ANALISIS SENTIMEN ULASAN PENGGUNA PLATFORM BELAJAR COURSERA MENGGUNAKAN MODEL DISTILBERT DENGAN AUGMENTASI DATA DAN PENANGANAN KETIDAKSEIMBANGAN DATA

Fadlillah Mukti Ayudewi, Arridho Ramadhan Firdaus
Teknologi Informasi, Universitas 'Aisyiyah Yogyakarta
Jl Ringroad Barat, Sleman, Daerah Istimewa Yogyakarta
fadlillahmuktiayudewi@unisayogya.ac.id

ABSTRAK

Klasifikasi sentimen terhadap ulasan pengguna berperan penting dalam memahami persepsi publik terhadap layanan digital. Penelitian ini bertujuan menerapkan model DistilBERT untuk mengklasifikasikan sentimen ulasan Coursera ke dalam tiga kategori: positif, netral, dan negatif. DistilBERT dipilih karena efisiensinya dalam memproses konteks linguistik yang kompleks tanpa mengorbankan akurasi. Tahapan penelitian mencakup praproses teks dan penerapan tiga teknik augmentasi data, yaitu: *Synonym Replacement, Contextual Word Embeddings, dan Back Translation* yang efektuf untuk menangani ketidakseimbangan kelas dan meningkatkan keragaman data latih. Evaluasi model menunjukkan bahwa penggunaan kombinasi teknik augmentasi berhasil meningkatkan akurasi secara signifikan, dengan hasil tertinggi mencapai 99,78%. Hasil ini menunjukkan bahwa augmentasi data mampu meningkatkan generalisasi dan kinerja model, bahkan pada data yang kompleks dan tidak seimbang. Penelitian ini membuktikan efektivitas integrasi teknik augmentasi dalam mendukung kinerja DistilBERT pada tugas klasifikasi sentimen. Untuk penelitian selanjutnya, disarankan mengeksplorasi model NLP lanjutan seperti XLNet, RoBERTa, atau pendekatan ensemble untuk hasil klasifikasi yang lebih stabil dan akurat.

Kata Kunci : *Data Tidak Seimbang, DistilBERT, Teknik Augmentasi*

1. PENDAHULUAN

Dalam era digital saat ini, platform pembelajaran daring seperti Coursera telah menjadi bagian penting dari pendidikan dan pengembangan karier. Platform ini menyediakan berbagai kursus yang diikuti oleh jutaan pengguna di seluruh dunia, dan ulasan pengguna menjadi sumber informasi yang sangat berguna. Ulasan ini tidak hanya membantu calon peserta kursus dalam membuat keputusan, tetapi juga memberikan umpan balik berharga bagi penyedia kursus dan pengembang platform. Analisis sentimen terhadap ulasan ini menjadi sangat penting untuk menggali opini pengguna dan meningkatkan kualitas layanan yang diberikan. Analisis sentimen memungkinkan kita untuk mengidentifikasi dan mengklasifikasikan opini atau perasaan yang terkandung dalam teks, seperti ulasan produk atau layanan [1]. Tantangan utama dalam analisis sentimen adalah menangani ketidakseimbangan data, di mana data dalam kelas mayoritas jauh lebih banyak dibandingkan dengan kelas minoritas, yang dapat mengurangi akurasi model dalam mendeteksi kelas minoritas [2]. Penggunaan teknik augmentasi data seperti Synonym Replacement, Contextual Word Embeddings, dan Back Translation sangat penting untuk meningkatkan keragaman data latih, yang dapat membantu model mengenali kelas minoritas dengan lebih baik [3]

Penelitian ini berfokus pada penggunaan model DistilBERT untuk klasifikasi sentimen pada ulasan pengguna Coursera, dengan tujuan mengatasi masalah ketidakseimbangan data yang sering terjadi dalam analisis sentimen. DistilBERT merupakan versi ringan dari model BERT yang dikembangkan melalui teknik distilasi pengetahuan (*knowledge distillation*). Dibandingkan dengan BERT yang memiliki 12 lapisan, DistilBERT hanya memiliki 6 lapisan, sehingga membutuhkan komputasi yang lebih ringan. Meskipun lebih ringkas, DistilBERT tetap mampu mempertahankan sekitar 97% akurasi BERT pada berbagai tugas NLP [4]. Dengan ukuran model yang lebih kecil, kecepatan pelatihan yang lebih tinggi, dan efisiensi komputasi yang lebih baik, DistilBERT sangat cocok diterapkan di lingkungan dengan sumber daya terbatas, seperti aplikasi dunia nyata yang memerlukan pemrosesan cepat dan efisien [5].

