LAPORAN PRAKTIKUM

PRAKTIK PEMROGRAMAN PYTHON

**PRAKTIKUM**

**WEB SCARPPING**



**Disusun oleh :**

Aulia Diva Sukmadevi (V3923004)

**Dosen**

Yusuf Fadila Rachman. S.Kom., M.Kom

**PS D-III TEKNIK INFORMATIKA**

**SEKOLAH VOKASI**

**UNIVERSITAS SEBELAS MARET**

**2024**

1. **HASIL DAN PEMBAHASAN**

**INPUT**

* 1. **DATA GAMBAR**

****

Kode menggunakan library requests dan BeautifulSoup untuk melakukan web scraping dari halaman web https://proxyway.com/news. Pertama, kode mengambil konten HTML dari halaman tersebut menggunakan requests.get(url), lalu menguraikan konten HTML tersebut menggunakan BeautifulSoup untuk memudahkan ekstraksi informasi. Kemudian, kode mencari semua elemen <img> di halaman web dengan menggunakan soup.select('img'). Untuk setiap elemen gambar yang ditemukan, kode mengekstrak atribut src (URL sumber) dan alt (teks alternatif) dari elemen tersebut. Informasi gambar-gambar ini kemudian disimpan dalam bentuk dictionary dan ditambahkan ke dalam list images\_list.

Selanjutnya, kode mencetak informasi setiap gambar (URL sumber dan teks alternatif) dari images\_list menggunakan loop for. Terakhir, kode menyimpan informasi gambar-gambar ke dalam file CSV dengan nama 'data gambar.csv' menggunakan csv.writer

* 1. **DATA SUBJUDUL**

****

Kode menggunakan library requests dan BeautifulSoup untuk melakukan web scraping dari halaman https://proxyway.com/news. Tujuannya adalah untuk mengekstrak teks dari semua elemen <h2> (yaitu subjudul) pada halaman tersebut, kemudian menyimpan teks subjudul ke dalam file CSV dengan nama 'subjudul.csv'. Pertama, kode melakukan permintaan HTTP GET ke URL 'https://proxyway.com/news' menggunakan requests.get(url). Setelah mendapatkan respons dari permintaan tersebut, kode memeriksa status kode respons HTTP (r.status\_code). Jika respons adalah 200 (artinya permintaan berhasil), maka kode menggunakan BeautifulSoup untuk menguraikan konten HTML halaman tersebut.

Selanjutnya, kode mencari semua elemen <h2> di dalam konten HTML menggunakan soup.find\_all('h2'). Informasi teks dari setiap elemen <h2> kemudian diambil dan disimpan dalam list paragraphs. Kemudian, kode membuka file CSV dengan mode write ('w') menggunakan csv.writer untuk menulis teks subjudul dari paragraphs ke dalam file CSV. Setiap teks subjudul ditulis sebagai satu baris dalam file CSV. Terakhir, jika terjadi kesalahan dalam mengambil halaman (status kode respons HTTP bukan 200), kode mencetak pesan bahwa halaman tidak dapat diambil.

