

## PERBANDINGAN KINERJA MODEL MACHINE LEARNING UNTUK SENTIMEN ANALISIS ISU KESEHATAN MENTAL DI FORUM KESEHATAN ONLINE

Lady Agustin Fitriana<sup>[1]</sup>; Deni Risdiansyah<sup>[2]</sup>

<sup>1</sup> Program Studi Sistem Informasi, Fakultas Teknik dan Informatika

<sup>2</sup> Program Studi Informatika, Fakultas Teknik dan Informatika

Universitas Bina Sarana Informatika

[Lady.lag@bsi.ac.id](mailto:Lady.lag@bsi.ac.id)

### INFO ARTIKEL

**Diajukan :**

04-11-2025

**Diterima :**

05-11-2025

**Diterbitkan:**

01-12-2025

**Kata Kunci :**

Analisis Sentimen, Machine Learning, Kesehatan Mental

### INTISARI

Kesehatan mental merupakan aspek penting dalam kehidupan manusia, namun sering kali kurang mendapat perhatian karena adanya stigma sosial dan keterbatasan akses layanan profesional. Di era digital, banyak individu akhirnya menjadikan forum daring seperti Quora dan Reddit sebagai ruang aman untuk mengekspresikan emosi, berbagi pengalaman, serta mencari dukungan. Penelitian ini hadir untuk memahami lebih dalam suara-suara tersebut melalui analisis sentimen, dengan tujuan menggambarkan kondisi emosional masyarakat sekaligus mengembangkan model machine learning yang mampu mengidentifikasi pola sentimen secara akurat. Menggunakan kerangka kerja CRISP-DM, penelitian ini mengumpulkan 2.000 ulasan terkait isu kesehatan mental, kemudian melakukan preprocessing, pelabelan sentimen berbasis leksikon InSet, dan pemodelan menggunakan lima algoritma: Bernoulli Naive Bayes, Decision Tree, Logistic Regression, XGBoost, dan IndoBERT. Hasil analisis menunjukkan bahwa percakapan daring didominasi oleh sentimen negatif seperti depresi, kecemasan, dan pikiran bunuh diri, yang menggambarkan adanya kebutuhan emosional yang besar dalam komunitas digital ini. Dari hasil evaluasi model, Logistic Regression menunjukkan akurasi tertinggi sebesar 72%, sementara XGBoost memberikan stabilitas terbaik melalui weighted F1-score tertinggi. IndoBERT unggul dalam mendeteksi sentimen positif berkat pemahaman konteks bahasa alami yang lebih mendalam. Temuan ini menegaskan bahwa analisis sentimen tidak hanya menjadi proses komputasional, tetapi juga sarana untuk memahami manusia emosi, beban, dan harapan mereka. Penelitian ini diharapkan dapat menjadi langkah awal menuju sistem pemantauan kesehatan mental yang lebih empatik dan responsif.

### I. PENDAHULUAN

Gangguan kesehatan mental merupakan salah satu masalah kesehatan masyarakat yang signifikan di seluruh dunia. Menurut World Health Organization, sekitar 1 dari 4 orang akan mengalami gangguan mental sepanjang hidupnya, yang berdampak besar pada kualitas hidup dan produktivitas individu serta masyarakat ((WHO), 2021). Meskipun prevalensinya tinggi, masalah kesehatan mental sering kali tidak terdeteksi atau tidak mendapatkan penanganan yang memadai akibat stigma sosial dan keterbatasan akses layanan kesehatan. Seiring dengan perkembangan teknologi digital dan meluasnya penggunaan internet, forum kesehatan mental daring dan media sosial menjadi tempat utama bagi individu untuk berbagi pengalaman, mencari dukungan, dan berdiskusi mengenai masalah psikologis mereka

(Usmani et al., 2022)(Ifdil et al., 2023). Data teks dari platform ini sangat berharga karena menggambarkan kondisi psikologis masyarakat secara real-time dan dapat dimanfaatkan untuk mendukung intervensi kesehatan mental yang lebih efektif.

Dalam penelitian ini, forum Quora dan Reddit dipilih sebagai sumber data karena keduanya menyediakan diskusi terbuka yang kaya akan ekspresi emosional, opini, dan pengalaman pribadi terkait kesehatan mental. Kedua platform ini banyak digunakan oleh individu dari berbagai latar belakang untuk membahas isu seperti kecemasan, depresi, stres, dan bentuk gangguan psikologis lainnya. Dengan memanfaatkan teks dari forum ini, penelitian dapat menggali wawasan yang lebih dalam mengenai persepsi dan kondisi emosional masyarakat secara nyata.

