

## **LAPORAN PRAKTIKUM DATA SCRAPING**

Disusun untuk Memenuhi Matakuliah Data Scraping  
Dibimbing oleh Pelsri Ramadar N.S. M.Kom



Oleh:

**Sidqi Ali Maghfur**

**1123102124**

**PROGRAM STUDI S1 TEKNIK INFORMATIKA  
SEKOLAH TINGGI ILMU KOMPUTER PGRI BANYUWANGI  
2024**

**LEMBAR PENGESAHAN**  
**LAPORAN PRAKTIKUM**

Matakuliah : Data Scraping  
Oleh : Sidqi Ali Maghfur  
NIM : 1123102124

**Telah disahkan pada**

Hari : Senin  
Tanggal : 2-12-2024

**Mengetahui/ Menyetujui :**  
Dosen Pengampu Mata Kuliah

**Pelsri Ramadar N.S. M.Kom**  
NIDN. 0704068408

## KATA PENGANTAR

Puji syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, sehingga saya dapat menyelesaikan tugas mata kuliah Data Scraping ini dengan baik. Mata kuliah ini dibimbing oleh Bapak Pelsri Ramadar N.S, M.Kom yang telah memberikan arahan dan ilmu yang sangat bermanfaat dalam memahami konsep dan implementasi data scraping.

Data scraping merupakan teknik yang sangat penting dalam era digital saat ini, di mana kebutuhan akan pengumpulan dan analisis data menjadi semakin crucial. Mata kuliah ini memberikan pemahaman mendalam tentang bagaimana mengekstrak data dari berbagai sumber web secara efisien dan etis, serta mengolahnya menjadi informasi yang berguna.

Dalam kesempatan ini, saya ingin menyampaikan terima kasih kepada:

1. Bapak Pelsri Ramadar N.S, M.Kom selaku dosen pengampu mata kuliah Data Scraping yang telah membimbing dan memberikan ilmu yang sangat berharga.
2. Seluruh teman-teman yang telah berpartisipasi aktif dalam diskusi dan berbagi pengetahuan selama perkuliahan.
3. Semua pihak yang telah membantu dalam penyelesaian tugas-tugas mata kuliah ini.

Saya menyadari bahwa dalam penyusunan tugas ini masih terdapat banyak kekurangan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan untuk perbaikan di masa mendatang.

Akhir kata, saya berharap ilmu yang didapatkan dalam mata kuliah Data Scraping ini dapat bermanfaat bagi pengembangan teknologi informasi dan dapat diimplementasikan dalam dunia kerja nantinya.

## DAFTAR ISI

Daftar Isi	Halaman
LEMBAR PENGESAHAN LAPORAN PRAKTIKUM.....	ii
KATA PENGANTAR .....	iii
DAFTAR ISI .....	iv
Modul Praktikum 2 .....	1
1.1 Capaian Praktikum Modul 2 .....	1
1.2 Indikator Capaian.....	1
1.3 Landasan Teori.....	1
1.4 Pelaksanaan Praktikum .....	2
1.4.1 Percobaan Pertama .....	2
a. Script / Setting Program .....	2
b. Langkah Uji coba.....	3
c. Hasil Uji Coba .....	4
d. Analisa Hasil.....	4
1.5 Kesimpulan .....	5
1.5.1 Kesimpulan Percobaan .....	5
BAB 2	
Penutup .....	8
2.1 Kesimpulan .....	8
2.2 Saran .....	8
DAFTAR PUSTAKA .....	9
LAMPIRAN.....	10

## DAFTAR GAMBAR

Daftar Gambar	Halaman
---------------	---------



# BAB 1

## MODUL 2 PRAKTIKUM DATA SCRAPPING

### 1.1 Capaian Praktikum Pertemuan 2

Memahami cara menggunakan Selenium untuk mengotomatisasi interaksi dengan browser web. Mempelajari cara melakukan web scraping untuk mengambil data dari halaman web. Mengimplementasikan teknik pengambilan data dari situs web dinamis yang memuat konten menggunakan JavaScript.

