

**PERBANDINGAN PENJUALAN PRODUK HALAL *LABELED*  
DAN *NON-LABELED* PADA *E-COMMERCE* TOKOPEDIA  
INDONESIA**



Disusun Oleh:

N a m a : Apriza Zicka Rizquina

NIM : 19523211

**PROGRAM STUDI INFORMATIKA – PROGRAM SARJANA  
FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS ISLAM INDONESIA**

**2023**

HALAMAN PENGESAHAN DOSEN PEMBIMBING

**PERBANDINGAN PENJUALAN PRODUK HALAL *LABELED*  
DAN *NON-LABELED* PADA *E-COMMERCE* TOKOPEDIA  
INDONESIA**

**TUGAS AKHIR**



N a m a : Apriza Zicka Rizquina  
NIM : 19523211

الجامعة الإسلامية  
الابستد الاندو

Yogyakarta, 12 Oktober 2023

Pembimbing,

( Chanifah Indah Ratnasari, S.Kom., M.Kom. )

## HALAMAN PENGESAHAN DOSEN PENGUJI

**PERBANDINGAN PENJUALAN PRODUK HALAL *LABELED*  
DAN *NON-LABELED* PADA *E-COMMERCE* TOKOPEDIA  
INDONESIA**

## TUGAS AKHIR

Telah dipertahankan di depan sidang penguji sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer dari Program Studi Informatika – Program Sarjana di Fakultas Teknologi Industri Universitas Islam Indonesia

Yogyakarta, 3 November 2023

Tim Penguji

Chanifah Indah Ratnasari, S.Kom., M.Kom.

**Anggota 1**

Galang Prihadi Mahardhika, S.Kom., M.Kom.

**Anggota 2**

Mukhammad Andri Setiawan, S.T., M.Sc., Ph.D.

Mengetahui,

Ketua Program Studi Informatika – Program Sarjana  
Fakultas Teknologi Industri  
Universitas Islam Indonesia



(Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D.)

**HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR**

Yang bertanda tangan di bawah ini:

Nama : Apriza Zicka Rizquina

NIM : 19523211

Tugas akhir dengan judul:

**PERBANDINGAN PENJUALAN PRODUK HALAL *LABELED*  
DAN *NON-LABELED* PADA *E-COMMERCE* TOKOPEDIA  
INDONESIA**

Menyatakan bahwa seluruh komponen dan isi dalam tugas akhir ini adalah hasil karya saya sendiri. Apabila di kemudian hari terbukti ada beberapa bagian dari karya ini adalah bukan hasil karya sendiri, tugas akhir yang diajukan sebagai hasil karya sendiri ini siap ditarik kembali dan siap menanggung risiko dan konsekuensi apapun.

Demikian surat pernyataan ini dibuat, semoga dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 12 Oktober 2023

A 10,000 Indonesian postage stamp with a QR code and a signature. The stamp is pink and white, featuring the Garuda Pancasila emblem and the text 'METERA 10000'. The signature 'Zicka' is written in black ink over the stamp.

( Apriza Zicka Rizquina )

## HALAMAN PERSEMBAHAN

*Alhamdulillah Rabbil'alamin*, tugas akhir ini saya persembahkan untuk kedua orang tua dan adik-adik saya. Saya bersyukur memiliki keluarga yang selalu memberikan dukungan moral dan motivasi. Terima kasih atas kesabaran dan pengertian yang diberikan. Terima kasih juga saya ucapkan kepada semua yang turut serta dalam penelitian ini. Semua kontribusi sangat berarti bagi kesuksesan penelitian ini.

**HALAMAN MOTO**

“Never been a natural, all I do is try, try, try”

TS

## KATA PENGANTAR

*Assalamualaikum warahmatullahi wabarakatuh.*

*Alhamdulillah Rabbil' alamin*, puji Syukur kehadiran Allah SWT, yang telah melimpahkan Rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan penyusunan tugas akhir ini dengan judul “Perbandingan Penjualan Produk Halal *Labeled* dan *Non-Labeled* pada *E-commerce* Tokopedia Indonesia”. Tugas akhir ini disusun sebagai salah satu syarat untuk mendapatkan gelar kesarjanaan pada jurusan Informatika Universitas Islam Indonesia. Penulis mengucapkan terima kasih kepada seluruh pihak yang telah memberikan dukungan, bimbingan, dan motivasi dalam perjalanan penulisan tugas akhir ini. Pada kesempatan ini penulis juga ingin mengucapkan rasa terima kasih sebesar-besarnya kepada:

1. Allah SWT atas segala nikmat dan rahmat-Nya yang diberikan selama ini.
2. Ayah dan Mama selaku orang tua yang selalu mendukung penuh secara materi dan non-materi, serta memberikan doa dan restu sehingga saya dapat menyelesaikan tugas akhir dengan baik.
3. Uti, Keisha, dan Qeela selaku adik-adik yang selalu memberikan semangat, menghibur, dan menjadi motivasi untuk menyelesaikan tugas akhir.
4. Bapak Prof. Fathul Wahid, S.T., M.Sc., Ph.D., selaku Rektor Universitas Islam Indonesia.
5. Bapak Dr. Raden Teduh Dirgahayu, S.T., M.Sc., selaku Ketua Jurusan Informatika Universitas Islam Indonesia.
6. Bapak Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D., selaku Ketua Progam Studi Informatika Program Sarjana Universitas Islam Indonesia.
7. Ibu Arrie Kurniawardhani, S.Si., M.Kom., selaku Dosen Pembimbing Akademik (DPA) yang telah membimbing saya selama perkuliahan di Informatika Universitas Islam Indonesia.
8. Ibu Chanifah Indah Ratnasari, S.Kom., M.Kom., selaku Dosen Pembimbing tugas akhir yang telah meluangkan waktu, selalu memberikan arahan serta dukungan dalam proses bimbingan.
9. Bapak dan Ibu dosen Jurusan Informatika yang telah memberikan ilmu dan pengetahuan yang bermafaat.
10. Keluarga besar yang selalu memberikan semangat sehingga saya dapat menyelesaikan tugas akhir dengan baik.

11. Diva, Sallu, Adis, Rania, Rima, Dyah, Minek, Via, Iiq, dan teman-teman saya yang telah menemani saya selama ini diwaktu senang maupun susah.
12. Teman-teman seperjuangan Informatika Angkatan 2019 yang sedang berjuang untuk memperoleh gelar S.Kom., yang telah memberikan pengalaman selama kuliah.
13. Terakhir untuk diri saya sendiri, Apriza Zicka Rizquina yang telah berjuang dan pantang menyerah untuk mendapatkan gelar S.Kom.

Saya menyadari bahwa tugas akhir ini masih ada kekurangan baik dari segi susunan kalimat maupun tata bahasa. Namun saya berharap semoga tugas akhir ini dapat diterima dan bermanfaat bagi prmbaca dan semua pihak.

***Wassalamu'alaikum warahmatullahi wabarakatuh.***

Yogyakarta, 12 Oktober 2023



( Apriza Zicka Rizquina )

## SARI

Beredarnya produk makanan dan minuman kemasan yang tanpa label halal di Indonesia berpotensi menjadi masalah di masa mendatang. Label halal menjadi penanda penting sebuah produk untuk menjamin produk tersebut telah melalui pengujian kehalalan yang sesuai. Terlebih dengan adanya *e-commerce* yang memiliki peran besar dalam transaksi jual-beli secara *online*. Tokopedia Indonesia merupakan salah satu platform *e-commerce* populer di Indonesia, menawarkan berbagai produk termasuk produk makanan dan minuman dalam kemasan. Selain itu, Tokopedia juga menjual produk tanpa label halal dengan tingkat penjualan yang tinggi. Oleh karena itu, penelitian ini penting untuk dilaksanakan guna membandingkan penjualan produk berlabel halal dan tidak pada produk makanan dan minuman kemasan. Penelitian ini menggunakan pengujian terhadap data yang diperoleh melalui metode *web scraping* dengan menerapkan teknik *HTML Parsing* dari *e-commerce* Tokopedia. Pengujian dilakukan dengan menggunakan 2 uji statistik yang terdiri dari uji normalitas Kolmogorov-Smirnov dan uji nonparametrik Mann-Whitney U dengan tujuan untuk menilai perbedaan penjualan antara dua kelompok yaitu produk berlabel halal dan tidak. Hasil penelitian menunjukkan bahwa  $p\text{-value} < 0.05$  yang mengindikasikan adanya bukti yang cukup kuat untuk menolak hipotesis nol ( $H_0$ ) yang menyatakan tidak terdapat perbedaan yang signifikan dalam hasil penjualan produk antara produk yang diberi label halal dan tidak di *e-commerce* Tokopedia Indonesia dan menerima hipotesis alternatif ( $H_a$ ) yang menyatakan terdapat perbedaan yang signifikan dalam hasil penjualan produk antara produk yang diberi label halal dan tidak di *e-commerce* Tokopedia Indonesia.

Kata kunci: Label Halal; Penjualan; *E-commerce*; *Web Scraping*; Uji Statistik

## GLOSARIUM

<i>Dataset</i>	Kumpulan data yang digunakan untuk analisis.
EDA	<i>Exploratory Data Analysis</i> .
<i>Library</i>	Kumpulan kode dengan fungsi tertentu.
<i>Preprocessing</i>	Tahapan untuk merapikan data.
<i>P-value</i>	Ukuran statistik yang digunakan dalam uji hipotesis.
<i>Web scraping</i>	Metode yang digunakan untuk mengestrak data dari sebuah website yang kemudian disimpan ke dalam sebuah <i>file</i> .

## DAFTAR ISI

HALAMAN JUDUL .....	i
HALAMAN PENGESAHAN DOSEN PEMBIMBING .....	ii
HALAMAN PENGESAHAN DOSEN PENGUJI .....	iii
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR .....	iv
HALAMAN PERSEMBAHAN .....	v
HALAMAN MOTO .....	vi
KATA PENGANTAR .....	vii
SARI .....	ix
GLOSARIUM .....	x
DAFTAR ISI .....	xi
DAFTAR TABEL .....	xiii
DAFTAR GAMBAR .....	xiv
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan Masalah .....	3
1.4 Tujuan Penelitian .....	3
1.5 Manfaat Penelitian .....	3
1.6 Sistematika Penulisan .....	3
<b>BAB II LANDASAN TEORI</b> .....	<b>5</b>
2.1 Tokopedia .....	5
2.2 Sertifikasi dan Labelisasi Halal .....	5
2.3 <i>Web Scraping</i> .....	7
2.3.1 Kajian Pustaka Teknik <i>Web Scraping</i> .....	7
2.3.2 Selenium WebDriver .....	10
2.4 Metode Statistik .....	11
2.5 Uji Statistik .....	11
2.5.1 Uji Normalitas .....	12
2.5.2 Uji Mann-Whitney U .....	13
2.6 <i>Preprocessing</i> .....	13
2.6.1 <i>Data Cleaning</i> .....	14
2.6.2 <i>Data Transformation</i> .....	14
2.7 <i>Exploratory Data Analysis</i> .....	14
2.8 <i>Interquartile Range (IQR)</i> .....	15
2.9 Python .....	15
2.10 SPSS .....	16
2.11 Kajian Pustaka Uji Statistik Nonparametrik .....	16
<b>BAB III METODOLOGI PENELITIAN</b> .....	<b>20</b>
3.1 Tahapan Penelitian .....	20
3.2 Pengumpulan Data .....	21
3.3 <i>Preprocessing</i> .....	23
3.4 <i>Exploratory Data Analysis (EDA)</i> .....	24
3.5 Uji Statistik .....	24
3.5.1 Uji Mann-Whitney U .....	24
3.5.2 Uji Kolmogorov-Smirnov .....	25
3.6 Pengujian menggunakan <i>software</i> SPSS .....	26
3.7 Analisis Hasil .....	26

BAB IV HASIL DAN PEMBAHASAN.....	27
4.1 Pengumpulan Data .....	27
4.2 <i>Preprocessing</i> .....	32
4.2.1 Iterasi 1 <i>Data Cleaning</i> .....	33
4.2.2 <i>Data Transformation</i> .....	35
4.2.3 Iterasi 2 <i>Data Cleaning</i> .....	35
4.3 <i>Exploratory Data Analysis</i> (EDA).....	37
4.4 Uji Statistik Menggunakan Python .....	44
4.4.1 Uji Mann-Whitney U.....	45
4.4.2 Uji Kolmogorov-Smirnov .....	46
4.5 Uji Statistik Menggunakan SPSS.....	47
4.5.1 Uji Mann-Whitney U.....	47
4.5.2 Uji Kolmogorov-Smirnov .....	48
4.6 Hasil Analisis .....	48
BAB V KESIMPULAN DAN SARAN .....	50
5.1 Kesimpulan .....	50
5.2 Saran.....	50
DAFTAR PUSTAKA.....	51
LAMPIRAN .....	58

**DAFTAR TABEL**

Tabel 2.1 Tabel Perbandingan Teknik dalam <i>Web scraping</i> .....	8
Tabel 2.2 Tabel Perbandingan Pengaplikasian Uji Statistik .....	12
Tabel 2.3 Tabel Perbandingan metode dalam uji statistik nonparametrik .....	18
Tabel 4.1 Lokasi Penjualan .....	41
Tabel 4.2 Hasil uji Mann-Whitney U .....	46
Tabel 4.3 Hasil uji Kolmogorov-Smirnov .....	47
Tabel 4.4 Hasil uji Mann-Whitney U SPSS .....	48
Tabel 4.5 Hasil uji Kolmogorov-Smirnov SPSS .....	48

## DAFTAR GAMBAR

Gambar 2.1 (a) Logo Halal LPPOM MUI; (b) Logo Halal BPJPH .....	6
Gambar 2.2 Alur proses sertifikasi halal.....	6
Gambar 3.1 Contoh video ulasan produk di TikTok .....	22
Gambar 4.1 Kode program untuk <i>import libraries</i> .....	27
Gambar 4.2 Kode program untuk membuka halaman Chrome .....	28
Gambar 4.3 Tampilan halaman situs web Shopee Indonesia .....	28
Gambar 4.4 Tampilan halaman situs web Tokopedia.....	29
Gambar 4.5 Kode sumber web.....	30
Gambar 4.6 Kode program untuk pengambilan data .....	31
Gambar 4.7 Kode program untuk pelabelan dengan menambah variabel baru.....	31
Gambar 4.8 Kode program untuk menyimpan data.....	31
Gambar 4.9 Contoh <i>web scraping</i> dalam format <i>file</i> CSV .....	32
Gambar 4.10 Kode program <i>data cleaning</i> .....	34
Gambar 4.11 <i>Output</i> kode program <i>data cleaning</i> .....	34
Gambar 4.12 <i>Output</i> kode program <i>data cleaning</i> .....	34
Gambar 4.13 Kode <i>data transformation</i> .....	35
Gambar 4.14 <i>Output</i> kode program <i>data transformation</i> .....	35
Gambar 4.15 Kode program interkuartil (IQR) .....	36
Gambar 4.16 (a) <i>Boxplot</i> variabel “Harga” sebelum dilakukan penghapusan <i>outlier</i> ; (b) <i>Boxplot</i> variabel “Harga” setelah dilakukan penghapusan <i>outlier</i> .....	36
Gambar 4.17 (a) <i>Boxplot</i> variabel “Terjual” sebelum dilakukan penghapusan <i>outlier</i> ; (b) <i>Boxplot</i> variabel “Terjual” setelah dilakukan penghapusan <i>outlier</i> .....	37
Gambar 4.18 Ringkasan statistik dari nilai numerik.....	37
Gambar 4.19 Kode program visualisasi <i>countplot</i> .....	38
Gambar 4.20 Perbandingan jumlah produk berlabel halal dan tidak.....	38
Gambar 4.21 Kode program visualisasi diagram.....	39
Gambar 4.22 Perbandingan total penjualan produk berlabel halal dan tidak .....	39
Gambar 4.23 Kode program visualisasi diagram.....	40
Gambar 4.24 Perbandingan rata-rata harga produk berdasarkan label .....	40
Gambar 4.25 Kode program cek lokasi .....	41
Gambar 4.26 <i>Output</i> kode program cek lokasi .....	41
Gambar 4.27 Kode program visualisasi <i>stacked bar plot</i> .....	42

Gambar 4.28 <i>Stacked bar plot</i> proporsi produk terjual berdasarkan label dan lokasi .....	43
Gambar 4.29 Kode program visualisasi <i>heatmap</i> .....	44
Gambar 4.30 <i>Heatmap</i> korelasi antarvariabel .....	44
Gambar 4.31 Kode program hapus variabel .....	44
Gambar 4.32 <i>Output</i> lima data dalam <i>dataframe</i> .....	45
Gambar 4.33 Kode program uji Mann-Whitney U .....	45
Gambar 4.34 Kode program uji Kolmogorov-Smirnov.....	47

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Indonesia merupakan negara berkembang dengan mayoritas penduduk muslim. Berdasarkan *website* [www.dataindonesia.org](http://www.dataindonesia.org) jumlah penduduk muslim di Indonesia mencapai 237 juta jiwa per 31 Desember 2021 yaitu setara dengan 89,9% dari populasi keseluruhan penduduk di Indonesia (Bayu, 2022). Seiring dengan banyaknya populasi penduduk muslim, maka kebutuhan akan permintaan produk dan layanan halal juga semakin meningkat (Juniwati & Listiawati, 2021). Untuk memudahkan dalam mengenali makanan yang dikonsumsi halal atau tidak, yaitu dengan melihat label halal yang tercantum pada produk kemasan (Murni & Fajrina, 2021). Selain label halal, terdapat sertifikasi halal yang menjadi pengakuan kehalalan produk yang dikeluarkan oleh Badan Penyelenggara Jaminan Produk Halal (BPJPH) (Warto & Samsuri, 2020). Pemberian label halal diperlukan untuk menarik serta meyakinkan konsumen sebagai dasar atau standar produk tersebut sudah sesuai syariat Islam serta sebagai standar jaminan kualitas, kebersihan, kesehatan, serta keselamatan atas apa yang mereka jual (Pramintasari & Fatmawati, 2017).

Kesadaran halal dalam memilih suatu produk yang akan dibeli sangat penting bagi seorang muslim (Alinda & Adinugraha, 2022). Kesadaran halal pada masyarakat memiliki pengaruh terhadap keputusan pembelian produk (Juliana et al., 2022). Saat ini ketersediaan produk di Indonesia yang sudah tersertifikasi halal hanya sekitar 20% (Setyaningsih & Marwansyah, 2019). Beberapa pelaku usaha juga masih menjual produk yang bahkan tidak memenuhi persyaratan berupa izin edar yang berpengaruh pada pengabaian hak yang dimiliki oleh pembeli seperti informasi yang benar dan kejelasan produk (Sagita & Saputra, 2022). Ditambah adanya produk impor yang kini masuk ke Indonesia dengan berbagai jenis kemasan yang menarik mengharuskan masyarakat perlu berhati-hati dalam memilih produk yang akan dikonsumsi (Al-Bara & Nasution, 2018). Oleh karena itu, penting adanya label halal dan izin edar dari BPOM untuk menjamin produk, terutama makanan. Namun demikian, belum semua masyarakat sadar terhadap adanya label halal pada kemasan tersebut (Al-Bara & Nasution, 2018).

