

Analisis Toko dan Asal Toko *Fashion* Pria di Shopee Menggunakan *Data Scrapping* dan *Exploratory Data Analysis*

Mezan el-Khaeri Kesuma¹, Rodzan Iskandar²

[Submission: 13-02-2022, Accepted: 04-04-2022]

Abstract— In the current era of technological development, e-commerce is required to be more developed, effective, efficient and become the main choice for buyers and sellers. The problem formulation of this journal is what stores are trending and the city of origin of the store. This is evidenced by the best-selling products by category and type of men's fashion products. This study aims to find out the types of fashion that are trending/loved by buyers in the shopee marketplace and can be a recommendation for what fashion products are suitable for exist seller or newcomers to Shopee Indonesia. This research was conducted on men's fashion trend shops which started to exist after their establishment of the shop until April 4th 2021. This research was carried out in 4 steps by applying Web Scrapping and Exploratory Data Analysis techniques: Observation, Data Collection, Information Extraction and Analysis. The results of this study: (1) The best-selling shop is Erigostore from Bandung which has sold more than 400,000 products, the types of products are t-shirts, chinos, jackets, trousers, sweater cardigans. (2) the best-selling products sold on Shopee.id are the top (clothes) product category with casual shirts and shirts as well as the pants product category with chino, slim fit jeans and boxer products.

Keyword— Shopee, Data Scrapping, *Fashion*, Shop, EDA.

Intisari— Pada era perkembangan teknologi saat ini *e-Commerce* dituntut untuk semakin berkembang, efektif, efisien serta menjadi pilihan utama bagi pembeli dan penjual. Rumusan masalah dari jurnal ini adalah toko apa yang menjadi tren dan kota asal toko. Hal ini dibuktikan produk yang paling laris terjual berdasarkan kategori dan jenis produk *fashion* pria. Penelitian ini bertujuan untuk mengetahui jenis *fashion* yang sedang tren/ digandrungi oleh pembeli di *marketplace* shopee dan bisa menjadi rekomendasi produk *fashion* apa yang cocok untuk pendatang baru di Shopee Indonesia. Penelitian ini dilakukan kepada toko trend *fashion* pria yang mulai eksis sejak berdirinya toko hingga tanggal 4 April 2021. Penelitian ini dilakukan 4 langkah dalam menerapkan teknik Web Scrapping dan Exploratory Data Analysis: Observasi, Pengumpulan Data, Ekstraksi Informasi dan Analisis. Hasil dari penelitian ini: (1) Toko yang paling laris produknya adalah Erigostore asal kota Bandung yang telah menjual lebih dari 400.000 produk, jenis

produknya berupa kaos, chino, jaket, celana panjang, sweater cardigan. (2) produk yang paling laris terjual di Shopee.id adalah kategori produk atasan (baju) dengan jenis produk kaos dan kemeja kasual serta kategori produk celana dengan jenis produk chino, slim fit jeans dan boxer.

Kata Kunci— Shopee, Data Scrapping, EDA, *Fashion*, Toko.

I. PENDAHULUAN

Dengan berkembangnya teknologi saat ini *e-Commerce* dituntut untuk semakin berkembang, efektif, efisien serta menjadi pilihan utama bagi pembeli dan penjual. Internet dapat membantu pengguna dalam mencari informasi dan layanan mengenai produk yang diinginkan. Teknologi Informasi membuka peluang secara luas dan berdampak besar terhadap kehidupan masyarakat untuk memberikan informasi yang aktual di era globalisasi ini[1]. Untuk meningkatkan servis (layananan) sumber informasi merupakan peran penting dari Teknologi Informasi[2].

Selama pandemi Covid-19 sejak tahun 2020 Indonesia khususnya mengalami kelumpuhan dan resesi ekonomi yang berkepanjangan. Namun hal ini merupakan peluang dan tantangan dalam mencari strategi permasalahan resesi ekonomi, salah satunya melalui strategi jual-beli dan pemasaran *online*[3]. Pemesanan barang dimanapun dan kapanpun dengan mudah dan cepat tanpa harus datang langsung ke lokasi sehingga pemesanan barang dilakukan cukup dilakukan dengan mengakses Internet [4].

Shopee adalah sebuah perusahaan *e-Commerce* yang dikelola oleh SEA Group (Garena) perusahaan internet di Asia Tenggara. Shopee pada tahun 2020 memiliki pengunduhan (download) ranking pertama di Indonesia tetapi menempati ranking empat berdasarkan jumlah visitor perbulannya[5]. Shopee menjadi *marketplace* yang banyak dimintai oleh pengguna dikarenakan adanya tingkat kepercayaan yang dibangun. Kepercayaan berupa kemudahan dan kualitas informasi adalah aspek penting bagi sebuah *marketplace e-Commerce* untuk menarik perhatian dan minat user[6] termasuk pelayanan toko kepada pelanggan didalamnya. Kepuasan pelanggan selalu menjadi prioritas agar dapat memenuhi kebutuhan pelanggan berdasarkan dari history penjualan yang sudah dilakukan sehingga dapat menjadi acuan dalam meningkatkan kualitas produk-produk berikutnya dalam melakukan perencanaan stok untuk menargetkan penjualan yang lebih baik sehingga mampu memberikan kepuasan bagi para pelanggan [7].

Sekarang Shopee merupakan salah satu *e-Commerce* yang perkembangannya cukup baik di Indonesia. Shopee adalah salah satu platform belanja *online* (*e-Commerce*) yang

p-ISSN:1693 – 2951; e-ISSN: 2503-2372

¹Dosen, Program Studi Ilmu Perpustakaan dan Informasi Islam, Fakultas Adab Universitas Islam Negeri Raden Intan Lampung, Jl. Letnan Kolonel H. Endro Suratmin, Sukarame, Bandar Lampung, Lampung 35151 INDONESIA (telp: 082280682703; e-mail: mezan@radenintan.ac.id)

