

IMPLEMENTASI MODEL *INDOBERT* UNTUK ANALISIS SENTIMEN PUBLIK TERHADAP ISU DUGAAN IJAZAH PALSU PRESIDEN JOKO WIDODO DALAM PERCAKAPAN DI *PLATFORM X* (TWITTER)

¹Naufal Rafif Ramadhani*, ²Sam Farisa Chaerul Haviana

^{1,2} Teknik Informatika, Fakultas Teknologi Industri, Universitas Islam Sultan Agung

*Corresponding Author:
Naufalrafifr012@std.unissula.ac.id
sam@unissula.ac.id

ABSTRAK

Penelitian ini membahas implementasi model IndoBERT untuk analisis sentimen publik terhadap isu dugaan ijazah palsu Presiden Joko Widodo di Platform X (Twitter). Tujuan penelitian ini adalah mengklasifikasikan opini publik menjadi tiga kategori sentimen, yakni positif, netral, serta negatif, untuk mengetahui kecenderungan opini masyarakat terhadap isu yang tengah berkembang. Metode yang dipergunakan yakni pendekatan kuantitatif dengan eksperimen model IndoBERT-base-p1 yang di-fine-tune menggunakan dataset berbahasa Indonesia. Data dikumpulkan melalui Twitter API v2 dengan kata kunci terkait isu tersebut, kemudian melalui tahapan preprocessing meliputi cleaning, case folding, normalisasi, tokenisasi, filtering, serta stemming. Sebanyak 2.113 tweet digunakan dan diklasifikasikan ke dalam data latih serta data uji dengan perbandingan 70:30. Model dilatih mempergunakan algoritma AdamW optimizer dengan fungsi Cross-Entropy Loss dan dievaluasi mempergunakan metrik accuracy, precision, recall, serta F1-score. Hasil uji memperlihatkan bahwasanya model IndoBERT memiliki akurasi sebesar 91,13% dengan nilai F1-score sejumlah 0,91, menandakan performa klasifikasi yang sangat baik serta stabil. Distribusi hasil analisis memperlihatkan dominasi sentimen negatif sejumlah 61%, disusul sentimen positif sejumlah 23% serta netral sejumlah 16%. Model ini diimplementasikan ke dalam aplikasi web berbasis Streamlit yang memungkinkan pengguna melaksanakan analisis sentimen secara langsung dan menampilkan hasil dalam bentuk visualisasi interaktif berupa pie chart, bar chart, dan wordcloud. Temuan studi ini menunjukkan bahwa IndoBERT mampu memahami konteks opini publik dalam Bahasa Indonesia secara efektif dan dapat dijadikan dasar pengembangan sistem analisis opini publik yang informatif dan aplikatif.

Kata Kunci: Analisis Sentimen, IndoBERT, Opini Publik, Media Sosial, Platform X

Abstract

This research discusses the implementation of the IndoBERT model for sentiment analysis of public opinion regarding the alleged fake diploma issue of President Joko Widodo on Platform X (Twitter). The purpose of this study is to classify public opinions into three sentiment categories, positive, neutral, and negative, in order to identify the general tendency of public perception toward the ongoing issue. The research adopts a quantitative experimental approach using the IndoBERT-base-p1 model, which has been fine-tuned with an Indonesian-language dataset. Data were collected through the Twitter API v2 using relevant keywords related to the diploma issue and then processed through several preprocessing stages, including cleaning, case folding, normalization, tokenization, filtering, and stemming. A total of 2,113 tweets were used, divided into training and testing sets with a 70:30 ratio. The model was trained using the AdamW optimizer algorithm with the Cross-Entropy Loss function and evaluated using accuracy, precision, recall, and F1-score metrics. The evaluation results show that the IndoBERT model achieved an accuracy of 91.13% and an F1-score of 0.91, indicating highly accurate and balanced classification performance. The sentiment distribution results reveal that negative sentiments dominate with 61%, followed by positive sentiments at 23% and neutral sentiments at 16%. The model was implemented in a web-based application using Streamlit, allowing users to perform real-time sentiment analysis and view interactive visualizations such as pie charts, bar charts, and word clouds. Overall, the findings indicate that IndoBERT effectively understands Indonesian-language public opinion and can serve as a foundation for developing an informative and practical public sentiment analysis system.

Keywords: Sentiment Analysis, IndoBERT, Public Opinion, Social Media, Platform X

1. PENDAHULUAN

Perkembangan media sosial di era digital telah mengubah pola komunikasi politik di Indonesia secara signifikan. Platform seperti Twitter, yang kini dikenal sebagai X, menjadi ruang publik virtual tempat masyarakat menyampaikan opini, kritik, serta dukungan terhadap berbagai isu nasional (Yusrizal dan Sasongko, 2024). Dalam konteks ini, Presiden Joko Widodo kerap menjadi pusat perbincangan publik, terutama ketika muncul isu-isu kontroversial yang menyita perhatian masyarakat. Salah satu isu yang ramai diperbincangkan di ruang digital adalah dugaan penggunaan ijazah palsu oleh Presiden. Percakapan mengenai isu tersebut berkembang pesat dan membentuk arus opini publik yang beragam, baik positif, negatif, maupun netral. Sejumlah penelitian menunjukkan bahwa dinamika opini di media sosial, khususnya Twitter, cenderung mencerminkan tingkat kepuasan atau ketidakpuasan masyarakat terhadap tokoh politik dan kebijakan pemerintahan (Fatra dkk., 2024).

Dalam komunikasi politik digital, *framing* atau pembingkaiian informasi memainkan peran penting dalam membentuk persepsi masyarakat. Informasi yang beredar di media sosial tidak hanya dipahami berdasarkan fakta, tetapi juga melalui narasi, sudut pandang, dan cara penyajian yang memengaruhi interpretasi publik (Wenando dkk., 2020). Oleh karena itu, untuk memahami kecenderungan sentimen masyarakat secara lebih objektif, diperlukan pendekatan berbasis data yang mampu mengolah percakapan digital dalam jumlah besar. Salah satu metode yang relevan adalah analisis sentimen, yaitu teknik pemrosesan bahasa alami (*Natural Language Processing/NLP*) yang digunakan untuk

mengklasifikasikan opini ke dalam kategori positif, negatif, atau netral. Metode ini telah banyak digunakan dalam studi politik untuk memetakan persepsi publik terhadap tokoh maupun isu tertentu melalui data media sosial (Utami dan Darmaiza, 2020).

