

PENERAPAN METODE *USER-BASED COLLABORATIVE FILTERING* UNTUK REKOMENDASI PRODUK *CLOTHING* PADA PLATFORM *E-COMMERCE SHOPEE*

Jelita Arba Sholeha^{1*}, Zunita Wulansari², Ukhianti Mawaddah³

^{1,2,3}Universitas Balitar, Blitar, Indonesia

*Corresponding author: jejelita29@gmail.com

Abstract: This study aims to implement the User-Based Collaborative Filtering (UBCF) method in a clothing product recommendation system based on e-commerce data. Web scraping techniques using the Tamper Monkey extension (JavaScript) are applied to extract user data from the Shopee platform, including usernames, ratings, review times, and comment descriptions. The extracted data are stored in CSV format and processed using the Python programming language with the Pandas, NumPy, Selenium, and BeautifulSoup libraries. Similarity analysis between users is performed using the Euclidean Distance method to determine users with similar preferences. The predicted rating results are calculated using the Weighted Sum Prediction method and tested with the Mean Absolute Error (MAE) metric to assess the level of accuracy. Based on the test results, the UBCF-based recommendation system is able to provide product recommendations with a good level of accuracy and high relevance to user preferences. Thus, the application of UBCF to e-commerce data is proven to improve users' online shopping experience and make a real contribution to the development of a more personalized and adaptive recommendation system.

Keywords: Collaborative Filtering, Web Scraping, Python, Shopee

Abstrak: Penelitian ini bertujuan untuk mengimplementasikan metode User-Based Collaborative Filtering (UBCF) dalam sistem rekomendasi produk clothing berbasis data e-commerce. Teknik web scraping menggunakan ekstensi Tamper Monkey (JavaScript) diterapkan untuk mengekstraksi data pengguna dari platform Shopee, termasuk nama pengguna, rating, waktu ulasan, serta deskripsi komentar. Data hasil ekstraksi disimpan dalam format CSV dan diproses menggunakan bahasa pemrograman Python dengan pustaka Pandas, NumPy, Selenium, dan BeautifulSoup. Analisis kemiripan antar pengguna dilakukan menggunakan metode Euclidean Distance untuk menentukan pengguna dengan preferensi serupa. Hasil prediksi rating dihitung menggunakan metode Weighted Sum Prediction dan diuji dengan metrik Mean Absolute Error (MAE) untuk menilai tingkat akurasi. Berdasarkan hasil pengujian, sistem rekomendasi berbasis UBCF mampu memberikan rekomendasi produk dengan tingkat akurasi yang baik dan relevansi tinggi terhadap preferensi pengguna. Dengan demikian, penerapan UBCF pada data e-commerce terbukti dapat meningkatkan pengalaman belanja daring pengguna dan memberikan kontribusi nyata bagi pengembangan sistem rekomendasi yang lebih personal dan adaptif.

Kata kunci: Penyaringan Kolaboratif, Pengikisan Web, Python, Shopee

Copyright (c) 2025 The Authors. This is an open-access article under the CC BY-SA 4.0 license (<https://creativecommons.org/licenses/by-sa/4.0/>)

PENDAHULUAN

Perkembangan teknologi informasi dan ilmu pengetahuan yang pesat telah mendorong perubahan besar dalam dunia bisnis, termasuk dalam sektor perdagangan

daring (Ahadiyah, 2023). Munculnya berbagai situs belanja online menyebabkan banyak pelaku usaha beralih menjual produknya secara digital (Aamer et al., 2025). E-commerce menjadi wadah transaksi jual-beli antara penjual dan pelanggan tanpa pertemuan langsung (Saprida et al., 2023). Namun, dengan semakin banyaknya variasi produk dan merek *clothing*, konsumen sering kali mengalami kebingungan dalam menentukan pilihan (Li, 2024). Setiap konsumen memiliki preferensi dan minat yang berbeda, sehingga dibutuhkan sistem rekomendasi yang dapat membantu mereka menemukan produk sesuai kebutuhan dan selera (Chang et al., 2019).

Sistem rekomendasi menjadi solusi penting untuk membantu pelanggan memilih produk yang relevan. Salah satu metode yang banyak digunakan dalam pengembangan sistem rekomendasi adalah *Collaborative Filtering* (Jaja et al., 2020), yakni metode yang memanfaatkan data perilaku pengguna seperti riwayat pembelian dan rating produk untuk memberikan saran yang bersifat personal. Penelitian yang dilakukan oleh (Pradana & Noorrizki, 2022) menunjukkan bahwa metode *Collaborative Filtering* menghasilkan tingkat akurasi yang baik dengan nilai MAE terbaik sebesar 0,23 pada satu pengguna dan 0,25 pada lima pengguna. Sementara itu, penelitian (Halim et al., 2022) memperoleh hasil sebesar 86% responden menyatakan sistem ini berfungsi dengan baik. Adapun penelitian Alma dkk. (2020) menggunakan metode *Apriori* dengan nilai minimum support 40% dan minimum *confidence* 70%, yang menghasilkan rekomendasi produk seperti hubungan antara pembelian Khimar dan Gamis dengan nilai *confidence* 78,57%. Dari perbandingan tersebut, metode *Collaborative Filtering* terbukti lebih unggul (Jin et al., 2024).

