

마기꾼들

권 효 은

박 재 원

심 재 만

정 진 명

이 승 희

최 규 광

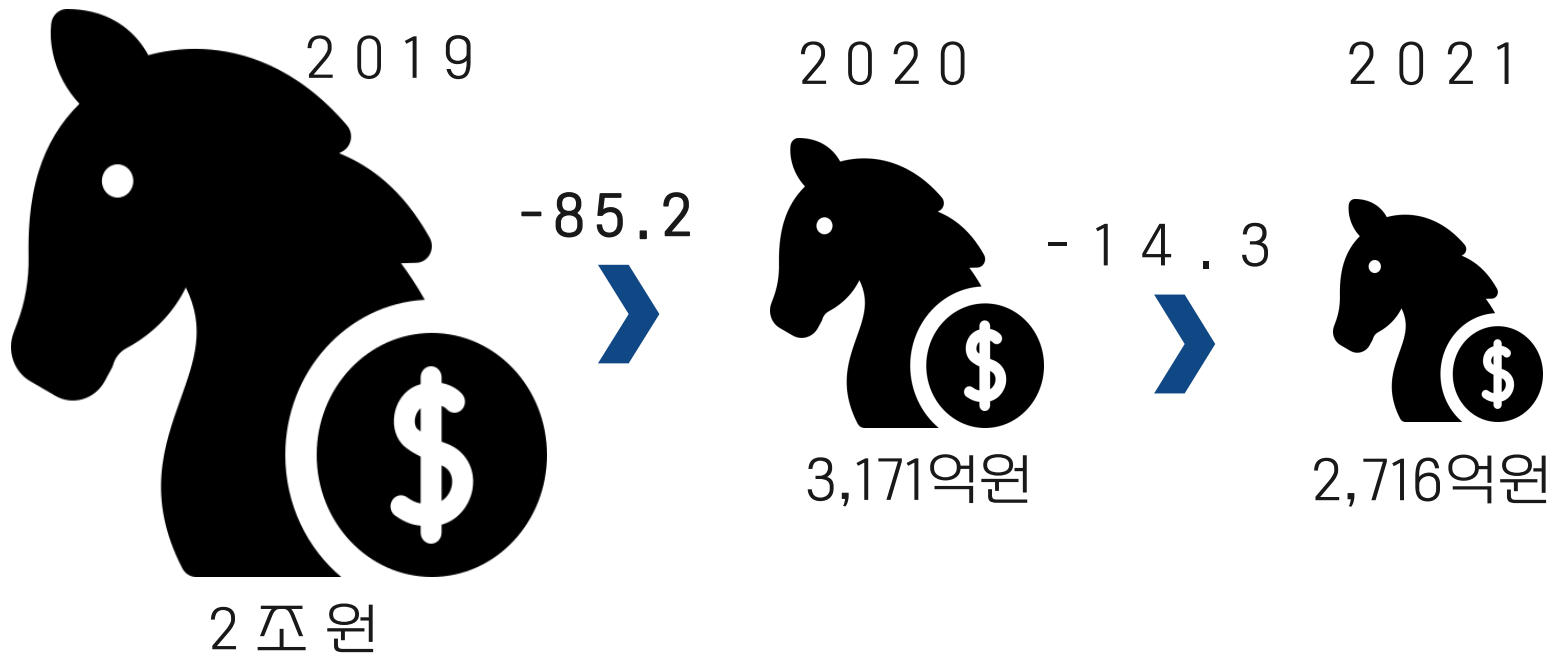


경마 데이터 분석을 통한 경마 경기 결과 예측

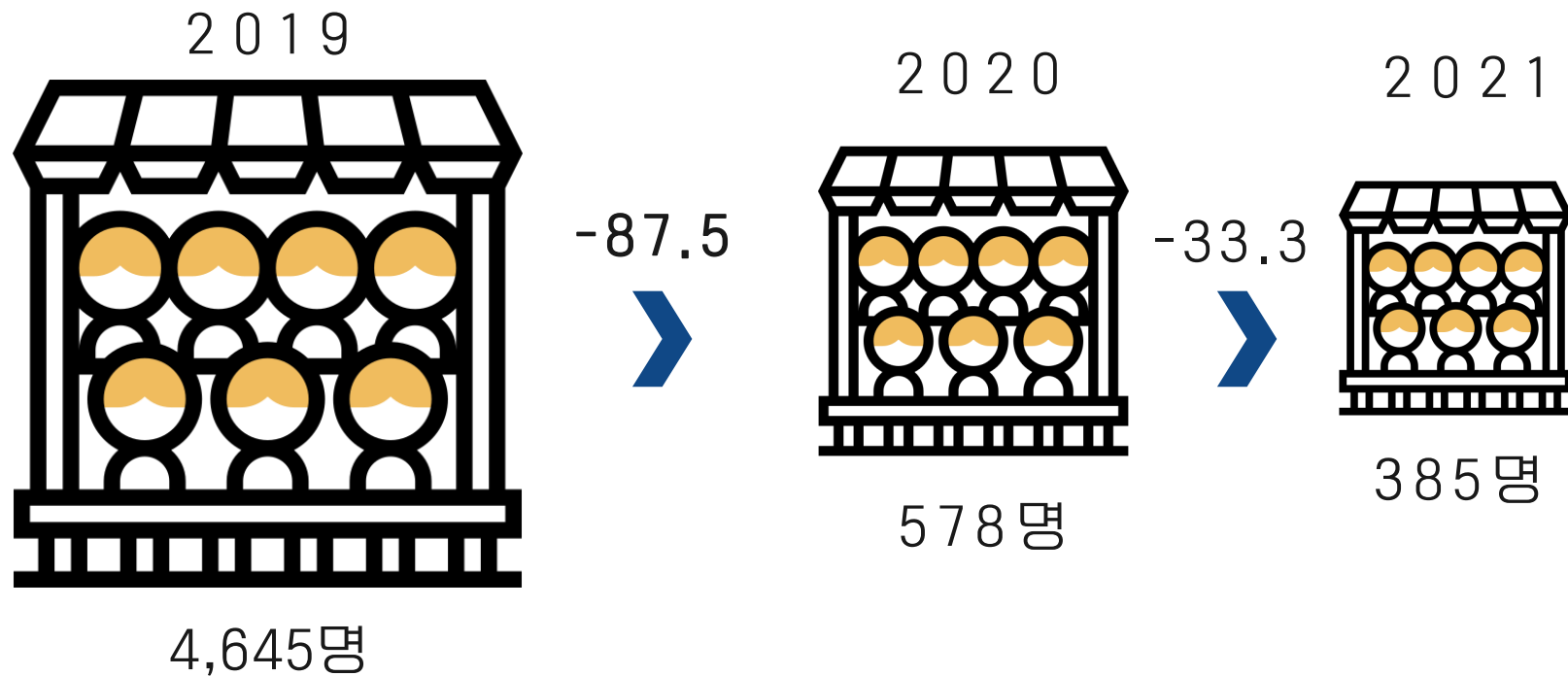


1922-2022

1922년부터
시행된
한국 경마산업은
2022년
100주년을 맞음



02



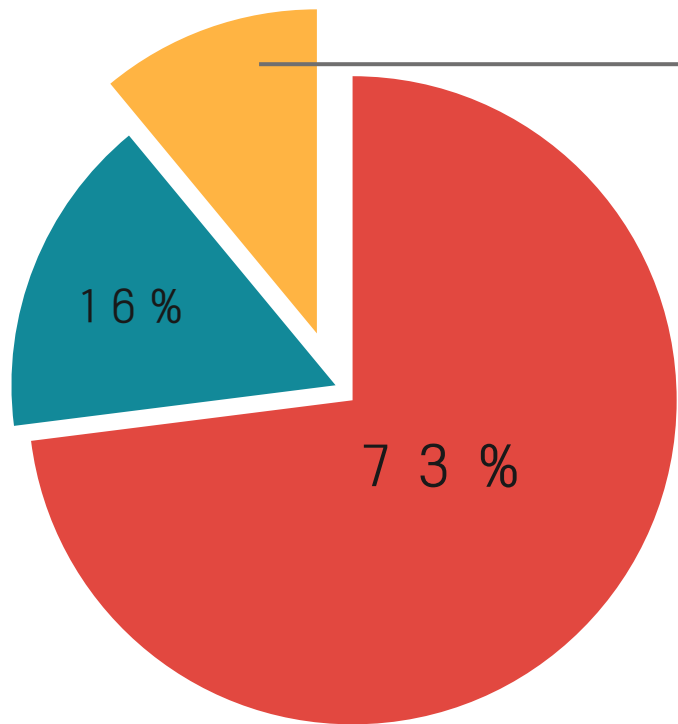
01 지난 100년간 경마,
축산, 수의분야 등
다양한 분야의
데이터 축적



데이터 활용 위한
미래사업부 신설

02 공공 데이터 개방도
100% 완료하여 빅데이터로
새로운 가치 창출을 시도

경마 산업 지출 구조



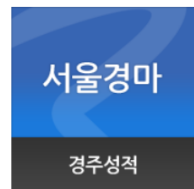
■ 환급금 ■ 세금 ■ 수득금

11%

경주 개최비용을 제외한
나머지는 법적으로
사업준비금 및
축산발전기금으로
적립해야 함

+

사행산업으로 조성되는 복지 기금
예산 또한 편성되지 못하는 상황



경주성적표

[홈](#) > [서울경마](#) > [경주성적](#) > [경주성적표](#)
[경주성적표 리스트](#) | [일자별 요약성적표](#) | [경주별 상세 성적표](#)

