

BAB II

TINJAUAN PUSTAKA DAN LANDASAN TEORI

2.1 Tinjauan Pustaka

Tinjauan pustaka diambil dari beberapa penelitian sebelumnya yang terkait dengan Analisis Sentimen dengan Analisis VADER. Penelitian sebelumnya dari (Vyas *et al.*, 2021) adalah untuk mengembangkan kerangka kerja otomatis yang dapat mengekstrak sentimen positif, negatif, dan netral dari tweet terkait pandemi COVID-19 dan mengklasifikasikannya melalui teknik pembelajaran mesin. Kerangka kerja yang dikembangkan adalah kerangka kerja hibrida yang menggabungkan teknik berbasis leksikon untuk analisis sentimen tweet dan pelabelan dengan teknik pembelajaran mesin yang diawasi untuk klasifikasi tweet. Tujuan dari penelitian ini adalah untuk meningkatkan pemahaman tentang bagaimana mengklasifikasikan sentimen orang yang berkembang biak di situs microblogging seperti Twitter selama peristiwa mendalam seperti pandemi COVID-19, adopsi blockchain pada usaha kecil dan menengah (UKM), mengidentifikasi faktor kunci dan motivasi untuk penerapannya.

Selanjutnya, Penelitian tersebut mengusulkan *pipeline* kerangka klasifikasi hibrida untuk analisis sentimen otomatis dan klasifikasi tweet yang terkait dengan pandemi COVID-19. Kerangka kerja yang diusulkan terdiri dari tiga tahap utama:

1. Prapengolahan Data: Mulai dari data tweet tekstual mentah, teknik pemrosesan bahasa alami digunakan untuk membersihkan teks dengan menghilangkan redundansi dan anomali. Kata-kata berhenti, karakter khusus, URL, sebutan pengguna, dan hashtag dihapus menggunakan toolkit pemrosesan bahasa alami berbasis Python (NLTK) dan ekspresi reguler (RegEx).
2. Analisis Sentimen: Tahap analisis sentimen melibatkan penggunaan teknik berbasis leksikon yang disebut VADER (*Valence Aware Dictionary and sEntiment Reasoner*) untuk menganalisis dan memberi label sentimen tweet COVID-19 sebagai positif, negatif, atau netral.

3. Klasifikasi: Tahap klasifikasi melibatkan penggunaan teknik pembelajaran mesin yang diawasi untuk mengklasifikasikan tweet berlabel ke dalam empat kategori: Informasi, Opini / Emosi, Berorientasi pada tindakan, dan Lainnya. Validasi silang lima kali lipat digunakan untuk menyetel parameter untuk mendapatkan kinerja terbaik dari Teknik pembelajaran mesin.

Penelitian berikutnya dari (D Abimanyu dkk., 2022) bertujuan untuk menganalisis sentimen pengguna Twitter terhadap game Apex Legends menggunakan algoritma VADER (*Valence Aware Dictionary and sEntiment Reasoner*). Tujuan utamanya adalah untuk mengidentifikasi persepsi dan reaksi komunitas pengguna Twitter terhadap Apex Legends, yang dapat mencakup aspek-aspek seperti kepuasan pemain, kritik, dan saran untuk pengembangan game lebih lanjut. Penelitian ini menggunakan metodologi analisis sentimen dengan mengandalkan VADER, sebuah tool yang dirancang untuk menganalisis teks bahasa Inggris dari media sosial untuk menentukan polaritas (positif, negatif, atau netral) dari ekspresi. Data dikumpulkan dari akun Twitter yang berkaitan dengan Apex Legends. Metode ini memungkinkan peneliti untuk mengolah data besar dari tweet secara otomatis dan efisien, menghasilkan insight tentang sentimen umum terhadap game tersebut. Analisis dilakukan dengan memproses teks tweet, kemudian mengklasifikasikannya ke dalam kategori sentimen berdasarkan skor yang dihasilkan oleh VADER. Hasil penelitian menunjukkan gambaran umum sentimen komunitas Twitter terhadap Apex Legends mencakup distribusi sentimen positif, negatif, dan netral, bersamaan dengan analisis tentang topik atau aspek game yang paling sering dikomentari oleh pengguna. Hasil ini dapat memberikan wawasan berharga bagi pengembang game tentang aspek-aspek apa yang paling diapresiasi oleh pemain, serta area-area yang mungkin memerlukan perbaikan atau penyesuaian.

Penelitian berikutnya dari (Balaji, P., dan Haritha, D. (2023)) menyajikan teknik canggih untuk analisis sentimen pesan singkat, ulasan di media sosial online, dan pesan situs jejaring sosial. Makalah ini memperkenalkan *Ensemble Multi-Layered Sentiment Analysis Model* (EMLSA) yang mengeksplorasi analisis sentimen berbasis kepercayaan untuk memberikan hasil yang akurat pada berbagai dataset

real-time. Penelitian ini menggunakan metode yang dikembangkan dengan berbagai model canggih seperti model DL (*Deep Learning*) prapelatihan model *stemming* untuk prapengolahan data, TF-IDF (*Term Frequency-Inverse Document Frequency*) untuk ekstraksi fitur, WLE (*Word-Level Embeddings*) untuk analisis teks, VADER untuk pelatihan dan RNN untuk klasifikasi data teks. Model EMLSA adalah pendekatan gabungan dengan VADER (*Valence Aware Dictionary and Sentiment Reasoned*) dan *Recurrent Neural Networks* (RNNs). Penelitian ini juga mencakup diagram arsitektur sistem yang menunjukkan proses implementasi langkah demi langkah.

Tinjauan pustaka diambil dari beberapa penelitian sebelumnya yang membahas artikel-artikel terbaru tentang temuan-temuan penting metode VADER dan BI-LSTM. Studi terbaru tentang beberapa penelitian sebelumnya disajikan pada Tabel 2.1.

Tabel 2.1 Studi tentang Analisis Sentimen, VADER dan LSTM

No	Key words	Judul	Tujuan	Metodologi Penelitian	Hasil
1	Analisis Senti men	Kavitha, G., Saveen, B., & Imtiaz, N. (2018). <i>Discovering Public Opinions by Performing Sentimental Analysis on Real Time Twitter Data</i> . 2018 International Conference on Circuits and Systems in Digital Enterprise Technology, ICCSDET 2018, 1-4. https://doi.org/10.1109/ICCSDET.2018.8821105	Mengusulkan metodologi untuk mengambil tweet real-time dan melakukan analisis sentimen tanpa memerlukan profesional yang terampil dan pertanyaan yang rumit. rendah dan akurasi tinggi	<ol style="list-style-type: none"> 1. Mengambil tweet real-time dari Twitter menggunakan streaming. 2. Preprocessing tweet dengan menghapus kata-kata berhenti, URL, dan karakter khusus. 3. Menerapkan pendekatan analisis sentimen untuk menentukan polaritas setiap tweet (positif, negatif, atau netral). 4. Mengkategorikan hasil output berdasarkan polaritas setiap tweet. 5. Memvisualisasikan aliran data langsung dan menghasilkan hasil 	<p>Membahas implementasi dan hasil yang diperoleh. Penelitian ini menganalisis data Twitter yang terkait dengan pertandingan IPL dan para pemainnya untuk menemukan opini dan peringkat publik. Analisis menentukan bahwa 88% orang men-tweet secara netral, 8% positif, dan 4% negatif tentang IPL. Representasi grafis</p>

