e- Jurnal Ilmiah Mahasiswa Sains UNISMA Malang (JIMSUM)
Volume: 3 / No.: 1 / Halaman 65 - 85 / Februari Tahun 2025

Analisis Sentimen pada Ulasan Konsumen Ayam Goreng Nelongso *di Google Maps* Menggunakan Metode *Bidirectional Long Short-Term Memory (Bi-LSTM)*

Sentiment Analysis of Consumer Reviews on Ayam Goreng Nelongso in Google Maps Using the Bidirectional Long Short-Term Memory (Bi-LSTM) Method

Yeremias Neno Fernando[1], Aviv Yuniar Rahman[2], Rangga Pahlevi Putra[3].

[123]Program Studi Teknik Informatika, Fakultas Teknik, Unversitas Widyagama Malang

ABSTRAK

Penelitian ini menganalisis sentimen ulasan konsumen terhadap Ayam Goreng Nelongso di Google Maps menggunakan metode Bidirectional Long Short-Term Memory (Bi-LSTM). Data yang digunakan mencakup 4.450 ulasan dengan rasio data latih dan uji bervariasi, mulai dari 90:10 hingga 10:90. Hasil evaluasi menunjukkan bahwa model Bi-LSTM memiliki performa sangat baik dengan rata-rata akurasi 98,33%, presisi 99,44%, recall 99,44%, dan F1-score 99,44%. Temuan ini menunjukkan bahwa Bi-LSTM mampu secara andal dan konsisten mengidentifikasi sentimen positif, negatif, dan netral pada data ulasan konsumen, memberikan wawasan yang bermanfaat untuk peningkatan layanan dan kualitas produk UMKM Ayam Goreng Nelongso.

Kata Kunci: *Analisis Sentimen,* Ayam Goreng Nelongso, *Google Maps, Bidirectional Long Short-Term Memory, Bi-LSTM.*

ABSTRACT

This study analyzes consumer review sentiments of Ayam Goreng Nelongso on Google Maps using the Bidirectional Long Short-Term Memory (Bi-LSTM) method. The data used consists of 4,450 reviews, with varying training and testing data ratios ranging from 90:10 to 10:90. The evaluation results show that the Bi-LSTM model performs excellently, with an average accuracy of 98.33%, precision of 99.44%, recall of 99.44%, and F1-score of 99.44%. These findings demonstrate that Bi-LSTM can reliably and consistently identify positive, negative, and neutral sentiments in consumer review data, providing valuable insights for improving the services and product quality of the MSME Ayam Goreng Nelongso.

Keywords*: Sentiment Analysis, Ayam Goreng Nelongso, Google Maps, Bidirectional Long Short-Term Memory, Bi-LSTM.*

I. PENDAHULUAN

Dalam era digital, ulasan konsumen telah menjadi salah satu sumber data penting yang dapat memberikan wawasan mendalam tentang persepsi pelanggan terhadap produk atau layanan. Google Maps, sebagai platform peta digital, tidak hanya memfasilitasi navigasi tetapi juga menyediakan ruang bagi pengguna untuk memberikan ulasan dan penilaian terhadap bisnis, termasuk di sektor kuliner.

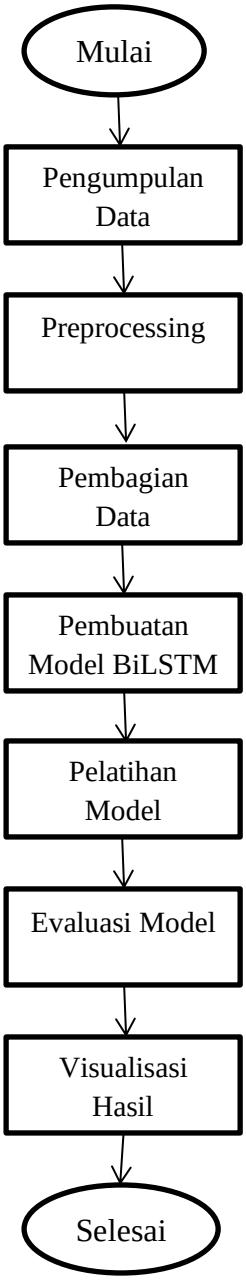
Ayam Goreng Nelongso, sebagai salah satu brand kuliner yang berkembang pesat di Indonesia, memanfaatkan ulasan pelanggan di Google Maps untuk memahami kebutuhan dan ekspektasi konsumen. Dengan lebih dari 4.000 ulasan yang mencakup berbagai aspek seperti makanan, layanan, harga, dan suasana, terdapat potensi besar untuk menganalisis sentimen pelanggan dan mengidentifikasi area yang perlu ditingkatkan.

Namun, tantangan utama adalah bagaimana mengekstraksi informasi yang bermanfaat dari data teks yang tidak terstruktur. Salah satu metode yang terbukti efektif dalam memahami sentimen dan konteks adalah Bidirectional Long Short-Term Memory (BiLSTM), yang dapat menangkap pola temporal dalam data ulasan secara lebih mendalam dibandingkan metode tradisional.

Penelitian ini bertujuan untuk menganalisis sentimen terhadap ulasan Ayam Goreng Nelongso dengan menggunakan metode BiLSTM. Dengan mennngunakan berbagai ratio perbandingan, penelitian ini diharapkan dapat memberikan wawasan yang lebih spesifik bagi pengelola bisnis untuk meningkatkan kepuasan pelanggan.

2. METODE PENELITIAN

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan machine learning untuk analisis sentimen. Data yang digunakan adalah data ulasan dari Ayam Goreng Nelongso di *Google Maps*, yang dianalisis menggunakan metode *Bidirectional Long Short-Term Memory (Bi-LSTM).* Desain Penelitian dapat dilihat pada gambar 1.



Gambar 1. Alur Penelitian

Berikut adalah penjelasan detail dari setiap poin dalam alur metode BiLSTM untuk analisis sentimen, mulai dari awal hingga selesai:

A. Mulai

Menandakan titik awal dari proses analisis sentimen dengan mempersiapkan segala kebutuhan dalam melakukan penelitian.

B. Pengumpulan Data

Pada tahap ini, data ulasan dikumpulkan dari Google Maps yang mencakup teks ulasan dan rating.

C. Pra-pemrosesan Data

Langkah ini bertujuan untuk membersihkan dan mempersiapkan data sehingga dapat digunakan oleh model. Tahapan pra-pemrosesan meliputi:

- *Text Cleaning:* Menghapus tanda baca, angka, dan karakter spesial lainnya.
- *Lowercasing*: proses mengubah semua huruf dalam teks menjadi huruf kecil.
- Penghapusan *Stop Words*: Menghapus kata-kata umum (contoh: "dan," "di") yang tidak menambah nilai semantik.
- *Stemming atau Lemmatization*: Mengurangi kata ke bentuk dasar untuk mengurangi variasi kata.
- *Tokenisasi*: Memecah teks menjadi token atau kata-kata terpisah yang dapat diolah lebih lanjut.

D. Pembagian Data (Data Splitting)

Data kemudian dibagi menjadi dua bagian: data pelatihan dan data pengujian.

E. Pembuatan Model BiLSTM

Di tahap ini, model BiLSTM dibangun dengan struktur sebagai berikut:

- Layer Embedding: Untuk mengonversi kata menjadi vektor embedding.
- Layer LSTM Bidirectional: Untuk menangkap konteks dari teks secara menyeluruh, baik dari depan ke belakang maupun sebaliknya.
- Dropout Layer: Untuk mencegah overfitting.
- Dense Layer: Untuk menghasilkan output berdasarkan jumlah kelas sentimen (misalnya: positif, netral, negatif).

