

Article

DotA 2 Match Outcome Prediction System Using Decision Tree Ensemble Algorithms

Sukhrob Yangibaev ¹, Jamolbek Mattiev ^{2,*} and Sello Mokwena ³

¹ Department of Artificial Intelligence, Al-Khwarizmi University, Abulgazi Bakhodirkhan Street 195, Urgench 220100, Uzbekistan; s.yangibaev@akhu.uz

² Computer Science and Artificial Intelligence Technologies Department, Urgench State University, Khamid Alimdjan 14, Urgench 220100, Uzbekistan

³ Computer Science Department, University of Limpopo, Polokwane 0727, South Africa; sello.mokwena@ul.ac.za

* Correspondence: jamolbek.mattiev@famnit.upr.si or mattiev.jamolbek@urdu.uz; Tel.: +998-99-585-4223

Abstract

This paper explores the replication of the DotA Plus prediction system using decision tree algorithms. The study implements and evaluates Extra Trees Classifier, Random Forest Classifier, and Hist Gradient Boosting Classifier, along with their combined average, for predicting the outcome of Defense of the Ancients (DotA) 2 matches. Data was collected using the OpenDotA API and the Steam API, and various features such as game duration, tower and barracks states, net-worth, assists, last hits, gold, level, gold per minute, and experience per minute were extracted for model training. Additionally, hero and item win rate features, derived from Dotabuff data, were incorporated to enhance the models' predictive accuracy. The models were trained on datasets with varying match durations, including segments for matches under 10 min, between 10 and 20 min, and over 20 min. The experimental results show that the Extra Trees Classifier consistently outperformed other individual models and performed comparably to the averaged models, achieving a peak performance of 98.6% test accuracy on matches longer than 20 min when using match duration segmentation and hero/item embeddings. The study highlights the effectiveness of decision tree-based methods for real-time match outcome prediction in DotA 2 and offers insights into feature importance. The combined average of Extra Trees Classifier, Random Forest Classifier, and Hist Gradient Boosting Classifier models provides a robust and reliable prediction of DotA 2 match outcomes, thus showing potential as a real-time prediction system.

Keywords: DotA 2; real-time analysis; OpenDotA API; feature engineering; item winrates; esports; prediction system



Academic Editor: Giuseppe Maria Luigi Sarnè

Received: 1 October 2025

Revised: 14 November 2025

Accepted: 22 November 2025

Published: 27 November 2025

Citation: Yangibaev, S.; Mattiev, J.; Mokwena, S. DotA 2 Match Outcome Prediction System Using Decision Tree Ensemble Algorithms. *Big Data Cogn. Comput.* **2025**, *9*, 302. <https://doi.org/10.3390/bdcc9120302>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The exponential growth of esports in the past decade has transformed competitive gaming into a global phenomenon, with titles such as DotA 2 drawing millions of active players and international audiences through tournaments like The International. This rise has not only popularized the spectator experience but also generated substantial research interest in predictive analytics, particularly in understanding and forecasting match outcomes. Predictive models can enhance engagement for viewers, provide insights for professional teams, and support third-party platforms in delivering analytical content.

One of the most notable innovations in this space is Valve's DotA Plus subscription service, which includes a real-time win probability prediction system. This system dynamically updates during a match, offering spectators and players a sense of how likely each team is to win as the game evolves. While impactful, the system is proprietary and closed-source, limiting the ability of researchers and practitioners to study or replicate its mechanisms.

The practical value of developing an open and interpretable DotA 2 match prediction system extends beyond academic curiosity. Such a system can provide real-time analytical insights to players and coaches for post-match review, support casters and broadcasters in explaining in-game momentum shifts to spectators, and assist tournament organizers and analysts in forecasting outcomes during live events. By making predictive mechanisms transparent, the system also promotes trust and educational value, enabling players and data scientists to understand how specific in-game factors such as hero selections, gold advantage, or item timings affect the likelihood of winning. This creates a strong incentive for the academic community to develop transparent, reproducible, and explainable systems that combine high predictive accuracy with practical usability, bridging the gap between proprietary esports analytics tools (e.g., DotA Plus) and open research-driven frameworks.

Previous studies have investigated match prediction in DotA 2 and similar multiplayer online battle arena (MOBA) games using a wide range of machine learning techniques. Akhmedov and Phan [1] applied machine learning models to predict DotA 2 outcomes, highlighting the potential of structured approaches for esports analytics. Deep learning models have also been explored, for example, in Ke et al. [2], who proposed a neural network framework to predict match outcomes based on simulated team fights. Other approaches, such as MOBA-Slice by Yu et al. [3], introduced time-slice frameworks to evaluate the relative advantage between teams, reflecting the importance of temporal segmentation. Do et al. [4] demonstrated that incorporating player-champion experience substantially improves predictive accuracy in League of Legends, highlighting the value of contextual features. Similarly, Chowdhury et al. [5] found that ensemble-based models outperform single classifiers for match prediction tasks, emphasizing the robustness of decision tree-based approaches in esports prediction research. These works collectively emphasize that time-evolving match dynamics and contextual feature selection are crucial to building accurate prediction models.

Despite their success, deep learning models often lack interpretability, posing challenges for esports analysts who wish to understand why a model produces certain predictions. In this regard, decision tree-based algorithms (e.g., Random Forest, Extra Trees, and Gradient Boosting) offer a strong alternative. They are highly effective for structured tabular data [6,7], computationally efficient [8–11], and provide interpretable outputs such as feature importance scores. These properties make decision trees particularly well-suited for DotA 2, where structured features (net-worth, gold per minute, hero levels, tower states) and contextual statistics (hero/item win rates) dominate the prediction landscape. The study [12,13] demonstrates the advantage of decision tree models on bigger datasets through experimental evaluation.

This study investigates the use of decision tree algorithms, Extra Trees Classifier (ETC), Random Forest (RF), and Histogram-based Gradient Boosting (HGB) for replicating the DotA Plus prediction system. To capture the evolving dynamics of gameplay, datasets were segmented by match duration (<10 min, 10–20 min, >20 min), mirroring the early, mid, and late phases of DotA 2 matches. Furthermore, the feature space was enriched with hero and item embeddings derived from Dotabuff, reflecting strategic decisions that strongly influence outcomes.

By combining publicly available match data from the OpenDota API, the Steam API, and external win rate statistics. The main contribution of this study lies in developing

a transparent and reproducible DotA 2 outcome prediction framework that replicates the functionality of Valve's DotA Plus system using decision tree ensembles. This work introduces the following methodological innovations:

- Temporal segmentation of datasets into early-, mid-game, and late-game phases, reflecting DotA 2's evolving match dynamics;
- Integration of hero and item embeddings, derived from Dotabuff win rates, to encode meta-strategic information;
- Evaluation of model interpretability and ensemble averaging, producing stable, accurate, and explainable predictions.

The experimental results show that the Extra Trees Classifier consistently outperforms other models, achieving accuracies of 91.4%, 97.3%, and 98.6% for early (<10 min), mid (10–20 min), and late (>20 min) game segments, respectively. When enhanced with hero and item embeddings, performance further improved to 94.6%, 94.1%, and 97.9%, highlighting the contribution of contextual features. The findings demonstrate that decision tree models not only achieve a high degree of predictive accuracy, demonstrating the potential to function at a level comparable to proprietary systems like Dota Plus, but also provide interpretable insights into which features most strongly determine success. In doing so, this research contributes to the growing body of literature on esports analytics, while offering practical tools and open-source methodologies that can be extended to other MOBA titles.

2. Related Works

Research on esports analytics and match outcome prediction has gained momentum in recent years, with DotA 2 serving as a central testbed due to its strategic complexity and data richness. Early approaches applied classical machine learning techniques to predict outcomes from structured features. For example, Akhmedov and Phan [1] investigated multiple learning models for DotA 2 prediction and demonstrated that feature selection and algorithm choice strongly influence predictive performance. Their work highlighted the feasibility of using publicly available datasets for accurate modeling of match outcomes, but lacked temporal and contextual components, which limited its adaptability to evolving game states.

