

KLASIFIKASI EMOSI PADA DATA TEKS PIDATO POLITIK MENGGUNAKAN METODE *RoBERTa*

Safitri Ramadhayanti Saputro¹, Deni Arifianto², Triawan Adi Cahyanto^{*3}

^{1,2,3}Universitas Muhammadiyah Jember, Jember

Email: ¹safitriramadhayanti@gmail.com, ²deniarifianto@unmuhjember.ac.id, ³triawanac@unmuhjember.ac.id

*Penulis Korespondensi

(Naskah masuk: 30 Oktober 2025, diterima untuk diterbitkan: 20 Desember 2025)

Abstrak

Analisis emosi dalam teks merupakan salah satu cabang penting dalam *Natural Language Processing* (NLP), khususnya dalam memahami pesan tersirat pada pidato politik. Pidato politik tidak hanya menyampaikan informasi, tetapi juga emosi yang bertujuan membentuk opini publik. Penelitian ini memanfaatkan model *RoBERTa* untuk mengklasifikasikan emosi dalam pidato Presiden Joko Widodo selama periode 2014–2024. Data diperoleh dari transkrip video resmi, menghasilkan 2952 paragraf yang telah dilabeli secara otomatis menggunakan model *pre-trained* 'Indonesian-roberta-base-emotion-classifier'. Langkah praproses dilakukan melalui tahapan *cleaning*, *lowercasing*, tokenisasi, dan *one-hot encoding*. Selanjutnya, model *RoBERTa* dilakukan *fine-tuning* menggunakan *batch size* 16, *learning rate* 1e-5, dan 3 *epoch*. Evaluasi performa dilakukan dengan *confusion matrix* dan metrik akurasi, presisi, *recall*, dan *F1-score*. Hasil menunjukkan model mampu mengklasifikasikan lima emosi (*anger*, *fear*, *happy*, *love*, dan *sadness*) dengan akurasi 90%, presisi 91%, *recall* 90%, dan *F1-score* 0,91. Temuan ini menunjukkan bahwa *RoBERTa* efektif digunakan untuk klasifikasi emosi dalam teks pidato politik berbahasa Indonesia dan memberikan kontribusi terhadap pengembangan NLP dalam konteks komunikasi politik.

Kata kunci: Bahasa Indonesia, *Fine-Tuning*, Klasifikasi Emosi, *Natural Language Processing*, Pidato Politik, *RoBERTa*

EMOTION CLASSIFICATION IN POLITICAL SPEECH TEXTS USING THE *ROBERTa* MODEL

Abstract

Emotion analysis in texts is a significant branch of Natural Language Processing (NLP), particularly in understanding implicit messages in political speeches. Political speeches not only convey information but also express emotions to shape public opinion. This study utilizes the RoBERTa model to classify emotions in the speeches of President Joko Widodo during the 2014–2024 period. The dataset was obtained from official video transcripts, resulting in 2952 paragraphs labeled automatically using the pre-trained model 'Indonesian-roberta-base-emotion-classifier'. The preprocessing stages included text cleaning, lowercasing, tokenization, and one-hot encoding. The RoBERTa model was fine-tuned using a batch size of 16, a learning rate of 1e-5, and 3 epochs. Performance evaluation was conducted using a confusion matrix and metrics such as accuracy, precision, recall, and F1-score. The results show that the model can classify five emotions (anger, fear, happy, love, and sadness) with 90% accuracy, 91% precision, 90% recall, and an F1-score of 0.91. These findings demonstrate that RoBERTa is effective for emotion classification in Indonesian political speech texts and contributes to the development of NLP in political communication contexts.

Keywords: Indonesian, *Fine-Tuning*, Emotional Classification, *Natural Language Processing*, Political Speech, *RoBERTa*

1. PENDAHULUAN

Teknologi informasi dan ilmu komputer menjadi salah satu bidang yang berkembang pesat dalam beberapa dekade terakhir. Cara orang memandang dan menganalisis data terutama data berbasis teks pun ikut berubah. Tidak hanya sekadar hal yang bersifat akademis, klasifikasi emosi pada

teks pidato politik saat ini mempunyai peran penting terhadap dinamika sosial dan politik di masyarakat. Salah satu bidang yang mengalami kemajuan pesat adalah *Natural Language Processing* (NLP) adalah bidang multidisipliner yang berfokus pada upaya membuat mesin mampu memahami, menafsirkan, dan menghasilkan bahasa manusia dengan cara yang berakna dan relevan secara konteks (Kinzler, Rayhan

and Rayhan, 2023). Salah satu aplikasi utama NLP adalah analisis emosi (*emotion analysis*), yaitu proses mengidentifikasi dan mengklasifikasikan emosi dalam suatu teks.

Klasifikasi emosi adalah proses untuk mengidentifikasi dan menetapkan jenis emosi tertentu yang terkandung dalam suatu kalimat atau paragraf dan juga kemampuan untuk memprediksi emosi dari data teks secara akurat yang memungkinkan memperoleh wawasan mendalam terhadap sentimen dan membuat keputusan yang lebih tepat berdasarkan ekspresi emosional yang dianalisis (Subramani, 2023). Selain itu, klasifikasi emosi pada data teks pidato politik juga memiliki relevansi yang tinggi dalam konteks sosial dan politik. Pidato politik sering kali digunakan sebagai alat untuk memengaruhi pendengar, menyampaikan pesan, dan membangkitkan reaksi tertentu (Makuta, 2025). Dengan membaca emosi dalam pidato, politisi bisa merancang strategi komunikasi yang lebih tepat sasaran dan mengukur dampaknya ke publik.

