

Analisis Data Produk Marketplace Menggunakan *Web Scraping* Menggunakan Metode *Beautiful Soup* Dan *Lxml*

I Gusti Bagus Indra Permana¹, L. Ahmad S. Irfan Akbar², Cipta Ramadhani³

^{1,3}Jurusan Teknik Elektro Universitas Mataram
Jl. Majapahit 62, Mataram, Lombok NTB, INDONESIA
¹Ipermana302@gmail.com

Abstrak

Perkembangan internet dalam menyediakan informasi yang cepat memberikan dampak positif bagi para penggiat bisnis untuk menawarkan atau menjual produk melalui internet. Salah satu manfaat yang diperoleh adalah kita dapat mengetahui atau mencari berbagai macam produk yang dibutuhkan melalui Internet. Hal ini mendorong banyaknya toko online yang bermunculan di Indonesia. Dibutuhkan sebuah aplikasi yang dapat membantu pengguna dalam mencari dan mengumpulkan data produk di internet. Tujuan dari penelitian ini adalah membuat analisis data produk dari hasil dengan *web scraping* dari halaman *website* Tokopedia menggunakan metode *Beautiful Soup (Bs4)* dan *Lxml*. Pada penelitian ini menggunakan *website* Tokopedia sebagai objek dari web scraping. Perancangan *website* dilakukan menggunakan modul *streamlit* dari python. Hasil dari penelitian ini adalah analisa data dari hasil *scraping* halaman *website* Tokopedia menggunakan metode *Beautiful Soup* dan *Lxml*, dengan kategori yang ditentukan antara lain: rentang harga, rating tertinggi dan terendah, penjualan tertinggi dan terendah, harga tertinggi dan terendah, persentase lokasi, persentase toko penjual dan rekomendasi produk.

Kata kunci: *Web Scraping*, *Beautiful Soup*, *lxml*, Analisa Data

1. PENDAHULUAN

Perkembangan teknologi informasi dari hari ke hari semakin cepat, terutama internet yang memiliki kelebihan untuk menyediakan informasi yang cepat hal ini meningkatkan pengguna internet di Indonesia. Menurut data dari Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), pada tahun 2020 jumlah pengguna internet di Indonesia mencapai 196,7 juta orang atau sekitar 73,7% dari total populasi. Pengguna dapat dengan mudah mengoperasikan dan menggunakan internet sepanjang waktu dengan *smartphone*, *tablet*, *notebook* dan komputer yang dimiliki. Kemudahan yang ditawarkan layanan internet memberikan dampak *positif* bagi para penggiat bisnis untuk menawarkan atau menjual

produk melalui internet. Salah satu manfaat yang diperoleh adalah kita dapat mengetahui atau mencari berbagai macam produk yang dibutuhkan melalui Internet. Hal ini mendorong banyaknya toko online yang bermunculan di Indonesia. *Marketplace* adalah salah satu media untuk menjalankan toko online.

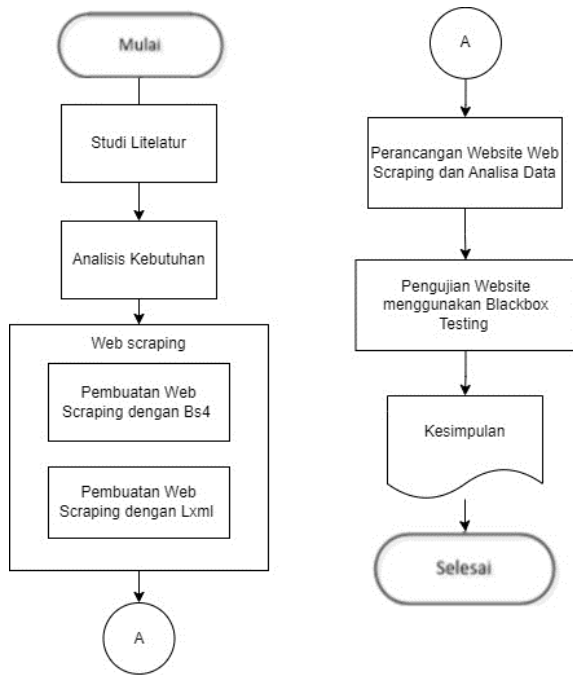
Dalam mencari data produk yang ingin di jual dari sebuah *website* atau *marketplace*, tentu akan menghabiskan banyak waktu jika mencari dan menyalin data produk satu persatu secara manual. Untuk memudahkan pengguna dalam mendapatkan data produk dari kategori yang diinginkan, dengan hasil penjualan terbaik, harga, lokasi ,serta perbandingan produk dari toko kompetitor, maka diperlukan sebuah media yang mampu mengumpulkan informasi data produk dari sebuah *website Marketplace*. Informasi tersebut berupa harga, nama produk, banyak produk terjual, rating produk, lokasi, toko yang menjual dan link produk.

Web scraping adalah teknik untuk mendapatkan informasi dari *website* secara otomatis tanpa harus menyalinnya secara manual. *Web scraping* berfokus dalam mendapatkan data dengan cara pengambilan dan ekstraksi. Manfaat dari web scraping adalah agar informasi yang dikeruk lebih terfokus sehingga memudahkan dalam melakukan pencarian sesuatu. *Web scraping* dapat menjadi salah satu alat bantu para pengusaha online dalam mengumpulkan data produk dalam waktu singkat.

Berdasarkan paparan diatas, maka dibuat rumusan masalah yaitu bagaimana menganalisis data produk dari hasil *scraping* pada halaman *website* Tokopedia menggunakan *Beautiful Soup* dan *Lxml* menjadi informasi bagi pengguna

2. METODE PENELITIAN

Penelitian ini dilakukan dari awal hingga akhir sesuai dengan alur penelitian yang digambarkan dalam Gambar 1. Alur penelitian digunakan oleh penulis dalam pelaksanaan penelitian ini agar hasil yang dicapai tidak menyimpang dari tujuan yang telah ditetapkan sebelumnya. Alur penelitian dapat dilihat dalam Gambar 1 berikut.



