

Sentiment Analysis Classification of Deepseek AI on Twitter (X) Big Data Using SOM-LSTM

Indra Irawan-1^{a*}, M.Hizbul Wathan-2^a, Roofik Aditya Erlangga-3^a

^aProdi D4 Teknologi Rekayasa Perangkat Lunak, Politeknik Manufaktur Negeri Bangka Belitung

*Email : indraa@polman-babelac.id

Abstrak

Urgensi penelitian ini adalah perlunya mengidentifikasi dan mengklasifikasikan sentimen publik terhadap DeepSeek AI di Twitter (X) secara akurat dan efisien menggunakan metode SOM-LSTM untuk memahami persepsi masyarakat berbasis big data. Analisis sentimen terhadap opini publik mengenai DeepSeek AI pada media sosial Twitter (X) penting dilakukan untuk mengidentifikasi kecenderungan persepsi pengguna terhadap perkembangan teknologi kecerdasan buatan. Penelitian ini bertujuan mengklasifikasikan sentimen publik terhadap DeepSeek AI menggunakan model hibrida Self-Organizing Map–Long Short-Term Memory (SOM–LSTM). Data penelitian terdiri atas 10.000 cuitan yang dikumpulkan melalui Twitter API menggunakan kata kunci “DeepSeek AI” pada periode Januari 2024–Juli 2025. Data melalui tahap pelabelan otomatis menggunakan BERT, preprocessing, tokenisasi, dan representasi kata menggunakan GloVe berdimensi 100 sebelum diproses menggunakan SOM–LSTM. SOM digunakan untuk melakukan pengelompokan fitur secara tidak terawasi, sedangkan LSTM digunakan untuk mempelajari hubungan sekuensial dalam teks. Evaluasi dilakukan menggunakan *accuracy*, *precision*, *recall*, dan F1-score dengan membandingkan kinerja SOM–LSTM terhadap LSTM sebagai model baseline. Hasil penelitian menunjukkan bahwa SOM–LSTM memperoleh *accuracy* sebesar 0,98, *precision* sebesar 0,99, *recall* sebesar 0,98, dan F1-score sebesar 0,98, sedangkan LSTM memperoleh *accuracy* sebesar 0,85. Hasil tersebut menunjukkan bahwa integrasi SOM dapat meningkatkan kinerja klasifikasi dibandingkan LSTM tunggal pada dataset yang digunakan. Secara substantif, hasil analisis menunjukkan dominasi sentimen negatif terhadap DeepSeek AI. Temuan penelitian ini dapat menjadi informasi pendukung bagi peneliti, pengembang, dan pemangku kepentingan dalam memahami persepsi publik serta mengidentifikasi isu yang berkaitan dengan penggunaan dan pengembangan teknologi AI.

Kata Kunci : *Analisis Sentimen, Big Data, Deepseek AI, SOM–LSTM, Twitter (X)*

1. Latar Belakang

Perkembangan kecerdasan buatan (*Artificial Intelligence*) dalam dekade terakhir menunjukkan peningkatan yang sangat signifikan, terutama dengan hadirnya berbagai model bahasa besar (*Large Language Model/LLM*) yang mampu memahami dan menghasilkan teks menyerupai bahasa manusia. Model-model

seperti ChatGPT, Gemini, dan DeepSeek AI telah berkontribusi dalam mempercepat transformasi digital di berbagai sektor, mulai dari pendidikan, bisnis, hingga penelitian ilmiah [1]. DeepSeek AI, yang dirilis secara terbuka pada awal tahun 2025, menandai tonggak penting dalam perkembangan AI terbuka (*open-source reasoning model*). Peningkatan penggunaan teknologi tersebut

mengindikasikan adanya perubahan preferensi pengguna terhadap teknologi AI yang bersifat kolaboratif, terbuka, dan dapat dikembangkan bersama komunitas global. Namun, seiring meningkatnya popularitas DeepSeek AI, muncul pula beragam opini publik di media sosial, terutama pada platform Twitter (X), yang berperan sebagai ruang diskusi global terkait berbagai isu dan sentimen publik terhadap perkembangan teknologi. Analisis terhadap opini publik tersebut penting untuk memahami persepsi masyarakat terhadap perkembangan AI melalui kecenderungan sentimen yang muncul pada media sosial [2], [3].

Penelitian-penelitian sebelumnya telah menggunakan berbagai pendekatan untuk menganalisis sentimen publik di media sosial. Selain pendekatan berbasis machine learning, analisis sentimen pada data media sosial juga dapat dilakukan menggunakan pendekatan berbasis aturan seperti vader [4], [5]. Metode berbasis machine learning telah digunakan secara luas untuk mengklasifikasikan sentimen pada data media sosial [6], [7]. Namun, pendekatan tersebut memiliki keterbatasan dalam menangkap konteks semantik dan hubungan sekuensial antarkata. Oleh karena itu, pendekatan berbasis *deep learning* seperti *Long Short-Term Memory* (LSTM) mulai banyak diterapkan karena kemampuannya dalam mengenali pola sekuensial pada teks [8]. Meskipun demikian, model LSTM masih menghadapi tantangan dalam menghasilkan representasi fitur yang optimal, terutama ketika diterapkan pada data dengan jumlah besar dan variasi konteks yang tinggi. Penelitian terdahulu menunjukkan bahwa *Self-Organizing Map* (SOM) dapat digunakan untuk melakukan clustering dan menstrukturkan representasi data, termasuk pada data berbasis teks [9], [10]. SOM berperan dalam melakukan *unsupervised clustering* untuk menstrukturkan representasi fitur sebelum diproses lebih lanjut oleh LSTM, sehingga dapat mendukung proses pembelajaran representasi teks [11], [12].

Meskipun penelitian sebelumnya telah menunjukkan efektivitas metode machine

learning, *deep learning*, LSTM, dan pendekatan hybrid dalam analisis sentimen media sosial, penerapan kombinasi SOM dan LSTM untuk menganalisis opini publik terhadap DeepSeek AI pada Twitter (X) masih terbatas. Penelitian terdahulu lebih banyak berfokus pada penggunaan model klasifikasi secara individual maupun pendekatan *deep learning* dan *transformer* untuk analisis sentimen, sedangkan pemanfaatan SOM sebagai tahap pengelompokan representasi fitur sebelum proses pembelajaran sekuensial menggunakan LSTM dalam konteks opini publik mengenai DeepSeek AI belum banyak dikaji. Selain itu, karakteristik data Twitter (X) yang singkat dan memiliki variasi konteks yang tinggi memerlukan pendekatan yang mampu mengelola representasi fitur secara efektif sebelum proses klasifikasi. Kondisi tersebut menunjukkan adanya kebutuhan untuk mengevaluasi penerapan model hybrid SOM–LSTM pada data Twitter (X) yang membahas DeepSeek AI.

