

TESIS
SISTEM REKOMENDASI PARFUM BERBASIS TEKS
MENGGUNAKAN *SIAMESE-BERT*



disusun oleh
GUSTIAN HERLAMBANG
22.55.2302
Konsentrasi : Business Intelligence

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025

TESIS
SISTEM REKOMENDASI PARFUM BERBASIS TEKS
MENGGUNAKAN *SIAMESE-BERT*

A TEXT-BASED FRAGRANCE RECOMMENDATION SYSTEM
USING SIAMESE-BERT

Diajukan untuk memenuhi salah satu syarat mencapai derajat Pascasarjana
Program Studi Magister Informatika



disusun oleh

GUSTIAN HERLAMBAANG

22.55.2302

Konsentrasi : Business Intelligence

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA

2025

HALAMAN PERSETUJUAN

**SISTEM REKOMENDASI PARFUM BERBASIS TEKS MENGGUNAKAN
*SIAMESE-BERT***

***A TEXT-BASED FRAGRANCE RECOMMENDATION SYSTEM USING
SIAMESE-BERT***

yang disusun dan diajukan oleh

Gustlan Herlambang

22.55.2302

telah disetujui oleh Dosen Pembimbing Tesis
pada tanggal 10 Desember 2025

Dosen Pembimbing,



Hanif Al Fatta, S.Kom., M.Kom., P.hD.
NIK. 190302096

HALAMAN PENGESAHAN

**SISTEM REKOMENDASI PARFUM BERBASIS TEKS MENGGUNAKAN
*SIAMESE-BERT***

***A TEXT-BASED FRAGRANCE RECOMMENDATION SYSTEM USING
*SIAMESE-BERT****

yang disusun dan diajukan oleh

Gustlan Herlambang

22.55.2302

Telah dipertahankan di depan Dewan Penguji
pada tanggal 10 Desember 2025

Susunan Dewan Penguji

Nama Penguji

Emha Taufiq Luthfi, S.T., M.Kom., Ph.D.

NIK. 190302125

Alva Hendi Muhammad, A.Md., S.T., M.Eng., Ph.D.

NIK. 190302493

Hanif Al Fatta, S.Kom., M.Kom., Ph.D.

NIK. 190302096

Tanda Tangan



Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer
Tanggal 10 Desember 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.

NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : **Gustian Herlambang**
NIM : **22.55.2302**

Menyatakan bahwa Tesis dengan judul berikut:

**SISTEM REKOMENDASI PARFUM BERBASIS TEKS MENGGUNAKAN
SIAMESE-BERT**

Dosen Pembimbing : Hanif Al Fatta, S.Kom., M.Kom., Ph.D.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 10 Desember 2025

Yang Menyatakan,

The image shows a square official stamp of Universitas AMIKOM Yogyakarta. The stamp contains the university's logo, the name 'UNIVERSITAS AMIKOM', and the year '1998'. Overlaid on the stamp is a handwritten signature in black ink.

Gustian Herlambang

HALAMAN PERSEMBAHAN

Alhamdulillahirabbil'alamin. Sujud syukur kepada Allah SWT atas segala rahmat, kekuatan, dan petunjuk-Nya sehingga Tesis ini dapat diselesaikan. Tesis ini saya kepada:

1. Ayahanda, Sersan Mayor (Purn) Siswoyo, Ibunda Dewi Eti Sukmiati dan kedua kakak saya Mbak Eka Trisnowati dan Mbak Sri Rahayu Pujiyastuti. Tesis ini sebagai tanda bakti, cinta, dan hormat tertinggi saya. Terimakasih atas dukungan, do'a mengetuk pintu langit dan motivasi demi kelancaran jalan studi saya. Gelar Master ini adalah mahkota kecil untuk membalas ridha kalian.
2. Keluarga Besar Atmodiharjo. Terima kasih atas segala dukungan moral, do'a, dan kehangatan yang selalu memantik semangat proses studi saya.
3. Bapak Hanif Al Fatta, S.Kom., M.Kom., Ph.D. selaku dosen pembimbing saya yang telah memberikan arahan, motivasi, dan bimbingan yang ramah dan positif dalam proses penyelesaian tesis ini, sehingga dapat terselesaikan dengan baik.
4. Sahabat seperjuangan saya Pak M. Husein Budiraharjo, Pak Sudarto, Pak Rendra Prasetya, Pak Aris Subadi, Pak Agus Sukarno dan Pak M. Lutfi yang telah memberikan semangat, dukungan dan do'a untuk tidak menyerah dalam berjuang.
5. Keluarga besar MTI Amikom Yogyakarta, khususnya Angkatan 8 - kelas PJJ MTI Konsentrasi *Business Intelligence* yang telah menjadi teman-teman saya semasa kuliah pasca sarjana. Serta semua pihak yang telah membantu dan mendukung saya yang tidak saya sebutkan satu per satu.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT, Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, penulis dapat menyelesaikan tesis yang berjudul **“SISTEM REKOMENDASI PARFUM BERBASIS TEKS MENGGUNAKAN *SIAMESE-BERT*”**. Penulisan tesis ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Magister Informatika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.

Penulis menyadari bahwa tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan tesis ini, sangatlah sulit bagi penulis untuk menyelesaikan tesis ini. Oleh karena itu, pada kesempatan ini, penulis ingin menyampaikan ucapan terima kasih dan penghargaan yang setinggi-tingginya kepada:

1. Kedua orang tua, Sersan Mayor (Purn.) Siswoyo dan Dewi Eti Sukmiati. Terima kasih atas segala doa, pengorbanan, dan dedikasi seumur hidup yang telah membentuk karakter penulis hingga hari ini. Tidak lupa, keluarga besar Atmodiharjo yang senantiasa memberikan doa, dukungan yang tak terhingga.
2. Kedua kakak, Eka Trisnowati dan Sri Rahayu Pujiyastuti. Terimakasih atas kasih sayang, do'a dan dukungan selama penyusunan Tesis ini.
3. Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas Amikom Yogyakarta atas dukungan dan fasilitas yang diberikan selama penulis menempuh pendidikan.
4. Prof. Dr. Kusriani, M.Kom., selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
5. Hanif Al Fatta, S.Kom., M.Kom., Ph.D., selaku dosen pembimbing yang telah menyediakan waktu, tenaga, dan pikiran untuk mengarahkan penulis dalam penyusunan Tesis ini.
6. Prof. Dr. Ema Utami, S.Si., M.Kom., Robert Marco, S.T., M.T., Ph.D. , Emha Taufiq Luthfi, S.T., M.Kom., Ph.D., dan Alva Hendi Muhammad,

A.Md., S.T., M.Eng., Ph.D. selaku Tim Dosen Penguji yang telah memberikan kritik dan saran demi kesempurnaan tesis ini.

7. Seluruh dosen dan staf Sekolah Pascasarjana Program Studi PJJ Magister Informatika yang telah banyak membantu dalam kelancaran proses akademik penulis.
8. Sahabat-sahabat penulis (Fahmi Noor Fiqri, M.Kom., Dwi Monita L, S.Kom., M. Agung Ardiansyah, S.Kom., Farhan Naufal, S.Kom., Imam Agus Faisal, S.Hum., Andri Ismawan, dan Raka S. Permono, S.Pd) yang senantiasa memberikan semangat, motivasi, serta menjadi teman diskusi yang baik selama penyusunan tesis ini.
9. Teman-teman seperjuangan PJJ MTI Angkatan 8 Universitas Amikom Yogyakarta.
10. Biro Hukum dan Sumber Daya Manusia – Universitas Pakuan yang telah mengizinkan saya melanjutkan studi Magister di Universitas Amikom Yogyakarta. Profesor, Staf dan Dosen Universitas Pakuan yang mendukung studi Magister penulis. Tidak lupa, seluruh staf dan jajaran Biro Hukum SDM dan Biro Umum dan Keuangan Universitas Pakuan.

Penulis menyadari bahwa dalam penyusunan Tesis ini masih banyak kekurangan, oleh karena itu dengan senang hati menerima kritik dan saran yang membangun dari pembaca. Akhir kata, penulis berharap Tuhan Yang Maha Esa berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga tesis ini membawa manfaat bagi penulis sendiri, yang membaca dan pengembangan ilmu pengetahuan di masa mendatang.

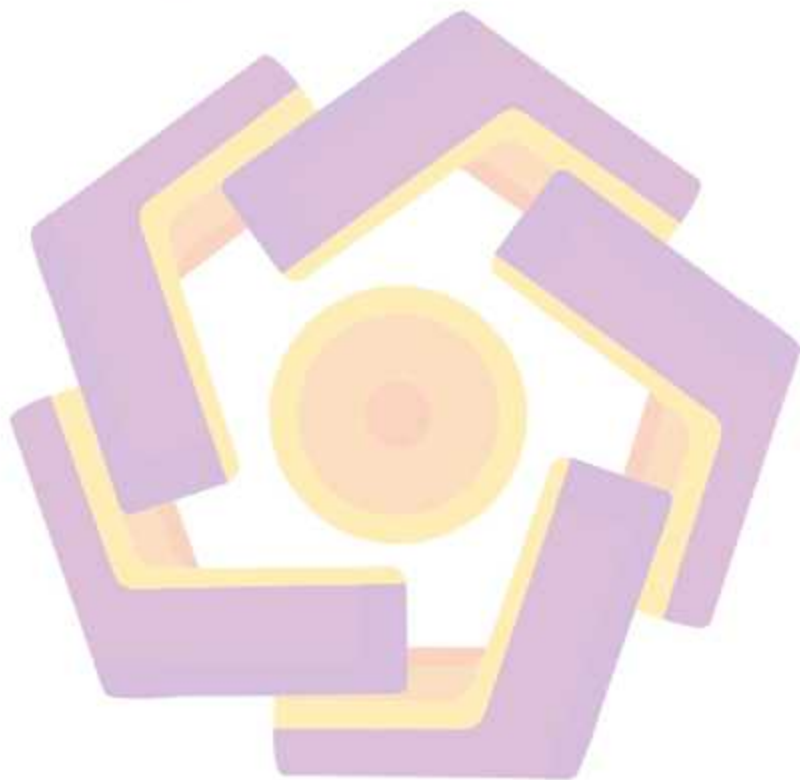
Yogyakarta, 10 Desember 2025

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN TESIS	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xii
INTISARI	xiii
<i>ABSTRACT</i>	xiv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah	6
1.4 Tujuan Penelitian	6
1.5 Manfaat Penelitian	7
BAB 2 TINJAUAN PUSTAKA	8
2.1 Tinjauan Pustaka	8
2.2 Keaslian Penelitian	14
2.3 Landasan Teori	24
BAB 3 METODE PENELITIAN	37
3.1 Jenis, Sifat, dan Pendekatan Penelitian	37
3.2 Metode Pengumpulan Data	37
3.3 Metode Analisis Data	38
3.4 Alur Penelitian	38
3.4.1 Pemahaman Masalah	38
3.4.2 Pemahaman dan Persiapan Data	39
3.4.3 Pemodelan	39
3.4.4 Evaluasi	42
BAB 4 HASIL PENELITIAN DAN PEMBAHASAN	46

4.1	Pemahaman Masalah	46
4.2	Pemahaman dan Persiapan Data	46
4.3	Pemodelan.....	52
4.4	Evaluasi.....	54
BAB 5	PENUTUP.....	70
5.1	Kesimpulan.....	70
5.2	Saran	71



DAFTAR TABEL

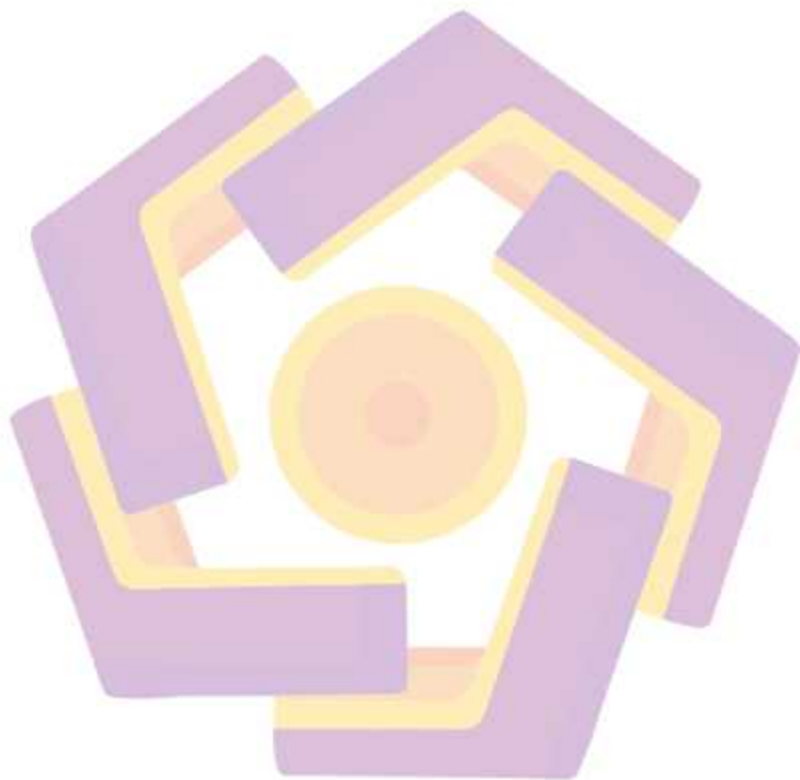
Tabel 1 Nilai <i>Fleiss' Kappa</i>	36
Tabel 2 Interpretasi Nilai <i>Fleiss' Kappa</i>	36
Tabel 3 Instrumen Penilaian Untuk Penilai (Kolektor dan <i>Enthusiast</i>).....	44
Tabel 4 Contoh 2 baris data parfum.....	46
Tabel 5 Contoh Data yang Telah Terstruktur.....	51
Tabel 6 Waktu Pelatihan Model.....	54
Tabel 7 Rekomendasi Parfum dari Ketiga Model.....	56
Tabel 8 Nilai <i>Cosine Similarity</i> Rekomendasi Berdasarkan 3 Model.....	57
Tabel 9 Rekomendasi Parfum dari Ketiga Model (<i>Berdasarkan Jarak Vektor Terdekat</i>).....	59
Tabel 10 Nilai <i>Cosine Similarity</i> Rekomendasi Berdasarkan 3 Model.....	60
Tabel 11 <i>Confusion Matrix</i> Model Berdasarkan Kueri.....	62
Tabel 12 Analisis <i>Fleiss' Kappa</i>	63
Tabel 13 Dua puluh Kueri untuk Menguji 4 arsitektur Model <i>Siamese-BERT</i>	66
Tabel 14 Perbandingan Performa Model <i>Siamese-BERT</i> pada Rekomendasi Parfum Berbasis Teks.....	67

DAFTAR GAMBAR

Gambar 1 Arsitektur <i>Transformer</i>	28
Gambar 2 (kiri) <i>Scaled Dot Attention</i> , (kanan) <i>Multi-Head Attention</i>	29
Gambar 3 Arsitektur SBERT dengan fungsi objektif klasifikasi, misalnya, untuk <i>fine-tuning</i> pada dataset SNLI. Kedua jaringan BERT memiliki bobot yang sama (arsitektur Siamese-BERT).	31
Gambar 4 Alur CRISP-DM	38
Gambar 5 Alur Evaluasi Komparatif Relevansi Rekomendasi.....	42
Gambar 6 Lima data teratas dari <i>Perfume Recommendation Dataset</i>	50
Gambar 7 Dataset yang terisi	50
Gambar 8 Proses Strukturisasi Data	51
Gambar 9 model <i>TF-IDF</i>	52
Gambar 10 Model <i>Word2Vec</i>	53
Gambar 11 Model <i>Siamese-BERT</i>	53
Gambar 12 Perbandingan Performa Model <i>Siamese-BERT</i> pada Rekomendasi Parfum Berbasis Teks.....	69

DAFTAR LAMPIRAN

Lampiran 1 Tampilan Form Fleiss' Kappa dan Hasil Anotator	79
---	----



INTISARI

Pemilihan parfum secara daring memiliki tantangan karena konsumen tidak dapat mencium aroma secara langsung dan hanya mengandalkan deskripsi yang bersifat subjektif serta metaforis. Pendekatan rekomendasi tradisional belum optimal dalam memahami makna kontekstual deskripsi parfum. Penelitian ini bertujuan untuk mengevaluasi kinerja arsitektur *Siamese-BERT* dibandingkan dengan metode leksikal (*TF-IDF* dan *Word2Vec*) dalam sistem rekomendasi parfum berbasis teks, serta menentukan model *Siamese-BERT* terbaik. Penelitian eksperimental ini menggunakan *dataset* publik sejumlah 2.191 data parfum dengan memanfaatkan kolom deskripsi dan *notes* (aroma). Evaluasi dilakukan secara kuantitatif melalui metrik *Cosine Similarity*, *Precision*, *Recall*, dan *F1-Score*, serta secara kualitatif menggunakan *Fleiss' Kappa* oleh tiga evaluator yang terdiri dari *enthusiast* dan kolektor parfum.

Hasil penelitian menunjukkan bahwa *Siamese-BERT* memberikan rekomendasi yang paling relevan berdasarkan persepsi pengguna. Evaluasi kualitatif *Fleiss' Kappa* membuktikan bahwa *Siamese-BERT* mencapai tingkat kesepakatan *Substantial Agreement* (0,69), mengungguli *TF-IDF* (0,47) dan *Word2Vec* (0,31). Selain itu, pengujian berbagai varian *Siamese-BERT* menunjukkan bahwa model *all-mpnet-base-v2* memiliki performa terbaik dengan nilai rata-rata *F1-Score* sebesar 0,95, *Precision* 0,94, dan *Recall* 0,93 karena mampu menangkap nuansa kompleks catatan aroma.

model *Siamese-BERT* terbukti lebih unggul dan efektif dalam menangkap makna semantik kalimat serta keseluruhan konteks pada kueri yang kompleks dibandingkan metode *baseline*. Model *all-mpnet-base-v2* dengan kapasitas parameter yang lebih besar sangat direkomendasikan untuk sistem pencarian berbasis teks pada domain yang bersifat deskriptif dan emosional seperti parfum.

Kata kunci: Sistem Rekomendasi, Parfum, *Siamese-BERT*, *Natural Language Processing*

ABSTRACT

Online perfume selection presents a challenge as consumers cannot smell the scent directly and must rely on subjective, metaphorical descriptions. Traditional recommendation approaches are suboptimal in understanding the contextual meaning of perfume descriptions. This study aims to evaluate the performance of the Siamese-BERT architecture against lexical methods (TF-IDF and Word2Vec) in a text-based perfume recommendation system, and to determine the best Siamese-BERT model. This experimental research utilized a public dataset of 2,191 perfume records, leveraging the description and notes columns. The evaluation was conducted quantitatively using Cosine Similarity, Precision, Recall, and F1-Score metrics, and qualitatively using Fleiss' Kappa assessed by three evaluators consisting of perfume enthusiasts and a collector.

The results showed that although Siamese-BERT's provided the most relevant recommendations based on user perception. The qualitative Fleiss' Kappa evaluation proved that Siamese-BERT achieved Substantial Agreement (0.69), outperforming TF-IDF (0.47) and Word2Vec (0.31). Furthermore, testing various Siamese-BERT variants revealed that the all-mpnet-base-v2 model performed the best, achieving an average F1-Score of 0.95, Precision of 0.94, and Recall of 0.93 due to its ability to capture complex olfactory notes.

