**CSGO: Predicting match winners using historical performance data**

Divye Anand Gupta, Vivek M Avlani, Mayank Kumar Daga, Amjad Hasnain,

Sudeep Kurian, Cassandra(Yachu) Liu, Matthew A. Lanham

Purdue University, Department of Management, 403 W. State Street, West Lafayette, IN 47907

[gupta505@purdue.edu](mailto:gupta505@purdue.edu); [vavlani@purdue.edu](mailto:vavlani@purdue.edu); [mdaga@purdue.edu](mailto:mdaga@purdue.edu); ahasnai@purdue.edu; [kurians@purdue.edu](mailto:kurians@purdue.edu); [liu1498@purdue.edu](mailto:liu1498@purdue.edu); lanhamm@purdue.edu

# 1. ABSTRACT

In this project, we conduct predictive analysis to estimate the Counter Strike: Global Offensive match results. According to Global 2018 Global Sports Market Report[[1]](#REFERENCES) , the global E-sports audience will reach 380 million this year, made up of 165 million E-sports Enthusiasts and 215 million Occasional Viewers, and the industry will reach $1.4 million by 2020. The CS: GO is one of the most popular E-sports which owns 29% of market share, therefore we would like to research in this game to get analytical findings. We used R language with ELO approach, five points moving average and feature engineering to process our data, then we utilized Machine Learning algorithm—Gradient Boosted Decision Tree, Random Forest and Linear Discriminant Regression—to train our models. Finally, we test the models to evaluate the accuracy. Our model is able to predict the match results around 70%.

# 2. INTRODUCTION

Esports or Electronic sports is a competition in which video games are played in tournaments, and players grouped within team compete against each other. Esports has taken the world by storm with the increasing no. of cyber-athletes, for example, the no. of players for the most popular game League of Legends is 100 million whereas for the seventh most popular game Counter-Strike Global Offensive, the no. of players is 11.9 million.[2] The worldwide esport audience has increased from 58 million in 2012 to 165 million in 2018 and is projected to reach 250 million by 2021, and this figure includes enthusiasts only, there are almost equal amount of occasional viewers.[3] This esport environment is supported not only by parent companies such as Valve, & Riot Games, but also by sponsors ranging from as diverse as Intel to Coca-Cola.[4]

Alongside the rise of esports, a new field of analytics has emerged, which seeks to analyze strategies and predict behavior. This new field of analytics is aptly called esport analytics. Matthias Schubert, Anders Drachen, & Tobias Mahlmann define esports analytics as the process of using esports related data, primarily behavioral telemetry but also other sources, to find meaningful patterns and trends in said data, and the communication of these patterns using visualization techniques to assist with decision-making processes.[5] This definition is comprehensive enough to capture run-time game behavior as well as the overall context around the games such as tournaments and competitions.

At a high level, Esports analytics can be considered as a subset of sports analytics and/or game analytics. These three analytics application domains have lot of fundamental principles in common also face similar challenges. Regardless, analytics has evolved into a major component of the overall esports environment. Currently, most professional teams hire analysts to observe competing teams and undercover their patterns and strategies, and build counterstrategies etc., akin to analysts working in other sports.

This field has wider applications in the sense that match-winner prediction would create buzz and enthusiasm among loyal fans. Also, this prediction result would be of interest for organizers who can use this to pit teams, which are equally capable, against each other, thus making the game or tournament more interesting. Finally, the recent judgment by Supreme Court legalizing betting in esports opens avenues for legal betting on match outcomes as well as skin gambling. Thus, betting industry is expected to grow by and hence we see huge potential in Esports Analytics in the days to come.

In this project, the focus is on Multi-Player Online Battle-Arena (MOBAs) game Counter-Strike: Global Offensive (CS: GO). MOBA games have grown to a major proportion of the esports environment in recent times. The term MOBA describes games where two teams of five players each compete within an enclosed virtual structure over a short period of time. For instance, CS: GO pits two teams against each other: the Terrorists and the Counter-Terrorists. Both sides are tasked with eliminating the other while also completing separate objectives, the Terrorists, depending on the game mode, must either plant the bomb or defend the hostages, while the Counter-Terrorists must either prevent the bomb from being planted, defuse the bomb, or rescue the hostages. There are eight game modes, all of which have distinct characteristics specific to that mode.[7]

In this project we attempt to utilize historical performance of players and teams to predict the performance in the current match. We believe that being a video game, player’s skills are translatable from one match to another and there is no strong reason to believe that performance can vary from match to match. The remainder of this paper is organized as follows: A review on the literature on various criteria and methods used for match winner prediction is presented in the next section. Post that we discuss about the datasets we received. In Section 4 the proposed methodology is presented, and the criteria formulation is discussed. In Section 5 we discuss about creating multiple models and testing them. Section 6 outlines the performance of our models. Section 7 concludes the paper with a discussion of the implications of this study, future research directions, and concluding remarks.

