



배틀그라운드 순위 예측



CONTENTS

01 개요

- 게임설명
- 데이터 설명
- 프로젝트 목표설정

02 데이터 전처리

- 결측치/이상치처리
- 필요데이터추출

03 변수 선택

- 다중공선성

04 모델 생성 및 평가

- 모델 생성
- 모델 평가

05 데이터 시각화

- 데이터 요약

06 결론

- 결론
- 한계점



01 개요



01 개요

배틀그라운드란?



살아남기 위한 생존 게임

- 배틀 로얄 형식의 슈팅 게임
-> 최대 100명의 플레이어들과 생존을 위해 경쟁
- 플레이어들은 생존전략과 기술을 활용
-> 은둔형/호전형, 저격/근거리 등 다양한 전략 존재
- 다양한 요인이 승리에 영향



01 개요

배틀그라운드란?



01 개요

데이터셋



Id	groupId	matchId	assists	boosts	damageDealt	DBNOs	headshots	heals	killPlace	killPoints	kills	killStreaks	longestKill	matchDur.	matchType	maxPlace	numGroup	rank	Points	revives	rideDistance	road
7f96b2f874d4b580da10357fd1			0	0	0	0	0	0	60	1241	0	0	0	1306	squad-fpp	28	26		-1	0	0	
eef90569b684d56564aeb375fc5			0	0	91.47	0	0	0	57	0	0	0	0	1777	squad-fpp	26	25		1484	0	0.0045	
1eaf90ac76a4a42c32110163d88			1	0	68	0	0	0	47	0	0	0	0	1318	duo	50	47		1491	0	0	
4616d365ca930a9c79f1f1f4ef41			0	0	32.9	0	0	0	75	0	0	0	0	1436	squad-fpp	31	30		1408	0	0	
315c96c26de04010b6dc8ff871a			0	0	100	0	0	0	45	0	1	1	58.53	1424	solo-fpp	97	95		1560	0	0	
ff79c12f32289a6836cbac52627a			0	0	100	1	1	0	44	0	1	1	18.44	1395	squad-fpp	28	28		1418	0	0	
95959be042c485a1aca8274e903			0	0	0	0	0	0	96	1262	0	0	0	1316	squad-fpp	28	28		-1	0	0	
311b84c6feaba5fcb7292611730			0	0	8.538	0	0	0	48	1000	0	0	0	1967	solo-fpp	96	92		-1	0	2004	
1a68204cc47cfbb04edf014fbee			0	0	51.6	0	0	0	64	0	0	0	0	1375	squad	28	27		1493	0	0	
e5bb5a43f759bb6f753d3031c79			0	0	37.27	0	0	0	74	0	0	0	0	1930	squad	29	27		1349	0	0	
2b574d43fc549efede2dd6ddb8			0	0	28.38	0	0	0	75	0	0	0	0	1811	squad-fpp	29	29		1475	0	0	
8de328a7f643df9df80170383c			0	0	137.9	1	0	0	64	0	0	0	0	1384	duo-fpp	48	46		1488	0	0	
ce4f6ac16da24cdbc9535b5dbd			0	0	0	0	0	0	37	0	0	0	0	1774	squad-fpp	29	28		1766	0	6639	
b780718643c08e46182c30ddf48			0	1	324.2	0	1	5	5	986	4	1	49.83	1886	solo-fpp	97	94		-1	0	1228	
8e244ac61d40d0c7d94e1c1cc4			0	1	122.8	1	0	2	25	1411	1	1	37.91	1458	squad-fpp	31	30		-1	1	1237	
12d8d4bdfes52d481b6fd9e765c			0	0	80.71	1	0	0	72	0	0	0	0	1313	duo-fpp	48	48		1495	1	0	
62f2f0917f61b698271d6cfe0f6f			0	2	81.71	1	0	14	25	0	1	1	9.158	1882	squad-fpp	27	26		1436	0	519.9	
92022479f2f2c33f54f07948d72f			0	3	254.3	0	0	12	13	0	2	1	36	1371	solo-fpp	96	95		1536	0	2367	
7bd224786dde607d733af30ccf			0	0	0	0	0	0	79	0	0	0	0	1301	squad-fpp	27	26		1355	0	0	
71cbdbc317b61f74b5a329ac994			0	1	65.28	0	0	1	48	1349	0	0	0	1322	squad-fpp	30	28		0	0	0	
02ace8c6ea4bc5480280f2b8448			0	4	269.1	0	1	8	18	0	2	1	7.438	1890	duo-fpp	48	48		1491	1	2734	
00341b1ccd661a2d1f3956286e			0	0	158.7	1	0	0	75	0	0	0	0	1853	squad-fpp	29	28		1522	0	0	
9b2961d4f799d0a4dfe833ca228			0	1	192.3	1	2	3	15	1022	2	1	280.6	1775	duo-fpp	49	48		-1	0	2332	
0b6fbdfb57a75c3e8f8b0a78c0c			0	6	1011	6	2	2	2	1564	9	2	186.6	1857	duo	50	45		-1	0	4860	
736eda9bd35e80e4c62f7be7260			0	3	327.6	4	1	1	3	1346	4	2	167.8	1454	squad-fpp	27	27		-1	0	0	

29(Columns) x 440만(rows)

