ANALISIS SENTIMEN MULTI-PLATFORM DIGITAL MENGGUNAKAN BERT DENGAN PENDEKATAN PENYEIMBANGAN DISTRIBUSI KELAS

Arridho Ramadhan Firdaus, Fadlillah Mukti Ayudewi, Arizona Firdonsyah
Program Studi Teknologi Informasi, Universitas Aisyiyah Yogyakarta
Jl. Siliwangi Jl. Ringroad Barat No.63, Mlangi, Nogotirto, Kec. Gamping,
Kabupaten Sleman, Daerah Istimewa Yogyakarta, Indonesia 55292
arridhoramadhanfirdaus@unisayogya.ac.id

ABSTRAK

Analisis sentimen merupakan bagian penting dalam Natural Language Processing (NLP) yang digunakan untuk memahami opini pengguna terhadap layanan digital melalui data teks. Meningkatnya jumlah ulasan pada berbagai platform daring menuntut metode klasifikasi sentimen yang akurat dan andal. Namun, permasalahan utama yang sering dihadapi adalah ketidakseimbangan distribusi kelas sentimen, di mana data positif mendominasi dibandingkan sentimen netral dan negatif, sehingga berpotensi menurunkan performa model. Penelitian ini bertujuan untuk mengevaluasi kinerja model BERT dalam klasifikasi sentimen ulasan multi-platform serta menganalisis pengaruh teknik augmentasi data dalam mengatasi ketidakseimbangan kelas. Metode penelitian meliputi tahap *pre-processing* teks, penerapan teknik augmentasi data berupa *Synonym Replacement, Contextual Word Embeddings,* dan *Back Translation,* serta pelatihan dan pengujian model BERT pada dataset *Coursera, Starbucks, dan Threads.* Evaluasi dilakukan menggunakan metrik akurasi, *precision, recall,* dan *F1-score.* Hasil penelitian menunjukkan bahwa penerapan kombinasi teknik augmentasi data mampu meningkatkan performa model secara signifikan, dengan akurasi tertinggi mencapai 99,83% pada dataset *Coursera,* serta meningkatkan kemampuan generalisasi model dalam klasifikasi sentimen.

Kata kunci : *Sentimen, BERT, Data Tidak Seimbang, Teknik Augmentasi Data, Coursera*

1. PENDAHULUAN

Analisis sentimen telah menjadi fokus utama dalam bidang *Natural Language Processing (NLP)* [1]. Klasifikasi sentimen memungkinkan sistem untuk memahami opini atau pemikiran yang terkandung dalam teks, seperti ulasan produk atau layanan [2]. Studi ini berfokus pada penggunaan model NLP terkini yaitu BERT untuk melakukan klasifikasi sentimen pada ulasan platform daring [3]. Penelitian ini juga mengeksplorasi penggunaan teknik *augmentasi* data untuk mengatasi masalah ketidakseimbangan kelas dalam kumpulan data [4]. Ketidakseimbangan data sering kali menyebabkan model lebih cenderung mengklasifikasikan data ke dalam kelas mayoritas yang mengurangi efektivitas model dalam mendeteksi kelas minoritas [5]. Teknik *augmentasi* seperti *Back Translation, Contextual Word Embeddings* dan *Synonym Replacement* diterapkan untuk meningkatkan kualitas dan akurasi model [6].

Seiring dengan pesatnya perkembangan platform digital, jumlah ulasan pengguna yang tersedia dalam bentuk teks mengalami peningkatan yang sangat signifikan. Hampir seluruh layanan digital saat ini, baik platform pembelajaran daring, layanan komersial, maupun media sosial, menyediakan ruang bagi pengguna untuk menyampaikan pengalaman, opini, serta tingkat kepuasan mereka terhadap layanan yang digunakan. Ulasan-ulasan tersebut menjadi sumber data yang sangat bernilai karena mencerminkan persepsi pengguna secara langsung, objektif, dan bersifat real-time. Informasi yang terkandung di dalamnya dapat dimanfaatkan untuk mengevaluasi kualitas layanan, mengidentifikasi kelebihan dan kekurangan sistem, serta mendukung pengambilan keputusan berbasis data. Oleh karena itu, analisis sentimen menjadi pendekatan yang strategis untuk mengekstraksi dan mengolah informasi tersebut secara otomatis dan berskala besar [1], [2].

Dalam perkembangannya, berbagai pendekatan telah diterapkan untuk melakukan klasifikasi sentimen, mulai dari metode berbasis *leksikon, machine learning* tradisional, hingga model *deep learning.* Salah satu model yang menunjukkan performa unggul dalam berbagai tugas NLP adalah BERT. Model ini dirancang untuk memahami konteks bahasa secara dua arah sehingga mampu menangkap makna semantik dan nuansa sentimen dalam teks secara lebih mendalam dibandingkan pendekatan sebelumnya [9], [13]. Keunggulan tersebut menjadikan BERT sangat relevan untuk digunakan dalam analisis sentimen pada data ulasan yang kompleks dan bervariasi, khususnya pada platform digital dengan gaya bahasa yang beragam.

