ANALISIS PERFORMA INDOBERTWEET DAN DISTILBERT PADA ANALISIS SENTIMEN DENGAN DATASET BERLABEL MANUAL DAN OTOMATIS

(*Performance Analysis of IndoBERTweet and DistilBERT on Sentiment Analysis Using Manually and Automatically Labeled Datasets*)

Lyudza Aprilia Kansha[1], I Gede Pasek Suta Wijaya[2], Fitri Bimontoro[3]
Dept Informatics Engineering, Mataram University
Jl. Majapahit 62, Mataram, Lombok NTB, INDONESIA
Email: lyudzaaprilia@gmail.com, [gpsutawijaya, bimo]@unram.ac.id

Abstract

Tourism in North Lombok, particularly the Gili Islands (Trawangan, Air, and Meno), plays a significant role in the regional economy. Understanding public sentiment through social media is crucial for improving tourism services and management. This study compares the performance of two transformer-based models IndoBERTweet and DistilBERT in sentiment analysis of tourism-related tweets from X (Twitter). The dataset used consists of 3,159 preprocessed Indonesian-language tweets, labeled through both manual annotation and automatic classification using DistilBERT. IndoBERTweet was evaluated on both manual and automatic labels, while DistilBERT was only applied to the manually translated dataset. Experimental results show that IndoBERTweet with manual labeling achieved an F1-score of 72.98% and demonstrated more balanced performance across all sentiment classes. Meanwhile, DistilBERT showed lower F1-scores overall (max. 57%) but performed efficiently in terms of computational time. Automatic labeling showed weak agreement with manual annotation (only 31.8% match), leading to bias in model learning, particularly the failure to detect neutral sentiment. Evaluation using new test sentences yielded 80% prediction accuracy, yet revealed that IndoBERTweet struggles with implicit sentiment or subtle dissatisfaction. This research highlights IndoBERTweet's effectiveness in Indonesian sentiment analysis and emphasizes the trade off between computational efficiency and contextual accuracy in lightweight models like DistilBERT.

Keywords: *DistilBERT, IndoBERTweet, X(Twitter), Sentiment Analysis, Tourism*

1. PENDAHULUAN

Pariwisata merupakan salah satu sektor dengan pertumbuhan tercepat di dunia dan memberikan kontribusi signifikan terhadap perekonomian global. Menurut *World Travel* & *Tourism Council* (WTTC), sektor ini menyumbang lebih dari 10% terhadap Produk Domestik Bruto (PDB) global serta menciptakan jutaan lapangan kerja [1]. Di Indonesia, pariwisata berperan penting dalam pembangunan ekonomi daerah, salah satunya adalah Pulau Lombok yang semakin populer di kalangan wisatawan domestik maupun internasional. Keindahan alam, kekayaan budaya, serta keramahan penduduk menjadikan Lombok sebagai destinasi yang kompetitif [2].

Kawasan Strategis Pariwisata Nasional (KSPN) Gili Tramena yang meliputi Gili Trawangan, Gili Air, dan Gili Meno di Kabupaten Lombok Utara (KLU), merupakan salah satu destinasi unggulan di Indonesia [3]. Persepsi dan tanggapan wisatawan terhadap destinasi ini berperan besar dalam membentuk citra pariwisata, sehingga penting untuk dipantau dan dianalisis sebagai dasar penyusunan strategi promosi dan peningkatan kualitas layanan [4].

Perkembangan teknologi informasi telah menyebabkan pertumbuhan media sosial yang sangat pesat. Media sosial kini menjadi salah satu sumber informasi paling populer dan banyak digunakan masyarakat untuk berinteraksi secara online serta berbagi pendapat, kritik, dan saran secara bebas [5]. Salah satu platform media sosial yang populer di Indonesia adalah X (Twitter), yang secara luas digunakan untuk menyampaikan berbagai pendapat, kritik, dan saran terkait berbagai isu, termasuk pariwisata [5],[6]. Dengan volume data yang besar dan sifatnya yang dinamis, platform ini memiliki potensi yang sangat besar sebagai sumber data untuk

menganalisis opini publik melalui pendekatan analisis sentimen.

Analisis sentimen merupakan bagian dari text mining yang bertujuan mengklasifikasikan opini menjadi sentimen positif, netral, atau negatif [8]. Dalam beberapa tahun terakhir, pendekatan berbasis deep learning telah menjadi metode yang unggul untuk analisis sentimen, terutama dengan hadirnya model berbasis transformer seperti BERT dan variannya [9].

Penelitian ini bertujuan untuk membandingkan performa dua model dengan arsitektur transformer learning, yaitu DistilBERT dan IndoBERTweet, dalam menganalisis sentimen wisatawan terhadap destinasi wisata Gili di KLU. DistilBERT adalah versi ringan dari BERT yang dirancang untuk efisiensi dalam pemrosesan teks berbahasa Inggris [10], sedangkan IndoBERTweet merupakan model yang dioptimalkan untuk memahami konteks bahasa Indonesia informal seperti yang banyak ditemukan di *X* (*Twitter*) [11].

Pemilihan model IndoBERTweet dan DistilBERT dalam penelitian ini didasarkan pada kebutuhan untuk melakukan analisis sentimen terhadap teks berbahasa Indonesia yang dihasilkan di media sosial, khususnya Twitter. IndoBERTweet dipilih karena kemampuannya untuk memahami konteks bahasa Indonesia yang informal, seperti yang banyak ditemukan dalam platform media sosial. Model ini telah dioptimalkan dengan data Twitter, yang memungkinkan pengenalan nuansa emosi dalam teks informal, slang, dan penggunaan bahasa sehari-hari yang khas. Dalam konteks ini, IndoBERTweet menawarkan keunggulan dalam menangani variasi dan kompleksitas bahasa Indonesia yang muncul dalam tweet, sehingga model relevan untuk analisis sentimen pada platform seperti Twitter yang sering kali penuh dengan ekspresi subjektif dan bahasa tidak baku [12].

Sementara itu, DistilBERT dipilih karena merupakan versi yang lebih ringan dan efisien dari BERT, yang terbukti sangat efektif dalam tugas-tugas pemrosesan bahasa alami (NLP), terutama dalam analisis sentimen bahasa Inggris. DistilBERT unggul dalam hal kecepatan pelatihan dan penggunaan memori yang lebih rendah, sehingga cocok digunakan saat efisiensi komputasi dibutuhkan tanpa mengorbankan performa secara signifikan. Selain itu, DistilBERT juga memungkinkan pengujian pada data berbahasa Inggris, memberikan perspektif berbeda dalam membandingkan model analisis sentimen antara bahasa Indonesia dan bahasa Inggris [13],[14].

