

Multiclass Sentiment Analysis of Electric Vehicle Incentive Policies Using IndoBERT and DeBERTa Algorithms

Muhammad Bayu Nugroho ^{1*}, Akhmad Khanif Zyen ^{2*}, Nur Aeni Widiasuti ^{3*}

*Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Nahdlatul Ulama Jepara
211240001153@unisnu.ac.id¹, khanif.zyen@unisnu.ac.id², nuraeniwidiasuti@unisnu.ac.id³

Article Info

Article history:

Received 2025-04-25

Revised 2025-05-07

Accepted 2025-05-09

Keyword:

*Sentiment Analysis,
IndoBERT,
DeBERTa,
Electric Vehicle,
Random Oversampling.*

ABSTRACT

The electric vehicle (EV) incentive policy in Indonesia has generated various public reactions, particularly on social media platforms. This study aims to classify public sentiment using the IndoBERT and DeBERTa transformer models. A total of 6,758 comments were collected from YouTube, filtered, preprocessed, and labeled into three sentiment categories: positive, negative, and neutral. From this, 1,711 clean data points were used and analyzed in two phases: before and after applying the Random Oversampling technique to address class imbalance. Model performance was evaluated using accuracy, precision, recall, F1-score, and training time. In the initial phase, IndoBERT achieved 96% accuracy with 603.71 seconds of training time, while DeBERTa reached 93% in 439.19 seconds. After balancing and applying 5-Fold Cross Validation, IndoBERT maintained 96% accuracy with balanced metric distribution, while DeBERTa recorded 93% accuracy. IndoBERT performed better in recognizing neutral sentiment, whereas DeBERTa was more time-efficient. These results highlight the effectiveness of local transformer models and data balancing techniques in improving sentiment classification performance on imbalanced datasets.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Pemerintah Indonesia telah mendorong akselerasi penggunaan kendaraan listrik sebagai upaya pengurangan emisi karbon dan ketergantungan terhadap bahan bakar fosil. Berbagai kebijakan insentif dikeluarkan seperti kebijakan seperti Peraturan Presiden No. 55 Tahun 2019 dan Peraturan Menteri Keuangan No. 38/PMK.010/2023 telah diterbitkan untuk mendorong percepatan adopsi kendaraan listrik. Namun demikian, penerapan kebijakan ini memicu respons publik yang beragam, khususnya dalam ruang digital seperti media sosial [1]. Komentar yang tersebar di platform YouTube mencerminkan persepsi masyarakat terhadap efektivitas dan relevansi kebijakan tersebut. Oleh karena itu, analisis terhadap data komentar publik dapat menjadi landasan penting dalam memahami sikap masyarakat dan sebagai masukan bagi evaluasi kebijakan [2].

Analisis sentimen menjadi salah satu metode populer untuk mengekstraksi opini dari teks. Model berbasis transformer seperti IndoBERT dan DeBERTa telah banyak digunakan dalam tugas pemrosesan bahasa alami karena kemampuannya

dalam memahami konteks secara mendalam [3]. IndoBERT secara khusus dirancang untuk bahasa Indonesia, sedangkan DeBERTa bersifat multilingual [4].

Namun, tantangan yang sering muncul dalam klasifikasi sentimen adalah ketidakseimbangan distribusi data antar kelas. Kelas mayoritas sering mendominasi, yang berdampak pada rendahnya performa model terhadap kelas minoritas. Untuk mengatasi masalah ini, diperlukan teknik penyeimbangan data seperti Random Oversampling, namun perlu dilakukan validasi untuk menghindari overfitting [5].

Berbagai studi sebelumnya telah menerapkan pendekatan machine learning maupun deep learning untuk melakukan klasifikasi sentimen pada data media sosial. Alhasil, misalnya, menggunakan algoritma Support Vector Machine (SVM) dan berhasil mencapai akurasi sebesar 87% [6]. Penelitian lainnya melaporkan peningkatan akurasi dari 79,43% menjadi 82,11% melalui penerapan algoritma Naive Bayes dengan penambahan seleksi fitur [7]. Di sisi lain, studi yang menganalisis opini masyarakat Indonesia terhadap pengungsi Rohingya di media sosial X menggunakan model Long Short-Term Memory (LSTM), menemukan bahwa dari 17.613 tweet

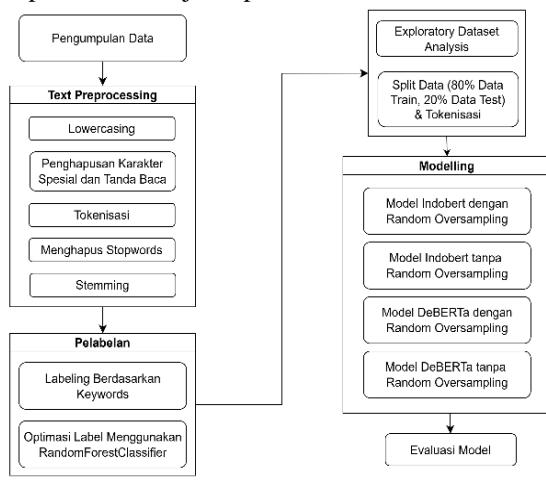
yang dianalisis, sekitar 73,04% mengandung sentimen negatif, sementara 26,96% menunjukkan sentimen positif [8]. Meski metode-metode tersebut menunjukkan kinerja optimal, keterbatasannya terletak pada kurangnya pemahaman terhadap kompleksitas bahasa Indonesia, terutama dalam konteks kalimat yang lebih rumit dan unik.

Hadirnya model transformer seperti IndoBERT dan DeBERTa menjadi alternatif baru dalam mengatasi keterbatasan tersebut. IndoBERT, yang dirancang khusus dengan pelatihan pada korpus bahasa Indonesia, telah menunjukkan performa yang menjanjikan dalam berbagai tugas Natural Language Processing (NLP) [9]. Sementara itu, DeBERTa yang mengandalkan pendekatan disentangled attention dinilai lebih unggul dalam menangkap hubungan semantik antar kata [4]. Penelitian lain menunjukkan bahwa IndoBERT yang dilatih pada data yang telah diseimbangkan dengan Random Oversampling mengalami peningkatan akurasi dari 84% menjadi 92%, serta perbaikan recall dan F1-score pada kelas minoritas [9]. Di sisi lain, DeBERTa tercatat memberikan performa yang lebih tinggi dibandingkan sejumlah model NLP lainnya seperti BERT, RoBERTa, dan XLNet, dengan akurasi mencapai 94,2%—lebih tinggi dibanding BERT (91,5%) dan RoBERTa (92,3%) [4].

