

# 최종 칼럼

주제: 2026 북중미 월드컵 예측 및 대한민국의 전술 분석

목차:

1. 주제 선정 배경
2. 국가 별 전력 및 전술 분석
  - 한국 전술 분석
  - 덴마크 전술 분석
  - 남아공 전술 분석
  - 멕시코 전술 분석
3. 데이터 선정
- └ ELO Rating
4. 예측모델 구현
  - 데이터 전처리
  - 데이터 분석
  - 모델학습 전략
  - TEST DATA 예측 결과
5. 최종 모델 결과
6. 월드컵 결과예측>결론
  - 최종 결과 예측

## 서론

### 1. 주제 선정 배경

칼럼 작성 시점 기준으로 4달밖에 남지 않은 2026 월드컵은 시대를 풍미한 스타들의 마지막 월드컵으로 예상되어 기대를 모으고 있는 월드컵이다. 한국의 경우 한국 축구 최대의 스타 손흥민의 마지막 월드컵이자, 동시에 축구협회의 근 4년간 연이어 벌어진 감독 선임 논란, 그리고 홍명보 감독의 절차적/역량적 적합성이 심판에 서게 되는 대회이기도 하다. 홍명보가 부임한 후, 한국 국가대표팀은 월드컵 예선 통과를 기록함에도 불구하고, 전술적 선택에 대한 여론의 수많은 의문과 비난을 받고 있다.

이번 칼럼에서는 'ELO Rating'과 최근 특실점을 통해 XGBoost를 활용한 예측모델을 구현하고, 해당 모델을 통해 2026 월드컵 조별 예선의 행방을 예측해보고자 하였다. 과연 위기의 한국 축구 국가대표팀이 비교적 '꿀조'로 여겨지는 그룹 A에서 조별 예선을 통과할 수 있을지 예상해보고, 대면하게 될 팀들에 대한 대응방안을 전력 분석과 통계적 수치를 기반으로 제안하고자 한다.

### 2. 국가 별 전력 및 전술 분석

-한국 전술 분석

-덴마크 전술 분석

-남아공 전술 분석

-멕시코 전술 분석

## 덴마크 전술분석

감독 브리안 리머 (Brian Riemer) 2024년 10월 부임 / 2024.11.16 네이션스리그 스페인전 데뷔

주장 피에르에밀 호이비에르 (Pierre-Emile Højbjerg)

부주장 카스페르 슈마이켈 (Kasper Schmeichel) 피파랭킹 21위

최근 대회 전적 월드컵 예선 3승 2무 1패 조2위(1위 스코틀랜드) 네이션스리그 8강 진출(포르투갈에게 8강에서 패배) 유로 조 2위(3무, 1위잉글랜드) 16강 진출(독일에게 16강에서 패배) 월드컵 조 4위 1무 2패 최근 5경기 3승 1무 1패(그리스, 벨라루스, 그리스, 벨라루스, 스코틀랜드)

## 주 포메이션 및 전술

베스트 11(예상)

GK 슈마이켈

RB 라스무스 크리스텐센 CB 크리스텐센 베스테르고르/안데르센 LB 도르구

CM 호이비에르 히울만

RW 이삭센 CAM 프로홀트 LW 담스고르

ST 호일룬

리머 감독은 부임 이후 4231, 433, 3421 등 다양한 포메이션으로 경기를 스타트했다.

덴마크는 3백을 사용하는 경기에서도 전통적인 의미의 정통 3백을 고수하지는 않는다. 윙백 두 자리 중 한쪽에는 풀백 성향의 선수를, 반대쪽에는 윙어 성향의 선수를 배치하는 비대칭 구조를 활용한다. 따라서 월드컵에서 덴마크가 3백을 가동하는 시나리오라면, 센터백 리인은 크리스텐센-베스터고르-안데르센 조합으로 구성될 가능성이 높다. 또한 윙백 구성 역시 비대칭 운용이 유력한데, 최근 맨체스터 유나이티드에서 윙어로도 활용되면서 수비 시에는 5백의 일원으로 내려서는 도르구가 상대적으로 수비적인 윙백 역할을 맡고, 반대편은 보다 공격 성향이 강한 윙어 유형 자원이 기용될 가능성이 높다.

## 전술적 포인트

-강한 전방 압박

리머 감독은 부임 초부터 상당히 강력한 전방압박을 고수했다. 수비형 미드필더에 해당하는 호이비에르를 중심으로 하는 이 압박은, 골키퍼에게 가하는 압박을 시작으로 상대가 미들블록에 들어온 후에도 꾸준히 시행된다. 또 풀백까지 상당히 높은 위치로 전진시키며 압박을 중요시 여긴다. 대한민국의 최근 a매치에서 전방 압박이 약한 팀으로는 비교적 강팀임에도 좋은 결과를 얻은 반면, 강한 전방 압박을 실시했던 팀들을 상대로는 좋은 결과를 얻지 못한 점을 고려하면 이에 대한 해결책이 필요할 것으로 보인다.







#### -맨체스터 유나이티드를 조심하라

최근 연이은 공격포인트를 기록하며 좋은 품을 보여주는 도르구, 2026년 들어 침체된 모습을 보이지만 나폴리에서 주전을 따낸 호일룬, 덴마크에 정신적 지주 에릭센. 도르구와 호일룬은 대표팀에서 좋은 호흡을 보여주며 덴마크 공격의 중심이 되어주고 있으며, 에릭센은 지금은 완벽한 주전은 아니지만, 덴마크의 최근 유로 등의 대항전 돌풍의 중심이었던 만큼 조커로 투입 시 각별한 주의가 필요하다.

#### -이삭센과 크리스텐센의 호흡

최근 덴마크의 경기에서 가장 핵심이 되는 선수로 이삭센을 꼽을 수 있다. 이삭센은 오른쪽 윙어포지션에서 중앙으로 들어오며 플레이메이킹과 변수 창출을 하며 위협적인 장면을 만들고, 빈 오른쪽 측면은 크리스텐센 혹은 우측 풀백이 올라와 크로스 기회를 노린다.

#### -북유럽 축구지만 피지컬 보다는 치력으로

북유럽 팀들의 이미지는 바이킹처럼 딴딴한 몸으로 피지컬 축구를 떠올리게 한다. 물론 덴마크의 경우에도 전체적으로 작은 선수가 존재하지 않고, 평균 이상의 신장을 가진다. 하지만 크로스와 피지컬적 공격만 생각하면 오산이다. 담스가르드와 뇌르가르드를 필두로 한 미드필더 진의 중거리는 상당히 위협적이며, 센터백 안데르센의 룽킥과 후방에서 뿌리는 뒷공간 패스는 정확도가 매우 높다.

#### 대한민국의 대응 방안

##### -세컨볼 및 얼리크로스

최근 덴마크의 실점 상황을 살펴 보면 세컨볼 집중도가 높지 않은 것으로 보인다. 중거리 슈팅 혹은 세트플레이 직후 상황에서 세컨볼을 따내지 못해 실점하는 경우가 종종 있었다. 이는 대한민국이 공략해야 할 포인트가 될 수 있을 것으로 보인다. 또한 낮은 크로스 및 얼리크로스를 높은 크로스에 비해 잘 처리하지 못하는 모습을 보인다. 홍명보 감독이 조규성 대신 오현규를 기용한다면 높은 크로스의 경합에서는 상대적으로 아쉽겠지만 낮은 얼리크로스를 활용한 공격에 초점을 주는 것이 좋을 듯 하다.



### -높은 점유율 그러나 실속 없는 축구

덴마크는 강한 전방압박을 펼치면서도 점유율을 상당히 높게 가져간다. 상대적으로 약팀을 상대할 경우 60% 넘는 점유율을 보여주는 경기가 많고, 독일과 스페인 등의 강팀을 상대하는 경기에서도 45% 정도의 점유율을 가져간다. 다만 xG값과 점유율에 비해 득점으로 전환되는 비율은 그다지 높지 않다. 가장 최근 경기이자 월드컵 직행에 실패하게 된 스코틀랜드와의 경기에서 70%의 점유율과 2.20의 xG값을 기록했음에도 4대2로 패배하였다. (스코틀랜드 xG 1.76) 대한민국의 전술 컨셉이 역습일 경우, 볼 탈취 후 강한 전방압박을 한 번만 끓어낸다면 손흥민과 황희찬 등을 필두로 매우 좋은 기회를 잡을 수 있을 것이다.