Dalam penelitian ini, Data dikumpulkan melalui proses *crawling* dari *X* (*Twitter*) dan mencakup tweet berbahasa Indonesia. Label sentimen diberikan dengan dua pendekatan, yaitu secara manual oleh annotator dan otomatis menggunakan model DistilBERT. Kombinasi dua jenis bahasa dan pelabelan ini memungkinkan analisis perbandingan model secara komprehensif.

Hasil dari penelitian ini diharapkan dapat memberikan wawasan bagi pemangku kepentingan di industri pariwisata untuk memantau sentimen wisatawan secara *real time*. Sentimen negatif terhadap aspek tertentu, seperti kebersihan atau fasilitas, dapat menjadi indikator untuk perbaikan cepat, sementara sentimen positif dapat digunakan untuk memperkuat promosi. Selain itu, penelitian ini juga berkontribusi pada pengembangan metode analisis sentimen dalam konteks pariwisata lokal di Indonesia, guna meningkatkan kualitas pelayanan berdasarkan opini publik.

2. TINJAUAN PUSTAKA

Penelitian mengenai analisis sentimen telah banyak dilakukan, baik menggunakan model berbasis deep learning maupun metode klasik. Secara khusus, model berbasis BERT dan turunannya telah menjadi pilihan utama dalam pengolahan bahasa alami.

Putri dalam studinya [15] membandingkan performa BERT dan DistilBERT dalam analisis sentimen terkait Pemilu Presiden Indonesia. Hasilnya menunjukkan bahwa meskipun akurasi DistilBERT sedikit di bawah BERT, model ini lebih efisien dari segi waktu eksekusi dan konsumsi memori, menjadikannya cocok untuk implementasi real-time.

Sharma dan kumar [16] menunjukkan bahwa DistilBERT merupakan model yang efisien dan akurat dalam tugas analisis sentimen teks bahasa inggris. Mereka menyoroti keunggulan DistilBERT dalam kecepatan dan efisiensi sumber daya, menjadikannya sangat ideal untuk implementasi pada sistem berskala besar dan *real time*.

Setiawan dan Hidayat [17] membandingkan kinerja dua model pra-latih, IndoBERT dan IndoBERTweet, dalam mendeteksi emosi pada teks berbahasa Indonesia. Penelitian ini mengkategorikan empat emosi yaitu marah, senang, takut, dan sedih. Hasil penelitian menunjukkan bahwa model yang tidak melalui tahapan *preprocessing*, khususnya penghapusan *stopwords* dan *stemming*, menghasilkan kinerja yang lebih baik dibandingkan model yang melalui tahapan tersebut. Hasil penelitian tersebut menunjukkan bahwa stopwords dan kata tanpa stemming berkontribusi terhadap performa model. Model IndoBERTweet menunjukkan akurasi tertinggi

sebesar 92,54%, sedangkan IndoBERT mencapai 88,81%.

Sementara itu Indriani et al. [18] melakukan studi perbandingan terhadap tiga model NLP populer: IndoBERT, IndoBERTweet, dan mBERT dalam klasifikasi multilabel umpan balik mahasiswa. Hasilnya menunjukkan bahwa IndoBERTweet memiliki performa terbaik karena kemampuannya dalam memahami konteks bahasa Indonesia informal secara efektif.

Penelitian oleh Claudia et al. [19] menerapkan *fine-tuning* pada model BERT untuk klasifikasi sentimen terhadap isu obesitas pada tweet berbahasa Indonesia. Model IndoBERTweet mencatatkan akurasi dan F1-score yang lebih tinggi dibandingkan IndoBERT, menunjukkan kemampuannya dalam menangkap nuansa emosional dalam bahasa Indonesia informal.

Berdasarkan penelitian-penelitian tersebut, dapat disimpulkan bahwa DistilBERT sangat cocok digunakan untuk teks berbahasa Inggris dengan kebutuhan efisiensi tinggi, sementara IndoBERTweet menjadi pilihan terbaik dalam pengolahan teks informal berbahasa Indonesia. Namun belum banyak penelitian yang secara langsung membandingkan kedua model tersebut dalam konteks pelabelan manual dan otomatis dalam konteks pariwisata.

1. DistilBERT

DistilBERT adalah model pembelajaran mesin yang lebih ringan dan cepat dibandingkan BERT. Meskipun memiliki struktur yang lebih sederhana, DistilBERT tetap mempertahankan kemampuan pemahaman bahasa alami yang serupa. Model ini menggunakan teknik yang disebut *knowledge distillation*, yang melatih model lebih kecil untuk meniru kinerja model lebih besar, sehingga menghasilkan waktu inferensi yang lebih cepat dan efisiensi memori yang lebih baik [20].

2. IndoBERTweet

IndoBERTweet adalah model BERT yang dioptimalkan untuk tweet berbabahasa Indonesia. Model ini dilatih dengan data dari berbagai sumber teks, termasuk media sosial, sehingga mampu memahami konteks lokal, gaya bahasa informal, dan nuansa sentimen khas pengguna Twitter di Indonesia. Kemampuan ini sangat penting dalam penerapan analisis sentimen [21].

3. X (Twitter)

X (Twitter) yang berasal dari opini dan komentar adalah sumber yang bisa dimanfaatkan guna mengkaji sentimen khalayak umum terhadap organisasi atau individu karena memuat sentimen yang dijadikan selaku indikator persepsi khalayak umum untuk evaluasi berikutnya [22].

4. Analisis Sentimen

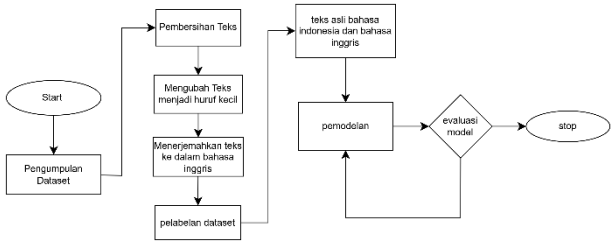
Analisis sentimen, atau *opinion mining*, adalah proses otomatis untuk memahami, mengekstrak, dan mengolah data teks guna memperoleh informasi tentang sikap atau emosi terhadap suatu entitas. Proses ini melibatkan beberapa tahap, termasuk pengumpulan data, praproses teks, dan penerapan algoritma untuk mengklasifikasikan sentimen ke dalam kategori positif, negatif, atau netral [23].

5. Pariwisata

Pariwisata merupakan sektor ekonomi yang melibatkan aktivitas yang melibatkan aktivitas perjalanan untuk keperluan rekreasi, bisnis, atau tujuan lain. Industri ini mencakup berbagai layanan seperti penginapa, transportasi restoran [24].

