1.75 billion in 2020. Due to the number of players and the large influence sphere of LoL, many third-party applications have been created to help or assist players. They range from recommending specific strategies and paths to which items to buy and collect data by accessing the Riot's database through an API.

In LoL, two teams compete simultaneously against each other and games last around 30 to 40 minutes in average. Players, however, have the option to forfeit, or surrender, as soon as the 15 minutes mark. Since games last a considerable amount of time and players have the option to forfeit, a potential market opportunity exists for a third-party software capable of collecting and processing game data to recommend players whether they should continue playing or surrender. Despite tech valuations being down in 2022 compared to previous years, a total addressable market of 180 million people is equivalent to nearly 30 times the size of Denmark. Even when considering that the average player would not be willing to pay for the service, running ads on the platform could potentially be enough. Nevertheless, despite possibly low conversion number the attractiveness of the opportunity calls for exploration.

This research will therefore analyze the predictability of match results by processing information collected until the 10-minute mark of high-skill LoL games. at the Diamond level to given. Hence, the research question is: "can information collected from the first 10 minutes of game accurately predict the winning team in League of Legends matches?". This question, however, does not cover pre-game data, data from after 10 minutes, or different skill levels. It also does not include additional methods other than descriptive analysis, decisions trees, and neural networks.

League of Legends: Gameplay

LoL is a multiplayer online battle arena (MOBA) in which two teams of five players compete to defeat the opponent, achieved by advancing into the opponent's base and destroying the Nexus.[1] To win, each of the ten players choose a character to play with, known as a champion, from a pool of over 160 options. Players cannot repeat champions in the same match and champion strength varies throughout time as the game changes, being therefore an important variable to outcome prediction.

Once the game starts, players move through the top, middle, and bottom paths shown in <u>Appendix A</u>, known as lanes. Two players from each team fight on the bottom lane, one on the top, one on the middle, and one player roams around. The champions that each player chooses can become stronger through leveling up, which comes from gathering experience (XP), or through items, which come from gathering gold. To reach the other team's base and destroy the Nexus, players must also destroy towers (or turrets) along the way. Destroying towers gives gold to the player(s) involved.

Gold is regarded as "one of the main ways for champions to increase their power over the course of a game" (League of Legends Wiki, n.d.), can be collected from different sources, and gold collection is highly correlated with XP collection. The main source of gold and XP in the game are minions, neutral monsters that are created automatically, or spawned, by the game every 30 seconds. Players collect gold only by successfully killing minions. The Herald is an additional neutral monster that spawns only once in the first 10 minutes, gives a small amount of gold and XP, but significantly impacts the game by assisting in taking down towers. The Dragon is the third type of neutral monster, spawns up to two times in the first 10 minutes, and gives gold, XP, and strength benefits to all players in the team that kills it.

Additionally, players collect gold and XP by taking down the enemy champions. Killing an enemy is an effective way of generating gold and XP even when you are not the main player involve, which counts as an assist. By dying, players don't lose any of their money, but rather the game itself gives the money to those involved in the kill. Lastly, an important aspect of the game is vision. Most of the map remains dark to the players due to fog of war and wards are the objects that increase vision. Their importance after the first 10 minutes and as the game progresses however is much higher.

[1] In similar fashion a game of Capture the Flag.

Conceptual Framework and Related Works

The most important concept to understanding the dynamics of winning and losing in LoL is that of gold. By amassing gold, players buy items that strengthen their champions and increase the odds of winning fights. After having won a fight, teams attempt to push the opponents further behind by progressing through the map, destroying turrets, and taking down the Dragon or Herald, all of which give more gold. By successfully executing this strategy, teams achieve a virtuous cycle of growth, also known as snowballing effect.

The expected relationship between variables is, therefore, that gold will be the main lever toward predicting game outcome from pre-10 minutes game data while most other variables will be affecting the probability through the effect of gold itself. One meaningful addition is that relative levels are just as, or even more, important than the absolute levels. One hypothesis to be tested is that the relative gold difference between teams is more correlated to winning than the total gold amount collected. Past research in similar games has corroborated with the importance of the gold variable.

Albeit not focusing on the same game, previous research has demonstrated the effectiveness of utilizing game data to predict the outcome in MOBAs. Hodge et al. (2017) conduct a literature review showcasing that most of previous works have relied on logistic regression (LR) for prediction while some incorporated random forests (RF), decision trees (DT), and meta-learning. In their work, the authors focus on DotA 2, a similar game to LoL, and utilize a mix of professional matches and public matches from high-skilled players. Identifying the insufficiency of data from professional matches, Hodge et al. (2017) demonstrate that there is only a slight loss of accuracy from incorporating non-professional data. Through LR and RF, the authors claim that the loss is only small through careful model and parameters evaluation, but it is important to mention that the results still possess low accuracy overall.

