ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN PLATFORM RUANGGURU MENGGUNAKAN METODE *LONG SHORT-TERM MEMORY* (LSTM)

TESIS

DELTA APRIZA

ENTERPRISE IT INFRASTRUCTURE

232420017

PROGRAM STUDI TEKNIK INFORMATIKA S-2

PROGRAM PASCA SARJANA

UNIVERSITAS BINA DARMA

TAHUN 2026

ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN PLATFORM RUANGGURU MENGGUNAKAN METODE *LONG SHORT-TERM MEMORY* (LSTM)

Tesis ini diajukan sebagai salah satu
syarat untuk memperoleh gelar

MAGISTER KOMPUTER

DELTA APRIZA
ENTERPRISE IT INFRASTRUCTURE
232420017

PROGRAM STUDI TEKNIK INFORMATIKA S-2
PROGRAM PASCA SARJANA
UNIVERSITAS BINA DARMA
TAHUN 2026

HALAMAN PENGESAHAN PEMBIMBING TESIS

Judul Tesis: ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN PLATFORM RUANGGURU MENGGUNAKAN METODE LONG SHORT-TERM MEMORY (LSTM)

Oleh DELTA APRIZA , NIM 232420017, Tesis ini telah disetujui dan disahkan oleh Tim Penguji Program Studi Teknik Informatika – S2 konsentrasi ENTERPRISE IT INFRASTRUCTURE, Program Pascasarjana Universitas Bina Darma pada 04 Maret 2026 dan telah dinyatakan LULUS.

Palembang, 04 Maret 2026

Mengetahui,
Program Studi Teknik Informatika – S2
Universitas Bina Darma
Ketua,

Universitas Bina Darma
Magister Teknik Informatika

.....
Dr. Usman Ependi, S.Kom., M.Kom.

Pembimbing,

.....
Dr. Usman Ependi, S.Kom., M.Kom.

HALAMAN PENGESAHAN PENGUJI TESIS

Judul Tesis: ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN PLATFORM RUANGGURU MENGGUNAKAN METODE LONG SHORT-TERM MEMORY (LSTM)

Oleh DELTA APRIZA , NIM 232420017, Tesis ini telah disetujui dan disahkan oleh Tim Penguji Program Studi Teknik Informatika – S2 konsentrasi ENTERPRISE IT INFRASTRUCTURE, Program Pascasarjana Universitas Bina Darma pada 04 Maret 2026 dan telah dinyatakan LULUS.

Palembang, 04 Maret 2026
Mengetahui,
Program Pascasarjana
Universitas Bina Darma
Direktur,

PROGRAM PASCASARJANA

Prof. Dr. Ir. Achmad Syarifudin, M.Sc.

Penguji I,

Dr. Usman Ependi, S.Kom., M.Kom.

Penguji II,

Dr. Yesi Novaria Kunang, S.T., M.Kom.

Penguji III,

Nita Rosa Damayanti, M.Kom., Ph.D.

SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini:

Nama : DELTA APRIZA

NIM : 232420017

Dengan ini menyatakan bahwa:

1. Karya tulis Saya (Tesis, Skripsi, Tugas Akhir) ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik baik (Magister, Sarjana, dan Ahli Madya) di Universitas Bina Darma;
2. Karya tulis ini murni gagasan, rumusan dan penelitian Saya sendiri dengan arahan tim pembimbing;
3. Dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dikutip dengan mencantumkan nama pengarang dan memasukkan ke dalam daftar pustaka;
4. Karena yakin dengan keasliaan karya tulis ini, Saya menyatakan bersedia Tesis/Skripsi/Tugas Akhir, yang Saya hasilkan di unggah ke internet;
5. Surat Pernyataan ini Saya tulis dengan sungguh-sungguh dan apabila terdapat penyimpangan atau ketidakbenaran dalam pernyataan ini, maka Saya bersedia menerima sanksi dengan aturan yang berlaku di perguruan tinggi ini.

Demikian Surat Pernyataan ini saya buat agar dapat dipergunakan sebagaimana mestinya.

Palembang, 04 Maret 2026
Yang Membuat Pernyataan,

DELTA APRIZA
NIM: 232420017

ABSTRAK

Platform pembelajaran daring Ruangguru menerima berbagai ulasan dari pengguna yang memuat opini terhadap beragam aspek layanan. Penelitian ini bertujuan untuk menerapkan analisis sentimen berbasis aspek pada ulasan platform Ruangguru menggunakan metode *Long Short-Term Memory* (LSTM). Tahapan penelitian meliputi pengumpulan data ulasan, pra-pemrosesan teks, ekstraksi aspek menggunakan *Latent Dirichlet Allocation* (LDA) dengan penentuan jumlah topik optimal berdasarkan nilai *coherence score*, serta pelabelan sentimen otomatis menggunakan model mBERT dan IndoBERT. Model LSTM dilatih menggunakan dua skenario, yaitu penggabungan seluruh aspek dalam satu model dan pelatihan terpisah untuk setiap aspek. Evaluasi kinerja model dilakukan menggunakan *confusion matrix* untuk menganalisis kemampuan klasifikasi sentimen positif, netral, dan negatif, serta menghitung nilai akurasi, *precision*, *recall*, dan *F1-score*. Hasil menunjukkan bahwa pelatihan gabungan seluruh aspek menghasilkan performa yang lebih stabil dan konsisten karena didukung jumlah data yang lebih besar dan distribusi sentimen yang lebih beragam. Sebaliknya, pelatihan terpisah per aspek cenderung bias terhadap kelas dominan akibat ketidakseimbangan distribusi data, sehingga menurunkan kemampuan generalisasi model. Selain itu, metode pelabelan otomatis berpengaruh signifikan terhadap performa LSTM. Penggunaan label mBERT menghasilkan akurasi dan konsistensi klasifikasi yang lebih tinggi dibandingkan IndoBERT.

Kata kunci: Analisis Sentimen, Berbasis Aspek, LSTM, Ruangguru, IndoBERT, mBERT

ABSTRACT

The online learning platform Ruangguru receives various user reviews containing opinions on multiple service aspects. This study aims to implement aspect-based sentiment analysis on Ruangguru platform reviews using the Long Short-Term Memory (LSTM) method. The research stages include collecting review data, text preprocessing, aspect extraction using Latent Dirichlet Allocation (LDA) with the optimal number of topics determined based on the coherence score, and automatic sentiment labeling using mBERT and IndoBERT models. The LSTM model was trained under two scenarios: combining all aspects into a single model and training separate models for each aspect. Model performance was evaluated using a confusion matrix to analyze the classification capability for positive, neutral, and negative sentiments, as well as by calculating accuracy, precision, recall, and F1-score. The results indicate that training on the combined aspects produces more stable and consistent performance due to the larger amount of data and more diverse sentiment distribution. In contrast, training separate models for each aspect tends to be biased toward the dominant class because of imbalanced data distribution, thereby reducing the model's generalization capability. Furthermore, the automatic labeling method significantly affects LSTM performance. The use of mBERT-generated labels yields higher accuracy and classification consistency compared to IndoBERT.

Keywords: *Sentiment Analysis, Aspect-Based, LSTM, Ruangguru, IndoBERT, mBERT*

MOTTO DAN PERSEMBAHAN

"Setiap langkah yang dijalani dengan niat yang baik, usaha yang sungguh-sungguh, dan doa yang tidak pernah terputus akan selalu menemukan jalannya menuju hasil terbaik menurut kehendak Allah SWT."

Alhamdulillah, dengan penuh rasa syukur kehadirat Allah SWT atas segala rahmat, karunia, dan kemudahan yang diberikan, sehingga penulis dapat menyelesaikan tesis ini. Penulis persembahkan Tesis ini kepada :

1. Orang Tua Tercinta, khususnya Ayahanda almarhum Mirhan Ibrahim, yang semasa hidupnya telah menjadi sumber inspirasi, teladan, dan motivasi bagi penulis hingga mampu menyelesaikan studi dan penelitian ini. Semoga segala amal dan jasa beliau mendapatkan balasan terbaik di sisi Allah SWT.
2. Orang Tua Ibunda Emilia Gustina, Ibu mertua Nurliah, dan Ayah mertua Yuniman, yang senantiasa memberikan doa, semangat, nasihat, dan dukungan moral selama penulis menempuh pendidikan hingga penyusunan tesis ini dapat diselesaikan dengan baik.
3. Pada Suami tercinta Muhammad Burlian Firmansyah, yang selalu hadir mendampingi, memberikan dukungan, pengertian, kekuatan dan doa dalam setiap proses perjuangan penulis selama menempuh studi magister ini.
4. Untuk Anakku Muhammad Ardan Delian Firmansyah, yang kehadirannya menjadi sumber kekuatan dan motivasi terbesar bagi penulis. Pada saat penyelesaian studi ini, usiamu masih 1 tahun. Semoga kelak, ketika engkau tumbuh dan memahami arti perjuangan, engkau dapat melihat bahwa pendidikan merupakan bekal penting dalam membentuk masa depan, sehingga kedua orang tuamu berupaya menempuh pendidikan setinggi yang kami mampu.
5. Pada keluarga Kakak, Adik, Ponakan yang juga dengan segenap hati memberikan dukungannya.
6. Pada Teman dan Rekan-rekan kerja yang mendukung dan membantu Penulis menyelesaian S2 ini.

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan penyusunan tesis yang berjudul "**Analisis Sentimen Berbasis Aspek pada Ulasan Platform Ruangguru Menggunakan Metode *Long Short-Term Memory* (LSTM)**". Tesis ini disusun sebagai salah satu syarat untuk menyelesaikan program studi Magister Teknik Informatika di Universitas Bina Darma. Penulis menyadari bahwa penyusunan Tesis ini tidak terlepas dari dukungan, bimbingan, dan doa dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Ibu Prof. Dr. Sunda Ariana, M.Pd.,M.M. Selaku Rektor Universitas Bina Darma Palembang.
2. Prof. Dr. Ir. Achmad Syarifudin, M.Sc. Selaku Direktur Pascasarjana Universitas Bina Darma Palembang.
3. Dr. Usman Ependi, S.Kom., M.Kom Selaku Pembimbing dan Ketua Program Studi Magister Teknik Informatika yang telah memberi saran, kritik, arahan dan memberi dorongan dalam penyusunan tesis ini.
4. Dr. Yesi Novaria Kunang, S.T., M.Kom. selaku penguji yang telah memberikan arahan dalam penulisan tesis ini
5. Nita Rosa Damayanti, M.Kom., Ph.D. selaku penguji yang telah memberikan motivasi dan arahan dalam penulisan tesis ini.
6. Pihak Sekretariat Pascasarjana Universitas Bina Darma Palembang yang telah memberikan bimbingan pelayanan dengan baik.

Penulis menyadari bahwa Tesis ini masih jauh dari sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun untuk penyempurnaan di masa yang akan datang. Semoga Tesis ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan dan teknologi.

Palembang, 06 Januari 2026

Delta Apriza

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi telah mengubah cara berpikir manusia dalam menawarkan kemudahan dan fleksibilitas untuk mengakses dan memperoleh pendidikan. Ruangguru merupakan salah satu platform pembelajaran daring terbesar di Indonesia yang menawarkan berbagai layanan edukasi termasuk layanan kelas virtual, platform ujian online, video belajar berlangganan, marketplace les privat, serta konten-konten pendidikan lainnya (Fitri et al., 2020). Sejak pandemi COVID-19, kebutuhan akan pembelajaran berbasis teknologi semakin mendesak. Sistem pendidikan di berbagai negara dipaksa untuk beradaptasi dengan metode pembelajaran daring dan Ruangguru menjadi salah satu aplikasi pendidikan yang paling banyak digunakan di Indonesia (Ali et al., 2025). Hal ini menunjukkan bahwa platform ini berhasil menarik perhatian publik secara luas.

Namun, seiring dengan meningkatnya jumlah pengguna, tantangan utama yang dihadapi adalah bagaimana menjaga dan meningkatkan kualitas layanan. Salah satu cara untuk mengevaluasi dan memahami kebutuhan pengguna adalah dengan menganalisis ulasan yang diberikan oleh pengguna. Ulasan produk memberikan informasi mengenai kualitas suatu produk serta memberi dampak besar bagi pengguna dan pengembang. Kepuasan pelanggan adalah isu penting yang menjadi tujuan perusahaan (Hasugian et al., 2023). Menurut survei yang dilakukan oleh situs TrustYou.com menyatakan bahwa sebanyak 93% orang melakukan pencarian ulasan terlebih dahulu sebelum memilih suatu produk (Cahyaningtyas, 2021).

Analisis terhadap ulasan ini dapat dilakukan melalui pendekatan analisis sentimen berbasis aspek yang bertujuan mengidentifikasi dan mengklasifikasikan opini pengguna ke dalam kategori positif dan negatif terkait berbagai aspek dalam suatu entitas (Sparta & Rimadias, 2024). Penerapan analisis sentimen berbasis aspek sangat relevan dalam konteks ulasan terhadap Ruangguru, karena pengguna sering kali menyampaikan opini mereka secara rinci terhadap fitur aplikasi yang

digunakan. Namun, kompleksitas gaya bahasa dalam ulasan seringkali ditemukan dan menjadi tantangan tersendiri, terutama saat teks mengandung makna implisit, ironi, sindiran, ataupun sarkasme. Oleh karena itu, analisis sentimen yang hanya berdasarkan pola kata tidak cukup efektif untuk memahami kompleksitas dalam opini pengguna (Warakmulty et al., 2025).

Penelitian tentang ABSA telah dilakukan menggunakan berbagai metode. Pada penelitian (Asmara et al., 2024) metode yang digunakan untuk klasifikasi aspek dan sentimen adalah metode *machine learning*, sedangkan pada penelitian (Cahyaningtyas, 2021) proses klasifikasi aspek dan sentimen berhasil menggunakan metode *deep learning*. Selain itu, proses analisis sentimen berbasis aspek dilakukan dalam beberapa domain seperti ulasan hotel, produk, dan aplikasi dengan berbagai aspek. Adapun data yang digunakan dapat berasal dari beberapa sumber, seperti pada penelitian (Nurhidayat & Dewi, 2023) menggunakan data yang sudah ada dari situs *kaggle* ataupun menggunakan metode *scraping* dibeberapa situs online seperti yang dilakukan oleh (Asmara et al., 2024). Selain itu, bahasa ulasan juga dapat menggunakan bahasa inggris seperti penelitian yang dilakukan oleh (Naquitasia et al., 2022) dan menggunakan bahasa indonesia seperti penelitian yang dilakukan oleh (Cahyaningtyas, 2021). Meskipun telah diterapkan pada beberapa domain, bahasa, dan metode, namun proses analisis sentimen berbasis aspek menggunakan data ulasan aplikasi Ruangguru berbahasa Indonesia dengan metode *Long Short-Term Memory* (LSTM) masih terbilang cukup baru dan berdasarkan penelitian sebelumnya yang dilakukan oleh (Ramadhan, 2023) dan (Cahyaningtyas, 2021) tentang analisis sentimen menggunakan metode *Long Short-Term Memory* (LSTM) menghasilkan akurasi yang tinggi.

Penelitian ini bertujuan untuk membuat sebuah model sentimen berbasis aspek pada data ulasan platform Ruangguru. Model ini dibuat dengan memanfaatkan teknologi *deep learning* berbasis *Long Short-Term Memory* (LSTM). Hasil dari penelitian ini diharapkan mampu melakukan analisa informasi terkait apa saja aspek yang dibahas pada ulasan platform Ruangguru agar dapat memahami bagaimana pengguna merespon berbagai fitur yang ditawarkan oleh Ruangguru sehingga informasi tersebut dapat digunakan untuk peningkatan

kualitas layanan yang diberikan kepada pengguna.

1.2 Identifikasi Masalah

Berdasarkan latar belakang masalah diatas, maka penulis merumuskan identifikasi masalah dalam penelitian ini yaitu sebagai berikut:

1. Banyaknya ulasan tidak terstruktur dari pengguna Ruangguru

   Platform Ruangguru menerima ribuan ulasan dari penggunanya di berbagai kanal seperti *Play Store*. Ulasan ini mengandung opini yang beragam terkait berbagai aspek layanan. Namun, data tersebut bersifat tidak terstruktur dan sulit dianalisis secara manual karena jumlahnya yang besar dan gaya bahasa yang bervariasi.

2. Analisis sentimen konvensional hanya memberikan hasil secara umum

   Kebanyakan analisis sentimen yang dilakukan selama ini hanya memfokuskan pada klasifikasi sentimen secara keseluruhan meliputi sentimen positif dan negatif terhadap suatu teks ulasan, tanpa mengidentifikasi aspek-aspek yang lebih spesifik. Hal ini menyebabkan berkurangnya nilai informasi yang dapat digunakan secara langsung oleh pengelola platform untuk pengambilan keputusan.

3. Kurangnya penelitian tentang analisis sentimen berbasis aspek untuk teks ulasan berbahasa Indonesia

   Sebagian besar penelitian tentang analisis sentimen berbasis aspek masih didominasi oleh data teks berbahasa inggris dengan banyaknya sumber daya dan model yang mendukung. Sebaliknya, ulasan berbahasa Indonesia belum banyak mendapatkan perhatian dalam riset analisis sentimen berbasis aspek, padahal memiliki karakteristik linguistik yang berbeda dan lebih kompleks, seperti struktur kalimat bebas dan penggunaan bahasa informal.

4. Kebutuhan akan model yang mampu memahami konteks urutan kata dalam teks

   Metode klasifikasi tradisional seperti Naive Bayes atau SVM memiliki keterbatasan dalam memahami konteks kalimat dan urutan kata dalam teks. Oleh karena itu, diperlukan model *deep learning* seperti *Long Short-Term Memory* (LSTM) yang mampu memproses data teks sebagai urutan (*sequence*), sehingga

dapat mengenali pola sentimen yang kompleks dalam ulasan pengguna.

1.3 Perumusan Masalah

Berdasarkan identifikasi masalah diatas, maka penulis merumuskan permasalahan dalam penelitian ini yaitu bagaimana penerapan model *Long Short-Term Memory (*LSTM) untuk melakukan klasifikasi sentimen berbasis aspek terhadap ulasan pengguna pada platform kursus online Ruangguru?

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dibuat, maka tujuan dari penelitian ini ialah untuk menerapkan model *Long Short-Term Memory (*LSTM) untuk melakukan klasifikasi sentimen berbasis aspek terhadap ulasan pengguna pada platform kursus online Ruangguru.

1.5 Kebaruan

Kebaruan yang ada pada penelitian ini meliputi beberapa aspek yang belum banyak diimplemetasikan dalam penelitian sebelumnya yaitu sebagai berikut:

1. Fokus pada Platform Lokal (Ruangguru) sebagai Objek Studi

Sebagian besar penelitian terdahulu dalam bidang analisis sentimen berbasis aspek lebih banyak berfokus pada platform global seperti Amazon, Yelp, Google Review, atau Coursera. Penelitian ini menghadirkan kebaruan dengan menjadikan Ruangguru sebagai salah satu platform kursus online terkemuka di Indonesia untuk objek kajian utama. Dengan demikian, penelitian ini memberikan kontribusi pada kajian analisis sentimen dalam konteks layanan edutech lokal berbahasa Indonesia, yang masih sangat terbatas dalam literatur ilmiah.

2. Penerapan analisis sentimen berbasis aspek dalam Bahasa Indonesia

Analisis sentimen berbasis aspek umumnya dilakukan dalam bahasa Inggris. Penelitian ini menawarkan kebaruan dengan mengadaptasi analisis berbasis aspek untuk teks berbahasa Indonesia yang memiliki tantangan tersendiri seperti struktur kalimat, imbuhan, dan kosakata informal.

3. Pemetaan aspek ke sentimen dalam konteks edukasi daring

Penelitian ini tidak hanya mengidentifikasi apakah ulasan bersifat positif atau negatif secara keseluruhan, melainkan juga mengevaluasi sentimen terhadap setiap aspek spesifik. Hal ini memberikan nilai tambah berupa analisis opini yang lebih terfokus, sehingga dapat digunakan langsung oleh pengambil keputusan untuk memperbaiki fitur atau strategi layanan kursus online.

1.6 Manfaat Penelitian

Manfaat-manfaat yang diperoleh dari hasil penelitian ini dapat dilihat dari beberapa sudut pandang, yaitu bagi penulis, pengguna sistem, dan peneliti lain.

1. Memberikan kontribusi ilmiah dalam pengembangan metode analisis sentimen berbasis aspek khususnya dalam konteks teks ulasan berbahasa Indonesia.
2. Mengidentifikasi opini positif, netral dan negatif terhadap masing-masing aspek secara spesifik, sehingga dapat menjadi bahan evaluasi dan perbaikan layanan yang lebih terarah.
3. Mendorong peningkatan kualitas layanan aplikasi Ruangguru melalui masukan langsung dari ulasan pengguna yang dianalisis secara sistematis.

1.7 Pembatasan Masalah

Agar penelitian ini lebih terarah, maka perlu adanya batasan masalah yaitu sebagai berikut:

1. Data yang digunakan merupakan data ulasan pengguna platform Ruangguru dari *Google play store* 5 tahun terakhir.
2. Sentimen yang digunakan adalah postif, negatif dan netral.
3. Model yang digunakan untuk klasifikasi sentimen adalah *Long Short-Term Memory* (LSTM) tanpa membandingkan dengan metode *deep learning* lainnya.
4. Penelitian hanya mencakup analisis terhadap ulasan dalam bentuk teks.

1.8 Sistematika Penulisan

Sistematika Penulisan proposal tesis ini dimaksudkan agar dapat memberikan garis besar secara jelas sehingga terlihat hubungan antara bab yang satu dengan bab yang lainnya. Susunan dan struktur proposal tesis dijabarkan dibawah ini sebagai

berikut:

BAB I PENDAHULUAN

Pada bab ini membahas tentang latar belakang, identifikasi masalah, batasan masalah, rumusan masalah, tujuan dan manfaat penelitian, ruang lingkup penelitian, serta susunan dan struktur proposal tesis.

BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI

Pada bab ini membahas tentang tinjauan pustaka, landasan teori yang relevan dari penelitian yang akan dilakukan.

BAB III METODOLOGI PENELITIAN

Pada bab ini pembahasannya yang terdiri dari desain dan jadwal penelitian, data penelitian, kemudian konsep dan metode penelitian yang digunakan, metode pengumpulan data serta teknik analisis data.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil analisis menggunakan model *Long Short-Term Memory (*LSTM), termasuk interpretasi hasil dan akurasi prediksi.

BAB V SIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari penelitian yang dilakukan, serta saran-saran yang dapat dikemukakan untuk penelitian lebih lanjut atau untuk penerapan model analisis sentimen berbasis aspek.

DAFTAR PUSTAKA

Bagian ini mencantumkan semua sumber referensi yang digunakan dalam penulisan penelitian, termasuk buku, artikel jurnal, dan sumber lainnya yang relevan dengan topik penelitian.

LAMPIRAN

BAB II
TINJAUAN PUSTAKA DAN LANDASAN TEORI

2.1 Tinjauan Pustaka

Penelitian mengenai analisis sentimen berbasis aspek telah mengalami perkembangan yang cukup pesat (Novirianto, 2023) khususnya dalam sektor komersil. Adapun *Aspect Based Sentiment Analysis* (ABSA) memberikan pendekatan yang lebih terfokus dibandingkan dengan analisis sentimen biasa karena mampu mengidentifikasi aspek-aspek spesifik dari suatu entitas serta polaritas opini terhadap masing-masing aspek tersebut. Pada beberapa penelitian, proses analisis sentimen berbasis aspek dilakukan dalam beberapa domain seperti ulasan hotel, tempat wisata, produk, dan aplikasi dengan berbagai aspek. Adapun data yang digunakan dapat berasal dari beberapa sumber, seperti pada penelitian (Nurhidayat & Dewi, 2023) menggunakan data yang sudah ada dari situs *kaggle* dan pada penelitian (Asmara et al., 2024) menggunakan metode *scraping* dibeberapa situs online. Selain itu, metode yang digunakan pada *Aspect Based Sentiment Analysis* (ABSA) dapat berbeda di setiap penelitian. Pada penelitian (Asmara et al., 2024) metode yang digunakan untuk klasifikasi aspek dan sentimen adalah metode *machine learning*, sedangkan pada penelitian (Cahyaningtyas, 2021) proses klasifikasi aspek dan sentimen berhasil menggunakan metode *deep learning*. Kemudian, bahasa ulasan juga dapat menggunakan bahasa inggris seperti peneltian (Naquitasia et al., 2022) dan menggunakan bahasa indonesia seperti penelitian yang dilakukan oleh (Cahyaningtyas, 2021). Meskipun telah diterapkan pada beberapa domain, bahasa, dan metode, namun proses *Aspect Based Sentiment Analysis* (ABSA) menggunakan data ulasan aplikasi Ruangguru berbahasa Indonesia dengan metode *Long Short-Term Memory* (LSTM) masih terbilang cukup baru.

Asmara et al. (2024) telah melakukan penelitian terkait Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi RuangGuru Menggunakan *Support Vector Machine* dan *Naive Bayes Classifier*. Penelitian ini menggunakan Pendekatan klasifikasi dua tahap secara sekuensial yaitu pertama mengidentifikasi aspek dalam ulasan, kemudian mengklasifikasikan sentimen (positif/negatif) berdasarkan aspek

tersebut. Data yang digunakan sebanyak 1865 ulasan berbahasa Indonesia dari *Google Play Store*. Proses meliputi *scraping* data, *preprocessing* (*stopword removal, stemming, tokenisasi*), dan ekstraksi fitur dengan TF-IDF. Adapun hasil rata-rata akurasi yang didapatkan SVM lebih tinggi daripada NBC yaitu 0,91%.

Pada penelitian Mubaroroh et al. (2022) metode yang dikembangkan ialah analisis sentimen menggunakan metode multinomial *Naive Bayes Classifier* dengan normalisasi kata menggunakan metode *Levenshtein Distance* untuk memperbaiki kesalahan penulisan atau typo secara otomatis, sehingga hasil analisis menjadi lebih akurat. Dataset yang digunakan terdiri dari 1000 ulasan paling relevan, yang berasal dari hasil *scraping Google Play*, dan telah disaring dari ulasan netral. Hasil klasifikasi menunjukkan bahwa 78,9% ulasan mengandung sentimen positif, sedangkan 21,1% ulasan bersentimen negatif. Evaluasi model dilakukan dengan metode 10-fold cross validation dan menghasilkan rata-rata akurasi sebesar 88,20%, dengan akurasi tertinggi pada fold ke-8 mencapai 94%. Penambahan proses *Levenshtein Distance* terbukti mampu meningkatkan akurasi model sebesar 0,6% dibandingkan dengan normalisasi tanpa metode tersebut.

Penelitian lain terkait metode *Long Short-Term Memory* (LSTM) pada domain kursus online Ruangguru adalah penelitian yang dilakukan oleh (Ramadhan, 2023). Penelitian ini bertujuan untuk mengevaluasi opini pengguna terhadap aplikasi Ruangguru melalui ulasan di *Google Play Store* menggunakan metode *deep learning*. Data diperoleh melalui teknik scraping yang menghasilkan 100.000 ulasan, kemudian difilter menjadi 65.184 setelah tahap pembersihan dan *text preprocessing*. Proses pelabelan dilakukan secara otomatis menggunakan pendekatan berbasis kamus (*lexicon-based*) dengan Indonesia Sentiment Lexicon (InSet). Selain itu, dalam penelitian ini juga membandingkan dengan algoritma *machine learning* klasik seperti SVM, Naive Bayes, Random Forest, dan K-Nearest Neighbor. Hasilnya, LSTM menunjukkan performa tertinggi dengan akurasi 94,75%, melampaui SVM (92,89%), Random Forest (91,51%), KNN (86,25%), dan Naïve Bayes (18,97%). Selain itu, validasi menunjukkan bahwa model LSTM tidak mengalami *overfitting*, ditunjukkan oleh kurva akurasi validasi yang stabil pada setiap *epoch*.

Nurhidayat & Dewi (2023) melakukan penelitian analisis sentimen berbasis aspek dengan ekstraksi fitur menggunakan N-Gram (unigram, bigram, trigram) yang bertujuan menggambarkan pola hubungan antar kata dalam ulasan teks secara lebih rinci seperti membantu mengidentifikasi kombinasi kata yang sering muncul bersama, yang tidak bisa ditangkap jika hanya menggunakan kata tunggal (unigram). Misalnya, kata "tidak" dan "bagus" secara terpisah mungkin tidak berarti negatif, tetapi dalam bigram "tidak bagus" maknanya jelas negatif.

Selain itu, Cahyaningtyas (2021) melakukan analisis sentimen pada Data Hotel Berbahasa Indonesia dengan membandingkan kinerj beberapa metode *deep learning* yaitu RNN, LSTM, GRU, BiLSTM, Attention BiLSTM, CNN, CNN-LSTM, dan CNN-BiLSTM. Hasilnya menunjukkan bahwa model GRU memberikan performa terbaik dalam klasifikasi aspek dengan akurasi 0,952 (top-1) dan 0,984 (top-2), sedangkan CNN-LSTM menjadi model terbaik untuk klasifikasi sentimen dengan akurasi rata-rata sebesar 0,911 jika diterapkan secara terpisah di setiap aspek. Penelitian ini menegaskan bahwa pendekatan klasifikasi sentimen per aspek (skenario kedua) memberikan hasil yang lebih akurat dibanding klasifikasi keseluruhan aspek. Akhirnya, model-model terbaik ini digunakan untuk menganalisis data ulasan nyata dari Hotel Pesonna Malioboro dan menghasilkan akurasi klasifikasi aspek sebesar 0,992 dan sentimen sebesar 0,976, menunjukkan keberhasilan metode yang diusulkan. Berikut merupakan penelitian terkait dapat dilihat pada Tabel 2.1.

Tabel 2.1 Penelitian Terkait

No	Judul dan penulis	Metode	Hasil	Keterangan
1	Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi RuangGuru Menggunakan *Support Vector*	*Support Vector Machine* dan *Naive Bayes Classifier*	*Accuracy score :* NBC – NBC : 70% NBC – SVM : 70% SVM – NBC : 60%	Pada penelitian ini representasi fitur hanya menggunakan TF-IDF yang bersifat statistik dan tidak mempertimbangkan makna kata secara

No	Judul dan penulis	Metode	Hasil	Keterangan
	Machine dan *Naive Bayes Classifier* Asmara et al. (2024)		SVM – SVM: 80%	semantik. Selain itu, penghapusan *stopword* pada saat preprocessing data teks dapat mengurangi akurasi karena beberapa kata penting justru terhapus.
2	Analisis Sentimen Data Ulasan Aplikasi Ruangguru Pada Situs Google Play Menggunakan Algoritma *Naive Bayes Classifier* Dengan Normalisasi Kata *Levenshtein Distance* Mubaroroh et al. (2022)	Multinomial *Naive Bayes Classifier* dan *Levenshtein Distance*	*Accuracy score :* Multinomial *Naive Bayes Classifier* : 87,60% Multinomial *Naive Bayes Classifier* dengan Normalisasi Kata Levenshtein Distance : 88,20%	Pada penelitian ini evaluasi performa model lebih difokuskan pada metrik akurasi, tanpa disertai analisis mendalam terhadap metrik lain seperti precision, recall, dan F1-score yang seharusnya digunakan untuk mengukur kinerja model pada data yang mungkin tidak seimbang. Selain itu, penelitian ini hanya melakukan analisis sentimen secara umum terhadap keseluruhan ulasan dimana label tidak didasarkan pada aspek, padahal dalam satu ulasan bisa

No	Judul dan penulis	Metode	Hasil	Keterangan
				terdapat campuran opini positif dan negatif untuk aspek yang berbeda.
3	Analisis Sentimen Ulasan Aplikasi Ruangguru Dengan Algoritma *Long Short-Term Memory (LSTM)* (Ramadhan, 2023)	*Long Short-Term Memory* (LSTM)	Accuracy score : LSTM : 94,75%, SVM : 92,89% Random Forest : 91,51% KNN : 86,25% Naive Bayes : 18,97%	Penelitian hanya mengklasifikasikan ulasan berdasarkan polaritas (positif, negatif, netral), tanpa mengidentifikasi aspek-aspek spesifik (misal: fitur video, harga, layanan guru), sehingga kurang memberikan insight mendalam bagi pengembangan aplikasi.
4	Penerapan Algoritma K-Nearest Neighbor Dan Fitur Ekstraksi N-Gram dalam Analisis Sentimen Berbasis Aspek Nurhidayat & Dewi (2023)	K-Nearest Neighbor (KNN) dan Fitur Ekstraksi N-Gram	*Accuracy score :* Aspek Aroma: 91,9% (gabungan unigram, bigram, trigram) Aspek Harga: 95,4% (unigram) Aspek Kemasan: 98,6% (trigram) Aspek Efektivitas: 88,8% (gabungan	Pada penelitian penggunaan trigram dalam beberapa aspek justru menunjukkan penurunan akurasi dibandingkan unigram atau bigram. Hal ini menandakan perlunya analisis lebih lanjut terhadap pemilihan fitur n-gram.

No	Judul dan penulis	Metode	Hasil	Keterangan
			unigram, bigram, trigram)	
5	*Deep Learning* pada *Aspect-based Sentiment Analysis* pada Data Hotel Berbahasa Indonesia Cahyaningtyas (2021)	RNN, LSTM, GRU, BiLSTM, *Attention* BiLSTM, CNN, CNN-LSTM, dan	*Accuracy score :* Akurasi klasifikasi aspek: **0,992** Akurasi klasifikasi sentimen: **0,976**	Pada penelitian ini Aspek-aspek yang dianalisis ditentukan terlebih dahulu dan tidak diekstraksi secara otomatis dari data. Hal ini membuat sistem kurang fleksibel terhadap domain baru atau perubahan struktur data yang memiliki aspek berbeda.

2.2 Landasan Teori

2.1.1 *Aspect-based Sentiment Analysis* (ABSA)

Aspect-Based Sentiment Analysis (ABSA) adalah suatu model yang digunakan untuk mengekstraksi aspek serta sentimen dari sebuah kalimat sehingga dapat diidentifikasi jumlah aspek yang terdapat di dalamnya dan ditentukan polaritas sentimen untuk setiap aspek tersebut dengan tujuan untuk mengurangi bias informasi (Amien et al., 2021). Sebagai contoh, dalam kalimat: “Materinya sangat menarik, tetapi antarmuka aplikasinya membingungkan”, opini positif diarahkan pada aspek materi, sedangkan opini negatif diarahkan pada aspek antarmuka aplikasi. ABSA memungkinkan sistem untuk memahami nuansa tersebut dengan menguraikan sentimen per aspek.

(Pontiki et al, 2016) menyatakan bahwa ABSA terdiri dari tiga tugas utama, yaitu identifikasi aspek, ekstraksi opini, dan klasifikasi polaritas sentimen terhadap aspek yang dimaksud. Pendekatan ini banyak digunakan dalam bidang layanan

publik, pemasaran, dan pendidikan untuk memahami secara komprehensif opini pengguna terhadap berbagai komponen layanan. Menurut (Cahyaningtyas, 2021) proses dalam mengidentifikasi aspek dapat dilakukan dengan berbagai cara. Beberapa penelitian memanfaatkan proses POS *Tagger* untuk mengidentifikasi kelas kata (Khong et al., 2018), (Jang et al., 2020), dan (Afzaal et al., 2019). Namun di beberapa penelitian, identifikasi kelas kata dalam proses ABSA menggunakan POS *Tagger* tidak digunakan karena dinilai membutuhkan *resources* yang cukup besar (Gu et al., 2018). Pada penelitian (Peng et al., 2018), (Ekawati & Khodra, 2017), dan (S. Wu et al., 2019) proses identifikasi kelas aspek maupun opini dilakukan menggunakan proses klasifikasi menggunakan pelabelan aspek maupun sentimen untuk setiap data. Klasifikasi sentimen yang dilakukan setelah proses klasifikasi aspek, dapat dilakukan menggunakan seluruh data aspek (Al-Smadi et al., 2019) maupun untuk setiap aspek (Ekawati & Khodra, 2017).

2.1.2 Ulasan Pengguna sebagai Sumber Data Teks

Ulasan pengguna (*user reviews*) merupakan salah satu bentuk data teks tidak terstruktur yang berasal dari pengalaman subjektif pengguna terhadap suatu produk atau layanan (Cahyaningtyas, 2021). Dalam konteks penelitian ini, ulasan pengguna terhadap platform Ruangguru menjadi sumber informasi utama yang akan dianalisis untuk mengetahui persepsi pengguna terhadap aspek-aspek layanan seperti kualitas materi, kinerja tutor, antarmuka aplikasi, serta kecepatan dan ketepatan layanan pelanggan. (Pang dan Lee, 2008) menyebutkan bahwa ulasan yang ditulis oleh pengguna merupakan bentuk opini eksplisit dan sangat bernilai dalam menggambarkan pengalaman aktual. Oleh karena itu, ulasan pengguna memiliki potensi yang besar untuk dieksplorasi dalam analisis sentimen dan studi kepuasan pelanggan.

2.1.3 Platform Ruangguru

Ruangguru merupakan salah satu *marketplace* bidang pendidikan yang ada di Indonesia. Ruangguru menghubungkan pelajar dengan pengajar dimana pelajar dapat mencari dan menemukan pengajar berdasarkan kebutuhannya begitupun sebaliknya, pengajar dapat menyalurkan ilmu yang dimilikinya(Rahadian et al.,

2019). Kegiatan belajar mengajar yang difasilitasi oleh Ruangguru menghadirkan sistem pengelolaan pembelajaran *e-learning* yang dapat dimanfaatkan oleh murid dan guru untuk mengatur proses pembelajaran secara *virtual*. (Rahadian et al., 2019).

Menurut (Lubis et al., 2021) penggunaan Ruangguru selama masa pandemi COVID-19 sangat efektif karena mampu menjawab kebutuhan pembelajaran jarak jauh secara fleksibel. Aplikasi ini dinilai efektif dalam membantu siswa memahami materi pelajaran secara mandiri, kapan pun dan di mana pun. Lebih lanjut, menurut (Yazid et al., 2019) tingkat kepuasan pengguna Ruangguru berkorelasi erat dengan faktor-faktor seperti *content*, keakuratan data, tampilan antarmuka sistem, kemudahan pengguna, dan ketepatan waktu. Pengguna yang merasa terbantu dengan pendekatan belajar yang disesuaikan akan lebih cenderung memberikan umpan balik positif.

2.1.4 *Natural Language Processing (NLP)*

Natural Language Processing (NLP) merupakan bidang penelitian dan penerapan yang mengkaji cara komputer dalam memahami serta mengolah teks berbahasa alami. (Astiningrum et al., 2018). Dalam penelitian ini, *Natural Language Processing* (NLP) digunakan dalam pra-pemrosesan (*preprocessing*) data teks ulasan dengan melakukan beberapa tahapan dasar yaitu sebagai berikut:

1. Membersihkan teks ulasan dari elemen yang tidak relevan.
2. Mengonversi kata ke bentuk dasarnya (*stemming/lemmatization*).
3. Menghapus kata-kata umum yang tidak bermakna (*stopword removal*).
4. Mengonversi teks ke dalam format numerik agar dapat diproses oleh model LSTM.

2.1.5 Pemodelan Topik (*Topic Modeling*)

Pemodelan topik adalah teknik yang umum digunakan untuk segmentasi data dan sistem rekomendasi dengan cara mengelompokkan data berdasarkan kesamaan ciri-cirinya (An et al., 2023). *Topic modeling* merupakan salah satu bentuk model berbasis probabilistik yang diterapkan dalam *machine learning* dan *Natural Language Processing* (NLP) untuk menemukan serta menggambarkan topik-topik

tersembunyi dalam kumpulan dokumen teks yang tidak terstruktur. (Sirkis & Maitland, 2023). *Topic modeling* digunakan untuk mengelompokkan sekumpulan data atau teks ke dalam beberapa topik tertentu. Pendekatan ini dapat digunakan untuk segmentasi produk serta penyusunan rekomendasi produk bagi pemangku kepentingan guna memperoleh manfaat berkelanjutan (Raditya et al., 2017).