• 29개 컬럼

-> ID(user, game, match) , 게임 내 플레이 데이터, 게임 종류(solo, squad, duo), 순위 기록

• 440만개 행

-> 수 많은 게임들에 대한 정보를 데이터화

- 데이터 출처 : Kaggle

01 개요

데이터 설명: 전체 컬럼 설명_1



변수 이름	데이터 유형	데이터 설명
ASSISTS	숫자	자신에 의해 사살되지는 못했지만 자신의 데미지 지분율이 가장 높은 적의 수
BOOSTS	숫자	도핑 아이템 사용 횟수(에너지 드링크, 진통제, 아드레날린 주사기)
DAMAGEDEALT	숫자	적에게 가한 전체 데미지
DBNOS	숫자	기절시킨 적의 수
GROUPID	문자열	경기 내 그룹 식별 ID
HEADSHOTKILLS	숫자	헤드샷으로 죽인 적의 수
HEALS	숫자	회복 아이템 사용 횟수(붕대, 구급상자, 의료용 키트)
ID	문자열	해당 데이터 ID
KILLPLACE	숫자	경기 내 적을 죽인 적 수의 순위
KILLPOINTS	숫자	유저의 ELO레이팅(죽인 적 수 기반)

01 개요

데이터 설명: 전체 컬럼 설명_2



변수 이름	데이터 유형	데이터 설명
KILLS	문자	경기 내 적을 죽인 수
KILLSTREAKS	숫자	짧은 시간 내에 연속으로 적을 죽인 수
LONGESTKILL	숫자	사살한 적까지의 거리의 최대값
MATCHDURATION	숫자	경기가 진행된 시간(단위:초)
MATCHID	문자열	경기 식별 ID
MATCHTYPE	숫자	경기 모드(ex: 솔로, 듀오, 스쿼드)
MAXPLACE	숫자	경기 내 총 인원 수
NUMGROUP	숫자	경기 내 실제 참여 인원 수
RANKPOINTS	숫자	유저의 ELO 레이팅
REVIVES	숫자	유저가 팀원을 부활시킨 횟수



변수 이름	데이터 유형	데이터 설명
RIDEDISTANCE	숫자	이동수단을 통해 이동한 거리(단위:m)
ROADKILLS	숫자	이동수단으로 살해한 적의 수
SWIMDISTANCE	숫자	수영으로 이동한 거리(단위:m)
TEAMKILLS	숫자	같은 팀원을 살해한 수
VEHICLEDESTROYS	숫자	파괴한 이동수단의 수
WALKDISTANCE	숫자	도보로 이동한 거리(단위:m)
WEAPONSACQUIRED	숫자	획득한 무기의 수
WINPOINTS	숫자	유저의 ELO 레이팅(승리 횟수 기반)
WINPLACEPERC	숫자	현재 경기에서의 백분위 기반 유저의 순위(종속변수)



게임 데이터를 활용한 순위 예측
&
Top 10 및 최후의 1인이 되기 위한 주요 전략 분석



02

데이터 전처리



02 데이터 전처리 분석 환경



Version 3.4.3



Version 3.6.5

Oracle
SQL Developer



ORACLE

Copyright © 2006, 2007 Oracle Corporation. All rights reserved.

Version 18.2.0



```
In [8]: game[game.winPlacePerc.isna()]
Out[8]:
```

	Id	groupId	...	winPoints	winPlacePerc
2744604	f70c74418bb064	12dfbede33f92b	...	0	NaN

[1 rows x 29 columns]

```
In [9]: game[game.winPlacePerc.isna()].loc[:, ['Id', 'winPlacePerc']]
Out[9]:
```

	Id	winPlacePerc
2744604	f70c74418bb064	NaN

결측치 처리

- 종속 변수인 winPlacePerc에 NaN이 존재
- 삭제 처리

02 데이터 전처리

이상치 삭제



ID	WALKDISTANCE	WEAPONSACQUIRED
1 0df7c355497d77	0	6
2 0a5036716cd6e0	0	1
3 72f70c6c074ca9	0	15
4 269108abb1ba4e	0	1
5 62f59fc192169d	0	2
6 bfb4ea17cb8c57	0	1
7 3fc123559fc935	0	3
8 7ae965d09812d8	0	6
9 89546537d383a0	0	1
10 5ea53b3bf3ec57	0	2
11 a687cd61cde2e1	0	1
12 854a8588d4ef5	0	1
13 a3cdb1386fe019	0	4
14 fd173cc3fa9c2b	0	2
15 5882729c06903b	0	3
16 d279587369ea17	0	1
17 ce4d3e2a1e83f5	0	1
18 f8d9246391c1a7	0	5
19 2e0ff0f55ed189	0	1
20 6dfb288f1b1a8b	0	1
21 82a8e0b14b9b03	0	1
22 a9dd9c28801581	0	3
23 bc558bc2c90812	0	4
24 Odd20907726050	0	1
25 0351565a7058e9	0	9
26 e23fbee17ece0e	0	2
27 36f5637d206f9f	0	1

이상치 처리

- 움직이지 않았는데, 무기를 소지한 유저는
불법 프로그램 사용자로 판단

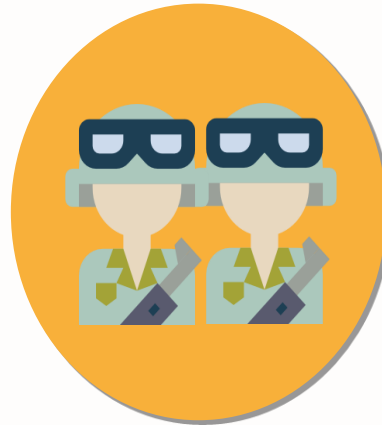
- 분석에 악영향을 줄 수 있으므로 삭제 처리
-> 약 1만 1천명



MATCHTYPE



Solo 



Duo 



Squad 

필요 데이터 추출_불필요 컬럼 제거



변수 이름	제외 이유
ID	구분 짓기 위한 변수들로서 종속 변수에 영향을 끼치지 않음
GROUPID	
MATCHID	
DBNOS	solo 게임에서는 발생하지 않음
REVIVES	
TEAMKILLS	
KILLPOINTS	과거 랭크 기록이므로 각 게임 순위 결정에 영향이 없음
RANKPOINTS	
WINPOINTS	
MATCHDURATION	MATCHID별로 동일한 값을 가지므로 순위에 영향을 끼치지 않음
MATCHTYPE	앞에서 SOLO만 추출했기 때문에 의미 없음
NUMGROUP	MAXPLACE와 비슷한 의미를 가지므로 제거