Beberapa kajian terkini juga menunjukkan potensi pendekatan hybrid dalam meningkatkan performa analisis sentimen pada data media sosial [13], [14]. Kajian pada data Twitter juga menunjukkan pentingnya pemilihan dan evaluasi metode yang sesuai dalam klasifikasi sentimen [15]. Pendekatan *hybrid* menggabungkan beberapa mekanisme pembelajaran dalam satu arsitektur untuk mendukung proses klasifikasi sentimen pada data teks yang memiliki karakteristik beragam. Temuan tersebut memperkuat relevansi penggunaan pendekatan hybrid dalam analisis sentimen, khususnya pada data media sosial yang memiliki variasi konteks yang tinggi.

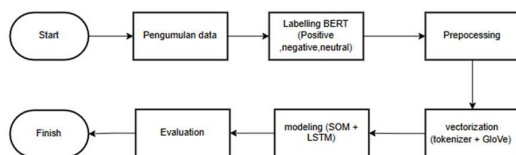
Berdasarkan kondisi tersebut, penelitian ini mengintegrasikan *Self-Organizing Map* (SOM) dan *Long Short-Term Memory* (LSTM) dalam klasifikasi sentimen publik terhadap DeepSeek AI pada data Twitter (X). SOM digunakan untuk melakukan pengelompokan representasi fitur secara tidak terawasi sebelum diproses oleh LSTM untuk mempelajari hubungan sekuensial dalam teks. Representasi kata dilakukan

menggunakan GloVe berdimensi 100 [16], sedangkan kinerja pendekatan *hybrid* dievaluasi melalui perbandingan dengan model LSTM tunggal pada dataset yang sama. Pendekatan tersebut diarahkan untuk mengkaji sejauh mana pengelompokan representasi fitur menggunakan SOM dapat mendukung kinerja LSTM dalam klasifikasi sentimen pada data opini publik mengenai DeepSeek AI.

Dengan demikian, penelitian ini bertujuan untuk mengklasifikasikan sentimen publik terhadap DeepSeek AI pada Twitter (X) menggunakan model *hybrid* SOM–LSTM serta mengevaluasi kinerjanya dibandingkan dengan model LSTM tunggal. Hasil klasifikasi selanjutnya digunakan untuk menggambarkan kecenderungan persepsi publik terhadap DeepSeek AI berdasarkan kategori sentimen positif, negatif, dan netral.

2. Metodologi

Metode penelitian ini dirancang untuk mengklasifikasikan sentimen terhadap Deepseek AI berdasarkan big data Twitter (X) dengan pendekatan berbasis deep learning. Tahapan penelitian melibatkan pengumpulan data, labelling, preprocessing, pemodelan sentimen menggunakan SOM–LSTM, evaluasi analisis hasil serta luaran dan publikasi. Berikut adalah tahapan yang dilakukan dalam penelitian ini.



Gambar 1. Metode Penelitian

Pengumpulan Data

Data dikumpulkan melalui Twitter API dengan kata kunci “DeepSeek AI” selama Januari 2024 – Juli 2025. Sebanyak 10.000 data dikumpulkan, mencakup opini positif, negatif, dan netral dari berbagai negara. Data duplikat, *retweet*, dan spam dihapus untuk menjaga validitas dataset. Validasi data dilakukan melalui pemeriksaan kualitas dan

relevansi dataset. Setelah proses pengumpulan menggunakan Twitter API, data diperiksa untuk mengidentifikasi dan menghapus data duplikat, retweet, spam, serta data yang tidak relevan dengan kata kunci “DeepSeek AI”. Proses validasi ini dilakukan untuk memastikan bahwa data yang digunakan memiliki relevansi terhadap tujuan penelitian dan tidak mengandung data berulang atau konten yang dapat mengganggu proses klasifikasi sentimen. Data yang telah melalui proses validasi selanjutnya digunakan pada tahap pelabelan sentimen menggunakan model BERT dan proses pemodelan SOM–LSTM.

Pelabelan Data

Pelabelan dilakukan menggunakan model BERT Sentiment Analysis untuk menghasilkan tiga label utama: positif, negatif, dan netral. Model ini dipilih karena kemampuannya memahami konteks linguistik yang kompleks dan menghasilkan label secara otomatis dengan akurasi tinggi [17].

Preprocessing

Tahap *preprocessing* dilakukan untuk membersihkan dan mempersiapkan data teks sebelum digunakan dalam proses klasifikasi [18], [19]. Tahap ini terdiri dari beberapa sub-tahapan sebagai berikut :

- Case Folding*: Mengubah seluruh teks menjadi huruf kecil untuk memastikan keseragaman dan mengurangi variasi yang disebabkan oleh kapitalisasi. Contoh: “DEEPSEEK AI IS AMAZING” → “deepseek ai is amazing”.
- Cleansing*: Menghapus karakter yang tidak diperlukan seperti URL, hashtags (#DeepSeekAI), mentions (@user), emoji, angka, dan tanda baca. seperti: “@deepseek is awesome! 😊 #AI2025” menjadi “deepseek is awesome”.
- Tokenizing*: Memecah kalimat menjadi kata-kata atau token terpisah. seperti: “deepseek ai is amazing” → [“deepseek”, “ai”, “is”, “amazing”].

- d. *Normalization*: Mengubah kata informal atau singkatan menjadi bentuk standarnya untuk menjaga konsistensi dan meningkatkan kemampuan model dalam memahami makna dengan akurat. seperti: “gonna” → “going to”, “btw” → “by the way”.
- e. *Stopword Removal*: Menghapus kata-kata yang sering muncul namun memiliki makna semantik yang rendah seperti “and”, “is”, “the”, “of”, dan “to”. seperti: “this is an AI model” → “AI model”. Langkah ini membantu mengurangi noise dan memfokuskan model pada kata-kata dengan bobot sentimen yang lebih tinggi
- f. *Lemmatization*: Mengubah kata menjadi bentuk dasarnya atau bentuk kamusnya untuk menyatukan variasi kata. seperti: “running”, “ran” → “run”; “better” → “good”.
- g. *Join Process*: Menggabungkan kembali token yang telah dibersihkan menjadi urutan teks yang koheren yang akan digunakan sebagai masukan untuk proses embedding dan model.

Rangkaian langkah ini secara signifikan meningkatkan kualitas teks, mengurangi dimensi data, serta memperkuat kemampuan pembelajaran model pada tahap selanjutnya.