Meskipun demikian, penerapan BERT pada data ulasan nyata masih menghadapi sejumlah tantangan. Salah satu tantangan utama adalah ketidakseimbangan distribusi kelas sentimen. Dalam banyak kasus, data ulasan didominasi oleh sentimen positif, sementara sentimen netral dan negatif memiliki jumlah yang jauh lebih sedikit [4], [5]. Kondisi ini dapat menyebabkan model pembelajaran mesin menjadi bias terhadap kelas mayoritas, sehingga performa prediksi pada kelas minoritas menurun. Padahal, sentimen negatif dan netral sering kali memuat informasi kritis yang penting dalam konteks evaluasi layanan, seperti

keluhan pengguna, kelemahan sistem, atau aspek layanan yang perlu diperbaiki [4].

Untuk mengatasi permasalahan tersebut, diperlukan strategi yang mampu meningkatkan representativitas data latih tanpa mengubah makna sentimen asli. Salah satu pendekatan yang banyak digunakan adalah teknik *augmentasi* data teks. Augmentasi data bertujuan untuk memperkaya variasi data latih dengan menghasilkan data sintetis yang tetap mempertahankan konteks dan polaritas sentimen [5]. Teknik *augmentasi* seperti *Synonym Replacement, Contextual Word Embeddings*, dan *Back Translation* memungkinkan terciptanya variasi linguistik yang lebih beragam, sehingga model dapat belajar dari pola bahasa yang lebih luas dan seimbang [6].

Dengan mengombinasikan kemampuan representasi kontekstual BERT dan teknik *augmentasi* data teks, penelitian ini diharapkan mampu meningkatkan performa klasifikasi sentimen secara signifikan. Pendekatan ini tidak hanya bertujuan untuk meningkatkan nilai akurasi secara keseluruhan, tetapi juga untuk memperbaiki kemampuan model dalam mengenali sentimen pada kelas minoritas. Hal ini menjadi sangat penting terutama pada analisis sentimen ulasan multi-platform digital yang memiliki karakteristik data kompleks, tidak seimbang, dan terus berkembang [4], [6].

2. TINJAUAN PUSTAKA

2.1. Penelitian terdahulu

Berbagai penelitian terdahulu telah mengkaji penerapan analisis sentimen pada data ulasan dan media sosial menggunakan beragam pendekatan dan model pembelajaran mesin. Ngoc et al. [1] melakukan analisis sentimen terhadap ulasan mahasiswa pada platform pembelajaran daring dengan pendekatan transfer learning. Hasil penelitian menunjukkan bahwa model berbasis deep learning mampu meningkatkan akurasi klasifikasi sentimen dibandingkan metode konvensional, terutama dalam memahami konteks ulasan akademik yang kompleks.

Turjaman dan Budi [2] menerapkan analisis sentimen berbasis aspek pada ulasan aplikasi dompet digital menggunakan data dari Twitter. Penelitian tersebut menegaskan bahwa analisis sentimen dapat memberikan wawasan strategis terkait persepsi pengguna terhadap layanan digital, meskipun masih menghadapi tantangan pada variasi bahasa dan ketidakseimbangan data. Studi serupa juga dilakukan oleh Mishra et al. [3] yang memanfaatkan data Twitter untuk ekstraksi sentimen, menunjukkan bahwa kompleksitas bahasa informal media sosial menuntut penggunaan model NLP yang lebih kontekstual.

Dalam konteks penggunaan model BERT, Fimoza et al. [9] membuktikan bahwa BERT memberikan performa yang unggul dalam klasifikasi sentimen ulasan film berbahasa Indonesia dibandingkan pendekatan deep learning lainnya. Keunggulan BERT juga diperkuat oleh Devlin et al. [13] yang memperkenalkan arsitektur BERT sebagai model pre-trained berbasis Transformer dengan kemampuan pemahaman konteks dua arah, sehingga sangat efektif untuk tugas klasifikasi teks. Penelitian lanjutan oleh Ananthakrishnan et al. [14] dan Sebastian et al. [15]. menunjukkan bahwa BERT mampu menangani teks dengan struktur kompleks dan variasi bahasa yang tinggi, termasuk pada data media sosial dan Bahasa Indonesia.

Namun demikian, beberapa penelitian menyoroti bahwa performa model BERT dapat menurun ketika dihadapkan pada dataset dengan distribusi kelas yang tidak seimbang. Suhaeni dan Yong [4] mengungkapkan bahwa ketidakseimbangan data merupakan permasalahan utama dalam analisis sentimen, karena model cenderung bias terhadap kelas mayoritas. Abonizio et al. [5] juga menegaskan bahwa tanpa penanganan khusus, model klasifikasi sentimen akan mengalami penurunan recall dan F1-score pada kelas minoritas.

Untuk mengatasi permasalahan tersebut, teknik augmentasi data teks banyak digunakan dalam penelitian terdahulu. Abonizio et al. [5] menunjukkan bahwa augmentasi data mampu meningkatkan kualitas data latih dan performa model secara signifikan. Nair et al. [6] mengevaluasi dampak augmentasi data teks menggunakan DistilBERT dan menemukan bahwa teknik seperti Synonym Replacement dan Back Translation efektif dalam meningkatkan performa klasifikasi, terutama pada dataset tidak seimbang. Hasil ini sejalan dengan temuan Ahmed et al. [8] dan Başarslan & Kayaalp [11] yang menyatakan bahwa pengayaan data latih berperan penting dalam meningkatkan kemampuan generalisasi model deep learning pada analisis sentimen.

