Journal of Advanced Zoology

ISSN: 0253-7214
*Volume **44** Issue **S-6 Year 2023** Page **1050:1055***

Predictive Analysis and Comparison of Various Models on Esports Competitions

Adarsh A. Nair[1*], Aniketh V. Hotagi[2], Abdul Mateen[3], Ankit Kumar[4], Kiran M[5]

[1,2,3,4,5]*School of Computer Science and Engineering, REVA University*
[1]*r18cs018@cit.reva.edu.in*
[2]*r18cs040@cit.reva.edu.in*
[3]*r18cs003@cit.reva.edu.in*
[4]*r18cs043@cit.reva.edu.in*
[5]*kiran.m@reva.edu.in*

**Corresponding author's E-mail: r18cs018@cit.reva.edu.in*

Article History

Received: 06 June 2023
Revised: 05 Sept 2023
Accepted: 30 Nov 2023

Abstract

eSports has emerged as a popular genre for players and viewers, promoting a global industry in entertainment. The study of eSports has grown to resolve the need for data driven feedback, which focuses on assessment, strategy, and prediction of cyber-athletes. The focus of this project is to create and compare various models to predict the likely winner for professional games based on the data recorded from various eSports tournament matches. Pro-games have the top industry and audience attention but are restricted in number. The project is dominant on Deep Learning and Machine Learning, where the predictions are made using the model that we will build. This project can play a big part in gauging which model is most suitable for predicting the results of a match.

Keywords: *Machine Learning; Predictive Analysis; eSports; Logistic Regression; XG Boost; Feed-Forward Neural Network*

1. Introduction

eSports is a term that refers to a medium of entertainment generally viewed by large audiences, and which involves professional competitive tournaments among highly qualified teams. [1]

With the increasing number of online streaming media outlets, eSports has grown in popularity as a common form of entertainment. League of Legends (LoL), Dota 2, and other professional esports tournaments are among the most successful eSports competitions. This is evident such that the success and acceptances of eSports tournaments has grown at a pace comparable to, if not greater than, that of physical conventional sports such as soccer or basketball [2]. eSports is perceived by investors as a sector that has a significant impact on game sales and the potential to develop into a stand-alone industry in the future.

eSports analysis [3] is described as: "The process of using eSports related data, to find meaningful patterns and trends in said data, and the communication of these patterns using visualization techniques to assist with decision-making processes."

Predictive analytics has traditionally played a significant part in sports and athletics. As part of each major professional sports team, an analytical division or specialist is standardized. The primary objective of these analytical experts is to perform predictive analysis in order to increase the team's win rate based on the statistical data obtained from the team and individual player's previous games.

As with conventional tournaments such as the NFL and NBA, eSports tournaments can also benefit from predictive analysis such as to discover better team compositions and to predict the outcome of eSports games. However, when compared to traditional sports, eSports prediction will include a greater amount of detail and complexity since it incorporates characteristics such as champion selections, team composition, counter picks and various other factors that could also potentially influence game results besides a players' statistics. This project will focus on predicting the likely winner of a LoL eSports tournament.

In this project, *League-of-Legends* has been selected as a sample for the eSport predictive analytics.

The League-of-Legends is a fast-paced, competitive, multiplayer virtual battle-field game developed and articulated by Riot Games. It combines the intensity and complexity of real-time strategy with elements of role-playing games. The official developers maintain a repository of open-source, accurate, and trustworthy data on champions, players and matches which can be used to construct predictive analytics.

Previous works and experiments in the field of predictive analysis of any competitive events have proved to be highly useful in terms of aspects ranging from recognizing the precision and general utility of the predictions themselves to providing analytical support in helping the audience, including but not limited to casual viewers understand the smaller events and their outcomes that take place throughout the game. Exploration in the field of predictive analysis of eSports has a lot of potential to benefit anyone involved in the sport. Compared to conventional sports event, a significant benefit of eSports estimation is that all the characteristics can be produced automatically and accurately in the course of the game.

This paper will aim at implementing various machine learning algorithms and compare them to find the most suitable model to predict the outcome of the matches.

