

FIFA

선수 추천 / 팀 예측 프로그램

 20230904 경북대학교 AI·빅데이터 전문가 양성과정 4기
사시옷 : 박소원 이승수 이승훈 최성진

INDEX

01 주제 선정 이유

02 프로그램 개요

02-1 선수 찾으시옷

02-2 팀 만드시옷

03 분석 데이터

04 예측 모델링

05 프로그램 시연

01

주제 선정 이유

01-1 주제 선정 이유

- 구단주의 입장에서 **Potential이 높은 선수를 Value가 낮을 때 영입하는 것이 유리함**
- 하지만, Value 대비 Potential이 높은 선수를 찾기 **어려움**
- 따라서, 회귀 모델을 통해 **미래의 선수가치를 예측** 하여 낮은 금액으로 높은 Potential의 선수를 미리 영입하는 것이 목표

01-2 업무 별 담당자

역할 구분	담당자	비고
주제 선정	전원	
자료 조사	전원	
데이터 전처리	전원	
모델링 1	최성진	Overall 예측 모델링
모델링 2	이승수	Potential 예측 모델링
모델링 3	박소원	Value 예측 모델링
모델링 4	이승훈	Wage 예측 모델링
크롤링	이승수	
함수 구현	이승훈, 최성진	
발표 자료 준비	박소원	

02

프로그램 개요

02-1 선수 찾으시옷

[선수 추천 프로그램]

- 이적 시장에서 성공적인 선수 영입 및 방출을 위해 원하는 능력치와 금액에 맞는 선수 추천
- 원하는 수준의 **Overall** 입력
 ⇒ 입력된 값에 적합한 선수 top 10을 추천

02-1 선수 찾으시옷

[주요 제공 항목]

- **포지션** : FW / DF / MID / GK
- **Overall** : 선수의 종합 능력치
- **Potential** : 선수의 잠재력, **높은** 순 추천
- **Value** : 선수의 현재 가치, **낮은** 순 추천
- **Wage** : 선수의 주급, **낮은** 순 추천
- 위 4가지 항목에 대한 그래프 제공

02-2 팀 만드시옷

[팀 Overall 예측 프로그램]

- [선수 찾으시옷]을 통해 11명의 선수 선택
⇒ 생성한 **팀의 Overall 예측 프로그램**
- 생성한 팀과 같은 Overall을 가진
실제 구단 및 실제 구단 기준 **순위**를 제공
⇒ 생성한 팀의 **Overall 수준 파악**

03

분석 데이터

03-1 데이터 출처

[데이터 출처]

■ 선수 데이터 파일 출처

- <https://www.kaggle.com/datasets/stefanoleone992/fifa-22-complete-player-dataset>
- 18~22년 데이터 파일

■ 팀 데이터 크롤링 출처

- <https://sofifa.com/teams>

03-2 데이터 전처리

[데이터 전처리]

■ 분석 특성 선정

- 세부능력치를 **분야별 종합능력치** 평균으로 분석
- 분야별 종합능력치 결측치 제거
- 이상치 / 중복값 없음
- 포지션 **FW, MID, DF, GK**로 통일

	name	age	overall	potential	value_eur	wage_eur	position	attacking	skill	movement	power	mentality	defending	goalkeeping
	L. Messi	34	93	93	78000000.0	320000.0	RW	85.8	79.000000	90.2	77.8	73.833333	28.25	10.8
R.	Lewandowski	32	92	92	119500000.0	270000.0	ST	86.0	68.500000	81.6	84.8	80.666667	35.00	10.2
istiano	Ronaldo	36	91	91	45000000.0	270000.0	ST	87.6	70.500000	85.4	87.2	74.333333	28.50	11.6
	Neymar Jr	29	91	91	129000000.0	270000.0	LW	80.6	75.166667	90.2	71.8	77.000000	33.25	11.8
	K. De Bruyne	30	91	91	125500000.0	350000.0	CM	81.4	74.000000	80.0	81.6	82.666667	62.50	11.2

