Predicting Player Performance in Valorant E-Sports using Random Forest Algorithm: A Data Mining Approach for Analyzing Match and Agent Data in Virtual Environments

Adi Suryaputra Paramita[1,*], Jusak Jusak[2]

[1]School of Information Technology, Universitas Ciputra, CBD Boulevard Citraland, Surabaya 60219, Indonesia

[2]James Cook University (Singapore Campus), 149 Sims Dr, Singapore 387380

ABSTRACT

This study presents a data-driven approach to predict player performance in Valorant, an increasingly popular e-sport, using a Random Forest machine learning model. As e-sports continue to evolve within the metaverse, the need for strategic optimization and player selection has become critical. By analyzing a dataset containing player statistics from the Valorant Champion Tour (VCT), we aimed to predict player Rating, a key performance indicator. The dataset includes various metrics such as Kills Per Round, Average Combat Score (ACS), Clutch Success Ratio, and Kills:Deaths. After preprocessing the data, which involved handling missing values and feature engineering, the dataset was split into training and testing sets (80% and 20%, respectively). The Random Forest model, with 100 estimators and a maximum depth of 10, was trained on the processed data. The model's performance was evaluated using regression metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2). The results demonstrated that the model could predict player performance with a high degree of accuracy, with an R^2 value of 0.8831, meaning it explained 88.31% of the variance in player ratings. Furthermore, Kills Per Round emerged as the most significant feature, followed by Kill, Assist, Trade, Survive Ratio and Average Damage Per Round. These insights suggest that key metrics like kills and damage output are crucial for predicting player success. This study not only provides a comprehensive framework for predicting Valorant player performance but also demonstrates the potential of data mining in optimizing e-sports strategies. The findings contribute to the growing body of research on virtual gaming environments and offer actionable insights for teams in the metaverse, enabling data-driven decision-making to enhance performance and strategic outcomes.

Keywords Valorant, E-Sports, Random Forest, Machine Learning, Player Performance

Submitted: 28 March 2025
Accepted: 5 June 2025
Published: 20 November 2025

Corresponding author
Adi Suryaputra Paramita,
adi.suryaputra@ciputra.ac.id

Additional Information and Declarations can be found on page 309

DOI: 10.47738/ijrm.v2i4.39

INTRODUCTION

The integration of e-sports within the metaverse has emerged as a significant factor in reimagining competitive virtual activities. At its core, the metaverse is defined as an interconnected virtual space that leverages advanced immersive technologies, such as virtual reality and augmented reality, enabling users to engage in continuous and immersive experiences [1]. This digital evolution facilitates the creation of complex, persistent online worlds were competitive

How to cite this article: A. S. Paramita and J. Jusak, "Predicting Player Performance in Valorant E-Sports using Random Forest Algorithm: A Data Mining Approach for Analyzing Match and Agent Data in Virtual Environments," *Int. J. Res. Metav.*, vol. 2, no. 4, pp. 292-311, 2025.

activities, including e-sports, can thrive. The metaverse's infrastructure ensures persistency and user-generated content, supporting the expansion of e-sports from traditional gaming platforms to more integrated and immersive virtual environments [2].

In recent years, e-sports have increasingly been redefined as multifaceted competitive engagements that blend elements of sport, entertainment, and social interaction. Research has shown that e-sports encompass diverse gaming genres through both professional and amateur competitions, involving sophisticated hardware and interactive platforms [3]. The concept of "MetaSport" reflects the influence of metaverse technologies on the sports domain, where digital avatars, virtual reality environments, and real-time competitive interactions create experiences that expand upon traditional sports [4]. Furthermore, classifying gamers into typologies such as achievement-oriented and competition-driven illustrates the diverse motivations behind e-sports participation, which are amplified in metaverse contexts [5].

The synergies between e-sports and the metaverse are further evidenced by empirical research demonstrating how such virtual frameworks contribute to competitive sports experiences. For instance, the deployment of metaverse-based virtual reality sport arenas has been shown to enhance both the performance and mental engagement of athletes and gamers, enriching the competitive experience [6]. Additionally, the metaverse serves not only as a platform for gameplay but also as a crucial tool for content creation, live-streaming events, and social interaction among participants, which are critical drivers for the growing popularity of competitive e-sports [7]. Consequently, the metaverse is redefining e-sports by providing a fully immersive and interactive domain that supports both gameplay and the significant economic and cultural infrastructure surrounding competitive gaming.

Valorant, developed by Riot Games, has rapidly ascended as one of the most prominent titles in the competitive gaming arena due to its intricate blend of first-person shooter mechanics, unique character abilities, and high-level strategic team play [8]. The game distinguishes itself through its dynamic gameplay, which requires a balanced approach between individual skill and coordinated team strategies, fostering a competitive environment that appeals to both casual and professional gamers [9]. This combination of engaging mechanics and competitive depth has solidified its position within traditional esports and has implications for emerging virtual ecosystems such as the metaverse.

In the context of the metaverse, Valorant's inherent design elements—such as immersive gameplay, interactive character elements, and strategic competitive scenarios—allow it to be integrated into a broader digital landscape that emphasizes persistent virtual experiences [10]. As the metaverse continues to evolve, it offers a platform where games like Valorant can be experienced in novel ways. These platforms enable gamers to participate in immersive arenas that simulate real-world competitive environments, thereby extending the reach and impact of Valorant's competitive ecosystem. The metaverse's capacity to support various modes of interaction, content creation, and social connectivity further underscores Valorant's relevance, as players can engage with the game beyond traditional gaming setups, thereby fostering a more interconnected and enriched community experience.

Moreover, Valorant serves as an example of how competitive virtual games can adapt to new technological paradigms. The game's popularity is bolstered by its psychological appeal—stemming from well-crafted game design that leverages competitive instincts and strategic thinking [8]—which aligns with the immersive, user-centric experiences that the metaverse strives to deliver. Additionally, career pathways in esports increasingly recognize titles such as Valorant as viable options, with educational and professional sectors beginning to acknowledge the game within the context of competitive and technological advancement in virtual environments [11]. This convergence of competitive gaming and metaverse capabilities indicates that Valorant is well-positioned to remain at the forefront of both esports and next-generation interactive experiences.

In the growing field of e-sports, particularly in games like Valorant, understanding player performance is crucial for both professional teams and fans. However, a significant gap exists in the predictive insights available regarding player behavior, performance, and match outcomes. Despite the wealth of match data, such as player kills, deaths, agent picks, and round results, the ability to predict individual player performance based on these variables remains underexplored. This lack of predictive modeling limits the potential for optimizing player strategies and improving team compositions, both of which are vital for success in competitive gaming tournaments like the Valorant Champions Tour (VCT).

The objective of this study is to address this gap by utilizing data mining techniques, specifically the Random Forest algorithm, to predict player performance in Valorant e-sports tournaments. By analyzing historical match data, including individual player stats such as kills per round, average combat score, and agent usage, the study aims to build a robust model that forecasts player performance based on past behaviors. The primary goal is to enable better decision-making for teams in terms of player selection, agent strategies, and overall match predictions, ultimately enhancing team performance during competitions.