### 한국이 어떤경기를 참고해야할까?

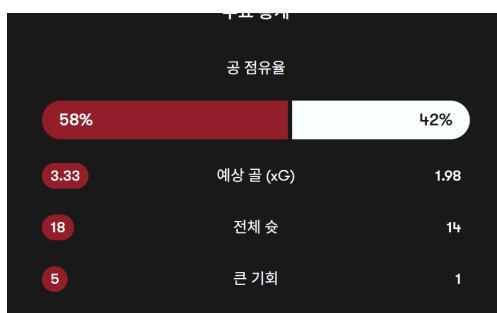
#### 스코틀랜드에게 4대2로 패배한 경기



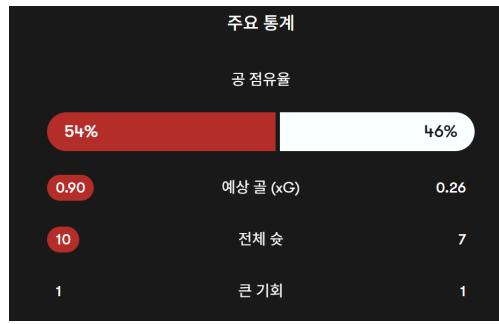
#### 벨라루스에게 2대2로 무승부를 거둔 경기



#### 포르투칼전 5대2 패배한 경기



#### 세르비아전 2대0으로 승리한 경기



아래 장면은 강한 압박 이후 상대가 직선적이 전방패스로 한 번에 뚫어내자 후방에 적은 수만 남으며 치명적인 역습을 당하는 장면이다.



한국은 이와 같이 높은 압박 라인에 잘 대응해서 뒷공간을 노릴 수 있어야겠다.

### 남아프리카 공화국 전술 분석

'Bafana Bafana!'라는 당찬 구호로 그들의 국가대표팀을 호명하는 남아프리카 공화국은 2010 월드컵을 주최해보기도 한 축구에 진심인 국가 중 하나이며, 그들에게 축구는 국가 내에서 제일 인기있는 스포츠 중 하나이다. 남아프리카 공화국의 축구 리그는 과거 카시 플라바(Kasi Flava)로도 불리던 개인기/서커스 같은 축구에서 벗어나 유럽의 선진적인 전술과 선수 육성, 아프리카의 신체적 우월성을 기반으로 성장하고 있으며, 국가대표팀 선수들의 대부분을 배출하고 있다. 이 파트에선 남아프리카 공화국의 전력 및 전술 분석과 함께 한국이 어떤 경기를 참고하여 대응법을 가져가야 할지 알아보고자 한다.

: 팀 소개.

감독: '휴고 브루스(Hugo Broos)'는 벨기에의 전설적인 선수 출신 인물이다. 과거의 잊혀지는 감독이 되어가는 듯 했으나 2017 카메룬의 AFCON 기적적인 우승을 이끈 이후로 부활, 이후 2021년부터 남아공 대표팀 감독으로 재임중이다. 2023년 남아공의 AFCON 4강 신화를 이끌며 남아프리카 공화국 대표팀을 성공적으로 이끄는 중이다.

주장: 론웬 윌리엄스(Ronwen Williams)는 뛰어난 승부차기 선방 능력과 발밑을 가진 현대적인 골키퍼로, 무엇보다 무서운 것은 그의 연륜에서 오는 노련함이다.

피파랭킹 60위

최근 대회 전적: (월드컵 예선: 5승 3무 2패 / 2025 AFCON 16강 진출 (카메룬에게 패배))

최근 전적: 5경기 3승 2패 (가나, 앙골라, 이집트, 짐바브웨, 카메룬)

: 전력 분석

BEST 11:

이름	포지션	출전수	나이	비고	주발	시장가치
론웬 윌리엄스 (Ronwen Williams)★	GK	56	34	주장, 승부차기/뛰어난 발밑	오른발	9억
오브리 모디바 (Aubrey Modiba)	LB	43	30	공격 시 영리한 공간 이해	왼발	12억
음베케젤리 음보 카지(Mbekezeli Mbokazi)	LCB	6	20	05년생, 남아공의 유망주이자 MLS 시카고 파이어 소속, 원발잡이 센터백.	왼발	6억
시아봉가 은게자 (Siyabonga Ngezana)	RCB	11	28	191cm의 거구, 남아공 리그 정상급 센터백	오른발	34억
쿨리소 무다우 (Khuliso Mudauf)	RB	31	30	남아공 불박이 유풀백, 센터백 멀티 자원	오른발	7억
데보호 모코에나 (Teboho Mokoena)★	LDM	45	29	불박이 수미, 가장 많은 패스와 기회 창출. 프리킥 능력	오른발	21억
페펠로 시틀 (Sphephelo Sithole)	RDM	25	26		오른발	6억
オス윈 아폴리스 (Oswin Appolis)	LW	26	24	아약스 유소년 출신, 현품 최상, 인버티드 윙어 기질	오른발	7억
시포 음블레 (Sipho Mbule)	CAM	12	27		오른발	5억
렐레보힐 모포켕 (Relebohile Mofokeng)	RW	11	21	남아공 리그의 유망주, 원래 포지션은 공미나 LW이나, 팀 여건상 RW에서 뛰는 것으로. 키가 작음 (170cm)	오른발	19억
라이 포스터(Lyle Foster)★	ST	27	25	번리 소속, PL 경험. 스피드 장점의 스트라이커이나, 소속팀에선 톱으로서의 강점을 보	오른발	170억

여주지 못해 윙어로 기용되는 경우 늘고 있음. 소속팀에서와 달리 국대 예선 불박이 톱으로 월드컵 본선 진출에 기여하기도 함.

추가적으로 주의해야 할 선수: 강팀과의 경기에 강한 서브 공격수 에비던스 막고파와, 최근 벨기에 클럽 브뤼헤에서 출전 수를 늘리고 있는 산드레 캠벨 등이 있다.

#### 스쿼드 구성:

과거 남아공의 국가대표 선수단의 다수는 마멜로디 선다운즈(Sundowns)라는 팀 소속이 대다수였다. 그러나 현재 휴고 브루스 체제에 들어서는 그 중심이 올랜도 파이어츠(Orlando Pirates) 위주로 재편되고 있다. 현재 선수단 전체 인원 중 파이어츠 소속이 무려 9명이기 때문이다. 같은 팀에서 뛰는 선수들에게서 뽑아져나오는 남다른 팀워크는 남아공의 핵심적인 전력이 된다.

또 다른 스쿼드 상의 특이점으로는 휴고 브루스 감독 체제 하에서의 센터백 조합 변경이다. 이번 AFCON에서 새로운 센터백 조합(음보카지와 은게자)을 시험했으며, 이 둘의 시너지는 성공적으로 보인다. 원발 센터백 유망주로서 포텐을 터트리고 있는 음보카지와, 압도적인 피지컬을 지닌 은게자의 조합은 남아프리카 공화국 빌드업의 초석이 되며, 그들의 듼든한 최후의 보루로서 기능하고 있다.

#### 남아공 전술 분석

##### -어떤 포메이션을 사용할까? 433 (공격 시 2-3-4-1)

남아공의 주력 포메이션은 4231이 분명하다. 수비의 경우 베스트11의 포백을 그대로 구성하고, 중원은 보통 더블 피봇 위에 공격형 미드필더를 운용한다. 그러나 가장 최근 경기인 AFCON 카메룬전과 같이 강팀을 상대할 땐 수비형 미들 자리에 센터백 출신인 Sibisi를 앵커로서 기용하여, 원볼란치를 기반으로 구성하는 3미들 체제를 사용함을 알 수 있다. 따라서 상대적 강팀인 한국과의 경기에서도 433으로 안정성을 택할 가능성 또한 크다고 볼 수 있다.

##### -전진하는 풀백을 조심하라구?

그러나 결국 남아공의 핵심은 4백을 통한 빌드업이다. 지공 상황에서 양쪽 풀백을 상당히 공격적으로 전진 배치한다. 양 윙어들은 중앙지향적 성향이 강하기 때문에, 풀백이 올라갈 때 윙어들은 하프스페이스 혹은 수비수들 사이로 이동한다. 측면에서 미들 1명, 풀백 1명, 윙어 1명이 유기적으로 필요한 상황에서 움직일 수 있으며, 이는 패스 측면에서 유리한 삼각형을 수비 주위에 구성하며, 순간적으로 상대보다 수적 우위를 가져갈 수 있다. 볼을 소유하다 선택지가 생기면 직선적인 패스로 공을 투입하고 빠르게 마무리하는 양상을 보인다.

