Jurnal Informatika: Jurnal pengembangan IT

Vol. 11, No. 2, 2026
ISSN : 2477-5126, DOI:10.30591/jpit.v11i2.10204

Analisis Sentimen dan Pemodelan Topik Ulasan Pelanggan Restoran Cepat Saji Berbasis BiLSTM dan BERTopic

Deni Gunawan [1], Andi Prasetyo Wicaksono [2], Gandung Triyono [3]
[1,2,3] Fakultas Teknologi Informasi, Magister Ilmu Komputer, Budi Luhur University, Indonesia
2211601576@student.budiluhur.ac.id, 2311601815@student.budiluhur.ac.id, gandung.triyono@budiluhur.ac.id

Info Artikel

Riwayat Artikel:

Received 2026-01-17
Revised 2026-06-28
Accepted 2026-07-15

Abstract *– Fast-food restaurants in Indonesia continue to grow, including Restaurant X, which competes to attract and retain customers. Reviews on digital platforms such as Google Maps have become a crucial provider of information for understanding customer satisfaction, complaints, and service aspects that require enhancement. The present study seeks to examine customer sentiment and identify key topics in customer reviews using a combination of Bidirectional Long Short-Term Memory (BiLSTM) and BERTopic. The originality of this research is reflected in the integratio of supervised and unsupervised learning approaches to provide a more holistic understanding of customer perceptions in the Indonesian fast-food restaurant industry. A total of 10,451 reviews from nine outlets with having the largest number of customer reviews during the 2016–2025 period were gathered and examined. The findings suggest that most reviews exhibit positive sentiment, with 7,762 reviews classified as positive. The BiLSTM model attained an accuracy rate of 78.6%, a precision value of 81.6%, a recall value of 83.5%, and an F1-score value of 82.5%, demonstrating strong classification performance. Meanwhile, BERTopic successfully identified dominant topics related to service quality, service speed, food spiciness, and menu variety. The research contribution lies in the development of a unified analytical framework capable of simultaneously identifying customer sentiment and uncovering the key issues influencing customer experiences. These findings can function as a data-based decision support system for enhancing service quality and customer satisfaction.*

Keywords: *BERTopic; BiLSTM; fast-food restaurant; sentiment analysis.*

Corresponding Author:

Andi Prasetyo Wicaksono

Email: 22311601815@student.budiluhur.ac.id

Abstrak – *Restoran cepat saji di Indonesia terus berkembang, termasuk Restoran X yang bersaing dalam menarik dan mempertahankan pelanggan. Ulasan pada platform digital seperti Google Maps menjadi sumber informasi penting untuk memahami tingkat kepuasan pelanggan, keluhan, serta aspek layanan yang perlu ditingkatkan. Penelitian ini bertujuan menganalisis sentimen dan mengidentifikasi topik utama dalam ulasan pelanggan menggunakan kombinasi metode BiLSTM dan BERTopic. Kebaruan penelitian ini terletak pada integrasi pendekatan supervised learning dan unsupervised learning untuk memperoleh pemahaman yang lebih komprehensif terhadap persepsi pelanggan restoran cepat saji di Indonesia. Sebanyak 10.451 ulasan dari sembilan gerai dengan jumlah ulasan terbanyak pada periode 2016–2025 dikumpulkan dan dianalisis. Hasil penelitian memperlihatkan jika mayoritas ulasan mempunyai sentimen positif, dengan 7.762 ulasan berada pada kategori positif. Model BiLSTM menghasilkan akurasi senilai 78,6%, precision 81,6%, recall 83,5%, dan F1-score 82,5%, yang menunjukkan kinerja klasifikasi yang baik. Sementara itu, BERTopic berhasil mengidentifikasi topik-topik dominan yang berkaitan dengan pelayanan, kecepatan layanan, tingkat kepedasan makanan, dan variasi menu. Kontribusi penelitian ini adalah pengembangan kerangka analisis yang mampu mengidentifikasi sentimen pelanggan sekaligus mengungkap isu-isu utama yang memengaruhi pengalaman pelanggan. Temuan penelitian bisa diterapkan sebagai landasan penentuan kebijakan berlandaskan informasi untuk meningkatkan kualitas layanan dan kepuasan pelanggan.*

Kata Kunci*: analisis sentimen; BERTopic; BiLSTM; restoran cepat saji.*

I. PENDAHULUAN

Industri makanan cepat saji di Indonesia terus mengalami pertumbuhan seiring perubahan gaya hidup masyarakat urban dan perkembangan teknologi digital [1]. Persaingan yang semakin ketat mendorong perusahaan untuk memahami kebutuhan dan persepsi pelanggan secara lebih mendalam guna meningkatkan kualitas layanan dan mempertahankan loyalitas pelanggan. Salah satu sumber informasi yang dapat dimanfaatkan adalah ulasan pelanggan yang tersedia pada platform digital seperti *Google Maps*. Ulasan tersebut memuat berbagai opini mengenai kualitas makanan, harga, pelayanan, kebersihan, dan kenyamanan yang secara langsung mencerminkan pengalaman pelanggan

[2], [3]. Namun, sebagian besar perusahaan masih belum memanfaatkan data ulasan tersebut secara baik sebagai sumber informasi strategis untuk mendukung pengambilan keputusan bisnis berbasis data [4].

Pemanfaatan teknik *Natural Language Processing* (NLP) dalam analisis ulasan pelanggan semakin berkembang, salah satunya melalui analisis sentimen yang berfungsi untuk mengidentifikasi kecenderungan opini ke dalam kategori positif, negatif, maupun netral [5]. Di sisi lain, pemodelan topik (*topic modeling*) berperan dalam mengungkap tema-tema dominan yang menjadi fokus pembahasan pelanggan pada suatu produk atau layanan [6]. Integrasi analisis sentimen dan pemodelan topik mampu menghasilkan pemahaman yang lebih menyeluruh karena tidak hanya menunjukkan arah sentimen yang muncul, tetapi juga menjelaskan isu-isu yang mendasari terbentuknya sentimen tersebut [7].

