

ANALISIS KOMPARATIF NAIVE BAYES, RANDOM FOREST: INDIKATOR DEPRESI TEKS PADA TWITTER

Ika Sari Romadon¹⁾, Chandra Kirana²⁾

Teknik Informatika ISB Atma Luhur

email : 2311500020@mahasiswa.atmaluhur.ac.id ¹⁾, chandra.kirana@atmaluhur.ac.id ²⁾

Abstraksi

Depresi merupakan isu kesehatan mental krusial pada remaja, di mana stigma sosial mendorong ekspresi emosional secara daring. Penelitian ini bertujuan melakukan studi komparatif antara algoritma Naïve Bayes (NB) dan Random Forest (RF) untuk mengidentifikasi model klasifikasi yang paling efektif dalam deteksi depresi berbasis teks. Data dikumpulkan melalui Twitter API (N = 100 pengguna) dengan prapemrosesan seperti normalisasi slang dan penanganan negasi, kemudian diubah menjadi vektor menggunakan TF-IDF. Hasil menunjukkan Random Forest mencapai akurasi 90%, mengungguli Naïve Bayes yang mencatat akurasi 80%. RF menunjukkan kinerja superior pada metrik diagnostik dengan F1-Score 89,3%, sementara NB memiliki Recall tinggi namun rentan terhadap *False Positive*. Random Forest direkomendasikan untuk sistem deteksi dini karena stabilitas akurasi dan presisi yang tinggi dalam menangani data teks media sosial yang tidak terstruktur.

Kata Kunci :

Depresi, Machine Learning, Random Forest, Naive Bayes, Media Sosial

Abstract

Depression is a crucial mental health issue among adolescents, where social stigma drives emotional expression online. This study aims to conduct a comparative study between Naïve Bayes (NB) and Random Forest (RF) algorithms to identify the most effective classification model for text-based depression detection. Data was collected via the Twitter API (N = 100 users) with preprocessing such as slang normalization and negation handling, then converted into vectors using TF-IDF. The result show that Random Forest achieved 90% accuracy, outperformance Naïve Bayes, which recorded 80% accuracy. RF demonstrated superior performance on diagnostic metrics with an F1-Score of 89.3%, while NB had high Recall but was prone to False Positives. Random Forest is recommended for early detection systems due to its stability in accuracy and high precision in handling unstructured social media text data.

Keywords :

Depression, Machine Learning, Random Forest, Naive Bayes, Social Media

Pendahuluan

Kesehatan mental kini telah bertransformasi menjadi salah satu isu kesehatan global yang paling krusial, di mana diperkirakan lebih dari 300 juta individu di seluruh dunia menderita gangguan depresi[1]. Di Indonesia, beban gangguan jiwa menunjukkan angka yang mengkhawatirkan, terutama pada kelompok remaja dan mahasiswa yang menghadapi tantangan hidup kompleks[2]. Data Riset Kesehatan Dasar (Riskesdas) terbaru mencatat prevalensi depresi pada penduduk usia 15-24 tahun mencapai 6,2%, di mana sekitar 61% di antaranya pernah memiliki pemikiran untuk mengakhiri hidup[3]. Ironisnya, meskipun prevalensi masalah kejiwaan tinggi, tingkat literasi kesehatan mental yang rendah menyebabkan hanya sekitar 10,4% penderita yang aktif mencari pertolongan profesional[3].

Tekanan akademik yang berat, kurangnya dukungan sosial, serta gaya hidup tidak sehat sering kali menjadi faktor utama pemicu depresi di lingkungan pendidikan[2]. Penelitian mengungkap bahwa faktor eksternal seperti perundungan (*bullying*), ekspektasi orang tua yang tinggi, hingga masalah finansial berkontribusi signifikan terhadap munculnya

perasaan tidak berharga dan kesepian pada remaja[4][5]. Situasi ini diperparah oleh keterkaitan antara kecanduan media sosial dengan peningkatan gangguan emosional, di mana remaja cenderung terisolasi dalam dunianya sendiri [6].

