



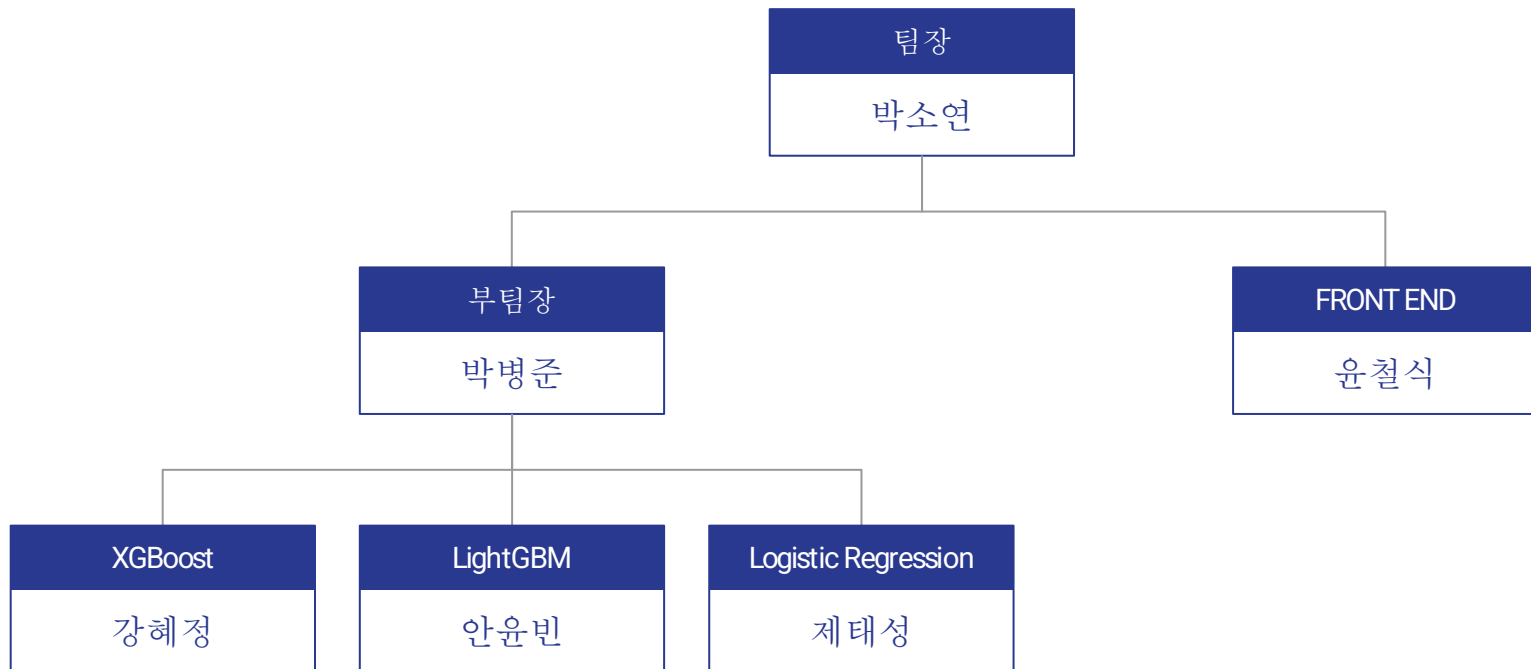
당신의
MONEY를
지켜드립니다!

