



Introduction







API-GENERATED DATA



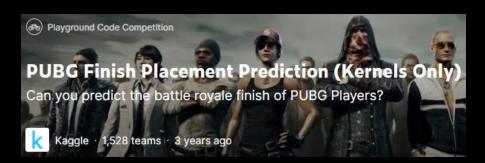
BEHAVIORS VS.
PERFORMANCE



FINAL PLACEMENT PERCENTAGE

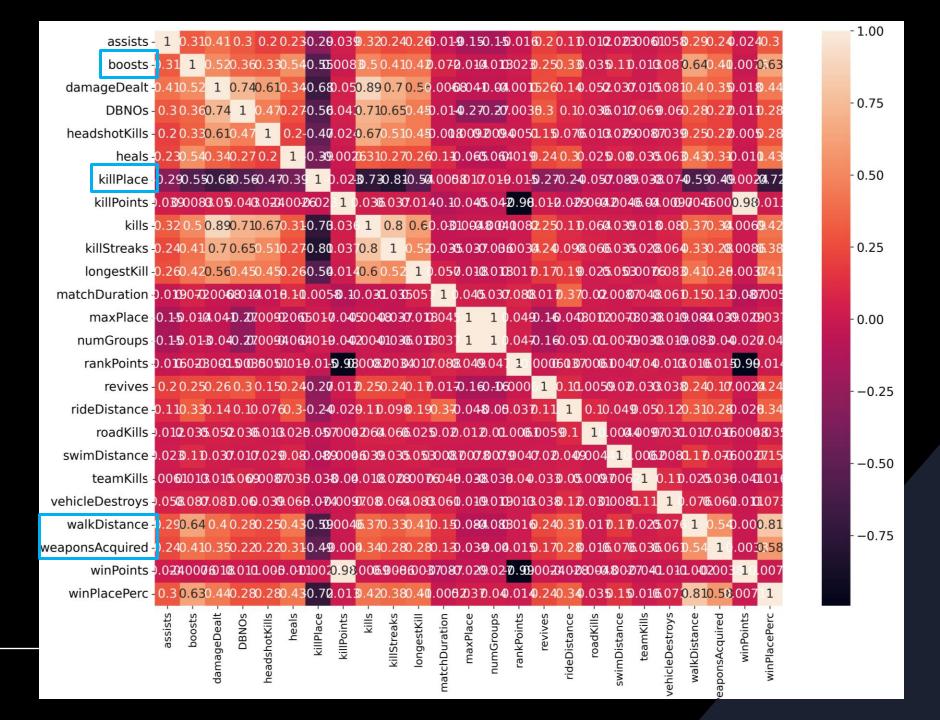


REGRESSION



EDA -Correlation

- Highly correlated
 - ✓ Kill rank (-)
 - ✓ Walk distance (+)
 - ✓ # of boosts (+)
 - ✓ # of weapons (+)





Remove

l record: missing target value

5 columns -

3 IDs: Id, matchId, groupId

2 inconsistent: numGroups, rankPoints



Unique Ids

Games are independent

No group structure

No time series

Data Preprocessing

Data Transformation

OneHotEncoder

MinMaxScaler

StandardScaler

matchType

killPlace

The rest 20

maxPlace

Data Set(s)

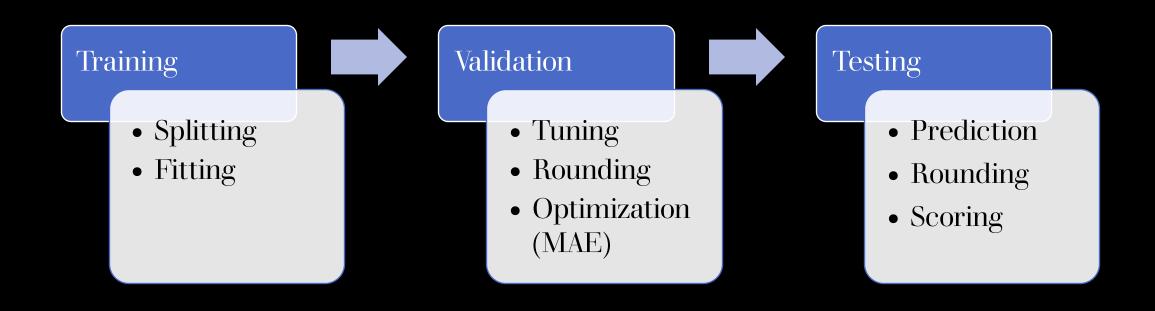
Original

- # of records: 4,446,966
- # of columns: 29
 - IDs: 3
 - Independent variables: 25
 - Target variable: l