This paper contributes to the understanding of virtual player dynamics within e-sports by applying data mining methods to Valorant match data. Using the Random Forest algorithm, which is well-known for its ability to handle large datasets with high accuracy, the study aims to uncover patterns in player behavior that can predict outcomes. By offering a method for forecasting player performance, this research not only deepens the understanding of player and team dynamics in competitive games but also provides valuable insights into how data-driven approaches can be leveraged within the metaverse and virtual gaming environments.

Literature Review

E-Sports Data Analytics

Previous works on analyzing player performance in e-sports have typically focused on a combination of physiological, cognitive, and behavioral metrics to derive insights from in-game data and external factors that may affect performance. In particular, studies have employed wearable sensors and real-time monitoring to gauge the impact of stress, heart rate variability, and psychological arousal on player performance during competitive play. For instance, Behnke et al. [12] examined cardiovascular responses and social challenges in competitive video gaming, highlighting how psychological stress can influence both physiological markers and performance outcomes. Similarly, Watanabe et al. [13] investigated the effects of competitive play on professional e-sports players' physiological state, linking sympathetic nervous system activation to in-game decision-making processes. Such approaches have provided a blueprint for quantitatively assessing performance dynamics in fast-paced formats, including first-person shooter games that share common competitive elements with Valorant.

Comparative approaches in e-sports data analytics have also integrated movement science principles and advanced computational models to parse the multifaceted nature of player performance. Iwatsuki et al. [14] demonstrated that leveraging movement science can optimize esports performance by analyzing biometric and in-game movement data, revealing insights that may have applicability for game titles like Valorant. Furthermore, systematic reviews like that by Dmello [15] highlight how data analytics frameworks, combined with real-time data capture, can aid coaches and analysts in identifying key indicators of both individual and team success in competitive settings. These methodologies underscore the need for nuanced performance analytics that incorporate both in-game behavior and physical readiness—essential for accurately assessing competitive performance in games with complex tactical demands.

Moreover, existing research has noted that while broad analyses of player performance have been successfully applied to several e-sports titles, there is a limited number of studies specifically focused on predictive analytics in Valorant. Trotter et al. [16] found an association between practice time and in-game rank, suggesting that performance analytics should incorporate a holistic perspective of player experience and training intensity—a recurring theme in many e-sports studies. Meanwhile, Sousa et al. [17] evaluated physiological and cognitive functions following competitive gaming sessions, proposing that transient cognitive states are influential in determining in-game performance outcomes. Although these studies primarily address general e-sports contexts or alternate FPS titles, their methodologies and findings provide valuable insights for developing tailored analytics models for Valorant. This highlights an opportunity for future work to adapt these analytical frameworks to capture the unique in-game mechanics, team dynamics, and individual skill sets characteristic of Valorant and similar competitive digital sports.

Data Mining in E-Sports

Recent research demonstrates that machine learning and data mining algorithms have been widely applied in sports analytics to predict performance, and these methodologies present promising opportunities for application in esports as well. Although most of the seminal works focus on traditional sports, their methodologies provide a solid foundation for adapting similar techniques to the unique environment of competitive gaming.

In traditional sports, data mining approaches have successfully predicted outcomes by modeling complex interactions and extracting critical performance features. For instance, Soto-Valero [18] applied support vector machines and comparative data mining methods to predict win-loss outcomes in Major League Baseball, showcasing the potential for selecting game strategies and forecasting performance through the analysis of relevant game situations. Such approaches emphasize the utility of sophisticated algorithms in handling high-dimensional data, a requirement that is equally vital in esports where player behavior and in-game events generate massive datasets.

Moreover, sports-related data mining studies such as [19], [20] have leveraged advanced ensemble methods, including random forest algorithms and gradient boosting methods, to optimize evaluation technologies and predict game outcomes. These studies underline how ensemble machine learning methods are capable of handling non-linear relationships and interactions between multiple features, thereby enhancing prediction accuracy. When applied to esports contexts, similar hybrid approaches could be tailored to capture the complex interplay of individual player skills, team dynamics, and in-game decision-making processes that define competitive gaming performance.

Further supporting this view, [21] reviewed the application of various machine learning algorithms for sports result prediction and emphasized the success of classification techniques in forecasting match outcomes. Although their work primarily focused on traditional sports, it reinforces the idea that the systematic use of historical performance data and classification algorithms could be transferred to the domain of esports. Given that esports share many similarities with conventional sports in terms of performance dynamics—albeit in a digital environment—the adaptation of these approaches could lead to robust predictive models in gaming scenarios.

While these studies provide robust methodologies and promising results in traditional sports, there is still a notable gap regarding their direct application to esports. The unique characteristics of gaming—such as virtual environments, rapid decision-making, and digital interaction mechanisms—demand specific adaptations of these data mining frameworks. The success of predictive analytics in traditional sports, however, suggests that with appropriate modifications and the integration of game-specific features, these methods can be effectively applied to predict player performance in competitive gaming environments.

Random Forest Algorithm

The Random Forest algorithm is an ensemble learning method that constructs multiple decision trees and aggregates their predictions to enhance classification accuracy and robustness compared to individual decision trees. It is particularly effective for managing high-dimensional data with numerous explanatory variables, which is often observed in esports performance metrics.

A Random Forest is created by generating a large number (T) of decision trees from bootstrapped samples of the original dataset. At each node in a tree, rather than examining all available features, the algorithm selects a random subset of features. This random feature selection coupled with bootstrap aggregation (bagging) helps to decrease the model's variance and mitigates the risk of overfitting [22].

In predicting player performance, especially in competitive esports such as Valorant, the Random Forest algorithm offers several advantages. Esports generate extensive performance data including in-game actions, decision-making metrics, physiological signals, and historical performance records.

Random Forests can effectively manage these diverse data sources by identifying the importance of various performance indicators through their inherent feature selection mechanism. Recent advancements have further refined the Random Forest framework into forms like multi-view rank-based Random Forest models, designed to capture intricate patterns in esports data where multiple perspectives (e.g., player behavior, team dynamics, and in-game events) are vital.

Additionally, the method's robustness to noise and capacity to generalize from bootstrapped samples make it particularly applicable to the often-variable and rapidly evolving scenarios in esports. Predictive studies, such as those performed in the context of Rocket League, have successfully employed Random Forests to identify key performance metrics that correlate strongly with match outcomes and player rankings [23]. This approach is adaptable to esports titles like Valorant, where the actions and strategies of various team members must be integrated to assess both individual and collective performance outcomes.

Related Work on Feature Selection

Research in sports performance analytics has consistently highlighted the importance of feature selection when attempting to predict outcomes, whether in traditional sports or in competitive digital gaming environments. Prior studies focusing on traditional sports have demonstrated that selecting key performance indicators (KPIs) significantly improves model accuracy by filtering out redundant or irrelevant data. For example, [24] investigated rugby league match-play, where correlation-based feature selection identified a set of key indicators—demonstrating that refined feature sets, such as technical-tactical performance metrics, yield improved classification outcomes. These findings are transferable to esports settings, where analogous KPIs such as kill/death (K/D) ratio and kills per round are instrumental in gauging player effectiveness.