Tahap pertama adalah Studi literatur, Studi literatur dibutuhkan sebagai penunjang dalam menyelesaikan penelitian, yaitu dengan mengumpulkan informasi-informasi yang berkaitan dengan topik penelitian. Pada studi literatur ini berkaitan dengan metode *Web Scraping* dengan *Bs4* dan *Lxml* yang bersumber dari buku-buku, jurnal, dan lain-lain sebagai referensi

Tahap kedua adalah analisis kebutuhan, adapun berbagai kebutuhan yang mendukung untuk dilakukan penelitian ini *hardware*, *software*, dan *library* dari *Python*. Software yang digunakan adalah *microsoft excel*, *microsoft word*, *website draw.io*, aplikasi *whimsical*, *visual studio code*. Library yang digunakan adalah *selenium*, *Bs4*, *lxml*, *pandas*, *matplotlib*, *time*, *document*, *streamlit*.

Tahap ketiga adalah penentuan target *scraping*, Target dalam penelitian ini adalah *website* Tokopedia yang diakses melalui aplikasi *browser Chrome*. Scraping akan difokuskan untuk mengambil data dari halaman pencarian dengan kata kunci yang diinginkan atau halaman produk dari toko yang diinginkan. Bagian pertama yang dibutuhkan *URL* dari halaman pencarian produk yang dicari.

Tahap keempat analisis struktur *html* Pada tahap ini penulis akan menganalisis element *HTML* dari informasi yang akan di *scraping*. *Web scraping* di penelitian ini menggunakan metode parsing *HTML*. *Beautiful Soup* akan membaca kode element *HTML* dari halaman *web Tokopedia* dan mengekstrak informasi dari element yang diinginkan. Informasi yang akan di ambil adalah nama produk, harga, terjual, lokasi, rating, dan link produk.

Tahap kelima adalah pembuatan dua program *web scraping* menggunakan *beautiful soup* dan *web scraping* menggunakan *lxml*, Pada tahap ini, setelah tahap menentukan target *scraping* dan analisis

struktur akan dirancang script untuk *scraping* halaman *website Tokopedia* dengan metode parsing menggunakan *Beautiful Soup* dan *lxml*. *Beautiful Soup* dan *lxml* akan mengekstraksi informasi dari element berdasarkan hasil dari tahap analisis struktur *HTML*.

Tahap keenam adalah perancangan *website web scraping* dan analisa data produk, pada tahap ini dirancang sebuah *website* yang dapat melakukan *web scraping* dari halaman *web tokopedia*, *scraping* dari halaman toko penjual dan memberikan analisa data produk dari hasil *scraping* tersebut.

Tahap terakhir pengujian aplikasi menggunakan *blackbox testing*, pada tahap ini Setelah perancangan aplikasi *website* selesai, aplikasi akan diuji dengan metode *blackbox testing*. Penulis melakukan pengujian fungsional dan akan memastikan fungsi dari *web scraping* dan analisa data sudah berjalan seperti yang diharapkan.

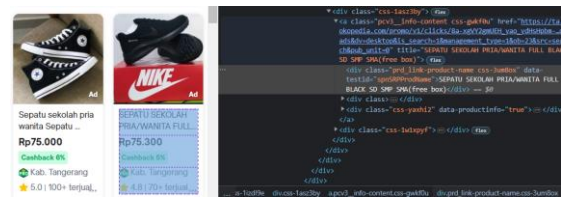
3. HASIL DAN PEMBAHASAN

3.1. Target *scraping*

Pada penelitian ini, tahap ini adalah tahapan dalam menentukan target *scraping* oleh pengguna dimana akan mengambil *url* dari halaman pencarian atau halaman toko pada *website Tokopedia*. *Url* yang akan diambil adalah *url* dari halaman pencarian dengan *keyword* ‘sepatu sekolah’ dengan *url* sebagai berikut :

https://www.tokopedia.com/search?navsource=&sh_op_tier=2%233%231&srp_component_id=04.06.00.00&srp_page_id=&srp_page_title=&st=&q=Sepatu%20sekolah

3.2 Analisis Struktur *HTML*



Gambar 1. Proses Analisa Astruktur *HTML*

Pada gambar 1 adalah tahap identifikasi element dari halaman pencarian dengan *keyword* “sepatu sekolah”. Informasi nama produk “Sepatu Sekolah Pria/Wanita Full Balck SD SMP SMA’ berada pada element *div class="prd_link-product-name css-3um8ox"*. Informasi harga produk berada pada element *div class="prd_link-product-price css-h66vau"*. Informasi terjual produk berada pada element *div class="prd_label-integrity css-1sgek4h"*. Informasi lokasi produk berada pada element *div class="prd_link-shop-loc css-1kdc32b flip"*. Informasi rating produk berada pada element *div class="prd_rating-average-text css-t70v7i"*. Informasi toko produk berada pada element *div class="prd_link-shop-name css-1kdc32b flip"*.

Informasi link produk berada pada element `a class="pcv3__info-content css-gwxf0u"`.

