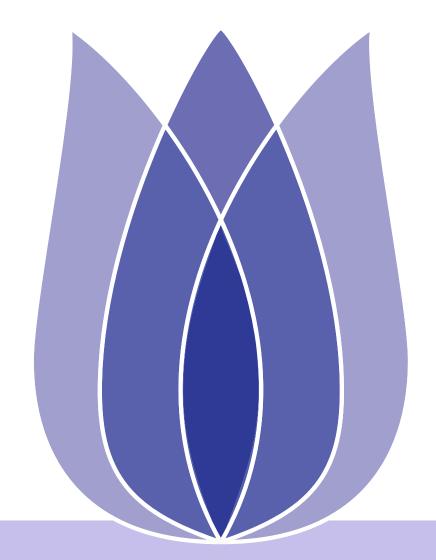
PUBG Game Data Analysis and prediction

Yang Cao

Deakin University

July 5, 2021





Overview

Problem Definition

Data Preprocess

Data Visualization

Feature selection

Modeling and Forecasting

Comparison and Conclusion

Problem Definition

Problem introduction

Data Preprocess

Data description

Dataset Description

Miss vlaue and NaN value

Data Visualization

Feature selection

Modeling and Forecasting

Model Selection

Comparison and Conclusion

Best Parameters

Comparison

Conclusion





Problem introduction

Data Preprocess

Data Visualization

Feature selection

Modeling and Forecasting

Comparison and Conclusion

Problem Definition





Problem introduction

Problem Definition

Problem introduction

Data Preprocess

Data Visualization

Feature selection

Modeling and Forecasting

Comparison and Conclusion

In a PUBG game, up to 100 players start in each match (matchId). Players can be on teams (groupId) which get ranked at the end of the game (winPlacePerc) based on how many other teams are still alive when they are eliminated. In game, players can pick up different munitions, revive downed-but-not-out (knocked) teammates, drive vehicles, swim, run, shoot, and experience all of the consequences — such as falling too far or running themselves over and eliminating themselves. Different game behaviors will lead to different final rankings, so the main purpose is to build a model to predicts players' finishing placement based on their final stats, on a scale from 1 (first place) to 0 (last place).

- A game team data analyst may be interested in the game actions that make game teams get higher rank than others.
- Players can also estimate their final ranking based on the current situation and make strategic decisions in advance (such as running away or fighting).



Description and Evaluation

Problem Definition

Problem introduction

Data Preprocess

Data Visualization

Feature selection

Modeling and Forecasting

Comparison and Conclusion

Mean Square Error: the average squared difference between the estimated values and the actual value

- Train dataset MSE
- Test Dataset MSE





Data Preprocess

Data description

Dataset Description

Miss vlaue and NaN value

Data Visualization

Feature selection

Modeling and Forecasting

Comparison and Conclusion

Data Preprocess





Data description

Problem Definition

Data Preprocess

Data description

Dataset Description

Miss vlaue and NaN value

Data Visualization

Feature selection

Modeling and Forecasting

Table 1: Data Field

Attributes	description
DBNOs	Number of enemy players knocked
Assists	Number of enemy players this player damaged that were killed by teammat
Boosts	Number of boost items used
damageDealt	Total damage dealt
headshotKills	Number of enemy players killed with headshots
heals	Number of healing items used
killPlace	Ranking in match of number of enemy players killed
killPoints	Kills-based external ranking of player
killStreaks	Max number of enemy players killed in a short amount of time
kills	Number of enemy players killed
longestKill	Longest distance between player and player killed at time of death
rankPoints	Elo-like ranking of player
revives	Number of times this player revived teammates





Dataset Description

Problem Definition

Data Preprocess

Data description

Dataset Description

Miss vlaue and NaN value

Data Visualization

Feature selection

Modeling and Forecasting

Comparison and Conclusion

- Train and Test Dataset
 - train_v2.csv
 - ◆ test_v2.csv

train_v2.csv

- There are 4446966 rows and 29 columns.
- ◆ 4446966 unique ID.
- ◆ 2026745 unique groupId

test_v2.csv

- ◆ There are 1934174 rows and 28 columns.
- ◆ 1934174 unique ID
- ◆ 886238 unique groupId





