22-2 중앙대학교 소프트웨어학부
빅데이터

Final Report

FIFA 랭킹과 FIFA 게임 시리즈 선수 데이터를 이용한
2022 FIFA World Cup 우승자 예측

1. 개요

FIFA 랭킹과 FIFA 게임 시리즈의 선수 데이터를 이용한 2022 World Cup 우승자 예측

A. 사용 Dataset

i. International football results from 1872 to 2022

ii. FIFA World Ranking 1992-2022

iii. FIFA World Ranking 1992-2022

iv. FIFA 22 complete player dataset

B. 학습 모델

i. Xgboost

ii. MultiOutputRegressor

C. 예측 결과

i. 16강 진출 팀

- A조: Netherlands, Senegal
- B조: England, United States
- C조: Argentina, Poland
- D조: France, Australia
- E조: Japan, Spain
- F조: Morocco, Croatia
- G조: Brazil, Switzerland
- H조: Portugal, South Korea

ii. 결승

- Senegal 우승

2. 문제 정의

A. 주제

과거 축구 경기 기록, 국가별 축구 랭킹과 선수 데이터를 이용해 2022 FIFA World Cup 우승자를 예측하는 예측모델이 프로젝트의 주제이다.

이를 위해 우리는 경기를 진행할 두 국가의 정보와 선수 데이터를 입력으로 주었을 때 점수를 예측하는 모델을 만들었다.

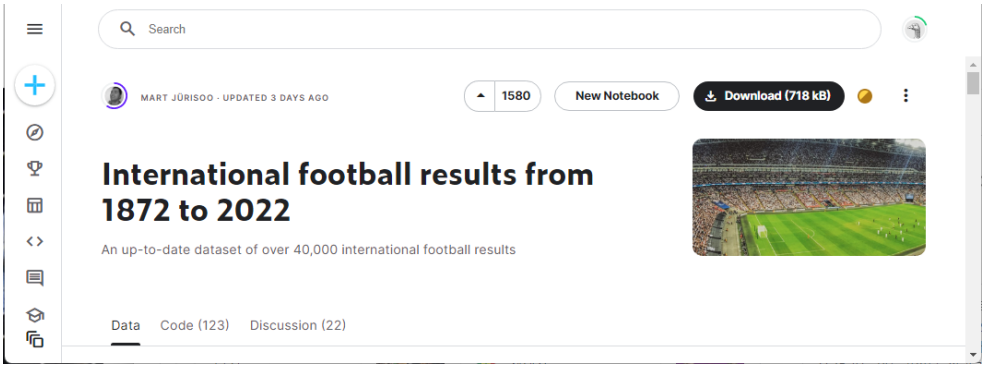
2022 FIFA World Cup 시뮬레이션에서 모든 경기에 대해 점수 예측모델을 이용해 예선전과 본선을 치러 우승 국가를 예측한다.

B. Dataset

i. 과거 축구 경기 기록

International football results from 1872 to 2022

(https://www.kaggle.com/datasets/martj42/international-football-results-from-1872-to-2017)



1872년도부터 현재까지 진행된 친선경기를 포함한 남자 국가대표 팀간 진행된 축구 경기 dataset이다.

	date	home_team	away_team	home_score	away_score	tournament	city	country	neutral
0	1872-11-30	Scotland	England	0.0	0.0	Friendly	Glasgow	Scotland	False
1	1873-03-08	England	Scotland	4.0	2.0	Friendly	London	England	False
2	1874-03-07	Scotland	England	2.0	1.0	Friendly	Glasgow	Scotland	False
3	1875-03-06	England	Scotland	2.0	2.0	Friendly	London	England	False
4	1876-03-04	Scotland	England	3.0	0.0	Friendly	Glasgow	Scotland	False

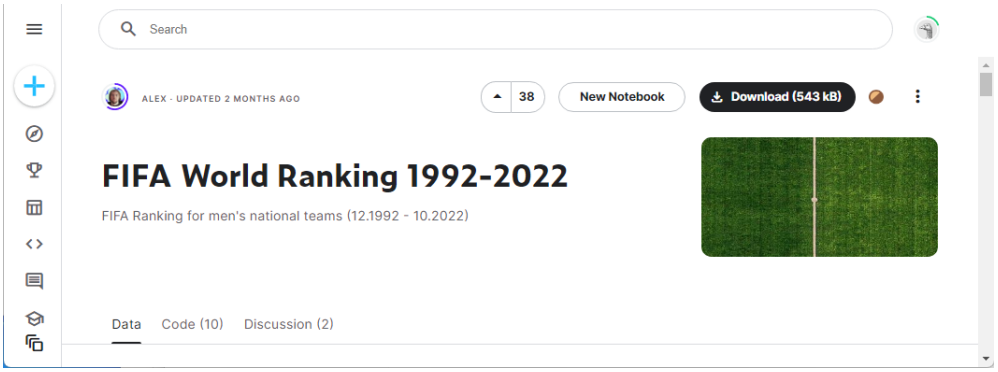
경기 진행 날짜, 홈팀명, 원정팀명, 홈팀 점수, 원정팀 점수, 대회명, 경기 진행

