

TESIS

**PERBANDINGAN ALGORITMA DNN DAN LSTM UNTUK
ANALISIS SENTIMEN LAYANAN TRANSPORTASI ONLINE
DI INDONESIA MENGGUNAKAN DATA TWITTER**



disusun oleh:

NAMA : JANDRI TAMPUBOLON

NIM : 23.55.2543

Konsentrasi : Digital Transformation Intelligence

**PROGRAM PASCASARJANA
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

TESIS
PERBANDINGAN ALGORITMA DNN DAN LSTM UNTUK
ANALISIS SENTIMEN LAYANAN TRANSPORTASI ONLINE
DI INDONESIA MENGGUNAKAN DATA TWITTER

SENTIMENT ANALYSIS OF USERS TOWARDS ONLINE
TRANSPORTATION SERVICES USING TWITTER DATA

Diajukan untuk memenuhi salah satu syarat mencapai derajat Pascasarjana
Program Studi *S2 PJJ Informatika*

HALAMAN JUDUL



disusun oleh

NAMA : JANDRI TAMPUBOLON
NIM : 23.55.2543
Konsentrasi : Digital Transformation Intelligence

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025

HALAMAN PERSETUJUAN

**PERBANDINGAN ALGORITMA DNN DAN LSTM UNTUK ANALISIS
SENTIMEN LAYANAN TRANSPORTASI ONLINE DI INDONESIA
MENGUNAKAN DATA TWITTER**

**COMPARISON OF DNN AND LSTM ALGORITHMS FOR SENTIMENT
ANALYSIS OF ONLINE TRANSPORTATION SERVICES IN INDONESIA USING
TWITTER DATA**

Dipersiapkan dan Disusun oleh

Jandri Tampubolon

NIM 23.55.2543

Telah disetujui oleh Tim Dosen Pembimbing Tesis
pada tanggal 10 November 2025

Pembimbing Utama



Prof. Dr. Kusriani, M.Kom.

NIK. 190302106

HALAMAN PENGESAHAN

PERBANDINGAN ALGORITMA DNN DAN LSTM UNTUK ANALISIS
SENTIMEN LAYANAN TRANSPORTASI ONLINE DI INDONESIA
MENGUNAKAN DATA TWITTER

COMPARISON OF DNN AND LSTM ALGORITHMS FOR SENTIMENT
ANALYSIS OF ONLINE TRANSPORTATION SERVICES IN
INDONESIA USING TWITTER DATA

yang disusun dan diajukan oleh

Jandri Tampubolon

23.55.2543

Telah dipertahankan di depan Dewan Penguji
pada tanggal, 10 November 2025.

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Hanif Al Fatta, S.Kom., M.Kom., Ph.D.

NIK. 190302096

I Made Artha Agastya, S.T., M.Eng., PhD

NIK. 190302352

Prof. Dr. Kusriani, M.Kom.

NIK. 190302106

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Tanggal 10 November 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.

NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : **Jandri Tampubolon**

NIM : **23.55.2543**

Menyatakan bahwa Tesis dengan judul berikut:

Perbandingan Algoritma DNN Dan LSTM Untuk Analisis Sentimen

Layanan Transportasi Online Di Indonesia Menggunakan Data Twitter

Dosen Pembimbing : Prof. Dr. Kuselni, M.Kom.

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian **SAYA** sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab **SAYA**, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini **SAYA** buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka **SAYA** bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 10 November 2025

Yang Menyatakan,

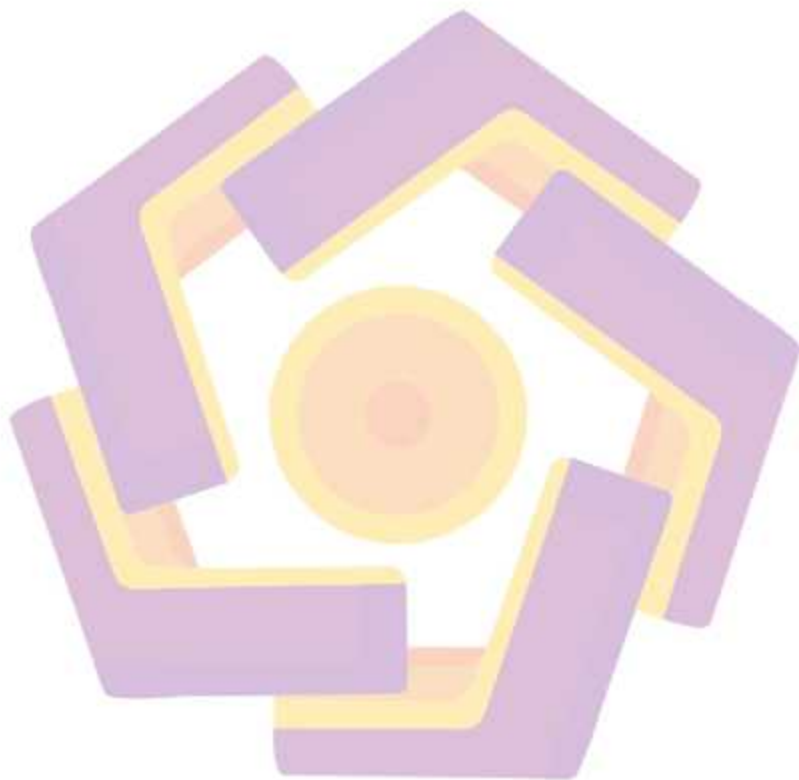


Jandri Tampubolon.

PERSEMBAHAN

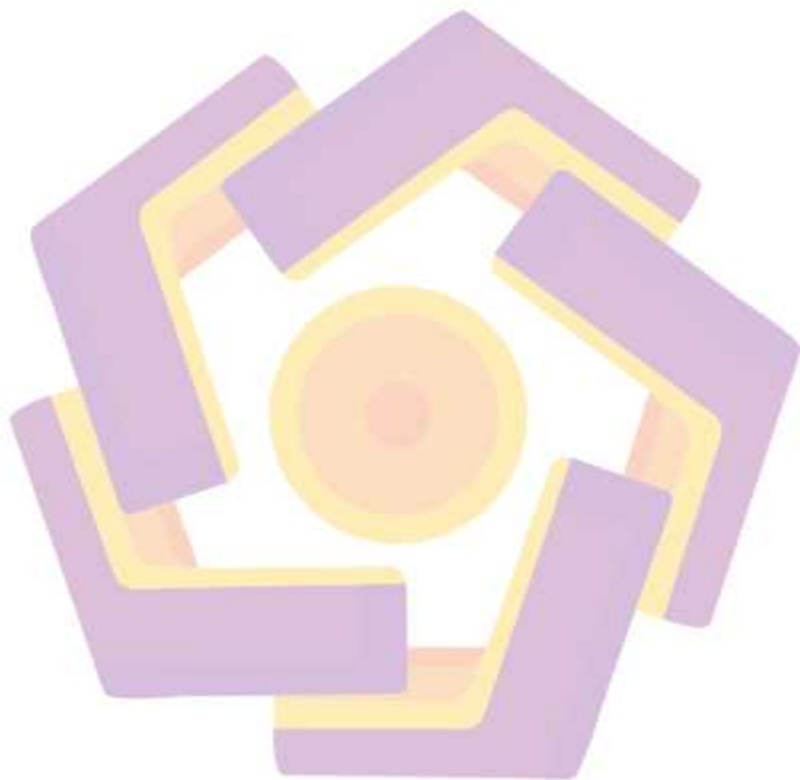
Untuk keluarga tercinta, fondasi, doa, dan keberanian.

Untuk sahabat seperjuangan, ruang bertumbuh dalam diskusi dan kerja
Semoga penelitian ini menjadi sumbangsih kecil yang dapat digunakan dan
dikembangkan lebih lanjut.



MOTTO

Pariurna dalam proses; lapang menerima hasil; tanpa penyesalan.



KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas limpahan rahmat dan karunia-Nya sehingga tesis berjudul “Perbandingan Algoritma Dnn Dan Lstm Untuk Analisis Sentimen Layanan Transportasi Online Di Indonesia Menggunakan Data Twitter” ini dapat diselesaikan dengan baik. Tesis ini disusun untuk memenuhi salah satu syarat mencapai derajat Pascasarjana Program Studi S2 PJJ Informatika, Konsentrasi Digital Transformation Intelligence, Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta. Pada kesempatan ini, penulis menyampaikan terima kasih yang setulus-tulusnya kepada:

1. Prof. Dr. Kusriani, M.Kom., selaku Pembimbing Utama, atas bimbingan, arahan, dan masukan yang sangat berharga sepanjang proses penelitian hingga penyusunan tesis ini.
2. Dewan Penguji: Hanif Al Fatta, S.Kom., M.Kom., Ph.D.; I Made Artha Agastya, S.T., M.Eng., Ph.D.; dan Prof. Dr. Kusriani, M.Kom., atas kritik dan saran konstruktif pada ujian tesis 10 November 2025.
3. Pimpinan Universitas AMIKOM Yogyakarta, Fakultas Ilmu Komputer, dan Program Studi S2 PJJ Informatika atas fasilitas dan dukungan akademik yang diberikan.
4. Rekan-rekan dan semua pihak yang telah membantu pengumpulan data, diskusi teknis, serta dukungan moral selama penelitian berlangsung.
5. Keluarga tercinta atas doa, pengertian, dan semangat yang tidak pernah putus.

Penulis menyadari bahwa tesis ini masih memiliki keterbatasan. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi penyempurnaan karya di masa mendatang. Semoga karya ini memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya di bidang analisis sentimen dan pembelajaran mendalam untuk data media sosial.

Penulis,

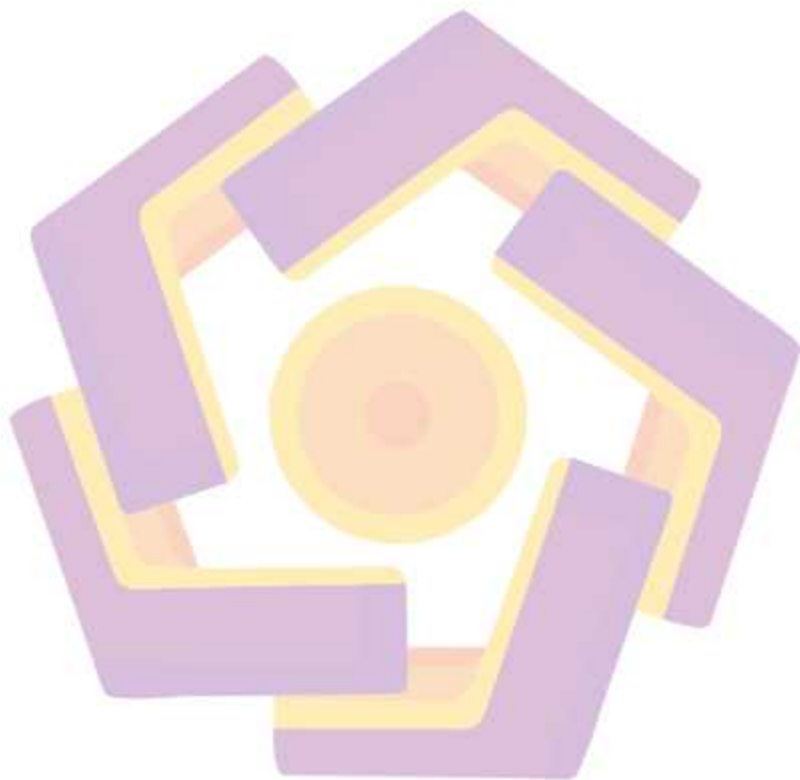
Jandri Tampubolon

NIM 23.55.2543

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN KEASLIAN TESIS	iv
PERSEMBAHAN	v
MOTTO	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR ISTILAH.....	xii
INTISARI.....	xv
ABSTRACT.....	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	5
1.3 Batasan Masalah.....	6
1.4 Tujuan Penelitian.....	6
1.5 Manfaat Penelitian.....	7
BAB II TINJAUAN PUSTAKA	8
2.1 Tinjauan Pustaka.....	8
2.2 Landasan Teori.....	11
2.3 Keaslian Penelitian.....	23
BAB III METODE PENELITIAN	47
3.1 Jenis, Sifat Dan Pendekatan Penelitian.....	47
3.2 Metode Pengumpulan Data.....	47
3.3 Metode Analisis Data.....	48
3.4 Alur Penelitian	52
BAB IV HASIL PENELITIAN DAN PEMBAHASAN	54
4.1 Pengumpulan Data	54
4.2 Pre-processing Data	55
4.3 Pelabelan Data.....	59
4.4 Pelatihan Model	62

4.5	Evaluasi Model.....	63
BAB V PENUTUP.....		74
5.1	Kesimpulan	74
5.2	Kontribusi Kebaruan (Novelty).....	74
Daftar Pustaka		75



DAFTAR TABEL

Tabel 2. 1 Matriks literatur review dan posisi penelitian	23
Tabel 4. 1 Hasil <i>Crawling</i>	54
Tabel 4. 2 Hasil penghapusan data tidak relevan	55
Tabel 4. 3 Hasil <i>Case folding</i>	56
Tabel 4. 4 Hasil <i>stopword removal</i>	56
Tabel 4. 5 Hasil <i>Tokenisasi</i>	57
Tabel 4. 6 Hasil <i>Stemming</i>	58
Tabel 4. 7 Hasil <i>Preprocessing Data</i>	58
Tabel 4. 8 Hasil <i>Labeling</i>	60
Tabel 4. 9 Hasil Contoh Data Setelah Dilakukan Drop Variabel	60
Tabel 4. 10 Perbandingan Auto-Label dan Label Manual per Kelas ($n=200$)	61
Tabel 4. 11 Hasil Uji Performa Model DNN dan LSTM	63
Tabel 4. 12 Hasil Uji Performa Model DNN dan LSTM	65
Tabel 4. 13 Hasil Evaluasi Model	68
Tabel 4. 14 Hasil Evaluasi Model	68
Tabel 4. 15 Hasil Evaluasi Model	69
Tabel 4. 16 Hasil Evaluasi Model	69

DAFTAR GAMBAR

Gambar 2. 1 Arsitektur Deep Neural Network (DNN).....	13
Gambar 2. 2 Arsitektur jaringan LSTM untuk data sekuensial.....	14
Gambar 2. 3 Confusion matrix.....	18
Gambar 3. 1 Tahapan Analisis Sentimen.....	48
Gambar 3. 2 Tahapan Alur Penelitian.....	52
Gambar 4. 1 Hasil Uji Performa Model DNN dan LSTM.....	63
Gambar 4. 2 Hasil Uji Performa Model DNN dan LSTM.....	66
Gambar 4. 3 Perbandingan Kinerja Model	68
Gambar 4. 4 Perbandingan Kinerja Model	68
Gambar 4. 5 Perbandingan Kinerja Model	69
Gambar 4. 6 Perbandingan Kinerja Model	70

DAFTAR ISTILAH

- Akurasi: Proporsi prediksi yang benar terhadap seluruh prediksi.
- Anotasi otomatis (auto-labeling): Pemberian label kelas secara otomatis oleh model untuk membangun data latih.
- Arsitektur model: Rancangan struktur jaringan (jenis layer, jumlah unit, urutan koneksi) pada sebuah model.
- Batch size: Jumlah contoh data yang diproses sekaligus sebelum pembaruan bobot.
- Case folding: Penyeragaman huruf (mis. ke huruf kecil) pada tahap prapemrosesan.
- Cohen's Kappa: Koefisien kesepakatan antara dua pelabel (atau pelabel otomatis vs manual) dengan mengoreksi peluang kesepakatan acak; rentang -1 s.d. 1 (semakin tinggi semakin baik).
- Cross-validation (K-Fold): Prosedur validasi dengan membagi data ke dalam K lipatan untuk pelatihan dan pengujian bergiliran.
- Data latih/validasi/uji: Pembagian dataset untuk melatih model, menyeting hiperparameter, dan mengukur kinerja final.
- Deep Neural Network (DNN): Jaringan saraf berlapis banyak tanpa mekanisme memori eksplisit atas urutan.
- Embedding: Representasi vektor berdimensi tetap dari teks sehingga dapat diproses oleh model.
- Epoch: Satu putaran penuh memindai seluruh data latih pada proses pelatihan.
- Early stopping: Teknik menghentikan pelatihan lebih awal saat performa validasi berhenti membaik, untuk mencegah *overfitting*.
- F1-score: Rata-rata harmonik antara presisi dan recall; menyeimbangkan keduanya pada evaluasi klasifikasi.
- Fine-tuning: Penyesuaian bobot model pralatih pada tugas/korpus spesifik.
- Fungsi aktivasi: Transformasi nonlinier pada neuron (mis. ReLU, Sigmoid, Softmax).

- Fungsi rugi (loss): Ukuran kesalahan yang diminimalkan selama pelatihan (mis. *cross-entropy*).
- Hyperparameter (hiperparameter): Parameter yang ditetapkan sebelum pelatihan (mis. laju belajar, batch size, jumlah layer).
- IndoBERT / IndoBERTweet: Varian model *Transformer* berbahasa Indonesia; pada penelitian ini digunakan terutama sebagai sumber embedding atau anotator otomatis.
- Inferensi (inference): Proses menghasilkan prediksi dari model terlatih.
- Kamus slang: Daftar padanan kata/ungkapan gaul-baku untuk normalisasi teks informal (mis. Twitter).
- Kebingungan (confusion matrix): Tabel ringkasan prediksi vs label sebenarnya per kelas.
- Kelas sentimen: Kategori polaritas (umumnya positif, negatif, netral).
- Laju belajar (learning rate): Besaran langkah pembaruan bobot pada optimisasi.
- Latensi: Waktu tunda dari masukan hingga keluaran saat inferensi.
- Long Short-Term Memory (LSTM): Varian RNN dengan *gating* untuk menjaga/melupakan informasi sehingga efektif menangkap ketergantungan urutan.
- MAE (Mean Absolute Error): Rata-rata nilai absolut selisih prediksi terhadap label; semakin kecil semakin baik.
- Normalisasi teks: Proses merapikan teks (mis. penghapusan URL, *mention*, angka berulang) agar konsisten untuk pemodelan.
- Optimizer: Algoritme pembaruan bobot (mis. Adam, SGD).
- Overfitting: Kondisi model terlalu menyesuaikan data latih sehingga kinerjanya menurun pada data baru.
- Padanan istilah (synonym mapping): Pemetaan kata sepadan (mis. *ojol* → *ojek online*) guna menyatukan variasi.
- Prapemrosesan (preprocessing): Langkah awal mengolah data mentah (tokenisasi, normalisasi, stopword, dsb.).
- Presisi (precision): Proporsi prediksi positif yang benar terhadap seluruh prediksi positif.

- Rekal (recall): Proporsi kasus positif yang berhasil terdeteksi terhadap seluruh kasus positif sebenarnya.
- Regularisasi: Teknik menekan kompleksitas model (mis. *dropout*, L2) untuk mengurangi *overfitting*.
- Resampling: Penyeimbangan distribusi kelas (mis. *oversampling* minoritas atau *undersampling* mayoritas).
- Stopword: Kata umum yang biasanya dihapus karena nilai informasinya rendah (mis. “dan”, “yang”).
- Tokenisasi: Pemenggalan teks menjadi unit lebih kecil (token/kata/sub-kata).
- Transformer: Arsitektur berbasis mekanisme *self-attention* yang kuat untuk representasi konteks.
- Tweet: Unit pesan di Twitter/X berisi teks singkat dan elemen metadata.
- Validasi: Pengujian sementara untuk menyetel model sebelum evaluasi final di data uji.
- Waktu pelatihan (training time): Durasi yang dibutuhkan model untuk menyelesaikan proses pelatihan.

INTISARI

Penelitian ini membandingkan dua arsitektur pembelajaran mendalam. Deep Neural Network (DNN) dan Long Short-Term Memory (LSTM), untuk analisis sentimen tweet berbahasa Indonesia terkait layanan transportasi online. Data primer berjumlah ± 10.000 tweet (2023) dihimpun melalui API Twitter dengan kata kunci merek layanan, kemudian direpresentasikan sebagai embedding IndoBERT untuk masukan ke kedua model.

Pelabelan awal dilakukan secara otomatis menggunakan model IndoBERT dan divalidasi pada 200 sampel acak: 185 label sesuai (akurasi 92,5%) dengan Cohen's Kappa 0,89, menunjukkan reliabilitas "sangat tinggi" untuk keperluan pelatihan.

Eksperimen dirancang dalam empat skenario (tanpa/ dengan resampling dan tanpa/ dengan kamus slang) dan dievaluasi menggunakan K-Fold Cross-Validation ($k=3$ dan $k=5$) serta metrik akurasi, F1-score, MAE, dan waktu pelatihan.

Hasil menunjukkan LSTM secara konsisten mengungguli DNN dalam akurasi dan F1-score, terutama pada skenario dengan resampling dan kamus slang (mencapai akurasi $\sim 93\%$ dan F1-score $\sim 93\%$, MAE $\sim 0,088$). Namun, LSTM memerlukan waktu pelatihan lebih lama (~ 475 detik) dibandingkan DNN (~ 56 detik). Resampling terbukti memberikan peningkatan performa paling signifikan, sedangkan penggunaan kamus slang memberikan peningkatan tambahan yang bersifat pendukung. Analisis berdasarkan panjang teks menunjukkan DNN lebih stabil untuk tweet ≤ 10 kata, sementara LSTM unggul pada tweet > 10 kata, yang mempertegas kelebihan LSTM dalam menangkap dependensi urutan pada teks panjang. Secara keseluruhan, LSTM direkomendasikan sebagai arsitektur utama untuk analisis sentimen pengguna layanan transportasi online berbahasa Indonesia, dengan DNN sebagai alternatif ketika efisiensi komputasi diutamakan.

Berdasarkan keseluruhan temuan, penelitian ini merekomendasikan LSTM sebagai arsitektur utama untuk analisis sentimen layanan transportasi online berbahasa Indonesia, dengan DNN sebagai alternatif saat efisiensi komputasi menjadi prioritas.

Kata kunci: analisis sentimen, Twitter, transportasi online, IndoBERT, DNN, LSTM, resampling, kamus slang, K-Fold.

