

機器學習期末專題 —PUBG數據分析

作者:4111056030孟令桓



PLAYERUNKNOWN'S
BATTLEGROUNDS

前言與動機

前言與動機

- 我對各式各樣的電腦遊戲充滿興趣，也玩過不少遊戲，會走入資訊這條路也有部分原因是為了一探究竟遊戲背後的運作原理。同時，電子競技活動在這些年逐漸形成完整且具規模的產業環境。基於上述理由，因此我決定以遊戲資料來做為這次的期末專題的主題。
- 在網路上搜尋遊戲相關資料時找到了以曾經紅極一時的PUBG為主題的資料集，我將以此資料集做名次迴歸與模式分類兩個研究



簡報流程圖

介紹

背景介紹

資料介紹與任務目標

資料觀察與清洗

方法

任務1. 名次預測迴歸

資料前處理

建立模型與調整

結果分析

任務2. 模式分類預測

資料前處理

建立模型與調整

結果分析

結果

總結



背景介紹

PUBG介紹

- PUBG全名為PlayerUnknown 's Battlegrounds，是一款線上槍戰大逃殺類型的遊戲，數十個玩家在同一張地圖上對戰，角逐最終贏家
- 開局時玩家選擇降落在地圖上的不同地方，遊戲過程中在場上撿取武器及裝備，並以武器擊殺其他隊伍的玩家獲得勝利，其餘玩家也會依出局順序排名。
- 分成多種模式(ex:單人、多人組隊、第一人稱、排名戰...)
- 多人模式下玩家被擊倒時可由隊友急救復活



資料介紹與任務目標

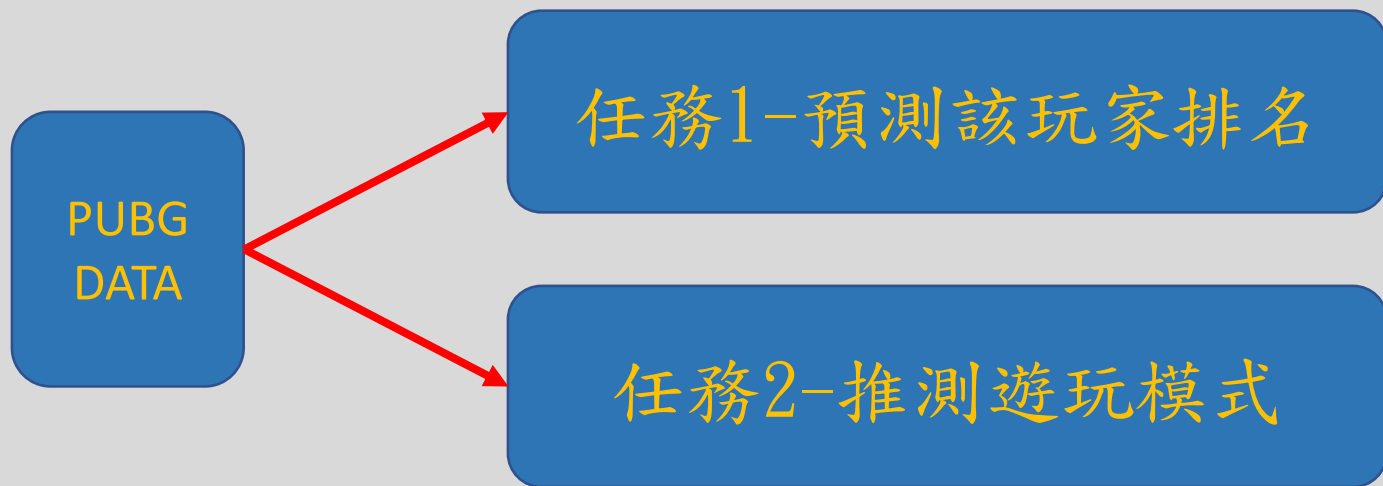
資料集來源

- 來源-- <https://www.kaggle.com/competitions/pubg-finish-placement-prediction/overview>
- 該資料集來源為kaggle上一比賽，原目標為根據遊戲數據預測該玩家結束時的名次



目標

- 本報告有兩個任務
 - 利用玩家的遊戲數據預測他該場取得的排名(kaggle原任務)
 - 利用遊戲數據與排名推測他遊玩的是什麼模式(自己額外增加的任務)
- 次要任務
 - 利用上述結果找出可能對電競產業有幫助的資訊



同個資料集做兩個任務



資料內容




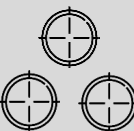
- 資料集內將近4.45百萬筆資料
- 內容可分成以下三部分
 - 玩家與隊伍識別資料
 - 遊戲內各項數據
 - 最終排名
 - 0為第一名，1為最後一名，中間名次則按百分比呈現


遊戲內各項數據


• 包含玩家行為、交戰與擊殺、裝備使用等多種數據，例如：


➤ assists-助攻數 

➤ kills-擊殺數 

➤ killStreaks-短時間內擊殺總數 

➤ longestKill-最遠擊殺距離 


➤ vehicleDestroys-載具摧毀數 


➤ DBNOs-被擊倒但未死亡次數 

➤ damageDealt-造成傷害值 

➤ heals-使用回血物品次數 

➤ walkDistance-行走距離 

➤ matchType-遊戲模式 

➤ matchDuration-遊戲持續時間 



所有feature(節錄自kaggle)

| | |
|---|---|
| <ul style="list-style-type: none">• 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.• matchId - 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. | <ul style="list-style-type: none">• 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. |
|---|---|



