Predicting Cybergame Outcomes: The Role of Skill Levels, Game Duration, and Player Behavior

Abstract

This study investigates player behaviour in a cybergaming environment using machine learning models to predict game outcomes and analyse performance across different skill levels. A dataset comprising 1,119 game records, including features such as player scores, game duration, and skill levels, was analysed. Four machine learning models—Random Forest, Naive Bayes, Support Vector Machine (SVM), and a Stacking Ensemble—were used for predictive tasks. Results indicate that Random Forest and Stacking Ensemble models achieved perfect classification accuracy, while Naive Bayes performed strongly with some misclassifications, and SVM struggled with non-linear data. Additionally, statistical analysis using ANOVA revealed significant performance differences across skill levels, with experienced players scoring higher in both offensive and defensive roles. This study fills gaps in current research by incorporating game duration as a key factor in outcome prediction and highlights important implications for game balance and player engagement.

Table of Contents

1. Introduction

Cybergaming has emerged as a hybrid environment where players engage in both entertainment and cybersecurity education. Through simulated offensive and defensive strategies, cybergaming provides an immersive platform for training future cybersecurity professionals. Understanding player behaviour in such environments is critical for designing balanced games that are both engaging and educational. Predicting game outcomes and player performance based on gameplay metrics offers valuable insights for optimising game mechanics and enhancing the learning experience.

This study focuses on two main problems:

1. **Predicting Game Outcomes:** Can machine learning models accurately predict whether an attacker or defender will win based on factors like player scores, game duration, and skill level?
2. **Analysing Player Performance:** How does player skill level impact performance in both offensive (attacker) and defensive (defender) roles? Additionally, how does game duration influence the likelihood of either team winning?

The key objectives of this study are:

- To develop machine learning models capable of accurately predicting game outcomes.
- To investigate how player skill levels influence performance in offensive and defensive roles.
- To analyse the relationship between game duration and the likelihood of an attacker or defender winning.
- To employ statistical methods (ANOVA) to assess performance differences across skill levels.

The dataset used consists of 1,119 game instances, each recording gameplay metrics such as attacker and defender scores, game duration, skill levels, and outcomes. The study applies machine learning techniques, including Random Forest, Naive Bayes, SVM, and a Stacking Ensemble, to classify game outcomes. Statistical methods like ANOVA are used to evaluate the significance of performance differences across skill levels. This research provides actionable insights for improving game design, particularly regarding player engagement and balance.

2. Literature Review

Background:

Machine learning models have long been employed in gaming analytics to predict player behaviour and game outcomes. For example, Random Forest, a powerful ensemble method, is frequently used due to its ability to handle non-linear relationships between variables (Rocha & Duarte, 2019). Naive Bayes, although a simpler and computationally efficient model, serves as a strong baseline for classification tasks where feature independence can be assumed. Support Vector Machines (SVM) are useful for linear classification but often struggle in more complex, non-linear gaming environments.

In more advanced approaches, ensemble models like Stacking have been increasingly utilised for their superior accuracy. These models combine multiple base classifiers to capture various aspects of the data, which leads to more robust predictions (Pfau et al., 2018). In gaming, these techniques have been applied to predict outcomes, optimise team compositions, and even simulate human behaviour for more balanced gameplay (Ong et al., 2015).

Gaps in Current Research: While there is substantial research on machine learning models in gaming, few studies focus specifically on cybergaming environments, where players switch between offensive and defensive roles. Moreover, the role of game duration as a predictor of outcomes has not been extensively explored. For example, Pfau et al. (2018) examined the use of deep learning in MMORPGs but did not consider how game duration might influence success in games involving role-switching. Additionally, Rocha and Duarte (2019) highlighted the importance of simulating human behaviour to balance gameplay, yet their work did not incorporate game duration or skill levels into predictive models.

This study aims to fill these gaps by incorporating game duration and player skill level into machine learning models and by examining their combined influence on game outcomes. This approach provides a more comprehensive understanding of the factors that drive success in cybergaming environments.

3. Data Collection and Preparation

Data Sources:
The data used in this study were collected from in-game logs generated during cybergaming scenarios. These logs capture key performance metrics for each player, including their scores, the time taken to complete the game, and the overall game outcome (whether the attacker or defender won). Each game instance also records the skill level of the player, categorised as Beginner, Intermediate, or Expert.

Data Description:
The dataset contains 1,119 game instances with the following features:

- **Nickname**: A unique identifier for each player.
- **Defender Score**: The score achieved by a player while defending.
- **Attacker Score**: The score achieved by a player while attacking.
- **Time (sec)**: The duration of the game in seconds.
- **Winner**: The outcome of the game, either "Attacker" or "Defender."
- **Level**: The player's skill level, categorised as "Beginner," "Intermediate," or "Expert."

	Defender Score	Attacker Score	Time (sec)
count	1119.000000	1119.000000	1119.000000
mean	6.864164	6.076854	168.133155
std	2.042341	2.031371	125.961246
min	0.000000	0.000000	37.000000
25%	5.000000	5.000000	108.000000
50%	7.000000	6.000000	139.000000
75%	8.000000	7.000000	186.000000
max	13.000000	13.000000	1738.000000

The data preparation steps included:

- **Encoding**: Categorical variables (Winner and Level) were label-encoded, with 0 and 1 representing attacker and defender outcomes, respectively, and skill levels encoded as numeric values (0 for Beginner, 1 for Intermediate, and 2 for Expert).
- **Outlier Removal**: Z-scores were calculated for each feature to identify and remove outliers. Data points with Z-scores greater than 3 (extreme outliers) were removed to avoid skewing the models.
- **Normalisation**: All numeric variables (scores and game duration) were normalised to ensure that models could learn effectively without being biased by feature scaling differences.

