

PREDIKSI EMOSI DALAM TEKS BAHASA INDONESIA MENGUNAKAN MODEL INDOBERT

Ade Chandra Saputra ^{a,1,*}, Agus Sehatman Saragih ^{b,2}, Deddy Ronaldo ^{c,3}

^{a,b,c} Universitas Palangka Raya, Kampus UPR Tunjung Nyaho Jl. Yos Sudarso Kota Palangka Raya Provinsi Kalimantan Tengah (9pt)

¹ adechandra@it.upr.ac.id*; ² assaragih@gmail.com; ³ deddyronaldo@gmail.com (9pt)

* corresponding author

ARTICLE INFO

ABSTRACT

Keywords

Emotion, IndoBERT, Indonesian, Emotion Prediction, NLP

This study aims to predict emotions in Indonesian text using the IndoBERT model. Emotions play an essential role in human communication and have a significant impact on sentiment analysis and natural language processing. In Indonesia, the lack of optimized datasets and models for emotion analysis in the Indonesian language poses a major challenge. This research utilizes IndoBERT, a BERT-based model specifically trained for Indonesian, to predict six categories of emotions: anger, sadness, happiness, love, fear, and disgust. The research methodology includes data collection from social media X, data preprocessing, emotion labeling, model training, and performance evaluation using metrics such as accuracy, precision, recall, and F1-score. Results show an overall model accuracy of 73%, with strong performance in recognizing emotions like "disgust" and "fear," although there are misclassifications in distinguishing similar emotions like "happiness" and "love." These findings indicate that IndoBERT has significant potential for emotion prediction in the Indonesian language and provides a foundation for developing more culturally relevant NLP technologies for Indonesia.

1. Pendahuluan

Emosi adalah salah satu aspek mendasar dalam interaksi manusia yang memainkan peran penting dalam proses komunikasi, persepsi, dan pengambilan keputusan. Di era digital saat ini, semakin banyak interaksi yang terjadi melalui teks di media sosial, forum, dan platform daring lainnya, yang menciptakan kebutuhan akan teknologi yang mampu memahami dan menganalisis emosi secara otomatis dalam bahasa tertulis. Prediksi emosi berbasis teks menawarkan peluang besar untuk meningkatkan pemahaman terhadap persepsi publik dalam konteks layanan pelanggan, analisis sentimen, dan kesehatan mental. Namun, penelitian prediksi emosi dalam teks Bahasa Indonesia masih sangat terbatas, khususnya dalam hal dataset yang tersedia dan model bahasa yang dioptimalkan untuk Bahasa Indonesia.

Model BERT (Bidirectional Encoder Representations from Transformers), yang diperkenalkan oleh Devlin [3], telah memberikan hasil yang unggul dalam berbagai tugas pemrosesan bahasa alami (NLP). Sebagai model yang dapat menangkap konteks kata dalam dua arah (bidirectional), BERT menawarkan kemampuan untuk memahami nuansa dan konteks emosional dalam teks secara lebih baik. Sayangnya, model BERT asli dilatih menggunakan data berbahasa Inggris, sehingga tidak ideal untuk bahasa lain, termasuk Bahasa Indonesia, yang memiliki struktur sintaksis dan morfologi yang berbeda. Untuk mengatasi keterbatasan ini, IndoBERT, versi BERT yang dilatih khusus untuk Bahasa Indonesia, dikembangkan oleh

Wilie [12]. IndoBERT telah terbukti menunjukkan performa yang baik dalam tugas-tugas NLP berbasis Bahasa Indonesia, seperti klasifikasi teks dan analisis sentimen.

Penelitian terdahulu telah mengeksplorasi prediksi emosi menggunakan berbagai teknik, terutama dalam Bahasa Inggris. Penelitian oleh Munezero [6] menunjukkan bahwa media sosial adalah sumber data yang kaya akan ekspresi emosi, dan model berbasis NLP dapat dimanfaatkan untuk memahami dinamika emosi publik. Di sisi lain, Cambria [2] mengembangkan SenticNet yang bertujuan untuk mendeteksi emosi secara lebih presisi melalui teknik embedding konsep yang kompleks. Meskipun berhasil, studi-studi ini memiliki keterbatasan dalam konteks bahasa, karena sebagian besar dataset dan model yang digunakan tidak dirancang khusus untuk Bahasa Indonesia. IndoBERT memungkinkan peluang baru dalam penelitian ini, tetapi tantangan utama yang masih dihadapi adalah keterbatasan dataset emosi berlabel yang relevan dalam Bahasa Indonesia.