Penelitian ini bertujuan membandingkan performa model IndoBERT dan DeBERTa dalam klasifikasi sentimen komentar publik mengenai kebijakan insentif kendaraan listrik di Indonesia, serta mengevaluasi dampak penerapan teknik Random Oversampling dalam meningkatkan performa model, terutama pada kelas yang kurang terwakili atau minoritas.

II. METODE

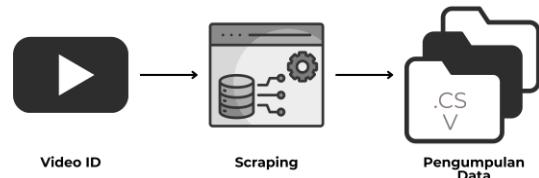
Penelitian ini terdiri dari sejumlah tahapan inti, meliputi proses pengumpulan data, pembersihan dan persiapan data (preprocessing), pemberian label sentimen, eksplorasi karakteristik data, penyeimbangan distribusi kelas, implementasi model IndoBERT dan DeBERTa, serta pengukuran performa model. Gambaran umum dari alur proses penelitian disajikan pada Gambar 1.



Gambar. 1 Alur Proses Penelitian

A. Pengumpulan Data

Pengambilan data dilakukan melalui proses scraping menggunakan YouTube API, yang memungkinkan akses terstruktur terhadap data dari platform tersebut. Seperti pada gambar 2, tahapan dimulai dengan konfigurasi logging untuk mencatat seluruh aktivitas selama proses berjalan, diikuti dengan pengaturan API key guna menjalin koneksi ke endpoint API. Setelah koneksi berhasil, data komentar dikumpulkan melalui permintaan API dan disimpan dalam format CSV yang kemudian data siap dianalisis lebih lanjut [10].



Gambar. 2 Alur Proses Pengumpulan Data

Data dikumpulkan melalui proses scraping menggunakan YouTube Data API v3 yang dijalankan di platform Google Colaboratory. Pengambilan komentar dilakukan berdasarkan sejumlah kata kunci relevan “kebijakan insentif kendaraan listrik”, “subsidi kendaraan listrik”, “motor listrik”, “EV”, “pajak kendaraan listrik”, dan “mobil listrik”. Komentar yang dianalisis berasal dari 70 video pilihan sesuai dengan tema objek, dengan maksimal 500 komentar per video. Dari proses tersebut diperoleh sebanyak 6.758 komentar. Data hasil scraping memiliki struktur kolom yang terdiri atas video_id, author, comment, dan published_at.

B. Filtering dan Validasi Data

Dari total 6.758 komentar yang diperoleh melalui scraping pada platform YouTube, dilakukan proses penyaringan untuk memastikan kualitas dan relevansi data yang akan dianalisis. Tahap awal mencakup penghapusan komentar yang tidak sesuai dengan konteks penelitian, seperti spam, promosi judi online, komentar kosong, penggunaan emoji atau karakter tunggal, serta duplikasi komentar. Selain itu, komentar yang terlalu singkat atau tidak mengandung informasi bermakna juga dieliminasi secara manual. Setelah proses ini, tersisa 5.562 komentar yang dinilai layak untuk diproses lebih lanjut.

Selanjutnya, dilakukan proses pelabelan sentimen dengan pendekatan semi-otomatis menggunakan kombinasi metode berbasis kata kunci dan di optimasi menggunakan algoritma Random Forest Classifier yang dijelaskan pada subbab D tentang pelabelan. Setelah keseluruhan proses filtering dan pelabelan, diperoleh 1.711 komentar yang digunakan sebagai dataset utama. Komentar yang masuk dalam kategori “unknown” dieliminasi karena tidak memenuhi kriteria klasifikasi.

C. Text Preprocessing

Text preprocessing merupakan langkah awal yang krusial dalam pemrosesan bahasa alami (Natural Language Processing/NLP). Tahapan ini bertujuan untuk

membersihkan, menormalkan, dan menyederhanakan teks agar dapat diproses secara efektif oleh algoritma. Melalui proses ini, elemen-elemen yang tidak diperlukan dapat dihilangkan tanpa mengurangi informasi penting dalam teks [11]. Ilustrasi alur preprocessing ditampilkan pada Gambar 3.

Tahapan Text Preprocessing



Gambar. 3 Tahapan Text Preprocessing

Adapun tahapan yang diterapkan dalam proses text preprocessing meliputi [6]:

- Case folding, mengubah seluruh huruf menjadi huruf kecil guna menjaga konsistensi teks.
- Pembersihan karakter non-alfabet, menghapus simbol, angka, dan tanda baca yang tidak relevan dengan konteks analisis.
- Tokenisasi, memecah kalimat menjadi unit-unit terkecil seperti kata atau frasa.
- Penghapusan stopwords mengeliminasi kata-kata umum seperti "dan", "yang", atau "di" yang tidak memiliki kontribusi signifikan dalam analisis.
- Stemming, mengembalikan kata ke bentuk dasarnya menggunakan algoritma Sastrawi yang telah diadaptasi untuk teks berbahasa Indonesia.

TABEL 1
IMPLEMENTASI TEXT PREPROCESSING

Proses	Teks Asli	Hasil
Case Folding	"Mobil listrik?? Sangat luar biasa!!"	"mobil listrik?? sangat luar biasa!!"
Penghapusan Simbol	"mobil listrik?? sangat luar biasa!!"	"mobil listrik sangat luar biasa"
Tokenisasi	"mobil listrik sangat luar biasa"	["mobil", "listrik", "sangat", "luar", "biasa"]
Penghapusan Stopwords	["mobil", "listrik", "sangat", "luar", "biasa"]	["mobil", "listrik", "luar", "biasa"]
Stemming	mobil listrik sangat luar biasa"	"mobil listrik luar biasa"

D. Pelabelan Data

Pelabelan data dalam penelitian ini dilakukan secara semi-otomatis dengan pendekatan pseudo-labeling untuk mengklasifikasikan komentar ke dalam tiga kategori sentimen (multiclass): positif, netral, dan negatif. Sentimen positif mencerminkan dukungan atau pandangan optimis terhadap kebijakan kendaraan listrik; negatif menunjukkan kritik, keluhan, atau ketidakpuasan; sedangkan netral bersifat objektif, deskriptif, atau tidak memuat muatan emosional.

Proses pelabelan terdiri dari dua tahap. Pada tahap pertama, sistem menganalisis komentar menggunakan metode berbasis kata kunci. Komentar yang cocok dengan kata kunci tertentu secara otomatis diberi label sesuai kategori sentimen.

Komentar yang tidak dapat diklasifikasikan melalui pendekatan ini diberi label "unknown" dan dikeluarkan dari proses pelabelan lanjutan untuk menjaga kualitas data yang digunakan dalam pelatihan model.