Penelitian analisis sentimen telah banyak dilakukan menggunakan berbagai algoritma *machine learning* maupun *deep learning*. Penelitian oleh Prasetyo dan Wahju memperlihatkan jika *Logistic Regression* menghasilkan performa klasifikasi sentimen terbaik pada analisis ulasan pengguna GoPay berbasis TF-IDF dengan akurasi senilai 88,16%, lebih tinggi dibandingkan *Support Vector Machine* (SVM) yang mendapatkan akurasi 87,5% juga *Random Forest* dengan akurasi 79,33% [8]. Hasil tersebut memperlihatkan jika pemilihan algoritma mempunyai pengaruh yang signifikan terhadap performa klasifikasi sentimen. Perkembangan penelitian di bidang analisis teks memperlihatkan jika metode *machine learning* tradisional belum sepenuhnya mampu merepresentasikan makna yang terkandung dalam hubungan antar kata dan konteks kalimat yang kompleks. Sebagai alternatif, pendekatan *deep learning* semakin banyak diterapkan karena mempunyai kompetensi yang lebih baik guna mempelajari representasi bahasa secara kontekstual.

Penelitian terdahulu terkait *Bidirectional Long Short-Term Memory* (BiLSTM) memperlihatkan jika metode ini mampu menghasilkan performa klasifikasi yang tinggi pada berbagai domain. Penelitian oleh Angga dan Rito pada analisis sentimen pengguna aplikasi MyPertamina memperlihatkan jika BiLSTM menghasilkan kinerja yang lebih baik dibandingkan *Long Short-Term Memory* (LSTM), dengan ketepatan senilai 90%, LSTM senilai 86,25% [9]. Selanjutnya, penelitian oleh Adi dan Galuh pada klasifikasi multi-kelas data aduan masyarakat memperoleh akurasi tertinggi senilai 70% dan memperlihatkan jika BiLSTM efektif dalam menangani variasi bahasa serta tugas klasifikasi teks multi-kelas [10]. Selain itu, penelitian oleh Aini dan Fadli pada pengkajian opini perolehan pemilihan umum pada platform sosial X memperlihatkan jika model BiLSTM memperoleh akurasi senilai 86,89% dengan kinerja evaluasi yang tinggi pada beberapa metrik, sehingga dinilai efektif dalam menangani data teks yang kompleks [11]. Secara keseluruhan, berbagai studi tersebut memperlihatkan jika BiLSTM mempunyai kecakapan yang baik dalam mempelajari pola sekuensial serta mempertahankan informasi kontekstual dibandingkan pendekatan konvensional.

BERTopic ialah teknik pemetaan tema mutakhir sehingga memadukan gambaran makna berlandaskan transformer serta berbasis kelas TF-IDF guna membentuk tema kian konsisten [8]. Beragam studi memperlihatkan jika teknik tersebut sanggup menghasilkan mutu tema kian unggul ketimbang metode konvensional, misalnya Latent Dirichlet Allocation (LDA, terutama pada kumpulan teks yang besar dan kompleks [11]. Selain itu, BERTopic telah diterapkan pada berbagai bidang, termasuk analisis pasar saham dan pemetaan perkembangan penelitian, karena mampu mengidentifikasi tema-tema utama dengan taraf akuirsi yang lebih tinggi [12].

Berdasarkan sejumlah studi sebelumnya, performa algoritma klasifikasi sentimen masih bervariasi dengan akurasi berkisar antara 70% hingga 90% tergantung pada metode dan karakteristik dataset [13]. kondisi ini memperlihatkan jika meskipun BiLSTM mempunyai kinerja yang cukup tinggi pada beberapa studi, konsistensi hasil masih belum tercapai terutama pada dataset dengan kompleksitas tinggi dan skenario multi-kelas [14]. Selain itu, sebagian besar penelitian masih menerapkan BiLSTM untuk klasifikasi sentimen dan BERTopic untuk pemodelan topik secara terpisah.

Penelitian yang berfokus pada analisis sentimen umumnya hanya menghasilkan polaritas opini tanpa menjelaskan isu yang melatarbelakanginya, sedangkan pemodelan topik hanya mengidentifikasi tema tanpa mengaitkannya dengan sentimen pengguna. Kondisi ini menunjukkan adanya *research gap* yaitu belum terintegrasinya analisis sentimen dan pemodelan topik dalam satu kerangka analisis yang komprehensif pada konteks ulasan pelanggan restoran cepat saji di Indonesia.

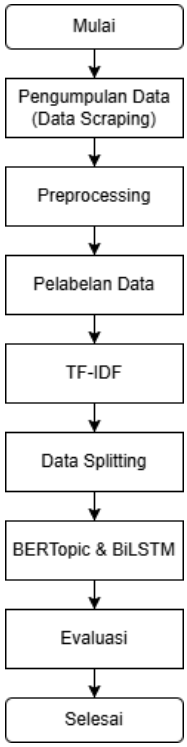
Berdasarkan gap tersebut, penelitian ini bertujuan untuk menganalisis sentimen pelanggan sekaligus mengidentifikasi topik utama pada ulasan Restoran X menggunakan metode BiLSTM dan BERTopic. Sebanyak 10.451 ulasan dari sembilan gerai pada periode 2016–2025 digunakan sebagai data penelitian. Temuan penelitian ini diupayakan bisa diterapkan guna mengidentifikasi persepsi pelanggan secara lebih menyeluruh sehingga dapat menjadi landasan dalam peningkatan layanan dan pengambilan keputusan berbasis data.