2.1.6 *Recurrent Neural Network (RNN)*

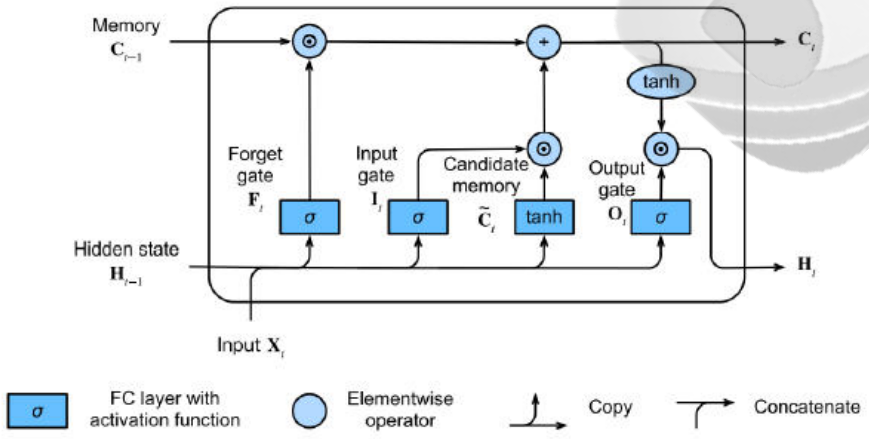
Recurrent Neural Network (RNN) merupakan jenis jaringan saraf yang dikembangkan untuk mengolah data berurutan dengan mempertahankan informasi dari tahap sebelumnya. Berbeda dengan jaringan saraf *feedforward*, RNN memiliki kinerja yang baik dalam pengolahan bahasa alami, analisis data deret waktu, dan pengenalan suara (Saputri & Baita, 2025). Kelebihan dari RNN bekerja dengan memanfaatkan hasil perhitungan sebelumnya untuk digunakan pada proses berikutnya, karena input kata diproses secara berurutan. Namun, RNN juga memiliki beberapa kelemahan. Salah satu tantangan utamanya adalah masalah *vanishing gradient*, di mana *gradien* yang dihitung selama pelatihan menjadi sangat kecil atau menghilang ketika data memiliki dependensi jangka panjang. Hal ini menyebabkan jaringan sulit belajar dari informasi historis yang jauh, sehingga kinerja model menurun untuk prediksi jangka panjang (Pascanu et al., 2013). Penelitian oleh (Baradja, 2023) menyoroti bahwa RNN memiliki keterbatasan dalam menghadapi prediksi dengan *frame* waktu yang panjang, yang dikenal sebagai masalah *long memory dependency.* Untuk mengatasi hal ini, digunakan arsitektur seperti LSTM dan GRU yang mampu menangani masalah *vanishing gradient* dan *gradient exploding*, sehingga lebih efektif dalam menangani, memprediksi, dan mengklasifikasikan data yang bersifat sekuensial.

2.1.7 *Long Short Term Memories* (LSTM)

LSTM merupakan salah satu metode *deep learning* yang cocok diaplikasikan untuk mengatasi data panjang karena masalah *vanishing gradient* yaitu ketika nilai gradien menjadi sangat kecil sehingga menyebabkan informasi dari waktu sebelumnya tidak bisa dipertahankan (Cahyaningtyas, 2021). LSTM memiliki struktur unik yang memungkinkan penyimpanan informasi untuk waktu yang lama,

sehingga dapat memproses urutan yang panjang dengan memanfaatkan sel memori untuk menyimpan informasi dalam jangka waktu lebih lama, sehingga dapat menangkap keterkaitan data yang melibatkan banyak langkah (Saputri & Baita, 2025). Dengan memanfaatkan mekanisme gerbang, LSTM mampu menentukan informasi yang perlu disimpan atau diabaikan, sehingga sangat efisien dalam mengatasi permasalahan yang melibatkan urutan data kompleks dengan rentang waktu yang panjang (Hochreiter & Schmidhuber, 1997).

Tidak seperti RNN yang cenderung mengalami kendala dalam mempertahankan informasi dalam jangka waktu panjang, LSTM dirancang untuk menyimpan informasi yang relevan lebih lama serta membuang data yang sudah tidak dibutuhkan. Kemampuan ini membuat LSTM lebih unggul dalam menangani, memprediksi, dan mengklasifikasikan data yang bersifat sekuensial atau berdasarkan urutan waktu tertentu. LSTM adalah salah satu metode *deep neural network* yang paling baik dalam menangani data sekuensial yang panjang dan mempelajari pengetahuan secara tersirat (Yusuf, 2023).



Gambar 2.1 Arsitektur *Long Short Term Memories* (LSTM)
(Sumber : (Yusuf, 2023))

Pada Gambar 2.1 menunjukkan gambar arsitektur *Long Short Term Memories* (LSTM) yang memiliki beberapa lapisan yaitu sebagai berikut

1. *Input gate*

   Input gate digunakan untuk mengontrol seberapa banyak informasi baru yang akan dimasukkan ke dalam sel memori (*cell state)* (Rolangon et al., 2023).

Adapun perhitungan *input gate* dapat menggunakan persamaan (2.1).

$$i_t = \sigma\ (W_i\ x_t + U_i + U_i\ h_{t-1} + b_i) \qquad (2.1)$$

2. *Forget gate*

Forget gate berfungsi untuk membuang informasi yang sudah tidak relevan dari sel memori (Rolangon et al., 2023). Adapun *input*nya adalah $ht-1$ dan xt dan nilai *output* antara (0,1). Jika *forget gate* = 1 informasi akan disimpan, jika *forget gate* = 0 informasi akan dilupakan. Hasil *forget gate* dapat menggunakan persamaan ke (2.2).

$$f_t = \sigma\ (W_f\ x_t + U_f\ h_{t-1} + b_f) \qquad (2.2)$$

3. *Output gate*

Output gate berfungsi untuk mengatur jumlah informasi yang akan dikeluarkan dari sel memori (Rolangon et al., 2023). Adapun untuk perhitungannya dapat menggunakan persamaan (2.3).

$$O_t = \sigma\ (W_o\ x_t\ + U_o\ h_{t-1} + b_o) \qquad (2.3)$$

4. *Hidden layer*

Hidden layer berperan sebagai lapisan utama dalam proses pengolahan yang menyimpan informasi dari urutan data sebelumnya (Yanti & Utami, 2025). Hasil perhitungannya dapat menggunakan persamaan (2.4).

$$H_t = O_t \odot \tanh(C_t) \qquad (2.4)$$

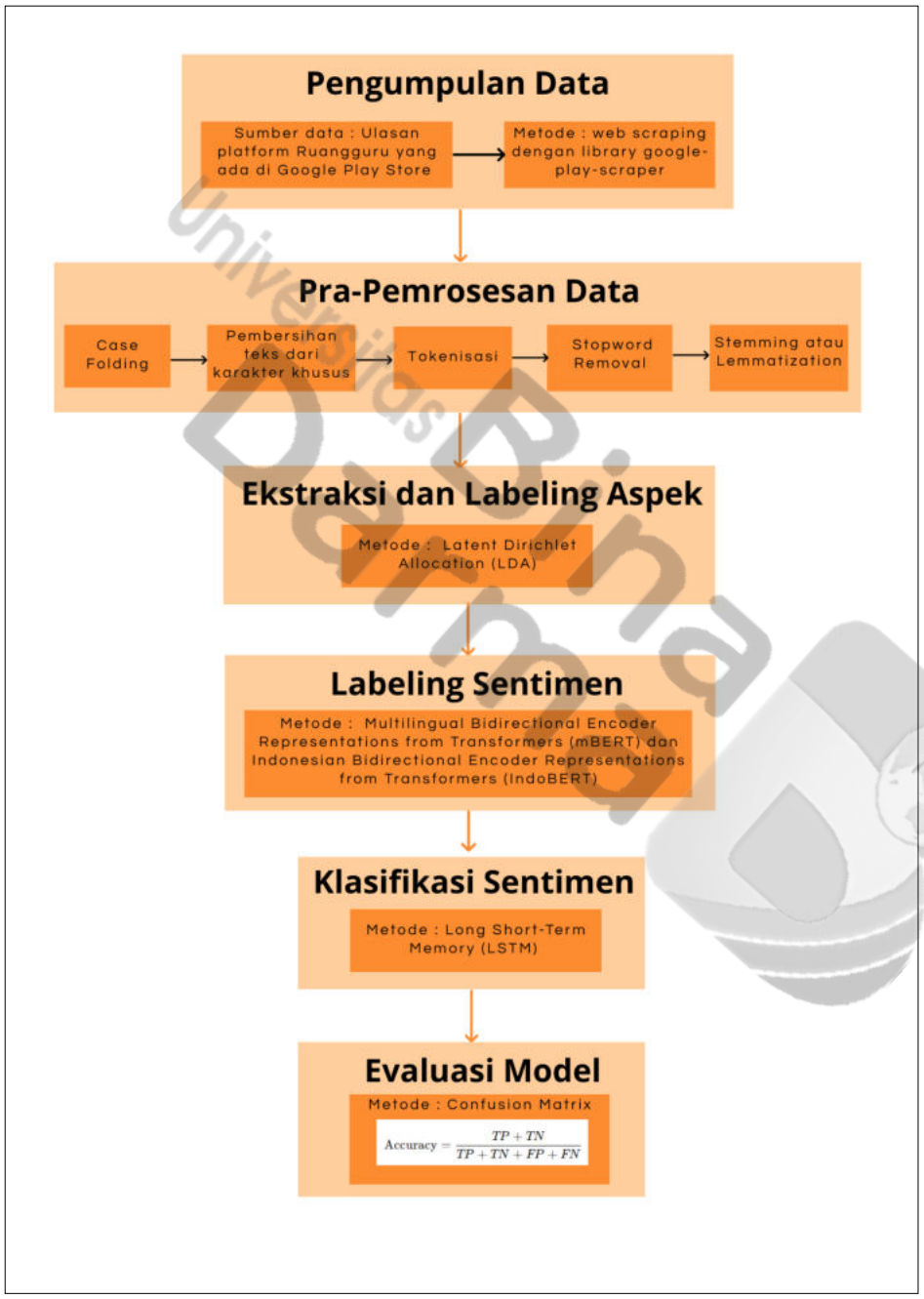
BAB III
METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan pada periode Agustus hingga Desember 2025. Seluruh proses penelitian dilakukan di tempat Laboratorium *Experimental Information Technologies* dengan perangkat keras yang mendukung untuk pengembangan model *deep learning*.

3.2 Prosedur Penelitan

Prosedur Penelitian ini dilakukan melalui serangkaian langkah sistematis yang bertujuan untuk untuk menganalisis sentimen berbasis aspek dari ulasan pengguna platform Ruangguru dengan menggunakan pendekatan *deep learning* yaitu model *Long Short-Term Memory* (LSTM). Prosedur penelitian ini disusun agar hasil yang diperoleh relevan dan dapat dijadikan dasar dalam pengambilan keputusan dalam peningkatan kualitas layanan aplikasi Ruangguru. Tahapan penelitian ini meliputi proses pengumpulan data dari ulasan pengguna platform Ruangguru di *google playstore* 5 tahun terakhir menggunakan metode *web scraping* dengan library google-play-scraper, pra-pemrosesan data yang meliputi proses *case folding,* pembersihan teks dari karakter khusus, *tokenisasi*, *stopword removal*, dan *stemming* atau *lemmatization*. Kemudian proses ekstraksi dan klasifikasi aspek menggunakan metode *Latent Dirichlet Allocation* (LDA), labeling data menggunakan metode BERT, Klasifikasi Sentimen dengan menggunakan metode *Long Short-Term Memory* (LSTM), dan evaluasi model menggunakan metode *Confusion Matrix* untuk mengukur seberapa baik model dapat mengklasifikasikan sentimen suatu teks. Adapun alur pada tahapan penelitian ini dapat dilihat pada Gambar 3.1.



Gambar 3. 1 Tahapan Penelitian

3.2.1 Pengumpulan Data Ulasan Pengguna Platform Ruangguru

Langkah pertama dalam prosedur penelitian ini adalah mengumpulkan data berupa ulasan pengguna platform Ruangguru. Data diambil dari sumber-sumber yang relevan dan *representative* yaitu melalui ulasan platform Ruangguru yang ada di *Google Play Store* dengan menggunakan *web scraping* dengan library google-play-scraper untuk memperoleh komentar dari pengguna yang telah mengunduh

dan menggunakan aplikasi Ruangguru. Adapun data yang dikumpulkan yaitu sebagai berikut:

a) Username ID pengguna

   Username atau ID pengguna adalah identitas unik dari pemberi ulasan yang berguna untuk mengidentifikasi pengguna yang sama, mencegah duplikasi data, dan dalam beberapa kasus dapat digunakan untuk analisis perilaku atau segmentasi pengguna.

b) Tanggal posting

   Tanggal ulasan menunjukkan waktu kapan pengguna memberikan komentarnya, yang berguna untuk mengetahui perkembangan atau perubahan sentimen dari waktu ke waktu serta memungkinkan analisis tren atau visualisasi distribusi sentimen secara kronologis.

c) Teks ulasan

   Bagian ini memuat komentar pengguna yang berisi opini, pengalaman, atau penilaian terhadap layanan, fitur, dan kualitas platform Ruangguru, yang nantinya menjadi data inti dalam proses analisis sentimen dan ekstraksi aspek.

Tahapan ini sangat penting karena kualitas data yang dikumpulkan akan berdampak langsung terhadap akurasi model dan validitas hasil analisis sentimen. Data mentah yang diperoleh umumnya masih berisi banyak *noise*, sehingga perlu dilakukan tahapan pra-pemrosesan data.

3.2.2 Pra-Pemrosesan Data

Tahapan Pra-Pemrosesan data dilakukan untuk memastikan bahwa data yang digunakan bersih, relevan, dan siap diproses lebih lanjut (Daffa et al., 2024). Pra-pemrosesan dilakukan secara bertahap dengan teknik sebagai berikut:

1) Case folding

Case folding merupakan tahap awal dalam pra-pemrosesan teks di mana semua huruf dalam setiap kata pada ulasan pengguna diubah menjadi huruf kecil guna memastikan konsistensi penulisan dan menghindari perbedaan representasi yang tidak perlu, karena dalam konteks analisis teks, kata “Bagus”, “BAGUS”, dan “bagus” sebenarnya memiliki makna yang sama dan seharusnya tidak

diperlakukan sebagai entitas yang berbeda oleh model pembelajaran mesin.

2) Pembersihan teks

Pembersihan teks adalah proses yang bertujuan untuk menghilangkan elemen-elemen yang tidak relevan atau mengganggu analisis seperti tanda baca (., !, ?), angka, simbol khusus (@, #, %, dll), emoji, URL, dan karakter non-alfabetik lainnya yang sering kali muncul dalam ulasan atau komentar pengguna, khususnya di media sosial atau platform digital. Semua elemen tersebut biasanya tidak memberikan kontribusi bermakna terhadap pemahaman semantik model tetapi dapat memperbesar kompleksitas data secara signifikan jika tidak dibersihkan.

3) Tokenisasi

Tokenisasi adalah tahap di mana setiap ulasan pengguna yang berbentuk kalimat panjang dipecah menjadi unit-unit kata (token) secara sistematis, dengan tujuan agar model dapat memahami struktur linguistik dari teks tersebut. Proses ini sangat penting karena dalam NLP, semua teknik analisis lanjutan seperti *word embedding*, *stemming*, dan klasifikasi bergantung pada struktur token sebagai satuan analisis dasar dari teks.

4) Stopword Removal

Stopword removal adalah proses eliminasi terhadap kata-kata umum yang sangat sering muncul dalam teks Bahasa Indonesia namun tidak membawa makna informatif dalam konteks analisis sentimen, seperti “yang”, “dan”, “itu”, “adalah”, dan sebagainya. Penghapusan kata-kata ini dilakukan agar model LSTM dapat lebih fokus pada kata-kata utama yang benar-benar mencerminkan opini atau emosi pengguna, sekaligus mengurangi dimensi input dan meningkatkan efisiensi pelatihan model.

5) *Stemming* atau *Lemmatization*

Stemming atau Lemmatization bertujuan untuk menemukan bentuk dasar dari sebuah kata (Subowo et al., 2022), misalnya kata “pengajar”, “mengajar”, dan “pengajaran” semuanya akan distem menjadi “ajar” sehingga model dapat mengenali bahwa ketiga kata tersebut merujuk pada konsep atau topik yang sama, yang sangat berguna dalam menyatukan variasi morfologis kata yang

sering ditemukan dalam bahasa alami, terutama dalam analisis sentimen berbasis aspek yang bergantung pada keakuratan makna kata.

3.2.3 Ekstraksi dan Labeling Aspek

Pada penelitian ini, proses ekstraksi aspek dilakukan secara otomatis dengan pendekatan *Unsupervised Topic Modeling* menggunakan metode *Latent Dirichlet Allocation* (LDA). Metode *Latent Dirichlet Allocation* (LDA) merupakan sebuah algoritma probabilistik yang secara luas digunakan untuk menemukan struktur laten dalam kumpulan dokumen teks. LDA bekerja dengan cara memodelkan data ulasan pengguna platform Ruangguru sebagai kombinasi dari beberapa topik, dan setiap topik sebagai kombinasi dari sejumlah kata. Dengan kata lain, LDA berusaha menemukan pola tersembunyi dari kata-kata yang sering muncul bersama dan mengelompokkannya ke dalam kelompok-kelompok topik yang merepresentasikan aspek-aspek tertentu.

Proses pelatihan model LDA akan menghasilkan distribusi probabilitas kata terhadap topik. Hasil ini memungkinkan peneliti untuk mengidentifikasi kata-kata kunci dari setiap topik, yang kemudian diinterpretasikan sebagai aspek-aspek spesifik. Sebagai contoh, jika dalam salah satu topik muncul kata-kata dominan seperti “harga”, “langganan”, “mahal”, dan “murah”, maka topik tersebut dapat diinterpretasikan sebagai aspek harga. Topik-topik ini memberikan gambaran secara statistik tentang apa saja yang menjadi perhatian utama pengguna ketika memberikan ulasan.

Penggunaan LDA dalam penelitian ini memberikan keuntungan besar karena bersifat *unsupervised*, artinya tidak memerlukan data berlabel atau anotasi manual yang memakan waktu dan tenaga. Selain itu, LDA mampu menyesuaikan diri dengan berbagai jenis dataset dan mampu mendeteksi topik-topik baru yang mungkin belum teridentifikasi sebelumnya. Dengan demikian, metode ini sangat cocok digunakan dalam skenario di mana ulasan pengguna sangat beragam dan dinamis, seperti dalam platform edukasi online Ruangguru. Melalui ekstraksi aspek otomatis ini, proses analisis sentimen menjadi lebih fokus, karena model tidak hanya menentukan sentimen umum, tetapi juga menyasar opini terhadap aspek

tertentu yang benar-benar penting dalam meningkatkan kualitas layanan.

3.2.4 Labeling Sentimen

Labeling Sentimen adalah suatu proses untuk memberikan label sentimen pada setiap kalimat yang mengandung salah satu aspek yang mencerminkan opini pengguna yaitu:

1) Positif

   Apabila ulasan pengguna menunjukkan kepuasan atau pengalaman baik dari penggunaan aplikasi.

2) Negatif

   Apabila ulasan pengguna berisi keluhan, kritik, atau pengalaman buruk dari penggunaan aplikasi

3) Netral

   Apabila ulasan pengguna tidak mengandung ekspresi emosional yang jelas, baik positif maupun negatif terhadap suatu aspek tertentu.

Pada penelitian ini, proses pelabelan tidak dilakukan secara manual, melainkan menggunakan pendekatan *automatic labeling* berbasis *deep learning*. Pendekatan ini dipilih karena jumlah data ulasan yang besar sehingga pelabelan manual berpotensi menimbulkan inkonsistensi, bias subjektif, serta membutuhkan waktu yang lama. Untuk memperoleh perbandingan performa, penelitian ini menggunakan dua model berbasis transformer, yaitu *Multilingual Bidirectional Encoder Representations from Transformers* (mBERT) dan *Indonesian Bidirectional Encoder Representations from Transformers* (IndoBERT). Hasil prediksi sentimen dari mBERT dan IndoBERT kemudian digunakan sebagai label referensi (*ground truth*) dalam dua skenario pelatihan model LSTM yang berbeda. Dengan demikian, penelitian ini dapat menganalisis bagaimana perbedaan kualitas pelabelan otomatis memengaruhi kinerja klasifikasi sentimen berbasis LSTM.

3.2.5 Klasifikasi Sentimen

Data yang sudah dibersihkan akan dipisah menjadi data *training* dan data *testing*. Setelah itu, data *training* akan diproses melalui tahapan klasifikasi sentimen berdasarkan aspek. Model yang sudah dilatih akan diuji kepada data *testing*

sehingga nantinya model tersebut dapat memprediksi sentimen pada setiap aspek dari ulasan platform Ruangguru. Model akan membaca data kemudian melakukan klasifikasi dan juga menentukan sentimen melalui polaritasnya (Naquitasia et al., 2022). Penelitian ini menggunakan metode *Long Short-Term Memory* (LSTM) sebagai pendekatan utama dalam membangun sistem klasifikasi sentimen berbasis aspek dari ulasan pengguna terhadap platform Ruangguru. Arsitektur model LSTM dirancang secara berlapis dan modular, sehingga mampu mengakomodasi karakteristik teks dalam Bahasa Indonesia, seperti struktur kalimat yang tidak selalu eksplisit dan kosakata yang bervariasi. Proses pemodelan LSTM terdiri dari beberapa tahapan sebagai berikut:

1) Persiapan data *input*

   Pada model *Long Short-Term Memory* (LSTM) data input harus mealui beberapa tahapan karena model *Long Short-Term Memory* (LSTM) tidak dapat langsung memahami teks dalam bentuk kalimat atau kata, melainkan memerlukan input berupa angka yang dapat diproses secara matematis. Oleh karena itu dibutuhkan proses representasi numerik yang bertujuan untuk mengubah teks ke dalam bentuk vektor berdimensi tetap yang mampu mencerminkan makna semantik dari kata-kata tersebut, sehingga model dapat mengenali pola dan hubungan antar kata dalam konteks urutan. Struktur model LSTM menerima data berupa urutan token numerik. Token ini adalah hasil dari proses *tokenisasi* dan *padding*, yang sebelumnya dilakukan pada teks ulasan. Panjang input ditentukan berdasarkan rata-rata panjang kalimat dalam dataset. Selanjutnya dilakukan *word embedding* untuk mengubah token (angka) menjadi vektor kata berdimensi tetap. *Word* e*mbedding* ini dapat menggunakan pre-trained Word2Vec yang mampu mengatur posisi kata dalam ruang semantik. Misalnya, kata “guru” dan “pengajar” akan memiliki vektor yang saling dekat karena digunakan dalam konteks yang sama. Selanjutnya, vektor *embedding* akan diproses oleh LSTM layer.

2) Desain arsitektur model

   Model LSTM terdiri dari beberapa lapisan sebagai berikut:

a. *Input Layer:* Memasukkan data *input* berupa urutan kata yang sudah direpresentasi dalam bentuk numerik.
b. *Embedding Layer :* Mengubah ID kata menjadi representasi vektor (*word embedding)* dan menangkap makna semantik dari teks.
c. LSTM *Layer* : Lapisan utama yang digunakan untuk memproses data teks secara berurutan dengan mengingat informasi penting dan melupakan informasi yang tidak relevan, agar model bisa memahami makna keseluruhan kalimat.
d. *Hidden Layer*: Digunakan untuk menyimpan hasil pola *temporal* dari kata-kata sebelumnya, agar model bisa memahami makna kalimat secara menyeluruh saat memproses kata berikutnya..
e. *Dropout Layer*: Digunakan untuk mengurangi risiko *overfitting* dengan mengabaikan sejumlah koneksi secara acak selama pelatihan. Selain itu, untuk mengurangi kemungkinan overfitting, penelitian ini menggunakan fungsi *earlystopping* untuk mendefinisikan jumlah epochs untuk setiap model (Cahyaningtyas, 2021).
f. *Dense Layer* (*Output*): Memberikan hasil prediksi klasifikasi sentimen dimana pada lapisan ini digunakan fungsi aktivasi softmax dengan tiga neuron sebagai representasi dari tiga kelas sentimen yaitu positif, netral dan negatif. Fungsi softmax mengubah hasil klasifikasi menjadi *probabilitas*, di mana nilai tertinggi akan menentukan label akhir yang diprediksi.

3) *Hyperparameter Tuning*

Parameter utama yang disesuaikan meliputi:

a. *Learning Rate*: Menentukan kecepatan pembaruan bobot selama pelatihan.
b. *Epochs*: Jumlah iterasi penuh atas *dataset* untuk meningkatkan akurasi.
c. *Batch Size*: Ukuran *subset* data yang diproses dalam satu iterasi.

4) Proses Pelatihan dan Validasi

a. Proses Pelatihan

Pada tahap pelatihan, model LSTM menggunakan data *training* yang sudah diproses menjadi urutan token numerik dan vektor embedding. Selama pelatihan, bobot pada setiap gate di LSTM (*Forget Gate, Input Gate, Output*

Gate) diperbarui melalui mekanisme *backpropagation* berdasarkan error prediksi dari tahap sebelumnya. Proses ini memungkinkan model belajar mengenali pola sentimen dalam data berurutan dan mempertahankan informasi penting yang relevan dengan konteks kalimat. Pelatihan dilakukan dengan dua skenario yaitu sebagai berikut:

- Pelatihan ke Seluruh Aspek

  Data yang digunakan mencakup seluruh aspek yang ada dalam dataset, seperti harga, performa, materi, pengajar dan lain sebagainya. Kemudian model LSTM dilatih untuk mengenali pola, kemudian menghasilkan prediksi sentimen positif, netral dan negatif dari seluruh aspek secara bersamaan. Bobot pada setiap *gate* meliputi *forget gate, input gate, output gate* diperbarui melalui backpropagation berdasarkan error dari prediksi tersebut. Parameter seperti *learning rate, epochs*, dan *batch size* disesuaikan untuk mengoptimalkan performa model selama pelatihan. Pada akhir proses, model diharapkan mampu mempelajari hubungan kompleks dalam dataset.

- Pelatihan Khusus per Aspek

  Dataset dipisahkan berdasarkan aspek tertentu, misalnya aspek harga terlebih dahulu, kemudian aspek performa aplikasi, dan seterusnya. Kemudian model dilatih secara terpisah pada setiap aspek, untuk mendapatkan pemahaman yang lebih tajam terhadap fitur dari aspek tertentu. Setelah pelatihan selesai, model akan menghasilkan prediksi sentimen untuk aspek tertentu dengan akurasi yang dievaluasi secara terpisah.

b. Validasi

Setelah proses pelatihan selesai, model divalidasi dengan menggunakan data *testing*. Prediksi yang dihasilkan dari model kemudian diuji performanya menggunakan *confusion matrix*, yang mengukur berapa banyak prediksi benar dan salah untuk kategori sentimen positif dan negatif

3.2.6 Evaluasi model

Setelah model dilatih, tahapan selanjutnya ialah melakukan evaluasi hasil prediksi yang dihasilkan oleh model. Evaluasi dilakukan dengan menganalisis kinerja model dalam memprediksi label sentimen pada data uji dengan menggunakan *Confusion Matrix*. *Confusion matrix* merupakan metode yang digunakan untuk mengukur performa dari model klasifikasi dan umumnya digunakan pada analisis sentimen untuk mengukur seberapa baik model dapat mengklasifikasikan sentimen suatu teks (Rolangon et al., 2023). *Confusion matrix* dipilih sebagai metode evaluasi utama karena mampu memberikan gambaran rinci mengenai kinerja klasifikasi pada setiap kelas sentimen, serta mengungkap pola kesalahan klasifikasi yang tidak dapat ditunjukkan oleh metrik evaluasi tunggal seperti akurasi atau *loss function* (Liu, 2012). Pendekatan ini dinilai lebih sesuai untuk analisis sentimen berbasis aspek yang melibatkan klasifikasi multikelas dan distribusi data yang tidak seimbang. *Confusion matrix* terdiri dari *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN) (Naquitasia et al., 2022). TP merupakan jumlah data yang benar diklasifikasikan sebagai positif, TN merupakan jumlah data yang benar diklasifikasikan sebagai negatif, FP merupakan jumlah data yang salah diklasifikasikan sebagai positif, dan FN merupakan jumlah data yang salah diklasifikasikan sebagai negatif. Evaluasi ini memberikan gambaran seberapa baik model dapat mengenali polaritas sentimen pengguna terhadap aspek-aspek layanan dalam platform Ruangguru. Untuk perhitungan akurasi hasil evaluasi dapat menggunakan persamaan (3.1).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (3.1)$$

BAB IV
HASIL DAN PEMBAHASAN

4.1 Pengumpulan Data

Pengumpulan data dilakukan dengan teknik *web scraping* pada data ulasan pengguna platform Ruangguru yang ada di *Google Play Store*. Proses *scraping* dilaksanakan dengan memanfaatkan library *google-play-scraper* pada bahasa pemrograman Python. Library ini digunakan untuk mengekstraksi elemen-elemen data ulasan yang diberikan oleh pengguna yang telah mengunduh dan menggunakan aplikasi Ruangguru. Dari proses *scraping* data, penelitian ini berhasil mendapatkan sebanyak 311.106 data ulasan pengguna aplikasi Ruangguru. Jumlah data tersebut menunjukkan tingginya partisipasi pengguna dalam memberikan umpan balik terhadap layanan yang disediakan oleh platform Ruangguru.

Setelah proses pengumpulan data, dilakukan tahap penyaringan data berdasarkan rentang waktu untuk menjaga relevansi temporal penelitian. Data ulasan diambil untuk periode waktu lima tahun terakhir, yaitu dari 31 Oktober 2020 hingga 31 Oktober 2025. Penyaringan ini bertujuan untuk memastikan bahwa data yang dianalisis merefleksikan kondisi, kebijakan, serta fitur aplikasi Ruangguru yang relatif mutakhir, sehingga hasil analisis lebih relevan dengan konteks saat ini.

Dari proses penyaringan data didapatkan sebanyak 45.765 data ulasan. Dari keseluruhan atribut yang tersedia, penelitian ini memfokuskan pada tiga variabel utama, yaitu tanggal ulasan, rating pengguna, dan isi ulasan. Variabel tanggal digunakan untuk mengidentifikasi periode waktu pemberian ulasan, rating digunakan sebagai indikator tingkat kepuasan pengguna, sedangkan isi ulasan menjadi sumber utama data teks yang dianalisis lebih lanjut. Seluruh data yang telah diseleksi kemudian disimpan dalam format CSV. Data ini selanjutnya digunakan sebagai input pada tahap pra-pemrosesan data.

4.2 Pra-Pemrosesan Data

Sebelum dilakukan proses ekstraksi dan klasifikasi aspek, dataset ulasan pengguna aplikasi Ruangguru terlebih dahulu diproses melalui tahapan pra-

pemrosesan (*preprocessing*) data. Tahapan ini bertujuan untuk membersihkan data dari *noise*, menyeragamkan format teks, serta meningkatkan kualitas data agar dapat diproses secara optimal pada saat pemodelan topik dan LSTM. Data ulasan yang diperoleh dari *Google Play Store* umumnya bersifat tidak terstruktur dan mengandung berbagai elemen informal seperti emoji, tanda baca berlebih, kata tidak baku, serta variasi penulisan huruf besar dan kecil. Oleh karena itu, tahapan pra-pemrosesan menjadi langkah krusial untuk meningkatkan akurasi dan stabilitas model. Tahapan pra-pemrosesan yang dilakukan dalam penelitian ini ialah sebagai berikut:

1. *Case Folding*

   Case Folding merupakan proses mengubah seluruh teks ulasan menjadi huruf kecil (lowercase). Tahap ini bertujuan untuk mencegah perbedaan representasi kata akibat penggunaan huruf kapital, sehingga kata dengan makna yang sama dapat diperlakukan sebagai satu entitas yang konsisten dalam analisis. Hasil pra-pemrosesan data pada tahap *case folding* ini ditunjukkan pada Tabel 4.1

   Tabel 4.1 Hasil Pra-Pemrosesan Data pada Tahap *Case Folding*

No	*Text*	*Case folding*
1	Alhamdulillah, sangat terbantu untuk belajar ulang kalau ada materi yang tidak paham di sekolah	alhamdulillah, sangat terbantu untuk belajar ulang kalau ada materi yang tidak paham di sekolah
2	aplikasi nya udah bagus walau ngeleg sedikit tapi tolong dong yang SMK ada jurusan teknik GEOMATIKA/GEOSPASIAL	aplikasinya sudah bagus, walau sedikit nge-lag. tapi tolong dong, untuk smk supaya ada jurusan teknik geomatika/geospasial
...	…	...
45765	RUANG GURU NYA BAGUS BIKIN AKU PINTER TERIMAKASIH RUANG GURU	ruangguru nya bagus, bikin aku pintar. terima kasih, ruangguru.

2. Pembersihan Karakter Khusus dan Emoji

Pembersihan teks adalah proses yang bertujuan untuk menghilangkan elemen-elemen yang tidak relevan atau mengganggu analisis seperti tanda baca (., !, ?), angka, simbol khusus (@, #, %, dll), emoji, URL, dan karakter non-alfabetik lainnya yang sering kali muncul dalam ulasan atau komentar pengguna, khususnya di media sosial atau platform digital. Emoji dan karakter khusus umumnya bersifat ekspresif, namun dalam konteks analisis topik dan ekstraksi aspek, elemen tersebut dianggap sebagai *noise* yang dapat mengganggu hasil pemodelan. Hasil pra-pemrosesan data pada tahap pembersihan karakter khusus dan emoji ditunjukkan pada Tabel 4.2

Tabel 4.2 Hasil Data pada Tahap Pembersihan Karakter Khusus dan Emoji

No	*Text*	Pembersihan Karakter Khusus dan Emoji
1	alhamdulillah, sangat terbantu untuk belajar ulang kalau ada materi yang tidak paham di sekolah 💗💗💗	alhamdulillah sangat terbantu untuk belajar ulang kalau ada materi yang tidak paham di sekolah
2	aplikasinya sudah bagus, walau sedikit nge-lag. tapi tolong dong, untuk smk supaya ada jurusan teknik geomatika/geospasial. 🙏	aplikasi nya udah bagus walau ngeleg sedikit tapi tolong dong yang smk ada jurusan teknik geomatika geospasial
...	...	...
45765	bagus, mudah dipahami. aplikasi ini sangat bermanfaat. 👍😁🤩🤩.	bagus mudah dipahami aplikasi ini sangat bermanfaat

3. Tokenisasi

Pada tahap tokenisasi, teks ulasan dipecah menjadi unit-unit kata atau token menggunakan metode *tokenization*. Tokenisasi bertujuan untuk memisahkan teks panjang menjadi kata-kata individual, sehingga memudahkan proses

analisis linguistik dan komputasi pada tahap selanjutnya. Hasil pra-pemrosesan data pada tahap tokenisasi ditunjukkan pada Tabel 4.3

Tabel 4.3 Hasil Pra-Pemrosesan Data pada Tahap Tokenisasi

No	*Text*	*Tokenisasi*
1	ruang guru sangat keren	['ruang', 'guru', 'sangat', 'keren']
2	paket berbayarnya sangat mahal	['paket', 'berbayarnya', 'sangat', 'mahal']
...	...	...
45765	aplikasi ini sangat membantu dalam materi pembelajaran	['aplikasi', 'ini', 'sangat', 'membantu', 'dalam', 'materi', 'pembelajaran']

4. *Stopword Removal*

Tahap berikutnya adalah penghapusan *stopword*, yaitu kata-kata umum dalam bahasa Indonesia yang sering muncul tetapi tidak memiliki kontribusi signifikan terhadap makna topik, seperti kata penghubung dan kata fungsi. Penghapusan *stopword* bertujuan untuk mengurangi dimensi data dan memfokuskan analisis pada kata-kata yang lebih informatif dan relevan secara semantik. Hasil pra-pemrosesan data pada tahap *stopword removal* ditunjukkan pada Tabel 4.4

Tabel 4.4 Hasil Pra-Pemrosesan Data pada Tahap *Stopword Removal*

No	*Text*	*Stopword Removal*
1	menurutku ini aplikasi belajar terbaik karena semua pelajaran dan materi ada di ruangguru bisa latihan soal juga walaupun masih banyak video yang belum bisa terbuka karena harus berlangganan tapi aku udah terbantu banget dengan ruangguru terima kasih ya	['menurutku', 'aplikasi', 'belajar', 'terbaik', 'semua', 'pelajaran', 'materi', 'ruangguru', 'latihan', 'soal', 'walaupun', 'banyak', 'video', 'terbuka', 'berlangganan', 'udah', 'terbantu', 'banget', 'ruangguru', 'terima', 'kasih',

	ruangguru	'ruangguru']
2	materi nya lengkap dan sangat mudah di pahami harga untuk bimbel nya juga lumayan murah	['materi', 'lengkap', 'sangat', 'mudah', 'pahami', 'harga', 'bimbel', 'lumayan', 'murah']
...	...	...
45765	bagus jadi mengerti dalam pelajaran videonya juga menarik	['bagus', 'jadi', 'mengerti', 'dalam', 'pelajaran', 'videonya', 'juga', 'menarik']

5. *Stemming*

Proses ini dilakukan untuk mengubah kata menjadi bentuk dasarnya. Dalam penelitian ini, *stemming* bertujuan untuk menyatukan variasi kata yang memiliki akar makna yang sama, sehingga dapat meningkatkan konsistensi representasi kata dalam dataset dan mengurangi redundansi istilah. Hasil pra-pemrosesan data pada tahap *stemming* ditunjukkan pada Tabel 4.5

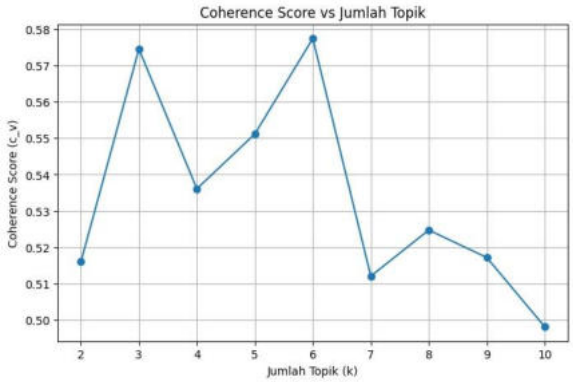
Tabel 4.5 Hasil Pra-Pemrosesan Data pada Tahap *Stemming*

No	*Text*	*Stemming*
1	bagus banget materinya juga pas kyk yg di sekolah	bagus banget materi sekolah
2	materi nya lengkap dan sangat mudah di pahami harga untuk bimbel nya juga lumayan murah	materi lengkap sangat mudah paham harga bimbel lumayan murah
...	...	...
45765	aku suka sekali dengan aplikasi ini karena setelah menggunakan aplikasi ini nilai ulangan aku menjadi bagus dan aku mendapatkan peringkat	suka aplikasi aplikasi nilai ulang bagus peringkat

4.3 Ekstraksi dan Labeling Data

Setelah data ulasan pengguna Ruangguru melalui tahap pra-pemrosesan data, tahapan selanjutnya dalam penelitian ini adalah melakukan ekstraksi dan klasifikasi aspek. Tahap ini bertujuan untuk mengidentifikasi aspek-aspek utama layanan yang menjadi fokus pembahasan pengguna dalam ulasan serta mengelompokkan setiap ulasan ke dalam aspek yang paling merepresentasikan isi ulasan tersebut. Pada tahapan ini digunakan pendekatan *topic modeling* berbasis *Latent Dirichlet Allocation* (LDA).

Latent Dirichlet Allocation (LDA) merupakan metode *topic modeling* berbasis *unsupervised learning* yang bekerja dengan mengasumsikan bahwa setiap dokumen merupakan kombinasi dari beberapa topik, dan setiap topik direpresentasikan oleh distribusi probabilitas kata tertentu. Melalui pendekatan ini, LDA mampu mengelompokkan kata-kata yang sering muncul bersama ke dalam topik-topik yang secara semantik saling berkaitan. Pemodelan topik dilakukan dengan menguji beberapa variasi jumlah topik untuk memperoleh konfigurasi model yang paling optimal. Pemilihan jumlah topik pada model *Latent Dirichlet Allocation* (LDA) dilakukan menggunakan *coherence score* karena metrik ini terbukti memiliki korelasi yang kuat dengan tingkat interpretabilitas topik oleh manusia (Röder et al., 2015). *Coherence score* mengevaluasi keterkaitan semantik antar kata dalam satu topik, sehingga lebih representatif dibandingkan metrik berbasis probabilistik. Untuk hasil evaluasi nilai *coherence score* berdasarkan variasi jumlah topik yang diuji pada pemodelan topik menggunakan metode *Latent Dirichlet Allocation* (LDA) ditunjukkan pada Gambar 4.1



Gambar 4.1 Hasil evaluasi nilai *coherence score* berdasarkan variasi jumlah topik

Penentuan jumlah topik optimal didasarkan pada nilai *coherence score* (C_v), yang digunakan untuk mengukur tingkat keterkaitan semantik antar kata dalam satu topik. Nilai *coherence* yang tinggi menunjukkan bahwa kata-kata dalam topik tersebut memiliki makna yang lebih konsisten dan mudah diinterpretasikan secara konseptual. Berdasarkan hasil pengujian iteratif terhadap beberapa jumlah topik, diperoleh bahwa model LDA dengan jumlah topik K = 6 menghasilkan nilai *coherence* tertinggi yaitu 0.579, sehingga jumlah tersebut ditetapkan sebagai jumlah topik optimal yang digunakan pada tahap ekstraksi dan klasifikasi aspek.