	D. Da Silva	21	47	55	90000.0	500.0	GK	12.4	11.666667	34.0	35.4	17.166667	10.00	46.2
	H. Singh	18	47	64	120000.0	500.0	CM	39.2	33.500000	62.4	48.2	47.166667	41.50	12.2
	M. O'Brien	18	47	61	110000.0	500.0	RM	41.0	34.500000	54.0	41.4	40.666667	25.25	10.8
	C. Porter	19	47	59	110000.0	500.0	CM	43.2	36.500000	60.0	48.8	46.500000	42.25	9.4
	N. Logue	21	47	55	100000.0	500.0	CM	40.0	36.833333	56.6	49.4	42.500000	42.75	7.4

◀ 데이터 전처리
완료 결과

04

예측 모델링

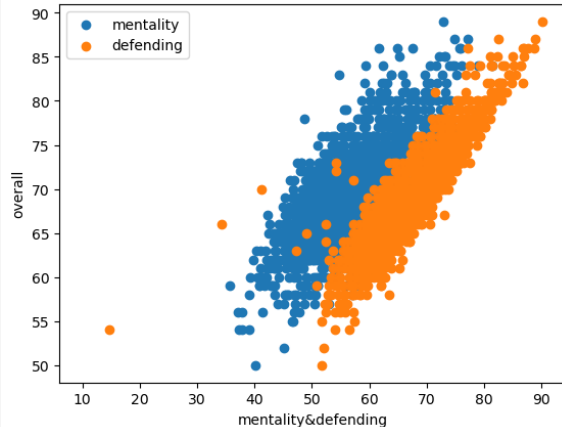
04-1 Overall



[선수의 능력치 예측]

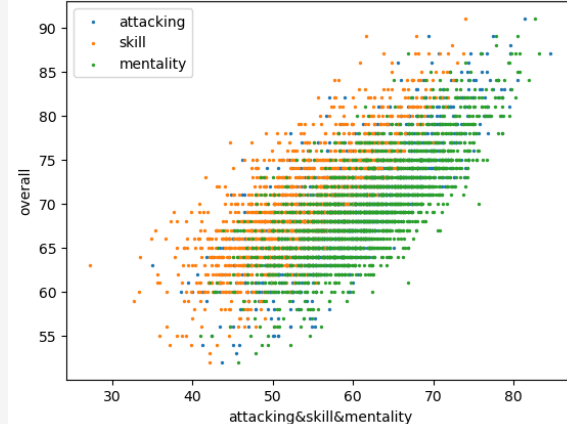
- 주요 학습 방법 : 회귀
- 세부 학습 방법 : **선형회귀(LinearRegression)**
- 특성 공학 : 골키퍼/수비수/미드필더/공격수 나눠서 분석