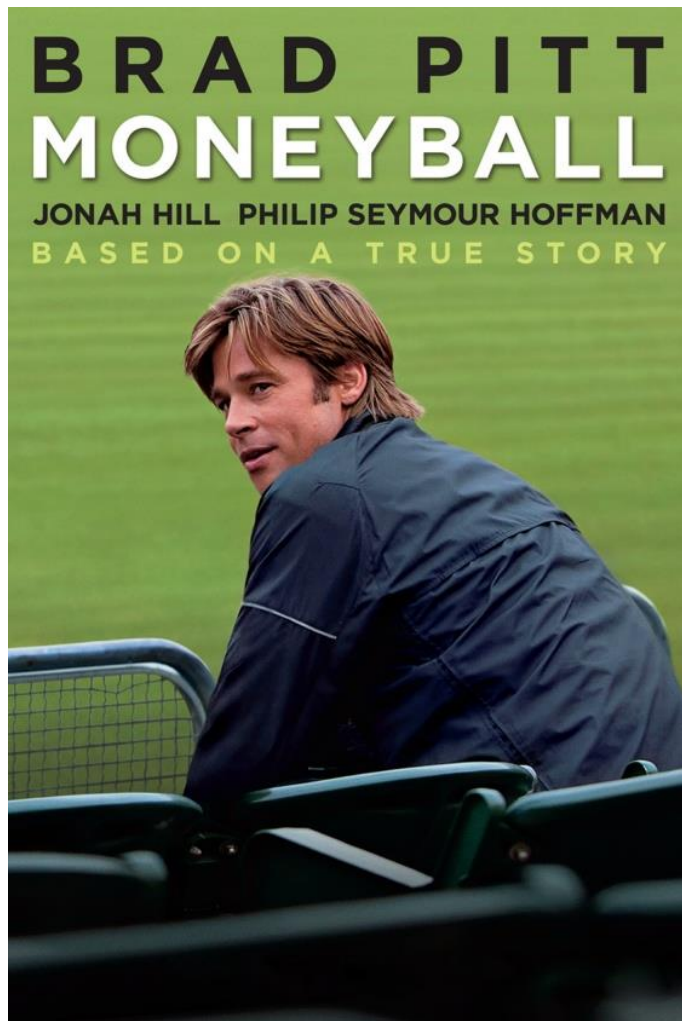
KBO 한국 프로야구 승부 예측

‘역대 기록(Records)’을 분석하여 오늘의 승패를 예측할 수 있는가

박소연 박병준 강혜정 안윤빈 윤철식 제태성

팀 정보





개요






1. 프로젝트 주제
2. 프로젝트 목적
3. 분석 과정 및 결과
4. 실제 예측 적용

—

1. 프로젝트 주제



2. 프로젝트 목적

	baseball_2015.csv	baseball_data
	baseball_2016.csv	baseball_data
	baseball_2017.csv	baseball_data
	baseball_2018.csv	baseball_data
	baseball_2019.csv	baseball_data
	baseball_2020.csv	baseball_data

KBO_softmax_2015_2020_vol5

localhost:8888/notebooks/KBO_softmax_2015_2020_vol5.ipynb

Logout

Jupyter KBO_softmax_2015_2020_vol5 (autosaved)

Python 3 (ipykernel)

File Edit View Insert Cell Kernel Widgets Help

Run

Code

```
0 20150328HHNE0 20150328 HH NE 0 T 54 42 4 4 ... 1 7 0 24 0.133333 15 2 0.352941
1 20150328HHNE0 20150328 NE HH 0 B 46 40 5 5 ... 0 7 1 18 0.090909 11 1 0.304348
2 20150328KTLT0 20150328 KT LT 0 T 46 35 8 9 ... 1 8 1 23 0.285714 14 4 0.511111
3 20150328KTLT0 20150328 LT KT 0 B 42 37 12 12 ... 1 8 1 17 0.437500 16 7 0.439024
4 20150328LGHT0 20150328 LG HT 0 T 34 29 1 1 ... 0 6 2 17 0.200000 10 2 0.333333
```

5 rows x 28 columns

```
In [7]: baseball_2016.head()
```

```
Out [7]:
```

	G_ID	GDAY_DS	T_ID	VS_T_ID	HEADER_NO	TB_SC	PA	AB	RBI	RUN	...	HP	KK	GD	LOB	P_HRA_RT	P_AB_CN	P_HIT_CN	
0	20160401HHLG0	20160401	HH	LG	0	T	52.0	46.0	4.0	4.0	...	0.0	10.0	1.0	24.0	0.200000	15	3	0.326
1	20160401HHLG0	20160401	LG	HH	0	B	47.0	42.0	4.0	5.0	...	0.0	11.0	0.0	15.0	0.333333	12	4	0.282
2	20160401HTNC0	20160401	HT	NC	0	T	38.0	34.0	3.0	4.0	...	1.0	10.0	1.0	17.0	0.100000	10	1	0.315
3	20160401HTNC0	20160401	NC	HT	0	B	36.0	30.0	5.0	5.0	...	0.0	9.0	1.0	14.0	0.142857	7	1	0.400
4	20160401KTSK0	20160401	KT	SK	0	T	42.0	36.0	7.0	8.0	...	1.0	7.0	0.0	17.0	0.285714	7	2	0.380

5 rows x 28 columns

```
In [8]: baseball_2017.head()
```

```
Out [8]:
```

	G_ID	GDAY_DS	T_ID	VS_T_ID	HEADER_NO	TB_SC	PA	AB	RBI	RUN	...	HP	KK	GD	LOB	P_HRA_RT	P_AB_CN	P_HIT_CN	O
0	20170331HHOB0	20170331	HH	OB	0	T	33.0	29.0	0.0	0.0	...	0.0	8.0	0.0	11.0	0.000000	4	0	0.218
1	20170331HHOB0	20170331	OB	HH	0	B	33.0	27.0	3.0	3.0	...	2.0	8.0	0.0	15.0	0.111111	9	1	0.242
2	20170331HTSS0	20170331	HT	SS	0	T	42.0	34.0	7.0	7.0	...	0.0	5.0	0.0	12.0	0.285714	7	2	0.357
3	20170331HTSS0	20170331	SS	HT	0	B	33.0	31.0	2.0	2.0	...	0.0	7.0	1.0	8.0	0.000000	3	0	0.250
4	20170331KTSK0	20170331	KT	SK	0	T	37.0	35.0	3.0	3.0	...	0.0	8.0	1.0	12.0	0.200000	5	1	0.297

