

MODEL *HYBRID* CNN DAN LEXICON-BASED UNTUK MEN-DETEKSI SENTIMEN PUBLIK TERHADAP DANANTARA

(*HYBRID CNN AND LEXICON-BASED MODEL FOR DETECTING PUBLIC SENTIMENT TOWARDS DANANTARA*)

Andri Wijaya¹⁾, dan Thomas Filikano²⁾

^{1,2)}Program Studi Sistem Informasi, Universitas Katolik Musi Charitas
Jl. Bangau No.60 Palembang 30113
e-mail: andri_wijaya@ukmc.ac.id¹⁾, thomasfilikano555@gmail.com²⁾

ABSTRAK

Danantara merupakan dana kekayaan negara yang mendapat perhatian luas di media sosial, termasuk di “X”. Berbagai opini publik tentang danantara muncul, baik dalam sentimen positif maupun negatif. Untuk memahami persepsi masyarakat secara lebih akurat, diperlukan metode analisis sentimen yang efektif. Metode Lexicon-Based sering digunakan tetapi memiliki keterbatasan dalam menangkap konteks, sementara Convolutional Neural Network (CNN) mampu mengenali pola dalam teks tetapi membutuhkan banyak data. Maka dari itu penelitian ini bertujuan untuk mengembangkan model hybrid CNN dan Lexicon-based guna meningkatkan akurasi analisis sentimen publik terhadap Danantara. Data dikumpulkan dari media sosial “X” menggunakan metode scraping atau API dengan hasil dataset sebanyak 4190 baris data dan terdiri dari 4 kolom, kemudian melalui proses preprocessing sebelum dianalisis menggunakan pendekatan hybrid. Evaluasi model dilakukan dengan membandingkan performa CNN, Lexicon-based dan hybrid model menggunakan metrik seperti akurasi, precision, recall dan F1-Score. Hasil penelitian ini memberikan model yang lebih akurat dalam memahami opini publik serta memberikan wawasan bagi pemangku kebijakan dalam merancang strategi komunikasi yang lebih efektif dengan tingkat akurasi sebesar 86%. Selain itu peneliti juga akan berkontribusi dalam pengembangan Natural Language Processing (NLP) dan analisis sentimen berbasis deep learning.

Kata Kunci: analisis sentimen, danantara, convolutional neural network (CNN), lexicon-based, natural language processing (NLP)

ABSTRACT

Danantara is a state-owned fund that has received widespread attention on social media, including on “X.” Various public opinions about Danantara have emerged, both positive and negative. To understand public perception more accurately, an effective sentiment analysis method is needed. The Lexicon-Based method is often used but has limitations in capturing context, while Convolutional Neural Network (CNN) is capable of recognizing patterns in text but requires a lot of data. Therefore, this study aims to develop a hybrid CNN and Lexicon-based model to improve the accuracy of public sentiment analysis towards Danantara. Data was collected from the social media platform “X” using scraping or API methods, resulting in a dataset of 4,190 rows of data consisting of 4 columns, which then underwent preprocessing before being analyzed using a hybrid approach. Model evaluation was carried out by comparing the performance of CNN, lexicon-based, and hybrid models using metrics such as accuracy, precision, recall, and F1-Score. The results of this study provide a more accurate model for understanding public opinion and provide insights for policy makers in designing more effective communication strategies with an accuracy rate of 86%. In addition, researchers will also contribute to the development of Natural Language Processing (NLP) and deep learning-based sentiment analysis.

Keywords: convolutional neural network (CNN), danantara, lexicon-based, natural language processing (NLP), sentiment analysis

I. PENDAHULUAN

Perkembangan teknologi digital telah mendorong meningkatnya interaksi masyarakat di media sosial seperti *Twitter* yang sudah berubah nama menjadi *X*, terutama dalam menyampaikan opini dan sentimen terhadap berbagai isu sosial, politik, dan ekonomi mereka secara *real-*

time dalam bentuk cuitan (*tweets*) [1][2]. Media sosial tidak hanya menjadi ruang diskusi publik, tetapi juga sumber data yang kaya untuk menganalisis persepsi masyarakat terhadap kebijakan dan program pemerintah.

Salah satu topik yang saat ini sedang menjadi perbincangan yang cukup banyak adalah **Danantara**, yang menimbulkan berbagai

pandangan publik yang beragam. Danantara Indonesia adalah dana kekayaan negara (*sovereign wealth fund*) terbaru yang diluncurkan oleh Presiden Prabowo Subianto pada tanggal 24 Februari 2025. Dana ini dirancang untuk mengelola dan mengoptimalkan aset negara dengan tujuan mendorong pertumbuhan ekonomi nasional. Dengan komitmen awal sebesar US\$20 Miliar, Danantara akan berinvestasi dalam sektor-sektor strategis seperti pemrosesan logam, kecerdasan buatan, energi terbarukan, dan produksi pangan. Model operasionalnya terinspirasi oleh Temasek Singapura, dengan rencana untuk mengelola aset senilai lebih dari US\$900 miliar, termasuk perusahaan-perusahaan milik negara utama seperti Bank Mandiri dan Pertamina. Kebijakan ini memunculkan beragam respons masyarakat, baik berupa dukungan maupun kekhawatiran terhadap transparansi dan tata kelolanya. Dinamika opini tersebut menunjukkan pentingnya analisis sentimen untuk memahami kecenderungan persepsi publik secara lebih objektif dan terukur. Meskipun menghadapi kekhawatiran publik terkait potensi mismanajemen, pemerintah menekankan bahwa Danantara akan beroperasi dengan transparansi penuh dan dapat diaudit oleh siapa saja [3].

Meskipun diskusi mengenai Danantara berkembang pesat di media sosial, belum banyak penelitian yang secara sistematis menganalisis sentimen publik terhadap kebijakan ini menggunakan pendekatan komputasional berbasis data. Hal ini menimbulkan kesenjangan penelitian dalam memahami pola opini masyarakat secara kuantitatif dan terstruktur. Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis sentimen publik terhadap Danantara menggunakan metode analisis sentimen berbasis machine learning. Adapun rumusan masalah dalam penelitian ini adalah bagaimana kecenderungan sentimen masyarakat terhadap Danantara yang diekspresikan melalui media sosial X. Sejalan dengan itu, pertanyaan penelitian yang diajukan adalah Bagaimana distribusi sentimen publik (positif, negatif, dan netral) terhadap Danantara di media sosial X? dan Metode analisis sentimen apa yang memberikan performa terbaik dalam mengklasifikasikan opini terkait Danantara?