도시 및 국가와 중립국에서 경기를 진행했는지 여부를 담고 있다. 1872년 11월 30일부터 2022년 11월 20일까지의 경기 기록을 갖고 있으며 매달 Dataset이 업데이트 된다. 현재는 2022 FIFA Qatar World Cup 경기 기록도 업데이트되었으나 해당 데이터는 제거했다.

ii. 국가별 축구 랭킹

FIFA World Ranking 1992-2022

(https://www.kaggle.com/datasets/cashncarry/fifaworldranking)



공식 FIFA 웹사이트에서 수집된 1993년 8월부터 2022년 10월까지의 사용 가능한 모든 FIFA 남자 국제 축구 순위이다

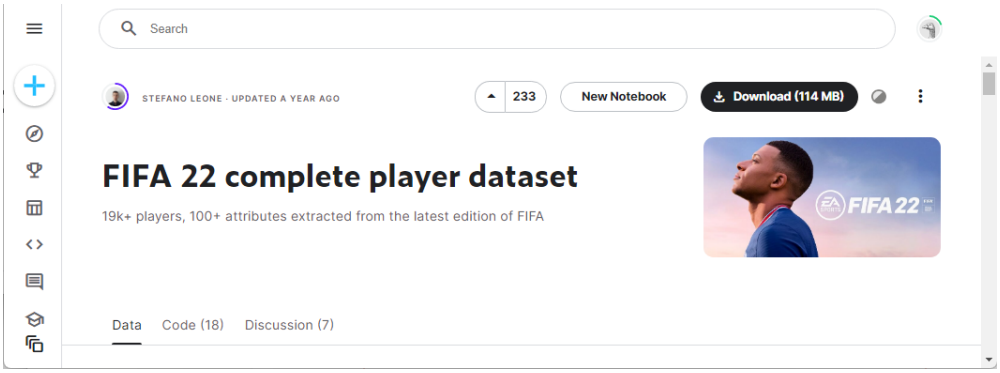
	rank	country_full	country_abrv	total_points	previous_points	rank_change	confederation	rank_date
0	1	Germany	GER	57.0	0.0	0	UEFA	1992-12-31
1	96	Syria	SYR	11.0	0.0	0	AFC	1992-12-31
2	97	Burkina Faso	BFA	11.0	0.0	0	CAF	1992-12-31
3	99	Latvia	LVA	10.0	0.0	0	UEFA	1992-12-31
4	100	Burundi	BDI	10.0	0.0	0	CAF	1992-12-31

랭크, 국가명, 국가 약어, Ranking 점수, 이전 버전 ranking 점수, rank 변동값, 소속 연합, rank 개시 날짜를 담고 있다.

iii. 선수 데이터

FIFA 22 complete player dataset

(https://www.kaggle.com/datasets/stefanoleone992/fifa-22-complete-player-dataset)

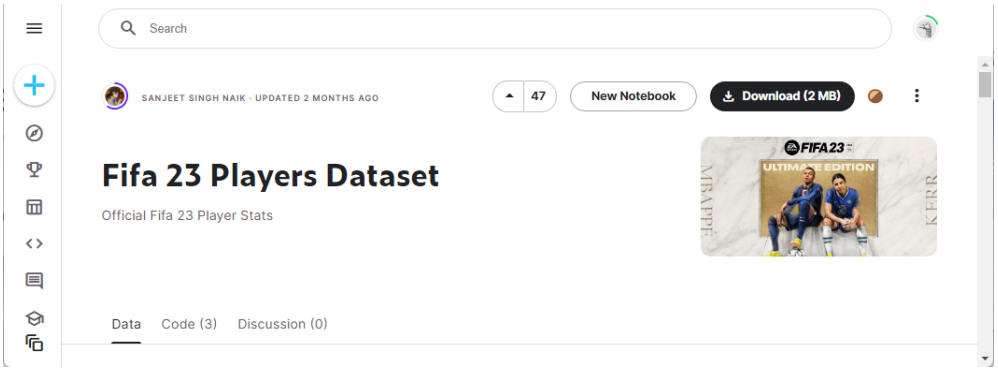


FIFA 게임 시리즈 FIFA 15부터 FIFA 22까지 게임에 존재하는 모든 선수의 Career Mode 선수 데이터셋이다.

각 선수마다 총 110개의 속성값을 갖는다. 선수 이름, 신체정보, 소속 클럽과 관련된 속성, 선수 가치와 임금, 총평 점수, 선수의 잠재력, 소속국가, 축구 능력치 관련 속성들이 존재한다. 결측값이 존재하는 문제가 있다.

Fifa 23 Players Dataset

(https://www.kaggle.com/datasets/sanjeetsinghnaik/fifa-23-players-dataset)



FIFA 게임 시리즈 중 FIFA 2에 존재하는 모든 선수의 Career Mode 선수 데이터셋이다.