5 rows x 28 columns

```
In [9]: baseball_2018.head()
```

```
Out [9]:
```

	G_ID	GDAY_DS	T_ID	VS_T_ID	HEADER_NO	TB_SC	PA	AB	RBI	RUN	...	HP	KK	GD	LOB	P_HRA_RT	P_AB_CN	P_HIT_CN	
0	20180324HHNE0	20180324	HH	NE	0	T	38.0	37.0	2.0	3.0	...	0.0	9.0	0.0	16.0	0.230769	13	3	0.366
1	20180324HHNE0	20180324	NE	HH	0	B	41.0	37.0	5.0	6.0	...	0.0	10.0	0.0	26.0	0.235294	17	4	0.390
2	20180324KTHT0	20180324	KT	HT	0	T	38.0	37.0	5.0	5.0	...	0.0	9.0	0.0	11.0	0.400000	5	2	0.315
3	20180324KTHT0	20180324	HT	KT	0	B	43.0	36.0	4.0	4.0	...	2.0	8.0	0.0	25.0	0.272727	11	3	0.390
4	20180324LGNC0	20180324	LG	NC	0	T	32.0	30.0	2.0	2.0	...	0.0	9.0	1.0	10.0	0.000000	4	0	0.250

PA	AB	RBI	RUN	HIT	H2
타석	타수	타점	득점	안타	2루타
H3	HR	SB	CS	SF	BB
3루타	홈런	도루	도루실패	희생플라이	볼넷
HP	KK	GD	LOB	OBP	OOO
홀드포인트	삼진	병살	잔루율	출루율	타율
win					
0:패, 1:승, 0.5:무					

팀 명	HT	KT	LG	LT	NC
	기아	KT	LG	롯데	NC
OB	SK	SS	WO	NE	HH
두산	SK	삼성	키움	넥센	한화

3. 분석 과정 및 결과

데이터 전처리

효율적인 데이터 셋을 찾아라!!

- koreabaseball.com
- statiz.co.kr
- kbreport.com
- dacon.io
- kaggle.com

상관관계 분석

So Many Features!!

