Performance Consistency and Playstyle Analysis of a Dota 2 Player Using OpenDota Match Data

Renaire Odarve
College of Computing and Information Technology
National University
Manila, Philippines
odarverb@students.national-u.edu.ph

***Abstract*—This study analyzes the performance consistency of a Dota 2 player using OpenDota match data collected between November 16, 2025, and February 6, 2026. The overall win rate was 51.52%. Winning matches exhibited higher mean values for KDA (4.58 vs. 2.05), gold per minute (539.67 vs. 455.98), and experience per minute (769.94 vs. 654.56). Independent-samples *t*-tests confirmed statistically significant differences in KDA ($t = 6.892$, $p < 0.001$), gold per minute ($t = 2.869$, $p = 0.005$), and experience per minute ($t = 2.672$, $p = 0.009$). Correlation analysis indicated a moderate relationship between win outcome and KDA ($r = 0.564$) and weaker correlations with gold ($r = 0.280$) and experience ($r = 0.262$), suggesting that combat effectiveness is the strongest contributor to match success.**

***Index Terms*—Dota 2, OpenDota, data science, exploratory data analysis, player performance analysis, gaming analytics, descriptive statistics, data visualization**

I. Introduction

Dota 2 is a multiplayer online battle arena (MOBA) game developed and published by Valve Corporation and released in 2013. It is one of the most played esports titles and provides detailed in-game statistics that allow systematic performance analysis. Evaluating player performance and playstyle is relevant for skill assessment, competitive analysis, and esports research. Public data platforms such as OpenDota collect match statistics from the official Dota 2 application programming interfaces (APIs) and make them available for analysis. OpenDota provides match histories and key performance metrics, including kills, deaths, assists, gold per minute (GPM), and win–loss outcomes. Documentation and structured data access are available through the OpenDota website [1] and API documentation [2].

The objective of this study is to explore and quantify how a single player's performance varies over a selected period and how different playstyle factors relate to match outcomes. This research applies descriptive statistics, exploratory data analysis (EDA), and data visualization techniques to provide insights into performance consistency across matches and to identify patterns in playstyle associated with wins and losses.

Specifically, the study aims to:

1) Evaluate the consistency of the selected player's performance over a one-week period.
2) Analyze key performance indicators (KPIs), including kill/death/assist ratios, gold per minute, experience per minute, and match duration.
3) Examine relationships between playstyle elements and match outcomes.
4) Present visual and statistical evidence highlighting performance trends over time.

By leveraging publicly available match data from OpenDota, this study provides insights into how performance metrics relate to match success.

II. Review of Related Works

Esports analytics has grown rapidly due to the availability of gameplay data and the strategic complexity of games like Dota 2. Many studies use data-driven methods to analyze player performance and predict match outcomes, showing that in-game metrics such as kills, gold, and experience can forecast results with high accuracy [3].

Professional esports teams, such as Team Liquid, also use analytics to review replays, track key performance indicators, and guide coaching, highlighting the value of descriptive analytics in competitive strategy [4]. Bibliometric analyses further show that esports research spans multiple fields, including computer science, sports management, and psychology, with growing interest in approaches beyond predictive models [5].

Despite this, few studies focus on descriptive, exploratory analyses of individual player performance over time using public platforms like OpenDota. This study addresses this gap by applying descriptive statistics and visual analysis to a week of match data, providing insights into playstyle and performance trends that complement predictive research.

III. Methodology

This section describes the complete data preparation and preprocessing workflow implemented for the study. The methodology follows a structured data science pipeline to ensure reproducibility and transparency. All preprocessing and analysis procedures were conducted using Python 3.12.12 within Visual Studio Code 1.109.2.

A. Participants

The participant in this study is an individual who provided informed consent for their gameplay data to be used for research purposes. The analysis was conducted using the participant's publicly available Dota 2 match records obtained from the OpenDota platform through its publicly accessible application programming interface (API).

No private or sensitive personal information was collected. The dataset was strictly limited to publicly available gameplay statistics, and the participant's identity was anonymized to ensure confidentiality.

B. Data Collection Methods

Match data were collected from the OpenDota platform, which provides structured Dota 2 match statistics through its official API. The dataset included match identifiers, hero selections, performance metrics, gold and experience statistics, match duration, and win/loss outcomes. Data extraction was performed programmatically to ensure consistency and reduce manual errors.

1) Variables Collected: The variables collected for this study are summarized in Table I on the next page. The table includes player identifiers, match metrics, hero and item information, and performance statistics.

