# Predictive Analysis of Player Performance and Game Outcomes in CS:GO Matches

Department of Computer Science, Emory University, Atlanta, GA, USA

Kiran Kumar Nimmakayala Sam Miller Aadarsh Eetikala

## Introduction

Counter-Strike: Global Offensive (CS:GO) has become a powerhouse in the esports industry, captivating audiences worldwide with its intense first-person shooter gameplay. As one of the premier titles in competitive gaming, CS:GO has not only gained immense popularity but has also shaped the landscape of digital sports. With millions of players and viewers, CS:GO tournaments draw attention for their strategic depth, skillful plays, and dynamic team dynamics.

This project aims to explore the intricate world of CS:GO by leveraging advanced data analytics to predict game outcomes and player performances. In an environment where data-driven insights are crucial for strategic decision-making, this study delves into over 1400 competitive matchmaking matches to uncover patterns and strategies that define winning teams and top-performing players. Through the lens of machine learning techniques, the project not only seeks to enhance our understanding of CS:GO dynamics but also aspires to provide actionable insights for players and teams striving for excellence in this highly competitive arena.

The objectives of this study are twofold. First, we aim to analyze the rich dataset, comprising 'Games,' 'Damage,' 'Grenade,' 'Players,' and 'Results' datasets, to gain a holistic view of CS:GO dynamics. Second, by employing machine learning models such as Linear Regression, Decision Trees, Random Forest, KNN, and Gradient Boosting, we seek to provide clear, actionable insights into the determinants of player performance and team success in CS:GO.

## Dataset Descriptions

### *Game Performance*

Our analysis begins with an exploration of game dynamics, drawing from the 'Games', 'Damage', and 'Grenade' datasets. The 'Games' dataset provides a macro view of each match, detailing the flow of rounds and map-specific strategies. The 'Damage' and 'Kills' datasets delve into the micro aspects of gameplay, chronicling every combat interaction, weapon choice, and player positioning. These datasets together offer a comprehensive look into the tactical and strategic elements that influence game outcomes, including how teams adapt to different maps and exploit combat scenarios to gain an advantage.

### *Player Performance*

The player-centric aspect of our analysis is illuminated by the 'Players' and 'Results' datasets. The 'Players' dataset captures individual performances, including statistics on kills, assists, deaths, and economic contributions, essential for assessing skill and impact in each match. The 'Results' dataset complements this by providing a broader perspective on match outcomes, team rankings, and map victories. This dual dataset approach allows for a nuanced understanding of how individual prowess translates into team success and shapes the competitive hierarchy in CS:GO.

## Methodology

### *Game Outcome Prediction*

#### *Data Merging*

The amalgamation of diverse datasets—'Games,' 'Kills,' 'Damage,' and 'Grenades'—was a pivotal task, considering the substantial dataset size. However, this meticulous merging process resulted in a robust master dataset, providing a comprehensive panorama of CS:GO game outcomes.

#### *Data Preprocessing*

To ensure the integrity of our analysis, a rigorous data preprocessing phase was executed:

- One-hot encoding: Applied to multiclass categorical data.
- Standard scaling: Implemented on numerical data.
- Imputation: Techniques were employed to handle missing values.
- Training-testing split: Executed with a 70/30 ratio.

#### *Feature Engineering*

A set of new features was introduced to capture nuanced gameplay aspects:

| **Dataset** | **Features** |
|---|---|
| **kills** | First blood |
| | Time to first blood |
| | Average time between kills |
| | Time to fist blood / round duration |
| | Average time between kills / round duration |
| | Min/max/avg distance between kills |
| | CT/T survivors |
| | CT/T trades |
| | Kills before/after planting |
| **Damage** | Damage / second |
| | Non-lethal damage instances |
| | Total non-lethal damage |
| | Assists |
| | Friendly fire instances |
| | First damage |
| | Time to first damage |
| | Time to first damage / round duration |
| **Grenades** | Total grenades thrown |
| | CT/T grenade ratio |
| | Types of grenades thrown |
| **Misc** | Round duration |
| | Is bomb planted? |

### *Classification Models*
Our predictive models included Logistic Regression, Random Forest, Decision Tree, Gradient Boosting, and XGBoost.