경주별 상세 성적표

[단](#) · [연](#) · [복](#) | [쌍승식](#) | [복연승식](#) | [삼복승식](#) | [삼쌍승식](#)
[경주영상](#) | [영상보기](#) | [YouTube](#) | [경주전 영상](#) | [예시장](#) | [경주로입장](#) | [이전경주](#) | [다음경주](#) | [인쇄하기](#) | [엑셀변환](#)

2022년 11월 13일 (일) 제1경주 서울				제 84일	호림	포화	18%	10:45
국6등급	1000M	별정A	일반	R0~0		2세 암		Rookie1
순위상금		33,000,000원	13,200,000원	8,400,000원	3,000,000원	2,400,000원		

- ☒ 중량앞에 +표시는 52kg미만의 말이 52kg으로 증량신청한 것을 의미합니다.
☒ 마주명앞에 ◆표시는 마주복색 등록마주를 의미합니다.
☒ 장구현황의 +는 신규장구, -는 해지장구를 의미합니다.(2016년 2월 성적표자료부터 표기합니다.)
☒ 굵은파란 표기마는 금주 출전마 입니다.

순위	마번	마명	산지	성별	연령	중량	레이팅	기수명	조교사명	마주명	도착차	마체중	단승	연승	장구현황
1	1	영영익스	한	암	2세	53		김동수	이신영	◆김창식		491(4)	7.4	1.4	
2	5	나을행복	한	암	2세	53		함환식	박영호	◆이경호a	2	503(2)	6.6	1.5	계란형큰*, 말사눈
3	10	라운엘파이어	한	암	2세	53		문세영	박종곤	◆손광섭	4	455(10)	1.3	1.0	말사눈
4	4	월가닥	한	암	2세	53		이동하	강성오	◆김철	2	455(-14)	37.6	3.2	말사눈
5	12	레드태풍	한	암	2세	53		송재철	안해양	◆제일우	5	450(17)	21.6	3.0	말사

경주 성적표, 부마/모마,
경주마 정보 등
사이트에 있는 데이터 중
필요한 데이터 크롤링

002

본론

데 이 터

컬럼 설명

	마번	조교사 점수	기수 점수	경주번호	거리	군(등급)	번호	출전두수	주요상태	기록	S1F기록	G3F기록	G1F기록	부담중량	마체중	순위 점수
0	46584	2703.103	1847.108	2211139	12	4	1	8	3	71.6	13.3	35.8	12.4	53	6.265301	0
1	44261	1150.096	8048.12	2211139	12	4	3	8	3	71.6	13.3	35.6	12.4	54	6.152733	6
2	45027	1927.085	3931.112	2211139	12	4	2	8	3	72.6	13.7	36.3	12.4	51	6.135565	9
3	45117	1553.087	1442.093	2211139	12	4	7	8	3	72.6	14.5	35.8	12.4	55.5	6.198479	0
4	46520	618.1131	698.1077	2211139	12	4	4	8	3	73.5	13.9	36.9	13.1	52	6.159095	0
5	44927	221.1118	691.099	2211139	12	4	6	8	3	74.2	13.6	37.7	13.5	57	6.144186	0
6	41091	2490.086	1438.104	2211139	12	4	5	8	3	74.4	14	37.6	13.4	55.5	6.182085	0
7	41147	2288.096	1424.164	2211139	12	4	8	8	3	75.5	13.4	39.5	14.7	52	6.111467	0
8	40560	2433.096	4251.106	2211138	20	7	6	14	3	125.3	13.5	37.4	13	55	6.122493	25
9	38511	970.0862	8048.12	2211138	20	7	3	14	3	125.6	13.7	37.5	13	57	6.228511	0
10	42917	213.0879	1847.108	2211138	20	7	4	14	3	126	13.4	38.4	13.5	57	6.248043	0
11	38733	650.1063	1442.093	2211138	20	7	1	14	3	126.2	14.7	37	12.9	57	6.214608	0
12	42916	2132.097	1971.082	2211138	20	7	5	14	3	126.6	14.4	37.5	12.7	57	6.269096	3
13	42593	1389.093	2080.094	2211138	20	7	13	14	3	126.7	13.8	37.8	13.1	57	6.159095	2
14	40633	170.1168	3931.112	2211138	20	7	2	14	3	126.7	14.1	38.6	13.5	57	6.161207	12
15	42254	650.1063	1438.104	2211138	20	7	9	14	3	127.2	14.1	37.9	12.8	57	6.240276	4
16	39422	4295.104	691.099	2211138	20	7	8	14	3	129.1	13.7	40.6	14.6	57	6.228511	4

002

본론

부마, 모마

말의 혈통, 특히 부마가 중요(경마는 혈통 스포츠)

조교사

말을 키우는 사람, 말을 얼마나 다양한 기수와 잘 적응할 수 있을지 조련

기수

말에 타는 사람. 말과의 교감이 중요.