2) Frequency of Data Logging: Match data were collected on a weekly basis from November 16, 2025, to February 6, 2026. Each week's matches were downloaded in comma-separated values (CSV) format and subsequently merged into a consolidated dataset for analysis.

3) Tools and Platforms Used: The following tools and platforms were utilized in this study:

1) VS Code 1.109.2 (computational environment)
2) Python 3.12.12 (programming language)
3) Pandas 2.3.3 (data manipulation)
4) NumPy 2.3.5 (numerical computing)
5) Matplotlib 3.10.0 (data visualization)
6) Seaborn 0.13.2 (statistical visualization)
7) SciPy 1.16.3 (scientific computing)
8) Requests 2.32.5 (HTTP requests)
9) Built-in Python modules: os, glob, re

C. Operational Definitions

To ensure clarity and replicability, the following operational definitions were applied.

1) Kills = Times the enemy heroes died by the player.
2) Deaths = Times the player died in a match.
3) Assists = Times the player helped killing an enemy
4) Gold per Minute (GPM) = Average gold earned per minute during a match.
5) Experience per Minute (XPM) = Average experience gained per minute during a match.
6) Kills per Minute (KPM) = Average kills achieved per minute.
7) Match Duration = Total length of the match in seconds.
8) Win Outcome = Binary indicator (1 = Win, 0 = Loss).

A derived metric was computed.

$$KDA = \frac{\text{kills} + \text{assists}}{\text{deaths}}$$

Fig. 1. KDA Formula

KDA serves as a composite performance indicator representing combat efficiency.

Performance consistency was measured using variability metrics such as standard deviation and coefficient of variation.

D. Data Cleaning and Preprocessing

Data preprocessing was conducted prior to analysis to ensure accuracy and reliability.

1) Handling Missing Values: Missing values were addressed based on the observed data and column relevance:

- A comprehensive check was performed to calculate the missing count and percentage for each column, displaying columns sorted by descending missing count and identifying which columns contained any missing values.
- Critical columns, such as `kills_per_min`, were handled by dropping rows with missing values to preserve the integrity of key performance metrics. The number of removed rows was tracked.
- Non-essential or redundant columns were removed manually, including `name`, `last_login`, `party_id`, `party_size`, `permanent_buffs`, `region`, `account_id`, `personaname`, `hero_variant`, `start_time_iso`, and others.

This approach ensured that missing data did not compromise the analysis of key performance metrics while retaining relevant columns for exploratory and statistical analysis.

2) Datetime Conversion: The `start_time` column was converted to `datetime` format to enable chronological and weekly trend analysis.

3) Feature Engineering: The KDA ratio was computed as a new feature. No extreme outliers were removed unless values were logically inconsistent.

The cleaned and consolidated dataset was stored as `combined_player_data.csv`.

E. Statistical Analysis

The following statistical methods were applied to evaluate player performance and match outcomes.

1) Descriptive Statistics: Descriptive statistics were computed to summarize central tendency and dispersion. The following measures were calculated:

- Mean
- Median
- Standard deviation
- Minimum and maximum values

These statistics were computed for the following variables:

- kills
- deaths
- assists
- gold_per_min

TABLE I
COLLECTED DOTA 2 MATCH VARIABLES

Player Info	Hero Info	Items / Purchases	Performance Metrics
account_id	hero_id	item_0	actions_per_min
player_slot	hero_variant	item_1	life_state_dead
party_id		item_2	cosmetics
permanent_buffs		item_3	
party_size		purchase_time	
team_number		first_purchase_time	
team_slot		item_win	
		item_usage	
		purchase_ward_observer	
		purchase_ward_sentry	
		purchase_tpscroll	

- xp_per_min
- kills_per_min
- KDA

2) Time-Series Analysis: Matches were ordered chronologically using the `start_time` variable to analyze weekly performance trends. The following metrics were evaluated over time:

- Average KDA
- Win rate
- Gold per minute

Line plots were generated to assess performance consistency and trend patterns across the observation period.

3) Comparative Analysis: Performance metrics were grouped according to match outcome (win or loss). Independent-samples t-tests were conducted to determine whether differences between wins and losses were statistically significant, with Cohen's d values indicating the magnitude of these differences.

4) Correlation Analysis: Pearson correlation coefficients were computed to measure relationships among key performance variables, including:

- KDA
- gold_per_min
- xp_per_min
- kills_per_min
- win outcome

A correlation matrix was generated to identify the performance metrics most strongly associated with match success.