II. STUDI PUSTAKA

Dalam beberapa tahun terakhir, berbagai metode telah diimplementasikan untuk mengklasifikasikan sentimen pada berbagai ulasan aplikasi maupun opini publik dari berbagai sumber data,

termasuk algoritma *Naive Bayes* [4], *Support Vector Machine* (SVM) [5] dan teknik berbasis pembelajaran mendalam (*deep learning*) [6][7][8] seperti *Long Short-Term Memory* (LSTM) [9][10] dan *Word2Vec*, seperti pada penelitian yang dilakukan oleh Shyahrin et al. menerapkan metode LSTM dan *Word2Vec* untuk menganalisis sentimen dari ulasan aplikasi Ferizy dengan tujuan mengidentifikasi kebutuhan pengguna dan potensi perbaikan layanan [11]. Selain itu, studi oleh Abdillah et al. Menunjukkan perbandingan efektivitas antara *Naive Bayes* dan *K-Nearest Neighbors* dalam menganalisis sentimen pengguna aplikasi Zenius, memberikan wawasan yang berharga untuk penelitian lebih lanjut dalam domain yang sama [12]. Dalam konteks aplikasi GoPay, Riskawati et al. Mencatat pentingnya analisis sentimen untuk umpan balik pengguna yang dapat digunakan untuk meningkatkan layanan [13]. Penelitian oleh Putri dan Ridwan menggunakan SVM untuk memodelkan dan mengklasifikasi sentimen ulasan aplikasi Pospay, menekankan penggunaan algoritma data mining dalam pengolahan data tidak terstruktur untuk mendapatkan efektivitas klasifikasi yang tinggi [14]. Selain itu dalam konteks teknik baru yang sedang berkembang, model *hybird* lebih menunjukkan hasil yang lebih menjanjikan, seperti pada penelitian yang dilakukan oleh Yang et al memperkenalkan model yang mengkombinasikan informasi dari lexicon sentimen dengan arsitektur CNN, dan hasilnya menunjukkan akurasi yang lebih tinggi dalam klasifikasi sentimen dibandingkan dengan metode tradisional yang hanya fokus pada satu aspek [15][15]. Lebih lanjut penelitian *model hybird* yang membahas permasalahan analisis sentimen pada bullet screen yaitu komentar video yang muncul secara real time di platform media sosial, untuk mengatasi masalah ini diusulkan menggunakan model *hybird* RoBERTA dengan CNN yang diintegrasikan dengan *Lexicon Based* sentimen [16][17] Selanjutnya penelitian dengan model *hybird* yang dilakukan oleh Kumar et al yang membahas tentang analisis sentimen tentang ulasan obat karena pada penelitian-penelitian sebelumnya menunjukkan bahwa analisis sentimen ulasan obat dapat memberikan data berharga bagi tenaga kesehatan dan industri farmasi dalam menilai keamanan obat setelah dipasarkan, informasi ini penting untuk melindungi pasien dan meningkatkan kepercayaan terhadap tenaga media.

Untuk itu pada penelitian ini digunakan model *hybird* metode *lexicon based* dan metode *machine learning* [18]. Selain itu penelitian yang dilakukan oleh Aljabri *et al* menekankan pentingnya pembelajaran yang terintegrasi dengan *lexicon* untuk meningkatkan akurasi dalam analisis sentimen di konteks bahasa arab [19]. Selanjutnya penelitian lain yang dilakukan oleh Fauziah *et al* menunjukkan bahwa pendekatan berbasis *lexicon* memiliki potensi untuk diperbaharui dan disesuaikan dengan model deep learning dalam konteks bahasa Indonesia, yang meningkatkan relevansi hasil analisis terhadap konteks spesifik [20]. Dengan demikian, kombinasi antara model berbasis CNN dan pendekatan *lexicon* tidak hanya memperkuat pemahaman terhadap emosi tersirat dalam teks, tetapi juga membuka tantangan baru untuk analisis sentimen di berbagai domain termasuk media sosial dan *e-commerce* [21].

Oleh karena itu dalam penelitian ini mengusulkan model *hybird* [22][23][24] terbaru yang menggabungkan CNN [25][26] dalam memahami pola teks secara mendalam dan pendekatan berbasis *lexicon* [27] karena dari beberapa penelitian sebelumnya model *hybird* [28] ini belum dilakukan sehingga menjadi model terbaru dalam menangkap makna sentimen dari kata-kata spesifik, dan diharapkan model yang dihasilkan mampu meningkatkan akurasi dalam klasifikasi sentimen publik terhadap Danantara.

III. METODE PENELITIAN

Pada bagian ini menjelaskan alur metodologi penelitian yang digunakan secara sistematis dan terstruktur untuk mencapai tujuan penelitian yang ditunjukkan pada Gambar 1. Proses dimulai dari tahap studi literatur sebagai landasan teoritis dan konseptual, dilanjutkan dengan pengumpulan data yang relevan dengan objek penelitian. Data yang diperoleh kemudian melalui tahap *preprocessing* untuk memastikan kualitas dan kesiapan data sebelum dianalisis. Selanjutnya dilakukan analisis sentiment dengan pendekatan CNN dan *Lexicon-based* yang kemudian dikembangkan lebih lanjut melalui pemodelan *hybrid* klasifikasi sentiment dengan CNN+*lexicon-based* dengan pembagian 90% data latih dan 10% data uji. Model klasifikasi sentimen dibangun menggunakan arsitektur Convolutional Neural Network (CNN) berbasis Keras Sequential. Model terdiri dari enam komponen utama: (1) Embedding layer dengan ukuran kosakata 10.000 dan dimensi embedding 100 serta panjang input 100 token, (2) Conv1D dengan 128

filter dan kernel size 5 menggunakan aktivasi ReLU, (3) GlobalMaxPooling1D, (4) Dense layer 64 unit dengan aktivasi ReLU, (5) Dropout sebesar 0,5 untuk mengurangi overfitting, dan (6) output Dense layer 3 unit dengan aktivasi softmax untuk klasifikasi tiga kelas sentimen (negatif, netral, positif). Model dilatih menggunakan optimizer Adam dengan loss function categorical cross-entropy dan metrik evaluasi accuracy. Kinerja model dievaluasi pada tahap evaluasi model diikuti dengan analisis dan interpretasi hasil untuk memahami temuan penelitian secara mendalam. Tahap akhir dari alur ini adalah penyusunan kesimpulan dan rekomendasi sebagai kontribusi ilmiah dan acuan untuk penelitian atau implementasi selanjutnya.