ABSTRACT

This study compares two deep learning architectures—Deep Neural Networks (DNN) and Long Short-Term Memory (LSTM)—for sentiment analysis of Indonesian-language tweets about online transportation services. The primary dataset comprised approximately ±10,000 tweets collected in 2023 via the Twitter API using brand-related keywords, then represented as IndoBERT embeddings to serve as inputs to both models. Initial labeling was performed automatically using an IndoBERT model and validated on 200 random samples: 185 labels matched (92.5% accuracy) with a Cohen's Kappa of 0.89, indicating “very high” reliability for training purposes. Experiments were designed across four scenarios (with/without resampling and with/without a slang dictionary) and evaluated using k-fold cross-validation (k = 3 and k = 5) with accuracy, F1-score, MAE, and training time as metrics. Results show that LSTM consistently outperforms DNN in accuracy and F1-score, especially in the scenario with both resampling and a slang dictionary (achieving ~93% accuracy and ~93% F1-score, MAE ~0.088). However, LSTM requires longer training time (~475 seconds) compared to DNN (~36 seconds). Resampling yielded the most substantial performance gains, while the slang dictionary provided additional supportive improvements. A length-based analysis indicates that DNN is more stable for tweets ≤10 words, whereas LSTM excels for tweets >10 words, underscoring LSTM's advantage in capturing sequential dependencies in longer text. Overall, LSTM is recommended as the primary architecture for sentiment analysis of Indonesian-language users of online transportation services, with DNN as an alternative when computational efficiency is prioritized. Based on the overall findings, this study recommends LSTM as the main architecture for sentiment analysis of Indonesian online transportation services, with DNN as an alternative when efficiency is the priority.

Keywords: *sentiment analysis, Twitter, online transportation, IndoBERT, DNN, LSTM, resampling, slang dictionary, k-fold.*

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Di era modern saat ini, kemajuan teknologi telah merambah ke berbagai sektor, termasuk transportasi. Salah satu terobosannya adalah munculnya moda transportasi online yang memberikan kemudahan bagi masyarakat dalam bepergian. Layanan aplikasi transportasi online makin menjamur di Indonesia, menawarkan berbagai layanan untuk mempermudah mobilitas penggunanya.

Twitter, sebagai salah satu platform media sosial paling populer dan banyak digunakan di dunia, termasuk di Indonesia, menjadi tempat bagi pengguna untuk aktif membagikan pengalaman, opini, dan kefuhan mereka terhadap berbagai layanan, termasuk transportasi online seperti Gojek, Grab, Maxim dan InDrive. Aliran data yang kaya dan beragam dari platform ini sangat berguna untuk analisis sentimen, (Zhu et al., 2023) menegaskan bahwa media sosial seperti Twitter telah menjadi sumber utama data dalam pengembangan model deep learning untuk analisis sosial. Data langsung dari pengguna ini memungkinkan perusahaan untuk mendapatkan wawasan yang lebih mendalam mengenai persepsi publik terhadap layanan yang mereka tawarkan.

Analisis sentimen telah banyak dimanfaatkan dalam berbagai sektor layanan, termasuk transportasi, untuk mengevaluasi persepsi pengguna dan meningkatkan kualitas layanan. Salah satu pendekatan yang efektif adalah menggunakan kerangka kerja deep learning yang terbukti mampu menangkap pola sentimen secara lebih akurat dalam konteks layanan transportasi (Zhang & Liu, 2023).

Penelitian (Safira et al., 2023) berjudul *Sentiment Analysis to Find Out Positive or Negative Opinions on Ride Hailing Application*. Penelitian ini menganalisis opini pengguna mengenai layanan Gojek yang diekspresikan melalui Twitter. Penelitian ini mengevaluasi sentimen pengguna terhadap berbagai aspek layanan Gojek seperti pengantaran makanan (GoFood), transportasi (GoRide dan GoCar), serta layanan terkait lainnya. Studi ini menunjukkan bagaimana opini

masyarakat dapat memengaruhi penilaian terhadap kualitas layanan. Data dikumpulkan menggunakan Twitter API dengan kata kunci terkait layanan Gojek. Sentimen pada tweet diklasifikasikan menggunakan metode Support Vector Machine (SVM). Data melewati tahapan preprocessing seperti konversi emotikon, pembersihan data, tokenisasi, case folding, penghapusan stopwords, dan stemming sebelum dilakukan klasifikasi. Model SVM yang digunakan mencapai akurasi 80,3% dalam mengklasifikasikan sentimen positif dan negatif pada tweet terkait Gojek. Model ini mampu memberikan gambaran yang akurat mengenai sentimen pengguna terhadap layanan yang diberikan oleh Gojek.

Penelitian (Rakshit, Sarkar, & Roy, 2024) membahas bagaimana pendekatan deep learning dapat digunakan untuk menganalisis sentimen dari data Twitter dalam jumlah besar, mencakup 1,6 juta tweet dengan sentimen positif dan negatif. Fokus utama penelitian adalah meningkatkan akurasi klasifikasi sentimen dengan mempertimbangkan pentingnya keberadaan stopwords. Data tweet yang diproses dengan Word2Vec untuk menghasilkan vektor kata. Model yang dibangun mencakup RNN, Bi-directional RNN, dan LSTM. Data dibagi dengan rasio 80:20 untuk pelatihan dan pengujian. Proses pelatihan menggunakan embedding layer dan dilakukan eksperimen dengan mempertahankan dan menghapus stopwords. Model LSTM menunjukkan akurasi tertinggi sebesar 83,5% dan F1 score 84,6%. Hasil menunjukkan bahwa mempertahankan stopwords dalam data meningkatkan performa model secara signifikan. Model yang dikembangkan dapat digunakan untuk mengklasifikasikan sentimen dengan baik pada data Twitter.

Penelitian (Pullanikkat et al., 2024) mengeksplorasi bagaimana Twitter dapat digunakan untuk mengidentifikasi keluhan spesifik mengenai sistem transportasi di kota-kota besar India, seperti keterlambatan, ketersediaan transportasi, dan sikap petugas. Hasilnya diharapkan dapat memberikan wawasan bagi pemerintah dan otoritas transportasi dalam merespons keluhan publik. Data dikumpulkan menggunakan Twitter Search API dari enam kota besar di India selama 15 hari pada Maret 2023. Data diklasifikasikan menggunakan model RoBERTa dan BERTweet untuk klasifikasi sentimen dan model multi-label untuk mengidentifikasi jenis keluhan. Sebanyak 12.000 tweet diolah untuk membentuk dataset yang digunakan dalam pelatihan model. Model berbasis CrisisTransformer-

M3 memberikan akurasi tertinggi dalam klasifikasi keluhan transportasi, dengan akurasi 91,4% dan F1 score 91,3%. Studi ini menunjukkan bahwa media sosial seperti Twitter dapat digunakan secara efektif untuk mengidentifikasi keluhan transportasi secara rinci, yang berguna bagi pemerintah untuk perencanaan transportasi.

Penelitian (Urolagin & Patel, 2024) mengembangkan metrik seperti User Sentiment Score (USS), Satisfaction Score (SATS), dan Social Data Influence Score (SDIS) menggunakan data Twitter. Penelitian ini membandingkan pendekatan baru dengan metrik tradisional seperti NPS (Net Promoter Score) dan NSS (Net Sentiment Score). Data Twitter dari lima platform e-learning—Coursera, Udemy, Udacity, Khanacademy, dan edX—digunakan untuk mengevaluasi kinerja model. Menggunakan model Deep Neural Network (DNN) untuk memprediksi loyalitas pelanggan berdasarkan skor USS. Model ini dibandingkan dengan beberapa algoritma lain, termasuk Random Forest Classifier (RFC), eXtreme Gradient Boosting (XGB), dan Support Vector Classifier (SVC). Data sentiment diolah menggunakan VADER untuk klasifikasi sentimen. Model DNN menunjukkan akurasi tertinggi dalam memprediksi loyalitas pelanggan (98,62%). Hasilnya menunjukkan bahwa metrik yang diusulkan dapat memberikan wawasan yang lebih baik tentang kepuasan dan pengaruh pelanggan di tingkat pengguna. Metrik SDIS membantu mengukur pengaruh keseluruhan pelanggan terhadap pengguna baru di platform.

Penelitian (Vyas et al., 2022) Menggunakan VADER (Valence Aware Dictionary and Sentiment Reasoner) untuk analisis sentimen awal dan pemberian label pada tweet. Tweet yang sudah diberi label ini kemudian digunakan untuk melatih model pembelajaran mesin seperti Gaussian Naïve Bayes, Logistic Regression, Random Forest, dan Long Short-Term Memory (LSTM). Evaluasi dilakukan menggunakan metrik akurasi, presisi, recall, dan F1 score. Model LSTM menunjukkan kinerja terbaik dengan akurasi klasifikasi 83%. Kerangka hibrida yang dikembangkan berhasil mengotomatisasi klasifikasi sentimen dalam volume besar tweet, memberikan wawasan yang berguna bagi pembuat kebijakan terkait sentimen masyarakat terhadap kebijakan pandemi.

Penelitian (Qi & Shabrina, 2023) mengeksplorasi bagaimana sentimen publik terhadap COVID-19, termasuk vaksinasi dan kebijakan lockdown, diungkapkan melalui Twitter. Analisis dilakukan pada tiga tahap lockdown untuk memahami perubahan sentimen dari waktu ke waktu. Dengan menggunakan pendekatan leksikon (TextBlob, VADER, SentiWordNet) dan pembelajaran mesin (Random Forest, Multinomial Naive Bayes, dan SVC). Data diambil dari Twitter selama tiga tahap lockdown di kota-kota besar di Inggris, dan dilakukan pembersihan data serta pemrosesan fitur teks dengan BoW, TF-IDF, dan Word2Vec. Pendekatan berbasis leksikon menunjukkan hasil yang bervariasi, dengan VADER lebih akurat dalam menangani bahasa tidak formal. Dari model pembelajaran mesin, SVC dengan representasi fitur TF-IDF memberikan akurasi terbaik (71%). Perubahan sentimen positif meningkat pada awal lockdown, kemudian menurun di tahap akhir seiring dengan kenaikan kasus COVID-19.

Berdasarkan latar belakang yang diangkat didapati Penelitian sebelumnya telah mengeksplorasi berbagai pendekatan dalam analisis sentimen menggunakan metode pembelajaran mesin dan pembelajaran mendalam (deep learning). Namun, belum ada penelitian yang secara langsung membandingkan kinerja algoritma Deep Neural Network (DNN) dan Long Short-Term Memory (LSTM) dalam konteks analisis sentimen terhadap layanan transportasi online di Indonesia. Oleh karena itu, penelitian ini bertujuan untuk mengisi celah tersebut.

Meskipun LSTM (Long Short-Term Memory) pada awalnya dirancang untuk memproses data time series, model ini juga cocok untuk data non-time series karena fleksibilitasnya dalam memproses urutan, banyak data non-time series memiliki urutan atau hubungan antar-elemen, meskipun tidak berbasis waktu. Misalnya, dalam teks, urutan kata sangat penting untuk menentukan arti atau sentimen. Menangkap dependensi jangka panjang, dalam analisis sentimen, konteks kata-kata sebelumnya dan sesudahnya dapat memengaruhi interpretasi sebuah kata. Misalnya, kalimat "Film ini tidak bagus" memiliki sentimen negatif karena kata "tidak" membalikkan arti dari kata "bagus". Menghasilkan representasi data yang kaya. Kunci utamanya adalah dengan merepresentasikan data non-time series sebagai urutan atau struktur yang relevan, sehingga LSTM dapat belajar pola atau hubungan antar-elemen dengan efektif.

Sementara itu DNN cocok untuk analisis sentimen karena kemampuannya untuk menangkap hubungan kompleks, fleksibilitas dalam arsitektur, dan performa yang tinggi pada dataset besar. Dengan representasi data yang kaya seperti word embeddings atau contextual embeddings, DNN dapat memahami konteks, nuansa, dan pola dalam data teks

Keunggulan masing-masing model menjadi alasan kuat untuk membandingkannya. LSTM unggul dalam memahami urutan dan dependensi antar-elemen, sedangkan DNN lebih fleksibel dalam menangani dataset besar dan menemukan hubungan kompleks. Perbandingan ini memberikan wawasan yang penting tentang bagaimana kedua pendekatan ini dapat digunakan secara optimal untuk memecahkan tantangan dalam analisis sentimen. Representasi data yang kaya dan kemampuan masing-masing untuk memahami pola mendalam dalam teks menjadikan LSTM dan DNN sebagai pilihan yang sangat layak untuk analisis ini.

1.2 Rumusan Masalah

Dari latar belakang tersebut, rumusan masalah yang akan dibahas pada penelitian ini, adalah :

- a. Berapakah nilai performa dari model algoritma Deep Neural Network (DNN) dan Long Short-Term Memory (LSTM) berdasarkan metrik evaluasi seperti akurasi, presisi, recall, F1-score, waktu pelatihan, dan Mean Absolute Error (MAE) ? Sebab penggunaan model yang kurang optimal tanpa analisis yang tepat dapat menghasilkan prediksi yang tidak akurat dan berpotensi menimbulkan keputusan yang salah dalam penerapan nyata?
- b. Apakah penerapan metode K-Fold Cross Validation dapat meningkatkan reliabilitas dalam evaluasi performa model DNN dan LSTM ? Sebab tanpa adanya metode validasi yang tepat evaluasi model dapat menyesatkan serta mengakibatkan generalisasi yang buruk pada data baru sehingga menurunkan keandalan model dalam implementasi praktis ?

1.3 Batasan Masalah

Batasan masalah dari penelitian adalah sebagai berikut :

- a. Menggunakan dataset yang berasal dari data postingan twitter berdasarkan kata kunci Gojek, Grab, Maxim dan InDrive dengan total data 10.000 data.
- b. Dataset yang digunakan merupakan data berformat text yang didapatkan melalui media API Twitter.
- c. Dataset yang digunakan bersumber dari postingan Twitter dengan komposisi data yang terbagi menjadi 3 bagian yaitu positif, negatif dan netral
- d. Jumlah pembagian dataset untuk data training dan testing adalah 20% untuk testing dan 80% untuk training (Muraina, 2022).
- e. Dataset yang digunakan adalah data postingan pendapat masyarakat Indonesia mengenai Gojek, Grab, Maxim dan InDrive tanpa memiliki kecenderungan dengan golongan atau instansi manapun
- f. Fokus penelitian adalah membandingkan algoritma DNN dan algoritma LSTM untuk klasifikasi.
- g. Analisis data dilakukan melalui pre-processing menggunakan teknik proses case folding, tokenizing, filtering, dan stemming.
- h. Algoritma yang digabungkan adalah algoritma DNN dan algoritma LSTM untuk klasifikasi.
- i. Model pengukuran validasi performa model algoritma klasifikasi menggunakan metode atau teknik K-Fold Cross-Validation
- j. Evaluasi performa berdasarkan akurasi, presisi, recall, F1-score, waktu pelatihan, dan MAE.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah disebutkan sebelumnya, tujuan penelitian ini adalah untuk mengetahui nilai performa berdasarkan hasil penelitian dan implementasi DNN dan algoritma LSTM berdasarkan nilai performa akurasi, presisi, f1 score, recall, waktu training dan MAE terhadap dataset twitter transportasi online.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini, antara lain:

- a. Penelitian ini akan memberikan wawasan baru tentang efektivitas algoritma Deep Neural Network (DNN) dan Long Short-Term Memory (LSTM) dalam analisis sentimen berbasis data Twitter, khususnya di Indonesia.
- b. Hasil penelitian dapat digunakan sebagai referensi bagi peneliti selanjutnya yang ingin mengeksplorasi topik analisis sentimen atau pengembangan algoritma untuk data media sosial.



BAB II TINJAUAN PUSTAKA

2.1 Tinjauan Pustaka

Penelitian (Safira et al., 2023) berjudul *Sentiment Analysis to Find Out Positive or Negative Opinions on Ride Hailing Application*. Penelitian ini menganalisis opini pengguna mengenai layanan Gojek yang diekspresikan melalui Twitter. Penelitian ini mengevaluasi sentimen pengguna terhadap berbagai aspek layanan Gojek seperti pengantaran makanan (GoFood), transportasi (GoRide dan GoCar), serta layanan terkait lainnya. Studi ini menunjukkan bagaimana opini masyarakat dapat mempengaruhi penilaian terhadap kualitas layanan. Data dikumpulkan menggunakan Twitter API dengan kata kunci terkait layanan Gojek. Sentimen pada tweet diklasifikasikan menggunakan metode Support Vector Machine (SVM). Data melewati tahapan preprocessing seperti konversi emotikon, pembersihan data, tokenisasi, case folding, penghapusan stopwords, dan stemming sebelum dilakukan klasifikasi. Model SVM yang digunakan mencapai akurasi 80,3% dalam mengklasifikasikan sentimen positif dan negatif pada tweet terkait Gojek. Model ini mampu memberikan gambaran yang akurat mengenai sentimen pengguna terhadap layanan yang diberikan oleh Gojek.

Penelitian (Rakshit, Sarkar, & Roy, 2024) berjudul "Hybrid Deep Learning Approach for Sentiment Analysis on Twitter Data" membahas bagaimana pendekatan deep learning dapat digunakan untuk menganalisis sentimen dari data Twitter dalam jumlah besar, mencakup 1,6 juta tweet dengan sentimen positif dan negatif. Fokus utama penelitian adalah meningkatkan akurasi klasifikasi sentimen dengan mempertimbangkan pentingnya keberadaan stopwords. Data tweet yang diproses dengan Word2Vec untuk menghasilkan vektor kata. Model yang dibangun mencakup RNN, Bi-directional RNN, dan LSTM. Data dibagi dengan rasio 80:20 untuk pelatihan dan pengujian. Proses pelatihan menggunakan embedding layer dan dilakukan eksperimen dengan mempertahankan dan menghapus stopwords. Model LSTM menunjukkan akurasi tertinggi sebesar 83,5% dan F1 score 84,6%. Hasil menunjukkan bahwa mempertahankan stopwords dalam data meningkatkan

performa model secara signifikan. Model yang dikembangkan dapat digunakan untuk mengklasifikasikan sentimen dengan baik pada data Twitter.

Penelitian (Pullanikkat et al., 2024) yang berjudul "Utilizing the Twitter social media to identify transportation-related grievances in Indian cities" mengeksplorasi bagaimana Twitter dapat digunakan untuk mengidentifikasi keluhan spesifik mengenai sistem transportasi di kota-kota besar India, seperti keterlambatan, ketersediaan transportasi, dan sikap petugas. Hasilnya diharapkan dapat memberikan wawasan bagi pemerintah dan otoritas transportasi dalam merespon keluhan publik. Data dikumpulkan menggunakan Twitter Search API dari enam kota besar di India selama 15 hari pada Maret 2023. Data diklasifikasikan menggunakan model RoBERTa dan BERTweet untuk klasifikasi sentimen dan model multi-label untuk mengidentifikasi jenis keluhan. Sebanyak 12.000 tweet diolah untuk membentuk dataset yang digunakan dalam pelatihan model. Model berbasis CrisisTransformer-M3 memberikan akurasi tertinggi dalam klasifikasi keluhan transportasi, dengan akurasi 91,4% dan F1 score 91,3%. Studi ini menunjukkan bahwa media sosial seperti Twitter dapat digunakan secara efektif untuk mengidentifikasi keluhan transportasi secara rinci, yang berguna bagi pemerintah untuk perencanaan transportasi.

Penelitian (Urolagin & Patel, 2024) berjudul "User-Specific Loyalty Measure and Prediction Using Deep Neural Network From Twitter Data" Menyiapkan basis data nyata dari Twitter yang berisi data dari lima platform pembelajaran online populer (Coursera, Udemy, Udacity, Khanacademy, dan edX). Menciptakan metrik pada tingkat pengguna dan produk untuk mengukur pengaruh antar pengguna, kepuasan, dan loyalitas pelanggan menggunakan data media sosial dan melakukan prediksi skor sentimen pengguna (User Sentiment Score/USS) untuk mengidentifikasi pelanggan yang loyal dan tidak loyal melalui jaringan saraf padat (Dense Neural Network/DNN) serta membandingkan kinerjanya dengan model pembelajaran mesin lainnya seperti Random Forest, XGBoost, dan SVC. Peneliti mendapati hasil Khanacademy memiliki Skor Pengaruh Data Sosial (SDIS) tertinggi, menunjukkan pengaruh positif yang kuat terhadap pengguna baru, sementara Udemy memiliki jumlah tweet terbanyak tetapi pengaruh positifnya lebih rendah dibandingkan yang lain, DNN mencapai akurasi klasifikasi tertinggi

sebesar 98,62% dalam memprediksi loyalitas pengguna, lebih unggul dibandingkan dengan model Random Forest, XGBoost, dan SVC.

Penelitian (Vyas et al., 2022) berjudul "Automated Classification of Societal Sentiments on Twitter With Machine Learning", mengembangkan kerangka kerja otomatis yang menggabungkan teknik berbasis leksikon dan pembelajaran mesin (ML) untuk mengklasifikasikan sentimen masyarakat pada tweet terkait COVID-19. Kerangka ini bertujuan untuk mengidentifikasi sentimen positif, negatif, dan netral dari tweet dan membantu memahami dinamika sentimen masyarakat selama peristiwa besar seperti pandemi COVID-19. Kerangka kerja hybrid yang dikembangkan menggabungkan analisis sentimen berbasis leksikon menggunakan VADER untuk labeling otomatis dengan teknik pembelajaran terawasi seperti LSTM untuk klasifikasi tweet. Hasil evaluasi menunjukkan bahwa model LSTM memberikan akurasi tertinggi sebesar 83% dibandingkan dengan teknik ML lainnya, seperti Gaussian Naïve Bayes dan Random Forest. Sebagian besar sentimen yang diidentifikasi adalah positif (38.5%) dan netral (34.7%). Ini menunjukkan respons masyarakat yang cenderung optimis terhadap situasi pandemi.

Penelitian (Qi & Shabrina, 2023) berjudul "Sentiment analysis using Twitter data: a comparative application of lexicon- and machine-learning-based approach", menganalisis sentimen publik terhadap Covid-19 menggunakan data Twitter selama lockdown ketiga di Inggris. Selain itu, studi ini membandingkan efektivitas metode analisis sentimen berbasis leksikon (TextBlob, VADER, SentiWordNet) dan metode berbasis pembelajaran mesin (Random Forest, Multinomial Naïve-Bayes, SVC) untuk membantu peneliti dalam memilih metode yang tepat untuk analisis NLP. Hasil penelitian didapati Pendekatan Leksikon: VADER unggul dalam menangani bahasa informal di Twitter, menghasilkan analisis sentimen yang lebih akurat dibandingkan TextBlob dan SentiWordNet. Analisis menunjukkan sentimen positif lebih dominan dalam topik terkait vaksin Covid-19. Pendekatan Pembelajaran Mesin: Model SVC dengan fitur TF-IDF memiliki akurasi tertinggi (71%). Model Word2Vec menunjukkan kinerja yang kurang optimal karena keterbatasan data pelatihan.