²Alumni, Teknik Fisika, Institut Teknologi Sepuluh Nopember Surabaya Jl. Teknik Kimia, Keputih, Kec. Sukolilo, Surabaya, Jawa Timur 60111 INDONESIA (telp: 031-5994251; e-mail: rodzan.iskandar@gmail.com)



menyediakan segala macam produk seperti kebutuhan rumah tangga, makanan, minuman, alat elektronik, pakaian, aksesoris hingga *fashion*. Selain itu banyak kemudahan ketika memakai platform ini antara lain: pertama, pengguna juga dapat mengembalikan dana atau produk yang telah dibeli jika produk tersebut tidak sesuai keinginan ataupun rusak ketika diterima. Kedua, pengguna dapat mencari produk dengan mudah dan gampang dengan cara memasukkan gambar produk yang sedang dicari melalui kolom pencarian (*searching*). Ketiga, sewaktu-waktu berdasarkan *event* atau tanggal tertentu shopee memberikan layanan gratis ongkos kirim (ongkir) untuk pengiriman barang. Kemudahan-kemudahan ini membuat shopee menjadi *platform* nomor empat di kategori pencarian *e-Commerce* yang paling diminati dengan jumlah pengunjung sebanyak 34 juta lebih sehingga menjadi Top 40 *e-Commerce*. Pengguna shopee yang terlalu banyak memungkinkan terjadi masalah pelanggan yang merasa kecewa terhadap pelayanan yang telah diberikan[8].

Di Indonesia shopee merupakan platform yang memiliki ciri khas yang penggunaannya mayoritas terdiri dari UMKM dan agen/gerai resmi dari banyak brand ternama. Namun sekarang ini Shopee juga telah meluncurkan platform baru yang bernama Shopee Mall. Shopee Mall adalah toko terpilih di Shopee yang memberikan pembeli pengalaman berbelanja *online* terbaik. Sampai saat ini, sudah ada sekitar 500 brand bergabung di dalam *marketplace* Shopee Mall. Tidak ada perbedaan akses untuk masuk ke Shopee Mall, karena sudah menyatu langsung dengan aplikasi mobile Shopee, di mana nanti ada banner khusus untuk masuk ke mall. Untuk menjadi penjual di Shopee Mall, Anda harus bisa memastikan untuk memberikan layanan operasional terbaik. Namun disini peneliti akan fokus kepada *marketplace* Shopee Indonesia (shopee.co.id) bukan Shopee Mall.

Semakin besar sebuah platform pastinya semakin kompleks pula masalah yang akan muncul. Sehingga penelitian ini dilakukan berdasarkan pada tanggapan pembeli dan penjual tentang perubahan sistem pengiriman peraturan oleh Shopee Indonesia dalam sistem *marketplace* mereka. Seperti suatu ketika pernah terjadi cukup mahalnya ongkos kirim yang dibebankan oleh penjual maupun pembeli yang diputuskan secara sepihak. Dikarenakan hal ini terjadi bahkan mendorong beberapa pengguna untuk beralih platform *marketplace* lain. Peristiwa ini pernah menjadi salah satu *thread viral* atau *trending topic* di media sosial twitter. Ide utamanya adalah untuk mendapatkan gambaran yang lebih besar tentang pasar *fashion* pria di shopee Indonesia dan membandingkannya dengan *platform marketplace* lain di Indonesia seperti tokopedia, buka lapak dll. Sehingga jika terjadi migrasi konsumen dari Shopee ke *marketplace* lainnya konsumen bisa menjual *fashion* pria yang tren (penjualan terlaris) yang belum ada ada di *marketplace* selain shopee.

Data penjualan yang ada di shopee Indonesia (shopee.co.id) dapat secara otomatis didapatkan dengan menggunakan web scrapping. Analisis yang dibangun menggunakan web scrapping ini bertujuan untuk mengambil data: nama produk, kategori produk, jenis produk, lokasi toko, harga produk, jumlah produk terjual, peringkat toko, jumlah review toko, jumlah produk toko, umur toko, jumlah pengikut toko.

Sebelum melakukan analisis data kuantitatif, lebih baik mempelajari karakter dari sebuah data seperti bagaimana

bentuk atau pola dari sebaran data. Tujuannya adalah agar metode analisis statistik yang dipilih adalah benar. Untuk mempelajari karakteristik sebuah data maka dapat dilakukan dengan analisis data eksploratif (*Explorative Data Analysis*)[9].

Tujuan dari penelitian ini adalah (1) Bagi peneliti untuk mengetahui toko apa yang paling laris (banyak terjual) produknya dan asal kota berapa banyak produk yang telah terjual, serta apa jenis produknya. (2) Bagi pembeli untuk mengetahui produk yang paling laris terjual di Shopee.id dengan rincian kategori dan jenis produk toko. (3) Bagi penjual lama, penjual baru ataupun rencana menjadi penjual bisa menjadi rekomendasi produk *fashion* apa yang cocok di Shopee Indonesia.

Penelitian sebelumnya yang terkait tentang Shopee dan *Web Scrapping* dilakukan oleh: (1) Mitra dkk Rancang Bangun Aplikasi *Web Scraping* untuk Korpus Paralel Indonesia-Inggris dengan Metode HTML DOM tahun 2017[10]. Penelitian ini sama-sama menggunakan aplikasi *web scrapping*, metode berupa *web scrapping* dan *cleaning*; (2) Ayani dkk dengan judul Implementasi *Web Scraping* untuk Pengambilan Data pada Situs *Marketplace* pada tahun 2019[11]. Penelitian ini sama-sama mengambil data (*Web Scrapping*) pada situs *marketplace* Bukalapak.com, Elevenia, dan JD.id; (3) Yoga Sahria dengan judul Implementasi Teknik *Web Scraping* pada Jurnal SINTA Untuk Analisis Topik Penelitian Kesehatan Indonesia pada tahun 2020[12]. Penelitian ini sama-sama menggunakan bahasa pemrograman Python dengan modul-modul *web scrapping* di dalamnya.

Yang membedakan penelitian ini dengan penelitian-penelitian sebelumnya ialah fokus penelitian. Pada penelitian sebelumnya, masih terfokus pada *Web Scrapping* pada Korpus dan *Marketplace*: Bukalapak, Elevenia dan JD.id. Sedangkan penelitian yang dilakukan ini fokus kepada *Web Scrapping* dan Analisis Data Eksploratif pada *Marketplace* Shopee khususnya produk *fashion* pria.. Serta untuk mengetahui lebih lanjut toko yang paling tren dan kota asal berdasarkan data produk yang paling banyak dibeli (paling banyak terjual).