Metode *Collaborative Filtering* bekerja dengan menganalisis data historis pelanggan, seperti rating produk dalam skala 0–5, untuk memprediksi preferensi terhadap produk lain ((Pradana & Noorrizki, 2022). Dalam penelitian ini, bahasa pemrograman *Python* digunakan untuk mengembangkan algoritma *Collaborative Filtering* karena kemampuannya dalam pengolahan data, dukungan pustaka yang lengkap, serta fleksibilitas integrasi dengan berbagai format data (Maesaroh et al., 2024). Data pelanggan disimpan dalam format file .CSV yang mudah diakses dan diolah. Hasil penelitian memang menawarkan fleksibilitas dan dukungan pustaka yang luas, namun sebagian besar riset hanya menekankan pada kemampuan teknis *Python* tanpa menggali bagaimana integrasi data khususnya data berbasis file .CSV dapat dioptimalkan untuk

meningkatkan akurasi rekomendasi (Roy et al., 2024).

Selain itu, banyak studi terdahulu belum membahas secara mendalam tantangan seperti variasi kualitas data dalam file .CSV, sparsity data rating, serta bagaimana penyesuaian parameter algoritma dapat memengaruhi performa Collaborative Filtering dalam konteks data pelanggan yang nyata (Rahman et al., 2024). Dengan demikian, masih terdapat kesenjangan penelitian dalam hal pengembangan model Collaborative Filtering yang lebih adaptif, terintegrasi, dan mampu mengatasi karakteristik data pelanggan yang heterogen, terutama ketika diimplementasikan menggunakan Python dengan sumber data .CSV.

Penelitian ini bertujuan menguji efektivitas penerapan metode *Collaborative Filtering* dalam memberikan rekomendasi produk *clothing* (Han, 2023). Evaluasi dilakukan melalui metrik seperti akurasi, presisi, dan recall (Xia, 2016). Hasil penelitian diharapkan dapat memberikan kontribusi praktis bagi pelaku bisnis fashion daring dalam meningkatkan pengalaman berbelanja pelanggan serta mendorong peningkatan penjualan (Sri, 2025).

METODE

Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan mulai Desember 2024 hingga Juli 2025, meliputi tahap perencanaan, pengumpulan data, pengembangan sistem rekomendasi, hingga evaluasi akhir. Studi literatur serta perancangan sistem dilakukan pada Desember 2024–Januari 2025 dengan fokus pada identifikasi kebutuhan, analisis metode *Collaborative Filtering*, serta persiapan perangkat keras dan lunak. Pengumpulan data melalui web scraping di platform Shopee berlangsung Februari–Maret 2025 untuk memperoleh data pengguna dan produk pakaian sebagai dasar model rekomendasi.

Tahap pengembangan sistem dilakukan April–Mei 2025 dengan menggunakan Python dan pustaka seperti Pandas, NumPy, Selenium, serta BeautifulSoup. Algoritma User-Based Collaborative Filtering diterapkan untuk menghitung kemiripan pengguna dan memprediksi rating produk. Evaluasi sistem dilaksanakan Juni 2025 dengan menggunakan metrik *Mean Absolute Error* (MAE) dan *Root Mean Square Error* (RMSE) untuk menilai akurasi serta relevansi rekomendasi. Penelitian ini berlokasi di Laboratorium Komputer Fakultas Teknologi Informasi, Universitas Islam Balitar, Blitar, yang dilengkapi fasilitas pendukung seperti komputer, jaringan internet, dan akses

sumber daya digital. Beberapa tahap, termasuk pengujian sistem, juga dilakukan secara daring guna menjaga fleksibilitas penelitian.

Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk menganalisis data numerik hasil scraping dan mengevaluasi kinerja sistem. Algoritma *User-Based Collaborative Filtering* digunakan untuk mengidentifikasi pola preferensi pengguna berdasarkan rating produk. Evaluasi sistem dilakukan dengan menggunakan metrik MAE untuk mengukur tingkat kesalahan prediksi terhadap data aktual, sebagaimana disarankan oleh (Zhang et al., 2014) dan (Herlocker et al., 2004).

Teknik Pengumpulan Data

Pengumpulan data dilakukan melalui *web scraping* pada Shopee, meliputi data pengguna, detail produk, rating, dan ulasan. Proses ini melibatkan empat tahap utama: sumber data, perangkat pengembangan, scraping, dan penyimpanan data. Shopee dipilih sebagai sumber karena menyediakan data publik berupa rating, ulasan, serta identitas pengguna yang relevan untuk membangun profil pelanggan (Nizamuddin et al., 2021). Data yang dikumpulkan bersifat kuantitatif (angka rating) dan kualitatif (teks ulasan).

Tahap pengumpulan dan pengolahan data dilakukan menggunakan bahasa pemrograman Python dengan pustaka Selenium, BeautifulSoup, Pandas, dan NumPy. Data hasil ekstraksi disimpan dalam format CSV untuk memudahkan analisis dan pra-pengolahan. Perangkat keras yang digunakan memiliki spesifikasi cukup untuk menangani komputasi intensif dan pemrosesan data dalam jumlah besar.

Scraping dan Penyimpanan Data

Proses web scraping menggunakan Selenium untuk mengakses halaman dinamis dan BeautifulSoup untuk mengekstrak elemen HTML seperti nama pengguna, rating, waktu ulasan, dan isi komentar. Sebelum dilakukan secara menyeluruh, skrip diuji menggunakan sampel kecil agar akurasi terjamin. Data yang diperoleh kemudian disimpan dalam format CSV yang berisi lima variabel utama: avatar pengguna, nama pengguna, rating, waktu ulasan, dan deskripsi komentar.