Gambar 1. Metode Penelitian

A. Studi Literatur

Tahap studi literatur merupakan langkah awal yang bertujuan untuk membangun landasan teoritis dan konseptual penelitian. Pada tahap ini dilakukan penelusuran dan kajian terhadap berbagai sumber ilmiah yang relevan, seperti jurnal nasional dan internasional, prosiding konferensi, buku teks serta laporan penelitian terdahulu yang berkaitan dengan analisis sentiment, *deep learning*, *convolutional neural network (CNN)*, dan pendekatan berbasis *lexicon-based*. Studi literatur ini digunakan untuk mengidentifikasi konsep, metode, algoritma, serta temuan-temuan penting yang telah ada, sekaligus menemukan celah penelitian (*research gap*). Hasil dari tahap ini menjadi dasar dalam penentuan metode penelitian, perancangan model *hybrid* CNN+*Lexicon*, serta penyusunan kerangka penelitian yang sistematis dan sesuai dengan tujuan penelitian.

B. Pengumpulan Data

Tahap pengumpulan data dilakukan untuk memperoleh dataset yang relevan dengan tujuan penelitian. Data dikumpulkan pada tahun 2025 melalui media sosial X dengan menggunakan kata kunci “**danantara**” pada fitur pencarian live. Proses pengambilan data dilakukan secara otomatis

menggunakan teknik web scraping, sehingga diperoleh tweet dari pengguna yang mengandung kata kunci tersebut. Dataset yang dihasilkan tersimpan dalam format comma separated value (CSV) dengan jumlah 4.190 baris dan 4 atribut, yaitu username, isi tweet, URL, dan timestamp. Data yang diperoleh masih bersifat mentah dan bervariasi, sehingga memerlukan tahapan preprocessing sebelum dilakukan analisis sentimen dan pemodelan klasifikasi.

C. Preprocessing Data

Tahap *preprocessing data* bertujuan untuk menyiapkan data mentah agar siap digunakan dalam proses analisis dan pemodelan. Tahap ini dilakukan pembersihan dan transformasi data seperti penghapusan *noise*, duplikasi dan data tidak relevan, normalisasi teks, *case folding*, *tokenisasi*, penghapusan *stopwords* serta proses *stemming* atau *lemmatization*.

D. Analisis Sentimen

Tahap analisis sentimen dilakukan untuk mengidentifikasi dan mengklasifikasikan opini atau persepsi dalam data teks ke dalam kategori sentimen tertentu, seperti positif, negatif, atau netral. Pada tahap ini digunakan pendekatan CNN dan leksikon sentimen, di mana CNN berperan dalam mengekstraksi fitur dan pola kompleks dari data teks, sementara leksikon sentimen digunakan untuk memperkuat informasi polaritas berdasarkan kata-kata bermuatan emosi. Hasil dari tahap ini menjadi dasar dalam pembentukan dan pengembangan model klasifikasi sentimen pada tahap selanjutnya.

E. Pemodelan Hybrid CNN + Lexicon-Based

Tahap pemodelan hybrid klasifikasi sentimen (CNN+Lexicon) bertujuan untuk mengintegrasikan keunggulan pendekatan *deep learning* dan berbasis pengetahuan. Pada tahap ini, fitur hasil ekstraksi CNN dikombinasikan dengan skor atau bobot sentimen dari metode leksikon untuk membentuk representasi data yang lebih komprehensif. Model hybrid ini diharapkan mampu menangkap pola kontekstual yang kompleks sekaligus memperhatikan polaritas kata secara eksplisit, sehingga dapat meningkatkan akurasi dan keandalan dalam mengklasifikasikan sentimen.

F. Evaluasi Model

Tahap evaluasi model dilakukan untuk mengukur kinerja dan efektivitas model klasifikasi sentimen yang telah dibangun. Pada tahap ini, per-

forma model dievaluasi menggunakan metrik evaluasi yang relevan, seperti akurasi, presisi, *recall*, dan *F1-score*. Hasil evaluasi digunakan untuk menilai kemampuan model dalam mengklasifikasikan sentimen secara tepat, serta sebagai dasar untuk membandingkan performa model hybrid CNN+Lexicon dengan pendekatan lain dan melakukan perbaikan jika diperlukan.

G. Analisis dan Interpretasi Hasil

Tahap analisis dan interpretasi hasil bertujuan untuk mengkaji dan memahami keluaran dari proses evaluasi model secara mendalam. Pada tahap ini dilakukan analisis terhadap nilai metrik evaluasi, pola kesalahan klasifikasi, serta kecenderungan hasil prediksi sentimen yang dihasilkan oleh model. Hasil analisis kemudian diinterpretasikan untuk menjelaskan makna temuan penelitian, menilai keunggulan dan keterbatasan model hybrid CNN+Lexicon, serta mengaitkannya dengan tujuan penelitian dan temuan pada studi literatur.

H. Kesimpulan dan Rekomendasi

Tahap kesimpulan dan rekomendasi merupakan tahap akhir yang merangkum seluruh hasil penelitian berdasarkan analisis dan interpretasi yang telah dilakukan. Pada tahap ini disusun kesimpulan utama terkait kinerja dan efektivitas model hybrid CNN+Lexicon dalam klasifikasi sentimen. Selain itu, diberikan rekomendasi yang mencakup saran pengembangan model, perluasan dataset, maupun penerapan metode pada konteks atau studi lanjutan, sehingga hasil penelitian dapat memberikan kontribusi praktis dan menjadi acuan bagi penelitian selanjutnya.

