

Perbandingan Bidirectional Encoder Representations from Transformers (BERT) Language Model pada Deteksi Emosi

Dwi Hosanna Bangkalang^{*1}, Nina Setiyawati²

¹Departemen Sistem Informasi, Universitas Kristen Satya Wacana, Indonesia

²Departemen Teknik Informatika, Universitas Kristen Satya Wacana, Indonesia

Email: ¹dwhosanna.bangkalang@uksw.edu, ²nina.setiyawati@uksw.edu

Abstrak

Informasi Tekstual menjadi salah satu cara untuk deteksi emosi. Namun, ekstraksi emosi menjadi tantangan tersendiri dikarenakan makna implisit dan eksplisit yang terkandung dalam teks. Pendekatan Ekstraksi makna emosi berbasis teks sudah banyak dilakukan dengan model deep learning. Meski begitu, performa komputasi dan akurasi model sering kali kontradiktif dikarenakan model yang kompleks. Oleh karena itu, dilakukan eksperimental model deep learning menggunakan BERT Model *Language* untuk deteksi emosi. Tujuan penelitian ini yaitu menghasilkan model deteksi emosi yang optimal dan akurat yang dapat memberikan performa komputasi yang rendah. Metode yang digunakan pada penelitian ini yaitu data collection, data *pre-processing*, *arsitektur* BERT, *BERT Model Comparison*, dan *Model Evaluation*. Model deteksi emosi terbaik ditemukan pada model DistilBERT dengan akurasi 0.9425 dan nilai F1 0.942. Berdasarkan evaluasi proses pembelajaran model DistilBERT, terdapat loss trend menurun sehingga model semakin mampu untuk melakukan prediksi emosi yang baik terhadap unseen data. Model deteksi emosi yang diusulkan, menghasilkan performa lebih unggul dibandingkan dengan model deteksi emosi menggunakan deep learning pada penelitian sebelumnya dimana model ini dapat meminimalisir kompleksitas model deteksi emosi, mengurangi komputasi namun tetap memberikan performa yang optimal.

Kata kunci: BERT, Deep Learning, Deteksi Emosi, Klasifikasi Emosi, Teks

Comparison of Bidirectional Encoder Representations From Transformers (BERT) Language Models on Emotion Detection

Abstract

Approaches to emotion detection can be done with textual information. However, emotion extraction is a challenge due to the implicit and explicit meanings contained in the text. The text-based emotion-meaning extraction approach has been widely used with deep learning models. However, computational performance and accuracy are often contradictory due to the complex model. Therefore, an experimental deep learning model was conducted using the BERT Model *Language* for emotion detection. The purpose of this study is to produce an optimal and accurate emotion detection model that can provide low computational performance. The methods used in this study are data collection, data *pre-processing*, BERT architecture, BERT Model Comparison, and Model Evaluation. The best emotion detection model was found in the DistilBERT model with an accuracy of 0.9425 and an F1 value of 0.942. Based on the evaluation of the DistilBERT model learning process, there is a decreasing loss trend so that the model is increasingly able to make good emotion predictions on occult data. The proposed emotion detection model produces superior performance compared to the emotion detection model using deep learning in previous studies where this model can minimize the complexity of the emotion detection model, reduce computation but still provide optimal performance.

Keywords: BERT, Deep Learning, Emotion Detection, Emotion Classification, Text

1. PENDAHULUAN

Deteksi emosi adalah sebuah metode untuk mengidentifikasi keadaan emosional seseorang [1] yang digunakan untuk mengenali ekspresi/opini/persepsi seseorang saat berkomunikasi dengan bantuan teknologi. Terdapat beberapa teknik untuk identifikasi emosi manusia salah satunya adalah informasi tekstual. Banyaknya sumber informasi tekstual dari sosial media [1] [2], membuka ruang analisis tidak hanya identifikasi emosi, namun

melalui data text kita juga dapat mengetahui mental state seseorang [3] [4]. Selain itu deteksi emosi juga dapat diterapkan pada business intelligence dengan mengimplementasikan model [4] sebagai salah satu strategi peningkatan marketing dan bisnis. Lebih lanjut, teknik tersebut sering digunakan di lembaga pemerintah dan analisis politik untuk pengambilan keputusan melalui analisis emosi dan opini masyarakat [5].