4.3.1 Ekstraksi Aspek

Tahap ekstraksi aspek bertujuan untuk mengidentifikasi aspek-aspek utama yang dibahas oleh pengguna Ruangguru dalam ulasan mereka. Setelah model LDA dengan enam topik terbentuk, setiap topik dianalisis lebih lanjut melalui kata-kata kunci (*keywords*) yang memiliki probabilitas kemunculan tertinggi. Untuk menjaga objektivitas akademik dalam proses penamaan aspek, penelitian ini tidak hanya mengandalkan interpretasi manual. Penelitian ini menerapkan pendekatan *Hybrid Methodology*, yaitu mengombinasikan hasil *unsupervised topic discovery* dari LDA dengan pendekatan *Zero-Shot Text Classification* berbasis model *transformer* DeBERTa.

Penamaan aspek diturunkan dari kerangka teoritis model DeLone dan Mclean (2003) yang membagi kesuskesan *Information System* (IS) menjadi 3 dimensi kualitas dan dampaknya yang meliputi kualitas sistem (*system quality*), kualitas informasi (*information quality*), kualitas layanan (*service quality)* dan manfaat bersih atau kepuasan pengguna (*net benefits*) (Petter et al., 2008). Model ini dipilih karena model kesuksesan sistem informasi yang dikemukakan oleh DeLone dan McLean berfokus pada evaluasi keberhasilan sistem informasi berdasarkan dampak dan manfaat yang dihasilkan dari penggunaan sistem (Urbach & Müller, 2012). Oleh karena itu, model ini dapat dipahami sebagai model evaluasi berbasis hasil (*outcome-based evaluation*) yang menekankan pengalaman dan persepsi pengguna setelah sistem digunakan.

Petter et al. (2008) menegaskan bahwa Model DeLone dan McLean menyediakan taksonomi aspek keberhasilan sistem informasi yang bersifat multidimensional dan berbasis evaluasi pengguna, sehingga relevan digunakan sebagai kerangka konseptual untuk menafsirkan persepsi dan pengalaman pengguna. Dalam konteks penelitian ini, kata kunci yang dihasilkan dari pemodelan topik menggunakan metode LDA merepresentasikan evaluasi dan pengalaman pengguna terhadap platform Ruangguru. Melalui pendekatan ini, pola kata kunci dari setiap topik dicocokkan ke dalam dalam kategori aspek yang telah ditetapkan sebelumnya berdasarkan kerangka teori DeLone dan McLean (2003), sehingga meminimalkan bias subjektif peneliti dan meningkatkan kredibilitas hasil ekstraksi aspek. Hasil pemetaan label ke teori model DeLone dan McLean (2003) ditunjukkan pada Tabel 4.6

Tabel 4.6 Hasil pemetaan label ke teori model DeLone dan McLean (2003)

Kerangka Teori model DeLone dan McLean (2003)	Label Aspek	Konteks di platform Ruangguru
Kualitas Informasi (*Information Quality*)	Kualitas Materi dan Media Pembelajaran	Pengguna menilai akurasi materi, kelengkapan soal, kualitas konten dan efektivitas pembelajaran yang berdampak pada pemahaman pengguna
Kualitas Sistem (*System Quality*)	Fitur dan Performa Aplikasi	Menunjukkan stabilitas sistem, performa fitur, serta pengalaman teknis pengguna
kualitas layanan (*Service Quality*)	Akses Layanan dan Harga	Ulasan berkaitan dengan pengalaman mengunduh aplikasi, mencoba layanan, serta persepsi terhadap skema gratis dan berbayar yang diterapkan Ruangguru

Dampak atau kepuasan pengguna (*Net Benefits*)	Capaian Akademik, Kepuasan Pengguna, Pendampingan Belajar	Menggambarkan peran sistem sebagai pendukung pembelajaran dan dampaknya pada motivasi serta capaian akademik pengguna serta ekspresi kepuasan pengguna

Setiap topik yang dihasilkan oleh model LDA direpresentasikan oleh serangkaian kata kunci (*keywords*) dengan probabilitas kemunculan tertinggi. Kata kunci tersebut mencerminkan tema utama pembahasan pengguna dalam kelompok ulasan tertentu. Proses interpretasi dilakukan dengan menganalisis keterkaitan semantik antar kata kunci dalam satu topik, kemudian mengaitkannya dengan konteks layanan pendidikan daring yang disediakan oleh Ruangguru. Hasil distribusi topik beserta kata kunci dominan dan label aspek yang diberikan ditunjukkan pada Tabel 4.7

Tabel 4.7 Distribusi Topik, Kata Kunci, dan Penamaan Aspek

Topik	Kata Kunci Dominan	Label Aspek
Topik 0	ruangguru, nilai, kelas, pakai, langgan, nama, moga, bahasa, sukses, matematika	Capaian Akademik
Topik 1	ajar, ruangguru, paham, mudah, materi, erti, video, senang, seru, manfaat	Kualitas Materi dan media pembelajaran
Topik 2	unduh, ruangguru, aplikasi, bayar, gratis, bintang, orang, maaf, coba, tv	Akses Layanan dan Harga
Topik 3	aplikasi, keren, video, baik, buka, jelek, fitur, roboguru, langgan, baru	Fitur dan Performa Aplikasi
Topik 4	ajar, ruangguru, bantu, aplikasi, anak, semangat, peringkat, sekolah, tambah, guna	Pendampingan Belajar
Topik 5	bagus, aplikasi, banget, suka, mantap, pintar, download, deh, pokok, juara	Kepuasan Pengguna

Berdasarkan Tabel 4.7 diperoleh enam aspek utama yang dibahas pada ulasan platform Ruangguru yaitu sebagai berikut:

1. Aspek Capaian Akademik

   Aspek Capaian Akademik (Topik 0) didominasi oleh kata kunci yaitu *ruangguru, nilai, kelas, pakai, langgan, nama, moga, bahasa, sukses, matematika*. Pola kata ini menunjukkan bahwa pengguna banyak membahas hasil belajar yang diperoleh, seperti nilai akademik, capaian pada mata pelajaran tertentu, serta keberhasilan penggunaan aplikasi Ruangguru dalam mendukung proses pembelajaran formal di kelas. Kata kunci yang merepresentasikan aspek ini ditunjukkan pada Gambar 4.2

Gambar 4.2 *Word cloud* Aspek Capaian Akademik

2. Aspek Kualitas Materi dan media pembelajaran

   Aspek Kualitas Materi dan media pembelajaran (Topik 1) memiliki kata kunci dominan antara lain *ajar, ruangguru, paham, mudah, materi, erti, video, senang, seru, manfaat*. Kata-kata tersebut merefleksikan persepsi pengguna terhadap kualitas konten pembelajaran, kemudahan memahami materi, serta efektivitas media pembelajaran seperti video dalam mendukung proses belajar. Kata kunci yang merepresentasikan aspek ini ditunjukkan pada Gambar 4.3

Gambar 4.3 *Word cloud* Aspek Kualitas Materi dan Media Pembelajaran

3. Aspek Akses Layanan dan Harga

   Aspek Akses Layanan dan Harga (Topik 2) memiliki kata kunci dominan antara lain *unduh, ruangguru, aplikasi, bayar, gratis, bintang, orang, maaf, coba, tv*. Kombinasi kata ini mengindikasikan adanya pembahasan terkait kemudahan akses layanan, proses pengunduhan aplikasi, serta persepsi pengguna terhadap skema harga, layanan gratis dan berbayar, serta pengalaman awal saat mencoba aplikasi Ruangguru. Kata kunci yang merepresentasikan aspek ini ditunjukkan pada Gambar 4.4

Gambar 4.4 *Word cloud* Aspek Akses Layanan dan Harga

4. Aspek Fitur dan Performa Aplikasi

   Aspek Fitur dan Performa Aplikasi (Topik 3) menunjukkan dominasi kata kunci *aplikasi, keren, video, baik, buka, jelek, fitur, roboguru, langgan, baru*. Kata-kata ini menggambarkan opini pengguna terkait kinerja teknis aplikasi, kualitas

fitur yang tersedia, stabilitas aplikasi saat digunakan, serta pengalaman pengguna terhadap fitur-fitur pendukung pembelajaran yang disediakan oleh Ruangguru. Kata kunci yang merepresentasikan aspek ini ditunjukkan pada Gambar 4.5

Gambar 4.5 *Word cloud* Aspek Fitur dan Performa Aplikasi

5. Aspek Pendampingan Belajar

Aspek Pendampingan Belajar (Topik 4) memiliki kata kunci dominan seperti *ajar, ruangguru, bantu, aplikasi, anak, semangat, peringkat, sekolah, tambah, guna*. Pola kata ini mencerminkan peran Ruangguru sebagai media pendamping belajar yang membantu siswa dalam memahami pelajaran, meningkatkan motivasi belajar, serta mendukung proses belajar anak baik di sekolah maupun di luar lingkungan sekolah. Kata kunci yang merepresentasikan aspek ini ditunjukkan pada Gambar 4.6

Gambar 4.6 *Word cloud* Aspek Pendampingan Belajar

6. Aspek Kepuasan Pengguna

Aspek Kepuasan Pengguna (Topik 5) memiliki kata kunci dominan antara lain *bagus, aplikasi, banget, suka, mantap, pintar, download, deh, pokok, juara*. Kata-kata tersebut merepresentasikan ekspresi kepuasan, apresiasi, dan penilaian positif pengguna terhadap pengalaman menggunakan aplikasi Ruangguru secara keseluruhan. Dengan demikian, topik ini diinterpretasikan sebagai aspek Kepuasan Pengguna. Kata kunci yang merepresentasikan aspek ini ditunjukkan pada Gambar 4.7

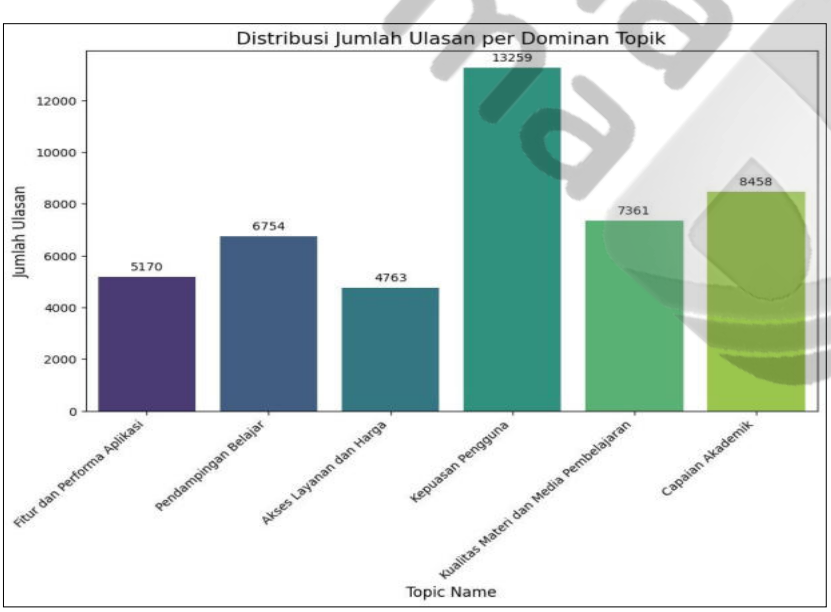
Gambar 4.7 *Word cloud* Aspek Kepuasan Pengguna

4.3.2 Labeling Aspek

Setelah proses ekstraksi aspek menggunakan metode *Latent Dirichlet Allocation* (LDA) dan validasi penamaan aspek dilakukan, tahap selanjutnya adalah labeling aspek. Labeling aspek bertujuan untuk memberikan label aspek tertentu pada setiap ulasan pengguna. Proses labeling aspek dilakukan dengan memanfaatkan distribusi probabilitas topik yang dihasilkan oleh model LDA. Setiap ulasan memiliki nilai probabilitas terhadap masing-masing topik, dan ulasan tersebut kemudian diberikan label aspek dengan nilai probabilitas tertinggi (*dominant topic*). Dengan pendekatan ini, setiap ulasan dipastikan memiliki satu aspek utama yang paling merepresentasikan isi ulasan tersebut. Labeling aspek yang digunakan pada penelitian ini ialah aspek capaian akademik, kualitas materi dan media pembelajaran, akses layanan dan harga, fitur dan performa aplikasi, pendampingan belajar, dan kepuasan pengguna. Pembagian aspek ini

mencerminkan pengalaman pengguna dalam menggunakan platform Ruangguru, baik dari sisi teknis maupun emosional.

Setelah setiap ulasan dilabeli ke dalam aspek tertentu, dilakukan analisis distribusi untuk mengetahui aspek mana yang paling banyak dibahas oleh pengguna. Analisis ini dilakukan dengan menghitung jumlah ulasan pada masing-masing aspek hasil labeling. Hasil distribusi labeling aspek menunjukkan bahwa beberapa aspek memiliki frekuensi kemunculan yang lebih tinggi dibandingkan aspek lainnya. Hal ini mengindikasikan bahwa aspek-aspek tersebut memiliki tingkat kepentingan yang lebih besar dalam pengalaman pengguna Ruangguru. Distribusi jumlah ulasan berdasarkan aspek yang dihasilkan dari proses pemodelan topik menggunakan metode *Latent Dirichlet Allocation* (LDA) ditunjukkan pada Gambar 4.8



Gambar 4.8 Jumlah ulasan platform Ruangguru berdasarkan aspek

Gambar 4.8 memperlihatkan bahwa aspek kepuasan pengguna merupakan aspek yang paling dominan dibahas oleh pengguna dengan jumlah ulasan tertinggi, diikuti oleh aspek capaian akademik serta kualitas materi dan media pembelajaran. Tingginya jumlah ulasan pada aspek-aspek tersebut mengindikasikan bahwa pengguna ruangguru cenderung memberikan perhatian besar terhadap pengalaman belajar secara keseluruhan, hasil akademik yang diperoleh, serta kualitas konten pembelajaran yang disediakan oleh platform.

Di sisi lain, aspek pendampingan belajar dan fitur dan performa aplikasi memiliki jumlah ulasan yang berada pada kategori menengah, yang menunjukkan bahwa meskipun tidak menjadi fokus utama, kedua aspek tersebut tetap memiliki peran penting dalam membentuk pengalaman belajar pengguna. Sementara itu, aspek akses layanan dan harga memiliki jumlah ulasan yang relatif lebih rendah dibandingkan aspek lainnya. Namun demikian, aspek ini tetap signifikan dalam konteks evaluasi layanan, mengingat isu terkait akses dan biaya sering kali menjadi faktor penentu dalam keputusan penggunaan layanan pendidikan berbasis digital. Dengan mengetahui aspek utama dari setiap ulasan, proses pelabelan sentimen dan pelatihan model *Long Short-Term Memory* (LSTM) dapat dilakukan secara lebih terarah dan spesifik. Pendekatan ini memungkinkan model tidak hanya mengenali polaritas sentimen secara umum, tetapi juga memahami sentimen pengguna terhadap aspek tertentu.

4.4 Labeling Sentimen

Setelah proses ekstraksi dan labeling aspek menggunakan metode *Latent Dirichlet Allocation* (LDA) dilakukan, tahap selanjutnya dalam penelitian ini adalah pelabelan sentimen pada data ulasan. Pelabelan sentimen bertujuan untuk memberikan label polaritas terhadap setiap teks ulasan yang telah dipetakan ke dalam aspek tertentu. Pada penelitian ini, proses pelabelan tidak dilakukan secara manual, melainkan menggunakan pendekatan *automatic labeling* berbasis *deep learning*. Pendekatan ini dipilih karena jumlah data ulasan yang besar sehingga pelabelan manual berpotensi menimbulkan inkonsistensi, bias subjektif, serta membutuhkan waktu yang lama. Selain itu, pelabelan otomatis memungkinkan konsistensi dan objektivitas dalam pemberian label sentimen. Untuk memperoleh perbandingan performa, penelitian ini menggunakan dua model berbasis transformer, yaitu *Multilingual Bidirectional Encoder Representations from Transformers* (mBERT) dan *Indonesian Bidirectional Encoder Representations from Transformers* (IndoBERT).

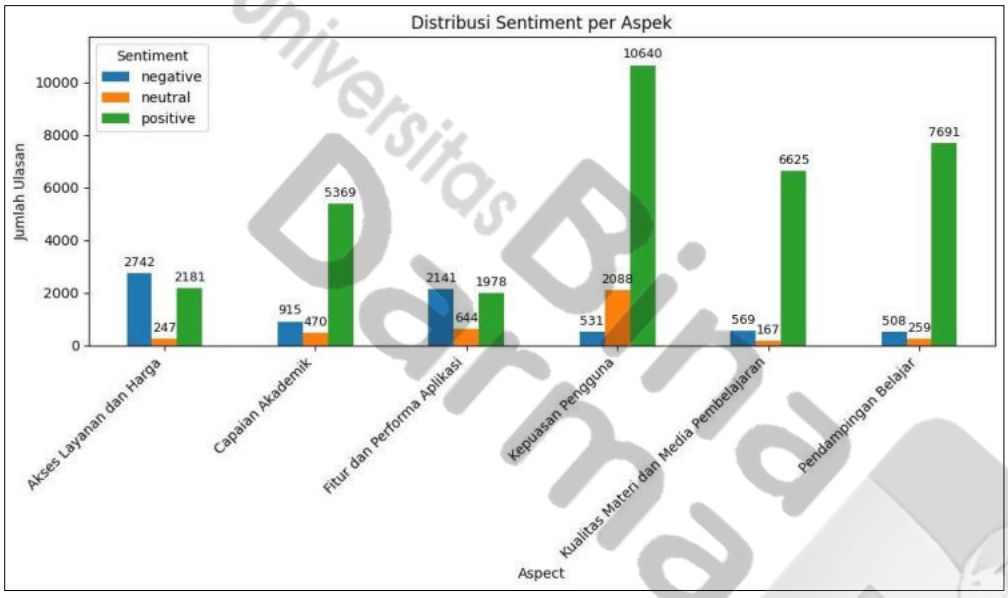
Model mBERT merupakan model BERT multilingual yang telah dilatih pada korpus teks dalam berbagai bahasa, termasuk Bahasa Indonesia. Keunggulan

mBERT terletak pada kemampuannya memahami teks multibahasa melalui mekanisme *bidirectional encoding*, yaitu proses pemahaman konteks kata berdasarkan kata sebelum dan sesudahnya secara simultan. Hal ini memungkinkan model menangkap makna sentimen secara kontekstual meskipun terdapat variasi bahasa, istilah informal, maupun struktur kalimat yang tidak baku yang umum ditemukan pada ulasan *Google Play Store*. Sementara itu, IndoBERT merupakan model bahasa berbasis transformer yang secara khusus telah dilatih menggunakan korpus teks berbahasa Indonesia. Karena dikembangkan dan pre-trained pada data Bahasa Indonesia, IndoBERT memiliki pemahaman linguistik yang lebih mendalam terhadap morfologi, struktur kalimat, serta variasi kosakata dalam Bahasa Indonesia. Karakteristik ini menjadikan IndoBERT lebih adaptif terhadap nuansa sentimen lokal dibandingkan model multilingual.

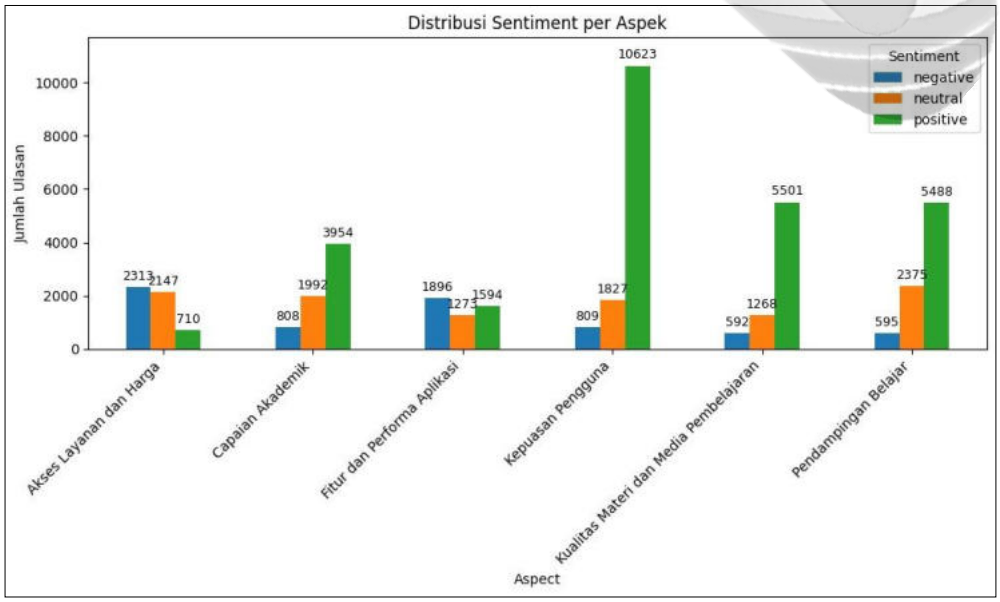
Dalam implementasinya, setiap teks ulasan yang telah melalui tahap pra-pemrosesan dan klasifikasi aspek diubah menjadi representasi numerik menggunakan *tokenizer* bawaan masing-masing model. Proses tokenisasi menghasilkan token ID dan *attention mask* yang kemudian diproses oleh model melalui mekanisme *bidirectional encoding*. Model selanjutnya menghasilkan probabilitas untuk tiga kelas sentimen, yaitu negatif, netral, dan positif. Label akhir ditentukan berdasarkan nilai probabilitas tertinggi (argmax). Hasil prediksi sentimen dari mBERT dan IndoBERT kemudian digunakan sebagai label referensi (*ground truth*) dalam dua skenario pelatihan model LSTM yang berbeda. Dengan demikian, penelitian ini dapat menganalisis bagaimana perbedaan kualitas pelabelan otomatis memengaruhi kinerja klasifikasi sentimen berbasis LSTM.

Pendekatan ini memungkinkan model LSTM untuk mempelajari pola urutan kata dan dinamika sentimen pada setiap aspek secara lebih efisien. Setelah dilatih, LSTM dapat melakukan klasifikasi sentimen pada data baru tanpa perlu menjalankan model transformer yang memiliki kompleksitas komputasi lebih tinggi. Dengan kata lain, model transformer berperan sebagai *teacher* model, sedangkan LSTM berperan sebagai model klasifikasi yang lebih ringan dalam tahap implementasi. Selain itu, untuk memperoleh gambaran awal mengenai kecenderungan sentimen pengguna terhadap setiap aspek, dilakukan analisis

distribusi sentimen berdasarkan hasil pelabelan otomatis menggunakan dua model, yaitu mBERT dan IndoBERT. Distribusi sentimen per aspek menggunakan mBERT ditunjukkan pada Gambar 4.9, sedangkan distribusi sentimen menggunakan IndoBERT ditunjukkan pada Gambar 4.10.



Gambar 4.9 Distribusi Sentimen Berdasarkan Aspek Menggunakan Metode mBERT



Gambar 4.10 Distribusi Sentimen Berdasarkan Aspek Menggunakan Metode IndoBERT

Berdasarkan Gambar 4.9 dan 4.10 dapat diamati bahwa secara umum, kedua metode menunjukkan pola yang sama, yaitu dominasi sentimen positif pada aspek

kepuasan pengguna, capaian akademik, kualitas materi dan media pembelajaran serta pendampingan belajar. Namun, pada aspek akses layanan dan harga serta fitur dan performa aplikasi, proporsi sentimen negatif relatif lebih tinggi dibandingkan aspek lainnya. Perbedaan utama terlihat pada distribusi jumlah masing-masing kelas. IndoBERT cenderung menghasilkan jumlah sentimen positif yang lebih besar pada beberapa aspek seperti Capaian Akademik dan Pendampingan Belajar, serta lebih tegas dalam mengidentifikasi sentimen negatif pada aspek tertentu. Sementara itu, mBERT menunjukkan distribusi yang relatif lebih seimbang antara kelas negatif dan netral pada beberapa aspek.

4.5 Klasifikasi Sentimen Berbasis Aspek

Pada penelitian ini, dilakukan proses klasifikasi sentimen berbasis aspek menggunakan metode *Long Short-Term Memory* (LSTM). Tahap ini bertujuan untuk mengukur kemampuan model *deep learning* dalam mengklasifikasikan polaritas sentimen ulasan pengguna Ruangguru pada masing-masing aspek yang telah diidentifikasi sebelumnya. Klasifikasi sentimen dilakukan setelah data ulasan memperoleh label sentimen hasil pelabelan otomatis dengan metode BERT, yang berperan sebagai label referensi (ground truth). Proses klasifikasi sentimen dengan menggunakan metode LSTM terdiri dari beberapa tahapan yaitu sebagai berikut:

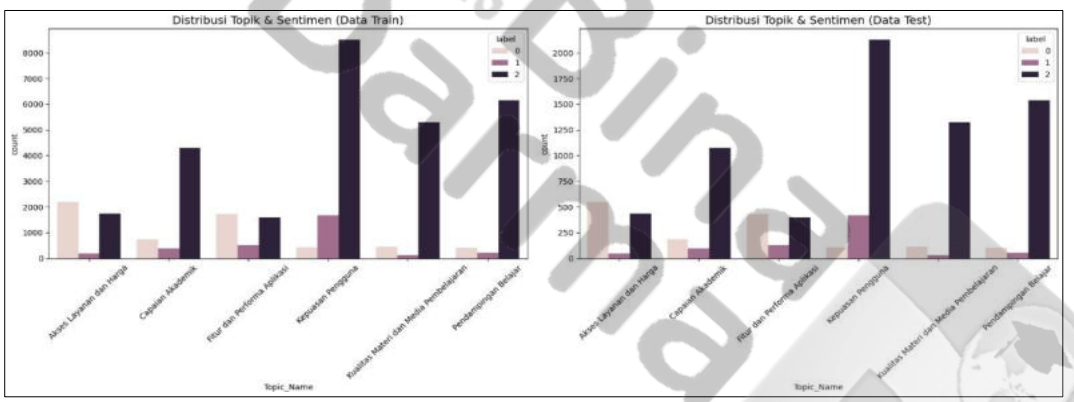
1. Persiapan Data Input

Sebelum proses pelatihan model dilakukan, label sentimen pada setiap teks ulasan dikonversi ke dalam bentuk numerik. Proses konversi ini diperlukan karena model LSTM hanya dapat memproses data dalam bentuk angka. Pada penelitian ini, sentimen negatif diberi label 0, sentimen netral diberi label 1, dan sentimen positif diberi label 2. Dengan demikian, setiap ulasan memiliki target kelas yang jelas dan dapat digunakan dalam proses pembelajaran terawasi (*supervised learning*).

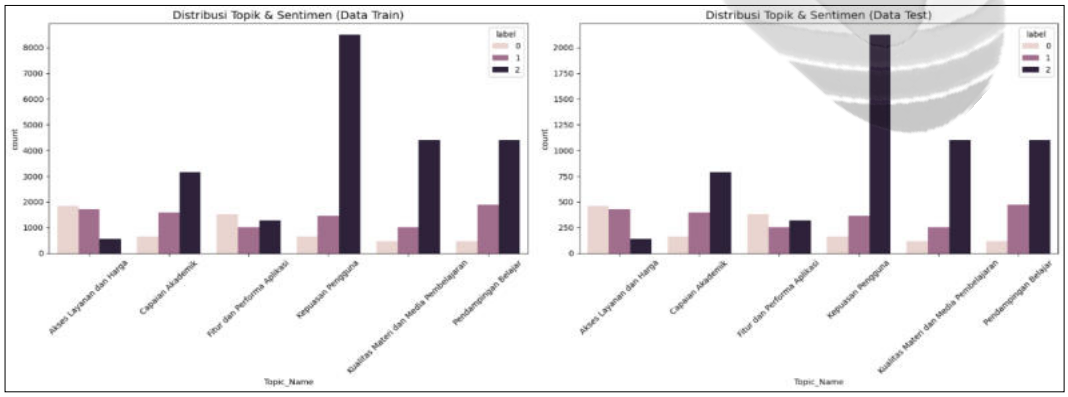
Setelah konversi label sentimen dilakukan, teks ulasan digabungkan dengan label aspek. Proses ini dilakukan dengan menyatukan aspek dan teks ulasan dalam satu representasi teks, sehingga aspek berfungsi sebagai konteks tambahan bagi model LSTM. Pendekatan ini bertujuan agar model tidak hanya mempelajari pola

sentimen dari isi ulasan, tetapi juga mempertimbangkan aspek layanan Ruangguru yang sedang dibahas, sehingga klasifikasi sentimen yang dihasilkan berbasis aspek.

Tahap berikutnya adalah pembagian data menjadi data pelatihan (*training*) dan data pengujian (*testing*). Pembagian data dilakukan dengan rasio 80:20 yaitu 80% untuk data *training* dan 20% untuk data *testing*. Pembagian data dengan rasio 80:20 menggunakan metode mBERT ditunjukkan pada Gambar 4.11, sedangkan Pembagian data dengan rasio 80:20 menggunakan metode IndoBERT ditunjukkan pada Gambar 4.12



Gambar 4.11 Data *Training* dan Data *Testing* menggunakan mBERT



Gambar 4.12 Data *Training* dan Data *Testing* menggunakan IndoBERT

Gambar 4.11 dan 4.12 mengambarkan jumlah data pada data *training* dan data *testing* yang digunakan dalam penelitian ini. Pembagian ini bertujuan untuk mengevaluasi kemampuan generalisasi model terhadap data yang belum pernah dilihat selama proses pelatihan. Secara rinci, jumlah data pada data *training* dan data *testing* dengan rasio 80:20 menggunakan metode mBERT ditunjukkan pada Tabel 4.8, sedangkan jumlah data pada data *training* dan data *testing* dengan rasio

80:20 menggunakan metode IndoBERT Tabel 4.9.

Tabel 4.8 Data *Training* dan Data *Testing* menggunakan Metode mBERT

Label	Aspek	Jumlah Data	
		Data *Training*	Data *Testing*
Negatif	Akses Layanan dan Harga	2194	548
	Capaian Akademik	732	183
	Fitur dan Performa Aplikasi	1713	428
	Kepuasan Pengguna	425	106
	Kualitas Materi dan media pembelajaran	455	114
	Pendampingan Belajar	406	102
Netral	Akses Layanan dan Harga	198	49
	Capaian Akademik	376	94
	Fitur dan Performa Aplikasi	515	129
	Kepuasan Pengguna	1670	418
	Kualitas Materi dan media pembelajaran	134	33
	Pendampingan Belajar	207	52
Positif	Akses Layanan dan Harga	1745	436
	Capaian Akademik	4295	1074
	Fitur dan Performa Aplikasi	1582	396
	Kepuasan Pengguna	8512	2128
	Kualitas Materi dan media pembelajaran	5300	1325
	Pendampingan Belajar	6153	1538

Tabel 4.9 Data *Training* dan Data *Testing* menggunakan Metode IndoBERT

Label	Aspek	Jumlah Data	
		Data *Training*	Data *Testing*
Negatif	Akses Layanan dan Harga	1850	463
	Capaian Akademik	646	162
	Fitur dan Performa Aplikasi	1517	379
	Kepuasan Pengguna	647	162
	Kualitas Materi dan media pembelajaran	474	118
	Pendampingan Belajar	476	119
Netral	Akses Layanan dan Harga	1718	429
	Capaian Akademik	1594	398
	Fitur dan Performa Aplikasi	1019	254
	Kepuasan Pengguna	1462	365
	Kualitas Materi dan media pembelajaran	1014	254
	Pendampingan Belajar	1900	475
Positif	Akses Layanan dan Harga	568	142
	Capaian Akademik	3163	791

Positif	Fitur dan Performa Aplikasi	1275	319
	Kepuasan Pengguna	8498	2125
	Kualitas Materi dan media pembelajaran	4401	1100
	Pendampingan Belajar	4390	1098

Selanjutnya dilakukan proses *tokenization*, yaitu mengubah teks menjadi representasi numerik. *Tokenizer* digunakan dengan batas kosakata maksimum sebanyak 20.000 kata dan hanya dilatih (*fit*) menggunakan data pelatihan untuk menghindari terjadinya *data leakage*. Setelah itu, teks ulasan pada data pelatihan dan pengujian diubah menjadi sekuens angka dan dilakukan *padding* dengan panjang maksimum 100 token. Proses *padding* bertujuan untuk menyeragamkan panjang sekuens agar seluruh data memiliki format input yang konsisten dan dapat diproses secara efektif oleh model LSTM.

2. Desain Arsitektur model

Model LSTM dirancang agar mampu mengenali pola bahasa yang mengklasifikasi sentimen positif, negatif, maupun netral. Model terdiri dari beberapa lapisan yaitu sebagai berikut:

a) *Input Layer*

Lapisan input menerima data teks ulasan yang telah diproses dan diubah ke dalam bentuk sekuens numerik. Berdasarkan arsitektur model, input memiliki format (None, 100) yang menunjukkan bahwa setiap ulasan direpresentasikan sebagai sekuens dengan panjang maksimum 100 token. Format ini sesuai dengan kebutuhan LSTM yang memproses data secara berurutan berdasarkan urutan kata dalam ulasan.

b) *Embedding Layer*

Lapisan *embedding* mengubah setiap token menjadi representasi vektor berdimensi 128 dimana setiap ulasan terdiri dari 100 token dan setiap token direpresentasikan sebagai vektor 128 dimensi. Lapisan ini membantu model memahami makna kata dan hubungan semantik antar kata, sehingga proses klasifikasi sentimen menjadi lebih efektif.

c) Masking (NotEqual)

Mekanisme Masking digunakan untuk mengabaikan token *padding* yang ditambahkan pada tahap pra-pemrosesan. Token *padding* tidak mengandung

informasi semantik, sehingga jika ikut diproses oleh model dapat menimbulkan bias dalam pembelajaran. Hal ini sangat penting dalam pemodelan sekuens, karena memungkinkan model fokus pada informasi relevan tanpa terganggu oleh nilai nol hasil padding.

d) LSTM *Layer*

Lapisan LSTM merupakan inti model yang berfungsi untuk mempelajari pola urutan kata dan konteks kalimat dalam ulasan. Pada arsitektur ini, keluaran LSTM menghasilkan output yang menunjukkan vektor fitur berdimensi 128 sebagai representasi konteks dari ulasan. Lapisan ini berperan penting dalam menangkap hubungan antar kata yang memengaruhi sentimen, baik dalam konteks jangka pendek maupun jangka panjang.

e) *Dropout Layer*

Lapisan *dropout* untuk mengurangi risiko *overfitting*. *Dropout* bekerja dengan menonaktifkan sebagian unit dipilih secara acak selama proses pelatihan, sehingga model tidak terlalu bergantung pada karakteristik tertentu saja dan mampu melakukan generalisasi dengan lebih baik.

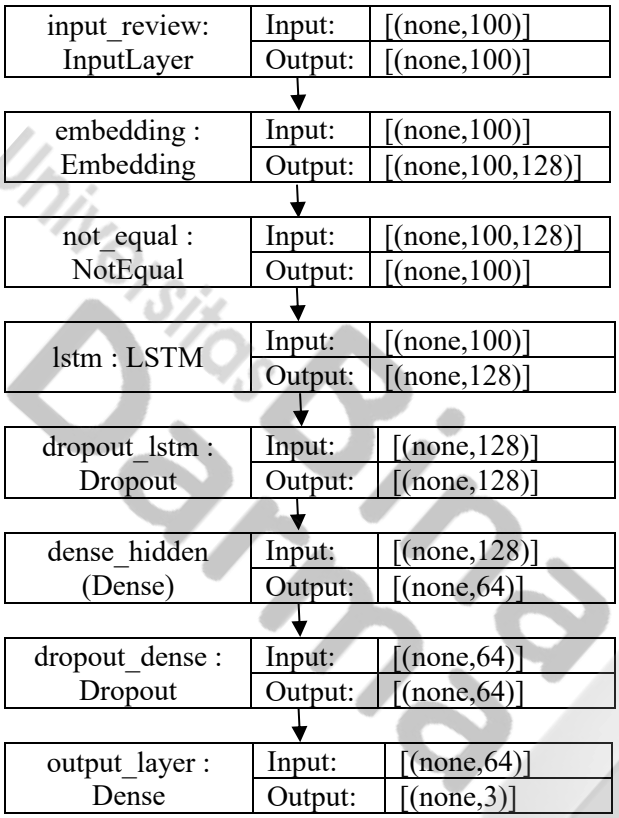
f) Dense *Hidden Layer*

Lapisan ini berfungsi sebagai lapisan perantara antara keluaran LSTM dan lapisan *output* (*softmax*). Tujuan utamanya adalah untuk mengolah dan menyaring fitur yang sudah dipelajari oleh LSTM agar klasifikasi sentimen menjadi lebih akurat.

g) *Output Layer (Dense Softmax*)

Lapisan output merupakan Dense layer dengan 3 neuron menggunakan fungsi aktivasi softmax dan menghasilkan output (None, 3). Tiga neuron tersebut merepresentasikan probabilitas prediksi terhadap tiga kelas sentimen, yaitu Negatif, Netral, dan Positif.

Arsitektur model *Long Short-Term Memory* (LSTM) yang digunakan pada penelitian ditunjukkan pada Gambar 4.11



Gambar 4.13 Arsitektur LSTM dalam Penelitian

Gambar 4.11 menunjukkan bahwa model menerima input berupa teks ulasan yang telah diproses dan ditransformasikan ke bentuk urutan angka dengan panjang maksimum 100 token melalui *InputLayer* berukuran (None, 100). Selanjutnya, data diteruskan ke lapisan *Embedding* yang mengubah setiap token menjadi vektor berdimensi 128, sehingga setiap kata direpresentasikan sebagai vektor yang mengandung informasi semantik. Setelah itu, diterapkan mekanisme *masking* atau *NotEqual* yang berfungsi untuk mengabaikan token *padding* bernilai nol, sehingga model tidak mempelajari pola dari token kosong yang ditambahkan hanya untuk menyamakan panjang input. Selanjutnya, data diproses oleh LSTM *layer* yang bertugas untuk menangkap pola sekuensial dan hubungan jangka panjang antar kata dalam kalimat. Kemudian, untuk mengurangi resiko *overfitting*, diterapkan *dropout layer* yang secara acak menonaktifkan sebagian neuron selama pelatihan. Selanjutnya, data diproses di *dense layer* yang berfungsi sebagai lapisan non-linear agar model mampu mempelajari pola non-linear untuk memperkaya representasi

sebelum proses klasifikasi akhir. Setelah itu, *dropout layer* kembali diterapkan untuk menjaga stabilitas generalisasi model. Pada tahap akhir, *output layer* dengan tiga neuron menghasilkan probabilitas untuk masing-masing kelas sentimen yaitu negatif, netral, dan positif.

3. Mekanisme Kerja LSTM

Pada lapisan LSTM, terdapat mekanisme internal yang memungkinkan model memahami konteks urutan kata dalam ulasan. Mekanisme tersebut terdiri atas tiga gerbang utama, yaitu:

a) *Forget Gate*, yang berfungsi menentukan informasi mana dari *cell state* sebelumnya yang perlu dipertahankan atau dilupakan.
b) *Input Gate*, yang berfungsi mengendalikan informasi baru yang relevan untuk dimasukkan ke dalam *cell state*.
c) *Output Gate*, yang menentukan informasi yang dikeluarkan sebagai *hidden state* pada *timestep* tertentu.

Selain tiga gerbang tersebut, terdapat dua aliran informasi penting dalam LSTM yaitu sebagai berikut:

a) *Cell State* (C_t) yang berperan sebagai jalur memori utama untuk membawa informasi jangka panjang.
b) *Hidden State* (h_t) yang membawa hasil pemrosesan pada setiap timestep dan menjadi representasi konteks yang diteruskan untuk proses berikutnya.

Kombinasi mekanisme gerbang dan aliran informasi ini membuat LSTM mampu menangkap hubungan antar kata dan konteks kalimat secara lebih baik, sehingga mendukung pengenalan pola sentimen dalam teks ulasan.

