

# Sentiment Classification of Indonesian-Language Roblox Reviews Using IndoBERT with SMOTE Optimization

Ferdiansyah<sup>1\*</sup>, Ryan Randy Suryono<sup>2\*</sup>

<sup>\*</sup>Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia

<sup>\*\*</sup>Magister Ilmu Komputer, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia

[ferdiansyahh@teknokrat.ac.id](mailto:ferdiansyahh@teknokrat.ac.id)<sup>1</sup>, [ryan@teknokrat.ac.id](mailto:ryan@teknokrat.ac.id)<sup>2</sup>

## Article Info

### Article history:

Received 2025-07-08

Revised 2025-07-25

Accepted 2025-08-09

### Keyword:

*Sentiment Classification,  
IndoBERT,  
SMOTE,  
Game Reviews,  
Roblox,  
Analysis,  
Natural Language Processing.*

## ABSTRACT

Roblox is a community-based gaming platform that is extremely popular among users of various age groups. Millions of user reviews available on the platform contain valuable information regarding user satisfaction, expectations, and criticisms of the gameplay experience. To extract insights from these reviews, a reliable natural language processing (NLP) approach tailored to the local language context is essential. This study aims to classify sentiments in Indonesian-language user reviews of Roblox into three categories: positive, negative, and neutral. The model used is IndoBERT, a transformer-based model specifically trained to understand the structure and vocabulary of the Indonesian language. One of the main challenges in this study is the imbalance in the number of data points across sentiment classes. To address this, the SMOTE (Synthetic Minority Over-sampling Technique) method is applied to strengthen the representation of minority classes. The dataset consists of thousands of reviews that have been manually labeled by annotators. Model performance is evaluated using accuracy, precision, recall, and F1-score metrics. The results show that the combination of IndoBERT and SMOTE provides significant improvements compared to the baseline approach without oversampling. This research contributes to the development of automated sentiment analysis systems in the Indonesian language, which can be applied across various digital platforms. The implementation of this model can assist game developers and product analysts in efficiently understanding user opinions, thereby driving improvements in service quality and user satisfaction in the future.



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

## I. PENDAHULUAN

Ketika datang video game di seluruh dunia. Indonesia berada di peringkat ketiga, dengan sekitar 94,5% pengguna internet berusia 16-64 tahun yang bermain video game pada Januari 2022. Laporan We Are Social ini menunjukkan besarnya minat masyarakat Indonesia terhadap game. Bahkan, Filipina dan Vietnam juga menduduki peringkat atas dengan proporsi pengguna internet yang bermain video game mencapai 96,5% dan 94,7% [1]. Industri game di Indonesia mengalami pertumbuhan signifikan, dengan jumlah pemain game yang cukup besar. Berdasarkan data, terdapat sekitar 53,4 juta pemain game PC dan 133,8 juta pemain game mobile di Indonesia. Hal ini menunjukkan

bahwa game menjadi salah satu bentuk hiburan populer di kalangan masyarakat Indonesia [2]. Roblox adalah salah satu game mobile yang populer. Pengguna platform game online Roblox dapat membuat dan memainkan berbagai game virtual. Platform ini digemari anak-anak dan remaja karena mudah digunakan dan menawarkan ruang untuk kreativitas. Dengan Roblox, pengguna bisa berkreasi sesuai keinginan mereka [3].

Keberadaan analisis sentimen dalam industri game semakin mendesak untuk diteliti karena beberapa faktor penting. Saat ini, konsumen semakin bergantung pada ulasan dan pendapat pengguna lain sebelum memutuskan untuk membeli atau memainkan sebuah game. Dengan demikian, analisis sentimen menjadi alat krusial untuk

memantau dan memahami opini publik terhadap suatu *game* secara langsung. Ulasan yang bernada positif maupun negatif dapat memberikan dampak signifikan terhadap performa *game* di pasar. Oleh karena itu, para pengembang *game* memanfaatkan analisis ini untuk mengidentifikasi aspek-aspek yang perlu diperbaiki atau dikembangkan, seperti kualitas grafik, mekanisme permainan, atau fitur tertentu yang sering menjadi sorotan dalam komentar pengguna [4]. Selain itu, analisis sentimen adalah komponen penting dari melakukan perbandingan dengan produk pesaing, sehingga perusahaan dapat mengevaluasi posisi produk mereka relatif terhadap kompetitor. Manfaat ini tidak hanya mendukung proses pengembangan produk, tetapi juga membantu meningkatkan pengalaman pengguna serta memperkuat loyalitas merek, dengan cara merespons kritik dan preferensi pengguna secara cepat dan proaktif [5]. Mengidentifikasi dan memperbaiki fitur-fitur yang kurang diminati oleh pemain, serta mengembangkan fitur-fitur yang disukai, merupakan hal yang penting. Ulasan dari pengguna memiliki peran besar, karena calon pemain umumnya mempertimbangkan ulasan, rating, dan peringkat *game* sebelum memutuskan untuk mengunduhnya [6].

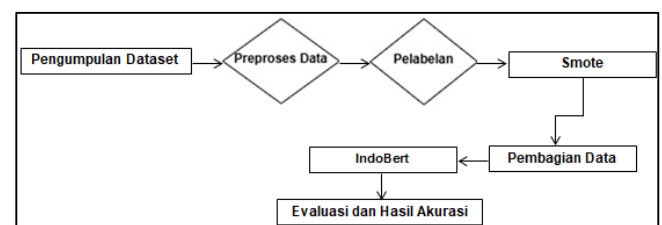
Analisis sentimen sangat bermanfaat bagi developer *game* yang ingin memahami perspektif pelanggan terhadap pengalaman bermain. Menelusuri ulasan dari berbagai platform merupakan cara yang efektif untuk menentukan langkah pengembangan dan penyempurnaan *game* di masa depan. Selain itu, analisis sentimen juga membantu dalam mengukur sejauh mana popularitas suatu produk [7]. Dalam penelitian sebelumnya oleh Aditian et al. [2024], digunakan dua metode yaitu algoritma *Naïve Bayes* dan *Support Vector Machine (SVM)* untuk menganalisis sentimen ulasan pada aplikasi *Roblox*. Hasil analisis menunjukkan bahwa sebagian besar ulasan memiliki nada positif sebesar 65,06%, sementara sentimen negatif berjumlah 34,94%. Di antara kedua algoritma, *SVM* terbukti menjadi metode yang paling akurat dalam analisis sentimen, dengan tingkat akurasi tertinggi mencapai 90% pada rasio data 90:10. Untuk kategori sentimen negatif, *SVM* menghasilkan *precision* sebesar 88%, *recall* 85%, dan *f1-score* 86%. Sementara itu, pada sentimen positif, *precision*, *recall*, dan *f1-score* masing-masing mencapai 91%, 93%, dan 92%. Sebaliknya, algoritma *Naïve Bayes* memperoleh akurasi tertinggi sebesar 72,4% pada rasio data yang sama. Pada sentimen negatif, *precision*, *recall*, dan *f1-score* yang dihasilkan adalah 88%, 34%, dan 49%. Sedangkan untuk sentimen positif, nilai *precision* mencapai 70%, *recall* 97%, dan *f1-score* 81% [8].

