

Project report.

Kyungjun Kang
github.com/kangjun205

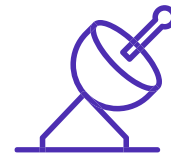
Mingyu Park
github.com/mg4432



Contents.

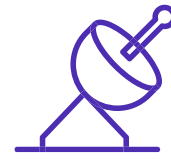
Contents of project reports.

Project description, project process,
our strategies, result & conclusion.



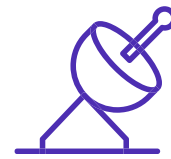
Project description

Project outline, purpose, weekly plan, etc.



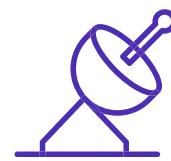
Project process

Hypotheses, EDA, feature engineering,
model selection, etc.



Our strategy

Strategies used to develop performance
and save time & memory.



Result & Conclusion

Our rank in LB, MAE
Most effective factors in predicting
finish placement.

Project Description

Project description.

Project outlinepurpose & weekly plan.

✓ 프로젝트 목적

- 기계학습을 통한 **매치 순위에 영향을 미치는 요인 파악**
- 2018년 대회 당시 리더보드 기준 **상위 5%(76/1528)**의 성능을 가진 모델 개발

✓ 주간 계획

Week 1 (2/27~3/5)

- ✓ 데이터 내 cheating 유저 파악
- ✓ EDA

Week 2 (3/6~3/12)

- ✓ Feature Engineering
- ✓ Baseline 모델 생성
- ✓ 분석 계획 수립

Week 3 (3/13~3/19)

- ✓ 분석 모델 선정
- ✓ 학습 및 성능 고도화

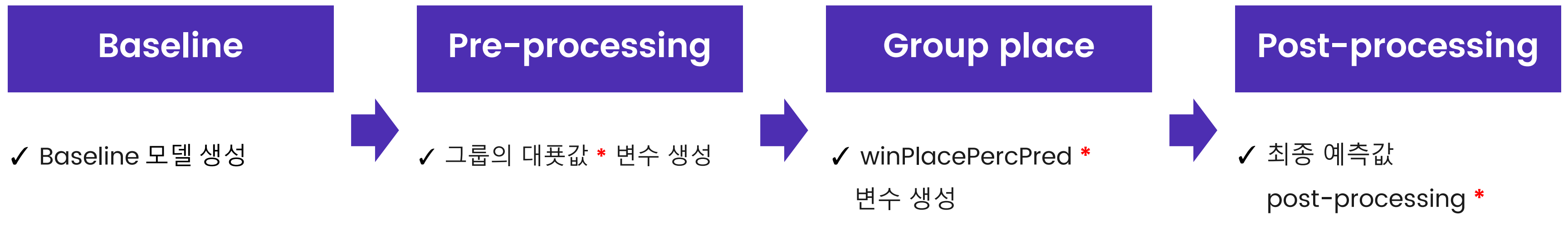
Week 4 (3/20~3/26)

- ✓ 분석 결과, 한계점 및 개선사항 정리
- ✓ 보고서 작성

Project Process

Project process.

Overall process.



* 해당 변수의 그룹별 평균값, 최댓값, 평균값의 순위, 최댓값의 순위를 의미함.

* winPlacePercPred 변수는 killPlace 변수를 바탕으로 추론한 그룹 순위를 의미함.

* post-processing은 continuous variable인 예측값을 이용해 그룹을 재정렬한 후 해당 매치에서 존재할 수 있는 순위로 재정렬한 것을 의미함.

Project process.

Hypotheses.

✓ Hypothesis 1

개인의 기록보다는 그룹 평균, 그룹의 최댓값 등 그룹의 대푯값이 최종 순위 예측에 더 효과적일 것임.

✓ Hypothesis 2

매치 내의 킬 순위를 나타내는 killPlace 변수의 변수중요도가 Baseline 모델에서 매우 높게 나타남. 따라서, 해당 변수가 최종 순위에 대해 중요한 정보를 갖고 있을 것임.