F. Pelatihan Model

Model dilatih menggunakan data pelatihan. Selama proses ini, model mencoba mempelajari pola dalam data ulasan yang berkaitan dengan sentimen. Proses pelatihan melibatkan beberapa epoch, dan metrik seperti akurasi dan loss dipantau untuk memastikan model berfungsi dengan baik.

G. Evaluasi Model

Setelah pelatihan, model dievaluasi menggunakan data pengujian. Metrik evaluasi seperti akurasi, precision, recall, dan F1-score digunakan untuk mengukur kinerja model dalam mengklasifikasikan sentimen ulasan. Evaluasi ini menunjukkan seberapa baik model dapat menangani data yang belum pernah dilihat.

Metrik Evaluasi terdiri atas *Accuracy, precision, recall, dan F1-score* untuk masing-masing kelas sentimen dengan rumus sbb:

Accuracy: Proporsi prediksi yang benar dari keseluruhan data.

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

Precision: Proporsi prediksi positif yang benar.

$$Precision = \frac{TP}{TP+FP}$$

Recall (Sensitivity): Proporsi sampel positif yang terdeteksi dengan benar.

$$Recall = \frac{TP}{TP + FN}$$

F1-Score: Perbandingan rata-rata antara *precision dan recall*.

$$F1 - Score = 2 * \frac{Precision * Recall}{(Precision) + (Recall)}$$

Confusion Matrix adalah sebuah tabel yang digunakan untuk mengevaluasi kinerja model klasifikasi, terutama pada masalah supervised learning. Matriks ini memberikan gambaran tentang bagaimana prediksi model dibandingkan dengan label sebenarnya (*ground truth*) dari data. Dalam *confusion matrix*, prediksi model akan dibandingkan dengan kategori kelas yang benar, biasanya dibagi ke dalam empat elemen utama:

1. *True Positive (TP):* Jumlah sampel yang diprediksi sebagai kelas positif, dan benar-benar merupakan kelas positif.
2. *True Negative (TN)*: Jumlah sampel yang diprediksi sebagai kelas negatif, dan benar-benar merupakan kelas negatif.
3. *False Positive (FP):* Jumlah sampel yang diprediksi sebagai kelas positif, tetapi sebenarnya merupakan kelas negatif (disebut juga sebagai *Type I Error*).
4. *False Negative (FN*): Jumlah sampel yang diprediksi sebagai kelas negatif, tetapi sebenarnya merupakan kelas positif (disebut juga sebagai *Type II Error*).

H. Visualisasi Hasil Akhir

Hasil evaluasi model dan distribusi sentimen divisualisasikan untuk analisis yang lebih mudah dipahami. Tahap visualisasi meliputi:

- Menampilkan hasil evaluasi performa model berdasarkan rasio pembagian data dalam proses pelatihan dan pengujian.
- Menampilkan Metrik Performa Mode Secara Keseluruhan.

3. HASIL DAN PEMBAHASAN

A. Deskripsi Data

Data penelitian terdiri dari 4.450 ulasan dan rating konsumen Ayam Goreng Nelongso yang diambil dari Google Maps selama 5 tahun terakhir (2020–2024). Data ini mencakup dua kolom utama: Review (teks ulasan) dan Rating (1 hingga 5). Dari total data, analisis sentimen dilakukan melalui pelabelan yang menghasilkan tiga kategori sentimen: positif (2.674 data), negatif (1.147 data), dan netral (630 data). Sentimen positif mendominasi, mengindikasikan mayoritas ulasan pelanggan memberikan tanggapan baik terhadap layanan.

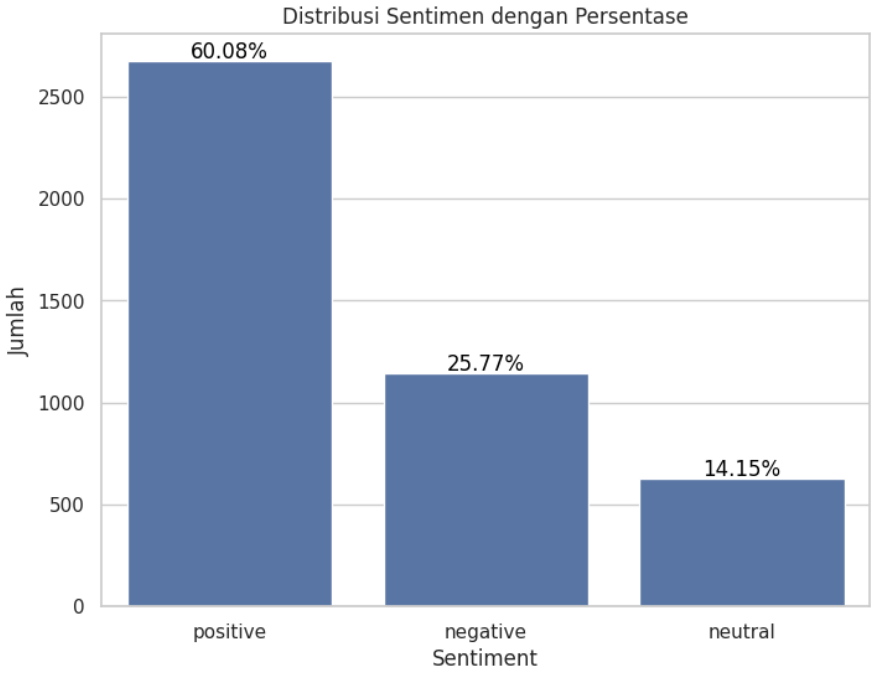
B. Pelabelan dan Distribusi Data

Dalam analisis sentimen ini, data ulasan di-label berdasarkan rating yang diberikan oleh pengguna, menggunakan fungsi label_sentiment untuk mengkategorikan rating menjadi tiga kelas sentimen: positif untuk rating 4 dan 5, negatif untuk rating 1 dan 2, serta netral untuk rating 3. Setelah pelabelan, distribusi kelas sentimen ditampilkan, menunjukkan bahwa terdapat 2.674 ulasan positif, 1.147 ulasan negatif, dan 630 ulasan netral. Visualisasi distribusi sentimen dilakukan dengan menggunakan grafik batang, di mana setiap kelas sentimen ditampilkan beserta persentase kontribusinya terhadap total ulasan. Grafik ini menunjukkan proporsi masing-masing sentimen secara jelas, memudahkan pemahaman terhadap pandangan umum pengguna terkait produk atau layanan yang dianalisis. Dengan total ulasan sebanyak 4.450, grafik memperlihatkan bahwa ulasan positif mendominasi, menyiratkan persepsi yang umumnya baik dari pengguna.Hasil Pelabelan sentimen dapat dilihat pada gambar 2 dan Distribusi Sentimen dapat di;iohat pada gambar 3.