Dalam penelitian lain oleh Braja dan Kodar (2023), digunakan dua model *pre-trained*, yaitu BERT base Multilingual dan IndoBERT Base, dengan dua pendekatan pelabelan berbeda berdasarkan skor dan TextBlob. Hasil penelitian menunjukkan bahwa model IndoBERT yang telah di *fine-tune* memberikan performa terbaik dalam pelabelan data menggunakan pendekatan TextBlob. Model

ini berhasil mencapai presisi maksimum dari 94% dengan konfigurasi *learning rate*  $2e-5$ , *batch size* 32, jumlah *epoch* sebanyak 5, dan waktu pelatihan selama 12 menit [9]. Dengan demikian, penelitian ini menegaskan bahwa model IndoBERT mampu memberikan kinerja yang sangat baik dalam analisis sentimen, konsisten dengan kesimpulan studi sebelumnya bahwa menunjukkan keunggulan model BERT dibandingkan metode tradisional dan beberapa model *deep learning* lainnya. Peneliti melakukan penyesuaian terhadap *hyperparameter* untuk memperoleh hasil yang optimal. *Hyperparameter* yang digunakan dalam penelitian ini meliputi *learning rate* sebesar 0,000003, dropout 0,1, *batch size* 16, dan jumlah *epoch* sebanyak 10. Dataset yang digunakan merupakan ulasan pengguna aplikasi *game Roblox* yang dikumpulkan dari *Google Play Store* berdasarkan ulasan terbaru. Karena dataset tersebut tidak memiliki label, peneliti melakukan pelabelan berdasarkan skor atau rating dan mengelompokkan data ke dalam tiga kategori sentimen: positif, netral, dan negatif. Model IndoBERT dipilih karena penelitian terdahulu menunjukkan akurasi yang tinggi, serta kemampuannya dalam menangani data ulasan berbahasa Indonesia dengan baik, mengingat model ini dirancang khusus untuk memproses teks dalam bahasa Indonesia [10].

## II. METODE

Prosedur yang digunakan dalam studi ini disusun mengamati diagram alur yang menggambarkan setiap tahapan yang dilakukan selama proses penelitian, sebagaimana ditampilkan pada Gambar 1 berikut



. Gambar 1. Tahapan Penelitian

### 2.1. Pengumpulan Data

Pada penelitian ini, informasi yang diperoleh dari platform *Google Play Store* untuk *game Roblox* dengan ulasan berbahasa Indonesia. Pengambilan data dilakukan melalui proses *scraping* menggunakan package yang tersedia di *Google Colab*, yang memanfaatkan API dalam bahasa pemrograman *Python*. Data yang diperoleh kemudian disimpan dalam format *CSV* dan *Excel*. Total Dataset yang digunakan dalam ada 10.000 entri dalam studi ini.

### 2.2. Tahap Preprocessing

Pra-pemrosesan adalah saat data mentah diubah menjadi data yang dapat digunakan untuk penambahan teks. Langkah ini sangat penting dalam penambahan teks

karena menyiapkan data yang berantakan dan tidak terorganisir untuk disusun [11]. Proses ini membersihkan data yang belum dalam format standar. Fase ini mencakup pembersihan, penghilangan huruf besar, tokenisasi, penghapusan kata-kata umum (stopwords), dan stemming. Tujuannya adalah untuk membuat dataset yang bersih dan terorganisir yang siap untuk analisis lebih lanjut [12]. Pembersihan melibatkan penghapusan angka, spasi ekstra, tanda baca, dan simbol yang tidak relevan dengan penelitian. Penghilangan huruf besar adalah metode yang mengubah semua kata dalam dataset menjadi huruf kecil. Ini menghilangkan perbedaan antara huruf besar dan huruf kecil, yang dapat membuat analisis lebih konsisten. Tokenisasi adalah proses memecah teks menjadi kata-kata individu dengan menghilangkan tanda baca, sehingga menciptakan kata-kata terpisah yang membentuk kalimat atau frasa [13]. Penghapusan stopwords adalah proses mengeluarkan kata-kata yang tidak memiliki arti banyak dan tidak diperlukan untuk pra-proses [14]. Kata-kata ini sering kali termasuk kata-kata seperti "yang," "di," "dan," "pada," dan lainnya seperti itu. Stemming adalah proses menghapus semua prefiks dan sufiks dari kata, sehingga kata-kata diubah menjadi bentuk dasar atau kata akar dalam sebuah kalimat.

#### A. Cleaning Data

Pembersihan data adalah proses menghapus atau memperbaiki data yang tidak berguna, tidak penting, atau dapat menurunkan kualitas data [15]. Ini mencakup hal-hal seperti URL, nama pengguna, retweet, kode HTML, dan tagar.

Tabel I. Cleaning Data

Sebelum Cleaning	Setelah Cleaning
Game bagus banget □□, udah Gitu ngilangin gabut. Suka banget main game Roblox □, karena seruuu. Permainan dan map <sup>2</sup> nya sangat bagus! tetapi, gamenya sangat seruuu!!	Game bagus banget udah Gitu ngilangin gabut. Suka banget main game Roblox karena seruuu. Permainan dan map nya sangat bagus tetapi

#### B. Case Folding

Untuk memastikan bahwa huruf kapital tidak menyebabkan masalah, proses pengubahan huruf besar menjadi kecil mengubah semua teks dalam dokumen menjadi huruf kecil [16]. Ini karena proses tersebut mengakui bahwa huruf kapital dan huruf kecil adalah berbeda.

Tabel II. Case Folding

Sebelum Casefolding	Setelah Casefolding
<b>Game</b> bagus banget udah <b>Gitu</b> ngilangin gabut. <b>Suka</b> banget main game <b>Roblox</b> karena seruuu. <b>Permainan</b> dan map nya sangat bagus.	<b>game</b> bagus banget udah <b>gitu</b> ngilangin gabut. <b>suka</b> banget main game <b>roblox</b> karena seruuu. <b>permainan</b> dan map nya sangat bagus.

#### C. Tokenize

Tokenisasi data adalah cara untuk mengubah teks atau data menjadi serangkaian token [17]. Token adalah bagian terkecil yang bermakna saat memproses teks. Ini biasanya berarti memecah teks menjadi kata-kata tunggal atau bagian bermakna lainnya, seperti tanda baca atau angka.

Tabel III. Tokenize

Sebelum Tokenizing	Sebelum Tokenizing
game bagus banget udah gitu ngilangin gabut. suka banget main game roblox karena seruuu. permainan dan map ya sangat bagus tetapi. gamenya sangat seruuu map nya banyak dan menarik. game nya bagus sih tapi ada bug bug	[game, bagus, banget, sudah, begitu, menghilangkan, an, gabut]. [suka, banget, main, game, roblox, karena, seruuu]. [permainan, dan, map, ya, sangat, bagus]. [game, ya, bagus, sih, tapi, ada, bug, bug, ya].