2.2 Landasan Teori

Transportasi online adalah layanan transportasi yang menggunakan platform berbasis internet untuk menghubungkan penumpang dengan pengemudi. Berbeda dengan metode tradisional, layanan ini menggunakan aplikasi digital untuk melakukan pemesanan, mengatur tarif, dan mengoordinasikan perjalanan (Mitropoulos, Kortsari, & Ayfantopoulou, 2021)

Twitter adalah platform media sosial yang memungkinkan pengguna untuk berbagi pesan singkat, yang dikenal sebagai tweet, dengan batas karakter tertentu (awalnya 140 karakter, sekarang 280 karakter). Pengguna dapat berbagi teks, gambar, video, atau tautan, dan berinteraksi dengan tweet pengguna lain melalui fitur seperti balasan, retweet, dan tanda suka. Platform ini sering digunakan sebagai sumber berita dan diskusi publik, karena memungkinkan informasi menyebar dengan cepat tanpa memerlukan persetujuan khusus dari pengguna lain (Safira et al., 2023)

API Twitter adalah antarmuka pemrograman aplikasi (Application Programming Interface) yang disediakan oleh Twitter untuk memungkinkan pengembang berinteraksi dengan data dan layanan Twitter secara programatik. API ini memungkinkan pengembang untuk mengakses tweet, mengambil data pengguna, memposting tweet, melacak tren, mengumpulkan tweet berdasarkan kata kunci atau hashtag, serta berbagai fungsi lainnya yang terkait dengan platform Twitter. (Twitter, 2024). Untuk mendapatkan akses ke API Twitter, dibutuhkan akun pengembang di Twitter dan mendaftarkan aplikasi yang akan menggunakan API tersebut.

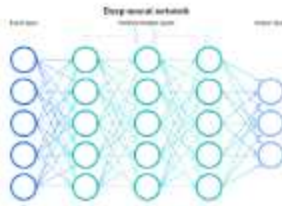
Menurut Pang, B., & Lee, L. (2008) dalam makalah mereka yang berjudul "Opinion mining and sentiment analysis", **analisis sentimen** adalah proses otomatisasi untuk mengidentifikasi, mengekstrak, dan mengklasifikasikan opini dari teks, yang berfokus pada penentuan sikap atau sentimen dari penulis terhadap suatu entitas, topik, atau kejadian tertentu. Analisis ini bertujuan untuk memahami apakah sikap yang diekspresikan dalam teks tersebut bersifat positif, negatif, atau netral.

Deep Neural Network (DNN) merupakan salah satu bentuk lanjutan dari jaringan saraf tiruan (Artificial Neural Network) yang memiliki lebih dari satu lapisan tersembunyi (hidden layer) antara lapisan input dan lapisan output. Model ini dikembangkan sebagai bagian dari pendekatan Deep Learning yang bertujuan untuk memungkinkan komputer belajar dari representasi data yang kompleks dan berlapis-lapis. Setiap neuron dalam jaringan saling terhubung secara terstruktur, di mana keluaran dari satu lapisan menjadi masukan bagi lapisan berikutnya.

Arsitektur DNN terdiri dari tiga komponen utama: lapisan input yang menerima data mentah (misalnya vektor kata, piksel gambar, atau sinyal audio), lapisan tersembunyi yang menangani proses transformasi nonlinier terhadap data, serta lapisan output yang menghasilkan prediksi akhir sesuai dengan tugas yang ditetapkan, seperti klasifikasi atau regresi. Proses pembelajaran dalam DNN dilakukan melalui algoritma backpropagation dengan bantuan fungsi aktivasi seperti ReLU, Sigmoid, atau Tanh untuk memperkenalkan nonlinieritas.

Setiap lapisan dalam DNN bertugas menangkap representasi fitur yang semakin abstrak dan kompleks seiring bertambahnya kedalaman jaringan. Lapisan awal biasanya mengekstrak fitur dasar, sementara lapisan-lapisan berikutnya mempelajari pola yang lebih tinggi atau konteks semantik yang lebih dalam. Kemampuan ini membuat DNN sangat efektif dalam berbagai bidang seperti klasifikasi gambar, pemrosesan bahasa alami (Natural Language Processing), pengenalan ucapan, hingga deteksi anomali.

Keunggulan utama DNN terletak pada kemampuannya dalam mempelajari representasi fitur secara hierarkis dari data masukan tanpa perlu rekayasa fitur manual yang ekstensif. Semakin dalam arsitektur jaringan, semakin tinggi pula kompleksitas fitur yang dapat dipelajari oleh model. Hal ini menjadikan DNN sangat efektif untuk tugas-tugas klasifikasi teks, sebagaimana ditunjukkan dalam kajian komprehensif oleh Zhang et al. (2023). Struktur umum dari Deep Neural Network ditampilkan pada Gambar 2.1 sebagai ilustrasi dari alur pemrosesan dan hubungan antar-lapisannya.

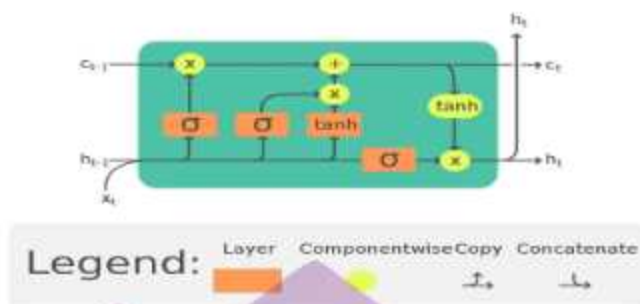


Gambar 2. 1 Arsitektur Deep Neural Network (DNN)

Long Short-Term Memory (LSTM) adalah salah satu jenis arsitektur jaringan saraf tiruan (neural network) yang dikembangkan sebagai penyempurnaan dari model Recurrent Neural Network (RNN). LSTM dirancang untuk mengatasi kelemahan utama RNN standar, yaitu ketidakmampuannya dalam mempertahankan informasi jangka panjang akibat fenomena yang dikenal sebagai vanishing gradient problem. Masalah ini membuat RNN sulit mempelajari ketergantungan jarak jauh dalam data sekuensial.

LSTM memiliki struktur khusus berupa unit memori (memory cell) yang dilengkapi dengan tiga gerbang utama, yaitu input gate, forget gate, dan output gate. Ketiga gerbang ini mengatur aliran informasi yang masuk, disimpan, dan dikeluarkan dari unit memori, sehingga memungkinkan LSTM untuk secara selektif mempertahankan atau melupakan informasi berdasarkan konteks. Dengan mekanisme tersebut, LSTM mampu menangani data berurutan dengan ketergantungan temporal yang kuat, seperti dalam kasus pemrosesan teks, pengenalan ucapan, analisis sentimen, dan prediksi deret waktu.

Kemampuan LSTM dalam menyimpan informasi selama periode yang lebih panjang membuatnya unggul dalam berbagai aplikasi Natural Language Processing (NLP), serta telah mengalami banyak kemajuan dalam menangani berbagai jenis data sekuensial, termasuk teks dan time series (Sun et al., 2023). Arsitektur lengkap LSTM, termasuk struktur gerbang dan alur data, ditunjukkan secara visual pada Gambar 2.2.



Gambar 2. 2 Arsitektur jaringan LSTM untuk data sekuensial.

Case folding adalah proses mengubah seluruh huruf dalam teks menjadi huruf kecil (lowercase). Tujuannya adalah untuk menyamakan bentuk huruf sehingga tidak ada perbedaan antara huruf besar dan huruf kecil, seperti "Data" dan "data" dianggap sama.

Contoh:

Input : "Machine Learning adalah Masa Depan Teknologi."

Output : "machine learning adalah masa depan teknologi."

Manfaat :

1. Mengurangi keragaman data karena perbedaan huruf.
2. Menstandarkan teks sebelum analisis lebih lanjut.

Tokenizing adalah proses memecah teks menjadi unit-unit kecil yang disebut token. Token ini biasanya berupa kata, frasa, atau karakter, tergantung pada kebutuhan analisis.

Contoh:

Input : "Teknologi AI sedang berkembang pesat."

Output : ["Teknologi", "AI", "sedang", "berkembang", "pesat"]

Manfaat :

1. Membagi teks menjadi elemen-elemen yang lebih kecil untuk dianalisis.
2. Mempermudah penghitungan kata atau analisis frekuensi.

Filtering adalah proses menyaring atau menghapus elemen-elemen yang tidak relevan dari teks. Biasanya, elemen yang dihapus adalah **stopwords** (kata-

kata umum seperti "dan", "di", "yang", "adalah") atau simbol tertentu yang tidak memiliki arti penting dalam analisis.

Contoh:

Input : ["Teknologi", "AI", "sedang", "berkembang", "pesat"]

Stopwords : ["sedang"]

Output : ["Teknologi", "AI", "berkembang", "pesat"]

Manfaat:

1. Mengurangi noise pada data.
2. Memfokuskan analisis pada kata-kata yang lebih bermakna.

Embedding adalah representasi numerik dari data, khususnya data teks atau kategori, dalam bentuk vektor berdimensi tetap yang digunakan dalam pembelajaran mesin dan pemrosesan bahasa alami (Natural Language Processing/NLP). Embedding memungkinkan komputer memahami dan memproses data non-numerik dengan cara yang lebih efisien.

IndoBERT adalah varian model BERT (Bidirectional Encoder Representations from Transformers) yang dirancang khusus untuk Bahasa Indonesia. Dibangun dengan arsitektur Transformer, IndoBERT menawarkan berbagai kelebihan dibandingkan static embedding seperti Word2Vec atau GloVe (Zhou et al., 2023).

Perbedaan mendasar antara **static embedding** dan **transformer embedding** seperti IndoBERT terletak pada cara keduanya merepresentasikan kata dalam model bahasa. Static embedding merepresentasikan kata secara **statis**, artinya satu kata akan selalu memiliki representasi vektor yang sama tanpa memandang konteks kalimat. Sebaliknya, IndoBERT sebagai bagian dari model transformer menghasilkan **representasi kontekstual**, di mana satu kata dapat memiliki makna vektor yang berbeda tergantung pada konteks kalimatnya. Sebagai contoh, kata "bank" akan tetap direpresentasikan sama dalam static embedding, tetapi dalam IndoBERT, kata tersebut akan direpresentasikan secara berbeda jika digunakan dalam konteks "menabung di bank" dibandingkan dengan "duduk di tepi bank sungai."

Dalam hal **pemahaman konteks**, static embedding seperti Word2Vec atau GloVe hanya mengandalkan **statistik hubungan kata** seperti *co-occurrence* dan

tidak memiliki mekanisme untuk menangkap makna konteks kalimat secara menyeluruh. Sementara itu, IndoBERT dibangun di atas arsitektur **Transformer** yang memungkinkan model memahami **hubungan antar kata di seluruh teks**, bukan hanya secara lokal. Kemampuan ini membuat IndoBERT unggul dalam menangani **polisemi**, yaitu kemampuan membedakan makna kata ganda berdasarkan konteks penggunaannya.

Dari segi **transfer learning**, static embedding tidak dirancang untuk digunakan ulang dalam berbagai tugas NLP secara fleksibel. Sebaliknya, IndoBERT sangat cocok untuk pendekatan **fine-tuning**, memungkinkan model yang sama diadaptasi untuk berbagai tugas seperti klasifikasi teks, analisis sentimen, dan ekstraksi entitas, hanya dengan pelatihan tambahan minimal.

Kelebihan IndoBERT sebagai Transformer Embedding

1. Representasi Kontekstual IndoBERT menghasilkan embedding kata yang kontekstual. Artinya, representasi kata dalam vektor embedding berubah tergantung pada kalimat atau konteksnya.

o *Contoh:*

- "Saya menyimpan uang di bank."
- "Saya duduk di tepi bank sungai."

Dalam IndoBERT, embedding untuk "bank" akan berbeda pada kedua kalimat ini, sesuai dengan maknanya.

2. Pretraining pada Bahasa Indonesia IndoBERT dilatih secara spesifik dengan korpus teks Bahasa Indonesia, seperti berita, buku, dan media sosial. Hal ini membuatnya lebih cocok untuk memahami struktur bahasa, idiom, atau karakteristik unik dari Bahasa Indonesia. Penelitian oleh Yang et al. (2023) menyoroti kemajuan signifikan dalam pengembangan model NLP untuk Bahasa Indonesia, termasuk tantangan dan kebutuhan akan model kontekstual seperti IndoBERT.

3. Kemampuan Generalisasi Karena menggunakan Transformer dengan mekanisme *self-attention*, IndoBERT dapat memahami hubungan antar kata dalam kalimat panjang, memungkinkan generalisasi yang lebih baik dibandingkan embedding statis.

4. Fleksibilitas Tugas NLP IndoBERT dapat digunakan dalam berbagai tugas NLP dengan fine-tuning:
 - o Klasifikasi teks.
 - o Analisis sentimen.
 - o Named Entity Recognition (NER).
 - o Question Answering.
5. Kemampuan Transfer Learning Model pretrained IndoBERT dapat digunakan pada dataset kecil dengan performa tinggi setelah dilakukan fine-tuning, sehingga menghemat waktu dan sumber daya.
6. Dukungan Konteks Polisemi dan Sinonim IndoBERT menangkap makna polisemik (kata dengan makna ganda) dan sinonim dengan lebih baik. Ini memungkinkan analisis teks lebih kaya.
 - o *Contoh Sinonim:*
 - "Pintar" dan "cerdas" akan memiliki embedding yang dekat dalam ruang vektor karena konteksnya mirip.
7. Performansi Tinggi dalam Bahasa Indonesia Di bandingkan embedding statis global seperti Word2Vec atau GloVe yang dilatih pada data multibahasa atau berbahasa Inggris, IndoBERT dirancang khusus untuk menangani nuansa Bahasa Indonesia, termasuk imbuhan (*ber-, di-, ke-, dll.*), akronim, atau struktur kalimat.

Kelemahan Static Embedding yang Diatasi IndoBERT

1. Tidak Kontekstual: Word2Vec atau GloVe menghasilkan representasi kata tetap (static), sehingga tidak mempertimbangkan konteks.
2. Sulit Menangani Polisemik: Kata dengan makna ganda (homonim) sulit diolah dengan static embedding.
3. Kurang Relevan untuk Bahasa Indonesia: Embedding statis seperti Word2Vec yang dilatih pada korpus berbahasa Inggris cenderung tidak optimal untuk teks berbahasa Indonesia.

Contoh Penerapan IndoBERT

Misalnya, untuk kalimat:

"IndoBERT sangat berguna dalam analisis teks berbahasa Indonesia."

IndoBERT menghasilkan embedding untuk kata-kata seperti "berguna" dan "analisis" yang mempertimbangkan seluruh kalimat. Representasi ini lebih kaya

dibandingkan static embedding, yang hanya mempertimbangkan frekuensi atau co-occurrence.

TensorFlow adalah framework open-source untuk machine learning dan deep learning yang dikembangkan oleh Google Brain sejak tahun 2015. Framework ini dirancang untuk memfasilitasi pembangunan, pelatihan, dan penerapan model kecerdasan buatan, mulai dari skala penelitian hingga produksi. Dengan konsep dasar tensor sebagai representasi data multidimensi dan flow sebagai aliran data melalui computational graph, TensorFlow mampu melakukan komputasi numerik secara efisien di berbagai perangkat, termasuk CPU, GPU, dan TPU. Kekuatan TensorFlow terletak pada fleksibilitasnya, dukungan lintas platform, serta ekosistem lengkap yang mencakup TensorBoard, TF-Data, TensorFlow Lite, hingga TensorFlow.js.

Confusion Matrix (Matriks kebingungan) adalah alat evaluasi yang penting dalam analisis klasifikasi yang memungkinkan kita untuk memvisualisasikan performa model klasifikasi dengan lebih detail. Matriks ini menggambarkan jumlah prediksi yang benar dan salah yang dibuat oleh model pada set data uji. Seperti ditunjukkan pada Gambar 2.3, biasanya, matriks kebingungan berisi empat sel,



The diagram shows a 2x2 confusion matrix. The vertical axis is labeled 'Nilai Prediksi' (Predicted Value) and the horizontal axis is labeled 'Nilai Aktual' (Actual Value). The matrix cells are: Top-Left (Positive Predicted, Positive Actual) is TP; Top-Right (Positive Predicted, Negative Actual) is FP; Bottom-Left (Negative Predicted, Positive Actual) is FN; Bottom-Right (Negative Predicted, Negative Actual) is TN.

	Positive	Negative
Positive	TP	FP
Negative	FN	TN

Gambar 2. 3 Confusion matrix

1. **True Positive (TP):** Jumlah sampel positif yang diprediksi dengan benar oleh model.

2. **False Positive (FP):** Jumlah sampel negatif yang salah diprediksi sebagai positif oleh model.
3. **False Negative (FN):** Jumlah sampel positif yang salah diprediksi sebagai negatif oleh model.
4. **True Negative (TN):** Jumlah sampel negatif yang diprediksi dengan benar oleh model.

Dengan matriks kebingungan, kita dapat menghitung berbagai metrik performa klasifikasi seperti akurasi, presisi, recall, dan F1-score. Matriks kebingungan sangat penting dalam mengevaluasi dan memahami performa model klasifikasi, terutama dalam konteks di mana distribusi kelas tidak seimbang. Dengan melihat matriks kebingungan, kita dapat mendapatkan wawasan yang lebih mendalam tentang kekuatan dan kelemahan dari model klasifikasi yang digunakan.

Akurasi adalah metrik yang paling sederhana dan umum digunakan untuk mengukur performa klasifikasi. Akurasi menghitung persentase dari prediksi yang benar dibandingkan dengan total jumlah sampel. Meskipun akurasi berguna untuk mengevaluasi performa secara keseluruhan, dapat menjadi tidak informatif jika distribusi kelas tidak seimbang.

Rumus Akurasi :

$$\text{Akurasi} = \frac{\text{Jumlah prediksi yang benar}}{\text{Total sampel}} \times 100\% \quad (2.1)$$

Presisi mengukur proporsi dari prediksi positif yang benar dibandingkan dengan total prediksi positif yang dibuat oleh model. Ini memberikan informasi tentang seberapa banyak dari prediksi positif yang benar-benar relevan.

Rumus Presisi :

$$\text{Presisi} = \frac{\text{Jumlah True Positiv}}{\text{Jumlah True Positiv} + \text{Jumlah False Positiv}} \quad (2.2)$$

Recall mengukur proporsi dari instance positif yang telah diidentifikasi dengan benar oleh model, dibandingkan dengan total jumlah instance positif yang ada dalam data. Ini memberikan informasi tentang seberapa efektif model dalam menemukan semua instance dari kelas yang diinginkan.

Rumus Recall :

$$\text{Recall (Sensitivity)} = \frac{TP}{(TP + FN)} \quad (2.3)$$

F1-Score adalah rata-rata harmonik dari presisi dan recall. Ini memberikan kompromi antara presisi dan recall, dan berguna ketika distribusi kelas tidak seimbang.

Rumus F1-Score :

$$\text{F1-Score} = \frac{2 (\text{Presisi} \times \text{Recall})}{(\text{Presisi} + \text{Recall})} \quad (2.4)$$

Mean Absolute Error (MAE) adalah metrik evaluasi yang digunakan untuk mengukur seberapa dekat hasil prediksi suatu model dengan nilai yang sebenarnya. MAE mengukur rata-rata dari selisih absolut antara nilai prediksi dan nilai sebenarnya. Semakin rendah nilai MAE, semakin baik kinerja model.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2.5)$$

di mana:

n adalah jumlah total data.

y_i adalah nilai aktual ke i .

\hat{y}_i adalah nilai prediksi ke i

$| \cdot |$ adalah fungsi nilai absolut (absolute value).

\sum menunjukkan penjumlahan dari setiap nilai dalam rentang dari 1 hingga n .

K-Fold Cross-Validation adalah teknik evaluasi model dalam machine learning yang digunakan untuk mengukur kemampuan generalisasi model secara efektif. Teknik ini membagi dataset menjadi beberapa subset (atau folds), dan model dilatih serta diuji secara bergantian menggunakan kombinasi subset tersebut. Pendekatan ini memberikan evaluasi kinerja yang lebih akurat dibandingkan pembagian data sederhana (train-test split), terutama pada dataset yang terbatas.

Proses K-Fold Cross-Validation

1. Pembagian Dataset

Dataset dibagi secara acak ke dalam k subset yang berukuran hampir sama. Tiap subset disebut sebagai satu fold. Jumlah folds (k) biasanya dipilih berdasarkan ukuran dataset, dengan nilai umum seperti 5 atau 10.

2. Iterasi Pengujian

Model dilatih dan diuji sebanyak k kali, di mana pada setiap iterasi:

Data Pelatihan: Gabungan dari $k-1$ subset digunakan untuk melatih model.

Data Pengujian: Satu subset yang tersisa digunakan untuk menguji model.

3. Perhitungan Performa

Kinerja model (misalnya, akurasi, presisi, atau F1-score) dihitung pada setiap iterasi. Setelah semua iterasi selesai, rata-rata dari metrik performa dihitung untuk memberikan gambaran keseluruhan kinerja model.

Keunggulan K-Fold Cross-Validation

1. Memaksimalkan Penggunaan Data

Semua data digunakan baik untuk pelatihan maupun pengujian, sehingga informasi dari dataset dimanfaatkan secara optimal.

2. Mengurangi Variasi Evaluasi

Dengan melakukan evaluasi pada berbagai subset data, K-Fold Cross-Validation mengurangi pengaruh pembagian data yang mungkin bias pada train-test split sederhana.

3. Cocok untuk Dataset Kecil

Pada dataset yang terbatas, pembagian train-test yang sederhana sering kali tidak cukup mewakili distribusi data. Teknik ini membantu mengatasi masalah tersebut.

4. Evaluasi yang Lebih Akurat

Karena model diuji pada semua data dengan berbagai kombinasi pelatihan dan pengujian, hasil evaluasi lebih representatif terhadap kemampuan generalisasi model.

Variasi K-Fold Cross-Validation

1. Stratified K-Fold Cross-Validation

Teknik ini memastikan distribusi label (pada data klasifikasi) tetap seimbang di setiap fold. Cocok untuk dataset yang memiliki distribusi kelas tidak seimbang.

2. Leave-One-Out Cross-Validation (LOOCV)

Variasi khusus di mana k sama dengan jumlah total data, sehingga pada setiap iterasi, hanya satu data digunakan untuk pengujian sementara sisanya untuk pelatihan. LOOCV memberikan evaluasi yang sangat detail tetapi memerlukan waktu komputasi yang lebih tinggi.

