

Analisis Model Klasifikasi Sentimen Publik Terhadap Kebijakan Keberlanjutan IKN Menggunakan BERT Sebagai Feature Extractor dan K-Nearest Neighbor (KNN)

Mohammad Hiqmal Fiqri*, Rudiman Rudiman, Naufal Azmi Verdikha

Sains dan Teknologi, Program Studi Teknik Informatika, Universitas Muhammadiyah Kalimantan Timur, Samarinda, Indonesia

Email: ^{1,*}2111102441052@umkt.ac.id, ²rud959@umkt.ac.id, ³nav651@umkt.ac.id

Email Penulis Korespondensi: 2111102441052@umkt.ac.id

Submitted: 04/08/2025; Accepted: 04/09/2025; Published: 05/09/2025

Abstrak—Penelitian ini bertujuan untuk mengevaluasi kinerja model klasifikasi sentimen publik terhadap kebijakan pemindahan Ibu Kota Negara (IKN) menggunakan kombinasi IndoBERT sebagai feature extractor dan K-Nearest Neighbor (KNN) sebagai classifier. Data penelitian berupa 1.274 komentar YouTube terkait IKN, yang dilabeli oleh seorang ahli di bidang sosiologi dan analisis teks. Proses preprocessing meliputi pembersihan teks dari angka, URL, emotikon, dan tanda baca, serta penghapusan stopword menggunakan Sastrawi. Representasi vektor yang dihasilkan IndoBERT berukuran 768 dimensi, kemudian diklasifikasikan menggunakan KNN dengan nilai $k=5$ dan Euclidean distance. Hasil evaluasi menggunakan 5-fold cross validation menunjukkan akurasi sebesar 73,31%. Namun, recall untuk kelas positif tergolong rendah (0,49), yang mengindikasikan adanya tantangan dalam mendeteksi komentar positif akibat distribusi kelas yang tidak seimbang (831 negatif, 294 positif, dan 149 netral). Temuan ini memperlihatkan bahwa kombinasi IndoBERT+KNN cukup efektif pada kelas mayoritas, tetapi masih lemah pada kelas minoritas. Kontribusi penelitian ini adalah memberikan analisis kritis terhadap keterbatasan model berbasis IndoBERT dalam konteks klasifikasi sentimen berbahasa Indonesia serta memberikan rekomendasi pengembangan selanjutnya, seperti balancing data dan fine-tuning model.

Kata Kunci: IndoBERT; K-Nearest Neighbor; Ibu Kota Negara; Youtube; Klasifikasi

Abstract—This study aims to evaluate the performance of sentiment classification models for public opinions regarding the relocation of Indonesia's new capital (IKN) using a combination of IndoBERT as a feature extractor and K-Nearest Neighbor (KNN) as a classifier. The dataset consisted of 1,274 YouTube comments related to IKN, which were annotated by an expert in sociology and text analysis. The preprocessing stage involved cleaning numbers, URLs, emojis, and punctuation, as well as removing stopwords using the Sastrawi library. IndoBERT produced 768-dimensional vector representations, which were then classified using KNN with $k=5$ and Euclidean distance. Evaluation with 5-fold cross validation achieved an accuracy of 73.31%. However, the recall for the positive class was relatively low (0.49), indicating challenges in detecting positive comments due to class imbalance (831 negative, 294 positive, 149 neutral). These findings suggest that the IndoBERT+KNN model performs well on majority classes but struggles with minority classes. The contribution of this research is to provide a critical analysis of the limitations of IndoBERT-based models in Indonesian sentiment classification and to recommend future directions, including data balancing and fine-tuning approaches.

Keywords: IndoBERT; K-Nearest Neighbor; Nusantara Capital City; YouTube; Classification

1. PENDAHULUAN

Ibu kota negara, Pemindahan Ibu Kota Negara (IKN) dari Jakarta ke Kalimantan Timur menimbulkan perdebatan luas di masyarakat Indonesia. Topik ini sering dibahas di media sosial, khususnya YouTube, sebagai wadah bagi publik untuk mengekspresikan opini mereka. Analisis opini publik melalui komentar daring penting dilakukan karena memberikan masukan terhadap pembuat kebijakan dan mencerminkan persepsi masyarakat. Dalam perspektif ilmu sosial komputasional, analisis sentimen bukan hanya menghitung opini positif atau negatif, tetapi juga mengevaluasi kemampuan model dalam memahami nuansa bahasa Indonesia dalam konteks kebijakan publik.

Sejumlah penelitian sebelumnya telah mengkaji analisis sentimen dengan beragam pendekatan. Pendekatan berbasis Support Vector Machine (SVM) [1] sering menghasilkan akurasi tinggi, namun terbatas dalam menangkap konteks semantik bahasa alami yang kompleks [2]. Naïve Bayes dan Random Forest [3] populer karena kesederhanaannya, tetapi performanya menurun saat menghadapi data berbahasa Indonesia yang sarat variasi kata dan konteks sosial [4]. Transformer berbasis BERT menunjukkan kinerja signifikan dalam NLP karena kemampuannya memahami konteks dalam kalimat [5]. Namun, sebagian besar studi hanya mengeksplorasi IndoBERT dengan layer klasifikasi standar (misalnya linear/softmax), dan belum banyak yang mengombinasikannya dengan metode non-parametrik seperti K-Nearest Neighbor (KNN)[6].

Untuk mengatasi keterbatasan tersebut, penelitian ini menggunakan IndoBERT, model berbasis BERT yang dilatih khusus untuk bahasa Indonesia. BERT memiliki keunggulan dalam memahami konteks dua arah dalam kalimat, sehingga lebih efektif menangkap makna yang kompleks dan variasi bahasa dalam komentar publik seperti di YouTube [7]. Studi sebelumnya [8] yang membandingkan tiga metode ekstraksi fitur FastText, TF-IDF, dan IndoBERT, menunjukkan bahwa IndoBERT memberikan performa terbaik pada data uji, mengatasi keterbatasan model tradisional dalam menangkap konteks semantik

Kombinasi IndoBERT sebagai feature extractor dan KNN sebagai classifier sangat menarik untuk dieksplorasi. IndoBERT menghasilkan representasi vektor 768 dimensi, sedangkan KNN melakukan klasifikasi berbasis jarak. Pendekatan hybrid ini memungkinkan evaluasi performa klasifikasi pada data dengan distribusi kelas tidak seimbang,

di mana kelas mayoritas (negatif) mendominasi dibandingkan kelas minoritas (positif dan netral). Dengan demikian, penelitian ini menawarkan perspektif baru dalam pemanfaatan model transformer pada analisis sentimen berbahasa Indonesia.

