SKANIKA: Sistem Komputer dan Teknik Informatika
Volume 8, Nomor 2, Juli 2025, Halaman 244-256 E-ISSN: 2721-4788

Analisis Sentimen Opini Publik Terhadap Kasus Korupsi Bahan Bakar Minyak Oplosan PT Pertamina dengan Hybrid Model Deep Learning

Muhammad Ramdhan Awali[1*], Sawali Wahyu[2], Anik Hanifatul Azizah[3]
[1,2,3]Fakultas Ilmu Komputer, Sistem Informasi, Universitas Esa Unggul, Jakarta, Indonesia
E-mail: [1*]ramdhangt@gmail.com, [2]sawaliwahyu@esaunggul.ac.id, [3]anik.hanifa@esaunggul.ac.id
(* : corresponding author)

Abstrak

Kasus dugaan korupsi bahan bakar minyak (BBM) oplosan yang melibatkan PT Pertamina menjadi isu nasional dan memicu berbagai respons publik. Analisis sentimen di media sosial memberikan wawasan bagi pemerintah dalam memahami persepsi masyarakat terhadap isu tersebut. Penelitian ini bertujuan menganalisis opini publik menggunakan pendekatan hybrid model deep learning. Data dikumpulkan dari Twitter (X) pada periode 24 Februari hingga 19 Maret 2025, menghasilkan 12.365 tweet setelah preprocessing. Jumlah data dianggap representatif karena mencerminkan persebaran opini selama puncak perbincangan, dengan topik dan ekspresi sentimen yang beragam. Preprocessing mencakup pembersihan teks, normalisasi, case folding, tokenisasi, stopword removal, dan stemming. Pelabelan dilakukan dengan pendekatan lexicon-based untuk mengklasifikasikan tweet ke dalam tiga kategori: positif, negatif, dan netral. Empat model diterapkan: IndoBERT, CNN, LSTM, dan hybrid IndoBERT-CNN-LSTM. Evaluasi menggunakan confusion matrix menunjukkan bahwa IndoBERT memiliki akurasi tertinggi (90%), diikuti CNN (86%), hybrid (84%), dan LSTM (69%). Skema validasi K-Fold memberikan hasil evaluasi yang lebih stabil dibandingkan metode Hold-Out. Sentimen publik didominasi oleh respons negatif (72%), sementara sentimen positif dan netral masing-masing sebesar 16%. Penelitian ini menunjukkan pentingnya pemilihan arsitektur model dan strategi validasi yang tepat dalam klasifikasi teks Bahasa Indonesia untuk pemetaan opini publik berbasis media sosial.

Kata kunci: Analisis Sentimen, Pertamina, Hybrid, Deep Learning, Twitter

Abstract

The alleged corruption case involving adulterated fuel (BBM) by PT Pertamina became a national issue and triggered various public responses. Sentiment analysis on social media can provide insights for the government and stakeholders to understand public perception of the issue. This study aims to analyze public opinion using a hybrid deep learning model. Data were collected from Twitter (X) between February 24 and March 19, 2025, resulting in 12,365 tweets after preprocessing. This amount is considered representative, as it reflects public discourse during the peak period of the topic, with diverse expressions and content. Preprocessing included text cleaning, normalization, case folding, tokenization, stopword removal, and stemming. Sentiment labels (positive, negative, neutral) were assigned using a lexicon-based approach tailored to the Indonesian language context. Four models were implemented: IndoBERT, CNN, LSTM, and a hybrid model combining IndoBERT-CNN-LSTM. Evaluation using a confusion matrix showed that IndoBERT achieved the highest accuracy (90%), followed by CNN (86%), hybrid (84%), and LSTM (69%). The K-Fold validation scheme yielded more stable evaluation results than the Hold-Out method. The sentiment distribution revealed a dominance of negative sentiment (72%), while positive and neutral sentiments each accounted for 16%. This study highlights the importance of appropriate model architecture and validation strategy in Indonesian text classification for mapping public opinion on social media.

***Keywords**: Sentiment Analysis, Pertamina, Hybrid, Deep Learning, Twitter*

1. PENDAHULUAN

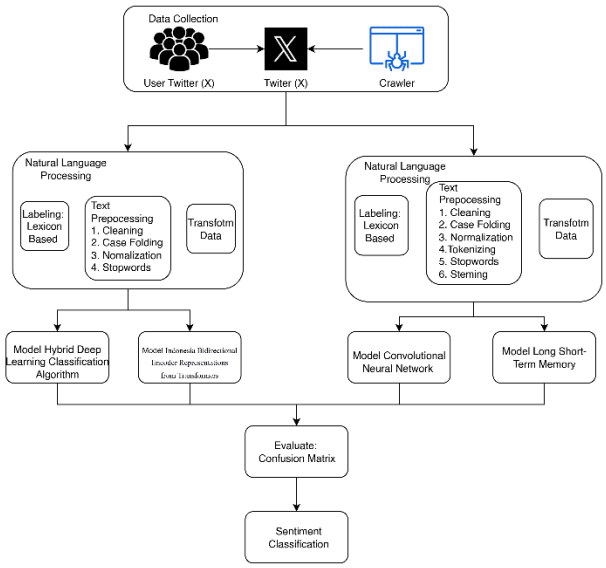
Kasus dugaan korupsi distribusi BBM oplosan Pertamax oleh PT Pertamina memicu reaksi luas di media sosial, khususnya di *platform Twitter* (X), dengan dominasi sentimen negatif [1]. ini menunjukan pentingnya analisis sentimen untuk memahami persepsi publik secara cepat dan terukur. Seperti disampaikan oleh [2], pelanggaran dalam distribusi BBM tidak hanya merugikan negara, tetapi juga menimbulkan ketidakadilan sosial, sehingga opini masyarakat perlu dikaji sebagai masukan dalam evaluasi kebijakan publik. Penelitian ini mengumpulkan 12.924 tweet berbahasa Indonesia terkait lima kata kunci utama: “Pertamax”, “Pertalite”, “Pertamina”, “BBM”, dan “korupsi”, yang kemudian dianalisis menggunakan pendekatan *deep learning*.

Dengan kemajuan teknologi kecerdasan buatan, pendekatan *hybrid deep learning* menjadi semakin relevan untuk meningkatkan akurasi klasifikasi teks. Penelitian sebelumnya oleh [3] menunjukkan bahwa kombinasi BERT dan Bi-LSTM atau TCN efektif dalam menangkap konteks semantik teks berbahasa Indonesia. Sementara itu, [4] menunjukkan bahwa gabungan IndoBERT dan LSTM memberikan akurasi tinggi dalam deteksi hoaks di *Twitter*. Berdasarkan temuan tersebut, penelitian ini menerapkan pendekatan *hybrid* yang menggabungkan IndoBERT, CNN, dan LSTM untuk melakukan klasifikasi sentimen ke dalam tiga kelas (positif, negatif, dan netral), serta membandingkan performa antara model hybrid dan model individual. Evaluasi dilakukan menggunakan dua skema validasi: hold-out dan k-fold cross validation, dengan tujuan menghasilkan pemetaan opini publik yang akurat terhadap isu sosial strategis.