##### -점유율 기반의 4백 운용과 플레이메이커 모코에나

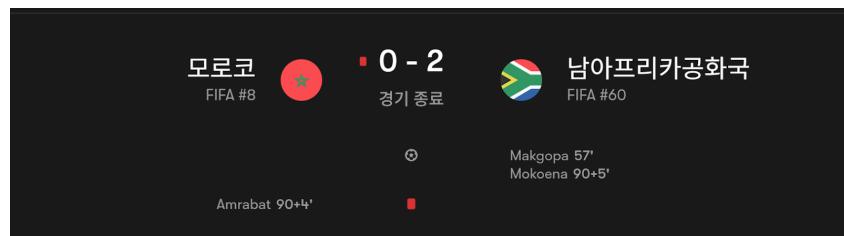
그리고 그러한 패스의 뿌리를 찾아가면 결국 데보호 모코에나이다. 아프리카 국가들 특유의 피지컬을 살린 역습 축구만을 보게 될 것이라 예상하는 것은 큰 오산이다. 최근 AFCON 경기들 통계를 보면 전부 **58%이상의 점유율**을 기록했다. 또한 모든 경기에서 슈팅 숫자가 대등하거나 더 많았음. 이는 새롭게 주전으로 들인 센터백 두명을 포함한 4백의 빌드업이 매우 안정적이고 성공적이었음을 의미하며, 테보호 모코에나라는 중원의 핵심을 딱 두고 경기를 운영하는 전술을 사용하기 때문이다. 모코에나는 마지막 경기였던 카메룬전을 제외하면 모든 경기에서 최다 패스를 기록했다. 점유율 유지는 물론 기회 창출을 위한 기술적 역량이 충분한 선수로 보인다.

최근 전적에서 특히 집중했을 때, 승리한 앙골라전과 짐바브웨전에선 xG값 대비 더 많은 골을 뽑아냈지만, 패배한 경기의 경우 이집트전 xG 0.78, 카메룬전 xG 2.08임에도 불구하고 1골밖에 넣지 못했다. 자신들보다 기술적, 신체적 역량이 강한 팀이 점유율을 신경쓰지 않고 역습을 노릴 때 취약한 모습을 보인다는 점에서 한국과 유사하다.

#### 한국은 어떤 경기를 참고해야 할까?

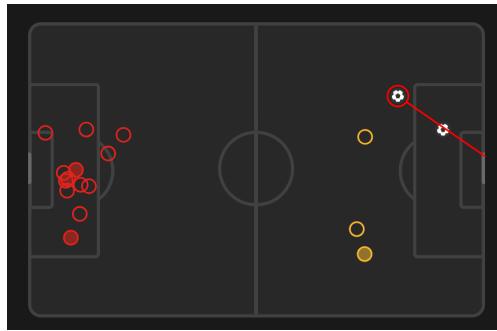
남아공의 근 2년간의 전적들은 대부분 월드컵 수준이 아닌 국가들과의 경기가 대다수였다. 또는 아프리카의 스타일, 즉 점유율 추구보단 직선적인 공격을 시도하는 팀들과의 경기가 많았다. 남아공이 그나마 한국과 유사한 팀, 즉 전력이 더 강하면서 점유율을 더 쟁기는 팀을 상대한 사례를 시간을 조금 더 거슬러 찾아볼 수 있었다.

2024년 1월 31일 펼쳐진 AFCON 2023의 16강전. 남아공은 모로코와 경기를 치르게 되었다. 당시 피파랭킹 8위 VS 60위의 경기이며, 카타르 월드컵에서 4강까지 도달했던 모로코의 열풍이 채 가시기 전이었기에, 모두가 모로코의 승리를 예상했다.



그러나 예상을 깨고 남아공이 승리를 기록했다. 모로코가 야신부누, 마즈라위, 로맹 사이스, 하키미, 암라바트, 엔니시리, 아들리 등 유럽 출신 멤버들을 내세웠지만, 남아공은 자국 리그 출신들로 끌어들 웅쳤다. 모로코가 13번의 슛팅 중 겨우 2번만을 유효슈팅으로 만들어냈으며, 그 두번 마저 키피의 선방으로 인해 막혔다. 반면 남아공은 직선적인 패스로 첫 골을 만들고, 모코에나의 직접 프리킥으로 빼기골을 박았다.





두 팀 다 의미없는 공돌리기로 패스수를 늘리기 보단ダイレクト하게 볼을 투입했던 것으로 보인다. 비교적 적은 총 패스수가 이를 의미하며, 남아공의 우주 방어를 수비 지표에서 확인할 수 있었다.

패스		수비	
425	패스	386	9
372 (88%)	정확한 패스	335 (87%)	9
209	내 팀 진영	213	0
163	상대 팀 진영	122	4
21 (57%)	정확한 긴 패스	21 (47%)	1
7 (23%)	정확한 크로스	0 (0%)	
14	던지기	10	
28	상대편 박스 내에서의 터치	9	44
4	오프사이드	4	31 (56%)
반칙		8 (57%)	24 (44%)
		13 (65%)	7 (35%)
		볼 경합	
		볼 경합 성공	
		31	
		지상 볼 경합 성공	24 (44%)
		공중 볼 경합 성공	7 (35%)
		드리블 성공	3 (43%)

결국 남아공은 이 때 최종 3위를 차지하는 기염을 토했다. 비록 올해는 16강 탈락이라는 고배를 마셨으나. 이 경기를 참고할 수 있는 이유는 라인업의 경우 공격진과 센터백 라인이 절반정도 물갈이되긴 했으나, 감독이 같으며 비슷한 전술적 접근을 가져갈 수 있는 환경이기 때문이다.

### 대한민국의 대응 방안

#### -상대의 좌측 수비 공략. (우측의 이강인, 설영우 활용)

: 좌측 풀백 모디바는 노련하지만 전술 움직임으로 인해 높이 전진하기 때문에 뒷공간이 필연적으로 발생한다. 또한 왼쪽 센터백 음보카지는 아직 유망주인 만큼 한국의 공격수들이 더 노련하게 그를 상대할 수 있을 것으로 보인다. 180cm이란 비교적 작은 신장도 공략 대상. 반면 우측은 센터백도 볼 수 있는 수비력의 무다우와, 좀 더 노련하고 191의 거구인 은제자나가 존재하기 때문에, 비교적 단단한 수비력을 자랑한다.

남아공 좌측의 수비 약세를 한국이 더 잘 활용할 수 있는 이유는 이강인과 설영우의 존재다. 이 둘로 구성된 한국의 우측면은 이재성(또는 황희찬)과 이명재로 구성될 좌측보다 돌파력과 기회 창출 측면에서 우세하다. 둘의 연계 또는 이강인의 돌파로 좌측 수비를 파훼하면 한국은 쉽게 공격 활로를 찾을 수 있다.

#### -모코에나를 압박, 상대의 측면 삼각 플레이 차단

: 모코에나는 남아공 빌드업의 핵심이고, 그에게 공간을 줄 시 위협적인 패스가 투입될 수 있다. 경합력 좋은 이재성을 필두로 2선이 모코에나를 압박하고, 윙어들과 수비들은 상대의 전진한 풀백과 윙어의 스위칭을 방해하고 착실하게 마킹해 모코에나의 선택지를 없애야 한다. 남아공은 지금 상황에서의 갑작스런 전진패스 투입에 상당한 강점을 가졌기 때문에, 이에 대해 무방비하다면 1승 제물은 한국이 될 수도 있을 것이다.

#### -빠른 선제골

: 모로코는 강팀이었으나 남아공의 짠물 수비에 고전하다가 중요한 찬스를 허용한 바 있다. 아직 그때 당시의 수비 및 미들진을 남아공은 아직 보유하고 있다. 한국의 경우도 빠른 선제골이 나오지 않으면 경기력이 상당히 답답한 모습을 많이 보였다. 따라서 한

국의 빠른 선제골이 절실하다. 그렇지 않을 시 남아공은 단단한 4백과 3미들로 내려앉아, 기회를 노릴 것이다.

#### -라일 포스터의 속력 견제.

: 라일 포스터는 프리미어리그에서 증명하지 못한 스트라이커이지만, 그럼에도 불구하고 속력에서는 분명한 강점을 가지는 스트라이커이다. 김민재의 빠른 커버 능력으로 그를 견제해야 할 필요성이 있다.