각 선수마다 총 89개의 속성값을 갖는다. 선수 이름, 신체정보, 소속 클럽과 관련된 속성, 선수 가치와 임금, 총평 점수, 선수의 잠재력, 소속국가, 축구 능력치 관련 속성들이 존재한다. 결측값이 존재하는 문제가 있다.

C. 주제 및 데이터 선정 이유

i. Kaggle의 많은 주제 중 가장 구체적으로 상황을 그릴 수 있는 데이터였다. 그렇기에 Data에 직접으로 드러나 있지 않은 연관성과 Data를 그려보기 적합했다.

ii. 운동 경기에는 수많은 변수들이 존재한다. 우리에게 주어진 데이터를 보며 어떤 변수가 Data에 들어가 있을지 고민하고 옳은 가정이었는지 Data를 분석하는 과정을 진행하기 좋은 Dataset이라고 생각했기에 위의 데이터셋과 주제를 선정했다.

iii. 위의 Dataset은 FIFA World Cup 우승자 예측에 관련된 Dataset 중 가장 오랜 기간의 Data를 가지면서 최신 Data를 포함하는 Dataset이었기에 해당 4개의 데이터셋을 선정했다.

3. 기존 접근법 분석

A. 접근법

i. 과거 경기 데이터(홈팀, 원정팀, 각 팀별 점수, 중립국 여부)와 국가별 Ranking Point만을 이용해 승부 예측모델을 만들었다.

B. 사용 예측 알고리즘

i. DNN, Logistic regressions, Random Forest, LightGBM

4. 차별점

A. 승부예측 시 선수 데이터셋을 활용한다.

i. Kaggle에 동일한 Dataset을 가지고 축구 경기 승부예측을 한 코드 중 선수 데이터를 포함한 경우는 없었다.

ii. 선수 데이터의 경우 결측값이 많고, 모든 경기에 대해 어떤 선수가 경기에 출전했는지 알 수 없는 부분, 나라별로 경기에 실제로 출전한 선수 숫자가 다르다는 점 등 많은 예외사항으로 인해 정형화하기 어렵기 때문이다.

iii. 이런 예외들을 모두 처리해 학습 데이터로 만들어 승부 예측에 사용했다.

5. 학습 Data 생성을 위한 Data 분석 및 처리

A. Data 전처리

i. 선수 데이터

- 축구 능력치와 관련 없는 속성값은 제거했다.

  전체 속성값 중 25개의 속성값을 제거했다.

- 선수 능력치 결측값

  FIFA 15에서 FIFA 22까지의 Dataset의 경우 포지션이 골키퍼인 선수의 경우 선수 능력치 중 pace, shooting, passing, dribbling, defending, physic 값이 없었다. 골키퍼에게 중요한 능력치는 goal keeping관련 능력치라 판단해 살펴본 결과 goal keeping관련 능력치에는 결측값이 없었기에 FIFA 23 Dataset 또한 골키퍼의 pace, shooting, passing, dribbling, defending, physic 값을 0으로 처리했다.

- 선수 Value, Wage 결측값

  인터넷 상의 데이터에서 Value, Wage 결측값을 찾아보았으나, 원래 게임 데이터 자체에 0 값으로 처리되어 있는 데이터라 결측값을 해결할 수 있는 방법을 찾을 수 없었다. 선수의 Value와 Wage가 경기 결과에 큰 영향을 줄 것으로 예측되지만 이를 처리할 수 없어 Dataset에서 제거했다.

ii. 통일된 국가 목록 사용

각 Dataset마다 Data에 존재하는 국가 목록이 모두 다르다. 학습을 위해 처리된 선수 Dataset을 기준으로, 선수 Dataset에 존재하는 국가의 경기 결과와 Ranking data만을 추출했다. 또한 각 Dataset 마다 동일 국가여도 표기명이 다른 경우가 존재했다. 이러한 경우 경기결과 Dataset을 기준으로 국가명을 통일해 주었다.

iii. 4개의 Dataset 연결성

경기 날짜에 따라 몇 년도 Ranking 값과 몇 년도 선수들이 연결되는지가 달라진다. 경기 날짜를 기준으로 그 당시 최신 Ranking 값과 해당 년도 선수들을 해당 경기 Data와 결합시켰다.

B. 학습을 위한 데이터 생성

선수의 결측값은 전처리 과정에서 처리했으나, 4개의 데이터셋을 하나의 모델 Input Data로 합치는 과정이 필요했다.

i. 경기 출전 선수 문제

4개의 Dataset을 통합할 때, 각 경기별로 국가마다 출전한 선수를 지정해줘야 했다. 하지만 모든 경기에 대해 출전한 선수 명단을 얻을 수 없었으며 국가마다 경기에 출

전한 선수의 수가 모두 다른 문제가 발생했다. 다음은 학습 데이터 생성 과정에서 존재한 선수관련 문제이며 이에 대한 처리이다.