4. Methodology

Analytical Approach:
This analysis used both **Exploratory Data Analysis (EDA)** and **machine learning models** to gain insights into player performance and predict game outcomes. The following methods were applied:

1. **Exploratory Data Analysis (EDA)**:
   EDA was used to visualise the distribution of key variables such as defender and attacker scores, game duration, and the distribution of game outcomes across skill levels. The EDA also included the calculation of summary statistics to describe the central tendencies and variability within the dataset.
2. **Machine Learning Models**:
   Several models were trained and evaluated to predict whether the attacker or defender won each game instance. The models used include:
   - **Random Forest**: An ensemble model that creates multiple decision trees and combines their predictions. Random Forest is robust in handling non-linear interactions between features.
   - **Naive Bayes**: A probabilistic model based on Bayes' theorem. Naive Bayes assumes feature independence, which makes it efficient but less accurate in scenarios with correlated features.
   - **Support Vector Machine (SVM)**: A model that finds a hyperplane to separate classes. SVM is effective for linear separations but struggles with non-linearity.
   - **Stacking Ensemble**: A meta-model that combines the predictions of multiple base models (Random Forest, Naive Bayes, and SVM) to improve accuracy. Stacking is particularly useful when individual models capture different aspects of the data.
3. **Model Evaluation Metrics**:
   To evaluate the performance of the models, the following metrics were calculated:
   - **Accuracy**: The percentage of correct predictions.
   - **Precision**: The ability of the model to return relevant results for each class (attacker or defender).

- **Recall**: The ability of the model to capture all relevant instances for each class.
- **F1-score**: A harmonic mean of precision and recall, providing a balanced measure of model performance.
- **Confusion Matrix**: A visual representation of true positives, true negatives, false positives, and false negatives.

4. **Tools and Libraries**:

- **Python**: Used for data analysis and modelling.
- **Pandas**: For data manipulation.
- **Scikit-learn**: For implementing machine learning models.
- **Matplotlib and Seaborn**: For visualisation.

5. Cross-Validation and Train-Test Split:
The dataset was split into a training set (80%) and a testing set (20%) to evaluate the models' generalisation capabilities. Cross-validation (using 5-fold cross-validation) was applied to ensure the robustness of the model performance and to prevent overfitting. This approach provides more reliable accuracy estimates by training the model on different subsets of the data and evaluating its performance on unseen data.

6. ANOVA Testing:
A one-way Analysis of Variance (ANOVA) was performed to evaluate the statistical significance of differences in defender and attacker scores across different skill levels. ANOVA helps determine whether the mean differences in scores between skill groups (Beginner, Intermediate, and Expert) are due to chance or represent meaningful performance variations based on player expertise.