	Rating	Review	Sentiment
0	4	Tempat nyaman luas,	positive
1	4	pelayanan ramah dan cepat.	positive
2	2	harusnya kan bersih ya? lantai kotor.	negative
3	4	Sambel dan lauknya enak serta harga nyaman dik...	positive
4	3	rasa makanan juga lumayan enak....	neutral
...	...	...	...
4446	3	Lumayan baguslah tenpatnya dan luas	neutral
4447	3	lumayan cepat,	neutral
4448	5	Makanannya enak pelayanannya ramah dan baik\nN...	positive
4449	1	Makananya sih lumayan rasanya ,harga minuman d...	negative
4450	1	Pelayanan buruk\nSuasana buruk\nNunggu sangat ...	negative

[4451 rows x 3 columns]

Gambar 2. Visualisasi hasil pelabelan sentimen



Gambar 3. Distribusi Sentimen

Dari hasil distribusi sentimen berdasarkan gambar 3. dapat dianalisis:

- Sentimen Positif (Positive): Terdapat 2.674 data ulasan yang dikategorikan sebagai sentimen positif. Jumlah ini menunjukkan bahwa sebagian besar ulasan memiliki opini atau tanggapan yang baik terhadap topik atau objek yang dianalisis. Hal ini mengindikasikan persepsi yang positif dari mayoritas pengguna terhadap aspek tertentu.
- Sentimen Negatif (Negative): Terdapat 1.147 data ulasan yang termasuk dalam kategori sentimen negatif. Jumlah ini menunjukkan adanya tanggapan atau opini yang kurang baik dari sebagian pengguna. Hal ini mencerminkan adanya aspek yang perlu diperbaiki atau dikaji lebih lanjut untuk meningkatkan kepuasan pengguna.
- Sentimen Netral (Neutral): Terdapat 630 data ulasan yang diklasifikasikan sebagai sentimen netral. Sentimen ini mencerminkan ulasan yang tidak menunjukkan kecenderungan kuat ke arah positif maupun negatif, melainkan bersifat netral atau objektif.

Data distribusi sentimen menunjukkan bahwa mayoritas ulasan bersifat positif, dengan sentimen negatif dan netral berada pada jumlah yang lebih kecil. Hasil ini memberikan gambaran umum mengenai persepsi atau opini pengguna secara keseluruhan, yang didominasi oleh pandangan positif.

C. *Preprocessing*

Tahapan *preprocessing* meliputi pembersihan teks, *lowercasing*, penghapusan kata-kata tidak relevan (*stopwords*), stemming, dan tokenisasi. Proses ini menghasilkan teks ulasan yang lebih bersih dan relevan untuk analisis. Langkah ini memastikan data siap untuk dimasukkan ke model dengan fokus pada elemen-elemen penting dari ulasan.

Berikut adalah hasil pembersihan teks dapat dilihat pada gambar 4 :

```
Contoh teks ulasan sebelum dan sesudah preprocessing:
Sebelum: Tempat nyaman luas,
Sesudah: tempat nyaman lua

Sebelum: pelayanan ramah dan cepat.
Sesudah: pelayanan ramah dan cepat

Sebelum: harusnya kan bersih ya? lantai kotor.
Sesudah: harusnya kan bersih ya lantai kotor

Sebelum: Sambel dan lauknya enak serta harga nyaman dikantong
Sesudah: sambel dan lauknya enak serta harga nyaman dikantong

Sebelum: rasa makanan juga lumayan enak....
Sesudah: rasa makanan juga lumayan enak
```

Gambar 4. Hasil Pembersihan Teks

Melalui langkah-langkah pembersihan dan preprocessing diatas, teks ulasan yang diperoleh menjadi lebih bersih dan siap untuk analisis lebih lanjut. Penerapan fungsi-fungsi tersebut memungkinkan penghilangan unsur-unsur yang tidak relevan sehingga fokus pada kata-kata penting yang dapat digunakan untuk analisis sentimen.

D. Pembagian Data

Pada penelitian ini, data dibagi ke dalam beberapa proporsi yang bervariasi antara data pelatihan dan data pengujian. Pembagian ini dilakukan untuk mengevaluasi kinerja model pada berbagai rasio data. Detail pembagian data yang digunakan dapat dilihat pada tabel 1.

No	Total Data	*Split ratio*	Data Pelatihan	Data Pengujian
1	4450	10:90	445	4005
2		20:80	890	3560
3		30:70	1335	3115
4		40:60	1780	2670
5		50:50	2225	2225
6		60:40	2670	1780
7		70:30	3115	1335
8		80:20	3560	890
9		90:10	4005	445

Tabel 1. Pembagian Data

E. Arsitektur dan Kinerja Model *Bi-LSTM*

Model BiLSTM digunakan karena kemampuannya menangkap konteks dari dua arah (maju dan mundur). Arsitektur model mencakup:

- Layer *embedding* dengan dimensi 32 untuk representasi teks.
- Layer BiLSTM dengan 128 unit untuk menangkap ketergantungan sekuensial.
- Layer *dropout* (33%) untuk mengurangi overfitting.
- Layer *dense* dengan 64 unit (*ReLU*) dan output layer (*softmax*) untuk klasifikasi tiga kelas sentimen.

Arsitektur *Bi-Lstm* dapat dilihat pada gambar 5:

```
# Membangun model
model = Sequential()
model.add(Embedding(input_dim=50000, output_dim=32, input_length=max_length))  # Definisikan input_length
model.add(Bidirectional(LSTM(128)))
model.add(Dropout(0.33))
model.add(Dense(64, activation='relu'))
model.add(Dropout(0.33))
model.add(Dense(3, activation='softmax'))

# Kompilasi model
model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])

# Bangun model berdasarkan input shape
model.build(input_shape=(None, max_length))

# Print model summary
print(model.summary())
```

Gambar 5. Arsitektur model *Bi-LSTM*

Berikut adalah penjelasan dari Arsitektur model *Bi-Lstm* pada gambar 5:

- *Sequential*: Ini adalah jenis model dalam Keras yang memungkinkan untuk menambahkan layer secara berurutan.
- *Embedding Layer:*
  - *input_dim*=50000: Ini menunjukkan bahwa model menganggap bahwa ada 50.000 kata dalam vocabulary (ukuran kosakata).
  - *output_dim*=32: Setiap kata akan direpresentasikan sebagai vektor berdimensi 32.
  - *input_length=max_length*: Panjang maksimum dari setiap input (jumlah kata dalam setiap ulasan) yang akan diterima oleh layer ini. Meskipun ada peringatan bahwa argumen input_length sudah deprecated (tidak lagi direkomendasikan), argumen ini masih bisa digunakan pada beberapa versi *TensorFlow/Keras*.
- *Bidirectional LSTM*:

  Menggunakan 128 unit untuk model *LSTM*. *Bidirectional LSTM* akan memproses data input dari dua arah (maju dan mundur), sehingga dapat memahami konteks dari kedua arah.
- *Dropout Layer:*

  Dropout(0.33): Ini adalah teknik regularisasi untuk mengurangi overfitting. Selama pelatihan, 33% neuron pada layer ini akan "dibuang" (dinonaktifkan) secara acak untuk meningkatkan generalisasi model.
- *Dense Layer:*

  Pertama: *Dense*(64, *activation='relu'*) — *Layer dense* ini memiliki 64 neuron dengan fungsi aktivasi *ReLU (Rectified Linear Unit).*

Kedua: *Dense(3, activation='softmax')* — *Layer output* yang memiliki 3 neuron, satu untuk setiap kelas sentimen (*positif, netral, negatif*) dengan fungsi aktivasi *softmax,* yang mengonversi keluaran menjadi probabilitas.

- *Loss Function:*

  categorical_crossentropy: Digunakan untuk menghitung kerugian antara keluaran model dan label sebenarnya untuk masalah klasifikasi multi-kelas.
- *Optimizer:*

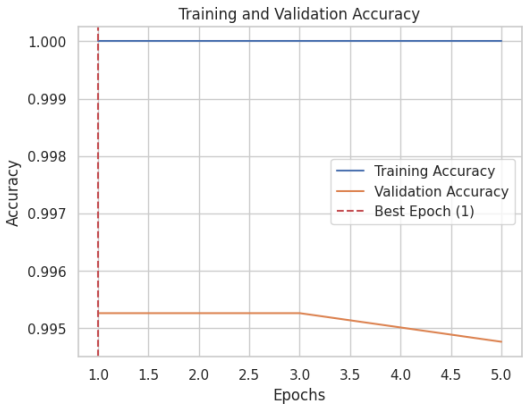
  adam: Algoritma optimasi yang efisien untuk pembaruan bobot.
- *Metrics:*

  accuracy: Metrik yang digunakan untuk mengevaluasi performa model.