Tahap kedua merupakan proses optimasi pelabelan dengan memanfaatkan algoritma Random Forest Classifier. Model ini dilatih menggunakan data hasil pelabelan awal dengan representasi fitur berbasis TF-IDF. Pendekatan ini sejalan dengan strategi semi-supervised learning seperti pseudo-labeling, sebagaimana digunakan dalam penelitian Ayuningtyas & Yustanti yang memungkinkan pemanfaatan data dalam skala besar meskipun pelabelan manual hanya dilakukan secara terbatas [12]. Evaluasi performa model dilakukan menggunakan pembagian data latih dan uji sebesar 80:20, dengan metrik akurasi dan F1-score. Model menunjukkan performa yang baik dengan akurasi 94% dan skor F1 makro sebesar 0.91.

TABEL 2
CLASSIFICATION REPORT HASIL OPTIMASI PELABELAN MENGGUNAKAN RANDOM FOREST CLASSIFIER

Sentimen	Precision	Recall	F1-Score	Support
Negatif	0.92	0.94	0.93	205
Netral	0.98	0.99	0.98	205
Positif	0.92	0.90	0.91	206
Accuracy			0.94	616
Macro avg		0.94	0.94	616
Weighted avg		0.94	0.94	616

Selain itu, untuk mengukur konsistensi pelabelan, dilakukan validasi manual oleh dua annotator independen pada 200 sampel komentar. Hasil perbandingan antar annotator menghasilkan nilai Cohen's Kappa sebesar 0.81, yang mengindikasikan tingkat kesepakatan yang tinggi atau berada pada kategori substantial agreement. Proses ini menghasilkan 1.711 komentar yang telah terkласifikasi secara sistematis, dengan distribusi seperti pada Tabel 3.

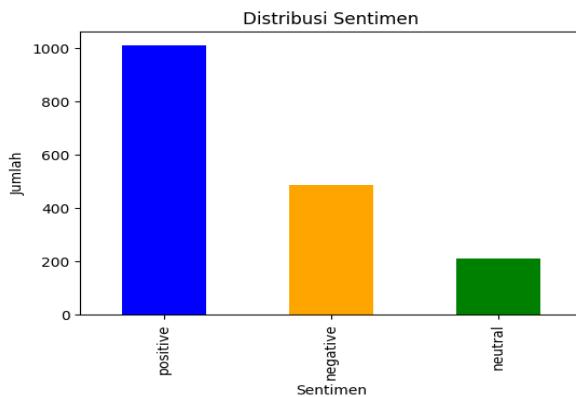
TABEL 3
DISTRIBUSI SENTIMEN TIAP KELAS

Sentimen	Jumlah (Komentar)
Positif	1013
Netral	210
Negatif	488

E. Exploratory Data Analysis

Exploratory Data Analysis (EDA) atau eksplorasi data dilakukan untuk memperoleh pemahaman yang mendalam mengenai struktur dan karakteristik dataset yang digunakan [13]. Tahapan ini mencakup analisis distribusi sentimen serta pengenalan terhadap pola kata yang dominan melalui berbagai visualisasi, seperti histogram, box plot, WordCloud, dan analisis frekuensi kata. Visualisasi-visualisasi tersebut berperan penting dalam mengungkap pola bahasa serta persebaran sentimen yang nantinya akan berpengaruh pada proses pemodelan [14]. Sebagaimana ditampilkan pada

Gambar 4, divisualisasikan distribusi sentimen hasil akhir pelabelan.



Gambar. 4 Distribusi Sentimen Hasil Akhir Pelabelan

Selanjutnya, Gambar 5, 6, dan 7 menyajikan visualisasi WordCloud untuk masing-masing kategori sentimen, sedangkan Gambar 8 memperlihatkan kata-kata yang paling sering muncul secara keseluruhan di semua kategori sentimen.



Gambar. 5 Word Cloud Sentimen Negatif

Word cloud pada gambar 5 diatas tersebut menampilkan kata-kata yang paling sering muncul pada komentar bernada negatif tentang kendaraan listrik di Indonesia. Kata-kata seperti "mahal," "harga," "mobil," dan "listrik" mendominasi, mengindikasikan bahwa biaya tinggi menjadi keluhan utama masyarakat.



Gambar. 6 Word Cloud Sentimen Netral

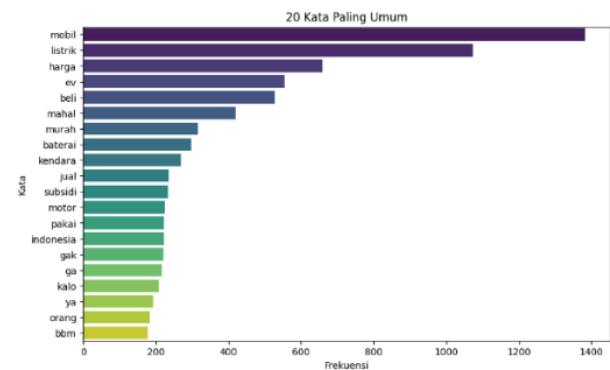
Gambar 6 diatas ini menampilkan kata-kata yang sering muncul dalam komentar dengan sentimen netral mengenai kendaraan listrik di Indonesia. Kata-kata seperti "moga,"

"mending," "beli," "mobil," dan "subsidi" mendominasi, menunjukkan adanya harapan, pertimbangan, dan perbandingan dalam keputusan pembelian kendaraan listrik.



Gambar. 7 Word Cloud Sentimen Positif

Word cloud pada gambar 7 menampilkan kata-kata yang sering muncul dalam komentar dengan sentimen positif mengenai kendaraan listrik di Indonesia. Dominasi kata seperti "mobil," "listrik," "harga," "murah," dan "untung" mencerminkan apresiasi publik terhadap manfaat, harga terjangkau, dan insentif yang mendukung penggunaan kendaraan listrik.



Gambar. 8 Top 20 Kata Paling Umum Muncul

Grafik batang pada gambar 8 menampilkan 20 kata yang paling sering digunakan dalam komentar mengenai kendaraan listrik. Kata-kata seperti "mobil," "listrik," dan "harga" muncul dengan frekuensi tertinggi, menandakan bahwa perhatian utama publik tertuju pada kendaraan listrik dan aspek harga. Selain itu, kata-kata seperti "ev," "beli," "mahal," dan "murah" mencerminkan fokus masyarakat pada proses pembelian dan pertimbangan biaya.