Tabel 2.1 Studi tentang Analisis Sentimen, VADER dan LSTM(Lanjutan)

No	Key words	Judul	Tujuan	Metodologi Penelitian	Hasil
		8.8821105	dikategorikan, yang menentukan opini publik untuk setiap penerapan topik yang meledak dan trendi dengan biaya rendah dan akurasi tinggi	keluaran yang dikategorikan.	dari analisis tweet memberikan wawasan yang berguna tentang topik yang meledak tentang tren saat ini yang bergerak viral di masyarakat. Oleh karena itu, makalah ini menyimpulkan bahwa analisis Twitter menentukan tren dan puncak dari data Twitter real-time dan memberikan implikasi tentang opini publik di domain apa pun
2	Analisis Senti men	Yuxing Qi1 · Zahratu Shabrina (2023), <i>Sentiment Analysis using Twitter data: a comparative Application of lexicon- and machine-learning-based Approach” Social Network Analysis and Mining</i> (2023) 13:31 https://doi.org/10.1007/s13278-023-01030-x	Penelitian ini bertujuan untuk membandingkan efektivitas dua pendekatan yang berbeda dalam analisis sentimen data Twitter: pendekatan berbasis leksikon dan pendekatan berbasis pembelajaran mesin. Tujuannya adalah untuk menentukan	Dua metodologi utama digunakan dalam penelitian ini. Pertama, pendekatan berbasis leksikon yang melibatkan penggunaan kamus kata-kata yang telah ditandai dengan polaritas sentimen (positif, negatif, atau netral). Kedua, pendekatan berbasis pembelajaran mesin yang menggunakan algoritma klasifikasi, seperti Support Vector Machines (SVM) atau	Menunjukkan bahwa salah satu metode memiliki keunggulan dalam kondisi tertentu atau bahwa kombinasi dari kedua metode memberikan hasil yang lebih baik. Penelitian ini juga mungkin mengidentifikasi kekuatan dan keterbatasan dari masing-masing pendekatan dalam

Tabel 2.1 Studi tentang Analisis Sentimen, VADER dan LSTM(Lanjutan)

No	Key words	Judul	Tujuan	Metodologi Penelitian	Hasil
			<p>pendekatan mana yang lebih akurat dalam mengklasifikasikan sentimen dari tweet, sehingga dapat memberikan wawasan yang lebih baik dalam pemahaman opini publik yang diekspresikan di media sosial.</p>	<p>Neural Networks, yang telah dilatih pada dataset yang terdiri dari tweet dengan sentimen yang telah ditandai. Kedua pendekatan ini diterapkan pada dataset Twitter yang sama dan hasilnya dibandingkan</p>	<p>sentimen data Twit</p>
3	VADER	<p><i>A Complete VADER-Based Sentiment Analysis of Bitcoin (BTC) Tweets during the Era of COVID-19</i>, Toni Pano and Rasha Kash</p>	<p>Tujuan dari penelitian ini adalah untuk mengembangkan strategi pra pengolahan data teks yang berbeda untuk mengkorelasikan skor sentimen dari teks Twitter dengan harga Bitcoin selama pandemi COVID-19.</p>	<p>Peneliti menggunakan metode analisis sentimen dan analisis korelasi untuk mengkorelasikan skor sentimen dari teks Twitter dengan harga Bitcoin selama pandemi COVID-19. Peneliti juga menggunakan beberapa teknik prepengolahan teks yang berbeda untuk memproses data Twitter sebelum melakukan analisis korelasi. Selain itu, penulis juga menggunakan model regresi vektor otomatis (VAR) untuk memprediksi harga penutupan Bitcoin.</p>	<p>Hasil dari penelitian ini menunjukkan bahwa terdapat korelasi antara skor sentimen dari teks Twitter dan harga Bitcoin selama pandemi COVID-19. Namun, hasil korelasi yang paling kuat terjadi ketika data Twitter diproses dalam jangka waktu satu hari, yang memiliki polaritas korelasi yang sulit diprediksi.</p>

Tabel 2.1 Studi tentang Analisis Sentimen, VADER dan LSTM(Lanjutan)

No	Key words	Judul	Tujuan	Metodologi Penelitian	Hasil
4	VADER	Balaji, P., & Haritha, D. (2023). <i>An Ensemble Multi-layered Sentiment Analysis Model (EMLSA) for Classifying the Complex Datasets</i> . <i>International Journal of Advanced Computer Science and Applications</i> , 14(3), 185–190. https://doi.org/10.14569/IJACSA.2023.0140320	Menyajikan teknik canggih untuk analisis sentimen pesan singkat, ulasan di media sosial online, dan pesan situs jejaring sosial. Makalah ini memperkenalkan <i>Ensemble Multi-Layered Sentiment Analysis Model (EMLSA)</i> yang mengeksplorasi analisis sentimen berbasis kepercayaan untuk memberikan hasil yang akurat pada berbagai dataset real-time.	Penelitian ini dikembangkan dengan berbagai model canggih seperti model DL pra-terlatih model stemming untuk pra-pemrosesan, TF-IDF untuk ekstraksi fitur, <i>Word-Level Embeddings (WLE)</i> untuk analisis teks, VADER untuk pelatihan dan RNN untuk klasifikasi data teks. Model EMLSA adalah pendekatan gabungan dengan VADER (<i>Valence Aware Dictionary and Sentiment Reasoner</i>) dan Recurrent Neural Networks (RNNs).	Model EMLSA yang diusulkan mencapai akurasi tinggi dalam analisis sentimen untuk berbagai dataset benchmark. Kesimpulannya pendekatan yang diusulkan dapat memproses kumpulan data yang kompleks berdasarkan ulasan pengguna dan mengekstrak setiap aspek tinjauan input dengan kinerja tinggi menggunakan kombinasi model penyematan kata TF-IDF dan Word2Vec dengan model DL.

Tabel 2.1 Studi tentang Analisis Sentimen, VADER dan LSTM(Lanjutan)

No	Key words	Judul	Tujuan	Metodologi Penelitian	Hasil
5	VADER dan LSTM	Vyas, P., Reisslein, M., Rimal, B. P., Vyas, G., Basyal, G. P., & Muzumdar, P. (2021). <i>Automated Classification of Societal Sentiments on Twitter With Machine Learning</i> . IEEE Transactions on Technology and Society, 3(2), 100–110. https://doi.org/10.1109/tts.2021.3108963	Tujuan dari Penelitian ini adalah untuk mengembangkan kerangka kerja otomatis yang dapat mengekstrak sentimen positif, negatif, dan netral dari tweet terkait pandemi COVID-19 dan mengklasifikasikannya melalui teknik pembelajaran mesin. Kerangka kerja yang dikembangkan adalah kerangka kerja hibrida yang menggabungkan teknik berbasis leksikon untuk analisis sentimen tweet dan pelabelan dengan teknik pembelajaran mesin yang diawasi untuk klasifikasi tweet.	Penelitian ini mengusulkan pipeline kerangka klasifikasi hibrida untuk analisis sentimen otomatis dan klasifikasi tweet yang terkait dengan pandemi COVID-19. Kerangka kerja yang diusulkan terdiri dari tiga tahap utama: Pra-pemrosesan Data: Mulai dari data tweet tekstual mentah kemudian Tahap analisis sentimen melibatkan penggunaan teknik berbasis leksikon yang disebut VADER untuk menganalisis dan memberi label sentimen tweet COVID-19 sebagai positif, negatif, atau netral. Terakhir Tahap klasifikasi melibatkan penggunaan teknik pembelajaran mesin yang diawasi untuk mengklasifikasikan tweet berlabel ke dalam empat kategori: Informasi, Opini / Emosi, Berorientasi pada tindakan, dan Lainnya.	Penelitian ini melaporkan bahwa pipeline kerangka kerja hibrida yang diusulkan mencapai hasil yang menjanjikan dalam hal analisis sentimen dan klasifikasi tweet terkait pandemi COVID-19. Teknik LSTM mengungguli semua teknik pembelajaran mesin lainnya, mencapai akurasi tes 83% dan skor recall 83%. Makalah ini menyimpulkan bahwa pipeline kerangka kerja hybrid yang diusulkan dapat digunakan sebagai alat yang berharga untuk menganalisis sentimen masyarakat di Twitter terkait pandemi COVID-19

