Team 14 aka Team KARMMA's Poster.

Predicting Clash Royale Match Outcomes: A Machine Learning Approach

Motivation/Introduction

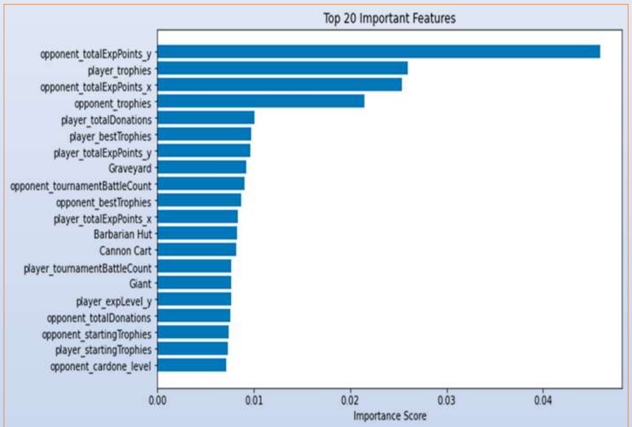
1. **The Problem:**
    1. *In 'Clash Royale,' players cannot predict the outcome of matches.*
    2. This leads to a trial-and-error approach, affecting their gaming experience and strategy development. This project aims to utilize the power of modern machine learning to create a model that helps players develop new strategies, predict game outcomes, and create an easily digestible visualization of these predictive outcomes.

2.Importance for Players:

1. *Predicting match outcomes can significantly enhance player strategy and decision-making capabilities*
2. It provides an edge in gameplay, helping players to refine their skills and approaches based on data-driven insights.

3.Broader Significance:

1. *Predictive capabilities are crucial for both player engagement and game evolution.*
2. For the developers, it offers insights for balancing the game, while for the community, it adds a deeper, more analytical aspect to the gaming experience.
3. The company also provides future revenue streams where they can create a betting/gambling platform where players get rewarded more when they beat the algorithms' predictions. Other players can bet on games about to start.



Who Won	%GT Number Of Battles	Number Of Battles
Player 1 Won	51.97%	51,974
Player 2 Won	48.03%	48,030
Total	**100.00%**	**100,004**

Team Approaches (Algorithm and Interactive Visualization)

What are they?

- *The team used interactive visualization to understand the dynamics of the games better and gather basic statistics on how commonly certain phenomena occurred. E.g., We learned that on average 92K players only win 48.6% of the time.*
- *The team utilized a range of ML algorithms, including decision trees, logistic regression models, SGD classification, dense neural networks (MLP), TabNet and XGBoost*

•How do they work?

- *Decision trees segment data into branches, leading to conclusive outcomes, while SGD classification optimizes decision-making through iterative updates.*
- *Dense neural networks analyze complex, non-linear relationships in data, and TabNet employs attention mechanisms to focus on relevant features.*
- *Logistic Regression assigns a win probability based on the combination of factors associated with each game. This probability is rounded to a 1 for values 0.5 and above and 0 otherwise.*
- *XGBoost incorporates L1 (LASSO) and L2 (ridge) regularization terms in its objective function. This helps prevent overfitting and improves the model's generalization to new, unseen data*

•Effectiveness for Solving the Problem:

- *The varied nature of these algorithms allows for a comprehensive analysis of player data, capturing different aspects of gameplay strategies.*
- *This multi-faceted approach increases the accuracy of predicting match outcomes, as evidenced by the improved accuracy scores in our tests.*
- *The most accurate approach was utilizing an XGBoost model after the team could identify the most essential features.*

•Innovations in the Approach:

- *Our innovation lies in combining diverse, cutting-edge ML techniques adapted explicitly for gaming data, a relatively unexplored application area.*
- *The interactive visualization tool is a unique addition, transforming complex analytical results into accessible insights for players.*

This shows us that games are competitive, as 67% of the games are only won by 1 crown.

