The Application of Machine Learning in Predicting the Results of Popular eSports Games: Win Rate Prediction in MOBA and FPS Games

Yulin Peng *

College of Arts and Science, Syracuse University, Syracuse, New York 13210, United States of America

* Corresponding author: ypeng01@syr.edu

Abstract. Electronic Sports (eSports) are one of the most prevalent entertainments in modern society. As eSports is gaining more and more popularity over time, a proper evaluation system that can accurately analyze and predict players' performance is indispensable. Therefore, machine learning is proposed as an effective prediction tool in the studying of data from eSports matches so that eSports teams can utilize it to ameliorate their future accomplishments. Two popular eSports game types (Multiplayer Online Battle Arena and First Person Shooter) are analyzed. Common features they share include communication, equipment, economy, and so on. Correct hero pick is the most predominant factor in MOBA games. Research shows that machine learning models like Logistic Regression, Support Vector Machine, and Neural Networks can effectively combine hero selection data to predict the outcome of a game. On the other hand, the most important factor in FPS games is the in-game equipment each round. Additional research proves that Logistic Regression, K-Nearest Neighbors, Random Forest, and Neural Networks are the most powerful models for predicting the team win rate. This paper can serve as an instruction for applying machine learning methods to making predictions in eSports games. Meanwhile, it can provide available techniques to help professional eSports teams improve their competitive level and future in-game understanding as well.

Keywords: Machine Learning, Classification, Neural Network, eSports, Win rate prediction.

1. Introduction

Electronic sports (eSports), which mostly refers to competitive gaming using video devices and computers under well-established rules, has bloomed within the past three decades and has been one of the most popular entertainments in society so far. A survey from Limelight Networks indicates that among male teenagers, watching online video gaming is far more preferred than watching traditional sports, and this trend will probably increase further as technology improves [1]. With the high prevalence it possesses, esports competitions attract a considerable number of sponsorships. The prize pool of Dota 2 's championship tournament once reached 40 million dollars, and Fortnite's World Cup also has a prize pool of over 30 million dollars [2]. The basic elements of an eSports tournament contain three parts: competitors—namely the eSports clubs and the players, officials—the organization that hosts the competition, provides judges and commentary groups, and millions of audiences. Same as any traditional sport, the outcome of an eSports competition depends on the performance of each team's players—the more they outplay their opponents, the more possibilities they can win, which means more prize money for the whole eSports club. Therefore, in order for the eSports clubs as well as the commentary table to improve their future performance or understanding of a competition, an appropriate system for analyzing and predicting players' performance is essential.

Under this context, machine learning is brought to attention. Machine learning by definition is a type of problem-solving method where the system can learn patterns that are specific to the problem from the training data, and then build relevant models for predicting and categorizing after analyzing the data [3]. It is a highly efficient tool in making accurate predictions especially when the patterns of the elements in the subject are unclear or undefined, which has been proven efficient on tasks such as making predictions on bankruptcy, weather forecasting, or neurosurgical outcome [4-6]. In the

context of video gaming, most of the information during a match is dynamic and hidden behind players' actions since elements like the interaction between the character picked for both sides of the kind of strategy each team plays are hard to quantify and define. Also, as eSports is too new to have a consensual way of data processing, so far, the data analysts in some first-tier clubs and most analysts in second-tier make decisions simply based on their own understanding of the usually visible information in a match. As a result, machine learning will be a powerful solution for helping analysts improve their teams.

Up until now, the eSports industry is still a new-born industry that lacks a lot of information, so there are only a few reliable researches done so far compared with other mature academic fields. Particularly, there is limited research that concentrates on the combination of machine learning and eSports. Therefore, this essay will focus on showing how machine learning completes the essential tasks for eSports analytical requirements and comparing different machine learning techniques on their potency and feasibility in the prediction of outcomes in eSports matches.

2. Background Information

2.1. Some Facts about eSports

In 2014, already more than half of the U.S. and European citizens knew what eSports is. During the outbreak of the COVID-19 pandemic, most people's everyday lives have been limited to being at home, and this is when video gaming and streaming reached their peak in recent years. By the end of 2023, it is estimated that eSports viewers will exceed 646 million [7].

