A machine learning approach to predict the result of League of Legends

Qiyuan Shen
Data Science, Duke Kunshan University, Kunshan, Jiangsu Province, 215300
* Corresponding author: qs31@duke.edu

***Abstract*—Nowadays, the MOBA game is the game type with the most audiences and players around the world. Recently, the League of Legends has become an official sport as an e-sport among 37 events in the 2022 Asia Games held in Hangzhou. As the development in the e-sport, analytical skills are also involved in this field. The topic of this research is to use the machine learning approach to analyze the data of the League of Legends and make a prediction about the result of the game. In this research, the method of machine learning is applied to the dataset which records the first 10 minutes in diamond-ranked games. Several popular machine learning (AdaBoost, GradientBoost, RandomForest, ExtraTree, SVM, Naïve Bayes, KNN, LogisticRegression, and DecisionTree) are applied to test the performance by cross-validation. Then several algorithms that outperform others are selected to make a voting classifier to predict the game result. The accuracy of the voting classifier is 72.68%.**

Keywords-League of Legends; machine learning; Voting classifier; MOBA

I. INTRODUCTION

Compared to the last decade, the views towards games have changed a lot in these years. Parents hold an open opinion towards the games rather than the spirit opium which is commonly believed in the past. As the development in the game industry, there are more and more types of games to choose like: RPG (Role-Playing Game), FPS (First-person shooting game), strategy game, and so on. And the game competition has also become an official sports event which is called e-sport. And the e-sport will also be one of the 37 sports played in the 2022 Asia Games held in Hangzhou. Among all these e-sports, the MOBA (Multiplayer Online Battle Arena) game is the most popular one. According to Mora-Cantallops and Sicilia's article, about 205 million people participate in the MOBA games (2018). The research topic League of Legends is the hottest MOBA game. In 2013, the League of Legends Championship had drawn 27 million online audiences to watch the competition. Correspondingly, money and attention have been spent on players' performance during the game. Data analysis by machine learning is one of the topics of this research.

The research topic is the application of machine learning skills in competitive games: predicting the result of the competition by game indicators and analyzing which indicator affects the game most. The research provides a relatively objective method to analyze and rate the game to avoid subjective influence. In addition, this research can also be applied to other e-sport games and sports games to help players and athletes improve their performance.

Then some recent works in the application of machine learning in League of Legends are checked for reference. In Xia, Wang, and Zhou's research, they mainly focus on another popular MOBA game: DOTA 2 (Defense of the Ancients 2). They choose R to analyze the data and they find out that the tactical awareness affects the multiplayer killing indicator and the result of the game more than the operations [2]. Another project also research data of the game DOTA 2. They present live win predictions in MOBA games by using standard machine learning models [3]. The third research mainly compares the results of 3 models: Linear Regression, Neural Network, and LSTM [4]. Another research mainly uses 3 learning models Naïve Bayes classifier, Logistic Regression, and Gradient Boost. And their result shows that the prediction oscillates depending on players' operation skills [6]. Then researches in League of Legends was cited to see the commonly used method in League of Legends' field, in a recent research, the researchers mainly use a deep neural network based on the players' previous experience on the champions they choose [5]. While another research mainly focuses on the pre-game factors' influence on the result of the game [7]. There is one research that carries recurrent neural networks on the data about League of Legends. And their results show that the accuracy of the model trained by 0-5 minutes data is worse than the data from 20-25 minutes [8]. In Lin's research, the researcher conducts several popular machine learning models which give some inspiration to this research [9].

II. METHODS

A. Game map introduction

Figure 1 below is the classical map for the MOBA game including League of Legends. There are three lanes on the map which are the top, middle, bottom lane. Each lane is separated into 2 parts for each team. Each lane has 2 defend towers. At the end of each lane, it is the base of each team and consists of 3 defend towers and a Nexus. Between each lane, it is the Jungle where there are jungle minions in it which can provide experience and gold. There are also some elite monsters in the jungle. They are dragons and heralds which can bring great benefits to the team who kills them.

978-1-6654-9567-7/22/$31.00 ©2022 IEEE
DOI 10.1109/MLKE55170.2022.00013

Figure 1. The general map for the MOBA games

B. *Dataset introduction*

Normally, one game of League of Legends will last for 30-45 minutes. To ensure the accuracy of the model and make the prediction available to be applied in e-sport games, the dataset used in this research records the data of the first 10 minutes of different games. All the games in the dataset are diamond-ranked. Moreover, to ensure enough data to train, there are nearly 10000 data in the set. In the dataset, 40 columns record the detailed game data in the first 10 minutes.

