

IMPLEMENTASI SISTEM REKOMENDASI PRODUK E-COMMERCE MENGGUNAKAN CONTENT-BASED FILTERING BERBASIS COSINE SIMILARITY

Stenly Ibrahim Adam¹, Wilsen Grivin Mokodaser*²

^{1,2} Informatika, Fakultas Ilmu Komputer, Universitas Klabat, Airmadidi, Sulawesi Utara
Email: stenly.adam@unklab.ac.id, wilsenm@unklab.ac.id

Pesatnya perkembangan e-commerce menghadirkan tantangan berupa banyaknya pilihan produk yang dapat menimbulkan information overload bagi konsumen. Untuk mengatasi permasalahan tersebut, penelitian ini mengembangkan sistem rekomendasi produk berbasis Content-Based Filtering dengan Cosine Similarity. Metode ini memanfaatkan kombinasi fitur teks (judul dan deskripsi produk) yang direpresentasikan dengan TF-IDF, serta fitur numerik (harga, rating, dan jumlah rating) yang dinormalisasi menggunakan StandardScaler. Selanjutnya, seluruh fitur digabungkan dan dihitung tingkat kesamaannya dengan cosine similarity untuk menghasilkan rekomendasi produk yang relevan. Hasil penelitian menunjukkan bahwa pendekatan ini mampu memberikan rekomendasi yang logis, di mana produk dengan spesifikasi serupa ditampilkan secara berurutan berdasarkan tingkat kesamaan. Analisis tambahan juga memperlihatkan bahwa mayoritas produk memiliki rating tinggi meskipun harga bervariasi, menunjukkan harga bukan satu-satunya indikator kualitas. Dengan demikian, sistem ini terbukti efektif membantu konsumen dalam menemukan produk sesuai preferensi sekaligus memberikan insight bagi pelaku e-commerce.

Kata Kunci— E-Commerce, Content-Based Filtering, Cosine Similarity, TF-IDF.

I. PENDAHULUAN

Dokumen Memasuki era perkembangan zaman yang diiringi oleh kemajuan teknologi, aktivitas masyarakat menjadi semakin mudah, termasuk dalam transaksi jual beli dan pemasaran. Perkembangan teknologi yang pesat mendorong masyarakat untuk beradaptasi dengan kemudahan yang ditawarkan, khususnya kemajuan teknologi berbasis internet. Berdasarkan data Kementerian Komunikasi dan Informatika, jumlah pengguna internet di Indonesia terus meningkat signifikan setiap tahunnya: tahun 2019 mencapai 172 juta pengguna, tahun 2020 naik menjadi 200 juta, tahun 2021 sebanyak 201 juta, tahun 2022 sebanyak 202 juta, dan hingga Januari 2023 mencapai 213 juta pengguna.[1] Angka

tersebut menunjukkan bahwa pemanfaatan internet telah menjadi kebutuhan masyarakat Indonesia.

Seiring dengan meningkatnya jumlah pengguna internet, sektor e-commerce di Indonesia juga mengalami pertumbuhan luar biasa dalam beberapa tahun terakhir. Survei *We Are Social* pada April 2021 menunjukkan bahwa 88,1% pengguna internet di Indonesia menggunakan layanan belanja online, menjadikannya salah satu tingkat adopsi tertinggi di dunia.[2] Fenomena ini dipengaruhi oleh kemunculan berbagai situs jual beli online atau *marketplace* yang secara positif mempengaruhi tingginya tingkat penggunaan internet dan memacu perkembangan e-commerce di Indonesia.

Keuntungan dari segi efisiensi waktu dan tenaga menjadi faktor utama konsumen memilih berbelanja online, di samping kualitas produk, harga yang bersaing, dan kemudahan transaksi. Desain dan kualitas halaman situs juga penting untuk menarik minat konsumen. Shopee sebagai salah satu *marketplace* paling populer di Indonesia tidak hanya menjadi platform e-commerce terkemuka, tetapi juga menjadi wadah penjualan online yang lengkap dengan berbagai macam produk.[3] Kehadirannya memberikan kontribusi signifikan terhadap dinamika pasar e-commerce di Indonesia.

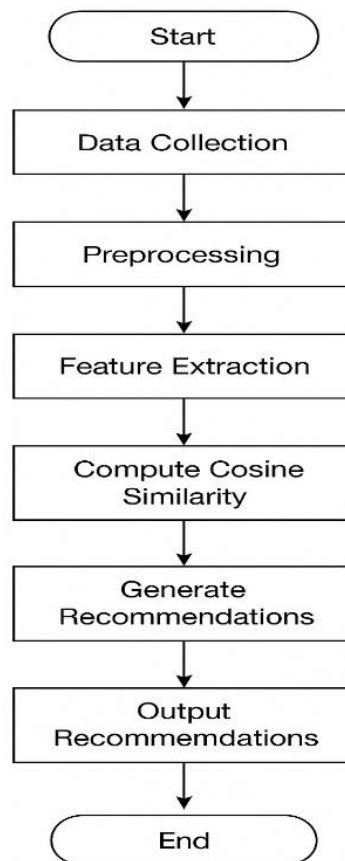
Pertumbuhan pesat e-commerce di Indonesia menunjukkan tingginya minat masyarakat dalam melakukan transaksi daring. Namun, perkembangan ini juga membawa tantangan baru, seperti meningkatnya jumlah pesaing di platform yang sama, kesulitan konsumen dalam memilih produk terbaik di tengah banyaknya pilihan, serta kebutuhan akan layanan yang cepat, aman, dan terpercaya.[4] Pelaku usaha dituntut untuk memahami preferensi konsumen secara lebih mendalam dan menyajikan penawaran yang relevan agar mampu bertahan dalam persaingan yang ketat. Di sisi konsumen, muncul kebutuhan akan solusi yang dapat membantu proses pengambilan keputusan pembelian, mulai dari perbandingan harga, kualitas, ulasan pengguna, hingga jaminan keamanan transaksi. Oleh karena itu, dibutuhkan pendekatan berbasis teknologi untuk