3. Repeated K-Fold Cross-Validation

Melakukan K-Fold Cross-Validation beberapa kali dengan **pembagian data** yang berbeda pada setiap iterasi. Teknik ini memberikan hasil evaluasi yang lebih stabil.

Contoh Proses dengan $K = 5$

Jika dataset terdiri dari 100 data dan $k = 5$:

Dataset dibagi menjadi 5 subset (masing-masing 20 data).

Iterasi 1: Train pada subset 2-5, tes pada subset 1.

Iterasi 2: Train pada subset 1, 3-5, tes pada subset 2.

Iterasi 3: Train pada subset 1-2, 4-5, tes pada subset 3.

Iterasi 4: Train pada subset 1-3, 5, tes pada subset 4.

Iterasi 5: Train pada subset 1-4, tes pada subset 5.

Rata-rata dari hasil evaluasi di semua iterasi menjadi skor akhir model.

Kekurangan K-Fold Cross-Validation

1. Waktu Komputasi

Teknik ini membutuhkan model untuk dilatih dan diuji sebanyak k kali, sehingga membutuhkan waktu lebih lama dibandingkan metode train-test split sederhana.

2. Kompleksitas

Untuk dataset besar atau algoritma dengan kompleksitas tinggi, K-Fold Cross-Validation dapat menjadi mahal secara komputasi.

2.3 Keaslian Penelitian

Tabel 2. 1 Matriks literatur review dan posisi penelitian
Analisis Sentimen Pengguna Terhadap Layanan Transportasi Online Menggunakan Data Twitter

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1	Hybrid Deep Learning Approach for Sentiment Analysis on Twitter Data	Pranati Rakshit, Pronit Sarkar, dan Shubhankar Roy Jurnal: Multimedia Tools and Applications, Springer Nature, 2024.	Mengembangkan model analisis sentimen menggunakan metode deep learning pada data Twitter. Tujuan utamanya adalah menghasilkan model yang kuat dengan akurasi tinggi untuk menganalisis sentimen tweet dengan jumlah data yang besar,	Model Long Short-Term Memory (LSTM) menunjukkan hasil terbaik dalam klasifikasi sentimen, dengan akurasi 83,5% dan skor F1 sebesar 84,6%. Penelitian ini menemukan bahwa menyertakan stopwords dapat meningkatkan akurasi model untuk	Tantangan dalam model ini meliputi masalah memori ketika menangani data dalam volume besar dan kebutuhan akan dukungan GPU untuk proses pelatihan. Sebagai pengembangan lanjutan, penelitian ini merekomendasikan penggunaan metode	Dataset , Penelitian ini menggunakan dataset besar sebanyak 1,6 juta tweet yang terdiri dari 800.000 tweet positif dan 800.000 tweet negatif. Ini jauh lebih besar dibandingkan penelitian sebelumnya, seperti Go et al. yang menggunakan data kecil dengan pendekatan distant learning.

termasuk eksperimen dengan mempertahankan kata-kata stopwords dalam preprocessing untuk hasil yang lebih akurat.

analisis sentimen, berbeda dengan pendekatan sebelumnya yang biasanya menghilangkan stopwords.

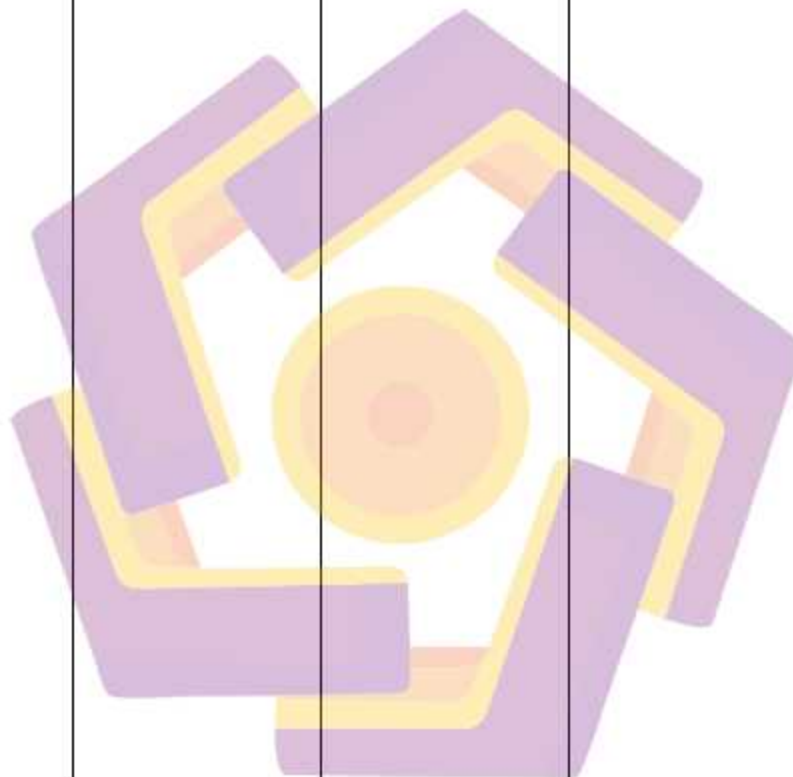
embedding seperti BERT untuk peningkatan akurasi, serta optimasi hyper-parameters.

Pendekatan Model,

Penelitian ini memanfaatkan model berbasis deep learning, yaitu RNN, Bi-directional RNN, dan LSTM. Hal ini berbeda dari penelitian awal seperti Barbosa dan Feng, yang hanya menggunakan fitur manual dan algoritma tradisional seperti Naive Bayes atau SVM.

Penggunaan Stopwords,

Penelitian ini mengevaluasi efek stopwords pada performa model, menemukan bahwa mempertahankan stopwords menghasilkan akurasi yang



lebih tinggi (83,5% dengan LSTM). Penelitian sebelumnya sering menghapus stopwords tanpa mengevaluasi dampaknya, seperti yang dilakukan oleh Hassan Saif et al.

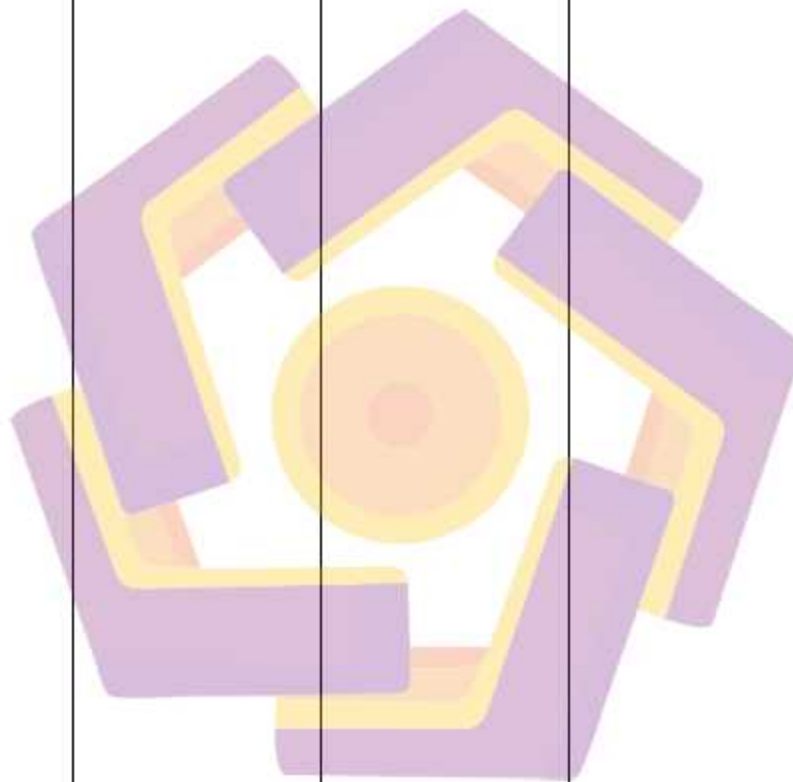
Akurasi dan Evaluasi,

Model LSTM dengan dataset besar menghasilkan akurasi 83,5% dan F1 score 84,6%. Ini lebih tinggi dibandingkan dengan pendekatan klasik seperti Semantic Smoothing (80,7%) atau model berbasis SVM (70%).

Penggunaan Teknologi Deep Learning, tidak

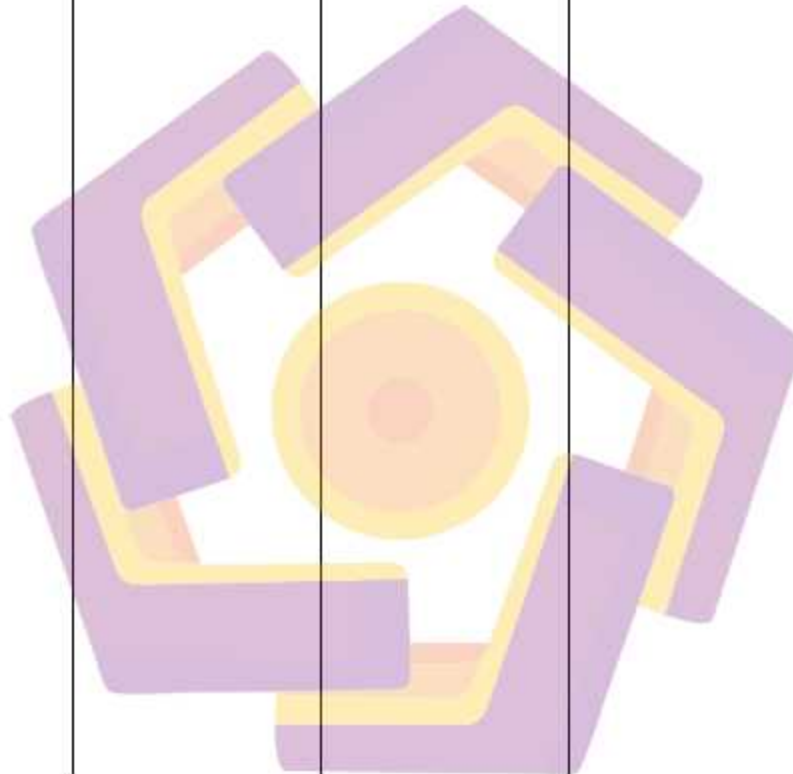
						seperti penelitian awal yang lebih fokus pada fitur manual, penelitian ini mengintegrasikan embedding Word2Vec dan arsitektur deep learning untuk mengatasi kekurangan dari pendekatan tradisional, seperti ketidakmampuan memahami konteks panjang dan ironi.
2	Utilizing the Twitter social media to identify transportation-related	Peneliti: Rahul Pullanikkat, Soham Poddar, Anik Das, Tushar Jaiswal, Vivek Kumar	Penelitian ini bertujuan untuk mengidentifikasi dan mengklasifikasikan keluhan masyarakat terkait transportasi di kota-kota India	Studi ini menunjukkan bahwa Twitter bisa menjadi sumber data yang efektif untuk mengidentifikasi keluhan spesifik terkait	Saran yang diberikan mencakup perlunya melakukan pengumpulan data berkala untuk analisis jangka panjang. Kelemahan yang	Cakupan Data dan Lingkup Analisis , Sebagian besar penelitian sebelumnya hanya fokus pada satu moda transportasi (contoh: metro di Delhi oleh Agrawal dan

<p>grievances in Indian cities</p>	<p>Singh, Moumita Basu, dan Saptarshi Ghosh. Media Publikasi: Social Network Analysis and Mining. Tahun: 2024</p>	<p>menggunakan data dari media sosial Twitter. Studi ini mengembangkan dataset yang mengklasifikasikan tweet berdasarkan sentimen dan kategori keluhan spesifik untuk mendukung perencanaan transportasi</p>	<p>transportasi di kota-kota besar India. Model berbasis Transformer, seperti Crisis Transformer, efektif dalam mengklasifikasikan tweet ke dalam kategori keluhan yang lebih mendetail. Hasil ini menunjukkan potensi penggunaan media sosial untuk menggantikan survei yang memerlukan biaya tinggi dalam mengumpulkan opini masyarakat</p>	<p>diidentifikasi meliputi keterbatasan pengumpulan data hanya selama 15 hari dan kemungkinan bias pada data karena hanya melibatkan pengguna Twitter usia muda. Selain itu, penelitian ini tidak dapat memverifikasi kebenaran keluhan di dalam tweet</p>	<p>Kuriakose, 2022) atau satu kota (contoh: Greater Mumbai oleh Das, 2021). Penelitian Ini, Menganalisis data dari berbagai moda transportasi di enam kota besar di India (Delhi, Mumbai, Bangalore, Kolkata, Hyderabad, Chennai) untuk memberikan gambaran yang lebih luas dan holistik mengenai keluhan transportasi. Fokus pada Keluhan Spesifik, Penelitian Sebelumnya, Sebagian besar studi hanya berfokus pada analisis sentimen umum (positif/negatif/neutral) atau</p>
------------------------------------	---	--	---	--	--



keluhan tingkat tinggi seperti kemacetan atau kepuasan umum. **Penelitian Ini,** Mengklasifikasikan keluhan secara lebih spesifik ke dalam 11 kategori, termasuk keterlambatan, kondisi fisik kendaraan, perilaku personel, dan ketersediaan transportasi. Ini merupakan pendekatan multi-label yang memungkinkan satu tweet mengandung beberapa kategori keluhan.

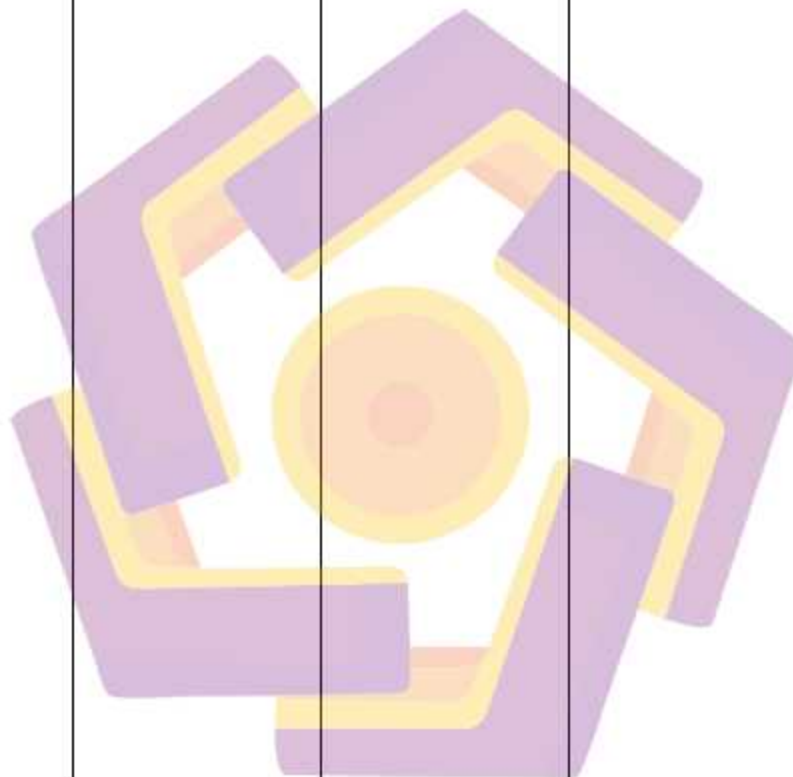
Dataset Baru, Penelitian Sebelumnya, tidak memiliki dataset yang dikhususkan untuk analisis keluhan transportasi dengan tingkat



kehalusan ini. **Penelitian Ini**, Membuat dataset pertama yang berisi tweet transportasi di India yang telah diberi label secara spesifik berdasarkan jenis keluhan.

Pendekatan Teknologi, Penelitian Sebelumnya, Menggunakan model pembelajaran mesin tradisional seperti Random Forest dan SVM (contoh: Das, 2021).

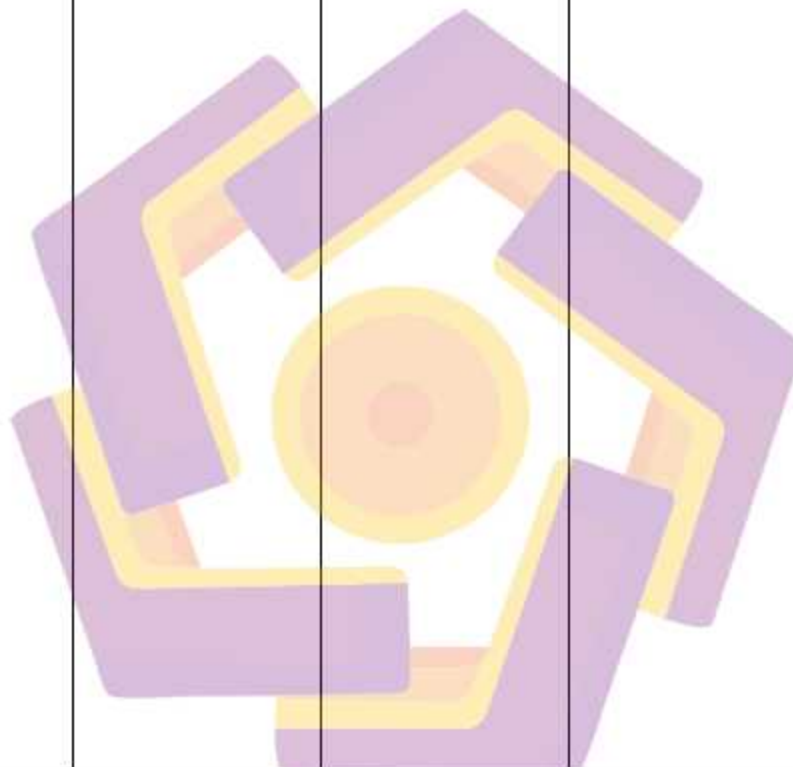
Penelitian Ini, Memanfaatkan model transformer seperti BERT, RoBERTa, dan



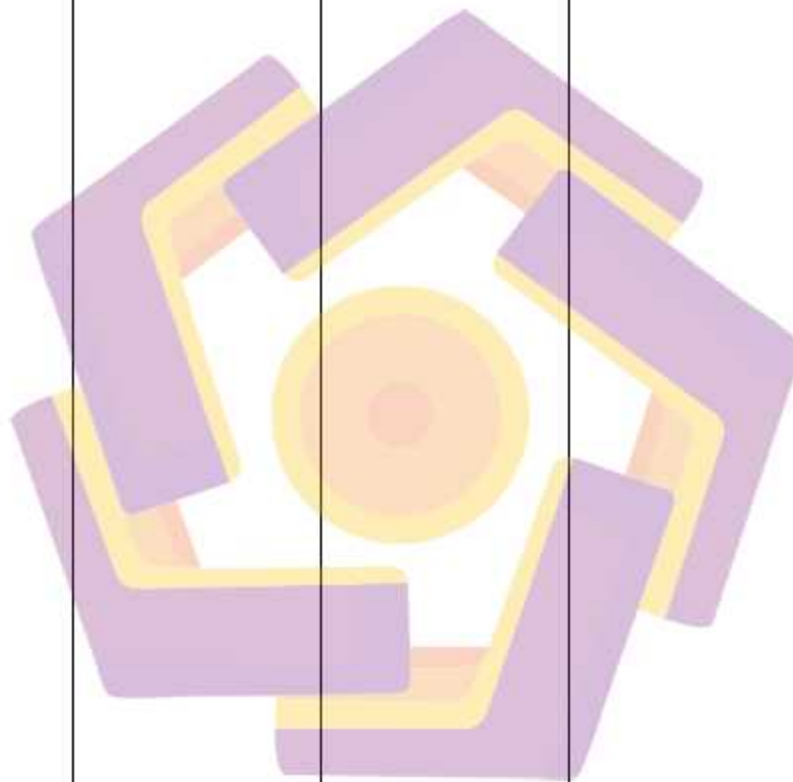
CrisisTransformer untuk klasifikasi sentimen dan keluhan secara lebih akurat. Bahkan, model GPT-3.5 digunakan sebagai pembanding nol-shot.

Output untuk Kebijakan Publik, Penelitian Sebelumnya, Hanya memberikan wawasan umum terkait sentimen dan masalah transportasi. **Penelitian Ini,** Memberikan rekomendasi yang lebih detail untuk perencanaan transportasi berdasarkan analisis kota-demi-kota dan kategori keluhan.

