

Perbandingan Kinerja Model IndoBERT, IndoBERTweet, dan Algoritma Klasik pada Analisis Sentimen Isu Indonesia Gelap

Fris Alvin, Nurul Anisa Sri Winarsih*

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia

Email: ¹111202213986@mhs.dinus.ac.id, ^{2,*}nurulanisasw@dsn.dinus.ac.id

Email Penulis Korespondensi: nurulanisasw@dsn.dinus.ac.id

Submitted: 03/11/2025; Accepted: 06/12/2025; Published: 08/12/2025

Abstrak—Penelitian ini bertujuan untuk membandingkan performa model berbasis Transformer, yaitu IndoBERT dan IndoBERTweet, dengan tiga algoritma klasik, yaitu Support Vector Machine (SVM), Logistic Regression, dan Random Forest dalam menganalisis sentimen publik terhadap isu Indonesia Gelap yang ramai diperbincangkan di media sosial. Data dikumpulkan melalui proses crawling komentar pengguna TikTok menggunakan kata kunci terkait isu tersebut sehingga menghasilkan 5.000 data komentar. Setelah melalui tahapan preprocessing, sebanyak 4.667 komentar dinyatakan layak digunakan untuk proses analisis dan pelabelan sentimen menggunakan pendekatan lexicon-based ke dalam tiga kategori, yaitu positif, negatif, dan netral. Untuk mengatasi ketidakseimbangan distribusi kelas, penelitian ini menerapkan tiga pendekatan penanganan data, yaitu tanpa oversampling, oversampling sebelum pembagian data, serta oversampling setelah pembagian data pada data latih. Setiap model dievaluasi menggunakan empat metrik, yaitu akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa penerapan oversampling sebelum pembagian data memberikan performa terbaik pada seluruh model, dengan IndoBERT memperoleh nilai F1-score tertinggi sebesar 0,93, diikuti IndoBERTweet sebesar 0,91 serta algoritma klasik dengan rata-rata 0,89–0,90. Sebaliknya, pendekatan tanpa oversampling maupun oversampling setelah pembagian data pada data latih menghasilkan performa lebih rendah dengan nilai rata-rata F1-score berada pada kisaran 0,70–0,78. Temuan ini membuktikan bahwa model Transformer lebih efektif dalam memahami konteks bahasa informal pada media sosial, serta teknik penyeimbangan data sebelum proses pelatihan mampu meningkatkan performa klasifikasi secara signifikan pada data dengan distribusi kelas yang tidak merata.

Kata Kunci: Analisis Sentimen; IndoBERT; IndoBERTweet; Oversampling; Algoritma Klasik

Abstract—This study aims to compare the performance of Transformer-based models, namely IndoBERT and IndoBERTweet, with three classical machine learning algorithms, namely Support Vector Machine (SVM), Logistic Regression, and Random Forest, in analyzing public sentiment regarding the “Indonesia Gelap” issue that has been widely discussed on social media. The dataset was collected using a crawling process on TikTok user comments containing keywords related to the issue, resulting in 5.000 comments. After the preprocessing stage, 4.667 comments were deemed suitable for analysis and were labeled into positive, negative, and neutral sentiment categories using a lexicon-based approach. To address the imbalance in class distribution, three oversampling strategies were applied: without oversampling, oversampling before data splitting, and oversampling after data splitting applied only to the training data. Each model was evaluated using four performance metrics: accuracy, precision, recall, and F1-score. The results show that oversampling before data splitting yielded the best performance across all models, with IndoBERT achieving the highest F1-score of 0.93, followed by IndoBERTweet with 0.91, while the classical algorithms achieved average F1-scores ranging from 0.89 to 0.90. Meanwhile, both the non-oversampling scenario and oversampling after data splitting on the training data resulted in lower performance, with average F1-scores ranging from 0.70 to 0.78. These findings indicate that Transformer-based models are more effective in capturing informal language characteristics commonly found in social media comments. Furthermore, balancing the dataset before model training significantly improves the stability and performance of sentiment classification on imbalanced data.

Keywords: Sentiment Analysis; IndoBERT; IndoBERTweet; Oversampling; Classical Algorithm

1. PENDAHULUAN

Pada awal tahun 2025, media sosial di Indonesia dipenuhi dengan perbincangan mengenai isu Indonesia Gelap, sebuah topik yang menimbulkan kekhawatiran besar terhadap dinamika sosial, politik, dan ekonomi nasional. Tagar #IndonesiaGelap menjadi ramai diperbincangkan di berbagai platform dengan lebih dari 14 juta unggahan hanya dalam 24 jam pertama, di mana sekitar 81% unggahan mengandung sentimen negatif terhadap kebijakan pemerintah [1], [2]. Isu ini muncul sebagai bentuk reaksi publik terhadap rencana pemangkasan anggaran di sektor esensial seperti pendidikan dan kesehatan demi mendanai program unggulan seperti Makan Bergizi Gratis (MBG) serta pembentukan badan baru bernama Danantara [3], [4]. Fenomena ini menunjukkan bahwa saat ini media sosial tidak lagi hanya menjadi sarana hiburan, melainkan juga wadah bagi masyarakat untuk menyalurkan opini, aspirasi, serta kritik terhadap kebijakan pemerintah [5].

Meskipun demikian, komentar-komentar yang muncul di media sosial umumnya bersifat tidak terstruktur, penuh dengan singkatan, istilah gaul, dan emotikon [6]. Hal ini menjadi tantangan bagi sistem komputasi dalam memahami makna sesungguhnya dari opini publik. Dalam upaya menyelesaikan permasalahan itu, analisis sentimen dinilai tepat digunakan karena memiliki kemampuan untuk mengidentifikasi dan memahami kecenderungan opini masyarakat secara cepat dan komprehensif berdasarkan data teks yang tersebar di media sosial [7]. Analisis sentimen termasuk dalam cabang *Natural Language Processing* (NLP) yang berfokus pada pengelompokan opini atau emosi pengguna menjadi kategori positif, negatif, dan netral [8].

Perkembangan pesat di bidang NLP dalam beberapa tahun terakhir menghadirkan model berbasis Transformer, seperti *Bidirectional Encoder Representations from Transformers* (BERT), yang terbukti unggul dalam memahami

makna kontekstual kata secara dua arah [9]. Model ini kemudian dikembangkan untuk bahasa Indonesia melalui IndoBERT, yang dilatih menggunakan korpus besar Bahasa Indonesia serta menunjukkan kinerja tinggi pada tugas analisis sentimen dan klasifikasi teks [10]. Penelitian oleh Widyananda dkk. (2025) menunjukkan bahwa model IndoBERT mampu memberikan kinerja terbaik pada analisis sentimen ulasan *e-commerce* berbahasa Indonesia dengan akurasi mencapai 94,1%, mengungguli SVM 89,5% dan *Naive Bayes* 84,2% [11]. Peningkatan performa juga ditunjukkan oleh penelitian Jayadianti dkk. (2022) yang menggabungkan IndoBERT dengan arsitektur *Recurrent Convolutional Neural Network* (RCNN), menghasilkan akurasi 95,16%, precision 94,05%, recall 92,74%, dan F1-score 93,27% [12]. Di sisi lain, IndoBERTtweet dikembangkan untuk menangani karakter bahasa informal yang banyak ditemukan di media sosial. Penelitian oleh Setiawan dkk. (2024) melaporkan bahwa IndoBERTtweet mencapai akurasi 80% dan F1-score 80% pada analisis sentimen pengguna TikTok, lebih tinggi dibandingkan model LSTM yang hanya memperoleh akurasi 78% dan F1-score 77% [13]. Hal ini menunjukkan kemampuan IndoBERTtweet dalam memahami ekspresi dan struktur bahasa yang tidak baku di media sosial.

Sementara itu, penelitian dengan algoritma klasik masih terus dilakukan karena keunggulannya dalam efisiensi komputasi. Penelitian oleh Rohman dan Trisnapradika (2025) membandingkan *Naive Bayes Classifier* (NBC), *Support Vector Machine* (SVM), dan *Logistic Regression* pada data Twitter bertagor #KaburAjaDulu dan menunjukkan bahwa SVM mencapai akurasi tertinggi sebesar 85,63%, meningkat menjadi 86,74% setelah penerapan SMOTE [14]. Temuan serupa juga dilaporkan oleh Saputra dan Suryono (2025) yang membandingkan SVM, *Random Forest*, dan *Naive Bayes* dalam analisis sentimen penipuan di Twitter, di mana SVM memperoleh akurasi tertinggi mencapai 88%, *Random Forest* 84%, dan *Naive Bayes* 76% [15]. Kedua studi ini menunjukkan bahwa algoritma klasik tetap relevan untuk analisis opini publik karena efisien dan ringan. Penerapan SMOTE juga terbukti dapat meningkatkan akurasi, meskipun performanya masih terbatas dalam memahami konteks bahasa informal yang kompleks.

