Exploring Key Variables Influencing the Outcome of Esports Matches

Submitted by

M. Hisham Zahid
23L-7827

Supervised by

Dr. Asif Mahmood Gilani

A thesis submitted in partial fulfillment of the requirements for the degree of
Masters (Computer Science)
at National University of Computer & Emerging Sciences

School of Computing
National University of Computer & Emerging Sciences

Lahore, Pakistan.

September 2024

Plagiarism Undertaking

I take full responsibility of the research work conducted during the Masters Thesis titled *Exploring Key Variables Influencing the Outcome of Esports Matches*. I solemnly declare that the research work presented in the thesis is done solely by me with no significant help from any other person; however, small help wherever taken is duly acknowledged. I have also written the complete thesis by myself. Moreover, I have not presented this thesis (or substantially similar research work) or any part of the thesis previously to any other degree awarding institution within Pakistan or abroad.

I understand that the management of National University of Computer and Emerging Sciences has a zero tolerance policy towards plagiarism. Therefore, I as an author of the above-mentioned thesis, solemnly declare that no portion of my thesis has been plagiarized and any material used in the thesis from other sources is properly referenced. Moreover, the thesis does not contain any literal citing of more than 70 words (total) even by giving a reference unless I have the written permission of the publisher to do so. Furthermore, the work presented in the thesis is my own original work and I have positively cited the related work of the other researchers by clearly differentiating my work from their relevant work.

I further understand that if I am found guilty of any form of plagiarism in my thesis work even after my graduation, the University reserves the right to revoke my Masters degree. Moreover, the University will also have the right to publish my name on its website that keeps a record of the students who plagiarized in their thesis work.

M. Hisham Zahid

Date: ____________________

Abstract

Dota 2, a premier title in the esports landscape, presents a complex and strategically rich environment where the interplay between team compositions, individual player performances, and in-game decisions critically impacts match outcomes. As the game's meta evolves, understanding the key variables that contribute to victory has become increasingly important for teams, analysts, and the broader esports community. This research leverages a dataset of approximately 180,000 professional matches (from 2014 to 2024) to analyze these variables, employing machine learning and statistical methods to uncover patterns and predictors of success. By developing predictive models that incorporate both pre-game factors (such as hero drafts) and in-game dynamics (such as player metrics, first bloods, and net worth), this study aims to provide actionable insights that can enhance decision-making in drafting and gameplay. Although centered on Dota 2, the variables and methods explored in this research are common across many MOBA (Multiplayer Online Battle Arena) games, making the findings potentially relevant to a broader context. The results have the potential to inform better strategies and contribute to the ongoing development of data-driven approaches in competitive esports.

Contents

Chapter 1

Introduction

Dota 2, a leading title in the esports industry, presents a complex environment where team compositions, player actions, and in-game decisions significantly impact match outcomes. This research aims to analyze these factors by leveraging a comprehensive dataset of professional Dota 2 matches, applying advanced machine learning and statistical techniques to predict match results. Although the focus is on Dota 2, the variables examined—such as player performance metrics, in-game events, and team dynamics—are common across other MOBA (Multiplayer Online Battle Arena) games, making the findings potentially relevant and applicable beyond just Dota 2.

1.1 Importance

Predicting match outcomes in Dota 2 holds immense strategic value. Accurate predictions can guide key decisions during drafting, team coordination, and in-game strategy, offering teams a competitive edge. As the esports industry grows, data-driven approaches like those explored in this research are becoming essential for optimizing team performance.

1.2 Motivation

The increasing complexity and competitiveness in professional esports highlight the need for data-driven decision-making tools. The potential to enhance team performance through predictive modeling motivates this research, which seeks to bridge the gap between raw data and actionable insights in a high-stakes environment.

1.3 Problem Statement

The central problem addressed by this research is the development of a robust predictive model for determining match outcomes in professional Dota 2 games. This model must account for the multifaceted and dynamic nature of the game, incorporating various in-game variables and performance metrics.

1.4 Research Questions

This research seeks to answer the following questions:

- What are the key variables that most significantly impact match outcomes in Dota 2?
- How can machine learning models be optimized to predict match results accurately?
- Can the predictive models developed for Dota 2 be generalized to other MOBA games?

Chapter 2

Literature Review

Research into outcome prediction in Dota 2 and similar games has utilized various machine learning models and statistical methods, with an emphasis on identifying critical factors that influence match outcomes.

2.1 Related Works

Research into outcome prediction in Dota 2 and similar games has explored a variety of machine learning models and statistical methods. Previous studies have demonstrated that variables like team compositions, early-game performance, and player statistics can be powerful predictors of match outcomes.

