

# Analisis Sentimen Masyarakat Menggunakan Algoritma Long Short Term Memory (LSTM) Pada Ulasan Aplikasi Halodoc

Nelvi Yulianti\*, M Afdal, Muhammad Jazman, Megawati Megawati, Anofrizen Anofrizen

Fakultas Sains dan Teknologi, Sistem Informasi, Universitas Islam Negeri Sultan Syarif Kasim, Riau, Indonesia

Email: <sup>1</sup>\*12150323457@students.uin-suska.ac.id, <sup>2</sup>m.afdal@uin-suska.ac.id, <sup>3</sup>jazman@uin-suska.ac.id,

<sup>4</sup>megawati@uin-suska.ac.id, <sup>5</sup>anofrizen@uin-suska.ac.id

Email Penulis Korespondensi: 12150323457@students.uin-suska.ac.id

Submitted: 30/04/2025; Accepted: 28/05/2025; Published: 02/09/2025

**Abstrak**—Halodoc merupakan platform layanan kesehatan digital yang memberikan kemudahan bagi pengguna dalam mengakses layanan medis secara online. Penelitian ini dilakukan untuk mengetahui sentimen masyarakat terhadap aplikasi Halodoc berdasarkan 1.416 ulasan pengguna yang dikumpulkan selama periode Juli hingga September 2024. Ulasan tersebut diklasifikasikan ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral, dengan menggunakan algoritma *Long Short-Term Memory* (LSTM). Sebelum dilakukan klasifikasi, teknik *Word2Vec* digunakan untuk mengubah kata-kata dalam ulasan menjadi representasi vektor numerik agar dapat diproses oleh model. Berdasarkan hasil analisis, ditemukan bahwa sebagian ulasan mengandung sentimen negatif, yang sebagian besar terkait dengan masalah keterlambatan pengiriman obat serta lambatnya respons dari layanan pelanggan. Evaluasi performa model menunjukkan bahwa penerapan algoritma LSTM yang dioptimasi menggunakan metode Adam (*Adaptive Moment Estimation*) serta regularisasi *dropout* sebesar 0.2 berhasil menghasilkan akurasi tertinggi sebesar 89,40% dan *F1-score* mencapai 88,63%. Hasil ini menunjukkan bahwa model mampu mengklasifikasikan sentimen dengan sangat baik dan dapat digunakan sebagai alat bantu dalam memahami kepuasan pengguna terhadap aplikasi Halodoc.

**Kata Kunci:** Google Play Store; Halodoc; LSTM; Analisis Sentimen; Word2Vec

**Abstract**—Halodoc is a digital healthcare platform that provides users with convenient access to medical services online. This study aims to analyze public sentiment toward the Halodoc application based on 1,416 user reviews collected during the period from July to September 2024. The reviews are categorized into three sentiment classes: positive, negative, and neutral, using the Long Short-Term Memory (LSTM) algorithm. Prior to classification, the Word2Vec technique is applied to transform the words in the reviews into numerical vector representations for processing by the model. The analysis revealed that a portion of the reviews expressed negative sentiments, mainly concerning delays in medication delivery and slow responses from customer service. Model performance evaluation shows that the implementation of the LSTM algorithm optimized with the Adam (*Adaptive Moment Estimation*) optimizer and a dropout rate of 0.2 achieved the highest accuracy of 89.40% and an F1-score of 88.63%. These results indicate that the model performs very well in classifying sentiments and can be used as a useful tool for understanding user satisfaction with the Halodoc application.

**Keywords:** Google Play Store; Halodoc; LSTM; Sentiment Analysis; Word2Vec

## 1. PENDAHULUAN

Layanan kesehatan digital berkembang pesat seiring dengan kemajuan teknologi, memungkinkan masyarakat untuk mengakses layanan medis secara daring dengan lebih mudah dan efisien. Halodoc adalah aplikasi yang menyediakan berbagai layanan kesehatan, termasuk pengiriman obat, konsultasi online melalui video call, pemeriksaan laboratorium, direktori rumah sakit dan dokter, serta fitur penjadwalan [1]. Aplikasi ini tidak hanya mencakup aspek medis tetapi juga inovasi yang bertujuan memberikan layanan kesehatan yang lebih mudah dan terjangkau bagi masyarakat [2]. Perkembangan teknologi menghadirkan beragam kegiatan dan layanan yang memudahkan pasien dalam melakukan konsultasi, perawatan dan pengobatan secara online tanpa pasien mendatangi fasilitas kesehatan. Berbagai kemudahan terus dikembangkan agar pasien dapat memperoleh pelayanan optimal, dengan memprioritaskan kesembuhan pasien Halodoc [3]. Dengan adanya aplikasi seperti Halodoc, masyarakat tidak harus datang langsung ke fasilitas kesehatan untuk mendapatkan layanan medis tertentu, sehingga lebih menghemat waktu dan tenaga. Banyak masyarakat yang memilih menggunakan Halodoc karena dapat memberikan kualitas informasi yang baik dan mudah digunakan bagi setiap kalangan [4].

Namun, seiring dengan meningkatnya jumlah pengguna, tantangan dalam mempertahankan kualitas layanan yang diberikan oleh Halodoc juga semakin besar. Pengguna memiliki ekspektasi yang tinggi terhadap berbagai aspek layanan, seperti kecepatan respons dalam menangani keluhan, ketepatan serta keandalan informasi medis yang diberikan oleh dokter maupun sistem aplikasi, serta kemudahan dalam mengakses layanan kesehatan yang tersedia. Salah satu fitur yang terdapat pada Google Play adalah fitur rating dan ulasan, di mana pengguna produk dapat memberikan opini mereka tentang produk yang telah digunakan, fitur ini berpengaruh pada calon pengguna karena mereka akan melihat kolom ulasan atau review sebagai acuan menilai kualitas suatu produk [5]. Kepuasan pengguna menjadi faktor krusial yang sangat memengaruhi citra perusahaan, terutama di tengah kompetisi yang semakin ketat dalam industri pelayanan kesehatan digital. Ulasan atau pengalaman pengguna, baik yang bersifat positif maupun negatif, dapat dengan cepat menyebar melalui berbagai platform media sosial ataupun situs ulasan aplikasi, yang pada akhirnya dapat memberikan dampak signifikan terhadap persepsi masyarakat terhadap Halodoc. Salah satu kendala utama yang sering dihadapi oleh pengguna adalah keterlambatan dalam pengiriman obat hingga sampai ke tangan pasien, di mana pasien sangat membutuhkan obat tersebut dalam waktu yang singkat dan obat yang diresepkan hanya

dapat ditebus melalui aplikasi Halodoc. Selain itu, customer service yang lambat dalam merespons layanan juga menjadi salah satu aspek yang sering dikeluhkan oleh pengguna, karena keterlambatan dalam menangani keluhan atau pertanyaan dapat berdampak pada pengalaman pengguna secara keseluruhan. Oleh karena itu, analisis sentimen menjadi sangat penting untuk memahami bagaimana masyarakat menilai layanan yang diberikan oleh aplikasi Halodoc, sehingga dapat memberikan wawasan yang lebih mendalam terkait aspek yang perlu diperbaiki guna meningkatkan kualitas layanan serta kepuasan pengguna.