3. METODE PENELITIAN

Penelitian ini dimulai dengan pengumpulan dataset melalui *crawling* di platform X (Twitter), yang menghasilkan sejumlah 6.030 tweet mentah. Selanjutnya, dilakukan tahap preprocessing yang mencakup pembersihan data duplikat, sehingga menghasilkan 3.159 tweet bersih yang akan digunakan dalam proses *fine tune*. Alur metodologi penelitian dapat dilihat pada Gambar 1.



Gambar 1. Alur Metodologi Penelitian

3.1 Pengumpulan Dataset

Pengumpulan dataset dilakukan dengan *crawling* pada aplikasi *X* (*Twitter*) dengan tweet terkait Gili di KLU dari 1 Januari 2020 hingga 5 November 2024. Menggunakan kata kunci yang relevan dengan topik sentimen seperti “Gili Island”, “ Gili lombok”, “Gili KLU”, “Gili Meno”, “Gili Trawangan”, ‘’Gili Air”, dan “Pariwisata Gili”. Proses pengumpulan menggunakan autentikasi token X untuk memastikan akses yang sah ke API dan data disimpan dalam format CSV.

3.2 *Cleaning* Dataset

Data mentah yang dikumpulkan masih mengandung elemen yang perlu dibersihkan (data *cleaning*), di mana dari total 6030 data mentah yang dikumpulkan, hanya tersisa 3159 data setelah melalui proses pembersihan. Pada tahap ini, dilakukan penghapusan berbagai elemen yang tidak diperlukan

seperti tanda baca, *emotikon*, angka, URL, *mention*, *hashtag*, karakter khusus, serta spasi berlebih yang dapat mengganggu analisis teks.

TABEL I. PEMBERSIHAN DATASET

No	Teks Asli	Teks
1	@IShowUpdates07 mending ke NTB gili trawangan lombok lebih better daripada ke papua agax susah	mending ke NTB gili trawangan lombok lebih better daripada ke papua agax susah
2	@IShowUpdates07 its better to go to NTB gili trawangan lombok than to go to papua which is a bit difficult	its better to go to NTB gili trawangan lombok than to go to papua which is a bit difficult
..	..	…

3.3 Preprocessing Dataset

3.3.1 *Lowercase*

Untuk menyeragamkan kata, setiap huruf dalam penelitian ini diubah menjadi huruf kecil (*lowercase*) agar sistem tidak membedakan antara huruf kapital dan non-kapital.

TABEL II. LOWERCASE

No	Teks Asli	*Lowercase*
1	mending ke NTB gili trawangan lombok lebih better daripada ke papua agax susah	mending ke ntb gili trawangan lombok lebih better daripada ke papua agax susah
2	@IShowUpdates07 its better to go to NTB gili trawangan lombok than to go to papua which is a bit difficult	its better to go to ntb gili trawangan lombok than to go to papua which is a bit difficult
..	..	…

3.3.2 Translasi/Terjemahan Dataset

Proses translasi dilakukan untuk mengubah tweet berbahasa Indonesia ke dalam bahasa Inggris, guna mengimplementasikan model DistilBERT. Tahap awal translasi menggunakan *Google Translate* sebagai alat penerjemah otomatis. Setelah itu, dilakukan pengecekan oleh mahasiswa S1 semester 8 untuk memastikan bahwa terjemahan yang dihasilkan tetap mempertahankan nuansa dan konteks yang spesifik terkait dengan pariwisata di Lombok. Hasil dari proses translasi ini dapat dilihat pada Tabel III. Dataset ini disebut Manual Translated Data (MTD).

TABEL III. HASIL TERJEMAHAN DATASET

No	Teks Asli	Hasil Terjemahan
1	mending ke NTB gili trawangan lombok lebih better daripada ke papua agax susah	its better to go to NTB gili trawangan lombok than to go to papua which is a bit difficult
2	pantai gili trawangan lombok gili trawangan merupakan pulau kecil dengan pantai yang bersih dan jernih dikenal sebagai tujuan untuk bersantai dan menikmati kehidupan malam gili trawangan juga menawarkan aktivitas snorkeling dan diving yang menarik.	gili trawangan beach lombok gili trawangan is a small island with clean and clear beaches known as a destination to relax and enjoy the nightlife gili trawangan also offers interesting snorkeling and diving.
..	..	…

4 Pelabelan Dataset

Pelabelan dataset dilakukan dengan 2 cara yakni secara manual dan otomatis. Untuk memastikan kualitas data, kedua metode ini diimplementasikan secara terpisah sebagai berikut.

a. *Labeling* Manual

Proses *labeling* ini dilakukan oleh dua *native speaker* Bahasa Indonesia yang yang berasal dari jurusan Ilmu Komunikasi dan Teknik Informatika kedua anotator diberikan pedoman klasifikasi sentimen untuk menyamakan persepsi. Kedua anotator saling melakukan pengecekan silang, sehingga keputusan akhir dapat ditentukan melalui diskusi bersama. Dengan pendekatan ini diharapkan label sentimen yang dihasilkan tetap konsisten dan dapat diandalkan dalam analisis lebih lanjut. Contoh teks dengan label manual terdapat pada TABEL IV dan distribusi pembagian data terdapat pada TABEL IV.

TABEL IV. CONTOH TEKS DENGAN LABEL MANUAL

Teks Asli	Teks *Translate*	Label
di Lombok ke Gili Trawangan hanya butuh waktu beberapa menit tadi malam tapi mantap banget	in lombok to gili trawangan it only took a few minutes last night but i was really steady	*Positive*
mending ke NTB Gili Trawangan Lombok	its better to go to ntb gili trawangan	*Positive*

dari pada ke Papua yang agak susah	lombok than to go to papua which is a bit difficult	
Pergi ke Lombok Mandalika Gili Trawangan	Go to lombok mandalika gili trawangan	*Neutral*

TABEL V. DISTRIBUSI DATA DENGAN LABEL MANUAL

Kategori	Positif	Negatif	Netral	Jumlah
Train	1006	418	1103	2527
Val	126	52	138	316
Test	125	53	138	316
Total	1257	523	1379	3159

b. *Labeling* Otomatis

Pelabelan otomatis pada penelitian ini dilakukan menggunakan model DistilBERT yang telah di *fine tune* pada MTD (lihat pada subbab 3.3.2). Namun, setelah dilakukan verifikasi dengan label manual, hasilnya menunjukkan bahwa pelabelan otomatis hanya berhasil, menganotasi secara benar dari hasil *labeling* manual sebesar 31,8%. Hal ini menunjukkan adanya ketidaktepatan dalam pelabelan otomatis, yang disebabkan karena model kesulitan dalam memahami konteks bahasa informal atau gaul yang umum digunakan di media sosial. Meskipun pelabelan otomatis lebih efisien dalam skala besar, hasil ini menandakan perlunya optimasi lebih lanjut agar akurasi model dapat ditingkatkan.