Permasalahan utama yang akan dijawab oleh penelitian ini adalah bagaimana meningkatkan akurasi model DistilBERT dalam mengklasifikasikan sentimen ulasan yang tidak seimbang, serta bagaimana teknik augmentasi dapat mempengaruhi performa model ini. Penelitian ini bertujuan untuk menganalisis pengaruh teknik augmentasi data terhadap performa model DistilBERT dalam mengklasifikasikan sentimen, serta membandingkan efektivitas teknik-teknik augmentasi seperti Synonym Replacement, Contextual Word Embeddings, dan Back Translation dalam meningkatkan performa model tersebut.

Untuk mencapai tujuan tersebut, penelitian ini menggunakan pendekatan eksperimen dengan model klasifikasi sentimen berbasis model DistilBERT. Dataset yang digunakan berasal dari ulasan pengguna di platform Coursera, yang bersumber dari Kaggle. Tahapan penelitian meliputi pengumpulan data, preprocessing, penerapan teknik augmentasi untuk menangani ketidakseimbangan data, dan evaluasi model menggunakan metrik akurasi, presisi, recall, dan F1-score.

Analisis sentimen adalah cabang dari text mining yang termasuk dalam bidang Natural Language Processing (NLP), dengan tujuan utama untuk mengidentifikasi dan menentukan perasaan atau opini yang terkandung dalam teks (Turjaman & Budi, 2022). Penelitian tentang BERT dan variasinya, seperti DistilBERT, telah banyak dilakukan untuk mengatasi masalah analisis sentimen pada dataset yang besar dan tidak seimbang. BERT, yang dikembangkan oleh [6], memiliki kemampuan untuk memahami konteks dua arah dalam teks, sehingga memberikan performa yang sangat baik dalam banyak tugas NLP. Namun, kekurangannya adalah dalam kebutuhan komputasi yang tinggi. Untuk mengatasi hal ini, DistilBERT, sebagai versi yang lebih efisien dari BERT, menawarkan pengurangan ukuran model tanpa mengorbankan kinerja [4]. Selain itu, teknik augmentasi data seperti Synonym Replacement, Contextual Word Embeddings, dan Back Translation telah terbukti efektif dalam menangani ketidakseimbangan data dan meningkatkan performa model dalam klasifikasi sentiment [3].

2. TINJAUAN PUSTAKA

DistilBERT adalah versi yang lebih ringan dan efisien dari BERT (Bidirectional Encoder Representations from Transformers), yang dikembangkan untuk mengurangi ukuran model dan mempercepat proses pelatihan, sambil tetap mempertahankan sebagian besar kemampuan

BERT dalam memahami konteks teks secara mendalam. DistilBERT mengurangi jumlah parameter model BERT hingga 40% dan mempercepat proses pelatihan hingga 60% [4]. Distilasi pengetahuan, yang digunakan dalam pembuatan DistilBERT, memungkinkan model untuk mempelajari informasi yang relevan dari model yang lebih besar (BERT) tanpa mengorbankan akurasi. Teknik ini mengurangi kompleksitas komputasi, menjadikan DistilBERT lebih efisien untuk digunakan dalam aplikasi dunia nyata yang memerlukan kecepatan dan efisiensi sumber daya [7].