Sampled & Preprocessed

- # of records: 44,470 (1%)
- Split
 - Training: 60%
 - Validation: 20%
 - Testing: 20%
- # of features: 38

Pipeline





Lasso

alpha: le-7, le-6, le-5, le-4, le-3, le-2, le-1

Ridge

alpha: le-7, le-6, le-5, le-4, le-3, le-2, le-1

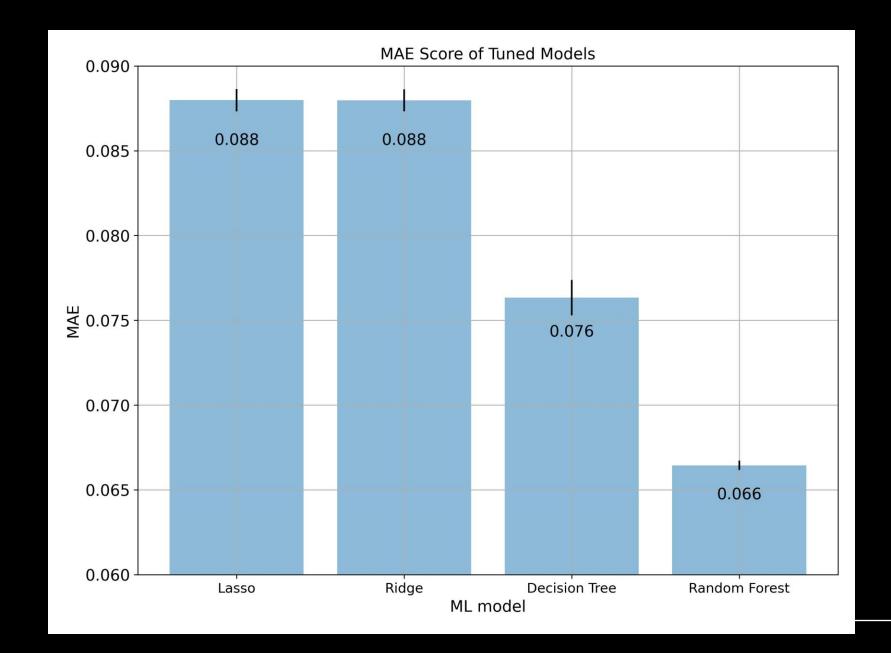
Decision Tree

splitter: best, random
max_depth: 7, 8, 9, 10, 11
max_features: auto, log2,
sqrt

Random Forest

max_depth: 10, 11, 12, 13, 14 n_estimators: 500, 600, 700, 800, 900

Model Tuning



Model Comparison

Model Evaluation

Baseline (median)

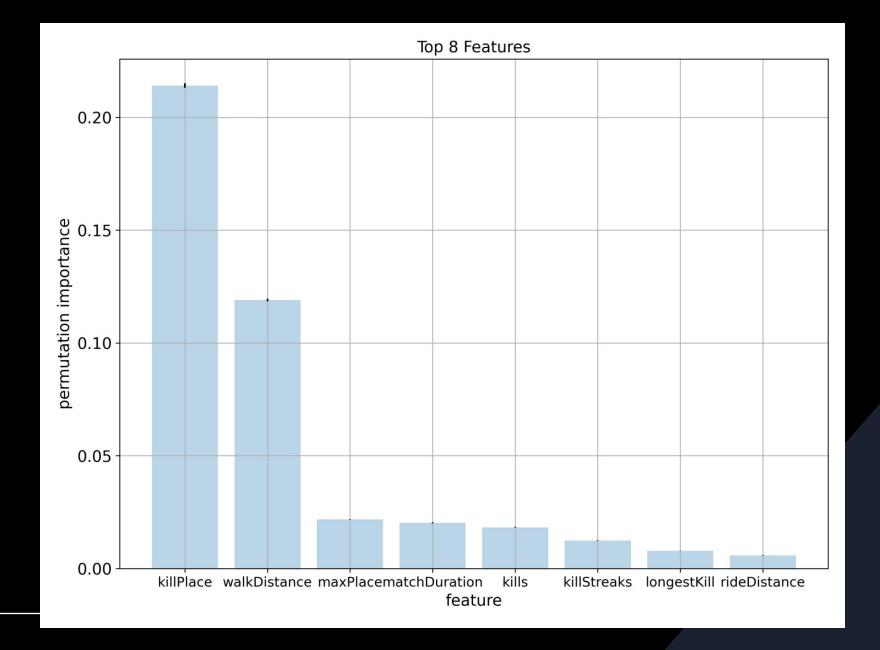
- MAE: 0.268 +/- 0.002
- 505 stds above RF

Random Forest

- MAE: 0.066 +/- 0.0004
- 101 stds below baseline

Feature Importances

- ✓ Top 2: highly correlated
- ✓ Kill-related (4)
- ✓ Distance-related (2)
- ✓ Ranking-related (2)
- ✓ match duration???