### *Model Training and Selection*
Hyperparameter tuning was conducted using 5-fold cross-validation via grid search. The optimal hyperparameter values for each model were determined.

| Model | Hyperparameter(s) | Search Space(s) | Optimal Value(s) |
|---|---|---|---|
| **Logistic Regression** | C (lambda) | [0.001, 0.01, 0.1, 1, 10, 100, 1000] | 1 |
| **Decision Tree** | Max depth | [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] | 5 |
| | Min samples per leaf | [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] | 3 |
| **Random Forest** | N-estimators | [50, 100, 150, 200] | 200 |
| | Max depth | [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] | 6 |
| | Min samples per leaf | [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] | 5 |
| **XGBoost** | N-estimators | [50, 100, 150, 200] | 150 |
| | Subsample | [0.8, 0.9, 1.0] | 0.8 |
| | Max depth | [3, 4, 5, 6, 7] | 7 |
| | Min child weight | [1, 2, 3] | 2 |
| | Eta | [0.01, 0.1, 0.2] | 0.1 |
| | Alpha | [0, 0.1, 0.5, 1] | 0 |
| | Gamma | [0, 0.1, 0.2, 0.5, 1] | 0.2 |

## ***Player Performance Prediction***
### *Data Preproccessing*
In the pursuit of a comprehensive analysis linking individual player performance metrics with team outcomes, our data merging methodology focused on utilizing common identifiers such as ‘match_id’, ‘_map’, and ‘team’. This meticulous approach resulted in a unified dataset that seamlessly combined individual player statistics with team performance metrics. The integration provides a holistic view, laying the foundation for in-depth analysis and predictive modeling.

The 'Results' dataset underwent a critical transformation to align it with the player data. This involved segregating the dataset based on team performance and restructuring the data to reflect rounds won and lost for each team, thereby providing a more granular view of match outcomes.

### *Feature Engineering*
To elevate the interpretability of our 'Players' dataset, we embarked on a feature engineering journey. This involved aggregating player performance data to create unified indicators, such as kills, assists, and deaths. Simultaneously, we implemented a unique numeric ID approach for

handling categorical data imbalances, steering clear of potential bias associated with high-cardinality features like team names. This dual-pronged approach in feature engineering aimed at creating a more comprehensive and interpretable dataset.

| DATASET | FEATURES | |
|---|---|---|
| | ***Before*** | ***After*** |
| **RESULTS** | date(yyyy-mm-dd) | Date (yyyy) |
| | Team_1, Team_2 | Team |
| | Map | Map |
| | result_1 | rounds_won |
| | result_2 | rounds_lost |
| **PLAYERS** | m1_kills, m1_assists, m1_deaths,<br>m1_hs, m1_flash_assists, m1_kast,<br>m1_kddiff, m1_adr, m1_fkdiff, m1_rating<br><br>m2_kills, m1_assists, …..<br><br>m3_kills, m3_assists, ….. | Kills<br>Assists<br>Deaths<br>HS<br>Flash_Assists<br>Kast<br>Kd_diff<br>ADR<br>Rating<br>Fk_diff |

*Regression Models*

In our pursuit of predictive insights, various regression models, including Linear Regression, Decision Trees, Random Forest, Gradient Boosting, and KNN, were employed. Rigorous model evaluations, encompassing metrics like RMSE, MSE, MAE, R2, and MAPE, provided a robust understanding of each model's performance. Hyperparameter tuning further optimized the models, and the incorporation of cross-validation methods ensured their resilience.

## Results and Analysis

### *Game Outcome Prediction*

The Linear Regression model emerges as a standout performer, showcasing an impressive combination of accuracy and precision on the test set. With remarkably low RMSE, MSE, and MAE, the model excels in minimizing prediction errors. The elevated R2 attests to the robust correlation between predicted and actual values. Notably, the low MAPE underlines the model's ability to maintain a minimal percentage difference in predictions, instilling confidence in its reliability for CS:GO game outcome forecasts.

The Decision Tree model, while demonstrating extraordinary accuracy on the training set with near-zero RMSE and MSE, prompts a nuanced examination of potential overfitting. Rigorous evaluation on both validation and test sets is imperative to discern its generalization capabilities. Despite its excellent performance in the training phase, the Decision Tree's real test lies in its adaptability to new data scenarios.

In line with the Decision Tree, the Random Forest model showcases robust performance during training, reflecting its ensemble strength in mitigating overfitting concerns. The model's proficiency in maintaining accuracy on the test set underscores its reliability for predicting CS:GO game outcomes. The ensemble nature of Random Forest ensures resilience against idiosyncrasies in the training data, contributing to its consistent predictive power.

| Game Outcome Results | | | | |
|---|---|---|---|---|
| **Model** | **Dataset** | **F1** | **F2** | **AUC** |
| Logistic Regression | Training | 0.9747 | 0.9378 | 0.9954 |
| | Validation | 0.9748 | 0.9738 | 0.9957 |
| | Test | 0.9827 | 0.9836 | 0.9823 |
| Decision Tree | Training | 0.9822 | 0.9808 | 0.9966 |
| | Validation | 0.9823 | 0.9807 | 0.9963 |
| | Test | 0.9836 | 0.9841 | 0.9832 |
| Random Forest | Training | 0.9781 | 0.9787 | 0.9968 |
| | Validation | 0.978 | 0.9784 | 0.9967 |
| | Test | 0.9788 | 0.9795 | 0.9783 |
| XGBoost | Training | 0.98 | 0.9805 | 0.997 |
| | Validation | 0.9775 | 0.9779 | 0.9965 |
| | Test | 0.9835 | 0.984 | 0.983 |

The Gradient Boosting model, leveraging ensemble techniques, competes effectively with its counterparts. By combining weak learners, it offers accurate predictions and holds potential for enhancing CS:GO game outcome forecasting. Concurrently, the kNN model, being sensitive to data characteristics, necessitates careful scrutiny to decipher its suitability for the CS:GO context.