mengatasi masalah-masalah tersebut secara efisien dan tepat sasaran.

Untuk menjawab kebutuhan tersebut, berbagai teknologi telah berkembang dan diadopsi oleh pelaku *e-commerce* maupun penyedia layanan pendukungnya. Teknologi analitik data besar (*big data analytics*) memungkinkan perusahaan menganalisis perilaku dan preferensi pelanggan secara real time.[5] Sistem rekomendasi berbasis *machine learning* membantu menampilkan produk sesuai kebutuhan dan minat pengguna. *Cloud computing* mendukung skala layanan yang lebih luas dengan biaya yang efisien, sementara teknologi pembayaran digital (*e-payment*) memudahkan dan mempercepat proses transaksi.[6] Selain itu, teknologi keamanan siber seperti enkripsi *end-to-end* dan otentikasi ganda digunakan untuk melindungi data pengguna. Kombinasi teknologi-teknologi ini bukan hanya meningkatkan pengalaman pengguna, tetapi juga membantu pelaku usaha menciptakan strategi pemasaran yang lebih tepat sasaran dan adaptif terhadap dinamika pasar *e-commerce*.

Pemilihan metode *Content-Based Filtering* berbasis *Cosine Similarity* pada penelitian ini didasarkan pada karakteristik data dan kebutuhan sistem rekomendasi di platform *e-commerce*. Pada *marketplace* seperti Shopee, data yang tersedia lebih banyak berupa atribut produk seperti judul, deskripsi, harga, rating, dan jumlah ulasan dibandingkan dengan data interaksi pengguna yang mendetail. Kondisi ini menjadikan pendekatan berbasis konten lebih tepat digunakan karena sistem hanya memerlukan informasi mengenai item atau produk itu sendiri tanpa ketergantungan besar pada histori perilaku pengguna sebagaimana pada metode *collaborative filtering*. [7] Selain itu, *Cosine Similarity* dinilai efektif untuk mengukur kemiripan antarproduk pada ruang vektor berukuran besar, seperti hasil representasi teks dengan TF-IDF.[8] Metode ini berfokus pada sudut antara vektor, bukan pada besarannya, sehingga lebih stabil terhadap perbedaan panjang judul atau deskripsi produk.[9] Dengan demikian, metode ini mampu menghasilkan rekomendasi yang relevan berdasarkan kemiripan atribut produk, mengatasi masalah data yang bersifat jarang (*sparse*), serta mudah diimplementasikan dengan waktu komputasi yang relatif cepat sehingga cocok untuk skala data yang besar. Pendekatan ini juga memberikan fleksibilitas bagi peneliti untuk menggabungkan fitur nonteks, misalnya harga, rating, dan jumlah penilaian, dengan fitur teks seperti judul atau deskripsi produk, sehingga kualitas rekomendasi yang dihasilkan menjadi lebih baik. Oleh karena itu, *Content-Based Filtering* berbasis *Cosine Similarity* dipandang sebagai metode yang tepat dan efisien untuk penelitian ini.

II. METODE PENELITIAN



Gambar 1 Tahapan Penelitian

A. Data Collection

Tahap pertama dimulai dengan pengumpulan dan pemuatan data dari Google Drive ke lingkungan Google Colab untuk mempermudah proses komputasi. Dataset berisi informasi produk *e-commerce* seperti judul, harga, jumlah rating, dan rating bintang.

B. Data Preprocessing

Tahap berikutnya adalah pra-pemrosesan data (*data pre-processing*). Pada tahap ini dilakukan pembersihan kolom harga dengan menghapus simbol mata uang (seperti ₹ atau \$) dan tanda baca sehingga hanya menyisakan angka. Nilai kosong pada kolom harga diisi dengan nilai median agar tidak mengganggu analisis.[10] Begitu pula nilai kosong pada kolom rating dan jumlah rating juga diisi dengan median untuk menjaga konsistensi data. Tahapan ini bertujuan menyiapkan data agar siap diproses pada tahap berikutnya.

C. Feature Extraction

Pada tahap ini informasi penting dari produk diubah menjadi vektor numerik: Judul/nama produk diubah menjadi representasi numerik menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*) untuk menangkap bobot pentingnya kata dalam judul produk.[11] Fitur numerik seperti harga, rating bintang, dan jumlah rating dinormalisasi menggunakan *StandardScaler* agar berada dalam skala yang sama. Hasil TF-IDF digabung dengan fitur numerik menjadi satu matriks fitur yang lengkap.