Subsequent research incorporated deep learning to capture temporal dependencies and team interactions. Ke et al. [2] developed a team-fight-based neural model that simulates combat dynamics to enhance predictive performance. Although their framework improved accuracy during mid-game and late-game stages, it required high computational resources and suffered from limited interpretability, an important drawback for real-time analytics. Similarly, Yu et al. [3] introduced MOBA-Slice, a time-slice based evaluation framework that quantified relative team advantages across different game phases, reflecting the importance of temporal segmentation in MOBA prediction. These works emphasized that outcome prediction benefits from contextualized, time-evolving features, rather than static aggregates. While it successfully modeled temporal progression, it remained descriptive rather than predictive, offering limited real-time applicability.

In parallel, the broader machine learning literature has produced a wide spectrum of algorithms and frameworks applicable to esports prediction. Rokach [14] provided a foundational overview of ensemble learning, illustrating how the aggregation of weak learners through methods such as bagging, boosting, and random subspace selection can significantly improve predictive accuracy and generalization. These ensemble strategies are particularly effective for complex, nonlinear datasets like DotA 2 match records, where interactions between features such as hero selection, itemization, and temporal performance evolve dynamically. However, while ensemble approaches tend to outperform single

learners, they often suffer from limited interpretability, making it difficult to trace individual decisions or feature contributions.

Subsequent advances in model optimization such as histogram-based estimation [15] and metaheuristic tuning strategies [16], have further extended the capabilities of traditional ensemble models. Histogram-based methods accelerate computation by discretizing continuous variables, yielding faster convergence and reduced memory usage for large-scale gaming datasets. Yet, this discretization can obscure subtle feature dependencies, reducing fine-grained interpretability. Conversely, metaheuristic algorithms, including genetic optimization and swarm-based tuning, often achieve higher accuracy through adaptive parameter search but are computationally expensive and impractical for real-time prediction contexts like live esports analysis.

Win prediction in esports has been widely investigated in live professional environments where the ability to update forecasts in real time is critical. Hodge et al. [17] conducted one of the earliest large-scale empirical studies on DotA 2 and League of Legends, demonstrating that machine learning models can effectively predict match outcomes using live in-game statistics streamed from professional tournaments. Their system highlighted the importance of incrementally updated features, such as team net-worth, kill differentials, and objective control, to capture the evolving state of play. This research established the conceptual foundation for dynamic, data-driven win probability systems that mirror the real-time analysis performed by proprietary engines such as DotA Plus. However, while Hodge et al.'s framework achieved strong predictive accuracy, it relied heavily on complex feature engineering and high-frequency data collection, making it computationally intensive and less accessible for open, reproducible research. Moreover, their models prioritized real-time responsiveness over interpretability, offering limited insight into the relative importance of contributing factors.

Beyond pure prediction accuracy, interpretability has emerged as a central concern in recent research. Losada-Rodríguez et al. [18] specifically explored the explainability of machine learning algorithms in the context of DotA 2. This interpretability focus is essential for esports analytics, where analysts and players must trust the system's insights. However, many explainability approaches remain post hoc rather than intrinsic, whereas the tree-based methods used in this study provide native interpretability through feature importance metrics.

Earlier research into DotA 2 outcome prediction primarily focused on applying conventional machine learning models to structured match data. Wang et al. [19] investigated several baseline algorithms, including logistic regression, decision trees, and support vector machines, to forecast match winners using in-game features such as kills, gold accumulation, experience gain, and tower destruction. Their study provided one of the first empirical validations that meaningful predictive signals exist in DotA 2's statistical data, establishing a foundation for subsequent analytical models. However, these early approaches were constrained by limited dataset diversity, static feature design, and the absence of temporal or contextual modeling, which reduced their ability to capture the evolving dynamics of a match.

In addition to interpretability, hybrid approaches that merge in-game and external data are gaining traction. Lee and Kim [20] introduced machine learning based indicators leveraging bounty mechanisms to detect comebacks, highlighting dynamic prediction beyond static win rates. The present study builds upon these ideas by integrating hero and item win rate embeddings with in-game statistics, an innovation that captures both meta-level strategies and real-time performance trends.

In contrast, the present study leverages the strengths of ensemble decision tree algorithms, specifically, Extra Trees, Random Forest, and Histogram-based Gradient Boosting

while mitigating the interpretability and computational drawbacks of prior approaches. Unlike deep or metaheuristic models, our framework provides transparent feature attribution, allowing analysts to understand which variables most strongly influence match outcomes. Moreover, by integrating temporal segmentation and meta-level embeddings (hero and item win rates), the proposed system achieves high accuracy (up to 98.6%) with substantially lower computational cost, offering a reproducible, interpretable, and efficient alternative to ensemble and deep learning methods.

3. Materials and Methods

The methodological framework of this study was designed to replicate the DotA Plus prediction system in a transparent and reproducible manner (Shown in Figure 1). The process involved four main stages: data acquisition, feature engineering, model development, and evaluation.

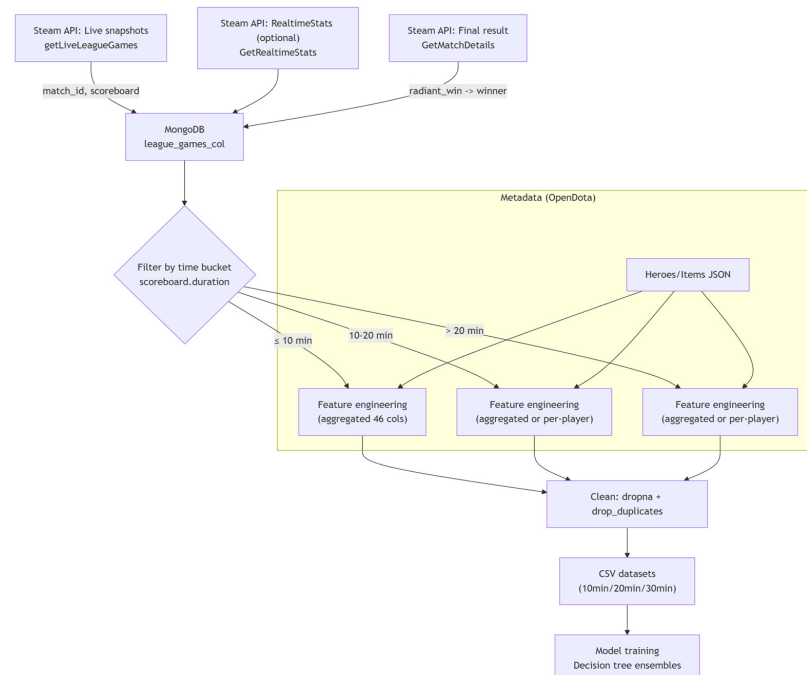


Figure 1. End-to-end data pipeline of the proposed methodology.

Figure 1 shows the data preprocessing pipeline to ensure full reproducibility from collection to model-ready datasets. Match data were obtained from the Steam API, which provided live league snapshots and final match outcomes, and from the OpenDota API. Each live game snapshot was synchronized with its corresponding final result using a unique match identifier. Once the final ground-truth winner became available, the record was updated accordingly and retained only if both teams' complete scoreboards (five players each) were present.

To improve temporal interpretability, snapshots were filtered. Two parallel feature sets were derived: (1) compact team-aggregate differences (46 variables, such as net-worth, experience, and tower states), and (2) rich per-player features (186 variables) capturing kills, deaths, assists, hero levels, and item information. Bitfield data representing tower and barracks states were expanded into binary indicator columns for each structure. Missing or incomplete rows were conservatively removed rather than imputed to avoid label distortion. All features were normalized. The binary target label (winner) was encoded as 1 for Radiant victory and 0 for Dire victory. This process produced stratified CSV datasets for each temporal bucket, which were used in the training phase. The following subsections discuss each of above-mentioned steps.