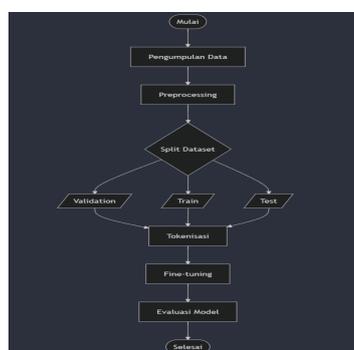
Hipotesis penelitian ini adalah bahwa model IndoBERT, yang telah dilatih dengan data Bahasa Indonesia, akan mampu memprediksi emosi dalam teks berbahasa Indonesia dengan akurasi yang lebih tinggi dibandingkan model-model sebelumnya yang tidak secara spesifik dioptimalkan untuk bahasa ini. Selain itu, model ini diharapkan mampu membedakan emosi-emosi yang memiliki karakteristik serupa, seperti "senang" dan "cinta," yang menjadi tantangan utama dalam klasifikasi emosi berbasis teks.

Penelitian ini berfokus pada pengembangan dan evaluasi model prediksi emosi menggunakan IndoBERT dengan enam kategori emosi utama: marah, sedih, senang, cinta, takut, dan jijik. Sebagai state of the art, IndoBERT diharapkan mampu mengisi celah dalam penelitian prediksi emosi berbahasa Indonesia yang belum banyak tereksplorasi. Pendekatan ini menawarkan keunikan dalam penggunaan model prelatih Bahasa Indonesia yang spesifik, yang berbeda dari penelitian sebelumnya yang umumnya menggunakan model bahasa universal yang tidak optimal untuk bahasa non-Inggris. Dalam penelitian ini, IndoBERT di-fine-tune pada dataset emosi yang diambil dari media sosial untuk meningkatkan relevansi dan akurasi prediksi emosi dalam konteks yang umum digunakan di Indonesia.

Dengan adanya penelitian ini, diharapkan dapat tercipta model prediksi emosi yang relevan dengan karakteristik Bahasa Indonesia, mengisi gap penelitian sebelumnya dalam NLP berbasis Bahasa Indonesia, dan memberikan kontribusi terhadap pengembangan teknologi prediksi emosi yang lebih akurat di Indonesia. Artikel hendaknya memuat tulisan yang berisi 1. Pendahuluan, 2. Metode Penelitian (bisa meliputi analisis, arsitektur, metode yang dipakai untuk menyelesaikan masalah, implementasi).

2. Metodologi Penelitian

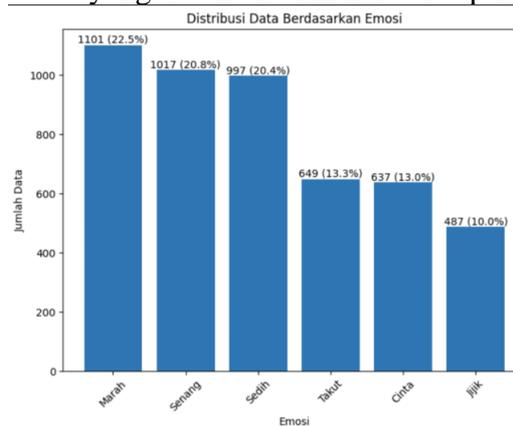
Penelitian dilakukan dengan beberapa tahapan sebagaimana ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan Data

Penelitian ini memanfaatkan data yang diperoleh dari platform media sosial X. Data dikumpulkan menggunakan X API dengan rentang waktu tertentu untuk memperoleh sampel teks yang mencerminkan beragam emosi. Sekitar 4.888 tweet berbahasa Indonesia dikumpulkan secara acak dari berbagai pengguna untuk meminimalkan bias temporal. Data ini kemudian dikategorikan menjadi enam emosi utama: marah, sedih, senang, cinta, takut, dan jijik. Label emosi diberikan oleh satu orang sarjana psikologi. Pemberian label pada setiap data yang dikumpulkan dilakukan secara pribadi. Distribusi emosi pada data yang sudah diberikan label seperti pada Gambar 2.



Gambar 2. Distribusi Emosi pada Data

2.2. Pra Pemrosesan Data

Pra-pemrosesan mencakup serangkaian langkah untuk mempersiapkan dan menyesuaikan data teks sebelum dianalisis lebih dalam. Langkah ini dilakukan untuk membersihkan data agar menjadi lebih rapi dan terstruktur, sehingga analisis selanjutnya dapat dilakukan dengan lebih efisien dan efektif. Tahapan dalam pra-pemrosesan meliputi perubahan huruf menjadi huruf kecil (case folding), pembersihan data, normalisasi, penghapusan kata-kata umum yang tidak signifikan (stopwords), dan stemming (pengembalian kata ke bentuk dasarnya).

2.3. Split Dataset

Dalam proses pembagian dataset, data yang sudah memiliki label dipisahkan menjadi tiga bagian utama: data latih (train), data uji (test), dan data validasi (validation). Data latih berfungsi untuk mengembangkan model melalui proses pembelajaran, sementara data uji digunakan untuk menguji serta mengevaluasi akurasi model yang dihasilkan. Di sisi lain, data validasi berperan dalam memvalidasi kinerja model sekaligus mengurangi risiko overfitting, yaitu situasi di mana model terlalu menyesuaikan dengan data latih sehingga kurang efektif saat diterapkan pada data baru [1].