資料觀察與清洗

資料集細節

- 僅一筆有缺失排名資料，移除後剩4446965筆
- 包含排名資料共29個feature
 - 4個feature為String格式
 - 其餘皆為數值型資料(int、float)

```
Index: 4446965 entries, 0 to 4446965  
Data columns (total 29 columns):
```

| id | groupId | matchId | assists | boosts | damageDealt | DBNOs | headshotKills | heals | killPlace | killPoints | kills | killStreak | longestKill | matchDuration | matchType | maxPlace | numGroup | rankPoints | revives | rideDistance | roadKills | swimDistance | teamKills | vehicleDestroyed | walkDistance | weaponsAcquired | winPoints | winPlace |
|-----------------------------|---------|---------|---------|--------|-------------|-------|---------------|-------|-----------|------------|-------|------------|-------------|---------------|-----------|----------|----------|------------|---------|--------------|-----------|--------------|-----------|------------------|--------------|-----------------|-----------|----------|
| 7f96b2f874d4b580da10357fd1 | | | 0 | 0 | 0 | 0 | 0 | 0 | 60 | 1241 | 0 | 0 | 0 | 1300 | squad-fpp | 28 | 26 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 244.8 | 1 | 1466 | 0.444 |
| eeef90569b684d5656aeb375fc5 | | | 0 | 0 | 91.47 | 0 | 0 | 0 | 57 | 0 | 0 | 0 | 0 | 177 | squad-fpp | 26 | 25 | 1484 | 0 | 0.0045 | 0 | 11.04 | 0 | 0 | 1434 | 5 | 0 | 0.64 |
| 1eaf90ac76a4a42c3f110163d8 | | | 1 | 0 | 68 | 0 | 0 | 0 | 47 | 0 | 0 | 0 | 0 | 1318 | duo | 50 | 47 | 1491 | 0 | 0 | 0 | 0 | 0 | 0 | 161.8 | 2 | 0 | 0.7755 |
| 4616d365a930a9c79f1f1f4ef4 | | | 0 | 0 | 32.9 | 0 | 0 | 0 | 75 | 0 | 0 | 0 | 0 | 1430 | squad-fpp | 31 | 30 | 1408 | 0 | 0 | 0 | 0 | 0 | 0 | 202.7 | 3 | 0 | 0.1667 |
| 315c96c2de04010b6dc8ff871 | | | 0 | 0 | 100 | 0 | 0 | 0 | 45 | 0 | 1 | 1 | 58.53 | 1424 | solo-fpp | 97 | 95 | 1560 | 0 | 0 | 0 | 0 | 0 | 0 | 49.75 | 2 | 0 | 0.1875 |
| ff79c12f3289a6836bac52627a | | | 0 | 0 | 100 | 1 | 1 | 0 | 44 | 0 | 1 | 1 | 18.44 | 1395 | squad-fpp | 28 | 28 | 1418 | 0 | 0 | 0 | 0 | 0 | 0 | 34.7 | 1 | 0 | 0.037 |
| 95959be02c485a1aca8274e90 | | | 0 | 0 | 0 | 0 | 0 | 0 | 96 | 1262 | 0 | 0 | 0 | 1310 | squad-fpp | 28 | 28 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 13.5 | 1 | 1497 | 0 |

少數為String的資料



人工資料篩選

- 以人工觀察剔除無用資料
 - 用於識別玩家與隊伍之資料
 - 與遊戲場次外部排名相關(對於單場遊戲成果關聯性較低)

識別用資料

擊殺外部排名

外部排名資料

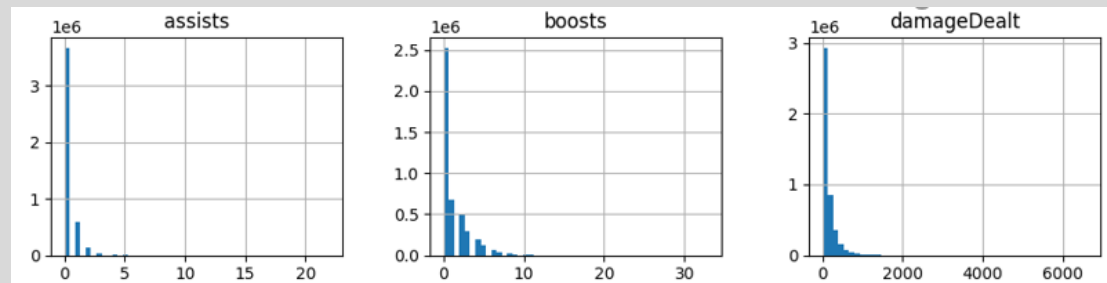
外部排名資料

| id | groupId | matchId | assists | boosts | damageDealt | DBNOs | headshots | heals | killPlace | killPoints | kills | killStreak | longestKill | matchDuration | matchType | maxPlace | numGroup | rankPoints | revives | rideDistance | roadKills | swimDistance | teamKills | vehicleDestroyed | walkDistance | weaponsUsed | winPoints | winPlace |
|------------------------------|---------|---------|---------|--------|-------------|-------|-----------|-------|-----------|------------|-------|------------|-------------|---------------|-----------|----------|----------|------------|---------|--------------|-----------|--------------|-----------|------------------|--------------|-------------|-----------|----------|
| 7196b2f874d4b580da10357fd1 | | | 0 | 0 | 0 | 0 | 0 | 0 | 60 | 1241 | 0 | 0 | 0 | 1306 | squad-fpp | 28 | 26 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 244.8 | 1 | 1466 | 0.444 |
| ce90569f684d5656aeb3756e2 | | | 0 | 0 | 91.47 | 0 | 0 | 0 | 57 | 0 | 0 | 0 | 0 | 1777 | squad-fpp | 26 | 25 | 1484 | 0 | 0.0045 | 0 | 11.04 | 0 | 0 | 1434 | 5 | 0 | 0.64 |
| 1eaf90a766a4a42c3711b163d8f | | | 1 | 0 | 68 | 0 | 0 | 0 | 47 | 0 | 0 | 0 | 0 | 1318 | duo | 50 | 47 | 1491 | 0 | 0 | 0 | 0 | 0 | 0 | 161.8 | 2 | 0 | 0.775 |
| 4616d365a5920e0e7f11f1f4ef41 | | | 0 | 0 | 32.9 | 0 | 0 | 0 | 75 | 0 | 0 | 0 | 0 | 1436 | squad-fpp | 31 | 30 | 1408 | 0 | 0 | 0 | 0 | 0 | 0 | 202.7 | 3 | 0 | 0.166 |
| 315e96c2fcd0401b6b6dc8ff871 | | | 0 | 0 | 100 | 0 | 0 | 0 | 45 | 0 | 1 | 1 | 58.53 | 1424 | solo-fpp | 97 | 95 | 1560 | 0 | 0 | 0 | 0 | 0 | 0 | 49.75 | 2 | 0 | 0.187 |
| ff79c12f3289a6836bac52627e | | | 0 | 0 | 100 | 1 | 1 | 0 | 44 | 0 | 1 | 1 | 18.44 | 1395 | squad-fpp | 28 | 28 | 1418 | 0 | 0 | 0 | 0 | 0 | 0 | 34.7 | 1 | 0 | 0.037 |
| 83959be02c485a1aca8274e98 | | | 0 | 0 | 0 | 0 | 0 | 0 | 96 | 1262 | 0 | 0 | 0 | 1316 | squad-fpp | 28 | 28 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 13.5 | 1 | 147 | 0 |

