

Content-Based Filtering Recommendation System for E-Commerce Products Using Sentence-BERT and Cosine Similarity

Muhammad Abbad Raid Alfath ^{1*}, Arif Nur Rohman ^{2*}

* Sistem Informasi, Universitas Amikom Yogyakarta
abbad2975@students.amikom.ac.id ¹, arifrahman@amikom.ac.id ²

Article Info

Article history:

Received 2026-05-03

Revised 2026-05-18

Accepted 2026-06-11

Keyword:

Content-Based Filtering, Cosine Similarity, Recommender System, Sentence-BERT, Watches.

ABSTRACT

The rapid growth of e-commerce catalogs complicates product discovery, particularly for items with complex technical specifications like luxury watches. Conventional keyword searches and TF-IDF methods often fail to capture underlying semantic relationships. The primary objective of this study is to propose a robust e-commerce recommendation system utilizing Content-Based Filtering enhanced with Sentence-BERT (SBERT) semantic embeddings. This study employs an experimental comparative research design. The methodology involves aggregating product attributes into descriptive sentences and pre-processing them to minimize representation bias. These sentences are transformed into high-dimensional embeddings using the lightweight all-MiniLM-L6-v2 SBERT model, with similarities calculated via the Cosine Similarity algorithm. The system's performance is comparatively evaluated against a baseline TF-IDF method. Main outcomes and experimental results across 20 testing scenarios demonstrate that SBERT significantly outperformed the baseline, achieving an average Precision@5 of 93.00%, a Recall@5 of 2.71%, and a highly efficient latency of 0.38 ms. In conclusion, SBERT provides a superior, scalable solution for recommending complex products by accurately capturing the semantic similarity of technical specifications and textual representations of visual characteristics. The approach inherently resolves the item cold-start problem, and its successful integration into a web application confirms its feasibility for real-time similarity computation in modern e-commerce platforms.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Ekosistem perdagangan digital saat ini mengalami pertumbuhan yang sangat pesat [1]. Beberapa kategori produk *e-commerce*, seperti jam tangan yang menjadi studi kasus pada penelitian ini, memiliki karakteristik data yang sangat teknis dan kompleks. Hal ini mencakup atribut spesifik seperti jenis penggerak, fitur komplikasi, hingga ketahanan air. Banyaknya variasi produk dan spesifikasi teknis ini seringkali menciptakan tantangan tersendiri bagi calon pembeli. Pengguna kerap kesulitan menemukan produk yang sesuai dengan preferensi mereka di antara ribuan pilihan yang ada, terutama jika mereka tidak mengetahui nama model atau seri tertentu secara spesifik [2]. Tantangan ini menuntut adanya sistem yang mampu menjembatani kebutuhan pengguna dengan spesifikasi produk secara lebih akurat dan efisien.

Untuk mengatasi tantangan penemuan produk tersebut, berbagai platform *e-commerce* telah menerapkan sistem pencarian dan rekomendasi. Pendekatan yang paling umum digunakan hingga saat ini adalah pencarian berdasarkan kata kunci atau pemfilteran kategori secara manual. Dalam konteks sistem rekomendasi otomatis, metode *Content-Based Filtering* klasik sering diimplementasikan menggunakan algoritma seperti TF-IDF (*Term Frequency-Inverse Document Frequency*). Metode ini bekerja dengan menghitung bobot frekuensi kata dalam deskripsi produk untuk menemukan item yang memiliki kemiripan tekstual dengan item yang sedang dilihat oleh pengguna [3].

Meskipun metode konvensional seperti TF-IDF cukup populer, pendekatan ini memiliki kelemahan dalam menangkap makna atau konteks sebuah deskripsi. TF-IDF hanya berfokus pada kecocokan leksikal (kata demi kata) dan

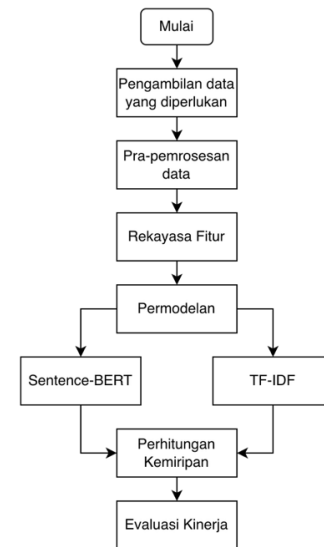
mengabaikan probabilitas hubungan makna antar kata[4]. Sebagai contoh pada spesifikasi produk *e-commerce* jam tangan, deskripsi seperti “*diver’s watch*” dan “*water resistance 300m*” memiliki korelasi makna yang sangat erat meskipun tidak mengandung irisan kata yang sama. Metode berbasis representasi frekuensi kata sering gagal mendeteksi relevansi laten ini, yang mengakibatkan hasil rekomendasi menjadi kurang akurat atau melewatkan produk yang sebenarnya sangat relevan berdasarkan spesifikasi teknis namun memiliki gaya penulisan deskripsi yang berbeda.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk membangun sistem rekomendasi jam tangan menggunakan metode *Content-Based Filtering* yang ditingkatkan dengan pendekatan *Deep Learning*, yaitu *Sentence-BERT (SBERT)*. Model SBERT varian *all-MiniLM-L6-v2* dipilih secara spesifik untuk mengubah deskripsi tekstual jam tangan menjadi representasi vektor (*embedding*) yang mampu menangkap konteks semantik secara mendalam, karena menawarkan keseimbangan optimal antara akurasi dan efisiensi komputasi untuk penerapan sistem secara *real-time*[5], [6]. Selain itu, pendekatan *Content-Based Filtering* menggunakan atribut teknis secara inheren memecahkan masalah *item cold-start*, di mana produk jam tangan rilis terbaru dapat langsung direkomendasikan tanpa harus menunggu data interaksi atau riwayat pembelian pengguna. Pemilihan arsitektur *all-MiniLM-L6-v2* juga didasari pada ukurannya yang ringan(kompak), memastikan skalabilitas yang tinggi apabila sistem diimplementasikan pada katalog *e-commerce* yang jauh lebih masif dengan jutaan produk. Tingkat kemiripan antar produk kemudian dihitung menggunakan metode *Cosine Similarity* untuk mengukur kedekatan sudut antar vektor tersebut. Berbeda dengan penelitian-penelitian sebelumnya yang umumnya menerapkan model SBERT pada domain dokumen akademik atau *chatbot* [7], [8], [6], naskah ini memberikan kontribusi dengan melakukan analisis komparatif langsung antara SBERT dan metode TF-IDF secara spesifik pada ranah rekomendasi produk *e-commerce* beratribut teknis tinggi. Melalui pendekatan ini, diharapkan sistem dapat memberikan rekomendasi yang tidak hanya didasarkan pada kesamaan kata, melainkan pada kesamaan karakteristik dan spesifikasi teknis yang sesungguhnya.