수비수



Feature : 수비력, 정신력
Target : 능력치

미드필더



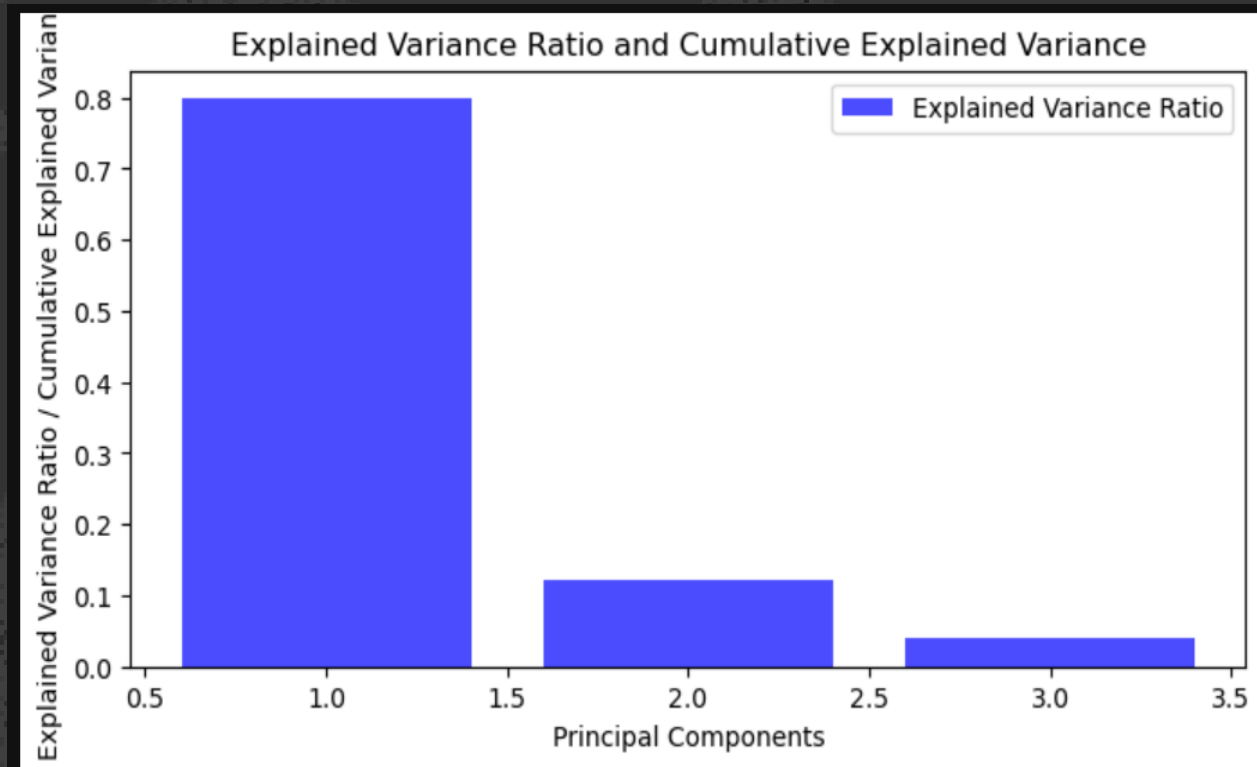
Feature : 공격력, 스킬, 정신력
Target : 능력치

04-2 Potential

- 주요 학습 방법 : 주성분 분석
- 세부 학습 방법 : **Random Forest Regressor**

특성과 타겟의 상관관계 비교 결과

관련 있는 변수가 없어서 주성분분석으로 차원 축소하여 분석함

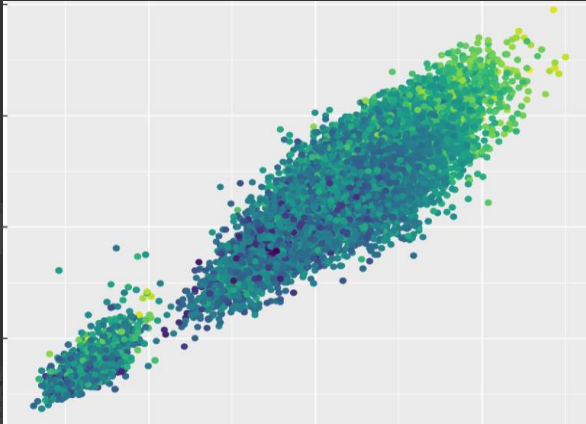


주성분 2개 결정
⇒ **2차원** 축소

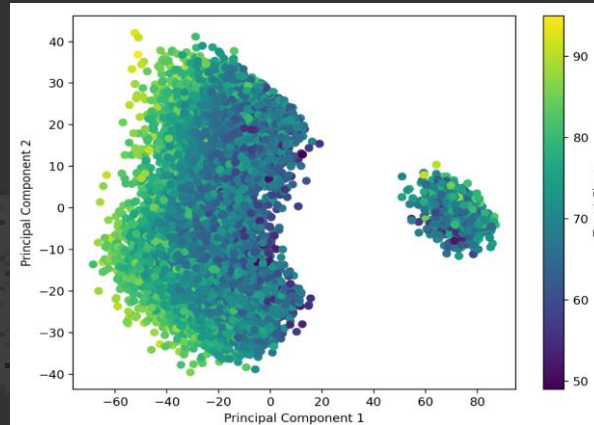
04-2 Potential



[실제 데이터 분포]