刪除這些feature



資料統計資訊



資料分布極度左偏

- 部分數值的min或max過於誇張
 - 擊殺數過高、武器獲得次數過多、遊戲時間過長/過短 等
 - 後續處理將去除離群值

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W |
|---|-------|----------|----------|-------------|----------|---------------|----------|-----------|----------|------------|-------------|---------------|----------|-----------------|--------------|-----------|--------------|-----------|------------------|--------------|-----------------|--------------|----------|
| 1 | | assists | boosts | damageDealt | DBNOs | headshotKills | heals | killPlace | kills | killStreak | longestKill | matchDuration | maxPlace | numGroupRevives | rideDistance | roadKills | swimDistance | teamKills | vehicleDestroyed | walkDistance | weaponsAcquired | winPlacePerc | |
| 2 | count | 4446965 | 4446965 | 4446965 | 4446965 | 4446965 | 4446965 | 4446965 | 4446965 | 4446965 | 4446965 | 4446965 | 4446965 | 4446965 | 4446965 | 4446965 | 4446965 | 4446965 | 4446965 | 4446965 | 4446965 | 4446965 | 4446965 |
| 3 | mean | 0.233815 | 1.106908 | 130.7172 | 0.657876 | 0.22682 | 1.370148 | 47.59936 | 0.924784 | 0.543955 | 22.9976 | 1579.507 | 44.50468 | 43.0076 | 0.164659 | 606.1158 | 0.003496 | 4.509323 | 0.023868 | 0.007918 | 1154.218 | 3.660488 | 0.472822 |
| 4 | std | 0.588573 | 1.715794 | 170.7806 | 1.145743 | 0.602155 | 2.679982 | 27.46293 | 1.558445 | 0.710972 | 50.97262 | 258.7388 | 23.8281 | 23.28949 | 0.472167 | 1498.344 | 0.073373 | 30.5022 | 0.167394 | 0.092612 | 1183.497 | 2.456543 | 0.307405 |
| 5 | min | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 133 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 25% | 0 | 0 | 0 | 0 | 0 | 0 | 24 | 0 | 0 | 0 | 1367 | 28 | 27 | 0 | 0 | 0 | 0 | 0 | 0 | 155.1 | 2 | 0.2 |
| 7 | 50% | 0 | 0 | 84.24 | 0 | 0 | 0 | 47 | 0 | 0 | 0 | 1438 | 30 | 30 | 0 | 0 | 0 | 0 | 0 | 0 | 685.6 | 3 | 0.4583 |
| 8 | 75% | 0 | 2 | 186 | 1 | 0 | 2 | 71 | 1 | 1 | 21.32 | 1851 | 49 | 47 | 0 | 0.191 | 0 | 0 | 0 | 0 | 1976 | 5 | 0.7407 |
| 9 | max | 22 | 33 | 6616 | 53 | 64 | 80 | 101 | 72 | 20 | 1094 | 2237 | 100 | 100 | 39 | 40710 | 18 | 3823 | 12 | 5 | 25780 | 236 | 1 |

