Analyzing Player Performance Metrics for Rank Prediction in Valorant Using Random Forest: A Data-Driven Approach to Skill Profiling in the Metaverse

Untung Rahardja[1], Qurotul Aini[2]

[1,2]Faculty of Science and Technology, University of Raharja, Tangerang 1511, Indonesia

ABSTRACT

This study explores the application of Random Forest, a powerful data mining technique, to predict player ranks in Valorant, a competitive first-person shooter. By analyzing a range of player performance metrics, including headshots, kills, damage received, and traded kills, the study identifies the key features that influence player rank determination. Using a dataset of player statistics, the model was trained to predict player ranks, achieving a prediction accuracy of 50.09%. The analysis revealed that headshots and traded kills were the most influential metrics in determining player rank, suggesting that skill-based metrics like accuracy and tactical gameplay are crucial for ranking in the game. These findings highlight the importance of understanding the relationship between various performance indicators and rank progression, offering valuable insights for both game developers and players. The results contribute to the growing body of research in gaming analytics, showcasing how data mining techniques can be used to analyze player behavior and improve competitive balance in games. The study underscores the potential of using data-driven approaches to enhance game design, providing developers with actionable insights to refine rank prediction systems, adjust in-game mechanics, and ensure a more balanced competitive environment. Looking ahead, future research can explore the use of alternative machine learning models, such as support vector machines (SVM), XGBoost, or neural networks, to improve the prediction accuracy and robustness of the model. Additionally, expanding the dataset to include more detailed player behaviors, match outcomes, and even temporal aspects of player performance could provide a more comprehensive understanding of the factors influencing player ranks. This can help further unravel the complexities of player behavior and performance in the metaverse, where virtual environments evolve dynamically based on player interactions.

Keywords Valorant, Random Forest, Rank Prediction, Gaming Analytics, Performance Metrics, Headshots, Traded Kills, Machine Learning, Metaverse, Data Mining

Submitted 5 January 2025
Accepted 10 April 2025
Published 1 June 2025

Corresponding author
Untung Rahardja,
untung@raharja.info

Additional Information and Declarations can be found on page 117

DOI: 10.47738/ijrm.v2i2.26

INTRODUCTION

In the burgeoning landscape of competitive gaming and the metaverse, we find ourselves on the precipice of a data revolution—one that is both dizzying in its scope and profound in its implications. This era of digital interconnectedness demands a recalibration of how we approach data analysis, as both competitive gaming and virtual worlds converge to offer unprecedented volumes of intricate data. Within this digital frontier, data is not merely collected but is intricately woven into the fabric of gameplay and user interaction, creating a kaleidoscope of possibilities for analysis.

How to cite this article: U. Rahardja and Q. Aini, "Analyzing Player Performance Metrics for Rank Prediction in Valorant Using Random Forest: A Data-Driven Approach to Skill Profiling in the Metaverse," *Int. J. Res. Metav.*, vol. 2, no. 2, pp. 102-120, 2025.

The domain of competitive gaming, colloquially known as eSports, epitomizes this data abundance with its rich tapestry of performance metrics, player interactions, and environmental variables. Such data streams, teeming with dynamic complexity, extend beyond simple records of scores and statistics, evolving into multidimensional constructs that capture the nuances of player behavior and in-game strategies. This environment, as elucidated by [1], calls for specialized machine learning techniques adept at navigating challenges like sensor anomalies and unbalanced datasets—underscoring the field's demand for methodological innovation.

Parallel to this is the rise of the metaverse, a boundless virtual ecosystem where data flows both voluminously and incessantly. Here, within these immersive realms, the interactions of avatars mirror real-world social and economic exchanges, generating vast datasets ripe for exploration. Scholars such as [2], emphasize the critical role of big data methodologies in synthesizing this information, elevating our understanding of virtual engagements and enabling new paradigms of insight. The datafication witnessed within the metaverse is not simply a trend; it represents a seismic shift where data's value amplifies through the connections it sparks—whether in shaping consumer experiences or redefining business strategies [3].

This confluence of competitive gaming and the virtual worlds of the metaverse beckons a new horizon in data analytics; one where AI, blockchain, and big data analytics converge to tackle the intricate dance of data management and interpretation [4], [5]. This synergistic interplay offers fertile ground for refining our grasp of player dynamics in eSports, while simultaneously innovating how enterprises engage with virtual consumers.

In the kaleidoscopic realm of competitive gaming and the metaverse, understanding player performance metrics emerges not merely as a prerequisite but as a catalyst for transformative change. These metrics function as a beacon, guiding developers and stakeholders alike through the intricate labyrinth of player dynamics and game mechanics. By harnessing these insights, the industry cultivates not only equitable gaming experiences but also engenders superior game design, laying the groundwork for a harmonious digital ecosystem.

Consider the notion of competitive balance—a cornerstone of any thriving multiplayer environment. Within the frenetic battlegrounds of Multiplayer Online Battle Arenas (MOBAs), equilibrium is more than a goal; it is an imperative. The exigencies of maintaining this balance demand unwavering vigilance, as emphasized by Huang and Bruda, who demonstrate that continuous recalibration based on player performance data is vital for sustaining a fair and engaging competition [6]. Such meticulous analysis empowers developers to address disparities in power dynamics—whether through rebalancing overpowering characters or refining dominant strategies—thereby ensuring that no competitor is left with an insurmountable advantage.

Beyond balancing acts, the interplay of player metrics in game design reveals itself with elegant complexity. Artificial Intelligence, wielded wisely, serves as both a lantern and a scalpel, illuminating potential pathways while dissecting intricate performance data. In this regard, Nalbant and Aydin reveal that AI not only evaluates player efficacy but also aids in crafting tactical innovations, offering coaches the precision tools needed to fine-tune player strategies [7]. Through this analytical lens, developers deliver adaptive gameplay experiences that mirror the evolving competencies of their audience, thus enhancing engagement and prolonging player involvement.

The confluence of performance metrics and strategic foresight extends its tendrils further into the corporate sphere. The monumental merger of Microsoft and Activision-Blizzard stands as a testament to the profound impact of data-driven insights, reshaping competitive dynamics and spurring industry innovation [8]. In this digital chessboard, companies leverage performance metrics to discern market trends and refine their offerings, solidifying their stature in the increasingly competitive gaming arena.

