



TEAM NO.3

# DATA DRIVEN SPORTS

실시간 승률을  
이용한 축구 중계

# 목차

---

1. 배경

2. 분석

3. 모델링

4. 한계

5. 향후 과제&발전

# 배경

---

- 해외축구(EPL) 경기를 시청하다 보면 그림(1) 같이 화면 좌상단에 승률이 나옴
- K-league 중계 플랫폼(쿠팡플레이)에선 중계 중 승률을 보여주지 않음
- 경기에서 누적된 데이터를 활용해 승률을 구해보면 어떨까?
- 승률을 중계 화면에 보여줌으로써 중계진들에게 다양성을 제시  
→ 중계 퀄리티 향상



<그림 1. 중계 화면>

# 배경

---

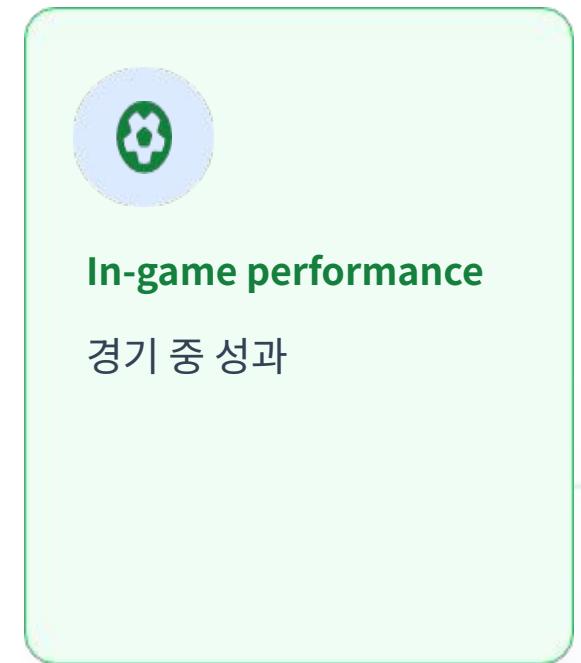
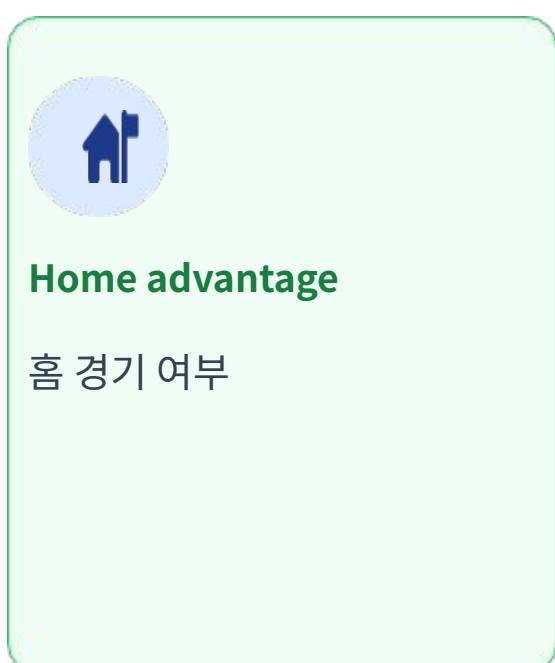
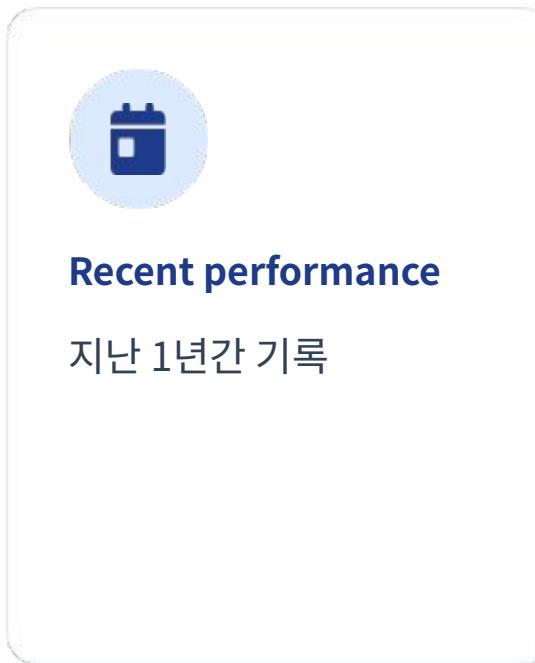
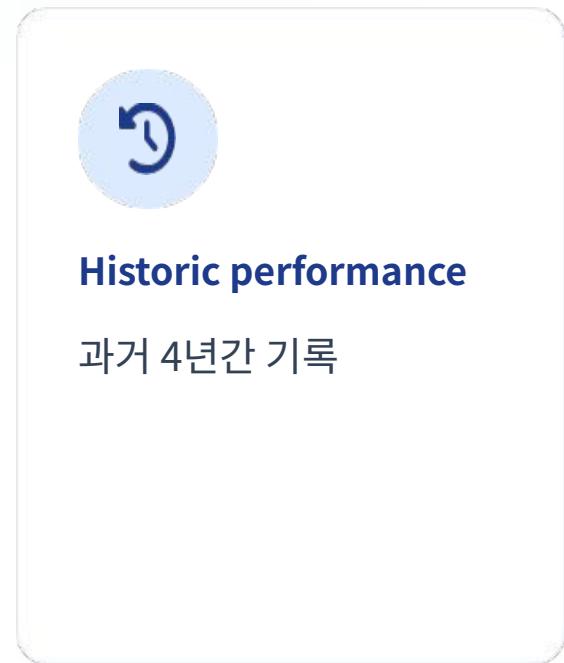
- ⚽ 승률은 어떻게 해서 구할 것인가?
- ⚽ 해외 통계 사이트&논문에선 어떻게 승률을 구하는 것일까?
- ⚽ 경기에서 누적된 데이터를 활용해 어떻게 승률을 구할까?
- ⚽ 모델링 결과를 중계화면에 적용해도 괜찮을까?
- ⚽ 향후 과제와 발전 가능성은 무엇이 있을까?