3.3. Web Scraping Menggunakan BeautifulSoup

Tahap ini adalah tahapan membuat program *web scraping* menggunakan *Beautiful Soup (Bs4)*. Terdapat beberapa modul yang digunakan dalam pembuatan program ini yaitu *selenium*, *Bs4* dan *pandas*. Hasil *url* dan identifikasi element dari tahapan sebelumnya ditambahkan pada program, *Bs4* akan melakukan ekstraksi dengan parsing data pada element *HTML* yang sudah ditentukan. *Pandas* akan melakukan pengolahan data dari hasil parsing data yang dilakukan *Bs4* dan akan ditampilkan dalam bentuk tabel.

| | A | Y | B | Y | C | Y | D | Y | E | Y | F | Y | G | Y |
|----|-------------------------------|---|---------|---|---------|---|--------|---|-------------------|---|----------------|---|------------------|---|
| 1 | Nama Produk | | Harga | | Terjual | | Rating | | Toko | | Lokasi | | Link | |
| 2 | Kankly Story Horjio - Sepatu | | 298.800 | | 3.000 | | 5 | | Sepatu Kankly | | Kab. Bandung | | https://ta.tokop | |
| 3 | Sepatu Converse All Star Peti | | 75.000 | | 30 | | 5 | | briantara_shoes | | Kab. Tangerang | | https://ta.tokop | |
| 4 | Spotec Sepatu Student Activ | | 252.900 | | 80 | | 5 | | Spotec Official S | | Jakarta Utara | | https://ta.tokop | |
| 5 | TOMKINS Scool Wednesday | | 249.000 | | 90 | | 5 | | TOMKINS Offici | | Bandung | | https://ta.tokop | |
| 6 | Sepatu Casual Kasogi Julio P | | 220.050 | | 80 | | 4.90 | | Kasogishoes | | Surabaya | | https://ta.tokop | |
| 7 | Kankly Story Horjio - Sepatu | | 298.800 | | 750 | | 5 | | Sepatu Kankly | | Kab. Bandung | | https://www.tok | |
| 8 | Sepatu Sekolah Warrior PK S | | 78.000 | | 3.000 | | 4.90 | | PASShoes | | Kab. Bandung | | https://www.tok | |
| 9 | Dr. Kevin Sepatu Sneakers Sp | | 119.900 | | 100 | | 4.90 | | Dr Kevin Shoes | | Beikasi | | https://www.tok | |
| 10 | Kankly Story Horjio - Sepatu | | 298.800 | | 3.000 | | 5 | | Sepatu Kankly | | Kab. Bandung | | https://www.tok | |
| 11 | Kankly Hiro Musashi - Sepatu | | 259.920 | | 500 | | 4.90 | | Sepatu Kankly | | Kab. Bandung | | https://www.tok | |
| 12 | Kankly Back To School Pria 35 | | 196.920 | | 250 | | 5 | | Sepatu Kankly | | Kab. Bandung | | https://www.tok | |
| 13 | Leedoo Sepatu Sneakers Pria | | 99.000 | | 3.000 | | 4.70 | | Leedoo | | Tangerang | | https://www.tok | |
| 14 | Sepatu Warrior/Warrior Spart | | 109.900 | | 1.000 | | 4.90 | | Mitra Sepatu | | Kab. Tangerang | | https://www.tok | |
| 15 | Sepatu Converse Allstar Altst | | 69.999 | | 500 | | 4.90 | | project_sneakers | | Kab. Tangerang | | https://www.tok | |
| 16 | SEPATU ANDO HITAM POLO | | 109.800 | | 100 | | 4.90 | | Toko Abadi Sho | | Tangerang | | https://www.tok | |
| 17 | Homyped Anak Sepatu Seko | | 99.900 | | 100 | | 4.90 | | Homyped Offici | | Tangerang | | https://ta.tokop | |
| 18 | sepatu sekolah Warrior tinggi | | 150.000 | | 1 | | 5 | | Ds_Storee21 | | Jakarta Timur | | https://ta.tokop | |
| 19 | TOMKINS Scool Sanditon - H | | 249.000 | | 100 | | 5 | | TOMKINS Offici | | Bandung | | https://ta.tokop | |

Gambar 2. Hasil Scraping Menggunakan BeautifulSoup

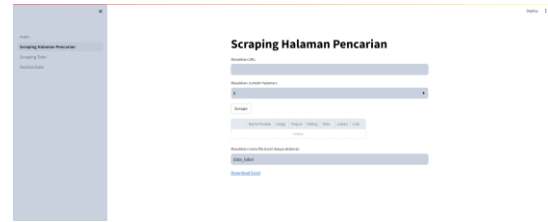
3.4 Web Scraping Menggunakan Lxml

Tahap ini adalah tahapan membuat program *web scraping* menggunakan *Lxml* Terdapat beberapa modul yang digunakan dalam pembuatan program ini yaitu *selenium*, *lxml* dan *pandas*. Hasil *url* dan identifikasi element dari tahapan sebelumnya ditambahkan pada program, *lxml* akan melakukan ekstraksi dengan parsing data pada element *HTML* yang sudah ditentukan. *Pandas* akan melakukan pengolahan data dari hasil parsing data yang dilakukan *lxml* dan akan ditampilkan dalam bentuk tabel.

