Predicting League of Legends Match Outcomes

1 Introduction

1.1 Context

In the past few years, video gaming and e-sports have transformed from a small niche into a popular entertainment form across the globe. In the past, e-sports was considered an outsider compared to traditional sporting culture, but it has now indisputably grown into a full outright industry. As e-sports grew increasingly prevalent in mainstream culture, it became apparent that the industry was teeming with lucrative venture opportunities, grabbing the attention of brands, media, and investors alike. Coming from humble origins in arcade gaming, e-sports has burst onto the global stage, projected to rake in $1.8 billion USD of revenue by 2022 (1) and has even been classified as a sport, garnering recognition from the International Olympic Committee. At the forefront of this revolution is Riot Games' League of Legends, a multiplayer online battle arena game with a long-established international competitive scene.

1.2 Aims

Possessing such a degree of economic and cultural significance, the growth of e-sports shows no signs of slowing down anytime soon, and attainment of any insight to winning competitions would provide substantial benefit to numerous stakeholders. For this reason, our group is investigating the key factors that influence a match outcome, and whether it would be possible to predict this with a level of statistical significance at some point before the match's conclusion. We believe possession of such predictions could also make games more interesting to watch, thus affecting the viewership and overall interest in esports.

1.3 Dataset

Our dataset has been sourced from Kaggle and contains several .csv files containing basic statistics for every professional match that took place between 2015 and 2018 (except those played in mainland China) (2).

2 Exploratory Analysis

Through exploratory analysis, we sought to answer the question of identifying the main factors that contribute a team's victory, with our target variable as winning team (and winning time/duration). As a team, we conducted research by playing the game, and became able to grasp which in-game statistics and factors were important. Through this, we managed to obtain a cursory understanding and tentatively identified several prediction variables; these were gold difference, kills, and structures taken. We also gave consideration to 'derivate variables', compounding these with conditions such as location, time, and frequency/sparseness. At this point, a potential plan we devised was splitting each match into time-based data, i.e. a set for each minute of each match.

2.1 Preliminary Exploration

We began our preliminary graphical exploration of a portion of the dataset through the investigation of gold difference and champion kill location, as shown in Figures 1 and 2, to provide us with an initial judgement for the potential of our investigation. Conventional wisdom alludes to the fact that gold difference between the two teams is often the best indicator of win probability, which is why it was selected in this first investigation. Kill location serves as an indicator of where conflict occurs, from which we can infer based on time in the match which nexus is destroyed, and thus which team emerges victorious.

As shown in the Figure 1, matches where gold difference favoured the blue side more often yielded a victory for blue, and vice versa. In Figure 2, we see that as the match progresses, champion kills occur closer to a team's nexus (a structure that if destroyed, ends the match). Based on this data, we postulated that match outcome could be predicted by the distance of recent kills from the blue nexus.

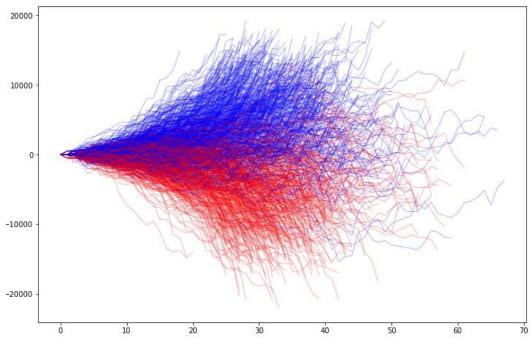

Figure 1: Plot of team gold difference relative to the blue team (vertical axis) against match duration (horizontal axis). Line colour represents which team ended up winning the match.

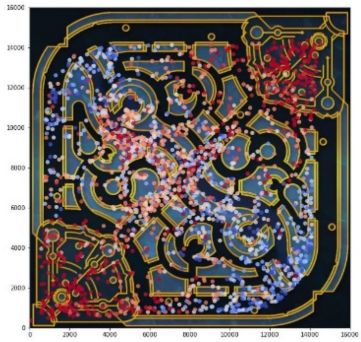

Figure 2: Location of champion kills throughout the match. Blue represents earlier in the match, and red later.