03

변수 선택





처음

```
var      vif
assists  1.10901682619096
boosts   2.38969509067496
damageDealt 10.8377163669476
headshotkills 2.01178210656268
heals    1.37779042073024
killPlace 7.1792432660015
kills    13.1875567851574
killstreaks 4.36619042497667
longestkill 1.89635559706679
maxPlace 1.09333893031163
rideDistance 1.18705689689649
roadkills 1.04158225881761
swimDistance 1.03841298448851
vehicleDestroys 1.01742904333295
walkDistance 2.54783470787003
weaponsAcquired 1.79236033000172
```

removed: kills 13.18756



결과

```
var      vif
assists  1.06977485990246
boosts   2.34932298062634
damageDealt 3.73468252560062
headshotkills 1.87684686020788
heals    1.37710285890932
killPlace 7.10534181614396
killstreaks 4.27019027718423
longestkill 1.85228526334138
maxPlace 1.09047315346495
rideDistance 1.18322259254492
roadkills 1.04015219117696
swimDistance 1.03840942342775
vehicleDestroys 1.01733631608423
walkDistance 2.53996054147261
weaponsAcquired 1.78383884620509
```

Result

```
"assists"      "boosts"      "damageDealt"  "headshotkills"
"heals"        "killPlace"   "killstreaks"  "longestkill"
"maxPlace"     "rideDistance" "roadkills"    "swimDistance"
"vehicleDestroys" "walkDistance" "weaponsAcquired"
```



처음

```
var      vif
assists  1.10901682619096
boosts   2.38969509067496
damageDealt 10.8377163669476
headshotKills 2.01178210656268
```

결과

```
var      vif
assists  1.06977485990246
boosts   2.34932298062634
damageDealt 3.73468252560062
```

하지만 서바이벌 게임 승률 예측에서
Kills를 뺄 수 없음



Kills와 다른 변수들의 상관 관계 확인

```
removed: kills 13.18756
```

```
weaponsAcquired 1.78383884620509
```

Result

```
"assists"      "boosts"      "damageDealt"  "headshotKills"
"heals"        "killPlace"   "killstreaks"  "longestKill"
"maxPlace"     "rideDistance" "roadKills"    "swimDistance"
"vehicleDestroys" "walkDistance" "weaponsAcquired"
```

03 변수선택

상관관계표 - HEAT MAP





VIF 확인 결과

kills 유지

damageDealt 제거

killPlace 제거

var	vif
assists	1.03324224062665
boosts	2.38405127471651
headshotKills	2.00034203486896
heals	1.37516559879596
kills	4.25638303142721
killstreaks	2.19670063004277
longestkill	1.89573504018442
maxPlace	1.00232873612825
rideDistance	1.17786931552879
roadkills	1.03625128116578
swimDistance	1.03716038137548
vehicleDestroys	1.01711671708709
walkDistance	2.19614947135184
weaponsAcquired	1.54373307024381

최종 변수 선택

"assists"	"boosts"	"headshotKills"	"heals"
"kills"	"killstreaks"	"longestkill"	"maxPlace"
"rideDistance"	"roadkills"	"swimDistance"	"vehicleDestroys"
"walkDistance"	"weaponsAcquired"		

03 변수 선택

최종 독립변수_14개



변수 이름	데이터 유형	데이터 설명
ASSISTS	숫자	자신에 의해 사살되지는 못했지만 자신의 데미지 지분율이 가장 높은 적의 수
BOOSTS	숫자	도핑 아이템 사용 횟수(에너지 드링크, 진통제, 아드레날린 주사기)
HEADSHOTKILLS	숫자	헤드샷으로 죽인 적의 수
HEALS	숫자	회복 아이템 사용 횟수(붕대, 구급상자, 의료용 키트)
KILLS	문자	경기 내 적을 죽인 수
KILLSTREAKS	숫자	짧은 시간 내에 연속으로 적을 죽인 수
LONGESTKILL	숫자	사살한 적까지의 거리의 최대값
MAXPLACE	숫자	경기 내 총 인원 수
RIDEDISTANCE	숫자	이동수단을 통해 이동한 거리(단위:m)
ROADKILLS	숫자	이동수단으로 살해한 적의 수
SWIMDISTANCE	숫자	수영으로 이동한 거리(단위:m)
VEHICLEDESTROYS	숫자	파괴한 이동수단의 수
WALKDISTANCE	숫자	도보로 이동한 거리(단위:m)
WEAPONSACQUIRED	숫자	획득한 무기의 수



04

모델 생성 및 평가





릿지

- Parameters :
alpha = [0.01, 0.05, 0.1, 0.5, 1, 5, 10, 50, 100]

라쏘

- Parameters :
alpha = [0.01, 0.05, 0.1, 0.5, 1, 5, 10, 50, 100]

랜덤 포레스트

- Parameters :
n_estimators = [50, 100, 200]
max_depth = [3, 4, 5]

그라디언트 부스트

- Parameters :
n_estimators = [50, 100, 200]
learning_rate = [0.01, 0.05, 0.1, 0.5, 1]
max_depth = [3, 4, 5]

04 모델 생성 및 평가

4가지 모델 비교



릿지

- Opt params : $\alpha = 50$
- 훈련 스코어(%) : 76.11%
- 테스트 스코어(%) : 76.25%
- RMSE : $66.9e-05$

라쏘

- Opt params : $\alpha = 0.01$
- 훈련 스코어(%) : 75.51%
- 테스트 스코어(%) : 75.62%
- RMSE : $72.4e-05$