- 국가마다 실제로 경기에 출전한 선수 수가 다른 문제

  A. 축구에는 예비선수와 선수 교체 시스템이 있어 경기별로, 팀별로 참여한 선수의 수가 다르다. 이는 학습 데이터 생성에 있어 데이터 형식을 정형화할 수 없는 문제를 발생시켰다.

  B. 이를 해결하기 위해 각 경기에는 국가마다 11명의 선수만 출전하며 예비 선수 및 선수 교체는 없었다고 가정한다.

- 선수 포지션과 축구 전술 문제

  A. 각 축구팀마다 펼치는 축구 전술(포메이션)이 다르기 때문에 출전 선수의 포지션도 동일한 11개의 포지션으로 나눌 수 없는 문제가 존재한다.

  B. 이를 위해 모든 국가는 수비수 4명, 미드필더 4명, 공격수 2명을 운용하는 전술(4-4-2 전술)을 펼친다 가정한다. 현대 대부분 팀들의 기본 전술 중 하나이기 때문에 해당 전술을 선택했다. 구체적인 포지션(ST, RW, LW, LM, RM 등)은 학습 요소에서 제거했다.

- 11명의 선수 선정 문제

  A. 국가마다 11명의 선수만 출전하며 예비 선수 및 선수 교체가 없다고 가정했기에 경기를 치를 11명의 선수를 추출해야 했다.

  B. 선수마다 각 포지션에서 가장 높은 능력치를 기준으로 포지션별로 능력치가 높은 선수를 추출했다. 공격수 2명을 예로 들면, ST, LW, LF, CF, RF, RW 포지션 능력치 중 가장 높은 값을 대표 공격수 능력치로 추출해 그 값이 가장 높은 2명을 공격수로 선정했다. 동일한 능력치 값을 갖는 경우 관련 포지션 능력치의 평균이 높은 쪽을 선정했다.

ii. 선수 데이터 부족 문제

위의 가정을 기반으로 선수 Dataset을 처리하는 과정에서 국가에 선수가 11명 미만인 국가와 골키퍼를 제외했을 때 10명 미만의 선수를 갖는 국가가 존재했다.

해당 국가의 선수들을 선수 Dataset 처리 과정에서 제거했다. 마찬가지로 해당 국가들은 Ranking Dataset에서도 Data를 제거했으며, 과거 축구 경기 기록의 경우 해당 국가가 출전한 경기 기록은 제외시켰다.

C. 학습에 사용할 선수 능력치 Feature 선정

Data 전처리 후 남은 선수의 축구 능력치 관련 Feature의 개수는 총 53개이다. 이를 모두 학습에 사용하고 싶었으나 데이터 사이즈가 너무 커져 모델 학습이 어려웠다. 이에 일부 Feature만 선정해서 학습을 진행했다.

총 14개의 features를 선정했다. 다음은 각 feature의 설명과 학습 feature로 선정된 이유이다.

i. overall

선수의 종합 능력치 값이다.

선수의 종합적인 능력치는 경기에 높은 영향을 준다 판단하여 학습 feature로 선정했다.

ii. preferred_foot

선수가 주로 사용하는 발을 의미한다.

왼발 잡이 선수는 오른발 잡이 선수보다 그 수가 적기 때문에 오른발 잡이 선수에 대항할 때 큰 이점이 존재한다. 따라서 선수가 주로 사용하는 발이 어느 쪽인지는 경기 진행에 경기에 높은 영향을 준다 판단해 학습 feature로 선정했다.

iii. pace

선수의 pace 관련 능력치의 종합 능력치 값이다. 선수의 최대 속력, 가속력이 종합된 값이다.

Pace 관련 많은 능력치 값이 있었으나 모델 학습에 어려움이 발생하여 이들을 대표할 수 있는 pace feature 값을 학습 feature로 선정했다.

iv. shooting

선수의 shooting 관련 능력치의 종합 능력치 값이다. 패널티 박스 안에서의 골 결정력, 중거리 슛 골 결정력, 패널티킥 정확도, 상대 수비수의 마킹을 따돌리는 능력치, 슛 정확도, 발리슛 정확도 및 결정력 등이 종합된 값이다.

Shooting 관련 많은 능력치 값이 있었으나 모델 학습에 어려움이 발생하여 이들을 대표할 수 있는 shooting feature 값을 학습 feature로 선정했다.

v. passing

선수의 passing 관련 능력치의 종합 능력치 값이다. 크로스, 롱패스, 짧은패스 각각의 passing 안정성 및 속도, 패스의 성공률, curve passing 능력치, 프리킥 정확도 등이 종합된 값이다.

Passing 관련 많은 능력치 값이 있었으나 모델 학습에 어려움이 발생하여 이들을 대표할 수 있는 passing feature 값을 학습 feature로 선정했다.

vi. dribbling

선수의 dribbling 관련 능력치의 종합 능력치 값이다. 민첩성, 밸런스, 볼 컨트롤, 드리블, 반응속도 등이 종합된 값이다.