Namun begitu, ekstraksi emosi manusia berdasarkan teks mempunyai tantangan tersendiri dikarenakan kompleksitas emosi manusia. Kompleksitas ini disebabkan karena masing-masing individu mempunyai rangsangan emosi, perilaku emosional [6] serta nilai emosional yang berbeda berdasarkan lingkungan sosial, jenis kelamin, usia, etnis, kondisi kesehatan, psikologi state individu, dll. Tantangan lainnya, ekstraksi emosi terkhusus emosi implisit menjadi kompleks [4] dan memerlukan pemahaman semantik dan pembelajaran kontekstual yang mendalam [6].

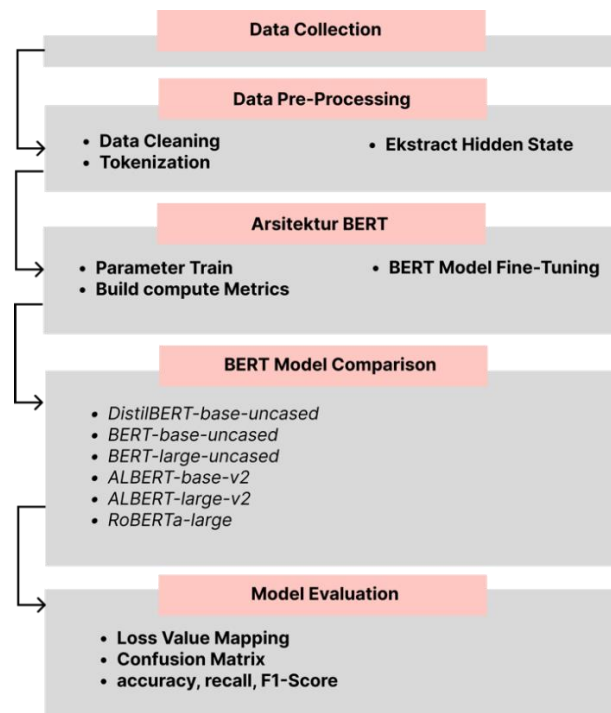
Deep learning adalah salah satu pendekatan model yang sering digunakan untuk melakukan klasifikasi text. Penelitian ini melakukan eksperimental pemodelan deteksi emosi berbasis teks menggunakan *Bidirectional Encoder Representations from Transformers* (BERT) [7], [8], [9]. BERT yang merupakan salah satu pre-trained model dengan basis *masked language* model dan *next sentence prediction* untuk merepresentasikan kata secara dua arah secara mendalam dari teks yang tidak berlabel dengan mempertimbangkan konteks kata kiri dan kanan di semua *layer* [5], [10], [11]. Arsitektur bidireksional pada BERT, memberikan kapabilitas pada BERT untuk memprediksi konteks sebelum dan sesudah pada kata secara paralel. Hal ini membuat model BERT mampu mempunyai pemahaman kontekstual mendalam pada kalimat, mempelajari hubungan semantic antar kata dengan baik sehingga dapat menemukan pola yang kompleks dalam kalimat yang dapat digunakan sesuai dengan *task* yang diberikan. Proses *pre-trained* BERT, membuat modifikasi model pada fine-tuning menjadi lebih efisien baik dari segi kebutuhan data training untuk spesifik task dan juga komputasi dikarenakan proses ini dilakukan dengan penambahan satu *output* layer tanpa melakukan modifikasi pada arsitektur dasar sehingga dapat menghasilkan model dengan komputasi dan *resource* yang rendah namun tetap memiliki performa yang baik terutama dalam implementasi dan integrasi pada aplikasi *third party* lainnya [7], [11]. Tokenisasi menggunakan BERT juga mampu mengetahui informasi eksplisit dan implisit dari data teks untuk diekstraksi berdasarkan task class yang diberikan [6].

Terdapat beberapa penelitian terdahulu yang membahas BERT. Pada penelitian yang berjudul “Fine-tune BERT based on Machine Learning Models For Sentiment Analysis” [12] mengimplementasikan BERT model language untuk melakukan ekstraksi token embedding kata. Hasil yang didapatkan pada model yang diusulkan BERT-GSVM mendapatkan akurasi tertinggi sebesar 0.91. Namun, model membutuhkan waktu eksekusi lebih lama dibanding Word2vec-GSVM yaitu 2 jam 38 menit. Penelitian ini menghasilkan model fine tune BERT sentiment analisis yang mempunyai performa lebih baik dibandingkan dengan algoritma machine learning traditional. Penelitian lainnya yang berjudul “Emotion Detection Via Bert-Based Deep Learning Approaches In Natural Language Processing” [13] membuat model deteksi emosi menggunakan model BERT dengan akurasi 0.87 pada empat emosi dasar yaitu anger, fear, joy, dan sadness. Model deteksi emosi ini menghasilkan performa yang baik. Pada class label sadness mendapatkan performa yang rendah dikarenakan *false* prediksi pada class label joy. Keterbatasan pada penelitian ini yaitu model BERT membutuhkan komputasi yang tinggi dan dataset yang besar sehingga membutuhkan pemrosesan yang lama. Selain itu, terdapat beberapa penelitian lainnya yang membahas tentang deteksi emosi menggunakan *deep learning* yang menyebutkan bahwa performa yang rendah disebabkan oleh kuantitas dan kualitas dataset. Oleh karena itu perlu dilakukan model ekstraksi fitur yang berbeda dan eksperimental pada hyperparameter untuk meningkatkan performa model [14] [15].