In conclusion, the Siamese-BERT model is proven to be superior and more effective in capturing the semantic meaning of sentences and overall context in complex queries compared to baseline methods. The all-mpnet-base-v2 model, with its larger parameter capacity, is highly recommended for text-based retrieval systems in descriptive and emotional domains such as perfumes.

Keyword: Recommendation System, Perfume, Siamese-BERT, Natural Language Processing.

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Parfum merupakan barang yang saat ini lekat dengan setiap orang, bahkan menjadi identitas dari pemakainya. Industri parfum merupakan salah satu sektor yang berkembang pesat di dunia. Nilai pasar global parfum mencapai US\$60 milyar dan diproyeksikan terus meningkat hingga tahun 2029 [1]. Pertumbuhan ini didorong oleh meningkatnya konsumen terhadap ekspresi diri dan personalisasi produk. Pemilihan parfum sering dipengaruhi oleh faktor gaya hidup, kesempatan acara, preferensi aroma, bahkan nostalgia [2], [3]. Dalam konteks pemilihan hingga belanja parfum secara daring, konsumen seringkali menghadapi keterbatasan karena tidak dapat mencium aroma secara langsung. Konsumen mengandalkan deskripsi parfum yang sifatnya subjektif dan metaforis.

Pada era digital saat ini keterbatasan pengguna untuk mendapatkan informasi atau rekomendasi spesifik telah dijumpai oleh adanya sistem rekomendasi (*recommender system*) [4]. Sistem rekomendasi secara luas telah diaplikasikan pada bidang pariwisata, sosial media, bahkan *e-commerce* [5]. Khususnya untuk parfum, sistem rekomendasi telah dikembangkan menggunakan pendekatan *Content Based Filtering* [6]. Penelitian tersebut berfokus pada masalah pemilihan parfum pria secara daring di mana konsumen tidak dapat mencium aroma secara langsung. Menerapkan metode *Multi-Criteria Decision Making (MCDM)*

untuk membangun sistem pendukung keputusan. Metodologi yang digunakan mengombinasikan dua metode. Pertama, *Analytical Hierarchy Process (AHP)* diterapkan untuk menentukan bobot prioritas dari setiap kriteria parfum, yang meliputi *Top Notes*, *Middle Notes*, *Base Notes*, *Longevity* (ketahanan), dan *Price* (harga). Kedua, setelah bobot kriteria diperoleh, metode *Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS)* digunakan untuk melakukan pemeringkatan terhadap alternatif parfum yang ada. Hasil dari penelitian tersebut adalah sebuah sistem rekomendasi yang mampu memberikan peringkat parfum terbaik secara terstruktur dan objektif berdasarkan preferensi kriteria pengguna, sehingga membantu mengatasi tantangan dalam pembelian parfum secara daring.

Penelitian serupa juga dilakukan oleh [7] dengan mengambil studi kasus pada toko "*Quality Perfume*". Untuk membantu pengguna membuat keputusan yang objektif, penelitian ini merancang sebuah sistem pendukung keputusan berbasis situs web. Metode yang digunakan adalah *Weighted Product (WP)*. Dalam penelitian tersebut, proses perhitungan untuk menentukan alternatif parfum terbaik didasarkan pada kriteria-kriteria yang telah ditentukan, yaitu ketahanan aroma, jenis aroma, kualitas, dan harga.

Lebih lanjut, penelitian dilakukan oleh [8] yang telah menguji *Latent Semantic Analysis (LSA)* dan *Word2Vec* pada sistem rekomendasi parfum. Dimana, sistem tersebut menganalisis deskripsi kualitatif yang diberikan oleh pengguna, seperti suasana hati atau acara. Hasil evaluasi dalam penelitian menunjukkan bahwa akurasi model *Doc2Vec* dalam memberikan rekomendasi

mencapai 91%, sementara akurasi model LSA adalah 84%. Metode *Doc2Vec*, secara efektif mampu memberikan rekomendasi yang lebih baik.

Meskipun berbagai pendekatan pada penelitian diatas telah membantu proses rekomendasi, terdapat keterbatasan yang masih perlu diatasi. Metode berbasis *ranking* dan kuis cenderung terbatas pada atribut yang telah ditentukan sebelumnya, sehingga kurang fleksibel dalam menangkap preferensi unik setiap individu. Pendekatan berbasis kata kunci seperti LSA dan *Word2Vec* mampu memproses teks, namun belum optimal dalam memahami makna kontekstual deskripsi parfum. Saat ini, deskripsi parfum sering kali menggunakan bahasa yang kaya metafora dan nuansa, yang memerlukan pemahaman semantik mendalam.

Perkembangan teknologi *Natural Language Processing* (NLP), khususnya model berbasis *Transformer* seperti *Bidirectional Encoder Representations from Transformers* (BERT), memungkinkan pemahaman semantik yang lebih dalam terhadap teks. Kompleksitas data berbentuk teks dapat ditangani BERT [9]. Oleh karena itu model BERT telah digunakan untuk peningkatan rekomendasi salah satunya adalah produk pada *e-commerce* [10], [11].

Meskipun BERT dapat menangani basis data berbentuk teks kompleks dengan baik, namun BERT memiliki kekurangan dalam proses pelatihan yang memerlukan waktu dan tugas perbandingan kemiripan kalimat [12]. Keterbatasan BERT melahirkan *Sentence-BERT* melalui arsitektur *Siamese-BERT* yang dipublikasikan pertama kali oleh [13] dimana tidak hanya memiliki satu bobot saja seperti pada BERT, melainkan 2 bobot dan berbagi bobot parameter yang sama.

Singkatnya, *Sentence-BERT* memiliki 2 BERT didalamnya untuk memproses kemiripan kalimat. *Siamese-BERT* dengan data teks telah digunakan dalam berbagai bidang kesehatan [14], [15] untuk personalisasi sistem rekomendasi. Model BERT juga digunakan untuk mengekstrak preferensi pengguna pada sistem rekomendasi percakapan [16].

Siamese-BERT dalam konteks parfum telah diteliti oleh [17] untuk sistem rekomendasi parfum berbasis data deskripsi dan famili produk. Peneliti menggunakan data dari situs ensiklopedia parfum terbesar yaitu *Fragrantica*. Data yang berhasil dikumpulkan sebanyak 855 produk parfum berasal dari 242 merek rentang tahun 1920-2022. Atribut data yang digunakan meliputi nama, merek, *accords*, nilai ulasan, jumlah ulasan, cuaca, deskripsi, pembuat parfum, pembentuk (*top notes, middle notes, base notes*), jenis kelamin, dan harga. Untuk memberikan rekomendasi berdasarkan basis data yang ada, peneliti menggunakan pendekatan berbasis kuis yang berjumlah 5 pertanyaan yaitu 1) gender, 2) cuaca, 3) waktu penggunaan, 4) ketahanan, dan 5) lama jejak parfum (*sillage*).

Penelitian serupa yaitu *Sentence Transformer* untuk rekomendasi parfum telah dilakukan oleh [18]. Menggunakan data parfum dari komunitas daring bernama *BaseNotes*, peneliti membangun basis data baru bernama *Perfume Notes and Descriptions* (PND). PND terdiri dari 62 parfum komersial dengan deskripsi dan 255 aroma parfum unik. Sistem rekomendasi dibangun menggunakan 3 model arsitektur *Sentence-BERT* yaitu *MPNet-Base*, *DistilRoBERTa*, dan *MiniLM* (6 dan 12 lapis). Sistem dibangun melalui 2 tahap. Pertama, *baseline-model* dengan akurasi *Hit@k* dan *Mean Reciprocal Rank* (MRR) hanya 37% hingga 41%. Kedua,

dilakukan *fine-tuning* yaitu model ditingkatkan kinerjanya menggunakan *optimizer AdamW*, *learning-rate*, *loss-function* menggunakan *CosineSimilarityLoss*, dan *BatchSize* 32 dengan *Epoch* sebanyak 3. Proses *fine-tuning* yang telah dilakukan, meningkatkan *Hit@k* dan *Mean Reciprocal Rank (MRR)* 69% hingga 73%.

Berdasarkan uraian kedua penelitian diatas yang telah mengimplementasikan *Siamese-BERT* dalam domain parfum, masih terdapat hal yang belum dieksplorasi lebih jauh, pertama adalah jumlah data yang masih sedikit di penelitian sebelumnya yaitu 62 data. Kedua, mengenai *fine-tuning Siamese-BERT* dan penggunaan lebih jauh arsitektur *MPNetBase* atau *paraphrase-MiniLM* dengan *dataset* yang lebih banyak. Oleh karena itu, penelitian ini mengusulkan pengembangan sistem rekomendasi parfum berbasis teks menggunakan *Siamese-BERT* yang ditingkatkan untuk menghasilkan rekomendasi yang lebih baik, relevan, dan personal dibandingkan penelitian sebelumnya.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah pada penelitian ini adalah:

- a. Apakah pendekatan berbasis semantik (*Siamese-BERT*) mampu memberikan tingkat relevansi yang lebih tinggi dibandingkan pendekatan leksikal (*TF-IDF*) dan statis (*Word2Vec*) berdasarkan evaluasi kesepakatan pakar (*Fleiss' Kappa*)?
- b. Manakah arsitektur model *Siamese-BERT* yang paling optimal (antara varian *mpnet-base*, *minilm*, dan model spesialis) untuk menangani deskripsi parfum berdasarkan metrik evaluasi *F1-Score*, *Precision*, dan *Recall*?

- c. Se jauh mana model terbaik hasil optimasi mampu menangani kompleksitas kueri bahasa alami (*natural language queries*) yang bersifat abstrak dan subjektif dalam implementasi pencarian semantik parfum?

1.3 Batasan Masalah

Agar penelitian lebih terarah, ruang lingkup penelitian ini dibatasi pada :

- a. Jenis data: Data deskripsi parfum diambil dari repositori *Kaggle* yaitu *Perfume Recommendation Dataset* dimana informasi teks yang digunakan adalah kolom deskripsi dan aroma parfum.
- b. Fokus model: Model yang digunakan adalah *Bidirectional Encoder Representations from Transformers* (BERT) dengan arsitektur *Stanses-BERT*.
- c. Jenis rekomendasi: Rekomendasi diberikan berdasarkan kemiripan deskripsi dan aroma parfum.
- d. Evaluasi: Pengujian model dilakukan menggunakan metrik evaluasi kesamaan teks *cosine similarity*, dan *precision@k*, *recall@k*, dan *F1 Score*.
- e. Lingkup pengguna: sistem diuji pada skenario simulasi berbasis basis data, belum diimplementasikan untuk uji coba langsung pada pengguna akhir secara komersial.

1.4 Tujuan Penelitian

Tujuan penelitian ini yaitu :

- a. Mengevaluasi keunggulan pendekatan *Siamese-BERT* dibandingkan metode *baseline* (*TF-IDF* dan *Word2Vec*) menggunakan uji validasi manusia (*Fleiss' Kappa*).
- b. Menganalisis kinerja komparatif berbagai arsitektur model *Siamese-BERT pre-trained* untuk menentukan model terbaik (*State-of-the-Art*) yang memiliki akurasi tertinggi.
- c. Menguji efektivitas implementasi model terbaik dalam menangkap konteks semantik pada kueri pengguna yang kompleks dan abstrak (seperti nuansa, suasana, atau momen) melalui studi kasus pencarian..

1.5 Manfaat Penelitian

Manfaat penelitian ini adalah :

- a. Secara akademis, memberikan kontribusi pada pengembangan sistem rekomendasi berbasis *Natural Language Processing* (NLP) dengan memanfaatkan arsitektur *Siamese-BERT*.
- b. Secara akademis, menjadi referensi bagi penelitian selanjutnya yang mengkaji penerapan *transformer-based models* pada domain rekomendasi parfum berbasis deskripsi teks.
- c. Secara praktis, membantu pengguna menemukan parfum yang sesuai dengan preferensi aroma mereka meskipun tidak dapat mencium aroma secara langsung.
- d. Secara praktis, memberikan alternatif solusi bagi industri parfum dalam mengembangkan layanan rekomendasi produk yang lebih personal dan relevan.

BAB 2

TINJAUAN PUSTAKA

2.1 Tinjauan Pustaka

Sistem rekomendasi merupakan salah satu bidang penting dalam *information retrieval* yang berfungsi membantu pengguna menemukan item sesuai preferensi. Secara umum, terdapat tiga pendekatan utama, yaitu *content-based filtering*, *collaborative filtering*, dan *hybrid*. *Content-based filtering* bekerja dengan mencocokkan karakteristik item dengan profil pengguna, sedangkan *collaborative filtering* memanfaatkan pola kesamaan preferensi antar pengguna. Keduanya memiliki keterbatasan, seperti tidak memiliki cukup data untuk memberikan rekomendasi (*cold start*) pada *collaborative filtering* dan keterbatasan variasi pada *content-based filtering*. Oleh karena itu, pendekatan *hybrid* banyak dikembangkan untuk mengatasi kelemahan tersebut. Misalnya, [19] menunjukkan bahwa integrasi *content-based* dan *collaborative filtering* dalam sistem rekomendasi kesehatan mampu meningkatkan presisi hingga 0,87 dan *recall* 0,91 melampaui metode tunggal yang digunakan secara terpisah. Penelitian lain oleh [20] juga menegaskan bahwa pendekatan *hybrid* menghasilkan rekomendasi yang lebih beragam dan akurat dibandingkan metode tunggal. Bahkan, studi komparatif oleh [21] pada sistem rekomendasi musik membuktikan bahwa *Hybrid-Filtering* mencapai presisi tertinggi dibandingkan *Content-Based* maupun *Collaborative Filtering* secara terpisah.

Sistem rekomendasi berbasis *Sentence-BERT* dengan arsitektur *Siamese-BERT* telah digunakan untuk domain kesehatan yaitu untuk harmonisasi data biomedis penyakit *Alzheimer* yaitu otomatisasi proses pencocokan variabel penyakit *Alzheimer*. Variabel tersebut adalah *GERAS-EU* (Eropa) dan *GERAS-JP* (Jepang), yang memiliki format dan penamaan variabel berbeda. Proses harmonisasi seringkali memakan waktu sehingga peneliti [22] memanfaatkan model bahasa besar (*Large Language Models/LLMs*) seperti *E5*, *Siamese-BERT (MPNet dan MiniLM)*, *BioLORD-2023*, serta *fuzzy matching* sebagai pembandingan. *Siamese-BERT* dibangun untuk menghasilkan *embedding* kalimat yang bermakna secara semantik. Pada penelitian ini pencocokan variabel diperlakukan sebagai masalah pemerinkatan, dan hasil dari berbagai metode NLP digabungkan menggunakan *ensemble learning* berbasis *Random Forest*. Hasilnya menunjukkan bahwa model *E5* memberikan performa terbaik di antara metode individual, namun model *Random Forest* secara konsisten melampaui *E5* dalam semua metrik evaluasi, dengan *HR-30* mencapai 0.986 dan *MRR* sebesar 0.744.

Penelitian yang masih berkaitan dilakukan oleh [23] dimana peneliti mengembangkan pendekatan evaluasi otomatis terhadap komunikasi interprofesional berbasis format SBAR (*Situation, Background, Assessment, Recommendation*) dengan memanfaatkan model pemrosesan bahasa alami, khususnya *Sentence-BERT (S-BERT)*. Penelitian ini bertujuan untuk mengukur kesamaan semantik antara teks *handover* mahasiswa dan referensi ideal, sehingga dapat memberikan umpan balik yang objektif dan efisien dalam konteks pendidikan profesi kesehatan. Model yang digunakan dioptimalkan melalui pembelajaran

kontrasif dan pembobotan token, menghasilkan korelasi tinggi dengan penilaian manusia, terutama pada komponen “*Situation*” dan “*Recommendation*”. Secara kuantitatif, model menunjukkan performa akurasi yang menjanjikan, dengan nilai *F1* mencapai 0.78 dan korelasi *Pearson* sebesar 0.72 terhadap penilaian evaluator manusia. Penelitian ini menyarankan pengembangan lebih lanjut dengan mengintegrasikan data multimodal dan memperluas cakupan data untuk meningkatkan generalisasi model. Namun, keterbatasan data yang hanya berasal dari satu institusi dan fokus pada teks tertulis menjadi kelemahan utama.

Dalam bidang hukum, *Siamese-BERT* diimplementasikan untuk pengelompokan dokumen hukum yang diteliti oleh [24]. Pada penelitian ini menunjukkan bahwa pendekatan berbasis *transformer* dalam *Natural Language Processing (NLP)* mampu meningkatkan akurasi pengelompokan dokumen hukum secara signifikan. Mereka membandingkan delapan model *transformer* yaitu *BERT*, *GPT-2*, *RoBERTa*, dan *LLaMA* yang masing-masing dilatih pada korpus umum dan korpus khusus pada bidang hukum dalam bahasa Portugis. Hasilnya menunjukkan bahwa model yang dibuat khusus untuk hukum, seperti *Sabia 7B Jud.*, menghasilkan tingkat kemiripan tertinggi antar dokumen berdasarkan *cosine similarity*, dengan nilai mencapai rata-rata 99%. Penelitian ini juga memperkenalkan strategi segmentasi token (*Slot N/K*) untuk menangani dokumen panjang, serta metode *fine-tuning* yang mempertimbangkan *masking* dan tokenisasi sesuai arsitektur model. Temuan ini relevan dalam konteks pengembangan sistem rekomendasi berbasis teks, karena menunjukkan pentingnya spesialisasi model dan

teknik representasi vektor dalam meningkatkan kualitas pengelompokan dan pencarian semantik.

Pada sektor wisata sejarah atau museum pintar, *Siamese-BERT* digunakan untuk mengungkap bahwa pendekatan *Statistical and Data Analysis* (97%) dan *AI/Machine Learning* (63.6%) menjadi metode dominan dalam memahami pola interaksi pengguna, yang kemudian digunakan untuk menyajikan konten yang dipersonalisasi secara *real-time*. Arsitektur seperti *Siamese-BERT*, yang mengandalkan pembelajaran representasi semantik dari pasangan teks, sangat cocok untuk mengidentifikasi preferensi pengguna berdasarkan deskripsi parfum atau ulasan. Dalam konteks museum, perilaku seperti pola pergerakan, keterlibatan interaktif, dan respons emosional dianalisis untuk menyesuaikan konten. Integrasi teori pembelajaran kontekstual (Falk & Dierking), teori motivasi (Deci & Ryan), dan konstruktivisme (Piaget) menunjukkan bahwa personalisasi yang efektif harus mempertimbangkan dimensi personal, sosial, dan kognitif pengguna.

Pendekatan peningkatan pemahaman semantik teks melalui *fine-tuning* model bahasa besar (*Large Language Models/LLMs*) berbasis arsitektur *LLaMA* menggunakan teknik *Low-Rank Adaptation* (LoRA) dilakukan oleh [25]. Penelitian ini berfokus pada tugas klasifikasi kemiripan pertanyaan menggunakan *dataset Quora Question Pairs* (QQP), dan membandingkan performa lima model *LLM* yang telah dioptimasi dengan pendekatan sebelumnya, yaitu *Siamese Convolutional Neural Network* (*S-CNN*) yang dikombinasikan dengan *BERT*. Hasilnya menunjukkan bahwa model *qLLaMA_LoRA-7B* yang di-*finetune* secara *supervised* menghasilkan *F1-score* tertinggi sebesar 84.9%, mengungguli model *S-*

CNN (*F1-score* 82.02%) dan model LLM berparameter lebih besar seperti *Alpaca_LoRA-65B*.