✓ Hypothesis 3

모든 기록은 그 자체의 기록보다 해당 매치에서의 순위가 최종 순위 예측에 더 효과적일 것임.

Ex. kill < 매치 내 kill 순위

Project process.

Hypotheses.

✓ Hypothesis 1

개인의 기록보다는 그룹 평균, 그룹의 최댓값 등
그룹의 대푯값이 최종 순위 예측에 더 효과적일 것임.

Result

그룹의 대푯값을 사용하는 것이 최종 순위 예측에 더 **효과적**임.

Reason

매치 순위와의 상관계수가 개인의 기록보다 그룹의 대푯값에서 더 높음.
예측 모델에서 그룹의 대푯값의 변수 중요도가 개인의 기록보다 우세함.

Project process.

Hypotheses.

✓ Hypothesis 2

매치 내의 킬 순위를 나타내는 killPlace 변수의 변수중요도가 Baseline 모델에서 매우 높게 나타남. 따라서, 해당 변수가 최종 순위에 대해 중요한 정보를 갖고 있을 것임.

Result

Kill이 같은 그룹 간에는 killPlace 변수를 통해 **순위 관계를 알 수 있음.**

Reason

Kill이 같은 그룹이 존재하는 경우 killPlace가 같지 않고,
매치 순위와 동일한 순서로 주어짐.

Project process.

Hypotheses.

✓ Hypothesis 3

모든 기록은 그 자체의 기록보다 해당 매치에서의 순위가 최종 순위 예측에 더 효과적일 것임.

Result

해당 매치에서의 순위를 사용하는 것이 최종 순위 예측에 더 **효과적**임.

Reason

각 매치마다 **양상이 다름**.

어떤 유저가 5 kill을 했다고 가정했을 때, 5 kill 이상을 기록한 유저가 많으면, 그만큼 순위가 낮아질 것으로 예상할 수 있으며, 5 kill 이상을 기록한 유저가 해당 매치에 존재하지 않으면, 그만큼 순위가 높아질 것으로 예상할 수 있음.

Project process.

Model selection.

✓ Model

- LightGBM regressor

✓ Why LightGBM?

High performance

✓ Purpose of competition
: high score
→ **boosting algorithm**

Fast

✓ Shows **high speed**
among boosting models.

Overfitting?

✓ Over 4,000,000 rows
→ **less sensitive to
overfitting**

Frequently used

✓ **Most frequently used**
in competitions
with xgboost, etc.

Our Strategy

Our Strategy.

Strategies used to develop performance and save time & memory.

✓ Reduce memory

메모리, 시간을 절약하기 위해 데이터 타입 변경

✓ Get ordered group places

killPlace 변수를 통해서 그룹 간의 순위 관계를 추론함.

✓ Post-processing

Continuous value인 예측값을 실제 그룹 수를 고려해

Discrete value로 변환함으로써 정확한 예측값을 계산함

Our Strategy.

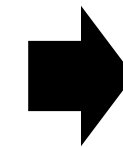
Reduce memory.

✓ description

- 변수 type 변경을 통한 memory & time 절약
- ex) float64 → float16

train

3339.02 MB

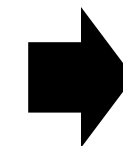


856.67 MB

Memory reduced
about **74.4%**

test

1446.14 MB



354.16 MB

Memory reduced
about **75.5%**

Our Strategy.

Get ordered group places.

- killPlace 변수를 통해서 그룹 간의 순위 관계를 추론함.

