PENGEMBANGAN METODE SENTIMEN ANALISIS BAHASA INDONESIA MENGGUNAKAN DISTILBERT DENGAN *FUZZY* SOM

Haleluya Noka [1], Hendry [2]
[1,2] Teknik Informatika, Universitas Kristen Satya Wacana
672022069@student.uksw.edu

ABSTRAK

Seiring lonjakan data diera digitalisasi, efisiensi komputasi menjadi aspek krusial dalam penerapan analisis sentimen Bahasa Indonesia. Hal ini didasari oleh kendala pada model bahasa standar yang cenderung berat dan lambat, sehingga sulit diterapkan pada skenario pemrosesan data skala besar secara *realtime* atau pada perangkat dengan spesifikasi terbatas. Penelitian ini bertujuan mengembangkan metode analisis sentimen teks Bahasa Indonesia dengan menyederhanakan arsitektur *deep learning* IndoBERT menjadi DistilBERT yang lebih ringan secara komputasi. Dataset Indo4B digunakan melalui tahapan *preprocessing*, pelabelan, dan *fine-tuning* dengan model DistilBERT. Selanjutnya, optimalisasi hasil klasifikasi model DistilBERT *fine-tuned* dilakukan menggunakan algoritma *Trapezoidal Fuzzy Membership* dan *Self Organizing Maps* (SOM) berdasarkan nilai *confidence* dan distribusi representasi *node*, agar performanya dapat menyamai IndoBERT. Hasil evaluasi menunjukkan peningkatan akurasi dari 79.85% menjadi 80.42%, serta peningkatan pada *Precision*, *Recall*, dan *F1-Score*. Hal ini membuktikan bahwa kombinasi DistilBERT dengan *Fuzzy* dan SOM menghasilkan klasifikasi sentimen multilevel yang lebih efisien dan akurat untuk teks Bahasa Indonesia.

Kata kunci : *Sentimen Analisis, DistilBERT, IndoBERT, Fuzzy, SOM, Bahasa Indonesia*

1. PENDAHULUAN

Bahasa Indonesia merupakan bahasa keempat yang paling sering digunakan di internet, dengan 171 juta pengguna. Namun, perkembangan penelitian mengenai Bahasa Indonesia dalam pemrosesan bahasa alami atau *Natural Language Processing* (NLP) berjalan lambat karena kurangnya sumber daya yang tersedia. Untuk mengatasi hal ini diperkenalkanlah model berbasis BERT untuk Bahasa Indonesia, yaitu IndoBERT. Model ini dilatih dengan dataset Indo4B, yang berisi sekitar 4 miliar kata dengan Bahasa Indonesia yang telah melalui *preprocess* sebesar 23GB. Dataset ini dikumpulkan dari berbagai sumber seperti berita, media sosial, *wikipedia*, artikel, subtitel video, dan kumpulan dataset paralel Bahasa Indonesia lainnya [7].

IndoBERT menggunakan arsitektur dasar BERT dalam memahami konteks dua arah atau *bidirectional* namun menuntut komputasi dan memori yang besar [1][28]. Oleh karena itu muncullah penelitian menggunakan DistilBERT, sebuah model hasil *knowledge distillation* dengan arsitektur lebih kecil, ringan, dan cepat. Penelitian mengenai DistilBERT menunjukkan bahwa model ini memiliki *transformer layer* 40% lebih kecil dan dapat mencapai kinerja serupa, sekaligus 60% lebih cepat dibandingkan model BERT aslinya [6][28].

Penelitian ini bertujuan memanfaatkan model DistilBERT sebagai alternatif untuk meningkatkan efektivitas kinerja model IndoBERT, yang memiliki dasar model BERT dalam arsitekturnya. Namun model DistilBERT masih memiliki kekurangan, dimana penelitiannya menunjukkan model ini hanya dapat mempertahankan 97% kinerja pemahaman teks dari model aslinya [6]. Model BERT asli memiliki 12 lapisan *transformer layer* dengan total 110 juta parameter. Hal ini menyebabkan model memiliki tingkat pemahaman teks yang tinggi dalam dataset namun membutuhkan komputasi yang besar [1]. Sebagai solusinya, penelitian ini mengusulkan penggunaan *Fuzzy* dan *Self Organizing Maps* (SOM) untuk mengoptimalisasi model DistilBERT agar dapat mendekati performa model IndoBERT [6][7]. Optimalisasi dilakukan dengan logika *Fuzzy* untuk menangani ketidakpastian bobot pada skor *confidence* hasil model [2][23]. Hasil tersebut kemudian diproses dalam SOM untuk memetakan data berdimensi tinggi ke dua dimensi, guna menemukan pola *cluster* yang berisiko pada hasil *confidence* klasifikasi sentimen [24]. Dengan demikian, hasil DistilBERT dapat dioptimalkan dan digunakan sebagai salah satu metode pengembangan model sentimen analisis untuk dataset Bahasa Indonesia.

2. TINJAUAN PUSTAKA

Sentimen analisis adalah penelitian yang meneliti bahasa yang dituliskan oleh individu untuk menentukan suatu pendapat, perasaan, sikap, penilaian, emosi, subjektivitas, atau perspektif mereka [15]. Cara melakukan penelitian untuk analisa sentimen dapat dilakukan dengan menggunakan model *machine learning* atau *deep learning*. Model terkini yang banyak digunakan dalam penelitian adalah model *deep learning* dikarenakan kemudahannya untuk diaplikasikan ke dalam teks untuk menemukan hasil sentimennya.

Beberapa model *deep learning* dengan fokus teks Bahasa Indonesia telah banyak dibahas, terutama penelitian mengenai pembuatan model

deep learning terbaru berbasis BERT yang difokuskan untuk teks Bahasa Indonesia, yaitu model IndoBERT [7]. Tantangan terbaru dalam sentimen analisis adalah bagaimana membuat model *deep learning* dari hasil penelitian sebelumnya, yaitu IndoBERT menjadi model dengan komputasi yang tidak terlalu besar [17]. Penelitian sebelumnya mengenai model IndoBERT dapat dimengerti bahwa model ini memiliki dasar dari model BERT di dalam arsitekturnya [7]. Model BERT ini memiliki 12 *transformer layer* dalam arsitektur modelnya untuk melakukan analisis pada suatu sentimen [1].

Oleh karena itu, penelitian ini bertujuan untuk mengatasi tantangan tersebut. Dengan cara adalah menggunakan model *deep learning* yang lebih ringan dari BERT, yaitu DistilBERT [6]. Model DistilBERT hanya memerlukan 6 *transformer layer* dalam arsitekturnya untuk mengolah suatu sentimen [20]. Hasil analisis sentimen dengan model DistilBERT kemudian dioptimalkan kembali dengan mengatur nilai *confidence* dari fungsi *Fuzzy* dan SOM agar dapat mengoptimalkan nilai akurasi klasifikasi sentimen dalam teks Bahasa Indonesia [23][26].