Klasifikasi emosi dalam pidato politik penting untuk mengungkap bagaimana politisi menyusun pesan untuk membangun hubungan emosional dengan masyarakat dan memperkuat citra kepemimpinan. Emosi seperti optimisme, kemarahan atau kesedihan dalam pidato dapat membentuk persepsi publik. NLP memang sudah banyak membantu untuk analisis emosi pada teks, namun pendekatan lama yang hanya mengandalkan leksikon sering gagal untuk menangkap nuansa emosi, apalagi dalam konteks pidato politik.

Saif M. Mohammad pernah menyoroti soal ini, bahwa pendekatan berbasis kamus mempunyai batasan terutama saat harus memahami makna emosional yang sangat bergantung pada konteks dan makna tersembunyi di balik kalimat. Ia juga menekankan bahwa penggunaan leksikon harus dilakukan dengan hati-hati karena asosiasi emosi dalam leksikon umumnya didasarkan pada makna dominan kata dan persepsi mayoritas (Mohammad, 2020). Penelitian dari (Adoma, Henry and Chen, 2020) membuktikan *RoBERTa* sebagai model dengan performa terbaik untuk klasifikasi emosi berbasis teks, mengungguli BERT, XLNet, dan DistilBERT pada dataset ISEAR bahasa Inggris. *RoBERTa* mencatat akurasi 74,31% lebih tinggi dari model lain dalam mengenali emosi dari teks.

Studi tersebut menyimpulkan bahwa *RoBERTa* adalah kandidat optimal untuk klasifikasi emosi. Penelitian tentang klasifikasi emosi berbasis teks dalam bahasa Indonesia masih relatif terbatas dibandingkan dengan bahasa Inggris. Penelitian yang dilakukan oleh (Basbeth, 2024) membandingkan performa tiga model transformer populer, yaitu BERT, *RoBERTa*, dan DistilBERT. Setelah diuji di dataset bahasa Indonesia (indo4B) dengan fine-tuning dan hyperparameter tuning, *RoBERTa* meraih akurasi 90,83% dan F1-score 91%. Angka ini melampaui BERT yang bernilai F1-score sebesar

90% dan DistilBERT dengan nilai F1-score sebesar 89%. Meskipun penelitian yang dilakukan oleh (Basbeth, 2024) membuktikan keunggulan *RoBERTa* dalam klasifikasi emosi teks berbahasa Indonesia secara umum, penerapannya dalam konteks teks politik terutama pidato presiden masih jarang dilakukan.

Penelitian soal klasifikasi emosi dalam teks pidato berbahasa Indonesia, apalagi pidato kenegaraan, masih sangat terbatas. Kebanyakan studi sebelumnya lebih fokus ke teks-teks umum, seperti ulasan produk, media sosial, atau berita. Belum banyak penelitian yang menyoroti terkait wacana politik formal. Selain itu belum banyak penelitian yang secara spesifik tentang emosi dalam pidato Presiden Indonesia menggunakan pendekatan transformer seperti *RoBERTa*, meskipun model ini terbukti efektif pada data berbahasa Indonesia di domain umum. *Transformer* muncul sebagai arsitektur model yang mempercepat proses pelatihan pembelajaran mesin, terutama untuk data berurutan seperti teks. Model ini tidak lagi tergantung pada mekanisme rekursi seperti Recurrent Neural Network (RNN) dan Long Short-Term Memory (LSTM) yang cenderung melambat saat memproses data secara berurutan (Vaswani et al., 2017).

Oleh karena itu memilih pidato Presiden sebagai objek penelitian menjadi kontribusi ilmiah yang signifikan untuk NLP berbasis bahasa Indonesia. Selain itu studi tentang klasifikasi emosi pada pidato politik Presiden di Indonesia masih sedikit, terutama yang menggunakan metode *transformer* salah satunya metode *RoBERTa*. Penelitian ini memanfaatkan metode *RoBERTa* sebagai pendekatan modern untuk menjawab kebutuhan klasifikasi emosi yang kontekstual dan akurat, khususnya dalam bahasa Indonesia yang digunakan untuk mengevaluasi muatan emosional dalam pidato Presiden Joko Widodo sepanjang dua periode kepemimpinannya. Penelitian ini memiliki kontribusi unik dengan menerapkan metode *RoBERTa* untuk pertama kalinya secara terfokus pada teks pidato politik berbahasa Indonesia Presiden Indonesia.

Novelty dalam penelitian ini yaitu menggunakan metode *RoBERTa* dalam mengklasifikasi emosi pada data teks pidato politik yang belum pernah dilakukandan juga dataset yang digunakan juga berasal dari hasil transkrip video menjadi teks yang diambil dari youtube sekretariat kabinet. Banyak penelitian sebelumnya hanya menganalisis sentimen dari dataset publik atau komentar media sosial dalam Bahasa Indonesia dengan metode *machine learning*, atau mengklasifikasi emosi pada dataset bahasa Inggris memakai *deep learning*. Di penelitian ini, pendekatan *Deep Learning* menggunakan *RoBERTa* dipakai untuk mengklasifikasi emosi pada dataset unik, yaitu berdasarkan transkrip video pidato politik yang diubah menjadi teks. Hal inilah yang menjadi kebaruan dari riset ini baik dari sisi metode maupun datanya.

2. METODE PENELITIAN

Penelitian ini dilaksanakan melalui enam tahapan, yaitu: pengumpulan data, pelabelan emosi secara otomatis, praproses data, pelatihan dan penyetelan lanjutan (*fine-tuning*) model *RoBERTa*, dan evaluasi model. Berikut penjelasan detail tiap tahapannya.