II. METODE

Metode penelitian ini dirancang secara sistematis agar dapat direplikasi secara ilmiah. Alur penelitian ini diawali dengan tahap pengumpulan data yang menjadi landasan utama analisis. Proses ini dilakukan dengan mengekstraksi dataset sekunder berjudul “*Luxury Watch Price Dataset*” dari repositori publik Kaggle. Secara kuantitatif, data yang dikumpulkan terdiri dari 507 entri produk jam tangan unik yang mencakup 39 merek(brands) global terkemuka. Distribusi data menunjukkan variasi yang representatif untuk kategori produk jam tangan, di mana dataset didominasi oleh merek-merek utama seperti IWC dengan 39 produk, Audemars Piguet dengan 38 produk, dan Patek Philippe

dengan 33 produk. Pada tahap ini, dilakukan pemilihan atribut kunci yang diperlukan untuk sistem rekomendasi, meliputi *Brand, Model, Case Material, Strap Material, dan Movement Type*. Penggunaan dataset dengan variasi merek dan ukuran yang memadai ini bertujuan untuk memastikan model mampu menggeneralisasi rekomendasi pada katalog *e-commerce* berskala menengah hingga besar, sebagaimana divisualisasikan pada alur penelitian di Gambar 1[7].



Gambar 1. Alur metode penelitian

Setelah tahap pengumpulan data selesai, penelitian dilanjutkan ke tahap pra-pemrosesan data. Tahap ini mencakup penyesuaian format kolom, imputasi nilai hilang (*missing values*), dan pembersihan karakter guna menjaga integritas data serta meningkatkan akurasi pembentukan pola *text mining*[9], [10]. Setelah data dipastikan bersih, dilakukan tahap rekayasa fitur (*feature engineering*) dengan menggabungkan atribut-atribut spesifikasi yang terpisah menjadi satu dokumen kalimat deskriptif (*rich semantic sentence*). Strategi penggabungan fitur ini bertujuan untuk memperkaya konteks informasi yang akan diproses, sehingga sistem tidak hanya mengenali kunci parsial tetapi juga menyentuh secara konteks semantik[9].

Masuk pada tahap permodelan, dokumen teks hasil rekayasa fitur diproses melalui dua jalur representasi vektor secara paralel untuk dibandingkan efektivitasnya. Jalur pertama menggunakan pendekatan *Proposed Embedding* berbasis *Sentence-BERT(SBERT)* varian *all-MiniLM-L6-v2*. Model ini memanfaatkan arsitektur *Deep Learning transformer* yang terbukti efektif menangkap hubungan makna antar kata dibandingkan metode statistik konvensional[4]. Proses matematis pembentukan vektor melalui fungsi model tersebut ditunjukkan pada Persamaan (1):

$$e_i = f_{SBERT}(T_i), ||e_i||_2 = 1 \quad (1)$$

Pada Persamaan (1), T_i merepresentasikan teks deskripsi jam tangan ke- i , sedangkan f_{SBERT} merupakan fungsi

embedding dari model SBERT yang menghasilkan vektor e_i pada ruang dimensi $d = 384$. Proses normalisasi dilakukan sehingga panjang vektor memenuhi $\|e_i\|_2 = 1$ guna menyeragamkan skala vektor agar perhitungan tidak dipengaruhi oleh besaran nilai absolut.

Secara paralel, jalur kedua menerapkan pendekatan *Baseline Embedding* berbasis algoritma TF-IDF untuk mengekstraksi bobot statistik berdasarkan frekuensi kemunculan kata (leksikal)[11]. Nilai bobot representasi vektor TF-IDF untuk setiap kata (t) dalam dokumen (d) dihitung secara matematis menggunakan Persamaan (2):

$$W_{t,d} = tf_{t,d} \times \log\left(\frac{N}{df_t}\right) \quad (2)$$

Pada Persamaan (2), $W_{t,d}$ merupakan bobot embedding TF-IDF akhir, $tf_{t,d}$ merepresentasikan *Term Frequency* (frekuensi kemunculan kata t dalam dokumen d), N adalah total jumlah keseluruhan dokumen dalam dataset, dan df_t merupakan *Document Frequency* (jumlah dokumen yang mengandung kata t tersebut).

