

Sentiment Classification of Health Education YouTube Comments Using IndoBERT Embeddings with Logistic Regression and Naïve Bayes

Andre Septa Wijaya^{1*}, Amiq Fahmi^{2*}, Yuventius Tyas Catur Pramudi^{3*}

*Informatin System Program, Faculty of Computer Science, University of Dian Nuswantoro, Semarang 50131
112202206859@mhs.dinus.ac.id¹, amiq_fahmi@dosen.dinus.ac.id², tyas.catur@dsn.dinus.ac.id³

Article Info

Article history:

Received 2026-05-16

Revised 2026-05-28

Accepted 2026-06-11

Keyword:

Sentiment analysis,
IndoBERT,
SMOTE,
Logistic Regression,
Naïve Bayes,
Text Clasification.

ABSTRACT

Class imbalance is a common issue in sentiment classification of social media data, particularly in mental health-related discussions where certain sentiment classes are underrepresented. This study focuses on sentiment classification of mental health-related YouTube comments by utilizing IndoBERT as a pre-trained language model to generate contextual text embeddings. Sentiment classification is subsequently performed using conventional machine learning algorithms, namely Logistic Regression and Naïve Bayes. The research framework includes data collection through the YouTube Data API, text preprocessing, semi-manual sentiment labeling into positive, neutral, and negative classes, and dataset partitioning using an 80:20 train-test split. To address class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) is applied exclusively to the training data to prevent data leakage. Feature representation is obtained from IndoBERT embeddings with a dimensionality of 768. Model performance is evaluated using accuracy, precision, recall, and F1-score. Experimental results show that Logistic Regression outperforms Naïve Bayes, achieving an accuracy of 78%, compared to 56% for Naïve Bayes. This indicates that Logistic Regression is more effective in handling dense contextual embeddings generated by transformer-based models. Overall, the findings demonstrate that combining contextual embeddings with data balancing techniques can improve sentiment classification performance in mental health-related social media analysis, particularly in low-resource language settings.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Perkembangan teknologi digital telah menstimulasi peningkatan pemanfaatan media sosial sebagai sarana penyebaran sekaligus pertukaran informasi, termasuk di bidang kesehatan. YouTube menjadi salah satu platform yang banyak digunakan masyarakat karena aksesnya mudah dan penyajian informasi dalam format audiovisual [1]. Selain berfungsi sebagai media distribusi, fitur komentar pada YouTube juga menyediakan ruang interaksi antar pengguna yang merefleksikan opini publik terhadap suatu konten [2]. Tingginya aktivitas pengguna menghasilkan data teks dalam jumlah besar yang dapat dimanfaatkan untuk memahami pandangan masyarakat terhadap berbagai isu, baik berupa sentimen positif, negatif, maupun netral. Kondisi ini menegaskan perlunya metode otomatis untuk mengenali dan

mengelompokkan opini melalui analisis sentimen. Analisis sentimen sendiri merupakan cabang *Natural Language Processing* (NLP) yang berfokus pada ekstraksi, pemahaman, dan klasifikasi opini dari data berbasis teks [3]. Hal ini menunjukkan bahwa media sosial tidak hanya berfungsi sebagai saluran informasi, tetapi juga sebagai cerminan dinamika opini masyarakat terhadap isu kesehatan.

Metode klasifikasi konvensional seperti Naïve Bayes dan Logistic Regression tetap populer dalam penerapan analisis sentimen karena menawarkan efisiensi komputasi yang tinggi serta kemampuan dalam mengelola data berdimensi besar [4]. Namun, pendekatan representasi teks tradisional seperti *Bag-of-Words* dan *Term Frequency-Inverse Document Frequency* (TF-IDF) memiliki keterbatasan dalam menangkap konteks linguistik secara mendalam, khususnya pada teks media sosial yang cenderung bersifat informal dan ambigu [5].

Kemajuan model berbasis *transformer*, seperti BERT, telah menghasilkan peningkatan performa yang signifikan dalam berbagai tugas NLP. Pada bahasa Indonesia, IndoBERT terbukti mampu membangun representasi teks yang lebih kontekstual sehingga meningkatkan akurasi klasifikasi sentimen dibandingkan pendekatan representasi tradisional [6],[7]. Penelitian terkini juga menegaskan bahwa model *transformer* berbahasa Indonesia, seperti IndoRoBERTa dan NusaBERT, memiliki kemampuan yang kuat dalam memahami konteks linguistik serta hubungan semantik pada data media sosial berbahasa Indonesia [8].

Berbagai penelitian terdahulu menegaskan bahwa pendekatan berbasis *transformer* mampu meningkatkan kinerja analisis sentimen pada beragam kasus di Indonesia [9],[10],[11]. Integrasi representasi teks berbasis BERT dengan algoritma *machine learning* konvensional juga terbukti efektif dalam memperkuat kualitas klasifikasi sentimen multi-kelas pada data media sosial [12],[13]. Selain itu, analisis sentimen telah banyak diterapkan pada data media sosial pemerintah maupun layanan publik sebagai sarana untuk memahami respons masyarakat terhadap kebijakan dan informasi publik [14].

Penelitian lain mengungkapkan bahwa penerapan IndoBERT tidak hanya efektif untuk klasifikasi sentimen umum, tetapi juga relevan pada *aspect-based sentiment analysis* serta analisis isu sosial dan politik di media sosial Indonesia [15]. Pendekatan *Explainable Artificial Intelligence* (XAI), seperti LIME, mulai digunakan untuk meningkatkan interpretabilitas hasil klasifikasi dari model *transformer* [16]. Hasil penelitian tersebut menunjukkan bahwa representasi teks kontekstual membantu model memahami hubungan semantik dan makna kalimat secara lebih mendalam.

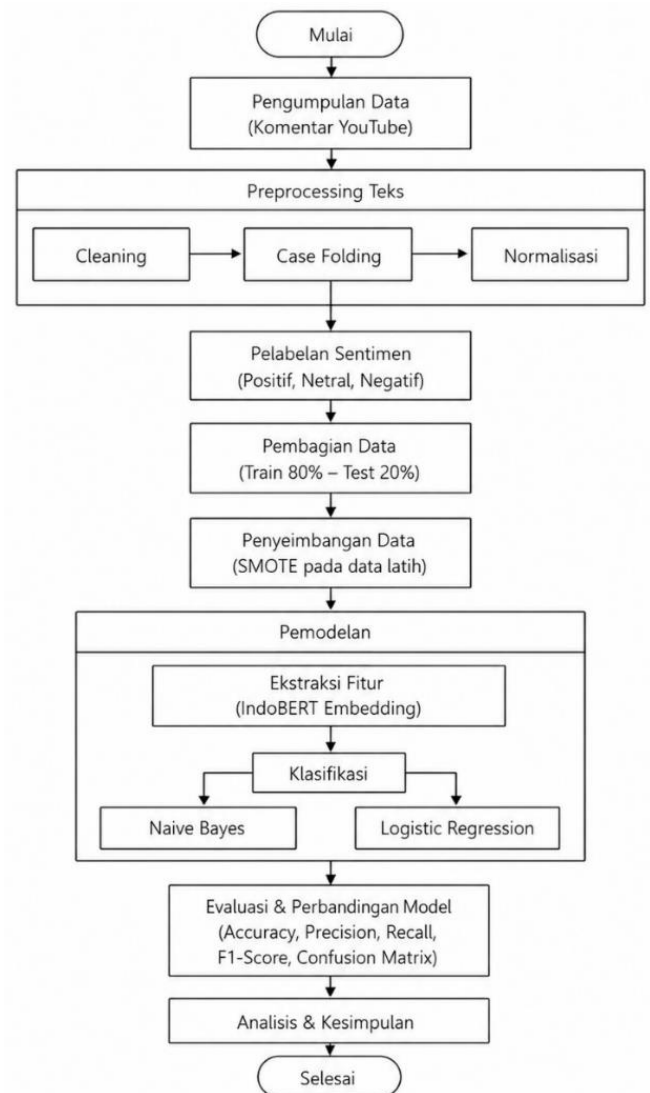
Selain pemilihan metode representasi teks, ketidakseimbangan distribusi data merupakan tantangan penting dalam proses klasifikasi sentimen. Distribusi yang tidak proporsional sering membuat model lebih cenderung memprediksi kelas mayoritas, sehingga performa pada kelas minoritas menjadi kurang optimal. Untuk mengatasi permasalahan tersebut, diperlukan pendekatan khusus seperti *Synthetic Minority Over-sampling Technique* (SMOTE) [17] yang berfungsi memperkuat representasi kelas minoritas selama tahap pelatihan model.

Berdasarkan kajian penelitian sebelumnya, sebagian besar studi berfokus pada proses *fine-tuning* model *transformer* secara langsung. Penelitian yang memanfaatkan IndoBERT sebagai ekstraktor fitur dan dikombinasikan dengan algoritma klasifikasi konvensional seperti Naïve Bayes dan Logistic Regression masih relatif terbatas, khususnya dalam konteks komentar YouTube terkait edukasi kesehatan. Selain itu, studi yang menyoroti klasifikasi sentimen multi-kelas dengan mempertimbangkan ketidakseimbangan distribusi data juga belum banyak ditemukan.

Kontribusi utama penelitian ini terletak pada integrasi IndoBERT sebagai ekstraktor fitur dengan algoritma klasifikasi tradisional Naïve Bayes dan Logistic Regression,

yang dipadukan dengan penerapan SMOTE untuk mengatasi ketidakseimbangan distribusi data. Pendekatan ini secara khusus ditujukan untuk meningkatkan akurasi klasifikasi sentimen multi-kelas pada komentar YouTube yang berkaitan dengan konten edukasi kesehatan. Dengan demikian, penelitian ini menawarkan kombinasi representasi berbasis *deep learning* dan algoritma konvensional yang belum banyak dieksplorasi, sekaligus memberikan kontribusi praktis dalam pemantauan opini publik di ranah kesehatan digital.

II. METODE



Gambar 1. Skema Alur Penelitian Analisis Sentimen

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk membandingkan performa algoritma Naïve Bayes dan Logistic Regression dalam klasifikasi sentimen komentar YouTube terkait konten edukasi kesehatan. Alur penelitian dilaksanakan secara terstruktur, meliputi pengumpulan data, praproses teks, pelabelan sentimen, pembagian data latih dan uji,

penyeimbangan kelas dengan SMOTE, ekstraksi fitur menggunakan IndoBERT, klasifikasi dengan dua algoritma, serta evaluasi model menggunakan metrik akurasi, presisi, recall, dan F1-score. Keseluruhan tahapan penelitian dirangkum pada Gambar 1.

A. Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa komentar pengguna pada platform YouTube yang berkaitan dengan konten edukasi kesehatan. Pemilihan YouTube sebagai sumber data didasarkan pada tingginya intensitas interaksi melalui kolom komentar, sehingga mampu merepresentasikan opini, persepsi, dan respons masyarakat terhadap informasi kesehatan yang disajikan dalam video. Dengan demikian, YouTube dipandang sebagai media yang relevan untuk menggambarkan dinamika opini publik di ranah kesehatan digital. Data komentar dikumpulkan dari tiga kanal YouTube yang secara konsisten menyajikan konten edukasi dan informasi kesehatan, yaitu Dr_tirta, Kok Bisa, dan Alodokter. Video yang dijadikan sumber data dipilih berdasarkan relevansinya terhadap topik kesehatan mental, sehingga komentar yang dikumpulkan merepresentasikan wacana publik yang sesuai dengan fokus penelitian.

Proses pengumpulan data dilakukan pada bulan Maret 2026 dengan memanfaatkan YouTube Data API untuk mengekstraksi komentar dari sejumlah video yang relevan dengan topik edukasi kesehatan mental. Setelah proses pengumpulan data, diperoleh sebanyak 2.918 komentar. Selanjutnya dilakukan tahap preprocessing dan filtering data, meliputi penghapusan komentar duplikat, komentar kosong, komentar nonbahasa Indonesia, serta penyaringan panjang teks komentar untuk mengurangi noise dan menjaga konsistensi data. Setelah proses pembersihan data, diperoleh sebanyak 1.848 komentar valid yang digunakan dalam proses pelabelan sentimen dan analisis sentimen.

Distribusi awal dataset menunjukkan adanya ketidakseimbangan antar kelas, di mana komentar dengan sentimen negatif lebih dominan dibandingkan sentimen positif maupun netral. Kondisi ini lazim terjadi dalam analisis sentimen media sosial karena pengguna cenderung lebih aktif menyampaikan kritik atau keluhan dibandingkan apresiasi. Ketidakseimbangan distribusi data tersebut berpotensi menurunkan performa model klasifikasi, sehingga diperlukan penanganan khusus melalui teknik oversampling.

Data yang telah dikumpulkan kemudian diproses lebih lanjut melalui tahap preprocessing untuk meningkatkan kualitas teks sebelum digunakan dalam analisis sentimen. Tahap ini menjadi krusial karena kualitas data yang baik akan mendukung representasi kontekstual IndoBERT dan meningkatkan akurasi klasifikasi pada tahap berikutnya. Seluruh proses pengolahan dilakukan dengan tetap memperhatikan aspek etika penelitian, yakni hanya menggunakan komentar yang bersifat publik, mengabaikan identitas personal, serta menjaga kerahasiaan pengguna agar analisis tetap sesuai dengan kaidah akademik dan tidak menyinggung individu tertentu.

B. Preprocessing Teks

Tahap preprocessing dilakukan untuk membersihkan sekaligus menyiapkan data teks sebelum dianalisis dalam penelitian sentimen. Komentar yang diperoleh dari platform YouTube umumnya mengandung berbagai elemen noisy yang tidak relevan, seperti simbol, tanda baca berlebihan, variasi penulisan kata yang tidak konsisten, serta duplikasi data yang berpotensi menurunkan kinerja model klasifikasi. Oleh karena itu, preprocessing diperlukan agar representasi teks menjadi lebih terstruktur, konsisten, dan siap digunakan pada tahap analisis berikutnya [18].

Secara umum, tahapan preprocessing dapat dirumuskan melalui persamaan (1).

$$T' = f(T) \quad (1)$$

Keterangan :

T : teks mentah (*raw text*)

f : fungsi *preprocessing*

T' : teks hasil *preprocessing*

Persamaan tersebut menunjukkan bahwa teks mentah T diproses melalui serangkaian tahapan preprocessing f sehingga menghasilkan teks baru T' yang lebih bersih dan terstruktur. Tahapan ini berperan penting dalam meningkatkan kualitas representasi fitur sekaligus mengurangi noise yang tidak relevan pada proses analisis sentimen [19].

Tahapan *preprocessing* yang diterapkan dalam penelitian ini meliputi:

1) *Cleaning*

Proses *cleaning* dilakukan dengan menghapus elemen yang tidak memiliki kontribusi signifikan terhadap analisis sentimen, seperti URL, emoji, angka, simbol, maupun karakter khusus lainnya. Langkah ini bertujuan meminimalisasi *noise* sehingga model dapat lebih terfokus pada informasi utama yang terkandung dalam teks. Selain itu, proses *cleaning* juga membantu mengurangi variasi teks informal yang umum ditemukan pada komentar media sosial.

2) *Case Folding*

Case folding merupakan tahapan normalisasi teks dengan cara mengubah seluruh huruf menjadi bentuk kecil (*lowercase*). Proses ini dilakukan untuk menghindari perbedaan representasi kata akibat penggunaan huruf kapital, sehingga kata dengan makna yang sama dapat dikenali sebagai satu bentuk konsisten. Secara matematis dapat dirumuskan sebagai (2).

$$T = lower(T) \quad (2)$$

Keterangan :

T : teks sebelum perubahan

$lower(T)$: fungsi untuk mengubah huruf menjadi lowercase.

Dengan demikian, kata “Sehat” dan “sehat” akan dikenali sebagai bentuk yang sama oleh model klasifikasi.

3) Normalisasi Teks

Normalisasi dilakukan untuk mengubah kata tidak baku, singkatan, maupun istilah informal menjadi bentuk yang lebih standar sesuai kaidah bahasa. Secara matematis dapat direpresentasikan sebagai (3).

$$w' = \text{norm}(w) \quad (3)$$

Keterangan :

w : kata non-standar

$\text{norm}(w)$: fungsi normalisasi

w' : kata standar hasil normalisasi

Normalisasi berperan penting dalam analisis data media sosial karena teks pengguna cenderung bersifat informal. Proses ini meningkatkan konsistensi makna kata sehingga representasi teks menjadi lebih baik.

4) Penghapusan Duplikasi

Tahap ini dilakukan untuk menghilangkan komentar identik agar tidak menimbulkan bias pada proses pelatihan model. Secara matematis dapat dinyatakan sebagai (4).

$$D' = D - x_i | x_i \in D, \exists x_j = x_i, i \neq j \quad (4)$$

Keterangan :

D : dataset awal

x_i : data ke- i

D' : dataset bebas duplikasi

Dengan menghapus data berulang, distribusi menjadi lebih representatif dan setiap komentar memberikan kontribusi informasi yang unik.

Melalui serangkaian tahapan preprocessing tersebut, data teks yang sebelumnya tidak terstruktur berhasil diubah menjadi representasi teks yang lebih bersih, konsisten, dan terorganisasi. Hasil preprocessing ini memastikan bahwa data siap digunakan pada tahap selanjutnya, yaitu pelabelan sentimen dan ekstraksi fitur menggunakan IndoBERT, sehingga mendukung peningkatan akurasi klasifikasi.

C. Pelabelan Sentimen

Pelabelan sentimen merupakan tahap penting dalam penelitian ini yang bertujuan untuk mengklasifikasikan setiap komentar ke dalam kategori tertentu sesuai dengan opini yang terkandung di dalamnya. Data yang telah melalui proses preprocessing kemudian diberi label sentimen dalam tiga kategori, yaitu positif, netral, dan negatif.

Secara matematis, proses pelabelan sentimen dapat dirumuskan sebagai (5).

$$y = g(x), y \in \text{positif}, \text{netral}, \text{negatif} \quad (5)$$

Keterangan :

x : teks komentar hasil preprocessing

g : fungsi pelabelan

y : label sentimen

Persamaan tersebut menunjukkan bahwa setiap teks (x) dipetakan ke dalam label sentimen (y) melalui fungsi (g). Dengan demikian, data teks yang awalnya tidak terstruktur dapat diubah menjadi data terstruktur yang siap digunakan pada proses klasifikasi.

1) Metode Pelabelan

Pelabelan pada penelitian ini dilakukan menggunakan pendekatan semi-manual, yaitu kombinasi antara pelabelan otomatis dan validasi manual. Pada tahap awal, sistem melakukan pelabelan secara otomatis untuk mempercepat proses anotasi data dalam jumlah besar [20]. Pelabelan otomatis tersebut dilakukan menggunakan pendekatan berbasis heuristik sederhana yang mempertimbangkan kecenderungan polaritas kata dalam konteks kalimat, tanpa melibatkan model klasifikasi pembelajaran mesin. Selanjutnya, peneliti melakukan validasi manual guna memastikan bahwa label yang diberikan sesuai dengan konteks dan makna sebenarnya dari setiap komentar [21]. Pendekatan semi-manual dipilih untuk menjaga keseimbangan antara efisiensi proses dan akurasi pelabelan, terutama pada data media sosial yang umumnya menggunakan bahasa informal, singkatan, serta ungkapan kontekstual yang sulit dipahami secara otomatis [22]. Validasi manual dilakukan dengan mempertimbangkan konteks kalimat, penggunaan kata, serta kecenderungan makna sentimen yang muncul pada setiap komentar.

2) Distribusi Data

Hasil pelabelan menghasilkan distribusi data sebagai berikut:

- a) Negatif: 2073 data
- b) Positif: 447 data
- c) Netral: 398 data

Distribusi awal menunjukkan adanya ketidakseimbangan kelas yang cukup signifikan, di mana jumlah data pada kategori negatif lebih dominan dibandingkan dengan kelas positif maupun netral.

3) Implikasi Ketidakseimbangan Data

Ketidakseimbangan distribusi kelas dapat menyebabkan model cenderung lebih sering memprediksi kelas mayoritas, sehingga kemampuan model dalam mengenali kelas minoritas menjadi menurun. Secara matematis, kondisi tersebut dapat dirumuskan sebagai (6).

$$P(y = \text{negatif}) \gg P(y = \text{positif}), P(y = \text{netral}) \quad (6)$$

Keterangan :

$P(y)$: probabilitas kemunculan kelas

$\text{negatif}, \text{positif}, \text{netral}$: kategori sentimen

Persamaan (6) menunjukkan bahwa probabilitas kemunculan kelas negatif jauh lebih besar dibandingkan kelas lainnya. Oleh karena itu, pada tahap berikutnya dilakukan penanganan ketidakseimbangan data menggunakan metode SMOTE untuk memperkuat representasi kelas minoritas.

D. Ekstraksi Fitur Menggunakan IndoBERT

Tahap ekstraksi fitur dilakukan untuk mengubah data teks menjadi representasi numerik sehingga dapat diproses oleh algoritma *machine learning*. Pada penelitian ini digunakan IndoBERT sebagai metode representasi teks berbasis *deep learning* yang mampu memahami makna kontekstual dari setiap kata dalam kalimat. IndoBERT dikembangkan berdasarkan arsitektur BERT dan telah dilatih menggunakan korpus berbahasa Indonesia, sehingga lebih efektif dalam menangkap struktur bahasa serta konteks penggunaan bahasa Indonesia dibandingkan metode representasi tradisional [6]. Pada penelitian ini, IndoBERT digunakan sebagai contextual embedding extractor untuk menghasilkan representasi fitur teks yang selanjutnya digunakan sebagai masukan bagi algoritma klasifikasi.