Dengan mempertimbangkan keunggulan IndoBERT dalam ekstraksi fitur dan efektivitas KNN sebagai metode klasifikasi, penelitian ini memfokuskan diri pada analisis model klasifikasi, bukan semata-mata pada hasil analisis sentimen. Fokus utama penelitian adalah mengevaluasi performa kombinasi IndoBERT dan KNN dalam menangani klasifikasi data opini publik terhadap kebijakan keberlanjutan IKN.

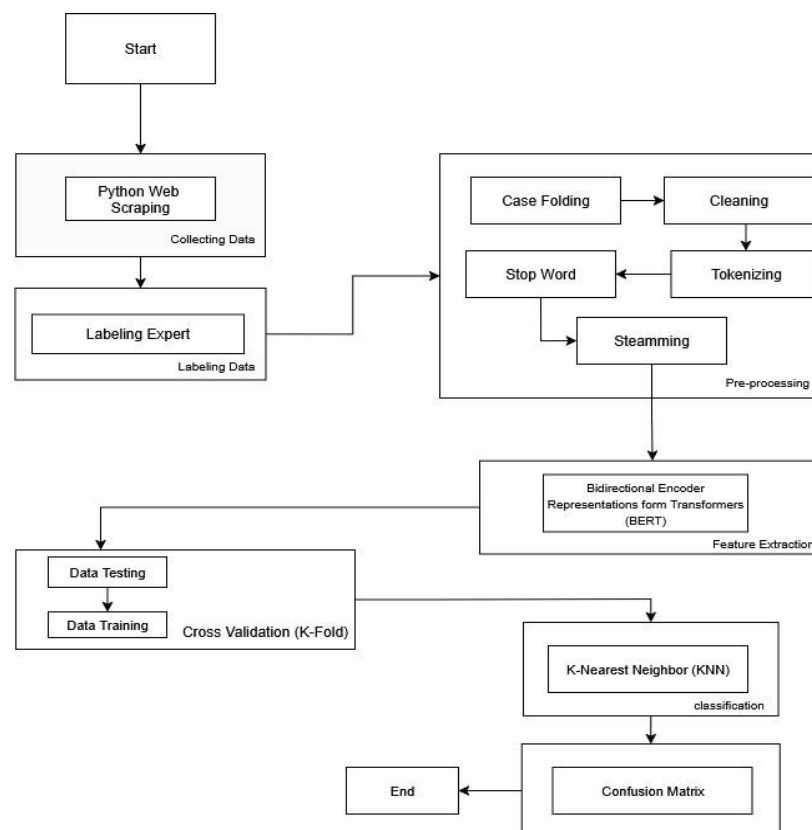
2. METODOLOGI PENELITIAN

Objek dalam penelitian ini adalah model klasifikasi sentimen terhadap opini publik yang diambil dari komentar video YouTube dengan topik pembangunan Ibu Kota Negara (IKN) Nusantara. Penelitian ini tidak hanya bertujuan untuk mengidentifikasi jenis sentimen (positif, negatif, netral), tetapi lebih menekankan pada evaluasi performa model klasifikasi dengan menggunakan kombinasi representasi fitur dari IndoBERT dan algoritma K-Nearest Neighbor (KNN). Komentar yang dianalisis berasal dari video berjudul "BOS IKN MENOLAK 'IK-END' | TOP ECONOMY." yang diunggah oleh akun MetroTV. Dengan menggunakan data tersebut, penelitian ini menguji seberapa baik model klasifikasi bekerja dalam mengelompokkan opini publik berdasarkan metrik seperti akurasi, precision, recall, dan F1-score.

Data yang digunakan dalam penelitian ini diambil pada tanggal 15 April 2025, dengan total 1.274 komentar yang dianalisis. Komentar tersebut diperoleh melalui YouTube Data API v3 dari video YouTube berjudul "[FULL] BOS IKN MENOLAK 'IK-END' | TOP ECONOMY" (<https://www.youtube.com/watch?v=bY6u0r3CdVE>). Komentar ini digunakan sebagai sumber data untuk membangun dan menguji model klasifikasi sentimen berbasis kombinasi fitur IndoBERT dan algoritma K-Nearest Neighbor (KNN).

2.1 Alur Penelitian

Prosedur penelitian adalah langkah-langkah yang diambil oleh peneliti untuk menghimpun data atau informasi untuk kemudian dianalisis secara ilmiah. Terdapat 9 tahapan yang menjadi dasar bagi sebuah penelitian yang ditampilkan pada Gambar 1.



Gambar 1 Alur Penelitian

Metode penelitian ini terdiri dari beberapa tahapan utama, dimulai dengan data dikumpulkan dengan cara scraping dari Youtube menggunakan Python. Setelah data terkumpul, dilakukan labelling oleh ahli, kemudian tahap pre-processing, yang meliputi case folding, cleaning, tokenizing, stop word, dan stemming untuk representasi teks. Kemudian, setelah itu diekstraksi menggunakan IndoBERT, Cross Validation (K-Fold) dan K-Nearest Neighbor

(KNN) untuk klasifikasi. Terakhir, untuk menilai kinerja model, digunakan confusion matrix yang menghasilkan metrik seperti precision, recall, dan F1-score, yang menunjukkan seberapa baik model dalam membuat prediksi.

2.2 Pengumpulan Data

Data diambil dari komentar YouTube terkait kebijakan IKN menggunakan YouTube Data API v3 [9]. Sebanyak 1.274 komentar dikumpulkan dari berbagai kanal berita resmi dan diskusi publik seputar IKN. Penggunaan API resmi memastikan legalitas dan konsistensi data, berbeda dengan teknik scraping manual yang sering melanggar kebijakan penggunaan [10]. Data yang dikumpulkan meliputi teks komentar, tanggal unggahan, dan metadata seperti tanggal posting, jumlah like, dan identitas pengguna. Untuk mendukung reproducible research, data disimpan dalam format CSV dan JSON yang dapat diakses melalui repository publik, selain versi Excel [11] sebagai referensi internal.

2.4 Labelling Data

Tahapan a Proses pelabelan data dilakukan secara manual oleh H. Irfan Abdul Hakim, S.Sos., M.A., seorang ahli dengan latar belakang pendidikan S1 dan S2 Sosiologi Universitas Gadjah Mada, serta berpengalaman dalam bidang sentiment annotation, NLP, text classification, contextual analysis, dataset preparation, linguistik bahasa Indonesia, serta research & data interpretation.