Berdasarkan kajian penelitian terdahulu tersebut, dapat disimpulkan bahwa model BERT memiliki potensi yang sangat besar dalam tugas klasifikasi sentimen. Namun, performa optimalnya sangat dipengaruhi oleh kualitas dan distribusi data latih. Oleh karena itu, penelitian ini memposisikan diri dengan mengintegrasikan model BERT dan teknik augmentasi data teks untuk menangani ketidakseimbangan kelas pada analisis sentimen ulasan multi-platform digital, sehingga diharapkan mampu menghasilkan model yang lebih akurat, robust, dan memiliki kemampuan generalisasi yang lebih baik.

2.2. Analisis Sentimen dalam Natural Language Processing

Analisis sentimen merupakan salah satu cabang penting dalam bidang Natural Language Processing (NLP) yang bertujuan untuk mengidentifikasi dan mengklasifikasikan opini, emosi, atau sikap pengguna terhadap suatu objek, layanan, maupun fenomena tertentu dalam bentuk teks [1]. Penerapan analisis sentimen banyak digunakan pada ulasan produk, media sosial, serta platform layanan digital karena mampu memberikan gambaran persepsi publik secara otomatis dan berskala besar [2].

Dalam konteks platform pembelajaran daring, analisis sentimen terhadap ulasan pengguna menjadi sangat relevan karena dapat membantu penyedia layanan memahami tingkat kepuasan pengguna serta mengidentifikasi aspek yang perlu ditingkatkan [1]. Berbagai penelitian menunjukkan bahwa klasifikasi sentimen berbasis teks mampu memberikan informasi strategis bagi pengambilan keputusan dan evaluasi kualitas layanan digital [3]

2.3. Permasalahan Ketidakseimbangan Data pada Analisis Sentimen

Salah satu tantangan utama dalam analisis sentimen adalah ketidakseimbangan distribusi kelas (class imbalance), di mana jumlah data pada kelas tertentu jauh lebih dominan dibandingkan kelas lainnya [4]. Kondisi ini sering terjadi pada ulasan daring, yang cenderung didominasi oleh sentimen positif, sementara sentimen negatif dan netral memiliki jumlah yang jauh lebih sedikit [5].

Ketidakseimbangan data dapat menyebabkan model pembelajaran mesin menjadi bias terhadap kelas mayoritas, sehingga performa klasifikasi pada kelas minoritas menurun secara signifikan [4]. Oleh karena itu, diperlukan pendekatan khusus untuk mengatasi masalah ini agar model dapat belajar secara lebih representatif dan menghasilkan prediksi yang adil pada seluruh kelas sentimen [5].

2.4. Teknik Augmentasi Data Teks

Augmentasi data merupakan salah satu pendekatan yang efektif untuk mengatasi ketidakseimbangan data dengan cara menambah jumlah data latih secara sintetis tanpa mengubah makna dasar teks [5]. Beberapa teknik augmentasi yang umum digunakan dalam analisis sentimen antara lain Synonym Replacement, Contextual Word Embeddings, dan Back Translation [6].

Synonym Replacement bekerja dengan mengganti kata tertentu dalam kalimat dengan sinonimnya untuk menciptakan variasi teks baru, sehingga memperkaya data latih tanpa mengubah konteks sentimen [6]. Contextual Word Embeddings memanfaatkan model berbasis konteks untuk menghasilkan variasi kata yang sesuai dengan makna kalimat secara keseluruhan, sehingga lebih adaptif terhadap konteks linguistik [12]. Sementara itu, Back Translation menghasilkan data baru dengan menerjemahkan teks ke bahasa lain dan kemudian menerjemahkannya kembali ke bahasa asal, yang terbukti efektif dalam meningkatkan keragaman struktur kalimat [5].

Penelitian sebelumnya menunjukkan bahwa penerapan teknik augmentasi teks mampu meningkatkan performa model klasifikasi sentimen secara signifikan, terutama pada dataset yang tidak seimbang [6].

2.5. Model BERT untuk Klasifikasi Sentimen

BERT (Bidirectional Encoder Representations from Transformers) merupakan model pre-trained language model berbasis arsitektur Transformer yang dirancang untuk memahami konteks bahasa secara dua arah [13]. Keunggulan utama BERT terletak pada kemampuannya menangkap hubungan semantik antar kata dalam kalimat dengan mempertimbangkan konteks sebelum dan sesudah kata tersebut [14].

Dalam tugas klasifikasi sentimen, BERT telah terbukti memberikan performa yang unggul dibandingkan metode tradisional maupun model deep learning konvensional [9], [11]. Model ini memanfaatkan representasi vektor dari token [CLS] sebagai ringkasan konteks keseluruhan kalimat yang kemudian digunakan untuk proses klasifikasi sentimen [13].

