

# LAPORAN PRAKTIKUM INFORMATIKA

( Tugas / Tes)

Nama : David Christian Priyanto, Alan Wijaya  
Kelas : XII-4 No 1 & 12  
Nama Program : Project Akhir Web Scraping  
Bahasa Pemrograman : Python  
Compiler : Python 3.12

## 1. Syntax

```
WEB SCRAPING , PARSING, FRAMING PROCESS
import selenium
from selenium import webdriver
from selenium.webdriver.chrome.service import Service
from webdriver_manager.chrome import ChromeDriverManager
from bs4 import BeautifulSoup
import time
import numpy as np
import csv
import matplotlib.pyplot as plt
import os
import pandas as pd

service = Service(executable_path=ChromeDriverManager().install())
driver = webdriver.Chrome(service=service)

url = "https://www.ruparupa.com/c/rak-dan-penyimpanan/rak/rak-
buku.html?itm_source=megamenu-rak-buku&itm_campaign=third-
level&itm_device=desktop"

driver.get(url)

time.sleep(5)

page_source = driver.page_source

soup = BeautifulSoup(page_source, 'lxml')

links = []
itemlink = soup.find_all('div', class_='row product-card')
for link in itemlink:
    for item in link.find_all('a', href=True):
        links.append(item['href'])
        print(item['href'])

AjazzList = []
products = soup.find_all('div', class_='container-card__product container-
card__product--mobile')
for product in products:
    a = product.find('span', class_='product_name')
    h = product.find('div', class_='price__initial')
    i = product.find('span', class_='ui-text-4 text__grey50')
```

```

        if a and h and i:
            AjazzList.append({'Nama': a.text.strip(), 'Harga': h.text.strip(),
                              'Rating': i.text.strip()})
        print(AjazzList)

df = pd.DataFrame(AjazzList, columns=['Nama', 'Harga', 'Rating'])
print(df)
directory = r'D:\coding\PYTHON\TIK KELAS 12'
filename = "Data-Rak.csv"
file_path = os.path.join(directory, filename)

df.to_csv(file_path, index=False)

print(f"File saved at {file_path}")
driver.quit

```

#### **PEMROSESAN DATA**

```

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

dataset = pd.read_csv(r'D:\coding\PYTHON\TIK KELAS 12\Web_Scraping\Data-
Rak.csv')
print(dataset)

NanData = pd.isnull(dataset["Nama"])
dataset[NanData]

NanData = pd.isnull(dataset["Harga"])
dataset[NanData]

NanData = pd.isnull(dataset["Rating"])
dataset[NanData]

duplicate = dataset[dataset.duplicated(["Nama"], keep=False)]

dropData = dataset.dropna()

print("Jumlah data Rak sebelum membuang data kosong : ", len(dataset))
print("Jumlah data Rak sesudah membuang data kosong : ", len(dropData))
print("Jumlah data kosong : ", len(dataset) - len(dropData))

#mencari rata-rata harga
dataset['Harga'] = dataset['Harga'].str.replace('Rp', '').str.replace('.', '',
'')
dataset['Harga'] = dataset['Harga'].astype(float)
dataMean = dataset["Harga"].mean()
print("Rata-rata Harga barang : ", "Rp ", dataMean)

#mencari rata-rata rating
dataset['Rating'] = dataset['Rating'].astype(float)
RatingMean = dataset["Rating"].mean()
print("Rata-rata Rating barang : ", "Rp ", RatingMean)

#mencari rating tertinggi
maxRating = dataset["Rating"].max()