3	User-Specific Loyalty Measure and Prediction Using Deep Neural Network From Twitter Data	<p>Peneliti: Siddhaling Urolagin dan Saifali Patel</p> <p>Media Publikasi: IEEE Transactions on Computational Social Systems</p> <p>Tahun: February 2024</p>	<p>Mengembangkan metrik loyalitas dan kepuasan pengguna pada level individu menggunakan data Twitter, serta mengukur pengaruh antar pengguna dalam media sosial. Penelitian ini juga bertujuan untuk memprediksi loyalitas pengguna menggunakan model Deep Neural Network (DNN) serta membandingkan kinerjanya dengan model lain seperti Random Forest, XGBoost, dan SVC</p>	<p>Penelitian ini berhasil mengembangkan metrik baru, seperti User Sentiment Score (USS) dan Social Data Influence Score (SDIS), yang dapat mengukur loyalitas dan kepuasan pelanggan serta pengaruhnya terhadap pengguna baru. Model DNN menghasilkan akurasi klasifikasi tertinggi (98,62%) dibandingkan model lain dalam memprediksi loyalitas pengguna</p>	<p>Studi ini terbatas pada platform Twitter sebagai sumber data sosial. Saran untuk penelitian selanjutnya adalah memperluas penggunaan data dari media sosial lain, seperti Facebook atau Reddit, untuk analisis yang lebih komprehensif</p>	<p>Penelitian ini memperkenalkan metrik baru seperti User Sentiment Score (USS), Social Data Influence Score (SDIS), Overall Satisfaction Score (OSATS), dan lainnya untuk mengukur loyalitas pelanggan berdasarkan data Twitter.</p> <p>Net Promoter Score (NPS), Kekurangan penelitian sebelumnya mengandalkan survei dan tidak menyediakan informasi pada tingkat individu. Penelitian ini menggunakan data media</p>
---	--	--	--	--	---	---



sosial untuk menciptakan metrik tingkat pengguna (user-level) yang lebih mendalam, seperti USS. **Net Brand Reputation (NBR)** dan **Net Sentiment Score (NSS)**, Kekurangan penelitian sebelumnya Fokus pada komunitas pengguna secara keseluruhan tanpa mempertimbangkan pengaruh antar pengguna (user-to-user influence). Penelitian ini menghitung pengaruh pengguna terhadap pengguna lain menggunakan faktor seperti retweet dan jumlah pengikut, sesuatu yang tidak dilakukan oleh



NBR dan NSS. Pendekatan Sentimen Klasifikasi,

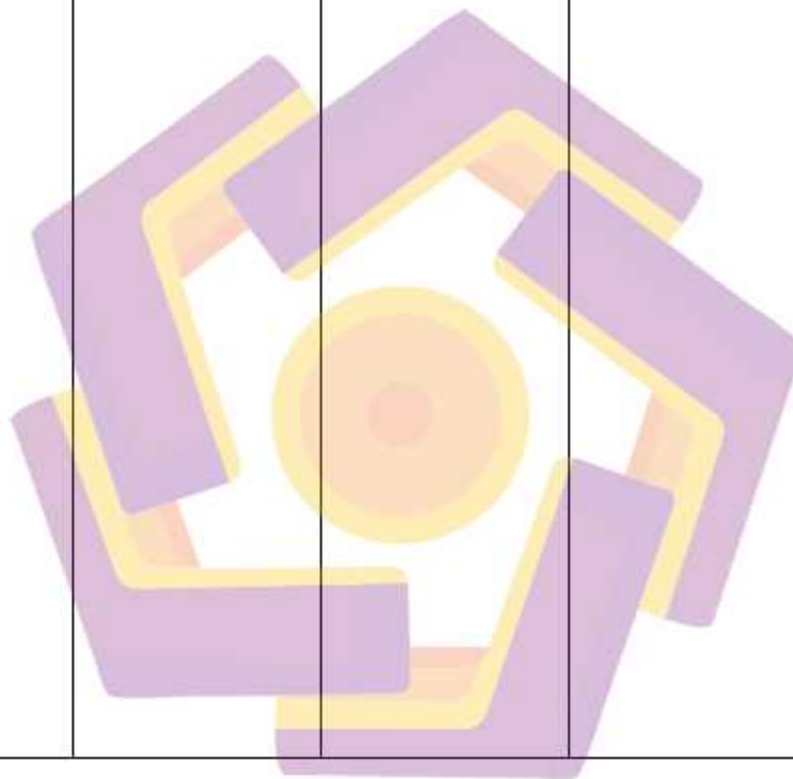
Penelitian sebelumnya hanya mengklasifikasikan sentimen tanpa menciptakan metrik yang mengukur pengaruh pengguna pada tingkat individu. Penelitian ini memanfaatkan skor sentimen tertimbang (Weighted Sentiment Score) dan VADER untuk analisis sentimen yang lebih kontekstual.

Pendekatan Machine Learning pada Loyalttas Pelanggan,

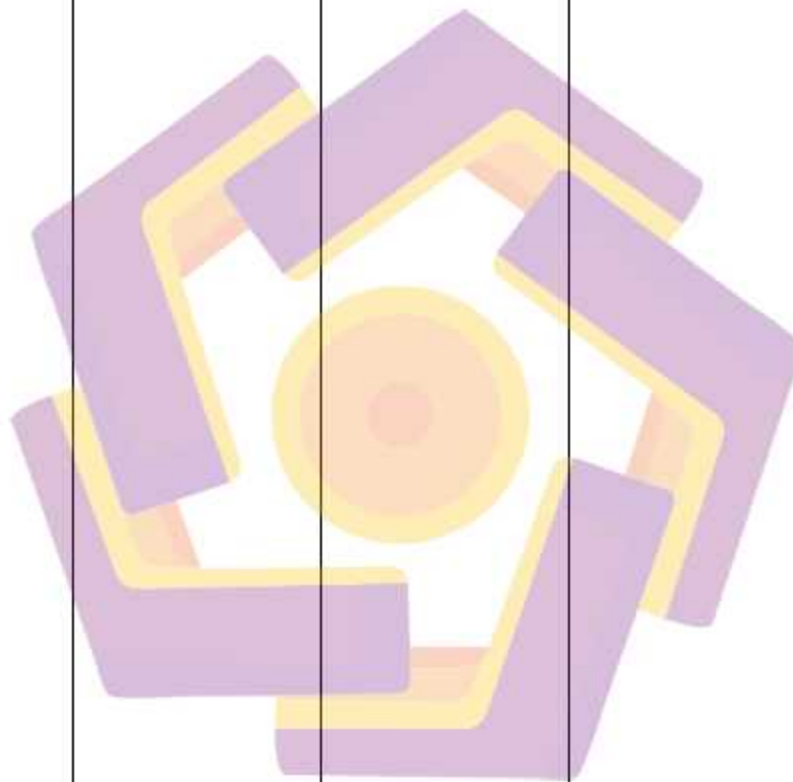
Penelitian sebelumnya fokus pada analisis data survei atau

						transaksi tanpa mempertimbangkan data sosial media secara real-time. Penelitian ini menggunakan deep neural network (DNN) untuk memprediksi loyalitas pelanggan dengan akurasi 98.62%, lebih tinggi dibanding metode lainnya seperti SVM atau random forest.
4	Sentiment analysis using Twitter data: a comparative application of	Peneliti: Yuxing Qi, Zahratu Shabrina Media Publikasi: Social Network	Penelitian ini bertujuan untuk mengeksplorasi sentimen publik terhadap Covid-19 selama lockdown ketiga di	Hasil dari pendekatan berbasis leksikon menunjukkan sentimen positif yang lebih tinggi dibandingkan negatif,	Keterbatasan penelitian meliputi jumlah data yang terbatas pada Inggris dan hanya melalui platform Twitter,	Topik Penelitian. Penelitian ini berfokus pada analisis sentimen menggunakan data Twitter dengan pendekatan berbasis

<p>lexicon- and machine-learning-based approach</p>	<p>Analysis and Mining, Springer Tahun: 2023</p>	<p>Inggris menggunakan data Twitter, dengan pendekatan analisis berbasis leksikon dan machine learning untuk membandingkan hasil dari kedua metode ini.</p>	<p>meskipun proporsi sentimen negatif meningkat setelah tahap kedua lockdown. Pada pendekatan machine learning, model Support Vector Classification (SVC) dengan metode BoW atau TF-IDF menunjukkan performa terbaik dengan akurasi hingga 71%. Penelitian ini juga menemukan bahwa metode Word2Vec kurang efektif karena keterbatasan data pelatihan.</p>	<p>sehingga hasil mungkin tidak sepenuhnya representatif. Untuk meningkatkan akurasi, disarankan untuk menggunakan data yang lebih besar, lebih banyak platform media sosial, dan waktu yang lebih panjang. Penggunaan deep learning atau metode anotasi manual yang lebih luas juga direkomendasikan.</p>	<p>leksikon dan pembelajaran mesin untuk mengamati sentimen terhadap COVID-19 selama lockdown ketiga di Inggris. Penelitian sebelumnya lebih umum mengkaji berbagai pendekatan analisis sentimen, sering kali terbatas pada bahasa Inggris formal, atau berfokus pada aspek tertentu, seperti perubahan sentimen terhadap pembelajaran online selama pandemi (Mostafa, 2021). Metode yang Digunakan, Penelitian ini membandingkan tiga pendekatan utama:</p>
---	--	---	--	--	---



TextBlob, VADER, dan SentiWordNet untuk metode berbasis leksikon, serta algoritma pembelajaran mesin (Random Forest, Multinomial Naïve Bayes, SVC) dengan berbagai representasi fitur seperti BoW, TF-IDF, dan Word2Vec. **Penelitian sebelumnya** biasanya hanya mengevaluasi satu pendekatan, misalnya pendekatan berbasis leksikon tanpa perbandingan langsung dengan model pembelajaran mesin (Boon-Itt & Skunkan, 2020).



Kontribusi Unik,

Penelitian Ini memberikan analisis perubahan sentimen di berbagai kota selama tiga tahap lockdown, menyertakan data geografis (geo-tagging) yang jarang digunakan sebelumnya.

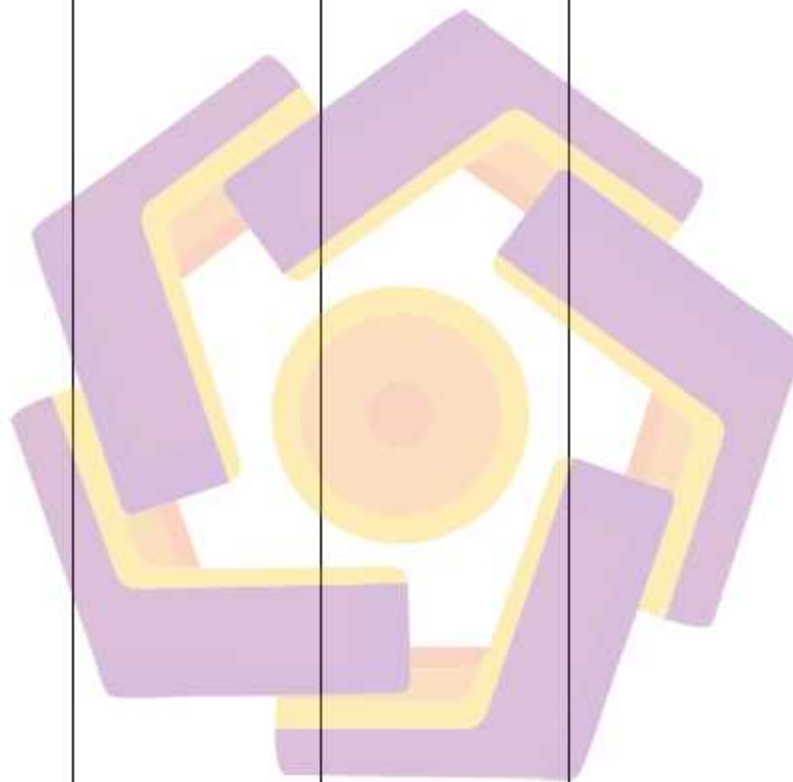
Penelitian sebelumnya, misalnya oleh Alamoodi et al. (2021), lebih banyak mengevaluasi tren sentimen global tanpa fokus lokal atau geo-tagging.

Keterbatasan yang Diatasi,

Penelitian Ini mengatasi bias leksikon dengan kombinasi metode dan labeling manual untuk

						meningkatkan akurasi pembelajaran mesin. Penelitian sebelumnya sering kali terbatas pada dataset atau pendekatan spesifik, seperti dominasi bahasa formal atau analisis menggunakan teknik Word2Vec tanpa evaluasi komprehensif.
5	Sentiment Analysis to Find Out Positive or Negative Opinions on Ride Hailing Application	Peneliti: Ain Nadia Safira, Eli Pujastuti, Hanafi, Bayu Setiaji, Donni Prabowo, Nuri Cahyono	Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap layanan aplikasi transportasi online, khususnya layanan dari aplikasi Gojek,	Penerapan metode SVM dalam analisis sentimen terhadap tweet terkait layanan Gojek berhasil mencapai akurasi 80.3% dalam mengklasifikasikan	Penelitian ini menyebutkan bahwa kualitas data dan kompleksitas dalam pengolahan data sentimen dari teks berkontribusi pada	Metode yang Digunakan, Penelitian ini menggunakan metode Support Vector Machine (SVM) untuk klasifikasi sentimen. Penelitian sebelumnya membandingkan SVM

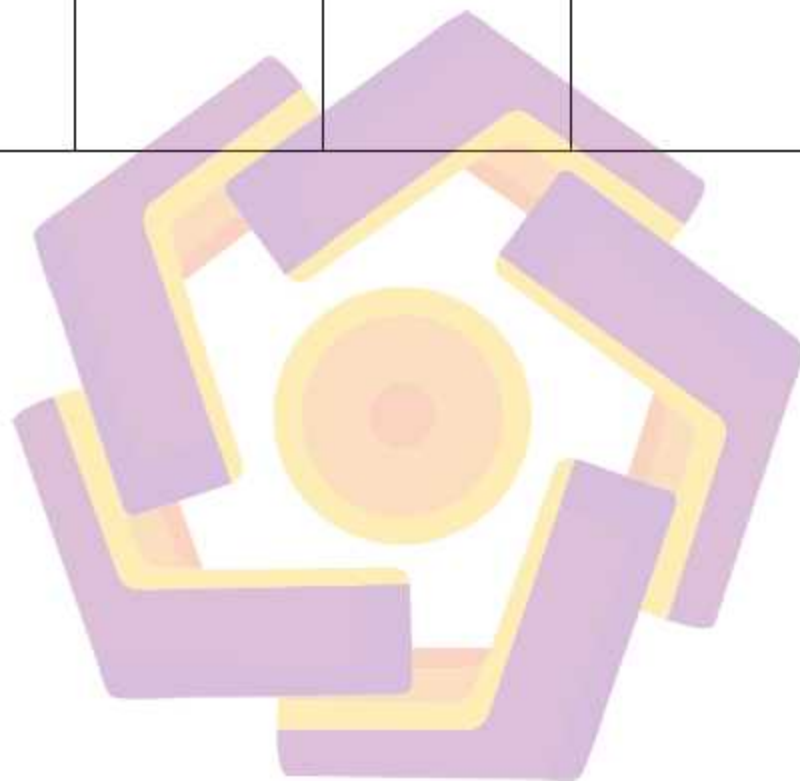
<p>Media Publikasi: 2023 International Conference on Informatics, Multimedia, Cyber and Information Systems (ICIMCIS) Tahun: 2023</p>	<p>menggunakan metode Support Vector Machine (SVM) dalam klasifikasi data sentimen dari media sosial Twitter.</p>	<p>sentimen positif atau negatif. Hal ini menunjukkan bahwa model ini mampu mengklasifikasikan dengan baik sentimen yang terkandung dalam tweet terkait Gojek.</p>	<p>tantangan yang dihadapi. Akan tetapi, saran eksplisit untuk pengembangan lebih lanjut atau kelemahan rinci tidak dijelaskan dalam dokumen.</p>	<p>dengan Naïve Bayes dan metode lainnya seperti Random Forest dan K-NN. SVM sering menunjukkan hasil akurasi yang lebih baik dibandingkan metode lain, seperti yang diungkapkan oleh Kurniawan et al. (2023) dan Kim (2014)</p> <p>Hasil yang Dicapai, Akurasi model SVM pada penelitian ini mencapai 80.3%, lebih tinggi dibandingkan dengan beberapa penelitian sebelumnya yang menggunakan teknik berbeda (misalnya, 78% pada penelitian oleh Kim)</p>
---	---	--	---	--



Fokus Studi, Penelitian ini difokuskan pada layanan ride-hailing, khususnya aplikasi Gojek, menggunakan data dari Twitter. Penelitian sebelumnya memiliki fokus beragam, seperti analisis sentimen pada merek produk, marketplace, dan isu sosial lainnya seperti minyak goreng.

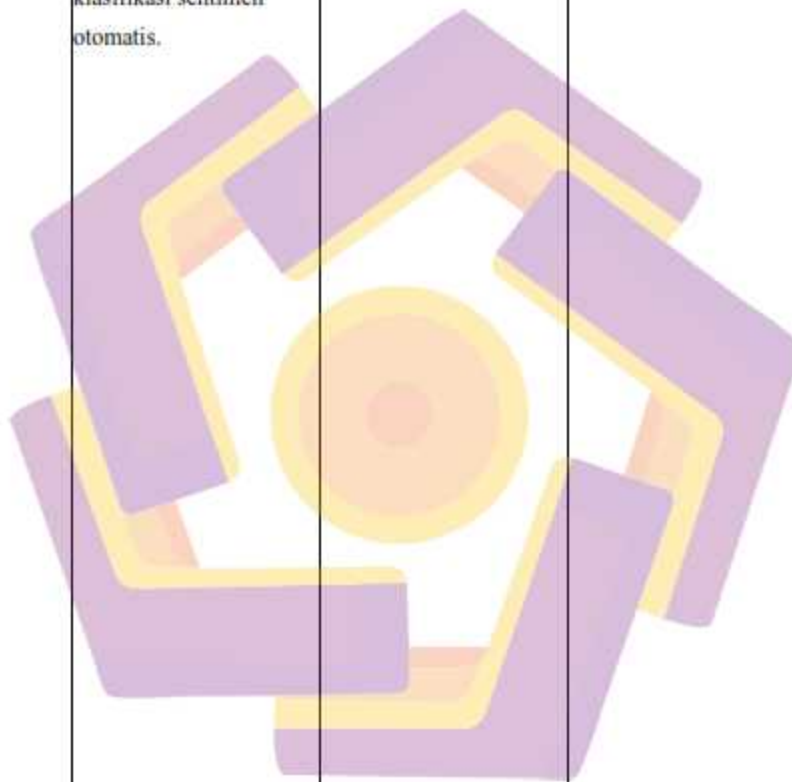
Pendekatan dan Proses, Penelitian ini menekankan pada tahap pre-processing data seperti stemming, tokenization, dan case folding, yang sebelumnya juga diidentifikasi sebagai

						langkah penting dalam beberapa penelitian terdahulu (misalnya, Balahur, 2012).
--	--	--	--	--	--	--



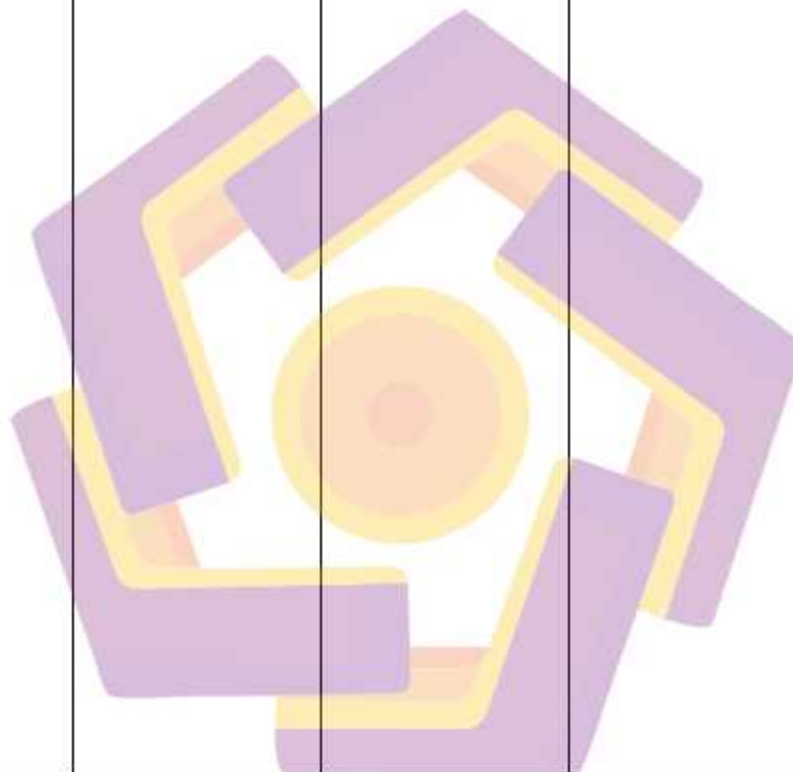
6	Automated Classification of Societal Sentiments on Twitter With Machine Learning	<p>Peneliti: Piyush Vyas, Martin Reisslein, Bhaskar Prasad Rimal, Gitika Vyas, Ganga Prasad Basyal, Prathamesh Muzumdar</p> <p>Media Publikasi: IEEE Transactions on Technology and Society</p> <p>Tahun: 2022</p>	<p>Tujuan penelitian ini adalah mengembangkan kerangka kerja otomatis yang menggabungkan analisis sentimen berbasis leksikon dan teknik pembelajaran mesin (ML) untuk mengklasifikasikan sentimen publik terhadap COVID-19 berdasarkan tweet di Twitter.</p> <p>Penelitian ini bertujuan untuk membantu memahami sentimen masyarakat dalam situasi penting, seperti pandemi COVID-19, melalui</p>	<p>Kerangka kerja hybrid yang dikembangkan berhasil mengklasifikasikan sentimen tweet terkait COVID-19 dengan akurasi 83% menggunakan model LSTM. Kerangka ini mampu mengklasifikasikan sentimen positif, negatif, dan netral dengan tingkat akurasi yang tinggi dan memiliki potensi untuk digunakan dalam pemantauan sentimen sosial secara otomatis selama peristiwa krisis.</p>	<p>Penelitian ini menyarankan peningkatan dalam penalaan hyperparameter dan penggunaan model pretrained seperti BERT untuk meningkatkan akurasi klasifikasi tweet.</p> <p>Penelitian juga menunjukkan perlunya analisis lebih lanjut terhadap sentimen emosional yang lebih luas, termasuk komponen emosi seperti takut, marah, dan senang, menggunakan leksikon emosi yang lebih luas.</p>	<p>Pendekatan yang digunakan, Penelitian sebelumnya cenderung menggunakan pendekatan berbasis lexicon atau supervised machine learning secara terpisah. Contohnya, [10] menggunakan VADER untuk analisis sentimen tanpa kombinasi dengan teknik pembelajaran mesin lain. Dalam penelitian ini, dikembangkan kerangka hybrid yang menggabungkan teknik lexicon-based (VADER) untuk analisis dan labeling tweet dengan teknik supervised machine learning</p>
---	--	--	---	---	---	---

klasifikasi sentimen otomatis.



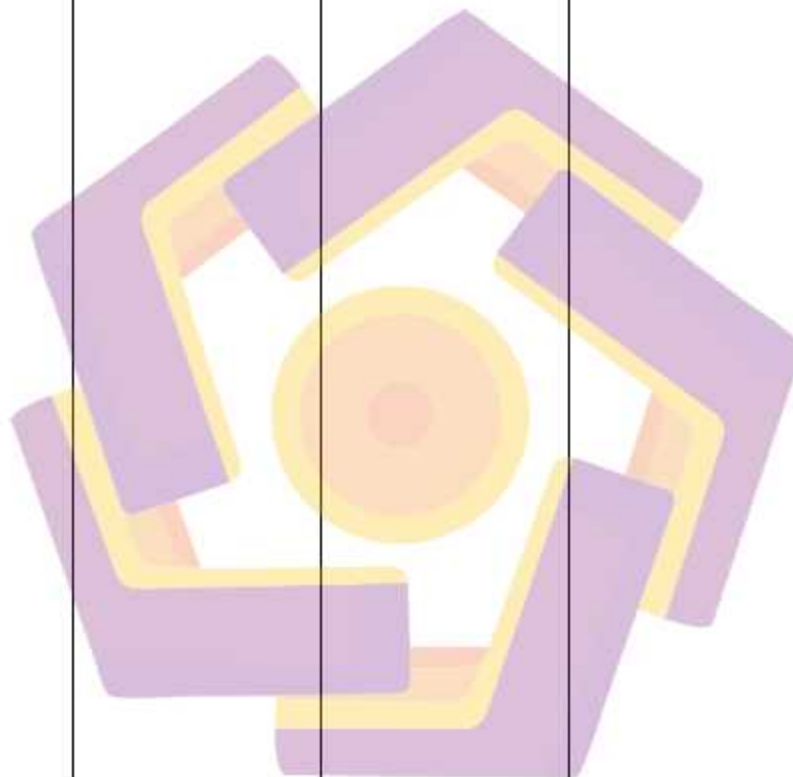
(termasuk LSTM) untuk klasifikasi sentimen, menciptakan proses otomatis tanpa bias manual dalam pelabelan. **Cakupan dataset**, Beberapa penelitian sebelumnya, seperti [10], hanya fokus pada demografi tertentu (Australia) atau [14] yang fokus pada platform Weibo (China). Penelitian ini memanfaatkan dataset yang lebih luas dengan tweet terkait COVID-19 dari seluruh dunia.