랜덤 포레스트

- Opt params : $\text{max_depth} = 5, \text{n_estimators} = 200$
- 훈련 스코어(%) : 86.69%
- 테스트 스코어(%) : 86.79%
- RMSE : $14.8e-05$

그라디언트 부스트

- Opt params : $\text{learning rate} = 0.1, \text{max_depth} = 5, \text{n_estimators} = 200$
- 훈련 스코어(%) : 88.47%
- 테스트 스코어(%) : 88.33%
- RMSE : $1.9e-05$

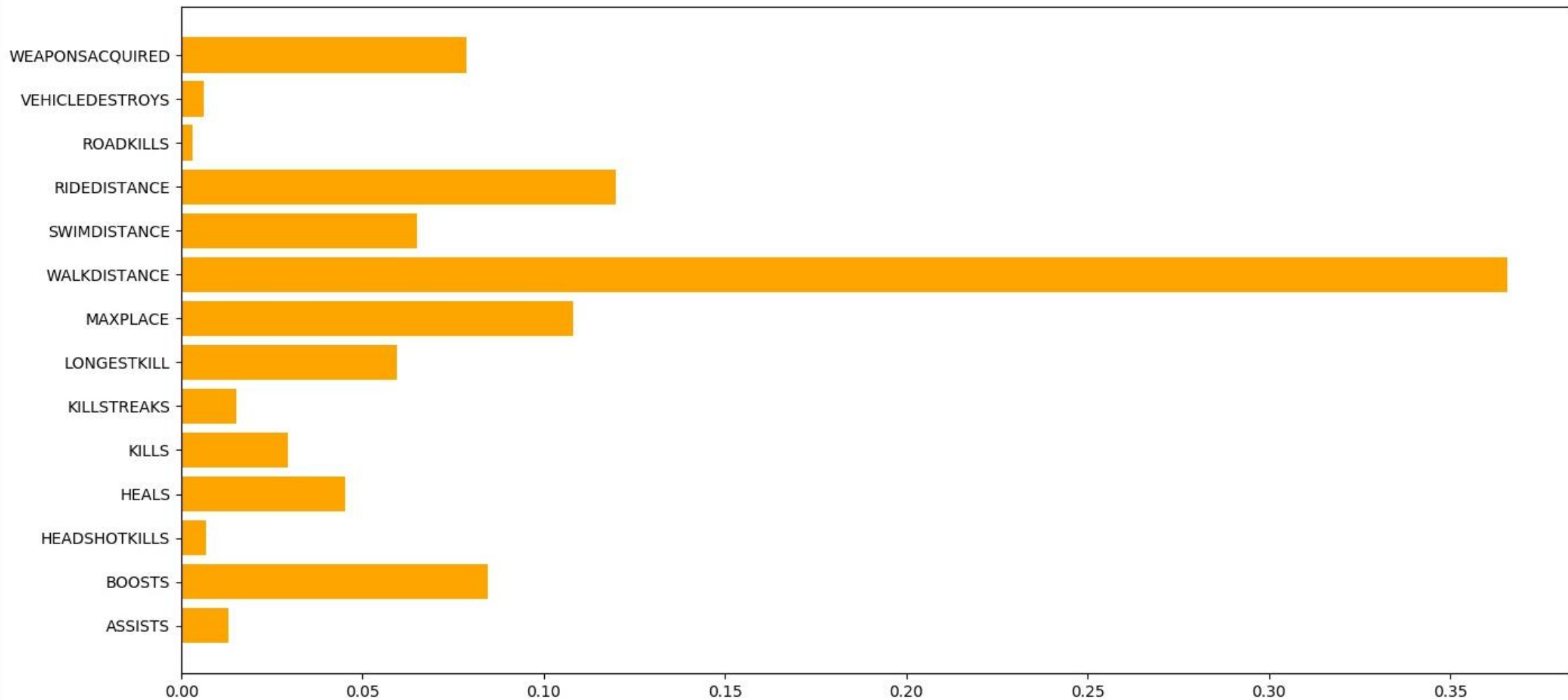


그라디언트 부스트

- Opt params : learning rate = 0.1
max_depth = 5
n_estimators = 200
- 훈련 스코어(%) : 88.47%
- 테스트 스코어(%) : 88.33%
- RMSE : $1.9e-05$

04 모델 생성 및 평가

변수 중요도 - 막대 그래프



04 모델 생성 및 평가

예측값 - 실제값 비교



데이터	Value1	Value2	Value3
ASSISTS	0	0	0
BOOSTS	5	4	6
HEADSHOTKILLS	0	0	0
HEALS	9	11	13
KILLS	3	1	5
KILLSTREAKS	0	0	1
LONGESTKILL	67	3	64
MAXPLACE	98	93	96
WALKDISTANCE	2750	2960	3580
SWIMDISTANCE	0	0	0
RIDEDISTANCE	9540	7960	8280
ROADKILLS	0	0	0
VEHICLEDESTROYS	0	0	0
WEAPONSACQUIRED	5	5	8
예측값	0.94	0.8913	0.9842
실제값	0.9285	0.8172	1