Agar hasil analisis sentimen terhadap teks berbahasa Indonesia lebih akurat, dibutuhkan model bahasa yang mampu memahami konteks lokal, struktur kalimat, serta nuansa linguistik khas Indonesia. Model *IndoBERT* hadir sebagai solusi karena merupakan model *deep learning* berbasis *transformer* yang telah dilatih menggunakan korpus bahasa Indonesia dalam jumlah besar (Putri dkk., 2024). Keunggulan *IndoBERT* terletak pada kemampuannya memahami relasi antar kata dalam konteks kalimat, sehingga mampu menangkap makna implisit, sindiran, maupun ekspresi emosional yang sering muncul dalam percakapan politik di media sosial. Dengan demikian, penerapan *IndoBERT* dalam sistem analisis sentimen dinilai relevan untuk mengklasifikasikan opini publik terkait isu dugaan ijazah palsu Presiden Joko Widodo di platform X (Twitter) (Kominfo, 2023).

Selain untuk mengelompokkan sentimen, analisis ini juga penting dalam mengidentifikasi potensi polarisasi sosial dan konflik digital. Isu politik yang sensitif sering kali memunculkan ujaran bernada tajam, provokatif, bahkan berpotensi mengandung hoaks (MAFINDO, 2023). Dalam kondisi literasi digital masyarakat yang masih perlu ditingkatkan, sebagaimana tercermin dari indeks literasi digital nasional yang belum maksimal, penelitian semacam ini memiliki urgensi akademik dan sosial. Studi ini tidak bertujuan menilai kebenaran atau aspek hukum dari isu yang beredar, melainkan berfokus pada bagaimana opini publik terbentuk dan tersebar di ruang digital.

Berdasarkan latar belakang tersebut, perumusan masalah dalam penelitian ini adalah bagaimana implementasi model *IndoBERT* dalam sistem analisis sentimen dapat memprediksi kategori sentimen positif, negatif, atau netral terhadap opini publik terkait isu dugaan ijazah palsu Presiden Joko Widodo di platform X (Twitter). Penelitian ini dibatasi pada *tweet* berbahasa Indonesia yang dikumpulkan pada periode ketika isu tersebut ramai diperbincangkan, menggunakan model *IndoBERT* sebagai metode klasifikasi, serta hanya mengelompokkan sentimen ke dalam tiga kategori utama.

Tujuan dari penelitian ini adalah untuk membangun dan mengimplementasikan sistem analisis sentimen berbasis *IndoBERT* guna memprediksi kecenderungan opini publik secara lebih terstruktur dan berbasis data. Adapun manfaat yang diharapkan adalah tersedianya gambaran objektif mengenai persepsi masyarakat terhadap isu tersebut, yang dapat menjadi referensi bagi peneliti, media, maupun lembaga terkait dalam memahami dinamika komunikasi politik di era digital. Selain itu, penelitian ini diharapkan dapat berkontribusi pada pengembangan teknologi analisis bahasa Indonesia berbasis *deep learning* serta mendorong peningkatan literasi digital masyarakat agar lebih kritis dan bijak dalam menyikapi informasi politik di media sosial.

2. METODE

Penelitian ini menggunakan jenis penelitian kuantitatif dengan pendekatan eksperimental yang berfokus pada pengujian performa model kecerdasan buatan dalam mengklasifikasikan sentimen opini publik. Pendekatan kuantitatif dipilih karena penelitian ini bertujuan mengukur tingkat ketepatan model *IndoBERT* secara objektif melalui perhitungan metrik evaluasi yang terukur. Sementara itu, pendekatan eksperimental digunakan karena peneliti melakukan proses pelatihan ulang (*fine-tuning*)

terhadap model yang sudah ada, kemudian menguji kinerjanya menggunakan dataset yang telah disiapkan. Dengan pendekatan ini, penelitian tidak hanya mendeskripsikan fenomena opini publik, tetapi juga menguji secara langsung kemampuan model dalam membaca, memahami, dan mengelompokkan teks berdasarkan kecenderungan sentimennya.

Teknik pengumpulan data dilakukan dengan mengambil data sekunder dari media sosial Platform X melalui *Application Programming Interface* versi 2. Data yang dikumpulkan berupa unggahan publik berbahasa Indonesia yang memuat kata kunci relevan dengan isu dugaan ijazah palsu Presiden Joko Widodo, seperti “ijazah Jokowi”, “ijazah palsu Presiden”, dan variasi istilah lain yang memiliki makna serupa. Pengambilan data dibatasi pada periode ketika isu tersebut ramai diperbincangkan agar dataset yang diperoleh benar-benar mencerminkan dinamika opini publik pada saat isu memuncak (Merinda Lestandy dkk., 2021). Proses pengambilan data dilakukan menggunakan bahasa pemrograman *Python* pada lingkungan komputasi berbasis *cloud*, sehingga memungkinkan pengolahan data dalam jumlah besar secara efisien. Setelah data terkumpul, dilakukan proses seleksi untuk menghapus duplikasi dan memastikan setiap teks benar-benar relevan dengan topik penelitian. Selanjutnya, setiap tweet diberi label sentimen secara manual ke dalam tiga kategori, yaitu positif, negatif, dan netral. Pelabelan manual dilakukan agar model memperoleh data latih yang lebih akurat dan sesuai konteks (Merinda Lestandy dkk., 2021).

Sebelum masuk ke tahap pelatihan model, data terlebih dahulu melalui proses pra-pemrosesan. Tahapan ini bertujuan untuk membersihkan teks dari unsur-unsur yang tidak diperlukan dan menyeragamkan format tulisan agar lebih mudah dipahami oleh sistem (Yuniar dkk., 2022).. Proses pra-pemrosesan meliputi pembersihan tanda baca, tautan, emoji, dan simbol khusus, pengubahan seluruh huruf menjadi huruf kecil, normalisasi kata tidak baku menjadi bentuk baku, tokenisasi untuk memecah kalimat menjadi kata-kata, penghapusan kata umum yang tidak memiliki makna penting, serta stemming untuk mengembalikan kata ke bentuk dasarnya. Tahap ini sangat penting karena kualitas data yang bersih dan terstruktur akan sangat mempengaruhi performa model dalam proses klasifikasi (Pahtoni dan Jati, 2024).