1) Representasi Teks dengan IndoBERT

Secara matematis, proses embedding dapat direpresentasikan formula (7).

$$v = \text{IndoBERT}(x) \quad (7)$$

Keterangan :

x : teks input (hasil preprocessing)

$\text{IndoBERT}(x)$: fungsi transformasi model IndoBERT

v : vektor embedding hasil representasi teks

Persamaan (7) menunjukkan bahwa setiap teks yang telah diproses diubah menjadi vektor numerik yang merepresentasikan makna semantik dan konteks linguistik.

Penelitian ini menggunakan model pretrained *indobenchmark/indobert-base-pl* melalui library *HuggingFace Transformers* untuk menghasilkan representasi kontekstual teks. Proses tokenisasi dilakukan menggunakan *AutoTokenizer* dengan panjang maksimum token sebesar 128. Embedding setiap komentar diperoleh dari representasi token [CLS] pada hidden layer terakhir (*last_hidden_state[:,0,:]*) yang menghasilkan vektor berdimensi 768. Representasi vektor tersebut kemudian digunakan sebagai fitur masukan pada proses klasifikasi menggunakan algoritma Logistic Regression dan Naïve Bayes.

2) Dimensi Representasi

Hasil dari proses embedding berupa vektor dengan dimensi tetap (8).

$$v \in R^{768} \quad (8)$$

Keterangan :

R^{768} : ruang vektor berdimensi 768

Setiap teks diubah menjadi vektor berdimensi 768 yang merepresentasikan makna semantik serta konteks kalimat.

Representasi ini memungkinkan model memahami hubungan antar kata secara lebih kompleks dibandingkan metode tradisional seperti *Bag-of-Words* atau *TF-IDF*.

Penggunaan IndoBERT dalam penelitian ini memiliki beberapa keunggulan utama:

- Kontekstual: mampu memahami makna suatu kata berdasarkan konteks kalimat tempat kata tersebut digunakan.
- Pretrained Model: telah dilatih menggunakan korpus bahasa Indonesia dalam jumlah besar sehingga tidak memerlukan proses pelatihan dari awal.
- Representasi Lebih Kaya: menghasilkan fitur yang lebih informatif dibandingkan metode ekstraksi fitur klasik.

3) Integrasi dengan Model Klasifikasi

Vektor embedding yang dihasilkan oleh IndoBERT kemudian digunakan sebagai input bagi algoritma klasifikasi Naïve Bayes dan Logistic Regression. Pendekatan ini memungkinkan pemanfaatan representasi kontekstual dari model deep learning tanpa melakukan proses fine tuning secara penuh pada model transformer. Alur transformasi data dalam proses klasifikasi dapat direpresentasikan sebagai (9).

$$x \rightarrow v \rightarrow y \quad (9)$$

Keterangan :

x : teks input

v : vektor embedding

y : label prediksi

Proses ini menunjukkan bahwa teks mentah diubah menjadi bentuk angka sebelum diproses oleh model klasifikasi untuk menghasilkan prediksi sentimen.

E. Pembagian Data dan SMOTE

Setelah proses pelabelan selesai, dataset dibagi menjadi dua bagian, yaitu data latih (*training data*) dan data uji (*testing data*). Pembagian ini bertujuan untuk melatih model sekaligus mengukur performanya terhadap data yang belum pernah digunakan sebelumnya. Secara matematis, proses pembagian data dapat dirumuskan sebagai (10).

$$D = D_{train} \cup D_{test} \quad (10)$$

Keterangan :

D : keseluruhan dataset

D_{train} : data latih

D_{test} : data uji

Proporsi pembagian dataset ditetapkan sebesar 80% untuk data latih dan 20% untuk data uji, yang secara matematis dapat dituliskan sebagai (11).

$$|D_{train}| = 0.8 |D|, |D_{test}| = 0.2 |D| \quad (11)$$

Dengan demikian, 80% data digunakan untuk melatih model, sedangkan 20% sisanya dimanfaatkan untuk proses

pengujian. Pembagian ini merupakan praktik umum dalam *machine learning* guna menjaga keseimbangan antara tahap pelatihan dan evaluasi model.

1) Ketidakseimbangan Data

Hasil pelabelan menunjukkan bahwa distribusi data tidak merata, dengan jumlah komentar pada kelas negatif jauh lebih dominan dibandingkan kelas positif maupun netral. Ketidakseimbangan ini berpotensi membuat model lebih cenderung memprediksi kelas mayoritas, sehingga performa dalam mengenali kelas minoritas menjadi kurang optimal.

2) Penerapan Synthetic Minority Over-sampling Technique (SMOTE)

Untuk mengatasi permasalahan tersebut, diterapkan metode SMOTE [17], [23], yaitu teknik oversampling yang menghasilkan data sintetis pada kelas minoritas. Penerapan metode ini bertujuan mengurangi ketidakseimbangan distribusi kelas sehingga model dapat mempelajari representasi setiap kelas secara lebih proporsional. Secara konseptual, SMOTE dapat direpresentasikan sebagai (12).

$$x_{new} = x_i + \lambda(x_{nn} - x_i) \quad (12)$$

Keterangan :

x_i : data minoritas asli

x_{nn} : tetangga terdekat dari x_i

λ : bilangan acak, $0 < \lambda < 1$

x_{new} : data sintetis yang dihasilkan

Persamaan (12) menunjukkan bahwa data sintetis x_{new} dihasilkan dari interpolasi antara data asli x_i dan tetangga terdekatnya x_{nn} . Dengan cara ini, SMOTE tidak hanya menggandakan data, tetapi menciptakan variasi baru yang lebih representatif.

3) Hasil Penerapan SMOTE

Penerapan SMOTE hanya dilakukan pada data latih (D_{train}) untuk mencegah terjadinya data *leakage*. Setelah proses SMOTE diterapkan, jumlah data pada kelas minoritas mengalami peningkatan signifikan, meskipun distribusi akhir tidak dibuat sepenuhnya seimbang. Strategi ini digunakan untuk mengurangi risiko *overfitting* akibat penggunaan data sintetis secara berlebihan, sekaligus meningkatkan kemampuan model dalam mengenali pola pada kelas yang kurang terwakili.

4) Alur Akhir Data

Setelah tahap ini, data latih dapat direpresentasikan sebagai (13):

$$D'_{train} = SMOTE D_{train} \quad (13)$$

Keterangan :

D_{train} : data latih sebelum SMOTE

D'_{train} : data latih setelah SMOTE

Data latih yang telah dimodifikasi dengan SMOTE kemudian digunakan sebagai masukan dalam pelatihan model klasifikasi.

F. Klasifikasi

Tahap klasifikasi merupakan proses untuk menentukan kategori sentimen pada setiap data berdasarkan fitur yang telah diekstraksi sebelumnya. Dalam penelitian ini digunakan dua algoritma klasifikasi, yaitu Naïve Bayes dan Logistic Regression, untuk membandingkan performa keduanya dalam mengklasifikasikan sentimen teks. Kedua algoritma tersebut dipilih karena memiliki pendekatan yang berbeda: Naïve Bayes menggunakan pendekatan probabilistik, sedangkan Logistic Regression menerapkan pendekatan diskriminatif dalam proses klasifikasi. Perbedaan pendekatan tersebut memungkinkan dilakukannya analisis komparatif terhadap performa kedua algoritma dalam mengklasifikasikan sentimen teks [4].

1) Naive Bayes

Naïve Bayes merupakan algoritma klasifikasi berbasis probabilitas yang memanfaatkan **Teorema Bayes** dengan asumsi bahwa setiap fitur bersifat independen satu sama lain. Dalam analisis sentimen, algoritma ini menggunakan distribusi fitur untuk menghitung kemungkinan suatu teks termasuk ke dalam kategori sentimen tertentu [24].

Secara umum, probabilitas suatu data berada pada kelas tertentu dapat dirumuskan sebagai (14).

$$P(y|v) = \frac{P(v|y)P(y)}{P(v)} \quad (14)$$

Keterangan :

$P(y | v)$: probabilitas kelas y diberikan fitur v

$P(v | y)$: likelihood

$P(y)$: prior probability

$P(v)$: probabilitas total

Dengan asumsi independensi antar fitur, likelihood dihitung sebagai (15).

$$P(v|y) = \prod_{i=1}^n P(v_i|y) \quad (15)$$

Keterangan :

v_i : fitur ke-i

n : jumlah fitur

Meskipun asumsi ini menyederhanakan perhitungan probabilitas, fitur cenderung saling berkorelasi pada data teks, terutama yang diwakili melalui embedding kontekstual seperti IndoBERT. Hal ini menyebabkan asumsi independensi menjadi kurang terpenuhi, yang dapat berdampak negatif pada kinerja model [7].

Selanjutnya, kelas dengan probabilitas tertinggi dipilih sebagai hasil klasifikasi (16).

$$\hat{y} = \underset{y}{\operatorname{arg\,max}} P(y|v) \quad (16)$$

Keterangan :

\hat{y} : hasil prediksi

$\operatorname{arg\,max}$: fungsi nilai maksimum.

Meskipun Naïve Bayes sederhana dan efisien, keterbatasan muncul ketika digunakan pada embedding kontekstual seperti IndoBERT, karena fitur-fitur cenderung saling berkorelasi sehingga asumsi independensi tidak sepenuhnya terpenuhi [25].

2) Logistic Regression

Logistic Regression merupakan metode klasifikasi yang digunakan untuk memprediksi probabilitas suatu kelas berdasarkan kombinasi linier dari fitur-fitur. Model ini termasuk dalam kelompok pembelajaran terawasi dan sering digunakan dalam analisis sentimen karena mampu mengelola representasi fitur berdimensi tinggi secara efektif serta menghasilkan prediksi probabilistik [4], [25].

Agar hasil dari kombinasi linier antara fitur dan bobot dapat diubah menjadi nilai probabilitas, Logistic Regression memanfaatkan fungsi aktivasi sigmoid. Fungsi ini mengonversi nilai masukan ke dalam rentang antara 0 dan 1, sehingga bisa dipahami sebagai probabilitas dari sebuah kelas. Fungsi sigmoid diungkapkan sebagai (17).

$$P(y = 1|v) = \frac{1}{1+e^{-(w^T v + b)}} \quad (17)$$

Keterangan :

$P(y = 1 | v)$: probabilitas data termasuk ke dalam kelas positif

w : vektor bobot (weight)

v : vektor fitur (embedding hasil IndoBERT)

b : bias

e : konstanta eksponensial

Fungsi sigmoid digunakan untuk mengubah hasil kombinasi linear antara fitur dan bobot menjadi nilai probabilitas dalam rentang 0 hingga 1. Nilai ini merepresentasikan tingkat keyakinan model terhadap suatu kelas.