3.1. Data Source and Acquisition

To build a reliable prediction framework, high-quality match data was required. In this study, match data were collected from two primary sources: the OpenDotA API and the Steam API, both of which provide structured access to live and historical DotA 2 data. For real-time and league-level information, custom Python-based data miners were developed. These miners continuously retrieved live match statistics and stored them in a MongoDB database, while additional processes updated the database with final match outcomes once games concluded. The dataset spans the 2021–2024 period, covering multiple DotA 2 patches and competitive leagues. Each match was uniquely identified by its `match_id`, ensuring strict separation between training and testing sets.

To enrich predictive features, hero and item win rate statistics were scraped from Dotabuff, a widely used third-party analytics platform. These statistics were processed into numerical embeddings representing both individual team compositions and aggregated win rate profiles, providing contextual knowledge beyond raw in-game statistics.

3.2. Dataset Composition

After acquisition, the collected data were organized into structured datasets suitable for machine learning. Two categories of datasets were used:

1. Public matches: ranging from 1 k to 40 k samples across different experiments.
2. League matches: larger datasets from 20 k to 116 k samples, including segmented subsets by match duration.

Each record contained hundreds of structured features:

- Game-level: duration, tower states, barracks states, team score, radiant/dire series wins;
- Player-level: kills, deaths, assists, last hits, denies, gold, level, gold per minute (GPM), experience per minute (XPM), and net-worth;
- Inventory-level: hero identifiers, item slots (0–6), and additional backpack/neutral items.

External hero and item win rates were scraped from Dotabuff and merged with the datasets to capture meta-level knowledge

3.3. Preprocessing and Feature Engineering

Raw data alone is insufficient for capturing the complexity of DotA 2 matches. Therefore, additional feature engineering steps were applied to transform in-game statistics and external knowledge into predictive inputs.

- Normalization: Continuous features such as GPM and XPM were scaled.
- Feature selection: Different feature subsets were tested, ranging from full-scale (186+ features) to reduced representations (~46 features), to evaluate model performance under computational constraints.
- Temporal segmentation: Matches were partitioned into three phases:
 1. Early game (<10 min)
 2. Mid game (10–20 min)
 3. Late game (>20 min)

This segmentation reflects DotA 2's evolving dynamics and improves prediction reliability.

- Embeddings: To incorporate meta-level strategic context, hero and item embeddings were generated using publicly available data from Dotabuff. Each hero and item was represented by two frequency-based attributes: global win rate and pick rate. These statistics were retrieved through Dotabuff's historical match data and aggregated to form contextual embeddings for every team composition. For each match, 10 hero slots (5 per team) and up to 6 item slots per player were encoded into a structured

vector, yielding fixed-dimensional inputs that capture both individual and collective performance potential. All embedding values were normalized using min–max scaling to the $[0, 1]$ range prior to model training. This encoding method preserves interpretability while enriching the model with data-driven meta-information on hero and item effectiveness.

3.4. Decision Tree Algorithms

The selection of Extra Trees, Random Forest, and Histogram Gradient Boosting models was guided by the need for transparent and reproducible ensemble methods that balance interpretability with accuracy. Although modern alternatives such as XGBoost and CatBoost often achieve higher raw accuracy, their gradient-boosting frameworks involve additional complexity and reduced transparency in feature attribution. The decision tree ensembles used in this study, implemented through Scikit-learn, provide straightforward feature-importance metrics and reproducible parameter settings, making them ideal for an open Dota 2 prediction framework.

The selection of these specific models was directly informed by our prior comparative analysis [21], which evaluated six different decision tree algorithms for Dota 2 outcome prediction. In that work, these three ensemble methods demonstrated the highest predictive accuracy and robustness, establishing them as the most promising candidates for the more advanced, time-segmented replication study presented in this paper. Their proven effectiveness, capacity to capture nonlinear relationships, and interpretability make them ideal for this task.

Extra Trees Classifier: The Extra Trees Classifier [16] constructs an ensemble of randomized decision trees, where both feature selection and split thresholds are determined randomly during training. This approach increases model diversity, reduces variance, and accelerates training compared with conventional Random Forests. In this study, the ETC was configured with 150 estimators (a relatively high number of trees (150) was selected to stabilize accuracy without excessive training time), using the entropy criterion to maximize class separability.

Random Forest Classifier: The Random Forest model aggregates multiple decision trees trained on random subsets of the dataset using bootstrap sampling. Each tree is built using a random subset of features to enhance generalization and reduce overfitting. In this study, the RF was implemented with 50 estimators (a moderate ensemble size achieved a strong balance between predictive accuracy and computation time) and the Gini impurity criterion, providing an interpretable and computationally efficient baseline for ensemble comparison.

Hist Gradient Boosting Classifier: The Histogram Gradient Boosting Classifier builds trees sequentially, with each new tree correcting the residual errors of the previous ensemble. Unlike traditional gradient boosting, HGB uses histogram-based binning of continuous features, which significantly improves computational efficiency on large-scale datasets. The classifier was configured with a learning rate of 0.2 and 100 iterations, providing rapid convergence while maintaining high accuracy.

Ensemble Methods (Combining Classifiers): In addition to evaluating the individual performance of each algorithm, this study explores the use of ensemble methods, specifically averaging the predictions of multiple classifiers, to achieve a more robust overall system [14]. By averaging the predictions from Extra Trees, Random Forest, and Hist Gradient Boosting models, this study aims to reduce variance and bias, leading to a more stable and accurate prediction of Dota 2 match outcomes.

4. Results

All models were developed using the Scikit-learn framework (version 1.5.2) and trained under identical data partitions with a 90/10 train–test split and a fixed random seed (random_state = 1) to ensure reproducibility.

The configuration of each model followed standard practices in ensemble learning, with hyperparameters tuned empirically through multiple iterations to balance interpretability, accuracy, and training efficiency. The performance of each model was evaluated individually, and their combined average was also analyzed to determine the effectiveness of an ensemble approach. The models were trained and tested on datasets segmented by match duration to assess the impact of game length on predictive accuracy. These datasets included matches under 10 min, between 10 and 20 min, and over 20 min. The features used for model training included game duration, tower and barracks states, net-worth, assists, last hits, gold, level, gold per minute (gpm), and experience per minute (xpm). Additionally, hero and item win rate features derived from Dotabuff were incorporated to enhance the models' predictive accuracy.

We summarize the main results obtained from experiments across varying dataset sizes and match duration segments.

The DotA Plus prediction system, which serves as a benchmark for this study, displays the win probability for each team in real-time during a match (see Figure 2). The panel shows three time-series provided by the in-game DotA Plus system for an example professional match: (i) win probability for the Radiant team (percentage), (ii) team net-worth differential (gold), and (iii) team experience differential (XP). The x-axis is match time (minutes). Curve colors follow the legend at the top; positive values indicate a Radiant advantage and negative values indicate a Dire advantage. The colored dots along the bottom are event markers produced by the game client (e.g., kills/objectives).



Figure 2. DotA Plus analytics for an example match: win probability, net-worth, and experience over time.

Figure 2 visualizes the DotA Plus analytics available during a match: win probability (%), net-worth (gold) differential, and experience (XP) differential as functions of time. We reference this panel to orient the reader to the signals that motivate our features. In our dataset, we construct snapshot features at fixed horizons (10/20/30 min), aggregating

team economy and experience indicators together with hero/item information; the label remains the final match outcome. We do not use the proprietary DotA Plus win-probability output as a target or as a direct input. The figure therefore serves only as an example of the temporal context from which our fixed-time features are derived.

To evaluate the performance of our models, we analyzed their predictions over the course of sample matches. Figure 3 shows the predictions of the Extra Trees Classifier throughout a match. The x-axis represents the match duration in minutes, and the y-axis reports $P(\text{Radiant win})$ in percent (0–100). The dashed line at 50% is the decision boundary: values above 50% indicate the model favors Radiant, while values below 50% indicate the model favors Dire ($P(\text{Dire win}) = 1 - P(\text{Radiant win})$). The red line indicates the Extra Trees Classifier's predictions, which fluctuate as the game state changes. Notably, the predictions stabilize as the match progresses, reflecting the increasing certainty of the outcome.

