

Sports Seminar

OT 및 데이터 수집



2024.1.16

Table of Contents

01 OT

02 세미나 진행 순서

03 사이트 방문

04 데이터 크롤링 체험

스포츠(야구, 농구, 배구)에 대한 관심 정도

- 1) 스포츠는 잘 모른다.
- 2) 특정 종목에는 관심이 있지만, 팬은 아니다.
- 3) 특정 종목에 관심을 가지고 있으며 호감가는 팀도 있다.
- 4) 종목에 관심을 가지고 있고, 통계적으로 경기를 볼려는 관심을 가지고 있다.
- 5) 진지하게 스포츠 통계에 흥미를 가지며, 졸업 이후 스포츠 기업 및 구단으로 취업을 하고자 하는 계획을 가지고 있다.

스포츠(야구, 농구, 배구)에 사용하는 데이터 관심 정도

- 1) 전혀 모른다.
- 2) 일반적인 기록은 알고 있다. (1차 스탯 : ex. 평균자책점, 평균득점, 공격성공률)
- 3) 통계적으로 지표를 만들 수 있다는 점을 인지하고 있다. (2차 스탯을 알고 있는지)
- 4) 지표에 대한 간략한 설명 or 동작 분석이 필요한 종목의 경우 모션에 대한 이해도가 높다.