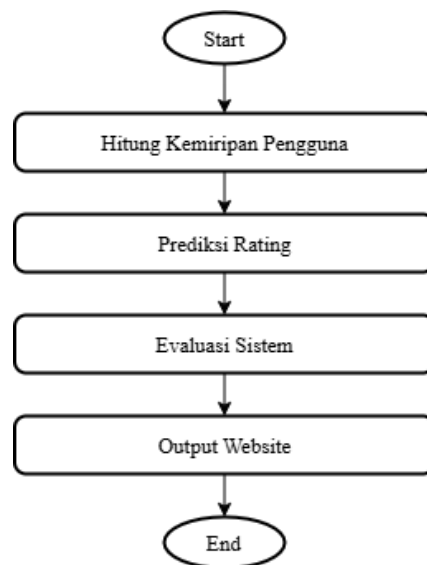
Jenis dan Instrumen Data

Jenis data yang digunakan adalah data sekunder dari Shopee yang terdiri atas lima variabel utama: avatar, author name, rating, waktu ulasan, dan deskripsi ulasan. Instrumen pengumpulan data dilakukan dengan memanfaatkan sumber ilmiah seperti

Google Books, Google Scholar, dan ResearchGate untuk memperkuat landasan teori (Nizamuddin et al., 2021).

Tahap-Tahap Penelitian

Penelitian dilakukan secara sistematis melalui tahapan: perencanaan, pengumpulan data, pra-pengolahan, penerapan algoritma, dan evaluasi sistem. Pada tahap perencanaan, dirumuskan tujuan dan batasan penelitian serta dilakukan studi literatur untuk memahami Collaborative Filtering dan metrik evaluasi.



Gambar 1. Flowchart Metode *Collaborative Filtering* di sini

Pra-pengolahan data mencakup pembersihan duplikasi, normalisasi rating, dan penanganan data kosong untuk meningkatkan kualitas data. Tahap implementasi dilakukan dengan menghitung kemiripan antar pengguna menggunakan Euclidean Distance $d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$, kemudian memprediksi rating dengan metode Weighted Sum $\hat{R}_{Xj} = \frac{\sum (similarity(X, i) \times R_{i,j})}{\sum |similarity(X, i)|}$.

Tabel 1. Data Pengguna

Pengguna	Jaket (Rating)	Dress (Rating)
Joseph Walker	4	3
Alice Smith	5	4

Tabel 2. Rating Tabel

Pengguna	Jaket	Dress	Kemeja	Rok	Celana
Joseph Walker	4	3	?	?	?
Alice Smith	?	5	4	?	?
Samuel Johson	?	?	3	2	?
Emily Davis	?	?	?	4	?
James Lopes	?	?	?	?	2

Nilai kemiripan dikonversi menjadi skor similarity menggunakan persamaan $Similarity(A, B) = \frac{1}{1+d(A, B)}$, sedangkan hasil prediksi diuji menggunakan Mean Absolute Error (MAE) $MAE = \frac{1}{N} \sum |r_i - \hat{r}^i|$ untuk mengukur ketepatan sistem rekomendasi.

Implementasi Metode Collaborative Filtering

Metode Collaborative Filtering diimplementasikan dalam bentuk sistem berbasis web yang menampilkan hasil rekomendasi produk clothing. Tahapan mencakup pengambilan data pengguna, perhitungan similarity, pembuatan rekomendasi, hingga penilaian ulang model berdasarkan hasil evaluasi sistem. Sistem ini diharapkan mampu meningkatkan pengalaman belanja pengguna melalui rekomendasi yang relevan dan personal.