Vektorisasi

Setelah tahap preprocessing, data teks yang telah dibersihkan diubah menjadi representasi numerik menggunakan Tokenizer untuk menghasilkan urutan kata. Selanjutnya, *Global Vectors for Word Representation* (GloVe) digunakan untuk menghasilkan representasi vektor kata berdasarkan hubungan semantik antarkata [19]. GloVe merepresentasikan setiap kata dalam ruang vektor berdimensi tertentu, sehingga kata-kata yang memiliki keterkaitan semantik cenderung memiliki representasi vektor yang berdekatan. Pada penelitian ini, digunakan GloVe dengan dimensi vektor 100. Representasi tersebut selanjutnya digunakan sebagai masukan pada tahap

pemodelan untuk mendukung proses klasifikasi sentimen.

Pemodelan Hybrid SOM–LSTM

Model *hybrid* yang diusulkan mengintegrasikan *Self-Organizing Map* (SOM) dan *Long Short-Term Memory* (LSTM) untuk meningkatkan akurasi serta stabilitas dalam klasifikasi sentimen. SOM berfungsi sebagai mekanisme *unsupervised clustering* yang memetakan fitur-fitur sentimen berdimensi tinggi ke dalam grid topologis, sedangkan LSTM berperan sebagai model pembelajaran sekuensial terawasi (*supervised sequential learning model*) yang mampu menangkap ketergantungan kontekstual dalam teks [6].

Self-Organizing Map (SOM)

Self-Organizing Map (SOM) yang dikembangkan oleh Kohonen dirancang untuk melakukan *clustering* dan reduksi dimensi melalui proses *competitive learning*. Setiap neuron pada SOM direpresentasikan oleh vektor bobot w_j yang akan bersaing untuk menjadi *Best Matching Unit* (BMU) bagi setiap vektor masukan x_i . Jarak Euclidean antara vektor masukan dan bobot setiap neuron dihitung dengan persamaan sebagai berikut:

$$D_j = \|X_i - W_j\| = \sqrt{\sum_{k=1}^n (X_{ik} - W_{jk})^2} \quad (1)$$

Neuron dengan jarak terkecil D_{jD_jDj} akan dipilih sebagai *Best Matching Unit* (BMU). Setelah BMU ditentukan, pembaruan bobot dilakukan berdasarkan aturan pembelajaran berikut:

$$W_j(t+1) = W_j(t) + \alpha(t) \cdot h_{bj}(t) \cdot (X_i - W_j(t)) \quad (2)$$

di mana : 1). $\alpha(t)$ = laju pembelajaran (*learning rate*) pada iterasi ke- t ; 2). $h_{bj}(t)$ = fungsi kedekatan (*neighborhood function*) antara BMU dan neuron ke- j ; 3). X_i = vektor masukan (*input vector*) pada iterasi ke- i ; 4). $W_j(t)$ = bobot neuron ke- j pada waktu ke- t .

Melalui proses pembaruan yang dilakukan secara iteratif, SOM membentuk

peta topologis di mana pola-pola sentimen yang serupa dikelompokkan dalam posisi yang berdekatan. Representasi hasil klasterisasi tersebut kemudian digunakan sebagai masukan terstruktur bagi lapisan LSTM pada tahap selanjutnya.

Long Short-Term Memory (LSTM)

Setelah proses *clustering* menggunakan SOM, hasil pengelompokan diteruskan ke model *Long Short-Term Memory* (LSTM) untuk melakukan klasifikasi sentimen secara berurutan [20]. LSTM merupakan salah satu jenis *Recurrent Neural Network* (RNN) yang dirancang untuk mempelajari ketergantungan informasi dalam jangka panjang (*long-term dependencies*). Kemampuan tersebut diperoleh melalui mekanisme *cell state* dan tiga gerbang utama, yaitu *forget gate*, *input gate*, dan *output gate*, yang mengatur informasi yang dipertahankan, ditambahkan, dan dikeluarkan pada setiap *time step*.

Pada tahap pertama, *forget gate* digunakan untuk menentukan informasi dari *cell state* sebelumnya yang masih relevan dan perlu dipertahankan. Perhitungannya dinyatakan pada Persamaan (3):

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3)$$

Pada Persamaan (3), (f_t) merupakan keluaran *forget gate* yang menentukan tingkat informasi yang akan dipertahankan dari *cell state* sebelumnya. Fungsi (σ) merupakan fungsi aktivasi sigmoid yang menghasilkan nilai antara 0 dan 1. (W_f) menunjukkan bobot pada *forget gate*, sedangkan (h_{t-1}) merupakan *hidden state* pada waktu sebelumnya dan (x_t) merupakan masukan pada *time step* saat ini. Sementara itu, (b_f) merupakan *bias* pada *forget gate*. Nilai keluaran yang mendekati 0 menunjukkan bahwa informasi sebelumnya cenderung dilupakan, sedangkan nilai yang mendekati 1 menunjukkan bahwa informasi tersebut dipertahankan.

Tahap berikutnya adalah *input gate*, yang berfungsi menentukan informasi baru yang perlu dimasukkan ke dalam *cell state*. Proses ini terdiri atas perhitungan *input gate*

dan *candidate cell state*. Nilai *input gate* dihitung menggunakan Persamaan (4a), sedangkan kandidat informasi baru dihitung menggunakan Persamaan (4b):

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (4a)$$

$$C_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (4b)$$

Nilai (i_t) menunjukkan seberapa besar informasi baru yang akan dimasukkan ke dalam *cell state*, dengan (W_i) dan (b_i) masing-masing merupakan bobot dan bias untuk *input gate*. Sementara itu, (C_t) merupakan *candidate cell state* yang berisi kandidat informasi baru yang diperoleh melalui fungsi aktivasi tanh. Bobot dan bias yang digunakan untuk menghasilkan kandidat tersebut masing-masing dinyatakan dengan (W_c) dan (b_c).

Setelah nilai *forget gate*, *input gate*, dan *candidate cell state* diperoleh, *cell state* diperbarui dengan mengombinasikan informasi lama dan informasi baru. Proses pembaruan tersebut dinyatakan sebagai berikut:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot C_t \quad (4c)$$

Pada Persamaan (4c), (C_t) merupakan *cell state* yang telah diperbarui, sedangkan (C_{t-1}) merupakan *cell state* pada *time step* sebelumnya. Komponen ($f_t \cdot C_{t-1}$) menunjukkan informasi lama yang dipertahankan, sedangkan ($i_t \cdot C_t$) menunjukkan informasi baru yang ditambahkan ke dalam *cell state*. Mekanisme tersebut memungkinkan LSTM mempertahankan informasi penting dari urutan sebelumnya sekaligus memperbaruinya berdasarkan masukan pada *time step* saat ini.