D. Compute Cosine Similarity

Setelah semua produk direpresentasikan dalam bentuk vektor, sistem menghitung tingkat kemiripan antarproduk menggunakan *cosine similarity*. Nilai ini menunjukkan seberapa mirip dua produk berdasarkan fitur gabungan (judul, harga, rating, jumlah rating). Semakin tinggi nilainya, semakin mirip produk tersebut.

E. Generate Recommendation

Sistem rekomendasi dibangun berdasarkan matriks kesamaan yang telah dihitung. Input judul produk yang dipilih pengguna. Proses mengambil skor *cosine similarity* antara produk yang dipilih dengan produk

lainnya. Output menampilkan daftar produk yang paling mirip (Top-N rekomendasi).

F. Visualisation

Untuk memperkuat analisis, dilakukan visualisasi data dan hasil rekomendasi. Visualisasi meliputi distribusi harga produk, distribusi rating bintang, hubungan harga dengan rating dalam bentuk scatter plot, grafik batang skor kemiripan produk yang direkomendasikan, serta *heatmap cosine similarity* untuk menampilkan peta kemiripan antarproduk secara keseluruhan.[12] Visualisasi ini membantu peneliti maupun pembaca memahami pola dan karakteristik produk dalam dataset.

III. HASIL DAN PEMBAHASAN

1. Data Collection

Tabel 1. Struktur data

Product Title	Product Price	Product Star Rating	Product Num Rating
Kaspersky Premium - Total Security (Ultimate Security) 1 Device 1 Year Email Delivery in 1 Hour	469.00	4.3	13324
K7 Security K7, Total Security, 1 User, 1 Year, Email Delivery In 30 Min, No Cd	370.00	4.4	2291
Microsoft Office 2021 Professional - Lifetime Validity - Windows only - Email delivery in 2 Hours- no CD - Email Bind Key	1,799.00	4.5	388
Bitdefender - 1 Device, 1 Year - Mobile Security Android Latest Version Email Delivery In 2 Hours- No Cd	94.00	4.1	9630
McAfee Total Protection 2025 1 Device, 3 Years Antivirus Internet Security Software Scam Protection & Identity Monitori	1,699.00	4.4	5783
Dst...			

Data yang akan digunakan untuk penelitian ini merupakan data transaksi dari salah satu E-Commerce dengan jumlah data pada dataset sebanyak 999 baris dan 12 kolom yang merupakan struktur data sebelum dilakukan preprocessing

2. Data Preprocessing

Tabel 2. Hasil Preprocessing

No	Product Price	Product Price Clean
1	469.00	469.0
2	370.00	370.0
3	1.799.00	1799.0
4	94.00	94.00
5	1.699.00	1.699.00

Pada tabel 2 menampilkan hasil dari tahapan preprocessing, untuk menghasilkan output tersebut digunakan bahasa pemrograman python untuk membersihkan data harga dari simbol mata uang maupun karakter lain yang tidak relevan sehingga hanya menyisakan angka desimal. Pertama, nilai dalam kolom *product_price* diubah menjadi tipe string menggunakan *astype(str)* agar bisa diproses dengan metode string. Setelah itu, fungsi *str.replace(r'^0-9.', '', regex=True)* dijalankan untuk mengganti semua karakter yang bukan angka atau tanda titik dengan string kosong. Ekspresi reguler *['^0-9.']* berarti mencari karakter selain angka 0–9 dan titik, sehingga simbol rupee (₹) maupun tanda koma pada ribuan akan dihapus. Hasilnya, harga yang semula ditulis sebagai ₹1,799.00 akan berubah menjadi 1799.00. Data bersih ini kemudian disimpan pada kolom baru bernama *product_price_clean*, sehingga lebih mudah digunakan untuk perhitungan numerik atau analisis lanjutan.

3. Feature Extraction

```
# fitur numerik: harga, rating, jumlah rating
num_features =
df[['product_price_clean','product_star_rating','product_num_ratings']]
scaler = StandardScaler()
num_features_scaled = scaler.fit_transform(num_features)
# gabungkan TF-IDF + fitur numerik
combined_features = hstack([tfidf_matrix,
```

Code Snippet Feature Extraction

Pada code snippet diatas digunakan untuk mempersiapkan data teks agar bisa diubah menjadi representasi numerik dengan metode TF-IDF (*Term Frequency – Inverse Document Frequency*). Prosesnya dapat dijelaskan sebagai berikut: Pertama, kolom *product_title* dan *product_description* digabungkan menjadi satu kolom baru bernama *combined_features*. Tujuannya adalah menggabungkan berbagai informasi teks yang relevan mengenai produk agar representasi data lebih kaya. Misalnya, jika hanya menggunakan judul, informasi bisa kurang lengkap, sehingga deskripsi ditambahkan agar model lebih memahami konteks produk.