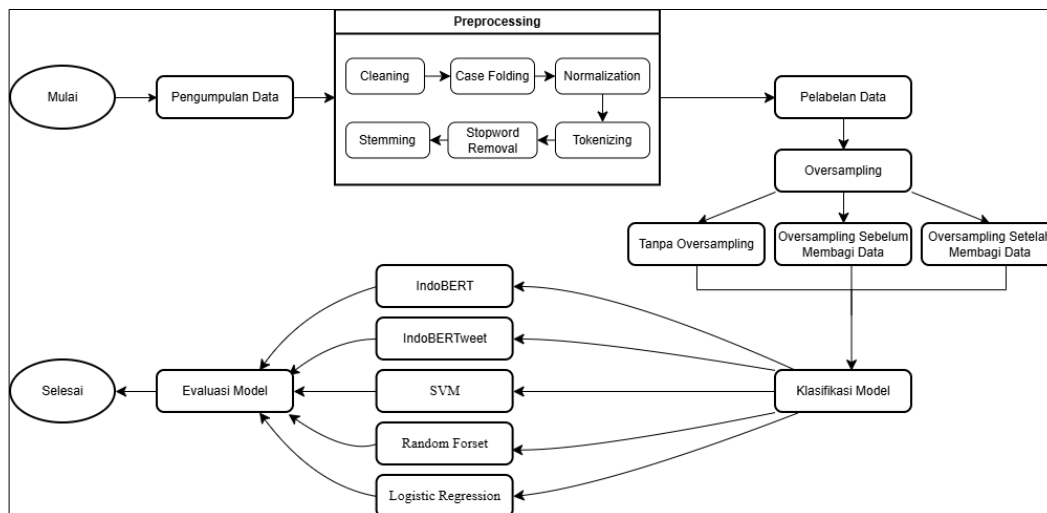
Meskipun berbagai penelitian telah menguji performa model Transformer maupun algoritma klasik pada tugas analisis sentimen, masih terdapat gap dari penelitian sebelumnya, yaitu belum adanya kajian yang secara komprehensif membandingkan kelima model yaitu (IndoBERT, IndoBERTtweet, SVM, *Random Forest*, dan *Logistic Regression*) sekaligus dalam satu penelitian secara bersamaan pada isu sosial viral “Indonesia Gelap”. Selain itu, penelitian terdahulu juga belum mengevaluasi pengaruh pendekatan untuk penanganan ketidakseimbangan data secara sistematis, khususnya melalui tiga pendekatan berbeda, yaitu tanpa oversampling, oversampling sebelum pembagian data, dan oversampling setelah pembagian data hanya pada data latih. Gap ini penting dikaji karena karakteristik model yang berbeda serta variasi pendekatan penanganan data tidak seimbang berpotensi menghasilkan perbedaan performa yang signifikan dalam klasifikasi sentimen pada data media sosial berbahasa Indonesia.

Berdasarkan gap tersebut, penelitian ini bertujuan untuk (1) membandingkan performa model Transformer (IndoBERT dan IndoBERTtweet) dan algoritma klasik (SVM, *Random Forest*, *Logistic Regression*) pada analisis sentimen isu Indonesia Gelap, serta (2) menganalisis pengaruh tiga pendekatan oversampling terhadap performa masing-masing model. Melalui pendekatan tersebut, penelitian ini diharapkan mampu menghasilkan pemahaman yang mendalam dan terstruktur terkait efektivitas setiap model dan pengaruh metode penyeimbangan data terhadap kinerja klasifikasi. Hasil penelitian juga diharapkan menjadi acuan berbasis data dalam pengembangan teknologi NLP berbahasa Indonesia serta mendukung perumusan kebijakan publik yang lebih tanggap terhadap opini masyarakat di media sosial.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dilaksanakan melalui serangkaian tahapan yang saling berkesinambungan untuk menghasilkan analisis sentimen yang optimal pada isu Indonesia Gelap. Proses dimulai dengan pengumpulan data komentar dari platform media sosial menggunakan teknik *crawling* berbasis kata kunci yang berkaitan dengan topik penelitian, kemudian data tersebut diproses melalui tahapan preprocessing untuk memastikan bahwa teks telah berada dalam kondisi yang konsisten dan layak digunakan pada tahap berikutnya. Setelah dibersihkan, data diberikan label sentimen menggunakan pendekatan *lexicon-based* untuk mengelompokkan komentar ke dalam kategori positif, negatif, dan netral secara sistematis. Tahap berikutnya mencakup penerapan tiga pendekatan untuk penanganan ketidakseimbangan data, yaitu pemodelan tanpa oversampling, oversampling yang dilakukan sebelum proses pembagian data, serta oversampling yang diterapkan setelah proses pembagian data hanya pada data latih guna melihat pengaruh pendekatan tersebut terhadap kinerja model. Selanjutnya, proses pelatihan dilakukan dengan memanfaatkan dua model Transformer (IndoBERT dan IndoBERTtweet) serta tiga algoritma klasik, yaitu SVM, *Random Forest*, dan *Logistic Regression*, sehingga diperoleh perbandingan kinerja yang lebih komprehensif. Evaluasi performa model dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score, yang kemudian dianalisis lebih lanjut menggunakan confusion matrix untuk mengidentifikasi pola prediksi pada setiap kelas sentimen. Keseluruhan alur penelitian ditampilkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Berdasarkan pada Gambar 1, tahapan penelitian disusun secara sistematis dan saling terhubung untuk memastikan data diproses secara terstruktur, konsisten, serta siap digunakan pada tahap pemodelan hingga evaluasi performa model.

2.2 Pengumpulan Data

Proses pengumpulan data dilakukan sebagai tahap awal untuk memperoleh opini publik yang akan diterapkan pada proses analisis sentimen. Data penelitian diperoleh dari platform TikTok dengan memanfaatkan teknik *crawling* sebagai metode pengumpulan data dengan bantuan layanan Apify. Proses pengambilan data dilakukan dengan memanfaatkan kata kunci yang berkaitan langsung dengan topik penelitian, lalu data mentah hasil *crawling* disimpan untuk keperluan untuk tahap analisis selanjutnya. Platform TikTok dipilih karena menjadi media sosial dengan tingkat interaksi tinggi serta mampu merepresentasikan opini masyarakat secara cepat dan luas terhadap isu yang sedang berkembang di Indonesia.

2.3 Preprocessing Data

Preprocessing merupakan tahapan penting yang bertujuan untuk memperbaiki kualitas data komentar yang diperoleh dari media sosial karena banyak mengandung *noise* serta penggunaan bahasa yang tidak baku [16]. Tahap ini berperan penting dalam memastikan setiap teks yang dianalisis memiliki struktur yang lebih rapi, konsisten, dan sesuai konteks bahasa sehingga informasi di dalamnya dapat diproses secara efektif oleh model klasifikasi sentimen. Dengan demikian, preprocessing membantu menghasilkan fitur yang lebih akurat dan mendukung peningkatan performa prediksi dalam melakukan analisis sentimen. Adapun tahapan preprocessing pada penelitian ini meliputi:

a. Cleaning

Tahapan *cleaning* merupakan proses awal dalam penyiapan data komentar yang diperoleh dari media sosial. Pada tahap ini, elemen-elemen yang tidak memberikan makna terhadap sentimen seperti URL, emoji, *mention*, *hashtag*, angka berlebihan, dan karakter yang tidak sesuai dengan struktur bahasa Indonesia dihapus agar informasi yang dianalisis tetap fokus pada opini yang disampaikan dalam teks [14].

b. Case Folding

Case folding diterapkan untuk menyeragamkan teks dengan mengubah setiap huruf menjadi bentuk huruf kecil [14]. Hal ini dilakukan agar model tidak memperlakukan kata yang sama namun berbeda bentuk penulisan seperti “Gelap” dan “gelap” sebagai istilah yang berbeda, sehingga proses analisis menjadi lebih konsisten.