Tabel 2.1 Studi tentang Analisis Sentimen, VADER dan LSTM (Lanjutan)

No	Key words	Judul	Tujuan	Metodologi Penelitian	Hasil
6	VADER	<p>Analisis Sentimen Akun Twitter Apex Legends Menggunakan VADER, Dicky Abimanyu1, Elvia Budianita2, Eka Pandu Cynthia3, Febi Yanto4, Yusra5</p> <p>Program Studi Teknik Informatika, Fakultas Sains Dan Teknologi1,2,3,4,5, Universitas Islam Negeri Sultan Syarif Qasim Riau</p>	<p>Penelitian ini bertujuan untuk menganalisis sentimen pengguna Twitter terhadap game Apex Legends menggunakan algoritma VADER (<i>Valence Aware Dictionary and sEntiment Reasoner</i>). Tujuan utamanya adalah untuk mengidentifikasi persepsi dan reaksi komunitas pengguna Twitter terhadap Apex Legends, yang dapat mencakup aspek-aspek seperti kepuasan pemain, kritik, dan saran untuk pengembangan game lebih lanjut.</p>	<p>Penelitian ini menggunakan metodologi analisis sentimen dengan mengandalkan VADER, sebuah tool yang dirancang untuk menganalisis teks bahasa Inggris dari media sosial untuk menentukan polaritas (positif, negatif, atau netral) dari ekspresi. Data dikumpulkan dari akun Twitter yang berkaitan dengan Apex Legends. Metode ini memungkinkan peneliti untuk mengolah data besar dari tweet secara otomatis dan efisien, menghasilkan insight tentang sentimen umum terhadap game tersebut.</p>	<p>Hasil penelitian menunjukkan gambaran umum sentimen komunitas Twitter terhadap Apex Legends mencakup distribusi sentimen positif, negatif, dan netral, bersamaan dengan analisis tentang topik atau aspek game yang paling sering dikomentari oleh pengguna. Hasil ini dapat memberikan wawasan berharga bagi pengembang game tentang aspek-aspek apa yang paling diapresiasi oleh pemain, serta area-area yang mungkin memerlukan perbaikan atau penyesuaian.</p>

SEKOLAH PASCASARJANA

Tabel 2.1 Studi tentang Analisis Sentimen, VADER dan LSTM (Lanjutan)

No	Key words	Judul	Tujuan	Metodologi Penelitian	Hasil
7	VADER	Itani, M., Roast, C., & Al-Khayatt, S. (2017). <i>Developing Resources for Sentiment Analysis of Informal Arabic Text in Social Media</i> . <i>Procedia Computer Science</i> , 117, 129–136. https://doi.org/10.1016/j.procs.2017.10.101	Tujuan dari makalah ini adalah untuk menggambarkan pengembangan sumber daya untuk analisis sentimen khusus untuk teks Arab di media sosial. Ini berfokus pada membangun korporat dan leksikon media sosial di DA, menggunakan Facebook sebagai sumber, dan membentuk leksikon korporat dan sentimen beranotasi. Makalah ini juga meninjau dan membahas hasil dari pekerjaan ini.	Makalah ini mengusulkan pendekatan baru untuk mengembangkan sumber daya untuk analisis sentimen teks Arab informal di media sosial. Pendekatan ini melibatkan membangun korporat dan leksikon media sosial di DA, menggunakan Facebook sebagai sumber, dan membentuk korpora beranotasi dan leksikon sentimen. Makalah ini juga meninjau dan membahas hasil dari pekerjaan ini. Oleh karena itu, metode yang digunakan dalam tulisan ini adalah kombinasi dari pengumpulan data, anotasi, dan analisis.	Penelitian ini menyajikan pengembangan sumber daya untuk analisis sentimen teks Arab informal di media sosial, khususnya menggunakan Facebook sebagai sumber. Para penulis membangun korpora dan leksikon media sosial, membentuk korpora beranotasi dan leksikon sentimen, dan meninjau serta mendiskusikan hasil pekerjaan ini. Penelitian ini menunjukkan bahwa meskipun domain berbeda, orang mungkin berharap untuk menemukan hasil yang sebanding ketika berhadapan dengan platform media sosial lainnya seperti Twitter atau Instagram.

2.2 Perbandingan dengan Studi Sebelumnya

Hasil penelitian ini dibandingkan dengan studi yang dilakukan oleh Mohamed Chiny et al. (2021) yang menggunakan pendekatan hibrid antara LSTM, VADER, dan TF-IDF. Hasil penelitian ini menunjukkan bahwa kombinasi VADER dan Bi-LSTM memberikan peningkatan akurasi yang signifikan tanpa memerlukan teknik tambahan seperti TF-IDF.

Penelitian ini memberikan kontribusi yang signifikan dalam bidang Sistem Informasi dan Analisis Sentimen. Secara teoretis, penelitian ini memperkaya literatur dengan mengadaptasi metode analisis sentimen yang lebih kompleks untuk konteks industri energi BUMN. Secara praktis, hasil ini memberikan panduan bagi perusahaan BUMN untuk:

- a. Meningkatkan komunikasi keberlanjutan dalam kampanye pemasaran.
- b. Mengadopsi teknologi analisis sentimen untuk memahami opini publik secara *real-time*.