Kesenjangan akses layanan kesehatan mental di Indonesia masih menjadi hambatan besar bagi deteksi dini. Rasio psikiater jauh dari standar ideal World Health Organization (WHO) sebesar 1 berbanding 30.000 [7]. Selain keterbatasan fisik, stigma sosial yang kuat di masyarakat sering kali menghambat individu untuk mengakui gejala depresi mereka secara langsung kepada tenaga ahli [8]. Akibatnya, banyak remaja beralih ke media sosial seperti X (Twitter) sebagai ruang anonim untuk mengekspresikan kondisi emosional, opini dan tanda-tanda depresi secara tersirat melalui unggahan teks [9].

Urgensi penelitian ini semakin meningkat seiring dengan tingginya penggunaan media sosial di kalangan remaja serta keterbatasan akses layanan kesehatan mental yang cepat dan responsif. Penggunaan teknologi *Machine Learning* menawarkan solusi inovatif untuk melakukan

deteksi dan klasifikasi pola bahasa yang mengindikasikan gangguan mental secara lebih efisien dibandingkan metode konvensional [10][11]. Berbagai algoritma klasik telah diuji kemampuannya dalam analisis sentimen kesehatan mental, namun pencarian model yang paling stabil dalam menangani teks bahasa Indonesia yang dinamis tetap menjadi tantangan utama.

Penelitian ini bertujuan untuk melakukan studi komparatif kinerja antara algoritma Naive Bayes dan Random Forest dalam deteksi/klasifikasi depresi pada pengguna media sosial X [12]. Fokus utama diberikan pada identifikasi model yang paling efektif dalam memberikan keseimbangan kinerja antara metrik Presisi dan Recall guna mendukung sistem deteksi dini [2]. Dengan memvalidasi model yang paling stabil untuk dataset N = 100 pengguna, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem pemantauan kesehatan mental yang responsif bagi institusi pendidikan di Indonesia [13].

Tinjauan Pustaka

Isu deteksi dini depresi pada remaja melalui unggahan teks di media sosial Twitter (X) telah menjadi fokus krusial dalam penelitian *Machine Learning* [12]. Berbagai studi terdahulu memberikan landasan mengenai efektivitas algoritma klasifikasi, namun juga menunjukkan adanya celah penelitian yang menjadi motivasi bagi pengembangan studi ini. Sebagaimana disajikan pada Tabel 1, berikut adalah perbandingan antara penelitian relevan.

Tabel 1 Perbandingan Penelitian terdahulu

Peneliti	Metode	Kelebihan	Kelemahan
Rahayu, dkk. (2023) [11]	RF, NB, SVM, KNN	RF mencapai akurasi 96% pada Twitter; tangguh menangani data bising.	Belum mendalami optimasi metrik Recall untuk mencegah <i>False Negative</i>
Faisti, dkk. (2025)[12]	NB, RF	RF menunjukkan kinerja stabil (89.3%) pada data teks tidak terstruktur.	Fokus pada teks bahasa Inggris, kurang menangani slang Indonesia.
Santoso, dkk. (2025) [14]	RF, FSA	Optimasi fitur Flamingo Search Algorithm (FSA) meningkatkan akurasi hingga 82.2%. Memberikan analisis efisiensi waktu detail;	Peningkatan akurasi dibayar dengan kompleksitas komputasi sistem yang tinggi.
Setiawan, dkk. (2025) [2]	RF, DT, NB, KNN	NB lebih cepat (0.13 detik) dari RF.	Menggunakan data tabular mahasiswa, bukan teks media sosial yang kompleks.

Majid, dkk. (2025)[13]	Ensemble Stacking	Meningkatkan akurasi hingga 88.13% melalui penggabungan algoritma NB dan RF	Metode ensemble kompleks cenderung sulit diinterpretasikan (<i>black-box</i>).
Situmorang & Purba (2024) [15]	IndoBERT	Model berbasis Transformer mencapai akurasi sangat tinggi sebesar 94.91%	Ketergantungan tinggi pada sumber daya komputasi (GPU).