Selanjutnya, *output gate* menentukan informasi yang akan dikeluarkan dari unit LSTM sebagai *hidden state*. Nilai *output gate* dihitung menggunakan Persamaan (5a), sedangkan *hidden state* diperoleh melalui Persamaan (5b):

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5a)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (5b)$$

Pada Persamaan (5a), (o_t) merupakan keluaran output gate, sedangkan (W_o) dan (b_o) masing-masing merupakan bobot dan bias pada output gate. Selanjutnya, (h_t) pada Persamaan (5b) merupakan hidden state yang dihasilkan berdasarkan output gate dan cell state yang telah diperbarui. Dalam konteks penelitian ini, hidden state menjadi representasi keluaran dari proses pembelajaran LSTM yang selanjutnya digunakan dalam proses klasifikasi sentimen.

Penggunaan LSTM setelah proses *clustering* SOM memungkinkan informasi hasil pengelompokan dipelajari secara berurutan. SOM berperan dalam mengelompokkan pola data berdasarkan karakteristik yang dimiliki, sedangkan LSTM memanfaatkan informasi tersebut untuk mempelajari hubungan antardata dalam urutan. Dengan demikian, kombinasi SOM–LSTM digunakan untuk memperoleh representasi pola sentimen dari hasil *clustering* sekaligus mempertahankan informasi urutan yang relevan dalam proses klasifikasi.

Dataset

Dataset yang digunakan dalam penelitian menjadi masukan utama pada tahapan pengolahan dan pemodelan. Dataset tersebut berisi data yang digunakan untuk proses clustering menggunakan SOM dan selanjutnya diteruskan ke model LSTM untuk proses klasifikasi sentimen. Pada tahap awal, dataset dipersiapkan dan diolah agar sesuai dengan kebutuhan masing-masing metode yang digunakan. Proses clustering dengan SOM bertujuan mengelompokkan data berdasarkan pola atau karakteristik tertentu, sedangkan LSTM digunakan untuk mempelajari pola sekuensial dalam data guna mendukung proses klasifikasi sentimen. Dengan demikian, keterkaitan kedua metode tersebut diharapkan dapat membantu menghasilkan klasifikasi sentimen yang lebih terstruktur sesuai dengan karakteristik data yang dianalisis.

id_str	full_text	created_at	tweet_url	user_id_str
0	DeepSeek Coder ranks as the fastest open-sour...	Fri Jan 26 03:58:26 +0000 2024	https://x.com/undefined/status/17502988773497...	1.714800e+18
1	Just Out: Tech Report On #DeepSeekCoder - An ...	Fri Jan 26 03:58:26 +0000 2024	https://x.com/undefined/status/1750298844396...	1.714800e+18
2	one of my fav models has a G7 that seems to w...	Fri Jan 26 03:49:45 +0000 2024	https://x.com/undefined/status/17502779125666...	5.833730e+07
3	this is so cool thank you for sharing your th...	Thu Jan 25 14:35:55 +0000 2024	https://x.com/undefined/status/175025750451546...	1.684410e+18
4	You can use the model for FREE by downloading ...	Thu Jan 25 14:30:43 +0000 2024	https://x.com/undefined/status/17502561759228...	1.638370e+18
5	@BulyCman Try @deepseek_ai for coding tasks...	Tue Jan 23 21:46:49 +0000 2024	https://x.com/undefined/status/17491188989877...	1.638330e+18
6	@gptmax @gptmaxcompute @NousResearch @deep...	Tue Jan 23 15:36:42 +0000 2024	https://x.com/undefined/status/1749184727957...	1.641380e+18
7	@misyong @gptmaxcompute @NousResearch @deep...	Tue Jan 23 06:53:55 +0000 2024	https://x.com/undefined/status/17490880343985...	1.570327e+18
8	@deepseek_ai Where's a model 7b?	Tue Jan 23 06:07:18 +0000 2024	https://x.com/undefined/status/17490751330299...	1.532770e+18
9	@misyong @gptmaxcompute @NousResearch @deep...	Mon Jan 22 18:38:02 +0000 2024	https://x.com/undefined/status/1748516380304...	1.683760e+18
10	Besides Bard two exciting new models are also ...	Mon Jan 22 17:42:14 +0000 2024	https://x.com/undefined/status/17484876144905...	1.641380e+18
11	A popular Chinese AI chatbot called #DeepSeek...	Mon Jan 22 14:28:48 +0000 2024	https://x.com/undefined/status/1748388068871...	1.538750e+18
12	DeepSeek AI: https://t.co/9T84JHEV Tele Chat...	Mon Jan 22 09:09:15 +0000 2024	https://x.com/undefined/status/17483713847055...	1.642400e+18
13	DeepSeek AI Proposes DeepSeekAI: An Innovati...	Sun Jan 21 12:45:45 +0000 2024	https://x.com/undefined/status/17480563116080...	3.511710e+18
14	The language model landscape is evolving rap...	Thu Jan 18 20:11:51 +0000 2024	https://x.com/undefined/status/17460751498768...	1.707860e+18
15	DeepSeek AI Proposes DeepSeekAI: An Innovati...	Thu Jan 18 18:20:07 +0000 2024	https://x.com/undefined/status/17460476327055...	7.75910e+17
16	DeepSeek AI's Innovative Language Model Arch...	Thu Jan 18 16:16:54 +0000 2024	https://x.com/undefined/status/17460468134192...	9.82910e+17
17	Explore the innovative DeepSeekAI architectu...	Thu Jan 18 15:26 +0000 2024	https://x.com/undefined/status/17460464327437...	2.889380e+18

Gambar 2. Dataset

Gambar 2 memperlihatkan dataset yang digunakan sebagai sumber data penelitian. Dataset tersebut menjadi dasar dalam proses pengolahan data sebelum memasuki tahapan *clustering* dan klasifikasi. Data yang telah melalui tahapan persiapan selanjutnya digunakan oleh SOM untuk memperoleh pola pengelompokan, kemudian hasil pengelompokan tersebut digunakan sebagai masukan pada model LSTM. Dengan alur tersebut, dataset tidak hanya berfungsi sebagai sumber data awal, tetapi menjadi bagian penting dalam keseluruhan proses pembentukan model SOM–LSTM untuk klasifikasi sentimen.

Evaluasi Model

Evaluasi dilakukan menggunakan metrik akurasi, precision, recall, dan F1-score [12]. Kinerja SOM–LSTM dibandingkan dengan model LSTM tunggal untuk mengukur peningkatan hasil.

