

**LAPORAN TEORI**  
**PEMROSESAN DATA MINGGU KE-12**

Web Scraping & Data Crawling

Dosen Pengampu: Edi Satriyanto S.Si., M.Si.



Oleh:

Bayu Kurniawan  
(3322600019)

**PROGRAM STUDI D4 SAINS DATA TERAPAN**  
**DEPARTEMEN TEKNIK INFORMATIKA DAN KOMPUTER**  
**POLITEKNIK ELEKTRONIKA NEGERI SURABAYA**  
**2023**

1. Lakukan scrapping produk pakaian di toped dan shopee (kriteria produk yg dicari sama atau serupa)

Untuk web scraping ini saya menggunakan produk Kaos Kessoku Band di tokopedia dan shopee

```
import time
from selenium import webdriver
from bs4 import BeautifulSoup as soup
import pandas as pd
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.common.by import By

url = 'https://www.tokopedia.com/search?q=sepatu+adidas&source=universe&st=product&srp_component_id=02.02.02.01'
driver = webdriver.Chrome()
driver.get(url)

WebDriverWait(driver, 5).until(EC.presence_of_all_elements_located((By.CSS_SELECTOR, '#zeus-root'))))
time.sleep(2)
for i in range(21):
    driver.execute_script('window.scrollTo(0, 250)')
    time.sleep(1)

driver.execute_script('window.scrollTo(50, 0)')
time.sleep(1)
soup1 = soup(driver.page_source, 'html.parser')

data = []
for item in soup1.findAll('div', class_='css-974ipl'):
    nama_produk = item.find('div', class_='prd_link-product-name css-3um8ox').text
    harga_produk = item.find('div', class_='prd_link-product-price css-1ksb19c').text

    rtg = item.findAll('span', class_='css-t70v7i')
    tj1 = item.findAll('span', class_='css-1duhs3e')
    if len(rtg) > 0:
        terjual = item.find('span', class_='css-1duhs3e').text
    else:
        terjual = '-'

    for item2 in item.findAll('div', class_='css-1rn0ir1'):
        lokasi = item2.findAll('span', class_='css-1kdc32b')[0].text

        data.append((nama_produk, harga_produk, terjual, lokasi))
toped = pd.DataFrame(data, columns=['Nama Barang', 'Harga', 'Terjual', 'Lokasi'])
toped.to_csv('C:/Users/bayuk/OneDrive/Documents/AI/pens/smtr2/Pemrosesan Data/12/toped.csv', index=False)
print('data telah tersimpan')
print(toped)
driver.close()
```