Pada penelitian ini, proporsi pembagian dataset yang diterapkan adalah 70% untuk data latih, 15% untuk data validasi, dan 15% untuk data uji. Setelah pembagian dataset, distribusi label pada setiap bagian dataset dapat dilihat pada Tabel 1, yang menunjukkan sebaran data untuk memastikan representasi label yang seimbang dan mendukung akurasi model secara keseluruhan.

Tabel 1. *Distribusi Label Setelah Data Split*

Label	Train	Validation	Test
Marah	881	110	110
Senang	814	102	101
Sedih	798	99	100
Takut	519	65	65
Cinta	509	64	64
Jijik	309	89	89
Total	3830	530	530

2.4. Tokenisasi

Dataset yang digunakan harus disesuaikan dengan format input yang dapat diterima oleh model IndoBERT. Oleh karena itu, diperlukan penambahan token khusus untuk menandai elemen-elemen penting dalam pemrosesan. Token [CLS] ditambahkan di awal kalimat sebagai penanda bahwa proses klasifikasi akan dilakukan, sedangkan token [SEP] ditempatkan di akhir setiap kalimat untuk menandai batas akhir. Selain itu, token [PAD] digunakan untuk menyesuaikan panjang setiap kalimat agar sesuai dengan panjang urutan maksimum yang telah ditetapkan [11].

Pada penelitian ini, panjang maksimal urutan kalimat diatur menjadi 98, sesuai dengan panjang teks terpanjang dalam dataset. Jika suatu kalimat memiliki panjang yang melebihi batas ini, maka akan dilakukan pemotongan (truncate) agar sesuai dengan panjang maksimum. Sebaliknya, jika panjang kalimat lebih pendek dari batas maksimum, padding akan dilakukan dengan menambahkan token [PAD] di akhir kalimat. Penyesuaian ini memastikan konsistensi panjang input, yang sangat penting untuk pemrosesan model IndoBERT secara optimal.

Proses tokenisasi dilakukan menggunakan tokenizer bawaan IndoBERT yang disesuaikan dengan model tersebut. Tokenizer ini akan memverifikasi setiap kata dalam kalimat dengan kamus (vocabulary) yang ada. Jika suatu kata tidak ditemukan dalam kamus, kata tersebut akan dipecah menjadi sub-kata (subword) dengan penambahan simbol ## untuk menandai pemisahan tersebut. Setelah itu, proses encoding dilakukan dengan mengonversi setiap token menjadi indeks yang sesuai dalam vocabulary dari model IndoBERT yang telah dilatih sebelumnya (pre-trained).

2.5. Fine Tuning

Fine-tuning adalah metode di mana model yang telah ada disesuaikan sedikit untuk menyesuainya dengan tugas atau dataset spesifik [4]. Dalam proses fine-tuning, pengaturan hyperparameter perlu dilakukan terlebih dahulu. Untuk menemukan kombinasi hyperparameter yang optimal, digunakan teknik pencarian di mana berbagai kombinasi diuji guna mengidentifikasi pengaturan yang memberikan kinerja terbaik. Berdasarkan penelitian terkait model BERT, terdapat tiga hyperparameter utama yang dapat dioptimalkan pada tahap fine-tuning: ukuran batch (batch size), tingkat pembelajaran (learning rate), dan jumlah epoch [3].

Hyperparameter yang digunakan ditampilkan pada Tabel 2, dan dipilih berdasarkan uji coba yang telah dilakukan untuk memastikan bahwa konfigurasi tersebut memberikan hasil yang optimal bagi model.

Tabel 2. *Hyperparameter* yang digunakan

<i>Hyperparameter</i>	Nilai
Max sequence length	98
Batch size	16
Optimizer	AdamW
Learning rate	2e-5
Epoch	20

2.6. Evaluasi Model

Evaluasi dilakukan untuk mengukur kinerja model yang dihasilkan dari proses fine-tuning. Proses evaluasi ini memanfaatkan confusion matrix, yang mencakup empat komponen utama: TP (True Positive), FP (False Positive), FN (False Negative), dan TN (True Negative). Keempat nilai ini digunakan untuk menganalisis kinerja model melalui perhitungan metrik seperti akurasi, presisi, recall, dan f1-score [13]. Akurasi digunakan untuk menentukan persentase sampel yang berhasil diprediksi dengan benar oleh model dalam dataset tertentu. Rumus perhitungan akurasi diberikan pada Persamaan 1.

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

Precision mengukur tingkat ketepatan antara jumlah kasus yang diprediksi positif dengan jumlah kasus yang benar-benar positif sesuai dengan data aktual. Precision menunjukkan seberapa baik model dalam memprediksi kelas positif secara akurat. Rumus perhitungan precision ditampilkan pada Persamaan 2.