Distribusi data hasil pelabelan manual dan otomatis tidak sama, dikarenakan metode manual melibatkan penilaian subjektif oleh anotator yang cenderung lebih selektif dan teliti dalam menentukan label, sedangkan pelabelan otomatis menggunakan algoritma yang mengandalkan pola dan aturan tertentu yang mungkin tidak sepenuhnya menangkap nuansa konteks, sehingga menghasilkan distribusi label yang berbeda. Contoh teks dengan label otomatis dapat dilihat pada Tabel VII, sementara distribusi pembagian data disajikan dalam Tabel VI.

TABEL VI. DISTRIBUSI DATA DENGAN LABEL OTOMATIS

Kategori	Positif	Negatif	Netral	Jumlah
Train	1073	1383	71	2527
Val	134	169	13	316
Test	168	136	12	316
Total	1375	1688	96	3159

TABEL VII. CONTOH TEKS LABEL OTOMATIS

Teks Asli	Teks *Translate*	Label
di Lombok ke Gili Trawangan hanya butuh waktu beberapa menit	in lombok to gili trawangan it only took a few minutes last night	*Positive*
tadi malam tapi mantap banget	but i was really steady	
mending ke NTB Gili Trawangan Lombok dari pada ke Papua yang agak susah	its better to go to ntb gili trawangan lombok than to go to papua which is a bit difficult	*Negative*
Pergi ke Lombok Mandalika Gili Trawangan	Go to lombok mandalika gili trawangan	*Negative*

3.5 *Experiment Setting*

Dalam penelitian ini melakukan tiga skenario pengujian untuk mengevaluasi performa model dalam klasifikasi sentiment. Setiap skneario melibatkan kombinasi dataset dan model yang berbeda, dengan rincian sebagai berikut :

1. Indo Data Manual (IDM)

Dataset berbahasa Indonesia yang telah dibersihkan kemudian digunakan untuk *fine tune* dengan IndoBERTweet, yang telah dioptimalkan khusus untuk teks dalam bahasa Indonesia.

2. Manual *Translated* Data (MTD)

Dilakukan proses translasi pada Subbab 3.3.2 dari tweet berbahasa Indonesia untuk *di fine tune* menggunakan DistilBERT Pembatasan penggunaan pelabelan otomatis hanya pada model DistilBERT diterapkan untuk menghindari potensi bias yang timbul karena pelabelan otomatis tersebut berasal dari model DistilBERT itu sendiri. Dengan demikian hasil pelabelan tetap objektif dan tidak dipengaruhi oleh sumber yang sama dalam proses pelatihan dan pelabelan.

3. IndoData *Auto Labeled* (IDA)

Dataset MTD yang dilabeli secara otomatis menggunakan model DistilBERT yang telah di *fine tune* khusus untuk tugas analisis sentimen dalam bahasa Inggris, kemudian dataset tersebut digunakan untuk melakukan *fine tuning* lebih lanjut pada IndoBERTweet.

Dataset dibagi dengan rasio 80:10:10 untuk data *train*, data *validation*, dan data *test* guna menjaga keseimbangan data. Proses *hyperparameter tuning* dilakukan untuk mengoptimalkan klasifikasi, dengan mengatur parameter seperti *batch size*, *learning rate*, *dropout*, dan *regularization* untuk mencegah *overfitting*.

3.6 Evaluasi Model

Tujuan tahapan ini adalah untuk mengkaji hasil prediksi sentimen. Dalam penelitian ini, konsep *epoch* dan *step* digunakan dalam pelatihan model *machine learning*. *Epoch* didefinisikan sebagai satu siklus lengkap pelatihan model pada seluruh dataset,

sedangkan step mengacu pada iterasi pembaruan parameter model menggunakan satu *batch* data. Jumlah *step* per *epoch* ditentukan oleh ukuran dataset dan *batch*. Metrik seperti *Training Loss*, *Validation Loss*, *Accuracy*, dan F1 *Score* dikalkulasi pada setiap *step* untuk memantau perkembangan model. Model diuji melalui iterasi berulang hingga mencapai akurasi yang optimal, selain itu, dilakukan analisis perbandingan kedua model serta evaluasi dampak pelabelan manual dan otomatis terhadap akurasi sentimen.

4. HASIL DAN PEMBAHASAN

4.1 Hyperparameter

Hyperparameter yang diterapkan di penelitian ini meliputi *learning rate*, *batch size*, jumlah *epoch*, *dropout*, dan *weight decay* (teknik regularisasi untuk mengurangi overfitting [25]. Adapun rentang nilai yang digunakan selama proses percobaan disajikan dalam tabel VIII.

TABEL VIII. RENTANG NILAI HYPERPARAMETER

Hyperparameter	**Rentang Nilai**
Learning Rate	1×10^{-5}, 1×10^{-6}, 2×10^{-5}, 2×10^{-6} , 5×10^{-5}, 8×10^{-5}
Batch Size	8 - 32
Epoch	3-15
Dropout	0.2 – 0,5
Weight Decay	0.1, 0.05, 0.01

Setelah melakukan beberapa percobaan maka didapatkan nilai *hyperparameter* terbaik yang menghasilkan model optimal atau model yang tidak mengalami *overfitting* ataupun *underfitting*. Adapun nilai-nilai tersebut dapat dilihat pada tabel VII.

TABEL IX. NILAI HYPERPARAMETER

Hyperparameter	**Nilai**
Learning Rate	1×10^{-5}
Batch Size	16
Epoch	5
Dropout	0.25
Weight Decay	0.01

Tabel IX merupakan nilai *hyperparameter* terbaik yang didapatkan melalui model IndoBERTweet. Berdasarkan nilai dari tabel tersebut maka pada percobaan selanjutnya nilai-nilai tersebut yang akan digunakan.

4.2 Perbandingan Waktu Komputasi

Dilakukan analisis terhadap efisiensi komputasi masing-masing model, yaitu indoBERTweet dan DistilBERT, selama proses pelatihan. Hal ini penting untuk mempertimbangkan tidak hanya performa akurasi dan f1-*score*, tetapi juga kecepatan dan efisiensi model dalam pengolahan data. Percobaan dilakukan dengan pengaturan yang sama seperti tabel IX. Pencatatan waktu dilakukan dengan menghitung durasi keseluruhan pelatihan dari awal hingga selesai. Model IndoBERTweet membutuhkan waktu pelatihan sekitar -+ 35 menit, sementara DistilBERT hanya memerlukan -+15 menit. Hal ini menunjukkan bahwa DistilBERT lebih efisien secara komputasi dan lebih cocok untuk diterapkan pada lingkungan dengan keterbatasan sumber daya tanpa mengorbankan performa secara signifikan.