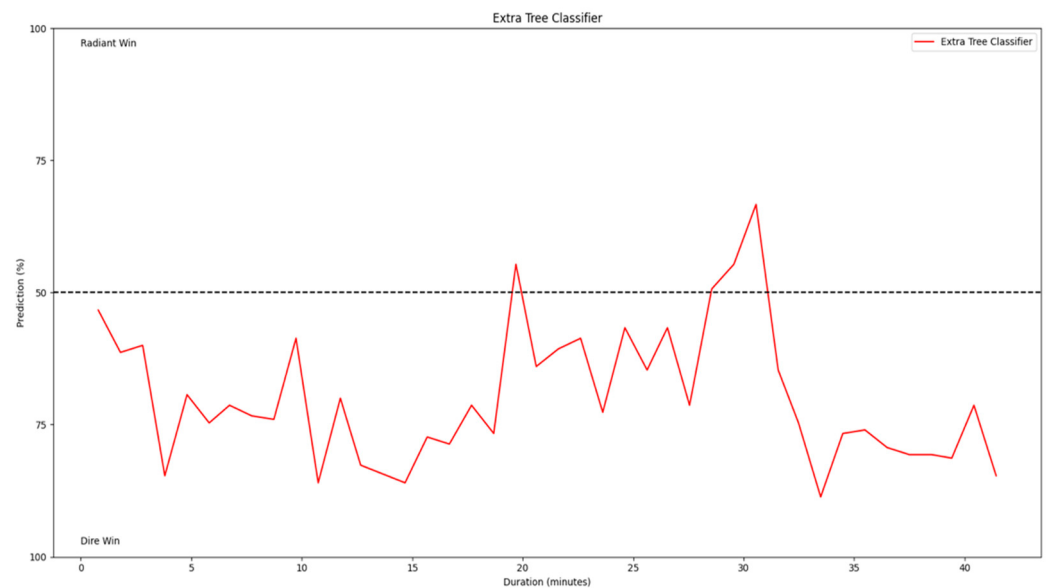


Figure 3. Predicted Radiant win probability over match duration for the Extra Trees Classifier.

Similarly, Figures 4 and 5 display the predictions of the Random Forest and Hist Gradient Boosting classifiers, respectively.

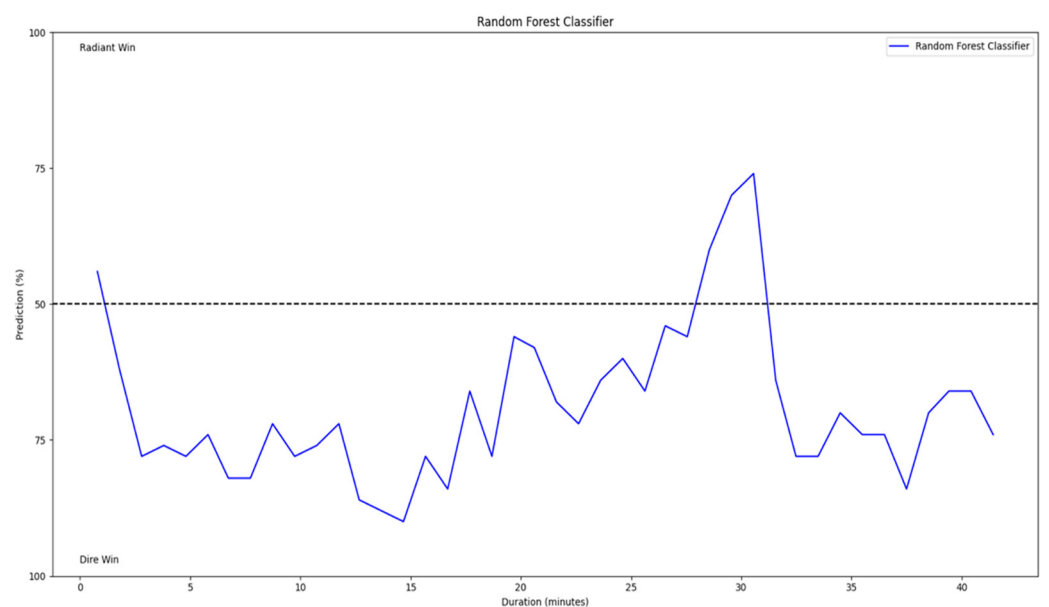


Figure 4. Predicted Radiant win probability over match duration for the Random Forest Classifier.

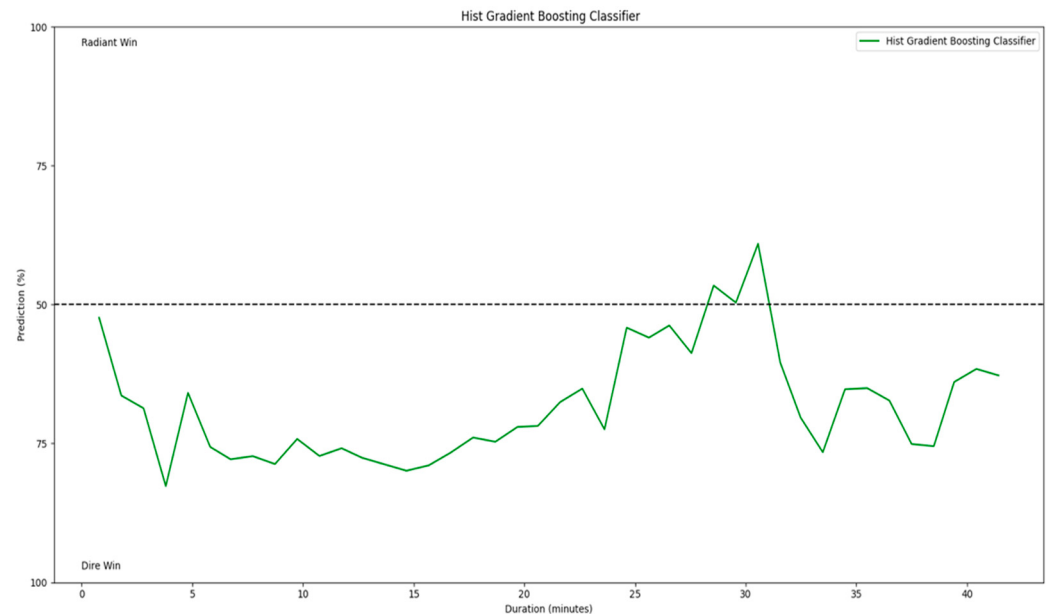


Figure 5. Predicted Radiant win probability over match duration for the Hist Gradient Boosting Classifier.

The blue line in Figure 4 represents the Random Forest predictions, while the green line in Figure 5 represents the Hist Gradient Boosting predictions. Both models exhibit similar trends to the Extra Trees Classifier, with predictions fluctuating in the early stages of the match and becoming more stable as the game progresses.

To explore the potential benefits of an ensemble approach, we calculated the average of the predictions from the Extra Trees, Random Forest, and Hist Gradient Boosting models. Figure 6 shows the average predictions over the course of the match. The black line represents the combined average, which demonstrates a smoother trend compared to the individual models. This suggests that averaging the predictions can reduce the variance and potentially improve the overall accuracy of the prediction system.