Miss vlaue and NaN value

Problem Definition

Data Preprocess

Data description

Dataset Description

Miss vlaue and NaN value

Data Visualization

Feature selection

Modeling and Forecasting

Id	0
groupId	0
matchId	0
assists	0
boosts	8
damageDealt	0
DBNOs	0
headshotKills	0
heals	0
killPlace	8
killPoints	8
kills	8
killStreaks	8
longestKill	0
matchDuration	0
matchType	0
maxPlace	0
numGroups	8
rankPoints	8
revives	8
rideDistance	0
roadKills	0
swimDistance	0
teamKills	0
vehicleDestroys	8
walkDistance	8
weaponsAcquired	8
winPoints	0
winPlacePerc	1
dtype: int64	



Data Preprocess

Data Visualization

Feature selection

Modeling and Forecasting

Comparison and Conclusion

Data Visualization





Game type proportion

Problem Definition

Data Preprocess

Data Visualization

Feature selection

Modeling and Forecasting



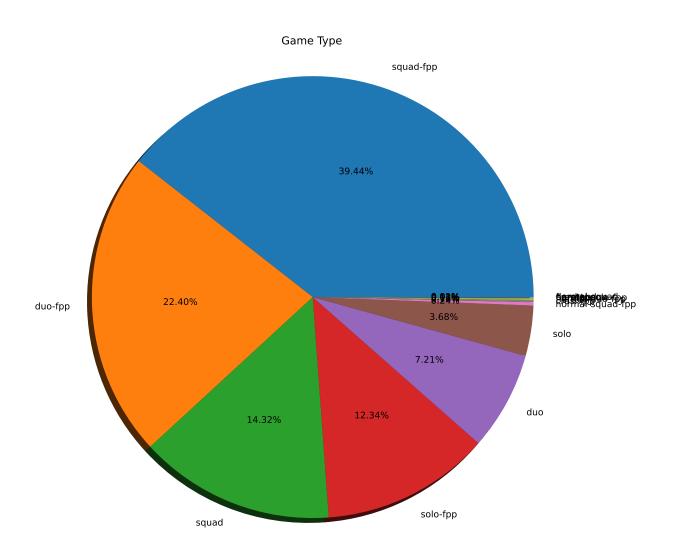


Figure 2: game type proportion



Walk distance with win place

Problem Definition

Data Preprocess

Data Visualization

Feature selection

Modeling and Forecasting

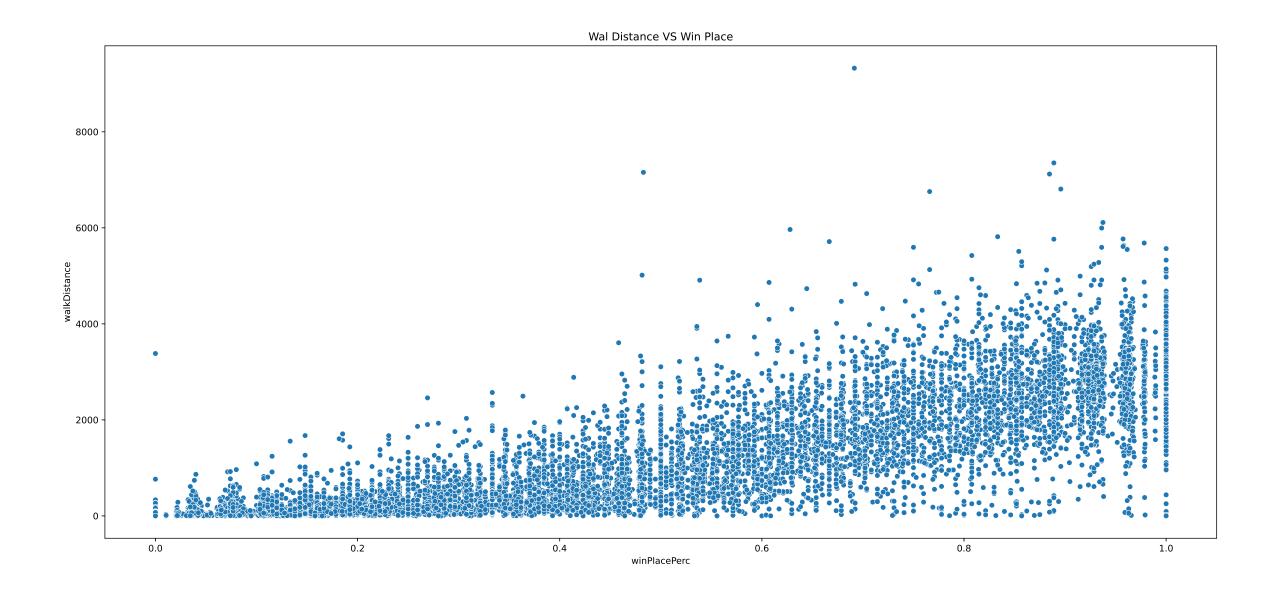


Figure 3: walking distance VS Win Place





Data Preprocess

Data Visualization

Feature selection

Modeling and Forecasting

Comparison and Conclusion

Feature selection





Correlation

Problem Definition

Data Preprocess

Data Visualization

Feature selection

Modeling and Forecasting

Comparison and Conclusion

assists -	1.00	0.31	0.41	0.30	0.20	0.23	-0.29	0.04	0.32	0.24	0.26	-0.02	-0.15	-0.15	-0.02	0.20	0.11	0.01	0.02	0.01	0.06	0.29	0.24	0.02	0.30
boosts -	0.31	1.00	0.52	0.36	0.33	0.54	-0.55	0.01	0.50	0.41	0.42	0.07	-0.01	-0.01	0.02	0.25	0.33	0.04	0.11	0.01	0.09	0.64	0.41	-0.01	0.63
damageDealt -	0.41	0.52	1.00	0.74	0.61	0.34	-0.68	0.05	0.89	0.70	0.56	-0.01	-0.04	-0.04	-0.00	0.26	0.14	0.05	0.04	0.02	0.08	0.40	0.35	0.02	0.44
