

Analisis Sentimen Ulasan Google Maps Menggunakan *Long Short-Term Memory* (LSTM) (Studi Kasus: Kafe di Surabaya)

Asmaul Khusna¹, Yuni Yamasari²

^{1,2} Program Studi Teknik Informatika, Fakultas Teknik, Universitas Negeri Surabaya

¹asmaul.21046@mhs.unesa.ac.id

²yuniyamasari@unesa.ac.id

Abstrak – Pertumbuhan pesat kafe di Surabaya telah menjadikan ulasan Google Maps sebagai sumber informasi krusial bagi konsumen, namun volume ulasan yang besar menyulitkan untuk analisis secara manual. Penelitian ini bertujuan melakukan analisis sentimen pada ulasan kafe tersebut menggunakan metode *Long Short-Term Memory* (LSTM) yang diperkuat dengan mekanisme *Attention*. Data ulasan dikumpulkan melalui teknik *scraping* menggunakan SerpAPI (114.960 ulasan dari 218 kafe) dan dibagi secara sekuensial (*Sequential split*) untuk pelatihan dan pengujian. Ulasan diklasifikasikan ke dalam tiga kelas sentimen: positif, netral, dan negatif. Model dibangun dengan arsitektur spesifik: *embedding layer* berdimensi 128, LSTM (64 unit), *attention layer*, *dropout* (0,5), dan *dense layer* dengan tiga *output neuron*. Evaluasi kinerja model, yang diukur menggunakan akurasi, presisi, *recall*, dan *F1-score*, menunjukkan hasil terbaik dengan akurasi 92,45%, presisi 91,84%, *recall* 92,45%, dan *F1-score* 92,01%. Model terbukti cukup efektif dalam mendeteksi sentimen ekstrem (positif dan negatif), tetapi kesulitan dalam mengklasifikasikan kelas netral akibat ketidakseimbangan distribusi data. Hal ini menunjukkan bahwa penyeimbangan data atau penyederhanaan kelas sentimen diperlukan untuk meningkatkan performa model secara keseluruhan pada penelitian mendatang.

Kata Kunci – Analisis Sentimen, LSTM, Google Maps, Ulasan Kafe, Deep Learning.

I. PENDAHULUAN

Perkembangan pesat teknologi digital telah membawa perubahan besar pada kehidupan masyarakat Indonesia, khususnya dalam mempermudah akses informasi melalui berbagai *platform* daring [1]. Salah satu platform yang banyak digunakan oleh masyarakat untuk mencari informasi terkait tempat usaha, restoran, dan kafe adalah Google Maps [2]. Melalui fitur ulasan (*review*), pengguna dapat berbagi pengalaman, menilai, serta memberikan pendapat mereka tentang tempat tersebut, mencakup kualitas pelayanan, suasana, maupun pilihan menu yang disajikan [3].

Sebagai salah satu kota metropolitan terkemuka, Surabaya menunjukkan pertumbuhan yang signifikan dalam industri kuliner, khususnya sektor kafe pada tahun 2020 hingga 2023 [4]. Kafe kini tidak lagi berfungsi sebatas tempat bersantai, tetapi juga sebagai ruang sosial bagi masyarakat untuk bekerja, belajar, dan berinteraksi. Melimpahnya pilihan kafe di Surabaya membuat ulasan pengguna di Google Maps menjadi sumber informasi krusial bagi calon pelanggan dalam mengambil keputusan [5]. Namun, jumlah ulasan yang sangat

besar dan sifatnya tidak terstruktur menimbulkan tantangan besar bagi pengguna untuk menganalisis opini tersebut secara manual [6]. Ulasan pelanggan memiliki sifat yang tidak terstruktur, penggunaan bahasa informal, dan seringnya mengandung campuran bahasa sehingga menjadikannya sulit untuk diinterpretasikan tanpa adanya dukungan teknologi. Kondisi data yang kompleks ini menciptakan kebutuhan mendesak akan pengembangan sistem yang memiliki kapabilitas untuk menganalisis opini pengguna secara otomatis dan menghasilkan pemetaan sentimen dengan tingkat akurasi yang tinggi.

Untuk mengatasi tantangan analisis data tersebut, diperlukan sebuah pendekatan berbasis analisis sentimen otomatis yang memiliki kapabilitas untuk mengklasifikasikan opini pengguna secara efisien dan akurat [7]. Analisis sentimen, sebagai salah satu disiplin ilmu dalam *Natural Language Processing* (NLP), bertujuan untuk mengevaluasi kecenderungan emosi atau opini yang terkandung dalam teks, dan selanjutnya mengelompokkannya ke dalam kategori positif, netral, atau negatif. Teknik ini telah terbukti efektif dan banyak dimanfaatkan untuk memahami persepsi publik terhadap produk maupun layanan tertentu [8].

Beberapa penelitian terdahulu terkait analisis sentimen ulasan Google Maps telah dilakukan dengan berbagai metode. Penelitian oleh Fa'iz Yahya et al. (2025) menggunakan metode *Long Short-Term Memory* (LSTM) untuk menganalisis ulasan rumah sakit dan memperoleh akurasi sebesar 96% [9]. Penelitian Sri Lestanti et al. (2024) mengombinasikan TextBlob dan Naïve Bayes dengan akurasi 94%, namun menghadapi kendala ketidakseimbangan data [10]. Selain itu, Oktaviani et al. (2024) membandingkan performa LSTM dan BERT pada ulasan *coffee shop* dan menemukan bahwa model BERT memberikan hasil lebih unggul dengan akurasi 99% [11]. Berdasarkan penelitian tersebut, LSTM tetap relevan untuk analisis sentimen teks berbahasa Indonesia karena kemampuannya menangkap konteks urutan kata secara efektif. Meskipun penelitian-penelitian tersebut menunjukkan hasil yang menjanjikan, sebagian besar menggunakan dataset dengan skala terbatas dan belum melakukan eksplorasi mendalam terhadap kombinasi *hyperparameter* untuk optimasi model LSTM.