Dalam penelitian ini, digunakan threshold sebesar 0.5 untuk mengklasifikasikan data ke dalam kelas tertentu, yang dirumuskan sebagai (18).

$$\hat{y} = \begin{cases} 1, & \text{jika } P(y = 1 | v) \geq 0.5 \\ 0, & \text{jika } P(y = 1 | v) < 0.5 \end{cases} \quad (18)$$

Keterangan simbol:

\hat{y} : hasil prediksi kelas

0.5 : threshold (batas keputusan)

Model mengklasifikasikan data ke dalam kelas tertentu berdasarkan nilai probabilitas yang dihasilkan. Jika

probabilitas melebihi ambang batas, maka data dikategorikan sebagai kelas positif, dan sebaliknya. Pada implementasinya, klasifikasi multi kelas pada penelitian ini dilakukan menggunakan pendekatan Softmax untuk menentukan probabilitas pada setiap kategori sentimen.

Karena penelitian ini melibatkan lebih dari dua kelas sentimen, Logistic Regression diperluas menggunakan fungsi Softmax untuk menangani klasifikasi multi-kelas (19).

$$P(y = k|v) = \frac{e^{w_k^T v}}{\sum_{j=1}^K e^{w_j^T v}} \quad (19)$$

Keterangan :

$P(y = k | v)$: probabilitas data termasuk ke kelas ke-k

k : indeks kelas

K : jumlah kelas (3 kelas sentimen)

w_k : bobot untuk kelas ke-k

Karena penelitian ini menggunakan tiga kategori sentimen, yaitu positif, netral, dan negatif, maka Logistic Regression dikembangkan menggunakan fungsi Softmax. Fungsi tersebut memungkinkan model menghitung probabilitas pada setiap kelas, kemudian memilih kelas dengan nilai probabilitas tertinggi sebagai hasil prediksi.

G. Evaluasi Model

Tahap evaluasi model dilakukan untuk menilai sejauh mana algoritma mampu mengklasifikasikan data sentimen dengan tepat [26]. Proses evaluasi menggunakan sejumlah metrik, yaitu accuracy, precision, recall, dan F1-score, yang dihitung berdasarkan confusion matrix. Penggunaan berbagai metrik tersebut bertujuan memberikan gambaran yang lebih komprehensif mengenai performa model [27].

1) Confusion Matrix

Confusion matrix merupakan sebuah tabel yang digunakan untuk membandingkan hasil prediksi model dengan label aktual pada data.

TABEL I
CONFUSION MATRIX

	Prediksi Positif	Prediksi Negatif
Aktual Positif	TP (True Positif)	FP (False Positif)
Aktual Negatif	TN (True Negatif)	FN (False Negatif)

- TP (True Positive): data positif yang berhasil diprediksi dengan benar sebagai positif.
- TN (True Negative): data negatif yang berhasil diprediksi dengan benar sebagai negatif.
- FP (False Positive): data negatif yang salah diprediksi sebagai positif.
- FN (False Negative): data positif yang salah diprediksi sebagai negatif.

2) Accuracy

Nilai accuracy digunakan untuk mengukur tingkat ketepatan model secara keseluruhan dalam melakukan

prediksi. Secara matematis, accuracy dapat dirumuskan sebagai (20).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (20)$$

Metrik ini mampu memberikan gambaran umum mengenai performa model secara keseluruhan. Namun, akurasi memiliki keterbatasan apabila diterapkan pada dataset dengan distribusi kelas yang tidak seimbang, karena model dapat menghasilkan nilai akurasi tinggi meskipun gagal mengenali kelas minoritas.

Oleh karena itu, penelitian ini juga menggunakan metrik tambahan, yaitu Precision, Recall, dan F1-Score, yang lebih sensitif terhadap ketidakseimbangan kelas.

- *Precision* mengukur proporsi prediksi positif yang benar dibandingkan seluruh prediksi positif.
- *Recall* mengukur proporsi prediksi positif yang benar dibandingkan seluruh data aktual positif.
- *F1-Score* merupakan harmonisasi antara precision dan recall, sehingga memberikan penilaian yang lebih seimbang.

Dengan kombinasi metrik tersebut, evaluasi model tidak hanya menyoroti performa secara agregat, tetapi juga menilai kemampuan model dalam mengenali kelas minoritas. Hal ini penting mengingat distribusi data sentimen pada komentar YouTube cenderung tidak merata, dengan dominasi kelas negatif.

III. HASIL DAN PEMBAHASAN

A. Karakteristik Data

Dataset penelitian ini bersumber dari komentar pengguna pada platform YouTube yang dikategorikan ke dalam tiga jenis sentimen, yaitu positif, netral, dan negatif. Data yang terkumpul berbentuk teks tidak terstruktur dengan ragam penggunaan bahasa yang luas, sehingga mampu merepresentasikan opini publik secara lebih alami. Dari hasil pelabelan awal, ditemukan adanya ketidakseimbangan distribusi data yang cukup signifikan. Setelah melalui proses preprocessing dan filtering data, diperoleh sebanyak 1.848 komentar valid yang digunakan dalam proses klasifikasi sentimen. Distribusi dataset menunjukkan adanya ketidakseimbangan antar kelas, di mana kategori positif memiliki jumlah data lebih dominan dibandingkan kategori negatif dan netral. Kondisi tersebut berpotensi memengaruhi performa model klasifikasi, khususnya dalam mengenali kelas dengan jumlah data yang lebih sedikit. Ketimpangan jumlah antar kelas ini berpotensi menurunkan kinerja model dalam proses klasifikasi sentimen.

Untuk mengatasi permasalahan distribusi yang tidak seimbang, penelitian ini menerapkan metode SMOTE pada data latih. Metode ini berfungsi menghasilkan data sintesis pada kelas minoritas sehingga distribusi antar kelas menjadi lebih proporsional. Setelah dilakukan penyeimbangan, proporsi data tiap kategori lebih merata dan memungkinkan model mempelajari karakteristik sentimen secara lebih

optimal. Penerapan SMOTE hanya dilakukan pada data latih, sedangkan data uji tetap mempertahankan distribusi asli agar evaluasi performa model tetap objektif dan mencerminkan kondisi data yang sesungguhnya.

B. Tahapan Pra-Pemrosesan Teks

Tahap pra-pemrosesan teks dilakukan untuk membersihkan dan menormalisasi data sebelum masuk ke proses ekstraksi fitur serta klasifikasi. Langkah ini memiliki peran krusial karena komentar pada platform YouTube sering kali mengandung berbagai bentuk *noise*, seperti penggunaan huruf kapital yang tidak konsisten, tanda baca berlebihan, simbol, maupun karakter lain yang tidak relevan terhadap analisis sentimen. Kondisi tersebut berpotensi menurunkan kualitas pengolahan data dan berdampak pada performa model.

Untuk mengatasi permasalahan tersebut, penelitian ini menerapkan beberapa tahapan preprocessing, yaitu *cleaning* dan *case folding*, agar teks menjadi lebih seragam serta mudah diproses. Pada tahap *cleaning*, dilakukan penghapusan elemen-elemen yang tidak memberikan kontribusi signifikan terhadap isi teks, seperti karakter khusus, angka, tanda baca, dan simbol tertentu. Selain itu, proses *cleaning* juga mencakup penghapusan emoji serta simbol nontekstual lain yang umum ditemukan pada komentar media sosial. Penelitian ini juga mempertahankan beberapa bentuk bahasa informal yang masih memiliki makna sentimen agar konteks opini pengguna tetap terjaga. Selanjutnya, *case folding* diterapkan dengan mengubah seluruh huruf menjadi bentuk kecil (*lowercase*) untuk menghindari perbedaan representasi kata yang sebenarnya memiliki makna sama.

TABEL II
HASIL PREPROCESSING TEKS

Index	Content	Final_text
0	13:35 aku mendiagnosa diri sendiri ADHD ternyata ke psikiater lebih bad news lagi malah BPD 😞	1335 aku mendiagnosa diri sendiri adhd ternyata ke psikiater lebih bad news lagi malah bpd
1	Salah satu faktor depresi yang kerap terjadi adalah Kesepian	salah satu faktor depresi yang kerap terjadi adalah kesepian
2	Loser, fake smile, im fine	loserfake smileim fine
3	Apakah kalian ada yang pernah ngalamin depresi? Kalo iya, gimana cara kalian ngatasin kondisi ini? Coba kasih tau di sini ya, siapa tau ada yang membutuhkan :)	apakah kalian ada yang pernah ngalamin depresi kalo iya gimana cara kalian ngatasin kondisi ini coba kasih tau di sini ya siapa tau ada yang membutuhkan
4	Salah satu faktor depresi yang kerap terjadi adalah Kesepian	salah satu faktor depresi yang kerap terjadi adalah kesepian
5	Saya skrang lagi mengalami 😞	saya skrang lagi mengalami

Melalui tahapan tersebut, data teks berhasil diubah menjadi lebih bersih, konsisten, dan terorganisasi. Hasil pra-pemrosesan ini memastikan bahwa data siap digunakan dalam proses ekstraksi fitur menggunakan IndoBERT, sehingga mendukung peningkatan akurasi dan reliabilitas pada tahap klasifikasi sentimen.

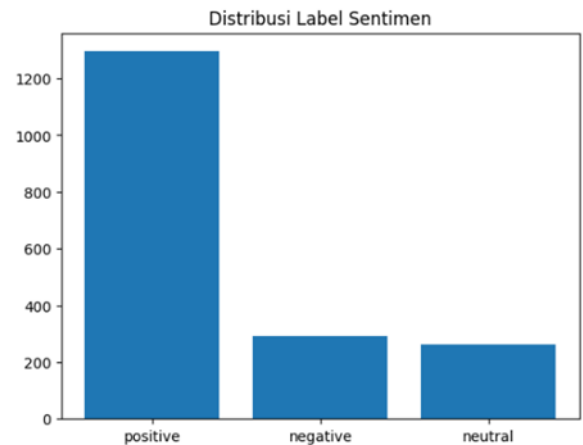
Berdasarkan Tabel II, proses pra-pemrosesan teks menunjukkan perubahan signifikan dari data mentah yang bersifat tidak terstruktur menjadi teks yang lebih terstandarisasi, di mana penghapusan karakter non-alfabet melalui *cleaning* serta normalisasi huruf dengan *case folding* menghasilkan representasi teks yang lebih konsisten dan seragam; transformasi ini penting karena mampu mengurangi *noise* pada data sekaligus meningkatkan kualitas fitur yang dihasilkan pada tahap ekstraksi, sehingga dengan teks yang lebih bersih dan terorganisasi, model klasifikasi diharapkan dapat mengenali pola sentimen secara lebih akurat dan optimal.