#### D. Stopword Removal

Penghapusan kata umum digunakan untuk menghilangkan kata-kata umum yang tidak menambah banyak makna pada teks. Ini dilakukan dengan menggunakan algoritma daftar kata untuk menghapus kata-kata yang dianggap tidak penting [18]. Ini juga dapat dilakukan dengan menggunakan daftar kata untuk mempertahankan kata-kata yang memiliki makna nyata.

Tabel IV. Stopword Removal

Sebelum Remove Stopword	Sesudah Remove Stopword
[game, bagus, banget, sudah, begitu, menghilangkan, an, gabut]. [suka, banget, main, game, roblox, karena, seruuu]. [permainan, dan, map, ya, sangat, bagus]. [game, ya, bagus, sih, tapi, ada, bug, bug, ya].	[game, bagus, banget, menghilangkan, gabut]. [suka, banget, main, game, roblox, seru]. [permainan, map, ya, bagus, semenjak, roblox]. [game, ya, bagus, sih, bug, bug, ya, gapapa].

### E. Stemming

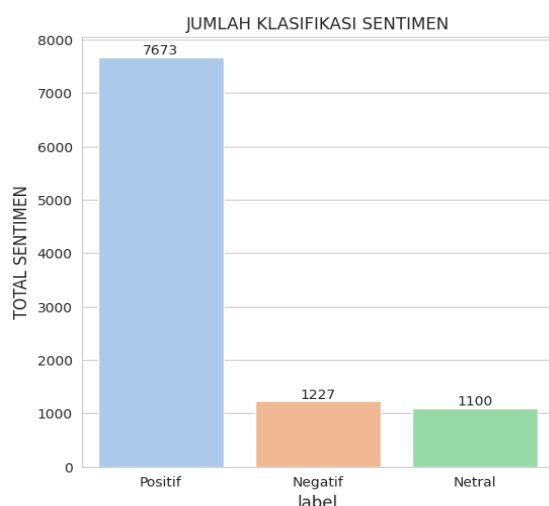
Selama proses pengakaran, setiap kata diubah menjadi bentuk dasarnya dengan menghilangkan awalan dan akhiran [19]. Proses pengakaran ini dilakukan menggunakan pustaka sastrawi, yang menangani fungsi pengakaran.

Tabel V. Stemming

Sebelum Steming	Setelah Steming
<b>permainan,</b> map, ya, bagus,semenjak,roblox.gam e,bagus,banget, <b>menghilangkan,</b> gabut,suka banget main game roblox karena seru. gamenya, seru, map, ya, <b>menarik, menjelajah,</b>	<b>main</b> map ya bagus semenjak roblox update map. game bagus banget <b>hilang</b> gabutkasih bintang. gamenya seru map ya <b>tarik jelajah</b> tualang main.

### 2.3 Labeling

Pelabelan merupakan prosedur pelabelan data untuk mengelompokkannya ke dalam kategori positif, negatif, atau netral. Dalam penelitian ini, pelabelan dilakukan menggunakan kamus lexicon, yang menghasilkan distribusi sentimen sebanyak 7.672 untuk kelas positif, 1227 untuk kelas negatif, dan 1.100 untuk kelas netral.



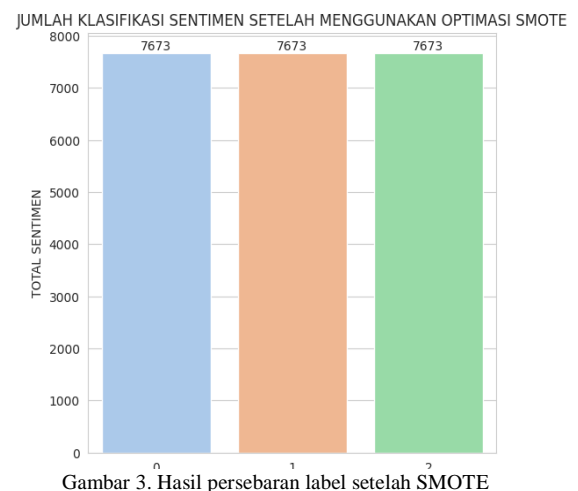
Gambar 2. Hasil Persebaran Label sebelum SMOTE

### 2.4 SMOTE

SMOTE (Synthetic Minority Oversampling Technique) merupakan teknik yang digunakan untuk menangani ketidakseimbangan kelas dalam suatu dataset [20]. Ketika salah satu kelas memiliki jumlah data yang jauh lebih sedikit dibandingkan kelas lainnya, model pembelajaran mesin berpotensi menjadi tidak seimbang dan cenderung bias terhadap kelas mayoritas. Untuk mengatasi hal ini, SMOTE menciptakan sampel sintetis bagi kelas minoritas melalui proses interpolasi antar data yang sudah ada.

Dengan begitu, proporsi kelas minoritas meningkat, memungkinkan model memahami pola dari kelas tersebut secara lebih baik dan meningkatkan akurasi prediksi. Dalam penelitian ini, dataset awal terdiri dari total 1000 data dengan komposisi sebagai berikut: 1227 komentar negatif, 7673 komentar positif, dan 1100 komentar netral. Pada kondisi ini, kelas minoritas adalah komentar positif. Oleh karena itu, SMOTE diterapkan dengan cara memilih sampel acak dari kelas minoritas, kemudian mencari tetangga terdekatnya, dan membuat data baru melalui interpolasi antara kedua sampel tersebut. Proses ini diulang hingga jumlah sampel sintetis yang diinginkan tercapai.

Setelah penerapan SMOTE, jumlah total data menjadi 4962, yang terdiri dari 1227 komentar negatif, 7673 komentar positif asli, 1100 komentar netral, serta 13.019 sampel sintetis tambahan. Dengan dataset yang telah diseimbangkan ini, model pembelajaran mesin dapat dilatih secara lebih efektif guna menghasilkan prediksi yang lebih tepat dan tidak bias.



Gambar 3. Hasil persebaran label setelah SMOTE

### 2.5 Pembagian Data

Dalam proses pembuatan model, membagi data menjadi dua kategori, yaitu data pelatihan dan data uji, sangatlah penting. Data pelatihan digunakan untuk melatih model agar dapat membedakan antara kategori positif dan negatif, sedangkan data uji digunakan untuk mengevaluasi kinerja model tersebut. Dalam penelitian ini, pembagian data dilakukan dengan alokasi 80% untuk pelatihan dan 20% untuk pengujian, yang diharapkan dapat menghasilkan hasil yang lebih optimal.