✓ 34.3s

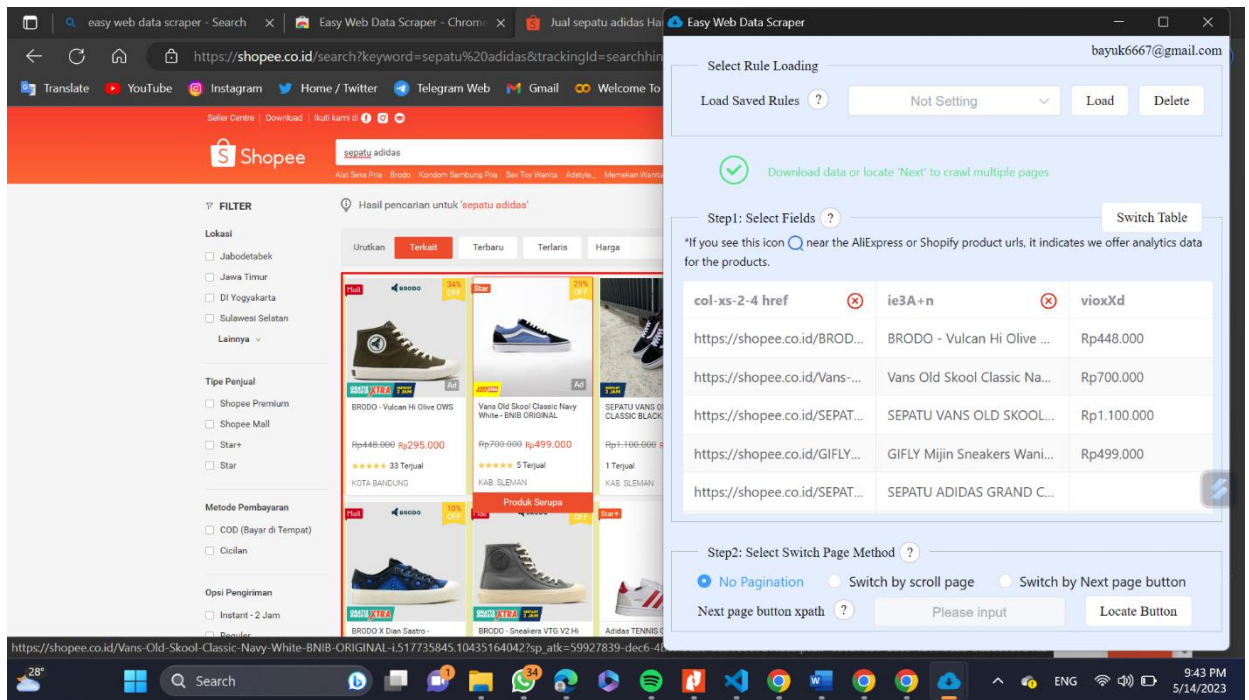
data telah tersimpan

	Nama Barang	Harga	\
0	SEPATU ADIDAS SPEZIAL SAMBA GRADE ORIGINAL 100...	Rp179.000	
1	SEPATU PRIA ADIDAS ZX 750 GREY WHITE ORIGINAL ...	Rp399.000	
2	sepatu pria / sepatu lari pria murah / sepatu ...	Rp249.000	
3	SALE TERMURAH !! ADIDAS SAMBA X PW FULL BLACK ...	Rp375.000	
4	Sepatu Sneakers Olahraga Pria Adidas Zoom Flyk...	Rp270.000	
..	...	...	
75	SEPATU CASUAL ADIDAS GAZELLE NAVY WHITE ORIGINAL	Rp299.000	
76	SEPATU ADIDAS GAZELLE SPEZIAL HANDBALL SUEDE S...	Rp90.000	
77	sepatu adidas beckenbauer allround black white...	Rp349.000	
78	SEPATU ADIDAS ZX RACER BLACK WHITE ORIGINAL - ...	Rp285.000	
79	Sepatu Sneakers Adidas City Canvas - HQ6929	Rp900.000	

	Terjual	Lokasi
0	Terjual 500+	Kab. Tangerang
1	Terjual 12	Kab. Tangerang
2	-	Kab. Tangerang
3	-	Kab. Tangerang
4	Terjual 7	Jakarta Barat
..	...	...
75	Terjual 4	Kab. Tangerang
76	Terjual 8	Kab. Tangerang
77	-	Tangerang
78	Terjual 22	Kab. Tangerang
79	-	Bandung

[80 rows x 4 columns]

Untuk web scrap pada tokopedia saya menggunakan koding dengan import bs4, pandas, time, dan selenium. Setelah scraping kemudian saya menyimpan hasilnya dalam bentuk csv dengan nama toped.csv



Untuk scraping pada shopee saya menggunakan extension untuk chromium edge karena saat saya melakukan web scraping shopee dengan koding terdapat error login false. Setelah scraping, hasilnya saya simpan dalam bentuk csv dengan nama shopee.csv

## 2. Lakukan analisa berdasarkan harga dll dari hasil scrapping anda secara statistik

Hal pertama yang saya lakukan adalah merapikan data csv tiap file dengan menyamakan jumlah data dan mengubah type data harga sehingga type data yang sama.

### Tokopedia



## Shopee

```

datashopee = pd.read_csv('C:/Users/bayuk/OneDrive/Documents/AI/pens/smt2/Pemrosesan Data/12/shopee.csv')
datashopee = datashopee.fillna(0)
datashopee = datashopee.rename(columns={'ie3A+n': 'Nama Barang', 'ZEgDH9': 'Harga', 'r6HknA': 'Terjual', 'zGGwiV': 'Lokasi'})
datashopee