c. Normalization

Normalisasi dilakukan untuk menyamakan penulisan kata dengan mengubah *slang*, singkatan, atau campuran bahasa menjadi bentuk baku sesuai kaidah Bahasa Indonesia [14]. Studi ini menerapkan kamus berisi kata-kata tidak baku yang diperoleh dari platform *Kaggle*, berisi pasangan kata *slang* dan padanan bakunya. Dengan tahapan ini, teks komentar menjadi lebih konsisten dan dapat diproses dengan lebih baik pada tahap pelabelan serta klasifikasi sentimen.

d. Tokenizing

Tahapan tokenisasi berfungsi untuk membagi teks menjadi unit-unit kata yang dikenal sebagai token agar dapat diproses secara individual dalam pemodelan bahasa. Dengan pemecahan ini, model dapat memahami setiap kata sebagai fitur atau bagian dari opini yang lebih spesifik dalam kalimat [14].

e. Stopword Removal

Proses *stopword removal* dilakukan untuk menyaring kata yang tidak memberikan pengaruh signifikan terhadap makna kontekstual, seperti ‘yang’, ‘dan’, serta ‘di’, karena tidak memberikan pengaruh berarti terhadap analisis

sentimen [14]. Dengan mengurangi kata yang tidak relevan, fitur yang dihasilkan lebih terfokus pada kata yang merepresentasikan opini, sehingga dapat meningkatkan akurasi model.

f. **Stemming**

Proses *stemming* diterapkan agar bisa mengubah setiap kata ke dalam bentuk kata dasar supaya tidak ada perbedaan makna antara kata yang sama namun memiliki tambahan imbuhan, sehingga model tidak memproses banyak bentuk kata yang sebenarnya memiliki arti serupa [8]. Dengan demikian, fitur yang dihasilkan menjadi lebih ringkas dan efisien dalam pemodelan.

2.4 Pelabelan Data

Tahap pelabelan data bertujuan untuk menentukan kategori sentimen pada setiap komentar yang telah melewati tahap preprocessing. Dalam penelitian ini digunakan pendekatan berbasis *lexicon-based* dengan mencocokkan kata dalam komentar terhadap kamus *lexicon* Bahasa Indonesia berisi daftar kata positif dan negatif yang diperoleh dari *repository* GitHub dan telah digunakan pada berbagai penelitian analisis sentimen. Setiap kata diberikan skor sesuai kategorinya, kemudian hasil penjumlahan skor menentukan label sentimen akhir berupa positif, negatif, atau netral. Pendekatan ini dipilih karena mampu melakukan pelabelan awal secara otomatis dan konsisten pada data media sosial dalam jumlah besar [17].

2.5 Oversampling

Tahapan oversampling merupakan teknik yang digunakan untuk menyeimbangkan proporsi data pada setiap kelas sentimen agar model dapat mempelajari karakteristik dari seluruh kelas secara setara dan tidak hanya berfokus pada kelas yang jumlahnya lebih besar [18]. Dalam penelitian ini, oversampling digunakan melalui tiga pendekatan sebagai bahan perbandingan, yaitu tanpa dilakukan oversampling, oversampling sebelum pembagian data, dan oversampling setelah pembagian data pada data latih. Pendekatan ini dipilih agar proses pelatihan model berlangsung seimbang pada setiap kategori sentimen, sehingga hasil evaluasi performanya menjadi lebih objektif.

2.6 Klasifikasi Model

Tahap klasifikasi model dilakukan untuk membandingkan kemampuan beberapa model dalam memprediksi tiga kategori sentimen, yaitu negatif, positif, dan netral. Lima model digunakan pada penelitian ini, yaitu IndoBERT dan IndoBERTweet sebagai model berbasis Transformer yang sesuai dengan karakteristik dalam memproses bahasa Indonesia dan teks media sosial [13]. Sementara itu SVM, *Logistic Regression*, serta *Random Forest* sebagai model pembanding dari algoritma klasik yang telah terbukti memiliki performa baik dalam pemrosesan teks dengan waktu komputasi yang relatif lebih cepat [7]. Seluruh model dilatih menggunakan pembagian data 80% untuk pelatihan dan 20% untuk pengujian dalam memastikan evaluasi yang lebih objektif. Penjelasan mengenai setiap model yang diterapkan dalam penelitian ini disajikan sebagai berikut:

2.6.1 IndoBERT

IndoBERT merupakan model berbasis Transformer yang dilatih menggunakan kumpulan data Bahasa Indonesia dalam jumlah yang besar sehingga mampu memahami konteks setiap kata dalam kalimat dengan lebih akurat [12]. Pada penelitian ini, alur penyelesaian klasifikasi sentimen menggunakan IndoBERT dimulai dari pemrosesan input teks melalui *WordPiece tokenizer*, kemudian setiap token diubah menjadi representasi numerik (*embedding*) yang menyimpan informasi semantik. Selanjutnya, *embedding* diproses oleh lapisan Transformer *encoder* untuk menangkap hubungan antarkata dalam kalimat, kemudian representasi akhir diarahkan ke lapisan klasifikasi untuk menentukan label sentimen. Model juga melalui proses *fine-tuning* menggunakan dataset komentar TikTok agar pola yang dipelajari relevan dengan karakteristik bahasa pada data penelitian, sehingga mampu menghasilkan klasifikasi sentimen berbasis konteks secara lebih tepat.

2.6.2 IndoBERTweet

IndoBERTweet merupakan pengembangan dari IndoBERT yang dilatih dengan korpus Twitter berbahasa Indonesia sehingga lebih adaptif terhadap karakteristik bahasa media sosial [13]. Proses penyelesaian tugas klasifikasi sentimen pada model ini dimulai dari tahapan tokenisasi menggunakan *SentencePiece*, kemudian setiap token diubah menjadi *embedding* dan diproses melalui arsitektur Transformer yang memanfaatkan *self-attention* untuk menangkap hubungan konteks antar kata dalam sebuah kalimat. Hasil pemrosesan tersebut kemudian diteruskan ke lapisan akhir untuk menentukan label sentimen. Perbedaan utama IndoBERTweet terletak pada data pelatihan dan desain tokenisasinya yang secara khusus disesuaikan dengan karakteristik bahasa pengguna media sosial, sehingga model ini lebih efektif dalam memahami pola penggunaan bahasa informal pada komentar publik sehingga prediksi sentimen yang dihasilkan lebih relevan dengan konteks data penelitian.

2.6.3 Support Vector Machine (SVM)

SVM merupakan metode pembelajaran mesin yang dimanfaatkan dalam menyelesaikan permasalahan klasifikasi dengan cara menentukan batas pemisah terbaik antara beberapa kelas dalam suatu ruang fitur [19]. Dalam analisis sentimen, SVM banyak digunakan karena mampu mengelola data teks dengan variasi penulisan yang tinggi dan tetap

menghasilkan pemisahan sentimen yang jelas pada komentar media sosial [14]. Kemampuan ini menjadikan SVM mampu secara akurat mengelompokkan komentar ke dalam tiga kategori sentimen negatif, positif, dan netral berdasarkan opini yang disampaikan pada teks tersebut. Fungsi keputusan SVM digunakan untuk menentukan kelas setiap komentar dan dapat dirumuskan sebagai berikut pada persamaan 1.

$$f(x) = w^T x + b \quad (1)$$

Pada persamaan tersebut, x merupakan fitur dari teks yang akan diklasifikasikan, w adalah bobot yang digunakan untuk menentukan arah pemisahan kelas, dan b merupakan nilai bias yang menyesuaikan posisi pemisah tersebut. Nilai $f(x)$ digunakan untuk menentukan kategori sentimen berdasarkan sisi pemisah yang ditempati oleh teks tersebut.

2.6.4 Logistic Regression

Logistic Regression merupakan algoritma dari pembelajaran mesin yang berfungsi menyelesaikan tugas klasifikasi dengan memprediksi probabilitas suatu data tergolong ke dalam kelas tertentu. Dalam analisis sentimen, model ini sering digunakan sebagai model pembanding karena proses pelatihannya sederhana, efisien, dan mampu memberikan pemahaman yang lebih jelas mengenai kontribusi setiap fitur terhadap keputusan klasifikasi [20]. *Logistic Regression* dapat mengkategorikan komentar sebagai negatif, positif, atau netral berdasarkan kecenderungan opini yang terdapat pada teks tersebut. Persamaan fungsi dari Logistic Regression dapat dirumuskan sebagai berikut pada persamaan 2.

$$P(y = 1 | x) = \frac{1}{1 + e^{-(w^T x + b)}} \quad (2)$$

Pada persamaan tersebut, x merepresentasikan fitur dari teks yang dianalisis, w merupakan bobot yang menggambarkan kontribusi tiap fitur dalam memprediksi sentimen, dan b merupakan nilai bias yang menyesuaikan posisi batas keputusan. Nilai $P(y = 1 | x)$ menunjukkan probabilitas suatu komentar termasuk dalam sentimen tertentu, dan proses klasifikasi dilakukan dengan membandingkan nilai tersebut terhadap ambang batas yang telah ditentukan sehingga komentar dapat dipetakan ke dalam kelas sentimen yang sesuai.