Moreover, the narrative spun by performance metrics weaves into the psychological tapestry of gaming culture. Games, akin to mirrors, reflect and influence player behaviors and mindsets, an assertion supported by studies like those of Etchells et al., which explore the ramifications of gaming on youth conduct and mental health [9]. Insights gleaned from these metrics can steer developers to craft environments that foster healthy competitiveness, mitigating the specter of toxic behaviors and cultivating a vibrant, supportive community.

The kaleidoscope of competitive gaming, with its intricate array of player interactions and performance metrics, presents a tantalizing challenge: can one predict a player's rank using these metrics through the lens of data mining? This study ventures into that very enigma, wielding the dual-edged sword of quantitative analysis and machine learning. By embarking on this analytical pilgrimage, we aspire not merely to forecast rank but to unravel the nuances of player dynamics—nuances that shape the competitive tapestry of gaming environments.

Historically, the application of data mining in sports analytics has been both groundbreaking and transformative. Kovalchik's exploration into sports prediction underscores the necessity of robust predictors, particularly for lesser-skilled players, positing that integrating psychological elements and fine-grained dynamics could elevate accuracy [10]. This narrative, while rooted in traditional sports, finds resonance in the digital arenas of gaming, where the variegated data streams offer fertile ground for deriving predictive insights.

Our ambition to navigate player rank prediction through data mining techniques is not confined to mundane statistical pursuits; it traverses into the realms of eSports, where the relevance of ranking systems extends its influence over match outcomes, as demonstrated by [11]. This empirical foundation solidifies the notion that video games, much like physical sports, harbor profound predictive potential within their coded veins—a potential (when unlocked) that offers strategic foresight and competitive advantage.

The narrative intensifies when one probes into the art of graph mining, adeptly utilized by Alobaidi et al., who illustrated its prowess in deciphering player strategies within real-time strategy games research [12]. Their deployment of classification and regression trees, bolstered by neural networks, exemplifies the sophisticated methodologies suitable for dissecting player performance. Such approaches dovetail with our study’s objective, spotlighting the predictive power latent in these analytical frameworks when applied to rank determination.

Yet, the predictive journey would be incomplete without acknowledging the critical role of behavior. The explorations of Dehpanah et al. into behavioral data integration yield compelling evidence; they reveal that embedding player conduct within predictive models enhances accuracy beyond conventional ratings [13]. This insight resonates deeply within our methodological core, championing a holistic approach that marries performance metrics with behavioral nuances.

In the intricate arena of competitive Valorant play, the quest to decode the determinants of player rank becomes an intriguing analytical endeavor. At the heart of this endeavor lies our thesis: by employing Random Forest, we can identify the key performance metrics that concretely determine player rank. This thesis is not an abstract proposition but a pragmatic assertion rooted in the proven capabilities of Random Forest, a machine learning technique renowned for its prowess in distilling clarity from the convolutions of complex datasets.

Random Forest, with its ensemble learning approach, has earned recognition across diverse analytical landscapes, including the realm of sports analytics. The illustrative work of Dijkhuis et al. demonstrates how such models can effectively forecast player performance in elite soccer, thus informing tactical decisions during pivotal game moments [14]. These findings illuminate the versatility of Random Forest in untangling performance metrics, a versatility that we harness to probe the stratagems of Valorant, seeking to unveil which metrics wield the most influence over player ranking.

Within the gaming sector, the adaptability of Random Forest has been further substantiated by Pengmatchaya's exploration within Dota 2, where the model adeptly discerned player skill by synthesizing myriad gameplay statistics [15]. This exemplifies how Random Forest's analytical might extends seamlessly into the domain of eSports, where the granular intricacies of performance metrics offer a rich tapestry for analysis, enabling nuanced predictions of player ranks in the strategic theaters of Valorant.

To further elevate our discourse, comparative analyses across a spectrum of machine learning models have consistently highlighted Random Forest's superior predictive accuracy. A comparative study examined gaming popularity trends and affirmed that Random Forest outperformed its algorithmic peers, underscoring its formidable capacity for predictive analytics [16]. This not only reinforces its aptness for our study but also underscores its potential to revolutionize our understanding of key performance metrics in Valorant's ranking paradigm.

The meticulous identification of performance indicators is pivotal to our thesis. Insights from Geurkink et al. underscore the criticality of isolating variables that significantly sway game outcomes—an imperative that echoes through our examination of Valorant player rankings [17]. By spotlighting the most pertinent metrics, Random Forest transcends mere statistical analysis, offering actionable insights that enhance player performance and fuel competitive balance, driving not just analysis but strategic evolution.

Literature Review

Data Mining in Gaming

As gaming evolves into an intricate tapestry of interactive experiences, data mining techniques have become the loom that weaves raw player data into rich insights, enhancing the gaming ecosystem. This literature review navigates through the vast expanse of gaming analytics, uncovering the multifaceted applications of data mining techniques that continue to redefine not only how games are played but also how they are conceived, balanced, and enhanced.

A cornerstone of these applications is the analysis of player behavior and performance, a domain where data mining has revealed its potent capability to distill clarity from chaos. The work of Alobaidi et al., through graph mining on real-time strategy games, illustrates this vividly; they uncover patterns that serve as a Rosetta Stone for distinguishing winning strategies from losing ones, thereby illuminating potential imbalances within the game mechanics that might otherwise remain obscured [12]. This insight into player strategy transcends mere observation, offering developers a lens through which to view potential enhancements to game design. Parallel to this, Li et al.'s research on elite tennis players exemplifies the power of data mining to classify rankings based on nuanced performance metrics, reinforcing the role of analytics in maintaining competitive equilibrium [18].

Yet, data mining's influence extends beyond performance metrics, delving into the realm of player engagement and preferences. Here, the ethical considerations spotlighted by Willson and Leaver become pertinent, as they explore the complexities of big data mining in social games like FarmVille. They caution that while data collection illuminates player psychology and improves user engagement, it is encumbered with the responsibility of maintaining player privacy [19]. This duality underlines the delicate balance between enhancing user experience and preserving user trust within the digital playground.

The pedagogical sphere of gaming isn't immune to the allure of data mining either. Game Learning Analytics (GLA) emerges as a key area of exploration, as noted by Alonso-Fernández et al., wherein the meticulous analysis of player interactions within educational games not only enhances educational outcomes but also customizes learning environments to better fit learner needs [20]. Through visual analytics tools, stakeholders can transform complex datasets into actionable insights, thereby tailoring educational content to individual learning paths—a testament to the transformative power of data-informed pedagogy.