### 1.2 Indikator Capaian

- Berhasil membuka halaman Tokopedia dengan kata kunci pencarian.
- Mengambil sumber HTML halaman menggunakan Selenium.
- Memproses HTML dengan BeautifulSoup untuk menemukan elemen produk.
- Berhasil mengekstrak dan mencetak data produk (nama, harga, dan penjual) ke konsol.
- Menutup browser secara otomatis.

### 1.3 Landasan Teori

- Web Scraping: Teknik mengambil data dari situs web. Digunakan untuk mencari elemen HTML tertentu (produk, harga, penjual) dengan BeautifulSoup.
- Automasi Browser: Selenium menjalankan browser secara otomatis untuk menangani halaman yang memuat konten dinamis menggunakan JavaScript.
- Parsing HTML: BeautifulSoup memproses HTML hasil Selenium, memudahkan pencarian elemen menggunakan struktur DOM.
- Query String: URL dengan parameter pencarian (?q=klembe banyuwangi) digunakan untuk mengakses data produk.
- Etika Penggunaan: Scraping harus sesuai dengan kebijakan situs dan dilakukan dengan tujuan yang sah..

### 1.4 Pelaksanaan Praktikum

#### 1.4.1 Percobaan

Pada percobaan ini kita melakukan web scraping pada website [www.tokopedia.com](http://www.tokopedia.com) untuk mengambil data 'klembe banyuwangi' Dengan menggunakan selenium.

#### a. Script / Setting Program

- **tugas.py**

```

• from bs4 import BeautifulSoup
• from selenium import webdriver
• import time
•
• URL = 'https://www.tokopedia.com/search'
•
• params = {
•     'q': 'klemben banyuwangi'
• }
•
• driver = webdriver.Chrome()
• fullURL = f"{URL}?q={params['q']}"
• driver.get(fullURL)
•
• time.sleep(10)
• html = driver.page_source
• soup = BeautifulSoup(html, 'html.parser')
•
• data = soup.find_all('div', {'class': 'css-5wh65g'})
•
• for i in range(len(data)):
•     nama_produk = data[i].find('span', {'class': '_0T8-
iGxMpV6NEsYEhwkqEg=='})
•     harga = data[i].find('div', {'class':
'_67d6E1xDKIzw+i2D2L0tjw=='})
•     penjual = data[i].find('span', {'class': 'T0rpy-LEwYNQifsgB-
3SQw== pC8DMVkBZGW7-egObcwMFQ== flip'})
•     if nama_produk and harga and penjual :
•         print("nama produk: "+nama_produk.text)
•         print("harga: "+harga.text)
•         print("penjual: "+penjual.text)
•
• print(data)
• driver.quit()

```

**b. Langkah Uji Coba**

**Persiapkan alat:**

- Instal pustaka selenium dan beautifulsoup4.
- Unduh dan pasang ChromeDriver yang sesuai dengan versi browser Chrome Anda.

**Jalankan kode:**

- Simpan kode ke file Python (misalnya, scraper.py) dan jalankan dengan python scraper.py.

**Periksa URL:**

- Pastikan URL yang dibentuk (fullURL) sesuai dengan pencarian di Tokopedia.

#### Cek elemen HTML:

- Gunakan inspect element di browser untuk memeriksa apakah kelas CSS yang digunakan (css-5wh65g, dll.) sesuai dengan yang ada di halaman Tokopedia.

#### Ubah kata kunci:

- Coba ganti kata kunci di params['q'] untuk memeriksa hasil pencarian lain.

#### Periksa output:

- Nama produk, harga, dan penjual harus tercetak. Jika tidak, cek struktur HTML atau tambahkan waktu tunggu lebih lama agar halaman termuat penuh.