### 2.6 Klasifikasi Model

Indo Metode Bidirectional Encoder Representations from Transformers (IndoBERT) adalah sebuah model bahasa yang dirancang untuk memahami dan memproses teks

secara mendalam [21]. BERT menggunakan arsitektur Transformer, yang terdiri dari lapisan-lapisan pengkode (encoder layers) yang saling terhubung, untuk mempelajari representasi teks yang kaya dan kontekstual. IndoBERT dilatih pada korpus teks yang sangat besar dan beragam, sehingga memperoleh pemahaman yang lebih baik tentang konteks kata-kata dalam sebuah kalimat [22]. Salah satu keunggulan utama IndoBERT adalah kemampuannya dalam memahami kata-kata yang ambigu atau frasa yang bergantung pada konteks tertentu.

### 2.7 Evaluasi dan Hasil

Akurasi Selanjutnya, model diuji menggunakan cross-validation K-fold ini melibatkan pemeriksaan kinerja model menggunakan pengukuran kesalahan untuk melihat seberapa akurat itu. Kinerja model diperiksa menggunakan matriks kebingungan untuk mendapatkan pengukuran seperti akurasi, presisi, recall, dan F1-Score. Matriks kebingungan membantu menganalisis kinerja model setiap kali. Matriks kebingungan memiliki empat bagian utama: True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) [23]. Bagian-bagian ini digunakan untuk menghitung nilai-nilai dalam matriks kebingungan. Nilai-nilai ini menunjukkan seberapa baik model mengklasifikasikan data. Menggunakan matriks ini, akurasi, presisi, recall, dan F1-score dapat dihitung dengan rumus-rumus ini:

Tabel VI. Kriteria Confusion Matrix

Actual Class	Predicted Class	
	Positif	Negatif
Positif	True Positif (TP)	False Negatif (FN)
Negatif	False Positif (FP)	True Negatif (TN)

Keterangan:

a. True Positive (TP) mengacu pada data yang memiliki kondisi aktual positif dan ramalan yang akurat.

b. True Negative (TN) merujuk pada data suram yang diantisipasi secara akurat.

c. False Positive (FP) terjadi ketika data yang seharusnya negatif diprediksi sebagai positif.

d. False Negative (FN) terjadi ketika data yang seharusnya positif diprediksi sebagai negatif.

Terdapat beberapa metrik evaluasi pada confusion matrix yaitu :

a. Accuracy

$$\text{accuracy} = \frac{TN + TP}{TN + FN + TP + FP} \times 100\%$$

b. Precision

$$\text{precision} = \frac{TP}{TP + FP} \times 100\%$$

c. Recall

$$\text{recall} = \frac{TP}{TP + FN} \times 100\%$$

d. F1 –score

$$\text{f1 – score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \times 100\%$$

## III. HASIL DAN PEMBAHASAN

### 3.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari ulasan pengguna game Roblox di Google Play Store. Pengambilan data dilakukan secara otomatis menggunakan teknik web scraping dengan bantuan library di Google Colab. Proses ini memungkinkan pengumpulan ulasan secara efisien dari platform tersebut. Rentang waktu pengambilan data mencakup periode dari Juni 2024 hingga Maret 2025, dengan total jumlah data yang berhasil dikumpulkan sebanyak 25.000 ulasan.

### 3.2 Tahap Pengujian

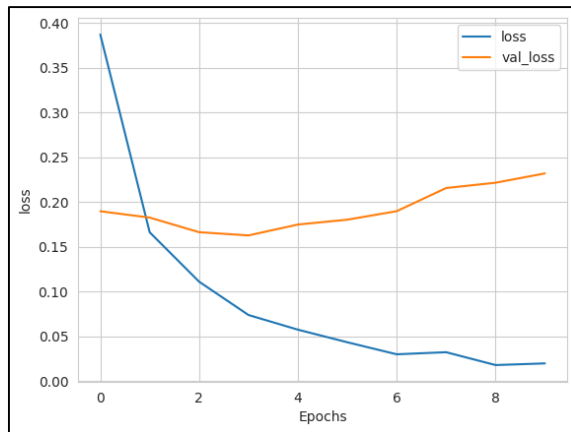
Indo Bidirectional Encoder Representations dari Transformer (IndoBERT) adalah model bahasa yang dapat memahami dan bekerja dengan teks secara mendetail [24]. IndoBERT menggunakan struktur Transformer, yang memiliki lapisan encoder yang terhubung. Ini membantunya belajar cara merepresentasikan teks dengan cara yang mendetail dan sadar konteks. IndoBERT dilatih menggunakan kumpulan besar teks yang berbeda, yang membantunya lebih memahami bagaimana kata-kata digunakan dalam kalimat [25]. Salah satu manfaat utama dari BERT adalah kemampuannya untuk memahami kata atau ungkapan yang tidak jelas karena tergantung pada situasi. Data pelatihan digunakan selama 10 epoch performa dapat dilihat pada Gambar 4.

\Epoch	No.	Train Accuracy	Train loss	Val Accuracy	Val loss
	1	0.855000	0.424585	0.932000	0.209238
	2	0.946500	0.176136	0.953000	0.155669
	3	0.964000	0.111737	0.950000	0.173398
	4	0.973375	0.086420	0.944000	0.201143
	5	0.979500	0.061957	0.946000	0.216837
	6	0.985375	0.048473	0.945000	0.216376
	7	0.987875	0.040102	0.947000	0.225210
	8	0.989750	0.033613	0.948000	0.223635
	9	0.992000	0.024410	0.948000	0.296477
	10	0.994125	0.019088	0.944000	0.305430

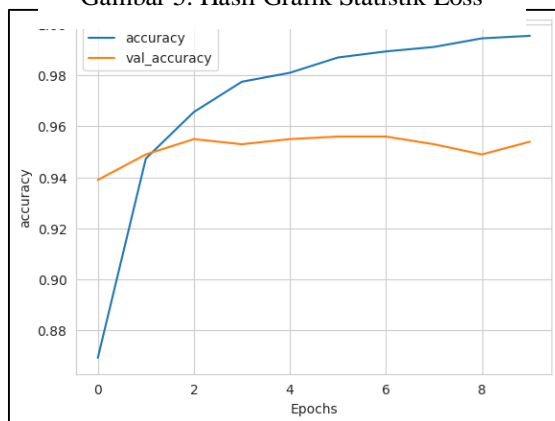
Gambar 4. Performa model BERT pada setiap Epoch

### 3.3 Hasil Fine Tuning Data Ulasan Roblox

Untuk menilai konsistensi kinerja model yang dikembangkan, evaluasi dilakukan dengan mengamati selisih antara nilai ketepatan dan kehilangan data dalam validasi. Pelatihan model dilakukan selama 10 epoch dan dilengkapi dengan penerapan fungsi early stopping guna mencegah terjadinya overfitting.