DBNOs -	0.30	0.36	0.74	1.00	0.47	0.27	-0.56	0.04	0.71	0.65	0.45	-0.01	-0.27	-0.27	-0.00	0.30	0.10	0.04	0.02	0.07	0.06	0.28	0.22	0.01	0.28
headshotKills -	0.20	0.33	0.61	0.47	1.00	0.20	-0.47	0.02	0.67	0.51	0.45	-0.02	0.01	0.01	0.01	0.15	0.08	0.01	0.03	0.01	0.04	0.25	0.22	0.00	0.28
heals -	0.23	0.54	0.34	0.27	0.20	1.00	-0.39	-0.00	0.31	0.27	0.26	0.11	-0.06	-0.06	0.02	0.24	0.30	0.02	0.08	0.04	0.06	0.43	0.31	-0.01	0.43
killPlace -	-0.29	-0.55	-0.68	-0.56	-0.47	-0.39	1.00	-0.02	-0.73	-0.81	-0.54	-0.01	0.02	0.02	-0.01	-0.27	-0.24	-0.06	-0.09	-0.04	-0.07	-0.59	-0.49	-0.00	-0.72
killPoints -	0.04	0.01	0.05	0.04	0.02	-0.00	-0.02	1.00	0.04	0.04	0.01	-0.10	-0.04	-0.04	-0.98	0.01	-0.03	-0.00	-0.00	-0.04	-0.01	0.00	-0.00	0.98	0.01
kills -	0.32	0.50	0.89	0.71	0.67	0.31	-0.73	0.04	1.00	0.80	0.60	-0.03	-0.00	-0.00	0.01	0.25	0.11	0.06	0.04	0.02	0.08	0.37	0.34	0.01	0.42
killStreaks -	0.24	0.41	0.70	0.65	0.51	0.27	-0.81	0.04	0.80	1.00	0.52	-0.03	-0.04	-0.04	0.00	0.24	0.10	0.07	0.03	0.03	0.06	0.33	0.28	0.01	0.38
longestKill -	0.26	0.42	0.56	0.45	0.45	0.26	-0.54	0.01	0.60	0.52	1.00	0.06	-0.02	-0.02	0.02	0.17	0.19	0.03	0.05	0.01	0.08	0.41	0.28	-0.00	0.41
matchDuration -	-0.02	0.07	-0.01	-0.01	-0.02	0.11	-0.01	-0.10	-0.03	-0.03	0.06	1.00	0.04	0.04	0.09	0.02	0.37	0.02	0.01	0.05	0.06	0.15	0.13	-0.09	-0.01
maxPlace -	-0.15	-0.01	-0.04	-0.27	0.01	-0.06	0.02	-0.04	-0.00	-0.04	-0.02	0.04	1.00	1.00	0.05	-0.16	-0.05	0.01	0.01	-0.04	-0.02	-0.08	-0.04	-0.03	0.04
numGroups -	-0.15	-0.01	-0.04	-0.27	0.01	-0.06	0.02	-0.04	-0.00	-0.04	-0.02	0.04	1.00	1.00	0.05	-0.16	-0.05	0.01	0.01	-0.04	-0.02	-0.08	-0.04	-0.03	0.04
rankPoints -	-0.02	0.02	-0.00	-0.00	0.01	0.02	-0.01	-0.98	0.01	0.00	0.02	0.09	0.05	0.05	1.00	-0.00	0.04	0.01	0.00	0.04	0.01	0.02	0.01	-0.99	0.01