Similarly, [25] emphasized that feature selection is essential for reducing the dimensionality of datasets, thereby lowering computational costs and mitigating the risk of overfitting in machine learning models. Their survey highlights that the careful removal of irrelevant variables is paramount, particularly in gaming analytics where vast amounts of in-game data can obscure critical performance markers. In the realm of esports, metrics such as K/D ratio, headshot accuracy, and assists are routinely employed as indicators of in-game performance; these metrics serve a role analogous to traditional sports performance indicators, enabling robust predictive and classification models.

Taken together, these studies underscore that in both traditional and esports domains, carefully selected features such as the K/D ratio, kills per round, and supplementary in-game performance metrics are crucial for developing accurate predictive models. The integration of feature selection techniques not only improves the performance of data mining and machine learning models but also provides a more transparent understanding of the underlying factors that drive competitive success. This synthesis of research establishes a clear precedent for the continued exploration of feature selection in esports analytics, particularly for fast-paced, competitive environments like those found in modern digital games.

Method

Figure 1 illustrates the overall workflow of the proposed Player Performance Prediction Pipeline (PPPP) used in this study. The flowchart depicts each major stage of the analytical process, starting from data loading, cleaning, and preprocessing, through feature engineering and model training, to model evaluation and output generation. It shows how the dataset is first inspected for percentage values and missing data, followed by transformations such as converting percentage values into numerical ratios, splitting compound features, and creating the Kill-Death Impact Index (KDI). The pipeline then proceeds to data partitioning, Random Forest model training, and performance evaluation using MAE, RMSE, and R^2 metrics. This structured flow, as shown in figure 1, ensures a consistent and reproducible procedure for predicting player performance based on game statistics.

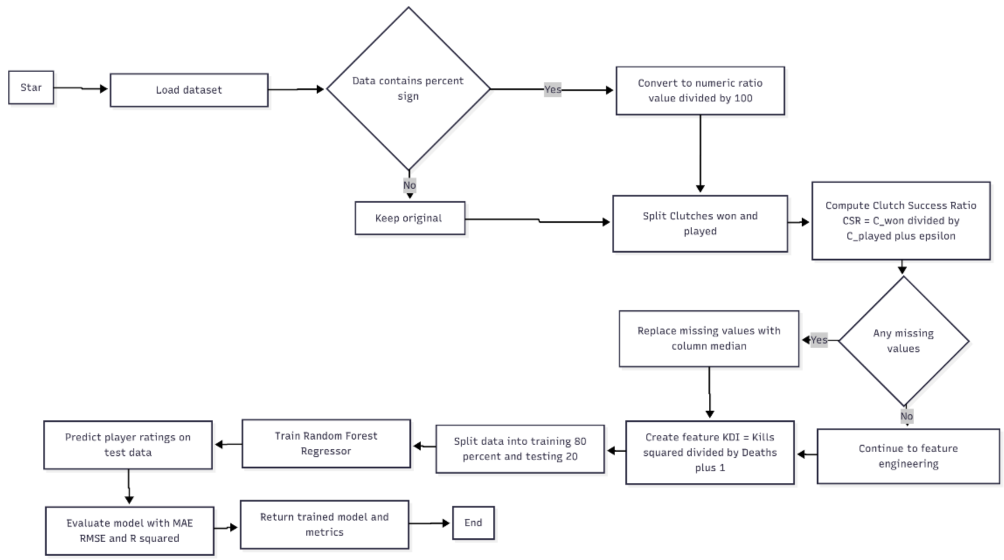

Figure 1 Research Flow

Data Loading and Preprocessing

The first step in the analysis involves loading the dataset, which consists of detailed player statistics, match data, and agent performance from the Valorant Champion Tour. The data is stored in the player_stats.csv file and is loaded into a pandas DataFrame. Upon loading, we check the integrity of the dataset by printing out an overview, which includes the first few rows and some summary information about the dataset. This initial inspection helps ensure that the dataset has been loaded correctly and contains the expected columns. If there are any issues, such as missing or malformed data, an error message is displayed, and the script exits to prevent further issues during the analysis.

Next, we filter the dataset to focus on the aggregated match statistics, which are critical for the prediction of player performance. These aggregated rows, which contain data for players who have performed across multiple agents (indicated by commas in the 'Agents' column), are essential for the analysis as they summarize overall match performance rather than individual agent statistics. If no aggregated rows are found, a fallback mechanism is used to proceed with the entire dataset, although this may reduce the interpretability and relevance of the results. This filtering ensures that we are analyzing data that represents the player's contribution in a match as a whole, rather than their performance with specific agents.

Exploratory Data Analysis (EDA)

Once the data was properly loaded and cleaned, an Exploratory Data Analysis (EDA) was conducted to identify structural patterns and potential relationships among the variables. Descriptive statistics were computed for all numerical features, such as *Kills*, *Deaths*, *Average Combat Score (ACS)*, and *Rating*. This analysis provided insight into the central tendency, variability, and spread of performance indicators.

Visual analysis was then employed to deepen understanding of data distribution and inter-variable associations. Histograms were generated to evaluate the skewness of the *Rating* variable, while scatter plots were used to explore how *Kills* and *Deaths* relate to player performance. In these plots, color gradients and marker sizes represented *Rating*, enabling an intuitive assessment of outliers and clustering patterns. A correlation heatmap was also developed to reveal the degree of linear association among all numerical variables, highlighting those most predictive of overall performance.

To capture non-linear relationships that standard correlations might overlook, the Mutual Information (MI) measure was computed between features. This information-theoretic approach quantifies dependency using the following expression:

$$MI(X,Y) = \sum_{x\in X}\sum_{y\in Y} p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right) \quad (1)$$

The MI value identifies hidden dependencies between features, offering a deeper understanding of the complex interactions influencing player outcomes.

Data Preprocessing and Feature Engineering

Before model training, the dataset underwent a series of preprocessing transformations to standardize and enrich its features. Columns containing percentage values, such as *Headshot %* or *Clutch Success %*, were converted into decimal ratios (e.g., 47% → 0.47) to ensure proper numerical treatment. Compound statistics such as *Clutches (won/played)* were split into two separate columns, *clutches_won* and *clutches_played*, which allowed for the creation of derived features like *Clutch Success Ratio*.

To manage incomplete data, missing numerical values were imputed using the median of their respective columns. This approach provides resilience to outliers and maintains the stability of model training. Formally, each missing value X_i was treated as follows:

$$X_i' = \begin{cases} \text{median}(X_i), & \text{if } X_i \text{ is missing} \\ X_i, & \text{otherwise} \end{cases} \quad (2)$$

Irrelevant variables such as *Tournament*, *Stage*, and *Match Type* were excluded, as they contribute little to performance prediction. Following cleaning, a new metric—the Kill-Death Impact Index (KDI)—was engineered to amplify the importance of high-impact players by penalizing frequent deaths:

$$KDI = \frac{Kills^2}{Deaths + 1} \tag{3}$$

This transformation highlights players who maintain strong kill performance while minimizing deaths, a valuable feature for competitive gaming analysis.