Saat ini, penjualan produk ataupun jasa dapat dengan mudah dilakukan dengan bantuan teknologi yaitu *e-commerce* (Lupi & Nurdin, 2016). Menurut Sulistiyawati & Widayani, (2020), *e-commerce* merupakan aplikasi yang digunakan untuk melakukan aktivitas jual beli secara *online* serta sebagai media untuk mempromosikan produk. Tokopedia merupakan salah satu *e-commerce* populer yang telah berdiri sejak 2009 dan menjadi salah satu *marketplace* yang paling sering dikunjungi di Indonesia (Kurniawan et al., 2021). Tokopedia sering kali melakukan promosi guna meningkatkan penjualan produk (Arsta & RsiRespati, 2021). Strategi promosi yang dilakukan yaitu seperti pemberian diskon, kupon, koin emas virtual, jaminan uang kembali, dan asuransi pengiriman gratis (Zhao et al., 2020).

Cahya et al. (2022) melakukan penelitian untuk menganalisis pengaruh sertifikasi halal serta kesadaran halal terhadap minat beli produk makanan Korinus Tok-poki menggunakan metode analisis regresi linier berganda, uji instrumen, analisis deskriptif, uji normalitas, uji asumsi klasik, dan uji hipotesis. Penelitian tersebut menghasilkan temuan yaitu sertifikasi halal dan kesadaran halal secara parsial memiliki pengaruh positif terhadap minat pembeli. Puspitasari (2019) berdasarkan penelitiannya menemukan bahwa label halal juga berpengaruh kepada peningkatan penjualan kosmetik. Sementara itu dalam penelitian Resmawati (2018) menggunakan metode kualitatif mendapatkan hasil rata-rata volume penjualan sebelum adanya sertifikasi halal yaitu sekitar 33,03% dan setelah adanya sertifikasi halal sebesar 21,11%. Namun hal itu tidak dapat dijadikan kesimpulan mengenai turunnya angka penjualan sebab sertifikasi halal tersebut baru berjalan sekitar 1 tahun.

Berdasarkan latar belakang yang telah disampaikan maka akan dilakukan penelitian yang bertujuan untuk mengetahui perbandingan penjualan produk label halal dan tidak menggunakan metode statistik. Dalam hal ini data yang digunakan adalah data *e-commerce* Tokopedia pada produk makanan dan minuman dalam kemasan yang dijual di Indonesia. Pengambilan data pada situs web *e-commerce* dalam jumlah yang besar dapat dilakukan menggunakan metode *web scraping*. Penggunaan *web scraping* bertujuan agar informasi yang diperoleh dapat lebih terfokus sehingga akan memudahkan dalam melakukan pencarian data yang diperlukan (A. Yani et al., 2019). Dengan demikian, hasil penelitian ini dapat dimanfaatkan oleh lembaga pemerintah terkait untuk mengambil kebijakan selanjutnya dan dapat dimanfaatkan oleh pihak industri dalam perencanaan strategi penjualan ataupun penelitian sejenis ke depannya.

## 1.2 Rumusan Masalah

Rumusan masalah pada penelitian ini, yaitu bagaimana perbandingan penjualan *online* produk berlabel halal dan tidak di Indonesia pada *e-commerce* Tokopedia Indonesia?

## 1.3 Batasan Masalah

Batasan masalah pada penelitian ini yaitu:

1. Data produk yang digunakan berasal dari *e-commerce* Tokopedia Indonesia.
2. Jenis produk yang digunakan yaitu produk makanan dan minuman dalam kemasan yang dijual di *e-commerce* Tokopedia Indonesia.
3. Produk berlabel halal harus sudah terdaftar dalam situs web MUI (<https://halalmui.org/>).

## 1.4 Tujuan Penelitian

Tujuan penelitian ini yaitu untuk mengetahui perbandingan penjualan *online* produk berlabel halal dan tidak pada *e-commerce* Tokopedia Indonesia, sehingga dapat diketahui kesadaran masyarakat Indonesia yang notabene negara berpenduduk mayoritas muslim terhadap pembelian produk berlabel halal. Hasil penelitian ini diharapkan dapat dimanfaatkan oleh lembaga atau instansi terkait ataupun penelitian sejenis ke depannya.

## 1.5 Manfaat Penelitian

Manfaat penelitian ini yaitu:

1. Memberikan fakta berdasarkan data terkini mengenai perbandingan penjualan produk *online* yang sudah berlabel halal dan tidak.
2. Bermanfaat untuk pemerintah maupun dalam bidang industri terkait atau penelitian sejenis.

## 1.6 Sistematika Penulisan

Sistematika penulisan dalam penyusunan laporan tugas akhir ini terdiri dari lima bab, yaitu:

## BAB I PENDAHULUAN

Bab ini berisi pembahasan latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penelitian.

## BAB II LANDASAN TEORI

Bab ini berisi pembahasan mengenai teori dasar terhadap penelitian yang pernah ada berhubungan dengan penelitian ini.

## BAB III METODOLOGI

Bab ini berisi uraian tentang langkah-langkah yang dilakukan dalam penelitian ini.

## BAB IV HASIL

Bab ini membahas tentang hasil dari penelitian yang sudah dilakukan serta analisis hasil dari penelitian tersebut.

## BAB V KESIMPULAN DAN SARAN

Bab ini merupakan bab terakhir yang akan membahas mengenai kesimpulan dan saran terhadap penelitian yang telah dilakukan.

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Tokopedia**

Tokopedia merupakan salah satu platform jual beli *online* yang menghubungkan antara penjual dan pembeli di Indonesia yang juga memiliki beberapa layanan yang dapat melengkapi kebutuhan konsumen (Nasri, 2020). Sejak berdiri pada tahun 2009, Tokopedia berhasil mencapai kesuksesan dengan pertumbuhan yang pesat, pendapatan yang terus meningkat, dan menjadi *e-commerce* paling populer yang sering dikunjungi masyarakat Indonesia (Kurniawan et al., 2021). Berdasarkan survei yang dilakukan Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) yang dilakukan terhadap 1000 UMKM dan korporasi yang tersebar di Indonesia dalam periode 24 Januari–9 Maret 2023, sebanyak 73,73% pelaku UMKM lebih memilih *e-commerce* Tokopedia sebagai lapak untuk berjualan (Septiani, 2023). Kepercayaan dan antusiasme yang tinggi dari masyarakat terhadap penggunaan layanan belanja melalui *e-commerce* membuktikan bahwa dengan menggunakan platform ini semakin memudahkan aktivitas jual-beli, serta memiliki sejumlah penawaran dengan keunggulan menarik, dan memberikan pengalaman belanja yang lebih modern dibandingkan dengan belanja konvensional (Wardhana et al., 2021).

#### **2.2 Sertifikasi dan Labelisasi Halal**

Menurut standar Lembaga Pengkajian Pangan, Obat-obatan, dan Kosmetika Majelis Ulama Indonesia (LPPOM MUI) dalam penelitian yang dilakukan oleh Wartyo dan Samsuri (2020), sertifikasi halal merupakan proses untuk mendapatkan sertifikat halal yang harus melalui beberapa tahapan pemeriksaan untuk pembuktian bahan baku, proses produksi, dan sistem jaminan halal produk yang sudah sesuai dengan standar yang telah ditetapkan. Sedangkan label halal digunakan untuk memberikan kepastian kehalalan suatu produk pangan, obat-obatan, dan kosmetika kepada pengguna produk tersebut (Al-Bara & Nasution, 2018).

Pada tanggal 10 Februari 2022, Badan Penyelenggara Jaminan Produk Halal (BPJPH) Kementerian Agama mengganti logo halal LPPOM MUI dengan logo halal baru setelah adanya peralihan otoritas sertifikasi halal di Indonesia berdasarkan UU No. 33 Tahun 2014 dan PP No. 39 Tahun 2021 (Rachman & Ulpah, 2022). Logo halal lama oleh LPPOM MUI dan logo halal baru oleh BPJPH dapat dilihat pada Gambar 2.1.



Gambar 2.1 (a) Logo Halal LPPOM MUI; (b) Logo Halal BPJPH

Sumber: (a) LPPOM MUI; (b) BPJPH

Alur proses sertifikasi halal di Indonesia dapat dilihat pada Gambar 2.2 yang bersumber dari situs web resmi LPPOM MUI. Adapun penetapan kehalalan produk tetap dilakukan oleh MUI dibuktikan dengan adanya rapat komisi fatwa MUI yang dilanjutkan dengan penerbitan sertifikasi halal oleh BPJPH.



Gambar 2.2 Alur proses sertifikasi halal

Sumber: LPPOM MUI

## 2.3 Web Scraping

*Web scraping* merupakan teknik *Natural Language Processing* (NLP) yang digunakan untuk mengekstrak data dalam jumlah besar dari situs web dan menyimpannya dalam format *file* lokal atau basis data dalam bentuk tabel yang dapat digunakan untuk analisis (Rizaldi & Arief, 2017). Proses pengambilan data dilakukan pada dokumen yang bersifat semi terstruktur dari internet yang berupa halaman-halaman web yang berbentuk HTML (Sahria, 2020). Pengambilan data menggunakan *web scraping* memiliki beberapa teknik yang dapat diaplikasikan mulai dari *web scraping* manual hingga *web scraping* otomatis (*automated web scraping*). Teknik *web scraping* tradisional seperti *traditional copy-pasting*, menggunakan pendekatan manual di mana data yang harus diekstraksi dari situs web disalin secara manual sebagai grup dan ditempelkan ke dalam dokumen (Thota & Ramez, 2021). Teknik ini biasa digunakan untuk pengumpulan data dalam jumlah yang sedikit (Darmawan et al., 2022). Selain itu, terdapat teknik *web scraping* otomatis yang dapat diterapkan untuk pengambilan data dengan jumlah yang banyak secara otomatis (Onyenwe et al., 2021).

### 2.3.1 Kajian Pustaka Teknik Web Scraping

Untuk menentukan teknik pengambilan data yang cocok digunakan dalam penelitian ini, dilakukan studi literatur terhadap penelitian terdahulu sebagai perbandingan teknik pengambilan data. Kata kunci yang digunakan dalam pencarian sumber di situs web Google Scholar adalah “*web scraping*”, “*web scraping Python*”, dan “*web scraping e-commerce*” dengan pengaturan filter dari tahun 2017 hingga saat ini. Berdasarkan literatur yang diperoleh, diambil lima penelitian yang paling sesuai untuk dirangkum dalam laporan ini.

Pertama, penelitian yang dilakukan oleh Rizaldi dan Arief (2017) yang berjudul “Perbandingan Metode *Web Scraping* Menggunakan *CSS Selector* dan *Xpath Selector*”. Penelitian ini melakukan pengambilan data menggunakan *web scraping* dari situs blog Detik (<http://blog.detik.com/>). Berdasarkan hasil evaluasi performa *web scraping*, diperoleh temuan bahwa metode Regex memiliki penggunaan CPU dan memori yang paling sedikit. Lalu, metode XPath memakan waktu tersingkat dalam *scraping* dibandingkan metode lainnya. Sementara itu, metode dengan penggunaan *bandwidth* paling kecil yaitu dengan menggunakan metode *CSS Selector*.

Kedua, penelitian yang dilakukan oleh Elveny et al. (2021) yang berjudul “*Web-based E-commerce Products Grouping*”. Dalam penelitian tersebut memiliki tahapan untuk pengambilan data produk *e-commerce* dengan menggunakan *HTML Parsing* yang dapat

mendeteksi nama produk, harga produk, angka ulasan, dan URL produk pertama hingga keempat. *Library* yang digunakan yaitu BeautifulSoup yang merupakan salah satu *library* Python yang memiliki kemampuan HTML *Parsing*.

Ketiga, penelitian yang dilakukan oleh Gunawan et al. (2019) berjudul “*Comparison of Web Scraping Techniques: Regular Expression, HTML DOM, and XPath*”. Penelitian ini bertujuan untuk membandingkan tiga teknik *web scraping* dalam pengambilan data situs web (<http://testing-ground.scraping.pro>). Parameter untuk perbandingan yang digunakan yaitu waktu proses, penggunaan memori, dan penggunaan data sebagai parameter pengukuran dalam percobaan. Percobaan dengan metode *Regular Expression* memiliki hasil penggunaan memori paling kecil dibandingkan dengan metode HTML DOM, dan XPath. Sedangkan HTML DOM membutuhkan waktu paling sedikit dan penggunaan data paling kecil dibandingkan dengan metode *Regular Expression* dan XPath.

Keempat, penelitian yang dilakukan oleh Nair et al. (2021) berjudul “*Comparative Review on Sentiment Analysis-Based Recommendation System*”. Penelitian ini melakukan analisis sentimen dari data *review e-commerce* melalui web media sosial. Untuk mendapatkan data tersebut dilakukan ekstraksi data atau pengumpulan data dari web yang dituju. Dijelaskan bahwa ekstraksi data dapat dilakukan dengan algoritma *Machine Learning (Rapier)* dan *Sequence Rules with Validation (SRV)* dengan menggunakan teknik HTML *Parsing* serta memiliki ketentuannya sendiri untuk mengekstrak data dari halaman web secara lebih efisien.

Kelima, penelitian Thota dan Ramez (2021) berjudul “*Web scraping of COVID-19 News Stories to Create Datasets for Sentiment and Emotion Analysis*” yang melakukan pengambilan data menggunakan *web scraping* yang akan digunakan untuk analisis sentimen dan emosi pada situs web berita CNN. Hasil yang didapat berdasarkan fleksibilitas, dan mengetahui kesulitan penggunaannya, dapat disimpulkan bahwa HTML *Parsing* adalah teknik yang paling baik digunakan pada web ini.

Tabel 2.1 Tabel Perbandingan Teknik dalam *Web scraping*

No	Tahun	Judul	Teknik	Keterangan
1	2017	Perbandingan Metode <i>Web scraping</i> Menggunakan CSS Selector dan XPath Selector	CSS Selector, XPath Selector	Metode XPath memakan waktu yang tersingkat. Sementara itu, untuk penggunaan <i>bandwidth</i> paling kecil

No	Tahun	Judul	Teknik	Keterangan
				yaitu dengan menggunakan metode <i>CSS Selector</i> .
2	2021	<i>Web-based E-commerce Products Grouping</i>	<i>HTML Parsing</i>	Pengambilan data menggunakan metode <i>web scraping</i> dengan teknik <i>HTML Parsing</i> .
3	2018	<i>Comparison of Web scraping Techniques: Regular Expression, HTML DOM, and XPath</i>	<i>Regular Expression, HTML DOM, XPath</i>	Hasil percobaan dengan metode <i>regex</i> paling kecil penggunaan memorinya dibandingkan dengan metode <i>HTML DOM</i> , dan <i>Xpath</i> . Sedangkan <i>HTML DOM</i> memakan waktu paling sedikit dan penggunaan data paling kecil dibandingkan dengan metode <i>Regular Expression</i> dan <i>XPath</i> .
4	2021	<i>Comparative Review on Sentiment Analysis-Based Recommendation System</i>	<i>HTML Parsing</i>	<i>Web scraping</i> dilakukan dengan algoritma <i>Machine Learning</i> dan <i>SRV</i> yang menerapkan

No	Tahun	Judul	Teknik	Keterangan
				teknik HTML <i>Parsing</i> .
5	2021	<i>Web Scraping of COVID-19 News Stories to Create Datasets for Sentiment and Emotion Analysis</i>	HTML <i>Parsing</i>	Berdasarkan fleksibilitas, dan mengetahui kesulitan penggunaannya, dapat diamati bahwa HTML <i>Parsing</i> adalah teknik yang paling baik digunakan pada web ini.

Berdasarkan perbandingan teknik *web scraping* dalam Tabel 2.1 yang telah dilakukan, teknik HTML *Parsing* memiliki keunggulan untuk ekstraksi data dengan menggunakan *library* Python yaitu BeautifulSoup. Dalam penerapannya, BeautifulSoup dapat mengekstrak informasi seperti teks, atribut, *link*, atau gambar dari halaman web. Dalam penelitian Elveny et al. (2021) yang menggunakan teknik HTML *Parsing* untuk pengambilan data produk dari *e-commerce* yang mana penelitian ini juga melakukan pengambilan data produk dari *e-commerce*. Penelitian ini menggunakan HTML *Parsing* sebagai teknik *web scraping* untuk pengambilan data dalam hal ini yaitu situs web *e-commerce* Tokopedia.

### 2.3.2 Selenium WebDriver

Selenium merupakan salah satu alat pengujian yang umum digunakan untuk menguji aplikasi (Nyamathulla et al., 2021). Salah satu versi Selenium yang disebut Selenium 2.0 atau Selenium WebDriver sangat berguna untuk melakukan *web scraping* karena memiliki kemampuan otomatisasi web dan interaksi dengan elemen-elemen web seperti melakukan klik pada tombol, mengisi formulir, membuka tab baru, membuka halaman web, dan sebagainya (Levi et al., 2020). Saat ini, Selenium WebDriver mendukung sebagian besar web *browser* populer yaitu Chrome, Firefox, Opera, dan lainnya.

## 2.4 Metode Statistik

Dalam statistik terdapat dua pendekatan yaitu statistik deskriptif dan statistik inferensial (Kaur et al., 2018). Statistik deskriptif adalah pendekatan statistik yang digunakan untuk memberikan gambaran mengenai objek yang diteliti melalui data dari sampel atau populasi yang terlibat dalam penelitian tersebut (Pandiangan et al., 2022). Statistik deskriptif menyajikan gambaran atau ringkasan data yang terdiri dari nilai rata-rata, deviasi standar, nilai minimum, dan nilai maksimum (Pandiangan et al., 2022). Sementara itu, statistika inferensial merupakan komponen dalam ranah ilmu statistik yang bertujuan untuk melakukan pengujian hipotesis dalam sebuah penelitian, dengan tujuan untuk mencapai kesimpulan yang valid (Mustafa, 2022).

## 2.5 Uji Statistik

Uji statistik merupakan bagian dari statistik inferensial yang terdiri dari dua kelompok, yaitu statistik parametrik dan nonparametrik (Mustafa, 2022). Statistik parametrik digunakan untuk menganalisis jenis data interval dan rasio yang diperoleh dari kelompok keseluruhan dengan populasi yang berdistribusi normal (Mustafa, 2022). Distribusi normal biasa ditandai dengan kurva simetris berbentuk lonceng (*bell-shaped curved*) (Najmi et al., 2021). Di sisi lain, statistik nonparametrik digunakan untuk menganalisis data yang tidak memerlukan persyaratan data mengikuti distribusi normal (Quraissy et al., 2021).

