ISSN (Online) 2747-0563
Volume 6 Tahun 2026

Perbandingan BERT-base dan DistilBERT untuk Ekstraksi Aspek dan Klasifikasi Sentimen

Muhammad Bayu Nashrullah[1], Rizky Parlika[2*], Budi Mukhamad Mulyo[3]

[1,2,3] Informatika, Universitas Pembangunan Nasional "Veteran" Jawa Timur
[1]21081010042@student.upnjatim.ac.id
[3]budi.m.mulyo.fasilkom@upnjatim.ac.id
*Corresponding author email: rizkyparlika.if@upnjatim.ac.id

Abstrak*— Aspect-Based Sentiment Analysis (ABSA) merupakan pendekatan analisis sentimen yang bertujuan untuk mengidentifikasi aspek tertentu dari suatu ulasan serta menentukan polaritas sentimen terhadap aspek tersebut. Penelitian ini bertujuan untuk membandingkan kinerja dua model berbasis transformer, yaitu BERT-base dan DistilBERT, dalam tugas Aspect Term Extraction (ATE) dan Aspect Sentiment Classification (ASC) pada dataset ulasan restoran. Dataset yang digunakan terdiri dari total 1006 kalimat, yang kemudian diproses untuk mengekstraksi aspek dan mengklasifikasikan sentimen. Hasil eksperimen menunjukkan bahwa DistilBERT memberikan performa yang sedikit lebih baik dibandingkan BERT-base pada kedua tugas. Pada ATE, DistilBERT memperoleh token F1-score sebesar 0,794, sedangkan BERT-base memperoleh 0,781. Pada ASC, DistilBERT mencapai akurasi sebesar 0,892 dan weighted F1-score sebesar 0,874, lebih tinggi dibandingkan BERT-base dengan akurasi 0,886 dan weighted F1-score 0,856. Selain itu, DistilBERT juga menunjukkan efisiensi waktu pelatihan yang jauh lebih baik dibandingkan BERT-base. Analisis confusion matrix menunjukkan bahwa kedua model lebih akurat dalam memprediksi sentimen positif, sementara performa pada kelas netral masih rendah akibat ketidakseimbangan distribusi data. Hasil penelitian ini menunjukkan bahwa DistilBERT dapat menjadi alternatif yang lebih efisien untuk penerapan ABSA dengan performa yang tetap kompetitif.

Kata Kunci— **Sentiment Analysis, Aspect-Based Sentiment Analysis, BERT, DistilBERT.**

I. PENDAHULUAN

Perkembangan teknologi digital mendorong meningkatnya jumlah data teks yang berisi opini dan pendapat pengguna, terutama dalam bentuk ulasan, komentar, maupun tanggapan terhadap suatu produk, layanan, atau fenomena tertentu. Data opini tersebut menyimpan informasi yang sangat berharga karena merefleksikan persepsi, kepuasan, serta kritik dari pengguna secara langsung [1]. Namun, besarnya volume data teks membuat proses analisis secara manual menjadi tidak efisien dan sulit dilakukan, sehingga diperlukan pendekatan otomatis untuk mengekstraksi informasi penting dari teks opini tersebut [2], [3].

Analisis sentimen menjadi salah satu pendekatan yang banyak digunakan untuk memahami kecenderungan opini dalam data teks. Pendekatan ini bertujuan mengklasifikasikan sentimen suatu teks ke suatu kategori yang telah ditentukan, seperti positif atau negatif [4]. Meskipun demikian, analisis sentimen pada tingkat dokumen atau kalimat sering kali bersifat terlalu umum dan kurang mampu menangkap detail informasi yang lebih spesifik. Dalam banyak kasus, satu teks dapat memuat lebih dari satu opini dengan sentimen yang berbeda-beda, tergantung pada aspek yang dibahas [5]. Kondisi ini menyebabkan hasil analisis sentimen secara keseluruhan menjadi kurang representatif terhadap isi teks yang sebenarnya. Untuk mengatasi keterbatasan tersebut, pendekatan *Aspect-Based Sentiment Analysis* (ABSA) diperkenalkan. ABSA memungkinkan analisis sentimen dilakukan pada tingkat aspek, sehingga sistem tidak hanya menentukan polaritas sentimen, tetapi juga mengidentifikasi aspek atau atribut tertentu yang menjadi fokus opini [6]. Dalam ABSA, terdapat dua tugas utama yang saling berkaitan, yaitu ekstraksi aspek (*aspect extraction*) dan klasifikasi sentimen terhadap aspek tersebut (*aspect sentiment classification*). Ekstraksi aspek bertujuan mengidentifikasi istilah atau frasa dalam teks yang merepresentasikan aspek tertentu, sedangkan klasifikasi sentimen bertugas menentukan polaritas sentimen yang terkait dengan setiap aspek yang telah ditemukan. Dengan pendekatan ini, hasil analisis menjadi lebih rinci dan informatif dibandingkan analisis sentimen konvensional [7], [8].

Seiring dengan perkembangan bidang pemrosesan bahasa alami, pendekatan berbasis *deep learning* semakin banyak digunakan untuk menyelesaikan tugas-tugas ABSA. Model-model dengan basis *transformer*, khususnya BERT (*Bidirectional Encoder Representations from Transformers*), telah menunjukkan kinerja yang sangat baik dalam berbagai tugas NLP, termasuk ekstraksi aspek dan klasifikasi sentiment [9]. Model BERT dasar, yaitu BERT-base memiliki keunggulan yang terletak pada kemampuannya memahami konteks dua arah, sehingga representasi kata yang dihasilkan menjadi lebih kaya dan kontekstual [10]. Hal ini menjadikan BERT-base sebagai salah satu model yang paling populer dan banyak dijadikan acuan dalam penelitian ABSA [10], [11].