Model Training and Evaluation

The predictive model selected for this research was the Random Forest Regressor, chosen for its robustness, interpretability, and ability to manage non-linear feature interactions. The dataset was divided into training and testing subsets, with 80% used for model fitting and 20% reserved for validation. During training, Out-of-Bag (OOB) estimation was used to internally evaluate model performance without the need for additional data, defined as:

$$E_{OOB} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_{OOB,i})^2 \tag{4}$$

Model performance was measured using multiple regression metrics to capture different error perspectives. Mean Absolute Error (MAE) was used to evaluate the average deviation of predictions, while Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) emphasized larger deviations. The model's explanatory power was quantified through the coefficient of determination, R^2, calculated as:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \tag{5}$$

A visualization comparing predicted and actual ratings was also generated. The closer the data points align to the diagonal reference line, the stronger the model's predictive consistency.

Results Visualization and Interpretation

Upon evaluation, the Random Forest model revealed the relative importance of each feature in predicting player performance. The visualization of feature importance scores demonstrated that *Kills*, *ACS*, and *Clutch Success Ratio* had the most substantial influence on the predicted *Rating*. These results indicate that both offensive and clutch abilities significantly contribute to competitive success.

To provide a more interpretable explanation of model behavior, Shapley Values (ϕ_i) were employed. Derived from cooperative game theory, this method quantifies the contribution of each feature to a given prediction:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!}[v(S \cup \{i\}) - v(S)] \tag{6}$$

By attributing partial contributions to each variable, Shapley analysis helps reveal which features most influence model predictions, offering transparency and explainability beyond global importance metrics.

Algorithm 1 Random Forest–Based Player Rating Estimation Algorithm (RF-PREA)

Step 1 – Data Loading
Load the dataset from file:

$$D \leftarrow \text{Load("player\_stats.csv")}$$

Then check data structure, column consistency, and completeness.

Step 2 – Data Cleaning
For each feature $x_i \in D$:
if x_i contains the character '%', then convert the value into a numeric ratio:

$$x_i' = \begin{cases} \frac{v_i}{100}, & \text{if } x_i = v_i\% \\ x_i, & \text{otherwise} \end{cases}$$

End if
This ensures that percentage values (e.g., 47%) are standardized to decimal format (e.g., 0.47).

Step 3 – Feature Splitting
Split the compound feature "Clutches (won/played)" into two separate numerical columns:

$$C_{\text{won}}, C_{\text{played}} \leftarrow \text{Split("Clutches(won/played)")}$$

Then compute the **Clutch Success Ratio (CSR)** for each player:

$$CSR = \frac{C_{\text{won}}}{C_{\text{played}} + \varepsilon}$$

where ε is a small constant to prevent division by zero.

Step 4 – Handling Missing Values
For each feature $x_i \in D$:
if x_i contains missing entries, then replace with the median value of that column:

$$x_i' = \begin{cases} \text{median}(x_i), & \text{if missing} \\ x_i, & \text{otherwise} \end{cases}$$

End if

Step 5 – Feature Engineering
For each player p_i:
compute a new feature called **Kill-Death Impact Index (KDI)** as

$$KDI_i = \frac{Kills_i^2}{Deaths_i + 1}$$

End for
This non-linear metric emphasizes kill efficiency while penalizing frequent deaths.

Step 6 – Data Partitioning
Split the preprocessed dataset into training and testing sets:

$$(D_{\text{train}}, D_{\text{test}}) \leftarrow \text{Split}(D, 0.8, 0.2)$$

Step 7 – Model Training
Train the **Random Forest Regressor** model on the training data:

$$M \leftarrow \text{Train}(\text{RandomForestRegressor}, (X_{\text{train}}, Y_{\text{train}}))$$

Step 8 – Prediction
Use the trained model M to predict the player ratings on the test data:

$$\hat{Y} = M(X_{\text{test}})$$

Step 9 – Model Evaluation
Compute the following performance metrics:

$$MAE = \frac{1}{m}\sum_{i=1}^{m} | \hat{y}_i - y_i |$$

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m} (\hat{y}_i - y_i)^2}$$

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

Then compare actual versus predicted values to visualize model accuracy.

Step 10 – Output
Return the trained model and its performance metrics:

$$\text{return } M^*, \{MAE, RMSE, R^2\}$$

Result and Discussion

Data Overview and EDA Result

The initial dataset, which contains 15,030 entries across 25 columns, was successfully loaded. Upon inspection, the dataset includes key player statistics such as Kills, Deaths, Assists, Rating, and Average Combat Score (ACS). The dataset is comprehensive, capturing performance metrics for players in the Valorant Champion Tour (VCT), with data spanning various tournaments, stages, and match types. After filtering the dataset to focus on aggregated match statistics (where players have stats across multiple agents), the number of rows was reduced to 4,061, ensuring that only meaningful match summaries were included in the analysis.

Exploratory Data Analysis (EDA) was performed on the filtered dataset to understand the distribution of key variables. The Rating variable, representing player performance, ranged from 0.2 to 2.13, with a mean of approximately 0.99, indicating a fairly balanced distribution of performance levels across players. The dataset also showed that Average Combat Score (ACS) had a wide range, with values between 69 and 393, highlighting significant variation in player efficiency. Other important metrics like Kills:Deaths, Average Damage Per Round, and Kills Per Round also exhibited substantial variability, with the Kills Per Round ranging from 0.21 to 1.44, reflecting different levels of player engagement and performance in matches.

Figure 2 depicts the distribution of Average Combat Score (ACS) for players. The plot shows a right-skewed distribution, with a peak around 200, indicating that most players tend to have an ACS between 150 and 250. The distribution has a slight tail on the right, suggesting that while the majority of players have a moderately high ACS, there are a few players with significantly higher scores. This pattern indicates that most players are performing at a competent level, but there are some outliers who perform exceptionally well. The accompanying kernel density estimate (KDE) curve further confirms the shape of the distribution, with the peak aligned around 200 and a gradual decline as ACS increases.

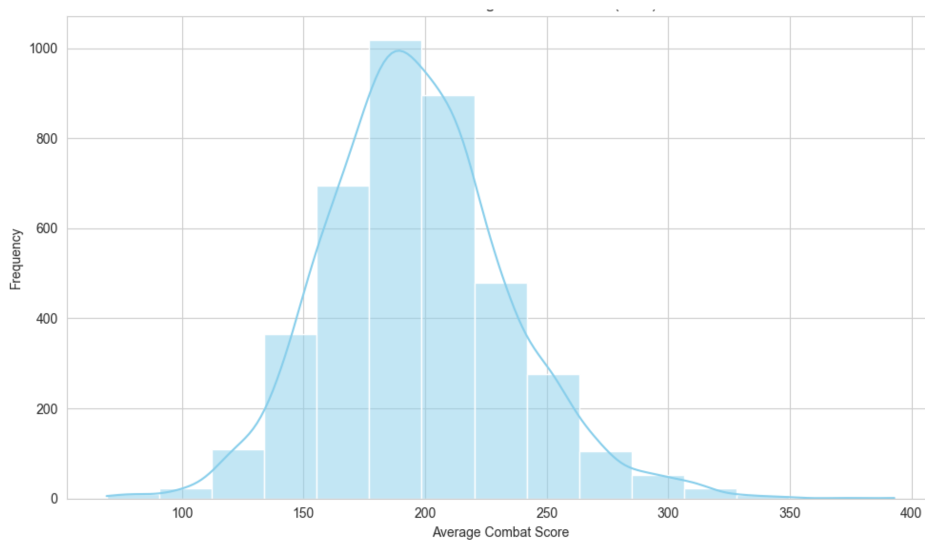

Figure 2 Distribution of Average Combat Score (ACS)

Figure 3 provides a heatmap showing the pairwise correlations between various numerical features in the dataset. The heatmap uses color gradients to represent correlation strength, with shades of blue indicating negative correlations and shades of red indicating positive correlations. For instance, Kills and Deaths exhibit a strong positive correlation of 0.99, suggesting that players with higher kills typically also have higher deaths. Kills Per Round and Rating are positively correlated (0.87), indicating that players with more kills per round tend to have higher ratings, which is a direct reflection of better performance. Clutch Success Ratio and Kills Per Round show a moderate positive correlation, signifying that players who succeed in clutch situations often have higher kill rates. These correlations provide a clear understanding of which metrics are most closely linked to player performance, with Kills, Deaths, and ACS being key predictors.