IV. HASIL DAN PEMBAHASAN

Pada bagian ini menyajikan hasil yang diperoleh dari seluruh tahapan penelitian, khususnya terkait proses analisis sentiment, pemodelan klasifikasi (CNN+Lexicon) dan kinerja model klasifikasi sentimen yang diusulkan. Pada bagian ini ditampilkan hasil analisis sentiment, pemodelan dan pengujian serta evaluasi model, diikuti dengan pembahasan yang menginterpretasikan temuan tersebut secara kritis. Pembahasan dilakukan dengan mengaitkan hasil penelitian dengan tujuan yang telah ditetapkan, metode yang digunakan, serta temuan pada studi literatur, sehingga dapat menjelaskan kelebihan, keterbatasan, dan implikasi dari model hybrid CNN+Lexicon yang dikembangkan.

A. Hasil Ekperimen

Pengumpulan data yang dilakukan melalui metode data *scraping* dari media social *twitter* dengan menggunakan *Application Programming Interface* (API). Dari *twitter* data yang dikumpulkan dalam bentuk format *comma separated values* (CSV) yang berisikan komentar-komentar publik yang biasa disebut dengan *tweet* terkait dengan. Data yang dikumpulkan dari media social mengandung kata kunci “*Danantara*” serta dengan waktu penarikan data tahun 2025. Data yang terkumpul sebanyak 4190 baris data dan terdiri dari 4 atribut. Data 10 teratas yang berhasil di *scraping* dapat dilihat pada gambar 2 berikut ini.

index	username	tweet	url	terakhir
0	gayoisme	Danantara menjadi inovasi dalam pengelolaan aset negara, agar lebih produktif dan berdaya saing optimal bagi perekonomian nasional.	https://x.com/gayoismefit/1925449543091622	2025-04-16T03:22:57.000Z
1	gayoisme	Danantara dibentuk untuk memperkuat ekonomi Indonesia di masa depan.	https://x.com/gayoismefit/1925449543091622	2025-04-16T03:22:57.000Z
2	gayoisme	Danantara adalah konsolidasi semua kekuatan ekonomi kita.	https://x.com/gayoismefit/1925449543091622	2025-04-16T03:21:57.000Z
3	gayoisme	Danantara merupakan langkah strategis untuk mengatur aset negara agar lebih efektif dan efisien.	https://x.com/gayoismefit/1925449543091622	2025-04-16T03:21:57.000Z
4	gayoisme	Pembentukan BPI Danantara merupakan hal positif karena diharapkan dapat meningkatkan kinerja Pertamina ke depan.	https://x.com/gayoismefit/1925449543091622	2025-04-16T03:17:56.000Z
5	gayoisme	Dari Danantara Indonesia mendunia.	https://x.com/gayoismefit/1925449543091622	2025-04-16T03:16:48.000Z
6	jawahin	Danantara. A name that sounds like a spell to unite the universe. But what it unites. Beams, of electricity and signaling. They say efficiency. But isn't bringing everything together the first step towards real chaos?	https://x.com/jawahin/1925449543091622	2025-04-16T03:16:21.000Z
7	gayoisme	BPI Danantara dapat mendorong investasi di berbagai sektor, termasuk dalam industri.	https://x.com/gayoismefit/1925449543091622	2025-04-16T03:15:48.000Z
8	katasapoes	Terkait sinnya, ya benar sih bloop hal yg skrg dilakukan ya g. Danantara, Koperasi udah diungkap di sin sebagai "solusi". Tp ada paradoks yg, yg dibuang sebagai "masalah" di sini (g. napaisme, penelestan) mesi kita banyar (di bahas terkait dgn jys-grmwa ya).	https://x.com/katasapoes/1925449543091622	2025-04-16T03:15:19.000Z
9	idc_channel	Presiden Prabowo Subianto mengklaim jika Qatar siap kerja sama dengan Danantara.	https://x.com/idc_channel/1925449543091622	2025-04-16T03:15:18.000Z

Gambar 2. Hasil Scraping dari Media Sosial

Selanjutnya pada preprocessing data dilakukan proses *translate data* terlebih dahulu yaitu dengan melakukan perubahan dari *tweet* bahasa Indonesia ke bahasa Inggris, dikarenakan kamus yang akan digunakan basisnya berbahasa Inggris. Hasil dari *translate* dapat dilihat pada gambar 3 berikut ini.

index	username	tweet	translate
0	gayoisme	Danantara menjadi inovasi dalam pengelolaan aset negara, agar lebih produktif dan berdaya saing optimal bagi perekonomian nasional.	Danantara is an innovation in managing state assets, to be more productive and optimal competitive for the national economy.
1	gayoisme	Danantara dibentuk untuk memperkuat ekonomi Indonesia di masa depan.	Danantara was formed to strengthen Indonesia's economy in the future.
2	gayoisme	Danantara adalah konsolidasi semua kekuatan ekonomi kita.	Danantara is the consolidation of all our economic power.
3	gayoisme	Danantara merupakan langkah strategis untuk mengatur aset negara agar lebih efektif dan efisien.	Danantara is a strategic step to regulate state assets to be more effective and directed.
4	gayoisme	Pembentukan BPI Danantara merupakan hal positif karena diharapkan dapat meningkatkan kinerja Pertamina ke depan.	The formation of BPI Danantara is a positive thing because it is expected to catalyze Pertamina's performance going forward.
5	gayoisme	Dari Danantara Indonesia mendunia.	from danantara Indonesia worldwide.
6	jawahin	Danantara. A name that sounds like a spell to unite the universe. But what it unites. Beams, of electricity and signaling. They say efficiency. But isn't bringing everything together the first step towards real chaos?	danantara. A name that sounds like a spell to unite the universe. but what it unites. beams, of electricity and signaling. they say efficiency. but isn't bringing everything together the first step towards real chaos?
7	gayoisme	BPI Danantara dapat mendorong investasi di berbagai sektor, termasuk dalam industri.	BPI Danantara can encourage investment in various sectors, including downstream industry.
8	katasapoes	Terkait sinnya, ya benar sih bloop hal yg skrg dilakukan ya g. Danantara, Koperasi udah diungkap di sin sebagai "solusi". Tp ada paradoks yg, yg dibuang sebagai "masalah" di sini (g. napaisme, penelestan) mesi kita banyar (di bahas terkait dgn jys-grmwa ya).	Regarding its contents, it's true that the things that are done now (E.G. Danantara, Cooperatives) have been revealed here as a "solution" but there is a paradox also, which is said to be a "problem" here (in g. napaisme, equity) still continues and even related to what?
9	idc_channel	Presiden Prabowo Subianto mengklaim jika Qatar siap kerja sama dengan Danantara.	President Prabowo Subianto claims that Qatar is ready to cooperate with danantara.

