

İstanbul İli Baraj Doluluk Tahmini Uygulaması (Web Kazıma Kısım)

Yazarlar :

Mehmet Kordon / Bilişim Sistemleri Mühendisliği / Kocaeli Üniversitesi / mehmetkordon09@gmail.com
Hakan Aytuğ Fırat / Bilişim Sistemleri Mühendisliği / Kocaeli Üniversitesi / hakanaytugfirat@gmail.com

Giriş ve Özet

Bu rapor İstanbul ilinin baraj doluluğunun tahmini uygulamasını yapmak için gereken veri toplama kısmını içermektedir. Belirlenen feature lar arasında ; ilçe bazlı su tüketim miktarı, ortalama rüzgar hızı, baraja düşen yağmur, barajların doluluk oranları, sıcaklık, nem, basınç, çiğlenme noktası ve deniz seviyesi basıncı gibi veriler vardır. Bu raporda bu verilerin otomatize bir şekilde lokale alınması konusu ele alınmıştır.

Introduction and Summary

This report covers the data collection process necessary to develop an application for estimating the reservoir levels in Istanbul. Among the selected features are data such as district-level water consumption, average wind speed, rainfall on reservoirs, reservoir occupancy rates, temperature, humidity, pressure, dew point, and sea level pressure. This report addresses the automated acquisition of these data into a local system.

Kullanılan Dil ve Kütüphaneler

Bu projede esnekliği ve kullanım kolaylığı açısından etkin bir dil olan Python tercih edilmiştir. Web sitelerinden ilgili verileri çekmek için Python'un Selenium kütüphanesini, çektiğimiz verileri lokale kaydetmek için de Pandas kütüphanesi kullandık.



Selenium Kütüphanesi

Selenium, web sayfalarıyla etkileşim kurmayı sağlayan popüler bir Python kütüphanesidir. Çoğunlukla, otomatik testler yazmak veya veri toplamak (web scraping) amacıyla kullanılır. Selenium, tarayıcı ile etkileşime geçmek için bir Web Driver arayüzünü kullanır. WebDriver, belirli bir tarayıcıya ait sürücüyü yükleyerek Selenium'un bu tarayıcıyı kontrol etmesini sağlar. Biz Chrome driver kullandık.

Kod Örnekleri:

İlk olarak web sitemizin url sini driver a atıyoruz

```
link=f"https://www.timeanddate.com/weather/turkey/istanbul/historic?month={ay}&year={yil}"  
driver.get(link)
```

Daha sonra bu web sitesinden çekmek istediğimiz tablodaki verilerin xpath ini f12 tuşuna basarak öğrendikten sonra bu xpath i(XML ve HTML belgelerindeki elementleri tanımlamak ve bulmak için kullanılan bir sorgulama dili) selenium kütüphanesinin find element metodunu kullanarak çektik

```
tablo = WebDriverWait(driver, 10).until(  
    EC.presence_of_element_located((By.XPATH,  
        "//table[contains(@class, 'zebra tb-wt fw tb-hover')]"))  
)
```

```
try:  
    dusuk_sicaklik = tablo.find_element(By.XPATH, ".//tr[2]/td[1]").text  
except NoSuchElementException:  
    dusuk_sicaklik = "Veri yok"
```

Verileri istediğimiz saniyede bir çekmesi için time sleep attık daha sonra her bir while döngüsünde bir listeye verileri kaydettik. Geriye pandas kütüphanesini kullanarak verileri indirmek kaldı

```
veriler.append([yil, ay, yuksek_sicaklik, yuksek_nem, yuksek_basinc,  
    dusuk_sicaklik, dusuk_nem, dusuk_basinc,  
    ortalama_sicaklik, ortalama_nem, ortalama_basinc])
```