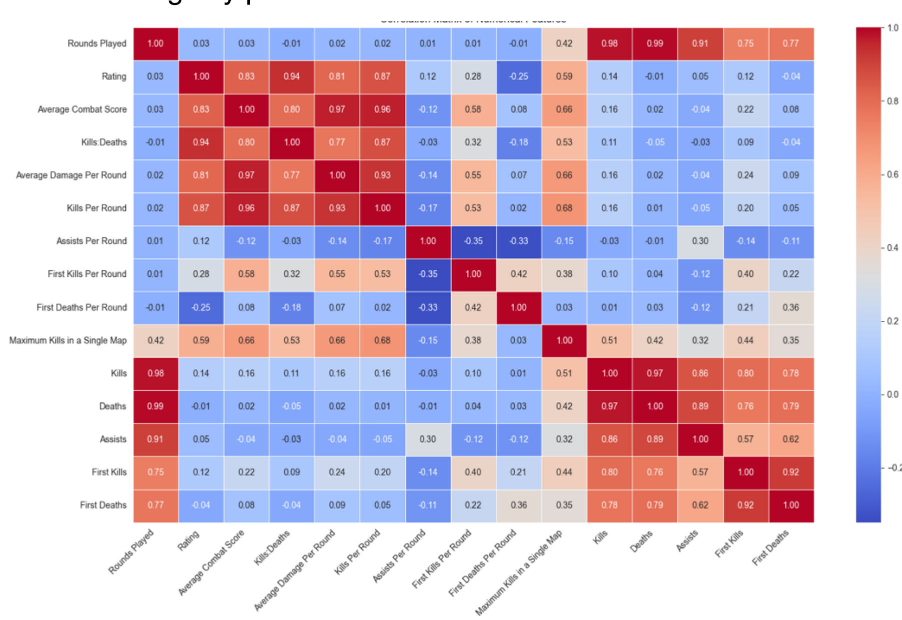

Figure 3 Correlation Matrix of Numerical Features

Figure 4 is a scatter plot that shows the relationship between the number of Kills and Deaths for players, with each point's color representing the player's Rating. The plot highlights a strong positive relationship between Kills and Deaths, where players with more kills generally have more deaths as well. This relationship is further emphasized by the color gradient, with lower-rated players (shown in purple) concentrated in the lower-left region, where kills are fewer, and deaths are relatively higher. Conversely, higher-rated players (shown in green and yellow) appear towards the upper right, where both kills and deaths are higher, but the Rating remains positively correlated with the number of kills. This visualization supports the idea that players who are highly engaged in the game (with more kills) tend to perform better, although they may also have more deaths.

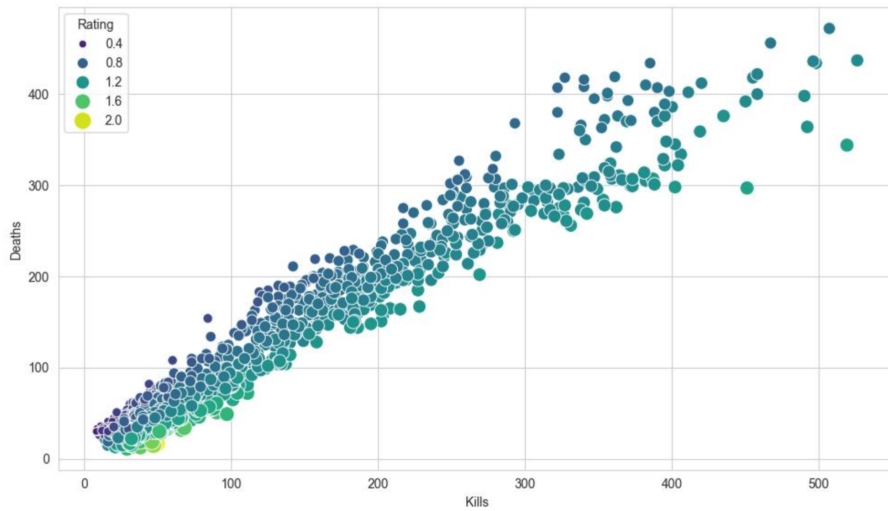

Figure 4 Scatter Plot of Kills vs Deaths (Colored by Rating)

Figure 5 shows the distribution of the Rating variable, which represents overall player performance. The distribution appears to be normal, with the majority of players concentrated around the middle range of the rating scale, near 1.0. The graph also shows a slightly skewed tail to the right, with fewer players exhibiting very low or very high ratings. The kernel density curve aligns with this, peaking near 1.0, which suggests that most players perform at a moderate level, with relatively fewer players reaching the extreme ends of the performance scale. The figure reinforces the idea that Valorant's player base is generally composed of average performers, with a smaller group achieving exceptionally high or low ratings.

Before model training, several preprocessing steps were carried out. The dataset contained missing values in various columns, such as Rating, Headshot Ratio, and Clutch Success Ratio. For these columns, missing values were handled by filling them with the median value of each respective column, ensuring that the imputation method was robust to outliers and did not distort the distribution of values. For example, the Rating column had 1,066 missing entries, which were imputed with the median value of 0.99. Similarly, other columns like Kill, Assist, Trade, Survive Ratio and Average Damage Per Round had missing values, which were filled with their respective median values of 0.72 and 129.00. After filling missing values, the dataset no longer contained any null entries, making it ready for model training.