Gambar 3. Hasil Translate Dataset

Kemudian dilakukan preprocessing lanjutan dengan *lowercase*, *hapus URL*, *hapus tanda mention*, *hapus tanda hashtag*, *non-ASCII*, membersihkan angka dan symbol, spasi kosong, serta dilakukan *tokenisasi*, *stopword* dan *lemmatization*. Dilanjutkan dengan pembersihan kolom yang memiliki nilai yang kosong (null values) dan penghapusan kata yang kosong (empty string). Hasil dari preprocessing ini dapat dilihat pada gambar 4 berikut ini.

index	username	tweet	translate	text_preprocessing
0	gayoisme	Danantara menjadi inovasi dalam pengelolaan aset negara, agar lebih produktif dan berdaya saing optimal bagi perekonomian nasional.	Danantara is an innovation in managing state assets, to be more productive and optimal competitive for the national economy.	danantara innovation manage state asset productive optimal competitive national economy
1	gayoisme	Danantara dibentuk untuk memperkuat ekonomi Indonesia di masa depan.	Danantara was formed to strengthen Indonesia's economy in the future.	danantara form strengthen indonesia economy future
2	gayoisme	Danantara adalah konsolidasi semua kekuatan ekonomi kita.	Danantara is the consolidation of all our economic power.	danantara consolidation economic power
3	gayoisme	Danantara merupakan langkah strategis untuk mengatur aset negara agar lebih efektif dan efisien.	Danantara is a strategic step to regulate state assets to be more effective and directed.	danantara strategic step regulate state asset effective direct
4	gayoisme	Pembentukan BPI Danantara merupakan hal positif karena diharapkan dapat meningkatkan kinerja Pertamina ke depan.	The formation of BPI Danantara is a positive thing because it is expected to catalyze Pertamina's performance going forward.	formation bpi danantara positive thing expect catalyze Pertamina performance going forward
5	gayoisme	Dari Danantara Indonesia mendunia.	from danantara Indonesia worldwide.	danantara Indonesia worldwide
6	jawahin	Danantara. A name that sounds like a spell to unite the universe. But what it unites. Beams, of electricity and signaling. They say efficiency. But isn't bringing everything together the first step towards real chaos?	danantara. A name that sounds like a spell to unite the universe. but what it unites. beams, of electricity and signaling. they say efficiency. but isn't bringing everything together the first step towards real chaos?	danantara name sound like spell unite universe. but what it unites beams of electricity signal say efficiency bring everything together first step towards real chaos?
7	gayoisme	BPI Danantara dapat mendorong investasi di berbagai sektor, termasuk dalam industri.	BPI Danantara can encourage investment in various sectors, including downstream industry.	bpi danantara encourage investment various sector include downstream industry
8	katasapoes	Terkait sinnya, ya benar sih bloop hal yg skrg dilakukan ya g. Danantara, Koperasi udah diungkap di sin sebagai "solusi". Tp ada paradoks yg, yg dibuang sebagai "masalah" di sini (g. napaisme, penelestan) mesi kita banyar (di bahas terkait dgn jys-grmwa ya).	Regarding its contents, it's true that the things that are done now (E.G. Danantara, Cooperatives) have been revealed here as a "solution" but there is a paradox also, which is said to be a "problem" here (in g. napaisme, equity) still continues and even related to what?	regard content true thing do g danantara cooperative reveal solution paradox also say problem g napaisme equity still continue even relate
9	idc_channel	Presiden Prabowo Subianto mengklaim jika Qatar siap kerja sama dengan Danantara.	President Prabowo Subianto claims that Qatar is ready to cooperate with danantara.	president prabowo subianto claim qatar ready cooperate danantara

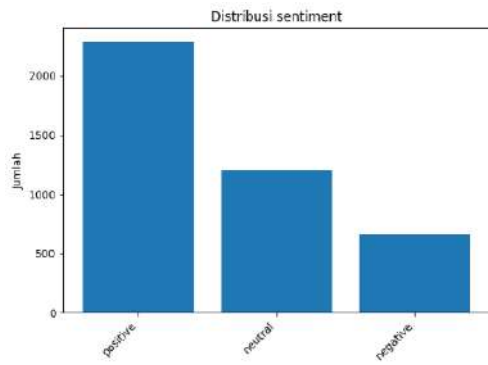
Gambar 4. Hasil Preprocessing Data

Tahapan yang utama yaitu analisis sentimen dengan sebagai masukan berupa dataset yang sudah bersih siap untuk dilakukan analisis data yaitu melakukan Analisis sentiment dengan melakukan labeling untuk setiap komentar yang ada termasuk komentar positif, netral atau negative. Hasil dari analisis sentiment ini memberikan penambahan kolom baru pada dataset yaitu kolom sentiment yang berisikan hasil analisis sentiment dari komentar. Pada gambar 5 menampilkan hasil analisis sentiment.