Data mining adalah pengetahuan tersembunyi dalam database yang diproses untuk menemukan pola dengan menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning guna mengekstraksi serta mengidentifikasi informasi pengetahuan dari database tersebut [6]. Teks mining merupakan bentuk variasi dari data mining yang bertujuan untuk menemukan pola menarik dari kumpulan data teks berjumlah besar [7]. *Artificial Intelligence* (AI) berperan penting dalam menganalisis data besar, dengan kemampuannya mendeteksi pola dan anomali secara cepat yang mendukung pengambilan keputusan lebih baik. Salah satu penerapannya adalah *Natural Language Processing* (NLP) [8]. *Natural Language Processing* dapat digunakan untuk menganalisis umpan balik pengguna secara lebih mendalam, membantu perusahaan memahami persepsi pengguna terhadap layanan yang diberikan, yang disebut sebagai analisis sentimen yang bertujuan untuk mengidentifikasi sentimen yang terkandung dalam teks berkonotasi positif, negatif, atau netral. Teknik ini sangat berguna dalam menilai kepuasan pengguna terhadap aplikasi Halodoc melalui ulasan atau *feedback* yang diberikan [9]. Ulasan pada Google Play Store berfungsi sebagai bagian penting dalam mengevaluasi kualitas layanan aplikasi [10]. Di Google Play Store, orang sering membaca ulasan yang ditulis oleh pengguna lain dan menilai reputasi aplikasi sebelum membeli mengunduh dan menggunakan aplikasi [11].

Deep Learning merupakan salah satu bagian dari Neural Network yang memiliki arsitektur lebih rumit dengan jumlah layer yang lebih banyak. Hal ini diharapkan dapat menangani masalah yang lebih kompleks dengan menggunakan data yang lebih banyak [12]. Deep Learning juga memiliki berbagai algoritma khusus seperti Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), dan Self Organizing Maps (SOM) [13]. Dalam melakukan klasifikasi sentimen, penelitian ini menggunakan algoritma LSTM (*Long Short-Term Memory*). *Long Short-Term Memory* merupakan salah satu jenis arsitektur model *deep learning* yang banyak digunakan dalam tugas-tugas *Natural Language Processing* (NLP) seperti analisis sentimen, klasifikasi teks, dan penerjemahan bahasa [14]. Keunggulan utama LSTM adalah mampu menangani masalah *vanishing gradient*, yang sering menjadi kendala ketika memproses data dengan urutan panjang. LSTM dirancang untuk mengatasi masalah urutan data dan dapat mempertahankan informasi penting dari konteks yang panjang [15]. Dalam analisis sentimen, kemampuan LSTM untuk mengingat konteks teks yang lebih panjang sangat bermanfaat, karena sentimen dalam sebuah teks sering kali dipengaruhi oleh kalimat-kalimat sebelumnya atau sesudahnya. Hal ini membuat LSTM ideal untuk menganalisis teks dengan konteks yang kompleks dan panjang seperti ulasan pengguna aplikasi Halodoc [16]. Dengan struktur gerbang yang ada dalam LSTM (*Long Short-Term Memory*) memungkinkan pengaturan aliran informasi dengan lebih efisien, memastikan bahwa informasi penting dipertahankan selama proses analisis [17].

Adapun pada penelitian sebelumnya telah dilakukan oleh Izzatul Azizah dkk (2023) untuk analisis sentimen pada ulasan pengguna aplikasi Shopee di Google Play Store menggunakan metode Word2Vec dan LSTM, dengan hasil akurasi sebesar 73% dan F1-Score 82% [18]. Penelitian yang dilakukan oleh Hendra & Fitriyani (2021) untuk analisis sentimen review Halodoc menggunakan Naive Bayes Classifier mendapatkan akurasi sebesar 81,68% dan AUC 0,756 yang termasuk kategori *fair classification* [19]. Penelitian yang dilakukan oleh Dashtipour dkk (2021) untuk analisis sentimen review *Persian Movie* dengan menggunakan dua algoritma *deep learning* yaitu CNN dan LSTM, menunjukkan bahwa LSTM memberikan akurasi yang lebih baik yaitu 93,65% [20]. Penelitian yang dilakukan oleh Mutmatimah dkk (2024) untuk analisis sentimen aplikasi PeduliLindungi dengan menggunakan LSTM, dengan hasil akurasi sebesar 92,51% [21]. Penelitian yang dilakukan oleh Ardian Pradana dkk (2023) untuk analisis sentimen pindahan Ibu Kota Indonesia pada media sosial Twitter menggunakan LSTM dan Word2Vec, dengan hasil akurasi sebesar 93% dan F1-Score 95% [22].

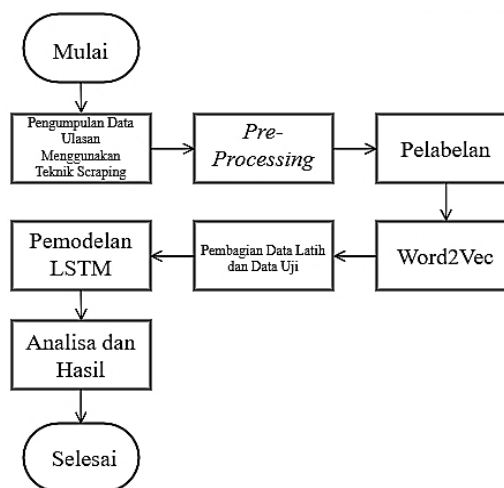
Tujuan dari penelitian ini ialah menganalisis sentimen masyarakat terhadap layanan aplikasi Halodoc, apakah positif, negatif, atau netral dengan menggunakan pendekatan *deep learning*. Untuk mencapai tujuan tersebut, penelitian ini dilakukan dengan menggunakan model *Long Short-Term Memory* (LSTM) dalam menganalisis sentimen masyarakat berdasarkan ulasan yang diberikan pada aplikasi Halodoc di Google Play Store. Melalui hasil analisis ini, diharapkan dapat memberikan pemahaman yang lebih mendalam dan gambaran yang lebih jelas terkait pandangan masyarakat terhadap layanan aplikasi Halodoc.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Untuk mencapai tujuan penelitian, dilakukan serangkaian tahapan sistematis untuk memastikan setiap langkah yang diambil menghasilkan data dan hasil yang akurat. Tahapan penelitian dimulai dari tahap pengumpulan data, tahap *pre-processing*, tahap pelabelan, tahap pemodelan LSTM (*Long Short-Term Memory*), tahap pembagian data, tahap

*Word2Vec*, hingga tahap analisa dan hasil. Adapun metodologi yang diterapkan dalam penelitian ini terdapat pada Gambar 1.