HASIL DAN PEMBAHASAN

Hasil

Proses Pengumpulan Data

Proses pengumpulan data pada penelitian ini dilakukan dengan menerapkan teknik web scraping menggunakan *extension Tamper Monkey* yang berbasis bahasa pemrograman JavaScript. Teknik ini digunakan karena kemampuannya dalam mengekstraksi data secara otomatis dari situs web tanpa perlu melakukan input manual. Dalam konteks penelitian ini, *web scraping* diarahkan untuk mengumpulkan data dari situs e-commerce Shopee, yang merupakan salah satu platform belanja daring dengan jumlah pengguna aktif tinggi di Indonesia. Data yang dikumpulkan meliputi avatar pengguna, nama pengguna (*username*), nilai rating terhadap produk, waktu ulasan diberikan, serta deskripsi komentar yang berisi opini pengguna terhadap produk yang telah dibeli. Proses ini dilakukan secara berulang pada beberapa halaman produk untuk memastikan keberagaman data yang diperoleh, sehingga hasil analisis dapat mewakili berbagai preferensi konsumen.

```
function scrapeProducts() {
    return Array.from(document.querySelectorAll('.shop-search-result-view_item')).map(item => ({
        name: item.querySelector('.line-clamp-2.break-words.text-sm')?.innerText.trim() || 'N/A',
        price: item.querySelector('.text-shopee-primary:not(.bg-shopee-pink)')?.innerText.trim() || 'N/A',
        discount: item.querySelector('.text-shopee-primary.bg-shopee-pink')?.innerText.trim() || 'No Discount',
        rating: item.querySelector('div.text-shopee-black87[class*="text-xs/spl4"]')?.innerText.trim() || 'No Rating',
        sold: item.querySelector('.truncate.text-shopee-black87')?.innerText.trim() || 'No Data',
        link: item.querySelector('a')?.href || 'N/A',
        imgSrc: item.querySelector('img.object-contain')?.src || 'N/A'
    }));
}

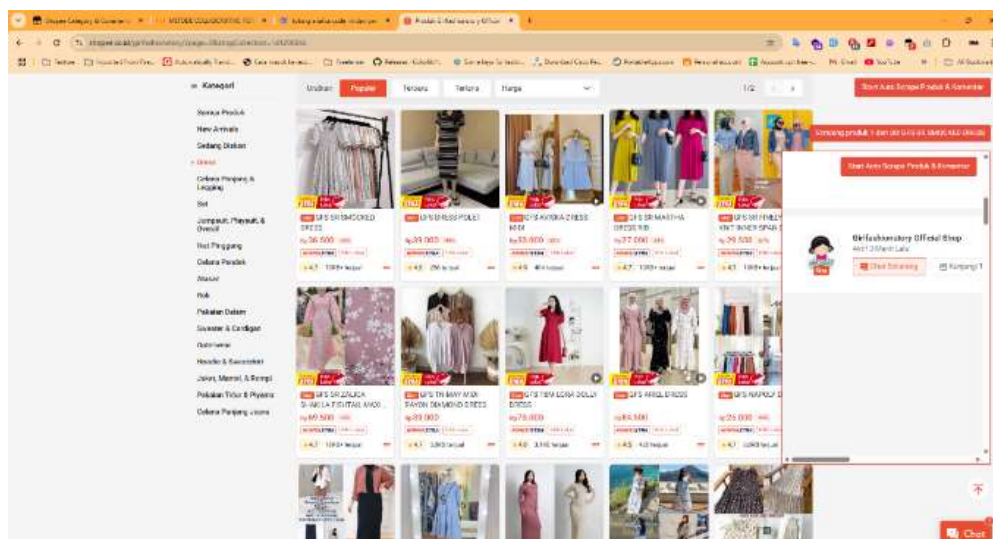
function scrapeComments() {
    const comments = [];
    document.querySelectorAll('.shopee-product-rating').forEach(commentElement => {
        const avatar = commentElement.querySelector('.shopee-product-rating_avatar img')?.src || 'No Avatar';
        const authorName = commentElement.querySelector('.shopee-product-rating_author-name')?.innerText.trim() || 'No Name';
        const rating = commentElement.querySelector('.icon-rating-solid').length;
        const time = commentElement.querySelector('.shopee-product-rating_time')?.innerText.trim() || 'No Time';
        const description = Array.from(commentElement.querySelectorAll('div[style*="margin-top"]'))
            .map(el => el.textContent.trim()).join(' ') || 'No Description';

        comments.push([avatar, authorName, rating, time, description]);
    });

    allComments = allComments.concat(comments);
    //console.table(comments);
}
}
```

Gambar 2. Source Code Scraping Data

Langkah-langkah pelaksanaan *web scraping* dilakukan secara sistematis. Peneliti terlebih dahulu memilih kategori produk yang akan dijadikan fokus, yaitu produk *clothing* seperti gamis, khimar, dan atasan wanita. Setelah kategori ditentukan, peneliti menjalankan *script Tamper Monkey* dengan menekan tombol “Start Auto Scraping Produk & Komentar” pada halaman Shopee. Proses ini akan berjalan secara otomatis, di mana *script* menelusuri setiap elemen HTML pada laman produk dan mengekstrak data sesuai atribut yang diinginkan. Tahap ini membutuhkan kestabilan jaringan internet serta pengaturan jeda waktu (*delay*) tertentu agar sistem Shopee tidak mendeteksi aktivitas otomatis sebagai *spam request*. Tampilan hasil *scraping* dapat dilihat melalui contoh toko GFS yang menjadi salah satu objek uji coba pengambilan data.



Gambar 3. Tampilan Web Scraping

Seluruh data hasil *scraping* kemudian disimpan dalam format CSV (*Comma-Separated Values*). Format ini dipilih karena memiliki struktur penyimpanan sederhana

berbasis teks yang mudah diakses melalui berbagai platform pengolahan data seperti *Microsoft Excel*, *Google Sheets*, hingga bahasa pemrograman seperti Python. File CSV menjadi dasar dalam tahap pra-pengolahan data sebelum diterapkan pada algoritma rekomendasi. Data yang disimpan terdiri dari beberapa kolom penting seperti *author name*, *rating*, *review time*, *product name*, serta *review description*. Pemilihan format CSV juga mendukung interoperabilitas data, artinya file tersebut dapat dengan mudah diimpor ke berbagai pustaka Python seperti Pandas dan NumPy untuk dilakukan analisis lanjutan.