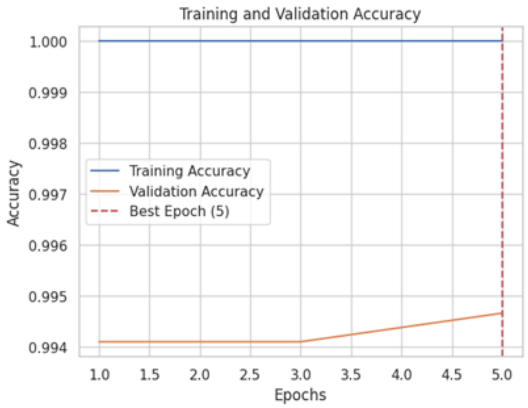
F. Pelatihan Model

Model Bi-LSTM dilatih menggunakan data pelatihan yang telah dibagi berdasarkan rasio tertentu. Pada tahap pelatihan, model mencoba untuk mempelajari pola dan hubungan dalam data ulasan yang berkaitan dengan sentimen (positif, negatif, dan netral). Proses pelatihan ini dilakukan dalam lima (5) epoch, di mana setiap epoch terdiri dari satu iterasi pelatihan penuh melalui data. Pada setiap epoch, model memproses data dalam batch, yang memungkinkan pelatihan menjadi lebih efisien. Selama pelatihan, metrik seperti akurasi dan loss dipantau untuk menilai performa model. Akurasi mengukur persentase prediksi yang benar dibandingkan dengan label yang sebenarnya, sementara loss mengukur seberapa besar kesalahan prediksi. Untuk menghindari overfitting, teknik regularisasi Dropout digunakan di beberapa lapisan model untuk mengurangi ketergantungan yang berlebihan pada fitur tertentu.

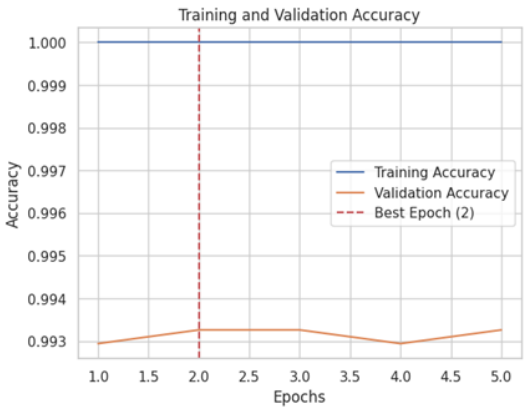
Berikut adalah hasil visualisasi selama proses pelatihan dari berbagai ratio.



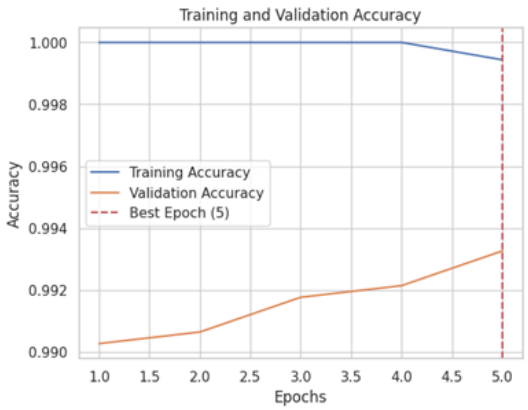
Gambar 6. Hasil training data ratio 10:90



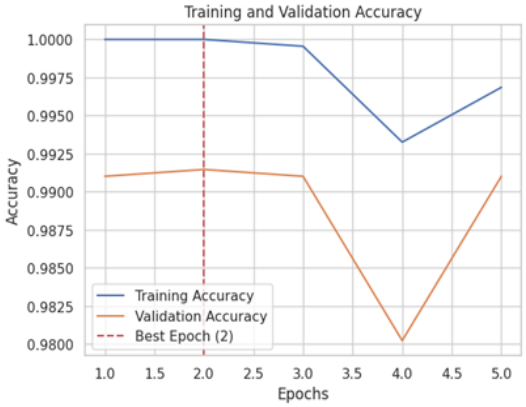
Gambar 7. Hasil training data ratio 20:80



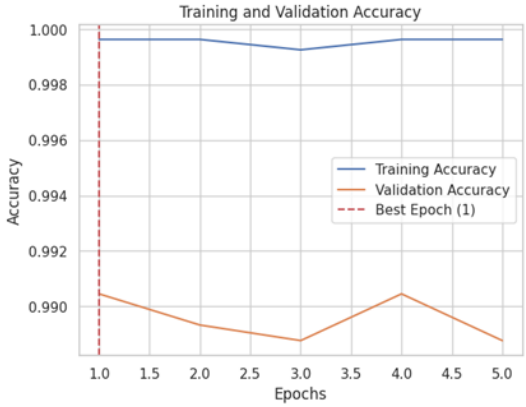
Gambar 8. Hasil training data ratio 30:70



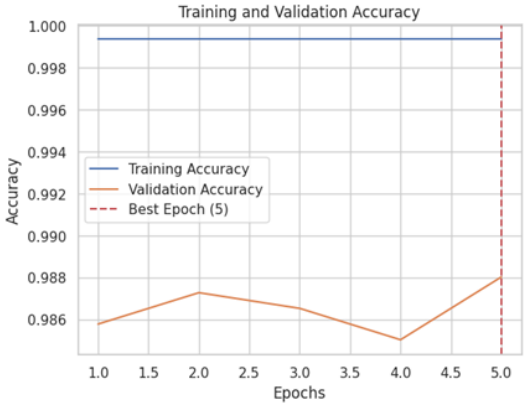
Gambar 9. Hasil training data ratio 40:60



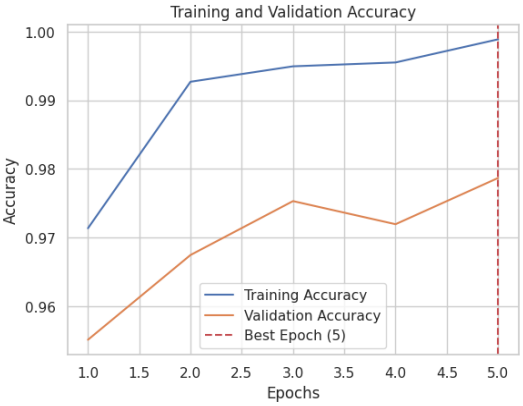
Gambar 10. Hasil training data ratio 50:50



Gambar 11. Hasil training data ratio 60:40



Gambar 12. Hasil training data ratio 70:30



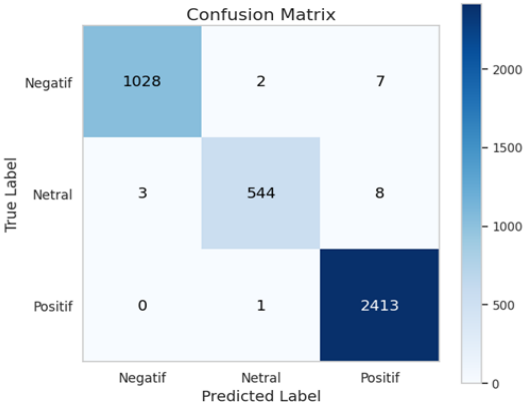
Gambar 13. Hasil training data ratio 80:90