2.3 Perbandingan dengan Literatur Tambahan

Berikut penelitian yang telah dilakukan sebelumnya:

Tabel 2.2 Perbandingan dengan literatur tambahan

Penelitian	Metode Utama	Temuan Utama	Referensi
Vyas et al. (2021)	VADER + LSTM	Akurasi 83%	Tersedia
Dicky Abimanyu et al. (2022)	VADER	Sentimen Twitter Apex Legends	Tersedia
Balaji & Haritha (2023)	VADER + RNN	Model EMLSA akurasi tinggi	Tersedia
Toni Pano et al. (2021)	VADER	Korelasi Bitcoin dan sentimen media	Tersedia

Berikut adalah penjelasan tabel perbandingan dengan penelitian sebelumnya, dengan gaya penulisan yang ilmiah, kronologis, dan mudah dipahami:

2.4 Analisis Perbandingan dengan Penelitian Terdahulu

Tabel 2.2 Perbandingan dengan literatur tambahan menyajikan perbandingan antara penelitian ini dengan beberapa penelitian terdahulu yang relevan. Perbandingan ini bertujuan untuk menunjukkan posisi dan kontribusi penelitian ini dalam bidang analisis sentimen, khususnya dengan pemanfaatan metode *lexicon-based* seperti VADER.

a. Studi Vyas et al. (2021)

Vyas et al. (2021) telah menggunakan kombinasi VADER dengan *Long Short-Term Memory* (BI-LSTM) untuk analisis sentimen. Penelitian ini berhasil mencapai akurasi sebesar 83% dalam klasifikasi sentimen. Hasil ini menunjukkan potensi penggabungan metode *lexicon-based* dengan pendekatan *deep learning* seperti BI-LSTM.

b. Studi Dicky Abimanyu et al. (2022)

Dicky Abimanyu et al. (2022) memfokuskan penelitiannya pada analisis sentimen terhadap gim Apex Legends di platform Twitter menggunakan VADER. Penelitian ini berkontribusi dalam penerapan VADER pada kasus-kasus spesifik di media sosial, seperti analisis sentimen terhadap produk atau *brand*.

c. Studi Balaji & Haritha (2023)

Balaji & Haritha (2023) mengembangkan model *Ensemble Multi-Level Sentiment Analysis* (EMLSA) dengan menggabungkan VADER dan *Recurrent Neural Network* (RNN). Model EMLSA yang diusulkan menunjukkan akurasi yang tinggi dalam analisis sentimen. Penelitian ini mengindikasikan bahwa integrasi VADER dengan RNN dapat meningkatkan performa analisis sentimen.

d. Studi Toni Pano et al. (2021)

Toni Pano et al. (2021) meneliti korelasi antara harga Bitcoin dan sentimen media sosial yang dianalisis menggunakan VADER. Hasil penelitian ini menunjukkan adanya hubungan antara sentimen publik di media sosial dengan fluktuasi harga Bitcoin. Studi ini memperluas aplikasi VADER dalam bidang ekonomi dan keuangan.

e. Tabel Penelitian terdahulu (*State of the Art*)

Penelitian terdahulu yang sudah dilakukan dan research gap penelitian dapat dilihat pada tabel 2.3.

Tabel 2.3 Penelitian terdahulu dan Research Gap

Peneliti/ Tahun	Objek Penelitian	Metode/ Algoritma	Hasil/Akurasi	Research Gap yang Diisi oleh Penelitian Promovendus
Elouardi ghi et al. (2017)	Komentar Facebook (Bahasa Arab Standar & Dialek)	Naive Bayes, Random Forest, SVM + TF/TF- IDF	Akurasi tertinggi ~78% (SVM dengan TF- IDF)	Bahasa & Domain: Bahasa Arab, politik. Metode: ML tradisional. Promovendus fokus pada Bahasa Indonesia, domain BUMN energi, dengan hybrid deep learning.
Asri et al. (2022)	Ulasan Aplikasi PLN Mobile	VADER + Naive Bayes	Akurasi 70%	Model: VADER hanya untuk labeling, klasifikasi sederhana. Promovendus integrasikan VADER dengan IndoBERT & Bi-LSTM

Tabel 2. 3 Penelitian terdahulu dan Research Gap (lanjutan)

Peneliti/ Tahun	Objek Penelitian	Metode/Algoritma	Hasil/Akurasi	Research Gap yang Diisi oleh Penelitian Promovendus
Nurhayati et al. (2023)	Analisis sentimen terhadap produk energi terbarukan di Indonesia	Ensemble Model (SVM + Random Forest + Naïve Bayes)	Presisi 87.9%, Akurasi 86.2%	Deep Learning vs Traditional ML: Menggunakan arsitektur DL canggih (Bi-LSTM dengan IndoBERT) yang mampu menangkap pola kompleks dan dependensi jangka panjang dalam teks. Lexicon Integration: Mengintegrasikan VADER untuk menangkap nuansa bahasa informal media sosial yang mungkin terlewat oleh model berbasis transformer
Gu et al. (2020)	Review Produk & Hotel (Bahasa Tionghoa)	MBGC V (BiGRU + CNN + VIB)	Akurasi hingga 94%	Arsitektur: Multichannel CNN+BiGRU. Promovendus gunakan BiLSTM + IndoBERT + VADER untuk konteks bahasa Indonesia & energi.

SEKOLAH PASCASARJANA

Tabel 2. 3 Penelitian terdahulu dan Research Gap (lanjutan)

Peneliti/ Tahun	Objek Penelitian	Metode/Algoritma	Hasil/Akurasi	Research Gap yang Diisi oleh Penelitian Promovendus
Febrianti & Chen (2024)	Analisis sentimen kampanye pemasaran green energy di Asia Tenggara	BERT (Multilingual BERT)	F1- Score 90.2 % untuk data Indonesia	Optimasi Bahasa Indonesia: Menggunakan Indo BERT yang secara khusus dioptimalkan untuk bahasa Indonesia, bukan model multilingual umum. Hybrid untuk Nuansa: Menggabungkan

Posisi dan Kontribusi Penelitian Promovendus:

Dari analisis *state of the art* di atas, penelitian ini menempati posisi yang unik dan advance dengan kontribusi berikut:

1. Trilogi Metodologi yang Sinergis: Ini menggabungkan tiga paradigma berbeda:
 - VADER: Aturan leksikon untuk deteksi sentimen eksplisit dan bahasa informal
 - IndoBERT: Transformer model untuk pemahaman kontekstual mendalam
 - Bi-LSTM: Jaringan saraf untuk menangkap dependensi sequential dua arah
2. Optimasi Spesifik Domain: Penelitian ini secara khusus:
 - Menggunakan IndoBERT yang dioptimalkan untuk bahasa Indonesia
 - Fokus pada anak perusahaan BUMN bidang energi (level operasional pemasaran)
 - Menganalisis konten pemasaran produk (bukan layanan atau isu korporat)
3. Arsitektur Hybrid yang Inovatif: Kombinasi VADER-IndoBERT-Bi-LSTM merupakan pendekatan hybrid yang belum banyak diterapkan dalam analisis sentimen pemasaran produk energi di Indonesia, terutama untuk data media sosial multi-platform.
4. Validasi Metodologi Komparatif: Dengan membandingkan ketiga metode tersebut, penelitian promovendus dapat memberikan insight tentang:

- Performa relatif masing-masing pendekatan
 - Sinergi optimal antara metode berbasis aturan dan deep learning
 - Efektivitas model untuk bahasa informal media sosial Indonesia
5. Kontribusi Praktis untuk Industri: Hasil penelitian dapat langsung dimanfaatkan oleh:
- Tim pemasaran anak perusahaan BUMN energi
 - Agency digital marketing di sektor energi
 - Regulator yang memantau komunikasi pemasaran BUMN

Penelitian ini mengisi celah antara teknikalitas NLP canggih dan aplikasi praktis di industri energi BUMN, dengan metodologi yang lebih komprehensif dibanding penelitian-penelitian sejenis yang ada

2.4 Landasan Teori

Analisis sentimen melibatkan berbagai teknik dan metode yang dapat digunakan untuk mengidentifikasi dan menganalisis sentimen dari data tekstual. Berikut adalah beberapa teknik dan metode yang umum digunakan dalam analisis sentimen.