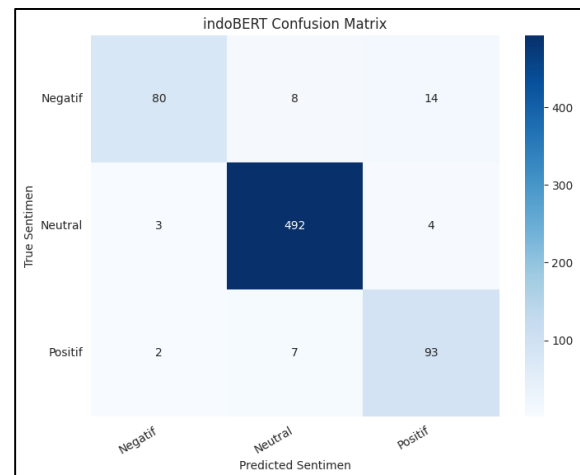
Gambar 5. Hasil Grafik Statistik Loss



Gambar 6. Hasil Grafik Statistik Accuracy

Didapatkan hasil berdasarkan Gambar 4 dan Gambar 5 memberikan ringkasan keluaran model BERT untuk setiap epoch dataset ulasan *roblox*. Sebuah pengamatan lebih dekat mengungkapkan bahwa baik akurasi pelatihan maupun akurasi validasi cenderung untuk menstabilkan diri dalam suatu rentang 85-99% kemampuan model untuk mengidentifikasi pola dalam data konsisten sepanjang epoch. Meskipun akurasi validasi tetap konstan sepanjang perjalanan epoch, ada indikasi overfitting, seperti yang terlihat dari meningkatnya kehilangan validasi. Namun, overfitting belum sangat mengganggu, seperti yang dibuktikan oleh perbedaan yang cukup terkontrol antara kehilangan pelatihan dan kehilangan validasi. Model telah membaca data pelatihan dengan sukses, seperti yang dibuktikan oleh akurasi pelatihan yang mencapai 99% dengan kehilangan pelatihan yang sangat rendah pada epoch

ke-10. Namun, akurasi validasi tetap di 94% seiring dengan



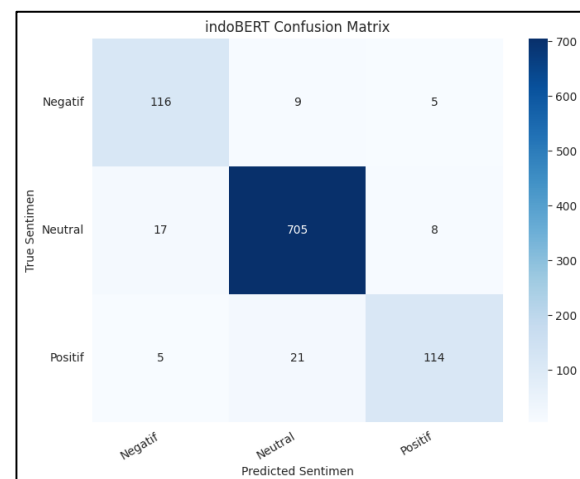
Gambar 7. Hasil Confusion Matrix indoBert sebelum SMOTE

meningkatnya kehilangan validasi, yang menunjukkan bahwa model tidak mampu melakukan generalisasi. dengan baik pada data validasi. Oleh karena itu, pemantauan terus-menerus terhadap tren *val loss* dan *val accuracy* menjadi penting dalam pengembangan model, serta mempertimbangkan strategi untuk mengatasi *overfitting* yang mungkin terjadi

### 3.4 Confusion Matrix

Tahap confusion matrix digunakan sebagai metode evaluasi untuk mengukur kinerja model yang telah diterapkan, dengan memberikan informasi tentang jumlah prediksi yang benar dan salah dari masing-masing kelas[26]. Hasil confusion matrix dapat kita lihat pada Gambar 7 dan Gambar 8.

Hasil evaluasi menggunakan confusion matrix



Gambar 8. Hasil Confusion Matrix indoBert setelah SMOTE

menunjukkan bahwa dari 130 ulasan berlabel negatif, model berhasil mengklasifikasikan 116 ulasan dengan benar (True Negatif). Namun, terdapat 9 ulasan yang keliru diklasifikasikan sebagai netral (False Neutral) dan 5 sebagai positif (False Positif). Sementara itu, dari 729 ulasan netral, model mampu mengklasifikasikan 705 ulasan dengan tepat (True Neutral), tetapi masih terjadi kesalahan pada 17 ulasan yang diklasifikasikan sebagai Negatif (False Negatif). Untuk ulasan berlabel Positif sebanyak 140, model berhasil mengidentifikasi 144 ulasan dengan benar (True Positif), namun masih terdapat 5 ulasan yang diklasifikasikan sebagai Negatif (False Negatif) dan 21 ulasan sebagai netral (False Neutral).

Evaluasi ini mencerminkan sejauh mana model mampu membedakan kelas sentimen dengan tepat serta menunjukkan area yang masih perlu diperbaiki.

	precision	recall	f1-score	support
0	0.84	0.89	0.87	130
1	0.96	0.97	0.96	730
2	0.90	0.81	0.85	140
accuracy			0.94	1000
macro avg	0.90	0.89	0.89	1000
weighted avg	0.94	0.94	0.93	1000

Gambar 9. Hasil Performa Model indoBert sebelum SMOTE

Berdasarkan Gambar 9 didapatkan hasil performa model menunjukkan performa yang baik dalam mengklasifikasikan sentimen positif, netral, dan negatif. Berdasarkan hasil evaluasi, model memiliki precision sebesar 90% untuk sentimen positif, 96% untuk netral, dan 84% untuk negatif, yang menunjukkan tingkat ketepatan prediksi yang tinggi. Recall model mencapai 81% untuk positif, 97% untuk netral, dan 89% untuk negatif, menandakan model mampu mengidentifikasi mayoritas ulasan dengan benar di setiap kategori. F1-score masing masing kategori adalah 85% untuk positif, 96% untuk netral, dan 87% untuk negatif, yang mencerminkan

	precision	recall	f1-score	support
0	0.94	0.78	0.86	102
1	0.97	0.99	0.98	499
2	0.84	0.91	0.87	102
accuracy			0.95	703
macro avg	0.92	0.89	0.90	703
weighted avg	0.95	0.95	0.95	703

Gambar 10. Hasil Performa Model indoBert setelah SMOTE

keseimbangan antara precision dan recall.

Akurasi keseluruhan model mencapai 94%, dengan macro average dan weighted average precision, recall, serta f1-score sebesar 93%, menunjukkan konsistensi performa di semua kategori. Jumlah data yang hampir merata untuk setiap kategori memastikan distribusi data yang seimbang, sehingga model dapat menangkap pola dengan baik tanpa bias yang signifikan.