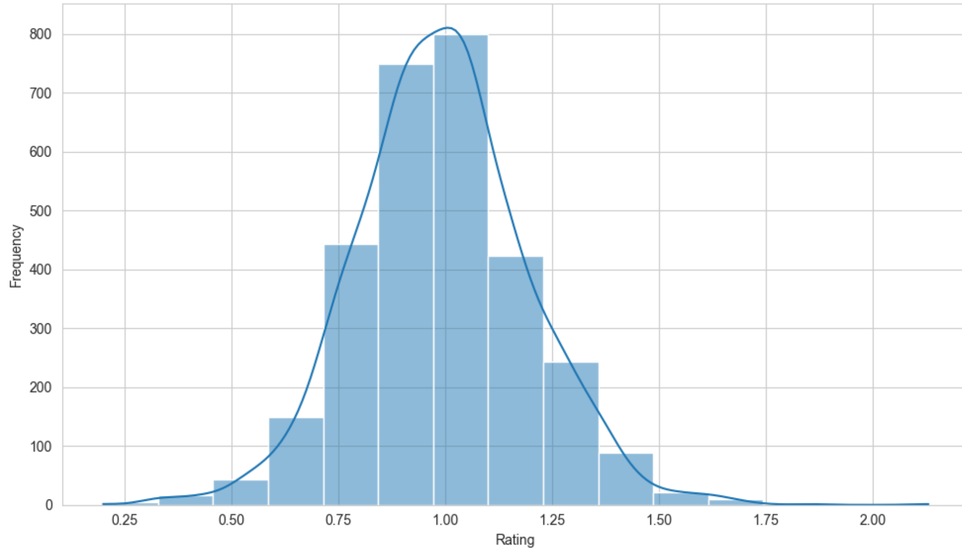

Figure 5 Distribution of Player Rating

Additionally, columns with non-numeric or redundant data, such as Tournament, Stage, and Match Type, were dropped as they were not needed for the predictive task. The final feature set consisted of 17 numerical features, including Kills Per Round, Clutch Success Ratio, and Assists Per Round, while the target variable, Rating, remained the focus for prediction. This preprocessing ensures that the data is clean and appropriately formatted for training a machine learning model, allowing for more accurate predictions of player performance.

Finding From Model Training and Evaluation

The Random Forest Regressor was trained to predict Rating, the target variable representing player performance, based on the selected features. The dataset was split into an 80% training set and a 20% testing set, with 3,248 rows used for training and 813 rows reserved for evaluation. The Random Forest model, with 100 estimators and a maximum depth of 10, was trained on the processed features, and its performance was evaluated using several regression metrics. The evaluation results were promising, with a Mean Absolute Error (MAE) of 0.0400, indicating that the model's predictions were within a narrow margin of error. The Mean Squared Error (MSE) was 0.0034, and the Root Mean Squared Error (RMSE) was 0.0583, both suggesting that the model was able to capture the underlying patterns in the data with minimal error.

The R-squared (R^2) value was 0.8831, which demonstrates that the model explains approximately 88.31% of the variance in player ratings, indicating a strong predictive capability. These metrics suggest that the Random Forest model was highly effective in predicting player performance based on the available features, providing reliable predictions that can be utilized for strategic decision-making in e-sports teams.

Figure 6 presents a comparison between the actual player ratings and the predicted ratings generated by the Random Forest model. Each point represents a player, with the x-axis showing the actual rating and the y-axis showing the predicted rating. The red dashed line represents the ideal scenario where the predicted rating matches the actual rating. As observed, the points are generally close to this line, indicating that the model has performed well, with most predicted ratings falling near their actual counterparts. This visual representation reinforces the strong predictive ability of the model, suggesting that the Random Forest algorithm was effective in capturing the patterns in player performance.

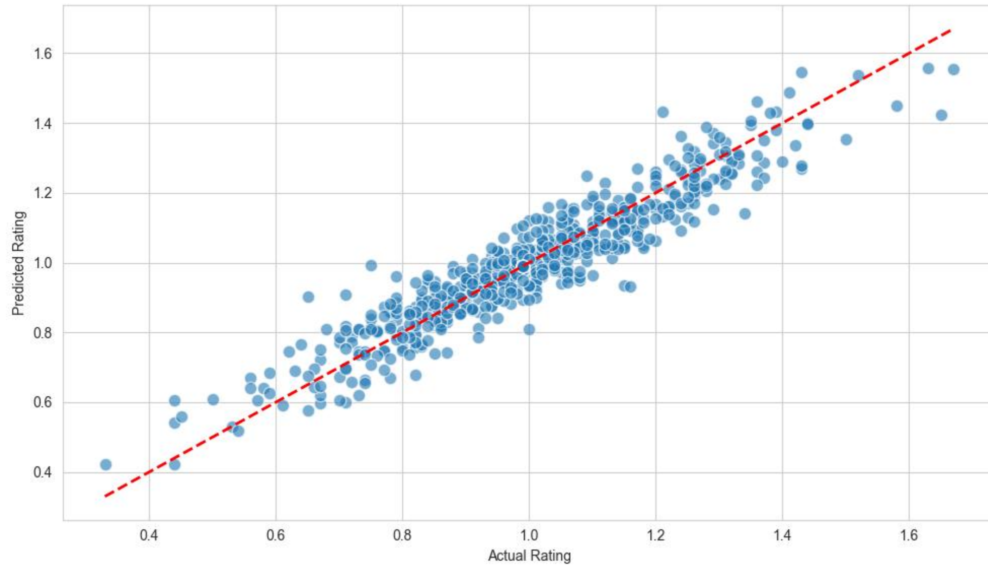

Figure 6 Actual vs. Predicted Player Rating

Feature Importance and Insights

Feature importance analysis revealed that certain variables played a much larger role in predicting player performance than others. The most important feature was Kills Per Round, which had an importance score of 0.5269, highlighting that the number of kills a player achieves per round is a strong predictor of their overall performance. Other influential features included Kill, Assist, Trade, Survive Ratio (0.2342), Average Damage Per Round (0.0549), and Clutches Played (0.0415), which indicate that factors such as player survival, engagement in clutch situations, and overall damage dealt are also significant in determining player success.

These findings suggest that players who consistently achieve higher kills per round, maintain strong damage output, and contribute in clutch scenarios are more likely to perform well in competitive Valorant matches. The feature importance plot further illustrates this, showing a clear distinction between the features that have the most impact and those that have minimal influence on the model's predictions. This analysis not only validates the relevance of certain player statistics for performance prediction but also offers valuable insights into which aspects of gameplay are most critical in competitive e-sports.

Figure 7 provides a bar plot of the importance scores of various features used in the Random Forest model for predicting player ratings. The most influential feature is Kills Per Round, with an importance score of 0.5269, indicating that the number of kills a player achieves per round has the largest impact on their overall performance prediction. Other significant features include Kill, Assist, Trade, Survive Ratio (0.2342) and Average Damage Per Round (0.0549). These features highlight that aspects such as a player's ability to engage in combat, deal damage, and contribute in multi-kill situations are critical to determining their rating. This analysis not only shows which features are most important for the model's predictions but also provides insights into the key factors that drive player performance in Valorant.