Pada domain parfum berbasis data teks metaforis sistem rekomendasi parfum telah diteliti oleh [17]. Sistem tersebut berhasil dibangun berbasis data deskripsi dan famili produk parfum. Peneliti memanfaatkan arsitektur *Siamese-BERT* dan data dari situs ensiklopedia parfum terbesar yaitu *Fragrantica*. Data yang berhasil dikumpulkan sebanyak 855 produk parfum berasal dari 242 merek rentang tahun 1920-2022. Atribut dataset yang digunakan meliputi nama, merek, *accords*, nilai rating, jumlah rating, cuaca, deskripsi, pembuat parfum, pembentuk (*top notes, middle notes, base notes*), jenis kelamin, dan harga. Untuk memberikan rekomendasi berdasarkan dataset yang ada, peneliti menggunakan pendekatan berbasis kuis yang berjumlah 5 pertanyaan yaitu 1) gender, 2) cuaca, 3) waktu penggunaan, 4) ketahanan, dan 5) lama jejak parfum (*sillage*).

Penelitian lanjutan dengan memanfaatkan *Siamese-BERT* untuk rekomendasi parfum telah dilakukan oleh [18]. Menggunakan data parfum dari komunitas daring bernama *BaseNotes*, peneliti membangun basis data baru bernama *Perfume Notes and Descriptions* (PND). PND terdiri dari 62 parfum komersial dengan deskripsi dan 255 aroma parfum unik. Sistem rekomendasi dibangun menggunakan 3 model arsitektur Sentence-BERT yaitu MPNet-Base, DistilRoBERTa, dan MiniLM (6 dan 12 lapis). Sistem dibangun melalui 2 tahap. Pertama, *baseline-model* dengan akurasi *Hit@k* dan *Mean Reciprocal Rank* (MRR) hanya 37% hingga 41%. Kedua, dilakukan *fine-tuning* yaitu model ditingkatkan kinerjanya menggunakan *optimizer AdamW*, *learning-rate*, *loss-function*

menggunakan *CosineSimilarityLoss*, dan *BatchSize* 32 dengan *Epoch* sebanyak 3. *Fine-tuning* yang telah dilakukan, meningkatkan *Hit@k* dan *Mean Reciprocal Rank (MRR)* 69% hingga 73%.

Berdasarkan tinjauan pustaka diatas, penerapan *Siamese-BERT* belum spesifik digunakan untuk rekomendasi produk khususnya parfum. Penelitian ini mengusulkan temuan / *novelty* berupa pengembangan sistem rekomendasi parfum menggunakan pendekatan *Siamese-BERT* seperti yang telah dilakukan oleh [17], [18]. Penelitian ini menjawab *gap* dari studi sebelumnya yang belum memaksimalkan arsitektur selain *MPNet* dan jumlah hingga kompleksitas data pelatihan.



2.2 Keaslian Penelitian

Tabel 2.1 Matriks literatur review dan posisi penelitian
Sistem Rekomendasi Parfum Berbasis Teks Menggunakan *Siamese-Bert*

No	Judul Penelitian	Nama Peneliti, Tahun, Index	Metode Penelitian	Hasil	Keunggulan dan Kelemahan	Perbandingan
1	<i>A natural language processing approach to support biomedical data harmonization: Leveraging large language models</i>	Li, Z. et al., PLOS One, 2025, Scopus Q1	<i>Large Language Models (LLMs) (E5, MPNet, MiniLM, BioLORD-2023) dan ensemble learning (Random Forest).</i>	Hasil eksperimen menunjukkan bahwa pendekatan ensemble memberikan performa superior dibandingkan model individu, dengan capaian rata-rata <i>Hit Ratio (HR-30)</i> sebesar 0,986 dan <i>Mean Reciprocal Rank (MRR)</i> sebesar 0,744. Di antara model individu, E5 tercatat sebagai	Keunggulan: Kombinasi metode representasi semantik dari berbagai arsitektur LLM (E5, MPNet, hingga BioLORD-2023) dan <i>ensemble learning</i> dalam menangani variasi penamaan variabel yang sangat heterogen pada dataset studi Alzheimer. Kelemahan: Sumber daya komputasi yang sangat besar karena menjalankan banyak	Kedudukan penelitian membahas mengenai kemiripan teks, namun penelitian yang diusulkan memiliki <i>scope</i> yang berbeda yaitu gaya hidup (rekomendasi parfum) sedangkan penelitian sebelumnya pada bidang biomedis. Penggunaan <i>Siamese-BERT</i> pada penelitian ini untuk menguji apakah efisiensi <i>Siamese-BERT</i> sudah cukup mumpuni untuk menangkap nuansa

No	Judul Penelitian	Nama Peneliti, Tahun, Index	Metode Penelitian	Hasil	Keunggulan dan Kelemahan	Perbandingan
				model dengan performa terbaik.	model sekaligus (<i>ensemble</i>)	deskripsi sensorik parfum tanpa perlu menggunakan banyak model seperti yang dilakukan dalam skenario harmonisasi data medis Alzheimer.
2	<i>Enhancing Interprofessional Communication in Healthcare Using Large Language Models: Study on Similarity Measurement Methods with Weighted Noun Embeddings</i>	Ji-Young Yeo, Sungkwan Youm, Kwang-Seong Shin, <i>Electronics (MDPI, Open Access Journal)</i> , 2025, Scopus Q1.	Sentence-BERT dengan arsitektur Siamese-BERT yang ditingkatkan menggunakan pendekatan <i>Contrastive Learning</i> . Metode Biasing unik yang melibatkan (1) integrasi jawaban referensi ke dalam korpus pelatihan.	proses <i>biasing</i> dan penggunaan bobot pada kata benda (<i>domain-critical tokens</i>) secara signifikan meningkatkan keselarasan antara skor otomatis dengan evaluasi manusia dibandingkan model standar. <i>Sentence-</i>	Keunggulan: <i>weighted noun embeddings</i> sangat efektif dalam menangani skema pembobotan kata benda yang memastikan istilah medis penting tidak hilang. Penggunaan <i>Contrastive Learning</i> pada arsitektur <i>Siamese</i> berhasil mempertajam batas antara jawaban yang benar secara kualitas	<i>Scope</i> penelitian yang berbeda yaitu pendidikan medis vs gaya hidup. Penelitian Yeo dkk. (2025) lebih menekankan pada evaluasi ketepatan (penyelarasan dengan jawaban pakar) dalam konteks pendidikan, sementara penelitian yang dilakukan berfokus pada sistem rekomendasi (penyelarasan antara

No	Judul Penelitian	Nama Peneliti, Tahun, Index	Metode Penelitian	Hasil	Keunggulan dan Kelemahan	Perbandingan
			<p>(2) penggabungan respons siswa berprestasi tinggi, dan (3) penerapan Weighted Noun Embeddings untuk memberikan bobot lebih tinggi pada kata benda teknis medis yang krusial. Pengukuran kemiripan dilakukan menggunakan <i>Cosine Similarity</i>, <i>Euclidean Distance</i>, dan</p>	<p><i>BERT</i> memberikan skor kemiripan yang lebih tinggi. Model <i>FastText</i> memberikan performa unggul dalam menangani terminologi medis yang kompleks.</p>	<p>dengan yang hanya mirip secara leksikal</p> <p>Kelemahan: Model masih kesulitan dalam menangani variasi linguistik luas, struktur narasi kurang konsisten, dan ukuran sampel yang masih kecil yaitu 13 peserta.</p>	<p>preferensi pengguna dan profil produk). Selain itu, Yeo dkk. menggunakan metode <i>Weighted Noun Embeddings</i> sebagai fitur tambahan untuk memperkuat kata kunci domain medis. Penelitian ini dapat mengambil posisi dengan mengadaptasi teknik pembobotan istilah (seperti <i>Top Notes</i>, <i>Heart Notes</i>, <i>Base Notes</i> pada parfum) ke dalam arsitektur <i>Siamese-BERT</i>, dimana dalam jurnal tersebut terbukti mampu meningkatkan akurasi.</p>

No	Judul Penelitian	Nama Peneliti, Tahun, Index	Metode Penelitian	Hasil	Keunggulan dan Kelemahan	Perbandingan
			<i>Manhattan Distance</i> untuk membandingkan vektor <i>embedding</i> antara input dan referensi			
3.	<i>Analysing similarities between legal court documents using natural language processing approaches based on transformers</i>	Raphael Souza de Oliveira & Erick Giovani Sperandio Nascimento, PLOS One, 2025. Scopus Q1.	Penelitian ini mengeksplorasi efektivitas delapan teknik <i>Natural Language Processing</i> (NLP) berbasis arsitektur <i>transformer</i> untuk mendeteksi kemiripan pada dokumen hukum di sistem peradilan Brasil	Hasil eksperimen membuktikan bahwa model berbasis <i>transformer</i> secara signifikan mengungguli teknik NLP tradisional seperti Word2Vec atau TF-IDF dalam menangkap kemiripan dokumen yang kompleks dalam kasus hukum.	Keunggulan: <i>fine-tuning</i> pada dataset hukum masif (210.000 kasus hukum) memastikan model sangat sensitif terhadap nuansa bahasa hukum yang kaku. Penggunaan teknik <i>overlapping</i> pada "Slot N/K" memberikan solusi praktis bagi batasan <i>context window</i> pada	Penelitian ini memposisikan pada optimasi sistem rekomendasi yang lebih dinamis. Sementara Oliveira dkk. berfokus pada pengelompokan (<i>clustering</i>) dokumen statis untuk efisiensi peradilan, Penelitian ini juga mengevaluasi bagaimana arsitektur <i>Siamese-BERT</i> dapat memberikan

No	Judul Penelitian	Nama Peneliti, Tahun, Index	Metode Penelitian	Hasil	Keunggulan dan Kelemahan	Perbandingan
			yaitu BERT, GPT-2, RoBERTa, dan LLaMA (khususnya varian Sabia 7B) yang dilatih pada korpus bahasa Portugis Brasil	Model LLaMA (Sabia 7B Jud.) yang telah di- <i>fine-tuning</i> secara spesifik untuk domain hukum Brasil mencapai tingkat akurasi tertinggi dibandingkan model lainnya.	arsitektur <i>transformer</i> konvensional. Kelemahan: Komputasi sumber daya yang sangat tinggi untuk <i>fine-tuning</i> dan inferensi LLaMA 7B. Karena dilatih khusus pada bahasa Portugis Brasil, generalisasi model ini terhadap bahasa atau sistem hukum negara lain memerlukan pelatihan ulang yang intensif.	rekomendasi produk yang presisi berdasarkan deskripsi aromatik dari data parfum.
4	<i>Analyzing Visitor Behavior to Enhance Personalized Experiences in Smart</i>	Ivanov, R., & Velkova, V., 2025, Scopus Q1.	Penelitian ini menggunakan metode <i>Systematic Literature Review</i> (SLR) terhadap 33	Studi ini mengungkap bahwa analisis statistik dan data merupakan	Keunggulan: memberikan taksonomi yang sangat komprehensif mengenai hubungan antara teknologi	Ivanov & Velkova menyoroti kurangnya analisis terhadap aspek subjektif (seperti emosi) dalam personalisasi, maka

No	Judul Penelitian	Nama Peneliti, Tahun, Index	Metode Penelitian	Hasil	Keunggulan dan Kelemahan	Perbandingan
	<i>Museums: A Systematic Literature Review.</i>		publikasi terpilih selama satu dekade terakhir untuk memetakan metodologi utama dalam analisis perilaku pengunjung di museum cerdas. Peneliti mengklasifikasikan literatur berdasarkan metodologi analisis data (statistik, AI/ML), teknologi yang digunakan (IoT, pelacakan lokasi),	metodologi yang dominan (97%), di mana teknologi AI dan <i>Machine Learning</i> (63,6%) menjadi pilar utama dalam sistem personalisasi waktu nyata. Hasil tinjauan menunjukkan bahwa fokus utama analisis perilaku saat ini adalah pola pergerakan dan keterlibatan interaktif, sementara aspek respons emosional dan dinamika kelompok	pelacakan perilaku dengan efektivitas personalisasi pengalaman pengguna, serta berhasil mengidentifikasi celah krusial terkait analisis emosi manusia. Kelemahan: terbatas pada sintesis data sekunder dari jumlah sampel yang relatif kecil (33 dokumen) dan belum menguji efektivitas arsitektur model AI tertentu secara eksperimental di lapangan.	penelitian ini mengisi celah tersebut dengan mengimplementasikan Siamese-BERT untuk menangkap kemiripan semantik pada deskripsi parfum yang bersifat subjektif-tekstual, sehingga memberikan kontribusi teknis pada sistem rekomendasi yang lebih sensitif terhadap nuansa deskripsi manusia.

No	Judul Penelitian	Nama Peneliti, Tahun, Index	Metode Penelitian	Hasil	Keunggulan dan Kelemahan	Perbandingan
			serta jenis perilaku pengunjung yang dianalisis untuk mendukung pengiriman konten personalisasi.	masih sangat jarang dieksplorasi dalam pengembangan konten yang dipersonalisasi.		
5	<i>An NLP-Based Perfume Note Estimation Based on Descriptive Sentences.</i>	J. Kim, K. Oh & B. Oh. 2024. Scopus Q2	<i>Natural Language Processing (NLP) berbasis model Transformer dengan arsitektur seperti BERT, RoBERTa, dan Electra.</i>	Penelitian oleh Kim, Oh, dan Oh (2024) mengusulkan pendekatan <i>Natural Language Processing (NLP)</i> berbasis model <i>Transformer</i> untuk mengestimasi komponen <i>perfume notes</i> (top, middle,	Keunggulan: Keunggulan utama penelitian ini terletak pada kemampuannya menjembatani celah antara persepsi sensoris manusia yang dituangkan dalam bahasa alami dengan komposisi teknis parfum Kelemahan:	Perbedaan mendasar antara penelitian Kim dkk. (2024) dengan penelitian ini terletak pada tujuan akhir dan arsitektur model; Kim dkk. berfokus pada estimasi konten (<i>content estimation</i>) menggunakan klasifikasi standar, sedangkan penelitian ini memposisikan diri pada

No	Judul Penelitian	Nama Peneliti, Tahun, Index	Metode Penelitian	Hasil	Keunggulan dan Kelemahan	Perbandingan
				<p>dan base notes) melalui analisis kalimat deskriptif subjektif. Dengan menggunakan dataset yang terdiri dari 1.250 deskripsi parfum, mereka mengevaluasi berbagai arsitektur seperti BERT, RoBERTa, dan Electra dalam tugas klasifikasi multi-label, di mana hasil eksperimen menunjukkan bahwa model BERT-base memberikan akurasi</p>	<p>keterbatasan pada fokus klasifikasi label tunggal yang belum mengeksplorasi hubungan relasional atau kemiripan antar produk secara mendalam.</p>	<p>aspek rekomendasi berbasis kemiripan (<i>similarity-based recommendation</i>). Dengan menerapkan arsitektur Siamese-BERT, penelitian ini mengoptimalkan representasi vektor (<i>embedding</i>) untuk menghitung kedekatan semantik antar deskripsi parfum, sehingga mampu memberikan rekomendasi produk yang lebih relevan secara personal dibandingkan sekadar pelabelan <i>note</i> secara mandiri.</p>

No	Judul Penelitian	Nama Peneliti, Tahun, Index	Metode Penelitian	Hasil	Keunggulan dan Kelemahan	Perbandingan
				tertinggi dalam memetakan narasi tekstual ke label aromatik yang objektif.		
6	Personalized Quiz-Based Perfume Recommender System Using Social Data	Lutan, R & Badica, C, 2023, Scopus Q1.	NLP dan Kuis	Studi ini menggunakan dataset komprehensif sebanyak 855 entri parfum yang dihimpun secara mandiri dari ensiklopedia daring <i>Fragrantica</i> , mencakup parameter seperti keluarga aroma (<i>Fragrance Families</i>), musim, waktu penggunaan,	<p>Keunggulan:</p> <p>Keunggulan utama penelitian ini adalah keberhasilannya mengintegrasikan aspek kontekstual (sosial/gaya hidup) ke dalam algoritma rekomendasi, yang sangat membantu dalam mengatasi masalah <i>cold-start</i> bagi pengguna awam.</p> <p>Kelemahan:</p>	Penelitian ini memiliki kebaruan dalam hal otomatisasi pemrosesan bahasa alami dibandingkan dengan sistem berbasis kuis milik Lutan dkk. (2023). Jika Lutan dkk. memerlukan interaksi aktif pengguna melalui kuis untuk memetakan kategori, penelitian ini menggunakan arsitektur Siamese-BERT untuk mengevaluasi kemiripan

No	Judul Penelitian	Nama Peneliti, Tahun, Index	Metode Penelitian	Hasil	Keunggulan dan Kelemahan	Perbandingan
				<p>daya tahan (<i>longevity</i>), dan jejak aroma (<i>sillage</i>). Hasil penelitian menunjukkan bahwa pendekatan berbasis kuis interaktif efektif dalam menghasilkan rekomendasi yang personal dengan cara menyalurkan ekspektasi kontekstual pengguna terhadap profil objektif parfum dalam dataset.</p>	<p>ketergantungan sistem terhadap input eksplisit pengguna melalui antarmuka kuis yang bersifat statis, serta belum mengeksplorasi hubungan semantik antar-deskripsi tekstual yang lebih kompleks di luar kategori yang sudah terdefinisi.</p>	<p>semantik secara langsung dari teks deskripsi parfum. Dengan demikian, penelitian ini memposisikan diri pada optimasi sistem rekomendasi yang lebih dinamis dan mampu menangkap nuansa subjektif dalam teks tanpa terbatas pada parameter kategori kaku (seperti musim atau keluarga aroma saja), sehingga memberikan tingkat presisi semantik yang lebih tinggi pada domain parfum.</p>

2.3 Landasan Teori

2.3.1 Sistem Rekomendasi

Sistem Rekomendasi (*Recommender System*) adalah kelas perangkat lunak yang cerdas. Sistem ini menyaring sejumlah besar data untuk menyajikan berbagai item (misalnya, produk, berita, film, musik) yang paling relevan dan menarik bagi pengguna. Tujuan utamanya adalah untuk mengatasi masalah kelebihan informasi, di mana pengguna kesulitan menemukan item yang mereka sukai dari kumpulan data yang sangat besar.

Tidak hanya bertujuan untuk memprediksi "peringkat" atau "preferensi" yang mungkin diberikan pengguna terhadap suatu item, sistem rekomender meningkatkan pengalaman pengguna secara keseluruhan. Sistem ini bekerja dengan cara menemukan pola dalam data historis, seperti item yang pernah dibeli, dilihat, atau diberi peringkat oleh pengguna. Terdapat 3 jenis sistem rekomendasi yaitu :

- a. ***Collaborative Filtering***. Metode ini merupakan salah satu yang paling populer dan banyak digunakan. Metode ini merekomendasikan item kepada pengguna berdasarkan selera dari pengguna lain yang "serupa". Sebagai contoh, jika dua orang pengguna menyukai item yang sama sebelumnya, kemungkinan besar mereka akan menyukai item yang sama berikutnya [19]. Metode ini tidak memerlukan analisis terhadap konten item itu sendiri. Metode ini dibagi menjadi 2 pendekatan utama yaitu:

- 1) **User-Based Collaborative Filtering.** Mencari pengguna lain yang memiliki riwayat penilaian (*rating*) atau interaksi yang mirip dengan pengguna target. Kemudian, sistem akan merekomendasikan item yang disukai oleh pengguna-pengguna serupa tersebut tetapi belum pernah dicoba oleh pengguna target.
- 2) **Item-Based Collaborative Filtering.** Metode ini menghitung kesamaan antar item. Misalnya, jika banyak pengguna yang menyukai film A juga menyukai film B, maka sistem akan merekomendasikan film B kepada pengguna lain yang baru saja menonton film A. Pendekatan ini sering digunakan di platform *e-commerce* besar karena lebih stabil dan skalabel dibandingkan *User-Based*.