4. *Hyperparameter Tuning*

Konfigurasi pelatihan model LSTM pada penelitian ini meliputi beberapa parameter utama yaitu sebagai berikut:

a) *Epochs*

Model dilatih dengan jumlah maksimum 100 *epoch* (*epochs*=100). Namun, pelatihan dapat berhenti sebelum mencapai 100 epoch karena diterapkan mekanisme penghentian dini (*early stopping*) untuk mencegah *overfitting*.

b) *Batch Size*

Nilai *batch_size*=128 digunakan selama pelatihan, yang berarti model memproses 128 data pada setiap iterasi. Penggunaan *mini-batch* membantu meningkatkan efisiensi komputasi dan menjaga kestabilan pembaruan bobot model.

c) *Loss Function*

Fungsi kerugian (*loss function*) yang digunakan adalah *Sparse Categorical Crossentropy*. *Loss* ini sesuai untuk klasifikasi multikelas (Negatif, Netral, Positif) dengan label kelas berupa angka, sehingga mampu mengukur kesalahan prediksi model terhadap label target secara tepat.

d) *Optimizer*

Model dioptimasi menggunakan (*optimizer="adam"*). *Optimizer* ini dipilih karena mampu mempercepat proses konvergensi dan menjaga kestabilan proses pelatihan melalui pembaruan bobot yang adaptif.

e) Penanganan Ketidakseimbangan Kelas

Untuk mengatasi ketidakseimbangan distribusi data antar kelas sentimen, penelitian ini menerapkan *Class Weight*. Pendekatan ini memberikan bobot lebih tinggi pada kelas minoritas dan bobot lebih rendah pada kelas mayoritas selama pelatihan, sehingga model tidak bias terhadap kelas dominan dan mampu meningkatkan performa pada kelas yang jumlah datanya lebih sedikit.

f) *EarlyStopping*

Diterapkan *callback EarlyStopping* dengan parameter *monitor='val_accuracy'*, *patience=10*, dan *restore_best_weights=True* dimana artinya pelatihan akan dihentikan apabila akurasi validasi tidak meningkat selama 10 *epoch* berturut-turut, dan bobot terbaik selama proses pelatihan akan dipulihkan secara otomatis.

g) *ModelCheckpoint*

Callback *ModelCheckpoint* digunakan untuk menyimpan model terbaik berdasarkan nilai *val_accuracy*. Model terbaik disimpan pada file *best_model.keras*, dengan *save_best_only=True* sehingga hanya model dengan performa validasi tertinggi yang disimpan.

h) Penyesuaian *Learning Rate* (*ReduceLROnPlateau*)

Callback ReduceLROnPlateau digunakan untuk menurunkan *learning rate*

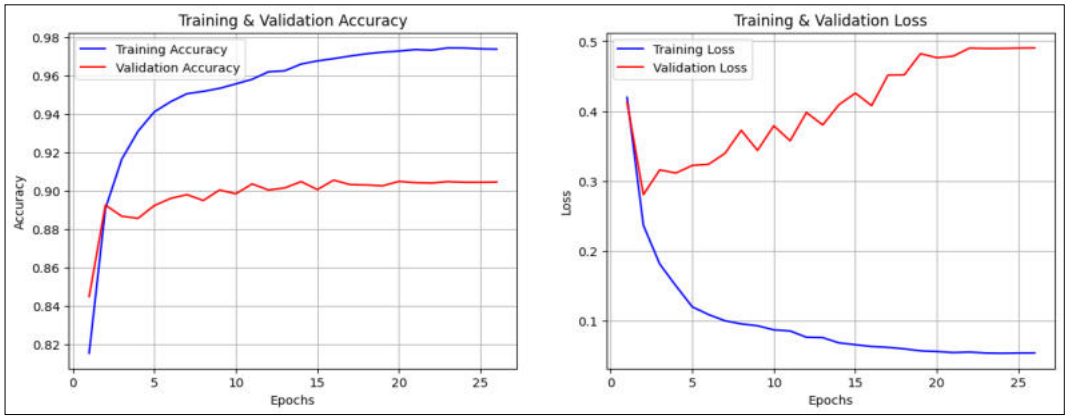
secara otomatis ketika performa model stagnan. Strategi ini membantu model tetap dapat meningkatkan performa ketika proses pelatihan mulai melambat atau tidak mengalami peningkatan.

5. Proses Pelatihan dan Validasi

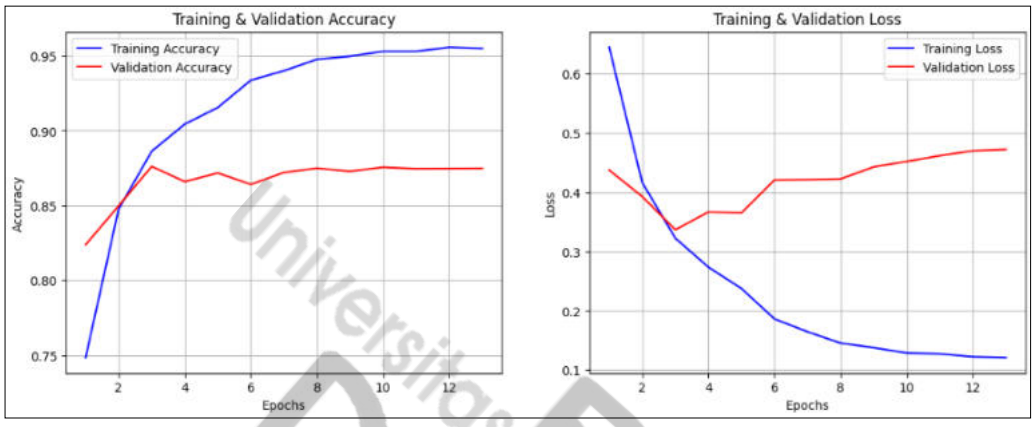
Proses pelatihan bertujuan untuk mengembangkan model klasifikasi sentimen berbasis aspek yang dapat mengidentifikasi pola sentimen negatif, netral, dan positif pada ulasan pengguna platform Ruangguru. Selama pelatihan, performa model dipantau melalui nilai akurasi dan *loss* pada data pelatihan dan data validasi untuk memastikan proses pembelajaran berlangsung secara efektif. Dalam penelitian ini, pelatihan dilakukan dengan dua skenario yaitu sebagai berikut:

a) Skenario Pertama (Pelatihan ke Seluruh Aspek)

Skenario pertama adalah pelatihan menggunakan seluruh data dari semua aspek. Pada skenario ini, satu model LSTM dilatih untuk mempelajari pola sentimen secara umum tanpa membedakan karakteristik masing-masing aspek. Skenario ini bertujuan untuk melihat kemampuan model dalam melakukan klasifikasi sentimen secara global. Selama proses pelatihan, performa model dipantau melalui nilai akurasi dan *loss* pada data pelatihan dan data pengujian. Grafik akurasi dan loss pada data pelatihan dan pengujian model LSTM dengan pelabelan otomatis menggunakan mBERT pada skenario pertama ditunjukkan pada Gambar 4.14, sedangkan Grafik akurasi dan loss pada data pelatihan dan pengujian model LSTM dengan pelabelan otomatis menggunakan IndoBERT pada skenario pertama ditunjukkan pada Gambar 4.15



Gambar 4.14 Grafik Akurasi dan Loss Model LSTM pada Skenario 1 dengan Pelabelan Otomatis mBERT



Gambar 4.15 Grafik Akurasi dan Loss Model LSTM pada Skenario 1 dengan Pelabelan Otomatis IndoBERT

Gambar 4.14 menunjukkan grafik akurasi pada data pelatihan dan pengujian yang menggambarkan akurasi model LSTM pada data pelatihan meningkat secara konsisten seiring bertambahnya epoch hingga mencapai kondisi stabil. Akurasi validasi juga mengalami peningkatan pada fase awal pelatihan dan kemudian cenderung stabil dengan fluktuasi yang relatif kecil. Perbedaan antara akurasi pelatihan dan validasi menunjukkan adanya kecenderungan overfitting ringan, namun stabilnya akurasi validasi mengindikasikan bahwa model tetap memiliki kemampuan generalisasi yang baik. Pada grafik loss, nilai loss pada data pelatihan menurun secara konsisten, sementara loss validasi meningkat setelah beberapa epoch. Pola ini mengindikasikan kecenderungan overfitting pada fase akhir pelatihan, sehingga mekanisme earlystopping diterapkan untuk menghentikan.

Gambar 4.15 menunjukkan peningkatan kinerja yang konsisten pada data pelatihan seiring bertambahnya *epoch*. Nilai akurasi pelatihan meningkat secara bertahap, sementara nilai *training loss* mengalami penurunan signifikan. Hal ini mengindikasikan bahwa model berhasil mempelajari pola pada data pelatihan dengan baik. Namun demikian, pola yang berbeda terlihat pada data pengujian. Akurasi validasi meningkat pada *epoch* awal, kemudian cenderung stagnan tanpa peningkatan yang berarti. Pada saat yang sama, *validation loss* sempat menurun pada awal pelatihan, tetapi kemudian mengalami peningkatan bertahap pada *epoch* selanjutnya. Hal ini menunjukkan adanya kecenderungan *overfitting*, di mana model semakin baik dalam mempelajari data pelatihan, tetapi kemampuan

generalisasinya terhadap data pengujian tidak mengalami peningkatan yang sebanding. Pola ini mengindikasikan kecenderungan *overfitting* pada fase akhir pelatihan, sehingga mekanisme *earlystopping* diterapkan untuk menghentikan proses pelatihan pada *epoch* optimal sebelum penurunan kemampuan generalisasi model terjadi.

Setelah proses pelatihan selesai, model LSTM dievaluasi pada seluruh aspek menggunakan data pengujian (*testing set*). *Classification report* menggunakan label hasil pelabelan otomatis mBERT ditunjukkan pada Gambar 4.13, sedangkan *classification report* menggunakan label hasil pelabelan otomatis IndoBERT ditunjukkan pada Gambar 4.14.

	precision	recall	f1-score	support
NEGATIF	0.78	0.83	0.80	1481
NETRAL	0.71	0.88	0.79	775
POSITIF	0.96	0.93	0.94	6897
acuracy			0.91	9153
macro avg	0.82	0.88	0.84	9153
weighted avg	0.91	0.91	0.91	9153

Gambar 4. 16 *Classification Report* Evaluasi Seluruh Aspek Skenario 1 Menggunakan Pelabelan Otomatis mBERT

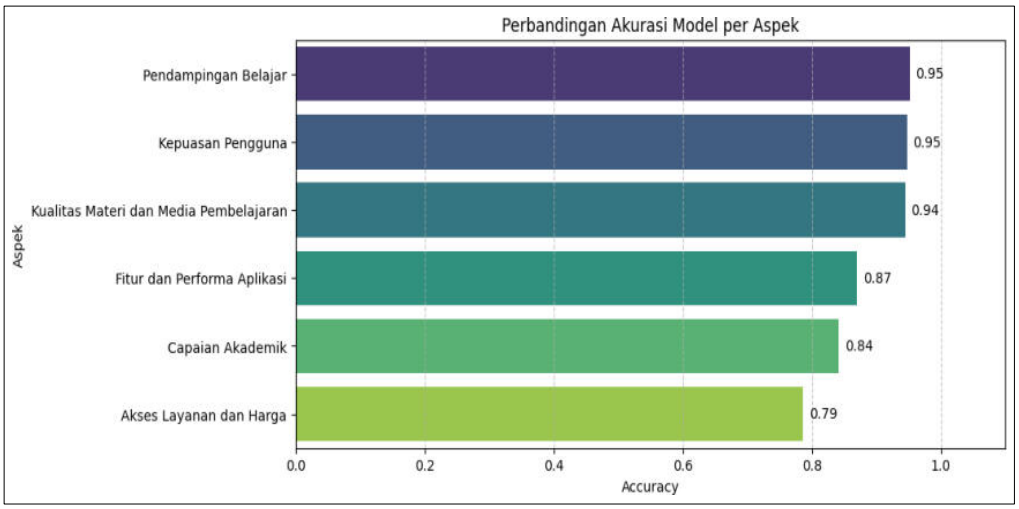
	precision	recall	f1-score	support
NEGATIF	0.77	0.77	0.77	1403
NETRAL	0.77	0.83	0.80	2175
POSITIF	0.95	0.93	0.94	5575
acuracy			0.88	9153
macro avg	0.83	0.84	0.84	9153
weighted avg	0.88	0.88	0.88	9153

Gambar 4.17 *Classification Report* Evaluasi Seluruh Aspek Skenario 1 Menggunakan Pelabelan Otomatis IndoBERT

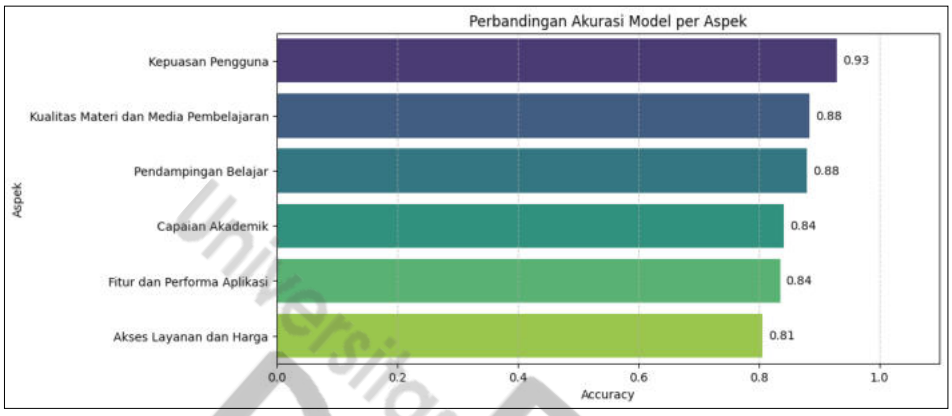
Gambar 4.16 menunjukkan hasil kinerja model Long Short-Term Memory (LSTM) pada data pengujian, yang mencakup nilai precision, recall, dan f1-score untuk masing-masing kelas sentimen, yaitu negatif, netral, dan positif. Hasil tersebut menunjukkan bahwa model memiliki performa yang baik dalam

mengklasifikasikan sentimen, dengan tingkat akurasi keseluruhan sebesar 0,91. Nilai weighted average yang tinggi mengindikasikan bahwa model mampu melakukan klasifikasi secara konsisten meskipun terdapat perbedaan jumlah data pada setiap kelas sentimen, sedangkan Gambar 4.17 menunjukkan hasil kinerja model *Long Short-Term Memory* (LSTM) pada data pengujian, yang mencakup nilai *precision*, *recall*, dan *f1-score* untuk masing-masing kelas sentimen, yaitu negatif, netral, dan positif. Hasil tersebut menunjukkan bahwa model memiliki performa yang baik dalam mengklasifikasikan sentimen, dengan tingkat akurasi keseluruhan sebesar 0,88. Nilai *weighted average* yang tinggi mengindikasikan bahwa model mampu melakukan klasifikasi secara konsisten meskipun terdapat perbedaan jumlah data pada setiap kelas sentimen.

Selain evaluasi pada seluruh aspek, dilakukan pula evaluasi kinerja model berdasarkan masing-masing aspek untuk mengetahui konsistensi performa model pada konteks aspek yang berbeda. Hasil evaluasi per aspek menunjukkan adanya variasi performa antar aspek yang dipengaruhi oleh karakteristik bahasa dan distribusi data pada setiap aspek. Perbandingan tingkat akurasi model berdasarkan masing-masing aspek menggunakan label hasil pelabelan otomatis mBERT ditunjukkan pada Gambar 4.18, sedangkan perbandingan tingkat akurasi model berdasarkan masing-masing aspek menggunakan label hasil pelabelan otomatis IndoBERT ditunjukkan pada Gambar 4.19



Gambar 4.18 Perbandingan Tingkat Akurasi Model Setiap Aspek Pada Skenario 1 Menggunakan Pelabelan Otomatis mBERT

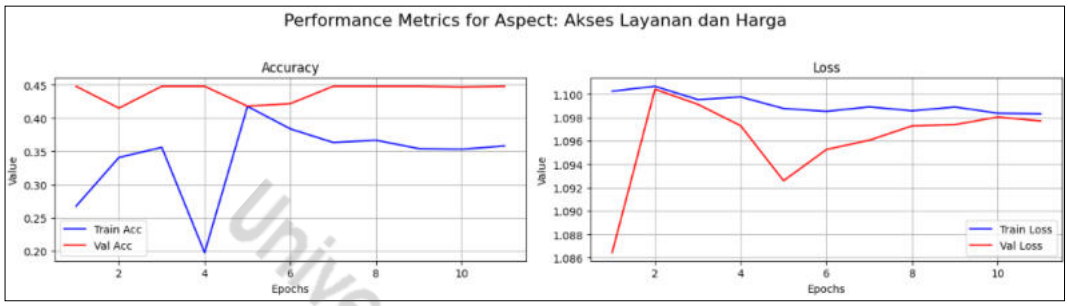


Gambar 4.19 Perbandingan Tingkat Akurasi Model Setiap Aspek pada Skenario 1 Menggunakan Pelabelan Otomatis IndoBERT

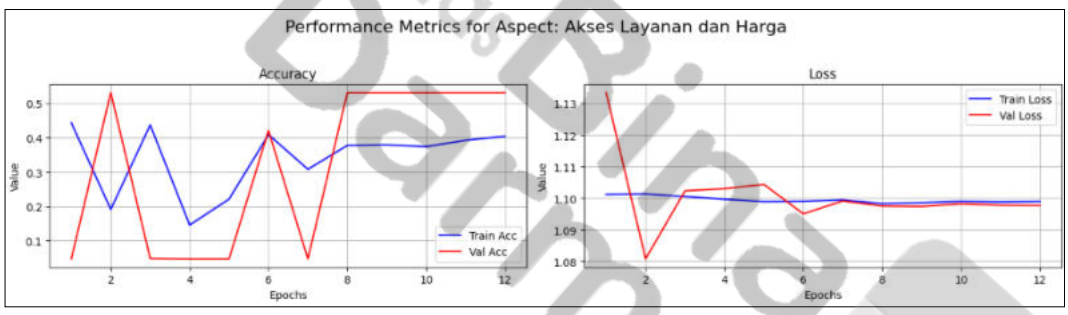
Gambar 4.18 dan 4.19 menunjukkan bahwa model lstm dengan pelabelan otomatis mbert menunjukkan akurasi tertinggi pada aspek pendampingan belajar dan kepuasan pengguna (0,95), sedangkan akurasi terendah terdapat pada aspek akses layanan dan harga (0,79). Sementara itu, menggunakan pelabelan indobert, akurasi tertinggi diperoleh pada aspek kepuasan pengguna (0,93) dan terendah pada aspek akses layanan dan harga (0,81). Secara umum, mbert menghasilkan akurasi yang sedikit lebih tinggi pada beberapa aspek utama, namun kedua metode menunjukkan pola yang konsisten, yaitu performa terbaik pada aspek kepuasan dan pendampingan belajar, serta performa relatif lebih rendah pada aspek akses layanan dan harga.

b) Skenario Kedua (Pelatihan khusus per Aspek)

Skenario kedua adalah pelatihan khusus untuk setiap aspek. Pada skenario ini, model LSTM dilatih secara terpisah untuk masing-masing aspek menggunakan data yang relevan. Pendekatan ini bertujuan untuk memungkinkan model mempelajari pola sentimen yang lebih spesifik sesuai dengan konteks tiap aspek yang ada di ulasan platform Ruangguru. Selama proses pelatihan, performa model dipantau melalui nilai akurasi dan *loss* pada data pelatihan dan data validasi. Grafik akurasi dan *loss* pelatihan dan pengujian model LSTM pada skenario kedua untuk aspek akses layanan dan harga dengan pelabelan otomatis mBERT ditunjukkan pada Gambar 4.20, sedangkan Grafik akurasi dan *loss* pelatihan dan pengujian model LSTM pada skenario kedua untuk aspek akses layanan dan harga dengan pelabelan otomatis IndoBERT ditunjukkan pada Gambar 4.21



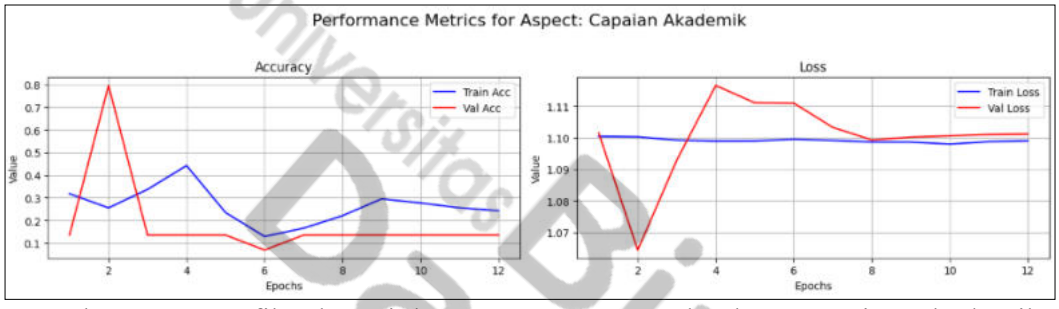
Gambar 4.20 Grafik Akurasi dan *Loss* LSTM Aspek Akses Layanan dan Harga (mBERT)



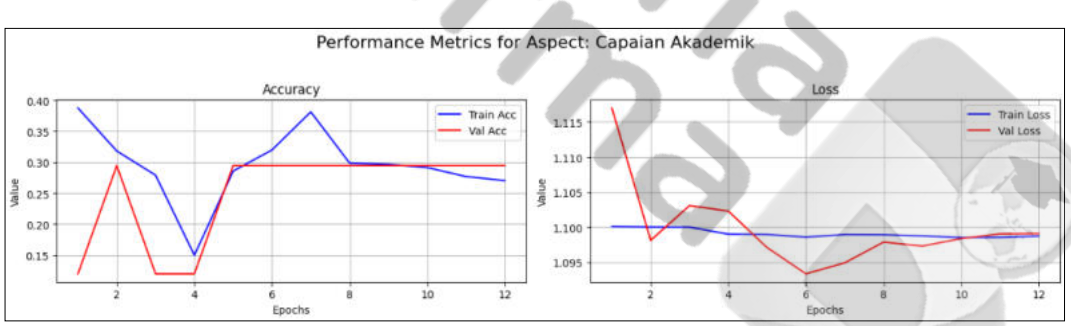
Gambar 4.21 Grafik Akurasi dan *Loss* LSTM Aspek Akses Layanan dan Harga (IndoBERT)

Gambar 4.20 menunjukkan fluktuasi akurasi yang cukup tinggi pada beberapa epoch awal, baik pada data pelatihan maupun validasi, yang menandakan model masih menyesuaikan bobot terhadap karakteristik data pada aspek tersebut. Seiring bertambahnya epoch, akurasi menjadi lebih stabil dan meningkat secara bertahap. Nilai loss pelatihan dan validasi relatif stabil setelah fase awal, sehingga tidak menunjukkan indikasi overfitting yang signifikan. Fluktuasi yang terjadi dipengaruhi oleh jumlah data pada aspek *Akses Layanan dan Harga* yang relatif lebih kecil dan tidak merata, sehingga model lebih sensitif terhadap perubahan bobot pada awal pelatihan. Sementara itu, Gambar 4.21 menunjukkan akurasi pelatihan berfluktuasi pada kisaran 0,20–0,42, sedangkan akurasi validasi relatif stabil di kisaran 0,43–0,45 setelah beberapa epoch awal. Nilai training loss dan validation loss juga berada pada kisaran yang hampir sama tanpa penurunan signifikan. Kondisi ini menunjukkan bahwa model belum mampu mempelajari pola sentimen secara optimal pada aspek tersebut, yang kemungkinan dipengaruhi oleh jumlah data yang terbatas atau distribusi sentimen yang tidak seimbang. Selanjutnya, Grafik akurasi dan *loss* pelatihan dan pengujian model LSTM pada

skenario kedua untuk aspek capaian akademik dengan pelabelan otomatis mBERT ditunjukkan pada Gambar 4.22, sedangkan Grafik akurasi dan *loss* pelatihan dan pengujian model LSTM pada skenario kedua untuk aspek capaian akademik dengan pelabelan otomatis IndoBERT ditunjukkan pada Gambar 4.23



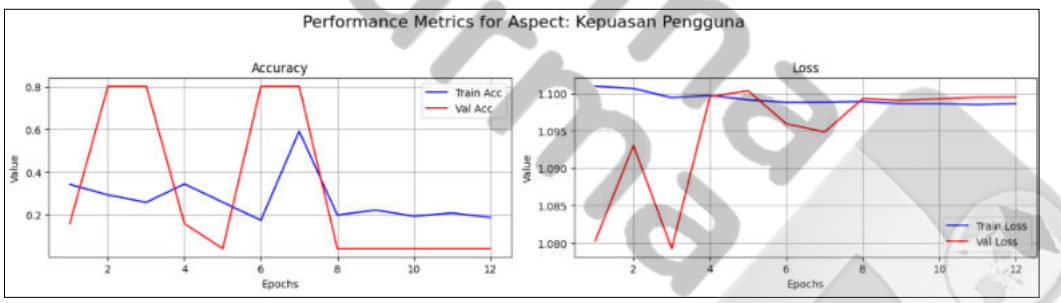
Gambar 4. 22 Grafik Akurasi dan Loss LSTM Aspek Akses Capaian Akademik (mBERT)



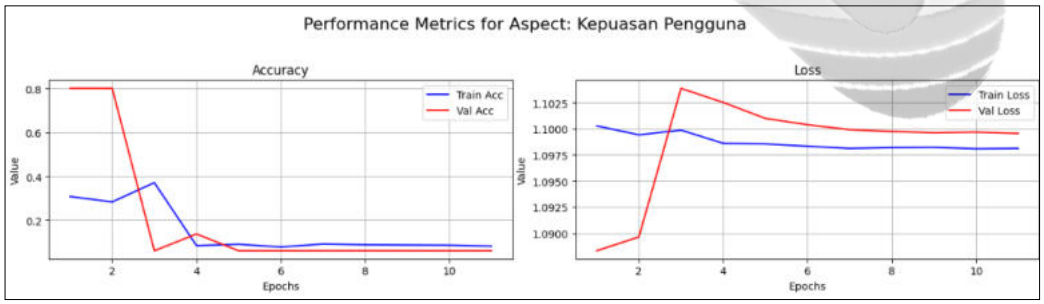
Gambar 4.23 Grafik Akurasi dan *Loss* LSTM Aspek Akses Capaian Akademik (IndoBERT)

Gambar 4.22 menunjukkan fluktuasi akurasi yang cukup tinggi pada *epoch* awal, dengan akurasi validasi mengalami peningkatan yang signifikan pada tahap awal pelatihan sebelum menurun dan stabil pada nilai yang lebih rendah. Pola ini menunjukkan model mengalami kesulitan dalam mempelajari pola sentimen secara konsisten pada aspek capaian akademik. Nilai *loss* pelatihan relatif stabil, sementara *loss* validasi berfluktuasi pada fase awal sebelum akhirnya mencapai kondisi konvergen. Secara umum, model tidak mengalami *overfitting* yang signifikan, namun performanya masih belum optimal dalam menangkap pola sentimen pada aspek tersebut. Sementara itu, Gambar 4.23 menunjukkan akurasi pelatihan berfluktuasi pada awal epoch, mencapai puncak pada epoch ke-7, kemudian menurun dan stabil di kisaran 0,27–0,30. Akurasi validasi relatif stabil dan berada pada kisaran yang hampir sama dengan pelatihan, sehingga tidak menunjukkan

overfitting yang signifikan meskipun tingkat akurasi masih rendah. Nilai training dan validation loss juga berdekatan tanpa penurunan berarti, yang mengindikasikan model belum mampu mempelajari pola sentimen secara optimal, kemungkinan karena keterbatasan data atau kompleksitas variasi bahasa pada aspek capaian akademik. Kemudian, Grafik akurasi dan *loss* pelatihan dan pengujian model LSTM pada skenario kedua untuk aspek kepuasan pengguna dengan pelabelan otomatis mBERT ditunjukkan pada Gambar 4.24, sedangkan Grafik akurasi dan *loss* pelatihan dan pengujian model LSTM pada skenario kedua untuk aspek kepuasan pengguna dengan pelabelan otomatis IndoBERT ditunjukkan pada Gambar 4.25



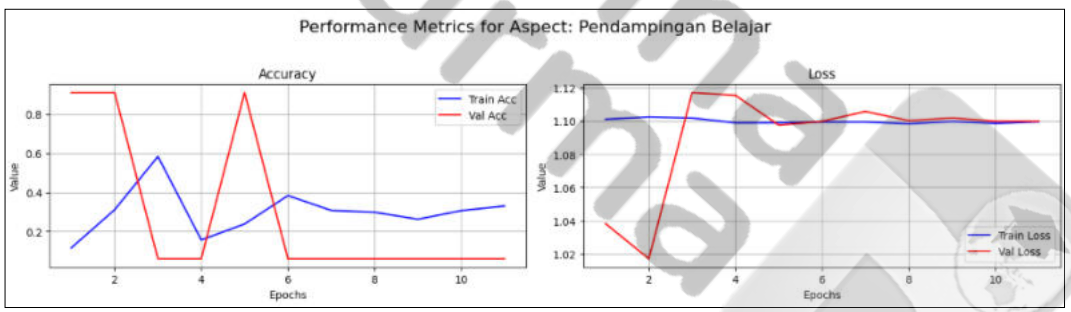
Gambar 4.24 Grafik Akurasi dan *Loss* LSTM Aspek Kepuasan Pengguna (mBERT)



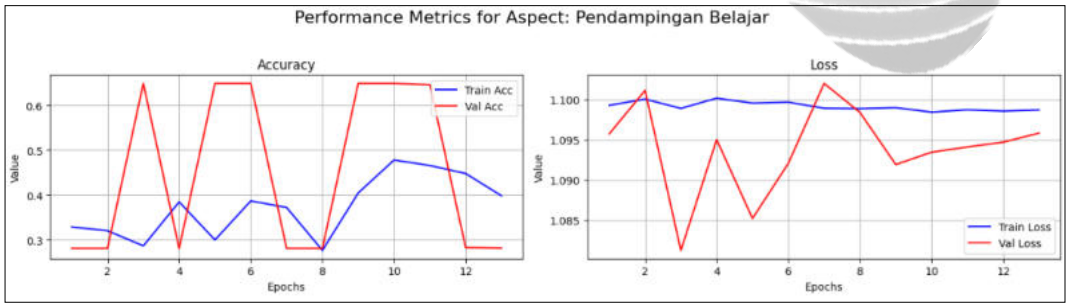
Gambar 4.25 Grafik Akurasi dan *Loss* LSTM Aspek Akses Kepuasan Pengguna (IndoBERT)

Gambar 4.24 menunjukkan menunjukkan fluktuasi nilai akurasi yang cukup tinggi pada beberapa epoch awal, terutama pada akurasi validasi. Kondisi ini mengindikasikan bahwa model mengalami ketidakstabilan pembelajaran pada fase awal pelatihan. Pada grafik loss, nilai loss pelatihan dan validasi relatif stabil setelah beberapa epoch yang menunjukkan bahwa model mencapai kondisi konvergen, sedangkan Gambar 4.25 menunjukkan akurasi pelatihan berada pada kisaran rendah dan menurun hingga stabil di sekitar 0,07–0,10, sementara akurasi validasi

berfluktuasi tajam pada awal epoch sebelum stabil pada nilai yang rendah. Nilai training loss dan validation loss relatif berdekatan dan stabil tanpa penurunan signifikan. Kondisi ini mengindikasikan bahwa model belum mampu mempelajari pola sentimen secara optimal dan tidak menunjukkan peningkatan performa yang berarti pada aspek tersebut. Selanjutnya, Grafik akurasi dan *loss* pelatihan dan pengujian model LSTM pada skenario kedua untuk aspek pendampingan belajar dengan pelabelan otomatis mBERT ditunjukkan pada Gambar 4.26, sedangkan Grafik akurasi dan *loss* pelatihan dan pengujian model LSTM pada skenario kedua untuk aspek pendampingan belajar dengan pelabelan otomatis IndoBERT ditunjukkan pada Gambar 4.27



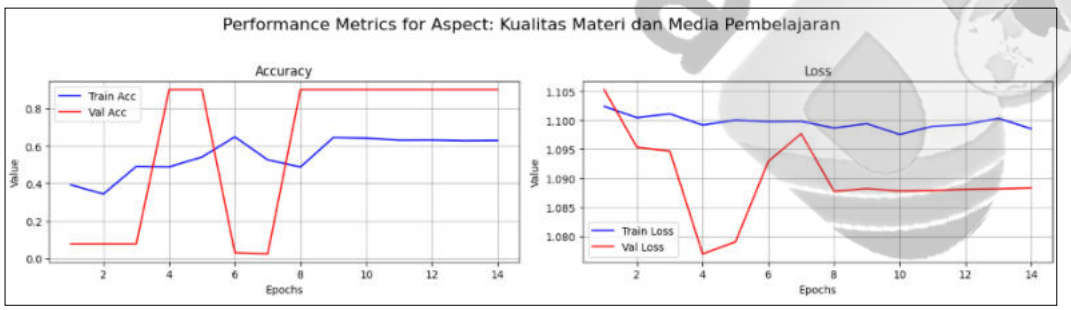
Gambar 4.26 Grafik Akurasi dan Loss LSTM Aspek Pendampingan Belajar (mBERT)



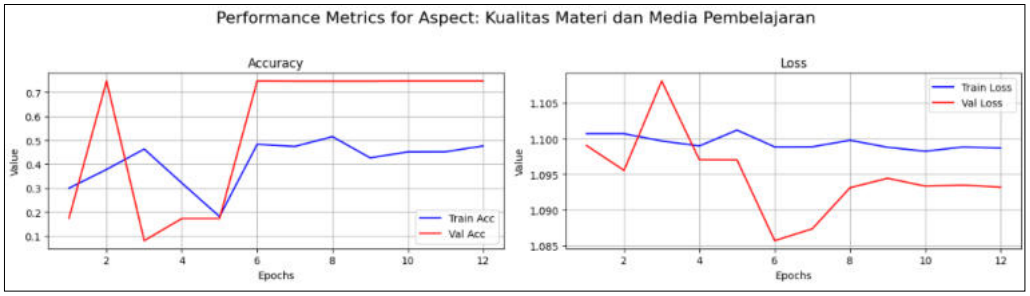
Gambar 4.27 Grafik Akurasi dan *Loss* LSTM Aspek Pendampingan Belajar (IndoBERT)

Gambar 4.26 menunjukkan fluktuasi tajam pada epoch awal, terutama pada akurasi validasi yang berubah ekstrem sebelum menurun dan stagnan pada nilai rendah. Hal ini menunjukkan model kesulitan mempelajari pola sentimen secara konsisten pada aspek pendampingan belajar, kemungkinan akibat keterbatasan dan ketidakseimbangan data. Nilai *loss* pelatihan dan validasi relatif stabil setelah fase awal dan tidak berbeda signifikan, menandakan model telah konvergen, namun

tanpa diikuti peningkatan akurasi yang optimal. Sementara itu, Gambar 4.27 menunjukkan akurasi pelatihan berfluktuasi pada awal epoch, kemudian meningkat hingga kisaran 0,45–0,48 sebelum sedikit menurun. Akurasi validasi menunjukkan fluktuasi yang lebih tajam dan tidak stabil, meskipun pada beberapa epoch mencapai nilai lebih tinggi dari pelatihan. Nilai training loss relatif stabil, sedangkan validation loss berfluktuasi pada fase awal sebelum mendekati nilai training loss. Pola ini menunjukkan adanya ketidakstabilan pembelajaran yang kemungkinan dipengaruhi oleh variasi atau ukuran dataset pada aspek tersebut. Kemudian, Grafik akurasi dan *loss* pelatihan dan pengujian model LSTM pada skenario kedua untuk aspek Kualitas Materi dan Media Pembelajaran dengan pelabelan otomatis mBERT ditunjukkan pada Gambar 4.28, sedangkan Grafik akurasi dan *loss* pelatihan dan pengujian model LSTM pada skenario kedua untuk aspek pendampingan belajar dengan pelabelan otomatis IndoBERT ditunjukkan pada Gambar 4.29



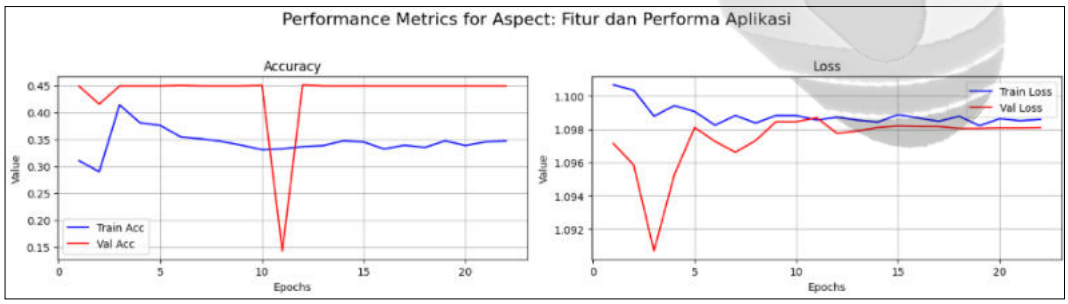
Gambar 4.28 Grafik Akurasi dan Loss LSTM Aspek Kualitas Materi dan Media Pembelajaran (mBERT)



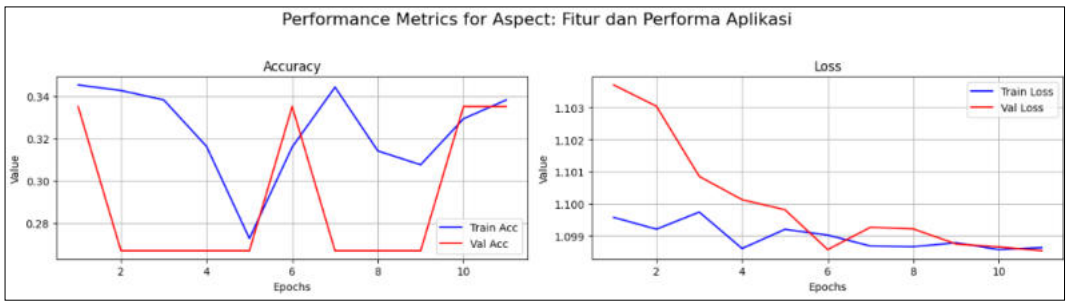
Gambar 4.29 Grafik Akurasi dan *Loss* LSTM Aspek Kualitas Materi dan Media Pembelajaran (IndoBERT)

Gambar 4.28 menunjukkan peningkatan nilai akurasi pelatihan secara bertahap, sementara akurasi validasi mengalami fluktuasi pada beberapa *epoch*

awal sebelum mencapai kondisi stabil. Pada grafik *loss*, nilai *loss* pelatihan dan validasi relatif stabil setelah fase awal pelatihan. Pola ini menunjukkan bahwa model mampu mempelajari pola sentimen pada aspek ini dengan cukup baik, meskipun stabilitas pembelajaran dipengaruhi oleh jumlah data yang terbatas. Sementara itu, Gambar 4.29 menunjukkan akurasi pelatihan berfluktuasi pada awal epoch, kemudian meningkat ke kisaran 0,45–0,52 dan stabil. Akurasi validasi mengalami fluktuasi pada tahap awal, kemudian setelah pertengahan epoch menjadi stabil dan sedikit lebih tinggi dari pelatihan, yang menunjukkan kemampuan generalisasi yang cukup baik. Pada grafik *loss*, nilai *loss* pelatihan relatif stabil, sedangkan loss validasi berfluktuasi di awal sebelum mendekati training loss, tanpa indikasi overfitting yang signifikan. Selanjutnya, Grafik akurasi dan *loss* pelatihan dan pengujian model LSTM pada skenario kedua untuk aspek Fitur dan Performa Aplikasi dengan pelabelan otomatis mBERT ditunjukkan pada Gambar 4.30, sedangkan Grafik akurasi dan *loss* pelatihan dan pengujian model LSTM pada skenario kedua untuk aspek Fitur dan Performa Aplikasi dengan pelabelan otomatis IndoBERT ditunjukkan pada Gambar 4.31



Gambar 4.30 Grafik Akurasi dan *Loss* LSTM Aspek Fitur dan Performa Aplikasi (mBERT)



Gambar 4.31 Grafik Akurasi dan *Loss* LSTM Aspek Fitur dan Performa Aplikasi (IndoBERT)

Gambar 4.20 menunjukkan nilai akurasi yang relatif rendah dan cenderung stagnan sepanjang proses pelatihan. Akurasi validasi mengalami fluktuasi tajam pada beberapa epoch awal sebelum kembali stabil pada nilai yang tidak jauh berbeda dari akurasi pelatihan. Pola ini mengindikasikan bahwa model LSTM belum mampu mempelajari pola sentimen secara optimal pada aspek fitur dan performa aplikasi. Nilai loss pelatihan dan validasi menunjukkan kecenderungan stabil setelah beberapa epoch awal dan tidak memperlihatkan perbedaan yang signifikan. Kondisi ini menunjukkan bahwa model telah mencapai konvergensi, namun rendahnya nilai akurasi mengindikasikan keterbatasan kemampuan model dalam membedakan kelas sentimen secara akurat pada aspek ini, sedangkan Gambar 4.31 menunjukkan akurasi pelatihan berada pada kisaran 0,27–0,34 dengan fluktuasi sepanjang epoch, sementara akurasi validasi juga tidak stabil dan mengalami penurunan pada beberapa epoch sebelum meningkat kembali di akhir pelatihan. Nilai training *loss* dan validation *loss* menurun secara bertahap lalu stabil dengan selisih yang kecil. Kondisi ini menunjukkan model tidak mengalami overfitting yang signifikan, namun kemampuan klasifikasinya masih terbatas.