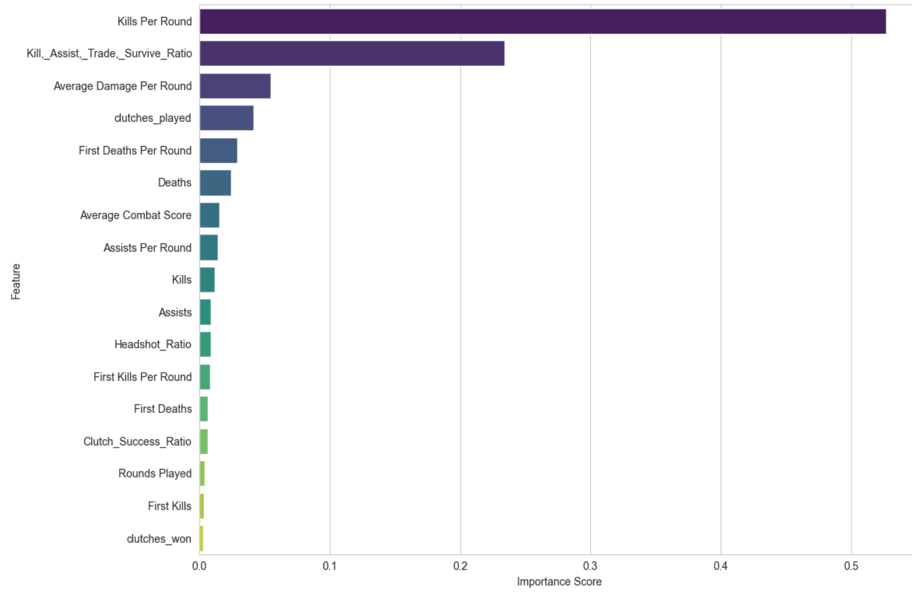

Figure 7 **Feature Importance from Random Forest**

In summary, the analysis demonstrates the effectiveness of the Random Forest Regressor in predicting player performance in the Valorant Champion Tour. By applying data mining techniques to the dataset, we were able to accurately predict player ratings based on performance metrics such as Kills Per Round, Clutch Success Ratio, and Average Damage Per Round. The model's evaluation metrics, particularly the R-squared value of 0.8831, suggest that the model can reliably predict player performance, offering valuable insights for e-sports teams looking to optimize their strategies and player selection. Additionally, the feature importance analysis sheds light on the key factors that drive player success, which could be used to inform training, team compositions, and in-game decisions.

Comparison with Previous Research

The integration of e-sports within the metaverse has been widely recognized as a transformative shift in the way competitive gaming is experienced and understood. Researchers such as [1], [2] highlight the metaverse's potential to offer immersive and interactive environments that blend virtual reality (VR), augmented reality (AR), and persistent digital worlds, reshaping the competitive landscape for e-sports. These studies argue that the metaverse enables a more integrated and engaging experience for players and fans alike, fostering a competitive environment that goes beyond traditional gaming platforms. In alignment with this, the current study on Valorant player performance prediction delves into the data-driven aspects of competitive e-sports, offering a predictive model to enhance player performance in a game that already exists within the metaverse. Unlike traditional e-sports research that primarily focuses on the cultural and technological impact of the metaverse, this paper introduces a quantitative approach, utilizing data mining techniques and the Random Forest algorithm to predict player performance based on match data. This integration of data-driven analysis into e-sports research provides a unique contribution by linking virtual world dynamics with predictive modeling, thereby aligning with the broader goals of the metaverse.

The role of Valorant in the metaverse has been explored extensively by [8], [9], [10], who discuss how the game's immersive design elements and strategic team play offer significant potential for integration into metaverse platforms. These scholars argue that Valorant's mechanics, which balance individual skills with coordinated team strategies, are key to its success in both traditional e-sports and evolving virtual ecosystems. This paper builds on those ideas by applying data mining to explore how specific player metrics, such as Kills Per Round, Clutch Success Ratio, and Average Combat Score (ACS), influence performance outcomes. The Random Forest model trained in this study to predict player ratings is based on these performance metrics and shows strong results, with an R-squared (R^2) value of 0.8831. This indicates that the model explains over 88% of the variance in player ratings, making it highly effective in predicting player performance. The evaluation of the model using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Squared Error (MSE) demonstrates that the model's predictions are accurate and reliable, providing a solid foundation for making data-driven decisions in team strategy and player selection.

Previous research on the metaverse and e-sports, such as [3], emphasizes the importance of virtual environments in enhancing competitive sports experiences. These studies highlight how such environments contribute to player engagement and mental performance. However, while these works focus on the broader immersive experience, this study adds a data mining perspective by examining specific performance factors that contribute to success in competitive e-sports. Feature importance analysis revealed that the most critical factors for predicting player performance include Kills Per Round, which had an importance score of 0.5269, and Kill, Assist, Trade, Survive Ratio (0.2342). This suggests that players who consistently achieve high kills per round, maintain strong damage output, and perform well in clutch situations are more likely to succeed in Valorant. These findings validate the relevance of certain gameplay aspects for performance prediction and offer actionable insights for gameplay optimization within the metaverse. The current study bridges the gap between the qualitative insights about the virtual experience in e-sports and quantitative performance modeling, contributing a new dimension to the research on how data-driven approaches can be used to enhance player performance in immersive digital ecosystems. This predictive modeling approach aligns with the goals of the metaverse by fostering an environment where competitive games like Valorant can evolve through data-backed optimization and strategic decision-making.

Conclusion

The predictive analysis of player performance in Valorant using the Random Forest Regressor yielded promising results, with the model explaining 88.31% of the variance in player ratings. Key features such as Kills Per Round, Kill, Assist, Trade, Survive Ratio, and Clutch Success Ratio were identified as the most influential factors in predicting a player's overall performance. The model was trained on a dataset consisting of over 4,000 rows of aggregated match data, and its evaluation metrics, including Mean Absolute Error (MAE) of 0.0400 and Root Mean Squared Error (RMSE) of 0.0583, indicated high accuracy and reliable predictions. The model's strong performance suggests that data-driven approaches can significantly enhance our understanding of player dynamics and be used for improving player strategies and team compositions in competitive e-sports.

The findings of this study have significant implications for the future of e-sports in the metaverse, where immersive virtual environments provide new opportunities for enhancing competitive gameplay. By utilizing data mining techniques, such as the Random Forest algorithm, to predict player performance, this research introduces a more analytical approach to virtual gaming. These predictive models can help e-sports teams optimize their strategies, select the best-performing players based on historical data, and make real-time decisions during matches. Furthermore, as the metaverse continues to evolve with more complex and interactive virtual worlds, integrating data-driven insights into these environments will provide a more personalized and strategic gaming experience. This could not only improve player engagement but also enhance the overall competitiveness and sustainability of e-sports within the metaverse, fostering a more dynamic and optimized virtual ecosystem.

While the study provides valuable insights into player performance prediction, it does have several limitations. One significant limitation is the missing data, particularly for specific performance metrics like Headshot Ratio and Clutch Success Ratio, which were incomplete for some players. This data sparsity could have affected the overall accuracy of the model. Additionally, the study used a Random Forest Regressor, which, while effective, has certain constraints in terms of capturing more complex relationships between features, particularly in highly dynamic environments like the metaverse. Future research could explore deep learning models, such as neural networks or recurrent neural networks (RNNs), which are better suited for handling complex, non-linear relationships in large datasets. Moreover, integrating real-time prediction systems into the metaverse could allow for more adaptive strategies and dynamic decision-making during live matches, creating a more immersive and responsive gaming environment. Further research could also focus on integrating player sentiment analysis and behavioral data to enhance the understanding of player motivations and interactions within the virtual gaming space.