#### 멕시코 전술 분석

감독 하비에르 아가레 (Javier AGUIRRE) 2024년 7월 부임 / 멕시코 대표팀에서 세 번째로 감독을 맡고 있음. 라리가, 국대 감독 경험 다수

주장 에드손 알바레스 (Edson ÁLVAREZ)

피파랭킹 16위

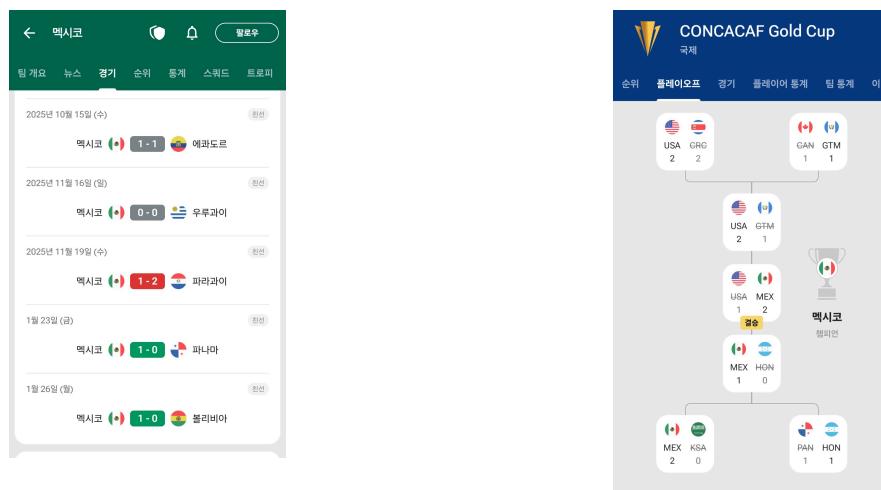
#### 최근 대회 전적

-2025 CONCACAF 골드컵 우승

-2024 코파아메리카 조별 3위 (1승 1무 1패) 탈락

-2022 카타르 월드컵 조별 3위 (1승 1무 1패) 탈락

최근 5경기 2승 2무 1패(볼리비아, 파나마, 파라과이, 우루과이, 에콰도르)



이름	소속 클럽
루이스 말라	클럽 아메리카
라울 랭헬	과달라하라
카를로스 아세베도	산토스 라구나
요한 바스케스	제노아
세사르 몬테스	로코모티프 모스크바
헤수스 가야르도	톨루카
빅토르 구스만	몬테레이
이스라엘 레예스	클럽 아메리카
마테오 차베스	AZ 알크마르
에드손 알바레스	페네르바체
루이스 차베스	디나모 모스크바

오르밸린 피네다	AEK 아테네
마르셀 루이스	톨루카
길베르토 모라	티후아나
오베드 바르가스	시애틀 사운더스
산티아고 히메네스	AC 밀란
라울 히메네스	풀럼
이르빙 로사노	샌디에이고 FC
헤르만 베르테라메	인터 마이애미
세사르 후에르타	안더레흐트

예상 BEST 11

라울 히메네스

알렉시스 베가

이르빙 로사노

루이스 차베스      마르셀 루이스

에드손 알바레스(C)

예수스 가야르도

이스라엘 레예스

요한 바스케스      세사르 몬테스

루이스 말라곤

최근 멕시코는 2026 월드컵을 앞두고 전통적인 4-3-3을 기본으로 하되, 경기 상황에 따라 변형 백쓰리를 혼용하는 유연한 전술을 보여주고 있다.

멕시코는 최근 하비에르 아기레 감독 체제에서 전술적 유연함을 강조하고 있다. 특히 수비 라인에서 정형화된 4백이나 3백에 얹매이지 않고, 양 측면 자원의 서로 다른 성향을 극대화하는 비대칭 구조를 활용한다.

만약 멕시코가 월드컵 본선에서 수비 안정감을 위해 3백(또는 5백)을 가동하는 시나리오라면, 센터백 라인은 요한 바스케스 – 세사르 몬테스 – 이스라엘 레예스 조합으로 구성될 가능성이 매우 높다. 이 조합에서 핵심은 양쪽 측면 자원의 역할 분담이다.

-왼쪽의 공격적 전진

왼쪽 윙백(혹은 풀백)으로 나서는 예수스 가야르도는 사실상 윙어에 가까운 높은 위치까지 전진하며 공격의 폭을 넓힌다. 이는 가야르도의 풍부한 경험과 공격 기담 능력을 활용해 상대 측면을 허물기 위함이다.

-오른쪽의 수비적 밸런스

반대편인 오른쪽에서는 이스라엘 레예스가 배치된다. 레예스는 본래 센터백까지 소화 가능한 자원으로, 왼쪽의 가야르도가 전진할 때 중앙으로 좁혀 들어오며 실질적인 '스토퍼' 역할을 수행한다. 이를 통해 수비 시에는 견고한 3백 혹은 5백의 형태를 갖추고, 공격 시에는 안정적인 후방 빌드업 숫자를 확보한다.

-중원의 에너자이저

이 비대칭 구조가 원활하게 돌아가기 위해서는 수비형 미드필더 에드손 알바레스의 역할이 절대적이다. 그는 포백 보호는 물론, 비대칭적으로 변하는 수비 라인의 빈 공간을 메우는 '포지셔닝의 핵심' 역할을 맡는다.

### 전술적 포인트

#### -신뢰의 트라이앵글

아기레 감독은 부임 후 후방의 안정감을 최우선으로 하며, 에드손 알바레스 – 요한 바스케스 – 세사르 몬테스로 이어지는 이른바 삼각형을 전술을 구축했다. 중원의 핵인 알바레스는 수비 시 센터백 사이로 내려가 변형 3백을 만들며, 상대의 전방 압박을 무력화하는 기점 역할을 한다. 최근 튜르키예 페네르바체로 팀을 옮긴 알바레스의 컨디션이 멕시코 후방 빌드업의 질을 결정짓는 핵심 요소다.

#### -비대칭 풀백 운용

멕시코는 4백을 기반으로 하되 공격 시에는 철저한 비대칭 구조를 활용한다. 왼쪽의 헤수스 가야르도는 윙어에 가까울 정도로 높게 전진하여 측면 수적 우위를 만드는 반면, 오른쪽의 이스라엘 레예스는 중앙으로 좁혀 들어와 사실상 세 번째 센터백처럼 움직인다. 이는 공격의 파괴력을 높이면서도 역습 시 숫자가 부족해지는 것을 방지하는 아기레 감독의 실용주의적 선택이다.

#### -17세 신성 길베르토 모라

현재 멕시코 축구계에서 가장 뜨거운 이름은 티후아나 소속의 길베르토 모라다. 2008년생인 그는 최근 A매치에서 라울 히메네스의 결승골을 득는 등 좋은 활약을 펼치고 있다. 여기에 프리미어리거 라울 히메네스의 노련한 포스트 플레이와 산티아고 히메네스의 결정력이 더해진 공격진은 위협적이다.

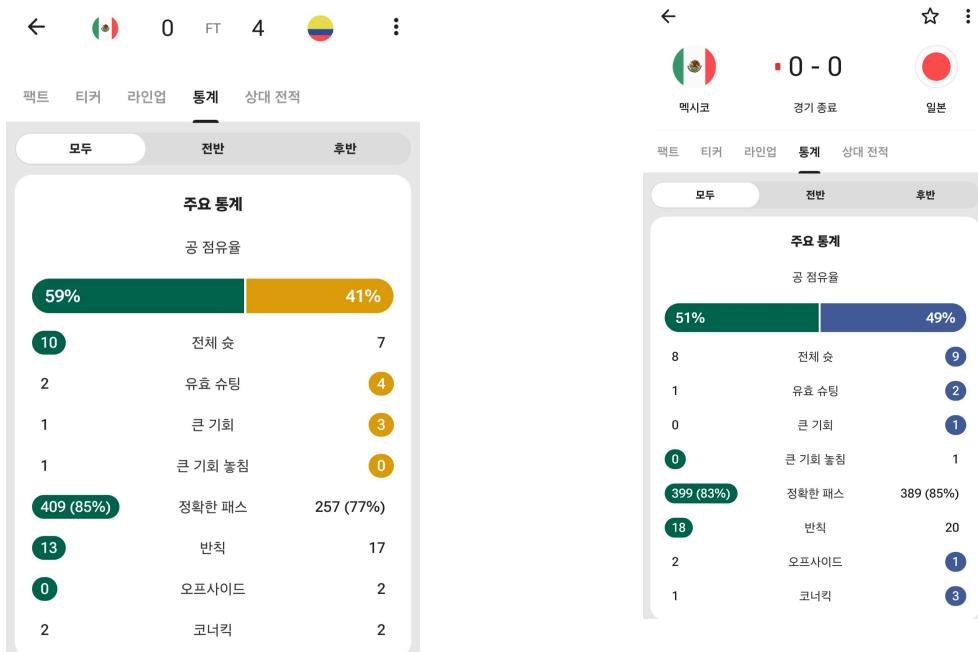
#### -기동성의 축구

과거 멕시코가 짧은 패스 위주의 아기자기한 축구를 했다면, 2026년의 멕시코는 훨씬 직선적이고 강한 압박을 병행한다. 특히 루이스 차베스와 마르셀 루이스가 버티는 미드필더진은 강력한 중거리 슛 능력을 갖추고 있어, 상대 수비가 내려앉을 때 변수를 만들 수 있다. 또한 센터백 요한 바스케스의 원발 롱킥은 단번에 전방의 이르링 로사노에게 기회를 제공하는 위협적인 라인이다.