Meskipun BERT-base menawarkan performa yang tinggi, model ini memiliki ukuran dan kompleksitas yang relatif besar, sehingga membutuhkan sumber daya komputasi yang cukup tinggi dalam proses pelatihan dan inferensi [12]. Kondisi ini mendorong munculnya varian model yang lebih ringan, salah satunya adalah DistilBERT. DistilBERT merupakan hasil proses *distillation* dari BERT yang dirancang untuk mempertahankan sebagian besar kemampuan BERT dengan jumlah parameter yang lebih sedikit dan waktu komputasi yang lebih efisien. Dengan karakteristik tersebut, DistilBERT

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

menjadi alternatif menarik, terutama pada skenario yang memiliki keterbatasan sumber daya [13], [14].

Dalam berbagai penelitian terkait ABSA, BERT-base dan DistilBERT telah banyak digunakan sebagai model dasar untuk menyelesaikan tugas ekstraksi aspek maupun klasifikasi sentimen. Keduanya menunjukkan performa yang kompetitif dan sering dijadikan acuan dalam eksperimen berbasis transformer [15]. Perbedaan karakteristik antara BERT-base yang lebih kompleks dan DistilBERT yang lebih ringkas menimbulkan perbedaan terkait bagaimana kedua model tersebut bekerja pada tugas-tugas ABSA yang memiliki tingkat kesulitan dan kebutuhan konteks yang berbeda. Oleh karena itu, analisis perbandingan kinerja kedua model menjadi relevan untuk memahami pola performa yang dihasilkan serta implikasinya terhadap penggunaan model dalam praktik.

Berdasarkan pertimbangan tersebut, penelitian ini melakukan studi perbandingan antara BERT-base dan DistilBERT pada dua tugas utama dalam Aspect-Based Sentiment Analysis, yaitu ekstraksi aspek dan klasifikasi sentimen. Evaluasi dilakukan menggunakan dataset ulasan restoran yang dianotasi oleh Naver Labs Europe dengan skema eksperimen yang konsisten pada kedua model, sehingga perbedaan hasil yang diperoleh dapat dianalisis secara lebih objektif. Penelitian ini tidak bertujuan untuk menghasilkan kesimpulan universal, melainkan menganalisis perilaku BERT-base dan DistilBERT pada skenario dataset berukuran terbatas. Hasil dari penelitian ini diharapkan dapat memberi *insight* yang lebih jelas mengenai karakteristik performa masing-masing model terhadap data yang berkarakter serta menjadi referensi dalam memilih model yang sesuai untuk tugas ekstraksi aspek dan klasifikasi sentimen [16].

II. TINJAUAN PUSTAKA

A. *Sentiment Analysis*

Sentiment analysis atau analisis sentimen merupakan teknik yang bertujuan untuk mengelompokkan opini atau sikap yang dituliskan dalam bentuk teks. Dengan menganalisis dokumen tertulis, seperti contohnya ulasan produk, komentar media sosial, atau umpan balik pelanggan, analisis sentimen dapat menentukan apakah sentimen yang terkandung dalam teks tersebut positif, negatif, atau netral. Proses ini dapat membantu untuk memahami opini publik terkait topik tertentu, memantau reputasi suatu merek, atau mengukur tingkat kepuasan pelanggan [17].

Menurut [18], analisis sentimen dapat dilakukan dalam beberapa tingkatan. Pada tingkat dokumen, seluruh isi dokumen dianalisis untuk menentukan satu polaritas sentimen yang mewakili keseluruhan dokumen, namun pendekatan ini kurang akurat ketika dokumen mengandung lebih dari satu jenis sentimen. Pada tingkat kalimat, setiap kalimat dianalisis secara terpisah sehingga variasi sentimen dalam satu dokumen dapat diidentifikasi dengan lebih baik. Selanjutnya, pada tingkat frasa, analisis difokuskan pada frasa-frasa yang mengandung opini dalam satu kalimat sehingga perbedaan sentimen dapat dikenali secara lebih rinci. Tingkatan yang paling detail adalah analisis sentimen berbasis aspek, di mana analisis difokuskan pada aspek-aspek tertentu dalam kalimat dan setiap aspek diberikan polaritas sentimen masing-masing, yang dikenal sebagai *aspect-based sentiment analysis*.

B. *Aspect-Based Sentiment Analysis*

Aspect-Based Sentiment Analysis (ABSA) adalah bentuk *advanced* dari *sentiment analysis* biasa yang tidak hanya menentukan polaritas keseluruhan teks, tapi juga menggali opini pengguna pada tingkat aspek tertentu dari suatu entitas, produk, atau layanan. Dengan kata lain, ABSA memberikan analisis sentimen yang lebih granular karena setiap ulasan dapat memiliki banyak aspek dengan polaritas yang berbeda-beda [19], [20].

ABSA bertujuan mengidentifikasi aspek yang disebutkan dalam sebuah teks, baik eksplisit maupun implisit, lalu mengaitkan setiap aspek tersebut dengan opini atau sentimen yang sesuai. Misalnya, dalam ulasan hotel, ABSA dapat membedakan opini pengguna mengenai kamar, layanan, atau lokasi, dan menentukan apakah masing-masing aspek dinilai positif, negatif, atau netral [21].

Secara umum, ABSA mencakup beberapa sub-tugas utama, yaitu *Aspect Term Extraction* (ATE) yang bertujuan mengidentifikasi aspek dalam teks, *Opinion Term Extraction* (OE) untuk mengenali kata atau frasa yang mengandung opini, serta *Aspect Sentiment Classification* (ASC) yang berfungsi menentukan polaritas sentimen dari aspek yang telah ditemukan. Selain itu, pada pendekatan yang lebih mutakhir, terdapat metode terintegrasi seperti *Aspect-Sentiment Triplet Extraction* (ASTE) yang secara simultan mengekstraksi aspek, opini, dan sentimen dalam satu proses [21], [22].