데이터 컬럼 설명

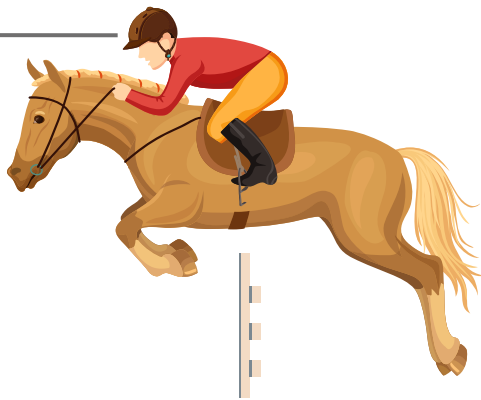
1

2

3



말에 타는 사람. 말과의 교감이 중요.



[경마는 과학이다]말과 기수의 환상 호흡을 찾아라

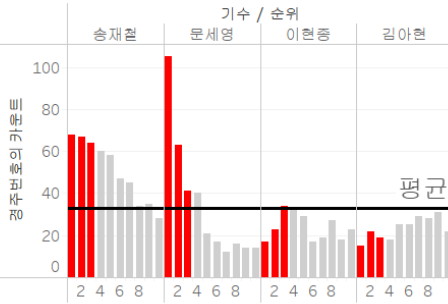
[경마는 과학이다]

권승주 전문 기자

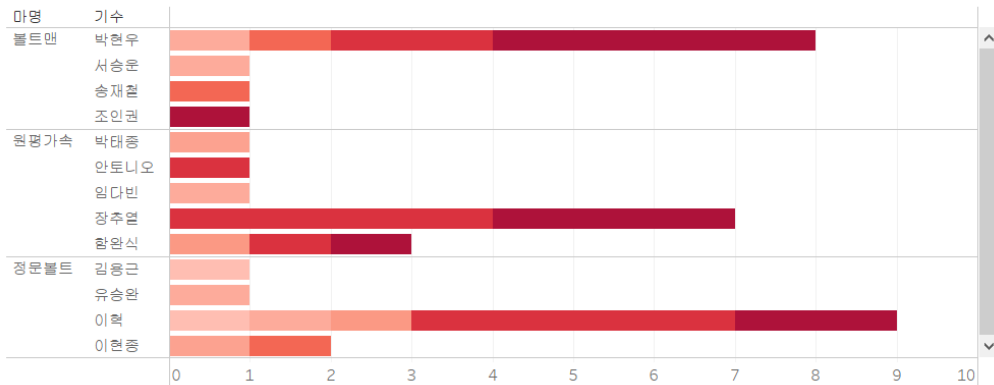
입력 2022.09.18 08:45

필자가 신인 기수시절 '삼학도'라는 말이 있었다. 이 말은 능력이 그다지 좋지 않은 하위 군에서 출전하는 말이었다. 여러 기수들이 타고 경주에 출전을 했지만 성적이 나지 않았다. 그러나 필자가 경주에 출전하여 2승을 한 적이 있었다. 이 말을 타고 경주에 나가면 말도 기분이 업 되는 느낌을 받았고 나 또한 자신감이 생기곤 했다. 그 결과가 좋은 성적으로 이어진 것이 아닌가 싶다. 경주 전개도 선형으로 가는 스타일이었는데 선형을 가지 않고 선입에서 차분하게 호흡을 내쉬면서 갔다. 그때의 느낌은 말이 나에게 의지하고 간다는 생각이 들었다. 이처럼 무엇인지는 모르지만 공감이 맞다는 생각이 들었다. 이러한 것을 찰떡공감이라고 이야기 하는 것이 아닐까.

300경기 이상된 기수들의 기록 평균값



기수와 말의 호흡



002 본론

경주번호

경주날짜+그 날 경주 회차

군(등급)

말 능력과 전적에 따라 국1~6, 외1~4 등급

출전 두수

출전한 말의 수

주로 상태

:"건조" : 0, "양호" : 1, "다습" : 2,
"포화" : 3, "불량" : 4

부담 중량

기수의 체중에 안장등의 무게를 합친 것으로 박진감 넘치는 경주를 만들기 위한 핸디캡과 같다.
쉽게 말해 밸런스 패치, 부담중량은 이전성적 혹은 상금 수득으로 결정되며 성적이 좋은 말일수록 부담중량이 증가한다. 최대 상한 60kg을 초과하여 부여되지 않는다. 부담중량은 경주에 큰 영향을 미친다.