C. Statistik Dataset

Dataset yang digunakan dalam penelitian ini terdiri atas 1.848 komentar yang telah dikategorikan ke dalam tiga jenis sentimen, yaitu positif, netral, dan negatif. Berdasarkan hasil pelabelan, distribusi dataset menunjukkan bahwa kelas positif memiliki jumlah data paling dominan dibandingkan kelas negatif dan netral. Secara keseluruhan, dataset terdiri atas 1.295 data positif, 291 data negatif, dan 262 data netral. Kondisi tersebut menunjukkan adanya ketidakseimbangan distribusi kelas (*imbalanced dataset*) yang berpotensi memengaruhi performa model klasifikasi, khususnya dalam mengenali kelas dengan jumlah data yang lebih sedikit.

Berdasarkan Gambar 2, distribusi kategori sentimen pada dataset setelah preprocessing dan filtering memperlihatkan dominasi kelas positif dibandingkan kelas negatif maupun netral. Perubahan distribusi ini dipengaruhi oleh proses pembersihan data, seperti penghapusan komentar duplikat, komentar kosong, komentar nonbahasa Indonesia, serta penyaringan teks yang tidak relevan. Kondisi ketidakseimbangan distribusi tersebut berpotensi menyebabkan bias pada model klasifikasi sehingga diperlukan penanganan khusus melalui metode SMOTE pada data latih.

Data kemudian dibagi menjadi data latih dan data uji menggunakan metode *train-test split* untuk mendukung proses pelatihan serta evaluasi model. Dari hasil pembagian, sebanyak 370 data ditetapkan sebagai data uji, sedangkan sisanya digunakan sebagai data latih. Distribusi pada data uji menunjukkan adanya 259 data positif, 53 data netral, dan 58 data negatif. Komposisi ini memperlihatkan dominasi kelas positif dibandingkan dua kelas lainnya, sehingga kondisi ketidakseimbangan data tetap terlihat pada proses evaluasi model.



Gambar 2. Distribusi Kelas Sentimen Setelah Preprocessing dan Filtering Data

Ketidakseimbangan tersebut berpotensi menurunkan performa model, khususnya dalam mengenali kelas dengan jumlah data yang lebih sedikit. Oleh karena itu, pada tahap pelatihan diterapkan metode SMOTE untuk memperkuat representasi kelas minoritas. Dengan penerapan teknik ini, model diharapkan mampu mempelajari pola data secara lebih optimal serta menghasilkan klasifikasi sentimen yang lebih akurat.

1. Pembagian Dataset

TABEL III
PEMBAGIAN DATASET MENJADI DATA LATIH DAN DATA UJI

Keterangan	Jumlah Data
Total Dataset	1.848
Data Latih	1.478
Data Uji	370

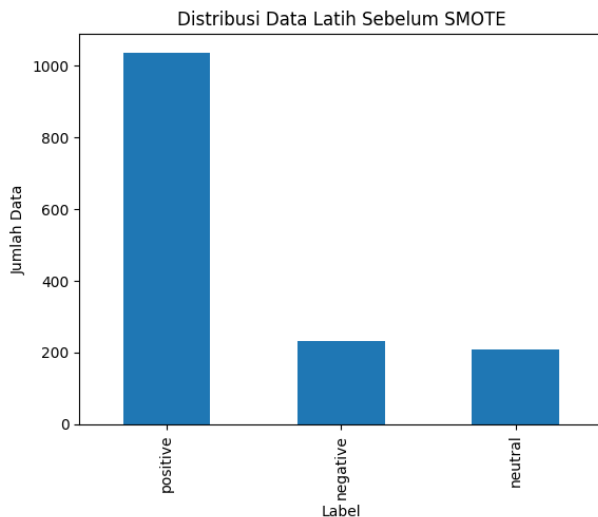
2. Distribusi Data Uji

TABEL IV
DISTRIBUSI DATA UJI BERDASARKAN KELAS SENTIMEN

Kelas Sentimen	Jumlah Data
Positif	259
Netral	53
Negatif	58
Total	370

Berdasarkan Tabel III, jumlah dataset yang digunakan dalam penelitian ini terdiri atas 1.848 komentar yang kemudian dibagi menjadi data latih dan data uji dengan metode *train-test split*, di mana 1.478 data dialokasikan sebagai data latih dan 370 data lainnya digunakan sebagai data uji; pembagian ini bertujuan agar model memperoleh data yang memadai untuk mempelajari pola yang ada sekaligus menyediakan data uji yang representatif pada tahap evaluasi. Selanjutnya, merujuk pada Tabel IV, distribusi data uji menunjukkan bahwa kelas positif memiliki jumlah paling dominan dengan 259 data, sedangkan kelas negatif berjumlah 58 data dan kelas netral 53 data, perbedaan distribusi antar kelas tersebut menegaskan adanya kondisi *imbalanced dataset* yang dapat memengaruhi performa model, khususnya

dalam mengenali kelas dengan jumlah data terbatas. Untuk mengatasi permasalahan tersebut, penelitian ini menerapkan metode SMOTE pada tahap pelatihan, yang berfungsi memperkuat representasi kelas minoritas sehingga model dapat mempelajari pola data secara lebih komprehensif dan optimal.



Gambar 3. Distribusi Kelas Sentimen Sebelum SMOTE

Berdasarkan Gambar 3, distribusi kategori sentimen pada dataset sebelum penerapan metode SMOTE menunjukkan bahwa jumlah data pada kategori positif jauh lebih besar dibandingkan dengan kategori netral maupun negatif, sehingga kondisi ini menimbulkan ketidakseimbangan distribusi yang berpotensi membuat model klasifikasi lebih cenderung memprediksi kelas dominan; situasi tersebut menjadi tantangan dalam proses klasifikasi sentimen, khususnya dalam meningkatkan kemampuan model mengenali kelas dengan jumlah data terbatas, sehingga pada tahap pelatihan diterapkan metode SMOTE untuk menyeimbangkan distribusi data latih agar model dapat mempelajari karakteristik setiap kategori sentimen secara lebih efektif dan optimal.

D. Ekstraksi Fitur Menggunakan IndoBERT

Dalam penelitian ini, proses ekstraksi fitur dilakukan dengan memanfaatkan IndoBERT untuk mengubah data teks menjadi representasi numerik berupa vektor. IndoBERT, sebagai model bahasa berbasis arsitektur Transformer yang dilatih menggunakan korpus berbahasa Indonesia, memiliki keunggulan dalam memahami konteks dan makna kata dibandingkan metode representasi tradisional. Setelah melalui tahap pra-pemrosesan, setiap teks dikonversi menjadi embedding berdimensi tinggi yang memungkinkan model menangkap hubungan semantik antar kata dalam kalimat; dengan demikian, informasi yang diperoleh tidak hanya bergantung pada frekuensi kata, tetapi juga mempertimbangkan konteks penggunaannya.

Proses ekstraksi fitur dilakukan dengan mengambil keluaran dari layer terakhir IndoBERT yang kemudian digunakan sebagai masukan bagi algoritma klasifikasi, seperti Naïve Bayes dan Logistic Regression. Representasi embedding diperoleh menggunakan output hidden state pada layer terakhir IndoBERT dengan pendekatan pooling berbasis token [CLS], sehingga setiap teks direpresentasikan menjadi satu vektor embedding berdimensi tetap. Pendekatan ini menjadikan setiap data teks tertransformasi menjadi vektor numerik yang siap diproses secara sistematis. Penggunaan IndoBERT bertujuan menghasilkan fitur yang lebih kontekstual dan berkualitas, sehingga diharapkan mampu meningkatkan performa model dalam klasifikasi sentimen. Kemampuan IndoBERT dalam memahami nuansa bahasa menjadikan representasi teks yang dihasilkan lebih informatif dan relevan dibandingkan pendekatan berbasis statistik, seperti TF-IDF, sekaligus memberikan landasan yang lebih kuat bagi analisis sentimen yang mendalam.

E. Analisis Performa Model Klasifikasi

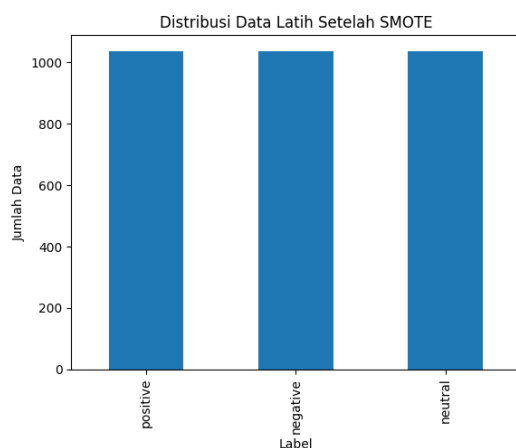
Pada tahap ini bertujuan mengevaluasi performa algoritma klasifikasi yang digunakan dalam penelitian, yaitu Logistic Regression dan Naïve Bayes, dalam mengelompokkan sentimen teks ke dalam kategori positif, netral, dan negatif. Evaluasi dilakukan menggunakan data uji yang sebelumnya dipisahkan melalui metode *train-test split*, sehingga hasil pengujian dapat menunjukkan kemampuan model dalam melakukan generalisasi terhadap data baru. Proses pelatihan model menggunakan data latih yang telah diseimbangkan dengan metode SMOTE untuk mengatasi permasalahan *imbalanced dataset*, sementara data uji tetap mempertahankan distribusi asli agar hasil evaluasi tetap objektif dan mencerminkan kondisi data yang sesungguhnya.

Dalam mengukur performa model, penelitian ini memanfaatkan sejumlah metrik evaluasi yang umum digunakan pada tugas klasifikasi, yaitu accuracy, precision, recall, dan F1-score. Accuracy digunakan untuk menilai tingkat ketepatan model secara keseluruhan dalam melakukan prediksi, sedangkan precision dan recall memberikan gambaran lebih detail mengenai kemampuan model dalam mengenali masing-masing kelas sentimen. Adapun F1-score berfungsi sebagai metrik gabungan yang menyeimbangkan antara precision dan recall, sehingga mampu memberikan penilaian yang lebih komprehensif, khususnya pada kondisi data yang tidak seimbang.

1) Dampak Smote

Penerapan metode Synthetic Minority Over-sampling Technique (SMOTE) dilakukan untuk mengatasi ketidakseimbangan distribusi kelas. Teknik ini bekerja dengan menghasilkan data sintesis pada kelas minoritas sehingga distribusi antar kelas menjadi lebih proporsional. Tujuan penggunaan SMOTE adalah agar model dapat mempelajari pola data dari setiap kategori sentimen secara lebih merata dan tidak hanya berfokus pada kelas mayoritas. Gambar 3 memperlihatkan distribusi data sebelum dilakukan

proses penyeimbangan, di mana kelas negatif tampak jauh lebih dominan dibandingkan kelas positif maupun netral. Ketimpangan ini berpotensi membuat model lebih sering memprediksi kelas mayoritas, sehingga kemampuan dalam mengenali kelas minoritas menjadi kurang optimal. Dengan penerapan SMOTE, representasi kelas minoritas diperkuat sehingga model memiliki peluang lebih besar untuk mengenali pola sentimen secara lebih efektif dan seimbang. Meskipun demikian, penerapan SMOTE pada embedding kontekstual tetap memiliki keterbatasan karena data sintetis yang dihasilkan belum tentu sepenuhnya merepresentasikan hubungan semantik alami antar teks. Oleh karena itu, penggunaan SMOTE dalam penelitian ini dibatasi hanya pada data latih guna meminimalkan potensi distorsi representasi fitur serta menghindari data leakage pada tahap evaluasi.