Tabel 3. Contoh Data Hasil *Scraping*

Author Name	Rating	Time	Produk	No. Description
f*****i	1	2024-12-15 13:35	ANDINI BLOUSE	Penampilan: sangat menarik dan baguss. Bahan: nyaman dipakai. Kenyamanan: adem dan sangat nyaman di pakai. bareng di kemas dengan baik, untuk kualitasnya baguss dan untuk jasa pengirimannya sangat cepat
g*****n	5	2024-12-16 15:59	ANDINI BLOUSE	Desain: modern. Bahan: tipis. Kenyamanan: lumbanyan lah. Lumbanyan lah dengan harga 30rb..bahannya bgs.warnanya jg bgs...sesuai yg aku pengen
deningcucu	5	2024-12-18 08:02	ANDINI BLOUSE	Desain: menarik. Bahan: tipis. Kenyamanan: nyaman. Alhamdulillah barang sudah sampai sesuai dengan pesanan pengiriman dan pengemasan baik, kemeja nya pas, bahannya tipis
ristimila	5	2024-12-19 22:10	ANDINI BLOUSE	Bahan: sedeng, lentur dan lembut bgt. Desain: elegan dn feminim. Pas untuk Acara: apapun .. Sdh dtang ya kak pket yg ditnggu 😊☐☐mksih seller blousemya nyaman bgt dipke gtrllu ngpress asli gx nyesel siih uk XL kya L hehe ga nerawang lagi . 🙏 "
i*****6	5	2024-12-19 19:01	ANDINI BLOUSE	Desain: baik. Bahan: agak nerawang dikit tapi gk begitu kelihatan. Kenyamanan: kain berbahan adem. Nyaman di pakai
thiiy_allty	5	2024-12-31 06:57	ANDINI BLOUSE	Bahan: nyaman. Desain: menarik. Pas untuk Acara:

Author Name	Rating	Time	Produk	No. Description
				resmi ataupun vsbdfđ. Cantik desainnya,beda darinkwmeja pafa imunnya,saya suka
mazvo5vskp	5	2024-12-20 18:49	ANDINI BLOUSE	Bahan: lumayan sih dengan harga segitu. Desain: Desainnya menarik sh menurutku cuma jahitannya kurang rapih dan banyak benang yg keluar jd hrs telaten ngguntingin benangnya,,setelah dicuci & disetrika sh lumayan bagus soalnya jd kelihatan lebih rapih ngga kusut soalnya bahannya rawan kusut menurutku jd hrs rajin nyetrika biar rapih terus
dewi75nov	5	2024-12-29 23:56	ANDINI BLOUSE	Bahan: nyaman. Desain: cantik dan unik. Pas untuk Acara: untuk kerja,santai. Tadi nya PO,ternyata vepat datangnya,dan bagus dwh

Tahap berikutnya yaitu *upload* data ke Google Drive, di mana seluruh file CSV disimpan ke dalam folder dengan nama *METODE_COLLABORATIVE_FILTERING-UNTUK_REKOMENDASI_PRODUK_CLOTHING*. Langkah ini dilakukan untuk memudahkan akses dan integrasi dengan *Google Colab*, platform berbasis *cloud* yang digunakan dalam tahap pengolahan dan analisis data. Dengan penyimpanan berbasis *cloud*, data lebih aman, mudah dibagikan antar perangkat, serta mendukung replikasi penelitian.



Gambar 4. Folder Google Drive

Setelah data berhasil disimpan, tahap *cleaning* data dilakukan untuk menghapus data duplikat, memperbaiki format penulisan, serta menghapus baris data yang tidak relevan. Proses *data cleaning* menggunakan *Google Colab* dengan bahasa pemrograman *Python*. Pada tahap ini, peneliti menjalankan sejumlah *script* untuk menormalkan struktur dataset, menghapus nilai kosong (*missing value*), serta melakukan konversi tipe data yang tidak sesuai. Tujuan utama dari tahap ini adalah memastikan kualitas data agar

sistem rekomendasi dapat bekerja dengan akurat. Hasil dari tahap ini menunjukkan bahwa dataset akhir terdiri atas beberapa variabel utama, yakni nama pengguna, nama toko, harga produk, jumlah produk terjual, rating produk, kategori produk, dan deskripsi komentar pengguna.

```
product_files = [f for f in os.listdir(data_folder) if
                  f.endswith('_products.csv')]
comment_files = [f for f in os.listdir(data_folder) if
                  f.endswith('_comments.csv')]
```

Gambar 5. Source Code Cleaning Data

```
{
  "data": {
    "Girlfashionstory Official Shop": {
      "Atasan": [
        {
          "nama": "GFS ANDINI BLOUSE",
          "price": 34500,
          "discount": -0.62,
          "rating": 3.3,
          "sold": 10584,
          "link": "https://shopee.co.id/GFS-SR-ANDINI-BLOUSE-i.5825746.29519977046",
          "image_source": "https://down-id.img.susercontent.com/file/id-11134258-7ra01-
mcag7vaa9cgw51",
          "comments": [
            {
              "avatar": "No Avatar",
              "author_name": "f*****i",
              "rating": 1.0,
              "time": "2024-12-15 13:35",
              "variation": "MAUVE PINK,STANDAR",
              "description": "No Description"
            },...
          ],...
        },...
      ],...
    }
  }
}
```

Gambar 6. Hasil *Cleaning Data*

Proses Perhitungan *Similarity*

Tahap berikutnya dalam penelitian ini adalah perhitungan *similarity matrix* atau matriks kemiripan antar pengguna. Tahapan ini merupakan inti dari metode *User-Based Collaborative Filtering*, di mana sistem akan mencari hubungan atau kesamaan antar pengguna berdasarkan pola rating terhadap produk yang sama. Dengan kata lain, sistem mengasumsikan bahwa pengguna dengan preferensi serupa di masa lalu cenderung memiliki kesamaan dalam menilai produk di masa depan. Perhitungan kemiripan antar pengguna dilakukan dengan menggunakan metode *Euclidean Distance*, yang merupakan salah satu pendekatan paling umum untuk mengukur jarak antar dua titik dalam ruang multidimensi. Dalam konteks ini, titik-titik tersebut merepresentasikan rating pengguna terhadap berbagai produk.