# 3. LITERATURE REVIEW

|  |  |  |  |
| --- | --- | --- | --- |
| **Topic** | **Research Paper Name** | **Paper description** | **How is this related to our work and what is different in our work** |
| **#Data Pre-**  **Processing**  **#ELO**  **#Rating system** | **Using ELO ratings for match result prediction in association Football** [**[8]**](#REFERENCES) | The paper incorporates the historical performance of the teams to predict results in football. Each team is given a rating called ELO rating which uses a mathematical algorithm taking past performance to rate every team before their current match.  The ELO system is very beneficial when the data is available for very short spans. | We have match level data and to predict an outcome of a match, we need to analyse the past performance of the team.  ELO Rating system helps us to rank the teams using numerical values and use the parameter in predicting the match winner. We will use the inspiration from this rating system and assign marks to the teams of CS:GO based on their past performance.  We consider this algorithm in our data pre-processing stage. |
| **#Data Pre-**  **processing**  **#Feature**  **Selection**  **#Descriptive**  **Analysis** | **Success in e-Sports: Does Country Matter?** [**[9]**](#REFERENCES) | This paper focuses on testing whether country differences exist in the eSports. The research classified the player statistics into country level. He add additional features related to country and combine all the statistics to a new data.  To examine the social-cultural characteristics influence success in eSports, the researcher interpret the country dummy variable coefficients in the equality as an indicator of the country of origin’s contribution to success in terms of total prizes.  Since variation occurs only across the countries, the researchers decided to use a correlation analysis rather than a regression to pinpoint the important indicators. | This research provides a new perspective to our study. Instead of traditional predicting the match result, this research looks into the association between certain features and match performance.  Even though this study is not well-conducted and contains some indexes which could not reflect the true measure, this study provides some other perspective our study design. We can do some analyses other than prediction analysis.  The research provides new insight on the study design. Our project can focus not only on predictive analysis but also feature and descriptive analysis. |
| **#Data Pre-**  **processing**  **#Mix-rank**  **Approach** | **Win Prediction in Esports: Mixed-Rank Match Prediction in Multi-player Online Battle Arena Games**∗ [[10]](#REFERENCES) | The research paper explores different approaches of using mixed rank datasets to predict the outcome of professional-level e-sports competitions in Online Battle Area games.  The paper discusses different algorithms and their achieved accuracies based on the type of data available to predict the outcome such as in-game statistics or historical time-series data | This paper is similar in what our team is exploring for the CS-GO team. Therefore, it sets the foundation of our approach for predicting team performance for CS:Go tournaments. Our team would strives to incrementally increase the accuracy of prediction from 55% to more.  We can use this approach as a benchmark algorithm for our work |
| **#Model Building**  **#Logistic**  **Regression**  **#Neural**  **Network** | **Real-time eSports Match Result Prediction** [**[11]**](#REFERENCES) | The paper explores how using real-time match data could improve the prediction of the match outcomes at the 40th minute of a Dota 2 game. The researchers had previously used prior (pre-match) features and predicted the outcome but in this paper, they used real-time features in the match as well as more prior features to predict which team is going to win.  They used logistic regression and neural networks for prior modeling – that is using the features at the start of the match. For the real-time data, they explored slicing the time series data by sliding window and training a logistic regression. | The researchers use Machine Learning techniques instead of traditional learning method to improve the model. They are able to predict the real-time eSports match result combining real-time data with prior features to get constantly changing result predictions.  Even though our project will not include real-time data, we can look into their prior features analysis which utilizes logistic regression and neural networks.  In addition, we can investigate on their real-time model to observe which features are strongest associated to the match results.  The study is important for the training method selection and model building. |
| **#Model Building**  **#SVM**  **#Boosting**  **#Lasso**  **#Naives Bayes** | **Predicting the Betting Line in NBA Games** [**[12]**](#REFERENCES) | The research focuses on predicting the outcome of NBA games by incorporating performance fluctuations for teams during the season. | This paper serves as an inspiration to simulate a similar model for CS:GO teams. The methodology discussed for implementing algorithms such as Support Vector Machines, Boosting, Lasso and Naives Bayes for team performance prediction can be learnt from the paper |
| **#Model Building**  **#Portfolio**  **Analysis** | **2018 (Decision Analysis) Kelly betting on horse races with uncertainty in probability estimates** [**[13]**](#REFERENCES) | The paper talks about the limitation of Keyle betting algorithm. The limitation of the algorithm is that it assumes exact knowledge of probabilities and payouts when estimating the returns. The paper allows the user to have room for uncertainties in the probability estimates. The solution is inspired from drawing analogy from horse betting | The second stage of the project is to do a portfolio analysis for the betting companies.  In order to optimize the portfolio return, we need to use Keyle algorithm which helps to maximum return using probability estimations. But to incorporate uncertainties in the environment, the research paper helps us to build an approach to modify the keyle algorithm incorporating uncertainties into our study. |
| **#Model Evaluation**  **#Model Accuracy Assessment** | **Performance of Machine Learning Algorithms in Predicting Game Outcome from Drafts in Dota 2** [**[14]**](#REFERENCES) | The researchers use Machine Learning algorithms—Naive Bayes Classifer, Logistic Regression, and Gradient Boosted Decision Trees to build the model. They also introduced Factorization Machines for best results.  More importantly, they focus on evaluating the model accuracy for testing and benchmarking. | We can refer to this paper about model evaluation.  To access the accuracy of our model is very important, and this article covers two algorithms we have built. |
| **#Future Scope**  **#Temporal**  **Analysis**  **#Run-time**  **Analysis** | **Esports Analytics Through Encounter Detection** [**[15]**](#REFERENCES) | In general, analytics in esports domain is focused on player and team behavior. This paper, however, discusses a technique for match segmentation, into spatiotemporally defined components referred to as encounters. The paper tries to perform analysis using this technique. In such games it is possible that two relevant actions take place on two separate parts of the map in parallel. This means that a technique for breaking down matches into analyzable sections of decisive game play is needed before player and team performance can be evaluated in detail. Also metrics for evaluating team performance during runtime is proposed. | This paper is relevant in the sense that,it reinforces that majority of the esport analytics focuses on the team behavior, the approach which we are taking. Also we don’t have relevant data for the spatio analysis and it can be added to future scope of work.  Once we have the location specific data we can utilize that to do spatio temporal analysis and run-time analysis to predict match winner. |