Nilai inovatif studi ini ada pada penggabungan BiLSTM serta BERTopic pada sebuah kerangka pengkajian untuk secara simultan mengidentifikasi sentimen dan topik utama. Kontribusi penelitian ini meliputi pengembangan kerangka analisis terpadu, evaluasi performa BiLSTM dalam klasifikasi sentimen, serta identifikasi topik dominan sebagai dasar pengambilan keputusan berbasis data.

II. METODE

A. Tahapan Penelitian

Metodologi penelitian disusun untuk secara sistematis menggambarkan tahapan-tahapan yang dilakukan demi mencapai tujuan serta hasil yang diharapkan[15]. Setiap langkah dirancang secara teratur dan berurutan agar proses penelitian berjalan sesuai rencana. Gambar 1 menyajikan visualisasi dari langkah-langkah yang akan di laksanakan pada studi ini.



Gambar 1. Tahapan Penelitian

Berlandaskan Gambar 1, studi ini diawali dengan penghimpunan data lewat scraping review pelanggan pada platform *Google Maps* dari sembilan gerai Restoran X. Data yang didapat kemudian lewat tahap *preprocessing* yang mencakup *case folding, cleansing, normalization, tokenizing, stopword removal*, dan *stemming* untuk menghasilkan data yang siap dianalisis. Selanjutnya, ulasan diberi label berdasarkan rating pengguna dan dikelompokkan ke dalam lima kategori sentimen. Data dokumen yang sudah dipersiapkan berikutnya diidentifikasi memanfaatkan TF-IDF untuk menghasilkan representasi numerik, lalu dibagi menjadi data pelatihan juga pengujian. Tahap analisis dilaksanakan guna mengaplikasikan BERTopic untuk mengidentifikasi topik-topik utama dalam ulasan pelanggan serta BiLSTM untuk mengklasifikasikan sentimen. Kinerja model dievaluasi menggunakan *confusion matrix*, sedangkan hasil pemodelan topik diinterpretasikan untuk mengungkap isu-isu utama yang memengaruhi pengalaman pelanggan. Seluruh tahapan tersebut bertujuan menghasilkan pemahaman yang lebih komprehensif mengenai persepsi pelanggan terhadap layanan Restoran X.

B. Pengumpulan Data

Data dikumpulkan dari ulasan publik yang tersedia pada *Google Maps* selama periode 2016–2025. Sebanyak 10.451 ulasan berhasil diperoleh dari sembilan gerai Restoran X yang mempunyai jumlah ulasan terbanyak. Untuk menjaga kualitas data, dilakukan proses validasi dengan menghapus data duplikat, ulasan kosong, serta karakter yang tidak relevan. Data yang digunakan merupakan data publik sehingga tidak mengandung informasi pribadi yang dapat mengidentifikasi pengguna secara langsung. Jumlah data yang besar serta rentang waktu pengumpulan yang panjang diharapkan mampu memberikan representasi yang memadai terhadap persepsi pelanggan.

C. Prapemrosesan Data

Prapemrosesan data diawali dengan tahap *data cleaning* untuk menghilangkan karakter atau elemen teks yang tidak relevan dan tidak memberikan informasi yang diperlukan dalam analisis [16]. Setelah itu, dilakukan *case folding* dengan mengonversi seluruh huruf menjadi huruf kecil guna menciptakan format teks yang seragam. Tahap selanjutnya ialah *stopword removal*, yakni menghilangkan kosa-kata yang sering muncul tetapi mempunyai kontribusi yang rendah terhadap makna teks. Selanjutnya, teks dipecah menjadi unit-unit kata melalui proses tokenisasi. Tahap

akhir adalah *stemming*, yaitu mengembalikan setiap token ke bentuk wujud istilah pokoknya agar arti pokok pada istilah bisa diidentifikasi secara lebih konsisten oleh model.

D. Pelabelan Data

Penentuan label sentimen pada penelitian ini didasarkan pada skor rating yang diberikan pengguna melalui *Google Maps*, karena nilai tersebut mencerminkan tingkat kepuasan pelanggan terhadap layanan yang diterima. Sentimen kemudian dikelompokkan ke lima kategori, yakni sangat negatif, negatif, netral, positif, dan sangat positif. Rating 1 merepresentasikan sentimen sangat negatif, rating 2 negatif, rating 3 netral, rating 4 positif, dan rating 5 sangat positif sebagaimana ditunjukkan pada Tabel 1. Penggunaan rating sebagai dasar pelabelan dipilih karena telah banyak diterapkan dalam penelitian analisis sentimen serta memungkinkan proses anotasi dilakukan secara lebih konsisten dan efisien pada dataset berskala besar.