05

데이터 시각화

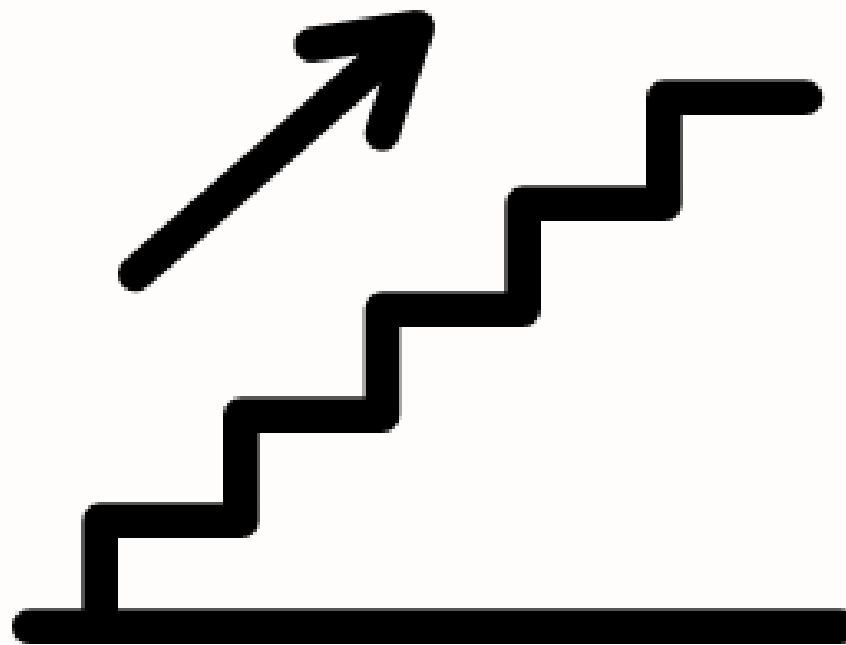




순위 백분위 구간화

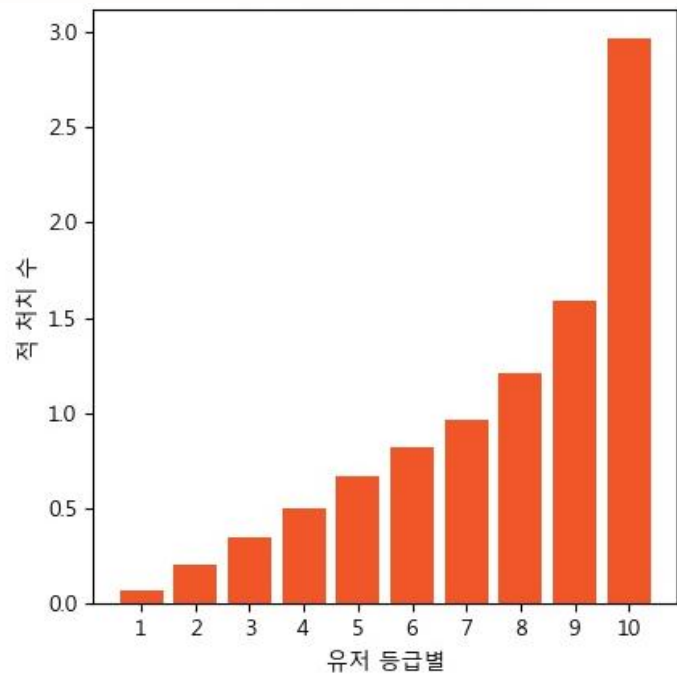
순위 백분위를 10개의 구간으로 나눔

- 1 = 0% ~ 10%
- 2 = 10% ~ 20%
- ⋮
- 10 = 90% ~ 100%

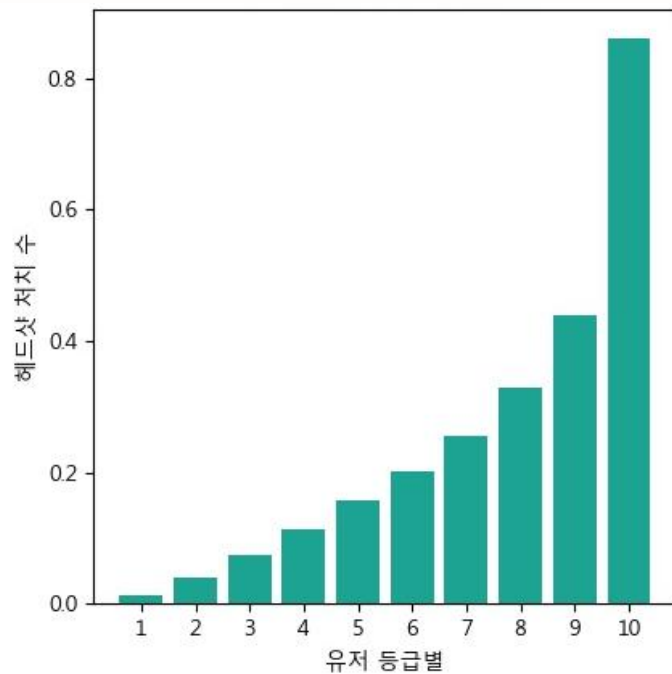




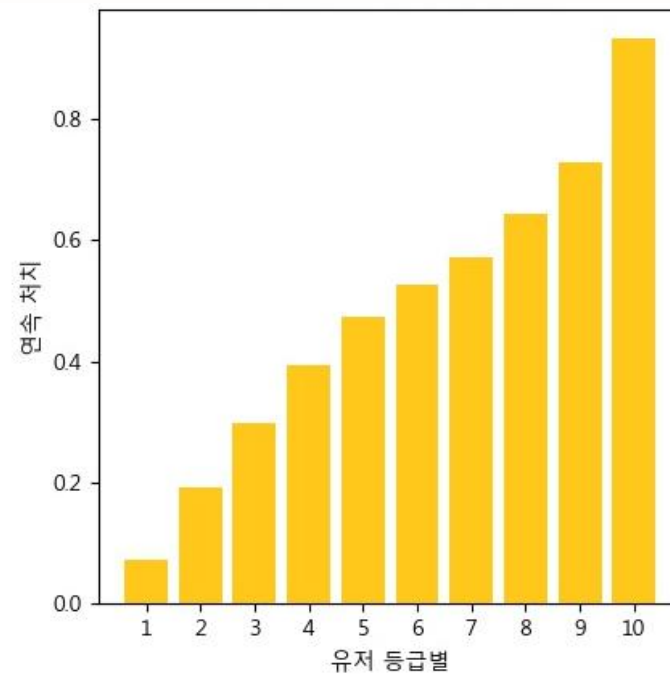
유저 등급별 KILL 횟수



헤드샷 횟수

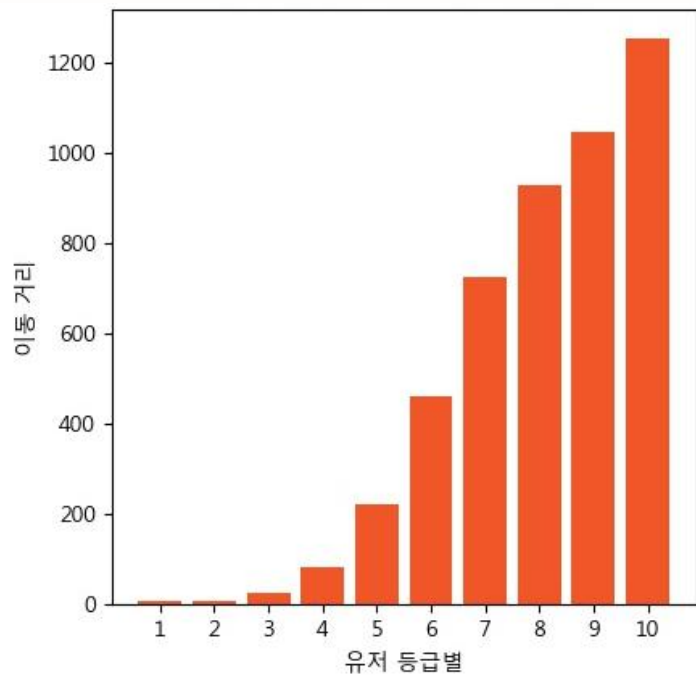


연속 KILL 횟수

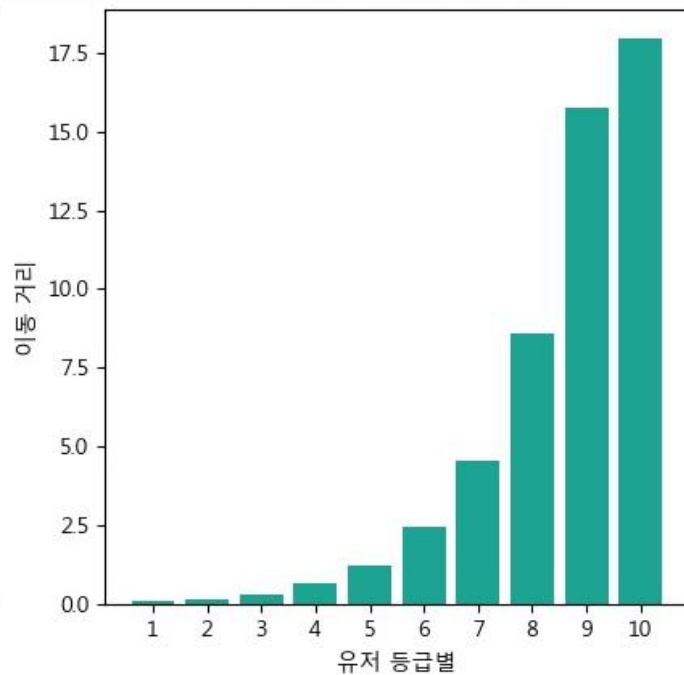




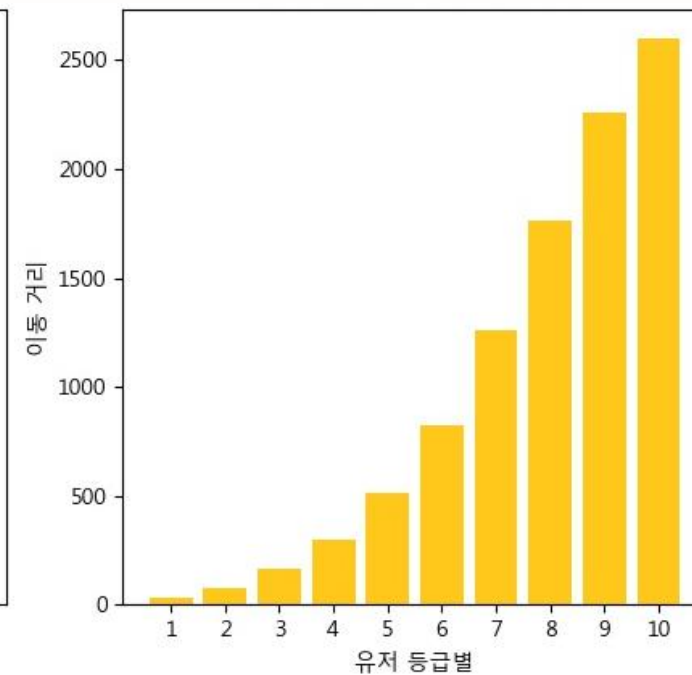