F. Splitting Data

Proses pembagian dataset menjadi data pelatihan dan data pengujian merupakan tahap penting dalam pengembangan dan evaluasi model. Langkah ini memungkinkan model untuk mempelajari pola dari sebagian data (data pelatihan) dan diuji pada bagian data lainnya (data pengujian), sehingga dapat diperoleh estimasi performa model yang lebih objektif. Dalam penelitian ini, digunakan pembagian dengan rasio 80% untuk data pelatihan dan 20% untuk data pengujian. Proses pembagian dilakukan menggunakan fungsi `train_test_split`

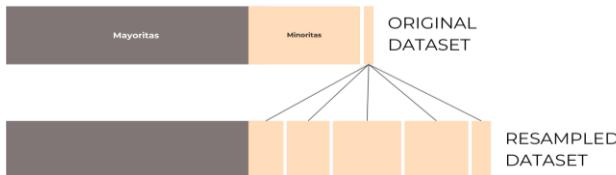
dari pustaka scikit-learn, dengan penetapan parameter random_state guna menjamin reproduksibilitas hasil [14]. Selanjutnya, setiap label sentimen dikodekan sebagai berikut:

- Label 0 untuk sentimen negatif
- Label 1 untuk sentimen netral
- Label 2 untuk sentimen positif.

G. Random Oversampling

Dalam Natural Language Processing (NLP), augmentasi data digunakan untuk meningkatkan kemampuan generalisasi model dengan menciptakan variasi pada teks. Pendekatan yang umum meliputi transformasi token seperti back-translation dan augmentasi kontekstual berbasis model deep learning [15]. Random Oversampling digunakan untuk mengatasi ketidakseimbangan kelas karena metode ini sederhana, cocok untuk dataset kecil, dan tidak mengurangi data kelas mayoritas. Teknik augmentasi semantik kurang sesuai karena komentar pendek dalam bahasa Indonesia memerlukan kurasi yang kompleks [16]. Meskipun efektif, Random Oversampling dapat memicu overfitting karena hanya menduplikasi data tanpa menambah variasi. Untuk menguji generalisasi model secara menyeluruh, evaluasi lanjutan seperti k-fold cross-validation sangat disarankan [17]. Gambar 9 merupakan penerapan oversampling.

Oversample Kelas Minoritas



Gambar. 9 Konsep Oversampling Kelas Minoritas

H. K-Fold Cross Validation

K-Fold Cross Validation adalah teknik validasi silang yang membagi dataset menjadi k bagian (fold) untuk mengevaluasi model secara lebih andal, khususnya saat data terbatas atau berisiko overfitting. Model dilatih dan diuji sebanyak k kali, dengan setiap fold bergiliran menjadi data uji. Rata-rata hasil evaluasi dari seluruh fold digunakan untuk mengukur performa model secara menyeluruh. K-Fold membantu menilai model secara adil dan mengurangi risiko overfitting yang mungkin timbul akibat Random Oversampling [18]. Berikut ini adalah ilustrasi penerapan K-Fold Cross Validation pada gambar 10.

	Test	Train	Train	Train	Train
Iteration 1					
Iteration 2	Train	Test	Train	Train	Train
Iteration 3	Train	Train	Test	Train	Train
Iteration 4	Train	Train	Train	Test	Train
Iteration 5	Train	Train	Train	Train	Test

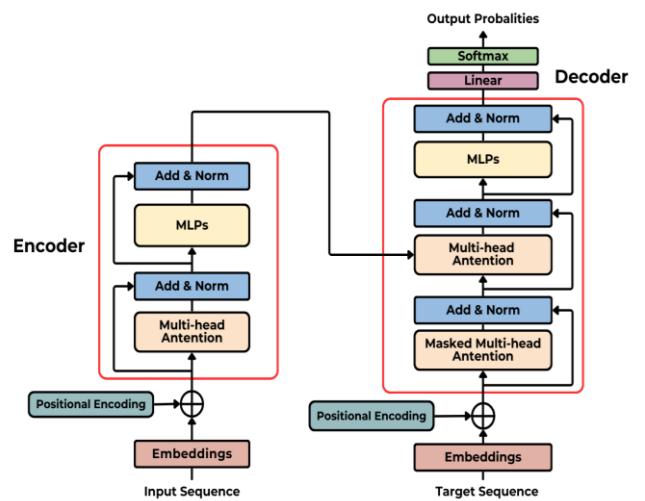
Gambar. 10 Penerapan K-Fold Cross Validation

I. Hyperparameter Tuning

Untuk mencapai kinerja model IndoBERT yang optimal, dilakukan penyesuaian terhadap sejumlah hyperparameter selama proses pelatihan. Learning rate disesuaikan dalam kisaran $1e-5$ hingga $5e-5$ guna menjaga stabilitas dan konvergensi selama proses pembelajaran. Batch size ditentukan antara 16 hingga 32, dengan mempertimbangkan keterbatasan kapasitas GPU yang digunakan. Model dilatih selama 3 hingga 5 epoch, yang dinilai cukup untuk mencapai akurasi terbaik tanpa memicu risiko overfitting. Panjang maksimum sequence diatur dalam rentang 128 hingga 512 token, sebagai upaya menyeimbangkan cakupan konteks dengan efisiensi komputasi. Untuk proses optimasi, digunakan optimizer AdamW yang dikenal stabil dalam memperbarui bobot model, serta diterapkan mekanisme dropout sebesar 0.1 sebagai langkah preventif terhadap overfitting [19].

J. Algoritma IndoBERT

IndoBERT IndoBERT adalah model berbasis arsitektur BERT yang dikembangkan khusus untuk tugas-tugas pemrosesan bahasa alami (NLP) dalam bahasa Indonesia. Model ini dilatih menggunakan korpus besar berbahasa Indonesia, sehingga mampu mengenali karakteristik khas seperti morfologi aglutinatif, struktur kalimat bebas, idiom lokal, serta kosakata informal (slang) yang sering muncul dalam komunikasi digital [1]. Seperti pada gambar 11, IndoBERT memanfaatkan dua teknik pretraining utama, yaitu Masked Language Modeling (MLM) dan Next Sentence Prediction (NSP), yang bertujuan untuk memahami hubungan antar kata dan antar kalimat dalam konteks lokal bahasa Indonesia. Secara arsitektural, model ini mengadopsi konfigurasi BERT-base, dengan 12 lapisan encoder, dimensi embedding 768, dan total parameter sekitar 110 juta [19].