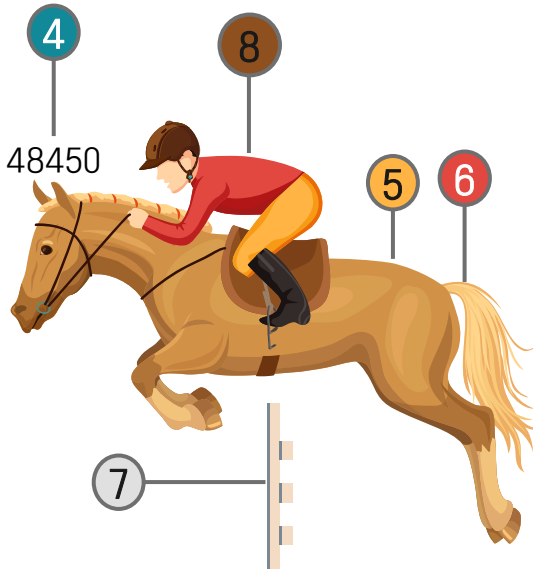
4

5

6

7

8



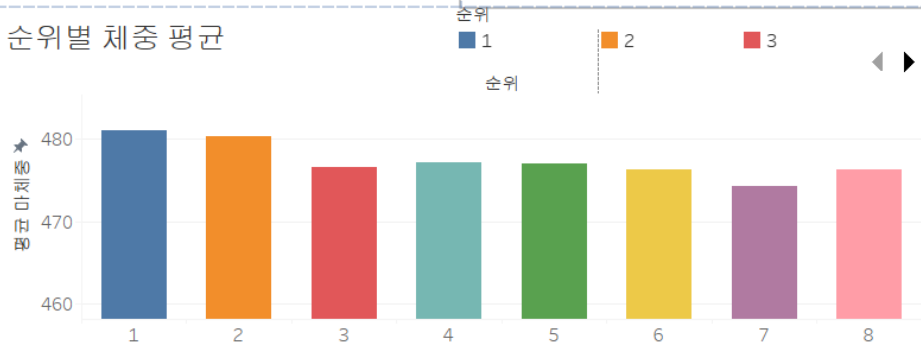
홈 > 경마 > 팻츠한 > [과학은 과학이다]

[경마는 과학이다]말의 체중변화에 따른 배팅전략은?

A. 권용 기자 · © 입력 2022.02.07 14:10 · © 수정 2022.02.16 10:34 · 📄 댓글 0

한국마사회 사이트에 접속해서 경마정보에 들어가 보면 경마일 당일 체중과 평균, 최고, 최저, 3위내 평균 체중 등을 공개한다. 여기서 더욱 눈여겨 보아야 할 것은 3위내 평균 체중이다. 3위내 평균체중을 참고로 계절 및 기온을 감안하여 적절한 체중이 어느 정도가 좋은지를 판단해야 한다. 여름에는 땅 배춧줄로 인해 체중이 좀 더 줄어든다고 겨울에는 조금 늘어나는 것이 일반적이다. 그렇다고 여름에 체중이 늘고 겨울에는 체중이 줄었다고 하여 문제가 되는 것은 아니다.

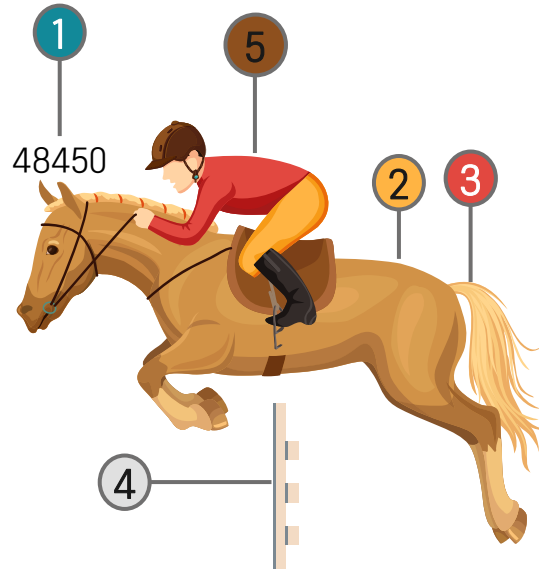
순위별 체중 평균



부담 중량

5

기수의 체중에 안장등의 무게를 합친 것으로 박진감 넘치는 경주를 만들기 위한 핸디캡과 같다. 쉽게 말해 밸런스 패치, 부담중량은 이전성적 혹은 상금 수득으로 결정되며 성적이 좋은 말일수록 부담중량이 증가한다. 최대 상한 60kg을 초과하여 부여되지 않는다. 부담중량은 경주에 큰 영향을 미친다.