# 4. DATA

The proprietary dataset was provided by a worldwide sports data provider. There were 12 tables in all containing details about Players, Team, & Match statistics. A brief description of each table is provided in the below table. With our understanding of Business and Data, we could remove 25% of the tables as they had redundant information. Of the remaining tables, we observed relationship among Players, Team, and Match data and consolidated them into one final dataset. The final dataset had approximately 40 features and around 2500 observations.

|  |  |  |
| --- | --- | --- |
| **Table Name** | **No. Of Features** | **Table Contents** |
| Match-odds | 5 | Provides probabilities of winning match b/w home and away team |
| Match\_Scores | 7 | Provides scores between teams displayed map-wise.  One match can have multiples maps. |
| Match\_Summary | 33 | Provides details of matches, such as teams playing, season, venue etc. |
| Player\_Metadata | 16 | Provides player details such as gender, age, country etc. |
| Player\_Statistics | 15 | Provides historical stats of player (assists, deaths, headshots) mapped with team. |
| Player\_Statistics\_Per\_Map\_In\_Match | 13 | Provides player stats at the map level within a match |
| Player\_Statistics\_Per\_Match | 12 | Provides player stats per match |
| Player\_Team\_Histories | 5 | Provides same data as Match-Odds. Title is misnomer. Table is redundant. |
| Player\_Social\_Medias | 3 | Provides a metric, 'No. Of Socials' for each player. Not applicable for this business problem. |
| Team\_Information | 7 | Provides information such as team country etc. |
| Team\_Statistics\_Per\_Match | 13 | Provides team stats (assists, deaths) per match |
| Team\_Stats\_Per\_Map\_In\_Match | 10 | Provides team stats (assists, deaths) per map in match |

# 5. METHODOLOGY

## 5.1 Data Preparation:

1. **Exploratory Data Analysis**: The data provided contained three levels of information: Match, Player and Team Levels. An initial exploratory data analysis was performed to identify relevant information that may be useful for match results prediction based on past performance of home teams
2. **Feature Engineering**:
   1. **ELO scores calculation:** Inspired by an algorithm for predicting winning probabilities for teams in a soccer game based on historical results, a similar methodology was devised by the team to develop a similar algorithm for the e-Sports dataset. ‘ELO scores’ were assigned to each team based on their previous performances.

γH = ELO ratings after the match for the home team

γA = ELO ratings after the match for the away team

l1H = Score rating of a team after match for home team

l0H = Score rating of the previous match for home team

c, d, k0, ω = tuning parameters used for scaling the ratings

αH = 1 if home team wins, 0 if away team wins

* l0A and l1A for away team are calculated in the same way as home team.
* Tuning parameters are set using hit and trial method.
  1. **5-point moving average game statistic generation:** 5-point moving average scores were calculated for each team that comprised of game statistics such as number of kills, deaths, assists and headshots, headshot percentage, etc.
  2. **Derived feature generation:** Based on the existing features in the existing data, new features that could be significant in the prediction analysis were explored. For example, difference between the home team metrics and away team metrics for each match.