5) Variability Measurement: The coefficient of variation (CV) was calculated to quantify performance stability across matches and to assess relative variability in key metrics.

IV. RESULTS

This section presents the statistical findings derived from the cleaned dataset `combined_player_data.csv`. The results are organized into descriptive statistics, time-series analysis, correlation analysis, and hypothesis testing. Interpretation of these findings is reserved for the Discussion section.

A. Dataset Overview

The final dataset consists of consolidated weekly match records collected between November 16, 2025, and February 6, 2026. Each row represents a single match played by the participant.

The dataset includes the following primary variables:

- Match identifiers and timing: `match_id`, `start_time`, `duration`
- Player and hero information: `player_slot`, `hero_id`, `hero_name`
- Performance metrics: `kills`, `deaths`, `assists`, `gold_per_min`, `xp_per_min`, `kills_per_min`
- Items carried: `item_0` through `item_5`, `backpack_0` through `backpack_2`, `item_neutral`, `item_neutral2`, along with their corresponding mapped item name columns
- Match outcome: `win` (derived from `radiant_win`)

This dataset structure enables comprehensive analysis of player performance, hero selection, item usage, and their relationship with match outcomes.

1) Overall Descriptive Statistics: Table I presents the overall descriptive statistics of the key performance metrics.

TABLE II
DESCRIPTIVE STATISTICS OF PERFORMANCE METRICS

Metric	Mean	Median	Std. Dev.	Min	Max
kills	10.41	9.0	6.22	1.0	34.0
deaths	9.84	10.0	4.46	3.0	24.0
assists	16.26	16.0	7.34	2.0	39.0
calculated_kda_ratio	3.35	2.67	2.25	0.75	12.67
gold_per_min	499.09	480.0	150.40	234.0	896.0
xp_per_min	714.00	673.0	221.51	312.0	1263.0
kills_per_min	0.24	0.20	0.13	0.02	0.76
duration	2556.28	2388.0	698.94	1503.0	4560.0

The dataset shows that `assists` (mean 16.26, max 39) consistently exceed `kills` (mean 10.41, max 34), while `deaths` remain close to `kills` (mean 9.84). Economic performance varies widely, with `gold_per_min` ranging from 234 to 896 and `xp_per_min` from 312 to 1263, whereas `kills_per_min` stays relatively low and match durations span from 1503 to 4560 seconds.

B. Distribution of Performance Metrics

To examine distribution patterns of key performance metrics, histograms were generated.

Figure 2 on the next page illustrates the distribution of the KDA ratio across matches.

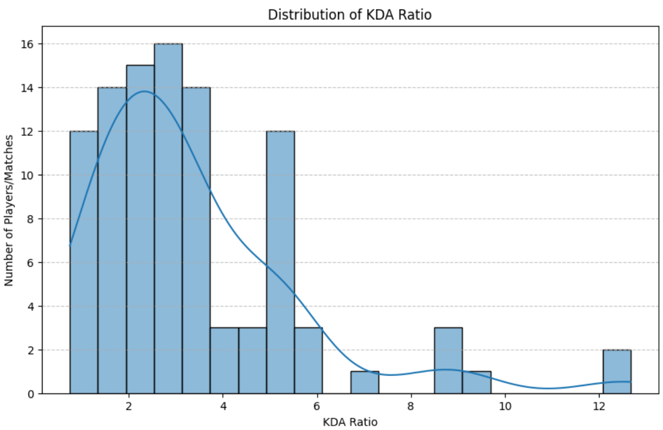

Fig. 2. Histogram of KDA Ratio

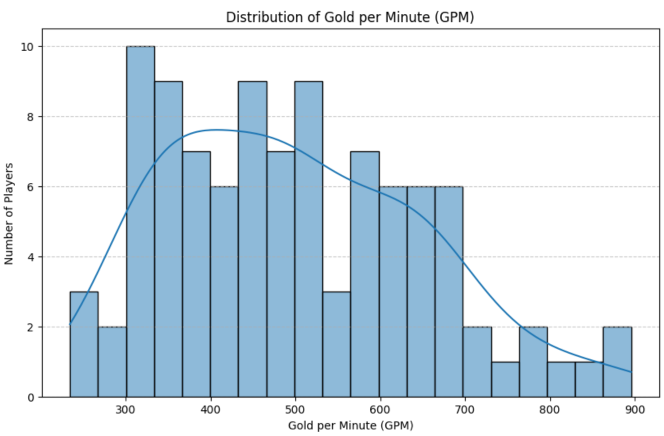

Fig. 3. Histogram of Gold per Minute (GPM)

Figure 3 presents the distribution of gold per minute (GPM).