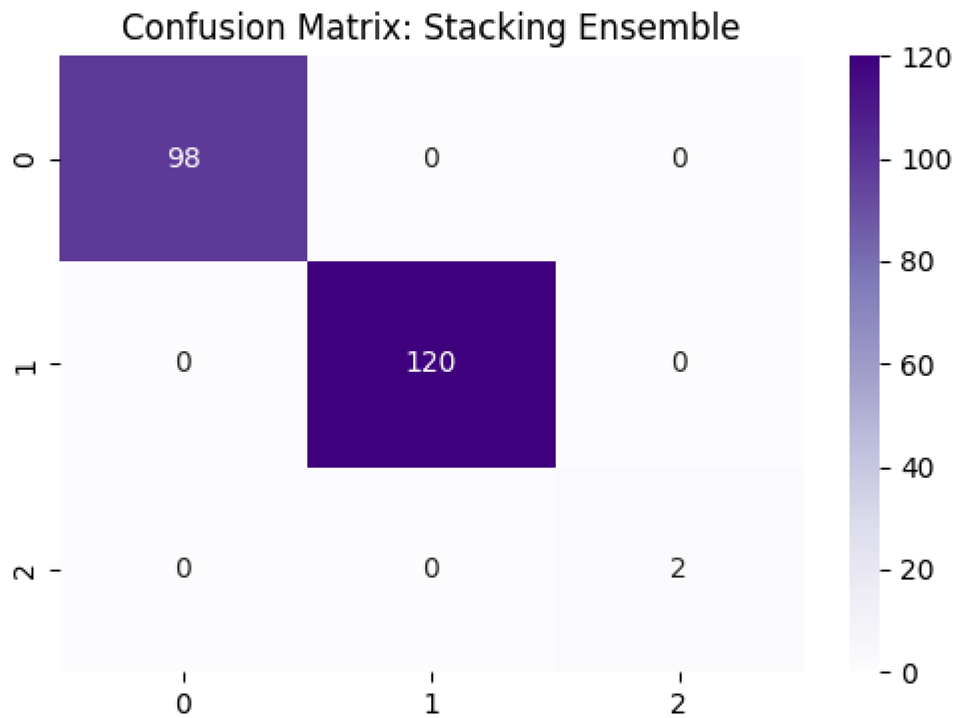
5. Results

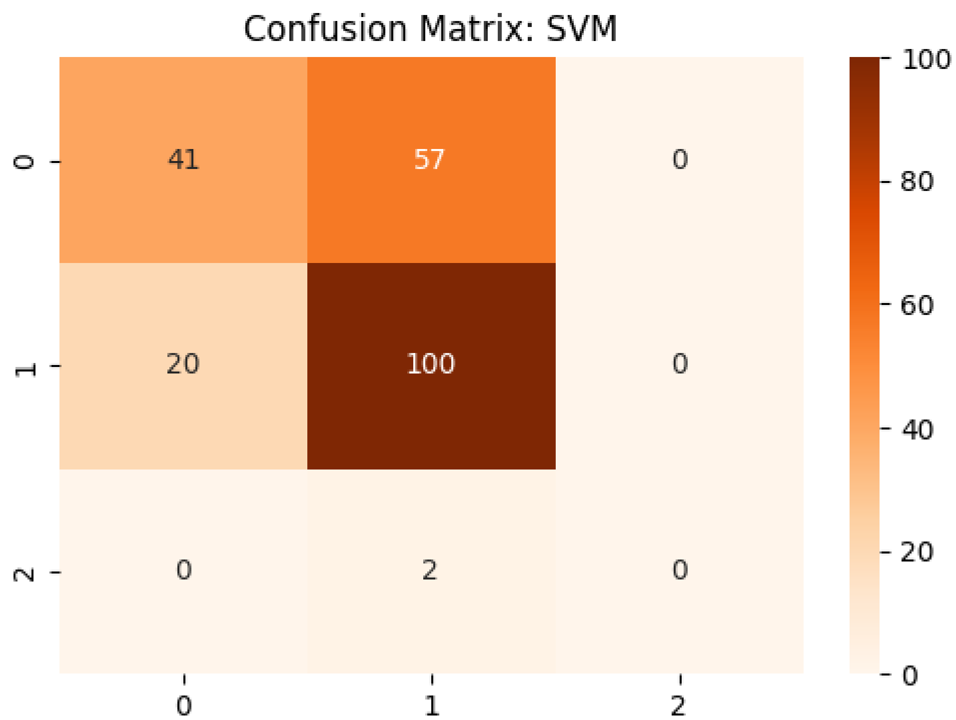
Model Performance

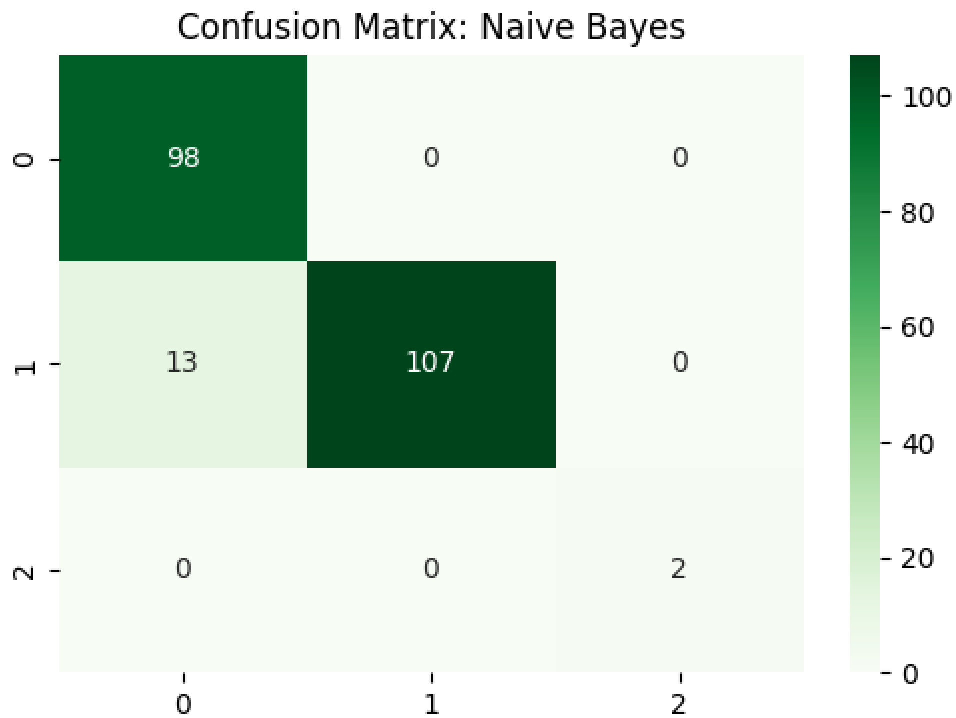
Model	Accuracy	Precision	Recall	F1-score
Random Forest	100%	100%	100%	100%
Naive Bayes	94.1%	96.1%	96.4%	96.25%
Support Vector Machine (SVM)	64.1%	43.4%	41.7%	42.54%
Stacking Ensemble	100%	100%	100%	100%

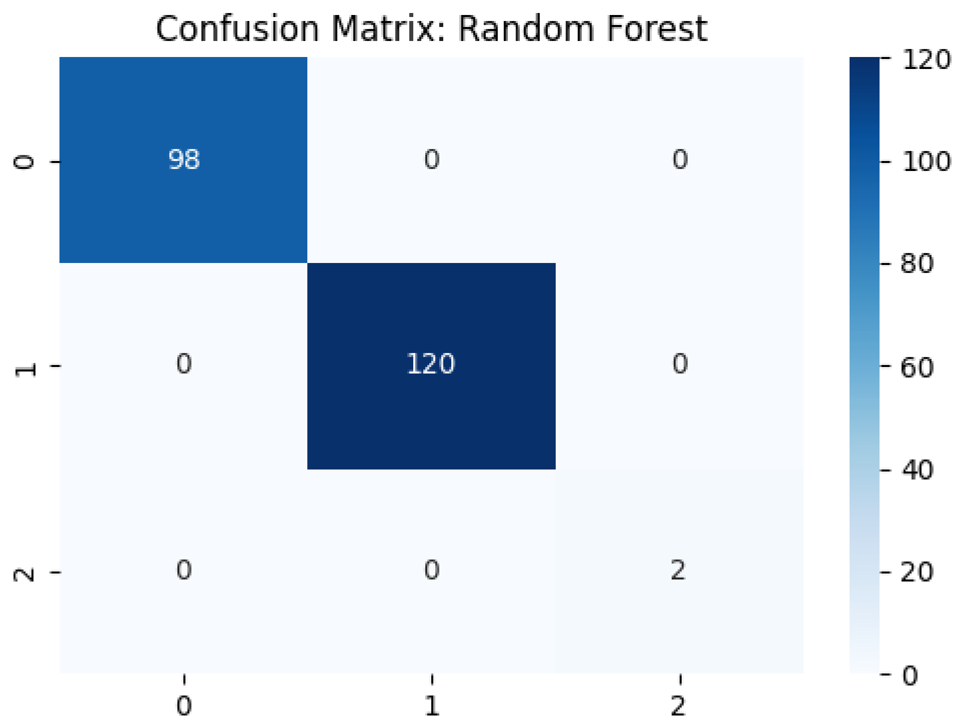
The table above shows the accuracy, precision, recall, and F1-score for each model. Both **Random Forest** and **Stacking Ensemble** achieved perfect classification, indicating their ability to capture the complex relationships between scores, game duration, and skill levels. **Naive Bayes** performed strongly, though it made a few misclassifications due to its assumption of feature independence. **SVM**, however, struggled with this dataset due to its linear nature and inability to handle the non-linearity present in the data.