Berdasarkan analisis pada Tabel 1, penelitian ini bersifat menyempurnakan penelitian terdahulu melalui beberapa argumen strategis:

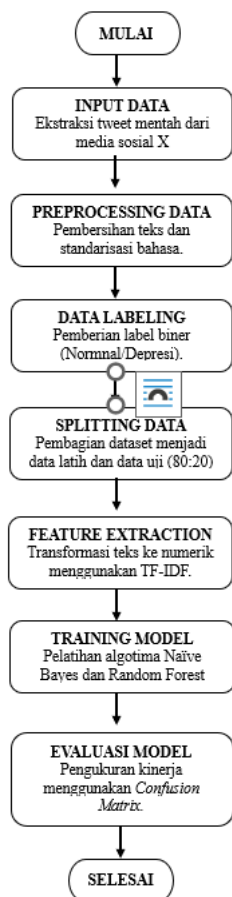
1. Optimasi Fitur Linguistik: Melanjutkan studi Muhammad Jazum Faisti, penelitian ini menerapkan *slang normalization* dan *negation handling* yang ketat guna menangkap makna emosional terdalam pada unggahan remaja di Indonesia.
2. Prioritas Sensitivitas Klinis (Recall): Sebagai penyempurnaan dari riset Kharisma Rahayu, penelitian ini memberikan prioritas pada nilai Recall guna memastikan tidak ada risiko depresi yang terabaikan oleh sistem (meminimalkan *False Negatives*).
3. Keseimbangan Performa dan Efisiensi: Mengacu pada temuan Ismail Setiawan, riset ini mencari titik keseimbangan antara efisiensi Naive Bayes dan kekuatan Random Forest untuk menghasilkan model yang praktis namun tetap akurat bagi layanan kesehatan primer.
4. Fokus pada Kelompok Rentan: Sesuai urgensi yang dipaparkan Mutiara Rahma, penelitian ini memfokuskan deteksi pada pola bahasa remaja yang merupakan kelompok paling terdampak oleh tekanan lingkungan digital.

Secara sistematis, tinjauan ini menunjukkan bahwa penelitian yang dilakukan memiliki posisi strategis untuk mengembangkan solusi deteksi dini depresi yang lebih sensitif, stabil, dan efisien dalam konteks lokal Indonesia.

Metode Penelitian

Bagian ini memuat penjelasan lengkap mengenai langkah-langkah implementatif yang dilakukan untuk memastikan penelitian ini dapat direproduksi kembali (*reproducible*).

Penelitian ini mengikuti kerangka kerja sistematis yang dimulai dari identifikasi masalah hingga evaluasi model.



Gambar 1. Diagram Alir Penelitian

Pengumpulan Data

Tahapan awal dari penelitian ini adalah pengumpulan data teks mentah dari media sosial Twitter. Dataset dikumpulkan secara spesifik menggunakan kueri pencarian yang difokuskan pada kata kunci indikator emosional negatif, seperti “depresi”, “hancur”, “nyerah”, dan “capek”. Total entitas data yang diekstraksi berjumlah 100 cuitan (N=100). Data tersebut kemudian dianotasi (dilabeli) menjadi dua kelas biner: label 1 untuk teks yang menunjukkan indikasi depresi berat atau keputusasaan, dan label 0 untuk teks yang bersifat keluhan harian biasa, opini umum, atau non-depresi. Pada tahap ini, dataset masih dalam bentuk *raw text* yang mengandung banyak *noise* seperti tautan, sebutan pengguna (mention), dan karakter tidak baku.

Prapemrosesan Teks (Text Preprocessing)

Tingkat *noise* yang tinggi paada data teks media sosial mengharuskan adanya prapemrosesan yang ketat agar mesin tidak memproses variabel sampah.

Rangkaian transformasi teks dieksekusi melalui lima tahapan berurutan:

1. Case Folding: menyeragamkan seluruh huruf menjadi huruf kecil (*lowercase*) serta menghapus URL, karakter non-alfabetik, dan tanda baca.
2. Slang Normalization: mengonversi leksikon tidak baku (bahasa gaul atau singkatan seperti “bgt”, “gk”, “knp”) menjadi kata baku bahasa Indonesia.
3. Negation Handling: ini adalah tahap paling krusial untuk menjaga polaritas kalimat. Kata negasi seperti “tidak”, “bukan”, atau “jangan” digabungkan secara paksa dengan kata sifat setelahnya (misal: “tidak pengaruh” menjadi token tunggal “tidak_pengaruh”). Tanpa penanganan ini, algoritma akan membaca kata “pengaruh” secara terpisah dan merusak konteks kalimat.
4. Stopword Removal: mengeliminasi kata hubung yang tidak memiliki bobot sentimen (seperti “dan”, “di”, “ke”).
5. Stemming: menggunakan *library* Sastrawi untuk memangkas imbuhan dan mengembalikan setiap token ke bentuk kata dasarnya.