- PA
- AB
- RUN
- OBP
- OOO

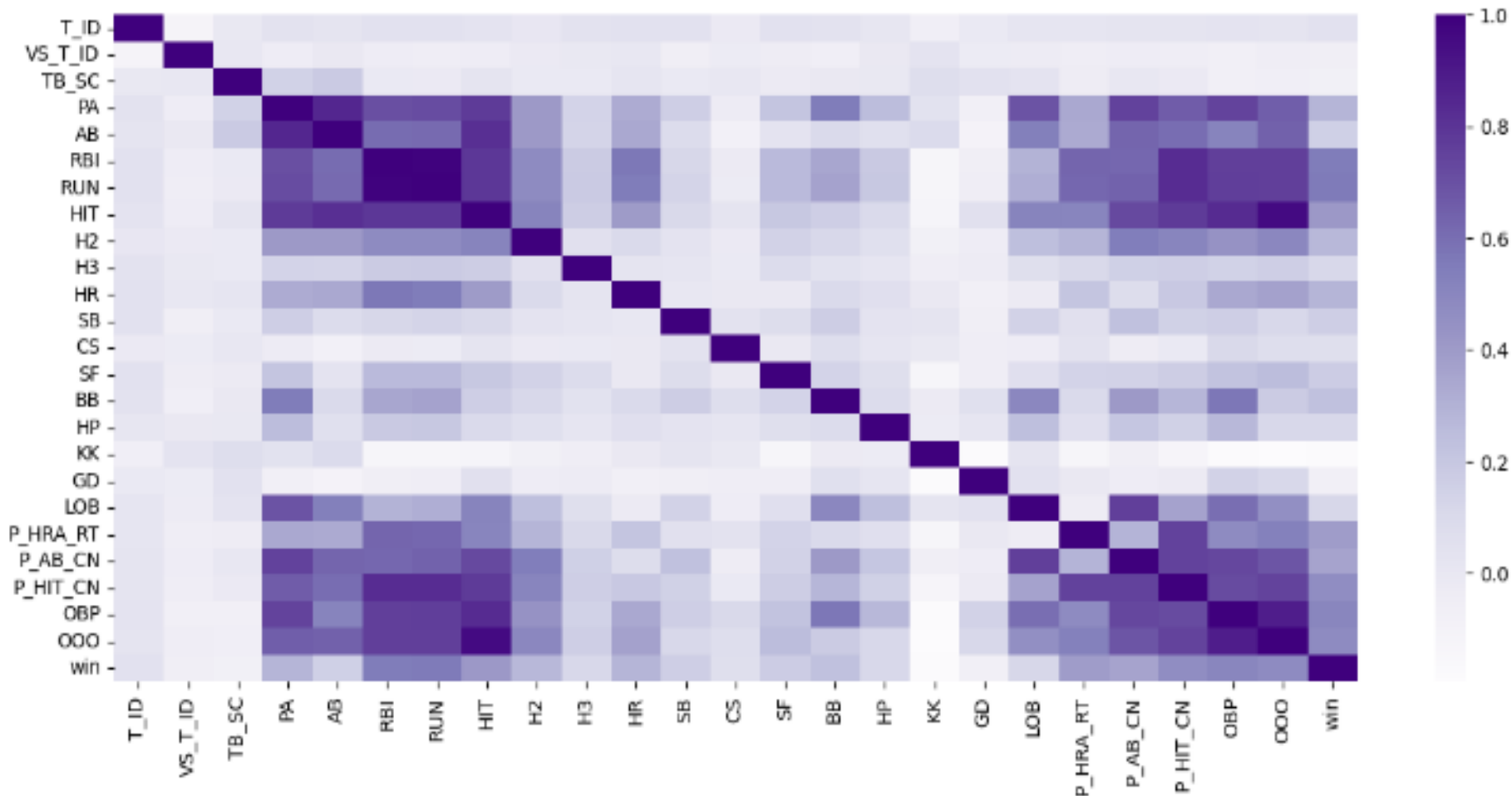
알고리즘 선정

정확도를 높여라!!

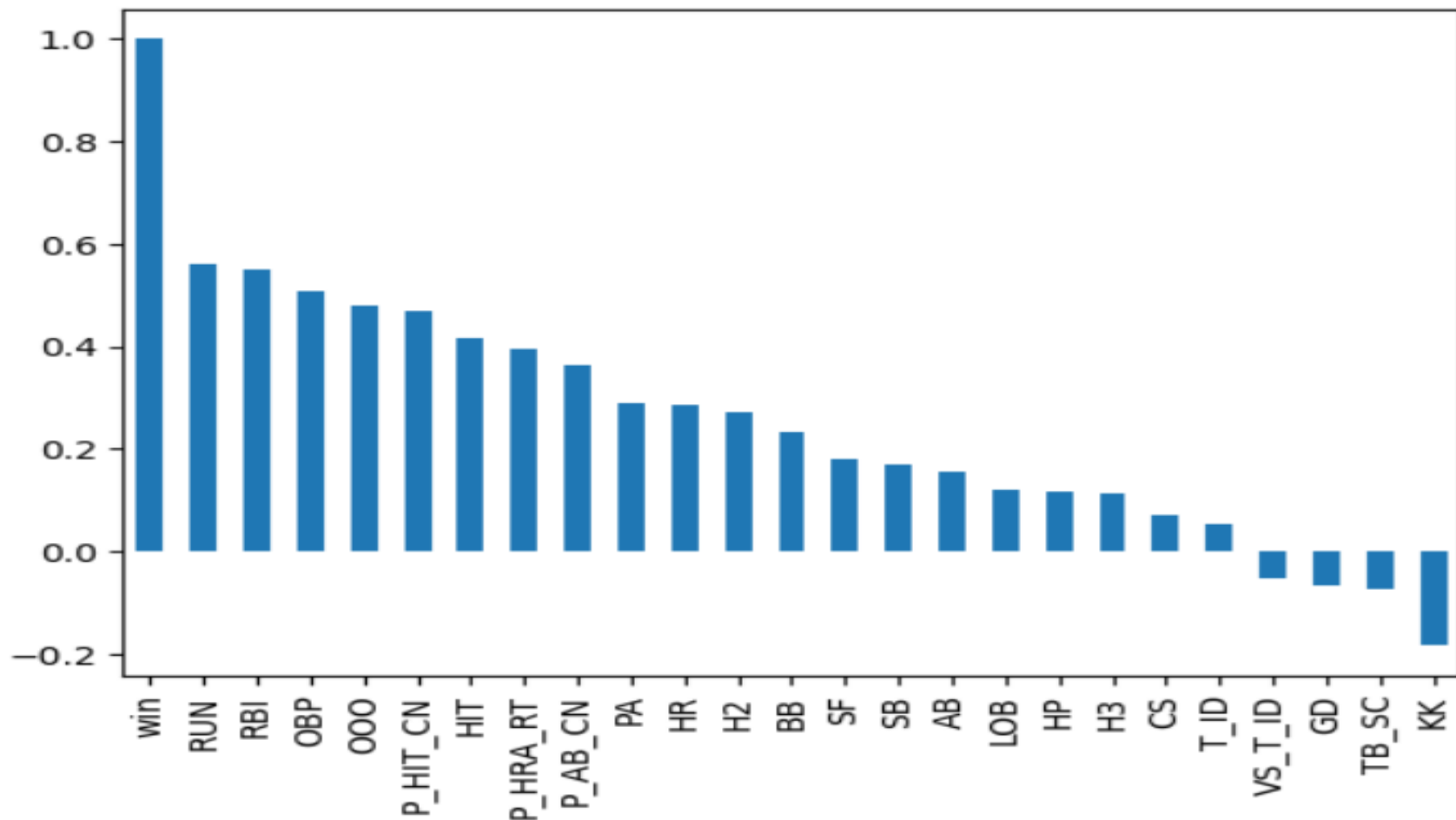
- RandomForest
- GBM
- LightGBM
- XGBosst
- LogisiticRegression