Tabel 1. Perbandingan penelitian sebelumnya

Judul Jurnal	Model	Tujuan	Gap
Bello et al., 2023 [1]	BERT	Mengoptimalkan analisis sentimen tweet berbasis pemahaman konteks dengan model BERT.	Arsitektur besar (12 *transformer layer*) butuh komputasi yang tinggi.
Gandhi et al., 2025 [6]	DistilBERT	Mewujudkan komputasi efisien dan akurat untuk aplikasi *realtime* pada perangkat terbatas dengan menggunakan model DistilBERT.	Komputasi berkurang 40%, namun performa sedikit menurun (97%) dibanding BERT asli.
Geni et al., 2023 [7]	IndoBERT	Menerapkan IndoBERT pada analisis persepsi pemilu 2024 demi transparansi kebijakan publik.	IndoBERT menggunakan arsitektur dasar BERT, sehingga membutuhkan komputasi tinggi.
Onan & Alhumyani, 2024 [4]	*Fuzzy* + BERT	Menggabungkan *FuzzyTM* (*Topic Modeling*) dan BERT untuk ringkasan teks.	Pendekatan *Fuzzy* belum mempertimba ngkan ambiguitas kata atau keanggotaan *Fuzzy* dalam topik.
Munshi et al., 2025 [24]	SOM	Menganalisis ekspresi kesejahteraan (PERMA+H) dalam diskusi longitudinal 20 tahun di Suomi24.	Hanya menerapkan algoritma SOM untuk analisis *cluster* psikologis, tanpa tujuan optimasi output model.
Kusumaningrum et al., 2023 [3]	CNN + LSTM	Mengembangkan aplikasi web sentimen analisis multilevel (ulasan hotel) Bahasa Indonesia.	Minimnya sumber daya teks Bahasa Indonesia untuk sentimen analisis multilevel.
Tujuan Penelitian	DistilBERT + *Fuzzy* + SOM	Mengganti IndoBERT dengan DistilBERT (komputasi ringan) dan mengoptimalkan performanya menggunakan SOM dan *Fuzzy Membership* agar setara dengan IndoBERT.	

Tabel 1 merangkum perbedaan mendasar antara penelitian yang sedang dilakukan dengan studi-studi terdahulu. Dalam konteks analisis sentimen Bahasa Indonesia, model IndoBERT menjadi standar umum karena kemampuannya menangkap konteks lokal secara akurat [7]. Namun, penggunaan model ini menuntut sumber daya komputasi yang sangat besar akibat arsitektur dasar BERT yang kompleks dengan 12 *layer transformer* [1]. Sebagai solusi efisiensi, penelitian lain menerapkan DistilBERT yang berhasil meringankan beban komputasi sebesar 40% dan mempercepat waktu pemrosesan hingga 60%, meskipun terdapat konsekuensi penurunan performa model yang hanya mampu menyamai sekitar 97% dari kemampuan model BERT aslinya [6].

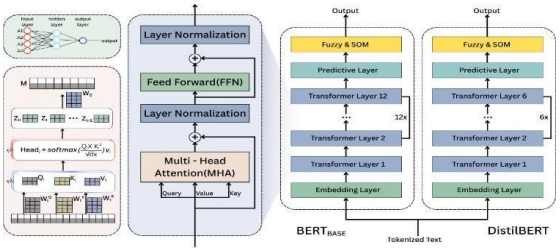
Selain tantangan komputasi, terdapat kendala ketersediaan data pada pengembangan analisis sentimen multilevel, di mana sumber daya teks Bahasa Indonesia masih sangat terbatas jika dibandingkan dengan teks Bahasa Inggris [3]. Dari sisi metodologi, penggunaan algoritma *Fuzzy* sebagai pembantu model juga masih menyisakan celah. Di mana belum diketahui secara pasti apakah algoritma *Fuzzy* mampu memberikan optimalisasi yang signifikan pada kasus tertentu [4]. Hal ini sama dengan pendekatan pada studi kesejahteraan psikologis yang menggunakan *Self Organizing Maps* (SOM), di mana algoritma tersebut sejauh ini hanya dimanfaatkan untuk tujuan *clustering* semata bukan untuk optimasi *output* prediksi pada suatu model [24].

Berbeda dengan keterbatasan-keterbatasan pada penelitian dalam jurnal sebelumnya, penelitian dalam jurnal ini mengusulkan untuk mengembangkan model IndoBERT dengan mengganti dasar model menggunakan DistilBERT. Penelitian ini akan difokuskan pada teks Bahasa Indonesia untuk melakukan sentimen analisis secara multilevel, yang juga akan dioptimalisasi dengan SOM dan algoritma *Trapezoidal Fuzzy Membership*

agar hasil performa DistilBERT dapat lebih menyerupai hasil dari model dasar IndoBERT.

3. METODE PENELITIAN

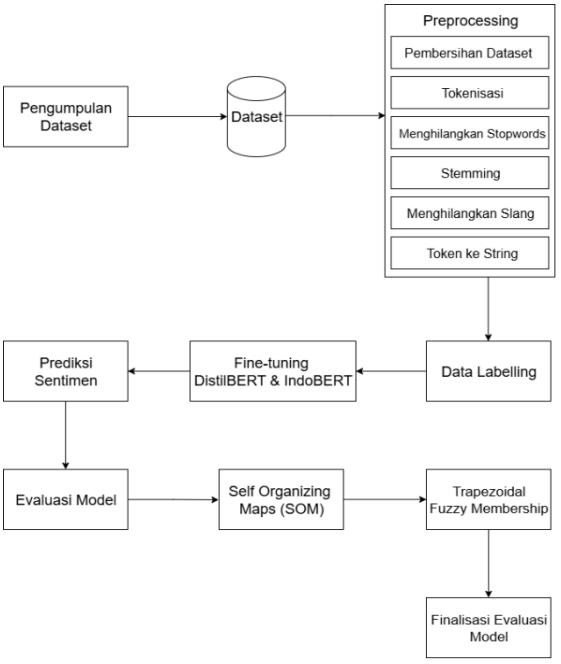
Metodologi yang diusulkan adalah mengganti dasar arsitektur model IndoBERT dengan DistilBERT. Arsitektur model yang diusulkan dapat dilihat pada Gambar 1 berikut.



Gambar 1. Arsitektur model BERT dan DistilBERT

Pada Gambar 1 dapat dilihat bahwa proses dataset teks yang diolah di dalam model BERT diperlukan terlebih dahulu untuk memasuki tahap *transformer layer* sebesar 12 kali setelah dari tahapan *embedding layer*. Fungsi dari *transformer layer* dalam model digunakan sebagai otak model yang memungkinkan pemahaman mendalam terhadap dataset teks yang diberikan [1]. Model DistilBERT mempunyai dasar model yang sama, yaitu BERT. Perbedaannya terdapat pada proses model BERT yang dilakukan proses *knowledge distillation* untuk mengurangi jumlah komputasi yang berat. Proses ini digunakan untuk melakukan kompresi model yang bertujuan membuat model yang lebih kecil dan efisien namun tetap memiliki performa yang mirip dari model aslinya [6].

Model DistilBERT hanya menggunakan 6 *transformer layer*, berbeda dengan jumlah dari model BERT aslinya. Walaupun model DistilBERT sudah dilakukan kompresi dan mempunyai komputasi yang kecil, hasil performa yang dihasilkan dari model tersebut masih tergolong lebih rendah dari model BERT [6]. Oleh karena itu, muncul ide untuk memberikan *layer* tambahan pada model DistilBERT berupa *layer* optimalisasi menggunakan *Self Organizing Maps* (SOM) dan *Fuzzy* berupa *Trapezoidal Fuzzy Membership*. *Layer* tambahan ini berfungsi agar hasil dari model DistilBERT dapat mendekati model BERT aslinya. Model BERT juga digunakan sebagai dasar model dari IndoBERT, dimana model *deep learning* ini biasa digunakan untuk mengolah dataset teks dengan fokus Bahasa Indonesia [7]. Model DistilBERT digunakan dalam penelitian ini sebagai model alternatif pengganti dari model IndoBERT, agar dapat digunakan sebagai metode pengembangan sentimen analisis dalam teks Bahasa Indonesia dengan proses komputasi yang lebih ringan namun hasil tidak berbeda jauh dari model IndoBERT aslinya. Metodologi dalam penelitian ini dapat dilihat pada Gambar 2 berikut.