3. Hasil Dan Pembahasan

Pada Rangkaian pengujian dalam penelitian ini dilakukan melalui beberapa tahap sistematis yang berkontribusi pada proses analisis sentimen menggunakan model hybrid SOM-LSTM. Tahapan meliputi pengumpulan data, pelabelan, praproses, vektorisasi, pemodelan, dan evaluasi sebagaimana dijelaskan berikut:

Pengumpulan Data

Data penelitian diperoleh dari media sosial Twitter (X) menggunakan API resmi dengan kata kunci “DeepSeek AI” pada periode Januari 2024 – juli 2025 sejumlah 10.000 data. Data yang dikumpulkan mencakup *full_text*, *user_id*, *created_at*, dan *tweet_url*. Cuitan duplikat, *retweet*, serta spam dihapus untuk menjaga kebersihan dataset.

Pelabelan Data

Pelabelan dilakukan secara otomatis menggunakan model BERT yang mengklasifikasikan sentimen menjadi tiga kategori: positif, negatif, dan netral. BERT mampu memahami konteks semantik sehingga hasil pelabelan lebih akurat dan relevan sebelum tahap pelatihan model.

id_str	cleaned	sentiment_label	labels
0	deepseek optimizing language models for dialogue	5 stars	positive
1	try talking with deepseek our new ai system wh...	5 stars	positive
2	deepseek optimizing language models for dialog...	4 stars	positive
3	thrilled to share that deepseek our new model ...	5 stars	positive
4	as of minutes ago released their new deepseek ...	5 stars	positive
5	just launched deepseek our new ai system which...	5 stars	positive
6	as of minutes ago released their new deepseek ...	5 stars	positive
7	deepseek coming out strong refusing to help me...	3 stars	neutral
8	just deployed a thing i ve been helping build ...	1 star	negative
9	research preview of our newest model deepseek ...	5 stars	positive
10	god damn it stop announcing things i am too busy	1 star	negative
11	deepseekai deepseek optimizing language models...	5 stars	positive
12	deepseek optimizing language models for dialog...	4 stars	positive
13	deepseekai announced deepseek a model optimize...	5 stars	positive
14	deepseekai deepseek optimizing language models...	5 stars	positive
15	deepseek optimizing language models for dialogue	5 stars	positive
16	deepseek optimizing language models for dialog...	5 stars	positive
17	models are set to become the search engines of...	3 stars	neutral
18	deepseekai deepseek optimizing language models...	5 stars	positive

Gambar 3. Pelabelan Data

Gambar 3 menunjukkan proses pelabelan data yang dilakukan secara otomatis menggunakan model BERT. Pada tahap ini, setiap data dianalisis berdasarkan konteks semantik yang terkandung di dalamnya, kemudian dikategorikan ke dalam salah satu dari tiga kelas sentimen, yaitu positif, negatif, atau netral. Hasil klasifikasi tersebut digunakan sebagai label sentimen pada dataset yang selanjutnya menjadi acuan dalam proses pembelajaran model. Pelabelan secara otomatis memungkinkan seluruh data diproses dengan mekanisme yang seragam sebelum digunakan pada tahap *clustering*

menggunakan SOM dan klasifikasi menggunakan LSTM.

Preprocessing Data

Data teks dibersihkan dan dinormalisasi melalui beberapa langkah: *case folding*, *cleansing*, *tokenizing*, *normalization*, *stopword removal*, *lemmatization*. Tahapan ini menghasilkan teks yang bersih, konsisten, dan siap untuk dikonversi ke bentuk numerik.

	full_text	casefold	cleaned	tokenized	normalized	stopwords	lemmatizer
0	DeepSeek: Optimizing Language Models for Dialo...	deepseek: optimizing language models for dialo...	deepseek optimizing language models for dialogue	[deepseek, optimizing, language, models, for, ...]	[deepseek, optimizing, language, models, for, ...]	[deepseek, optimizing, language model dialogue]	deepseek optimizing language model dialogue
1	Try talking with DeepSeek, our new AI system w...	try talking with deepseek, our new ai system w...	try talking with deepseek our new ai system wh...	[try, talking, with, deepseek, our, new, ai, s...	[try, talking, with, deepseek, our, new, ai, s...	[try, talking, deepseek, new, ai, system, opti...	try talking deepseek new ai system optimized d...
2	DeepSeek: Optimizing Language Models for Dialo...	deepseek: optimizing language models for dialo...	deepseek optimizing language models for dialog...	[deepseek, optimizing, language, models, for, ...]	[deepseek, optimizing, language, models, for, ...]	[deepseek, optimizing, language, models, dialo...	deepseek optimizing language model dialogue re...
3	THRILLED to share that DeepSeek, our new model ...	thrilled to share that deepseek, our new model ...	thrilled to share that deepseek our new model ...	[thrilled, to, share, that, deepseek, our, new...	[thrilled, to, share, that, deepseek, our, new...	[thrilled, share, deepseek, new, model, optimi...	thrilled share deepseek new model optimized al...
4	As of 2 minutes ago, @DeepSeekAI released thei...	as of 2 minutes ago, @deepseekai released thei...	as of minutes ago released their new deepseek ...	[as, of, minutes, ago, released, their, new, d...	[as, of, minutes, ago, released, their, new, d...	[minutes, ago, released, new, deepseek, nand...	minute ago released new deepseek nand use right
5	Just launched DeepSeek, our new AI system whic...	just launched deepseek, our new ai system whic...	just launched deepseek our new ai system whic...	[just, launched, deepseek, our, new, ai, syste...	[just, launched, deepseek, our, new, ai, syste...	[launched, deepseek, new, ai, system, optimiz...	launched deepseek new ai system optimized dial...

Gambar 4. Preprocessing Data

Gambar 4 memperlihatkan tahapan *preprocessing* data teks yang dilakukan sebelum data digunakan dalam proses pemodelan. Data terlebih dahulu melalui *case folding* untuk menyeragamkan penggunaan huruf, kemudian dilakukan *cleansing* untuk menghilangkan unsur atau karakter yang tidak diperlukan. Selanjutnya, teks dipecah menjadi token melalui proses *tokenizing*, diikuti *normalization* untuk menyeragamkan bentuk kata, *stopword removal* untuk mengurangi kata-kata yang kurang memberikan informasi, serta *lemmatization* untuk memperoleh bentuk dasar kata. Rangkaian tahapan tersebut menghasilkan data teks yang lebih bersih dan konsisten sehingga dapat diproses pada tahap berikutnya dan dikonversi menjadi representasi numerik sebagai masukan model.

