

2020-2 프로젝트 8팀 예비보고서

# 잉글랜드 프리미어 리그 경기력 예측 (EPL 경기력 예측)



Premier  
League

# 목차

---

Step 1

주제 선정 계기

Step 2

프로젝트 개요

Step 3

기대 효과

Step 4

References

Q. 왜 이 문제를 해결하고 싶은가?



컴퓨터가 발달이 덜 된 1990년대 이전에는 대용량의 축구 자료들을 기존 전문가들이 조사와 직감만을 통해서 축구 승패를 예측하고 전략적 작전을 구상해왔었다. 많은 경험과 노하우를 통해 시합 날 개별 선수들과 상대방 팀의 컨디션 등을 보고 그때 그때 작전을 변경할 수 있어 더 큰 효과를 불러일으킬 수 있지만, 그와 동시에 직감만으로 인해 변경되는 것이므로 위험한 결과를 일으킬 수도 있을 뿐더러 방대한 양을 사람이 다 참고하고 외우기엔 불가능한 일이었다.

1990년대 이후 컴퓨터의 급속한 발달로 인하여 대용량의 자료를 처리하는 것이 가능하게 되었다. 급속도로 늘어나고 있는 축구에 대한 자료들을 데이터 마이닝을 통해 내재된 의미 있는 상관 관계, 패턴, 경향 등을 쉽고 빠르게 찾아낼 수 있게 되었다.

최근 스포츠 경기의 승패를 예측하는 시스템이 많이 도입되면서 예측결과를 얻을 수 있게 되었다. 이를 통해 더 나은 전략적 작전을 구상하고 경기력을 향상시킬 수 있도록 많은 도움이 되었지만, **실질적으로는 데이터 분석을 통한 스포츠의 승부 예측이 높은 정확성을 가지고 있는지는** 아직도 의문이 있다. 이러한 점을 통해 과연 분석과 예측이 실질적인 효과를 불러일으키는지에 대해 의문이 들어 **축구 데이터들을 통해 미래의 승패를 예측하는 것이 가능한 지에 대해 알고 싶었다.**

## Step 2

# 프로젝트 개요

Q 데이터를 어떻게 가져오고 처리할 것인가?



우리가 측정하고자 하는 스포츠 경기 승패 예측은 잉글랜드 프리미어 리그의 승패율이다. 이것에 대한 자료는 잉글랜드 프리미어 리그의 모든 통계자료를 사용할 것이다.

<https://hducc.handong.edu/em/5fb38303c203a>

경기 승패 예측은 다양한 변수를 고려하는데, 대표적인 feature로는, 홈 경기와 원정 경기에서의 경기력 등이 있다. 이러한 변수들에 대한 자료는 모두 유럽 프리미어 리그 공식 홈페이지에서 제공받을 수 있다. 그리고 그 자료를 보기 좋게 정리한 사이트가 있다.

<https://1xbet.whoscored.com/Regions/252/Tournaments/2/England-Premier-League>

### Home feature

HomeTeam: 홈 팀 이름  
FTHG: Full time 홈 팀 득점 기록  
HTHG: Halt time 홈 팀 득점 기록  
HS: 홈 팀 슈팅 수  
HST: 홈 팀 유효슈팅 수  
HF: 홈 팀 반칙 수  
HC: 홈 팀 코너킥 수  
HY: 홈 팀 경고 수  
HR: 홈 팀 퇴장 수

### Away feature

AwayTeam: 어웨이 팀 이름  
FTAG: Full time 어웨이 팀 득점 기록  
HTAG: Halt time 어웨이 팀 득점 기록  
AS: 어웨이 팀 슈팅 수  
AST: 어웨이 팀 유효슈팅 수  
AF: 어웨이 팀 반칙 수  
AC: 어웨이 팀 코너킥 수  
AY: 어웨이 팀 경고 수  
AR: 어웨이 팀 퇴장 수



## Q 사용하고자 하는 알고리즘은 무엇인가?

우리는 EPL(유럽 프리미어 리그)의 다양한 변인들의 데이터를 학습시키기 위하여 다음과 같은 알고리즘을 사용하려고 한다. 다음의 알고리즘은 선행 연구에서 제시된 알고리즘들이다.

<https://medium.com/@bjmoon.korea/ai-x-%EB%94%A5%EB%9F%AC%EB%8B%9D-fianl-assignment-84e66d7e451d>

### Deep Learning 알고리즘

#### 1 ANN

인공신경망은 시냅스의 결합으로 네트워크를 형성한 인공뉴런(노드)가 학습을 통해 시냅스의 결합 세기를 변화시키고 문제 해결 능력을 가지는 모델 전반을 가리킨다. 우리는 다중 인공신경망을 사용하여 20개가 넘는 분류데이터 간의 시냅스를 조절하고 문제를 해결할 것이다.