| | A | Y | B | Y | C | Y | D | Y | E | Y | F | Y | G | Y |
|----|-------------------------------|---|-----------|---|--------------|---|--------|---|-------------------|---|----------------|---|-------------------|---|
| 1 | Nama Produk | | Harga | | Terjual | | Rating | | Toko | | Lokasi | | Link | |
| 2 | Kankly Story Horjio - Sepatu | | Rp298.800 | | 3rb+ terjual | | 5.0 | | Sepatu Kankly | | Kab. Bandung | | [https://ta.tokop | |
| 3 | Sepatu Converse All Star F | | Rp75.000 | | 30+ terjual | | 5.0 | | briantara_shoes | | Kab. Tangerang | | [https://ta.tokop | |
| 4 | Spotec Sepatu Student Activ | | Rp252.900 | | 80+ terjual | | 5.0 | | Spotec Official S | | Jakarta Utara | | [https://ta.tokop | |
| 5 | TOMKINS Scool Wednesday | | Rp249.000 | | 90+ terjual | | 5.0 | | TOMKINS Offici | | Bandung | | [https://ta.tokop | |
| 6 | Sepatu Casual Kasogi Julio | | Rp220.050 | | 80+ terjual | | 5 | | Kasogishoes | | Surabaya | | [https://ta.tokop | |
| 7 | Kankly Story Horjio - Sepatu | | Rp298.800 | | 750+ terjual | | 5.0 | | Sepatu Kankly | | Kab. Bandung | | [https://www.to | |
| 8 | Sepatu Sekolah Warrior Pj | | Rp78.000 | | 3rb+ terjual | | 5 | | PASShoes | | Kab. Bandung | | [https://www.to | |
| 9 | Dr. Kevin Sepatu Sneakers | | Rp119.900 | | 100+ terjual | | 5 | | Dr Kevin Shoes | | Beikasi | | [https://www.to | |
| 10 | Kankly Story Horjio - Sepatu | | Rp298.800 | | 3rb+ terjual | | 5.0 | | Sepatu Kankly | | Kab. Bandung | | [https://www.to | |
| 11 | Kankly Hiro Musashi - Sepatu | | Rp259.920 | | 500+ terjual | | 5 | | Sepatu Kankly | | Kab. Bandung | | [https://www.to | |
| 12 | Kankly Back To School Pria | | Rp196.920 | | 250+ terjual | | 5.0 | | Sepatu Kankly | | Kab. Bandung | | [https://www.to | |
| 13 | Leedoo Sepatu Sneakers Pria | | Rp99.000 | | 3rb+ terjual | | 5 | | Leedoo | | Tangerang | | [https://www.to | |
| 14 | Sepatu Warrior/Warrior Spart | | Rp109.900 | | 1rb+ terjual | | 5 | | Mitra Sepatu | | Kab. Tangerang | | [https://www.to | |
| 15 | Sepatu Converse Allstar Altst | | Rp69.999 | | 500+ terjual | | 5 | | project_sneakers | | Kab. Tangerang | | [https://www.to | |
| 16 | SEPATU ANDO HITAM POLO | | Rp109.800 | | 100+ terjual | | 5 | | Toko Abadi Sho | | Tangerang | | [https://www.to | |
| 17 | Homyped Anak Sepatu Seko | | Rp99.900 | | 100+ terjual | | 5 | | Homyped Offici | | Tangerang | | [https://ta.tokop | |
| 18 | sepatu sekolah Warrior tinggi | | Rp150.000 | | 1 terjual | | 5.0 | | Ds_Storee21 | | Jakarta Timur | | [https://ta.tokop | |
| 19 | TOMKINS Scool Sanditon - H | | Rp249.000 | | 100+ terjual | | 5.0 | | TOMKINS Offici | | Bandung | | [https://ta.tokop | |

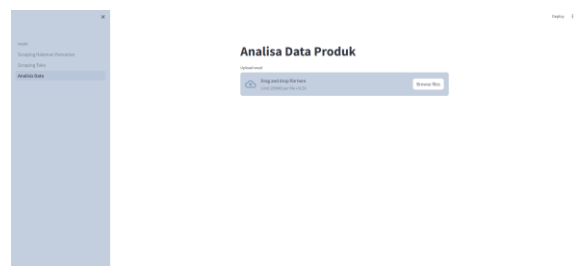
Gambar 3. Hasil Scraping Menggunakan Lxml

3.5 Hasil Percangan Website



Gambar 4. Halaman Menu Scraping Halaman Pencarian

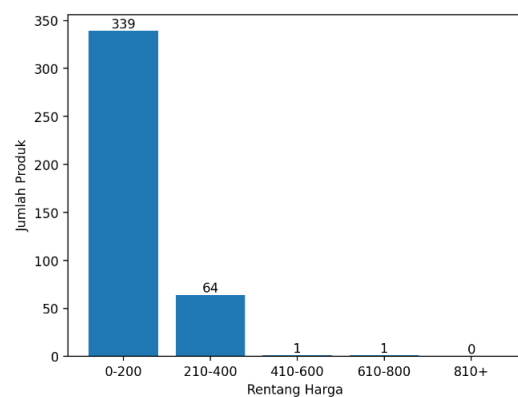
Gambar 4 merupakan halaman dari menu *scraping* halaman pencarian. Pada halaman ini *user* akan diminta memasukan alamat *url* dari halaman *website* Tokopedia yang akan di *scrape*, kemudian *user* memasukan jumlah halaman yang akan di *scrape*. Setelah itu *user* dapat menekan tombol “Scrape” dan sistem akan menjalankan proses *scraping*. *Website* akan menampilkan hasil *scraping*, lalu *user* dapat mengunduh hasil *scraping* dengan format *excel* dengan menekan tombol “Download Excel”.



Gambar 5. Halaman Menu Analisa Data Produk

Gambar 5 merupakan tampilan dari menu analisis data produk. Pada halaman ini, *user* diminta untuk mengunggahnya file hasil *scraping* yang sudah di unduh *user* sebelumnya dari menu *scraping* halaman pencarian .

Jumlah Produk Berdasarkan Rentang Harga



Gambar 6. Halaman Menu Analisa Data Produk

Gambar 6 merupakan tampilan dari menu rentang harga pada halaman analisis data produk, dapat dilihat terdapat sebuah diagram batang yang menampilkan jumlah produk berdasarkan rentang harga, dimana sistem akan membaca kolom ‘Harga’ dan akan mengkategorikan setiap nilai. Dapat dilihat pada diagram batang diatas terdapat 339 produk

dengan harga antara Rp. 0 – Rp.200.000 dan 64 produk pada rentang Rp. 210.000 – Rp.400.000.

