



20220501

Participants	
Created By	
Created	@2022년 4월 24일 오후 9:22
Last Edited Time	@2022년 4월 26일 오전 8:34
Host	

1. Worldcupgoals

2. Son Heung-min's goals for TOT

3. Occupancy Detection

1. Worldcupgoals

첨부된 'Worldcupgoals' 데이터셋은 1930년부터 2018년까지 FIFA 월드컵에서 득점한 선수 관련 데이터입니다.

- 1) 골득점을 한 선수가 가장 많은 나라 상위 5개 국가와 그 선수 숫자를 출력하세요.
- 2) Brazil 국가의 월드컵 출전횟수가 5회 이상인 선수들을 내림차순으로 정렬하세요.

2. Son Heung-min's goals for TOT

첨부된 'Son Heung-min's goals for TOT' 데이터셋은 손흥민 선수의 Season 15/16 - 21/22 골 획득 관련 데이터입니다.

- 1) 시즌별 획득한 골의 개수를 내림차순으로 정렬하세요.
- 2) Goal을 가장 많이 넣은 시즌과 가장 적게 넣은 시즌의 Goal 개수 차이를 구하세요.



3. Occupancy Detection

첨부된 'Occupancy Detection' 데이터를 활용하여 사람의 재실 여부를 예측하는 모델을 만드세요.

- **Imputation**
 1. 결측치가 있는 변수를 확인한 뒤
 2. 결측치 변수를 첫 행부터 80%까지의 값의 중위수로 대체하세요.
- **Feature Engineering**
 1. 'date'변수를 제외한 모든 피처에 min-max scaling을 시행하세요.
- **Sampling**
 1. 타겟(Occupancy)에 불균형이 있는지 확인한 뒤
 2. 언더 샘플링 기법 중 단순 임의 추출법(Simple Random sampling)을 사용하여 불균형을 완화하세요.
- **Modeling**
 1. 나이브 베이즈 모델을 만드세요.
 - Test set: 30%
 - Target: **Occupancy**
 - Feature: 'date'를 제외한 모든 변수 활용
 - seed 번호: 2022
 - 변수설명 (ref. <https://elapser.github.io/projects/2019/02/21/Projects-Naver.html>)

date	날짜	년, 월, 일, 시, 분, 초
Temperature	사무실 온도	섭씨
Humidity	사무실 습도	%
Light	사무실 빛의 밝기	Lux
CO2	사무실 이산화탄소 농도	ppm
HumidityRatio	습도 비율	온도와 습도를 이용하여 kgwater-vapor/kg-air 방식으로 계산됨
Occupancy	사람 재실 여부	1: 재실중, 0: 사람 없는 상태