Penelitian sebelumnya umumnya fokus pada klasifikasi biner dan belum banyak menerapkan pendekatan hybrid untuk sentimen multi-kelas dalam isu nasional. Gap penelitian ini ada pada penggunaan hybrid IndoBERT-CNN-LSTM untuk klasifikasi opini publik di media sosial. IndoBERT dipilih karena unggul dalam memahami konteks Bahasa Indonesia, CNN untuk menangkap fitur lokal, dan LSTM untuk memproses urutan kata. Kombinasi ini diharapkan dapat meningkatkan akurasi klasifikasi dibandingkan model individual.

2. METODE PENELITIAN

Proses dimulai dengan tahapan preprocessing teks, seperti case folding, tokenisasi, dan normalisasi. Setelah itu, teks diproses oleh IndoBERT untuk menghasilkan embedding kontekstual yang merepresentasikan makna tiap kata dalam konteks kalimat. Embedding ini kemudian diteruskan ke arsitektur CNN dan LSTM secara berurutan sebelum masuk ke layer klasifikasi akhir. Arsitektur lengkap dari model hybrid ini ditampilkan pada Gambar 1 berikut:



Gambar 1. Arsitektur Analisis Sentimen

Gambar 1 menunjukkan alur proses analisis sentimen, dimulai dari pengumpulan data Twitter menggunakan *Twitter API v2* dan *Python library tweepy*. Data mentah kemudian melalui tahapan preprocessing dengan menggunakan Python dan pustaka Sastrawi, re, dan NLTK, yang mencakup *case folding, stopword removal*, tokenisasi, dan *stemming*. Proses pelabelan dilakukan secara otomatis dengan pendekatan lexicon-based menggunakan daftar kata sentimen dari Kamus Sentimen Indonesia. Selanjutnya, data dilatih dengan tiga pendekatan model: IndoBERT (menggunakan transformers dari HuggingFace), CNN dan LSTM (dibangun menggunakan TensorFlow dan Keras), baik secara individu maupun dalam bentuk *hybrid*. Seluruh hasil klasifikasi dievaluasi menggunakan confusion matrix dan *classification report* dari sklearn.metrics untuk menentukan label akhir: positif, negatif, atau netral..

2.1. Pengumpulan Data

dalam penelitian ini data di *crawling* dari *Twitter* (x) menggunakan teknik *crawling* melalui *Twitter API* dengan *Python* selama periode 24 Februari hingga 19 Maret 2025. Proses ini mengikuti metode otomatisasi yang efektif untuk pengambilan data dari media sosial [5] , dengan kata kunci seperti "Pertamax", "BBM", "Korupsi Pertamina", "Pertalite", dan "Pertamina" untuk menjaga relevansi topik. Dari 12.924 tweet yang dikumpulkan, sebanyak 12.365 tweet tersisa setelah proses pembersihan. Penelitian ini berfokus pada atribut *full_text* karena memuat opini pengguna secara lengkap dan relevan untuk analisis sentimen [6]. *Twitter* (x) dipilih sebagai sumber data karena kemampuannya merepresentasikan opini publik secara cepat dan padat makna [7].

2.2 *Preprocessing* Data

Pemrosesan data merupakan tahap krusial dalam menyiapkan teks mentah sebelum analisis sentimen berbasis *NLP*. Tahapan ini mencakup deteksi duplikasi dengan *Pandas*, *cleaning*, *case folding*, *normalize* kata tidak baku, *tokenized*, *stopword removal*, dan *stemming* menggunakan Sastrawi. *Preprocessing* yang dilakukan secara cermat sangat memengaruhi validitas dan akurasi analisis, sebagaimana dijelaskan dalam [8] [9].

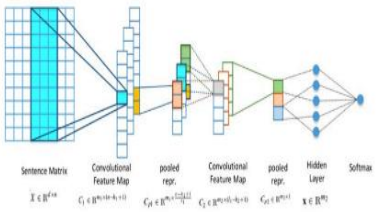
2.3 Pelabelan Data

Pelabelan data dilakukan secara semi-otomatis dengan pendekatan *lexicon-based*, yang mencocokkan kata dalam teks dengan kamus sentimen berpolaritas positif, negatif, atau netral [10]. *Lexicon-based* yang digunakan mencakup pendekatan berbasis kamus maupun korpus, seperti *SentiWordNet* [11]. Hasil pelabelan diverifikasi secara manual untuk meningkatkan akurasi, khususnya pada teks ambigu. Metode ini dipilih karena sederhana, tidak memerlukan banyak data latih, dan efektif untuk teks informal di media sosial [12].

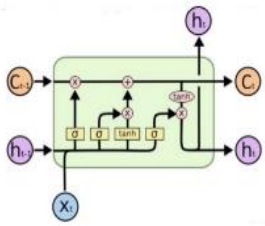
2.4 Pembagian Data

Dalam penelitian ini, pembagian data dilakukan untuk mengevaluasi performa model secara objektif menggunakan dua pendekatan, yaitu *hold-out validation* dan *K-Fold cross validatio*n. Metode *hold-out validation* membagi data sebesar 80% untuk pelatihan dan 20% untuk pengujian, dipilih karena efisien pada dataset yang besar [13] Sementara itu, *K-Fold cross validation* digunakan untuk memperoleh evaluasi yang lebih stabil dengan membagi data menjadi beberapa *fold* dan mengujinya secara bergantian, sehingga meminimalkan risiko *overfitting* dan bias [14].

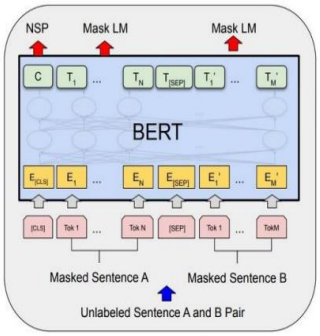
2.5 Arsitektur Model



Gambar 2. Arsitektur CNN



Gambar 3. Arsitektur LSTM



Gambar 4. Arsitektur indoBERT

Gambar 2 menunjukkan arsitektur dasar CNN yang bekerja secara bertingkat untuk mengekstraksi fitur penting dari teks. Melalui lapisan konvolusi dan pooling, CNN menyaring pola kata, kemudian merangkumnya di lapisan fully connected untuk klasifikasi sentimen secara efisien [15], Sementara itu, Gambar 3 menunjukkan arsitektur LSTM yang memanfaatkan mekanisme gerbang untuk mempertahankan konteks penting dalam data sekuensial. Struktur ini memungkinkan LSTM memahami urutan kata secara efektif, sehingga cocok untuk tugas analisis sentimen berbasis teks [16], dan Gambar 3 menunjukkan arsitektur IndoBERT yang mengubah teks menjadi representasi numerik dan memprosesnya menggunakan

transformer. Model ini dilatih dengan dua tugas utama, yaitu *Masked Language Modeling* dan *Next Sentence Prediction* untuk memahami makna dan hubungan antar kalimat [17].