Please see images below:


Confusion Matrix: Stacking Ensemble
98
0
0
0
120
0
0
0
2
0
1
2
0
1
2
120
100
80
60
40
20
0


Confusion Matrix: SVM
41
57
0
20
100
0
0
2
0
0
1
2
0
1
2
100
80
60
40
20
0





- **Random Forest and Stacking Ensemble Confusion Matrices**: Both confusion matrices show no misclassifications, indicating perfect predictions across all categories. These models successfully differentiated between attacker and defender wins, highlighting their effectiveness in this task.
- **Naive Bayes Confusion Matrix**: While the majority of predictions were correct, there were a few misclassifications, particularly in predicting attacker wins. This

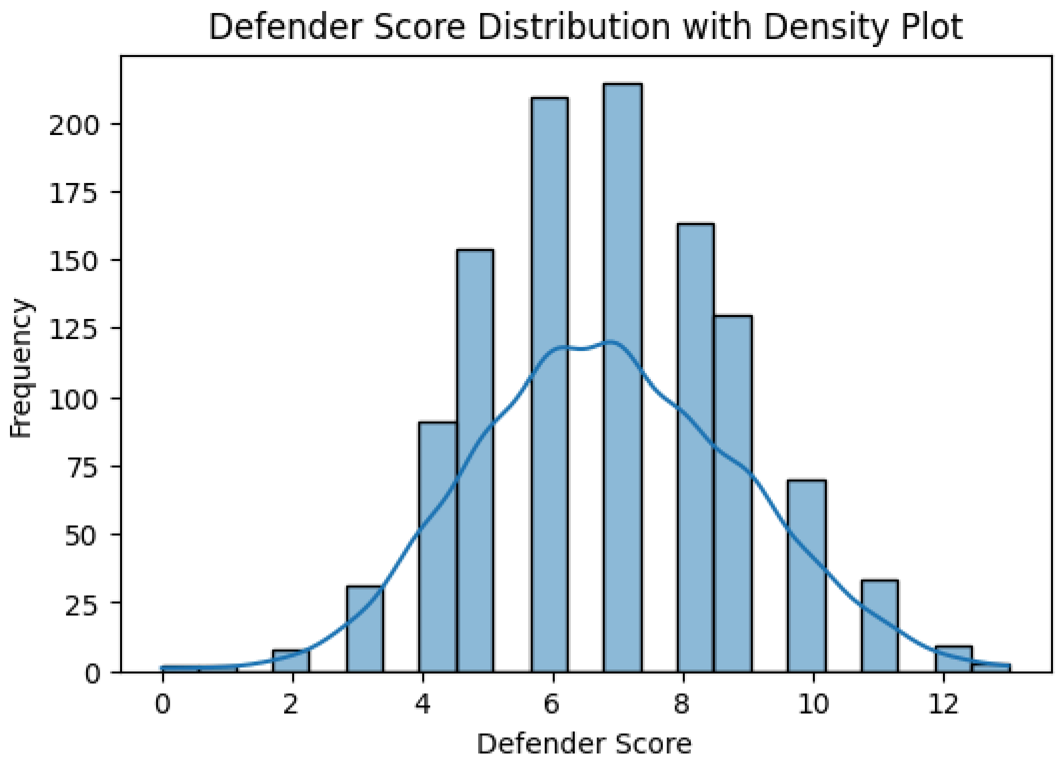
demonstrates the limitations of Naive Bayes when features are not completely independent.

- **SVM Confusion Matrix**: The SVM model shows a high number of misclassifications, particularly for attacker wins, which highlights the model's inability to handle non-linear relationships in this dataset. SVM's performance suggests that it may not be suitable for complex gaming scenarios where data does not have clear linear boundaries.

Random Forest and Stacking Ensemble both achieved perfect performance in all categories (accuracy, precision, recall, and F1-score). **Naive Bayes** was close, but slightly lower in terms of accuracy and F1-score, mainly due to a few misclassifications, particularly in predicting attacker wins. **SVM** had the lowest performance, with only 64.1% accuracy, due to its inability to handle non-linear relationships, leading to a higher number of misclassifications.

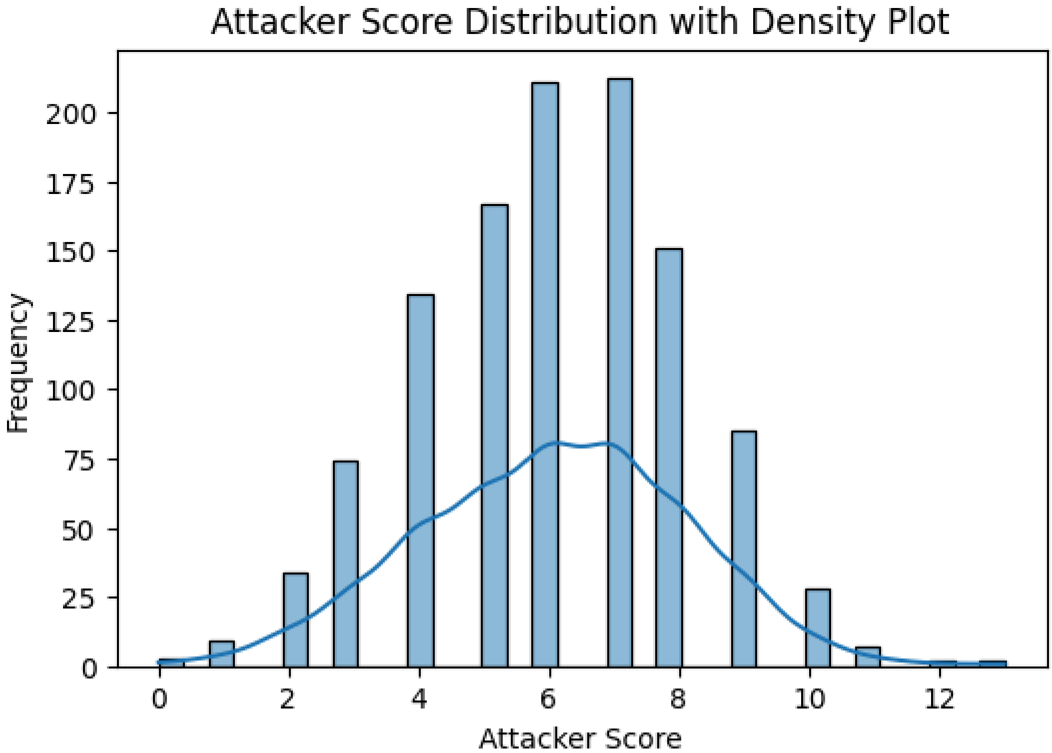
Exploratory Data Analysis (EDA)