Klasifikasi sentimen, Penelitian [41] menggunakan klasifikasi biner (positif dan negatif)



untuk data sentimen, sedangkan penelitian ini melakukan klasifikasi multikelas (positif, netral, dan negatif), yang memberikan pemahaman lebih mendalam tentang distribusi sentimen masyarakat.

Evaluasi model, Model dalam penelitian ini menggunakan evaluasi komprehensif berbasis akurasi, precision, recall, dan F1-score dengan hasil terbaik pada LSTM (akurasi 83%). Penelitian sebelumnya lebih banyak bergantung

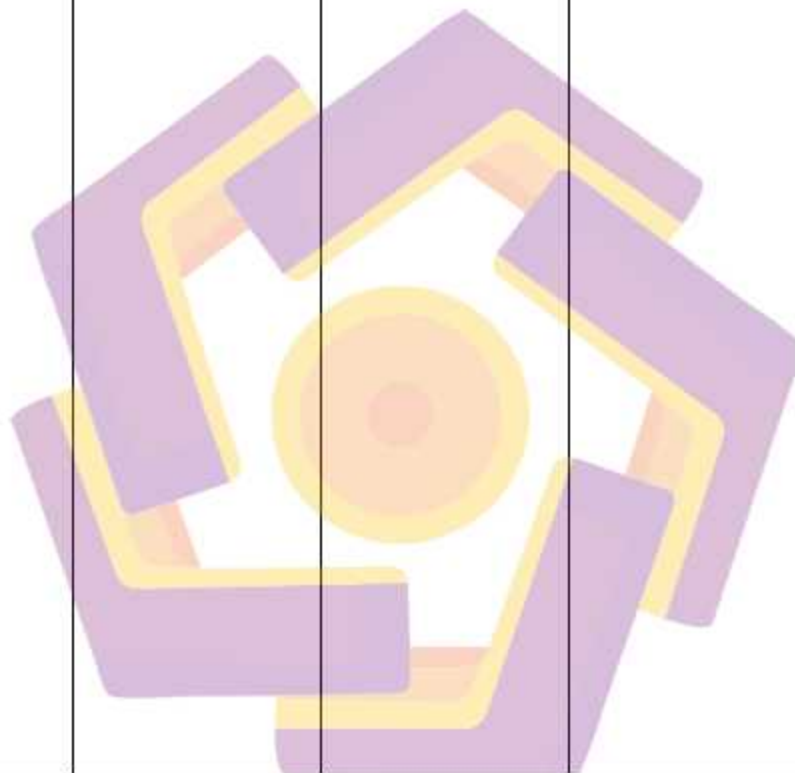


pada satu atau dua metrik evaluasi saja.

Kecepatan pemrosesan,

Teknik hybrid yang diusulkan mempercepat klasifikasi dengan menggunakan ML dibandingkan VADER saja, meningkatkan throughput hingga 10 kali lipat, memungkinkan analisis volume besar dalam waktu singkat.

Kontribusi baru, Penelitian ini mengintegrasikan lexicon-based labeling dengan teknik ML secara unik untuk membangun



sistem otomatis. Ini belum ditemukan pada penelitian sebelumnya.

BAB III METODE PENELITIAN

3.1 Jenis, Sifat Dan Pendekatan Penelitian.

Penelitian ini termasuk dalam jenis penelitian eksperimen, di mana data dikumpulkan dari media sosial Twitter dengan menggunakan kata kunci "Gojek, Grab, Maxim, dan InDrive." Data yang diperoleh berjumlah sekitar 10.000 tweet berbahasa Indonesia dari tahun 2023. Data ini kemudian digunakan untuk membandingkan tingkat akurasi dari dua metode machine learning, yaitu Deep Neural Network (DNN) dan Long Short-Term Memory (LSTM). Eksperimen dilakukan untuk mengetahui performa terbaik antara kedua metode tersebut berdasarkan dataset yang sama.

Dari segi sifat penelitian, penelitian ini bersifat deskriptif, karena bertujuan untuk menggambarkan dan menganalisis perbandingan performa antara algoritma DNN dan LSTM. Parameter evaluasi yang digunakan dalam penelitian ini mencakup akurasi, presisi, recall, F1-score, waktu pelatihan, dan mean absolute error (MAE). Selain itu, penelitian ini juga ingin mengetahui sejauh mana proses pre-processing data dan jumlah dataset yang digunakan dalam validasi ulang (cross-validation sebanyak lima kali) berpengaruh terhadap kualitas hasil identifikasi sentimen dari masing-masing algoritma.

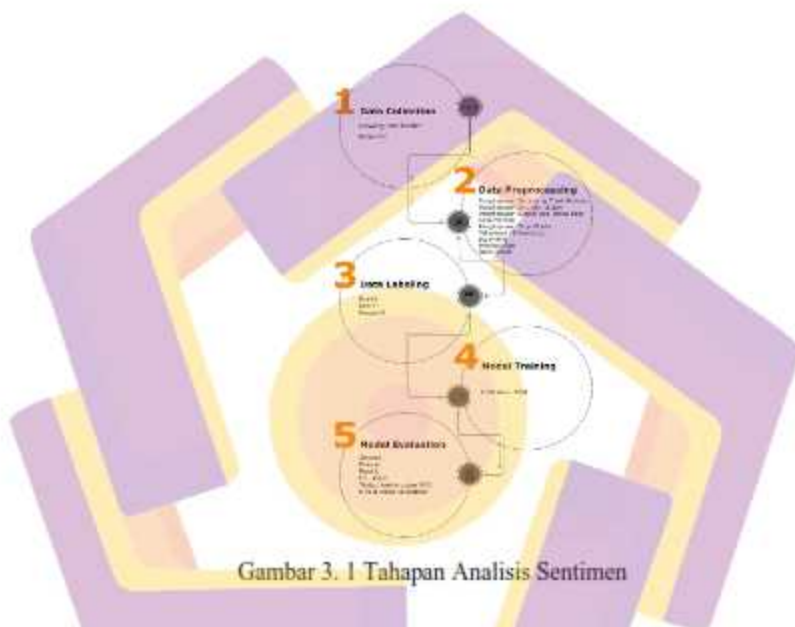
Adapun pendekatan penelitian yang digunakan adalah pendekatan kuantitatif. Hasil penelitian dianalisis secara numerik melalui pengolahan data eksperimen dan disajikan dalam bentuk angka, Tabel, dan diagram. Proses pengumpulan data dilakukan melalui eksperimen langsung yang melibatkan perbandingan performa kedua algoritma. Analisis dilakukan dengan membandingkan nilai-nilai metrik evaluasi untuk menentukan metode mana yang paling optimal dalam melakukan analisis sentimen terhadap data Twitter yang telah dikumpulkan.

3.2 Metode Pengumpulan Data

Pada tahap ini, peneliti akan mengumpulkan data yang dibutuhkan untuk mengerjakan penelitian ini. Data tersebut berupa file excel dari proses **crawling** pada data Twitter menggunakan Tweepy Pustaka python. Informasi yang

dibutuhkan baik untuk membantu pengolahan data dan pemecahan masalah pada penelitian ini akan dikumpulkan pada langkah ini. Data yang digunakan adalah data dari Twitter dengan kata kunci “Gojek, Grab, Maxim, dan InDrive” id lang = Indonesia waktu bulan 01 Januari 2023 sampai dengan 31 Desember 2023.

3.3 Metode Analisis Data



Gambar 3. 1 Tahapan Analisis Sentimen

Sebagaimana ditunjukkan pada Gambar 3.1, data yang telah didapat dari tahap sebelumnya subbab 3.2 yang masih merupakan data mentah seperti terlihat pada Tabel 3.1, akan diolah dengan tahap selanjutnya yaitu Pre-Processing. Langkah-langkah Pre-Processing yang akan dilakukan dalam penelitian ini, dipaparkan sebagai berikut : **Penghapusan Data yang Tidak Relevan**, Menghapus tweet yang tidak mengandung sentimen atau mengandung lebih dari satu sentimen (ambigu) dan menghapus simbol retweet ("RT") untuk memastikan hanya tweet asli yang dianalisis. **Penghapusan URL dan Tautan**, Semua URL atau tautan dalam tweet dihapus karena tidak memberikan nilai dalam analisis sentimen.

Penghapusan Simbol dan Tanda Baca, Menghapus simbol-simbol seperti ~!@#S%^&*()_+={ }[]|;""<>., untuk menghindari noise dalam data. **Case Folding**, Semua teks diubah menjadi huruf kecil (non-kapital) untuk menyeragamkan data dan menghindari perbedaan akibat sensitivitas huruf kapital. **Penghapusan Stop Words**, Menghapus kata-kata umum yang tidak memberikan kontribusi besar terhadap analisis sentimen, seperti "dan", "di", "yang", dsb. **Tokenisasi (Tokenizing)**, Memecah teks menjadi kata per kata untuk mempermudah analisis lebih lanjut, seperti penghitungan frekuensi kata. **Stemming** mengembalikan kata-kata ke bentuk dasar (akar kata), misalnya "berjalan" menjadi "jalan", untuk menyederhanakan analisis, pada penelitian ini stemming menggunakan Pustaka **Sastrawi**. **Pembersihan Data Ganda**, Menghapus duplikasi atau entri yang sama agar dataset hanya berisi data unik.

Tabel 3.1 Tweet Mentah

No	Tweet Mentah
1	RT @user123: Pelayanan driver Grab sangat memuaskan, ramah, dan tepat waktu! 🍻 🍻 https://t.co/example
2	Driver Gojek terlambat datang, perjalanan jadi terganggu! 🚫 #Gojek #TransportasiOnline https://gojek.com/error
3	Indrive memberikan promo yg menarik untuk perjalanan jarak jauh, sangat membantu 🍻 !! RT @IndrivePromo: 🍻 🍻
4	Tarif Maxim naik terus, jd ga worth it utk jarak dekat! #MaximTransport
5	Layanan transportasi online cukup memadai, tapi msh ada yg bisa diperbaiki. 🍻
6	Grab Food datang lebih cepat dr estimasi, makanannya msh hangat, mantap! https://grab.com/food/promo
7	Kesulitan menemukan driver Gojek! Aplikasi error terus, mengecewakan! 🍻 #GojekDown https://t.co/err404
8	Maxim punya fitur lumayan bagus, tp sering bgt kendala pembayaran #TransportasiOnline 🍻
9	Promo Indrive bantu bgt kondisi dompet yg lagi tipis! Thx Indrive 🍻 🍻 https://indrive.com/promo
10	Driver ga ramah dan perjalanan sgt tdk menyenangkan. Ini parah banget! RT @usercomplain: 🍻

Setelah melalui tahap pre-processing, langkah berikutnya adalah pelabelan yang bertujuan untuk mengategorikan tweet ke dalam tiga kelas sentimen, yaitu positif, netral, dan negatif. Proses pelabelan dilakukan dengan menggunakan pendekatan automated labeling berbasis IndoBERTweet yang diimplementasikan melalui pustaka Python Hugging Face Transformers. Pelabelan otomatis diperlukan karena skala data yang besar membuat anotasi manual tidak realistis dari sisi waktu dan biaya. Kualitas label tetap dijaga melalui validasi sampel dan pemeriksaan reliabilitas, sehingga cukup andal untuk pelatihan serta evaluasi model.

Untuk memastikan kualitas data, hasil pelabelan otomatis akan dilakukan pengecekan secara acak (Random Sampling) secara manual sebanyak 200 tweet. Apabila ditemukan data yang ambigu atau tidak konsisten, data tersebut akan dihapus dari dataset. Data yang telah dilabel kemudian diberikan kolom tambahan berupa label sentimen (lihat Tabel 3.2), dan selanjutnya disimpan dalam format CSV.

Penelitian ini menempatkan IndoBERTweet semata sebagai alat bantu anotasi otomatis (automatic labeler) untuk menghasilkan data berlabel, bukan sebagai model utama yang dievaluasi. Meskipun IndoBERT terbukti kuat pada berbagai tugas NLP bahasa Indonesia, performanya dapat menurun pada analisis sentimen teks pendek dan informal seperti Twitter: korpus pralatih dominan bersumber dari teks formal; ukuran model yang besar menuntut sumber daya komputasi tinggi dan berisiko overfitting ketika data terbatas; serta pemodelan urutan yang tidak rekuren berpotensi kurang menangkap dinamika temporal yang relevan bagi penentuan polaritas. Berdasarkan pertimbangan tersebut, penelitian ini memfokuskan perbandingan pada arsitektur yang lebih ringan DNN dan LSTM yang peka terhadap urutan kata dan dapat dilatih secara efisien pada data terbatas yang telah disesuaikan dengan domain media sosial. Dengan demikian, peran IndoBERTweet berhenti pada penyediaan anotasi yang andal, sedangkan tujuan utama penelitian adalah mengevaluasi dan membandingkan kinerja DNN dan LSTM dalam analisis sentimen. Dengan demikian, peran IndoBERTweet terbatas pada tahap penyediaan data berlabel yang akurat, sedangkan tujuan utama penelitian adalah mengevaluasi dan membandingkan performa DNN dan LSTM dalam melakukan analisis sentimen.

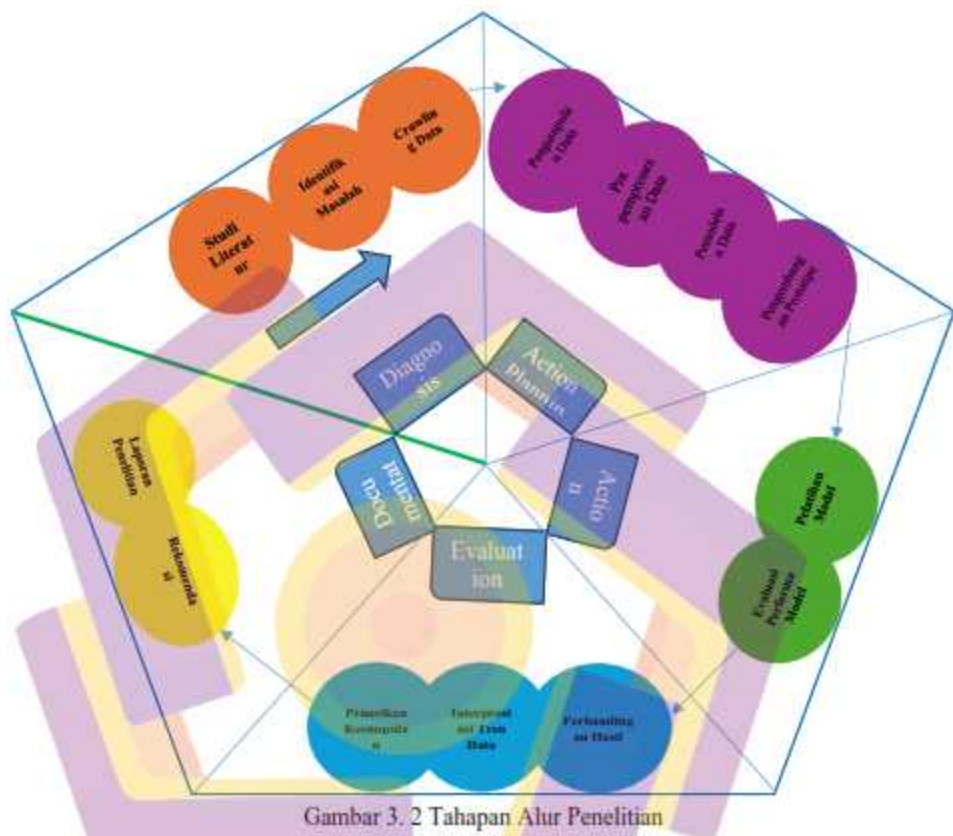
Tabel 3.2 Tweet Hasil Pelabelan

Tweet	Sentimen
"Pelayanan driver Grab sangat memuaskan, ramah, dan tepat waktu!"	Positif
"Driver Gojek terlambat datang, perjalanan jadi terganggu!"	Negatif
"Indrive memberikan promo yang menarik untuk perjalanan jarak jauh, sangat membantu."	Positif
"Tarif Maxim akhir-akhir ini naik terus, jadi tidak worth it untuk jarak dekat."	Negatif
"Layanan aplikasi transportasi online cukup memadai, tapi masih ada ruang untuk perbaikan."	Netral
"Grab Food datang lebih cepat dari estimasi, makanannya masih hangat, mantap!"	Positif
"Kesulitan menemukan driver Gojek, aplikasi terus error, sangat mengecewakan."	Negatif
"Maxim memiliki fitur yang lumayan bagus, tapi sering ada kendala kecil dalam pembayaran."	Netral
"Promo dari Indrive sangat membantu, apalagi saat kondisi dompet lagi tipis!"	Positif
"Driver tidak ramah dan perjalanan sangat tidak menyenangkan, benar-benar mengecewakan."	Negatif

Selanjutnya Setelah data diberi label, data digunakan untuk melatih model machine learning DNN dan LSTM. Kedua Model ini akan belajar mengenali pola dalam teks untuk mengklasifikasikan sentimen dari data baru yang tidak diberi label.

Setelah seluruh proses training dijalankan, maka akan dilakukan proses analisis terhadap model yang dihasilkan. Model yang dihasilkan akan dievaluasi akurasi, presisi, recall, f1- score, waktu training dan MAE. Analisis tren kenaikan atau penurunan data pada performa algoritma DNN dan algoritma LSTM dengan pengujian terhadap sekumpulan data uji. Proses ini dilakukan untuk mengukur sejauh mana model mampu klasifikasi teks dengan tepat. Untuk evaluasi ini langkah terperinci dapat dilihat pada bab 4.

3.4 Alur Penelitian



Gambar 3.2 Tahapan Alur Penelitian

Sebagaimana ditunjukkan pada Gambar 3.2, secara garis besar alur penelitian yang digunakan dalam penelitian ini adalah sebagai berikut :

Diagnosis. Pada tahapan ini peneliti melakukan studi pustaka dengan membaca jurnal, buku, makalah dan laporan penelitian lainnya yang terkait dengan objek yang akan diteliti Selanjutnya melakukan studi lapangan dengan melakukan crawling data melalui API Twitter untuk mendapatkan data primer

Action Planning. Pada tahapan ini peneliti melakukan analisis data, setelah data didapatkan lalu dilakukan verifikasi data dalam hal ini ada beberapa tahapan pre-processing untuk mendapatkan nilai atau bobot dari setiap data. Selanjutnya

membuat pemodelan data dan algoritma ke dua metode dengan data yang tersedia dan membuat prototype sistem analisis sentimen dengan satu sistem satu metode, sehingga terdapat dua prototype.

Action Taking, Pada tahapan ini peneliti melakukan eksperimen pengujian tingkat performa melalui pengujian serta membandingkan performa kedua metode malui parameter pengujian akurasi presisi, recall, F1 score, waktu training dan MAE serta K-Fold Cross-Validation untuk mendapatkan nilai performa terbaik antara kedua algoritma.

Evaluation, Pada tahapan yang terakhir, peneliti melakukan evaluasi hasil perbandingan dua metode yang digunakan yaitu DNN dan LSTM. Selanjutnya dihitung hasil yang sesuai untuk mendapatkan persentase performa pada masing-masing metode. Tahap terakhir diperoleh kesimpulan, metode mana yang paling baik dalam melakukan analisis sentimen data twitter melalui kata kunci "Gojek, Grab, Maxim dan InDrive".

Menyusun laporan akhir penelitian yang mencakup hasil temuan, kesimpulan, serta rekomendasi untuk penelitian selanjutnya..



BAB IV
HASIL PENELITIAN DAN PEMBAHASAN

4.1 Pengumpulan Data

Total data ulasan yang diperoleh melalui proses crawling data Twitter, terdapat 10.000 twitt. Berikut contoh data yang berhasil di angkat menggunakan metode crawling sebagaimana ditunjukkan pada Tabel 4.1.

Tabel 4. 1 Hasil *Crawling*

<i>Create_at</i>	<i>Full_text</i>	<i>Username</i>
Mon Jan 30 23:47:14 +0000 2023	@GrabID Makasih kak respon nya tadi barusan udah dibatalin kok. Aku udah hubungi cs lewat grab nya makasih ya	
Mon Feb 27 23:57:07 +0000 2023	hoki pagi pagi kalo dapetnya gini mah tiap hari aja naik gojek.	
Thu Mar 30 23:48:29 +0000 2023	TERLAMPAU MURAHHH NI . Rugi kalau tak GRAB HARGA RAHMAH . Pelbagai design dan warna yang menarik untuk anda !• @E• https://t.co/68oaHJ4wxh https://t.co/QAbQD5hOSQ	
Sat Apr 29 23:58:17 +0000 2023	@jogmfis Waktu liburan akhir tahun kemarin mau naik maxim dari teras Malioboro sampe st tugu karena hujan ditolak terus. Sampe pada akhirnya di telp salah satu driver ditawarin harganya 100k mau ga? Langsung shock	
Tue May 30 23:58:35 +0000 2023	ya Allah pagi pagi dapet grab selalu yg jelet	

4.2 Pre-processing Data

Selanjutnya dilakukan penghapusan data tidak relevan, seperti tweet tanpa sentimen atau ambigu, serta penghapusan URL dan tautan yang tidak diperlukan. Selain itu, simbol dan tanda baca seperti `!@#S%^&()*` juga dihapus agar data lebih bersih, sebagaimana ditunjukkan pada table 4.2.