2.2 Initial Logistic Regression Model

With this preliminary exploration completed, we proceeded to undertaking a more formal modelling of the data to aid us in identifying which factors are most pivotal in effecting a victory for either team. As our dependent variable is a binary datum – either a win or loss – we opted to create a logistical regression model to provide us with an odds ratio. We modelled the probability of blue team winning using the predictive variables summarised in Appendix B. Predicated on our intuition and anecdotal experience, we tried to anticipate the variables that would contribute the greatest impact to a match outcome. Across the predictor variables, we hypothesised that gold difference and structures taken/lost to have the greatest impact.

Using this model, we were able to make a prediction for this event under the assumption that a probability of >0.5 is equivalent to a blue team victory. For the initial exploratory model, we sampled every minute from the first 100 games of our dataset, to construct a derived dataset containing 3759 game state instances. These game states excluded the kill location variable as the data for each game was too sparse to construct meaningful minute by minute data. Additionally, because kills rarely occurred before 10 minutes into a game, we were unable to fill the gaps in the data before the first kill occurred, though we still wanted to be able to make predictions for this time. Excluding kill location as a predictor variable, we then performed an 80/20 train-test split on this exploratory dataset, arriving at the model results shown in Figures 3 and 4 below.

	coefficient	std	p-value	[0.025	0.975]
intercept	0.279	0.051	0.000	0.178	0.380
Unnamed: 0	0.000	0.048	1.000	-0.094	0.094
Minute	-0.196	0.807	0.808	-1.778	1.386
Blue kills	0.415	0.202	0.040	0.019	0.811
Blue deaths	-1.439	0.210	0.000	-1.851	-1.028
Blue kill freq	0.238	0.101	0.018	0.041	0.436
Blue death freq	-0.222	0.105	0.035	-0.428	-0.016
Red structures taken	0.174	0.192	0.365	-0.202	0.550
Blue structures lost	1.033	0.187	0.000	0.665	1.400
Blue structure freq	0.000	0.103	1.000	-0.203	0.203
Red structure freq	-0.175	0.095	0.064	-0.361	0.010
Blue gold	0.000	0.987	1.000	-1.935	1.935
Gold diff	1.967	0.224	0.000	1.526	2.407
Delta gold diff	-0.026	0.111	0.816	-0.244	0.192

```
Confusion Matrix (total:3007)    Accuracy:        0.774
  TP: 1459 | FN: 205
  FP: 474 | TN: 869
```

Figure 3: Exploratory logistic regression model, training split

	coefficient	std	p-value	[0.025	0.975]
intercept	0.279	0.105	0.008	0.072	0.486
Unnamed: 0	0.000	0.088	1.000	-0.172	0.172
Minute	-0.196	1.627	0.904	-3.390	2.997
Blue kills	0.415	0.429	0.333	-0.428	1.258
Blue deaths	-1.439	0.438	0.001	-2.300	-0.579
Blue kill freq	0.238	0.222	0.284	-0.198	0.675
Blue death freq	-0.222	0.208	0.287	-0.631	0.187
Red structures taken	0.174	0.383	0.650	-0.578	0.925
Blue structures lost	1.033	0.412	0.012	0.223	1.842
Blue structure freq	0.000	0.201	1.000	-0.394	0.394
Red structure freq	-0.175	0.223	0.433	-0.613	0.263
Blue gold	0.000	2.012	1.000	-3.949	3.949
Gold diff	1.967	0.473	0.000	1.039	2.894
Delta gold diff	-0.026	0.235	0.913	-0.487	0.435

```
Confusion Matrix (total:752)     Accuracy:        0.758
  TP: 370 | FN: 57
  FP: 125 | TN: 200
```

Figure 4: Exploratory logistic regression model, test split

From the training model summary, we see that the variables that are most significant in determining a match outcome for the blue team – whether favourable or adverse – blue deaths, blue structures lost, and gold difference; each with a p-value of <0.0005. This was followed by blue kill frequency, blue death frequency and blue kills. Compared with our initial proposition, structures lost and gold difference demonstrated their expected impacts. However, we overvalued structures taken, and undervalued deaths, kills, and their frequencies. This initial training model yielded an accuracy of 77.4%, alongside a precision and recall of 0.755 and 0.877, respectively.

3 Predictive Analysis

3.1 Logistic Regression Model

With the 13 variables of our match dataset, we could theoretically construct over 8000 different models. However rather than testing very model by brute force, we chose to utilise a LASSO regression to eliminate insignificant variables. This was used for both the exploratory model (with a penalty of C = 0.1) and our predictive logistic regression model. We decided to test various penalties, starting out with a large penalty, (C = 0.01) resulting in a model only predicting based on gold difference and delta gold difference, while still obtaining accuracies of around 65%. In the end, we settled on a lasso select with a penalty of C = 0.2, as this provided a 0.1% increase in accuracy while training. Taking 21,577 instances, we once again carried out an 80/20 train-test split and constructed a logistic regression on our training split as above before verifying its reliability on the test set.