Outlook

Evaluation Metric

- MAPE
 - target in [0,1]

Computing Power

- Larger Sample
- More Algorithms
- Better Tuning

THANKYOU!

Any question?

Data Dictionary

- . DBNOs Number of enemy players knocked.
- · assists Number of enemy players this player damaged that were killed by teammates.
- boosts Number of boost items used.
- · damageDealt Total damage dealt. Note: Self inflicted damage is subtracted.
- · headshotKills Number of enemy players killed with headshots.
- heals Number of healing items used.
- Id Player's Id
- · killPlace Ranking in match of number of enemy players killed.
- killPoints Kills-based external ranking of player. (Think of this as an Elo ranking where only kills matter.) If there is a value other than
 -1 in rankPoints, then any 0 in killPoints should be treated as a "None".
- 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. This may be misleading, as downing a player and driving away may lead to a large longestKill stat.
- matchDuration Duration of match in seconds.
- . matchld ID to identify match. There are no matches that are in both the training and testing set.
- matchType String identifying the game mode that the data comes from. The standard modes are "solo", "duo", "squad", "solo-fpp",
 "duo-fpp", and "squad-fpp"; other modes are from events or custom matches.

Data Dictionary

- rankPoints Elo-like ranking of player. This ranking is inconsistent and is being deprecated in the API's next version, so use with caution. Value of -1 takes place of "None".
- · revives Number of times this player revived teammates.
- rideDistance Total distance traveled in vehicles measured in meters.
- roadKills Number of kills while in a vehicle.
- swimDistance Total distance traveled by swimming measured in meters.
- teamKills Number of times this player killed a teammate.
- vehicleDestroys Number of vehicles destroyed.
- walkDistance Total distance traveled on foot measured in meters.
- · weaponsAcquired Number of weapons picked up.
- winPoints Win-based external ranking of player. (Think of this as an Elo ranking where only winning matters.) If there is a value other than -1 in rankPoints, then any 0 in winPoints should be treated as a "None".
- groupId ID to identify a group within a match. If the same group of players plays in different matches, they will have a different groupId each time.
- numGroups Number of groups we have data for in the match.
- maxPlace Worst placement we have data for in the match. This may not match with numGroups, as sometimes the data skips over placements.
- winPlacePerc The target of prediction. This is a percentile winning placement, where 1 corresponds to 1st place, and 0 corresponds
 to last place in the match. It is calculated off of maxPlace, not numGroups, so it is possible to have missing chunks in a match.

Data Set(s)

Original

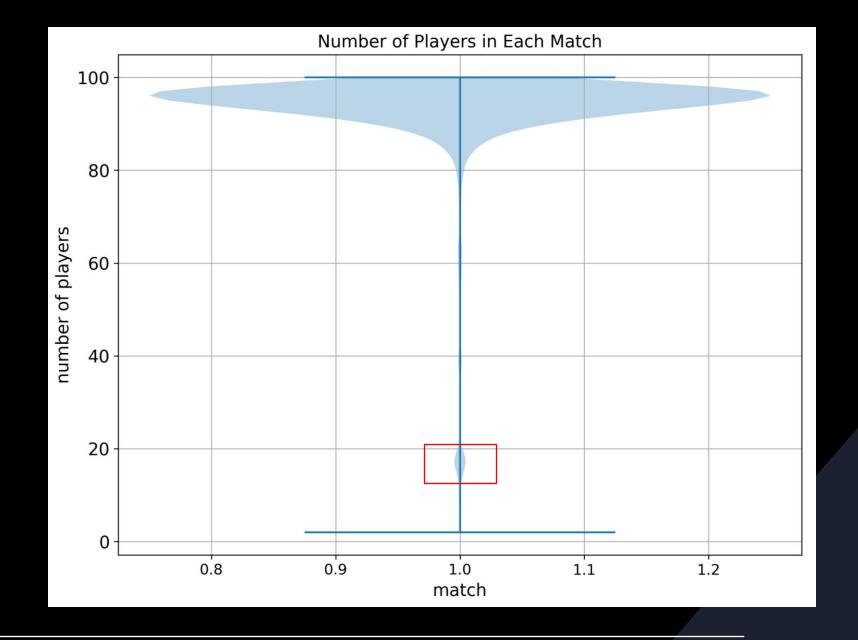
- # of records: 4,446,966
- # of columns: 29
 - IDs: 3
 - Independent variables: 25
 - Target variable: 1