Berdasarkan penelitian terdahulu, model deteksi emosi menggunakan pendekatan deep learning dapat menjawab kebutuhan pengenalan emosi. Tetapi, model-model yang dihasilkan membutuhkan komputasi yang tinggi dan waktu pemrosesan yang lama. Oleh karena itu, penelitian ini akan mencoba untuk meminimalisasi kebutuhan komputasi pada model yang deteksi emosi yang juga dapat beradaptasi dengan baik dengan multiclass label, dan tetap menghasilkan model yang optimal. Penelitian sebelumnya juga belum melakukan eksplorasi perbandingan performa model versi BERT. Pada penelitian ini, model deteksi emosi yang dibangun tidak menggunakan spesifik ekstraksi *features* namun mencoba memaksimalkan potensi BERT dalam memanfaatkan informasi tersembunyi dari multi dimensi untuk memperoleh prediksi class label emosi yang optimal [8] berdasarkan proses *transfer learning* dari *pre-trained model* [5]. Sehingga, akan dilakukan perbandingan performa pada 6 model versi BERT.

Rumusan masalah pada penelitian ini yaitu bagaimana melakukan eksperimental model BERT untuk menghasilkan model deteksi emosi. Spektrum emosi yang digunakan adalah enam emosi manusia yaitu *sadness*, *happiness*, *anger*, *love*, *fear* dan *surprise*. Tujuan penelitian ini yaitu menghasilkan model deteksi emosi yang optimal dan akurat yang dapat memberikan performa komputasi yang rendah. Penelitian ini juga diharapkan dapat menghasilkan model deteksi emosi berbasis teks yang dapat beradaptasi dengan baik dengan meminimalisasi kebutuhan resource dalam aplikatif integrasi model ke aplikasi humanoid *Artificial Intelligence (AI)* [16] maupun *service third party* lainnya.

2. METODE PENELITIAN



Gambar 1. Tahapan Penelitian

Pada Gambar 1 terlihat bahwa penelitian ini dibagi dalam 4 tahapan utama, yaitu:

2.1. Data Collection

Pada tahap ini dilakukan pencarian literature review terkait penelitian yang dilakukan seperti: pemrosesan emosi manusia, klasifikasi text, arsitektur, jenis dan cara kerja BERT language model, deteksi emosi, kebutuhan ekosistem library utama hugging Face. Penelitian ini menggunakan open dataset dari kaggle.com yang telah dimodifikasi, memiliki enam label class emosi yang berjumlah 20000 data yang terlihat pada tabel 1.

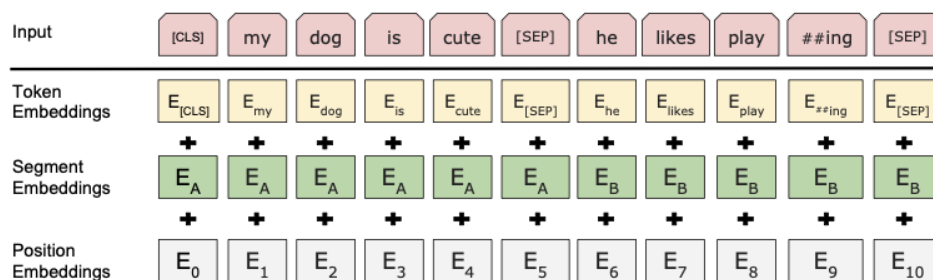
Tabel 1. Distribusi dataset pada tiap label

Label	Jumlah
anger	2159
fear	1937
happiness	5362
love	1304
sadness	4666
surprise	572