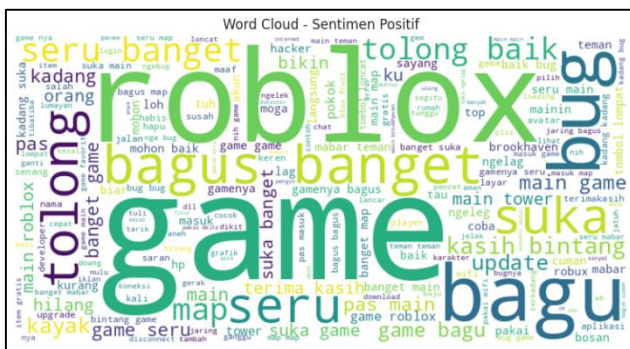
Berdasarkan Gambar 10 didapatkan hasil performa model menunjukkan kinerja yang baik dalam mengklasifikasikan sentimen positif, netral, dan negatif. Berdasarkan hasil evaluasi, model tersebut memiliki precision sebesar 84% untuk sentimen positif, 97% untuk netral, dan 94% untuk negatif, yang menunjukkan tingkat ketepatan prediksi yang tinggi. Recall model mencapai 91% untuk positif, 99% untuk netral, dan 78% untuk negatif, menandakan model mampu mengidentifikasi mayoritas ulasan dengan benar di setiap kategori. F1-score masing masing kategori adalah 87% untuk positif, 98% untuk netral, dan 86% untuk negatif, yang mencerminkan keseimbangan antara precision dan recall.

Akurasi keseluruhan model mencapai 95%, dengan macro average dan weighted average precision, recall, serta f1-score sebesar 95%, menunjukkan konsistensi performa di semua kategori. Jumlah data yang hampir merata untuk setiap kategori memastikan distribusi data yang seimbang, sehingga model dapat menangkap pola dengan baik tanpa bias yang signifikan.

### 3.5 Visualisasi Wordcloud

Wordcloud merupakan representasi visual yang menunjukkan istilah yang muncul paling sering dalam sebuah teks. Dalam konteks analisis sentimen ulasan game Roblox, wordcloud digunakan untuk memvisualisasikan kata-kata yang dominan berdasarkan kategori sentimen, seperti positif, netral, dan negatif. Proses ini mencakup tahapan pengolahan teks, pemilihan kata kunci, hingga penyajian visual. Selain berfungsi untuk mengevaluasi kinerja model sentimen, wordcloud juga memberikan gambaran yang lebih mendalam mengenai kata-kata dan isu utama yang sering diungkapkan oleh pengguna, menjadikannya alat visualisasi yang efektif dalam analisis sentimen teks.

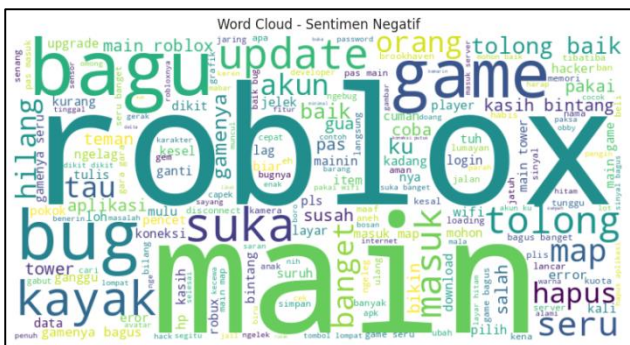




Gambar 11. Wordcloud Ulasan Positif Game Roblox

Berdasarkan Gambar 11, terlihat wordcloud yang merepresentasikan ulasan dengan label sentimen positif. Kata-kata yang muncul secara dominan mencerminkan tingkat kepuasan pengguna terhadap game ini. Beberapa kata yang paling sering ditemukan antara lain: bagus, roblox, game, seru, tolong, keren, lebih, map, sangat, grafik, mantap, banget, makin, suka, event, bintang, sekarang, dan mantap.

Analisis ini menunjukkan bahwa ulasan positif terhadap game Roblox di Google Play Store mencerminkan apresiasi pengguna terhadap berbagai aspek, terutama gameplay, pembaruan konten, keamanan, kualitas grafis, serta keberagaman event yang dihadirkan. Kata-kata yang sering muncul menandakan tingkat kepuasan yang tinggi dan pengalaman bermain yang menyenangkan. Temuan ini memperkuat pandangan bahwa game Roblox telah berhasil memenuhi ekspektasi mayoritas penggunanya serta memberikan pengalaman bermain yang positif.

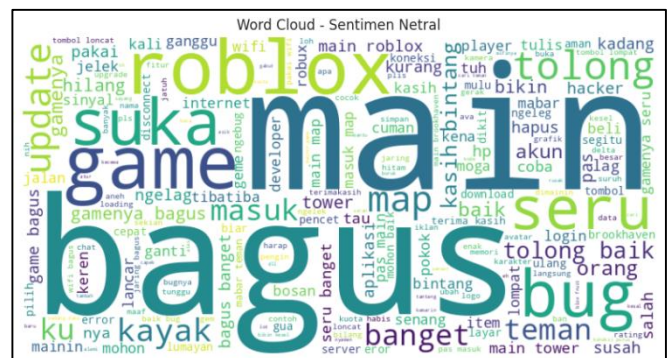


Gambar 12. Wordcloud Ulasan Negatif Game Roblox

Berdasarkan Gambar 12, ditampilkan wordcloud untuk ulasan dengan label sentimen negatif. Kata-kata yang dominan dalam wordcloud tersebut mencerminkan berbagai keluhan dan masalah yang dialami oleh pengguna saat bermain game Roblox. Beberapa kata yang paling sering muncul antara lain: game, main, roblox, enggak, koneksi, bagus, malah, kalah, parah, kurang, susah, online, map, download, lambat, hapus, lag, lama, jelek, gameplay, bug, dan delay. Dari analisis ini, dapat disimpulkan bahwa

ulasan negatif terhadap game Roblox umumnya berkaitan dengan performa permainan, kendala koneksi internet, pembaruan yang dianggap tidak memuaskan, serta keberadaan bug dan lag. Masalah-masalah ini menjadi sumber ketidakpuasan utama di kalangan pengguna. Informasi dari wordcloud ini dapat dimanfaatkan oleh pengembang sebagai masukan penting untuk mengidentifikasi dan memperbaiki aspek-aspek yang sering dikeluhkan. Dengan demikian, pengembang dapat meningkatkan kualitas dan stabilitas game, serta memberikan pengalaman bermain yang lebih baik dan memuaskan bagi pengguna.

Berdasarkan Gambar 13, wordcloud untuk ulasan dengan



Gambar 13. Wordcloud Ulasan Netral Game Roblox

label sentimen netral memperlihatkan kata-kata yang mencerminkan pandangan pengguna yang relatif seimbang. Beberapa kata yang paling dominan dalam visualisasi tersebut antara lain: main, game, bagus, baik, update, koneksi, suka, enggak, tambah, lama, bug, grafik, dan fitur. Dari analisis ini dapat disimpulkan bahwa ulasan netral mencerminkan pengakuan terhadap elemen-elemen positif dalam game, seperti kualitas grafis dan keberagaman fitur. Namun, ulasan ini juga menyoroti beberapa kekurangan yang masih perlu diperbaiki, seperti masalah konektivitas dan bug yang mengganggu pengalaman bermain.