Declarations

Author Contributions

Conceptualization, A.S.P. and J.J.; Methodology, A.S.P. and J.J.; Software, A.S.P. and J.J.; Validation, A.S.P. and J.J.; Formal Analysis, A.S.P.; Investigation, A.S.P. and J.J.; Resources, J.J.; Data Curation, A.S.P.; Writing—Original Draft Preparation, A.S.P.; Writing—Review and Editing, J.J.; Visualization, A.S.P. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the

corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S. Mystakidis, “Metaverse,” *Encyclopedia*, vol. 2, no. 1, pp. 486–497, 2022, doi: 10.3390/encyclopedia2010031.

[2] S. Malik and A. Rana, “Exploring current technologies, applications, and challenges in the metaverse: A comprehensive survey,” *J. Graph. Era Univ.*, vol. 2024, no. 1, pp. 1–12, 2024, doi: 10.13052/jgeu0975-1416.1217.

[3] J. Hamari and M. Sjöblom, “What is eSports and why do people watch it?,” *Internet Res.*, vol. 27, no. 2, pp. 211–232, 2017, doi: 10.1108/intr-04-2016-0085.

[4] K. A. Akgül, A. Erdoğan, and E. F. Şirin, “Examination of metaverse awareness levels of university students who have sports education,” *Turk. J. Sport Exerc.*, vol. 26, no. 2, pp. 265–274, 2024, doi: 10.15314/tsed.1448116.

[5] P. Matuszewski, P. Dobrowolski, and B. Zawadzki, “The association between personality traits and eSports performance,” *Front. Psychol.*, vol. 11, no. 6, pp. 1–12, 2020, doi: 10.3389/fpsyg.2020.01490.

[6] Z. Huang, D. Choi, B. Lai, Z. Lü, and H. Tian, “Metaverse-based virtual reality experience and endurance performance in sports economy: Mediating role of mental health and performance anxiety,” *Front. Public Health*, vol. 10, no. 11, pp. 1–12, 2022, doi: 10.3389/fpubh.2022.991489.

[7] N. Mehta and N. Joshi, “Metaverse is emerging!,” *Braz. J. Sci.*, vol. 2, no. 11, pp. 1–12, 2023, doi: 10.14295/bjs.v2i11.381.

[8] Y. Xu, “Applications of psychological principles in game design: Unveiling the dynamics and psychology of Valorant,” *Lect. Notes Educ. Psychol. Public Media*, vol. 2024, no. 6, pp. 1–12, 2024, doi: 10.54254/2753-7048/62/20241731.

[9] M. A. Putra and L. Manurizal, “Pre-service physical education teachers’ perceptions of eSports as a potential career pathway,” *J. Sport Area*, vol. 9, no. 2, pp. 1–12, 2024, doi: 10.25299/sportarea.2024.vol9(2).13895.

[10] P. Sahoo, A. Saurav, L. K. Vemali, and R. Gandla, “Understanding crucial factors of metaverse acceptance among gamers: Evidence from empirical studies,” *EATP*, vol. 30, no. 4, pp. 1–12, 2024, doi: 10.53555/kuey.v30i4.2008.

[11] S. Kelly and D. van der Leij, “A new frontier: Alcohol sponsorship activation through eSports,” *Mark. Intell. Plan.*, vol. 38, no. 5, pp. 603–616, 2020, doi: 10.1108/mip-03-2020-0101.

[12] M. Behnke, M. Kosakowski, and Ł. D. Kaczmarek, “Social challenge and threat predict performance and cardiovascular responses during competitive video gaming,” *Psychol. Sport Exerc.*, vol. 46, no. 1, pp. 1–12, 2020, doi: 10.1016/j.psychsport.2019.101584.

[13] K. Watanabe, N. Saijo, S. Minami, and M. Kashino, “The effects of competitive and interactive play on physiological state in professional eSports players,” *Heliyon*, vol. 7, no. 6, pp. 1–12, 2021, doi: 10.1016/j.heliyon.2021.e06844.

[14] T. Iwatsuki, G. Hagiwara, and M. E. Dugan, “Effectively optimizing eSports performance through movement science principles,” *Int. J. Sports Sci. Coach.*, vol. 16, no. 4, pp. 805–817, 2021, doi: 10.1177/17479541211016927.

[15] J. R. Dmello, “The impact of data analytics on player performance in professional sports: A systematic review,” *Int. Res. J. Mod. Eng. Technol. Sci.*, vol. 2023, no. 5, pp. 1–12, 2023, doi: 10.56726/irjmets37597.

[16] M. G. Trotter, T. J. Coulter, P. Davis, D. Poulus, and R. Polman, “The association between eSports participation, health and physical activity behaviour,” *Int. J. Environ. Res. Public Health*, vol. 17, no. 19, pp. 7329–7340, 2020, doi: 10.3390/ijerph17197329.

[17] A. Sousa, F. A. Silva, D. Moreira, and J. Santos, “Physiological and cognitive functions following a discrete session of competitive eSports gaming,” *Front. Psychol.*, vol. 11, no. 5, pp. 1–12, 2020, doi: 10.3389/fpsyg.2020.01030.

[18] C. Soto-Valero, “Predicting win-loss outcomes in MLB regular season games – A comparative study using data mining methods,” *Int. J. Comput. Sci. Sport*, vol. 15, no. 2, pp. 71–87, 2016, doi: 10.1515/ijcss-2016-0007.

[19] C. Wang, “Optimization of sports effect evaluation technology from random forest algorithm and elastic network algorithm,” *PLoS One*, vol. 18, no. 8, pp. 1–12, 2023, doi: 10.1371/journal.pone.0292557.

[20] W. J. Chen, M.-J. Jhou, T. Lee, and C.-J. Lu, “Hybrid basketball game outcome prediction model by integrating data mining methods for the National Basketball Association,” *Entropy*, vol. 23, no. 4, pp. 1–12, 2021, doi: 10.3390/e23040477.

[21] S. Lotfi and M. Rebbouj, “Machine learning for sport results prediction using algorithms,” *Int. J. Inf. Technol. Appl. Sci.*, vol. 3, no. 3, pp. 1–12, 2021, doi: 10.52502/ijitas.v3i3.114.

[22] K. U. Birant and D. Birant, “Multi-objective multi-instance learning: A new approach to machine learning for eSports,” *Entropy*, vol. 25, no. 1, pp. 1–12, 2022, doi: 10.3390/e25010028.

[23] T. D. Smithies, M. J. Campbell, N. Ramsbottom, and A. J. Toth, “A random forest approach to identify metrics that best predict match outcome and player ranking in the esport Rocket League,” *Sci. Rep.*, vol. 11, no. 1, pp. 1–12, 2021, doi: 10.1038/s41598-021-98879-9.

[24] V. E. Adeyemo, A. Palczewska, B. Jones, D. Weaving, and S. Whitehead, “Optimising classification in sport: A replication study using physical and technical-tactical performance indicators to classify competitive levels in rugby league match-play,” *Sci. Med. Footb.*, vol. 6, no. 3, pp. 251–259, 2022, doi: 10.1080/24733938.2022.2146177.

[25] S. Mohammad, V. N. Madishetti, Y. Nitin, B. P. Yadav, B. S. Sree, and B. M. Moses, “Survey on machine learning based game predictions,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 981, no. 2, pp. 1–12, 2020, doi: 10.1088/1757-899X/981/2/022052.