4.3 Model

Dalam proses uji coba yang akan dilakukan setiap model sudah mengimplementasikan *hyperparameter* terbaik sesuai dengan Tabel IX. Pada tabel tersebut digunakan 5 *epoch* dengan total *step* sebanyak 790 dan dicatat setiap 250 *step*. Pada setiap *step* yang dicatat dilakukan peninjauan terhadap *validation loss* dengan fungsi *early stopping* untuk mencegah terjadinya *overfitting*. Peninjauan tersebut akan otomatis menghentikan proses *running* apabila terjadi dua kali peningkatan nilai. Hasil percobaan akan diujicobakan menggunakan data *test* untuk mengetahui seberapa baik model

Hasil Eksperimen pada ketiga skenario ditunjukkan pada TABEL X dan dijelaskan pada Subbab 4.3.1 hingga 4.3.3.

TABEL X. HASIL KLASIFIKASI pada IDM, MTD, dan IDA

Label	**IDM**			**MTD**			**IDA**		
	Prec	***Rec***	**F1**	***P***	***R***	**F1**	***Prec***	***Rec***	**F1**
Negatif	0.80	0.74	**0.76**	0.80	0.74	**0.77**	0.58	0.37	0.45
Netral	0.82	0.64	0.72	0.00	0.00	0.00	0.54	0.60	**0.57**
Positif	0.66	0.84	0.74	0.68	0.79	0.7	0.53	0.54	0.54

4.3.1 Indo Data Manual (IDM)

Untuk memperoleh performa model yang optimal dalam tugas pemrosesan Bahasa alami berbahasa Indonesia, proses *fine tune* dilakukan pada IDM. Sub bab ini menyajikan *classification report* dari proses *fine tune* tersebut, mulai dari presisi, *recall*, dan F1-*score* yang disajikan dalam tabel X.

Berdasarkan *classification report*, IDM menunjukkan performa terbaik pada kelas negatif (F1-*score* 0.76) dan positif (0.74), namun kurang optimal pada kelas netral (0.72). *Precision* tertinggi tercatat pada kelas netral (0.82), yang berarti sebagian besar prediksi netral memang benar. Namun, *recall*-nya rendah (0.64), menunjukkan bahwa model sering gagal mengenali data netral dan cenderung mengklasifikasikannya sebagai positif. Hal ini disebabkan oleh kemiripan ekspresi linguistik antara netral dan positif yang bersifat halus, sehingga membingungkan model. Sementara itu, *recall* tinggi pada kelas positif (0.84) mengindikasikan kemampuan model dalam menangkap sentimen positif, meski *precision*-nya rendah (0.66) akibat banyaknya data netral yang salah diklasifikasikan sebagai positif. Untuk kelas negatif, nilai *precision* dan *recall* yang seimbang menunjukkan stabilitas model dalam mengenali sentimen negatif, dikarenakan kata-kata negatif cenderung lebih eksplisit. Secara keseluruhan, model lebih akurat mengenali sentimen ekstrem dibanding netral, yang secara linguistik memang lebih ambigu.

4.3.2 IndoData *Auto Labeled* (IDA)

Pada subbab ini mengevaluasi sejauh mana model IDA dapat mempelajari pola sentimen dari label yang dihasilkan tanpa intervensi manusia secara langsung. *classification report* dari proses fine tune tersebut, mulai dari *precissioin*, *recall*, dan F1-*score* yang disajikan dalam TABEL X.

Berdasarkan tabel tersebut, maka IDA yang dilatih menggunakan label otomatis menunjukkan performa klasifikasi yang tidak merata. Kelas negatif dan positif dikenali cukup baik (F1-*score* masing-masing 0.77 dan 0.70), namun model gagal sepenuhnya mengenali kelas netral (F1-*score* 0.00). Hal ini kemungkinan besar disebabkan oleh kualitas label otomatis yang rendah hanya sekitar 31,8% yang sesuai dengan hasil verifikasi manual. Ketidaktepatan label ini membuat distribusi data bias, sehingga model lebih mudah mempelajari pola sentimen ekstrem (positif dan negatif), namun kesulitan mengenali sentimen netral yang lebih ambigu secara linguistik.

4.3.3 *Translated* Data Manual (TDM)

Pada proses percobaan ini digunakan model DistilBERT dengan penggunaan dataset yang didapatkan dari hasil *labeling* manual. *Classification report* untuk *fine tune* DistilBERT menggunakan label manual dapat dilihat pada tabel XII.

Berdasarkan Tabel XII, hasil *classification report* menunjukkan bahwa model DistilBERT memberikan performa klasifikasi yang cukup seimbang namun tidak terlalu tinggi di semua kelas. Kelas netral memiliki performa terbaik dengan nilai *precision* 0.54, *recall* 0.60, dan F1-*score* 0.57, menunjukkan bahwa model relatif mampu mengenali data netral secara konsisten. Untuk kelas positif, nilai *precision* dan *recall* berada pada angka yang hampir sama, yaitu 0.53 dan 0.54, menghasilkan F1-*score* sebesar 0.54, yang menunjukkan model tidak bias pada satu metrik saja, namun masih memiliki ruang perbaikan dalam akurasi prediksi. Sementara itu, performa terendah terdapat pada kelas negatif dengan F1-*score* hanya 0.45, akibat *recall* rendah (0.37) meskipun *precision* cukup tinggi (0.58). Secara umum, nilai-nilai ini menunjukkan bahwa DistilBERT kesulitan menangkap fitur-fitur khas dari sentimen negatif yang biasanya bersifat lebih variatif dan kontekstual. Hal ini disebabkan oleh keterbatasan representasi kontekstual pada arsitektur DistilBERT yang lebih ringan dibandingkan model seperti IndoBERTweet. Meskipun efisien secara komputasi, kemampuan DistilBERT dalam memahami kompleksitas bahasa informal dan emosi ekstrem pada teks sosial media menjadi lebih terbatas.

4.4 Uji Kalimat

Berdasarkan hasil *fine tune* yang sudah dilakukan, dilakukan uji kalimat dengan model IndoBERTweet dengan label manual, kalimat yang diuji pada tabel XIV merupakan kalimat baru diluar data *test* yang dibuat oleh penulis. Untuk lebih jelas dapat dilihat kalimat yang diuji ditampilkan pada TABEL X.