Gambar. 11 Arsitektur Umum pada Model Transformer

K. Algoritma DeBERTa

DeBERTa (Decoding-enhanced BERT with disentangled attention) adalah model transformer lanjutan yang

dikembangkan untuk mengatasi keterbatasan representasi pada model BERT standar. Keunggulan utama DeBERTa terletak pada mekanisme disentangled attention, di mana informasi posisi kata dan konten semantik diproses secara terpisah. Hal ini memungkinkan model menangkap hubungan antar kata dengan lebih presisi, terutama pada kalimat yang memiliki struktur kompleks. Selain itu, DeBERTa menggunakan enhanced mask decoder, yang meningkatkan kemampuan dalam memprediksi kata-kata yang disembunyikan selama pelatihan (masked language modeling) [4].

Dengan kombinasi kedua fitur tersebut, DeBERTa mampu mempelajari konteks linguistik baik secara global (antar kalimat) maupun lokal (dalam kalimat) secara bersamaan. Hal ini menjadikan DeBERTa baik dalam berbagai tugas NLP, terutama pada bahasa-bahasa yang memiliki struktur fleksibel atau kalimat panjang. Namun, karena kompleksitas arsitekturnya lebih tinggi dibandingkan model seperti IndoBERT, waktu pelatihan DeBERTa cenderung lebih cepat pada dataset kecil, tetapi dapat lebih berat secara komputasi saat diterapkan pada data skala besar [4].

L. Evaluasi

Dalam proses klasifikasi sentimen, confusion matrix merupakan salah satu metode evaluasi yang sering digunakan. Metode ini disajikan dalam bentuk tabel yang menggambarkan distribusi hasil prediksi model, baik yang tepat maupun yang keliru, terhadap masing-masing kelas target. Confusion matrix sangat berguna untuk mengidentifikasi area kesalahan klasifikasi serta keberhasilan model dalam mengenali setiap kategori, sehingga dapat dijadikan dasar untuk perbaikan model pada tahapan iterasi berikutnya [20].

		A	B	C
True Class	A	TP	FN	FN
	B	FP	TN	FN
	C	FP	FN	TN
Predicted Class				

Gambar. 12 Konsep Confusion Matrix

Dimana label-label pada confusion matrix meliputi:

- True Positive (TP), yaitu prediksi positif yang sesuai
- True Negative (TN), prediksi negatif yang benar
- False Positive (FP), prediksi positif yang salah
- False Negative (FN), prediksi negatif yang keliru.

Berdasarkan confusion matrix tersebut, performa model kemudian dievaluasi menggunakan sejumlah metrik klasifikasi umum [21], di antaranya:

- Akurasi (Accuracy) mengukur proporsi prediksi benar terhadap seluruh prediksi dan dihitung dengan rumus persamaan 1.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

- Presisi (Precision) menunjukkan ketepatan model dalam memprediksi kelas positif, penting saat kesalahan positif berdampak besar. Dirumuskan pada persamaan 2.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

- Recall mengukur kemampuan model mendeteksi seluruh data positif, relevan saat kesalahan negatif perlu diminimalkan. Dirumuskan pada persamaan 3.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

- F1-score adalah rata-rata harmonik antara presisi dan recall, berguna dalam kondisi data tidak seimbang. Dirumuskan pada persamaan 4.

$$F1 - Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (4)$$

III. HASIL DAN PEMBAHASAN

Setelah melalui proses pelabelan dan filtering, diperoleh 1.711 data bersih dengan distribusi sebagai berikut: 1.013 positif, 488 negatif, dan 210 netral. Berdasarkan hasil pengujian, model IndoBERT dan DeBERTa menunjukkan performa yang sangat baik dalam mengklasifikasikan sentimen publik terhadap kebijakan insentif kendaraan listrik. Evaluasi dilakukan melalui dua skenario, yakni tanpa penerapan teknik Random Oversampling dan dengan penerapan teknik tersebut sebagai metode penyeimbang distribusi data. Dengan menerapkan hyperparameter tuning seperti pada tabel 4.

TABEL 4
HYPERPARAMETER TUNING MODEL INDOBERT DAN DEBERTA

Hyperparameter	Rentang/Nilai			
	Tanpa Random Oversampling		Dengan Random Oversampling	
	Indo BERT	De BERTa	Indo BERT	De BERTa
Learning Rate	3e-6	2e-5	2e-5	1e-5
Batch Size	16	16	16	16
Epoch	5	5	5	5
Max Sequence Length	512	512	512	512
Optimizer	AdamW	AdamW	AdamW	AdamW

A. Fase 1: Evaluasi Model Tanpa Random Oversampling

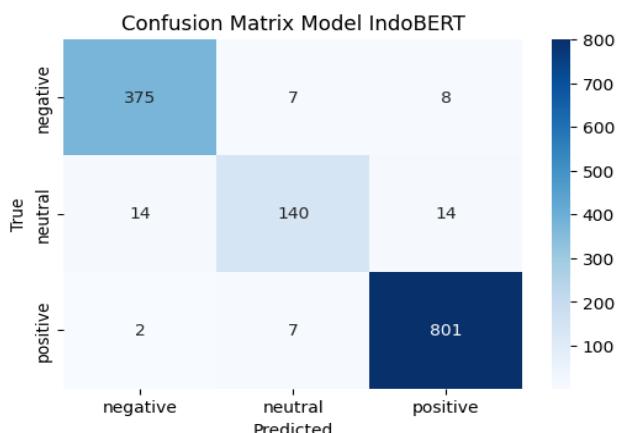
Pada tahap awal, model IndoBERT dan DeBERTa diuji menggunakan data yang belum diseimbangkan, yang telah melalui proses preprocessing dan pelabelan dalam tiga kategori sentimen, yaitu positif, netral, dan negatif. Rincian performa masing-masing model berdasarkan kategori sentimen tanpa penerapan Random Oversampling dapat dilihat pada Tabel 5, 6, 7 dan gambar 13, 14, 15 berikut.

TABEL 5
PERBANDINGAN EVALUASI MODEL TANPA RANDOM OVERSAMPLING

Model	Accuracy	Precision	Recall	F1-Score
IndoBERT	0.96	0.96	0.96	0.96
DeBERTa	0.93	0.93	0.93	0.93

TABEL 6
CLASSIFICATION REPORT MODEL INDOBERT TANPA RANDOM
OVERSAMPLING

Sentimen	Precision	Recall	F1-Score	Support
Negatif	0.96	0.96	0.96	390
Netral	0.91	0.83	0.87	168
Positif	0.97	0.99	0.98	810
Accuracy			0.96	1368
Macro avg	0.95	0.93	0.94	1368
Weighted avg	0.96	0.96	0.96	1368