Selanjutnya, pada tahap pengujian rekomendasi, tingkat kemiripan antar produk dari kedua jenis *embedding* tersebut dihitung menggunakan perhitungan *Cosine Similarity* sebagaimana ditunjukkan pada Persamaan (3), yang mengukur kedekatan sudut antar dua vektor untuk menghasilkan rekomendasi berdasarkan nilai kemiripan tertinggi:

$$\text{Cosine Similarity}(A, B) = \frac{A \cdot B}{\|A\| \|B\|} \quad (3)$$

Nilai Similaritas yang dihasilkan berkisar antara 0 hingga 1, di mana nilai yang mendekati 1 mengindikasikan bahwa kedua produk memiliki korelasi spesifikasi teknis dan karakteristik material fisik yang sangat identik berdasarkan representasi semantik tekstualnya.

Tahap akhir penelitian adalah evaluasi kinerja komparatif menggunakan 20 skenario produk acuan untuk menghindari bias pengujian. Kinerja sistem diukur menggunakan 3 metrik utama: *Precision*, *Recall*, dan latensi sistem, untuk memastikan efisiensi pada lingkungan aplikasi nyata. Mengingat ketiadaan data interaksi pengguna eksplisit, *Ground Truth* (relevansi) didefinisikan secara heuristik: sebuah rekomendasi dianggap relevan jika produk hasil sistem memiliki tipe mesin (*movement*) yang sama serta kesamaan pada salah satu material komponen (*case* atau *strap*) dengan produk acuan. *Precision@K* dihitung berdasarkan rasio jumlah rekomendasi yang relevan (*True Positive*) terhadap total rekomendasi yang diberikan sesuai dengan Persamaan (4)[12]:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (4)$$

Dalam konteks pengujian sistem rekomendasi ini, nilai *True Positive* mewakili jumlah rekomendasi produk yang

relevan. Sementara itu, total dari *True Positive* dan *False Positive* merupakan total keseluruhan produk yang direkomendasikan kepada pengguna. Karena sistem membatasi batas rekomendasi sebanyak 5 produk teratas ($K = 5$), maka total penyebut (TP + FP) adalah 5. Oleh karena itu, metrik ini secara spesifik dievaluasi sebagai *Precision@5*.

Selain *Precision@5*, penelitian ini juga mengukur *Recall@5* untuk mengevaluasi kemampuan sistem dalam memanggil kembali seluruh item relevan yang tersedia di dalam dataset. Kinerja *Recall* dihitung menggunakan Persamaan (5)[12]:

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (5)$$

Pada Persamaan (5), nilai *True Positive* mewakili item relevan yang berhasil ditemukan pada 5 rekomendasi teratas, sedangkan *False Negative* mewakili sisa item relevan di dalam dataset yang tidak masuk ke dalam daftar rekomendasi tersebut. Dengan kata lain, penjumlahan antara *True Positive* dan *False Negative* merepresentasikan total keseluruhan item relevan yang ada di dalam database untuk produk acuan tersebut.

Nilai *Precision* dan *Recall* yang tinggi, ditambah dengan pengukuran latensi waktu respons dalam milidetik, memberikan gambaran komprehensif mengenai perbandingan efektivitas dan efisiensi model dalam menyaring informasi yang tidak relevan serta menyajikan rekomendasi produk yang akurat secara *real-time*.

III. HASIL DAN PEMBAHASAN

Berdasarkan alur penelitian yang telah dirancang, tahap awal dimulai dengan pengambilan data yang diperlukan. Sumber data utama dalam penelitian ini adalah dataset dengan judul "*Luxury Watch Price Dataset*" yang diperoleh dari repositori publik Kaggle. Dataset ini mencakup berbagai atribut spesifikasi teknis seperti *Brand*, *Model*, *Case Material*, *Strap Material*, *Movement Type*, *Water Resistance*, *Case Diameter*, *Case Thickness*, *Band Width*, *Dial Color*, *Crystal Material*, *Complications*, dan *Power Reserve* yang merepresentasikan karakteristik fisik dan fungsional produk[3]. Penggunaan dataset dengan variasi merek dan ukuran yang memadai ini bertujuan untuk memastikan model mampu menggeneralisasi rekomendasi pada katalog *e-commerce* berskala menengah hingga besar [13].

Data mentah ini kemudian diproses melalui tahap pra-pemrosesan data (*pre-processing*) untuk menjamin kualitas input sebelum diolah lebih lanjut. Tahap ini, mencakup *case folding* (konversi seluruh teks menjadi huruf kecil) serta penghapusan karakter non-alfanumerik dan *stopwords* untuk mengurangi *noise* pada teks. *Pre-processing* terbukti berpengaruh signifikan terhadap kualitas representasi *embedding* dan performa sistem berbasis kemiripan semantik[14]. Contoh perbandingan data sebelum dan sesudah pra-pemrosesan data dapat dilihat pada Tabel I.

TABEL I
SAMPSEL DATA SEBELUM DAN SESUDAH PRA-PROCESSING

Atribut	Data Mentah (sebelum)	Data Bersih (sesudah)
Brand	Rolex	rolex
Model	Submariner	submariner
Case Material	Stainless Steel	stainless steel
Strap Material	Stainless Steel	stainless steel
Movement Type	Automatic	automatic
Water Resistance	300 meters	300 meters
Case Diameter(mm)	40.0	40.0
Case Thickness(mm)	13.0	13.0
Band Width(mm)	20.0	20.0
Dial Color	Black	black
Crystal Material	Sapphire	sapphire
Complications	Date	date
Power Reserve	48 hours	48 hours
Price	9,500	9,500

Tabel I menyajikan perbandingan perbandingan visual antara data mentah (*raw data*) dengan data yang telah melalui tahapan pra-pemrosesan. Pada kolom "Data Mentah", terlihat bahwa atribut spesifikasi jam tangan memiliki format penulisan yang bervariasi, termasuk penggunaan huruf kapital yang tidak konsisten. Sebaliknya, pada kolom "Data Bersih", seluruh atribut teks yang dinormalisasikan sehingga perbedaan format tidak lagi diperlakukan sebagai entitas yang berbeda oleh model *embedding*. Normalisasi ini penting karena model berbasis *transformer* yang sensitif terhadap variasi token yang berpotensi memecah makna semantik.

