

Penerapan Metode *Content-Based Filtering* Berbasis *Multilingual Sentence-BERT* untuk Sistem Rekomendasi Menu (Studi Kasus: Depot Mie Gemes)

Larasati Romadhani Yunita Putri¹, Amri Muhaimin², Trimono³

^{1,2,3}*Sains Data, Universitas Pembangunan Nasional "Veteran" Jawa Timur*

¹22083010053@student.upnjatim.ac.id

²amri.muhamin@stat@upnjatim.ac.id

³trimono.stat@upnjatim.ac.id

Corresponding author email: amri.muhamin@stat@upnjatim.ac.id

ABSTRAK

Abstrak— Persaingan industri kuliner yang semakin ketat menuntut pelaku usaha untuk memahami preferensi pelanggan secara lebih mendalam, salah satunya melalui pemanfaatan sistem rekomendasi menu. Penelitian ini berfokus pada pengembangan model *Content-Based Filtering* (CBF) berbasis representasi teks menggunakan *Multilingual Sentence-BERT* (mSBERT) dengan *cosine similarity* untuk menghasilkan rekomendasi menu yang relevan secara semantik. Data yang digunakan berupa deskripsi menu makanan, minuman, dan dimsum dari Depot Mie Gemes di Porong, Sidoarjo, yang memuat karakteristik rasa, cara penyajian, serta komposisi produk. Proses penelitian meliputi tahap *preprocessing* data, ekstraksi fitur menggunakan mSBERT, perhitungan kemiripan antar-menu dengan *cosine similarity*. Evaluasi dilakukan melalui *user preference study* terhadap 20 responden untuk menilai relevansi dan daya tarik rekomendasi. Hasil evaluasi menunjukkan bahwa sistem mampu menghasilkan rekomendasi dengan tingkat penerimaan yang baik, ditunjukkan oleh rata-rata skor penilaian 4.31 dari skala 5. Temuan ini menunjukkan bahwa mSBERT mampu menangkap konteks semantik deskripsi menu secara efektif, sehingga rekomendasi yang dihasilkan lebih sesuai dengan preferensi pengguna. Selain itu, penelitian ini menghasilkan sebuah prototipe sistem rekomendasi berbasis web menggunakan Streamlit, yang menampilkan rekomendasi menu secara interaktif dan mudah dipahami. Sistem ini berpotensi menjadi solusi pendukung keputusan bagi pelaku usaha kuliner dalam menyajikan rekomendasi menu yang lebih relevan bagi pelanggan.

Keywords: *Content-Based Filtering*, *Multilingual-Sentence BERT*, Rekomendasi Menu

I. PENDAHULUAN

Industri ekonomi kreatif telah menjadi salah satu sektor utama dalam perekonomian Indonesia. Menurut Kementerian Pariwisata dan Ekonomi Kreatif, subsektor kuliner, fashion, dan kerajinan tangan berkontribusi rata-rata 75% terhadap Produk Domestik Bruto (PDB) ekonomi kreatif setiap tahunnya [1]. Di antara sektor-sektor tersebut, subsektor kuliner memberikan kontribusi terbesar, sekitar 42% [2]. Angka-angka ini menyoroti peran strategis sektor kuliner dalam mendorong pertumbuhan ekonomi. Namun, dominasi ini juga mencerminkan persaingan yang ketat di industri ini, di mana bisnis harus terus berinovasi untuk mempertahankan minat dan loyalitas pelanggan.

Jumlah usaha makanan dan minuman yang terus bertambah semakin mempertegas tekanan kompetitif ini. Data dari Badan Pusat Statistik (BPS) pada 2024 menunjukkan bahwa jumlah usaha di sektor ini mencapai 4,85 juta pada 2023, meningkat 21,13% dibandingkan 2016 [3]. Pertumbuhan signifikan ini menunjukkan bahwa semakin banyak pelaku bisnis yang masuk ke industri kuliner, sehingga semakin sulit bagi usaha untuk menonjol. Untuk tetap kompetitif, bisnis kuliner tidak hanya

perlu meningkatkan efisiensi operasional tetapi juga mengadopsi teknologi digital yang dapat membantu mereka memahami dan merespons kebutuhan pelanggan dengan lebih baik.

Situasi ini dapat terlihat pada Depot Mie Gemes, salah satu UMKM kuliner yang beroperasi di Sidoarjo. Berdasarkan hasil analisis data transaksi periode Juli–September 2025, usaha ini mengalami penurunan rata-rata penjualan sebesar 0,54% pada bulan Juli–September. Penurunan ini kontras dengan kondisi pada periode April–Juni 2025, ketika penjualan justru meningkat rata-rata sekitar 22%. Pergeseran tren dari pertumbuhan menuju penurunan ini mengindikasikan bahwa persaingan yang semakin padat mulai berdampak pada kemampuan usaha dalam mempertahankan pelanggan.

Untuk menghadapi kondisi tersebut, pemanfaatan teknologi digital menjadi semakin relevan. Salah satu langkah yang dapat dilakukan adalah menerapkan sistem rekomendasi menu, yang tidak hanya membantu pelanggan menemukan pilihan sesuai preferensi mereka, tetapi juga berpotensi meningkatkan nilai transaksi melalui rekomendasi item serupa atau komplementer. Dari sisi bisnis, rekomendasi menu dapat menyederhanakan proses pemilihan dan mendorong pelanggan menambah pembelian ketika ditawarkan produk yang relevan. Penelitian sebelumnya juga menunjukkan bahwa rekomendasi produk komplementer mampu meningkatkan volume pembelian secara signifikan [4].