正常狀況很難擊殺數十個玩家

時間過短: 玩家尚未跳傘完畢
時間過長: 一段時間後安全區消失, 剩餘玩家不斷扣血直到死亡無法存活過久

獲取武器次數過多

PLAYERUNKNOWN'S
BATTLEGROUNDS

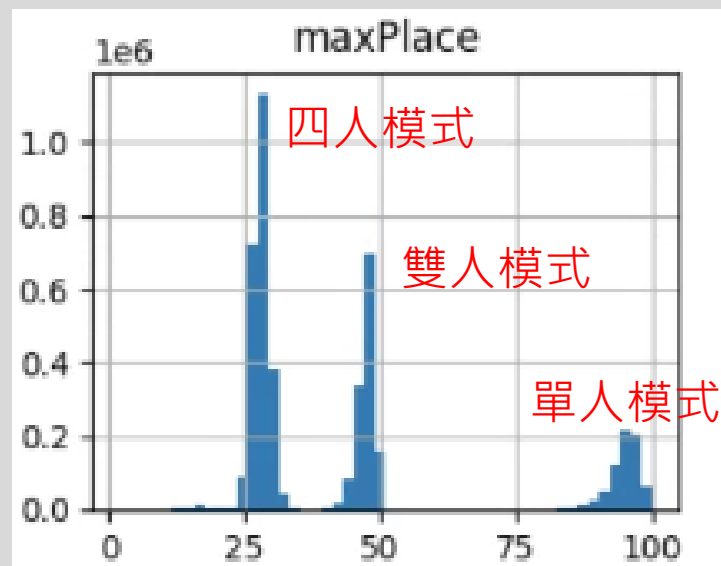
特徵關聯性

- 找出彼此關聯性高的feature
- 以關聯性0.9做門檻去除
- 將maxPlace去除

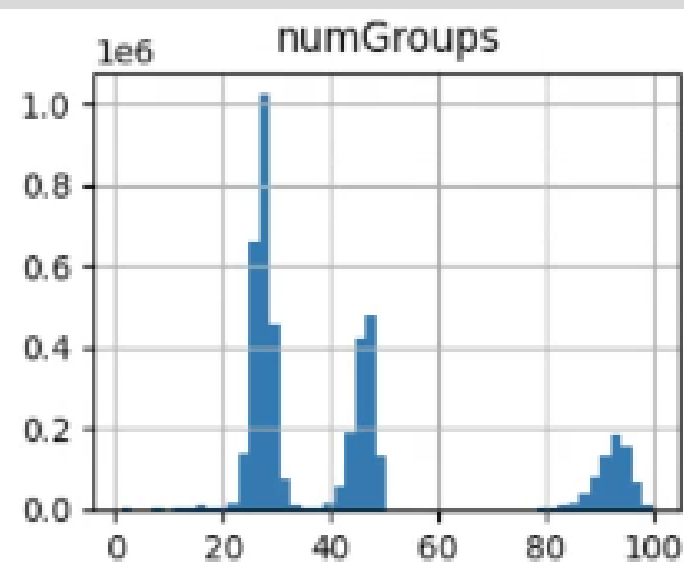
```
maxPlace  
0.9978997344931734  
numGroups  
0.9978997344931734
```

兩者相關性

該場遊戲最後一名的排名



該場遊戲隊伍數



正常狀況下最後一名與隊伍數相等
且不同模式有不同隊伍數



遊戲模式

- matchType分成右邊幾種模式
- 模式解釋
 - squad: 四人一組
 - duo: 兩人一組
 - solo: 一人一組
 - fpp: 第一人稱
 - normal: 非排位戰
 - flare/crash: 其他特殊模式

| | |
|------------------|---------|
| matchType | |
| squad-fpp | 1756186 |
| duo-fpp | 996691 |
| squad | 626526 |
| solo-fpp | 536761 |
| duo | 313591 |
| solo | 181943 |
| normal-squad-fpp | 17174 |
| crashfpp | 6287 |
| normal-duo-fpp | 5489 |
| flaretpp | 2505 |
| normal-solo-fpp | 1682 |
| flarefpp | 718 |
| normal-squad | 516 |
| crashtpp | 371 |
| normal-solo | 326 |
| normal-duo | 199 |

任務1 —— 名次預測迴歸

任務概述

- 利用玩家遊玩的各項數據，推測該場遊戲他取得的名次百分比
- 任務類型
 - 迴歸任務
- 評估標準
 - R2 Score

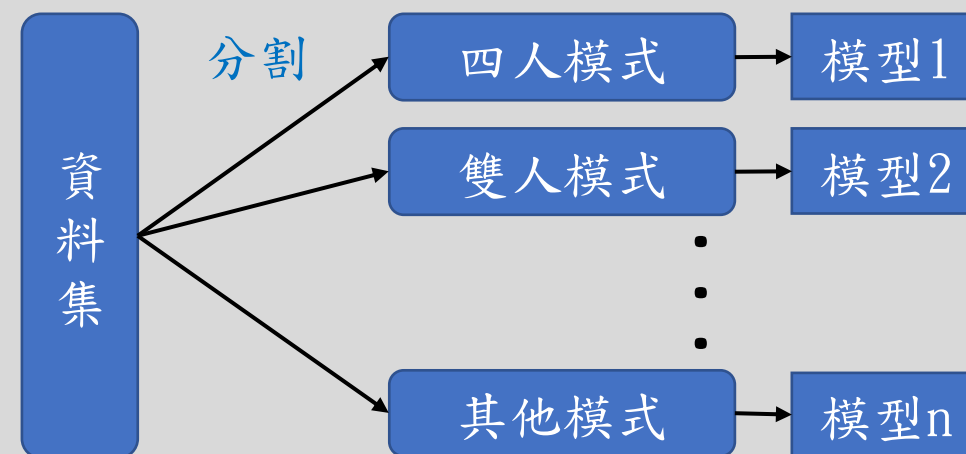
資料前處理

資料分割

模式分割完後
各子集資料量

```
squadf:1773360  
duof:1002180  
solof:538443  
squad:627042  
duo:313790  
solo:182269  
other:9881
```

- 不同模式下戰術與行為會有所不同
- 比起整個資料集丟進去訓練，將相近模式合併建立子集，依模式分別建模，以減少不同戰術對模型造成的影響
- 根據模式分成七個子集
 - 四人模式第一人稱(squadf)
 - 雙人模式第一人稱(duof)
 - 單人模式第一人稱(solof)
 - 四人模式(squad)
 - 雙人模式(duo)
 - 單人模式(solo)
 - 其他特殊模式(other)
- 每個子集80%做訓練集，10%做驗證集，10%做測試集



PR99.99的數值

離群值去除

- 不需處理之feature
 - 排名類、隊伍數量類
- 僅對訓練集操作
- 剔除含有超過99.99百分位的sample
- 剔除遊戲時間低於0.01百分位的sample
 - 時間過短也不正常

```
assists:7.0
boosts:13.0
damageDealt:2147.243099999614
DBNOs:15.0
headshotKills:9.0
heals:33.0
killPlace:100.0
kills:21.0
killStreaks:5.0
longestKill:693.4160299994983
matchDuration:2214.0
numGroups:100.0
revives:5.0
rideDistance:15320.0
roadKills:3.0
swimDistance:778.0486199999228
teamKills:3.0
vehicleDestroys:2.0
walkDistance:9154.243099999614
weaponsAcquired:39.0
winPlacePerc:1.0
```