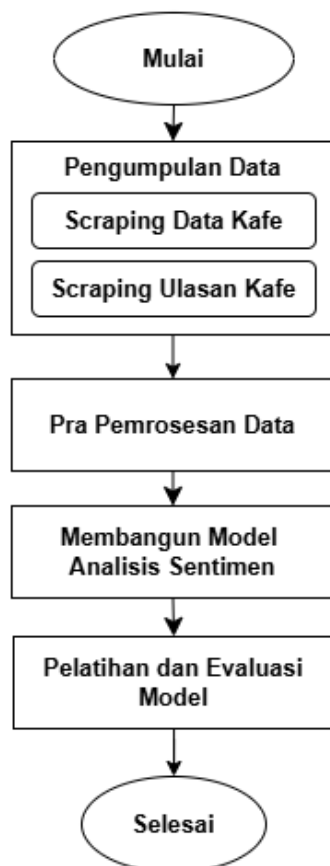
Berdasarkan latar belakang dan penelitian terdahulu, penelitian ini menerapkan metode *Long Short-Term Memory* (LSTM) yang diintegrasikan dengan mekanisme *Attention* untuk melakukan analisis sentimen pada ulasan Google Maps

kafe di Kota Surabaya. Data yang digunakan dikumpulkan melalui SerpApi, mencakup total 114.960 ulasan dari 218 kafe. Proses pelatihan model menggunakan pendekatan *sequential split*, di mana pengujian dilakukan terhadap beberapa parameter kunci. Parameter yang dieksplorasi meliputi rasio pembagian data (mulai dari 90:10 hingga 50:50), *batch size* (8 hingga 256), *epoch* (10 hingga 50), dan *learning rate* (0.01 hingga 0.0005). Evaluasi performa model dilakukan menggunakan metrik standar seperti akurasi, presisi, recall, dan F1-score untuk mengidentifikasi konfigurasi parameter optimal. Kontribusi utama penelitian ini meliputi: penggunaan dataset skala besar yang representatif untuk analisis sentimen kafe di Surabaya, evaluasi komprehensif terhadap kombinasi *hyperparameter* model LSTM, dan validasi model menggunakan data terpisah untuk memastikan generalisasi performa pada data *real-world*.

II. METODOLOGI PENELITIAN

Penelitian ini terdiri dari beberapa tahapan utama untuk membangun model analisis sentimen ulasan kafe berbasis *Long Short-Term Memory* (LSTM). Metodologi yang diterapkan mencakup pengumpulan data, prapemrosesan (*preprocessing*), pembangunan arsitektur model, serta tahapan dan evaluasi performa. Secara keseluruhan, alur penelitian ini dapat dilihat pada Gbr 1.

A. Alur Penelitian



Gbr 1. Diagram Alur Penelitian

Gbr 1 menunjukkan alur penelitian yang terbagi menjadi 4 tahapan utama. Tahap awal adalah pengumpulan data melalui *scraping* informasi kafe dan ulasan pelanggan yang bersumber dari Google Maps. Tahap kedua adalah prapemrosesan data, yang bertujuan untuk membersihkan dan menyiapkan teks ulasan agar siap diolah. Selanjutnya, tahap ketiga berfokus pada pembangunan model analisis sentimen dengan mengadopsi arsitektur *Long Short-Term Memory* (LSTM) yang dilengkapi dengan *Attention*. Tahap akhir yaitu pelatihan model, tahap ini meliputi pengujian berbagai konfigurasi *hyperparameter* dan di akhir dengan evaluasi menggunakan metrik klasifikasi standar.

B. Pengumpulan Data

Data ulasan kafe yang digunakan dalam penelitian ini dikumpulkan dari Google Maps melalui teknik *scraping* yang terbagi menjadi dua tahapan. Tahapan pertama melibatkan pengambilan informasi umum kafe di Kota Surabaya menggunakan ekstensi *Instant Data Scraper* dengan kata kunci “kafe di Surabaya”. Informasi yang berhasil dikumpulkan meliputi *link* Google Maps kafe, nama kafe, alamat, *rating*, jumlah ulasan, harga, jam operasional, dan jenis tempat.

Tahap kedua adalah pengambilan ulasan pelanggan dari masing-masing kafe menggunakan SerpAPI dan bahasa pemrograman Python. Proses ini dilakukan secara iteratif dengan mengidentifikasi setiap kafe melalui *Place ID* Google Maps hingga seluruh ulasan berhasil dikoleksi, mencakup nama pengulas, teks ulasan, *rating*, waktu ulasan, dan gambar yang dicantumkan.

C. Prapemrosesan Data

Setelah proses pengumpulan data ulasan selesai, langkah selanjutnya adalah tahap prapemrosesan untuk membersihkan dan menyiapkan data. Tahap ini diterapkan meliputi langkah-langkah berikut:

1. Penghapusan Kolom Tidak Relevan
Menghapus kolom-kolom yang tidak berpengaruh terhadap proses analisis sentimen.
2. Pembersihan Data
Menghapus data kosong, duplikat, karakter khusus, angka yang tidak relevan, dan emoji.
3. Penyaringan Data
Menghapus ulasan atau komentar yang terlalu pendek, khususnya yang memiliki kurang dari 10 kata.
4. *Case Folding*
Mengubah semua teks menjadi huruf kecil untuk memastikan konsistensi dalam analisis teks.
5. Tokenisasi
Memecahkan teks ulasan menjadi individual agar dapat dianalisis dengan lebih baik.
6. Penghapusan *Stopwords*
Menghapus kata-kata umum yang tidak memiliki makna signifikan, seperti “yang”, “dan”, “atau”.
7. Pelabelan Sentimen
Setiap ulasan diberi label sentimen berdasarkan deskripsi ulasan. Seperti ulasan “Ini kafe yg enak makanannya, plattingnya bagus, coffeenya mantap. Cocok utk wfc atau nongkrong bareng teman.” termasuk

label positif, dan ulasan “maaf bangetttt 🙏 rasa makanan terutama untuk menu yang seperti dimsum nggak banget 🙏🙏 waktu di gigit dalamnya gamateng, kecutt dab jujur baunya ganggu banget sampai ga kemakann 🙏🙏🙏” termasuk label negatif.