Basic categories of eSports consist of First Person Shooter (FPS), Multiplayer Online Battle Arena (MOBA), fighting games, real-time strategy (RTS), and card games. Data shows that MOBA and FPS games are the most popular types of games based on prize pool and peak viewers. There are more than 50% of MOBA games and 30% of FPS games in every top ten video game each year for the last five years [8]. A well-known FPS game series is Counter-Strike and its successor CS: GO (now CS2). However, for most top FPS games, games like Overwatch, Apex Legends, Quake, or Valorant also contain a considerable amount of MOBA elements as well. Therefore, MOBA games are the most preferred games in the environment, among which League of Legends and Dota 2 are the most representative ones. FPS and MOBA games have their own characteristics, so there are many distinct factors that could help a team win a match based on game type. For instance, hero selection is unique in MOBA, and controlling the crosshair is unique in FPS. Moreover, there are a large number of factors that both types of games could share as well, including players' physical and mental health, relationship with teammates, positions on the map, economy, kill-death rate, etc. As a result, later parts of this paper will focus on how to use machine learning methods to build reasonable models considering some important factors of each game type and predict the outcome of a game.

2.2. Machine Learning Models

Common models used in the prediction of outcomes in eSports competitions involve Logistic Regression, K-Nearest Neighbors, Random Forest, Support Vector Machine, Neural Network, Lasso Logistic Regression, and K-Means Clustering.

(1) Logistic Regression--LR

Logistic Regression is a powerful model in machine learning, which can be further categorized into Binary Logistic Regression and Multinomial Logistic Regression. A Binary Logistic Regression is a classification model that takes single or multiple inputs on a binary outcome (usually 0 and 1, referring to true and false), while a Multinomial Logistic Regression predicts the probability of one of more than or equal to three outcomes, like what type of toothpaste people prefer. Logistic Regression is a predictive analysis, and it can predict a given subject about which category it belongs to by computing its probability over the logistic function.

(2) K-Nearest Neighbors—KNN

K-Nearest Neighbors is an effective non-parametric supervised learning algorithm that can be applied to both regression and classification problems. Since it is non-parametric, it is widely applied in real-world situations as it does not presume the data follow any kind of distribution. As written in its name, the algorithm functions in a way that each data point's value is determined by its k nearest neighbors, these values will be used in the final prediction. This algorithm is preferred since it has a relatively low complexity.

(3) Random Forest--RF

Random Forest is a supervised learning algorithm that is strong enough to process all kinds of data. The 'forest' it refers to is a collection of decision trees (that are subsets from the training set with replacement) that are usually trained with the bagging method. Decision trees algorithm is an algorithm that aims to find the most suitable split of the data by making either-or selections in each decision node. The average of individual decision trees will be used for regression analysis, and the most frequent categorical variables will be used for classification analyses. The final prediction is drawn by using a cross-validation method [9].

(4) Support Vector Machine--SVM

Support Vector Machine is a machine learning model designed for regression and classification. It aims at finding a hyperplane in a multidimensional space that best separates different categories of data, and uses this to predict which category a new data would most probably belong to. There are several types of this best-fit hyperplane for different distributions of data. For linear analysis, one type of hyperplane could be decided by maximizing the distance from the closest data point of each side, named as a maximum-margin hyperplane. For non-linear analysis, usually, this hyperplane is found by using the Kernel trick where the inner product of the image of all pairs of data is computed with the Kernel function to find the pattern similarity.

(5) Neural Network—NN

Neural Network is one of the most popular and most powerful machine learning algorithms currently. NN can achieve the analysis of patterns in a dataset even when the structure of data is not explicitly stated. It is widely used to train data with nonlinear relationships or complex implicit relationships, especially in fields with image recognition, natural language processing, facial recognition, or medical analysis. Neural Network algorithm is built on layers of nodes, inside which there is always an input layer, an output layer, and one or multiple hidden layers in the middle. Each node works as a neuron in the human brain and connects to other nodes with reasonable weights and thresholds which determine whether information should be delivered to the next node. Each layer works as classifiers that categorizes data and passes them to the next layer selectively. Eventually, results are produced by the output layer. There are usually two types of NN: Convolutional Neural Networks (CNN), which are usually used in image pattern identification, and Recurrent Neural Networks (RNN), which are usually used in dealing with time-series data.

