Team 14 AKA Team KARMMA
Kai Klienbard, Andrew Ranon, Roberto Pasquier, Mai Nguyen, Michael Daniels, Ameerkumar Upadhyay

# Final Write-Up: Revolutionizing Strategy in Clash Royale with Analytics

## Introduction

### Motivation

Competitive gaming, particularly in popular titles like "Clash Royale," is a landscape of constant evolution and strategy. Players continuously seek new methods to gain an edge over their opponents but are often hindered by the unpredictable nature of these games. Recognizing this challenge, our project is motivated by the ambition to elevate the player's experience through technology.

### Project Goal

We aim to develop a sophisticated tool that utilizes modern machine learning techniques to predict outcomes of "Clash Royale" matches. This tool aims to:

1. **Predict Match Outcomes**: Offer players a reliable forecast of game results, transforming how they approach each match.
2. **Enhance Strategic Planning**: Provide players with data-driven insights to refine gameplay strategies, moving beyond intuition to informed decision-making.
3. **Create Intuitive Visualizations**: Develop an interface that translates complex data into understandable and actionable information, making advanced analytics accessible to all players.

Significance for Players and Developers

The benefits of this project extend to various aspects of the gaming ecosystem:

1. **Improved Player Experience**: By equipping players with foresight into game outcomes, we aim to enhance their engagement and enjoyment.
2. **Informed Game Development**: Offer game developers like Supercell valuable insights for balancing gameplay and improving overall game design.
3. **New Economic Opportunities**: Explore the potential for novel revenue streams, such as betting platforms based on predictive models, adding an exciting dimension to the gaming experience.

## Problem Definition

**Jargon-Free Objective (Heilmeier's Question #1)**

What are we trying to do? In simple terms, we are building a tool that acts like a weather forecast, but instead of predicting rain or sunshine, it predicts who is likely to win or lose in a "Clash Royale" game. This tool will help players make better decisions during their games and better understand their chances of winning. For the game's creators, it is like giving them a map that shows where the game is balanced and where it might need some adjustments.

**Formal Problem Definition**

The project is centered on developing a predictive analytics model for "Clash Royale," a popular mobile strategy game. The model aims to:

1. Accurately forecast the outcomes of matches based on historical and real-time gameplay data.
2. Utilize machine learning algorithms to analyze and interpret complex datasets, identifying patterns and factors that significantly influence game results.
3. Provide a user-friendly visualization interface for players, translating the analytical predictions into easily understandable and actionable insights.
4. Aid game developers in identifying balance issues within the game mechanics, contributing to a more equitable gaming environment.
5. Investigate the feasibility of integrating predictive models into new business models, such as in-game betting platforms, enhancing player engagement and potential revenue channels.

## Literature Surveys

1. Research presents a cluster regression analysis model to score basketball players' performance based on past games. Introduces a data-driven method for player evaluation, using clustering techniques for enhanced predictive accuracy. Limited datasets and basketball-specific approaches may hamper broader applicability; diverse datasets and advanced techniques could enrich future insights.