Hasil dari tahap prapemrosesan ini adalah dataset ulasan yang telah dibersihkan, dinormalisasi, dan telah memiliki label sentimen. Dataset ini siap digunakan untuk tahap eksperimen model sentimen.

D. Arsitektur Model LSTM

Model analisis sentimen pada penelitian ini dikembangkan menggunakan arsitektur *Long Short-Term Memory* (LSTM) yang diintegrasikan dengan mekanismen *Attention*. LSTM dipilih karena unggul dalam menangkap ketergantungan (dependesi) jangka panjang dalam data sekuensial, sebuah kemampuan krusial untuk memahami konteks pada teks ulasan yang kompleks. Penambahan mekanisme *Attention* lbih lanjut memungkinkan model untuk memfokuskan perhatian pada kata-kata yang paling berkontribusi dalam menentukan polaritas sentimen.

Sebelum data teks dimasukkan kedalam model, dilakukan preprocessing tambahan berupa tokenisasi dan *padding*. Teks ulasan dikonversi menjadi representasi numerik menggunakan *Tokenizer Keras*, dengan pembatasan *vocabulary* sebanyak 10.000 kata teratas. Kata-kata di luar batas *vocabulary* dipresentasikan dengan token *<OOV>*. Setiap urutan token kemudian diseragamkan panjangnya menjadi 100 token mealui proses *padding* atau pemotongan, guna menyelaraskan dimensi input.

Selain itu, label sentimen (Positif, Netral, Negatif) yang semula berupa teks dikonversi menjadi format numerik (0,1,2) menggunakan *LabelEncoder* dari *scikit-learn* agar sesuai untuk pemrosesan oleh model *neural network*. Arsitektur model yang digunakan terdiri dari beberapa layer yang tersusun secara berurutan untuk memproses data teks dan mengklasifikasikannya ke dalam tiga kategori sentimen. Detail arsitektur yang diimplementasikan dapat dilihat pada Tabel 1.

Tabel I
Arsitektur Model LSTM dengan Attention

Layer	Type	Parameter	Output Shape
1	Input	shape=(100,)	(None, 100)
2	Embedding	vocab_size=10000, output_dim=128	(None, 100, 128)
3	LSTM	units=64, return_sequences=True	(None, 100, 64)
4	Attention	Custom Layer	(None, 64)
5	Dropout	rate=0.5	(None, 64)
6	Dense	Units=3, activation=softmax	(None, 3)

Model dikompilasi menggunakan *optimizer Adam* dengan *loss function 'sparse_categorical_crossentropy'* dan metrik evaluasi akurasi. Penambahan *Attention* membantu model untuk fokus pada kata-kata penting dalam menentukan sentimen ulasan.

E. Pelatihan dan Evaluasi Model

Pelatihan dan evaluasi model dioptimalkan melalui serangkaian eksperimen terstruktur. Eksperimen ini menggunakan teknik *grid search* untuk secara sistematis menguji semua kombinasi empat *hyperparameter*, yang meliputi rasio pembagian data (data latih dan data uji), *batch size*, jumlah *epoch*, dan *learning rate*.

Tabel II menunjukkan nilai-nilai *hyperparameter* yang dikombinasikan dalam proses *grid search*. Total kombinasi yang diuji adalah $5 \times 6 \times 5 \times 4 = 600$ konfigurasi berbeda.

Tabel II
Kombinasi Nilai Hyperparameter

No.	Hyperparameter	Nilai yang Diuji
1	Rasio Data Latih:Uji	90:10, 80:20, 70:30, 60:40, 50:50
2	Batch Size	8, 16, 32, 64, 128, 256
3	Epoch	10, 20, 30, 40, 50
4	Learning Rate	0.01, 0.001, 0.0001, 0.0005

Setiap konfigurasi *hyperparameter* diuji menggunakan dataset utama dengan teknik pembagian data *sequential split*. Model dilatih pada data latih dan kinerjanya diverifikasi pada data uji. Evaluasi performa dilakukan menggunakan metrik klasifikasi multi-kelas standar yaitu akurasi, presisi, *recall*, dan *F1-score*.

Akurasi mengukur proporsi prediksi yang benar dari seluruh data, dihitung menggunakan rumus:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

Presisi mengukur tingkat ketepatan model dalam memprediksi kelas tertentu, yaitu proporsi prediksi positif yang benar:

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Recall mengukur kemampuan model dalam mendeteksi semua data yang termasuk dalam kelas tertentu:

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

F1-score merupakan rata-rata harmonik anatar presisi dan *recall*, memberikan keseimbangan antara kedua metrik:

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (4)$$

III. HASIL DAN PEMBAHASAN

A. Hasil Pengumpulan Data

Dataset yang digunakan dalam penelitian ini merupakan hasil *scraping* dari Google Maps, yang dikumpulkan dalam dua tahap berbeda. Tahap pertama adalah pengumpulan informasi umum mengenai kafe, sedangkan tahap kedua adalah pengumpulan ulasan dari masing-masing kafe. Total dataset yang diperoleh mencapai 114.960 ulasan dari 218 kafe di

surabaya, dataset ini kemudian dibagi menjadi dua set dengan tujuan berbeda. Dataset utama (111.823 ulasan dari 212 kafe) digunakan untuk pelatihan dan pengujian model. Dataset ini dibagi menjadi data latih dan data uji dengan berbagai rasio (90:10 hingga 50:50) menggunakan metode *sequential split*. Pemilihan *sequential split* bertujuan untuk mempertahankan urutan temporal data, di mana data terbaru digunakan sebagai data uji, mensimulasikan skenario prediksi di masa depan. Sementara itu dataset validasi (3.137 dari 6 kafe) disisihkan secara independen untuk menguji kemampuan generalisasi model pada data yang baru.