Kedua, dilakukan ekstraksi fitur teks menggunakan TF-IDF *Vectorizer*. Objek *TfidfVectorizer* dibuat dengan parameter *stop_words='english'*, artinya kata-kata umum (seperti "and", "the", "is") yang tidak memiliki makna penting akan diabaikan. Kemudian, fungsi *fit_transform()* dijalankan pada kolom *combined_features*, sehingga teks diubah menjadi sebuah matriks TF-IDF. Matriks ini menyimpan bobot numerik untuk setiap kata, di mana

bobot lebih tinggi menunjukkan kata tersebut lebih penting dalam menggambarkan produk dibandingkan kata lain yang sering muncul di seluruh dokumen. Dengan demikian, hasil akhir dari kode ini adalah sebuah *tfidf_matrix*, yaitu representasi numerik dari teks produk yang siap digunakan untuk berbagai keperluan analisis, seperti sistem rekomendasi, clustering, atau klasifikasi.

4. Compute Cosine Similarity

```
# cosine similarity
cosine_sim =
cosine_similarity(combined_features,
combined_features)

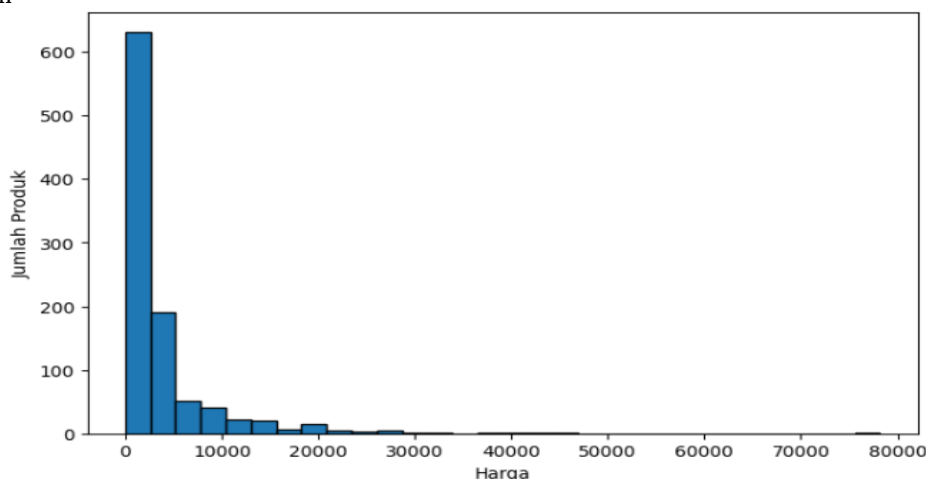
# mapping index->judul
indices = pd.Series(df.index,
index=df['product_title']).drop_duplicates(
)

# fungsi rekomendasi
def recommend_products(title,
cosine_sim=cosine_sim, top_n=5):
idx = indices[title]
```

Code Snippet proses *Compute Cosine Similarity*

Pada code snippet diatas digunakan untuk membangun sebuah sistem rekomendasi produk berbasis *hybrid*, yaitu dengan menggabungkan informasi teks dari judul produk serta fitur numerik berupa harga, rating, dan jumlah rating. Pertama, judul produk diubah menjadi representasi numerik menggunakan *TF-IDF Vectorizer*, sehingga setiap kata dalam judul diberi bobot sesuai tingkat kepentingannya. Selanjutnya, fitur numerik seperti harga, rating bintang, dan jumlah rating distandarasi menggunakan *StandardScaler* agar berada pada skala yang sebanding. Kedua jenis fitur ini kemudian digabungkan menjadi satu matriks yang merepresentasikan setiap produk secara lebih lengkap. Dari matriks tersebut dihitung *cosine similarity*, yaitu ukuran kemiripan antarproduk berdasarkan gabungan fitur teks dan numerik. Untuk memudahkan pencarian, dibuat pemetaan antara judul produk dan indeks data. Terakhir, didefinisikan fungsi *recommend_products()* yang menerima judul produk sebagai input, mencari produk tersebut dalam data, menghitung tingkat kemiripannya dengan produk lain, lalu mengembalikan sejumlah

6. Visualisation



Gambar 2. Distribusi harga produk

rekomendasi produk paling mirip beserta informasi harga, rating, dan jumlah ratingnya. Dengan cara ini, sistem dapat memberikan saran produk yang relevan kepada pengguna berdasarkan kombinasi deskripsi teks dan karakteristik numerik produk.