Dribbling 관련 많은 능력치 값이 있었으나 모델 학습에 어려움이 발생하여 이들을 대표할 수 있는 dribbling feature 값을 학습 feature로 선정했다.

vii. defending

선수의 defending 관련 능력치의 종합 능력치 값이다. 헤딩 정확도, 상대팀의 패스 차단율, 공격수 마킹 능력치, 태클 정확도, 슬라이딩 태클 정확도 및 성공률 등이 종합된 값이다.

Defending 관련 많은 능력치 값이 있었으나 모델 학습에 어려움이 발생하여 이들을 대표할 수 있는 defending feature 값을 학습 feature로 선정했다.

viii. physic

선수의 physic 관련 능력치의 종합 능력치 값이다. 플레이 공격성, 점프 높이, 스태미나, 근력 등이 종합된 값이다.

Physic 관련 많은 능력치 값이 있었으나 모델 학습에 어려움이 발생하여 이들을 대표할 수 있는 physic feature 값을 학습 feature로 선정했다.

ix. goalkeeping_diving

골키퍼 능력치의 한 종류이다. 떠 있는 공을 다이빙에서 세이브할 수 있는가를 결정하는 능력치 값이다.

골키퍼 능력치는 종합적인 값이 존재하지 않아 각각의 값을 feature로 선정했다.

x. goalkeeping_handling

골키퍼 능력치의 한 종류이다. 골키퍼가 공을 쳐내는가, 잡기를 시도하는가를 결정하는 값으로, 값이 높을수록 잡기를 선호함을 의미한다.

xi. goalkeeping_kicking

골키퍼 능력치의 한 종류이다. 골키퍼의 짧은패스, 긴패스, 시야 능력의 종합 값으로 골키퍼가 멀리 있는 선수에게 긴패스를 했을 때 공이 얼마나 안정적으로 같은 팀 선수에게 도착하는가를 결정한다.

xii. goalkeeping_positioning

골키퍼 능력치의 한 종류이다. 상대팀 선수가 슛을 성공시킬 각도를 제한하는 능력치이며 스위퍼 키퍼로써의 능력을 의미한다.

xiii. Position_Best_Rating

대표 포지션 능력치이다. 공격수일 경우 ST, LW, LF, CF, RF, RW 포지션 능력치 중 가장 높은 값이 Position_Best_Rating 값이 된다.

각 선수가 얼마나 자신의 포지션을 잘 수행할 수 있을지에 대한 값으로 여겨 경기에 큰 영향을 준다 판단했다. 이에 학습 feature로 선정했다.

D. 경기 결과와 각 Data feature 의 상관관계 분석

경기 결과 예측을 위해서 어떤 data feature가 중요도를 높게 가지는 지 분석하기 위해서 Random Forest Model 을 사용했다. Random Forest Model 은 앙상블 학습법에 해당하며, 다수의 의사결정 트리들로 구성되어 있어서 그 영향력이 줄어들게 되어 일반화에 적합하다. Random Forest 모델을 학습한 후, 여러 개의 요인들 중에서 예측력이 높은 요인을 파악하여 이후 월드컵 경기 예측 모델의 학습 데이터를 만드는데 반영했다.

```
y_es = df_label.to_numpy() # home_team_score, away_team_score
x_es = df_temp.to_numpy() # dataset for training
X_train, X_test , y_train, y_test = train_test_split(x_es, y_es, random_state=28)
```

Random Forest Model에 넣은 학습 X_train, X_test에 대한 데이터셋의 형태는 홈팀 국가명, 원정팀 국가명, 토너먼트 이름, 중립국여부, 홈팀 total points, 원정팀 away points, home 팀의 11명의 선수들에 대한 정보, away 팀의 11명의 선수에 대한 정보이다. 데이터는 학습을 위해 numeric 한 값으로 바꾸었다. 아래의 사진은 X_train, X_test에 사용한 데이터 프레임이다.