Setelah data dibersihkan, proses dilanjutkan ke tahap rekayasa fitur (*feature engineering*). Tahap ini bertujuan untuk memperkaya konteks informasi yang akan dipelajari oleh model dengan menggabungkan atribut-atribut spesifikasi yang terpisah menjadi satu kalimat deskriptif utuh (*rich semantic sentence*)[15]. Pendekatan ini bertujuan untuk mempertahankan hubungan kontekstual antar-atribut sehingga model dapat memahami karakter produk secara holistik, misalnya hubungan antara "Rubber Strap" dan "300m Water Resistance" sebagai indikator jam tangan penyelam[9]. Dalam proses transformasi menjadi deskripsi kalimat, sistem menangani ketiadaan atribut (*missing values*) dengan menghilangkannya secara dinamis dari struktur kalimat agar tidak memicu bias representasi (*representation bias*) pada ruang vektor. Strategi penggabungan atribut terbukti meningkatkan performa sistem rekomendasi berbasis konten karena memperkaya informasi semantik yang dipelajari model[8], [13].

Dokumen teks ini kemudian dikonversi menjadi representasi vektor (*embedding*) menggunakan model *Sentence-BERT* dengan arsitektur *all-MiniLM-L6-v2*. Model ini memetakan setiap jam tangan ke dalam ruang vektor berdimensi 384. *Sentence-BERT* dirancang untuk menghasilkan *embedding* kalimat yang optimal dalam pengukuran kesamaan semantik menggunakan metrik berbasis jarak vektor, seperti *Cosine Similarity*[6]. Untuk

mengukur tingkat kemiripan antar produk, digunakan metrik *Cosine Similarity* karena mampu mengukur kedekatan arah vektor tanpa dipengaruhi oleh besar magnitudo, sehingga sesuai untuk membandingkan *embedding* hasil *transformer*[16]. Metrik ini umum digunakan pada sistem rekomendasi berbasis *embedding* karena efisien dan stabil pada ruang vektor berdimensi tinggi[17].

Pengujian dilakukan secara spesifik terhadap lima skenario produk acuan yang mewakili berbagai kategori gaya jam tangan. Berikut adalah paparan hasil pengujian untuk masing-masing skenario:

TABEL II
HASIL PENGUJIAN PADA TISSOT LE LOCLE

Peringkat	Brand dan Model	Skor	Status
1	Tissot Le Locle	0.9947	Relevan
2	Breguet Classique	0.8782	Relevan
3	Breguet Classique	0.8759	Relevan
4	Breguet Classique	0.8758	Relevan
5	Breguet Classique	0.8757	Relevan

Pada Tabel II menunjukkan hasil pengujian dengan acuan produk "Tissot Le Locle", sistem menunjukkan kemampuan generalisasi yang baik dengan merekomendasikan "Breguet Classique" pada peringkat kedua hingga kelima. Tinjauan terhadap atribut spesifikasi memperlihatkan bahwa meskipun Tissot Le Locle menggunakan material *Stainless Steel* dan Breguet Classique menggunakan *18k White Gold*, algoritma memberikan skor kemiripan yang signifikan (0.8782). Hal ini mengindikasikan bahwa model mampu mengasosiasikan karakteristik visual *silver-tone* dari kedua material tersebut sebagai fitur yang setara dalam konteks estetika. Lebih lanjut, kesamaan pada elemen desain fundamental seperti *dial* berwarna terang dan penggunaan *leather strap* diklasifikasikan oleh sistem ke dalam satu kluster semantik "jam tangan formal". Temuan ini membuktikan bahwa sistem beroperasi secara agnostik terhadap tingkatan harga (*price-agnostik*), memprioritaskan kesesuaian gaya (*style alignment*) di atas segmentasi pasar[8].

TABEL III
HASIL PENGUJIAN PADA ROLEX SUBMARINER

Peringkat	Brand dan Model	Skor	Status
1	Rolex Submariner	1.0000	Relevan
2	Sinn U50	0.9035	Relevan
3	Rolex Datejust	0.9008	Relevan
4	Rolex Datejust	0.8998	Relevan
5	Rolex Datejust	0.8987	Relevan

Hasil pengujian pada skenario "Rolex Submariner" memvalidasi kapasitas sistem dalam mengidentifikasi atribut fungsional produk. Rekomendasi "Sinn U50" dengan skor 0.9035 yang muncul di urutan kedua memiliki relevansi kontekstual yang tinggi. Analisis komparatif menunjukkan bahwa deskripsi fitur "German Submarine Steel" dan ketahanan air "500 meters" pada Sinn U50 memiliki pendekatan vektor dengan deskripsi "Oystersteel" dan "300 meters" pada Rolex Submariner. Kedua entitas tersebut

dipetakan sebagai *professional tool watch* dengan spesifikasi durabilitas tinggi. Kemunculan rekomendasi ini menegaskan bahwa model *Sentence-BERT* berhasil mengekstraksi nuansa ketangguhan (*ruggedness*) dan utilitas spesifik dari teks deskripsi, sehingga rekomendasi yang dihasilkan sangat presisi bagi pengguna yang berorientasi pada performa teknis.