K-league 경쟁력 향상

# 배경

---

스포츠 통계 사이트 OPTA

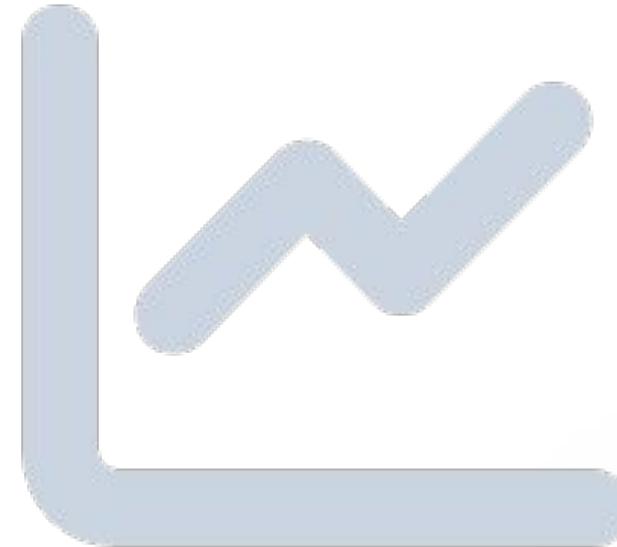


그 외에도 여러가지 변수들이 존재 함을 확인  
→ **홈 여부와, 경기중 성과를 이용해 승률 도출**

# 분석

---

- ❶ 축구는 득점이 적은 스포츠이며 득점과 연관된 지표가 필요
- ❷ 득점 외에도 경기력 또한 중요하다 판단
- ❸ 득점과 경기력을 통해 경기중 성과를 정량화