### 대한민국의 대응 방안

#### -높은 라인 배후 공간과 전환 시의 수비 정체성 상실

2025년 10월 12일 콜롬비아전에서 멕시코는 59%의 점유율을 기록하고도 대패했다. 가장 큰 원인은 공격 시 과도하게 라인을 올렸다가 볼을 탈취당했을 때 벌어지는 수비수 사이의 공간이었다. 특히 세사르 몬테스와 요한 바스케스 등 중앙 수비진은 피지컬은 뛰어나지만 뒷공간으로 침투하는 빠른 공격수를 추격하는 속도에서 약점을 드러냈다. 대한민국이 볼 탈취 후 손흥민, 황희찬, 배준호 등을 활용해 직선적인 역습을 가한다면 위협적일 것이다.



-점유율만 높은 '실속 없는 축구' 유도

2025년 9월 7일 일본전에서 멕시코는 경기를 주도했음에도 유효 슈팅을 단 1개밖에 기록하지 못할 정도로 답답한 공격력을 보여줬다. 아기레 감독의 멕시코는 점유율은 높게 가져가지만, 상대가 내려앉아 조직적인 미들 블록을 형성할 경우 이를 정교하게 뚫어내는 세부 전술이 부족하다. 대한민국이 수비 간격을 좁히고 중원을 밀집시켜 백쓰리 전술을 꺼내들어 실수를 줄인다면, 멕시코는 xG(기대 득점) 값만 높을 뿐 득점하지 못하는 비효율적인 경기를 반복할 가능성이 크다.

## 경계해야 할 부분

### 1. 실용적 축구와 거친 압박

최근 파나마전과 볼리비아전에서 나타난 멕시코의 특징은 매우 거칠고 실용적이라는 점이다. 과거의 기술 축구가 아닌, 중원의 에드손 알바레스를 중심으로 상대 미드필더를 강하게 압박하며 흐름을 끊는 전술을 구사한다. 파나마전에서 경기 내내 파울로 상대를 괴롭히다 경기 막판 결승골로 승리한 것이 전형적인 사례다. 우리 미드필더진이 멕시코의 거친 압박에 당황해 후방에서 실책을 범한다면 곧바로 위기를 맞이하게 된다.

### 2. 길베르토 모라

최근 멕시코 축구의 희망으로 떠오른 길베르토 모라는 경계대상 1호다. 좁은 공간에서 상대를 허무는 턴 동작과 박스 안으로 찔러 넣는 칼패스는 멕시코 공격진의 단순함을 해결해 주는 유일한 열쇠다. 비록 최근 부상 변수가 있으나, 그가 조커로 투입될 경우 대한민국 수비진은 그에게 쏠리는 시선을 경계하고 배후에서 침투하는 라울 히메네스나 이르빙 로사노를 놓치지 말아야 한다.

### 3. 세트피스

멕시코는 필드골이 터지지 않을 때 루이스 차베스의 원발 킥에 의존한다. 차베스는 직접 프리킥은 물론 코너킥 상황에서 좋은 킥을 보이며, 신체 조건이 좋은 몬테스나 바스케스의 헤더로 이어진다. 박스 인근에서 불필요한 파울을 내주는 것을 조심해야 한다.

### 4. 멕시코 홈경기

멕시코의 홈에서 열리는 경기이자, 경기가 열리는 구장의 고도가 엄청 높기 때문에 한국 선수들의 적응 문제도 중요하다. 우리나라 멕시코와 월드컵에서 두 번 만나 두 번을 졌다. 그리고, 아시아팀은 5전 5패를 기록하고 있다. 멕시코와의 경기는 항상 어렵게 져왔다. 한국은 실수를 줄이고 특유의 붉은 악마 정신의 기동력과 역습 한두방을 노릴 필요가 있고, 최근 손흥민의 프리킥 또한 중요시된다.

## 결론

멕시코는 강한 압박과 높은 점유율을 구사하지만, 뒷공간 수비와 결정력에 확실한 문제를 안고 있다. 대한민국은 멕시코의 압박을 빠르게 벗겨내고 역습을 노리는 것이 승리 확률을 높이는 가장 확실한 길이다.

## II. 본론

## 3. 데이터 선정

### ↳ ELO Rating

<https://www.eloratings.net/>에서 카타르 월드컵 이후, 북중미 월드컵 진출국들의 A매치 기록 및 Elo rating을 수집하였다.

### ↳ 피파랭킹 및 국제경기 데이터

<https://www.kaggle.com>에서 현재 피파 랭킹 및 국제경기 데이터(득실 기록/홈 어웨이 등 기록 포함)를 수집하였다.

## 4. 예측모델 구현

### 4-1. 데이터 전처리

먼저 데이터를 수집한 파일은 다음과 같다.

월드컵통합.xlsx

월드컵 예측 모델을 만드는 것이 목표이기 때문에, 월드컵 국가들 간의 경기들을 학습 데이터로 사용하였고, 최근 경기 득실점 계산 시에는 월드컵 비진출국 간의 경기를 포함하였다.

데이터 전처리는 다음과 같은 과정으로 진행하였다.

```
library(tidyverse)
library(readxl)
library(zoo)
library(lubridate)
library(writexl)

# 1. 48개국 리스트
wc_countries <- c(
  "Denmark", "Mexico", "South Korea", "South Africa",
  "Canada", "Italy", "Switzerland", "Qatar",
  "Brazil", "Morocco", "Haiti", "Scotland",
  "United States", "Paraguay", "Australia", "Turkey",
  "Germany", "Curaçao", "Ivory Coast", "Ecuador",
  "Netherlands", "Japan", "Ukraine", "Tunisia",
  "Belgium", "Egypt", "Iran", "New Zealand",
  "Spain", "Cape Verde", "Saudi Arabia", "Uruguay",
  "France", "Senegal", "Iraq", "Norway",
  "Argentina", "Algeria", "Austria", "Jordan",
  "Portugal", "Democratic Republic of Congo", "Uzbekistan",
  "Colombia", "England", "Croatia", "Ghana", "Panama"
)

# 2. 데이터 불러오기
file_path <- "C:/Users/LEEBG/Desktop/월드컵통합.xlsx"

raw_data <- read_excel(file_path, col_types = "text") %>%
  distinct(date, home_team, away_team, .keep_all = TRUE) %>%
  mutate(
    date_num = suppressWarnings(as.numeric(date)),
    date_final = case_when(
      !is.na(date_num) ~ as.Date(date_num, origin = "1899-12-30"),
      TRUE ~ ymd(date)
    ),
    # 수치형은 수치형으로, 논리형은 논리형으로
    home_score = as.numeric(home_score),
    away_score = as.numeric(away_score),
    elo_home_rating = as.numeric(elo_home_rating),
    elo_away_rating = as.numeric(elo_away_rating),
    neutral = as.logical(neutral)
  ) %>%
  # 날짜 변화
  mutate(date = date_final) %>%
  # 결측치 제거
  drop_na(date, home_score, away_score)

# 총 데이터 개수
print(nrow)
```

```

# 3. 모든 경기를 활용한 '최근 5경기' 변수 계산

# 전체 경기 Double Entry
all_matches <- bind_rows(
  # Home 관점
  raw_data %>%
    select(date, Team = home_team, Opponent = away_team,
           GF = home_score, GA = away_score,
           Elo = elo_home_rating, Elo_Opp = elo_away_rating, neutral) %>%
    mutate(Venue = ifelse(neutral == TRUE, 2, 0)),

  # Away 관점
  raw_data %>%
    select(date, Team = away_team, Opponent = home_team,
           GF = away_score, GA = home_score,
           Elo = elo_away_rating, Elo_Opp = elo_home_rating, neutral) %>%
    mutate(Venue = ifelse(neutral == TRUE, 2, 1))
) %>%
  arrange(Team, date) # 팀, 날짜 순 정렬

# 최근 날짜 계산
stats_all <- all_matches %>%
  group_by(Team) %>%
  mutate(
    # 직전 5경기 평균
    Recent_GF_5 = rollmean(lag(GF, 1), k = 5, fill = NA, align = "right"),
    Recent_GA_5 = rollmean(lag(GA, 1), k = 5, fill = NA, align = "right")
  ) %>%
  ungroup()

print(nrow(stats_all))

# 4. 필터링 -> 월드컵 진출국끼리 간의 경기만

target_matches <- stats_all %>%
  filter(Team %in% wc_countries & Opponent %in% wc_countries) %>%
  drop_na(Recent_GF_5, Recent_GA_5) # 5경기 데이터 없는 초반 경기 제거

# 5. 상대 전적 불이기
final_dataset <- target_matches %>%
  inner_join(target_matches,
             by = c("date" = "date", "Team" = "Opponent"),
             suffix = c("", "_Opp")) %>%
  mutate(
    Result_Class = case_when(
      GF > GA ~ 0, # 승
      GF == GA ~ 1, # 무
      GF < GA ~ 2 # 패
    )
  ) %>%
  select(
    date, Team, Opponent, Venue,
    Elo, Elo_Opp,