Studi sebelumnya menunjukkan bahwa pendekatan *Content-Based Filtering* (CBF) dan penggunaan *embedding* cukup efektif dalam sistem rekomendasi. Penelitian [5] misalnya, memanfaatkan TF-IDF dan *cosine similarity* untuk merekomendasikan resep, tetapi hasilnya masih terbatas karena metode tersebut belum mampu menangkap makna semantik secara mendalam. Penelitian lain [6] juga menggunakan TF-IDF untuk rekomendasi menu dan memperoleh 77,3% umpan balik positif, namun tetap menghadapi kendala serupa karena hanya mengandalkan pencocokan kata kunci. Seiring berkembangnya penelitian, mulai muncul tren baru yang mengarah pada penggunaan arsitektur berbasis transformer. Penelitian [7] menunjukkan bahwa *Sentence-BERT* (SBERT) mampu melakukan deduplikasi produk *e-commerce* dengan sangat baik, mencapai akurasi 98,10% dan presisi 100%, dan secara jelas mengungguli metode pencocokan string maupun TF-IDF. Temuan tersebut sejalan dengan penelitian [8] yang memperlihatkan bahwa model berbasis BERT secara konsisten lebih baik dibandingkan *Universal Sentence Encoder* (USE) dan InferSent dalam mengukur kesamaan semantik. Rangkaian hasil penelitian ini memperkuat bahwa *embedding* berbasis mSBERT sangat relevan dan lebih unggul untuk tugas rekomendasi menu yang membutuhkan pemahaman semantik yang lebih mendalam.

Meskipun demikian, metode CBF konvensional seperti TF-IDF masih memiliki sejumlah keterbatasan. TF-IDF hanya mengandalkan bobot kata dan tidak mampu menangkap makna kontekstual dari deskripsi menu, sehingga kualitas rekomendasi yang dihasilkan sering kali kurang optimal [9][10]. Tantangan ini semakin terlihat pada bisnis kuliner skala kecil hingga menengah, karena pada level UMKM umumnya tidak tersedia data pengguna yang teridentifikasi, seperti *user ID*, yang menjadi prasyarat utama bagi penerapan *collaborative filtering*. Selain itu, *collaborative filtering* juga menghadapi masalah *cold-start* ketika jumlah riwayat interaksi masih sangat terbatas. Oleh karena itu, diperlukan pendekatan yang mampu menghasilkan representasi kalimat yang lebih kaya dan bermakna tanpa bergantung pada data pengguna yang terpersonalisasi.

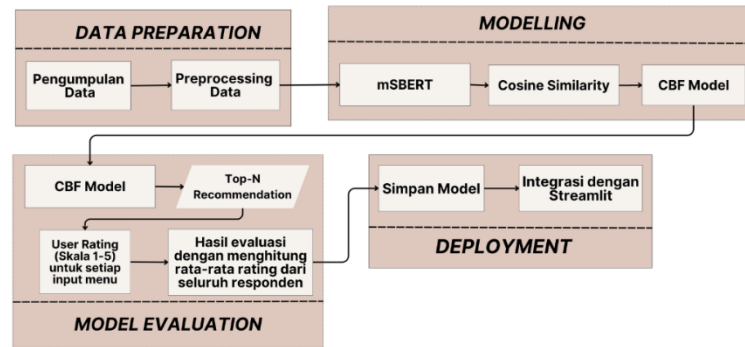
Sebagai solusi, penelitian ini menerapkan sistem rekomendasi menu berbasis *Content-Based Filtering* (CBF) yang diperkuat dengan *embedding Multilingual Sentence-BERT* (mSBERT). BERT memang dirancang untuk memahami konteks bahasa secara mendalam, namun tidak secara langsung menghasilkan representasi kalimat yang efisien untuk perbandingan. SBERT memodifikasi arsitektur BERT dengan menerapkan arsitektur *siamese* agar dapat menghasilkan *embedding* kalimat yang dapat dibandingkan secara efektif menggunakan *cosine similarity*. Varian mSBERT, yang dilatih secara multibahasa, mampu menangkap makna semantik deskripsi menu dalam bahasa Indonesia secara lebih akurat. Dengan *embedding* yang lebih kontekstual, sistem dapat mengukur kesamaan antar-menu secara lebih efektif dan mengatasi keterbatasan CBF konvensional [11][12], [13].

Penerapan *embedding* mSBERT dalam kerangka kerja CBF diharapkan dapat menghasilkan rekomendasi menu yang lebih relevan karena mampu menangkap karakteristik semantik dari setiap deskripsi menu. Berbeda dengan pendekatan berbasis kata kunci yang sering melewatkan makna kontekstual, metode berbasis *embedding* ini dirancang untuk mencerminkan hubungan semantik yang lebih akurat antar-item menu. Dengan rekomendasi yang lebih tepat dan sesuai konteks, pelanggan dapat mengambil keputusan pembelian dengan lebih cepat, sementara pelaku usaha berpotensi meningkatkan kepuasan pelanggan, nilai transaksi, dan daya saing di pasar kuliner. Sehingga, model yang diusulkan tidak hanya memperkuat aspek teknis dari sistem rekomendasi, tetapi juga

menawarkan nilai praktis bagi UMKM yang memiliki keterbatasan sumber daya.

II. METODOLOGI PENELITIAN

Penelitian ini terdiri dari empat tahapan utama yang ditunjukkan pada Gambar 1, yaitu *data preparation*, pemodelan, evaluasi model, serta tahap *deployment*.



Gambar 1. Diagram Alir Penelitian

II.1 Data Preparation

Data dalam studi ini diperoleh dari Depot Mie Gemes yang berlokasi di Kecamatan Porong, Sidoarjo. Data tersebut diperoleh dalam format file *.xlsx* yang berisi informasi seperti nama menu, kategori menu, dan deskripsi menu. Data tersebut distandardisasi melalui langkah-langkah pra-pemrosesan seperti mengubah huruf menjadi huruf besar dan menghapus spasi ganda untuk menghindari gangguan dalam proses pemodelan.