Building on their previous work, Hodge et al. (2021) further investigate the case of prediction for professional games with real-live data. By incorporating real-live data, the authors find an increase in accuracy when compared to previous works, and through similar fashion (LR and RF models). The highest accuracy reported was of 74.59% (Hodge et al., 2021). Highlighting an additional piece of the puzzle, Yang et al. (2016) showcase the importance of adding pre-game information to the model when predicting for DotA 2. As discussed earlier in this research, the relative strength of each champion matters to the game result. By feeding similar data to a LR model, Yang et al. (2016) observed an increase in accuracy from 58,69% to 71,49%.

Methodology

Data Set

The dataset utilized contains 9.879 observations collected between March and May 2020. Each consists of information generated in the first 10 minutes of a game played by high-skill players and all games lasted more than 15 minutes, most likely to control for players who quit after the game had started. They were collected through the Riot API by a Kaggle user and made available online (Lan Ma, 2020). The dataset features a unique game identifier, the outcome of the match, and 19 performance variables for each team about gold, experience, wards, neutral monsters killed, towers, and variables regarding kills, deaths, and assists. The distributions can be seen in Appendix B. There were no missing values. It is important to note that gameplay data quickly becomes irrelevant as Riot issues new versions of LoL, called patches, approximately every two weeks. The models fitted would therefore only be meaningful for games that happened in that timeframe or shortly after.

The process for analyzing the data after collection followed the classic steps of collecting, transforming, cleaning, and modeling. After collection, the first step was to focus on one of the sides as the effects will be the same for either. The blue observations were kept and the variables *blueCSPerMin*, *blueGoldPerMin*, and *blueTotalExperience* were excluded for being just transformations of other variables. When analyzing the independent variables, the distribution of *blueWardsPlaced* showcased an unusually long tail, most likely from games where players give up and keep placing wards as protest. Therefore, observations more than three standard deviations from the mean were dropped, reducing total observations to 9.589. Following the same logic as *blueGoldDiff*, *blueMinionDiff* was created from combining the blue and red data sets and represents the distance in total minions killed between the teams.

As the models applied afterwards have different assumptions and some require scaled data, these transformations were made prior to each model drawing from the base data set called *blue*. Still in the pre-processing, a factor variable was created from the *blueWins* to assist in descriptive statistics and as it would ease future operations. The final summary of *blue* can be seen on Appendix C.

Modeling, Methods, and Tools

As the analysis conducted is one of predictability, DT was the first utilized model. Given how the dataset is comprised of several different players and games spanning over two months, the hypothesis of a single data generating process would be very strong. DTs are therefore useful as they do not require such assumption nor assume linear data sets and as the dependent variable is categorical. Prior to fitting it, categorical variables were transformed into factor variables and the data set was split into a train set, with 70% of the total data, and a test set with 30% through the *createDataPartition()* function. The model was then fitted to the train data through the *rpart()* function and plotted through *rpart.plot()*. The result of the first model, however, was a DT containing only the root node. To force the tree to have more nodes, *rpart.control()* was applied to change parameters. The accuracy of both models was then measured against the sample they were trained on, which was the same, and the test sample subsequently. The *confusionMatrix()* function was critical to assess the accuracy. Finally, a 10-fold cross validation was applied through the *train()* function and the model was assessed following the same steps as previously.

In addition to DTs, neural networks (NNs) were applied to investigate if there would be a gain in accuracy. Due to NNs being universal function approximators, the lack of assumptions about data generating processes fits well with this data. Prior to fitting the model, it was necessary to scale the continuous variables with *scale()* and to build a *model.matrix()* from the categorical variables. The data was partitioned as with the DTs. A first model was fitted with the *neuralnet()* function and (2,1) neurons in the hidden layers. A second model was then fitted with (3,3) neurons in the hidden layers and, lastly, a model was applied through 4-fold cross validation. For the last model, however, it was necessary to utilize *nnet()* due to package limitation.

Limitations, Alternate Methods, and Cost / Benefit Analysis

The main limitation of the data set is the lack of information about the champions. Riot continuously releases patches that change the relative strength (among other attributes) of champions and past works have highlighted the high importance of it. Additionally, they have also mentioned the benefits from including data about the players' past performance. Some players perform better with specific champions, so this information could potentially increase model accuracy. Reflecting on the fact that most players do not reach the high-skill level, the dataset has limited applicability. It would be interesting to collect data from all different skill levels and study the clustered predictions versus general predictions.

An additional limitation is the lack of late-game actions as explanatory variables. As expected from previous research in similar games, early-game actions matter less to the match result than those in the later stages (Hodge et al., 2021). However, unlike the authors, this research focus on providing a recommendation for surrendering or not as soon as possible, therefore adding late-game data might be helpful but defeats the purpose.

As mentioned, past research utilized LRs and focused on algorithm improvement and careful choices of parameters and model specification. This exercise focused, on the other hand, on analyzing the predictive power of early-game data. Going forward, incorporating LRs would be a natural step to rank independent variables according to importance and further fine-tune the models to achieve the highest possible amount of accuracy. As the perspective taken in this work is that of a potential product, a discussion of alternate methods is key to ensure the approach taken is the one that offers most accuracy while keeping the costs associated with processing the data at an acceptable threshold.