Vektorisasi dan Embedding

Teks hasil praproses dikonversi menjadi urutan numerik menggunakan Tokenizer, kemudian direpresentasikan menggunakan

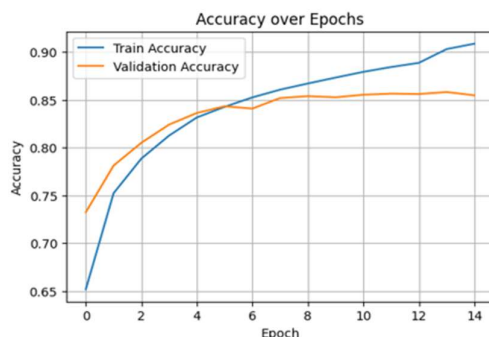
GloVe (Global Vectors for Word Representation) dengan dimensi 100. Teknik ini memungkinkan model memahami hubungan semantik antar kata, misalnya “AI” dan “intelligence” berada dekat dalam ruang vektor [21].

Implementasi Model Hybrid SOM-LSTM

Model SOM digunakan untuk melakukan *unsupervised clustering* terhadap fitur sentimen, sehingga menghasilkan peta topologis dua dimensi yang merepresentasikan kedekatan semantik antar data. Hasil *clustering* tersebut kemudian dijadikan masukan (*input feature*) bagi model LSTM, yang berfungsi menganalisis hubungan sekuensial dan konteks kalimat untuk menentukan kategori sentimen. Untuk mengetahui peningkatan kinerja model, dilakukan perbandingan performa antara LSTM murni dan SOM + LSTM hybrid, sebagaimana ditunjukkan pada hasil evaluasi berikut.

Performa LSTM

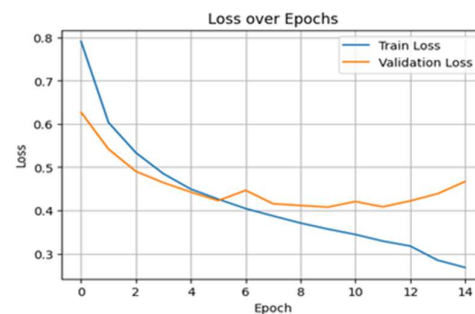
Model LSTM digunakan sebagai *baseline* untuk mengukur performa dasar sebelum dilakukan integrasi dengan metode SOM. Arsitektur model terdiri atas satu lapisan embedding berukuran 100 dan satu lapisan LSTM dengan 128 unit memori, dilatih menggunakan data hasil preprocessing dan pelabelan otomatis berbasis BERT.



Gambar 5. Accuracy Over Epoch LSTM

Gambar 5 menunjukkan perubahan nilai akurasi model LSTM pada data pelatihan (*train accuracy*) dan data validasi (*validation accuracy*) selama proses pembelajaran. Pada

awal pelatihan, akurasi *train* berada pada sekitar 0,65 dan meningkat secara bertahap hingga mencapai sekitar 0,91 pada epoch ke-14. Sementara itu, akurasi validasi meningkat dari sekitar 0,73 pada awal pelatihan hingga mencapai kisaran 0,85 pada beberapa epoch terakhir. Perbedaan antara akurasi pelatihan dan validasi mulai terlihat lebih jelas pada epoch akhir, ketika akurasi pelatihan terus meningkat sedangkan akurasi validasi cenderung stabil. Pola tersebut menunjukkan bahwa model mampu mempelajari pola pada data pelatihan dengan baik, sementara peningkatan kinerja pada data validasi mulai melandai pada tahap akhir pelatihan.



Gambar 6. Loss over epoch LSTM

Berdasarkan Gambar 6, grafik *Loss over Epochs* memperlihatkan penurunan nilai loss baik pada data pelatihan maupun validasi. Nilai training loss terus menurun hingga sekitar 0,25, sedangkan *validation loss* cenderung stabil di kisaran 0,35 setelah epoch ke-10. Kondisi ini menunjukkan bahwa model telah mencapai titik konvergensi yang baik. Hasil evaluasi kuantitatif disajikan pada tabel metrik berikut.

Tabel 1. Hasil Evaluasi Model LSTM

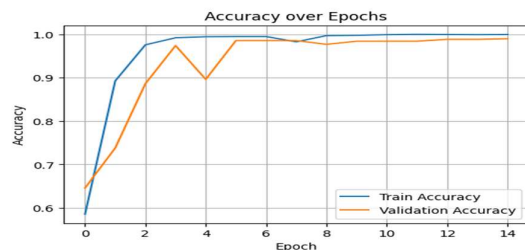
Sentiment	Precision	Recall	F1-Score
Negative	0,90	0,93	0,91
Neutral	0,73	0,73	0,73
Positive	0,89	0,82	0,85

Rata-rata *macro average* untuk precision, recall, dan F1-score masing-masing sebesar 0.84, 0.83, dan 0.83, sedangkan *weighted average* mencapai 0.85. Secara keseluruhan, model LSTM murni memperoleh tingkat akurasi sebesar 0.85.

Hasil ini menunjukkan bahwa meskipun model LSTM mampu mengenali pola sentimen dengan baik, performanya masih dapat ditingkatkan khususnya pada kategori netral, yang memiliki nilai F1-score terendah. Hal ini menjadi dasar dilakukannya pengembangan model hybrid SOM–LSTM pada tahap berikutnya untuk memperbaiki ketepatan klasifikasi sentimen.

Performa Hybrid SOM + LSTM

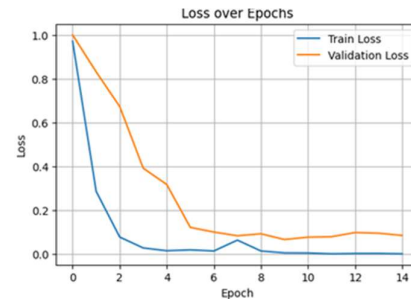
Pada tahap berikutnya, dilakukan pengembangan model *hybrid Self-Organizing Map – Long Short-Term Memory* (SOM–LSTM) dengan tujuan meningkatkan akurasi dan stabilitas klasifikasi sentimen. Model SOM berperan dalam melakukan unsupervised *clustering* terhadap fitur sentimen sehingga menghasilkan representasi topologis dua dimensi. Hasil dari proses ini digunakan sebagai input feature tambahan bagi model LSTM untuk memperkuat pemahaman konteks antar kata dalam teks.