Top 10 Data Rating Tertinggi

| | Nama Produk | Harga | Terjual | Rating | Toko |
|----|---|---------|---------|--------|-----------------------|
| 0 | Kanky Story Honjo - Sepatu Sneakers Casual Sport Seki | 277,884 | 750 | 5 | Sepatu Kanky |
| 1 | Sepatu Converse All Star Petir Tinggi Sepatu Sekolah Lc | 75,000 | 30 | 5 | brilantara_shoes |
| 2 | Kanky Story Honjo - Sepatu Sneakers Casual Sport Seki | 298,800 | 3,000 | 5 | Sepatu Kanky |
| 7 | Kanky Story Honjo - Sepatu Sneakers Casual Sport Seki | 277,884 | 750 | 5 | Sepatu Kanky |
| 10 | Kanky Story Honjo - Sepatu Sneakers Casual Sport Seki | 298,800 | 3,000 | 5 | Sepatu Kanky |
| 14 | Kanky Back To School Pria 39-44 - Sepatu Sneakers Cas | 196,920 | 250 | 5 | Sepatu Kanky |
| 15 | sepatu sekolah Warrior tinggi sparta-hc Original | 150,000 | 1 | 5 | Ds_Store21 |
| 16 | TOMKINS Scool Sanditon - Hitam Sepatu Sekolah | 249,000 | 100 | 5 | TOMKINS Official Shop |
| 17 | sepatu sekolah Warrior pendek hitam polos original | 145,000 | 1 | 5 | Ds_Store21 |
| 21 | TOMKINS Scool Sanditon - Hitam Sepatu Sekolah | 249,000 | 100 | 5 | TOMKINS Official Shop |

Top 10 Rating Terendah

| | Nama Produk | Harga | Terjual | Rating | Toko |
|-----|--|---------|---------|--------|--------------|
| 19 | sepatu sekolah Warrior tinggi Ventela | 235,000 | 0 | 0 | Ds_Store21 |
| 91 | MJC Indonesia 37-45 Sepatu Sekolah hitam Pria Wanita Fabr | 199,000 | 0 | 0 | dropstore77 |
| 150 | SEPATU SNEAKERS CASUAL KERJA KANTOR PROYEK KULIAH | 129,999 | 0 | 0 | Mamoru_Store |
| 155 | Sepatu Sekolah Pantofel Kanvas Kets Sneaker Pria Kerja Len | 399,000 | 0 | 0 | SUPER KIDZ88 |
| 185 | sepatu sekolah pendek merk DALLAS campus original | 130,000 | 1 | 0 | jessy shoes |
| 225 | sepatu kets/sekolah merk DUFF original | 220,000 | 2 | 0 | Ezalea Shoes |
| 227 | Sepatu ADIO VICEL ORIGINAL casual sekolah | 185,000 | 0 | 0 | Antarez Shop |
| 228 | Sepatu Compass Original Gazelle Vintage Compas Sepatu Se | 373,000 | 0 | 0 | thara jaya |
| 231 | SAL15 - Sepatu EAGLE AMARIS Size 38-45 - Sekolah Laki-laki | 220,000 | 0 | 0 | SarahAstuti |
| 249 | Sepatu sekolah bts STRADE NINE ORIGINAL MICHIGAN ALL B | 169,900 | 0 | 0 | lcconnex |

Gambar 7. Tampilan Top 10 Rating Tertinggi dan Terendah

Gambar 7 merupakan halaman analisis data yang menampilkan menu Top rating. Sistem menampilkan dua buah tabel, tabel pertama adalah tabel rating tertinggi yang menampilkan urutan data berdasarkan rating tertinggi. Sistem membaca kolom 'Rating' dan mengurutkan nilai dari yang terbesar ke terkecil. Sedangkan tabel kedua adalah tabel rating terendah yang menampilkan urutan data berdasarkan rating terendah. Sistem membaca kolom 'Rating' dan mengurutkan nilai dari yang terkecil ke terbesar.

Top 10 Harga Terendah

| | Nama Produk | Harga | Terjual | Rating | Toko |
|-----|--|--------|---------|--------|------------------|
| 309 | Sepatu Vans sepatu casual sepatu sekolah termurah be | 9,150 | 50 | 4.7 | dunia sneakers21 |
| 122 | Sepatu Converse sekolah Sneakers high casual termura | 9,270 | 30 | 4.9 | dunia sneakers21 |
| 226 | sepatu vans hitam pria cowok cowo olahraga sekolah s | 24,900 | 50 | 5 | Toko49 SHOES |
| 11 | SEPATU PRIA COWOK COWO SEKOLAH TOKO49 SHOES | 28,500 | 2,000 | 4.6 | Toko49 SHOES |
| 210 | Sepatu ApStar AP Star Sepatu Sekolah Anti Air Tahan Ai | 28,888 | 21 | 5 | wonzila |
| 138 | ATT ABS05 Sepatu Pria Pantofel Karet Sekolah Kerja Hit | 30,000 | 500 | 4.9 | Toko Buka |
| 153 | Sepatu Karet Pria Hitam / Sepatu Sneakers Cowok untu | 31,500 | 70 | 4.9 | Floppyshopy |
| 101 | sepatu sneakers olahraga pria cowok cowo hitam sekol | 32,300 | 250 | 4.7 | Toko49 SHOES |
| 13 | 6RR sepatu sekolah pria cowok cowo wanita cewek cev | 34,500 | 250 | 4.8 | Toko49 SHOES |
| 334 | SEPATU CONVERSE WHITE LOW PRIA DAN WANITA COCI | 35,500 | 60 | 4.6 | Nauli sneakers |