2.6.5 Random Forest

Random Forest merupakan pendekatan *ensemble* yang memadukan beberapa pohon keputusan (*decision tree*) dengan tujuan meningkatkan akurasi dan meminimalkan resiko *overfitting* pada data teks, termasuk dalam analisis sentimen Bahasa Indonesia [21]. Pendekatan ini sangat cocok untuk data media sosial yang memiliki variasi struktur bahasa dan fitur yang kompleks karena setiap pohon dalam *ensemble* memberikan prediksi, kemudian keputusan akhir ditentukan berdasarkan suara mayoritas dari seluruh pohon. Fungsi prediksi algoritma *Random Forest* dapat dirumuskan sebagai berikut pada persamaan 3.

$$\hat{y} = \text{majority_vote}\{h_t(x)\}_{t=1}^T \quad (3)$$

Pada persamaan tersebut, $h_t(x)$ merupakan hasil prediksi dari pohon keputusan ke- t , T adalah jumlah seluruh pohon yang digunakan dalam model, dan \hat{y} adalah hasil prediksi akhir berdasarkan suara terbanyak. Dengan cara ini, Random Forest dapat memberikan keputusan klasifikasi yang lebih konsisten meskipun data memiliki variasi penulisan yang tinggi seperti pada komentar media sosial.

2.7 Evaluasi Model

Proses evaluasi dimaksudkan untuk menilai performa model klasifikasi agar dapat membedakan komentar ke dalam kategori sentimen negatif, positif, dan netral secara akurat dan konsisten. Evaluasi dilakukan menggunakan metrik standar klasifikasi, yaitu akurasi, presisi, recall, dan F1-score. Keempat metrik ini dimanfaatkan untuk mengevaluasi performa model dalam memprediksi setiap kelas sentimen secara tepat dan seimbang, sekaligus menjadi acuan untuk membandingkan performa antar model yang diterapkan dengan berbagai teknik *oversampling*. Penilaian ini penting agar model tidak hanya menunjukkan akurasi tinggi, tetapi juga mampu menghasilkan prediksi yang seimbang di seluruh kategori sentimen. Rumus akurasi ditunjukkan pada Persamaan 4.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Kemudian presisi dan recall masing-masing dihitung melalui Persamaan 5 dan 6.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (6)$$

Sedangkan F1-score dirumuskan sebagai rata-harmonik antara presisi dan recall pada Persamaan 7.

$$\text{F1score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$



Secara keseluruhan, keempat metrik tersebut digunakan secara bersamaan untuk menilai kualitas hasil klasifikasi model. Akurasi mengukur seberapa besar proporsi prediksi yang benar dari keseluruhan data uji, sementara presisi menggambarkan tingkat ketepatan model dalam memprediksi komentar yang benar pada suatu kelas tertentu. Recall menilai kemampuan model dalam mengenali seluruh data yang benar dari tiap kelas, sedangkan F1-score mengombinasikan presisi dan recall guna menghasilkan ukuran performa yang seimbang untuk memastikan kinerja model tetap stabil meskipun terjadi ketidakseimbangan jumlah data antar kelas [22]. Dengan menggunakan perhitungan keempat metrik ini, performa setiap model dapat dievaluasi secara lebih menyeluruh, sehingga model dengan hasil terbaik benar-benar mencerminkan kemampuan dalam menganalisis sentimen pada komentar media sosial.

3. HASIL DAN PEMBAHASAN.

3.1 Hasil Pengumpulan Data

Proses pengumpulan data melalui teknik *crawling* dengan layanan Apify berhasil memperoleh sekitar 5.000 komentar dari TikTok selama periode Januari hingga Maret 2025, menggunakan kata kunci yang relevan dengan topik penelitian. Dataset mentah yang diperoleh berisi berbagai atribut penting seperti *avatarThumbnail*, *cid*, *createTimeISO*, *diggCount*, *replyCommentTotal*, *videoWebUrl*, *uniqueId*, dan *text*. Meskipun dataset memiliki banyak kolom, penelitian ini berfokus hanya pada kolom *text* sebagai sumber utama analisis sentimen, karena memuat opini dan ekspresi pengguna terhadap isu yang sedang dibahas. Oleh karena itu, pada bagian penyajian data ini hanya ditampilkan kolom teks komentar, dengan tujuan menjaga fokus penelitian. Hasil pengumpulan data yang diperoleh melalui proses *crawling* selanjutnya disajikan pada Tabel 1 berikut.

Tabel 1. Pengumpulan Data

No	Text
1 5000	yg melihat Indonesia gelap yg masih punya nurani dan kejujuran, yg tidak gila kekuasaan. apa kabar kasus pagar laut

Berdasarkan Tabel 1, dapat dilihat bahwa data komentar yang dikumpulkan merepresentasikan opini pengguna yang bersifat responsif, emosional, dan ekspresif terhadap isu yang dibahas. Komentar yang diperoleh memiliki variasi gaya penulisan, mulai dari bahasa informal, penggunaan emotikon, hingga pernyataan yang mengandung sentimen eksplisit. Karakteristik tersebut menjadi alasan penting dilakukannya tahap *preprocessing* sebelum masuk ke proses analisis sentimen, guna memastikan data siap digunakan pada tahap klasifikasi model.

3.2 Hasil Preprocessing Data

Tahap *preprocessing* dilakukan untuk meningkatkan kualitas data dengan menghilangkan komentar yang tidak relevan serta menyeragamkan teks sebelum dilakukan proses analisis sentimen. Dari total sekitar 5.000 komentar hasil *crawling* pada platform TikTok, terdapat komentar yang harus dihapus karena hanya berisi karakter khusus, spam, atau tidak mengandung opini. Setelah melalui tahapan *cleaning*, *case folding*, *normalization*, *tokenizing*, *stopword removal*, dan *stemming*, diperoleh sebanyak 4.667 komentar yang layak digunakan untuk klasifikasi sentimen. Proses *preprocessing* ini bertujuan untuk menghasilkan data teks yang lebih bersih, terstruktur, dan memiliki makna yang relevan tanpa menghilangkan konteks utama dari opini yang disampaikan. Untuk menunjukkan alur perubahan data, digunakan satu sampel komentar dari dataset yang ditampilkan pada setiap tahapan *preprocessing* yang ditampilkan pada Tabel 2.

Tabel 2. Penerapan Preprocessing

Tahapan	Hasil
Data Awal	yg melihat Indonesia gelap yg masih punya nurani dan kejujuran, yg tidak gila kekuasaan.
<i>Cleaning</i>	yg melihat Indonesia gelap yg masih punya nurani dan kejujuran yg tidak gila kekuasaan
<i>Case Folding</i>	yg melihat indonesia gelap yg masih punya nurani dan kejujuran yg tidak gila kekuasaan
<i>Normalization</i>	yang melihat indonesia gelap yang masih punya nurani dan kejujuran yang tidak gila kekuasaan
<i>tokenizing</i>	['yang', 'melihat', 'indonesia', 'gelap', 'yang', 'masih', 'punya', 'nurani', 'dan', 'kejujuran', 'yang', 'tidak', 'gila', 'kekuasaan']
<i>Stopword Removal</i>	['indonesia', 'gelap', 'nurani', 'kejujuran', 'gila', 'kekuasaan']
<i>Stemming</i>	indonesia gelap nurani jujur gila kuasa

Pada Tabel 2 menunjukkan bahwa perubahan setelah *preprocessing* berhasil mengarahkan fokus teks pada konteks isu yang sedang dianalisis. Dengan demikian, fitur teks yang dihasilkan menjadi lebih representatif dan siap digunakan dalam proses pelabelan sentimen maupun klasifikasi model pada tahap berikutnya.