	groupId	kills	killPlace	winPlacePerc	groupId	kills	killPlace	winPlacePerc
0	3fe42857a5d0e3	0	40	1.0000	edbce070d4ee9d	0	68	0.2963
1	5112ebf44e9575	0	41	0.9630	1d31d15e045c4d	0	71	0.2593
2	69c3ba1a23df2a	0	44	0.9259	c92eb4405f4d3a	0	72	0.2222
3	74898d59ac4879	0	46	0.8519	2f8336736f3be2	0	74	0.1852
4	7e5fb71e05a678	0	48	0.8148	a27667dbecd7d7	0	76	0.1481
5	e93dfe8291d05c	0	49	0.7778	432b8256f23c7e	0	78	0.1111
6	168d1fc8ebe03c	0	52	0.7407	80d9f9014c6e68	0	82	0.0741
7	2af5c87d86546d	0	54	0.5556	a269f08371068b	0	89	0.0370
8	654c638629b8fc	0	55	0.4815	d7d1cd054b0f5b	0	94	0.0000
9	4d4b580de459be	0	60	0.4444	30e393959e45ac	1	27	0.8889
10	5ddf679672a972	0	63	0.4074	64903e9ba2e503	1	31	0.6296
11	e3b971e00acee4	0	65	0.3704	2cdf256e3c2c1d	1	32	0.5926
12	92a09f0a596787	0	67	0.3333	0b4bf93ca082a4	1	34	0.5185

Our Strategy.

Get ordered group places.

✓ case 1 : Can be inferred exactly

Kills	Group order
0	A, B, C, D
1	B, E, C
2	F, A
3	A

✓ 0 kill
A > B > C > D
✓ 1 kill
A > B > E > C > D
✓ 2 kill
F > A > B > F > C > D
✓ 3 kill
G > A > B > F > C > D



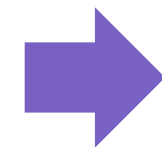
Group	winPlacePercPred
A	0.8
B	0.6
C	0.2
D	0
E	0.4
F	1

Our Strategy.

Get ordered group places.

✓ case 2 : Cannot be inferred exactly

Kills	Group
0	A, B, C, D, E
1	B, F, D

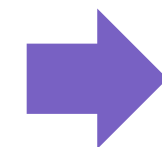


✓ maximum rank of F is 3

$A > B > F > C > D > E$

✓ minimum rank of F is 4

$A > B > C > F > D > E$



순위 관계가 명확하지 않은
F 그룹의 경우 3과 4의 평균인
3.5를 순위로 부여함

→ 순위관계가 명확하지 않은 경우, 다른 그룹과의 관계를 통해 최대 순위, 최소 순위를 구한 후,
두 값의 평균을 순위로 부여한 값을 변수로 사용했음.

Our Strategy.

Get ordered group places.

✓ case 2 : Cannot be inferred exactly

Group	Ordered rank
A	1
B	2
C	3
D	4
E	5
F	3.5



Group	winPlacePercPred
A	1
B	0.8
C	0.4
D	0.2
E	0
F	0.6

Our Strategy.

Post-processing.

✓ **Continuous → Discrete**

- ex. 매치 내 6개의 그룹이 존재하는 경우

Group	Predicted winPlacePerc
A	0.9856
B	0.7854
C	0.4249
D	0.2189
E	0.0014
F	0.6514

“Convert”