2.2. Data Pre-processing

Tahapan ini untuk mempersiapkan dan memastikan kebutuhan dataset awal memenuhi kriteria pemrosesan sebelum pre-trained BERT model. Sebelum dilakukan pre-trained deteksi emosi model menggunakan BERT, dilakukan data cleaning. Data cleaning sebelum pre-trained dapat meningkatkan representasi pembelajaran model deteksi emosi yang dibangun dengan mengurangi bias dan random noise pada dataset yang disebabkan oleh missing data, duplikasi data, spelling error, informasi tidak relevan yang tidak mempunyai arti semantik dan inkonsistensi data. Meskipun BERT mempunyai kemampuan yang sudah teruji terhadap random noise [17], data *cleaning* dapat memaksimalkan proses pre-trained BERT dalam menghasilkan model yang andal dan akurat. Dalam prosesnya, dilakukan pre-analysis untuk penentuan cleaning seperti distribusi data, perhitungan kata tiap class label, dll. Pada representasi input, BERT melakukan tokenisasi dengan algoritma WordPiece. Tokenisasi dilakukan dengan encode string pada teks menjadi numerik vector. Token pertama pada tiap urutan adalah token spesial klasifikasi [CLS]. Pada proses tokenisasi kalimat dibedakan menjadi dua bagian. Pertama, tiap kalimat dipisahkan dengan token spesial [SEP]. Kedua, ditambahkan representasi vector hasil pembelajaran pada tiap

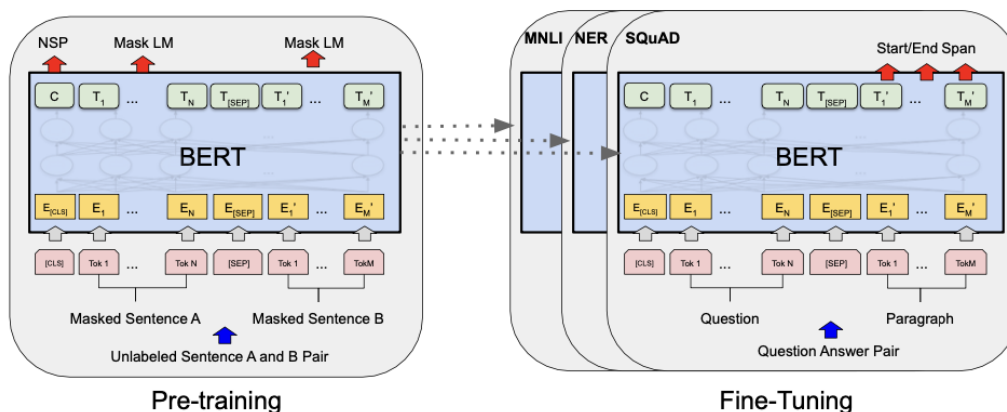
token untuk mengindikasikan korelasi tiap token ke kalimat [10]. Representasi input pada BERT memiliki *token*, *segment* dan *position embeddings* dan terlihat pada Gambar 2. Model BERT mempunyai panjang maksimum 512 token. Pada model BERT tiap urutan harus mempunyai jumlah token yang sama. Pada model yang dibangun, `max_len_word` berjumlah 74 kata sehingga tidak diperlukan proses truncated namun perlu dilakukan proses padding [PAD] agar jumlah token sama [18]. Agar model tetap dapat membedakan token yang original token dan sintetik token dilakukan proses attention mask [MASK]. Selanjutnya, tahapan terakhir sebelum pre-trained dilakukan encode dan embedded lalu dilakukan encoder data dengan output hidden state. Hidden state feature mempunyai agregat pada token [CLS] yang digunakan sebagai input pembelajaran model deteksi emosi.



Gambar 2. Representasi input data pada BERT

2.3. Arsitektur BERT

Penelitian ini menggunakan pendekatan eksperimental pada deteksi emosi model menggunakan DistilBERT. Arsitektur BERT terbagi menjadi dua bagian yaitu pre-training dan fine-tuning yang terlihat pada Gambar 3. DistilBERT merupakan model lite dari BERTbase yang menawarkan kecepatan dan komputasi ringan namun tetap mempunyai kinerja yang hampir menyamai BERTbase [9]. Perbedaan mendasar pada DistilBERT dengan BERTbase adalah pada ukuran, kecepatan dan performa. Jumlah ukuran pada jumlah layer dan parameter lebih kecil dari versi BERTbase sehingga kebutuhan resource pemrosesan lebih sedikit dan tetap memiliki downstream kurang lebih 97% sama dengan BERT. Pada tahap ini juga dilakukan eksperimental pada enam versi model BERT lainnya dan tuning hyperparameter.



Gambar 3. Arsitektur BERT

2.4. BERT Model Comparison

Pada tahap ini, dilakukan analisis perbandingan performa pada versi model BERT yaitu DistilBERT, BERTbase, BERT-large-uncased, ALBERT-base-v2, ALBERT-large-v2 dan RoBERTa-large untuk melihat perbedaan performa akurasi dari masing-masing model. Matrik perbandingan dilihat berdasarkan loss, akurasi, runtime dan steps.