Further analyzing the costs and benefits of big data analytics, the discussion would have taken a different angle had it been about evaluating performance or explaining phenomena. In these situations, companies risk becoming disconnected from the situations that lead to the results observed in the data and going overboard with data-driven decision making. Additionally, it is expensive to hire, train, and keep a data analytics department. The analysis for the case of this work is distinct, as the costs and benefits merge with the feasibility hypothesis of the product in its core. Therefore, it would be important to incorporate market size, pricing, and potential upside to execute a comprehensive comparison with the costs.

Results

The first interesting result stems from the hypothesis that gold difference between teams would be more relevant than absolute level of gold. The result is especially interesting as many commentators and analysts of the game choose to focus on absolute and not relative levels. Despite the strong correlation between the two of 0.89, the variable with the highest correlation with winning was gold difference with 0.51. The Correlation Analysis can be found in Appendix D.

Figure 1

Title: Density According to Difference in Gold

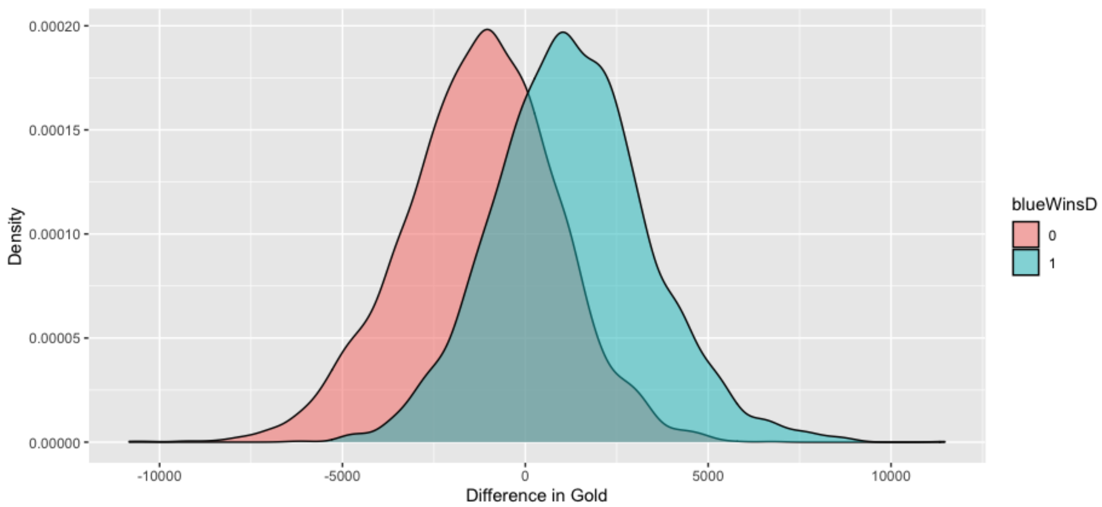
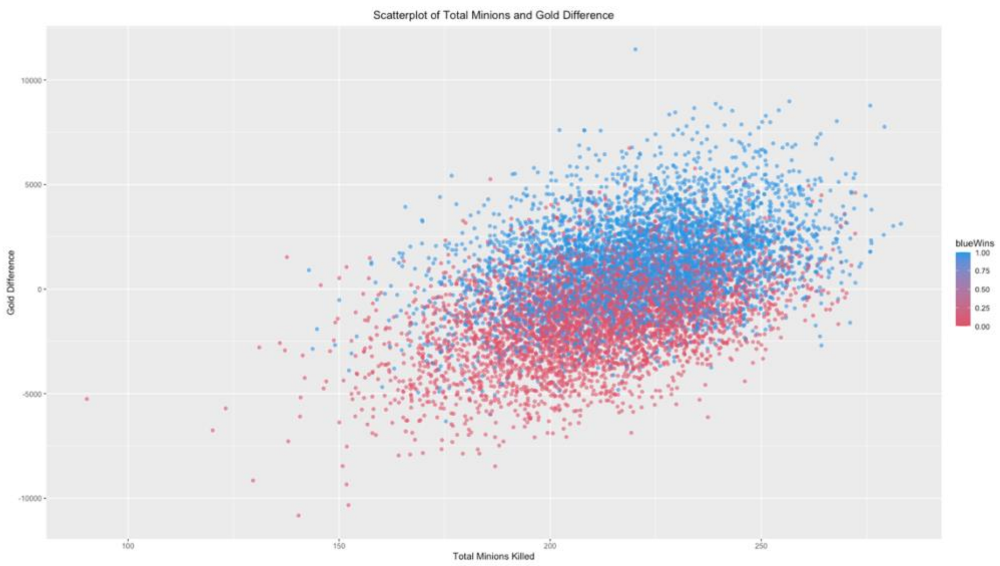