Sebanyak 1.274 komentar dilabeli ke dalam tiga kategori sentimen, negatif (831 komentar), positif (294 komentar), dan netral (149 komentar). Pelabelan dilakukan dengan mempertimbangkan konteks sosial-linguistik dari setiap komentar. Meskipun hanya melibatkan satu annotator, kredibilitas ahli dipandang memadai karena sesuai dengan bidang kompetensinya. Untuk meningkatkan transparansi, distribusi data per kelas dilaporkan secara detail, dan proses labeling ini mengikuti praktik umum pada penelitian NLP berbasis bahasa Indonesia[12].

2.5 Preprocessing

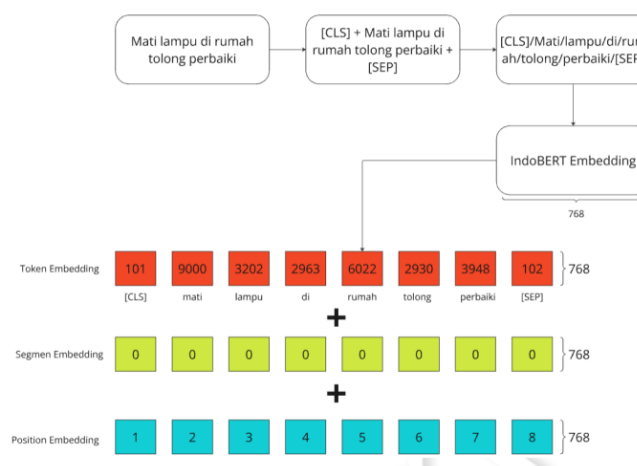
Tahap preprocessing mencakup case folding, cleaning, stopword removal, dan stemming menggunakan library Sastrawi [13]. Pada model berbasis BERT, penggunaan stemming sering diperdebatkan karena IndoBERT telah dilatih pada korpus bahasa Indonesia tanpa stemming. Beberapa penelitian menunjukkan bahwa stemming [14] berpotensi mengurangi konteks semantik kata dan menurunkan performa[15].

Namun, pada penelitian ini stemming tetap diterapkan. Alasannya adalah komentar YouTube sebagai sumber data seringkali mengandung bahasa informal, singkatan, serta variasi kata yang sangat beragam. Dengan menerapkan stemming, variasi kata tersebut dapat dipetakan ke bentuk dasar sehingga representasi teks menjadi lebih konsisten. Hasil uji coba awal menunjukkan bahwa performa klasifikasi lebih stabil dengan teks yang telah distem. Oleh karena itu, meskipun terdapat pandangan bahwa stemming tidak diperlukan untuk BERT, keputusan penelitian ini adalah tetap menggunakan stemming karena sesuai dengan karakteristik data yang dianalisis.

2.6 Fitur Ekstraksi (IndoBERT)

Ekstraksi Pada tahap ini digunakan IndoBERT-base sebagai model pre-trained language model untuk menghasilkan representasi vektor dari setiap komentar. IndoBERT dipilih karena merupakan model transformer yang telah dilatih secara luas pada korpus bahasa Indonesia [16], sehingga mampu menangkap konteks linguistik dengan baik [17]. Dalam penelitian ini, IndoBERT digunakan hanya sebagai feature extractor tanpa dilakukan fine-tuning pada bobot model. Hal ini dipilih karena keterbatasan komputasi serta fokus penelitian yang diarahkan pada evaluasi efektivitas algoritma K-Nearest Neighbor (KNN) dalam mengklasifikasikan representasi fitur hasil IndoBERT.

Setiap komentar diproses melalui tokenizer IndoBERT untuk menghasilkan token, kemudian dimasukkan ke dalam model IndoBERT. Vektor representasi diambil dari hidden layer terakhir berupa 768 dimensi untuk setiap teks. Representasi ini kemudian digunakan sebagai input bagi algoritma klasifikasi KNN.



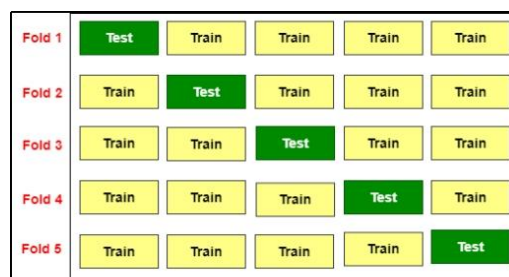
Gambar 2 Diagram Proses Ekstraksi IndoBERT

Proses ekstraksi fitur dengan IndoBERT ditunjukkan pada Gambar 2, yang menggambarkan alur input teks hingga menghasilkan representasi vektor. Tahapan dimulai dengan memasukkan kalimat komentar seperti “Mati lampu di rumah tolong perbaiki” ke dalam model. Sebelum diproses oleh IndoBERT, kalimat ini terlebih dahulu ditambahkan token khusus berupa [CLS] di awal dan [SEP] di akhir, sehingga menjadi “[CLS] Mati lampu di rumah tolong perbaiki [SEP]”. Setelah itu, teks mengalami proses tokenisasi menggunakan tokenizer IndoBERT, yang memecah kalimat menjadi token-token dan mengubahnya ke dalam bentuk angka berdasarkan vocabulary bawaan IndoBERT [18].

Setiap token kemudian dipetakan ke dalam token embedding berdimensi 768. Selain itu, segment embedding digunakan untuk membedakan antar bagian teks (dalam konteks ini bernilai 0 karena hanya satu segmen), dan position embedding diberikan untuk menyandikan posisi masing-masing token dalam urutan kalimat. Ketiga embedding tersebut (token, segment, dan position) dijumlahkan untuk membentuk representasi akhir setiap token sebelum dimasukkan ke dalam model IndoBERT. Hasil akhirnya adalah vektor representasi dari setiap token, namun pada umumnya hanya vektor dari token [CLS] yang digunakan sebagai representasi keseluruhan kalimat. Representasi ini kemudian digunakan sebagai input fitur dalam proses klasifikasi sentimen pada tahap selanjutnya. Proses ini memungkinkan model untuk menangkap konteks makna kata dalam kalimat secara lebih akurat dibandingkan dengan metode ekstraksi fitur konvensional.

2.7 Cross Validation (K-Fold)

Untuk mengevaluasi model secara menyeluruh dan menghindari bias akibat pembagian data acak, penelitian ini menerapkan metode validasi silang (cross validation) menggunakan teknik K-Fold Cross Validation. Teknik ini membagi data menjadi K bagian (fold), di mana setiap bagian secara bergantian digunakan sebagai data uji, sementara bagian lainnya digunakan sebagai data latih. Pada penelitian ini digunakan 5-Fold Cross Validation, karena memberikan keseimbangan antara waktu komputasi dan reliabilitas hasil [19]. Dengan metode ini, seluruh data berkesempatan menjadi data uji dan data latih secara bergantian, sehingga hasil evaluasi lebih objektif.