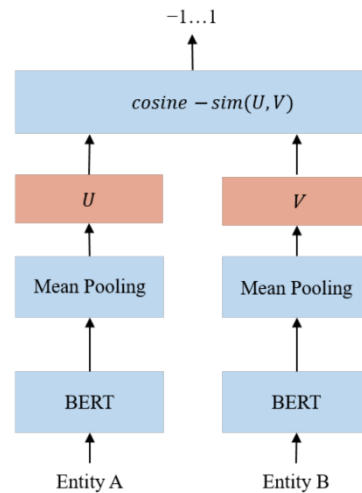
Tabel 1. Data Setelah tahap *Preprocessing*

No	Kode	Kategori	Nama	Deskripsi
0	A001	makanan	mie kalem	mie gurih asin ini menghadirkan cita rasa yang...
...	...	...	...	...
56	C025	dimsum	dimsum ekado	dimsum goreng isi ayam cincang yang dibungkus dengan..

Seperti yang ditunjukkan dalam Tabel 1, Depot Mie Gemes menawarkan 57 item menu yang diklasifikasikan ke dalam tiga kategori utama: makanan, minuman, dan dimsum. Setiap kategori mencakup berbagai item menu. Setiap item menu dijelaskan dalam Bahasa Indonesia dan deskripsi setiap item menu mewakili konten atau karakteristik produk, seperti rasa, bahan, dan tingkat kepedasan.

II.2 Modelling

Pada tahap pemodelan, sistem rekomendasi dikembangkan menggunakan pendekatan *Content-Based Filtering* (CBF) yang diperkuat dengan embedding mSBERT. Berikut adalah bagaimana model mSBERT bekerja untuk mencari kemiripan kalimat:



Gambar 2. Cara Kerja mSBERT

Secara umum, proses pembentukan *sentence embedding* pada mSBERT yang ditunjukkan pada Gambar 2 diawali dengan tahapan *encoding* kalimat yang bekerja dengan prinsip yang sama seperti pada BERT. Pada tahap ini, setiap kalimat diproses untuk menghasilkan embedding kontekstual pada level token, kemudian dilakukan proses agregasi untuk mendapatkan satu vektor representasi yang bersifat tetap bagi keseluruhan kalimat. Setelah embedding diperoleh, tingkat kesamaan antar kalimat dapat dihitung melalui jarak vektor. Rangkaian proses tersebut dijelaskan sebagai berikut [12][14][15]:

- a. Pembentukan *embedding* kontekstual dengan *encoder* BERT

Tahap awal pada SBERT mengikuti proses *embedding* yang digunakan dalam BERT. Setiap kalimat terlebih dahulu diuraikan menjadi token-token menggunakan *WordPiece tokenizer*, dengan penambahan token khusus [CLS] di bagian awal dan [SEP] di akhir. Token-token tersebut kemudian dikonversi menjadi vektor numerik melalui tiga jenis *embedding*: *token embedding*, *segment embedding*, dan *position embedding*. Seluruh *embedding* ini diproses melalui sejumlah lapisan *Transformer encoder* yang memanfaatkan mekanisme *self-attention* untuk memahami hubungan antar kata dalam arah dua arah (*bidirectional*). Pada tahap ini, setiap token memperoleh representasi kontekstual yang mempertimbangkan konteks dari seluruh kata dalam kalimat, sehingga keluaran *encoder* untuk sebuah kalimat $S = [w_1, w_2, \dots, w_n]$, berupa deretan vektor $H = [h_1, h_2, \dots, h_n]$, di mana h_i merupakan vektor kontekstual dari token ke- i .

- b. Agregasi *embedding* token menjadi representasi kalimat (*Mean Pooling*)

Setelah *embedding* token terbentuk, mSBERT melakukan proses agregasi untuk menghasilkan satu representasi tetap yang menggambarkan keseluruhan makna kalimat. Proses ini dilakukan menggunakan teknik *mean pooling*, yaitu menghitung rata-rata dari seluruh *embedding* token. Melalui operasi ini diperoleh vektor U dan V yang menjadi *sentence embedding* berdimensi tetap dan berfungsi sebagai representasi semantik kalimat tersebut. Tahapan ini dilakukan untuk setiap kalimat secara terpisah agar model dapat membandingkan makna antar kalimat secara konsisten.

- c. Mengukur kemiripan antar kalimat menggunakan *cosine similarity*

Tahap berikutnya adalah mengukur tingkat kemiripan semantik antar kalimat berdasarkan *embedding* yang telah dihasilkan. Model mSBERT menggunakan *cosine similarity* sebagai metrik pengukuran, di mana nilai kesamaan dihitung berdasarkan sudut antara dua vektor. Nilai *cosine similarity* berada pada rentang 0 hingga 1. Nilai yang mendekati 1 menunjukkan bahwa kedua kalimat memiliki makna yang sangat serupa, sedangkan nilai yang mendekati 0 atau

negatif menunjukkan perbedaan makna yang signifikan. Dengan proses ini, mSBERT mampu memberikan estimasi kedekatan semantik antar kalimat secara lebih efektif dibandingkan penggunaan *embedding* token secara langsung.

Pada tahap awal pemrosesan, model mSBERT memerlukan data deskripsi menu sebagai sumber utama untuk membentuk representasi semantik setiap item. Deskripsi inilah yang akan melalui proses tokenisasi, *encoding*, dan pembentukan *sentence embedding* pada tahapan sebelumnya. Adapun rincian teks deskriptif yang menjadi input bagi model disajikan pada Tabel II, yang menggambarkan isi konten setiap menu sebelum diproses lebih lanjut oleh mSBERT.

Tabel 2. Input mSBERT

No	Deskripsi
0	mie gurih asin ini menghadirkan cita rasa yang lembut tanpa sensasi pedas, cocok bagi penikmat rasa original. disajikan tanpa kuah, mie ini berpadu sempurna dengan topping ayam kering yang gurih, keripik yang renyah, serta pangsit yang memberikan tekstur garing di setiap suapan.
...	...
56	dimsum goreng isi ayam cincang yang dibungkus dengan kulit tahu dan diikat seperti kantong. setelah digoreng, bagian luarnya menjadi renyah sementara isiannya tetap lembut dan juicy.