TABEL 1
KATEGORI LABEL SENTIMEN

Rating	Ulasan
1	Sangat Negatif
2	Negatif
3	Netral
4	Positif
5	Sangat Positif

E. TF-IDF

TF-IDF (*Term Frequency–Inverse Document Frequency*) merupakan pendekatan ekstraksi fitur yang banyak digunakan dalam *Natural Language Processing* (NLP) untuk mengubah data teks menjadi representasi numerik berbentuk vektor berbobot sehingga bisa diolah memakai metode pembelajaran mesin [17]. Teknik tersebut memadukan sepasang unsur pokok, yakni Term Frequency (TF) yang memperlihatkan tingkat munculnya sebuah istilah pada berkas serta Inverse Document Frequency (IDF) yang menilai derajat keterbatasan istilah itu di seluruh kumpulan berkas [18]. Melalui kombinasi kedua komponen tersebut, kata yang sering muncul pada suatu berkas namun relatif jarang ditemukan pada berkas lain akan memperoleh nilai yang lebih besar sebab dianggap mempunyai informasi yang lebih penting. Perhitungan TF-IDF pada penelitian ini mengacu pada Persamaan 1.

$$TF - IDF(w, d, D) = TF(w, d) \times log\ log\ \left(\frac{N}{DF(w)}\right) \quad (1)$$

Dalam rumus TF-IDF, w merujuk pada kata atau token yang dianalisis dalam teks, sementara d adalah dokumen tempat kata tersebut muncul. D merupakan kumpulan dokumen atau korpus yang menjadi basis perhitungan. Nilai "TF" (w,d) mengatakan frekuensi muncul kata w dalam dokumen d, dan "DF" (w) menunjukkan total berkas dalam korpus D yang mengandung kata w. Adapun N adalah total berkas dalam keseluruhan korpus[19].

Kian tinggi nilai TF kata dalam dokumen, dan kian rendah nilai DF kata dalam korpus, maka bobot TF-IDF akan semakin tinggi. Bobot ini digunakan untuk membentuk vektor representasi dokumen, yang kemudian dapat diproses lebih lanjut dalam klasifikasi, pengelompokan, maupun analisis kemiripan antar teks.

F. Pembagian Data

Pada tahapan ini, dataset yang sudah melalui proses prapemrosesan dibagi menjadi data training senilai 80% dan data testing senilai 20% menggunakan Scikit-learn [20]. Pembagian ini bertujuan agar model BiLSTM dapat dilatih untuk mengenali pola sentimen dari data training, kemudian diuji pada data testing untuk mengevaluasi kemampuannya dalam merepresentasikan kondisi data sebenarnya. Sementara itu, BERTopic diterapkan pada keseluruhan data ulasan karena bersifat unsupervised sehingga tidak bergantung pada pembagian dataset. Data training digunakan untuk membangun model BiLSTM dalam mengklasifikasikan sentimen seperti positif, negatif, dan netral, sedangkan data testing dimanfaatkan untuk mengukur performa prediksi model serta menguji generalisasi terhadap data yang belum pernah dilihat sebelumnya. Maka, proses ini memastikan model mempunyai kinerja yang stabil baik pada data pelatihan maupun data pengujian.

G. Pemodelan

Dalam penelitian ini, pemodelan topik dilakukan menggunakan BERTopic yang termasuk dalam metode *unsupervised learning*[21]. Pendekatan ini membentuk representasi semantik dokumen dengan memanfaatkan model berbasis transformer seperti BERT, kemudian mengelompokkan dokumen berdasarkan kesamaan makna melalui algoritma clustering berbasis density, yaitu HDBSCAN. Setelah proses pengelompokan selesai, BERTopic menggunakan pendekatan *class-based Term Frequency–Inverse Document Frequency* (c-TF-IDF) untuk

mengidentifikasi kata-kata yang paling representatif pada setiap kelompok topik yang terbentuk. Perhitungan c-TF-IDF untuk topik t dan kata w dapat dilihat pada Persamaan 2.

$$c - TF - IDF_{t,w} = \frac{f_{t,w}}{\sum_{w'} \quad f_{t,w'}} \cdot log\ log\ \left(\frac{N}{n_w}\right) \qquad (2)$$

Di mana t adalah indeks topik tertentu, w adalah kata (token) dalam dokumen, $f_{t,w}$ adalah frekuensi kemunculan kata *w* dalam seluruh dokumen pada topik t, $\sum_{w'} \quad f_{t,w'}$ adalah total frekuensi semua kata dalam topik t, N adalah jumlah total topik, dan n_w adalah jumlah topik yang mengandung kata w. Dengan demikian, semakin tinggi nilai c-TF-IDF suatu kata, semakin besar pula kontribusi dan kekhasan kata tersebut dalam merepresentasikan topik tertentu.

Pada penelitian ini, parameter BERTopic diatur menggunakan model embedding berbasis Sentence-BERT "all-MiniLM-L6-v2" untuk menghasilkan representasi semantik dokumen. Proses reduksi dimensi menggunakan UMAP dengan parameter n_neighbors=15, n_components=5, dan metric cosine untuk mempertahankan struktur lokal data. Selanjutnya, algoritma clustering HDBSCAN digunakan dengan parameter minimum *cluster size* senilai 10 untuk membentuk kelompok topik yang stabil. Selain itu, parameter top_n_words ditetapkan senilai 10 untuk menampilkan kata kunci utama pada setiap topik, serta calculate_probabilities diaktifkan untuk menghitung probabilitas keanggotaan dokumen terhadap masing-masing topik.

TABEL 2
HYPERPARAMETER MODEL BERTOPIC

Parameter	Nilai
Embedding	all-MiniLM-L6-v2
UMAP n_neighbors	15
UMAP n_components	5
UMAP metric	cosine
HDBSCAN min_cluster_size	10
Top N Words	10
Calculate Probabilities	True
Language	multilingual

Hasil dari pemodelan ini berupa sekumpulan topik yang masing-masing disertai kata-kata kunci yang paling dominan, mencerminkan tema utama dari ulasan pelanggan seperti *pelayanan cepat*, *harga terjangkau*, *rasa pedas*, atau *tempat bersih*.

H. Klasifikasi Model

Klasifikasi model dilakukan dengan pendekatan *supervised learning* menggunakan algoritma *Bidirectional Long Short-Term Memory* (BiLSTM). BiLSTM merupakan pengembangan dari arsitektur LSTM, yang memungkinkan pemrosesan sekuens data dari dua arah sekaligus (maju dan mundur), sehingga lebih efektif dalam memahami konteks kata dalam kalimat, model terdiri dari tiga lapisan utama[22].