2.5. Model Evaluation

Pada tahap ini dilakukan analisis dan evaluasi pada model BERT yang diusulkan. Selain melihat akurasi model, dilakukan evaluasi F1-Score. F1-score adalah matrik evaluasi yang mengevaluasi kinerja model dengan menyeimbangkan rata-rata harmonik presisi dan recall[19]. Berbeda dengan akurasi yang mengevaluasi keseluruhan dataset, F1-Score memberikan evaluasi model berdasarkan kemampuan prediksi model pada tiap class. Evaluasi confusion matrix juga dilakukan untuk melihat performa prediksi dari tiap class label. Selain itu, dilakukan analisis melihat steps logging dan loss value mapping untuk memantau performa, proses pembelajaran dan kinerja model pada saat fine-tuning.

3. HASIL DAN PEMBAHASAN

3.1. Hasil

Pada penelitian ini, dilakukan beberapa eksperimental versi model BERT dan hyperparameter fine tuning model. Eksperimental berfokus pada empat training parameter yaitu batch size, epoch, learning rate dan weight decay. Dilakukan eksperimen batch size 32 dan 64, learning rate 2e-5, 3e-5, weight decay 0,01, 0,05 dan 0,1. Berdasarkan hasil eksperimen, didapatkan final Parameter final seperti yang terlihat pada Tabel 2. Nilai parameter training tersebut memberikan hasil akurasi yang baik pada ketiga versi model BERT yang terlihat pada Tabel 3.

Tabel 2. Hyperparamater final

Training Parameter	Value
per_device_train_batch_size	32
num_train_epochs	2
learning_rate	3e-5
weight_decay	0.01

Tabel 3. Perbandingan Akurasi Bert

Bert Model	Loss	Acc	F1-Score	Runtime (m)	Samples (s)	Steps (s)
bert-base-uncased	0.1428892165	0.938	0.9380544839	1.9198	1041.801	32.817
bert-large-uncased	0.1209011078	0.9345	0.9343371548	6.199	322.635	10.163
distilbert-base-uncased	0.1362385005	0.943	0.9429628426	1.0671	1874.245	59.039
albert-base-v2	1.580403566	0.352	0.1832899408	2.1349	936.798	29.509
albert-large-v2	1.581650972	0.352	0.1832899408	6.8689	291.168	9.172
roberta-large	1.581780314	0.352	0.1832899408	6.2281	321.127	10.116

Tabel 3 merupakan perbandingan hasil akurasi model training. Hasil yang didapatkan yaitu DistilBERT mendapatkan akurasi tertinggi yaitu 0.943 dengan runtime yang cukup singkat yaitu 1.0671 menit menggunakan GPU. Adapun performa model dengan pemrosesan CPU mendapatkan runtime 480.01 menit. Hal ini menunjukkan bahwa komputasi menggunakan GPU jauh lebih efisien dibandingkan menggunakan CPU.

BERT model versi ALBERT base/large dan RoBERTa large menghasilkan akurasi yang rendah. Selain nilai parameter yang tidak fit, hal ini juga dipengaruhi proses tokenisasi menggunakan model DistilBERT sehingga proses embedding saat training model tidak optimal. Penyebab lain juga dikarenakan masing-masing versi model BERT mempunyai proses penanganan format input yang berbeda. Meski begitu, perbedaan tokenizer pada model BERT base dan BERT large, tidak berpengaruh signifikan karena menghasilkan akurasi cukup tinggi 0.938 dan 0.934.

Pada basis model BERT large, menghasilkan runtime yang lebih lama dikarenakan ukuran model, perbedaan jumlah lapisan transformer, ukuran hidden layer dan attention heads yang lebih besar dari DistilBERT. Sehingga, model DistilBERT memberikan performa kecepatan proses lebih unggul[20], [21] dibandingkan dengan model BERT base atau large lainnya. Berdasarkan perbandingan tersebut, secara keseluruhan didapatkan versi model DistilBERT memberikan model yang paling optimal.

Evaluasi pada final model deteksi emosi berbasis BERT dilakukan dengan melakukan analisis pada training loss dengan melihat logging step. Analisis Loss value mapping pada tabel 4 memperlihatkan semakin kecil nilai loss, model semakin mampu untuk melakukan prediksi emosi yang baik terhadap unseen data. Terjadi false prediksi pada true label happiness namun prediksi model sadness dengan nilai Loss 6.65. Hal ini terjadi pada step awal antara 0-100. Pada proses pembelajaran model, ditemukan tren penurunan nilai loss pada step 100-800 yang berarti model fit dengan batch data training. Namun, step 800-1000 terjadi sedikit kenaikan 0.12 ke 0.14 dan penurunan kembali ke 0.13. Detail Logging step dapat dilihat pada Tabel 5.