Top 10 Harga Tertinggi

| | Nama Produk | Harga | Terjual | Rating | Toko |
|-----|--|---------|---------|--------|-------------------|
| 392 | SEPATU PRIA ORI New Balance 500 v2 SNEAKERS SEKOLAH # | 675,000 | 0 | 0 | okshoes_store |
| 390 | Big Sale Hcd10 Promo Sepatu Puma Olahraga Hangout Sepa | 405,000 | 0 | 0 | HRC Storc |
| 298 | PATROBAS X ONIC BLACK - LIMITED EDITION - SEPATU GAME | 400,000 | 0 | 0 | shop&shop |
| 155 | Sepatu Sekolah Pantofel Kanvas Kets Sneaker Pria Kerja Len | 399,000 | 0 | 0 | SUPER KIDZ88 |
| 287 | Sepatu Sekolah Mills Revolt Beta Black / Black 9700901 | 399,000 | 0 | 0 | victoria sport dk |
| 277 | MUTI177- Kanky Hiro Musashi - Sepatu Sneakers Casual Spor | 385,000 | 0 | 0 | MutiYara |
| 228 | Sepatu Compass Original Gazelle Vintage Compas Sepatu Se | 373,000 | 0 | 0 | thara jaya |
| 261 | Sepatu Sekolah Mills Evander Black / Black 9700801 | 349,000 | 1 | 5 | victoria sport dk |
| 269 | Sepatu Dewasa Pria Kasogi Rajut Sneakers Casual putih seki | 321,237 | 6 | 4.8 | Kasogishoes |
| 377 | Sepatu Sneakers pria full black SPOTEC Versa /Sepatu sekol | 310,000 | 1 | 5 | 1000 SHOES |

Gambar 8. Tampilan Top 10 Harga Tertinggi dan Terendah

Gambar 8 merupakan halaman analisis data yang menampilkan menu Top Harga. Sistem menampilkan dua buah tabel, tabel pertama adalah tabel harga terendah yang menampilkan urutan data berdasarkan harga terendah. Sistem membaca kolom 'Harga' dan mengurutkan nilai dari yang terkecil ke terbesar. Sedangkan tabel kedua adalah tabel harga tertinggi yang menampilkan urutan data berdasarkan harga tertinggi. Sistem membaca kolom 'Harga' dan mengurutkan nilai dari yang terbesar ke terkecil.

Top 10 Penjualan Tertinggi

| | Nama Produk | Harga | Terjual | Rating | Toko |
|-----|--|---------|---------|--------|----------------------|
| 46 | Leedoo Sepatu Sneakers Pria Sepatu Olahraga Lari Ci | 99,000 | 5,000 | 4.6 | Leedoo |
| 2 | Kanky Story Honjo - Sepatu Sneakers Casual Sport Se | 298,800 | 3,000 | 5 | Sepatu Kanky |
| 8 | Sepatu Sekolah Warrior PK Star (Menerima Pesanan S | 78,000 | 3,000 | 4.9 | PASShoes |
| 10 | Kanky Story Honjo - Sepatu Sneakers Casual Sport Se | 298,800 | 3,000 | 5 | Sepatu Kanky |
| 20 | Leedoo Sepatu Sneakers Pria Sepatu Sekolah Kasual | 99,000 | 3,000 | 4.7 | Leedoo |
| 82 | Leedoo Sepatu Sneakers Pria Sepatu Sekolah Kasual | 99,000 | 3,000 | 4.7 | Leedoo |
| 11 | SEPATU PRIA COWOK COWO SEKOLAH TOKO49 SHOE | 28,500 | 2,000 | 4.6 | Toko49 SHOES |
| 81 | SEPATU VANS OLD SKOOL SCHOOL SNEAKERS SEKOL | 104,900 | 2,000 | 4.8 | rajalelaki |
| 93 | SEPATU VANS OLD SKOOL SCHOOL SNEAKERS SEKOL | 104,900 | 2,000 | 4.8 | rajalelaki |
| 113 | Sepatu Joemen J 49 Original Import Pria Wanita Kerji | 120,000 | 2,000 | 4.7 | JOEMEN OFFICIAL SHOP |

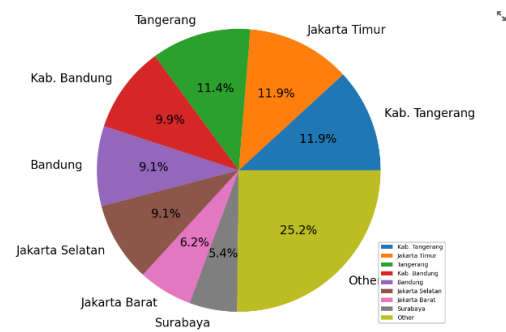
Top 10 Penjualan Terendah

| | Nama Produk | Harga | Terjual | Rating | Toko | Lokasi |
|-----|---|---------|---------|--------|--------------|----------|
| 19 | sepatu sekolah Warrior tinggi Ventela | 235,000 | 0 | 0 | Ds_Store21 | Jakarta |
| 91 | MJC Indonesia 37-45 Sepatu Sekolah hitam Pria Wani | 199,000 | 0 | 0 | dropstore77 | Jakarta |
| 150 | SEPATU SNEAKERS CASUAL KERJA KANTOR PROYEK | 129,999 | 0 | 0 | Mamoru_Store | Malang |
| 155 | Sepatu Sekolah Pantofel Kanvas Kets Sneaker Pria Ke | 399,000 | 0 | 0 | SUPER KIDZ88 | Jakarta |
| 227 | Sepatu ADIO VICEL ORIGINAL casual sekolah | 185,000 | 0 | 0 | Antarez Shop | Kab. Tan |
| 228 | Sepatu Compass Original Gazelle Vintage Compas Se | 373,000 | 0 | 0 | thara jaya | Jakarta |
| 231 | SAL15 - Sepatu EAGLE AMARIS Size 38-45 - Sekolah La | 220,000 | 0 | 0 | SarahAstuti | Surabaya |
| 249 | Sepatu sekolah bts STRADE NINE ORIGINAL MICHIGA | 169,900 | 0 | 0 | lcconnex | Bandung |
| 272 | NewRaven Sepatu Pria Sneakers Sepatu Sekolah Sep | 120,000 | 0 | 0 | ary sport | Malang |
| 277 | MUTI177- Kanky Hiro Musashi - Sepatu Sneakers Casu | 385,000 | 0 | 0 | MutiYara | Jakarta |