Tabel 4. Dataset Rating Pengguna

Pengguna	ANDINI BLOUSE	IKO RAYON DIAMOND SHIRT	BESLI KAOS OVERSIZE	ARILA ROMPI KEMEJA BORDIR	CELIA OVERSIZE BAHAN KATUN
f*****i	1	-	-	-	-
g*****n	5	-	-	-	-
deningcucu	5	5	-	-	-
ristimila	5	-	-	-	-
i*****6	5	-	-	-	-
thiyy_allty	5	-	-	-	-
mazvo5vskp	5	-	-	-	-
dewi75nov	5	-	-	-	5

Secara matematis, rumus Euclidean Distance dirumuskan sebagai berikut:

$$Distance(deningcucu, dewi75nov) = \sqrt{(5 - 5)^2} = 0$$

$$Simmilarity(deningcucu, dewi75nov) = \frac{1}{1 + 0} = 1$$

Nilai *similarity* akan berkisar antara 0 hingga 1, di mana semakin mendekati angka 1 menunjukkan tingkat kemiripan yang semakin tinggi antar pengguna. Berdasarkan hasil perhitungan, pengguna deningcucu dijadikan sebagai pengguna target dalam sistem rekomendasi. Untuk memprediksi rating produk *CELIA OVERSIZE BAHAN KATUN*, peneliti menggunakan data dari dua pengguna lain, yaitu dewi75nov dan fitriastuti179, yang juga memberikan rating pada produk serupa. Hasil perhitungan menunjukkan bahwa:

$$Distance(deningcucu, fitriastuti179)$$

$$= \sqrt{(5 - 4)^2 + (5 - 5)^2} = \sqrt{1^2 + 0} = 1$$

$$Simmilarity(deningcucu, fitriastuti179) = \frac{1}{1 + 1} = 0,5$$

Tabel 5. Hasil Kemiripan dengan *Euclidean Distance*

Pasangan Pengguna	Euclidean Distance	Similarity
deningcucu – dewi75nov	0	1.0
deningcucu – fitriastuti179	1	0.5

Dari hasil di atas, dapat disimpulkan bahwa dewi75nov memiliki tingkat kemiripan tertinggi dengan deningcucu, dengan nilai *similarity* sebesar 1.0, sedangkan fitriastuti179 memiliki kemiripan sedang sebesar 0.5. Artinya, sistem akan lebih memprioritaskan data rating dari pengguna dewi75nov dalam proses prediksi karena tingkat kemiripannya lebih tinggi.

Proses Prediksi Rating

Tahap terakhir adalah prediksi rating menggunakan metode *Weighted Sum Prediction*, di mana sistem menghitung nilai rating yang belum diberikan oleh pengguna target terhadap produk tertentu. Proses ini dilakukan dengan menjumlahkan seluruh rating dari pengguna-pengguna yang memiliki kemiripan, lalu dikalikan dengan bobot kemiripan masing-masing pengguna. Tujuan dari metode ini adalah untuk menghasilkan prediksi yang lebih realistis dan mendekati preferensi pengguna sebenarnya.

Tabel 6. Rating Data untuk Prediksi

Pengguna	ANDINI BLOUSE	IKO RAYON DIAMOND SHIRT	CELIA OVERSIZE BAHAN KATUN
deningcucu	5	5	?
dewi75nov	5	-	5
fitriastuti179	4	5	5

Simbol tanda tanya “?” dalam tabel menunjukkan bahwa pengguna target belum memberikan rating pada produk tersebut, sehingga sistem perlu memperkirakannya berdasarkan pola perilaku pengguna lain yang serupa. Formula *Weighted Sum* dapat dituliskan sebagai berikut:

$$\hat{r}_{u,i} = \bar{r}_u + \frac{\sum_{v \in Nsim(u,v)} (r_{v,i} - \bar{r}_v)}{\sum_{v \in Nsim(u,v)} 1}$$

$$\hat{r}_{denningcucu, CELIA} = 5,0 + \frac{[1 \times (5 - 5,0)] + [0,5 \times (5 - 4,67)]}{1 + 0,5} = 5,0 + \frac{[0] + [0,5 \times (0,5 - 0,33)]}{1,5}$$

$$= 5,0 + \frac{0,165}{1,5} = 5,0 + 0,11 = 5,11$$

Dengan menggunakan rumus tersebut, sistem menghasilkan prediksi rating akhir untuk pengguna target. Nilai prediksi yang tinggi menunjukkan bahwa produk tersebut berpotensi besar disukai oleh pengguna, sehingga layak direkomendasikan. Melalui serangkaian pengujian dan validasi, sistem *Collaborative Filtering* yang dikembangkan dalam penelitian ini terbukti mampu memberikan rekomendasi produk *clothing* secara personal dan akurat, serta membantu calon konsumen dalam menentukan pilihan produk yang sesuai dengan preferensinya.