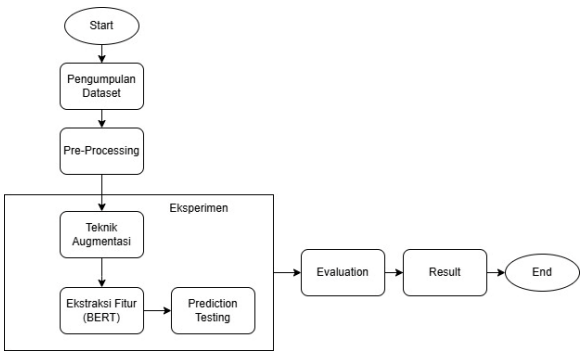
Beberapa penelitian menunjukkan bahwa BERT sangat efektif dalam menangani teks ulasan yang kompleks dan bervariasi, termasuk ulasan media sosial dan platform digital [14], [15]. Oleh karena itu, BERT menjadi pilihan utama dalam penelitian ini untuk mengklasifikasikan sentimen ulasan pengguna pada platform daring.

2.6. Integrasi BERT dan Augmentasi Data

Integrasi antara model BERT dan teknik augmentasi data telah banyak diteliti sebagai solusi untuk meningkatkan performa klasifikasi sentimen pada data tidak seimbang [5], [6]. Augmentasi data membantu memperkaya variasi teks, sementara BERT berperan dalam mengekstraksi representasi fitur kontekstual yang mendalam dari teks tersebut.

Penelitian terdahulu menunjukkan bahwa penggunaan augmentasi data secara terpadu mampu meningkatkan akurasi, precision, recall, dan F1-score secara signifikan dibandingkan model tanpa augmentasi [4], [6]. Dengan demikian, kombinasi BERT dan teknik augmentasi diharapkan mampu menghasilkan model klasifikasi sentimen yang lebih akurat, robust, dan memiliki kemampuan generalisasi yang lebih baik pada berbagai jenis ulasan pengguna.

3. METODE PENELITIAN



Gambar 1. Alur Penelitian

Penelitian ini menggunakan dataset *review* dari platform *Coursera* untuk menguji kinerja model BERT [4]. Dataset ini terdiri dari *review* pengguna yang diberi rating 1-5, yang kemudian dikategorikan menjadi tiga kelas: negatif (rating 1-2), netral (rating 3), dan positif (rating 4-5) [7]. Alur tahapan penelitian sistem menggunakan teknik augmentasi disajikan melalui Gambar 1.

3.1. Dataset

Dataset terdiri dari *Coursera*. Penelitian ini dapat menguji kinerja model dalam situasi yang lebih nyata dan beragam. Penelitian ini juga membuka peluang untuk mengevaluasi kemampuan model dalam mengklasifikasikan sentimen dalam konteks bahasa yang lebih luas dan berbagai gaya penulisan yang mungkin ada. Dataset yang digunakan relatif besar, yaitu 10.000 ulasan untuk *Coursera*. Dataset dalam pembagiannya dibagi menjadi tiga subset diantaranya, data pelatihan (80%), data validasi (10%), dan data pengujian (10%).

3.2. *Pre-Processing*

Pre-processing yang dilakukan *Case Folding, Remove Punctuation, Tokenization* dan *Label Encoding* yang merupakan langkah standar yang sangat penting dalam menyiapkan data untuk analisis sentimen [8]. *Case Folding* bertujuan untuk mengubah huruf besar menjadi huruf kecil pada teks [9]. Remove Punctuation bertujuan menghilangkan tanda baca sangat tepat karena dalam analisis sentimen, simbol atau emoji mungkin tidak memberikan informasi yang relevan, dan pemrosesan ini akan membantu model fokus pada kata dan frasa [10]. Tokenisasi dan *label encoding* akan memfasilitasi pemrosesan teks dan meminimalkan kesalahan saat model mempelajari data [11].

Tahapan *pre-processing* telah selesai dilakukan, dan selanjutnya dataset akan diproses pada tahap augmentasi data. Hasil dari tahapan *pre-processing* disajikan pada Tabel 1.

Tabel 1. *Pre-Processing*

Proses	Hasil
Data Awal	Sorry to say, but among the four courses of this specialization, this one really disappointed me.
Case folding	sorry to say, but among the four courses of this specialization, this one really disappointed me.
Remove punctuation	sorry to say but among the four courses of this specialization this one really disappointed me
Tokenization	['sorry', 'to', 'say', 'but', 'among', 'the', 'four', 'courses', 'of', 'this', 'specialization', 'this', 'one', 'really', 'disappointed', 'me']
Data akhir	sorry to say but among the four courses of this specialization this one really disappointed me

3.3. Teknik Augmentasi

Penelitian ini menggunakan teknik augmentasi data berupa *Synonym Replacement, Contextual Word Embeddings* dan *Back Translation* yang bertujuan untuk menyeimbangkan data. Teknik *Synonym Replacement* sangat berguna untuk menambahkan variasi teks tanpa mengubah maknanya, yang penting dalam menangani data yang tidak seimbang [6], terutama dalam kumpulan data dengan lebih banyak ulasan positif daripada ulasan negatif atau netral. *Contextual Word Embeddings* memungkinkan model untuk memahami konteks yang lebih dalam dari setiap ulasan [12] dan *Back Translation* memberikan variasi frasa baru sambil mempertahankan makna aslinya [5], sehingga meningkatkan keragaman data tanpa kehilangan informasi. Penerapan ketiga teknik penambahan ini akan memperkaya kumpulan data dan memberi model lebih banyak contoh untuk dipelajari, sehingga meningkatkan kemampuan prediktif model terhadap ulasan minoritas. Hasil dari augmentasi seperti ditunjukkan pada Tabel 2.