Figure 2

Scatterplot of Total Minions and Gold Difference

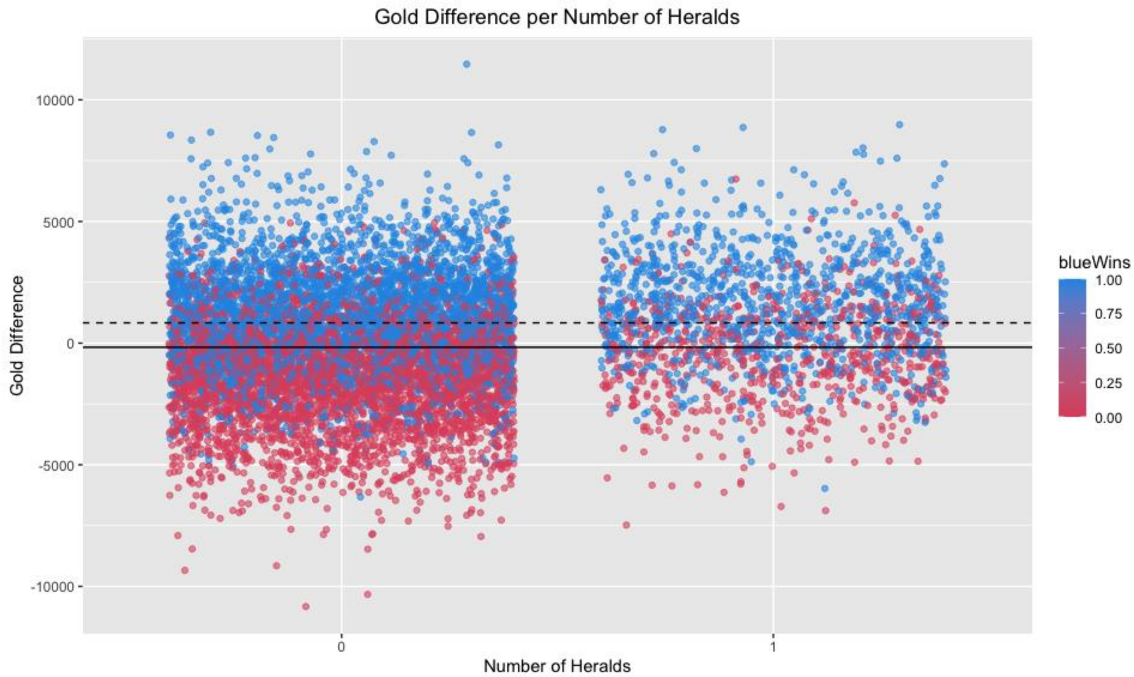


By plotting Figure 1, gold difference clearly reveals itself as a strong predictor for the winning chances of a team. One of the main ways to collect gold, by killing minions offers more perspective. In Figure 2, it is possible to see how the two are related. The meaningful result here is that teams should strive to collectively kill the most minions possible, with number over 225 resulting in significantly more victories. Because of the very similar correlation with difference in gold, it is possible to argue that the total number of minions killed by one team is just as relevant gold difference as the difference in minions between teams.

Another meaningful fact from the descriptive analysis, better described as fascinating and hidden from plain sight, is the impact of killing the Herald. Despite the extremely low correlation of 0.09 with the dependent variable, applying conditional probabilities yields results as seen in Figure 3: the teams that killed the Herald had statistically significant increases in the average of both total gold and gold difference. The t-tests had 95% CI [-671.9, -512.8] and [-1131.5, -888.8] respectively, and full test results are exhibited in Appendix E.

Figure 3

Gold Difference per Number of Heralds

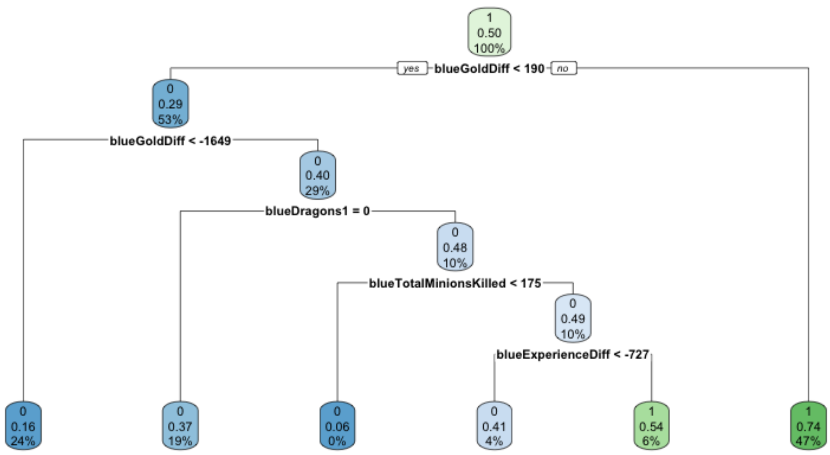


Note: Line represents mean for no Herald and Dashed Line represents mean for one Herald.

After fitting the models, the first insight comes from the one that provided the highest level of accuracy. Reaching 74.31% with 95% CI [72.67%, 75.89%], the 10-fold cross-validation DT model confirmed expectations that iterations helped in providing a better specification for the tree. The tree had six leaves in the end with gold difference being the root node as seen in Figure 4. In unexpected fashion, however, all three NN models lost accuracy when predicting results from out of the train sample. It indicates that the overfit was persistent even for the iterative version, which provided the highest accuracy for the test data with 72.85% and 95% CI [71.18%, 74.47%].