```

	col-xs-2-4 href	Nama Barang	vioxXd	Harga	Terjual	Lokasi
0	https://shopee.co.id/BRODO-Vulcan-Hi-Olive-OWS...	BRODO - Vulcan Hi Olive OWS	Rp448.000	295.000	33 Terjual	KOTA BANDUNG
1	https://shopee.co.id/Vans-Old-Skool-Classic-Na...	Vans Old Skool Classic Navy White - BNIB ORIGINAL	Rp700.000	499.000	5 Terjual	KAB. SLEMAN
2	https://shopee.co.id/SEPATU-VANS-OLD-SKOOL-CLA...	SEPATU VANS OLD SKOOL CLASSIC BLACK WHITE RESM...	Rp1.100.000	999.999	1 Terjual	KAB. SLEMAN
3	https://shopee.co.id/GIFLY-Mijin-Sneakers-Wani...	GIFLY Mijin Sneakers Wanita Sepatu Sport Shoes...	Rp499.000	169.000	288 Terjual	KAB. TANGERANG
4	https://shopee.co.id/SEPATU-ADIDAS-GRAND-COURT...	SEPATU ADIDAS GRAND COURT SE FULLWHITE ORIGINA...		0 280.000	126 Terjual	KOTA TANGERANG SELATAN
5	https://shopee.co.id/BRODO-X-Dian-Sastro-Sneak...	BRODO X Dian Sastro - Sneakers VTG V2 Wanita C...	Rp549.000	494.000	56 Terjual	KOTA BANDUNG
6	https://shopee.co.id/BRODO-Sneakers-VTG-V2-Hi-...	BRODO - Sneakers VTG V2 Hi Limestone Grey IS	Rp688.000	429.000	86 Terjual	KOTA BANDUNG
7	https://shopee.co.id/Adidas-TENNIS-Grand-Court...	Adidas TENNIS Grand Court SE Shoes Pria Putih ...	Rp1.100.000	495.000	3 Terjual	KOTA YOGYAKARTA
8	https://shopee.co.id/Adidas-TENNIS-Sepatu-Cour...	Adidas TENNIS Sepatu Court Lifestyle Vegan Nov...	Rp1.100.000	407.000	9 Terjual	KOTA YOGYAKARTA
9	https://shopee.co.id/ADIDAS-TENNIS-H02062-GRAN...	ADIDAS TENNIS H02062 GRAND COURT		0 540.750	4 Terjual	KOTA SURABAYA
10	https://shopee.co.id/BRODO-Vantage-V2-Hi-Natur...	BRODO - Vantage V2 Hi Natural White WS	Rp445.000	375.000	3RB Terjual	KOTA BANDUNG
11	https://shopee.co.id/Adidas-TENNIS-Grand-Court...	Adidas TENNIS Grand Court Floudfoam Lifestyle ...	Rp1.100.000	484.000	11 Terjual	KOTA YOGYAKARTA
12	https://shopee.co.id/adidas-RUNNING-Sepatu-Gal...	adidas RUNNING Sepatu Galaxy 6 GW4141	Rp750.000	637.500	180 Terjual	KOTA BEKASI
13	https://shopee.co.id/Adidas-SKATEBOARDING-Sepa...	Adidas SKATEBOARDING Sepatu Lifestyle Modern Fu...	Rp1.000.000	490.000	3 Terjual	KOTA YOGYAKARTA
14	https://shopee.co.id/adidas-RUNNING-Sepatu-Sho...	adidas RUNNING Sepatu Showtheway 2.0 GX1707		0 800.000	133 Terjual	KOTA BEKASI

Untuk scraping shopee karena kolom datanya masih bernama class dari div maka saya menggantinya agar sesuai.

```

# perbandingan harga
print('Tokopedia'.center(27, '='))
print(f"Harga tertinggi\t: Rp{round(max(datatoped['Harga']))}.000")
print(f"Harga terendah\t: Rp{round(min(datatoped['Harga']))}.000\n")
print('Shopee'.center(27, '='))
print(f"Harga tertinggi\t: Rp{round(max(datashopee['Harga']))}.000")
print(f"Harga terendah\t: Rp{round(min(datashopee['Harga']))}.000")

```

```

=====Tokopedia=====
Harga tertinggi : Rp4899.000
Harga terendah  : Rp25.000

=====Shopee=====
Harga tertinggi : Rp1000.000
Harga terendah  : Rp169.000

```

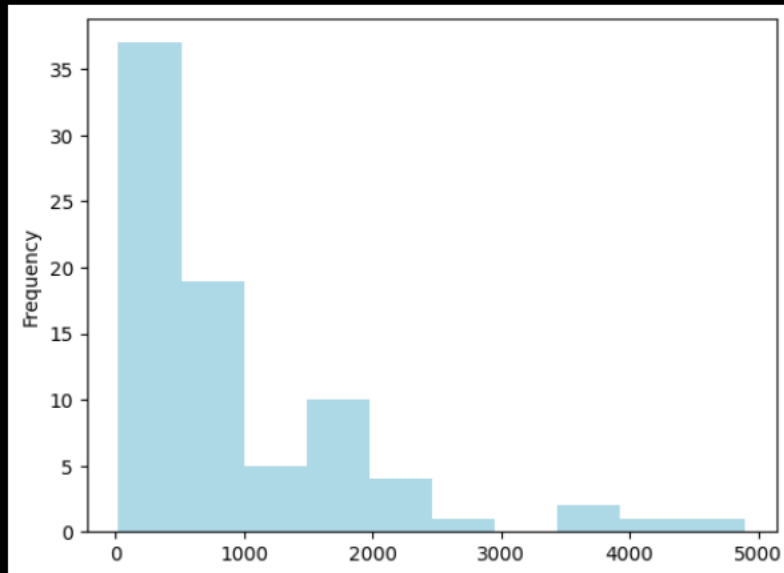
Untuk Harga tertinggi dan terendah berada di tokped dengan harga tertinggi sebesar 4,899 jt rupiah dan harga terendah sebesar 25 ribu rupiah

```
meantoped = datatoped['Harga'].mean()
print(f"Mean harga sepatu adidas di Tokopedia adalah Rp{round(meantoped)}.000")
datatoped['Harga'].plot.hist(color = 'lightblue')
```