1) *Embedding Layer*: Mengubah kata menjadi vektor berdimensi tetap.
2) *BiLSTM Layer*: mempelajari pola sekuens dari dua arah.
3) *Dense Output Layer*: menghasilkan skor probabilitas untuk masing-masing kelas sentimen (positif, negatif, netral).

Pada penelitian ini, model BiLSTM dibangun menggunakan tiga lapisan *Bidirectional* LSTM dengan masing-masing 128, 64, dan 32 unit neuron. Representasi kata menggunakan pre-trained GloVe embedding berdimensi 100 yang bersifat non-trainable. Untuk mengurangi risiko overfitting, digunakan lapisan Dropout senilai 0,6 setelah dense layer. Proses pelatihan model dilakukan menggunakan optimizer Adam dengan learning rate 0,001, batch size 64, dan maksimum 20 epoch. Selain itu, diterapkan Early Stopping dengan nilai patience 3 serta ReduceLROnPlateau dengan factor 0,5 dan patience 2 untuk meningkatkan stabilitas proses pelatihan.

TABEL 3
HYPERPARAMETER MODEL BILSTM

Parameter	Nilai
Embedding	Glove
Dimensi	100
BiLSTM Layer 1	128 Unit
BiLSTM Layer 2	64 Unit
BiLSTM Layer 3	32 Unit
Dense Layer	64
Dropout	0,6
Optimizer	Adam
Learning Rate	0,001
Epoch	20
Batch Size	64
Early Stopping	Patience = 3
ReduceLROnPlateau	Factor = 0,5; Patience = 2

Fungsi aktivasi yang digunakan pada lapisan output adalah softmax, yang dirumuskan pada Persamaan 3.

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad (3)$$

Dimana z_i = skor logit kelas ke-i, K = merupakan jumlah total kelas yang diprediksi. Model dioptimasi menggunakan fungsi kerugian categorical cross-entropy, yang digunakan untuk mengukur perbedaan antara distribusi label sebenarnya dan prediksi model, dirumuskan pada Persamaan 4.

$$L = -\sum_{i=1}^{N} \sum_{j=1}^{K} y_{ij} \, log \, log \left(\widehat{y_{ij}}\right) \quad (4)$$

Dimana y_{ij} = label sebenarnya (1 jika benar di kelas ke-j, 0 jika tidak), $\widehat{y_{ij}}$= probabilitas prediksi pada kelas ke-j, N = jumlah sampel, K = jumlah kelas.

I. Evaluasi

Evaluasi dilakukan untuk mengukur performa model klasifikasi sentimen terhadap data uji[23]. Digunakan metode Confusion Matrix yang membandingkan hasil prediksi dengan label sebenarnya dan menghasilkan beberapa metrik evaluasi utama yang dirumuskan pada Persamaan 5,6,7 dan 8.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$Presisi = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$F1 = 2 \cdot \frac{Presisi \cdot Recall}{Presisi+Recall} \quad (8)$$

Dimana TP menunjukkan jumlah prediksi yang benar ketika model mengklasifikasikan data sebagai positif dan memang benar positif, TN (True Negative) adalah jumlah prediksi benar untuk data negatif, FP mengacu pada jumlah kesalahan prediksi saat model mengklasifikasikan data sebagai positif padahal sebenarnya negatif, FN (False Negative) menunjukkan jumlah data positif yang salah diklasifikasikan sebagai negatif oleh model. Dengan evaluasi ini, performa model dalam mengklasifikasikan sentimen dapat dinilai secara objektif dan menyeluruh berdasarkan keseimbangan ketepatan dan cakupan klasifikasi.

III. HASIL DAN PEMBAHASAN

A. Hasil Pengumpulan Data

Data penelitian ini diperoleh dari ulasan pelanggan pada sembilan outlet Restoran X yang tersebar di wilayah Jabodetabek melalui platform *Google Maps*. Secara keseluruhan, terdapat 10.451 ulasan yang digunakan sebagai dataset penelitian. Untuk melihat distribusi jumlah data berdasarkan lokasi pengambilan data, dilakukan rekapitulasi jumlah ulasan pada masing-masing outlet. Hasil rekapitulasi tersebut ditunjukkan pada Tabel 4.

TABEL 4
REKAPITULASI DATA ULASAN BERDASARKAN OUTLET RESTORAN X

No	Nama Outlet	Jumlah Ulasan
1	Restoran X – Cempaka Putih	1401
2	Restoran X – Cimone	1010
3	Restoran X – Daan Mogot	1200
4	Restoran X – Hankam	1100
5	Restoran X – Jatiwaringin	1110
6	Restoran X – Kisamuan	960
7	Restoran X – Margonda	1330
8	Restoran X – Pekayon	1310
9	Restoran X – Transyogi	1030

Berdasarkan Tabel 4, dapat diketahui bahwa jumlah ulasan pada setiap outlet mempunyai variasi yang berbeda. Hal ini menunjukkan adanya perbedaan tingkat aktivitas pelanggan di masing-masing lokasi. Outlet dengan jumlah ulasan tertinggi adalah Restoran X – Cempaka Putih, sedangkan jumlah terendah terdapat pada Restoran X – Kisamaun. Untuk memberikan gambaran mengenai bentuk data yang digunakan dalam penelitian, ditampilkan contoh struktur dataset yang terdiri dari atribut nama pengguna, isi ulasan, rating, serta tahun penulisan tersebut ditunjukkan pada Tabel 5.