TABEL IV
HASIL PENGUJIAN PADA SEIKO PROSPEX

Peringkat	Brand dan Model	Skor	Status
1	Seiko Prospex	0.9806	Relevan
2	Seiko Prospex	0.9704	Relevan
3	Seiko Prospex	0.9564	Relevan
4	Seiko Prospex	0.9543	Relevan
5	Bell & Ross Br 03-92	0.9186	Relevan

Pada kategori *sport watch*, sistem mendemonstrasikan kemampuan rekomendasi lintas merek (*cross-brand recommendation*) yang efektif. Selain merekomendasikan varian internal Seiko, sistem mengidentifikasi “Bell & Ross Br 03-92” dengan skor 0.9186 sebagai produk yang relevan. Meskipun Bell & Ross secara tradisional diasosiasikan dengan aviasi, rekomendasi ini muncul akibat adanya interseksi fitur fisik yang signifikan, yakni penggunaan *Black Dial*, *Stainless Steel Case* yang masif, dan *Rubber Strap*. Kombinasi atribut ini diterjemahkan oleh model sebagai karakteristik jam tangan instrument yang maskulin. Hal ini menunjukkan bahwa sistem tidak terikat pada klasifikasi kategori konvensional, melainkan memanfaatkan kesamaan berbasis materialitas dan desain produk[17].

TABEL V
HASIL PENGUJIAN PADA BREITLING NAVITIMER

Peringkat	Brand dan Model	Skor	Status
1	Breitling Navitimer	1.0000	Relevan
2	Breitling Avenger II	0.8968	Relevan
3	Vacheron Constantin Overseas	0.8825	Relevan
4	Breguet Type Xx	0.8823	Relevan
5	Patek Philippe Nautilus	0.8814	Relevan

Hasil pengujian pada “Breitling Navitimer” menyoroti sensitivitas sistem terhadap fitur komplikasi (*complications*). Hasil rekomendasi yang mencakup “Breitling Avenger II” dan “Breguet Type XX” didorong oleh keberadaan fitur *Chronograph* yang dominan pada deskripsi produk. Dalam ruang vektor semantik, terminologi teknis seperti “Chronograph” memiliki bobot informasi tinggi dalam ruang vektor sehingga menjadi faktor dominan dalam pengelompokan produk. Hal ini sejalan dengan temuan bahwa fitur teknis eksplisit memiliki kontribusi besar terhadap kedekatan *embedding* semantik[6].

TABEL VI
HASIL PENGUJIAN PADA LONGINES MASTER COLLECTION

Peringkat	Brand dan Model	Skor	Status
1	Longines Master Collection	0.9997	Relevan
2	Longines Master Collection	0.9903	Relevan
3	Longines Master Collection	0.9842	Relevan
4	Longines Master Collection	0.9812	Relevan
5	Longines Conquest	0.9493	Relevan

Pengujian pada “Longines Master Collection” merefleksikan stabilitas (*robustness*) model yang tinggi dalam menangani variasi produk internal. Skor kemiripan yang konsisten di atas 0.98 untuk varian dalam seri yang sama menunjukkan bahwa sistem mampu mendeteksi pola deskriptif naratif yang seragam. Skor kemiripan yang tinggi antar varian menunjukkan bahwa sistem mampu mempertahankan koherensi rekomendasi dalam satu keluarga produk meskipun terdapat variasi minor pada ukuran atau komplikasi tambahan.

Setelah analisis per skenario, dilakukan evaluasi kuantitatif kinerja sistem secara keseluruhan menggunakan metrik *Precision at K (P@5)* dengan ambang batas (*threshold*) relevansi 0.5. Rekapitulasi hasil evaluasi disajikan pada Tabel VII.

TABEL VII
HASIL EVALUASI DENGAN PRECISION

Produk	K	True Positive	False Positive	Precision
Tissot Le Locle	5	5	0	100%
Rolex Submariner	5	5	0	100%
Seiko Prospex	5	5	0	100%
Breitling Navitimer	5	5	0	100%
Longines Master Collection	5	5	0	100%

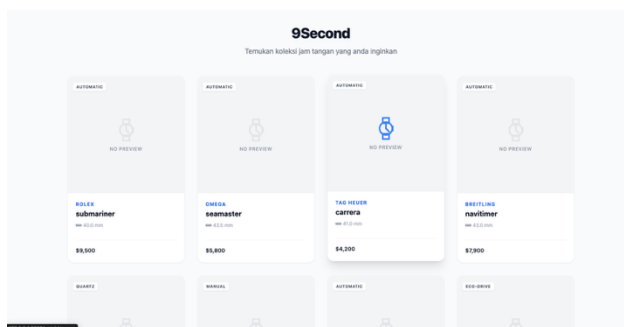
Tabel VII mengilustrasikan kinerja presisi sistem yang mencapai angka sempurna (100%) di lima skenario sampel tersebut. Skenario awal ini sengaja dipilih untuk mendemonstrasikan batas atas (*upper bound*) kemampuan model dalam mengenali kluster gaya yang sangat kontras (misalnya antara *Diver* murni dan *Dress* murni). Namun, untuk menghindari indikasi *overfitting* dan menguji keandalan sistem pada variasi kueri yang lebih kompleks, evaluasi kuantitatif kemudian diperluas menjadi 20 skenario pengujian acak dan dibandingkan secara langsung dengan pendekatan *Baseline Embedding* berbasis algoritma TF-IDF[11]. Perbandingan kinerja diukur berdasarkan parameter *Precision@5*, *Recall@5*, dan rata-rata latensi komputasi sebagaimana disajikan pada Tabel VIII.