```

```

print("Rating tertinggi: ", maxRating)

#mencari rating terendah
minRating = dataset["Rating"].min()
print("Rating terendah: ", minRating)

# Mencari korelasi pearson antara harga dan rating
corr_coef = dataset['Harga'].corr(dataset['Rating'])
print("Koefisien korelasi antara Harga dan Rating: ", corr_coef)
if corr_coef <= 0:
    print("""harga dan Rating tidak ada hubungan.
          Naiknya harga tidak memengaruhi Rating.""")
elif 0.01 <= corr_coef <= 0.09 :
    print("Harga dan Rating memiliki hubungan namun sangat lemah")
elif 0.1 <= corr_coef <= 0.29 :
    print("Harga dan Rating memiliki hubungan lemah")
elif 0.3 <= corr_coef <= 0.49:
    print("Harga dan Rating memiliki hubungan sedang")
elif 0.5 <= corr_coef <= 0.69:
    print("Harga dan Rating memiliki hubungan kuat")
elif 0.7 <= corr_coef <= 0.89:
    print("Harga dan Rating memiliki hubungan yang sangat kuat")
elif corr_coef >= 0.9:
    print("""Harga dan Rating memiliki hubungan yang sangat kuat.""")
else:
    print("No numbers")

# visualisasi distribusi harga
plt.figure(figsize=(16, 6))
plt.hist(dataset['Harga'], bins=50, alpha=0.7, color='skyblue')
plt.title("Distribusi Harga")
plt.xlabel("Harga (Rp)")
plt.ylabel("Frequency")
plt.xticks(dataset['Harga'].unique(), rotation=90, fontsize=6)
plt.show()

# visualisasi hubungan 'Harga' dan 'Rating'
plt.figure(figsize=(10, 6))
sns.scatterplot(x='Harga', y='Rating', data=dataset)
plt.title("Hubungan antara Harga dan Rating")
plt.xlabel("Harga (Rp)")
plt.ylabel("Rating")
plt.xticks(dataset['Harga'].unique(), rotation=90, fontsize=6)
plt.show()

#visualisasi data dengan diagram lingkaran
rating_counts = dataset['Rating'].value_counts()
plt.figure(figsize=(8, 6))
plt.pie(rating_counts, labels=rating_counts.index, autopct=lambda p :
' {:.1f}% ( {:.0f}) '.format(p,p * sum(rating_counts)/100), textprops={'size':
'11'})
plt.title("Distribusi Rating")
plt.show()

```

## 2. Penjelasan Syntax

### **PENJELASAN WEB SCRAPING**

Pada kode ini, dilakukan web scraping menggunakan Python untuk mengambil data produk dari situs Ruparupa.com. Program diawali dengan mengimport library yang diperlukan, yaitu selenium untuk mengotomatisasi browser, webdriver\_manager untuk mengelola driver Chrome, BeautifulSoup dari bs4 untuk parsing HTML, serta beberapa pustaka tambahan seperti time, numpy, csv, matplotlib.pyplot, os, dan pandas. Pustaka ini mendukung proses scraping, pengelolaan data, dan penyimpanan dalam file CSV.

Selanjutnya, driver Chrome diinisialisasi menggunakan webdriver.Chrome() yang dikonfigurasi dengan ChromeDriverManager untuk secara otomatis menginstal driver Chrome terbaru. Kemudian halaman web Ruparupa yang berisi produk rak buku diakses dengan driver.get(url). Setelah halaman terbuka, program menunggu selama 5 detik dengan time.sleep(5) untuk memastikan halaman termuat sepenuhnya sebelum konten diambil. Konten HTML dari halaman web kemudian diekstraksi dengan driver.page\_source dan diparsing menggunakan BeautifulSoup agar data dapat diambil dari elemen-elemen HTML yang spesifik.

Program selanjutnya mengekstrak link produk dengan mencari elemen div yang memiliki kelas 'row product-card', di mana link produk ditemukan di dalam elemen a yang memiliki atribut href. Semua tautan produk disimpan dalam list links. Setelah itu, nama produk, harga, dan rating diekstrak dari elemen HTML yang sesuai, dan data tersebut disimpan ke dalam list AjazzList sebagai dictionary yang berisi "Nama", "Harga", dan "Rating" untuk setiap produk.

Data yang telah dikumpulkan dalam list kemudian diubah menjadi DataFrame menggunakan pustaka pandas untuk memudahkan pengelolaan dan analisis data. Data ini selanjutnya dicetak ke konsol sebagai output sementara, dan disimpan dalam file CSV dengan nama "Data-Rak.csv" di direktori yang telah ditentukan. Penyimpanan file dilakukan dengan fungsi df.to\_csv(), dan setelah file tersimpan, browser otomatis yang digunakan untuk scraping ditutup dengan driver.quit(). Kode ini memberikan hasil berupa data terstruktur dalam format CSV, yang dapat digunakan untuk analisis lebih lanjut atau keperluan lain.