* 1. **DATA KETERANGAN**

****

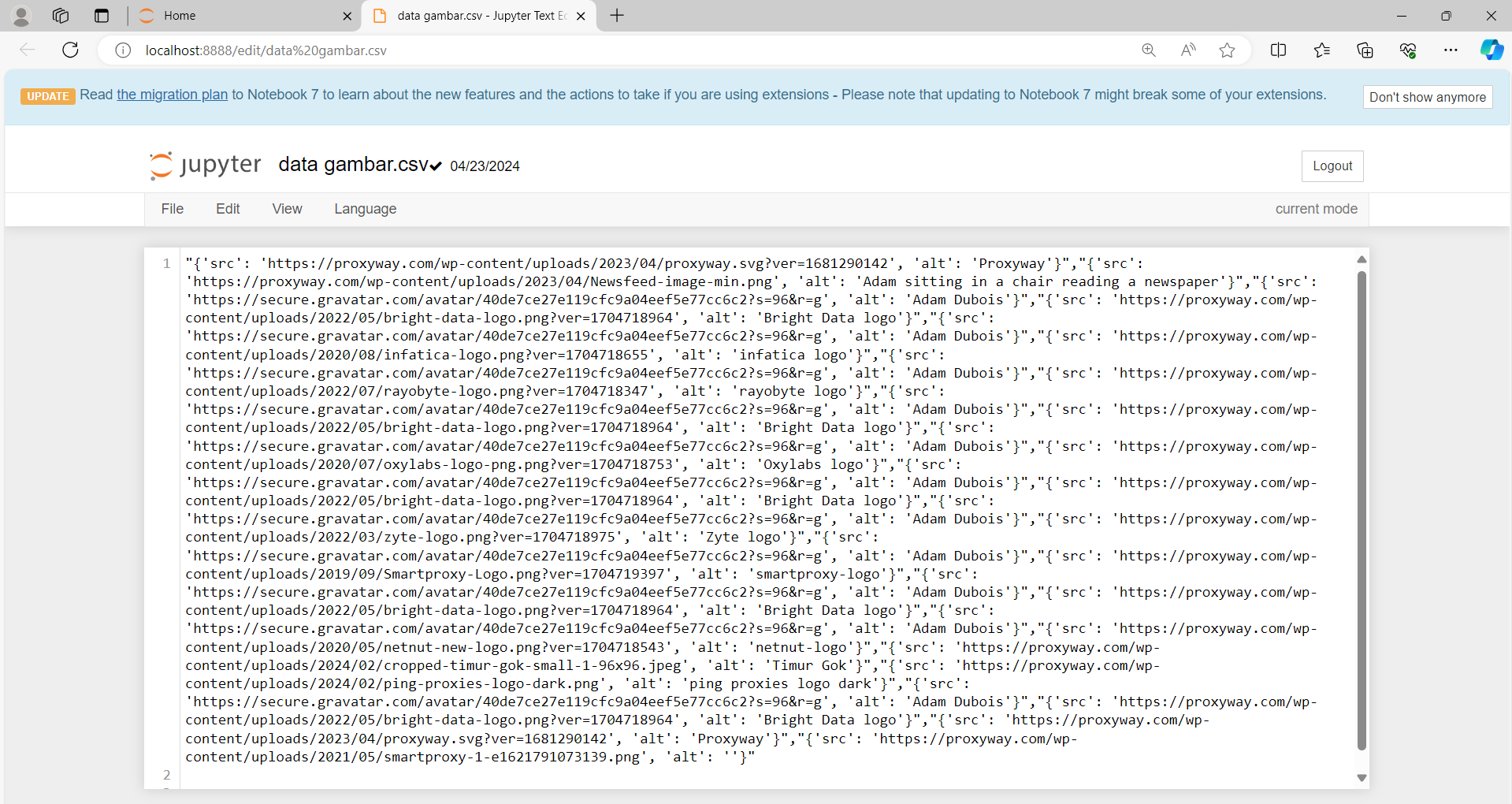
Kode digunakan untuk melakukan web scraping dari halaman https://proxyway.com/news dengan menggunakan library requests dan BeautifulSoup. Tujuan dari kode ini adalah untuk mengekstrak teks konten dari elemen <div> yang memiliki atribut khusus (data-widget\_type='theme-post-excerpt.default') pada halaman tersebut, dan kemudian menyimpan teks tersebut ke dalam file CSV dengan nama 'keterangan.csv'. Prosesnya dimulai dengan mengirimkan permintaan HTTP GET ke URL target menggunakan requests.get(url). Jika permintaan berhasil (status kode 200), maka konten HTML dari halaman tersebut diuraikan menggunakan BeautifulSoup. Selanjutnya, kode mencari semua elemen <div> yang memiliki atribut data-widget\_type='theme-post-excerpt.default' menggunakan soup.find\_all('div', attrs={'data-widget\_type': 'theme-post-excerpt.default'}). Setelah itu, untuk setiap elemen <div> yang ditemukan, kode mengekstrak teks konten dari elemen <div> dalamnya menggunakan .find('div', class\_='elementor-widget-container'). Jika teks konten berhasil ditemukan, maka teks tersebut ditambahkan ke dalam file CSV dengan menggunakan csv.writer.writerow([text\_content]). Selama proses ekstraksi dan penyimpanan, teks konten juga dicetak ke konsol dengan print(text\_content). Jika ada kesalahan dalam mengambil halaman (misalnya status kode HTTP bukan 200), maka pesan kesalahan akan dicetak di konsol.