revives -	0.20	0.25	0.26	0.30	0.15	0.24	-0.27	0.01	0.25	0.24	0.17	0.02	-0.16	-0.16	-0.00	1.00	0.11	0.01	0.02	0.03	0.04	0.24	0.17	0.00	0.24
rideDistance -	0.11	0.33	0.14	0.10	0.08	0.30	-0.24	-0.03	0.11	0.10	0.19	0.37	-0.05	-0.05	0.04	0.11	1.00	0.10	0.05	0.05	0.12	0.31	0.28	-0.03	0.34
roadKills -	0.01	0.04	0.05	0.04	0.01	0.02	-0.06	-0.00	0.06	0.07	0.03	0.02	0.01	0.01	0.01	0.01	0.10	1.00	0.00	0.01	0.03	0.02	0.02	-0.00	0.03
swimDistance -	0.02	0.11	0.04	0.02	0.03	0.08	-0.09	-0.00	0.04	0.03	0.05	0.01	0.01	0.01	0.00	0.02	0.05	0.00	1.00	0.01	0.01	0.17	0.08	-0.00	0.15
teamKills -	0.01	0.01	0.02	0.07	0.01	0.04	-0.04	-0.04	0.02	0.03	0.01	0.05	-0.04	-0.04	0.04	0.03	0.05	0.01	0.01	1.00	0.11	0.02	0.04	-0.04	0.02
vehicleDestroys -	0.06	0.09	0.08	0.06	0.04	0.06	-0.07	-0.01	0.08	0.06	0.08	0.06	-0.02	-0.02	0.01	0.04	0.12	0.03	0.01	0.11	1.00	0.08	0.06	-0.01	0.07
walkDistance -	0.29	0.64	0.40	0.28	0.25	0.43	-0.59	0.00	0.37	0.33	0.41	0.15	-0.08	-0.08	0.02	0.24	0.31	0.02	0.17	0.02	0.08	1.00	0.54	-0.00	0.81
weaponsAcquired -	0.24	0.41	0.35	0.22	0.22	0.31	-0.49	-0.00	0.34	0.28	0.28	0.13	-0.04	-0.04	0.01	0.17	0.28	0.02	0.08	0.04	0.06	0.54	1.00	-0.00	0.58
winPoints -	0.02	-0.01	0.02	0.01	0.00	-0.01	-0.00	0.98	0.01	0.01	-0.00	-0.09	-0.03	-0.03	-0.99	0.00	-0.03	-0.00	-0.00	-0.04	-0.01	-0.00	-0.00	1.00	0.01
winPlacePerc -	0.30	0.63	0.44	0.28	0.28	0.43	-0.72	0.01	0.42	0.38	0.41	-0.01	0.04	0.04	0.01	0.24	0.34	0.03	0.15	0.02	0.07	0.81	0.58	0.01	1.00
	assists -	boosts -	mageDealt -	DBNOs -	adshotKills -	heals -	killPlace -	killPoints -	Kills -	killStreaks -	longestKill -	chDuration -	maxPlace -	ոսmGroups -	rankPoints -	revives -	deDistance -	roadKills -	imDistance -	teamKills -	:leDestroys -	الا	nsAcquired -	winPoints -	inPlacePerc -