2. Highlighting big data's transformative impact on the online gaming industry, particularly its role in driving mobile gaming revenues and its strategic importance for companies like Microsoft. Emphasizes big data's potential to decode user behaviors, improve game design, personalize advertising, and support the freemium business model. Lacks a detailed exploration of specific extensive data methodologies or techniques, focusing more on high-level industry trends.
3. The article accentuates the role of analytics in mobile game development, offering tools and strategies to optimize game mechanics and enhance player experience. Provides practical insights into how analytics can assist developers in understanding player behavior, crafting personalized content, and refining game designs, with case studies like "Crossy Road" and "Candy Crush Saga." Though it discusses ethical considerations, it may not delve deeply into potential challenges or limitations of the mentioned analytic tools.
4. Application of DNN and RNN deep learning algorithms to predict Premier League football match outcomes, highlighting RNN with LSTM cells as the most accurate. Introduces advanced deep learning techniques and model comparisons for predicting sports outcomes. Limited to football; applicability to other contexts or games remains uncertain.
5. Prediction of NFL game winners using machine learning and deep learning, considering factors like team statistics, weather, injuries, and betting odds; deep learning methods prove superior. Provides a comprehensive set of features for game prediction and demonstrates the strength of deep learning in such predictions. Analysis is limited to football games and may not apply to Clash Royale.
6. Predicting outcomes of ranked matches in League of Legends using a machine learning model centered on a deep neural network, emphasizing player-champion experience as the key feature. The topics discussed are relevant because they are tailored to online gaming predictions and introduce the concept of player-character proficiency as a predictive factor. The specificity of the player-champion experience feature may limit the model's generalizability outside League of Legends.
7. A deep learning framework for football match outcome predictions, integrating various features and leveraging both deep neural networks and artificial neural networks for model construction. Demonstrates a multifaceted deep learning approach, which might be adaptable to different sports or game prediction challenges. May require modifications for broader applications.
8. Predicting Clash Royale deck win rates using an SVR (Support Vector Regression) model and a clustered greedy selection strategy. Provides a targeted model specific to Clash Royale. Relies solely on SVR; other potentially more suitable models might be overlooked.
9. Basketball game outcomes are predicted using various models; Naïve Bayes has a 70.5% accuracy rate. Offers a diverse set of modeling ideas potentially adaptable to Clash Royale. Approach might not apply to Clash Royale.
10. Importance of model choice and validation in basketball game outcome predictions, with KNN and cross-validation standing out. The document imparts valuable insights on modeling and validation techniques. Basketball-centric methods might not be directly transferable to Clash Royale's context.
11. MOBA games' win prediction after 5 minutes of live gameplay reaches an 85% accuracy mark. Establishes a high accuracy benchmark for live gameplay predictions in Clash Royale. Not immediately applicable because the team will not perform real-time analysis.
12. Analysis of gameplay through the lens of risks, rewards, and affordances to decode game balance and player proficiency. Endorses the notion of integrating prediction-challenging mechanics for enriching player experience in Clash Royale. Perceived affordances might be challenging to test empirically.
13. MLB game outcomes predicted using multiple methods. Introduces simple, potentially adaptable models for Clash Royale, setting an accuracy goal. Limited dataset and potential inapplicability due to MLB's differing dynamics compared to Clash Royale.

14. Analysis of basketball's offensive tactics using graph-based link prediction focusing on ball-passing chains. Offers insights into tactical strategy that could be translated into game mechanics or AI strategy. Data collection could be arduous, and basketball's tactical dynamics might not correspond with Clash Royale's.
15. A method to predict when players might quit mobile games, utilizing survival analysis. Presents a comprehensive churn prediction approach crucial for game longevity and monetization. Defining churn is challenging without complete logging, and survival analysis can get intricate.
16. Classifying player behaviors in mobile games using clustering based on behavioral data. Understanding player behaviors can inform game updates and event strategies for enhanced player retention. The method's specificity to two mobile games might hamper its broad applicability.
17. The article presents a supervised learning technique, specifically Random Forest, to predict the outcome of upcoming NCAA Basketball games, also detailing evaluation metrics for the model. Provides insights into the application of supervised learning for predicting basketball game outcomes and introduces relevant model evaluation metrics. Focuses solely on Random Forest, neglecting potential exploration of other machine learning models or techniques that could offer better results.
18. The study introduces an enhanced sports outcome prediction process by incorporating adaptive weighted features and machine learning algorithms for basketball score forecasting. Offers a comprehensive method that integrates adaptive feature engineering with multiple machine learning models, leading to improved prediction results. The method may be too complex to implement.

## Proposed Method

### Intuition

Our approach stands out from the state of the art due to its unique blend of advanced machine learning techniques tailored specifically for the dynamic environment of "Clash Royale." Unlike conventional methods that might rely on static or generic models, our solution delves into the intricacies of player behavior and game mechanics, utilizing various algorithms like XGBoost, logistic regression, and deep neural networks. This multifaceted approach increases the robustness and accuracy of our predictions and ensures adaptability to the game's ever-evolving strategies. Furthermore, our innovative use of interactive visualizations makes complex data insights accessible to all players, enhancing their decision-making process. Our project transcends traditional boundaries by combining innovative technology with a deep understanding of the gaming landscape, promising a more accurate, dynamic, and user-friendly predictive tool for "Clash Royale."

### Detailed Description of Approaches

Our project employs a comprehensive approach that combines several advanced algorithms and user-friendly interfaces to predict "Clash Royale" game outcomes. Here is a detailed description of our methodologies:

### Algorithms Used

**XGBoost**: At the core of our model is XGBoost, a robust machine-learning algorithm known for its efficiency and effectiveness. It incorporates L1 and L2 regularization to prevent overfitting, ensuring our model generalizes well to new, unseen data. This feature is crucial for accurately predicting game outcomes in the constantly evolving landscape of "Clash Royale."