```

```

Recent_GF_5, Recent_GA_5,
Recent_GF_5_Opp, Recent_GA_5_Opp,
Result_Class
) %>%
distinct(date, Team, Opponent, .keep_all = TRUE) # 최종 종복 제거

# 결과 확인
print(nrow(final_dataset))
head(final_dataset)

```

최종적으로 만들어진 변수들은 다음과 같다.

Team: 팀 A

Opponent: 팀 B

Venue:

- 0: A팀의 홈경기
- 1: B팀의 홈경기
- 2: 중립 구장

Elo: A팀의 Elo(경기 시점 기준)

Elo: B팀의 Elo(경기 시점 기준)

Recent\_GF\_5: A팀의 최근 5경기 득점

Recent\_GA\_5: A팀의 최근 5경기 실점

Recent\_GF\_5\_Opp: B팀의 최근 5경기 득점

Recent\_GA\_5\_Opp: B팀의 최근 5경기 실점

Result\_Class(타겟 변수):

- 0: A팀 승리
- 1: 무승부
- 2: B팀 승리

총 관측치의 개수는 650개이며, 이는 실제 325개의 데이터를 대칭적으로 하여 구성하였다. 이러한 데이터 증강 및 구조화 방식을 택한 이유는 다음과 같다.

### 1. 데이터 부족 문제 해결

월드컵 및 A매치 데이터는 그 특성 상 샘플 수가 제한적이다. 325개의 데이터만으로는 머신러닝 모델 패턴을 충분히 학습하기에 부족하여 과소 적합이 발생할 우려가 있다. 이를 해결하기 위해 A vs B 경기와 B vs A 경기를 별개의 관측치로 취급하여 학습 데이터의 양을 2배로 확보하였다.

### 2. 위치 편향 제거

머신러닝 모델은 입력 변수의 순서에 영향을 받을 수 있다. 예를 들어, 학습 데이터에 Team: 강팀 vs Opponent: 약팀 순서만 존재한다면, 모델은 "Team 변수에 들어온 국가가 무조건 이긴다"는 잘못된 편향을 학습할 수 있다.

따라서 모든 경기에 대해 관점 전환을 적용하여, Team과 Opponent의 위치를 바꾼 데이터를 모두 학습시켰다. 이를 통해 모델이 특정 위치가 아닌, 두 팀 간의 '전력 차이'에 집중하여 승패를 판단하도록 유도하였다.

### 3. 타겟 변수의 대칭성 확보

입력 데이터의 관점이 바뀌면 결과 또한 논리적으로 반전되어야 한다.

입력과 출력이 동시에 대칭적으로 변환된 데이터를 학습함으로써, 모델의 일반화 성능을 강화하였다.

## 4-2 모델 학습 전략

본 프로젝트에서는 월드컵 경기 결과를 예측하기 위해 XGBoost (Extreme Gradient Boosting) 알고리즘을 최종 모델로 선정하였다. 최적의 모델을 선정하기 위해 SVM(Support Vector Machine)과 Random Forest를 베이스라인 모델로 설정하여 성능을 비교 검증하였으며, 그 결과 XGBoost가 가장 우수한 예측 성능을 보였다.

### -활용 모델: XGBoost

XGBoost는 여러 개의 약한 예측 모형, 즉 의사결정 나무를 순차적으로 결합하여 강력한 예측 성능을 내는 부스팅 양상을 기법이다.

- 핵심 원리: 이전 단계의 트리가 예측하지 못한 오차에 가중치를 두어 다음 트리가 이를 학습하게 함으로써, 점진적으로 오차를 줄여나가는 방식(Gradient Descent)을 사용한다.
- 특징: 기존의 Gradient Boosting Machine(GBM) 대비 병렬 처리를 지원하여 학습 속도가 빠르며, 자체적인 규제 기능을 포함하고 있어 과적합을 방지하는 데 탁월하다.

### -모델 선정 이유

이번 월드컵 예측 프로젝트에서 XGBoost를 채택한 주된 이유는 다음과 같다.

- 정형 데이터(Tabular Data)에서의 압도적 성능:  
승패 예측과 같은 정형 데이터 분류 문제에서 XGBoost는 딥러닝이나 다른 머신러닝 알고리즘 대비 SOTA(State-of-the-art) 수준의 성능을 보여주는 것으로 널리 알려져 있다.
- 과적합 방지 및 소규모 데이터 대응:  
본 프로젝트의 학습 데이터는 약 650개로, 딥러닝과 같은 대규모 모델을 적용하기에는 데이터 수가 제한적이다. XGBoost는 다양한 하이퍼 파라미터를 통해 모델의 복잡도를 제어하고 과적합을 효과적으로 막을 수 있어, 제한된 데이터 내에서도 일반화된 성능을 확보하기 유리하다.
- 비선형 관계 학습:  
축구 경기의 승패는 단순히 Elo Rating의 차이만으로 결정되지 않는다. 홈/어웨이 변수, 최근 5경기 득실점 흐름 등 다양한 변수 간의 복잡한 비선형 상호작용을 트리 구조를 통해 효과적으로 포착할 수 있다.

### -모델 구현

#### ① 경기 단위 층화 추출

일반적인 무작위 분할대신 경기 match\_id를 기준으로 데이터를 분할하였다.

- 데이터 누수 방지: 한 경기가 'A vs B'와 'B vs A'로 증강되어 있으므로, 동일 경기가 Train과 Test 세트에 동시에 들어가는 현상을 차단했다.
- 일관된 평가: 특정 경기에 속한 두 데이터가 반드시 같은 세트에 포함되게 하여, 모델이 완전히 새로운 경기에 대해 예측하는 능력을公正하게 평가하도록 설계했다.
- 무승부 비율 유지: 전체 데이터셋의 무승부 비율을 Train과 Test 세트에서도 동일하게 유지하여 데이터 불균형으로 인한 편향을 방지했다.

#### ② 하이퍼파라미터 최적화

데이터 규모를 고려하여 과적합 방지와 일반화 성능 강화에 집중했다.

- 학습률 (eta = 0.05): 가중치 업데이트 속도를 늦추어 모델이 데이터의 패턴을 보다 세밀하게 찾아내도록 유도했다.
- 나무 깊이 (Max Depth = 4): 개별 결정 나무의 복잡도를 제한하여, 데이터 내의 노이즈까지 학습해버리는 과적합을 방지했다.
- 확률적 샘플링:
  - subsample = 0.7: 매 단계에서 데이터의 70%만 활용
  - colsample\_bytree = 0.8 변수의 80%만 활용

#### ③ 조기 종료 및 목적 함수 설정

- 목적 함수 (Softmax Cross Entropy): 단순히 승무패를 맞히는 것을 넘어, 각 결과가 나타날 확률값을 산출했다.
- 평가 지표: 오분류율을 기준으로 모델을 평가하여 실질적인 정확도를 확보했다.
- Early Stopping : 1,000회의 학습을 설정하되, 50회 연속으로 성능 개선이 없을 경우 학습을 자동 중단하여 최적의 성능 지점에서 모델을 보존했다.