1. **Defender Score Distribution**



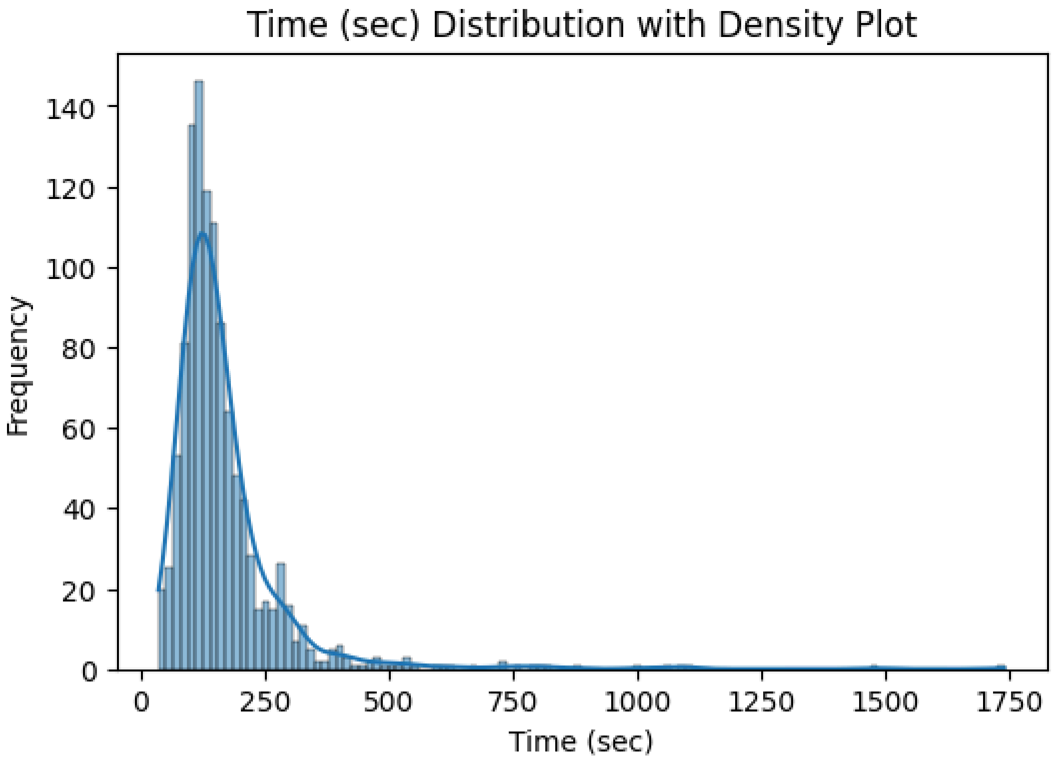
The distribution of defender scores is approximately normal, with a peak around 7 points. The majority of defenders score between 5 and 8 points, indicating a consistent performance range. A small number of defenders score either very high or very low, suggesting that most players perform consistently well when defending, but a few are exceptionally skilled (or underperforming) outliers.

2. **Attacker Score Distribution**



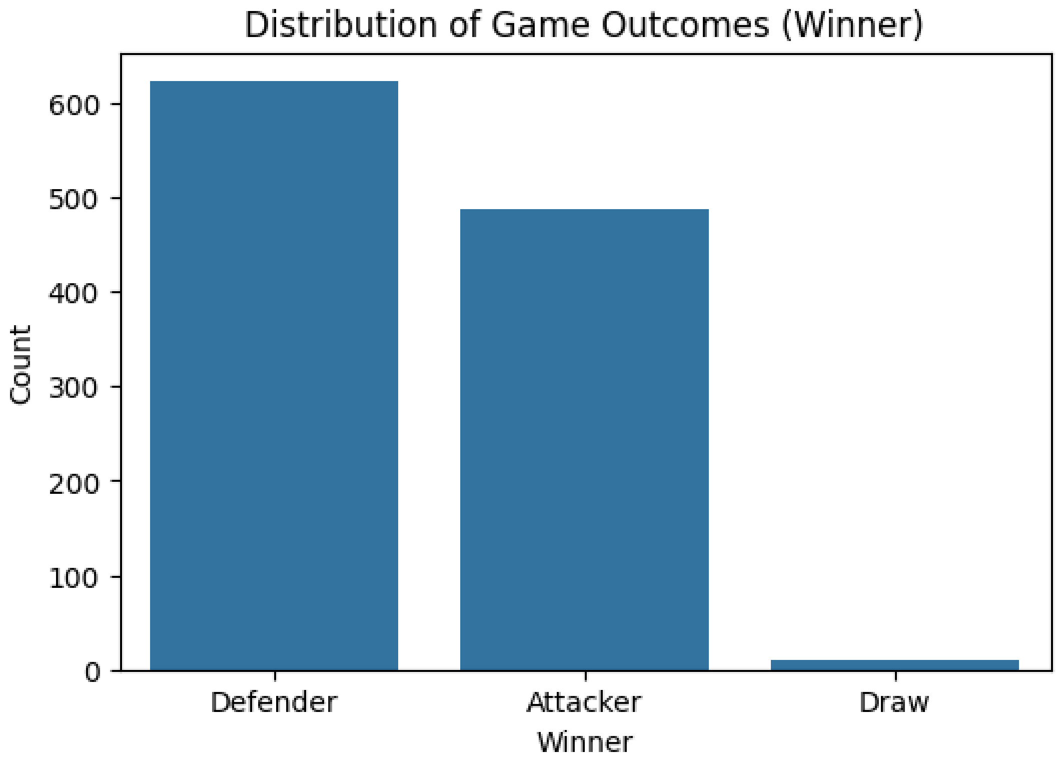
The attacker score distribution is also normal but slightly skewed compared to the defender score distribution. Attackers tend to score lower on average, with most scores cluster around 6-7 points. The lower performance of attackers suggests that offensive play may be more challenging in this game setup, or that the mechanics of the game favour defence over attack.