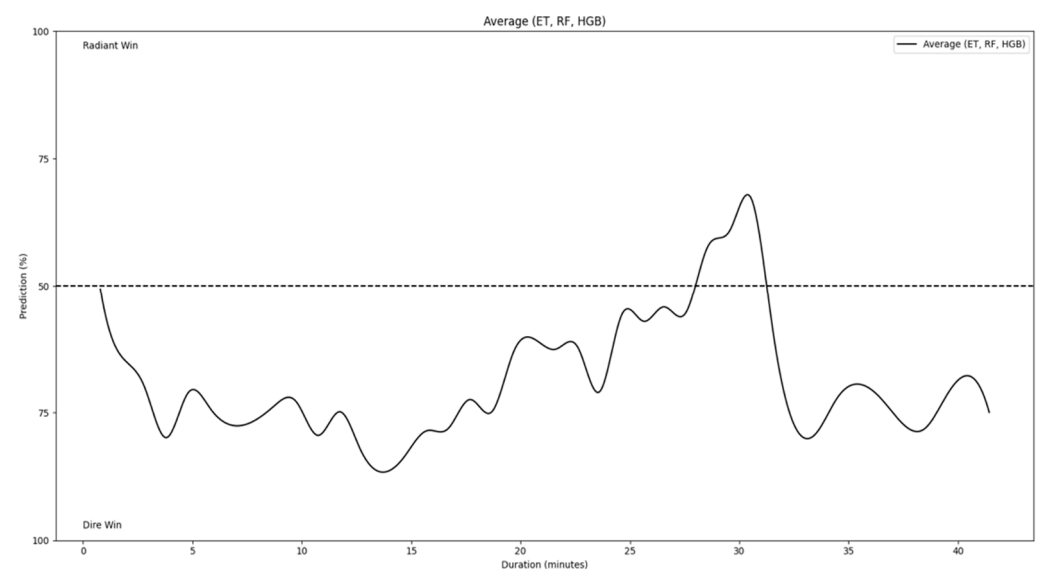


Figure 6. Average Model Predictions: Combined average of the predicted win probabilities for the Radiant team from the Extra Trees, Random Forest, and Hist Gradient Boosting Classifiers during the first match.

Figure 7 provides a comparative overview of the predictions from all three individual models and their combined average for the first analyzed match. The red, blue, green, and

black lines represent the Extra Trees, Random Forest, Hist Gradient Boosting, and average predictions, respectively. This comparison highlights the similarities and differences in the models' predictions and demonstrates the potential of the ensemble approach to provide a more robust and reliable prediction.

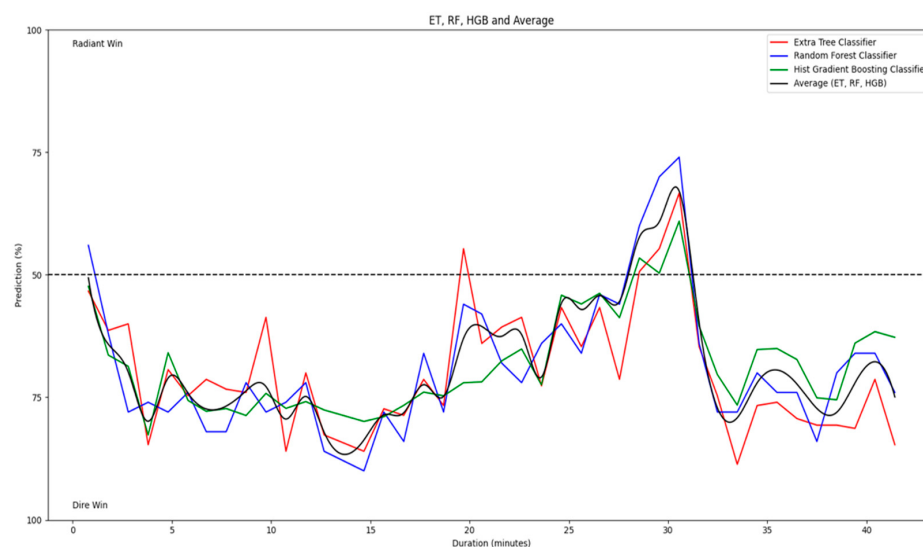


Figure 7. Example match timeline with the model's evolving prediction (First Match).

To further validate the models, a second match was analyzed using the same methodology. Figure 8 illustrates the DotA Plus prediction system for this second match, showcasing how the win probabilities change as the game unfolds. Figure 9 presents the predictions of the Extra Trees, Random Forest, and Hist Gradient Boosting classifiers, along with their combined average, for this second match.



Figure 8. Example match timeline with the model's evolving prediction.



Figure 9. Example of a match timeline with the model’s evolving prediction (Second Match).

Figure 8 summarizes one match as a timeline. The green line is the ensemble model’s win probability for Radiant computed at each step from the feature vector (Section 3), and it should be read on the 0–100% scale on the left. The yellow and blue lines are context signals that the model consumes team net-worth and experience differentials shown here after simple scaling so that their trends are visible on the same plot. These curves do not change the reported probabilities; they are included to make the drivers of probability shifts interpretable. Vertical grid lines mark 10 min intervals. The colored tick marks at the bottom indicate salient events (green for Radiant, red for Dire), such as team fights and objective captures, which often precede sharp probability movements. In the pictured match, a decisive fight around minute ~22 produces simultaneous spikes in net-worth and experience and a corresponding increase in the model’s confidence in a Radiant victory.

Similar to the first match, the models’ predictions fluctuate in the early game but tend to converge as the match progresses. The combined average again exhibits a smoother trend, suggesting its potential for increased stability and accuracy. In this match, the combined average remains more stable than the individual models.

The results demonstrate that the Extra Trees Classifier consistently outperformed the Random Forest and Hist Gradient Boosting classifiers across all datasets. Notably, the Extra Trees Classifier achieved a peak accuracy of 98.6% on matches longer than 20 min when using match duration segmentation and hero/item embeddings. This suggests that the Extra Trees algorithm is particularly well-suited for predicting the outcome of longer Dota 2 matches, where more data is available, and the game state is more stable. The combined average of the three models also performed well, achieving comparable accuracy to the Extra Trees Classifier. This indicates that the ensemble approach can provide a robust and reliable prediction system, leveraging the strengths of each individual model.

Tables 1 and 2 report the classification accuracies of the evaluated models across different dataset sizes and match-duration segments based on random hold-out split using scikit-learn’s `train_test_split` with a 90%/10% train/test split and a fixed random seed (`random_state = 1`); Each dataset contains one row per match at the corresponding time horizon, and models were fit and evaluated within that dataset only, so no multiple time slices from the same match are mixed across the train/test split. The only label transformation was encoding the binary match outcome; no preprocessing steps were fit on the full dataset before splitting. The accuracy values reported in Tables 1 and 2 correspond

to the classification accuracy metric, defined as the ratio of correctly predicted outcomes to the total number of test instances. Accuracy was computed using the “Scikit-learn `accuracy_score()`” function. Three-time intervals are considered: matches shorter than 10 min, matches lasting between 10 and 20 min, and matches longer than 20 min. The dataset size for each segment is approximately 38 k, 33 k, and 42 k records, respectively.

Table 1. Accuracy of models trained on league match data segmented by match duration.

Dataset Size	Duration Segment	Extra Trees	Random Forest	Hist Grad Boost
~38 k	<10 min	91.4%	85.6%	79.9%
~33 k	10–20 min	97.3%	96.1%	90.0%
~42 k	>20 min	98.6%	98.6%	96.1%

Table 2. Accuracy comparison of models trained with hero and item embeddings across different match durations.

Dataset Size	Duration Segment	Extra Trees	Random Forest	Hist Grad Boost
~83 k	<10 min	94.6%	84.4%	85.5%
~72 k	10–20 min	94.1%	93.0%	88.5%
~98 k	>20 min	97.9%	97.6%	93.2%

The results show that for short matches (<10 min), the Extra Trees classifier achieved the highest accuracy (91.4%), followed by Random Forest (85.6%), while Histogram-based Gradient Boosting (Hist Grad Boost) performed the weakest (79.9%). For medium-length matches (10–20 min), all models improved significantly, with Extra Trees reaching 97.3% and Random Forest 96.1%. Hist Grad Boost also improved but remained lower at 90.0%. For long matches (>20 min), both Extra Trees and Random Forest achieved equal and highest performance (98.6%), while Hist Grad Boost again trailed behind with 96.1%.