C. *Transformer*

Transformer merupakan arsitektur jaringan saraf yang diperkenalkan oleh [23] sebagai pendekatan baru dalam pemrosesan data berurutan dengan sepenuhnya mengandalkan mekanisme *attention*, tanpa menggunakan jaringan rekuren seperti RNN atau LSTM maupun jaringan konvolusional seperti CNN. Arsitektur ini dirancang untuk mengatasi keterbatasan RNN dalam menangkap ketergantungan jarak jauh serta kesulitan dalam melakukan komputasi secara paralel [24]. Secara umum, *transformer* berdiri atas dua komponen utama, yakni *encoder* dan *decoder*, di mana encoder berfungsi mengubah urutan input menjadi representasi internal, sedangkan *decoder* menggunakan representasi tersebut untuk menghasilkan urutan output. Setiap encoder dan decoder tersusun atas beberapa lapisan berulang yang mencakup mekanisme *multi-head attention* dan *feed-forward network*, serta dilengkapi dengan *residual connection* dan *layer normalization* untuk menjaga stabilitas proses pelatihan [23].

Salah satu komponen utama dalam *transformer* adalah *self-attention*, yaitu mekanisme yang memungkinkan setiap token dalam sekuens memperhatikan token lain untuk membentuk representasi yang kontekstual. Berbeda dengan RNN yang

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

bekerja secara sekuensial, *self-attention* memungkinkan pemrosesan paralel sehingga hubungan antar kata, termasuk ketergantungan jarak jauh, dapat ditangkap dengan lebih efisien. Mekanisme ini dikembangkan menjadi *multi-head attention* agar model dapat mempelajari berbagai jenis hubungan linguistik secara simultan [23]. Karena *transformer* tidak memiliki informasi urutan, digunakan *positional encoding* untuk menyisipkan informasi posisi ke dalam embedding token, sementara *feed-forward network*, *residual connection*, dan *layer normalization* digunakan untuk memperkaya representasi serta menjaga stabilitas dan efektivitas proses pelatihan [23].

D. BERT

BERT (*Bidirectional Encoder Representations from Transformers*) merupakan model representasi bahasa yang sudah dilatih sebelumnya (*pre-trained*) secara bidirectional dengan memanfaatkan arsitektur *encoder transformer*, sehingga mampu memahami konteks kiri dan kanan secara bersamaan pada setiap lapisan. Tujuan utama BERT adalah menghasilkan representasi bahasa yang bersifat umum dan kontekstual, yang selanjutnya dapat di-*fine-tune* untuk berbagai tugas. Keunggulan utama BERT dibandingkan model tradisional terletak pada kemampuannya menggabungkan konteks dua arah dalam representasi token, sehingga informasi yang dihasilkan menjadi lebih kaya dan akurat [10]. Secara arsitektural, BERT hadir dalam dua konfigurasi utama, yaitu BERT-base dan BERT-large, yang dibedakan berdasarkan jumlah lapisan *encoder*, dimensi representasi, dan jumlah *attention heads* [10]. Dalam pemrosesan input, BERT menggunakan embedding token yang dikombinasikan dengan *positional encoding* serta *segment embeddings* untuk membedakan pasangan kalimat [10].

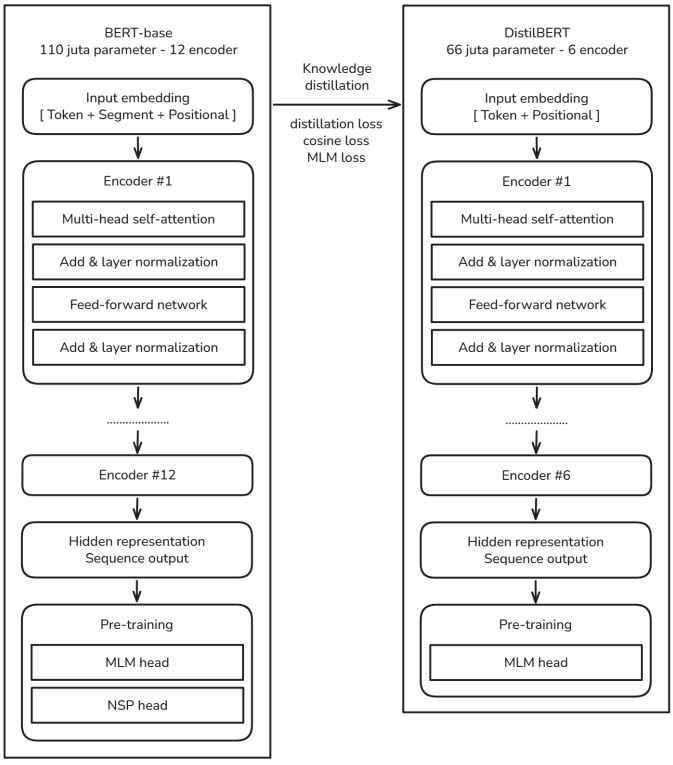
Dalam proses *pre-train*, BERT menggunakan skema tokenisasi *WordPiece* untuk memecah kata menjadi sub-kata, sehingga mampu menangani kata jarang maupun kata baru secara efisien [10]. *Pre-train* dilakukan melalui dua tugas utama, yaitu *Masked Language Modeling* (MLM) yang melatih model memprediksi token yang disamarkan berdasarkan konteks dua arah, serta *Next Sentence Prediction* (NSP) yang bertujuan memahami hubungan antar kalimat [10]. Setelah tahap pra-pelatihan, BERT dapat di-*fine-tune* pada tugas tertentu sesuai kebutuhan, seperti klasifikasi atau ekstraksi informasi, dan melatih ulang seluruh parameter model. Proses *fine-tuning* ini umumnya lebih efisien karena BERT telah membawa pengetahuan bahasa yang kuat, meskipun pengaturan hiperparameter tetap berperan penting terhadap performa akhir model [10].