Setiap deskripsi diproses oleh BERT untuk menghasilkan vektor padat berukuran tetap yang mewakili makna keseluruhan kalimat. Proses ini melibatkan pemecahan teks menjadi unit kata (subkata), pemetaan ke vektor awal (*embedding* pencarian), dan kemudian memasukkannya ke dalam arsitektur Transformer yang menggunakan perhatian diri untuk menyerap konteks antara kata-kata. Hasil akhir adalah vektor kalimat berdimensi 384, yang diperoleh melalui *mean-pooling* dari vektor output token. Hasil *embedding* disusun ke dalam matriks *embedding* berdimensi $n \times d$, di mana n adalah jumlah item dan d adalah dimensi *embedding*. Untuk mengukur kesamaan antara menu, digunakan *Cosine Similarity*. Berikut adalah rumus *cosine similarity*:

$$\cos_{(A,B)} = \frac{U \cdot V}{|U| \cdot |V|} = \frac{\sum_{i=1}^n U_i \times V_i}{\sqrt{\sum_{i=1}^n (U_i)^2} \times \sqrt{\sum_{i=1}^n (V_i)^2}} \quad (1)$$

Hasil yang diperoleh dari persamaan (1) akan berupa nilai dalam rentang [0,1]. Semakin besar nilai yang diperoleh, semakin mirip kedua vektor yang dibandingkan. Sebaliknya, semakin kecil nilai yang diperoleh, semakin berbeda/tidak mirip kedua vektor yang dibandingkan. Untuk menghasilkan matriks kemiripan, perhitungan dilakukan secara berpasangan pada semua pasangan menu yang tersedia dalam dataset. Hasilnya adalah matriks kemiripan berukuran $n \times n$ di mana setiap sel mencerminkan nilai kemiripan antara dua item. Berdasarkan nilai-nilai ini, sistem akan memilih Top-N item dengan skor tertinggi untuk item yang dipilih oleh pengguna untuk ditampilkan sebagai rekomendasi.

II.2 Model Evaluation

Evaluasi dilakukan melalui *user preference study*, yaitu metode evaluasi subjektif yang menilai kualitas rekomendasi berdasarkan penilaian, persepsi, dan tingkat ketertarikan pengguna terhadap item yang direkomendasikan. Metode ini digunakan karena penelitian ini tidak memiliki data interaksi pengguna sebagai *ground truth*, sehingga evaluasi objektif seperti *precision* atau *recall* tidak dapat diterapkan [16][17]. Teknis evaluasi ini adalah dengan menyebar kuesioner kepada 20 responden. Setiap responden akan menilai 5 hasil rekomendasi dengan input menu yang berbeda beda. Setiap hasil rekomendasi akan dinilai dengan skala 1-5, di mana 1 berarti variasi rekomendasi sangat tidak menarik untuk dibeli dan 5 berarti variasi rekomendasi sangat menarik untuk dibeli. Penyebaran kuesioner menggunakan platform *Google Form* sebagai berikut:

Input menu: Mie Yamin Level 1 *

Rekomendasi Menu Lainnya

Mie Yamin (0) Mie Yamin Lu 2 Mie Kalem

Mie Yamin Lu 3 Mie Yamin Lu 4 Mie Gemes Lu 1

Mie Ayam Ceker Mie Ayam Pangsit Mie Ayam Komploit

Mie Ayam

1 2 3 4 5

Variasi rekomendasinya sangat tidak menarik untuk dibeli

Input Menu: Cwie Mie *

Rekomendasi Menu Lainnya

Mie Ayam Pangsit Mie Ayam Pangsit Pangsit Rebus

Cele Mie Pangsit Mie Kalem Mie Yamin (0)

Cele Mie Bakso Pangsit Goreng Mie Ayam Ceker

Dimsum Siomay Jamur

1 2 3 4 5

Variasi rekomendasinya sangat tidak menarik untuk dibeli

Input Menu: Mie Ayam *

Rekomendasi Menu Lainnya

Mie Ayam Pangsit Mie Ayam Ceker Cele Mie

Mie Yamin (0) Mie Kalem Mie Yamin Lu 2

Mie Yamin (0) Mie Kalem Mie Yamin Lu 2

Cele Mie Pangsit Mie Gemes Lu 1 Mie Ayam Bakso

Mie Yamin Lu 1

1 2 3 4 5

Variasi rekomendasinya sangat tidak menarik untuk dibeli

Input Menu: Thai Tea *

Rekomendasi Menu Lainnya

Green Thai Tea Es Teh Teh Hangat

Es Kopi Susu Es Beras Kencur Es Lemon Tea

Es Jeruk Goodday Cappuccino Es Kopi Susu Rum Begit

Es Kopi Milo

1 2 3 4 5

Variasi rekomendasinya sangat tidak menarik untuk dibeli

Input Menu: Siomay Mozarella *

Rekomendasi Menu Lainnya

Dimsum Siomay Keju Dimsum Tim Tahu Dimsum Siomay Rangkaian

Dimsum Dado Dimsum Bola Keju Mie Kalem

Pangsit Goreng Mie Yamin (0) Dimsum Bola Rambutan

Mie Gemes Lu 1

1 2 3 4 5

Variasi rekomendasinya sangat tidak menarik untuk dibeli

Gambar 3. Tampilan *Google Form* untuk Evaluasi

Berdasarkan Gambar 3 di atas, input menu yang digunakan untuk menilai hasil rekomendasi adalah Mie Yamin Level 1 yang mewakili varian menu mie tidak berkuah, Cwie Mie yang mewakili varian mie berkuah gurih, Mie Ayam yang mewakili varian mie berkuah manis, Thai Tea yang mewakili varian minuman dan Dimsum Siomay Mozarella yang mewakili varian dimsum.

II.3 Deployment

Tahap *Deployment* merupakan langkah terakhir dalam pengembangan sistem rekomendasi, yang berfokus pada penyimpanan dan penyajian model kepada pengguna melalui antarmuka aplikasi. Setelah model rekomendasi dibangun dan dievaluasi, langkah pertama pada tahap ini adalah menyimpan model yang telah dilatih. Dalam konteks *Content-Based Filtering* berbasis mSBERT, penyimpanan ini mencakup matriks kemiripan yang dihasilkan dari transformasi deskripsi menu oleh model mSBERT dan *cosine similarity*. Komponen ini disimpan dalam format CSV untuk pemanggilan yang efisien selama inferensi.