Table 2 presents the accuracy of models trained with hero and item embeddings, segmented by match duration. The datasets are considerably larger than in Table 1, with approximately 83 k samples for matches under 10 min, 72 k samples for 10–20 min matches, and 98 k samples for matches longer than 20 min.

It can be seen from the results that for short matches (<10 min), Extra Trees achieved the highest accuracy (94.6%), significantly outperforming both Random Forest (84.4%) and Histogram-based Gradient Boosting (85.5%). In medium-duration matches (10–20 min), Extra Trees and Random Forest performed comparably well, with 94.1% and 93.0% accuracy, respectively, while Histogram-based Gradient Boosting trailed slightly at 88.5%. For long matches (>20 min), both Extra Trees (97.9%) and Random Forest (97.6%) achieved almost identical, near-perfect accuracies, while Histogram-based Gradient Boosting performed strongly but lower at 93.2%.

Overall, the models demonstrate improved predictive power as match duration increases, similar to the trend in Table 1. Extra Trees consistently achieves the best results across all match durations, while Random Forest narrows the performance gap in longer matches. Histogram-based Gradient Boosting remains the weakest model but still benefits from longer match durations.

To ensure the reliability of the reported performance, we applied repeated stratified 5-fold cross-validation (3 repeats; 15 resamples) to all decision tree models on 10 min, 20 min, and 30 min snapshots datasets. For each model, the mean accuracy, standard deviation (SD), and 95% confidence interval (CI) were computed from the resampled accuracies. In addition, paired two-sided t-tests were conducted between the best model

and all others on identical resampling splits. The best-performing model for each temporal segment is highlighted in bold.

The results (Table 3) confirm that Histogram Gradient Boosting achieved the highest accuracy on the 10 min dataset, reaching an average of $65.3\% \pm 0.19\%$ (95% CI [65.23%, 65.43%]), marginally outperforming Extra Trees (64.87%) and Random Forest (64.4%). For the 20 min dataset, Extra Trees achieved the best performance with $74.9\% \pm 0.15\%$ (95% CI [74.86%, 75.02%]), surpassing Random Forest (73.3%) and Histogram Gradient Boosting (72.0%). On the 30 min dataset, Extra Trees again demonstrated the highest stability and predictive power, reaching $86.4\% \pm 0.07\%$ (95% CI [86.36%, 86.43%]), followed by Random Forest (82.1%) and Histogram Gradient Boosting (76.3%). Across all configurations, the performance differences were found to be statistically significant (paired *t*-test, all $p < 1 \times 10^{-10}$), confirming that the ensemble-based decision tree models generalize reliably across multiple temporal segments. The narrow confidence intervals (below $\pm 0.2\%$) further highlight the robustness and reproducibility of the proposed approach under repeated resampling. Near-real-time deployability refers to inference latencies; all per-sample prediction latencies are below 0.02 ms (e.g., Extra Trees 0.013–0.016 ms, Hist Gradient Boosting $\approx$ 0.001–0.002 ms, Random Forest $\approx$ 0.003–0.004 ms).

Table 3. Cross-validation performance of tree-based classifiers (mean accuracy $\pm$ standard deviation, 95% CI) for early-, mid-, and late-game datasets (the best result is highlighted with bold).

Dataset	Class Distribution	Model	Mean Acc (%) $\pm$ SD	95% CI
10 min Snapshots (414 k)	Radiant-win: 51.9% Dire-win: 48.1%	Extra Trees Classifier	64.7 ± 0.0015	[0.6458, 0.6473]
		Random Forest	64.3 ± 0.0016	[0.6417, 0.6433]
		Hist Grad Boost	65.3 ± 0.0019	[0.6523, 0.6543]
20 min Snapshots	Radiant-win: 51.5% Dire-win: 48.5%	Extra Trees Classifier	74.9 ± 0.0015	[0.7486, 0.7502]
		Random Forest	73.3 ± 0.0015	[0.7327, 0.7342]
		Hist Grad Boost	72.0 ± 0.0016	[0.7192, 0.7208]
30 min Snapshots	Radiant-win: 49.8% Dire-win: 50.2%	Extra Trees Classifier	86.4 ± 0.0007	[0.8636, 0.8643]
		Random Forest	82.1 ± 0.0012	[0.8204, 0.8216]
		Hist Grad Boost	76.3 ± 0.0011	[0.7456, 0.7467]

To quantify and visualize the contribution of individual in-game metrics to match outcome prediction, two complementary analyses were conducted using the trained Random Forest models across the early (≤ 10 min), mid (10–20 min), and late (> 20 min) match phases. The first method, impurity-based (Gini) importance (Shown in Figure 10), measures how frequently each feature contributes to node splits that reduce classification error within the ensemble. The second method, permutation-based importance (Shown in Figure 11), evaluates the decrease in test accuracy observed when the values of a given feature are randomly shuffled, providing a model-agnostic estimate of its true predictive influence. Both approaches were applied to the held-out test set using the same random split employed in model evaluation to ensure consistent comparability.

Figure 10 illustrates the impurity-based (Gini) feature importance derived from the Random Forest classifier across the early (≤ 10 min), mid (10–20 min), and late (> 20 min) match phases. In all three segments, net-worth difference, gold lead, and experience difference emerged as the most dominant predictors, highlighting their consistent role in shaping match outcomes. During the early game, economic indicators such as last hits, denies, and hero level differentials were particularly influential, reflecting the importance of initial resource accumulation. In the mid-game phase, tower status and teamfight partic-

ipation metrics gained higher importance, indicating the growing influence of territorial control and coordinated engagements. By the late-game stage, item-based advantages and kill differentials became more decisive, signifying a transition from macroeconomic to strategic dominance. The overall similarity in feature ranking across time segments demonstrates the model's stability and the persistent predictive relevance of a core subset of gameplay variables.

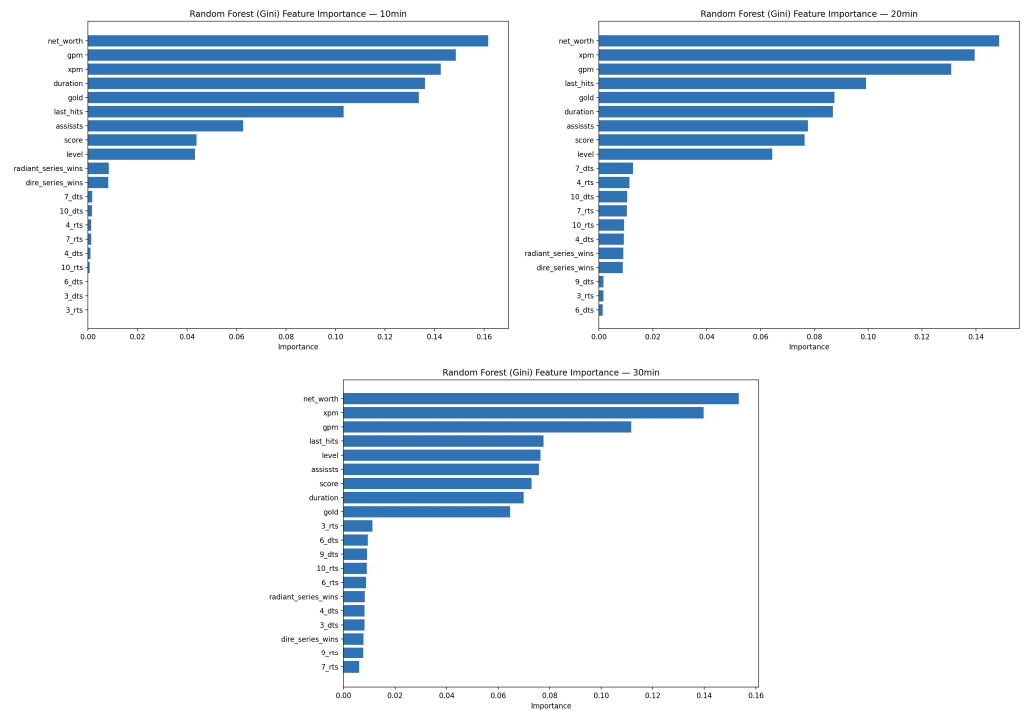


Figure 10. Impurity-based (Gini) feature importance by match phase.