Gambar 7. Accuracy Over Epoch SOM+LSTM

Gambar 7 menunjukkan perubahan akurasi model SOM+LSTM pada data pelatihan (*train accuracy*) dan data validasi (*validation accuracy*) selama proses pembelajaran. Pada awal pelatihan, akurasi *train* berada sekitar 0,58 dan meningkat secara signifikan hingga mendekati 0,98 pada epoch ke-3. Setelah itu, akurasi pelatihan terus meningkat dan berada pada kisaran 0,99 hingga 1,00 pada epoch berikutnya. Akurasi validasi juga menunjukkan peningkatan dari sekitar 0,65 pada awal pelatihan hingga mencapai sekitar 0,97 pada epoch ke-3, kemudian berfluktuasi sebelum berada pada kisaran 0,98–0,99 pada epoch akhir. Perbedaan antara akurasi

pelatihan dan validasi pada beberapa epoch menunjukkan adanya variasi kinerja selama proses pembelajaran, tetapi kedua nilai tersebut tetap berada pada tingkat yang tinggi pada akhir pelatihan. Hasil ini menunjukkan bahwa kombinasi SOM dan LSTM mampu menghasilkan kinerja klasifikasi yang tinggi dan relatif stabil pada data validasi.



Gambar 8. Accuracy over epoch SOM+LSTM

Berdasarkan Gambar 8, grafik *Loss over Epochs* memperlihatkan hasil yang stabil, dengan penurunan *training loss* hingga di bawah 0.1 dan *validation loss* yang konvergen di kisaran 0.15. Perbedaan yang sangat kecil antara *train loss* dan *validation loss* menandakan bahwa model memiliki generalisasi yang baik terhadap data uji [22].

Hasil evaluasi kuantitatif menunjukkan peningkatan signifikan dibandingkan LSTM murni, sebagaimana ditunjukkan pada tabel berikut:

Tabel 2. Hasil Evaluasi Model SOM+ LSTM

Sentiment	Precision	Recall	F1-Score
Negative	0,98	0,99	0,99
Neutral	0,99	0,97	0,98
Positive	0,99	0,98	0,98

Rata-rata *macro average* untuk *precision*, *recall*, dan *F1-score* masing-masing sebesar 0,99, 0,98, dan 0,98, sedangkan *weighted average* juga mencapai 0,98. Secara keseluruhan, akurasi model SOM–LSTM mencapai 0,98.

Hasil ini membuktikan bahwa integrasi SOM sebelum LSTM secara signifikan meningkatkan kemampuan model dalam menangkap pola semantik dan hubungan non-linear antar kata. Model *hybrid* ini tidak hanya lebih akurat, tetapi juga menunjukkan

stabilitas pelatihan yang lebih baik dibandingkan model LSTM murni.

Perbandingan Performa Model LSTM dan SOM–LSTM

Untuk menilai peningkatan kinerja yang dihasilkan oleh penerapan metode hybrid, dilakukan perbandingan antara model LSTM murni dan model SOM + LSTM berdasarkan metrik evaluasi precision, recall, F1-score, dan akurasi. Hasil komparasi ditunjukkan pada tabel 3 berikut:

Tabel 3. Hasil komparasi Evaluasi Performa

Model	Pesisi (%)	Recall (%)	F1-Score (%)	Akurasi (%)
LSTM	84	83	83	85
SOM+ LSTM	99	98	98	98

Dari tabel di atas terlihat bahwa model SOM–LSTM memberikan peningkatan performa yang signifikan dibandingkan LSTM murni. Nilai akurasi meningkat dari 85% menjadi 98%, dengan peningkatan serupa pada seluruh metrik lainnya.

Peningkatan tersebut menunjukkan bahwa penerapan *Self-Organizing Map* (SOM) sebelum LSTM membantu model dalam menemukan representasi fitur yang lebih optimal melalui proses clustering unsupervised. Kombinasi ini membuat model lebih sensitif terhadap variasi konteks dalam teks dan mampu mengurangi kesalahan klasifikasi, terutama pada kategori sentimen netral yang sebelumnya memiliki performa rendah. Dengan demikian, integrasi SOM terbukti mampu memperkuat kemampuan LSTM dalam menganalisis pola semantik secara mendalam dan meningkatkan generalisasi model terhadap data uji.

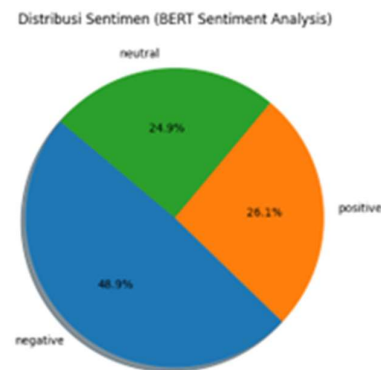
Analisis dan Interpretasi Data

Bagian ini menjelaskan hasil pengolahan data yang diperoleh melalui tahapan pemodelan dan analisis sentimen terhadap DeepSeek AI menggunakan pendekatan BERT *Sentiment Analysis* serta model SOM–LSTM. Analisis dilakukan untuk menggambarkan persepsi publik terhadap DeepSeek AI berdasarkan data besar (big

data) yang dikumpulkan dari platform media sosial Twitter (X). Tahapan analisis mencakup tiga bagian utama. Pertama, analisis distribusi sentimen yang bertujuan mengidentifikasi proporsi opini publik dalam kategori positif, negatif, dan netral. Kedua, analisis kontekstual berbasis *WordCloud* dan frekuensi kata, yang digunakan untuk mengetahui kata-kata dominan pada masing-masing kategori sentimen sehingga dapat memberikan gambaran konteks opini pengguna. Ketiga, interpretasi hasil dan implikasi, yang membahas makna dari temuan tersebut serta hubungannya dengan kinerja model SOM–LSTM dan dampaknya terhadap pemahaman publik terhadap teknologi kecerdasan buatan.

Distribusi Sentimen

Berdasarkan hasil klasifikasi awal menggunakan model BERT *Sentiment Analysis*, diperoleh distribusi sentimen publik terhadap DeepSeek AI sebagaimana ditampilkan pada gambar 9. Dari keseluruhan data yang dianalisis, sebanyak 48,9 % cuitan tergolong negatif, 26,1% bersentimen positif, dan 24,9% bersentimen netral.



Gambar 9. Distribusi Sentiment

Hasil ini menunjukkan bahwa persepsi publik terhadap DeepSeek AI di media sosial Twitter (X) masih didominasi oleh opini bernada negatif. Kecenderungan ini kemungkinan besar dipengaruhi oleh kekhawatiran pengguna terkait isu keamanan data, bias algoritma, serta reliabilitas sistem AI yang masih sering menjadi perdebatan di ruang digital.

dan akurat. Hal ini menunjukkan bahwa integrasi SOM membantu LSTM memahami konteks semantik dengan lebih baik sehingga memperkuat analisis terhadap opini publik. Temuan ini dapat menjadi bahan pertimbangan bagi pengembang, peneliti, maupun regulator dalam memahami persepsi masyarakat terhadap teknologi kecerdasan buatan. Pemahaman terhadap persepsi ini penting untuk meningkatkan kepercayaan, transparansi, serta etika penggunaan AI di masa depan.