TABEL VIII
PERBANDINGAN KINERJA EMBEDDING SBERT DAN TF-IDF

Metrik Evaluasi	Baseline Embedding TF-IDF	Proposed Embedding SBERT
Rata-rata <i>Precision@5</i>	89,00%	93,00%
Rata-rata <i>Recall@5</i>	2,56%	2,71%
Rata-rata Latensi	2,51 ms	0,38 ms

Berdasarkan Tabel VIII, hasil perbandingan menunjukkan bahwa *embedding* SBERT secara konsisten mengungguli TF-IDF pada seluruh metrik pengujian. SBERT mencapai rata-rata *Precision@5* sebesar 93,00% untuk 20 skenario, membuktikan keunggulannya dalam memahami korelasi material dan spesifikasi teknis jam tangan secara semantik dibandingkan TF-IDF yang hanya mengandalkan kecocokan leksikal[4]. Dari aspek daya panggil, SBERT mencatatkan skor *Recall@5* sebesar 2,71%, sedikit lebih tinggi dari TF-IDF. Angka recall yang tampak rendah pada kedua model merupakan implikasi matematis alami dari Batasan evaluasi *Top-K*, di mana sistem hanya diizinkan menampilkan 5 item terbaik dari total seluruh item yang tersedia di dalam dataset [18]. Poin penemuan paling signifikan dalam evaluasi ini terletak pada metrik efisiensi sistem. *Embedding* SBERT mencatatkan rata-rata latensi sebesar 0,38 ms, jauh lebih cepat dan efisien dibandingkan metode TF-IDF yang membutuhkan 2,51 ms. Hasil ini mengonfirmasi bahwa pendekatan *Content-Based Filtering* menggunakan representasi semantik SBERT tidak hanya unggul dari segi akurasi, tetapi juga sangat optimal untuk diimplementasikan pada kebutuhan aplikasi *e-commerce real-time* berskala menengah hingga besar [13].

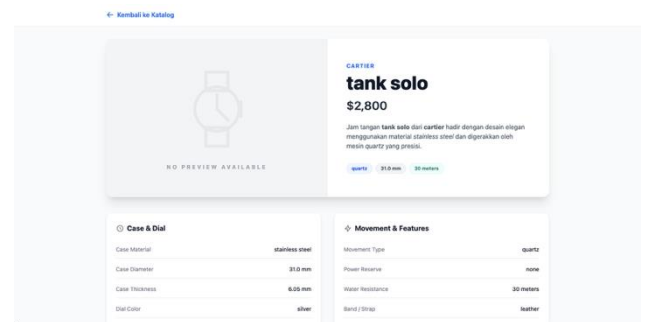
Sebagai tahap akhir validasi, sistem diimplementasikan ke dalam aplikasi berbasis *web* menggunakan *framework* Laravel. Antarmuka halaman utama ditampilkan pada Gambar 2.



Gambar 2. Halaman utama sistem rekomendasi jam tangan "9Second"

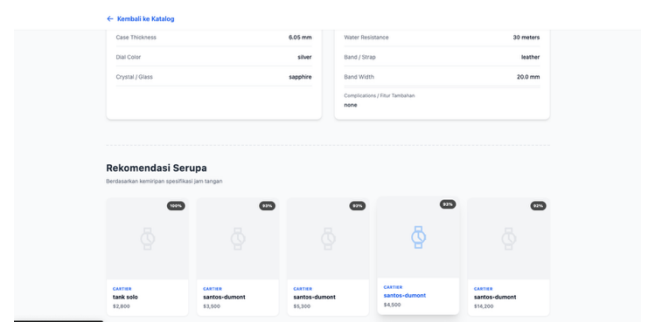
Gambar 2 memperlihatkan antarmuka utama sistem (*landing page*) yang berfungsi sebagai katalog produk. Halaman ini menerapkan tata letak *grid* untuk menampilkan koleksi jam tangan secara terstruktur. Setiap kartu produk (*product card*) memuat informasi visual berupa foto jam tangan beserta atribut identitas utama seperti merek, model,

dan harga dari jam tangan yang ditampilkan. Namun, karena ketidaktersediaan gambar jam tangan dalam dataset yang digunakan, *Card* menampilkan *placeholder* yang menggantikan gambar jam tangan yang tidak tersedia.



Gambar 3. Halaman detail jam tangan

Gambar 3 menyajikan antarmuka halaman detail produk yang akan diakses ketika pengguna memilih salah satu item dari katalog. Pada halaman ini, sistem secara komprehensif menampilkan penjabaran spesifikasi teknis produk yang sedang ditinjau. Informasi yang disajikan mencakup berbagai atribut fundamental, mulai dari material *case*, jenis *strap*, tipe mesin penggerak (*movement*), hingga tingkat ketahanan air (*water resistance*). Penyajian data spesifikasi yang mendetail ini memegang peranan krusial sebagai bentuk transparansi informasi (*information transparency*). Hal ini bertujuan untuk memberikan konteks yang jelas dan edukatif bagi pengguna mengenai karakteristik esensial yang melekat pada produk tersebut. Pemahaman atribut ini sangat penting karena data spesifikasi inilah yang selanjutnya akan menjadi landasan parameter bagi sistem dalam menjalankan proses komputasi pencarian kemiripan produk.