Tabel 4. 2 Hasil penghapusan data tidak relevan

<i>Create_at</i>	<i>Full_text</i>	<i>Username</i>
Mon Jan 30 23:47:14 +0000 2023	GrabID Makasih kak respon nya tadi barusan udah dibatalin kok. Aku udah hubungi cs lewat grab nya makasih ya	
Mon Feb 27 23:57:07 +0000 2023	hoki pagi pagi kalo dapetnya gini mah tiap hari aja naik gojek.	
Thu Mar 30 23:48:29 +0000 2023	TERLAMPAU MURAHHH NI. Rugi kalau tak GRAB HARGA RAHMAH. Pelbagai design dan warna yang menarik untuk anda	
Sat Apr 29 23:58:17 +0000 2023	jogmfs Waktu liburan akhir tahun kemarin mau naik maxim dari teras Malioboro sampe st tugu karena hujan ditolak terus. Sampe pada akhirnya di telp salah satu driver ditawarkan harganya 100k mau ga? Langsung shock	
Tue May 30 23:58:35 +0000 2023	ya Allah pagi pagi dapet grab selalu yg lelet	

Tahapan selanjutnya adalah Case folding diterapkan untuk mengubah semua teks menjadi huruf kecil. Seperti terlihat pada Tabel 4.3.

Tabel 4. 3 Hasil *Case folding*

<i>create_at</i>	<i>full_text</i>	<i>username</i>
mon jan 30 23:47:14 +0000 2023	grabid makasih kak respon nya tadi barusan udah dibatalin kok. aku udah hubungi es lewat grab nya makasih ya	
mon feb 27 23:57:07 +0000 2023	hoki pagi pagi kalo dapetnya gini mah tiap hari aja naik gojek.	
thu mar 30 23:48:29 +0000 2023	terlampau murahhh ni. rugi kalau tak grab harga rihmah, pelbagai design dan warna yang menarik untuk anda	
sat apr 29 23:58:17 +0000 2023	jomngs waktu liburan akhir tahun kemaren mau naik maxim dari teras malioboro sampe st tugu karena hujan ditolak terus. sampe pada akhirnya di telp salah satu driver ditawarkan harganya 100k mau ga? langsung shock	
tue may 30 23:58:35 +0000 2023	ya allah pagi pagi dapet grab selalu yg lelet	

Setelahnya dilakukan stopword removal, digunakan untuk menghapus kata-kata umum yang tidak memiliki makna sentimen, seperti "dan", "di", dan "yang", sebagaimana dapat dilihat dalam Tabel 4.4.

Tabel 4. 4 Hasil *stopword removal*

<i>create_at</i>	<i>full_text</i>	<i>username</i>
mon jan 30 23:47:14 +0000 2023	makasih kak respon barusan dibatalin hubungi es grab makasih	
mon feb 27 23:57:07 +0000 2023	hoki pagi pagi dapetnya gini tiap hari naik gojek.	

thu mar 30 23:48:29 +0000 2023	terlampau murahhh ni rugi grab harga rahmah pelbagai design warna menarik	
sat apr 29 23:58:17 +0000 2023	liburan akhir tahun kemaren naik maxim teras malioboro sampe tugu hujan ditolak telp driver ditawarkan harga langsung shock	
tue may 30 23:58:35 +0000 2023	allah pagi pagi dapat grab lelet	

Berikutnya, teks **ditokenisasi** dengan memecahnya menjadi kata-kata individual, seperti ditunjukkan pada Tabel 4.5.

Tabel 4.5 Hasil *Tokenisasi*

<i>create_at</i>	<i>full_text</i>	<i>username</i>
mon jan 30 23:47:14 +0000 2023	"makasih", "kak", "respon", "barusan", "dibatalin", "hubungi", "cs", "grab", "makasih"	
mon feb 27 23:57:07 +0000 2023	"hoki", "pagi", "pagi", "dapatnya", "gini", "tiap", "hari", "naik", "gojek"	
thu mar 30 23:48:29 +0000 2023	"terlampau", "murahhh", "ni", "rugi", "grab", "harga", "rahmah", "pelbagai", "design", "warna", "menarik"	
sat apr 29 23:58:17 +0000 2023	["liburan", "akhir", "tahun", "kemaren", "naik", "maxim", "teras", "malioboro", "sampe", "tugu", "hujan", "ditolak", "telp", "driver", "ditawarin", "harga", "langsung", "shock"	
tue may 30 23:58:35 +0000 2023	"allah", "pagi", "pagi", "dapat", "grab", "lelet"	

Selanjutnya **stemming** dilakukan menggunakan pustaka Sastrawi untuk mengembalikan kata ke bentuk dasarnya, sebagaimana dapat dilihat dalam Tabel 4.6.

Tabel 4. 6 Hasil *Stemming*

<i>create_at</i>	<i>full_text</i>	<i>username</i>
mon jan 30 23:47:14 +0000 2023	"makasih", "kak", "respon", "baru", "batal", "hubung", "cs", "grab", "makasih"	
mon feb 27 23:57:07 +0000 2023	"hoki", "pagi", "pagi", "dapat", "gini", "tiap", "hari", "naik", "gojck"	
thu mar 30 23:48:29 +0000 2023	"lampau", "murah", "ni", "rugi", "grab", "harga", "rahmah", "bagai", "design", "warna", "tarik"	
sat apr 29 23:58:17 +0000 2023	"libur", "akhir", "tahun", "kemut", "naik", "maxim", "teras", "malioboro", "sampai", "tugu", "hujan", "tolak", "telp", "driver", "tawar", "harga", "langsung", "shock"	
tue may 30 23:58:35 +0000 2023	"allah", "pagi", "pagi", "dapat", "grab", "lelet"	

Hasil akhir dari proses preprocessing ditampilkan pada Tabel 4.7.

Tabel 4. 7 Hasil *Preprocessing Data*.

<i>Tweet Asli</i>	<i>Setelah Preprocessing</i>	<i>username</i>
@GrabID Makasih kak respon nya tadi barusan udah dibatalin kok. Aku udah hubungi es lewat grab nya makasih ya	makasih, kak, respon, baru, batal, hubung, cs, grab, makasih	
hoki pagi pagi kalo dapetnya gini mah tiap hari aja naik gojek.	hoki, pagi, pagi, dapat, gini, tiap, hari, naik, gojek	

TERLAMPAU MURAHHH NI . Rugi kalau tak GRAB HARGA RAHMAH . Pelbagai design dan warna yang menarik untuk anda • ä€• https://t.co/68oaHJ4wxh https://t.co/QAbQD5hOSQ	lampau, murah, ni, rugi, grab, harga, rahmah, bagai, design, warna, tarik	
@jogmfs Waktu liburan akhir tahun kemarin mau naik maxim dari teras Malioboro sampai tugu karena hujan ditolak terus. Sampai pada akhirnya di telp salah satu driver ditawarkan harganya 100k mau ga? Langsung shock	libur, akhir, tahun, kemarin, naik, maxim, teras, malioboro, sampai, tugu, hujan, tolak, telp, driver, tawar, harga, langsung, shock	
ya Allah pagi pagi dapat grab selalu yg lelet	allah, pagi, pagi, dapat, grab, lelet	

4.3 Pelabelan Data

Proses pelabelan data dilakukan dengan mengkategorikan sentimen tweet menjadi tiga kelas, yaitu positif, negatif, dan netral, secara otomatis menggunakan model IndoBERTweet dari pustaka Transformers milik Hugging Face, sebagaimana ditunjukkan pada Tabel 4.8. Namun karena hasil pelabelan otomatis menunjukkan distribusi yang tidak seimbang dengan proporsi kelas *netral* yang jauh lebih tinggi dibandingkan kelas positif dan negatif, maka perlu dilakukan penanganan terhadap ketidakseimbangan data (*class imbalance*). Pendekatan yang dapat diterapkan adalah Metode Safe Resample merupakan sebuah pendekatan resampling yang lebih *hati-hati* dibandingkan dengan Random Undersampling (RUS) atau oversampling konvensional. Jika RUS hanya sekadar menghapus sampel dari kelas mayoritas secara acak, dan oversampling biasa hanya menambahkan data pada kelas minoritas, Safe Resample berusaha menjaga keseimbangan data dengan tetap mempertahankan kualitas distribusi dataset. Dengan kata lain, metode ini bukan hanya memanipulasi jumlah data, tetapi juga memastikan agar representasi data tidak terdistorsi secara berlebihan.

Pendekatan ini memastikan bahwa oversampling tidak menghasilkan terlalu banyak duplikasi data yang dapat memicu masalah overfitting. Sebaliknya, saat

melakukan undersampling, Safe Resample berupaya untuk tidak menghapus terlalu banyak informasi penting dari kelas mayoritas. Selain itu, metode ini juga menekankan konsistensi, yakni menjaga agar proses resampling tetap dapat direproduksi (reproducible) dengan mengontrol nilai random seed. Dengan kontrol semacam ini, hasil eksperimen lebih dapat dipercaya karena stabil dan tidak mudah berubah hanya akibat variasi data acak. Dengan cara ini, model DNN dan LSTM tidak akan terlalu bias terhadap kelas mayoritas dan dapat mempelajari pola sentimen secara lebih optimal.

Tabel 4. 8 Hasil *Labeling*.

<i>Tweet</i>	<i>Kategori Sentimen</i>	<i>username</i>
makasih, kak, respon, baru, batal, hubungi, cs, grab, makasih	positif	
hoki, pagi, pagi, dapat, gini, tiap, hari, naik, gojek	positif	
lampau, murah, ni, rugi, grab, harga, rahmah, bagai, design, warna, Tarik	positif	
libur, akhir, tahun, kemar, naik, maxim, teras, malioboro, sampai, tugu, hujan, tolak, telp, driver, tawar, harga, langsung, shock	negatif	
allah, pagi, pagi, dapat, grab, lelet	negatif	

Setelah proses pelabelan selesai, langkah selanjutnya adalah menghapus variabel yang tidak diperlukan, sehingga hanya menyisakan dua variabel utama, yaitu content dan sentimen. Contoh hasil data setelah dilakukan penghapusan terhadap variabel yang tidak relevan ditampilkan pada Tabel 4.9.

Tabel 4. 9 Hasil Contoh Data Setelah Dilakukan Drop Variabel

<i>No</i>	<i>Tweet</i>	<i>Kategori Sentimen</i>

1	makasih, kak, respon, baru, batal, hubung, cs, grab, makasih	positif
2	hoki, pagi, pagi, dapat, gini, tiap, hari, naik, gojek	positif
3	lampau, murah, ni, rugi, grab, harga, rahmah, bagai, design, warna, Tarik	positif
4	libur, akhir, tahun, kemar, naik, maxim, teras, malioboro, sampai, tugu, hujan, tolak, telp, driver, tawar, harga, langsung, shock	negatif
5	allah, pagi, pagi, dapat, grab, lelet	negatif

Sebagaimana dijelaskan pada Bab III Untuk menilai reliabilitas hasil pelabelan otomatis yang dilakukan oleh model IndoBERT, dilakukan validasi terhadap 200 sampel data yang dipilih secara acak. Proses validasi dilakukan dengan membandingkan label otomatis dengan label manual yang diberikan oleh peneliti. Berdasarkan hasil perbandingan, sebanyak 185 dari 200 data memiliki kesesuaian label, dengan tingkat akurasi sebesar 92,5%. Nilai Cohen's Kappa sebesar 0,89 menunjukkan tingkat kesepakatan yang sangat tinggi antara pelabel otomatis dan pelabel manual seperti ditunjukkan pada tabel 4.10. Hal ini mengindikasikan bahwa hasil pelabelan otomatis cukup reliabel untuk digunakan sebagai dasar pelatihan model klasifikasi.

Selanjutnya, dilakukan analisis terhadap pengaruh koreksi label terhadap performa model DNN dan LSTM. Setelah dilakukan koreksi pada 200 data tersebut, model DNN mengalami peningkatan akurasi sebesar 0,8%, sedangkan model LSTM meningkat sebesar 0,9%. Dengan demikian, dapat disimpulkan bahwa error kecil dalam pelabelan otomatis tidak secara signifikan memengaruhi performa keseluruhan model, sehingga pendekatan semi-otomatis ini efisien dan valid untuk penelitian berbasis data berukuran besar.

Tabel 4. 10 Perbandingan Auto-Label dan Label Manual per Kelas (n=200)

Label	Jumlah Auto-Label	Jumlah Label Manual	Jumlah Cocok	Akurasi Label	Ketidaksesuaian
Positif	80	82	76	92.7%	6
Negatif	70	68	63	92.6%	5
Netral	50	50	46	92.0%	4
Total	200	200	185	92.5%	15

4.4 Pelatihan Model

Proses pelatihan dalam penelitian ini dilakukan menggunakan bahasa pemrograman Python dengan pustaka utama TensorFlow dan Transformers dari Hugging Face. Tujuan utama dari pelatihan adalah untuk membandingkan performa dua model pembelajaran mendalam, yaitu Deep Neural Network (DNN) dan Long Short-Term Memory (LSTM), dalam melakukan klasifikasi sentimen berbahasa Indonesia yang telah direpresentasikan melalui IndoBERT-base-p1.

Langkah pertama yang dilakukan adalah pra-pemrosesan data dan pengkodean label. Dataset yang digunakan terdiri dari data teks sentimen dari berbagai platform transportasi online (seperti Gojek, Grab, Maxim, dan InDrive), yang telah dibersihkan dan diproses sebelumnya. Label kategori sentimen dikonversi ke bentuk numerik menggunakan algoritma LabelEncoder agar dapat digunakan dalam proses klasifikasi.

Setelah data disiapkan, proses dilanjutkan dengan ekstraksi embedding dari IndoBERT. Representasi vektor dari setiap teks dihasilkan melalui tokenizer dan model pre-trained IndoBERT. Vektor-vektor ini berfungsi sebagai fitur masukan bagi model DNN dan LSTM, memungkinkan pemanfaatan representasi kontekstual yang lebih dalam dan relevan terhadap Bahasa Indonesia.

Selanjutnya, dilakukan konfigurasi parameter pelatihan model. Pengaturan eksperimen mencakup penggunaan panjang token maksimal sebesar 128, ukuran batch 64, learning rate $2e-5$, serta pelatihan selama 3 epoch. Untuk model LSTM, digunakan dua lapisan dengan ukuran tersembunyi 512, serta dropout 0.3 untuk menghindari overfitting. Seluruh proses pelatihan dan evaluasi dijalankan pada perangkat GPU jika tersedia, yang secara signifikan mempercepat komputasi, terutama pada model LSTM.

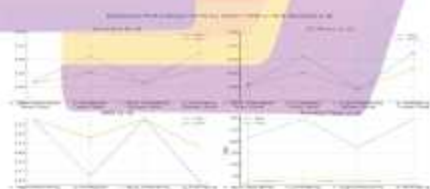
Penilaian performa model dilakukan dengan pendekatan K-Fold Cross Validation sebanyak x lipatan ($k = x$), guna meningkatkan reliabilitas hasil evaluasi dan menghindari bias akibat pembagian data yang tetap. Dalam skema ini, data dibagi menjadi lima bagian, dan pelatihan dilakukan sebanyak lima kali, di mana setiap lipatan bergiliran digunakan sebagai data validasi.

Selama proses pelatihan, kedua model (DNN dan LSTM) dievaluasi menggunakan sejumlah metrik utama, yaitu Accuracy, Precision, Recall, F1-Score, dan Mean Absolute Error (MAE). Selain itu, dicatat pula waktu pelatihan sebagai indikator efisiensi. Hasil dari setiap fold dirata-rata untuk memperoleh performa keseluruhan.

Akhirnya, hasil evaluasi divisualisasikan dalam bentuk grafik, yang menunjukkan tren performa model di setiap lipatan untuk setiap metrik. Hal ini memberikan pemahaman yang lebih mendalam mengenai kestabilan dan efektivitas masing-masing model dalam kondisi yang berbeda. Visualisasi ini menjadi dasar utama dalam menarik kesimpulan mengenai model mana yang paling cocok diterapkan dalam konteks klasifikasi opini konsumen transportasi online berbahasa Indonesia.

4.5 Evaluasi Model

Evaluasi model dilakukan sebagaimana dijelaskan pada Subbab 3.3 *Metode Analisis Data*, dengan menggunakan skema K-Fold ($K = 3$). Penilaian performa model mengacu pada beberapa metrik, yaitu: Accuracy dihitung menggunakan persamaan (2.2.1), Precision dihitung menggunakan persamaan (2.2.2), Recall dihitung menggunakan persamaan (2.2.3), F1-Score dihitung menggunakan persamaan (2.2.4), dan MAE dihitung menggunakan persamaan (2.2.5). Hasil evaluasi tersebut ditunjukkan pada Gambar 4.1 dan Tabel 4.10.



Gambar 4. 1 Hasil Uji Performa Model DNN dan LSTM

Tabel 4. 11 Hasil Uji Performa Model DNN dan LSTM

Skenario	Model	Accuracy	Precision	Recall	F1-Score	MAE	Training Time (s)
a. Tanpa resampling dan tanpa kamus slang	DNN	0.8156	0.8056	0.8073	0.8053	0.2306	28.46
	LSTM	0.8223	0.8119	0.8126	0.8108	0.2277	214.71
b. Dengan Resampling dan tanpa kamus slang	DNN	0.8528	0.8555	0.8533	0.8524	0.1905	32.40
	LSTM	0.9130	0.9137	0.9131	0.9129	0.1095	294.49
c. Tanpa Resampling dengan kamus slang	DNN	0.8130	0.7989	0.7851	0.7902	0.2318	26.77
	LSTM	0.8179	0.8025	0.7921	0.7954	0.2309	172.69
d. Dengan Resampling dan dengan Kamus Slang	DNN	0.8666	0.8683	0.8668	0.8666	0.1697	35.55
	LSTM	0.9249	0.9254	0.9251	0.9249	0.0972	295.59

Dalam evaluasi awal dengan skema K-Fold Cross Validation sebanyak tiga kali ($K = 3$), secara umum, hasil pengujian menunjukkan bahwa LSTM unggul dalam semua metrik performa dibandingkan DNN. Pada skenario dasar (a), performa kedua model relatif seimbang dengan akurasi sekitar 0.82 untuk LSTM dan 0.81 untuk DNN. Namun, perbedaan signifikan mulai terlihat pada skenario (b) dan (d), di mana resampling diterapkan. LSTM mampu mencapai akurasi lebih dari 0.91 pada skenario (b) dan meningkat hingga 0.92 pada skenario (d). Sementara itu, DNN juga mengalami peningkatan dengan resampling, tetapi tetap berada di bawah performa LSTM. Hal ini menunjukkan bahwa resampling merupakan faktor kunci dalam meningkatkan kualitas klasifikasi, terutama pada model LSTM yang lebih sensitif terhadap distribusi data.

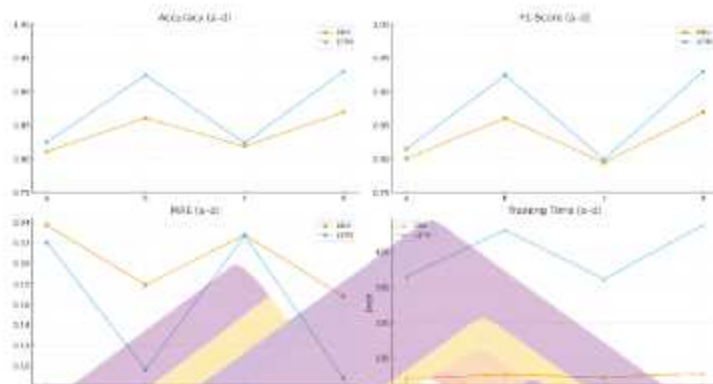
Penggunaan kamus slang memberikan dampak yang relatif kecil dibandingkan resampling. Pada skenario (c), baik LSTM maupun DNN tidak menunjukkan peningkatan signifikan, bahkan performa cenderung stagnan jika dibandingkan dengan skenario (a). Namun, ketika digabungkan dengan resampling (skenario d), kamus slang memberikan tambahan perbaikan, meskipun kontribusinya tidak sebesar resampling. Dengan demikian, dapat disimpulkan

bahwa kamus slang lebih berperan sebagai pendukung daripada faktor utama dalam meningkatkan performa model.

Selain dari sisi akurasi, hasil pengujian juga memperlihatkan adanya trade-off yang jelas antara performa dan efisiensi waktu pelatihan. LSTM membutuhkan waktu pelatihan yang jauh lebih lama, berkisar antara 172 hingga 295 detik, sedangkan DNN hanya memerlukan sekitar 26 hingga 36 detik. Artinya, DNN mampu beradaptasi lebih cepat dengan data meskipun mengorbankan tingkat akurasi. Sebaliknya, LSTM lebih lambat tetapi mampu memberikan hasil prediksi yang lebih konsisten, presisi, dan dengan kesalahan yang lebih rendah (MAE lebih kecil). Selanjutnya dilakukan Pelatihan model dengan K-Fold sama dengan 5 didapati hasil Evaluasi Model seperti pada Gambar 4.2 dan Tabel 4.11.

Tabel 4. 12 Hasil Uji Performa Model DNN dan LSTM

Skenario	Model	Accuracy	Precision	Recall	F1-Score	MAE	Training Time (s)
a. Tanpa resampling dan tanpa kamus slang	DNN	0.8113	0.8009	0.8015	0.8005	0.2374	42.56
	LSTM	0.8253	0.8199	0.8134	0.8149	0.2206	329.84
b. Dengan resampling dan tanpa kamus slang	DNN	0.8611	0.8629	0.8611	0.8612	0.1791	53.22
	LSTM	0.9246	0.9252	0.9247	0.9246	0.0963	461.9
c. Tanpa resampling dengan kamus slang	DNN	0.8189	0.8062	0.7873	0.7945	0.228	45.84
	LSTM	0.824	0.8119	0.795	0.7994	0.2273	323.02
d. Dengan resampling dan dengan kamus slang	DNN	0.8695	0.871	0.8697	0.8694	0.1678	55.97
	LSTM	0.9302	0.9308	0.9303	0.9301	0.0879	475.49



Gambar 4. 2 Hasil Uji Performa Model DNN dan LSTM

Berdasarkan hasil evaluasi dengan teknik K-Fold Cross Validation sebanyak tiga kali lipat ($K = 5$), model Long Short-Term Memory (LSTM) kembali menunjukkan performa yang secara konsisten lebih unggul dibandingkan Deep Neural Network (DNN) dalam tugas klasifikasi sentimen terhadap data Twitter layanan transportasi online. Seluruh metrik utama yang digunakan untuk mengevaluasi kinerja model memperlihatkan keunggulan LSTM, meskipun selisih yang ditunjukkan bersifat moderat.

memperlihatkan perbandingan performa model LSTM dan DNN pada empat skenario pengujian, yaitu (a) tanpa resampling dan tanpa kamus slang, (b) dengan resampling dan tanpa kamus slang, (c) tanpa resampling dan dengan kamus slang, serta (d) dengan resampling dan kamus slang. Setiap metrik utama—Accuracy, F1-Score, MAE, dan Training Time—ditampilkan secara berdampingan untuk memberikan gambaran menyeluruh mengenai trade-off antara performa prediktif dan efisiensi komputasi.