2.6 Implementasi Model

Implementasi model dilakukan dengan empat pendekatan: CNN, LSTM, IndoBERT, dan *hybrid* IndoBERT-CNN-LSTM. Masing-masing diterapkan pada klasifikasi sentimen pada tiga kelas: positif, negatif, dan netral. CNN mengekstraksi fitur spasial, LSTM memahami urutan kata, dan IndoBERT memberikan representasi kontekstual dua arah untuk Bahasa Indonesia [15]. Model *hybrid* mengintegrasikan ketiganya untuk meningkatkan performa klasifikasi [18].

2.7 Evaluasi

Evaluasi dilakukan menggunakan *confusion matrix* untuk mengukur performa klasifikasi berdasarkan perbandingan label aktual dan prediksi. *Matrix* ini menunjukkan detail prediksi benar dan salah untuk tiap kelas (positif, negatif, netral). [19] *confusion matrix* penting dalam evaluasi *Deep learning* karena menyajikan informasi yang menyeluruh. Selain itu, metrik turunan seperti akurasi, presisi, dan recall memberikan wawasan lebih dalam terhadap kinerja model [20].
Dalam penelitian ini, metrik evaluasi dihitung menggunakan rumus berikut:

a. *Accuracy*

$$accuracy = \frac{Jumlah\ Prediksi\ Benar}{Total\ Data} \times 100\% \quad (1)$$

b. *Precsion Per kelas* :

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (2)$$

c. *Recall Per Kelas:*

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

2.8 Visualisasi

Visualisasi digunakan untuk memperjelas hasil klasifikasi sentimen, dengan dua jenis utama: *barchart* untuk menunjukkan jumlah sentimen per kategori, dan *wordcloud* untuk menggambarkan *frekuensi* kemunculan kata-kata dominan dalam data.. Menurut [1], visualisasi membantu memperjelas tren dan distribusi data, sementara [20] menyebutkan bahwa *wordcloud* menyederhanakan informasi kompleks dan memudahkan interpretasi. Pendekatan ini mempermudah pemahaman pola opini publik secara intuitif.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pengumpulan data

Data dikumpulkan melalui *crawling* menggunakan *Twitter API* dan aplikasi *Tweet-Harvest* v2.6.1 selama 24 Februari sampai dengan 19 Maret 2025, dengan lima kata kunci utama: “Pertamax”, “BBM”, “Korupsi Pertamina”, “Pertalite”, dan “Pertamina”. Sebanyak 12.924 *tweet* berhasil dikumpulkan dan disimpan dalam format CSV untuk tahap *preprocessing* dan analisis sentimen.

3.2 Hasil *Preprocessing* Data

Data yang diperoleh melalui lima kata kunci utama menghasilkan total 12.924 *tweet*. Setelah melalui proses identifikasi dan pembersihan terhadap data duplikat

Dalam text *preprocessing* terdapat 6 (Enam) proses yang masing-masing diuraikan sebagai berikut:

a. *Cleaning*

Pada tahap ini, data tweet dibersihkan dari unsur yang tidak relevan seperti *mention*, *URL*, dan karakter khusus agar teks menjadi rapi dan siap dianalisis, sebagaimana ditunjukkan pada Tabel 1.

Tabel 1. Hasil *Cleaning*

No	Sebelum *Cleaning*	Sesudah *Cleaning*
1	RT @user: Harga BBM naik lagi! https://bit.ly	Harga BBM naik lagi
2	#Pertamina mahal banget nih... @pertaminaid	Pertamina mahal banget nih

b. *Case Folding*

Setelah proses pembersihan, dilakukan *case folding* untuk menyeragamkan seluruh teks menjadi huruf kecil agar kata seperti "Pertamina" dan "pertamina" diperlakukan sama demi menjaga konsistensi data, sebagaimana ditunjukkan pada Tabel 2.

Tabel 2. Hasil *Case Folding*

No	Sebelum *Case Folding*	Setelah *Case Folding*
1	Pertamax isi nya pertalite pelakunya udah di amankan Tinggal Khong guan isi rengginang yang belum di tangkap	Pertamax isi nya pertalite pelakunya udah di amankan tinggal khong guan isi rengginang yang belum di tangkap
2	Penjabat tp rasa buzzer pertamax rasa pertalite ya oplosan	penjabat tp rasa buzzer pertamax rasa pertalite ya oplosan

c. *Normalization*

Normalization dilakukan untuk mengubah kata tidak baku menjadi bentuk standar guna meningkatkan konsistensi teks sebelum dianalisis, sebagaimana ditunjukkan pada Tabel 3.

Tabel 3. Hasil *Normalization*

No	Sebelum *Normalization*	Sesudah *Normalization*
1	sejujurnya klo gw nih tinggal di ibukota dan tetep make pertamina krn dr awal gw emg belinya pertalite bukan pertamax syukur2 klo gw beli pertalite dioplos pertamax wkwkwk tp tetep gw ga merasa ketipu dan pindah ke pom lain krn emg dr awal gw slalu beli Pertalite	sejujurnya kalau gw nih tinggal di ibukota dan tetep make pertamina karena dari awal gw memang belinya pertalite bukan pertamax syukur2 kalau gw beli pertalite dioplos pertamax wkwkwk tapi tetep gw ga merasa ketipu dan pindah ke pom lain karena memang dari awal gw slalu beli pertalite
2	klo emg bener ya buat pasar bebas aja hapus monopoli pertamina bebaskan perusahaan minyak lain utk masuk amp ngimpor sendiri bbmnya klo perlu kerjasama ama china buat nyambungin pipa bbm sampai ke indo krn mrk dah nyambung dg pipa rusia gk ada cara instan lain	kalau memang bener ya buat pasar bebas saja hapus monopoli pertamina bebaskan perusahaan minyak lain untuk masuk amp ngimpor sendiri bbmnya kalau perlu kerjasama ama china buat nyambungin pipa bbm sampai ke indo karena mrk dah nyambung dengan pipa rusia gk ada cara instan lain

d. *Tokenized*

Tokenizing dilakukan dengan memecah teks pada kolom *full_text* menjadi daftar kata (token) untuk mempermudah ekstraksi fitur dan proses klasifikasi dalam pemodelan NLP, sebagaimana ditunjukkan pada Tabel 4.

Tabel 4. Hasil *Tokenized*

No	Sebelum *Tokenized*	Sesudah *Tokenized*
1	jokowi terlibat skandal pertamina??	'jokowi', 'terlibat', 'skandal', 'pertamina??'
2	juara liga korupsi di indonesia dimenangkan oleh pertamina pertamina s lewat orang2senang	'juara', 'liga', 'korupsi', 'di', 'indonesia', 'dimenangkan', 'oleh', 'pertamina', 'pertamina', 's', 'lewat', 'orang2senang'

e. *Stopwords*

Stopwords removal dilakukan dengan menghapus kata-kata umum seperti "yang" dan "dan" dari token menggunakan daftar NLTK yang telah diperluas, sehingga teks menjadi lebih ringkas dan relevan untuk dianalisis, sebagaimana ditunjukkan pada Tabel 5.