C. *data processing and visualization*

The first thing in the data processing part is to clean the missing and duplicate values. Fortunately, there are no data missing or duplicating in this dataset. Because there is so much data in the dataset. Choosing proper features is especially important in this research. Table 1 below shows some features that are useful in this project.

TABLE I. SOME USEFUL INFORMATION IN THE DATASET

blueWins	The final goal for the research. The feature blue wins is used as the determent (1 means blue wins, 0 means red wins).
blueWardsPlaced/ blueWardsDestroied	The warding totems are an important item in League of Legends. When it is placed in an invisible area, it can reveal an area for the team that places the warding totems. This feature shows the wards placed/destroyed by the blue team.
blueKills/ blueDeath/ blueAssists	These 3 features are the basic data for the game which record the kills and death between different teams.
blueEliteMonsters/ blueDragons/ blueHeralds	These 3 features record the data about the elite monsters: Dragon and Heralds which can give great benefit to the team killing them in a certain period.
blueTowerDestroyed	This feature records the defending tower destroyed by the blue team which is also an important factor deciding the result of the game.
blueTotalGold/ blueAvgLevel/ blueTotalExperience	The total gold and experience earned by the team can represent the development of the champion in that period.
blueTotalMinionsKilled/ blueTotalJungleMinionsKilled	This feature is closely related to the gold and experience earned by the team.

Different teams have different game paces. Moreover, some champions can dominant the game at the early stage which may affect the accuracy of the model. So, several features are standardized to make it easier to compare their differences. The main standardized function is shown below:

$$\text{Standardized ratio} = \frac{blue\ data - red\ data}{blue\ data + red\ data}$$

In this research, this function is used to standardize the following features: blueKills, blueAssists, blueWardsPlaced, blueWardsDestroyed, blueTotalGold, blueTotalExperience, blueAvgLevel, blueTotalJungleMinionsKilled, blueTotalMinionsKilled, redKills, redAssists, redWardsPlaced, redWardsDestroyed, redTotalGold, redTotalExperience, redAvgLevel, redTotalJungleMinionsKilled, redTotalMinionsKilled. Because the features of elite monsters, dragons, heralds and first blood are so close to each other. So, they haven't been standardized it.

Then the processed data is visualized in the following figures.