Tabel 2. Hasil Teknik Augmentasi

Teknik Augmentasi	Kalimat asli	Hasil
Synonym Replacement	coursea does not allow me to **quit** the class also i cannot do the homework or watch **video** at my **own** pace	coursea does not allow me to **fall by the wayside** the class also i can not do the homework or watch **video recording** at my **possess** pace
Contextual Embedding	coursea does not allow me to quit the class also i **cannot** do the homework or watch video at my own pace	coursea does not allow me to quit the class also i **can not** do the homework or watch video at my own pace
Back Translations	**coursea** does not allow me to **quit** the class also i cannot do the homework or watch video at my own pace	**kursa** does not allow me to **finish** the class also I can not do the homework or watch video at my own pace

3.4. Eksperimen

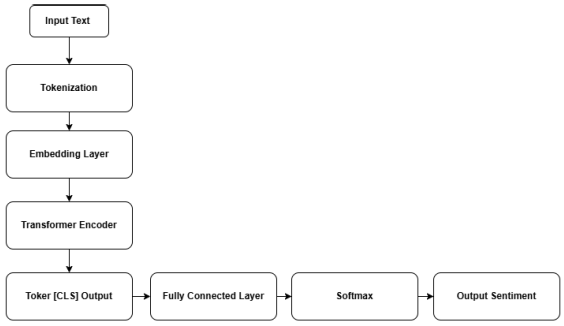
Tahap eksperimen yang dilakukan dengan membagi data menjadi tiga yaitu 80% data latih, 10% data validasi, dan 10% data tes, merupakan praktik yang baik untuk memastikan bahwa model yang dilatih dapat dievaluasi secara objektif dan transparan. Eksperimen yang didukung oleh teknik augmentasi, peneliti dapat lebih yakin hasilnya akan memberikan gambaran yang lebih akurat tentang kinerja model dalam kondisi dunia nyata. Data yang sudah dibagi menjadi 3 subset akan dilakukan *pre-trained* hingga klasifikasi sentimen menggunakan model BERT. Skenario pengujian dapat dilihat pada Tabel 3.

Tabel 3. Eksperimen

Parameter	Nilai Parameter Yang Diujikan
Model Bahasa	BERT
Teknik Augmentasi	*Synonym Replacement, Contextual Word Embeddings, Back Translation*
Dataset	Review *Coursera*, Review *Starbucks*, dan Review *Threads*
Eksperimen 1	BERT + *Synonym Replacement*, BERT + *Contextual Word Embeddings*, BERT + *Back Translation*
Eksperimen 2	BERT + *Synonym Replacement, Contextual Word Embeddings, Back Translation*
Eksperimen 3	BERT + Tanpa augmentasi

3.5. Ekstraksi Fitur dan Model

Ekstraksi fitur menggunakan BERT sangat cocok untuk masalah analisis sentimen, karena BERT ini terbukti efektif dalam memahami data teks yang kompleks serta menangkap konteks [13]. Model menawarkan kelebihannya sendiri dengan kemampuannya memahami konteks dua arah [14]. Dengan memanfaatkan model ini, studi ini berpotensi memperoleh hasil optimal untuk klasifikasi sentimen dari kumpulan data yang ada. BERT *(Bidirectional Encoder Representations from Transformers)* menggunakan pendekatan transformer untuk ekstraksi fitur dalam teks [15]. Berikut rancangan *Step By Step* model BERT ditunjukkan pada Gambar 2.



Gambar 2. Rancangan *step by step* model BERT

Gambar 2. menampilkan rancangan proses analisis sentimen menggunakan model BERT dimulai dengan memasukkan teks mentah seperti ulasan pelanggan ke dalam model. Teks ini kemudian dipecah menjadi token menggunakan tokenisasi, dengan penambahan token khusus seperti [CLS] di awal teks untuk merepresentasikan keseluruhan kalimat dan [SEP] di akhir teks untuk menunjukkan batas teks. Setiap token diubah menjadi representasi vektor melalui *embedding layer*, yang mencakup 3 jenis embedding yaitu *token embedding* yang berguna untuk mengonversi setiap kata atau token menjadi vektor yang mencerminkan makna kata tersebut dalam konteks model, sedangkan *position embedding* memberikan informasi mengenai posisi token dalam kalimat, yang memungkinkan model untuk memahami urutan kata dan konteks yang bergantung pada posisi kata, dan *segment embedding* memberikan informasi tentang bagian mana dari input yang merupakan kalimat pertama dan mana yang merupakan kalimat kedua.