Moreover, data mining's commercial implications are underscored in Mäntymäki et al.'s research, which elucidates how analytics can revolutionize business models for small and medium-sized game developers by optimizing player experiences and in-game purchases [21]. This commercial aspect highlights the potential of data-driven insights to not only elevate gameplay satisfaction but also to bolster financial sustainability in a competitive market.

Intriguingly, the confluence of data mining techniques is not confined to gaming alone. The realm of sports analytics, as explored by Hong and Deng, offers a parallel narrative where data-driven insights into tennis matches inform coaching strategies and player development [22]. Such cross-disciplinary fertilization between traditional sports and gaming serves to amplify the utility and adaptability of data mining methodologies across domains.

Performance Metrics and Ranking Systems

The intricate dance between player performance metrics and ranking systems captivates both scholars and practitioners, driving a deeper understanding of how skill is quantified and celebrated in competitive arenas. Recent years have witnessed a surge in studies dissecting this relationship, particularly as it pertains to the dual realms of sports and gaming. This synthesis navigates through such studies, illuminating the methodologies that capture performance and their ramifications on ranking structures.

A paradigmatic example emerges in the work of Oukil and Govindaluri, who innovate a framework synergizing Data Envelopment Analysis (DEA) with Ordered Weighted Averaging (OWA) to meticulously rank football players based on objective performance determinants [23]. This methodological tapestry underscores the pivotal role of quantifiable metrics, offering a lens that transcends traditional, subjective evaluations. By anchoring rankings in objective indicators, this framework minimizes bias, thus enhancing the credibility and robustness of ranking systems—a principle that resonates across both sporting and gaming landscapes.

Shifting to the court, Kovalchik's exploration into tennis rankings unveils a dynamic system harnessing a player's year-long win-loss record as a bellwether for their prowess [10]. This rolling weighted approach encapsulates recent performance, mirroring the player's current form and adapting in real-time—a necessity in the ever-evolving theatre of competitive eSports. Such adaptive, responsive ranking systems are the vanguard of ensuring that rankings reflect true on-field—or in-game—abilities.

The Australian Football League (AFL) Player Ranking, scrutinized by Sullivan et al., reveals telling insights into the disconnect between existing metrics and actual performance outcomes [24]. Their study ventures into the chasm between draft order and player efficacy, advocating for more layered, discerning metrics that encompass the multifarious dimensions of gameplay. This underscores a continual call for refinement, demanding metrics that can navigate and encapsulate the nuances of performance across various dimensions.

In the world of tennis, Li et al. traverse the career trajectories of athletes, uncovering how player maturation subtly, yet significantly, influences competitive performance and rank ascension [18]. This emphasizes the potency of weaving developmental considerations into ranking methodologies, arguing for a holistic approach that appreciates the athlete's journey from nascent talent to seasoned veteran.

Team-based games such as League of Legends offer another layer of complexity, as Horne's research delves into the interplay between personality traits and performance metrics [25]. This study unearths the psychological undercurrents that meander through ranking systems, suggesting that metrics capturing psychological and skill-based factors compose a more complete picture of player success.

Lastly, the investigative work of Morales et al. into the temporal dynamics of performance rankings unveils universal patterns that persist across varied ranking systems [26]. This reinforces the importance of temporal elements in performance analysis, as player capabilities ebb and flow over time, revealing the rhythm of improvement and regression that informs ranking methodologies.

Random Forest Algorithm

In the diverse and intricate domain of predictive analytics, the Random Forest algorithm stands as a beacon of methodological innovation and prowess, weaving simplicity with sophistication. This ensemble learning technique, renowned for its versatility in both classification and regression tasks, finds itself indispensable in fields as varied as gaming, where predictive precision is paramount. The algorithm's suitability for such tasks is underscored by its defining characteristics—traits that simultaneously enhance performance and bolster interpretability.

At its core, the Random Forest algorithm orchestrates an ensemble of decision trees, each birthed through a process of bagging and randomization. It distills the collective wisdom of these trees into definitive outputs: the mode for classification tasks or the mean for regression. This ensemble approach inherently mitigates overfitting—a notorious pitfall in machine learning—by averaging out the peculiarities of individual trees, thus fostering robustness in high-dimensional spaces. Bennett et al. illuminate this trait, emphasizing the stability of predictor rankings as a testament to the algorithm's resilience [27]. In the unpredictable landscapes of gaming analytics, where player performance data often oscillates wildly, such robustness becomes a cornerstone of reliability.

The algorithm's prowess is exemplified in its applications within the gaming ecosystem, as demonstrated by Smithies et al. in their exploration of Rocket League. Here, Random Forest efficiently dissected player performance metrics to unravel the complex tapestry of factors influencing match outcomes and player rankings [28]. This example underscores the algorithm’s capability to transform chaotic performance data into coherent and actionable insights, affirming its status as an ideal instrument for predicting player ranks within multifaceted gaming environments.

One of Random Forest's lauded advantages is its facility with large, feature-rich datasets—a common scenario in gaming, where metrics range from raw gameplay actions to nuanced behavioral patterns. Zhou and Qiu extol the algorithm's adaptability across diverse domains, from computer vision to medical imaging, crediting its flexibility and robustness [29]. In gaming analytics, this versatility allows researchers to grapple with complex datasets, extracting nuanced insights that inform both strategic gameplay and game design.

The feature importance measurements intrinsic to Random Forest further amplify its interpretability, allowing researchers to distill which metrics most significantly sway outcomes. This insight is invaluable in clarifying the drivers behind player performance, ultimately shaping more informed game design and strategic decision-making processes. Although Alobaidi et al. provide a broader context of data mining in gaming rather than specifics on Random Forest, the inclusion of other supportive studies underscores the algorithm's definitive utility in gaming analytics [12].

Random Forest’s efficacy is not static; it evolves with methodological synergies such as feature selection. Sá et al. illustrate how coupling Random Forest with label ranking elevates feature selection, refining the model's capacity to capture subtle performance nuances [30]. This fusion of techniques ensures that models are not only accurate but finely attuned to the rhythms of player dynamics, enhancing predictive accuracy.

Relevant Formula

In the intricate domain of data analytics, the Random Forest algorithm emerges not only as a predictive powerhouse but also as a nuanced evaluator of feature significance. Central to this evaluative prowess is the equation that calculates feature importance within the Random Forest framework:

$$\text{Importance}(X_i) = \sum_{t \in RF} I_t(X_i)$$

Here, $I_t(X_i)$ symbolizes the importance of feature X_i within a specific decision tree t of the Random Forest ensemble. This mathematical articulation aggregates the importance scores across the multitude of trees, delivering a holistic measure of each feature's contribution to the model's predictive efficacy. It serves as a beacon, illuminating which features wield the greatest influence, thus guiding analysts in deciphering complex datasets.