### 1 Uji Statistik Parametrik

Uji statistik parametrik adalah jenis uji yang mengasumsikan bahwa variabel mengikuti distribusi normal dan melibatkan perhitungan nilai rata-rata dan standar deviasi. Selain itu, terdapat persyaratan yang harus dipenuhi agar nilai setiap kelompok yang dibandingkan tidak terpengaruh terhadap nilai-nilai yang ekstrem (*outliers*) (Emerson, 2023). Jenis uji statistik parametrik meliputi uji T dan *Analisis of Variance* (ANOVA).

### 2 Uji Statistik Nonparametrik

Uji statistik nonparametrik dikembangkan untuk mencapai uji perbandingan yang sama dengan uji statistik parametrik tanpa bergantung dengan nilai rata-rata dan standar deviasi (Emerson, 2023). Dalam kata lain, uji statistik nonparametrik tidak diperlukan adanya asumsi tentang bentuk atau parameter distribusi (Hu et al., 2017). Jenis uji statistik nonparametrik meliputi uji Mann-Whitney U, uji Wilcoxon, dan uji Kruskal-Wallis.

Pengaplikasian uji statistik yang akan dilakukan juga memperhatikan jumlah kelompok yang akan diuji. Perbandingan penggunaan uji statistik berdasarkan jumlah kelompok dapat dilihat pada Tabel 2.2 (Henry, 2021).

Tabel 2.2 Tabel Perbandingan Pengaplikasian Uji Statistik

	2 kelompok independen	2 kelompok dependen	>2 kelompok independen	>2 kelompok dependen
Parametrik	Uji T	Uji T	Uji one-way ANOVA	Uji Repeated-measures ANOVA
Non parametrik	Uji Mann-Whitney U	Uji Wilcoxon	Uji Kruskal-Wallis	Uji Friedman

### 2.5.1 Uji Normalitas

Pemilihan metode statistik yang akan digunakan dalam penelitian bergantung pada apakah distribusi data mengikuti distribusi normal atau tidak (Ostrowski & Menyhárt, 2020). Normalitas data dapat diuji dengan dua metode yang umum digunakan, yaitu uji Kolmogorov-Smirnov dan uji Shapiro-Wilk (Orcan, 2020). Uji Kolmogorov-Smirnov digunakan ketika memiliki data yang berukuran besar (Orcan, 2020). Sementara itu uji Shapiro-Wilk digunakan apabila data memiliki jumlah di bawah 50 (Razali & Yap, 2011). Rumus perhitungan untuk uji normalitas dengan menggunakan Kolmogorov-Smirnov ditunjukkan pada persamaan 2.1. (Razali & Yap, 2011).

$$T = \sup_x |F^*(x) - F_n(x)| \quad (2.1)$$

Keterangan:

$T$  = Nilai uji statistik KS

$F^*(x)$  = Fungsi hipotesis distribusi

$F_n(x)$  = Fungsi distribusi kumulatif

Uji statistik dihitung sebagai *supremum* ( $\sup_x$ ) yaitu perbedaan terbesar antara dua fungsi distribusi kumulatif. Statistik ini mengukur sampel yang diamati sesuai dengan distribusi yang diasumsikan. *Supremum* yang lebih kecil mengindikasikan kesesuaian yang lebih baik antara sampel dan distribusi yang diasumsikan, sementara *supremum* yang lebih besar menunjukkan kesesuaian yang lebih buruk (Razali & Yap, 2011).

### 2.5.2 Uji Mann-Whitney U

Uji Mann-Whitney U digunakan untuk membandingkan apakah ada perbedaan nilai untuk dua kelompok independen dari populasi yang sama (Karadimitriou & Marshall, 2015). Pada uji T, hipotesis nol menyatakan nilai rata-rata dari kedua kelompok tersebut sama (Gerald & Patson, 2021). Keadaan tersebut berbanding terbalik dengan uji Mann-Whitney U yang mengaitkan dengan hipotesis nol, menyatakan bahwa perbedaan sebenarnya tidak ada antara dua kelompok data, yang mana sampel dari data ini tidak berkaitan satu dengan lainnya (Mubarok et al., 2021). Rumus perhitungan statistik U dalam uji Mann-Whitney U untuk membandingkan dua kelompok independen ditunjukkan pada persamaan 2.2 (Gerald & Patson, 2021).

$$U_1 = \frac{n_1(n_1+2n_2 + 1) - 2R_1}{2} \quad (2.2)$$

$$U_2 = \frac{n_2(2n_1+n_2 + 1) - 2R_2}{2}$$

Keterangan:

$U_1$  dan  $U_2$  = Nilai statistik U

$n_1$  dan  $n_2$  = Jumlah sampel

$R_1$  dan  $R_2$  = Jumlah peringkat

Perhitungan melibatkan penentuan statistik  $U_1$  dan  $U_2$  untuk setiap kelompok data. Kemudian nilai U diperoleh dengan membandingkan nilai terkecil dari  $U_1$  dan  $U_2$ .

## 2.6 Preprocessing

*Preprocessing* merupakan tahapan awal yang dibutuhkan untuk merapikan data sebelum data tersebut diolah. Informasi yang redundan, tidak terkait, atau mengandung *noise* akan menyebabkan data tersebut tidak berkualitas sehingga menghasilkan analisis yang kurang terepresentasi dengan baik (Alexandropoulos et al., 2019). Tujuan *preprocessing* ini yaitu untuk mengurangi ukuran data, mencari hubungan antara dua data, normalisasi data, menghilangkan *outliers*, dan mengekstraksi fitur. *Preprocessing* memiliki beberapa metode seperti *data cleaning*, *data integration*, *data transformation*, dan *data reduction* (Alasadi, 2017). Pemilihan metode yang akan digunakan tersebut disesuaikan dengan karakteristik dan kebutuhan data. Penelitian ini hanya membutuhkan dua metode yaitu *data cleaning* dan *data transformation*.

### 2.6.1 *Data Cleaning*

*Data cleaning* atau pembersihan data merupakan salah satu metode yang bekerja untuk mendeteksi data yang rusak, menghapus bagian data yang tidak relevan dan mengganti data yang salah (Sahoo et al., 2019). Proses pembersihan data mencakup tugas-tugas seperti pencocokan data, mengidentifikasi ketidakakuratan data, pengurutan data, identifikasi data *outlier*, pemeriksa ejaan data tekstual, dan peningkatan kualitas data (Nongthombam, 2021). Dalam penelitian ini, akan dilakukan metode *data cleaning* berupa penanganan *missing value*, penghapusan duplikasi data, mendeteksi *outlier*, dan memperbaiki ketidaksesuaian data.

*Missing value* atau nilai kosong dapat menyebabkan masalah analisis data yaitu data dapat menjadi bias yang disebabkan oleh perbedaan antara nilai data yang hilang dan lengkap (Emmanuel et al., 2021). Pendekatan yang dapat dilakukan untuk menangani nilai kosong yaitu dengan cara penghapusan data, yang mana setiap data yang memiliki nilai kosong akan dihapus. Penghapusan data dianggap sebagai pendekatan paling sederhana karena tidak perlu mencoba dan memperkirakan nilai (Emmanuel et al., 2021). Sementara itu, *outlier* atau nilai ekstrem merupakan suatu nilai yang berbeda secara eksplisit dari nilai lainnya dalam *dataset* yang dapat mempengaruhi hasil analisis (Boukerche et al., 2020). Oleh karena itu, penghapusan *outlier* dilakukan dalam proses *data cleaning* dalam *dataset* agar dapat menghasilkan model statistik yang lebih baik (Boukerche et al., 2020).

### 2.6.2 *Data Transformation*

*Data transformation* atau transformasi data adalah proses modifikasi representasi data sehingga mereka memenuhi syarat untuk diolah (Canchen, 2019). Transformasi data juga memiliki definisi yaitu sebuah teknik untuk mengonversi data agar mendekati persyaratan teknik atau metode yang akan diterapkan atau digunakan (Adikaram et al., 2015).

## 2.7 *Exploratory Data Analysis*

*Exploratory Data Analysis* (EDA) merupakan tahapan dalam projek *data science* yang krusial (Peng et al., 2021). EDA membantu menganalisis kumpulan data untuk meringkas karakteristik statistik yang berfokus pada empat aspek utama, yaitu ukuran tendensi sentral (rata-rata, modus, dan median), ukuran penyebaran (standar deviasi dan varians), bentuk distribusi, dan keberadaan *outlier* (Sahoo et al., 2019). EDA memiliki tujuan utama untuk menganalisis data sebelum mengemukakan asumsi-asumsi yang beragam dan juga dapat membantu dalam penemuan hubungan antarvariabel (Maringka & Kusnawi, 2021). Hubungan

antarvariabel dapat diidentifikasi melalui analisis korelasi, di mana matriks korelasi menghitung tingkat korelasi antara setiap pasangan kolom, yang mengungkapkan hubungan antara fitur yang berkorelasi dengan target dan juga antara fitur yang saling berkorelasi (Peng et al., 2021). Sebagian besar teknik EDA berbentuk grafis dengan beberapa teknik kuantitatif (Radhi et al., 2022). Dengan melakukan EDA, dapat membantu pemahaman terhadap *dataset* untuk analisis selanjutnya (Samosir et al., 2021). Pemahaman terkait kondisi *dataset* dapat ditinjau setidaknya melalui poin-poin berikut (Samosir et al., 2021):

1. Menyaring variabel yang relevan dan mengabaikan variabel yang tidak memiliki nilai signifikan.
2. Mengenali nilai ekstrem (*outliers*), nilai yang tidak lengkap (*missing values*), atau kesalahan manusia (*human error*).
3. Menggambarkan hubungan antarvariabel.
4. Meningkatkan pemahaman atas situasi data dan mengurangi risiko kesalahan di masa mendatang.

## 2.8 Interquartile Range (IQR)

*Interquartile Range*/Rentang Interkuartil merupakan teknik yang digunakan untuk mendeteksi *outlier* dengan data yang tidak memiliki distribusi normal (Nair & Kashyap, 2019). Teknik ini menggunakan perbedaan antara peringkat nilai sebesar 25% dan 75% dalam *dataset* (Yang et al., 2019). *Outlier* dapat terdeteksi melalui penggunaan *boxplot* yaitu grafik visualisasi yang didasarkan pada IQR (Vinutha et al., 2018). Rumus dari IQR akan ditunjukkan pada persamaan 2.3 (Vinutha et al., 2018):

$$IQR = Q_3 - Q_1 \quad (2.3)$$

Keterangan:

IQR = Interkuartil

$Q_1$  = Kuartil pertama

$Q_3$  = Kuartil ketiga

## 2.9 Python

Python merupakan bahasa pemrograman yang populer sebagai alternatif untuk MATLAB dan R (Zhao et al., 2020). Python memiliki sejumlah keunggulan dibandingkan bahasa pemrograman lainnya (Nongthombam, 2021). Salah satu keunggulannya terletak pada bahasa pemrograman tingkat tinggi yang memungkinkan kode dengan mudah dibaca sehingga dapat

dipahami dan digunakan baik oleh *programmer* ataupun pengguna dari berbagai latar belakang (Nongthombam, 2021). Selain itu, Python menyediakan berbagai pustaka dan fungsi untuk keperluan statistik dan analisis numerik.

## 2.10 SPSS

*Statistical Package for the Social Science* (SPSS) merupakan aplikasi untuk perhitungan statistik yang dikembangkan oleh IBM (Wahyuni et al., 2021). Dalam menjalankan program, SPSS memiliki beberapa keterbatasan antara lain (Yang, 2013) : (1) Terbatas pada metode langkah-langkah dan tidak memiliki kemampuan untuk melakukan regresi dengan semua kemungkinan subset; (2) Terbatas dalam hal statistik optimalitas untuk pemilihan variabel, dan kriteria yang ada berbentuk uji signifikansi yang rentan terhadap kesalahan (*type I error* dan *type II error*); (3) Tidak memiliki kemampuan secara otomatis mengidentifikasi dan mengatasi kasus-kasus yang mencolok; (4) Tidak memiliki kemampuan melakukan penggabungan model untuk meningkatkan prediksi; (5) Tidak memiliki kemampuan berinteraksi dengan program SPSS *Server* untuk bekerja dengan data yang besar.

## 2.11 Kajian Pustaka Uji Statistik Nonparametrik

Dalam pembahasan ini mengambil 5 penelitian terkait uji statistik nonparametrik sebagai perbandingan dari setiap metode. Kata kunci yang digunakan dalam pencarian sumber di situs web Google Scholar adalah “uji statistik nonparametrik”, “*nonparametric statistical testing*”, “*Mann-Whitney U test*”, dan “*Kruskal-Wallis test*” dengan pengaturan filter dari tahun 2017 hingga saat ini. Berdasarkan literatur yang diperoleh, diambil lima penelitian yang paling sesuai untuk dirangkum dalam laporan ini.

Pertama, penelitian yang dilakukan oleh Mubarok et al. (2021) berjudul “*Mann Whitney Test in Comparing The Students’ Consultation Results of Entrepreneurial Practice Between Male and Female Lecturers in Economic Faculty of Pamulang University*”. Penelitian ini bertujuan untuk mengetahui apakah ada perbedaan hasil bimbingan praktik kewirausahaan mahasiswa antara dosen laki-laki dengan dosen perempuan pada Fakultas Ekonomi Universitas Pamulang. Metode yang digunakan untuk pengujian hipotesis ini yaitu uji Mann-Whitney U dengan program SPSS. Hasil penelitian yang didapatkan dari pengujian dengan sampel sejumlah 66 kelas yaitu tidak terdapat perbedaan yang signifikan antara hasil bimbingan praktik mahasiswa antara dosen laki-laki maupun perempuan.

Kedua, penelitian yang dilakukan oleh Utomo (2021) berjudul “Perbandingan Tabel Mortalita Indonesia dan Tabel Mortalita CSO Menggunakan Uji Mann-Whitney U dan Uji Kruskal-Wallis”. Penelitian ini membahas tabel mortalita yang merupakan instrumen yang penting pada studi aktuarial, demografi, dan epidemiologi. Penelitian ini menguji apakah tabel mortalita mewakili perubahan demografis yang nyata menggunakan nonparametrik yaitu uji Mann-Whitney dan Kruskal-Wallis. Hasil yang diperoleh yaitu keempat tabel mortalita milik Indonesia tidak menunjukkan perbedaan yang nyata pada tingkat kematian. Penelitian tersebut juga menguji tabel mortalita milik Amerika Serikat dan hasil yang didapatkan yaitu sama seperti Indonesia yang tidak menunjukkan perbedaan yang signifikan pada tingkat kematian.

Ketiga, penelitian yang dilakukan oleh Syamsuar dan Megayani (2022) berjudul “Perbandingan Pengaruh *Brand Ambassador* BTS dengan Blackpink Terhadap Minat Beli Pada Tokopedia Menggunakan Mann-Whitney U-Test”. Penelitian ini berisi analisis dari dua *brand ambassador* yaitu BTS dan Blackpink terhadap minat beli pada *e-commerce* Tokopedia. Penelitian ini mengkaji pengaruh antara variabel independen yaitu dua *brand ambassador* terhadap variabel dependen yaitu minat beli. Metode yang digunakan dalam penelitian ini yaitu menggunakan Mann-Whitney U *test*. Hasil yang diperoleh yaitu kedua nilai besaran median BTS dan Blackpink secara statistik berpengaruh terhadap minat beli pada *e-commerce* Tokopedia.

Keempat, penelitian yang dilakukan oleh Quraisy et al. (2021) berjudul “Analisis Kruskal-Wallis Terhadap Kemampuan Numerik Siswa”. Penelitian ini bertujuan untuk membandingkan kemampuan numerik siswa dari empat kelas. Metode yang digunakan yaitu uji nonparametrik Kruskal-Wallis. Hasil yang diperoleh yaitu tidak terdapat perbedaan signifikan dari kemampuan numerik siswa dari keempat kelas tersebut.

Kelima, penelitian yang dilakukan oleh Wibowo dan Sutandi (2018) berjudul “Analisa Rasio Keuangan Garuda Indonesia Airlines, Singapore Airlines, dan Thailand Airlines dengan Uji Non-Parametrik”. Penelitian ini bertujuan untuk membandingkan kinerja keuangan tiga maskapai penerbangan yaitu Garuda Indonesia Airlines, Singapore Airlines, dan Thailand Airlines. Metode analisis yang digunakan yaitu uji statistik nonparametrik Kruskal-Wallis dan Mann-Whitney U. Pengujian hipotesis yang dilakukan mencakup analisis terhadap berbagai rasio keuangan, seperti rasio likuiditas, aktivitas, solvabilitas, profitabilitas, serta menggunakan metode *Du Pont System*. Dari hasil penelitian ini, ditemukan adanya perbedaan yang signifikan dalam perbandingan rasio keuangan, meskipun ada beberapa yang tidak menunjukkan perbedaan yang signifikan.

Tabel 2.3 Tabel Perbandingan metode dalam uji statistik nonparametrik

No	Tahun	Judul	Metode	Keterangan
1	2021	<i>Mann Whitney Test in Comparing The Students' Consultation Results of Entrepreneurial Practice Between Male and Female Lecturers in Economic Faculty of Pamulang University</i>	Mann-Whitney U	Hasil yang diperoleh menunjukkan nilai $p\text{-value} > 0,05$ . Hipotesis nol diterima, yang mengindikasikan bahwa tidak terdapat perbedaan yang signifikan antara dua kelompok variabel yang sedang diuji.
2	2021	Perbandingan Tabel Mortalita Indonesia dan Tabel Mortalita CSO Menggunakan Uji Mann Whitney dan Uji Kruskal-Wallis	Mann-Whitney U, Kruskal-Wallis	Metode uji Mann-Whitney U digunakan untuk membandingkan rata-rata tingkat kematian dua tabel mortalita. Sedangkan uji Kruskal-Wallis digunakan untuk membandingkan rata-rata tingkat kematian lebih dari dua tabel.
3	2022	Perbandingan Pengaruh BTS Brand Ambassador Dengan Blackpink Terhadap Minat Beli Pada Tokopedia Menggunakan Mann-Whitney U-Test	Mann-Whitney U	Hasil yang diperoleh yaitu kedua kelompok memiliki pengaruh yang signifikan terhadap minat beli di platform <i>e-commerce</i> Tokopedia.
4	2021	Analisis Kruskal-Wallis Terhadap	Kruskal-Wallis	Hasil yang diperoleh dengan metode uji Kruskal-Wallis yaitu tidak terdapat

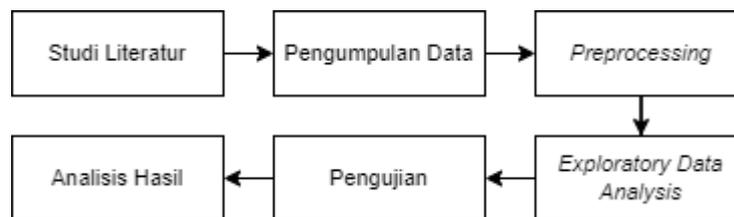
No	Tahun	Judul	Metode	Keterangan
		Kemampuan Numerik Siswa		perbedaan signifikan dari kemampuan numerik siswa dari 4 kelas tersebut.
5	2018	Analisa Rasio Keuangan Garuda Indonesia Airlines, Singapore Airlines, dan Thailand Airlines dengan Uji Non-Parametrik	Mann-Whitney U, Kruskal-Wallis	Uji Kruskal-Wallis digunakan untuk membandingkan signifikansi antar tiga negara dengan indikator yang telah ditentukan. Sementara uji Mann-Whitney U digunakan untuk membandingkan signifikansi antar dua negara.