Tabel III
Jumlah Dataset

No.	Jenis Data	Jumlah Kafe	Jumlah Ulasan	Keterangan Penggunaan
1	Data Utama	212	111.823	Digunakan untuk pelatihan dan pengujian model
2	Data Validasi	6	3.137	Digunakan untuk validasi performa model
Total		218	114.960	

Pembagian dataset dengan menggunakan kafe yang berbeda untuk validasi bertujuan untuk memastikan bahwa model dapat digeneralisasi dengan baik pada data dari kafe yang tidak pernah dilihat selama proses pelatihan. Hal ini penting untuk aplikasi praktis dimana model diharapkan dapat memberikan prediksi akurat pada ulasan kafe baru yang belum pernah dianalisis sebelumnya.

B. Hasil Prapemrosesan Data

Proses ini merupakan tahap penting untuk membersihkan dan menyiapkan teks agar siap digunakan untuk pelatihan model analisis sentimen. Prapemrosesan dilakukan menggunakan Python dengan bantuan pustaka *pandas*, *nltk*, *Sastrawi*, *re*, dan lainnya. Berikut merupakan hasil dari tahapan prapemrosesan:

1. Penghapusan Kolom Tidak Relevan
Menghapus kolom-kolom yang tidak berpengaruh terhadap analisis sentimen, seperti: *page*, *link*, *thumbnail*, *images*, dan *local_guide*.
2. Penghapusan Ulasan Kosong
Ulasan yang kosong (baik NaN maupun string kosong) dihapus dari dataset karena tidak memberikan informasi yang bisa dianalisis oleh model. Berikut merupakan jumlah ulasan sementara pada proses ini:

Tabel IV
Jumlah Dataset Penghapusan Ulasan Kosong

	Jumlah Ulasan Awal	Jumlah Ulasan Kosong	Jumlah Ulasan Setelah Bersih
Data Utama	111.823	33.818	78.005

	Jumlah Ulasan Awal	Jumlah Ulasan Kosong	Jumlah Ulasan Setelah Bersih
Data Validasi	3.137	1.129	2.008

3. Penyaringan Ulasan Pendek

Proses ini dilakukan dengan menghitung jumlah kata pada setiap ulasan menggunakan metode pemisahan berdasarkan spasi (*whitespace tokenization*). Hanya ulasan yang memiliki ≥ 10 kata yang dipertahankan. Tabel berikut menunjukkan jumlah data sebelum dan sesudah proses penyaringan ulasan pendek:

Tabel V
Jumlah Data Penyaringan Ulasan Pendek

	Jumlah Ulasan Sebelum Penyaringan	Jumlah Ulasan Pendek (<10 kata)	Jumlah Ulasan Setelah Penyaringan
Data Utama	78.005	36.674	41.331
Data Validasi	2.008	908	1.100

4. Penghapusan Ulasan Duplikat

Ulasan yang memiliki isi yang sama dihapus berdasarkan kolom snippet untuk mencegah bias pada proses pelatihan model. Tabel berikut menunjukkan jumlah data sebelum dan sesudah proses penghapusan ulasan duplikat:

Tabel VI
Jumlah Ulasan Duplikat yang Dihapus

	Jumlah Ulasan Sebelum Penghapusan	Jumlah Ulasan Duplikat	Jumlah Ulasan Setelah Penghapusan
Data Utama	41.331	35	41.296
Data Validasi	1.100	1	1.099

5. Pembersihan Teks (*Cleaning Text*)

Pembersihan ini bertujuan untuk menyederhanakan struktur teks agar mudah dianalisis. Hasil dari tahapan pembersihan teks ditampilkan pada Gbr 2.

snippet	cleaned
Pasti ada alasan dong kenapa Mandailing masih ...	Pasti ada alasan dong kenapa Mandailing masih ...
Br pertama nyoba makan berat di sini. Nasi nya...	Br pertama nyoba makan berat di sini Nasi nya ...
Cozy banget tempatnya\nPorsi makanannya banyak...	Cozy banget tempatnya Porsi makanannya banyak ...
Resto yang mengusung No msg No preservatives k...	Resto yang mengusung No msg No preservatives k...
Rasa makanan cukup enak\nUntuk porsi sedang s...	Rasa makanan cukup enak Untuk porsi sedang saj...
cafe favorit, selain karena suasananya nyaman ...	cafe favorit selain karena suasananya nyaman d...
Place was great, food doesn't come with the sa...	Place was great food doesnt come with the same...
suka curry nya kalo mampir kesini pasti beli m...	suka curry nya kalo mampir kesini pasti beli m...
Ke sini lagi setelah sekian lama, pilihan menu...	Ke sini lagi setelah sekian lama pilihan menu ...
Tempatnya kurang luas dan kurang instragramable...	Tempatnya kurang luas dan kurang instragramable...

Gbr 2. Hasil Pembersihan Teks

6. Case Folding

Seluruh teks diubah menjadi kecil (*lowercase*) agar kata dengan huruf kapital tidak dianggap berbeda. Hasil dari tahapan *case folding* ditampilkan pada Gbr 3.

Cleaned	casefolded
Pasti ada alasan dong kenapa Mandailing masih ...	pasti ada alasan dong kenapa mandailing masih ...
Br pertama nyoba makan berat di sini Nasi nya ...	br pertama nyoba makan berat di sini nasi nya ...
Cozy banget tempatnya Porsi makanannya banyak ...	cozy banget tempatnya porsi makanannya banyak ...
Resto yang mengusung No msg No preservatives K...	resto yang mengusung no msg no preservatives k...
Rasa makanan cukup enak Untuk porsi sedang saj...	rasa makanan cukup enak untuk porsi sedang saj...
cafe favorit selain karena suasananya nyaman d...	cafe favorit selain karena suasananya nyaman d...
Place was great food doesnt come with the same...	place was great food doesnt come with the same...
suka curry nya kalo mampir kesini pasti beli m...	suka curry nya kalo mampir kesini pasti beli m...
Ke sini lagi setelah sekian lama pilihan menu ...	ke sini lagi setelah sekian lama pilihan menu ...
Tempatnya kurang luas dan kurang instragamable...	tempatny kurang luas dan kurang instragamable...