Untuk mengolah data teks dari platform tersebut, digunakan pendekatan teknik analisis sentimen yang berasal dari bidang Natural Language Processing (NLP) memungkinkan ekstraksi informasi emosional dari teks, sehingga mampu mengidentifikasi pola-pola sentimen positif maupun negatif dalam percakapan daring (Md Suhaimin et al., 2023). Dalam penelitian ini dapat membantu mengidentifikasi tanda-tanda awal gangguan seperti depresi, kecemasan, dan stres melalui analisis ujaran dan tulisan pengguna forum daring (Xie et al., 2021)(Gao et al., 2021).

Beberapa studi terdahulu telah menggunakan berbagai model machine learning untuk menganalisis sentimen dan kondisi mental dari data sosial media dan forum kesehatan. Misalnya, (Rohan Dixit et al., 2022) menggunakan algoritma NLP dan Lambda Architecture untuk analisis sentimen dari data sensor dan media sosial, guna memprediksi depresi, stres, dan kecemasan.

Hasilnya menunjukkan pendekatan ini efektif memantau kondisi mental secara real-time dengan tetap menjaga privasi pengguna (Md Suhaimin et al., 2023). Penelitian selanjutnya dilakukan oleh (Lase et al., 2023) mengaplikasikan algoritma Naïve Bayes untuk analisis sentimen komentar TikTok terkait kesehatan mental, dengan akurasi 80,95%. Sementara itu, (Oyebode et al., 2020) mengevaluasi 104 aplikasi kesehatan mental dengan analisis sentimen 88.125 ulasan, dan klasifikasi terbaik capai F1-score 89,42% untuk mengungkap 50 tema positif-negatif yang memengaruhi efektivitas aplikasi .

Lebih lanjut, penelitian (Lu et al., 2023) model berbasis Transformer seperti BERT (Bidirectional Encoder Representations from Transformers) berbasis deep learning untuk ekstraksi entitas dari rekam medis elektronik, dan hasilnya mengungguli model lain dalam tugas pengenalan entitas medis. Meskipun demikian, penelitian komparatif yang menguji performa berbagai model machine learning klasik dan modern khususnya pada data forum kesehatan mental masih sangat terbatas.

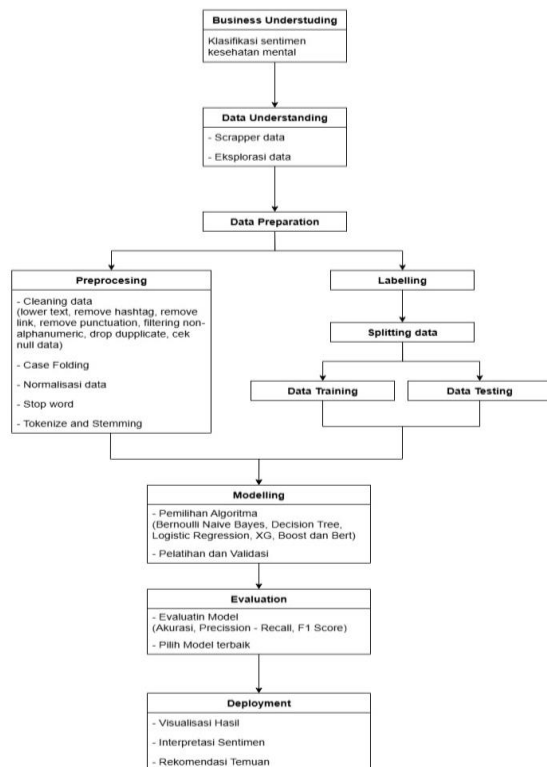
Selain pentingnya pengembangan model yang akurat, penelitian ini juga relevan dalam kerangka inovasi komputasi untuk kesejahteraan global, yaitu pemanfaatan teknologi komputasi untuk meningkatkan kesehatan dan kualitas hidup manusia secara luas (Yuda & Kühner, 2023). Pemahaman lebih baik tentang sentimen dan kondisi kesehatan mental dari data forum daring dapat menjadi landasan pengembangan sistem monitoring otomatis yang memberikan peringatan dini dan rekomendasi intervensi yang tepat waktu. Hal ini sangat penting mengingat kondisi kesehatan mental seringkali dinamis dan membutuhkan respons yang cepat dan akurat (Putler & Krider, 2020).