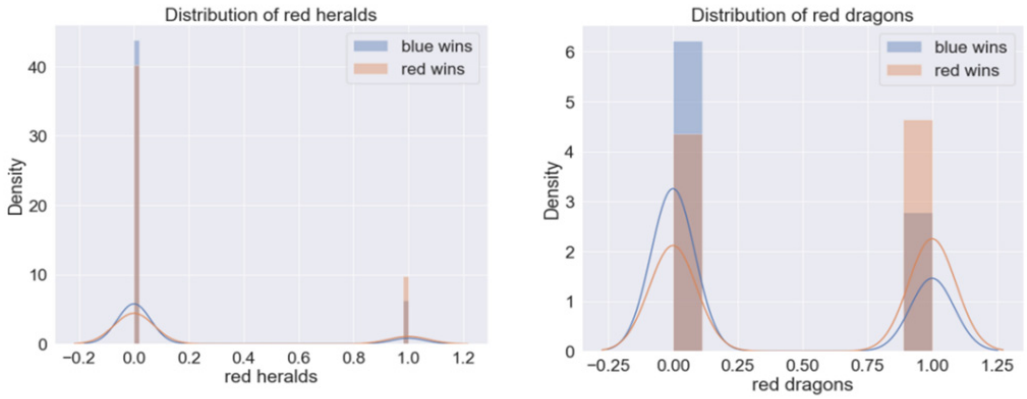

Figure 2. The distribution of features

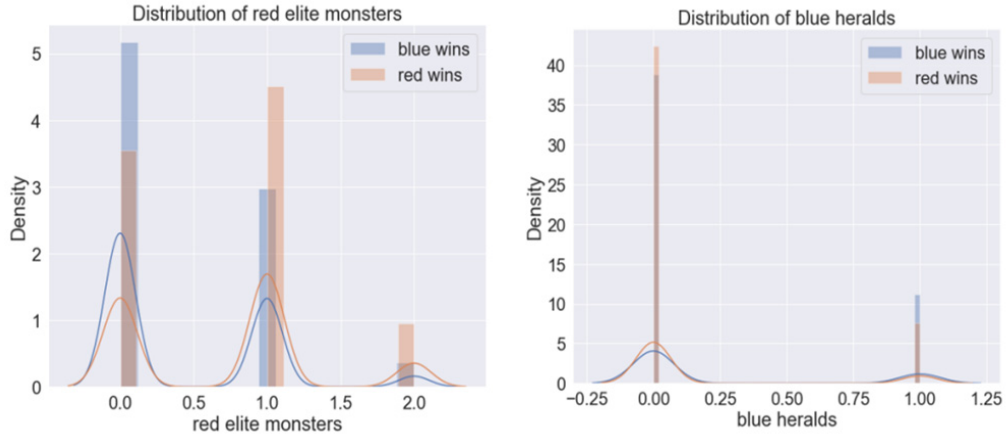

Figure 3. The distribution of features

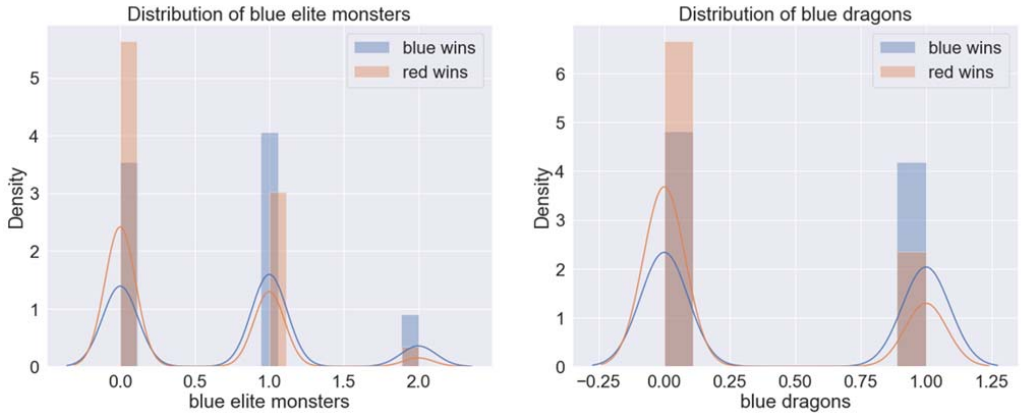

Figure 4. The distribution of features

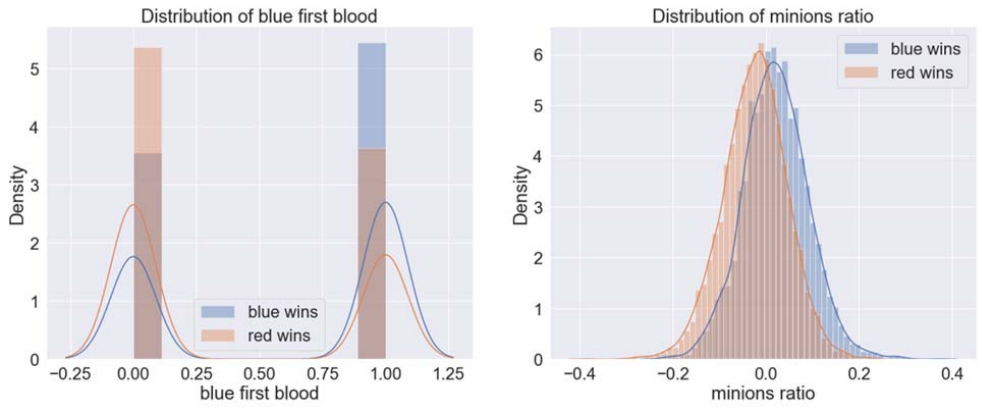

Figure 5. The distribution of features

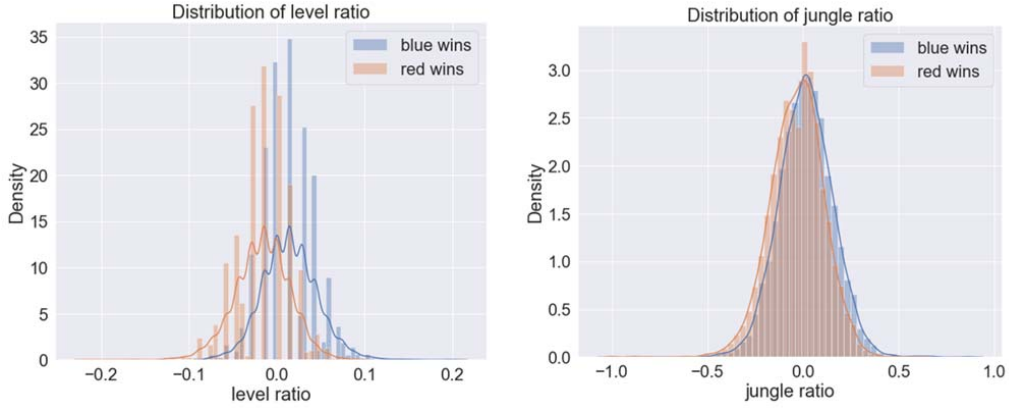

Figure 6. The distribution of features

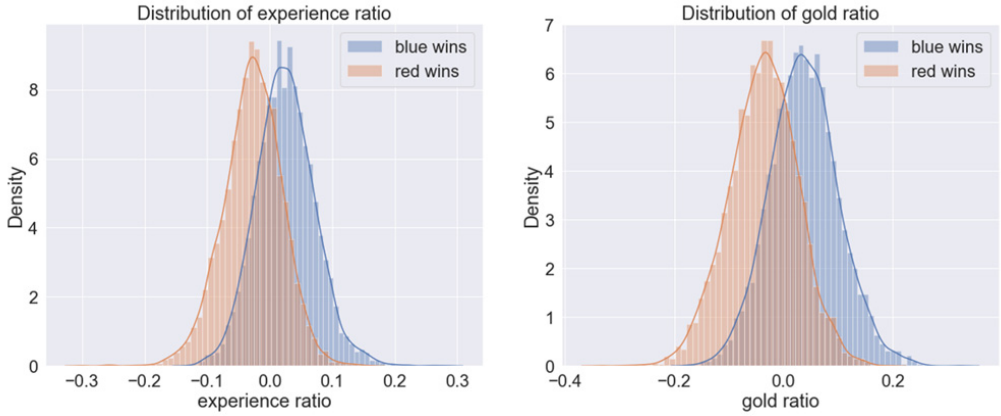