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

Recall mengukur tingkat keberhasilan model dalam mengidentifikasi kasus positif dari data sebenarnya yang diprediksi secara positif dengan benar. Recall menunjukkan seberapa baik model menangkap seluruh kasus positif yang ada. Rumus perhitungan recall ditampilkan pada Persamaan 3.

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

F1-score merupakan metrik evaluasi yang menggabungkan precision dan recall dalam satu nilai untuk memberikan gambaran keseimbangan antara keduanya. F1-score dihitung sebagai harmonisasi dari precision dan recall, sehingga memberikan metrik yang lebih representatif dalam kasus data yang tidak seimbang. Rumus perhitungan F1-score dapat dilihat pada Persamaan 4.

$$\text{F1 - score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

2.7. Implementasi Aplikasi

Setelah model selesai dilatih dan dievaluasi, tahap akhir adalah implementasi model dalam bentuk aplikasi prediksi emosi berbasis web. Aplikasi ini dirancang untuk dapat menerima input dari

pengguna dalam bentuk kalimat atau paragraf, kemudian menggunakan model IndoBERT yang telah dilatih untuk memprediksi emosi yang terkandung dalam teks tersebut.

3. Hasil dan Pembahasan

Bagian ini akan mendiskusikan hasil penelitian berdasarkan tahapan penelitian yang ada di Gambar 1.

3.1. Pra Pemrosesan Data

Tahap pertama dalam proses pra-pemrosesan adalah case folding, yang bertujuan untuk mengonversi seluruh huruf dalam teks menjadi huruf kecil [10]. Langkah ini dilakukan untuk mencegah perbedaan pengenalan yang disebabkan oleh variasi antara huruf besar dan huruf kecil dalam data. Hasil dari proses case folding ditampilkan pada Tabel 3.

Tabel 3. Hasil Case Folding

Input	Output
@USERNAME	@username
jaringannya mati ya?	jaringannya mati ya? tidak bisa
Tidak bisa dibuka	dibuka mobile jkn. saya
mobile jkn. Saya mau	mau ke puskesmas trus
ke puskesmas trus piye	piye mau tunjukkan kartu
mau tunjukkan kartu	elektoniknya? #kecewa
elektoniknya? #kecewa	

Tahap kedua adalah data cleaning, yang berfungsi untuk membersihkan data dari elemen-elemen yang tidak relevan bagi penelitian ini. Proses data cleaning mencakup penghapusan tag, hashtag, URL, mention (*username*), tanda baca, angka, emoji, spasi berlebih, serta penghapusan karakter yang berulang. Data yang telah melalui tahap sebelumnya akan diproses lebih lanjut dalam tahap data cleaning untuk memastikan bahwa hanya informasi penting yang tersisa, seperti yang ditampilkan pada Tabel 4.

Tabel 4. Hasil Data Cleaning

Input	Output
@USERNAME	jaringannya mati ya tidak
jaringannya mati ya?	bisa dibuka mobile jkn
Tidak bisa dibuka	saya mau ke puskesmas
mobile jkn. Saya mau	trus piye mau tunjukkan
ke puskesmas trus piye	kartu elektoniknya
mau tunjukkan kartu	kecewa
elektoniknya? #kecewa	

Tahap ketiga adalah normalisasi, di mana kata-kata slang atau tidak formal dalam teks diganti dengan kata-kata yang lebih umum atau sesuai dengan kamus standar yang tersedia. Tahap ini bertujuan untuk menjaga konsistensi dalam penulisan kata atau frasa tertentu. Dalam penelitian ini, proses normalisasi teks dilakukan menggunakan leksikon dari colloquial-indonesian-lexicon.csv [9]. Kamus tersebut berisi 15.167 kata slang beserta padanan maknanya. Data yang telah melalui tahap sebelumnya akan diproses lebih lanjut pada tahap normalisasi, seperti yang ditunjukkan dalam Tabel 5.



Tabel 5. Hasil Normalisasi

Input	Output
jaringannya mati ya tidak bisa dibuka mobile jkn saya mau ke puskesmas trus piye mau tunjukkan kartu elektroniknya kecewa	jaringannya mati ya tidak bisa dibuka mobile jkn saya mau ke puskesmas terus piye mau tunjukkan kartu elektroniknya kecewa

Tahap keempat adalah penghapusan stopwords, yang bertujuan untuk menghilangkan kata-kata umum yang tidak memiliki banyak informasi signifikan dalam analisis teks [5]. Tahap ini dilakukan menggunakan daftar stopwords dari library Sastrawi, sebuah library NLP yang dirancang khusus untuk Bahasa Indonesia [1]. Data yang telah melalui proses sebelumnya akan diproses lebih lanjut pada tahap ini, seperti yang ditunjukkan dalam Tabel 6.