Gbr 3. Hasil Case Folding

7. Tokenisasi

Teks yang sudah bersihkan kemudian dipecah menjadi token (kata-kata individual). Proses ini membantu midel dalam mengenail struktur kalimat secara kata per kata. Hasil dari tahapan tokenisasi ditampilkan pada Gbr 4.

casefolded	tokenized
pasti ada alasan dong kenapa mandailing masih ...	[pasti, ada, alasan, dong, kenapa, mandailing,...
br pertama nyoba makan berat di sini nasi nya ...	[br, pertama, nyoba, makan, berat, di, sini, n...
cozy banget tempatnya porsi makanannya banyak ...	[cozy, banget, tempatnya, porsi, makanannya, b...
resto yang mengusung no msg no preservatives K...	[resto, yang, mengusung, no, msg, no, preserva...
rasa makanan cukup enak untuk porsi sedang saj...	[rasa, makanan, cukup, enak, untuk, porsi, sed...
cafe favorit selain karena suasananya nyaman d...	[cafe, favorit, selain, karena, suasananya, ny...
place was great food doesnt come with the same...	[place, was, great, food, doesnt, come, with, ...
suka curry nya kalo mampir kesini pasti beli m...	[suka, curry, nya, kalo, mampir, kesini, pasti...
ke sini lagi setelah sekian lama pilihan menu ...	[ke, sini, lagi, setelah, sekian, lama, pilih...
tempatny kurang luas dan kurang instragamable...	[tempatny, kurang, luas, dan, kurang, instrag...

Gbr 4. Hasil Tokenisasi

8. Normalisasi Kata Tidak Baku

Menormalisasi ulasan tidak baku menggunakan kamu slang (*slang.txt*), misalnya “gk” menjadi “tidak”, “bgt” menjadi “banget”. Hasil dari tahapan normalisasi ditampilkan pada Gbr 5.

tokenized	normalized
[pasti, ada, alasan, dong, kenapa, mandailing,...	[pasti, ada, alasan, dong, kenapa, mandailing,...
[br, pertama, nyoba, makan, berat, di, sini, n...	[br, pertama, nyoba, makan, berat, di, sini, n...
[cozy, banget, tempatnya, porsi, makanannya, b...	[cozy, banget, tempatnya, porsi, makanannya, b...
[resto, yang, mengusung, no, msg, no, preserva...	[resto, yang, mengusung, no, msg, no, preserva...
[rasa, makanan, cukup, enak, untuk, porsi, sed...	[rasa, makanan, cukup, enak, untuk, porsi, sed...

Gbr 5. Hasil Normalisasi Kata Tidak Baku

9. Penghapusan Stopwords

Stopwords adalah kata-kata umum yang sering muncul namun tidak memiliki makna penting. Kata-kata tersebut dihapus menggunakan daftar *stopwords* Bahasa Indonesia dari pustaka *nlTK*. Hasil dari tahapan ini ditampilkan pada Gbr 6.

normalized	stopword_removed
[pasti, ada, alasan, dong, kenapa, mandailing,...	[alasan, mandailing, buka, cafe, ga, bertahan,...
[br, pertama, nyoba, makan, berat, di, sini, n...	[br, nyoba, makan, berat, nasi, nya, keras, di...
[cozy, banget, tempatnya, porsi, makanannya, b...	[cozy, banget, tempatnya, porsi, makanannya, e...
[resto, yang, mengusung, no, msg, no, preserva...	[resto, mengusung, no, msg, no, preservatives,...
[rasa, makanan, cukup, enak, untuk, porsi, sed...	[makanan, enak, porsi, harga, ya, lumayan, rel...

Gbr 6. Hasil Penghapusan Stopwords

10. Pelabelan Sentimen

Proses pelabelan dilakukan secara manual dengan membaca satu per satu ulasan. Pendekatan manual ini dilakukan untuk memastikan akurasi dan relevansi label sentimen, karena dalam beberapa kasus *rating* numerik tidak selalu mencerminkan isi ulasan secara utuh. Misalnya, ada ulasan dengan rating tinggi namun isi teksnya mengandung keluhan, atau sebaliknya.

C. Proses Pelatihan dan Evaluasi Model

Tahapan pelatihan dan evaluasi model dilaksanakan untuk mengidentifikasi konfigurasi *hyperparameter* terbaik yang dapat mencapai performa klasifikasi sentimen paling optimal. Proses optimasi ini memanfaatkan pendekatan *grid search* (sebagaimana diuraikan dalam Bab II), yang menguji total 600 kombinasi *hyperparameter* yang berbeda.

Seluruh proses pelatihan diimplementasikan menggunakan *framework TensorFlow* dan *Keras* dalam lingkungan Google Colaboratory dengan dukungan GPU untuk mempecepat komputasi. Dataset utama dibagi menggunakan metode *sequential split*, di mana pembagian tersebut untuk data latih dan data uji.

Performa setiap konfigurasi dievaluasi secara komprehensif menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*. Tabel VII menampilkan hasil perbandingan beberapa konfigurasi representatif yang menunjukkan variasi performa model terhadap perubahan *hyperparameter*.

Tabel VII
Hasil Eksperimen

No	Rasio Data (D.latih:D. Uji)	Batch Size	Epoch	Learning Rate	Akurasi
1	90:10	8	10	0,01	0,9163
2	90:10	8	10	0,001	0,9035
3	90:10	8	10	0,0001	0,9172
...
41	90:10	32	10	0,01	0,9245
...
598	50:50	256	50	0,001	0,9026
599	50:50	256	50	0,0001	0,9026

No	Rasio Data (D.latih:D. Uji)	Batch Size	Epoch	Learning Rate	Akurasi
600	50:50	256	50	0,0005	0,8963

Berdasarkan sebagian hasil yang ditampilkan pada Tabel VII, akurasi model LSTM berada dalam kisaran 0.8826 hingga 0,9245. Pola kinerja menunjukkan bahwa rasio pembagian data 90:10 secara konsisten menghasilkan akurasi yang lebih tinggi dibandingkan dengan rasio 50:50. Hal ini disebabkan oleh kuantitas data latih yang lebih besar pada rasio 90:10, memungkinkan model untuk mempelajari dan menggeneralisasi pola sentimen secara lebih efektif.