The significance of this feature importance calculation is underscored extensively in the scholarly literature, revealing its reputation as a robust alternative to linear models and decision trees. Guo et al. delve into how Random Forest surmounts the limitations endemic to these models, adeptly navigating complex dependencies and providing stable, reliable feature rankings research [31]. This capability is indispensable for practitioners seeking to comprehend the underlying architecture of their datasets, offering clarity in a field where obscurity often reigns.

Moreover, Liu and Zhao expound on the operational utility of this feature ranking, highlighting how it refines the analytical process by affording the possibility to eliminate superfluous features [32]. Through iterative refinement and focus on potent variables, the Random Forest algorithm streamlines the predictive endeavor, refining performance and bolstering interpretative clarity.

In an intriguing pivot, Sun et al. introduce an innovative methodology by leveraging the Banzhaf power index within the Random Forest context, thus framing feature importance as a cooperative game theory problem [33]. This novel perspective deepens the understanding of feature interactions, going beyond conventional metrics to offer a rich, multidimensional appreciation of feature dynamics.

Compounding this complexity, the technique known as Random Forest Importance (RFI), discussed by Mustapha, quantifies feature significance through reduction in average impurity [34]. By measuring how the removal of a feature impacts the impurity reduction, this method adheres to the foundational tenet of Random Forest, underscoring efforts to minimize impurity while concurrently enhancing predictive performance.

Method

Data Collection

The dataset used in this study contains player performance metrics from the game Valorant, which were utilized to predict player rank (tier). The dataset comprises 2,694 rows and 10 columns, including both numeric and categorical data. The key features in the dataset include assists, damage_received, headshots, tier, traded, kills, matches, deaths, and damage. The assists and damage_received columns initially contained commas, which were removed, and the values were converted to numeric types for further analysis. The tier column, which indicates the player's rank (e.g., "iron," "bronze"), was encoded into numeric values to facilitate its use in machine learning models.

Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) was performed to understand the distribution of the data and to uncover any patterns or relationships between the variables. Initially, we checked for missing values, ensuring that the dataset was complete. Basic statistical summaries were generated using the `describe()` function, revealing the central tendencies and spread of the numeric features. For instance, the kills column had a mean of 6,685 and a standard deviation of 1,431, suggesting significant variation in player performance. This variation is crucial for distinguishing players of different ranks.

Data Visualization

The dataset was then visualized using histograms and scatter plots. The histograms helped identify the distribution of key performance metrics like headshots, kills, and damage, while scatter plots illustrated relationships between variables such as kills and damage, with attributes like headshots and matches incorporated as size and color attributes. These visualizations provided valuable insights into the patterns within the data and guided the feature selection process. Figure 1 provide insights into the distribution of several key player performance metrics from the dataset.

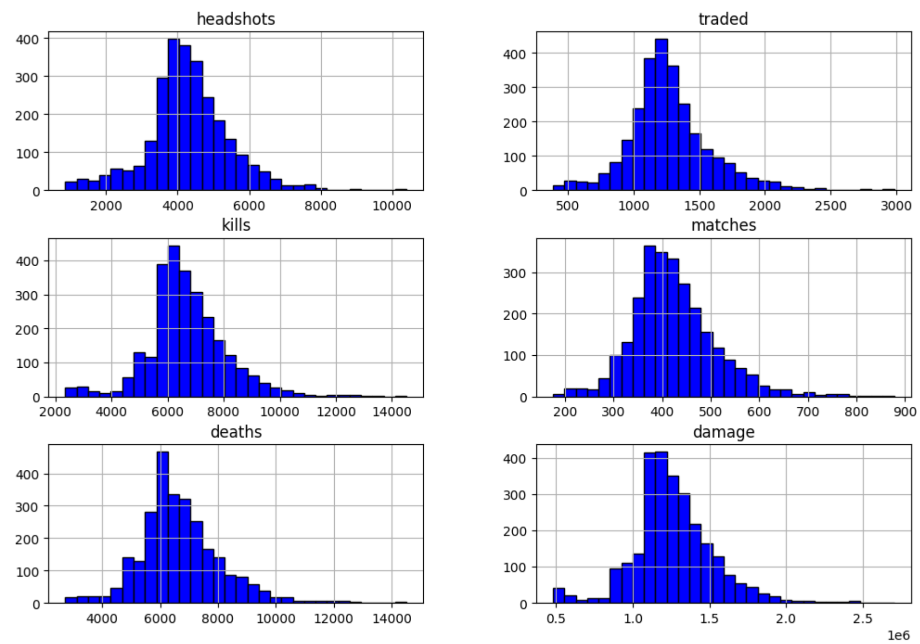

Figure 1 Histogram of Key Metrics

The headshots histogram displays a relatively normal distribution, with most players having headshot counts clustered around 4,000, with fewer players achieving very high or very low counts. This suggests that, for the majority of players, headshot performance is within a certain range, indicating a moderate level of skill. The traded kills metric shows a similar distribution, with most values concentrated around 1,000 and a long tail to the right, meaning a few players have much higher traded kill counts. This reflects the tactical nature of the game, where most players perform moderately, but some engage in more strategic trades. The kills histogram shows a distribution that is skewed to the right, with most players scoring between 5,000 and 7,000 kills. This indicates that most players are average in terms of kills, but fewer players achieve very high kill counts, highlighting the top performers.

The matches histogram reveals a distribution with most players having between 350 and 450 matches, with a peak around 400. This suggests that player experience, as measured by the number of matches played, is fairly uniform for the majority of the players. The deaths histogram has a distribution similar to that of kills, indicating that the number of deaths for most players falls within a moderate range, with fewer players experiencing very high or low deaths. The damage histogram indicates a slightly skewed distribution, with most players dealing between 1 million and 1.4 million damage, but with a significant number of players having lower or higher values, showing variation in damage output across the dataset. These visualizations offer a glimpse into player performance distribution and suggest areas of focus for improving competitive balance, such as adjusting for players who perform significantly above or below average in specific metrics.