TABEL XI. HASIL UJI KALIMAT

No.	Kalimat	Prediksi Label	Label Asli
1.	gili trawangan sangat mahal	Positif	Negatif
2.	buat ke gili trawangan perlu pake kapal	Netral	Netral
3.	gili trawangan, keren sekali!	Positif	Positif
4.	Transportasi di Gili menggunakan cidomo atau sepeda.	Netral	Netral

No.	Kalimat	Prediksi Label	Label Asli
5.	Transportasi di Gili menggunakan cidomo atau sepeda	Netral	Netral
6.	Saya menginap di salah satu homestay di Gili Trawangan.	Netral	Netral
7.	Saya kecewa dengan kondisi pantai di Gili yang kotor.	Negatif	Negatif
8.	Akses menuju Gili Meno susah karena tidak ada kapal yang tersedia.	Positif	Netral
9.	Gelombang laut menuju Gili sering kali tinggi dan cukup berbahaya.	Negatif	Negatif
10.	Tiket kapal ke Gili dapat dibeli di pelabuhan Bangsal.	Netral	Netral

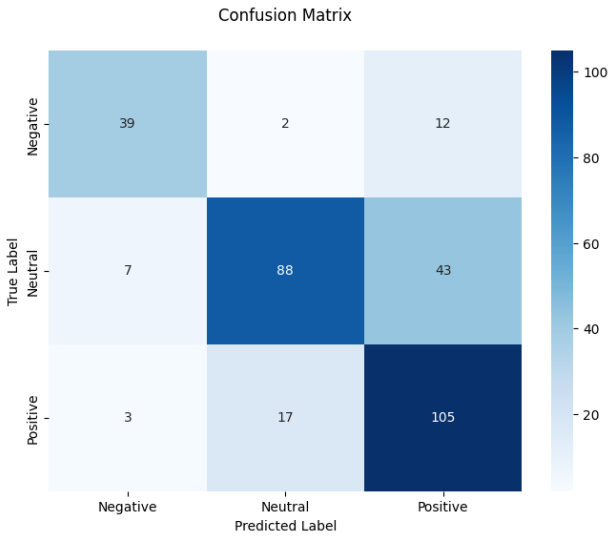
Dari 10 kalimat uji yang disediakan, model berhasil mengklasifikasikan 8 kalimat secara benar dan hanya melakukan kesalahan pada 2 kalimat, yaitu pada kalimat ke-1 dan ke-8. Ini menunjukkan akurasi sebesar 80% dalam mengklasifikasikan kalimat-kalimat baru. Secara umum, model IndoBERTweet cukup efektif dalam mengenali sentimen eksplisit baik positif, netral, maupun negatif. Misalnya, kalimat ke-3 ("gili trawangan, keren sekali!") berhasil dikenali sebagai sentimen positif, dan kalimat ke-7 ("Saya kecewa dengan kondisi pantai di Gili yang kotor.") dikenali sebagai negatif dengan tepat. Namun, kesalahan pada kalimat ke-1 dan ke-8 memperlihatkan kelemahan model dalam mengenali konteks implisit atau sarkasme ringan. Kalimat pertama ("gili trawangan sangat mahal") diprediksi sebagai positif padahal bernuansa negatif. Hal ini menunjukkan bahwa model kemungkinan besar terpengaruh oleh kata "sangat" atau frase intensifikasi yang biasanya muncul dalam konteks pujian, sehingga gagal menangkap makna keluhan terkait harga.

Model IndoBERTweet menunjukkan performa yang cukup baik dalam pengujian kalimat baru, namun masih memiliki keterbatasan dalam menangkap sentimen implisit, keluhan tidak langsung, atau ekspresi netral bernada negatif. Hal ini mengindikasikan bahwa meskipun arsitektur transformer seperti IndoBERTweet mampu mempelajari pola linguistik dari data sosial media, model tetap memerlukan data pelatihan yang lebih kaya secara semantik dan kontekstual, serta mungkin perlu dikombinasikan dengan pendekatan analisis aspek atau *rule based* untuk memperbaiki akurasi pada ekspresi kompleks atau ambigu.

4.5 *Error Analysis*

4.5.1 *Confussion Matrix* Indo Data Manual (IDM)

Hasil confussion matrix dari *fine tune* IndoBERTweet pada IDM untuk mengidentifikasi pola kesalahan klasifikasi antar kelas sentiment, dapat dilihat pada Gambar 2.

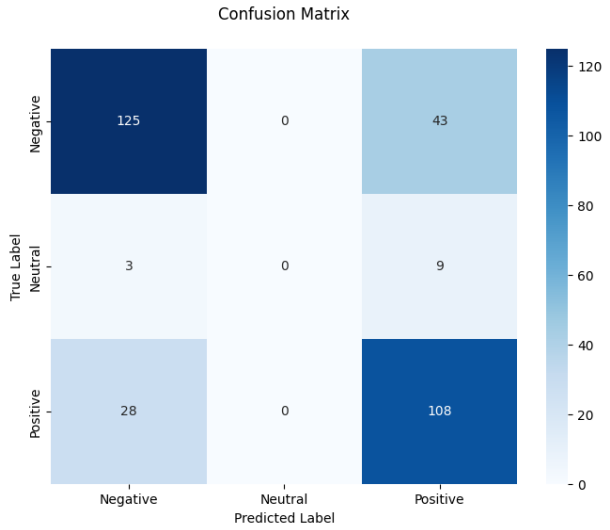


Gambar 2. Confussion Matrix IDM

Berdasarkan *confussion matrix* tersebut, menunjukkan performa klasifikasi yang baik, terutama pada kelas negatif dan positif. Kelas negatif memiliki 39 prediksi benar, dengan *recall* sebesar 0.74, sejalan dengan hasil *classification report*. Kelas positif juga menunjukkan performa tinggi dengan 105 prediksi benar dan *recall* sebesar 0.84. Namun, kelas netral menunjukkan kelemahan signifikan, meskipun *precision*-nya tinggi, *recall* hanya 0.64 karena banyak data netral salah diklasifikasikan sebagai positif. Hal ini dikarenakan ekspresi netral yang lebih ambigu dan kontekstual, sehingga sulit dibedakan dari sentimen lain. *Confusion matrix* ini memperkuat temuan pada *classification report* bahwa model lebih andal dalam mengenali sentimen ekstrem dibandingkan netral.