**Gambar 1.** Metodologi Penelitian

## 2.2 Tahap Pengumpulan Data

Penelitian ini berfokus pada analisis ulasan pengguna aplikasi Halodoc. Data yang digunakan diperoleh dari ulasan pengguna di Google Play Store. Keberhasilan penelitian diukur berdasarkan keterkaitan dan relevansi data terhadap permasalahan yang diteliti. Proses pengumpulan data dilakukan dengan metode web scraping menggunakan bahasa pemrograman Python melalui platform Google Play Store. Rentang waktu pengambilan data dimulai dari bulan Juli hingga September 2024. Dari proses ini, terkumpul sebanyak 1.416 ulasan pengguna yang kemudian dianalisis lebih lanjut untuk menggambarkan sentimen masyarakat terhadap layanan yang disediakan oleh aplikasi Halodoc.

## 2.3 Tahap *Pre-Processing*

Setelah data dikumpulkan maka selanjutnya masuk pada tahap *pre-processing*. Tahap ini dilakukan untuk mengubah teks yang tidak terstruktur atau acak menjadi format yang terstruktur dengan menghindari terbentuknya kumpulan data yang tidak lengkap atau tidak konsisten untuk mempermudah pemrosesan dokumen. Tahap *pre-processing* dimulai dari proses *cleaning* untuk menghapus karakter khusus seperti tanda baca, *case folding* untuk mengonversi semua karakter yang ada di dalam teks menjadi huruf kecil (lowercase), *tokenizing* untuk membagi teks menjadi unit-unit yang disebut token, *stopword* untuk menghapus kata-kata yang tidak memiliki makna atau informasi yang berkaitan dengan topik, dan *stemming* untuk mengubah setiap kata yang memiliki imbuhan menjadi kata dasar [23]. Dengan melakukan *pre-processing* yang tepat, kualitas data yang digunakan dapat ditingkatkan, sehingga menghasilkan model yang lebih akurat dan dapat diandalkan [24].

## 2.4 Tahap Pelabelan

Data hasil scraping awalnya belum memiliki label sentimen. Oleh karena itu, dilakukan proses pelabelan secara manual oleh seorang pakar untuk menjamin kualitas dan konsistensi data dalam menentukan sentimen pengguna. Setiap ulasan yang dikumpulkan dari aplikasi Halodoc akan diklasifikasikan ke dalam salah satu dari tiga kategori sentimen, yaitu positif, negatif, atau netral. Proses ini penting untuk memastikan bahwa data yang digunakan dalam analisis memiliki akurasi dan validitas yang tinggi.

## 2.5 Tahap Word2Vec

Tahap ekstraksi fitur dilakukan untuk menghasilkan atribut berupa bobot kata yang dapat digunakan dalam proses klasifikasi data. Dalam penelitian ini, metode Word2Vec diterapkan untuk mengubah setiap kata menjadi representasi vektor numerik dengan dimensi tertentu, sehingga dapat diproses oleh algoritma pembelajaran mesin [25].

## 2.6 Tahap Data Latih dan Data Uji

Data dibagi menjadi dua bagian, yaitu 80% untuk data latih dan 20% untuk data uji. Pembagian ini bertujuan agar model dapat belajar dari data latih dan kemudian diuji kemampuannya pada data uji. Selanjutnya, proses klasifikasi dilakukan menggunakan algoritma *Long Short-Term Memory* (LSTM) untuk menentukan sentimen pada ulasan.

## 2.7 Tahap Pemodelan

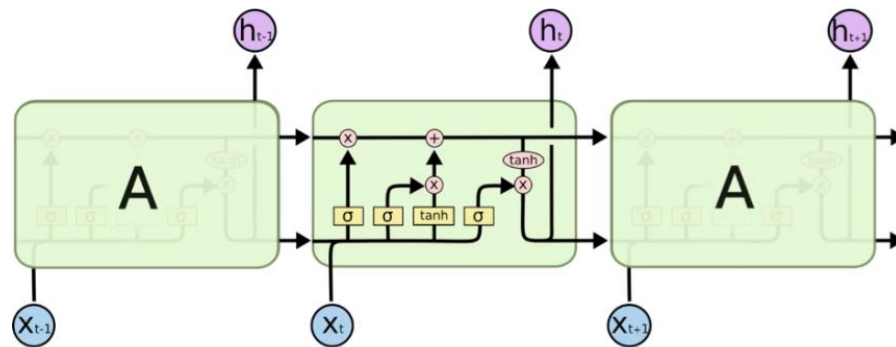
### a. Parameter

Beberapa parameter utama yang memengaruhi kinerja model LSTM dalam penelitian ini antara lain penggunaan optimizer Adam, yang berfungsi untuk mengoptimalkan model dengan menyesuaikan langkah pembelajaran guna

meminimalkan nilai loss. Learning rate yang digunakan sebesar 0.001, karena nilai ini dianggap optimal dan sering memberikan hasil baik dalam berbagai jenis pelatihan model. Ukuran batch ditentukan sebanyak 64, sementara jumlah epoch yang digunakan adalah 15. Untuk mengukur tingkat kesalahan dalam prediksi, digunakan fungsi loss *Categorical Crossentropy* sebagai alat evaluasi performa model.

b. *Long Short-Term Memory* (LSTM)

*Long Short-Term Memory (LSTM)* dirancang untuk menangani ketergantungan jangka panjang dalam data dengan memanfaatkan memori internal yang kuat. Setiap sel LSTM memiliki mekanisme untuk mengatur proses penyimpanan, penghapusan, dan pembacaan informasi melalui tiga gerbang utama: *input*, *output*, dan *forget*. Gerbang-gerbang ini menggunakan fungsi *softmax* untuk mengontrol aliran informasi, menormalkan keluaran menjadi bentuk *probabilistik*, serta membantu dalam proses *backpropagation*. Dengan demikian, LSTM dapat menyimpan dan mengolah informasi secara lebih efisien. Struktur LSTM terdiri dari modul berulang yang menggabungkan empat lapisan jaringan saraf yang saling berinteraksi dengan cara unik [26]. Struktur ini ditampilkan pada Gambar 2.