Figure 7. The distribution of features

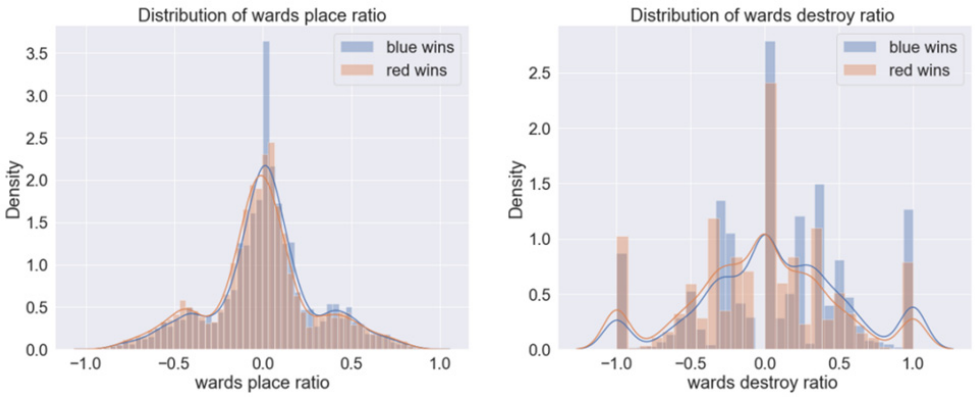

Figure 8. The distribution of features

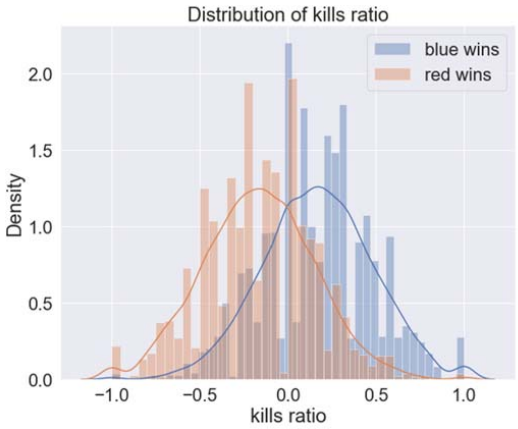

Figure 9. The distribution of features

From the figures above, the conclusion is that the level, experience, gold and kills ratio have a great influence on the result of the game. While for the none-standardized feature, the dragons, elite monsters, and first blood can influence the result of the game to some extent.