Several studies have applied machine learning models to predict outcomes in Dota 2 matches. Akhmedov and Phan (2021) [1] focused on building predictive models using Linear Regression (LR), Neural Networks (NN), and Long Short-Term Memory (LSTM) networks to identify match outcomes in real-time. The study utilized a data collection server through Game State Integration (GSI) to track real-time player data. After performing exploratory feature analysis and hyperparameter tuning, the models were tested on players with varying levels of experience. The study found that the LSTM model achieved the highest accuracy, with an average prediction accuracy of 93 percent, followed by NN with 88 percent, and LR with 82 percent in the best-case scenario.

Similarly, Wang et al. (2021) [2] explored quantifying heroes based on 17 different attributes and introduced a priority table for hero selection. This study offered a more nuanced method for modeling team compositions. The evaluation of various machine learning methods demonstrated that this approach significantly enhanced the accuracy.

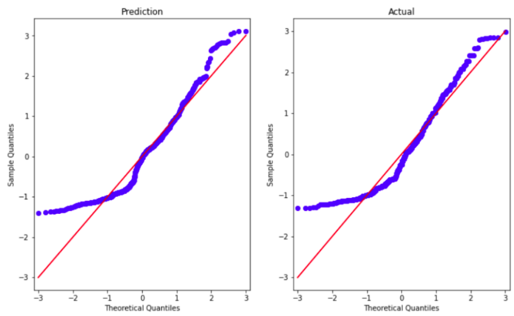

Figure 2.1: Performance of the Linear Regression model

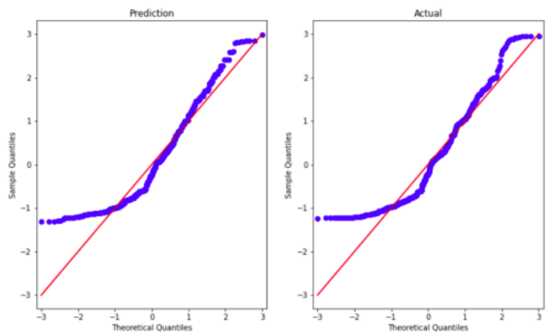

Figure 2.2: Performance of the Long Short-Term Memory (LSTM) model

Predictive modeling in professional Dota 2 matches often requires understanding the complexities of hero selection and its impact on game outcomes Zhang et al. (2020) [3] tackled this challenge by developing an improved bidirectional Long Short-Term Memory (LSTM) neural network model specifically designed for Dota 2 lineup recommendations. This model leveraged the Continuous Bag Of Words (CBOW) approach within the Word2Vec framework to generate hero vectors, where a ”word” is represented as a hero and a ”sentence” as a lineup. By predicting the final hero selection based on the first four heroes chosen, the model offered a method to optimize team composition strategies. This approach aligns well with the goals of professional-level predictive modeling, where understanding and predicting the impact of team composition is crucial. The study highlights the potential for advanced machine learning techniques to guide strategic decisions in real-time.

Song et al. (2019) [4] developed a predictive model using logistic regression and decision trees to determine the winning side of matches, highlighting the role of metrics like gold and experience differences. Similarly, Wang and Shang (2017) [5] utilized a Na’ive Bayes classifier to predict match outcomes.

2.2 Research Gap

While existing studies on predictive modeling in Dota 2 have made significant contributions, many focus primarily on specific in-game metrics, often derived from matchmaking data. This approach, while valuable, may not fully capture the dynamics of professional play, where the stakes and strategies can differ significantly. Additionally, while these models are often tailored specifically to Dota 2, there is an opportunity to explore the broader applicability of the variables involved, which are relevant across the MOBA genre.

This research seeks to address these areas by focusing on a dataset composed of professional Dota 2 matches. Although the dataset is specific to Dota 2, the general nature of the variables it includes—such as player performance metrics, team dynamics, and in-game events—suggests potential for broader applicability within the MOBA genre. This study aims to develop models that are not only robust in the context of Dota 2 but also potentially generalizable to other games within the MOBA genre.

Chapter 3

Data

3.1 Dataset Description

The dataset used for this research consists of a comprehensive collection of approximately 180,000 professional Dota 2 matches, each represented by a wide array of variables. These variables encompass various aspects of the game, from match metadata to detailed player performance metrics. The columns in this dataset are selected based on the general presence across all games in the MOBA genre. The dataset includes the following key columns:

- **league_id, league_tier, league_start_date_time, league_end_date_time:** These columns provide information about the league in which the match took place, including the league's unique identifier, its competitive tier, and the start and end dates of the league.
- **series_id, series_type:** These fields denote the series in which the match occurred, including the series' unique identifier and type (e.g., best of three).
- **match_id, match_start_date_time, match_duration_seconds:** This set of columns captures the specific match details, including the unique match identifier, the start time of the match, and its duration in seconds.
- **first_blood_time_seconds:** This column records the time, in seconds, at which the first blood (first kill) occurred during the match.
- **radiant_team_id, radiant_team_name, dire_team_id, dire_team_name, winner_id:** These columns provide details about the teams involved in the match, including their identifiers, names, and the winning team.