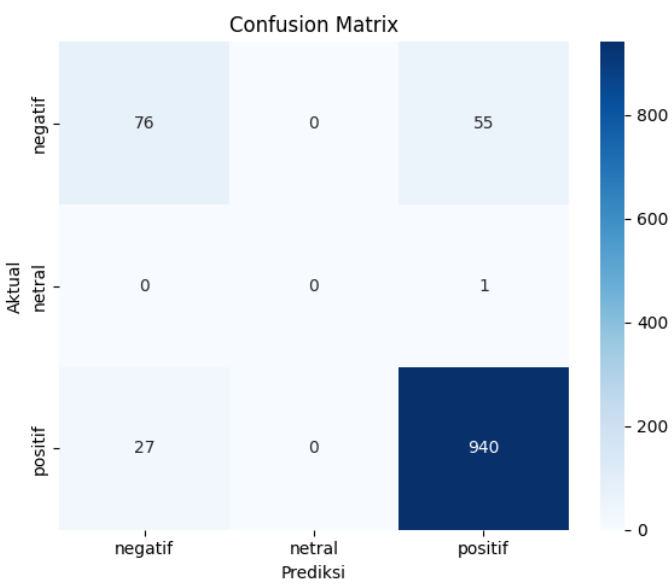
Selain rasio data, *batch size* dan *learning rate* juga memengaruhi performa model secara signifikan. Sebagai contoh, kombinasi *batch size* 32 dan *learning rate* 0,01 menunjukkan kinerja yang stabil dengan akurasi diatas 0,91. Kinerja ini kontras dengan konfigurasi pada rasio 50:50 yang cenderung memiliki akurasi lebih rendah, yaitu kisaran 0,89-0,90.

Setelah mengevaluasi seluruh eksperimen, konfigurasi optimal teridentifikasi pada percobaan ke-41. Konfigurasi ini, yang menggunakan rasio data 90:10, *batch size* 32, *epoch* 10, dan *learning rate* 0,01, berhasil mencapai akurasi tertinggi sebesar 0,9245. Konfigurasi ini selanjutnya ditetapkan sebagai parameter paling optimal untuk model LSTM ini.

Model dengan konfigurasi terbaik tersebut kemudian digunakan untuk analisis lebih lanjut termasuk pembuatan *confusion matrix* dan evaluasi per kelas sentimen untuk menilai akurasi model terhadap masing-masing label sentimen (positif, netral, negatif)

D. Analisis Confusion Matrix

Untuk memperoleh pemahaman yang lebih mendalam mengenai kinerja model, evaluasi dilakukan menggunakan *confusion matrix* yang memperlihatkan distribusi prediksi model terhadap label aktual untuk setiap kelas sentimen. Gbr 7 menyajikan *confusion matrix* dari model dengan konfigurasi terbaik (90:10, *batch size* 32, *epoch* 10, *learning rate* 0.01) ketika diterapkan pada dataset validasi independen.



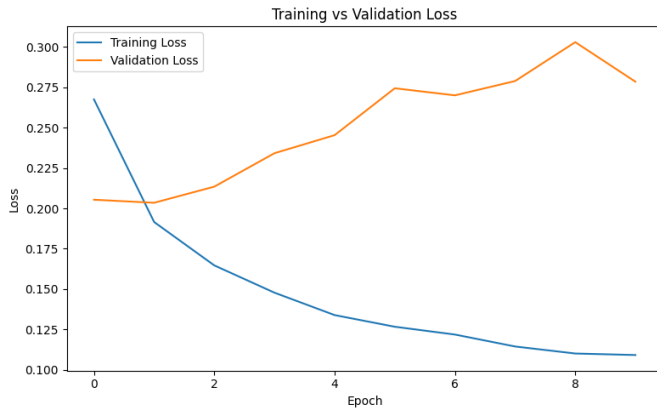
Gbr 7. Confusion Matrix Model LSTM dengan Attention

Berdasarkan *confusion matrix* tersebut, terlihat bahwa model menunjukkan kemampuan prediksi yang sangat kuat untuk kelas positif, di mana 940 data positif berhasil diklasifikasikan dengan tepat. Namun, terdapat bias yang jelas terhadap kelas positif, karena hampir semua data dari kelas lain (netral dan sebagian negatif) juga diprediksi sebagai positif.

Kelas netral menunjukkan kinerja terburuk, karena tidak ada satupun data netral yang diklasifikasikan dengan benar (1 data netral bahkan diklasifikasikan sebagai positif). Hal ini bisa disebabkan oleh jumlah data netral yang sedikit (kelas minoritas) atau fitur yang kurang mampu membedakan nuansa netral dari positif/negatif.

Untuk kelas negatif, kinerja model tergolong masih bisa mengenali sebagian besar dengan benar (76 benar dari total 131 data negatif), meskipun terdapat 55 data negatif yang keliru diklasifikasikan sebagai netral atau positif. Kesalahan klasifikasi pada kelas negatif ini umumnya disebabkan oleh pengguna bahasa yang ambigu, sindiran, atau sarkasme dalam ulasan, yang sulit dideteksi secara eksplisit oleh model berbasis *deep learning*.

E. Analisis Kurva Training dan Validation Loss



Gbr 8. Training vs Validation Loss

Gbr 8 menampilkan perbandingan antara *training loss* dan *validation loss* selama 10 epoch pelatihan model LSTM dengan mekanisme *Attention*. Grafik menunjukkan bahwa nilai *training loss* (garis biru) mengalami penurunan stabil, dari sekitar 0,27 pada *epoch* pertama menjadi 0,10 pada *epoch* terakhir, mengidentifikasi bahwa model berhasil mempelajari pola dalam data latih secara efektif. Sebaliknya, *validation loss* (garis orange) tetap stabil hingga *epoch* ketiga, kemudian menunjukkan peningkatan bertahap, mencapai nilai sekitar 0,30 pada *epoch* ke-10.