Tabel 5. Hasil *Stopwords*

No	Sebelum *Stopwords*	Sesudah *Stopwords*
1	['yang', 'ini', 'bahkan', 'lebih', 'lucu', 'daripada', 'yang', 'bilang', 'pertamax', 'itu', 'bukan', 'pertalite', 'oplosan', 'buta', 'nya', 'gak', 'ketolong', 'dan', 'yang', 'ini', 'bahkan', 'lebih', 'buta', 'lagi', 'yang', 'di', 'phk', 'massal', 'itu', 'banyak', 'banget', 'yang', 'pengangguran', 'saja', 'dah', 'banyak', 'banget', 'ditambah', 'lagi', 'sama', 'ini', 'dan', 'jawabannya', 'bisa', 'sesantai', 'itu']	['lucu', 'bilang', 'pertamax', 'pertalite', 'oplosan', 'buta', 'nya', 'gak', 'ketolong', 'buta', 'phk', 'massal', 'banget', 'pengangguran', 'dah', 'banget', 'ditambah', 'jawabannya', 'sesantai']

2	['artinya', 'kondisi', 'pertamax', 'yang', 'ada', 'sudah', 'bagus', 'dan', 'sudah', 'sesuai', 'dengan', 'standar', 'yang', 'ada', 'di', 'pertamina', 'ucapnya', 'terlebih', 'kata', 'dia', 'bbm', 'adalah', 'barang', 'yang', 'habis', 'pakai', 'ia', 'mengatakan', 'bahwa', 'stok', 'kecukupan', 'bbm', 'hanya', 'sekitar', '21', '23', 'hari', 'sehingga', 'bbm', 'yang', 'dipasarkan', 'pada', 'tahun']	['kondisi', 'pertamax', 'bagus', 'sesuai', 'standar', 'pertamina', 'bbm', 'barang', 'habis', 'pakai', 'stok', 'kecukupan', 'bbm', '21', '23', 'bbm', 'dipasarkan']

f. Hasil *Stemming*

Stemming dilakukan untuk mengubah token ke bentuk dasar menggunakan Sastrawi, seperti kata "menyediakan" yang diubah menjadi "sedia". Proses ini dilakukan setelah *stopwords removal*, dan hasilnya disimpan dalam bentuk daftar token untuk analisis selanjutnya, sebagaimana ditunjukkan pada Tabel 6.

Tabel 6. Hasil *Stemming*

No	Sebelum *Stemming*	Sesudah *Stemming*
1	"['azab', 'pengoplos', 'bbm']"	['azab', 'oplos', 'bbm']
2	['cegah', 'kecurangan', 'bbm', 'jalur', 'mudik', 'wilayah', 'kabupaten', 'banjar', 'spbu', 'disidak']	['cegah', 'curang', 'bbm', 'jalur', 'mudik', 'wilayah', 'kabupaten', 'banjar', 'spbu', 'disidak']

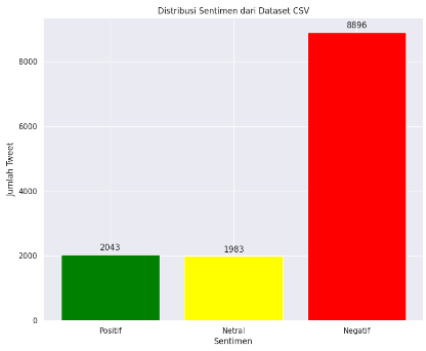
3.3 Hasil *Labeling*

Pelabelan sentimen dilakukan secara terpisah untuk model IndoBERT dan CNN-LSTM, menyesuaikan alur preprocessing masing-masing. Pada IndoBERT, labeling dilakukan tanpa normalisasi atau tokenisasi manual, sedangkan pada CNN dan LSTM dilakukan setelah semua tahap preprocessing, sebagaimana tercantum pada Tabel 7.

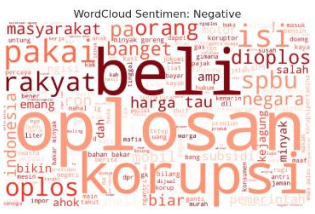
Tabel 7. Hasil *Labeling*

Full Text	Sentimen	Skor
pertalite rasa pertamax minyak kita 1 l isinya 0 8 l beras 5 kg isi 4 7 kg sebenere mereka itu goblok apa emang maling sehingga rakyat terus yg selalu jadi korban belum lagi motor telat bayar 2 disita	negatif	-1
kalo pertamina abang nya kya gini smpe 5tahun kedepan jg gue gpp ngisi pertamax oplosan pertalite	Netral	0
konsumsi pertalite dan pertamax diprediksi naik 11 persen di lebaran 2025	Positif	1

3.4 Hasil Visualisasi



5 (a)



5 (b)

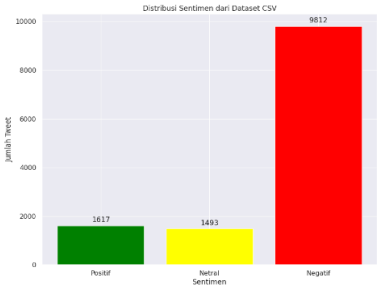
5 (c)

5 (d)

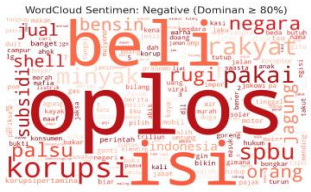
Gambar 5. (a, b, c dan d) Hasil Visualisasi Labeling Untuk Model *hybrid* dan indoBERT

Pada Gambar 5 (a, b, c dan d) hasil visualisasi labeling untuk model hybrid dan indoBERT menunjukkan bahwa dari total 12.365 *tweet*, sebanyak 8.896 (72%) tergolong sentimen negatif,

2.043 (16%) positif, dan 1.983 (16%) netral, dengan *wordcloud* menampilkan kata dominan seperti "oplosan", "beli", dan "korupsi" pada sentimen negatif yang mencerminkan ketidakpuasan publik terhadap isu yang dianalisis.

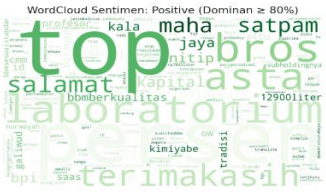


6 (a)



6 (b)

6 (c)



6 (a)

Gambar 6 (a, b, c dan d) Hasil Visualisasi Labeling Untuk Model CNN dan LSTM

Visualisasi pada Gambar 6 (a, b, c dan d) menunjukkan bahwa dari total 12.365 *tweet*, 9.612 (78%) tergolong sentimen negatif, 1.617 (13%) positif, dan 1.493 (9%) netral, dengan *wordcloud*memperlihatkan kata dominan seperti "oplos", "rugi", dan "korupsi" yang mencerminkan kritik publik terhadap isu BBM Pertamina.