建立模型與調整

候選模型

- ✗ • SVR
 - 包括資料標準化
 - 訓練時間過久且迭代次數仍不夠提升成效因此最終不採用

- ✓ • 隨機森林
 - 可計算特徵重要性進一步分析哪些feature較重要，並運用在電競訓練上
 - 為了保留原本feature因此前處理不包括降維

- ✓ • Gradient boost
 - 迴歸常用boosting模型

※線性迴歸與多項式迴歸較難應對較多feature，因此沒有納入考量

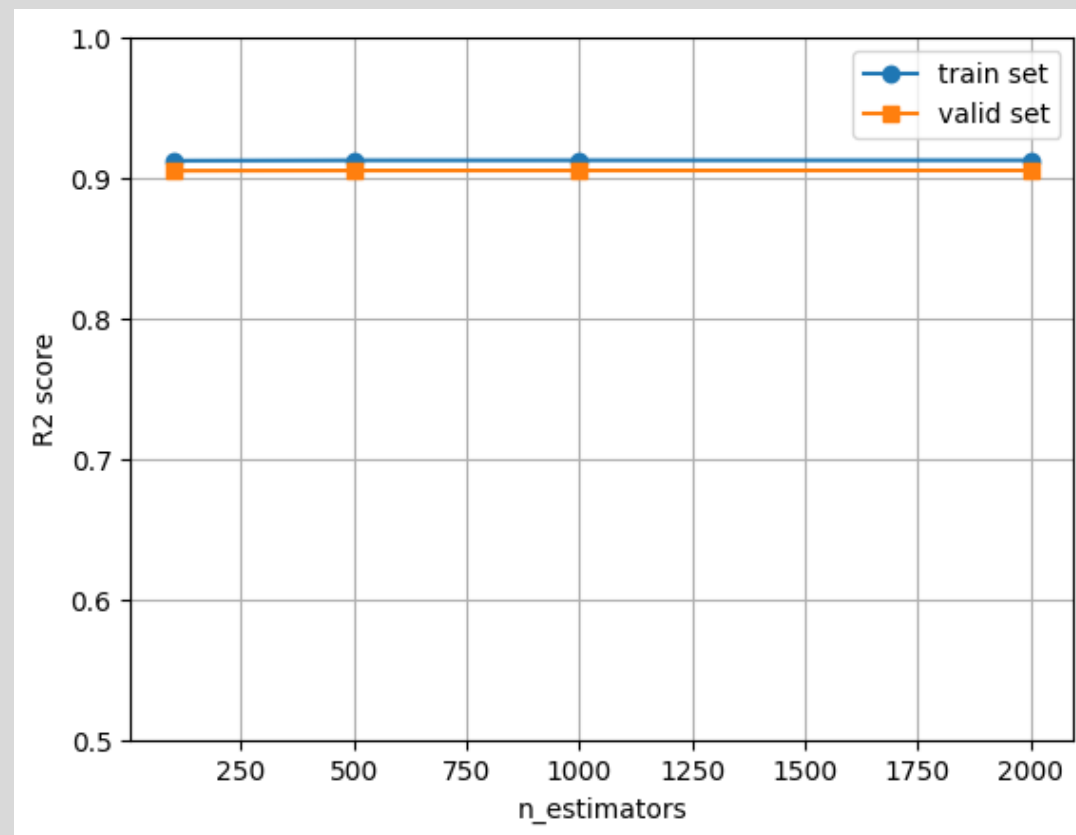


隨機森林

- 調整超參數
 - 樹數量(`n_estimators`)
 - 每顆樹的取樣數(`max_samples`)
- 皆以四人模式的子集做測試
 - 以節省訓練測試時間
 - 所有子集最終皆以同樣超參數進行訓練
- 手動進行超參數調整
 - 網格搜尋需交叉驗證與測試所有參數間組合，較花時間

隨機森林-調整樹數量

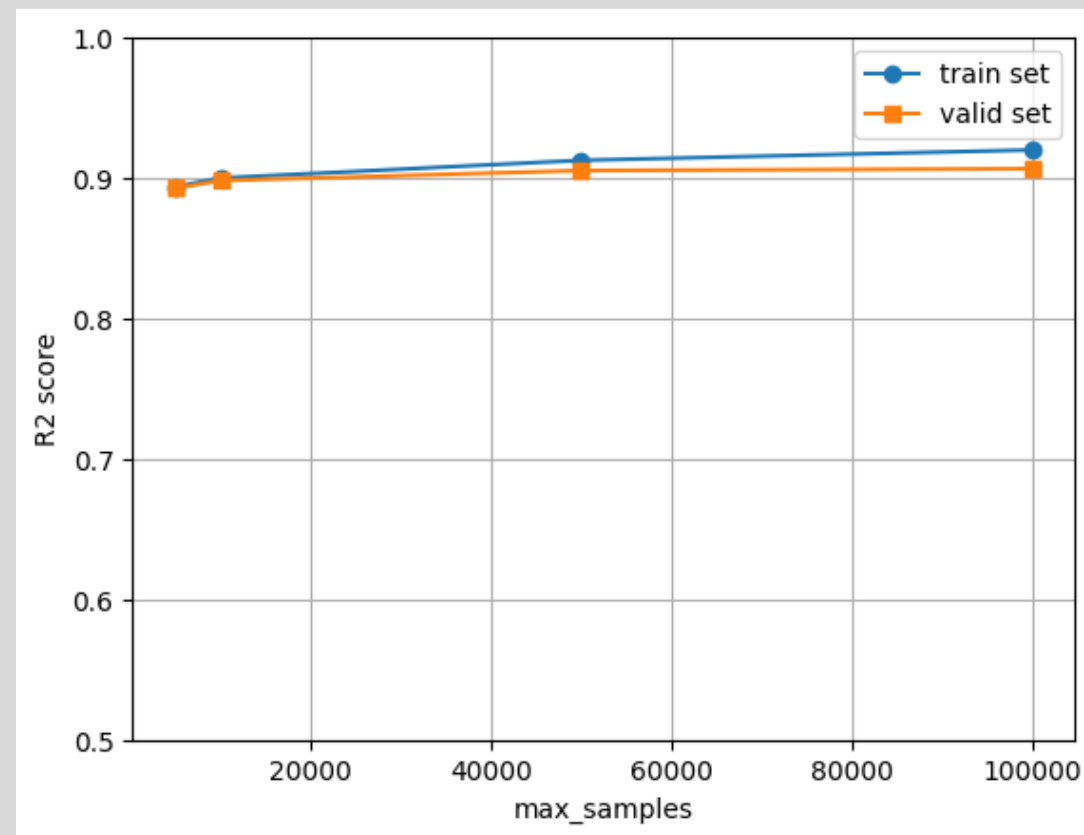
- 固定超參數
 - 取樣數:50000
 - 每葉最少樣本數:7
- 改變超參數
 - 樹的數量(100, 500, 1000, 2000)
- 結果
 - 達到一定數量後對結果僅些微差別
 - 以500作為最終超參數



| 樹數量 | Train set | Valid set |
|------|-----------|-----------|
| 100 | 0.91223 | 0.90508 |
| 500 | 0.91250 | 0.90520 |
| 1000 | 0.91253 | 0.90524 |
| 2000 | 0.91255 | 0.90527 |