2.4.1 Metode VADER

Metode VADER (*Valence Aware Dictionary and Sentiment Reasoner*) adalah metode analisis sentimen berbasis *lexicon* yang dikembangkan oleh (Hutto dan Gilbert, 2014). Metode ini menggunakan daftar kata dan frase untuk menghitung polaritas (positif, negatif, atau netral) dari teks. Daftar kata dan frase ini dilengkapi dengan bobot untuk mengukur kekuatan sentimen.

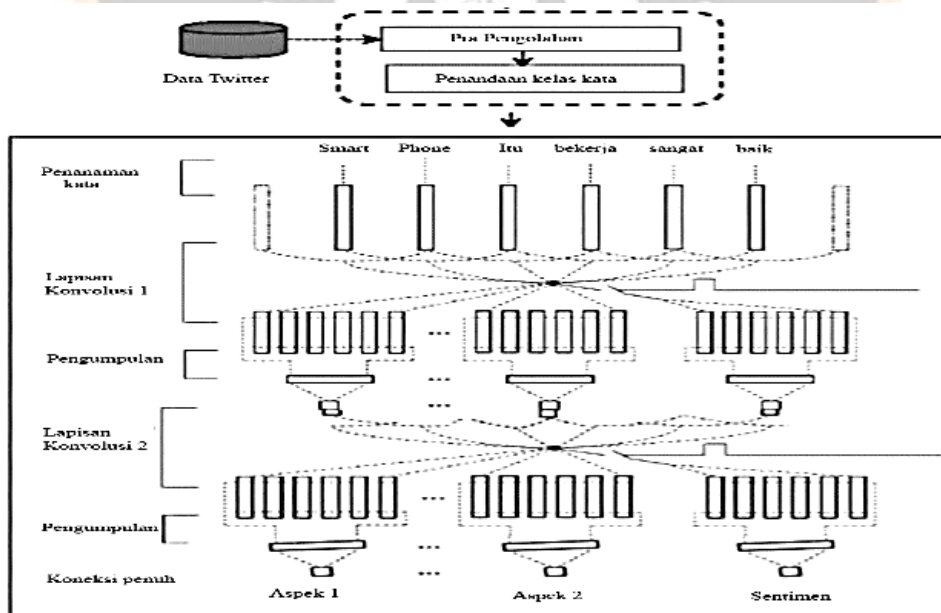
Metode VADER memiliki beberapa kelebihan, antara lain:

1. Mudah digunakan dan diimplementasikan
2. Teliti dan akurat
3. Bersifat umum

Metode VADER telah banyak digunakan untuk berbagai aplikasi analisis sentimen, termasuk analisis ulasan produk, analisis media sosial, dan analisis opini publik.

2.4.2 Metode berbasis aturan

Metode berbasis aturan adalah pendekatan dalam analisis sentimen yang menggunakan aturan atau aturan heuristik yang telah ditentukan sebelumnya. Aturan-aturan ini digunakan untuk mengklasifikasikan teks berdasarkan sentimen yang terkandung di dalamnya (Amalia *et al.*, 2018) Metode berbasis aturan adalah salah satu jenis metode dalam analisis data yang menggunakan aturan-aturan atau kaidah-kaidah yang telah ditentukan sebelumnya untuk mengklasifikasikan atau mengambil keputusan terhadap data baru. Metode ini didasarkan pada pendekatan logika, di mana aturan-aturan yang telah ditentukan digunakan untuk menginterpretasikan pola-pola dalam data. Dalam metode berbasis aturan, aturan-aturan ini dapat berupa aturan logika sederhana atau berupa rangkaian kondisi dan tindakan yang harus diikuti. Aturan-aturan ini dapat diperoleh melalui pengetahuan domain, pengalaman praktis, atau dengan menggunakan teknik pembelajaran aturan dari data yang ada. Penerapan metode berbasis aturan dapat melibatkan proses penggalian data (*data mining*) untuk menemukan aturan-aturan yang relevan dan signifikan dari data. Aliran informasi Metode Berbasis Aturan dapat dilihat pada Gambar 2.1.



Gambar 2.1 Metode Berbasis Aturan

Simber: (Ray dan Chakrabarti, 2022)

Aturan-aturan ini kemudian dapat digunakan untuk mengklasifikasikan atau mengambil keputusan terhadap data baru berdasarkan kondisi yang ada (Steenwinckel *et al.*, 2021). Metode berbasis aturan cenderung lebih transparan dan dapat diinterpretasikan, namun memerlukan upaya yang lebih besar dalam pembuatan aturan. Metode berbasis aturan dalam analisis sentimen cenderung lebih transparan dan dapat diinterpretasikan. Metode ini melibatkan pembuatan aturan atau aturan heuristik yang digunakan untuk mengklasifikasikan teks berdasarkan sentimen yang terkandung di dalamnya. Aturan-aturan ini dibuat berdasarkan pengetahuan dan pemahaman manusia tentang bahasa dan sentimen (Yadav *et al.*, 2022).

Keuntungan utama dari metode berbasis aturan adalah transparansinya. Karena aturan-aturan yang digunakan dalam analisis sentimen ditentukan secara eksplisit, hasil analisis dapat dijelaskan dan dipahami oleh manusia. Hal ini memungkinkan interpretasi dan validasi oleh pengguna atau ahli domain yang terkait. Namun, metode berbasis aturan juga memiliki beberapa keterbatasan. Salah satunya adalah upaya yang lebih besar dalam pembuatan aturan. Membuat aturan yang akurat dan mencakup berbagai kasus dapat memerlukan waktu dan pengetahuan yang mendalam tentang domain yang sedang dianalisis. Pembuatan aturan yang lengkap dan terperinci dapat menjadi tugas yang kompleks dan memakan waktu. Selain itu, metode berbasis aturan mungkin kurang fleksibel dalam menangani variasi dan kompleksitas dalam bahasa dan sentimen manusia. Setiap perubahan dalam bahasa atau konteks dapat memerlukan pembuatan ulang atau penyesuaian aturan yang ada. Metode ini juga mungkin tidak dapat mengatasi ambiguitas atau pemahaman konteks yang lebih kompleks dalam bahasa manusia (Krishnamoorthy, 2018).

2.4.3 Metode berbasis klasifikasi

Metode berbasis klasifikasi adalah pendekatan dalam analisis sentimen yang menggunakan algoritma pembelajaran mesin untuk melakukan klasifikasi otomatis terhadap teks berdasarkan sentimen yang terkandung di dalamnya. Metode berbasis klasifikasi adalah salah satu jenis metode dalam pembelajaran mesin yang digunakan untuk mengklasifikasikan data ke dalam kategori atau kelas yang telah

ditentukan sebelumnya. Tujuan utama dari metode ini adalah untuk mengidentifikasi pola-pola dalam data yang memungkinkan pengklasifikasian yang tepat (Charbuty dan Abdulazeez, 2021).

Metode berbasis klasifikasi melibatkan pembuatan model atau classifier yang dapat mempelajari hubungan antara fitur-fitur yang ada dalam data dengan label kelas yang telah ditentukan. Fitur-fitur ini dapat berupa atribut-atribut numerik, kategorikal, atau bahkan teks. Model tersebut akan menggunakan pola-pola yang telah dipelajari dari data latihan untuk mengklasifikasikan data baru yang belum diberi label ke dalam kelas yang sesuai. Beberapa contoh metode berbasis klasifikasi yang umum digunakan adalah *Naive Bayes*, *Decision Tree*, *Random Forest*, *Support Vector Machines (SVM)*, dan *k-Nearest Neighbors (k-NN)*. Setiap metode memiliki kelebihan dan kelemahan masing-masing tergantung pada karakteristik data dan tujuan analisis (Vali *et al.*, 2020).