Gambar 9. Tampilan Top 10 Penjualan Tertinggi dan Terendah

Gambar 9 merupakan halaman analisis data yang menampilkan menu Top Terjual. Sistem menampilkan dua buah tabel, tabel pertama adalah tabel penjualan tertinggi yang menampilkan urutan data berdasarkan penjualan tertinggi. Sistem membaca kolom 'Terjual' dan mengurutkan nilai dari yang terbesar ke terkecil. Sedangkan tabel kedua adalah tabel penjualan terendah yang menampilkan urutan data berdasarkan penjualan terendah. Sistem membaca kolom 'Terjual' dan mengurutkan nilai dari yang terkecil ke terbesar.

Persentasi Lokasi Penjual



Data Lokasi:

| Lokasi | Frekuensi |
|-------------------|-----------|
| 0 Kab. Tangerang | 48 |
| 1 Jakarta Timur | 48 |
| 2 Tangerang | 46 |
| 3 Kab. Bandung | 40 |
| 4 Bandung | 37 |
| 5 Jakarta Selatan | 37 |
| 6 Jakarta Barat | 25 |
| 7 Surabaya | 22 |
| 8 Kab. Mojokerto | 14 |
| 9 Kab. Bekasi | 13 |

Gambar 10. Tampilan Diagram Persentasi Lokasi Penjual

Gambar 10 merupakan halaman analisis data yang menampilkan menu Persentase Lokasi. Sistem akan membaca kolom 'Lokasi' dan mengidentifikasi nilai yang sama dan menggabungkannya, lalu dihitung persentasinya dari gabungan nilai yang sama dan ditampilkan dalam bentuk diagram lingkaran dimana jika ada dua atau lebih nilai bernilai dibawah 5% maka akan digabungkan pada bagian 'other'. Selain itu terdapat tabel yang menampilkan urutan nilai dengan frekuensi terbanyak.

Gambar 11 merupakan halaman analisis data yang menampilkan menu Persentase Toko. Sistem akan membaca kolom 'Toko' dan mengidentifikasi nilai yang sama dan menggabungkannya, lalu dihitung persentasinya dari gabungan nilai yang sama dan di tampilkan dalam bentuk diagram lingkaran dimana jika ada dua atau lebih nilai bernilai dibawah 5% maka akan digabungkan pada bagian 'other'. Selain itu terdapat tabel yang menampilkan urutan nilai dengan frekuensi terbanyak.

Rekomendasi Produk

Rekomendasi Produk Menggunakan Rumus Normalisasi Min-Max

| Nama Produk | Rekomendasi | Harga Norm | Terjual Norm | Rating Norm |
|--|-------------|------------|--------------|-------------|
| 46 Leedoo Sepatu Sneakers Pria Sepatu Olahraga L | 2.7851 | 0.8651 | 1 | 0.92 |
| 8 Sepatu Sekolah Warrior PX Star (Menerima Pesa | 2.4766 | 0.8966 | 0.6 | 0.98 |
| 20 Leedoo Sepatu Sneakers Pria Sepatu Sekolah Ki | 2.4051 | 0.8651 | 0.6 | 0.94 |
| 82 Leedoo Sepatu Sneakers Pria Sepatu Sekolah Ki | 2.4051 | 0.8651 | 0.6 | 0.94 |
| 11 SEPATU PRIA COWOK COWO SEKOLAH TOKO49 | 2.2909 | 0.9709 | 0.4 | 0.92 |
| 116 Leedoo Sepatu Sneakers Pria Sepatu Olahraga L | 2.2501 | 0.9101 | 0.4 | 0.94 |
| 93 SEPATU VANS OLD SKOOL SCHOOL SNEAKERS S | 2.2162 | 0.8562 | 0.4 | 0.96 |
| 81 SEPATU VANS OLD SKOOL SCHOOL SNEAKERS S | 2.2162 | 0.8562 | 0.4 | 0.96 |
| 218 Sepatu Joemen J 49 Original Import Pria Wanita | 2.1735 | 0.8335 | 0.4 | 0.94 |
| 113 Sepatu Joemen J 49 Original Import Pria Wanita | 2.1735 | 0.8335 | 0.4 | 0.94 |

harga max = 675000

harga min = 9150

Terjual max = 5000

terjual min = 9150

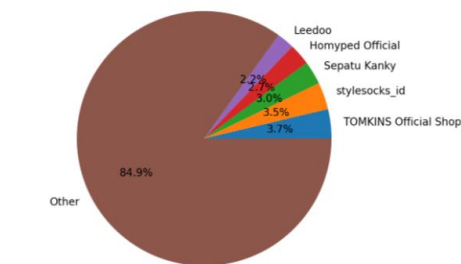
rating max = 5.0

rating min = 0.6

Gambar 12. Tampilan Rekomendasi Produk

Gambar 12 merupakan halaman analisis data yang menampilkan menu Rekomendasi Produk. Menu rekomendasi produk berfokus pada 3 kolom yaitu kolom Harga, Terjual dan Rating, kemudian dilakukan perhitungan *Min-max normalization* pada ketiga kolom tersebut. Hasil normalisasi dari kolom Harga, Terjual dan Rating dijumlahkan kemudian hasilnya ditampilkan dalam bentuk tabel dan diurutkan nilai rekomendasi dari yang terbesar ke terkecil.