- **radiant_kills, dire_kills:** The total number of kills achieved by the Radiant and Dire teams during the match.
- **radiant_player_X_id, radiant_player_X_name, radiant_player_X_hero_id, radiant_player_X_hero, radiant_player_X_position, radiant_player_X_lane, radiant_player_X_role, radiant_player_X_kills, radiant_player_X_deaths, radiant_player_X_assists, radiant_player_X_networth:**
  These fields detail the performance metrics of each of the five players on the Radiant team. They include each player's ID, name, chosen hero, in-game position, lane, role, and statistics such as kills, deaths, assists, and net worth.
- **dire_player_X_id, dire_player_X_name, dire_player_X_hero_id, dire_player_X_hero, dire_player_X_position, dire_player_X_lane, dire_player_X_role, dire_player_X_kills, dire_player_X_deaths, dire_player_X_assists, dire_player_X_networth:**
  Similar to the Radiant team, these columns capture the same performance metrics for each of the five players on the Dire team.
- **game_version_id:** This column indicates the version of the game during which the match was played, which is essential for understanding the context of the match, as gameplay mechanics and balance can change between versions.

The dataset for this research includes various columns that are critical for analyzing match outcomes in Dota 2, yet these variables are equally applicable to other popular MOBAs, such as League of Legends and Mobile Legends: Bang Bang (MLBB). The following key variables exemplify their universality across the genre:

3.1.1 Hero and Player Metrics

Each match records player-specific data, including kills, deaths, assists, and net worth. These metrics reflect individual performance, which is a cornerstone in all MOBAs. In League of Legends, similar metrics are used to evaluate champions, indicating their impact on the game. Studies have shown that a player's performance is closely correlated with match outcomes, regardless of the specific game being analyzed.

3.1.2 Team Composition and Strategy

Variables related to team composition, such as player roles and positions, influence strategic decisions. Each MOBA employs a diverse roster of characters, where the selection of heroes can significantly sway match dynamics. Research has highlighted that effective team compositions lead to higher win rates, demonstrating a common strategic framework across different MOBAs.

3.1.3 Match Context

The inclusion of match context variables, such as league tier and match duration, is vital. Competitive play in MOBAs is characterized by distinct tiered leagues, influencing the strategies and performance metrics of players. The significance of contextual data in predicting match outcomes is consistent across MOBAs, establishing a framework for analysis that extends beyond Dota 2.

sectionResearch Supporting Variable Relevance Research in the esports domain has explored these key variables, underscoring their significance in predicting match outcomes. For instance:

- One study identified the following key variables [1]:
  - **Player Statistics**: Kills, deaths, assists, gold per minute, experience per minute.
  - **Team Dynamics**: Team composition, hero synergy, and player roles.
  - **Game Events**: First blood, tower destructions, Roshan kills.
- Another study on League of Legends highlighted these variables [6]:
  - **Team Structures**: Number of towers and inhibitors destroyed.
  - **Player Interactions**: Number of kills, deaths, and assists.
  - **Neutral Objectives**: First Baron Nashor, first drake, and Rift Herald kills.
  - **Gold Earned**: Total gold earned by the team.
- It is important to note that neutral objectives like Baron Nashor, Roshan and Dragons etc differ across multiple games in terms of type and quantity. Therefore, to ensure our model is generalized and applicable across different MOBA games, we will not include these variables.

After going through these papers we find out the general variables that are involved in MOBAs and effect the outcome of matches confirming our dataset is generalized for different MOBAs and not just Dota 2.

This research emphasizes the importance of these variables not just within Dota 2 but also across the broader MOBA landscape, underscoring their relevance in informing predictive modeling approaches that could be beneficial for developers and analysts within the genre.

3.2 Report on Classification Model Performance

3.2.1 Overview

In this report, we present the performance metrics of four different classification models used to predict match outcomes in MOBA games. The models tested include Logistic Regression, Decision Tree and Random Forest. h.

3.2.2 Model Performance

1. Logistic Regression

- **Accuracy:** 98.71%
- **Classification Report:**
  - Precision: 0.99 (Radiant), 0.99 (Dire)
  - Recall: 0.99 (Radiant), 0.99 (Dire)
  - F1-Score: 0.99 (Radiant), 0.99 (Dire)
  - Support: 10,617 (Radiant), 10,752 (Dire)

Logistic Regression achieved very high accuracy and balanced performance across both classes, indicating a robust model with excellent ability to differentiate between winners.