D. Machine learning

However, there are still 15 columns in the dataset which is quite a lot of data. In addition, from the visualizing result, there are some features less related to the result of the game which may decrease the accuracy of the model. So, the first thing before the machine learning part is dimensionality reduction. Firstly, a heatmap is made based on the correlation map between features and features shown below.

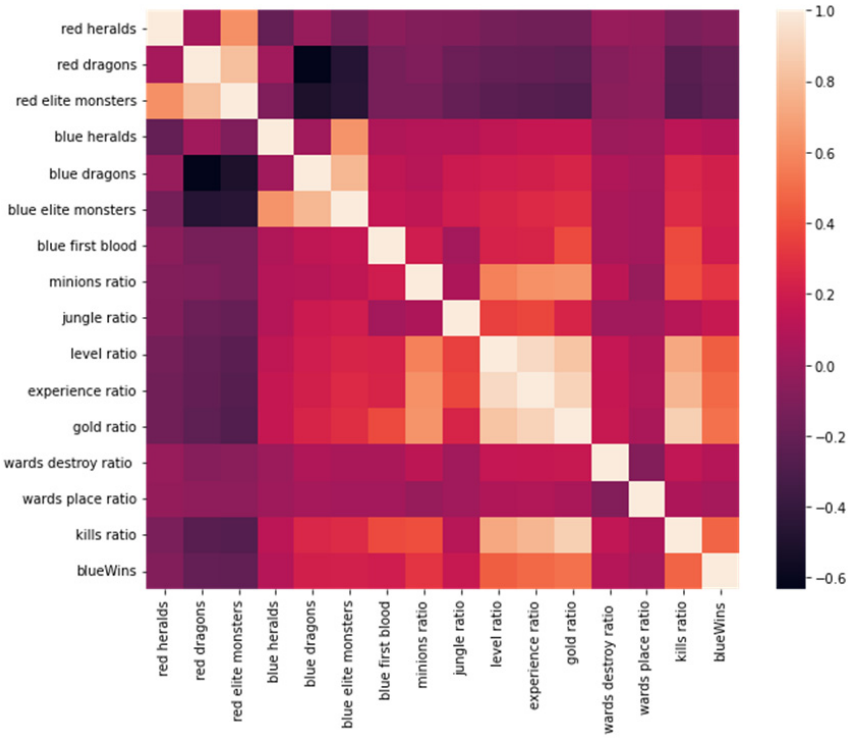

Figure 10. Heatmap of the correlation matrix among all the features

From the heat map above, the feature gold ratio, experience ratio, kills ratio, level ratio, and minions ratios have a stronger relationship with the feature blue wins. Then, the dataset is separated into train dataset and test dataset for cross-validation. After that, several popular machine learning models (AdaBoost, GradientBoost, RandomForest, ExtraTree, SVM, Naïve Bayes, KNN, LogisticRegression, and DecisionTree) were chosen to test the accuracy of the training result. Based on the result, several models will be chosen to make a voting classifier to make the result more precise. The result will be shown in the next part.

III. Result and Discussion

After training the model, the accuracy of the training is shown below as well as the confusion matrix.