Penerapan metode berbasis klasifikasi dalam analisis sentimen dapat dilakukan dengan menggunakan data teks, di mana teks diubah menjadi fitur-fitur yang dapat digunakan untuk mengklasifikasikan sentimen (positif, negatif, atau netral) dari teks tersebut (Zhang *et al.*, 2018). Model yang telah dilatih menggunakan data latihan dengan sentimen yang telah diketahui dapat digunakan untuk mengklasifikasikan sentimen pada data teks baru. Metode berbasis klasifikasi sangat berguna dalam berbagai aplikasi seperti analisis sentimen, klasifikasi dokumen, deteksi spam, pengenalan pola, dan banyak lagi. Dengan kemampuan mereka dalam mengklasifikasikan data secara otomatis, metode ini memungkinkan pengambilan keputusan yang lebih efisien dan akurat dalam berbagai domain (Pratiwi dan Adiwijaya, 2018).

Metode ini memanfaatkan teknik pembelajaran mesin untuk mempelajari pola-pola yang terdapat dalam *dataset* pelatihan, dan kemudian menerapkan pola-pola tersebut untuk mengklasifikasikan teks yang belum dikenal (Veluchamy *et al.*, 2018). Metode berbasis klasifikasi memiliki keunggulan-keunggulan berikut:

1. Klasifikasi otomatis

Metode ini dapat melakukan klasifikasi sentimen secara otomatis tanpa intervensi manusia. Algoritma pembelajaran mesin dapat mempelajari pola-

pola sentimen yang kompleks dan membuat prediksi secara cepat dan efisien.

2. Akurasi yang tinggi

Dengan menggunakan teknik pembelajaran mesin yang canggih, metode berbasis klasifikasi dapat mencapai tingkat akurasi yang tinggi dalam mengklasifikasikan sentimen. Algoritma pembelajaran mesin dapat mempelajari fitur-fitur penting yang mendukung pengklasifikasian yang akurat.

3. Skalabilitas

Metode berbasis klasifikasi dapat diterapkan pada *dataset* yang besar dengan efisiensi yang baik. Algoritma pembelajaran mesin dapat dipelajari dengan menggunakan *dataset* pelatihan yang luas, sehingga mampu menangani jumlah data yang besar.

Namun, metode berbasis klasifikasi juga memiliki beberapa kelemahan, seperti:

1. Ketergantungan pada *dataset* pelatihan: Metode ini memerlukan *dataset* pelatihan yang representatif dan berkualitas tinggi untuk memperoleh hasil klasifikasi yang akurat. Proses pengumpulan dan anotasi *dataset* pelatihan bisa memerlukan waktu dan upaya yang besar.
2. Kurangnya interoperabilitas: Meskipun metode berbasis klasifikasi dapat memberikan hasil klasifikasi yang akurat, interpretasi dan pemahaman terhadap sentimen yang diekstrak mungkin sulit dilakukan. Algoritma pembelajaran mesin cenderung kompleks dan sulit untuk diinterpretasikan secara langsung oleh manusia.

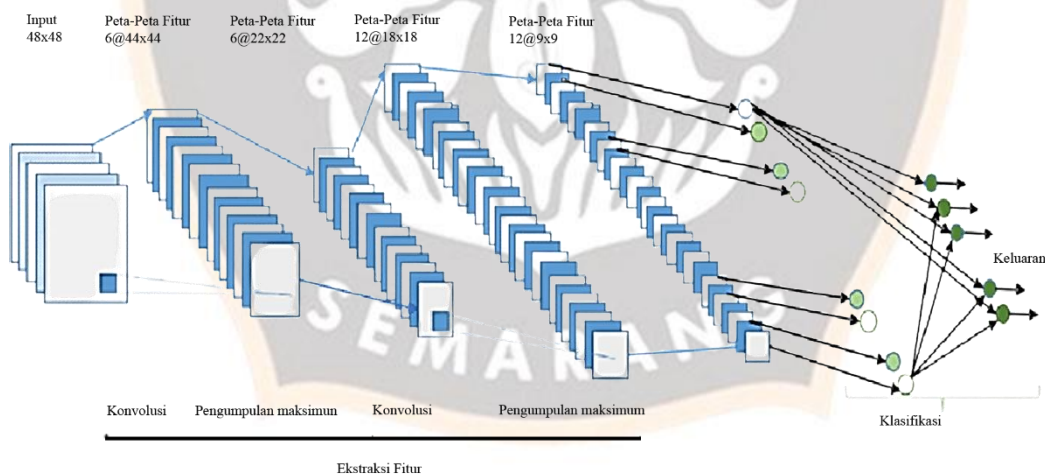
2.4.4 Metode berbasis pembelajaran mesin

Metode berbasis pembelajaran mesin adalah pendekatan dalam analisis sentimen yang menggunakan teknik-teknik dari bidang pembelajaran mesin untuk mengolah dan menganalisis data tekstual (Mittal *et al.*, 2019). Metode berbasis pembelajaran mesin (*Machine Learning*) digunakan untuk memproses dan menganalisis data teks guna mengidentifikasi dan memahami sentimen atau opini dalam teks tersebut. Metode ini melibatkan pelatihan sebuah model menggunakan data pelatihan yang berisi teks-teks yang telah diberi label sentimen (positif, negatif, atau netral). Model ini akan belajar pola dan korelasi antara fitur-fitur dalam teks

dengan sentimen yang terkait (Xu et al., 2020).

Metode berbasis pembelajaran mesin dapat menggunakan algoritma klasifikasi seperti *Support Vector Machine (SVM)*, *Naive Bayes*, atau *Random Forest* untuk mengklasifikasikan teks ke dalam kategori sentimen yang sesuai. Proses ini melibatkan ekstraksi fitur dari teks, seperti kata-kata kunci, frase, atau n-gram, dan penggunaan algoritma klasifikasi untuk mengenali pola-pola yang berkaitan dengan sentimen (Shahi et al., 2022).

1. Jaringan Saraf Tiruan Konvolusional (*Convolutional Neural Networks-CNN*) adalah jenis sistem saraf yang khususnya efektif dalam memproses informasi spasial, terutama dalam konteks pengolahan gambar. Lapisan konvolusional dalam *CNN* menggunakan asosiasi khusus dari lapisan sebelumnya, di mana neuron terdekat terhubung dengan neuron pada lapisan berikutnya. Hal ini memungkinkan lapisan untuk memahami perspektif umum tentang sumber informasi yang dihadapinya. Salah satu contoh arsitektur *CNN* dapat ditunjukkan pada Gambar 2.3.