Secara keseluruhan, fluktuasi pada grafik akurasi dan *loss* selama proses pelatihan dan validasi model LSTM terutama disebabkan oleh jumlah dataset yang relatif kecil dan tidak merata pada setiap aspek, serta keragaman karakteristik bahasa pada ulasan pengguna. Keterbatasan jumlah data menyebabkan model menjadi lebih sensitif terhadap perubahan bobot pada setiap *epoch*, sehingga nilai akurasi dan *loss* cenderung berfluktuasi, khususnya pada fase awal pelatihan. Selain itu, perbedaan kompleksitas konteks sentimen antar aspek, seperti adanya sentimen campuran dan ekspresi bahasa yang ambigu, turut memengaruhi stabilitas proses pembelajaran model.

Setelah proses pelatihan pada setiap aspek selesai, dilakukan evaluasi model gabungan seluruh aspek. *Classification report* menggunakan label hasil pelabelan otomatis mBERT ditunjukkan pada Gambar 4.31, sedangkan *classification report* menggunakan label hasil pelabelan otomatis IndoBERT ditunjukkan pada Gambar 4.33.

	precision	recall	f1-score	support
NEGATIF	0.49	0.66	0.56	1481
NETRAL	0.00	0.00	0.00	775
POSITIF	0.85	0.88	0.86	6897
acuracy			0.77	9153
macro avg	0.45	0.51	0.48	9153
weighted avg	0.72	0.77	0.74	9153

Gambar 4.32 Classification Report Evaluasi Seluruh Aspek Skenario 2 Menggunakan Pelabelan Otomatis mBERT

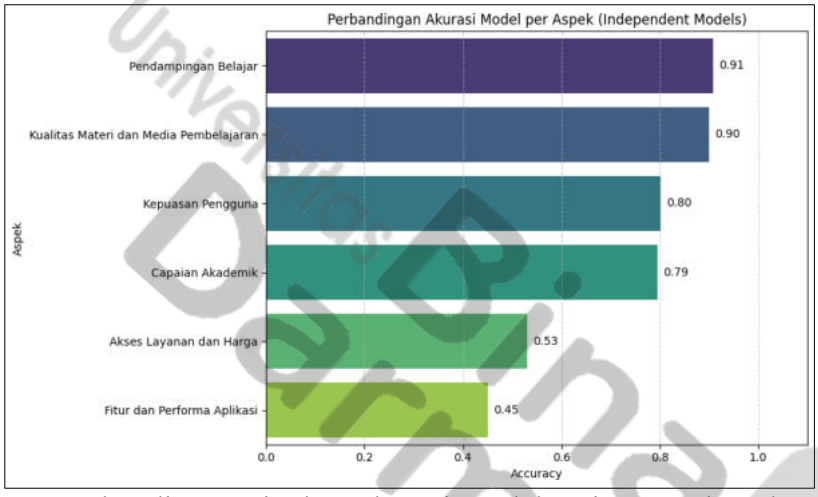
	precision	recall	f1-score	support
NEGATIF	0.45	0.33	0.38	1403
NETRAL	0.29	0.18	0.23	2175
POSITIF	0.69	0.83	0.75	5575
acuracy			0.60	9153
macro avg	0.48	0.45	0.45	9153
weighted avg	0.56	0.60	0.57	9153

Gambar 4.33 *Classification Report* Evaluasi Seluruh Aspek Skenario 2 Menggunakan Pelabelan Otomatis IndoBERT

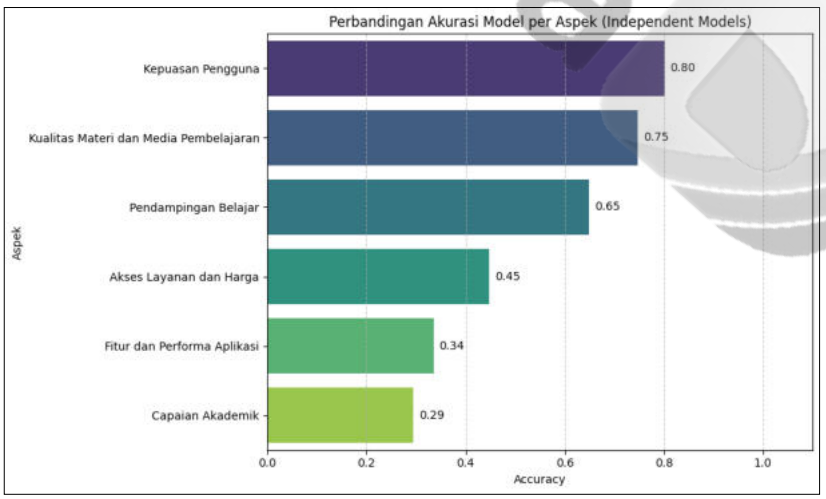
Gambar 4.32 menunjukkan hasil kinerja model *Long Short-Term Memory* (LSTM) yang mencakup nilai *precision*, *recall*, dan *f1-score* untuk masing-masing kelas sentimen, yaitu negatif, netral, dan positif. Hasil tersebut menunjukkan tingkat akurasi keseluruhan sebesar 0,77, sedangkan Gambar 4.33 menunjukkan hasil kinerja model *Long Short-Term Memory* (LSTM) yang mencakup nilai *precision*, *recall*, dan *f1-score* untuk masing-masing kelas sentimen, yaitu negatif, netral, dan positif. Hasil tersebut menunjukkan tingkat akurasi keseluruhan sebesar 0,60.

Selain evaluasi model pada seluruh aspek, dilakukan pula evaluasi kinerja model berdasarkan masing-masing aspek model LSTM yang dilatih secara terpisah. Evaluasi ini bertujuan untuk menilai kinerja model dalam mengklasifikasikan sentimen pada setiap aspek secara spesifik. Perbandingan tingkat akurasi model berdasarkan masing-masing aspek menggunakan label hasil pelabelan otomatis

mBERT ditunjukkan pada Gambar 4.34, sedangkan perbandingan tingkat akurasi model berdasarkan masing-masing aspek menggunakan label hasil pelabelan otomatis IndoBERT ditunjukkan pada Gambar 4.35



Gambar 4.34 Perbandingan Tingkat Akurasi Model Setiap Aspek Pada Skenario 2 Menggunakan Pelabelan Otomatis mBERT



Gambar 4.35 Perbandingan Tingkat Akurasi Model Setiap Aspek Pada Skenario 2 Menggunakan Pelabelan Otomatis IndoBERT

Gambar 4.34 dan 4.35 menunjukkan bahwa pada skenario 2, model lstm dengan pelabelan otomatis mBERT menunjukkan akurasi tertinggi pada aspek pendampingan belajar (0,91) dan kualitas materi dan media pembelajaran (0,90), sedangkan akurasi terendah terdapat pada aspek fitur dan performa aplikasi (0,45). Sementara itu, menggunakan pelabelan IndoBERT, akurasi tertinggi diperoleh pada aspek kepuasan pengguna (0,80), diikuti kualitas materi dan media pembelajaran

(0,75), dan akurasi terendah pada aspek capaian akademik (0,29).

Secara umum, pelabelan mBERT menghasilkan akurasi yang lebih tinggi pada sebagian besar aspek dibandingkan IndoBERT. Namun, kedua metode menunjukkan pola yang konsisten, yaitu performa relatif lebih baik pada aspek kualitas layanan dan kepuasan, serta lebih rendah pada aspek fitur dan performa aplikasi, akses layanan dan harga, serta capaian akademik. Rendahnya akurasi pada ketiga aspek ini mengindikasikan bahwa model mengalami kesulitan dalam membedakan sentimen secara akurat. Hal ini dipengaruhi oleh beberapa faktor, antara lain jumlah dataset yang relatif lebih sedikit, distribusi data yang tidak seimbang antar kelas sentimen, serta karakteristik ulasan yang cenderung ambigu dan mengandung sentimen ganda, khususnya pada pembahasan harga dan performa teknis aplikasi.

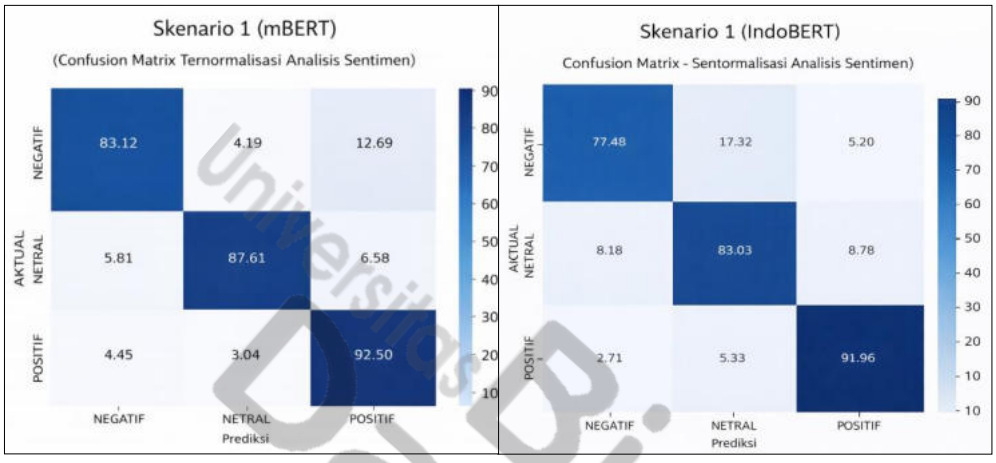
4.6 Evaluasi Model

Evaluasi model dilakukan menggunakan *confusion matrix* untuk menganalisis secara lebih mendalam pola kesalahan klasifikasi yang dihasilkan oleh model *Long Short-Term Memory* (LSTM). Berbeda dengan validasi kinerja model yang menekankan pada nilai akurasi dan metrik evaluasi secara ringkas, evaluasi menggunakan *confusion matrix* bertujuan untuk menunjukkan distribusi prediksi benar dan salah pada setiap kelas sentimen, yaitu negatif, netral, dan positif baik dalam bentuk nilai persentase (ternormalisasi). Dengan demikian, *confusion matrix* memberikan gambaran yang lebih komprehensif mengenai perilaku model dalam mengklasifikasikan sentimen. Evaluasi model ini dilakukan pada dua skenario pelatihan, yaitu sebagai berikut:

1. Skenario pertama

Skenario pertama ialah evaluasi pada model yang dilatih menggunakan data ulasan dari seluruh aspek secara bersamaan. Hasil evaluasi ini memberikan gambaran umum kemampuan model dalam mengklasifikasikan sentimen negatif, netral, dan positif secara keseluruhan, tanpa membedakan aspek secara spesifik. *Confusion matrix* ternormalisasi berdasarkan hasil evaluasi skenario pertama menggunakan data seluruh aspek dengan pelabelan otomatis mBERT dan

IndoBERT ditunjukkan pada Gambar 4.23

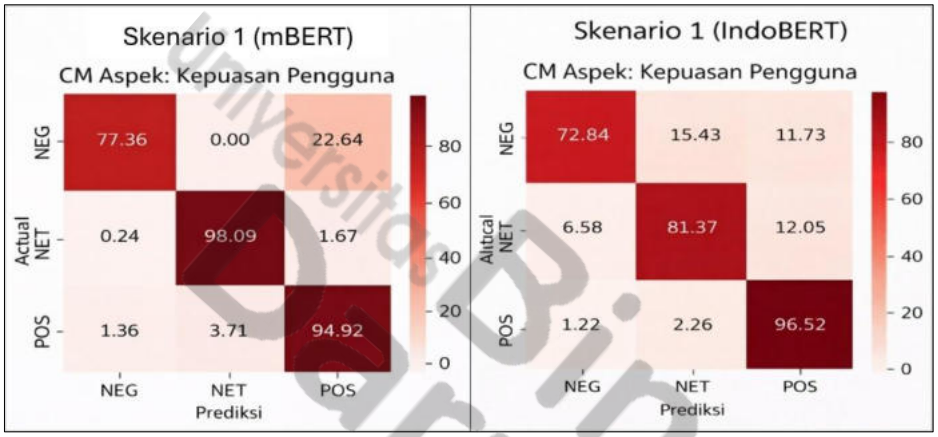


Gambar 4.36 *Confusion Matrix* Skenario 1 pada Seluruh Aspek

Gambar 4.36 menunjukkan pada pelabelan menggunakan mBERT, tingkat ketepatan tertinggi terdapat pada kelas positif (92,50%), diikuti oleh kelas netral (87,61%) dan negatif (83,12%). Kesalahan klasifikasi paling banyak terjadi antara kelas negatif dan positif (12,69%), yang menunjukkan adanya ambiguitas dalam membedakan kedua kelas tersebut pada sebagian data. Sementara itu, pada pelabelan menggunakan IndoBERT, tingkat ketepatan tertinggi juga terdapat pada kelas positif (91,96%), diikuti oleh netral (83,03%) dan negatif (77,48%). Dibandingkan dengan mBERT, nilai akurasi kelas negatif pada IndoBERT lebih rendah dan menunjukkan peningkatan kesalahan klasifikasi ke kelas netral (17,32%).

Secara keseluruhan, kedua metode menghasilkan pola klasifikasi yang serupa dengan performa terbaik pada kelas positif. Namun, pelabelan mBERT menunjukkan distribusi prediksi yang sedikit lebih stabil pada kelas negatif dan netral dibandingkan IndoBERT pada skenario pertama. Perbedaan ini menunjukkan bahwa karakteristik label hasil pelabelan otomatis berpengaruh terhadap pola kesalahan klasifikasi model LSTM. Selain evaluasi model LSTM menggunakan seluruh aspek secara bersamaan, dilakukan pula evaluasi kinerja model LSTM berdasarkan masing-masing aspek untuk mengetahui konsistensi performa model dalam konteks aspek yang berbeda. Evaluasi ini dilakukan menggunakan *confusion matrix* ternormalisasi pada setiap aspek, sehingga memberikan gambaran rinci mengenai distribusi prediksi benar dan kesalahan klasifikasi pada setiap kelas

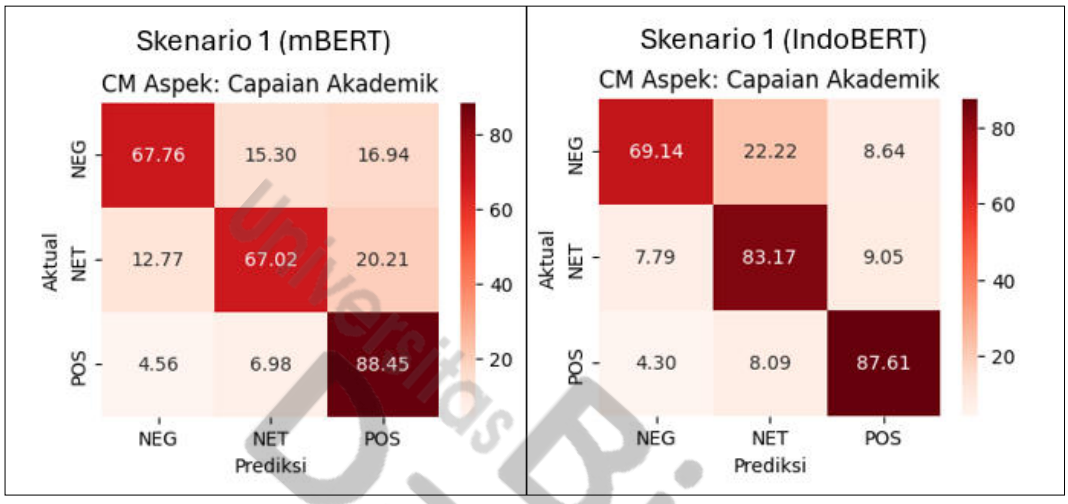
sentimen. *Confusion matrix* ternormalisasi skenario pertama pada aspek kepuasan pengguna menggunakan pelabelan otomatis mBERT dan IndoBERT ditunjukkan pada Gambar 4.37



Gambar 4.37 *Confusion Matrix* Skenario 1 pada Aspek Kepuasan Pengguna

Gambar 4.37 menujukkan pada pelabelan menggunakan mBERT, tingkat ketepatan tertinggi terdapat pada kelas positif (95,85%), diikuti oleh kelas negatif (83,33%) dan netral (75,76%). Kesalahan klasifikasi paling terlihat pada kelas netral yang masih cukup sering diprediksi sebagai negatif (15,15%) dan positif (9,09%). Sementara itu, pada pelabelan menggunakan IndoBERT, akurasi tertinggi juga terdapat pada kelas positif (92,64%), diikuti oleh kelas netral (79,13%) dan negatif (68,64%). Dibandingkan dengan mBERT, IndoBERT menunjukkan tingkat kesalahan yang lebih besar pada kelas negatif yang diprediksi sebagai netral (18,64%) serta pada kelas netral yang diprediksi sebagai positif (15,35%).

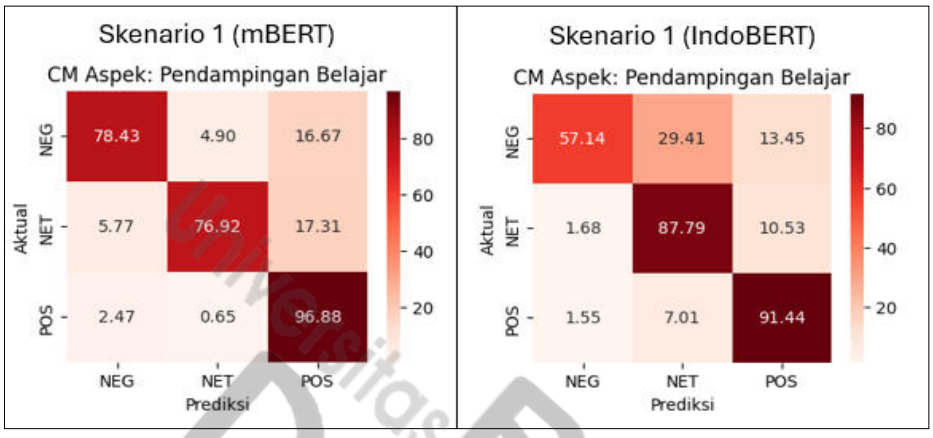
Secara keseluruhan, kedua metode menunjukkan kemampuan klasifikasi yang baik pada kelas positif, namun pelabelan mBERT menghasilkan tingkat ketepatan yang relatif lebih tinggi dan distribusi kesalahan yang lebih kecil pada aspek ini dibandingkan IndoBERT. Hal ini menunjukkan bahwa pada aspek kualitas materi dan media pembelajaran, label hasil mBERT memberikan pembelajaran yang lebih stabil bagi model LSTM pada Skenario 1. Selanjutnya, *Confusion matrix* ternormalisasi skenario pertama pada aspek capaian akademik menggunakan pelabelan otomatis mBERT dan IndoBERT ditunjukkan pada Gambar 4.38



Gambar 4.38 *Confusion Matrix* Skenario 1 pada Aspek Capaian Akademik

Gambar 4.38 menujukkan bahwa pada pelabelan menggunakan mBERT, tingkat ketepatan tertinggi terdapat pada kelas positif (88,45%), sedangkan kelas negatif dan netral memiliki tingkat akurasi yang relatif lebih rendah, masing-masing sebesar 67,76% dan 67,02%. Kesalahan klasifikasi cukup terlihat pada kelas netral yang diprediksi sebagai positif (20,21%) serta kelas negatif yang diprediksi sebagai positif (16,94%). Hal ini menunjukkan adanya kecenderungan model untuk mengarahkan sebagian data ke kelas positif. Sementara itu, pada pelabelan menggunakan IndoBERT, performa klasifikasi menunjukkan peningkatan pada kelas netral (83,17%) dibandingkan mbert. Kelas positif juga tetap tinggi (87,61%), sedangkan kelas negatif berada pada 69,14%. Meskipun demikian, masih terdapat kesalahan klasifikasi pada kelas negatif yang diprediksi sebagai netral (22,22%) dan pada kelas netral yang diprediksi sebagai positif (9,05%).

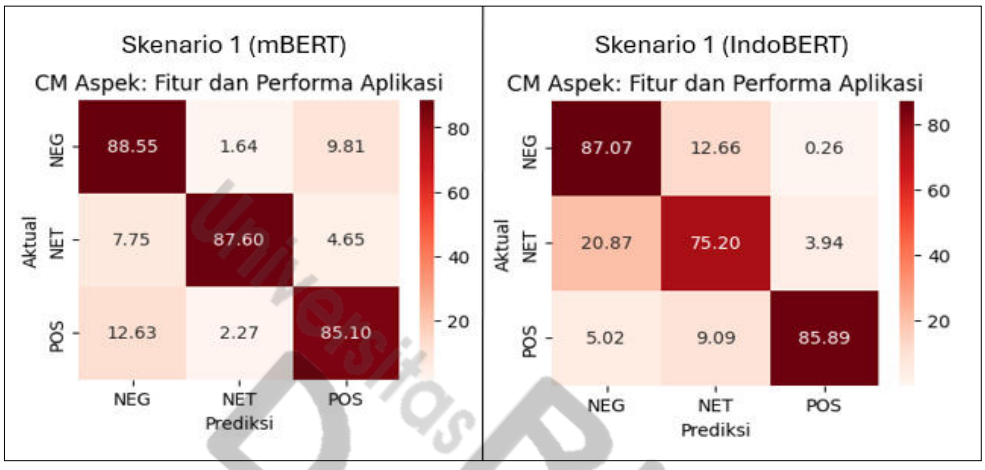
Secara keseluruhan, pelabelan IndoBERT menghasilkan distribusi klasifikasi yang lebih seimbang dan peningkatan akurasi pada kelas netral dibandingkan mBERT pada aspek capaian akademik. namun, kedua metode tetap menunjukkan tantangan dalam membedakan kelas negatif dan netral secara konsisten, yang mengindikasikan kompleksitas variasi sentimen pada aspek tersebut. Kemudian, *Confusion matrix* ternormalisasi skenario pertama pada aspek pendampingan belajar menggunakan pelabelan otomatis mBERT dan IndoBERT ditunjukkan pada Gambar 4.39



Gambar 4.39 *Confusion Matrix* Skenario 1 pada Aspek Pendampingan Belajar

Gambar 4.39 menujukkan bahwa pada pelabelan menggunakan mBERT, tingkat ketepatan tertinggi terdapat pada kelas positif (96,88%), diikuti oleh kelas negatif (78,43%) dan netral (76,92%). Kesalahan klasifikasi terlihat pada kelas negatif dan netral yang cukup sering diprediksi sebagai positif, masing-masing sebesar 16,67% dan 17,31%. Hal ini menunjukkan kecenderungan model untuk mengklasifikasikan sebagian data ke dalam kelas positif. Sementara itu, pada pelabelan menggunakan IndoBERT, kelas positif juga menunjukkan akurasi tinggi (91,44%) dan kelas netral meningkat menjadi 87,79%. Namun, akurasi kelas negatif relatif lebih rendah (57,14%), dengan tingkat kesalahan yang cukup besar pada kelas negatif yang diprediksi sebagai netral (29,41%). Pola ini menunjukkan bahwa indobert lebih sensitif dalam membedakan sentimen netral, tetapi masih mengalami kesulitan dalam memisahkan kelas negatif secara konsisten.

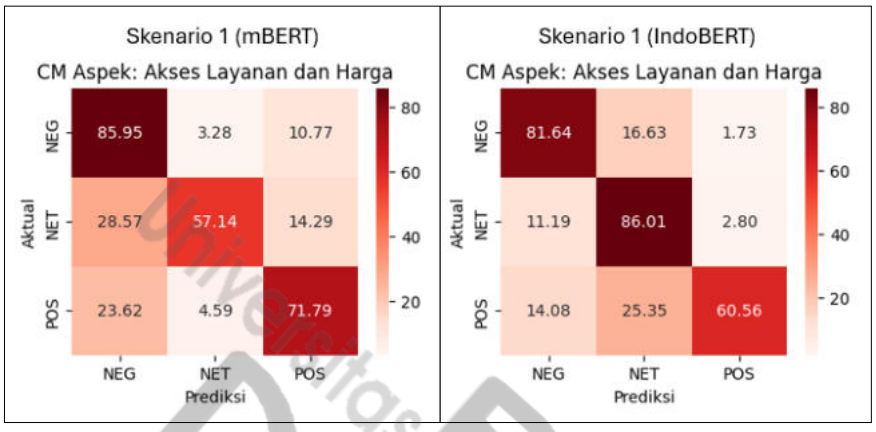
Secara keseluruhan, pelabelan mBERT menunjukkan distribusi klasifikasi yang lebih stabil pada kelas negatif dan netral dibandingkan IndoBERT pada aspek pendampingan belajar. Namun, kedua metode tetap menunjukkan performa terbaik pada kelas positif, yang mengindikasikan dominasi sentimen positif pada aspek tersebut. Selanjutnya, *Confusion matrix* ternormalisasi skenario pertama pada aspek fitur dan performa aplikasi menggunakan pelabelan otomatis mBERT dan IndoBERT ditunjukkan pada Gambar 4.40



Gambar 4.40 *Confusion Matrix* Skenario 1 pada Aspek Fitur dan Performa Aplikasi

Gambar 4.40 menujukkan bahwa pada pelabelan menggunakan mBERT, tingkat ketepatan tertinggi terdapat pada kelas negatif (88,55%) dan netral (87,60%), sedangkan kelas positif berada pada 85,10%. Kesalahan klasifikasi relatif kecil dan tersebar, meskipun masih terlihat kecenderungan sebagian data positif diprediksi sebagai negatif (12,63%). Secara keseluruhan, distribusi kesalahan pada mBERT relatif seimbang. Sementara itu, pada pelabelan menggunakan IndoBERT, kelas negatif menunjukkan akurasi yang tinggi (87,07%) dan kelas positif juga cukup baik (85,89%). Namun, kelas netral memiliki tingkat akurasi yang lebih rendah (75,20%) dibandingkan mbert. Kesalahan terbesar terlihat pada kelas netral yang diprediksi sebagai negatif (20,87%), menunjukkan adanya ambiguitas dalam membedakan sentimen netral dan negatif pada aspek ini.

Secara keseluruhan, pelabelan mBERT menunjukkan performa yang lebih stabil dan seimbang pada ketiga kelas dibandingkan IndoBERT pada aspek fitur dan performa aplikasi. Meskipun demikian, kedua metode tetap mempertahankan tingkat akurasi yang relatif tinggi pada kelas positif dan negatif, yang menunjukkan kemampuan model dalam mengidentifikasi sentimen ekstrem dengan lebih konsisten dibandingkan sentimen netral. Kemudian, *Confusion matrix* ternormalisasi skenario pertama pada aspek layanan dan harga menggunakan pelabelan otomatis mBERT dan IndoBERT ditunjukkan pada Gambar 4.41

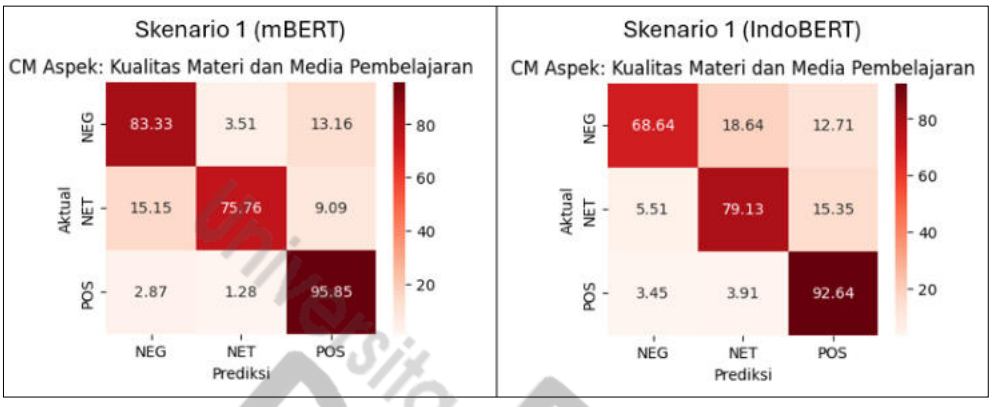


Gambar 4.41 *Confusion Matrix* Skenario 1 pada Aspek Akses Layanan dan Harga

Gambar 4.41 menujukkan bahwa pada pelabelan menggunakan mBERT, kelas negatif memiliki tingkat ketepatan yang tinggi (85,95%), namun kelas netral menunjukkan akurasi yang relatif rendah (57,14%). Kelas positif juga memiliki tingkat ketepatan yang sedang (71,79%). Kesalahan klasifikasi cukup besar terjadi pada kelas netral yang diprediksi sebagai negatif (28,57%) serta pada kelas positif yang diprediksi sebagai negatif (23,62%). Pola ini menunjukkan adanya kecenderungan model untuk mengarahkan sebagian data ke kelas negatif.

Sementara itu, pada pelabelan menggunakan IndoBERT, kelas netral menunjukkan peningkatan akurasi yang signifikan (86,01%) dibandingkan mBERT. Kelas negatif tetap memiliki akurasi tinggi (81,64%), namun kelas positif mengalami penurunan akurasi (60,56%). Kesalahan terbesar terlihat pada kelas positif yang diprediksi sebagai netral (25,35%), yang menunjukkan kesulitan model dalam membedakan sentimen positif dan netral pada aspek ini.

Secara keseluruhan, pelabelan mBERT menunjukkan performa yang lebih baik pada kelas negatif dan positif, sedangkan IndoBERT lebih unggul dalam mengidentifikasi kelas netral. Perbedaan ini menunjukkan bahwa karakteristik label hasil pelabelan otomatis berpengaruh terhadap pola klasifikasi model LSTM, terutama pada aspek yang memiliki distribusi sentimen lebih kompleks seperti akses layanan dan harga. Selanjutnya, *Confusion matrix* ternormalisasi skenario pertama pada aspek kualitas materi dan media pembelajaran menggunakan pelabelan otomatis mBERT dan IndoBERT ditunjukkan pada Gambar 4.42



Gambar 4.42 *Confusion Matrix* Skenario 1 pada Aspek Kualitas Materi dan Media Pembelajaran

Gambar 4.42 menujukkan bahwa Gambar 4.28 menujukkan bahwa pada pelabelan menggunakan mBERT, kelas positif memiliki tingkat ketepatan tertinggi (95,85%), diikuti oleh kelas negatif (83,33%) dan netral (75,76%). Kesalahan klasifikasi relatif kecil dan sebagian besar terjadi pada kelas netral yang diprediksi sebagai negatif (15,15%) atau positif (9,09%). Pola ini menunjukkan bahwa mBERT mampu mempertahankan distribusi klasifikasi yang cukup stabil pada ketiga kelas. Sementara itu, pada pelabelan menggunakan IndoBERT, kelas positif juga menunjukkan akurasi tinggi (92,64%) dan kelas netral sedikit lebih tinggi dibandingkan mBERT (79,13%). Namun, akurasi kelas negatif lebih rendah (68,64%) dan cukup banyak diprediksi sebagai netral (18,64%). Hal ini menunjukkan bahwa IndoBERT cenderung mengalami kesulitan dalam membedakan sentimen negatif secara konsisten pada aspek ini. Secara keseluruhan, kedua metode menunjukkan performa terbaik pada kelas positif, yang mengindikasikan dominasi sentimen positif pada aspek kualitas materi dan media pembelajaran. Namun, mBERT memberikan hasil klasifikasi yang lebih stabil dan seimbang pada kelas negatif dibandingkan IndoBERT dalam skenario 1.

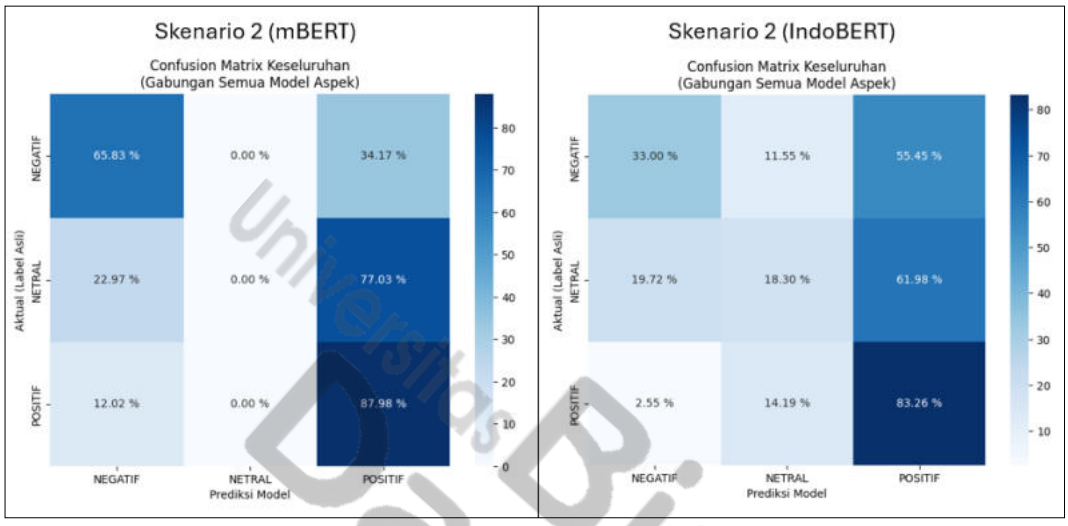
Berdasarkan hasil evaluasi *confusion matrix* pada masing-masing aspek, secara umum model LSTM mampu mengklasifikasikan sentimen dengan baik, terutama pada kelas positif yang secara konsisten menunjukkan nilai tertinggi pada diagonal matriks. Hal ini terlihat dari dominasi prediksi benar pada kelas positif di hampir seluruh aspek, yang mengindikasikan bahwa model lebih stabil dan akurat dalam mengenali ekspresi apresiatif atau ulasan dengan polaritas yang jelas dan eksplisit.

Namun demikian, variasi performa antar aspek tetap terlihat. Pada beberapa aspek seperti pendampingan belajar, kualitas materi dan media pembelajaran, serta kepuasan pengguna, proporsi prediksi benar relatif tinggi dan kesalahan klasifikasi lebih terkendali.

Sebaliknya, pada aspek seperti akses layanan dan harga, fitur dan performa aplikasi, serta capaian akademik, distribusi prediksi menunjukkan tingkat misklasifikasi yang lebih besar, khususnya pada kelas negatif dan netral. Kondisi ini dipengaruhi oleh beberapa faktor, antara lain ketidakseimbangan distribusi data yang menyebabkan dominasi kelas positif, adanya sentimen campuran dalam satu ulasan, kemiripan pola linguistik antar kelas terutama pada kelas negatif dan netral, serta banyaknya ulasan singkat yang minim konteks. Selain itu, penggunaan bahasa informal dan variasi istilah dalam ulasan juga meningkatkan ambiguitas semantik. Perbedaan karakteristik bahasa pada masing-masing aspek juga berkontribusi terhadap variasi performa, di mana aspek dengan ekspresi emosional yang eksplisit lebih mudah dikenali dibandingkan aspek yang bersifat teknis atau deskriptif.

2. Skenario kedua

Skenario kedua ialah evaluasi pada model yang dilatih secara terpisah untuk setiap aspek pada ulasan platform Ruangguru. Evaluasi kinerja model LSTM ini menggunakan gabungan model seluruh aspek untuk mengetahui konsistensi performa model secara keseluruhan. Hasil evaluasi ini memberikan gambaran umum kemampuan model dalam mengklasifikasikan sentimen negatif, netral, dan positif secara keseluruhan, tanpa membedakan aspek secara spesifik. *Confusion matrix* ternormalisasi berdasarkan hasil evaluasi skenario kedua menggunakan data seluruh aspek dengan pelabelan otomatis mBERT dan IndoBERT ditunjukkan pada Gambar 4.43

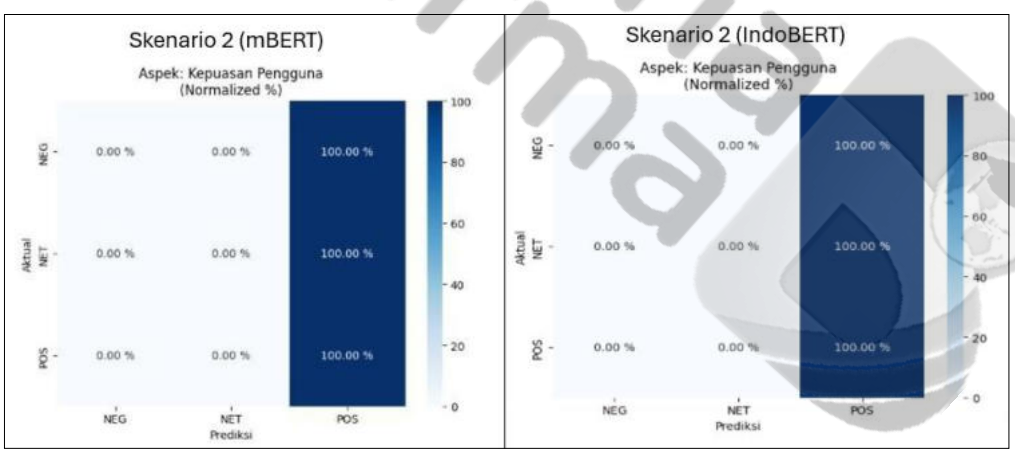


Gambar 4.43 *Confusion Matrix* Skenario 2 pada Seluruh Aspek

Gambar 4.43 menunjukkan bahwa pada skenario ini terlihat pola klasifikasi yang lebih tidak seimbang, khususnya dominasi prediksi pada kelas positif. Pada pelabelan menggunakan mBERT, kelas positif menunjukkan tingkat ketepatan tertinggi (87,98%). Namun, kelas netral tidak teridentifikasi dengan baik (0,00%), karena seluruh data netral diprediksi sebagai kelas positif (77,03%) atau negatif (22,97%). Kelas negatif masih memiliki ketepatan cukup baik (65,83%), tetapi sebagian besar kesalahan terjadi ketika data negatif diprediksi sebagai positif (34,17%). Pola ini menunjukkan kecenderungan model untuk mengklasifikasikan data ke kelas positif. Sementara itu, pada pelabelan menggunakan IndoBERT, distribusi klasifikasi terlihat lebih menyebar dibandingkan mBERT. Kelas positif tetap memiliki ketepatan tertinggi (83,26%), namun kelas netral mulai teridentifikasi meskipun masih rendah (18,30%). Kelas negatif memiliki ketepatan sebesar 33,00%, dengan kesalahan terbesar terjadi ketika data negatif diprediksi sebagai positif (55,45%). Hal ini menunjukkan bahwa meskipun IndoBERT menghasilkan distribusi yang lebih variatif, tingkat kesalahan antar kelas masih cukup tinggi.