### c. Hasil Uji Coba

```
PS C:\Users\hp\OneDrive\WORD\DATA SCRAPING> & C:/Users/hp/AppData/Local/Programs/Python/Python313/python.exe "c:/Users/hp/OneDrive/WORD/DA
TA SCRAPING/tugas.py"

DevTools listening on ws://127.0.0.1:54796/devtools/browser/9f861fea-7f40-4a49-b432-6cf36c405586
nama produk: Kue Klemben Original Oleh Oleh Khas Banyuwangi 250 gr
harga: Rp10.000
penjual: ANISA KUE BANYUWANGI
nama produk: Bolu kering istimewa roti kue kering klemben camilan khas banyuwangi
harga: Rp20.900
penjual: ADA2001shop
nama produk: Bolu Kering Klemben Bolu Jadul Bolu Kijing Kue Jadoel Tweety Khas Banyuwangi
harga: Rp18.000
penjual: Cherry2810
nama produk: kk_ Kue Klemben Original Oleh Oleh Khas Banyuwangi 250 gr
harga: Rp13.600
penjual: Azarein99
nama produk: Bolu klemben bolu kering jadul khas Banyuwangi
harga: Rp64.500
penjual: Bake my Day Tebet
nama produk: Kue Bolu Klemben Pandan Kering khas Banyuwangi Bolu Kering Jadul Snack
harga: Rp19.722
penjual: Al Fath Herba dan Carica Khas Dieng
nama produk: Kue Bolu Klemben Kering khas Banyuwangi Snack Camilan Jadul Lebaran
harga: Rp19.722
penjual: Al Fath Herba dan Carica Khas Dieng
nama produk: Bolu klemben/bolu kenong/bolu jadul khas Banyuwangi packing plastik mika.
harga: Rp35.000
penjual: Bake my Day Tebet
nama produk: Bolu Klemben Khas Banyuwangi Dua Mandiri
harga: Rp20.000
penjual: UMMK DUA MANDIRI
nama produk: Savage| Kue Klemben Original Oleh Oleh Khas Banyuwangi 250 gr
harga: Rp19.000
```

### d. Analisa Hasil

#### Tujuan Kode:

- Kode ini digunakan untuk mengambil data produk (nama, harga, dan penjual) dari halaman pencarian Tokopedia berdasarkan kata kunci klemben banyuwangi.

#### Hasil yang Diharapkan:

- Nama produk, harga, dan penjual tercetak di konsol.
- Informasi lengkap produk akan diambil dan disimpan dalam variabel data.

#### Kemungkinan Hasil dan Analisisnya:



a. Data Berhasil Diambil:

- Informasi seperti nama produk, harga, dan penjual ditampilkan di konsol.
- Artinya, kelas CSS yang digunakan dalam kode cocok dengan struktur HTML halaman Tokopedia.

b. Data Tidak Ditemukan (data Kosong):

- Penyebab: Struktur HTML halaman Tokopedia berubah, sehingga kelas CSS yang digunakan tidak lagi sesuai.
- Solusi: Gunakan inspect element di browser untuk menemukan kelas CSS terbaru dan sesuaikan kode.

c. Error Saat Menjalankan Kode:

- ChromeDriver tidak cocok: Pastikan ChromeDriver sesuai dengan versi browser Chrome.
- Timeout: Halaman belum termuat sepenuhnya dalam 10 detik. Solusi: Tambahkan waktu tunggu atau gunakan Selenium wait.

**Catatan untuk Hasil Kosong atau Salah:**

Tambahkan debugging untuk memeriksa elemen HTML:

python

Salin kode

```
print(len(data)) # Memeriksa jumlah elemen yang ditemukan
for item in data:
    print(item.prettify()) # Menampilkan struktur HTML elemen
```

**Kesimpulan:**

- Jika kode berhasil, informasi produk tercetak.
- Jika gagal, kemungkinan besar kelas CSS yang dicari tidak sesuai dengan struktur HTML terkini halaman Tokopedia. Debugging diperlukan untuk menyesuaikan kode dengan struktur terbaru..