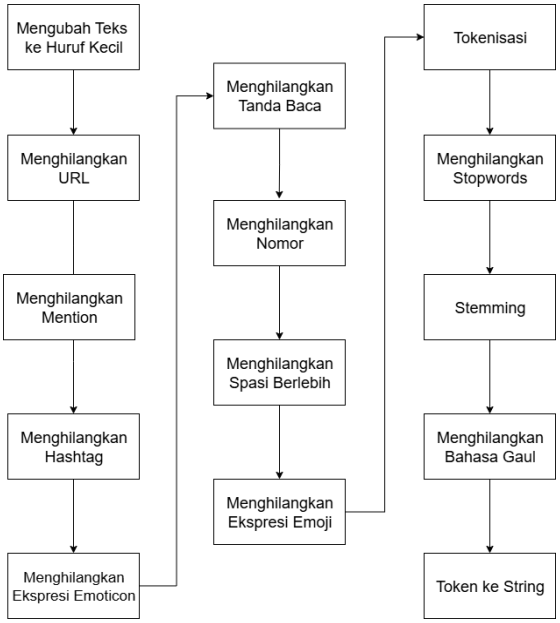


Gambar 2. Tahapan metode penelitian

Tahapan awal pada penelitian adalah pengumpulan data. Penelitian ini menggunakan data sekunder dari penelitian sebelumnya, dengan dataset bernama Indo4B yang digunakan juga sebagai dataset untuk model *pre-train* dari IndoBERT. Data tersebut merupakan kumpulan data teks Bahasa Indonesia dari berbagai sumber seperti berita, media sosial, *wikipedia*, artikel, subtitel dari rekaman video, dan kumpulan data teks pararel Bahasa Indonesia lainnya yang telah dikumpulkan menjadi satu sebagai sumber daya dalam pemrosesan bahasa alami teks Bahasa Indonesia [7]. Setelah dataset selesai dikumpulkan, dataset teks baru masuk ke dalam tahapan *preprocessing*. Tahapan *preprocess* ini berguna untuk mengubah kumpulan data teks kotor atau mentah menjadi data teks bersih yang siap untuk digunakan dalam proses selanjutnya [5]. Setelah *preprocessing*, data masuk ke dalam tahap *data labelling* yang fungsinya untuk melakukan pelabelan sentimen secara multilevel pada data teks mentah yang sebelumnya tidak ada label murni atau *true label* [9].

Setelah dilakukan pelabelan, tahapan seterusnya adalah melakukan *fine-tuning* pada model DistilBERT dan IndoBERT agar model dapat belajar dari dataset teks yang sudah melalui tahapan-tahapan sebelumnya. Setelah *fine-tuning* model, dilakukan klasifikasi sentimen pada dataset agar nantinya model dari DistilBERT dan IndoBERT dapat dilakukan perbandingan evaluasinya untuk melihat hasil performa model. Setelah klasifikasi sentimen dan evaluasi, proses dilanjutkan ke dalam proses *Self Organizing Maps* (SOM) untuk melakukan *clustering* dari nilai *confidence* tiap hasil klasifikasi sentimen yang sudah dihasilkan sebelumnya [25].

Setelah proses SOM, barulah diberi aturan untuk menghilangkan ambiguitas pada hasil klasifikasi sentimen dari *fine-tuning* model DistilBERT dan IndoBERT dengan cara memberikan aturan nilai dari fungsi *Trapezoidal Fuzzy Membership* terhadap nilai *confidence* hasil sentimen [23][22]. Pada tahapan terakhir, baru dilakukan hasil evaluasi setelah dilakukan fungsi dari algoritma *Fuzzy* dan SOM pada hasil klasifikasi. Diharapkan setelah dilakukan evaluasi dapat dilihat bahwa hasil klasifikasi model DistilBERT dapat semakin mendekati performa model IndoBERT jika dibandingkan dengan hasil evaluasi sebelum diterapkan fungsi *Fuzzy* dan SOM. Gambar 3 adalah tahapan di dalam *preprocess* untuk mempersiapkan dataset terlebih dahulu, mulai dari pembersihan dataset hingga normalisasi kata pada dataset lebih lanjut.



Gambar 3. Tahapan persiapan dataset

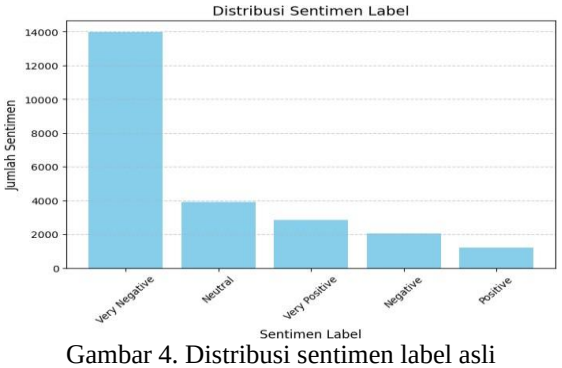
Dalam Gambar 3 pada tahapan pertama, dataset teks diubah terlebih dahulu ke huruf kecil semua. Setelah diubah ke huruf kecil, tahapan dilanjutkan ke proses menghilangkan ciri khas dari

URL *website*. Setelah itu, proses dilanjutkan kembali untuk menghilangkan beberapa huruf atau simbol seperti *mention*, *hashtag*, *emoticon*, tanda baca, nomor, spasi yang berlebih, dan terakhir emoji [5].

Setelah dataset dibersihkan, dataset teks diolah menjadi berbentuk *token* melalui proses tokenisasi. Setelah tokenisasi proses dilanjutkan pada tahap menghilangkan *stopwords*, dimana proses ini menghilangkan kata sambung yang sering tidak merepresentasikan suatu sentimen. Setelah *stopwords* proses dilanjut ke tahap *stemming*, dimana tahapan ini melakukan pengubahan kata ke dalam kata dasar untuk kata kerja dengan imbuhan di awal atau di akhir kalimat.

Setelah proses *stemming*, tahapan proses dilanjut ke tahap menghilangkan bahasa gaul dengan bantuan dari dataset *kamus-alay* Nalsabila (2021) dan *NLP bahasa resources* Louis Owen (2021) yang keduanya tersedia pada *GitHub* [29][30]. Terakhir, setelah melalui semua tahapan dari pembersihan data dan pengolahan data, *token* yang sudah diolah tersebut dijadikan satu kembali dengan berbentuk kalimat bertipe data *string*.

Dikarenakan dataset dari Indo4B tidak memiliki label sentimen asli yang disediakan, perlu dilakukan pelabelan mandiri dalam datasetnya [7]. Dalam penelitian ini digunakan fungsi pelabelan dari model BERT yang sudah dilatih untuk melakukan pelabelan dengan dataset multibahasa, yaitu model dari *nlptown/bert-base-multilingual-uncased-sentiment*. Fungsi pelabelan ini digunakan untuk membuat label asli atau *true label* yang berbentuk sentimen secara multilevel, untuk nantinya digunakan sebagai pembanding dengan hasil klasifikasi untuk membuat evaluasi dari model yang sudah dilakukan *fine-tuning*. Dalam tahap proses *labelling* sentimen dengan model *pre-train* yang digunakan dalam penelitian ini, hasilnya dapat dilihat pada Gambar 4.