Gambar 2. Pengulangan LSTM Empat Layer

## 2.8 Tahap Evaluasi dan Hasil

*Confusion matrix* digunakan untuk evaluasi yang memungkinkan perhitungan tingkat prediksi yang benar dan salah. Tabel *confusion matrix* digunakan untuk pengklasifikasian multi-class, yang mengandung tiga kategori berbeda seperti *accuracy* untuk mengukur berapa banyak prediksi benar, *precision* untuk mengukur prediksi positif yang benar, *recall* untuk mengukur data prediksi positif dengan data positif yang sebenarnya dan *f1-score* untuk menghitung rata-rata dan melihat ketidakseimbangan antara *precision* dan *recall*. Adapun rumus persamaan tersebut ialah:

$$\text{Accuracy} = \frac{\sum_{i=1}^c TP_i}{\sum_{i=1}^c (TP_i + FP_i + FN_i)} \tag{1}$$

$$\text{Precision}_{\text{weighted}} = \sum_{i=1}^n \left( \frac{N_i}{N_i} \times \frac{TP_i}{TP_i + FP_i} \right) \tag{2}$$

$$\text{Recall}_{\text{weighted}} = \sum_{i=1}^n \left( \frac{N_i}{N_i} \times \frac{TP_i}{TP_i + FN_i} \right) \tag{3}$$

$$F1_{\text{weighted}} = \sum_{i=1}^n \left( \frac{N_i}{N_i} \times \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \right) \tag{4}$$

## 3. HASIL DAN PEMBAHASAN

### 3.1 Pengumpulan Data

Tahap pengumpulan data dilakukan untuk memperoleh informasi yang relevan dengan mempertimbangkan jenis dan sumber data sesuai kebutuhan penelitian. Dalam studi ini, data yang dikumpulkan berupa ulasan pengguna aplikasi Halodoc yang tersedia di Google Play Store. Proses pengambilan data dilakukan menggunakan teknik web scraping yang diimplementasikan dengan bahasa pemrograman *Python*. Pengumpulan data difokuskan pada periode waktu antara Juli hingga September 2024. Dari proses tersebut, diperoleh total sebanyak 1416 ulasan pengguna yang nantinya akan dianalisis untuk mengetahui sentimen masyarakat terhadap layanan yang diberikan oleh aplikasi Halodoc.

### 3.2 Pre-Processing

Setelah pengumpulan data, tahap pre-processing dilakukan untuk mempersiapkan data agar lebih mudah dianalisis dan menghasilkan model yang lebih efektif. *Pre-processing* bertujuan untuk mengubah teks yang acak menjadi format

terstruktur dilakukan dengan menghindari terbentuknya kumpulan data yang tidak lengkap atau tidak konsisten untuk mempermudah pemrosesan dokumen [23]. Beberapa proses pre-processing ialah:

**3.2.1 Cleaning**

Pada proses ini dilakukan penghapusan karakter khusus seperti tanda baca, angka, serta simbol-simbol lain (seperti seperti \*, #, %, dan sebagainya). Hasil *cleaning* dapat dilihat pada Tabel 1.

**Tabel 1.** Hasil *Pre-Processing* Data

Data Awal	Cleaning
Gk efektif. Diberikan resep obat, cuma bisa tebus pakai Aplikasi. Terus pilihan pengiriman cuman ada reguler, 3-4 hari. Lha gimana orang butuh obatnya sekarang malah mesti nunggu berhari2. Padahal apotiknya di dalam kota aja...tp pengirimannya lambannnn... harap diperbaiki lagi	Gk efektif Diberikan resep obat cuma bisa tebus pakai Aplikasi Terus pilihan pengiriman cuman ada reguler hari Lha gimana orang butuh obatnya sekarang malah mesti nunggu berhari Padahal apotiknya di dalam kota ajatp pengirimannya lambannnn harap diperbaiki lagi

**3.2.2 Case Folding**

Pada proses ini semua karakter dikonversi kedalam teks menjadi huruf kecil (*lowercase*) untuk mempermudah pencarian karena tidak semua dokumen teks menggunakan huruf kapital secara konsisten. Hasil case folding dapat dilihat pada Tabel 2.

**Tabel 2.** Hasil Case Folding Data

Data Awal	Case folding
Gk efektif. Diberikan resep obat, cuma bisa tebus pakai Aplikasi. Terus pilihan pengiriman cuman ada reguler, 3-4 hari. Lha gimana orang butuh obatnya sekarang malah mesti nunggu berhari2. Padahal apotiknya di dalam kota aja...tp pengirimannya lambannnn... harap diperbaiki lagi	gk efektif diberikan resep obat cuma bisa tebus pakai aplikasi terus pilihan pengiriman cuman ada reguler hari lha gimana orang butuh obatnya sekarang malah mesti nunggu berhari padahal apotiknya di dalam kota ajatp pengirimannya lambannnn harap diperbaiki lagi

**3.2.3 Tokenizing**

Proses ini dilakukan dengan membagi teks menjadi unit-unit yang disebut token agar dapat dianalisis lebih lanjut dan memisahkan kata-kata dengan mengacu pada setiap spasi sebagai pemisah. Hasil tokenizing dapat dilihat pada Tabel 3.

**Tabel 3.** Hasil Tokenizing Data

Data Awal	Tokenizing
Gk efektif. Diberikan resep obat, cuma bisa tebus pakai Aplikasi. Terus pilihan pengiriman cuman ada reguler, 3-4 hari. Lha gimana orang butuh obatnya sekarang malah mesti nunggu berhari2. Padahal apotiknya di dalam kota aja...tp pengirimannya lambannnn... harap diperbaiki lagi	['gk', 'efektif', 'diberikan', 'resep', 'obat', 'cuma', 'bisa', 'tebus', 'pakai', 'aplikasi', 'terus', 'pilihan', 'pengiriman', 'cuman', 'ada', 'reguler', 'hari', 'lha', 'gimana', 'orang', 'butuh', 'obatnya', 'sekarang', 'malah', 'mesti', 'nunggu', 'berhari', 'padahal', 'apotiknya', 'di', 'dalam', 'kota', 'ajatp', 'pengirimannya', 'lambannnn', 'harap', 'diperbaiki', 'lagi']

**3.2.4 Stopword**

Proses ini dilakukan dengan menghapus kata-kata yang tidak memiliki makna atau informasi penting yang berkaitan dengan topik atau sentimen. Hasil stopword dapat dilihat pada Tabel 4.