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa teks pidato politik Presiden Joko Widodo pada periode 2014-2024. Sumber data adalah dari kanal Youtube Sekretariat Kabinet, khusus video rapat terbatas, total ada 273 video. Semua video tersebut kemudian ditranskripsi otomatis menjadi teks menggunakan aplikasi Sound Type AI, cukup tempel link dari video saja. Setelah semua video berubah menjadi teks, kemudian pecah setiap videonya ke dalam paragraf-paragraf dengan bantuan library Indonesian-sbert-large (Ihsan, 2023). Dari sini, terkumpul 1645 paragraf. Mengingat distribusi data yang belum seimbang untuk kebutuhan klasifikasi, dilakukan proses data augmentation dengan menghasilkan data sintesis dari kelas minoritas menggunakan metode yang dijelaskan oleh (Sánchez Caparrós, 2023). Setelah proses tersebut, jumlah total paragraf meningkat menjadi 2.952 Dataset tema rapat terbatas dipilih karena hanya Presiden Joko Widodo yang berbicara di situ. Kalau mengambil tema video lain seperti siaran pers, bisa saja ada pembicara lain, dan itu tidak sesuai dengan fokus penelitian, yaitu klasifikasi emosi pidato Presiden Jokowi. Untuk pembagian dataset menggunakan skema 80% *training*, 10% *validation* dan 10% *testing*.

2.2 Labeling Emosi Otomatis

Tahap berikutnya melakukan *labeling* emosi otomatis dengan model Indo RoBERTa Emotion Classifier (Limcorn, 2021) lewat pipeline *text-classification* dari huggingface. Model ini menghasilkan lima label emosi: *anger*, *fear*, *happy*, *love* dan *sadness*. *Labeling* otomatis ini sangat menghemat waktu untuk ribuan data teks. Namun *labeling* otomatis juga memiliki keterbatasan yang perlu dicermati, salah satunya soal potensi bias model. Bias model ini dapat muncul karena pada pidato politik ini seringkali menggunakan Bahasa formal yang bisa saja ditafsirkan berbeda oleh model yang dilatih dengan konteks percakapan umum atau teks social media. Untuk meminimalisir kesalahan *labeling* dilakukan pengecekan kembali secara manual pada seluruh data.

2.3 Praproses Data

Setelah data terlabeli, berikutnya melanjutkan ke praproses agar data siap diproses model. Ada beberapa tahap:

1. *Cleaning text* untuk membersihkan teks dari berbagai karakter khusus, tanda baca, angka yang gak relevan, simbol tidak penting, dan menghapus

spasi berlebih menggunakan *library spaCy* (Matthew et al., 2020).

2. *Lowercasing* mengubah seluruh huruf besar pada teks menjadi huruf kecil.
3. Tokenisasi untuk Langkah awal dalam proses praproses NLP yang melibatkan pemecahan teks menjadi unit-unit yang lebih kecil, biasanya berupa kata atau sub kata yang disebut token (Kinzler, Rayhan and Rayhan, 2023).

Langkah selanjutnya adalah tokenisasi yang digunakan untuk penerapan *RoBERTa*. Setelah diubah kembali menjadi *string* kemudian dilakukan teknik *one hot encoding* yaitu mengubah label emosi menjadi format vektor biner 0 dan 1 yang menampilkan ada atau tidak ada kategori dalam sampel data tersebut untuk memudahkan proses pembelajaran model (Andika Surya, Cahyanto and Muharom, 2025).

2.4 Pelatihan dan *Fine-tuning* model *RoBERTa*

Dalam proses klasifikasi emosi menggunakan *RoBERTa* diperlukan *library transformers* (Huggingface, 2019), termasuk komponen *RobertaTokenizer*, *RoBERTaForSequenceClassification*, *Trainer*, dan *TrainerArguments*. Proses *training* menggunakan *Torch* (Facebook AI Research, 2017) sebagai *framework* utama serta *AdamW* (Loshchilov and Hutter, 2019) untuk optimizer dan *get linear schedule with warmup* untuk pengaturan *learning rate*, *softmax* dari *scipy.special* digunakan untuk menghitung probabilitas dari output model. Pada tahap ini dilakukan pelatihan dan *fine-tuning* model *RoBERTa* yaitu *Roberta base* dengan spesifikasi 12 *hidden layers* yaitu lapisan dalam jaringan yang mengubah *input* melakukan *transformasi nonlinier* sehingga jaringan dapat mempelajari pola kompleks dari data (Goodfellow, Bengio and Courville, 2016). 12 model *multi head attention* adalah kombinasi dari beberapa *attention heads* yang bekerja secara parallel setiap *head* belajar menyoroti informasi yang berbeda dari *input* lalu semua hasilnya digabungkan untuk membentuk representasi akhir yang lebih kompleks (Vaswani et al., 2017), 768 model *hidden size*, 512 panjang token dan 50265 *vocab size* (Liu et al., 2019). *Fine-tuning* yaitu teknik model menggunakan model yang sudah terlatih untuk tugas tertentu dan menyatukan beberapa *layer* terakhir dari model tersebut untuk memecahkan tugas baru dan *fine-tuning* bisa meningkatkan kinerja model (Sun et al., 2019). *Fine-tuning* akan dilakukan dengan menambahkan *output layer* dan *loss function cross entropy*, dengan *hyperparameter batch size* 16, *learning rate* 1e-5, dan *epoch* 3.