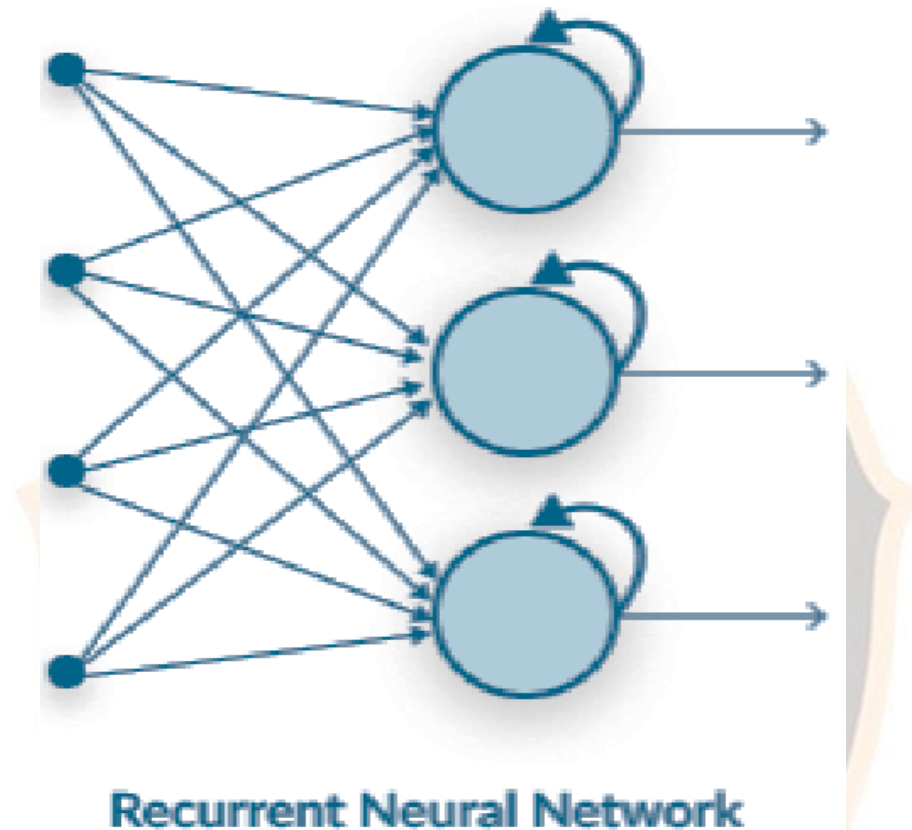


Gambar 2.2 Contoh Arsitektur Convolutional Neural Networks-CNN

Sumber: (Kalitha, 2022)

2. Jaringan Saraf Tiruan Berulang (*Recurrent Neural Networks-RNN*) adalah metode yang digunakan dalam analisis opini, dengan mempertimbangkan perhitungan masa lalu dan menggunakan data yang berurutan. *RNN*

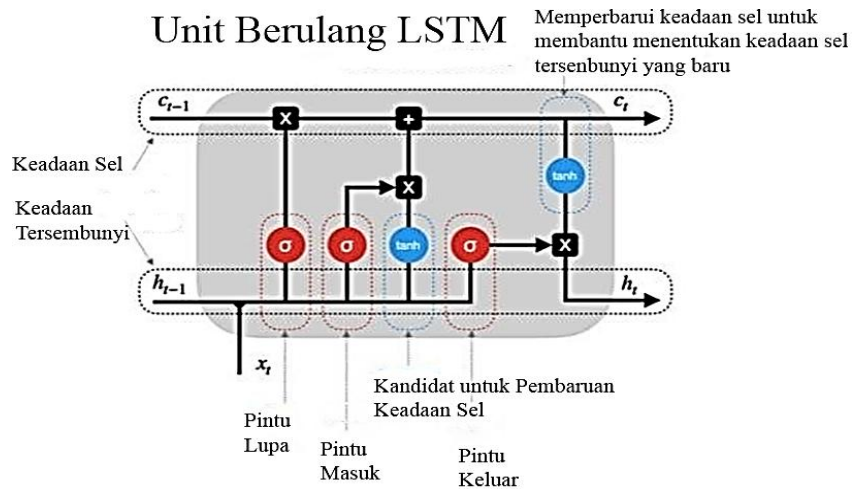
membantu dalam tugas-tugas Pemrosesan Bahasa Alami, seperti perkiraan kata dalam kalimat. *RNN* secara efektif dapat memodelkan hubungan kontekstual antara kata-kata dalam kalimat. Salah satu contoh arsitektur *RNN* dapat ditunjukkan pada Gambar 2.4.



Gambar 2.3 Contoh Arsitektur Recurrent Neural Networks-RNN

Sumber: (Dobilas, 2022)

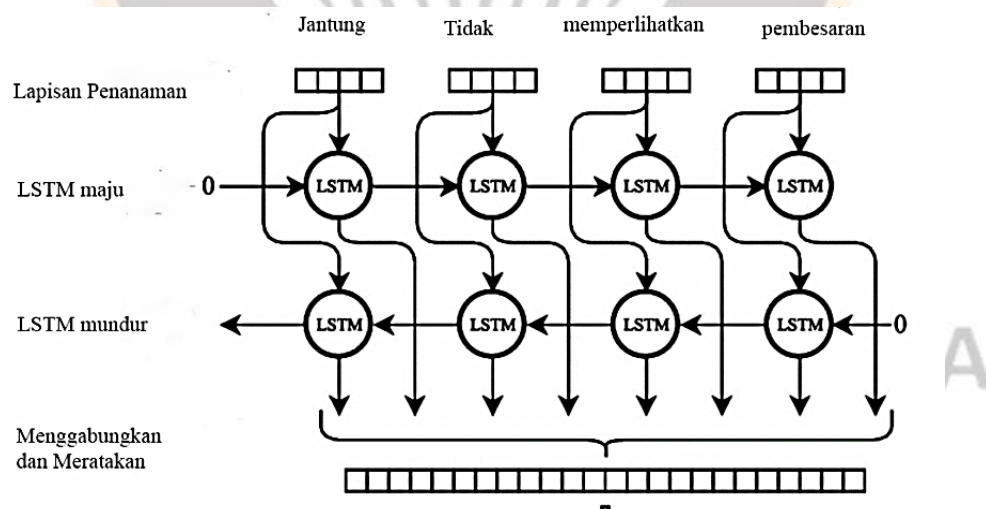
3. *Long Short Term Memory*-LSTM merupakan varian dari *RNN* yang dirancang khusus untuk mengatasi masalah kehilangan data historis dalam urutan yang panjang. *BI-LSTM* menggunakan gerbang memori untuk mempertahankan informasi historis yang relevan, sehingga menghindari kehilangan data yang disebabkan oleh pelatihan yang buruk pada *RNN*. Dengan demikian, LSTM memungkinkan analisis yang lebih baik pada data dengan urutan panjang.



Gambar 2.4 Contoh Arsitektur Long Short-Term Memory-LSTM

Sumber: (Dobilas, 2022)

4. *Bidirectional Long Short-Term Memory-Bi-LSTM* adalah pengembangan dari LSTM yang membawa keuntungan dalam konteks yang lebih kompleks, seperti transfer psikososial. *Bi-LSTM* menggabungkan keadaan tersembunyi maju dan tersembunyi mundur, sehingga memungkinkan aliran informasi dalam kedua arah. Hal ini membantu dalam memodelkan hubungan kontekstual lebih baik dan meningkatkan kinerja dalam situasi yang membutuhkan informasi dari kedua arah. Salah satu contoh arsitektur *Bi-LSTM* dapat ditunjukkan pada Gambar 2.6.



Gambar 2.5 Contoh Arsitektur Bidirectional Long Short-Term Memory-Bi-LSTM

Sumber: (Malviya et al., 2020).