Gambar 3 Skema K-Fold Cross Validation

Skema pembagian data dengan 5-Fold Cross Validation ditunjukkan pada Gambar 3, di mana setiap fold berperan secara bergantian sebagai data uji (blok hijau), sementara fold lainnya digunakan sebagai data latih (blok kuning).

2.8 K-Nearest Neighbors (KNN)

KNN menghitung jarak antara vektor uji dan seluruh data latih, kemudian mengambil K tetangga terdekat untuk menentukan label. Nilai K dalam penelitian ini ditentukan secara eksperimen dan literatur, yaitu K = 5 . Jarak antar vektor dihitung menggunakan rumus Euclidean:

$$euc = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \tag{1}$$

Dimana euc merepresentasikan jarak Euclidean antara vektor fitur komentar xxx dan xix_ixi, dengan nnn adalah jumlah dimensi fitur, yaitu 768 pada representasi IndoBERT. Setelah semua jarak dihitung, dipilih k tetangga terdekat, kemudian kelas dengan frekuensi terbanyak dari tetangga tersebut ditetapkan sebagai label untuk data uji. Pada penelitian ini digunakan nilai k =5 dengan fungsi jarak Euclidean Distance, karena secara empiris kombinasi ini menghasilkan performa yang cukup baik pada data berukuran menengah dan berdimensi tinggi . Representasi fitur IndoBERT dengan dimensi 768 digunakan sebagai input ke dalam algoritma KNN, sehingga setiap komentar direpresentasikan sebagai vektor numerik dalam ruang berdimensi tinggi. Dengan pendekatan ini, diharapkan KNN mampu membedakan komentar positif, negatif, dan netral berdasarkan pola kedekatan vektor.

2.9 Evaluasi

Evaluasi performa model dilakukan dengan menggunakan metrik akurasi, presisi, recall, dan F1-score. Akurasi menggambarkan tingkat keseluruhan prediksi yang tepat. Presisi berfokus pada seberapa tepat model dalam memprediksi kelas positif, sedangkan recall menunjukkan seberapa baik model menangkap seluruh data positif yang sebenarnya. F1-score digunakan sebagai ukuran kompromi antara presisi dan recall, khususnya pada data yang tidak seimbang antar kelas [20].



Akurasi merupakan ukuran yang menunjukkan seberapa dekat nilai prediksi dengan nilai sebenarnya dalam suatu model klasifikasi. Semakin tinggi nilai akurasi, semakin baik performa model dalam mengklasifikasikan data. Rumus untuk menghitung akurasi dalam persamaan

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

Dengan kata lain, akurasi menunjukkan proporsi data yang berhasil diklasifikasikan dengan tepat oleh model terhadap keseluruhan data uji.

Presisi mengukur ketepatan model dalam memprediksi kelas positif, yaitu seberapa besar proporsi dari prediksi positif yang benar-benar positif.

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

Presisi yang tinggi menunjukkan bahwa sebagian besar prediksi positif yang dihasilkan oleh model adalah benar. Recall mengukur kemampuan model dalam menangkap seluruh instance positif yang sebenarnya. Dengan kata lain, ini adalah proporsi dari data yang benar-benar positif yang berhasil dikenali oleh model.

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

Recall tinggi menunjukkan bahwa model tidak melewatkan banyak kasus positif. F1-score merupakan rata-rata harmonik dari presisi dan recall. Skor ini digunakan untuk menyeimbangkan keduanya, terutama saat terdapat ketidakseimbangan distribusi kelas dalam dataset.

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \tag{5}$$

F1-score memberikan ukuran yang lebih baik ketika presisi dan recall memiliki nilai yang berbeda secara signifikan. Definisi simbol yang digunakan dalam perhitungan metrik evaluasi dapat dilihat pada Tabel 1.

Tabel 1 Definisi Simbol

Simbol	Keterangan
TP	True Positive, jumlah data positif yang diprediksi benar sebagai positif
TN	True Negative, jumlah data negatif yang diprediksi benar sebagai negatif
FP	False Positive, jumlah data negatif yang salah diprediksi sebagai positif
FN	False Negative, jumlah data positif yang salah diprediksi sebagai negatif

Berdasarkan tabel tersebut, Precision dihitung sebagai perbandingan antara TP dengan (TP + FP), yang menunjukkan proporsi prediksi positif yang benar. Recall merupakan perbandingan antara TP dengan (TP + FN), yang mengukur seberapa baik model mendeteksi data positif. F1-Score merupakan rata-rata harmonik antara Precision dan Recall, sehingga memberikan gambaran performa yang lebih seimbang terutama pada dataset yang tidak seimbang. Sementara itu, Accuracy dihitung sebagai rasio jumlah prediksi benar (TP + TN) terhadap seluruh data. Keempat metrik ini dipilih karena mampu memberikan gambaran yang lebih komprehensif terkait kinerja model, khususnya dalam mengatasi distribusi kelas yang tidak seimbang pada dataset komentar YouTube terkait kebijakan IKN [15].

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Dalam penelitian ini, hasil dari data komentar youtube yang diambil dari video “BOS IKN MENOLAK “IK-END” | TOP ECONOMY”. Sebanyak 1.274 komentar yang terkandung kata IKN berhasil dikumpulkan dan disimpan dalam bentuk format xls. Hasil *crawling* komentar YouTube terkait kebijakan IKN disajikan pada Tabel 2.

Tabel 2 Hasil Crawling

No.	Username	Waktu	Komentar
1.	@sumini3316	2025-04-07	Alhamdullillah IKN jalan terus,karena keponakan saya jadi mandor di sana ,Ramadhan Libur tapi habis lebaran jalan lagi
Total		1.274	

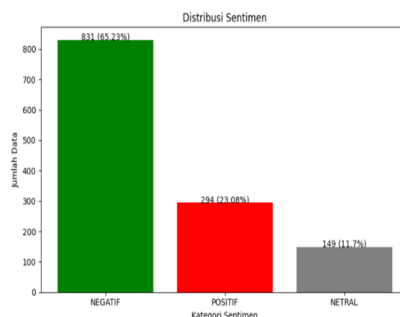
Pemilihan YouTube sebagai sumber data didasarkan pada alasan bahwa platform ini merupakan salah satu media sosial terbesar di Indonesia, dengan tingkat partisipasi publik yang tinggi dalam isu-isu politik dan kebijakan. Oleh karena itu, komentar pada kanal berita resmi maupun diskusi publik dapat merepresentasikan opini publik terkait kebijakan IKN.