Pembahasan

Berdasarkan hasil penelitian ini, implementasi metode *User-Based Collaborative Filtering* untuk rekomendasi produk *clothing* menunjukkan bahwa teknik pengumpulan data via web scraping menggunakan ekstensi *Tamper Monkey (JavaScript)* efektif dalam mengekstraksi secara otomatis atribut pengguna seperti nama, rating, dan deskripsi

ulasan. Data yang terkumpul kemudian diolah dan disimpan dalam format CSV, yang memfasilitasi tahap prapemrosesan dan penggunaan selanjutnya dalam algoritma rekomendasi. Teknik ini mempercepat proses pengumpulan data dalam skala besar dan memungkinkan pengolahan data numerik dan teks deskriptif secara simultan.

Salah satu aspek penting dari penelitian ini adalah perhitungan kemiripan antar pengguna menggunakan metode *Euclidean Distance*. Proses ini menentukan pengguna mana yang memiliki preferensi serupa berdasarkan rating yang mereka berikan terhadap produk-produk *clothing* yang sama. Hasil menunjukkan bahwa semakin rendah jarak *Euclidean* antar pengguna, semakin tinggi kemiripannya, sesuai dengan teori bahwa pengguna yang memberikan rating yang mirip cenderung memiliki selera yang sama. Misalnya, dalam contoh perhitungan, pengguna “deningcucu” dan “dewi75nov” memiliki kemiripan nilai 1 (jarak 0), sedangkan dengan “fitriastuti179” nilai kemiripannya 0,5 (jarak 1). Temuan ini sejalan dengan penelitian oleh (Hu et al., 2018), yang juga menggunakan data nyata dari e-commerce besar dan menunjukkan bahwa ukuran kemiripan yang tepat (menggunakan frekuensi pengguna atau *inverse user frequency*) dapat meningkatkan performa sistem rekomendasi, terutama dalam presisi dan *recall*.

Proses prediksi rating dalam penelitian ini dibuat berdasarkan rating rata-rata dari pengguna-pengguna yang memiliki kemiripan tinggi dengan pengguna target. Prediksi rating tersebut kemudian digunakan untuk membuat daftar rekomendasi produk yang paling cocok. Evaluasi dengan metrik *Mean Absolute Error* (MAE) menunjukkan sistem memiliki tingkat kesalahan yang rendah, mengindikasikan bahwa prediksi mendekati nilai aktual pengguna. Hasil ini sejalan dengan penelitian oleh (Liu et al., 2022) yang menggunakan dataset *MovieLens*, di mana metode CF berdasarkan atribut pengguna dan skor item menghasilkan MAE yang lebih rendah dibandingkan metode CF tradisional.

Secara keseluruhan, penelitian ini memperlihatkan bahwa penerapan *User-Based Collaborative Filtering* cukup efektif dalam memberikan rekomendasi produk *clothing* yang bersifat personal dan relevan. Sistem mampu memanfaatkan data rating dari platform e-commerce seperti Shopee untuk mengenali pola kemiripan preferensi antar pengguna dan menebak rating pada item yang belum mereka ulas. Namun demikian terdapat beberapa keterbatasan, seperti hanya menggunakan atribut rating dan ulasan saja, serta terbatas pada satu platform (Shopee). Penelitian lain yang menggunakan

berbagai sumber data dan atribut tambahan (misalnya metadata produk, fitur konten) menunjukkan bahwa kombinasi atribut tersebut dapat menambah akurasi rekomendasi. Sebagai contoh, (Hu et al., 2018) mengkombinasikan *similarity* berdasarkan frekuensi pengguna dan pengurangan matriks co-occurrence untuk memperbaiki performa, sementara (Liu et al., 2022) menunjukkan bahwa CF dengan atribut pengguna dan skor item memiliki MAE lebih baik dibanding CF tradisional tanpa atribut tambahan.

SIMPULAN

Penelitian ini menunjukkan bahwa penerapan metode *User-Based Collaborative Filtering* efektif dalam menghasilkan rekomendasi produk *clothing* yang personal dan relevan berdasarkan pola kesamaan perilaku pengguna. Teknik *web scraping* dengan *Tamper Monkey* terbukti efisien dalam pengumpulan data secara otomatis dari *platform Shopee*, sedangkan implementasi algoritma UBCF dengan *Python* memberikan hasil prediksi yang akurat berdasarkan penghitungan kemiripan menggunakan *Euclidean Distance*. Evaluasi sistem menggunakan metrik MAE menunjukkan tingkat kesalahan prediksi yang rendah, menandakan performa yang baik. *Collaborative Filtering* dapat diandalkan untuk sistem rekomendasi berbasis preferensi pengguna. Meskipun demikian, keterbatasan penelitian ini terletak pada ruang lingkup data yang masih terbatas pada atribut rating dan ulasan dari satu *platform* saja. Penelitian lanjutan disarankan untuk menggabungkan lebih banyak sumber data dan atribut tambahan seperti metadata produk, kategori, atau gaya pengguna guna meningkatkan akurasi dan personalisasi sistem rekomendasi secara lebih komprehensif.