# 분석

---

## raw data

`time_seconds` (시간)

`start_x, y / end_x, y` (좌표)

`result_name` (액션의 결과)

`action_id` (경기내 이벤트 순서)

`type_name` (어떤 액션)

## match\_info

`home_score, away_score` (홈, 원정 점수)

`home_team_id, away_team_id` (홈 원정팀의 id)

## 공통

`game_id` (경기 번호)

# 분석

---

## 분석 타임라인

1

EDA

2

XG, XA, XT

3

누적&최근  
10분차이

4

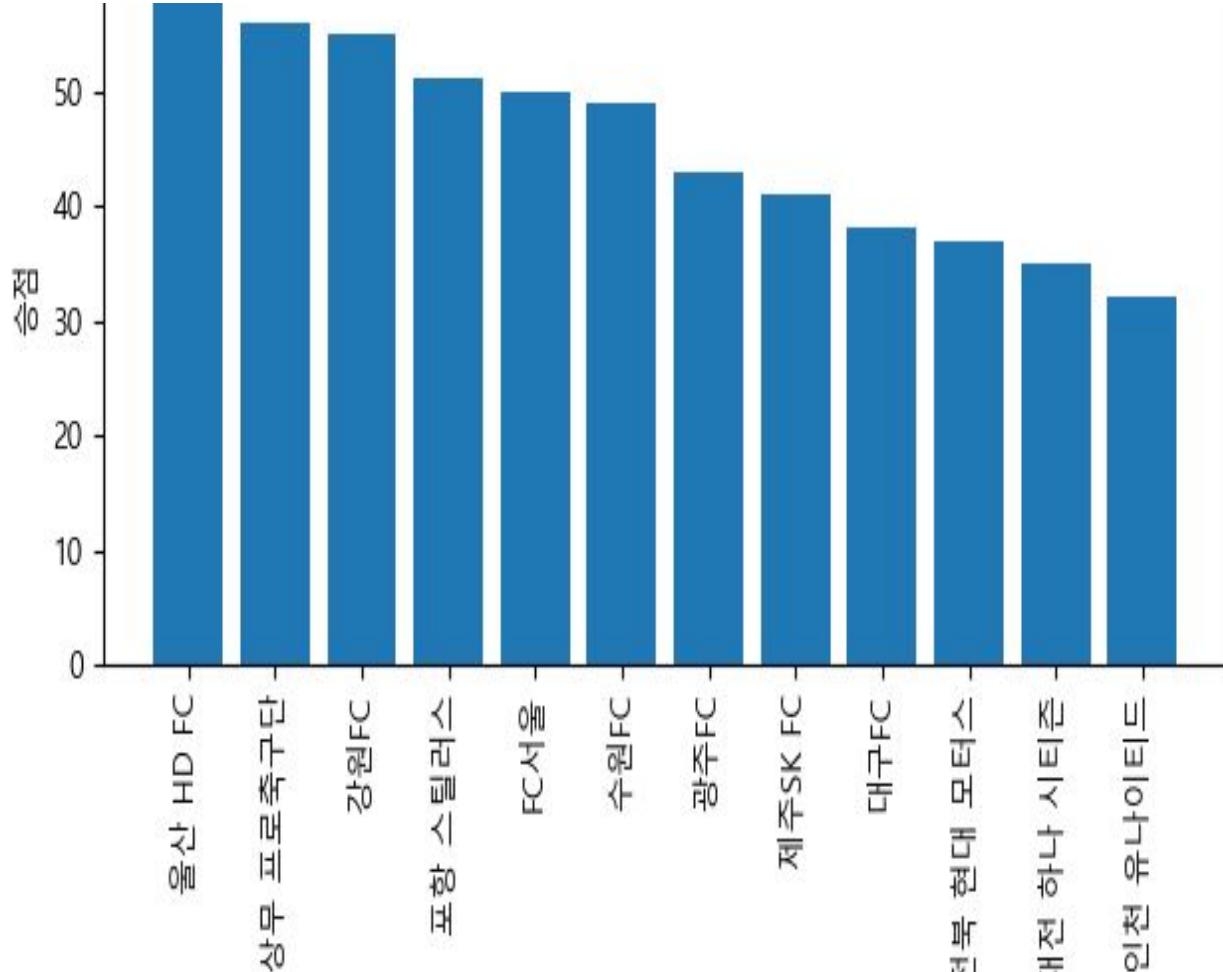
모델링

5

결과 해석

# 분석

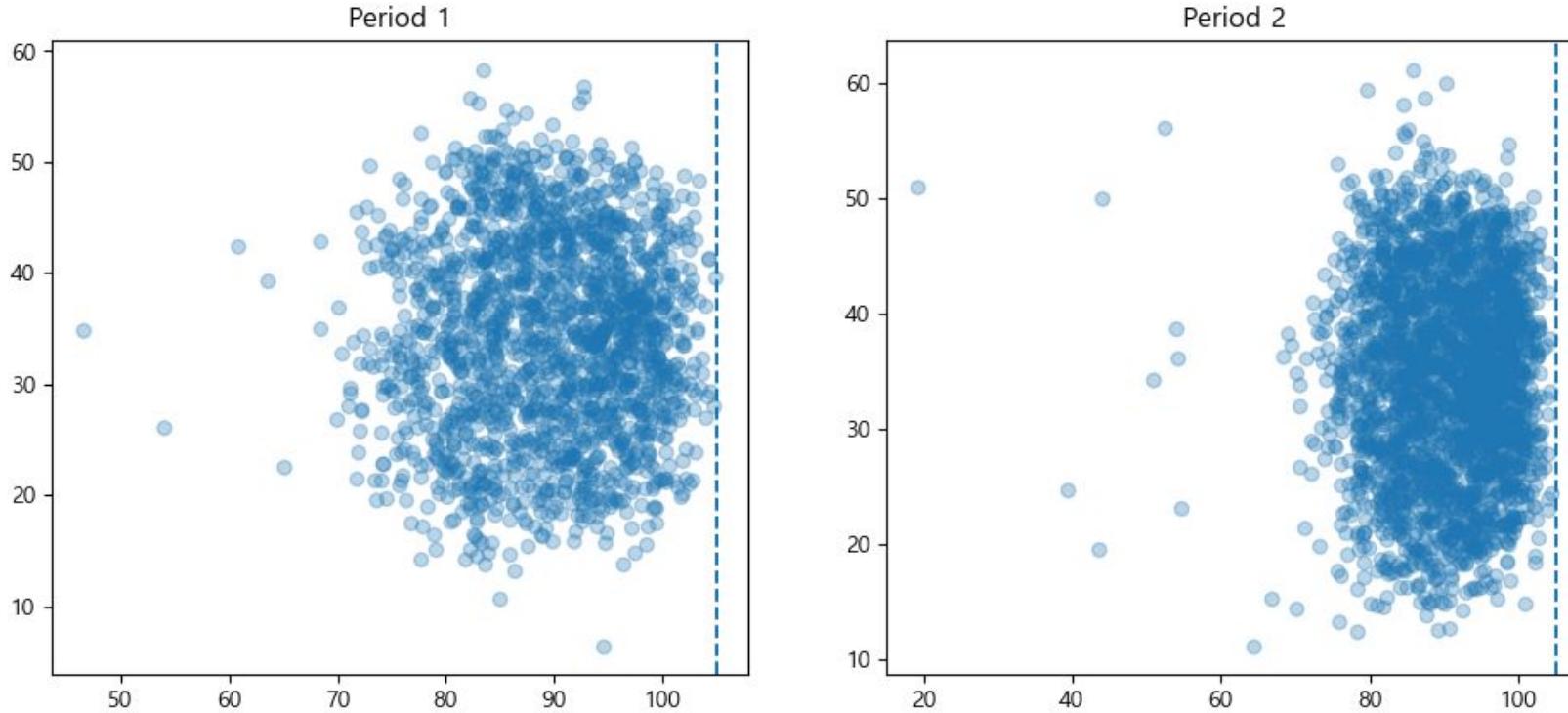
- ⚽ 24시즌 33라운드 기준