3.2 Labelling Data

Proses pelabelan dilakukan oleh H. Irfan Abdul Hakim, S.Sos., M.A., seorang ahli sosiologi dengan latar belakang NLP, text classification, dan linguistik bahasa Indonesia. Hasil pelabelan menunjukkan distribusi kelas yang tidak seimbang, dengan dominasi komentar bernada negatif.

Tabel 3 Hasil Labelling Expert

No.	Komentar	Sentimen
1.	IKN tak begitu urgen yang sangat urgen adalah lapangan kerja , Proyek strategis ini adalah politik ambisiJokowi didukung oleh group sangat2 ,Sengkuni2 Duryodana2 yang duduk di elit politik .	Negatif
	Total	1.274



Gambar 4 Distribusi Data

Hasil labeling, yaitu 831 negatif, 294 positif, dan 149 netral Distribusi ini menunjukkan bahwa sentimen negatif terhadap kebijakan IKN lebih dominan dibandingkan sentimen positif maupun netral. Kondisi ini penting untuk dipertimbangkan dalam interpretasi hasil model, karena ketidakseimbangan kelas berpotensi memengaruhi performa klasifikasi.

3.3 Preprocessing

Sebelum data komentar digunakan, tahap pre-processing dilakukan untuk mendapatkan data bersih. Tahapan yang dilakukan antara lain case folding, cleaning, tokenizing, stopword removal, dan stemming. Berikut ini adalah sebelum dan setelah dilakukan pre-processing.

a. Case Folding

Tahap pertama dalam preprocessing adalah case folding. Case folding adalah proses mengonversi teks menjadi huruf kecil (lowercase) untuk menghilangkan variasi antara huruf besar dan huruf kecil dalam analisis teks. Hasil dari data case folding pada Tabel 4

Tabel 4. Hasil Case Folding

No.	Case Folding
1.	menurut mereka yang merasa paling hebat, ikn adalah simbol kesejahteraan dan kemajuan indonesia, biar tekor asal kesohor, biar miskin yang penting gaya, biar susah yang penting gaya ,susah kalau sdm rendah
Total	1.274

b. Cleaning

Tahap berikutnya adalah membersihkan teks dari karakter yang tidak dibutuhkan. Tujuan dari tahapan ini adalah untuk menghilangkan elemen-elemen yang tidak relevan dan berpotensi mengganggu proses analisis. Komentar berbahasa Indonesia kerap membuat simbol, kata-kata tidak baku, angka, hastag, maupun URL yang perlu disaring terlebih dahulu penghapusan elemen yang tidak relevan dengan analisis sentimen, meliputi: Angka (misalnya 123, 2024), URL (https://...),Hashtag (#IKN, #pindahibukota),Mention (@user),Simbol (misalnya %, \$, &, *), Emotikon, Tanda baca berlebih (!!!, ???). Hasil pembersihan teks (cleaning) ditunjukkan pada Tabel 5.

Tabel 5 Hasil Cleaning

No.	Cleaning
1.	menurut mereka yang merasa paling hebat ikn adalah simbol kesejahteraan dan kemajuan indonesia biar tekor asal kesohor biar miskin yang penting gaya biar susah yang penting gaya susah kalau sdm rendah
Total	1.274

c. Tokenizing

Tahap selanjutnya dalam preprocessing adalah tokenizing, yaitu memecah kalimat menjadi kata-kata individual atau token. Tujuannya adalah untuk mempersiapkan teks agar dapat dianalisis secara lebih mendalam pada tingkat kata. Hasil proses tokenizing ditampilkan pada Tabel 6.

Tabel 6. Hasil Tokenizing

No.	Tokenizing
1.	['menurut', 'mereka', 'yang', 'merasa', 'paling', 'hebat', 'ikn', 'adalah', 'simbol', 'kesejahteraan', 'dan', 'kemajuan', 'indonesia', 'biat', 'tekor', 'asal', 'kesohor', 'biar', 'miskin', 'yang', 'penting', 'gaya', 'biar', 'susah', 'yang', 'penting', 'gaya', 'susah', 'kalau', 'sdm', 'rendah']
Total	1.274

d. Stopword Removal

Tahap berikutnya adalah Stopword Removal digunakan untuk menghapus kata-kata yang tidak memiliki makna penting dalam teks, sehingga hanya tersisa kata-kata yang relevan dan signifikan untuk dianalisis. Stopword dihapus menggunakan daftar default library Sastrawi, yang berisi kata-kata umum seperti “yang,” “dan,” “atau,” “dengan.”. Daftar kata yang dihapus pada tahap stopwords removal dapat dilihat pada Tabel 7.

Tabel 7. Stopword Removal

No.	Stopword Removal
1.	['merasa', 'paling', 'hebat', 'ikn', 'simbol', 'kesejahteraan', 'kemajuan', 'indonesia', 'biat', 'tekor', 'asal', 'kesohor', 'biar', 'miskin', 'penting', 'gaya', 'biar', 'susah', 'penting', 'gaya', 'susah', 'kalau', 'sdm', 'rendah']
Total	1.274

e. Stemming

Tahap *stemming* merupakan proses untuk mengubah kata yang memiliki imbuhan menjadi bentuk dasarnya. alam tahap ini, setiap kata dianalisis untuk menemukan akar katanya dengan memanfaatkan library Sastrawi di Python. Proses stemming dilakukan menggunakan Sastrawi Stemmer. Setiap kata diubah ke bentuk dasar. Hasil stemming ditunjukkan pada Tabel 8.