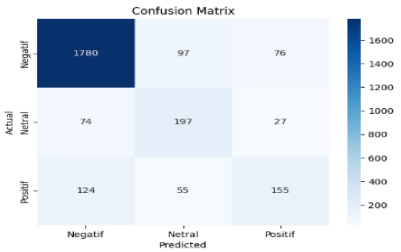
3.5 *Modeling*

Pemodelan dimulai dengan pembagian data menggunakan *hold-out validation* (80:20) dan K-*Fold cross validation*. Setiap model IndoBERT, CNN, LSTM, dan *hybrid* dilatih terpisah, dengan data latih udan data uji untuk evaluasi performa secara objektif.

a. Modeling *Hybrid*

	precision	recall	f1-score	support
Negatif	0.90	0.91	0.91	1953
Netral	0.56	0.66	0.61	298
Positif	0.60	0.46	0.52	334
accuracy			0.82	2585
macro avg	0.69	0.68	0.68	2585
weighted avg	0.82	0.82	0.82	2585

7 (a)

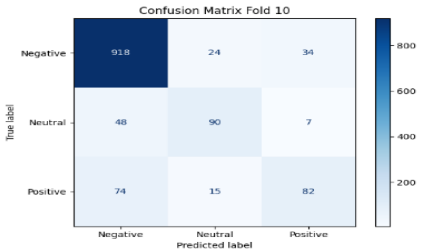


7 (b)

Gambar 7. (a dan b) Hasil Model *Hybrid* dengan Metode *Hold Out*

	precision	recall	f1-score	support
Negative	0.88	0.94	0.91	976
Neutral	0.70	0.62	0.66	145
Positive	0.67	0.48	0.56	171
accuracy			0.84	1292
macro avg	0.75	0.68	0.71	1292
weighted avg	0.83	0.84	0.84	1292

8 (a)



8 (b)

Gambar 8. (a dan b) Hasil Model *Hybrid* dengan Metode *K-Fold*

Hasil klasifikasi ditampilkan pada Gambar 7 (a dan b) dan Gambar 8 (a dan b) menunjukan Model hybrid menggabungkan IndoBERT, CNN, dan LSTM untuk menangkap konteks, fitur lokal, dan urutan kata. Arsitektur dibangun dengan Transformers dan TensorFlow.

Tabel 8. Evaluasi Model *Hybrid* dengan Metode *Hold Out*

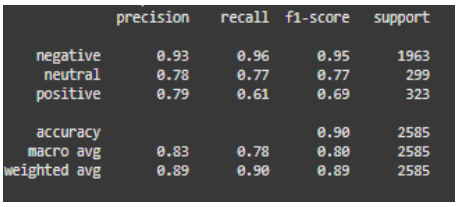
	Negatif (0)	Netral (1)	Positif (2)
Precision	$\frac{1780}{1780+198} = 0{,}90$	$\frac{197}{196+101} = 0{,}56$	$\frac{155}{155+103} = 0{,}60$
Recall	$\frac{1780}{1780+73} = 0{,}91$	$\frac{197}{197+101} = 0{,}66$	$\frac{155}{155+179} = 0{,}46$
F1-score	$2 \times \frac{0{,}90 \times 0{,}91}{0{,}90+0{,}91} = 0{,}91$	$2 \times \frac{0{,}56 \times 0{,}66}{0{,}56+0{,}66} = 0{,}61$	$2 \times \frac{0{,}60 \times 0{,}46}{0{,}60+0{,}46} = 0{,}52$

Tabel 9. Evaluasi Model *Hybrid* dengan Metode K-Fold

	Negatif (0)	Netral (1)	Positif (2)
Precision	$\frac{918}{918+122} = 0{,}88$	$\frac{90}{90+39} = 0{,}70$	$\frac{82}{82+41} 0{,}67$
Recall	$\frac{918}{918+58} = 0{,}94$	$\frac{90}{90+55} = 0{,}62$	$\frac{82}{82+89} = 0{,}48$
F1-score	$2 \times \frac{0{,}88 \times 0{,}94}{0{,}88+0{,}94} = 0{,}91$	$2 \times \frac{0{,}70 \times 0{,}62}{0{,}70+0{,}62} = 0{,}66$	$2 \times \frac{0{,}67 \times 0{,}48}{0{,}67+0{,}48} = 0{,}56$

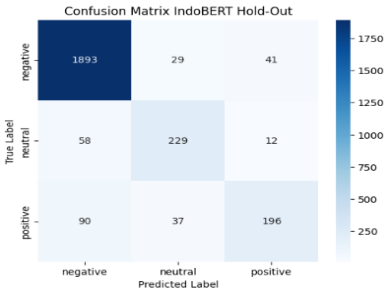
Pada Gambar 7 dan 8, serta Tabel 8 dan 9, dijelaskan model *hybrid* yang menunjukkan akurasi 82% pada metode *hold-out* dan meningkat menjadi 84% dengan *k-fold*, dengan performa tertinggi pada kelas negatif dan hasil yang lebih seimbang menggunakan validasi silang.

b. *Modeling* indoBERT



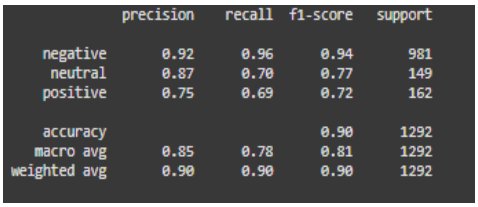
	precision	recall	f1-score	support
negative	0.93	0.96	0.95	1963
neutral	0.78	0.77	0.77	299
positive	0.79	0.61	0.69	323
accuracy			0.90	2585
macro avg	0.83	0.78	0.80	2585
weighted avg	0.89	0.90	0.89	2585

9 (a)



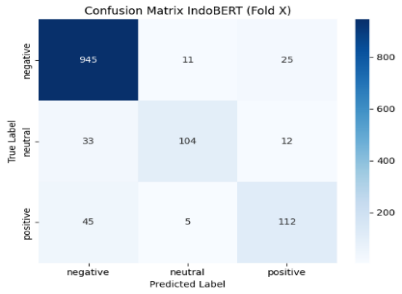
9 (b)

Gambar 9. (a dan b) Hasil Model indoBERT dengan Metode Hold Out



	precision	recall	f1-score	support
negative	0.92	0.96	0.94	981
neutral	0.87	0.70	0.77	149
positive	0.75	0.69	0.72	162
accuracy			0.90	1292
macro avg	0.85	0.78	0.81	1292
weighted avg	0.90	0.90	0.90	1292

10 (a)



10 (b)

Gambar 10. (a dan b) Hasil Model indoBERT dengan Metode K-Fold

Model IndoBERT digunakan untuk klasifikasi tiga kelas sentimen dengan arsitektur pre-trained dari Transformers, tanpa preprocessing tambahan karena memiliki tokenizer internal. Hasil klasifikasi ditampilkan pada Gambar 9 (a dan b) dan Gambar 10 (a dan b).