E. DistilBERT

DistilBERT merupakan varian ringan dari BERT yang dikembangkan melalui teknik *knowledge distillation*, di mana model BERT berperan sebagai *teacher* dan DistilBERT sebagai *student*. Tujuan utama dari DistilBERT adalah mempertahankan sebagian besar kemampuan representasi bahasa milik BERT, namun dengan ukuran model yang lebih kecil dan waktu inferensi yang lebih cepat. Model ini tetap berbasis arsitektur Transformer dan dilatih menggunakan korpus data yang sama dengan BERT secara *self-supervised*, dengan target untuk meniru distribusi probabilitas keluaran BERT selama proses pra-pelatihan. Dengan pendekatan ini, DistilBERT mampu mempertahankan sekitar 97% performa BERT, meskipun jumlah parameternya berkurang secara signifikan hingga sekitar 66 juta parameter, sehingga lebih efisien dari segi komputasi dan memori [13].

F. Perbedaan BERT dan DistilBERT

BERT-base dan DistilBERT memiliki struktur dasar yang hamper sama karena keduanya menggunakan arsitektur *transformer encoder*. Setiap *encoder* tersusun atas komponen *multi-head self-attention*, *feed-forward network*, serta *add & layer normalization* [10], [13]. Perbedaan keduanya terletak pada jumlah lapisan *encoder* dan beberapa modifikasi yang diterapkan pada DistilBERT untuk menghasilkan model yang lebih ringkas dan efisien. Perbandingan arsitektur kedua model dapat dilihat pada Gambar 1.



Gbr. 1. Arsitektur BERT dan DistilBERT

Perbedaan utama antara BERT-base dan DistilBERT terletak pada kompleksitas model. DistilBERT mengurangi jumlah lapisan *encoder* dari 12 menjadi 6 yang menyebabkan parameternya turun dari 110 juta ke 66 juta parameter, menghilangkan *segment embedding* juga tidak menggunakan *next sentence prediction* pada tahap *pre-training* karena tidak

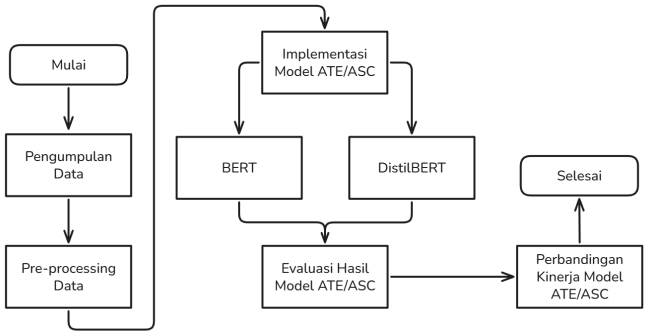
ISSN (Online) 2747-0563
Volume 6 Tahun 2026

lagi memanfaatkan informasi pasangan kalimat seperti pada objektif NSP [13], [14]. Serta melalui pendekatan *knowledge distillation*, dimana proses ini bekerja dengan melatih DistilBERT menggunakan kombinasi tiga fungsi *loss* secara bersamaan.

Pertama, *distillation loss* yang mendorong DistilBERT untuk meniru distribusi probabilitas keluaran BERT pada setiap token. Kedua, *masked language modeling loss* yang digunakan untuk memprediksi token yang disembunyikan. Ketiga, *cosine embedding loss* yang bertujuan menyelaraskan representasi internal DistilBERT dengan representasi yang dihasilkan oleh BERT [14]. Kombinasi ketiga fungsi *loss* tersebut memungkinkan DistilBERT menyerap pengetahuan dari BERT secara lebih komprehensif, baik pada tingkat *output* maupun representasi internal. Dengan pendekatan ini, DistilBERT mampu mempertahankan sekitar 97% performa BERT meskipun menggunakan arsitektur yang lebih ringkas dan waktu inferensi yang lebih cepat [13], [14].

III. METODOLOGI

Tahapan yang dilakukan dalam penelitian ini ditunjukkan pada bagan di Gambar 2. Secara umum, proses penelitian ini terdiri dari empat tahap utama, yakni pengumpulan data, pra-pemrosesan data, pelatihan model, serta evaluasi model. Kedua model yang digunakan, yaitu BERT-base dan DistilBERT.



Gbr. 2. Tahapan Penelitian

A. *Pengumpulan Data*

Dataset atau kumpulan data yang dipakai dalam penelitian ini berasal dari *review* pengguna pada platform Foursquare yang dikumpulkan dan dianotasi oleh Naver Labs Europe [25]. Data terdiri dari 1006 kalimat dari ulasan pengguna berbahasa Inggris yang membahas restoran dari berbagai wilayah di dunia.

Setiap ulasan dalam dataset telah dilengkapi dengan anotasi aspek yang menunjukkan bagian kalimat yang menjadi target opini, kategori aspek, serta polaritas sentimen yang terkait dengan aspek tersebut. Informasi anotasi disimpan dalam format XML yang memuat posisi karakter awal dan akhir dari istilah aspek di dalam kalimat. Dataset ini terdiri dari 6 kategori aspek yaitu *food, service, drinks, restaurant, ambience, dan location* dengan distribusi yang telah dijabarkan pada Tabel I.

TABEL I
DISTRIBUSI KATEGORI

Kategori	Jumlah
Food	527
Service	113
Drinks	93
Restaurant	77
Ambience	62
Location	9

Jumlah data pada tugas ASC tidak sama dengan jumlah kalimat pada dataset awal disebabkan karena setiap kalimat dapat mengandung lebih dari satu aspek yang memiliki polaritas sentimen yang berbeda. Oleh karena itu, pada tahap transformasi data, setiap pasangan *aspect-sentence* diperlakukan sebagai satu kesatuan klasifikasi yang terpisah. Berdasarkan proses ini diperoleh total 881 data, yang sesuai dengan jumlah anotasi aspek pada dataset. Distribusi polaritas ditampilkan pada Tabel II, yang menunjukkan jumlah polaritas dalam dataset.