### *Player Performance Prediction*

Our predictive modeling work focused on player performance, applying multiple machine learning models to understand the complex dynamics that contribute to CS:GO success. The following results provide a complete evaluation of the models used, offering light on their ability to predict player performance.

Random Forest stood out among the models, with the lowest RMSE (Root Mean Square Error) and MAE (Mean Absolute Error), indicating its accuracy in predicting player performance. The strong R2 value (0.9762) emphasizes the model's capacity to explain variance in our data, highlighting its robustness.

Linear Regression showcased commendable performance, providing a solid baseline for player performance prediction. With an R2 score of 0.9667, it demonstrates a strong correlation between the predicted and actual values. While Decision Tree displayed respectable performance, it exhibits a slightly higher error rate compared to Random Forest and Linear Regression. The trade-offs between interpretability and accuracy should be considered when opting for this model.

KNN and Gradient Boosting, with their intermediary performance metrics, position themselves as reliable alternatives. These models strike a balance between accuracy and interpretability, catering to different analytical preferences.

Our results not only demonstrate these models' predictive abilities but also provide actionable information for CS:GO, players, and teams. Using these forecasts, players can improve specific parts of their gaming, such as economic efficiency, strategic adaptability, or kill involvement, thereby improving their overall competitive success. As the esports landscape evolves, these prediction models become an important tool for continual improvement and strategic refinement.

| Player Performance Results | | | | | | |
|---|---|---|---|---|---|---|
| **Model** | **Dataset** | **RMSE** | **MSE** | **MAE** | **R2** | **MAPE** |
| **Linear Regression** | Train | 0.060443196 | 0.00365338 | 0.046491652 | 0.966814741 | 4.904732518 |
| | Validation | 0.060704876 | 0.003685082 | 0.046772594 | 0.966195284 | 4.906930436 |
| | Test | 0.060522753 | 0.003663004 | 0.046557109 | 0.966469298 | 4.896220774 |
| **Decision Tree** | Train | 1.12841181 | 1.273313225 | 8.7315559 | 1.00 | 9.371502 |
| | Validation | 0.073987669 | 0.005474175 | 0.057718596 | 0.949783225 | 6.013014527 |
| | Test | 0.074185715 | 0.00550352 | 0.057911186 | 0.949621426 | 6.052339759 |
| **Random Forest** | Train | 0.019216542 | 0.000369275 | 0.014934863 | 0.996645708 | 1.56737215 |
| | Validation | 0.051542238 | 0.002656602 | 0.040383909 | 0.975629935 | 4.218384655 |
| | Test | 0.051637519 | 0.002666433 | 0.040411376 | 0.975591784 | 4.232490515 |
| **Gradient Boosting** | Train | 0.051454719 | 0.002647588 | 0.040440521 | 0.975950791 | 4.257506034 |
| | Validation | 0.052146827 | 0.002719292 | 0.040946942 | 0.975054862 | 4.285067863 |
| | Test | 0.05211748 | 0.002716232 | 0.040878032 | 0.975135936 | 4.293991423 |
| **kNN** | Train | 0.046160857 | 0.002130825 | 0.036128229 | 0.980644781 | 3.771249292 |
| | Validation | 0.057149655 | 0.003266083 | 0.044644539 | 0.970038927 | 4.640019437 |
| | Test | 0.056732239 | 0.003218547 | 0.044406375 | 0.970537801 | 4.616550676 |

## Discussion

### *Important Features*

In predicting CS:GO game outcomes, features such as "First blood," "Time to first blood," and "Average time between kills" emerged as pivotal in determining match results. These features provide insights into the early dynamics of a match, emphasizing the importance of securing the initial advantage. Additionally, factors like "Total grenades thrown" and "CT/T grenade ratio" underscore the strategic significance of utility usage, highlighting the impact of tactical decision-making on team success.

For player performance predictions, the range of metrics, including kills, assists, deaths, headshots, flash assists, and more, showcases the multifaceted nature of CS:GO gameplay. The Random Forest model demonstrated exceptional accuracy in discerning the nuanced interplay of these diverse metrics. Beyond fragging capabilities, metrics like "ADR" and "Kd_diff" reflect economic efficiency, while "Kast" captures strategic contributions, emphasizing the holistic nature of success in CS:GO.