2. Decision Tree

- **Accuracy:** 91.66%
- **Classification Report:**
  - Precision: 0.92 (Radiant), 0.91 (Dire)
  - Recall: 0.91 (Radiant), 0.92 (Dire)
  - F1-Score: 0.92 (Radiant), 0.92 (Dire)
  - Support: 10,617 (Radiant), 10,752 (Dire)

The Decision Tree model performed well but was less accurate compared to Logistic Regression. The precision and recall scores are still quite strong, though slightly less balanced.

3. Random Forest

- **Accuracy:** 96.57%
- **Classification Report:**
  - Precision: 0.97 (Radiant), 0.97 (Dire)
  - Recall: 0.97 (Radiant), 0.97 (Dire)
  - F1-Score: 0.97 (Radiant), 0.97 (Dire)
  - Support: 10,617 (Radiant), 10,752 (Dire)

Random Forest demonstrated high accuracy and very balanced performance, making it one of the best models in this evaluation for predicting match outcomes.

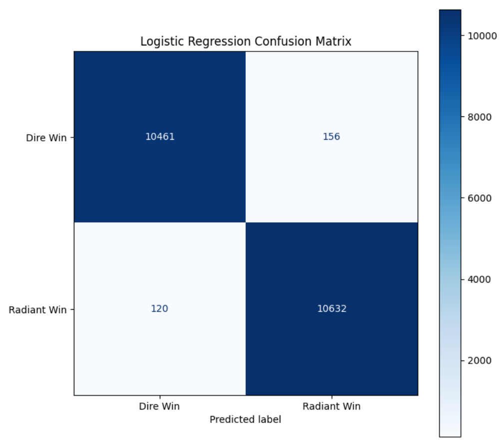

Figure 3.1: Performance of the Logistic Regression model

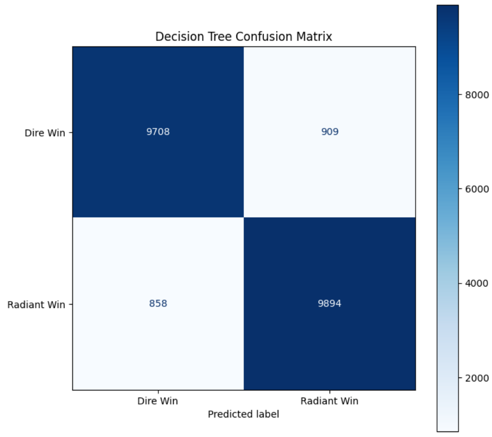

Figure 3.2: Performance of the Decision Tree model

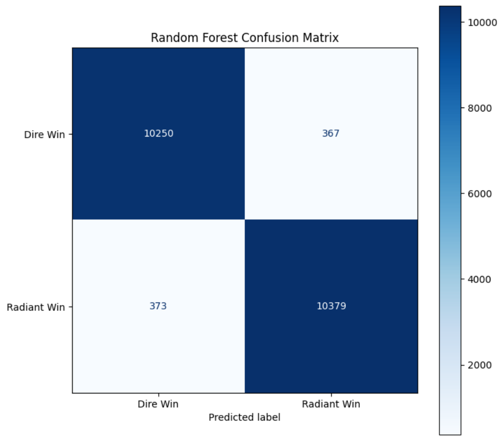

Figure 3.3: Performance of the Random Forest model

References

[1] K. Akhmedov and A. H. Phan, "Machine learning models for dota 2 outcomes prediction," *arXiv preprint arXiv:2106.01782*, 2021.

[2] N. Wang, L. Zhang, F. Li, and X. Liu, "Outcome prediction of dota 2 using machine learning methods," in *Proceedings of the 2018 International Conference on Mathematics and Artificial Intelligence*, (Beijing, China), 2018.

[3] L. Zhang, C. Xu, Y. Gao, Y. Han, X. Du, and Z. Tian, "Improved dota 2 lineup recommendation model based on a bidirectional lstm," *Tsinghua Science and Technology*, vol. 25, no. 6, pp. 712–720, 2020.

[4] K. Song, T. Zhang, and C. Ma, "Predicting the winning side of dota 2," *Journal of Computer Science and Technology*, vol. 34, no. 4, pp. 567–580, 2019.

[5] K. Wang and W. Shang, "Outcome prediction of dota 2 based on naïve bayes classifier," in *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, (Kobe, Japan), pp. 123–130, IEEE, 2017.

[6] A. M. Castellanos and G. B. Corps, "Variables related to the outcome of an esports professional tournament: Case study of 2019 league of legends world championship series," *International Journal of Esports*, vol. 2, no. 2, 2021.