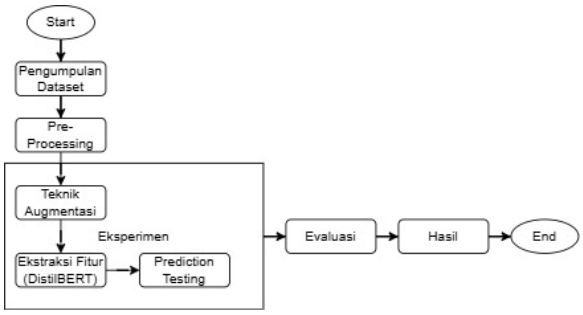
Arsitektur DistilBERT serupa dengan BERT, tetapi dengan beberapa perbedaan utama. DistilBERT menggunakan hanya 6 lapisan transformer, sedangkan BERT menggunakan 12 lapisan, sehingga mengurangi jumlah lapisan encoder dan mempercepat proses inferensi. Meskipun demikian, DistilBERT tetap mempertahankan kekuatan dalam memahami konteks dua arah [6]. Dalam tugas analisis sentimen, DistilBERT sering menunjukkan performa yang hampir setara dengan BERT namun dengan efisiensi yang lebih tinggi, seperti yang terlihat pada eksperimen yang dilakukan oleh [4], di mana DistilBERT menunjukkan kinerja yang sebanding pada dataset SQuAD v2 dibandingkan dengan BERT, tetapi dengan waktu pelatihan yang lebih singkat.

Salah satu keuntungan utama dari DistilBERT adalah kemampuannya untuk tetap mempertahankan performa tinggi dalam berbagai tugas NLP, termasuk analisis sentimen, sambil mengurangi penggunaan memori dan sumber daya komputasi [8]. Selain itu, teknik-teknik augmentasi data, seperti Synonym Replacement, Contextual Word Embeddings, dan Back Translation, dapat diterapkan pada DistilBERT untuk menangani masalah ketidakseimbangan data, yang terbukti meningkatkan performa model dalam tugas klasifikasi sentiment [3]. Penelitian oleh Fajri et al (2022) menunjukkan bahwa DistilBERT memberikan hasil akurasi yang lebih tinggi dibandingkan BERT pada beberapa dataset, termasuk dalam analisis sentimen pada tweet terkait Covid-19, dengan akurasi mencapai 97% [9].

Secara keseluruhan, DistilBERT menawarkan keseimbangan yang baik antara efisiensi komputasi dan akurasi tinggi, menjadikannya pilihan yang sangat baik untuk aplikasi yang memprioritaskan efisiensi dalam pengolahan teks besar, seperti analisis sentimen pada ulasan pengguna platform daring.

3. METODE PENELITIAN

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



Gambar 1. Alur Penelitian

Gambar 1. Menunjukan alur penelitian yang pertama dilakukan yaitu, pengumpulan dataset untuk dataset, selanjutnya pre-processing yaitu melakukan pengolahan kalimat, setelahnya teknik augmentasi yang bertujuan untuk merubah kalimat menjadi lebih beragam, ekstrasi fitur dengan model DistilBERT setelah melalui model ada prediction testing yang bertujuan untuk menghasilkan klasifikasi dari model yang digunakan, selanjutnya evaluasi dan terakhir hasilnya.

3.1. Pengumpulan Dataset

Pengumpulan Dataset dalam penelitian ini mencakup *Coursera.* 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-langkah dasar namun sangat penting dalam menyiapkan data untuk analisis sentimen [10]. *Case Folding* bertujuan unuk mengubah huruf besar menjadi huruf kecil dalam teks, sehingga membantu menjaga konsistensi data [11]. Remove Punctuation berfungsi untuk menghapus tanda baca dari teks, yang sangat tepat karena dalam analisis sentimen, simbol atau emoji mungkin tidak memberikan informasi yang relevan karena simbol atau emoji dalam analisis sentiment sering kali tidak memberikan informasi yang berguna, dan pemrosesan ini akan membantu model fokus pada kata dan frasa [12]. Tokenisasi dan *label encoding* akan bertujuan untuk mempermudah pemrosesan teks dan mengurangi potensi kesalahan saat model mempelajari data [13].