Analisis data dilakukan dengan menerapkan model *IndoBERT* yang telah dilatih ulang menggunakan dataset penelitian. Dataset yang sudah diproses kemudian dibagi menjadi dua bagian, yaitu 70 persen data latih dan 30 persen data uji. Data latih digunakan untuk mengajarkan model mengenali pola sentimen, sedangkan data uji digunakan untuk mengukur kemampuan model dalam mengklasifikasikan data baru yang belum pernah dilihat sebelumnya. Proses pelatihan model menggunakan algoritma optimasi *AdamW* dan fungsi kehilangan *Cross Entropy* untuk meminimalkan kesalahan prediksi. Setelah proses pelatihan selesai, model dievaluasi menggunakan confusion matrix untuk menghitung nilai akurasi, presisi, *recall*, dan *F1-score*. Metrik-metrik ini digunakan untuk mengetahui seberapa tepat dan konsisten model dalam mengelompokkan sentimen ke dalam kategori positif, negatif, dan netral. Seluruh proses analisis dilakukan menggunakan bahasa pemrograman *Python* dengan dukungan pustaka pemrosesan bahasa alami, dan hasil akhirnya diimplementasikan dalam bentuk aplikasi web berbasis *Streamlit* agar pengguna dapat melakukan analisis sentimen secara langsung dan melihat visualisasi hasilnya dengan mudah.

3. HASIL DAN PEMBAHASAN

HASIL PENELITIAN

A. Pengumpulan Data

Tahap pertama penelitian ini adalah pengumpulan data (*data collection*) berupa tweet dari Platform X (Twitter) yang membahas isu dugaan ijazah palsu Presiden Joko Widodo. Data yang diambil merupakan unggahan publik dan dikumpulkan menggunakan Twitter API v2 melalui bahasa pemrograman *Python* di *Google Colaboratory*. Penggunaan API memungkinkan pengambilan data secara terstruktur berdasarkan kata kunci, rentang waktu, dan bahasa. Data dikumpulkan pada periode 12 Januari 2025 hingga 14 Agustus 2025, saat isu tersebut ramai diperbincangkan. Kata kunci yang digunakan antara lain “ijazah Jokowi”, “ijazah palsu Presiden”, “ijazah Presiden Joko Widodo”, dan “ijazah palsu Jokowi”. Proses *scraping* menghasilkan 2.262 tweet mentah yang disimpan dalam *file* dataset.xlsx dengan atribut seperti waktu unggah, isi teks, jumlah suka, *retweet*, dan tautan. Setelah dilakukan pembersihan data untuk menghapus spam, duplikasi, dan tweet non-Bahasa Indonesia, diperoleh 2.113 tweet yang layak dianalisis dengan atribut utama seperti *tweet_id*, *created_at*, *full_text*, *favorite_count*, *retweet_count*, dan *tweet_url*.

Tabel 1. Jumlah Tweet Berdasarkan Bulan Pengambilan Data

Bulan	Jumlah Tweet	Persentase (%)
Januari 2025	162	7,7
Februari 2025	183	8,6
Maret 2025	202	9,5
April 2025	258	12,2
Mei 2025	471	22,3
Juni 2025	326	15,4
Juli 2025	301	14,2
Agustus 2025	210	10,1
Total	2.113	100

Merujuk pada tabel tersebut, tampak bahwasanya jumlah *tweet* terbanyak diperoleh pada bulan Mei 2025, yaitu sebanyak 471 *tweet* (22,3%). Peningkatan tersebut menunjukkan bahwa puncak perbincangan isu dugaan ijazah palsu Presiden Joko Widodo terjadi pada bulan tersebut. Sementara itu, jumlah *tweet* menurun secara bertahap setelah bulan Juni hingga Agustus, yang menandakan berkurangnya intensitas diskusi publik terhadap isu tersebut.

B. Implementasi Sistem Model

Pra Pemrosesan Data

Tahap pra-pemrosesan dilakukan untuk membersihkan data mentah agar siap dianalisis oleh model *IndoBERT*. Proses ini meliputi penghapusan elemen non teks seperti *mention*, *hashtag*, URL, angka, tanda baca, dan emoji agar teks menjadi lebih bersih, seragam, dan mudah dipahami secara semantik oleh model.

Tabel 2. Contoh Hasil Pembersihan Teks

Teks Asli	Hasil Pra-Pemrosesan
“RT @user: Jokowi palsu?? https://t.co/xxxx ”	“jokowi palsu”
“Jokowi Hadiri Klarifikasi Dugaan Ijazah Palsu”	“jokowi hadir klarifikasi duga ijazah palsu”
“Megawati singgung ijazah Jokowi palsu?”	“megawati singgung ijazah jokowi palsu”

Tahap selanjutnya meliputi *case folding* (mengubah seluruh huruf menjadi kecil), normalisasi kata tidak baku ke bentuk baku, tokenisasi (memecah kalimat menjadi kata), penghapusan *stopword*, serta stemming menggunakan algoritma Nazief–Adriani (Sastrawi) untuk mengembalikan kata ke bentuk dasar. Seluruh proses ini menghasilkan teks yang lebih bersih dan konsisten untuk analisis.

Tabel 3. Contoh Hasil Akhir Pra-Pemrosesan Teks

No	Teks Asli	Hasil Akhir
1	Jokowi Hadiri Undangan Klarifikasi Terkait Laporan Ijazah Palsu	jokowi hadir undang klarifikasi kait laporan ijazah palsu
2	Megawati Singgung Isu Ijazah Jokowi Palsu	megawati singgung isu ijazah jokowi palsu
3	Isu Ijazah Palsu Jokowi Tidak Benar	isu ijazah palsu jokowi tidak benar