002

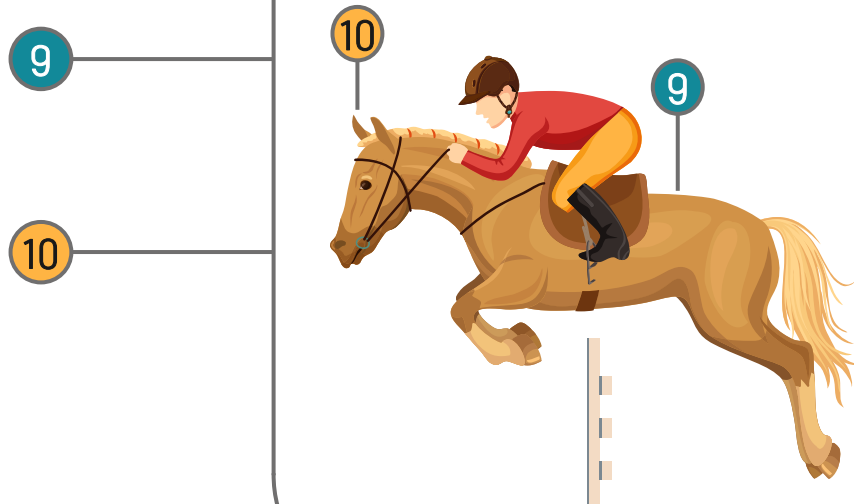
본론

마체중

log변환 이유: 숫자가 다른 변수에 비해
너무 큰 값이어서

군(등급)

경주 당시, 그 말의 능력치를 수치화한
것 (수치화 방법은: 과거 1,2,3위한
횃수의 전적에 각각 3,2,1 가중치를
부여(곱)해서 수치화했다.)



002 본론

데 이 터 컬럼 설명

마번	조교사 점수	기수 점수	경주번호	거리	군(등급)	출전두수	주조상태	기록	부담중량	마체중	순위 점수
46584	2703.103	1847.108	2211139	12	4	8	3	71.6	53	6.265301	0
44261	1150.096	8048.12	2211139	12	4	8	3	71.6	54	6.152733	6
45027	1927.085	3931.112	2211139	12	4	8	3	72.6	51	6.135565	9
45117	1553.087	1442.093	2211139	12	4	8	3	72.6	55.5	6.198479	0
46520	618.1131	698.1077	2211139	12	4	8	3	73.5	52	6.159095	0
44927	221.1118	691.099	2211139	12	4	8	3	74.2	57	6.144186	0
41091	2490.086	1438.104	2211139	12	4	8	3	74.4	55.5	6.182085	0
41147	2288.096	1424.164	2211139	12	4	8	3	75.5	52	6.111467	0
40560	2433.096	4251.106	2211138	20	7	14	3	125.3	55	6.122493	25
38511	970.0862	8048.12	2211138	20	7	14	3	125.6	57	6.228511	0
42917	213.0879	1847.108	2211138	20	7	14	3	126	57	6.248043	0
38733	650.1063	1442.093	2211138	20	7	14	3	126.2	57	6.214608	0
42916	2132.097	1971.082	2211138	20	7	14	3	126.6	57	6.269096	3
42593	1389.093	2080.094	2211138	20	7	14	3	126.7	57	6.159095	2
40633	170.1168	3931.112	2211138	20	7	14	3	126.7	57	6.161207	12

+

부담중량1	마체중1	순위 점수1	조교사 점수1	기수 점수1	부마성적1	모마성적1
51	6.2186001	0	0.398586391	0.00490165	3	6.2
57	6.1800167	3	0.086481304	0.09469662	20.16667	10.22685
57	6.1800167	3	0.086481304	0.09469662	20.16667	10.22685
57	6.1800167	3	0.086481304	0.09469662	20.16667	10.22685
57	6.1800167	3	0.086481304	0.09469662	20.16667	10.22685
57	6.1800167	3	0.086481304	0.09469662	20.16667	10.22685
57	6.1800167	3	0.086481304	0.09469662	20.16667	10.22685
57	6.1800167	3	0.086481304	0.09469662	20.16667	10.22685
57	6.1800167	3	0.086481304	0.09469662	20.16667	10.22685
57	6.1800167	3	0.086481304	0.09469662	20.16667	10.22685
57	6.1800167	3	0.086481304	0.09469662	20.16667	10.22685
57	6.1800167	3	0.086481304	0.09469662	20.16667	10.22685
57	6.1800167	3	0.086481304	0.09469662	20.16667	10.22685
57	6.1800167	3	0.086481304	0.09469662	20.16667	10.22685
57	6.1800167	3	0.086481304	0.09469662	20.16667	10.22685
56.5	6.2519039	0	0.817387636	0.36643735	9.142857	16
51	6.2441669	0	0.040653422	0.00490165	22.09091	10.22685
51	6.2441669	0	0.040653422	0.00490165	22.09091	10.22685

X 12

- 한 개의 로우가 한 경기의 데이터: 비교할 다른 말의 능력치도 필요함
- Ex) 12마리가 출전하는 경기면 제일 주축이 되는 1번 말과 그 옆에는 2번부터 12번말의 능력치 데이터를 옆으로 쌓음.
- 옆으로 쌓은 다른 말의 능력치는 (부담중량, 마체중, 순위점수, 조교사/기수 점수, 부마/모마 성적)

경기 데이터+ 말 데이터

말2 데이터

말3 데이터

말4 데이터

말5 데이터

...