### **LANGKAH ANALISIS DATA**

Pertama, pustaka pandas, matplotlib.pyplot, dan seaborn diimpor untuk pengelolaan data, visualisasi, dan analisis statistik. Data produk rak buku diambil dari file CSV yang disimpan di direktori lokal menggunakan pd.read\_csv(), dan kemudian dicetak untuk dilihat isinya. Program ini melakukan pengecekan apakah ada nilai kosong (missing values) pada kolom "Nama", "Harga", dan "Rating" dengan menggunakan pd.isnull(), dan data yang mengandung nilai kosong ditampilkan. Selain itu, data duplikat dalam kolom "Nama" dicari dengan

menggunakan fungsi `dataset.duplicated()`.

Langkah selanjutnya adalah menghapus data yang memiliki nilai kosong dengan `dataset.dropna()`, kemudian menghitung dan mencetak jumlah data sebelum dan sesudah data kosong dibuang, serta menghitung jumlah data yang kosong. Setelah itu, program melakukan konversi kolom "Harga" dari format string (yang mengandung simbol "Rp" dan titik) menjadi tipe data numerik (float) dengan mengganti simbol "Rp" dan titik, kemudian menghitung rata-rata harga produk. Rata-rata harga tersebut dicetak dengan format rupiah. Selanjutnya, kolom "Rating" juga diubah menjadi tipe data float, dan rata-rata rating dihitung serta dicetak.

Untuk mengetahui rating tertinggi dan terendah, program menggunakan fungsi `max()` dan `min()` pada kolom "Rating". Setelah itu, program menghitung koefisien korelasi Pearson antara harga dan rating produk menggunakan `corr()`. Berdasarkan nilai koefisien korelasi yang dihasilkan, program memberikan interpretasi hubungan antara harga dan rating, apakah tidak ada hubungan, hubungan sangat lemah, lemah, sedang, kuat, atau sangat kuat.

Program juga membuat beberapa visualisasi data. Pertama, histogram distribusi harga dibuat menggunakan `plt.hist()` untuk menunjukkan frekuensi harga produk. Grafik scatterplot (`sns.scatterplot()`) dibuat untuk memvisualisasikan hubungan antara harga dan rating. Terakhir, diagram lingkaran (`plt.pie()`) digunakan untuk menunjukkan distribusi rating produk. Setiap visualisasi dilengkapi dengan judul, label sumbu, dan penyesuaian tampilan untuk memastikan data mudah dibaca.

### 3. Output

	Nama	Harga	Rating
0	Informa Rana Rak Buku 10 Kompartemen - Putih	Rp1.999.000	5
1	Informa Galvin Rak Buku 4 Tingkat - Putih Oak	Rp1.499.000	5
2	Informa Pablo Rak Buku 6 Tingkat - Cokelat Oak	Rp1.499.000	4.6
3	Masterspace Rak Buku 5 Tingkat - Hitam	Rp999.900	4.9
4	Masterspace Rak Buku 5 Tingkat - Putih	Rp999.900	5
5	Informa Abby Rak Buku Anak 2 Tingkat	Rp2.099.000	4.5
6	Informa Abby Rak Buku Anak 4 Tingkat	Rp1.799.000	5
7	Informa Rana Rak Buku 10 Kompartemen - Cokelat...	Rp1.999.000	5
8	Informa Pablo Rak Buku 6 Tingkat - Putih	Rp1.499.000	4.7
9	Informa Dallas Rak Buku 12 Kompartemen - Cokel...	Rp1.999.000	5
10	Informa Pablo Rak Buku 6 Tingkat & Keranjang P...	Rp1.827.000	5
11	Informa Dallas Rak Buku 9 Kompartemen - Cokela...	Rp1.599.000	5
12	Masterspace Rak Buku 4 Tingkat - Putih	Rp699.900	5