隨機森林-調整取樣數

- 固定超參數
 - 樹數量:500
 - 每葉最少樣本數:7
- 改變超參數
 - 每次取樣數
(5000, 10000, 50000, 100000)
- 結果
 - 取樣越多時分數些微提升且趨於平緩
 - 以100000作為最終超參數



| | Train set | Valid set |
|--------|-----------|-----------|
| 5000 | 0.8935 | 0.8926 |
| 10000 | 0.8998 | 0.8980 |
| 50000 | 0.9125 | 0.9052 |
| 100000 | 0.9198 | 0.9065 |

Gradient boost

- 固定超參數
 - estimators:1000
 - 每顆樹深(max_depth):5
- 額外實作early stopping
 - 因此不對estimators數進行手動調整

結果分析

各模式子集R2分數

- 僅列出valid分數

隨機森林

| 模式 | R2分數 |
|--------|--------|
| 四人第一人稱 | 0.9141 |
| 雙人第一人稱 | 0.9452 |
| 單人第一人稱 | 0.9599 |
| 四人 | 0.9065 |
| 雙人 | 0.9393 |
| 單人 | 0.9363 |
| 其他 | 0.8836 |

Gradient boost

| 模式 | R2分數 |
|--------|--------|
| 四人第一人稱 | 0.9124 |
| 雙人第一人稱 | 0.9468 |
| 單人第一人稱 | 0.9624 |
| 四人 | 0.9060 |
| 雙人 | 0.9417 |
| 單人 | 0.9266 |
| 其他 | 0.8850 |

“其他”模式因樣本數較少且是多個模式混和，因此有些許Overfitting



名次預測最終預測器

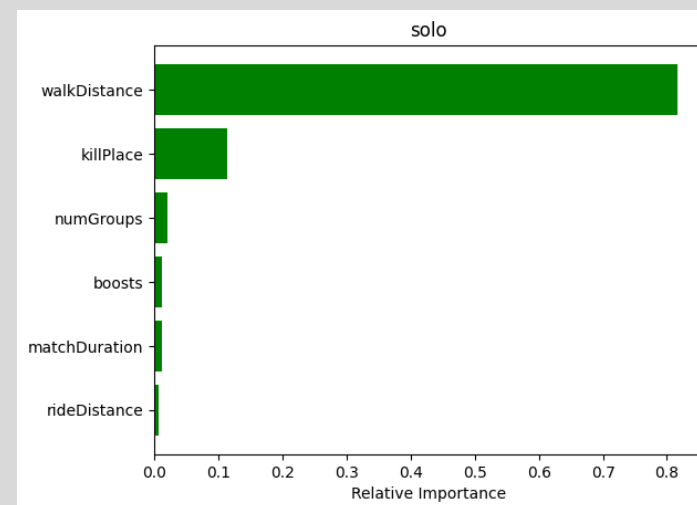
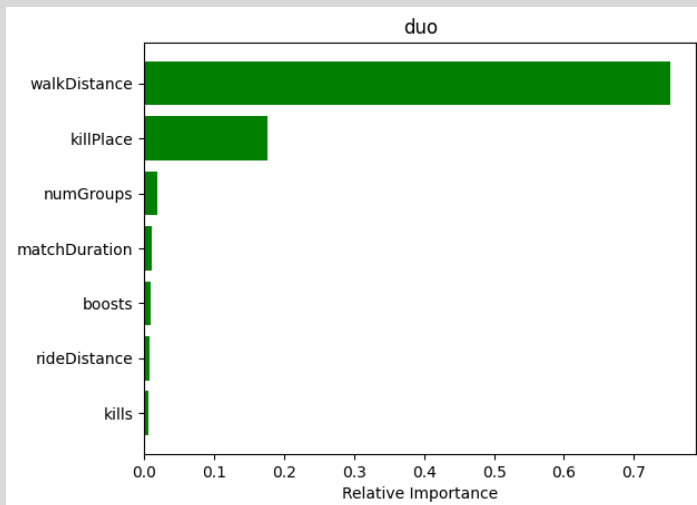
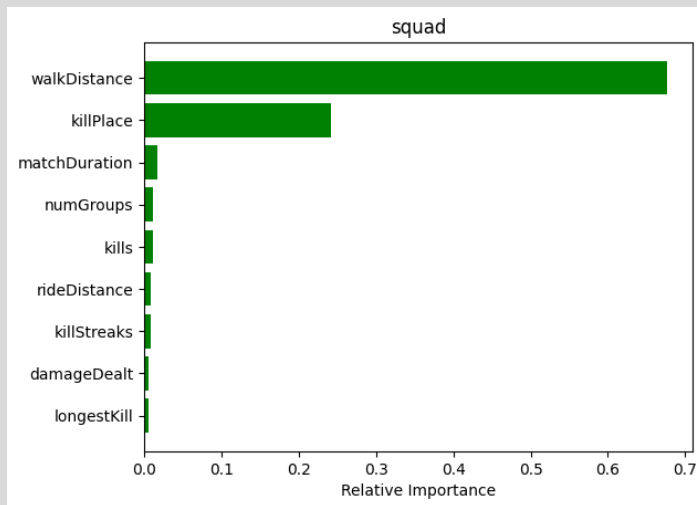
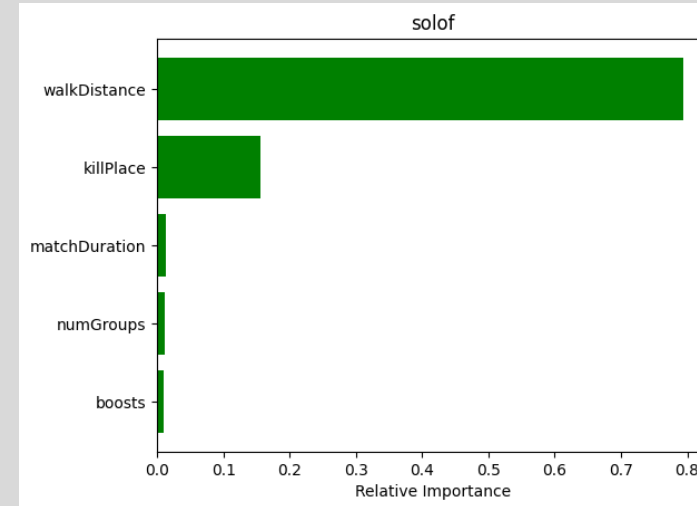
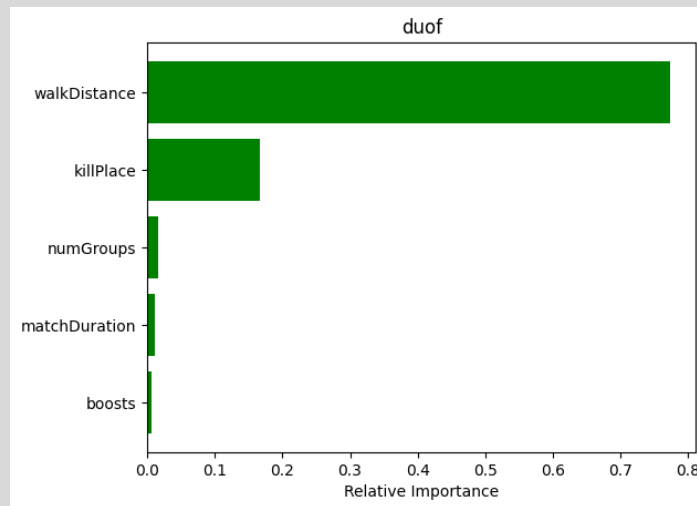
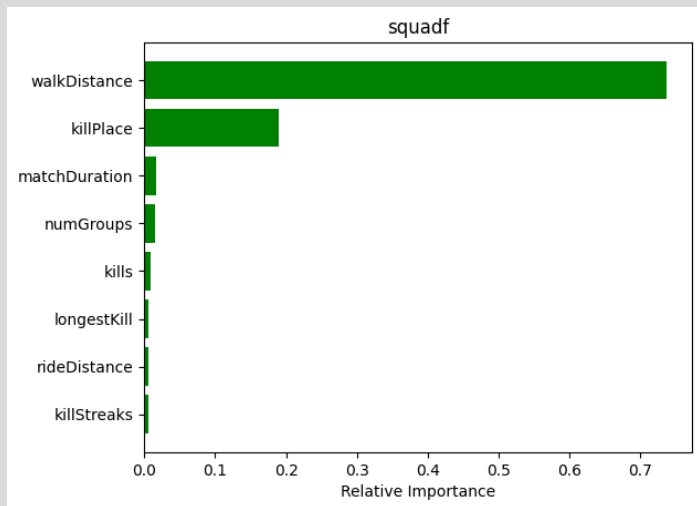
- 由隨機森林與gradient boost集成而來
 - 因為兩者成效皆不錯
 - 兩者結果取平均
- 用所有test data進行測試
- 整體R2分數為0.9277