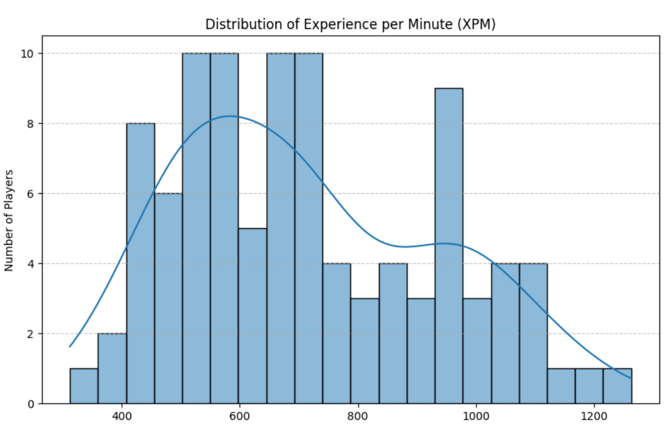

Fig. 4. Histogram of Experience per Minute (XPM)

Figure 4 shows the distribution of exp per minute (XPM).

C. Time-Series Trends

To analyze performance consistency, weekly averages were computed using the `start_time` variable.

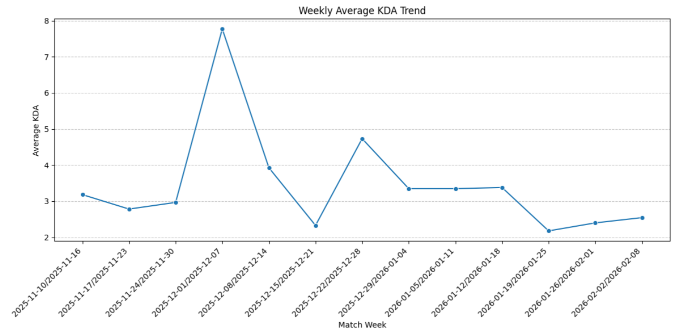

Fig. 5. Weekly Average KDA

Figure 5 illustrates erratic fluctuations in weekly combat efficiency across the observation period.

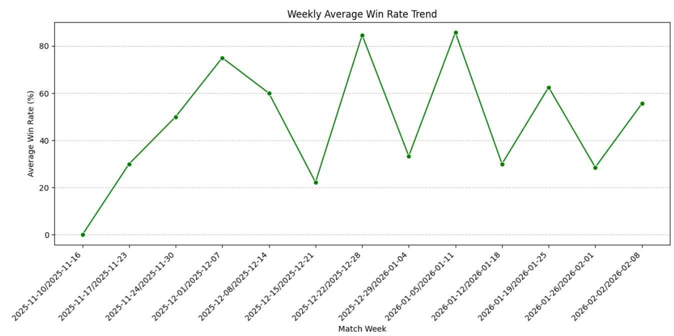

Fig. 6. Weekly Win Rate Trend

Figure 6 shows consistentcy in fluctuations of win rate over time.

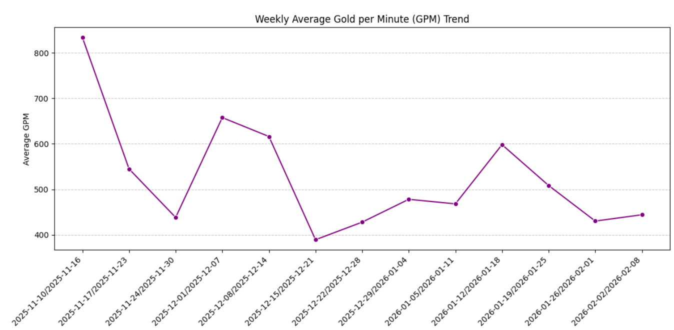

Fig. 7. Weekly Gold per Minute Trend

Figure 7 demonstrates similarities to Figure 5 in similar observation period.

D. Correlation Analysis

Pearson correlation coefficients were computed to identify relationships among key performance variables.

Figure 8 visualizes the correlation coefficients between the metrics.