13	Stora Rak Penyimpanan 3 Tingkat - Cokelat	Rp219.900	5
14	Stora Rak Penyimpanan 4 Tingkat - Cokelat	Rp289.900	5
15	Masterspace Rak Buku 4 Tingkat - Hitam	Rp699.900	4.9
16	Selma Ciara Rak Penyimpanan 4 Tingkat - Cokela...	Rp199.000	5
17	Selma Ciara Rak Penyimpanan 3 Tingkat - Cokela...	Rp199.000	5
18	Informa Paulson Rak Buku 5 Tingkat - Putih	Rp899.000	4.9
19	Selma Ciara Rak Penyimpanan 3 Tingkat - Putih	Rp199.000	5
20	Stora Rak Penyimpanan 5 Tingkat - Putih	Rp319.900	5
21	Selma Ciara Rak Penyimpanan 4 Tingkat - Putih	Rp199.000	5
22	Stora Rak Penyimpanan 1 Tingkat Stackable	Rp149.900	5
23	Selma Ciara Rak Penyimpanan 2 Tingkat - Putih	Rp159.000	5
24	Selma Ciara Rak Penyimpanan 2 Tingkat - Cokela...	Rp159.000	5
25	Informa Abby Rak Buku Anak 3 Tingkat	Rp2.799.000	4.7
26	Informa Milton Rak Buku - Putih	Rp2.599.000	4.9
27	Masterspace 60 Cm Rak Stairwell - Oak	Rp999.900	5
28	Informa Pablo Rak Buku 3 Tingkat - Putih	Rp599.000	4.9
29	Informa Pablo Rak Buku 3 Tingkat L40 - Putih	Rp499.000	5
30	Informa Pablo Rak Buku 3 Tingkat L40 - Cokelat...	Rp499.000	4.2
31	Informa Mac Rak Buku - Cokelat/putih	Rp2.599.000	5
32	Informa Naco Rak Buku - Putih	Rp3.699.000	4
33	Informa Logan Rak Buku Pintu Kaca - Putih	Rp2.799.000	5
34	Informa Pablo Rak Buku 3 Tingkat & Kotak Penyi...	Rp637.000	5
35	Informa Pablo Rak Buku 3 Tingkat & Kotak Penyi...	Rp637.000	5

	Nama	Harga	Rating
0	Informa Rana Rak Buku 10 Kompartemen - Putih	Rp1.999.000	5.0
1	Informa Galvin Rak Buku 4 Tingkat - Putih Oak	Rp1.499.000	5.0
2	Informa Pablo Rak Buku 6 Tingkat - Cokelat Oak	Rp1.499.000	4.6
3	Masterspace Rak Buku 5 Tingkat - Hitam	Rp999.900	4.9
4	Masterspace Rak Buku 5 Tingkat - Putih	Rp999.900	5.0
5	Informa Abby Rak Buku Anak 2 Tingkat	Rp2.099.000	4.5
6	Informa Abby Rak Buku Anak 4 Tingkat	Rp1.799.000	5.0
7	Informa Rana Rak Buku 10 Kompartemen - Cokelat...	Rp1.999.000	5.0
8	Informa Pablo Rak Buku 6 Tingkat - Putih	Rp1.499.000	4.7
9	Informa Dallas Rak Buku 12 Kompartemen - Cokel...	Rp1.999.000	5.0
10	Informa Pablo Rak Buku 6 Tingkat & Keranjang P...	Rp1.827.000	5.0
11	Informa Dallas Rak Buku 9 Kompartemen - Cokela...	Rp1.599.000	5.0
12	Masterspace Rak Buku 4 Tingkat - Putih	Rp699.900	5.0
13	Stora Rak Penyimpanan 3 Tingkat - Cokelat	Rp219.900	5.0
14	Stora Rak Penyimpanan 4 Tingkat - Cokelat	Rp289.900	5.0
15	Masterspace Rak Buku 4 Tingkat - Hitam	Rp699.900	4.9
16	Selma Ciara Rak Penyimpanan 4 Tingkat - Cokela...	Rp199.000	5.0
17	Selma Ciara Rak Penyimpanan 3 Tingkat - Cokela...	Rp199.000	5.0
18	Informa Paulson Rak Buku 5 Tingkat - Putih	Rp899.000	4.9
19	Selma Ciara Rak Penyimpanan 3 Tingkat - Putih	Rp199.000	5.0
20	Stora Rak Penyimpanan 5 Tingkat - Putih	Rp319.900	5.0