## 1.5 Kesimpulan

### 1.5.1 Kesimpulan Percobaan

**Hasil Berhasil:**

Jika nama produk, harga, dan penjual tercetak di konsol:

- Kode berhasil mengekstrak informasi produk dari Tokopedia berdasarkan kata kunci "klemben banyuwangi".
- Halaman termuat dengan benar, dan kelas CSS yang digunakan sesuai dengan struktur HTML halaman.

#### **Hasil Gagal:**

Jika data kosong atau tidak ada informasi yang tercetak:

- Struktur HTML halaman mungkin telah berubah, sehingga kelas CSS yang digunakan tidak cocok.
- Halaman belum sepenuhnya termuat dalam waktu 10 detik, menyebabkan data tidak ditemukan.
- Query pencarian mungkin tidak menghasilkan produk.

#### **Analisis Debugging:**

- Jika kode gagal, gunakan debugging untuk memeriksa:
- Apakah data kosong.
- Apakah kelas CSS untuk nama produk, harga, atau penjual sudah sesuai.
- Tambahkan waktu tunggu jika halaman membutuhkan waktu lebih lama untuk memuat.

#### **Kesimpulan Akhir:**

- Kode ini mampu melakukan scraping data produk dari Tokopedia jika kelas CSS dan waktu pemuatan halaman sesuai.
- Jika gagal, diperlukan pembaruan kode sesuai dengan perubahan struktur HTML halaman atau lingkungan runtime.

**Mengetahui:**

Dosen Pengampu Mata Kuliah

**Pelsri Ramadar N.S. M.Kom**

NIDN. 0704068408

**\*NB. Halaman ini dibuat sebanyak pertemuan yang diberikan**

## **BAB 2**

### **PENUTUP**

#### **2.1 Kesimpulan**

kode ini berhasil mengambil data produk dari Tokopedia menggunakan Selenium dan BeautifulSoup, dengan pencarian untuk "klemben banyuwangi". Jika berhasil, nama produk, harga, dan penjual akan dicetak. Jika tidak ada data, kemungkinan kelas CSS yang digunakan sudah berubah atau halaman membutuhkan waktu lebih lama untuk memuat. Pembaruan kode sesuai struktur HTML atau menambah waktu tunggu bisa diperlukan untuk memastikan hasil yang diinginkan.

#### **2.2 Saran**

- Gunakan WebDriverWait daripada `time.sleep()` untuk menunggu elemen yang dimuat secara dinamis.
- Periksa secara berkala apakah struktur HTML halaman berubah, terutama kelas CSS yang digunakan.
- Tambahkan pengecekan validitas data sebelum mencetak, dan gunakan try-except untuk penanganan error.
- Pastikan untuk selalu menutup browser dengan `driver.quit()` setelah selesai.
- Lakukan scraping secara etis dan sesuai dengan kebijakan situs web yang diakses.

## DAFTAR PUSTAKA

### 1.BeautifulSoup:

- Richardson, L. (2007). BeautifulSoup: A Python Library for Pulling Data Out of HTML and XML Files. Available at:  
<https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

### 2.Selenium:

- Burnette, E. (2016). Automate the Boring Stuff with Python: Practical Programming for Total Beginners. No Starch Press.
- Selenium Contributors. (n.d.). Selenium Documentation. Retrieved from <https://www.selenium.dev/documentation/en/>

### 3.Web Scraping:

- Mitchell, R. (2018). Web Scraping with Python: Collecting Data from the Modern Web. O'Reilly Media.
- Pradeep, S. (2017). Python Web Scraping. Packt Publishing.

### 4.Time Module:

- Python Software Foundation. (2024). time — Time access and conversions. Python 3 documentation. Retrieved from <https://docs.python.org/3/library/time.html>

### 5.Selenium WebDriver:

- Selenium Contributors. (2024). WebDriver API. Retrieved from <https://www.selenium.dev/documentation/en/webdriver/>

## **LAMPIRAN**