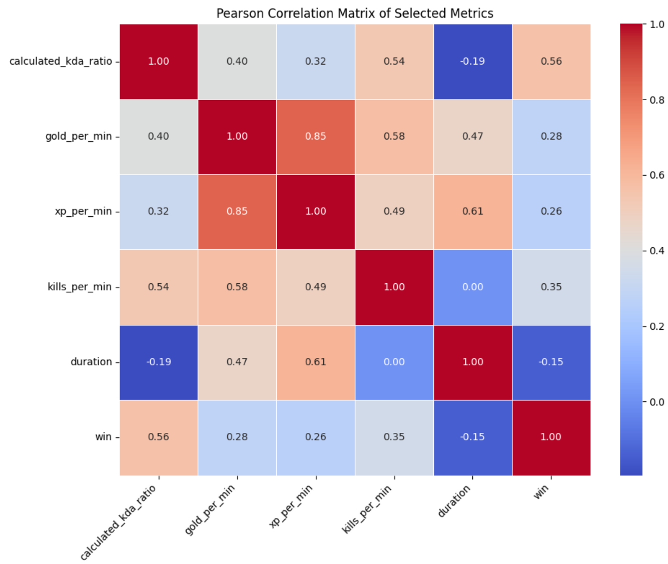

Fig. 8. Correlation Matrix of Performance Metrics

E. Win vs. Loss Comparative Analysis

Performance metrics were grouped by match outcome to compare differences between wins and losses.

TABLE III
MEAN PERFORMANCE METRICS BY MATCH OUTCOME

Metric	Win (Mean)	Loss (Mean)
KDA	4.58	2.05
gold_per_min	539.67	455.98
xp_per_min	769.94	654.56
kills_per_min	0.29	0.20

Table III shows the mean of each metric, grouped by win and losses.

1) Statistical Testing: Independent-samples t-tests were conducted to determine whether observed differences were statistically significant.

TABLE IV
INDEPENDENT-SAMPLES t-TEST RESULTS

Metric	t-value	p-value	Cohen's d
KDA	6.892	0.000	1.351
gold_per_min	2.869	0.005	0.577
xp_per_min	2.672	0.009	0.537

Table IV shows the statistical significant differences ($p < 0.05$) between wins and losses for KDA, gold per minute, and experience per minute, with Cohen's d values indicating medium to large effect sizes.

F. Performance Consistency Metrics

To quantify performance stability, the coefficient of variation (CV) was computed for key metrics.

Table V shows the standard deviation and coefficient of variation for key metrics.

TABLE V
STANDARD DEVIATION AND COEFFICIENT OF VARIATION FOR CORE METRICS

Metric	SD	CV (%)
KDA	1.48	42.73
gold_per_min	123.80	23.55
xp_per_min	147.33	19.75

V. DISCUSSION

This section interprets the statistical findings presented in Section IV and discusses their implications for gameplay behavior, performance consistency, and performance-driven match outcomes.

A. Interpretation of Results

The results indicate that combat efficiency and economic performance are both significantly associated with match outcomes. The overall win rate of 51.52% suggests relatively balanced performance across the observation period. However, winning matches demonstrated substantially higher averages in calculated KDA (4.58 vs. 2.05), gold per minute (539.67 vs. 455.98), and experience per minute (769.94 vs. 654.56). Among these, KDA showed the strongest statistical separation between wins and losses ($t = 6.892$, $p < 0.001$, Cohen's $d = 1.351$), indicating a large effect size. Correlation analysis further supports this finding, with win outcome moderately correlated with KDA ($r = 0.564$), while gold per minute ($r = 0.280$) and experience per minute ($r = 0.262$) showed weaker but positive relationships.

Time-series analysis revealed noticeable weekly fluctuations in KDA and win rate, indicating variability in short-term combat performance. In contrast, economic metrics displayed relatively smoother trends over time. This pattern suggests that while farming efficiency remains somewhat stable, combat execution and decision-making vary more significantly across matches.

The coefficient of variation (CV) results reinforce this observation. KDA exhibited the highest variability (CV = 42.73%), compared to gold per minute (23.55%) and experience per minute (19.75%). This indicates that combat-related performance is less consistent than economic accumulation. Overall, the findings suggest that although stable farming contributes to match success, strong combat effectiveness serves as the primary differentiating factor between wins and losses.

B. Comparison to Related Work

Prior gaming analytics studies have highlighted the importance of economic efficiency in MOBA games, showing that gold and experience accumulation increase win probability. The present findings partially align with this, as gold per minute and experience per minute were significantly higher in wins ($p < 0.05$), although their correlations with match outcome were relatively weak ($r = 0.280$ and $r = 0.262$, respectively). In contrast, KDA demonstrated a stronger relationship with winning ($r = 0.564$) and a large effect size.