Data Acquisition

•**Method of Collection:** Utilized the Clash Royale API to obtain data, employing a Python script with the requests library to scrape various data points.

•**Data Integration:** Merged data from the battle log, card data, and player rankings libraries into a Pandas DataFrame, ensuring a comprehensive dataset for analysis.

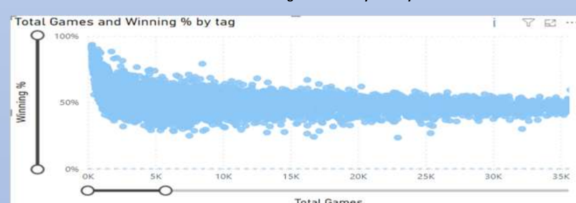
Data Characteristics

•**Volume and Structure:** We collected hundreds of thousands of rows over the experiment, culminating in a dataset with 100,000 records (139MB) and 100 features we had to filter through.

•**Content Details:** The dataset features granular details such as battle time, arena, game mode, player names, crowns indicating wins, and cards used.

•**Player-Specific Data:** Included player historical rankings, total wins, and losses, integrating this data with battle features for each player and their opponents.

•**Data Preparation:** Performed data cleaning and one-hot encoding to convert string-based data into a numerical format suitable for machine learning, focusing on predicting player win likelihood based on these features.



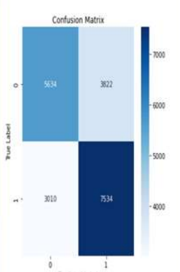
Shows that the more games a player plays, they level out at about a 50%-win rate, which demonstrates the balance of the game.

Logistic Regression Results

Logistic Regression Results

# of Correct Predictions	# of Records	% Accuracy
13,168	20,000	65.84%

Logistic Regression Results

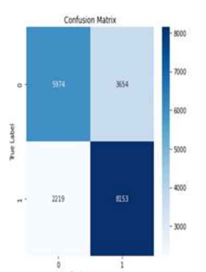


Dense Neural Network Results

Dense Neural Network Results

# of Correct Predictions	# of Records	% Accuracy
14,127	20,000	70.64%

Dense Neural Network Results

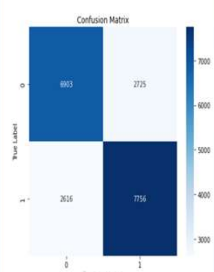


XGBoost Results

XGBoost Results

# of Correct Predictions	# of records	% Accuracy
14,659	20,000	73.30%

XGBoost Confusion Matrix



Experiments and Results

Evaluation of Approaches

•**Metric Selection:** Employed metrics such as accuracy, precision, AUC/ROC, and recall to assess the performance of various ML algorithms.

•**Defining Success:** These metrics quantified the predictive success, with accuracy indicating overall prediction correctness and AUC gauging the model's discriminative ability.

•**Precision and Recall Metrics:** Provided insights into the positive predictive value and the model's sensitivity to detect actual positives.

Results of the Experiments

•**XGBoost:** Achieved the highest accuracy at 73%,

•**Dense Neural Network Outcome:** Achieved an accuracy of 71%, demonstrating its strength in complex pattern recognition and classification.

•**Baseline Model Results:** Decision trees and SGD classification served as baseline comparisons, each achieving an accuracy of 63%.

•**TabNet's Performance:** Showcased the utility of attention mechanisms in tabular data with 69% accuracy.

•**Logistic Model Performance:** Achieved an accuracy of 65% with the logistic regression model.

Comparative Analysis

•**Model Superiority:** XGBoost outperformed all the other models, which we recommend integrating.

•**Foundational vs. Advanced Models:** Demonstrated the increased effectiveness of the XGBoost/dense neural network over simpler models like decision trees and SGD classification.

•**Innovative Approach:** Highlighted the strategic advantage of using complex models for predictive analytics in gaming data, contributing to a more nuanced understanding of game dynamics.