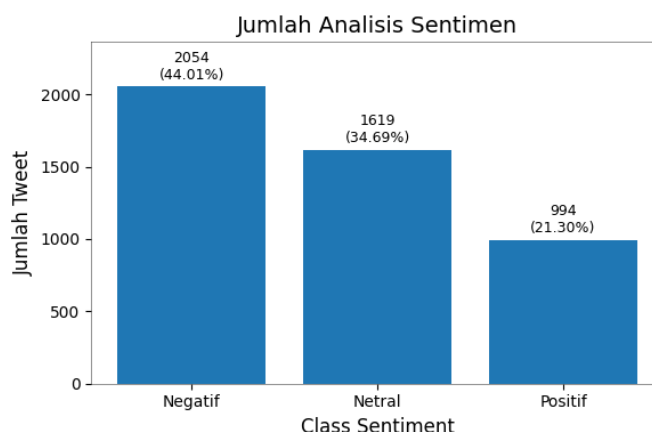
3.3 Hasil Pelabelan Data

Pelabelan sentimen diterapkan melalui metode berbasis kamus (*lexicon-based*), yaitu sistem memberikan skor pada setiap komentar berdasarkan kecocokan kata dengan kamus *lexicon* berbahasa Indonesia yang memuat daftar dari kata positif dan negatif. Setelah skor dihitung, komentar diklasifikasikan menjadi tiga kategori sentimen, yaitu positif, negatif, atau netral. Hasil pelabelan beberapa komentar ditampilkan pada Tabel 3.

Tabel 3. Pelabelan Data

Sebelum	Sesudah	Skor	Label
beli Pertamina d kasih pertalite	beli pertamax kasih pertalite	0	Netral
contohnya kader pdi banyak ketangkap 🤔	contoh kader pdi tangkap	-1	Negatif
presiden its just nonsense	Presiden omong kosong	-1	Negatif
Semoga Sukses Prabowo Gibran	moga sukses prabowo gibran	2	Positif
apa kabar kasus pagar laut	kabar pagar laut	0	Netral
kan emng gunanya efisiensi anggaran itu sebagian untuk membenerin fasilitas pendidikan	guna efisiensi anggar membenerin fasilitas didik	2	Positif

Berdasarkan hasil pelabelan, diperoleh sebanyak 994 komentar positif (21,30%), 1.619 komentar netral (34,69%), dan 2.054 komentar negatif (44,01%) dari total 4.667 komentar yang telah melalui tahap preprocessing. Dari distribusi tersebut terlihat bahwa sentimen negatif mendominasi komentar pada isu yang sedang dianalisis, dengan menunjukkan tingginya ekspresi ketidakpuasan atau kritik dari masyarakat terhadap topik yang dibahas. Visualisasi distribusi pelabelan sentimen ditampilkan pada Gambar 2.



Gambar 2. Hasil Pelabelan Data

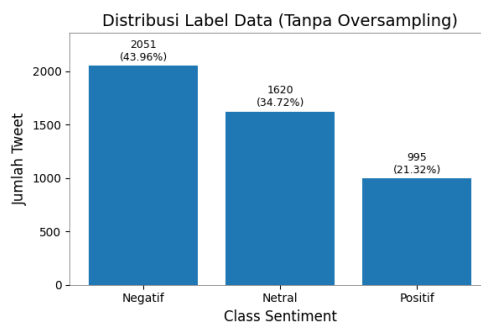
Terlihat pada Gambar 2 menunjukkan dominasi dari kelas negatif, hal ini perlu menjadi perhatian khusus karena dapat menyebabkan model lebih cenderung memprediksi sentimen ke arah negatif, sehingga penanganan ketidakseimbangan kelas diperlukan pada tahap pemodelan berikutnya agar hasil prediksi tidak bias terhadap kelas mayoritas.

3.4 Hasil Penanganan Ketidakseimbangan Data

Ketidakseimbangan data merupakan salah satu masalah yang umum muncul pada analisis sentimen, terutama pada data media sosial yang menunjukkan distribusi sentimen yang tidak merata antar kelas. Berdasarkan hasil pelabelan sebelumnya, data dengan sentimen negatif memiliki jumlah yang jauh lebih tinggi dibandingkan sentimen positif dan netral. Ketidakseimbangan ini berpotensi menurunkan kinerja model karena model cenderung lebih sering memprediksi kelas dengan jumlah data terbanyak. Untuk mengatasi hal tersebut, penelitian ini menerapkan tiga metode penanganan data tidak seimbang, yaitu tanpa oversampling, oversampling sebelum pembagian data, dan oversampling setelah pembagian data pada data latih.

3.3.1 Tanpa Oversampling

Pada metode pertama, data digunakan secara langsung tanpa melalui proses oversampling sehingga distribusi kelas dibiarkan apa adanya sesuai kondisi awal dataset. Pendekatan ini dilakukan untuk menggambarkan bagaimana performa model ketika berhadapan dengan data yang tidak seimbang, di mana kelas tertentu memiliki jumlah yang jauh lebih besar dibandingkan kelas lainnya. Dengan tidak melakukan penyeimbangan jumlah data, model berpotensi lebih sering mengenali pola dari kelas mayoritas, sehingga hasil yang diperoleh dapat memberikan gambaran awal mengenai dampak ketidakseimbangan tersebut terhadap performa model. Distribusi kelas yang digunakan pada metode ini ditunjukkan pada Gambar 3.

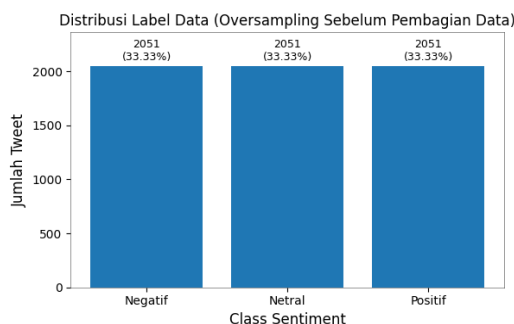


Gambar 3. Hasil Pelabelan Tanpa Oversampling

Pada Gambar 3 menunjukkan bahwa kelas negatif masih menjadi yang paling dominan dengan jumlah 2.051 data (43,96%), diikuti kelas netral sebanyak 1.620 data (34,72%), dan positif sebanyak 995 data (21,32%). Kondisi ini menggambarkan adanya ketidakseimbangan data yang cukup besar antara kelas mayoritas dan minoritas. Ketidakseimbangan seperti ini dapat menyebabkan model lebih sering memprediksi sentimen negatif karena data tersebut lebih banyak muncul dalam proses pelatihan.

3.3.2 Oversampling Sebelum Pembagian Data

Pada metode kedua, proses oversampling dilakukan sebelum data dibagi menjadi data latih dan data uji. Teknik ini bertujuan untuk memastikan bahwa distribusi kelas menjadi seimbang sejak awal, sehingga setiap kelas memiliki proporsi data yang sama dalam keseluruhan dataset. Dengan demikian, model diharapkan mampu mempelajari pola sentimen secara merata pada seluruh kelas. Hasil penyeimbangan kelas ditunjukkan pada Gambar 4.

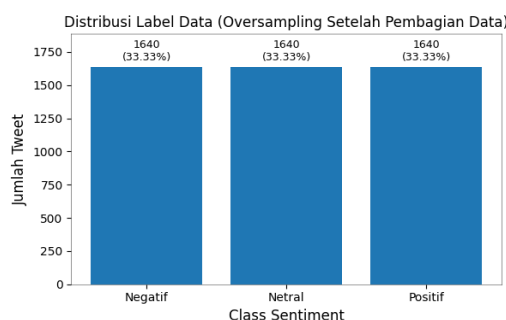


Gambar 4. Hasil Pelabelan Oversampling Sebelum Pembagian Data

Gambar 4 memperlihatkan bahwa setelah dilakukan oversampling, setiap kelas yaitu positif, netral, dan negatif masing-masing berjumlah 2.051 data (33,33%). Dengan pendekatan ini, model pelatihan diharapkan dapat belajar dengan lebih adil terhadap seluruh kelas sentimen karena tidak ada perbedaan jumlah data antar kelas. Namun, metode ini berpotensi menimbulkan risiko kebocoran informasi (*data leakage*) karena oversampling dilakukan sebelum pemisahan data latih dan uji.

3.3.3 Oversampling Setelah Pembagian Data

Pada metode ketiga, proses oversampling dilakukan hanya pada data latih setelah proses pembagian data dilakukan, sementara data uji tetap dipertahankan dalam kondisi aslinya. Pendekatan ini dipilih untuk mencegah terjadinya *data leakage* dari data uji ke data latih, sehingga evaluasi performa model tetap relevan dan mencerminkan kemampuan sesungguhnya dalam menangani data yang tidak seimbang. Distribusi hasil oversampling setelah pembagian data ditunjukkan pada Gambar 5.