Secara keseluruhan, pada skenario 2 dengan pelabelan otomatis mBERT dan IndoBERT menunjukkan kecenderungan dominan dalam memprediksi kelas positif, sehingga kemampuan model dalam membedakan kelas netral dan negatif menjadi lebih terbatas. Kondisi ini menunjukkan bahwa pendekatan model yang

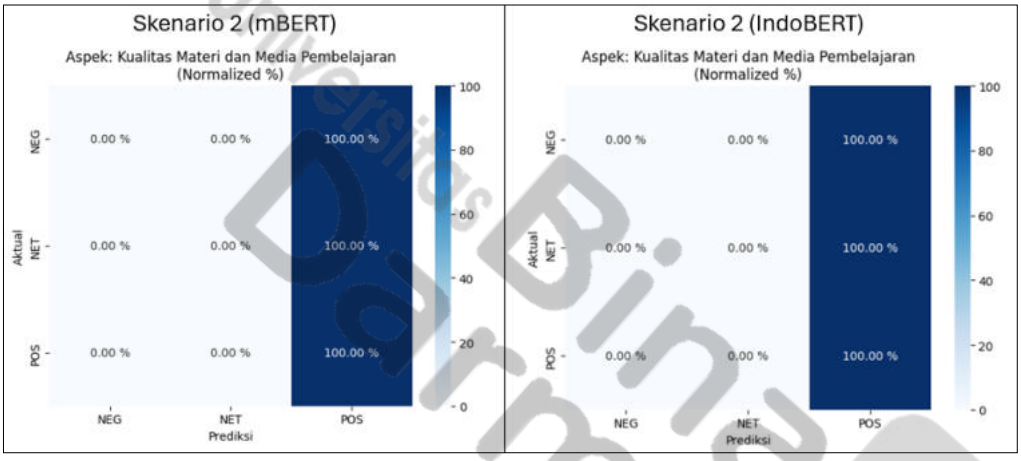
dilatih secara terpisah untuk setiap aspek pada skenario 2 memiliki tingkat kompleksitas yang lebih tinggi dalam menjaga keseimbangan klasifikasi dibandingkan pendekatan model yang dilatih menggunakan seluruh aspek pada skenario 1. Selain evaluasi model LSTM menggunakan seluruh aspek secara bersamaan, dilakukan pula evaluasi kinerja model LSTM berdasarkan masing-masing aspek untuk mengetahui konsistensi performa model dalam konteks aspek yang berbeda. Evaluasi ini dilakukan menggunakan *confusion matrix* ternormalisasi pada setiap aspek, sehingga memberikan gambaran rinci mengenai distribusi prediksi benar dan kesalahan klasifikasi pada setiap kelas sentimen. *Confusion matrix* ternormalisasi skenario kedua pada aspek kepuasan pengguna menggunakan pelabelan otomatis mBERT dan IndoBERT ditunjukkan pada Gambar 4.44



Gambar 4.44 *Confusion Matrix* Skenario 2 pada Aspek Kepuasan Pengguna

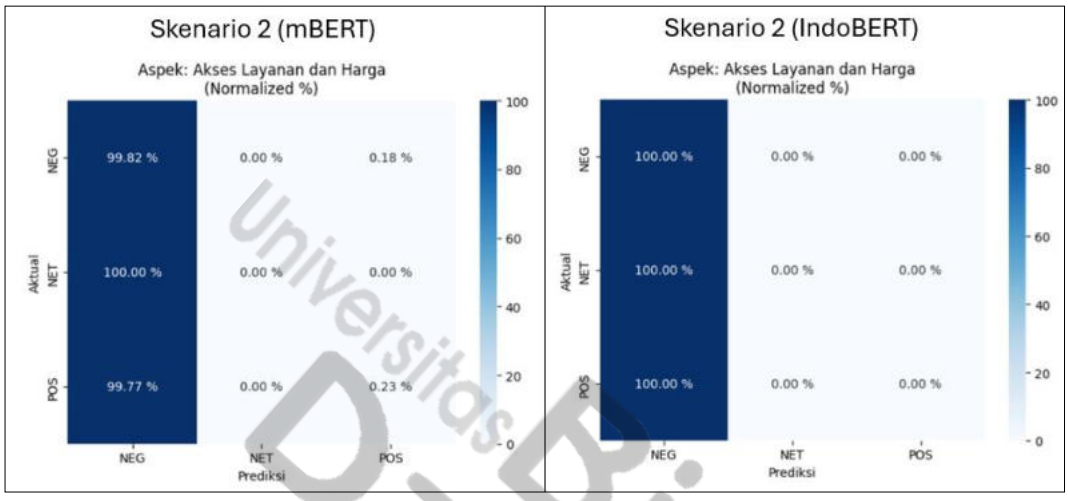
Gambar 4.44 menujukkan bahwa pada kedua metode pelabelan otomatis menggunakan mBERT dan IndoBERT, seluruh data dari ketiga kelas sentimen yaitu negatif, netral, dan positif diprediksi sebagai kelas positif (100%). Tidak terdapat nilai pada diagonal utama untuk kelas negatif maupun netral, yang menunjukkan bahwa model sama sekali tidak mampu mengidentifikasi kedua kelas tersebut. Semua data aktual, baik negatif maupun netral, sepenuhnya diklasifikasikan sebagai positif. Pola ini menunjukkan adanya bias prediksi yang sangat kuat terhadap kelas positif. Kondisi ini mengindikasikan bahwa pada skenario 2, model mengalami ketidakseimbangan klasifikasi yang ekstrem pada aspek kepuasan pengguna. Hal tersebut dapat disebabkan oleh dominasi jumlah data positif dalam dataset atau keterbatasan variasi pola bahasa pada kelas negatif

dan netral, sehingga model gagal membedakan ketiga kelas secara efektif. Selanjutnya, *Confusion matrix* ternormalisasi skenario kedua pada aspek kualitas materi dan media pembelajaran dengan pelabelan otomatis mBERT dan IndoBERT ditunjukkan pada Gambar 4.45



Gambar 4.45 *Confusion Matrix* Skenario 2 pada Aspek Kualitas Materi dan Media Pembelajaran

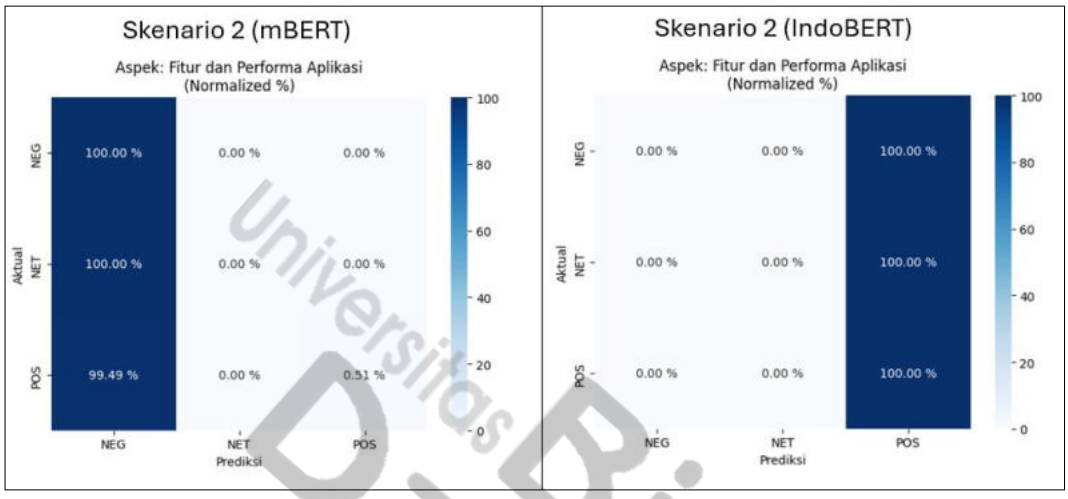
Gambar 4.45 menujukkan bahwa pada aspek kualitas materi dan media pembelajaran menggunakan pelabelan otomatis mBERT dan IndoBERT, seluruh data dari semua kelas sentimen yaitu negatif, netral, dan positif diprediksi sebagai kelas positif dengan nilai 100%. Tidak terdapat nilai pada diagonal utama untuk kelas negatif maupun netral, yang menunjukkan bahwa model tidak mampu mengidentifikasi kedua kelas tersebut. Seluruh data aktual, baik negatif maupun netral, sepenuhnya diklasifikasikan sebagai positif. Pola ini menunjukkan bahwa model mengalami kecenderungan bias terhadap kelas positif dan tidak mampu membedakan sentimen negatif maupun netral pada aspek ini. Hal ini kemungkinan dipengaruhi oleh dominasi sentimen positif dalam dataset atau keterbatasan variasi pola bahasa pada kelas negatif dan netral, sehingga model gagal membedakan distribusi sentimen secara efektif. *Confusion matrix* ternormalisasi skenario kedua pada aspek akses layanan dan harga dengan pelabelan otomatis mBERT dan IndoBERT ditunjukkan pada Gambar 4.46



Gambar 4.46 *Confusion Matrix* Skenario 2 pada Aspek Akses Layanan dan Harga

Gambar 4.46 menujukkan bahwa pada aspek akses layanan dan harga menggunakan pelabelan otomatis mBERT dan IndoBERT terlihat pola klasifikasi yang sangat tidak seimbang, dengan dominasi prediksi pada satu kelas tertentu. Pada pelabelan menggunakan mBERT, hampir seluruh data aktual dari ketiga kelas (negatif, netral, dan positif) diprediksi sebagai kelas negatif, dengan nilai mendekati 100% (negatif 99,82%, netral 100%, dan positif 99,77%). Kelas netral dan positif hampir tidak pernah diprediksi dengan benar. Hal ini menunjukkan bias prediksi yang sangat kuat ke arah kelas negatif. Sementara itu, pada pelabelan menggunakan IndoBERT, seluruh data aktual dari semua kelas sepenuhnya diprediksi sebagai negatif (100%). Tidak terdapat prediksi untuk kelas netral maupun positif. Kondisi ini menunjukkan ketidakseimbangan klasifikasi yang lebih ekstrem dibandingkan mBERT.

Secara keseluruhan, pada aspek akses layanan dan harga dalam skenario 2, kedua metode gagal membedakan ketiga kelas sentimen secara efektif dan menunjukkan kecenderungan kuat untuk memusatkan prediksi pada satu kelas dominan. Hal ini mengindikasikan adanya masalah ketidakseimbangan data atau keterbatasan generalisasi model pada pendekatan model yang dilatih secara terpisah untuk setiap aspek. Selanjutnya, *Confusion matrix* ternormalisasi skenario kedua pada aspek fitur dan performa aplikasi dengan pelabelan otomatis mBERT dan IndoBERT ditunjukkan pada Gambar 4.47

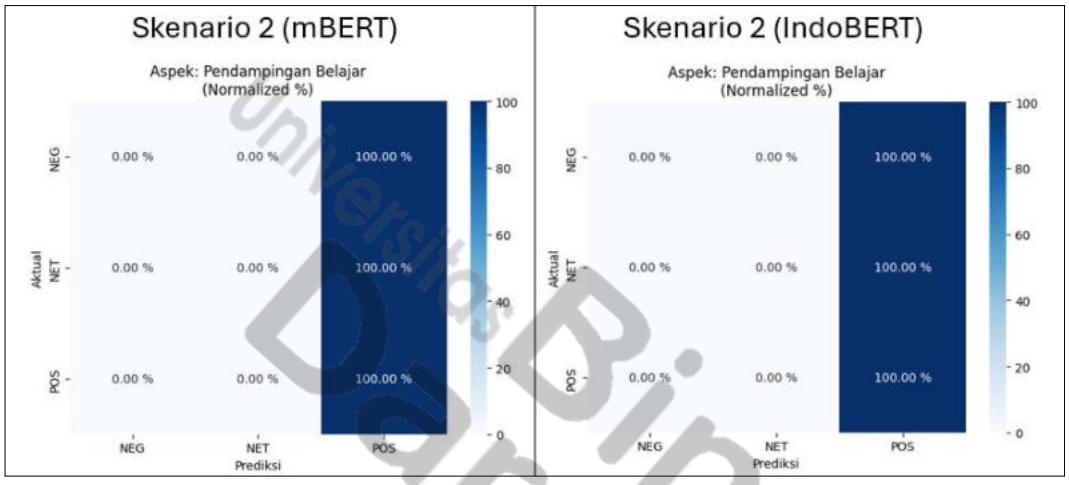


Gambar 4.47 *Confusion Matrix* Skenario 2 pada Aspek Fitur dan Performa Aplikasi

Gambar 4.47 menujukkan bahwa pada aspek fitur dan performa aplikasi menggunakan pelabelan otomatis mBERT dan IndoBERT terlihat pola klasifikasi yang sangat tidak seimbang, dengan dominasi prediksi pada satu kelas tertentu. Pada pelabelan menggunakan mBERT, hampir seluruh data aktual dari ketiga kelas diprediksi sebagai kelas negatif (100% untuk negatif dan netral, serta 99,49% untuk positif). Hanya sebagian sangat kecil data positif yang diprediksi benar (0,51%). Hal ini menunjukkan bahwa model mengalami bias kuat terhadap kelas negatif dan gagal membedakan kelas netral serta positif. Sementara itu, pada pelabelan menggunakan IndoBERT, seluruh data aktual dari semua kelas sepenuhnya diprediksi sebagai kelas positif (100%). Tidak terdapat prediksi untuk kelas negatif maupun netral. Kondisi ini menunjukkan bias klasifikasi yang ekstrem dan ketidakmampuan model dalam mengidentifikasi variasi sentimen pada aspek tersebut.

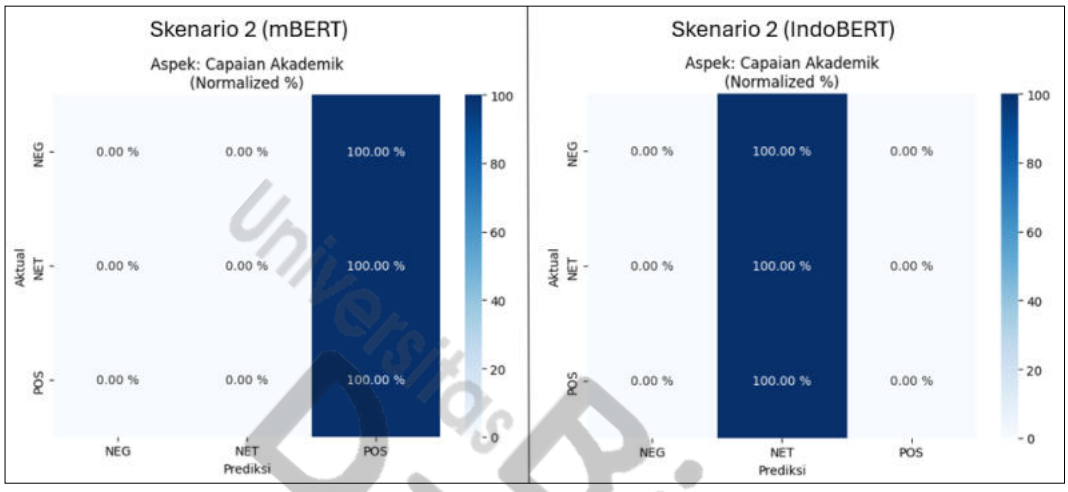
Secara keseluruhan, pada aspek fitur dan performa aplikasi dalam skenario 2, kedua metode tidak mampu melakukan klasifikasi secara proporsional dan menunjukkan kecenderungan prediksi tunggal pada satu kelas dominan. Hal ini mengindikasikan adanya permasalahan ketidakseimbangan data atau keterbatasan generalisasi model yang dilatih secara terpisah untuk setiap aspek dalam membedakan distribusi sentimen secara efektif. Kemudian, *Confusion matrix* ternormalisasi dan *confusion matrix absolut* skenario kedua pada aspek

pendampingan belajar dengan pelabelan otomatis mBERT dan IndoBERT ditunjukkan pada Gambar 4.48



Gambar 4.48 *Confusion Matrix* Skenario 2 pada Aspek Pendampingan Belajar

Gambar 4.48 menujukkan bahwa pada aspek pendampingan belajar menggunakan pelabelan otomatis mBERT dan IndoBERT terlihat seluruh data dari semua kelas sentimen yaitu negatif, netral, dan positif diprediksi sebagai kelas positif dengan nilai 100%. Tidak terdapat nilai pada diagonal utama untuk kelas negatif maupun netral, yang menunjukkan bahwa model tidak mampu mengidentifikasi kedua kelas tersebut. Semua data aktual, baik negatif maupun netral, sepenuhnya diklasifikasikan sebagai positif. Pola ini menunjukkan bias klasifikasi yang sangat kuat terhadap kelas positif. Kondisi ini mengindikasikan bahwa pada pendekatan model yang dilatih secara terpisah untuk setiap aspek pada skenario 2 mengalami ketidakseimbangan klasifikasi yang ekstrem pada aspek pendampingan belajar. Hal tersebut kemungkinan dipengaruhi oleh dominasi data positif dalam dataset atau keterbatasan variasi pola bahasa pada kelas negatif dan netral, sehingga model gagal membedakan ketiga kelas sentimen secara efektif. Selanjutnya, *Confusion matrix* ternormalisasi skenario kedua pada aspek capaian akademik dengan pelabelan otomatis mBERT dan IndoBERT ditunjukkan pada Gambar 4.35



Gambar 4.49 *Confusion Matrix* Skenario 2 pada Aspek Capaian Akademik

Gambar 4.49 menujukkan bahwa pada aspek capaian akademik menggunakan pelabelan otomatis mBERT dan IndoBERT terlihat pola klasifikasi yang sangat tidak seimbang dengan dominasi prediksi pada satu kelas tertentu. Pada pelabelan menggunakan mBERT, seluruh data aktual dari ketiga kelas sentimen (negatif, netral, dan positif) diprediksi sebagai kelas positif (100%). Tidak terdapat prediksi yang benar untuk kelas negatif maupun netral, sehingga diagonal utama hanya terisi pada kelas positif. Hal ini menunjukkan bahwa model mengalami bias ekstrem terhadap kelas positif dan gagal membedakan variasi sentimen pada aspek ini. Sementara itu, pada pelabelan menggunakan IndoBERT, seluruh data aktual dari semua kelas diprediksi sebagai kelas netral (100%). Tidak terdapat prediksi untuk kelas negatif maupun positif. Kondisi ini menunjukkan adanya bias klasifikasi yang sangat kuat, dengan kecenderungan prediksi yang terfokus pada kelas netral. secara keseluruhan, pada aspek capaian akademik dalam skenario 2, kedua metode tidak mampu melakukan klasifikasi secara proporsional dan menunjukkan kecenderungan prediksi tunggal pada satu kelas dominan. hal ini mengindikasikan adanya keterbatasan model dalam menangkap distribusi sentimen yang beragam, kemungkinan akibat ketidakseimbangan data atau kompleksitas variasi bahasa pada aspek tersebut.

Berdasarkan hasil evaluasi dengan menggunakan *confusion matrix* pada skenario model LSTM yang dilatih secara terpisah untuk setiap aspek menunjukkan bahwa sebagian besar model mengalami bias yang sangat kuat terhadap satu kelas

sentimen dominan, baik positif, netral atau negatif. Fenomena ini disebabkan oleh ketidakseimbangan distribusi data sentimen yang ekstrem, keterbatasan jumlah data pada beberapa aspek, serta kurangnya variasi ekspresi sentimen dalam data pelatihan. Temuan ini menunjukkan bahwa pelatihan model secara terpisah per aspek tanpa dukungan jumlah data yang memadai dan seimbang berpotensi menurunkan kemampuan generalisasi model dalam klasifikasi sentimen multikelas.

4.7 Analisis Hasil Evaluasi Model LSTM

Berdasarkan hasil evaluasi yang telah dilakukan terhadap model LSTM pada dua skenario pelatihan, diperoleh sejumlah temuan yang signifikan yaitu sebagai berikut:

1. Perbandingan Kinerja Dua Skenario Pelatihan

Berdasarkan hasil pengujian, model LSTM yang dilatih menggunakan seluruh data ulasan dari semua aspek dalam satu model (Skenario 1) menunjukkan performa yang lebih tinggi dan konsisten dibandingkan model yang dilatih secara terpisah pada masing-masing aspek (Skenario 2). Skenario pertama mampu mencapai tingkat akurasi yang lebih baik serta nilai precision, recall, dan F1-score yang lebih seimbang antar kelas. Hal ini menunjukkan bahwa pelatihan dengan data yang lebih besar dan lebih beragam memungkinkan model mempelajari representasi sentimen secara lebih komprehensif. Sebaliknya, pada skenario kedua, performa model cenderung menurun dan tidak stabil akibat keterbatasan jumlah data pada tiap aspek, sehingga model kurang mampu membangun pola klasifikasi yang kuat.

2. Pengaruh Jumlah dan Distribusi Data

Jumlah dan distribusi data terbukti menjadi faktor yang sangat menentukan dalam kinerja model LSTM. Pada skenario pelatihan gabungan, model memperoleh keuntungan dari jumlah data yang lebih besar serta variasi konteks bahasa yang lebih luas. Hal ini meningkatkan kemampuan generalisasi model terhadap berbagai pola ekspresi sentimen. Sebaliknya, pada pelatihan terpisah per aspek, ketidakseimbangan distribusi kelas sentimen menyebabkan model cenderung bias terhadap kelas mayoritas. Ketimpangan ini berdampak pada menurunnya kemampuan model dalam mengenali kelas minoritas, khususnya sentimen netral

dan negatif pada aspek tertentu. Dengan demikian, keseimbangan data menjadi prasyarat penting dalam pengembangan model klasifikasi multikelas.

3. Kemampuan Klasifikasi berdasarkan Kelas Sentimen

Model menunjukkan kemampuan yang lebih baik dalam mengklasifikasikan sentimen positif dibandingkan sentimen netral dan negatif. Hal ini disebabkan oleh dominasi jumlah data positif serta karakteristik ekspresi positif yang cenderung lebih eksplisit dan mudah dikenali. Sentimen negatif memiliki tingkat kesalahan yang lebih tinggi, terutama pada ulasan yang mengandung kritik ringan atau sentimen campuran. Sementara itu, sentimen netral merupakan kelas yang paling sulit diklasifikasikan secara konsisten karena sering kali memiliki batas yang samar antara ekspresi evaluatif dan deskriptif. Kondisi ini menunjukkan bahwa klasifikasi multikelas dalam analisis sentimen memerlukan distribusi data yang seimbang dan variasi konteks yang cukup untuk menghasilkan performa optimal.

4. Stabilitas Proses Pelatihan

Dari sisi proses pelatihan, skenario pertama menunjukkan pola pembelajaran yang relatif stabil. Kurva akurasi dan loss memperlihatkan peningkatan performa secara bertahap hingga mencapai titik konvergensi, meskipun terdapat indikasi overfitting ringan pada beberapa epoch akhir. Penerapan mekanisme early stopping terbukti efektif dalam mencegah overfitting berlebihan. Sebaliknya, pada skenario kedua, proses pelatihan menunjukkan fluktuasi yang lebih besar, terutama pada fase awal, akibat keterbatasan jumlah data pada masing-masing aspek. Meskipun model tetap mengalami konvergensi, stabilitas tersebut tidak selalu diikuti dengan peningkatan kemampuan klasifikasi, yang mengindikasikan bahwa model cenderung mengarah pada solusi bias terhadap kelas dominan.

5. Variasi Performa Antar Aspek

Hasil evaluasi juga menunjukkan adanya variasi performa antar aspek layanan. Aspek yang memiliki ekspresi sentimen lebih eksplisit dan dominasi kelas positif menunjukkan tingkat akurasi yang lebih tinggi. Sebaliknya, aspek yang berkaitan dengan fitur teknis, harga, atau akses layanan cenderung memiliki variasi bahasa yang lebih kompleks dan ambigu, sehingga tingkat kesalahan klasifikasi menjadi lebih tinggi. Perbedaan ini menunjukkan bahwa karakteristik linguistik setiap aspek

memengaruhi kemampuan model dalam membangun representasi sentimen. Oleh karena itu, pendekatan pelatihan yang mempertimbangkan kompleksitas bahasa dan distribusi data pada tiap aspek menjadi faktor penting dalam meningkatkan kinerja model secara menyeluruh.

4.8 Analisis Pengaruh Pelabelan Otomatis mBERT dan IndoBERT terhadap Kinerja LSTM

Performa model *deep learning* tidak hanya ditentukan oleh kompleksitas arsitektur yang digunakan, tetapi juga sangat dipengaruhi oleh kualitas data yang menjadi dasar proses pembelajaran. Dalam pendekatan *supervised learning*, termasuk *deep learning* berbasis *Long Short-Term Memory* (LSTM), label referensi (*ground truth*) berperan sebagai sinyal utama dalam proses optimasi bobot jaringan. Setiap pembaruan parameter model dilakukan berdasarkan selisih antara prediksi dan label yang diberikan. Oleh karena itu, konsistensi dan akurasi label menjadi faktor krusial dalam menentukan efektivitas pembelajaran. Hasil analisis perbandingan antara metode pelabelan otomatis mBERT dan IndoBERT terhadap kinerja LSTM ialah sebagai berikut:

1. Kualitas *Ground Truth* sebagai Faktor Penentu Kinerja Model

Dalam pendekatan *supervised learning*, kualitas *ground truth* merupakan faktor fundamental yang menentukan keberhasilan model dalam mempelajari pola klasifikasi. Model tidak hanya belajar dari data teks yang diberikan, tetapi juga dari label yang digunakan sebagai referensi selama proses pelatihan. Walaupun arsitektur LSTM yang digunakan sama, namun performa yang dihasilkan berbeda. Hal ini menunjukkan bahwa variasi hasil bukan disebabkan oleh perubahan struktur model, melainkan oleh karakteristik label referensi yang digunakan saat pelatihan. Oleh karena itu, apabila label yang digunakan memiliki tingkat akurasi dan konsistensi yang tinggi, maka model akan mampu membangun hubungan yang lebih tepat antara fitur teks dan kelas sentimen yang diprediksi.

Hasil penelitian menunjukkan bahwa model LSTM yang dilatih menggunakan label dari mBERT menghasilkan nilai akurasi dan F1-score yang lebih tinggi dibandingkan model yang dilatih menggunakan label dari IndoBERT. Hal ini

menunjukkan bahwa label yang dihasilkan oleh mBERT memiliki konsistensi yang lebih baik dalam merepresentasikan polaritas sentimen pada ulasan pengguna. Sebaliknya, apabila label referensi mengandung inkonsistensi atau kesalahan klasifikasi, model akan mempelajari pola yang tidak tepat sehingga menurunkan performa keseluruhan model.

2. Konsistensi Polaritas Sentimen

Analisis *confusion matrix* memberikan gambaran yang lebih rinci mengenai bagaimana model melakukan klasifikasi terhadap setiap kelas sentimen. Berdasarkan hasil evaluasi, terlihat bahwa pelabelan berbasis IndoBERT memiliki tingkat kesalahan silang yang lebih tinggi antara kelas negatif dan netral dibandingkan dengan pelabelan berbasis mBERT. Hal ini menunjukkan adanya kecenderungan *sentiment smoothing*, yaitu kondisi di mana sentimen negatif yang seharusnya diklasifikasikan sebagai kelas negatif justru digeser ke dalam kelas netral. Hal ini dapat terjadi ketika model pelabelan otomatis tidak mampu membedakan secara tegas antara ekspresi sentimen yang bersifat negatif dengan ekspresi yang bersifat netral atau ambigu.

Dalam konteks pelatihan model LSTM, kondisi ini dapat menimbulkan masalah karena model akan mempelajari pola yang tidak sepenuhnya merepresentasikan polaritas sentimen yang sebenarnya. Ketika sejumlah ulasan negatif diberi label netral, maka karakteristik linguistik yang membedakan kedua kelas tersebut menjadi kurang jelas bagi model. Sebaliknya, pelabelan menggunakan mBERT menunjukkan distribusi kesalahan yang lebih kecil dan lebih terpusat pada diagonal confusion matrix, yang menunjukkan bahwa sebagian besar data diklasifikasikan dengan benar. Hal ini menunjukkan bahwa polaritas sentimen pada label mBERT lebih konsisten sehingga memudahkan model dalam mempelajari batas pemisah antar kelas.

3. Dampak *Noise* Label terhadap Pembentukan *Decision Boundary*

Dalam pembelajaran terawasi, model membangun *decision boundary* atau batas pemisah antar kelas berdasarkan label yang diberikan selama proses pelatihan. *Decision boundary* ini berfungsi untuk memisahkan data yang memiliki karakteristik berbeda sehingga model dapat menentukan kelas yang tepat untuk

setiap data baru. Namun, apabila label referensi mengandung kesalahan atau *noise*, maka proses pembentukan *decision boundary* akan terganggu. Data yang sebenarnya berasal dari kelas tertentu dapat ditempatkan pada kelas yang berbeda, sehingga model akan membangun batas pemisah yang tidak optimal.

Pada pelabelan berbasis IndoBERT, tingkat kesalahan silang yang relatif tinggi menunjukkan bahwa kemungkinan *noise* label pada dataset juga lebih besar. Ketika sejumlah data dengan sentimen negatif diberi label netral, perbedaan karakteristik linguistik antara kedua kelas menjadi kurang jelas. Akibatnya, batas pemisah antara kelas negatif dan netral menjadi kabur dan saling tumpang tindih (*overlap*). Sebaliknya, pelabelan berbasis mBERT menghasilkan distribusi kesalahan yang lebih kecil sehingga membantu model LSTM membangun representasi fitur yang lebih jelas. Dengan label yang lebih konsisten, model dapat membentuk *decision boundary* yang lebih tegas sehingga meningkatkan kemampuan model dalam membedakan antar kelas sentimen.

4. Peran Distribusi Kelas dalam Stabilitas Model

Distribusi kelas merupakan faktor penting dalam proses pelatihan model klasifikasi. Dataset yang memiliki distribusi kelas yang seimbang memungkinkan model mempelajari karakteristik masing-masing kelas secara optimal. Sebaliknya, dataset yang tidak seimbang dapat menyebabkan model lebih fokus pada kelas dominan dan kurang mampu mengenali kelas minoritas. Hasil analisis distribusi data menunjukkan bahwa pelabelan menggunakan IndoBERT menghasilkan proporsi kelas netral yang lebih tinggi dibandingkan pelabelan menggunakan mBERT. Ketidakseimbangan distribusi kelas ini dapat menyebabkan model lebih sering mempelajari pola dari kelas netral, sehingga kemampuan model dalam mengenali kelas negatif atau positif menjadi berkurang. Sebaliknya, distribusi kelas pada pelabelan mBERT terlihat lebih proporsional sehingga membantu model mempelajari karakteristik setiap kelas sentimen secara lebih seimbang. Kondisi ini berkontribusi terhadap stabilitas pembelajaran model serta meningkatkan kemampuan model dalam melakukan klasifikasi yang lebih akurat.

5. Pengaruh *Stemming* pada Tahap Pra-Pemrosesan Data

Tahap pra-pemrosesan data merupakan salah satu proses penting yang dapat

memengaruhi performa metode pelabelan otomatis dalam analisis sentimen. Pada penelitian ini dilakukan proses *stemming* untuk mengubah kata berimbuhan menjadi bentuk dasar sebelum tahap analisis sentimen berbasis aspek menggunakan metode LSTM. Proses *stemming* tersebut bertujuan untuk mengurangi variasi kosakata serta menyederhanakan bentuk kata dalam dataset sehingga representasi teks menjadi lebih seragam dan memudahkan proses pembelajaran model. Namun, dalam bahasa Indonesia, imbuhan memiliki peran penting dalam membentuk makna kata. Ketika proses *stemming* dilakukan, sebagian informasi morfologis yang terkandung dalam kata berimbuhan dapat hilang. Kondisi ini dapat memengaruhi kemampuan model dalam memahami konteks sentimen yang sebenarnya.

IndoBERT yang dilatih secara khusus pada korpus bahasa Indonesia cenderung mempertahankan informasi morfologis kata dalam proses representasi bahasa. Ketika kata-kata dalam dataset telah melalui proses *stemming*, sebagian informasi linguistik yang digunakan oleh IndoBERT untuk memahami konteks sentimen menjadi berkurang. Sebaliknya, mBERT yang dilatih pada korpus multibahasa yang sangat besar cenderung lebih mampu menangani variasi bentuk kata yang telah disederhanakan. Hal ini dapat menjadi salah satu faktor yang menyebabkan pelabelan otomatis menggunakan mBERT menghasilkan label sentimen yang lebih stabil dibandingkan IndoBERT pada dataset yang telah melalui proses *stemming*.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini menunjukkan bahwa metode *Long Short-Term Memory* (LSTM) efektif diterapkan dalam analisis sentimen berbasis aspek terhadap ulasan pengguna Ruangguru. Proses penelitian meliputi pengumpulan data, pra-pemrosesan teks, ekstraksi dan labeling aspek, serta pelabelan sentimen otomatis sebagai dasar klasifikasi. Dalam penelitian ini, dilakukan dua skenario pelatihan model LSTM yaitu pelatihan gabungan seluruh aspek dan pelatihan terpisah per aspek. Hasil menunjukkan bahwa pelatihan gabungan seluruh aspek menghasilkan performa yang lebih stabil dan konsisten karena didukung jumlah data yang lebih besar dan distribusi sentimen yang lebih beragam. Sebaliknya, pelatihan terpisah per aspek cenderung bias terhadap kelas dominan akibat ketidakseimbangan distribusi data, sehingga menurunkan kemampuan generalisasi model.

Selain itu, metode pelabelan otomatis berpengaruh signifikan terhadap performa LSTM. Penggunaan label mBERT menghasilkan akurasi dan konsistensi klasifikasi yang lebih tinggi dibandingkan IndoBERT. Pelabelan IndoBERT menunjukkan kecenderungan *sentiment smoothing*, yaitu pelemahan sentimen negatif menjadi netral yang meningkatkan kesalahan antar kelas dan potensi *noise* label. Sebaliknya, pelabelan mBERT menghasilkan kesalahan yang lebih rendah dan pemisahan antar kelas yang lebih tegas, sehingga meningkatkan kemampuan generalisasi model. Dengan demikian, keberhasilan LSTM tidak hanya ditentukan oleh arsitektur dan strategi pelatihan, tetapi sangat dipengaruhi oleh kualitas dan konsistensi metode pelabelan otomatis yang digunakan sebagai *ground truth*.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, penelitian selanjutnya disarankan untuk menggunakan dataset dengan distribusi yang lebih seimbang pada setiap aspek dan kelas sentimen agar kinerja model LSTM dapat lebih optimal, khususnya pada aspek dengan jumlah data terbatas. Selain itu, pemanfaatan

arsitektur *deep learning* yang lebih kontekstual, seperti BiLSTM, *Attention-based* LSTM, atau model berbasis transformer, dapat dipertimbangkan untuk meningkatkan kemampuan model dalam menangkap sentimen implisit dan konteks kalimat yang kompleks. Dari sisi implementasi, hasil penelitian ini dapat dikembangkan menjadi sistem analisis sentimen berbasis aspek secara *real-time* dalam bentuk *dashboard* evaluasi layanan, serta diperluas dengan memanfaatkan sumber data ulasan yang lebih beragam agar analisis yang dihasilkan lebih komprehensif.

DAFTAR PUSTAKA

Ali, M., Dina, M., Fitri, M., Chrisman, D., Shella, L., Saptiany, G., Asrul, Djamil, M., Nur, M., & Raini, Y. (2025). *Teknologi Pendidikan Teori Dan Aplikasi* (M. I. Al Kutsi, Ed.; Cetakan Pertama). Azzia Karya Bersama.

Amien, S., Perdana, P., Bharata Aji, T., & Ferdiana, R. (2021). Aspect Category Classification dengan Pendekatan Machine Learning Menggunakan Dataset Bahasa Indonesia (Aspect Category Classification with Machine Learning Approach Using Indonesian Language Dataset). *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi* |, *10*(3). https://doi.org/https://doi.org/10.22146/jnteti.v10i3.1819

An, Y., Oh, H., & Lee, J. (2023). Marketing Insights from Reviews Using Topic Modeling with BERTopic and Deep Clustering Network. *Applied Sciences (Switzerland)*, *13*(16). https://doi.org/10.3390/app13169443

Asmara, Y., Rudyanto Arief, M., & kusrini. (2024). Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi RuangGuru Menggunakan Support Vector Machine dan Naive. *Februari*, *2024*(2), 209–215. https://doi.org/https://doi.org/10.36774/sisiti.v13i2.1616

Astiningrum, M., Saputra, Y. P., & Rohmah, S. M. (2018). Implementasi NLP dengan Konversi Kata Pada Sistem Chatbot Konsultasi Laktasi. *Jurnal Informatika Polinema*, *5*, 46–52. https://doi.org/10.33795/jip.v5i1.262

Baradja, A. (2023). Pemanfaatan Recurrent Neural Network (RNN) Untuk Meningkatkan Akurasi Prediksi Mata Uang Pada Forex Trading. In *Journal of Software Engineering Ampera* (Vol. 4, Number 2). https://doi.org/10.51519/journalsea.v4i2.505

Cahyaningtyas, S. (2021). *Deep Learning pada Aspect-based Sentiment Analysis pada Data Hotel Berbahasa Indonesia*. Universitas Islam Indonesia.

Daffa, M., Fahreza, A., Luthfiarta, A., Rafid, M., Indrawan, M., & Nugraha, A. (2024). Analisis Sentimen: Pengaruh Jam Kerja Terhadap Kesehatan Mental Generasi Z. *Journal Of Applied Computer Science And Technology (Jacost)*, *5*(1), 16–14535. https://doi.org/10.52158/jacost.715

Fitri, E., Yuliani, Y., Rosyida, S., & Gata, W. (2020). Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine. *Transformtika*, *18*(1), 71–80. https://doi.org/https://doi.org/10.26623/transformatika.v18i1.2317

Hasugian, A. H., Fakhriza, M., & Zukhoiriyah, D. (2023). Analisis Sentimen Pada Review Pengguna E-Commerce Menggunakan Algoritma Naive Bayes. *Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD*, *6*, 98–107. https://doi.org/https://doi.org/10.53513/jsk.v6i1.7400

Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.

Lubis, M. A., Azizan, N., Rasyid, A., & Marina, N. (2021). Aplikasi Ruangguru untuk Pembelajaran Di era Covid-19. *Prosiding Webinar Nasional Prodi PGMI IAIN Padangsidimpuan*, 19–28.

Mubaroroh, H. H., Yasin, H., & Rusgiyono, A. (2022). Analisis Sentimen Data Ulasan Aplikasi Ruangguru Pada Situs Google Play Menggunakan Algoritma Naïve Bayes Classifier Dengan Normalisasi Kata Levenshtein Distance. *Jurnal Gaussian*, *11*(2), 248–257. https://doi.org/10.14710/j.gauss.v11i2.35472

Naquitasia, R., Hatta Fudholi, D., & Iswari, L. (2022). Analisis Sentimen Berbasis Aspek Pada Wisata Halal Dengan Metode Deep Learning. *Jurnal Teknoinfo*, *16*(2), 156–164. https://doi.org/10.33365/jti.v16i2.1516

Novirianto, I. F. (2023). *Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Mybluebird Dengan Implementasi N-Gram Dan Algoritma Logistic Regression*. UIN Syarif Hidayatullah.

Nurhidayat, R., & Dewi, K. E. (2023). Penerapan Algoritma K-Nearest Neighbor Dan Fitur Ekstraksi N-Gram Dalam Analisis Sentimen Berbasis Aspek. *Komputa : Jurnal Ilmiah Komputer Dan Informatika*, *12*(1), 91–100. https://doi.org/10.34010/komputa.v12i1.9458

Pascanu, R., Mikolov, T., & Bengio, Y. (2013). *On the difficulty of training Recurrent Neural Networks*. http://arxiv.org/abs/1211.5063

Petter, S., DeLone, W., & McLean, E. (2008). Measuring information systems

success: Models, dimensions, measures, and interrelationships. *European Journal of Information Systems*, *17*(3), 236–263. https://doi.org/10.1057/ejis.2008.15

Raditya, B., Listyawan, N., Setiawan, N. Y., & Saputra, M. C. (2017). Topic Modelling Pada Aktivitas Pengembangan Perangkat Lunak Menggunakan BERTopic. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *1*(1), 1–11. http://j-ptiik.ub.ac.id

Rahadian, D., Rahayu, G., & Oktavia, R. R. (2019). Teknologi Pendidikan: Kajian Aplikasi Ruangguru Berdasarkan Prinsip dan Paradigma Interaksi Manusia dan Komputer. *Jurnal PETIK*, *5*(1), 11–21.

Ramadhan, A. R. (2023). *Analisis Sentimen Ulasan Aplikasi Ruangguru dengan Algoritma Long Short Term Memory*. UIN Syarif Hidayatullah.

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. *WSDM 2015 - Proceedings of the 8th ACM International Conference on Web Search and Data Mining*, 399–408. https://doi.org/10.1145/2684822.2685324

Rolangon, A., Weku, A., & Sandag, G. A. (2023). Perbandingan Algoritma LSTM Untuk Analisis Sentimen Pengguna Twitter Terhadap Layanan Rumah Sakit Saat Pandemi Covid-19. *Jurnal TeIKa*, *13*, 31–40. https://doi.org/https://doi.org/10.36342/teika.v13i01.3063

Saputri, R., & Baita, A. (2025). Analisis Sentimen Review Film Avatar 2 pada Platform IMDb Menggunakan LSTM dan GRU. *Jurnal Teknik Informatika Dan Terapan*, *3*(1), 24–36. https://doi.org/10.62951/router.v3i1.395

Sirkis, T., & Maitland, S. (2023). Monitoring real-time junior doctor sentiment from comments on a public social media platform: a retrospective observational study. *Postgraduate Medical Journal*, *99*(1171), 423–427. https://doi.org/10.1136/pmj-2022-142080

Sparta, & Rimadias, S. (2024). *Optimalisasi Artificial Intelligence (AI) Pada Platform SINTA*.

Subowo, E., Adi Artanto, F., Putri, I., & Umaedi, W. (2022). *BLTSM untuk analisis sentimen berbasis aspek pada aplikasi belanja online dengan cicilan*. *12*, 132–

140. https://doi.org/https://doi.org/10.37859/jf.v12i2.3759

Urbach, N., & Müller, B. (2012). *The Updated DeLone and McLean Model of Information Systems Success* (pp. 1–18). https://doi.org/10.1007/978-1-4419-6108-2_1

Warakmulty, P., Yeffry, D., & Putra, H. (2025). Optimalisasi Manajemen Sentimen di Media Sosial Universitas melalui Machine Learning dan AI: Studi Kasus pada Komentar Instagram. *Jurnal Tata Kelola Dan Kerangka Kerja Teknologi Informasi*, *11*(1), 31–38.