Berdasarkan studi literatur yang telah dilakukan yang telah dibandingkan dalam Tabel 2.3, metode uji statistik nonparametrik yang paling cocok digunakan yaitu uji Mann-Whitney U. Penggunaan uji Mann-Whitney U bertujuan untuk mengetahui apakah ada perbedaan pada variabel dependen antara dua kelompok independen yang berasal dari populasi yang sama (Karadimitriou & Marshall, 2015). Sementara itu metode uji Kruskal-Wallis digunakan ketika terdapat lebih dari dua kelompok sampel (Quraisy et al., 2021). Berdasarkan arah penelitian yang direncanakan yang melibatkan perbandingan antara dua kelompok independen, metode uji statistik nonparametrik yang tepat untuk digunakan adalah uji Mann-Whitney U.

## BAB III METODOLOGI PENELITIAN

### 3.1 Tahapan Penelitian

Penelitian ini akan dilakukan melalui beberapa tahapan yaitu: studi literatur, pengumpulan data, *Exploratory Data Analysis*, pembersihan data, pengujian, dan penarikan kesimpulan. Gambar 3.1. menunjukkan tahapan dari penelitian.



Gambar 3.1 Tahapan Penelitian

Tahap awal penelitian dimulai dengan melakukan studi literatur. Dalam proses ini, dilakukan pencarian, pengumpulan, dan pemahaman informasi yang telah dipublikasikan dalam berbagai bentuk, seperti buku, jurnal, atau makalah. Informasi ini merupakan sumber tertulis yang relevan dengan penelitian atau topik yang sedang diselidiki, dan akan digunakan sebagai landasan untuk memilih metode penelitian yang paling sesuai. Selanjutnya dilakukan proses pengumpulan data dengan metode *web scraping e-commerce* Tokopedia dan disimpan dalam bentuk *file* CSV. Setelah itu dilakukan *preprocessing* untuk menyiapkan data yang akan diolah. Tahapan *preprocessing* meliputi *data cleaning* untuk meningkatkan kualitas data dan *data transformation* untuk menangani ketidaksesuaian tipe data serta menjaga konsistensi data. Dilanjutkan dengan analisis data visual dengan *Exploratory Data Analysis* (EDA) untuk mendapatkan karakteristik data sebelum dapat dilakukan pengujian statistik. Langkah terakhir yaitu uji statistik untuk menguji hipotesis yang dilakukan dalam dua tahapan. Pertama, uji normalitas yaitu uji Kolmogorov-Smirnov untuk mengecek distribusi data. Kedua, metode uji statistik nonparametrik yang telah dipilih yaitu uji Mann-Whitney U untuk membandingkan dua kelompok independen. Setelah itu dilakukan analisis untuk mengkaji temuan atau hasil yang diperoleh dari penelitian ini.

### 3.2 Pengumpulan Data

Pada tahap ini, data yang digunakan merupakan data produk makanan dan minuman berlabel halal dan tidak pada platform *e-commerce* Tokopedia yang diambil pada bulan Juli 2023. Proses pengumpulan data menggunakan *web scraping* ditunjukkan melalui *pseudocode* pada Gambar 3.2.

```

START
  SET url = url tokopedia
  OPEN browser and navigate to url
  SET data as an empty list

  FOR i in range 3
    SCROLL_DOWN to load more content
    FOR item in items on the page
      EXTRACT nama, harga, rating, jumlah barang terjual,
        lokasi
      APPEND extracted data to data list
    GO_TO_NEXT_PAGE

  CREATE a DataFrame from data with column names: "Nama Barang", "Harga",
    "Terjual", "Rating", "Lokasi"

  ADD a new column to the DataFrame named "Label"

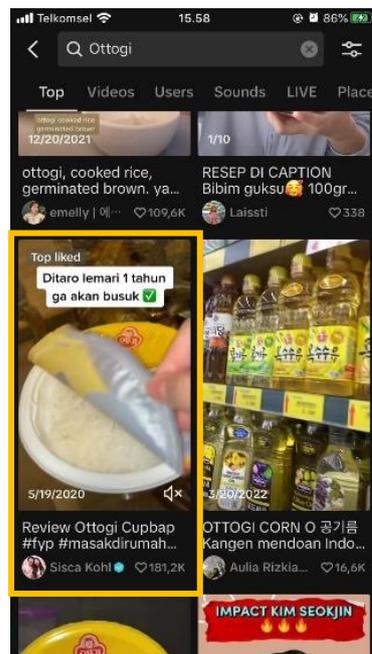
  PRINT the DataFrame
  CLOSE the browser
END

```

Gambar 3.2 *Pseudocode Web Scraping*

Pengambilan data menggunakan metode *web scraping* dengan menerapkan teknik *HTML Parsing*. *HTML Parsing* bekerja dengan melibatkan *library* Python yaitu `BeautifulSoup` untuk mengekstrak data yang dibutuhkan dengan merujuk pada elemen-elemen menggunakan fungsi *CSS Selector*. Proses *web scraping* berdasarkan *pseudocode* dimulai dengan mengatur URL dan membuka *browser* dengan *ChromeDriver*. *ChromeDriver* merupakan digunakan untuk mengotomatisasi *browser* untuk interaksi dengan halaman web menggunakan URL produk. Data yang diekstrak dapat berupa teks atau atribut dari elemen HTML, seperti nama produk, harga, jumlah produk terjual, ulasan, dan lokasi penjual. Proses ekstraksi dilakukan sebanyak 3 halaman yang tiap halamannya memiliki sekitar 80 produk dan dimasukkan kedalam *list* yang telah dibuat sebelumnya. Hasil ekstraksi disimpan kedalam *dataframe* yang digunakan untuk proses penyimpanan data. Sebelum data disimpan, dilakukan proses pelabelan dengan membuat variabel baru untuk mengelompokkan data yang telah berlabel halal dan tidak. Kemudian data disimpan dalam format CSV dan *ChromeDriver* dapat ditutup. Proses ini dilakukan berulang kali yang disesuaikan dengan jumlah produk yang akan diambil.

Pemilihan produk makanan dan minuman dalam kemasan yang digunakan dalam penelitian ini merupakan produk yang populer serta telah melalui seleksi atau pengecekan kehalalan melalui situs [halalmui.org/search-product/](http://halalmui.org/search-product/). Sementara untuk produk tidak berlabel halal, diperoleh melalui pengamatan sejumlah konten pada aplikasi media sosial TikTok yang banyak memberikan ulasan tentang produk makanan dan minuman dalam kemasan impor tanpa label halal dengan jumlah *like* yang cukup signifikan. Seperti pada produk Ottogi yang merupakan produk makanan impor yang memiliki video ulasan dengan jumlah *like* terbanyak yaitu sekitar 181.200 *likes* seperti Gambar 3.1.



Gambar 3.1 Contoh video ulasan produk di TikTok

Selain itu, produk yang tidak berlabel halal juga dikumpulkan melalui pengamatan toko yang menjual produk impor di situs web *e-commerce* Tokopedia. Kriteria pemilihan produk ini juga mempertimbangkan penjualan yang mencapai halaman 3. Pemilihan batas tiga halaman didasarkan pada fakta bahwa penjualan produk tidak berlabel halal rata-rata hanya mencapai tiga halaman.

Sejumlah 30 produk, terdiri dari 15 produk berlabel halal dan 15 produk tidak yang digunakan sebagai sampel dalam penelitian ini, yang mana data produk tersebut ditunjukkan pada Tabel 3.1.

Tabel 3.1 Daftar nama produk berlabel halal dan tidak

No	Berlabel Halal	No	Tidak Berlabel Halal
1	Indomie	1	Ottogi
2	Koko krunch	2	Post oreo
3	Yupi	3	Permen kelinci
4	Chitato	4	Wulama
5	Malkist roma	5	Vicenzovo
6	Saus abc	6	Gochujang taeyangcho
7	Pronas	7	Spam
8	Morin	8	Lotus biscoff spread
9	Sariwangi	9	Lotte chilsung
10	Sprite	10	Lotte cantata
11	Nescafe	11	Hata ramune
12	Polaris	12	French classic yellow mustard
13	Cocacola	13	Monin
14	Kellogs	14	Borges
15	Sirup marjan	15	Dongsuh

### 3.3 Preprocessing

*Preprocessing* dilakukan untuk memperoleh data yang bersih sehingga siap untuk proses analisis. Pada penelitian ini terdapat 2 langkah yang akan dilakukan dalam tahapan ini yaitu *data cleaning* dan *data transformation*.

#### a. Data Cleaning

Proses yang akan dilakukan pada tahapan ini yaitu penanganan *missing value*, penghapusan duplikasi data, mendeteksi *outlier*, dan memperbaiki ketidaksesuaian data. Dalam memperbaiki ketidaksesuaian data, digunakan Regex untuk menyamakan format data. Deteksi *outlier* yang dilakukan menggunakan metode *Interquartile Range* dengan menghitung selisih antara kuartil atas dan kuartil bawah dalam suatu distribusi data.

#### b. Data Transformation

Proses yang dilakukan dalam tahap ini adalah mengonversi tipe data yang tidak sesuai dengan yang seharusnya.

### 3.4 *Exploratory Data Analysis (EDA)*

Pada tahapan ini *Exploratory Data Analysis (EDA)* digunakan untuk menggali lebih dalam karakteristik *dataset* yang akan digunakan. EDA melibatkan penggunaan ringkasan statistik, visualisasi data, dan representasi grafis untuk membantu pemahaman tentang distribusi data, kecenderungan pusat, dispersi (penyebaran data), korelasi antarvariabel, dan kemungkinan adanya anomali dalam *dataset*. Analisis ini dilakukan untuk memastikan bahwa data telah siap untuk tahapan analisis yang lebih mendalam, seperti pengujian hipotesis.

### 3.5 Uji Statistik

Uji statistik yang digunakan pada penelitian ini meliputi uji statistik nonparametrik dan uji normalitas. Kedua uji ini memiliki proses yang sama yakni melibatkan langkah-langkah sebagai berikut:

1. Menentukan hipotesis

Secara umum, setiap uji statistik melibatkan hipotesis yang digunakan sebagai asumsi. Hipotesis biasanya mencakup hipotesis nol ( $H_0$ ) dan hipotesis alternatif ( $H_a$ ).

2. Menentukan nilai signifikansi

Tingkat signifikansi *alpha* menggambarkan probabilitas pengambilan keputusan yang salah ketika hipotesis nol benar. Secara umum, tingkat *alpha* dalam pengujian hipotesis biasanya bernilai 0.05 (5%). Pemilihan nilai 5% sebagai batasan dalam statistik didasarkan pada distribusi normal dan konsep *outlier*, sehingga memungkinkan untuk membuat keputusan yang konsisten dalam analisis data dan pengujian hipotesis (Andrade, 2019).

3. Uji hipotesis

Pada tahap ini data akan diuji sesuai dengan metode pengujian yang sesuai untuk menghasilkan *p-value* yang nantinya akan digunakan sebagai dasar pengambilan keputusan.

#### 3.5.1 Uji Mann-Whitney U

Berdasarkan arah penelitian yang direncanakan yang melibatkan perbandingan antara dua kelompok independen, metode uji statistik nonparametrik yang sesuai untuk digunakan adalah uji Mann-Whitney U dimana metode ini tidak memperhatikan normalitas distribusi data. Dalam uji Mann-Whitney U hipotesis yang diajukan adalah:

1.  $H_0$ : Tidak terdapat perbedaan yang signifikan dalam hasil penjualan produk antara produk yang diberi label halal dan tidak di *e-commerce* Tokopedia Indonesia.
2.  $H_a$ : Terdapat perbedaan yang signifikan dalam hasil penjualan produk antara produk yang diberi label halal dan tidak di *e-commerce* Tokopedia Indonesia.

Dalam konteks ini, terdapat dua asumsi yang dapat digunakan untuk menginterpretasi hasil:

1. Jika nilai signifikansi ( $p > 0.05$ ) maka  $H_0$  diterima, dan menolak  $H_a$ . Berdasarkan hasil uji, tidak terdapat bukti yang cukup untuk menolak hipotesis nol. Sehingga dapat disimpulkan bahwa tidak ada perbedaan signifikan dalam hasil penjualan produk antara kedua kelompok.
2. Jika nilai signifikansi ( $p < 0.05$ ) maka  $H_a$  diterima, dan menolak  $H_0$ . Berdasarkan hasil uji, memiliki yang cukup untuk menolak hipotesis nol. Sehingga dapat disimpulkan bahwa ada perbedaan signifikan dalam hasil penjualan produk antara dua kelompok.

Dalam bahasa pemrograman Python, untuk melakukan uji Mann-Whitney U dapat menggunakan *library* statistik yang disebut ``scipy`` dengan fungsi ``mannwhitneyu`` yang terdapat dalam modul ``scipy.stats``. Dari hasil dari uji Mann-Whitney U dengan tingkat signifikansi tersebut digunakan sebagai penentuan untuk pengambilan keputusan.

### 3.5.2 Uji Kolmogorov-Smirnov

Uji normalitas yang digunakan pada penelitian ini yaitu uji Kolmogorov-Smirnov yang dilakukan untuk melihat distribusi data. Dalam uji Kolmogorov-Smirnov hipotesis yang diajukan adalah:

1.  $H_0$ : data terdistribusi normal
2.  $H_a$ : data tidak terdistribusi normal

Dalam konteks ini, terdapat dua asumsi yang dapat digunakan untuk menginterpretasi hasil:

1. Jika nilai signifikansi ( $p > 0.05$ ) maka  $H_0$  diterima, dan menolak  $H_a$ . Berdasarkan hasil uji, tidak terdapat bukti yang cukup untuk menolak hipotesis nol. Sehingga dapat disimpulkan bahwa data mengikuti distribusi normal.
2. Jika nilai signifikansi ( $p < 0.05$ ) maka  $H_a$  diterima, dan menolak  $H_0$ . Berdasarkan hasil uji, memiliki cukup bukti untuk menolak hipotesis nol. Sehingga dapat disimpulkan bahwa data tidak mengikuti distribusi normal.

Dalam bahasa pemrograman Python, untuk melakukan uji Kolmogorov-Smirnov dapat menggunakan *library* statistik yang disebut ``scipy`` dan fungsi ``k2_2samp`` yang terdapat dalam modul ``scipy.stats``. Dengan demikian, hasil dari uji Kolmogorov-Smirnov dengan

tingkat signifikansi yang telah ditentukan digunakan sebagai penentuan metode pengujian statistik selanjutnya.

### **3.6 Pengujian menggunakan *software* SPSS**

Dalam penelitian ini juga akan dilakukan pengujian statistik menggunakan *software* SPSS guna membandingkan hasil yang diperoleh dengan pengujian melalui pemrograman Python untuk meningkatkan validitas dari hasil penelitian.

### **3.7 Analisis Hasil**

Pada bagian ini, akan dilakukan analisis dengan mengkaji hasil yang diperoleh yaitu nilai *p-value* guna mendapatkan pemahaman yang lebih mendalam.

## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1 Pengumpulan Data

Dalam proses pengumpulan data menggunakan metode *web scraping*, peneliti melakukan eksperimen terhadap dua *e-commerce* terpopuler di Indonesia yaitu Shopee dan Tokopedia pada bulan Juli 2023. Percobaan ini dilakukan dengan menggunakan teknik *HTML Parsing* untuk mengetahui apakah teknik tersebut dapat digunakan pada platform *e-commerce* yang berbeda.