Gambar 5. Hasil Pelabelan Oversampling Setelah Pembagian Data

Berdasarkan Gambar 5, distribusi data latih menjadi seimbang dengan masing-masing kelas berjumlah 1.640 data (33,33%). Pendekatan ini memungkinkan model belajar secara seimbang tanpa mengubah karakter asli pada data uji, sehingga hasil pengujian lebih objektif.

Ketiga metode ini digunakan untuk membandingkan pengaruh dari penyeimbangan data terhadap performa model klasifikasi pada tahap berikutnya. Dengan adanya variasi tersebut, dapat dianalisis sejauh mana oversampling berkontribusi dalam mengatasi bias model terhadap kelas mayoritas dan meningkatkan kemampuan klasifikasi terhadap sentimen positif maupun netral.

3.5 Hasil Klasifikasi Model

Tahap klasifikasi model dilakukan untuk mengevaluasi performa lima algoritma yang digunakan dalam analisis sentimen pada penelitian ini, yaitu IndoBERT, IndoBERTtweet, *Support Vector Machine* (SVM), *Logistic Regression*, dan *Random Forest*. Model IndoBERT dan IndoBERTtweet dilatih menggunakan pendekatan *fine-tuning* pada *pretrained* model “*indobenchmark/indobert-base-pl*” dan “*indolem/indoberttweet-base-uncased*” dengan parameter pelatihan *learning rate* $2e-5$, *batch size* 8, *epoch* 9, dan *weight decay* 0.01 untuk mengoptimalkan pemahaman konteks bahasa Indonesia. Sementara itu, model klasik seperti SVM, *Logistic Regression*, dan *Random Forest* dilatih menggunakan representasi teks berbasis *TF-IDF Vectorizer* dan dilakukan optimasi dengan *GridSearchCV* untuk mencari parameter terbaik melalui *3-fold cross-validation*. Evaluasi dilakukan dengan membagi data menjadi 80% untuk pelatihan dan 20% untuk pengujian pada tiga kondisi berbeda, yaitu tanpa oversampling, oversampling sebelum pembagian data, dan oversampling setelah pembagian data. Kinerja setiap model diukur menggunakan metrik akurasi, presisi, recall, dan F1-score untuk menilai kemampuan masing-masing model dalam mengenali tiga kategori sentimen, yaitu positif, netral, dan negatif.

3.4.1 Hasil Tanpa Oversampling

Penerapan tanpa oversampling dilakukan untuk menilai performa dasar setiap model terhadap distribusi kelas yang tidak seimbang. Kondisi ini mencerminkan situasi alami data media sosial yang didominasi oleh sentimen negatif. Hasil pengujian menunjukkan bahwa model IndoBERT memberikan hasil yang relatif stabil dengan akurasi 0.78, sedangkan IndoBERTtweet menunjukkan performa paling rendah di antara seluruh model. *Logistic Regression* menjadi model terbaik dengan akurasi 0.80, diikuti SVM dan IndoBERT yang memiliki nilai serupa, sementara *Random Forest* mencatat akurasi 0.76. Perbandingan nilai presisi, recall, F1-score, dan akurasi dari kelima model tanpa oversampling disajikan pada Tabel 4.

Tabel 4. Hasil Evaluasi Tanpa Oversampling

Model	Sentimen	Precision	Recall	F1-Score	Accuracy
IndoBERT	Negatif	0.85	0.82	0.83	0.78
	Netral	0.69	0.78	0.73	
	Positif	0.80	0.70	0.75	
IndoBERTtweet	Negatif	0.83	0.83	0.83	0.76
	Netral	0.68	0.72	0.70	
	Positif	0.74	0.66	0.70	
SVM	Negatif	0.83	0.87	0.85	0.78
	Netral	0.71	0.69	0.70	
	Positif	0.80	0.75	0.77	
Logistic Regression	Negatif	0.87	0.85	0.86	0.80
	Netral	0.70	0.76	0.73	
	Positif	0.82	0.74	0.78	
Random Forest	Negatif	0.79	0.83	0.81	0.76
	Netral	0.72	0.72	0.72	
	Positif	0.74	0.66	0.69	

Berdasarkan hasil pada Tabel 4, terlihat bahwa ketidakseimbangan data menyebabkan performa model belum optimal, terutama pada kelas positif dan netral yang jumlahnya lebih sedikit dibandingkan kelas negatif. Model klasik seperti *Logistic Regression* dan SVM masih mampu mempertahankan kinerja yang stabil, sedangkan model berbasis Transformer seperti IndoBERT menunjukkan potensi yang baik meskipun dipengaruhi distribusi data yang tidak merata. Untuk melihat sejauh mana teknik penyeimbangan data dapat meningkatkan performa model, tahap berikutnya dilakukan dengan menerapkan oversampling sebelum pembagian data.

3.4.2 Hasil Oversampling Sebelum Pembagian Data

Penerapan oversampling sebelum pembagian data dilakukan untuk menyeimbangkan jumlah data pada ketiga kelas sentimen. Pendekatan ini bertujuan agar seluruh model memperoleh proporsi data yang setara saat proses pelatihan, sehingga model tidak bias terhadap kelas dengan jumlah data yang lebih banyak. Hasil pengujian menunjukkan adanya peningkatan performa yang cukup signifikan dibandingkan tahap sebelumnya tanpa oversampling. Model berbasis

Transformer seperti IndoBERT mencapai performa tertinggi dengan akurasi 0.93, diikuti oleh IndoBERTweet dengan akurasi 0.91. Model klasik seperti SVM, *Logistic Regression*, dan *Random Forest* juga menunjukkan peningkatan stabil dengan akurasi yang sama, yaitu sekitar 0.90. Nilai F1-score pada seluruh model menunjukkan konsistensi di atas 0.88, menandakan bahwa oversampling sebelum pembagian data berhasil membantu model mengenali ketiga kelas sentimen secara seimbang. Perbandingan hasil evaluasi setiap model dapat dilihat pada Tabel 5, yang menampilkan nilai presisi, recall, F1-score, dan akurasi dari masing-masing model.

Tabel 5. Hasil Evaluasi Oversampling Sebelum Pembagian Data

Model	Sentimen	Precision	Recall	F1-Score	Accuracy
IndoBERT	Negatif	0.94	0.92	0.93	0.93
	Netral	0.90	0.92	0.91	
	Positif	0.95	0.95	0.95	
IndoBERTweet	Negatif	0.92	0.89	0.90	0.91
	Netral	0.86	0.90	0.88	
	Positif	0.95	0.94	0.94	
SVM	Negatif	0.91	0.88	0.90	0.90
	Netral	0.86	0.86	0.86	
	Positif	0.92	0.95	0.93	
Logistic Regression	Negatif	0.93	0.88	0.90	0.90
	Netral	0.86	0.87	0.87	
	Positif	0.91	0.95	0.93	
Random Forest	Negatif	0.91	0.86	0.88	0.90
	Netral	0.87	0.89	0.88	
	Positif	0.91	0.94	0.92	

Pada Tabel 5 menunjukkan bahwa penerapan oversampling sebelum pembagian data berhasil meningkatkan keseimbangan kinerja seluruh model pada ketiga kelas sentimen. Model IndoBERT mengalami peningkatan paling signifikan dibandingkan tahap sebelumnya, sementara model klasik seperti SVM, *Logistic Regression*, dan *Random Forest* tetap menunjukkan kestabilan performa. Hasil yang didapat, metode ini terbukti efektif dalam mengurangi bias kelas dan meningkatkan kinerja model dalam mendeteksi seluruh kategori sentimen secara seimbang.

3.4.3 Hasil Oversampling Setelah Pembagian Data

Penerapan oversampling setelah pembagian data dilakukan untuk menyeimbangkan distribusi kelas hanya pada data latih, sedangkan data uji tetap menggunakan proporsi asli. Pendekatan ini bertujuan untuk meningkatkan kemampuan model dalam mengenali kelas minoritas tanpa mengubah distribusi data validasi, sehingga hasil evaluasi tetap mencerminkan kondisi nyata. Berdasarkan hasil pengujian, performa seluruh model menunjukkan sedikit penurunan dibandingkan dengan pendekatan menggunakan oversampling sebelum pembagian data. Model IndoBERT memperoleh akurasi tertinggi sebesar 0.77, diikuti oleh SVM dan *Logistic Regression* dengan akurasi masing-masing 0.75, sedangkan *Random Forest* mencatat 0.72 dan IndoBERTweet menjadi yang terendah dengan 0.70. Rata-rata nilai F1-score untuk seluruh model berada pada kisaran 0.70 hingga 0.77, di mana IndoBERT memperoleh nilai tertinggi dan IndoBERTweet menjadi yang terendah. Hal ini menunjukkan bahwa meskipun oversampling membantu menyeimbangkan distribusi kelas pada data latih, dampaknya terhadap peningkatan performa keseluruhan tidak sebesar pada pendekatan oversampling sebelum pembagian data. Perbandingan hasil evaluasi perbandingan nilai presisi, recall, F1-score, dan akurasi untuk setiap model dapat dilihat pada Tabel 6.