Tabel 10. Evaluasi Model indoBERT dengan Metode Hold Out

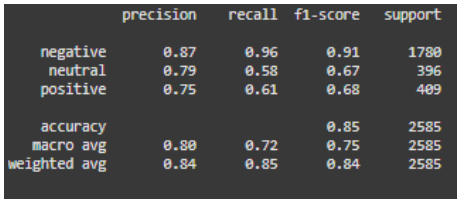
	Negatif (0)	Netral (1)	Positif (2)
Precision	$\frac{1893}{1893+148} = 0{,}93$	$\frac{229}{229+66} = 0{,}78$	$\frac{196}{196+53} 0{,}79$
Recall	$\frac{1893}{1893+70} = 0{,}96$	$\frac{229}{229+70} = 0{,}77$	$\frac{196}{196+127} = 0{,}61$
F1-score	$2 \times \frac{0{,}93 \times 0{,}96}{0{,}93+0{,}96} = 0{,}95$	$2 \times \frac{0{,}78 \times 0{,}77}{0{,}78+0{,}77} = 0{,}77$	$2 \times \frac{0{,}79 \times 0{,}61}{0{,}79+0{,}61} = 0{,}69$

Tabel 11. Evaluasi Model indoBERT dengan Metode K-Fold

	Negatif (0)	Netral (1)	Positif (2)
Precision	$\frac{945}{945+78} = 0{,}92$	$\frac{104}{104+16} = 0{,}87$	$\frac{112}{112+37} 0{,}75$
Recall	$\frac{945}{945+36} = 0{,}96$	$\frac{104}{104+45} = 0{,}70$	$\frac{112}{112+50} = 0{,}69$
F1-score	$2 \times \frac{0{,}92 \times 0{,}96}{0{,}92+0{,}96} = 0{,}94$	$2 \times \frac{0{,}87 \times 0{,}70}{0{,}87+0{,}70} = 0{,}77$	$2 \times \frac{0{,}75 \times 0{,}69}{0{,}75+0{,}69} = 0{,}72$

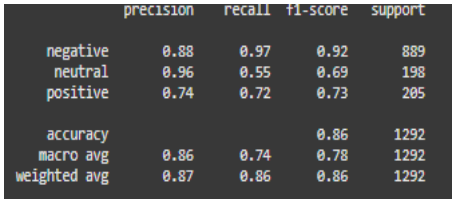
Pada Gambar 9 dan 10, dan tabel 10 dan 11, model IndoBERT menunjukkan akurasi tinggi pada kedua skema validasi, yaitu 90% pada *hold-out* dan 90% pada *k-fold*, dengan performa terbaik pada kelas negatif dan hasil yang seimbang di seluruh metrik evaluasi.

c. *Modeling* CNN



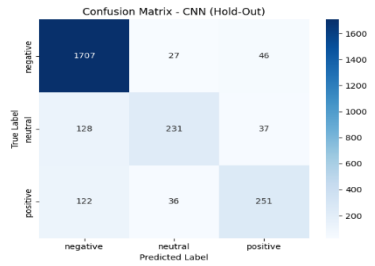
	precision	recall	f1-score	support
negative	0.87	0.96	0.91	1780
neutral	0.79	0.58	0.67	396
positive	0.75	0.61	0.68	409
accuracy			0.85	2585
macro avg	0.80	0.72	0.75	2585
weighted avg	0.84	0.85	0.84	2585

11 (a)

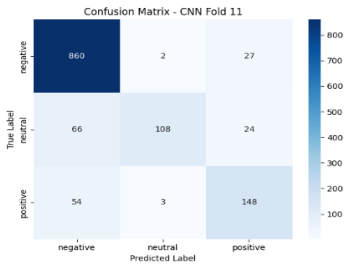


	precision	recall	f1-score	support
negative	0.88	0.97	0.92	889
neutral	0.96	0.55	0.69	198
positive	0.74	0.72	0.73	205
accuracy			0.86	1292
macro avg	0.86	0.74	0.78	1292
weighted avg	0.87	0.86	0.86	1292

12 (a)



11 (b)



12 (b)

Gambar 11. (a dan b) Hasil Model CNN dengan Metode Hold Out

Gambar 12. (a dan b)Hasil Model CNN dengan Metode K-Fold

Model CNN dibangun secara sequential dengan TensorFlow Keras dan dikompilasi menggunakan *sparse categorical crossentropy*, optimizer Adam, dan metrik akurasi pada Gambar 11 (a dan b) dan Gambar 12 (a dan b).

Tabel 12. Evaluasi Model CNN dengan Metode Hold Out

	Negatif (0)	Netral (1)	Positif (2)
Precision	$\frac{1707}{1707+250} = 0{,}87$	$\frac{231}{231+63} = 0{,}79$	$\frac{251}{251+83} 0{,}75$
Recall	$\frac{1707}{1707+73} = 0{,}96$	$\frac{231}{231+165} = 0{,}58$	$\frac{251}{251+158} = 0{,}61$
F1-score	$2 \times \frac{0{,}87 \times 0{,}96}{0{,}87+0{,}96} = 0{,}91$	$2 \times \frac{0{,}79 \times 0{,}58}{0{,}79+0{,}58} = 0{,}67$	$2 \times \frac{0{,}75 \times 0{,}61}{0{,}75+0{,}61} = 0{,}68$

Tabel 13. Evaluasi Model CNN dengan Metode K-Fold

	Negatif (0)	Netral (1)	Positif (2)
Precision	$\frac{860}{860+120} = 0{,}88$	$\frac{108}{108+5} = 0{,}96$	$\frac{148}{148+51} 0{,}74$
Recall	$\frac{860}{860+29} = 0{,}97$	$\frac{108}{108+90} = 0{,}55$	$\frac{148}{148+57} = 0{,}72$
F1-score	$2 \times \frac{0{,}88 \times 0{,}97}{0{,}88+0{,}97} = 0{,}92$	$2 \times \frac{0{,}96 \times 0{,}55}{0{,}96+0{,}55} = 0{,}69$	$2 \times \frac{0{,}74 \times 0{,}72}{0{,}74+0{,}72} = 0{,}73$

Pada penelitian ini, CNN mencapai akurasi 85% dan 86% di tunjukan Tabel 12 dan 13, dengan performa stabil pada kelas negatif dan peningkatan akurasi pada validasi silang.

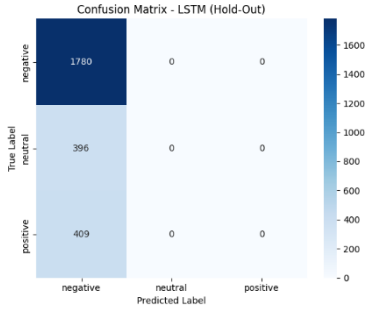
d. *Modeling* LSTM

	precision	recall	f1-score	support
negative	0.69	1.00	0.82	1780
neutral	0.00	0.00	0.00	396
positive	0.00	0.00	0.00	409
accuracy			0.69	2585
macro avg	0.23	0.33	0.27	2585
weighted avg	0.47	0.69	0.56	2585

10 (a)

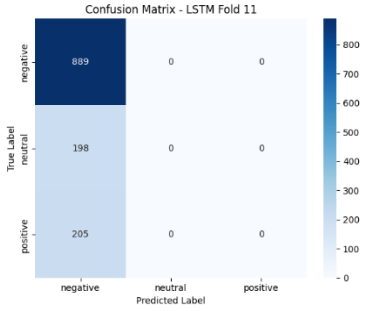
	precision	recall	f1-score	support
negative	0.69	1.00	0.82	889
neutral	0.00	0.00	0.00	198
positive	0.00	0.00	0.00	205
accuracy			0.69	1292
macro avg	0.23	0.33	0.27	1292
weighted avg	0.47	0.69	0.56	1292

11 (a)



10 (b)

Gambar 12. (a dan b) Hasil Model LSTM dengan Metode Hold Out



12 (a)

Gambar 13. (a dan b) Hasil Model LSTM dengan Metode *K-Fold*

Model LSTM dibangun secara sequential dengan TensorFlow Keras menggunakan embedding, LSTM, dan dense layer untuk klasifikasi tiga kelas sentimen. Dikompilasi dengan loss sparse categorical crossentropy dan optimizer Adam, model ini dirancang untuk memahami urutan kata. Hasil klasifikasi ditampilkan pada Gambar 12 (a dan b) dan Gambar 13 (a dan b).