Pembagian Data (Splitting Data).

Untuk mencegah terjadinya *overfitting* pada model klasifikasi, dataset yang telah bersih kemudian dipartisi menggunakan metode *Train-Test Split* dengan rasio pembagian 80:20.

1. Data Latih (80%): sebanyak 80 sampel dialokasikan untuk melatih algoritma agar mengenali pola kata antara kelas depresi dan non-depresi.
2. Data Uji (20%): sebanyak 20 sampel diisolasi (tidak dilibatkan dalam proses pelatihan) untuk bertindak sebagai alat evaluasi buta guna menguji keandalan prediksi model.

Feature Extraction (Ekstraksi Fitur TF-IDF)

Karena algoritma pembelajaran mesin hanya mengkalkulasi vektor matematis, data teks ditransformasi menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF). TF-IDF tidak sekadar menghitung seberapa sering sebuah kata muncul, melainkan juga menekan bobot kata yang muncul di hampir seluruh cuitan dan meningkatkan bobot (skor) kata-kata unik yang menjadi pembeda kuat indikator depresi (misalnya kata “mati”, “hancur”, “nyerah”). Output dari fase ini adalah matriks berdimensi tinggi yang hanya menjadi variabel independen pada saat klasifikasi.

Training Model dan Evaluasi Kinerja

Berikut pada gambar 2 adalah representasi matriks evaluasi dari data uji menggunakan algoritma Naive Bayes. Dari 20 data teks yang diuji secara acak, model berhasil memprediksi 16 data dengan benar dengan mendapat hasil akurasi 80%.

```

=== HASIL EVALUASI NAIVE BAYES ===
Akurasi: 80.00%

      precision    recall  f1-score   support

 0         0.79         0.85         0.81         11
 1         0.81         0.74         0.78          9

 accuracy         0.80
 macro avg         0.80         0.79         0.80
 weighted avg         0.80         0.80         0.80
    
```

Gambar 2. Pengujian klasifikasi Naive Bayes

Berikut pada gambar 3 adalah representasi matriks evaluasi dari data uji menggunakan algoritma Random Forest. Dari 20 data teks yang diuji secara acak, model mampu mengklasifikasikan 18 data secara tepat dan meminimalkan kesalahan *False Positive*.

```

=== HASIL EVALUASI RANDOM FOREST ===
Akurasi: 90.00%

      precision    recall  f1-score   support

 0         0.91         0.91         0.91         11
 1         0.89         0.89         0.89          9

 accuracy         0.90
 macro avg         0.90         0.90         0.90
 weighted avg         0.90         0.90         0.893
    
```

Gambar 3. Pengujian klasifikasi Random Forest

Matriks TF-IDF dari 80 data latih diumpungkan ke dalam algoritma Naive Bayes dan *Random Forest*. Setelah model mempelajari distribusi pola data, keandalan keduanya diuji menggunakan 20 data uji. Berdasarkan log komputasi, diperoleh perbandingan kinerja sebagai berikut:

1. Kinerja Naive Bayes (NB): Algoritma ini mencatatkan akurasi sebesar 80,00%. Dari 20 data uji, Nb berhasil memprediksi 16 data dengan benar. Meski mampu beroperasi dengan baik, model ini terbukti memiliki kelemahan mendasar: asumsi independensi fitur (naive) pada teorema Bayes gagal memetakan kompleksitas dan ketergantungan antar-kata dalam struktur kalimat bahasa gaul yang sering memutarbalikkan konteks, sehingga menurunkan tingkat *recall* pada kelas depresi menjadi 74%.
2. Kinerja Random Forest (RF): Algoritma ini menunjukkan superioritas yang signifikan dengan tingkat akurasi mencapai 90,00% dan *Weighted Average F1-Score* sebesar 89,3%. RF mampu memprediksi 18 dari 20 data uji secara tepat. Melalui mekanisme *ensemble learning* (penggabungan banyak pohon keputusan secara

acak), algoritma ini terbukti lebih presisi dalam membedakan interaksi antar-fitur kata yang ambigu. Model RF sukses menekan angka *False Positive* tanpa mengorbankan sensitivitas dalam mendeteksi cuitan dengan indikasi depresi nyata.