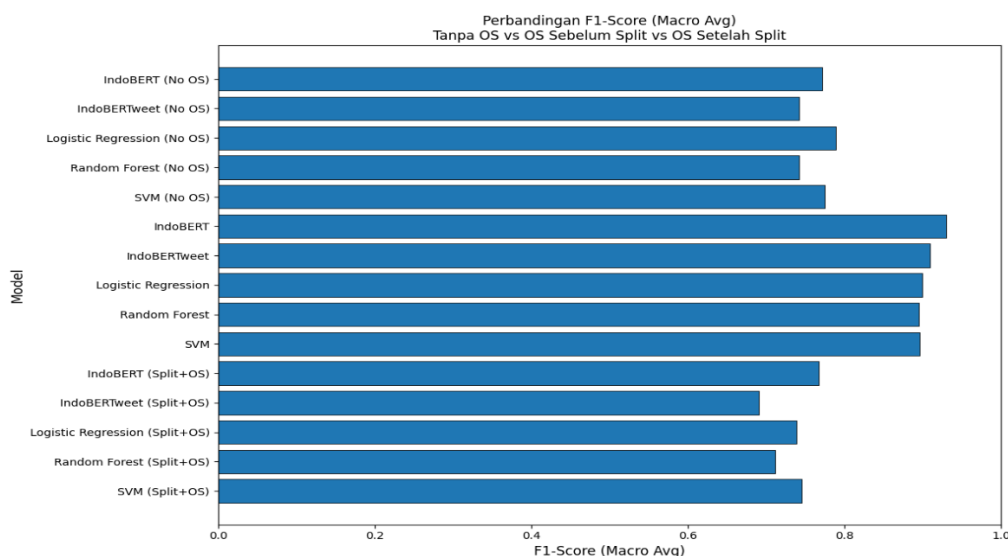
Tabel 6. Hasil Evaluasi Oversampling Setelah Pembagian Data

Model	Sentimen	Precision	Recall	F1-Score	Accuracy
IndoBERT	Negatif	0.83	0.82	0.82	0.77
	Netral	0.71	0.72	0.71	
	Positif	0.76	0.77	0.77	
IndoBERTweet	Negatif	0.80	0.74	0.77	0.70
	Netral	0.62	0.66	0.64	
	Positif	0.66	0.67	0.66	
SVM	Negatif	0.85	0.80	0.82	0.75
	Netral	0.65	0.69	0.67	
	Positif	0.72	0.75	0.74	
Logistic Regression	Negatif	0.85	0.80	0.82	0.75
	Netral	0.65	0.69	0.67	
	Positif	0.72	0.72	0.72	
Random Forest	Negatif	0.82	0.74	0.78	0.72
	Netral	0.66	0.69	0.67	
	Positif	0.65	0.72	0.68	

Hasil pada Tabel 6 menunjukkan bahwa penerapan oversampling setelah pembagian data tidak memberikan peningkatan performa yang signifikan, bahkan cenderung menurun jika dibandingkan dengan hasil oversampling sebelum pembagian data. Nilai akurasi dan F1-score pada sebagian besar model kembali mendekati hasil tanpa oversampling, atau bahkan sedikit lebih rendah, menandakan bahwa penerapan oversampling setelah pembagian data tidak efektif dalam menyeimbangkan distribusi kelas pada tahap pelatihan. Model IndoBERT masih menunjukkan kestabilan terbaik di antara model lainnya, sedangkan IndoBERTweet dan *Random Forest* mengalami penurunan paling besar. Secara keseluruhan, strategi oversampling setelah pembagian data kurang optimal dibandingkan pendekatan sebelumnya yang dilakukan sebelum pembagian data.

3.6 Hasil Perbandingan Evaluasi

Hasil evaluasi dilakukan terhadap lima model, yaitu IndoBERT, IndoBERTweet, *Logistic Regression*, *Random Forest*, dan SVM dengan membandingkan tiga pendekatan penanganan ketidakseimbangan data yaitu tanpa oversampling, oversampling sebelum pembagian data, dan oversampling setelah pembagian data pada data latih. Nilai F1-score dipilih sebagai metrik utama karena mampu menggambarkan keseimbangan presisi dan recall pada kondisi distribusi kelas yang tidak seimbang seperti komentar media sosial. Perbandingan nilai F1-score dari ketiga pendekatan dapat dilihat pada Gambar 6 berikut.



Gambar 6. Hasil Perbandingan Keseluruhan Model

Berdasarkan pada Gambar 6 dari hasil perbandingan tiga pendekatan, performa terbaik diperoleh pada strategi oversampling yang dilakukan sebelum pembagian data, di mana model IndoBERT mencapai nilai F1-score tertinggi sebesar 0.93, diikuti IndoBERTweet sebesar 0.91, serta model klasik seperti *Logistic Regression*, SVM, dan *Random Forest* yang berada pada kisaran 0.89–0.90. Kondisi ini memperlihatkan bahwa penyeimbangan distribusi kelas sebelum proses pembagian data membantu model mengenali pola sentimen secara lebih proporsional dan stabil di seluruh kelas. Sebaliknya, pada pendekatan tanpa oversampling, seluruh model menunjukkan penurunan performa dengan rata-rata F1-score hanya berkisar antara 0.74–0.78 akibat ketidakseimbangan kelas yang menyebabkan bias pada prediksi. Adapun penerapan oversampling setelah pembagian data tidak menunjukkan peningkatan yang signifikan, bahkan beberapa model mengalami penurunan performa hingga mendekati hasil tanpa oversampling, seperti pada IndoBERTweet dan *Random Forest* dengan F1-score di bawah 0.72. Secara keseluruhan, hasil ini menegaskan bahwa strategi oversampling yang diterapkan sebelum pembagian data merupakan pendekatan paling efektif dalam meningkatkan keseimbangan performa antar kelas sentimen.

4. KESIMPULAN

Penelitian ini membandingkan performa model berbasis Transformer, yaitu IndoBERT dan IndoBERTweet, dengan tiga algoritma klasik, yaitu *Logistic Regression*, SVM, dan *Random Forest* dalam menganalisis sentimen publik terhadap isu Indonesia Gelap. Hasil penelitian menunjukkan bahwa penerapan teknik oversampling berpengaruh signifikan terhadap performa klasifikasi sentimen, terutama pada data dengan distribusi kelas yang tidak seimbang. Berdasarkan hasil evaluasi, strategi oversampling sebelum pembagian data memberikan hasil terbaik dibandingkan dua pendekatan lainnya, di mana model IndoBERT mencapai nilai rata-rata F1-score dan akurasi tertinggi sebesar 0.93, diikuti IndoBERTweet dengan 0.91, serta model klasik yang berada pada kisaran 0.89–0.90. Hal ini membuktikan bahwa penyeimbangan data sebelum proses pelatihan membantu model mengenali karakteristik sentimen secara lebih proporsional dan konsisten di semua kelas. Sebaliknya, pendekatan tanpa oversampling maupun

oversampling setelah pembagian data menghasilkan performa yang lebih rendah dengan rata-rata F1-score berkisar antara 0.70–0.78, menandakan bahwa ketidakseimbangan data berdampak langsung terhadap bias prediksi, terutama pada kelas minoritas. Secara keseluruhan, model IndoBERT dengan penerapan oversampling sebelum pembagian data merupakan pendekatan paling efektif dalam analisis sentimen komentar media sosial berbahasa Indonesia, serta memperkuat efektivitas model berbasis Transformer dalam memahami konteks bahasa informal dengan hasil yang lebih akurat dibandingkan model klasik. Meskipun demikian, penelitian ini memiliki keterbatasan pada jumlah data yang terbatas serta penggunaan pendekatan oversampling sederhana yang berpotensi menimbulkan overfitting. Oleh karena itu, penelitian selanjutnya disarankan untuk menerapkan teknik penyeimbangan data yang lebih adaptif seperti ADASYN atau SMOTE, serta memperluas korpus data agar hasil analisis sentimen dapat lebih representatif terhadap opini publik di media sosial Indonesia.