2.5 Evaluasi Model

Evaluasi model dilakukan untuk menilai performa model dalam mengklasifikasikan emosi pada teks pidato. Evaluasi ini menggunakan *confusion matrix* yang berisi hasil kelas prediksi dan

hasil kelas sebenarnya yang menghitung semua klasifikasi pada setiap model (Xu, Zhang and Miao, 2020). *Confusion matrix* adalah struktur tabel berbentuk matriks persegi yang digunakan untuk mengevaluasi kinerja model klasifikasi yang barisnya menunjukkan label actual model sedangkan kolom menunjukkan prediksi model. Model *confusion matrix* yang digunakan adalah model *multiclass confusion matrix* karena memiliki lebih dari 2 kelas (Sathyanarayanan and Tantri, 2024). Pada *multiclass confusion matrix* mempunyai 4 kombinasi berbeda dari nilai prediksi yaitu *true positive*. Ketika nilai sebenarnya dan prediksi positif, *false positive* ketika prediksi bernilai positif tetapi nilai sebenarnya negatif, *true negative* ketika nilai sebenarnya dan prediksi negatif, *false negative* ketika nilai prediksi negatif tetapi nilai sebenarnya positif. Pada evaluasi model ini menggunakan *confusion matrix* dan empat metrik utama yaitu akurasi, presisi, *recall* dan, *f1-score*.

3. HASIL DAN PEMBAHASAN

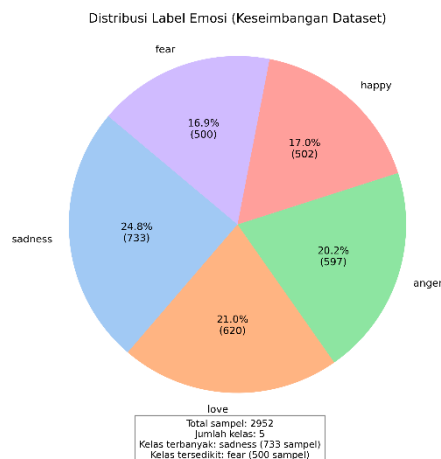
Bagian ini menyajikan hasil dari setiap tahapan penelitian, mulai dari pengumpulan data, praproses data, pelatihan dan penyetelan (*fine-tuning*) model *RoBERTa* hingga evaluasi model serta hasil analisis klasifikasi emosi. Berikut penjelasan secara lengkap.

3.1 Hasil Pengumpulan Data

Data penelitian terdiri dari 2952 paragraf teks hasil penyeimbangan (data balancing) sebagaimana dijelaskan pada subbab 2.1. Data bersumber dari transkrip video pidato politik Presiden Joko Widodo dalam forum rapat terbatas pada periode 2014-2024. Kemudian dataset dibagi menjadi tiga bagian 80% untuk *training*, 10% untuk *validation* dan 10% untuk *testing*. Distribusi jumlah data per label emosi disajikan pada Tabel 1 berikut.

Label	Training	Validation	Testing
Anger	477	60	60
Fear	400	50	50
Happy	402	50	50
Love	496	62	62
Sadness	586	73	74

Dataset akhir terdiri atas lima kategori emosi, dengan distribusi sebagai berikut: *sadness* merupakan label terbanyak dengan 733 paragraf (24.8%) sedangkan *fear* menjadi yang paling sedikit dengan 500 paragraf (16.9%). Label yang lain antara lain: *happy* (17%), *anger* (20.25%) dan *love* (21%). Distribusi ini divisualisasikan pada Gambar 1, yang memperlihatkan keseimbangan relatif antar label emosi.



Gambar 1 Keseimbangan data per label

Data hasil pengumpulan dapat diakses pada laman zenodo (Saputro, Arifianto and Cahyanto, 2025)

3.2 Hasil Praproses Data

Sebelum melakukan *training*, langkah awal yang dilakukan yaitu praproses *text*. Hasil dari praproses *text* terdapat pada tabel 2. Pada kolom paragraf asli berisi salah satu paragraf sebelum dilakukan praproses dan kolom praproses *text* sesudah dilakukan praproses. Praproses dilakukan dengan cara *cleaning text*, *lowercasing* dan tokenisasi seperti yang telah dijelaskan pada subbab 2.3. Contoh hasil praproses disajikan pada Tabel 2.

No	Paragraf asli	Praproses <i>text</i>
1	Assalamualaikum warahmatullahi wabarakatuh Rapat terbatas pada siang hari ini akan kita bahas lagi mengenai LRT di Palembang di Bandung dan di Jabodetabek Ini adalah sebuah rapat yang ke6 Artinya perlu harus ratas dan harus saya putuskan	assalamualaikum warahmatullahi wabarakatuh rapat terbatas pada siang hari ini akan kita bahas lagi mengenai lrt di palembang di bandung dan di jabodetabek ini adalah sebuah rapat yang ke6 artinya perlu harus ratas dan harus saya putuskan

Setelah praproses selesai, label emosi diubah menjadi representasi numerik melalui teknik *one-hot-encoding*, sebagaimana ditunjukkan pada Tabel 3. *One-hot-encoding* adalah teknik representasi numerik untuk data kategorikal, setiap kategori dipresentasikan sebagai vektor biner.

Label Emosi	<i>One hot encoding</i>				
	Anger	Fear	Happy	Love	Sadness
anger	1	0	0	0	0

Teknik *one-hot-encoding* digunakan agar data kategorikal seperti label emosi dapat diolah secara matematis oleh pemodelan pembelajaran mesin.

3.3 Hasil Hyperparameter dan Fine-tuning

Proses pelatihan menggunakan model RoBERTa dengan konfigurasi *hyperparameter* ditunjukkan pada Tabel 4.

