

(۱) موضوع داده تفکیک توییت های استقلال و پرسپولیس می باشد و زبان داده فارسی است.

(۲) اینکه بتوانیم از روی حرف زدن و توییت گذاشتن فرد به صورت خودکار بفهمیم شخص مقابل استقلالیست یا پرسپولیسی برایم جالب است و در آمار گرفتن هم کمک می کند و ...

(۳) منابع داده :

پیوند

پیوند

(۴) با استفاده از scrape کردن اکانت های باشگاه های استقلال و پرسپولیس با بررسی HTML Tag های توییت. (ابتدا با استفاده از library پر کاربرد selenium کل صفحه HTML توییت را load کرده سپس با استفاده از BeautifulSoup تگ های مورد نظر را پیدا کردم و در آخر متن توییت آن ها را در فایل [teamname].txt ذخیره کردم.

روش استفاده از کد

قبل از اجرای این کد ها باید library های زیر را نصب کنید.

لینوکس

```
sudo pip install selenium
sudo pip install BeautifulSoup4
sudo pip install io
sudo pip install csv
sudo pip install time
```

ویندوز (powershell را به صورت administrator اجرا کنید).

```
py -3 -m pip install selenium
py -3 -m pip install BeautifulSoup4
py -3 -m pip install io
py -3 -m pip install csv
py -3 -m pip install time
```

سپس کافیه به پوشه ی مورد نظر رفته سپس دستور های زیر را run کنید :

لینوکس

```
python3 esteghlal_tweetter_scrapper.py
python3 perspolis_tweetter_scrapper.py
```

ویندوز (powershell را به صورت administrator اجرا کنید).

```
py -3 esteghlal_tweeter_scrapper.py  
py -3 perspolis_tweeter_scrapper.py
```

پس از اجرای این دستور ها دو تا فایل متنی با نام های esteghlal.txt و perspolis.txt ایجاد می شوند که اولی توییت های باشگاه استقلال هست و دومی توییت های باشگاه پرسپولیس.