Pengumpulan data melalui *web scraping* pada situs web *e-commerce* membutuhkan ChromeDriver yang merupakan perangkat lunak (*software*) yang merupakan implementasi dari Selenium WebDriver. ChromeDriver berfungsi sebagai alat otomatisasi browser Google Chrome. Proses instalasi ChromeDriver dapat dilakukan melalui link <https://chromedriver.chromium.org/downloads>. Setelah instalasi berhasil dilakukan *file* yang berisi perangkat lunak ChromeDriver ditempatkan dalam folder yang sama dengan *file* kode yang digunakan untuk menjalankan *web scraping* dengan bahasa pemrograman Python. Diperlukan beberapa *library* Python untuk *web scraping* yang dapat dilihat pada Gambar 4.1. Penulisan kode dilakukan dengan menggunakan Jupyter Notebook dalam IDE Visual Studio Code dengan bahasa pemrograman Python.

```
#import library

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

Gambar 4.1 Kode program untuk *import libraries*

*Library* ``webdriver`` berfungsi sebagai alat otomatisasi *browser*. *Library* ``WebDriverWait`` merupakan bagian dari selenium yang menyediakan mekanisme menunggu di mana kondisi harus terpenuhi sebelum melanjutkan eksekusi. *Library* ``EC`` berfungsi untuk menampung kondisi yang telah ditentukan yang dapat digunakan dengan *library* ``WebDriverWait``. *Library* ``EC`` digunakan untuk mencari elemen pada halaman web. *Library* ``time`` digunakan untuk menambah waktu/jeda dalam eksekusi kode yang memungkinkan

halaman web dimuat sepenuhnya sebelum melakukan *scraping*. *Library* `BeautifulSoup` digunakan untuk ekstraksi data dari HTML. *Library* `Pandas` digunakan untuk membuat *dataframe* untuk menyimpan data struktur tabular.

Setelah mengimpor *library* yang diperlukan, perlu dilakukan inisiasi url yang akan di-*scraping* dan menyimpannya dalam variabel url. Dilanjutkan dengan menentukan *path* atau lokasi dari *file* berisikan *software* ChromeDriver yang telah diinstal sebelumnya. Halaman Chrome akan terbuka sesuai dengan url yang telah ditentukan. Kode program untuk proses ini ditunjukkan seperti pada Gambar 4.2.

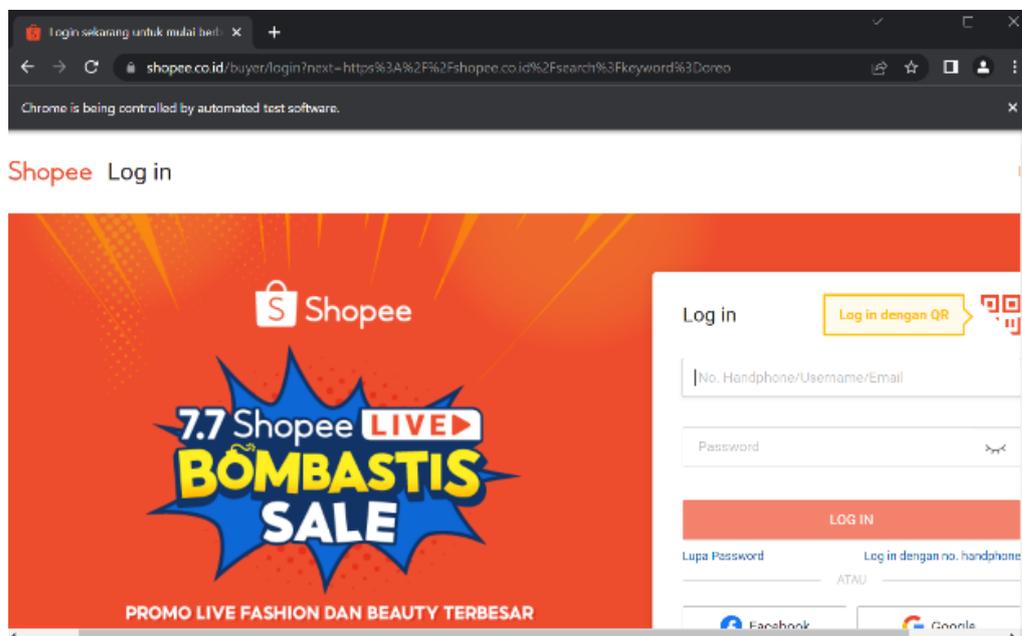
```
#otomatisasi webdriver

url = "url situs web ecommerce"
driver = webdriver.Chrome(executable_path=
r"D:\python\skripsi\tokopedia\chromedriver.exe")
driver.get(url)
```

Gambar 4.2 Kode program untuk membuka halaman Chrome

Berikut merupakan percobaan *web scraping* pada *e-commerce* Shopee dan Tokopedia Indonesia.

a. Percobaan *web scraping* menggunakan *E-commerce* Shopee Indonesia

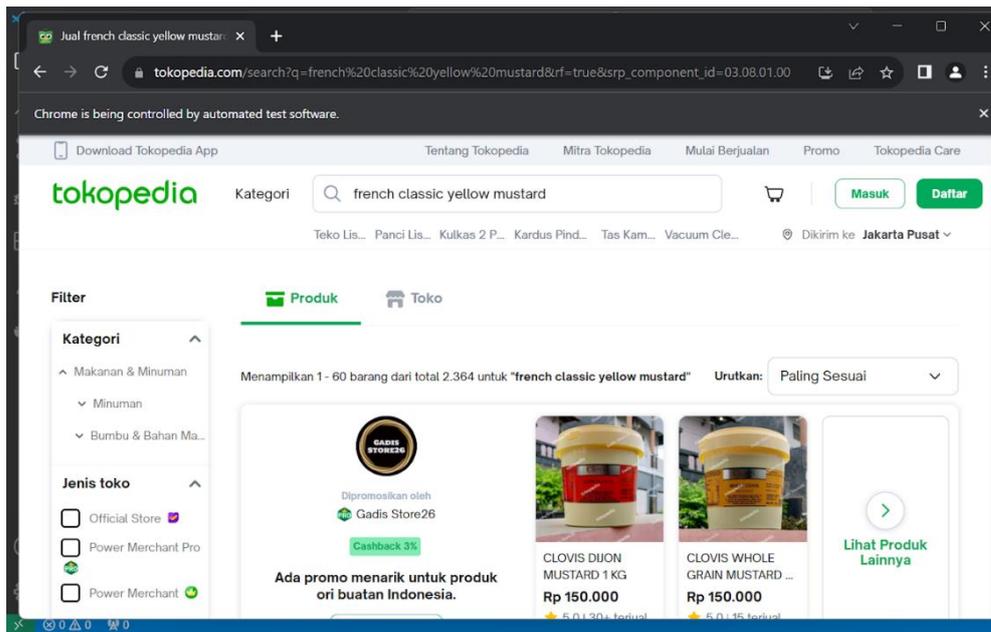


Gambar 4.3 Tampilan halaman situs web Shopee Indonesia

Pengambilan data menggunakan metode *web scraping* dengan teknik HTML Parsing yang dilakukan melalui *e-commerce* Shopee Indonesia tidak dapat dilakukan. Halaman web yang

terbuka tidak sesuai dengan url yang dimasukkan seperti Gambar 4.3. Dalam hal ini Shopee menggunakan deteksi bot yang mencegah akses otomatis seperti kemampuan Selenium WebDriver untuk berinteraksi dengan web. Oleh karena itu, dilakukan percobaan *web scraping* pada *e-commerce* lain, dalam hal ini Tokopedia.

b. Percobaan *web scraping* menggunakan *E-commerce* Tokopedia Indonesia



Gambar 4.4 Tampilan halaman situs web Tokopedia

Selenium WebDriver berhasil mengakses situs web Tokopedia sebagaimana dapat dilihat pada Gambar 4.4. Hal ini dapat dilanjutkan untuk pengambilan data melalui platform *e-commerce* Tokopedia. Produk yang digunakan dalam penelitian ini adalah data produk makanan dan minuman dalam kemasan berlabel halal dan tidak. Proses pengambilan data dari halaman web melibatkan elemen-elemen dari kode sumber web. Gambar 4.5 menunjukkan contoh beberapa elemen yang akan digunakan pada situs web yang dituju.

```

<div class="prd_link-product-name css-3um8ox"
data-testid="spnSRPProdName">Oreo Biskuit
Blackpink Pink Cookie 123.5g 3 Pack - Limited
Edition</div>
▼ <div class>
  <div class="prd_link-product-price css-1ksb19
c" data-testid="spnSRPProdPrice">Rp22.275
</div> flex

```

Gambar 4.5 Kode sumber web

Kode untuk *scraping* seperti pada Gambar 4.6 dijalankan. Proses *scraping* dilakukan hingga halaman ketiga dengan tiap pengambilan akan dilakukan secara otomatis. Proses *scrolling* otomatis sebanyak 20 kali dilakukan untuk memastikan halaman terbuka secara sempurna sehingga informasi dapat diekstrak dan disimpan dalam *dataframe*.

```

data = []
for i in range(3):
    WebDriverWait(driver,
5) .until(EC.presence_of_element_located((By.CSS_SELECTOR, "#zeus-root")))
    time.sleep(2)

# Scrolling otomatis
for j in range(20):
    driver.execute_script("window.scrollTo(0, 250)")
    time.sleep(1)
    driver.execute_script("window.scrollTo(50, 0)")
    time.sleep(2)

# Scraping
soup = BeautifulSoup(driver.page_source, "html.parser")
for item in soup.findAll('div', class_='css-974ipl'):
    nama = item.find('div', class_='prd_link-product-name css-
3um8ox').text
    harga = item.find('div', class_='prd_link-product-price css-
1ksb19c').text
    terjual = item.find('span', class_='prd_label-integrity css-
1duhs3e').text
    rating = item.find('span', class_='prd_rating-average-text css-
t70v7i').text

    for item2 in item.findAll('div', class_='css-1rn0irl'):
        lokasi = item2.find('span', class_='prd_link-shop-loc css-
1kdc32b flip').text

        data.append(
            (nama, harga, terjual, rating, lokasi)
        )
    time.sleep(2)

# Pindah halaman
driver.find_element(By.CSS_SELECTOR, "button[aria-label^='Laman
berikutnya']").click()
time.sleep(2)

```

```
# Simpan dalam dataframe
df = pd.DataFrame(data, columns=["Nama Barang", "Harga", "Terjual",
"Rating", "Lokasi"])
print(df)

driver.close()
```

Gambar 4.6 Kode program untuk pengambilan data

Tahap pelabelan data dilakukan sebelum data disimpan dalam format *file* CSV. Tujuan dari pelabelan ini untuk mengelompokkan data yaitu produk berlabel halal dan tidak berlabel yang dilakukan dengan membuat variabel baru yang bernama "Label" seperti kode program yang ditunjukkan pada Gambar 4.7.

```
# Membuat variabel baru (labeling)
df["Label"] = "no" #value 'no' untuk produk tidak berlabel halal
df["Label"] = "yes" #value 'yes' untuk produk berlabel halal
```

Gambar 4.7 Kode program untuk pelabelan dengan menambah variabel baru

Pengambilan data melalui metode *web scraping* dilakukan secara berulang kali sesuai dengan produk yang telah ditentukan yaitu 30 produk. Untuk menjaga konsistensi data yang ada maka data yang telah berhasil diambil disimpan dalam format *file* CSV. Kode program untuk penyimpanan *dataset* dalam format *file* CSV dapat dilihat pada Gambar 4.8.

```
# Menyimpan data ke dalam file csv
with open('tokopedia_fix.csv', 'a', newline='', encoding='utf-8') as f:
    df.to_csv(f, header=False)
```

Gambar 4.8 Kode program untuk menyimpan data

*Web scraping* pada *e-commerce* Tokopedia berhasil dilakukan, maka dari itu dalam penelitian ini ditentukan *e-commerce* sebagai tempat pengambilan data produk adalah Tokopedia. Hasil pengumpulan data melalui *web scraping* yang dilakukan dari tiap produk yang berlabel halal dan tidak yaitu sebanyak 240-246 data. Sehingga hasil akhir pengumpulan data dari 30 produk yaitu sebanyak 7.234 data dengan 7 variabel yaitu "unnamed column", "Nama Barang", "Harga", "Terjual", "Rating", "Lokasi", dan "Label" yang telah di simpan kedalam *file* CSV seperti Gambar 4.9. *Dataset* disimpan dalam format *file* CSV juga dikarenakan akan digunakan untuk analisis melalui *software* SPSS guna membandingkan hasil yang diperoleh.

	A	B	C	D	E	F	G
1		Nama Barang	Harga	Terjual	Rating	Lokasi	Label
2		0 Indomie Satuan/Lusinan/Karton	Rp2.900	Terjual 8 rb+	4.9	Jakarta Selatan	yes
3		1 indomie goreng 5 pc - goreng rica	Rp15.800	Terjual 100+	5	Kab. Bekasi	yes
4		2 Indomie Goreng 1 Dus Isi 40 pcs 85G	Rp115.000	Terjual 500+	5	Tangerang Selatan	yes
5		3 INDOMIE GORENG INSTAN 85 GRAM	Rp2.900	Terjual 10 rb+	4.8	Jakarta Barat	yes
6		INDOMIE GORENG - ALL VARIANT - 4 PCS/DOZEN/CARTON	Rp3.100	Terjual 3 rb+	4.9	Jakarta Selatan	yes
7		INDOMIE MIE GORENG SPECIAL 5 INDOMIE GORENG	Rp2.850	Terjual 10 rb+	4.9	Jakarta Timur	yes
8		6 INDOMIE MIE RASA SOTO MIE 70GR	Rp2.650	Terjual 5 rb+	5	Jakarta Timur	yes

Gambar 4.9 Contoh *web scraping* dalam format *file CSV*

Terdapat enam variabel bertipe *object* dan satu variabel bertipe *float*. Rincian nama variabel serta tipe data dapat dilihat pada Tabel 4.1.

Tabel 4.1 Daftar nama variabel beserta tipe data

Nama Variabel	Tipe Data
<i>Unnamed column</i>	<i>Object</i>
Nama produk	<i>Object</i>
Harga	<i>Object</i>
Terjual	<i>Object</i>
Rating	<i>Float</i>
Lokasi	<i>Object</i>
Label	<i>Object</i>

## 4.2 *Preprocessing*

Hasil dari pengambilan data menggunakan metode *web scraping* masih banyak yang tidak konsisten seperti tipe data yang tidak sesuai serta terdapat teks dan simbol yang tidak dibutuhkan. *Preprocessing* dilakukan melibatkan serangkaian langkah untuk membersihkan, mengubah, dan mempersiapkan data mentah sebelum dilakukan analisis lebih lanjut. Dalam tahapan ini data yang ada hanya membutuhkan dua proses, yaitu *data cleaning* dan *data transformation*. Dalam implementasinya, *preprocessing* akan dilakukan secara iteratif antara proses *data cleaning* dan *data transformation*. Contoh data hasil *web scraping* sebelum dilakukan *preprocessing* terdapat pada Tabel 4.2.

Tabel 4.2 Data hasil *web scraping* produk *e-commerce* Tokopedia

	Nama Barang	Harga	Terjual	Rating	Lokasi	Label
0	Indomie Satuan/Lusinan/Karton	Rp2.900	Terjual 8 rb+	4.9	Jakarta Selatan	yes
1	indomie goreng 5 pc - goreng rica	Rp15.800	Terjual 100+	5	Kab. Bekasi	yes
2	Indomie Goreng 1 Dus Isi 40 pcs 85G	Rp115.000	Terjual 500+	5	Tangerang Selatan	yes

#### 4.2.1 Iterasi 1 *Data Cleaning*

Tahapan *data cleaning* dimulai dengan menghapus kolom pertama yang berisi angka yang tidak relevan dalam analisis ini. Selanjutnya, langkah ini melibatkan eliminasi data duplikat dan data yang memiliki nilai kosong. Dalam proses tersebut, data yang tidak memenuhi kriteria yang sesuai dengan variabel “Lokasi” juga akan dihapus. Tujuannya adalah untuk memastikan bahwa *dataset* yang digunakan dalam analisis selanjutnya terbebas dari data yang tidak terkait yang dapat mempengaruhi hasil akhir. Selanjutnya, yaitu mengubah tipe data yang tidak cocok dengan yang semestinya dan menyamakan pola yang sesuai. *Regex (regular expression)* digunakan untuk menyamakan pola dalam variabel “Harga” dan “Terjual” di mana terdapat karakter-karakter yang tidak diperlukan atau mengganggu dalam data. Kode program untuk membersihkan data pada iterasi pertama di tunjukkan dalam Gambar 4.10.

Dalam kode program untuk membersihkan data ini melibatkan beberapa fungsi. Fungsi `drop()` yang digunakan untuk menghapus baris atau kolom tertentu dari *dataframe*. Fungsi `drop_duplicates()` digunakan untuk menghapus baris yang memiliki nilai duplikat. Fungsi `dropna()` digunakan untuk menghapus baris yang memiliki nilai yang kosong. Fungsi `replace()` yang digunakan untuk mengganti nilai tertentu dengan nilai baru.

```
# Hapus kolom pertama
df.drop(df.columns[0], axis=1, inplace=True)

# Hapus data duplikat
df = df.drop_duplicates()

# Hapus data kosong
df=df.dropna()

# Hapus lokasi = dilayani tokopedia
df.drop(df[df['Lokasi'] == 'dilayani tokopedia'].index, inplace=True)

# Regex
import re
```

```
df['Harga'] = df['Harga'].str.replace(r'^Rp|\.', '', regex=True)
df['Terjual'] = df['Terjual'].str.replace(r'\bTerjual\b', '',
regex=True).str.replace(r'\+', '', regex=True).str.replace(r'\rb',
'000').str.replace(" ", "")
df.head()
```

Gambar 4.10 Kode program *data cleaning*

*Output* dari tahapan *data cleaning* iterasi pertama yaitu mengurangi data sebanyak 2.220 data (sehingga menjadi 5.014 data) dan satu kolom kosong (*unnamed column*). Selain itu, tahapan ini menghasilkan data yang bersih dari data duplikat dan *missing value* seperti yang ditunjukkan dalam Gambar 4.11.

```
<class 'pandas.core.frame.DataFrame'>
Index: 5014 entries, 0 to 7233
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Nama Barang     5014 non-null   object
1   Harga           5014 non-null   object
2   Terjual         5014 non-null   object
3   Rating          5014 non-null   float64
4   Lokasi          5014 non-null   object
5   Label           5014 non-null   object
dtypes: float64(1), object(5)
memory usage: 274.2+ KB
```

Gambar 4.11 *Output* kode program *data cleaning*

Pada variabel “Harga” dan “Terjual” telah dilakukan penyetaraan pola dengan menghilangkan karakter “Rp”, “.”, “Terjual”, “+”, serta mengganti “rb” dengan “000” seperti yang ditunjukkan pada Gambar 4.12.

	Nama Barang	Harga	Terjual	Rating	Lokasi	Label
1	indomie goreng 5 pc - goreng rica	15800	100	5.0	Kab. Bekasi	yes
10	Paket Bundling Indomie Mie Instant Rasa Soto P...	12200	7	4.9	Tangerang	yes
11	Paket Bundling Indomie Kuah	59600	6	4.9	Tangerang	yes
19	INDOMIE KARI AYAM 5 PCS	14500	250	4.9	Jakarta Barat	yes
23	Indomie Soto Paket 10 Pcs	29500	250	4.8	Jakarta Barat	yes

Gambar 4.12 *Output* kode program *data cleaning*

### 4.2.2 Data Transformation

Tahapan *data transformation* dilakukan dengan mengonversi tipe data dari dua variabel yaitu “Harga” dan “Terjual” menjadi tipe data *integer* agar data tersebut dapat digunakan untuk perhitungan. Tahapan ini dilakukan dengan menjalankan kode program dalam Gambar 4.13.

```
# Data transformation
df[["Harga", "Terjual"]] = df[["Harga", "Terjual"]].astype(int)
df.info()
```

Gambar 4.13 Kode *data transformation*

Hasil dari tahapan *data transformation* dapat dilihat pada Gambar 4.14. Kedua variabel yang bertipe *object* telah diubah menjadi *integer*. Dalam tahapan ini tidak ada pengurangan jumlah data.