Tabel 4. Loss Value Mapping

Text	Label	Predicted Label	Loss	Mapping
i feel like im alone in missing him and because of that	sadness	sadness	0.002845	True
i feel a bit foolish for missing him as much as i do	sadness	sadness	0.002907	true
i gotta tell you for a while i been feeling gloomed and doomed and some ugly grey clouds been hanging round me	sadness	sadness	0.002925	True
i must say that i feel a little depressed because everything i know could be completely meaningless	happiness	sadness	6.65509	false
i called myself pro life and voted for perry without knowing this information i would feel betrayed but moreover i would feel that i had betrayed god by supporting a man who mandated a barely year old vaccine for little girls putting them in danger to financially support people close to him	happiness	fear	5.940544	false
im lazy my characters fall into categories of smug and or blas people and their foils people who feel inconvenienced by smug and or blas people	anger	happiness	4.942703	false
i had applied for a job and they had assured me that the exams would take place a few months later a week later i went to obtain some more information and they told me that the exams had already taken place	love	sadness	4.560087	false
i feel that he was being overshadowed by the supporting characters				

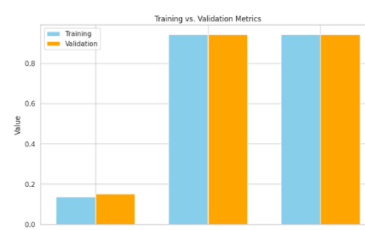
Tabel 5. Logging Step Loss

Step	Training Loss
100	1.1638
200	0.5098
300	0.3054
400	0.2628
500	0.2221
600	0.154
700	0.1378
800	0.1265
900	0.1448
1000	0.1302

Pada Tabel 6 terlihat bahwa metrik loss pada validation masih lebih besar dibandingkan nilai loss pada data training, meskipun selisih tidak signifikan. Demikian halnya nilai F1 dan akurasi pada data training lebih tinggi dibandingkan pada data validasi. Berdasarkan, hasil evaluasi dapat diambil kesimpulan model deteksi emosi yang dibangun tidak secara signifikan mengalami overfitting dan model mampu melakukan deteksi emosi yang baik pada unseen data. Representasi metrik evaluasi terlihat pada Gambar 4.

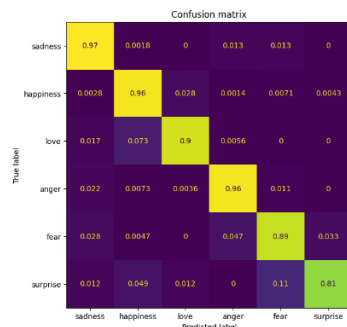
Tabel 6. Evaluation Model

	loss	accuracy	F1-Score
Training	0.1362385005	0.943	0.9429628426
Validation	0.1519172788	0.9425	0.9423331928



Gambar 4. Training dan Validation Matrix

Model deteksi emosi yang dibangun berbasis DistilBERT Language model ini mempunyai 6 class label emosi. Berdasarkan confusion matrix pada Gambar 5, terlihat label sadness menghasilkan akurasi tertinggi sebesar 0.97. Class label happiness dan joy sering tertukar. Hasil prediksi label emosi happiness 0.96 dimana kesalahan prediksi model happiness terbesar pada label joy. Begitu pula sebaliknya, class label joy mempunyai akurasi 0.9 dan model melakukan kesalahan prediksi ke label happiness. Hal yang sama juga terjadi pada class label fear dan surprise. Dengan akurasi label fear pada 0.89 dan emosi surprise 0.81. Hasil evaluasi recall dan F1 pada label emosi sadness mendapatkan nilai tertinggi sebesar 0.97 dan evaluasi presisi tertinggi pada label emosi happiness sebesar 0.96. Hasil recall, presisi, F1 terendah ada pada label emosi surprise yaitu 0.81, 0.86, 0.84. Matrik evaluasi terlihat pada Tabel 7.