Gambar 14. Hasil training data ratio 90:10

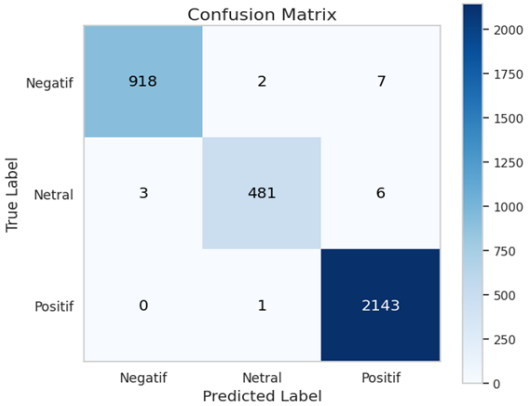
G. Evaluasi Model (Pengujian Model)

Setelah pelatihan selesai, model diuji menggunakan data pengujian yang tidak digunakan selama pelatihan. Data pengujian bertujuan untuk mengevaluasi kemampuan model dalam menggeneralisasi pengetahuan yang diperoleh pada data pelatihan ke data yang belum pernah dilihat sebelumnya. Proses ini mengukur sejauh mana model dapat mengklasifikasikan sentimen pada data baru.

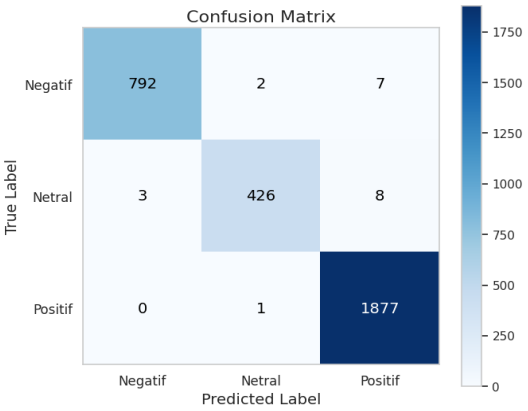
Pada tahap pengujian, model memprediksi sentimen (positif, netral, atau negatif) untuk setiap data uji dan dibandingkan dengan label yang benar. Hasil evaluasi ditampilkan dalam bentuk confusion matrix, yang menunjukkan seberapa banyak prediksi model yang benar dan salah untuk masing-masing kelas (positif, netral, negatif). Metrik evaluasi yang digunakan adalah precision, recall, dan F1 score. Precision mengukur seberapa tepat model dalam memprediksi kelas positif, recall mengukur seberapa baik model dalam mendeteksi kelas positif, dan F1 score adalah rata-rata harmonik antara precision dan recall, yang memberikan gambaran umum tentang kinerja model. Hasil prediksi pada confusion matrix diantaranya :



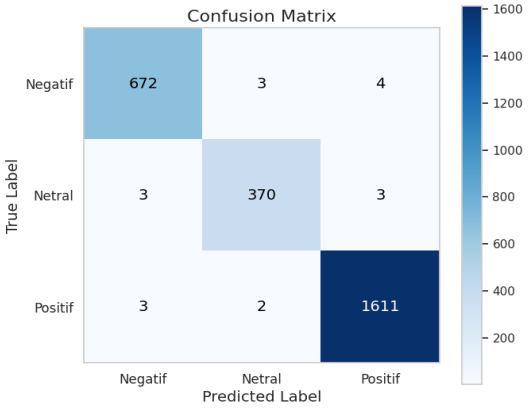
Gambar 15. *Confusion Matrix Ratio* 10:90



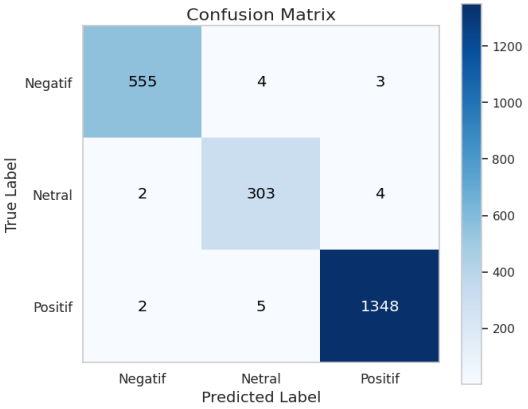
Gambar 16. *Confusion Matrix Ratio* 20:80



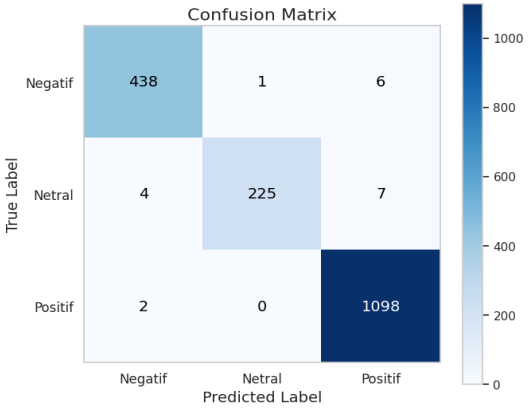
Gambar 17. *Confusion Matrix Ratio* 30:70



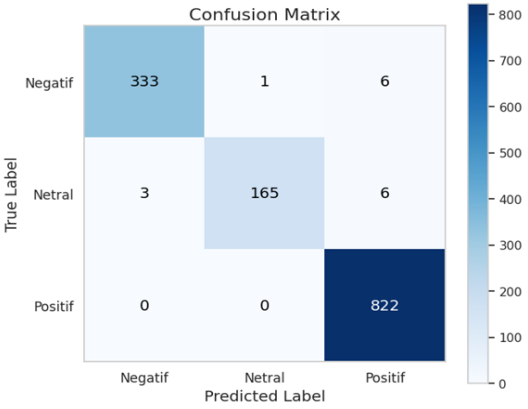
Gambar 18. *Confusion Matrix Ratio* 40:60



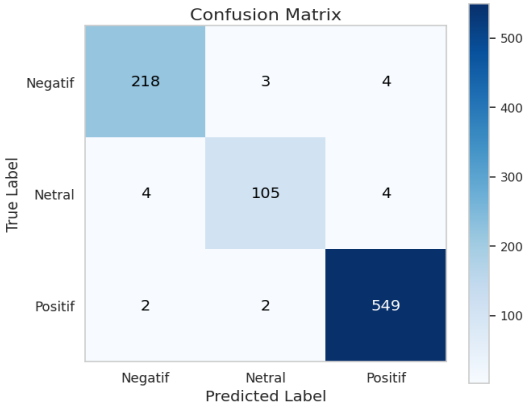
Gambar 19. *Confusion Matrix Ratio* 50:50



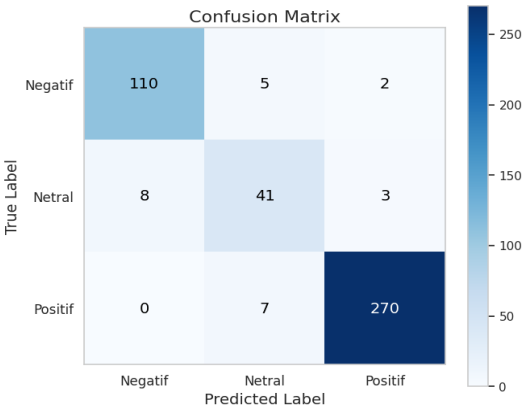
Gambar 20. *Confusion Matrix Ratio* 60:40