TABEL II
DISTRIBUSI POLARITAS

Polaritas	Jumlah
Positive	758
Negative	108
Neutral	15

B. *Pre-processing Data*

Tahap pra-pemrosesan data dilakukan untuk menyiapkan data ulasan agar dapat digunakan secara optimal dalam pelatihan model ekstraksi aspek dan klasifikasi sentimen. Proses ini diawali dengan pembersihan teks (*text cleaning*) untuk menghilangkan elemen-elemen yang tidak relevan, seperti tanda baca berlebih, karakter khusus, dan spasi ganda, tanpa mengubah makna utama dari kalimat. Selanjutnya, dilakukan normalisasi teks dengan mengubah seluruh huruf menjadi huruf kecil guna menjaga konsistensi representasi kata. Pada tahap ini tidak dilakukan penghapusan *stopword* maupun *stemming*, karena model berbasis *transformer* memanfaatkan konteks kata secara penuh dalam proses pembelajaran.

Setelah tahap pembersihan dan normalisasi, data dipersiapkan dalam dua format yang berbeda sesuai dengan tugas yang dikerjakan. Untuk tugas ekstraksi aspek (ATE), setiap kalimat diubah ke dalam format pelabelan berurutan menggunakan skema BIO (*Beginning, Inside, Outside*). Setiap token dalam kalimat diberi label B jika merupakan awal dari suatu aspek, I jika merupakan lanjutan aspek, dan O jika token tersebut bukan bagian dari aspek apa pun. Formati ini dilengkapi dengan kategori dari setiap aspeknya menjadi B-KATEGORI, I-KATEGORI, dan O. Format BIO ini memungkinkan model mempelajari posisi, aspek aktegori, dan batas istilah aspek

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

secara kontekstual dalam sebuah kalimat. Distribusi pelabelan ini dapat dilihat pada Tabel III.

TABEL III
DISTRIBUSI LABEL BIO

Label	Jumlah
O	6606
B-FOOD	514
I-FOOD	453
B-SERVICE	112
I-SERVICE	4
B-DRINKS	92
I-DRINKS	58
B-RESTAURANT	77
I-RESTAURANT	36
B-AMBIENCE	59
I-AMBIENCE	12
B-LOCATION	9
I-LOCATION	1

Sementara itu, untuk tugas klasifikasi sentimen aspek (ASC), data ditransformasikan ke dalam bentuk pasangan aspek-kalimat (*aspect-sentence pair*). Setiap aspek yang muncul dalam satu kalimat diperlakukan sebagai satu data tersendiri, dengan kalimat utuh sebagai konteks dan label sentimen sebagai target. Dengan demikian, satu kalimat dapat menghasilkan lebih dari satu data ASC apabila mengandung beberapa aspek dengan sentimen yang berbeda. Hasil dari tahap pra-pemrosesan ini berupa dua dataset terpisah, yaitu dataset berbasis BIO tagging untuk pelatihan model ATE dan dataset pasangan aspek–kalimat untuk pelatihan model ASC.

C. *Pelatihan Model*

Pada penelitian ini digunakan dua model berbasis transformer, yaitu BERT dan DistilBERT, untuk menyelesaikan dua tugas utama dalam ABSA, yakni *aspect term extraction* (ATE) dan *aspect sentiment classification* (ASC). Kedua tugas tersebut dievaluasi secara terpisah untuk menganalisis kemampuan model pada masing-masing sub-tugas ABSA secara independen, tanpa mempertimbangkan propagasi kesalahan antar sub-tugas.

Untuk tugas ATE, implementasi dilakukan sebagai permasalahan *sequence labeling* dengan pendekatan *token classification*. Setiap kalimat yang telah melalui tahap pra-pemrosesan dan diberi label BIO digunakan sebagai input model. Tokenisasi dilakukan menggunakan *fast tokenizer* bawaan masing-masing model, dengan mekanisme penyelarasan label (*label alignment*) agar setiap token hasil pemecahan subword tetap mempertahankan label yang sesuai. Model dilatih untuk memprediksi label BIO pada setiap token, sehingga mampu mengidentifikasi posisi dan batas istilah aspek dalam kalimat secara kontekstual.

Sementara itu, tugas ASC diimplementasikan sebagai permasalahan *sequence classification*. Setiap data direpresentasikan dalam bentuk pasangan aspek–kalimat (*aspect-sentence pair*) yang dibentuk berdasarkan anotasi aspek (*gold aspect*), di mana aspek dan kalimat dipisahkan sesuai dengan format input standar pada model transformer. Model dilatih untuk memprediksi polaritas sentimen dari setiap pasangan aspek–kalimat berdasarkan konteks kalimat secara keseluruhan. Pendekatan ini memungkinkan model untuk membedakan sentimen yang berbeda pada aspek yang berlainan, meskipun berasal dari kalimat yang sama.

Pada kedua tugas, proses pelatihan dilakukan dengan skema *fine-tuning* menggunakan optimizer AdamW dan *learning rate scheduler* linear dengan *warmup*. Untuk menjaga konsistensi eksperimen dan mengontrol variabel eksternal, seluruh pengaturan pelatihan seperti *learning rate*, jumlah *epoch*, ukuran *batch*, panjang maksimum token, serta pembagian data latih dan validasi dibuat sama pada kedua model, pengaturan pelatihan tersebut dijabarkan pada Tabel IV. Karena ukuran dataset relatif terbatas, penelitian ini menggunakan pembagian *train-validation* 80:20 tanpa *test set* agar jumlah data pelatihan tetap memadai dan model dapat belajar secara lebih efektif. Selain itu, jumah *epoch* sebanyak 3 dipilih untuk mengurangi resiko *overfitting* pada dataset kecil ini. Meskipun setiap model memiliki karakteristik optimasi yang berbeda, pendekatan ini bertujuan untuk memastikan bahwa perbedaan performa yang muncul terutama dipengaruhi oleh perbedaan arsitektur model, bukan oleh variasi skema pelatihan.