username	tweet	translate	text_preprocessing	sentiment
0	gayoisme	Danantara menjadi inovasi dalam pengelolaan as...	danantara innovation manage state asset produc...	positive
1	gayoisme	Danantara dibentuk untuk memperkuat ekonomi in...	danantara form strengthen indonesia economy fu...	positive
2	gayoisme	Danantara adalah konsolidasi semua kekuatan ek...	danantara consolidation economic power...	neutral
3	gayoisme	Danantara merupakan langkah strategis untuk me...	danantara strategic step regulate state asset ...	positive
4	gayoisme	Pembentukan BPI Danantara merupakan hal posit...	formation bpi danantara positive thing expect...	positive
5	gayoisme	Dari Danantara Indonesia mendunia.	danantara Indonesia worldwide	neutral
6	jawahin	Danantara. A name that sounds like a spell ...	danantara name sound like spell unite universe...	positive
7	gayoisme	BPI Danantara dapat mendorong investasi di ber...	bpi danantara encourage investment various sec...	positive
8	katasapoes	Terkait sinnya, ya benar sih bloop hal yg skrg ...	regard content true thing do e g danantara coo...	positive
9	idc_channel	Presiden Prabowo Subianto mengklaim jika Qatar...	president prabowo subianto claim qatar ready c...	positive

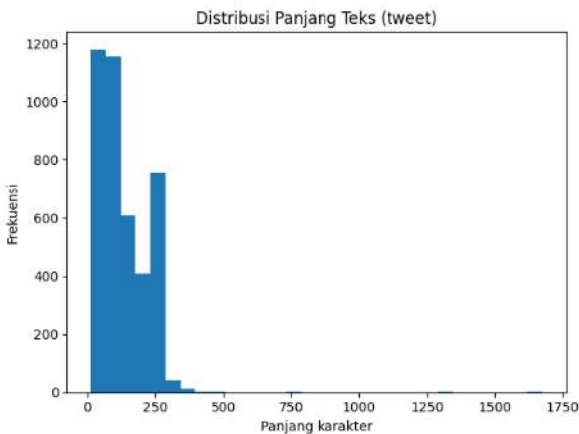
Gambar 5. Hasil Analisis Sentiment 10 Teratas

Pada gambar 6 adalah Grafik distribusi sentimen menunjukkan bahwa mayoritas data berada pada kategori **positif**, dengan jumlah yang paling dominan dibandingkan kategori lainnya, yang mengindikasikan bahwa persepsi atau opini terhadap topik yang dianalisis cenderung mengarah pada respons yang mendukung atau bernada baik. Kategori **netral** menempati urutan kedua dengan jumlah yang cukup signifikan, mencerminkan adanya sebagian data yang bersifat informatif atau tidak menunjukkan kecenderungan emosi yang kuat. Sementara itu, kategori **negatif** memiliki jumlah paling sedikit, yang menunjukkan bahwa respons bernada kritik, ketidakpuasan, atau sentimen buruk relatif lebih rendah dibandingkan sentimen positif dan netral. Secara keseluruhan, pola ini mengindikasikan kecenderungan sentimen publik yang lebih konstruktif dan stabil terhadap topik yang dikaji.



Gambar 6. Hasil Distribusi Sentiment

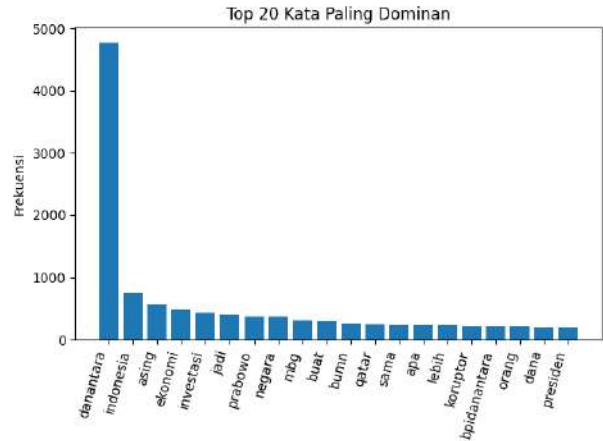
Gambar 7 merupakan grafik distribusi panjang teks menunjukkan bahwa sebagian besar data memiliki panjang teks yang relatif pendek hingga menengah, dengan konsentrasi frekuensi tertinggi berada pada rentang karakter yang rendah. Hal ini mengindikasikan bahwa mayoritas teks berupa cuitan atau pernyataan singkat yang ringkas dan langsung pada inti pesan, sesuai dengan karakteristik media sosial. Seiring bertambahnya panjang karakter, frekuensi data menurun secara signifikan, menunjukkan bahwa hanya sedikit teks yang memiliki panjang sangat panjang. Pola distribusi yang condong ke kanan (right-skewed) ini menegaskan bahwa dataset didominasi oleh teks pendek, yang penting untuk dipertimbangkan dalam tahap praproses dan pemodelan, khususnya dalam penentuan batas panjang input pada metode analisis berbasis *Natural Language Processing* (NLP).



Gambar 7. Hasil Distribusi Panjang Tweet

Pada gambar 8 adalah grafik Top 20 kata paling sering menunjukkan bahwa kata “danantara” muncul dengan frekuensi paling tinggi dan sangat dominan dibandingkan kata lainnya, yang mengindikasikan bahwa topik utama dalam dataset berfokus kuat pada isu atau entitas tersebut. Kemunculan kata-kata tematik seperti “indonesia”, “asing”, “ekonomi”, “investasi”, dan “negara”

menunjukkan bahwa pembahasan dalam teks berkaitan erat dengan isu ekonomi, kebijakan, dan konteks nasional. Pola ini memberikan gambaran awal mengenai fokus diskusi dan kata kunci dominan dalam dataset, yang dapat dimanfaatkan lebih lanjut pada tahap analisis topik, ekstraksi fitur, maupun pemodelan sentimen berbasis *Natural Language Processing* (NLP).



Gambar 8. Kata Paling Sering Muncul

Dilanjutkan dengan proses pemodelan klasifikasi menggunakan Algoritma CNN dan Lexicon-Based, berikut hasil dari klasifikasi sentiment pada gambar 9.