말12 데이터

경기에서 12마리가 뒀 경기인데도 1마리의 데이터만 있는 경우가 있었다.

- 이런 경우는 다른 11마리의 말이 은퇴한 경우였기 때문에 전처리과정에서 제외
- 경주번호를 셋을 때 한 경기 내에서 존재하는 말의 데이터의 개수가 6개 미만이면 제외함

그럼에도 불구하고, 경기마다 출전두 수, 즉 출전하는 말의 수가 달라 데이터를 쌓을 때 길이가 다른 문제 있음.

경기 데이터+ 말 데이터

말2 데이터

말3 데이터

...

002 본론

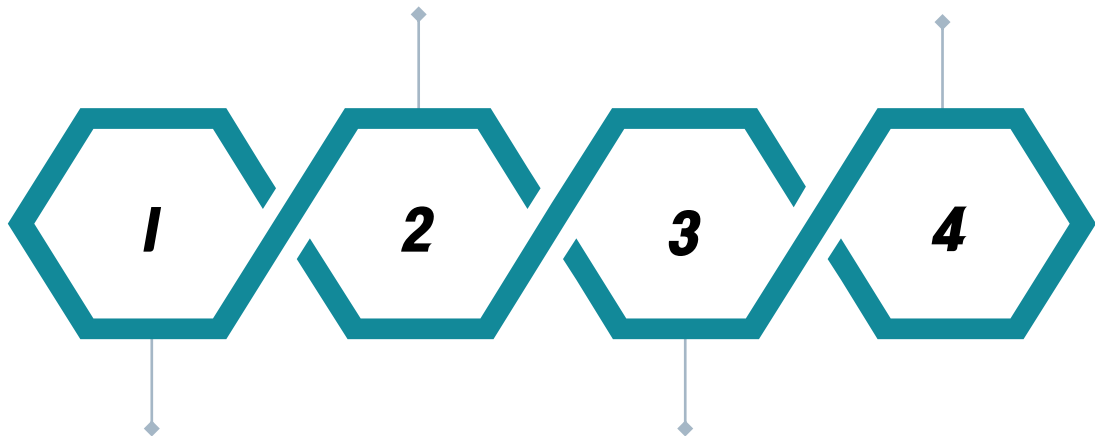
데 이 터 전 처 리



컬럼	채울 값	컬럼	채울 값
부담중량	평균	조교사 점수	평균
마체중	평균	기수 점수	평균
순위점수	0	모마 성적	평균
		부마 성적	평균

데이터 특성을 고려했을 때 의미
없는 피처를 탈락시키는
트리구조의 모델이 좋다고 판단

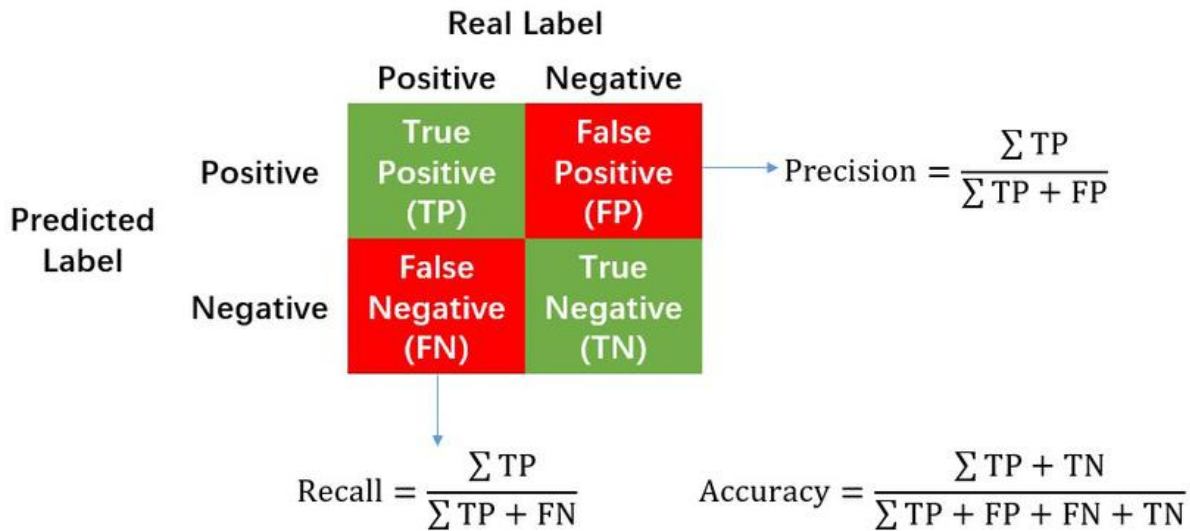
Light GBM, XGBoost,
RandomForest 선택



Null 값이 많음

tree베이스인 모델로
학습할 경우 성능이 좋을 것