### *Range of Player Performance*

The analysis of player performance spanned a spectrum of metrics, encapsulating kills, assists, deaths, headshots, flash assists, and more. The diversity in these metrics reflects the multifaceted nature of CS:GO gameplay, where individual contributions vary widely. Notably, the Random Forest model exhibited exceptional accuracy in predicting player performance, highlighting its capability to discern the nuanced interplay of these diverse metrics. The range of player performance metrics also emphasizes the holistic nature of success in CS:GO. Beyond fragging capabilities, factors such as economic efficiency (reflected in metrics like "ADR" and "Kd_diff") and strategic contributions (captured by "Kast" - Kill, Assist, Survive, Trade) play pivotal roles. This comprehensive approach to player evaluation mirrors the multifactorial nature of success in professional CS:GO.

*Practical Implications*

The predictive models presented in this study offer more than analytical prowess—they provide actionable insights for CS:GO players and teams. By understanding the determinants of game outcomes and individual player performances, teams can strategically focus on specific aspects of their gameplay. For instance, attention to early-game dynamics and effective use of utility can significantly impact match results. Similarly, players can tailor their training regimens based on identified performance metrics, optimizing their strengths and addressing weaknesses. As esports continue to evolve, these prediction models serve as valuable tools for continual improvement and strategic refinement, aligning with the dynamic nature of competitive CS:GO.

## Previous Work

This study contributes to the growing body of esports analytics research by taking a comprehensive approach to understanding CS:GO dynamics. While previous works have explored specific aspects of game outcomes or player performances, our study stands out in its holistic examination, incorporating diverse datasets and utilizing a range of machine learning models. The meticulous merging of 'Games,' 'Damage,' 'Grenade,' 'Players,' and 'Results' datasets provides a macro and micro view, capturing both the strategic depth and individual nuances that shape CS:GO matches. The inclusion of features like early-game dynamics, utility usage, and economic contributions sets our work apart, offering a nuanced perspective on the determinants of success in competitive gaming.

## Future Scope

Looking ahead, the future focus of this research will be further improving and expanding the created prediction models. While the current study provides a solid foundation, future research could look at new datasets or go deeper into certain areas of gameplay. Stricter feature selection, as indicated in the discussion, is an important avenue for improvement, ensuring that the models are resistant to potential overfitting. Furthermore, the idea of a real-time prediction system opens the door to dynamic applications during live matches, providing players and teams with quick information. As esports analytics continue to improve, our work serves as a springboard for future academics to investigate new dimensions and refine the application of machine learning in the competitive scene of CS:GO.

## Conclusion

In navigating the intricate world of CS:GO esports, this study has delved into over 1400 competitive matches, wielding advanced data analytics to illuminate the strategic nuances that define success. From early-game dynamics and utility usage to the multifaceted spectrum of player performance metrics, our analysis has unearthed pivotal insights. The predictive models, particularly the Linear Regression model for game outcomes and the Random Forest model for player performances, have showcased impressive accuracy during testing, offering a glimpse into the future of esports analytics.

As players and teams aim for excellence in the fiercely competitive landscape of CS:GO, the importance of these findings cannot be overstated. Beyond analytical curiosity, the predictive models presented here serve as invaluable tools for strategic refinement and continual improvement. The prospect of a real-time prediction system adds a dynamic layer to the application, providing on-the-fly insights as matches unfold. Looking forward, refinement of feature selection processes and a vigilant eye on potential overfitting open doors for future exploration and innovation. In the ever-evolving arena of esports, where precision and adaptability are paramount, this study marks not just a culmination but a stepping stone towards a more insightful and strategic future in competitive gaming.

## Codework:

https://github.com/kiran-001/CSGO-Predictive-Analysis.git

## References


1. CS:GO Competitive Matchmaking Data
   https://www.kaggle.com/datasets/skihikingkevin/csgo-matchmaking-damage

2. CS:GO Professional Matches
   https://www.kaggle.com/datasets/mateusdmachado/csgo-professional-matches

3. Bahrololloomi, F., Klonowski, F., Sauer, S. et al. E-Sports Player Performance Metrics for Predicting the Outcome of League of Legends Matches Considering Player Roles. SN COMPUT. SCI. 4, 238 (2023). https://doi.org/10.1007/s42979-022-01660-6

4. W. Xu Huang, J. Wang and Y. Xu, "Predicting Round Result in Counter-Strike: Global Offensive Using Machine Learning,"2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China, 2022, pp. 1685-1691, doi:10.1109/ICSP54964.2022.9778597.

5. Xenopoulos, P., Doraiswamy, H., & Silva, C. (2020). Valuing player actions in Counter-Strike: Global Offensive. 2020 IEEE International Conference on Big Data (Big Data). https://doi.org/10.48550/arXiv.2011.01324