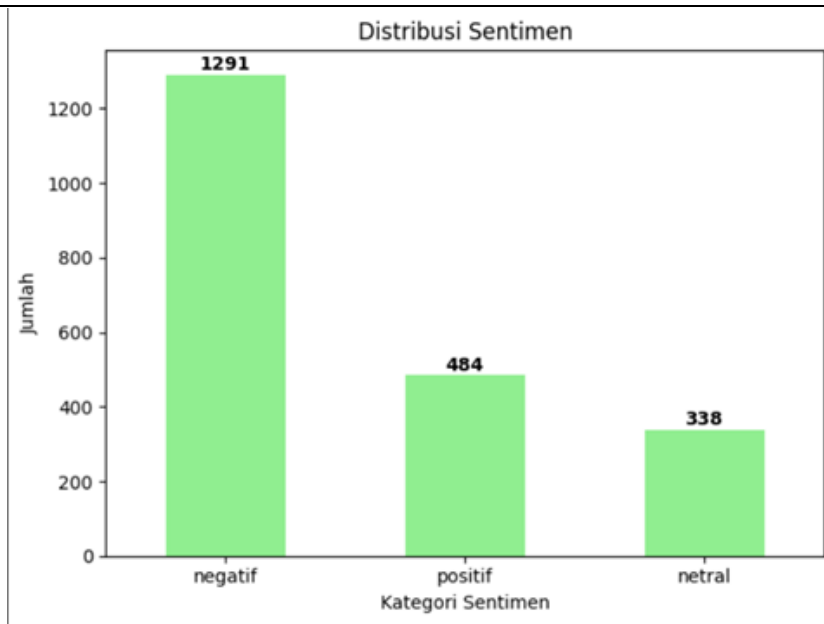
Tahapan pra-pemrosesan ini menghasilkan file akhir hasil_preprocessing.csv yang menjadi dasar untuk proses pelabelan dan pelatihan model *IndoBERT*.

Pelabelan Data

Tahap pelabelan dilakukan untuk mengklasifikasikan tweet ke dalam sentimen positif, negatif, atau netral menggunakan metode *lexicon-based* berdasarkan daftar kata positif dan negatif. Hasil pelabelan otomatis kemudian diverifikasi secara manual untuk memastikan kesesuaian konteks. Distribusi akhir sentimen ditampilkan pada Tabel 4.

Tabel 4. Distribusi Data Sentimen Publik terhadap Isu Dugaan Ijazah Palsu Presiden Joko Widodo

Kategori Sentimen	Jumlah Tweet	Persentase (%)	Keterangan
Negatif	1.291	61,1	Opini yang berisi kritik, ketidakpercayaan, atau tuduhan terhadap isu
Positif	484	22,9	Opini yang bersifat dukungan atau pembelaan terhadap Presiden
Netral	338	16,0	Pernyataan informatif tanpa opini emosional
Total	2.113	100	—



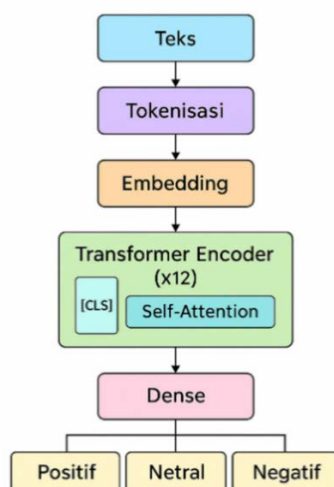
Gambar 1. Distribusi Sentimen Publik di Platform X

Gambar 1 menunjukkan distribusi sentimen publik terhadap isu dugaan ijazah palsu Presiden Joko Widodo. Sentimen negatif mendominasi dengan 1.291 *tweet*, diikuti sentimen positif sebanyak 484 *tweet* dan netral sebanyak 338 *tweet*. Hasil ini menggambarkan bahwa opini publik di Platform X cenderung bersifat negatif terhadap isu yang beredar.

Arsitektur Model *IndoBERT*

Model yang digunakan adalah *IndoBERT-base-p1*, model *transformer* berbahasa Indonesia yang mampu memahami konteks kalimat secara dua arah. Model ini terdiri dari 12 lapisan *encoder* dengan mekanisme *self-attention* untuk menangkap hubungan antar kata secara kontekstual. Teks diproses melalui tokenisasi *WordPiece*, *embedding*, lalu melewati lapisan *transformer*, dan *output* token [CLS] digunakan pada lapisan klasifikasi untuk memprediksi sentimen positif, netral, atau negatif.

Arsitektur Model *IndoBERT*
untuk Analisis Sentimen



Gambar 2. Arsitektur Model *IndoBERT* untuk Analisis Sentimen

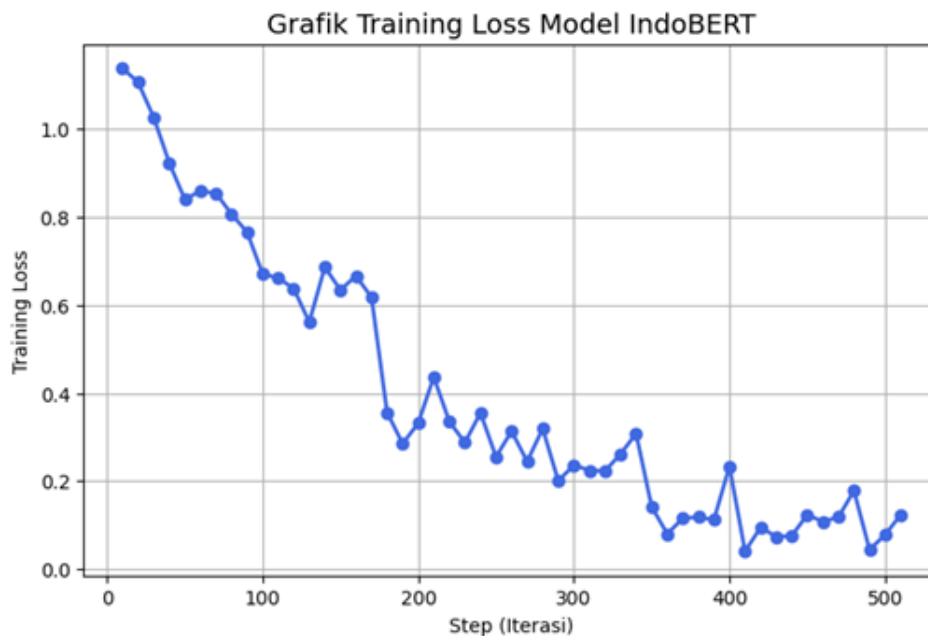
Gambar tersebut menunjukkan alur *IndoBERT* mulai dari teks masukan, tokenisasi dan *embedding*, pemrosesan melalui *Transformer* dengan *self-attention*, hingga lapisan klasifikasi yang menghasilkan label sentimen. Keunggulannya terletak pada kemampuan memahami konteks dan hubungan semantik secara menyeluruh, sehingga efektif untuk analisis sentimen isu sosial-politik di media sosial.

Pelatihan Model *IndoBERT*

Tahap berikutnya adalah *fine-tuning IndoBERT* menggunakan dataset penelitian agar sesuai dengan konteks isu yang dianalisis. Data dibagi menjadi 70% data latih dan 30% data uji. Pelatihan dilakukan di *Google Colab* dengan GPU menggunakan *Python* dan pustaka *HuggingFace* serta *PyTorch*. Optimasi memakai *AdamW* dan fungsi *loss Cross-Entropy* untuk klasifikasi tiga kelas sentimen.