Tabel 14. Evaluasi Model LSTM dengan Metode *Hold Out*

	Negatif (0)	Netral (1)	Positif (2)
Precision	$\frac{1780}{1780+805} = 0{,}69$	$\frac{0}{0+0} = 0{,}00$	$\frac{0}{0+0} 0{,}00$
Recall	$\frac{1780}{1780+0} = 1{,}00$	$\frac{0}{0+396} = 0{,}00$	$\frac{0}{0+409} = 0{,}00$
F1-score	$2 \times \frac{0{,}69 \times 1{,}00}{0{,}69+1{,}00} = 0{,}82$	$2 \times \frac{0{,}00 \times 0{,}00}{0{,}00+0{,}00} = 0{,}00$	$2 \times \frac{0{,}00 \times 0{,}00}{0{,}00+0{,}00} = 0{,}00$

Tabel 15. Evaluasi Model LSTM dengan Metode K-Fold

	Negatif (0)	Netral (1)	Positif (2)
Precision	$\frac{889}{889+403} = 0{,}69$	$\frac{0}{0+0} = 0{,}00$	$\frac{0}{0+0} 0{,}00$
Recall	$\frac{889}{889+0} = 1{,}00$	$\frac{0}{0+198} = 0{,}00$	$\frac{0}{0+205} = 0{,}00$
F1-score	$2 \times \frac{0{,}69 \times 1{,}00}{0{,}69+1{,}00} = 0{,}82$	$2 \times \frac{0{,}00 \times 0{,}00}{0{,}00+0{,}00} = 0{,}00$	$2 \times \frac{0{,}00 \times 0{,}00}{0{,}00+0{,}00} = 0{,}00$

Pada penelitian ini, Tabel 14 dan 15 model LSTM menunjukkan akurasi 69% pada *hold-out* dan 69% pada *k-fold*, dengan performa baik hanya pada kelas negatif dan kegagalan klasifikasi pada kelas netral dan positif di kedua skema validasi.

3.6 *Prediksi Kalimat Pada Setiap Model*

Untuk menguji performa masing-masing model dalam memahami konteks opini publik, digunakan kalimat asli dari data sebagai contoh berikut:
"sebaiknya pertamina menghentikan produk pertamax ron 92 selama setahun sebagai bentuk tanggung jawab menyisakan pertalite dan pertamax turbo ini untuk menghilangkan trauma masyarakat setelah ditipu produk pertamax setelah setahun munculkan produk baru untuk ron 92".

Berikut hasil prediksi sentimen dari masing-masing model adalah:

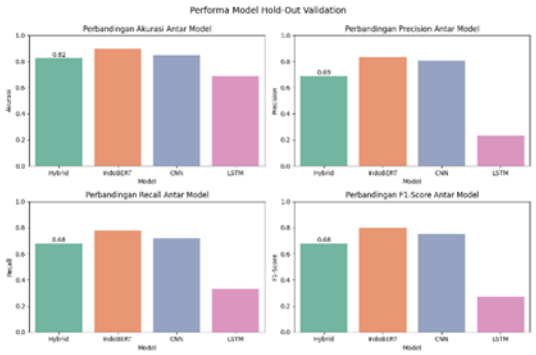
Tabel 16. Uji kalimat Pada Setiap Model

No	Model	Hasil Prediksi Sentimen
1	Algoritma indoBERT	Negatif
2	Algoritma CNN	Negatif
3	Algoritma LSTM	Netral
4	Hybrid Model	Negatif

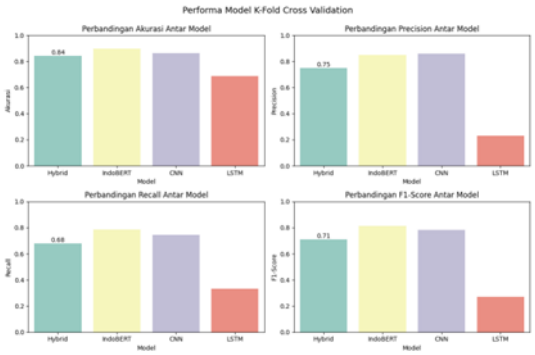
Berdasarkan Tabel 16, tampak bahwa model IndoBERT, CNN, dan hybrid berhasilmenangkap muatan sentimen negatif dari kalimat tersebut. Sebaliknya, model LSTM menghasilkan prediksi netral, menunjukkan keterbatasannya dalam mengenali kritik implisit dalam kalimat panjang dan bersifat saran. Hal ini menguatkan temuan bahwa model berbasis transformer seperti IndoBERT lebih mampu memahami konteks kalimat kompleks dalam bahasa Indonesia.

Performa model *hybrid c*enderung lebih rendah karena arsitektur yang digunakan belum optimal. Integrasi antar komponen belum sepenuhnya mampu menggabungkan kekuatan masing-masing model, sehingga hasil klasifikasinya belum melampaui model individual seperti IndoBERT.

3.7 Evaluasi Model



Gambar 14. Perbandingan Evaluasi antar Model *Hold-Out Validation*



Gambar 15. Perbandingan Evaluasi antar Model *K-Fold Cross Validation*

Evaluasi performa pada gambar 14 dan 15 dilakukan terhadap empat model IndoBERT, *Hybrid*, CNN, dan LSTM menggunakan metrik *Accuracy, Precision, Recall,* dan *F1-Score*. Pengujian dilakukan pada 12.365 *tweet* yang telah diproses, dengan dua metode validasi: *hold-out (80:20) dan K-fold cross validation.*

Evaluasi menggunakan metode *hold-out* menunjukkan bahwa model IndoBERT memiliki kinerja tertinggi, dengan akurasi 90%, presisi 89%, recall 90%, dan F1-score 89. Model CNN mencatat performa cukup baik dengan akurasi 85%, serta nilai *precision* dan *recall* masing-masing 85%, dan F1-score sebesar 84%. Model *Hybrid* mencatat akurasi 82%, *precision* 82%, *recall* 82%, dan *F1-score* 82%. Sementara itu, LSTM hanya mencapai akurasi 69%, dengan *precision*, *recall*, dan *F1-score* masing-masing sebesar 47%.

Pada metode *k-fold cross validation*, IndoBERT kembali unggul dengan akurasi 90%, *precision* 90%, *recall* 90%, dan *F1-score* 90%. CNN meraih akurasi 86%, *precision* 87%, *recall* 86%, dan *F1-score* 86%. *Hybrid* mencapai akurasi 84%, *precision* 83%, *recall* 83%, dan *F1-score* 83%. Sedangkan LSTM tetap memiliki performa paling rendah dengan akurasi 69% dan nilai *precision*, *recall*, serta *F1-score* sebesar 47%.