Gambar 21. *Confusion Matrix Ratio* 70:30



Gambar 22. *Confusion Matrix Ratio* 80:20



Gambar 23. *Confusion Matrix Ratio* 90:10

Confusion matrix dari berbagai rasio data di atas menggambarkan performa model klasifikasi sentimen (negatif, netral, positif) untuk dataset dengan distribusi *training* dan *testing* yang berbeda. Matriks ini menunjukkan hasil prediksi model dibandingkan dengan kelas sebenarnya, yang terdiri dari tiga kategori: *Negatif* (0), *Netral* (1), dan *Positif* (2). Berikut adalah analisis singkat berdasarkan matriks pada tiap rasio:

1. Ratio 10:90
    - Model memiliki performa terbaik pada kelas Positif (2), dengan jumlah prediksi benar sebesar 2413 (true positives), dibandingkan dengan kelas Netral (544) dan Negatif (1028).
    - Kesalahan terbesar terdapat pada prediksi kelas Netral (1), dengan 8 prediksi salah ke kelas Positif.
    - Prediksi untuk kelas Negatif cukup akurat, dengan hanya 7 kesalahan ke kelas Positif.
2. Ratio 20:80
    - Prediksi kelas Positif tetap paling akurat, dengan 2143 prediksi benar.
    - Pada kelas Netral, terdapat 6 kesalahan prediksi ke kelas Positif, menunjukkan sedikit kesalahan klasifikasi antara dua kelas ini.
    - Kesalahan pada kelas Negatif menurun dibandingkan ratio 10:90, yaitu hanya 7 prediksi salah ke kelas Positif.
3. Ratio 30:70
    - Model terus mempertahankan performa tinggi pada kelas Positif, dengan 1877 prediksi benar.
    - Kesalahan prediksi kelas Negatif ke kelas Positif tetap konstan (7 kesalahan).
    - Kelas Netral mengalami sedikit penurunan performa, dengan hanya 426 prediksi benar dan 8 kesalahan ke kelas Positif.
4. Ratio 40:60
    - Model menunjukkan sedikit penurunan performa di semua kelas, tetapi prediksi kelas Positif tetap dominan dengan 1611 prediksi benar.

- Kesalahan pada kelas Negatif ke kelas Positif berkurang menjadi 4, mencerminkan peningkatan presisi untuk kelas Negatif.
- Kelas Netral memiliki kesalahan 3 prediksi ke kelas Positif, yang lebih kecil dibandingkan rasio sebelumnya.

5. Ratio 50:50
    - Kinerja model tetap stabil dengan prediksi benar untuk kelas Positif sebanyak 1348, meskipun jumlah data untuk kelas ini menurun.
    - Kesalahan prediksi kelas Negatif ke kelas Positif sangat kecil, hanya 3 kesalahan.
    - Kesalahan pada kelas Netral cukup kecil dengan 4 prediksi salah ke kelas Positif.
6. Ratio 60:40
    - Model mempertahankan akurasi tinggi untuk kelas Positif (1098 prediksi benar), meskipun data training semakin berkurang.
    - Kesalahan prediksi kelas Negatif sedikit meningkat (6 kesalahan ke kelas Positif).
    - Kelas Netral memiliki kesalahan 7 prediksi ke kelas Positif, menunjukkan tantangan klasifikasi pada kelas ini.
7. Ratio 70:30
    - Prediksi benar pada kelas Positif tetap dominan (822 prediksi benar), tetapi penurunan performa terlihat pada semua kelas karena data training lebih kecil.
    - Kelas Netral memiliki 6 kesalahan ke kelas Positif, sedangkan kelas Negatif memiliki 6 kesalahan ke kelas Positif, menandakan tantangan dalam membedakan sentimen Positif dari kelas lainnya.
8. Ratio 80:20
    - Dengan data training yang semakin sedikit, performa model menurun secara keseluruhan. Kelas Positif memiliki 549 prediksi benar, tetapi kesalahan meningkat pada kelas Negatif (4 kesalahan ke kelas Positif) dan Netral (4 kesalahan ke kelas Positif).
    - Kesalahan pada kelas Netral dan Negatif mengindikasikan sulitnya membedakan sentimen yang kurang dominan.
9. Ratio 90:10
    - Dengan rasio data training yang paling kecil, performa model mengalami penurunan signifikan. Kelas Positif memiliki 270 prediksi benar, tetapi terdapat 7 kesalahan ke kelas Netral.
    - Kelas Netral menunjukkan 8 kesalahan ke kelas Negatif, sedangkan kelas Negatif memiliki 2 kesalahan ke kelas Positif, menandakan bahwa model kurang mampu menangkap pola dengan data training yang terbatas.

Dari semua rasio, performa model paling baik untuk kelas Positif karena distribusinya yang dominan dalam dataset. Namun, model cenderung mengalami kesalahan klasifikasi antara kelas Netral dan Positif. Rasio data training yang lebih besar (misalnya 10:90) menghasilkan performa terbaik, sedangkan rasio yang lebih kecil (90:10) menurunkan kemampuan model untuk memprediksi sentimen

dengan akurasi tinggi. Hal ini menegaskan pentingnya jumlah data training yang cukup untuk membangun model yang lebih andal.

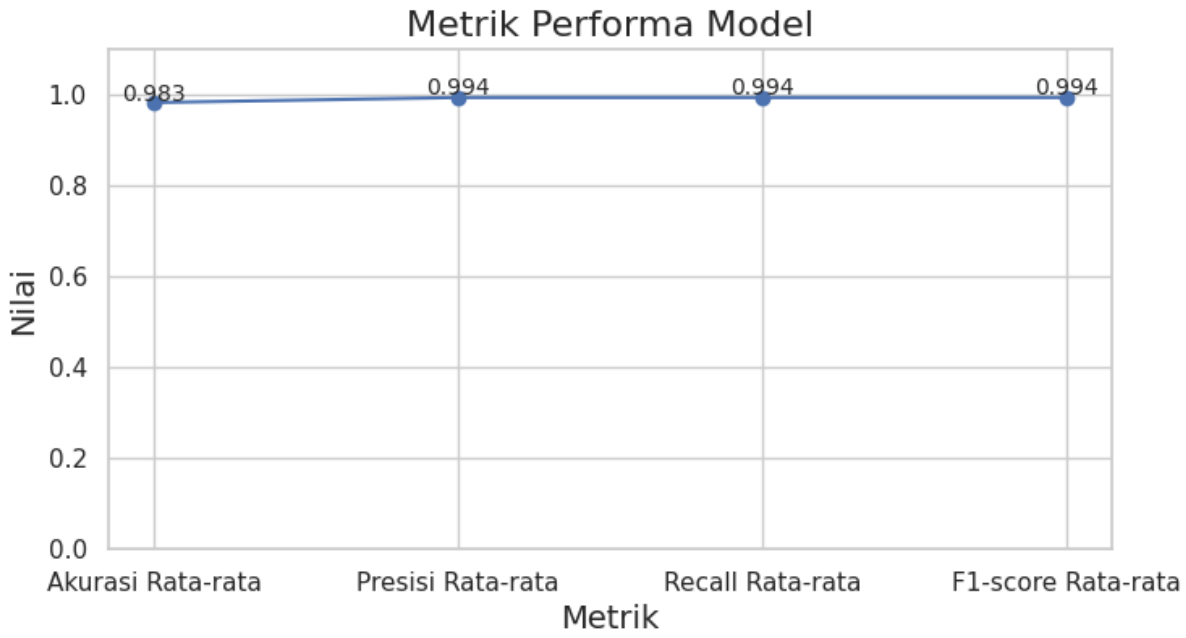
H. Hasil Akhir dan Performa Model Secara Keseluruhan

Ratio	*Accurasi*	*Precision*	*Recall*	*F1 Score*
0.1	0.94	0.98	0.97	0.98
0.2	0.98	0.99	0.99	0.99
0.3	0.99	0.99	1.00	0.99
0.4	0.99	1.00	1.00	0.99
0.5	0.99	0.99	0.99	1.00
0.6	0.99	1.00	1.00	1.00
0.7	0.99	1.00	1.00	1.00
0.8	0.99	1.00	1.00	1.00
0.9	0.99	1.00	1.00	1.00

Tabel di atas menunjukkan hasil evaluasi performa model berdasarkan rasio pembagian data dalam proses pelatihan dan pengujian. Rasio di kolom pertama menggambarkan proporsi data uji dalam proses validasi, sedangkan kolom lainnya mencantumkan metrik evaluasi kinerja model: *Akurasi, Presisi, Recall, dan F1-Score*.