TABEL 5
CONTOH DATA ULASAN

Nama	Ulasan	Rating	Tahun
Dimas Tri	Abis re-opening bukannya cepet malah tambah lama, payah pelayanan lelet	1	2024
Endang Dwi	Tambah stafnya, menunggu terlalu lama	2	2023
Rizky Saputra	juru parkir maksa minta bikin males	3	2021
Ryan Huda	aduhhh ini lama banget dtg makannya,	4	2024
Sekar Fitri	Bersih ramah bagus	5	2025

Berdasarkan Tabel 5, data ulasan tersebut kemudian digunakan sebagai input utama dalam proses analisis. Sebelum dilakukan pemodelan, data terlebih dahulu melalui tahap praproses untuk membersihkan dan menstandarkan teks. Setelah tahap praproses, data direpresentasikan ke dalam bentuk numerik menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) untuk menghasilkan bobot kata yang merepresentasikan tingkat kepentingan suatu term dalam dokumen.

B. TF-IDF

Setelah tahap prapemrosesan selesai dilakukan, metode *Term Frequency–Inverse Document Frequency* (TF-IDF) diterapkan untuk mengubah teks ulasan menjadi representasi numerik berbobot. Teknik ini menentukan tingkat pentingnya suatu kata dengan mempertimbangkan seberapa sering kata tersebut muncul dalam sebuah dokumen serta seberapa jarang kata tersebut ditemukan pada keseluruhan korpus. Nilai yang dihasilkan kemudian digunakan sebagai fitur utama dalam proses analisis lanjutan, seperti klasifikasi sentimen dan pemodelan topik. Hasil perhitungan skor TF-IDF tersebut disajikan pada Tabel 6.

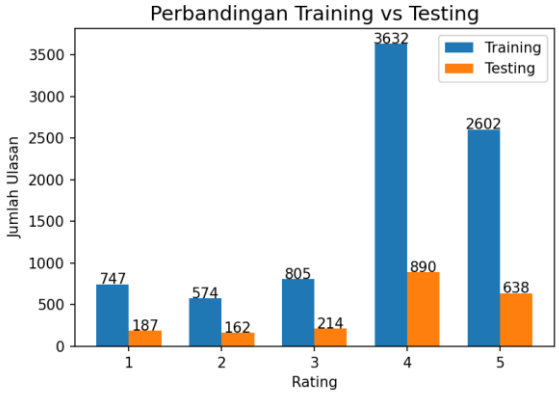
TABEL 6
SKOR TF-IDF KATA DOMINAN

Kata	Skor TF-IDF
sangat	0.0803
bagus	0.0375
enak	0.0312
tempat	0.0272
dan	0.0267

Berdasarkan Tabel 6 menunjukkan lima kata dengan skor TF-IDF tertinggi yang merepresentasikan kata paling informatif dalam korpus ulasan pengguna. Selanjutnya, matriks fitur TF-IDF digunakan sebagai representasi data untuk tahap pembagian dataset ke dalam data latih dan data uji.

C. Pembagian Data

Pada tahap ini, data ulasan dibagi menjadi dua bagian utama, yaitu data pelatihan (*training data*) dan data pengujian (*testing data*). Pembagian dilakukan dengan proporsi 80% untuk data pelatihan dan 20% untuk data pengujian guna memastikan model dapat dilatih dan dievaluasi secara optimal. Hasil pembagian data ditunjukkan pada Gambar 2.



Gambar 2. Hasil Pembagian Data

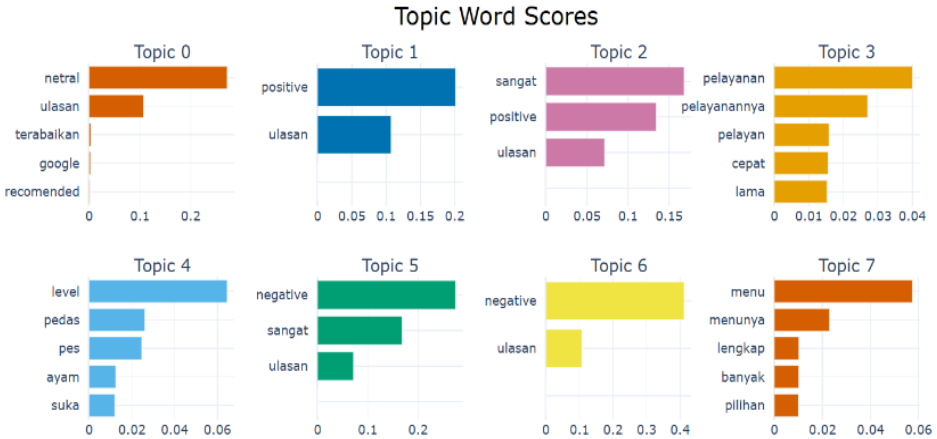
Berdasarkan Gambar 2, terlihat bahwa distribusi data pada data pelatihan dan pengujian telah terbagi sesuai dengan proporsi yang ditetapkan. Selain itu, dapat diamati bahwa distribusi rating didominasi oleh rating 4 dan 5 pada kedua subset data. Hasil ringkasan statistik dapat dilihat pada Tabel 7.

TABEL 7
HASIL RINGKASAN STATISTIK

Variable	Value
Total Ulasan	10451
Rata-rata Rating	3.80
Rating yang Paling Sering Muncul	Rating 4
Data Pelatihan (Training Data)	8360
Data Pengujian (Testing Data)	2091

D. Pemodelan

Tahapan pemodelan dalam penelitian ini menggunakan algoritma BERTopic untuk mengidentifikasi topik utama dari data ulasan pelanggan. BERTopic memanfaatkan embedding berbasis transformer serta clustering untuk mengelompokkan dokumen berdasarkan kemiripan semantik, serta menggunakan class-based TF-IDF (c-TF-IDF) untuk mengekstraksi kata kunci pada setiap topik. Hasil pemodelan menghasilkan sejumlah topik beserta kata kunci dominannya. Hasil ditunjukkan pada Gambar 3.