Gambar 4. Distribusi sentimen label asli

4. HASIL DAN PEMBAHASAN

4.1. *Fine-tuning* DistilBERT dan IndoBERT

Pada tahapan *fine-tuning*, dipilih model DistilBERT terlebih dahulu untuk melakukan analisis multilevel sentimen. Alasan digunakannya model DistilBERT sendiri dikarenakan memang tujuan penelitian dari awal yang ingin mengembangkan model IndoBERT menjadi model *deep learning* yang lebih ringan dalam komputasi

saat model dijalankan. Ditahapan ini model akan dilatih dengan menggunakan dataset yang telah melalui tahapan-tahapan sebelumnya. Setelah dataset disiapkan, label sentimen akan dijadikan data numerik kembali dengan cara melalui tahapan *encoder* untuk dilakukan pembagian dataset nantinya atau *data splitting*.

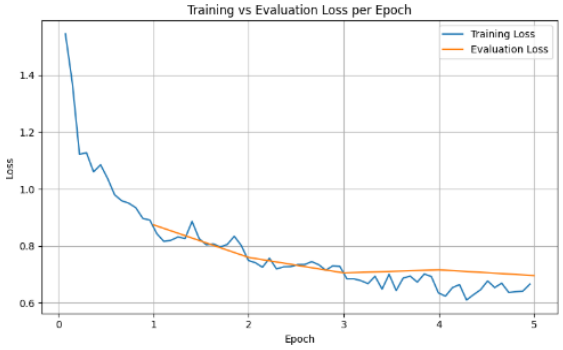
Pada proses pembagian dataset untuk dilanjutkan pada tahapan *fine-tuning*, dataset untuk label yang sudah dirubah menjadi numerik dan dataset teks yang sudah dibersihkan sebelumnya dipisah menjadi dua bagian sebesar 90% data *train*

dan 10% data *validation*. Setelah data dipisah, dilakukan persiapan model berbasis DistilBERT dari model *distilbert/distilbert-base-uncased* yang tersedia pada *Hugging Face*. Pada tahapan akhir, baru proses untuk *fine-tuning* dengan basis DistilBERT dilakukan dan hasilnya akan disimpan dengan berbentuk model *pre-train*. Hasil model yang diperoleh pada tahapan *fine-tuning* dengan DistilBERT ada pada Tabel 2 berikut.

Tabel 2. Performa hasil *fine-tuning* DistilBERT

Epoch	Training Loss	Validation Loss
1	0.8908	0.8742
2	0.7492	0.7598
3	0.7282	0.7053
4	0.6356	0.7162
5	0.6658	0.6962

Seperti yang dapat dilihat pada tabel di atas, selama proses *fine-tuning* dengan model DistilBERT, model menunjukan kinerja konsisten yang dapat dilihat pada nilai *training loss* dan *evaluation loss* yang tidak timpang jauh dari tiap *epoch* yang sudah dijalankan. Ini menandakan bahwa model yang telah dilakukan *fine-tuning* tidak mengalami *overfitting* maupun *underfitting*. Oleh karena itu dapat disimpulkan bahwa model DistilBERT dapat belajar dengan baik dari dataset yang diberi. Hal ini dapat dilihat pada nilai dari *training* dan *evaluation loss* yang tidak saling timpal balik. Ini menunjukan model sudah mempunyai kemampuan untuk melakukan klasifikasi sentimen secara multilevel dengan sangat baik. Untuk hasil pemahaman lebih mendalam mengenai hasil *fine-tuning* dengan model DistilBERT dapat dilihat pada Gambar 5 berikut.



Gambar 5. Visualisasi *fine-tuning* model DistilBERT

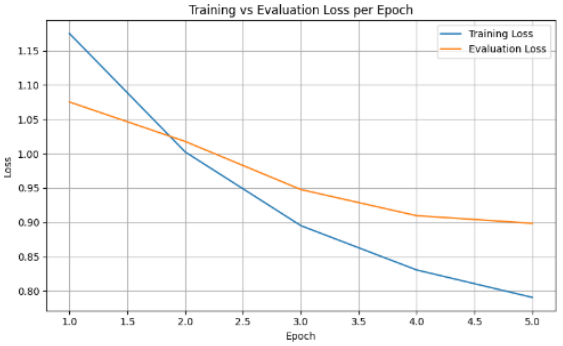
Dari Gambar 5, dapat dipahami dari hasil visualisasi tiap *epoch* dalam *fine-tuning* dengan model DistilBERT, bahwa dari setiap *epoch* model memiliki nilai *loss* yang tidak timpal balik antara *training* dengan *evaluation* berdasarkan dari nilai *epoch* yang ditentukan. Dapat dilihat pula bahwa model pada tahap *fine-tuning* berdasarkan *epoch* yang telah ditentukan, model dapat belajar dengan baik dari dataset yang diberikan pada saat *fine-tuning* model. Berdasarkan Gambar 5, hasil *training loss* menunjukan adanya penurunan dan juga kenaikan pada hasil *evaluation loss* yang tidak terlalu ekstrem. Ini menunjukan model dapat belajar dengan baik dari dataset yang diberikan.

Setelah model DistilBERT dilakukan *fine-tuning* berdasarkan dataset yang diberi, maka selanjutnya adalah melakukan *fine-tuning* pada model IndoBERT. Tahap ini berfungsi untuk mendapatkan hasil perbandingan dengan model IndoBERT dan juga alternatif yang ditawarkan dalam penelitian ini yaitu DistilBERT. Tahapan untuk melakukan *fine-tuning* pada model IndoBERT masih sama seperti tahapan *fine-tuning* pada model DistilBERT. Namun, letak pembedanya hanya terdapat pada *load* model yang digunakan, yaitu model *pre-train* dari *indolem/indobert-base-uncased* yang tersedia pada *Hugging Face*. Pada tahap terakhir, hasil dari *fine-tuning* dengan model IndoBERT disimpan pada lokal sebagai model *pre-train*. Hasil model yang diperoleh pada tahapan *fine-tuning* dengan IndoBERT ada pada Tabel 3 berikut.

Tabel 3. Performa hasil *fine-tuning* IndoBERT

Epoch	Training Loss	Validation Loss
1	1.175	1.0752
2	1.0025	1.0178
3	0.8951	0.9477
4	0.8306	0.9096
5	0.7906	0.8985

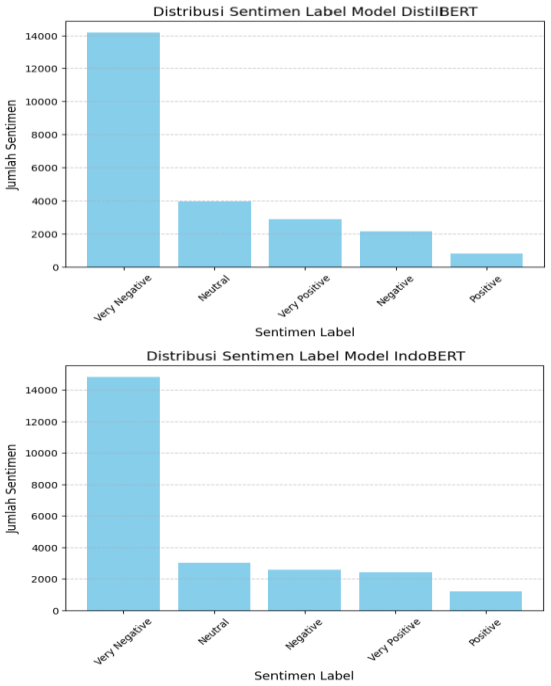
Seperti yang dapat dilihat pada tabel di atas, hasil dari *fine-tuning* model IndoBERT menunjukan kinerja konsisten pada nilai *training loss* dan *validation loss*, dimana nilai dari keduanya tidak jauh berbeda dari tiap *epoch* yang sudah dijalankan. Hasil tersebut dengan model IndoBERT menunjukan bahwa hasil *fine-tuning* seimbang dan tidak mengalami *overfitting* maupun *underfitting*. Untuk melakukan penghitungan waktu komputasi antara kedua model tersebut sebagai tolak ukur besar kecilnya sebuah proses dijalankan masih belum ditentukan pada tahap *fine-tuning*. Untuk pemahaman lebih mendalam pada hasil dari *fine-tuning* dengan model IndoBERT, dapat dilihat pada Gambar 6 berikut.