**Logistic Regression**: We have utilized logistic regression to assign probabilities to potential game outcomes. This algorithm is handy for its simplicity and interpretability, providing clear insights into how varied factors weigh in on the game's result.

**Deep Neural Networks (DNN)**: We have employed deep neural networks to capture the complex, non-linear relationships in the game data. These networks excel in identifying intricate patterns and have shown remarkable performance, especially in classifying winners and losers based on diverse gameplay elements.

**TabNet**: This novel deep learning model, explicitly designed for tabular data, uses attention mechanisms to focus on the most relevant features for predicting game outcomes. TabNet's unique approach is well-suited to handle the multifaceted nature of our dataset.
**Decision Trees and SGD (Stochastic Gradient Descent) Classification**: As our baseline model, decision trees segment data into branches for clear outcomes, while SGD classification optimizes decision-making through iterative updates. These models provide a foundational understanding and comparison point for the more complex algorithms.

### User Interface

**Interactive Visualization Tool using Power BI**: An essential aspect of our approach is making complex data insights accessible to all players. We have developed an interactive visualization interface that translates analytical predictions into easy-to-understand formats. This tool allows players to quickly grasp their odds of winning and adjust their strategies accordingly.
**Data Presentation using Power BI**: The interface presents data in various formats, including graphs, charts, and heatmaps, to accommodate different user preferences and enhance the interpretability of results.
Customizable Dashboards: Players can customize their dashboards to focus on the metrics most relevant to them, whether it's historical performance, card usage statistics, or opponent analysis.
**Accessibility and Ease of Use**: The interface focuses on user experience, ensuring that even those without a technical background can easily navigate and benefit from the tool.
In conclusion, our approach marries sophisticated algorithms with a user-centric interface, providing a robust, accurate, and accessible predictive tool for "Clash Royale" players and developers. This blend of advanced analytics and practical application sets our project apart in competitive gaming technology.

## Experiments/Evaluation

### Description of the Testbed

Our testbed for evaluating the "Clash Royale" predictive model is a carefully structured environment designed to assess the performance and accuracy of our algorithms. It consists of the following elements:
**Dataset**: The primary component of our testbed is a comprehensive dataset obtained via the Clash Royale API. This dataset includes over 100,000 records, encompassing detailed gameplay metrics such as player historical rankings, win/loss records, battle times, arenas, game modes, and card usage.
**Computational Resources**: We utilize a high-performance computing setup capable of handling large datasets and running complex machine-learning algorithms. This includes powerful CPUs, sufficient RAM, and GPUs for deep learning tasks.
**Development Environment**: The models are developed and tested using Python, leveraging libraries like Pandas for data manipulation, Scikit-learn for machine learning models, and TensorFlow/Keras for deep learning models.
**Evaluation Metrics**: The testbed includes tools and scripts to measure various evaluation metrics like accuracy, precision, recall, and AUC/ROC, providing a comprehensive analysis of each model's performance.

### Questions Addressed by the Experiments

Our experiments were designed to answer several critical questions, including:
**Accuracy of Predictions**: How accurately can each model predict the outcomes of "Clash Royale" matches? This is measured by the percentage of predictions that match the actual outcomes.
**Feature Relevance**: Which features (e.g., player rankings, card choices, game mode) most significantly impact the model's predictions? Understanding this helps in fine-tuning the models for better accuracy.
**Model Comparisons**: How do different models (XGBoost, logistic regression, DNN, etc.) compare in terms of predictive performance? This comparison helps in selecting the best model or a combination of models.
By answering these questions, our experiments aim to validate our predictive model's effectiveness, reliability, and user-friendliness, ensuring it meets the needs of "Clash Royale" players and developers.

## Detailed Description of the Experiments

Our experimental setup was meticulously designed to evaluate the performance of our predictive models for "Clash Royale." Here's a breakdown of the experiments and the critical observations:

**Experiment 1: Model Accuracy and Performance**

- **Objective**: To assess the accuracy and performance of different machine learning models, including XGBoost, logistic regression, deep neural networks, TabNet, decision trees, and SGD classification.
- **Methodology**: We divided our dataset into training and testing sets, ensuring a representative distribution of game scenarios. Each model was trained on the same dataset and evaluated on the test set.
- **Observations**:
    - XGBoost yielded the highest accuracy at 73%, demonstrating its effectiveness in handling complex datasets.
    - The deep neural network followed closely with 71% accuracy, excelling in pattern recognition.
    - Logistic regression, while less accurate at 65%, provided valuable insights into the importance of features.
    - Decision trees and SGD classification served as solid baseline models, achieving around 63% accuracy.
    - TabNet's performance at 69% highlighted the utility of attention mechanisms in tabular data.