Figure 4: Correlation



- 0.25

- -0.25

- -0.50



High Correlation

Problem Definition

Data Preprocess

Data Visualization

Feature selection

Modeling and Forecasting

boosts -	0.31	1.00	0.52	0.36	0.33	0.54	-0.55	0.01	0.50	0.41	0.42	0.07	-0.01	-0.01	0.02	0.25	0.33	0.04	0.11	0.01	0.09	0.64	0.41	-0.01	0.63	1.0	.00
damageDealt -	0.41	0.52	1.00	0.74	0.61	0.34	-0.68	0.05	0.89	0.70	0.56	-0.01	-0.04	-0.04	-0.00	0.26	0.14	0.05	0.04	0.02	0.08	0.40	0.35	0.02	0.44	- 0.	.75
heals -	0.23	0.54	0.34	0.27	0.20	1.00	-0.39	-0.00	0.31	0.27	0.26	0.11	-0.06	-0.06	0.02	0.24	0.30	0.02	0.08	0.04	0.06	0.43	0.31	-0.01	0.43		. 50
killPlace -	-0.29	-0.55	-0.68	-0.56	-0.47	-0.39	1.00	-0.02	-0.73	-0.81	-0.54	-0.01	0.02	0.02	-0.01	-0.27	-0.24	-0.06	-0.09	-0.04	-0.07	-0.59	-0.49	-0.00	-0.72	- 0.1	.50
kills -	0.32	0.50	0.89	0.71	0.67	0.31	-0.73	0.04	1.00	0.80	0.60	-0.03	-0.00	-0.00	0.01	0.25	0.11	0.06	0.04	0.02	0.08	0.37	0.34	0.01	0.42	- 0.:	.25
killStreaks -	0.24	0.41	0.70	0.65	0.51	0.27	-0.81	0.04	0.80	1.00	0.52	-0.03	-0.04	-0.04	0.00	0.24	0.10	0.07	0.03	0.03	0.06	0.33	0.28	0.01	0.38		
longestKill -	0.26	0.42	0.56	0.45	0.45	0.26	-0.54	0.01	0.60	0.52	1.00	0.06	-0.02	-0.02	0.02	0.17	0.19	0.03	0.05	0.01	0.08	0.41	0.28	-0.00	0.41	- 0.	.00
rideDistance -	0.11	0.33	0.14	0.10	0.08	0.30	-0.24	-0.03	0.11	0.10	0.19	0.37	-0.05	-0.05	0.04	0.11	1.00	0.10	0.05	0.05	0.12	0.31	0.28	-0.03	0.34	(-0.25
walkDistance -	0.29	0.64	0.40	0.28	0.25	0.43	-0.59	0.00	0.37	0.33	0.41	0.15	-0.08	-0.08	0.02	0.24	0.31	0.02	0.17	0.02	0.08	1.00	0.54	-0.00	0.81		
weaponsAcquired -	0.24	0.41	0.35	0.22	0.22	0.31	-0.49	-0.00	0.34	0.28	0.28	0.13	-0.04	-0.04	0.01	0.17	0.28	0.02	0.08	0.04	0.06	0.54	1.00	-0.00	0.58	(-0.50
winPlacePerc -	0.30	0.63	0.44	0.28	0.28	0.43	-0.72	0.01	0.42	0.38	0.41	-0.01	0.04	0.04	0.01	0.24	0.34	0.03	0.15	0.02	0.07	0.81	0.58	0.01	1.00	(-0.75
	assists -	boosts -	mageDealt -	- DBNOs -	adshotKills -	heals -	killPlace -	killPoints -	kills -	killStreaks -	longestKill -	chDuration -	maxPlace -	numGroups -	rankPoints -	revives -	deDistance -	roadKills -	imDistance -	teamKills -	cleDestroys -	alkDistance -	nsAcquired -	winPoints -	inPlacePerc –		