Secara keseluruhan, hasil komparasi ini membuktikan bahwa mekanisme pemungutan suara mayoritas (*majority voting*) pada arsitektur *Random Forest* jauh lebih stabil dan andal untuk mengklasifikasikan polaritas teks media sosial berbahasa Indonesia yang tidak terstruktur jika dibandingkan dengan pendekatan probabilistik linier dari *Naive Bayes*.

Hasil dan Pembahasan

Hasil Eksperimen dan Metrik Kinerja

Hasil pengujian terhadap N = 100 data tweet menunjukkan perbedaan kinerja yang signifikan antara kedua algoritma (Tabel 2).

Tabel 2 Eksperimen dan Metrik Kinerja

Performa Klasifikasi	Naive Bayes (NB)	Random Forest (RF)
Akurasi	80%	90%
Presisi	70%	95.4%
Recall	87.5%	84%
F1-Score	77.7%	89.3%

Analisis Performa Random Forest

Random Forest mencatat akurasi tertinggi sebesar 90%. Secara ilmiah, keunggulan RF disebabkan oleh kemampuannya menangani fitur teks yang bising melalui mekanisme ensemble learning. Namun, terdapat kasus *False Negative* pada RF. Hal ini disebabkan oleh ketidakmampuan model berbasis pohon dalam menangani representasi teks yang sangat jarang (*sparse*) atau tweet yang terlalu singkat sehingga pola kebahasaan tidak terdeteksi secara optimal.

Analisis Performa Naive Bayes

Naive Bayes mencatat akurasi sebesar 80%. Meskipun akurasi keseluruhan lebih rendah, NB memiliki keunggulan pada sensitivitas deteksi (*Recall*). Namun, NB menghasilkan nilai *False Positive* yang tinggi. Hal ini terjadi akibat asumsi independensi fitur yang "naif", di mana model mengabaikan urutan kata dan konteks semantik, sehingga kata-kata tertentu yang sering muncul di kelas depresi memicu prediksi positif meskipun konteksnya netral atau sarkasme.

Kesimpulan dan Saran

Berdasarkan seluruh hasil analisis dan pembahasan yang telah dilakukan, penelitian ini menyimpulkan bahwa algoritma Random Forest memiliki

keunggulan akurasi (90%) dan stabilitas yang lebih baik dibandingkan dengan Naive Bayes (80%) dalam klasifikasi teks depresi pada media sosial Twitter (X). Random Forest sangat direkomendasikan untuk sistem yang membutuhkan presisi tinggi guna menghindari kesalahan label (*False Positive*), sementara Naive Bayes tetap relevan sebagai alat skrining awal yang sensitif karena nilai Recall-nya yang tinggi.

Saran untuk penelitian selanjutnya adalah melakukan ekspansi jumlah sampel data untuk memperkaya variasi linguistik remaja Indonesia. Penggunaan model bahasa yang lebih canggih seperti IndoBERT juga patut dipertimbangkan untuk mengatasi hambatan makna tersirat pada teks media sosial.