Yanti, I., & Utami, E. (2025). Analisis Sentimen Pemindahan Ibu Kota Indonesia Menggunakan Word Embedding Word2Vec dan Metode Long Short-Term Memory. *Jurnal Teknik Informatika (Jutif)*, *6*(1), 149–158. https://doi.org/10.52436/1.jutif.2025.6.1.2712

Yazid, M. A., Wijoyo, S. H., & Rokhmawati, R. I. (2019). Evaluasi Kualitas Aplikasi Ruangguru Terhadap Kepuasan Pengguna Menggunakan Metode EUCS (End-User Computing Satisfaction) dan IPA (Importance Performance Analysis). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *3*(9), 8496–8505.

Yusuf, B. P. (2023). *Analisis Sentimen Berbasis Aspek Terhadap Ulasan Aplikasi Market Place Menggunakan Bidirectional Long Short-Term Memory*. Universitas Komputer Indonesia.

DAFTAR RIWAYAT HIDUP

Nama	: Delta Apriza
Tempat/ tanggal lahir	: Manna /12-04-1997
Jenis Kelamin	: Perempuan
Agama	: Islam
Alamat	: Perumahan Green Center Park Keeya 3 No 9, Palembang.
Status Pernikahan	: Menikah
Handphone	085226406872
Email	: deltaaprz@gmail.com / deltaaprz@radenfatah.ac.id

PENDIDIKAN

SD	: SD Negeri 02 Bengkulu Selatan
SMP	: SMP Negeri 01 Bengkulu Selatan
SMA	: SMA Negeri 01 Bengkulu Selatan
S1	: Teknik Informatika Universitas Islam Indonesia Yogyakarta (2015-2019)
S2	: Magister Teknik Informatika Universitas Bina Darma Palembang (2023-2026)

PENGALAMAN KERJA

1. Pusat Data dan Teknologi Informasi BPOM RI (Tenaga IT) – (2021-2022)
2. UIN Raden Fatah Palembang (Ahli Pertama Pranata Komputer) – (2022 – Sekarang).

PROGRAM PASCASARJANA
TEKNIK INFORMATIKA – S2
UNIVERSITAS BINA DARMA

LEMBAR KONSULTASI TESIS

Nama : Delta Apriza
Nim : 232420017
Konsentrasi : ENTERPRISE IT INFRASTRUCTUR
Judul : Analisis Sentimen Berbasis Aspek pada Ulasan Platform Ruangguru Menggunakan Metode Long Short-Term Memory (LSTM)

Pembimbing : Dr. Usman Ependi, M.Kom.

No	Tanggal	Uraian Materi Konsultasi	Paraf
1.	11 Maret 2025	Tema Penelitian dan metode yang akan digunakan	
2.	1 April 2025	Revisi Tema Penelitian	
3	8 Mei 2025	Latar belakang, rumusan masalah, batasan Penelitian, SAP penelitian, Tujuan dan Manfaat penelitian	
4	20 Juni 2025	Tinjauan Pustaka dan landasan teori meliputi pembahasan Penelitian terkait	

PROGRAM PASCASARJANA
TEKNIK INFORMATIKA – S2
UNIVERSITAS BINA DARMA

LEMBAR KONSULTASI TESIS

Nama : Delta Apriza
Nim : 232420017
Konsentrasi : ENTERPRISE IT INFRASTRUCTUR
Judul : Analisis Sentimen Berbasis Aspek pada Ulasan Platform Ruangguru Menggunakan Metode Long Short-Term Memory (LSTM)

Pembimbing : Dr. Usman Ependi, M.Kom.

No	Tanggal	Uraian Materi Konsultasi	Paraf
5.	17 Juli 2025	Metodologi Penelitian meliputi Prosedur Penelitian dan Jadwal Penelitian	
6.	4 Agustus 2025	Acc Proposal Acc daftar Ujian Proposal Lanjutkan Bab 4 dan 5.	

PROGRAM PASCASARJANA
TEKNIK INFORMATIKA – S2
UNIVERSITAS BINA DARMA

LEMBAR KONSULTASI TESIS

Nama : Delta Apriza
Nim : 232420017
Konsentrasi : ENTERPRISE IT INFRASTRUCTUR
Judul : Analisis Sentimen Berbasis Aspek pada Ulasan Platform Ruangguru Menggunakan Metode Long Short-Term Memory (LSTM)

Pembimbing : Dr. Usman Ependi, M.Kom.

No	Tanggal	Uraian Materi Konsultasi	Paraf
1.	1 - 12 - 2025	Hasil Pre-Processing data dan Penentuan Aspek pada Ulasan platform Ruangguru	
2.	12 - 01 - 2026	Hasil Analisis Sentimen berbasis Aspek menggunakan Metode LSTM	
3.	27 - 01 - 2026	Laporan Bab IV hasil dan pembahasan serta Bab V kesimpulan dan saran	
4.	09 - 02 - 2026	Acc daftar ujian tesis lengkapi berkas sesuaikan dengan dokumen.	

PROGRAM PASCASARJANA
TEKNIK INFORMATIKA – S2
UNIVERSITAS BINA DARMA

LEMBAR KONSULTASI TESIS

Nama : Delta Apriza
Nim : 232420017
Konsentrasi : ENTERPRISE IT INFRASTRUCTUR
Judul : Analisis Sentimen Berbasis Aspek pada Ulasan Platform Ruangguru Menggunakan Metode Long Short-Term Memory (LSTM)

Pembimbing : Dr. Usman Ependi, M.Kom.

No	Tanggal	Uraian Materi Konsultasi	Paraf
	18-02-2026	Hasil Evaluasi Model Menggunakan Pelabelan Sentimen Otomatis dengan IndoBert	
	23-02-2026	Laporan tesis sesudah Revisi / Perbaikan Semhas	
	25-02-2026	draft artikel untuk publikasi	
	28-02-2026	Acc Ujian tesis beserta Revisi.	

SURAT KEPUTUSAN
DIREKTUR PROGRAM PASCASARJANA
NOMOR: 386/SK/PPs-UBD/MTI/X/2025

TENTANG
PEMBIMBING MAHASISWA
PROGRAM STUDI TEKNIK INFORMATIKA JENJANG STUDI STRATA DUA (S2)
PROGRAM PASCASARJANA UNIVERSITAS BINA DARMA
DIREKTUR PROGRAM PASCASARJANA
UNIVERSITAS BINA DARMA

Menimbang : a. Bahwa mahasiswa semester akhir diharuskan membuat Tesis sebagai salah satu syarat untuk menyelesaikan studi pada Program Pascasarjana Program Studi TEKNIK INFORMATIKA – S2;
b. Bahwa untuk kelancaran pelaksanaan kegiatan dimaksud, dipandang perlu untuk menunjuk Pembimbing bagi setiap mahasiswa;
c. Bahwa untuk memenuhi butir-butir di atas, perlu diterbitkan Surat Keputusan sebagai landasan hukumnya.

Mengingat : 1. Undang-undang Nomor 20 tahun 2003;
2. Undang-undang Nomor 12 tahun 2012;
3. Peraturan Mentri Pendidikan Nasional Nomor 15 tahun 2005;
4. Surat Keputusan Direktur Jenderal Pendidikan Tinggi Depdiknas;
5. Akte Pendirian Yayasan Nomor 95 tanggal 28 Desember 2003;
6. Statuta Universitas Bina Darma;
7. Surat Keputusan Rektor Universitas Bina Darma Nomor: 078/SK/Univ-BD/VI/2009 tanggal 1 Juni 2009.
8. Surat Keputusan Rektor Universitas Bina Darma Nomor: 0001/SK/Univ-BD/I/2019 tanggal 10 Januari 2019.

MEMUTUSKAN

Menetapkan :

PERTAMA : Menunjuk dan menugaskan saudara:

Dr. Usman Ependi, S.Kom., M.Kom

sebagai Pembimbing dalam penyusunan Tesis bagi mahasiswa dibawah ini:

Nama	:	DELTA APRIZA
NIM	:	232420017
Konsentrasi	:	ENTERPRISE SOFTWARE ENGINEERING
Judul Tesis	:	ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN PLATFORM RUANGGURU MENGGUNAKAN METODE LONG SHORT-TERM MEMORY (LSTM)

KEDUA : Surat Keputusan ini berlaku 6 (enam) bulan sejak tanggal ditetapkan dan apabila dalam waktu tersebut mahasiswa belum menyelesaikan, maka akan diterbitkan Surat Keputusan Pembimbing yang baru, dengan ketentuan apabila dikemudian hari terdapat kekeliruan dalam penetapan ini, akan diperbaiki sebagaimana mestinya;

KETIGA : Surat Keputusan asli ini diberikan kepada mahasiswa yang bersangkutan untuk dilaksanakan dan diindahkan sebagaimana mestinya.

Ditetapkan di: Palembang
Pada Tanggal: 15 Oktober 2025
Direktur,

Universitas Bina Darma
PROGRAM PASCASARJANA

Prof. Dr. Ir. Achmad Syarifudin, M.Sc.

Tembusan:
1. Pembimbing
2. Arsip

Universitas Bina Darma ISO 9001 : 2000	**FORMULIR** **Perbaikan Proposal Tesis**	NomorDok : NomorRevisi : Tgl. Berlaku : Klausa ISO :	

Nama : DELTA APRIZA

Nim : 232420017

Konsentrasi : ENTERPRISE SOFTWARE ENGINEERING

JudulTesis : Analisis Sentimen Berbasis Aspek pada Ulasan Platform Ruangguru Menggunakan Metode Long Short-Term Memory (LSTM)

Dosen Pembimbing : Dr. Usman Ependi, S.Kom., M.Kom

Tanggal Seminar : 15 Oktober 2025

Telah diperbaiki dan dikonsultasikan dengan Pembimbing/Penguji Proposal Tesis.

No.	NamaDosenpembimbing/Penguji	Tanggal	TandaPersetujuan
1.	Dr. Usman Ependi, S.Kom., M.Kom.		1. .....
2.	Dr. Yesi Novaria Kunang, S.T., M.Kom.		2. .....
3.	Nita Rosa Damayanti, M. Kom., Ph.D.		3. .....

Palembang, 27 oktober 2025
Program Studi Teknik Informatika – S2
Ketua,

Dr. Usman Ependi, M.Kom.

NB.
Harap segera menyerahkan foto copi formulir perbaikan proposal tesis sebanyak 2 lembar ke Sekretariat Pascasarjana, apabila telah di tanda tangan oleh penguji dan disahkan oleh Ketua Program Studi untuk proses pembuatan SK Pembimbing

Universitas Bina Darma ISO 9001 : 2000	**FORMULIR** **Perbaikan Hasil Tesis**	Nomor Dok : Nomor Revisi : Tgl. Berlaku : Klausa ISO :	

Nama : DELTA APRIZA

Nim : 232420017

Konsentrasi : ENTERPRISE IT INFRASTRUCTURE

Judul Tesis setelah : ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN PLATFORM RUANGGURU MENGGUNAKAN METODE *LONG SHORT-TERM MEMORY* (LSTM)

(Kepada mahasiswa harap di isi sesuai dengan hasil ujian hasil tesis).

Dosen Pembimbing : Dr. Usman Ependi, S.Kom., M.Kom

Tanggal Seminar : 20 Februari 2026

Telah diperbaiki dan dikonsultasikan dengan Pembimbing/Penguji Hasil Tesis.

No.	Nama Dosen pembimbing/Penguji	Tanggal	Tanda Persetujuan
1.	Dr. Usman Ependi, S.Kom., M.Kom		1. .....
2.	Dr. Yesi Novaria Kunang, S.T., M.Kom.		2. .....
3.	Nita Rosa Damayanti, M.Kom., Ph.D.		3. .....

Palembang, 23 Februari 2026
Program Studi Teknik Informatika – S2
Ketua,

Dr. Usman Ependi, S.Kom., M.Kom.

NB. Harap segera menyerahkan foto copy formulir perbaikan hasil tesis sebanyak 2 lembar ke Sekretariat Pascasarjana, apabila telah di tanda tangan oleh penguji dan disahkan oleh Ketua Program Studi untuk proses pembuatan SK Pembimbing

Universitas Bina Darma ISO 9001 : 2000	Formulir Perbaikan Tesis	Nomor Dok :	
		Nomor Revisi :	
		Tgl. Berlaku :	
		Klausa ISO :	

Nama : DELTA APRIZA

NIM : 232420017

Konsentrasi : ENTERPRISE IT INFRASTRUCTURE

Judul Tesis : ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN PLATFORM RUANGGURU MENGGUNAKAN METODE LONG SHORT-TERM MEMORY (LSTM)

Dosen Pembimbing : Dr. Usman Ependi, S.Kom., M.Kom.

Tanggal Ujian : 04 Maret 2026

Telah diperbaiki dan dikonsultasikan dengan Pembimbing/Penguji Tesis.

No.	Nama Dosen Penguji	Tanggal	Tanda Persetujuan
1.	**Dr. Usman Ependi, S.Kom., M.Kom.**		1.....
2.	**Dr. Yesi Novaria Kunang, S.T., M.Kom.**		2.....
3.	**Nita Rosa Damayanti, M.Kom., Ph.D.**		3.....

*Nb.
Pembimbing harap memeriksa kembali format dari tesis yang telah diperbaiki dan keabsahan tanda tangan penguji

Palembang, 04 Maret 2026
Program Studi Teknik Informatika – S2
Ketua,

Universitas Bina Darma
Magister Teknik Informatika

Dr. Usman Ependi, S.Kom., M.Kom.

Aspect-Based Sentiment Analysis of Ruangguru Platform Reviews Using Long Short-Term Memory (LSTM)

1st Delta Apriza
Master of Informatics Engineering
Bina Darma University
Palembang, Indonesia
deltaaprz@gmail.com

2nd Usman Efendi
Master of Informatics Engineering
Bina Darma University
Palembang, Indonesia
u.ependi@binadarma.ac.id

***Abstract*—User-generated reviews on online learning platforms constitute a valuable source of information for evaluating service quality and user experience across various service aspects. However, conventional sentiment analysis approaches remain limited in capturing contextual dependencies and aspect-level sentiment patterns within sequential textual data, thereby reducing interpretability and analytical reliability. To overcome this limitation, this study proposes an aspect-based sentiment analysis framework for Ruangguru platform reviews by integrating Latent Dirichlet Allocation for automatic aspect identification and Long Short-Term Memory networks for sentiment classification. The proposed framework classifies sentiment polarity into positive, neutral, and negative categories based on Indonesian-language user reviews. Experimental results demonstrate that training the LSTM model using a combined dataset of all identified aspects produces more stable and consistent classification performance compared to aspect-specific training, achieving an accuracy of 0.91 in the primary evaluation scenario. In contrast, separate aspect-level models tend to exhibit bias toward dominant sentiment classes due to data imbalance. The findings indicate that the integrated LDA-LSTM approach enhances classification robustness while providing structured insights into user perceptions across different service aspects. This study contributes to improving interpretability in aspect-level sentiment modeling and supports data-driven evaluation strategies for educational technology platforms.**

Keywords—Aspect Based Sentiment Analysis; Long Short-Term Memory; Deep Learning; Educational Technology; Sentiment Classification.

I. INTRODUCTION

The rapid advancement of information technology has transformed the accessibility and delivery of education, enabling flexible and technology-driven learning environments. In Indonesia, Ruangguru has become one of the largest online learning platforms, offering virtual classes, online examinations, subscription-based video learning, and other digital educational services [1]. Since the COVID-19 pandemic, the demand for online learning has increased significantly, positioning Ruangguru as one of the most widely adopted educational applications in the country [2].

As user adoption grows, maintaining service quality and understanding user satisfaction become critical challenges. User-generated reviews published on platforms such as Google Play Store contain valuable insights into user experiences and perceptions of different service aspects. Online reviews strongly influence decision-making behavior, as a substantial proportion of users consult reviews before selecting a product or service [3]. Consequently, systematic analysis of user feedback is essential for improving service performance and sustaining competitiveness [4].

Conventional sentiment analysis methods generally focus on overall polarity classification without distinguishing sentiment toward specific aspects of a service. Such approaches limit interpretability and reduce the practical value of the analysis for strategic decision-making. Aspect-Based Sentiment Analysis (ABSA) addresses this limitation by identifying and classifying opinions toward distinct service aspects [5]. However, user reviews often contain informal expressions, implicit meanings, and complex linguistic structures that challenge traditional word-based classification techniques [6]. Moreover, conventional machine learning models such as Naïve Bayes and Support Vector Machines are limited in capturing contextual dependencies and sequential word relationships within textual data.

Previous studies have applied ABSA using both machine learning and deep learning approaches across various domains [7], [3]. Data sources range from publicly available datasets to scraped online reviews [8]. Although LSTM-based models have demonstrated high accuracy in sentiment classification tasks [9], research focusing on Indonesian-language reviews of local educational technology platforms remains limited. In particular, the application of LSTM for aspect-based sentiment classification in the context of Ruangguru reviews has not been extensively explored.

To fill this gap, this study proposes the implementation of a Long Short-Term Memory model for aspect-based sentiment classification of Indonesian-language Ruangguru user reviews. By leveraging LSTM's capability to model sequential dependencies, this research aims to generate structured and interpretable insights into user perceptions across different service aspects. The findings are expected to support data-driven evaluation and strategic improvement of educational technology services.

II. METHODOLOGY RESEARCH

This study applies a Natural Language Processing and Aspect-Based Sentiment Analysis approach to systematically analyze user perceptions of the Ruangguru platform from

digital review data. The methodology is organized into sequential stages to ensure robust and interpretable sentiment classification, as illustrated in Figure 1.

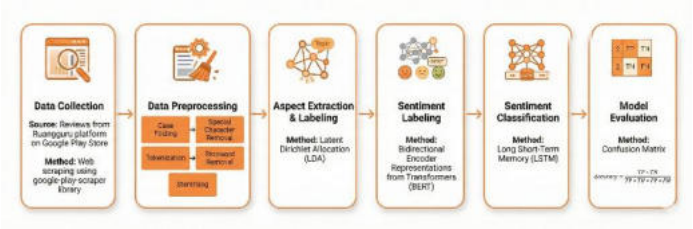


Fig. 1. Research Flow

A. Data Collection

User review data were collected from the Ruangguru application available on the Google Play Store using a web scraping technique implemented through the *google-play-scraper* library in Python. The scraping process extracted key attributes, including user ID, review date, rating, and review text. A total of 311,106 raw reviews were initially obtained, reflecting high user participation in providing feedback. To ensure temporal relevance, the dataset was filtered to include reviews posted within the last five years, from October 31, 2020 to October 31, 2025. After filtering, 45,765 reviews were retained for analysis. This study focuses on two primary variables. They are review date to capture temporal distribution and review text as the main input for aspect extraction and sentiment classification.

B. Preprocessing Data

The preprocessing stage is conducted to ensure that the review data used in the study is clean, structured, and ready for the subsequent stages of aspect extraction and sentiment classification. Raw data obtained from the Google Play Store is generally unstructured and contains various forms of noise, such as inconsistent capitalization, punctuation marks, emojis, special symbols, and informal word variations. Therefore, a series of preprocessing steps is applied, including case folding to standardize text into lowercase format, removal of special characters and emojis to eliminate irrelevant elements, tokenization to split text into individual word units, stopword removal to discard common non-informative words, and stemming or lemmatization to reduce words to their base forms. These steps aim to enhance the consistency of text representation, reduce data complexity, and improve the accuracy and stability of the model during topic modeling and sentiment classification stages.

C. Aspect Extraction and Labeling

Aspect extraction was conducted using an unsupervised topic modeling approach based on Latent Dirichlet Allocation (LDA). This probabilistic model represents each review as a mixture of latent topics, where each topic is characterized by a specific word distribution. Through this mechanism, frequently co-occurring terms are grouped into latent topics that are subsequently interpreted as service aspects discussed by users. To determine the optimal number of topics, several topic configurations were evaluated using the coherence score (C_v), a metric that has been shown to correlate strongly with human interpretability of topics [10]. Based on the evaluation results, the model with six topics achieved the highest coherence value and was therefore selected as the optimal configuration. Coherence score across different number of topics as show in figure 2.

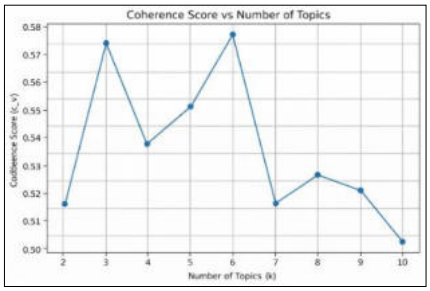


Fig. 2 Coherence Score Across Different Numbers of Topics

To enhance objectivity in aspect naming, this study employed a hybrid approach combining LDA-based topic modeling with Zero-Shot Text Classification using the DeBERTa transformer model. The extracted topics were mapped to the DeLone and McLean Information Systems Success Model, which encompasses information quality, system quality, service quality, and net benefits or user satisfaction [11]. This framework was selected because it emphasizes system success evaluation based on user-perceived impact and outcomes [12]. The extraction process identified six primary aspects: Academic Achievement, Content and Learning Media Quality, Service Access and Pricing, Application Features and Performance, Learning Support, and User Satisfaction, enabling aspect-level sentiment analysis rather than overall polarity classification.

Each review was subsequently labeled based on the highest topic probability generated by the LDA model. The distribution analysis shows that User Satisfaction was the most dominant aspect, followed by Academic Achievement and Content and Learning Media Quality. Learning Support and Application Features and Performance appeared at moderate frequency, while Service Access and Pricing had relatively lower frequency but remained relevant for service evaluation. The distribution of user reviews across dominant topics is presented in Figure. 3.

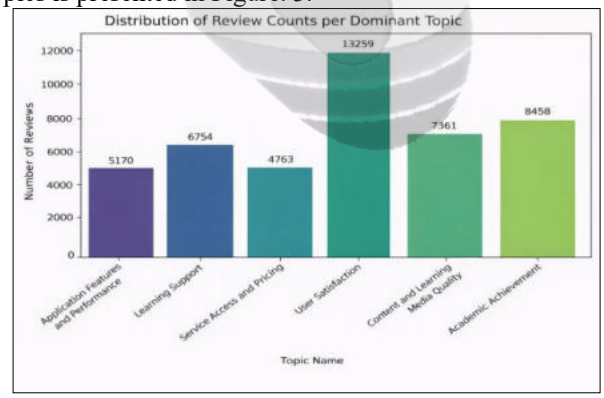


Fig. 3 Distribution of User Reviews Across Dominant Topics

D. Sentiment Labeling

Sentiment labeling was conducted to assign polarity labels to each review that had been mapped to a specific aspect. This study employed three sentiment categories: positive, negative, and neutral. The labeling process was performed automatically using a pre-trained multilingual Bidirectional Encoder Representations from Transformers (BERT) model to ensure consistency and objectivity across the large dataset. The selected model is capable of processing Indonesian-language reviews and capturing contextual meaning through bidirectional encoding, enabling more accurate sentiment prediction compared to conventional lexicon-based or classical machine learning approaches.

The sentiment labels generated by BERT were used as ground truth for training the Long Short-Term Memory (LSTM) model. Distribution analysis indicates that positive sentiment dominates most aspects, while neutral and negative sentiments are also present, providing sufficient class variation for effective model training.

E. Sentiment Classification

Aspect-based sentiment classification was conducted using the Long Short-Term Memory (LSTM) model to identify sentiment polarity (negative, neutral, and positive) for each review that had been assigned an aspect label and sentiment label generated by BERT as the ground truth. The review text was combined with its corresponding aspect label to provide contextual information, enabling the model to perform aspect-aware sentiment classification. The dataset was then split into training and testing sets using an 80:20 ratio to evaluate the model's generalization capability. Distribution of Aspects and Sentiment in Training and Testing Data as shown ini Tabel 1.

Tabel 1. Distribution of Aspects and Sentiment in Training and Testing Data

Label	Aspect	Number of Samples	
		Training Data	Testing Data
Negative	Service Access and Pricing	2194	548
	Academic Achievement	732	183
	Application Features and Performance	1713	428
	User Satisfaction	425	106
	Content and Learning Media Quality	455	114
	Learning Support	406	102
Neutral	Service Access and Pricing	198	49
	Academic Achievement	376	94
	Application Features and Performance	515	129
	User Satisfaction	1670	418
	Content and Learning Media Quality	134	33
	Learning Support	207	52
Positive	Service Access and Pricing	1745	436
	Academic Achievement	4295	1074
	Application Features and Performance	1582	396
	User Satisfaction	8512	2128
	Content and Learning Media Quality	5300	1325
	Learning Support	6153	1538

The proposed LSTM architecture consists of an Input Layer receiving sequences of length 100, followed by an Embedding Layer with 128-dimensional vector representations. A masking mechanism is applied to ignore padded tokens during training. The core LSTM layer then processes the sequential data and produces a contextual feature vector of 128 dimensions. To reduce overfitting, dropout layers are incorporated after the LSTM and dense hidden layers. A fully connected dense layer with ReLU activation refines the learned features, and the final output layer employs a softmax activation function to generate probability distributions over the three sentiment classes. LSTM Model Architecture as shown in Figure 4.

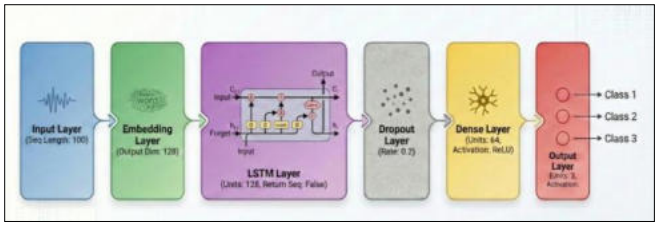


Fig. 4 LSTM Model Architecture

During training, the LSTM model processed data transformed into numerical token sequences and embedding vectors. Model parameters, including the forget, input, and output gate weights, were updated through backpropagation to learn sequential sentiment patterns and preserve contextual information. Two training scenarios were applied. In the first scenario, the model was trained on all aspects simultaneously to capture general sentiment patterns across the dataset. In the second scenario, separate LSTM models were trained for each individual aspect to learn more fine-grained sentiment patterns within specific service dimensions. Key hyperparameters, including the loss function, optimizer, imbalance handling, earlystopping, modelcheckpoint, learning rate, batch size, and number of epochs, were optimized to enhance performance.

F. Model Evaluation

After training, model validation was conducted using the testing dataset to evaluate generalization capability. The model's predictive performance was assessed using a confusion matrix, which provides a detailed evaluation of classification results across sentiment classes. The confusion matrix measures True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) predictions, offering deeper insights into classification behavior beyond single metrics such as accuracy or loss [13]. This evaluation approach is particularly suitable for aspect-based sentiment analysis involving multi-class classification and potentially imbalanced data distributions [14], [15]. Model accuracy was calculated using the following equation:

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

III. Result and Analysis

This section presents the results of the model evaluation, with a focus on the practical application of the methodology used. The search for the best model in sentiment classification was conducted by comparing the two applied training scenarioes.

A. Training Stability

In Scenario 1, where the model was trained on all aspects simultaneously, the accuracy and loss curves showed stable and gradual convergence. Training accuracy increased consistently while loss decreased in a controlled manner, with mild overfitting effectively managed through early stopping. These results indicate strong learning stability and generalization performance. In contrast, Scenario 2, which involved aspect-specific training, exhibited greater instability, particularly during the early epochs. The accuracy and loss curves fluctuated more noticeably, and improvements in validation accuracy were less consistent. This instability is likely due to smaller dataset sizes and class imbalance within each aspect-specific subset, limiting the

model's ability to capture sentiment dynamics effectively. Overall, the findings emphasize the importance of dataset size and diversity in ensuring stable LSTM training and robust generalization. The corresponding accuracy and loss curves are shown in Figures 5–11.

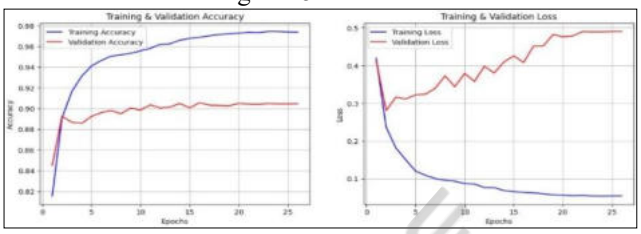


Fig. 5 Accuracy and Loss Curves in Scenario 1

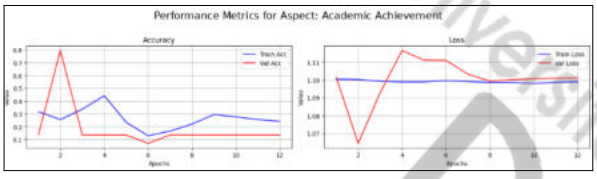


Fig. 6 Accuracy and Loss Curves for the Academic Achievement Aspect in Scenario 2

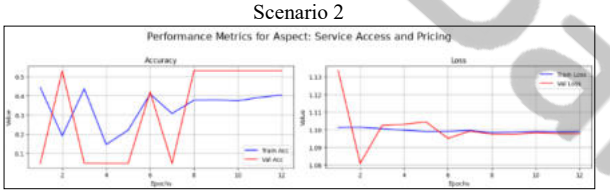


Fig. 7 Accuracy and Loss Curves for the Service Access and Pricing in Scenario 2

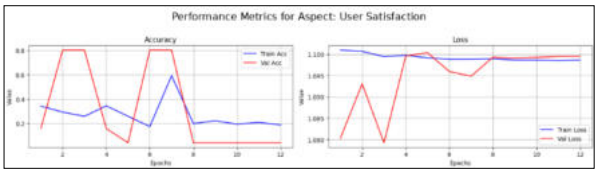


Fig. 8 Accuracy and Loss Curves for the User Satisfaction Aspect in Scenario 2

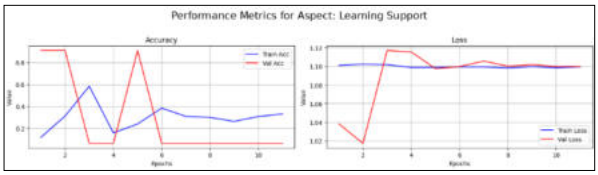


Fig. 9 Accuracy and Loss Curves for the Learning Support Aspect in Scenario 2

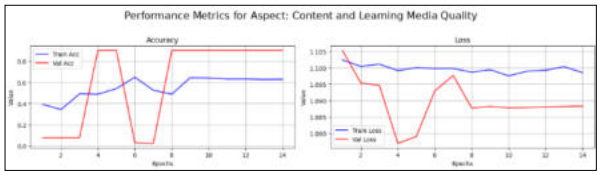


Fig. 10 Accuracy and Loss Curves for the Content and Learning Media Quality Aspect in Scenario 2

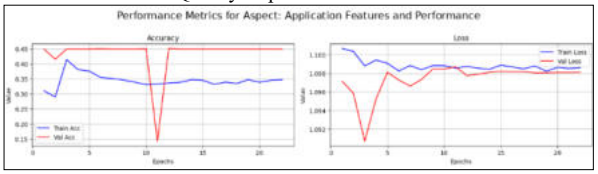


Fig. 11 Accuracy and Loss Curves for Application Features and Performance Aspect in Scenario 2

B. *Model Performance*

The comparison between the two training scenarios reveals a clear performance difference under combined validation. In Scenario 1, the LSTM model was trained on all aspects and validated on the combined testing dataset, achieving an overall accuracy of 0.91, which reflects strong generalization capability with a larger and more diverse dataset. In contrast, Scenario 2 involved aspect-specific training followed by validation on the combined dataset, resulting in a lower overall accuracy of 0.77 and reduced generalization performance. Additionally, Scenario 1 consistently outperformed Scenario 2 in terms of precision, recall, and F1-score. The detailed comparison of performance metrics for both scenarioes are presented in Table 2.

Tabel 2. LSTM Model Performance Validated on the Combined Testing Dataset Across All Aspects

Scenario	Precision	recall	f1-score	accuracy
1	0.82	0.88	0.84	0.91
2	0.45	0.51	0.48	0.77

When validated at the aspect level, Scenario 1 was trained on all aspects and evaluated separately for each aspect. The model achieved accuracy scores of 0.95 for Learning Support, 0.95 for User Satisfaction, 0.94 for Content and Learning Media Quality, 0.87 for Application Features and Performance, 0.84 for Academic Achievement, and 0.79 for Service Access and Pricing. These results indicate that the globally trained model maintained strong and relatively balanced performance across most service aspects. In contrast, Scenario 2 was trained separately for each aspect and validated on the corresponding aspect. Under this configuration, the accuracy scores were 0.95 for Learning Support, 0.95 for User Satisfaction, 0.94 for Content and Learning Media Quality, 0.87 for Application Features and Performance, 0.84 for Academic Achievement, and 0.79 for Service Access and Pricing. Although performance varied across aspects, the observed differences suggest that aspect-specific training may influence the model's generalization capability across heterogeneous sentiment patterns. A comparison of LSTM performance metrics validated separately for each corresponding aspect in both scenarioes are presented in Table 3.

Tabel 3. The Comparison of Aspect-Level Accuracy Between Scenario 1 and Scenario 2

Accuracy value	Scenario	
	1	2
Learning Support	0.95	0.91
User Satisfaction	0.95	0.80
Content and Learning Media Quality	0.94	0.90
Application Features and Performance	0.87	0.45
Academic Achievement	0.84	0.79
Service Access and Pricing	0.79	0.53

C. *Evaluation of Performance by Sentiment Class and Aspect*

The evaluation results using the confusion matrix indicate a significant difference in classification performance across sentiment classes between Scenario 1 and Scenario 2. In Scenario 1, the model trained using all identified aspects simultaneously demonstrates a more balanced classification capability across the three sentiment classes. The negative class was correctly classified at approximately 83.12%, the neutral class at 87.61%, and the positive class at 92.50%. Although some misclassifications were observed, particularly between negative and positive classes, the overall distribution of predictions remained relatively proportional. These results indicate that the model was able to effectively distinguish sentiment patterns when trained on a larger and more diverse dataset. In contrast, Scenario 2 exhibits a noticeable imbalance in class-level performance. The neutral

class was not optimally detected, showing nearly 0% correct classification. A substantial portion of neutral and negative instances was misclassified as positive. Although the positive class maintained a high classification rate, this dominance suggests a strong bias toward the majority class. This phenomenon indicates that aspect-specific training limits the model's ability to capture comprehensive sentiment distributions when evaluated on the combined dataset across all aspects. Confusion matrix for all aspects in scenario 1 and scenario 2 as shown in figure 12.

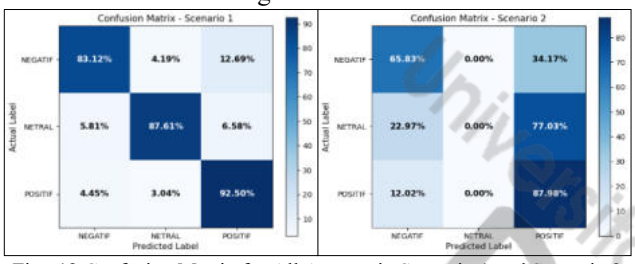


Fig. 12 Confusion Matrix for All Aspects in Scenario 1 and Scenario 2

The confusion matrix evaluation across individual aspects further reinforces the previous findings that Scenario 1 demonstrates more stable performance than Scenario 2. In Scenario 1, where the model was trained simultaneously on all aspects, it was able to classify negative, neutral, and positive sentiments in a relatively balanced manner across most aspects. Although some misclassifications occurred between classes, the prediction distribution remained proportional and did not exhibit extreme bias toward any single class. This indicates that training with a larger and more diverse dataset enables the model to develop more general and consistent sentiment representations across different aspect contexts. In contrast, Scenario 2 shows a strong tendency toward dominant-class bias in several aspects. In some cases, the model predicted nearly all instances as a single sentiment class, reflecting limited discriminative capability among sentiment categories. This pattern suggests that aspect-specific training increases the model's sensitivity to local data distributions within each aspect and reduces its generalization ability when encountering more heterogeneous sentiment patterns. Confusion matrix for Aspect-Level in scenario 1 and scenario 2 as shown in figure 15-20.

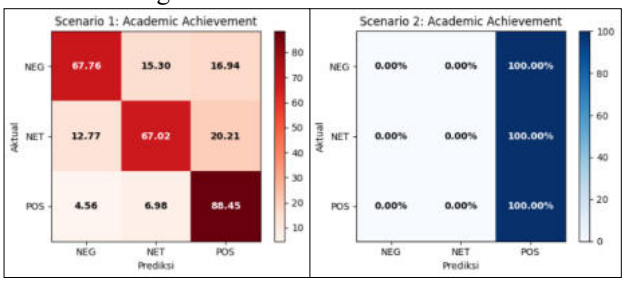


Fig. 13 Confusion Matrix for the Academic Achievement Aspect

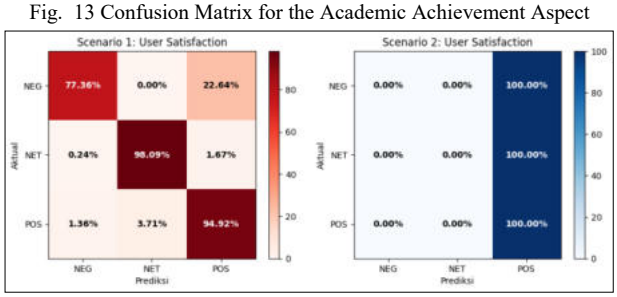


Fig. 14 Confusion Matrix for the User Satisfaction Aspect

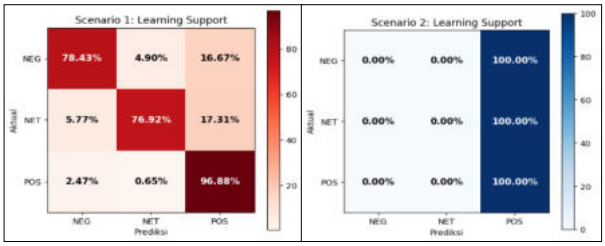


Fig. 15 Confusion Matrix for the Learning Support Aspect

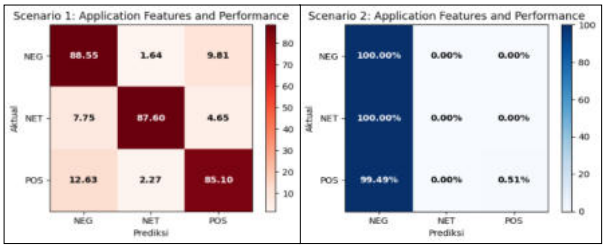


Fig. 16 Confusion Matrix for the Features and Application Performance Aspect

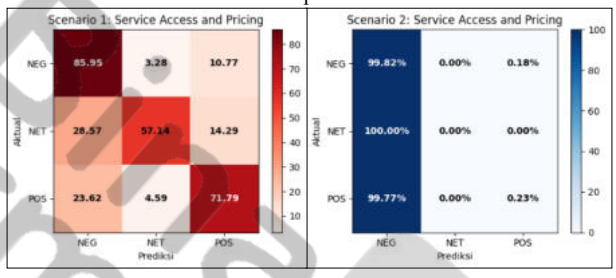


Fig. 17 Confusion Matrix for the Service Access and Pricing Aspect

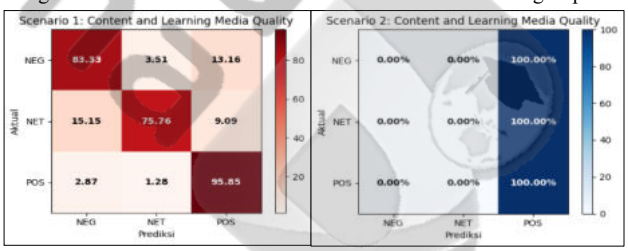


Fig. 18 Confusion Matrix for the Content and Learning Media Quality Aspect

D. *Discussion*

The results confirm that Scenario 1 achieves superior performance compared to Scenario 2, indicating that training on a larger and more diverse multi-aspect dataset enhances the generalization capability of LSTM models. Joint training across aspects enables the model to capture shared semantic patterns, resulting in more stable and balanced sentiment classification. Conversely, aspect-specific training in Scenario 2 reduces contextual diversity and increases sensitivity to class imbalance, leading to dominant-class bias and weaker discrimination, particularly for the neutral class. Since LSTM relies on contextual variation to optimize its gated memory mechanisms, limited data per aspect constrains representation learning.

Theoretically, this study contributes to ABSA research by demonstrating that integrated multi-aspect learning improves sentiment representation in neural models. The findings highlight the influence of dataset structure and class distribution on LSTM performance. Practically, the results suggest that digital education platforms benefit more from unified sentiment modeling to ensure stable feedback monitoring and data-driven service improvement.

IV. Conslusion

This study confirms that the Long Short-Term Memory (LSTM) model can be effectively applied to aspect-based sentiment analysis of user reviews on the Ruangguru platform. By integrating review data collection, text preprocessing, aspect extraction using Latent Dirichlet Allocation (LDA), and automatic sentiment labeling with BERT, the proposed framework successfully performed aspect-level sentiment classification. Comparative results from two training scenarios show that training the LSTM model using all aspects simultaneously yields more stable and consistent performance than training separate models per aspect. The combined-aspect approach achieved better generalization due to larger and more diverse data distribution, whereas aspect-specific training suffered from class imbalance and bias toward dominant sentiment classes. These findings emphasize the importance of sufficient data volume and balanced sentiment distribution in improving LSTM-based sentiment classification performance.Authors and Affiliations

References

[1] E. Fitri, Y. Yuliani, S. Rosyida, and W. Gata, "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine," *Transformtika*, vol. 18, no. 1, pp. 71–80, Jul. 2020.