Tabel hasil evaluasi menunjukkan performa model berdasarkan berbagai rasio pembagian data antara pelatihan dan pengujian. Rasio tersebut menggambarkan proporsi data uji terhadap total data, seperti pada rasio 0.1 yang berarti 10% data digunakan untuk pengujian dan 90% untuk pelatihan, serta seterusnya hingga rasio 0.9. Model menunjukkan kinerja yang sangat baik di semua rasio, dengan akurasi di atas 94%, bahkan mencapai nilai mendekati sempurna (99%) pada rasio yang lebih besar. Presisi model berada di kisaran 0.98 hingga 1.00, yang menunjukkan kemampuan model dalam membuat prediksi positif dengan sangat akurat, tanpa banyak kesalahan. Recall juga berada di angka tinggi, yaitu 0.97 hingga 1.00, mengindikasikan bahwa model dapat menangkap hampir semua data yang relevan. Selain itu, F1-Score yang konsisten mendekati 1 menunjukkan keseimbangan optimal antara presisi dan recall, terutama pada rasio 0.6 hingga 0.9, di mana semua metrik mencapai nilai sempurna. Kesimpulannya, model ini memiliki performa yang konsisten dan andal dalam memprediksi data, meskipun proporsi data uji meningkat, sehingga sangat layak diterapkan pada data yang lebih luas.

Performa Model Secara Keseluruhan



Grafik diatas menunjukkan nilai rata-rata dari berbagai metrik evaluasi model dalam analisis sentimen, yaitu *akurasi, presisi, recall, dan F1-score.*

1. *Akurasi* Rata-rata (0.9833) Akurasi menggambarkan proporsi prediksi yang benar dari seluruh prediksi yang dilakukan oleh model. Nilai rata-rata akurasi sebesar 0.9833 (atau 98,33%) menunjukkan bahwa secara keseluruhan, model mampu memprediksi dengan benar sebanyak 98,33% dari total data yang diuji. Ini menandakan performa model yang sangat baik.
2. *Presisi* Rata-rata (0.9944) Presisi menunjukkan seberapa baik model menghindari kesalahan dalam memprediksi kelas tertentu, terutama positif. Nilai rata-rata sebesar 0.9944 (atau 99,44%) menunjukkan bahwa hampir semua prediksi positif yang dihasilkan oleh model adalah benar. Model sangat akurat dalam membuat prediksi untuk setiap kelas.
3. *Recall* Rata-rata (0.9944) Recall mengukur sejauh mana model mampu menangkap semua data yang benar-benar relevan. Nilai recall sebesar 0.9944 (atau 99,44%) berarti model berhasil mengenali hampir semua data relevan dalam dataset, baik untuk prediksi kelas positif maupun negatif.
4. *F1-Score* Rata-rata (0.9944) F1-Score adalah rata-rata harmonis antara presisi dan recall, yang berguna ketika ada ketidakseimbangan antara kelas positif dan negatif. Nilai F1-Score sebesar 0.9944 menunjukkan keseimbangan yang sangat baik antara presisi dan recall, sehingga model memiliki performa yang konsisten dan andal Secara keseluruhan, performa model sangat tinggi pada setiap metrik (*precision, recall, f1-score, dan akurasi*). Ini menunjukkan bahwa model bekerja dengan sangat baik untuk klasifikasi data sentimen menggunakan model *BiLSTM.*

4. KESIMPULAN

Berdasarkan hasil evaluasi model analisis sentimen, diperoleh nilai rata-rata dari berbagai metrik, yaitu *akurasi, presisi, recall, dan F1-score*, yang menunjukkan performa model yang sangat baik. Nilai *akurasi* rata-rata sebesar 0,9833 atau 98,33% menunjukkan bahwa model mampu membuat prediksi yang benar pada sebagian besar data yang diuji. *Presisi* rata-rata sebesar 0,9944 atau 99,44% mencerminkan kemampuan model dalam menghindari kesalahan pada prediksi kelas tertentu, terutama kelas positif. Selain itu, *recall* rata-rata sebesar 0,9944 atau 99,44% menunjukkan bahwa model berhasil mengenali hampir seluruh data yang relevan dalam dataset. *F1-score* rata-rata sebesar 0,9944 menunjukkan keseimbangan yang sangat baik antara presisi dan recall, menjadikan model ini andal dan

konsisten dalam menangani data. Secara keseluruhan, performa model menunjukkan tingkat keakuratan dan keandalan yang sangat tinggi dalam menganalisis data sentimen.

DAFTAR PUSTAKA

Adi, F. A. (2022). Pengaruh Kualitas Pelayanan Terhadap Tingkat Kepuasan Konsumen dalam Membentuk Loyalitas (Studi Pada Resto Ayam Goreng Nelongso Soekarno Hatta Malang). *Jurnal Administrasi Bisnis*, *16*(1), 107–116. https://profit.ub.ac.id

Adzani, M. F., Puspalohita, N. D., Yulianti, V. R., & Kusuma, R. M. (2024). *Article History: Received: June 14*. *2*(4), 1262–1268.

Ainii, N. N. (2022). *Klasifikasi Sentimen, Topik, Dan Detail Topik Dari Ulasan Kai Access Menggunakan Multilayer Perceptron (Mlp) Dan Bidirectional Long Short Term Memory (Bilstm)*. https://dspace.uii.ac.id/handle/123456789/40914%0Ahttps://dspace.uii.ac.id/bitstream/handle/123456789/40914/18523252.pdf?sequence=1&isAllowed=y

Alghifari, D. R., Edi, M., & Firmansyah, L. (2022). Implementasi Bidirectional LSTM untuk Analisis Sentimen Terhadap Layanan Grab Indonesia. *Jurnal Manajemen Informatika (JAMIKA)*, *12*(2), 89–99. https://doi.org/10.34010/jamika.v12i2.7764

Cholissodin, I., & Soebroto, A. A. (2021). *AI , MACHINE LEARNING & DEEP LEARNING (Teori & Implementasi)*. *December*.