Daftar Pustaka

- [1] M. Fadhilla, R. Wandri, A. Hanafiah, P. R. Setiawan, Y. Arta, and S. Daulay, "Analisis Performa Algoritma Machine Learning Untuk Identifikasi Depresi Pada Mahasiswa," *J. Informatics Manag. Inf. Technol.*, vol. 5, no. 1, pp. 40–47, 2025, doi: 10.47065/jimat.v5i1.473.
- [2] Ismail Setiawan, I. Fatah Yasin, and Y. Tri Desianti, "Komparasi Kinerja Algoritma Random Forest, Decision Tree, Naive Bayes, dan KNN dalam Prediksi Tingkat Depresi Mahasiswa menggunakan Student Depression Dataset," *J. Ilmu Komput. dan Teknol.*, vol. 6, no. 1, pp. 47–58, 2025, doi: 10.35960/ikomti.v6i1.1756.
- [3] S. Zacksavira, C. A. Br Sebayang, J. Imanuel, and G. Br Sibarani, "Implementasi Algoritma Decision Tree untuk Deteksi Depresi Pada Pelajar," *Innov. J. Soc. Sci. Res.*, vol. 5, no. 3, pp. 5915–5934, 2025, doi: 10.31004/innovative.v5i3.19793.
- [4] M. Rahma, M. Fikry, and Y. Afrillia, "Prediksi Kesehatan Mental Remaja Berdasarkan Faktor Lingkungan Sekolah Menggunakan Machine Learning," *J. Inform. J. Pengemb. IT*, vol. 10, no. 2, pp. 382–390, 2025, doi: 10.30591/jpit.v10i2.8556.
- [5] N. Supriani, S. Wahyu Ningrum, and A. Gusti Nugraheni, "Penerapan Algoritma Naive Bayes Dalam Machine Learning Untuk Klasifikasi Faktor-Faktor Yang Mempengaruhi Kesehatan Mental Siswa," *Technol. Informatics Insight J.*, vol. 4, no. 2, pp. 19–29, 2025, doi: 10.32639/y35c0d63.
- [6] H. Aulia *et al.*, "Mental Pada Media Sosial Twitter Dengan Menggunakan," vol. 10, no. 2, pp. 75–81, 2024.
- [7] Tiya Muthia, N. Nurrahma, and Yudi Eka Putra, "Perbandingan Akurasi Model Pembelajaran Mesin SVM, KNN, Decision Tree, dan Naive Bayes pada Klasifikasi Gangguan Kesehatan Mental," *Electr. J. Rekayasa dan Teknol. Elektro*, vol. 18, no. 3, pp. 363–368, 2024, doi: 10.23960/elc.v18n3.2758.
- [8] R. Alfarezy, E. Ermatita, and R. M. B. Wadu, "Implementasi Algoritma Naive Bayes Untuk Analisis Klasifikasi Survei Kesehatan Mental (Studi Kasus: Open Sourcing Mental Illness)," *Inform. J. Ilmu Komput.*, vol. 19, no. 1, pp. 1–10, 2023, doi: 10.52958/iftk.v19i1.4696.
- [9] J. Sibarani *et al.*, "Detection of Mental Health Tendencies Using Naive Bayes Based on Social Media Activity Article Info ABSTRAK," *JOMLAI J. Mach. Learn. Artif. Intell.*, vol. 4, no. 2, pp. 2828–9099, 2025, doi: 10.55123/jomlai.v4i2.5959.
- [10] R. N. S. Amriza, D. Supriyadi, P. DI JI Panjaitan No, K. Purwokerto Selatan, K. Banyumas, and J. Tengah, "Komparasi Metode Machine Learning dan Deep Learning untuk Deteksi Emosi pada Text di Sosial Media," *J. JUPITER*, vol. 13, no. 2, pp. 130–139, 2021, doi: 10.5281/3603.jupiter.2021.10.
- [11] K. Rahayu, V. Fitria, D. Septhya, and ..., "Klasifikasi Teks untuk Mendeteksi Depresi dan Kecemasan pada Pengguna Twitter Berbasis Machine Learning: Text Classification for Detecting Depression and ...," *Indones. J. ...*, vol. 3, no. October, pp. 108–114, 2023, [Online]. Available: <https://journal.irpi.or.id/index.php/malcom/article/view/780>
- [12] M. J. Faisti, R. H. Kusumodestoni, and G. W. N. Wibowo, "Mental Health Classification Using Naive Bayes and Random Forest Algorithms," *J. Appl. Informatics Comput.*, vol. 9, no. 4, pp. 1740–1750, 2025, doi: 10.30871/jaic.v9i4.10144.
- [13] Annisa Maulana Majid, Karina Imelda, and Ismasari Nawangsih, "Komparasi Algoritma Klasifikasi Machine Learning Dengan Penerapan Metode Ensemble Stacking untuk Menganalisa Sentimen terhadap Kesehatan Mental," *SKANIKA Sist. Komput. dan Tek. Inform.*, vol. 8, no. 2, pp. 293–304, 2025, doi: 10.36080/skanika.v8i2.3561.
- [14] Z. R. Santoso, A. H. Wigena, and A. Kurnia, "Mendeteksi Unsur Depresi pada Unggahan Media Sosial Menggunakan Metode Machine Learning dengan Optimasi Berbasis Inspirasi Alam," vol. 6, no. 2, pp. 128–143, 2025, doi: 10.20956/ejsa.v6i2.45516.
- [15] G. F. Situmorang and R. Purba, "Deteksi Potensi Depresi dari Unggahan Media Sosial X Menggunakan IndoBERT," *Build. Informatics, Technol. Sci.*, vol. 6, no. 2, pp. 649–661, 2024, doi: 10.47065/bits.v6i2.5496.