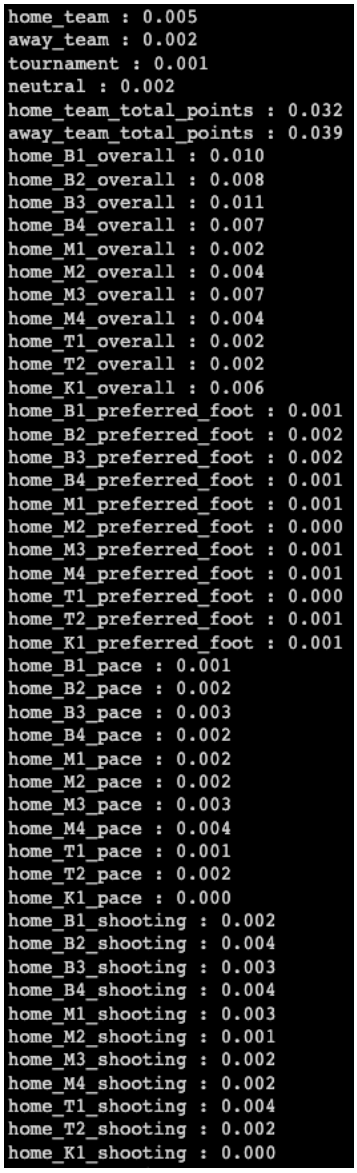
	home_team	away_team	tournament	neutral	home_team_total_points	away_team_total_points	home_B1_overall	home_B2_overall	home_B3_overall	home_B4_overall	...	a
0	1	10	91	1	664.0	710.0	71	72	70	70	...	
1	2	58	91	1	849.0	421.0	72	72	71	71	...	
2	3	1	91	0	554.0	664.0	72	72	69	70	...	
3	4	25	92	1	558.0	461.0	74	74	72	72	...	
4	3	10	91	0	576.0	701.0	72	72	69	70	...	
...	...	...	...	...	...	...	...	...	...	...	...	...
2540	85	16	92	0	1243.8	1335.36	67	66	63	59	...	
2541	82	27	92	0	961.23	1434.68	59	56	55	63	...	
2542	4	34	92	0	1473.04	1405.6	77	75	74	74	...	
2543	55	6	92	0	1425.59	1509.61	75	73	74	73	...	
2544	56	73	92	0	1384.04	1341.03	75	73	71	71	...	

2545 rows × 317 columns

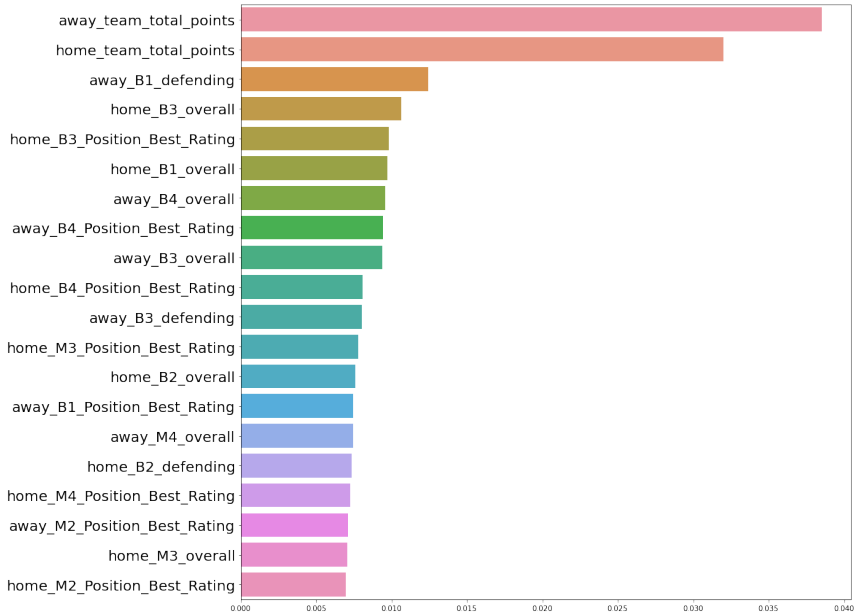
또한 y_train, y_test 에 대한 데이터셋은 홈팀 스코어와 원정팀 스코어 값이다. 아래의 사진은 y_train, y_test 에 사용한 데이터 프레임이다.

	home_score	away_score
0	2	1
1	1	0
2	1	1
3	0	3
4	1	3
...	...	...
2540	0	0
2541	0	5
2542	1	1
2543	0	0
2544	1	0

모델 학습 후 중요도 출력 예시는 아래와 같다. 'feature명 : 중요도'를 출력한 일부분이다.



총 feature가 316개여서, 한번에 확인하기 어려운 점을 개선하기 위하여 중요도가 높은 상위 20개 요인들을 추출하여 막대그래프로 시각화했다.



이에 의하면 각 팀의 ranking에 해당하는 total points가 경기 결과에 가장 큰 영향을 준다는 것을 알 수 있다.

다른 요인들은 차이가 적지만, 팀 내에서 수비 포지션과 미드필더 포지션을 하는 선수의 능력치가 중요한 것으로 나타났다.

또한 해당 포지션의 Best Rating 값이 반복적으로 나타나므로 중요한 요인이라는 것을 알 수 있다.