Coker, C., Greene, E., Shao, J., Enclave, D., Tula, R., Marg, R., Jones, L., Hameiri, S., Cansu, E. E., Initiative, R., Maritime, C., Road, S., Çelik, A., Yaman, H., Turan, S., Kara, A., Kara, F., Zhu, B., Qu, X., … Tang, S. (2018). No 主観的健康感を中心とした在宅高齢者における 健康関連指標に関する共分散構造分析Title. In *Transcommunication* (Vol. 53, Issue 1). http://www.tfd.org.tw/opencms/english/about/background.html%0Ahttp://dx.doi.org/10.1016/j.cirp.2016.06.001%0Ahttp://dx.doi.org/10.1016/j.powtec.2016.12.055%0Ahttps://doi.org/10.1016/j.ijfatigue.2019.02.006%0Ahttps://doi.org/10.1016/j.matlet.2019.04.024%0A

Gao, Y., Zhao, J., Qin, C., Yuan, Q., Zhu, J., Sun, Y., Lu, C., Federal, U., Cear, D. O., Ci, C. D. E., Agr, N., Ci, E. M., Alimentos, T. D. E., Lopes, S., Oliveira, G. O. D. E., Afifah, I., & Sopiany, H. M., Psicologia, P. D. E. P. E. M., Orrico Junior, M., Santos, H. D. S., …

Augusto, K. V. O. N. Z. (2023). No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する共分散構造分析Title. *Aleph*, *87*(1,2), 149–200. https://repositorio.ufsc.br/xmlui/bitstream/handle/123456789/167638/341506.pdf?sequence=1&isAllowed=y%0Ahttps://repositorio.ufsm.br/bitstream/handle/1/8314/LOEBLEIN%2C LUCINEIA CARLA.pdf?sequence=1&isAllowed=y%0Ahttps://antigo.mdr.gov.br/saneamento/proees

Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi. *Jurnal Teknoinfo*, *14*(2), 115. https://doi.org/10.33365/jti.v14i2.679

Hendra, A., & Fitriyani, F. (2021). Analisis Sentimen Review Halodoc Menggunakan Nai¨ve Bayes Classifier. *JISKA (Jurnal Informatika Sunan Kalijaga)*, *6*(2), 78–89. https://doi.org/10.14421/jiska.2021.6.2.78-89

Lestari, S., & Saepudin, S. (2021). Analisis Sentimen Vaksin Sinovac Pada Twitter Menggunakan Algoritma Naive Bayes. *SISMATIK (Seminar Nasional Sistem Informasi Dan Manajemen Informatika)*, 163–170.

Maulidah, I., Widodo, J., & Zulianto, M. (2019). Pengaruh Kualitas Produk Dan Kualitas Pelayanan Terhadap Kepuasan Konsumen Di Rumah Makan Ayam Goreng Nelongso Jember. *JURNAL PENDIDIKAN EKONOMI: Jurnal Ilmiah Ilmu Pendidikan, Ilmu Ekonomi Dan Ilmu Sosial*, *13*(1), 26. https://doi.org/10.19184/jpe.v13i1.10416

Nafsi, F. S. (2023). *Penerapan Digital Marketing pada E-commerce dan Media Sosial dalam Upaya Peningkatan Penjualan Produk PT Behaestex*. *1*(3), 156–166.

Ningrum, A. A., Syarif, I., Gunawan, A. I., Satriyanto, E., & Muchtar, R. (2021). Algoritma Deep Learning-LSTM untuk Memprediksi Umur Transformator. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *8*(3), 539–548. https://doi.org/10.25126/jtiik.2021834587

Nur Oktavia, A., Iqbal, M., Saputra, R. W., Zulfikar, M. I., & Saifudin, A. (2024). Implementasi Metode Natural Language Processing Dalam Studi Analisis. *Jurnal Ilmiah Ilmu Komputer Dan Multimedia*, *2*(1), 154–159. https://jurnalmahasiswa.com/index.php/biikma

Pambudi, M. S., Wiska, M., Purwanto, K., & Gusteti, Y. (2023). Analisis Pemanfaatan Google Maps Sebagai Sarana Promosi Terhadap Penjualan Usaha Mikro Kecil Menengah (UMKM) di Nagari Koto Padang. *Innovative: Journal Of Social Science Research*, *3*(5), 1562–1571. https://j-innovative.org/index.php/Innovative/article/view/5024

Prameta, M. I. (2021). Pengaruh Kualitas Layanan Dan Harga Terhadap Keputusan Pembelian Ayam Goreng Nelongso Cabang Ketintang. *Jurnal Pendidikan Tata Niaga (JPTN)*, *9*(1), 1062–1068.

Pratiwi, B. P., Handayani, A. S., & Sarjana, S. (2021). Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi Wsn Menggunakan Confusion Matrix. *Jurnal Informatika Upgris*, *6*(2), 66–75. https://doi.org/10.26877/jiu.v6i2.6552

Safitri, A. I., Sasongko, T. B., & Yogyakarta, U. A. (2024). Sentiment Analysis of Cyberbullying Using Bidirectional Long short Term Memory Algorithn on Twitter. *Jurnal Teknik Informatika (JUTIF)*, *5*(2), 615–620.

Sari, F. V., & Wibowo, A. (2019). Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi. *Jurnal SIMETRIS*, *10*(2), 681–686.

Sentoso, A., Lestari, M., & Nur Arafah, N. (2023). Strategi Pengembangan Pemasaran Digital Terhadap Umkm M_Two Capcay Solo. *Jurnal EK&BI*, *6*, 2620–7443. https://doi.org/10.37600/ekbi.v6i1.797

Sifa Amalia, B., Umaidah, Y., Mayasari, R., Karawang Jl HSRonggo Waluyo, S., Telukjambe Timur, K., & Karawang, K. (2021). Analisis Sentimen Review Pelanggan Restoran Menggunakan Algoritma Support Vector Machine Dan K-Nearest Neighbor. *SITEKIN: Jurnal Sains, Teknologi Dan Industri*, *19*(1), 28–34.

Sintia Amelia, D., Cahyana Aminuallah, N., & Informasi, S. (2023). Teks Dan Analisis Sentimen Pada Chat Grup Whatsapp Menggunakan Long Short Term Memory (Lstm). *Jurnal Teknologi Terkini*, *3*(2), 1. http://teknologiterkini.org/index.php/terkini/article/view/354

Ulhaq, M. Z. (2021). Implementasi Etika Bisnis Islam Perspektif Maqasid Syariah Pada Rumah Makan Hayaku Dan Rumah Makan Ayam Goreng Nelongso Yogyakarta. *Skripsi*, hlm.91. https://dspace.uii.ac.id/handle/123456789/31233

Ummah, M. S. (2019). No 主観的健康感を中心とした在宅高齢者における 健康関連指標に関する共分散構造分析Title. In *Sustainability (Switzerland)* (Vol. 11, Issue 1). http://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf?sequence=12&isAllowed=y%0Ahttp://dx.doi.org/10.1016/j.regsciurbeco.2008.06.005%0Ahttps://www.researchgate.net/publication/305320484_SISTEM_PEMBETUNGAN_TERPUSAT_STRATEGI_MELESTARI

Undap, M., Rantung, V. P., & Rompas, P. T. D. (2021). Analisis Sentimen Situs Pembajak Artikel Penelitian Menggunakan Metode Lexicon-Based. *Jointer - Journal of Informatics Engineering*, *2*(02), 39–46. https://doi.org/10.53682/jointer.v2i02.44

Widyanto, T., Ristiana, I., & Wibowo, A. (2023). Komparasi Naïve Bayes dan SVM Analisis Sentimen RUU Kesehatan di Twitter. *SINTECH (Science and Information Technology) Journal*, *6*(3), 147–161. https://doi.org/10.31598/sintechjournal.v6i3.1433

Yahyadi, A., & Latifah, F. (2022). Analisis Sentimen Twitter TerhadapKebijakan Ppkm Di Tengah Pandemi Covid-19Menggunakan Mode Lstm. *Journal of Information System, Applied, Management, Accounting and Research.*, *6*(2), 464–470. https://doi.org/10.52362/jisamar.v6i2.791

Zia Insani, R., Setianingsih, C., & Waluyo Purboyo, T. (2023). Analisis Sentimen Komentar Berdasarkan Geo Tagged Menggunakan Algoritma Bilstm. *E-Proceeding of Engineering*, *10*(1), 211.