Persentasi Toko



Data Toko:

| Toko | Frekuensi |
|-------------------------|-----------|
| 0 TOMKINS Official Shop | 15 |
| 1 stylesocks_id | 14 |
| 2 Sepatu Kanky | 12 |
| 3 Homyped Official | 11 |
| 4 Leedoo | 9 |
| 5 Juraganlocal | 8 |
| 6 johnsport | 8 |
| 7 Toko Abadi Shoes | 7 |
| 8 bilantara_shoes | 6 |
| 9 Ezalea Shoes | 6 |

Gambar 11. Tampilan Diagram Persentasi Toko Penjual

4. KESIMPULAN DAN SARAN

Analisis data produk dari hasil web scraping pada halaman website Tokopedia yang dihasilkan untuk user penjual dengan kategori yang ditentukan antara lain: kategori data berdasarkan rentang harga, pengurutan data dengan rating tertinggi dan terendah, pengurutan data dengan harga tertinggi dan terendah, pengurutan data dengan penjualan tertinggi dan terendah, persentasi lokasi penjual, persentasi toko penjual. Analisis data produk berupa rekomendasi produk dapat digunakan *user* pembeli dalam memilih produk dengan harga, penjualan dan rating terbaik.

REFERENSI

- [1] community@algorit.ma. (2022, April 13). Cara Cepat Copy Paste Data Python ke Excel. <https://Algorit.Ma/Blog/Copy-Paste-Data-Python-Ke-Excel/#:~:Text=Pertama%2Dtama%2C%20pastikan%20Anda%20sudah,Ini%20bekerja%20dengan%20dokumen%20Excel>.
- [2] community@algorit.ma. (2022, April 22). MENGENAL PYTHON DAN KEGUNAANNYA. <https://Algorit.Ma/Blog/Data-Science/Apa-Itu-Python-2022/>.
- [3] Dede Brahma. (2020, February 23). Scraping Web dengan Python. <https://Medium.Com/@dede.Brahma2/Scraping-Web-Dengan-Python-F686d3716028>.
- [4] dqlab@umn.ac.id. (2022, March 22). Tutorial Membuat Portofolio Data dari Python Streamlit. <https://Dqlab.Id/Tutorial-Membuat-Portofolio-Data-Dari-Python-Streamlit>.
- [5] Gendis Widodari. (2020). Mengambil Data dengan BeautifulSoup. <https://Learn.Nural.Id/Course/Data-Science/Web-Scraping-Beautifulsoup/Index>.
- [6] Hermawan Riyadi. (2021, June 11). Apa itu Shopee? Keunggulan Apa Saja yang Dimiliki oleh Shopee? <https://Www.Nesabamedia.Com/Apa-Itu-Shopee/>.
- [7] Marcellinus, S. (2021). Ekstraksi Fitur Peneliti Menggunakan Numpy dan Pandas.
- [8] Maulana, A. A., Susanto, A., & Kusumaningrum, D. P. (2019). Rancang Bangun Web Scraping Pada Marketplace di Indonesia. *JOINS (Journal of Information System)*, 4(1), 41–53. <https://doi.org/10.33633/joins.v4i1.2544>
- [9] Nurul Huda. (2022, November 22). Visual Studio Code: Pengertian, Fitur, Keunggulan dan Jenisnya. <https://Www.Dewaweb.Com/Blog/Mengenal-Visual-Studio-Code/>.
- [10] Priadana, A., & Murdiyanto, A. W. (2020). Pemantauan tren hashtag pada instagram menggunakan web scraping instagram hashtag trend monitoring using web scraping. *Jurnal Pekommas*, 5(1), 23–30.
- [11] Rony Setiawan. (2021, November 17). Black Box Testing Untuk Menguji Perangkat Lunak. <https://Www.Dicoding.Com/Blog/Black-Box-Testing/>.
- [12] Sahria, Y. (2020). Implementasi Teknik Web Scraping pada jurnal SINTA untuk analisis topik penelitian kesehatan Indonesia. *Prosiding University Research Colloquium*, 297–306.
- [13] Satriajati, S., Panuntun, S. B., & Pramana, S. (2020). Implementasi web scraping dalam pengumpulan berita kriminal pada masa pandemi COVID-19. *Seminar Nasional Official Statistics*, 2020(1), 300–308.
- [14] Trivusi, (2022, September 16). Normalisasi Data: Pengertian, Tujuan, dan Metodanya. [https://www.trivusi.web.id/2022/09/normalisasi-data.html#:~:text=2.%20Normalisasi%20Min%20Dmax,Mana%20yang%20lebih&text=Metode%20normalisasi%20min%20Dmax%20mengubah,%20hingga%201%20\(max\)](https://www.trivusi.web.id/2022/09/normalisasi-data.html#:~:text=2.%20Normalisasi%20Min%20Dmax,Mana%20yang%20lebih&text=Metode%20normalisasi%20min%20Dmax%20mengubah,%20hingga%201%20(max)).
- [15] Wibowo, F. R., Rusdianto, D. S., & Arwan, A. (2019). Pengembangan Sistem Pengumpulan Promo E-Commerce Berbasis Website Dengan Menerapkan Teknik Web Scraping Dalam Proses Pengambilan Data Promo. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer E-ISSN*, 2548, 964X.
- [16] Yani, D. D. A., Pratiwi, H. S., & Muhardi, H. (2019). Implementasi web scraping untuk pengambilan data pada situs marketplace. *JUSTIN (Jurnal Sistem Dan Teknologi Informasi)*, 7(4), 257–262.