Model Building

For the task of model building, our primary objective was to predict the player rank, more specifically the player's tier, based on a variety of performance metrics collected during gameplay. We chose to employ a Random Forest Classifier for this undertaking, owing to its robustness and effectiveness in handling classification problems, particularly when working with datasets that may include non-linear relationships between inputs and outputs.

Prior to training the model, we performed data preprocessing, which included encoding categorical variables to numerical values. The tier column, which was the target variable in our dataset, was initially a categorical variable. To make it suitable for use in machine learning algorithms, it was necessary to convert these categories into a numerical format. This was achieved through the process of label encoding, which assigned a unique integer to each category.

In terms of feature selection, a set of performance metrics were identified as predictors for the model. These features included Assists, which measure the number of times a player assisted in eliminating an opponent, and Damage_received, which accounts for the total amount of damage the player received during the matches. Another key feature is Headshots, denoting the count of successful headshots landed by the player. The feature Traded captures instances where the player’s kill was immediately followed by their own death, indicating a trade situation. Kills represent the total number of kills achieved by the player, while Matches track the number of matches the player participated in. Deaths document the total number of deaths the player experienced, and Damage reflects the cumulative damage dealt by the player throughout the matches. These carefully selected features served as the input variables (independent variables), while the tier served as the target variable (dependent variable) that the model aimed to predict.

For model evaluation and validation, the dataset was split into two sets: training and testing sets, using an 80/20 split ratio. This means 80% of the data was allocated for training the model, giving it ample data to learn from, while the remaining 20% was reserved for testing. This division ensured that the model's performance could be evaluated on unseen data, providing a reliable measure of its predictive capabilities. The Random Forest Classifier, initialized with 100 trees (n_estimators=100), was meticulously trained using the training data. The choice of 100 trees allowed the model to balance between computational efficiency and prediction accuracy. Each tree in the forest was constructed using a different subset of the training data, and the model's final prediction was determined by aggregating the predictions of all these individual trees, typically through majority voting.

Evaluation Metrics

To thoroughly evaluate the model's performance across different dimensions, a comprehensive set of evaluation metrics was employed, including accuracy, precision, and recall. These metrics are fundamental to understanding not only the model's general effectiveness but also its capability to make reliable predictions. Accuracy is perhaps the most straightforward metric, measuring the overall proportion of correct predictions made by the model. It provides a broad overview of the model's performance by indicating how often the model was right in its predictions. High accuracy is often desirable, but it alone does not paint a complete picture, as it doesn't account for the distribution of classes or the importance of different types of errors.

Precision offers a more nuanced view by calculating the proportion of true positive predictions out of all positive predictions made. This metric is crucial when the cost of false positives is high. For instance, in a setting where predicting a player as higher-ranked is consequential (perhaps involving promotion to a competitive league), precision becomes vital. High precision indicates that when the model predicts a high rank, it is often correct, thus reflecting the reliability and trustworthiness of those predictions. Recall, alternatively known as sensitivity, evaluates the model's ability to identify all relevant instances within a dataset. It measures the proportion of true positive predictions out of all actual positive cases. Recall is especially important in scenarios where missing a true positive has significant repercussions. For example, if the model fails to identify a top-performing player, crucial opportunities might be missed, such as offering advanced training or career advancement. Hence, recall provides insights into the model’s thoroughness in capturing all instances of interest.

Result and Discussion

Model Performance

The Random Forest model achieved a prediction accuracy of 50.09% on the test dataset, as shown in table 1. This indicates that the model's ability to correctly predict the player rank (tier) was slightly better than random chance, given the multiple classes involved. The overall precision of the model was 0.5359, meaning that, on average, the model's positive predictions were 53.59% accurate. The recall was 0.5009, which reflects the model's ability to identify 50.09% of all true positive instances of the various ranks in the test set.

Table 1. Model Performance Results

Metric	Score
Model Accuracy	0.5009
Model Precision	0.5359
Model Recall	0.5009

The detailed classification report provided further insights into how the model performed for each rank, shown in table 2. The precision and recall for each rank varied significantly, with some ranks performing better than others. For instance, the model achieved a high precision of 0.89 and recall of 0.75 for the rank corresponding to label ‘1’, suggesting that players in this tier were easier to predict accurately. However, ranks like ‘0’ and ‘2’ showed weaker performance, with precision and recall values below 0.40, indicating that the model struggled to predict these ranks accurately.

Table 2. Model Performance for Each Class

Class	Precision	Recall	F1-Score	Support
0	0.36	0.45	0.4	98
1	0.89	0.75	0.81	52
2	0.4	0.36	0.38	85
3	0.36	0.5	0.42	64
4	0.82	0.62	0.71	113
5	1.0	0.89	0.94	19
6	0.34	0.39	0.37	56
7	0.37	0.29	0.32	52

The macro average precision of 0.57 and recall of 0.53 provides a more general view of the model's performance across all ranks, while the weighted average precision and recall of 0.54 and 0.50, respectively, reflect the model's tendency to perform slightly better on more frequent ranks in the dataset.

The analysis of feature importance revealed several key metrics that played a significant role in predicting player rank. The most important feature was headshots, with an importance score of 0.1720, highlighting the relevance of accuracy in shooting for determining rank. This suggests that players with higher headshot accuracy tend to rank higher in the game. The second most influential feature was traded (with an importance of 0.1470), indicating that players who excel in trading kills—helping teammates by eliminating enemies after being killed themselves—also tend to perform better in the rankings.

Table 3. Feature Importance Results

Feature	Importance
headshots	0.172029
traded	0.14697
damage_received	0.141419
deaths	0.119048
damage	0.118127
matches	0.107254
kills	0.101831
assists	0.093321

Other notable features included damage_received (importance of 0.1414) and deaths (importance of 0.1190), which are critical factors in determining player rank. While damage_received indicates how well players can manage their health during a match, deaths serves as a negative indicator, with fewer deaths often correlating with higher skill levels. Damage (importance of 0.1181) also emerged as an important predictor, reflecting that players who deal more damage are generally ranked higher. Matches (importance of 0.1073) and kills (importance of 0.1018) were also found to contribute significantly, suggesting that experience (as measured by the number of matches played) and the ability to secure kills both influence rank prediction. Finally, assists (importance of 0.0933) appeared to have the least impact on the model's predictions, although it still played a role in helping identify skilled players who contribute to their team beyond just kills.

The feature importance results underscore that performance metrics such as headshots, traded kills, damage dealt, and deaths are crucial in determining player rank. These findings can inform game developers about the factors that players might need to focus on to improve their performance and rankings in competitive play. Furthermore, these results suggest potential avenues for improving the model, such as incorporating additional features, adjusting model parameters, or exploring different algorithms.