From Figure 5, we see that the lasso selection has only deemed the blue structures lost variable superfluous, as opposed to blue structure frequency and blue gold for the exploratory model. From the test results in Figure 6, the most significant variables are now gold difference, and blue kill and death frequency, though there was also a drop in accuracy from 0.774 in the exploratory model to 0.724, However, it is worth noting that this value increases for the test split, indicating there was no overfitting.

	coefficient	std	p-value	[0.025	0.975]
intercept	0.073	0.019	0.000	0.036	0.110
Minute	-0.232	0.179	0.196	-0.583	0.120
Blue kills	0.011	0.062	0.854	-0.109	0.132
Blue deaths	-0.343	0.063	0.000	-0.466	-0.219
Blue kill freq	0.350	0.037	0.000	0.279	0.422
Blue death freq	-0.335	0.037	0.000	-0.407	-0.263
Red structures taken	-0.023	0.062	0.712	-0.145	0.099
Blue structures lost	0.000	0.058	1.000	-0.113	0.113
Blue structure freq	0.087	0.036	0.017	0.016	0.158
Red structure freq	-0.104	0.035	0.003	-0.173	-0.035
Blue gold	0.441	0.208	0.034	0.034	0.848
Gold diff	1.269	0.065	0.000	1.140	1.397
Delta gold diff	0.035	0.040	0.392	-0.045	0.114

Confusion Matrix (total:17261) Accuracy: 0.724
TP: 7320 | FN: 1657
FP: 3109 | TN: 5175

Figure 5: Predictive logistic regression model, training split

	coefficient	std	p-value	[0.025	0.975]
intercept	0.073	0.038	0.052	-0.001	0.148
Minute	-0.232	0.356	0.515	-0.930	0.467
Blue kills	0.011	0.120	0.925	-0.224	0.247
Blue deaths	-0.343	0.124	0.006	-0.587	-0.099
Blue kill freq	0.350	0.071	0.000	0.212	0.489
Blue death freq	-0.335	0.073	0.000	-0.479	-0.191
Red structures taken	-0.023	0.124	0.853	-0.265	0.219
Blue structures lost	0.000	0.117	1.000	-0.229	0.229
Blue structure freq	0.087	0.072	0.230	-0.055	0.229
Red structure freq	-0.104	0.070	0.137	-0.242	0.033
Blue gold	0.441	0.406	0.277	-0.355	1.237
Gold diff	1.269	0.131	0.000	1.011	1.526
Delta gold diff	0.035	0.081	0.668	-0.123	0.193

Confusion Matrix (total:4316) Accuracy: 0.735
TP: 1867 | FN: 405
FP: 739 | TN: 1305

Figure 6: Predictive logistic regression model, test split

A limitation of breaking the up data into minute chunks, and in using the test-train split function, is that each segment of every game becomes scrambled together, and we cannot necessarily understand how it would function in a real-world scenario. It would be exceedingly more intuitive to visualise how our model's prediction evolves over the course of single matches, to understand how quickly it could be used to determine which team will win. We randomly extracted three matches from our dataset where our model: makes a virtually perfect prediction, makes a mostly accurate prediction, and one where the predicted victor fluctuates wildly. For each of these, there is a chart with the raw true/false data, and one with the numerical prediction data. Figures 7 and 8 show two of these plots, while the rest are in Appendix C.

For a match in which the winner can be determined relatively early, which often is typical of professional matches where players make fewer mistakes, our model is highly accurate.

Figure 8 depicts an output that continually alternates between predicting a blue or a red victory throughout the match. We concluded that it would be matches like these where our inaccuracy drop originates from. It is also interesting to note how despite the gold difference shown is typically in the red team's favour, our model still predicts a blue victory more times than red overall. This highlights how there are more factors than just gold difference influencing the predictions made by our model.

Having said that, if one were to solely examine the green line in in Figure 8, which tracks the cumulative average predictions, they could still conclude that the blue team will eventually win.