[2] M. Ali *et al.*, *Teknologi Pendidikan Teori Dan Aplikasi*, Cetakan Pertama. Azzia Karya Bersama, 2025.

[3] S. Cahyaningtyas, "Deep Learning pada Aspect-based Sentiment Analysis pada Data Hotel Berbahasa Indonesia," Universitas Islam Indonesia, 2021.

[4] A. H. Hasugian, M. Fakhriza, and D. Zukhoiriyah, "Analisis Sentimen Pada Review Pengguna E-Commerce Menggunakan Algoritma Naive Bayes," *Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD*, vol. 6, pp. 98–107, Jan. 2023.

[5] Sparta and S. Rimadias, "Optimalisasi Artificial Intelligence (AI) Pada Platform SINTA," 2024.

[6] P. Warakmulty, D. Yeffry, and H. Putra, "Optimalisasi Manajemen Sentimen di Media Sosial Universitas melalui Machine Learning dan AI: Studi Kasus pada Komentar Instagram," *Jurnal Tata Kelola dan Kerangka Kerja Teknologi Informasi*, vol. 11, no. 1, pp. 31–38, May 2025.

[7] Y. Asmara, M. Rudyanto Arief, and kusrini, "Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi RuangGuru Menggunakan Support Vector Machine dan Naive," *Februari*, vol. 2024, no. 2, pp. 209–215, 2024.

[8] R. Nurhidayat and K. E. Dewi, "Penerapan Algoritma K-Nearest Neighbor Dan Fitur Ekstraksi N-Gram Dalam Analisis Sentimen Berbasis Aspek," *Komputa : Jurnal Ilmiah Komputer dan Informatika*, vol. 12, no. 1, pp. 91–100, 2023, doi: 10.34010/komputa.v12i1.9458.

[9] A. R. Ramadhan, "Analisis Sentimen Ulasan Aplikasi Ruangguru dengan Algoritma Long Short Term Memory," UIN Syarif Hidayatullah, Jakarta, 2023.

[10] M. Röder, A. Both, and A. Hinneburg, "Exploring the space of topic coherence measures," in *WSDM 2015 - Proceedings of the 8th ACM International Conference on Web Search and Data Mining*, Association for Computing Machinery, Feb. 2015, pp. 399–408. doi: 10.1145/2684822.2685324.

[11] S. Petter, W. DeLone, and E. McLean, "Measuring information systems success: Models, dimensions, measures, and interrelationships," *European Journal of Information Systems*, vol. 17, no. 3, pp. 236–263, 2008, doi: 10.1057/ejis.2008.15.

[12] N. Urbach and B. Müller, "The Updated DeLone and McLean Model of Information Systems Success," 2012, pp. 1–18. doi: 10.1007/978-1-4419-6108-2_1.

[13] B. Liu, "Sentiment Analysis and Opinion Mining," Morgan & Claypool Publishers, May 2012.

[14] A. Rolangon, A. Weku, and G. A. Sandag, "Perbandingan Algoritma LSTM Untuk Analisis Sentimen Pengguna Twitter Terhadap Layanan Rumah Sakit Saat Pandemi Covid-19," *Jurnal TeIKa*, vol. 13, pp. 31–40, 2023.

[15] R. Naquitasia, D. Hatta Fudholi, and L. Iswari, "Analisis Sentimen Berbasis Aspek Pada Wisata Halal Dengan Metode Deep Learning," *Jurnal Teknoinfo*, vol. 16, no. 2, pp. 156–164, Jul. 2022.

2026 International Conference on Information Management and Technology (ICIMTech) : Submission (36) has been edited.

Inbox ×

Microsoft CMT <noreply@msr-cmt.org>
to me

Mon, Mar 9, 12:04 PM

Hello,

The following submission has been edited.

Track Name: ICIMTech2026

Paper ID: 36

Paper Title: Aspect-Based Sentiment Classification of Education Platform Reviews Using Long Short-Term Memory

Abstract:
User-generated reviews on online learning platforms constitute a valuable source of information for evaluating service quality and user experience across various service aspects. However, conventional sentiment analysis approaches remain limited in capturing contextual dependencies and aspect-level sentiment patterns within sequential textual data, thereby reducing interpretability and analytical reliability. To overcome this limitation, this study proposes an aspect-based sentiment analysis framework for Learning platform reviews by integrating Latent Dirichlet Allocation for automatic aspect identification and Long Short-Term Memory networks for sentiment classification. The proposed framework classifies sentiment polarity into positive, neutral, and negative categories based on Indonesian-language user reviews. Experimental results demonstrate that training the LSTM model using a combined dataset of all identified aspects produces more stable and consistent classification performance compared to aspect-specific training, achieving an accuracy of 0.91 in the primary evaluation scenario. In contrast, separate aspect-level models tend to exhibit bias toward dominant sentiment classes due to data imbalance. The findings indicate that the integrated LDA-LSTM approach enhances classification robustness while providing structured insights into user perceptions across different service aspects. This study contributes to improving interpretability in aspect-level sentiment modeling and supports data-driven evaluation strategies for educational technology platforms.

Created on: Mon, 09 Mar 2026 04:58:46 GMT

Last Modified: Mon, 09 Mar 2026 05:03:46 GMT

Reply Forward

SURAT KETERANGAN

Nomor : 050/PPs-UBD/MTI/III/2026

Direktur Program Pascasarjana Universitas Bina Darma menerangkan bahwa :

Nama	:	DELTA APRIZA
NIM	:	232420017
Konsentrasi	:	ENTERPRISE IT INFRASTRUCTURE

Telah menyelesaikan studinya di Program Pascasarjana Program Studi Teknik Informatika – S2 Universitas Bina Darma dan dinyatakan **LULUS** pada hari **Rabu**, tanggal **4 Maret 2026** dengan tesis berjudul :

"ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN PLATFORM RUANGGURU MENGGUNAKAN METODE LONG SHORT-TERM MEMORY (LSTM)"

Dan yang bersangkutan juga telah berhak untuk menggunakan gelar akademik Strata – 2 (S2) dengan sebutan **MAGISTER KOMPUTER (M.KOM)**.

Demikian Surat Keterangan ini dibuat untuk dipergunakan sebagaimana mestinya.

Dikeluarkan di : Palembang
Pada Tanggal : 4 Maret 2026
Direktur,

Universitas Bina Darma
PROGRAM PASCASARJANA

Prof. Dr. Ir. Achmad Syarifudin, M.Sc.

Cc. Arsip

ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN PLATFORM RUANGGURU MENGGUNAKAN METODE LONG SHORT-TERM MEMORY (LSTM)

ORIGINALITY REPORT

25%	20%	17%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Sriwijaya University Student Paper	2%
2	repository.binadarma.ac.id Internet Source	1%
3	dspace.uii.ac.id Internet Source	1%
4	etheses.uin-malang.ac.id Internet Source	1%
5	repository.dinamika.ac.id Internet Source	1%
6	Masriah Masriah, Wahyu Tisno Atmojo. "Analisis Sentimen Terhadap Ulasan pada Aplikasi Astro Menggunakan Algoritma Support Vector Machine", Jurnal Komtika (Komputasi dan Informatika), 2025 Publication	1%
7	jurnal.unai.edu Internet Source	<1%
8	ejurnal.seminar-id.com Internet Source	<1%
9	journal.aptii.or.id Internet Source	<1%

10	Mohammad Bayu Anggara. "Comparison of Naïve Bayes and SVM Methods in Sentiment Analysis of User Reviews on the RSUD AL IHSAN Mobile Application", Competitive, 2025 Publication	<1 %
11	jurnal.polibatam.ac.id Internet Source	<1 %
12	Paolo Ferro, Harinadh Vemanaboina, Chander Prakash. "Computational Techniques and Smart Manufacturing", CRC Press, 2026 Publication	<1 %
13	Submitted to Universitas Putera Batam Student Paper	<1 %
14	etd.repository.ugm.ac.id Internet Source	<1 %
15	digilib.uinsby.ac.id Internet Source	<1 %
16	repository.upi.edu Internet Source	<1 %
17	jurnal.tau.ac.id Internet Source	<1 %
18	repository.its.ac.id Internet Source	<1 %
19	kc.umn.ac.id Internet Source	<1 %
20	repositori.telkomuniversity.ac.id Internet Source	<1 %
21	repository.umsu.ac.id Internet Source	<1 %
22	repository.unissula.ac.id Internet Source	

		<1 %
23	eprints.amikompurwokerto.ac.id Internet Source	<1 %
24	Zahrotun Ni'mah -, Bambang Irawan -, Nur Ariesanto Ramdhan -. "ANALISIS SENTIMEN ULASAN PENGGUNA TERHADAP GAME ZENLESS ZONE ZERO MENGGUNAKAN METODE BI-DIRECTIONAL LSTM", Jurnal Informatika dan Teknik Elektro Terapan, 2026 Publication	<1 %
25	adoc.pub Internet Source	<1 %
26	ejournal.kresnamediapublisher.com Internet Source	<1 %
27	repository.unsri.ac.id Internet Source	<1 %
28	hohpublica.uni-hohenheim.de Internet Source	<1 %
29	repositori.unsil.ac.id Internet Source	<1 %
30	www.ejurnal.dipanegara.ac.id Internet Source	<1 %
31	hrcak.srce.hr Internet Source	<1 %
32	id.123dok.com Internet Source	<1 %
33	journal.amikveteran.ac.id Internet Source	<1 %

34	Eti Kurniawati, Ade Irma Purnamasari, Irfan Ali, Rudi Kurniawan, Odi Nurdiawan. "ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI FLO DI GOOGLE PLAY STORE DENGAN MENGGUNAKAN ALGORITMA NAIVE BAYES", Jurnal Informatika dan Teknik Elektro Terapan, 2026 Publication	<1 %
35	Fransesko Indrajid -, I Nyoman Saputra Wahyu Wijaya -. "MODEL KLASIFIKASI TWEET TERKAIT ISU #INDONESIAGELAP MENGGUNAKAN SUPPORT VECTOR MACHINE BERBASIS DISTRIBUSI TOPIK LDA", Jurnal Informatika dan Teknik Elektro Terapan, 2026 Publication	<1 %
36	Submitted to Universitas Muhammadiyah Sumatera Utara Student Paper	<1 %
37	docplayer.info Internet Source	<1 %
38	eprints.uns.ac.id Internet Source	<1 %
39	journal.fkpt.org Internet Source	<1 %
40	journal.institutpendidikan.ac.id Internet Source	<1 %
41	www.jurnal.polgan.ac.id Internet Source	<1 %
42	Ichsani Mursidah, Remi Sanjaya, Bambang Yulianto, Dhian Sweetania, Puji Sularsih. "Klasifikasi Sentimen Google Play Store Aplikasi ChatGPT Berbahasa Indonesia	<1 %

Berbasis IndoBERT", Jurnal Minfo Polgan, 2025
Publication

43	cdn.repository.uisi.ac.id Internet Source	<1 %
44	informatics.uii.ac.id Internet Source	<1 %
45	dspace.mackenzie.br Internet Source	<1 %
46	Submitted to Institut Pertanian Bogor Student Paper	<1 %
47	Ningrum Astriawati. "Development of interactive media based on videoscribe with realistic mathematics education approach to navigation", Math Didactic: Jurnal Pendidikan Matematika, 2020 Publication	<1 %
48	Submitted to Universitas Brawijaya Student Paper	<1 %
49	Submitted to Universitas Muhammadiyah Sukabumi Student Paper	<1 %
50	jutif.if.unsoed.ac.id Internet Source	<1 %
51	Submitted to Universitas Islam Indonesia Student Paper	<1 %
52	jurnal.ppjb-sip.org Internet Source	<1 %
53	elib.unikom.ac.id Internet Source	<1 %

www.researchsquare.com

54	Internet Source	<1 %
55	Ananda Rizki Dani, Irma Handayani. "Classification of Yogyakarta Batik Motifs Using GLCM and CNN Methods", Jurnal Teknologi Terpadu, 2024 Publication	<1 %
56	eprints.upj.ac.id Internet Source	<1 %
57	repository.ubpkarawang.ac.id Internet Source	<1 %
58	Submitted to UPN Veteran Yogyakarta Student Paper	<1 %
59	indojurnal.com Internet Source	<1 %
60	publikasi.teknokrat.ac.id Internet Source	<1 %
61	Submitted to Perpustakaan Student Paper	<1 %
62	kr.co.id Internet Source	<1 %
63	Submitted to UNIVERSITAS BUDI LUHUR Student Paper	<1 %
64	Submitted to Universitas Jember Student Paper	<1 %
65	Submitted to Universitas Raharja Student Paper	<1 %
66	eprints.umpo.ac.id Internet Source	<1 %
67	jip.polinema.ac.id	

	Internet Source	<1 %
68	jurnal.ugj.ac.id Internet Source	<1 %
69	repository.unej.ac.id Internet Source	<1 %
70	journal.literasisains.id Internet Source	<1 %
71	journal.ppmi.web.id Internet Source	<1 %
72	ojs.trigunadharma.ac.id Internet Source	<1 %
73	repository.unmuhjember.ac.id Internet Source	<1 %
74	www.jurnalmahasiswa.com Internet Source	<1 %
75	Dewi Arumsari, Kharisma, Ulfi Saidata Aesyi. "Sistem Chatbot Layanan Informasi Mahasiswa Menggunakan Algoritma Long Short-Term Memory", INDONESIAN JOURNAL ON DATA SCIENCE, 2025 Publication	<1 %
76	Submitted to Fakultas Ekonomi Universitas Indonesia Student Paper	<1 %
77	dokumen.pub Internet Source	<1 %
78	ejournal3.undip.ac.id Internet Source	<1 %
79	journal.ipm2kpe.or.id Internet Source	<1 %

80	lib.unnes.ac.id Internet Source	<1 %
81	Citra Fathia Palembang, Emanuella M. C. Wattimena, Vyarlita Isqama Fataruba. "Analisis Sentimen Dosen terhadap Kebijakan Tunjangan Kinerja di Perguruan Tinggi Negeri Menggunakan Algoritma Naive Bayes", ALGORITHM: Journal of Computer Science and Computational Intelligence, 2025 Publication	<1 %
82	Ari Fullah Author, Muhammad Arif Septian Author, Muhammad Rizky Haritama Putra Author, Muhammad Iqbal Fadiatama Author et al. "ANALISIS SENTIMEN ISU PENYALAHGUNAAN DATA PADA LAYANAN PINJAMAN ONLINE MENGGUNAKAN SUPPORT VECTOR MACHINE DI PLATFORM X", Jurnal Informatika dan Teknik Elektro Terapan, 2025 Publication	<1 %
83	Muhammad Irfani, Siti Khomsah. "Analisis Sentimen Berbasis Aspek pada EDOM Pembelajaran Menggunakan Metode CNN dan Word2vec", Jurnal Sistem dan Teknologi Informasi (JustIN), 2024 Publication	<1 %
84	ojs.stmik-banjarbaru.ac.id Internet Source	<1 %
85	Hendri Mahmud Nawawi, Agung Baitul Hikmah, Ali Mustopa, Ganda Wijaya. "Model Klasifikasi Machine Learning untuk Prediksi Ketepatan Penempatan Karir", Jurnal SAINTEKOM, 2024 Publication	<1 %

86	Khirzin, Fazaa Daffa Al. "Strategi Pembelajaran Remedial Teaching Dalam Mengembangkan Mata Pelajaran Al Qur'an Hadits (Studi Kasus Di Mts Ma'arif Mandiraja, Banjarnegara).", Universitas Islam Negeri Saifuddin Zuhri (Indonesia) Publication	<1 %
87	Submitted to Universitas Bina Darma Student Paper	<1 %
88	journals.usm.ac.id Internet Source	<1 %
89	Desi Permata Sari, Bambang Irawan. "Implementasi CNN Mobilenetv2 untuk Klasifikasi Kanker Kulit Dermatoskopi Digital Medis", RIGGS: Journal of Artificial Intelligence and Digital Business, 2026 Publication	<1 %
90	Harshit Shah, Dhruvil Shah, Nilesh Kumar Jadav, Rajesh Gupta et al. "Deep Learning-Based Malicious Smart Contract and Intrusion Detection System for IoT Environment", Mathematics, 2023 Publication	<1 %
91	Submitted to Universitas Islam Riau Student Paper	<1 %
92	backend.orbit.dtu.dk Internet Source	<1 %
93	ejournal-binainsani.ac.id Internet Source	<1 %
94	pengumuman-property.blogspot.com Internet Source	<1 %

95	Kasih Delayana, Setia Mangiring Marpaung, Syukur Nimei Iman Gulo, Sardo Sipayung. "Analisis Sentimen Publik terhadap Barak Militer Menggunakan Naive Bayes pada Rapid MinerJudul", remik, 2026 Publication	<1 %
96	Submitted to Universitas Muhammadiyah Jember Student Paper	<1 %
97	e-journal.upr.ac.id Internet Source	<1 %
98	repo.iain-padangsidimpuan.ac.id Internet Source	<1 %
99	repository.stipram.ac.id Internet Source	<1 %
100	sekarars15.blogspot.com Internet Source	<1 %
101	Annisa Maizaliyanti, Khothibul Umam, Wenty Dwi Yuniarti, Maya Rini Handayani. "Implementasi Algoritma Random Forest dalam Klasifikasi Ulasan Pengunjung Mall Semarang untuk Pengambilan Keputusan Layanan", Edumatic: Jurnal Pendidikan Informatika, 2025 Publication	<1 %
102	Submitted to Badan PPSDM Kesehatan Kementerian Kesehatan Student Paper	<1 %
103	Dedi Irawan, Yayu Sri Rahayu, Mohamad Farozi. "KLASIFIKASI SENTIMEN PELANGGAN RESTORAN KOKI SUNDA MENGGUNAKAN	<1 %

ALGORITMA NAIVE BAYES", Jurnal Manajemen Informatika dan Sistem Informasi, 2025
Publication

104	Nanda Dwi Husna Sadikin, Sari Susanti. "Analisis Sentimen Publik Terhadap Kampanye Pengurangan Sampah Plastik Menggunakan Algoritma Naïve Bayes", JURNAL FASILKOM, 2025 Publication	<1 %
105	Rahandya Cahyo, Alda Cendekia Siregar, Barry Ceasar Octariadi. "Implementasi Ant Colony Optimization Untuk Rute Terpendek Pada Pengiriman Barang J&T", Jurnal CoSciTech (Computer Science and Information Technology), 2025 Publication	<1 %
106	www.coursehero.com Internet Source	<1 %
107	Nelsi Silaban, Marliza Ganefi Gumay. "Analisis Sentimen Pengguna Aplikasi Pinjaman Online Melalui Media Sosial X (Twitter) Menggunakan Metode Support Vector Machine", Jurnal Minfo Polgan, 2025 Publication	<1 %
108	Submitted to Sim University Student Paper	<1 %
109	Singgih Jatmiko, Charles Dometian. "Analisis Sentimen Ulasan Google Play Store: Studi Komparatif Algoritma SVM, Naïve Bayes, dan Logistic Regression", JURNAL FASILKOM, 2026 Publication	<1 %
110	Submitted to Syiah Kuala University Student Paper	<1 %

111	eprints.binadarma.ac.id Internet Source	<1 %
112	repo.darmajaya.ac.id Internet Source	<1 %
113	repository.uin-suska.ac.id Internet Source	<1 %
114	repository.unand.ac.id Internet Source	<1 %
115	Aang Alim Murtopo, Maulana Aditdya, Pingky Septiana Ananda, Gunawan Gunawan. "PENERAPAN COMPUTER VISION UNTUK MENDETEKSI KELENGKAPAN ATRIBUT SISWA MENGGUNAKAN METODE CNN", PROSISKO: Jurnal Pengembangan Riset dan Observasi Sistem Komputer, 2024 Publication	<1 %
116	Bela Agustina, Auliya Rahman Isnain. "Penerapan Deep Learning pada Sistem Klasifikasi Tumor Otak Berbasis Citra CT Scan", Indonesian Journal of Computer Science, 2024 Publication	<1 %
117	Submitted to Ciputra University Student Paper	<1 %
118	Nazzel Maulana Mustofa, Ahmad Muharram Alfarisi, Abu Tholib. "Penerapan Algoritma Apriori Untuk Menentukan Pola Pembelian Konsumen", JSITIK: Jurnal Sistem Informasi dan Teknologi Informasi Komputer, 2025 Publication	<1 %
119	Raisya Nadzira Zahirahtush Shafa, Asti Herliana. "ANALISIS SENTIMEN PENGGUNA	<1 %

APLIKASI SAPAWARGA - JABAR SUPER APPS MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE", Djtechno: Jurnal Teknologi Informasi, 2025
Publication

120	Submitted to Universitas Bengkulu Student Paper	<1 %
121	Submitted to Universitas Pancasila Student Paper	<1 %
122	Submitted to University of East London Student Paper	<1 %
123	jumanji.unjani.ac.id Internet Source	<1 %
124	Submitted to UIN Walisongo Student Paper	<1 %
125	Submitted to Universitas Gadjah Mada Student Paper	<1 %
126	Wajdi, Bina Arumbinang. "Pengaruh lamanya perendaman terhadap absorpsi, ketahanan aus, dan kuat tekan paving block", Universitas Islam Sultan Agung (Indonesia), 2023 Publication	<1 %
127	core.ac.uk Internet Source	<1 %
128	eprints.uny.ac.id Internet Source	<1 %
129	radarjogja.jawapos.com Internet Source	<1 %
130	Muhammad Raffi, Aries Suharso, Iqbal Maulana. "Analisis Sentimen Ulasan Aplikasi Binar Pada Google Play Store Menggunakan	<1 %

Algoritma NaÃ¯ve Bayes", INTECOMS: Journal of Information Technology and Computer Science, 2023
Publication

131	Sebastian Ubaidillah Royan, Nana Suarna, Irfan Ali, Dodi Solihudin. "ANALISIS SENTIMEN ULASAN PRODUK SKINCARE DI SHOPEE UNTUK MENINGKATKAN KUALITAS PRODUK MENGGUNAKAN METODE SUPPORT VECTOR MACHINE", Jurnal Informasi dan Komputer, 2025 Publication	<1 %
132	digilib.itb.ac.id Internet Source	<1 %
133	dspace.univ-guelma.dz Internet Source	<1 %
134	ejurnal.umri.ac.id Internet Source	<1 %
135	jti.aisyahuniversity.ac.id Internet Source	<1 %
136	repo.undiksha.ac.id Internet Source	<1 %
137	D. Diffran Nur Cahyo, Andi Sunyoto. "Analisis Perbandingan Klasifikasi dalam Data Mining pada Prediksi Hujan dengan menggunakan Algoritma LSTM dan GRU", Jurnal Sains dan Informatika, 2025 Publication	<1 %
138	Fauzan Baehaqi, Nuri Cahyono. "Analisis Sentimen Terhadap Cyberbullying Pada Komentar Di Instagram Menggunakan	<1 %

Algoritma Naïve Bayes", Indonesian Journal of Computer Science, 2024
Publication

139	Fityan Hanif Assalmi, Wahyu Syaifullah Jauharis Saputra, Amri Muhaimin. "Analisis Prediksi Kata Kunci Situs Web MonsterMAC dengan Metode Long Short-Term Memory (LSTM)", Jurnal Teknologi Terpadu, 2024 Publication	<1 %
140	Muhammad Sholeh, Uniing Lestari, Dina Andayati. "Classification of Customer Opinions on the Quality of Cooperative Minimarket Services Using the Lexicon Approach", International Journal of Engineering and Computer Science Applications (IJECSA), 2026 Publication	<1 %
141	Yeni Kustiyahningsih, Yohan Permana. "Penggunaan Latent Dirichlet Allocation (LDA) dan Support-Vector Machine (SVM) Untuk Menganalisis Sentimen Berdasarkan Aspek Dalam Ulasan Aplikasi EdLink", Teknika, 2024 Publication	<1 %
142	blog.ambisius.com Internet Source	<1 %
143	eprints.walisongo.ac.id Internet Source	<1 %
144	garuda.kemdiktisaintek.go.id Internet Source	<1 %
145	journal.stmikjayakarta.ac.id Internet Source	<1 %
146	kth.diva-portal.org Internet Source	<1 %

147	nlp.cs.aueb.gr Internet Source	<1 %
148	rapik.pubmedia.id Internet Source	<1 %
149	repository.machung.ac.id Internet Source	<1 %
150	repository.unhas.ac.id Internet Source	<1 %
151	repository.unpar.ac.id Internet Source	<1 %
152	text-id.123dok.com Internet Source	<1 %
153	Beny Alphon Tondang, Muhammad Rizqan Fadhil, Muhammad Nugraha Perdana, Akhmad Fauzi, Ugra Syahda Janitra. "Analisis pemodelan topik ulasan aplikasi BNI, BCA, dan BRI menggunakan latent dirichlet allocation", INFOTECH : Jurnal Informatika & Teknologi, 2023 Publication	<1 %
154	Mohamad Jamil, Rosihan Rosihan, Masdar S Hanafi. "Analisis Sentimen Calon Kepala Daerah Maluku Utara dengan Metode CRISP-DM", bit-Tech, 2025 Publication	<1 %
155	Murtiwiyati Murtiwiyati, Rischa Fitriaswhalia Arini, Leli Safitri, Santi Widianti, Iwan Setiadi. "Implementasi Algoritma K-Nearest Neighbors untuk Klasifikasi Sentimen pada Ulasan Pakaian Anak di Platform TikTok Shop", Jurnal Minfo Polgan, 2025 Publication	<1 %

156	Nishu Gupta, Sandeep S. Joshi, Milind Khanapurkar, Asha Gedam, Nikhil Bhave. "Recent Advances in Science, Engineering and Technology (RASET-2023) - Proceedings of the International Conference on Recent Advances in Science, Engineering & Technology, 29–30 September 2023", CRC Press, 2024 Publication	<1 %
157	Nurhayati Nurhayati, Lili Tanti, Budi Triandi. "Optimasi Support Vector Machine Menggunakan Particle Swarm Optimization pada Analisis Sentimen Program Efisiensi Anggaran Pemerintah", Jurnal Minfo Polgan, 2026 Publication	<1 %
158	Revi Setiawan, Bayu Priyatna, Elfina Novalia, Baenil Huda. "Klasifikasi Tingkat Penjualan Produk pada Toko Jati Karebet Menggunakan Algoritma Naïve Bayes", JURNAL FASILKOM, 2025 Publication	<1 %
159	addi.ehu.es Internet Source	<1 %
160	aimos.ugm.ac.id Internet Source	<1 %
161	docplayer.net Internet Source	<1 %
162	gudangilmu2016.wordpress.com Internet Source	<1 %
163	journal.amikindonesia.ac.id Internet Source	<1 %
164	journal.institercom-edu.org	

	Internet Source	<1 %
165	journal.irpi.or.id Internet Source	<1 %
166	jurnal.fmipa.unila.ac.id Internet Source	<1 %
167	jurnal.itscience.org Internet Source	<1 %
168	proceeding.unpkediri.ac.id Internet Source	<1 %
169	repo.uinsatu.ac.id Internet Source	<1 %
170	repository.pnj.ac.id Internet Source	<1 %
171	repository.uinsu.ac.id Internet Source	<1 %
172	sabynuzbunyw.blogspot.com Internet Source	<1 %
173	shemutcantik.blogspot.com Internet Source	<1 %
174	123dok.com Internet Source	<1 %
175	Andi Nur Halim, Rudiman Rudiman, Nauval Azmi Verdikha. "Analisis Sentimen Opini Publik Terhadap Peristiwa Bitcoin Halving Pada Data Teks Twitter Menggunakan Metode Naïve Bayes Dan Pembobotan Fitur TF-IDF", RIGGS: Journal of Artificial Intelligence and Digital Business, 2025 Publication	<1 %

176 Dufan Yuwana, Pulung Andono, Hendy Kurniawan. "Comparison of Effectiveness of Machine Learning Methods in Predicting Chemical Compound Toxicity Enhance Pharmaceutical Product Safety", INOVTEK Polbeng - Seri Informatika, 2025 <1%
Publication

177 Melania Velisita Lau, Aryani Dwi Handayani, Miyosagusti Yudo Widodo, Rani Irma Handayani, Risca Lusiana Pratiwi. "Implementasi Metode Random Forest untuk Klasifikasi Sampah Organik dan Anorganik Berbasis Citra Digital", RIGGS: Journal of Artificial Intelligence and Digital Business, 2025 <1%
Publication

178 Mukhtar Nashir, Dian Ade Kurnia, Yudhistira Arie Wijaya, Ade Irma Purnama Sari, Nisa Dienwati Nuris. "PERBANDINGAN KINERJA SVM DAN NAÏVE BAYES PADA ANALISIS SENTIMEN KOMENTAR DEMONSTRASI DPR 25 AGUSTUS 2025", Jurnal Mahasiswa Sistem Informasi (JMSI), 2025 <1%
Publication

179 Raihan Raihan, Cecep Nurul Alam, Wildan Budiawan Zulfikar. "Deteksi Pneumonia pada Citra Akhir X – Ray Dada Menggunakan Convolutional Neural Networks Berdasarkan Fitur Prewitt Operator", INTERNAL (Information System Journal), 2025 <1%
Publication

180 Randi Afif Afif, Aji Supriyanto, Rr. Fitri Damaryanti, Wahyu Prasetya Adi. "Analisis Sentimen Aplikasi Adiraku di Google Play <1%

Store Menggunakan Metode Support Vectore Machine", JURNAL FASILKOM, 2025
Publication

181	Rheza Timothy Tedjo, Alwin M. Sambul, Arie S.M. Lumenta. "Klasifikasi Gambar Bahan Makanan untuk Penderita Buta Warna", Jurnal Teknik Elektro dan Komputer, 2022 Publication	<1 %
182	Sekar Cinta Amaria, Nurtriana Hidayati. "ANALISIS PROGRAM MAKAN BERGIZI GRATIS DENGAN SUPPORT VECTOR MACHINE (SVM) PADA APLIKASI X", JSiI (Jurnal Sistem Informasi), 2025 Publication	<1 %
183	Tarwoto, Rizki Nugroho, Najmul Azka, Wakhid Sayudha Rendra Graha. "Analisis Sentimen Ulasan Aplikasi Mobile JKN di Google PlayStore Menggunakan IndoBERT", Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi), 2025 Publication	<1 %
184	Tatang Rohana, Euis Nurlaelasari, Elsa Elvira Awal, Hilda Yulia Novita. "Kajian Model Jaringan Syaraf Tiruan Untuk Memprediksi Secara Dini Tingkat Kelulusan Mahasiswa", Technologia : Jurnal Ilmiah, 2024 Publication	<1 %
185	Submitted to Universitas Maritim Raja Ali Haji Student Paper	<1 %
186	Wiwin Juliyanti. "GAYA HIDUP, RELIGIUSITAS DAN LITERASI KEUANGAN SYARIAH TERHADAP KEPUTUSAN PENGGUNAAN E-WALLET LINKAJA SYARIAH PADA MASYARAKAT	<1 %

KOTA SURAKARTA", AMAL: Jurnal Ekonomi Syariah, 2024
Publication

187	Zilvi Azus Sriyanti, Dhian Satria Yudha Kartika, Abdul Rezha Efrat Najaf. "IMPLEMENTASI MODEL BERT PADA ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP AKSI BOIKOT PRODUK ISRAEL", Jurnal Informatika dan Teknik Elektro Terapan, 2024 Publication	<1 %
188	digilib.iain-palangkaraya.ac.id Internet Source	<1 %
189	documents.mx Internet Source	<1 %
190	e-journal.hamzanwadi.ac.id Internet Source	<1 %
191	e-journal.uajy.ac.id Internet Source	<1 %
192	ejournal.staidarussalamlampung.ac.id Internet Source	<1 %
193	eprints.unpak.ac.id Internet Source	<1 %
194	hukum.ub.ac.id Internet Source	<1 %
195	ijins.umsida.ac.id Internet Source	<1 %
196	jtera.polteksmi.ac.id Internet Source	<1 %
197	jurnal.uts.ac.id Internet Source	<1 %

198	kurniadilabeouf.blogspot.com Internet Source	<1 %
199	library.gunadarma.ac.id Internet Source	<1 %
200	mafiadoc.com Internet Source	<1 %
201	negociaaoforexitabunai.blogspot.com Internet Source	<1 %
202	repositori.usu.ac.id:8080 Internet Source	<1 %
203	repository.ittelkom-pwt.ac.id Internet Source	<1 %
204	repository.ub.ac.id Internet Source	<1 %
205	repository.unpkediri.ac.id Internet Source	<1 %
206	repository.upnjatim.ac.id Internet Source	<1 %
207	rianbayristian.blogspot.com Internet Source	<1 %
208	www.amity.edu Internet Source	<1 %
209	www.journal-isi.org Internet Source	<1 %
210	www.journal.lembagakita.org Internet Source	<1 %
211	Irfandi Rusdiansyah, Ridwan Pangestu, Devina Azalia, Muhammad Faiz Zhafran, Ferdy Saputra, Fachri Amsury. "Integrasi Model	<1 %

Klasifikasi Tingkat Stress Mahasiswa Berbasis Natural Language Processing", RIGGS: Journal of Artificial Intelligence and Digital Business, 2025
Publication

212 Muh. Falach Achsan Yusuf. "Klasifikasi Gambar Burung Konservasi di Wilayah Papua Barat Menggunakan Transfer Learning", Indonesian Journal of Computer Science, 2024 <1 %
Publication

213 Novandra Hanifah Najlawarni, Tukino Paryono, Fitria Nurapriani, Baenil Huda. "Klasifikasi Dan Prediksi Pada Data Ulasan Traveloka Menggunakan Algoritma LSTM", INTECOMS: Journal of Information Technology and Computer Science, 2025 <1 %
Publication

214 Nufia Alfi Rohyana, Aris Wahyu Murdiyanto, Kharisma. "Analisis Sentimen Di Media Sosial Twitter Dengan Studi Kasus Vaksinasi Covid-19", INDONESIAN JOURNAL ON DATA SCIENCE, 2023 <1 %
Publication

215 Rizky Saputra Cendikiawan, Ali Ibrahim, Mira Afrina, Rizka Dhini Kurnia. "Sentiment Analysis of Zalora Products on Google Play Store Using Random Forest Method", Indonesian Journal of Innovation Studies, 2025 <1 %
Publication

216 Ahmad Nursodiq, Ronald Parsaulian Simanjuntak, Anandriyan Aditya Zaky, Dues Fernando Kopong Duli, Rivaldi Afdholu Dzikri. "Pengembangan Sistem Prediksi Kelulusan <1 %

Mahasiswa Tepat Waktu Berdasarkan Data Akademik Menggunakan Metode Jaringan Syaraf Tiruan Backpropagation Berbasis Python", Jurnal Pendidikan Tambusai, 2026
Publication

---

217 Al Khaidar Author. "ANALISIS SENTIMEN DI INSTAGRAM TERHADAP MENTERI KEUANGAN PURBAYA YUDHI SADEWA MENGGUNAKAN METODE LOGISTIC REGRESSION", Jurnal Informatika dan Teknik Elektro Terapan, 2025 <1%
Publication

---

218 Amelia Mar'atusholihat, Nuk Ghurroh Setyoningrum, Dede Rizal Nursamsi. "Perbandingan Kinerja Algoritma Decision Tree Dan Naïve Bayes Dalam Analisis Sentimen Publik Terhadap Kebijakan Pemerintah Mengenai Batasan Penggunaan Media Sosial Anak", Informatics and Digital Expert (INDEX), 2025 <1%
Publication

---

219 Dimas Thaqif Attaulah, Dewi Soyusiawaty. "Analisis Sentimen Program Makan Siang Gratis di Twitter/X menggunakan Metode BI-LSTM", Edumatic: Jurnal Pendidikan Informatika, 2025 <1%
Publication

---

220 Edy Subowo, Fenilinas Adi Artanto, Isna Putri, Wahyu Umaedi. "Algoritma Bidirectional Long Short Term Memory untuk Analisis Sentimen Berbasis Aspek pada Aplikasi Belanja Online dengan Cicilan", JURNAL FASILKOM, 2022 <1%
Publication

---

221 Erna Nurmawati, Adielia Amanda. "ANALISIS SENTIMEN DAN PEMODELAN TOPIK PADA <1%

TWEET TERKAIT DATA BADAN PUSAT STATISTIK", Jurnal Sistem Informasi dan Informatika (Simika), 2023
Publication

222 Fathoni Fathoni, Ali Ibrahim, Putri Mutiara Arinie, Arvhi Randita Setia, Dian Febriansyah. "Analisis Perilaku FOMO Generasi Z dan Peran Influencer dalam Pembelian Produk Kecantikan Menggunakan Random Forest", Jurnal Sains dan Informatika, 2025 <1%
Publication

223 Junaedi, Alexius Hendra Gunawan, Verri Kuswanto, Jonathan. "Tinjauan Support Vector Machine dalam Text-Mining untuk Analisis Sentimen di Sektor Pariwisata", bit-Tech, 2024 <1%
Publication

224 Muhamad Ali Zaenal Abidin. "Evaluasi Sentimen Ulasan Pengguna CGV Cinemas Indonesia Menggunakan Metode Naïve Bayes dan Support Vector Machine", Indonesian Journal on Software Engineering (IJSE), 2025 <1%
Publication

225 Muhammad Samsul Ma'arif, Jajam Haerul Jaman, Agung Susilo Yuda Irawan. "ANALISIS SENTIMEN ULASAN APLIKASI INVESTASI MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE BERBASIS PARTICLE SWARM OPTIMIZATION", Jurnal Informatika dan Teknik Elektro Terapan, 2024 <1%
Publication

226 Putri Yefani, Rini Asmara. "Pengukuran kepuasan pengguna terhadap layanan perpustakaan menggunakan metode <1%

servqual dan importance performance analysis (ipa) di Dinas Kearsipan dan Perpustakaan Daerah Provinsi Sumatera Barat", Pustaka Karya : Jurnal Ilmiah Ilmu Perpustakaan dan Informasi, 2025
Publication

227 Riko Angga Bayu Kusuma -, Bambang Irawan -, Abdul Khamid -. "KLASIFIKASI PENYAKIT KULIT WAJAH MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK EFFICIENTNET-B3", Jurnal Informatika dan Teknik Elektro Terapan, 2026 <1%
Publication

228 Rizki Fegiyanto, Arief Hermawan, Farida Ardiani. "Prediksi Harga Crypto dengan Algoritma Jaringan Saraf Tiruan", Jurnal Indonesia : Manajemen Informatika dan Komunikasi, 2024 <1%
Publication

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		

ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN PLATFORM RUANGGURU MENGGUNAKAN METODE LONG SHORT-TERM MEMORY (LSTM)

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14

PAGE 15

PAGE 16

PAGE 17

PAGE 18

PAGE 19

PAGE 20

PAGE 21

PAGE 22

PAGE 23

PAGE 24

PAGE 25

PAGE 26

PAGE 27

PAGE 28

PAGE 29

PAGE 30

PAGE 31

PAGE 32

PAGE 33

PAGE 34

PAGE 35

PAGE 36

PAGE 37

PAGE 38

PAGE 39

PAGE 40

PAGE 41

PAGE 42

PAGE 43

PAGE 44

PAGE 45

PAGE 46

PAGE 47

PAGE 48

PAGE 49

PAGE 50

PAGE 51

PAGE 52

PAGE 53

PAGE 54

PAGE 55

PAGE 56

PAGE 57

PAGE 58

PAGE 59

PAGE 60

PAGE 61

PAGE 62

PAGE 63

PAGE 64

PAGE 65

PAGE 66

PAGE 67

PAGE 68

PAGE 69

PAGE 70

PAGE 71

PAGE 72

PAGE 73

PAGE 74

PAGE 75

PAGE 76

PAGE 77

PAGE 78

PAGE 79

PAGE 80

PAGE 81

PAGE 82

PAGE 83

PAGE 84

PAGE 85

PAGE 86

PAGE 87

PAGE 88

PAGE 89

PAGE 90

PAGE 91

PAGE 92

PAGE 93

PAGE 94

PAGE 95

PAGE 96

PAGE 97

PAGE 98

PAGE 99

PAGE 100

PAGE 101

PAGE 102

PAGE 103

PAGE 104

PAGE 105

PAGE 106

PAGE 107

PAGE 108

PAGE 109

PAGE 110

PAGE 111

PAGE 112

PAGE 113

PAGE 114

PAGE 115

PAGE 116

PAGE 117

PAGE 118


Universitas
Bina
Darma