Visualizations

Figure 2 illustrates the relationship between kills and damage. The size of each point is proportional to the number of matches played by the player, while the color of the points represents the number of headshots.

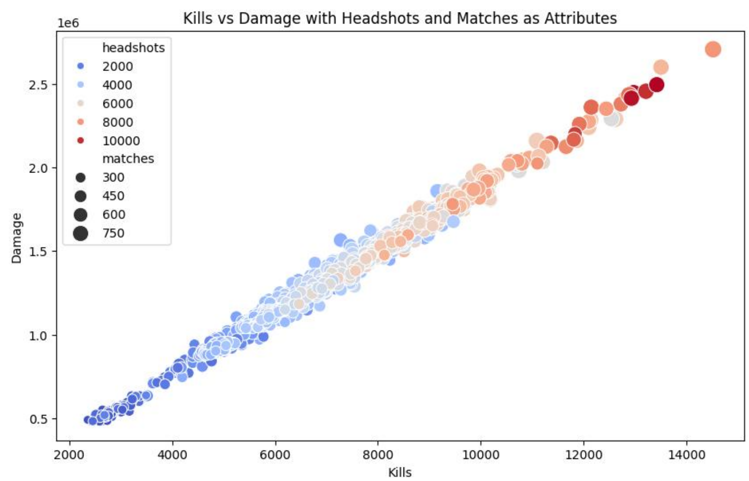

Figure 2 Kill vs Damage Scatter Plot

From figure 2, we can observe a clear positive correlation between kills and damage, indicating that players who deal more damage generally have more kills. The color gradient shows that players with more headshots (shown by the darker shades of red) also tend to have higher kills and damage, which is consistent with the expectation that accuracy (headshots) contributes to better performance. The variation in point size (representing matches) highlights that more experienced players, who have participated in a larger number of matches, tend to achieve higher kills and damage, reinforcing the importance of experience in gameplay.

Figure 3 show the feature importance scores derived from the Random Forest model. The chart ranks the most influential features in predicting player rank (tier).

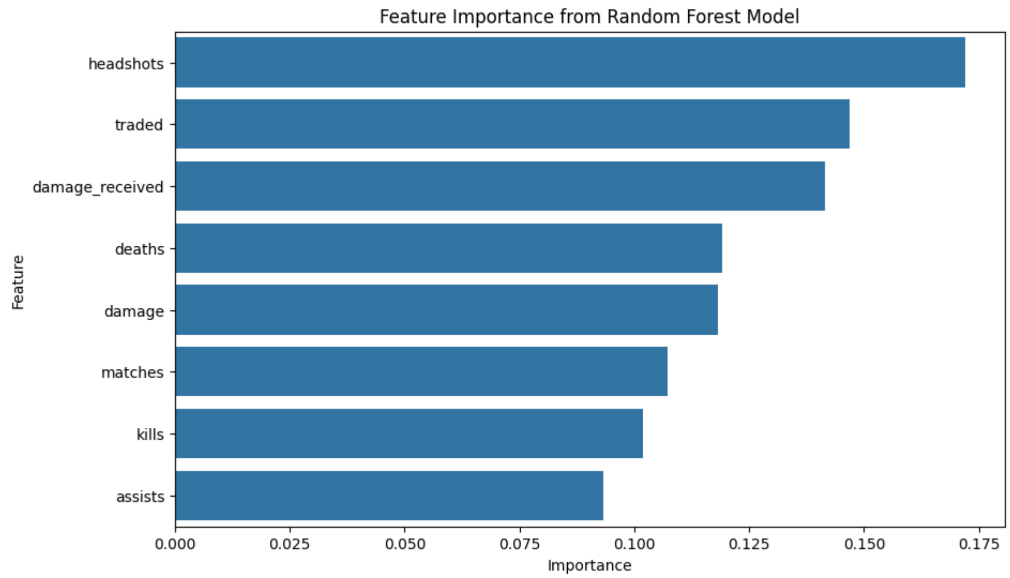

Figure 3 Feature Importance from Random Forest Model

The feature headshots stands out as the most important predictor, with the highest importance score of 0.172. This suggests that headshots are a crucial factor in determining a player's skill level, likely because players who can land more headshots tend to perform better in combat, thereby achieving higher ranks.Other significant features include traded (0.147), damage_received (0.141), and deaths (0.119). These features are closely related to player survival, combat efficiency, and overall impact in matches. Damage (0.118) and matches (0.107) also play important roles, but their contributions are slightly lower compared to features like headshots and traded. Kills (0.101) and assists (0.093) were identified as less influential in rank prediction, but they still contribute to the model's understanding of player performance.

Comparison with Literature

The findings from this study align with several key insights from existing literature on gaming analytics, particularly in the context of performance metrics and ranking systems. As demonstrated in the work of Alobaidi et al. [12], data mining techniques, including the Random Forest algorithm, are effective in analyzing player performance and uncovering patterns that enhance competitive balance and game design. In a similar vein, our study utilized Random Forest to predict player rank in Valorant and identified several key performance metrics such as headshots, traded kills, damage received, and kills as significant predictors of player rank. This finding mirrors the work of Smithies et al. [28] in Rocket League, where Random Forest successfully revealed the complex relationships between performance metrics and match outcomes. Both studies highlight the ability of data mining to transform raw gameplay data into actionable insights for improving game design and balancing competitive play.

Furthermore, our results on headshots being the most important predictor of rank are consistent with findings from other studies. For example, Kovalchik's exploration of tennis rankings [10] revealed the significance of a player's form and performance in determining their rank, akin to the way headshots indicate a player's skill and precision in Valorant. This reinforces the importance of skill-based metrics, such as accuracy, in determining player rank across various competitive domains.

In terms of feature importance, our results closely align with the research by Bennett et al. [27], who emphasized the robustness and interpretability of Random Forest's feature ranking. In our case, traded kills and damage received emerged as crucial features, illustrating how tactical gameplay and damage management influence rank prediction. These findings support the assertion by Zhou and Qiu [29] that Random Forest excels in handling complex, high-dimensional datasets, making it ideal for analyzing diverse player performance metrics.

While our study primarily focused on skill-based performance metrics, it also touched upon broader concepts, such as the experience factor, represented by matches played. This aligns with Li et al.'s research [18], which found that player maturation and experience contribute significantly to ranking in sports, emphasizing the importance of considering both skill and experience when designing ranking systems.