- ⚽ 팀별 승점으로 순위 도출



<그림 2. 33라운드 기준 순위표>

# 분석

## 전, 후반 슛팅 분포도 비교



<그림 3. 전&후반 슛팅 좌표 비교>

## 후반전이 골대 근처 슛팅이 많음

# 분석

---

## XG, XA, XT?

### XG

#### 슈팅 시 득점 확률

1. 슛을 찬 위치에서 골대 까지의 거리와 각도가 얼마인지, 패널티 박스 안 밖이냐, 패널트킥 여부인가에 따라서 골이 되었을 확률을 도출
2. XG값으로 골 결정력에 대해 이야기 가능
3. 공격의 핵심 지표

### XA

#### 슈팅 직전 패스의 득점 확률

1. 슛팅 직전의 패스가 얼마나 득점에 기여하였는가
2. 해당 패스가 득점이 될 확률
3. 슛팅 이외의 보이지 않는 기여도 반영

### XT

#### 경기 전체 영향력 지표

1. 축구장을 12x6 사이즈로 만든 후 전진 후진 여부(패스, 드리블 등)를 가지고 판단
2. 전진시 +, 후진시 - 값으로 도출
3. 경기 흐름을 파악

# 분석

## XG, XA, XT

Action ID	xg	xa	xt
6	0.0	0.0	0.004
7	0.0	0.0	0.010
12	0.0	0.0	0.042
...	...	...	...

<표 1. XG, XA, XT값>

# 분석

---

## 누적 & 10분 차이

- ⚽ XG, XA, XT 값을 누적 시킨후 10분간 차이
- ⚽ Home team 기준으로 우세하면 (+), 불리하면 (-)
- ⚽ 승률에 필요한 변수

	home_xgdiff	home_xadiff	home_xtdiff
	0.000000	0.000000	-0.008
	0.000000	0.000000	0.053
	-0.913295	-0.392972	-0.113
	0.489410	0.489410	0.139