Tabel 5. Parameter Pelatihan Model *IndoBERT*

Parameter	Nilai / Keterangan
Model	<i>IndoBERT-base-p1</i>
Optimizer	<i>AdamW</i>
Learning Rate	$5e-5$
Batch Size	16
Epoch	5
Training Split	70% data latih
Validation Split	30% data uji
Loss Function	<i>Cross Entropy Loss</i>
Evaluation Metric	<i>Accuracy, Precision, Recall, F1-score</i>

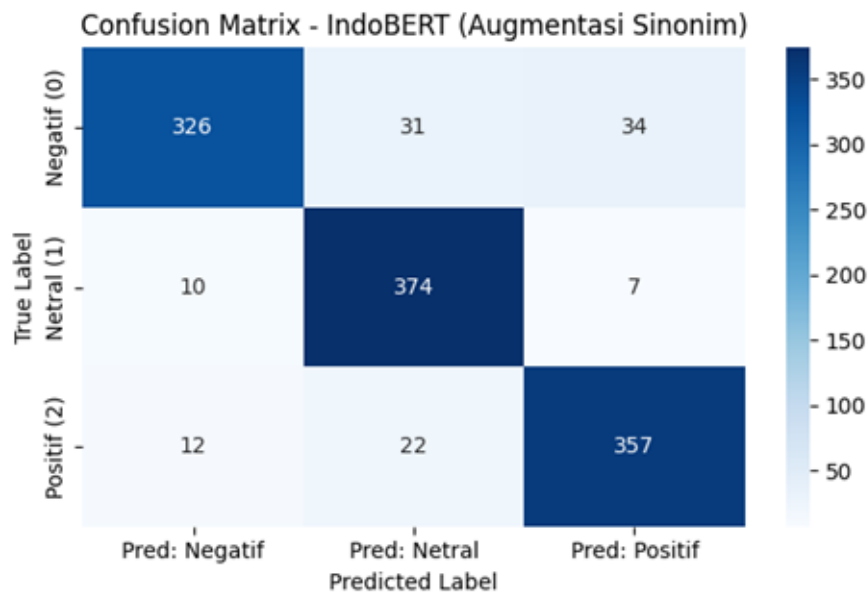


Gambar 3. Grafik Perubahan Nilai *Training Loss* Selama Pelatihan Model *IndoBERT*

Grafik menunjukkan *training loss IndoBERT* menurun dari sekitar 1,1 menjadi 0,1 secara konsisten selama pelatihan. Penurunan ini menandakan model belajar dengan baik, proses pelatihan stabil, dan tidak menunjukkan indikasi *overfitting*.

Evaluasi Model

Tahap evaluasi dilakukan untuk mengukur kinerja model *IndoBERT* dalam mengklasifikasikan sentimen positif, netral, dan negatif menggunakan data uji yang tidak dilibatkan dalam pelatihan. Evaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasilnya menunjukkan akurasi sebesar 91,13% dengan rata-rata *F1-score* 0,9011, menandakan performa model yang tinggi dan seimbang pada ketiga kelas. Kelas negatif memiliki *precision* tertinggi (0,9368), kelas netral menunjukkan *recall* tertinggi (0,9566), dan kelas positif juga stabil dengan *F1-score* 0,9094. Secara keseluruhan, model mampu mengklasifikasikan sentimen secara akurat dan konsisten, yang juga diperkuat melalui analisis matriks konfusi untuk melihat pola kesalahan prediksi.



Gambar 4. *Confusion Matrix* Hasil Evaluasi Model IndoBERT

Matriks konfusi menunjukkan sebagian besar data pada tiga kelas sentimen berhasil diklasifikasikan dengan benar, terlihat dari dominasi nilai pada diagonal utama. Model mampu memprediksi 326 data negatif, 374 netral, dan 357 positif secara tepat, dengan kesalahan kecil terutama antara kelas negatif dan netral. Secara keseluruhan, hasil ini menegaskan bahwa *IndoBERT* memiliki performa tinggi (akurasi >90%) dan mampu menggeneralisasi data dengan baik untuk analisis sentimen media sosial.

C. Implementasi Antar Muka Menggunakan *Streamlit*

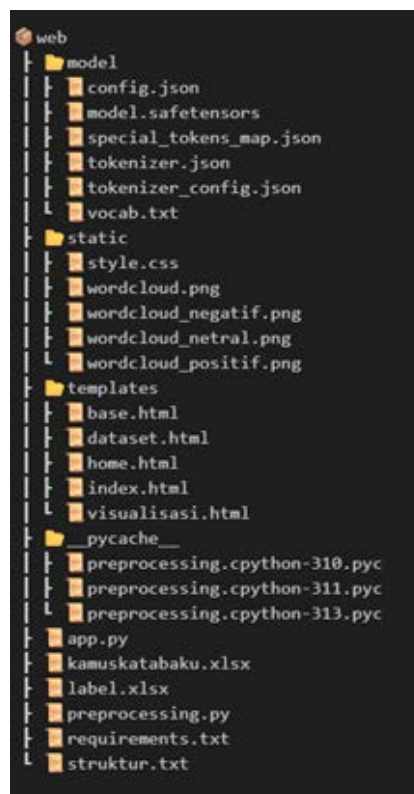
Penggunaan *Streamlit* dalam penelitian ini bertujuan menyediakan antarmuka pengguna yang sederhana, responsif, dan mudah dioperasikan sehingga model *IndoBERT* yang telah di-*fine-tune* dapat digunakan tanpa memerlukan kemampuan pemrograman lanjutan. *Streamlit* memungkinkan integrasi langsung antara logika *Python*, model machine learning, dan tampilan web, serta mendukung pemrosesan teks dan penampilan hasil analisis secara *real-time*. Dengan demikian, penelitian ini tidak hanya menghasilkan model yang akurat, tetapi juga sistem aplikasi yang aplikatif dan siap digunakan.