Gambar 5. Confusion Matrix

Tabel 7. Classification Report

Emotion	precision	recall	f1-score	support
Sadness	0.9674502712	0.9727272727	0.9700815956	550
Happiness	0.969740634	0.9559659091	0.9628040057	704
Love	0.8797814208	0.904494382	0.891966759	178
Anger	0.9326241135	0.9563636364	0.9443447038	275
Fear	0.8867924528	0.8867924528	0.8867924528	212
Surprise	0.8684210526	0.8148148148	0.8407643312	81
accuracy			0.943	
macro avg	0.9174683242	0.915193078	0.9161256414	2000
weighted avg	0.9431049424	0.943	0.9429628426	2000

3.2 Diskusi

Tabel 7. Perbandingan Evaluasi F1 pada Model Deteksi Emosi

No.	Judul Penelitian	Jumlah Dataset	Spectrum emosi	Model	F1
1	Emotion Detection Via Bert-Based Deep Learning Approaches In Natural Language Processing” [13]	2518	anger, fear, joy, dan sadness	BERT	0.8685
2	Emotion Detection in Textual Data using Deep Learning[14]	7666	Joy, fear, anger, sadness, guilt, shame, disgust	LSTM BiLSTM GRU	0.5705 0.59 0.59
3	Emotion Detection of Contextual Text using Deep learning[15]	2226	Happy, sad, angry	Word2Vec+Bi-Long Short-term Memory FastText+Bi-Long Short-term Memory Glove+Bi-Long Short-term Memory	66.14 67.08 69.63
4	Penelitian ini	20000	sadness, happiness, anger, love, fear dan surprise.	DistilBERT	0.9429628426

Tabel 7 merupakan tabel perbandingan evaluasi penelitian ini dengan penelitian sebelumnya. Penelitian 2 dan 3 menerapkan model neural network dengan memaksimalkan kombinasi model dengan metode ekstraksi fitur dengan maksimal performa 67.08. Penelitian 1 menggunakan model BERT namun menggunakan jumlah dataset lebih sedikit yaitu 2518 dengan empat spektrum emosi, sedangkan model yang diusulkan menggunakan 20000 dataset dan enam spektrum emosi. Model deteksi emosi menggunakan DistilBERT yang diusulkan pada penelitian ini secara keseluruhan mendapatkan performa class lebih baik yaitu 0.94 dibandingkan dengan tiga penelitian terdahulu yang menerapkan deep learning model.

4. KESIMPULAN

Pemodelan deteksi emosi dilakukan dengan fine-tuning pada versi BERT model Language. Berdasarkan perbandingan dengan beberapa versi model BERT, model terbaik ditemukan pada model DistilBERT dengan rata-rata F1 sebesar 0.942. Berdasarkan hasil eksperimental dengan model BERT lainnya, model deteksi emosi berbasis DistilBERT dapat mengurangi kebutuhan komputasi yang terlihat dari proses runtime yang lebih singkat dan menghasilkan model yang *less* kompleks dengan dataset kecil namun tetap mempertahankan performa yang optimal. Model DistilBERT mempunyai akurasi tertinggi (0.94) dengan nilai F1 dan Recall tertinggi ada pada label emosi sadness (0.97) dan nilai presisi tertinggi ada pada label emosi happiness (0.97). Label emosi yang mempunyai Presisi, Recall, dan F1 terkecil adalah surprise (0.87, 0.82, dan 0.84). Hasil perbandingan akurasi data training dan data validasi pada model DistilBERT tidak secara signifikan menunjukkan overfitting serta mampu melakukan generalisasi pada unseen data. Hal ini terlihat dari proses pembelajaran model deteksi emosi berbasis DistilBERT yang menunjukkan loss trend menurun. Saran untuk penelitian selanjutnya adalah perlu menambah signifikansi ketahanan model pada overfitting dengan cara eksplorasi pada augmentasi data dan tuning hyperparameter lainnya seperti optimizer, warmup steps, dll.