5. Generate Recommendation

Tabel 3. Hasil Rekomendasi produk

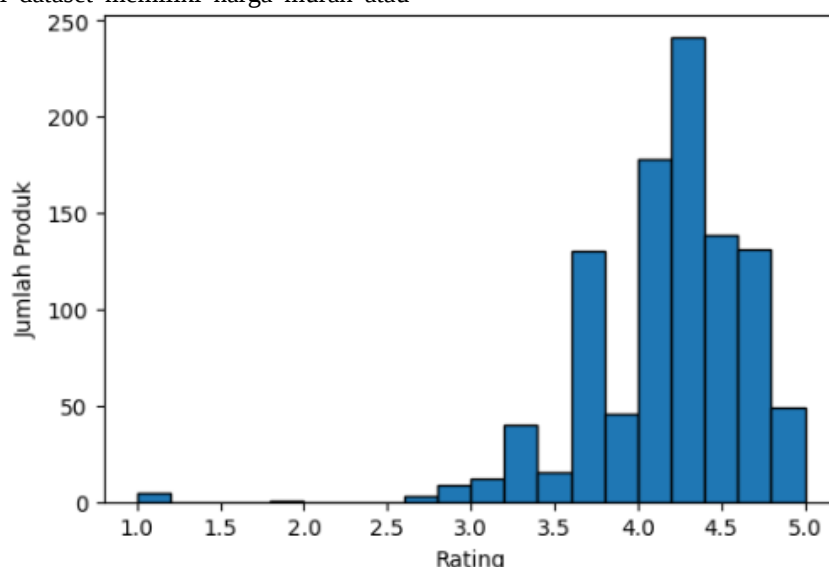
Product Acuan: Kaspersky Premium – Total Security (Ultimate Security) 1 Device 1 Year Rekomendasi produk mirip:			
Product Title	Product Price	Product Start Rating	Product Num Rating
Kaspersky Total Sec	469.00	4.3	9953.0
Kaspersky Internet Sec	199.00	4.2	119060
MacAfee 1 Device, 3 Year	750.00	4.2	16839.0
MacAfee 1 Device, 1 Year	399.00	4.1	17965.0
Quick Heal Total Sec	1.182.00	4.4	14912.0

Pada tabel 3 menunjukkan hasil rekomendasi tersebut menampilkan hasil penelitian sistem rekomendasi produk e-commerce menggunakan pendekatan *Content-Based Filtering* berbasis *Cosine Similarity*. Produk acuan yang digunakan adalah *Kaspersky Premium - Total Security (Ultimate Security) | 1 Device | 1 Year | Email Delivery in 1 Hour*.

Sistem kemudian menghitung tingkat kemiripan antara deskripsi produk acuan dengan produk lain berdasarkan representasi teksnya, sehingga diperoleh daftar rekomendasi produk yang paling mirip. Hasil rekomendasi menunjukkan produk-produk antivirus serupa, seperti *Kaspersky Total Security*, *Kaspersky Internet Security for Android*, *McAfee Antivirus 2024*, dan *Quick Heal Total Security*. Tabel yang ditampilkan tidak hanya memuat judul produk, tetapi juga informasi pendukung berupa harga, rating bintang, dan jumlah rating pengguna, sehingga pembeli dapat membandingkan relevansi, kualitas, serta popularitas produk. Dari hasil tersebut terlihat bahwa sistem berhasil memberikan rekomendasi yang konsisten karena produk yang ditawarkan memiliki fungsi dan karakteristik yang sejenis, misalnya sama-sama berupa software keamanan dengan lisensi satu perangkat dan durasi langganan tahunan.

Pada gambar 2 menunjukkan batang paling kiri (tinggi) menunjukkan ada ratusan produk dengan harga yang sangat rendah (mendekati 0–2000). Semakin ke kanan (harga lebih tinggi), batangnya makin rendah semakin sedikit produk di rentang harga tinggi. Ada beberapa produk dengan harga ekstrem (*outlier*) hingga mendekati 80.000, tapi jumlahnya sangat sedikit. Sebagian besar produk yang ada di dataset memiliki harga murah atau

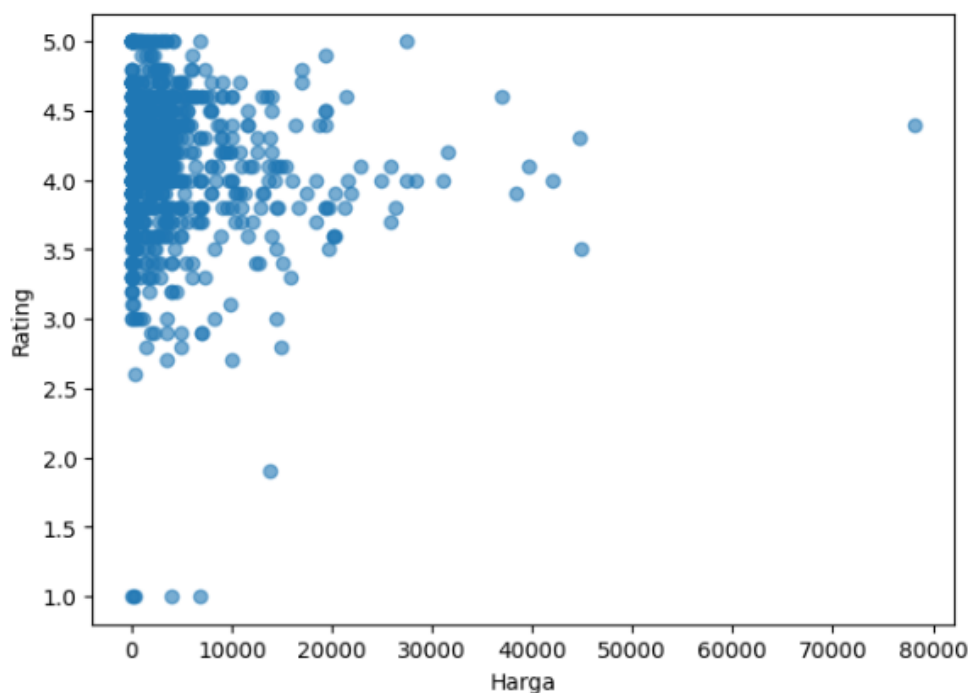
menengah (mungkin langganan software, aplikasi digital, dsb.). Hanya sedikit produk dengan harga mahal (produk premium atau paket langganan jangka panjang). Distribusi ini bersifat *right-skewed* (condong ke kanan): mayoritas harga berada di rentang rendah, dengan ekor panjang ke kanan (produk harga tinggi).