Meskipun sangat efektif, metode ini memiliki beberapa tantangan yang dikenal sebagai masalah *data sparsity* (data yang jarang, di mana banyak pengguna belum memberi peringkat pada banyak item) dan *cold start* (kesulitan memberikan rekomendasi untuk pengguna baru atau item baru yang belum memiliki data interaksi).

- b. **Content-based Filtering.** Metode ini memberikan rekomendasi berdasarkan analisis atribut atau karakteristik dari item itu sendiri dan profil preferensi pengguna. Sistem akan merekomendasikan item yang memiliki kemiripan atribut dengan item-item yang pernah disukai oleh pengguna sebelumnya. Untuk mengekstrak fitur dari item (seperti teks atau gambar), teknik-teknik seperti TF-IDF (*Term Frequency-Inverse Document Frequency*) dan *word embedding* (misalnya *Word2Vec* atau BERT) sering digunakan untuk

memahami konten secara semantik. Kesamaan antar item kemudian dihitung menggunakan metrik seperti *Cosine Similarity*. Kelebihan utama dari metode ini adalah kemampuannya untuk merekomendasikan item baru (mengatasi masalah *cold start* untuk item) dan tidak memerlukan data dari pengguna lain. Namun, kelemahannya adalah kecenderungan untuk menghasilkan rekomendasi yang terlalu mirip (*overspecialization*) dan kurangnya elemen kejutan (*serendipity*) (Putra & Rachman, 2024).

- c. **Hybrid Recommender System.** Sistem ini menggabungkan dua atau lebih teknik rekomendasi untuk menghasilkan rekomendasi yang lebih akurat dan kuat. Langkah yang biasanya digunakan adalah menggunakan 1) *Weighted Hybrid* yaitu menggabungkan hasil skor dari beberapa sistem rekomendasi yang berbeda dengan bobot tertentu, 2) *Switching Hybrid* yaitu sistem beralih antara beberapa metode rekomendasi tergantung pada konteks atau situasi tertentu (misalnya, jika pengguna baru, gunakan *Content-Based*; jika sudah memiliki banyak data, gunakan *Collaborative Filtering*) dan 3) *Feature Combination* yaitu mengintegrasikan fitur-fitur dari satu pendekatan ke pendekatan lainnya. Contohnya, memasukkan atribut item (dari *Content-Based*) ke dalam model *Collaborative Filtering* untuk membantu mengatasi masalah *cold start* [20].

2.3.2 *Natural Language Processing (NLP)*

NLP merupakan cabang dari kecerdasan buatan yang berfokus pada interaksi antara komputer dan bahasa manusia. Tujuan utama NLP adalah memungkinkan mesin untuk memahami, memproses, dan menghasilkan bahasa

alami secara bermakna [26]. Dalam konteks sistem rekomendasi berbasis teks, NLP berperan penting untuk mengekstraksi informasi semantik dari deskripsi produk, ulasan pengguna, maupun preferensi yang dinyatakan dalam bentuk teks.

Menurut [27] representasi teks dalam NLP dibagi menjadi beberapa tahap yaitu :

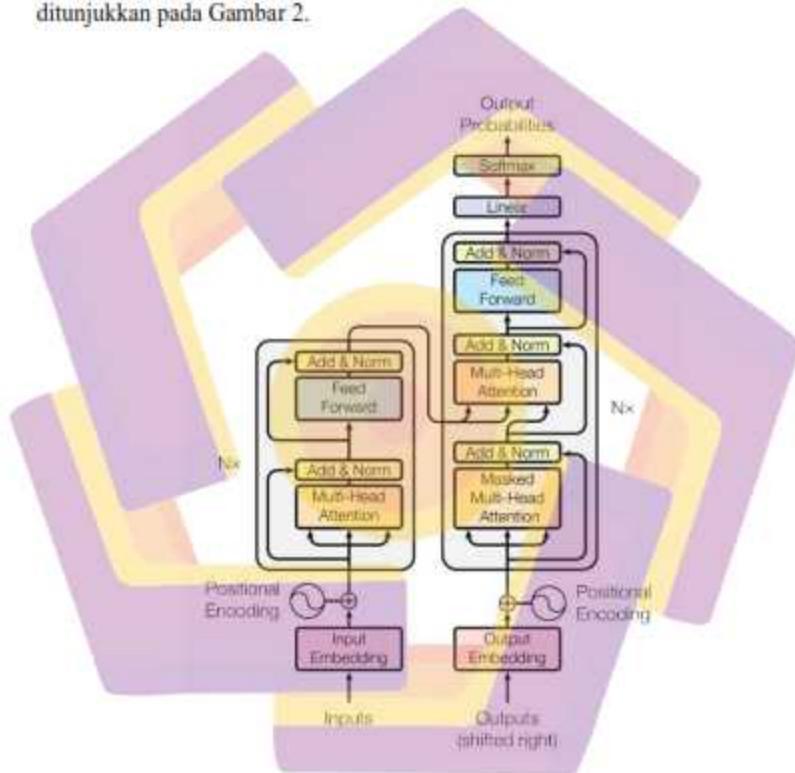
- a. **Representasi tradisional:** *Bag of Words (BoW)* dan *TF-IDF* yang hanya memperhatikan frekuensi kata tanpa konteks.
- b. **Distributed representation:** *Word2Vec* dan *GloVe* yang memetakan kata ke dalam vektor berdimensi rendah berdasarkan konteks kemunculannya.
- c. **Contextual embeddings:** model berbasis *Transformer* seperti BERT yang mampu menangkap makna kata sesuai konteks kalimat.

Representasi berbasis konteks ini sangat relevan untuk domain parfum, karena istilah seperti *fresh*, *woody*, atau *floral* dapat memiliki nuansa berbeda tergantung kombinasi kata lain dalam deskripsi.

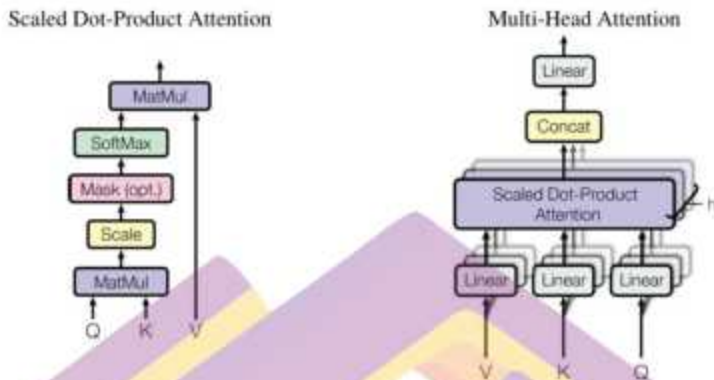
2.3.3 Transformer

Arsitektur Transformer diperkenalkan pertama kali oleh [28] para peneliti di Google. Arsitektur ini mengatasi masalah pada arsitektur sebelumnya yaitu *Recurrent Neural Networks (RNNs)*. *Transformer* sepenuhnya mengandalkan mekanisme *attention* khususnya *self-attention* dimana arsitekturnya ditunjukkan pada Gambar 1. *Attention* dalam *Transformer* adalah mekanisme yang memungkinkan model “memperhatikan” kata-kata lain dalam sebuah kalimat ketika sedang memproses satu kata tertentu.

Jadi, *attention* membaca kata demi kata secara berurutan seperti RNN, *Transformer* menggunakan *attention* untuk menghitung seberapa relevan setiap kata terhadap kata yang sedang diproses. Hasilnya, representasi kata tidak hanya bergantung pada dirinya sendiri, tetapi juga pada konteks seluruh kalimat. Arsitektur *Transformers* ditunjukkan pada Gambar 2.



Gambar 1 Arsitektur *Transformer*



Gambar 2 (kiri) *Scaled Dot Attention*, (kanan) *Multi-Head Attention*
 Komponen inti *Transformer* adalah *Scaled Dot-Product Attention*, yang dirumuskan sebagai berikut:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) \quad (1)$$

Dengan Q (*query*), K (*key*), dan V (*value*) merupakan representasi vektor dari token, serta d_k adalah dimensi *key*. Mekanisme ini memungkinkan setiap token memperhatikan token lain dalam sekuens, sehingga konteks global dapat dipertimbangkan secara bersamaan. Untuk memperkaya representasi, *Transformer* menggunakan *Multi-Head Attention*, yaitu menjalankan beberapa mekanisme atensi secara paralel pada sub-ruang representasi yang berbeda.

2.3.4 *Bidirectional Transformers for Language Understanding (BERT)*

BERT diperkenalkan oleh [12] merupakan model representasi bahasa berbasis *Transformer Encoder* [28]. Berbeda dari model sebelumnya (misalnya *GPT* yang *unidirectional*), *BERT* dilatih secara *bidirectional* sehingga mampu

memanfaatkan konteks dari kiri dan kanan secara simultan. Tujuan utamanya adalah menghasilkan representasi kontekstual yang dapat di-*fine-tune* untuk berbagai tugas NLP tanpa arsitektur khusus [12].

BERT menggunakan *multi-layer bidirectional Transformer encoder*. Dua konfigurasi utamanya adalah *BERT Base* dengan 12 lapisan, 768 *hidden-size*, 12 *attention heads* (110M parameter) dan *BERT Large* dengan 24 layer, 1024 hidden size, 16 *attention heads* (340M parameter). Input direpresentasikan dengan kombinasi *token embeddings*, *segment embeddings*, dan *position embeddings* [12].

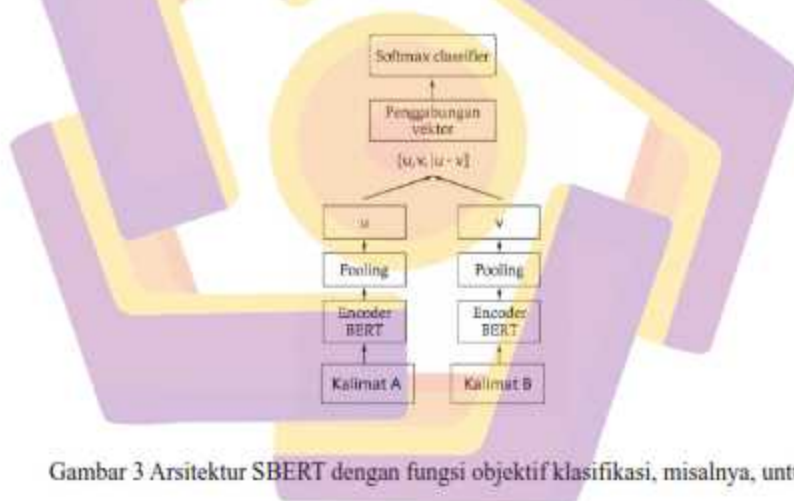
Proses *pre-training BERT* dilakukan dengan dua tujuan utama: *Masked Language Model* (MLM), yaitu memprediksi token yang disembunyikan berdasarkan konteks kiri dan kanan, serta *Next Sentence Prediction* (NSP), yaitu memprediksi hubungan antar kalimat. Setelah tahap ini, *BERT* dapat di-*fine-tune* untuk berbagai tugas NLP seperti klasifikasi teks, *sequence labeling*, dan *question answering* [12].

Keunggulan *BERT* terletak pada fleksibilitas dan performanya yang mencapai *state-of-the-art* di berbagai benchmark *NLP*. Model ini menjadi fondasi penting bagi pengembangan arsitektur lanjutan seperti *RoBERTa*, *ALBERT*, dan juga varian khusus seperti *Siamese-BERT* yang berfokus pada pengukuran kesamaan antar kalimat [12].

2.3.5 *Siamese-BERT*

Siamese-BERT (SBERT) adalah pengembangan dari *BERT* yang dirancang untuk menghasilkan representasi kalimat yang bermakna secara semantik. *BERT*

standar menggunakan *cross-encoder* sehingga tidak efisien untuk pencarian kesamaan kalimat dalam skala besar. Reimers & Gurevych (2019) mengusulkan arsitektur *siamese* yang memproses dua kalimat secara independen dengan bobot berbagi satu sama lain, lalu menghasilkan *sentence embeddings* tetap melalui proses *pooling*. Representasi ini dapat dibandingkan menggunakan *cosine similarity* atau metrik jarak lain, sehingga memungkinkan aplikasi seperti *semantic search*, *clustering*, dan *retrieval-based QA* berjalan jauh lebih cepat. Evaluasi menunjukkan SBERT mengungguli metode sebelumnya seperti *InferSent* dan USE, baik dari sisi akurasi maupun efisiensi komputasi.



Gambar 3 Arsitektur SBERT dengan fungsi objektif klasifikasi, misalnya, untuk *fine-tuning* pada dataset SNLI. Kedua jaringan BERT memiliki bobot yang sama (arsitektur Siamese-BERT).

Pada Gambar 3, terdapat input 2 kalimat. A dan B adalah dua kalimat yang akan dibandingkan. Lalu masing-masing kalimat di proses menggunakan BERT *encoder* secara terpisah. Diberi nama Siamese-BERT dikarenakan dua input

diproses oleh model yang identik (berbagi bobot, dalam contoh ini adalah *softmax*). Hasil keluarannya adalah vektor representasi kalimat, u untuk kalimat A dan v untuk kalimat B. BERT menghasilkan vektor token per posisi. Untuk mendapatkan satu vektor tetap per kalimat, digunakan strategi **pooling**. Selanjutnya vektor u dan v digabungkan menjadi satu vektor.

$$(u, v, |u - v|) \quad (2)$$

Dimana:

u = representasi kalimat A

v = representasi kalimat B

$|u - v|$ = nilai absolut dari perbedaan elemen vektor (menangkap jarak semantik)

Selanjutnya, vektor gabungan tersebut dimasukkan ke dalam *fully connected layer* dan *softmax classifier*. Hasil keluarannya adalah prediksi kelas.

2.3.6 Parfum

Parfum merupakan campuran senyawa pewangi yang dilarutkan dalam pelarut sesuai karakteristiknya. Bahan pewangi dapat berasal dari sumber alami seperti minyak atsiri maupun bahan sintetis. Komposisi utama parfum terdiri dari tiga komponen, yaitu zat pelarut, zat pewangi, dan zat pengikat. Zat pelarut yang umum digunakan adalah alkohol atau etanol, berfungsi melarutkan senyawa pewangi. Zat pewangi merupakan komponen utama yang menentukan aroma, sedangkan zat pengikat berfungsi mengurangi tingkat penguapan serta meningkatkan stabilitas komponen volatil (Maudhy et al., 2023).

Kualitas parfum umumnya ditentukan melalui uji daya tahan aroma dan kejernihannya. Salah satu cara meningkatkan daya tahan aroma adalah dengan menambah persentase bibit parfum dalam formulasi (Hikmah et al., 2023). Selain komposisi kimia, parfum juga diklasifikasikan berdasarkan tingkatan aroma yang muncul secara berurutan [31], yaitu:

- a. *Top Note*. Aroma pertama yang tercium, mudah menguap, dan cepat memudar.
- b. *Middle Note*. Aroma inti yang muncul setelah *top note*, bertahan sekitar satu jam.
- c. *Base Note*. Aroma dasar yang bertahan paling lama, muncul setelah beberapa jam pemakaian. Umumnya wangi khas seseorang ditempat publik tercium dari fase akhir (*base notes*) ini.

Dalam praktiknya, deskripsi parfum tidak hanya menjelaskan komposisi kimia, tetapi juga menggunakan bahasa naratif dan subjektif. Misalnya, istilah seperti *fresh*, *woody*, *floral*, *oriental*, atau metafora seperti "parfum segar tapi manis untuk sehari-hari" sering digunakan untuk menggambarkan karakter aroma. Variasi istilah dan gaya penulisan ini menimbulkan tantangan dalam pemrosesan teks, karena makna yang sama dapat diekspresikan dengan kata-kata berbeda.

Oleh karena itu, pemahaman terhadap domain parfum tidak hanya penting dari sisi kimiawi, tetapi juga dari sisi linguistik. Representasi semantik berbasis NLP diperlukan untuk mengekstraksi makna dari deskripsi parfum yang beragam, sehingga dapat dimanfaatkan dalam sistem rekomendasi berbasis teks.

2.3.7 Cosine Similarity

Cosine similarity merupakan salah satu metrik kesamaan yang paling banyak digunakan dalam pemrosesan bahasa alami dan sistem temu kembali informasi. Metrik ini menghitung nilai kosinus dari sudut antara dua vektor dalam ruang berdimensi tinggi, sehingga menekankan kesamaan arah vektor tanpa dipengaruhi oleh besarnya [32]. Secara matematis, dirumuskan sebagai berikut:

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

Dimana $A \cdot B$ adalah hasil kali titik (*dot product*) ini berarti mengukur seberapa besar komponen-komponen A searah dengan B . Semakin besar *dot product*, semakin kecil sudut di antara vektor. Sedangkan $\|A\|$, $\|B\|$ adalah panjang (norma) masing-masing vektor.

Konsep ini pertama kali dipopulerkan dalam bidang *information retrieval* melalui *Vector Space Model* yang diperkenalkan oleh [33]. Dalam model tersebut, dokumen dan kueri direpresentasikan sebagai vektor kata, dan *cosine similarity* digunakan untuk mengukur tingkat kesamaan antara keduanya. Sejak saat itu, *cosine similarity* menjadi metrik standar dalam berbagai aplikasi, mulai dari pencarian teks hingga sistem rekomendasi berbasis konten.

2.3.8 Confusion Matrix

Matriks kebingungan (*confusion matrix*) adalah cara untuk mengevaluasi model atau menggambarkan kinerja model pada data uji yang sudah diketahui

kebenarannya. Matriks ini dapat menangkap prediksi benar dan prediksi salah [34]. Matriks ini biasanya diukur dengan *Precision*, *Recall*, dan *F1-Score* [35]

Presisi mengukur seberapa akurat tebakan positif model. Presisi akan menjawab pertanyaan: dari semua yang *ditebak* sebagai "A", berapa banyak yang *benar-benar* "A"? Dengan demikian, metrik presisi ini fokus untuk meminimalkan kesalahan identifikasi (*False Positive*).

Selanjutnya, metrik *Recall* mengukur seberapa lengkap model dalam menemukan semua kasus positif yang ada; dari semua yang *sebenarnya* "A", berapa banyak yang berhasil ditemukan oleh model? Metrik *Recall* fokus untuk meminimalkan kasus yang terlewat (*False Negative*). Karena Presisi dan *Recall* seringkali bertentangan, *F1 Score* digunakan sebagai nilai tunggal yang mencari titik keseimbangan harmonis antara keduanya, dimana hal tersebut memberikan gambaran performa model secara keseluruhan.