Group	Predicted winPlacePerc
A	1
B	0.8
C	0.4
D	0.2
E	0
F	0.6

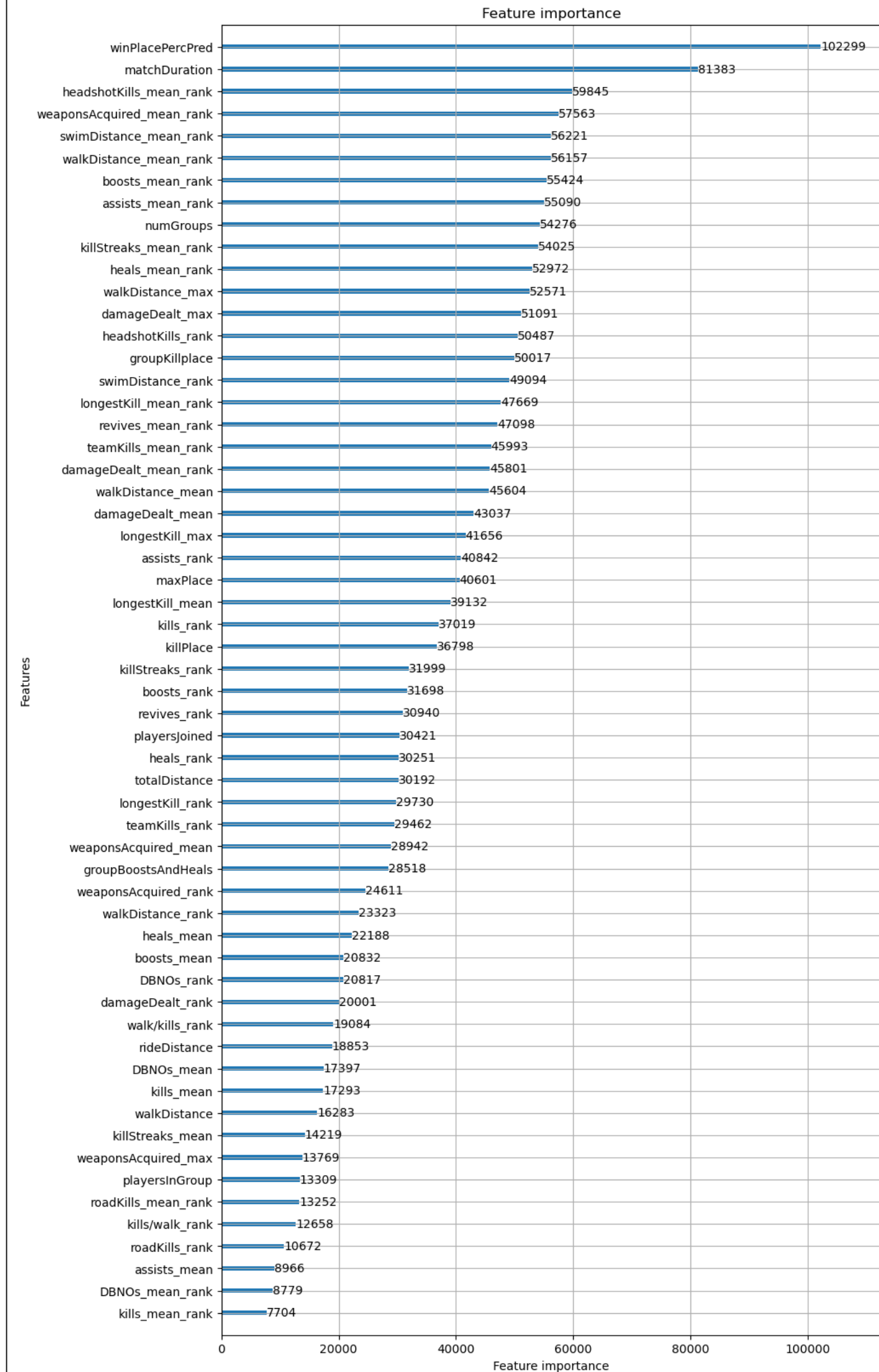
Result & Conclusion

Result & Conclusion.

Most effective factors in predicting finish placement.