Gambar 6. Visualisasi *fine-tuning* model IndoBERT

4.2. Klasifikasi Sentimen

Setelah dilakukan tahapan *fine-tuning* dengan model DistilBERT dan IndoBERT, baru dilakukan klasifikasi sentimen pada dataset dengan menggunakan model yang sudah dilatih sebelumnya. Proses tahapan ini digunakan agar nantinya hasil klasifikasi dari model dapat dibandingkan dengan label asli atau *true label* dari dataset yang sudah diberi pelabelan pada tahapan proses sebelumnya. Tujuannya agar dapat menghasilkan hasil evaluasi dari performa model DistilBERT dan IndoBERT yang sudah dilakukan *fine-tuning* dalam klasifikasi sentimen pada tingkat multilevel. Hasil untuk klasifikasi sentimen dalam dataset dengan menggunakan model DistilBERT dan IndoBERT yang sudah dilakukan *fine-tunned* sebelumnya, dapat dilihat pada Gambar 7 berikut.



Gambar 7. Distribusi klasifikasi sentimen label dengan model DistilBERT dan IndoBERT

4.3. *Self Organizing Maps* (SOM)

Proses dari *Self Organizing Maps* (SOM) dipakai dalam penelitian ini untuk melakukan pengelompokan pada nilai token khusus yang digunakan sebagai pembobotan kata pada nilai *confidence* dari model DistilBERT dan IndoBERT yang sudah dilakukan *fine-tuning*. Ukuran *maps* dari SOM yang dipakai dalam eksperimen ini adalah ukuran 10 X 10 dimana hasilnya adalah sebanyak 100 *nodes* atau titik representasi yang mirip. Titik-titik representasi ini, atau *nodes*, digunakan untuk mencari nilai titik representasi dengan nilai pembobotan dari nilai *confidence* dengan nilai yang terlalu beresiko untuk diklasifikasikan menjadi sebuah sentimen. Nantinya titik-titik atau *nodes* yang beresiko ini, akan digabungkan dengan teknik dari *Trapezoidal Fuzzy Membership* untuk ditingkatkan atau direndahkan dari hasil klasifikasi sentimen dengan nilai yang beresiko. Hasil SOM pada model DistilBERT dan IndoBERT dalam mencari *nodes* atau titik-titik representasi yang beresiko pada eksperimen penelitian ini, dapat dilihat pada Tabel 4 berikut.

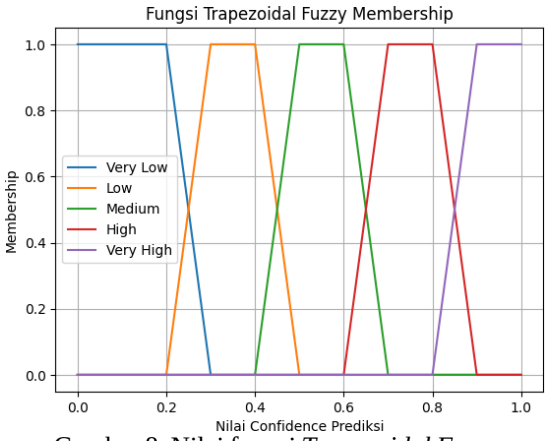
Tabel 3. Performa hasil *fine-tuning* IndoBERT

Epoch	Training Loss	Validation Loss
1	1.175	1.0752
2	1.0025	1.0178
3	0.8951	0.9477
4	0.8306	0.9096
5	0.7906	0.8985

4.4. *Trapezoidal Fuzzy Membership*

Arti dari *Trapezoidal Fuzzy Membership* ini adalah fungsi keanggotaan dalam logika *Fuzzy* dengan berbentuk trapesium. Alasan digunakannya fungsi *Fuzzy Membership* berbentuk trapesium adalah sebagai skala nilai *confidence* dari klasifikasi sentimen berdasarkan model DistilBERT dan IndoBERT pada pembenaran untuk nilai *confidence* agar hasil klasifikasi dapat semakin optimal.

Pada eksperimen ini, ditentukan nilai dari kelima kategori untuk nilai *confidence* mulai yang terkecil sampai terbesar (*very low*, *low*, *medium*, *high*, dan *very high*). Representasi dari nilai fungsi *Trapezoidal Fuzzy Membership* yang ditentukan dapat dilihat pada Gambar 8 berikut.



Gambar 8. Nilai fungsi *Trapezoidal Fuzzy Membership*

Pengaturan nilai *confidence* menggunakan fungsi *Trapezoidal Fuzzy Membership* memiliki fungsi untuk mengklasifikasikan nilai *confidence* ke dalam lima kategori yang berbeda seperti *very low*, *low*, *medium*, *high*, dan *very high*. Kategori *very low* mencakup nilai *confidence* pada rentang 0 hingga 0.2, dengan fungsi keanggotaan yang menurun hingga 0 pada nilai 0.3. Kategori *low* dimulai dari nilai 0.2, mencapai puncak pada 0.3 sampai 0.4, kemudian menurun dan berakhir pada nilai 0.5. Selanjutnya, kategori *medium* dimulai dari 0.4, mencapai maksimum pada rentang 0.5 sampai 0.6, dan menurun hingga 0 pada nilai 0.7. Kategori *high*

dimulai pada nilai 0.6, mencapai nilai maksimum pada rentang 0.7 sampai 0.8, lalu menurun hingga 0 pada nilai 0.9. Terakhir, kategori *very high* dimulai dari nilai 0.8 dan mencapai puncak pada rentang 0.9-1.0. Rentang nilai tersebut mencerminkan tingkat keyakinan model terhadap hasil klasifikasi, mulai dari 0 yang menandakan tidak akurat hingga 1 yang menandakan sangat akurat.

Setelah ditentukan pengaturan nilai pada fungsi *Trapezoidal Fuzzy Membership*, baru dilakukan pembenaran pada nilai *confidence* dari klasifikasi sentimen yang dihasilkan berdasarkan model DistilBERT dan IndoBERT yang sudah dilakukan *fine-tuning* sebelumnya. Contoh dari pembenarannya, jika nilai *confidence* rendah sebesar kurang dari 0.6 maka kelas klasifikasi sentimen akan diturunkan kelasnya sebesar satu kelas pada kelas sentimen *Very Positive* sampai sentimen *Very Negative*. Lalu jika nilai *confidence* sudah tinggi sebesar lebih dari 0.95 atau lebih sama dengan 0.85 maka pertahankan kelas sentimen tersebut. Lalu untuk nilai *confidence* sudah sangat tinggi dan tidak beresiko dari penentuan nilai SOM sebelumnya maka pastikan kelas sentimen tersebut tidak berubah sama sekali. Terakhir jika klasifikasi sentimen sudah sama dengan *true label* atau label aslinya maka pertahankan juga kelas hasil klasifikasi sentimen tersebut.