2.3.9 Fleiss Kappa

Fleiss' Kappa (κ) adalah sebuah ukuran statistik yang digunakan untuk menilai tingkat reliabilitas atau kesepakatan (*inter-rater reliability*) di antara tiga atau lebih penilai (*raters*). Ukuran statistik ini mengukur seberapa besar kesepakatan yang dicapai oleh para penilai, setelah memperhitungkan faktor kebetulan. Fleiss' Kappa mengukur konsistensi penilaian, bukan validitas (hal benar). Interpretasi nilai Fleiss' Kappa berkisar dari -1 sampai dengan +1 [36]. Nilai Fleiss' Kappa ditunjukkan pada Tabel 1. Interpretasi nilai Fleiss' Kappa secara umum dinotasikan dalam bentuk skala, yang ditunjukkan pada Tabel 2.

Tabel 1 Nilai *Fleiss' Kappa*

Nilai <i>Fleiss' Kappa</i>	Keterangan
$\kappa = +1$	Kesepakatan sempurna dari semua penilai
$\kappa = 0$	kesepakatan yang terjadi tidak lebih baik dari kebetulan
$\kappa < 0$	kesepakatan yang terjadi lebih buruk daripada kebetulan

Tabel 2 Interpretasi Nilai *Fleiss' Kappa*

Skala <i>Fleiss' Kappa</i>	Keterangan
$< 0,40$	Buruk
$0,40 - 0,75$	Cukup Baik
$> 0,75$	Sangat Baik

BAB 3 METODE PENELITIAN

3.1 Jenis, Sifat, dan Pendekatan Penelitian

Jenis penelitian ini adalah penelitian eksperimental dan menggunakan data primer. Dimana penelitian ini melakukan pengujian performa arsitektur *Siamese-BERT* terhadap data parfum berjumlah 2191 entri. Penelitian ini menggunakan pendekatan kuantitatif. Penelitian akan menghasilkan model sistem rekomendasi parfum berbasis teks menggunakan arsitektur Siamese-BERT. Sistem akan dilakukan evaluasi apakah dapat merekomendasikan item parfum dengan akurat atau tidak.

3.2 Metode Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah *Perfume Recommendation Dataset* berasal dari repositori publik bernama Kaggle (<https://www.kaggle.com/datasets/nandini1999/perfume-recommendation-dataset>). Data terdiri dari 2191 entri parfum dan berbahasa Inggris. Setiap entri berisi kolom Nama Parfum, Merek, *Notes* atau Aroma, Deskripsi Aroma dan Url Gambar. Kolom yang digunakan untuk pelatihan model adalah kolom “Deskripsi” dan “Notes”. Deskripsi ini bersifat teks bebas yang mencerminkan karakteristik aroma, sedangkan kolom *Notes* berisi teks yang menjelaskan bahan baku aroma seperti jeruk bergamot (*bergamot*), kapulaga (*cardamom*), akar wangi (*vetiver*) dan sebagainya.

3.3 Metode Analisis Data

Bagian ini memuat penjelasan secara lengkap dan terinci tentang metode dan alat yang digunakan untuk analisis data.

3.4 Alur Penelitian

Alur penelitian ini dikembangkan dengan pendekatan CRISP-DM (*Cross Industry Standard Process for Data Mining*) [37]. CRISP-DM merupakan metodologi standar yang banyak digunakan dalam proyek penambangan data karena bersifat fleksibel, iteratif, dapat diterapkan lintas domain. Tahapan-tahapan pada CRISP-DM diadaptasi secara sederhana dan diuraikan pada poin-poin di bawah ini disesuaikan dengan konteks penelitian ini. Kelima tahapan tersebut adalah : 1) *Business Understanding* atau pemahaman masalah, 2) *Data Understanding and Preparation* atau pemahaman dan persiapan data, 3) *Modelling* atau pemodelan, dan 4) *Evaluation* atau evaluasi yang ditunjukkan pada Gambar 1.



Gambar 4 Alur CRISP-DM

3.4.1 Pemahaman Masalah

Pada tahap ini, ditetapkan bahwa permasalahan utama adalah bagaimana memberikan rekomendasi parfum berbasis teks yang relevan secara semantik berdasarkan deskripsi dan aroma. Pengguna sering kali kesulitan memilih parfum karena karakteristik aroma yang bersifat subjektif dan deskriptif.

3.4.2 Pemahaman dan Persiapan Data

Basis data yang digunakan dalam penelitian ini adalah *Perfume Recommendation Dataset* berasal dari repositori publik bernama *Kaggle* (<https://www.kaggle.com/datasets/nandini1999/perfume-recommendation-dataset>). Basis data terdiri dari 2191 entri parfum dan berbahasa Inggris. Setiap entri berisi kolom “Nama Parfum”, “Merek”, “Notes atau Aroma”, “Deskripsi Aroma” dan “Url Gambar”. Kolom yang digunakan untuk pelatihan model adalah kolom “Deskripsi” dan “Notes/Aroma”. Deskripsi ini bersifat teks bebas yang mencerminkan karakteristik aroma, sedangkan kolom “Notes/Aroma” berisi teks yang menjelaskan bahan baku aroma seperti jeruk bergamot (*bergamot*), kapulaga (*cardamom*), akar wangi (*vetiver*) dan sebagainya.

Sebelum digunakan dalam proses pemodelan, data dilakukan praproses dasar yaitu mengecek data kosong (*null*) dan reduksi data kosong. Pada tahap praproses dilakukan reduksi sebanyak 80 baris dikarenakan terdapat data kosong didalamnya. Sehingga data yang digunakan untuk pelatihan sebanyak 2111 data. Tidak dilakukan *lowercasing*, *stemming* maupun penghapusan *stopwords*, mengingat model bahasa yang digunakan sudah dilatih dengan struktur kalimat lengkap.

3.4.3 Pemodelan

Representasi semantik dari setiap deskripsi dan aroma parfum dibangun menggunakan model *Sentence Transformer* salah satunya adalah *all-MiniLM-L6-v2* yang merupakan varian ringan dari arsitektur *Bidirectional Transformers for Language Understanding (BERT)*. Namun sebagai langkah awal, penelitian akan

diuji coba dengan model klasik yaitu *Term Frequency-Inverse Document Frequency (TF-IDF)* dan *Word2Vec* lalu model yang lebih canggih yaitu *Siamese-BERT*. Model TF-IDF fokus ke kata kunci, sedangkan Word2Vec adalah model yang memahami makna semantik pada tingkat kata.

Ketiga model akan dilihat performanya untuk menangani masalah pemahaman dalam tingkat kalimat. Perbandingan performa ini sangat penting untuk menunjukkan secara kuantitatif seberapa besar peningkatan performa dari model klasik ke model yang lebih baru. Hal ini diperlukan untuk menjustifikasi adopsi model *deep learning* berbasis *transformer* yang lebih kompleks untuk mengatasi deskripsi parfum yang subjektif dan metaforis.

Model *Siamese-BERT* pada penelitian ini tentunya akan di eksplorasi lebih lanjut karena model tersebut memang dirancang secara khusus untuk pemecahan masalah (*task*) dalam representasi kalimat (*sentence embedding*). Dalam hal ini tentu model lebih mengerti makna semantik lebih baik pada level kalimat.

3.4.3.1 Representasi Vektor Leksikal: *Term Frequency-Inverse Document (TF-IDF)*

TF-IDF digunakan sebagai *baseline* leksikal. Model ini bekerja dengan menghitung bobot statistik untuk setiap kata dalam deskripsi parfum relatif terhadap seluruh korpus. Kelemahan utama yang akan diuji di sini adalah ketidakmampuan TF-IDF dalam menangani fenomena *vocabulary mismatch* (misalnya: model gagal menghubungkan kata "floral" dengan "mawar" karena tidak ada kesamaan karakter teks).

3.4.3.2 Representasi Vektor Statis: *Word2Vec*

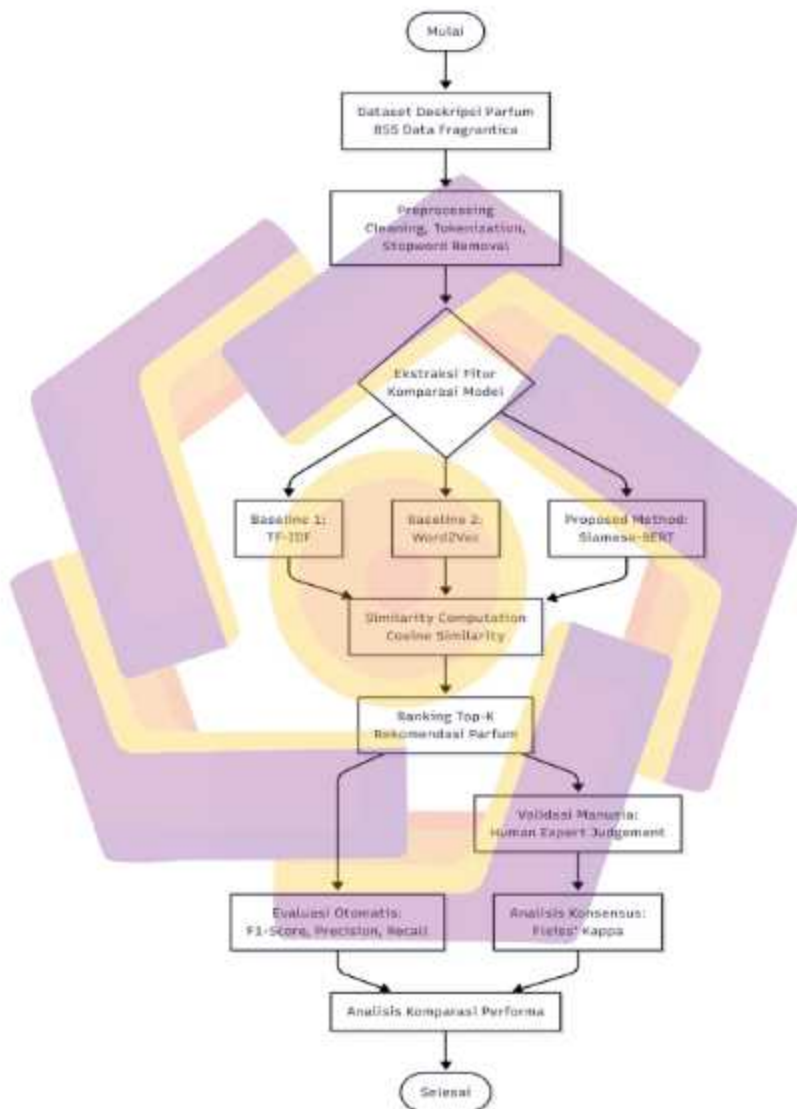
Word2Vec diterapkan untuk menghasilkan *embedding* pada level kata dengan menangkap relasi sintaktis dan semantik lokal. Pada penelitian ini, representasi dokumen parfum diperoleh dengan melakukan rata-rata (*averaging*) terhadap vektor kata-kata yang menyusun deskripsinya. Model ini berfungsi sebagai pembandingan untuk melihat apakah model *transformer* (*Siamese-BERT*) memberikan peningkatan signifikan dibanding *embedding* statis yang tidak sensitif pada konteks urutan kata.

3.4.3.3 Arsitektur *Siamese-BERT* (S-BERT)

Model inti penelitian ini menggunakan struktur siamese yang terdiri dari dua jaringan BERT identik yang berbagi bobot parameter yang sama (*shared weights*).

- a. Proses *Encoding*: Dua teks (misal: kueri pengguna dan deskripsi parfum) dimasukkan ke masing-masing *twin network*.
- b. *Pooling Strategy*: Mengingat BERT menghasilkan output pada level token, dilakukan strategi *Mean Pooling* untuk mentransformasikan output tersebut menjadi satu vektor tunggal berdimensi tetap (*sentence embedding*).
- c. *Similarity Computation*: Kedekatan semantik diukur menggunakan *Cosine Similarity* pada ruang vektor tersebut. Pendekatan ini memungkinkan sistem memahami nuansa metaforis dalam deskripsi parfum (misal: memahami aroma "segar seperti embun pagi" meskipun tidak mengandung kata kunci spesifik).

3.4.4 Evaluasi



Gambar 5 Alur Evaluasi Komparatif Relevansi Rekomendasi

Pada Gambar 5 menyajikan alur evaluasi sistem yang dirancang untuk mengukur efektivitas arsitektur *Siamese-BERT*. Proses evaluasi dimulai dengan ekstraksi fitur dari teks deskripsi parfum menggunakan tiga pendekatan berbeda: *TF-IDF* sebagai *baseline* leksikal, *Word2Vec* sebagai *baseline* statis, dan *Siamese-BERT* sebagai model usulan. Hasil dari masing-masing model kemudian dihitung kemiripan semantiknya menggunakan *Cosine Similarity* untuk menghasilkan daftar rekomendasi. Validasi dilakukan secara ganda, yakni melalui metrik komputasi (*F1-Score*) dan validasi manusia menggunakan *Fleiss' Kappa* untuk memastikan bahwa relevansi yang dihasilkan model selaras dengan persepsi sensoris manusia. *Fleiss' Kappa* digunakan untuk menilai model dalam merekomendasikan parfum dengan kesepakatan penilai (ahli). Penilai ahli yang ditunjuk adalah kolektor parfum dan penggemar parfum yang memang sudah sangat familiar dengan parfum dan aroma.

Jika model *Siamese-BERT* terbukti memberikan performa dan relevansi yang lebih baik berdasarkan performa *Confusion Matrix* dan *Fleiss' Kappa*, maka lebih jauh model *Siamese-BERT* akan dilakukan *inferencing* dengan menerapkan beberapa skenario misalnya mengganti arsitektur model dengan *embedding* yang lebih besar untuk meningkatkan performa lebih tinggi.

Skenario ini bertujuan untuk membuktikan hipotesis bahwa representasi vektor dari model *transformer* mampu menangkap nuansa deskripsi parfum lebih baik daripada model tradisional.

Evaluasi dilakukan melalui 3 skenario dimana ketiga skenario tersebut untuk memastikan objektivitas dan relevansi sistem:

- a. Skenario A (*Benchmarking* Otomatis): Membandingkan skor kemiripan antara *Siamese-BERT* dengan TF-IDF dan Word2Vec menggunakan metrik *F1-Score* dan korelasi *Pearson*. Tujuannya adalah memvalidasi secara statistik model mana yang paling konsisten dengan pola data.
- b. Skenario B (*Human-Centered Validation*): Melibatkan pakar/pengguna untuk menilai 5-10 rekomendasi teratas. Skor kesepakatan diukur menggunakan Fleiss' Kappa. Tujuannya adalah memastikan bahwa "kemiripan" menurut mesin selaras dengan persepsi penciuman manusia.
- c. Skenario C (Kueri Kompleksitas): Pengujian sistem menggunakan kueri bahasa alami yang bersifat subjektif (contoh: "*parfum untuk kencan romantis di musim hujan*"). Tujuannya untuk menguji ketahanan model terhadap teks yang tidak memiliki kata kunci aroma eksplisit.

Instrumen penilaian pada Tabel x dirancang untuk mengukur relevansi kueri dengan hasil yang ditunjukkan oleh sistem yang divalidasi oleh kolektor dan parfum *enthusiast*.

Tabel 3 Instrumen Penilaian Untuk Penilai (Kolektor dan *Enthusiast*)

Instruksi untuk Evaluator: Mohon berikan penilaian terhadap tingkat relevansi antara Kueri dengan Hasil Rekomendasi yang diberikan oleh sistem. Penilaian menggunakan skala Likert 1-4.

No	Kueri	Model	Hasil Rekomendasi	Hasil dari Penilai (Skala Likert 1 s.d 4)
1	Parfum yang cocok untuk hari hujan, yang wanginya menyenangkan dan 'cozy' seperti sedang membaca buku di dekat jendela	<i>TF-IDF</i>	Parfum A	1 s.d 4
2	Saya butuh wangi untuk liburan ke pantai di Bali. Harus segar, tropis, sedikit wangi kelapa, tapi tetap terasa mewah	<i>Word2Vec</i>	Parfum B	1 s.d 4

Definisi bobot skor berdasarkan skala likert:

- Skor 1 (Sangat Tidak Relevan): Karakter aroma pada produk sama sekali tidak mencerminkan deskripsi pada kueri.
- Skor 2 (Kurang Relevan): Terdapat kemiripan kecil pada salah satu *notes*, namun secara keseluruhan nuansa aroma meleset dari kueri.
- Skor 3 (Relevan): Karakter aroma produk sesuai dengan mayoritas deskripsi kueri secara semantik.
- Skor 4 (Sangat Relevan): Produk sangat akurat merepresentasikan seluruh aspek deskripsi (*top*, *heart*, dan *base notes*) yang diminta dalam kueri.

BAB 4

HASIL PENELITIAN DAN PEMBAHASAN

4.1 Pemahaman Masalah

Pada tahap ini, ditetapkan bahwa permasalahan utama adalah bagaimana memberikan rekomendasi parfum yang relevan secara semantik berdasarkan deskripsi aroma. Pengguna sering kali kesulitan memilih parfum karena karakteristik aroma yang bersifat subjektif dan deskriptif.

4.2 Pemahaman dan Persiapan Data

Basis data yang digunakan dalam penelitian ini adalah *Perfume Recommendation Dataset* berasal dari repositori publik bernama *Kaggle* (<https://www.kaggle.com/datasets/nandini1999/perfume-recommendation-dataset>). Data terdiri dari 2191 entri parfum dan berbahasa Inggris. Setiap entri berisi kolom "Nama Parfum", "Merek", "Notes atau Aroma", "Deskripsi Aroma" dan "Url Gambar". Contoh 2 baris data pada basis data *Perfume Recommendation Dataset* ditunjukkan pada Tabel 4.