세미나에서 사용하게 될 종목 데이터



야구



배구

세미나에서 사용하게 될 종목 데이터

RunExpectancy (KBO07_11) baseball-in-play.com

base/out	0 out	1 out	2 out
0	0.533	0.292	0.111
1	0.890	0.560	0.243
2	1.247	0.739	0.350
3	1.477	1.056	0.395
12	1.566	0.980	0.479
13	1.855	1.227	0.534
23	1.984	1.457	0.626
123	2.419	1.650	0.829

야구



배구

1~4번째 세미나

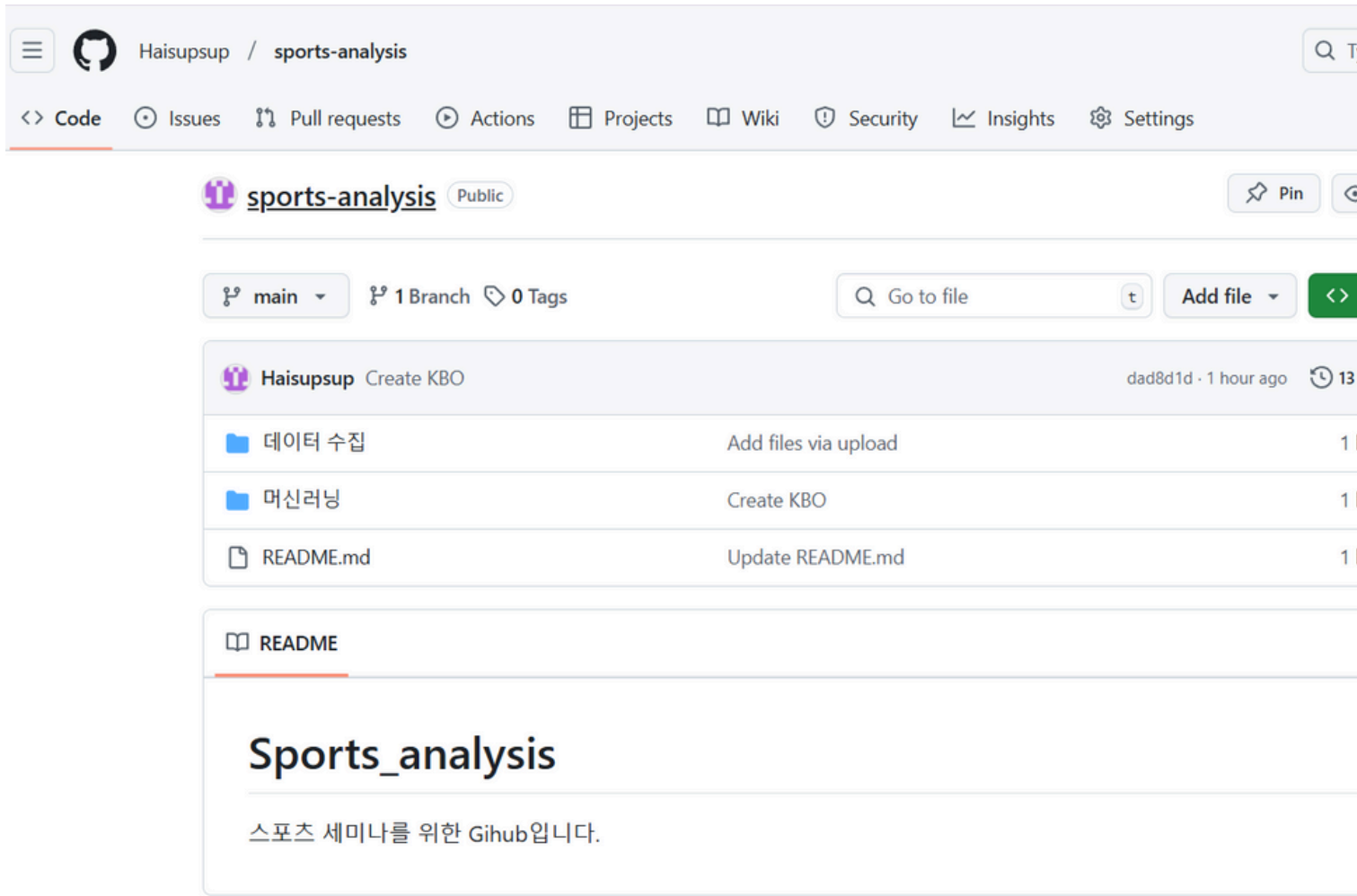
- 1) 데이터 사이트에 들어가서 데이터 수집&데이터 크롤링
- 2) 수집한 데이터로 기계학습을 이용한 KBO 팀 승률 예측 및 변수 파악
- 3) 컴퓨터 비전을 이용한 배구 동작인식(간단히)
- 4) 데이터 공부 or 상관관계 분석을 이용한 히트맵 그리기