**Experiment 2: Feature Relevance Analysis**

- **Objective**: To identify the most influential features in determining match outcomes.
- **Methodology**: We used feature importance evaluation tools available in XGBoost and logistic regression models and analyzed the weights in neural networks. We also used Partial Dependence Analysis tool to understand the relationship between a feature and the predicted outcome.
- **Observations**:
    - Player historical rankings, specific card usage, and game mode were among the most influential features.
    - Unusual patterns, like certain card combinations, also significantly influenced predictions.

**Experiment 3: Scalability and Performance**

- **Objective**: To test the scalability of the models with increasing data volumes.
- **Methodology**: The models were subjected to varying-sized datasets to observe their performance and resource utilization.
- **Observations**:
    - XGBoost and DNNs demonstrated robust scalability, maintaining high accuracy even with increased data sizes.
    - Logistic regression showed excellent performance with minimal computational resource requirements, making it suitable for lower-end devices.

**Summary of Key Findings**

- **Model Efficacy:** XGBoost emerged as the most effective model, closely followed by DNNs.
- **Feature Insights**: Player behavior and specific in-game choices are critical in determining match outcomes.
- **Real-Time Application**: Models showed promising potential for real-time predictions, a crucial aspect for in-game strategy adjustments. We could tune the model to adjust for real-time predictions with additional work.
- **User Interface**: The visualization tool successfully translated complex data into accessible insights, enhancing user experience.
- **Scalability and Performance**: Advanced models displayed robust scalability and performance, indicating their suitability for large-scale deployment.

**Addressing Limitations and Challenges**

- **Limitations in Data and Model Biases**: Our experiments faced limitations in data diversity and potential model biases. Future iterations could incorporate more diverse datasets and explore methods to mitigate biases.
- **Computational Constraints**: Some models and model's hyper-parameters fine tuning faced computational constraints, suggesting a need for more efficient algorithms or optimized computing resources.

**Deeper Analysis and Implications**

- **Impact on Gaming Analytics**: Our findings will hopefully advance the field of predictive analytics in gaming, providing new avenues for game strategy development and enhanced player experiences.
- **Contextualization with Existing Studies**: Compared to existing predictive tools, our models show improved accuracy and adaptability.
- **Broader Implications**: The project's implications extend to game design and the broader gaming industry, suggesting new methods for engaging players and enhancing game dynamics.

These experiments and observations collectively validate our models' effectiveness and the user interface's practicality, indicating a significant advancement in predictive analytics for "Clash Royale."

## Conclusions and Discussions:

### Project Summary

Our project successfully developed a predictive analytics tool for "Clash Royale," leveraging advanced machine learning models to forecast game outcomes. The most effective model, XGBoost, achieved an impressive 73% accuracy, closely followed by deep neural networks. These models provided critical insights into game dynamics, significantly enhancing player strategies and decision-making. Complementing the analytical prowess, our interactive visualization interface made complex data accessible to players of all skill levels, democratizing advanced game analytics.

Despite these advancements, the project faced limitations, particularly in real-time prediction capabilities, which restricted its use in pre-match analyses. Future work could integrate real-time predictions, refine model accuracy, and broaden the dataset to encompass a broader range of game scenarios. This evolution has the potential to further transform player experiences in "Clash Royale," offering real-time strategic guidance and paving the way for new interactive gaming dimensions.

### Group Efforts:

Each team member played a pivotal role in our project, contributing equally to various aspects of the work. We all shared responsibilities in data extraction, project management, write-ups, and the development of visualizations, ensuring a balanced distribution of effort across these critical areas. While the collective action was uniform, it is noteworthy that Mai distinguished herself by developing the most effective model, which significantly enhanced the project's overall success. Her contribution in this area was particularly remarkable and deserves special recognition. However, this achievement does not diminish the equally important contributions made by every team member in their respective roles.