TABEL IV
PARAMETER PELATIHAN

Parameter	Nilai
Rasio Train:Validation	80:20
Optimizer	AdamW
Batch	16
Max Sequence Length	128
Epoch	3
Learning Rate	2×10^{-5}
Learning Rate Scheduler	Linear Warmup
Warmup Ratio	0,1

Penelitian ini dilakukan menggunakan *Google Colab* sebagai *enivronment* komputasi dengan GPU T4 dan 12GB RAM. Implementasi model menggunakan framework *PyTorch* dengan memanfaatkan *library HuggingFace Transformers* untuk mengakses dan melatih model BERT dan DistilBERT.

D. *Evaluasi Model*

Evaluasi performa model pada penelitian ini dilakukan secara terpisah untuk masing-masing tugas, yaitu ATE dan ASC dengan menggunakan metrik yang sesuai dengan karakteristik setiap tugas. Seluruh evaluasi dilakukan pada data validasi yang displit secara tetap untuk memastikan perbandingan performa antara model BERT dan DistilBERT konsisten.

Pada tugas ATE, evaluasi dilakukan menggunakan *precision, recall, dan weighted F1-score* pada tingkat token, serta *entity-level F1-score*. Evaluasi *token-level* dilakukan berdasarkan skema BIO untuk mengukur kemampuan model dalam memprediksi label setiap token dalam kalimat. *Precision*

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

digunakan untuk mengukur ketepatan model dalam memberikan label token aspek, sedangkan *recall* mengukur kemampuan model dalam menemukan seluruh token aspek yang relevan. F1-score digunakan sebagai metrik utama karena mampu merepresentasikan keseimbangan antara *precision* dan *recall*. Selain evaluasi pada tingkat token, digunakan juga *entity-level F1-score* untuk menilai kemampuan model dalam mengekstraksi aspek secara utuh. Evaluasi token-level hanya menilai ketepatan label pada setiap kata secara individual, sedangkan entity-level mengevaluasi apakah kalimat yang membentuk suatu aspek berhasil dikenali dengan benar sebagai satu entitas. Dengan demikian, evaluasi entity-level memberikan gambaran yang lebih representatif terhadap kemampuan model dalam melakukan ekstraksi aspek secara lengkap.

Sementara itu, pada tugas ASC, evaluasi dilakukan menggunakan *accuracy, macro F1-score, dan weighted F1-score*. *Accuracy* digunakan untuk memberikan gambaran umum mengenai tingkat ketepatan model dalam mengklasifikasikan polaritas sentimen aspek. *Weighted* F1-score digunakan untuk menilai performa model secara keseluruhan dengan mempertimbangkan proporsi jumlah data pada setiap kelas, sedangkan *macro F1-score* digunakan untuk melihat pemerataan performa model pada seluruh kelas sentimen tanpa dipengaruhi oleh dominasi kelas mayoritas. Penggunaan kedua metrik ini penting karena distribusi data sentimen pada dataset penelitian tidak seimbang antar kelasnya.

IV. HASIL DAN PEMBAHASAN

Pada penelitian ini dilakukan evaluasi terhadap dua model berbasis transformer, yaitu BERT-base dan DistilBERT. Evaluasi dilakukan menggunakan data validasi yang telah dipisahkan sebelumnya.

A. *Aspect Term Extraction*

ATE dilakukan menggunakan BERT dan DistilBERT dengan pendekatan *token classification*. Evaluasi dilakukan menggunakan metrik *precision, recall, dan F1-score* pada tingkat token dan entitas (tingkat token ditandai dengan huruf 'T', sementara tingkat entitas ditandai dengan huruf 'E'). Tabel V menunjukkan hasil evaluasi kedua model.

TABEL V
HASIL ASPECT TERM EXTRACTION

Metrik	BERT-base	DistilBERT
Precision (T)	0,758	0,764
Recall (T)	0,814	0,825
F1-score (T)	0,781	0,794
F1-score (E)	0,419	0,411
Train Time	46,5 detik	29,8 detik

Berdasarkan hasil evaluasi pada Tabel V, kedua model menunjukkan performa yang cukup baik pada tingkat token. DistilBERT memperoleh nilai weighted F1-score sebesar 0,7941, sedikit lebih tinggi disbanding BERT dengan nilai 0,7817. Namun, pada tingkat entitas, nilai F1-score keduanya masih relatif rendah, yaitu 0,419 untuk BERT, dan 0,411 untuk DistilBERT. Hal ini menujukkan bahwa meski model dapat mengenali token yang berkaitan dengan aspek, model masih kesulitan dalam mengidentifikasi batasan entitas aspek. Hal ini bisa disebabkan oleh kurangnya jumlah dan kemerataan data, sehingga model tidak memiliki cukup informasi. Tapi di llain sisi, model yang lebih ringan seperti DistilBERT dapat menghasilkan performa yang mirip pada dataset berukuran kecil ini.

Sementara itu, dari sisi efisiensi komputasi, DistilBERT menunjukkan waktu pelatihan yang jauh lebih cepat, yaitu hanya 29,8 detik, dibandingkan BERT yang membutuhkan 46,5 detik. Hal ini menunjukkan bahwa arsitektur DistilBERT yang lebih ringan mampu memberikan hasil yang kompetitif dengan kebutuhan waktu komputasi yang jauh lebih cepat.