3. The Application of Machine Learning in eSports

3.1. Machine Learning Applications in Predicting the Outcome of MOBA Games

One unique characteristic of MOBA is that every game with MOBA elements consists of a large variety of distinct characters available for use. Classical examples are League of Legends and DOTA2. Since every character has his or her own ability pools, this abundancy of characters provide a limitless combination of abilities where players can best use to fit their roles in a team. As there is a counter and anticounter relationship among the diverse abilities, consequently, the match up of the heroes selected for the two teams can largely affect the win rate beforehand. Therefore, experiences show that a good hero selection is the key factor in winning a match. Also, in MOBA games, there are other elements that can be used for measurement, including gold and experience per minute, map control, teammate relationship etc. These factors will be considered collaboratively in the building of machine learning models. In the next part, paper [10] and [11] highlight the significance of correct hero

selection in MOBA games, and compare the efficiency of various machine learning models in predicting the outcome of a match.

First, Paper [10] gives a thorough analysis about the importance of hero pick in Dota2. The filtered dataset consists of 87204 matches that average between 37 and 45.7 minutes long, which is accord with real life situation, so the training dataset is not biased. There are several indexes for a match – difficulty, hero score, and experience score. The hero score is computed by doing dimensional reduction algorithm on the original attributes of each hero and trained in a Logistic Regression model since the data transformation in this case will be linear. This method is effective. It avoids the impractical complexity of computing every combination of 5 in the total of 66 heroes. In the meantime, it avoids the ambiguity if the heroes are defined solely by the position they are played in each fight. After doing ANOVA analysis using R, paper [10] indicates that hero selection is consistent in all levels of difficulty, with significant impact to the team's win rate of a match. Also, by doing Mann-Whitney test for mean difference of hero scores for the two teams, and using log-linear analysis in ANOVA to test the interaction among heroes, it also shows that the strongest hero selected and the combination of the five heroes are significant factor to success as well. Based on the significance of hero selections, some further analyses propose horizontal comparisons for the efficiency among different machine learning models on the win rate prediction using data containing more comprehensive elements. In paper [11], Porokhnenko, Polezhaev, and Shukhman compare the efficiency of plenty of machine learning techniques in predicting match results in Dota 2 including Logistic Regression, Linear Support Vector Classification, Random Forest, Gradient Boosting Classification, Neural Network and so on. The multivariable factors measured in the models include player skills, match time, and hero selections. By tuning the best hyperparameters using the Gridsearch algorithm, LR, LSVC, and Neural Network with activation function Softplus and Sigmoid yield the highest ROC AUC value equal to 0.77. Paper [12] studies the efficiency of machine learning methods in predicting outcomes in League of Legends by comparing LR, Linear and Quadratic Discriminant Analysis, Lasso Regression, SVM, and Gaussian Process. The dimension of the models is based on measuring 12 elements including Team, Side, Position, Player, Gold per 10 minutes etc. It concludes that all of the models have an accuracy at slightly over 80%, and LDA is the most consistent and well-performed model in this analysis. Other machine learning methods like Naïve Bayes and K-Nearest Neighbors are widely used in another research as well [13-14].

3.2. Machine Learning Applications in Predicting the Outcome of FPS Games

This section will focus on the analysis of how machine learning methods achieve reasonable predictions in the second most popular eSports game type – FPS games. Unlike MOBA games, FPS games usually does not include a hero selection part where different player in a team can have distinct hero abilities, and FPS games always involve players' manipulation on their mouse in controlling a crosshair to aim enemy targets while playing. Thus, the perspective in analyzing and predicting FPS games is to some extent different.