**OUTPUT**

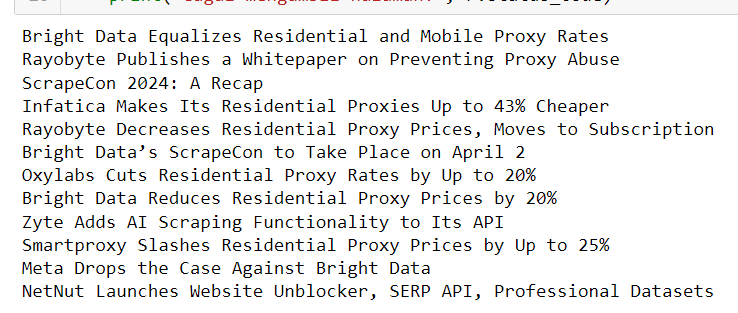
* + 1. **DATA GAMBAR**

****

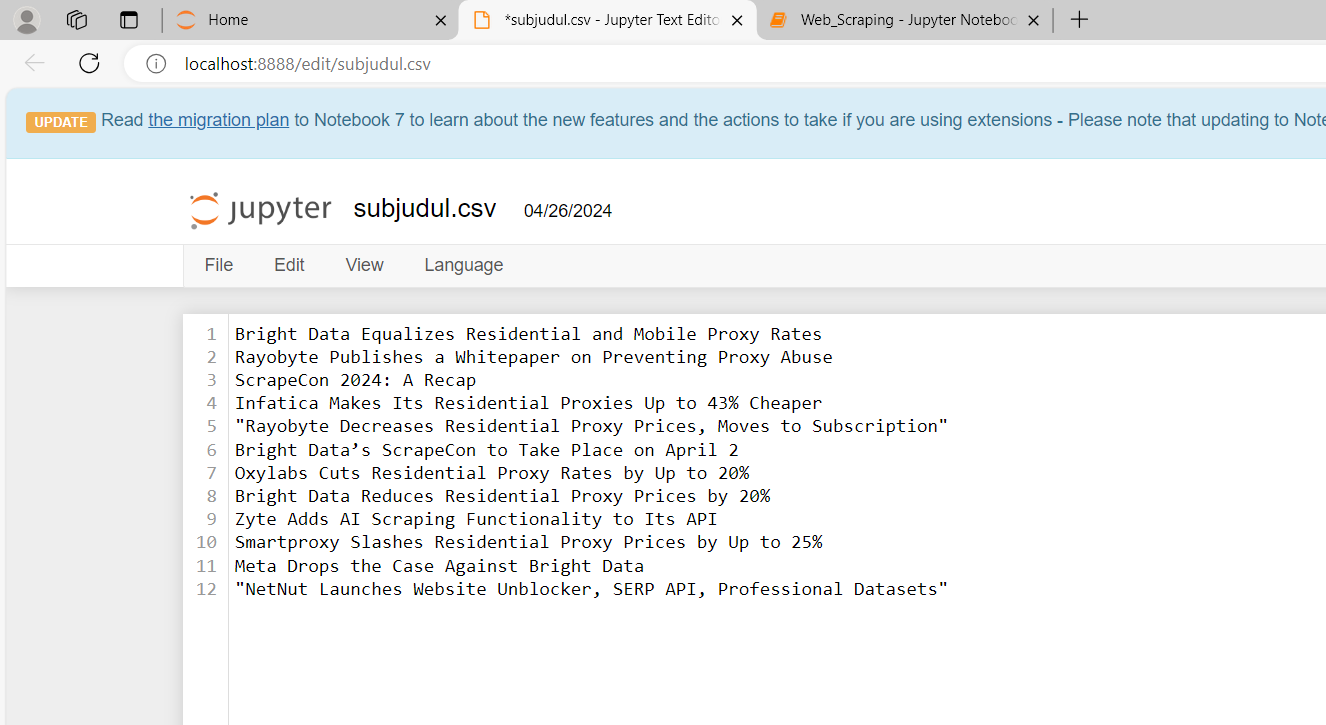
Dalam Bentuk CSV



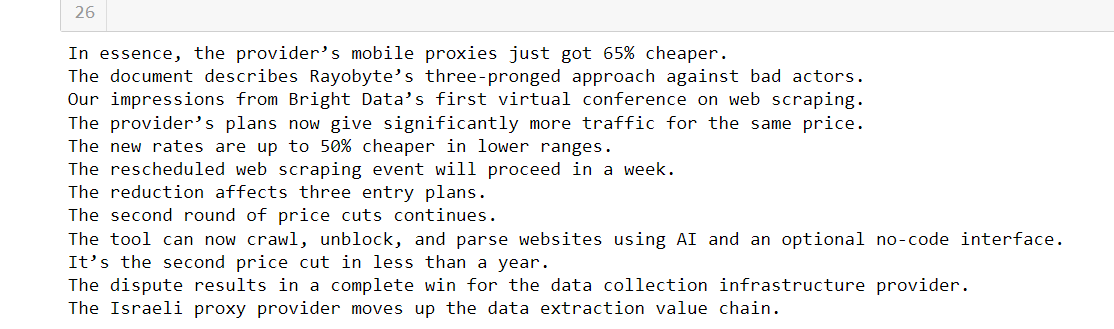
* + 1. **DATA SUBJUDUL**

****

Dalam bentuk CSV



* + 1. **DATA KETERANGAN**

****

Dalam Bentuk CSV