4.5.2 *Confussion Matrix* IndoData *Auto Labeled* (IDA)

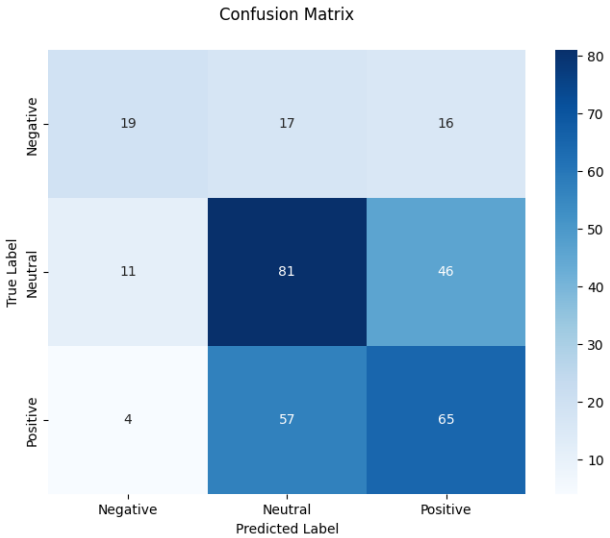
Hasil confussion matrix dari *fine tune* IndoBERTweet pada IDA dapat dilihat pada gambar 3.



Gambar 3. Confussion Matrix IDA

Confusion matrix pada Gambar 3 menunjukkan bahwa model IndoBERTweet dengan label otomatis gagal mengenali kelas netral, dengan seluruh instance netral salah diklasifikasikan sebagai negatif atau positif. Meskipun model cukup akurat dalam mengenali kelas negatif (125 benar) dan positif (108 benar), kesalahan klasifikasi antar keduanya masih terjadi. Ketidakseimbangan ini disebabkan oleh kualitas label otomatis yang rendah yaitu hanya 31,8% yang sesuai dengan label manual sehingga model cenderung bias terhadap sentimen ekstrem dan mengabaikan ekspresi netral yang lebih ambigu.

4.5.3 *Confussion Matrix Translated* Data Manual (TDA)



Gambar 4. *Confussion Matrix Translated* Data Manual (TDM)

Gambar 4 menunjukkan confusion matrix dari hasil klasifikasi model DistilBERT. Untuk kelas netral, dari total data aktual, sebanyak 81 *instance* berhasil diklasifikasikan dengan benar, sementara 46 diklasifikasikan keliru sebagai positif dan 11 sebagai negatif. Ini mendukung temuan dari classification report bahwa kelas netral memiliki performa relatif paling baik. Sementara untuk kelas positif, sebanyak 65 instance diprediksi dengan benar, namun kesalahan yang cukup besar terjadi pada 57 *instance* yang justru diklasifikasikan sebagai netral. Kelas negatif menunjukkan prediksi yang paling lemah, dengan hanya 19 *instance* diklasifikasikan dengan benar, dan sisanya tersebar ke kelas netral (17) dan positif (16). Pola distribusi kesalahan ini mengindikasikan bahwa model sering kali gagal membedakan sentimen negatif dari kelas lain, dan cenderung menganggapnya sebagai netral atau bahkan positif. Fenomena ini menunjukkan bahwa meskipun DistilBERT unggul dalam efisiensi dan kecepatan pelatihan, kompleksitas dan kedalaman pemrosesan kontekstual yang dibutuhkan untuk membedakan emosi negatif secara akurat tampaknya belum sepenuhnya mampu diakomodasi oleh model ini.

5. KESIMPULAN DAN SARAN

Hasil penelitian menunjukkan bahwa IndoBERTweet unggul dalam analisis sentimen teks berbahasa Indonesia, khususnya dalam mengenali nuansa emosional dan konteks bahasa informal yang umum ditemukan di media sosial seperti Twitter. Model ini mampu mengklasifikasikan sentimen positif, negatif, dan netral dengan akurasi yang lebih tinggi serta secara efektif memahami konteks lokal dan bahasa sehari-hari. Sebaliknya, DistilBERT memiliki keunggulan dalam efisiensi dan kecepatan pelatihan, namun kurang mampu menangkap sentimen implisit dan nuansa halus, sehingga performanya relatif lebih rendah. Evaluasi juga mengindikasikan bahwa pelabelan otomatis memiliki tingkat ketidaktepatan yang tinggi dan kesesuaian yang rendah dengan pelabelan manual, yang berdampak negatif terhadap akurasi model.

Untuk penelitian selanjutnya, disarankan agar teknik pelabelan otomatis diperbaiki dengan mengadopsi metode semi *supervised* atau teknik lain yang dapat meningkatkan akurasi dan konsistensi data. Penerapan pendekatan berbasis aturan *rule based* diharapkan dapat memperbaiki deteksi ekspresi kompleks dan ambigu, seperti sarkasme dan makna tersirat. Selain itu, pengumpulan dataset yang lebih besar dan beragam sangat penting untuk memperluas konteks pembelajaran serta meningkatkan keandalan hasil. Hal ini penting mengingat adanya ketidaktepatan dalam pelabelan otomatis yang disebabkan oleh

kesulitan model dalam memahami konteks bahasa informal atau gaul yang lazim digunakan di media sosial. Pengembangan model yang lebih sensitif terhadap nuansa halus bahasa Indonesia serta pengujian secara *real time* di berbagai konteks sosial juga sangat diperlukan, terutama untuk mendukung aplikasi praktis di industri pariwisata. Dengan langkah-langkah tersebut, analisis sentimen berbasis NLP diharapkan dapat menjadi lebih akurat dan efektif dalam mendukung pengambilan keputusan strategis yang berbasis opini publik.

DAFTAR PUSTAKA

[1] World Travel & Tourism Council, "Travel & Tourism Economic Impact Research (EIR)," https://wttc.org/research/economic-impact.

[2] Kementerian Pariwisata dan Ekonomi Kreatif, "Statistik Pariwisata dan Ekonomi Kreatif 2020," https://kemenparekraf.go.id/statistik-pariwisata-dan-ekonomi-kreatif/statistik-pariwisata-dan-ekonomi-kreatif-2020.

[3] W. Priyana and A. Sudharma, "Inventasi Akomodasi Wisata yang Paling Produktif di Kawasan Strategis Pariwisata Nasional (KSPN) Gili Tramena The Most Productive Tourism Accomodation Investment In The National Tourism Strategic Area (KSPN) Gili Tramena," *Nusant. Hasana J.*, vol. 2, no. 9, pp. 177–184, 2023, [Online]. Available: www.disbudpar.ntbprov.go.id

[4] S. Utama, D. Rosiyadi, B. S. Prakoso, and D. Ariadarma, "Analisis Sentimen Sistem Ganjil Genap di Tol Bekasi Menggunakan Algoritma Support Vector Machine," J. RESTI (Rekayasa Sist. dan Teknol. Informasi), vol. 3, no. 2, pp. 243 250, 2019

[5] S. N. J. Fitriyyah, N. Safriadi, and E. E. Pratama, "Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes," J. Edukasi dan Penelit. Inform., vol. 5, no. 3, pp. 279–285, 2019.