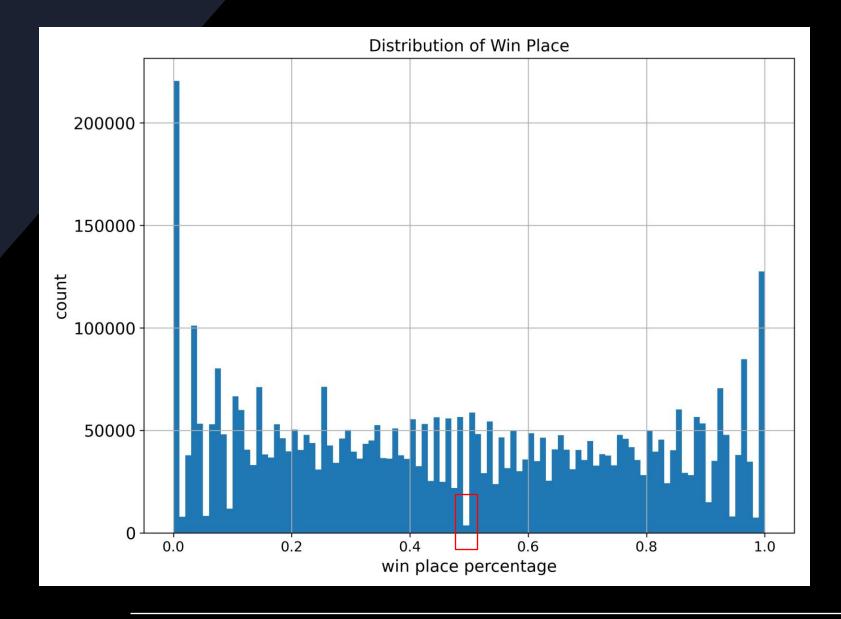
Preprocessed

• Loading ...

EDA - Match

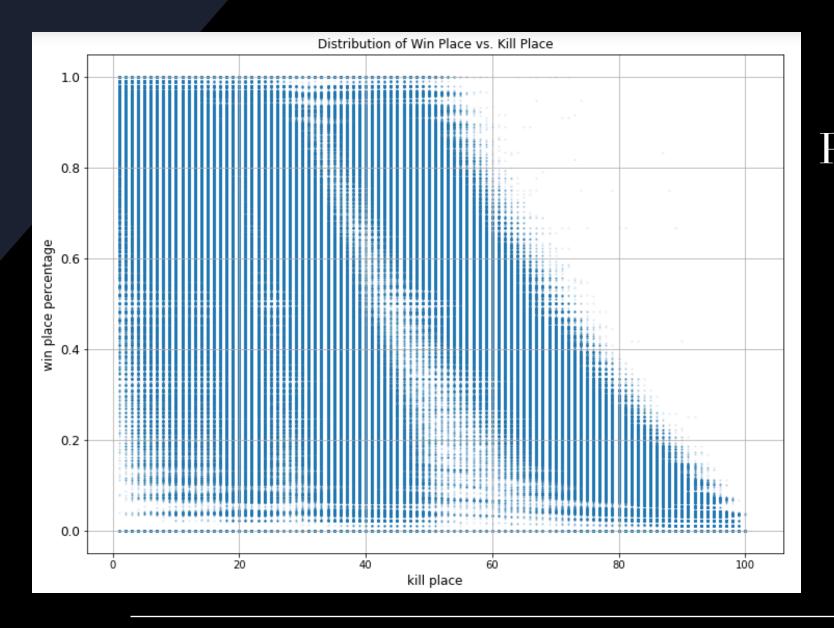
- Majority
 - 80 to 100 players
- ❖ A small group
 - around 20 players





EDA - Placement

- **❖** Bin size: 100
- * Target variable
 - o Obtained by calculation
 - Not the exact rank
 - Abnormal distribution



EDA – Placement VS. Kill Rank

- **❖** Negative correlation
 - ➤ Kill rank: ▼
 - Placement: 1
- * Top right
 - > Extremely low density

Data Split

