Analysis and Prediction of PUBG Player Behavior

Yuheng Zhu*
yzhu63@ncsu.edu
NC State University
Raleigh, North Carolina, USA

Mengzhe Wang*
mwang39@ncsu.edu
NC State University
Raleigh, North Carolina, USA

Jiayuan Huang*
jhuang52@ncsu.edu
NC State University
Raleigh, North Carolina, USA

shuai Chen*
schen76@ncsu.edu
NC State University
Raleigh, North Carolina, USA

1 Dataset Description

https://www.kaggle.com/datasets/skihikingkevin/pubg-match-deaths

1.1 PlayerUnknown's Battlegrounds

PUBG is a first/third-person shooter battle royale style game that matches over 90 players on a large island where teams and players fight to the death until one remains. Players are airdropped from an airplane onto the island where they are to scavenge towns and buildings for weapons, ammo, armor and first-aid. Players will then decide to either fight or hide with the ultimate goal of being the last one standing.

1.2 Dataset

This dataset provides two zips: aggregate and deaths.

- In deaths, the files record every death that occurred within the 720k matches. That is, each row documents an event where a player has died in the match.
- In aggregate, each match's meta information and player statistics are summarized (as provided by pubg). It includes various aggregate statistics such as player kills, damage, distance walked, etc as well as metadata on the match itself such as queue size, fpp/tpp, date, etc.

The uncompressed data is divided into 5 chunks of approximately 2gb each. For details on columns, please see the file descriptions.

2 Project Idea

PUBG is a popular online multiplayer game that involves up to 100 players fighting against each other in a shrinking map until only one player or team remains. However, although the number of players is considerable, not much is known about what factors influence the outcome of a match and how players can improve their chances of winning. Besides, the existence of cheaters heavily hurts the experience of normal players. This project aims to identify cheaters, as well as analyze and give specific suggestions to players that lead to the highest probablity of winning based on their statistics, such as kills, assists, damage, survival time, etc., using machine learning techniques.

We plan to first perform exploratory data analysis on the PUBG dataset to gain a better understanding of player behavior and gameplay patterns. This may include analyzing factors such as player rankings, game modes, and weapon usage. We will also investigate correlations between different attributes in the data to identify any useful patterns.

To identify cheaters, we plan to use cluster models to find out outliers who are relevant to strange behavors. As for game suggestions, we plan to use machine learning algorithms such as Random Forests and Support Vector Machine to analyze player data and detect anomalous behavior. We will compare the performance of different classification models and feature representation techniques, such as one-hot encoding and word embeddings, to determine the most effective approach. Besides, we also plan to use the data to predict recommended strategies for gamers in the game. This may involve analyzing successful strategies used by top players and using this information to recommend tactics for other players to improve their performance.

In our vision, there will be basically two majors of suggestions from the model: relative-player-objective map scene analysis and prediction, and relative-player-subjective player behavior suggestions. For example. the former will give the players the approximate area of poison circle. The latter, on the contrary, will give personalized current game tactic suggestions, such as whether the player should go to battle or search for supplies and equipments.

Overall, this project aims to use machine learning techniques to analyze player behavior in PUBG, and predict recommended strategies for gamers. By doing so, we hope to contribute to a fair and enjoyable gaming experience for all players.

3 Software Specification

- Visual Studio Code
- Python3
- Jupyter

---

*Both authors contributed equally to this research.

	Data Prepossessing	Model Construction & Evaluation	Output Analyze and Explanation
Yuheng Zhu [yzhu63]	O	O	X
Mengzhe Wang [mwang39]	X	O	O
Jiayuan Huang [jhuang52]	O	X	O
Shuai Chen[schen76]	X	X	O

Table 1. Work Division

- PyTorch
- Scikit-Learn
- Pandas
- TensorFlow
- Kera
- Matplotlib

4 Work Division

See Table 1.

5 Paper to Read

1. Ghazali N F, Sanat N, As' ari M A. Esports Analytics on PlayerUnknown's Battlegrounds Player Placement Prediction using Machine Learning[J]. International Journal of Human and Technology Interaction (IJHaTI), 2021, 5(1): 17-28.
2. Rokad B, Karumudi T, Acharya O, et al. Survival of the Fittest in PlayerUnknown BattleGround[J]. arXiv preprint arXiv:1905.06052, 2019.
3. Indulkar Y. PUBG Winner Ranking Prediction using R Interface 'h2o'Scalable Machine Learning Platform[C] //2021 International Conference on Emerging Smart Computing and Informatics (ESCI). IEEE, 2021: 300-305.
4. Mamulpet M M. Pubg Winner Placement Prediction Using Artificial Neural Network[J]. Int. J. Eng. Appl. Sci. Technol, 2019, 3(12): 107-118.
5. Gałka P, Strzelecki A. How randomness affects player ability to predict the chance to win at PlayerUnknown's Battlegrounds (PUBG)[J]. The Computer Games Journal, 2021, 10: 1-18.

6 Midterm Milestone

The midterm milestones have been divided into above 3 parts in Work Division. In our assumption, by the midterm, the entire first section "Data Prepossessing" and half of second section "Model Construction & Evaluation" should be done. To be more specific, the model should be able to give readable and explanable result based on data, although which may not be useful enough for the purpose of this project.