Pada grafik Accuracy dan F1-Score, terlihat jelas bahwa LSTM secara konsisten unggul dibandingkan DNN. Peningkatan performa paling signifikan terjadi pada skenario dengan resampling (b dan d). Misalnya, akurasi LSTM melonjak dari sekitar 0.82 (skenario a) menjadi lebih dari 0.92 (skenario b dan d).

Hal ini menunjukkan bahwa resampling efektif memperbaiki distribusi data yang tidak seimbang, sehingga meningkatkan kemampuan generalisasi model, terutama untuk LSTM yang lebih kompleks.

Grafik MAE (Mean Absolute Error) mendukung temuan ini dengan menunjukkan bahwa LSTM selalu menghasilkan kesalahan prediksi yang lebih kecil dibandingkan DNN. Nilai MAE LSTM menurun tajam pada skenario dengan resampling, mencapai titik terendah pada skenario (d) dengan nilai sekitar 0.09. Sebaliknya, DNN tetap memiliki MAE lebih tinggi di semua kondisi, meskipun juga mengalami sedikit perbaikan dengan resampling.

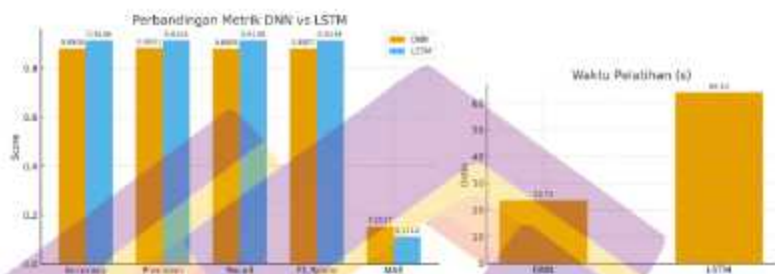
Namun, grafik Training Time menegaskan adanya trade-off signifikan antara akurasi dan efisiensi. DNN terbukti jauh lebih cepat dengan waktu pelatihan sekitar 40–55 detik di semua skenario, sementara LSTM membutuhkan waktu 6–10 kali lebih lama, berkisar antara 320–475 detik. Untuk memahami kondisi optimal penggunaan algoritma Deep Neural Network (DNN) dan Long Short-Term Memory (LSTM), dilakukan skenario uji tambahan Berdasarkan Panjang Teks.

Pada skenario uji ini dataset dibagi menjadi 2 bagian, Dimana didapati sebanyak 2.900 tweet dengan Panjang lebih kecil atau sama dengan sepuluh kata dan 7.100 tweet yang memiliki Panjang lebih dari sepuluh kata. tweet diklasifikasikan ke dalam tiga kelas target, dan seluruh eksperimen menggunakan parameter tetap untuk memastikan konsistensi. Konfigurasi yang digunakan meliputi panjang maksimum tokenisasi (`max_length`) sebanyak 128 token, ukuran mini-batch (`batch_size`) sebesar 64, laju pembelajaran (`learning_rate`) sebesar $2e-5$, dan jumlah epoch pelatihan sebanyak 3. Selain itu, evaluasi performa model dilakukan menggunakan skema *x-fold cross-validation* untuk menjaga validitas hasil dan menghindari bias akibat pembagian data.

Untuk tweet kurang dari atau sama dengan 10 kata dan menggunakan $K\text{-Fold} = 3$ didapati hasil seperti pada Gambar 4.3 dan Tabel 4.13.

Tabel 4. 13 Hasil Evaluasi Model

Model	Accuracy	Precision	Recall	F1-Score	MAE	Training Time (s)
DNN	0.8806	0.8831	0.8805	0.8807	0.1537	23.73
LSTM	0.9138	0.9144	0.9139	0.9138	0.1112	64.10

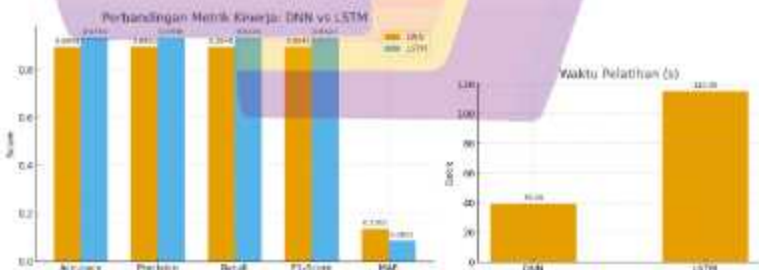


Gambar 4. 3 Perbandingan Kinerja Model

Untuk tweet kurang dari atau sama dengan 10 kata dan menggunakan K-Fold = 5
 didapati hasil seperti pada Gambar 4.4 dan Tabel 4.14.

Tabel 4. 14 Hasil Evaluasi Model

Model	Accuracy	Precision	Recall	F1-Score	MAE	Training Time (s)
DNN	0.8948	0.8960	0.8948	0.8945	0.1362	39.46
LSTM	0.9330	0.9330	0.9328	0.9327	0.0855	115.35

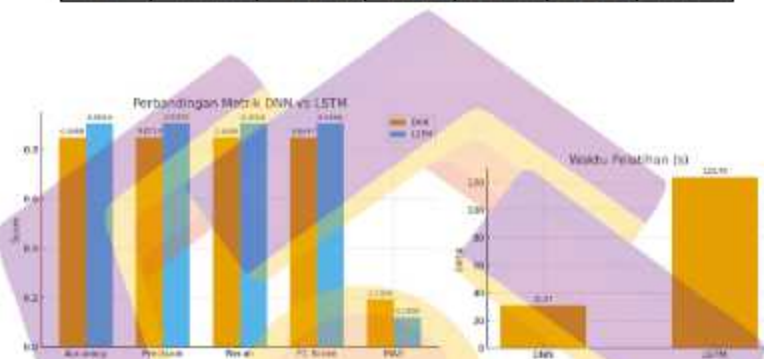


Gambar 4. 4 Perbandingan Kinerja Model

Untuk tweet lebih dari 10 kata dan menggunakan K-Fold = 3 didapati hasil seperti pada Gambar 4.5 dan Tabel. 4.15.

Tabel 4. 15 Hasil Evaluasi Model

Model	Accuracy	Precision	Recall	F1-Score	MAE	Training Time (s)
DNN	0.8498	0.8514	0.8499	0.8497	0.1926	31.07
LSTM	0.9060	0.9072	0.9058	0.9056	0.1208	123.78

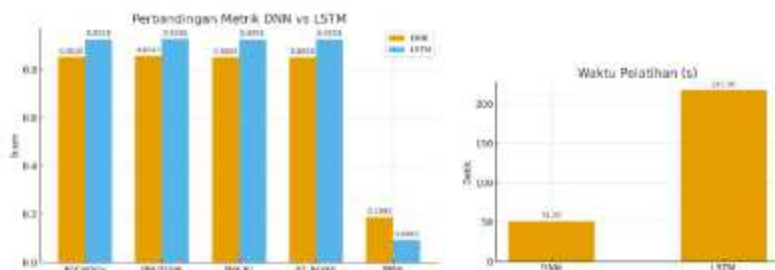


Gambar 4. 5 Perbandingan Kinerja Model

Untuk tweet lebih dari 10 kata dan menggunakan K-Fold = 5 didapati hasil seperti pada Gambar 4.6 dan Tabel. 4.16.

Tabel 4. 16 Hasil Evaluasi Model

Model	Accuracy	Precision	Recall	F1-Score	MAE	Training Time (s)
DNN	0.8515	0.8547	0.8504	0.8508	0.1892	51.22
LSTM	0.9258	0.9265	0.9255	0.9253	0.0943	217.30



Gambar 4. 6 Perbandingan Kinerja Model

Hasil ini menunjukkan bahwa untuk sentimen dengan panjang lebih kecil atau sama dengan 10 kata, kedua model baik DNN atau LSTM, mencapai performa yang tinggi, namun LSTM secara konsisten menunjukkan keunggulan dalam seluruh metrik evaluasi. Pada skema validasi silang 3 fold, LSTM mencatat akurasi 91,38% dan F1-score 0,9138, sedikit lebih tinggi dibanding DNN yang mencapai akurasi 88,06% dan F1-score 0,8807. Pola serupa terlihat pada 5 fold, di mana LSTM mencapai akurasi 93,30% dan F1-score 0,9327, melampaui DNN dengan akurasi 89,48% dan F1-score 0,8945. Selisih performa antara kedua model berada pada kisaran 3 sampai dengan 4 poin persentase, dan Mean Absolute Error (MAE) pada LSTM juga lebih rendah, menandakan prediksi yang lebih akurat secara numerik. Walaupun demikian, model DNN menonjol dari sisi efisiensi, dengan waktu pelatihan jauh lebih cepat dari pada LSTM, menjadikannya pilihan yang lebih ringan untuk teks pendek dalam konteks dengan keterbatasan komputasi.

Sementara pada sentimen panjang kata lebih dari 10 kata, LSTM menunjukkan keunggulan yang semakin jelas. Arsitektur LSTM, yang dirancang untuk memproses urutan dan mempertahankan konteks jangka panjang, lebih efektif dalam menangkap dependensi antar kata pada teks panjang. Hal ini tercermin dari performa LSTM yang meningkat tajam dibanding DNN. Pada validasi 5 fold, LSTM memperoleh akurasi 92,58% dan F1-score 0,9253, sedangkan DNN hanya mencapai akurasi 85,15% dan F1-score 0,8508. MAE LSTM juga jauh lebih rendah (0,0943 dibanding 0,1892 pada DNN), menunjukkan bahwa prediksi LSTM lebih dekat terhadap label sebenarnya. Perbedaan ini mengindikasikan bahwa model DNN mengalami penurunan efektivitas saat

menangani teks yang lebih panjang, diduga karena keterbatasannya dalam memodelkan urutan kata.

Penelitian ini juga menerapkan dua skema validasi silang 3 fold dan 5 fold cross validation, untuk mengukur generalisasi model. Hasil evaluasi menunjukkan bahwa penggunaan 5 fold secara umum menghasilkan metrik yang sedikit lebih tinggi dan konsisten dibanding 3 fold, terutama pada model LSTM. Hal ini menunjukkan bahwa validasi silang dengan jumlah fold lebih banyak memberikan estimasi performa yang lebih stabil dan mendekati kondisi nyata, terutama saat ukuran data terbatas. Namun, peningkatan jumlah fold juga menyebabkan waktu pelatihan meningkat secara proporsional, terutama pada LSTM yang memiliki arsitektur lebih kompleks.

Selain metrik klasifikasi seperti akurasi dan F1-score, penelitian ini menggunakan mean absolute error (MAE) sebagai indikator tambahan untuk mengevaluasi kualitas prediksi. MAE memberikan informasi tentang rata-rata selisih prediksi terhadap label aktual secara numerik, dan hasil menunjukkan bahwa LSTM secara konsisten menghasilkan MAE yang lebih rendah daripada DNN di seluruh skenario. Hal ini memperkuat bukti bahwa LSTM tidak hanya lebih baik dalam mengklasifikasi, tetapi juga lebih presisi secara kuantitatif.

Secara keseluruhan, hasil analisis memperlihatkan bahwa pemilihan arsitektur model perlu disesuaikan dengan karakteristik panjang input teks dan tujuan evaluasi. LSTM unggul dalam akurasi dan konsistensi, terutama pada teks panjang, namun membutuhkan waktu pelatihan lebih lama. Sebaliknya, DNN dapat menjadi pilihan alternatif yang efisien pada teks pendek dengan kompromi kecil dalam performa.

4.5.1 Perbandingan Hasil dengan Penelitian Sebelumnya

Penelitian ini selaras dengan hasil-hasil penelitian terdahulu, sebagaimana ditunjukkan pada Tabel 2.1, yang menggunakan pendekatan serupa, bahwa baik model Deep Neural Network (DNN) maupun Long Short-Term Memory (LSTM) memiliki performa yang kompetitif tergantung pada karakteristik data yang digunakan, khususnya panjang teks input. Untuk menilai signifikansi temuan ini,

dilakukan perbandingan dengan beberapa penelitian terdahulu yang relevan di bidang klasifikasi teks dan pemrosesan bahasa alami.

Penelitian oleh Rakshit et al. (2024) menemukan bahwa model LSTM mampu mencapai akurasi sebesar 83,5% dan F1-score sebesar 84,6% dalam tugas klasifikasi teks dengan data berukuran besar. Penelitian tersebut juga menyebutkan bahwa mempertahankan stopwords justru dapat meningkatkan performa model, yang menunjukkan bahwa LSTM efektif dalam memahami konteks teks yang kaya dan panjang. Hal ini mendukung hasil dalam penelitian ini, di mana LSTM menunjukkan keunggulan pada data dengan panjang >10 kata, karena mampu menangkap urutan dan relasi antar kata secara lebih baik.

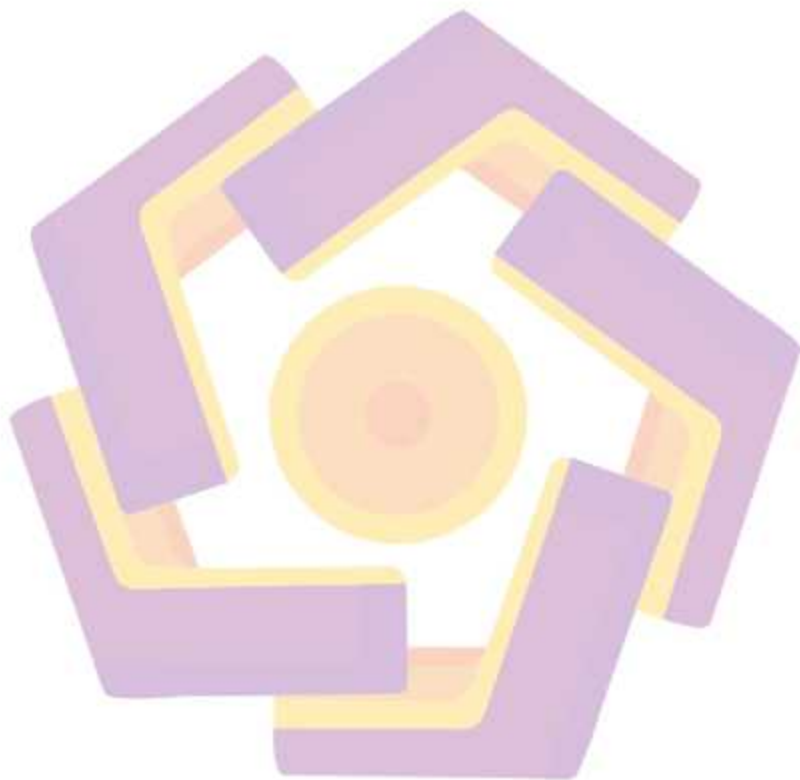
Sementara itu, Urolagin dan Patel (2024) menyoroti keunggulan DNN, yang mencapai akurasi sangat tinggi sebesar 98,62% dalam klasifikasi status sosial pengguna berbasis fitur sosial. Mereka menyatakan bahwa DNN unggul dalam efisiensi dan kecepatan prediksi, terutama pada fitur yang tidak memiliki urutan semantik kompleks. Temuan ini sejalan dengan hasil penelitian ini yang menunjukkan bahwa DNN lebih unggul pada data pendek (≤ 10 kata), karena strukturnya yang sederhana cocok untuk mengolah representasi statis seperti hasil embedding dari IndoBERT tanpa ketergantungan terhadap urutan kata.

Selain itu, Vyas et al. (2022) melaporkan bahwa LSTM menghasilkan akurasi tertinggi (83%) dalam klasifikasi multikelas sentimen dari data Twitter. Ini menguatkan hasil bahwa LSTM cenderung memberikan performa terbaik pada data sosial media yang panjang dan bervariasi secara semantik, sama seperti yang ditemukan dalam eksperimen terhadap data teks >10 kata dalam penelitian ini.

Sebagai pembandingan dari pendekatan klasik, Safira et al. (2023) menggunakan algoritma Support Vector Machine (SVM) dan memperoleh akurasi sebesar 80,3%. Namun, mereka menyimpulkan bahwa pendekatan deep learning seperti DNN dan LSTM lebih unggul, terutama dalam menangani konteks kalimat yang lebih kompleks. Hasil ini juga konsisten dengan temuan penelitian ini yang menunjukkan bahwa baik DNN maupun LSTM mengungguli model klasik, serta memiliki keunggulan masing-masing tergantung kondisi input.

Dari perbandingan ini, dapat disimpulkan bahwa temuan penelitian ini tidak hanya konsisten dengan studi-studi sebelumnya, tetapi juga memberikan sudut

pandang baru terkait pemilihan arsitektur model berdasarkan panjang teks dan kompleksitas konteks, yang belum banyak dibahas secara eksplisit dalam literatur terdahulu. Dengan demikian, kontribusi penelitian ini tidak hanya pada evaluasi empiris, tetapi juga dalam memberikan panduan praktis bagi pemilihan model yang lebih tepat pada data berbahasa Indonesia.



BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil dan analisis yang telah dilakukan, kesimpulan dari penelitian ini dapat dirumuskan sebagai berikut:

1. Berdasarkan hasil komparasi antara arsitektur Deep Neural Network (DNN) dan Long Short-Term Memory (LSTM), dapat disimpulkan bahwa LSTM menunjukkan kinerja yang lebih unggul dalam analisis sentimen teks berbahasa Indonesia. LSTM lebih efektif menangkap konteks dan urutan kata, sehingga menghasilkan akurasi dan skor F1 yang lebih tinggi dibandingkan DNN, khususnya pada teks yang lebih panjang. Dengan demikian, penelitian ini merekomendasikan penggunaan LSTM sebagai arsitektur utama dalam pengembangan sistem analisis sentimen layanan transportasi online di Indonesia berbasis data Twitter, sementara DNN tetap relevan ketika efisiensi komputasi menjadi prioritas.
2. Penerapan metode K-Fold Cross Validation berhasil meningkatkan reliabilitas dan generalisasi evaluasi model. Dengan membagi data ke dalam beberapa lipatan, variabilitas antarpercobaan dapat ditekan, dan performa model dievaluasi lebih stabil. Hal ini juga memperkuat keandalan perbandingan antara model DNN dan LSTM dalam penelitian ini.

5.2 Kontribusi Kebaruan (Novelty)

Kontribusi kebaruan dari penelitian ini terletak pada temuan bahwa model LSTM, jika dilatih secara cukup dan mendalam, memberikan hasil klasifikasi sentimen yang lebih baik dibandingkan DNN pada teks pendek dalam Bahasa Indonesia yang diolah dengan representasi IndoBERT. Penelitian ini juga menambahkan bukti empiris bahwa pemanfaatan K-Fold Cross Validation relevan untuk data kecil dengan distribusi tidak seimbang.

DAFTAR PUSTAKA

- Mehta, N., & Pandit, A. (2018). Concurrence of big data analytics and healthcare: A systematic review. *International Journal of Medical Informatics*, 114, 57–65.
- Mitropoulos, L., Kortsari, A., & Ayfantopoulou, G. (2021). Factors affecting drivers to participate in a carpooling to public transport service. *Sustainability*, 13(16), 9129. <https://doi.org/10.3390/su13169129>
- Muraina, I. O. (2022). Ideal dataset splitting ratios in machine learning algorithms: General concerns for data scientists and data analysts [Conference paper]. *ResearchGate*. <https://www.researchgate.net/publication/363844470>
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135. <https://doi.org/10.1561/1500000011>
- Pullanikkat, R., Poddar, S., Das, A., Jaiswal, T., Singh, V. K., & Basu, M. (2024). Utilizing the Twitter social media to identify transportation-related grievances in Indian cities. *Social Network Analysis and Mining*, 14, 118. <https://doi.org/10.1007/s13278-024-01278-x>
- Qi, Y., & Shabrina, Z. (2023). Sentiment analysis using Twitter data: A comparative application of lexicon- and machine-learning-based approach. *Social Network Analysis and Mining*, 13, 31. <https://doi.org/10.1007/s13278-023-01030-x>
- Rakshit, P., Sarkar, P., & Roy, S. (2024). Hybrid deep learning approach for sentiment analysis on Twitter data. *Multimedia Tools and Applications*. <https://doi.org/10.1007/s11042-024-19555-4>
- Safira, A. N., Pujastuti, E., Hanafi, H., Setiaji, B., Prabowo, D., & Cahyono, N. (2023). Sentiment analysis to find out positive or negative opinions on ride hailing application. In *Proceedings of the 2023 International Conference on Informatics, Multimedia, Cyber and Information Systems (ICIMCIS)* (pp. 341–348). <https://doi.org/10.1109/ICIMCIS60089.2023.10349083>
- Sun, T., Wang, Y., Liu, R., & Zhao, H. (2023). LSTM networks for time series analysis: Recent advances. *Information Sciences*, 587, 91–108.

- Twitter, Inc. (2024). Twitter API documentation. <https://developer.twitter.com/en/docs>
- Urolagin, S., & Patel, S. (2024). User-specific loyalty measure and prediction using deep neural network from Twitter data. *IEEE Transactions on Computational Social Systems*, 11(1), 1046–1059. <https://doi.org/10.1109/TCSS.2023.3239523>
- Vyas, P., Reisslein, M., Rimal, B. P., Vyas, G., Basyal, G. P., & Muzumdar, P. (2022). Automated classification of societal sentiments on Twitter with machine learning. *IEEE Transactions on Technology and Society*, 3(2), 100–112. <https://doi.org/10.1109/TTS.2021.3108963>
- Yang, L., Hartono, A., Siregar, R., & Susanto, Y. (2023). Natural language processing in Indonesian: Current progress and future directions. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 22(3), 1–28.
- Zhang, Q., Chen, M., & Wu, L. (2023). Deep neural networks for text classification: A comprehensive review. *Knowledge-Based Systems*, 259, 110162.
- Zhang, Y., & Liu, X. (2023). Sentiment analysis in transportation services: A deep learning framework. *Transportation Research Part E: Logistics and Transportation Review*, 171, 102966.
- Zhou, W., Prabowo, R., & Kurniawan, A. (2023). BERT-based models for Indonesian language processing: A comprehensive study. *Neural Computing and Applications*, 35(8), 5678–5694.
- Zhu, L., Tan, Y., & Wei, J. (2023). Deep learning for social media analysis: Recent advances and applications. *Information Processing & Management*, 60(2), 103123.