상관관계

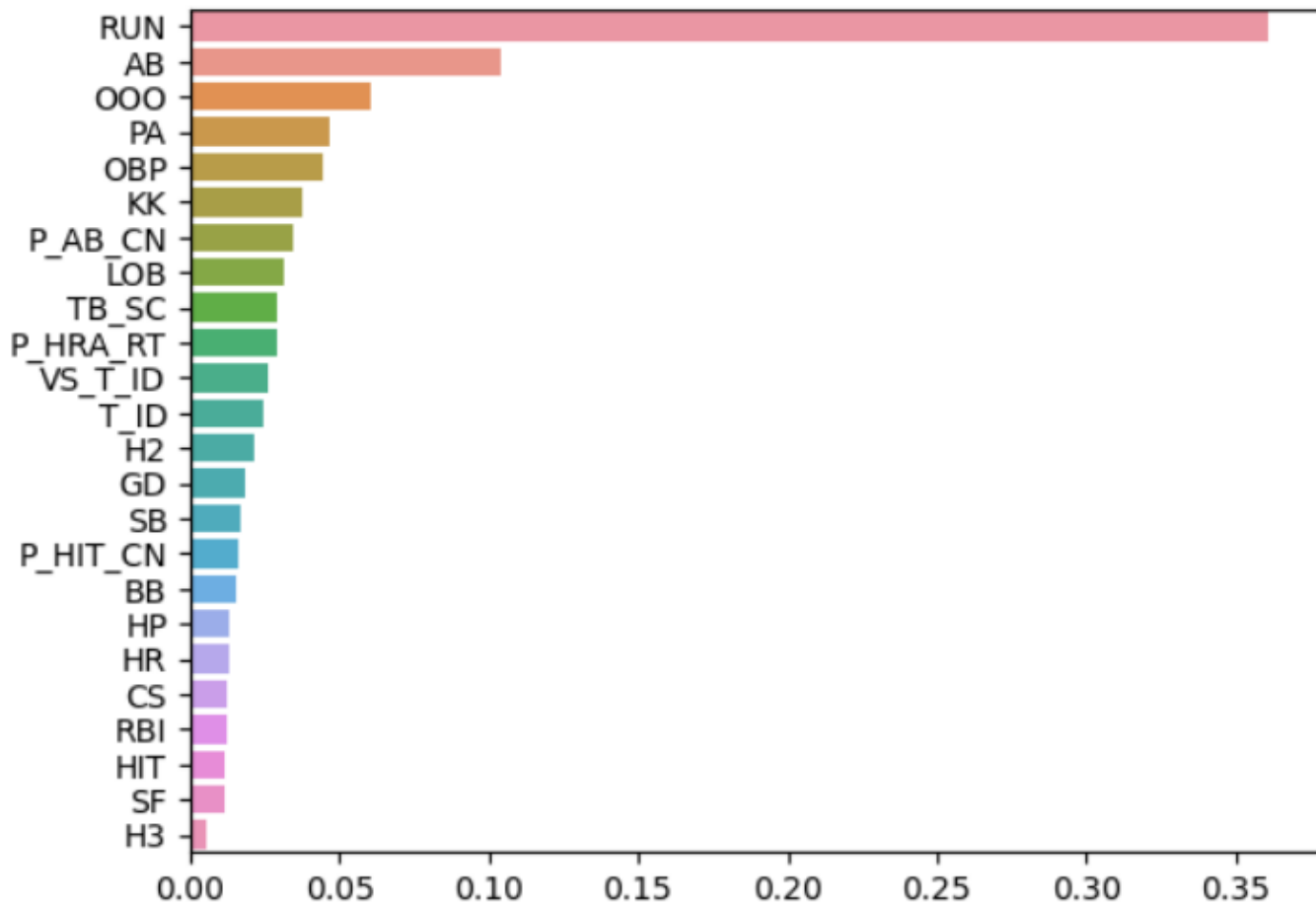
Features 상관계 히트맵



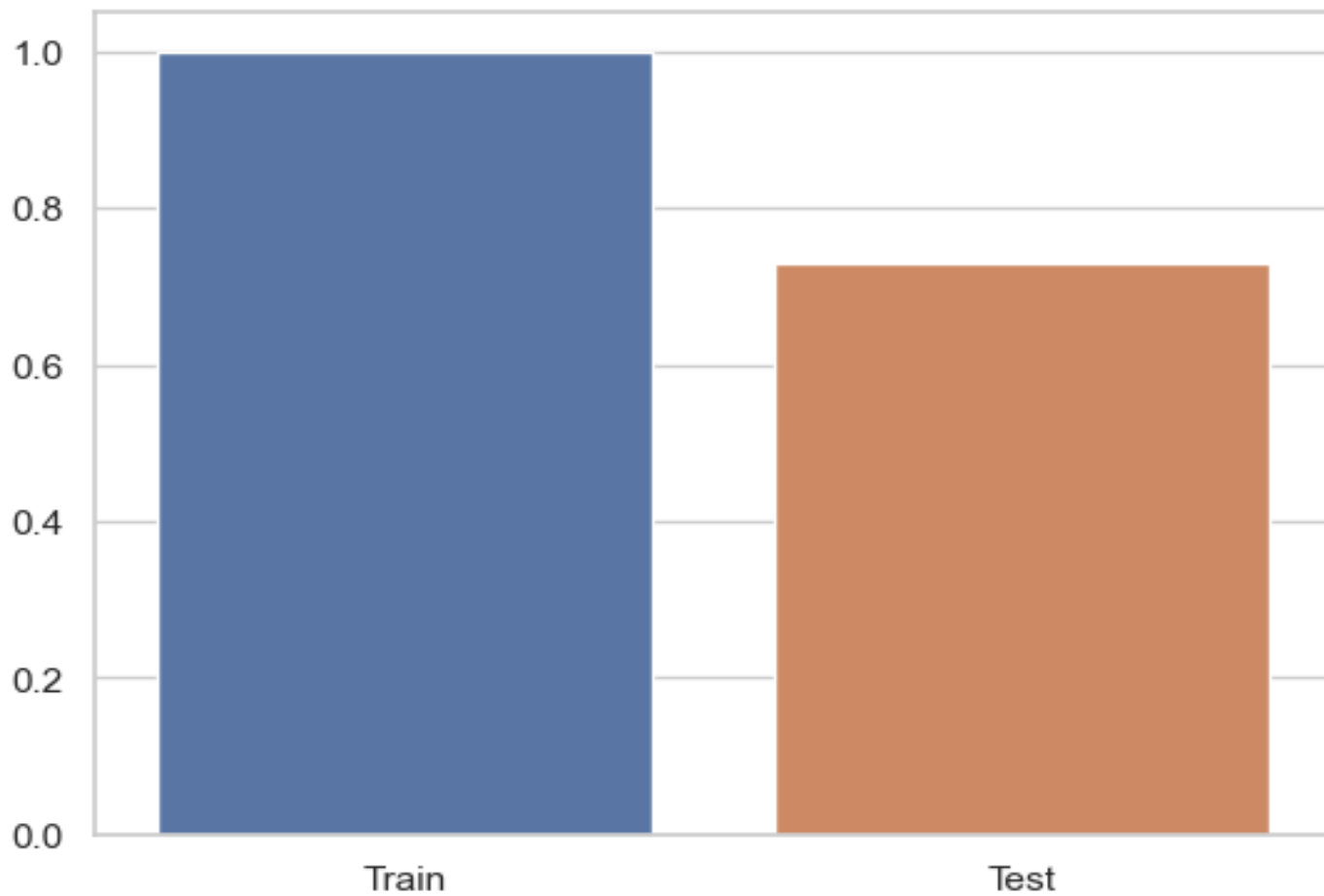
Win에 대한 상관관계 막대그래프



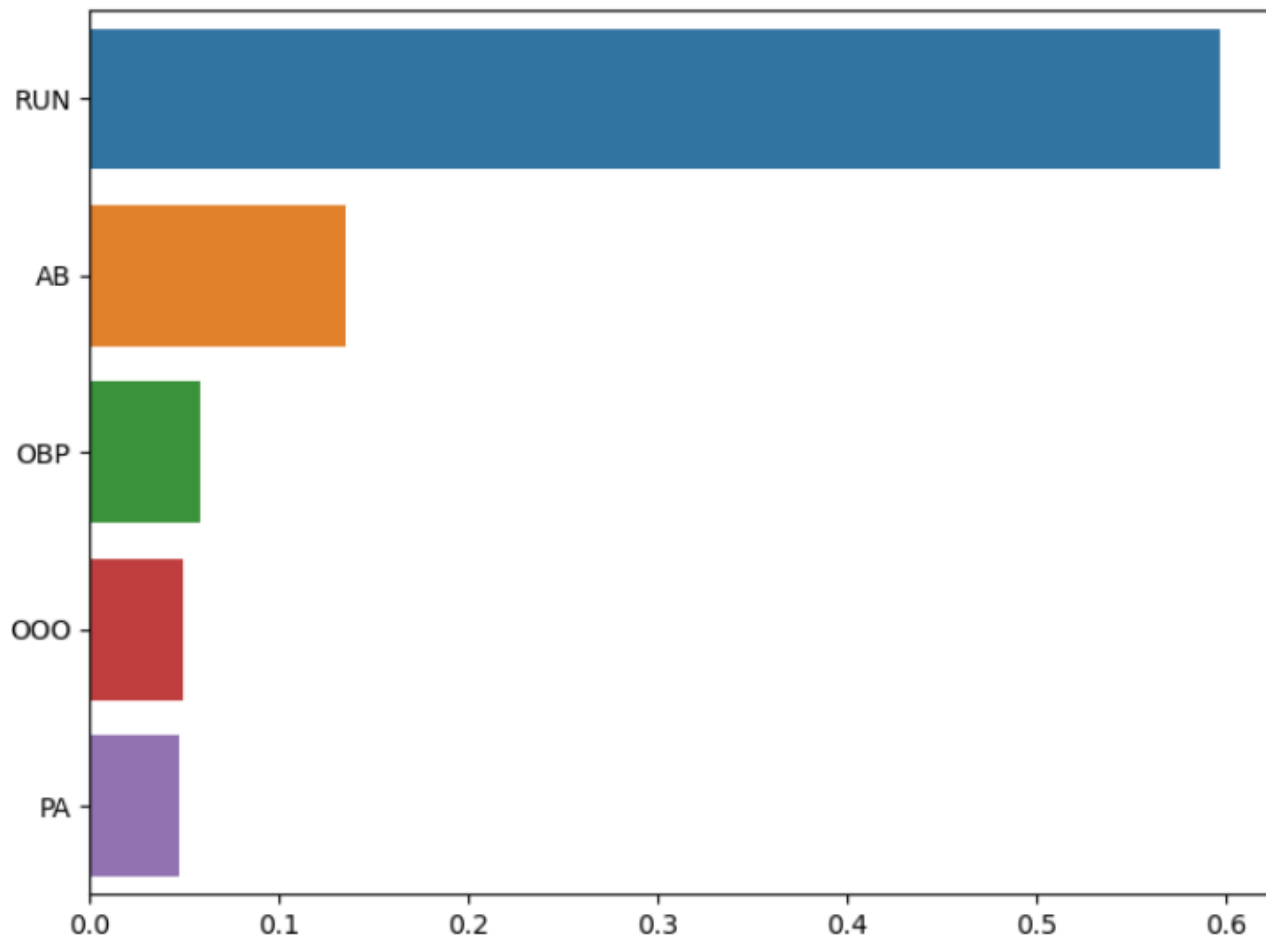
Decision Tree 모델 Features 중요도



Decision Tree 과대적합



GBM 모델 feature_importances_



분석 방향성 판단

과대적합 문제

부스팅 알고리즘

우수한 예측 성능

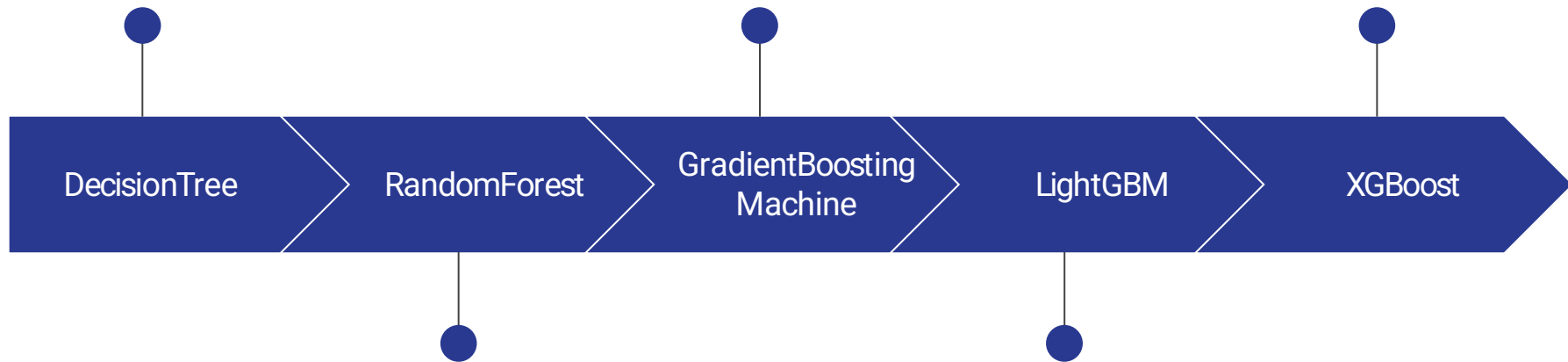
대용량 데이터 취약

예측 오차 최소화

교차 검증 및 성능 평가

`feature_importances` 시각화

조기 중단 기능



대표적인 앙상블 학습

배깅 알고리즘

하이퍼 파라미터

많은 시간 소요