Tabel 4 Arsitektur *Hyperparameter* RoBERTa

Model	Batch Size	Dropout	Att Dropout	Epoch	Learning Rate
RoBERTa	16	0.2	0.2	3	1e-5

Pemilihan *batch size* sebesar 16 didasarkan pada kestabilan pelatihan di perangkat CPU dengan kapasitas terbatas, sekaligus efisien untuk dataset yang kurang dari 10.000 data. Untuk mengatur seberapa cepat model belajar dari setiap data digunakan nilai *learning rate* dalam *fine-tuning* yang optimal berada pada *range* 1e-5 hingga 5e-5 (Devlin et al., 2018). Nilai yang terlalu tinggi dapat membuat model belajar terlalu cepat sedangkan nilai yang terlalu rendah bisa membuat proses belajar menjadi terlalu lambat dan kurang efektif. Dalam kasus ini karena dataset yang digunakan tergolong kecil jadi menggunakan *learning rate* 1e-5 untuk mencegah adanya indikasi *overfitting*. Selanjutnya *Epoch* 3 digunakan karena dengan dataset 2952 dan *batch size* 16, setiap 1 *epoch*nya ada 184 iterasi, artinya jika dilakukan sebanyak 3 *epoch* maka totalnya 552 iterasi yang berarti sudah cukup untuk mencapai konvergensi tanpa kehilangan generalisasi model. Arsitektur penyetalan lanjutan model ditampilkan pada Tabel 5.

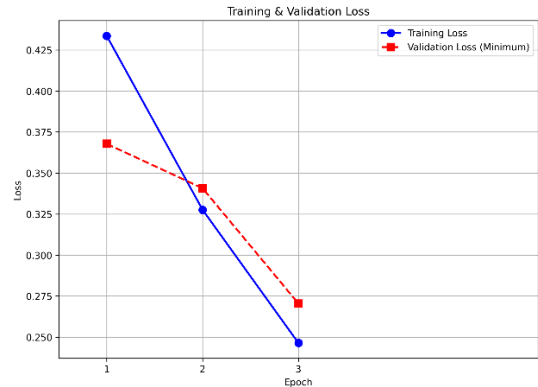
Tabel 5 Arsitektur *Fine-tuning* RoBERTa

Model	Layer
RoBERTa	<i>Input ids</i> = 256
	<i>Attention mask</i> = 256
	RoBERTa = 768
	<i>Output layer</i> = 5

3.4 Hasil Kinerja Training Hyperparameter dan Fine-tuning RoBERTa

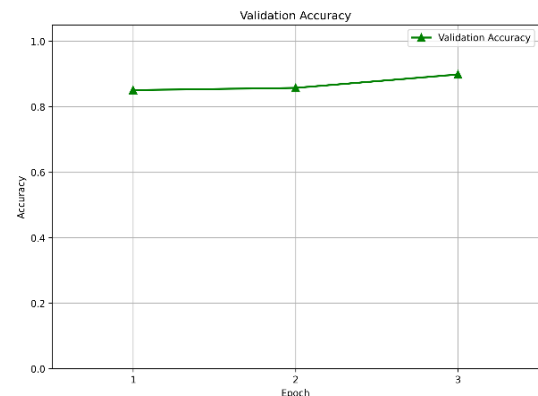
Selama proses pelatihan model menunjukkan performa yang stabil dan konvergen. Gambar 2 memperlihatkan penurunan nilai *loss* pada data pelatihan dan validasi secara konsisten di setiap *epoch*. Pada *epoch* pertama terjadi penurunan tajam yang menunjukkan peningkatan kemampuan model dalam mengenali pola. Memasuki *epoch* kedua laju peningkatan mulai melambat namun tetap konsisten, model terus membaik meskipun tidak secepat pada *epoch* pertama. Pola ini terlihat pada kedua garis yang terus menurun dengan jarak antar garis yang tetap seimbang, yang artinya proses stabil tanpa ada indikasi *overfitting*. Pada *epoch* ketiga garis validasi tetap menurun, hal ini menunjukkan bahwa model benar-benar memahami pola dasar dari data yang diberikan.

Disisi lain, Gambar 3 menampilkan perkembangan nilai *validation accuracy*. *Epoch* pertama model sudah mencapai hasil yang cukup baik, ini mengindikasikan bahwa sejak awal model sudah mampu menangkap pola dasar dari data.



Gambar 2 Training dan validation loss

Memasuki *epoch* kedua, peningkatannya tidak terlalu besar, ini menunjukkan bahwa model tetap belajar dan memperbaiki dirinya. Kemudian pada *epoch* ketiga terlihat ada kenaikan kecil dalam performa yang menunjukkan bahwa model membaik. Pada garis grafik ini menunjukkan perkembangan yang stabil dan konsisten. Tidak ada gejala penurunan performa atau ketidakseimbangan yang biasanya terjadi saat model terlalu menghafal data.



Gambar 3 Validation accuracy

3.5 Hasil Klasifikasi Emosi dan Evaluasi Kinerja Model

Hasil klasifikasi emosi pada RoBERTa dapat dilihat pada Gambar 4 (*confusion matrix*), berikut penjelasannya :

a. *Anger*

Sebanyak 43 data menunjukkan label *anger* dengan benar (*True Positive*) tetapi terdapat 17 data yang seharusnya diklasifikasikan sebagai *anger* menjadi emosi lain yaitu 3 *fear*, 5 *love* dan 9 *sadness* (*False Negative*). Sedangkan 2 data dari label lain sebagai *anger* (*False Positive*) yaitu 2 *love*, 234 data memang bukan *anger* berhasil diprediksi dengan benar (*True Negative*).

b. *Fear*

Sebanyak 43 data menunjukkan emosi *fear* dengan benar (*True Positive*), 7 data yang seharusnya *fear* tapi salah diklasifikasikan ke label lain 5 *love* dan 2 *sadness* (*False Negative*). 6 data yang bukan *fear* yang masing-masing berasal dari 3 *anger*, 1 *love* dan 2 *sadness* salah diprediksi sebagai *fear* (*False*

Positive) dan sebanyak 240 data yang bukan *fear* sebagai bukan *fear* (TN).