Gambar 4. Halaman detail jam tangan lanjutan

Lebih lanjut, Gambar 4 memvisualisasikan luaran utama dari komputasi algoritma *Content-Based Filtering* yang diimplementasikan pada sistem, yang letaknya terintegrasi di bagian bawah halaman detail produk. Pada antarmuka ini, sistem menghadirkan sebuah seksi khusus dengan judul "Rekomendasi Serupa". Seksi ini bertugas untuk menyajikan daftar produk alternatif yang telah dikalkulasi dan diurutkan secara dinamis berdasarkan tingkat relevansi semantik tertinggi terhadap produk utama yang sedang dilihat. Salah satu elemen antarmuka (*User Interface*) kunci yang

ditunjukkan pada bagian ini adalah kehadiran label persentase kemiripan (*similarity score*) yang disematkan secara presisi pada setiap kartu rekomendasi. Visibilitas metrik kemiripan ini tidak hanya berfungsi untuk memvalidasi akurasi operasional dari model *Sentence-BERT* dan *Cosine Similarity* di latar belakang, tetapi juga memberikan justifikasi rasional yang sangat membantu pengguna dalam mengambil keputusan perbandingan antar produk.

IV. KESIMPULAN

Penelitian ini menyimpulkan bahwa penerapan metode *Content-Based Filtering* yang dioptimasi menggunakan model *Sentence-BERT* (SBERT) terbukti sangat efektif dalam membangun sistem rekomendasi produk *e-commerce*, dengan jam tangan sebagai studi kasus aplikatifnya. Pendekatan berbasis representasi representasi vektor semantik ini berhasil memecahkan keterbatasan metode pencocokan kata kunci konvensional. Sistem terbukti mampu menangkap konteks informasi secara mendalam seperti korelasi spesifikasi teknis dan karakteristik material sehingga tingkat kemiripan antar produk dapat diidentifikasi secara lebih akurat dan representatif. Melalui pengujian komparatif, keunggulan komputasi *embedding* ini sejalan dengan berbagai literatur terdahulu yang membuktikan bahwa arsitektur berbasis *transformer* memberikan kualitas rekomendasi yang lebih superior dibandingkan metode statistik leksikal tradisional (*baseline*) seperti TF-IDF, terutama pada domain data yang memiliki atribut teknis kompleks[4], [11].

Evaluasi kinerja sistem secara kuantitatif menunjukkan bahwa SBERT secara konsisten mengungguli TF-IDF. Sistem SBERT mencatatkan *Precision@5* sebesar 93,00% dan *Recall@5* sebesar 2,71%. Pencapaian ini mengindikasikan kapabilitas model dalam memetakan produk ke dalam klaster yang relevan sekaligus secara signifikan meminimalkan kemunculan rekomendasi yang tidak sejalan dengan intensi kueri pengguna[15], [19]. Di samping itu, SBERT membuktikan efisiensi yang sangat tinggi dengan rata-rata latensi hanya 0,38 ms. Keberhasilan implementasi komputasi latensi rendah pada aplikasi berbasis *web* ini membuktikan bahwa integrasi SBERT dan *Cosine Similarity* sangat layak (*feasible*) diterapkan pada arsitektur sistem informasi nyata untuk menyajikan rekomendasi yang *real-time*. Keterbukaan informasi melalui penyajian metrik persentase kemiripan pada antarmuka aplikasi juga secara efektif mendukung aspek keterjelasan rekomendasi (*explainability*), yang merupakan faktor krusial dalam membangun kepercayaan pengguna terhadap luaran sistem.

Meskipun model yang dikembangkan telah menunjukkan performa yang optimal dalam pencocokan spesifikasi produk secara teknis, penelitian ini menyadari adanya Batasan berupa ketiadaan pengujian berbasis pengguna nyata (*user study*). Pengembangan lebih lanjut masih sangat terbuka untuk meningkatkan aspek personalisasi dan skalabilitas. Penelitian di masa depan direkomendasikan untuk beralih pada pendekatan *Hybrid Filtering* atau *Collaborative Filtering*,

yaitu dengan mengintegrasikan data interaksi dinamis pengguna seperti riwayat penelusuran atau metrik *rating* guna melengkapi batasan analisis atribut produk yang saat ini masih bersifat statis [20]. Selain itu, perluasan volume dataset dengan variasi merek yang lebih masif, serta penambahan fitur penyaringan (*filtering*) parameter secara manual pada antarmuka sistem, akan memberikan kendali interaktif yang lebih besar bagi pengguna. Optimalisasi tersebut diharapkan mampu menjadikan platform sistem rekomendasi ini semakin adaptif dan komprehensif dalam mengakomodasi kompleksitas preferensi konsumen modern.