II. TINJAUAN PUSTAKA

A. *Marketplace* Shopee

Shopee Indonesia (shopee.co.id) merupakan halaman elektronik daring yang bersifat komersial berupa pasar yang sering disebut dengan *marketplace*. Pada tahun 2009 Shopee pertama kali didirikan dan pada tahun 2015 launching (diluncurkan) oleh Forrest Li. Perusahaan shopee memiliki kantor utama di Singapore dan pada tahun 2021 memiliki pegawai lebih dari 20.000 orang[13].

Berbagai cara digunakan untuk mendapatkan loyalitas pembeli di *marketplace* shopee seperti analisis yang mendalam sampe kuisioner, dan wawancara[8]. Ada juga faktor lain seperti tampilan web, harga produk serta kepercayaan pembeli[5]. Bahkan bagi pembeli muslim di Indonesia kehalalan produk menjadi variable utama selain kepercayaan, kemudahan, kualitas informasi[6].

Pemilik UMKM dapat memanfaatkan *marketplace* Shopee secara maksimal untuk meningkatkan penjualan secara *online* dan memaksimalkan pelayanan yang diberikan kepada para konsumen yang membeli melalui *marketplace* shopee. Perlu

memaksimalkan dalam memanfaatkan seluruh fitur yang ada di Shopee baik fitur dari segi konsumen maupun penjual. Pemanfaatan ini dapat dilakukan dengan cara menambahkan produk baru maupun memberikan kelengkapan pada setiap keterangan dari produk[14].

B. Web Scrapping

Web Scrapping merupakan metode pengumpulan data melalui halaman web internet menggunakan API (Application Programming Interface)[12]. Web Scrapping digunakan untuk mendapatkan suatu data atau informasi pada suatu situs web berupa teks, tautan, video, audio ataupun dokumen secara otomatis[15]. Proses ini berupa pengambilan data, informasi, dokumen tadi yang berupa halaman-halaman situs (web) yang dibuat secara setengah (semi) terstruktur dalam Bahasa markup contohnya seperti HTML dan XHTML[10].

Josi & Suryayusra menjelaskan bahwa Web Scrapping sering dikenal sebagai Screen Scraping. Web scraping tidak dapat dimasukkan dalam bidang data mining karena data mining menyiratkan upaya untuk memahami pola semantik atau tren dari sejumlah besar data yang telah diperoleh[16].

Manfaat sistem informasi Web Scrapping antara lain: (1)mempermudah pelaku UMKM untuk mencari barang dan mengisi deskripsi produk dengan mudah dan cepat; (2)Pelaku UMKM juga dapat mengelola produk pada *marketplace* baik shopee maupun *marketplace* lainnya; (3)dapat mengelola data pengguna seperti data pribadi, lokasi, jumlah scrapping yang telah terdaftar dan total gabungan scrapping. Namun ada kekurangan yang terjadi pada SI web scrapping di *marketplace* tertentu seperti: (1)Scrapping data pada Alibaba tidak dapat scrapping gambar yang ada pada deskripsi produk; (2)pada *marketplace* Jd.id tidak dapat scrapping data produknya karena terdapat markup class banyak yang sama dan keamanan pada java script [17].

C. Analisis Data Eksploratif (Exploratory Data Analysis)

Analisis Data Eksploratif atau biasa disebut dengan Exploratory Data Analysis merupakan sebuah cara (metode) penjelajahan lebih banyak data (eksplorasi) dengan teknik aritmatika dan visual grafis dalam meringkas data yang diamati[9]. Metode ini digunakan untuk mencari tahu pattern pola sebaran, meringkas serta memvisualisasikan data dalam berbagai bentuk grafik, plot dan table dengan tujuan disajikan secara menyeluruh ringkasan statistik secara visual[18].

Dengan tersedianya data sedemikian besar, biasanya terdapat informasi yang terpendam dan dapat digali, teknik yang dapat digunakan bisa menggunakan statistic deksriptif biasa hingga teknik Data Mining dan Exploratory Data Analysis (EDA)[19].

Pada langkah terakhir pada penelitian ini dilakukan Exploratory Data Analysis (EDA), dimana dalam tahap ini dilakukan penggalan informasi apa saja yang dapat diperoleh dari data yang ada. Analisis ini dilakukan dengan bantuan bahasa pemrograman Python. Python merupakan sebuah bahasa pemrograman yang berorientasi obyek dinamis, yang kaya dengan pustaka- pustaka standar yang dapat diperluas dan dapat digunakan untuk bermacam macam pengembangan perangkat lunak[20].

III. METODE PENELITIAN

Penelitian ini dilakukan pada sebuah perusahaan di bidang *e-Commerce* yaitu Shopee Indonesia (shopee.co.id) khususnya kepada toko trend fashion pria yang mulai eksis sejak berdirinya toko hingga tanggal 4 April 2021 yang bertepatan dengan musim pandemi Covid-19. Objek penelitian ini adalah semua toko dan kota asal toko produk *fashion* pria di shopee Indonesia yang paling tren sehingga perlu dibuktikan dengan berdasarkan penjualan produk terbanyak (laris). Penelitian ini dilakukan dengan menerapkan teknik Web Scrapping yang dilakukan dengan cara observasi pada halaman *Marketplace* Shopee Indonesia.

Adapun metode penelitian digambarkan pada Gambar 1:



Gambar 1: Metode Penelitian

Tahapan pertama yakni membuka halaman web shopee Indonesia yang kemudian dibuat menjadi sumber data dan informasi. Pada tahap ini dilakukan dengan mengakses filtrasi halaman dengan kata kunci “*fashion* pria” dengan objek lebih dari 2.000 produk *fashion* pria. List *fashion* pria tersebut dapat dilihat pada url <https://shopee.co.id/Pakaian-Pria-cat.33?page=0&sortBy=sales>.

Tahap kedua, data dikumpulkan menggunakan teknik web *scrapping* dengan Bahasa pemrograman Python Versi 3.7.9 lebih spesifik menggunakan *library Selenium* dan *BeautifulSoup*.

Tahap ketiga yakni meringkas/ proses ekstraksi informasi yang bertujuan untuk mempermudah pengolahan data. Data yang sebelumnya tidak terstruktur menjadi terstruktur melalui proses *scrapping*.