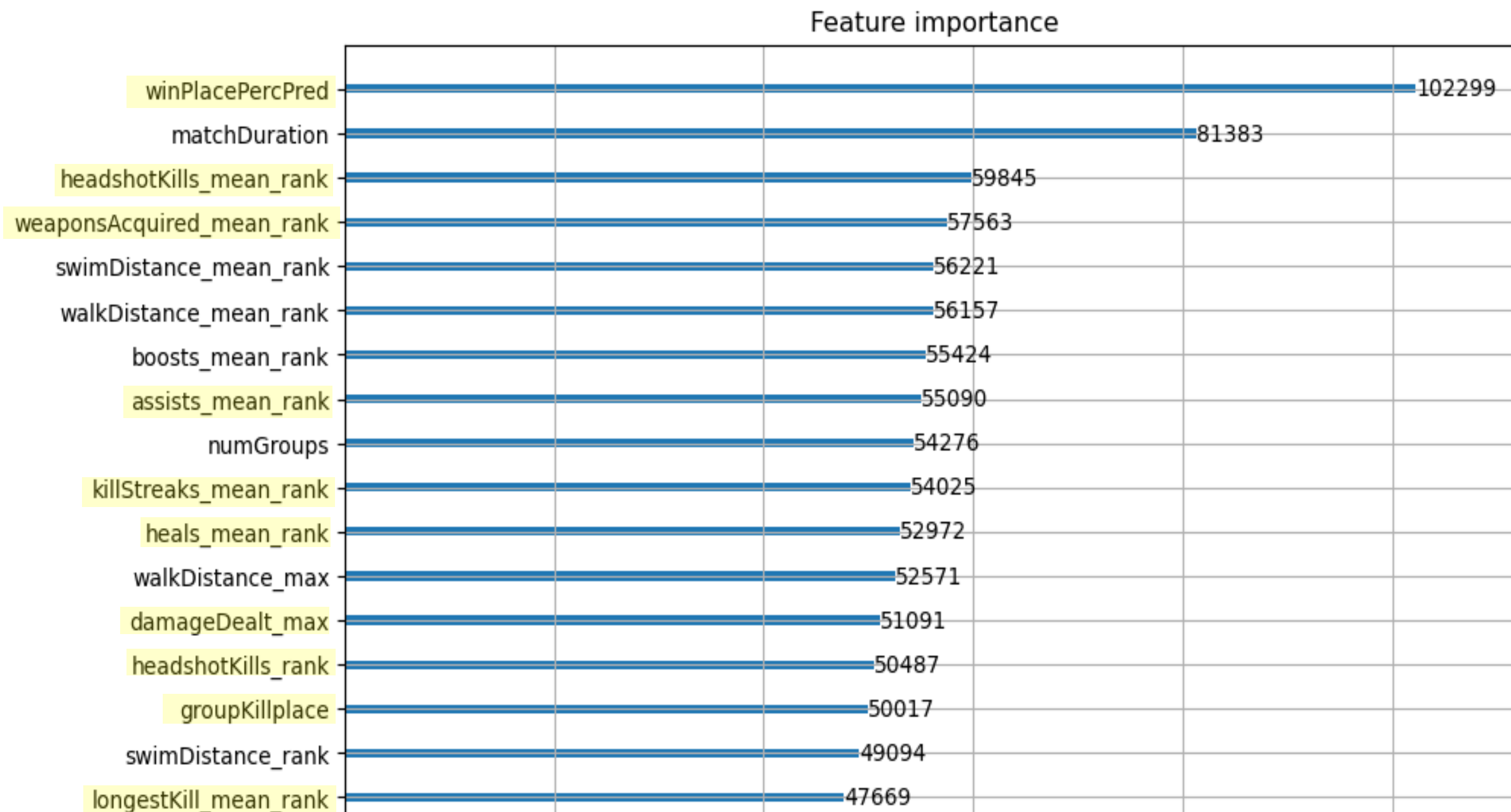
✓ **winPlacePercPred**가 가장 중요한 변수로 작용함.

✓ 변수중요도 순서로 변수 선택을 수행한 결과, 개인의 기록보다 **그룹의 기록**이 상대적으로 중요한 변수로 작용함.



Result & Conclusion.

Most effective factors in predicting finish placement.