Tabel 4 Contoh 2 baris data parfum

Nama	Merek	Deskripsi	Aroma	Image URL
Tihota Eau de Parfum	Indult	<i>Rapa Nui for sugar, Tihota is, quite simply, The One. The One that will call to you every moment you're not smelling it,</i>	<i>Vanilla bean; musks</i>	

		<p><i>The One that you've only had hints of in other vanilla perfumes, The One that lasts and lasts, The One that has perfectly captured the essence of the pure nature of fresh vanilla beans and has harnessed it beyond your wildest dreams. It inspires obsessive devotion with its fragrance of smooth, vanilla bean pods dipped in honeyed water and left to steep. The result is the pure magic, an unbridled vanilla, sweet, raw and achingly desirable. Of course, Tihota isn't the first perfume to focus on the dark sweetness of vanilla... perfumery is</i></p>		
--	--	--	--	--

		<p><i>filled with vanillas, but this smells like the dream of a master perfumer who was obsessed with vanilla and was finally given free rein to worship the note with no apologies. It is, without a sliver of a doubt, the finest pure vanilla we've ever had the honor of carrying, and it's so beautiful it pains us to put the bottle down."</i></p>		
Sola Parfume	Di Ser	<p><i>A tribute to the expanse of space extending from the sky to the flower and fruit laden earth below; Sola; the newest offering from DI SER and perfumer Yasuyuki Shinohara; showcases the natural flora of Hokkaido;</i></p>	<p><i>Lavender; Yuzu; Lemongrass; Magnolia; Geranium; Jasmine; Frankincense; Myrrh</i></p>	

		<p><i>Japan with locally sourced and sustainably harvested ingredients from DI SERs own gardens. Starting with aromatic notes of freshly harvested Hokkaido lavender; lemongrass and yuzu; the uplifting scent moves to a floral middle of Japanese rose; magnolia; geranium and jasmine before settling on a spiritual bed of frankincense and myrrh. The spiritual and meditative scent evokes an imagery of a peaceful universe in complete harmony with the bounties of nature and of peace and contentment with its</i></p>		
--	--	--	--	--

Strukturisasi data sangat penting dalam praproses mempersiapkan data. Langkah ini menerapkan pelabelan eksplisit agar model mengerti dan mengenali bahwa kolom “deskripsi” memuat deskripsi parfum, sedangkan “notes” adalah aroma parfum. Hal ini penting dilakukan untuk meningkatkan kedalaman semantik teks. Proses ini ditunjukkan pada Gambar 8.

```
import pandas as pd

def construct_semantic_string(row):
    # 1. ambil data dan pilih data yang tidak null
    brand = str(row['brand']).strip() if pd.notna(row['brand']) else "nama brand"
    name = str(row['name']).strip() if pd.notna(row['name']) else "nama nama"

    # mendapatkan kolom "notes" atau "aroma" sangat susah, karena ini penting
    notes = str(row['notes']).strip() if pd.notna(row['notes']) else ""

    # deskripsi panjang dimasukkan diakhir
    description = str(row['description']).strip() if pd.notna(row['description']) else ""

    # 2. strukturisasi dengan label eksplisit
    # menggunakan teknik "prompting" sederhana untuk labeling,
    # kita beri tahu model mana yang "merk", dan mana yang "nama".
    # format: "Merek: [X], Nama: [Y], Aroma: [Z], Deskripsi: [A]"
    structured_text = f"Merek: {brand}, Nama: {name}, Komposisi Aroma: {notes}, detail: {description}"

    return structured_text

# Memanggil ke dataframe
# "data" adalah dataframe yang sudah di-load
data['structured_text'] = data.apply(construct_semantic_string, axis=1)

# Cek hasil
print("Contoh data terstruktur:")
print(data['structured_text'])
```

Gambar 8 Proses Strukturisasi Data

Contoh data yang telah terstruktur ditunjukkan pada Tabel 5. Kolom digabungkan menjadi satu baris namun setiap isi kolom merepresentasikan nama kolom sehingga dataset jauh lebih terstruktur dan kaya.

Tabel 5 Contoh Data yang Telah Terstruktur

Merek: Indult. **Nama:** Tihota Eau de Parfum. **Komposisi Aroma:** Vanilla bean; musks. **Detail:** Rapa Nui for sugar; Tihota is; quite simply; The One. The One that will call to you every moment you're not smelling it; The One that you've only had hints of in other vanilla perfumes; The One that lasts and lasts; The One that has perfectly captured the essence of the pure nature of fresh vanilla beans and has harnessed it beyond your wildest dreams. It inspires obsessive devotion with its fragrance of smooth; vanilla bean pods dipped in honeyed water and left to steep. The result is the pure magic; an unbridled vanilla; sweet; raw and

achingly desirable. Of course; Tihota isn't the first perfume to focus on the dark sweetness of vanilla... perfumery is filled with vanillas; but this smells like the dream of a master perfumer who was obsessed with vanilla and was finally given free rein to worship the note with no apologies. It is; without a sliver of a doubt; the finest pure vanilla we've ever had the honor of carrying; and it's so beautiful it pains us to put the bottle down.

4.3 Pemodelan

Model *baseline Siamese-BERT* yang telah dibangun dibandingkan dengan model *baseline TF-IDF* dan *Word2Vec*. Proses perbandingan yang telah dilakukan memberikan gambaran awal performa 3 model tersebut dalam menangani permasalahan representasi kalimat (*sentence embedding*). Agar urutan perbandingan lebih rapi, proses pembangunan model diurutkan dari *TF-IDF*, *Word2Vec*, dan terakhir adalah *Siamese-BERT*. Model *TF-IDF* ditunjukkan pada Gambar 9, model *Word2Vec* ditunjukkan pada Gambar 10, dan model *Siamese-BERT* ditunjukkan pada Gambar 11.

```
tfidf = TfidfVectorizer(stop_words='english')
data['notes'] = data['notes'].fillna("")
data['combined'] = data['description'] + " " + data['notes']
tfidf_matrix = tfidf.fit_transform(data['combined'])

def rekomendasi_tfidf(query, top_k=3):
    query_vec = tfidf.transform([query])
    similarity = cosine_similarity(query_vec, tfidf_matrix).flatten()
    top_indices = similarity.argsort()[::-1].topk(top_k)[1]
    return data.iloc[top_indices][['name', 'brand', 'description', 'notes']], similarity[top_indices]
```

Gambar 9 model *TF-IDF*

Model *TF-IDF* telah berhasil dibangun seperti yang ditunjukkan pada Gambar 9. Pada proses pembangunan ini dataset telah dilakukan proses ekstraksi fitur menggunakan *TfidfVectorizer*. Fitur yang digunakan merupakan gabungan dari kolom "Deskripsi" dan "Aroma" untuk memperkaya informasi kontekstual produk.

Setelah representasi vektor terbentuk, tingkat kemiripan antara *query* pengguna dengan dokumen dihitung menggunakan metrik *Cosine Similarity*. Hasil dari perhitungan tersebut kemudian diurutkan berdasarkan skor tertinggi untuk memberikan rekomendasi sebanyak *k* item yang paling relevan bagi pengguna.

```
w2v_model = Word2Vec(
    sentences=tokenized_deskripsi,
    vector_size=300,
    window=5,
    min_count=1,
    workers=4,
    epochs=50
)

# 5. Fungsi Rekomendasi
def rekomendasi_w2v(query, top_k=3):
    query_vec = vectorize_w2v(query).reshape(1, -1)
    similarity = cosine_similarity(query_vec, w2v_model).flatten()
    top_indices = similarity.argsort()[::-1][0:top_k]
    return data.iloc[top_indices][['name', 'brand', 'description', 'notes']], similarity[top_indices]
```

Gambar 10 Model *Word2Vec*

Model *Word2Vec* pada Gambar 10 dibangun menggunakan ukuran *vector_size=300* agar lebih kaya fitur serta menggunakan *epoch=50* agar model mengerti konteks teks semantik-nya. Lalu model akan di evaluasi rekomendasinya menggunakan nilai cosine similarity berdasarkan top *k* rekomendasi.

```
model = SentenceTransformer('all-MiniLM-L6-v2')
embedding_matrix = model.encode(data['combined']).tolist()

def rekomendasi_siamese_bert(query, top_k=3):
    query_vec = model.encode([query])
    similarity = cosine_similarity(query_vec, embedding_matrix).flatten()
    top_indices = similarity.argsort()[::-1][0:top_k]
    return data.iloc[top_indices][['name', 'brand', 'description', 'notes']], similarity[top_indices]
```

Gambar 11 Model *Siamese-BERT*

Tahap pemodelan yang ditunjukkan pada Gambar 11 menggunakan arsitektur

Sentence-BERT (SBERT) untuk menghasilkan representasi vektor padat (*dense embeddings*) dari data teks parfum. Model yang diimplementasikan adalah *pre-trained model 'all-MiniLM-L6-v2'*, yang dipilih karena efisiensinya dalam menyeimbangkan kecepatan komputasi dan akurasi semantik. Berbeda dengan pendekatan berbasis frekuensi kata (seperti TF-IDF), SBERT menggunakan struktur *Siamese Network* untuk memetakan teks ke dalam ruang vektor berdimensi tinggi sehingga teks dengan makna serupa akan berada pada posisi yang berdekatan. Proses diawali dengan melakukan *encoding* pada seluruh data gabungan deskripsi dan *notes* menjadi sebuah matriks *embedding*. Ketika sistem menerima masukan pengguna (*query*), masukan tersebut dikonversi menjadi vektor menggunakan model yang sama, kemudian tingkat kemiripan dihitung menggunakan metrik *Cosine Similarity* untuk menghasilkan rekomendasi berdasarkan kedekatan makna semantik, bukan sekadar kecocokan kata kunci.

4.4 Evaluasi

Penelitian ini menerapkan strategi evaluasi bertingkat (*multi-stage evaluation*). Untuk menganalisa performa model leksikal dibandingkan dengan *Siamese-BERT*, model dibangun secara berurutan mulai dari *TF-IDF*, *Word2Vec*, lalu *Siamese-BERT*. Model yang telah dibangun memiliki waktu proses pelatihan yang berbeda-beda seperti ditunjukkan pada Tabel 6.

Tabel 6 Waktu Pelatihan Model

Metode	Waktu Proses Pelatihan (detik)
<i>TF-IDF</i>	0

<i>Word2Vec</i>	10
<i>Siamese-BERT</i>	240

Pada Tabel 6 ditunjukkan pelatihan model *TF-IDF* lebih cepat yaitu 0 detik, *Word2Vec* 10 detik dan yang paling membutuhkan waktu adalah *Siamese-BERT* yaitu 4 menit. Proses pelatihan model *Siamese-BERT* membutuhkan waktu sedikit lebih lama dibandingkan dengan *TF-IDF* dan *Word2Vec* dikarenakan arsitektur *Siamese-BERT* memiliki 2 encoder dan 2 bobot yang bekerja secara paralel untuk menghitung kemiripan semantik antar kalimat.

Setelah proses pelatihan, model sudah siap untuk proses pengujian dengan memasukkan kueri yang diinginkan. Pada proses pengujian ini model telah diuji dengan memasukkan 2 kueri berbahasa Inggris yaitu:

"A perfume that is suitable for rainy days, which has a calming and 'cozy' scent like reading a book by the window" - Parfum yang cocok untuk hari hujan, yang wanginya menenangkan dan 'cozy' seperti sedang membaca buku di dekat jendela.

dan

"I needed a scent for a beach holiday in Bali. It needed to be fresh, tropical, with a hint of coconut, but still feel luxurious" - Saya butuh wangi untuk liburan ke pantai di Bali. Harus segar, tropis, sedikit wangi kelapa, tapi tetap terasa mewah.

Kedua kueri di atas sengaja dibuat untuk menguji model dalam menangkap teks yang metaforis dan semantik. Kueri tersebut dibuat berdasarkan deskripsi parfum dan keinginan pengguna yang secara umum sangat metaforis. Hasil

rekomendasi ditampilkan sebanyak *top-3* atau 3 rekomendasi teratas menurut *TF-IDF*, *Word2Vec* dan *Siamese-BERT*. Hasil yang ditampilkan memuat 3 komponen hasil parfum yaitu nama parfum, merek parfum, deskripsi, dan aroma/*notes*. Kompleksitas ini memberikan ruang untuk mempertimbangkan dan menilai hasil rekomendasi apakah sesuai atau tidak secara keseluruhan. Hal tersebut dimaksudkan karena parfum sangat subjektif, nama parfum seringkali menipu pengguna. Nama parfum seringkali terkesan segar namun nyatanya tidak.

Kueri pertama yaitu “Saya butuh wangi untuk liburan ke pantai di Bali. Harus segar, tropis, sedikit wangi kelapa, tapi tetap terasa mewah” (*I needed a scent for a beach holiday in Bali. It needed to be fresh, tropical, with a hint of coconut, but still feel luxurious*). Ketiga model memberikan hasil seperti ditunjukkan pada Tabel 7 dan Nilai *Cosine Similarity* ditunjukkan pada Tabel 8.

Tabel 7 Rekomendasi Parfum dari Ketiga Model

(Berdasarkan Jarak Vektor Terdekat)

Kueri	Model	Rekomendasi Parfum ke-1	Rekomendasi Parfum ke-2	Rekomendasi Parfum ke-3
Saya butuh wangi untuk liburan ke pantai di Bali. Harus segar, tropis, sedikit wangi kelapa, tapi tetap terasa mewah	TF-IDF	<i>Coco Coco Eau de Parfum</i> by Coqui Coqui - Kelapa, aroma rumput	<i>Beach Hut Man Eau de Parfum</i> by Amouage - mint, bunga jeruk, tanaman galbanum, akar wangi, lumut, <i>ivy</i> , nilam, getah pohon Myrrh, kayu kering	<i>Soleil Blanc Eau de Parfum</i> by Tom Ford Private Blend - Getah pohon konifer (<i>amber</i>), kelapa, aroma bunga laut
	Word2Vec	Monyette Paris Eau de	Not a Perfume Eau	Totally White Eau de Parfum by

		Parfum Spray by Monyette Paris - Bunga gardenia, bunga lily perancis, bunga vanili	de Parfum by Juliette Has a Gun - <i>Ambroxan, Cetalox</i>	Parle Moi de Parfum - Lilac, Wisteria, Hawthorn
Siamese-BERT		Vittoria Apuana Eau de Parfum by Profumi del Forte - Sweet Brazilian orange; Sicilian mandarin; Coconut; Tahitian Gardenia; Banana; Madagascan Vanilla; Amber	<i>Coco Coco Eau de Parfum</i> by Coqui Coqui - Kelapa, aroma rumput	Beachwood Eau de Parfum by Costamor - Bergamot; Mandarin; Neroli; Jasmine; Coconut; Cedar wood; Sandalwood; Amber; Musk

Tabel 8 Nilai *Cosine Similarity* Rekomendasi Berdasarkan 3 Model

Kueri	Model	Rekomendasi Parfum ke-1	Rekomendasi Parfum ke-2	Rekomendasi Parfum ke-3
Saya butuh wangi untuk liburan ke pantai di Bali. Harus segar, tropis, sedikit wangi kelapa, tapi	TF-IDF	<i>Coco Coco Eau de Parfum</i> by Coqui Coqui - 0,22	<i>Beach Hut Man Eau de Parfum</i> by Amouage - 0,14	<i>Soleil Blanc Eau de Parfum</i> by Tom Ford Private Blend - 0,14
	Word2Vec	Monyette Paris Eau de Parfum Spray by Monyette Paris	Not a Perfume Eau de Parfum	Totally White Eau de Parfum by

tetap terasa mewah		- 0,99	by Juliette Has a Gun - 0,99	Parle Moi de Parfum - 0,99
	Siamese- BERT	Vittoria Apuana Eau de Parfum by Profumi del Forte - 0,68	<i>Coco Coco</i> <i>Eau de</i> <i>Parfum</i> by Coqui Coqui - 0,67	Beachwood Eau de Parfum by Costamor - 0,64

Berdasarkan hasil rekomendasi 3 model pada Tabel 7 dan nilai *cosine similarity* pada Tabel 8, model *TF-IDF* merekomendasikan 1 parfum yaitu *Coco Coco Eau de Parfum by Coqui Coqui* dengan nilai *cosine similarity* sebesar 0,22. Melihat hasil rekomendasi tersebut, *TF-IDF* berhasil merekomendasikan sesuai dengan kueri dimana aroma yang diberikan adalah aroma “kelapa” dan “aroma rumput”. Aroma tersebut cenderung manis kelapa dan segar sesuai kueri meskipun nilai *cosine similarity* nya rendah.

Berbeda dengan *TF-IDF*, *Word2Vec* memberikan hasil dengan nilai *cosine similarity* yang sangat tinggi yaitu 0,99 dengan parfum *Monyette Paris Eau de Parfum Spray by Monyette Paris*. Namun, melihat *notes* aroma parfum tersebut tidak menunjukkan hasil yang sesuai dengan kueri “kelapa” dan “segar” karena rekomendasi dominan aroma “bunga”. Dalam hal ini *Word2Vec* dikatakan belum berhasil memberikan rekomendasi yang cocok karena nilai *cosine* tinggi bukan menjadi jaminan.

Model *Siamese-BERT* dengan nilai *cosine similarity* 0,68 merekomendasikan *Vittoria Apuana Eau de Parfum by Profumi del Forte* dengan

aroma kelapa dan jeruk brazil. Hasilnya terlihat sesuai dengan kueri baik secara *notes* aroma maupun deskripsi.

Kueri kedua yaitu “Parfum yang cocok untuk hari hujan, yang wanginya menyenangkan seperti sedang membaca buku di dekat jendela” (*A perfume that is suitable for rainy days, the scent is calming and 'cozy' like reading a book by the window*). Ketiga model memberikan hasil seperti ditunjukkan pada Tabel 7 dan Nilai *Cosine Similarity* ditunjukkan pada Tabel 8.

Tabel 9 Rekomendasi Parfum dari Ketiga Model (*Berdasarkan Jarak Vektor Terdekat*)

Kueri	Model	Rekomendasi Parfum ke-1	Rekomendasi Parfum ke-2	Rekomendasi Parfum ke-3
Parfum yang cocok untuk hari hujan, yang wanginya menyenangkan dan 'cozy' seperti sedang membaca buku di dekat jendela”	TF-IDF	<i>Black Phantom Eau de Parfum</i> by Killian - Rum; sugar cane; dark chocolate; coffee; caramel; almond; heliotrope; sandalwood	<i>Ambra Nobile Eau de Parfum</i> by Nobile 1942 - Neroli; Bergamot; Ambergriis; Labdanum; Patchouly; Tonka Bean; Sandalwood; Cedarwood; Vetiver	<i>Biblioteca de Babel Eau de Parfum</i> by Fueguia 1833 - Cedar; cabreuva (woody note);
	Word2Vec	Not A Perfume Superdose Eau de Parfum by Juliette Has a Gun - Ambroxan, Cetalox	Muskara Phero J Eau de Parfum by Fueguia 1833 - Musk	Soleil Blanc Eau de Parfum by TOM FORD Private Blend - Amber; coconut; beachy floral notes

	Siamese-BERT	Erdenstern Eau de Parfum by April Aromatics - Botanical musk accord; botanical ambergris accord; tonka bean; vetiver; tabac; cacao; opoponax.	Nightingale Extrait de Parfum by Zoologist - Bergamot; lemon; saffron; Japanese plum blossom; rose; violet; oud; patchouli; sandalwood; moss; frankincense; white musk; labdanum; ambergris	Koiame Eau de Parfum by J-Scent - Apple; peach; bergamot; mandarin; lemon; cassis; green notes; rose; jasmine; magnolia; violet; water lily; musk; cedar; moss
--	--------------	---	---	--

Tabel 10 Nilai *Cosine Similarity* Rekomendasi Berdasarkan 3 Model

Kueri	Model	Rekomendasi Parfum ke-1	Rekomendasi Parfum ke-2	Rekomendasi Parfum ke-3
Parfum yang cocok untuk hari hujan, yang wanginya menenangkan dan 'cozy' seperti sedang membaca buku di dekat jendela"	TF-IDF	Black Phantom Eau de Parfume by Killian - 0,14	Ambra Nobile Eau de Parfum by Nobile 1942 - 0,10	Biblioteca de Babel Eau de Parfum by Fueguia 1833 - 0,09
	Word2Vec	Not A Perfume Superdose Eau de Parfum by Juliette Has a Gun - 0,99	Muskara Phero J Eau de Parfum by Fueguia 1833 - 0,98	Soleil Blanc Eau de Parfum by TOM FORD Private Blend - 0,99
	Siamese-BERT	Erdenstern Eau de Parfum by April Aromatics -	Nightingale Extrait de Parfum by Zoologist -	Koiame Eau de Parfum by J-Scent -

		0,62	0,62	Apple; peach; bergamot; mandarin; lemon; cassis; green notes; rose; jasmine; magnolia; violet; water lily; musk; cedar; moss - 0,60
--	--	------	------	--

Berdasarkan hasil rekomendasi 3 model pada Tabel 9 dan nilai *cosine similarity* pada Tabel 10, model *TF-IDF* merekomendasikan parfum *Black Phantom EDP by Kilian* dengan nilai *cosine similarity* sebesar 0,14. Melihat hasil rekomendasi tersebut, *TF-IDF* kurang tepat dalam merekomendasikan sesuai dengan kueri dimana aroma yang diinginkan adalah nuansa aroma “hujan” sedangkan aroma yang direkomendasikan justru memiliki aroma kopi dan makanan.