Selanjutnya, model yang disimpan diintegrasikan ke dalam antarmuka aplikasi menggunakan Streamlit, sebuah kerangka kerja *open-source* berbasis Python yang memudahkan pengembangan aplikasi web interaktif dengan cepat [18]. Integrasi dilakukan dengan memuat ulang data serta model kemiripan, kemudian membangun antarmuka yang memungkinkan pengguna memilih sebuah menu dan langsung memperoleh top-10 rekomendasi menu lainnya dalam tampilan visual yang intuitif.

Dalam implementasinya, antarmuka Streamlit tidak hanya menampilkan hasil rekomendasi berdasarkan menu yang dipilih, tetapi juga menyediakan tampilan gambar dan deskripsi menu. Pendekatan ini bertujuan untuk meningkatkan pengalaman pengguna dan memastikan sistem praktis bagi pengguna akhir.

III. HASIL DAN ANALISIS

III.1 Hasil *Modeling* dan Prototipe

a. Hasil *Embedding* mSBERT

Pada tahap pertama, *embedding* mSBERT berhasil dibentuk untuk seluruh item menu seperti yang ditampilkan pada Tabel 3.

Tabel 3. Hasil *Embedding* Tiap Menu

No	Nama	Embeddings (1×384)
1	Mie Kalem	-0.0612982 , -0.06863055, 0.08280633,..., 0.1250856, 0.18859644, 0.05932246
2	Mie Gemes Lv 1	-0.05407445, -0.10197259, 0.05272446, ..., 0.11094202, 0.07293275, 0.0541538
...	...	...
57	Dimsum Ekado	0.02522939, 0.02632627, -0.03934159, ..., -0.12710539, 0.37726292, 0.1522841

Setiap deskripsi menu diproses melalui model mSBERT sehingga menghasilkan vektor berdimensi 384 yang mewakili makna semantik dari deskripsi tersebut. *Embedding* ini bukan sekadar representasi kata secara literal, tetapi mencerminkan hubungan konteks antar kata yang menggambarkan karakteristik menu, misalnya rasa, bahan utama, bumbu, dan gaya penyajian. Perbedaan nilai-nilai vektor antar-item menunjukkan bahwa setiap menu memiliki representasi semantik yang unik, namun tetap memungkinkan pengukuran kedekatannya ketika deskripsinya mirip. Sebagai contoh, *embedding* untuk menu seperti *Mie Kalem* dan *Mie Gemes Level 1* tampak memiliki pola vektor yang relatif serupa, sehingga mengindikasikan bahwa kedua menu tersebut memang memiliki kedekatan dari sisi konsep tekstual, baik dari deskripsi rasa maupun komposisi bahan yang digunakan. Proses *embedding* ini menjadi fondasi utama sistem rekomendasi karena seluruh perhitungan kemiripan antar-menu sepenuhnya bergantung pada kualitas representasi vektor yang dihasilkan.

b. Hasil Matriks Kemiripan

Tahap berikutnya adalah pembentukan matriks kemiripan seperti yang ditunjukkan pada Tabel 4.

Tabel 4. Hasil Matriks Kemiripan antar Menu

Nama	Mie Kalem	Mie Gemes Lv 1	...	Dimsum Ekado
Mie Kalem	1.00	0.91		0.72

Mie Gemes Lv 1	0.91	1.00	...	0.62
...	...	...	...	...
Dimsum Ekado	0.72	0.62	...	1.00

Matriks ini berisi nilai *cosine similarity* antar seluruh pasangan menu dalam dataset. Nilai kemiripan 1.00 pada diagonal menunjukkan bahwa setiap menu memiliki kesamaan sempurna dengan dirinya sendiri, sementara nilai di luar diagonal mencerminkan tingkat kedekatan semantik antar-menu yang berbeda. Pola nilai pada tabel menunjukkan bahwa menu-menu dengan kategori serupa cenderung memiliki tingkat kemiripan tinggi. Misalnya, *Mie Kalem* memiliki kemiripan 0.91 terhadap *Mie Gemes Level 1*, yang menunjukkan bahwa kedua menu tersebut memiliki deskripsi yang sangat mirip, baik dari sisi bahan dasar mie, gaya penyajian, maupun elemen rasa. Sebaliknya, nilai kemiripan antara *Mie Kalem* dan *Dimsum Ekado* lebih rendah, yaitu sekitar 0.72, mengindikasikan bahwa meskipun terdapat beberapa elemen kesamaan dalam deskripsi (misalnya rasa gurih atau proses pengolahan tertentu), keduanya tetap mewakili kategori makanan yang berbeda sehingga tidak ditempatkan terlalu dekat dalam ruang vektor semantik. Matriks kemiripan ini berfungsi sebagai inti dari mekanisme rekomendasi, karena daftar rekomendasi untuk setiap menu disusun berdasarkan nilai *cosine similarity* tertinggi yang ditunjukkan pada tabel tersebut.