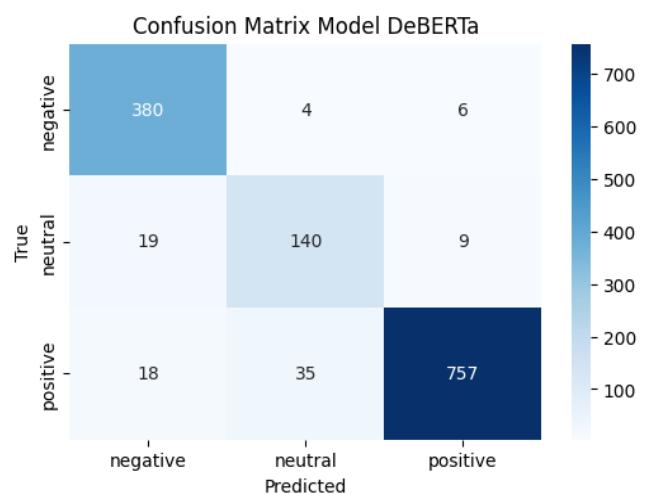


Gambar. 13 Confusion Matrix Model IndoBERT

Model IndoBERT menghasilkan akurasi tinggi sebesar 96%, dengan performa sangat baik pada kategori positif (f1-score 0.98) dan negatif (f1-score 0.96). Namun, performa pada kategori netral masih kurang optimal (f1-score 0.87), kemungkinan besar karena jumlah data netral yang lebih sedikit dan sifatnya yang sulit dibedakan secara semantik. Namun banyak data netral diklasifikasikan keliru sebagai positif atau negatif.

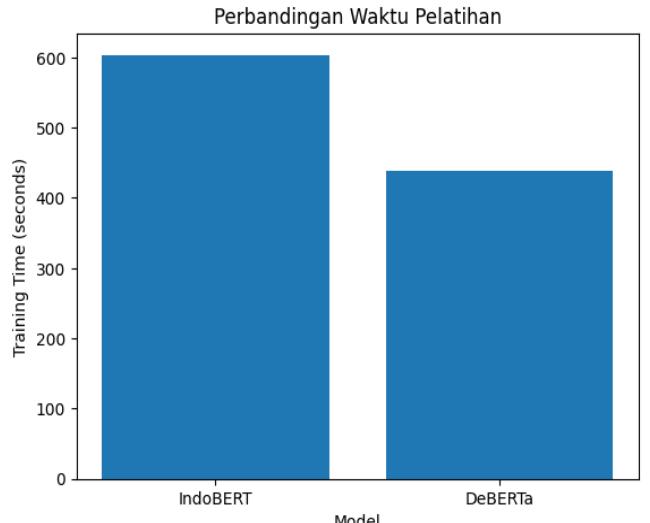
TABEL 7
CLASSIFICATION REPORT MODEL DEBERTA TANPA RANDOM
OVERSAMPLING

Sentimen	Precision	Recall	F1-Score	Support
Negatif	0.91	0.97	0.94	390
Netral	0.78	0.83	0.81	168
Positif	0.98	0.93	0.96	810
Accuracy			0.93	1368
Macro avg	0.89	0.91	0.90	1368
Weighted avg	0.94	0.93	0.93	1368



Gambar. 14 Confusion Matrix Model DeBERTa

Model DeBERTa menunjukkan akurasi 93%, dengan performa paling baik pada kelas positif (f1-score 0.96), disusul kelas negatif (f1-score 0.94), dan terendah pada kelas netral (f1-score 0.81). Meski kinerjanya secara umum baik, model masih mengalami kesulitan dalam mengenali sentimen netral, yang tercermin dari confusion matrix terdapat 35 prediksi salah dari positif ke netral, 19 dari negatif ke netral, dan 9 data netral justru diklasifikasi sebagai positif.



Gambar. 15 Perbandingan Waktu Pelatihan Model tanpa Random
Oversampling

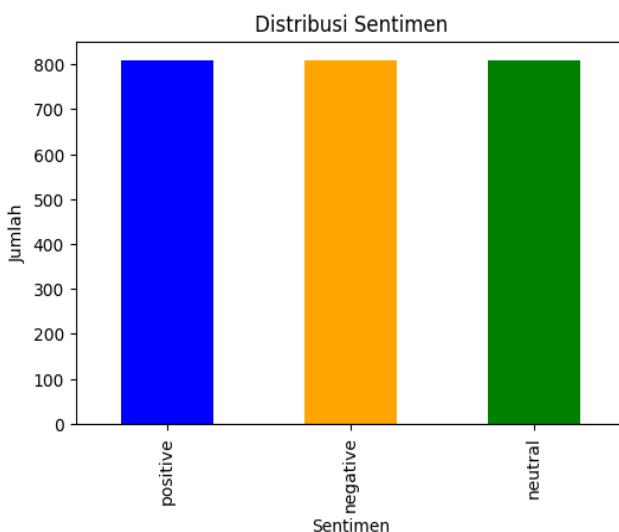
Berdasarkan grafik perbandingan waktu pelatihan, IndoBERT memerlukan waktu 603,71 detik, sedangkan DeBERTa hanya 439,19 detik. Hal ini menunjukkan bahwa DeBERTa lebih efisien secara waktu, dengan pelatihan sekitar 27% lebih cepat.

Meskipun waktu pelatihan DeBERTa lebih singkat, hasil evaluasi performa menunjukkan bahwa IndoBERT lebih unggul dari segi akurasi dan konsistensi klasifikasi, terutama pada kelas netral yang cenderung lebih sulit. IndoBERT

mencatat akurasi keseluruhan sebesar 96%, dibandingkan 93% pada DeBERTa, dengan f1-score yang lebih tinggi secara rata-rata dan distribusi prediksi yang lebih stabil. Perlu digaris bawahi juga distribusi data yang tidak seimbang pada fase ini menyebabkan bias ke kelas dominan (positif), sehingga diperlukan tahap lanjutan dengan teknik penyeimbangan data.

B. Fase 2: Evaluasi Model Dengan Random Oversampling

Pada fase kedua, dilakukan penyeimbangan data menggunakan teknik Random Oversampling, di mana jumlah data pada kelas minoritas (netral dan negatif) ditambah hingga setiap kelas memiliki jumlah data yang seimbang terlihat pada gambar 16, tiap kelas sentimen berjumlah 810 komentar sehingga total keseluruhan 2430 komentar.



Gambar. 16 Distribusi Label Setelah Penerapan Random Oversampling

Adapun performa model dapat dilihat pada Tabel 8, 9, 10 dan Gambar 17 dan 18.