Aplikasi *Streamlit* yang dikembangkan memiliki fitur utama berupa deteksi sentimen otomatis berbasis *IndoBERT*. Proses dimulai dari input teks pengguna, dilanjutkan dengan pra-pemrosesan (*cleaning*, normalisasi, *stopword removal*, dan tokenisasi),

kemudian tokenisasi khusus *IndoBERT* untuk mengubah teks menjadi token ID. Selanjutnya model melakukan inferensi dan menghasilkan probabilitas untuk tiga kelas sentimen (positif, netral, negatif), lalu sistem menampilkan label dengan probabilitas tertinggi secara langsung di halaman web.

Aplikasi dijalankan pada lingkungan lokal menggunakan server bawaan *framework* dan dapat diakses melalui localhost (<http://127.0.0.1:5050>). Sebelum dijalankan, seluruh dependensi seperti *Torch* dan *Transformers* diinstal sesuai *requirements.txt*. Metode ini memudahkan proses pengujian dan debugging sebelum aplikasi dideploy ke platform *cloud* agar dapat diakses secara luas.

Tahap akhir penelitian adalah membangun antarmuka web interaktif menggunakan *Streamlit* untuk menampilkan dan menguji hasil analisis sentimen *IndoBERT*. Aplikasi ini mengintegrasikan pemrosesan teks, inferensi model, dan visualisasi hasil dalam satu sistem yang dijalankan secara lokal melalui <http://localhost:8501>



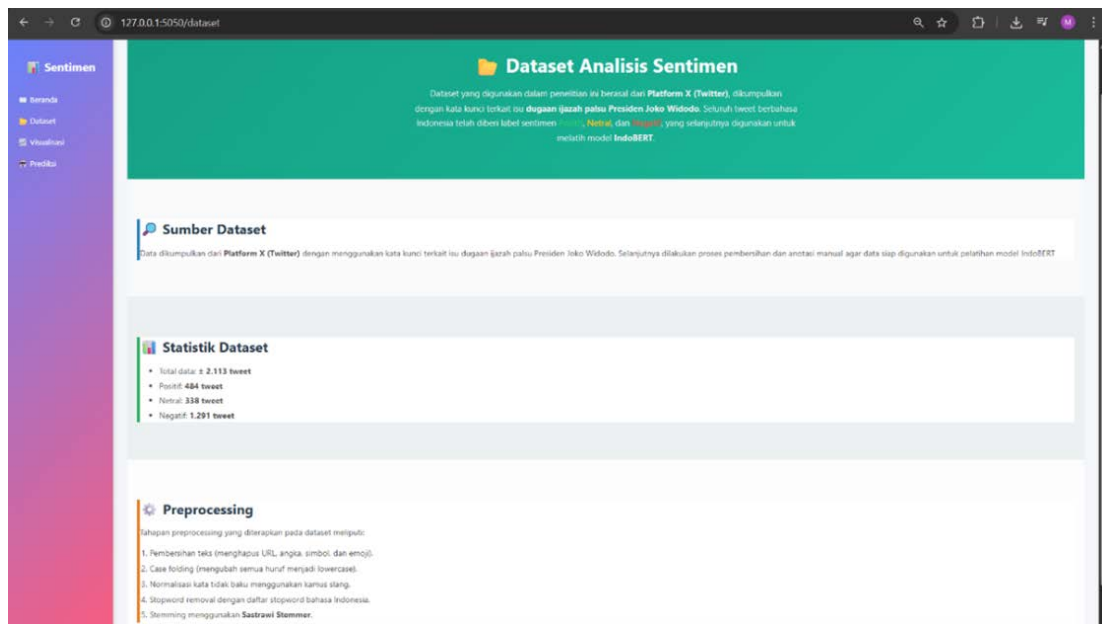
Gambar 5. Struktur Folder Aplikasi Web Analisis Sentimen

Struktur direktori terdiri dari beberapa komponen utama, antara lain folder *model/* yang berisi file hasil pelatihan *IndoBERT*, folder *templates/* yang menyimpan file HTML tampilan aplikasi, dan folder *static/* yang berisi berkas gambar dan file gaya (*style.css*). File utama *app.py* mengatur alur logika aplikasi dan pemanggilan model, sedangkan *preprocessing.py* menangani pembersihan teks sebelum dilakukan analisis sentimen.



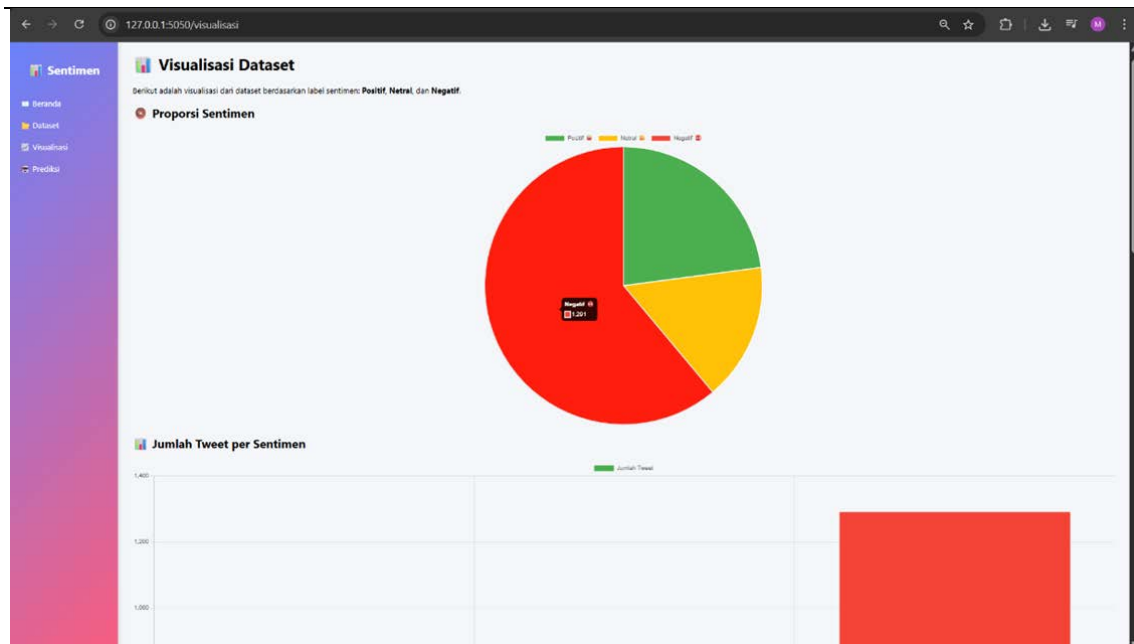
Gambar 6. Halaman Beranda Aplikasi Analisis Sentimen

Halaman Beranda menampilkan ringkasan penelitian, tujuan, latar belakang isu, dan metodologi yang digunakan. Tersedia sidebar navigasi untuk mengakses menu Dataset, Visualisasi, dan Prediksi Sentimen. Desainnya sederhana dan profesional agar mudah dipahami pengguna.