<표 2. Home 기준 차이값>

# 모델링

---

## 모델링 과정

- ➊ XG, XA, XT 차이, home여부, Time을 이용해 모델링 진행 예정



- ➋ 경기 action마다 데이터가 기록 → 데이터 압축 진행



- ➌ 다중 로지스틱, 포아송이용 → 승 무 패 확률 도출



- ➍ 시작화



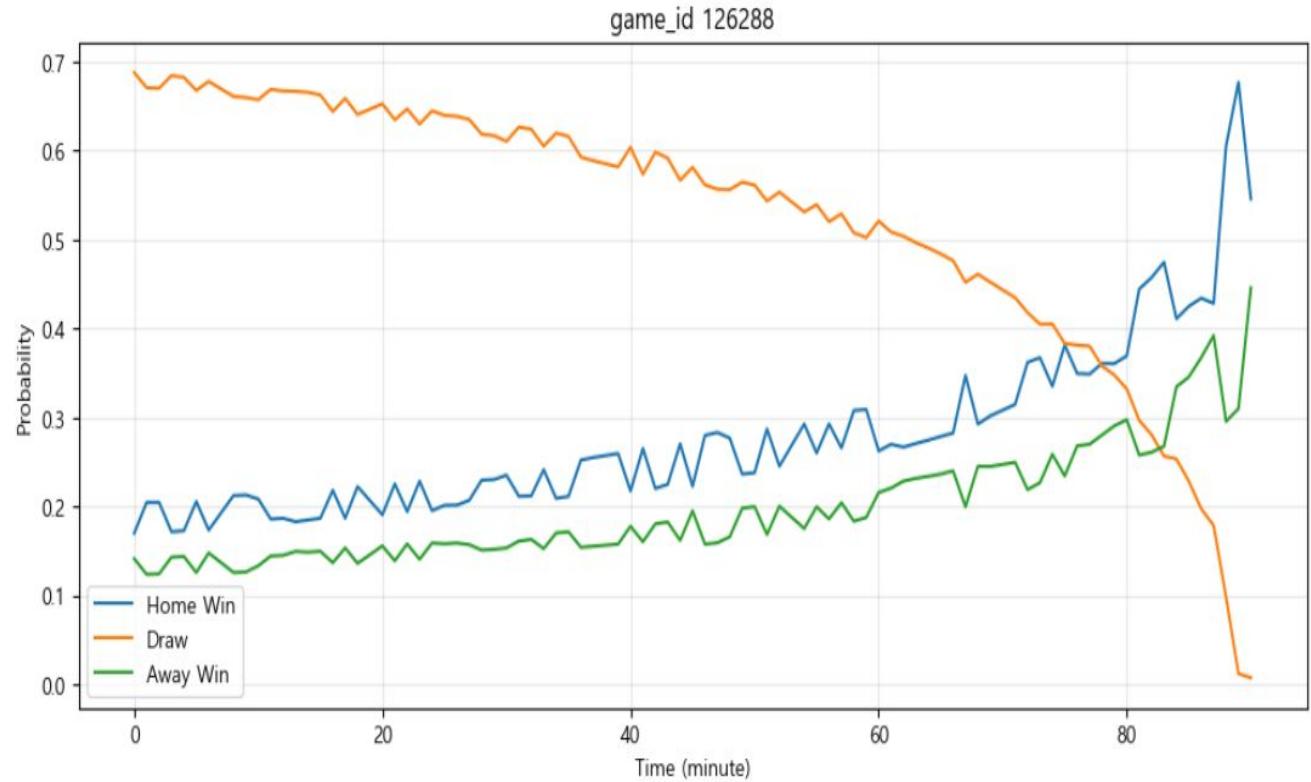
- ➎ 두 모델 성능 비교



# 모델링

## Multiple Logistic Regression

- ⚽ 전체적으로 완만한 그래프
- ⚽ 실제 경기와 비교 = 원정팀 승(0 : 1)
- ⚽ 점수 적용시 모델이 0, 1값으로 극단적으로 치우치게 됨
- ⚽ 점수 변수 제외후 모델 학습 → 경기력만으로 승률 판단
- ⚽ 보완 작업 필요