Oleh karena itu, sangat penting untuk melakukan perbandingan kinerja model-model machine learning dalam konteks analisis sentimen isu kesehatan mental di forum kesehatan online. Penelitian ini bertujuan untuk mengisi kekosongan tersebut dengan membandingkan kinerja empat model, Bernoulli Naive Bayes, Decision Tree, Logistic Regression, XGBoost dan Bert dalam mengklasifikasikan sentimen pada data teks yang berhubungan dengan kesehatan mental. Pelabelan sentimen pada data ini akan dilakukan menggunakan metode lexicon-based, dengan InSet (Indonesian Sentiment Lexicon), untuk mengidentifikasi sentimen positif dan negatif, serta pembobotan sentimen untuk menangkap intensitas emosional seperti kecemasan, depresi, dan stres.

Dengan membandingkan kinerja model-model ini, penelitian ini diharapkan dapat memberikan wawasan yang lebih mendalam mengenai keefektifan masing-masing model dalam memahami perasaan masyarakat terkait isu kesehatan mental. Hasil penelitian ini juga dapat menjadi referensi penting bagi para profesional kesehatan mental dan pembuat kebijakan dalam merancang strategi yang lebih baik untuk menangani masalah kesehatan mental berdasarkan analisis sentimen yang akurat dan efektif.

## II. BAHAN DAN METODE

Penelitian ini menerapkan kerangka kerja CRISP-DM (Cross Industry Standard Process for Data Mining) yang terdiri atas enam tahapan utama, yaitu, business understanding, data understanding, data preparation, modeling, dan evaluation. Adapun tahapan penelitian dapat dijabarkan pada Gambar 1, sebagai berikut :



Gambar 1. Tahapan Penelitian

### 2.1 Business Understanding

Pada tahapan ini, penulis fokus untuk memahami permasalahan dan tujuan utama penelitian, yaitu melakukan klasifikasi sentimen terkait kesehatan mental. Tahap ini bertujuan menentukan bagaimana analisis sentimen dapat digunakan untuk mengidentifikasi persepsi, emosi, atau kondisi psikologis pengguna forum daring terkait isu kesehatan mental. Pemahaman ini menjadi dasar dalam merancang alur penelitian, mulai dari pemilihan sumber data, metode pengolahan teks, algoritma yang digunakan, hingga evaluasi hasil, sehingga output penelitian dapat memberikan informasi yang relevan bagi pemantauan dan intervensi kesehatan mental.

### 2.2 Data Understanding

Penulis melakukan proses pengenalan dan pemahaman karakteristik data yang akan digunakan dalam penelitian. Langkah ini mencakup pengambilan data melalui proses scraping dari sumber yang telah ditentukan (misalnya forum kesehatan daring seperti Reddit atau Quora) serta eksplorasi data untuk melihat struktur, format, dan kualitasnya. Eksplorasi ini meliputi identifikasi jumlah data, distribusi sentimen, panjang teks, keberadaan data kosong atau duplikat, serta potensi noise seperti simbol, tautan, atau karakter non-alfanumerik. Tujuan tahap ini adalah untuk memastikan bahwa data yang diperoleh sesuai

dengan kebutuhan penelitian dan siap diproses lebih lanjut pada tahap data preparation.

### 2.3. Data Preparation

Pada tahapan data preparation, penulis mempersiapkan data agar siap digunakan dalam proses pelatihan dan pengujian model. Tahap ini mencakup tiga proses utama. Pertama, preprocessing, yang meliputi pembersihan data (mengubah teks menjadi huruf kecil, menghapus hashtag, tautan, tanda baca, karakter non-alfanumerik, data duplikat, serta memeriksa data kosong), case folding, normalisasi kata, penghapusan stopwords, serta proses tokenisasi dan stemming untuk memecah teks menjadi kata dasar. Kedua, labelling, yaitu memberikan label sentimen (positif atau negatif) pada setiap data teks sesuai metode yang digunakan, misalnya berbasis leksikon. Ketiga, splitting data, yaitu membagi dataset menjadi data training dan data testing dengan proporsi tertentu untuk memastikan model dapat dilatih dan diuji secara objektif. Tahapan ini penting untuk menjamin kualitas dan konsistensi data sebelum masuk ke tahap modelling.