Gambar 3. Distribusi rating produk

Pada gambar 3 menunjukkan puncak histogram terlihat di rating 4.0 – 4.5 yang artinya paling banyak produk mendapatkan rating pada kisaran ini. Ada sedikit produk yang memiliki rating rendah (1.0 – 2.0). Semakin mendekati rating 5.0, jumlah produk menurun tetapi masih cukup banyak. Mayoritas produk dalam dataset

memiliki rating tinggi (di atas 3.5). Distribusi ini menunjukkan bahwa kebanyakan produk yang dijual atau direview merupakan produk dengan kualitas yang relatif baik menurut pengguna. Produk dengan rating rendah jumlahnya sedikit sekali, sehingga distribusi condong ke kanan (*left-skewed*).



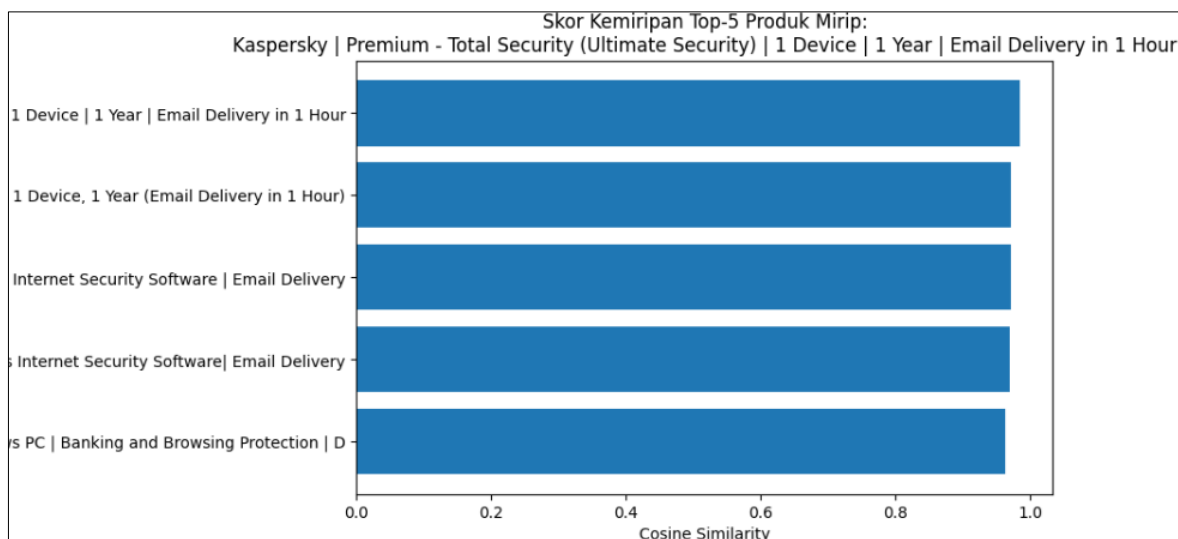
Gambar 4. Sebaran harga dan rating produk

Pada gambar 4 menunjukkan titik terkonsentrasi sangat padat di harga rendah (0 – 10.000) dengan rating tinggi (sekitar 4–5). Ada beberapa produk mahal (hingga 80.000) yang juga punya rating bervariasi, tetapi

jumlahnya jauh lebih sedikit. Tidak terlihat pola linear yang jelas antara harga dan rating produk mahal tidak selalu rating tinggi, dan produk murah tidak selalu rating rendah. Sebagian besar produk yang ada di dataset

berharga relatif rendah tetapi ratingnya tinggi (4–5). Produk berharga tinggi jumlahnya sedikit, dan ratingnya tersebar dari rendah sampai tinggi jadi tidak ada

hubungan kuat antara harga dan rating. Hal ini bisa menunjukkan bahwa konsumen memberi rating lebih berdasarkan kualitas atau kepuasan, bukan harga.