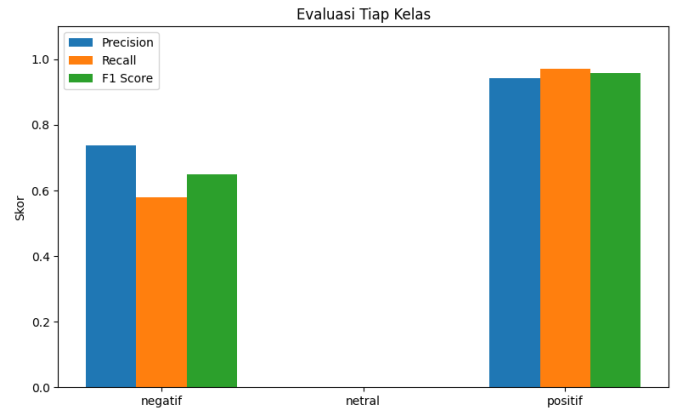
F. Evaluasi Per Kelas Sentimen

Evaluasi performa per kelas dilakukan untuk mengukur kemampuan model secara lebih spesifik dalam mengenali masing-masing kategori sentimen, yaitu positif, netral, dan negatif. Pengukuran ini menggunakan tiga metrik utama: presisi, *recall*, dan *F1-score*. Ketiganya metrik ini secara kolektif memberikan gambaran mengenai ketepatan (*precision*), kelengkapan (*recall*), dan keseimbangan performa model (*F1-score*) dalam setiap kelas sentimen. Hasil evaluasi per kelas disajikan pada Tabel VIII berikut:

Tabel VIII
Evaluasi Per Kelas Sentimen

Kelas Sentimen	Presisi	Recall	F1-score	Jumlah Sampel
Negatif	0.74	0.58	0.65	131
Netral	0.00	0.00	0.00	1
Positif	0.94	0.97	0.96	967
Weighted Avg	0.92	0.92	0.92	1.099

Gbr 9 menampilkan perbandingan nilai presisi, *recall*, dan *F1-score* pada setiap kelas sentimen.



Gbr 9. Perbandingan Presisi, Recall, dan F1-score per Kelas Sentimen

Hasil evaluasi per kelas menunjukkan bahwa model mencapai performa tertinggi pada kelas sentimen positif, ditandai dengan presisi 0.94, *recall* 0.97, dan *F1-score* 0.96. angka ini membuktikan bahwa model sangat efektif dalam mengidentifikasi ulasan positif dan memiliki tingkat kesalahan prediksi yang minimal untuk kategori ini.

Sebaliknya, kelas netral menunjukkan performa terendah, dengan ketiga metrik bernilai 0.00. Kondisi ini mengindikasikan bahwa model gagal mengidentifikasi ulasan netral, yang hampir seluruhnya diklasifikasikan ke dalam kelas lain (dominan ke positif). Faktor utama penyebabnya adalah ketidakseimbangan data yang ekstrem, mengingat sampel ulasan netral pada *dataset* validasi hanya terdiri dari satu data. Selain itu, ulasan netral sering kali memiliki fitur linguistik yang mirip dengan ulasan positif ringan, mempersulit diskriminasi oleh model.

Sementara itu, performa model pada kelas negatif berada pada tingkat menengah, dengan presisi 0.74 dan *recall* 0.58. Nilai *recall* yang relatif rendah menunjukkan adanya sejumlah data negatif yang keliru diklasifikasikan sebagai positif. Fenomena ini dapat timbul akibat penggunaan bahasa yang ambigu, sarkasme, atau bentuk penilaian tidak langsung dalam ulasan, yang sulit dideteksi secara eksplisit oleh model berbasis teks.

Secara keseluruhan, evaluasi per kelas mengonfirmasi adanya kecenderungan bias model terhadap kelas positif akibat dominasi data tersebut dalam *dataset*. Meskipun demikian, nilai *weighted average F1-score* sebesar 0.92 menegaskan bahwa model tetap mempertahankan performa klasifikasi yang stabil dan akurat secara umum. Untuk mencapai keseimbangan klasifikasi yang lebih baik di masa depan, penelitian lanjutan dapat mempertimbangkan penerapan teknik penyeimbangan data (*data balancing*) atau augmentasi teks.

IV. KESIMPULAN

Secara keseluruhan, penelitian ini telah berhasil dalam mengembangkan sebuah model analisis sentimen yang andal untuk mengolah ulasan kafe di Google Maps, khususnya di area Surabaya. Kita menggunakan pendekatan canggih, yaitu *Long Short-Term Memory* (LSTM) yang diperkuat dengan mekanisme *Attention*, untuk memproses data ulasan berbahasa

Indonesia yang sudah melalui proses pembersihan dan prapemrosesan yang ketat.

Dari total 114.960 ulasan yang berhasil dikumpulkan, proses optimasi model menjadi kunci. Dengan menerapkan skema *grid search* yang komprehensif, peneliti berhasil menemukan konfigurasi *hyperparameter* terbaik (rasio data 90:10, *batch size* 32, *epoch* 10, dan *learning rate* 0,01) yang menghasilkan performa yang sangat memuaskan. Model ini mencapai akurasi 92,45% dan nilai F1-score yang tinggi sebesar 92,01%. Keunggulan model tampak jelas dalam mengidentifikasi sentimen positif, dengan F1-score mencapai 96%. Meskipun masih ada ruang perbaikan untuk sentimen negatif dan terutama kelas netral (yang terdampak parah oleh ketidakseimbangan data ekstrem), kemampuan model untuk menggeneralisasi sentimen pada ulasan kafe baru yang belum pernah dilihat sebelumnya terbukti konsisten dan baik.

Penemuan ini memiliki implikasi praktis yang signifikan. Sistem analisis sentimen yang kita kembangkan ini menawarkan solusi otomatis dan efisien bagi pemilik kafe. Dengan mengolah volume besar *feedback* pelanggan, model dapat membantu mereka mengidentifikasi kekuatan utama yang harus dipertahankan dan kelemahan spesifik yang menuntut perbaikan segera. Kesuksesan penerapan metode LSTM-Attention ini menegaskan bahwa pendekatan ini sangat efektif untuk analisis sentimen ulasan berbahasa Indonesia, membuka jalan bagi aplikasi serupa di berbagai sektor lain.