Berbeda dengan *TF-IDF*, *Word2Vec* memberikan hasil dengan nilai *cosine similarity* yang sangat tinggi yaitu 0,99 dengan parfum *Not A Perfume Superdose Eau de Parfum by Juliette Has a Gun*. Namun, melihat *notes* aroma parfum tersebut tidak menunjukkan hasil yang sesuai dengan kueri nuansa “hujan” karena rekomendasi dominan aroma tebal dari *amber*.

Model *Siamese-BERT* dengan nilai *cosine similarity* 0,62 merekomendasikan *Erdenstern Eau de Parfum April Aromatics* dengan aroma

“musk” dan “ambergris”. Hasilnya terlihat cukup baik sesuai dengan kueri berdasarkan deskripsi.

Berdasarkan 2 kueri yang telah dilakukan ketiga model di evaluasi dengan *confusion matrix*, hasilnya ditunjukkan pada Tabel 11.

Tabel 11 *Confusion Matrix* Model Berdasarkan Kueri

Kueri	Model	Precision	Recall	F1 Score
Parfum yang cocok untuk hari hujan, yang wanginya menenangkan dan 'cozy' seperti sedang membaca buku di dekat jendela	TF-IDF	0,67	0,50	0,57
	Word2Vec	0,00	0,00	0,00
	Siamese-BERT	0,67	0,50	0,57
Saya butuh wangi untuk liburan ke pantai di Bali. Harus segar, tropis, sedikit wangi kelapa, tapi tetap terasa mewah	TF-IDF	0,67	0,50	0,57
	Word2Vec	0,00	0,00	0,00
	Siamese-BERT	1,00	0,75	0,86

Berdasarkan Tabel 11 pada kueri “*Parfum yang cocok untuk hari hujan, yang wanginya menenangkan dan 'cozy' seperti sedang membaca buku di dekat jendela*” model TF-IDF dan Siamese-BERT memberikan nilai *confusion matrix* yang sama. Hal ini membuktikan bahwa kedua model bisa dikatakan cukup baik dalam performa walaupun rekomendasi parfum yang diberikan jauh berbeda, dimana Siamese-BERT memberikan hasil yang lebih mendekati kueri.

Pada kueri kedua yaitu “*Saya butuh wangi untuk liburan ke pantai di Bali. Harus segar, tropis, sedikit wangi kelapa, tapi tetap terasa mewah*”, model *Siamese-BERT* memberikan nilai *F1-Score* lebih tinggi dibandingkan 2 metode lainnya. Dapat dikatakan bahwa *Siamese-BERT* memberikan hasil rekomendasi yang lebih baik sesuai dengan kueri.

Nilai *Confusion Matrix* pada model *Word2Vec* dapat membuktikan bahwa model tersebut gagal memberikan rekomendasi, namun perlu dikaji ulang dengan mengecek kembali tahap praproses apakah dilakukan dengan benar atau tidak. Terutama pada proses penggunaan *pooling* dan *lowercasing, tokenizing*.

Berdasarkan analisis *Confusion Matrix* pada Tabel 11 didapatkan model *Siamese-BERT* menunjukkan performa terbaik dalam analisis representasi kalimat rekomendasi parfum. Namun, untuk memberikan gambaran bagaimana sistem dinilai dari sisi penilai maka dilakukan analisis *Fleiss' Kappa*. Analisis ini dilakukan dengan menilai hasil rekomendasi parfum yang dihasilkan dari ketiga model berdasarkan 2 kueri. Hasil penilaian oleh anotorator (1 kolektor dan 2 parfum antusias) ditunjukkan pada Lampiran 1. Hasil *Fleiss' Kappa* telah dihitung yang hasilnya ditunjukkan pada Tabel 12.

Tabel 12 Analisis *Fleiss' Kappa*

Model dan Kueri	Rata-rata Kesepakatan	Ekspektasi Acak	Fleiss' Kappa	Interpretasi
<i>TF-IDF</i> (Hujan)	0.44	0.28	0.22	<i>Fair agreement</i>
<i>Word2Vec</i> (Hujan)	0.56	0.27	0.40	<i>Moderate agreement</i>
<i>Siamese-BERT</i> (Hujan)	0.67	0.29	0.54	<i>Moderate agreement</i>
<i>TF-IDF</i> (Pantai)	0.61	0.26	0.47	<i>Moderate agreement</i>

<i>Word2Vec</i> (Pantai)	0.50	0.28	0.31	<i>Fair agreement</i>
<i>Siamese-BERT</i> (Pantai)	0.78	0.29	0.69	<i>Substantial agreement</i>

Berdasarkan Tabel 10 hasil perhitungan *Fleiss' Kappa* terhadap penilaian tiga evaluator (dua *enthusiast* dan satu kolektor) terhadap rekomendasi parfum dari tiga model (*TF-IDF*, *Word2Vec*, dan *Siamese-BERT*), ditemukan variasi tingkat kesepakatan antar penilai yang cukup signifikan. Model *TF-IDF* menunjukkan tingkat kesepakatan yang berada pada kategori *fair* hingga *moderate*, dengan nilai Kappa berkisar antara 0.22 hingga 0.47. *Word2Vec* cenderung menghasilkan kesepakatan yang lebih rendah, terutama pada kueri rekomendasi parfum untuk ke pantai, dengan nilai Kappa hanya mencapai 0.31, yang masih berada dalam kategori *fair agreement*.

Sebaliknya, model *Siamese-BERT* menunjukkan performa paling konsisten, dengan nilai Kappa mencapai 0.54 untuk kueri rekomendasi parfum saat hujan dan 0.69 untuk kueri rekomendasi parfum untuk ke pantai, yang masing-masing masuk dalam kategori *moderate* dan *substantial agreement*. Temuan ini mengindikasikan bahwa model berbasis representasi semantik seperti *Siamese-BERT* lebih mampu menangkap nuansa deskriptif metaforis dalam kueri subjektif, sehingga menghasilkan rekomendasi yang lebih selaras dengan persepsi para evaluator.

Nilai *Fleiss' Kappa* sebesar 0,54 yang menunjukkan tingkat kesepakatan moderat (*moderate agreement*) mengindikasikan adanya variasi persepsi di antara para ahli dan kolektor parfum terhadap hasil rekomendasi. Hal ini wajar terjadi

dalam domain olfaktori (penciuman), di mana interpretasi terhadap deskripsi teks seperti 'mewah', 'segar', atau 'hangat' sangat dipengaruhi oleh latar belakang subjektif dan pengalaman sensorik masing-masing individu.

Berdasarkan hasil evaluasi *Confusion Matrix* dan *Fleiss Kappa* Hal ini menunjukkan bahwa *Siamese-BERT* mampu memahami konteks dan makna kalimat secara menyeluruh, tidak hanya berdasarkan frekuensi atau kedekatan vektor kata. Dalam lingkup seperti parfum, di mana deskripsi dan aroma seringkali bersifat subjektif, metaforis, dan naratif pendekatan model yang dapat memahami teks semantik secara lebih mendalam menjadi sangat penting. *Siamese-BERT* dapat menangkap nuansa deskriptif yang kompleks tersebut, sehingga menghasilkan rekomendasi yang lebih personal dan sesuai ekspektasi.

Untuk melihat lebih jauh performa model *Siamese-BERT* dalam menangani rekomendasi parfum berbasis teks, model ini telah diterapkan 4 model arsitektur yaitu *all-mpnet-base-v2*, *all-MiniLM-L6-v2*, *paraphrase-mpnet-base-v2*, dan *multi-qa-mpnet-base-cos-v1* untuk melihat arsitektur mana yang paling memberikan performa terbaik.

Perbandingan ini berdasarkan *ground-truth* yang dibuat menggunakan metode *Pool-Based Relevance Judgment*. Mengingat besarnya basis data (>2000 baris), tidak memungkinkan untuk melabeli seluruh dokumen secara manual. Oleh karena itu, diambil *Top-5* hasil prediksi dari model untuk setiap kueri, kemudian melakukan anotasi manual (*human verification*) untuk menentukan relevansi dokumen tersebut. Dokumen yang terverifikasi relevan kemudian dijadikan sebagai

Ground Truth final. Keempat arsitektur model *Siamese-BERT* diuji secara serentak berdasarkan 20 kueri yang ditunjukkan pada Tabel 13. Selanjutnya, Hasil pengujian keempat model ditunjukkan pada Tabel 14.

Tabel 13 Dua puluh Kueri untuk Menguji 4 arsitektur Model *Siamese-BERT*

Kueri 1	<i>I needed a scent for a beach holiday in Bali. It had to be fresh, tropical, with a hint of coconut, but still feel luxurious</i>
Kueri 2	<i>A perfume that is suitable for rainy days, the scent is calming and 'cozy' like reading a book by the window</i>
Kueri 3	<i>A woody and spicy fragrance for formal occasions</i>
Kueri 4	<i>An unisex fragrance for everyday use</i>
Kueri 5	<i>A fragrance for daily use</i>
Kueri 6	<i>A fresh citrusy fragrance for sport</i>
Kueri 7	<i>A fragrance that consist lavender notes</i>
Kueri 8	<i>Fragrance for Man with sandalwood and vanilla notes. It must be good longevity</i>
Kueri 9	<i>Fragrance full of rose for women</i>
Kueri 10	<i>Unisex fragrance with amber notes</i>
Kueri 11	<i>Fragrance for summer day, dominated with mint and citrus notes</i>
Kueri 12	<i>A women fragrance with luxury tuberose notes</i>
Kueri 13	<i>Fragrance for office occasion that smell light</i>
Kueri 14	<i>A sensual fragrance for dating</i>
Kueri 15	<i>A fragrance for vacation in a beach</i>

Kueri 16	<i>Fragrance with high sillage</i>
Kueri 17	<i>Unisex fragrance suits for all season</i>
Kueri 18	<i>Fragrance that smells like a rain</i>
Kueri 19	<i>A fragrance that smell clean after shower</i>
Kueri 20	<i>A powdery fragrance for unisex</i>

Tabel 14 Perbandingan Performa Model Siamese-BERT pada Rekomendasi Parfum Berbasis Teks

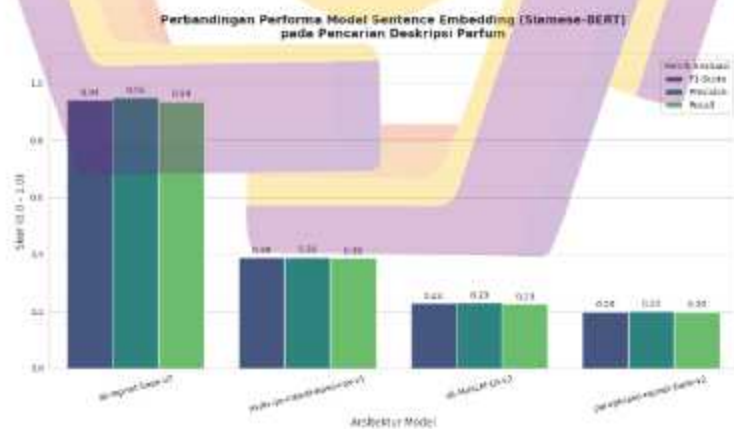
Model	Mean-F1 Score	Mean Precislion	Mean Recall
<i>all-mpnet-base-v2</i>	0,94	0,95	0,93
<i>multi-qa-mpnet-base-cos-v1</i>	0,38	0,39	0,38
<i>all-MiniLM-L6-v2</i>	0,22	0,23	0,22
<i>paraphrase-mpnet-base-v2</i>	0,19	0,20	0,19

Hasil eksperimen menunjukkan bahwa arsitektur *all-mpnet-base-v2* memberikan performa tertinggi dibandingkan model lainnya. Hal ini dapat diatribusikan pada jumlah parameter yang lebih besar (110M) dan strategi *pre-training* yang lebih kompleks, yang memungkinkan model menangkap dependensi semantik yang lebih dalam pada deskripsi parfum yang bersifat naratif. Namun, perlu dicatat adanya *trade-off* antara akurasi dan efisiensi komputasi; model ini memerlukan waktu inferensi yang lebih lama dibandingkan *all-MiniLM-L6-v2*.

Temuan menarik terlihat pada rendahnya performa model spesialis *multi-qa* (F1: 0.38) dan *paraphrase* (F1: 0.19). Hal ini mengindikasikan bahwa karakteristik

pencarian parfum bukanlah masalah *Question Answering* (tanya-jawab fakta) maupun deteksi parafrase murni. Pencarian parfum lebih menitikberatkan pada kemiripan semantik global (suasana, nuansa, komposisi). Oleh karena itu, model *all-mpnet-base-v2* yang dilatih untuk *general semantic similarity* terbukti paling *robust*.

Berdasarkan *F1-Score* rata-rata tertinggi (0.94), serta keseimbangan antara *Precision* (0.95) dan *Recall* (0.93), maka model *all-mpnet-base-v2* ditetapkan sebagai model terbaik yang akan diimplementasikan pada sistem akhir. Tingginya *Precision* menunjukkan sistem minim memberikan rekomendasi yang tidak relevan, sementara tingginya *Recall* menunjukkan sistem mampu menarik hampir seluruh variasi parfum yang sesuai kueri. Hasil Perbandingan Performa Model Siamese-BERT pada Rekomendasi Parfum Berbasis Teks ditunjukkan pada Gambar 12.



Gambar 12 Perbandingan Performa Model *Siamese-BERT* pada Rekomendasi Parfum Berbasis Teks

Berdasarkan observasi terhadap hasil rekomendasi, ditemukan beberapa kasus di mana model memberikan hasil yang kurang relevan, terutama pada kueri yang sangat singkat atau menggunakan terminologi yang bersifat kiasan (metafora). Sebagai contoh, kueri 'aroma hujan di pagi hari' terkadang menghasilkan rekomendasi yang didominasi oleh aroma laut (*aquatic*) namun melewatkan nuansa tanah (*earthy*) yang seharusnya ada. Hal ini menunjukkan bahwa meskipun *Siamese-BERT* unggul dalam menangkap kemiripan semantik umum, pemahaman terhadap nuansa spesifik dalam industri parfum masih memerlukan *fine-tuning* lebih lanjut pada korpus data yang lebih spesifik atau penggunaan teknik pembobotan yang lebih baik.

BAB 5 PENUTUP

5.1 Kesimpulan

Penelitian ini berhasil mengimplementasikan *Siamese-BERT* untuk rekomendasi parfum berbasis teks. *Siamese-BERT* mampu memahami konteks kolom dataset deskriptif dan aroma yang bersifat subjektif dan naratif, seperti aroma “segar”, “lembut”, “maskulin”, “kayu” atau “balsam” yang sulit ditangkap oleh pendekatan berbasis statistik biasa.

Hasil eksperimen perbandingan model baseline leksikal (*TF-IDF* dan *Word2Vec*) dengan *Siamese-BERT*, menunjukkan bahwa meskipun model *Word2Vec* menghasilkan skor *cosine similarity* yang lebih tinggi secara numerik, namun model *Siamese-BERT* justru memberikan hasil rekomendasi yang paling relevan berdasarkan persepsi pengguna dengan nilai *Cosine Similarity* rata-rata sebesar 0,67 dan *F1-Score* sebesar 0,86. Evaluasi kuantitatif menggunakan *Fleiss Kappa* oleh evaluator (parfum *enthusiast* dan kolektor parfum) juga mendukung bahwa *Siamese-BERT* lebih bisa memahami dan menangkap makna semantik dari deskripsi yang diberikan. Hal ini ditunjukkan nilai rata-rata kesepakatan sebesar 0,78 (*substansial agreement*). Perbandingan antara ketiga model yang telah dibangun dan dievaluasi memperkuat bahwa kemampuan *Siamese-BERT* dalam memahami keseluruhan kalimat dan konteks menjadikannya pendekatan yang lebih unggul untuk sistem rekomendasi berbasis teks, terutama pada domain yang bersifat deskriptif dan emosional seperti parfum.

Lebih jauh, setelah dilakukan proses optimasi Siamese-BERT menggunakan 4 jenis arsitektur berbeda yaitu *all-mpnet-base-v2*, *all-MiniLM-L6-v2*, *paraphrase-mpnet-base-v2*, dan *multi-qa-mpnet-base-cos-v1*. Hasil menunjukkan *Siamese-BERT* dengan arsitektur *all-mpnet-base-v2* memberikan hasil yang luar biasa dengan skor *F1* rata-rata sebesar 94%. Hal ini menunjukkan bahwa kedalaman arsitektur model berpengaruh signifikan terhadap akurasi pencarian semantik pada domain parfum. Model *all-mpnet-base-v2* (110M parameter) mampu menangkap nuansa kompleks catapan aroma (*olfactory notes*) jauh lebih baik dibandingkan *all-MiniLM-L6-v2* (22M parameter) yang merupakan versi distilasi. Model *mp-net* cocok digunakan untuk akurasi yang lebih tinggi dengan komputasi tinggi sedangkan *miniLM* lebih direkomendasikan untuk aplikasi real-time dengan sumber daya komputasi yang terbatas.

5.2 Saran

Berdasarkan keterbatasan dan temuan yang diperoleh dalam penelitian ini, terdapat beberapa rekomendasi untuk pengembangan penelitian di masa mendatang yaitu untuk meningkatkan presisi hasil rekomendasi pada peringkat teratas (*top-K*), penelitian selanjutnya dapat mempertimbangkan penerapan arsitektur *two-stage retrieval*. Tahap pertama menggunakan *Siamese-BERT (Bi-Encoder)* untuk menyaring kandidat secara cepat, kemudian diikuti oleh tahap kedua menggunakan *Cross-Encoder* untuk melakukan *re-ranking*. Pendekatan ini diharapkan dapat menutupi kelemahan *Bi-Encoder* dalam menangkap interaksi kata yang lebih detail antara kueri pengguna dan deskripsi parfum, sehingga menghasilkan rekomendasi yang jauh lebih baik.

Berikut 3 aspek yang dapat dijadikan pertimbangan untuk saran penelitian selanjutnya:

1. Aspek pertama yaitu Aspek Metodologi (Model & Training)

a. Penerapan *Domain-Adaptive Fine-Tuning*: Penelitian ini masih menggunakan model *pre-trained* umum. Penelitian selanjutnya disarankan untuk melakukan proses *fine-tuning* lanjutan pada model *mpnet-base* menggunakan dataset spesifik parfum dengan fungsi *Triplet Loss* atau *Cosine Similarity Loss*. Hal ini bertujuan agar model lebih sensitif terhadap terminologi spesifik industri wewangian lebih dalam lagi (seperti *sillage*, *dry-down*, atau *olfactory families*).

b. Eksplorasi *Large Language Models* (LLM) Generatif: Mengembangkan sistem dari sekadar *retrieval* (pencarian) menjadi *Retrieval-Augmented Generation* (RAG). Penelitian selanjutnya dapat mengintegrasikan model *embedding* ini dengan LLM (seperti GPT atau *Llama*) untuk tidak hanya mencari parfum yang mirip, tetapi juga "menjelaskan" alasan kemiripan tersebut kepada pengguna dalam bahasa alami.