1. **Pre-processing**: All missing values were imputed/ removed based their assessed relevance in predicting probabilities.All numerical features were standardized using Z-score standardization technique before the Data Modeling phase
2. **Data-Partitioning:** Train and test sets were created to carry out the validation set approach for our prediction model. createDataPartition function was used from the Caret library to partition 80% data into train set and rest 20% as test set.

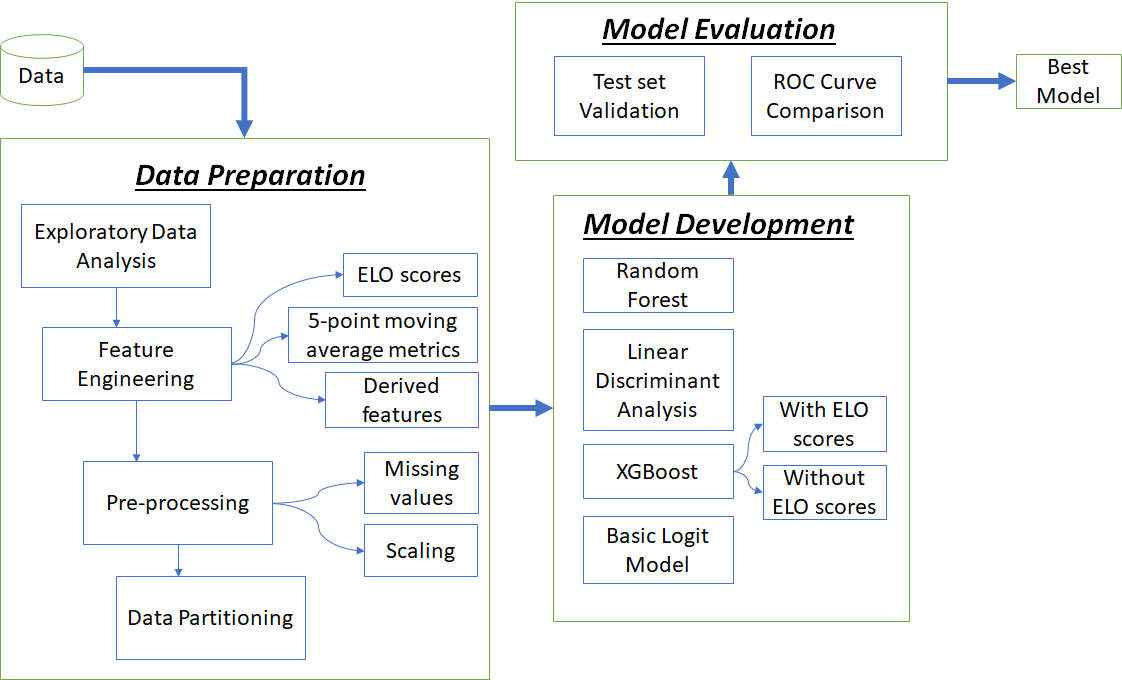
## 5.2 Model Development:

1. **Random Forest Model:** The cleaned data was fed into a Random Forest prediction algorithm to predict the winning probabilities. 5-fold cross validation method was used to obtain a best fit from the given data.
2. **Linear Discriminant Analysis Model:** Before feeding the data into this model, multicollinearity was eliminated from the dataset. Again, a 5-fold cross validation approach was carried out to get the best fit.
3. **XGBoost Algorithm:** This algorithm was chosen to improve the accuracy of the model since it’s a more sophisticated approach in machine learning. Tuning parameter eta was set between 0.3 and 0.4 and the maximum 3 inner layers were specified to fit the model according to the existing train dataset.
4. **Alternate basic models** were prepared based on the features given in the raw data such as probability to win. Some of these models are listed below:
   1. Logit model on just the raw probabilities in the dataset
   2. XGBoost model without ELO scores to compare against XGBoost model with ELO scores

## 5.3 Model Evaluation:

To evaluate our choice of models for predicting probability of home win, the following techniques were performed:

1. **Test dataset validation:** All the trained models were fit into the test dataset to find the candidate models and eliminate models that over-fit on the train-set data
2. **ROC Curve:** ROC curves were plotted for each of our candidate models to find which is the best model to predict home win on the test datasets. Since our model is a perfect example of classification problem, therefore ROC/ AUC values were used as a metric to identify the best model among the existing models.