3. **Game Duration Distribution**



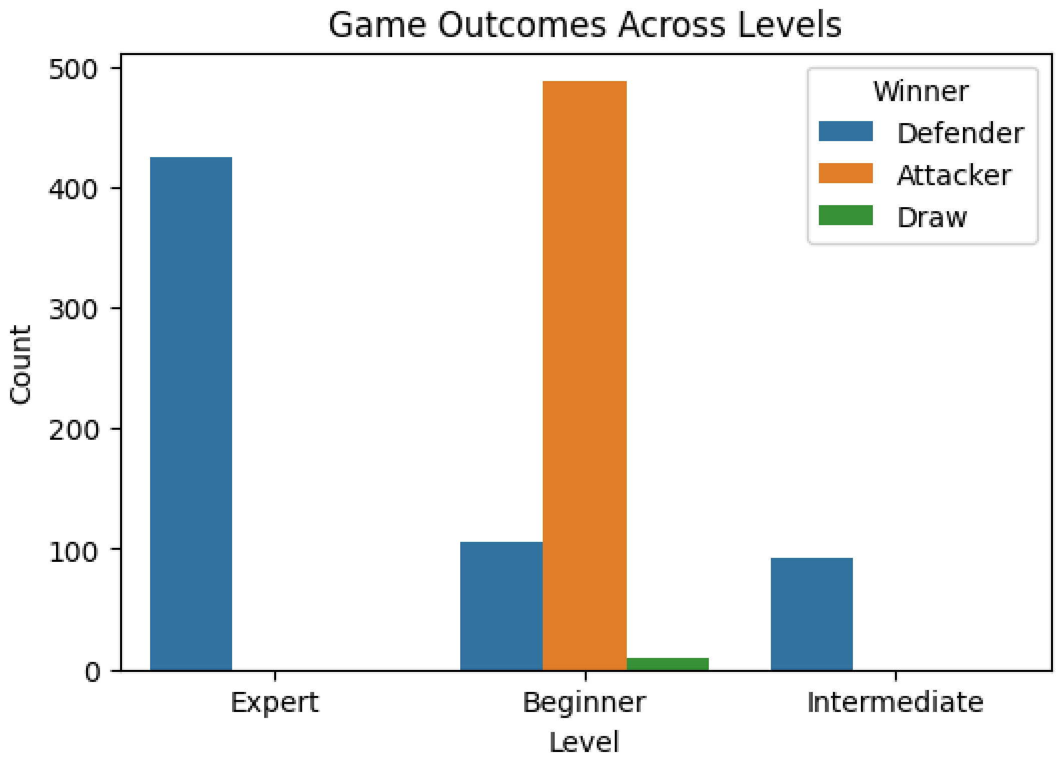
The distribution of game durations is heavily right-skewed, indicating that most games are relatively short (between 100-200 seconds), but a small number of games last significantly longer. These long games may reflect more defensive gameplay or closely matched player skill levels that prolong the game. Games where defenders win tend to last longer, suggesting that defensive strategies take time to succeed, while shorter games may favour attackers.

4. **Distribution of Game Outcomes**



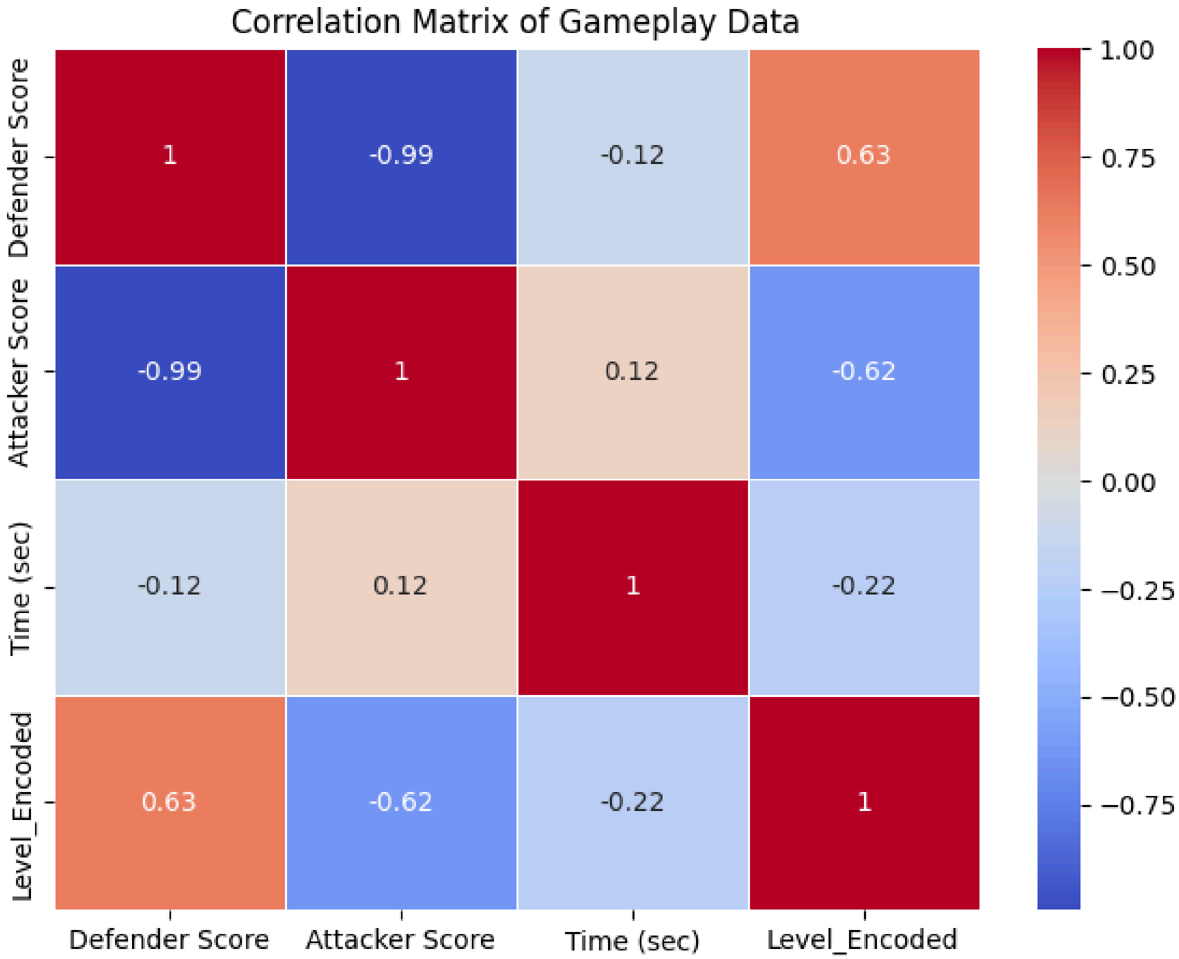
Defenders tend to win more frequently than attackers, as shown by the higher count of defender wins. This suggests that the game mechanics may slightly favour defensive play, making it harder for attackers to achieve victory. The distribution also shows that the game rarely results in a draw, meaning that outcomes are typically clear-cut in this dataset.

