

Soccer Player Performance Prediction

2021 혁신_인공지능
채민석

주제 선정 배경

1) 첫 프로젝트 : 데이터 전처리에 높은 비중

- 방대한 Dataset -> 필요 데이터 분류 및 추출
- 데이터 가공 및 전처리 스킬 향상

2) 주제 선정시 고려항목

- 수치화 가능한 데이터 보유
- 유의미한 결과값 도출 가능
- 실생활과의 연관성 및 흥미유발

3) 선정 결과

축구선수 스탯 분석을 통한 능력치 예측

개요

1) Datasets : <https://www.kaggle.com/hugomathien/soccer>









The ultimate Soccer database for data analysis and machine learning

What you get:

- +25,000 matches
- +10,000 players
- 11 European Countries with their lead championship
- Seasons 2008 to 2016
- Players and Teams' attributes* sourced from EA Sports' FIFA video game series, including the weekly updates
- Team line up with squad formation (X, Y coordinates)
- Betting odds from up to 10 providers
- Detailed match events (goal types, possession, corner, cross, fouls, cards etc...) for +10,000 matches

Data Explorer

298.59 MB

- ▼  database.sqlite
 -  Country
 -  League
 -  Match
 -  Player
 -  Player_Attributes
 -  Team
 -  Team_Attributes


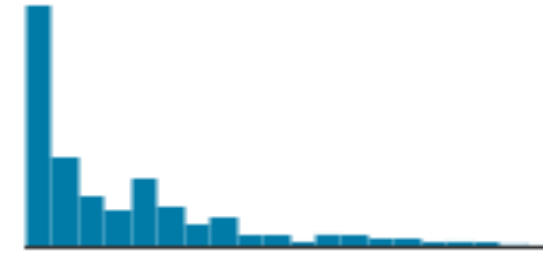
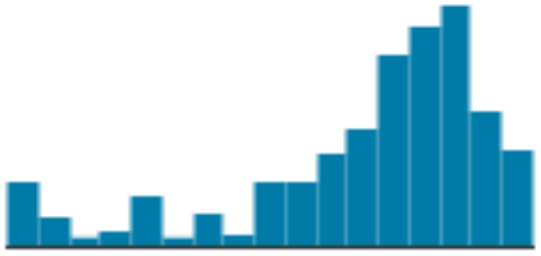
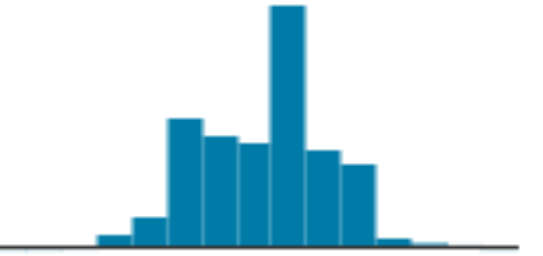
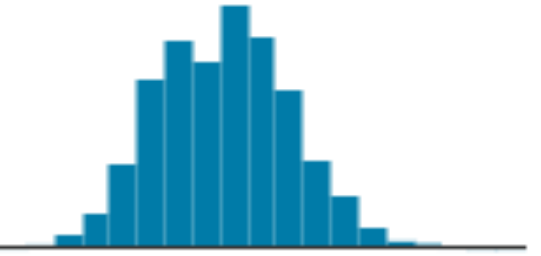
2) Player 관련 데이터 추출

2-1) Player

Data Explorer

298.59 MB

- database.sqlite
 - Country
 - League
 - Match
 - Player
 - Player_Attributes
 - Team
 - Team_Attributes

< Player (11.1k rows)								
Detail Compact Column							7 of 7 columns	
# id	# player_api_id	▲ player_name	# player_fifa_api_id	▲ birthday	# height	# weight		
 1 11.1k	 2625 751k	10848 unique values	 2 234k	5762 unique values	 157 208	 117 243		
1	505942	Aaron Appindangoye	218353	1992-02-29 00:00:00	182.88	187		
2	155782	Aaron Cresswell	189615	1989-12-15 00:00:00	170.18	146		
3	162549	Aaron Doran	186170	1991-05-13 00:00:00	170.18	163		
4	30572	Aaron Galindo	140161	1982-05-08 00:00:00	182.88	198		
5	23780	Aaron Hughes	17725	1979-11-08 00:00:00	182.88	154		
6	27316	Aaron Hunt	158138	1986-09-04 00:00:00	182.88	161		
7	564793	Aaron Kuhl	221280	1996-01-30 00:00:00	172.72	146		
8	30895	Aaron Lennon	152747	1987-04-16 00:00:00	165.1	139		
9	528212	Aaron Lennox	206592	1993-02-19 00:00:00	190.5	181		
10	101042	Aaron Meijers	188621	1987-10-28 00:00:00	175.26	170		
11	23889	Aaron Mokoena	47189	1980-11-25 00:00:00	182.88	181		
12	231592	Aaron Mooy	194958	1990-09-15 00:00:00	175.26	150		
13	163222	Aaron Muirhead	213568	1990-08-30 00:00:00	187.96	168		
14	40719	Aaron Niguez	183853	1989-04-26 00:00:00	170.18	143		
15	75489	Aaron Ramsey	186561	1990-12-26 00:00:00	177.8	154		
16	597948	Aaron Splaine	226014	1996-10-13 00:00:00	172.72	163		
17	161644	Aaron Taylor-Sinclair	213569	1991-04-08 00:00:00	182.88	176		