c. *Happy*

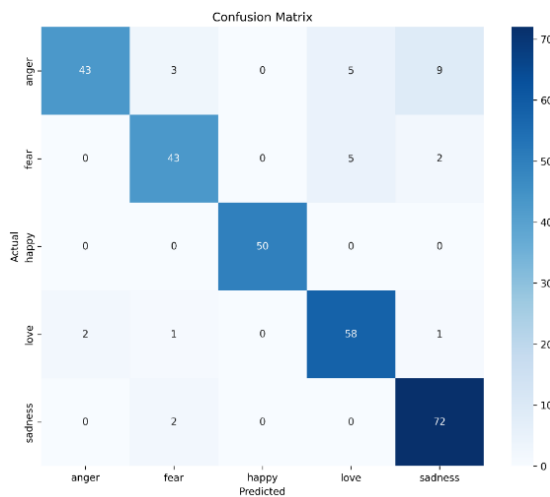
Sebanyak 50 data menunjukkan emosi *happy* dengan benar (*True Positive*), tidak ada data salah yang seharusnya diklasifikasikan ke emosi lain (*False Negative*). Tidak ada data dari label lain salah diprediksi sebagai *happy* (*False Positive*), dan sebanyak 246 data yang bukan *happy* berhasil diklasifikasi bukan *happy* (*True Negative*).

d. *Love*

Sebanyak 58 data menunjukkan label *love* dengan benar (*True Positive*), ada 4 data label *love* yang diklasifikasikan sebagai emosi lain yaitu 2 *anger*, 1 *fear*, 1 *sadness* (*False Negative*). Ada 10 data dari label lain yang salah di prediksi sebagai label *love* yaitu 5 *anger* dan 5 *fear* (*False Positive*), sebanyak 224 data yang bukan label *love* berhasil diklasifikasi dengan benar sebagai bukan label *love* (*True Negative*).

e. *Sadness*

Sebanyak 72 data menunjukan label *sadness* dengan benar (*True Positive*), ada 2 data yang sebenarnya *sadness* tetapi salah diklasifikasikan ke emosi lain 2 *fear* (*False Negative*). Terdapat 12 data dari label lain 9 *anger*, 2 *fear*, 1 *love* yang salah diklasifikasikan sebagai *sadness* (*False Positive*), dan sebanyak 210 data yang bukan *sadness* berhasil diklasifikasi dengan benar sebagai bukan *sadness* (*True Negative*).



Gambar 4 Confusion matrix

Pada Gambar 5 menyajikan *classification report* yang mencakup *precision*, *recall*, dan *f1-score* untuk masing-masing label. Label *happy* memperoleh skor sempurna di semua metrik yaitu *precision*, *recall*, dan *f1-score* sebesar 1.00, menunjukkan bahwa model sangat efektif dalam mengenali label *happy*. Label *sadness* juga menunjukkan performa sangat baik dengan *f1-score* sebesar 0.91. Rata-rata nilai makro untuk *precision*, *recall*, dan *f1-score* masing-masing mencapai 0.91, 0.90, dan 0.90 sedangkan akurasi keseluruhan model adalah 90%. Hal ini menunjukkan bahwa model bekerja cukup optimal dalam

mengenali emosi dalam teks pidato politik berbahasa Indonesia.

	precision	recall	f1-score	support
anger	0.96	0.72	0.82	60
fear	0.88	0.86	0.87	50
happy	1.00	1.00	1.00	50
love	0.85	0.94	0.89	62
sadness	0.86	0.97	0.91	74
accuracy			0.90	296
macro avg	0.91	0.90	0.90	296
weighted avg	0.90	0.90	0.90	296

Gambar 5 Classification report

Pada tabel 6 ditampilkan berbagai pasangan emosi yang sering mengalami kekeliruan dalam proses pengenalan emosi oleh model. Jumlah kesalahan yang tercatat berasal dari hasil analisis yang ditunjukkan pada gambar 4 *confusion matrix*. Kesalahan yang paling banyak terjadi antara label emosi *anger* dan *sadness*, untuk lebih jelasnya dapat dilihat pada tabel berikut.

Tabel 6 Pasangan Kesalahan Emosi

Pasangan Emosi	Kesalahan	%	Penyebab Utama
Anger↔Sadness	9 (FN)	15%	<ul style="list-style-type: none"> Model mungkin kurang sensitif terhadap nada tegas Tidak ada kata eksplisit seperti "marah"
Anger↔Love	5 (FN)	8,3%	<ul style="list-style-type: none"> Model kurang peka terhadap sarkasme. Kritik terselubung sering salah klasifikasi
Anger↔Fear	3 (FN)	5%	<ul style="list-style-type: none"> Model lebih sensitif terhadap kata-kata yang mencerminkan ancaman atau resiko Lebih dekat dengan <i>fear</i> dibandingkan dengan mengenali emosi marah yang tersirat secara halus.
Fear↔Love	5 (FN)	10%	<ul style="list-style-type: none"> Model cenderung keliru karena tidak mampu mengenali emosi implisit dalam bahasa diplomatis, teknokratik, dan solutif Terlalu mengandalkan kata kunci positif atau negative

Pasangan Emosi	Kesalahan	%	Penyebab Utama
Fear↔Sadness	2 (FN)	4%	<ul style="list-style-type: none"> Model mengalami kesulitan membedakan antara kekhawatiran dan kesedihan
Love↔Sadness	1 (FN)	1,6%	<ul style="list-style-type: none"> Model tidak dilatih cukup pada emosi <i>love</i> yang berbahasa formal jadi gagal mengenali <i>love</i> dalam konteks institusional seperti pidato politik.
Love↔Anger	2 (FN)	3,2%	<ul style="list-style-type: none"> Banyak model emosi terutama jika tidak dilatih secara kontekstual salah menafsirkan kalimat tegas sebagai marah
Sadness↔Fear	2 (FN)	2,7%	<ul style="list-style-type: none"> Model kesulitan mengenali konteks kesedihan secara implisit dan mencari alternatif lain seperti <i>fear</i> karena kalimat disusun dalam gaya naratif formal bukan emosional

3.6 Analisis Berdasarkan Karakteristik Linguistik

3.6.1 Kesalahan pada Kalimat Kompleks

Contoh paragraf : “Assalamualaikum warahmatullahi wabarakatuh. Rapat terbatas pada siang hari ini akan kita bahas lagi mengenai LRT di Palembang, di Bandung, dan di Jabodetabek. Ini adalah sebuah rapat yang ke-6. Artinya perlu harus ratas dan harus saya putuskan.”