E. 학습 데이터

i. 최종적으로 생성한 학습 데이터의 형태는 다음과 같다.

	home_team	away_team	home_score	away_score	tournament	neutral	home_team_total_points	away_team_total_points	home_B1_overall
0	Mali	Nigeria	2	1	African Nations Championship	True	664.00	710.00	71
1	Ghana	Congo	1	0	African Nations Championship	True	849.00	421.00	72
2	South Africa	Mali	1	1	African Nations Championship	False	554.00	664.00	72
3	Norway	Poland	0	3	Friendly	True	558.00	461.00	74
4	South Africa	Nigeria	1	3	African Nations Championship	False	576.00	701.00	72
...	...	...	...	...	...	...	...	...	...
2540	Luxembourg	Bulgaria	0	0	Friendly	False	1243.80	1335.36	67

(2545, 316) 사이즈의 데이터셋으로, 2014년도부터 2022년도까지 게임의 결과와 랭킹 포인트, 출전 선수11명의 능력치 값을 합쳤다. 또한 결측값을 drop하여 결측값이 없도록 데이터를 처리했다.

ii. 테스트 데이터의 형태는 다음과 같다.

	team	total_points	B1_overall	B2_overall	B3_overall	B4_overall	M1_overall	M2_overall	M3_overall	M4_overall	...	B2_Position_Best_Rating	B3_Position_Best_Rating	B4_Position_Best_Rating	M
0	Netherlands	1694.51	85	84	82	80	87	90	82	80	...	84	83	82	
1	Senegal	1584.38	77	76	76	73	82	87	77	77	...	76	76	75	
2	Ecuador	1464.39	71	72	69	69	78	79	73	70	...	71	71	71	
3	Qatar	1439.89	71	68	69	66	72	72	71	70	...	70	69	68	
4	England	1728.47	84	85	83	84	85	87	84	86	...	85	84	84	
5	United States	1627.48	78	76	74	74	77	77	75	76	...	77	76	76	
6	Iran	1564.61	72	68	66	65	73	72	72	70	...	70	66	65	
7	Wales	1569.82	75	73	72	73	79	78	76	77	...	74	74	73	
8	Argentina	1773.88	83	81	81	81	86	85	84	84	...	83	82	81	

실제 카타르 2022 월드컵에 출전하는 32개국 나라들의 total_points, 수비수 1, 2, 3, 4의 능력치, 미드필더 1, 2, 3, 4의 능력치, 공격수 1, 2의 능력치, 골키퍼의 능력치를 처리한 데이터셋이다.

6. 학습 모델

A. XGBoost

i. 해당 모델은 Gradient Boosting을 사용한다. Gradient Boosting은 성능이 높다고 알려져있고, 특히 테이블 형식의 데이터에 대해 높은 성능을 보여준다. 다만, 높은 정확도를 위해 상당한 계산량을 요구한다. 보다 정확한 결과를 도출하기 위해 XGBoost 모델을 선택했다.

B. MultiOutputRegressor

i. 본 프로젝트는 결승에서 어떤 팀이 승리할지, 또한 각 팀이 어떤 스코어를 낼지 예측한다. 각 팀의 스코어를 예측하기 위해, 다중 출력 알고리즘을 사용하는 MultiOutputRegressor를 선택했다.

7. 결과

A. 예선전

i. 예선전 진행 방법

- 8개의 조로 이루어진 32개의 팀을 이용하여 예선전 결과 예측을 진행한다. 32개의 국가는 다음과 같이 편성된다.

  A. 네덜란드, 세네갈, 에콰도르, 카타르

  B. 잉글랜드, 미국, 이란, 웨일스

  C. 아르헨티나, 폴란드, 멕시코, 사우디아라비아

  D. 프랑스, 오스트레일리아, 튀니지, 덴마크

E. 일본, 스페인, 독일, 코스타리카,

F. 모로코, 크로아티아, 벨기에, 캐나다

G. 브라질, 스위스, 카메룬, 세르비아

H. 포르투갈, 대한민국, 우루과이, 가나

- 각 조 내에서 두 팀 씩 짝지어 경기를 진행한다. 모델이 학습한 데이터를 기반으로 경기의 승패를 예측한다. 예측 결과에 기반하여 16강에 진출할 팀을 선정한다.