R2 Score=0.9277025160783797



特徵重要性分析

- 僅列出決策樹中重要性超過0.05的特徵



六種基本模式中，皆是行走距離與擊殺數排名對最終名次特別重要

任務2 - 模式分類預測

任務概述

- 利用各項數據及最終排名，推測他遊玩的模式
- 任務類型
 - 分類任務
- 評估標準
 - Accuracy(調參與選擇模型)
 - Weighted F1 Score(選擇模型)



資料前處理

資料前處理

- 遊戲模式處理
 - 僅分成”四人模式” ”雙人模式” ”單人模式” ”其他模式”
 - 將模式處理後拿出來做標籤
- 去除離群值
- 去除能明顯識別模式的特徵
 - ex: 隊伍數量
- 資料集80%做訓練集，10%做驗證集，10%做測試集
- 訓練集資料平衡
 - 將較少的類別以SMOTE創造樣本
 - 將較多的類別刪去部分以達到平衡
 - 後續判斷原資料與平衡資料優劣

| matchType | |
|-----------|---------|
| squad | 1917820 |
| duo | 1052035 |
| solo | 575448 |
| other | 7695 |

平衡前



| matchType | |
|-----------|---------|
| squad | 1000000 |
| duo | 1000000 |
| solo | 1000000 |
| other | 1000000 |

平衡後

建立模型與調整

候選模型

- 決策樹
- 隨機森林
- Adaboost
 - 分類常用boosting模型

決策樹

- 調整超參數
 - 每葉最少樣本數(2, 3, 5, 7, 9, 10, 15, 20, 25, 30)
- 原資料集與平衡資料集皆進行測試
- 利用網格搜尋法套件

| 原資料集 | Train set | Valid set |
|----------|-----------|-----------|
| 最佳超參數:15 | 0.8740 | 0.8294 |

平衡過的資料集嚴重過擬和，推測是平衡時產生大量過於相似的樣本導致，因此後續不採用此資料集做訓練

| 平衡後資料集 | Train set | Valid set |
|---------|-----------|-----------|
| 最佳超參數:9 | 0.8926 | 0.7546 |

Accuracy



隨機森林

- 調整超參數
 - 樹數量(`n_estimators`)
 - 每顆樹的取樣數(`max_samples`)
 - 每葉最少樣本數(`min_samples_leaf`)
- 僅對原資料集進行測試
- 手動進行超參數調整