#### 4. 최종 예측 및 양상들

단일 예측의 불안정성을 제거하기 위해 양방향 확률 평균화 기법을 적용했다.

- Symmetric Ensemble: [A vs B] 예측 결과와 [B vs A] 예측 결과를 각각 독립적으로 수행한다.
- 확률 보정: 'A가 이길 확률'과 'B가 질 확률'을 산술 평균하여 최종 확률을 도출한다.

구현한 코드는 다음과 같다.

```
# 2026 월드컵 결과 예측 모델: XGBoost 기반 학습 및 양상을 예측

library(xgboost)
library(caret)
library(tidyverse)
library(readr)

# 1. # 학습용 데이터 불러오기 및 파생변수 생성
worldcup <- read.csv("C:/Users/LEEBG/Desktop/월드컵_학습데이터.csv", header = T)

# Elo 점수 차이 및 경기 고유 ID 생성
worldcup <- worldcup %>%
  mutate(
    Elo_Diff = Elo - Elo_Opp,
    match_id = paste(date, pmin(Team, Opponent), pmax(Team, Opponent), sep = "_")
  )

# 2. 경기 단위 층화 추출 (Stratified Split by Match)
# Train/Test 세트 내 '무승부 경기'의 비율을 동일하게 유지

# (1) 경기별 결과 태입 요약표 생성
match_summary <- worldcup %>%
  group_by(match_id) %>%
  summarise(
    # 해당 경기 결과 중 하나라도 무승부가 있으면 "Draw", 아니면 "Decisive"
    Match_Type = ifelse(any(Result_Class == 1), "Draw", "Decisive")
  )

# 전체 경기 태입 비율 확인
print(prop.table(table(match_summary$Match_Type)))

# (2) 경기 태입 비율을 유지하며 70% 추출
set.seed(2026)
train_idx <- createDataPartition(match_summary$Match_Type, p = 0.7, list = FALSE)

# (3) 추출된 인덱스 기반 Match ID 리스트 확보
train_match_ids <- match_summary$match_id[train_idx]

# (4) 실제 데이터 분할 (학습용 / 테스트용)
train_data <- worldcup %>% filter(match_id %in% train_match_ids)
test_data <- worldcup %>% filter(!match_id %in% train_match_ids)

# 데이터셋 별 승/무/패 비율 검증
print(prop.table(table(train_data$Result_Class)))
```

```

print(prop.table(table(test_data$Result_Class)))

# 3. XGBoost 모델 학습 (확률 계산 모드)
# 학습에 사용할 변수 정의
features <- c("Venue", "Elo_Diff", "Recent_GF_5", "Recent_GA_5", "Recent_GF_5_Opp",
"Recent_GA_5_Opp")

# XGBoost 전용 행렬 생성
dtrain <- xgb.DMatrix(
  data = as.matrix(train_data %>% select(all_of(features)) %>% mutate(Venue = as.numeric(as.character(Venue))),
  label = as.numeric(as.character(train_data$Result_Class))
)

dtest <- xgb.DMatrix(
  data = as.matrix(test_data %>% select(all_of(features)) %>% mutate(Venue = as.numeric(as.character(Venue))),
  label = as.numeric(as.character(test_data$Result_Class))
)

# 하이퍼파라미터 설정
params1 <- list(
  booster = "gbtree",
  objective = "multi:softprob", # 결과별 확률 출력
  num_class = 3,               # 승/무/패 3개 클래스
  eval_metric = "merror",      # 오분류율 기준으로 평가
  eta = 0.05,                 # 학습률
  max_depth = 4,              # 나무 깊이
  subsample = 0.7,             # 데이터 샘플링 비율
  colsample_bytree = 0.8       # 변수 샘플링 비율
)

# 모델 학습 시작
bst_model1 <- xgb.train(
  params = params1,
  data = dtrain,
  nrounds = 1000,
  evals = list(test = dtest),
  early_stopping_rounds = 50, # 50회 이상 성능 미개선 시 조기 종료
  verbose = 0
)

# 성능 확인
pred_probs <- predict(bst_model1, dtest, reshape = TRUE)
pred_labels <- max.col(pred_probs) - 1
cm <- confusionMatrix(as.factor(pred_labels), as.factor(test_data$Result_Class))

print(cm$table)
print(paste("정확도(Accuracy):", round(cm$overall['Accuracy'] * 100, 2), "%"))

# 4. 2026 월드컵 정방향 예측 (Team vs Opponent)
target_path <- "C:/Users/LEEBG/Desktop/월드컵_조별리그_경기.csv"

```

```

real_matches <- read.csv(target_path, header = T)

# 입력 데이터 전처리
real_data_processed <- real_matches %>%
  mutate(
    Elo_Diff = Elo - Elo_Opp,
    Venue = as.numeric(as.character(Venue))
  )

dreal <- xgb.DMatrix(data = as.matrix(real_data_processed %>% select(all_of(features)))))

# 예측 수행 및 결과 행렬 변환
pred_vec <- predict(bst_model1, dreal)
pred_probs <- matrix(pred_vec, ncol = 3, byrow = FALSE)
colnames(pred_probs) <- c("Win_Prob", "Draw_Prob", "Loss_Prob")

# 정방향 결과 임시 저장
final_result <- real_matches %>%
  bind_cols(as.data.frame(pred_probs)) %>%
  mutate(
    Win_Prob = round(Win_Prob * 100, 1),
    Draw_Prob = round(Draw_Prob * 100, 1),
    Loss_Prob = round(Loss_Prob * 100, 1)
  )

write_excel_csv(final_result, "C:/Users/LEEBG/Desktop/2026_WorldCup_Prediction_Result.csv")

# 5. 2026 월드컵 역방향 예측 (Part 2: Opponent vs Team)
target_path2 <- "C:/Users/LEEBG/Desktop/월드컵_조별리그_경기2.csv"
real_matches2 <- read.csv(target_path2, header = T)

# 입력 데이터 전처리
real_data_processed2 <- real_matches2 %>%
  mutate(
    Elo_Diff = Elo - Elo_Opp,
    Venue = as.numeric(as.character(Venue))
  )

dreal2 <- xgb.DMatrix(data = as.matrix(real_data_processed2 %>% select(all_of(features)))))

# 예측 수행 및 결과 행렬 변환
pred_vec2 <- predict(bst_model1, dreal2)
pred_probs2 <- matrix(pred_vec2, ncol = 3, byrow = FALSE)
colnames(pred_probs2) <- c("Win_Prob", "Draw_Prob", "Loss_Prob")

# 역방향 결과 임시 저장
final_result2 <- real_matches2 %>%
  bind_cols(as.data.frame(pred_probs2)) %>%
  mutate(
    Win_Prob = round(Win_Prob * 100, 1),

```

```

    Draw_Prob = round(Draw_Prob * 100, 1),
    Loss_Prob = round(Loss_Prob * 100, 1)
  )

write_excel_csv(final_result2, "C:/Users/LEEBG/Desktop/2026_WorldCup_Prediction_Result2.csv")

# 6. 최종 양상을 (Final Ensemble): 양방향 확률 평균화
df1 <- read.csv("C:/Users/LEEBG/Desktop/2026_WorldCup_Prediction_Result.csv")
df2 <- read.csv("C:/Users/LEEBG/Desktop/2026_WorldCup_Prediction_Result2.csv")

final_ensemble <- df1 %>%
  mutate(
    # 최종 승리 확률 = (정방향 승리 확률 + 역방향 패배 확률) / 2
    Final_Win_Prob = (Win_Prob + df2$Loss_Prob) / 2,

    # 최종 무승부 확률 = (정방향 무승부 확률 + 역방향 무승부 확률) / 2
    Final_Draw_Prob = (Draw_Prob + df2$Draw_Prob) / 2,

    # 최종 패배 확률 = (정방향 패배 확률 + 역방향 승리 확률) / 2
    Final_Loss_Prob = (Loss_Prob + df2$Win_Prob) / 2,

    # 확률 합계 재조정 -> 100 % 기준으로 맞추기
    Total = Final_Win_Prob + Final_Draw_Prob + Final_Loss_Prob,
    Final_Win_Prob = round(Final_Win_Prob / Total, 3),
    Final_Draw_Prob = round(Final_Draw_Prob / Total, 3),
    Final_Loss_Prob = round(Final_Loss_Prob / Total, 3),

    # 최종 판정 로직
    Final_Prediction = case_when(
      Final_Win_Prob >= Final_Draw_Prob & Final_Win_Prob >= Final_Loss_Prob ~ "Team
      승",
      Final_Loss_Prob >= Final_Win_Prob & Final_Loss_Prob >= Final_Draw_Prob ~ "Opp
      onent 승",
      TRUE ~ "무승부"
    ),

    # 토너먼트 진출팀 판정
    Final_Tournament_Pick = ifelse(Final_Win_Prob >= Final_Loss_Prob, "Team 진출",
    "Opponent 진출")
  ) %>%
  select(Team, Opponent, Final_Prediction, Final_Win_Prob, Final_Draw_Prob, Final_L
oss_Prob, Final_Tournament_Pick)