```
<class 'pandas.core.frame.DataFrame'>
Index: 5014 entries, 0 to 7233
Data columns (total 6 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Nama Barang  5014 non-null   object
1   Harga        5014 non-null   int32
2   Terjual      5014 non-null   int32
3   Rating       5014 non-null   float64
4   Lokasi       5014 non-null   object
5   Label        5014 non-null   object
dtypes: float64(1), int32(2), object(3)
memory usage: 235.0+ KB
```

Gambar 4.14 Output kode program *data transformation*

### 4.2.3 Iterasi 2 Data Cleaning

Pada tahapan *data cleaning* iterasi kedua, dilakukan untuk menangani nilai *outliers*. Untuk mendeteksi *outlier* dalam data numerik dengan menggunakan *boxplot* yang menerapkan metode Interkuartil. Data numerik dalam *dataset* ini yaitu terdapat pada variabel “Harga” dan “Terjual”. Penanganan *outlier* yang dilakukan adalah dengan menghilangkan data yang dianggap *outlier* dengan menjalankan kode program pada Gambar 4.15. Dalam kode program interkuartil tersebut akan dicari nilai kuartil 1 dan 3. Kemudian data yang memiliki nilai diluar dari nilai kuartil 1 dan 3 atau yang disebut *outlier* akan dihapus.

Hasil pada tahap ini ditunjukkan pada Gambar 4.16 (a) *boxplot* variabel “Harga” sebelum dan (b) *boxplot* variabel “Harga” setelah dilakukan penghapusan data *outliers*. Dalam *boxplot*, data *outlier* ditandai dengan titik-titik yang menunjukkan data yang terletak jauh dari nilai-nilai

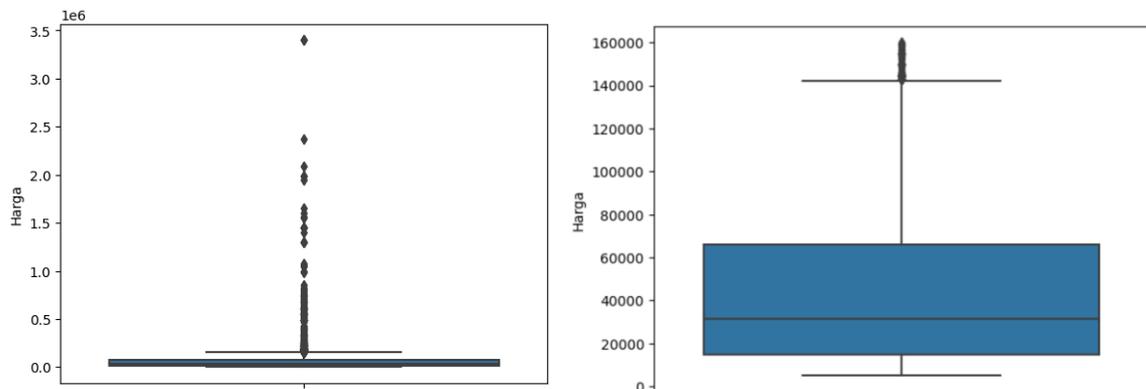
lain dalam distribusi. Namun, dalam grafik tersebut masih terdapat data *outlier* setelah dilakukan penghapusan.

```
# IQR
lowerlimit = df['Harga'].quantile(0.05)
print(lowerlimit)
len(df[df['Harga'] < lowerlimit])

upperlimit = df['Harga'].quantile(0.95)
print(upperlimit)
len(df[df['Harga'] > upperlimit])

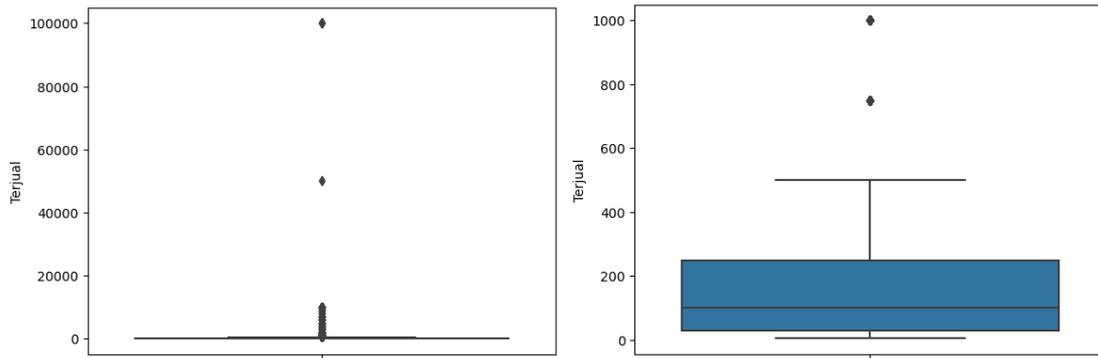
df = df[(df['Harga'] > lowerlimit) & (df['Harga'] < upperlimit)]
print(df)
```

Gambar 4.15 Kode program interkuartil (IQR)



Gambar 4.16 (a) *Boxplot* variabel “Harga” sebelum dilakukan penghapusan *outlier*; (b) *Boxplot* variabel “Harga” setelah dilakukan penghapusan *outlier*

Penanganan *outlier* dengan cara penghapusan data *outlier* pada variabel “Terjual” juga masih menyisakan *outlier* seperti yang dapat dilihat pada Gambar 4.17 (a) *boxplot* variabel “Terjual” sebelum dan (b) *boxplot* variabel “Terjual” setelah dilakukan penghapusan *outlier*. Hal tersebut dapat disebabkan oleh distribusi data dalam kedua variabel tersebut memiliki kemiringan (*skewness*) dengan nilai data sangat ekstrem. Jumlah data yang dihapus pada tahapan penanganan *outliers* ini yaitu sebanyak 1.255 data, sehingga data yang tersisa adalah sebanyak 3.759 data.



Gambar 4.17 (a) *Boxplot* variabel “Terjual” sebelum dilakukan penghapusan *outlier*; (b) *Boxplot* variabel “Terjual” setelah dilakukan penghapusan *outlier*

### 4.3 Exploratory Data Analysis (EDA)

Untuk melihat karakteristik dari *dataset* yang digunakan, dilakukan analisis data menggunakan *Exploratory Data Analysis* (EDA). Dilakukan perhitungan beberapa data statistik seperti nilai minimum, maksimum, persentil, nilai rata-rata, dan standar deviasi dari nilai numerik dengan menggunakan fungsi ``describe()`` yang berguna untuk mendapatkan gambaran mengenai *dataset*.

	Harga	Terjual	Rating
count	3759	3759	3759
mean	45740	209	5
std	39570	290	0
min	4999	5	2
25%	15000	29	5
50%	30500	100	5
75%	66000	250	5
max	159900	1000	5

Gambar 4.18 Ringkasan statistik dari nilai numerik

Gambar 4.18 merupakan *dataframe* yang menampilkan ringkasan statistik dari *dataset* nilai masing-masing variabel seperti jumlah baris (*count*), nilai rata-rata (*mean*), standar deviasi (*std*), nilai paling rendah (*min*), nilai paling tinggi (*max*), kuartil bawah (25%), kuartil tengah/median (50%), dan kuartil atas (75%). Dari ringkasan tersebut dapat diketahui bahwa tiap variabel memiliki nilai yang cukup berbeda. Pada variabel "Rating", standar deviasi yang

diperoleh adalah nol, menunjukkan bahwa variabel tersebut memiliki nilai yang identik atau dengan kata lain tidak terdapat keberagaman data.

Selain itu, dalam tahapan ini akan dilakukan beberapa visualisasi untuk memperjelas gambaran terhadap *dataset* tersebut. Gambar 4.20 merupakan *countplot* yang digunakan untuk memvisualisasikan jumlah produk berlabel halal dan tidak pada data yang telah melalui tahapan *preprocessing*. Kode program untuk menghasilkan visualisasi tersebut terdapat pada Gambar 4.19. Terdapat dua *library* yang digunakan yaitu ``seaborn`` dan ``pyplot`` yang digunakan untuk visualisasi. Hasilnya yaitu produk berlabel halal memiliki jumlah data sebanyak 2.133 dan pada produk tidak berlabel halal sebanyak 1.626 data.

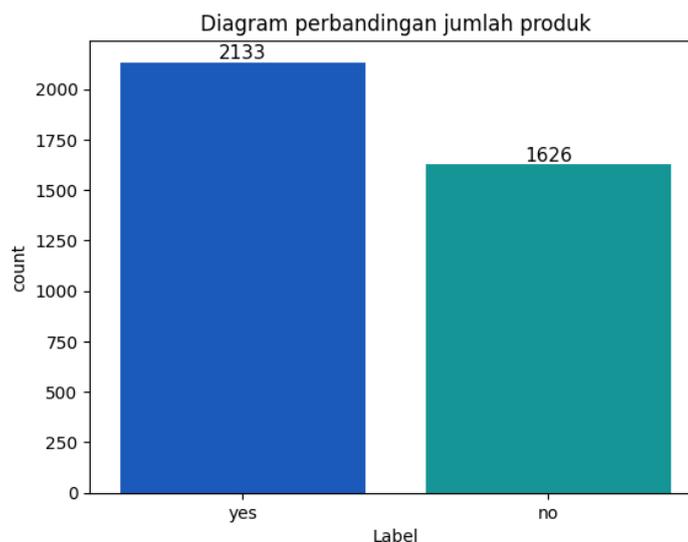
```
# import seaborn as sns
# from matplotlib import pyplot as plt

# Menggambarkan diagram
sns.countplot(x='Label', data=df, palette='winter')

# Menambahkan label pada setiap batang dalam diagram
for p in plt.gca().patches:
    plt.gca().annotate(f'{int(p.get_height())}', (p.get_x() + p.get_width()
/ 2, p.get_height()),
                      ha='center', va='center', fontsize=11,
                      color='black', xytext=(0, 5),
                      textcoords='offset points')

# Menampilkan plot
plt.show()
```

Gambar 4.19 Kode program visualisasi *countplot*



Gambar 4.20 Perbandingan jumlah produk berlabel halal dan tidak

Gambar 4.22 merupakan diagram yang digunakan untuk memvisualisasikan perbandingan jumlah penjualan produk berlabel halal dan tidak pada data yang telah melalui tahapan *preprocessing*. Kode program dapat dilihat pada Gambar 4.21. Hasilnya yaitu produk berlabel halal memiliki total penjualan sebanyak 514.655 dan pada produk tidak berlabel halal sebanyak 271.252. Sehingga produk berlabel halal memiliki total penjualan yang lebih banyak dari produk yang tidak berlabel halal.

```
# Menghitung jumlah produk yang terjual berdasarkan label
grouped_data = df.groupby('Label')['Terjual'].sum()

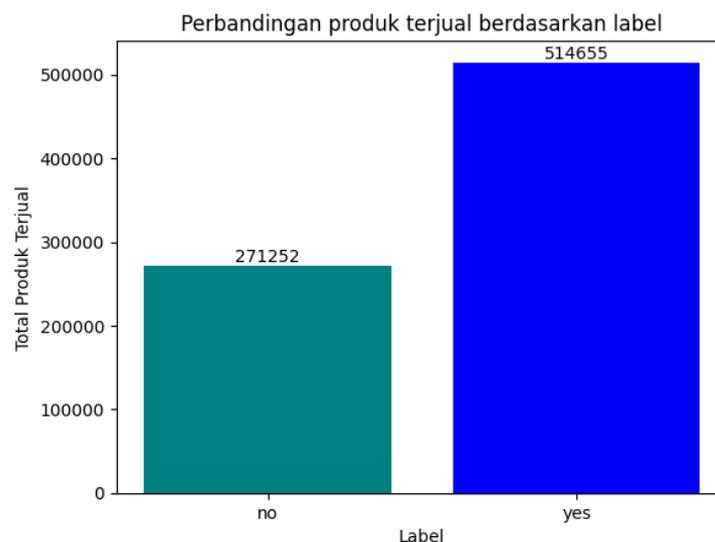
# Plotting diagram
colors = ["teal", "blue"]
bars = plt.bar(grouped_data.index, grouped_data, color=colors)

# Menambahkan nilai histogram di atas setiap batang
for bar in bars:
    yval = bar.get_height()
    plt.text(bar.get_x() + bar.get_width()/2, yval, str(yval), ha='center',
             va='bottom', fontsize=10, color='black')

# Menambah label
plt.xlabel('Label')
plt.ylabel('Total Produk Terjual')
plt.title('Histogram produk terjual berdasarkan label')

# Show plot
plt.show()
```

Gambar 4.21 Kode program visualisasi diagram



Gambar 4.22 Perbandingan total penjualan produk berlabel halal dan tidak

Gambar 4.24 merupakan diagram yang digunakan untuk memvisualisasikan perbandingan rata-rata harga terhadap produk berlabel halal dan tidak. Kode program ditunjukkan pada Gambar 4.23. Hasilnya yaitu produk berlabel halal memiliki rata-rata harga yang lebih kecil yaitu sebesar 36.675 rupiah dari pada produk tidak berlabel halal yaitu sebesar 57.633 rupiah.

```
# Menghitung jumlah produk yang terjual berdasarkan label
grouped_data = df.groupby('Label')['Harga'].mean()

grouped_data = grouped_data.round(0).astype(int)

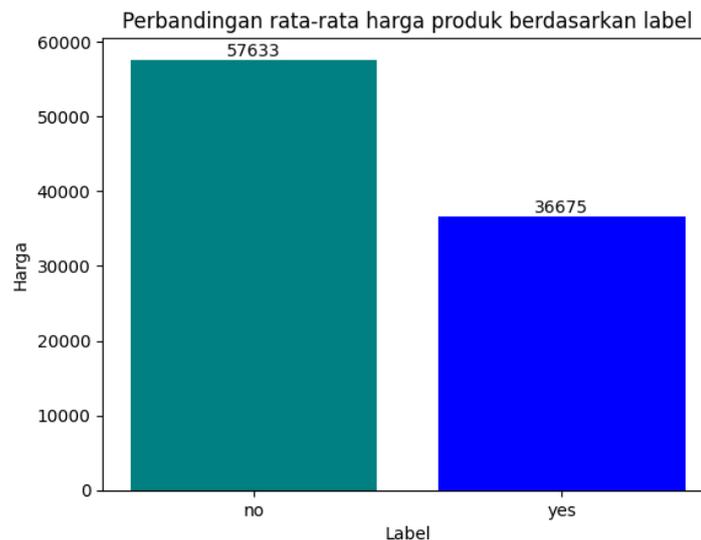
# Plotting diagram batang
colors = ["teal", "blue"]
bars = plt.bar(grouped_data.index, grouped_data, color=colors)

# Menambahkan nilai histogram di atas setiap batang
for bar in bars:
    yval = bar.get_height()
    plt.text(bar.get_x() + bar.get_width()/2, yval, str(yval), ha='center',
va='bottom', fontsize=10, color='black')

# Menambah label
plt.xlabel('Label')
plt.ylabel('Harga')
plt.title('Histogram rata-rata harga produk berdasarkan label')

plt.show()
```

Gambar 4.23 Kode program visualisasi diagram



Gambar 4.24 Perbandingan rata-rata harga produk berdasarkan label

Dalam dataset yang digunakan terdapat variabel “Lokasi” yang menunjukkan asal/daerah penjual produk tersebut. Untuk mengecek daerah tersebut digunakan fungsi ``unique()`` seperti ditunjukkan pada Gambar 4.25 dan *output* dari kode tersebut pada Gambar 4.26.

```
df.Lokasi.unique()
```

Gambar 4.25 Kode program cek lokasi

```
array(['Kab. Bekasi', 'Tangerang Selatan', 'Tangerang', 'Jakarta Barat',
      'Malang', 'Kab. Sidoarjo', 'Jakarta Utara', 'Jakarta Pusat',
      'Jakarta Selatan', 'Jakarta Timur', 'Bandung', 'Kab. Tangerang',
      'Bekasi', 'Depok', 'Surabaya', 'Medan', 'Denpasar', 'Kab. Bogor',
      'Bogor', 'Makassar', 'Semarang', 'Kab. Bandung', 'Palembang',
      'Cimahi', 'Pekanbaru', 'Kab. Garut', 'Batam', 'Kab. Sleman',
      'Kab. Magelang', 'Kab. Mojokerto', 'Yogyakarta', 'Bandar Lampung',
      'Pontianak', 'Tanjung Pinang', 'Kab. Malang', 'Kab. Bandung Barat',
      'Kab. Cirebon', 'Surakarta', 'Serang', 'Kab. Tuban',
      'Kab. Gianyar', 'Kab. Badung', 'Cirebon', 'Salatiga', 'Balikpapan',
      'Kab. Sukoharjo'], dtype=object)
```

Gambar 4.26 *Output* kode program cek lokasi

Terdapat 46 daerah yang terdaftar di *e-commerce* Tokopedia yang ditunjukkan dalam Tabel 4.1. Berdasarkan data tersebut, dapat disimpulkan bahwa sebagian besar penjual produk makanan dan minuman berasal dari Jakarta Barat.

Tabel 4.1 Lokasi Penjualan

No	Lokasi	Jumlah	No	Lokasi	Jumlah
1	Jakarta Barat	822	24	Kab. Sidoarjo	7
2	Jakarta Selatan	571	25	Palembang	6
3	Jakarta Utara	431	26	Tanjung Pinang	6
4	Jakarta Timur	355	27	Surakarta	4
5	Jakarta Pusat	236	28	Cimahi	4
6	Tangerang Selatan	220	29	Yogyakarta	3
7	Kab. Tangerang	208	30	Kab. Cirebon	3
8	Tangerang	193	31	Kab. Mojokerto	3
9	Surabaya	172	32	Kab. Tuban	3
10	Bandung	105	33	Kab. Badung	2
11	Bekasi	105	34	Bandar Lampung	2
12	Depok	99	35	Pontianak	2
13	Semarang	35	36	Kab. Bandung Barat	2
14	Medan	28	37	Kab. Magelang	1
15	Kab. Bekasi	26	38	Balikpapan	1
16	Malang	17	39	Salatiga	1

17	Bogor	15	40	Cirebon	1
18	Makassar	13	41	Pekanbaru	1
19	Kab. Bogor	12	42	Kab. Gianyar	1
20	Batam	11	43	Serang	1
21	Kab. Bandung	10	44	Kab. Malang	1
22	Kab. Sleman	9	45	Kab. Garut	1
23	Denpasar	9	46	Kab. Sukoharjo	1

Gambar 4.28 merupakan *stacked bar plot* yang menunjukkan proporsi produk terjual berdasarkan label dan lokasi, dengan fokus pada lima lokasi teratas dari penjual produk. Kode program dapat dilihat pada Gambar 4.27. Dari grafik tersebut, terlihat bahwa penjual yang berasal dari Jakarta Barat memiliki tingkat penjualan tertinggi dengan jumlah penjualan produk berlabel halal lebih besar dibandingkan produk yang tidak berlabel. Daerah lainnya seperti Jakarta Selatan, Jakarta Utara, Jakarta Timur, dan Kab. Tangerang juga menunjukkan jumlah penjualan produk berlabel halal lebih besar dibandingkan produk yang tidak berlabel. Hasil visualisasi ini menunjukkan bahwa daerah Jakarta secara keseluruhan memiliki penjualan tertinggi.