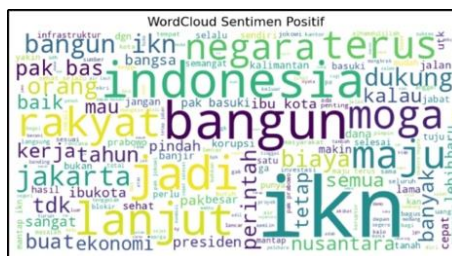
Tabel 8. Hasil Stemming

No.	Stemming
1.	rasa paling hebat ikn simbol sejahtera maju indonesia biat tekor asal kesohor biar miskin penting gaya biar susah penting gaya susah kalau sdm rendah
Total	1.274

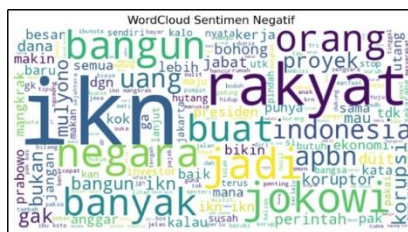
Meskipun IndoBERT secara teori tidak memerlukan stemming karena telah dilatih pada korpus besar bahasa Indonesia, penelitian ini tetap menerapkan stemming. Hal ini diputuskan setelah uji coba awal menunjukkan bahwa komentar YouTube cenderung menggunakan bahasa informal, singkatan, dan ejaan tidak konsisten. Dengan stemming, variasi kata seperti “bangunin”, “bangun”, atau “ngebangun” dapat dipetakan ke bentuk dasar “bangun”, sehingga representasi teks menjadi lebih konsisten. Dampaknya, performa klasifikasi berbasis IndoBERT-KNN menjadi lebih stabil. Dengan demikian, meskipun terdapat pandangan bahwa stemming dapat menurunkan performa BERT, pada penelitian ini justru terbukti membantu karena sesuai dengan karakteristik data yang dianalisis.

3.4 Hasil Visualisasi

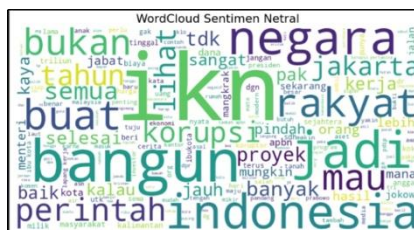
Kata-kata yang paling sering muncul dalam teks ulasan divisualisasikan dalam bentuk wordcloud. Untuk menghasilkan visualisasi ini, peneliti memanfaatkan pustaka Python bernama wordcloud. Visualisasi ini membantu mengidentifikasi kata-kata dominan berdasarkan kategori sentimen [14]. Gambar 5 menunjukkan wordcloud untuk ulasan dengan sentimen positif, Gambar 6 memperlihatkan visualisasi untuk ulasan negatif, dan Gambar 7 menampilkan wordcloud untuk ulasan dengan sentimen netral.



Gambar 5. Wordcloud Sentimen Positif



Gambar 6. Wordcloud Sentimen Negatif



Gambar 7. Wordcloud Sentimen Netral

3.5 Feature Extraction (IndoBERT)

Tahap selanjutnya dalam pemrosesan data adalah *feature extraction* menggunakan model IndoBERT (Indonesian BERT), yaitu versi BERT yang dilatih khusus untuk bahasa Indonesia. Proses ini bertujuan untuk mengubah teks yang telah dibersihkan menjadi representasi numerik (vektor fitur) yang dapat digunakan dalam proses klasifikasi. Pada penelitian ini, digunakan model IndoBERT yang disediakan oleh *Hugging Face Transformers*, dengan bantuan pustaka *transformers* dan *torch* dalam Python. IndoBERT karena kemampuannya dalam memahami konteks teks secara mendalam melalui representasi *bidirectional*.

Vektor ke-1 (10 fitur pertama):	[0.3064 3.2493 -0.4819 -0.0715 1.9914 -0.8849 0.0225 -0.6624 0.5495 0.5617]
Vektor ke-2 (10 fitur pertama):	[0.6857 1.7637 0.499 -0.0982 2.706 -1.9829 0.4727 -1.4455 0.1975 1.3475]
Vektor ke-3 (10 fitur pertama):	[1.014 2.2494 0.3485 -0.0191 2.3269 -1.8388 -0.1789 -0.9249 -0.037 1.6872]
. . .	
Vektor ke-1273 (10 fitur pertama):	[0.6842 0.7408 -0.1562 -0.0116 1.812 -1.6455 -0.1729 -1.0702 0.3615 1.0522]
Vektor ke-1274 (10 fitur pertama):	[0.7036 2.7815 0.1561 -0.643 2.6758 -1.0519 -0.3772 -0.1881 -0.6682 0.5588]

Gambar 8. Hasil Ekstraksi IndoBERT

Gambar 8 memperlihatkan hasil output berupa vektor *embedding* yang dihasilkan oleh model IndoBERT dari sebuah teks masukan. Pada visualisasi ini, ditampilkan 10 fitur pertama dari total 768 fitur (dimensi) untuk beberapa contoh vektor, yaitu vektor ke-1 hingga ke-3, serta vektor ke-1273 dan ke-1274, guna mewakili awal dan akhir dari keseluruhan data berjumlah 1274 entri. Sebagai contoh, vektor ke-1 memiliki nilai sebagai berikut (10 fitur pertama), [0.3064, 3.2493, -0.4819, -0.0715, 1.9914, -0.8849, 0.0225, -0.6624, 0.5495, 0.5617]. Nilai-nilai ini merupakan hasil dari proses *embedding* teks oleh model IndoBERT.

Dalam kasus ini, model IndoBERT menggunakan 768 dimensi sebagai panjang vektor output dari token [CLS] yang digunakan untuk representasi keseluruhan kalimat atau dokumen, sehingga setiap teks direpresentasikan sebagai sebuah vektor berdimensi 768. Dengan kata lain, 768 adalah jumlah fitur numerik atau dimensi representasi dari setiap teks setelah diproses oleh IndoBERT.

3.6 Hasil Evaluasi

Tahap Evaluasi dilakukan untuk mengukur performa model klasifikasi sentimen publik terhadap kebijakan IKN menggunakan kombinasi fitur IndoBERT dan algoritma K-Nearest Neighbor (KNN). Dataset berjumlah 1.274 komentar YouTube yang telah dilabel oleh ahli bahasa. Untuk menghindari overfitting dan bias evaluasi, digunakan teknik 5-Fold Cross Validation. Setiap fold digunakan secara bergantian sebagai data uji dan sisanya sebagai data latih. Hasil akurasi tiap fold ditunjukkan pada Tabel 9, berikut:

Tabel 9. Hasil Evaluasi K-Fold

Fold	Akurasi
1	73,33 %
2	75,69%

Fold	Akurasi
3	70,59%
4	75,69%
5	71,26%
Rata-rata	73.31%

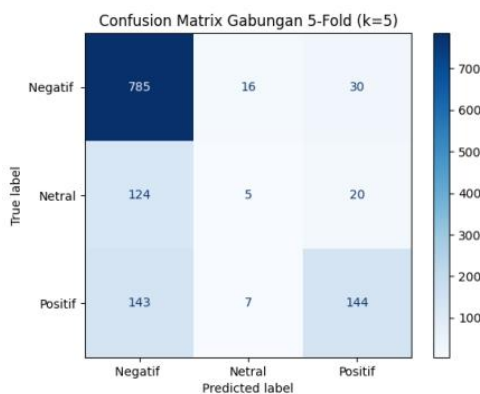
Dari kelima hasil evaluasi tersebut, diperoleh rata-rata akurasi sebesar 0.7331 atau 73.31%, yang menunjukkan bahwa model mampu memprediksi sentimen dengan tingkat ketepatan yang cukup baik. Nilai akurasi ini mencerminkan bahwa model yang dikembangkan dapat bekerja secara konsisten dalam mengenali pola-pola sentimen yang terkandung dalam teks opini masyarakat terkait kebijakan pembangunan Ibu Kota Nusantara (IKN).