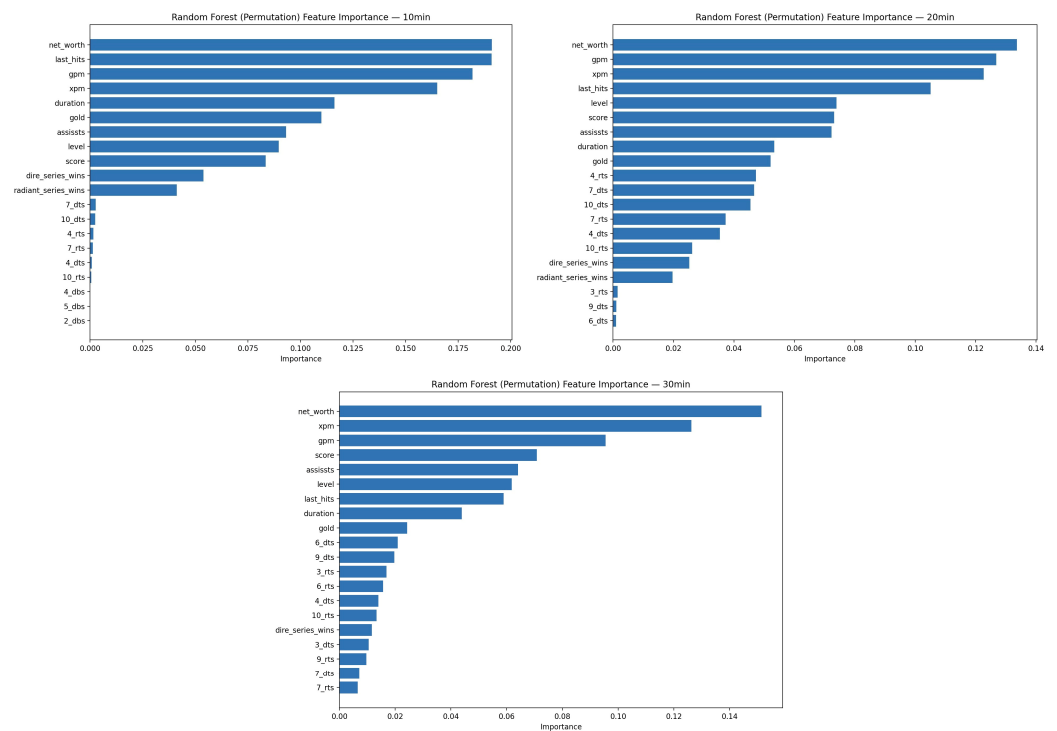


Figure 11. Permutation-based feature importance by match phase.

Figure 11 presents the permutation-based feature importance results for the same temporal segmentation, providing a model-agnostic evaluation of each variable's real predictive contribution. The permutation analysis confirms and strengthens the insights derived from impurity-based measures, showing that net-worth difference, experience gain, and tower advantage consistently produce the largest decreases in test accuracy when perturbed. This indicates that these features are central to the model's ability to distinguish winning and losing teams. Notably, in the mid- and late-game phases, item completion counts and hero level progression became more prominent, underscoring the increasing strategic complexity as matches evolve. The relatively smaller gaps between top-ranked features also suggest overlapping synergies among economic, territorial, and combat-related variables. These results collectively validate the interpretability of the decision tree ensembles and provide quantitative evidence of which in-game metrics drive predictive performance at each phase of the match.

These experimental results provide strong evidence for the effectiveness of decision tree-based algorithms in replicating the DotA Plus prediction system. The high accuracy achieved by the Extra Trees Classifier and the combined average of the models demonstrates the potential of these methods for real-time match outcome prediction in DotA 2. The consistent performance across two different matches further strengthens the validity of the approach, suggesting its generalizability to a wider range of DotA 2 games.

5. Discussion

While earlier studies such as MOBA-Slice and Hodge et al. demonstrated the potential of temporal segmentation and continuous prediction in esports analytics, the present work extends these concepts through a transparent and reproducible ensemble-based implementation. The novelty of this study lies in combining interpretable decision tree ensembles with temporal and meta-level contextual embeddings to achieve real-time prediction accuracy comparable to proprietary systems, while maintaining full methodological openness.

The Extra Trees Classifier emerged as the strongest individual performer, confirming its ability to handle high-dimensional structured data with robustness and efficiency. Its peak performance of 98.6% accuracy in late-game predictions demonstrates its capability to capture decisive game-state features.

The successful replication of the DotA Plus prediction system using decision tree algorithms has several implications. First, it provides a transparent and interpretable approach to real-time match analysis, offering insights into the factors driving the predictions. Second, it demonstrates the potential of machine learning techniques to enhance the viewer and player experience in esports by providing accurate and dynamic outcome predictions. Thirdly, it is computationally less demanding, enabling near real-time deployment. It also provides interpretable feature importance scores, allowing analysts and players to understand which variables—such as gold per minute, experience per minute, and net-worth—drive predictions. This interpretability is a crucial advantage over black-box neural networks, particularly in competitive environments where transparency fosters trust. Finally, the open-source nature of the methods employed in this study allows for further research and development in the field of esports analytics.

Feature importance analysis identified net-worth, gold per minute (gpm), and experience per minute (xpm) as key predictors of match outcomes. These findings align with the general understanding of DotA 2 gameplay, where resource advantage and experience levels are crucial factors in determining a team's success.

The segmentation of datasets by match duration revealed that predictive accuracy generally increases as the game progresses. This is likely due to the accumulation of more data and the stabilization of the game state, making it easier to discern the likely

winner. Furthermore, the inclusion of hero and item win rate features, derived from Dotabuff, significantly improved the models' performance, underscoring the importance of incorporating external knowledge sources to enhance prediction accuracy in complex domains like esports.

The integration of hero and item win rate embeddings demonstrates that blending in-game data with external meta-knowledge can further enhance prediction accuracy. This hybrid approach highlights the value of community-driven statistics (e.g., Dotabuff) in refining esports analytics models. Future work could extend this by incorporating player-specific statistics, drafting strategies, or patch-version data to account for evolving metagames.

Future extensions of this research could incorporate optimization-driven frameworks for model refinement. For instance, multi-objective optimization approaches, such as the improved non-dominated sorting genetic algorithm used in surveillance and control systems [22] provide structured strategies for balancing competing objectives like accuracy, interpretability, and computational efficiency. Integrating such frameworks with data-driven modeling could enable automated hyperparameter tuning and dynamic model adaptation for evolving esports environments.

From a deployment perspective, the decision tree-based models used in this study are well suited for real-time prediction. Inference times remained below 35 ms per instance on a standard CPU, enabling probability updates at 5–10 s intervals during a live match. Because predictions rely only on the current game state, the system can operate effectively in a streaming setting without requiring sequential processing of earlier states. Although the models were trained offline in this study, the framework could be adapted to online or incremental learning, allowing periodic retraining to accommodate evolving match dynamics or new balance patches.

6. Conclusions and Future Work

This study investigated the feasibility of replicating the DotA Plus prediction system using decision tree-based algorithms, specifically Extra Trees, Random Forest, and Hist Gradient Boosting classifiers. The experimental results demonstrate that these algorithms, particularly the Extra Trees Classifier, can effectively predict the outcome of DotA 2 matches in real-time with high accuracy. The combined average of the three models also showed strong predictive performance, suggesting the potential benefits of an ensemble approach for enhanced robustness and reliability.