Setelah pembaruan kelas sentimen dari fungsi *Trapezoidal Fuzzy Membership* ditentukan, nilai-nilai *confidence* tersebut dikelompokan ke dalam fungsi SOM kembali. Proses terakhir, fungsi dari SOM digunakan untuk melakukan koreksi dari fungsi *Fuzzy* sebelumnya. Jika nilai *confidence* tetap rendah walaupun sudah dilakukan fungsi *Fuzzy*, maka kelas sentimen dibenarkan melalui pengelompokan *nodes* dengan fungsi SOM.

4.5. Perbandingan Hasil Evaluasi

Setelah semua tahapan dalam penelitian ini selesai, tahapan dilanjut pada tahap evaluasi yang digunakan untuk meneliti hasil model yang sudah dilakukan dari beberapa tahapan sebelumnya. Pada tahap ini, digunakan untuk membandingkan hasil performa *fine-tuning* dari model yang sudah dijalankan dari hasil sebelum dilakukan fungsi *Trapezoidal Fuzzy Membership* dengan *Self Organizing Maps* (SOM) dan dibandingkan dengan sesudahnya. Selain itu juga, pada tahap ini akan dilakukan perbandingan model DistilBERT dan IndoBERT untuk melihat hasil performa dari DistilBERT yang dioptimalkan agar dapat menyeimbangi hasil performa model IndoBERT.

Pada hasil awalan ditemukan evaluasi pada model *fine-tuning* sebelum fungsi *Fuzzy* dan SOM seperti pada Tabel 5.

Tabel 5. Hasil evaluasi model *fine-tuning* DistilBERT dan IndoBERT

Model	***Accuracy***	***Precision***	***Recall***	**F1-*Score***	**Waktu Eksekusi**
DistilBERT	0.7985	0.80	0.80	0.80	9j 04 m 55d
IndoBERT	0.7239	0.72	0.72	0.72	22j 04m 55d

Dapat dilihat hasil dari model *fine-tuning* yang sudah dijalankan untuk model DistilBERT menghasilkan akurasi sebesar 0.7985 yang mengindikasikan model sudah dapat belajar dengan baik dari dataset yang diberikan. *Precision* sebesar 0.80 yang mengindikasikan model sudah dapat melakukan klasifikasi sentimen dengan tepat. *Recall* sebesar 0.80 yang mengindikasikan model rata-rata menemukan nilai yang tepat dari hasil klasifikasi. Terakhir *F1-Score* sebesar 0.80 yang mengindikasikan hasil dari klasifikasi model ini sudah seimbang. Hasil ini menunjukan bahwa model dari DistilBERT yang sudah dilakukan *fine-tuning* dapat belajar dengan baik dari dataset yang diberikan.

Evaluasi ini telah menjawab sebagian tujuan awal penelitian, yaitu mengembangkan alternatif model IndoBERT yang lebih cepat namun tetap memiliki performa sepadan. Efisiensi ini terlihat jelas pada perbandingan waktu eksekusi, di mana IndoBERT membutuhkan waktu sekitar 22 jam untuk tahapan *fine-tuning*. Durasi yang panjang ini sangat dipengaruhi oleh spesifikasi perangkat keras yang digunakan, yaitu prosesor *Intel(R) Core(TM) i5-11300H* Generasi 11 (*4 Cores, 8 Threads, 3.10 GHz*) dengan RAM 8.00 GB. Selain itu, penggunaan kartu grafis terintegrasi *Intel Iris Xe Graphics* menyebabkan sistem tidak dapat memanfaatkan akselerasi *CUDA* (*Compute Unified Device Architecture*), sehingga beban komputasi dialihkan sepenuhnya ke *CPU*. Pada lingkungan komputasi yang terbatas, DistilBERT membuktikan efisiensinya dengan hanya membutuhkan waktu sekitar 9 jam untuk proses yang sama dan jauh lebih singkat jika dibandingkan dengan model IndoBERT.

Oleh karena itu, tujuan dari penelitian ini ingin menggantikan dasar model dari IndoBERT menjadi DistilBERT. Model ini hanya membutuhkan komputasi kecil saat menjalankan model, juga hasil performa tidak kalah dari model aslinya. Namun model DistilBERT masih perlu untuk dioptimalkan kembali dengan cara melakukan implementasi fungsi *Fuzzy* dan SOM, untuk mendapatkan hasil performa yang lebih baik dari hasil tahapan *fine-tuning* untuk kedua model tersebut.

Untuk mendukung hasil evaluasi pada Tabel 5, terdapat *Confusion* Matrix pada tahap evaluasi ini, yang dapat dilihat pada Gambar 9 berikut.

Gambar 9. *Confusion Matrix* hasil *fine-tuning* DistilBERT dan IndoBERT

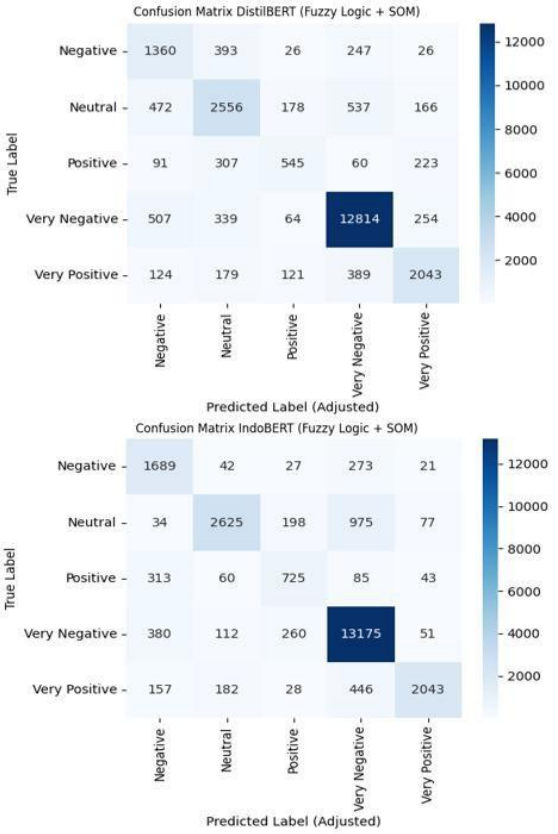
Pada *Confusion Matrix* dari kedua model tersebut dapat dilihat bahwa rata-rata hasil yang sudah dilakukan klasifikasi pada 5 kelas label sentimen sudah tepat. Dapat dilihat pula kelas label sentimen yang paling tepat dari kedua model tersebut adalah kelas label *very negative*.

Setelah evaluasi model *fine-tuning*, kedua model ini dioptimalkan kembali dalam tahap selanjutnya dengan fungsi *Fuzzy* dan SOM. Maka hasil evaluasi setelah tahap optimalisasi dengan fungsi *Fuzzy* dan SOM dapat dilihat pada Tabel 6.

Tabel 6. Hasil evaluasi model *fine-tuning* DistilBERT setelah Fuzzy dan SOM

Model	***Accuracy***	***Precision***	***Recall***	**F1-*Score***
DistilBERT	0.8042	0.81	0.80	0.80
IndoBERT	0.8433	0.85	0.84	0.84

Setelah hasil dari model *fine-tuning* dioptimalkan, dapat dilihat pada model DistilBERT hasil naik sebesar 0.71% dan model IndoBERT naik sebesar 16.51%. Contohnya, dari model DistilBERT hasil akurasi naik menjadi 0.8042, *Precision* naik menjadi 0.81, *Recall* naik menjadi 0.80, dan terakhir *F1-Score* naik menjadi 0.80. Untuk model IndoBERT hasil akurasi naik menjadi 0.8433, *Precision* naik menjadi 0.85, *Recall* naik menjadi 0.84, dan *F1-Score* naik menjadi 0.84. Ini menunjukan, kedua model sudah dapat lebih baik dalam melakukan klasifikasi sentimen yang tepat dalam hasil pembelajaran model dengan DistilBERT dan IndoBERT. Untuk mendukung hasil evaluasi pada tahap setelah dilakukan optimalisasi, dilakukan *Confusion Matrix* yang dapat dilihat pada Gambar 10 berikut.