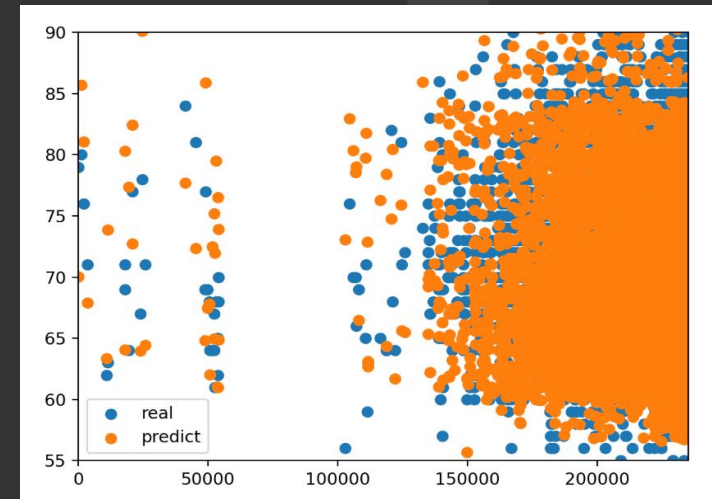
[차원축소 데이터 분포]



↓

주성분 분석한 특성으로
랜덤포레스트 회귀분석

→



[실제 값과 예측 값]

04-3 Value



[선수의 가치 예측]

- 주요 학습 방법 : 회귀 분석
- 세부 학습 방법 : **Bagging Regressor**
- 모델 선정 과정

1. 모델 선정 함수를 통해 5개의 모델 선별

```
('HistGradientBoostingRegressor', 0.95, 0.94, 0.92),  
( 'MLPRegressor', 0.95, 0.94, 0.92),  
( 'KNeighborsRegressor', 0.92, 0.94, 0.93),  
( 'BaggingRegressor', 0.91, 0.95, 0.95),  
( 'RandomForestRegressor', 0.91, 0.95, 0.95),  
( 'GradientBoostingRegressor', 0.9, 0.95, 0.94),  
( 'DecisionTreeRegressor', 0.89, 0.95, 0.95),
```

2. 각 모델 학습 후 결정 계수(R Square)가 가장 좋은 모델 선정

⇒ **Bagging Regressor (score : 0.9252)**

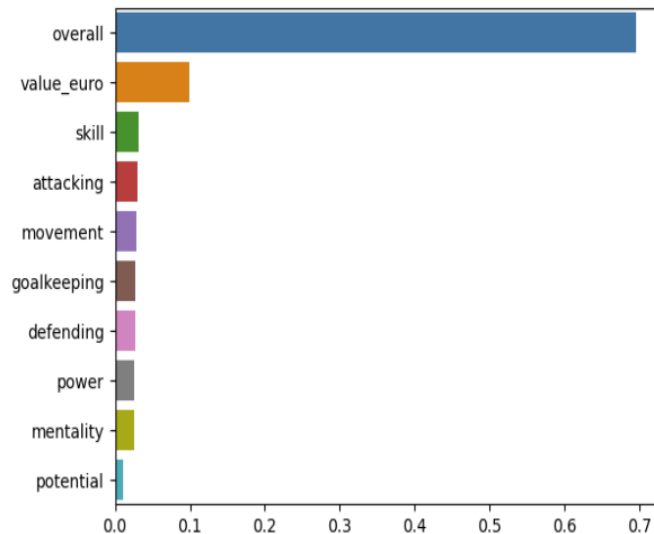
04-4 Wage



[선수의 가치 예측]

- 주요 학습 방법 : 회귀
- 세부 학습 방법 : 랜덤포레스트(Regressor)