이동수단(탈것) 이동거리



수영 이동거리

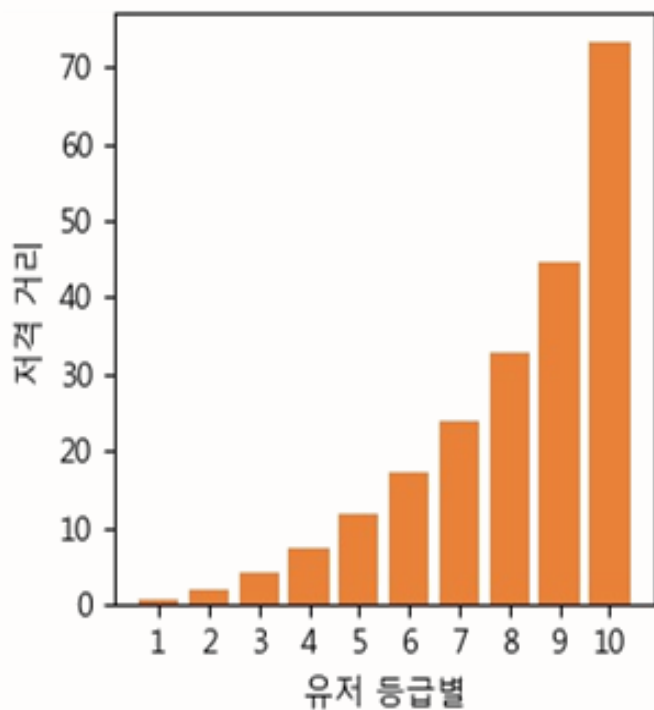


도보 이동거리

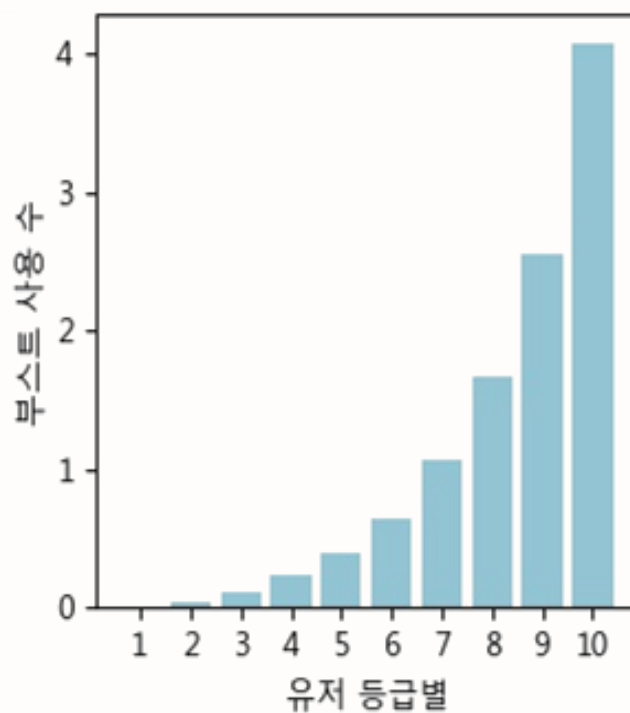




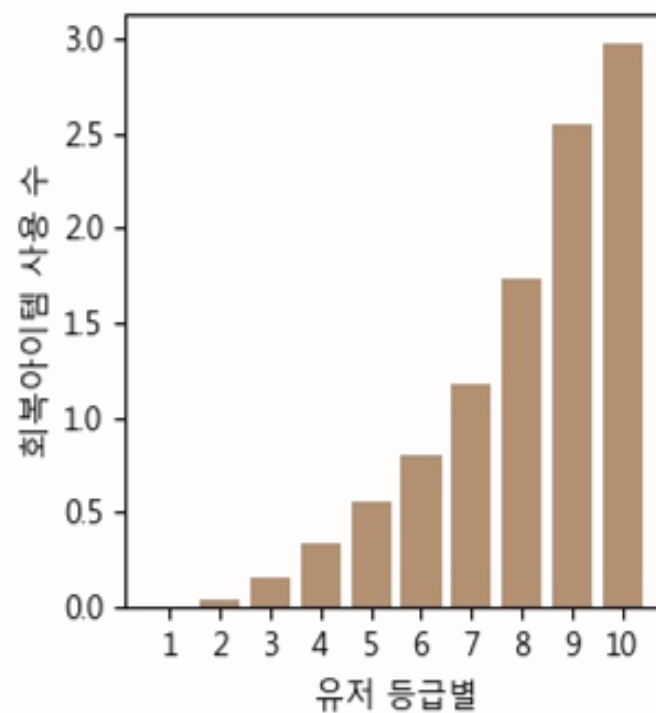
최장 저격 거리



부스트 사용 수

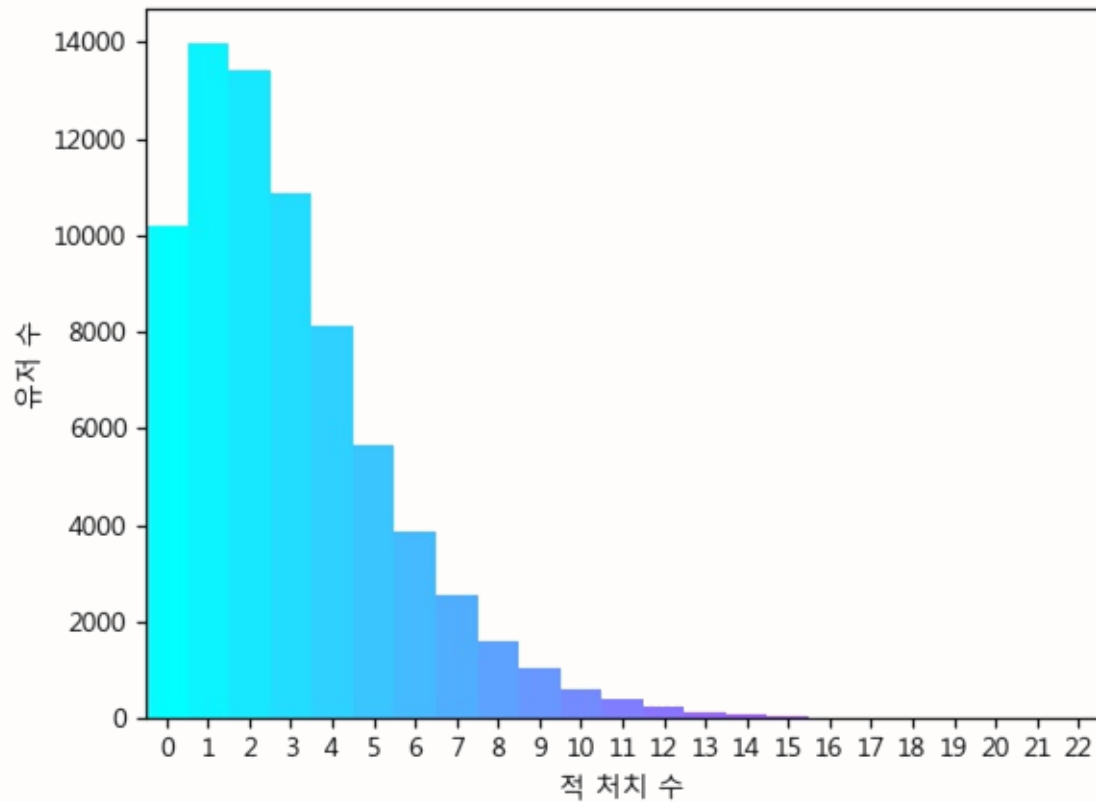


회복 아이템 사용 수

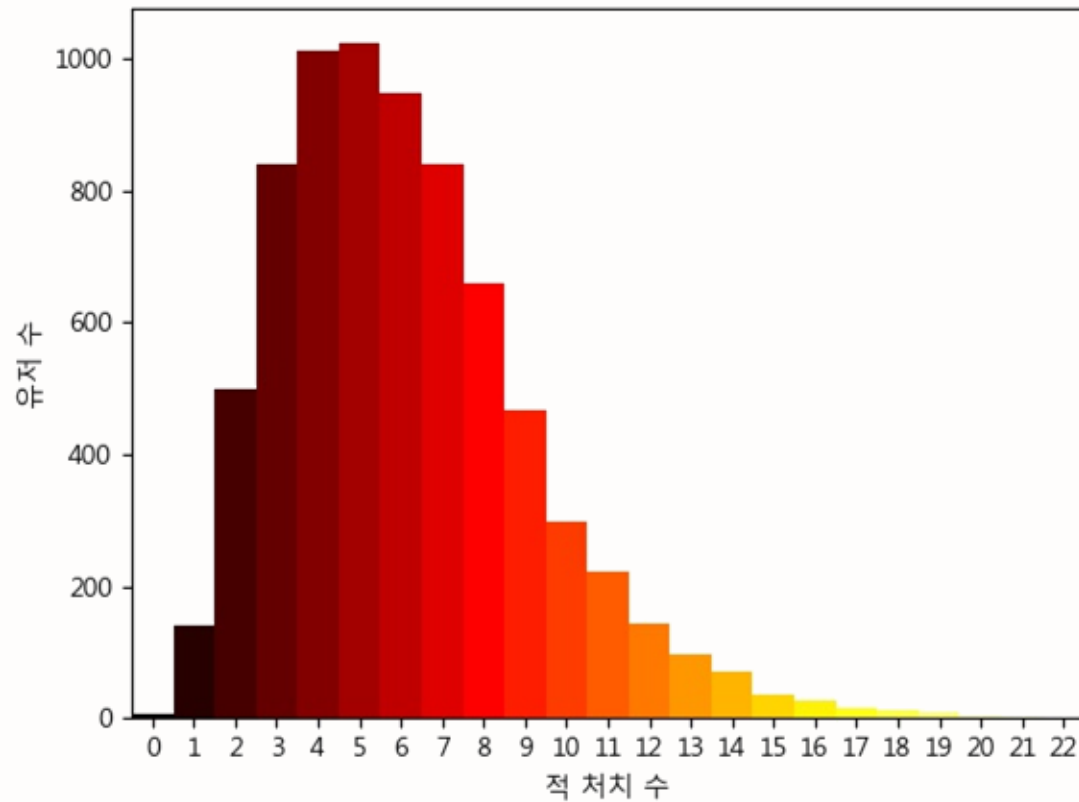




상위 10% 유저들의 KILLS 수



1위 유저들의 KILLS 수



05 데이터 시각화

변수별 최고치



업적	ID	KILLS	ROAD KILLS	HEADSHOT KILLS	HEALS	BOOSTS	WALK DISTANCE	RIDE DISTANCE	SWIM DISTANCE