TABEL 8

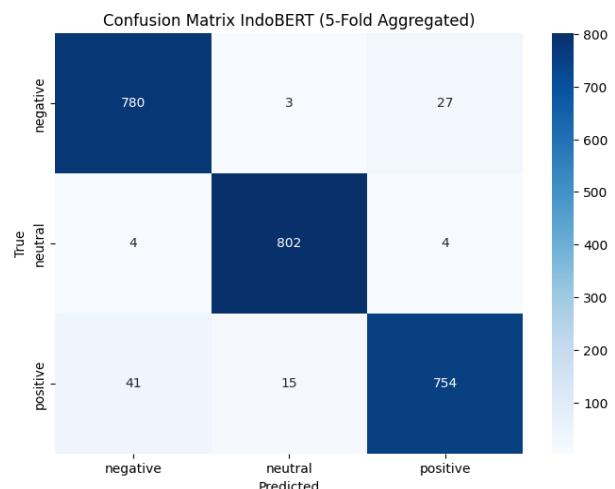
PERBANDINGAN EVALUASI MODEL DENGAN RANDOM OVERSAMPLING

Model	Accuracy	Precision	Recall	F1-Score
IndoBERT	0.9922	0.9923	0.9922	0.9922
DeBERTa	0.9320	0.9324	0.9320	0.9320

TABEL 9

CLASSIFICATION REPORT INDOBERT DENGAN RANDOM OVERSAMPLING

Sentimen	Precision	Recall	F1-Score	Support
Negatif	0.95	0.96	0.95	810
Netral	0.98	0.99	0.98	810
Positif	0.96	0.93	0.95	810
Accuracy			0.96	2430
Macro avg	0.96	0.96	0.96	2430
Weighted avg	0.96	0.96	0.96	2430

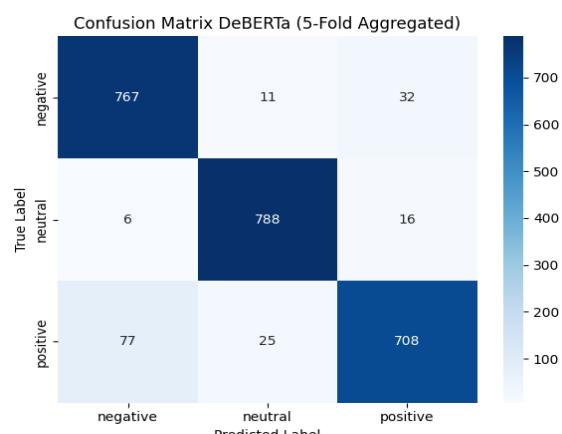


Gambar. 17 Confusion Matrix Model IndoBERT

Setelah diterapkan random oversampling untuk menangani ketidakseimbangan kelas serta 5-Fold Cross Validation untuk meningkatkan reliabilitas evaluasi, model IndoBERT menunjukkan performa yang sangat baik dalam klasifikasi sentimen. Model mencapai akurasi 96% dan rata-rata F1-score 0.96, dengan distribusi performa yang seimbang di seluruh kelas (negatif, netral, positif). Hasil confusion matrix menunjukkan bahwa model mampu mengklasifikasikan data dengan akurat, terutama pada kelas netral yang hampir seluruhnya diprediksi dengan benar (160 dari 162). Kesalahan antar kelas juga sangat kecil, seperti hanya 10 data positif yang diklasifikasikan sebagai negatif.

TABEL 10
CLASSIFICATION REPORT DEBERTA DENGAN RANDOM OVERSAMPLING

Sentimen	Precision	Recall	F1-Score	Support
Negatif	0.90	0.95	0.92	810
Netral	0.96	0.97	0.96	810
Positif	0.94	0.87	0.90	810
Accuracy			0.93	2430
Macro avg	0.93	0.93	0.93	2430
Weighted avg	0.93	0.93	0.93	2430



Gambar. 18 Confusion Matrix Model DeBERTa

Confusion matrix pada gambar 18, setelah dilakukan Random Oversampling dan validasi menggunakan 5-Fold Cross Validation, model DeBERTa mengalami peningkatan performa yang signifikan. Akurasi keseluruhan tercatat 93%, dengan distribusi nilai precision, recall, dan F1-score yang cukup merata untuk setiap kategori sentimen. Meskipun demikian, kesalahan klasifikasi masih terjadi, terutama pada kelas positif yang cukup sering diklasifikasikan sebagai negatif, serta kelas netral yang memiliki nilai recall sedikit lebih rendah (0.97).

IV. KESIMPULAN

Hasil penelitian ini mengindikasikan bahwa model IndoBERT dan DeBERTa memiliki kemampuan yang baik dalam mengklasifikasikan sentimen masyarakat terhadap kebijakan insentif kendaraan listrik. Pada pengujian awal tanpa teknik penyeimbangan data, IndoBERT mencatat akurasi tertinggi sebesar 96%, mengungguli DeBERTa yang mencapai 93%. Meski demikian, kedua model mengalami penurunan kinerja pada kelas netral akibat distribusi data yang tidak merata. Dari sisi efisiensi, DeBERTa menunjukkan waktu pelatihan yang lebih singkat dibanding IndoBERT.

Setelah penerapan Random Oversampling dan validasi dengan 5-Fold Cross Validation, performa kedua model mengalami peningkatan. IndoBERT tetap unggul dengan akurasi 96% dan menunjukkan distribusi metrik evaluasi (precision, recall, dan F1-score) yang seimbang pada semua kelas sentimen. Sementara itu, performa DeBERTa juga membaik dengan akurasi 93% dan seimbang di semua metrik evaluasi, meskipun masih terdapat sejumlah kesalahan klasifikasi antar kategori.

Teknik Random Oversampling terbukti efektif dalam membantu model mengenali kelas minoritas, yang sebelumnya cenderung diabaikan. Sementara itu, penggunaan validasi silang memberikan evaluasi yang lebih komprehensif dan memperkuat reliabilitas hasil penelitian. Secara keseluruhan, IndoBERT tampil lebih akurat dan stabil dalam klasifikasi, sedangkan DeBERTa menawarkan keunggulan dari sisi kecepatan pelatihan. Kedua teknik penyeimbangan data dan validasi silang berperan penting dalam meningkatkan keandalan dan keadilan performa model.

Untuk pengembangan selanjutnya, disarankan untuk menambahkan tahap normalisasi kata dalam proses prapemrosesan, termasuk penanganan terhadap istilah tidak baku dan singkatan, guna meminimalkan noise dalam data. Penelitian mendatang juga bisa mengeksplorasi model transformer lain yang disesuaikan dengan bahasa Indonesia, seperti RoBERTa atau XLNet, serta membandingkan berbagai pendekatan penyeimbangan data seperti SMOTE, ADASYN, atau penggunaan weighted loss, agar hasil klasifikasi menjadi lebih optimal.