DAFTAR PUSTAKA

- Aamer, N., Mittal, T., Muthusamy, N. K., Battula, R., S, A., & Soudagar, M. (2025). Comparing Collaborative Filtering Algorithms in AI-Powered Recommendation Systems for E-Commerce. *2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL)*, 1–6. <https://doi.org/10.1109/icsadl65848.2025.10933448>
- Ahadiyah, F. (2023). Perkembangan Teknologi Infomasi Terhadap Peningkatan Bisnis Online. *INTERDISIPLIN: Journal of Qualitative and Quantitative Research*, 1, 41–49. <https://doi.org/10.61166/interdisiplin.v1i1.5>
- Chang, D., Gui, H., Fan, R., Fan, Z., & Tian, J. (2019). Application of Improved Collaborative Filtering in the Recommendation of E-commerce Commodities. *Int. J. Comput. Commun. Control*, 14, 489–502. <https://doi.org/10.15837/ijccc.2019.4.3594>
- Halim, F., Wijaya, A., & Wiyono, W. (2022). Analisis dan Perancangan E-Commerce

- Berbasis Web Dengan Penerapan Sistem Perekomendasi Menggunakan Metode Collaborative Filtering Serta Metode Up, Down, Cross Selling. *Algor*, 4, 28–43. <https://doi.org/10.31253/algor.v4i1.1516>
- Han, X. (2023). *Research on clothing personalized recommendation algorithm based on improved collaborative filtering algorithm*. 12708, 127080. <https://doi.org/10.1117/12.2683852>
- Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, 22(1), 5–53.
- Hu, Zhi-Hua, Li, Xiang, Wei, Chen, & Zhou, Hong-Lei. (2018). Examining collaborative filtering algorithms for clothing recommendation in e-commerce. *Textile Research Journal*, 89(14), 2821–2835. <https://doi.org/10.1177/0040517518801200>
- Jaja, V., Susanto, B., & Sasongko, L. (2020). Penerapan Metode Item-Based Collaborative Filtering Untuk Sistem Rekomendasi Data MovieLens. *D’CARTESIAN*, 9, 78. <https://doi.org/10.35799/dc.9.2.2020.28274>
- Jin, Z., Ye, F., Nedjah, N., & Zhang, X. (2024). A comparative study of various recommendation algorithms based on E-commerce big data. *Electron. Commer. Res. Appl.*, 68, 101461. <https://doi.org/10.1016/j.eierap.2024.101461>
- Li, L. (2024). Research on Personalized Recommendation System for E-Commerce Products Based on Collaborative Filtering Algorithm. *2024 IEEE 3rd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)*, 876–880. <https://doi.org/10.1109/eebda60612.2024.10485710>
- Liu, C., Kong, X., Li, X., & Zhang, T. (2022). Collaborative Filtering Recommendation Algorithm Based on User Attributes and Item Score. *Scientific Programming*, 2022(1), 4544152. <https://doi.org/https://doi.org/10.1155/2022/4544152>
- Maesaroh, S., Afiyati, A., Hakim, L., Sari, Y., Yusuf, M., Budi Perkasa, E., Utami, W., Saptadi, N., Mutmainah, S., Khairunnas, K., Harahap, E., Alamin, Z., Karima, I., Saputra, A., Mubarak, R., & Pustaka, S. (2024). *Bahasa Pemrograman Python*.
- Nizamuddin, H., Azan, K., Khairul Anwar, M. S. I., Muhammad Ashoer, S. E., Nuramini, A., Irlina Dewi, M. H., Abrory, M., Pebriana, P. H., Jafar Basalamah, S. E., & Sumianto, M. P. (2021). *Metodologi Penelitian; Kajian Teoritis Dan Praktis Bagi Mahasiswa*. CV. Dotplus Publisher.
- Pradana, M., & Noorrizki, R. (2022). Analisis Motif Persuasi Affiliator dalam Menunjang Pemasaran Platform Binary Option. *Flourishing Journal*, 2, 375–383. <https://doi.org/10.17977/um070v2i52022p375-383>
- Rahman, M. S., Sarkar, T. D., Mitasha, U. T., Mia, M. S., & Karthikeyan, S. (2024). E-commerce-based Smart Recommendation System using element-by-element collaborative filtering following with the Machine Learning Technology. *2024 International Conference on Communication, Computing and Internet of Things (IC3IoT)*, 1–6. <https://doi.org/10.1109/ic3iot60841.2024.10550386>
- Roy, S. S., Kumar, A., & Kumar, R. S. (2024). Metadata and Review-Based Hybrid Apparel Recommendation System Using Cascaded Large Language Models. *IEEE*

Access, 12, 140053–140071. <https://doi.org/10.1109/access.2024.3462793>

Saprida, S., Umari, Z. F., & Raya, F. (2023). Legalitas transaksi jual beli online di Indonesia. *Ekonomika Sharia: Jurnal Pemikiran Dan Pengembangan Ekonomi Syariah*, 8(2), 315–336.

Sri, P. H. (2025). Real-Time Recommendations for E-Commerce. *INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*. <https://doi.org/10.55041/ijrem49375>

Xia, J. (2016). E-Commerce Product Recommendation Method Based on Collaborative Filtering Technology. *2016 International Conference on Smart Grid and Electrical Automation (ICSGEA)*, 90–93. <https://doi.org/10.1109/icsgea.2016.81>

Zhang, J., Peng, Q., Sun, S., & Liu, C. (2014). Collaborative filtering recommendation algorithm based on user preference derived from item domain features. *Physica A: Statistical Mechanics and Its Applications*, 396, 66–76. <https://doi.org/10.1016/j.physa.2013.11.013>