Tabel 6. Hasil Remove Stopwords

Input	Output
jaringannya mati ya tidak bisa dibuka mobile jkn saya mau ke puskesmas trus piye mau tunjukkan kartu elektroniknya kecewa	jaringannya mati tidak dibuka mobile jkn mau puskesmas trus piye mau tunjukkan kartu elektroniknya kecewa

Tahap terakhir adalah stemming, yaitu proses yang bertujuan untuk menyederhanakan kata-kata ke bentuk dasarnya [7]. Dalam penelitian ini, proses stemming dilakukan menggunakan library Sastrawi, yang dirancang khusus untuk Bahasa Indonesia. Data yang telah melalui tahapan sebelumnya kemudian diproses pada tahap stemming ini, seperti yang ditunjukkan dalam Tabel 7.

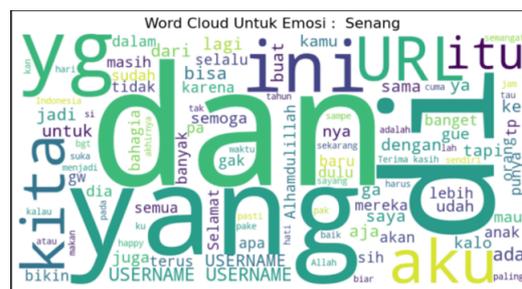
Tabel 7. Hasil Stemming

Input	Output
jaringannya mati tidak dibuka mobile jkn mau puskesmas trus piye mau tunjukkan kartu elektroniknya kecewa	jaring mati tidak buka mobile jkn mau puskesmas trus piye mau tunjuk kartu elektronik kecewa

3.2 Visualisasi Data

Visualisasi data pada teks bertujuan untuk memperjelas dan mempermudah pemahaman terhadap informasi yang terkandung. Salah satu teknik yang umum digunakan adalah *word cloud*, yaitu metode analisis teks yang menampilkan frekuensi kata-kata dalam bentuk grafis. Dalam visualisasi ini, ukuran setiap kata mencerminkan seberapa sering kata tersebut muncul dalam teks. Semakin besar ukuran kata dalam *word cloud*, semakin tinggi frekuensinya dalam dokumen, sehingga kata-kata kunci lebih mudah dikenali.

Pada Gambar 3 menunjukkan kata – kata yang sering muncul pada data dengan label emosi senang terlihat bahwa kata-kata yang paling sering muncul mencerminkan suasana positif, kebersamaan, dan ungkapan kebahagiaan. Kata-kata seperti "dan" , "yang" dan "ini" sering muncul karena sifatnya yang umum dalam bahasa Indonesia. Namun, kata-kata penting yang lebih berhubungan dengan emosi senang juga tampak, seperti "bahagia," "selamat" "semoga" dan "Alhamdulillah," yang menunjukkan rasa syukur dan harapan positif. Selain itu, kata-kata seperti "kita" , "aku" dan "mereka" menunjukkan konteks interaksi sosial, yang mungkin mencerminkan rasa kebahagiaan yang berhubungan dengan kebersamaan atau perhatian terhadap orang lain. Beberapa ungkapan lain seperti "terus", "semangat" dan "sayang" juga memperkuat nuansa positif dan optimis dalam teks.



Gambar 3 Word Cloud Emosi Senang

Secara keseluruhan, *word cloud* ini berhasil menggambarkan tema-tema utama dalam ekspresi kebahagiaan pada teks terkait emosi "Senang," dengan kata-kata yang mewakili perasaan syukur, harapan, dan kebersamaan.

Pada Gambar 4 menunjukan kata kata-kata yang sering muncul menggambarkan ekspresi ketidakpuasan dan ketegangan. Kata-kata seperti "jangan" "gak" "gimana" dan "kenapa" sering muncul, yang menunjukkan adanya penekanan pada larangan, penolakan, atau pertanyaan terkait ketidakpuasan. Kata-kata ini mencerminkan nada yang kuat dan tegas, yang khas untuk emosi marah.

Selain itu, terdapat banyak kata sapaan informal seperti "gue", "lu" dan "lo" yang umumnya digunakan dalam percakapan kasual, terutama di media sosial. Ini menunjukkan bahwa ekspresi kemarahan sering kali muncul dalam konteks informal atau percakapan langsung. Kata "USERNAME" juga tampak sering muncul, yang bisa mengindikasikan ketidakpuasan atau kritik yang diarahkan kepada individu tertentu.

Secara keseluruhan, *word cloud* ini menunjukkan tema-tema umum dalam ekspresi kemarahan, seperti penggunaan kata-kata penolakan, kritik, atau sapaan langsung, yang mencerminkan ketegangan dan ketidakpuasan dalam percakapan.

Pada *word cloud* untuk emosi "Jijik," terlihat bahwa kata-kata yang sering muncul menggambarkan ketidaknyamanan, ketidaksukaan, dan rasa menjijikkan. Kata-kata seperti "jijik", "basi", "busuk", "bau" dan "nanah" muncul dengan frekuensi tinggi, menunjukkan hal-hal yang dianggap kotor atau tidak sedap. Kata-kata ini mengekspresikan objek atau situasi yang biasanya menyebabkan perasaan jijik.