Literature Survey

eSports, an acronym for "electronic sports", extends online video gaming into a competitive gaming activity. The concept is similar to that of attending a professional conventional sports match, with the exception that audiences get to see professional gaming teams play against one another in a virtual online arena instead of a conventional physical sporting match. The popularity of eSports has risen dramatically over the past few years along with the overall revenue that flows in the eSports industry. These games, similar to that of conventional sports, are divided into professional leagues and tournaments. [4]

Millions of gamers, viewers, and sports enthusiasts from all over the world participate in the event as a spectator either virtually through live-streaming platforms such as Twitch or by visiting large public eSports events to watch their favorite team compete for the massive prize pool and to win the championship. By the 2010s, eSports was a substantial aspect in the video-game industry, with various game developers enthusiastically designing and affording subsidy for play-offs and other events and by 2023, it is anticipated to be approximately 300+ million users of eSports globally, an immense growth from the 173+ million in the year 2018. [5] [6]

Additionally, the whole eSports market is anticipated to grow from 694.2 million US Dollars in 2017 and is estimated to reach 2,174.8 million US Dollars by 2023". Reference [7] stated "The viewership of 2016 world finals of popular video game League of Legends (43 million viewers) surpassed the viewership of NBA Finals Game 7 (31 million viewers) of the same year by 12 million viewers".

The most popular eSports tournament franchises include League of Legends, Super Smash Bros, Fortnite, Street Fighter, Dota 2, Counter-Strike, Overwatch. Genres most commonly associated with eSports competitions are Team Battle Royale, First-Person Shooters (FPS), Real-Time Strategy (RTS), and Multiplayer Online Battle Arena (MOBA). [4]

Predictive Analysis

Machine learning is domain of computer science that studies the systems which "change their behavior in a way that makes them perform better in the future". [8]

Deep learning is a subset of a larger family unit of machine learning techniques that are based-upon neural networks with feature-learning. It is designed to simulate the way the human brain processes data in order to identify objects, recognize expression, translate languages, and make decisions.

Initially, various Traditional Regression and classification techniques will be applied to gauge the effectiveness of each model from the accuracy of their respective predictions. Then Deep Learning Techniques such as Neural Networks are used to see if then would perform better in this scenario.

A. Data used in Analysis

The dataset used in training the models is provided by Tim Sevenhuysen of OraclesElixir.com [9]. It contains thousands of matches played each year in csv format along with the unique GameHash of each match which can be used to verify the legitimacy of the data.

The dataset contains of around 70 attributes and 5000 games, which make up around 45,000 records captured from the recent matches in 2021, which can be used to predict the outcome of the match, which is indicated by the binary labels - 1 (for win) and 0 (for loss) in the target variable *'result'*.

B. Pre-processing the Data

First, we remove attributes which does not have any effect on the prediction like 'PlayerName', 'GameID', 'Date', etc. Then we remove the records where the data is incomplete (given by a data completeness attribute).

Now we need to encode non-numerical/categorical values such as *'TeamName'*. For this purpose, we use Label Encoder from the scikit-learn package. The same encoder is also used for encoding the target labels, i.e., *'results'* with value between 0 and (no. of classes - 1).

Min-Max Scaler is also applied on the input variables which is a type of Feature Scaling. It transmutes features by scaling each one to a particular scale. This scale and transforms each feature independently so that it falls within the training set's defined range. We then split the dataset into training and testing in a 75:25 ratio using the sklearn package for training the model.

We can see the correlation between the different attributes using the Heatmap as shown in Fig. 1.

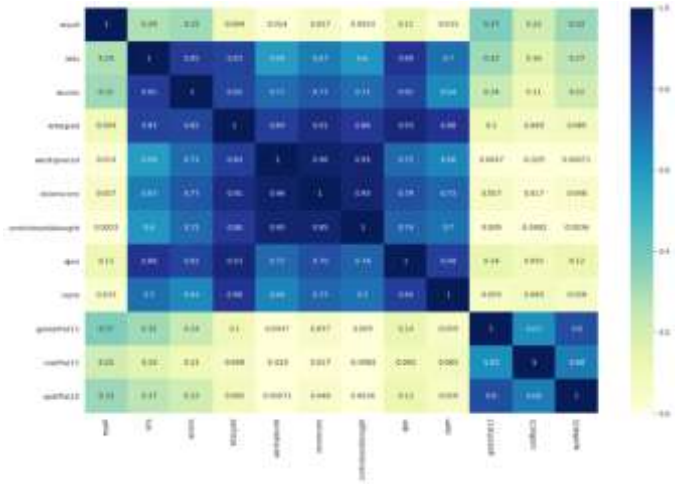

Fig. 1. A Heatmap of various different variables

C. Traditional Machine Learning Algorithms

1) Regression

We will apply various regression algorithms on our dataset, namely Lasso, Ridge and Logistic. we import those regression models from the external libraries such as sklearn package.