**Tabel 4.** Hasil Stopword Data

Data Awal	Stopword
Gk efektif. Diberikan resep obat, cuma bisa tebus pakai Aplikasi. Terus pilihan pengiriman cuman ada reguler, 3-4 hari. Lha gimana orang butuh obatnya sekarang malah mesti nunggu berhari2. Padahal apotiknya di dalam kota aja...tp pengirimannya lambannnn... harap diperbaiki lagi	['gk', 'efektif', 'resep', 'obat', 'tebus', 'pakai', 'aplikasi', 'pilihan', 'pengiriman', 'cuman', 'reguler', 'orang', 'butuh', 'obatnya', 'mesti', 'nunggu', 'berhari', 'apotiknya', 'kota', 'ajatp', 'pengirimannya', 'lambannnn', 'harap', 'diperbaiki']

**3.2.5 Stemming**

Proses ini mengubah setiap kata yang memiliki imbuhan diawal atau diakhir kata menjadi kata dasar. Hasil stemming dapat dilihat pada Tabel 5.

Tabel 5. Hasil Stemming Data

Data Awal	Stemming
Gk efektif. Diberikan resep obat, cuma bisa tebus pakai Aplikasi. Terus pilihan pengiriman cuman ada reguler, 3-4 hari. Lha gimana orang butuh obatnya sekarang malah mesti nunggu berhari2. Padahal apotiknya di dalam kota aja...tp pengirimannya lambannnn... harap diperbaiki lagi	['gk', 'efektif', 'resep', 'obat', 'tebus', 'pakai', 'aplikasi', 'pilih', 'kirim', 'cuman', 'reguler', 'orang', 'butuh', 'obat', 'mesti', 'nunggu', 'hari', 'apotik', 'kota', 'ajalp', 'kirim', 'lambannnn', 'harap', 'baik']

### 3.3 Pelabelan

Berdasarkan hasil pelabelan manual yang dilakukan oleh pakar terhadap total 1.416 data ulasan pengguna aplikasi Halodoc selama periode Juli hingga September, diperoleh sebanyak 1.190 ulasan yang bersentimen positif, 151 ulasan bersentimen negatif, dan 74 ulasan netral, yang mencerminkan persepsi beragam dari pengguna. Hasil pelabelan data dapat dilihat pada Tabel 6 berikut.

Tabel 6. Hasil Pelabelan Data

No	Komentar	Sentimen
1	['buruk', 'mana', 'orang', 'pesan', 'karna', 'butuh', 'cepat', 'aplikasi', 'lama', 'hari', 'tentu', 'obat', 'pesan', 'chat', 'denga', 'cs', 'guna', 'slow', 'respon', 'status', 'bayar', 'hasil', 'ubah', 'gagal', 'komplain', 'aplikasi', 'nya', 'halodoc', 'bad']	Negatif
2	['gk', 'efektif', 'resep', 'obat', 'tebus', 'pakai', 'aplikasi', 'pilih', 'kirim', 'cuman', 'reguler', 'orang', 'butuh', 'obat', 'mesti', 'nunggu', 'hari', 'apotik', 'kota', 'ajalp', 'kirim', 'lambannnn', 'harap', 'baik']	Negatif
3	['apk', 'alami', 'error', 'mana', 'konfirmasi', 'bayar', 'ulang', 'kali', 'otp', 'gagaltidak', 'masuk', 'apk', 'halodoc', 'laku', 'batal', 'bayar', 'sdh', 'masuk', 'refund', 'bayar', 'apk', 'halodoc']	Negatif
4	['rekomendasi', 'cs', 'nya', 'beli', 'obat', 'konfirmasi', 'jam', 'apotik', 'sampe', 'tutup', 'kunjung', 'kirim', 'jam', 'kurir', 'sampe', 'ganti', 'berulangkali', 'sayang', 'ga', 'menu', 'batal']	Neutral
5	['minggu', 'updatesudah', 'update', 'lagiheran', 'gk', 'aplikasi', 'nyasih', 'kirain', 'update', 'fitur', 'terbaruternyata', 'tidakjadi', 'update', 'tolong', 'perhati', 'update']	Negatif
...	...	...
1416	['respon', 'cepat']	Positive

### 3.4 Word2Vec

Proses ekstraksi fitur dilakukan dengan tujuan menghasilkan serangkaian atribut yang memiliki bobot kata tertentu dan dapat dihitung untuk membantu dalam proses klasifikasi data. Metode *Word2Vec* digunakan untuk memberikan bobot pada setiap istilah dalam dokumen dan mengonversinya menjadi representasi vektor numerik berdimensi tertentu yang dapat dipahami oleh model pembelajaran mesin [25]. Hasil yang diperoleh setelah penerapan proses ini terdapat pada Tabel 7.

Tabel 7. Pembobotan Word2Vec

Word	0	1	2	3	...	99
dokter	-0.015889738	0.11501964	0.117209785	0.22503224	...	-0.08211101
halodoc	-0.028197644	0.14419165	0.14600597	0.26763588	...	-0.09446312
obat	-0.027085884	0.13853703	0.14217883	0.27262986	...	-0.08619978
aplikasi	-0.024621999	0.107893646	0.11382428	0.20064656	...	-0.07559562
...	...	...	...	...	...	...
konsultasi	-0.008757221	0.115763165	0.12905519	0.24115056	...	-0.07212756

### 3.5 Pembagian Data Latih dan Data Uji

Pembagian data dalam penelitian ini dilakukan menjadi dua bagian, yaitu 80% sebagai data latih dan 20% sebagai data uji. Proses pembagian dilakukan menggunakan bahasa pemrograman Python dengan memanfaatkan fungsi `train_test_split` dari pustaka `sklearn.model_selection`. Dari proses tersebut, diperoleh sebanyak 1.133 data latih yang digunakan untuk melatih model dan 283 data uji yang digunakan untuk mengevaluasi performa model dalam melakukan klasifikasi sentimen terhadap ulasan pengguna aplikasi Halodoc. Pembagian data ini bertujuan untuk memastikan bahwa model dapat belajar dari sebagian besar data, sekaligus diuji kemampuannya dalam memprediksi data baru secara akurat dan objektif.