Tahapan *pre-processing* selesai Langkah berikutnya adalah memproses dataset pada tahap augmentasi. Hasil dari tahapan *pre-processing* disajikan pada Tabel 1.

Tabel 1. Hasil 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']

Tabel 1. Menunjukan hasil pre-processing dari kalimat yang ada pada dataset coursera, untuk prosesnya mulai dari data awal original masuk, selanjutnya di case folding yang merubah kalimat menjadi huruf kecil, remove punctuation yang berfungsi menghapus tanda baca, tokenization yang merubah kalimat menjadi sub token agar mudah dibaca oleh model nantinya, dan data akhir merupakan hasil dari serangkaian proses yang dilakukan saat pre-processing data.

3.3. Teknik Augmentasi

Penelitian ini mengaplikasikan teknik augmentasi data berupa *Synonym Replacement, Contextual Word Embeddings* dan *Back Translation* yang bertujuan untuk menyeimbangkan data. Teknik *Synonym Replacement* sangat efektif untuk menambahkan variasi teks tanpa mengubah maknanya, yang penting dalam menangani data yang tidak seimbang [14], terutama dalam kumpulan data dengan lebih banyak ulasan positif daripada ulasan negatif atau netral. *Contextual Word Embeddings* memungkinkan model untuk memahami konteks yang lebih mendalam dari setiap ulasan [15] sementara *Back Translation* memberikan variasi frasa baru sambil mempertahankan makna aslinya [16], sehingga memperkaya keragaman data tanpa kehilangan informasi. Penerapan ketiga teknik penambahan ini akan memperkaya kumpulan data tanpa 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*

Tabel 2. Menunjukan untuk kolom kalimat asli yang kata diberi tanda tebal(Bold) merupakan kata yang digunakan untuk augmentasi sehingga pada kolom hasil juga menunjukan kata yang diberi tanda tebal(bold) itu merupakan hasil augmentasi dari tanda kata yang diberi tebal(bold) pada kalimat asli, sehingga berdasarkan hasilnya terdapat keragaman bahasa yang lebih banyak.

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 DistilBERT. Skenario pengujian ditunjukan pada Tabel 3.

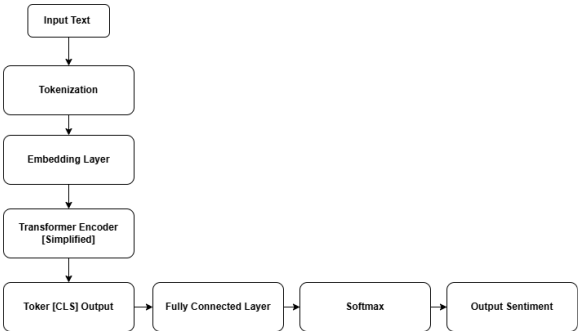
Tabel 3. Eksperimen

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

Tabel 3. Menunjukan skenario eksperimen yang akan dilakukan penelitian ini, mulai dari eksperimen 1, dimana model DistilBERT akan diuji dengan masing-masing teknik augmentasi, eksperimen 2, model diuji dengan tiga teknik augmentasi sekaligus, selanjutnya eksperimen 3, model diuji menggunakan dataset asli tanpa augmentasi.

3.5. Ekstraksi Fitur dan Model

Ekstraksi fitur menggunakan DistilBERT sangat cocok untuk masalah analisis sentimen, karena DistilBERT ini terbukti efektif dalam memahami data teks yang kompleks serta menangkap konteks [17]. Model menawarkan kelebihannya sendiri dengan kemampuannya memahami konteks dua arah [18]. Dengan memanfaatkan model ini, studi ini berpotensi memperoleh hasil optimal untuk klasifikasi sentimen dari kumpulan data yang ada. DistilBERT *(Distilled Bidirectional Encoder Representations from Transformers)* menggunakan pendekatan transformer untuk ekstraksi fitur dalam teks [19]. Berikut rancangan *Step By Step* model DistilBERT ditunjukkan pada Gambar 2.