✓ 변수 중요도 top10*

winPlacePercPred

headshotKills_mean_rank

weaponsAcquired_mean_rank

assists_mean_rank

killStreaks_mean_rank

heals_mean_rank

damageDealt_max

headshotKills_rank

groupKillplace

longestKill_mean_rank

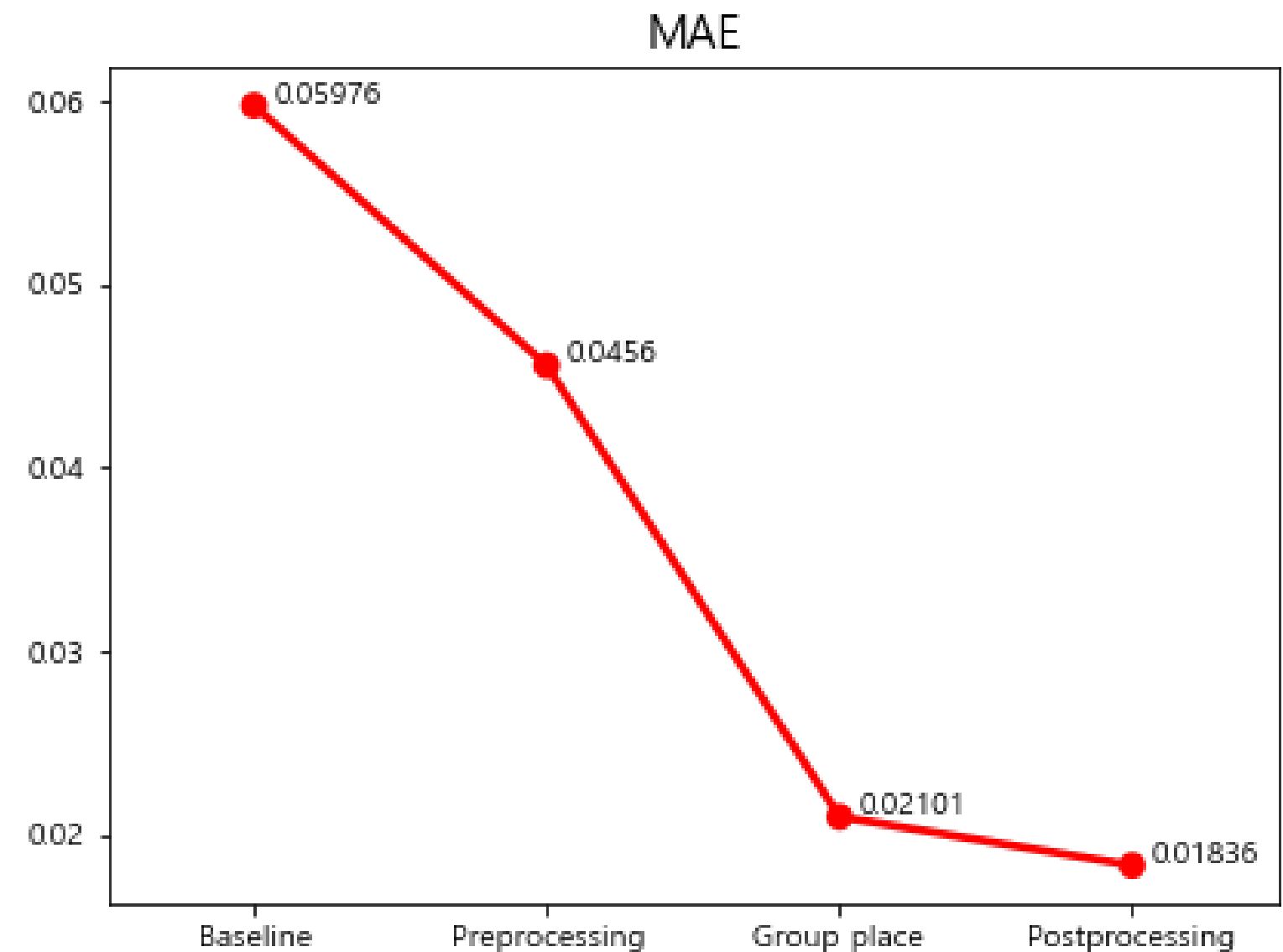
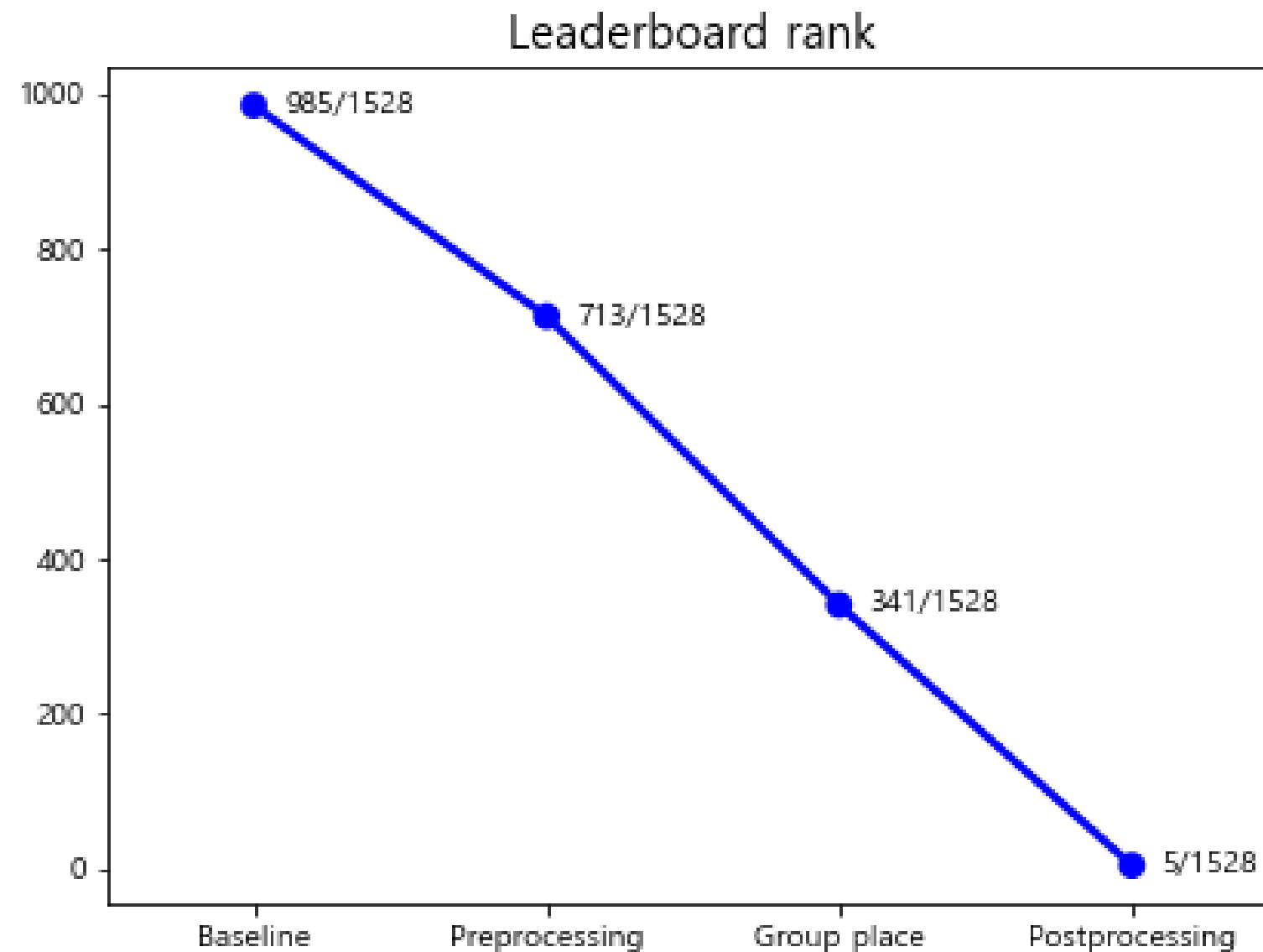
→ 높은 등수를 차지한 그룹의 경우,
전투와 관련된 기록이 높게 나타남.

→ 매치 순위에 영향을 주는 요인은
그룹의 전투력이라고 할 수 있음.

* matchDuration, Distance 등, 매치가 진행됨에 따라
필연적으로 증가할 수밖에 없는 변수들은 고려하지 않음.

Result & Conclusion.

Our rank in Leaderboard & MAE.



Result & Conclusion.

Conclusion.

✓ 최종 성적 : 5 / 1528 (top 0.3%)

✓ 결론

높은 순위를 차지하기 위해서 유저들이 취하는 방법에는 크게 2가지가 있음.

1. 교전을 통해서 적을 제거하는 방법
2. 교전을 피하면서 오래 살아남는 방법

분석 결과, 높은 등수를 차지한 그룹은 교전을 피하는 그룹이 아니라 전투력이 높고 교전을 많이 진행한 그룹이었음을 확인할 수 있었음.

따라서, **매치 순위에 영향을 주는 요인**은 **그룹의 전투력**이라고 할 수 있음.

Contact Us



Kyungjun Kang

kangjun205@gmail.com



Mingyu Park

mg4432@naver.com