Figure 4

Decision Tree From Model With Highest Accuracy (10-Fold Cross-Validation)



When reflecting about potential actionable insights and valuable outcomes of the research, it is necessary to incorporate two additional lenses to the original one of verifying the feasibility of predicting outcome to build third-party software. The second lens reflects LoL professional teams as winning games increases visibility and directly translates to financial prizes and sponsorship deals. Which metrics and actions should the teams focus on to ensure a win from the performance of the first 10 minutes? One additional lens is Riot as the producer of League of Legends. There is a fine balancing between creating an experience that rewards team for good performance in the beginning of the game and one that eliminates any chance of a comeback. How can Riot leverage game data to balance the game experience?

The first insight for professional teams is taking the next sentences with a grain of salt given that the data stems from high skill games, but not professional level games. That said, the data clearly indicates that focusing on total gold amassed and on increasing the gold gap between teams is key. Therefore, teams should instruct players to focus specifically on that for the first 10 minutes, including on farming well, destroying turrets, and acting on opportunities to deny gold to the enemy team. Additionally, teams should highlight the importance of the Herald, a neutral monster killed in around one fifth of the games and weakly correlated to winning outcomes. Nevertheless, conditional probability demonstrates that average gold is significantly higher if the Herald is killed. Wards should be placed to grant vision, but teams should not be focusing on this attribute in the first 10 minutes.

For Riot, the company should double down on tweaking if the goal is creating pleasant gaming experiences for winners and losers. Since it is possible to predict with over 70% accuracy just from the first 10 minutes, the company could create strategies to delay the transformation from gold to advantages in the game (such as making items more expensive, for example) or could build creative alternatives for losing teams to bridge the gap. The latter is known to be true, as Riot introduced Objective Bounties in late 2021, granting additional gold to the losing team in case the gold difference in the game surpasses a threshold. Curiously, this information also corroborates the relatively higher importance of gold difference.

Lastly, the most important perspective given the focus of this research. For someone interested in pursuing the third-party software route, the results from this research are promising. With data pertaining only to in-game attributes, the models were capable to predict the game result with 70% accuracy. Clear actionable goals follow. As a first step and following past research from other games, incorporating data from champion attributes and from players' past games should result in substantive increases in accuracy. Not only that, but it would also be potentially interesting to collect full game data and explore paths to use to train the data while keeping up to 10 minutes as inputs for the final prediction model. In the long run and with enough data to accurately train, test, and predict past outcomes, third-party companies should then move toward processing real-time data in a mid-term outlook. As the game is tweaked approximately every two weeks by Riot, it is imperative that companies move toward only needing two or three days of data for their third-party software to be current and therefore valuable to players. With 180 million active monthly players throughout the world, the underlying assumption here is that such time horizon would be feasible.

Discussion

This research set out to analyze whether information collected from the first 10 minutes of game could accurately predict the winning team in League of Legends matches. After fitting six models from two different approaches, it reached an accuracy level of 74.31% with 95% CI [72.67%, 75.89%]. This result demonstrates the power of predictive analysis, even when done in preliminary studies such as this. However, for the use case of building a product that would recommend actions to players, a higher accuracy level would be necessary to generate satisfaction.

The results and actionable insights nevertheless open doors to future works. The implications for research are that there are many similarities between past research conducted for analysis of DotA 2 games and the LoL literature should build from those hypotheses rather than start from scratch. It also implicates that achieving high levels of accuracy, if more data is included, should be a feasible task, especially as newer research in adjacent fields explore real-time data prediction.

For practice, its main implication is showcasing the business opportunity existent from exploring publicly available data. It demonstrates how simple models, both in terms of complexity and of inputs, can achieve meaningful results. For the teams, it illustrates the power of big data analytics to create valuable insights about which are the best predictors of wins and uncover hidden relationships between in-game and pre-game variables. By unlocking the potential of big data analysis, teams can leverage players' expertise from countless hours of dedication couple with trends from all around the world and all different skill levels.

This study, however, has several limitations as mentioned before. It was executed in a short timeframe given the period of the class and relied upon limited knowledge from the researcher. The problem formulation already delimited the focus of the study and the data utilized also had important limitations such as the lack of pre-game variables. It would have been interesting, for example, to understand which champions work well with which other champions and create a team synergy variable. Additionally, an important limitation was the goal of the study, focusing on descriptive statistics and fitting the models rather than dedicating extensive amounts of time to carefully choosing parameters and calibrating algorithms as previous researchers have done.

This research was a valuable learning experiences from a diverse set of angles. First and foremost, it enabled an academic yet interesting perspective to a subject of high personal importance. It showcased the need for hands-on exercises to solidify learnings in Big Data

Analytics. This research also illustrated the value of online forums, open-source projects, and the collaboration within the field. It was only possible to conduct the analysis with the assistance of the classes, readings, and questions answered on Stack Overflow. Finally, it opened the door to predictive analysis and gave the motivation to continue exploring the topic afterwards.