Visualisasi perbandingan hasil antar model ditampilkan pada Gambar 14 dan 15 menyajikan visualisasi perbandingan model, yang menunjukkan dominasi performa IndoBERT pada seluruh indikator evaluasi. Secara keseluruhan, metode *k-fold validation* menunjukkan hasil yang lebih stabil dan akurat *dibandingkan hold-out validation.*

3. KESIMPULAN

Penelitian ini menunjukkan bahwa model IndoBERT memberikan performa terbaik dalam klasifikasi sentimen dengan akurasi 90%, diikuti CNN (86%), *hybrid* IndoBERT-CNN-LSTM (84%), dan LSTM (69%). Sentimen negatif mendominasi opini publik sebesar 72%, mencerminkan persepsi negatif terhadap kasus korupsi BBM Pertamina. Meskipun pendekatan *hybrid* diterapkan, hasilnya belum melampaui model individual, menunjukkan pentingnya desain arsitektur yang tepat. Penelitian ini menegaskan keunggulan model transformer dalam memahami teks bahasa Indonesia dan peran analisis sentimen dalam menangkap opini publik secara digital.

Untuk pengembangan selanjutnya, disarankan menggunakan metode labeling tambahan seperti *TextBlob* atau *CoreNLP*, serta menerapkan *POS Tagging* untuk meningkatkan akurasi. Model LSTM dapat ditingkatkan dengan pendekatan seperti *Bidirectional LSTM* atau *Attention*, dan struktur *hybrid* perlu dioptimalkan agar benar-benar saling melengkapi.

DAFTAR PUSTAKA

[1] R. D. Pebrianti, "Analisis Sentimen Masyarakat Platform X," *JITET (Jurnal Inform. dan Tek. Elektro Ter.*, vol. 13, no. 2, 2025.

[2] N. Shahnaz, "Penegakan Hukum Terhadap Penyalahgunaan Pengangkutan Dan Niaga Bahan Bakar Minyak (BBM) Bersubsidi," vol. 4, no. 1, pp. 223–236, 2025.

[3] C. H. Lin and U. Nuha, "Sentiment analysis of Indonesian datasets based on a hybrid deep-learning strategy," *J. Big Data*, vol. 10, no. 1, 2023, doi: 10.1186/s40537-023-00782-9.

[4] D. Y. Yefferson, V. Lawijaya, and A. S. Girsang, "Hybrid model: IndoBERT and long short-term memory for detecting Indonesian hoax news," *IAES Int. J. Artif. Intell.*, vol. 13, no. 2, pp. 1911–1922, 2024, doi: 10.11591/ijai.v13.i2.pp1913-1924.

[5] K. Chen, Z. Duan, and S. Yang, "Twitter as research data," *Polit. Life Sci.*, vol. 41, no. 1, pp. 114–130, 2022, doi: 10.1017/pls.2021.19.

[6] S. Kumar, A. K. Kar, and P. V. Ilavarasan, "Applications of text mining in services management: A systematic literature review," *Int. J. Inf. Manag. Data Insights*, vol. 1, no. 1, p. 100008, 2021, doi: 10.1016/j.jjimei.2021.100008.

[7] L. L. Wang and K. Lo, "Text mining approaches for dealing with the rapidly expanding literature on COVID-19," *Brief. Bioinform.*, vol. 22, no. 2, pp. 781–799, 2021, doi: 10.1093/bib/bbaa296.

[8] L. Hickman, S. Thapa, L. Tay, M. Cao, and P. Srinivasan, "Text Preprocessing for Text Mining in Organizational Research: Review and Recommendations," *Organ. Res. Methods*, vol. 25, no. 1, pp. 114–146, 2022, doi: 10.1177/1094428120971683.

[9] A. Jakhotiya, H. Jain, B. Jain, and C. Chaniyara, "Text Pre-Processing Techniques in Natural Language Processing: A Review," *Int. Res. J. Eng. Technol.*, vol. 9, no. 2, pp. 878–880, 2022.

[10] Y. Qi and Z. Shabrina, "Sentiment analysis using Twitter data: a comparative application of lexicon- and machine-learning-based approach," *Soc. Netw. Anal. Min.*, vol. 13, no. 1, pp. 1–14, 2023, doi: 10.1007/s13278-023-01030-x.

[11] R. Catelli, S. Pelosi, and M. Esposito, "Lexicon-Based vs. Bert-Based Sentiment Analysis: A Comparative Study in Italian," *Electron.*, vol. 11, no. 3, 2022, doi: 10.3390/electronics11030374.

[12] J. Tao and X. Fang, "Toward multi-label sentiment analysis: a transfer learning based approach," *J. Big Data*, vol. 7, no. 1, pp. 1–26, 2020, doi: 10.1186/s40537-019-0278-0.

[13] M. A. A. Halim, M. T. A. Rahman, N. A. Rahim, A. Rahman, A. F. A. Hamid, and N. A. M. Amin, "Analysis on current flow style for vehicle alternator fault prediction," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 670, no. 1, 2019, doi: 10.1088/1757-899X/670/1/012042.

[14] M. R. A. Nasution and M. Hayaty, "Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter," *J. Inform.*, vol. 6, no. 2, pp. 226–235, 2019, doi: 10.31311/ji.v6i2.5129.

[15] D. T. Hermanto, A. Setyanto, and E. T. Luthfi, "Algoritma LSTM-CNN untuk Binary Klasifikasi dengan Word2vec pada Media Online," *Creat. Inf. Technol. J.*, vol. 8, no. 1, p. 64, 2021, doi: 10.24076/citec.2021v8i1.264.

[16] S. F. Handayani, R. W. Pratiwi, D. Dairoh, and D. I. Af'idah, "Analisis Sentimen pada Data Ulasan Twitter dengan Long-Short Term Memory," *JTERA (Jurnal Teknol. Rekayasa)*, vol. 7, no. 1, p. 39, 2022, doi: 10.31544/jtera.v7.i1.2022.39-46.

[17] R. A. P. Romadhony, "Identifikasi Similar Question dengan IndoBERT (Studi Kasus," vol. 2, no. 1, pp. 12–17, 2024.

[18] G. Kaur and A. Sharma, "A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis," *J. Big Data*, vol. 10, no. 1, 2023, doi: 10.1186/s40537-022-00680-6.

[19] M. Heydarian, T. E. Doyle, and R. Samavi, "MLCM: Multi-Label Confusion Matrix," *IEEE Access*, vol. 10, pp. 19083–19095, 2022, doi: 10.1109/ACCESS.2022.3151048.

[20] B. W. Rauf, "Sentimen Analisis Pertambangan Di Konawe Utara Dengan Metode Naïve Bayes," *Pros. Semin. Nas. Pemanfaat. Sains dan Teknol. Inf.*, vol. 1, no. 1, pp. 97–102, 2023, [Online]. Available: https://epublikasi.digitallinnovation.com/index.php/sempatin/article/view/98