Token *embedding* diproses oleh beberapa lapisan *encoder Transformer* yang berjumlah 12 lapisan. Mekanisme *self-attention* dalam encoder ini menghitung hubungan antar token untuk memahami konteks kedua arah, sedangkan *feed-forward network* menambahkan pola non-linear untuk meningkatkan pemahaman [1]. Representasi dari token [CLS] yang dihasilkan oleh encoder digunakan sebagai vektor utama yang mencerminkan konteks keseluruhan teks. Vektor ini kemudian dilewatkan ke lapisan *fully connected*, yang mentransformasikannya menjadi representasi yang lebih sederhana untuk keperluan prediksi [11]. Pada tahap akhir, fungsi aktivasi seperti *softmax* diterapkan untuk menghasilkan probabilitas sentimen, dan kelas sentimen ditentukan berdasarkan probabilitas tertinggi, misalnya positif, negatif, atau netral.

3.6. Evaluasi

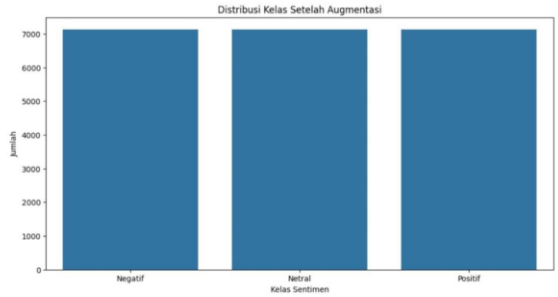
Tahap evaluasi yang menggunakan akurasi sebagai tolak ukur utama untuk mengevaluasi kinerja model dalam klasifikasi sentimen merupakan langkah yang tepat. Evaluasi yang lebih mendalam dengan menggunakan metrik lain seperti presisi, recall, dan F1-score dapat menjadi tambahan yang sangat berguna untuk menggali lebih dalam kekuatan dan kelemahan masing-masing model dalam berbagai konteks data. Hasil perhitungan dari *confusion matrix* akan digunakan sebagai acuan untuk menilai menggunakan BERT dapat memberikan hasil yang baik. Penjelasan mengenai metrik evaluasi beserta persamaan ditunjukkan pada Tabel 4.

Tabel 4. Metrik Evaluasi

Metrik	Deskripsi	Rumus
Akurasi	Menggambarkan seberapa akurat model dapat mengklasifikasikan dengan benar	$\frac{TP+TN}{TP+FP+FN+TN}$ (1)
Precision	Menggambarkan prediksi antara data yang diminta dengan hasil dari prediksi yang diberikan oleh model.	$\frac{TP}{TP+FP}$ (2)
Re-Call	Menggambarkan seberapa banyak kelas psoitif yang diperdiksi dengan benar.	$\frac{TP}{TP+FN}$ (3)
F1-Score	Rata-rata precision dan Recall.	$\frac{2*(Precision*Recall)}{Precison+Recall}$ (4)

4. HASIL DAN PEMBAHASAN

Data yang digunakan berjumlah 10.000 data dengan distribusi label yang tidak seimbang. Terdapat 7.134 label positif, 470 label netral, dan 396 label negatif. Setelah dilakukan augmentasi menggunakan teknik *Synonym Replacement*, *Contextual Word Embedding* dan *Back Translation* menjadi seimbang dengan distribusi yang disajikan pada Gambar 3.



Gambar 3. Distribusi Dataset Setelah Augmentasi

Data kemudian dibagi menjadi tiga diantaranya 80% latih, 10% validasi, dan 10% tes. Pembagian ini bertujuan untuk menjaga distribusi kelas pada setiap subset data tetap merata, memastikan bahwa model dapat diuji secara merata pada kelas yang seimbang. Hasil pembagian dapat dilihat pada tabel 5.

Tabel 5. Hasil Pembagian Data

Dataset	Positif	Negatif	Netral	Persentase
Training	5.707	5.707	5.707	80%
Validation	714	714	714	10%
Testing	714	714	714	10%

Data setelah dibagi, selanjutnya dilakukan tahapan penelitian berdasarkan alur yang telah dibuat, dan ini adalah hasil dari 3 percobaan yang telah dilakukan untuk percobaan 1 akan menggunakan BERT yang menggunakan tiga teknik augmentasi yang berbeda *(Synonym Replacement*, *Contextual Word Embeddings* dan *Back Translation*) pada dataset (*Coursera*). Berikut ini adalah hasil eksperimen 1 yang ditunjukkan pada Tabel 6.

Tabel 6. *Eksperimen 1*

Dataset	*Synonym Replacement*	*Contextual Word Embeddings*	Back Translation
Coursera	0.9673	0.9678	0.9692

Tabel 6 eksperimen 1, yang menguji tiga teknik augmentasi secara terpisah, BERT mencatat performa terbaik pada teknik *Back Translation* di dataset *Coursera* dengan F1-Score sebesar 0.9874, yang menunjukkan bahwa model ini sangat efektif dalam memahami konteks meskipun inputnya telah mengalami perubahan struktur kalimat. Hal ini membuktikan kemampuan representasi kontekstual BERT yang kuat, terutama saat menghadapi data terstruktur seperti ulasan *Coursera*. Eksperimen 2 ditunjukkan pada tabel 7.