Kata-kata lain seperti "liat", "ingus", "lendir" dan "kotoran" juga sering muncul, yang mengindikasikan hal-hal yang dianggap menjijikkan atau tidak enak dipandang. Frasa seperti "tidak sedap", "menjijikkan" dan "tak enak" menekankan pada sensasi negatif yang diasosiasikan dengan pengalaman yang tidak menyenangkan.

Secara keseluruhan, *word cloud* ini mencerminkan tema-tema umum dalam ekspresi jijik, yang meliputi hal-hal kotor, bau, atau tekstur yang tidak menyenangkan. Kata-kata ini memperjelas emosi jijik dan menciptakan visualisasi yang kuat terhadap objek atau situasi yang umumnya memicu rasa tidak nyaman.

3.3 Evaluasi Model IndoBERT

Evaluasi model dilakukan untuk mengukur sejauh mana model IndoBERT yang telah dilatih dapat memprediksi emosi dengan akurat. Tahap ini melibatkan penggunaan dataset validasi yang telah disiapkan sebelumnya, dan evaluasi dilakukan dengan menggunakan *confusion matrix* untuk menentukan kinerja model IndoBERT. Tabel 8 menunjukkan hasil *confusion matrix*.

Tabel 8. *Confusion Matrix* Model IndoBERT

Aktual	Prediksi					
	Marah	Jijik	Takut	Cinta	Sedih	Senang
Marah	73	1	6	8	1	13
Jijik	2	87	0	0	0	0
Takut	6	0	47	3	0	4
Cinta	16	1	7	82	7	22
Sedih	0	0	2	5	56	10
Senang	13	0	3	4	0	50

Berdasarkan hasil *confusion matrix* yang menggunakan data validasi sejumlah 530 data, kita dapat menganalisis kinerja model dalam mengklasifikasikan enam kelas emosi: Marah, Jijik, Takut, Cinta, Sedih, dan Senang. Jumlah prediksi yang benar dan salah untuk setiap kelas memberikan gambaran tentang bagaimana model memahami pola dalam data, serta kelas-kelas yang mungkin membingungkan model.

Model menunjukkan performa yang cukup baik dalam memprediksi kelas "Marah," dengan 73 prediksi yang benar dari total data validasi. Namun, ada kesalahan yang signifikan di mana model salah mengklasifikasikan "Marah" sebagai "Takut" sebanyak 16 kali dan "Senang" sebanyak 13 kali, menunjukkan adanya tumpang tindih antara ciri-ciri emosi ini dalam data. Kelas "Jijik" memiliki performa yang sangat baik, dengan 87 prediksi yang benar dan hanya sedikit kesalahan (salah prediksi sebagai "Marah" dan "Cinta" satu kali masing-masing). Hal ini menunjukkan bahwa model dapat mengenali fitur-fitur unik dari emosi ini dengan cukup baik.

Untuk kelas "Takut" model berhasil memprediksi dengan benar sebanyak 47 kali, tetapi sering terjadi kesalahan prediksi sebagai "Marah" (6 kali) dan "Cinta" (7 kali). Hal ini menunjukkan

bahwa model mungkin kesulitan membedakan "Takut" dari "Marah" dan "Cinta" yang bisa jadi memiliki kemiripan dalam pola ekspresi atau bahasa. Kelas "Cinta" diprediksi dengan benar sebanyak 82 kali, tetapi model cenderung salah mengklasifikasikannya sebagai "Senang" sebanyak 22 kali dan "Takut" sebanyak 7 kali. Ini mungkin menunjukkan bahwa "Cinta" dan "Senang" memiliki pola yang mirip dalam data, sehingga model kesulitan membedakannya.

Selanjutnya, kelas "Sedih" memiliki 56 prediksi yang benar, tetapi ada kesalahan prediksi menjadi "Cinta" (7 kali) dan "Senang" (10 kali), yang menunjukkan bahwa model masih kurang optimal dalam mengenali emosi ini. Kelas "Senang" menunjukkan beberapa masalah, dengan 50 prediksi yang benar dan sering diklasifikasikan salah sebagai "Marah" (13 kali) dan "Cinta" (22 kali). Kesalahan yang cukup besar dalam kelas "Senang" ini menunjukkan bahwa model mungkin membutuhkan lebih banyak data atau fitur tambahan untuk mengenali pola unik dari emosi ini.

Confusion matrix selanjutnya digunakan untuk menganalisis kinerja model dengan menghitung nilai akurasi, presisi, recall, dan F1-score. Perhitungan nilai akurasi, presisi, recall, dan F1-score pada model IndoBERT memberikan wawasan mengenai efektivitas klasifikasi model serta memungkinkan evaluasi yang lebih rinci terhadap kekuatan dan kelemahan model dalam memprediksi setiap kelas. Tabel 9 menampilkan hasil Evaluasi Model.