Research [15] published in 2018 gives a decent example of how machine learning is applied to the most representative FPS game-CS: GO (which stands for Counter Strike: Global Offensive, and currently is updated to a new version called CS2). The dataset is obtained by web scraping key features through the CS: GO demo files from an independent platform called FACEIT. The parser employs concurrent programming to achieve speedup in effectively parsing a large amount of data. Since it is impossible to track every feature that could potentially affect the outcome of a match, only players' team roles are considered in the models during the first test. However, noticing that the prediction rate would vary since the dimension of features are too low, the number of essential features tracked in the clustering increases to 50 after several more tests, which is a relatively high number of features compared with other research. Examples of features are players' kills and deaths, type of weapon when scoring a kill, type of kills, equipment, position, crosshair movement before scoring a kill, etc. The paper uses several algorithms. First it applies K-Means Clustering and Evolution Algorithm to find an optimal clustering, and then applies Neural Network with binary

cross-entropy cost function and logistic and RELU activation functions to predict team win rate. The maximum prediction accuracy reaches a 65.11% using Neural Network method and 32x clusters. Another research on the prediction of win rate per round and per match in CS: GO uses some other machine learning methods. Like [15], paper [16]'s dataset is gained from parsing match demos from a similar independent platform called HLTV.org, In this paper, 15 features across the 10 players in a match along with 2 general features are taken into account, including Team Type, Balance, Current Equipment Value, Score, etc. Unlike [15] using Neural Network, [16] implements three Classification models in training: Random Forest, XGBoost, and Multi-layer Perception. After tunning hyperparameters, Random Forest has the highest accuracy of 0.6407 when predicting the round winner, and XGBoost has the highest accuracy of 0.6237 when predicting the match winner.

There are not many research studying machine learning methods' effect on win rate of FPS games compared with MOBA games. Considering other aspects that machine learning could potentially contribute in predicting the win rate, there are research studying how machine learning methods can be used to measure the level of skills for each player in an FPS game [17-18]. Also, some research pay attention to predict the tiredness and behavioral features that could potentially be influential factors of winning a match as well [19-20]. For example, paper [19] uses Logistic Regression, K-Nearest Neighbors, and GRU Neural Network to measure outside-game features like chair movements, pulse, muscle activity, electroencephalogram, and environmental temperature. It can successfully predict whether a player will win or lose the next encounter fight with accuracy at around 70-73.5 percent. Paper [20] uses a new data collection method called Smart Chair that can detect players' physical and psychological changes while sitting on this chair. It then implements Logistic Regression, SVM with RBF kernel, Random Forest with 100 estimators and maximum tree depth 2, and K-Nearest Neighbors with k=3 in the model analysis. Random Forest has the best result with 77% prediction accuracy and 0.88 ROC AUC curve.

4. Conclusion

This paper elucidates elements of eSports competition and the mechanics of machine learning algorithms, gives a broad overview of how these algorithms can be applied to the prediction of win rates in MOBA and FPS games, and exhibits and compares the efficiency of different machine learning models. In the studying of MOBA games, hero selection is always a key influential factor to the win rate in various papers, but there is not a consistently dominant factor that can cause the win rate of a team to fluctuate in FPS games according to several research; the feature that can affect win rate the most is the equipment or equipment value. Across all the games stated in this paper, common factors like Kill-Death-Rate, gold per time period, and player experiences are the determining factors to either team win rate or individual player's skill level, and across all the machine learning models proposed in the paper, Logistic Regression, Random Forest, and Neural Network are the most consistent and effective models among all the cases that being analyzed, which usually yield an average win rate prediction accuracy of 60%-75%. This value of average accuracy is usually not considered satisfied in real-life situations, so improvements should be furthered in future studies. Since machine learning algorithms themselves are good enough to complete the tasks, and the features that can or potentially can impact a game is too versatile, one improvement in future analysis should definitely be a better categorization of features based on the characteristic of the game and better real-time data scraping method that can cover more perspectives both inside-game and outside-game including professional players' physical or mental status, muscle tension, heart rate, etc. With these data, there will be less bias and differences in training the models and more accordant outputs will be performed. Additionally, current research mostly focuses on predicting the result of a match or a round using the data from before the match or in real-time, but this will not help the eSports clubs directly since no direct actions for improvement can be concluded from merely a win rate or a player's skill level. Hence, systems that focus on what hero selections, hero bans, or hero combinations (for MOBA games), what equipment should each player buy in current round, what angle or map position

should a player hold in the current round, as well as what strategies (like aggressive or defensive) should a team play against different opponents in different rounds could be the center of future studies so that clubs and companies are able to directly employ these techniques to improve their teams timely and make eSports industry more prosperous.

