PUBG Final Placement Prediction

Jiayi Liu Yimin Xu Zeyu Liu

Introduction

Objectives for today:

- Introduction of game PUBG.
- Feature selection.
- Random Forest and LightGBM.
- Final result.

Introduction of game PUBG

PUBG rules introduction

- A game about survival of the fittest.
- Can enter with solo, duo or quadrant team.
- Search for weapons heals and boosts to ensure survival from the field and even from others.
- Survive as long as possible to win the game.

Data Set Description

- Data Input: Contain 28 features and 4446966 observations.
- Data Output: Indicate the final placement of the PUBG game(1 denotes the first place while 0 denotes the last)

Output data Distribution

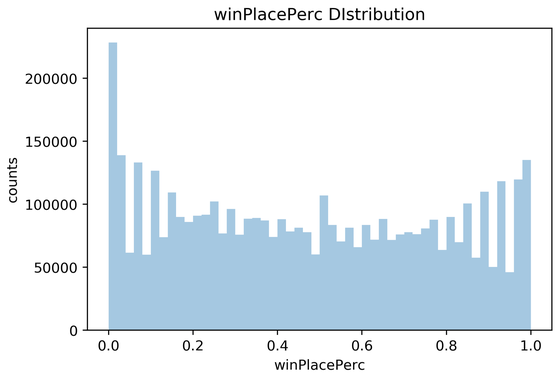

Figure 1: Final Placement D

Feature Selection

For such a big data set, it will be definitely difficult to do the training with all the features. Simultaneously, some of the features themselves have no correlation with the final placement. Hence, doing prediction with all features will not only cause more time but also make the result unconvincing. To drop out that meaningless features to the training data, we plot a heat map for the correlation between the different features and the result:

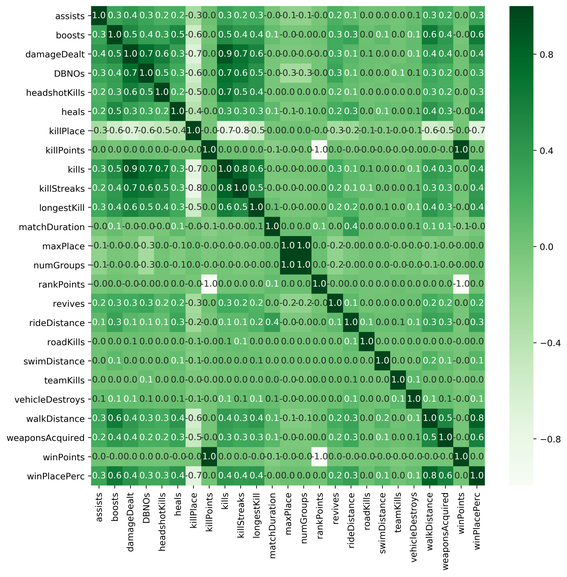

Figure 2: correlation heatmap

From the correlation map, it is easy to drop the features which have 0 correlation with the final placement. Then we transform data from shape $(4446966, 28)$ to $(4446996, 16)$ which becomes more convincing to be processed.

Random Forest LightGBM

- **Random Forest**

The training algorithm for random forests applies the general technique of bootstrap aggregating, or bagging, to tree learners. Given a training set $X = x_1, ..., x_n$ with responses $Y = y_1, ..., y_n$, bagging repeatedly (B times) selects a random sample with replacement of the training set and fits trees to these samples:
For $b = 1, ..., B$:

1. Sample, with replacement, n training examples from X, Y; call these X_b, Y_b.
2. Train a classification or regression tree f_b on X_b, Y_b.

After training, predictions for unseen samples x' can be made by averaging the predictions from all the individual regression trees on x':

$$\hat{f} = \frac{1}{B}\sum_{b=1}^{B} f_b(x')$$

And a pattern of random forest is shown as following:

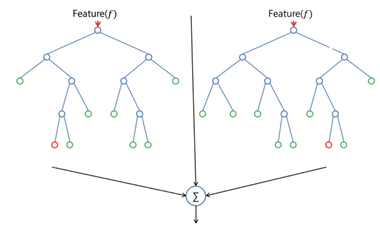

Figure 3: Random Forest Pattern[1]

- **LightGBM**[2]

A newly published algorithm to deal with a large data set. A fast, distributed, high performance gradient boosting (GBDT, GBRT, GBM or MART) framework based on decision tree algorithms, used for ranking, classification and many other machine learning tasks.

Final Result

- **Error rate**

The distribution of the error rate is shown as following:

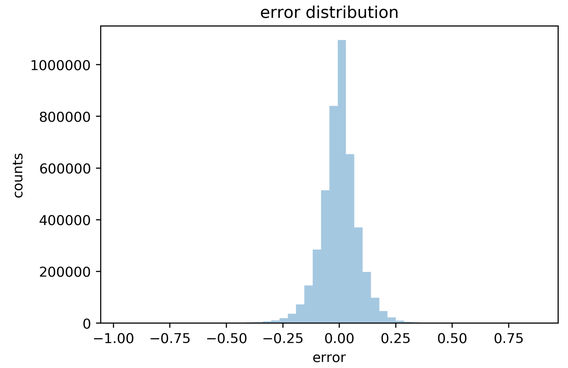

Figure 4: Distribution of error(mean error is 0.085)

- **Strategy for winning**

To have more chance to win in PUBG, it may be wiser to focus on features with more importance. To evaluate the importance of different features, a figure is derived as following:

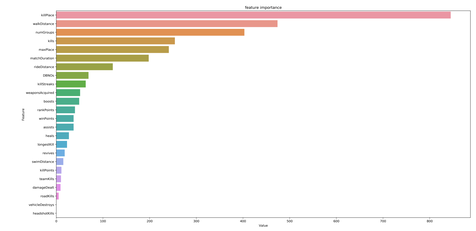

Figure 5: Feature importance

From the feature importance figure, it is obvious that the more players kill in game, the more probability they can win in the final.

Reference

[1] Donges, Niklas, and Niklas Donges."The Random Forest Algorithm." Towards Data Science, Towards Data Science, 22 Feb. 2018, towardsdatascience.com/the-random-forest-algorithm-d457d499ffcd.

[2] Ke, Guolin, et al. "Lightgbm: A highly efficient gradient boosting decision tree." Advances in Neural Information Processing Systems. 2017.