Tahapan keempat yaitu analisis data. Hasil analisis penelitian ini visualisasi data dengan menggunakan Numpy, Pandas, Matplotlib, Seaborn. Tujuan Penggambaran/ visualisasi data yakni agar informasi yang didapat dapat dikomunikasikan dengan jelas, efektif dan efisien, yang kemudian dapat dianalisis kepada user (pengguna) dengan memilih grafis informasi yang sesuai. Hasil analisis dari penelitian ini peneliti berharap dapat membantu pembeli maupun penjual dalam mengambil keputusan. Dari analisis yang dilakukan akan terlihat dari toko paling tren dan kota asal toko yang paling banyak penjualan (laris) berdasarkan kategori dan jenis pakaian pria.

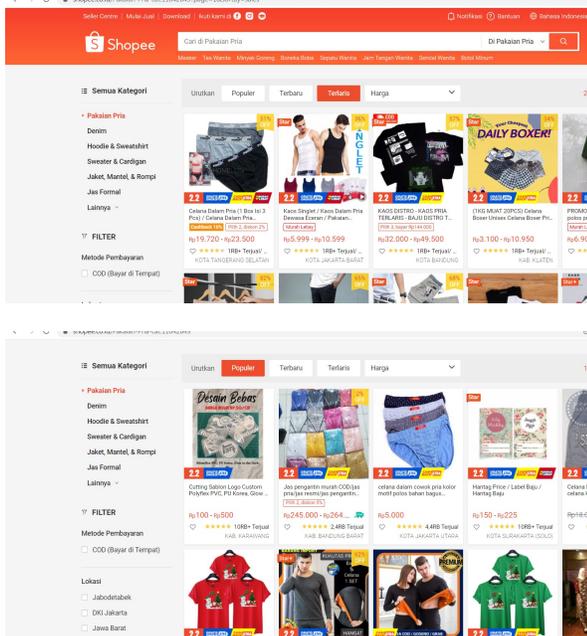


IV. PEMBAHASAN DAN ANALISIS SISTEM

Hasil dari penelitian ini yaitu penerapan teknik web scrapping pada *marketplace* shopee dengan menggunakan Bahasa pemrograman python dan hasil scrapping data ditampung di Ms. Excel dengan format .CSV kemudian data dianalisis untuk mengetahui toko paling tren dan kota asal toko berdasarkan tingkat penjualan produk yang paling tinggi.

A. Akses Shopee Indonesia

Berikut ini halaman web Shopee Indonesia yang di scrapping berdasarkan filtrasi *fashion* pria yang dikumpulkan datanya yang terdiri lebih dari 10.000 produk *fashion* pria. Berikut tampilan halaman web pada gambar dibawah ini:



Gambar 2: Halaman web shopee Indonesia (<https://shopee.co.id/Pakaian-Pria-cat.11042849?page=1&sortBy=sales>)

Gambar 2. menunjukkan tampilan toko *fashion* pria pada *marketplace* shopee Indonesia yang akan dilakukan teknik web *scrapping*.

B. Scrapping data

Sebelum menggunakan teknik web *scrapping*, sebaiknya harus peneliti ketahui dulu struktur HTML dari halaman situs (web) shopee untuk menentukan Tag HTML yang terdapat data informasi inti yang dipresentasikan dipresentasikan. Tujuannya agar dapat mengelola data toko *fashion* pria dan asal tokonya dengan mudah[17]. Sktruktur HTM tersebut dapat dilihat pada struktur *coding* pada tampilan pada link berikut <view-source:https://shopee.co.id/Pakaian-Pria-cat.11042849?page=0&sortBy=sales>. Pada tampilan link tersebut menunjukkan struktur coding HTML yang digunakan dengan Bahasa pemrograman python. Pada *coding* tersebut dilakukan tag HTML di halaman situs shopee Indonesia yaitu www.shopee.co.id.

Ketika peneliti sudah mengetahui tag halaman web yang akan diakuisisi. Peneliti membuat coding menggunakan Bahasa pemrograman python dalam menggunakan teknik scrapping yang bertujuan untuk memudahkan proses scrapping. Alamat

URL yang menjadi objek scrapping data adalah web shopee Indonesia: <https://shopee.co.id/Pakaian-Pria-cat.33?page=0&sortBy=sales>.

Kemudian untuk melihat peluang pasar dalam kategori yang sama. Peneliti membaca lebih dari 2000 postingan produk dalam mode sales man dari [link](#) yang sama dan menjawab beberapa pertanyaan seperti apa sub-kategori pasar yang cocok untuk dijual oleh penjual lama maupun pendatang baru di Shopee Indonesia dan toko apa yang paling tren dibuktikan dengan penjualan produk yang paling banyak (Laris).

Pengikisan Web dilakukan terhadap lebih dari 2000 posting produk dalam mode *sales man* dari [link](#) yang sama dalam bentuk list kriteria *fashion* pria setiap posting yang digunakan di penelitian ini diperlihatkan pada tabel:

TABEL 1
NAMA FIELD BASIS DATA YANG TERKAIT TAG HTML

No	Nama Field	Keterangan
1.	nama	nama produk
2.	kategori	Kategori produk
3.	jenis	lebih detail di bagian kategori
4.	lokasi	lokasi toko
5.	harga	harga produk
6.	terjual	penjualan
7.	peringkat	peringkat (rating) toko
8.	jumlah_review	jumlah review
9.	toko	nama toko
10.	n_produk_toko	jumlah produk di toko tertentu
11.	umur_toko	umur toko
12.	pengikut_toko	pengikut toko

Dari target *wrap scrapping fashion* pria di *marketplace* shopee di atas dapat ditentukan berdasarkan nama *field* basis data. Cuplikan dari coding Bahasa pemrograman python dalam proses scrapping data dapat dilihat di bawah ini:

TABEL 2
CODING SCRAPPING

```
from bs4 import BeautifulSoup
import pandas as pd
from selenium import webdriver
from selenium.webdriver.chrome.options import Options
import time
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.common.by import By
chrome_options = Options()
#chrome_options.add_argument('--headless')
chrome_options.add_argument('log-level=2')
chrome_options.add_argument("force-device-scale-factor=0.75");
chrome_options.add_argument("high-dpi-support=0.75");
path = 'D:\Data Science\Spyder\Web Scrapping\chromedriver.exe'
driver = webdriver.Chrome(path, options=chrome_options)

url = 'https://shopee.co.id/Pakaian-Pria-cat.33?page=0&sortBy=sales'
page=1
links = []
while True:
    if page <=50:
        try:
            driver.get(url)
            time.sleep(8)
            driver.implicitly_wait(30)
            driver.execute_script('window.scrollTo(0, 1500);')
            driver.implicitly_wait(30)
            driver.execute_script('window.scrollTo(0, 3500);')
            driver.implicitly_wait(30)
            soup_a = BeautifulSoup(driver.page_source, 'html.parser')
```

```

products = soup_a.find('div', class_='row shopee-search-item-
result__items')
for link in products.find_all('a'):
    links.append(link.get('href'))
except:
    print('gagal cok')
try:
    driver.find_element_by_css_selector('[class="shopee-icon-
button shopee-icon-button--right"]').click()
    url = 'https://shopee.co.id/Pakaian-Pria-
cat.33?page={}&sortBy=sales'.format(page)
    print('page ' + str(page))
    page+=1
except:
    print('udah nih')
    break
else:
    print('dapat ' + str(len(links)) + ' links')
    break
data = pd.DataFrame({'nama':[], 'kategori':[], 'jenis':[], 'lokasi':[],
'harga':[],
                    'terjual':[], 'rating':[], 'jumlah_review':[],
                    'toko':[], 'n_produk_toko':[], 'umur_toko':[],
                    'pengikut_toko':[]})
progres = 1
gagal = 1

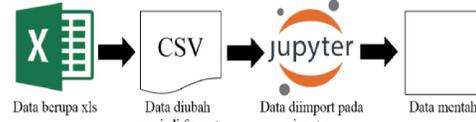
for link in links:
    try:
        url = 'http://www.shopee.co.id' + link
        driver.get(url)
        time.sleep(5)
        driver.execute_script('window.scrollTo(0, 500);')
        driver.implicitly_wait(30)
        WebDriverWait(driver, 30).until(
            EC.visibility_of_element_located((By.CLASS_NAME,
'attM6y'))))
        soup_b = BeautifulSoup(driver.page_source, 'html.parser')
        nama = soup_b.find('div', class_='attM6y').text
        kategori = soup_b.find_all('a', class_='3YDLCj')[2].text
        lokasi = soup_b.find_all('div', class_='aPKXeO')[-1].text
        harga = soup_b.find('div', class_='3e_UQT').text
        terjual = soup_b.find('div', class_='aca9MM').text
        rating = soup_b.find('div', class_='OitLRu_1mYa1t').text
        n_review = soup_b.find_all('div', class_='OitLRu')[-1].text
        toko = soup_b.find('div', class_='3uf2ae').text
        n_produk = soup_b.find_all('span', class_='zw2E3N
_2fK6RA')[1].text
        umur = soup_b.find_all('span', class_='zw2E3N')[-2].text
        pengikut = soup_b.find_all('span', class_='zw2E3N')[-1].text
        try:
            jenis = soup_b.find_all('a', class_='3YDLCj')[3].text
        except:
            jenis = None
        data = data.append({'nama':nama, 'kategori':kategori, 'jenis':jenis,
'lokasi':lokasi, 'harga':harga,
                    'terjual':terjual,
                    'rating':rating, 'jumlah_review':n_review, 'toko':toko,
                    'n_produk_toko':n_produk, 'umur_toko':umur,
                    'pengikut_toko':pengikut}, ignore_index=True)
        print('berhasil ' + str(progres) + '/' + str(len(links)))
        progres +=1
    except:
        print('gagal ' + str(gagal))
        gagal +=1
        pass
    data = data.iloc[1:]
#data.to_csv('pakaian pria shopee sales 6421.csv', index=False)
    
```

Tabel 2. menunjukan struktur coding yang digunakan untuk web scrapping berdasarkan nama_field basis data antara lain: nama, kategori, jenis, lokasi, harga, terjual, peringkat, MK Kesuma: Analisis Toko dan Asal...

jumlah_review, toko, n_produk_toko, umur_toko dan pengikut_toko. Sehingga setelah di-compile akan muncul hasil *web scrapping*. *Web scrapping* menunjukkan tampilan (sebagian) hasil *scrapping* yang didapat dengan 2.433 data lengkap dengan rincian nama produk, kategori produk, jenis produk (lebih detail di bagian kategori), lokasi toko, harga produk, total penjualan produk, peringkat (rating) toko, jumlah review, nama toko, jumlah produk di toko, umur toko, pengikut toko (follower).

C. Analisis

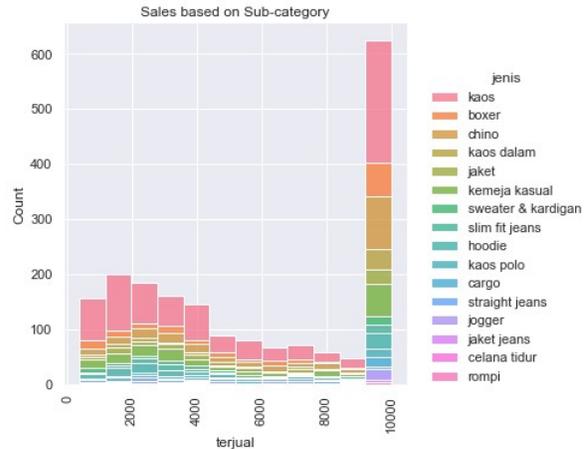
Tahap analisis ini berisikan tentang penggambaran (visualisasi) hasil dari data scrapping yang dikumpulkan di *marketplace* Shopee dengan langkah-langkah pada Gambar 6:



Gambar 3: Load Data

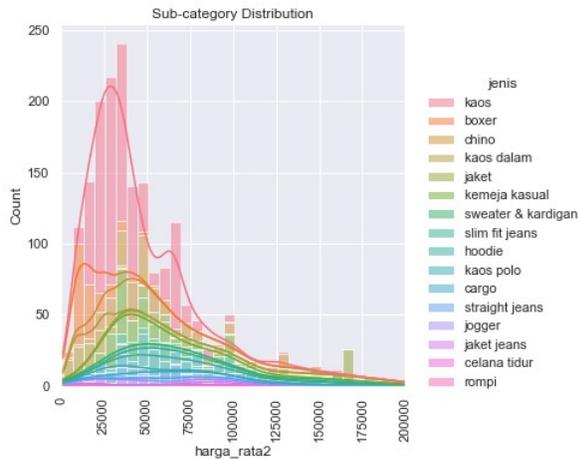
Pada gambar 3. menjelaskan tahapan analisis data yang dapat terbaca oleh *tools* python seperti jupyter. Sebelumnya data berbentuk format Ms. Excel (.xls) berubah menjadi Nilai Berbatas Koma (.csv). Adapun data yang di load adalah data keseluruhan hasil scrapping pada halaman web shopee Indonesia kategori *fashion* pria.