Label asli : *Anger* (0,349)

Label prediksi : *Sadness* (0,811)

Analisis :

Paragraf ini memuat gabungan antara informasi dan keputusan tegas. Kalimat "sudah rapat ke-6" dan "harus saya putuskan" menunjukkan adanya ketegasan tinggi yang mengandung emosi marah. Karena bentuknya berupa kalimat informatif dan tidak menggunakan kata-kata kasar atau ekspresif, model tidak secara langsung mengenali nada "*anger*" dan pada bagian "bahas lagi", "rapat ke-6" yang terdengar lebih ke arah lelah atau putus asa lalu ditafsirkan sebagai *sadness*.

3.6.2 Kesalahan pada Bahasa implisit

Emosi "*anger*" pada contoh paragraf diatas tidak disampaikan melalui kata emosional secara

eksplisit tetapi melalui struktur kalimat dan konteks keputusan, seperti "ini adalah sebuah rapat yang ke-6" menunjukkan kejenuhan atau tekanan, lalu "harus saya putuskan" menyiratkan ketegasan atau tekanan. Namun karena paragraf ditulis dalam gaya formal, netral dan tanpa ekspresi emosional langsung model sulit mengenali emosi sebenarnya.

3.6.3 Kesalahan pada konteks Panjang

Dalam penelitian ini mempunyai rata-rata panjang token yang benar 61,50 token dan rata-rata panjang token yang salah 66,89 token. Perbedaan ini menunjukkan bahwa semakin panjang teks semakin besar kemungkinan model melakukan kesalahan klasifikasi. Hal ini bisa terjadi karena emosi tersebar dalam paragraf yang panjang tidak langsung muncul di awal, model kesulitan mempertahankan konteks emosional secara utuh pada paragraf yang panjang, adanya perpaduan emosi sehingga membuat model kesulitan.

4. KESIMPULAN

Penelitian ini bertujuan untuk melakukan klasifikasi emosi pada teks pidato politik berbahasa Indonesia dengan memanfaatkan model *RoBERTa*. Data penelitian terdiri atas 2952 paragraf pidato Presiden Joko Widodo yang diperoleh dari transkrip rapat terbatas periode 2014-2024. Lima kategori emosi yang digunakan meliputi *anger*, *fear*, *happy*, *love* dan *sadness*. Pelabelan awal dilakukan secara otomatis menggunakan *library* Indonesian-roberta-base-emotion-classifier yang tersedia pada platform *Hugging Face*.

Tahapan pengolahan data meliputi praproses teks, penyeimbangan distribusi data, dan konversi label ke dalam bentuk *one hot encoding*. Model *RoBERTa* dilatih dengan konfigurasi *hyperparameter* dengan *batch size* sebesar 16, *dropout* 0,2, *attention dropout* 0,1, *learning rate* 1e-5 dan jumlah *epoch* sebanyak 3.

Hasil pelatihan menunjukkan bahwa model memiliki performa stabil dan tidak mengalami gejala *overfitting*. Berdasarkan hasil evaluasi, diperoleh akurasi sebesar 90% dengan rata-rata *precision* sebesar 91%, *recall* sebesar 90% dan *F1-score* sebesar 90%. Label *happy* bahkan mendapatkan skor sempurna 1,00 pada seluruh metrik evaluasi, sementara label lainnya juga memperoleh skor yang tinggi dan konsisten.

Secara umum, hasil penelitian ini membuktikan bahwa pendekatan deep learning berbasis transformer, khususnya model *RoBERTa*, efektif mengenali emosi tersirat pada data teks pidato politik berbahasa Indonesia yang bersifat formal dan kontekstual. Temuan ini menjadi kontribusi penting dalam pengembangan kajian Natural Language Processing (NLP), khususnya yang berfokus pada analisis emosi dan wacana politik dalam bahasa Indonesia.

Selain itu, untuk bidang keilmuan, penelitian ini memberikan kontribusi terhadap NLP berbasis *transformer* khususnya *RoBERTa* dan menambah literatur tentang NLP dalam analisis teks politik terutama dalam konteks bahasa Indonesia. Penelitian ini bersifat akademik yang memberikan manfaat dalam analisis pidato politik. Metode *RoBERTa* yang digunakan dapat dapat membantu peneliti maupun lembaga pemerintah dalam mengidentifikasi pola emosi dalam teks pidato politik.

Saran untuk penelitian selanjutnya adalah melakukan validasi ulang yang dilakukan oleh ahli bahasa untuk meningkatkan akurasi dan mengurangi potensi bias. Selain itu, penelitian mendatang disarankan untuk menambahkan kategori emosi lain yang lebih beragam mencakup berbagai konteks komunikasi politik seperti debat publik, wawancara, atau pidato kebijakan, untuk meningkatkan kemampuan generalisasi model terhadap variasi bahasa dan gaya tutur. Selanjutnya perlu dilakukan eksplorasi terhadap model *transformer* lain seperti *IndoBERT*, *XLM-RoBERT* atau *DistilBERT* guna mengetahui kinerja dan arsitektur yang paling optimal terhadap karakteristik bahasa Indonesia.