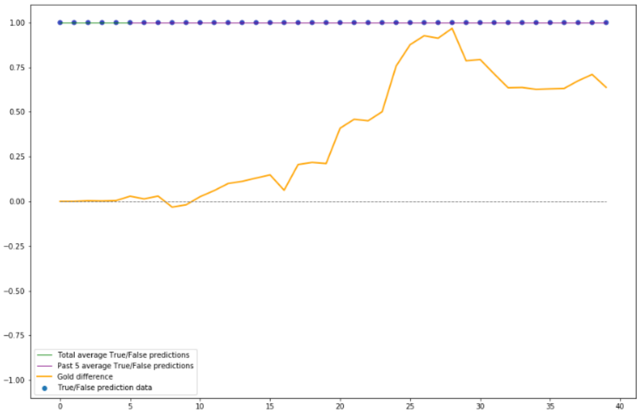

Figure 7a: Our model exhibiting perfect accuracy through one match. Match time (x-axis) plotted against various numerical data samples normalised for a single axis spanning -1 to 1. (Shown in the legend)

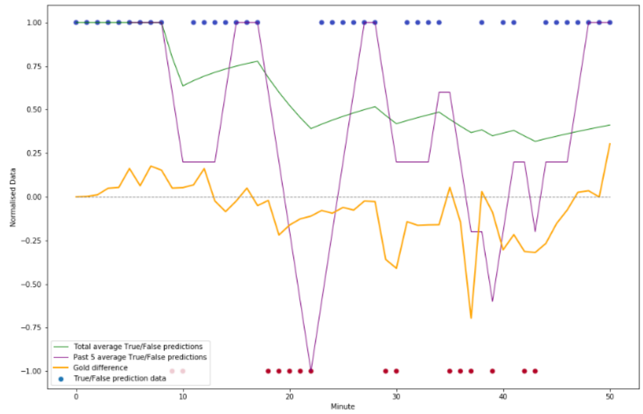

Figure 8a: Our model's predictions (raw true/false) for an extremely close game, where no clear winner was visible till the end. Note how the predictions flick between both teams.

It is interesting to note that for both Figures 7 and 8, the model starts from minute zero predicting a blue victory. This shows that the model has noticed a slight inherent bias in the data, whereby blue typically won more of the matches in our dataset (54.5%).

3.2 Other Model Types

We tested two other types of classification models to compare them against the performance of our logistic regression model: a decision tree using *sklearn.tree.DecisionTreeClassifier*, and a neural net using *sklearn.neural_network MLPClassifier*. Feature correlation testing was also undertaken, plotted in a seaborn heatmap (in Appendix E).

3.2.1 Decision Tree

A variety of decision trees were tested to see how much of an impact the size of the tree had on accuracy. Given the results of the logistic regression model, the conclusion was not surprising: the first node in the tree was always gold difference and subsequent nodes had minimal impact. Some other items of note are:

- The highest accuracy scores obtained for training and testing were 0.757 and 0.734, respectively.
- An extremely simplified model was discovered with an accuracy of roughly 0.72 on both test and training: *if the gold difference at any given point (blue minus red) is less than –65.5, blue is more likely to win (where False is a blue win)*. This proves just how important the gold difference is as a factor for match outcome, and also the bias favouring blue to win.

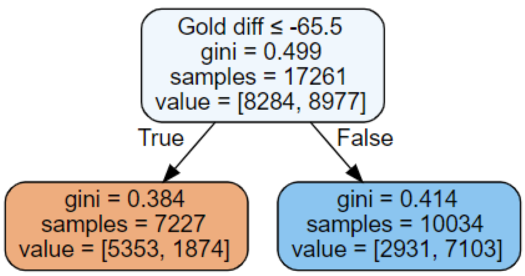

Figure 9: Small decision tree with an accuracy of 0.729 on testing

- By adjusting the *min_impurity_decrease* of the *DecisionTreeClassifier* model, we can adjust the size of the trees. Presented in Appendix D are three sizes of decision trees: 'small', 'large', 'comically large', with their corresponding accuracy. This shows that once a tree has progressed passed the gold difference, its accuracy does not improve significantly. Increasing the number of layers past the 'comically large' size results in overfitting (training accuracy goes up, but testing accuracy goes down).