Setelah *load data* selanjutnya memvisualisasikan topik *fashion* pria berdasarkan jenis pakaian pria tersebut ke dalam bentuk grafik.



Gambar 4: Penjualan Fashion Pria berdasarkan jenis fashion pria dan jumlah barang yang terjual

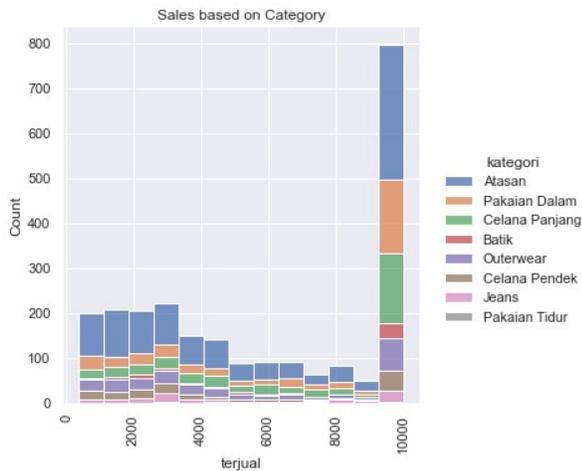




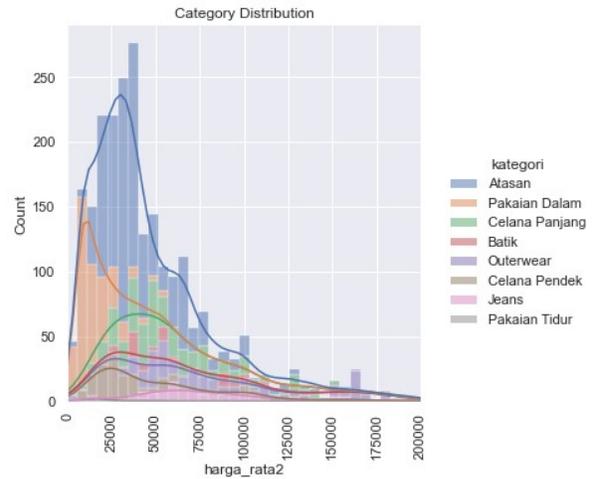
Gambar 5: Penjualan Fashion Pria berdasarkan jenis fashion pria dan harga rata-rata barang yang terjual

Pada gambar 4 dan 5 menjelaskan fashion pria yang telah terjual sebanyak lebih dari 10.000 produk dan harganya sekitar Rp. 25.000-50.000. Semua jenis fashion tersebut berupa produk berjenis: kaos, boxer, chino, kaos dalam, jaket dll.

adapun hasil visualisasi bentuk grafik fashion pria yang di scrapping berdasarkan kategori fashion pria dapat dilihat pada Gambar 6 dan 7:



Gambar 6: Penjualan Fashion Pria berdasarkan kategori fashion pria dan jumlah barang yang terjual



Gambar 7: Penjualan Fashion Pria berdasarkan kategori fashion pria dan harga rata-rata barang yang terjual

Pada gambar 6 dan 7 menjelaskan fashion pria yang telah terjual sebanyak lebih dari 10.000 produk dan harganya sekitar Rp. 25.000-50.000. Semua jenis fashion tersebut berupa produk berkategori: Atasan, Pakaian dalam, celana panjang, batik Outwear dll.

Dengan demikian dapat dilihat beberapa kategori dan jenis fashion pria yang menonjol/ paling banyak dibeli (laris). Sehingga untuk akurasi dan efisien nama toko dan kota asal toko yang redudan akan dikelompokkan menjadi dua klaster.

Pembersihan Data (Data Cleaning).

Data Cleaning adalah tahapan menyaring data, seperti penghapusan kata yang cukup panjang yang diatur oleh batas yang ditentukan Setelah scrapping, peneliti perlu membersihkannya agar data tersebut dapat digunakan untuk langkah selanjutnya, yaitu *Exploratory Data Analysis* (EDA). berikut beberapa perubahan yang peneliti lakukan:

TABEL 3

LANGKAH PERUBAHAN DATA UNTUK DATA CLEANING

No	Langkah
1.	kota atau kabupaten yang diuraikan di luar kolom lokasi(lokalasi)
2.	menghapus baris tanpa harga
3.	ubah kisaran harga menjadi harga rata-rata dan tambahkan nilainya di kolom harga_rata2(harga rata-rata)
4.	mengubah RB(K) atau JT(M) menjadi bilangan bulat ribuan atau jutaan dalam terjual(penjualan), n_review(jumlah ulasan), n_produk(jumlah produk) dan pengikut_toko(pengikut) kolom
5.	mengubah umur(umur) menjadi bilangan bulat dalam satuan bulan
6.	mengganti nilai shopee pada kolom jenis(sub-category) menjadi None
7.	periksa dan jatuhkan tabel yang digandakan

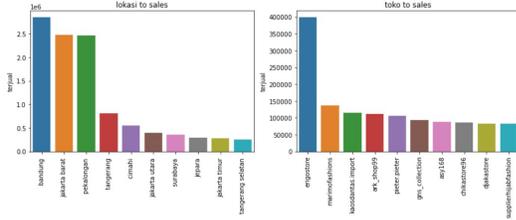
Pada tabel 3 setelah melakukan *cleaning* (pembersihan) data, sehingga informasi data yang cukup jelas untuk dilakukan EDA.