Untuk menilai performa model klasifikasi yang digunakan dalam memprediksi sentimen dari teks, dilakukan evaluasi dengan menghitung metrik akurasi. Selanjutnya, persentase setiap kategori dihitung berdasarkan total keseluruhan data. Nilai akurasi model diperoleh dari hasil prediksi yang dilakukan pada data uji, sebagaimana ditampilkan pada Gambar 9.

Akurasi 5-Fold (k=5): 0.7331

Gambar 9. Hasil Klasifikasi KNN

Akurasi dari model diatas adalah akurasi yang digunakan peneliti menggunakan data yang diberi label oleh ahli bahasa sebesar 0.7331 atau 73,31%. Untuk tampilan Confusion Matrix pada Gambar 10.



Gambar 10. Hasil Confusion Matrix

	precision	recall	f1-score	support
Negatif	0.75	0.94	0.83	831.00
Netral	0.18	0.03	0.06	149.00
Positif	0.74	0.49	0.59	294.00
accuracy	0.73	0.73	0.73	0.73
macro avg	0.56	0.49	0.49	1274.00
weighted avg	0.68	0.73	0.69	1274.00

Gambar 11. Confusion Matrix Hasil Klasifikasi Sentimen

Tabel 10. Hasil Confusion Matrix

Kelas Sentimen	Precision	Recall	F1-Score	Support
Negatif	0.75	0.94	0.83	831
Netral	1.00	0.13	0.24	149
Positif	0.95	0.60	0.73	294
Accuaracy			0.73	1274
Macro Avg	0.56	0.49	0.49	1274
Weighted Avg	0.68	0.73	0.69	1274

Tabel 10 menunjukkan hasil evaluasi performa model K-Nearest Neighbor (KNN) dalam mengklasifikasikan data opini publik ke dalam tiga kelas sentimen: Negatif, Netral, dan Positif. Evaluasi dilakukan dengan menggunakan beberapa metrik utama, yaitu precision, recall, F1-score, dan accuracy. Model menunjukkan performa yang cukup baik pada kelas negatif, dengan precision sebesar 0.75, recall 0.94, dan F1-score 0.83. Hal ini berarti model mampu mengenali sebagian besar data dengan sentimen negatif (recall), meskipun masih terdapat beberapa kesalahan klasifikasi (precision).

Dikelas netral performa model terlihat cukup rendah, dengan recall hanya sebesar 0.13, precision sempurna (1.00), dan F1-score 0.24. Meskipun semua prediksi yang diklasifikasikan sebagai netral adalah benar (precision

tinggi), model sangat jarang mengklasifikasikan data sebagai netral (recall rendah), sehingga banyak data netral yang tidak terdeteksi dengan baik. Untuk kelas positif, model mencapai precision sebesar 0.95, recall 0.60, dan F1-score 0.73. Artinya, sebagian besar prediksi positif adalah benar (precision tinggi), tetapi model belum sepenuhnya mampu mendeteksi semua data positif (recall masih menengah).

Berdasarkan confusion matrix, model cenderung lebih akurat dalam mengklasifikasikan data dengan label negatif, sementara kelas netral mengalami banyak kesalahan klasifikasi (hanya 5 yang terklasifikasi benar dari 149 data netral). Hal ini menunjukkan bahwa distribusi data yang tidak seimbang dan kemiripan fitur antara komentar netral dan negatif menjadi tantangan dalam klasifikasi. Adapun perhitungan berdasarkan hasil Gambar 11 pada tabel 11.

Tabel 11. Confusion Matrix Detail Data Uji (IndoBERT + KNN)

Kategori	Prediksi Negatif	Prediksi Netral	Prediksi Positif
Aktual Negatif	785	16	30
Aktual Netral	124	5	20
Aktual Positif	143	7	144

Hasil pengujian model menggunakan data pelabelan ahli menunjukkan akurasi sebesar 0.7331 atau 73,31% pada data uji. Artinya, dari total 1.274 data yang diuji, model mampu melakukan prediksi dengan benar sebanyak 934 kasus. Akurasi ini diperoleh dari perhitungan jumlah prediksi yang benar dibagi dengan total jumlah data uji yang digunakan.

Selain akurasi dan metrik evaluasi lainnya, efisiensi model juga diamati melalui kecepatan proses klasifikasi. Berdasarkan pengujian yang dilakukan menggunakan perangkat laptop standar (AMD Ryzen 5, RAM 8 GB), proses klasifikasi seluruh data uji (1.274 komentar) oleh model KNN hanya memerlukan waktu sekitar 0.24 detik. Hal ini menunjukkan bahwa model ini memiliki kompleksitas komputasi yang rendah dan sangat efisien digunakan bahkan tanpa perangkat keras yang canggih.

4. KESIMPULAN

Penelitian ini berhasil mengimplementasikan kombinasi IndoBERT sebagai ekstraksi fitur dan K-Nearest Neighbor (KNN) sebagai algoritma klasifikasi untuk analisis sentimen terhadap opini publik mengenai kebijakan pemindahan Ibu Kota Negara (IKN). Hasil penelitian menunjukkan akurasi sebesar 73,31% dengan nilai macro-average F1-score sebesar 0,49, yang menandakan performa model masih terbatas dalam mendeteksi kelas minoritas, khususnya pada sentimen positif dan netral. Temuan ini memperlihatkan bahwa meskipun IndoBERT mampu menghasilkan representasi teks yang kaya konteks, penggunaan KNN sebagai classifier cenderung bias terhadap kelas mayoritas. Dengan demikian, kontribusi penelitian ini terletak pada evaluasi efektivitas pendekatan hybrid IndoBERT–KNN dalam konteks analisis sentimen berbahasa Indonesia, sekaligus memberikan gambaran tantangan ketika berhadapan dengan data yang tidak seimbang. Penelitian mendatang disarankan untuk mengeksplorasi model pembandingan seperti SVM, Random Forest, atau fine-tuning IndoBERT, serta memperluas jumlah annotator dan dataset agar dapat meningkatkan reliabilitas dan generalisasi model.