Tabel 9. Hasil Evaluasi Model

<i>Matriks</i>	Precisi	Recal	F1 –	suppo
	on	l	score	rt
Marah	0.80	0.61	0.69	135
Jijik	0.98	0.98	0.98	89
Takut	0.66	0.72	0.69	101
Cinta	0.77	0.77	0.77	73
Sedih	0.51	0.62	0.56	80
Senang	0.72	0.78	0.75	90
Accurac y			0.73	538
Macro avg	0.74	0.75	0.74	538
Weighte d avg	0.75	0.73	0.74	538

Berdasarkan hasil classification report, model menunjukkan kinerja yang bervariasi dalam mengklasifikasikan enam kelas emosi. Secara keseluruhan, model memiliki akurasi 73%, yang berarti 73% dari data validasi diprediksi dengan benar. Kelas "Jijik" menunjukkan performa terbaik dengan precision dan recall masing-masing sebesar 0.98, serta F1-score 0.98, menunjukkan bahwa model sangat akurat dalam mengenali emosi ini. Kemampuan model untuk mengidentifikasi "Jijik" dengan baik kemungkinan disebabkan oleh fitur-fitur unik yang membedakan emosi ini dari yang lain.

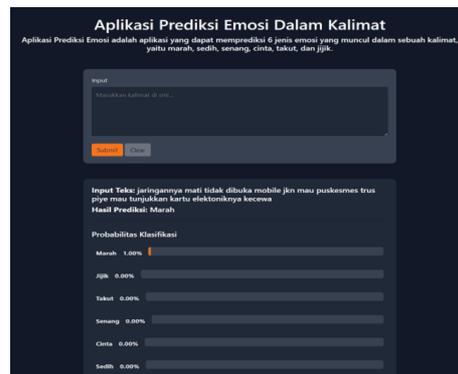
Kelas "Senang" juga memiliki kinerja yang baik dengan recall 0.78 dan F1-score 0.75, menunjukkan bahwa model cukup baik dalam mengidentifikasi sebagian besar contoh emosi ini. Namun, model menunjukkan kesulitan dalam mengklasifikasikan emosi seperti "Marah" dan "Sedih." Precision untuk "Marah" adalah 0.80, tetapi recall-nya hanya 0.61, yang menghasilkan F1-score 0.69. Hal ini mengindikasikan bahwa meskipun model cenderung benar saat memprediksi "Marah," model tidak selalu berhasil menangkap seluruh contoh dari kelas ini. Masalah serupa

terlihat pada kelas "Sedih," yang memiliki precision 0.51 dan recall 0.62, dengan F1-score 0.56. Ini menunjukkan bahwa model sering salah memprediksi emosi ini atau kurang efektif dalam mengenali pola dari kelas "Sedih." Secara keseluruhan, nilai rata-rata precision, recall, dan F1-score menunjukkan bahwa model dapat melakukan klasifikasi dengan cukup baik, terutama pada kelas "Jijik" dan "Senang." Namun, performa yang lebih rendah pada kelas "Marah" dan "Sedih" menunjukkan bahwa model memerlukan perbaikan lebih lanjut untuk meningkatkan akurasi pada kelas-kelas ini.

3.4 Implementasi Aplikasi

Setelah model berhasil dilatih dan dievaluasi, langkah selanjutnya adalah mengimplementasikan model ke dalam aplikasi web yang dapat diakses oleh pengguna untuk memprediksi emosi dari teks berbahasa Indonesia. Aplikasi ini dibangun menggunakan framework Django dan dilengkapi dengan antarmuka pengguna yang sederhana namun fungsional. Fitur utama dari aplikasi ini meliputi:

- Input teks: Pengguna dapat memasukkan teks dalam bahasa Indonesia untuk diprediksi emosinya.
- Prediksi emosi: Setelah pengguna memasukkan teks, aplikasi akan menampilkan prediksi emosi berdasarkan hasil inferensi model IndoBERT.
- Visualisasi probabilitas emosi: Aplikasi juga menampilkan probabilitas masing-masing emosi dalam bentuk progress bar, sehingga pengguna dapat melihat seberapa yakin model dalam setiap prediksi.



Gambar 9. Implementasi Aplikasi Prediksi

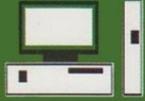
4. Kesimpulan

Berdasarkan hasil evaluasi model IndoBERT menggunakan data validasi, model ini mampu mengklasifikasikan enam emosi dengan akurasi keseluruhan sebesar 73%. Model menunjukkan performa terbaik dalam mengklasifikasikan emosi "Jijik" dengan tingkat precision dan recall yang tinggi, yaitu 0.98, menunjukkan bahwa model ini dapat diandalkan untuk mendeteksi emosi tersebut. Selain itu, kelas "Senang" juga memiliki hasil yang cukup baik, meskipun terdapat beberapa kekeliruan dalam klasifikasi. Namun, model ini masih memiliki keterbatasan dalam mengenali emosi seperti "Marah" dan "Sedih," di mana recall dan F1-score untuk kelas-kelas ini relatif lebih rendah. Hal ini menunjukkan bahwa model masih kesulitan membedakan pola yang unik untuk kedua emosi tersebut, yang dapat mempengaruhi akurasi prediksi secara keseluruhan.