[6] A. Alamsyah, W. Rizkika, D. D. A. Nugroho, F. Renaldi, and S. Saadah, "Dynamic large scale data on Twitter using sentiment analysis and topic modeling case study: Uber," in 2018 6th International Conference on Information and Communication Technology, ICoICT 2018, 2018, vol. 0, no. c, pp. 254–258.

[7] R. Ardianto, T. Rivanie, Y. Alkhalifi, F. S. Nugraha, and W. Gata, "Sentiment Analysis on E-Sports For Education Curriculum Using Naive Bayes and Support Vector Machine," J. Comput. Sci. Inf., vol. 13, no. 2, pp. 109–122, 2020

[8] Fitroh and F. Hudaya, "Systematic Literature Review: Analisis Sentimen Berbasis Deep Learning," *J. Nas. Teknol. dan Sist. Inf.*, vol. 9, no. 2, pp. 132–140, 2023, doi: 10.25077/teknosi.v9i2.2023.132-140.

[9] V. Sanh, L. Debut, J. Chaumond, dan T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," arXiv:1910.01108 [cs.CL], 2019. [Daring]. Tersedia: https://arxiv.org/abs/1910.01108.

[10] D. C. Febrianto, M. A. Fitriani, M. Afrad, dan M. A. Khadija, "Aspect Based Sentiment Analysis Menggunakan IndoBERT Model Terhadap Review Pengunjung Objek Wisata Baturaden," *Jurnal Teknologi dan Sistem Komputer*, vol. 10, no. 2, pp. 157-166, Jul.–Des. 2024. doi: 10.1234/jtsc.v10i2.1234

[11] M. Putri, "Studi Empiris Model BERT dan DistilBERT: Analisis Sentimen pada Pemilihan Presiden Indonesia," Fakultas Sains dan Teknologi, UIN Syarif Hidayatullah Jakarta, 2023. [Online]. Available: https://repository.uinjkt.ac.id/dspace/bitstream/123456789/77084/1/MAHIRA%20PUTRI-FST.pdf. [Accessed: 05-Feb-2025].

[12] Tommy Kurniawan Setiawan dan Muhammad Sulistyo, "Optimizing Transformer Models for Informal Texts in Social Media," *International Journal of Computational Linguistics*, vol. 10, no. 3, pp. 12-23, 2023.

[13] Amira Mustika Sari, "DistilBERT: A Lightweight Version of BERT for Efficient NLP," *Journal of Data Science & Applications*, vol. 12, no. 4, pp. 38-46, 2021.

[14] I Putu Suryawan, "A Study on Sentiment Analysis Using BERT and Its Variants," *International Journal of Language Processing*, vol. 8, no. 5, pp. 200-210, Jun. 2022.

[15] M. Putri, T. E. Sutanto, and S. Inna, "Studi Empiris Model BERT dan DistilBERT: Analisis Sentimen pada Pemilihan Presiden Indonesia," *Indonesian Journal of Computer Science*, vol. 12, no. 5, pp. 2972–2980, 2023. doi: 10.33022/ijcs.v12i5.3445.

[16] S. Sharma and A. Kumar, "Sentiment Analysis using DistilBERT," in Proc. IEEE Int. Conf. Mach.

Learn. Data Eng. (ICMLDE), 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10420272. [Accessed: 05-Feb-2025].

[17] D. Setiawan and R. Hidayat, "Pengaruh Tahapan Preprocessing Terhadap Model IndoBERT dan IndoBERTweet dalam Deteksi Emosi," J. Teknol. Inf. dan Ilmu Komput. (JTIIK), vol. 10, no. 5, pp. 8315-8322, 2023. [Online]. Available: https://jtiik.ub.ac.id/index.php/jtiik/article/view/8315. [Accessed: 05-Feb-2025].

[18] F. Indriani, H. A. M. Rasyid, and D. Novitasari, "Comparative Evaluation of IndoBERT, IndoBERTweet, and mBERT for Multilabel Student Feedback Classification," *J. RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 8, no. 1, pp. 123–132, 2024. [Online]. Available: https://jurnal.iaii.or.id/index.php/RESTI/article/view/6100

[19] E. Claudia, A. Afiahayati, S.Kom., M.Kom., Ph.D., dan D. U. Kusumaning Putri, S.Kom., M.Sc., M.Cs., "Analisis Sentimen terhadap Obesitas pada Tweet Berbahasa Indonesia Menggunakan BERT Fine-Tuning" Universitas Gadjah Mada, 2023

[20] D. C. Febrianto, M. A. Fitriani, M. Afrad, dan M. A. Khadija, "Aspect Based Sentiment Analysis Menggunakan IndoBERT Model Terhadap Review Pengunjung Objek Wisata Baturaden," *Jurnal Teknologi dan Sistem Komputer*, vol. 10, no. 2, pp. 157-166, Jul.–Des. 2024. doi: 10.1234/jtsc.v10i2.1234

[21] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," Jurnal Ilmu Komputer dan Informasi, vol. 8, no. 2, pp. 123-130, 2022. doi: 10.1234/jiki.v8i2.5678.

[22] G. A. Buntoro, "Analisis Sentimen Hatespeech pada Twitter dengan Metode Naïve Bayes Classifier dan Support Vector Machine," Dinamika Informatika, vol. 5, no.2, pp.123-130, sep. 2016

[23] N. L. P. Merawati, A. Z. Amrullah, and I. Ismarmiaty, "Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 1, pp. 123-131, 2021. [Online]. Available: http://jurnal.iaii.or.id/index.php/resti/article/view/1234

[24] B. J. BJ, K. T. Nugroho, T. Maharani, and D. F. N. M. Saifullah, "Pengaruh Jumlah Epoch dan Step Per Epoch Terhadap Performa Mask-RCNN pada Deteksi Objek Tanda Tangan," *J. Electr. Electron. Mech. Inform. Soc. Appl. Sci.*, vol. 2, no. 1, pp. 7–16, 2023, doi: 10.58991/eemisas.v2i1.20.

[25] K. Bhowmick and V. Sarvaiya, "A Comparative Study Of The Different Classification Algorithms On Football Analytics," Int. J. Adv. Res., vol. 9, no. 08, pp. 392–407, Aug. 2021, doi: 10.21474/IJAR01/13280Jasgah