Gambar 2. Rancangan *step by step* model DistilBERT

Gambar 2. menunjukan rancangan proses analisis sentimen menggunakan model DistilBERT 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 *Simplified* (Disederhanakan) disbanding BERT yaitu berjumlah 6 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 [20]. 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 [13]. 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 DistilBERT dapat memberikan hasil yang baik. Penjelasan mengenai metrik evaluasi beserta persamaan ditunjukkan pada Tabel 4.

Tabel 4. Metrik Evaluasi

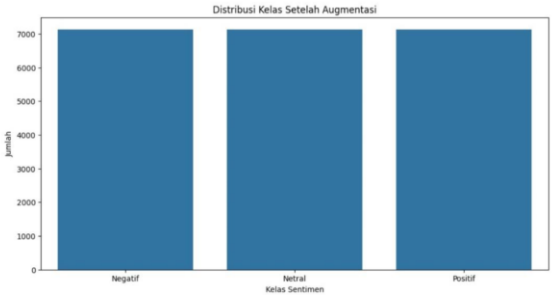
Metrik	Deskripsi	Rumus
Akurasi	Menggambarkan seberapa akurat model dapat mengklasifikasikan dengan benar	$\frac{TP+TN}{TP+FP+FN+TN}$
Precision	Menggambarkan prediksi antara data yang diminta	$\frac{TP}{TP+FP}$

Metrik	Deskripsi	Rumus
	dengan hasil dari prediksi yang diberikan oleh model.	
Re-Call	Menggambarkan seberapa banyak kelas psoitif yang diperdiksi dengan benar.	$\frac{TP}{TP+FN}$
F1-Score	Rata-rata precision dan Recall.	$\frac{2*(Precision*Recall)}{Precison+Recall}$

4. HASIL DAN PEMBAHASAN

4.1. Distribusi Dataset dan Proses Augmentasi

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

Gambar 3. Menunjukan hasil bahwa dataset yang tidak seimbang setelah diagmentasi menjadi seimbang dengan masing-masing kelas berjumlah 7.134 data ulasan coursera. 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%

Tabel 5. Menunjukan data setelah dibagi dengan seimbang sesuai masing-masing kelas.

4.2. Eksperimen 1 Teknik Augmentasi

Eksperimen pertama menggunakan model DistilBERT yang diterapkan pada tiga teknik augmentasi yang berbeda secara terpisah: *Synonym Replacement, Contextual Word Embeddings, dan Back Translation*. Hasil eksperimen ini ditunjukkan pada Tabel 6.

Tabel 6. Eksperimen 1

Dataset	*Synonym Replacement*	*Contextual Word Embeddings*	Back Translation
Coursera	0.9869	0.9696	0.9678

Tabel 6. eksperimen 1, menguji tiga teknik augmentasi secara terpisah, DistilBERT mencatat performa terbaik pada teknik Synonym Replacement dengan akurasi sebesar 0.9869, yang menunjukkan bahwa model ini sangat efektif dalam memahami konteks meskipun inputnya telah mengalami perubahan struktur kalimat. Hal ini membuktikan kemampuan representasi kontekstual DistilBERT yang kuat.

4.3. Eksperimen 2 Kombinasi Teknik Augmentasi

Eksperimen kedua menggunakan kombinasi ketiga teknik augmentasi secara bersamaan. Hasilnya disajikan pada Tabel 7.

Tabel 7. Eksperimen 2

Dataset	Kombinasi
Coursera	0.9978

Tabel 7. menunjukan hasil eksperimen 2, di mana ketiga teknik augmentasi digabungkan, DistilBERT berhasil mencapai akurasi tertinggi sebesar 0.9978. Peningkatan performa ini menunjukkan bahwa kombinasi augmentasi memberikan dampak positif yang signifikan terhadap model DistilBERT.

4.4. Eksperimen 3 Tanpa Teknik Augmentasi

Eksperimen ketiga dilakukan tanpa teknik augmentasi apa pun. Hasilnya disajikan pada Tabel 8.