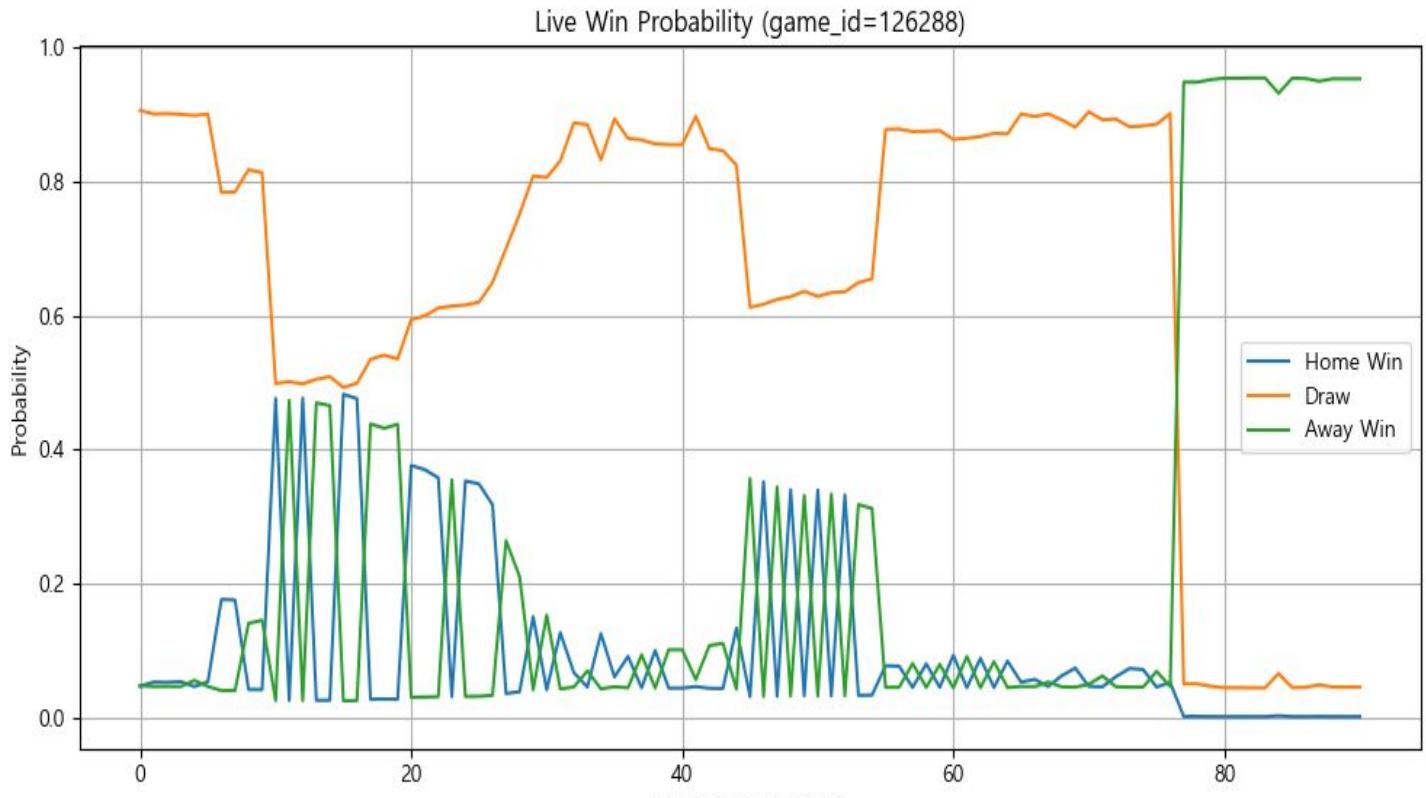


<그림 4. 다중 로지스틱>

# 모델링

## Poisson

- 경기력을 가지고 미래 득점 확률(승률)을 계산
- 득점&남은 시간에 따라 승률 변동
- 득점 시점부터 확률 극단



<그림 5. 포아송 분포>

# 모델링

---

## Multiple Logistic Regression VS Poisson

- 두 모델중 어떤 것이 최선일지 logloss와 brier score로 판단

	Model	LogLoss	Brier Score
0	Multinomial Logistic	0.929847	0.537226
1	Poisson	0.161838	0.057189