DAFTAR PUSTAKA

- ADOMA, A.F., HENRY, N.M. AND CHEN, W., 2020. Comparative Analyses of Bert, Roberta, Distilbert, and Xlnet for Text-Based Emotion Recognition. *2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing, ICCWAMTIP 2020*, (November), pp.117–121. <https://doi.org/10.1109/ICCWAMTIP51612.2020.9317379>.
- ANDIKA SURYA, I.M., CAHYANTO, T.A. AND MUHAROM, L.A., 2025. Deep Learning dengan Teknik Early Stopping untuk Mendeteksi Malware pada Perangkat IoT. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 12(1), pp.21–30. <https://doi.org/10.25126/jtiik.2025128267>.
- BASBETH, F., 2024. Klasifikasi Emosi Pada Data Text Bahasa Indonesia Menggunakan. 8(April), pp.1160–1170. <https://doi.org/10.30865/mib.v8i2.7472>.
- DEVLIN, J., CHANG, M.-W., LEE, K., GOOGLE, K.T. AND LANGUAGE, A.I., 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Naacl-Hlt 2019*, (Mlm), pp.4171–4186. <https://doi.org/10.48550/arXiv.1810.04805>.
- FACEBOOK AI RESEARCH, T., 2017. *Torch*. [online] Available at: <http://torch.ch/> [Diakses 10 Mei 2025].
- GOODFELLOW, I., BENGIO, Y. AND COURVILLE, A., 2016. *Example Notation for Deep Learning*. [online] Available at: https://github.com/goodfeli/dlbook_notation [Diakses 10 Mei 2025].
- HUGGINGFACE, T., 2019. *Transformer*. [online] Available at: <https://huggingface.co/docs/transformers/> [Diakses 10 Mei 2025].
- IHSAN, N., 2023. *indonesian-sbert-large*. [online] Available at: <https://huggingface.co/naufalihsan/indonesian-sbert-large> [Diakses 10 Mei 2025].
- KINZLER, R., RAYHAN, A. AND RAYHAN, R., 2023. Natural Language Processing: Transforming How Machines Understand Human Language. (August). <https://doi.org/10.13140/RG.2.2.34900.99200>.
- LIMCORN, S., 2021. *Indo RoBERTa Emotion Classifier*. [online] Available at: <https://huggingface.co/StevenLimcorn/indonesian-roberta-base-emotion-classifier> [Diakses 10 Mei 2025].
- LIU, Y., OTT, M., GOYAL, N., DU, J., JOSHI, M., CHEN, D., LEVY, O., LEWIS, M., ZETTLEMOYER, L. AND STOYANOV, V., 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. (1). <https://doi.org/10.48550/arXiv.1907.11692>.
- LOSHCHILOV, I. AND HUTTER, F., 2019. Decoupled weight decay regularization. *7th International Conference on Learning Representations, ICLR 2019*. <https://doi.org/10.48550/arXiv.1711.05101>.
- MAKUTA, V.P., 2025. Analisis Gaya Bahasa dan Retorika dalam Pidato Joko Widodo pada Kongres 6 PAN 2024 di Channel Youtube Tribunnews. *BLAZE: Jurnal Bahasa Dan Sastra Dalam Pendidikan Linguistik dan Pengembangan*, 3(1), pp.59–65. <https://doi.org/doi.org/10.59841/blaze.v3i1.2185>.
- MATTHEW, H., MONTANI, I., VAN LANDEGHEM, S. AND ADRIANE, B., 2020. *spaCy Industrial-strength Natural Language Processing in Python*. [online] spaCy. Available at: <https://spacy.io/> [Diakses 10 Mei 2025].
- MOHAMMAD, S.M., 2020. Practical and Ethical Considerations in the Effective use of Emotion and Sentiment Lexicons. <https://doi.org/10.48550/arXiv.2011.03492>.
- SÁNCHEZ CAPARRÓS, M., 2023. *Guidelines for the use of ChatGPT and text generative AI in Justice*. [online] Available at: https://www.researchgate.net/publication/377925463_Guidelines_for_the_use_of_ChatGPT_and_text_generative_AI_in_Justice [Diakses 10 Mei 2025].
- SAPUTRO, S., ARIFANTO, D. AND CAHYANTO, T.A., 2025. *Dataset*. <https://doi.org/10.5281/zenodo.16522640>.
- SATHYANARAYANAN, S. AND TANTRI, B.R., 2024. Confusion Matrix-Based Performance Evaluation Metrics. (December). <https://doi.org/10.53555/AJBR.v27i4S.4345>.

- SUBRAMANI, K., 2023. Enhanced Text Emotion Prediction Algorithm: A Comparative Study with Support Vector Machines for Emotion Prediction in Text. *International Research Journal of Modernization in Engineering Technology and Science*, (July). <https://doi.org/10.56726/irjmets42842>.
- SUN, C., QIU, X., XU, Y. AND HUANG, X., 2019. How to Fine-Tune BERT for Text Classification? *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11856 LNAI(2), pp.194–206. https://doi.org/10.1007/978-3-030-32381-3_16.
- VASWANI, A., SHAZEER, N., PARMAR, N., USZKOREIT, J., JONES, L., GOMEZ, A.N., KAISER, Ł. AND POLOSUKHIN, I., 2017. Attention is all you need. *Advances in Neural Information Processing Systems*, p.15. <https://doi.org/10.48550/arXiv.1706.03762>.
- XU, J., ZHANG, Y. AND MIAO, D., 2020. Three-way confusion matrix for classification: A measure driven view. *Information Sciences*, 507, pp.772–794. <https://doi.org/10.1016/j.ins.2019.06.064>.

Halaman ini sengaja dikosongkan