Tabel 8. Eksperimen 3

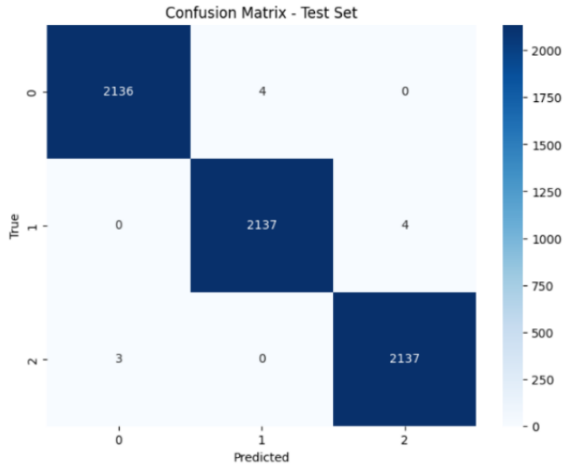
Dataset	Tanpa Augmentasi
Coursera	0.9140

Tabel 8. Menunjukan hasil eksperimen 3, yaitu saat tidak menggunakan augmentasi sama sekali, performa DistilBERT mengalami penurunan yang cukup nyata, meskipun masih tetap unggul dibandingkan model lain. akurasi yang didapatkan 0.9140. Penurunan ini menunjukkan bahwa meskipun DistilBERT cukup andal, kehadiran augmentasi tetap dibutuhkan untuk meningkatkan akurasi.

4.5. Evaluasi Model dengan Confusion Matrix

Hasil menunjukan bahwa model DistilBERT 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) menunjukkan bahwa DistilBERT secara keseluruhan memiliki performa terbaik di semua dataset, terutama dengan hasil akurasi hampir sempurna 0.9978 pada

hasil dari eksperimen 2 yang menggunakan tiga teknik augmentasi atau kombinasi dan langsung digunakan dalam satu model yaitu DistilBERT, penggunaan metode augmentasi berhasil membantu model mempertahankan performa yang tinggi dengan kemampuan generalisasi yang lebih baik yang ditunjukkan pada Gambar 4.



Gambar 4. *Confusion Matrix Coursera*-Kombinasi

Gambar 4. Menunjukan hasil confusion matrix bahwa model memiliki kinerja yang sangat tinggi pada eskperimen 2. Dengan metrik seperti akurasi, F1-Score, precision, dan Recall yang mencapai nilai yang sama 0.9978, confusion matrix memperlihatkan bahwa hampir seluruh prediksi dilakukan dengan sangat akurat.

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

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa model DistilBERT menunjukkan performa paling unggul dalam tugas klasifikasi sentimen, khususnya pada dataset Coursera, dengan akurasi tertinggi mencapai 99,78% ketika digunakan bersama kombinasi teknik augmentasi seperti *Back Translation, Contextual Word Embeddings, dan Synonym Replacement.* Hasil ini menunjukkan bahwa penggabungan teknik augmentasi mampu meningkatkan keberagaman dan kualitas data, sehingga menghasilkan model yang lebih akurat dan andal dalam menangani data yang tidak seimbang maupun bervariasi. Untuk penelitian selanjutnya, disarankan untuk mengeksplorasi model NLP yang lebih canggih seperti RoBERTa atau XLNet, serta mempertimbangkan pendekatan ensemble model guna memperoleh performa klasifikasi yang lebih stabil dan optimal pada berbagai jenis data.

DAFTAR PUSTAKA

[1] Abonizio, H.Q., Paraiso, E.C. & Barbon, S., 2022, Toward Text Data Augmentation for Sentiment Analysis, *IEEE Transactions on Artificial Intelligence*, 3, 5, 657–668.

[2] Adel, H., Dahou, A., Mabrouk, A., Elaziz, M.A., Kayed, M., El-Henawy, I.M., Alshathri, S. & Ali, A.A., 2022, Improving Crisis Events Detection Using DistilBERT with Hunger Games Search Algorithm, Mathematics, 10, 3, 1–22.

[3] Ahmed, S.S., Uppalapati, P.J., Ayesha, S., Hussain, S.M., Narasimharao, K. & Silpa, N., 2023, Assessing Public Sentiment towards Digital India through Twitter Sentiment Analysis: A Comparative Study, Proceedings of the 7th International Conference on Intelligent Computing and Control Systems, ICICCS 2023, 955–959.