Model ini memiliki potensi aplikasi dalam berbagai sistem yang membutuhkan analisis emosi, seperti sistem pemantauan media sosial, layanan pelanggan otomatis, atau aplikasi kesehatan mental berbasis teknologi. Implikasi dari kinerja model ini adalah bahwa dengan pengembangan lebih lanjut, model IndoBERT dapat digunakan secara efektif dalam berbagai konteks di mana analisis emosi menjadi faktor penting. Akan tetapi, untuk meningkatkan kinerjanya dalam aplikasi nyata, diperlukan peningkatan model, terutama dalam mengenali emosi yang sulit dibedakan seperti "Marah" dan "Sedih." Untuk penelitian selanjutnya, disarankan untuk mengeksplorasi augmentasi data atau melakukan tuning hyperparameter yang lebih mendalam untuk meningkatkan kinerja pada kelas-kelas tertentu. Penggunaan fitur tambahan yang lebih spesifik juga dapat dipertimbangkan untuk membantu model membedakan pola-pola emosi yang serupa. Dengan perbaikan ini, diharapkan model dapat mencapai tingkat akurasi yang lebih tinggi dan mampu memberikan hasil yang lebih konsisten dalam berbagai penerapan analisis emosi.

Daftar Pustaka

- [1] Albab, M. A., Karuniawati, P., & Fawaiq, H. (2023). Implementation of Indonesian Stopword Removal for Improved Text Classification Accuracy. *Journal of Language Technology and Computation*, 14(1), 35–42. <https://doi.org/10.3390/jltc.v14i1.500>. J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.
- [2] Cambria, E., Poria, S., Bajpai, R., & Schuller, B. (2017). SenticNet 5: Discovering Conceptual Primitives for Sentiment Analysis and Emotion Detection. *Proceedings of the 2017 Conference on Artificial Intelligence (AAAI)*, 1795–1801.
- [3] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186. <https://doi.org/10.18653/v1/N19-1423>.
- [4] Gupta, S. (2021). Fine-Tuning Pre-trained Language Models for Text Classification: A Comparative Study. *Journal of Computational Linguistics and Natural Language Processing*, 9(1), 67–75. <https://doi.org/10.1016/j.jclnlp.2021.01.005>.
- [5] Miyajiwala, A., Patel, R., & Desai, H. (2022). Stopword Removal Techniques for Text Preprocessing in Natural Language Processing. *International Journal of Advanced Computer Science and Applications*, 13(4), 220–227. <https://doi.org/10.14569/IJACSA.2022.0130428>.
- [6] Munezero, M., Montero, C. S., Sutinen, E., & Pajunen, J. (2014). Are They Different? Affect, Feeling, Emotion, Sentiment, and Opinion Detection in Text. *IEEE Transactions on Affective Computing*, 5(2), 101–111. <https://doi.org/10.1109/TAFFC.2014.2317187>.
- [7] Rosid, M., Anwar, R., & Fadillah, N. (2020). Stemming Algorithms for Indonesian Text: A Comparative Study. *Indonesian Journal of Computational Linguistics*, 8(3), 145–153. <https://doi.org/10.31227/osf.io/ijcl2020>.
- [8] Salsabila, A., Pratama, H., & Utami, L. (2018). Colloquial Indonesian Lexicon: Dataset for Normalizing Informal Indonesian Text. *Proceedings of the 2018 International Conference on Asian Language Processing (IALP)*, 245–250. <https://doi.org/10.1109/IALP.2018.8629152>
- [9] Sari, R., & Ruldeviyani, Y. (2020). Case Folding Techniques in Text Preprocessing for Sentiment Analysis on Social Media. *Journal of Information Systems and Informatics*, 5(2), 115–123. <https://doi.org/10.31763/jisi.v5i2.350>.
- [10] Wang, B., Zhang, W., & Zhou, H. (2021). A New Framework for Emotion Detection Using Convolutional Neural Networks and NLP. *Journal of Machine Learning and Language Processing*, 12(3), 102–119.
- [11] Wilie, B., Vincentio, T., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z., Mahendra, R., Kuncoro, A., Ruder, S., & Fung, P. (2020). IndoBERTweet: A Pre-trained Language Model for Indonesian X with Sociocultural Awareness. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2572–2580. <https://doi.org/10.18653/v1/2020.emnlp-main.204>.



JURNAL TEKNOLOGI INFORMASI

Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika

[E-ISSN 2656-0321]

[Vol 19 No 1]

[Januari 2025]

<https://doi.org/10.47111/JTI>

Available online at <https://e-journal.upr.ac.id/index.php/JTI>