2. Aspek yang kedua adalah Aspek Data (Multimodal & Ekspansi)

a. Pengembangan *Multimodal Retrieval*: Persepsi terhadap parfum tidak hanya dipengaruhi oleh deskripsi teks, tetapi juga visual (bentuk botol, warna cairan). Penelitian masa depan dapat menggabungkan *text embedding* dengan *image embedding* (misalnya menggunakan CLIP) untuk membangun sistem pencarian hibrid (teks dan citra).

b. Integrasi Ulasan Pengguna (*User Reviews*): Dataset saat ini hanya mengandalkan deskripsi resmi (*official description*). Menambahkan data ulasan pengguna (*user-generated content*) dapat menangkap persepsi subjektif yang lebih nyata, seperti ketahanan aroma (*longevity*) dan daya sebar (*projection*) yang sering kali berbeda dari klaim pemasaran.

3. Aspek Implementasi Sistem

a. Pengembangan Menjadi Sistem Rekomendasi Hibrida: Karena fokus penelitian ini adalah *Content-Based Filtering* (Pencarian Semantik), saran pengembangan logis berikutnya adalah menggabungkannya dengan metode *Collaborative Filtering* (berbasis riwayat interaksi pengguna). Penggabungan ini akan membentuk sistem rekomendasi yang utuh dan terpersonalisasi.

b. Optimasi Efisiensi Komputasi untuk Skalabilitas: Jika jumlah dataset parfum bertambah menjadi jutaan, pencarian *Cosine Similarity* standar akan menjadi lambat. Disarankan untuk menerapkan metode pengindeksan vektor yang lebih efisien seperti FAISS (*Facebook AI Similarity Search*) atau HNSW (*Hierarchical Navigable Small World*) untuk mempercepat waktu respons pencarian (*inference time*) pada skala industri.

DAFTAR PUSTAKA

- [1] STATISTA, "Market Insight : Fragrance Worldwide." Accessed: Dec. 12, 2024. [Online]. Available: <https://www.statista.com/outlook/cmo/beauty-personal-care/fragrances/worldwide>
- [2] R. Lutan and C. Badica, "Personalized Quiz-Based Perfume Recommender System Using Social Data," 2023, pp. 30–43. doi: 10.1007/978-3-031-41774-0_3.
- [3] A. Holschuh, "Make Sense of Scents: Fragrance Trends Now and Beyond | Mintel," <https://www.mintel.com/insights/beauty-and-personal-care/make-sense-of-scents-current-and-future-fragrance-trends/>. Accessed: Oct. 19, 2025. [Online]. Available: <https://www.mintel.com/insights/beauty-and-personal-care/make-sense-of-scents-current-and-future-fragrance-trends/>
- [4] X. Ma, M. Li, and X. Liu, "Advancements in recommender systems: a comprehensive analysis based on data, algorithms, and evaluation," *IJIO (International Journal of Industrial Optimization)*, vol. 6, no. 1, pp. 47–70, Oct. 2024, doi: 10.1298/ijio.v6i1.11107.
- [5] B. Sabiri, A. Khtira, B. El Asri, and M. Rhanoui, "Hybrid Quality-Based Recommender Systems: A Systematic Literature Review," *J. Imaging*, vol. 11, no. 1, Jan. 2025, doi: 10.3390/jimaging11010012.
- [6] Adline Twice Tobing *et al.*, "Men's Perfume Recommendation System Using Analytic Hierarchy Process (AHP) and Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) Method," *International Journal of Science, Technology & Management*, vol. 4, no. 6, pp. 1552–1559, 2023, doi: 10.46729/ijstm.v4i6.1002.
- [7] I. Ramadhani, A. Zakir, and A. M. Elhanafi, "Implementasi Metode Weighted Product Untuk Menentukan Alternatif Parfume Terbaik (Studi Kasus : Quality Parfume)," *International Journal of Data Science and Visualization*, vol. 3, no. 1, pp. 1–11, 2024.

- [8] N. S. Agashe, "Product Recommender Chat Bot," *International Journal of Engineering Research & Technology (IJERT)*, vol. 10, no. 06, pp. 702–705, 2021.
- [9] D. Wu and X. Li, "A Systematic Literature Review of Financial Product Recommendation Systems," *Information (Switzerland)*, vol. 16, no. 3, Mar. 2025, doi: 10.3390/INFO16030196.
- [10] M. Park and J. Oh, "Enhancing E-Commerce Recommendation Systems with Multiple Item Purchase Data: A Bidirectional Encoder Representations from Transformers-Based Approach," *Applied Sciences (Switzerland)*, vol. 14, no. 16, Aug. 2024, doi: 10.3390/app14167255.
- [11] Y. Su, Y. Li, and Z. Zhang, "Two-Tower Structure Recommendation Method Fusing Multi-Source Data," *Electronics (Switzerland)*, vol. 14, no. 5, Mar. 2025, doi: 10.3390/ELECTRONICS14051003.
- [12] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," May 2019, [Online]. Available: <http://arxiv.org/abs/1810.04805>
- [13] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," Aug. 2019, [Online]. Available: <http://arxiv.org/abs/1908.10084>
- [14] J. Y. Yeo, S. Youm, and K. S. Shin, "Enhancing Interprofessional Communication in Healthcare Using Large Language Models: Study on Similarity Measurement Methods with Weighted Noun Embeddings," *Electronics (Switzerland)*, vol. 14, no. 11, Jun. 2025, doi: 10.3390/ELECTRONICS14112240.
- [15] A. Bayani, A. Ayotte, and J. N. Nikiema, "Transformer-Based Tool for Automated Fact-Checking of Online Health Information: Development Study," *JMIR Infodemiology*, vol. 5, 2025, doi: 10.2196/56831.
- [16] W. S. Kim, S. Lim, G. W. Kim, and S. M. Choi, "Extracting Implicit User Preferences in Conversational Recommender Systems Using Large

- Language Models," *Mathematics*, vol. 13, no. 2, Jan. 2025, doi: 10.3390/MATH13020221.
- [17] R. Lutan and C. Badica, "Personalized Quiz-Based Perfume Recommender System Using Social Data," 2023, pp. 30–43. doi: 10.1007/978-3-031-41774-0_3.
- [18] J. Kim, K. Oh, and B. S. Oh, "An NLP-Based Perfume Note Estimation Based on Descriptive Sentences," *Applied Sciences (Switzerland)*, vol. 14, no. 20, 2024, doi: 10.3390/app14209293.
- [19] S. Ahmad *et al.*, "Hybrid recommender system for personalized healthcare: integration of content-based and collaborative filtering approaches," *International Journal of System Assurance Engineering and Management*, 2025, doi: 10.1007/s13198-025-02965-4.
- [20] R. Widayanti, M. Heru, R. Chakim, C. Lukita, U. Rahardja, and N. Lutfiani, "Improving Recommender Systems using Hybrid Techniques of Collaborative Filtering and Content-Based Filtering," *Journal of Applied Data Sciences*, vol. 4, no. 3, pp. 289–302, 2023.
- [21] K. R. Putra and M. A. Rachman, "Perbandingan Metode Content-based, Collaborative dan Hybrid Filtering pada Sistem Rekomendasi Lagu," *MIND Journal*, vol. 9, no. 2, pp. 179–193, Dec. 2024, doi: 10.26760/mindjournal.v9i2.179-193.
- [22] Z. Li *et al.*, "A natural language processing approach to support biomedical data harmonization: Leveraging large language models," *PLoS One*, vol. 20, no. 7 July, Jul. 2025, doi: 10.1371/JOURNAL.PONE.0328262.
- [23] J. Y. Yeo, S. Youm, and K. S. Shin, "Enhancing Interprofessional Communication in Healthcare Using Large Language Models: Study on Similarity Measurement Methods with Weighted Noun Embeddings," *Electronics (Switzerland)*, vol. 14, no. 11, Jun. 2025, doi: 10.3390/ELECTRONICS14112240.

- [24] R. S. de Oliveira and E. G. S. Nascimento, "Analysing similarities between legal court documents using natural language processing approaches based on transformers," *PLoS One*, vol. 20, no. 4 April, Apr. 2025, doi: 10.1371/JOURNAL.PONE.0320244.
- [25] S. Han, L. Shi, and F. Tsui, "Enhancing semantical text understanding with fine-tuned large language models: A case study on Quora Question Pair duplicate identification," *PLoS One*, vol. 20, no. 1, Jan. 2025, doi: 10.1371/JOURNAL.PONE.0317042.
- [26] A. Krzywicki, M. Bain, and W. Wobkef, "Editorial: Natural language processing for recommender systems," B. Goethals, Ed., Adelaide: *Frontiers in Big Data*, Mar. 2025. doi: 10.5220/0001422901350145.
- [27] I. Syahdana, R. Hidayat, and M. Khadafi, "A Web-Based Laptop Purchase Recommendation Model Using Natural Language Processing (NLP) on Marketplace Reviews," *Journal of Artificial Intelligence and Software Engineering (J-AISE)*, vol. 4, no. 2, p. 116, Nov. 2024, doi: 10.30811/jaise.v4i2.6133.
- [28] A. Vaswani *et al.*, "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Nips, pp. 5999–6009, 2017.
- [29] C. F. Maudhy, I. Sulaiman, and Murlida. Eva, "Pengaruh Perbedaan Konsentrasi Pelarut (Solvent) terhadap Daya Tahan Secara Sensori," *Jurnal Ilmiah Mahasiswa Pertanian*, vol. 8, no. 2, pp. 7–14, 2023.
- [30] W. Hikmah, Y. Aisyah, and F. -, "Pengaruh Perbandingan Konsentrasi Minyak Nilam Terhadap Kualitas Parfum Secara Sensori," *Jurnal Ilmiah Mahasiswa Pertanian*, vol. 8, no. 1, pp. 254–261, 2023, doi: 10.17969/jimfp.v8i1.23912.
- [31] M. Chairunnisa, K. Vidya Kusala, U. Putri Lestari, R. Fatoni, and K. Harismah, "PEMBUATAN DAN EVALUASI PARFUM EAU DE TOILETTE DARI MINYAK NILAM (Pogostemon cablin) DAN ADAS (Foeniculum vulgare) DENGAN WEWANGIAN BUAH," *Simposium Nasional RAPI XXI-2023 FT UMS*, pp. 1–6, 2023.

- 
- [32] D. Ortiz Martes, E. Gunderson, C. Neuman, and N. N. Kachouie, "Transformer Models for Paraphrase Detection: A Comprehensive Semantic Similarity Study," *Computers*, vol. 14, no. 9, Sep. 2025, doi: 10.3390/COMPUTERS14090385.
- [33] G. Salton, A. Wong, and C. S. Yang, "A Vector Space Model for Automatic Indexing," *Commun. ACM*, vol. 18, no. 11, pp. 613–620, Nov. 1975, doi: 10.1145/361219.361220;PAGE:STRING:ARTICLE/CHAPTER.
- [34] S. J. J. Guesné, T. Hanser, S. Werner, S. Boobier, and S. Scott, "Mind your prevalence!," *J. Cheminform.*, vol. 16, no. 1, Dec. 2024, doi: 10.1186/S13321-024-00837-W.
- [35] G. Zeng, "Invariance Properties and Evaluation Metrics Derived from the Confusion Matrix in Multiclass Classification," *Mathematics*, vol. 13, no. 16, Aug. 2025, doi: 10.3390/MATH13162609.
- [36] G. Sivaramakrishnan, M. Almuqahwi, S. Ansari, M. Lubbad, E. Alagawawy, and K. Sridharan, "Assessing the power of AI: a comparative evaluation of large language models in generating patient education materials in dentistry," *BDJ Open*, vol. 11, no. 1, Dec. 2025, doi: 10.1038/S41405-025-00349-1.
- [37] P. Chapman *et al.*, "CRISP-DM 1.0 Step-by-step data mining guide," DaimlerChrysler, 1999.

Lampiran 1 Tampilan Form Fieiss' Kappa dan Hasil Anotator



← Preview mode

U. Pahlawan 100% [Coba rekomendasi ini](#)

Formulir Evaluasi Kualitas Sistem Rekomendasi Parfum Berbasis AI

Informasi Responden

Nama *

Your phone *

Phone *

Program Studi (jangan di ubah)

Program Rekomendasi (jangan di ubah)

Program Rekomendasi (jangan di ubah)

Back Next [Lanjut ke halaman berikutnya](#) Page 2 of 4 [Close form](#)

© 2023. All rights reserved. [Privacy Policy](#) [Terms of Service](#)

3. Aplikasi
100% Zoom

Link dan Informasi Produk berisikan 2 Matri (T-01, T-02), dan
Survei (SWT) dengan penjabaran: "Produk yang sudah siap dan siap yang
menyebutkan mengenai fitur yang dapat digunakan pada 8 jenis produk

Produk yang sudah siap dan siap yang mengenai informasi dan siap yang
 sedang memiliki fitur dan informasi. Model T-01 dan informasi dan 3
 produk berikut

Link dan Informasi berisikan data:

1. Black Panther Tea & Perfume by Elita

Notes: Four eggs come with chocolate, coffee, oatmeal, almond, vegetable,
 & chocolate.

Description: Through a glass, daily Black Panther is a whimsy mood perfume
 that will please just about everyone. Its fragrance has a dangerous glitter to them.
 The glass of perfume is sure to get you going, sending sparks of excited
 notes, caramel, honey, and vanilla all the way up. Once the honey (honey
 here) comes, a little perfume comes together to become a wonderfully soft, truly
 addictive that comes as if caramel and you do the dip to the very rich.
 All about all the ingredients, Black Panther is not like or even that
 through with notes of chocolate, honey, and almond coffee. Black Panther is
 also a perfume that is sure to get you going. Imagine a cup of perfume with a
 soft, light feel. The one making a good and relaxing, truly soft with a drop of
 salt on the side to give you more, delicate, elegant of perfume, each perfume
 chocolate, vanilla, and so on. In the end, you'll see the paper vanilla smell of
 your look and the little almond notes of perfume on each table. Now imagine
 the undertone of a really good chocolate that smells like salty butter and sweet
 cream. It's a really good and better than that. It was been released online.

c. Preview mode



Tabel Hasil Penilaian Anotator

Kueri	Model	Hasil Rekomendasi	Nilai Enthusiast 1	Nilai Enthusiast 2	Kolektor 1
Parfum yang cocok untuk hari hujan, yang wanginya menenangkan dan 'cozy' seperti sedang membaca buku di dekat jendela	TF-IDF	<i>Black Phantom Eau de Perfume by Killian</i>	2	4	4
Parfum yang cocok untuk hari hujan, yang wanginya menenangkan dan 'cozy' seperti sedang membaca buku di dekat jendela	TF-IDF	<i>Ambra Nobile Eau de Parfum by Nobile 1942</i>	3	2	3
Parfum yang cocok untuk hari hujan, yang wanginya menenangkan dan 'cozy' seperti sedang membaca buku di dekat jendela	TF-IDF	<i>Biblioteca de Babel Eau de Parfum by Fuegoia 1833</i>	3	4	4
Parfum yang cocok untuk hari hujan, yang wanginya menenangkan dan 'cozy' seperti sedang membaca buku di dekat jendela	Word2Vec	<i>Not A Perfume Superdose Eau de Parfum by Juliette Has a Gun</i>	4	2	3
Parfum yang cocok untuk hari hujan, yang wanginya menenangkan dan 'cozy' seperti	Word2Vec	<i>Muskara Phero J Eau de Parfum by Fuegoia 1833</i>	4	2	3

Kueri	Model	Hasil Rekomendasi	Nilai Enthusiast 1	Nilai Enthusiast 2	Kolektor 1
sedang membaca buku di dekat jendela					
Parfum yang cocok untuk hari hujan, yang wanginya menenangkan dan 'cozy' seperti sedang membaca buku di dekat jendela	<i>Word2Vec</i>	<i>Soleil Blanc Eau de Parfum by TOM FORD Private Blend</i>	4	1	2
Parfum yang cocok untuk hari hujan, yang wanginya menenangkan dan 'cozy' seperti sedang membaca buku di dekat jendela	<i>Siamese-BERT</i>	<i>Erdenstern Eau de Parfum by April Aromatics</i>	2	4	4
Parfum yang cocok untuk hari hujan, yang wanginya menenangkan dan 'cozy' seperti sedang membaca buku di dekat jendela	<i>Siamese-BERT</i>	<i>Nightingale Extrait de Parfum by Zoologist</i>	4	1	3
Parfum yang cocok untuk hari hujan, yang wanginya menenangkan dan 'cozy' seperti sedang membaca buku di dekat jendela	<i>Siamese-BERT</i>	<i>Koame Eau de Parfum by J-Scent</i>	4	4	4
Saya butuh wangi untuk liburan ke pantai di Bali. Harus segar, tropis,	<i>TF-IDF</i>	<i>Coco Coco Eau de Parfum by Coqui Coqui</i>	2	3	4

Kuerl	Model	Hasil Rekomendasi	Nilai Enthuslast 1	Nilai Enthuslast 2	Kolektor 1
sedikit wangi kelapa, tapi tetap terasa mewah					
Saya butuh wangi untuk liburan ke pantai di Bali. Harus segar, tropis, sedikit wangi kelapa, tapi tetap terasa mewah	<i>TF-IDF</i>	<i>Beach Hut Man Eau de Parfum by Amouage</i>	4	2	3
Saya butuh wangi untuk liburan ke pantai di Bali. Harus segar, tropis, sedikit wangi kelapa, tapi tetap terasa mewah	<i>TF-IDF</i>	<i>Soleil Blanc Eau de Parfum by TOM FORD Private Blend</i>	2	4	4
Saya butuh wangi untuk liburan ke pantai di Bali. Harus segar, tropis, sedikit wangi kelapa, tapi tetap terasa mewah	<i>Word2Vec</i>	<i>Monyette Paris Eau de Parfum Spray by Monyette Paris</i>	3	1	4
Saya butuh wangi untuk liburan ke pantai di Bali. Harus segar, tropis, sedikit wangi kelapa, tapi tetap terasa mewah	<i>Word2Vec</i>	<i>Not a Perfume Eau de Parfum by Juliette Has a Gun</i>	2	1	2
Saya butuh wangi untuk liburan ke pantai di Bali. Harus segar, tropis, sedikit wangi kelapa, tapi tetap terasa mewah	<i>Word2Vec</i>	<i>Totally White Eau de Parfum by Parle Moi de Parfum</i>	2	1	3
Saya butuh wangi untuk liburan ke pantai di Bali. Harus segar, tropis, sedikit wangi kelapa, tapi tetap terasa mewah	<i>Siamese-BERT</i>	<i>Vittoria Apuana Eau de Parfum by Profumi del Forte</i>	4	4	4

Kuerl	Model	Hasil Rekomendasi	Nilai Enthuslast 1	Nilai Enthuslast 2	Kolektor 1
Saya butuh wangi untuk liburan ke pantai di Bali. Harus segar, tropis, sedikit wangi kelapa, tapi tetap terasa mewah	<i>Siamese-BERT</i>	<i>Coco Coco Eau de Parfum by Coqui Coqui</i>	4	3	4
Saya butuh wangi untuk liburan ke pantai di Bali. Harus segar, tropis, sedikit wangi kelapa, tapi tetap terasa mewah	<i>Siamese-BERT</i>	<i>Beachwood Eau de Parfum by Costamor</i>	3	3	4