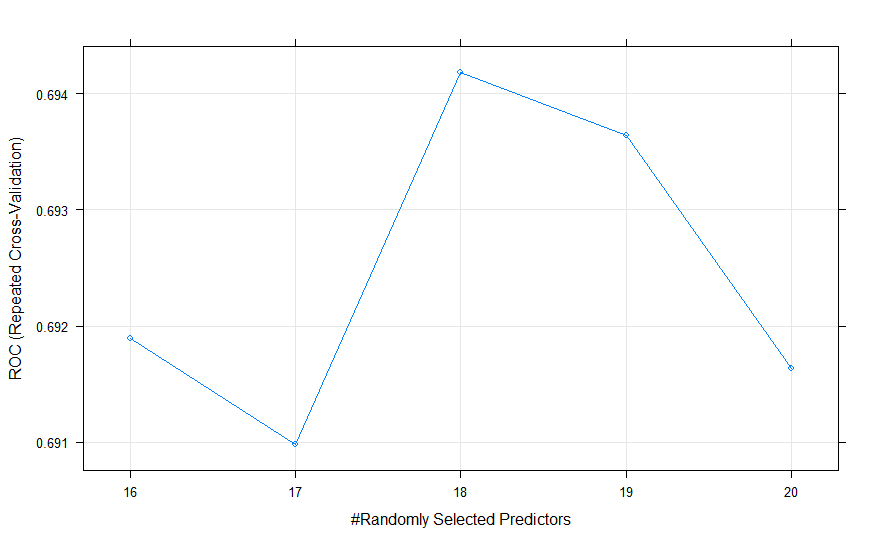
**Figure 5.1:** Analytics Workflow

# 6. MODEL(s)

Since the problem involved doing predictions for a binary classification problem, we could narrow down to a set of models which are optimized for this. We considered random forest, logistic regression, linear discriminant analysis, and xgboost. To implement these algorithms, we utilized the caret package in R which provides a common interface for a variety of modeling algorithms. The rationale behind choosing random forest was that it is a robust model which can handle multicollinearity very well. It was the baseline performance against which we judged the performance of the other models.

6.1.1 Random forest: Random forest grows multiple decision trees and when we need to predict the class of a data point, we pass it to each of the trees which independently predict the class. The final prediction is obtained after each tree votes for a particular class. The random forest would predict the class which has the most votes from these individual trees. Random forests are generally highly accurate, resistant to overfitting, able to handle categorical data very well and easy to tune. These make them ideal as an accurate prediction baseline. Often, they may outperform many algorithms even without the need to tune them. The only tuning parameter that we may need to tune is the number of columns it should consider at each split of the decision node.

Tuning parameter used: *mtry = 18*



**Figure 6.1:** Showing the results of tuning the parameter for random forest. We get the best ROC at 18

6.1.2 Linear discriminant analysis: For linear discriminant analysis, we model the distribution of predictor X in each of the class variables. Then we use Bayes theorem to get the probability of

Linear discriminant analysis is popular when we have multi-class prediction problems but can also be used in 2 class classification problems. There are no tuning parameters for an LDA in the caret package.

6.1.3 Logistic regression: Here we model the probabilities using the logit function where we transform the linear relation using a link function like logit. This transforms the equation into something which ranges from 0 to 1 which is where want our probabilities to lie. The equation below shows us that function which we can use to get the log odds of the probabilities:



In the next equation we see how the probability itself is directly involved with dependent variables:



We are only using the logistic regression for the base model which only contains the betting probabilities.

6.1.4 XGBoost: This is a new framework which falls under the family of gradient boosting algorithms. It can handle parallel tree boosting which helps it yield accurate and quick results. Although very powerful, it may be a tough model to train because it requires a lot of parameter tuning to get it working just right. The most common parameters are general parameters which specify whether we should use tree based boosting or linear boosting, booster parameters, learning task parameters and command line parameters. This method often yields the best accuracy in prediction challenges in Kaggle.