隨機森林-調整超參數

- 樹數量測試
 - 固定:每顆樹的取樣數=100000、每葉最少樣本數=7
 - 改變:樹數量(100, 500, 1000, 2000)
- 取樣數測試
 - 固定:樹數量=500、每葉最少樣本數=7
 - 改變:每顆樹的取樣數(50000, 100000, 500000, 1000000)
- 每葉最少樣本數測試
 - 固定:樹數量=500、每顆樹的取樣數=100000
 - 改變:每葉最少樣本數(5, 7, 10, 15, 20)

隨機森林-調整超參數

| 樹數量 | Train set | Valid set |
|------|-----------|-----------|
| 100 | 0.6958 | 0.6914 |
| 500 | 0.6982 | 0.6936 |
| 1000 | 0.6985 | 0.6938 |
| 2000 | 0.6990 | 0.6941 |

| 每葉最少樣本 | Train set | Valid set |
|--------|-----------|-----------|
| 5 | 0.7775 | 0.7264 |
| 7 | 0.7607 | 0.7224 |
| 10 | 0.7457 | 0.7183 |
| 15 | 0.7318 | 0.7124 |
| 20 | 0.7232 | 0.7082 |

Accuracy

| 取樣數 | Train set | Valid set |
|---------|-----------|-----------|
| 50000 | 0.6858 | 0.6839 |
| 100000 | 0.6982 | 0.6936 |
| 500000 | 0.7365 | 0.7153 |
| 1000000 | 0.7607 | 0.7224 |

- 樹的數量對成果影響不大
- 每棵樹取樣數越高成效越好
- 每葉樣本數越少成效越好，但逐漸過擬和
- 隨機森林最終模型
 - 樹數量=500
 - 取樣數=1000000
 - 每葉最少樣本=5



Adaboost

- 固定超參數
 - `n_estimators`:1000
 - 樹最大深度:10
- 僅對原資料集進行測試
- 該套件自帶early stopping

| Train set | Valid set |
|-----------|-----------|
| 0.6745 | 0.6680 |

Accuracy



結果分析

模式預測最終預測器

- 以決策樹做最終預測器
 - Accuracy與F1 score皆高於其他兩者

決策樹

accuracy:0.8294700199686977

f1 score:0.8286023038451655

隨機森林

accuracy:0.7264243438213971

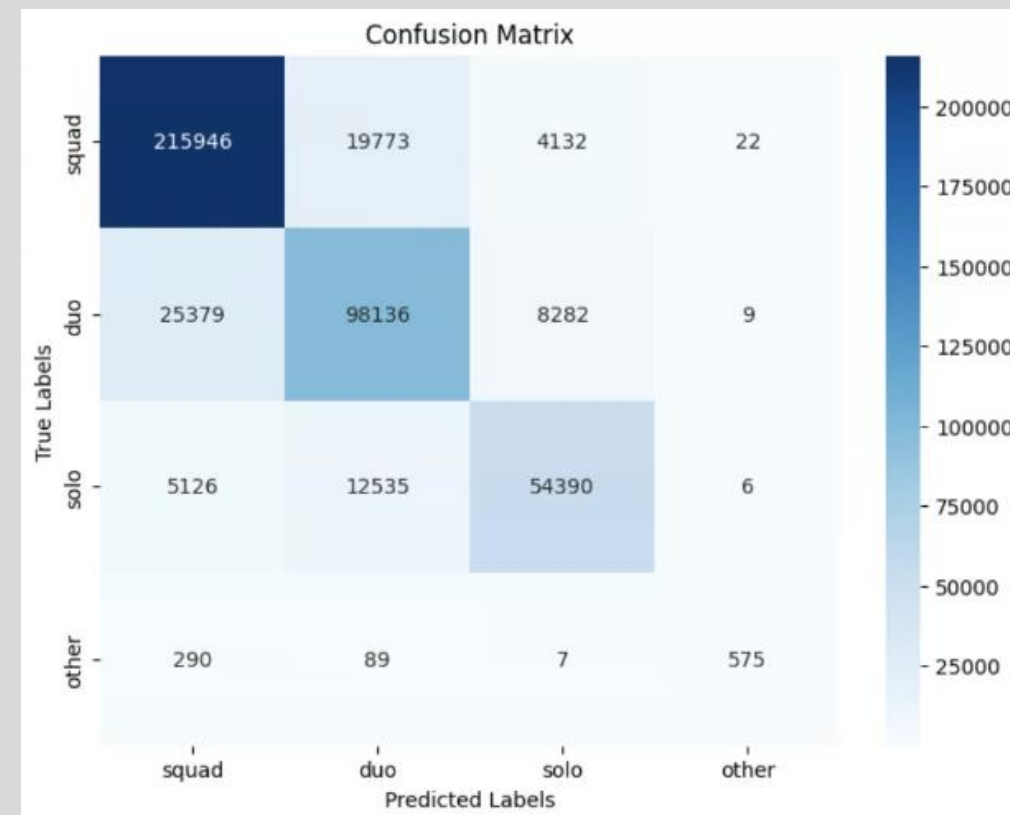
f1 score:0.7097229499680151

adaboost

accuracy:0.6680878622699552

f1 score:0.6355624443911883

Validation data



accuracy:0.8298841683213514

f1 score:0.82900859113023

決策樹預測test data



總結

DEMO影片:

<https://www.youtube.com/watch?v=GQMpzWUALYU>

總結

- 預測最終排名
 - 先依模式分出多個子集
 - 使用隨機森林與gradient boost取平均
 - Test data的R2分數達0.9277
- 預測遊玩模式
 - 使用決策樹
 - Test data的F1分數達0.8298
- 對於選手訓練
 - 根據特徵重要性發現移動距離與擊殺排名對名次影響大
 - 移動距離
 - 存活越久越有可能移動較遠
 - 多次轉移陣地
 - 擊殺排名
 - 擊敗越多玩家越有機會躋身前面名次



簡報到此結束



※備註:kaggle該比賽好像關閉上傳功能無法上傳評分了

PLAYERUNKNOWN'S
BATTLEGROUNDS