B. *Aspect Sentiment Classification*

Pada tahap ASC dilakukan klasifikasi polaritas sentimen terhadap aspek yang telah diidentifikasi sebelumnya. Evaluasi dilakukan menggunakan metrik *accuracy, macro F1-score,* dan *weighted F1-score*. Hasil evaluasi ditunjukkan pada Tabel VI.

TABEL VI
HASIL ASPECT SENTIMENT CLASSIFICATION

Metrik	BERT-base	DistilBERT
Accuracy	0,886	0,892
Macro F1-score	0,456	0,487
Weighted F1-score	0,856	0,874
Train Time	53,4 detik	28 detik

Berdasarkan hasil pada Tabel VI, DistilBERT menunjukkan performa yang sedikit lebih unggul daripada BERT dalam tugas klasifikasi sentimen. DistilBERT memperoleh akurasi sebesar 0,892, sedangkan BERT memperoleh 0,886. Nilai weighted F1 DistilBERT sebesar 0,874 juga lebih unggul dibandingkan BERT sebesar 0,856. Hal ini menunjukkan bahwa DistilBERT mampu melakukan klasifikasi sentimen dengan baik meskipun memiliki jumlah parameter yang lebih sedikit.

Namun demikian, nilai macro F1-score pada kedua model masih rendah. Hal ini menunjukkan bahwa performa model belum merata pada semua kelas sentimen. Model cenderung lebih baik dalam memprediksi kelas positif, sementara performa pada kelas netral masih sangat rendah. Untuk melihat distribusi prediksi model secara lebih rinci, digunakan confusion matrix untuk masing-masing model yang ditunjukkan pada Tabel VII dan Tabel VIII.

TABEL VII
CONFUSION MATRIX BERT_BASE

Actual/Predicted	Negative	Neutral	Positive
Negative	59	0	49
Neutral	0	0	15
Positive	16	0	742

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

TABEL VIII
CONFUSION MATRIX DISTILBERT

Actual/Predicted	Negative	Neutral	Positive
Negative	63	0	45
Neutral	0	0	15
Positive	12	0	746

Berdasarkan confusion matrix pada Tabel VII dan Tabel VIII, kedua model cenderung lebih baik dalam memprediksi kelas positif yang merupakan kelas dengan jumlah data terbanyak daam dataset. Sebaliknya, untuk kelas netral, kedua model memprediksinya sebagai kelas positif. Hal ini menunjukkan bahwa model tidak bisa membedakan sentiment netral dan positif. Selain itu, pada kelas negative, masih terdapat bebearapa kesalahan klasifikasi dimana Sebagian data negatif diprediksi sebagai positif. Ketidakseimbangan distribusi kelas pada dataset sangat berpengaruh terhadap performa model. Kelas positif mendominasi sebagian besar data, sedangkan kelas netral hanya memiliki jumlah sampel yang sangat sedikit. Kondisi ini menyebabkan model cenderung bias terhadap kelas mayoritas dan mengalami kesulitan dalam mengenali pola pada kelas minoritas.

V. KESIMPULAN

Penelitian ini melakukan perbandingan kinerja BERT dan DistilBERT dalam tugas *Aspect-Based Sentiment Analysis* yang terdiri dari *Aspect Term Extraction* dan *Aspect Sentiment Classification* pada dataset ulasan restoran. Hasil eksperimen menunjukkan bahwa kedua model mampu memberikan performa yang baik dalam mengekstraksi aspek dan mengklasifikasikan sentimen. Pada tugas ATE, DistilBERT memperoleh nilai F1-score token sebesar 0,794, sedikit lebih tinggi dibandingkan BERT yang memperoleh 0,781. Sementara itu, pada tugas ASC, DistilBERT juga menunjukkan performa yang lebih baik dengan akurasi sebesar 0,892 dan weighted F1-score sebesar 0,874, dibandingkan BERT yang memperoleh akurasi 0,886 dan weighted F1-score 0,856. DistilBERT juga menunjukkan efisiensi komputasi yang jauh lebih baik, dengan waktu pelatihan yang lebih cepat dibandingkan BERT pada kedua tugas. Namun demikian, hasil penelitian juga menunjukkan bahwa performa kedua model pada beberapa kelas aspek dan sentimen masih terbatas, terutama pada kelas dengan jumlah data yang sedikit. Oleh karena itu, penelitian selanjutnya dapat difokuskan pada peningkatan kualitas dataset, penyeimbangan distribusi kelas, serta eksplorasi metode penanganan data tidak seimbang untuk meningkatkan kinerja model secara keseluruhan.

UCAPAN TERIMA KASIH

Penulis mengucapkan banyak terima kasih kepada Program Studi Teknik Informatika UPN "Veteran" Jawa Timur yang telah memberikan fasilitas dan dukungan akademik. Penulis juga mengucapkan terima kasih kepada dosen pembimbing atau rekan peneliti yang telah memberikan masukan dan arahan dalam penyusunan penelitian ini. Penulis juga mengucapkan terima kasih kepada reviewer dan panitia penyelenggara SANTIKA atas masukan yang berharga serta dukungan dalam proses penelaahan dan publikasi artikel ini.

REFERENSI

[1] Jasmine Aulia Mumtaz, Kinaya Khairunnisa Komariansyah, Wildan Holik, Muhammad Galuh Gumelar, Reza Pratama, and Humannisa Rubina Lestari, "Analisis Sentimen Ulasan Aplikasi HeyJapan di Google Play Store Menggunakan Algoritma NLP," *Pragmatik J. Rumpun Ilmu Bhs. dan Pendidikan* , vol. 3, no. 3, pp. 157–167, 2025, doi: 10.61132/pragmatik.v3i3.1801.

[2] S. E. Safitri, W. D. Yuniarti, M. R. Handayani, and K. Umam, "User Opinion Mining on the Maxim Application Reviews Using BERT-Base Multilingual Uncased," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 14, no. 3, pp. 365–372, 2025, doi: 10.32736/sisfokom.v14i3.2391.