4. Kesimpulan

Penelitian ini mengembangkan model *hybrid Self-Organizing Map-Long Short-Term Memory* (SOM-LSTM) untuk klasifikasi sentimen publik terhadap DeepSeek AI berdasarkan data Twitter (X). Hasil pengujian menunjukkan bahwa model SOM-LSTM memperoleh *accuracy* sebesar 0,98, *precision* sebesar 0,99, *recall* sebesar 0,98, dan *F1-score* sebesar 0,98, sedangkan model LSTM memperoleh *accuracy* sebesar 0,85. Hasil tersebut menunjukkan bahwa integrasi SOM sebagai tahap pengelompokan representasi fitur sebelum proses pembelajaran sekuensial menggunakan LSTM memberikan kinerja klasifikasi yang lebih tinggi pada dataset penelitian.

Hasil analisis sentimen menunjukkan bahwa data didominasi oleh sentimen negatif, yang menggambarkan adanya kecenderungan opini negatif terhadap DeepSeek AI, meskipun terdapat pula sentimen positif yang mencerminkan pandangan terhadap potensi pemanfaatan teknologi tersebut. Temuan ini memberikan gambaran empiris mengenai kecenderungan persepsi publik terhadap DeepSeek AI pada Twitter (X).

Penelitian ini memiliki keterbatasan pada penggunaan 10.000 data yang dikumpulkan berdasarkan kata kunci “DeepSeek AI” pada periode Januari 2024–Juli 2025 serta penggunaan pelabelan otomatis berbasis BERT. Penelitian selanjutnya dapat memperluas jumlah dan periode data, menggunakan sumber media sosial yang lebih beragam, serta melakukan

validasi label melalui anotasi manual. Eksplorasi metode representasi fitur dan arsitektur model lainnya juga dapat dilakukan untuk menguji konsistensi kinerja SOM-LSTM pada dataset dan domain yang berbeda.

Daftar Pustaka

- [1] D. Guo et al., “DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning,” *Nature*, vol. 645, no. 8081, pp. 633–638, 2025, <http://dx.doi.org/10.1038/s41586-025-09422-z>.
- [2] Y. Zhang, S. Wang, and B. Liu, “Deep learning for sentiment analysis: A survey,” *WIREs Data Mining and Knowledge Discovery*, vol. 8, no. 4, 2018. <https://doi.org/10.1002/widm.1253>.
- [3] S. Rosenthal et al., “SemEval-2015 Task 10: Sentiment analysis in Twitter,” *Proceedings of the 9th International Workshop on Semantic Evaluation*, pp. 451–463, 2015. <https://doi.org/10.18653/v1/S15-2078>.
- [4] S. Nakov et al., “SemEval-2016 Task 4: Sentiment analysis in Twitter,” *Proceedings of the 10th International Workshop on Semantic Evaluation*, pp. 1–18, 2016. <https://doi.org/10.18653/v1/S16-1001>.
- [5] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- [6] M. Vesanto and E. Alhoniemi, “Clustering of the self-organizing map,” *IEEE Transactions on Neural Networks*, vol. 11, no. 3, pp. 586–600, 2000. <https://doi.org/10.1109/72.846731>.
- [7] A. M. Isa, S. Kallimani, and V. S. Lee, “Using self-organizing map for text document clustering,” *Expert Systems with Applications*, 2008.

- <https://doi.org/10.1016/j.eswa.2008.07.082>.
- [8] M. Jahin et al., “A hybrid transformer and attention-based BiLSTM model for sentiment analysis of tweets,” *Scientific Reports*, 2024. <https://doi.org/10.1038/s41598-024-76079-5>
- [9] J. Pennington, R. Socher, and C. D. Manning, “GloVe: Global vectors for word representation,” *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1532–1543, 2014. <https://doi.org/10.3115/v1/D14-1162>.
- [10] J. Devlin et al., “BERT: Pre-training of deep bidirectional transformers for language understanding,” *Proceedings of NAACL-HLT*, pp. 4171–4186, 2019. <https://doi.org/10.18653/v1/N19-1423>.
- [11] D. Jurafsky and J. H. Martin, *Speech and Language Processing*, 3rd ed., online manuscript, 2026.
- [12] S. Islam et al., “A survey on sentiment analysis: Challenges and future directions,” *Artificial Intelligence Review*, 2023. <https://doi.org/10.1007/s10462-022-10386-z>.
- [13] D. Q. Nguyen et al., “BERTweet: A pre-trained language model for English Tweets,” *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 9–14, 2020. <https://doi.org/10.18653/v1/2020.emnlp-demos.2>.
- [14] F. Barbieri et al., “TweetEval: Unified benchmark and comparative evaluation for tweet classification,” *Findings of the Association for Computational Linguistics: EMNLP 2020*, pp. 1644–1650, 2020. <https://doi.org/10.18653/v1/2020.findings-emnlp.148>.
- [15] C. Hutto and E. Gilbert, “VADER: A parsimonious rule-based model for sentiment analysis of social media text,” *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8, no. 1, pp. 216–225, 2014. <https://doi.org/10.1609/icwsm.v8i1.14550>.
- [16] B. Pang and L. Lee, “Opinion mining and sentiment analysis,” *Foundations and Trends in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008. <https://doi.org/10.1561/1500000011>.
- [17] B. Liu, *Sentiment Analysis and Opinion Mining*. Cham, Switzerland: Springer, 2012. <https://doi.org/10.2200/S00416ED1V01Y201204HLT016>.
- [18] A. Severyn and A. Moschitti, “Twitter sentiment analysis with deep convolutional neural networks,” *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 959–962, 2015. <https://doi.org/10.1145/2766462.276783>.
- [19] P. Nakov, “Sentiment analysis in Twitter: A SemEval perspective,” *Proceedings of the 10th International Workshop on Semantic Evaluation*, 2016. <https://doi.org/10.18653/v1/W16-0427>.
- [20] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [21] V. Deska, A. Armanto, and E. Elmayati, “Deteksi Serangan Man-In-The-Middle (MITM) Berbasis Machine Learning pada Dataset CIC IIOT 2025,” *djtech*, vol. 6, no. 2, pp. 168–184, Jul. 2026. <https://doi.org/10.59563/djtech.v6i2.317>
- [22] A. . Syavira, “Empirical Comparison: Analysis of Initial Set-up of BGP Dynamic Routing and Static Routing in Three Administrative Domain Network Interconnection,” *djtech*, vol. 6, no. 2, pp. 115–128, May 2026. <https://doi.org/10.59563/djtech.v6i2.379>