Gambar 4. Distribusi Kelas Sentimen Setelah SMOTE

Gambar 4 memperlihatkan distribusi data setelah penerapan metode SMOTE pada data latih. Proses penyeimbangan ini menghasilkan jumlah data yang lebih proporsional antar kelas dibandingkan kondisi sebelum SMOTE, sehingga distribusi terlihat lebih seimbang. Perubahan signifikan terjadi terutama pada kelas yang sebelumnya memiliki jumlah data terbatas; menariknya, setelah penyeimbangan, kelas positif justru menjadi kategori dengan jumlah data terbanyak, menggantikan dominasi kelas negatif. Hal ini menunjukkan bahwa SMOTE memberikan pengaruh besar terhadap komposisi data, khususnya dalam memperkuat representasi kelas minoritas.

Secara keseluruhan, penerapan SMOTE terbukti efektif dalam mengurangi permasalahan *imbalanced dataset* dengan menambah jumlah data sintetis pada kelas minoritas. Perubahan distribusi ini diharapkan membantu model mempelajari pola sentimen secara lebih merata dan optimal, sehingga meningkatkan kemampuan klasifikasi, terutama dalam mengenali kelas yang sebelumnya kurang terwakili. Hasil penyeimbangan data tersebut kemudian digunakan dalam tahap pelatihan model klasifikasi untuk memperkuat performa dalam mengidentifikasi setiap kategori sentimen.

2) Evaluasi Model Logistic Regression

Model Logistic Regression menunjukkan performa yang cukup baik dalam melakukan klasifikasi sentimen teks. Berdasarkan hasil pengujian, model mencapai tingkat akurasi sebesar 78%, yang menandakan bahwa sebagian besar data uji berhasil diklasifikasikan dengan tepat.

Jika ditinjau lebih lanjut pada masing-masing kategori sentimen, performa terbaik ditunjukkan oleh kelas positif dengan nilai precision sebesar 0,89, recall sebesar 0,85, dan F1-score sebesar 0,87. Hasil ini menegaskan bahwa model memiliki kemampuan yang sangat baik dalam mengenali sentimen positif. Namun demikian, performa pada kelas netral dan negatif masih relatif lebih rendah dibandingkan kelas positif. Pada kelas netral, nilai recall mencapai 0,68, yang menunjukkan bahwa model cukup mampu mendeteksi data netral, tetapi nilai precision yang hanya sebesar 0,53 mengindikasikan adanya kesulitan dalam membedakan sentimen netral dari kategori lainnya. Sementara itu, pada kelas negatif, model memperlihatkan nilai precision dan recall yang relatif seimbang, meskipun performanya tetap berada di bawah kelas positif. Kondisi ini menunjukkan bahwa model cukup baik dalam mengenali sentimen negatif, tetapi masih terdapat sejumlah kesalahan klasifikasi yang perlu diperhatikan.

Seluruh hasil evaluasi tersebut diperoleh dari model yang dilatih menggunakan data latih yang telah diseimbangkan dengan metode SMOTE, sedangkan pengujian tetap dilakukan menggunakan data uji dengan distribusi asli tanpa penyeimbangan. Dengan demikian, hasil evaluasi dapat dianggap objektif dan merepresentasikan kondisi data yang sesungguhnya, sekaligus memberikan gambaran nyata mengenai kekuatan dan keterbatasan Logistic Regression dalam tugas klasifikasi sentimen.

TABEL V
HASIL EVALUASI MODEL LOGISTIC REGRESSION

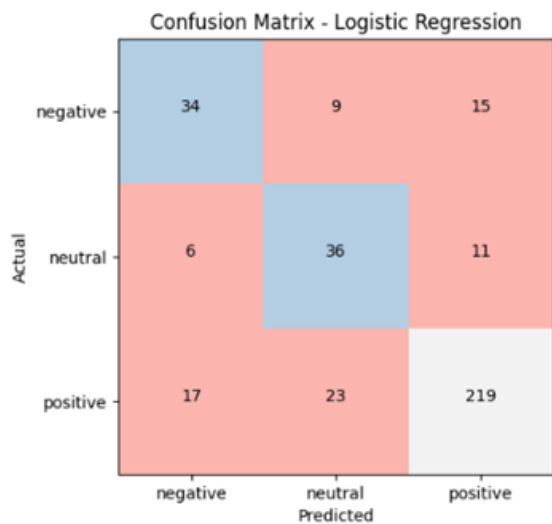
Kelas	Precision	Recall	F1-Score
Negatif	0,60	0,59	0,59
Netral	0,53	0,68	0,60
Positif	0,89	0,85	0,87
Accuracy			0,78

Berdasarkan Tabel V, model Logistic Regression menunjukkan performa yang cukup baik dengan tingkat akurasi sebesar 0,78. Performa tertinggi ditunjukkan pada kategori sentimen positif, dengan nilai precision, recall, dan F1-score yang tinggi, masing-masing sebesar 0,89, 0,85, dan 0,87. Temuan ini menegaskan bahwa model memiliki kemampuan yang sangat baik dalam mengidentifikasi sentimen positif.

Sebaliknya, performa pada kategori netral dan negatif masih relatif lebih rendah. Pada kategori netral, nilai recall yang cukup tinggi belum diimbangi dengan precision yang sepadan, sehingga masih terjadi kesalahan dalam proses klasifikasi. Sementara itu, pada kategori negatif, nilai precision dan recall yang relatif seimbang menunjukkan

bahwa model cukup mampu mengenali kelas tersebut, meskipun hasilnya belum sepenuhnya optimal.

Secara keseluruhan, hasil evaluasi ini menegaskan bahwa Logistic Regression lebih unggul dalam mengklasifikasikan kelas mayoritas dibandingkan kelas minoritas. Kondisi ini mengindikasikan perlunya strategi tambahan untuk meningkatkan performa pada kategori dengan jumlah data yang lebih sedikit, sehingga model dapat mencapai keseimbangan yang lebih baik dalam mengenali seluruh kelas sentimen.



Gambar 5. Confusion Matrix Model Logistic Regression

Gambar 5 menampilkan *confusion matrix* dari model Logistic Regression. Berdasarkan hasil tersebut, model menunjukkan performa yang cukup baik dalam mengidentifikasi kelas positif, dengan jumlah prediksi benar mencapai 219 data, sehingga menegaskan tingkat ketepatan yang tinggi dalam mengenali sentimen positif.

Namun demikian, masih terdapat sejumlah kesalahan klasifikasi pada kelas netral dan negatif. Pada kelas negatif, sebagian data diprediksi sebagai netral maupun positif, sedangkan pada kelas netral beberapa data salah diklasifikasikan ke kategori lain. Selain itu, terdapat pula data positif yang keliru diprediksi sebagai negatif ataupun netral. Kondisi ini menunjukkan bahwa model masih memiliki keterbatasan dalam membedakan antar kelas sentimen, terutama pada kategori dengan jumlah data yang relatif lebih sedikit.

Secara keseluruhan, analisis *confusion matrix* mengindikasikan bahwa Logistic Regression lebih unggul dalam mengklasifikasikan kelas mayoritas dibandingkan kelas minoritas. Temuan ini konsisten dengan karakteristik dataset yang tidak seimbang, sekaligus menegaskan perlunya strategi tambahan untuk meningkatkan kemampuan model dalam mengenali kelas dengan representasi data yang terbatas.

3) Evaluasi Model Naïve Bayes

Model Naïve Bayes digunakan sebagai pembanding dalam penelitian ini untuk mengklasifikasikan sentimen teks ke dalam tiga kategori, yaitu positif, netral, dan negatif. Berdasarkan hasil pengujian menggunakan data uji, model ini memperoleh tingkat akurasi sebesar 0,56, yang menunjukkan bahwa performanya masih lebih rendah dibandingkan dengan Logistic Regression.

Tabel V menyajikan hasil evaluasi model Naïve Bayes berdasarkan metrik precision, recall, dan F1-score pada setiap kategori sentimen. Performa terbaik ditunjukkan pada kategori positif, dengan nilai precision sebesar 0,80. Namun demikian, nilai recall yang hanya mencapai 0,55 mengindikasikan bahwa masih terdapat cukup banyak data positif yang belum berhasil dikenali secara optimal. Pada kategori netral, model memperoleh nilai recall sebesar 0,62, menandakan kemampuan cukup baik dalam mendeteksi sentimen netral, tetapi nilai precision yang hanya sebesar 0,42 menunjukkan adanya kesalahan klasifikasi dengan kategori lain. Sementara itu, pada kategori negatif, model menghasilkan nilai recall sebesar 0,53, namun precision yang relatif rendah, yaitu 0,27, memperlihatkan bahwa model masih sering melakukan kesalahan dalam memprediksi sentimen negatif.

Secara keseluruhan, hasil evaluasi menegaskan bahwa performa Naïve Bayes belum optimal jika dibandingkan dengan Logistic Regression, terutama dalam menjaga keseimbangan antara precision dan recall pada masing-masing kategori sentimen. Kondisi ini menunjukkan bahwa meskipun Naïve Bayes mampu mengenali pola dasar pada data, model masih memerlukan strategi tambahan untuk meningkatkan akurasi dan konsistensi klasifikasi, khususnya pada kelas dengan jumlah data yang lebih sedikit.

TABEL VI
HASIL EVALUASI MODEL NAÏVE BAYES

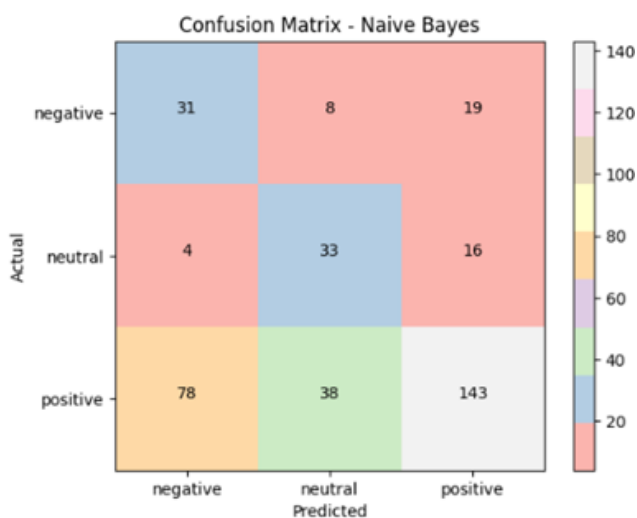
Kelas	Precision	Recall	F1-Score
Negatif	0,27	0,53	0,36
Netral	0,42	0,62	0,50
Positif	0,80	0,55	0,65
Accuracy			0,56

Berdasarkan Tabel VI, model Naïve Bayes menunjukkan performa yang belum optimal dengan tingkat akurasi sebesar 0,56. Meskipun pada kelas positif model memperoleh nilai precision yang cukup tinggi, nilai recall yang lebih rendah mengindikasikan bahwa model belum mampu mengenali seluruh data positif secara efektif.