3.2.2 Neural Net

For fun, several neural-net-based classifiers were also generated to see if there were more complex relationships between predictor variables. Some items of note are:

- No accuracy over 0.73 was ever obtained.
- “Autoencoder” style hidden layer setups seemed to work the best, implying that the outcome is related to just a few parameters.
- A model with 1 hidden layer containing 2 neurons was able to obtain a score of 0.72, further confirming the previous implication.
- Most MLP models exhibited a high recall score compared to other model types (0.85+).

3.3.3 Correlation Matrix

Appendix E shows the correlation between the different predictor variables in our dataset. Things to note are:

- All items except Gold difference and Delta gold difference have a strong, positive correlation to time (Minute)
- Nearly all items have a strong positive correlation to each other, likely because of the last point (as a match progresses, all values increase except the differences)
- Most ‘bad’ events that happen to the blue team (deaths, structures lost) had a negative correlation to the gold difference (and vice versa) as these events mean a gold gain for the red team. Note the positive -> negative -> positive etc bands on the right-hand side over the two gold difference variables.

The final point confirms what our models ends up predicting: the gold difference ‘encodes’ a significant amount of information about who is currently winning a match.

4 Conclusion

With a testing accuracy of 73.5% the logistic regression model was the most accurate model, with the ‘comically large’ decision tree trailing by 0.1%. As predicted initially, gold difference was in fact the most decisive factor in determining a match outcome. However, structures taken and lost did not have as high an impact as we expected, and instead kill and death frequency seemingly took their place. It is important to bear in mind that this data used was solely from professional games, and it is unlikely for the model to be as accurate for casual games between regular players. We anticipate that the predictions would sway more between the two teams as amateur players tend to make more mistakes and are less able to exploit advantages that would otherwise be pivotal in professional matches. Furthermore, all models achieved an accuracy of about 73% in testing. This is likely because they were all trained on the same predictor variables and our data relies on human interaction, which is inherently difficult to predict. Additionally, the models had to make predictions early into the game where there is little to no disparity between the teams in terms of all predictor variables, reducing the accuracy. In conclusion, the ability to predict which team might win a match depends very strongly on how contested the match is. In games where a winner is immediately clear from the model predictions, this result would be likely just as apparent to human spectators. However, the model could still serve some functionality in providing a quantified percentage chance of a team winning at a given moment.

Appendix A: References

1) Soto Reyes M. *Esports Ecosystem Report 2020: The key industry players and trends growing the esports market which is on track to surpass $1.5B by 2023*. Available from: https://www.businessinsider.com/esports-ecosystem-market-report [Accessed December 2020]
2) Ephron, C (2018). *League of Legends*. Version 7. Dataset. https://www.kaggle.com/chuckephron/leagueoflegends

Appendix B: Considered Variables

Table 1: Explanation of variables considered and the justification for how they contribute to match result

Variable	Description	Justification
Minute	Time from start of the match.	Different predictor variables will potentially be weighted differently depending on how far the match has progressed.
Blue kills	Cumulative number of kills made by the blue team.	Kills earn you gold and take the opponent out of the match for a certain amount of time, giving your team an advantage.
Blue deaths	Cumulative number of deaths attributed to the blue team.	*If the blue team is dying more, they are probably losing* -T.G. 2020
Blue kill frequency	Blue kills per minute over the past 5 minutes.	As a match progresses there usually is an increase in the frequency of kills and deaths. If the blue team has a higher kill frequency, they are more likely to win. Conversely, if they have a higher death frequency, they are more likely to lose.
Blue death frequency	Blue deaths per minute over the past 5 minutes.	
Red structures taken	Cumulative number of red structures destroyed by the blue team.	Structures are the barriers protecting the main objective, they give a good indication of how close a team is to the enemy team's nexus. The more structures a team captures, the more likely they are to win.
Blue structures lost	Cumulative number of blue structures destroyed by the red team.	
Blue structure frequency	Structures destroyed by the blue team per minute, over the past five minutes.	As a match progresses there is an increase in the frequency of structures being destroyed and captured. If the blue team has a higher structure frequency, they are more likely to win. Conversely, if the red team has a higher structure frequency, the blue team is more likely to lose.
Red structure frequency	Structures destroyed by the red team per minute, over the past five minutes.	
Blue gold	Cumulative gold earned by the blue team.	Gold allows the players to buy items, which enhances their characters' stats and abilities. Therefore, the gold difference can be said to be analogous to the statistical strength difference between the teams.
Gold difference	Difference in gold earned between the two teams. (Blue minus red)	
Delta gold difference	Change of gold difference between current and previous value.	The derivative of gold difference, i.e. where the strength difference of the teams is heading.
Kill location	Distance of the previous kill from the blue nexus. (Bottom left of the map, see Figure 2)	We expect the average kill location to shift closer towards the side that is losing.