✓ 0.3s

Mean harga sepatu adidas di Tokopedia adalah Rp949.000

<Axes: ylabel='Frequency'>

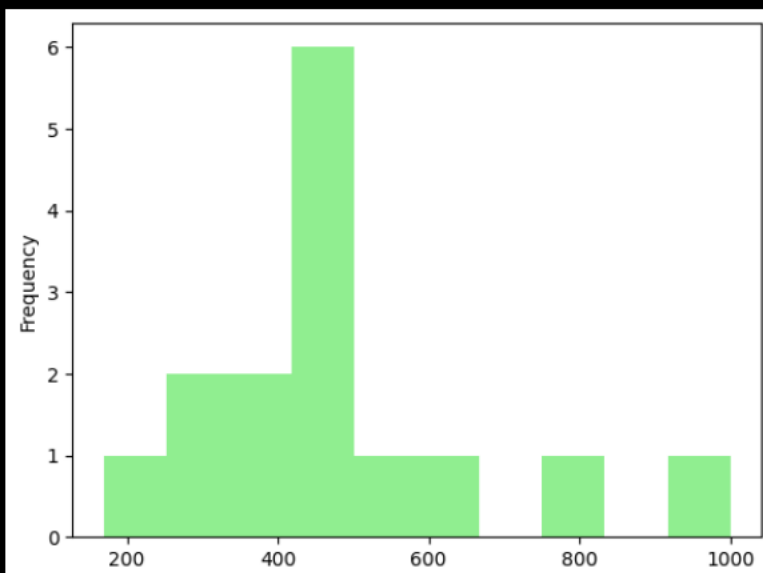


```
meanshopee = datashopee['Harga'].mean()
print(f"Mean harga sepatu adidas di Shopee adalah Rp{round(meanshopee)}.000")
datashopee['Harga'].plot.hist(color = 'lightgreen')
```

✓ 0.3s

Mean harga sepatu adidas di Shopee adalah Rp493.000

<Axes: ylabel='Frequency'>



Untuk rata-rata harga di tokopedia sebesar 949 ribu dan pada shopee sebesar 493 ribu.

```
# membandingkan harga dengan mean
if meantoped > meanshoopee:
|   print('Harga di Shopee lebih murah')
elif meantoped < meanshoopee:
|   print('Harga di Tokopedia lebih murah')
else:
|   print('Harga di kedua market place sama')
✓ 0.0s
```

Harga di Shopee lebih murah

Jika dilihat dari rata-rata harga, harga barang Kaos Kessoku Band di shopee cenderung lebih murah.