References

[1] Lynch J. As NFL ratings drop, a new internet study says young men like watching eSports more than traditional sports. Retrieved October 30, 2018, from https: //www.businessinsider.com/nfl-ratings - drop - study - young - men - watch - esports - more - than - traditional - sports - 2017 - 9.

[2] Heath, J.10 biggest prize pools in esports (updated 2023). Dot Esports. https: //dotesports.com/general/news/biggest-prize-pools-esports - 14605.

[3] Janiesch, C., Zschech, P., & Heinrich, K. Machine learning and deep learning. Electronic Markets, 2021, 31 (3), 685 - 695.

[4] Barboza, F., Kimura, H., & Altman, E. Machine learning models and bankruptcy prediction. Expert Systems with Applications, 2017, 83, 405 - 417.

[5] Bochenek, B., & Ustrnul, Z. Machine learning in weather prediction and climate analyses—applications and perspectives. Atmosphere,2022, 13 (2), 180.

[6] Senders, J. T., Staples, P. C., Karhade, A. V., Zaki, M. M., Gormley, W. B., Broekman, M. L., ... & Arnaout, O. Machine learning and neurosurgical outcome prediction: a systematic review. World neurosurgery, 2018,109, 476 - 486.

[7] Block, S., & Haack, F. eSports: a new industry. In SHS Web of Conferences. EDP Sciences, 2021, 92, 04002.

[8] Most popular esports games 2023 | esports charts. (n.d.). https: //escharts.com/top-games?order=peak.

[9] Zhang, Y. Prediction of Esports Game Results Using Early Game Datasets. In ICMLCA 2021; 2nd International Conference on Machine Learning and Computer Application VDE, 201, 1 - 6.

[10] Pobiedina, N., Neidhardt, J., Moreno, M. D. C. C., Grad-Gyenge, L., & Werthner, H. On successful team formation: Statistical analysis of a multiplayer online game. In 2013 IEEE 15th Conference on Business Informatics. IEEE, 2021, 55 - 62.

[11] Porokhnenko, I., Polezhaev, P., & Shukhman, A. Machine learning approaches to choose heroes in dota 2. In 2019 24th Conference of Open Innovations Association (FRUCT)IEEE, 2021, 345 - 350.

[12] Bailey, K. Statistical learning for Esports match prediction.2020.

[13] Chan, A. S., Fachrizal, F., & Lubis, A. R. Outcome prediction using Naïve Bayes algorithm in the selection of role hero mobile legend. In Journal of Physics: Conference Series, IOP Publishing. 2020, 1566 (1), 012041.

[14] Shen, Q. A machine learning approach to predict the result of League of Legends. In 2022 International Conference on Machine Learning and Knowledge Engineering (MLKE). IEEE. 2022, 38 – 45.

[15] Björklund, A., Johansson Visuri, W., Lindevall, F., & Svensson, P. Predicting the outcome of CS: GO games using machine learning. 2018.

[16] Rubin, A. Predicting Round and Game Winners in CSGO. 2022.

[17] Zhang, L., & Wang, W. Multi-Modal Machine Learning for Assessing Gaming Skills in Online Streaming: A Case Study with CS: GO. arXiv preprint arXiv: 2023, 2307. 12236.

[18] Buckley, D., Chen, K., & Knowles, J.Rapid skill capture in a first-person shooter. IEEE Transactions on Computational Intelligence and AI in Games,2015, 9 (1), 63 - 75.

[19] Smerdov, A., Somov, A., Burnaev, E., Zhou, B., & Lukowicz, P. Detecting video game player burnout with the use of sensor data and machine learning. IEEE Internet of Things Journal, 2021, 8 (22), 16680 - 16691.

[20] Smerdov, A., Burnaev, E., & Somov, A. esports pro-players behavior during the game events: Statistical analysis of data obtained using the smart chair. In 2019 IEEE Smart World, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big

Data Computing, Internet of People and Smart City Innovation (Smart World/SCALCOM/UIC/ATC/CBDCom/IOP/SCI) (pp. 1768 - 1775). IEEE. 2019.