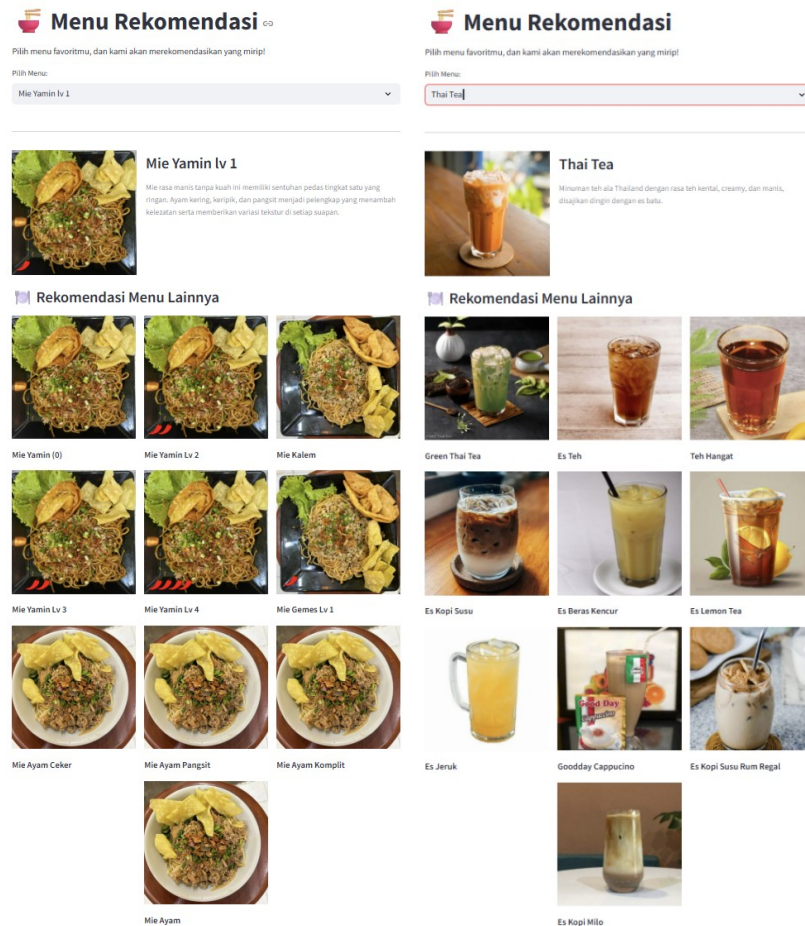
Melalui pola hubungan kemiripan yang ditunjukkan pada Tabel 4, sistem kemudian memanfaatkan nilai *cosine similarity* sebagai dasar untuk menentukan menu-menu yang paling relevan bagi setiap item. Seluruh nilai kemiripan yang telah dihitung disortir secara menurun sehingga menu dengan kedekatan semantik tertinggi ditempatkan pada urutan teratas. Dari proses tersebut, sistem mengambil sepuluh nilai tertinggi untuk setiap menu sebagai rekomendasi akhir.

c. Hasil Prototipe Streamlit

Hasil sistem rekomendasi yang ditampilkan di Streamlit pada Gambar 4 menunjukkan bagaimana model mSBERT mampu menangkap kedekatan semantik antar-menu berdasarkan deskripsi, bahan, dan konteks penyajian. Pada contoh pertama, ketika pengguna memilih *Mie Yamin Lv 1*, sistem menampilkan rekomendasi yang seluruhnya berasal dari kelompok menu mie dengan karakteristik serupa. Rekomendasi seperti *Mie Yamin Lv 0*, *Mie Kalem*, *Mie Yamin Lv 2–4*, hingga *Mie Ayam Ceker* dan *Mie Ayam Pangsit* menunjukkan bahwa model berhasil memahami bahwa deskripsi menu tersebut memiliki kesamaan konsep kuliner, yaitu hidangan mie ayam dengan variasi level pedas, *topping*, rasa, atau teknik penyajian yang tidak jauh berbeda. Pola rekomendasi ini menegaskan bahwa *embedding* mSBERT bekerja dengan baik dalam memetakan kedekatan makna antar-deskripsi, bukan sekadar mencocokkan kata kunci seperti “mie” atau “ayam”. Misalnya, kemunculan *Mie Kalem* atau *Mie Gemes Lv 1* tidak sekadar karena memiliki nama yang mirip, tetapi karena model menilai bahwa struktur deskripsi, profil penyajian, dan jenis hidangan berada dalam ruang semantik yang sangat dekat dengan *Mie Yamin Lv 1*. Ini menjadi indikator bahwa sistem mampu memberikan rekomendasi yang relevan dan konsisten terhadap preferensi rasa serta kategori hidangan yang dipilih pengguna.

Pada contoh kedua, yaitu *Thai Tea*, sistem memberikan rekomendasi yang mengarah pada kelompok minuman dengan profil rasa yang serupa, terutama dalam ranah minuman manis, *creamy*, atau berbasis teh. Rekomendasi seperti *Green Thai Tea*, *Es Teh*, *Teh Hangat*, *Es Lemon Tea*, *Es Teh*, *Es Beras Kencur*, maupun minuman lain seperti *Es Kopi Susu* dan *Godday Cappucino* menunjukkan bahwa model tidak hanya membatasi diri pada kategori minuman teh, tetapi juga memperluas kemungkinan ke minuman dengan pengalaman konsumsi yang mirip yaitu, manis, segar, dan mudah ditemukan di konteks kuliner yang sama. Pola ini memperlihatkan bahwa mSBERT menangkap hubungan semantik yang lebih abstrak, seperti kesamaan fungsi menu (minuman pendamping makanan), gaya minuman modern, serta kemiripan deskripsi bahan seperti susu, teh, atau gula. Namun, rekomendasi seperti *Es Kopi Susu* atau *Godday Cappucino* juga menunjukkan karakteristik model yang lebih fleksibel, di mana ia tidak membatasi rekomendasi hanya pada kategori ketat (kategori teh), melainkan menghubungkannya pada minuman dengan sentimen deskriptif atau preferensi konsumsi yang

beririsan. Hal ini justru menguatkan kemampuan *embedding* kontekstual dalam memetakan kesamaan makna, bukan sekadar kesamaan kata.