Appendix C: Whole-Match Analysis Graphs

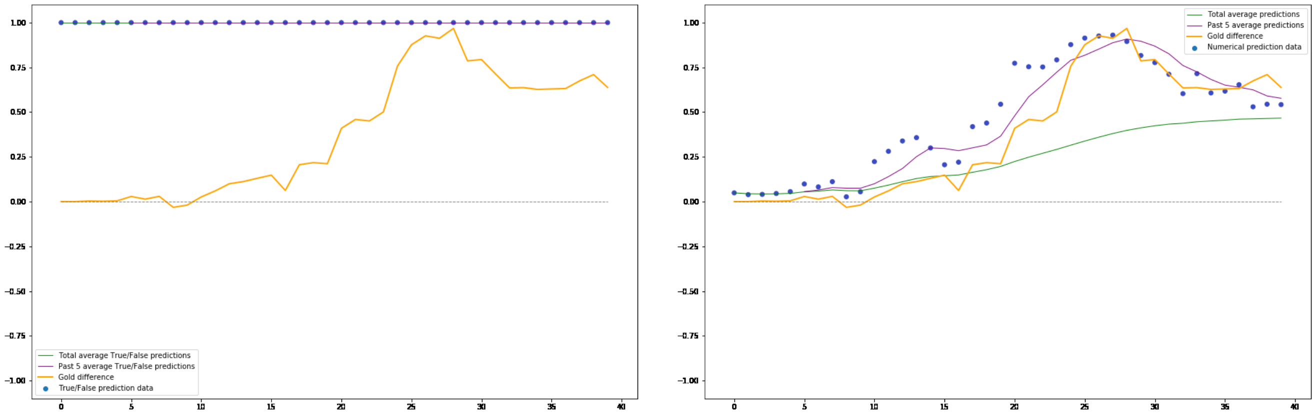

Figure 7a, 7b: Our model exhibiting perfect accuracy through one match. Match time (x-axis) plotted against various data samples normalised for a single axis spanning -1 to 1 (shown in the legend)

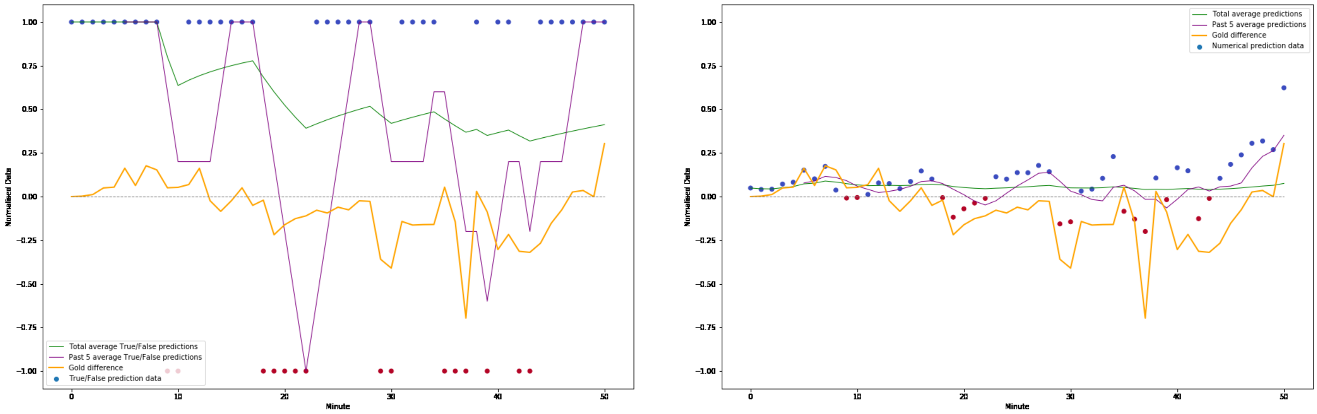

Figure 8a, 8b: Our model's predictions in a close game, where no clear winner was visible till the end. Note how the predictions flick between both teams, depicted as the purple line frequently crossing the horizontal zero axis.