V. SARAN

Berdasarkan hasil penelitian dan berbagai tantangan yang masih ditemukan, berikut beberapa saran yang dapat menjadi panduan bagi penelitian selanjutnya:

- penelitian selanjutnya perlu menerapkan teknik penyeimbangan data seperti *SMOTE* (*Synthetic Minority Over-sampling Technique*), *ADASYN*, atau teknik augmentasi teks untuk meningkatkan representasi kelas minoritas. Alternatif yang lebih sederhana adalah menjadikan klasifikasi menjadi dua kelas (positif,= dan negatif).
- Untuk mengoptimalkan akurasi, penelitian dapat fokus pada arsitektur yang lebih bagus, seperti *Bidirectional LSTM* (Bi-LSTM), GRU, atau *transformer-based models* seperti BERT (IndoBERT untuk bahasa Indonesia). Model *ensemble* yang menggabungkan beberapa arsitektur juga dapat dipertimbangkan untuk meningkatkan akurasi.
- Penelitian lanjutan dapat mengembangkan model untuk melakukan analisis sentimen berbasis aspek, yang mampu mengidentifikasi sentimen terhadap aspek spesifik seperti kualitas makanan, pelayanan, suasana, harga, dan kebersihan.
- Agar hasil penelitian lebih umum, model perlu diuji generalisasinya dengan memperluas cakupan geografis ke kota-kota besar lain atau memperluas domain ke jenis bisnis lain. Selain itu, sangat penting untuk memastikan dataset validasi independen memiliki distribusi kelas yang seimbang guna memastikan evaluasi performa model dilakukan secara objektif.

Penerapan saran-saran di atas diharapkan dapat menghasilkan model analisis sentimen yang lebih akurat, tangguh, dan aplikatif untuk mendukung pengambilan keputusan bisnis di industri kuliner, khususnya sektor kafe di Indonesia.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Pihak Universitas dan dosen pembimbing yang telah memberikan bimbingan, arahan, serta dukungan selama proses penelitian ini berlangsung. Ucapan terma kasih juga disampaikan kepada rekan-rekan penulis yang turut membantu dalam proses pengumpulan data dan pengolahan data, sehingga penelitian ini dapat terselesaikan dengan baik.

REFERENSI

- [1] T. A. Berutu, D. L. R. Sigalingging, G. K. V. Simanjuntak, and F. Siburian, "Pengaruh Teknologi Digital terhadap Perkembangan Bisnis Modern," *Neptunus: Jurnal Ilmu Komputer Dan Teknologi Informasi*, vol. 2, no. 3, pp. 358–370, Jul. 2024, doi: 10.61132/neptunus.v2i3.258.
- [2] H. Rizqian, I. Novianto, G. P. Permana, and A. C. Wahyudi, "Analisis Pemanfaatan Google Maps sebagai Fungsi Promosi dan Media EWOM pada Tempat Usaha Kuliner," *Journal of Trends Economics and Accounting Research*, vol. 5, no. 4, pp. 366–372, Jun. 2025, doi: 10.47065/jtear.v5i4.2054.
- [3] A. F. Hidayat, D. Putra Rakhmadani, and S. Van Marsally, "Pengaruh Rating Dan Review Google Maps Terhadap Purchase Decision Melalui E-Trust Sebagai Mediasi Pada Bisnis Kuliner Di Purwokerto," *Universitas Telkom*, vol. 12, no. 4, p. 4446, Aug. 2025.
- [4] Badan Pusat Kota Surabaya, *Kota Surabaya dalam Angka 2024*, vol. 36. 2024.
- [5] H. Rizqian, I. Novianto, G. Argo Permana, and A. Candra Wahyudi, "Analisis Pemanfaatan Google Maps sebagai Fungsi Promosi dan Media EWOM pada Tempat Usaha Kuliner," *Journal of Trends Economics and Accounting Research*, vol. 5, no. 4, pp. 366–372, Jun. 2025, doi: 10.47065/jtear.v5i4.2054.
- [6] M. R. Amalsyah, D. Kurniawan, A. Rifai, and P. Sari, "Analisis Sentimen Ulasan Pengguna Aplikasi Fintech menggunakan Framework CRISP-DM dalam Penentuan Prioritas Pengembangan Produk," *Sistemasi: Jurnal Sistem Informasi*, vol. 14, no. 2, pp. 813–825, 2025, [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- [7] D. Anggraini, S. Rahmawati, and R. Kurniawan, "Natural Language Processing For Automatic Sentiment Analysis In Social Media Data," *International Journal of Information Engineering and Science*, vol. 1, no. 1, pp. 16–19, Feb. 2024, doi: 10.62951/ijies.v1i2.54.
- [8] Y. Mao, Q. Liu, and Y. Zhang, "Sentiment analysis methods, applications, and challenges: A systematic

- literature review,” *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 4, 2024, doi: 10.1016/j.jksuci.2024.102048.
- [9] F. Yahya, D. Eka Ratnawati, and B. Rahayudi, “Analisis Sentimen Ulasan Pengguna dari Google Maps Menggunakan Metode Long Short-Term Memory (Studi Kasus: Rumah Sakit Gatoel),” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 4, pp. 2548–964, 2025, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [10] Saiful Nur Budiman, Sri Lesanti, and Erwan, “Analisis Sentimen Berdasarkan Hasil Review Lokasi Google Map Menggunakan Natural Language Toolkit TextBlob dan Naïve Bayes,” *JAMI: Jurnal Ahli Muda Indonesia*, vol. 5, no. 2, pp. 114–126, Dec. 2024, doi: 10.46510/jami.v5i2.311.
- [11] O. E. Putri *et al.*, “Analisis Sentimen Berbasis Aspek dan Deteksi Emosi pada Coffee Shop Palangka Raya Menggunakan Deep Learning,” *JOINTECOMS (Journal of Information Technology and Computer Science) p-ISSN: 2798-284X*, vol. 4, no. 3, pp. 2798–3862, Sep. 2024.