Figure 5: High Correlation





Data Preprocess

Data Visualization

Feature selection

Modeling and Forecasting

Model Selection

Comparison and Conclusion

Modeling and Forecasting





Model Selection

Problem Definition

Data Preprocess

Data Visualization

Feature selection

Modeling and Forecasting

Model Selection

- Linear Regression
- Decision Tree





Linear Regression

Problem Definition

Data Preprocess

Data Visualization

Feature selection

Modeling and Forecasting

Model Selection

Comparison and Conclusion

■ Sklearn linear regression parameters on grid search and cross validation.

Table 2: linear regression parameters and cross validation

Parameters	Values	CV
fit_intercept	True/False	3
normalize	True/False	3





Decision Tree

Problem Definition

Data Preprocess

Data Visualization

Feature selection

Modeling and Forecasting

Model Selection

Comparison and Conclusion

■ Sklearn Decision Tree parameters on grid search and cross validation.

Table 3: Decision Tree parameters and cross validation

Parameters	Values						
criterion	"mse", "friedman_mse", "mae"	3					
min_samples_leaf	1,2	3					





Data Preprocess

Data Visualization

Feature selection

Modeling and Forecasting

Comparison and Conclusion

Best Parameters

Comparison

Conclusion





Best Parameters

Problem Definition

Data Preprocess

Data Visualization

Feature selection

Modeling and Forecasting

Comparison and Conclusion

Best Parameters

Comparison

Conclusion

- Best parameters by grid search
- Linear Regression: copy_X=True, fit_intercept=True, n_jobs=None, normalize=True
- Decision Tree: ccp_alpha=0.0, criterion='mse', max_depth=None, max_features=None, max_leaf_nodes=None, min_impurity_decrease=0.0, min_impurity_split=None, min_samples_leaf=2, min_samples_split=2, min_weight_fraction_leaf=0.0, presort='deprecated', random_state=None, splitter='best'



Comparison

Problem Definition

Data Preprocess

Data Visualization

Feature selection

Modeling and Forecasting

Comparison and Conclusion

Best Parameters

Comparison

Conclusion

■ Compare MSE result between Linear regression and Decision Tree

Table 4: Linear regression VS Decision Tree

	train MSE	test MSE
linear regression	0.01564124116618947	0.015303007019988265
decision tree	0.01859312767771217	0.17048691321544765





Conclusion

Problem Definition

Data Preprocess

Data Visualization

Feature selection

Comparison and Conclusion

Modeling and Forecasting

Best Parameters

Comparison

Conclusion

- Both training and testing data shows that linear model get lower mean square error value.
- Most players choose to play squad-fpp and duo-fpp
- More walking distance always can bring higher win place.





Questions?

Problem Definition

Data Preprocess

Data Visualization

Feature selection

Modeling and Forecasting

Comparison and Conclusion

Best Parameters

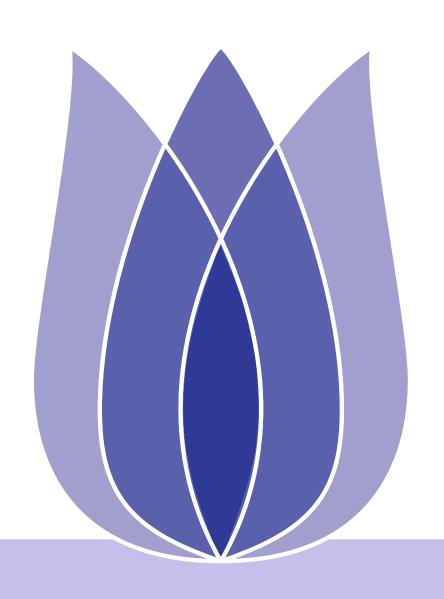
Comparison

Conclusion





Contact Information



YANG CAO

Deakin University, Australia



CAOYANG@DEAKIN.EDU.AU



TEAM FOR UNIVERSAL LEARNING AND INTELLIGENT PROCESSING