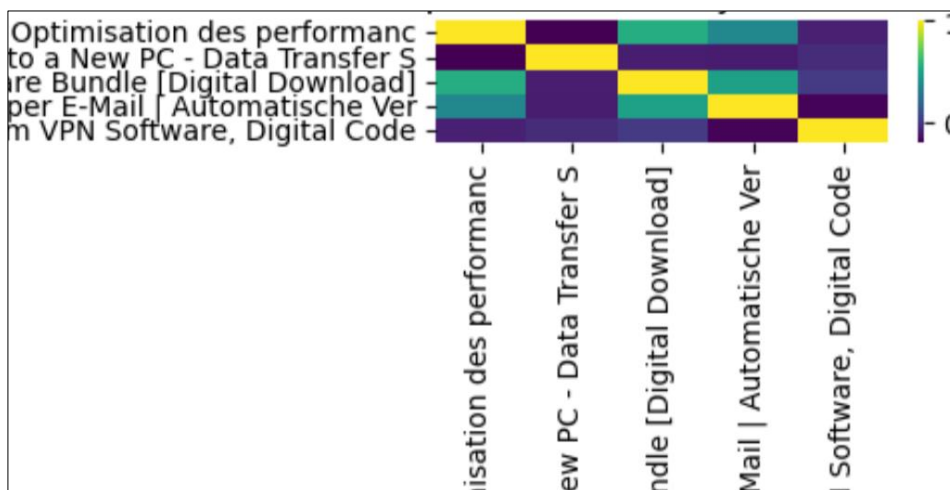


Gambar 5. Skor Kemiripan produk

Gambar 5 menampilkan visualisasi skor kemiripan (*Cosine Similarity*) antara produk acuan dan lima produk lain yang direkomendasikan. Produk acuan yang digunakan adalah *Kaspersky Premium - Total Security (Ultimate Security) | 1 Device | 1 Year | Email Delivery in 1 Hour*. Diagram batang horizontal menunjukkan Top-5 produk dengan tingkat kemiripan tertinggi berdasarkan analisis teks deskripsi produk.

Dari grafik terlihat bahwa produk yang paling mirip adalah *Kaspersky Total Security | 1 Device | 1 Year |*

Email Delivery in 1 Hour, diikuti oleh *Kaspersky Internet Security for Android Mobile*, *McAfee Antivirus 2024 (3 Year dan 1 Year)*, serta *Quick Heal Total Security*. Semua produk memiliki skor *cosine similarity* yang mendekati 1.0, yang berarti tingkat kemiripannya sangat tinggi terhadap produk acuan. Hal ini konsisten karena seluruh produk termasuk dalam kategori software antivirus/security dengan fitur dan jangka waktu langganan yang sejenis.



Gambar 6. Random 5 heatmap Cosine Similarity

Pada gambar 6 menunjukkan sumbu X dan sumbu Y masing-masing berisi nama-nama produk yang dipilih secara acak dari dataset (20 produk). Setiap sel pada kotak berwarna menunjukkan skor *cosine similarity* antara produk di baris tertentu dan produk di kolom tertentu. Warna menunjukkan tingkat kemiripan: Warna mendekati kuning/terang = kemiripan tinggi (skor mendekati 1). Warna mendekati hijau tua/gelap = kemiripan rendah (skor mendekati 0).

Diagonal utama (kiri atas ke kanan bawah) semuanya bernilai 1 (warna paling terang) karena setiap produk dibandingkan dengan dirinya sendiri. Karena datasetnya mencakup berbagai macam produk (software, VPN,

game, tools migrasi PC, dll.), terlihat banyak sel gelap ke produk-produk itu berbeda konten. Ada beberapa sel yang terang di luar diagonal utama dan menunjukkan ada pasangan produk yang judul, harga, atau ratingnya mirip (misalnya sama-sama software keamanan atau sama-sama alat migrasi PC).

IV. KESIMPULAN DAN SARAN

Perkembangan e-commerce yang semakin pesat menghadirkan tantangan bagi konsumen dalam memilih produk yang sesuai dengan kebutuhan dari ribuan bahkan jutaan pilihan yang tersedia. Dari latar belakang tersebut,

penelitian ini berangkat untuk menjawab permasalahan informasi berlebih (*information overload*) dengan membangun sebuah sistem rekomendasi produk berbasis Content-Based Filtering menggunakan Cosine Similarity. Hasil penelitian menunjukkan bahwa pendekatan ini efektif dalam menghubungkan konsumen dengan produk yang memiliki kemiripan konten, baik dari sisi teks (judul/deskripsi produk) maupun fitur numerik (harga, rating, jumlah rating). Tahap preprocessing seperti normalisasi harga, penghapusan simbol mata uang, serta pengolahan teks menggunakan TF-IDF berperan penting dalam menghasilkan data yang konsisten dan representatif.

Dari segi hasil, sistem rekomendasi mampu menampilkan produk-produk yang relevan terhadap produk acuan dengan tingkat kesamaan yang logis. Misalnya, produk antivirus dengan harga dan rating tertentu direkomendasikan bersama produk antivirus lain yang memiliki spesifikasi serupa. Analisis visualisasi data juga memperlihatkan pola menarik, seperti distribusi harga produk yang cenderung lebih dominan di kelas harga rendah, mayoritas rating yang tinggi menunjukkan kepuasan konsumen, serta lemahnya hubungan antara harga dan rating yang menandakan bahwa harga bukan satu-satunya indikator kualitas di mata pengguna. Hal ini memperkuat pandangan bahwa sistem rekomendasi tidak hanya membantu konsumen dalam menemukan produk yang sesuai, tetapi juga memberikan insight kepada pelaku e-commerce mengenai karakteristik produk yang dianggap mirip oleh sistem. Dengan demikian, penelitian ini menegaskan bahwa penerapan Content-Based Filtering berbasis Cosine Similarity merupakan solusi yang tepat dan efisien dalam menghadapi tantangan pemilihan produk di *platform e-commerce*.