In future works, it is necessary to add complementary data about champions and their relative strength or weakness depending on the patch; historic data about players, their ability with different champions, and their in-game habits; and data about different skill levels, both above in the professional level and below in the low skill levels. One important caveat to these hypotheses is that of data availability. Future work should build on identifying new sources of information. Additionally, it is imperative to dedicate time and effort extensively to improve the fit of the models. Initial paths are understanding why NN models have been overfitted, fixing the issue, and expanding to other models such as LR.

Conclusion

This research set out to investigate whether in-game data from until 10 minutes would be able to successfully predict game outcome in League of Legends. By utilizing a dataset containing over 9k observations, six models were fitted in decision trees and neural networks. In the end, the model with highest accuracy was the 10-fold cross-validation decision tree which reached an accuracy level of 74.31% with 95% CI [72.67%, 75.89%]. Insights were also provided such as the importance of the difference in gold rather than the absolute level of gold, how conditional probability helped uncover a relationship between Heralds and chances of winning that was hidden in the correlation matrix, and how teams must strive to amass gold while still farming well. Actionable outcomes were suggested to those interested in building a product out of game prediction and were similar to the recommendations for future work: increasing the size of the data and the number of variables covered, investing in fine-tuning the models, and learn from past research in similar games. Additionally, recommendations were suggested to Riot Games as the game developer and to professional teams on how to achieve their goals. There are many avenues for the growth of the game prediction field within League of Legends and this study was one step in providing an understanding of the relationship between the variables and their predictive power.

References

Hodge, V., Devlin, S., Sephton, N., Block, F., Cowling, P. I., & Drachen, A. (2021). Win Prediction in Multiplayer Esports: Live Professional Match Prediction. *IEEE Transactions on Games*, *13*(4), 368–379. https://doi.org/10.1109/TG.2019.2948469

Hodge, V., Devlin, S., Sephton, N., Block, F., Drachen, A., & Cowling, P. (2017). *Win Prediction in Esports: Mixed-Rank Match Prediction in Multi-player Online Battle Arena Games **. https://doi.org/https://doi.org/10.48550/arXiv.1711.06498

Lan Ma, Y. (2020). *League of Legends Diamond Ranked Games (10 min)*. https://www.kaggle.com/datasets/bobbyscience/league-of-legends-diamond-ranked-games-10-min

League of Legends Wiki. (n.d.). *Gold (League of Legends) | League of Legends Wiki | Fandom*. Retrieved November 23, 2022, from https://leagueoflegends.fandom.com/wiki/Gold_(League_of_Legends)

Straits Research. (2022, August 31). *Esports Market Size is projected to reach USD 5.74 Billion by 2030, growing at a CAGR of 21.9%: Straits Research*. https://finance.yahoo.com/news/esports-market-size-projected-reach-161000368.html

Yang, Y., Qin, T., & Lei, Y.-H. (2016). Real-time eSports Match Result Prediction. *30th Conference on Neural Information Processing Systems* .