DAFTAR PUSTAKA

- [1] S. I. Adam dan W. Grivin, "Implementasi Sistem Rekomendasi Produk E-Commerce Menggunakan Content-Based Filtering Berbasis Cosine Similarity," *Jurnal Sistem Informasi dan Teknik Komputer*, vol. 10, no. 2, 2025.
- [2] S. Rahmadhani dkk., "Sistem Rekomendasi Penelusuran Buku Berbasis Content-Based Filtering dengan Pembobotan TF-RF," *JIP (Jurnal Informatika Polinema)*, 2024.
- [3] A. Nurcahya dan S. Supriyanto, "Content-based recommender system architecture for similar e-commerce products," *Jurnal Informatika*, vol. 14, no. 3, hlm. 90, Sep 2020, doi: 10.26555/jifo.v14i3.a18511.
- [4] M. P. Syah, A. P. Wardani, M. Idhom, dan Trimono, "Perbandingan Representasi Teks TF-IDF Dan BERT Terhadap Akurasi Cosine Similarity Dalam Penilaian Otomatis Jawaban Berbasis Teks," *Data Sciences Indonesia (DSI)*, vol. 5, no. 1, hlm. 47–59, Jul 2025, doi: 10.47709/dsi.v5i1.6021.
- [5] M. D. S. Antariksa, A. Sugiharto, dan B. Surarso, "BERT Model Fine-tuned for Scientific Document Classification and Recommendation," *Jurnal RESTI*, vol. 9, no. 4, hlm. 754–764, Agu 2025, doi: 10.29207/resti.v9i4.6789.
- [6] R. M. Holis, P. E. P. Utomo, dan B. F. Hutabarat, "Semantic FAQ Chatbot Using SBERT (Sentence-BERT) and Cosine Similarity for Academic Services," *Brilliance: Research of Artificial Intelligence*, vol. 5, no. 2, hlm. 915–922, Okt 2025, doi: 10.47709/brilliance.v5i2.7027.
- [7] F. T. Sabillillah, S. Winarno, dan R. B. Abiyi, "Implementasi BERT dan Cosine Similarity untuk Rekomendasi Dosen Pembimbing berdasarkan Judul Tugas Akhir," *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 2, hlm. 585–594, Des 2024, doi: 10.29408/edumatic.v8i2.27791.
- [8] A. P. Putra, D. P. S. Putri, dan A. C. Wiranatha, "Scientific Paper Recommendation System: Application of Sentence Transformers and Cosine Similarity Using arXiv Data," *Journal of Applied Informatics and Computing (JAIC)*, vol. 9, no. 4, 2025.
- [9] M. A. Palomino dan F. Aider, "Evaluating the Effectiveness of Text Pre-Processing in Sentiment Analysis," *Applied Sciences (Switzerland)*, vol. 12, no. 17, Sep 2022, doi: 10.3390/app12178765.
- [10] C. Y. Ariyanto dan A. S. Aji, "Pengembangan Sistem Deteksi Plagiarisme Dokumen Jurnal Berbasis Bidirectional Encoder Representations from Transformers Dan Cosine Similarity," *Jurnal Riset Komputer*, vol. 12, no. 6, hlm. 942–948, 2025, doi: 10.30865/jurikom.v12i6.9325.
- [11] M. Z. F. Johari dan A. D. Laksito, "The Hybrid Recommender System of the Indonesian Online Market Products using IMDb weight rating and TF-IDF," *Jurnal RESTI*, vol. 5, no. 5, hlm. 977–983, Okt 2021, doi: 10.29207/resti.v5i5.3486.
- [12] R. Alana dan A. Hartanto, "Implementasi Algoritma Content Based Filtering dalam Sistem Rekomendasi Komik," *Sistemasi: Jurnal Sistem Informasi*, vol. 13, no. 4, hlm. 1344–1355, 2024, [Daring]. Tersedia pada: <http://sistemasi.ftik.unisi.ac.id>
- [13] M. W. Putri, A. Muchayan, dan M. Kamisutara, "Sistem Rekomendasi Produk Pena Eksklusif Menggunakan Metode

- Content-Based Filtering dan TF-IDF,” *JOINTECS (Journal of Information Technology and Computer Science)*, vol. 3, no. 1, hlm. 229–236, 2018.
- [14] A. Setiawan, Z. Abidin, dan M. Imamudin, “Impact of Preprocessing on Indonesian Extractive Summarization Using LexRank, TextRank, DivRank, and Cosine Similarity,” *G-Tech: Jurnal Teknologi Terapan*, vol. 9, no. 4, hlm. 2311–2321, Okt 2025, doi: 10.70609/g-tech.v9i4.8306.
- [15] H. Hartatik dan A. Syafrianto, “Penerapan Model Sentence-BERT Untuk Sistem Rekomendasi Buku Berbasis Konten Di Perpustakaan Digital,” *Jurnal Dialektika Informatika (Detika)*, vol. 6, no. 1, hlm. 12–19, Nov 2025, doi: 10.24176/detika.v6i1.15916.
- [16] M. F. Abdurrafi, D. Handayani, dan U. Ningsih, “Content-based filtering using cosine similarity algorithm for alternative selection on training programs,” *Journal of Soft Computing Exploration*, 2023, doi: 10.52465/josce.v4i3.232.
- [17] X. Yin, W. Zhang, W. Zhu, S. Liu, dan T. Yao, “Improving sentence representations via component focusing,” *Applied Sciences (Switzerland)*, vol. 10, no. 3, Feb 2020, doi: 10.3390/app10030958.
- [18] N. Azizah dan A. F. Rozi, “Sistem Rekomendasi Produk Something Menggunakan Metode Content-based Filtering,” *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 6, no. 3, hlm. 461–468, Jul 2024, doi: 10.47233/jteksis.v6i3.1411.
- [19] M. P. Nabila, J. D. Irawan, dan A. Faisal, “Sistem Rekomendasi Judul Skripsi Menggunakan Cosine Similarity Pada JATI ITN Malang,” *JUSIFOR: Jurnal Sistem Informasi dan Informatika*, vol. 4, no. 2, hlm. 243–252, Des 2025, doi: 10.70609/jusifor.v4i2.8340.
- [20] R. Astri, L. P. Hung, S. Binti Sura, dan A. Kamal, “Improving the Accuracy of Tourism Recommendation System Based on Neural Collaborative Filtering,” *Jurnal RESTI*, vol. 9, no. 4, hlm. 886–893, 2025, doi: 10.29207/resti.v9i4.6516.