Pandangan netral ini memberikan gambaran yang lebih holistik mengenai penerimaan pengguna, khususnya dari mereka yang tidak memiliki kecenderungan kuat terhadap sentimen positif maupun negatif. Informasi ini dapat menjadi dasar pertimbangan penting bagi pengembang untuk menjaga keunggulan yang sudah ada sekaligus melakukan perbaikan pada aspek-aspek yang masih menjadi perhatian pengguna.



#### IV. KESIMPULAN

Mengingat temuan tersebut eksperimen yang dilakukan dalam penelitian ini, dengan demikian bisa dikatakan bahwa penerapan model *IndoBERT* dalam melakukan klasifikasi sentimen terhadap ulasan game *Roblox* memberikan kinerja yang luar biasa. Model *IndoBERT* mampu mencapai tingkat akurasi sebesar 94%, dengan nilai *presisi*, *recall*, dan *F1-Score* masing-masing sebesar 99%, menunjukkan kemampuan yang tinggi dalam mengklasifikasikan sentimen secara tepat. Proses pengumpulan data dilakukan melalui crawling dari Google Play Store menggunakan library di *Google Colab*, sehingga data yang diperoleh cukup representatif untuk digunakan sebagai dasar pelatihan model. Selama proses pelatihan, model menunjukkan meskipun ada tren peningkatan akurasi yang stabil di awal epoch, juga terdapat tanda-tanda kemungkinan overfitting pada berbagai tahap pelatihan, terutama ketika model mulai belajar data pelatihan dengan terlalu tepat, yang mengurangi kemampuannya untuk menggeneralisasi ke data baru. Pemantauan terhadap tren akurasi dan *loss* selama pelatihan menjadi sangat penting untuk mengidentifikasi titik optimal pemberhentian pelatihan dan menghindari overfitting. Secara keseluruhan, *IndoBERT* terbukti efektif dalam memahami konteks teks dan memberikan prediksi sentimen yang akurat.

Saran untuk penelitian berikutnya lakukan Penggunaan Teknik Regularisasi untuk mengatasi masalah overfitting yang terjadi selama pelatihan, disarankan untuk menerapkan teknik regularisasi seperti dropout layer tambahan, early stopping, atau penggunaan learning rate scheduler agar model tidak terlalu memori terhadap data latih. Evaluasi dengan Metrik Tambahan: Disarankan untuk menggunakan metrik evaluasi tambahan seperti confusion matrix atau visualisasi t-SNE/UMAP untuk memahami distribusi kelas dan area kesalahan model secara lebih mendalam. Perbandingan dengan Model Lain: Sebagai langkah lanjutan, disarankan untuk membandingkan performa *IndoBERT* dengan model bahasa lain seperti *XLM-RoBERTa*, *mBERT*, atau model berbasis LSTM untuk melihat apakah ada peningkatan signifikan atau alternatif yang lebih efisien.

Dengan demikian, meskipun model *IndoBERT* telah menunjukkan hasil yang sangat baik dalam analisis sentimen ulasan game *Roblox*, tetap dibutuhkan pengoptimalan dan validasi lebih lanjut agar model dapat digunakan secara maksimal dalam skenario dunia nyata.

#### DAFTAR PUSTAKA

- [1] M. Haris, A. Suharso, E. H. Nurkifli, P. S. Informatika, U. S. Karawang, And T. Timur, "Analisis Sentimen Pada Game Efootball Di Google Play Store Menggunakan Algoritma Indobert," Vol. 8, No. 6, Pp. 12108–12121, 2024, [Online]. Available: <https://ejournal.itn.ac.id/index.php/jati/article/view/11810>
- [2] Z. Muttaqqin, F. Rahman, P. S. Fisioterapi, F. I. Kesehatan, And U. Muhammadiyah, "Kontribusi Status Ergonomi Terhadap Keterbatasan Fungsional Neck Pada Esports Contribution Of Ergonomic Status To Functional Neck Limitations In Esports," Vol. 17, No. 1, Pp. 175–188, 2025, [Online]. Available: <https://jurisikes.com/index.php/jrk/article/view/2800/775>
- [3] I. Sopiandi And D. Susanti, "Menganalisis Informasi Metaverse Pada Game Online Roblox Secara Garis Besar," *J. Petisi (Pendidikan Teknol. Informasi)*, Vol. 3, No. 1, Pp. 1–4, 2022, Doi: 10.36232/Jurnalpetisi.V3i1.2021.
- [4] T. Thiraviyam, *Artificial Intelligence Marketing*, Vol. 19, No. 4, 2018. [Online]. Available: [https://www.researchgate.net/profile/Budi-Harto/publication/373043823\\_Artificial\\_Intelligence\\_Marketing/links/64d522131290c3cce86a43f/Artificial-Intelligence-Marketing.Pdf](https://www.researchgate.net/profile/Budi-Harto/publication/373043823_Artificial_Intelligence_Marketing/links/64d522131290c3cce86a43f/Artificial-Intelligence-Marketing.Pdf)
- [5] D. Budianto, "Analisis Kepuasan Konsumen Terhadap Kualitas Pelayanan Dan Harga Produk Dengan Menggunakan Metode Customer Satisfaction Index (Csi) Dan Importance Performance Analysis (Ipa) (Studi Kasus : Minimarket Garuda Pekanbaru)," *Skripsi, Fak. Sains Dan Teknol. Univ. Islam Negeri Sultan Syarif Kasim Pekanbaru*, Pp. 1–113, 2018, [Online]. Available: <http://repository.uin-suska.ac.id/1204/>
- [6] M. R. Manoppo, I. C. Kolang, D. N. Fiat, R. Michelly, And C. Mawara, "Analisis Sentimen Publik Di Media Sosial Terhadap Kenaikan Ppn 12 % Di Indonesia Menggunakan Indobert Analysis Of Public Sentiment On Social Media Towards The 12 % Ppn Increase In Indonesia Using Indobert," Vol. 4, No. 2, Pp. 152–163, 2025, Doi: <https://doi.org/10.69916/jkbt.v4i2.322>.
- [7] R. Kusnadi, Y. Yusuf, A. Andriantony, R. Ardian Yaputra, And M. Caintan, "Analisis Sentimen Terhadap Game Genshin Impact Menggunakan Bert," *Rabit J. Teknol. Dan Sist. Inf. Univrab*, Vol. 6, No. 2, Pp. 122–129, 2021, Doi: 10.36341/Rabit.V6i2.1765.
- [8] A. F. Alkindi And N. Nasution, "Analisis Sentimen Ulasan Pengguna Pada Game Roblox Dengan Metode Support Vector Machine Dan Naive Bayes," *J-Com (Journal Comput.)*, Vol. 4, No. 2, Pp. 164–177, 2024, Doi: 10.33330/J-Com.V4i2.3319.
- [9] M. Fuad Nasvian And R. Afif, "Public Opinion On Facebook Rebrand To Meta: A Twitter Big Data Analysis On The First 24 Hours After Meta Launched Opini Publik Terhadap Rebrand Facebook Ke Meta: Analisis Data Besar Twitter Pada 24 Jam Pertama Setelah Meta Diluncurkan," *J. Ilmu Pengetah. Dan Teknol. Komun.*, Vol. 24, No. 1, Pp. 1–19, 2022, [Online]. Available: <https://pdfs.semanticscholar.org/4d94/1d98dc2668bd4e8e0419019bc544506f5024.Pdf>
- [10] S. Imron, E. I. Setiawan, And J. Santoso, "Deteksi Aspek Review E-Commerce Menggunakan Indobert Embedding Dan Cnn," *J. Intell. Syst. Comput.*, Vol. 5, No. 1, Pp. 10–16, 2023, Doi: 10.52985/Insyst.V5i1.267.
- [11] R. R. Pratama, R. R. Suryono, S. Informasi, And U. T. Indonesia, "Performance Comparison Of Naive Bayes , Support Vector Machine And Random Forest Algorithms For Apple Vision Pro Sentiment Analysis Perbandingan Performa Algoritma Naive Bayes , Support Vector Machine Dan Random Forest Untuk Analisis Sentimen Apple," Vol. X, No. December, Pp. 1–9, 2023, Doi: <https://doi.org/10.52436/1.Jutif.2025.6.1.4035>.
- [12] R. Forest And D. A. N. K. N. Dalam, "Analisis Sentimen : Perbandingan Performa Algoritma Naive Bayes , Support Vector