Gambar 4. Hasil Sistem Rekomendasi berbasis Streamlit

Jika dilihat secara keseluruhan, pola rekomendasi pada kedua contoh tersebut menggambarkan bahwa pendekatan *embedding* berbasis mSBERT efektif dalam meniru logika “kedekatan konsep” seperti yang dilakukan manusia. Ketika pengguna memilih menu makanan, model cenderung mengembalikan rekomendasi makanan serupa; ketika memilih minuman, sistem mengembalikan minuman lain yang relevan. Tingkat konsistensi ini memperlihatkan bahwa representasi vektor yang dihasilkan mSBERT cukup stabil dalam mengelompokkan menu secara semantik. Namun, beberapa rekomendasi yang tampak agak melebar, seperti variasi kopi pada input *Thai Tea*, memberikan ruang untuk diskusi lebih dalam. Hal tersebut menunjukkan bahwa model memahami kemiripan konteks konsumsi, tetapi belum sepenuhnya membedakan preferensi kategori secara ketat. Meski demikian, untuk kasus UMKM kuliner yang menu-menunya sangat beragam tetapi deskripsinya sederhana, kemampuan model untuk menangkap kedekatan makna secara fleksibel justru memberi keuntungan dalam memperluas opsi rekomendasi tanpa keluar terlalu jauh dari preferensi pengguna.

III.2 Hasil Evaluasi Sistem Rekomendasi

Hasil penilaian dari 20 responden yang disajikan pada Tabel 4 menunjukkan bahwa sistem memperoleh skor rata-rata keseluruhan sebesar 4,31, yang mengindikasikan bahwa rekomendasi yang diberikan dianggap relevan, menarik, dan dapat diterima oleh mayoritas pengguna. Dengan rentang skor 1 hingga 5, nilai ini menempatkan performa sistem pada kategori baik hingga sangat

baik, sehingga secara umum dapat disimpulkan bahwa pendekatan mSBERT mampu menangkap kesamaan semantik antar-menu dengan tingkat akurasi yang cukup tinggi dari perspektif pengguna.

Tabel 4. Hasil *User Preference Study*

No	Responden	Mie Yamin Level 1	Cwie Mie	Mie Ayam	Thai Tea	Dimsum Siomay Mozarella
1	Affa Lelira I	4	5	3	5	5
2	Shafira Amanda	4	5	5	5	5
3	Marsha Ardelia	5	5	4	5	5
4	Anggi	4	5	5	5	3
5	Hertina Esti	5	5	5	5	5
6	Wahyu Melinda	4	5	4	5	5
7	Mirechelin Kristanaya	3	4	4	4	5
8	Melinda	5	4	5	4	4
9	Alfaro Alamsyah Muchlisun	5	5	4	3	3
10	Tsabita	4	1	3	4	1
11	Daffin_TW	5	4	5	4	4
12	Yunita Ajeng	5	3	4	5	3
13	Serlinda Mareta Putri	5	5	5	5	5
14	Selayanti	5	5	5	5	5
15	Ajeng	5	4	5	4	5
16	Naura Ulayya Nariswari	5	4	5	5	4
17	Hanifah	4	3	3	4	4
18	Yuliana	4	2	4	5	5
19	Deannisa Syafira Putri	4	4	5	4	5
20	Tasyafa	3	4	5	4	2
Rata-Rata tiap Input Menu		4,4	4,1	4,4	4,5	4,15
Rata-Rata Keseluruhan		4,31				

Apabila ditinjau lebih dalam pada setiap menu input, terlihat bahwa performa sistem bervariasi antar kategori. Menu Thai Tea memperoleh skor rata-rata tertinggi sebesar 4,5, menunjukkan bahwa rekomendasi untuk kategori minuman dinilai paling sesuai dengan ekspektasi pengguna. Tingginya skor ini dapat disebabkan oleh deskripsi minuman yang cenderung lebih homogen sehingga *embedding* yang dihasilkan model lebih stabil dan mudah memetakan kesamaan antar-item. Sementara itu, *Mie Yamin Level 1* dan *Mie Ayam* memperoleh skor yang sama, yaitu 4,4, yang menunjukkan bahwa sistem cukup konsisten dalam memahami kemiripan antar-menu mie. Rekomendasi untuk menu-menu tersebut dinilai menarik dan rasional oleh sebagian besar responden. Di sisi lain, *Dimsum Siomay Mozarella* memiliki skor rata-rata 4,15 dan *Cwie Mie* menjadi menu dengan skor terendah sebesar 4,1 meskipun tetap berada pada kategori baik. Variasi ini mengindikasikan bahwa kategori menu dengan karakteristik lebih unik atau deskripsi yang lebih bervariasi cenderung menghasilkan rekomendasi yang tidak selalu dianggap sepadan oleh semua pengguna.

Perbedaan penilaian antarresponden juga menunjukkan dinamika preferensi yang cukup menarik. Meskipun sebagian besar responden memberikan skor tinggi, terdapat beberapa individu

yang memberikan nilai rendah pada menu tertentu. Hal ini menegaskan bahwa persepsi terhadap relevansi rekomendasi tidak hanya dipengaruhi oleh kesamaan semantik yang ditangkap model, tetapi juga oleh preferensi personal yang tidak bisa direpresentasikan melalui pendekatan *Content-Based Filtering* murni. Ketidakhadiran data preferensi historis pengguna memang tidak menjadi hambatan bagi CBF, tetapi membuat sistem tidak mampu menghasilkan rekomendasi yang benar-benar dipersonalisasi untuk setiap individu. Sehingga, variasi penilaian antarresponden menunjukkan bahwa meskipun sistem telah bekerja dengan baik, terdapat potensi pengembangan lebih lanjut ke arah pendekatan *hybrid* ketika data pengguna tersedia.

Secara keseluruhan, hasil evaluasi ini menunjukkan bahwa sistem rekomendasi berbasis mSBERT yang dikembangkan memiliki performa yang baik dalam konteks CBF dan mampu menghasilkan rekomendasi yang dihargai positif oleh pengguna. Skor rata-rata di atas 4,0 pada seluruh menu input memperlihatkan bahwa proses *embedding* berhasil menangkap kemiripan konten antar-menu secara konsisten. Namun, adanya perbedaan skor pada menu tertentu juga memberi gambaran bahwa peningkatan dapat dilakukan, misalnya dengan memperkaya atribut konten, memperhalus deskripsi menu, atau memanfaatkan model *embedding* yang lebih besar untuk menangkap konteks yang lebih kompleks. Meski demikian, capaian skor rata-rata 4,31 menunjukkan bahwa sistem telah memenuhi tujuan penelitian, yaitu menghasilkan rekomendasi yang relevan menurut persepsi pengguna, sehingga dapat dianggap berhasil pada tahap pengujian subjektif ini.