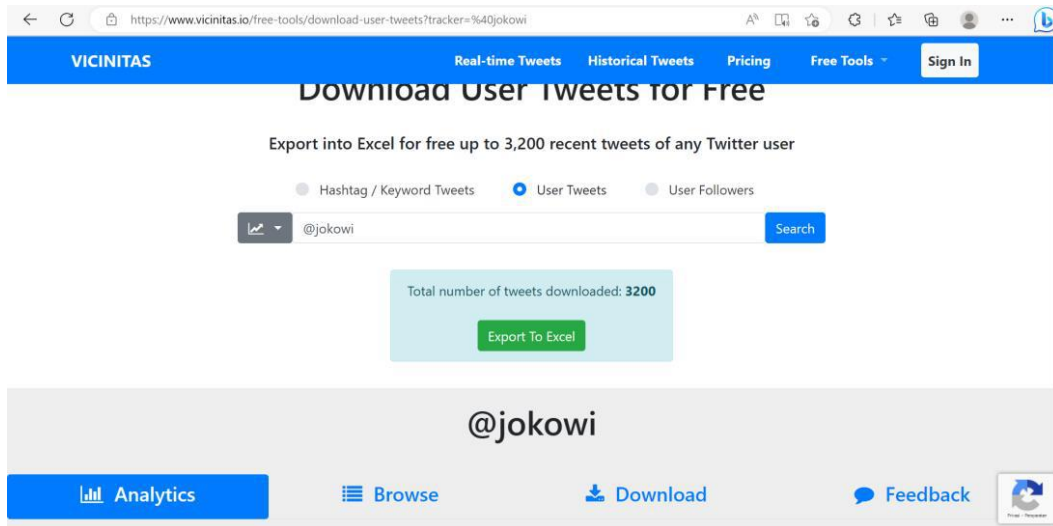
```
# kesesuaian kata kunci
keyword_count = datatoped['Nama Barang'].str.contains('Adidas').sum()
print('Kata kunci muncul pada Tokopedia sebanyak ', keyword_count, 'kali')
keyword_count2 = datashoopee['Nama Barang'].str.contains('Adidas').sum()
print('Kata kunci muncul pada Tokopedia sebanyak ', keyword_count2, 'kali')
✓ 0.0s
```

Kata kunci muncul pada Tokopedia sebanyak 53 kali  
Kata kunci muncul pada Tokopedia sebanyak 4 kali

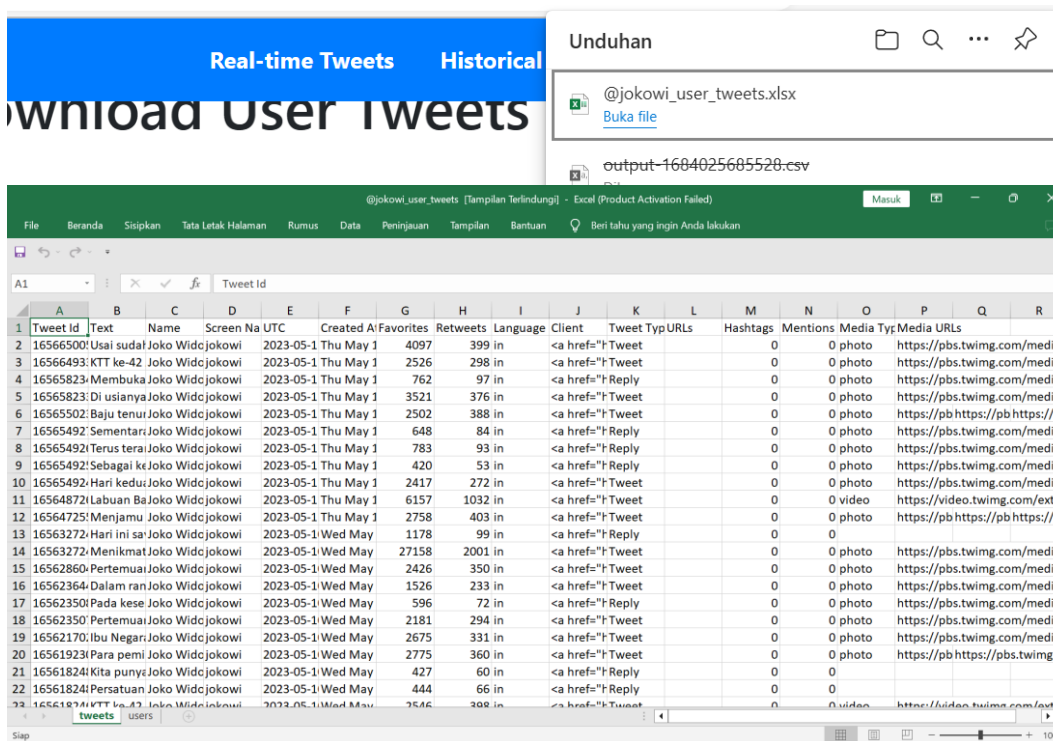
Terlihat bahwa Tokopedia menampilkan hasil yang lebih relevan dengan katakunci Kessoku muncul sebanyak 53 kali, dibandingkan dengan shopee yang hanya memunculkan kata kunci Kessoku sebanyak 4 kali.

- Lakukan Crawling twitter dari twitter org terkenal atau yg lagi viral saat ini, simpan hasil crawling tsb kedlm file csv

Pertama, sign up dengan akun twitter lalu masukkan username yang ingin dilakukan data crawling



setelah selesai, export ke file excel untuk mendapatkan hasilnya



Setelah dibuka, tampilannya akan menjadi seperti ini