- Machine ,” 2025, [Online]. Available: <https://bitsjournal.researchfloor.org/archives/>
- [13] Rahayu Deny Danar Dan Alvi Furwanti Alwie, A. B. Prasetyo, R. Andespa, P. N. Lhokseumawe, And K. Pengantar, “Tugas Akhir Tugas Akhir,” *J. Ekon. Vol. 18, Nomor 1 Maret 201*, Vol. 2, No. 1, Pp. 41–49, 2020, [Online]. Available: <http://repository.uin-suska.ac.id/id/eprint/85541>
- [14] A. Kunaefi, Z. Abidin, And R. Kusumawati, “Klasifikasi Berita Hoaks Bahasa Indonesia Menggunakan Indobert,” Vol. 10, No. 2, Pp. 1706–1714, 2025, [Online]. Available: <https://jurnal.stkipppgurutlungagung.ac.id/index.php/jipi/article/view/7811>
- [15] N. Widaad, D. Anggraini, E. Faculty, And U. Gunadarma, “Sentiment Analysis Of Chatgpt App User Reviews Using Svm And Cnn,” Vol. 5, No. 6, Pp. 1687–1700, 2024, Doi: <https://doi.org/10.52436/1.Jutif.2024.5.6.4010>.
- [16] A. Agung, A. Daniswara, I. Kadek, And D. Nuryana, “Data Preprocessing Pola Pada Penilaian Mahasiswa Program Profesi Guru,” *J. Informatics Comput. Sci.*, Vol. 05, Pp. 97–100, 2023, Doi: <https://doi.org/10.26740/jinacs.V5n01.P97-100>.
- [17] T. Review, P. Objek, And W. Baturraden, “Aspect Based Sentiment Analysis Menggunakan Indobert Model Melek It,” Vol. 10, No. 2, Pp. 157–166, 2024, Doi: <https://doi.org/10.30742/Melekitjournal.V10i2.358>.
- [18] T. Akhir, “Perbandingan Model Indobert Dan,” 2024, [Online]. Available: [https://cdn.ar-raniry.ac.id/aps-test/teknologi\\_informasi/pembimbing\\_di\\_usk\\_a.n\\_hendri\\_tugas\\_akhir\\_ulfia\\_khairani\\_1908107010068.pdf](https://cdn.ar-raniry.ac.id/aps-test/teknologi_informasi/pembimbing_di_usk_a.n_hendri_tugas_akhir_ulfia_khairani_1908107010068.pdf)
- [19] W. M. Baihaqi And A. Munandar, “Sentiment Analysis Of Student Comment On The College Performance Evaluation Questionnaire Using Naïve Bayes And Indobert,” *Juita J. Inform.*, Vol. 11, No. 2, P. 213, 2023, Doi: [10.30595/juita.V11i2.17336](https://doi.org/10.30595/juita.V11i2.17336).
- [20] U. Khairani, V. Mutiawani, And H. Ahmadian, “Pengaruh Tahapan Preprocessing Terhadap Model Indobert Dan Indobertweet Untuk Mendeteksi Emosi Pada Komentar Akun Berita Instagram,” *J. Teknol. Inf. Dan Ilmu Komput.*, Vol. 11, No. 4, Pp. 887–894, 2024, Doi: [10.25126/jtiik.1148315](https://doi.org/10.25126/jtiik.1148315).
- [21] F. Arifadilah *Et Al.*, *Frido Arifadilah-Fst*. 2023. [Online]. Available: <https://repository.uinjkt.ac.id/dspace/handle/123456789/76873>
- [22] E. Setia Budi, A. Nofriyaldi Chan, P. Priscillia Alda, And M. Arif Fauzi Idris, “Resolusi : Rekayasa Teknik Informatika Dan Informasi Optimasi Model Machine Learning Untuk Klasifikasi Dan Prediksi Citra Menggunakan Algoritma Convolutional Neural Network,” *Media Online*, Vol. 4, No. 5, P. 509, 2024, [Online]. Available: <https://djournals.com/resolusi>
- [23] M. F. Faturrian *Et Al.*, “Analisis Sentimen Terhadap Game Palworld Di Steam Menggunakan Algoritma Bidirectional Encoder,” Vol. 9, No. 1, Pp. 898–905, 2025, Doi: <https://doi.org/10.36040/jati.V9i1.12518>.
- [24] N. L. Sabili And F. R. Umbara, “Klasifikasi Penyakit Diabetes Menggunakan Algoritma Categorical Boosting Dengan Faktor Risiko Diabetes,” Vol. 8, No. 6, Pp. 11391–11398, 2024, Doi: <https://doi.org/10.36040/jati.V8i6.11447>.
- [25] H. B. Tambunan And T. W. D. Hapsari, “Analisis Opini Pengguna Aplikasi New Pln Mobile Menggunakan Text Mining,” *Petir*, Vol. 15, No. 1, Pp. 121–134, 2021, Doi: [10.33322/petir.V15i1.1352](https://doi.org/10.33322/petir.V15i1.1352).
- [26] R. Adolph, “Analisis Sentimen Terhadap Aksi Boikot Produk Pro Israel Pada Komentar Youtube Dengan Menggunakan Algoritma Naïve Bayes Classification (Nbc) & Support Vector Machine (Svm) Disusun,” Pp. 1–23, 2016, [Online]. Available: <https://repository.uinjkt.ac.id/dspace/handle/123456789/81397>