Implications for Gaming

The insights gained from this study can have several implications for both game design and competitive strategies in Valorant and other similar games. By identifying headshots and traded kills as the most significant predictors of rank, game developers can focus on refining these aspects in gameplay mechanics to enhance skill development and competitive balance. For instance, improving the feedback mechanisms for players' accuracy and tactical plays, such as trading kills, can motivate players to focus on these critical aspects of gameplay.

Incorporating these findings into game design can also help address potential imbalances in competitive play. As Alobaidi et al. [12] pointed out, identifying patterns of successful gameplay allows developers to adjust game mechanics that may inadvertently favor certain player strategies over others. Our study's emphasis on damage received and deaths could inspire game developers to fine-tune the game's damage mechanics, ensuring that high skill levels are consistently rewarded while preventing overpowered strategies or imbalances in gameplay.

Additionally, the insights into player experience gained through metrics like matches played suggest that ranking systems could be adapted to account for a player's progression over time. This could help create more dynamic and fair ranking systems that consider not just raw performance but also the journey of players as they improve through experience, much like the adaptive ranking systems discussed by Kovalchik [10]. By including such considerations, game developers could create more nuanced and realistic ranking structures that better reflect player skill and potential.

From a competitive strategy standpoint, players could use these insights to focus their training on the most impactful performance metrics, such as improving headshot accuracy and traded kills, to elevate their rank. Understanding the factors that influence rank prediction can provide players with a clear path for improvement, enabling them to refine their strategies and gameplay techniques to climb the ranks more efficiently.

Conclusion

This study utilized a Random Forest algorithm to analyze performance metrics in Valorant and predict player rank. The key findings from the analysis revealed that headshots, traded kills, damage received, and kills were the most influential factors in determining player rank. The model achieved a prediction accuracy of 50.09%, which indicates that while the model performed better than random chance, there is room for improvement in accuracy. These findings are consistent with existing literature on performance metrics and ranking systems in gaming, highlighting the importance of skill-based metrics like accuracy and damage management. The study also demonstrated the effectiveness of Random Forest in handling complex datasets with multiple features, providing valuable insights into player behavior and performance.

While this study focused on Valorant and used Random Forest for rank prediction, future research could explore the use of other data mining techniques such as support vector machines (SVM), XGBoost, or neural networks to improve prediction accuracy and model robustness. Additionally, future studies could expand the dataset by including more features, such as player behavior, match outcomes, or in-game decision-making, to provide a more holistic view of factors affecting player rank. Another promising direction would be to explore the integration of temporal dynamics in player performance, analyzing how performance evolves over time and influences rank progression, similar to the adaptive ranking systems used in traditional sports.

The application of predictive analytics, particularly in the realm of gaming, offers tremendous potential to enhance both player engagement and game development. By utilizing techniques like Random Forest, game developers can gain valuable insights into player behavior and skill progression, leading to more balanced and enjoyable competitive experiences. Moreover, as games increasingly become part of the metaverse, where players interact in expansive virtual worlds, these data-driven insights can help create dynamic and immersive environments that adapt to individual player preferences and behaviors. Ultimately, predictive analytics has the power to transform not only how players engage with games but also how games are designed and balanced in the metaverse.

Declarations

Author Contributions

Conceptualization: U.R., Q.A.; Methodology: Q.A.; Software: U.R.; Validation: U.R., Q.A.; Formal Analysis: U.R., Q.A.; Investigation: U.R.; Resources: Q.A.; Data Curation: Q.A.; Writing—Original Draft Preparation: U.R., Q.A.; Writing—Review and Editing: Q.A., U.R.; Visualization: U.R. 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] A. Korotin, N. Khromov, A. Stepanov, A. Lange, E. Burnaev, and A. Somov, "Towards Understanding of eSports Athletes' Potentialities: The Sensing System for Data Collection and Analysis," 2019, doi: 10.1109/smartworld-uic-atc-scalcom-iop-sci.2019.00319.

[2] T. Huynh-The et al., "Blockchain for the Metaverse: A Review," Future Generation Computer Systems, vol. 143, pp. 401–419, 2023, doi: 10.1016/j.future.2023.02.008.

[3] S. Kaddoura and F. A. Husseiny, "The Rising Trend of Metaverse in Education: Challenges, Opportunities, and Ethical Considerations," Peerj Computer Science, vol. 9, p. e1252, 2023, doi: 10.7717/peerj-cs.1252.

[4] K. M. Hou, X. Diao, H. Shi, H. Ding, H. Zhou, and C. d. Vaulx, "Trends and Challenges in AIoT/IIoT/IoT Implementation," Sensors, vol. 23, no. 11, p. 5074, 2023, doi: 10.3390/s23115074.

[5] V. Ahsani, A. Rahimi, M. Letafati, and B. H. Khalaj, "Unlocking Metaverse-as-a-Service the Three Pillars to Watch: Privacy and Security, Edge Computing, and Blockchain," 2023, doi: 10.48550/arxiv.2301.01221.

[6] C. Huang and S. D. Bruda, "Improved Balance in Multiplayer Online Battle Arena Games," Acta Universitatis Sapientiae Informatica, vol. 12, no. 2, pp. 183–204, 2020, doi: 10.2478/ausi-2020-0011.

[7] K. G. Nalbant and S. AYDIN, "Literature Review on the Relationship Between Artificial Intelligence Technologies With Digital Sports Marketing and Sports Management," Indonesian Journal of Sport Management, vol. 2, no. 2, pp. 135–143, 2022, doi: 10.31949/ijsm.v2i2.2876.

[8] Y. Ying, "The Impact of the Microsoft and Activision-Blizzard Merger on the Gaming Industry: A Theoretical and Analytical Examination," Advances in Economics Management and Political Sciences, vol. 19, no. 1, pp. 372–377, 2023, doi: 10.54254/2754-1169/19/20230162.

[9] P. Etchells, S. H. Gage, A. Rutherford, and M. R. Munafò, "Prospective Investigation of Video Game Use in Children and Subsequent Conduct Disorder and Depression Using Data From the Avon Longitudinal Study of Parents and Children," Plos One, vol. 11, no. 1, p. e0147732, 2016, doi: 10.1371/journal.pone.0147732.

[10] S. Kovalchik, "Searching for the GOAT of Tennis Win Prediction," Journal of Quantitative Analysis in Sports, vol. 12, no. 3, 2016, doi: 10.1515/jqas-2015-0059.