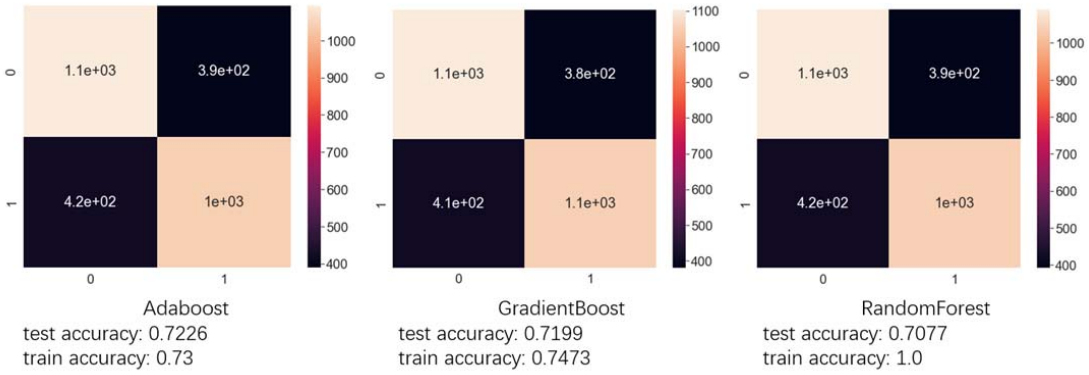

Figure 11. The training result of the chosen algorithm

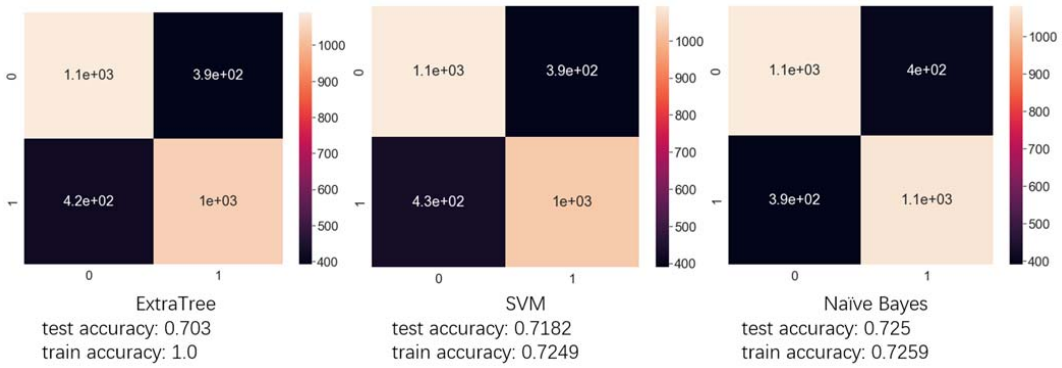

Figure 12. The training result of the chosen algorithm

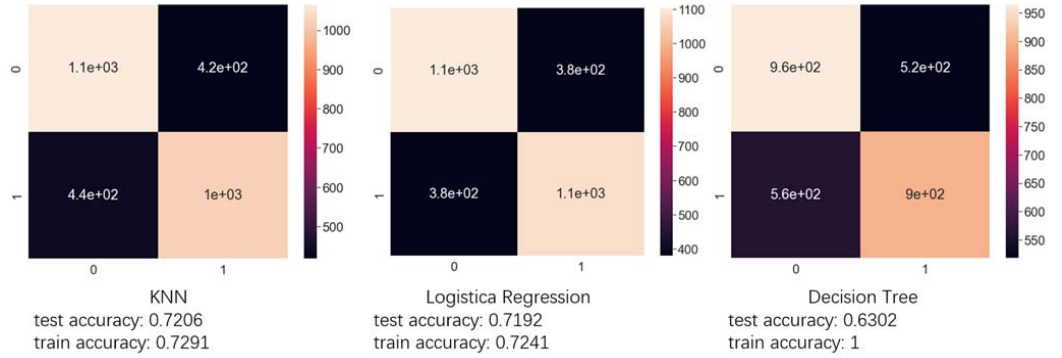

Figure 13. The training result of the chosen algorithm

From the figure above, KNN, Naïve Bayes, GradientBoost, and AdaBoost were chosen to make the Voting classifier. Moreover, to make the training result of the algorithm more precise, RandomForest is added into the Voting classifier. The training and test result is shown below.

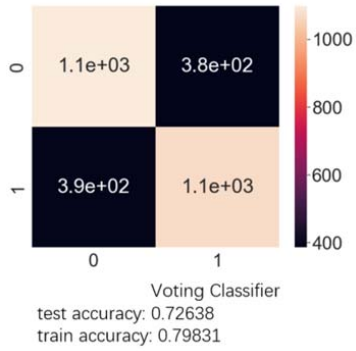

Figure 14. Training and test accuracy of the Voting Classifier

Judging from the confusion matrix about the voting classifier and the 9 algorithms above, it can conclude that the algorithm related to decision trees always performs well in the training dataset, but not so good in the test dataset which results in the situation of overfitting. In the algorithms related to decision trees, the random forest performs best. While in other algorithms the Naïve Bayes performs best in the test dataset. The Voting Classifier in the last step does improve the test and train accuracy of the whole model which is 0.72638 and 0.79831 respectively.