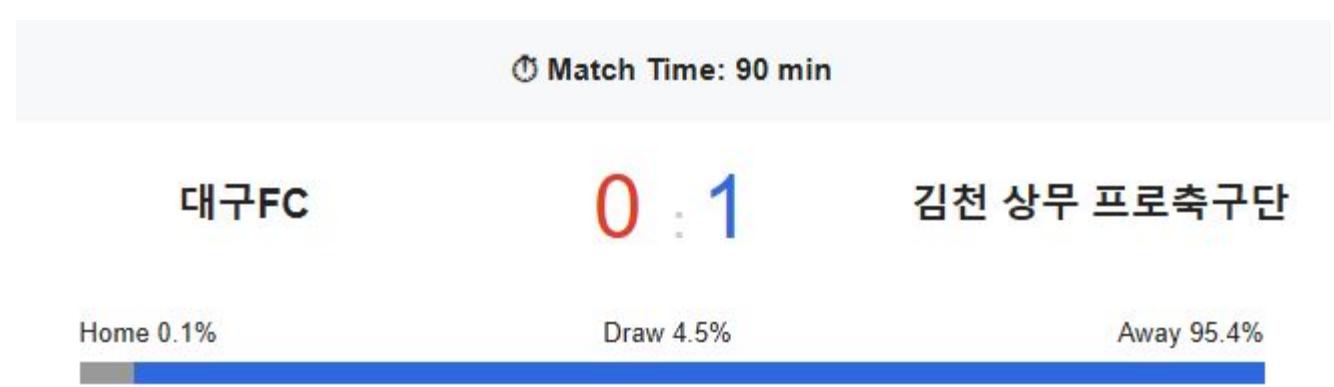
<그림 6. 다중로지스틱 과 포아송 계수>

- 다중 로지스틱이 logloss와 brier score가 높게 나옴
- 포아송 계수가 안정적 → **포아송으로 대시보드 생성**

# 모델링

## 대시보드

❶ Poisson 기반 대시보드 생성



❷ Match Time 90 min 기준 1분마다 승률이 변화



❸ XT, XA, XG에 따라 공격 흐름 우세여부 판단

<그림 6. 대시보드>

# 향후 과제&발전

## 한계

### 득점의 한계

축구는 득점의 영향이 큰  
스포츠, 모델이 득점을  
인식시 남은시간&경기력  
상관 없이 승률 과다  
→ 득점 의존도 조정 필요

### 중요 변수 적용 X

퇴장, 부상 등 중요 변수  
반영 X  
→ 새로운 변수 필요

### 기준 모델의 한계

다중 로지스틱, 포아송  
분포는 득점의 한계 발생  
→ 새로운 모델 적용

### 거시적인 XG

상대 수비의 좌표와 슛팅  
찬 선수의 주 발 여부 등을  
고려하지 않은 XG  
→ 거시적인 XG, 다양한  
변수들 적용 필요 XA, XT  
도 포함

### 과거 데이터

OPTA에서 승률을 구할 때  
과거 팀의 경기력을 반영  
→ 24시즌 33라운드의  
기록만 가지고 분석

# 향후 과제&발전

---

## 과제



### 과제 1

득점의 영향력을 최소화  
→ 승률 극단화 예방



### 과제 2

다양한 변수를 적용 해  
모델 성능UP



### 과제 3

과거 데이터를 적용  
→ 승률 고도화 작업



### 과제 4

기존 Multiple Logistic  
Regression, Poisson  
→ Bayesian 으로 확장

# 향후 과제&발전

---

## 발전



### 고도화된 승률 도출

경기 중 실시간 데이터를 이용하여  
특정 시점 승률 도출



### 데이터 공급처와 소통

데이터 공급처와 중계사 간의 유연한  
소통 체계를 구축하여 안정적인  
공급망 확보



### 방송 콘텐츠 활용

승률 그래프를 하프타임 분석 및  
풀타임 이후 프리뷰 쇼 등 다양한 방송  
콘텐츠로 활용

# 출처 및 참조

---

## ⓐ 출처

[https://www.researchgate.net/figure/An-example-of-the-Win-Probability-feature-during-the-live-broadcast-of-the-English\\_fig1\\_392529556](https://www.researchgate.net/figure/An-example-of-the-Win-Probability-feature-during-the-live-broadcast-of-the-English_fig1_392529556) - 그림 (1)

## ▣ 참조

Robberechts, P., Van Haaren, J., & Davis, J. (2021).

*A Bayesian approach to in-game win probability in soccer.*

Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '21), 10 pages.

<https://doi.org/10.1145/3447548.3467194>

<https://theanalyst.com/> (OPTA)