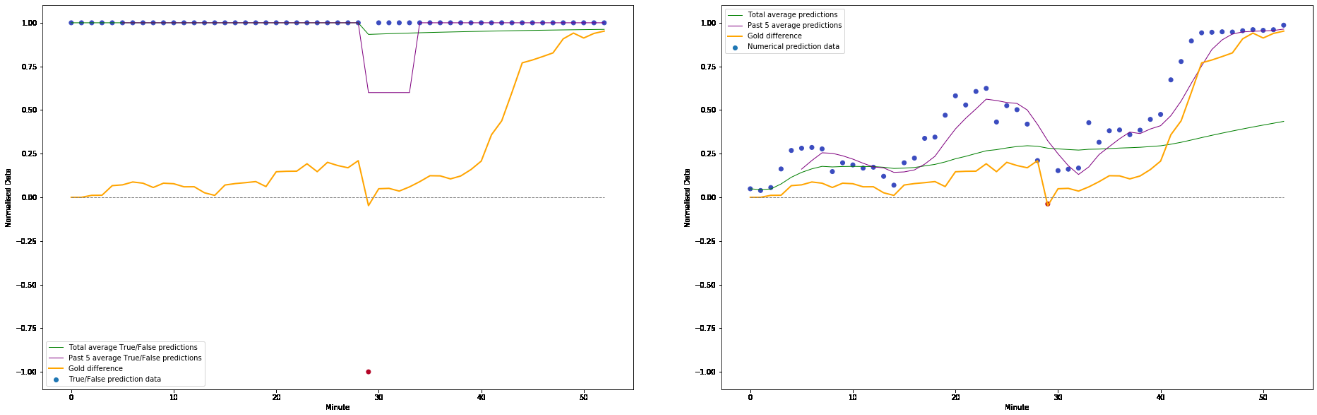

Figure 10a, 10b: Our model exhibiting a mostly accurate, near-perfect prediction through one match