### 3.6 Pemodelan

Pemodelan dilakukan dengan menggunakan algoritma *Long Short-Term Memory* (LSTM) untuk mengklasifikasikan sentimen positif, negatif dan netral untuk melihat tingkat akurasi model terhadap sentimen ulasan aplikasi Halodoc. Pada layer pertama model menggunakan variasi jumlah unit yang berbeda agar dapat menghasilkan model yang baik, jumlah unit yang diuji ialah 50, 100, dan 150. Pada layer kedua di gunakan dropout untuk menghindari overfitting,

nilai dropout yang digunakan adalah 0.2 yang artinya 20% dari jumlah unit secara acak akan di nonaktifkan agar model tidak terlalu kompleks mempelajari data. Pada layer terakhir yaitu dense layer berisi unit sebesar 3 dan fungsi aktivasi yang digunakan adalah *softmax*. Nilai *learning rate* yang digunakan adalah 0.001 untuk optimasi dengan Adam (*Adaptive Moment Estimation*). Untuk melakukan pelatihan model dijalankan dengan menggunakan batch 64 dengan jumlah epoch 15. Hasil akurasi pelatihan model dapat dilihat pada Tabel 4.

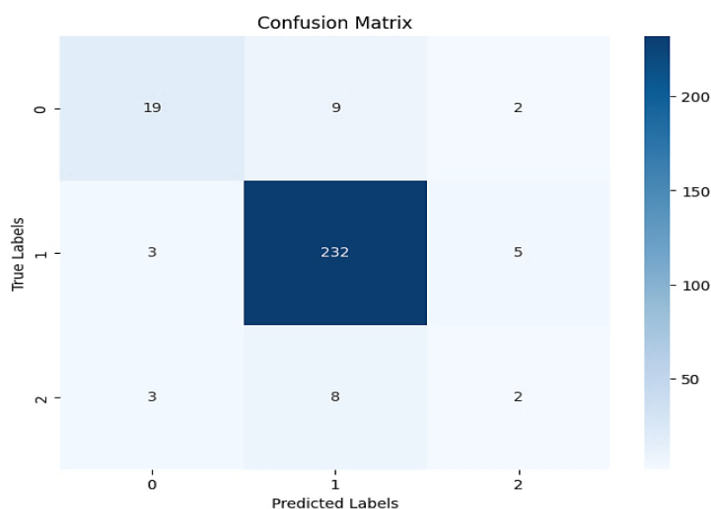
Tabel 8. Hasil Akurasi dan Loss

Data	Batch Size	Jumlah Unit	Akurasi	Loss
Ulasan Aplikasi Halodoc	64	50	0.8940	0.3483
		100	0.8763	0.4239
		150	0.8763	0.4333

Model LSTM yang diuji dengan berbagai konfigurasi unit dan menggunakan batch size 64 menunjukkan hasil yang berbeda-beda. Pengujian dengan 50 unit menghasilkan akurasi tertinggi yaitu 89,4% dengan loss 0,3483. Selanjutnya, pengujian dengan 100 unit memperoleh akurasi 87,6% dengan loss 0,4239, sementara pengujian dengan 150 unit menghasilkan akurasi yang sama, yaitu 87,6%, namun dengan loss yang sedikit lebih tinggi, yaitu 0,4333. Dari hasil-hasil tersebut, model LSTM dengan 50 unit menunjukkan performa paling optimal, karena memberikan akurasi tertinggi dengan loss yang lebih rendah. Hal ini menunjukkan bahwa model dengan 50 unit dapat memberikan prediksi yang lebih akurat dengan kompleksitas yang lebih rendah dibandingkan dengan konfigurasi unit lainnya. Model ini lebih efisien dan memberikan hasil yang lebih baik dibandingkan pilihan unit lainnya.

### 3.7 Evaluasi dan Hasil

Pengujian model dilakukan dengan menggunakan data uji, kemudian performa model dievaluasi melalui *confusion matrix*. *Confusion matrix* merupakan alat yang umum digunakan dalam data mining untuk menghitung akurasi serta mengevaluasi performa klasifikasi secara menyeluruh [27]. Hasil *confusion matrix* dapat dilihat pada Gambar 3.



Gambar 3. Hasil *Confusion Matrix*

*Confusion matrix* pada Gambar 3 menunjukkan hasil prediksi model terhadap sejumlah data ulasan terbagi menjadi tiga kategori: negatif, positif dan netral. Ulasan negatif dari total 30 data aktual, model berhasil memprediksi dengan benar 19 ulasan sebagai negatif (*True Negative*). Ulasan positif dari 240 data aktual, model berhasil mengklasifikasikan dengan benar 232 ulasan sebagai positif (*True Positive*). Ulasan netral dari total 13 data aktual, hanya 2 ulasan yang diprediksi dengan benar sebagai netral (*True Netral*). Hasil ini menunjukkan bahwa model memiliki performa yang baik dalam mengidentifikasi ulasan positif, namun masih mengalami kesulitan dalam membedakan ulasan netral dari kategori lainnya. Namun, jumlah prediksi yang benar jauh lebih tinggi pada kategori positif dan negatif yang mengindikasikan bahwa model memiliki kinerja yang cukup baik secara umum.

Berikutnya dilakukan perhitungan untuk memperoleh hasil menggunakan *weighted average* berdasarkan matriks dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score* sebagai berikut.

$$\begin{aligned}
 \text{a. Accuracy} &= \frac{19 + 232 + 2}{(19+6+11)+(232+17+8)+(2+7+11)} \\
 &= \frac{253}{36+257+20} = \frac{253}{313} = 89.40\%
 \end{aligned}$$

$$\begin{aligned}
 \text{b. Precision}_{\text{weighted}} &= \left( \frac{30}{283} \times 0.76 \right) + \left( \frac{240}{283} \times 0.9317 \right) + \left( \frac{13}{283} \times 0.2222 \right) \\
 &= 0.0804 + 0.7901 + 0.0104 = 88.09\%
 \end{aligned}$$

$$c. \text{ Recall}_{weighted} = \left(\frac{30}{283} \times 0.6333\right) + \left(\frac{240}{283} \times 0.9667\right) + \left(\frac{13}{283} \times 0.1538\right)$$

$$= 0.0671 + 0.8197 + 0.0072 = 89.40\%$$

$$d. \text{ F1}_{weighted} = \left(\frac{30}{283} \times 0.69\right) + \left(\frac{30}{283} \times 0.9490\right) + \left(\frac{13}{283} \times 0.1818\right)$$

$$= 0.0732 + 0.8047 + 0.0084 = 88.63\%$$

Setelah dilakukan perhitungan menggunakan persamaan weighted average untuk multi-class didapatkan hasil *accuracy*, *precision*, *recall*, dan *f1-score*. Dapat dilihat pada Tabel 9.