Overall, the ensemble-based decision tree methods demonstrated strong and reliable predictive performance across all temporal datasets. Statistical analysis using paired t-tests confirmed that Extra Trees achieved the highest accuracy on mid- and late-game matches (20–30 min), with performance differences statistically significant at $p < 0.001$. However, in the early-game phase (≤ 10 min), Random Forest and Histogram Gradient Boosting achieved similar accuracy levels (differences $< 0.3\%$, $p > 0.05$), indicating that when less temporal and item information is available, simpler ensembles perform comparably. These findings suggest that while Extra Trees offers a strong balance of accuracy and interpretability, model choice may depend on the game phase and available contextual information, supporting adaptive ensemble configurations in future implementations.

The high accuracy and interpretability of these models make them a valuable tool for real-time match analysis and offer a promising avenue for enhancing the understanding and enjoyment of DotA 2 and other esports. The findings contribute to the growing body of research on esports analytics and pave the way for further advancements in this exciting field.

Future work could explore the use of other machine learning algorithms, such as deep learning models, to further improve prediction accuracy. Additionally, incorporating more

granular data, such as player-specific statistics and in-game events, could enhance the models' ability to capture the nuances of DotA 2 gameplay. Investigating the application of these methods to other esports titles could also broaden the impact of this research.

Author Contributions: Conceptualization, S.Y. and J.M.; methodology, S.Y. and J.M.; software, S.Y.; validation, S.Y. and J.M.; formal analysis, S.Y., J.M. and S.M.; resources, S.Y., J.M. and S.M.; writing—original draft preparation, S.Y. and J.M.; writing—review and editing, J.M. and S.M.; supervision, J.M. and S.M.; funding acquisition, J.M. and S.M. All authors have read and agreed to the published version of the manuscript.

Funding: J.M. acknowledges funding by the Agency for “Innovative Development” of the Republic of Uzbekistan; grant: UZ-N39. S.M. acknowledge University of Limpopo for financial support and the availability of computer resources.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used in this paper can be found at <https://github.com/sukhrobyangibaev/thesis-pub> (accessed on 30 September 2025).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DotA	Defense of the Ancients
MOBA	multiplayer online battle arena
AC	Associative Classification
RF	Random Forest
ETC	Extra Trees Classifier
HGB	Histogram-based Gradient Boosting
gpm	gold per minute
xpm	experience per minute
CV	Cross-validation
CI	Confidence interval
SD	Standard deviation

References

1. Akhmedov, K.; Phan, A.H. Machine learning models for DOTA 2 outcomes prediction. *arXiv* **2021**, arXiv:2106.01782. [\[CrossRef\]](#)
2. Ke, C.H.; Deng, H.; Xu, C.; Li, J.; Gu, X.; Yadamsuren, B. DOTA 2 match prediction through deep learning team fight models. In Proceedings of the 2022 IEEE Conference on Games (CoG), Beijing, China, 21–24 August 2022; pp. 96–103. [\[CrossRef\]](#)
3. Yu, L.; Zhang, D.; Chen, X.; Xie, X. MOBA-Slice: A Time Slice Based Evaluation Framework of Relative Advantage Between Teams in MOBA Games. In Proceedings of the Computer Games, Workshop on Computer Games 2018, Stockholm, Sweden, 13 July 2018; Cazenave, T., Saffidine, A., Sturtevant, N., Eds.; Springer: Cham, Switzerland, 2018; Volume 1017, pp. 23–40. [\[CrossRef\]](#)
4. Do, T.D.; Wang, S.I.; Yu, D.S.; McMillian, M.G.; McMahan, R.P. Using Machine Learning to Predict Game Outcomes Based on Player–Champion Experience in League of Legends. In Proceedings of the 16th International Conference on the Foundations of Digital Games (FDG), Montreal, QC, Canada, 3–6 August 2021; ACM: New York, NY, USA, 2021. [\[CrossRef\]](#)
5. Chowdhury, S.; Ahsan, M.; Barraclough, P. Applications of Linear and Ensemble-Based Machine Learning for Predicting Winning Teams in League of Legends. *Appl. Sci.* **2025**, *15*, 5241. [\[CrossRef\]](#)
6. Quinlan, J.R. Induction of decision trees. *Mach. Learn.* **1986**, *1*, 81–106. [\[CrossRef\]](#)
7. Mienye, I.D.; Jere, N. A Survey of Decision Trees: Concepts, Algorithms, and Applications. *IEEE Access* **2024**, *12*, 86716–86727. [\[CrossRef\]](#)
8. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [\[CrossRef\]](#)
9. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
10. Tangniyom, N.; Boonma, P. A Study on Using Machine Learning to Predict Winner in Multiplayer Online Battle Arena (MOBA) Game. *ASEAN J. Sci. Tech. Report.* **2024**, *27*, e252289. [\[CrossRef\]](#)

11. Birant, K.U.; Birant, D. Multi-Objective Multi-Instance Learning: A New Approach to Machine Learning for eSports. *Entropy* **2023**, *25*, 28. [[CrossRef](#)] [[PubMed](#)]
12. Mattiev, J.; Kavšek, B. CMAC: Clustering Class Association Rules to Form a Compact and Meaningful Associative Classifier. In Proceedings of the International Conference on Machine Learning, Optimization, and Data Science, 6th International Conference, LOD 2020, Siena, Italy, 19–23 July 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 372–384.
13. Mattiev, J.; Kavšek, B. Distance based clustering of class association rules to build a compact, accurate and descriptive classifier. *Comput. Sci. Inf. Syst.* **2021**, *18*, 791–811. [[CrossRef](#)]
14. Rokach, L. Ensemble-based classifiers. *Artif. Intell. Rev.* **2010**, *33*, 1–39. [[CrossRef](#)]
15. Poosala, V. Histogram-Based Estimation Techniques in Database Systems. PhD Thesis, The University of Wisconsin-Madison, Madison, WI, USA, 1997.
16. Sharaff, A.; Gupta, H. Extra-Tree Classifier with Metaheuristics Approach for Email Classification. In *Advances in Computer Communication and Computational Sciences*; Bhatia, S., Tiwari, S., Mishra, K., Trivedi, M., Eds.; Advances in Intelligent Systems and Computing; Singapore, 2019; Volume 924, pp. 189–197. [[CrossRef](#)]
17. Hodge, V.J.; Devlin, S.M.; Sephton, N.J.; Block, F.O.; Cowling, P.I.; Drachen, A. Win Prediction in Multi-Player Esports: Live Professional Match Prediction. *IEEE Trans. Games* **2021**, *13*, 368–379. [[CrossRef](#)]
18. Losada-Rodríguez, J.; Castillo, P.A.; Mora, A.; García-Sánchez, P. The Explainability of Machine Learning Algorithms for Victory Prediction in the Video Game Dota 2. *Comput. Sci. Math. Forum* **2025**, *11*, 26. [[CrossRef](#)]
19. Wang, N.; Li, L.; Xiao, L.; Yang, G.; Zhou, Y. Outcome Prediction of DOTA2 Using Machine Learning Methods. In Proceedings of the 2018 International Conference on Mathematics and Artificial Intelligence (ICMAI 2018), Chengdu, China, 20–22 April 2018; pp. 61–67. [[CrossRef](#)]
20. Lee, J.; Kim, N. Development of Machine Learning-Based Indicators for Predicting Comeback Victories Using the Bounty Mechanism in MOBA Games. *Electronics* **2025**, *14*, 1445. [[CrossRef](#)]
21. Yangibaev, S.R.; Yangibaeva, M.R. Comparative Analysis of Decision Tree Algorithms for DotA 2 Match Outcome Prediction. In Proceedings of the 2025 IEEE 26th International Conference of Young Professionals in Electron Devices and Materials (EDM), Altai, Russian Federation, 27 June–1 July 2025; pp. 2290–2294. [[CrossRef](#)]
22. Liao, R.; Zhang, Y.; Wang, H.; Zhao, T.; Wang, X. Multi-objective optimisation of surveillance camera placement for bridge–ship collision early-warning using an improved non-dominated sorting genetic algorithm. *Adv. Eng. Inform.* **2025**, *69*, 103918. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.