Selain itu, pada kelas netral dan negatif, perbedaan yang cukup mencolok antara nilai precision dan recall menunjukkan masih tingginya tingkat kesalahan klasifikasi. Pada kelas netral, recall relatif baik tetapi precision rendah, sehingga banyak data salah terklasifikasi ke kategori lain. Sementara itu, pada kelas negatif, recall moderat tidak diimbangi dengan precision yang memadai, menandakan

model masih sering keliru dalam memprediksi sentimen negatif.

Secara keseluruhan, hasil evaluasi memperlihatkan bahwa performa Naïve Bayes masih kurang konsisten dalam mengklasifikasikan setiap kelas sentimen. Temuan ini menegaskan bahwa meskipun model mampu menangkap pola dasar, diperlukan strategi tambahan untuk meningkatkan keseimbangan antara precision dan recall, khususnya pada kategori dengan jumlah data yang lebih sedikit, agar performa klasifikasi menjadi lebih stabil dan representatif.



Gambar 6. Confusion Matrix Model Naïve Bayes

Gambar 6 menampilkan *confusion matrix* dari model Naïve Bayes. Matriks tersebut diperoleh dari model yang dilatih menggunakan data latih yang telah diseimbangkan dengan metode Synthetic Minority Over-sampling Technique (SMOTE), sedangkan pengujian dilakukan menggunakan data uji tanpa penyeimbangan.

Hasil pada gambar menunjukkan bahwa performa model masih belum optimal, khususnya pada kelas positif. Dari seluruh data positif, hanya 143 data yang berhasil diklasifikasikan dengan benar, sementara sebagian lainnya salah diprediksi sebagai kelas negatif sebanyak 78 data dan kelas netral sebanyak 38 data. Kondisi ini menandakan bahwa model masih mengalami kesulitan dalam mengenali sentimen positif secara akurat.

Untuk kelas negatif, terdapat 31 data yang berhasil diprediksi dengan tepat, namun masih ditemukan kesalahan klasifikasi ke kelas netral maupun positif. Sementara itu, pada kelas netral, model menunjukkan kinerja yang cukup baik dengan 33 prediksi benar, meskipun masih terdapat sejumlah kesalahan klasifikasi ke kategori lain.

Secara keseluruhan, analisis *confusion matrix* ini menunjukkan bahwa model Naïve Bayes memiliki tingkat kesalahan klasifikasi yang lebih tinggi dibandingkan Logistic Regression, terutama dalam membedakan kelas positif dari kategori sentimen lainnya. Temuan ini menegaskan bahwa meskipun Naïve Bayes mampu mengenali sebagian pola dasar, model masih memerlukan strategi tambahan untuk

meningkatkan akurasi, khususnya pada kelas dengan jumlah data yang lebih terbatas.

4) Perbandingan Performa Model Klasifikasi

Dalam penelitian ini dilakukan analisis komparatif antara model Logistic Regression dan Naïve Bayes dalam tugas klasifikasi sentimen teks. Perbandingan dilakukan menggunakan sejumlah metrik evaluasi, yaitu accuracy, precision, recall, dan F1-score, sehingga diperoleh gambaran yang lebih komprehensif mengenai performa masing-masing model.

Hasil pengujian menunjukkan bahwa Logistic Regression memiliki kinerja lebih unggul dibandingkan Naïve Bayes. Logistic Regression berhasil mencapai tingkat akurasi sebesar 0,78, sedangkan Naïve Bayes hanya memperoleh 0,56. Perbedaan ini menegaskan bahwa Logistic Regression lebih efektif dalam melakukan klasifikasi dengan tingkat kesalahan yang lebih rendah. Jika ditinjau berdasarkan masing-masing kategori sentimen, Logistic Regression menunjukkan keseimbangan yang lebih baik antara precision dan recall, khususnya pada kelas positif yang merupakan kelas mayoritas. Sebaliknya, Naïve Bayes cenderung memperlihatkan ketidakseimbangan antara kedua metrik tersebut, terutama pada kelas negatif dan netral, sehingga tingkat kesalahan klasifikasi pada kedua kategori tersebut masih cukup tinggi.

Selain itu, hasil analisis *confusion matrix* memperlihatkan bahwa Logistic Regression mampu menghasilkan jumlah prediksi benar yang lebih tinggi pada kelas positif dibandingkan Naïve Bayes. Di sisi lain, Naïve Bayes masih menunjukkan tingkat kesalahan klasifikasi yang relatif tinggi, terutama pada data positif yang sering diprediksi ke dalam kategori sentimen lain.

Secara keseluruhan, dapat disimpulkan bahwa Logistic Regression memiliki kinerja yang lebih unggul dibandingkan Naïve Bayes dalam tugas klasifikasi sentimen pada dataset penelitian ini. Temuan tersebut menunjukkan bahwa Logistic Regression lebih efektif dalam mempelajari pola data yang kompleks, terutama ketika dipadukan dengan representasi fitur berbasis IndoBERT, sehingga memberikan hasil klasifikasi yang lebih akurat dan konsisten.

TABEL VII
PERBANDINGAN KINERJA MODEL LOGISTIC REGRESSION DAN NAÏVE BAYES

Model	Accuracy	F1-Score (Avg)
Logistic Regression	0,78	0,79
Naïve Bayes	0,56	0,59

Berdasarkan Tabel VII, model Logistic Regression memperlihatkan performa yang lebih unggul dibandingkan Naïve Bayes, baik dari segi akurasi maupun F1-score. Temuan ini mengindikasikan bahwa Logistic Regression lebih mampu melakukan klasifikasi data secara konsisten serta menjaga keseimbangan kinerja pada setiap kategori sentimen.

Sebaliknya, Naïve Bayes menunjukkan hasil yang kurang memuaskan, terutama ketika berhadapan dengan kelas minoritas, sehingga tingkat kesalahan klasifikasi menjadi lebih tinggi. Peningkatan kinerja Logistic Regression juga dipengaruhi oleh penerapan SMOTE, yang berkontribusi pada penyeimbangan distribusi data latih dan memungkinkan model mempelajari pola sentimen secara lebih merata.

Perbedaan performa yang cukup signifikan antara Logistic Regression dan Naïve Bayes menunjukkan bahwa karakteristik fitur hasil embedding IndoBERT cenderung bersifat padat, berdimensi tinggi, dan saling berkorelasi. Kondisi ini tidak sepenuhnya sesuai dengan asumsi independensi antar fitur yang menjadi dasar Naïve Bayes, sehingga performanya menjadi kurang optimal. Sebaliknya, Logistic Regression tidak mengandalkan asumsi independensi fitur dan lebih mampu memanfaatkan hubungan linear antar dimensi embedding, sehingga dapat mempelajari pola sentimen secara lebih efektif. Temuan ini menjelaskan mengapa Logistic Regression menghasilkan performa yang lebih unggul dan stabil pada dataset penelitian ini.

5) Analisis Kesalahan (Error Analysis)

Analisis kesalahan dilakukan untuk mengidentifikasi pola klasifikasi yang tidak sesuai dengan label sebenarnya. Berdasarkan hasil evaluasi, masih ditemukan sejumlah data yang salah diklasifikasikan, terutama pada kelas dengan kemiripan konteks, seperti netral dan negatif. Kondisi ini menunjukkan bahwa model menghadapi tantangan dalam membedakan sentimen yang memiliki nuansa makna berdekatan, sehingga kesalahan klasifikasi lebih sering terjadi pada kategori tersebut.

Secara keseluruhan, temuan ini menegaskan pentingnya analisis kesalahan sebagai bagian dari evaluasi performa model. Dengan memahami pola kesalahan yang muncul, penelitian dapat mengarahkan strategi perbaikan, misalnya melalui pemilihan fitur yang lebih diskriminatif atau penyesuaian algoritma, agar kemampuan model dalam membedakan kelas sentimen menjadi lebih akurat dan konsisten.

Berdasarkan Tabel VIII, kesalahan klasifikasi yang terjadi tidak hanya dipengaruhi oleh keterbatasan model dalam melakukan prediksi, tetapi juga oleh karakteristik teks yang kompleks dan beragam. Beberapa teks mengandung ambiguitas, misalnya kalimat yang mencampurkan sentimen positif dan negatif dalam satu konteks. Selain itu, teks berbentuk pertanyaan atau yang tidak memiliki ekspresi emosi yang jelas juga menyulitkan model dalam menentukan label yang tepat. Penggunaan unsur humor, bahasa tidak formal, serta konteks percakapan sehari-hari turut menjadi faktor yang menyebabkan terjadinya kesalahan klasifikasi.

Temuan ini mengindikasikan bahwa model masih memiliki keterbatasan dalam memahami makna kontekstual secara mendalam, terutama pada komentar yang mengandung ambiguitas, humor, mixed sentiment, maupun bahasa informal yang umum digunakan pada media sosial. Meskipun telah memanfaatkan representasi kontekstual berbasis

IndoBERT dan data latih yang diseimbangkan menggunakan SMOTE, beberapa pola bahasa masih sulit dikenali secara konsisten oleh model klasifikasi. Oleh karena itu, analisis kesalahan menjadi penting untuk mengarahkan strategi perbaikan, seperti pemilihan fitur yang lebih diskriminatif maupun pengembangan pendekatan klasifikasi yang lebih adaptif, agar kemampuan model dalam membedakan sentimen dapat menjadi lebih akurat dan konsisten.