모델 시각화



중요도가 가장 높은
Overall, value, skill 을
Feature 변수로 선정



훈련점수 : **0.82**
테스트점수 : **0.79**

05

프로그램 시연

05-1 선수 찾으시옷

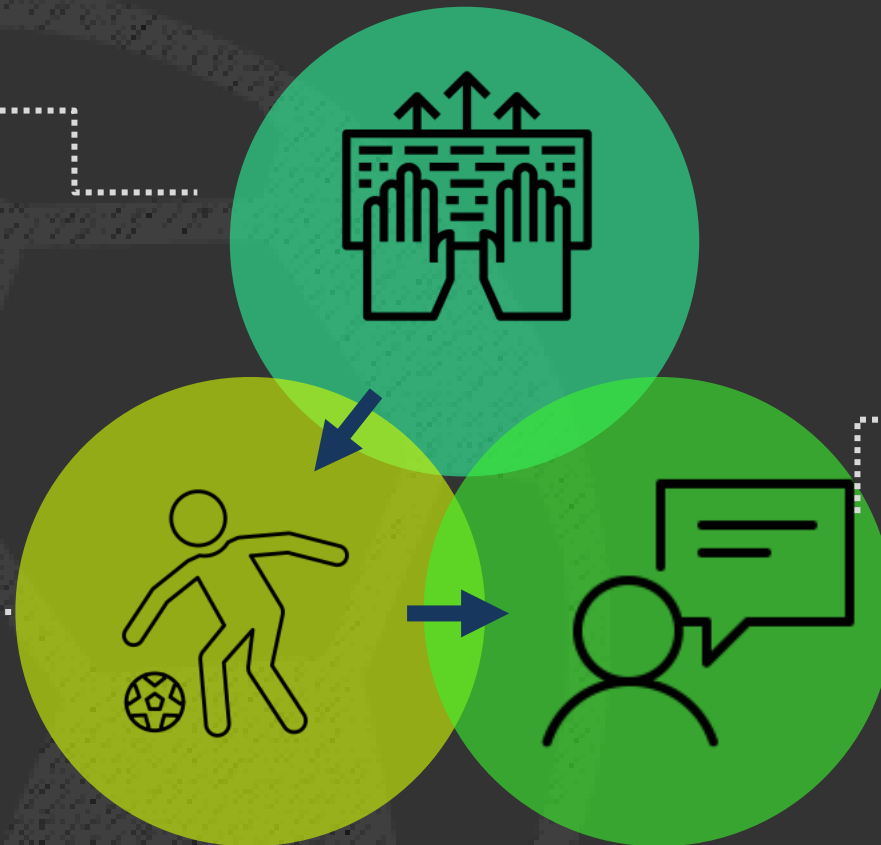
선수 Overall
입력



선수 입력 시
선수 정보 URL 팝업



해당 Overall 가진
선수 10명 출력

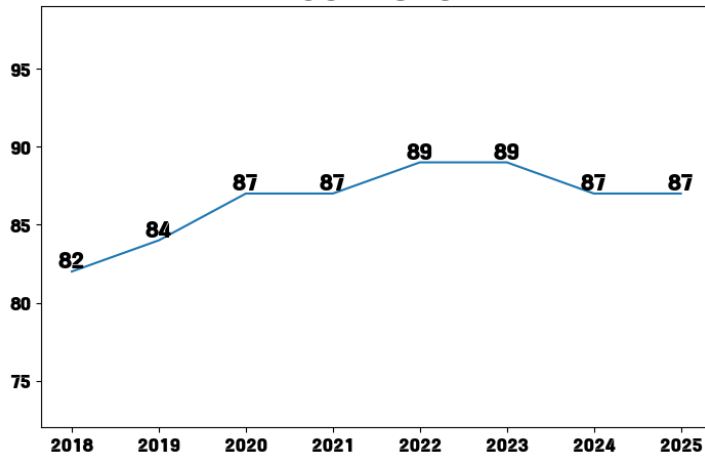


05-1 선수 찾으시요

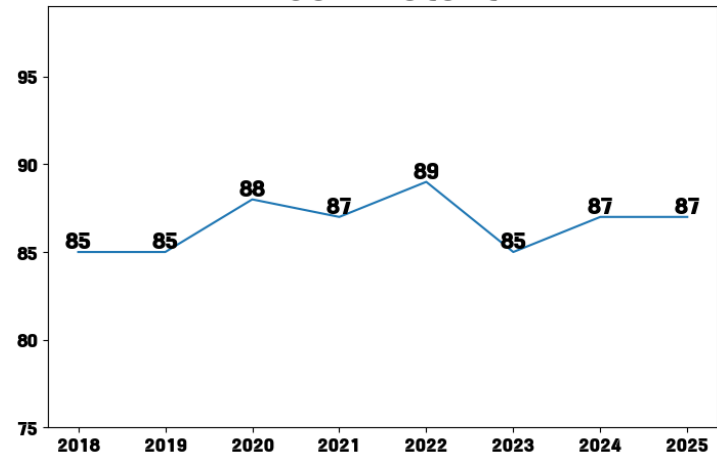


[선수의 변화량 그래프]