Gambar 3. Hasil Pemodelan Topic Word Scores

E. Klasifikasi Model

Hasil klasifikasi sentimen terhadap 10.451 ulasan memperlihatkan jika mayoritas ulasan berada pada kategori positif. Sebanyak 4.522 ulasan terklasifikasi sebagai Positif dan 3.240 sebagai Sangat Positif, sehingga total 7.762 ulasan (sekitar 74,3%) menunjukkan kecenderungan sentimen positif. Sementara itu, 736 ulasan termasuk Negatif dan 934 ulasan Sangat Negatif, dengan total 1.670 ulasan (sekitar 16,0%) menunjukkan sentimen negatif. Sisanya sebanyak 1.019 ulasan (sekitar 9,7%) berada pada kategori Netral. Hasil distribusi sentimen ini ditunjukkan pada Tabel 8.

TABEL 8
DISTRIBUSI HASIL KLASIFIKASI SENTIMEN

Sentimen	Jumlah	Persentase
Sangat Positif	3.240	31.0%
Positif	4.522	43.3%
Netral	1.019	9.7%
Negatif	736	7.0%
Sangat Negatif	934	8.9%

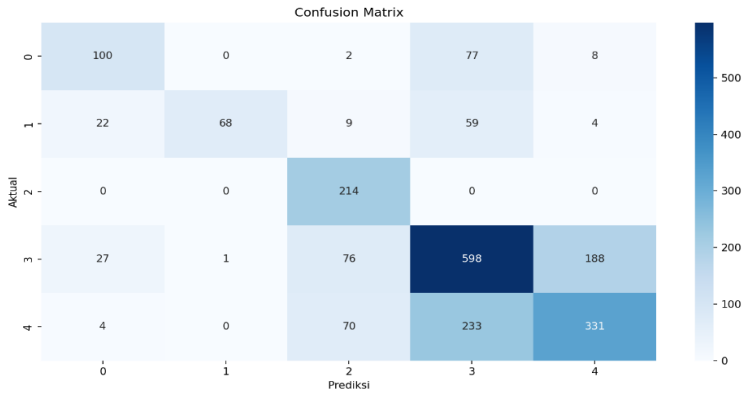
Untuk memperkuat analisis, dilakukan perhitungan distribusi rating pengguna. Hasil memperlihatkan jika mayoritas pengguna memberikan rating tinggi, yaitu 43,3% memberikan rating 4 dan 31,0% memberikan rating 5. Sementara itu, rating 3 senilai 9,8%, rating 1 senilai 8,9%, dan rating 2 senilai 7,0%. Hal ini memperlihatkan jika persepsi pengguna terhadap layanan cenderung positif dan konsisten dengan hasil klasifikasi sentimen. Hasil distribusi rating ditunjukkan pada Tabel 9.

TABEL 9
DISTRIBUSI RATING PENGGUNA

Rating	Jumlah / Persentase
5	3.240 (31.0%)
4	4.522 (43.3%)
3	1.019 (9.7%)
2	736 (7.0%)
1	934 (8.9%)

F. Evaluasi

Evaluasi model dilakukan menggunakan confusion matrix. Berdasarkan hasil pengujian diperoleh akurasi 78,6%, precision 81,6%, recall 83,5%, dan F1-score 82,5%. Hasil evaluasi menunjukkan performa model yang cukup optimal dalam klasifikasi sentimen. Hasil confusion matrix ditunjukkan pada Gambar 4.



Gambar 4. Hasil Confusion Matrix

IV. SIMPULAN

Penelitian ini difokuskan pada analisis sentimen dan mengidentifikasi topik utama pada ulasan pelanggan Restoran X menggunakan pendekatan BiLSTM dan BERTopic. Hasil eksperimen memperlihatkan jika model BiLSTM berhasil memperoleh tingkat akurasi senilai 78,6% pada tugas klasifikasi sentimen., sementara BERTopic berhasil mengekstraksi topik utama yang merepresentasikan isu dominan berupa kualitas layanan, kecepatan penyajian, tingkat kepedasan makanan, dan variasi menu. Temuan ini memperlihatkan jika kombinasi BiLSTM dan BERTopic efektif dalam menghasilkan analisis yang tidak hanya bersifat klasifikatif tetapi juga tematik, sehingga memberikan pemahaman yang lebih terhadap opini pelanggan. Secara praktis, pendekatan ini dapat dimanfaatkan sebagai dasar pengembangan sistem pemantauan ulasan pelanggan untuk mendukung peningkatan kualitas layanan restoran. Adapun keterbatasan dalam penelitian ini terletak pada penggunaan data yang hanya berasal dari satu sumber, yakni platform *Google Maps*, sehingga generalisasi model pada platform atau domain lain masih terbatas, selain itu model juga mempunyai tantangan dalam menangani variasi bahasa informal pada ulasan serta belum dilakukannya evaluasi lintas domain. Untuk penelitian selanjutnya, disarankan penggunaan model berbasis *transformer* seperti IndoBERT, perluasan dataset lintas platform dan lintas domain, serta penggunaan skema pelabelan yang lebih robust untuk meningkatkan kualitas dan generalisasi hasil analisis.

