CAPSTONE PROJECT

BETTING ON eSPORTS

Jonathan Pena

1. Problem Statement

This project aims to build a machine learning model that predicts eSport matches results and compares it to the odds provided by different betting websites so that bets can be placed in an optimal way.

2. Background

As Esports become more popular, many industries surrounding it have also grown in popularity. One of them is the eSports betting market which is expected to reach 205 billion dollars by 2027[1]. As internet access and investment increase, video games are becoming very popular, increasing live streaming of games and supporting infrastructure for tournaments. As a result of the lockdowns, gamblers also turned from traditional betting on live sports to eSports during the pandemic.

Taking advantage of this industry would be a fantastic opportunity, and that is what this project seeks to achieve. Many eSports are worth analyzing. Online games have the advantage of digital information, which can be updated more quickly and with greater accuracy than live games. This project will examine LoL games for 2021. This analysis was conducted using data collected by Tim Sevenhuysen from the Oracle Elixir website. It contains match data from different League of Legends leagues for 2021. It is updated daily in CSV format and is available for download. It includes information about 10934 matches, including team performance information. Our model was designed to identify if the red team is the winner of a game. In order to accomplish this, several models were built. The target distribution is balanced with 47% of winnings and 53% of losses.

Preprocessing of this dataset included filtering out all team information and matches played in patch 11, which is the version of the game that was used to conduct the official matches. Duplicates and missing values were also removed from the dataset. In exploratory data analysis (EDA), the distributions of the features were examined, most of which were close to normal. Additionally, characteristics of a winning team were identified. Teams that win tend to secure more goals through the game and have better teamwork.

1 https://www.prnewswire.com/news-releases/global-esports-betting-market-expected-to-exceed-205-billion-by-2027-301319074.html

3. Modeling Summary

First, I built a logistic regression model with the features in the dataset. Due to the number of features that are used to describe matches, we got a very high score in our test set. A model like this is only applicable to matches that are already over, but it cannot be used to predict future games.

Model	Accuracy Score
Logistic Regression	98.37%

This first model could be used by esport commentators to provide audience members with information such as probabilities for each team during the game. In this case, we can feed information into the model that has already passed to calculate the red team's chances of winning.

But our goal was to predict future games. To achieve our goal, we created features that help us to create a team profile based on previous games. In order to develop a KPI for the next game, we selected some of the indicators from the last ten games.

The following features were created:

- Kills per Minute: This feature divides the total number of kills made by a team in the last 10 games by the length of those games.
- Gold per Minute: This feature divides the total gold made by a team in the last 10 games by the length of those games.
- Minion Kills Per Minute: This feature divides the number of Minions kills made by a team in the last 10 games by the total length of those games.
- Towers Destroyed per Minute: This feature divides the total number of Towers destroyed by a team in the last 10 games by the total length of the games.

With those features, I could predict outcoming games. I built different classification models to find the one with the highest accuracy score. A summary of these results follows below.

Model	Mean Validation Accuracy Score
Logistic Regression	62.49%
Decision Tree	62.17%
Random Forest	62.81%
XGBoost	63.19%
KN Neighbors	57.72%

In comparison with our baseline, which is 47%, the Xgboost model gave us an accuracy score of 64% in the test set. This highlights the importance of feature engineering for predicting future games.

4. Conclusion

By building our model, we outperformed random guesses by 17%. We learned that feature engineering is crucial for building better models. Now we have a model that can help us place bets more effectively. We still need to improve our model. In order to describe a winning team more precisely, more features can be added.

The scope of the project is very broad; I have examined one major eSport, but there are many others. It is possible to automate the process of getting information and comparing the results with the bookers. It is possible to improve each model by taking into account additional factors and generating more attributes that are better suited to describe a winning team. Last but not least, I would like to explore other aspects of the eSport industry, such as clustering the player's behavior to find the next star or creating tools to coach the players.