Tahap selanjutnya adalah EDA sehingga data toko fashion pria dan asal tokonya didapat lebih visual dan tergambar dengan jelas dengan teknik aritmatika[9]. Dengan visualisasi

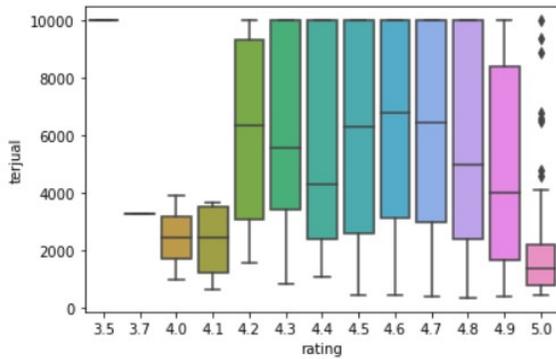
yang jelas maka peneliti dapat menjawab pertanyaan antara lain:

1. Apa nama toko yang penjualan terbanyak di pasar kategori *fashion pria*?
2. Kota asal toko yang penjualan terbanyak di pasar kategori *fashion pria*?

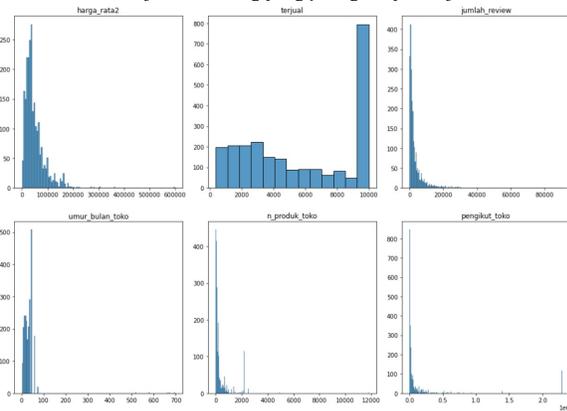
Disini peneliti juga melihat distribusi data dan korelasi antara data dengan visualisasi. Di bawah ini adalah beberapa sorotan dari EDA.



Gambar 8: Diagram grafik batang (bar plot) kategori dan jenis barang yang paling banyak terjual



Gambar 9: Diagram Kolom Discrete (Discrete Columns) tentang rating toko dan jumlah barang yang paling banyak terjual



Gambar 10: Diagram Kolom Kontinu (Countinous Columns) tentang rating toko dan jumlah barang yang paling banyak terjual

Pada gambar 8, 9, dan 10 dapat peneliti simpulkan bahwa “jumlah_review” satu-satunya fitur yang mempunyai korelasi positif dengan “terjual” Peneliti membuat 10 list toko yang paling laris dan kota asal toko yang dibuktikan dengan penjualan terbanyak:

TABEL 4
DATA YANG DIDAPAT SETELAH DATA CLEANING

MK Kesuma: Analisis Toko dan Asal...

No	Fashion Pria	
	Nama Toko	Kota Asal Toko
1.	Erigostore	Bandung
2.	Marimofashions	Jakarta barat
3.	Kaosdantas.import	Pekalongan
4.	Ark_shop99	Tangerang
5.	Pieter.pieter	Cimahi
6.	Gns_collection	Jakarta utara
7.	Asy168	Surabaya
8.	Chikastore96	Jepara
9.	Djakastore	Jakarta timur
10.	supplierhijabfashion	Tangerang selatan

TABEL 5
DATA 10 TOKO, PRODUK DAN JUMLAH PENJUALAN

No	Fashion Pria		
	Nama Toko	Produk (Jenis)	Jumlah Penjualan (item)
1.	Erigostore	kaos, chino, jaket, celana panjang, sweater cardigan	400.000
2.	Marimofashions	kemeja kasual, chino, boxer, slimfitjeans	140.000
3.	Kaosdantas.import	Kaos, kaos polo	120.000
4.	Ark_shop99	Kaos, chino	110.000
5.	Pieter.pieter	Kemeja kasual, slim fit jeans, chino	102.000
6.	Gns_collection	kaos	95.000
7.	Asy168	kaos	90.000
8.	Chikastore96	Kaos dalam	85.000
9.	Djakastore	Chino, kemeja kasual	80.000
10.	supplierhijabfashion	Kaos. kaos dalam, boxer	75.000

Dari tabel 4 dan 5 menjelaskan bahwa toko yang paling laris dari awal mula toko itu eksis (terdaftar di Shopee.id) hingga tanggal 4 April 2021 adalah pertama, toko Erigostore asal kota Bandung yang telah menjual lebih dari 400.000 produk. Jenis produknya berupa kaos, chino, jaket, celana panjang, sweater cardigan. Kedua, toko Marimofashions asal Jakarta barat yang telah menjual produk sebanyak 140.000 item, dengan jenis produknya berupa kemeja kasual, chino, boxer, slimfitjeans. Ketiga, toko Kaosdantas.import asal kota Pekalongan yang telah menjual lebih 120.000, dengan produk unggulannya berupa kaos dan kaos polo.

Sehingga hasil penelitian ini: pertama, dapat memberi masukan rekomendasi kepada toko/penjual baik yang lama maupun yang baru (*new commers*) untuk menjual *fashion pria* dengan kategori produk atasan (baju) dengan jenis produk kaos dan kemeja kasual serta kategori produk celana dengan jenis produk chino, slim fit jeans dan boxer. Kedua, ada p-ISSN:1693 – 2951; e-ISSN: 2503-2372



kemungkinan bagi pembeli akan mendapatkan potongan/ gratis ongkos kirim dikarenakan berasal dari kota Bandung, Jakarta Barat, Pekalongan.

V. KESIMPULAN

Penelitian ini berbeda dengan penelitian-penelitian sebelumnya karena belum ada penelitian yang melakukan kolaborasi antara metode *Web Scrapping* dan *Exploratory Data Analysis*. Penelitian yang dilakukan ini fokus kepada *marketplace* Shopee khususnya produk fashion pria. Serta untuk mengetahui lebih lanjut toko yang paling tren dan kota asal berdasarkan data produk yang paling banyak dibeli (paling banyak terjual)

Hasilnya adalah: pertama, toko yang paling laris (banyak terjual) produknya adalah Erigostore asal kota Bandung yang telah menjual lebih dari 400.000 produk dengan jenis produknya berupa kaos, chino, jaket, celana panjang, sweater cardigan. Kedua, produk yang paling laris terjual di Shopee.id adalah kategori produk atasan (baju) dengan jenis produk kaos dan kemeja kasual serta kategori produk celana dengan jenis produk chino, *slim fit jeans* dan boxer.