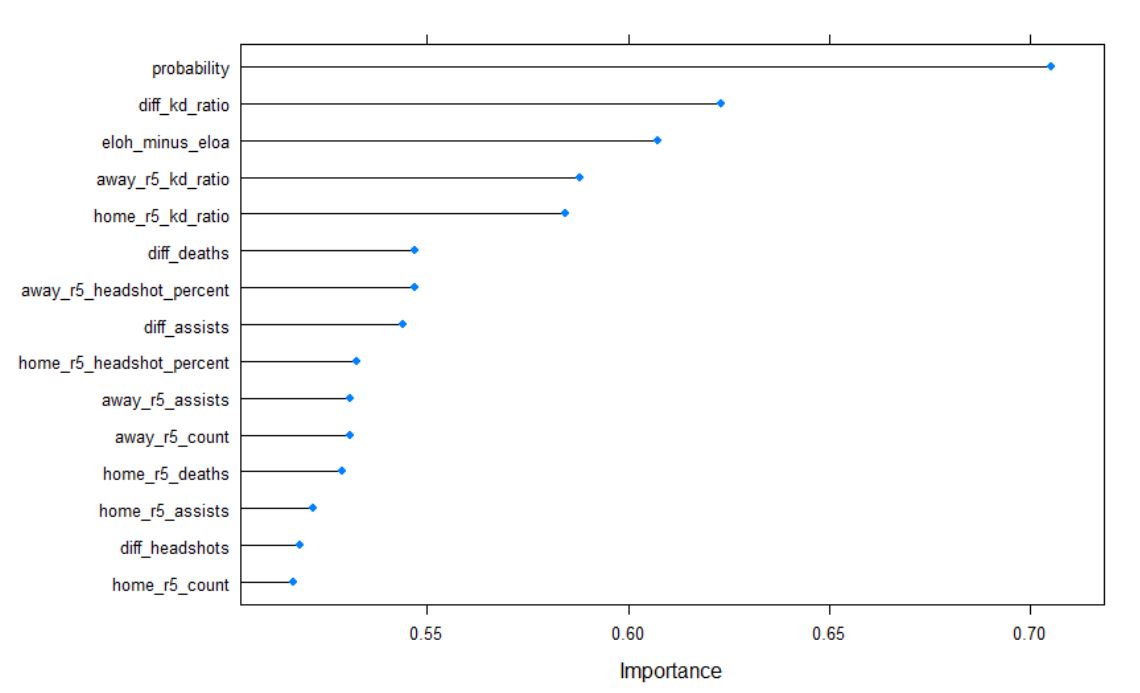
Tuning parameter used were *nrounds = 50, max\_depth = 3, eta = 0.3, gamma = 0, colsample\_bytree =*

*0.8, min\_child\_weight = 1 and subsample = 0.75*

We chose our tuning parameters because it is giving us consistently better results across the index.

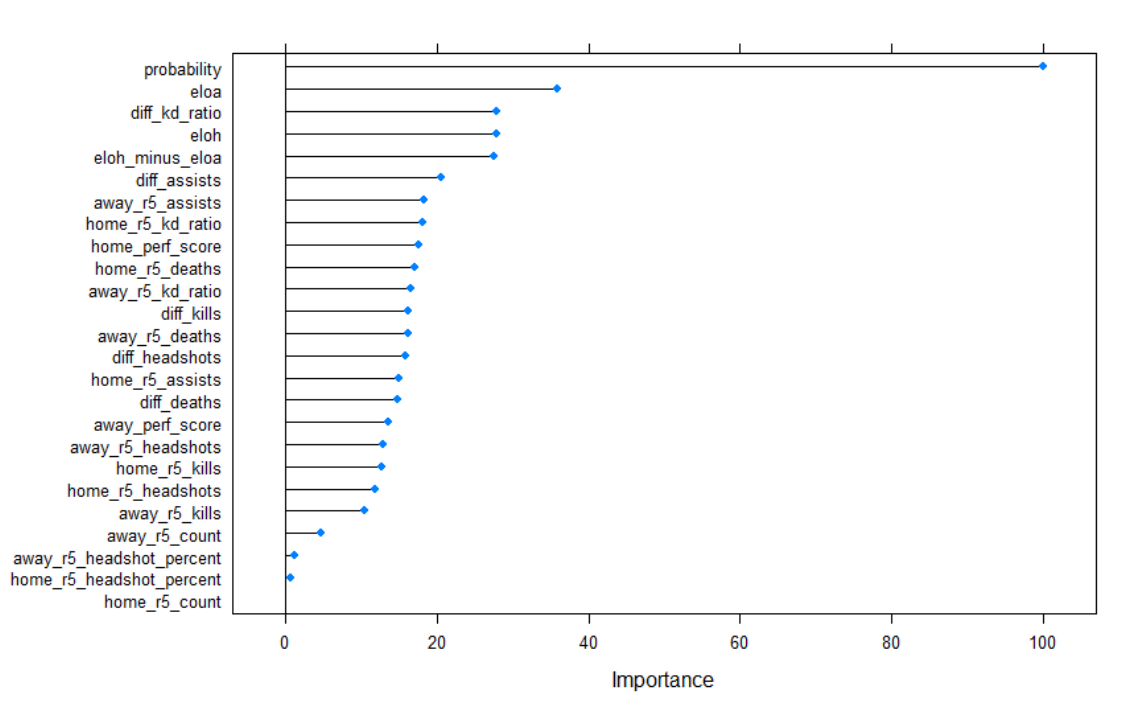
## 6.2 Identifying the important variables through modeling

The most importance variables in prediction are probability, difference in kills to death ratio (Home team – Away team), ELO home – ELO away. The figure which follows shows 16 of the most important variables arranged in descending order of their relative importance:



**Figure 6.2.1:** Variable importance identified through LDA

### Variable Importance through Random Forest:



**Figure 6.2.2**: Variable importance through random forest

# 7. RESULTS

After using the above models, we used many different performance matrices to evaluate the different models.