[4] Amal, M.I., Rahmasita, E.S., Suryaputra, E. & Rakhmawati, N.A., 2022, Analisis Klasifikasi Sentimen Terhadap Isu Kebocoran Data Kartu Identitas Ponsel di Twitter, Jurnal Teknik Informatika dan Sistem Informasi, 8, 3, 645–660.

[5] Ananthakrishnan, G., Jayaraman, A.K., Trueman, T.E., Mitra, S., Abinesh, A.K. & Murugappan, A., 2022, Suicidal Intention Detection in Tweets Using BERT-Based Transformers, 3rd IEEE 2022 International Conference on Computing, Communication, and Intelligent Systems, ICCCIS 2022, 322–327.

[6] Baguio, J.D.S., Lu, B.A. & Pena, C.F., 2023, Text Classification of Climate Change Tweets using Artificial Neural Networks, FastText Word Embeddings, and Latent Dirichlet Allocation, 2023 International Conference in Advances in Power, Signal, and Information Technology, APSIT 2023, 688–692.

[7] BAŞARSLAN, M.S. & KAYAALP, F., 2021, Sentiment Analysis on Social Media Reviews Datasets with Deep Learning Approach, Sakarya University Journal of Computer and Information Sciences, 4, 1, 35–49.

[8] Basiri, M.E., Nemati, S., Abdar, M., Asadi, S. & Acharrya, U.R., 2021, A novel fusion-based deep learning model for sentiment analysis of COVID-19 tweets, Knowledge-Based Systems, 228, 107242. https://doi.org/10.1016/j.knosys.2021.107242,.

[9] Devlin, J., Chang, M.W., Lee, K. & Toutanova, K., 2019, BERT: Pre-training of deep bidirectional transformers for language understanding, NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, 1, Mlm, 4171–4186.

[10] Fajri, F., Tutuko, B. & Sukemi, S., 2022, Membandingkan Nilai Akurasi BERT dan DistilBERT pada Dataset Twitter, JUSIFO (Jurnal Sistem Informasi), 8, 2, 71–80.

[11] Fimoza, D., Amalia, A. & Henny Febriana Harumy, T., 2021, Sentiment Analysis for Movie Review in Bahasa Indonesia Using BERT, 2021 International Conference on Data Science, Artificial Intelligence, and Business Analytics, DATABIA 2021 - Proceedings, 27–34.

[12] Nair, A.R., Singh, R.P., Gupta, D. & Kumar, P., 2024, Evaluating the Impact of Text Data Augmentation on Text Classification Tasks using DistilBERT, Procedia Computer Science, 235, 2023, 102–111. https://doi.org/10.1016/j.procs.2024.04.013,.

[13] Ngoc, T.V., Thi, M.N. & Thi, H.N., 2021, Sentiment Analysis of Students' Reviews on Online Courses: A Transfer Learning Method, Proceedings of the International Conference on Industrial Engineering and Operations Management, 306–314.

[14] Sanh, V., Debut, L., Chaumond, J. & Wolf, T., 2019, DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter, , 2–6. http://arxiv.org/abs/1910.01108,.

[15] Sebastian, D., Purnomo, H.D. & Sembiring, I., 2022, BERT for Natural Language Processing in Bahasa Indonesia, 2022 2nd International Conference on Intelligent Cybernetics Technology and Applications, ICICyTA 2022, 204–209.

[16] Turjaman, R.M. & Budi, I., 2022, Analisis Sentimen Berbasis Aspek Marketing Mix Terhadap Ulasan Aplikasi Dompet Digital (Studi Kasus: Aplikasi Linkaja Pada Twitter), Jurnal Darma Agung, 30, 2, 266.

[17] Wu, Y. & He, J., 2021, Sentiment analysis of barrage text based on ALBERT-ATT-BiLSTM model, 2021 4th International Conference on Pattern Recognition and Artificial Intelligence, PRAI 2021, 152–156.