First, we use the Least Absolute Shrinkage and Selection Operator (LASSO) Regression, it is a particular kind of linear regression algorithm which makes use of shrinkage. This shrinkage occurs when data values are compressed in the direction of a certain key point, such as the mean. This procedure encourages the use of simple models with fewer parameters [10]. The lasso model is run with an alpha of 0.0001 and maximum iterations of 1,00,000.

Next, we make use of the Ridge Regression model, it is a technique of assessing the co-efficient of multiple-regression models in scenarios where independent variables are highly correlated [11]. This model has the Normalization parameter set as true with an alpha of 0.1.

Finally, we use the Logistic Regression model on our dataset which is a probability estimation model centered on the cumulative logistic distribution that analyses data and determines a categorical outcome utilizing one or more independent variables. We use maximum iteration of 8,000 on this model.

We observe and log the accuracy scores of all these regression algorithms.

2) Random Forest

It is a meta-estimator that equips a large number of decision-tree classifiers on different sub-samples of the dataset and then uses averaging to increase the predictive precision and avoid overfitting. A parameter controls the size of the subsample; otherwise, each tree is constructed using the entire dataset.

We import Random Forest Regressor from the sklearn package. Then, set the number of trees in a forest (n_estimators parameter) as 500.

The score is observed and logged for further analysis.

3) XG Boost

XGBoost is a library that uses distributed gradient boosting which has been developed to be extremely powerful, scalable, and portable. It utilizes Gradient Boosting to incorporate machine learning algorithms. It enables concurrent tree boosting. [12]. We import the XGBoost from the external xgboost package. After running the model, we observe that the accuracy is approximately 95%.

To improve the accuracy of XGBoost model, we apply feature selection using In-built functions such as Feature Importance to find out the importance of each variable to XGBoost Classifier model as shown in Fig. 2.

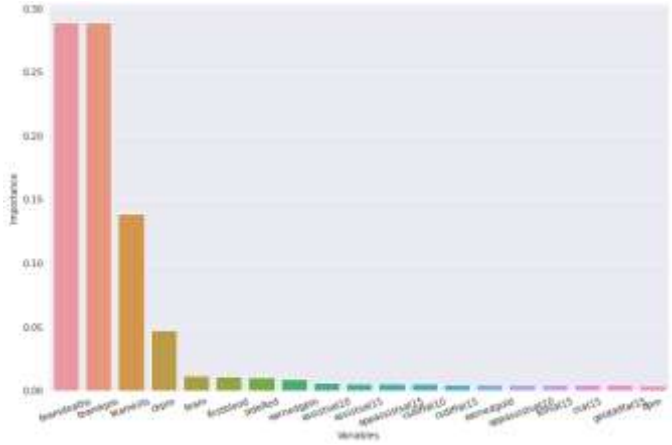

Fig. 2. Importance of each variable for the XGBoost Classifier Model

We use the few variables with the most importance according to the XGBoost model and optimize our input variables. On running the model again, the accuracy is slightly increased to 96%.

Deep Learning Algorithms

1) Feed Forward Neural Network

We will use Deep Neural Network (DNN) which is a type of Feed-forward neural network wherein there are several tiers between input and output layers. It comes under the ANN (Artificial neural network) where connections between the nodes do not form a loop. In this type of neural networks, data only moves uni-directionally, from the input nodes to the output nodes through the hidden nodes (if present). [13]

We import DNN classifier model from Tensorflow 2.0 package. We set the class attribute as 2 and we define Four layers in our neural networks with number of hidden units per layer as 512, 256, 125, 50, respectively. We set the batch size parameter of the classifier as 512.

3. Results and Discussion

Experimental Results

From the observations, we can see the accuracy levels of various algorithms we used in predictive analysis shown in the following table.