Appendix A

Appendix A. Map of League of Legends

Appendix B

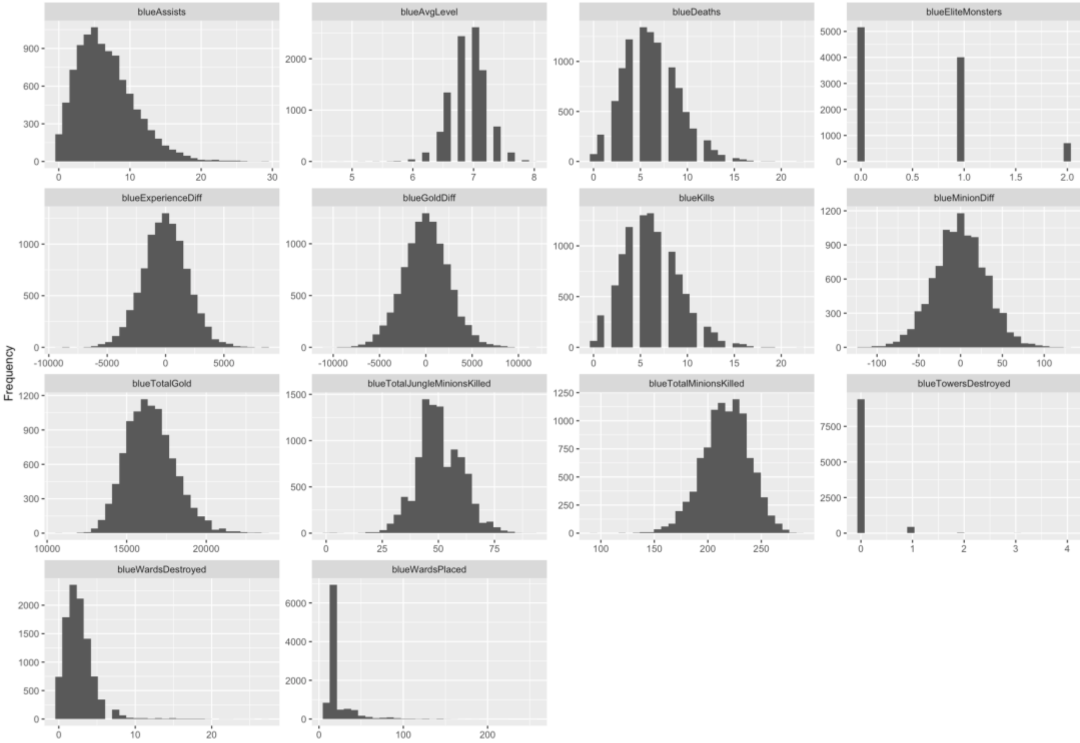


Figure B1. Histogram of Continuous Variables

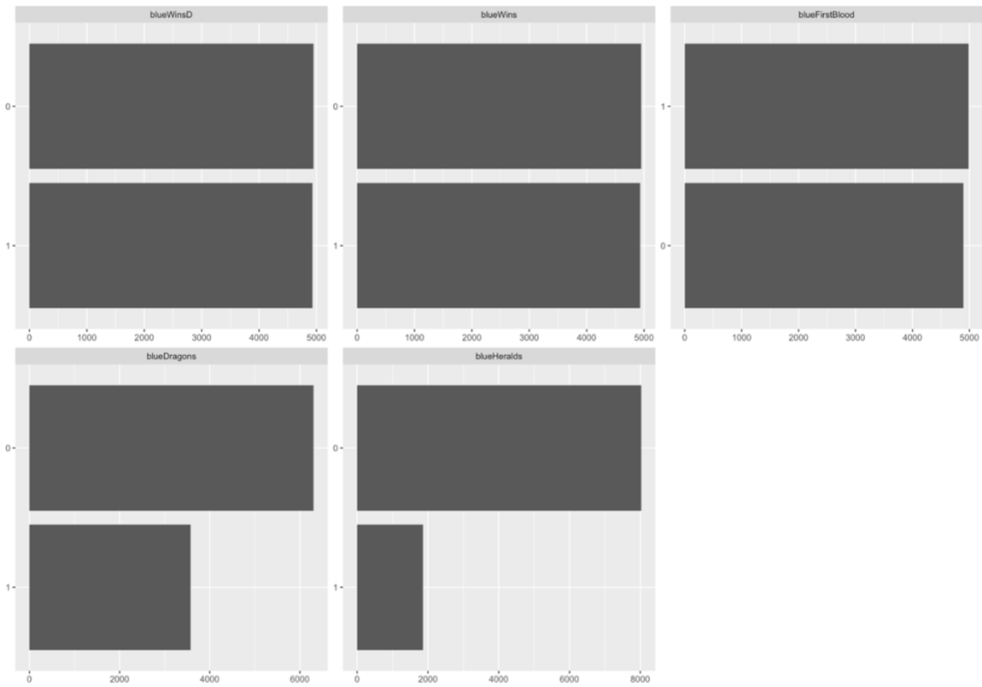


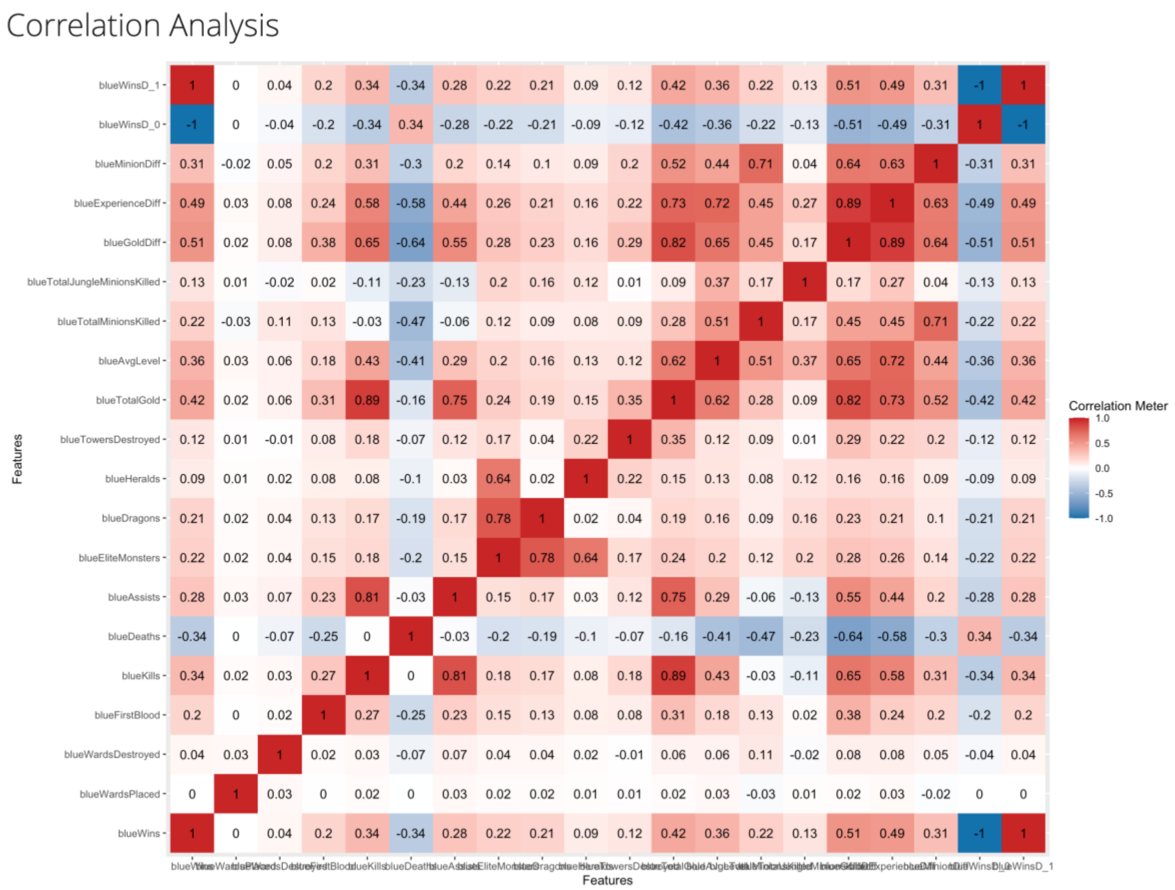
Figure B2. Frequency of Discrete Variables