Appendix D: Decision Tree Size Comparisons

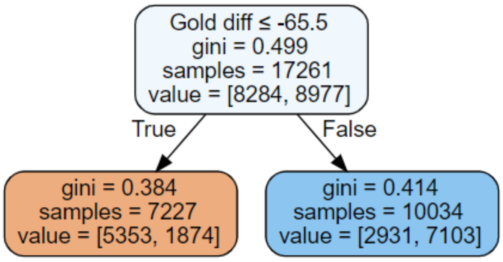

Figure 9: Small decision tree with an accuracy of 0.729 on testing

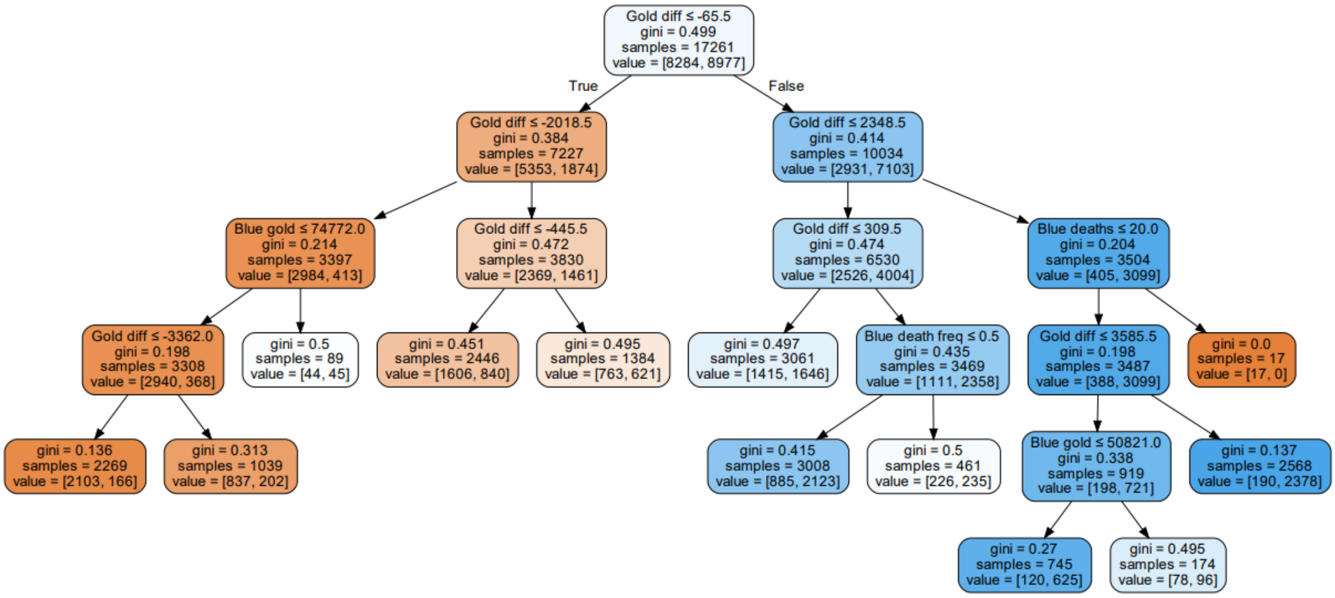

Figure 11: Large decision tree with an accuracy of 0.729 on testing

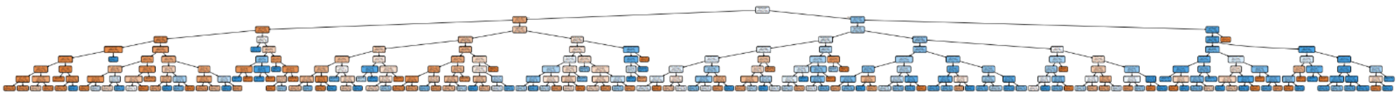

Figure 12: "Comically large" decision tree with an accuracy of 0.734 on testing

Appendix E: Correlation Matrix

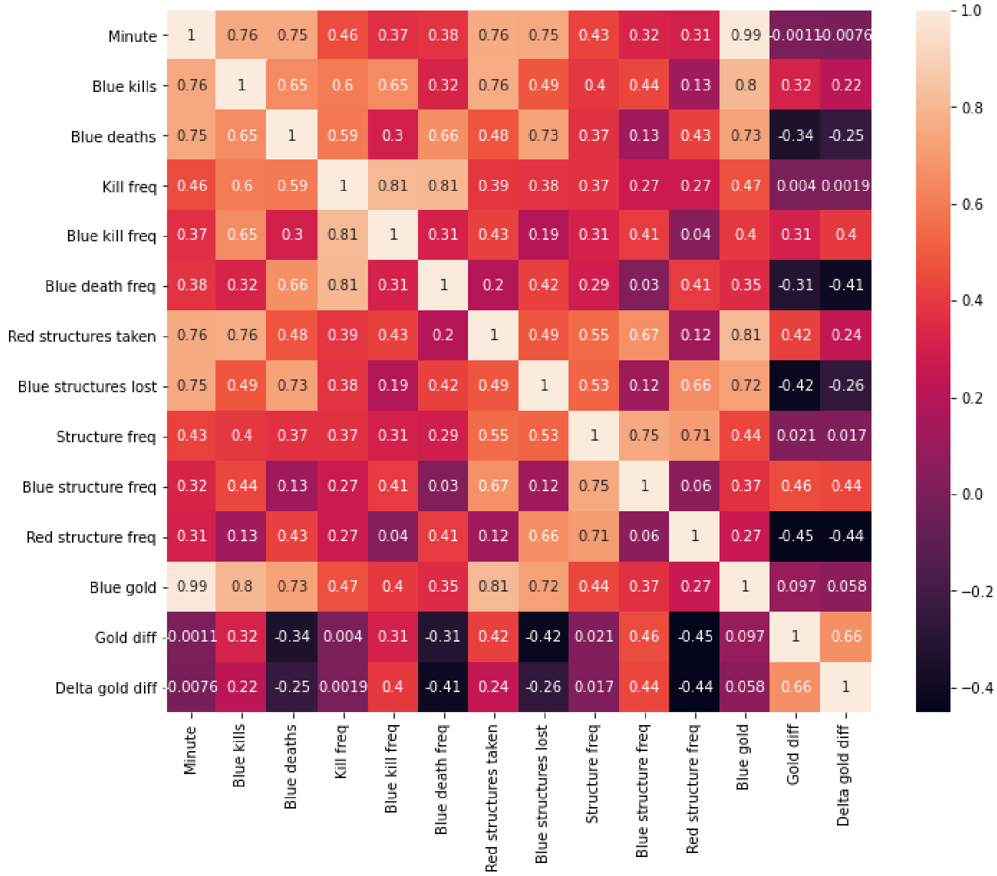

Figure 13: Seaborn correlation matrix