Determining eSports Team Potential Investment Value Using Machine Learning

Project Statement of Work V1

AI Algorithms – AIDI 1002

Course Facilitator – Marcos Bittencourt

Izza Godinez (100556078)

December 1, 2020

Executive Summary

In recent years, eSports has been growing into a multi-million dollar industry [2, 5]. Many people and companies are looking to invest in this industry, but due its relatively new nature, there is very little information about which teams or players to invest in. This provides a need for a system which can assign values to teams based on performance and popularity to determine which teams are worth investing in. In this project, I propose using data from players' past performance in games to create a model that will assign determine if a player has a high chance of being successful in professional games. The final deliverable will be a model where the user can input information on a player, such as their game performance history, and the model will be able to determine if the player will have a good chance of being successful in professional games.

Rationale Statement

Using artificial intelligence, we can determine which League of Legends eSports teams are the most valuable for investors. This can help the business grow even more with proper funding and can be helpful for those looking to invest their money.

Problem Statement

eSports is a growing industry, especially in recent years. One of the most prominent eSports games is League of Legends, whose championship events and tournaments bring in hundreds of thousands of viewers every year. For these events to continue to grow, there needs to be investment and sponsorships from people and companies. Currently, popular investors do not have much experience or expertise in eSports due to it being a relatively new and somewhat niche industry. Therefore, it is important for these people and companies to know which teams to choose that will bring in the most value to their investment.

In this project, I propose using Artificial Intelligence in predicting which teams have the most potential and bring the most value for investors. This can be based on many different factors, such as the team's history in tournaments and sporting events, the following of each team on various social media websites, the average viewership of players when they livestream, the viewership of each team's events, and other metrics. This would require plenty of data from different sources about each team eSports team, and even about each individual player since the players on the team could affect the revenue and competition results.

Data Requirements

For this project, a dataset is required that has different metrics that can be correlated to the success that each player might bring. In order to develop a model, it is important to have information on different players' performances in games and which factors are the highest contributors to winning a game. For the purpose

of this project, there are limitations that we need to impose on the dataset to keep the scope realistic. Some assumptions and limitations that will be added will be that the model will only look at players of the same region so that the dataset will not be too large. Also, this will make things more streamlined since players from different regions tend to have differing playstyles, which means that the factors that contribute to a win might be different depending on the region. We will also be looking at players who play the same position, so it is easy to compare each player with other similar players.

Dataset

The dataset in this project comes from a website called Oracle's Elixir [6], which provides more in-depth information about the games and teams. This includes information of every game played during the season from 2014 to 2020. The dataset includes information on each player, what roles they play, and various statistics that may be analyzed to determine if it is a contributing factor to a winning result.

Data Exploration and Analysis

Each year from 2014-2020 had its own dataset file, which needed to be cleaned up by relevant data and combined into one large dataset. Firstly, the data was all combined into one large dataset, so that all data cleaning only needed to be performed once. The data contains information that is relevant to each player, as well as some information that relates to the whole team. Since it is common for players to transfer between teams, we will only be looking at individual player performance.

This dataset had 105 columns of different statistics for each game, so this needed to be reduced. Some of the information, such as the statistics relating to opponent objectives or opponent kills were not relevant to the goal, so these columns were dropped. Although it is important to know the opponent statistics to determine the context of the game, this can be determined more effectively by using the differences between the player and the opponent. Some columns also contain information that are only relevant to the team as a whole, so these columns were removed.

Next, we specified the dataset so that it only shows data from the North American region, and only for players that play the "mid" position. Factors that relate to a win for a mid position player may differ from other positions, so for different positions, new models may need to be trained.

Finally, we look at which factors are more likely to contribute to a win by creating a correlation matrix. These factors include:

- Earned gold per minute

- Team kills per minute

- Team kills

- Assists

- Kills

- Earned gold

- Inhibitors

- Double kills

- Gold difference at 15

- Total gold

- Team deaths

- Deaths

All of these columns were put into a new dataframe, along with the result of the game. This dataframe can be further used when training the model.

Testing

In order to verify the plausibility of this project, we will need to determine if the results are accurate. The outcome of this project will be a result of the value of the player in terms of performance. Although there are other factors of success, performance and the number of wins are some of the most easily measurable metrics. The data will be divided into training and testing data so that it can later be compared to each other to determine the accuracy of the model. The user will then be able to input information about a new player and compare them to the players from the model database to determine whether that player has a higher or lower chance of winning the game. This will allow the user to make a decision on if they would like to invest in the player.

Project Plan

From this point in the project schedule, the next steps toward project completion are as follows:

1. Researching available algorithms and models related to predicting sports performance.
2. Performing initial exploratory data analysis on the datasets to make sure they are in the same format

3. Perform feature analysis to determine which metrics are most correlated to the success and performance of each player
4. Finalize which model will be used to score the value of each player depending on performance and popularity
5. Train model using the datasets collected to create a model that can predict the performance based on player
6. Test data against dataset to determine accuracy and see if there are any flaws in the model or if any improvements can be made to the accuracy
7. Tune the model to fix and flaws
8. Continue training, testing, and analyzing the model until the error is a reasonably small amount
9. Evaluate the model against other datasets, such as teams from other regions

Throughout the process, the model will also be evaluated to determine if the model is appropriate or if there are any flaws in the model. If there are any problems, other models may be considered and tested to see if they provide more accurate results.

References:

[1] Escharts.com. "Esports Teams and Events Stats: Dota 2, LoL, CS:GO." Esports Teams and Events Stats: Dota 2, LoL, CS:GO | Esports Charts, escharts.com/.

[2] Gough, Christina. "Global ESports Market Revenue 2023." Statista, 13 Oct. 2020, www.statista.com/statistics/490522/global-esports-market-revenue/.

[3] King, Jim. "Using Machine Learning to Predict Fantasy Football Points." Medium. Towards Data Science, 22 June 2020. Web. 02 Dec. 2020.

[4] "League of Legends (LoL) Data Feed API." SportsDataIO, sportsdata.io/developers/data-dictionary/lol.

[5] "League of Legends Esports Deep Dive: Worlds Viewership and How Format Changes Contributed to the LEC's Growth." Newzoo, 18 Dec. 2019, newzoo.com/insights/articles/league-of-legends-esports-deep-dive-worlds-viewership-and-how-format-changes-contributed-to-the-lecs-growth/.

[6] Oracle's Elixir - LoL Esports Stats, oracleselixir.com/.

[7] Twitch Tracker, https://twitchtracker.com/