Tabel 7. Eksperimen 2

Dataset	Kombinasi
Coursera	0.9983

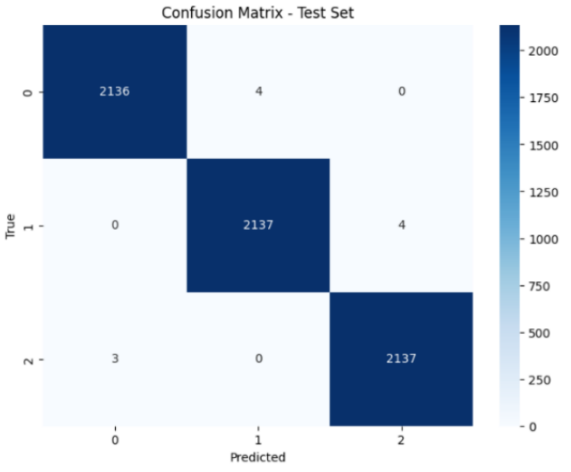
Table 7 merupakan eksperimen 2, di mana ketiga teknik augmentasi digabungkan, BERT berhasil mencapai F1-Score sebesar 0.9983 di dataset *Coursera*. Peningkatan performa ini menunjukkan bahwa kombinasi augmentasi memberikan dampak positif yang signifikan terhadap model BERT, memungkinkan model menangkap lebih banyak variasi linguistik dan meningkatkan kemampuan generalisasi pada dataset yang berbeda. Eksperimen 3 ditunjukkan pada Tabel 8.

Tabel 8. Eksperimen 3

Dataset	Tanpa Augmentasi
Coursera	0.9190

Tabel 8 eksperimen 3, yaitu saat tidak menggunakan augmentasi sama sekali, performa BERT mengalami penurunan yang cukup nyata, dengan *F1-Score* yang diperoleh 0.9190 pada dataset *Coursera*. Penurunan ini menunjukkan bahwa meskipun BERT cukup andal, kehadiran augmentasi tetap krusial untuk meningkatkan performa.

Hasil menunjukan bahwa model BERT terbukti sebagai model yang paling konsisten, akurat, dan responsif terhadap teknik augmentasi, menjadikannya pilihan utama dalam tugas klasifikasi sentimen, terutama ketika augmentasi diterapkan secara optimal. Hasil eksperimen dengan kombinasi teknik augmentasi (*Synonym Replacement*, *Contextual Word Embeddings*, dan *Back Translation*) pada dataset *Coursera* menunjukkan bahwa **BERT** secara keseluruhan memiliki performa terbaik di semua dataset, terutama dengan hasil *F1-Score* hampir sempurna **0.9983** Pada dataset ini, penggunaan metode augmentasi berhasil membantu model mempertahankan performa yang tinggi dengan kemampuan generalisasi yang lebih baik pada *Coursera* yang ditunjukkan pada Gambar 4.



Gambar 4. *Confusion Matrix Coursera*-Kombinasi

Berdasarkan analisis performa model yang telah dijelaskan, hasil *confusion matrix* dari dataset *Coursera* mendukung pernyataan bahwa model memiliki kinerja yang sangat tinggi. Dengan metrik seperti akurasi, *F1-Score*, precision, dan *Recall* yang mencapai nilai yang sama 0.9983, *confusion matrix* memperlihatkan bahwa hampir seluruh prediksi dilakukan dengan sangat akurat. Ditunjukkan pada Gambar 3.2 model mampu memprediksi 2136 dari 2140 data kelas negatif dengan benar, hanya ada 4 kesalahan prediksi pada kelas netral dan tidak ada kesalahan untuk kelas positif. Hal ini menunjukkan bahwa model mampu menangkap pola dengan sangat baik pada dataset yang terstruktur seperti *Coursera*.

5. KESIMPULAN DAN SARAN

Penelitian yang dilakukan, dapat disimpulkan model BERT menunjukkan performa paling unggul dalam tugas klasifikasi sentimen, pada dataset *Coursera* dengan kombinasi teknik augmentasi seperti *Back Translation, Contextual Word Embeddings* dan *Synonym Replacement*. Arsitektur BERT yang kompleks dan kemampuannya dalam memahami konteks secara mendalam menjadikannya lebih unggul dibandingkan model lain. Hasil terbaik diperoleh dari model BERT dengan kombinasi teknik augmentasi pada dataset *Coursera*, yang menghasilkan akurasi tertinggi sebesar 99,83%. Hal ini menunjukkan bahwa penggabungan berbagai teknik augmentasi mampu meningkatkan keberagaman dan kualitas data, sehingga menghasilkan model yang lebih akurat dan robust dalam menghadapi data tidak seimbang maupun beragam.