Gambar 7. Halaman Dataset Analisis Sentimen

Halaman Dataset menampilkan ringkasan data yang digunakan untuk pelatihan dan evaluasi model. Informasi yang disajikan meliputi sumber data dari Platform X (Twitter), total 2.113 tweet (484 positif, 338 netral, 1.291 negatif), serta tahapan *preprocessing* seperti *cleaning*, normalisasi, *stopword removal*, dan *stemming*. Tampilan dirancang informatif dan terstruktur agar mudah dipahami.



Gambar 8. Halaman Visualisasi Dataset (*Pie Chart* dan *Bar Chart*)

Halaman Visualisasi menampilkan distribusi sentimen dalam bentuk *pie chart* dan *bar chart*. Hasil menunjukkan dominasi sentimen negatif (1.291 tweet/61%), diikuti positif (484/23%) dan netral (338/16%). Grafik dirancang sederhana agar perbandingan data mudah dipahami secara visual.

Dashboard Analisis Sentimen Publik (Twitter/X)

Studi Kasus: Isu Dugaan Israr Palsu Presiden Joko Widodo

Masukkan Teks Opini Publik (Twitter/X):

tjerah Jokowi palsu

Analisis Sentimen

Hasil Analisis Sentimen: 63 negatif

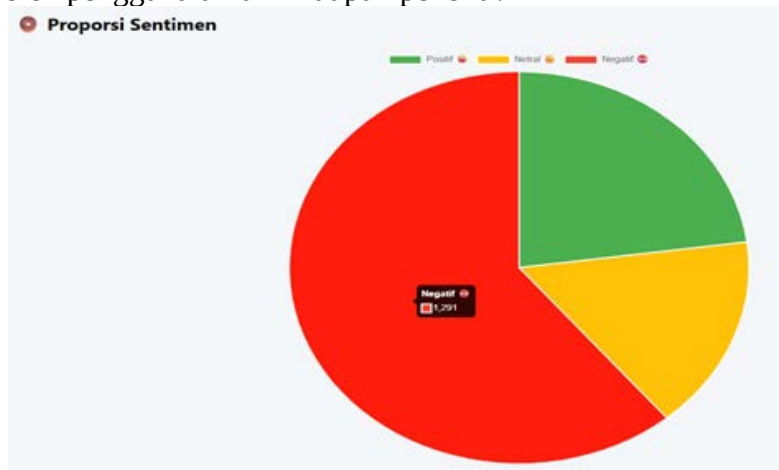
[Kembali](#)

Gambar 9. Halaman Prediksi Sentimen IndoBERT

Halaman Prediksi Sentimen memungkinkan pengguna memasukkan teks untuk dianalisis langsung oleh model *IndoBERT*. Sistem otomatis melakukan *preprocessing*, tokenisasi, dan inferensi, lalu menampilkan label sentimen (positif, netral, atau negatif) berdasarkan probabilitas tertinggi secara *real-time*. Antarmuka dirancang modern dan interaktif, sehingga aplikasi dapat berfungsi sebagai dashboard analisis opini publik yang mudah digunakan.

PEMBAHASAN

Berdasarkan hasil implementasi, aplikasi web berbasis *Streamlit* berhasil mengintegrasikan model *IndoBERT* hasil *fine-tuning* dengan antarmuka yang interaktif dan mudah digunakan. Sistem mampu melakukan analisis sentimen secara *real-time*, mulai dari input teks hingga menampilkan label hasil prediksi. Setiap halaman memiliki fungsi yang saling melengkapi, mulai dari pengantar penelitian, transparansi dataset dan *preprocessing*, visualisasi distribusi sentimen, hingga fitur prediksi langsung terhadap teks baru. Seluruh proses berjalan dalam satu sistem terintegrasi yang efisien dan modular, dengan desain antarmuka sederhana serta navigasi intuitif sehingga mudah dioperasikan oleh pengguna umum maupun peneliti.



Gambar 10. Visualisasi Proporsi Sentimen

Visualisasi proporsi sentimen dalam Gambar 4.10 memperlihatkan bahwasanya sentimen negatif mendominasi dengan jumlah sebanyak 1.291 data, diikuti oleh sentimen positif sejumlah 484 data, serta sentimen netral sejumlah 338 data. Hal ini menunjukkan bahwa opini publik terhadap isu dugaan ijazah palsu Presiden Joko Widodo di media sosial didominasi oleh pandangan negatif. Mayoritas pengguna lebih banyak mengekspresikan ketidakpercayaan, kekecewaan, serta kritik terhadap isu tersebut, sedangkan sebagian kecil lainnya menampilkan dukungan atau sikap netral. Hasil ini menandakan bahwa persepsi publik secara umum masih cenderung skeptis terhadap isu yang berkembang di ruang digital.



Gambar 11. Visualisasi Hasil Wordcloud Per Sentimen

Visualisasi *wordcloud* menunjukkan kata-kata dominan pada tiap kategori sentimen. Pada sentimen negatif, muncul kata seperti “ijazah”, “palsu”, “gugat”, dan “hukum” yang mencerminkan kritik dan ketidakpercayaan publik. Sentimen positif menampilkan kata seperti “asli”, “percaya”, “bukti”, dan “UGM” yang mengarah pada dukungan dan klarifikasi. Sementara itu, sentimen netral didominasi kata informatif seperti “lapor”, “sidang”, dan “dokumen” yang menggambarkan pemberitaan faktual tanpa emosi kuat. Secara keseluruhan, hasil visualisasi menunjukkan dominasi opini negatif, meskipun tetap terdapat sentimen positif dan netral yang menyeimbangkan diskusi publik. Temuan ini menegaskan bahwa model *IndoBERT* mampu mengklasifikasikan sentimen secara akurat serta memetakan persepsi publik baik secara kuantitatif maupun kualitatif melalui grafik dan *wordcloud*.