Precision



Precision

우리의 목적: 경마 경기 결과 예측

즉 예측 결과가 얼마나 정확한지가 중요

Recall은 실제 3위 중 정확한 예측이 얼마나 있는지의 비율
이는 중요하지 않음

우리의 예측이 얼마나 잘 맞는지가 중요, 이와 관련된 성능을 높여야 함

RandomGridSearch

성능이 너무 떨어짐
-> 너무 한쪽에 몰림...? 멀라...

Optuna

성능 향상이 큼
다만 overfit의 위험 있음
-> 조정

Random Forest

Baseline:
파라미터 조정:
최종 성능:

XGBoost

Baseline:
파라미터 조정:
최종 성능:

LightGBM

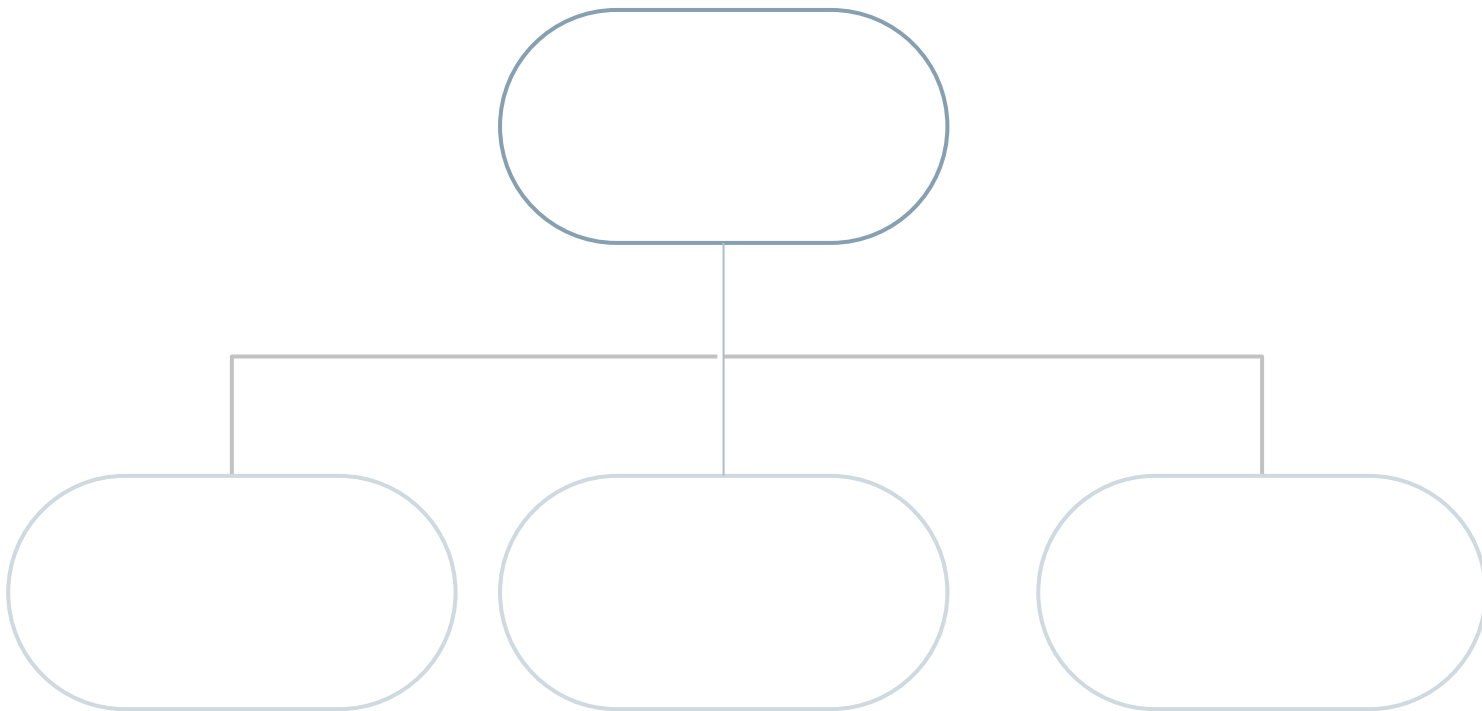
Baseline:
파라미터 조정:
최종 성능:

002

본론

모델 선택

Hard Voting



002 본론

실 예 제 측

지 난 주 경 기
데 이 터

결

과

실제 결과

.....,

예측값

.....

예측값

.....

한계

.....

의의

.....

발전방향

.....