5~8번째 세미나

- 1) 데이터 공부(야구위주 : 세이버메트릭스)
- 2) 기계학습을 이용해 세이버메트릭스 데이터 분석 및 모델 설정
- 3) 자체 학습 진행(운동종목은 자유, 필요한 데이터로 분석)
- 4) 발표 후 글 등록

- 1) matplotlib or seaborn을 이용한 데이터 시각화
- 2) YOLO(컴퓨터 비전)를 이용한 모션 동작 분석
- 3) OLS 모델(회귀분석)을 이용해 결정계수 높은 모델 제안하기
- 4) RandomForest를 이용한 feature 특징 찾기

참고 사이트



<https://github.com/Haisupsup/sports-analysis>
<https://blog.naver.com/tnv90>

MLB



fangraph



baseball reference



statcast

KBO



kbo



statiz



kbreport

배구



data volley 4



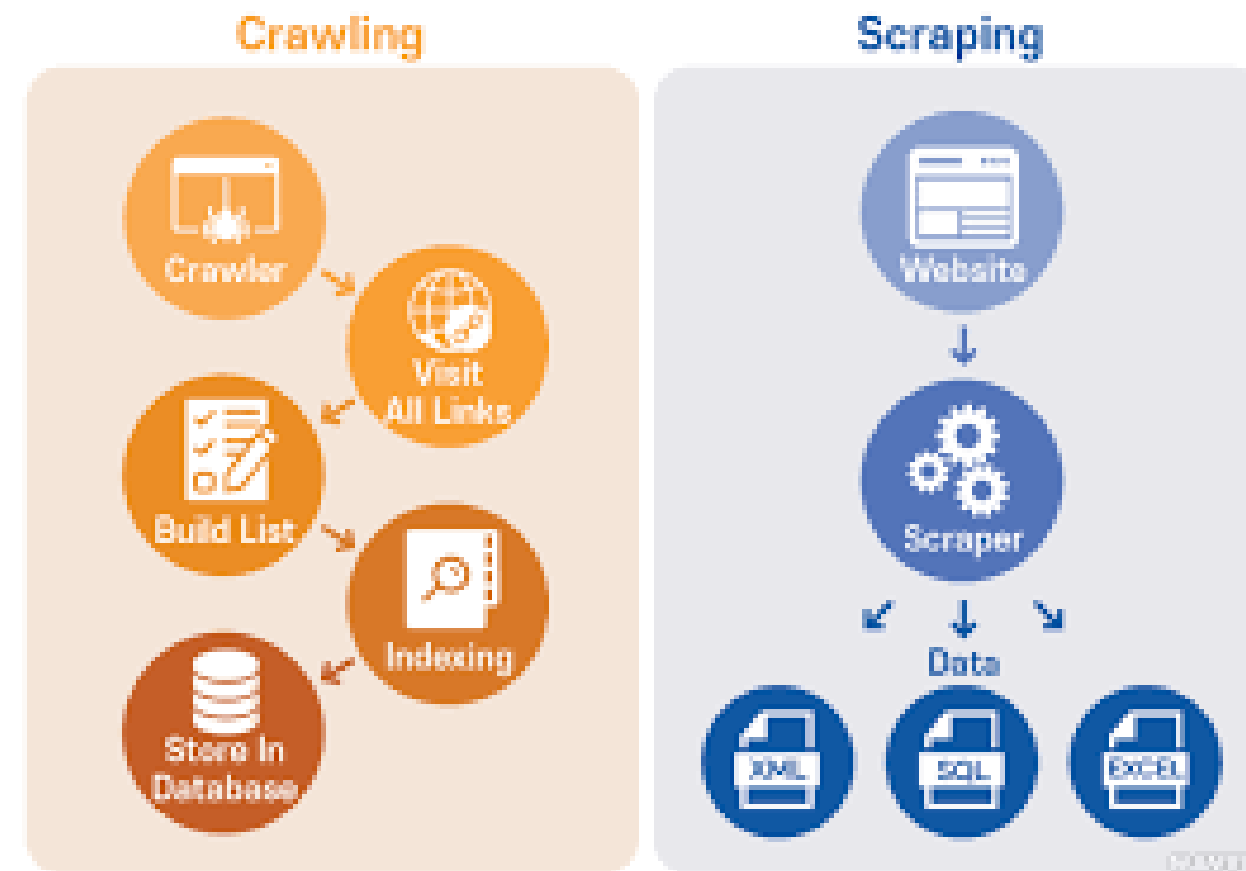
kovo

일부 사이트는 오픈 소스로 데이터를 제공하지 않음.

데이터를 수집할 수 있는 방법?

1. 돈을 지불해서 데이터를 구매(많은 돈을 지출하게 됨)
2. 사이트에 들어가서 데이터를 긁어오는 방법

2번의 방법으로 사이트의 데이터를 가져오는 방식을 “크롤링”이라고 한다



오늘 진행할 크롤링 사이트는 KBO, 스탯티즈 일부

1. chrome.exe를 설치해야 chrome 상에서 데이터 크롤링이 가능!

<https://googlechromelabs.github.io/chrome-for-testing/>

사이트에 들어가면 channel 중 Stable을 클릭.

[Chrome for Testing availability](#) 

This page lists the latest available cross-platform Chrome for Testing versions and assets per Chrome release channel.

Consult [our JSON API endpoints](#) if you're looking to build automated scripts based on Chrome for Testing release data.

Last updated @ 2025-01-14T08:09:58.964Z

Channel	Version	Revision	Status
Stable	131.0.6778.264	r 1368529	✓
Beta	132.0.6834.83	r 1381561	✓
Dev	133.0.6943.6	r 1402768	✓
Canary	134.0.6955.0	r 1405487	✓

2. 클릭 후 chromedriver를 설치.

<https://storage.googleapis.com/chrome-for-testing-public/131.0.6778.264/win64/chromedriver-win64.zip>

-> 설치 후 C드라이브 중 파일 하나를 만들어서 해당 파일에 넣어둡니다.
(이후로는 코드로 진행)

감사합니다.

정보통계학과 이현섭