빠른 학습 속도

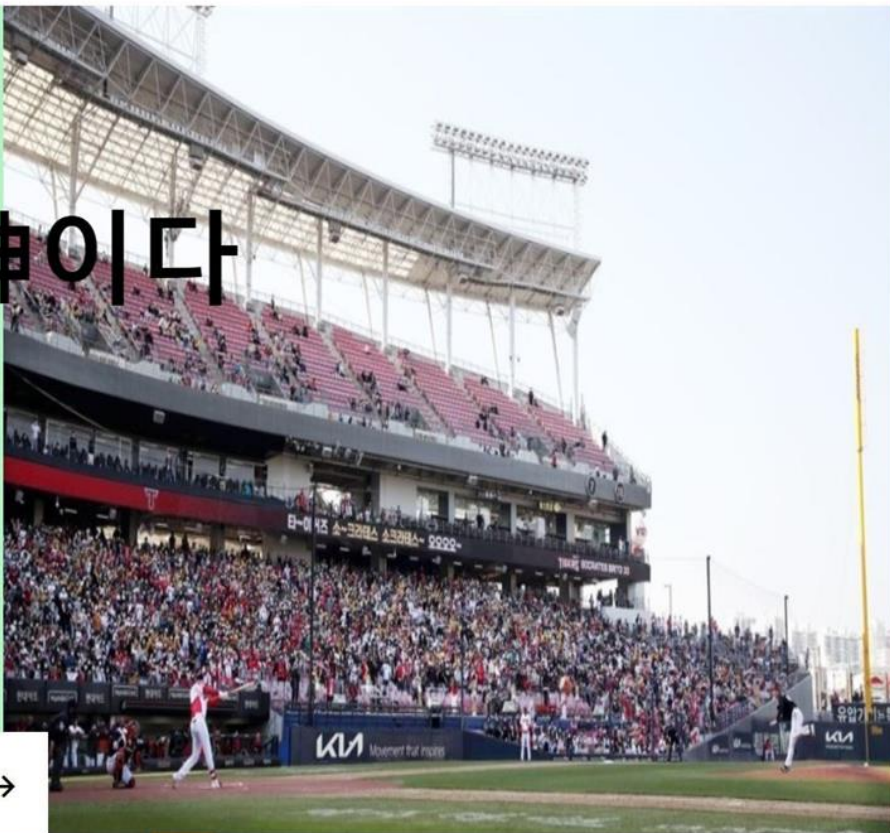
범주형 데이터 지원

나는 야신野神이다

프로야구 승부예측 올킬에 도전한다!

승부예측

← 1/3 →



승부 예측

어제의 승리팀은...??

proto



프로토 (승부식)

39회차

1787253277

1502-41398-1

E853-80B4-8829-79D

2023/ 4/ 6 09:38:30

최종경기일자: 2023/ 4/ 6
3경기 선택

경기	홀 팀	: 원정팀	예상	배당률
*227	두산	: NC	패	2.00
*233	삼성	: 한화	승	1.70
*239	키움	: LG	승	1.87

예상 적중배당률 : 6.40배
개별투표금액 : 5,000원
예상 적중금 : 32,000원

총투표금액 5,000원

1787253277

E853-80B4-8829-79D



```
array([[0.51225582, 0.48774418],  