DAFTAR PUSTAKA

- [1] R. Merdiansah And A. Ali Ridha, "Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan Indobert," *Jurnal Ilmu Komputer Dan Sistem Informasi (Jikomsi)*, Vol. 7, No. 1, Pp. 221–228, 2024.
- [2] M. H. Albab, A. D. F. Sari, S. N. Asrizal, And R. Kurniawan, "Analisis Sentimen Penggunaan Kendaraan Listrik Terhadap Lingkungan Di Indonesia Dengan Pendekatan Machine Learning," *Prosiding Seminar Nasional Sains Data*, 2024.
- [3] F. Mulya, S. Putra, S. Rakasiwi, And N. Ariyanto, "Twitter Sentiment Classification Towards Telecommunication Provider Users In Indonesia," *Journal Of Applied Informatics And Computing (Jaic)*, Vol. 9, No. 2, Pp. 314–321, Apr. 2025.
- [4] J. Ye, R. Biswas, And N. Mishra, "Deberta-Based Social Media Sentiment Analysis: A Cryptocurrency Case Study," *2024 International Conference On Intelligent Computing And Emerging Communication Technologies (Icec)*, Pp. 1–5, 2024.
- [5] L. N. Wakhidah, A. K. Zyen, And B. B. Wahono, "Evaluation Of Telecommunication Customer Churn Classification With Smote Using Random Forest And Xgboost Algorithms," *Journal Of Applied Informatics And Computing*, Vol. 9, No. 1, Pp. 89–95, 2025.
- [6] M. I. Alhari, O. N. Pratiwi, And M. Lubis, "Sentiment Analysis Of The Public Perspective Electric Cars In Indonesia Using Support Vector Machine Algorithm," In *2022 International Conference Of Science And Information Technology In Smart Administration (Icsintesa)*, 2022, Pp. 155–160. Doi: 10.1109/Icsintesa56431.2022.10041604.
- [7] S. Alfarizi, D. Gunawan, H. Basri, A. R. Mulyawan, And N. Ichsan, "Optimasi Naïve Bayes Menggunakan Seleksi Fitur Forward Selection Untuk Analisis Sentimen Kendaraan Listrik," *Jurnal Teknik Komputer*, 2024.
- [8] A. Pratama And M. Rosyda, "Analisis Sentimen Dalam Aplikasi X Terhadap Pengungsi Rohingya Dengan Lstm," *Skanika: Sistem Komputer Dan Teknik Informatika*, Vol. 8, No. 1, P. 95, 2025.
- [9] M. Haris, A. Suharso, And E. H. Nurkifli, "Analisis Sentimen Pada Game Efootball Di Google Play Store Menggunakan Algoritma Indobert," *Jati (Jurnal Mahasiswa Teknik Informatika)*, Vol. 8, No. 6, Pp. 12108–12121, 2024.
- [10] A. Karimah, G. Dwilestari, And M. Mulyawan, "Analisis Sentimen Komentar Video Mobil Listrik Di Platform Youtube Dengan Metode Naive Bayes," *Jati (Jurnal Mahasiswa Teknik Informatika)*, 2024.
- [11] R. M. R. W. P. K. Atmaja And W. Yustanti, "Analisis Sentimen Customer Review Aplikasi Ruang Guru Dengan Metode Bert (Bidirectional Encoder Representations From Transformers)," *Jeisbi (Journal Of Emerging Information Systems And Business Intelligence)*, Vol. 02, No. 3, P. 2021, Jul. 2021.
- [12] N. Ayuningtyas And W. Yustanti, "Semi-Supervised Learning Pada Pelabelan Dalam Klasifikasi Multi-Label Data Teks," *Journal Of Informatics And Computer Science*, Vol. 06, No. 1, Pp. 240–248, 2024.
- [13] S. M. Fauzi, R. Ramdani, And R. Cahyana, "Analisis Sentimen Pemilu Dalam Text Mining Terhadap Hasil Real Count 2024," *Indonesian Journal Of Computer Science And Engineering*, 2024.
- [14] F. Nur Salsabilla And A. Witanti, "Analisis Sentimen Akhir Masa Jabatan Presiden Jokowi Pada Media Sosial X Menggunakan Naïve Bayes," *Skanika: Sistem Komputer Dan Teknik Informatika*, Vol. 8, No. 1, Pp. 106–115, 2025.
- [15] L. R. Aini, E. Nurfadhilah, A. Jarin, A. Santosa, And M. T. Uliniansyah, "Enhancing Sentiment Analysis Models Through Multi-Technique Data Augmentation: A Study With Indobert," *2023 International Conference On Computer, Control, Informatics And Its Applications (Ic3ina)*, Pp. 137–142, 2023, Doi: 10.1109/Ic3ina60834.2023.10285775.
- [16] I. A. Rahma And L. H. Suadaa, "Penerapan Text Augmentation Untuk Mengatasi Data Yang Tidak Seimbang Pada Klasifikasi Teks Berbahasa Indonesia," *Jurnal Teknologi Informatika Dan Ilmu Komputer*, Vol. 10, No. 6, Pp. 1329–1340, Dec. 2023, Doi: 10.25126/Jtiik.2023107325.
- [17] S. Diantika, "Penerapan Teknik Random Oversampling Untuk Mengatasi Imbalance Class Dalam Klasifikasi Website Phishing

Menggunakan Algoritma Lightgbm,” *Jati (Jurnal Mahasiswa Teknik Informatika)*, 2023.

[18] T. Ridwansyah, “Klik: Kajian Ilmiah Informatika Dan Komputer Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier,” *Klik: Kajian Ilmiah Informatika Dan Komputer*, Vol. 2, No. 5, Pp. 178–185, 2022, [Online]. Available: <Https://Djournals.Com/Klik>

[19] Anugerah Simanjuntak *Et Al.*, “Research And Analysis Of Indobert Hyperparameter Tuning In Fake News Detection,” *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, Vol. 13, No. 1, Pp. 60–67, Feb. 2024, Doi: 10.22146/Jnteti.V13i1.8532.

[20] R. Mas, R. W. Panca, K. Atmaja1, And W. Yustanti2, “Analisis Sentimen Customer Review Aplikasi Ruang Guru Dengan Metode Bert (Bidirectional Encoder Representations From Transformers),” *Journal Of Emerging Information Systems And Business Intelligence (Jeisbi)*, Vol. 02, No. 03, Pp. 55–62, 2021.

[21] R. N. Tanaja, A. Widjaya, A. Agung, S. Gunawan, K. Eka, And Setiawan, “Evaluating Public Opinion On The 2024 Indonesian Presidential Election Candidate: An Indobert Approach To Twitter Sentiment Analysis,” *2024 10th International Conference On Smart Computing And Communication (Icscc)*, Pp. 88–94, 2024, Doi: 10.1109/Icscc62041.2024.10690796.