Penelitian selanjutnya, penggunaan teknik augmentasi *Back Translation* sangat direkomendasikan karena mampu memperkaya data tanpa mengubah makna inti kalimat. Selain itu, untuk peningkatan performa lebih lanjut, disarankan untuk mengeksplorasi model NLP lain yang lebih canggih seperti RoBERTa, XLNet, atau menerapkan pendekatan ensemble model, yang berpotensi menghasilkan klasifikasi yang lebih stabil dan akurat pada berbagai jenis data.

DAFTAR PUSTAKA

[1] T. V. Ngoc, M. N. Thi, and H. N. Thi, "Sentiment Analysis of Students' Reviews on Online Courses: A Transfer Learning Method," *Proc. Int. Conf. Ind. Eng. Oper. Manag.*, pp. 306–314, 2021, doi: 10.46254/ap01.20210122.

[2] R. M. Turjaman and I. Budi, "Analisis Sentimen Berbasis Aspek Marketing Mix Terhadap Ulasan Aplikasi Dompet Digital (Studi Kasus: Aplikasi Linkaja Pada Twitter)," *J. Darma Agung*, vol. 30, no. 2, p. 266, 2022, doi: 10.46930/ojsuda.v30i2.1672.

[3] C. Mishra, E. Bansal, A. H. Victoria, C. Mishra, and A. H. Victoria, "Tweet Sentiment Extraction," 2022.

[4] C. Suhaeni and H. S. Yong, "Mitigating Class Imbalance in Sentiment Analysis through GPT-3-Generated Synthetic Sentences," *Appl. Sci.*, vol. 13, no. 17, 2023, doi: 10.3390/app13179766.

[5] H. Q. Abonizio, E. C. Paraiso, and S. Barbon, "Toward Text Data Augmentation for Sentiment Analysis," *IEEE Trans. Artif. Intell.*, vol. 3, no. 5, pp. 657–668, 2022, doi: 10.1109/TAI.2021.3114390.

[6] A. R. Nair, R. P. Singh, D. Gupta, and P. Kumar, "Evaluating the Impact of Text Data Augmentation on Text Classification Tasks using DistilBERT," *Procedia Comput. Sci.*, vol. 235, no. 2023, pp. 102–111, 2024, doi: 10.1016/j.procs.2024.04.013.

[7] Y. A. Singgalen, "Penerapan Metode TOPSIS Sebagai Pendukung Keputusan Pemilihan Layanan Akomodasi di Destinasi Wisata Pulau," *J. Media Inform. Budidarma*, vol. 7, no. 3, p. 1386, 2023, doi: 10.30865/mib.v7i3.6530.

[8] S. S. Ahmed, P. J. Uppalapati, S. Ayesha, S. M. Hussain, K. Narasimharao, and N. Silpa, "Assessing Public Sentiment towards Digital India through Twitter Sentiment Analysis: A Comparative Study," *Proc. 7th Int. Conf. Intell. Comput. Control Syst. ICICCS 2023*, pp. 955–959, 2023, doi: 10.1109/ICICCS56967.2023.10142882.

[9] D. Fimoza, A. Amalia, and T. Henny Febriana Harumy, "Sentiment Analysis for Movie Review in Bahasa Indonesia Using BERT," *2021 Int. Conf. Data Sci. Artif. Intell. Bus. Anal. DATABIA 2021 - Proc.*, pp. 27–34, 2021, doi: 10.1109/DATABIA53375.2021.9650096.

[10] J. D. S. Baguio, B. A. Lu, and C. F. Pena, "Text Classification of Climate Change Tweets using Artificial Neural Networks, FastText Word Embeddings, and Latent Dirichlet Allocation," *2023 Int. Conf. Adv. Power, Signal, Inf. Technol. APSIT 2023*, pp. 688–692, 2023, doi: 10.1109/APSIT58554.2023.10201782.

[11] M. S. BAŞARSLAN and F. KAYAALP, "Sentiment Analysis on Social Media Reviews Datasets with Deep Learning Approach," *Sak. Univ. J. Comput. Inf. Sci.*, vol. 4, no. 1, pp. 35–49, 2021, doi: 10.35377/saucis.04.01.833026.

[12] Y. Wu and J. He, "Sentiment analysis of barrage text based on ALBERT-ATT-BiLSTM model," *2021 4th Int. Conf. Pattern Recognit. Artif. Intell. PRAI 2021*, pp. 152–156, 2021, doi: 10.1109/PRAI53619.2021.9551040.

[13] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, no. Mlm, pp. 4171–4186, 2019.

[14] G. Ananthakrishnan, A. K. Jayaraman, T. E. Trueman, S. Mitra, A. K. Abinesh, and A. Murugappan, "Suicidal Intention Detection in Tweets Using BERT-Based Transformers," *3rd IEEE 2022 Int. Conf. Comput. Commun. Intell. Syst. ICCCIS 2022*, pp. 322–327, 2022, doi: 10.1109/ICCCIS56430.2022.10037677.

[15] D. Sebastian, H. D. Purnomo, and I. Sembiring, "BERT for Natural Language Processing in Bahasa Indonesia," *2022 2nd Int. Conf. Intell. Cybern. Technol. Appl. ICICyTA 2022*, pp. 204–209, 2022, doi: 10.1109/ICICyTA57421.2022.10038230.