[3] A. Amrullah, "Advanced Sentiment Analysis Using Deep Learning: A Comprehensive Framework for High-Accuracy and Interpretable Models," *Intellithings*, vol. 1, no. 1, p. p, 2025, [Online]. Available: https://doi.org/xxxxxxxxx

[4] R. Parlika, S. I. Pradika, A. M. Hakim, and K. R. N. M, "Analisis Sentimen Twitter Terhadap Bitcoin dan Cryptocurrency Berbasis Python TextBlob," vol. 2, pp. 33–37, 2020.

[5] N. Wijaya and E. S. Panjaitan, "Analisis Sentimen Ulasan Aplikasi Instagram di Google Play Store: Pendekatan Multinomial Naive Bayes dan Berbasis Leksikon," *Build. Informatics, Technol. Sci.*, vol. 6, no. 2, pp. 921–929, 2024, doi: 10.47065/bits.v6i2.5615.

[6] A. N. Hasanah, B. N. Sari, U. S. Karawang, T. Timur, and J. Barat, "Jasa Ojek Online Maxim Pada Google Play," vol. 12, no. 1, pp. 90–96, 2024.

[7] I. K. Najibulloh, I. Tahyudin, and D. I. S. Saputra, "Analisis Sentimen Ulasan Co-Pilot Google Play dengan SVM, Neural Network, dan Decision Tree," *Edumatic J. Pendidik. Inform.*, vol. 9, no. 1, pp. 275–283, 2025, doi: 10.29408/edumatic.v9i1.29673.

[8] Ardiansyah, Adika Sri Widagdo, Krisna Nuresa Qodri, F. E. N. Saputro, and Nisrina Akbar Rizky P, "Analisis sentimen terhadap pelayanan Kesehatan berdasarkan ulasan Google Maps menggunakan BERT," *J. Fasilkom*, vol. 13, no. 02, pp. 326–333, 2023, doi: 10.37859/jf.v13i02.5170.

[9] A. Karimi, L. Rossi, and A. Prati, "Improving BERT Performance for Aspect-Based Sentiment Analysis," *ICNLSP 2021 - Proc. 4th Int. Conf. Nat. Lang. Speech Process.*, pp. 39–46, 2021.

[10] 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.

[11] H. Xu, L. Shu, P. S. Yu, and B. Liu, "Understanding Pre-trained BERT for Aspect-based Sentiment Analysis," *COLING 2020 - 28th Int. Conf. Comput. Linguist. Proc. Conf.*, pp. 244–250, 2020, doi: 10.18653/v1/2020.coling-main.21.

[12] Y. Wu, Z. Jin, C. Shi, P. Liang, and T. Zhan, "Research on the application of deep learning-based BERT model in sentiment analysis," *Appl. Comput. Eng.*, vol. 71, no. 1, pp. 14–20, 2024, doi: 10.54254/2755-2721/71/2024ma.

[13] S. Ling, "Fine-Tuning distilBERT for Enhanced Sentiment Classification," *J. Big Data Comput.*, vol. 2, no. 4, pp. 108–112, 2024, doi: 10.62517/jbdc.202401417.

[14] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," pp. 2–6, 2020, [Online]. Available: http://arxiv.org/abs/1910.01108

[15] I. Perikos and A. Diamantopoulos, "Explainable Aspect-Based Sentiment Analysis Using Transformer Models," *Big Data Cogn. Comput.*, vol. 8, no. 11, 2024, doi: 10.3390/bdcc8110141.

[16] D. Jayakody *et al.*, "Aspect-based Sentiment Analysis Techniques: A Comparative Study," *Moratuwa Eng. Res. Conf. MERCon*, pp. 205–210, 2024, doi: 10.1109/MERCon63886.2024.10688631.

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

[17] J. Serrano-Guerrero, J. A. Olivas, F. P. Romero, and E. Herrera-Viedma, "Sentiment analysis: A review and comparative analysis of web services," *Inf. Sci. (Ny).*, vol. 311, pp. 18–38, 2015, doi: 10.1016/j.ins.2015.03.040.
[18] M. Wankhade, A. C. S. Rao, and C. Kulkarni, *A survey on sentiment analysis methods, applications, and challenges*, vol. 55, no. 7. Springer Netherlands, 2022. doi: 10.1007/s10462-022-10144-1.
[19] M. Zhang *et al.*, "Span-level Aspect-based Sentiment Analysis via Table Filling," *Proc. Annu. Meet. Assoc. Comput. Linguist.*, vol. 1, pp. 9273–9284, 2023, doi: 10.18653/v1/2023.acl-long.515.
[20] S. K and F. F, "Survey on Aspect-level sentiment analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 3, pp. 813–830, 2016.
[21] Y. C. Hua, P. Denny, J. Wicker, and K. Taskova, *A systematic review of aspect-based sentiment analysis: domains, methods, and trends*, vol. 57, no. 11. Springer Netherlands, 2024. doi: 10.1007/s10462-024-10906-z.
[22] W. Zhang, X. Li, Y. Deng, L. Bing, and W. Lam, "A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 11, pp. 11019–11038, 2023, doi: 10.1109/TKDE.2022.3230975.
[23] K. Mohiuddin *et al.*, "Attention Is All You Need," *Int. Conf. Inf. Knowl. Manag. Proc.*, no. Nips, pp. 4752–4758, 2023, doi: 10.1145/3583780.3615497.
[24] D. Bahdanau, K. H. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–15, 2015.
[25] C. Brun and V. Nikoulina, "Aspect Based Sentiment Analysis into the Wild," *WASSA 2018 - 9th Work. Comput. Approaches to Subj. Sentim. Soc. Media Anal. Proc. Work.*, pp. 116–122, 2018, doi: 10.18653/v1/P17.