However, there are lots of champions in League of Legends who are weak in the early stage and dominant in the further stage. So, if only the first 10 minutes of data the result is analyzed, some errors may occur. So, this is the reason why the testing and training results can't be more precise. In addition, since the champions will not be fixed in one lane in League of Legends, this attribute will confuse the positions of opponents. Although the data about the pick and ban in diamond ranked games were used, based on this property, this data can't be used properly. So, there is no choice but to drop these features.

IV. Conclusion

The test accuracy of the voting classifier is 0.72638 which has spaces for improvement. For further research, a wider dataset that clearly records the data with every stage of the game will be applied in the research. A comparison is conducted on every stage to determine which factor plays the most important role in that stage. To analyze the relationship between different champions, the dataset will dig deep into the games which record the performance when facing another champion. An algorithm will then be developed to provide advice at any stage using the current data of the game.

Acknowledgment

Firstly, I am extremely grateful to professors in my university who give me helpful advice at the algorithm

selecting stage and help me solve some coding problems. Secondly, I am very grateful to our teaching assistant Kunyu An who helps me decide the topic of this research: a machine learning approach to analyze League of Legends. Last but not least, I must also thank my family members and friends who cheer me up when I met troubles and problems without their help I will not complete this research.

REFERENCES

[1] Mora-Cantallops, M., & Sicilia, M.-Á. (2018). MOBA games: A literature review. Entertainment Computing, 26, 128–138. https://doi.org/10.1016/j.entcom.2018.02.005

[2] Xia, B., Wang, H., & Zhou, R. (2019). What contributes to success in MOBA games? an empirical study of defense of the ancients 2. Games and Culture, 14(5), 498-522. https://doi.org/10.1177/1555412017710599

[3] Hodge, V. J., Devlin, S., Sephton, N., Block, F., Cowling, P. I., & Drachen, A. (2021). Win prediction in multiplayer esports: Live professional match prediction. IEEE Transactions on Games, 13(4), 368-379. https://doi.org/10.1109/TG.2019.2948469

[4] Akhmedov, K., & Phan, A. H. (2021, June 3). Machine learning models for DOTA 2 outcomes prediction. arXiv.org. Retrieved January 9, 2022, from https://arxiv.org/abs/2106.01782v1

[5] Do, T. D., Wang, S. I., Yu, D. S., McMillian, M. G., & McMahan, R. P. (2021). Using machine learning to predict game outcomes based on player-champion experience in League of Legends. The 16th International Conference on the Foundations of Digital Games (FDG) 2021. https://doi.org/10.1145/3472538.3472579

[6] Semenov, A., Romov, P., Korolev, S., Yashkov, D., & Neklyudov, K. (2017). Performance of machine learning algorithms in predicting game outcome from drafts in Dota 2. Communications in Computer and Information Science, 26–37. https://doi.org/10.1007/978-3-319-52920-2_3

[7] Costa, L. M., Mantovani, R. G., Monteiro Souza, F. C., & Xexeo, G. (2021). Feature analysis to league of legends victory prediction on the picks and bans phase. 2021 IEEE Conference on Games (CoG). https://doi.org/10.1109/cog52621.2021.9619019

[8] Silva, A. L. C., Pappa, G., & Chaimowicz, L. (1970, January 1). continuous outcome prediction of League of Legends Competitive matches using recurrent neural networks: Semantic scholar. Semantic Scholar. Retrieved January 10, 2022, from https://www.semanticscholar.org/paper/Continuous-Outcome-Prediction-of-League-of-Legends-Silva-Pappa/85a507426e3df16c0e5624fba0ad1f8572ce1874

[9] Lin, L. (2016). League of Legends Match Outcome Prediction. Retrieved January 10, 2022, from http://cs229.stanford.edu/proj2016/poster/Lin-LeagueOfLegendsMatchOutcomePrediction-poster.pdf

[10] Michel's fanboi, "League of Legends Diamond Ranked Games (10 min)", Kaggle, April 13, 2020, https://www.kaggle.com/bobbyscience/league-of-legends-diamond-ranked-games-10-min