```

- TEST DATA 예측 결과

	Reference		
Prediction	0	1	2
0	48	21	14
1	6	16	7
2	13	21	46

Accuracy(정확도) : 57.29%

-승리 예측

Recall(재현율) : 71.64 %

Precision(정밀도) : 57.83 %

-무승부 예측

Recall(재현율) : 27.59%

Precision(정밀도) : 55.17 %

-패배

Recall(재현율) : 68.66 %

Precision(정밀도) : 57.5 %

## 예측 성능 종합 평가

본 모델은 전체 정확도(Accuracy) 57.3%를 기록하며 무작위 예측 대비 약 1.7배 높은 성능을 보였다. 특히 클래스별 지표 분석을 통해 모델의 실전 예측 성향을 다음과 같이 정의할 수 있다.

- 승/패 예측의 높은 재현율:** 승리와 패배에 대한 재현율이 각각 71.6%와 68.7%로 나타났다. 이는 모델이 실제 승패가 갈리 는 경기의 패턴을 약 70% 수준으로 정확히 포착해내고 있음을 의미한다.
- 승/패 정밀도의 상대적 하락 원인:** 모델이 무승부 발생 가능성이 있는 경기에서도 승/패로 공격적인 예측을 내리는 경향이 있어 정밀도는 약 57%대를 유지하고 있다. 이는 무승부라는 변수를 최대한 배제하고 확실한 전력 차이를 바탕으로 승부를 예측하려는 모델의 특성이 반영된 결과이다.
- 무승부의 낮은 재현율(27.6%):** 축구 종목 특성상 예측 난이도가 가장 높은 무승부에 대해 모델은 매우 보수적인 기준을 적용하고 있다.
- 높은 무승부 정밀도(55.2%)의 함의:** 재현율은 낮지만, 무승부 정밀도가 55.2%에 달한다는 점은 매우 유의미한 결과이다. 이는 모델이 무승부를 남발하지 않으며, "모델이 무승부라고 예측한 경기 2판 중 1판 이상은 실제로 비겼다"는 것을 뜻한다. 따라서 모델의 무승부 예측은 발생 가능성이 매우 높은 상황에서만 도출한다고 판단된다.
- 

## 5. 최종 결과 예측

이 모델을 활용하여 모든 조별 예선을 예측하여, 다음과 같은 결론을 도출하였다.

A조

순위	팀	승	무	패	승점
1	멕시코	3	0	0	9
2	덴마크	2	0	1	6
3	대한민국	1	0	2	3
4	남아프리카 공화국	0	0	3	0

B조

순위	팀	승	무	패	승점
1	스위스	3	0	0	9
2	이탈리아	2	0	1	6
3	캐나다	1	0	2	3
4	카타르	0	0	3	0

C조

순위	팀	승	무	패	승점
1	브라질	3	0	0	9
2	모로코	1	1	1	4
3	스코틀랜드	1	1	1	4
4	아이티	0	0	0	0

D조

순위	팀	승	무	패	승점
1	터키	3	0	0	9
2	미국	2	0	1	6
3	호주	1	0	2	3
4	파라과이	0	0	3	0

E조

순위	팀	승	무	패	승점
1	에콰도르	2	1	0	7
2	독일	2	1	0	7
3	코트디부아르	1	0	2	3
4	퀴라소	0	0	3	0

F조

순위	팀	승	무	패	승점
1	네덜란드	3	0	0	9
2	일본	2	0	1	6
3	우크라이나	1	0	2	3
4	튀니지	0	0	3	0

G조

순위	팀	승	무	패	승점
1	벨기예	2	1	0	7
2	이란	1	2	0	5
3	이집트	1	1	1	4
4	뉴질랜드	0	0	3	0

H조

순위	팀	승	무	패	승점
1	스페인	3	0	0	9
2	우루과이	2	0	1	6
3	사우디아라비아	0	1	2	1
4	카보베르데	0	1	2	1

I조

순위	팀	승	무	패	승점
1	프랑스	3	0	0	9
2	세네갈	2	0	1	6

3	노르웨이	1	0	2	3
4	이라크	0	0	3	0

J조

순위	팀	승	무	패	승점
1	아르헨티나	3	0	0	9
2	오스트리아	2	0	1	6
3	요르단	1	0	2	3
4	알제리	0	0	3	0

K조

순위	팀	승	무	패	승점
1	콜롬비아	2	1	0	7
2	포르투갈	2	1	0	7
3	우즈베키스탄	1	0	2	3
4	콩고민주공화국	0	0	3	0

L조

순위	팀	승	무	패	승점
1	잉글랜드	3	0	0	9
2	크로아티아	2	0	1	6
3	파나마	1	0	2	3
4	가나	0	0	3	0

[2026\\_WorldCup\\_Prediction\\_Result.csv](#)

[2026\\_WorldCup\\_Prediction\\_Result2.csv](#)

[2026\\_WorldCup\\_FINAL\\_ENSEMBLE.csv](#)

위 파일들을 통해 각 경기 예측 결과 및 구체적인 확률값까지 확인할 수 있다.

## 6. 결론

본 프로젝트는 2026 북중미 월드컵 예측의 정확도를 높이기 위해 카타르 월드컵 이후의 경기 데이터만을 선별하여 학습에 활용했다. 장기간의 누적 데이터를 배제함에 따라 학습 데이터의 절대량이 부족해져 모델의 성능 지표는 기존 방식보다 낮게 측정되었다. 그러나 이는 과거의 기록이 아닌 현재 시점의 전력을 가장 정밀하게 반영하기 위한 의도적인 선택이었으며, 이를 통해 현재의 축구 흐름을 투영한 북중미 월드컵 맞춤형 모델로서의 정체성을 확보했다.

모델 구축 과정에서 A대표팀의 스쿼드 가치나 선수 개별의 시장 가치와 같은 정성적 변수들을 학습 데이터로 변환하는 데 현실적인 어려움이 따랐다. 이러한 변수들은 실제 경기 결과에 큰 영향을 미치는 요소임에도 불구하고, 일관된 수치로 산출하여 모델에 직접 반영하기에는 데이터의 가공 및 결합 과정에서 한계가 존재했다. 이는 향후 프로젝트에서 다양한 외부 지표를 정량화하여 모델의 복합적인 분석 능력을 강화해야 한다는 과제를 남겼다.

이번 분석은 경기 스코어가 아닌 승무패 결과 예측에 집중했기 때문에, 조별 예선의 순위 결정 방식인 다득점 원칙을 적용할 수 없

는 구조적 특성을 지닌다. 특히 승점과 맞대결 결과가 모두 동률인 상황에서 발생하는 순위 산정의 공백을 해결하기 위해 별도의 논리적 장치를 마련했다. 다득점이라는 불확실한 변수를 임의로 생성하는 대신, 모델이 계산한 상대 팀 대비 승리 확률을 비교하여 승률이 더 높은 팀을 상위 순위에 배치했다. 이러한 방식은 스코어 예측의 부재 속에서도 통계적 근거를 바탕으로 순위를 결정할 수 있는 가장 합리적인 대안이었다.

결론적으로 본 프로젝트는 데이터의 양적 풍부함보다는 질적인 최신성에 주목하여 2026년 월드컵의 판도를 분석했다는 데 의의가 있다. 비록 모델의 수치적 성능과 변수 활용 면에서 개선의 여지가 존재하지만, 승리 확률 기반의 순위 산정 로직을 통해 예측의 논리적 일관성을 유지하고자 노력했다. 이번 프로젝트는 제한된 환경 속에서도 데이터의 특성을 고려한 최적의 예측 모델을 설계하는 유의미한 경험이 되었다.