```
import pandas as pd

# Buat pivot table
pivot_table = df.pivot_table(index='Lokasi', columns='Label',
                              values='Terjual', aggfunc='sum')

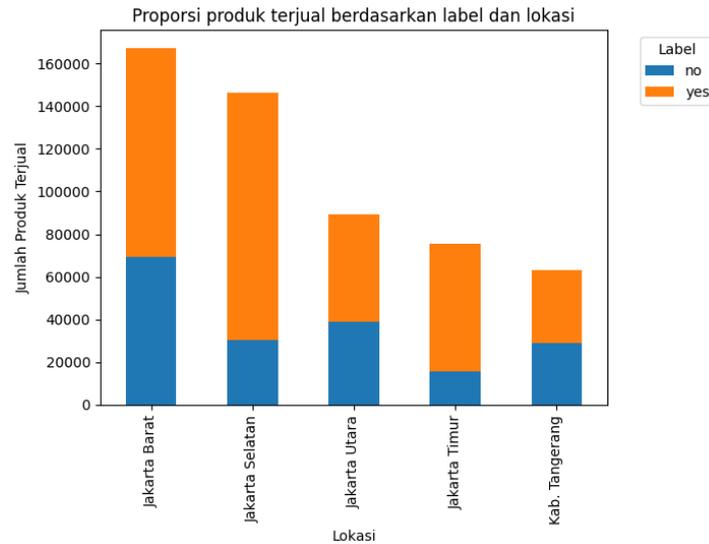
# Urut 5 lokasi berdasarkan penjualan tertinggi dan label
sorted_locations =
pivot_table.sum(axis=1).sort_values(ascending=False).head(5).index
pivot_table_top5 = pivot_table.loc[sorted_locations]

# Plotting stacked bar
pivot_table_top5.plot(kind='bar', stacked=True)

# Menambah label
plt.title('Proporsi produk terjual berdasarkan label dan lokasi')
plt.xlabel('Lokasi')
plt.ylabel('Jumlah Produk Terjual')
plt.legend(title='Label', bbox_to_anchor=(1.05, 1), loc='upper left')

plt.show()
```

Gambar 4.27 Kode program visualisasi *stacked bar plot*



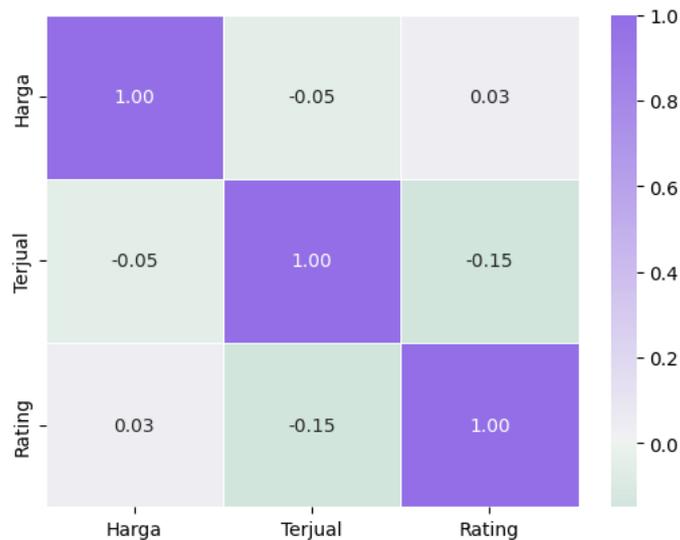
Gambar 4.28 *Stacked bar plot* proporsi produk terjual berdasarkan label dan lokasi

Korelasi merupakan metode statistik yang digunakan untuk mengukur keterkaitan antarvariabel. Korelasi antarvariabel dalam sebuah *dataset* dapat divisualisasikan menggunakan *heatmap*. *Heatmap* merupakan representasi visual dari matriks korelasi yang mana kekuatan dan arah korelasi diwakili dengan intensitas warna pada sel. Jenis korelasi yang digunakan adalah korelasi Spearman dapat diterapkan pada data yang tidak berdistribusi normal. Gambar 4.30 merupakan *heatmap* dari tiap variabel numerik dalam *dataset* yaitu “Harga”, “Terjual”, dan “Rating”. Kode program dapat dilihat pada Gambar 4.29. Hasil yang didapat yaitu *heatmap* dari hubungan ketiga variabel tidak memiliki intensitas warna yang tinggi. Variabel “Harga” dan “Terjual” memiliki korelasi negatif yang berarti ketika nilai salah satu variabel meningkat, nilai variabel lainnya cenderung menurun. Begitu pula dengan korelasi antara variabel “Terjual” dan “Rating” yang memiliki korelasi negatif. Sementara itu, nilai korelasi antara variabel “Harga” dan “Rating” menunjukkan tidak adanya korelasi, yang artinya tidak ada hubungan linear antara kedua variabel tersebut. Nilai korelasi ini tidak bertujuan untuk mengetahui sebab-akibat, namun korelasi hanya mengukur sejauh mana arah gerak dari dua variabel secara statistik.

```
# Menentukan warna
colors = sns.diverging_palette(150, 275, s=80, l=55, n=9, as_cmap=True)

# Membuat heatmap
columns_to_exclude = [0, 4, 5]
sns.heatmap(df.drop(df.columns[columns_to_exclude], axis =
1).corr(method='spearman'), center=0, cmap=colors, robust=True)

plt.show()
```

Gambar 4.29 Kode program visualisasi *heatmap*Gambar 4.30 *Heatmap* korelasi antarvariabel

#### 4.4 Uji Statistik Menggunakan Python

Uji statistik dilakukan untuk menguji hipotesis terhadap dua kelompok independen dalam hal ini produk yang berlabel halal dan tidak dengan menggunakan bahasa pemrograman Python pada Jupyter Notebook. Uji yang akan dilakukan yaitu uji nonparametrik Mann-Whitney U dan uji normalitas Kolmogorov-Smirnov untuk mengetahui distribusi data. Sebelumnya, variabel yang tidak digunakan dalam uji ini akan dihapus sehingga menyisakan variabel yang digunakan yaitu variabel “Terjual” dan “Label” dengan megeksekusi kode program pada Gambar 4.31 dan *output* pada Gambar 4.32.

```
# Hapus variabel
columns_to_drop = [0, 1, 3, 4]
df.drop(df.columns[columns_to_drop], axis=1, inplace=True)
df.head()
```

Gambar 4.31 Kode program hapus variabel

	Terjual	Label
0	100	yes
1	500	yes
2	1000	yes
3	1000	yes
4	500	yes

Gambar 4.32 *Output* lima data dalam *dataframe*

#### 4.4.1 Uji Mann-Whitney U

Pengujian menggunakan uji statistik nonparametrik yaitu uji Mann-Whitney U digunakan sebagai metode untuk mengetahui apakah ada perbedaan penjualan produk yang signifikan dari dua kelompok yang di uji. Kode program untuk uji Mann-Whitney U dapat dilihat pada Gambar 4.33.

```
# Import library
from scipy.stats import mannwhitneyu

# Pengelompokkan berdasarkan label
group1 = df[df["Label"] == "yes"]["Terjual"]
group2 = df[df["Label"] == "no"]["Terjual"]

# Uji Mann-Whitney U
if len(group1) > 0 and len(group2) > 0:
    statistic, p_value = mannwhitneyu(group1, group2, alternative='two-
sided')

    print(f"Test Statistic: {statistic}")
    print(f"P-value: {p_value}")

    alpha = 0.05

# Interpretasi hasil uji Mann-Whitney U
if p_value < alpha:
    print("Reject the null hypothesis: There is a significant
difference between the groups.")
else:
    print("Fail to reject the null hypothesis: There is no significant
difference between the groups.")
else:
    print("Cannot perform Mann-Whitney U test. One or both groups have zero
size.")
```

Gambar 4.33 Kode program uji Mann-Whitney U

Uji Mann-Whitney U ini menghasilkan temuan seperti yang ditunjukkan pada Tabel 4.2. Hasil yang didapatkan yaitu nilai *p-value* sebesar 1.7161226164364606e-17 atau sama dengan

0.00 yang artinya bahwa *p-value* yang dihasilkan sangat kecil dan kurang dari nilai *alpha* 0.05. Berdasarkan hasil uji, memiliki bukti untuk menolak hipotesis nol ( $H_0$ ) dan menerima hipotesis alternatif ( $H_a$ ). Sehingga dapat disimpulkan bahwa ada perbedaan signifikan dalam hasil penjualan produk antara dua kelompok.

Tabel 4.2 Hasil uji Mann-Whitney U

Jenis Uji	Nilai Uji Statistik	<i>p-value</i>	Hasil
Mann-Whitney U	2013136.5	1.7161226164364606e-17	Menolak $H_0$

#### 4.4.2 Uji Kolmogorov-Smirnov

Uji normalitas menggunakan Kolmogorov-Smirnov dilakukan di kedua kelompok label yang telah di kelompokkan menjadi group1 yang berisi data produk yang berlabel halal dan group2 berisi data produk yang tidak berlabel halal. Hal ini dilakukan untuk mengetahui distribusi yang dimiliki oleh kedua kelompok. Kode program uji Kolmogorov-Smirnov dapat dilihat pada Gambar 4.34. Dalam pengujian ini, terdapat dua parametrik yang digunakan yaitu nilai uji statistik dan *p-value*.

```
#import library
from scipy import stats

# Pengelompokkan berdasarkan label
group1 = df[df["Label"] == "yes"]["Terjual"]
group2 = df[df["Label"] == "no"]["Terjual"]

# Uji Kolmogorov-Smirnov pada kedua kelompok
statistic1, p_value1 = stats.kstest(group1, 'norm')
statistic2, p_value2 = stats.kstest(group2, 'norm')

# Hasil uji group1 => produk berlabel halal
print("Kolmogorov-Smirnov Test for Group 1")
print('Test statistics:', statistic1)
print('p-value:', p_value1)

# Interpretasi hasil uji Kolmogorov-Smirnov
alpha = 0.05 # Significance level
if p_value1 > alpha:
    print("p-value >", alpha, "=> Fail to reject the null hypothesis (Data
in Group 1 is approximately normally distributed)")
else:
    print("p-value <", alpha, "=> Reject the null hypothesis (Data in Group
1 is not normally distributed)")

# Hasil uji group2 => produk tidak berlabel halal
print("\nKolmogorov-Smirnov Test for Group 2")
print('Test statistics:', statistic2)
print('p-value:', p_value2)