[11] A. Fayomi, R. Majeed, A. Algarni, S. Akhtar, F. Jamal, and J. A. Nasir, "Forecasting Tennis Match Results Using the Bradley-Terry Model," International Journal of Photoenergy, vol. 2022, pp. 1–12, 2022, doi: 10.1155/2022/1898132.

[12] I. A. Alobaidi, J. L. Leopold, A. A. Allami, N. W. Eloe, and D. Tanksley, "Predictive Analysis of real-time Strategy Games: A Graph Mining Approach," Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery, vol. 11, no. 2, 2020, doi: 10.1002/widm.1398.

[13] A. Dehpanah, M. F. Ghori, J. Gemmell, and B. Mobasher, "Behavioral Player Rating in Competitive Online Shooter Games," 2022, doi: 10.48550/arxiv.2207.00528.

[14] T. Dijkhuis, M. Kempe, and K. Lemmink, "Early Prediction of Physical Performance in Elite Soccer Matches—A Machine Learning Approach to Support Substitutions," Entropy, vol. 23, no. 8, p. 952, 2021, doi: 10.3390/e23080952.

[15] M. Pengmatchaya, "Identifying Player Skill of Dota 2 Using Machine Learning Pipeline," Discover Artificial Intelligence, vol. 4, no. 1, 2024, doi: 10.1007/s44163-024-00139-y.

[16] K. V. V. -, "Machine Learning Models in Predicting Gaming Popularity: A Comparative Analysis," International Journal for Multidisciplinary Research, vol. 6, no. 3, 2024, doi: 10.36948/ijfmr.2024.v06i03.14362.

[17] Y. Geurkink, J. Boone, S. Verstockt, and J. Bourgois, "Machine Learning-Based Identification of the Strongest Predictive Variables of Winning and Losing in Belgian Professional Soccer," Applied Sciences, vol. 11, no. 5, p. 2378, 2021, doi: 10.3390/app11052378.

[18] P. Li, J. R. Weissensteiner, J. Pion, and V. D. Bosscher, "Predicting Elite Success: Evidence Comparing the Career Pathways of Top 10 to 300 Professional Tennis Players," International Journal of Sports Science & Coaching, vol. 15, no. 5–6, pp. 793–802, 2020, doi: 10.1177/1747954120935828.

[19] M. Willson and T. Leaver, "Zynga's FarmVille, Social Games, and the Ethics of Big Data Mining," Communication Research and Practice, vol. 1, no. 2, pp. 147–158, 2015, doi: 10.1080/22041451.2015.1048039.

[20] C. Alonso-Fernández, A. Calvo-Morata, I. Martínez-Ortiz, and B. Fernández-Manjón, "Game Learning Analytics:," Journal of Learning Analytics, vol. 9, no. 3, pp. 32–49, 2022, doi: 10.18608/jla.2022.7633.

[21] M. Mäntymäki, S. Hyrynsalmi, and A. Koskenvoima, "How Do Small and Medium-Sized Game Companies Use Analytics? An Attention-Based View of Game Analytics," Information Systems Frontiers, vol. 22, no. 5, pp. 1163–1178, 2019, doi: 10.1007/s10796-019-09913-1.

[22] H. Hong and R. Deng, "Analysis Technology of Tennis Sports Match Based on Data Mining and Image Feature Retrieval," Complexity, vol. 2020, pp. 1–15, 2020, doi: 10.1155/2020/8877161.

[23] A. Oukil and S. M. Govindaluri, "A Systematic Approach for Ranking Football Players Within an Integrated DEA-OWA Framework," Managerial and Decision Economics, vol. 38, no. 8, pp. 1125–1136, 2017, doi: 10.1002/mde.2851.

[24] C. Sullivan, T. Kempton, P. Ward, and A. J. Coutts, "The Efficacy of Talent Selection Criteria in the Australian Football League," Journal of Sports Sciences, vol. 38, no. 7, pp. 773–779, 2020, doi: 10.1080/02640414.2020.1734309.

[25] G. Horne, "Narcissism, Big Five Personality Traits, and Performance in Temporary Teams: An Investigation of League of Legends," Journal of Electronic Gaming and Esports, vol. 1, no. 1, 2023, doi: 10.1123/jege.2022-0022.

[26] J. A. Morales et al., "Generic Temporal Features of Performance Rankings in Sports and Games," Epj Data Science, vol. 5, no. 1, 2016, doi: 10.1140/epjds/s13688-016-0096-y.

[27] M. A. Bennett, N. E. Bezodis, D. A. Shearer, D. Locke, and L. P. Kilduff, "Descriptive Conversion of Performance Indicators in Rugby Union," Journal of Science and Medicine in Sport, vol. 22, no. 3, pp. 330–334, 2019, doi: 10.1016/j.jsams.2018.08.008.

[28] 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," Scientific Reports, vol. 11, no. 1, 2021,

doi: 10.1038/s41598-021-98879-9.
[29] Y. Zhou and G. Qiu, “Random Forest for Label Ranking,” Expert Systems With Applications, vol. 112, pp. 99–109, 2018, doi: 10.1016/j.eswa.2018.06.036.
[30] C. R. d. Sá, C. Soares, A. Knobbe, and P. Cortez, “Label Ranking Forests,” Expert Systems, vol. 34, no. 1, 2016, doi: 10.1111/exsy.12166.
[31] X. Guo, C. Zhang, W. Luo, J. Yang, and M. Yang, “Urban Impervious Surface Extraction Based on Multi-Features and Random Forest,” Ieee Access, vol. 8, pp. 226609–226623, 2020, doi: 10.1109/access.2020.3046261.
[32] Y. Liu and H. Zhao, “Variable Importance-weighted Random Forests,” Quantitative Biology, vol. 5, no. 4, pp. 338–351, 2017, doi: 10.1007/s40484-017-0121-6.
[33] J. Sun, G. Zhong, K. Huang, and J. Dong, “Banzhaf Random Forests: Cooperative Game Theory Based Random Forests With Consistency,” Neural Networks, vol. 106, pp. 20–29, 2018, doi: 10.1016/j.neunet.2018.06.006.
[34] S. M. F. D. Syed Mustapha, “Predictive Analysis of Students’ Learning Performance Using Data Mining Techniques: A Comparative Study of Feature Selection Methods,” Applied System Innovation, vol. 6, no. 5, p. 86, 2023, doi: 10.3390/asi6050086.