DAFTAR PUSTAKA

- [1] A. Deo, D. Thakore, R. Mistry, V. Iyer, and P. Tawde, "A Review of Emotion Detection Systems," in *2022 5th International Conference on Contemporary Computing and Informatics (IC3I)*, Uttar Pradesh, India: IEEE, 2022, pp. 32–37. doi: 10.1109/IC3I56241.2022.10073423.
- [2] P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," Dec. 01, 2021, *Springer*. doi: 10.1007/s13278-021-00776-6.
- [3] M. Singh, A. K. Jakhar, and S. Pandey, "Sentiment analysis on the impact of coronavirus in social life using the BERT model," *Soc Netw Anal Min*, vol. 11, no. 1, Dec. 2021, doi: 10.1007/s13278-021-00737-z.
- [4] M. Alrasheedy, R. Muniyandi, and F. Fauzi, "Text-Based Emotion Detection and Applications: A Literature Review," in *International Conference on Cyber Resilience (ICCR)*, Dubai, United Arab Emirates, Dubai, United Arab Emirates: IEEE, 2022, pp. 1–9. doi: 10.1109/ICCR56254.2022.9995902.
- [5] S. Kusal, S. Patil, · Jyoti Choudrie, · Ketan Kotecha, · Deepali Vora, and I. Pappas, "A Review on Text-Based Emotion Detection-Techniques, Applications, Datasets, and Future Directions."
- [6] N. Alswardan and M. E. B. Menai, "A survey of state-of-the-art approaches for emotion recognition in text," *Knowl Inf Syst*, vol. 62, no. 8, pp. 2937–2987, Aug. 2020, doi: 10.1007/s10115-020-01449-0.
- [7] M. Munikar, S. Shakya, and A. Shrestha, "Fine-grained Sentiment Classification using BERT," 2019. Accessed: May 21, 2025. [Online]. Available: <https://ieeexplore.ieee.org/document/8947435/>
- [8] A. R. Abas, I. Elhenawy, M. Zidan, and M. Othman, "Bert-cnn: A deep learning model for detecting emotions from text," *Computers, Materials and Continua*, vol. 71, no. 2, pp. 2943–2961, 2022, doi: 10.32604/cmc.2022.021671.
- [9] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," Oct. 2019, [Online]. Available: <http://arxiv.org/abs/1910.01108>
- [10] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," 2019. [Online]. Available: <https://github.com/tensorflow/tensor2tensor>
- [11] Koroteev MV, "BERT: A Review of Applications in Natural Language Processing and Understanding," 2021. Accessed: May 21, 2025. [Online]. Available: <https://arxiv.org/abs/2103.11943>
- [12] N. Smairi, H. Abadlia, H. Brahim, and W. L. Chaari, "Fine-tune BERT based on Machine Learning Models For Sentiment Analysis," in *Procedia Computer Science*, Elsevier B.V., 2024, pp. 2390–2399. doi: 10.1016/j.procs.2024.09.531.

-
- [13] D. Zulfikar ASLAN, "Emotion Detection Via Bert-Based Deep Learning Approaches In Natural Language Processing." Accessed: May 21, 2025. [Online]. Available: <https://dergipark.org.tr/en/download/article-file/4105316>
- [14] D. Yohanes, J. S. Putra, K. Filbert, K. M. Suryaningrum, and H. A. Saputri, "Emotion Detection in Textual Data using Deep Learning," in *Procedia Computer Science*, Elsevier B.V., 2023, pp. 464–473. doi: 10.1016/j.procs.2023.10.547.
- [15] U. Rashid, M. W. Iqbal, M. A. Skiandar, M. Q. Raiz, M. R. Naqvi, and S. K. Shahzad, "Emotion Detection of Contextual Text using Deep learning," in *4th International Symposium on Multidisciplinary Studies and Innovative Technologies, ISMSIT 2020 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., Oct. 2020. doi: 10.1109/ISMSIT50672.2020.9255279.
- [16] F. Mohammad *et al.*, "Text Augmentation-Based Model for Emotion Recognition Using Transformers," *Computers, Materials and Continua*, vol. 76, no. 3, pp. 3523–3547, 2023, doi: 10.32604/cmc.2023.040202.
- [17] D. Zhu, M. A. Hedderich, F. Zhai, D. Ifeoluwa Adelani, and D. Klakow, "Is BERT Robust to Label Noise? A Study on Learning with Noisy Labels in Text Classification," 2022.
- [18] I.-A. Albu and S. Spînu, "Emotion Detection From Tweets Using A Bert And Svm Ensemble Model," *U.P.B. Sci. Bull., Series C*, vol. 84, no. 1, 2022, [Online]. Available: <https://pypi.org/project/demoji>
- [19] S. R. Kalmegh, M. Bhushan, and R. Padar, "Empirical Study on Evaluation Metrics for Classification Algorithms," *International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)*, vol. 3, no. 2, 2023, doi: 10.48175/568.
- [20] A. Saputra, A. Saragih, and D. Ronaldo, "Perbandingan Nilai Akurasi Distilbert Dan BERT Pada Dataset Analisis Sentimen Lembaga Kursus," *Jurnal Teknologi Informasi*, 2024, doi: <https://doi.org/10.47111/JTI>.
- [21] F. Fajri *et al.*, "Membandingkan Nilai Akurasi BERT dan DistilBERT pada Dataset Twitter," *Jurnal Sistem Informasi*, vol. 8, no. 2, pp. 71–80, 2022.