### 2.4. Modelling

Pada tahapan modelling, penulis membangun dan melatih model machine learning untuk melakukan klasifikasi sentimen terkait kesehatan mental. Proses ini diawali dengan pemilihan algoritma yang akan digunakan, meliputi Bernoulli Naive Bayes, Decision Tree, Logistic Regression, XGBoost, dan BERT. Selanjutnya, dilakukan pelatihan model menggunakan data training yang telah melalui tahap data preparation, diikuti dengan proses validasi untuk mengukur kinerja model pada data uji internal. Tahapan ini bertujuan untuk menemukan model yang mampu memberikan prediksi sentimen dengan performa terbaik, sehingga dapat digunakan pada tahap evaluasi dan implementasi.

### 2.5. Evaluation

Pada tahapan evaluation, penulis mengukur kinerja setiap model machine learning yang telah dibangun untuk menentukan model dengan performa terbaik dalam mengklasifikasikan sentimen terkait kesehatan mental. Evaluasi dilakukan menggunakan metrik seperti akurasi, precision, recall, dan F1-score untuk menilai kemampuan model dalam memprediksi sentimen positif dan negatif secara tepat. Hasil evaluasi ini digunakan untuk membandingkan keunggulan dan kelemahan tiap algoritma, sekaligus memastikan bahwa model yang dipilih memiliki keseimbangan antara tingkat ketepatan prediksi dan kemampuan generalisasi terhadap data baru.

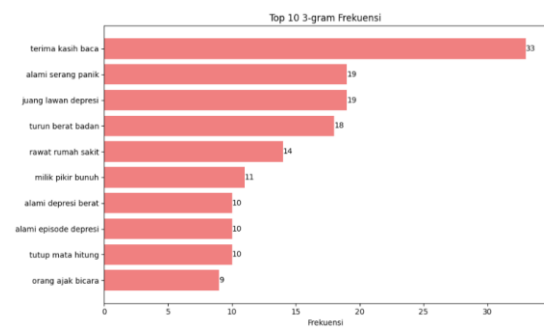
### 2.6 Deployment

Pada tahapan deployment, penulis menerapkan model machine learning terbaik yang telah diperoleh dari proses evaluasi untuk digunakan pada data nyata atau lingkungan aplikasi. Tahap ini mencakup visualisasi hasil analisis sentimen dalam bentuk grafik atau diagram agar lebih mudah dipahami, serta interpretasi temuan untuk menggambarkan pola sentimen yang muncul terkait isu kesehatan mental di forum daring. Selain itu, penulis juga menyusun rekomendasi berdasarkan hasil klasifikasi sentimen, sehingga output penelitian dapat dimanfaatkan sebagai bahan pertimbangan bagi pihak terkait, seperti profesional kesehatan mental atau pembuat kebijakan, dalam merancang strategi intervensi yang tepat.

## III. HASIL DAN PEMBAHASAN

### 3.1 Business Understanding

Pada tahap business understanding, penelitian ini bertujuan memahami arah permasalahan yang akan diselesaikan, yaitu klasifikasi sentimen terkait isu kesehatan mental pada forum daring Quora dan Reddit. Hasil eksplorasi awal terhadap data menunjukkan bahwa sebagian besar pengguna mengekspresikan sentimen negatif seperti depresi, stres, dan kecemasan, dengan intensitas tinggi pada frasa yang menggambarkan kondisi emosional berat. Temuan ini mengindikasikan perlunya penerapan model machine learning untuk menganalisis persepsi publik secara otomatis dan efisien, guna mendukung pemahaman terhadap kondisi psikologis masyarakat digital.



Gambar 2. Top 10 Kata Paling Sering Muncul

Gambar ini menampilkan sepuluh frasa paling sering muncul dalam data forum daring, seperti “alami serang panik”, “juang lawan depresi”, dan “milik pikir bunuh diri”. Dominasi frasa dengan muatan negatif menunjukkan bahwa topik kesehatan mental di forum online banyak diwarnai dengan keluhan emosional dan ekspresi psikologis yang rentan, sehingga memperkuat urgensi penelitian ini dalam membangun sistem analisis sentimen untuk memantau kondisi emosional masyarakat secara otomatis.