2-2) Player_Attributes

Data Explorer

298.59 MB

▼

database.sqlite

Country

League

Match

Player

Player_Attributes

Team

Team_Attributes

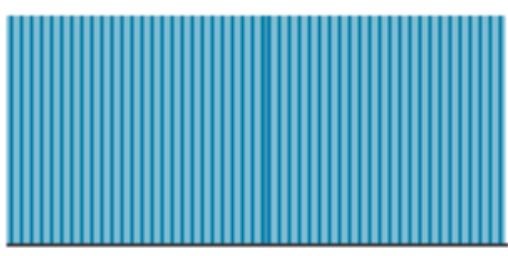
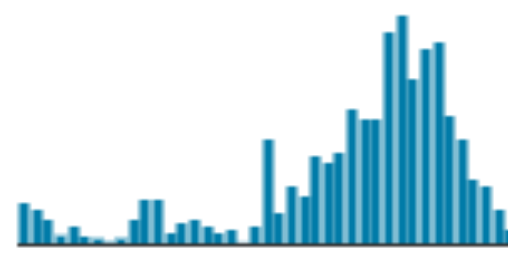
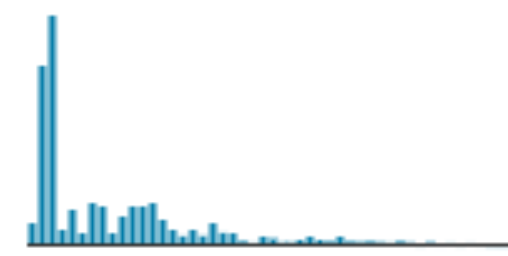
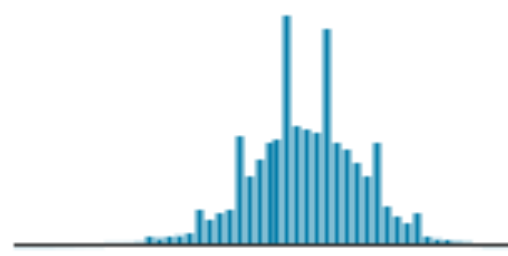
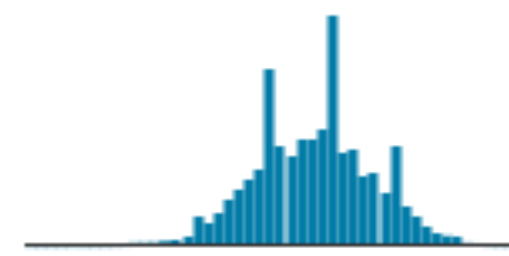
< Player_Attributes (184k rows)

Detail

Compact

Column

10 of 42 columns ▼

# id	# player_fifa_api_id	# player_api_id	▲ date	# overall_rating	# potential	▲ preferred_foot	▲ attacking_work_r...	▲ defensive_work_r...
 1184k	 2234k	 2625751k	2007-02-22 00:00:00 6% 2013-09-20 00:00:00 4% Other (165641) 90%	 3394	 3997	right 75% left 24% Other (836) 0%	medium 68% high 23% Other (16085) 9%	medium high Other (26091)
1	218353	505942	2016-02-18 00:00:00	67	71	right	medium	medium
2	218353	505942	2015-11-19 00:00:00	67	71	right	medium	medium
3	218353	505942	2015-09-21 00:00:00	62	66	right	medium	medium
4	218353	505942	2015-03-20 00:00:00	61	65	right	medium	medium
5	218353	505942	2007-02-22 00:00:00	61	65	right	medium	medium
6	189615	155782	2016-04-21 00:00:00	74	76	left	high	medium
7	189615	155782	2016-04-07 00:00:00	74	76	left	high	medium
8	189615	155782	2016-01-07 00:00:00	73	75	left	high	medium
9	189615	155782	2015-12-24 00:00:00	73	75	left	high	medium
10	189615	155782	2015-12-17 00:00:00	73	75	left	high	medium
11	189615	155782	2015-10-16 00:00:00	73	77	left	high	medium
12	189615	155782	2015-09-25 00:00:00	74	78	left	high	medium
13	189615	155782	2015-09-21 00:00:00	73	77	left	medium	medium
14	189615	155782	2015-01-09 00:00:00	71	75	left	medium	medium
15	189615	155782	2014-12-05 00:00:00	71	77	left	medium	medium
16	189615	155782	2014-11-07 00:00:00	71	77	left	medium	medium
17	189615	155782	2014-09-18 00:00:00	70	77	left	medium	medium

3) x, y값 설정

y = overall_rating

x = Stat 관련 46종

height, wieght, birthday, potential, crossing, finishin, heading_accuracy, short_passing, volleys, dribbling, curve, free_kick_accuracy, long passing, ball_control, acceleration, sprint_speed, agility, reactions, balance, shot_power, jumping, stamina, strength, long_shots, aggression, interceptions, positioning, vision, penalties, marking, standing_tackle, sliding_tackle, gk_diving, gk_handling, gk_kicking, gk_positioning, gk_reflexes, preferred_foot, attacking_work_rate_high, attacking_work_rate_low, attacking_work_rate_medium, defensive_work_rate_high, defensive_work_rate_low, defensive_work_rate_medium

기대성과

1) sql -> df (pd.read_sql_query)

- select 구문 학습

2) 46종 Stat 전처리

- 다양한 value 수치화(binary, one-hot encoding 등)
- column별 scaling

=> 데이터 가공 및 전처리 스킬 향상