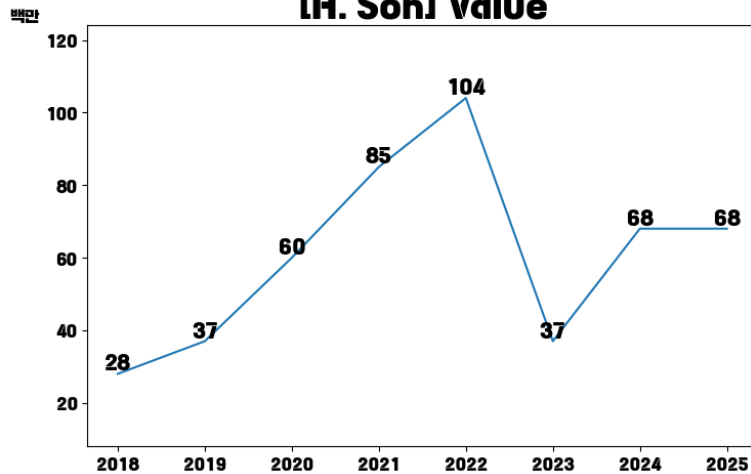
[H. Son] Overall



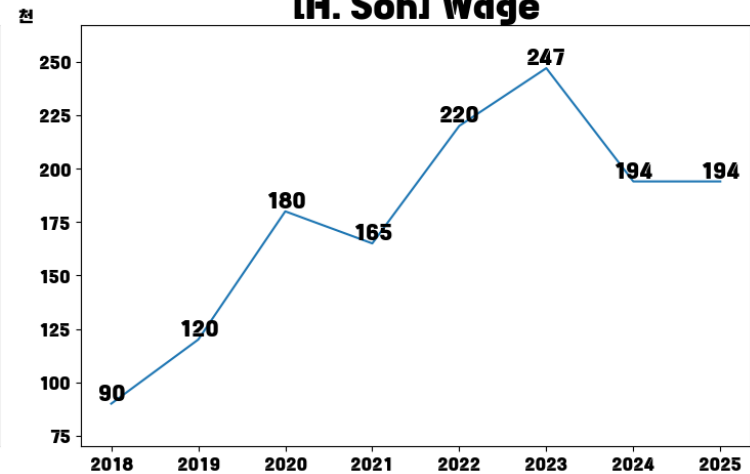
[H. Son] Potential



[H. Son] Value



[H. Son] Wage



05-2 팀 만드시옷



[11명 선수 ID 입력]

- 입력한 **선수 정보** 출력
- 내가 뽑은 선수들의 **능력치 평균**과 같은 **팀** 출력

선수 정보

sofifa_id	name	overall	value_eur	wage_eur	potential	value_bg
193041	K. Navas	75	15500000	130000	87.60	12386667
216267	A. Robertson	84	83500000	175000	79.22	19539666
231281	T. Alexander-Arnold	82	114000000	150000	78.74	17983769
239818	Rúben Dias	84	102500000	170000	79.87	20048407
189332	Jordi Alba	80	47000000	200000	80.15	15130274
215914	N. Kanté	77	100000000	230000	82.82	16247359
200145	Casemiro	79	88000000	310000	79.09	14590409
212622	J. Kimmich	82	108000000	160000	81.75	24480986
239085	E. Haaland	81	137500000	110000	79.02	17983769
208830	J. Vardy	83	33000000	180000	77.43	17584895
179813	E. Cavani	84	26000000	190000	80.27	20048407

비슷한 팀 순위

	team	TOTAL_MEAN	rank
12	Tottenham Hotspur	81	13
13	Borussia Dortmund	81	13
14	RB Leipzig	81	13
15	Milan	81	13



THANK YOU



20230904 경북대학교 시·빅데이터 전문가 양성과정 4기
사시옷 : 박소원 이승수 이승훈 최성진