최다 킬 수	f9d14251d99ecf	23	0	14	3	5	3324.0	2211.0	417.8
최다 뺑소니 수	c3e444f7d1289f	14	14	0	0	0	1277.0	0.005	5.297
최다 헤드샷 수	8e9eb1ce0e0135	21	0	19	2	7	3507.0	808.1	0.0
최다 회복 횟수	1b1d77a24be1e1	0	0	0	57	5	3003.0	2583.0	24.97
최다 도핑 횟수	5b7d6f8755297b	0	0	0	13	28	2488.0	0.0	0.0
최대 도보 이동	3304d0c68e27f5	0	0	0	0	0	25780.0	0.0078	381.9
최대 승차 이동	afe34618b04d77	0	0	0	0	0	560.1	40710.0	0.0
최대 수영 이동	3e0ae6d0e5c266	0	0	0	0	0	18.68	0.0	1974.0



06

결론





결론

- Kaggle에서 제공한 데이터셋을 이용하였으며, 29개의 변수 중에서 Match Type과 다중공선성에 의거하여 14개의 변수를 선택
- 각 모델의 성능을 평가하여, 가장 성능이 우수한 그라디언트 부스트를 선택
- 선택된 변수와 모델에 대해서 외부 게임 데이터를 입력한 결과, 실제 예측이 가능함을 확인
- 데이터를 시각화하여 분석한 결과, 상위 10퍼센트의 유저는 은둔형 플레이와 호전형 플레이가 비슷한 분포를 보이나, 최후의 1인은 호전형 플레이가 더 비율이 높았음을 확인



한계점

- 다양한 맵이 존재함에도 불구하고 맵에 대한 정보가 부재
- 게임 내에서 전투구역을 제한시키는 요소(자기장)에 대한 정보 부재



감사합니다
Q&A