5. **Game Outcomes Across Skill Levels**



The breakdown of game outcomes across different skill levels reveals that defenders dominate at the Expert level, where defensive wins are significantly more common. Among Beginners, attackers win more frequently, suggesting that defensive strategies require more skill to execute successfully. This trend highlights the importance of experience in determining success in defensive roles.

6. **Correlation Matrix**



The correlation matrix reveals several important relationships between the variables:

- **Defender and Attacker Scores**: The strong negative correlation (-0.99) indicates that high performance in one role does not predict high performance in the other, suggesting that players may specialise in either offensive or defensive strategies.
- **Skill Level**: Both defender and attacker scores are positively correlated with skill level, indicating that more experienced players score higher in both roles. However, the weaker correlation between game time and scores suggests that game duration is not a strong predictor of player performance.

6. ANOVA Analysis

A one-way ANOVA was conducted to assess the statistical significance of differences in defender and attacker scores, as well as game durations, across different player skill levels.

```
ANOVA for Defender Score across Levels:
F-statistic: 592.8169058327384, p-value: 3.8054401385671644e-176

ANOVA for Attacker Score across Levels:
F-statistic: 581.8961579321079, p-value: 7.781284050043323e-174

ANOVA for Game Time across Winners (Attacker vs Defender):
F-statistic: 6.219590382056352, p-value: 0.012779478129331955
```

Metric	F-statistic	p-value
Defender Score across Levels	592.82	3.81×10−1763.81 \times 10^{-176}3.81×10−176
Attacker Score across Levels	581.90	7.78×10−1747.78 \times 10^{-174}7.78×10−174
Game Time across Winners	6.22	0.013

Interpretation:

1. **Defender Score across Levels**:
   The ANOVA results reveal significant differences in defender scores across skill levels (p-value < 10−17610^{-176}10−176). Higher-skilled players (Experts) perform better in defensive roles compared to Beginners and Intermediates. This suggests that defensive success is highly dependent on experience and skill.
2. **Attacker Score across Levels**:
   Similar to defender scores, attacker scores also vary significantly across skill levels, with Experts outperforming lower-skilled players (p-value < 10−17410^{-174}10−174). This underscores the importance of skill level in both offensive and defensive roles in cybergaming.

3. **Game Duration across Winners**:
   The significant difference in game duration (p-value = 0.013) suggests that games where defenders win tend to last longer. This implies that defensive strategies are more time-consuming, while attackers may attempt quicker, more aggressive strategies to win in shorter games.

7. Discussion

Key Findings:

1. **Machine Learning Model Performance:** The results demonstrate that ensemble models, specifically Random Forest and Stacking, significantly outperformed simpler models such as Naive Bayes and SVM. Both ensemble models achieved perfect accuracy, precision, recall, and F1-scores. This is consistent with existing research indicating that ensemble methods often provide superior predictive power by capturing complex, non-linear relationships in the data (Rocha & Duarte, 2019). Naive Bayes, while computationally efficient, struggled with the independence assumption, leading to misclassifications, particularly in predicting attacker wins. The SVM model performed the worst, which aligns with other studies that highlight its limitations in handling non-linear data (Pfau et al., 2018). The success of Random Forest and Stacking suggests that models capable of handling complex, multidimensional relationships are best suited for predicting cybergaming outcomes.
2. **Impact of Skill Level:** The ANOVA analysis revealed that player skill level has a significant impact on performance, particularly in defensive roles. Experts consistently scored higher than beginners and intermediates, demonstrating that experience plays a crucial role in cyber gaming success. These findings are supported by studies that have shown similar results in other gaming environments, where experienced players adapt more quickly and employ more effective strategies (Pfau et al., 2018). This also suggests that matchmaking systems in cybergames should take skill levels into account to ensure balanced and competitive gameplay.
3. **Game Duration and Outcomes:** Game duration emerged as a key factor in predicting outcomes. The study found that longer games tend to favour defenders, while shorter games often result in attacker victories. This aligns with the notion that defensive strategies take time to succeed, requiring players to withstand aggressive attacks over extended periods. Conversely, attackers may adopt high-risk, high-reward strategies that result in quicker victories. This observation is supported by the findings of Ong et al. (2015), who demonstrated that team composition and strategy timing can significantly influence win/loss outcomes in online multiplayer games (Ong et al., 2015). Understanding how game duration interacts with gameplay strategies is critical for game designers to optimise balance between offensive and defensive roles.