REFERENSI

- [1] M. el-K. Kesuma, I. Yunita, J. Fitra, N. A. Sholiha, and H. Oktaria, "Penerapan Slims Pada Layanan Sirkulasi Di Perpustakaan Instidla," *Al Maktab.*, vol. 6, no. 2, pp. 103–114, 2021.
- [2] I. Yunita, R. Pitri, and M. el-K. Kesuma, "Evaluasi Repositori UIN Raden Intan Lampung dengan Pendekatan Technology Acceptance Model (TAM) Evaluation," *N-JILS Nusant. J. Inf. Libr. Stud.*, vol. 4, no. 1, pp. 61–74, 2021.
- [3] M. Lucky and N. Rosmadi, "Penerapan Strategi B isnis di Masa Pandemi Covid-19 Jurnal IKRA-ITH Ekonomika Vol 4 No 1 Bulan Maret 2021," *J. IKRA-ITH Ekon.*, vol. 4, no. 1, pp. 122–127, 2021.
- [4] G. M. Arya Sasmita, P. A. Bayupati, and Gerlan A. Manu, "Sistem Transaksi Penjualan Kain Gringsing Berbasis Mobile (Mobile Commerce)," *Commun. ACM*, vol. 8, no. 1, pp. 91–96, 2009, [Online]. Available: <http://portal.acm.org/citation.cfm?doid=953460.953483>.
- [5] E. Japarianto and S. Adelia, "Pengaruh Tampilan Web Dan Harga Terhadap Minat Beli Dengan Kepercayaan Sebagai Intervening Variable Pada E-Commerce Shopee," *J. Manaj. Pemasar.*, vol. 14, no. 1, pp. 35–43, 2020, doi: 10.9744/pemasaran.14.1.35-43.
- [6] H. Rohmah and Z. A. Fataron, "Effect Analysis of Trust, Ease, Information Quality, Halal Product on Online Purchase Decision of 2016-2018 Batch Students of Islamic Economics Study Program in UIN Walisongo at Shopee Marketplace," *J. Digit. Mark. Halal Ind.*, vol. 1, no. 1, p. 1, 2020, doi: 10.21580/jdmhi.2019.1.1.4772.
- [7] A. K-means, Y. D. Putra, M. Sudarma, I. Bagus, and A. Swamardika, "Cluster ing History Data Penjualan Menggunakan," *Maj. Ilm. Teknol. Elektro*, vol. 20, no. 2, 2021.
- [8] M. Mirza and H. Ali, "Strategi E-Commerce Shoppe: Dalam Konteks Bisnis Etik," *J. Manajemen, Vol 10, No 2*, vol. 7, no. 1, pp. 99–104, 2020.
- [9] T. Nawangsari, "Analisis Data Eksploratif Prestasi Mahasiswa Pada Mata Kuliah Statistika Matematika II," *Pros. Semin. Nas. Has. Penelit. dan Pengabd. Kpd. Masy. I*, no. September, 2017.
- [10] V. Mitra, H. Sujaini, and A. B. P. Negara, "Rancang Bangun Aplikasi Web Scraping Untuk Korpus Paralel Indonesia - Inggris dengan Metode HTML DOM," *JUSTIN (Jurnal Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 36–41, 2017.
- [11] D. D. Ayani, H. S. Pratiwi, and H. Muhardi, "Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace," *JUSTIN (Jurnal Sist. dan Teknol. Informasi)*, vol. 7, no. 4, p. 257, 2019, doi: 10.26418/justin.v7i4.30930.
- [12] Y. Sahria, "Implementasi Teknik Web Scraping pada Jurnal SINTA Untuk Analisis Topik Penelitian Kesehatan Indonesia," in *URECOL (University Research Colloquium)*, 2020, pp. 297–306, [Online]. Available: <http://repository.urecol.org/index.php/proceeding/article/view/1079>.
- [13] M. el-K. Kesuma, "Hasil analisis berita melalui google dan website shopee," <http://shopee.com/>, 2022. <http://shopee.com/>.
- [14] E. S. Sulistiyawati and A. Widayani, "Marketplace Shopee Sebagai Media Promosi Penjualan UMKM di Kota Blitar," *J. Pemasar. Kompetitif*, vol. 4, no. 1, p. 133, 2020, doi: 10.32493/jpkpk.v4i1.7087.
- [15] V. A. Flores, P. A. Permatasari, and L. Jasa, "Penerapan Web Scraping Sebagai Media Pencarian dan Menyimpan Artikel Ilmiah Secara Otomatis Berdasarkan Keyword," *Maj. Ilm. Teknol. Elektro*, vol. 19, no. 2, p. 157, 2020, doi: 10.24843/mite.2020.v19i02.p06.
- [16] A. Josi, L. A. Abdillah, and Suryayusra, "Penerapan teknik web scraping pada mesin pencari artikel ilmiah," *J. Sist. Inf.*, vol. 5, no. 1, pp. 159–164, 2014, [Online]. Available: <http://arxiv.org/abs/1410.5777>.
- [17] D. F. Setiawan, T. Tristiyanto, and A. Hijriani, "Aplikasi Web Scraping Deskripsi Produk," *J. Teknoinfo*, vol. 14, no. 1, p. 41, 2020, doi: 10.33365/jti.v14i1.498.
- [18] L. Hayati, Hapipi, and S. Azmi, "Pelatihan Pemanfaatan Perangkat Lunak Tinkerplots Untuk Analisis Data Eksploratif Pada Pembelajaran Statistika Bagi Guru-Guru Matematika SMP/ SMA Di Gunung Sari Lombok Barat," *Angew. Chemie Int. Ed. 6(11)*, 951–952., vol. 1, no. 2, pp. 5–24, 2018.
- [19] Beny, "Exploratory Data Analysis Pada Database Sistem Informasi Akademik Stikom Dinamika Bangsa," *J. Ilm. Media Process.*, vol. 11, no. 2, pp. 752–761, 2016.
- [20] I. H. Muna, "Exploratory Data Analysis Grup Chat Whatsapp Mata Kuliah Manajemen Pembiayaan Pendidikan (Studi Kasus di Pascasarjana Universitas Negeri Semarang)," *J. Litbang Eduaintechn*, vol. 2, no. 1, pp. 44–49, 2021.