       [0.65866678, 0.34133322],  
       [0.39885111, 0.6011489 ],  
       [0.87020795, 0.12979204],  
       [0.80865611, 0.1913439 ],  
       [0.71545524, 0.28454476]])
```

=> OB vs NC : NC 승

=> SS vs HH : HH 승

=> WO vs LG : WO 승

한화 승 문동주  8 : 1  삼성 패 백정현

LG 승 강효종  5 : 0  키움 패 장재영

NC 패 이준호  2 : 6  두산 승 김동주

그렇다면...
오늘의 승리팀은...??

```
array([[0.79001257, 0.20998742],  
       [0.75785758, 0.24214243],  
       [0.509688   , 0.490312   ],  
       [0.70583774, 0.29416226],  
       [0.16851304, 0.83148695],  
       [0.79396656, 0.20603344],  
       [0.74650047, 0.25349954],  
       [0.50005818, 0.49994182],  
       [0.6468026  , 0.3531974  ],  
       [0.81166141, 0.18833859]])
```

- KT vs LT : LT 승
- OB vs KIA : OB 승
- SSG vs HH : SSG 승
- SS vs LG : LG 승
- WO vs NC : WO 승





**It's unbelievable
how much
you don't know
about the game
you've been playing
all your life.**

- Mickey Mantle