#### 2 SVM

SVM은 이진선형분류를 실행할 때, 자료가 분류선과 지극히 근접할 경우, 근접한 각 데이터 간의 근접성을 평가하는 알고리즘이다. 두 개의 데이터가 분류선에서 각각의 다른 평면상에 존재할 때, 그 두 데이터, 양측으로 분류된 데이터 들 중 가장 가까운 데이터들 사이의 간격을 측정한다. 이 간격이 높을수록 오차 범위가 적고 좋은 분류라 할 수 있습니다. 이 알고리즘을 사용하여 우리는 각 데이터 간의 오차범위를 분석하고 알고리즘의 정확성을 높일 것이다.

#### 3 Logistic Regression

로지스틱 회귀는 종속 변수의 예측 결과가 여러 개 일 때 사용하는 분류이다.우리는 축구 경기의 승/무/패의 3개의 값을 예측하기 위하여 이 알고리즘을 사용할 것이다.

#### 4 K-Nearest Neighbor

K-최근접 이웃 알고리즘은, 분류선을 기준으로 서로 좌우로 근접한 두 이웃에게 가중치를 추가로 더하여 이들 값이 전체 평균에 더 큰 영향을 미치도록 하는 알고리즘이다. 이 알고리즘은 가장 간단한 기계학습 알고리즘으로 알려져 있다. 우리는 이 알고리즘을 통해 기계학습을 진행할 것이다.

#### 5 Naive Bayes

나이브 베이즈 알고리즘은 분류기를 만들 수 있는 간단한 기술입니다. 이것은 단일 알고리즘을 통한 훈련이 아닌 일반적인 원칙에 근거한 여러 알고리즘을 이용하여 훈련됩니다. 이 알고리즘은 무엇을 무엇이라고 특정하게 분류 가능하게 하는 특성들 가운데 연관성을 분석하고 각각 특성들이 한 객체에 독립적으로 기여하는 특성임을 학습시킨다. 이 알고리즘으로 우리는 승, 무, 패에서 독립적으로 기여하는 특성이 무엇인지 발견할 것이다.

#### 6 Decision Tree

결정 트리 학습법은 어떤 항목에 대한 관측값과 목표값을 연결시켜 주는 예측 모델이다. 이 알고리즘은 결정 트리를 사용한다. 결정트리는 의사 결정 과정과 결정된 의사를 시각적이고 명시적인 방법으로 보여주는데 사용한다. 또한, 자료 자체를 표현하는데도 사용할 수 있다. 우리는 결정 트리로 자료가 분류된 양상을 확인하고 자료를 시각화 할 것이다.

#### 7 Keras

Step 3

## 기대효과



Q. 이 프로젝트는 어떤 의미가 있는가?

- ▶ 현장의 감독들에게는 팀 경기력에 대한 정확한 평가를 바탕으로 전략적 작전을 구상할 수 있고 선수들에게는 쉽고 빠르게 자신들의 경기 기록요인별 평가 자료를 보면서 경기력을 향상시킬 수 있다.
- ▶ 사람들이 쉽고 간편하게 많은 양의 자료들을 정리해서 볼 수 있어서 축구의 인기가 높아질 수 있다.
- ▶ EPL 데이터로 학습시킨 알고리즘을 한국축구에 적용시킨다면 한국축구의 경기력에 대해 보다 정확히 결과를 예측할 수 있기 때문에, 구단 운영에 도움이 될 수 있다. 이를 통해 다시 한 번 2002년 월드컵 경기처럼 한국의 축구 선수팀들이 좋은 결과를 낸다면 한국의 경제도 긍정적인 양상을 보일 수 있다.



Step 4

## References



### 코드 출처

[AI:X 딥러닝 Final Assignment. 데이터를 통해 분석하는 유럽 축구 리그 | by 문봉준 / 예측 코드](#)

<https://www.sports.re.kr/pyxis-api/1/digital-files/c4783e68-99d8-4e4d-a18e-92124e9f7f5a>

예측 코드에 대한 설명 pdf

### 추가 자료

[https://gist.github.com/leeth5225/8b3c5f59df541088ea6ce7719be21737#file-epl\\_analysys-ipynb](https://gist.github.com/leeth5225/8b3c5f59df541088ea6ce7719be21737#file-epl_analysys-ipynb)

예측 코드

<https://www.premierleague.com/stats>

잉글랜드 프리미어리그의 모든 통계자료들

<https://1xbet.whoscored.com/Regions/252/Tournaments/2/England-Premier-League>

잉글랜드 프리미어리그의 통계자료를 보기 쉽게 정리해 놓은 사이트