IV. KESIMPULAN

Berdasarkan hasil sistem rekomendasi dan evaluasi preferensi pengguna, dapat disimpulkan bahwa pendekatan *Content-Based Filtering* berbasis *embedding* mSBERT mampu menghasilkan rekomendasi menu yang relevan, konsisten, dan selaras dengan persepsi mayoritas pengguna. Hasil visualisasi pada prototipe Streamlit menunjukkan bahwa model dapat memahami kedekatan semantik antar-menu, baik pada kategori makanan maupun minuman, serta memberikan rekomendasi yang logis berdasarkan deskripsi konten. Evaluasi dari 20 responden dengan skor rata-rata 4,31 mengonfirmasi bahwa rekomendasi yang dihasilkan dinilai menarik dan dapat diterima, meskipun masih terdapat variasi penilaian pada beberapa kategori menu yang disebabkan oleh keberagaman preferensi personal dan kompleksitas deskripsi tertentu. Secara keseluruhan, performa sistem memenuhi tujuan penelitian, menegaskan bahwa mSBERT merupakan pilihan yang tepat untuk UMKM kuliner yang memiliki keterbatasan data pengguna, dan menunjukkan potensi untuk dikembangkan lebih lanjut menuju sistem *hybrid* apabila data preferensi individual tersedia.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak yang telah mendukung proses penelitian ini, khususnya dosen pembimbing atas arahan yang diberikan selama penyusunan paper. Dukungan tersebut berperan penting dalam kelancaran penyelesaian penelitian dan penulisan naskah ini.

REFERENSI

- [1] Kementerian Pariwisata dan Ekonomi Kreatif Republik Indonesia, “INFOGRAFIS DATA STATISTIK INDIKATOR MAKRO PARIWISATA,” Kementerian Pariwisata dan Ekonomi Kreatif Republik Indonesia.
- [2] A. Rachman Putra, T. Seno Anjanarko, and E. Retnowati, “Creative Economy Development Efforts in Culinary Business,” 2022.
- [3] M. dan Minuman, “Statistik Penyediaan BADAN PUSAT STATISTIK BPS-STATISTICS INDONESIA,” 2023.
- [4] G. Duan, Y. Du, and Y. Shang, “Research on Personalized Recommendation of Complementary Products Based on Demand Cross-Elasticity and Hypergraphs,”

- Electronics (Switzerland)*, vol. 13, no. 23, Dec. 2024, doi: 10.3390/electronics13234851.
- [5] S. Chhipa, V. Berwal, T. Hirapure, and S. Banerjee, "Recipe Recommendation System Using TF-IDF," *ITM Web of Conferences*, vol. 44, p. 02006, 2022, doi: 10.1051/itmconf/20224402006.
- [6] A. Pramono and T. Satrio Setyo Ardi Wolayan, "IMPLEMENTATION OF CONTENT BASED FILTERING METHOD IN RESTAURANT MENU ORDERING RECOMMENDATION SYSTEM", [Online]. Available: <https://return.publikasikupublisher.com>
- [7] R. Prashanth, "Semantic similarity estimation for domain specific data using BERT and other techniques," Jun. 2025, [Online]. Available: <http://arxiv.org/abs/2506.18602>
- [8] H. Yulianton and R. Candra Noor Santi, "Product Matching using Sentence-BERT: A Deep Learning Approach to E-Commerce Product Deduplication," *Engineering and Technology Journal*, vol. 09, no. 12, Dec. 2024, doi: 10.47191/etj/v9i12.14.
- [9] R. Huang, "Improved content recommendation algorithm integrating semantic information," *J Big Data*, vol. 10, no. 1, Dec. 2023, doi: 10.1186/s40537-023-00776-7.
- [10] S. M. H. Danish, S. M. E. Hasnain, H. Ashraf, and R. Rukaiya, "Comparative Analysis of BERT and TF-IDF for Textual Semantic Similarity Assessment," in *Proceedings of the IEEE International Multi Topic Conference, INMIC*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/INMIC64792.2024.11004377.
- [11] A. Rogers, O. Kovaleva, and A. Rumshisky, "A Primer in BERTology: What we know about how BERT works," Nov. 2020, [Online]. Available: <http://arxiv.org/abs/2002.12327>
- [12] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," Aug. 2019, [Online]. Available: <http://arxiv.org/abs/1908.10084>
- [13] T. T. Ha, V. N. Nguyen, K. H. Nguyen, K. A. Nguyen, and Q. K. Than, "Utilizing SBERT For Finding Similar Questions in Community Question Answering," in *Proceedings - International Conference on Knowledge and Systems Engineering, KSE*, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/KSE53942.2021.9648830.
- [14] N. Reimers and I. Gurevych, "Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation," Oct. 2020, [Online]. Available: <http://arxiv.org/abs/2004.09813>
- [15] A. Muhaimin, I. Taufik, and D. Daniswara, "Pendeteksian Spam pada E-mail menggunakan Pendekatan Natural Language Processing," *PROSIDING SEMINAR NASIONAL SAINS DATA*, vol. 3, no. 1, pp. 116–121, Nov. 2023.
- [16] M. D. Ekstrand, F. M. Harper, M. C. Willemsen, and J. A. Konstan, "User perception of differences in recommender algorithms," in *RecSys 2014 - Proceedings of the 8th ACM Conference on Recommender Systems*, Association for Computing Machinery, Oct. 2014, pp. 161–168. doi: 10.1145/2645710.2645737.
- [17] M. Hertzum, "Understanding preference: A meta-analysis of user studies," *International Journal of Human Computer Studies*, vol. 195, Jan. 2025, doi: 10.1016/j.ijhcs.2024.103408.
- [18] Turn your data scripts into shareable web apps in minutes, "Streamlit."