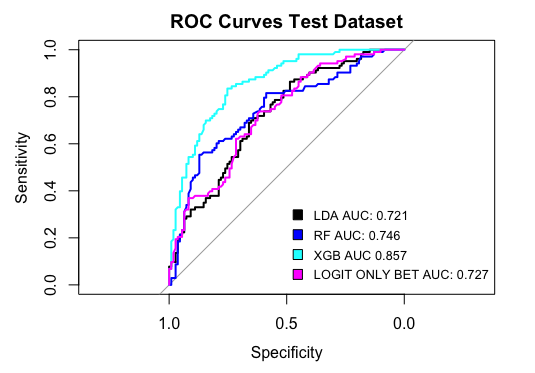
## Cross Validation:

* 1. We used a 5 fold repeated cross validation, repeated 3 times on a 80-20 split for train data. Specified the following parameters in the trainControl function of caret package
     1. **method** = “repeatedcv” (Repeated Cross Validation)
     2. **number** = 5 (k=5, 5-fold cross validation)
     3. **repeats** = 3
     4. **classProbs** = TRUE
     5. **summaryFunction** = twoClassSummary (Since it is a binomial function)
     6. **allowParallel** = TRUE (this allows parallel computing on the 5 datasets)
  2. Performance figures for “ROC” on cross validated train set and test are as follows:

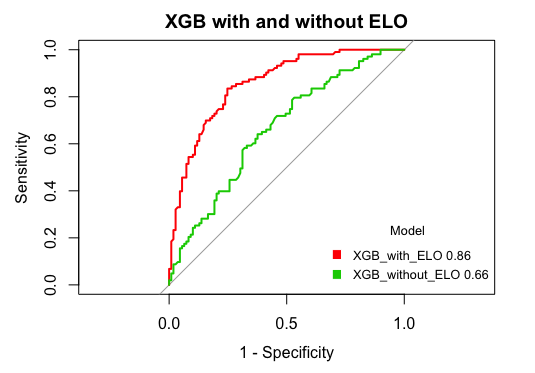
|  |  |  |  |
| --- | --- | --- | --- |
| **Datasets** | **Train CV-ROC** | **Test-Accuracy** | **Test-Balanced Accuracy** |
| **Random Forest** | 0.6819 | 0.6792 | 0.6785 |
| **LDA** | 0.7059 | 0.6557 | 0.6547 |
| **XGBTree** | 0.7166 | 0.7689 | 0.7688 |
| **XGBTree\_w/o\_ELO** | 0.7034 | 0.6179 | 0.6194 |
| **Logit-Bet\_Probs\_Only** | 0.7054 | 0.6651 | 0.6636 |

**Fig 7.1: Table formulating the results for different models**

7.2 AUC-ROC: Area under the ROC curve is a widely used performance measure indicator.



**Fig 7.2.1: Area under ROC curve is highest for XGBTree**



**Fig 7.2.2: ROC curve shows the importance of ELO variable in the study**

## 7.3 Mathews Correlation Coefficient:

|  |  |
| --- | --- |
| **Datasets** | **MCC** |
| Random Forest | 0.3575 |
| LDA | 0.3101 |
| XGBTree | 0.5375 |
| XGBTree\_w/o\_ELO | 0.2396 |
| Logit-Bet\_Probs\_Only | 0.3292 |

**Fig 7.3: Table formulating the results for different models**

## 7.4 Summary of Results:

1. None of the five models discussed above overfits on the train data
2. XGBTree is the best predictor of the home team win among the five models and performs best on cross validated data set and test set. XGB should be used to make decision on whom to bet because it is predicting accurately in 163 out of 212 matches as compared to second best model Random Forest which predicted accurately in 144 matches and much better than logistic model using only match odds to predict accurately in 141 matches.
3. A decision maker can use our model to predict the winner and place his money on the winning team and he will be enjoying the profits in 22 more matches than people who are simply using betting probabilities.
4. The biggest lesson learned from the above modelling methodology is that the present mindset and form of a team is highly correlated to the past performance of the team in matches as the predictions got better after using ELO ratings.

|  |  |  |
| --- | --- | --- |
| **Random Forest** | **Actual Yes** | **Actual No** |
| **Predicted Yes** | 77 | 36 |
| **Predicted No** | 32 | 67 |

**Fig 7.4.1: Table formulating the results for different models**

|  |  |  |
| --- | --- | --- |
| **XGBTree** | **Actual Yes** | **Actual No** |
| **Predicted Yes** | 84 | 24 |
| **Predicted No** | 25 | 79 |