Untuk penelitian selanjutnya, terdapat beberapa hal yang dapat dikembangkan agar sistem rekomendasi menjadi lebih optimal. Pertama, sistem dapat dikombinasikan dengan pendekatan Collaborative Filtering atau Hybrid Filtering yang memanfaatkan data interaksi pengguna, seperti riwayat pembelian atau klik, sehingga hasil rekomendasi menjadi lebih personal. Kedua, perlu dilakukan uji coba pada dataset yang lebih besar dan beragam agar performa sistem dapat dievaluasi pada skala yang lebih realistis seperti di *marketplace* besar (Shopee, Tokopedia, dsb.). Ketiga, integrasi algoritma pemrosesan bahasa alami (NLP) yang lebih canggih, misalnya Word2Vec, BERT, atau Sentence Transformers, dapat meningkatkan pemahaman konteks teks sehingga kemiripan produk dihitung lebih akurat. Selain itu, evaluasi kuantitatif menggunakan metrik seperti *Precision*, *Recall*, MAP (*Mean Average Precision*), dan NDCG (*Normalized Discounted Cumulative Gain*) dapat memberikan gambaran lebih objektif terkait kualitas rekomendasi yang dihasilkan. Terakhir, sistem ini dapat diimplementasikan langsung dalam bentuk aplikasi atau API sehingga dapat dimanfaatkan secara praktis oleh pelaku *e-commerce* maupun konsumen.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih yang sebesar-besarnya kepada Universitas Klabat atas dukungan yang diberikan dalam pelaksanaan penelitian ini. Dukungan berupa fasilitas, akses sumber daya, serta lingkungan akademik yang kondusif telah memberikan kontribusi penting bagi kelancaran proses penelitian hingga terselesaikannya karya ini. Tanpa adanya dorongan dan kesempatan yang diberikan oleh Universitas Klabat, penelitian ini tidak akan dapat terlaksana dengan baik.

DAFTAR PUSTAKA

- [1] M. I. Syarif, M. Hannum, and S. Wahyuni, "Potensi Perkembangan E-Commerce Dalam Menunjang Bisnis di Indonesia," *J. Comput. Digit. Bus.*, vol. 2, no. 1, pp. 11–14, 2023.
- [2] E. M. Asih, "Analisis pada Shopee sebagai e-commerce terpopuler di Indonesia," *J. Ekon. Bisnis Antart.*, vol. 2, no. 1, pp. 73–79, 2024.
- [3] Z. Alamin, R. Missouri, S. Sutriawan, F. Fathir, and K. Khairunnas, "Perkembangan e-commerce: analisis dominasi Shopee sebagai primadona marketplace di Indonesia," *J-ESA (Jurnal Ekon. Syariah)*, vol. 6, no. 2, pp. 120–131, 2023.
- [4] R. Sari, "Pengaruh Penggunaan Paylater TERHADAP perilaku impulse buying pengguna e-commerce di Indonesia," *J. Ris. Bisnis Dan Investasi*, vol. 7, no. 1, pp. 44–57, 2021.
- [5] D. V. N. Hasibuan and M. I. P. Nasution, "Penerapan Big Data dalam Pemasaran Digital: Studi Kasus pada Industri E-commerce di Indonesia," *J. Ilm. Nusantara*, vol. 1, no. 4, pp. 776–783, 2024.
- [6] W. Wulan et al., "Tinjauan Ancaman dan Risiko pada Sistem Keamanan Internet of Things, Berbasis Cloud Computing dalam Penggunaan E-Commerce dan Rencana Strategis," *J. Kewirausahaan dan Multi Talent.*, vol. 2, no. 2, pp. 126–137, 2024.
- [7] U. Javed, K. Shaukat, I. A. Hameed, F. Iqbal, T. M. Alam, and S. Luo, "A review of content-based and context-based recommendation systems," *Int. J. Emerg. Technol. Learn.*, vol. 16, no. 3, pp. 274–306, 2021.
- [8] L. Wu, X. He, X. Wang, K. Zhang, and M. Wang, "A survey on accuracy-oriented neural recommendation: From collaborative filtering to information-rich recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 5, pp. 4425–4445, 2022.
- [9] J. T. Hoe, K. W. Ng, T. Zhang, C. S. Chan, Y.-Z. Song, and T. Xiang, "One loss for all: Deep hashing with a single cosine similarity based learning objective," *Adv. Neural Inf. Process. Syst.*, vol. 34, pp. 24286–24298, 2021.
- [10] A. Apriani, H. Zakiyudin, and K. Marzuki, "Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF System Penerimaan Mahasiswa Baru pada Kampus Swasta," *J. Bumigora Inf. Technol.*, vol. 3, no. 1, pp. 19–27, 2021.
- [11] S. Sintia, S. Defit, and G. W. Nurcahyo, "Product codification accuracy with cosine similarity and weighted term frequency and inverse document frequency (TF-IDF)," *J. Appl. Eng. Technol. Sci.*, vol. 2, no. 2, pp. 14–21, 2021.
- [12] Y. Nagai and K. Katayama, "Multivariate curve resolution combined with estimation by cosine similarity mapping of analytical data," *Analyst*, vol. 146, no. 16, pp. 5045–5054, 2021.