Classification Report:

	precision	recall	f1-score	support
negative	0.82	0.60	0.69	68
neutral	0.83	0.88	0.85	120
positive	0.88	0.92	0.90	229
accuracy			0.86	417
macro avg	0.84	0.80	0.82	417
weighted avg	0.85	0.86	0.85	417

Accuracy: 0.8561151079136691

Gambar 9. Hasil Klasifikasi Sentiment dengan CNN+Lexicon-Based

Setelah proses pemodelan dilanjutkan dengan evaluasi dari model menggunakan Confusion matrix pada gambar 10 menunjukkan bahwa model klasifikasi sentimen mampu mengklasifikasikan sebagian besar data ke dalam kelas yang benar, terutama pada kelas positif dan netral. Hal ini terlihat dari nilai diagonal yang relatif tinggi, yaitu sebanyak 210 data positif dan 106 data netral yang berhasil diprediksi dengan tepat. Pada kelas negatif, sebanyak 41 data berhasil diklasifikasikan secara benar, meskipun jumlahnya lebih rendah dibandingkan dua kelas lainnya. Secara umum, nilai diagonal yang dominan menunjukkan bahwa

sektor ekonomi dan investasi. Hal ini menunjukkan adanya fokus diskusi yang cukup terpusat dan konsisten.

WordCloud ini juga menampilkan kata-kata bernuansa kebijakan dan tata kelola seperti “korupsi”, “aset”, “pemerintah”, dan “bumn”, yang mencerminkan adanya diskusi kritis terkait transparansi, pengelolaan sumber daya, serta peran pemerintah dan badan usaha milik negara. Kehadiran kata-kata tersebut menunjukkan bahwa topik yang dibahas tidak hanya bersifat ekonomis, tetapi juga mencakup aspek tata kelola dan akuntabilitas.

Secara keseluruhan, visualisasi WordCloud ini memberikan gambaran awal yang komprehensif mengenai fokus utama dan arah diskursus dalam dataset. Informasi ini sangat berguna sebagai dasar untuk analisis lanjutan, seperti pengelompokan topik (topic modeling), analisis sentimen tematik, maupun ekstraksi fitur untuk pemodelan berbasis Natural Language Processing (NLP), sehingga dapat memperdalam pemahaman terhadap pola dan isu utama yang berkembang dalam data.

V. KESIMPULAN [3]

Berdasarkan hasil analisis yang telah dilakukan, penelitian ini menunjukkan bahwa sentimen publik terhadap Danantara di media sosial X didominasi oleh sentimen positif dan netral, dengan diskursus yang kuat terkait isu ekonomi nasional, investasi, dan kebijakan pemerintah. Hasil pemodelan menggunakan arsitektur Convolutional Neural Network (CNN) menunjukkan kinerja klasifikasi yang baik dengan tingkat akurasi yang tinggi, terutama dalam mengidentifikasi sentimen positif dan netral, meskipun deteksi sentimen negatif masih menjadi tantangan. Kontribusi ilmiah dari penelitian ini terletak pada penerapan pendekatan deep learning berbasis CNN untuk menganalisis sentimen publik terhadap kebijakan strategis nasional yang relatif baru, serta penyajian analisis berbasis data sebagai dasar evaluasi persepsi masyarakat secara kuantitatif. Penelitian ini memperkaya kajian analisis sentimen dalam konteks kebijakan publik di Indonesia dengan memanfaatkan data real-time dari media sosial sebagai sumber informasi yang dinamis dan terukur. Secara praktis, temuan penelitian ini dapat menjadi referensi bagi pembuat kebijakan dalam memahami kecenderungan opini publik serta mengidentifikasi potensi isu sensitif yang berkembang di masyarakat. Untuk penelitian selanjutnya, disarankan untuk memperluas periode dan variasi sumber data, serta mengembangkan

model berbasis embedding kontekstual dan teknik penyeimbangan data guna meningkatkan performa klasifikasi, khususnya pada kelas sentimen negatif.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan dan kontribusi dalam pelaksanaan penelitian ini, baik secara langsung maupun tidak langsung. Ucapan terima kasih juga disampaikan kepada institusi saya Unika Musi Charitas, rekan sejawat, serta pihak-pihak yang telah membantu dalam penyediaan data, diskusi ilmiah, dan masukan yang berharga, sehingga penelitian ini dapat diselesaikan dengan baik.

REFERENSI

- L. Ellyanti, Y. Ruldeviyani, L. E. Pradana, and A. Harjanto, “Sentiment Analysis of Twitter Users to the PeduliLindungi Using Naïve Bayes Algorithm,” *Jurnal RESTI*, vol. 7, no. 2, pp. 414–421, Apr. 2023, doi: 10.29207/resti.v7i2.4684.
- L. Kothamasu and E. Kannan, “Sentiment analysis on twitter data based on spider monkey optimization and deep learning for future prediction of the brands,” *Concurr. Comput.*, vol. 34, Jun. 2022, doi: 10.1002/cpe.7104.
- Stefanno Sulaiman and Ananda Teresia, “New Indonesia sovereign wealth fund to invest \$20 billion in projects,” *https://www.reuters.com/world/asia-pacific/indonesias-prabowo-officially-establishes-new-sovereign-wealth-fund-2025-02-24/*, Jakarta, Feb. 24, 2025.
- K. Kevin, M. Enjeli, and A. Wijaya, “Analisis Sentimen Penggunaan Aplikasi Kinemaster Menggunakan Metode Naive Bayes,” *Jurnal Ilmiah Computer Science*, vol. 2, no. 2, pp. 89–98, Jan. 2024, doi: 10.58602/jics.v2i2.24.
- A. Larasati, Y. R. W. N. Susanto, E. Mohamad, and A. R. Purnama, “A Sentiment Analysis Using Fuzzy Support Vector Machine Algorithm,” *Buletin Ilmiah Sarjana Teknik Elektro*, vol. 5, no. 4, pp. 467–474, Dec. 2023, doi: 10.12928/biste.v5i4.9363.
- L. Yang, Y. Li, J. Wang, and R. S. Sherratt, “Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning,” *IEEE Access*, vol. 8, pp. 23522–23530, 2020, doi: 10.1109/ACCESS.2020.2969854.
- A. Chugh *et al.*, “Spider Monkey Crow Optimization Algorithm with Deep Learning for Sentiment Classification and Information Retrieval,” *IEEE Access*, vol. 9, pp. 24249–24262, 2021, doi: 10.1109/ACCESS.2021.3055507.
- F. Fitroh and F. Hudaya, “Systematic Literature Review: Analisis Sentimen Berbasis Deep Learning,” *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 9, no. 2, pp. 132–140, Aug. 2023, doi: 10.25077/teknosi.v9i2.2023.132-140.