## Bibliography:

1. Ashley. "Unleashing the Power of Analytics: Elevating Mobile Game Design and Performance." AppSamurai Blog, May 22, 2023. https://appsamurai.com/blog/unleashing-the-power-of-analytics-elevating-mobile-game-design-and-performance/

2. Rands, Kevin. "How Big Data is Disrupting the Gaming Industry." CIO, January 26, 2018. https://www.cio.com/article/228327/how-big-data-is-disrupting-the-gaming-industry.html
3. Xu, Jin, and Chao Yi. "The Scoring Mechanism of Players after Game Based on Cluster Regression Analysis Model." Mathematical Problems in Engineering, vol. 2021, Article ID 5524076. Published March 13, 2021. https://doi.org/10.1155/2021/5524076
4. Ahmed Amr Awadallah and Raghav Khandelwal, "Football Match Prediction using Deep Learning (Recurrent Neural Network)," Stanford University, 2020.
5. Pablo Bosch, "Predicting the winner of NFL-games using Machine and Deep Learning," Vrije Universiteit Amsterdam, 2018.
6. Zhenyu Zhang, Yifan Zhang, and Zhiyong Feng, "Using Machine Learning to Predict Game Outcomes Based on Player-Champion Experience," arXiv preprint arXiv:2108.02799, 2021.
7. Md. Rafsanjani Muin, Md. Rafiul Hassan, Md. Ashraful Alam Khan, and Md. Kamrul Hasan, "A deep learning framework for football match prediction," SN Applied Sciences 2, no. 1 (2020): 1-13
8. Cen, L., Q. H. Vu, and D. Ruta. "Efficient Support Vector Regression with Reduced Training Data." In 2019 Federated Conference on Computer Science and Information Systems (FedCSIS), Leipzig, Germany, 2019, pp. 15-18. doi: 10.15439/2019F362.
9. Alonso, Roger Poch, and Marina Bagić Babac. "Machine Learning Approach to Predicting a Basketball Game Outcome." International Journal of Data Science 7, no. 1 (2022): 60-77. doi: 10.1504/IJDS.2022.124356.
10. Horvat, Tomislav, Ladislav Havaš, and Dunja Srpak. 2020. "The Impact of Selecting a Validation Method in Machine Learning on Predicting Basketball Game Outcomes" Symmetry 12, no. 3: 431. https://doi.org/10.3390/sym12030431
11. Hodge, Victoria J. et al. "Win Prediction in Multi-Player Esports: Live Professional Match Prediction." *IEEE Transactions on Games* (In Press, Accepted: 2019). ISSN 2475-1510.
12. Pereira de Moraes Neto, Ernando. "Affordances and Structures of Risks and Rewards: A Case Study of Clash Royale." *SBC – Proceedings of SBGames 2017*, Art & Design Track – Short Papers, ISSN: 2179-2259.
13. Li, Shu-Fen, Mei-Ling Huang, and Yun-Zhi Li. "Exploring and Selecting Features to Predict the Next Outcomes of MLB Games." Entropy (Basel) 24, no. 2 (2022): 288. Published online 2022 Feb 17. doi: 10.3390/e24020288. PMCID: PMC8871522. PMID: 35205582. Academic Editors: Andrea Prati, Carlos A. Iglesias, Luis Javier García Villalba, and Vincent A. Cicirello.
14. Zhang, Tao, Gongzhu Hu, and Qi Liao. "Analysis of Offense Tactics of Basketball Games Using Link Prediction." IEEE Xplore. Accessed September 30, 2023. https://ieeexplore.ieee.org/document/6607842.
15. Bertens, Paul, Anna Guitart, and África Periáñez. "Games and Big Data: A Scalable Multi-Dimensional Churn Prediction Model." In 2017 IEEE Conference on Computational Intelligence and Games (CIG). IEEE, 2017. Accessed September 30, 2023. doi: 10.1109/CIG.2017.8080412.
16. Kwon, Hyoungjin, Wooyoung Jeong, Dae-Wook Kim, and Seong-Il Yang. "Clustering Player Behavioral Data and Improving Performance of Churn Prediction from Mobile Game." In 2018 International Conference on Information and Communication Technology Convergence (ICTC). IEEE, 2018. Accessed September 30, 2023. doi: 10.1109/ICTC.2018.8539514.
17. Sasson, Joe. "Machine Learning | Sports Predictions." Analytics Vidhya, February 2, 2021. Accessed October 5, 2023. URL: https://medium.com/analytics-vidhya/machine-learning-sports-predictions-f01c1736ef73.

18. Lee, Fan. "Deep Learning in Sports Prediction." Trinity University, 2022. Accessed October 5, 2023. URL: https://digitalcommons.trinity.edu/compsci_honors/65.