REFERENCES

- [1] A. Wijayanto, "Analisis Sentimen Komentar Youtube Mengenai Vaksin Covid-19 Menggunakan Support Vector Machine," *J. PILAR Teknol. J. Ilm. Ilmu Ilmu Tek.*, vol. 7, no. 1, pp. 24–31, 2022, doi: 10.33319/piltek.v7i1.118.
- [2] Fransiscus and A. S. Girsang, "Sentiment Analysis of COVID-19 Public Activity Restriction (PPKM) Impact using BERT Method," *Int. J. Eng. Trends Technol.*, vol. 70, no. 12, pp. 281–288, 2022, doi: 10.14445/22315381/IJETT-V70I12P226.
- [3] D. N. Larasakti, A. Aziz, and D. Aditya, "Analisis Sentimen Komentar Video YouTube dengan Metode K-Nearest Neighbor," *J. Ilm. Wahana Pendidik.*, vol. 9, no. 5, pp. 132–142, 2023, doi: 10.5281/zenodo.7728573.
- [4] N. P. I. Maharani, A. Purwarianti, Y. Yustiawan, and F. C. Rochim, "Domain-Specific Language Model Post-Training for Indonesian Financial NLP," *Proc. Int. Conf. Electr. Eng. Informatics*, 2023, doi: 10.1109/ICEEI59426.2023.10346625.
- [5] D. G. Mandhasiya, H. Murfi, and A. Bustamam, "The Hybrid of BERT and Deep Learning Models for Indonesian Sentiment Analysis," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 33, no. 1, pp. 591–602, 2024, doi: 10.11591/ijeecs.v33.i1.pp591-602.
- [6] M. Thomson, H. Murfi, and G. Ardaneswari, "BERT-Based Hybrid Deep Learning with Text Augmentation for Sentiment Analysis of Indonesian Hotel Reviews," in *Proceedings of the 12th International Conference on Data Science, Technology and Applications (DATA 2023)*, SCITEPRESS, 2023, pp. 468–473. doi: 10.5220/0012127400003541.
- [7] C. R. Tarumingkeng, *Analisis Sentimen Menggunakan Text Mining*, vol. 3, no. 1–2. Bogor, 2024. doi: 10.1016/0010-4655(90)90107-C.
- [8] N. Husin, "Komparasi Algoritma Random Forest, Naïve Bayes, dan BERT Untuk Multi-Class Classification Pada Artikel Cable News Network (CNN)," *J. Esensi Infokom J. Esensi Sist. Inf. dan Sist. Komput.*, vol. 7, no. 1, pp. 75–84, 2023, doi: 10.55886/infokom.v7i1.608.
- [9] S. Putatunda, A. Bhowmik, G. Thiruvankadam, and R. Ghosh, "A BERT-Based Ensemble Approach for Sentiment Classification of Customer Reviews and Its Application to Nudge Marketing in E-Commerce," 2023. [Online]. Available: <http://arxiv.org/abs/2311.10782>
- [10] M. S. Sayeed, V. Mohan, and K. S. Muthu, "BERT: A Review of Applications in Sentiment Analysis," *HighTech Innov. J.*, vol. 4, no. 2, pp. 453–462, 2023, doi: 10.28991/HIJ-2023-04-02-015.



- [11] I. Nuttakwa, Rudiman, and F. Yulianto, “Maps Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan Samarinda Menggunakan Metode K-Nearest Neighbour (KNN),” *J. Teknol. Inf.*, vol. 18, no. 2, pp. 100–116, 2024.
- [12] A. N. Azhar and L. M. Khodra, “Fine-Tuning Pretrained Multilingual BERT Model for Indonesian Aspect-Based Sentiment Analysis,” in *Proceedings of the 11th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA 2024)*, IEEE, 2024.
- [13] R. Z. Suchrady and A. Purwarianti, “Indo LEGO-ABSA: A Multitask Generative Aspect Based Sentiment Analysis for Indonesian Language,” *Proc. Int. Conf. Electr. Eng. Informatics*, 2023, doi: 10.1109/ICEEI59426.2023.10346852.
- [14] B. Hakim, “Analisa Sentimen Data Text Preprocessing Pada Data Mining Dengan Menggunakan Machine Learning,” *JBASE - J. Bus. Audit Inf. Syst.*, vol. 4, no. 2, pp. 16–22, 2021, doi: 10.30813/jbase.v4i2.3000.
- [15] J. R. Jim, M. A. R. Talukder, P. Malakar, M. M. Kabir, K. Nur, and M. F. Mridha, “Recent Advancements and Challenges of NLP-Based Sentiment Analysis: A State-of-the-Art Review,” *Nat. Lang. Process. J.*, vol. 6, p. 100059, 2024, doi: 10.1016/j.nlp.2024.100059.
- [16] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, “IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP,” in *COLING 2020 - 28th International Conference on Computational Linguistics, Proceedings of the Conference*, Association for Computational Linguistics (ACL), 2020, pp. 757–770. doi: 10.18653/v1/2020.coling-main.66.
- [17] A. Jazuli, Widowati, and R. Kusumaningrum, “Optimizing Aspect-Based Sentiment Analysis Using BERT for Comprehensive Analysis of Indonesian Student Feedback,” *Appl. Sci.*, vol. 15, no. 1, pp. 1–28, 2025, doi: 10.3390/app15010172.
- [18] I. Daqiqil, H. Saputra, Syamsudhuha, R. Kurniawan, and Y. Andriyani, “Sentiment Analysis of Student Evaluation Feedback Using Transformer-Based Language Models,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 36, no. 2, pp. 1127–1139, 2024, doi: 10.11591/ijeecs.v36.i2.pp1127-1139.
- [19] R. Zulcharnain, G. Abdurrahman, and D. Daryanto, “Analisis Sentimen Ulasan Duolingo dengan Metode Algoritma Multinomial Naive Bayes,” *J. Inform. dan Teknol. Pendidik.*, vol. 5, no. 1, pp. 1–16, 2025, doi: 10.59395/jitp.v5i1.113.
- [20] M. D. A. R. Dzakwan and Subektiningsih, “Klasifikasi Tingkat Risiko Kesehatan Ibu Hamil Menggunakan Algoritma Support Vector Machine,” *Indones. J. Comput. Sci.*, vol. 12, no. 5, pp. 2798–2807, 2023, doi: 10.33022/ijcs.v12i5.3372.