TABEL VIII
CONTOH KESALAHAN KLASIFIKASI MODEL

Teks	Label Asli	Prediksi	Keterangan
“dok cara memberi hidup agar ada gairah lagi gimana 🤔”	Negatif	netral	Kalimat berupa pertanyaan sehingga sentimen tidak eksplisit
“kalau stres nya anak kecil gimana kak”	Netral	positif	Tidak mengandung ekspresi emosi yang jelas
“gak seru bang hidup kyak cipung...”	Negatif	netral	Mengandung cerita panjang dengan konteks tidak langsung
“depresi karena ga dibolehin emak buat beli bakso 🤔”	Negatif	positif	Terdapat unsur humor meskipun mengandung kata negatif
“ending yg dapat menyebabkan depresi... but thanks for education dok”	Netral	positif	Mengandung campuran sentimen negatif dan positif

IV. KESIMPULAN

Hasil penelitian menunjukkan adanya perbedaan performa yang cukup signifikan antara model Logistic Regression dan Naïve Bayes dalam klasifikasi sentimen pada komentar YouTube. Logistic Regression memberikan performa lebih baik dengan tingkat akurasi sebesar 78%, sedangkan Naïve Bayes memperoleh akurasi sebesar 56%. Perbedaan tersebut mengindikasikan bahwa Logistic Regression lebih mampu memanfaatkan representasi fitur berbasis IndoBERT dalam mengenali pola sentimen pada data penelitian.

Penerapan metode Synthetic Minority Over-sampling Technique (SMOTE) turut membantu mengurangi ketidakseimbangan distribusi data pada kelas minoritas. Penyeimbangan data latih memungkinkan model mempelajari karakteristik setiap kategori sentimen secara lebih merata sehingga performa klasifikasi menjadi lebih stabil selama proses pengujian.

Meskipun demikian, hasil klasifikasi masih menunjukkan keterbatasan pada komentar yang bersifat ambigu, mengandung mixed sentiment, humor, maupun bahasa

informal yang umum ditemukan pada media sosial. Kondisi tersebut menyebabkan sebagian data masih mengalami kesalahan prediksi, khususnya pada kelas dengan karakteristik sentimen yang saling berdekatan.

Secara keseluruhan, kombinasi representasi fitur berbasis IndoBERT dengan algoritma klasifikasi tradisional mampu memberikan performa yang cukup baik dalam analisis sentimen teks berbahasa Indonesia. Namun, penelitian ini masih memiliki keterbatasan karena hanya menggunakan data komentar YouTube dan belum melakukan perbandingan dengan pendekatan transformer end-to-end. Selain itu, penggunaan SMOTE pada embedding transformer masih berpotensi menghasilkan representasi sintesis yang belum sepenuhnya merepresentasikan struktur bahasa alami.

Oleh karena itu, penelitian selanjutnya dapat mempertimbangkan penggunaan dataset lintas platform serta membandingkan pendekatan feature extraction berbasis IndoBERT dengan metode fine-tuning transformer end-to-end seperti IndoBERT atau IndoBERTtweet untuk memperoleh evaluasi performa yang lebih komprehensif dan meningkatkan kemampuan model dalam memahami konteks bahasa secara lebih mendalam.

DAFTAR PUSTAKA

- [1] Y. Puspitasari, "Edukasi Pentingnya Screening Informasi Youtube Sebagai Sarana Mencari Informasi Kesehatan," *Genitri J. Pengabd. Masy. Bid. Kesehat.*, vol. 2, no. 2, pp. 132–136, 2023.
- [2] J. Kristiyono and N. Dwi Hermawan, "Analisis Komunikasi Interaktif Brando Franco dengan Penontonnya dalam Live Streaming di Kanal YouTube Windah Basudara," *JCommsci - J. Media Commun. Sci.*, vol. 6, no. 2, pp. 11–19, May 2023, doi: 10.29303/jcommsci.v6i2.206.
- [3] G. P. Ertansyah, R. T. C. Kusuma, and A. A. Sari, "Analisis Sentimen Pada Media Sosial Menggunakan Teknik Natural Language Processing," *Pros. Semin. Nas. Teknol. Inf. Dan Bisnis*, pp. 183–189, Jul. 2025, doi: 10.47701/qgcey104.
- [4] M. Monica and A. Purwanto, "Comparative Performance Evaluation of Naïve Bayes and Logistic Regression for Indonesian YouTube Comment Sentiment Classification," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 6, no. 2, pp. 934–943, May 2026, doi: 10.57152/malcom.v6i2.2639.
- [5] M. P. Syah, A. P. Wardani, M. Idhom, and Trimono, "Perbandingan Representasi Teks Tf-Idf Dan Bert Terhadap Akurasi Cosine Similarity Dalam Penilaian Otomatis Jawaban Berbasis Teks," *Data Sci. Indones. DSI*, vol. 5, no. 1, pp. 47–59, Jul. 2025, doi: 10.47709/dsi.v5i1.6021.
- [6] B. Wilie *et al.*, "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," 2019.
- [7] D. Marutho and V. G. Utomo, "Benchmarking IndoBERT and Transformer Models for Sentiment Classification on Indonesian E-Government Service Reviews," *Jurnal Transformatika*, vol. 23, no. 1, pp. 85–95, 2025.
- [8] D. Ramdani, M. Irfan, and N. Lukman, "Perbandingan Kinerja IndoBERT, IndoRoBERTa dan NusaBERT dalam Analisis Sentimen Isu LGBT di Media Sosial X," *INTERNAL Journal*, vol. 8, no. 2, pp. 91–107, 2025.
- [9] V. D. Setiawan, D. U. Iswavigra, and E. Anggiratih, "Implementation of IndoBERT for Sentiment Analysis of the Constitutional Court's Decision Regarding the Minimum Age of Vice Presidential Candidates," *Scientific Journal of Informatics*, vol. 12, no. 3, pp. 397–406, 2025, doi: 10.15294/sji.v12i3.26360.
- [10] L. Geni, E. Yulianti, and D. I. Sensuse, "Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia Using IndoBERT Language Models," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 9, no. 3, pp. 746–757, 2024, doi: 10.26555/jiteki.v9i3.26490.
- [11] R. Merdiansah and A. A. Ridha, "Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT," vol. 7, pp. 221–228, 2024.
- [12] S. Uyun, R. P. Rosalin, L. V. Sari, and H. H. Sucinta, "A Hybrid Classification Model Based on BERT for Multi-Class Sentiment Analysis on Twitter," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 11, no. 2, pp. 194–205, 2025, doi: 10.26555/jiteki.v11i2.30665.
- [13] and M. L. D. Al Akhdaan, T. E. Sutanto, "Confident Learning on IndoBERT: Enhancing Sentiment Classification Performance," *The Indonesian Journal of Computer Science*, vol. 13, no. 5, pp. 8350–8359, 2024.
- [14] M. A. Gumilang, F. Abdillah, M. Y. Amin, and M. Hasan, "Sentiment Analysis of Indonesian Ministries' Social Media: Citizen Responses Utilizing TextBlob Analyzer," vol. 23, no. 2, pp. 203–216, 2024.
- [15] V. E. Sidauruk and W. Herowati, "IndoBERT-Based Sentiment Analysis of Political Discourse on Platform X: The Case of Prabowo-Gibran Administration," *Journal of Applied Informatics and Computing*, vol. 10, no. 1, pp. 673–683, 2026.
- [16] A. Sehatman, D. Ronaldo, and A. Chandra, "Analisis Sentimen Publik Indonesia terhadap Konflik Israel-Iran di Media Sosial Menggunakan IndoBERT dan Explainable AI (LIME)," *Jurnal Teknologi Informasi*, vol. 20, no. 1, pp. 78–87, 2026.
- [17] F. Rodriguez-Torres, J. A. Carrasco-Ochoa, and J. F. Martínez-Trinidad, "Deterministic oversampling methods based on SMOTE,"

- J. Intell. Fuzzy Syst.*, vol. 36, no. 5, pp. 4945–4955, Jan. 2019, doi: 10.3233/JIFS-179041.
- [18] H. Ma'rifah, A. P. Wibawa, and M. I. Akbar, "Klasifikasi Artikel Ilmiah Dengan Berbagai Skenario Preprocessing," *Sains Apl. Komputasi Dan Teknol. Inf.*, vol. 2, no. 2, p. 70, Apr. 2020, doi: 10.30872/jsakti.v2i2.2681.
- [19] D. Rifaldi, A. Fadlil, and Herman, "Teknik Preprocessing Pada Text Mining Menggunakan Data Tweet 'Mental Health,'" *Decode J. Pendidik. Teknol. Inf.*, vol. 3, no. 2, pp. 161–171, Apr. 2023, doi: 10.51454/decode.v3i2.131.
- [20] S. W. Iriananda, R. P. Putra, and A. Farhan, "Kinerja Auto Labelling pada Analisis Sentimen terhadap Pasangan Calon Presiden 2024 di Media Sosial X," *Conf. Innov. Appl. Sci. Technol. CIASTECH*, pp. 618–633, Dec. 2023, doi: 10.31328/ciastech.v6i1.5354.
- [21] A. A. Pratiwi and M. Kamayani, "Perbandingan Pelabelan Data dalam Analisis Sentimen Kurikulum Proyek di platform TikTok: Pendekatan Naïve Bayes," *J. Eksplora Inform.*, vol. 14, no. 1, pp. 96–107, Sep. 2024, doi: 10.30864/eksplora.v14i1.1093.
- [22] P. Ayuningtyas, S. Khomsah, and S. Sudiarto, "Pelabelan Sentimen Berbasis Semi-Supervised Learning menggunakan Algoritma LSTM dan GRU," *JISKA J. Inform. Sunan Kalijaga*, vol. 9, no. 3, pp. 217–229, Sep. 2024, doi: 10.14421/jiska.2024.9.3.217-229.
- [23] N. Mqadi, N. Naicker, and T. Adeliyi, "A SMOTe based Oversampling Data-Point Approach to Solving the Credit Card Data Imbalance Problem in Financial Fraud Detection," *Int. J. Comput. Digit. Syst.*, vol. 10, no. 1, pp. 277–286, Feb. 2021, doi: 10.12785/ijcds/100128.
- [24] S. Rabbani, D. Safitri, F. T. P. Siregar, R. Rahmaddeni, and L. Efrizoni, "Evaluation of Support Vector Machine, Naive Bayes, Decision Tree, and Gradient Boosting Algorithms for Sentiment Analysis on ChatGPT Twitter Dataset," *Indones. J. Artif. Intell. Data Min.*, vol. 7, no. 1, pp. 11–21, Nov. 2023, doi: 10.24014/ijaidm.v7i1.24662.
- [25] N. Z. Zahra, S. Farhanatussaidah, N. N. Afifah, L. Muthoharoh, A. Satria, and M. C. T. Manullang, "Benchmarking Logistic Regression, SVM, Naive Bayes, and IndoBERT Fine-Tuning for Sentiment Analysis on Indonesian Product Reviews," May 05, 2026, *arXiv: arXiv:2605.03439*. doi: 10.48550/arXiv.2605.03439.
- [26] D. Powers, "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation," *Journal of Machine Learning Technologies*, pp. 37–63.
- [27] A. El Hannani, R. Errattahi, F. Z. Salmam, T. Hain, and H. Ouahmane, "Evaluation of the effectiveness and efficiency of state-of-the-art features and models for automatic speech recognition error detection," *J. Big Data*, vol. 8, no. 1, p. 5, Jan. 2021, doi: 10.1186/s40537-020-00391-w.