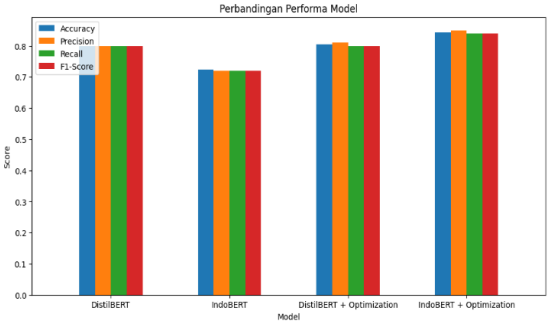
Gambar 10. Confusion Matrix hasil optimalisasi *fine-tuning* DistilBERT dan IndoBERT

Pada *Confusion Matrix* setelah dilakukan optimalisasi pada kedua model tersebut, dapat dilihat bahwa nilai pada sentimen *Negative*, *Neutral*, dan *Positive* pada klasifikasi sentimen yang tepat naik hasilnya. Pada model DistilBERT dari sebelum dilakukan optimalisasi, sentimen *Negative* hanya sebesar 1297, sentimen *Neutral* hanya sebesar 2524, dan *Positive* hanya sebesar 502. Sesudah optimalisasi sentimen *Negative* naik menjadi 1360, *Neutral* naik menjadi 2556, dan *Positive* naik menjadi 545. Hal ini dapat mendukung alasan dibalik meningkatnya hasil performa pada evaluasi setelah optimalisasi.

Analisis komprehensif terhadap Tabel 5 dan Tabel 6 menunjukkan perbandingan performa yang signifikan di antara kedua model. Pada tahap *fine-tuning* awal, DistilBERT justru unggul dengan akurasi 0.7985 jika dibandingkan IndoBERT yang hanya mencapai 0.7239. Namun setelah penerapan

metode optimalisasi menggunakan *Trapezoidal Fuzzy Membership* dan SOM, model IndoBERT menunjukkan lonjakan performa yang drastis, di mana akurasinya meningkat tajam menjadi 0.8433, melampaui DistilBERT yang berada di angka 0.8042.

Hal ini mengindikasikan bahwa arsitektur IndoBERT yang lebih kompleks dengan 12 lapisan *transformer layer* memiliki kapasitas belajar yang lebih besar untuk menyerap fitur-fitur dari optimalisasi *Fuzzy* dan SOM, sehingga mampu mendeteksi kelas sentimen dengan presisi tertinggi. Meskipun secara evaluasi performa IndoBERT pada Tabel 6 lebih unggul, DistilBERT tetap membuktikan keberhasilannya sebagai model alternatif yang efisien. Dengan akurasi sebesar 0.8042, DistilBERT mampu mendekati performa model aslinya dengan selisih yang relatif kecil sebesar 4% namun dengan keunggulan waktu komputasi untuk melakukan *fine-tuning* pada model yang jauh lebih ringkas selama 9 jam berbanding 22 jam dengan model IndoBERT. Sehingga, kombinasi ini menawarkan titik temu yang seimbang antara akurasi yang kompetitif dan efisiensi pada sumber daya komputasi. Untuk mempermudah analisis perbandingan evaluasi performa dari kedua model IndoBERT dan DistilBERT setelah dan sebelum dilakukan optimalisasi dengan *Fuzzy* dan SOM, dibuatlah grafis perbadingan pada Gambar 11 berikut.



Gambar 11. Evaluasi perbandingan performa model sebelum dan sesudah optimalisasi

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil menunjukkan bahwa penggunaan arsitektur *deep learning* DistilBERT sebagai pengganti IndoBERT merupakan solusi efektif dalam pengembangan metode analisis sentimen Bahasa Indonesia yang lebih ringan. Dengan pemangkasan menjadi 6 *transformer layer*, DistilBERT terbukti unggul dalam efisiensi komputasi untuk dataset dalam skala besar dibandingkan model aslinya. Penurunan performa yang biasanya terjadi pada model distilasi berhasil diatasi melalui teknik optimalisasi dengan algoritma *Trapezoidal Fuzzy Membership* dan *Self Organizing Maps* (SOM). Hasil evaluasi mengonfirmasi bahwa kombinasi ini mampu meningkatkan akurasi dari 79.85% menjadi 80.42%, sehingga performa model yang lebih ringkas ini dapat menyamai bahkan melampaui kemampuan model IndoBERT original.

Meskipun menunjukkan hasil positif, penelitian ini memiliki keterbatasan pada cakupan sumber daya komputasi yang hanya menggunakan sampel dataset *bppt* dari korpus Indo4B. Penggunaan data yang lebih spesifik ini menjadi catatan penting dalam proses pelatihan model IndoBERT maupun DistilBERT. Oleh karena itu, penelitian selanjutnya diharapkan dapat memperluas cakupan penggunaan seluruh dataset Indo4B untuk memvalidasi ketangguhan model secara lebih komprehensif. Secara keseluruhan, penelitian ini memberikan landasan kuat bahwa pendekatan efisiensi melalui DistilBERT yang dioptimalkan dengan fungsi *Fuzzy* dan SOM sangat potensial untuk klasifikasi sentimen multilevel yang akurat dan cepat.

DAFTAR PUSTAKA

[1] Bello, A., Ng, S. C., & Leung, M. F. (2023). A BERT Framework to Sentiment Analysis of Tweets. *Sensors*, *23*(1). https://doi.org/10.3390/s23010506

[2] Roudani, M., Elkari, B., El Moutaouakil, K., Ourabah, L., Hicham, B., & Chellak, S. (2024). Twitter-sentiment analysis of Moroccan diabetic using Fuzzy C-means SMOTE and deep neural network. *Mathematical Modeling and Computing*, *11*(3), 835–847. https://doi.org/10.23939/mmc2024.03.835

[3] Kusumaningrum, R., Nisa, I. Z., Jayanto, R., Nawangsari, R. P., & Wibowo, A. (2023). Deep learning-based application for multilevel sentiment analysis of Indonesian hotel reviews.Heliyon, 9(6). https://doi.org/10.1016/j.heliyon.2023.e17147

[4] Onan, A., & Alhumyani, H. A. (2024). FuzzyTP-BERT: Enhancing extractive text summarization with fuzzy topic modeling and transformer networks. *Journal of King Saud University - Computer and Information Sciences*, *36*(6). https://doi.org/10.1016/j.jksuci.2024.102080

[5] Reza, R. F., Muhmmad Thoriq, & Rd. Imam Saepul Millah. (2025). Sentiment Analysis of Marketplace Review with Islamic Perspective using Fine-Tuning DistilBERT. *Khazanah Journal of Religion and Technology*, *2*(2), 45–54. https://doi.org/10.15575/kjrt.v2i2.1118

[6] Gandhi, J. N., Guru, K. V., Kannan, A. R., Sudhan, R. A., Kumar, S. A., & Bharathvaj, M. (2025). *Efficient Sentiment Classification using DistilBERT for Enhanced NLP Performance* (pp. 1500–1511). https://doi.org/10.2991/978-94-6463-718-2_125

[7] Geni, L., Yulianti, E., & Sensuse, D. I. (2023).

Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia Using IndoBert Language Models. *Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika*, *9*(3), 746–757. https://doi.org/10.26555/jiteki.v9i3.26490

[8] Akdik, B., & Sarıman, G. (2025). Hate Speech Detection In Social Media with Deep Learning And Language Models. Bitlis Eren Üniversitesi Fen Bilimleri Dergisi, 14(2), 1077–1095. https://doi.org/10.17798/bitlisfen.1637822

[9] Jojoa, M., Eftekhar, P., Nowrouzi-Kia, B., & Garcia-Zapirain, B. (2024). Natural language processing analysis applied to COVID-19 open-text opinions using a distilBERT model for sentiment categorization. *AI and Society*, *39*(3), 883–890. https://doi.org/10.1007/s00146-022-01594-w

[10] Khatib Sulaiman, J., Putri, M., Edy Sutanto, T., & Inna, S. (n.d.). Studi Empiris Model BERT dan DistilBERT: Analisis Sentimen pada Pemilihan Presiden Indonesia. *Indonesian Journal of Computer Science Attribution*, *12*(5), 2023–2972.

[11] Alzaid, M., & Fkih, F. (2023). Sentiment Analysis of Students' Feedback on E-Learning Using a Hybrid Fuzzy Model. *Applied Sciences (Switzerland)*, *13*(23). https://doi.org/10.3390/app132312956

[12] Costa, E. L. R., Braga, T., Dias, L. A., Albuquerque, É. L. de, & Fernandes, M. A. C. (2022). Analysis of Atmospheric Pollutant Data Using Self-Organizing Maps. *Sustainability (Switzerland)*, *14*(16). https://doi.org/10.3390/su141610369

[13] Dwi Purnomo, T., & Sutopo, J. (2024). COMPARISON *OF PRE-TRAINED BERT-BASED TRANSFORMER MODELS FOR REGIONAL LANGUAGE TEXT SENTIMENT ANALYSIS IN INDONESIA*. *3*, 11–21. https://doi.org/10.56127/ijst.v3i3.1

[14] Dwiyansaputra, R., Ika Murpratiwi, S., & Aranta, A. (2025). ANALISIS SENTIMEN PENGGUNA PLATFORM MEDIA SOSIAL X PADA TOPIK PEMILIHAN PRESIDEN 2024 MENGGUNAKAN PERBANDINGAN MODEL MONOLINGUAL DAN MULTILINGUAL BERT. In Jurnal Mahasiswa Teknik Informatika) (Vol. 9, Issue 1).

[15] Faisal, M. R., Fitriani, K. E., Mazdadi, M. I., Indriani, F., Turianto Nugrahadi, D., & Prastya, S. E. (2025). *ShareAlike 4.0 International License (CC BY-SA 4.0). Enhancing Natural Disaster Monitoring: A Deep Learning Approach to Social Media Analysis Using Indonesian BERT Variants*. *7*(1), 77–89. https://doi.org/10.35882/ijeeemi.v7i1.38

[16] Fransiscus, & Girsang, A. S. (2022). Sentiment Analysis of COVID-19 Public Activity Restriction (PPKM) Impact using BERT Method. *International Journal of Engineering Trends and Technology*, *70*(12), 281–288. https://doi.org/10.14445/22315381/IJETT-V70I12P226

[17] Gandhi, J. N., Guru, K. V., Kannan, A. R., Sudhan, R. A., Kumar, S. A., & Bharathvaj, M. (2025). *Efficient Sentiment Classification using DistilBERT for Enhanced NLP Performance* (pp. 1500–1511). https://doi.org/10.2991/978-94-6463-718-2_125

[18] Guagliardi, I., Astel, A. M., & Cicchella, D. (2022). Exploring Soil Pollution Patterns Using Self-Organizing Maps. *Toxics*, *10*(8). https://doi.org/10.3390/toxics10080416

[19] Kofi Akpatsa, S., Lei, H., Li, X., Kofi Setornyo Obeng, V.-H., Mensah Martey, E., Clement Addo, P., & Dodzi Fiawoo, D. (2022). Online News Sentiment Classification Using DistilBERT. *Journal of Quantum Computing*, *4*(1), 1–11. https://doi.org/10.32604/jqc.2022.026658

[20] Putri, N. R. A., Trimono, T., & Damaliana, A. T. (2024). Sentiment Analysis on Digital Korlantas POLRI Application Reviews Using the Distilbert Model. *Journal of Renewable Energy, Electrical, and Computer Engineering*, *4*(2), 83–89. https://doi.org/10.29103/jreece.v4i2.17197

[21] Sayarizki, P., & Nurrahmi, H. (2024). Implementation of IndoBERT for Sentiment Analysis of Indonesian Presidential Candidates. *Journal on Computing*, *9*(2), 61–72. https://doi.org/10.34818/indojc.2024.9.2.934

[22] Sharma, D. K., Singh, B., Agarwal, S., Pachauri, N., Alhussan, A. A., & Abdallah, H. A. (2023). Sarcasm Detection over Social Media Platforms Using Hybrid Ensemble Model with Fuzzy Logic. *Electronics (Switzerland)*, *12*(4). https://doi.org/10.3390/electronics12040937

[23] Soelistijanto, B. (2022). Construction of Optimal Membership Functions for a Fuzzy Routing Scheme in Opportunistic Mobile Networks. *IEEE Access*, *10*, 128498–128513. https://doi.org/10.1109/ACCESS.2022.3227213

[24] Munshi, M., Pentikäinen, V., Agarwal, V., & Lagus, K. (2025). *Applying Self-Organizing Maps (SOM) to PERMA+H Framework: Analysing Well-Being Expressions in Suomi24 Online Discussions*. http://urn.fi/urn:nbn:fi:lb-2021101527

[25] Yang, Z., Guo, Y., Pang, H., & Yin, F. (2024). Performance Analysis of a Self-Organized Network Dynamics Model for Public Opinion Information. *IEEE Access*, *12*,

55521–55530. https://doi.org/10.1109/ACCESS.2024.3389104

[26] Yerkin, A., Shamoi, P., & Kadyrgali, E. (2024). Sentiment and Emotion-Aware Multi-Criteria Fuzzy Group Decision Making System. *Frontiers in* Artificial Intelligence *and Applications*, *398*, 9–19. https://doi.org/10.3233/FAIA241397

[27] Zhang, Z., Gu, Y., Wang, Z., Luo, S., Sun, S., Wang, S., & Feng, G. (2023). Application of the Self-Organizing Map Method in February Temperature and Precipitation Pattern over China: Comparison between 2021 and 2022. *Atmosphere*, *14*(7). https://doi.org/10.3390/atmos14071182

[28] Şentürk, A., Albayrak, A., & Arpacı, S. (2026). Multi-Class News Classification with BERT, DistilBERT, RoBERTa, and ELECTRA Natural Language Processing Models. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, *14*(1), 117–129. https://doi.org/10.29130/dubited.1737003

[29] louisowen6. GitHub - louisowen6/NLP_bahasa_resources: A Curated List of Dataset and Usable Library Resources for NLP in Bahasa Indonesia. GitHub. Published 2022. https://github.com/louisowen6/NLP_bahasa_resources nasalsabila. GitHub - nasalsabila/kamus-alay. GitHub. Published 2020. https://github.com/nasalsabila/kamus-alay