21	Selma Ciara Rak Penyimpanan 4 Tingkat - Putih	Rp199.000	5.0
22	Stora Rak Penyimpanan 1 Tingkat Stackable	Rp149.900	5.0
23	Selma Ciara Rak Penyimpanan 2 Tingkat - Putih	Rp159.000	5.0
24	Selma Ciara Rak Penyimpanan 2 Tingkat - Cokela...	Rp159.000	5.0
25	Informa Abby Rak Buku Anak 3 Tingkat	Rp2.799.000	4.7
26	Informa Milton Rak Buku - Putih	Rp2.599.000	4.9
27	Masterspace 60 Cm Rak Stairwell - Oak	Rp999.900	5.0
28	Informa Pablo Rak Buku 3 Tingkat - Putih	Rp599.000	4.9
29	Informa Pablo Rak Buku 3 Tingkat L40 - Putih	Rp499.000	5.0
30	Informa Pablo Rak Buku 3 Tingkat L40 - Cokelat...	Rp499.000	4.2
31	Informa Mac Rak Buku - Cokelat/putih	Rp2.599.000	5.0
32	Informa Naco Rak Buku - Putih	Rp3.699.000	4.0
33	Informa Logan Rak Buku Pintu Kaca - Putih	Rp2.799.000	5.0
34	Informa Pablo Rak Buku 3 Tingkat & Kotak Penyi...	Rp637.000	5.0
35	Informa Pablo Rak Buku 3 Tingkat & Kotak Penyi...	Rp637.000	5.0

Jumlah data Rak sebelum membuang data kosong : 36

Jumlah data Rak sesudah membuang data kosong : 36

Jumlah data kosong : 0

Rata-rata Harga barang : Rp 1182669.4444444445

Rata-rata Rating barang : Rp 4.894444444444444

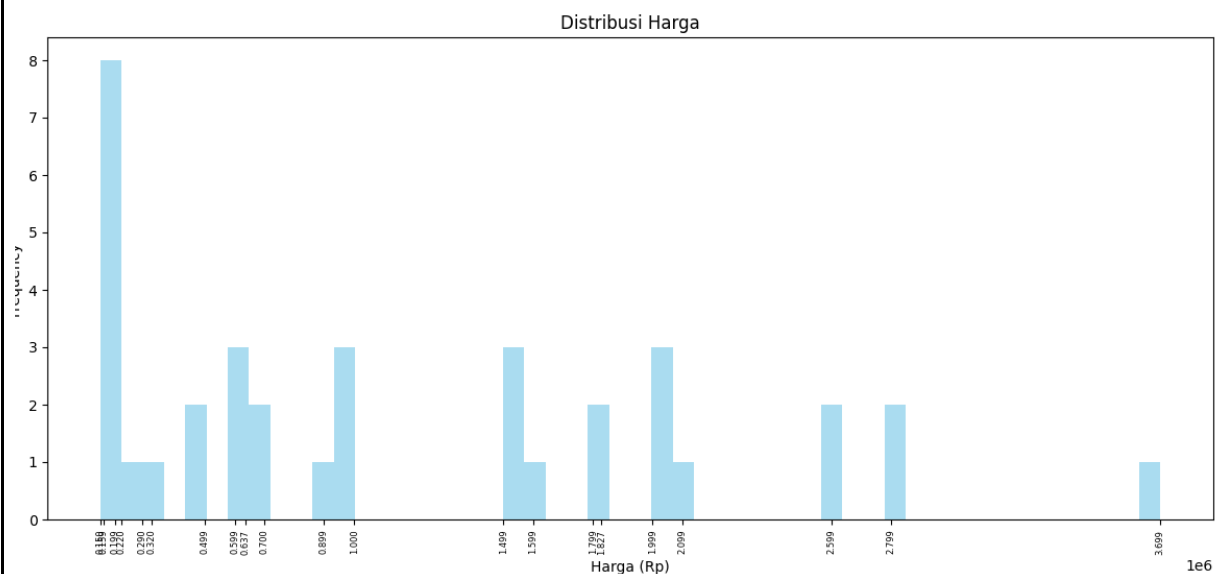
Rating tertinggi: 5.0

Rating terendah: 4.0

Koefisien korelasi antara Harga dan Rating: -0.404592292673446

harga dan Rating tidak ada hubungan.

Naiknya harga tidak memengaruhi Rating.







Penjelasan Grafik:

Harga dan rating tidak punya hubungan sama sekali. Perubahan nominal harga tidak mempengaruhi rating begitupula perubahan rating tidak mempengaruhi harga

Dari diagram lingkaran, 69.4% produk memiliki rating 5.0, 13.9% memiliki rating 4.9, 5.6% memiliki rating 4.7, 2.8% produk memiliki rating 4.6 , 4.5, 4.2, dan 4.0