Accuracy Levels of Algorithms using predictive analysis

Sl. No.	Algorithm	Accuracy
1.	Lasso Regression	82.02 %
2.	Ridge Regression	78.38 %
3.	Logistic Regression	97.01 %
4.	Random Forest Classifier	94.44 %
5.	XG Boost	96.01 %
6.	Feed-Forward Neural Network	95.95 %

4. Conclusion

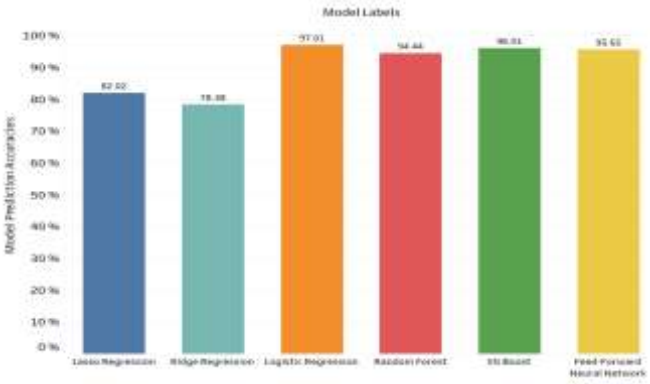

Fig. 3. Bar-Chart showing the accuracy levels of all the models

From the bar chart shown in Fig. 3., we can conclude that Logistic regression has a slight advantage over the other models and provides the best accuracy in this scenario for predictive analysis.

Predictive analysis as demonstrated in this paper can help eSports teams and the stakeholders develop advanced tactics and fine tune the team composition based on the performance with respect to any other known eSports team.

Furthermore, the concepts and procedures applied in this project can be extended, adapted, and applied to any other MOBA games such as Dota 2, Smite, etc. with the best compatibility, in addition, similar predictive analysis can also be performed on eSports games from other genres such as FPS games, which often shares the criteria of team-based competition, kills-death-assists ratio (KDA), and so on. For instance, Counter-Strike: Global Offensive is a highly competitive 5v5 game like that of most MOBA games, which also has a great presence in eSports tournaments worldwide.

In this project, we explored and compared the accuracies of various machine learning and deep learning algorithms. This helps us infer which models perform better in this scenario and could further be used for predictions on the Real-World data.

A. Future Implementations

To get insights in real-time, we can use Deep Learning models such as Long Short-Term memory (LSTM) which is a kind of Convolutional Neural network (CNN) and Recurrent Neural network (RNN) along with Object Detection and Computer Vision to do Real-Time Analysis of the matches to get predictions about the outcome of the match while the match is still in progress.

This can help the professional teams as well as individuals to get insights of paramount importance about which key areas to focus on and their weaknesses to be worked upon.

It can also help the audience to know which actions taken by the players can lead to the change in the outcome of the matches.

Acknowledgement

We would like to thank REVA University and the faculty members for guiding us and providing all the necessary support for the team in completing the technical paper.

References:

J. Hamari and M. Sjoblom, “What is eSports and why do people watch it?” Internet research, vol. 27, no. 2, pp. 211–232, 2017.

Krush, A. (2020, November 17). eSports may soon overtake traditional sports(analysis) ObjectStyle. https://www.objectstyle.com/agile/eSports-may-soon-overtake-traditional-sports-analysis

M. Schuert, A. Drachen and T. Mahlmann, “ESports analytics through encounter detection,” in Procs of the MIT Sloan Sports Analytics Conference, 2016.

Esports - https://en.wikipedia.org/wiki/Esports

Chapman, J. (2017, July 6). ESports: A Guide to Competitive Video Gaming. Toptal Finance Blog. https://www.toptal.com/finance/market-research-analysts/eSports

Gough, C. (2020, July 28). eSports market - Statistics & Facts. Statista. https://www.statista.com/topics/3121/esports-market/

Pei, A. (2019, April 14). This eSports giant draws in more viewers than the Super Bowl, and it’s expected to get even bigger. CNBC. https://www.cnbc.com/2019/04/14/league-of-legends-gets-more-viewers-than-super-bowlwhats-coming-next.htmlLoL eSports. (2017). About North American LCS. LoL eSports. http://www.loleSports.com/en_US/na-lcs/na_2017_summer/about

I. H. Witten and E. Frank, Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. San Francisco: Morgan Kaufmann, 2000.

Sevenhuysen, T. (2021, January 1–April 27). Oracle's Elixir - LoL ESports Stats [Dataset]. Oracle's Elixir. https://oracleselixir.com

Glen, S. (2015, September 24). Lasso Regression: Simple Definition. Statistics How To. https://www.statisticshowto.com/lasso-regression/

Ridge regression. (2021, March 13). In Wikipedia. https://en.wikipedia.org/wiki/Ridge_regression

XGBoost 1.5.0 documentation. (n.d.). XGBoost Documentation. https://xgboost.readthedocs.io/en/latestFeedforward neural network. (2021, March 23). In Wikipedia. https://en.wikipedia.org/wiki/Feedforward_neural_network