4. KESIMPULAN

Sesuai hasil penelitian serta analisis yang sudah dilaksanakan terkait *Implementasi Model IndoBERT untuk Analisis Sentimen Publik terhadap Isu Dugaan Ijazah Palsu Presiden Joko Widodo dalam Percakapan di Platform X (Twitter)*, dengan demikian bisa diambil sejumlah simpulan meliputi:

1. Proses pengumpulan data berhasil dilaksanakan melalui *web scraping* menggunakan Twitter API v2 dengan kata kunci “ijazah Jokowi”, “ijazah palsu Presiden”, “ijazah Presiden Joko Widodo”, dan “ijazah palsu Jokowi”. Data mencakup periode waktu 12 Januari 2025 hingga 14 Agustus 2025, menghasilkan 2.262 tweet mentah, yang setelah pembersihan dan seleksi menghasilkan 2.113 tweet berbahasa Indonesia yang relevan dengan topik penelitian.
2. Proses pelabelan sentimen dilaksanakan melalui metode *lexicon-based* yang menggunakan dua kamus leksikon (*positive.csv* dan *negative.csv*) serta disempurnakan melalui koreksi manual menggunakan *hasil_pelabelan_perbaikan.xlsx*. Distribusi sentimen publik menunjukkan 61,1% tweet negatif, 22,9% positif, dan 16% netral, yang berarti opini publik terhadap isu dugaan ijazah palsu Presiden Joko Widodo cenderung bersifat negatif.
3. Tahapan *preprocessing* data yang meliputi *cleaning*, *case folding*, *normalisasi*, *tokenisasi*, *filtering*, serta *stemming* berhasil menghasilkan data teks yang bersih dan terstruktur, siap dipergunakan pada proses pelatihan model *IndoBERT*.
4. Model *IndoBERT-base-p1* yang dipergunakan di studi ini terbukti efektif untuk analisis sentimen teks berbahasa Indonesia. Hasil evaluasi menunjukkan nilai akurasi 91,13%, *precision* 90,50%, *recall* 91,00%, serta *F1-score* 0,91, yang mengindikasikan performa sangat baik dan seimbang antara presisi serta sensitivitas model terhadap tiap kelas sentimen.
5. Implementasi antarmuka web berbasis *Flask* dan *HTML-CSS* berhasil mengintegrasikan model *IndoBERT* ke dalam sistem analisis sentimen yang interaktif. Aplikasi dapat menampilkan dataset, hasil visualisasi berupa *pie chart*, *bar chart*, dan *wordcloud*, serta melakukan prediksi sentimen terhadap teks baru secara langsung. Hal ini membuktikan bahwa model *IndoBERT* bukan saja efektif

secara teknis, tapi juga bisa diimplementasikan dalam bentuk aplikasi yang informatif serta mudah digunakan.

Secara keseluruhan, studi ini memperlihatkan bahwasanya model IndoBERT bisa dipergunakan secara efektif untuk memahami persepsi publik terhadap isu sosial-politik di Indonesia, khususnya dalam menganalisis opini masyarakat di media sosial berbahasa Indonesia yang bersifat tidak formal dan dinamis.

DAFTAR PUSTAKA

- Fatra, E., Tri, T., Manguma, F., Bisnis, F., Sosial, T., Studi, P., Informatika, T., Bisnis, F., Sosial, T., Madani, U. A., dan Joko, P. (2024). *Analisis Framing Berita Metro Tv (Studi Kasus : Netralitas Presiden Joko Widodo Dalam Pemilihan Calon Presiden 2024-2029 Di Indonesia)*. 2(1), 1–12.
- Kominfo. (2023). *Siaran pers No. 49/HM/KOMINFO/04/2023 tentang tingkatan literasi digital sektor pemerintahan lewat ASN makin cakap digital*. Kementerian Komunikasi dan Informatika Republik Indonesia. <https://www.komdigi.go.id/berita/siaran-pers/detail/siaran-pers-no-49-hm-kominfo-04-2023-tentang-tingkatkan-literasi-digital-sektor-pemerintahan-lewat-asn-makin-cakap-digital>
- MAFINDO. (2023). *Hoaks Pemilu Naik 2x Lipat, Ganggu Demokrasi Indonesia*. Gatra.com. <https://www.gatra.com/news-591850-politik-mafindo-hoaks-pemilu-naik-2x-lipat-ganggu-demokrasi-indonesia.html>
- Merinda Lestandy, Abdurrahim Abdurrahim, dan Lailis Syafa'ah. (2021). Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 5(4), 802–808. <https://doi.org/10.29207/resti.v5i4.3308>
- Putri, D. I., Alfian, A. N., Putra, M. Y., dan Mulyo, P. D. (2024). IndoBERT Model Analysis: Twitter Sentiments on Indonesia's 2024 Presidential Election. *Journal of Applied Informatics and Computing*, 8(1), 7–12. <https://doi.org/10.30871/jaic.v8i1.7440>
- Syifa, N., Nugraha, B., dan Zulfadhilah, M. (2024). *Online Analisis Sentimen pada Pemilihan Umum Presiden di Kota Banjarmasin 2024*. 7(6), 1511–1519.
- Tandijaya, J. H., Liliana, dan Sugiarto, I. (2021). Klasifikasi dalam Pembuatan Portal Berita Online dengan Menggunakan Metode BERT. *Jurnal Infra*, 09(02), 320–325.
- Utami, W. W., dan Darmaiza, D. (2020). Hate Speech, Agama, dan Kontestasi Politik di Indonesia. *Indonesian Journal of Religion and Society*, 2(2), 113–128. <https://doi.org/10.36256/ijrs.v2i2.108>
- Wenando, F. A., Hayami, R., dan Anggrawan, A. J. (2020). Analisis Sentimen Pada

Pemerintahan Terpilih Pada Pilpres 2019 Ditwitter Menggunakan Algoritme Naïvebayes. *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, 7(1), 101–106. <https://doi.org/10.33330/jurteksi.v7i1.851>

Yusrizal, M., dan Sasongko, T. B. (2024). Analisis Sentimen Masyarakat Terhadap Presiden dan Calon Presiden Terpilih 2024 Menggunakan Naïve Bayes. *Jurnal Media Informatika Budidarma*, 8(3), 1673. <https://doi.org/10.30865/mib.v8i3.7882>

Yuniar, E., Utsalinah, D. S., dan Wahyuningsih, D. (2022). Implementasi Scrapping Data Untuk Sentiment Analysis Pengguna Dompot Digital dengan Menggunakan Algoritma Machine Learning. *Jurnal Janitra Informatika dan Sistem Informasi*, 2(1), 35–42. <https://doi.org/10.25008/janitra.v2i1.145>