REFERENCES

- [1] H. Yaputra, “Tagar Indonesia Gelap Dapat 81 Persen Sentimen Negatif,” *Tempo*. Accessed: Nov. 02, 2025. [Online]. Available: <https://www.tempo.co/politik/tagar-indonesia-gelap-dapat-81-persen-sentimen-negatif-1221552>
- [2] D. T. Aswan, “Indonesia Gelap Trending X Capai 14 Juta Cuitan 24 Jam, Perkara Aksi Demo Mahasiswa Gegara Efisiensi,” *Tribunnews*. Accessed: Nov. 02, 2025. [Online]. Available: <https://sultra.tribunnews.com/2025/02/18/indonesia-gelap-trending-x-capai-14-juta-cuitan-24-jam-perkara-aksi-demo-mahasiswa-gegara-efisiensi>
- [3] I. Khozen, “Indonesia Gelap: Ketakutan publik atas negara yang direbut,” Publication: Fakultas Ilmu Administrasi Universitas Indonesia (FIA UI). Accessed: Nov. 02, 2025. [Online]. Available: <https://fia.ui.ac.id/indonesia-gelap-a-public-fear-over-a-seized-country/>
- [4] A. Firdaus, “Apa itu ‘Indonesia Gelap’, tajuk unjuk rasa mahasiswa di beberapa daerah?,” *BenarNews*. Accessed: Nov. 02, 2025. [Online]. Available: <https://www.benarnews.org/indonesian/berita/apa-itu-indonesia-gelap-02212025063250.html>
- [5] D. Andzani, D. Virgin, B. Pristica, and D. L. Dwihadiah, “Analisis Peran Media Sosial Dalam Proses Mediatisasi Politik: Perspektif Komunikasi Politik dan Partisipasi Publik,” *Jurnal Ilmiah Manajemen Bisnis dan Inovasi Universitas Sam Ratulangi (JMBI UNSRAT)*, vol. 11, no. 1, pp. 1003–1011, Apr. 2024, doi: <https://doi.org/10.35794/jmbi.v11i1.55526>.
- [6] H. B. Mardikantoro, M. B. Siroj, E. S. Utami, and E. Kurniati, “Investigating Indonesian language varieties in social media interactions: Implications to teaching practices,” *Indonesian Journal of Applied Linguistics*, vol. 13, no. 2, pp. 306–316, Sep. 2023, doi: <https://doi.org/10.17509/ijal.v13i2.63069>.
- [7] J. R. Jim, M. A. R. Talukder, P. Malakar, M. M. Kabir, K. Nur, and M. F. Mridha, “Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review,” *Natural Language Processing Journal*, vol. 6, pp. 1–30, Mar. 2024, doi: <https://doi.org/10.1016/j.nlp.2024.100059>.
- [8] G. Mario Conroy Paridy Man, A. Aristo Jansen Sinlae, and E. Ngaga, “Analisis Sentimen di Media Sosial X tentang IKN dengan Naïve Bayes,” *JIP (Jurnal Informatika Polinema)*, vol. 11, no. 4, pp. 417–425, Aug. 2025, doi: <https://doi.org/10.33795/jip.v11i4.7246>.
- [9] M. B. M. Amin *et al.*, “Deteksi Spam Berbahasa Indonesia Berbasis Teks Menggunakan Model Bert,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 6, pp. 1291–1301, Dec. 2024, doi: <https://doi.org/10.25126/jtiik.2024118121>.
- [10] Muh. Hasbi Ashiddiq and A. Witanti, “Analisis Sentimen tentang Penundaan Pengangkatan CPNS 2025 pada Platform X Menggunakan Metode IndoBERT,” *Jurnal teknika*, vol. 17, no. 2, pp. 97–108, Sep. 2025, doi: [10.30736/jt.v17i2.1448](https://doi.org/10.30736/jt.v17i2.1448).
- [11] W. Widyandana, Maskur, and A. Fauzi, “Machine Learning and Transformer-based Model for Sentiment Analysis of Indonesian E-Commerce Reviews,” *The Indonesian Journal of Computer Science*, vol. 14, no. 4, pp. 6262–6271, Aug. 2025, doi: <https://doi.org/10.33022/ijcs.v14i4.4980>.
- [12] H. Jayadianti, W. Kaswidjanti, A. T. Utomo, S. Saifullah, F. A. Dwiyanto, and R. Drezewski, “Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN,” *ILKOM Jurnal Ilmiah*, vol. 14, no. 3, pp. 348–354, Dec. 2022, doi: <https://doi.org/10.33096/ilkom.v14i3.1505.348-354>.
- [13] J. C. Setiawan, K. M. Lhaksana, and Bunyamin, “Sentiment Analysis of Indonesian TikTok Review Using LSTM and IndoBERTweet Algorithm,” *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 3, pp. 774–780, Sep. 2023, doi: <https://doi.org/10.29100/jupi.v8i3.3911>.
- [14] A. Annur Rohman, G. Alfa Trisnapradika, and K. Kunci, “Perbandingan Algoritma NBC, SVM, Logistic Regression untuk Analisis Sentimen Terhadap Wacana KaburAjaDulu di Media Sosial X,” *Building of Informatics, Technology and Science (BITS)*, vol. 7, no. 1, pp. 169–178, Jun. 2025, doi: <https://doi.org/10.47065/bits.v7i1.7261>.
- [15] R. Herdian Saputra and R. Randy Suryono, “Perbandingan Algoritma SVM, Random Forest, dan Naive Bayes Terhadap Kasus Scam di Media Sosial Twitter,” *Technology and Science (BITS)*, vol. 7, no. 2, pp. 907–919, Sep. 2025, doi: <https://doi.org/10.47065/bits.v7i2.7236>.
- [16] U. Khairani, V. Mutiawani, and H. Ahmadian, “Pengaruh Tahapan Preprocessing Terhadap Model Indobert Dan Indobertweet Untuk Mendeteksi Emosi Pada Komentar Akun Berita Instagram,” *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 11, no. 4, pp. 887–894, Aug. 2024, doi: <https://doi.org/10.25126/jtiik.1148315>.
- [17] M. Fernanda Naufal Fathoni, E. Yulia Puspaningrum, and A. Nugroho Sihananto, “Perbandingan Performa Labeling Lexicon InSet dan VADER pada Analisa Sentimen Rohingya di Aplikasi X dengan SVM,” *Jurnal Informatika dan Sains Teknologi*, vol. 1, no. 3, pp. 62–76, Jul. 2024, doi: <https://doi.org/10.62951/modem.v2i3.112>.
- [18] M. Rafly Gusmansyah, H. Hendrawan, Rahmadden, and Rohid, “Peningkatan Kinerja Analisis Sentimen pada Ulasan Aplikasi Identitas Kependudukan Digital (IKD) di Indonesia Menggunakan Algoritma Support Vector Machine (SVM) dan Smote,” *Jurnal Instek*, vol. 10, no. 1, pp. 185–196, May 2025, doi: <https://doi.org/10.24252/instek.v10i1.55292>.
- [19] I. Sari B, Muh. Rafli Rasyid, F. Wajidi, and K. Kunci, “Implementasi Support Vector Machine Untuk Analisis Sentimen Robot Polisi Humanoid,” *Jurnal Sistem Informasi dan Teknik Komputer (SIMTEK)*, vol. 10, no. 2, pp. 329–335, Oct. 2025, doi: <https://doi.org/10.51876/simtek.v10i2.1623>.



- [20] A. Mu'amar Wahid, K. Adi Nugroho, T. Safitri, Darmono, and F. Setyo Utomo, "Optimasi Logistic Regression dan Random Forest untuk Deteksi Berita Hoax Berbasis TF-IDF," *Jurnal Pendidikan dan Teknologi Indonesia (JPTI)*, vol. 4, no. 8, pp. 381–392, Jan. 2025, doi: 10.52436/1.jpti.602.
- [21] S. Rihastuti and A. Rosyidi, "Analisis Sentimen Pengguna Tiktok Tentang Progres Pembangunan IKN Dengan Metode Random Forest," *Journal of Computer Science and Technology JCS-TECH*, vol. 5, no. 1, pp. 19–23, May 2025, doi: <https://doi.org/10.54840/jcstech.v5i1.345>.
- [22] S. Syakira Rambe, Asriyanik, and Prajoko, "Penerapan Model Convolutional Neural Network (CNN) Berbasis MobilenetV2 Untuk Klasifikasi Tingkat Kesegaran Ikan Nila," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 3, pp. 2234–2246, Jul. 2025, doi: <https://doi.org/10.23960/jitet.v13i3.6744>.