Appendix C

Statistic	N	Mean	St. Dev.	Min	Max
blueWins	9,598	0.500	0.500	0	1
blueWardsPlaced	9,598	19.905	10.490	5	76
blueWardsDestroyed	9,598	2.824	2.173	0	27
blueFirstBlood	9,598	0.505	0.500	0	1
blueKills	9,598	6.178	3.010	0	22
blueDeaths	9,598	6.127	2.938	0	22
blueAssists	9,598	6.643	4.061	0	29
blueEliteMonsters	9,598	0.550	0.626	0	2
blueDragons	9,598	0.362	0.481	0	1
blueHeralds	9,598	0.188	0.391	0	1
blueTowersDestroyed	9,598	0.051	0.242	0	4
blueTotalGold	9,598	16,497.930	1,532.363	10,730	23,701
blueAvgLevel	9,598	6.915	0.304	4.600	8.000
blueTotalMinionsKilled	9,598	216.763	21.886	90	283
blueTotalJungleMinionsKilled	9,598	50.508	9.905	0	92
blueGoldDiff	9,598	11.892	2,453.875	-10,830	11,467
blueExperienceDiff	9,598	-35.053	1,920.773	-9,333	8,348
blueMinionDiff	9,598	-0.650	30.970	-120	127

Appendix C. Summary of *blue*

Appendix D



Appendix D. Correlation Analysis

Appendix E

```
        Welch Two Sample t-test

data:  blueTotalGold by as.factor(blueHeralds)
t = -14.595, df = 2629, p-value < 2.2e-16
alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
95 percent confidence interval:
 -671.9393 -512.7757
sample estimates:
mean in group 0 mean in group 1
       16386.53        16978.89

> wilcox.test(blueTotalGold ~ as.factor(blueHeralds), data = blue)

        Wilcoxon rank sum test with continuity correction

data:  blueTotalGold by as.factor(blueHeralds)
W = 5509778, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

> t.test(blueGoldDiff ~ as.factor(blueHeralds), data = blue)

        Welch Two Sample t-test

data:  blueGoldDiff by as.factor(blueHeralds)
t = -16.323, df = 2772.9, p-value < 2.2e-16
alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
95 percent confidence interval:
 -1131.4736  -888.7857
sample estimates:
mean in group 0 mean in group 1
      -178.0731        832.0565

> wilcox.test(blueGoldDiff ~ as.factor(blueHeralds), data = blue)

        Wilcoxon rank sum test with continuity correction

data:  blueGoldDiff by as.factor(blueHeralds)
W = 5384191, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

Appendix E. T-Test and Wilcoxon Test for Herald Significance on Total Gold and Gold Difference

Appendix F

Accuracy of the Models

	DT 0	**DT 1**	**DT K-Fold**
Train Data Set Accuracy (IC)	72.22% *(71.11% - 73.29%)*	72.61% *(71.53% - 73.67%)*	72.72% *(71.63% - 73.78%)*
Test Data Set Accuracy (IC)	73.61% *(71.96% - 75.21%)*	73.61% *(71.96% - 75.21%)*	74.31% *(72.67% - 75.89%)*

Table F1. Accuracy for Decision Tree Models With Train Data and Test Data

	NN 0	**NN 1**	**NN K-Fold**
Train Data Set Accuracy *(IC)*	73.91% *(72.84% - 74.95%)*	75.32% *(74.27% - 76.35%)*	73.73% *(72.66% - 74.78%)*
Test Data Set Accuracy *(IC)*	72.43% *(70.76% - 74.06%)*	71.81% *(70.12% - 73.44%)*	72.85% *(71.18% - 74.47%)*

Table F2. Accuracy for Neural Networks Models With Train Data and Test Data