**Tabel 9.** Hasil Pengujian

<i>Accuracy</i>	89.40%
<i>Precision</i>	88.09%
<i>Recall</i>	89.40%
<i>F1 Score</i>	88.63%

Hasil evaluasi model menunjukkan akurasi keseluruhan sebesar 89,40%, yang berarti 89,40% dari semua prediksi yang dilakukan oleh model sesuai dengan label yang benar. *Precision* sebesar 88,09% menunjukkan bahwa sebagian besar prediksi yang dikategorikan sebagai positif memang benar-benar positif, menandakan bahwa model mampu meminimalkan kesalahan tipe false positive dengan baik. *Recall* sebesar 89,40% mengindikasikan bahwa model mampu mengidentifikasi hampir semua data yang seharusnya positif, menandakan deteksi yang sangat baik terhadap kasus positif. *F1-Score* sebesar 88,63% mencerminkan keseimbangan yang optimal antara *precision* dan *recall*, mengindikasikan performa yang sangat baik dalam klasifikasi sentimen secara keseluruhan. Secara keseluruhan, model menunjukkan kemampuan yang sangat baik dalam mengklasifikasikan sentimen dengan akurasi dan keseimbangan yang tinggi antara ketepatan dan kemampuan deteksi.

#### 4. KESIMPULAN

Hasil evaluasi model analisis klasifikasi sentimen yang dilakukan menunjukkan bahwa analisis sentimen pada ulasan pengguna aplikasi Halodoc dapat memberikan wawasan yang sangat berharga mengenai persepsi masyarakat terhadap layanan yang disediakan. Dengan menerapkan algoritma *Long Short-Term Memory* (LSTM), penelitian ini berhasil mengklasifikasikan sentimen ulasan menjadi tiga kategori, yaitu positif, negatif, dan netral, dengan tingkat akurasi yang sangat optimal. Hasil evaluasi menunjukkan bahwa model LSTM dengan konfigurasi 50 unit memberikan kinerja terbaik, dengan akurasi sebesar 89,40% dan loss 0,3483. Keunggulan LSTM dalam menangani ketergantungan jangka panjang pada data teks berkontribusi signifikan dalam memahami konteks ulasan pengguna secara lebih mendalam. Proses *pre-processing* yang dilakukan, termasuk *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming*, juga turut berperan dalam meningkatkan kualitas data sebelum dilakukan pemodelan. Selain itu, penggunaan teknik *Word2Vec* dalam ekstraksi fitur berhasil mengubah teks menjadi representasi numerik yang lebih bermakna dan mudah dipahami oleh model. Hasil klasifikasi sentimen menunjukkan bahwa mayoritas ulasan pengguna bersifat positif, meskipun terdapat keluhan terkait keterlambatan pengiriman obat dan respons customer service yang masih menjadi tantangan utama. Penelitian ini juga membuktikan bahwa pemanfaatan kecerdasan buatan dalam analisis teks dapat memberikan solusi yang lebih efektif dalam menilai kepuasan pelanggan. Di masa depan, diharapkan penelitian ini dapat menjadi referensi yang berguna untuk pengembangan layanan telemedicine yang lebih baik dan responsif. Penelitian ini memiliki keterbatasan pada cakupan data yang digunakan hanya dari satu platform serta kemungkinan subjektivitas dalam pelabelan manual oleh ahli yang dapat mempengaruhi hasil model yang digunakan. Peneliti selanjutnya disarankan untuk menggunakan data dari berbagai sumber dan mengeksplorasi algoritma lain seperti BERT untuk memperoleh hasil yang lebih mendalam.

#### REFERENCES

- [1] U. I. Wakhida and S. Sanaji, "Peran Perceived Usefulness dan Perceived Risk sebagai Variabel Pemeditasi pada Pengaruh Perceived Ease of Use dan e-WOM Negatif terhadap Niat Pembelian Para Pengguna Aplikasi Layanan Kesehatan Halodoc," *J. Ilmu Manaj.*, vol. 8, no. 4, p. 1158, 2020, doi: 10.26740/jim.v8n4.p1158-1174.
- [2] Y. E. Kristianto, "Strategy of Technology Acceptance Model Utilization for Halodoc, a Telehealth Mobile Application with Task Technology Fit as Moderator Variable," *Int. J. Innov. Sci. Res. Technol.*, vol. 6, no. 8, 2021, [Online]. Available: [www.ijisrt.com](http://www.ijisrt.com)192
- [3] H. Herwando and T. H. Sitompul, "Evaluasi Manfaat Penerapan Telemedicine di Negara Kepulauan: Systematic Literature Review," *Indones. Heal. Inf. Manag. J.*, vol. 9, no. 2, pp. 91–101, 2021, doi: 10.47007/inohim.v9i2.261.
- [4] E. Setiawan and J. S. Suroso, "Analisis Faktor-Faktor yang Mempengaruhi Penggunaan dan Kepuasan Pengguna Aplikasi Halodoc," *J. Pendidik. dan Konseling Vol.*, vol. 4, no. 5, pp. 4850–4862, 2022, [Online]. Available: <https://journal.universitaspahlawan.ac.id/index.php/jpdk/article/view/7397>
- [5] P. Aditiya, U. Enri, and I. Maulana, "Analisis Sentimen Ulasan Pengguna Aplikasi Myim3 Pada Situs Google Play Menggunakan Support Vector Machine," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 4, p. 1020, 2022, doi:



- 10.30865/jurikom.v9i4.4673.
- [6] D. P. Utomo and M. Mesran, “Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung,” *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 437, 2020, doi: 10.30865/mib.v4i2.2080.
  - [7] D. A. C. Rachman, R. Goejantoro, and F. D. T. Amijaya, “Implementasi Text Mining Pengelompokan Dokumen Skripsi Menggunakan Metode K-Means Clustering,” *Eksponensial*, vol. 11, no. 2, p. 167, 2021, doi: 10.30872/eksponensial.v11i2.660.
  - [8] T. P. Nugrahanti, N. Puspitasari, and I. R. Andaningsih, “Transformasi Praktik Akuntansi Melalui Teknologi: Peran Kecerdasan Buatan, Analisis Data, dan Blockchain dalam Otomatisasi Proses Akuntansi,” *J. Akunt. Dan Keuang. West Sci.*, vol. 2, no. 03, pp. 213–221, 2023, doi: 10.58812/jakws.v2i03.644.
  - [9] Ernianti Hasibuan and Elmo Allistair Heriyanto, “Analisis Sentimen Pada Ulasan Aplikasi Amazon Shopping Di Google Play Store Menggunakan Naive Bayes Classifier,” *J. Tek. dan Sci.*, vol. 1, no. 3, pp. 13–24, 2022, doi: 10.56127/jts.v1i3.434.
  - [10] D. Diandra Audiansyah, D. Eka Ratnawati, and B. Trias Hanggara, “Analisis Sentimen Aplikasi MyXL menggunakan Metode Support Vector Machine berdasarkan Ulasan Pengguna di Google Play Store,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 8, pp. 3987–3994, 2022, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/11484>
  - [11] M. M. Maarif and N. Setiyawati, “Analisis Sentimen Review Aplikasi LinkedIn di Google Play Store Menggunakan Support Vector Machine,” *Progresif J. Ilm. Komput.*, vol. 20, no. 1, p. 454, 2024, doi: 10.35889/progresif.v20i1.1614.
  - [12] N. Giarsyani, “Komparasi Algoritma Machine Learning dan Deep Learning untuk Named Entity Recognition : Studi Kasus Data Kebencanaan,” *Indones. J. Appl. Informatics*, vol. 4, no. 2, p. 138, 2020, doi: 10.20961/ijai.v4i2.41317.
  - [13] M. R. S. Alfarizi, M. Z. Al-farish, M. Taufiqurrahman, G. Ardiansah, and M. Elgar, “Penggunaan Python Sebagai Bahasa Pemrograman untuk Machine Learning dan Deep Learning,” *Karya Ilm. Mhs. Bertauhid (KARIMAH TAUHID)*, vol. 2, no. 1, pp. 1–6, 2023.
  - [14] I. P. Ramayasa, I. G. Ayu, D. Saryanti, and I. K. Dharmendra, “Perbandingan Metode Vektorisasi Pada Analisa Sentiment, Studi Kasus : Cyberbullying Pada Komentar Instagram,” *J. Teknol. Inf. dan Komput.*, vol. 9, pp. 505–512, 2023.
  - [15] I. A. Pradana, A. D. Rahajoe, and A. N. Sihananto, “Pengembangan Aplikasi Pendeteksi Keretakan Jalan Berbasis Android Dengan Implementasi Algoritma Hybrid CNN-LSTM,” vol. 5, no. 2, pp. 1–10, 2024.
  - [16] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
  - [17] Z. W. Farida and N. Rochmawati, “Analisis Sentimen Masyarakat terhadap Fenomena Childfree Menggunakan Metode Long Short Term Memory dan Bidirectional Encoder Representations from Transformers di Twitter,” *J. Informatics Comput. Sci.*, vol. 5, no. 03, pp. 369–376, 2024, [Online]. Available: <https://ejournal.unesa.ac.id/index.php/jinacs/article/view/58142%0Ahttps://ejournal.unesa.ac.id>
  - [18] I. Azizah, I. Cholissodin, and N. Yudistira, “Analisis Sentimen Ulasan Pengguna Aplikasi Shopee di Google Play menggunakan Metode Word Embedding dan Long Short Term Memory (LSTM),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 5, pp. 2453–2459, 2023.
  - [19] A. Hendra and F. Fitriyani, “Analisis Sentimen Review Halodoc Menggunakan Naive Bayes Classifier,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 6, no. 2, pp. 78–89, 2021, doi: 10.14421/jiska.2021.6.2.78-89.
  - [20] K. Dashtipour, M. Gogate, A. Adeel, H. Larijani, and A. Hussain, “Sentiment analysis of persian movie reviews using deep learning,” *Entropy*, vol. 23, no. 5, pp. 1–16, 2021, doi: 10.3390/e23050596.
  - [21] S. Mutmatimah, Khairunnas, and Khairunnisa, “Metode Deep Learning LSTM dalam Analisis Sentimen Aplikasi PeduliLindungi,” *J. Comput. Sci. Informatics*, vol. 1, no. 1, pp. 9–19, 2024, doi: 10.34304/scientific.v1i1.231.
  - [22] Y. Ardian Pradana, I. Cholissodin, and D. Kurnianingtyas, “Analisis Sentimen Pemindahan Ibu Kota Indonesia pada Media Sosial Twitter menggunakan Metode LSTM dan Word2Vec,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 5, pp. 2389–2397, 2023, [Online]. Available: <http://j-ptiik.ub.ac.id>
  - [23] T. Astuti and Y. Astuti, “Analisis Sentimen Review Produk Skincare Dengan Naive Bayes Classifier Berbasis Particle Swarm Optimization (PSO),” *J. Media Inform. Budidarma*, vol. 6, no. 4, p. 1806, 2022, doi: 10.30865/mib.v6i4.4119.
  - [24] S. F. Pane and J. Ramdan, “Pemodelan Machine Learning : Analisis Sentimen Masyarakat Terhadap Kebijakan PPKM Menggunakan Data Twitter,” *J. Sist. Cerdas*, vol. 5, no. 1, pp. 12–20, 2022, doi: 10.37396/jsc.v5i1.191.
  - [25] W. Andriyani, Y. Astuti, and B. A. Wisesa, “Analisis Sentimen pada Ulasan Produk dengan SVM dan Word2Vec Sentiment Analysis on Product Reviews with SVM and Word2Vec,” *JIKO (Jurnal Inform. dan Komputer)*, vol. 9, no. 1, pp. 173–185, 2024, doi: 10.26798/jiko.v8i1.1498.
  - [26] Y. Romadhoni and K. F. H. Holle, “Analisis Sentimen Terhadap PERMENDIKBUD No.30 pada Media Sosial Twitter Menggunakan Metode Naive Bayes dan LSTM,” *J. Inform. J. Pengemb. IT*, vol. 7, no. 2, pp. 118–124, 2022, doi: 10.30591/jpit.v7i2.3191.
  - [27] M. F. Rahman, D. Alamsah, M. I. Darmawidjadja, and I. Nurma, “Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode Bayesian Regularization Neural Network (RBNN),” *J. Inform.*, vol. 11, no. 1, p. 36, 2017, doi: 10.26555/jifo.v11i1.a5452.