**Fig 7.4.2: Table formulating the results for different models**

|  |  |  |
| --- | --- | --- |
| **Logistic with Bet Odds Only** | **Actual Yes** | **Actual No** |
| **Predicted Yes** | 78 | 40 |
| **Predicted No** | 31 | 63 |

**Fig 7.4.3: Table formulating the results for different models**

# 8. CONCLUSIONS

## 8.1 Summary

Multi-Player Online Battle Arena (MOBA) games such as CS:GO form the major chunk of the digital games being played today, and form a prominent component of the global esports environment. Ability to predict winners in such matches will be of interest for multiple stake-holders. The data can be used to generate enthusiasm among the fans and loyalist. Also, the tournament organizers can utilize the data to pit teams of comparable capability. Finally, the match-winner prediction results will be of interest to betting agencies. The methodology discussed above can be applied to any team-based, competitive esport.

## 8.2 Assumptions and Limitations:

1. We have assumed that historical performance is reflective of present and future performance.
2. The prediction algorithm works for only those team who have had at least one match before. The algorithm removes the teams who have not played any match in the past and hence don’t predict for their matches.
3. The data used had around 1000 rows and the modeling is done on 80:20 split. More richness in the data in terms of quantity can certainly lead to better understanding and predictions.

## 8.3 Future Scope of Study:

1. Spatial-Temporal Analysis can be done in order to consider the other parameters affecting the game results. It consists of broadly two concepts:
   1. Real time data can help us to change predictions in real time. For example, the performance in first 15 minutes of the game can be a good indicator of the final winner. This is called temporal analysis.
   2. In spatial analysis, we consider the location and geographical data from the map and use parameters like ground covered, number of hide spots in the map etc., to support the models to predict better.
2. Analyze the emotions and confidence levels of players during the match using image processing and use the inputs to build a better predictor model.

# 

# REFERENCES

[1] <https://newzoo.com/insights/trend-reports/global-esports-market-report-2018-light/>

[2] <https://www.statista.com/statistics/506923/esports-games-number-players-global/>

[3] <https://www.statista.com/statistics/490480/global-esports-audience-size-viewer-type/>

[4] <http://www.sloansportsconference.com/wp-content/uploads/2016/02/1458.pdf>

[5] <https://www.businessinsider.com/top-esports-sponsors-gaming-sponsorships-2018-1>

[6] <https://www.pcgamer.com/us-supreme-court-opens-the-door-to-legal-esports-gambling/>

[7] <https://en.wikipedia.org/wiki/Counter-Strike:_Global_Offensive>

[8] Arntzen, Halvard and Havttum, Lars Magnus; “Using ELO ratings for match result prediction in

association football”. *International Journal of Forecasting. Volume 26, Issue 3*(2010): 460-470. print.

[9] Parshakov, Petr and Zavertiaeva, Marina, “Success in eSports: Does Country Matter?” (September 18,

2015). Available at SSRN: <https://ssrn.com/abstract=2662343> or <http://dx.doi.org/10.2139/ssrn.2662343>

[10] Hodge, Victoria; Devlin, Sam; Sephton, Nick; Block Florian; Drachen Anders, and Cowling, Peter;

“Win Prediction in Esports: Mixed-Rank Match Prediction in Multi-player Online Battle Arena Games∗”.

Available at: <http://eprints.whiterose.ac.uk/124333/>

[11] Yang, Yifan; Qin, Tian and Lei, Yu-heng; “Real-time eSports Match Result Prediction”(December

2016). *arXiv: 1701.03162v1 [stat.AP]*

[12] Cheng, Bryan; Dade, Kevin; Lipman, Michael and Mills, Cody; “Predicting the Betting Line in NBA

Games”.Available at :http://cs229.stanford.edu/proj2013/ChengDadeLipmanMills-

PredictingTheBettingLineInNBAGames.pdf

[13] Metel, Michael R. “Kelly betting on horse races with uncertainty in probability estimates”. *Decision*

*Analysis, Volume 15 Issue 1 (Mar 2018): 47-52.*

[14] Semenov, Aleksandr; Romov, Peter; Neklyudov, Kirill; Korolev, Sergey and Yashkov Danlil;

“Performance of Machine Learning Algorithms in Predicting Game Outcome from Drafts in Dota 2”.

*Conference Paper, ISSN 1865-0929 (Apr 2016).*

[15] Schubert, Matthais; Drachen, Anders and Mahlmann, Tobias*; “Esports Analytics Through*

*Encounter Detection”. Proceedings of the MIT Sloan Sports Analytics Conference 2016(2016);*

[16] <https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm#intro>

[17] Introduction to Statistical Learning – Hastie, Tibsharani

[18] <http://web.pdx.edu/~newsomj/usp534/ho_logistic.doc>

[19] <https://rdrr.io/cran/caret/man/models.html>

[20] <https://xgboost.readthedocs.io/en/latest/>