DAFTAR PUSTAKA

[1] Ariyani et al, "Pengaruh Harga, Lokasi Dan Kualitas Layanan Terhadap Kepuasan Konsumen," *Jurnal Ekonomi Dan Manajemen*, pp. 23–28, Jun. 2023.
[2] Huang et al, "Analysis and Recognition of Food Safety Problems in Online Ordering Based on Reviews Text Mining," *Wirel Commun Mob Comput*, vol. 2022, 2022, doi: 10.1155/2022/4209732.
[3] Verma et al, "Opinion and Suggestion Mining on Customer Reviews," *Int J Res Appl Sci Eng Technol*, vol. 11, no. 5, pp. 2524–2535, 2023, doi: 10.22214/ijraset.2023.52148.
[4] Rigobon et al, "Business Closures and (Re)Openings in Real-Time Using Google Places: Proof of Concept," *Journal of Risk and Financial Management*, vol. 15, no. 4, 2022, doi: 10.3390/jrfm15040183.
[5] Wang et al, "Multi-task Learning with BERT in NLP."
[6] Royani et al, "Topic Modelling Latent Dirichlet Allocation untuk Klasifikasi Komentar pada Layanan Streaming Platform," *JST (Jurnal Sains dan Teknologi)*, vol. 12, no. 3, Jan. 2024, doi: 10.23887/jstundiksha.v12i3.68492.
[7] Tahyudin et al, "Penerapan Teknik Heuristik untuk Meningkatkan Rekomendasi Produk dari Data Transaksi di Toko Jaffamart Menggunakan Query SQL," vol. 17, no. 1, pp. 50–71, 2025.
[8] R. A. Prasetyo, "Perbandingan Algoritma Logistic Regression, Support Vector Machine, dan Random Forest untuk Analisis Sentimen Ulasan Pengguna GoPay," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 10, no. 4, pp. 1178–1188, 2025.
[9] Angga et al, "Analisis Sentimen Pengguna Aplikasi MyPertamina Menggunakan Algoritma Bidirectional Long Short Term Memory," *Jurnal Eksplora Informatika*, vol. 13, no. 2, pp. 156–163, Mar. 2024, doi: 10.30864/eksplora.v13i2.973.
[10] A. Muzakir and U. Suriani, "Model Deteksi Berita Palsu Menggunakan Pendekatan Bidirectional Long Short Term Memory (BiLSTM)," J. Comput. Inf. Syst. Ampera, vol. 4, no. 2, pp. 93–105, May 2023, doi: 10.51519/journalcisa.v4i2.397.
[11] Q. Aini, M. N. F. Hidayat, and A. Tholib, "Analisis Sentimen Hasil Pemilu (Quick Count) Calon Presiden dan Wakil Presiden 2024 di Media Sosial X Menggunakan Metode Bidirectional Long Short-Term Memory (BiLSTM)," Journal of System and Cybernetics, vol. 5, no. 3, 2024, doi: 10.47065/josyc.v5i3.5223.
[12] Grootendorst, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," Mar. 2022, [Online]. Available: http://arxiv.org/abs/2203.05794
[13] Zhu et al, "BERTopic-Driven Stock Market Predictions: Unraveling Sentiment Insights," Apr. 2024, [Online]. Available: http://arxiv.org/abs/2404.02053
[14] Nursyahrina et al, "Metode BERTopic dan LDA untuk Analisis Tren Penelitian Bidang Ilmu Komputer," *Jurnal KomtekInfo*, pp. 332–341, Sep. 2024, doi: 10.35134/komtekinfo.v11i4.580.
[15] Franco et al, "Modelling the structure of the sports management research field using the bertopic approach," pp. 648–663, 2023.
[16] Wahyuni et al, "Development of Authentic Assessment Models in Research Methods Courses," 2022.
[17] Álvarez et al, "Streamlining Text Pre-Processing and Metrics Extraction," in Frontiers in Artificial Intelligence and Applications, *IOS Press BV*, Oct. 2022, pp. 55–58. doi: 10.3233/FAIA220314.
[18] R. K. Dey and A. K. Das, "Modified term frequency-inverse document frequency based deep hybrid framework for sentiment analysis," *Multimed Tools Appl*, vol. 82, no. 21, pp. 32967–32990, Sep. 2023, doi: 10.1007/s11042-023-14653-1.
[19] Z. Labd, S. Bahassine, K. Housni, F. Z. A. H. Aadi, and K. Benabbes, "Text classification supervised algorithms with term frequency inverse document frequency and global vectors for word representation: A comparative study," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 1, pp. 589–599, Feb. 2024, doi: 10.11591/ijece.v14i1.pp589-599.
[20] F. Lan, "Research on Text Similarity Measurement Hybrid Algorithm with Term Semantic Information and TF-IDF Method," *Advances in Multimedia*, vol. 2022, 2022, doi: 10.1155/2022/7923262.
[21] Jain et al, "A Diagnostic Approach to Assess the Quality of Data Splitting in Machine Learning," Jun. 2022, [Online]. Available: http://arxiv.org/abs/2206.11721
[22] George et al, "An integrated clustering and BERT framework for improved topic modeling," *International Journal of Information Technology (Singapore)*, vol. 15, no. 4, pp. 2187–2195, Apr. 2023, doi: 10.1007/s41870-023-01268-w.
[23] Xu et al, "Supplementary Features of BiLSTM for Enhanced Sequence Labeling," Jun. 2023, [Online]. Available: http://arxiv.org/abs/2305.19928
[24] Areshey et al, "Transfer Learning for Sentiment Classification Using Bidirectional Encoder Representations from Transformers (BERT) Model," *Sensors*, vol. 23, no. 11, Jun. 2023, doi: 10.3390/s23115232.
[25] R. A. Prasetyo, "Perbandingan Algoritma Logistic Regression, Support Vector Machine, dan Random Forest untuk Analisis Sentimen Ulasan Pengguna GoPay," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 10, no. 4, pp. 1178–1188, 2025.