5. BERT (*Bidirectional Encoder Representations from Transformers*) adalah salah satu inovasi terbesar dalam bidang Pemrosesan Bahasa Alami (Natural Language Processing - NLP). Diperkenalkan oleh Google pada tahun 2018, BERT memanfaatkan arsitektur Transformer, yang memungkinkan model memahami konteks kata secara bidirectional, baik dari arah kiri maupun kanan dalam sebuah teks. Pendekatan ini berbeda dari model sebelumnya yang hanya memproses teks secara unidirectional, sehingga sering kehilangan nuansa konteks

Pada dasarnya, metode berbasis pembelajaran mesin bekerja dengan memberikan model komputer sejumlah besar data latihan yang terdiri dari input dan output yang diinginkan. Model ini akan mempelajari pola-pola yang ada dalam data tersebut dan membangun hubungan antara input dan output. Setelah proses pelatihan selesai, model dapat digunakan untuk melakukan prediksi atau klasifikasi pada data baru yang belum diberi label (Malviya *et al.*, 2020). Metode ini memiliki beberapa keunggulan dalam menangani kompleksitas data dalam analisis sentimen, diantaranya adalah (Verma dan Thakur, 2018)

1. Fleksibilitas

Metode berbasis pembelajaran mesin memberikan fleksibilitas dalam mengatasi kompleksitas data. Algoritma pembelajaran mesin mampu mempelajari pola-pola yang rumit dan non-linear yang terdapat dalam data tekstual, sehingga dapat mengenali sentimen yang lebih halus dan nuansa yang kompleks.

2. Generalisasi

Metode ini mampu melakukan generalisasi dari data pelatihan ke data baru yang belum pernah dilihat sebelumnya. Dengan mempelajari pola-pola umum dari data pelatihan, algoritma pembelajaran mesin dapat mengenali sentimen yang sama pada data baru dengan tingkat akurasi yang tinggi.

3. Skalabilitas

Metode berbasis pembelajaran mesin dapat diterapkan pada dataset yang besar dan kompleks. Algoritma pembelajaran mesin dapat mengatasi

dimensi tinggi dan volume besar data, sehingga mampu mengolah dan menganalisis data yang kompleks secara efisien.

Namun, penggunaan metode berbasis pembelajaran mesin juga memiliki beberapa tantangan:

1. Pemahaman yang mendalam

Implementasi metode berbasis pembelajaran mesin membutuhkan pemahaman yang mendalam tentang pemodelan dan pemrosesan data. Pemilihan algoritma yang tepat, pengolahan data yang efisien, dan pengaturan parameter yang optimal memerlukan pengetahuan dan pengalaman yang cukup.

2. Pemrosesan data yang kompleks

Metode ini dapat mengatasi kompleksitas data, tetapi pada saat yang sama memerlukan pengolahan data yang lebih rumit. Hal ini meliputi pra-pengolahan data, seperti pembersihan teks, penghilangan *stopwords*, vektorisasi teks, dan pemilihan fitur yang relevan.

3. Ketergantungan pada jumlah data

Metode pembelajaran mesin memerlukan jumlah data yang memadai untuk memperoleh model yang baik dan akurat. Keterbatasan data pelatihan dapat mempengaruhi kinerja dan kemampuan generalisasi model.

Perkembangan terbaru dalam analisis sentimen telah melibatkan integrasi teknik-teknik analisis sentimen dengan teknologi lain seperti pemrosesan bahasa alami, jaringan sosial, dan analisis media sosial (Birjali *et al.*, 2021). Integrasi ini memungkinkan pemrosesan yang lebih canggih dan pemahaman yang lebih mendalam terhadap opini dan sentimen pengguna. Pengolahan bahasa alami yang lebih lanjut menggunakan teknik-teknik *Neuro Linguistic Programming-NLP* berbasis *transformer* telah meningkatkan akurasi analisis sentimen dengan pemahaman konteks yang lebih kompleks. NLP berbasis *transformer* adalah jenis arsitektur model yang dirancang untuk menangani urutan data, seperti teks. Selain itu, integrasi dengan jaringan sosial memungkinkan penggunaan informasi interaksi dan hubungan sosial dalam analisis sentimen (Gandhi *et al.*, 2023). Analisis media

sosial juga menjadi fokus, di mana data dari *platform* mediasosial digunakan untuk memahami sebaran opini dan pengaruh pengguna. Selain itu, analisis sentimen juga semakin melibatkan data multimodal seperti teks, gambar, dan video, dengan pengembangan teknik analisis sentimen multimodal yang memadukan informasi dari berbagai sumber data. Peningkatan dalam teknik klasifikasi dan prediksi seperti penggunaan *deep learning* memberikan hasil yang baik dalam klasifikasi sentimen. Integrasi teknologi lain dengan analisis sentimen membuka peluang baru dalam pemahaman opini dan sentimen pengguna, memberikan analisis yang lebih holistik dan akurat melalui berbagai kanal dan media komunikasi.

Dalam media sosial khusus *Engagement* dikenal istilah-istilah untuk menunjukkan ukuran ketertarikan pembaca terhadap media sosial sebuah perusahaan atau produk.

Berikut adalah penjelasan masing-masing istilah yang digunakan dalam grafik analisis media sosial:

1. *Likes*

- a. *Definisi*: Jumlah "suka" atau "likes" yang diterima sebuah unggahan.
- b. *Makna*: Menunjukkan tingkat apresiasi atau ketertarikan audiens terhadap konten. Semakin tinggi jumlah likes, semakin positif respons audiens terhadap konten tersebut.

2. *Replies*

- a. *Definisi*: Jumlah balasan atau komentar yang diberikan pengguna terhadap sebuah unggahan.
- b. *Makna*: Mengindikasikan tingkat interaksi langsung antara audiens dan pengunggah. Replies juga mencerminkan bagaimana audiens terlibat secara aktif dalam diskusi terkait konten.

3. *Retweet*

- a. *Definisi*: Jumlah unggahan ulang (retweet) konten oleh pengguna lain.

- b. *Makna*: Menunjukkan seberapa banyak konten dibagikan kembali ke jaringan audiens lain, memperluas jangkauan konten tersebut. Ini juga merupakan indikator bagaimana audiens menganggap konten itu relevan atau bermanfaat.

4. *Post Link Clicks*

- a. *Definisi*: Jumlah klik pada tautan yang disertakan dalam sebuah unggahan.
- b. *Makna*: Mengukur seberapa menarik tautan dalam konten tersebut untuk mendorong audiens mengambil tindakan, seperti mengunjungi halaman web, membaca artikel, atau membeli produk.

5. *Other Post Clicks*

- a. *Definisi*: Jumlah klik pada bagian lain dari unggahan selain tautan utama, misalnya gambar, video, atau elemen teks.
- b. *Makna*: Menunjukkan tingkat rasa ingin tahu atau eksplorasi audiens terhadap elemen visual atau bagian non-teks dari konten.

6. *Other Engagement*

- a. *Definisi*: Interaksi lain yang tidak termasuk likes, replies, retweets, atau clicks, seperti menonton video, menyimpan unggahan, atau tindakan non-klik lainnya.
- b. *Makna*: Memberikan gambaran tambahan tentang bagaimana audiens terlibat secara pasif atau tidak langsung dengan konten