Unlike large-scale population studies, this research adopts a single-player longitudinal approach. Rather than aiming for generalization, it emphasizes individualized performance consistency and playstyle profiling, offering a focused, micro-level application of statistical gaming analytics.

C. Limitations

Several limitations should be acknowledged:

- **Single-subject design** (n = 1)**:** The dataset represents only one player; therefore, findings cannot be generalized to the broader Dota 2 population.
- **Short observation period:** The data cover approximately three months. A longer timeframe may yield more stable trend estimates.
- **Limited contextual variables:** Factors such as team composition, opponent skill level, role assignment, and patch updates were not included.
- **Public data constraints:** The dataset relies exclusively on publicly available OpenDota metrics, limiting deeper behavioral tracking.
- **Unmeasured external influences:** Academic workload, stress, and physical fatigue were not measured but may affect gameplay performance.

D. Recommendations and Future Work

Future studies may extend the observation period to assess long-term performance consistency. Incorporating additional variables such as role classification, matchmaking rating, and patch changes could provide deeper contextual insights. Expanding the analysis to multiple players would allow comparative evaluation of playstyle consistency.

More advanced statistical methods, including regression or time-series modeling, may also be applied to predict performance trends. Consistent and structured data collection remains essential for improving analytical reliability in similar student-led projects.

E. Overall Interpretation

The findings indicate that both combat effectiveness and economic efficiency contribute to match success, with KDA emerging as the strongest differentiator between wins and losses. While gold per minute and experience per minute were significantly higher in wins, their correlations with match outcome were weaker compared to KDA. Economic metrics also showed lower variability, suggesting more consistent farming performance across matches.

Overall, the study demonstrates how descriptive statistics and inferential analysis can convert personal gaming data into quantifiable insights on performance consistency, efficiency, and gameplay behavior.

VI. Conclusion

This study analyzed the performance consistency and playstyle characteristics of a Dota 2 player using publicly available OpenDota match data. Through statistical analysis of weekly matches collected between November 16, 2025 and February 6, 2026, the research examined how combat efficiency, economic performance, and temporal trends relate to match outcomes.

The results show that while gold per minute and experience per minute were significantly higher in wins, the calculated KDA ratio demonstrated the strongest relationship with match success. Winning matches recorded substantially higher KDA values, supported by a large effect size and moderate positive correlation. Economic metrics also differed significantly between wins and losses, though their relationships with outcome were comparatively weaker. Variability analysis indicated that farming-related metrics were more stable over time than combat-related metrics.

Time-series trends revealed weekly performance fluctuations, suggesting that short-term consistency varies across the observation period. Correlation analysis further highlighted the strong association between economic growth and combat output, particularly the high relationship between gold and experience per minute.

From a practical perspective, the findings suggest that improving combat efficiency while maintaining consistent farming fundamentals may enhance overall match performance. Although limited to a single participant and a short timeframe, this study demonstrates how structured data science methods can convert personal gameplay data into measurable insights on performance consistency and strategic behavior.

References

[1] OpenDota, "OpenDota," 2026. [Online]. Available: https://www.opendota.com/. [Accessed: Feb. 15, 2026].

[2] OpenDota, "OpenDota API Documentation," 2026. [Online]. Available: https://docs.opendota.com/. [Accessed: Feb. 15, 2026].

[3] D. Zhang, Z. Xu, and M. Rashid, "DotA 2 Match Outcome Prediction System Using Decision Tree Ensemble Algorithms," *Intelligence* (Special Issue on Games), MDPI, vol. 9, no. 12, Dec. 2023. [Online]. Available: https://www.mdpi.com/2504-2289/9/12/302. [Accessed: Feb. 15, 2026].

[4] V. Chugani, "The Analytics Behind Esports: Data-Driven Strategies in Competitive Gaming," Statology, Sep. 26, 2024. [Online]. Available: https://www.statology.org/the-analytics-behind-esports-data-driven-strategies-in-competitive-gaming. [Accessed: Feb. 15, 2026].

[5] W. Chiu, T. C. M. Fan, S.-B. Nam, and P.-H. Sun, "Knowledge Mapping and Sustainable Development of eSports Research: A Bibliometric and Visualized Analysis," *Sustainability*, vol. 13, no. 18, p. 10354, Sep. 2021. [Online]. Available: https://www.mdpi.com/2071-1050/13/18/10354. [Accessed: Feb. 15, 2026].