- [9] M. Kamyab, G. Liu, and M. Adjeisah, "Attention-Based CNN and Bi-LSTM Model Based on TF-IDF and GloVe Word Embedding for Sentiment Analysis," *Applied Sciences (Switzerland)*, vol. 11, no. 23, Dec. 2021, doi: 10.3390/app112311255. [25]
- [10] D. Y. Liliana, N. N. Hikmah, and M. Harjono, "Pengembangan Sistem Pemantauan Sentimen Berita Berbahasa Indonesia Berdasarkan Konten Dengan Long Short-Term Memory The Development Of Content-Based Indonesian News Sentiment Monitoring System Using Long Short-Term Memory," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 8, no. 5, pp. 995–1002, Oct. 2021, doi: 10.25126/jtiik.202184624. [26]
- [11] M. V. Shyahrin, Y. Sibaroni, and D. Puspandari, "Penerapan Metode Long Short-Term Memory dan Word2Vec dalam Analisis Sentimen Ulasan pada Aplikasi Feri-Techno.COM," *Techno.COM*, vol. 22, no. 4, pp. 833–842, Nov. 2023. [27]
- [12] T. Abdillah, U. Khaira, and B. F. Hutabarat, "Komparasi Metode Naive Bayes dan K-Nearest Neighbors Terhadap Analisis Sentimen Pengguna Aplikasi Zenius," *Jurnal PROCESSOR*, vol. 19, no. 1, May 2024, doi: 10.33998/processor.2024.19.1.1596. [28]
- [13] Riskawati, Fatihanursari, Iin, and A. R. Rinaldi, "Penerapan Metode Naive Bayes Classifier Pada Analisis Sentimen Aplikasi Gopay," *Jurnal Mahasiswa Teknik Informatika (JATI)*, vol. 8, no. 1, Feb. 2024.
- [14] D. S. Putri and T. Ridwan, "Analisis Sentimen Ulasan Aplikasi Pospay dengan Algoritma Support Vector Machine," *Jurnal Ilmiah Informatika (JIF)*, vol. 11, no. 1, Mar. 2023.
- [15] L. Yang, Y. Li, J. Wang, and R. S. Sherratt, "Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning," *IEEE Access*, vol. 8, Jan. 2020, doi: 10.1109/ACCESS.2020.2969854.
- [16] Y. Liu, S. Wang, and S. Yu, "A Bullet Screen Sentiment Analysis Method That Integrates the Sentiment Lexicon with RoBERTa-CNN," *Electronics (Switzerland)*, vol. 13, no. 20, Oct. 2024, doi: 10.3390/electronics13203984.
- [17] Y. Liu, S. Wang, and S. Yu, "A Bullet Screen Sentiment Analysis Method That Integrates the Sentiment Lexicon with RoBERTa-CNN," *Electronics (Switzerland)*, vol. 13, no. 20, Oct. 2024, doi: 10.3390/electronics13203984.
- [18] A. Kumar and S. Shekhar, "Hybrid Framework Integrating Lexicon and Learning Methods for Enhancing Sentiment Analysis Based on Patients' Tweets on Medicines," *Journal Electrical Systems*, vol. 20, no. 6, pp. 2786–2810, 2024.
- [19] M. Aljabri *et al.*, "Sentiment analysis of arabic tweets regarding distance learning in saudi arabia during the covid-19 pandemic," *Sensors*, vol. 21, no. 16, Aug. 2021, doi: 10.3390/s21165431.
- [20] Y. Fauziah, B. Yuwono, and A. S. Aribowo, "Lexicon Based Sentiment Analysis in Indonesia Languages : A Systematic Literature Review," *RSF Conference Series: Engineering and Technology*, vol. 1, no. 1, pp. 363–367, Dec. 2021, doi: 10.31098/cset.v1i1.397.
- [21] M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowl. Based. Syst.*, vol. 226, Aug. 2021, doi: 10.1016/j.knosys.2021.107134.
- [22] A. Kumar and S. Shekhar, "Hybrid Framework Integrating Lexicon and Learning Methods for Enhancing Sentiment Analysis Based on Patients' Tweets on Medicines," 2024.
- [23] A. P. P. Sitanggang, "Hybrid CNN And RNNS Model For Sentiment Analysis," *PROXIES*, vol. 8, no. 1, 2024.
- S. J. and K. U., "Sentiment analysis of amazon user reviews using a hybrid approach," *Measurement: Sensors*, vol. 27, p. 100790, 2023, doi: <https://doi.org/10.1016/j.measen.2023.100790>.
- Y. Cheng, L. Yao, G. Xiang, G. Zhang, T. Tang, and L. Zhong, "Text Sentiment Orientation Analysis Based on Multi-Channel CNN and Bidirectional GRU with Attention Mechanism," *IEEE Access*, vol. 8, pp. 134964–134975, 2020, doi: 10.1109/ACCESS.2020.3005823.
- T. Yuan, H. H. Li, H. Y. Zhao, Q. H. Cai, H. Liu, and X. H. Hu, "Multi-Channel Convolutional Neural Network for Targeted Sentiment Classification," in *Proceedings - International Conference on Machine Learning and Cybernetics*, IEEE Computer Society, Jul. 2019. doi: 10.1109/ICMLC48188.2019.8949286.
- L. Yang, Y. Li, J. Wang, and R. S. Sherratt, "Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning," *IEEE Access*, vol. 8, pp. 23522–23530, 2020, doi: 10.1109/ACCESS.2020.2969854.
- B. Shin, T. Lee, and J. D. Choi, "Lexicon Integrated CNN Models with Attention for Sentiment Analysis," in *Proceedings of the EMNLP Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, Cornell University, Oct. 2016. [Online]. Available: <http://arxiv.org/abs/1610.06272>