Yıl	Ay	İksek Sıcaklık	Üçsek Nem	İksek Basınç	Düşük Sıcaklık	Düşük Nem	Düşük Basınç	Sıcaklık Ortalama	Nem Ortalama	Basınç Ortalama
2010	1	18 °C (1 Oc 100% (18 (1037 mbar -6 °C (26 Oc 42% (5 Oc 997 mbar 7 °C	76%	1016 mbar						
2010	2	19 °C (20 Ş 100% (1 Ş 1028 mbar -3 °C (6 Ş 43% (20 Ş 997 mbar 8 °C	76%	1011 mbar						
2010	3	20 °C (23 N 100% (1 M 1034 mbar -2 °C (18 N 19% (18 M 1000 mbar 9 °C	72%	1018 mbar						
2010	4	24 °C (23 N 100% (3 N 1025 mbar 6 °C (5 Nis 24% (4 Nis 1004 mbar 14 °C	67%	1016 mbar						
2010	5	30 °C (13 N 100% (2 M 1018 mbar 9 °C (1 Ma 20% (13 M 1001 mbar 19 °C	63%	1012 mbar						
2010	6	33 °C (16 T 100% (7 H 1015 mbar 15 °C (3 H 29% (1 Ha 1001 mbar 23 °C	67%	1010 mbar						
2010	7	33 °C (19 T 100% (9 T 1016 mbar 17 °C (9 Te 29% (7 Te 1001 mbar 26 °C	68%	1011 mbar						
2010	8	36 °C (17 A 94% (6 A 1020 mbar 22 °C (25 A 27% (29 A 1000 mbar 28 °C	64%	1011 mbar						
2010	9	31 °C (26 E 100% (20 E 1020 mbar 16 °C (1 Ey 26% (3 Ey 1002 mbar 22 °C	65%	1014 mbar						
2010	10	23 °C (18 E 100% (14 E 1029 mbar 7 °C (31 Ek 37% (4 Eki 1002 mbar 16 °C	72%	1017 mbar						
2010	11	26 °C (15 K 100% (1 K 1029 mbar 8 °C (3 Kas 31% (15 K 999 mbar 16 °C	76%	1015 mbar						
2010	12	22 °C (3 Ar 100% (27 / 1026 mbar -1 °C (12 A 43% (4 Ar 993 mbar 10 °C	71%	1014 mbar						

Verileri Pandas kütüphanesinin Data frame metodunu kullanıp yazdırmaya uygun hale geldikten sonra yukarıdaki gibi excelex ya da csv dosyasına yazdırdık.

```
df = pd.DataFrame(veriler, columns=["Yıl", "Ay", "Yüksek Sıcaklık", "Yüksek Nem", "Yüksek Basınç",  
                                "Düşük Sıcaklık", "Düşük Nem", "Düşük Basınç",  
                                "Ortalama Sıcaklık", "Ortalama Nem", "Ortalama Basınç"]  
df.to_excel("istanbul_weather_summary.xlsx", index=False)  
print("Veriler Excel dosyasına yazıldı.")
```

Kaynakça(Araştırma Kısmı)

Makale :

<https://www.mdpi.com/2071-1050/16/17/7696>

Selenium:

<https://medium.com/deep-learning-turkiye/selenium-ile-web-scraping-kazima-cb5cbb48c7bf>

Pandas:

https://python-land.translate.goog/data-processing/process-excel-data-in-python-and-pandas? x tr sl=en& x tr tl=tr& x tr hl=tr& x tr_pto=tc

Kaynakça (Veri Çekmek İçin)

Baraj doluluk oranları :

<https://data.ibb.gov.tr/dataset/istanbul-barajlari-gunluk-doluluk-oranlari/resource/af0b3902-cfd9-4096-85f7-e2c3017e4f21>

Basınç, nem sıcaklık, rüzgar hızı:

<https://www.timeanddate.com/weather/turkey/istanbul/historic?month=1&year=2014>

Baraja Düşen Yağmur Miktarı:

https://data.ibb.gov.tr/tr/dataset/istanbul-barajlarina-dusen-gunluk-toplam-yagis-miktari/resource/762b802e-c5f9-4175-a5c1-78b892d9764b?inner_span=True

Deniz Seviyesi Basıncı ve çiğlenme noktası :

<https://www.wunderground.com/history/weekly/tr/istanbul/LTBA/date/2020-1-14>