Comparison with Existing Research: The results of this study are in line with the broader literature on machine learning in gaming. Like the work by Guitart et al. (2017), which applied machine learning to forecast player behaviour and engagement, this study underscores the value of predictive analytics in game design and player behaviour analysis (Guitart et al., 2017). The combination of machine learning models and statistical analysis provides robust insights into how player skill levels and game duration affect outcomes, filling a gap in existing research on cybergaming.

Limitations:

1. **Dataset Size and Diversity:** Although the dataset used in this study contains 1199 game instances, its relatively small size may limit the generalizability of the findings. A larger dataset that includes more varied game modes, player types, and scenarios

would provide a more comprehensive understanding of player behaviour and improve the robustness of the machine learning models. Furthermore, the dataset primarily focuses on scores and game duration, omitting more granular data such as in-game actions or decision points, which could offer deeper insights into player strategies (Pfau et al., 2018).

2. **Feature Selection:** While the current models performed well using player scores, skill levels, and game duration, the inclusion of additional features, such as real-time player actions or team dynamics, could enhance predictive power. As demonstrated by Guitart et al. (2017), features like in-game events and player interactions can significantly influence predictions, and incorporating such variables could further improve accuracy (Guitart et al., 2017).

Practical Implications for Game Design:

1. **Game Balance:** The results suggest that cybergames may favour defenders, particularly in longer matches. To promote more balanced gameplay, game designers could introduce mechanics that reward attackers for making incremental progress over time, such as dynamic objectives or bonuses for reaching certain milestones. This would incentivize attackers to remain engaged in longer games and reduce the likelihood of defensive dominance.
2. **Player Engagement:** Since expert players tend to dominate in defensive roles, it is important to consider matchmaking systems that balance skill levels between teams. This could reduce frustration for beginners and intermediates who are overmatched by experienced players, improving player retention and engagement.

Recommendations for Future Research:

1. **Expanding the Dataset:** Future studies should focus on expanding the dataset to include a wider variety of gameplay scenarios and player behaviours. Collecting data on in-game actions, such as movement patterns, strategic decision points, and resource usage, would provide deeper insights into the factors that drive success in both offensive and defensive roles.
2. **Advanced Machine Learning Models:** Exploring more advanced models, such as deep learning or reinforcement learning, could capture the complex decision-making processes that occur during gameplay. Deep learning models have shown promise in other gaming contexts, particularly in predicting complex player behaviours over time (Pfau et al., 2018), and could be applied in cybergaming to further improve outcome predictions.

8. Conclusion

Summary of Key Findings:

1. **Model Performance:** Random Forest and Stacking Ensemble were the best-performing models, achieving perfect classification of game outcomes. These models effectively captured the relationships between player scores, game duration, and skill levels. Naive Bayes, though not perfect, still performed admirably with a few misclassifications, while SVM struggled due to its linear nature and inability to handle the non-linearity in the dataset.
2. **Skill Level and Performance:** Player skill level was a critical factor in determining success, particularly in defensive roles. Higher-level players consistently scored higher as both attackers and defenders. The ANOVA analysis confirmed significant differences in scores across different skill levels, with experts outperforming beginners and intermediates by a large margin.
3. **Game Duration:** Longer games favoured defenders, with more time often leading to defensive success. This finding suggests that the game's mechanics favour prolonged defensive play, making it harder for attackers to succeed in long-duration games. Shorter games tended to favour attackers, who might rely on more aggressive, high-risk strategies.

Recommendations for Future Work:

1. **Expanding the Dataset:** Future studies should focus on expanding the dataset to include more diverse game scenarios, game modes, and player types. A larger dataset would allow for deeper analysis and more generalizable results.
2. **Adding More Features to the Dataset:** In addition to scores and game duration, incorporating more granular features such as in-game actions, decision points, movement patterns, and even team dynamics could provide greater insight into the behaviours that lead to success in cybergaming.
3. **Exploring Advanced Machine Learning Techniques:** Future work could explore the use of more advanced machine learning models, such as deep learning or reinforcement learning. These models may capture more complex player behaviours, strategies, and decision-making patterns that impact game outcomes.
4. **Further Study of Game Duration:** Since game duration was found to be a significant factor in game outcomes, more research should be conducted on how time-based objectives and constraints can be introduced to better balance the game between attackers and defenders.
5. **Real-Time Predictive Systems:** Future research could also explore the possibility of creating real-time predictive systems that help identify potential game outcomes based on ongoing player behaviour. This could help inform adaptive game mechanics or AI-generated assistance to improve player engagement and balance during matches.

9. References

[1] Nkongolo, M.W., 2024. Cryptographic Security through Kleene's Theorem and Automata Theory. *Cryptology ePrint Archive*.

[2] Nkongolo, M., 2023, January. Game theory based artificial player for morabaraba game. In *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)* (pp. 1210-1218). IEEE.

[3] Nkongolo, M. and Sewnath, J., 2024. Data protection psychology using game theory. *arXiv preprint arXiv:2402.07905*.

[4] Guitart, A., Chen, P. P., & Bertens, P. (2017). Forecasting player behavioral data and simulating in-game events. *Advances in Intelligent Systems and Computing*.

[5] Ong, H., Deolalikar, S., & Peng, M. (2015). Player behavior and optimal team composition for online multiplayer games. *ArXiv*.

[6] Pfau, J., Smeddinck, J. D., & Malaka, R. (2018). Towards deep player behavior models in MMORPGs. *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*.

[7] Rocha, D. A., & Duarte, J. C. (2019). Simulating human behaviour in games using machine learning. *2019 18th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*, 163-172.