# Interpretasi hasil uji Kolmogorov-Smirnov
```

```

if p_value2 > alpha:
    print("p-value >", alpha, "=> Fail to reject the null hypothesis (Data
in Group 2 is approximately normally distributed)")
else:
    print("p-value <", alpha, "=> Reject the null hypothesis (Data in Group
2 is not normally distributed)")

```

Gambar 4.34 Kode program uji Kolmogorov-Smirnov

Uji Kolmogorov-Smirnov menghasilkan temuan seperti yang terdapat pada Tabel 4.3. Kedua kelompok yang diujikan memiliki nilai *p-value* 0.0 yang artinya bahwa *p-value* yang dihasilkan kurang dari nilai *alpha* yaitu sebesar 0.05. Berdasarkan hasil uji, memiliki bukti untuk menolak hipotesis nol ( $H_0$ ) dan menerima hipotesis alternatif ( $H_a$ ). Sehingga dapat disimpulkan bahwa data tidak memiliki distribusi normal. Sehingga dapat disimpulkan bahwa ada perbedaan signifikan dalam hasil penjualan produk antara dua kelompok.

Tabel 4.3 Hasil uji Kolmogorov-Smirnov

Jenis Uji	Kelompok	Nilai uji statistik	<i>p-value</i>	Hasil
Kolmogorov-Smirnov	Berlabel	0.9999997133484281	0.0	Menolak $H_0$
Kolmogorov-Smirnov	Tidak Berlabel	0.9999997133484281	0.0	Menolak $H_0$

#### 4.5 Uji Statistik Menggunakan SPSS

Uji statistik juga dilakukan melalui *software* SPSS yang memiliki tujuan untuk menguji hipotesis terhadap dua kelompok independen dalam hal ini produk yang berlabel halal dan tidak. Sama seperti pada pengujian dengan menggunakan pemrograman Python, uji yang akan dilakukan melibatkan uji nonparametrik Mann-Whitney U dan uji normalitas Kolmogorov-Smirnov.

##### 4.5.1 Uji Mann-Whitney U

Uji Mann-Whitney U dengan menggunakan SPSS menghasilkan temuan yang sama seperti yang ditunjukkan pada Tabel 4.4 yaitu *p-value* yang dihasilkan kurang dari 0.05 yang berarti memiliki bukti untuk menolak hipotesis nol ( $H_0$ ) dan menerima hipotesis alternatif ( $H_a$ ). Sehingga dapat disimpulkan bahwa ada perbedaan signifikan dalam hasil penjualan produk antara dua kelompok.

Tabel 4.4 Hasil uji Mann-Whitney U SPSS

	Hasil
Mann-Whitney U	1455121.500
Sig.( $<0.05$ )	$<.001$

#### 4.5.2 Uji Kolmogorov-Smirnov

Uji Kolmogorov-Smirnov dengan menggunakan SPSS menghasilkan temuan yang sama seperti yang ditunjukkan pada Tabel 4.5 yaitu *p-value* yang dihasilkan kurang dari 0.05 yang berarti memiliki bukti untuk menolak hipotesis nol ( $H_0$ ) dan menerima hipotesis alternatif ( $H_a$ ). Sehingga dapat disimpulkan bahwa data tidak mengikuti distribusi normal.

Tabel 4.5 Hasil uji Kolmogorov-Smirnov SPSS

	Produk berlabel halal	Produk tidak berlabel halal
N	2133	1626
Nilai Uji Statistik	0.320	0.354
Sig.( $<0.05$ )	$<.001$	$<.001$

#### 4.6 Hasil Analisis

Pengujian dalam penelitian ini dilakukan menggunakan dua *tools*, yaitu pemrograman Python dan SPSS guna meningkatkan validitas atau tingkat keakuratan dari hasil pengujian tersebut. Hasil uji menggunakan kedua *tools* untuk mengetahui perbandingan antar penjualan produk berlabel halal dan tidak menunjukkan bahwa terdapat perbedaan yang signifikan dalam hasil penjualan antara kedua kelompok tersebut. Ditunjukkan oleh nilai *p-value* yang dihasilkan kurang dari nilai *alpha* yaitu 0.05. Kemudian, perbedaan nilai statistik Mann-Whitney U yang diperoleh antara kedua *tools* disebabkan oleh implementasi algoritma yang mungkin berbeda untuk perhitungan statistik meskipun konsep dasar yang digunakan sama.

Dalam penelitian ini juga dilakukan uji normalitas untuk mengevaluasi distribusi data guna menentukan apakah data dapat diuji menggunakan uji statistik parametrik. Hasil dari pengujian menunjukkan bahwa data tidak memiliki distribusi normal dimana *p-value* yang dihasilkan kurang dari nilai *alpha*. Hal ini mengindikasikan bahwa data tidak memenuhi asumsi normalitas sehingga tidak dapat di uji menggunakan metode uji statistik parametrik.

Terdapat dua kendala dalam penelitian ini. Pertama, data yang diolah memiliki nilai kemiringan (*skewness*) yang sangat tinggi yang menyebabkan data tersebut memiliki nilai ekstrem (*outlier*) yang sangat banyak yang menyebabkan nilai signifikansi atau *p-value* yang

dihasilkan sangat kecil. Kedua, data yang diambil tidak memiliki catatan waktu transaksi sehingga tidak dapat dilakukan analisis lebih lanjut untuk memprediksi pengaruh label halal tersebut dalam penjualan produk tersebut dimasa yang akan datang.

## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **5.1 Kesimpulan**

Bagian ini akan merangkum kesimpulan yang diperoleh dari penelitian ini. Hasil penelitian ini dapat memberikan jawaban terhadap rumusan masalah sebagaimana yang diuraikan pada BAB I, yakni terkait perbandingan penjualan *online* produk berlabel halal dan tidak di Indonesia. Uji Mann-Whitney U yang dilakukan menggunakan dua *tools* (Pemrograman Python dan SPSS) menghasilkan hasil yang sama yaitu adanya perbedaan penjualan produk yang signifikan antara kelompok produk yang berlabel halal dan produk yang tidak berlabel halal di Indonesia karena keduanya memiliki nilai  $p < 0.05$ . Selain itu, temuan ini diperkuat oleh hasil tahapan EDA yang memvisualisasikan perbandingan penjualan produk berlabel halal dan tidak. Visualisasi tersebut menggambarkan perbandingan jumlah penjualan produk berlabel halal yang memiliki nilai lebih tinggi dibandingkan jumlah penjualan produk tidak berlabel halal. Perbandingan rata-rata harga dari produk yang tidak berlabel halal lebih tinggi dari pada produk yang berlabel halal. Selain itu, visualisasi hasil penjualan produk tertinggi di lima daerah pada kedua kelompok menunjukkan bahwa secara keseluruhan, Jakarta menjadi daerah dengan penjualan tertinggi. Penelitian ini juga menunjukkan bahwa produk berlabel halal memiliki penjualan yang lebih tinggi dari produk yang tidak berlabel halal secara tidak langsung mengindikasikan tingginya kesadaran akan produk halal di Indonesia yang notabene penduduk muslim.

#### **5.2 Saran**

Berdasarkan analisis yang dilakukan terhadap hasil penelitian ini serta kesimpulan yang telah ditarik maka disarankan untuk menerapkan langkah lain untuk memperoleh data yang bersih dan memiliki distribusi normal. Sehingga dapat dilakukan uji statistik parametrik yang harus memenuhi beberapa asumsi yang salah satunya yaitu normalitas, guna membandingkan hasil yang telah diperoleh melalui penelitian ini. Untuk melanjutkan penelitian kedepannya, dibutuhkan catatan data transaksi untuk membuat model prediksi terkait pengaruh label halal tersebut dalam penjualan produk tersebut dimasa yang akan datang. Dengan demikian, akan membantu memperbaiki kualitas analisis serta menambah kontribusi dalam penelitian ini.

## DAFTAR PUSTAKA

- A. Yani, D. D., Pratiwi, H. S., & Muhardi, H. (2019). Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace. *Jurnal Sistem Dan Teknologi Informasi (JUSTIN)*, 7(4), 257. <https://doi.org/10.26418/justin.v7i4.30930>
- Adikaram, K. K. L. B., Hussein, M. A., Effenberger, M., & Becker, T. (2015). Data transformation technique to improve the outlier detection power of grubbs' test for data expected to follow linear relation. *Journal of Applied Mathematics*, 2015. <https://doi.org/10.1155/2015/708948>
- Al-Bara, A.-B., & Nasution, R. (2018). Analisis Pengaruh Label Halal Pada Produk Makanan Terhadap Keputusan Pembelian Konsumen Masyarakat Kota Medan. *Intiqad: Jurnal Agama Dan Pendidikan Islam*, 10(2), 227–248. <https://doi.org/10.30596/intiqad.v10i2.1987>
- Alasadi, S. and B. W. (2017). *Review of Data Preprocessing Techniques in Data Mining*. *Journal of Engineering and Applied Sciences* 12 (16).
- Alexandropoulos, S. A. N., Kotsiantis, S. B., & Vrahatis, M. N. (2019). Data preprocessing in predictive data mining. In *Knowledge Engineering Review* (Vol. 34). <https://doi.org/10.1017/S026988891800036X>
- Alinda, R., & Adinugraha, H. H. (2022). Pengaruh Logo Halal, Kesadaran Halal, Dan Sikap Konsumen Untuk Kembali Membeli Produk Makanan Dan Minuman Kemasan. *SOSEBI: Jurnal Penelitian Mahasiswa Ilmu Sosial, Ekonomi, Dan Bisnis Islam*, 2(2), 153–168. <https://doi.org/10.21274/sosebi.v2i2.6298>
- Andrade, C. (2019). The P Value and Statistical Significance: Misunderstandings, Explanations, Challenges, and Alternatives. *Indian Journal of Psychological Medicine*, 41(2), 138–143. <https://doi.org/10.4103/IJPSYM.IJPSYM>
- Arsta, I. A. D. P., & RsiRespati, N. N. (2021). The Effect Of Sales Promotion On Purchase Decisions Mediated By Brand Image (Study On E-Commerce Tokopedia In Bali). *American Journal of Humanities and Social Sciences Research*, 5(12), 205–215. [www.ajhssr.com](http://www.ajhssr.com)
- Bayu, D. (2022). *Sebanyak 86,9% Penduduk Indonesia Beragama Islam*. [Dataindonesia.Id. https://dataindonesia.id/ragam/detail/sebanyak-869-penduduk-indonesia-beragama-islam](https://dataindonesia.id/ragam/detail/sebanyak-869-penduduk-indonesia-beragama-islam)
- Boukerche, A., Zheng, L., & Alfandi, O. (2020). Outlier Detection: Methods, Models, and Classification. *ACM Computing Surveys*, 55(3). <https://doi.org/10.1145/3381028>

- Cahya, R. T., Diana, N., & Alrasyid, H. (2022). Analisis Pengaruh Halal Certification dan Halal Awareness Terhadap Minat Beli Produk Makanan Korinus Tok-Poki (Studi Pada Mahasiswa Muslim Pecinta Drama Korea). *Islamic Economics and Finance Journal*, 3(1), 1–13. <http://riset.unisma.ac.id/index.php/laswq/article/view/17593/13307>
- Canchen, L. (2019). *Preprocessing Methods and Pipelines of Data Mining: An Overview*. June, 1–7. <http://arxiv.org/abs/1906.08510>
- Darmawan, I., Maulana, M., Gunawan, R., & Widiyasono, N. (2022). Evaluating Web Scraping Performance Using XPath, CSS Selector, Regular Expression, and HTML DOM With Multiprocessing Technical Applications. *International Journal on Informatics Visualization*, 6(4), 904–910. <https://doi.org/10.30630/joiv.6.4.1525>
- Elveny, M., Hardi, S. M., Jaya, I., & Gundari, P. (2021). Web-based E-Commerce Products Grouping. *Journal of Physics: Conference Series*, 1898(1). <https://doi.org/10.1088/1742-6596/1898/1/012018>
- Emerson, R. W. (2023). Mann-Whitney U test and t-test. *Journal of Visual Impairment and Blindness*, 117(1), 99–100. <https://doi.org/10.1177/0145482X221150592>
- Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., & Tabona, O. (2021). A survey on missing data in machine learning. In *Journal of Big Data* (Vol. 8, Issue 140). Springer International Publishing. <https://doi.org/10.1186/s40537-021-00516-9>
- Gerald, B., & Patson, T. F. (2021). Parametric and Nonparametric Tests: A Brief Review. *International Journal of Statistical Distributions and Applications*, 7(3), 78. <https://doi.org/10.11648/j.ijds.20210703.12>
- Gunawan, R., Rahmatulloh, A., Darmawan, I., & Firdaus, F. (2019). *Comparison of Web Scraping Techniques : Regular Expression, HTML DOM and Xpath*. 2(IcoIESE 2018), 283–287. <https://doi.org/10.2991/icoiese-18.2019.50>
- Henry, K. (2021). Importance of Web Scraping in E-Commerce and E-Marketing. *SSRN Electronic Journal*, January, 1–10. <https://doi.org/10.2139/ssrn.3769593>
- Hu, D., Yu, X., & Wang, J. (2017). Statistical Inference in Rough Set Theory Based on Kolmogorov-Smirnov Goodness-of-Fit Test. *IEEE Transactions on Fuzzy Systems*, 25(4), 799–812. <https://doi.org/10.1109/TFUZZ.2016.2578344>
- Juliana, J., Wafa Rizaldi, M., Azizah Al-Adawiyah, R., & Marlina, R. (2022). Halal Awareness : Pengaruhnya Terhadap Keputusan Pembelian Konsumen Shopee dengan Religiositas Sebagai Variabel Moderasi. *Coopetition : Jurnal Ilmiah Manajemen*, 13(2), 169–180. <https://doi.org/10.32670/coopetition.v13i2.1423>

- Juniwati, & Listiawati, F. (2021). Ketika Konsumen Muslim Berminat Membeli Ulang Makanan Faktor Apa yang Mempengaruhinya. *Equator Journal of Management and Entrepreneurship (EJME)*, 9(4), 221–239. <https://doi.org/10.26418/ejme.v9i4.52855>
- Karadimitriou, S. M., & Marshall, E. (2015). Mann-Whitney U Test. *Dictionary of Statistics & Methodology*. <https://doi.org/10.4135/9781412983907.n1115>
- Kaur, P., Stoltzfus, J., & Yellapu, V. (2018). Descriptive Statistics. *International Journal of Academic Medicine*, 4(1). [https://doi.org/10.1007/978-3-030-69901-7\\_12](https://doi.org/10.1007/978-3-030-69901-7_12)
- Kurniawan, M. I., Maryana, M., Subroto, P., & Ulfah, Y. (2021). The Impact of Merger Company on the Value of Case Study Stocks on Merger Gojek and Tokopedia. *Proceedings on Engineering Sciences*, 3(4), 425–432. <https://doi.org/10.24874/PES03.04.006>
- Levi, M., Palit, H. N., & Rostianingsih, S. (2020). *Perbandingan Performa Tools Web Scraping pada Website dengan Data Statis dan Dinamis*.
- Lupi, F. R., & Nurdin. (2016). Analisis strategi pemasaran dan penjualan. *Jurnal Elektronik Sistem Informasi Dan Komputer*, 2(1), 20–31.
- Maringka, R., & Kusnawi, K. (2021). Exploratory Data Analysis Faktor Pengaruh Kesehatan Mental di Tempat Kerja. *CogITo Smart Journal*, 7(2), 215–226. <https://doi.org/10.31154/cogito.v7i2.312.215-226>
- Mubarok, A., Sahroni, & Sunanto. (2021). MANN WHITNEY TEST IN COMPARING THE STUDENTS' CONSULTATION RESULTS OF ENTREPRENEURIAL PRACTICE BETWEEN MALE AND FEMALE LECTURERS IN ECONOMIC FACULTY OF PAMULANG UNIVERSITY. *Procuration*, 9(2), 9–15.
- Murni, S., & Fajrina, N. (2021). Pengaruh Labelisasi Halal Terhadap Keputusan Pembelian Makanan Kemasan. *Jurnal Ilmiah Akuntansi*, 8(2), 47.
- Mustafa, P. S. (2022). Statistika Inferensial meliputi Uji Beda dalam Pendidikan Jasmani: Sebuah Tinjauan. *DIDAKTIKA: Jurnal Pemikiran Pendidikan*, 28(2(1)), 71–86. [https://doi.org/10.30587/didaktika.v28i2\(1\).4166](https://doi.org/10.30587/didaktika.v28i2(1).4166)
- Nair, A., Paralkar, C., Pandya, J., Chopra, Y., & Krishnan, D. (2021). Comparative Review on Sentiment analysis-based Recommendation system. *2021 6th International Conference for Convergence in Technology, I2CT 2021*, 2–7. <https://doi.org/10.1109/I2CT51068.2021.9418222>
- Nair, P., & Kashyap, I. (2019). Hybrid Pre-processing Technique for Handling Imbalanced Data and Detecting Outliers for KNN Classifier. *Proceedings of the International*

- Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Perspectives and Prospects, COMITCon 2019*, 460–464.  
<https://doi.org/10.1109/COMITCon.2019.8862250>
- Najmi, A., Balakrishnan, S., & Ray, A. (2021). How to choose and interpret a statistical test? An update for budding reserchers. *Journal of Family Medicine and Primary Care*.  
<https://doi.org/10.4103/jfmpe.jfmpe>
- Nasri, H. (2020). PENGARUH BRAND AMBASSADOR BTS TERHADAP PURCHASE INTENTION YANG DIMEDIASI OLEH BRAND AWARENESS TOKOPEDIA DI INDONESIA. *Jurnal Ilmiah Mahasiswa FEB Universitas Brawijaya*, 8.  
<http://nylonthailand.com/asianbrandambassador>
- Nongthombam, K. (2021). Data Analysis Using Python. *International Journal of Engineering Research and Technology (IJERT)*, 10(07), 463–468.
- Nyamathulla, S., Ratnababu, P., Shaik, N. S., & N, B. L. (2021). A Review on Selenium Web Driver with Python. *Annals of the Romanian Society for Cell Biology*, 25(4), 16760–16768. <http://annalsofrscb.ro>
- Onyenwe, I., Onyedinma, E., Nwafor, C., & Agbata, O. (2021). Developing Products Update-Alert System for E-Commerce Websites Users using Html Data and Web Scraping Technique. *International Journal on Natural Language Computing*, 10(5), 01–07.  
<https://doi.org/10.5121/ijnlc.2021.10501>
- Orcan, F. (2020). Parametric or Non-parametric: Skewness to Test Normality for Mean Comparison. *International Journal of Assessment Tools in Education*, 7(2), 255–265.  
<https://doi.org/10.21449/ijate.656077>
- Ostrowski, J. G., & Menyhárt, J. (2020). Statistical analysis of machinery variance by python. *Acta Polytechnica Hungarica*, 17(5), 151–168.  
<https://doi.org/10.12700/APH.17.5.2020.5.8>
- Pandiangan, S. M. T., Oktafani, F., Panjaitan, S. R., Shifa, M., & Jefri, R. (2022). Analysis of Public Ownership and Management Ownership on the Implementation of the Triple Bottom Line in the Plantation Sector Listed on the Indonesia Stock Exchange. *Budapest International Research and Critics Institute-Journal*, 5(1), 3489–3497.  
<https://doi.org/10.33258/birci.v5i1.4016>
- Peng, J., Wu, W., Lockhart, B., Bian, S., Yan, J. N., Xu, L., Chi, Z., Rzeszotarski, J. M., & Wang, J. (2021). DataPrep.EDA: Task-Centric Exploratory Data Analysis for Statistical Modeling in Python. *Proceedings of the ACM SIGMOD International Conference on*

- Management of Data*, 2271–2280. <https://doi.org/10.1145/3448016.3457330>
- Pramintasari, T. R., & Fatmawati, I. (2017). Pengaruh Keyakinan Religius, Peran Sertifikasi Halal, Paparan Informasi, dan Alasan Kesehatan Terhadap Kesadaran Masyarakat Pada Produk Makanan Halal. *Jurnal Manajemen Bisnis*, 8(1), 1–33. <https://journal.umy.ac.id/index.php/mb/article/view/3922/3384>
- Puspitasari, R. H. U. (2019). Analisis Pengaruh Keputusan Pembelian Pada Kosmetik Berlabel Halal. *BMAJ: Business Management Analysis Journal*, 2(1), 68–77. <https://doi.org/10.24176/bmaj.v2i1.3212>
- Quraisy, A., Wahyuddin, & Hasni, N. (2021). Analisis Kruskal-Wallis Terhadap Kemampuan Numerik Siswa. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 3(3), 156–161. <https://doi.org/10.35580/variensiunm29957>
- Rachman, A., & Ulpah, M. (2022). Desain Baru Logo Halal Badan Penyelenggara. *El-Buhuth*, 4(2), 245–262.
- Radhi, M., Amalia, A., Sitompul, D. R. H., Sinurat, S. H., & Indra, E. (2022). Analisis Big Data Dengan Metode Exploratory Data Analysis (Eda) Dan Metode Visualisasi Menggunakan Jupyter Notebook. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima (JUSIKOM PRIMA)*, 4(2), 23–27. <https://doi.org/10.34012/journalsisteminformasidanilmukomputer.v4i2.2475>
- Razali, N. M., & Yap, B. W. (2011). Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. *Journal of Statistical Modeling and Analytics*, 2(1), 13–14.
- Resmawati, R. (2018). Analisis perbandingan volume penjualan sebelum dan sesudah mendapatkan sertifikasi halal. *E-Proceeding of Applied Science*, 4(2), 267–273.
- Rizaldi, T., & Arief, H. (2017). Perbandingan Metode Web Scraping Menggunakan CSS Selector dan Xpath Selector. *Teknika*, 6(1), 43–46. <https://doi.org/10.34148/teknika.v6i1.56>
- Sagita, A. M., & Saputra, A. (2022). Perlindungan Hukum Terhadap Konsumen Makanan Impor Tanpa Izin Edar Yang Dijual Melalui Aplikasi Shopee. *Ajudikasi : Jurnal Ilmu Hukum*, 6(1), 73–88. <https://doi.org/10.30656/ajudikasi.v6i1.4670>
- Sahoo, K., Samal, A. K., Pramanik, J., & Pani, S. K. (2019). Exploratory data analysis using python. *International Journal of Innovative Technology and Exploring Engineering*, 8(12), 4727–4735. <https://doi.org/10.35940/ijitee.L3591.1081219>
- Sahria, Y. (2020). Implementasi Teknik Web Scraping pada Jurnal SINTA Untuk Analisis

- Topik Penelitian Kesehatan Indonesia. *URECOL (University Research Colloquium)*, 297–306. <http://repository.urecol.org/index.php/proceeding/article/view/1079>
- Samosir, F. V. P., Mustamu, L. P., Anggara, E. D., Wiyogo, A. I., & Widjaja, A. (2021). Exploratory Data Analysis terhadap Kepadatan Penumpang Kereta Rel Listrik. *Jurnal Teknik Informatika Dan Sistem Informasi*, 7(2), 449–467. <https://doi.org/10.28932/jutisi.v7i2.3700>
- Septiani, L. (2023). *Survei APJII: UMKM Lebih Pilih Tokopedia dan Lazada*. <https://katadata.co.id/desysetyowati/digital/64645ca9cf3cf/survei-apjii-umkm-lebih-pilih-tokopedia-dan-lazada>
- Setyaningsih, E. D., & Marwansyah, S. (2019). The Effect of Halal Certification and Halal Awareness through Interest in Decisions on Buying Halal Food Products. *Syi'ar Iqtishadi: Journal of Islamic Economics, Finance and Banking*, 3(1), 65. <https://doi.org/10.35448/jiec.v3i1.5515>
- Sulistiyawati, E. S., & Widayani, A. (2020). Marketplace Shopee Sebagai Media Promosi Penjualan UMKM di Kota Blitar. *Jurnal Pemasaran Kompetitif*, 4(1), 133. <https://doi.org/10.32493/jpkpk.v4i1.7087>
- Syamsuar, G., & Megayani. (2022). *Perbandingan Pengaruh Brand Ambassadors BTS dengan Blackpink Terhadap Minat Beli pada Tokopedia Menggunakan Mann Whitney U-Test*. 1–5.
- Thota, P., & Ramez, E. (2021). Web Scraping of COVID-19 News Stories to Create Datasets for Sentiment and Emotion Analysis. *ACM International Conference Proceeding Series*, 306–314. <https://doi.org/10.1145/3453892.3461333>
- Utomo, H. (2021). PERBANDINGAN TABEL MORTALITA INDONESIA DAN TABEL MORTALITA CSO MENGGUNAKAN UJI MANN-WHITNEY DAN UJI KRUSKAL-WALLIS. *Jurnal Ilmiah Indonesia*, 6(3).
- Vinutha, H. P., Poornima, B., & Sagar, B. M. (2018). Detection of outliers using interquartile range technique from intrusion dataset. In *Advances in Intelligent Systems and Computing* (Vol. 701). Springer Singapore. [https://doi.org/10.1007/978-981-10-7563-6\\_53](https://doi.org/10.1007/978-981-10-7563-6_53)
- Wahyuni, S., Ananti, Y., & Issabella, C. M. (2021). Hubungan Anemia Kehamilan Dengan Kejadian Berat Badan Lahir Rendah (Bblr): Systematic Literatur Review. *Journal of Health (JoH)*, 8(2), 94–104. <https://doi.org/10.30590/joh.v8n2.p94-104.2021>
- Wardhana, A., Pradana, M., Shabira, H., Aji Buana, D. M., Nugraha, D. W., & Sandi, K. (2021). The Influence of Consumer Behavior on Purchasing Decision Process of

- Tokopedia E-commerce Customers in Indonesia. *Proceedings of the International Conference on Industrial Engineering and Operations Management*, 5929–5934.
- Warto, W., & Samsuri, S. (2020). Sertifikasi Halal dan Implikasinya Bagi Bisnis Produk Halal di Indonesia. *Al Maal: Journal of Islamic Economics and Banking*, 2(1), 98. <https://doi.org/10.31000/almaal.v2i1.2803>
- Wibowo, S., & Sutandi. (2018). Analisa Rasio Keuangan Garuda Indonesia Airlines, Singapore Airlines Dan Thailand Airlines Dengan Uji Non-Parametrik (Periode: 2010 – 2014). *ECo-Buss*, 1(2), 66–72.
- Yang, H. (2013). The case for being automatic: introducing the automatic linear modeling (LINEAR) procedure in SPSS statistics. *Multiple Linear Regression Viewpoints*, 39(2), 27–37.
- Yang, J., Rahardja, S., & Fränti, P. (2019). Outlier detection: How to threshold outlier scores? *ACM International Conference Proceeding Series*, 1–6. <https://doi.org/10.1145/3371425.3371427>
- Zhao, Z., Wang, J., Sun, H., Liu, Y., Fan, Z., & Xuan, F. (2020). What Factors Influence Online Product Sales? Online Reviews, Review System Curation, Online Promotional Marketing and Seller Guarantees Analysis. *IEEE Access*, 8, 3920–3931. <https://doi.org/10.1109/ACCESS.2019.2963047>

## LAMPIRAN

Lampiran A. Hasil Uji Kolmogorov-Smirnov dengan menggunakan SPSS

### One-Sample Kolmogorov-Smirnov Test

		labeled	nonlabeled	
N		2133	1626	
Normal Parameters <sup>a,b</sup>	Mean	241.28	166.82	
	Std. Deviation	312.955	251.923	
Most Extreme Differences	Absolute	.320	.354	
	Positive	.320	.354	
	Negative	-.225	-.260	
Test Statistic		.320	.354	
Asymp. Sig. (2-tailed) <sup>c</sup>		<.001	<.001	
Monte Carlo Sig. (2-tailed) <sup>d</sup>	Sig.	<.001	<.001	
	99% Confidence Interval	Lower Bound	.000	.000
		Upper Bound	.000	.000

Lampiran B. Hasil Uji Mann-Whitney U dengan menggunakan SPSS

### Test Statistics<sup>a</sup>

	Terjual
Mann-Whitney U	1455121.50 0
Wilcoxon W	2777872.50 0
Z	-8.512
Asymp. Sig. (2-tailed)	<.001