

# 11월 6일 - 주제 좁히기, 사용할 데이터 추리기

[주제 좁히기]

1. 시대별 축구 전략 스타일 변화 확인
2. 포지션의 중요도 변화
3. 승리에 주요한 Event 지표 설정

필요한 데이터 혹은 변수 정의

- 경기별 주요 Event 모두 수집
  - tuckyHSD
  - 분석 방법론 search, 참고문헌 search
- 참고문헌
    1. 축구 경기의 경기력 분석 (2015)

▼ 요약

## 1. 서론

이 연구는 축구 경기에서 승패에 영향을 미치는 요인을 과학적으로 분석하고자 합니다. 과거에는 기술의 한계로 비디오 분석에만 의존했으나, IT 기술이 발전함에 따라 위치 정보 등을 포함한 구체적인 데이터를 수집할 수 있게 되었습니다. 이러한 기술을 활용해 단순한 기록 이상으로 경기력을 분석하려는 시도가 본 연구의 목표입니다.

## 2. 연구내용

### 2.1 자료설명

연구는 2013년 K-리그에서 승패가 결정된 51경기를 대상으로 합니다. 연구에 사용된 데이터는 경기장에서 선수들의 위치와 움직임을 실시간으로 기록한 시스템을 통해 수집되었습니다. 주요 변수로는 득점, 실점뿐만 아니라 스프린트 거리, 패스 횟수, 수비 시 선수 간 거리 등을 설정하였습니다.

### 2.2 변수의 정의 및 설명

경기 결과를 분석하기 위해 각 변수는 다음과 같이 정의되었습니다:

- **sprint dist**: 팀 전체의 스프린트 거리
- **block**: 골키퍼 외 수비수의 슛 블록 횟수
- **pe shots**: 페널티 지역에서의 슛 횟수
- **def depth**: 수비 시 최전방 공격수와 최후방 수비수 간 거리
- **ave wide**: 경기 중 좌우 선수 간 거리

### 3. 연구결과

#### 3.1 중요변수의 평균검정

평균검정을 통해 승리한 팀과 패배한 팀 간의 경기력 차이를 분석했습니다. 승리 팀은 스프린트 거리, 페널티 지역 슛 횟수 등이 유의미하게 높았고, 패배 팀은 평균 패스 횟수와 좌우 선수 간 거리가 높게 나타났습니다. 이 결과는 패배 팀이 공격 및 수비 간격을 잘 유지하지 못했음을 시사합니다.

#### 3.2 예측모형을 이용한 승패요인 분석

##### 3.2.1 의사결정나무

의사결정나무 모델을 사용해 승패를 예측했으며, 정확도는 72.55%로 나타났습니다. 분석 결과 좌우 폭이 35.5m 이상일 때 패배할 가능성이 높았으며, 블록 횟수가 많을수록 승리할 가능성이 증가했습니다.

##### 3.2.2 로지스틱 회귀

로지스틱 회귀모형은 승리할 확률을 예측하는 데 사용되었고, 정확도는 74%였습니다. 수비 시 공격수와 수비수 간 거리가 넓을수록 패배할 가능성이 높아졌으며, 블록 횟수가 많을수록 승리 가능성이 증가하는 것으로 나타났습니다.

##### 3.2.3 판별분석

판별분석 모델은 정확도가 74.5%로 가장 높았습니다. 분석 결과 패스 횟수, 수비 시 선수 간 거리, 블록 횟수가 승패에 큰 영향을 미치는 요인으로 확인되었습니다.

#### 3.3 득점을 이용한 공격력 분석

득점 요인을 분석하기 위해 포아송 회귀모형을 사용했습니다. 분석 결과 논스톱 패스와 템포(공 처리 속도)가 득점에 큰 영향을 미치는 요인으로 나타났으며, 공격 속도가 느려지면 득점 기회가 줄어들음을 확인할 수 있었습니다.

### 3.4 실점을 이용한 수비력 분석

실점 요인 분석에서도 포아송 회귀모형을 사용했으며, 수비 시 선수 간 거리와 좌우 폭이 실점에 중요한 영향을 미치는 변수로 나타났습니다. 수비 간격이 넓으면 실점 확률이 증가했습니다.

## 4. 논의 및 결론

본 연구에서는 다양한 예측 모형을 사용해 축구 경기의 승패 요인을 분석했습니다. 결과적으로 수비 요인이 승패에 중요한 역할을 한다는 것을 확인할 수 있었으며, 특히 수비 간격과 좌우 폭이 좁을수록 실점 확률이 낮아짐을 알 수 있었습니다. 공격력 측면에서는 논스톱 패스와 공격 속도가 중요한 요인으로 작용했습니다.

- 데이터 분석

3943077event\_data

3943077.json 파일에서 얻을 수 있는 축구 경기 데이터를 엑셀로 변환

1. 이벤트 타이밍과 점유율

- 이벤트 타이밍 : 분, 초 단위로 기록 (timestamp, minute&second)
- 점유율 : 이벤트 발생 시점 공을 점유하고 있는 팀 명시(possession, duration)

2. 이벤트 유형 및 결과

- 이벤트 종류 : 패스, 슈트, 볼 리커버리 등의 이벤트 → 경기 중 각 팀의 공격/수비 패턴, 득점 기회 생성 방법 등 분석 가능 (pass, shot, ball\_recovery) (pass.length, pass.angle 등 세부사항도 포함)
- 이벤트 결과(성공/실패) : 특정 플레이 스타일의 성공률 계산, 팀/선수의 효율성 평가 가능(짧은 패스, 긴 패스 성공률 등)

3. 선수 행동 및 교체(substitution.outcome.id&substitution.outcome.name, substitution.replacement.id & substitution.replacement.name)

4. 공간 데이터(위치 좌표)

- 이벤트 위치 : 이벤트 발생한 위치 → 주요 공격 루트, 수비 위치, 득점 상황 시각적 표현, 공격 성공률 분석 가능

- 패스 네트워크 분석 → 패스맵 구현, 패스 빈도, 키플레이어 식별 가능

#### 5. 플레이 중단 정보

- 부상 및 반칙으로 인한 경기 중단 시점(injury\_stoppage.in\_chain)
- 카드 정보 (**bad\_behaviour.card.id&bad\_behaviour.card.name**)

#### 6. 이벤트 연계

- 연관 이벤트 : 이벤트끼리 연계되어 있어 복잡한 플레이 시퀀스 재구성 가능, 패스-슛-골 (related\_events)
- 상호작용 분석 : 공격/수비 시퀀스 비교 분석