ii. 32강 승패를 예측 결과

	game_id	group	team_1	team_2	label_win	predicted_win	accuracy	labe_win_team	predicted_win_team
0	0	A	Netherlands	Senegal	0	0.0	True	Netherlands	Netherlands
1	1	A	Netherlands	Ecuador	2	0.0	False	Draw	Netherlands
2	2	A	Netherlands	Qatar	0	2.0	False	Netherlands	draw
3	3	A	Senegal	Ecuador	0	0.0	True	Senegal	Senegal
4	4	A	Senegal	Qatar	0	2.0	False	Senegal	draw
5	5	A	Ecuador	Qatar	0	0.0	True	Ecuador	Ecuador
6	6	B	England	United States	2	0.0	False	Draw	England
7	7	B	England	Iran	0	0.0	True	England	England
8	8	B	England	Wales	0	0.0	True	England	England
9	9	B	United States	Iran	0	0.0	True	United States	United States
10	10	B	United States	Wales	2	0.0	False	Draw	United States
11	11	B	Iran	Wales	0	0.0	True	Iran	Iran
12	12	C	Argentina	Poland	0	0.0	True	Argentina	Argentina
13	13	C	Argentina	Mexico	0	0.0	True	Argentina	Argentina
14	14	C	Argentina	Saudi Arabia	1	0.0	False	Saudi Arabia	Argentina
15	15	C	Poland	Mexico	2	0.0	False	Draw	Poland
16	16	C	Poland	Saudi Arabia	0	0.0	True	Poland	Poland
17	17	C	Mexico	Saudi Arabia	0	0.0	True	Mexico	Mexico
18	18	D	France	Australia	0	0.0	True	France	France
19	19	D	France	Tunisia	1	0.0	False	Tunisia	France
20	20	D	France	Denmark	0	0.0	True	France	France
21	21	D	Australia	Tunisia	0	0.0	True	Australia	Australia
22	22	D	Australia	Denmark	0	2.0	False	Australia	draw
23	23	D	Tunisia	Denmark	2	0.0	False	Draw	Tunisia
24	24	E	Japan	Spain	0	2.0	False	Japan	draw
25	25	E	Japan	Germany	0	0.0	True	Japan	Japan
26	26	E	Japan	Costa Rica	1	0.0	False	Costa Rica	Japan
27	27	E	Spain	Germany	2	0.0	False	Draw	Spain
28	28	E	Spain	Costa Rica	0	0.0	True	Spain	Spain
29	29	E	Germany	Costa Rica	0	0.0	True	Germany	Germany
30	30	F	Morocco	Croatia	2	0.0	False	Draw	Morocco
31	31	F	Morocco	Belgium	0	0.0	True	Morocco	Morocco
32	32	F	Morocco	Canada	0	0.0	True	Morocco	Morocco
33	33	F	Croatia	Belgium	2	0.0	False	Draw	Croatia
34	34	F	Croatia	Canada	0	0.0	True	Croatia	Croatia
35	35	F	Belgium	Canada	0	0.0	True	Belgium	Belgium
36	36	G	Brazil	Switzerland	0	0.0	True	Brazil	Brazil
37	37	G	Brazil	Cameroon	1	0.0	False	Cameroon	Brazil
38	38	G	Brazil	Serbia	0	0.0	True	Brazil	Brazil
39	39	G	Switzerland	Cameroon	0	0.0	True	Switzerland	Switzerland
40	40	G	Switzerland	Serbia	0	0.0	True	Switzerland	Switzerland
41	41	G	Cameroon	Serbia	2	0.0	False	Draw	Cameroon
42	42	H	Portugal	South Korea	1	0.0	False	South Korea	Portugal
43	43	H	Portugal	Uruguay	0	0.0	True	Portugal	Portugal
44	44	H	Portugal	Ghana	0	0.0	True	Portugal	Portugal
45	45	H	South Korea	Uruguay	2	2.0	True	Draw	draw
46	46	H	South Korea	Ghana	1	0.0	False	Ghana	South Korea
47	47	H	Uruguay	Ghana	0	0.0	True	Uruguay	Uruguay

- 총 48번의 예선 경기에 대한 승패를 예측했다. label_win의 경우, 실제 승패 결과이다. 1은 team_1이 우승한 경우, 2는 team_2가 우승한 경우, 0는 비긴 경우이다. predicted_win 칼럼은 모델이 예측한 승패이다. 0, 1, 2의 의미는 동일하다. accuracy 칼럼의 True는 예측이 맞았을 경우, False는 예측이 틀렸을 경우를 나타낸다. 본 프로젝트에 사용된 모델은 32강에 대하여 60.42%의 정확도를 기록했다.
- 32강 예측 결과, 본선에 진출하는 팀은 다음과 같다.
  - A팀: Netherlands, Senegal
  - B팀: England, United States
  - C팀: Argentina, Poland
  - D팀: France, Australia
  - E팀: Japan, Spain
  - F팀: Morocco, Croatia
  - G팀: Brazil, Switzerland
  - H팀: Portugal, South Korea

B. 본선

i. 예측 결과

- 각각의 경기에서 모델이 경기의 승패를 예측한다. 승리한 팀이 다음 리그에 진출한다

- 16강

```
Round of 16

Netherlands vs United States
  Netherlands wins

Argentina vs Australia
  Argentina wins

Japan vs Croatia
  Japan wins

Brazil vs South Korea
  Brazil wins

Senegal vs England
  Senegal wins

Poland vs France
  Poland wins
```

- 8강

```
Quater Final Matches

Netherlands vs Argentina
  Netherlands wins

Japan vs Brazil
  Japan wins

Senegal vs Poland
  Senegal wins

Spain vs Switzerland
  Spain wins
```

- 4강

```
Semi Final Matches

Netherlands vs Japan
  Netherlands wins

Senegal vs Spain
  Senegal wins
```

- 결승

```
FIFA FINAl

[predicted_score]  3 : 2
Senegal vs Netherlands
  Senegal are the Winners


Third Place match

Spain vs Japan
  Spain Wins the 3rd Place



Winner is Senegal
Runner-up is Netherlands
3rd Place is Spain
```