Gambar 3. Kerangka Bisnis Penelitian

Gambar ini menggambarkan kerangka konseptual tahap business understanding, yang menghubungkan permasalahan utama berupa tingginya ekspresi negatif di forum daring dengan kebutuhan solusi berbasis machine learning. Diagram ini menjelaskan arah penelitian mulai dari identifikasi masalah hingga pencapaian tujuan akhir, yaitu menentukan model klasifikasi sentimen terbaik untuk isu kesehatan mental.

### 3.2 Data Understanding

Dataset penelitian ini diperoleh melalui proses scraping dari forum daring Quora dan Reddit yang membahas isu kesehatan mental seperti stres, depresi, dan kecemasan. Berdasarkan hasil eksplorasi atau scarping data, ditemukan 2.000 ulasan. Analisis distribusi awal menunjukkan bahwa sebagian besar data memiliki sentimen negatif, diikuti oleh sentimen netral dan positif, menandakan dominasi ekspresi emosional negatif dalam percakapan pengguna. Hasil eksplorasi juga mengungkapkan kemunculan kata-kata dominan seperti “depresi”, “cemas”, “panik”, dan “takut”, yang menggambarkan kondisi psikologis pengguna forum. Selain itu, ditemukan adanya noise berupa tautan, simbol, dan emotikon yang perlu dibersihkan pada tahap preprocessing.



Gambar 4. Wordcloud

Gambar 4 menunjukkan hasil visualisasi word cloud pada data dengan kategori sentimen negatif. Berdasarkan hasil analisis, kata-kata seperti “pikir”, “bunuh”, “hidup”, “mati”, “depresi”, “takut”, dan “sakit” menjadi kata yang paling dominan muncul dalam percakapan pengguna forum daring. Dominasi kata-kata tersebut mengindikasikan adanya intensitas emosi negatif yang tinggi serta kecenderungan topik seputar tekanan psikologis, keputusan, dan kondisi mental yang terganggu. Pola ini memperkuat temuan sebelumnya bahwa forum kesehatan mental banyak digunakan sebagai wadah ekspresi bagi individu yang mengalami masalah emosional dan psikologis.

### 3.3 Data Preparation

Pada penelitian ini, tahap data preparation mencakup tiga proses utama, yaitu preprocessing, labelling, dan splitting data, sebagaimana dijelaskan berikut:

#### 1. Preprocessing

Proses ini dilakukan untuk membersihkan menormalkan dataset yang telah diperoleh dengan tujuan menghilangkan elemen-elemen yang tidak relevan agar dapat diolah secara efektif oleh model machine learning.

##### a. Cleaning

Tahap langkah ini mencakup penghapusan data duplikat, lalu penerapan fungsi seperti `remove_URL()` untuk menghapus tautan, `remove_html()` untuk tag HTML, `remove_emoji()` untuk emotikon, serta `remove_symbols()` dan `remove_numbers()` untuk menghapus tanda baca dan angka. Fungsi `remove_usernames()` juga digunakan untuk menghapus mention @username. Setelah itu dilakukan case folding, yaitu mengubah seluruh teks menjadi huruf kecil agar konsisten.

Sehingga menghasilkan ulasan seperti berikut :

cleaning	case_folding
TAKUT MENYERAH KHAWATIR YA BAGAIMANA YA	takut menyerah khawatir ya bagaimana ya
Bantu mentalku Beneran kayak gini Tadi siang a...	bantu mentalku beneran kayak gini tadi siang a...
kok banyak banget yang kirim pesan minta bikin...	kok banyak banget yang kirim pesan minta bikin...
Menari di tepi kegelisahan	menari di tepi kegelisahan
Kalau kamu lagi sendirian di rumah kamu suka g...	kalau kamu lagi sendirian di rumah kamu suka g...

Gambar 5. Hasil Cleaning

##### b. Normalisasi kata

Pada tahap normalisasi kata bertujuan untuk memecah teks menjadi kata per kata, kemudian mencocokkannya dengan kamus kata tidak baku yang diunduh dari GitHub (`kamuskatabaku.xlsx`) dan dikonversi menjadi dictionary `kamus_tidak_baku_dict`. Jika sebuah kata

ditemukan dalam kamus, kata tersebut diganti dengan padanan bakunya, seperti “gk” menjadi “tidak” atau “bgt” menjadi “banget”. Hasil penggantian disimpan dalam beberapa variabel seperti `kalimat_baku`, `kata_diganti`, dan `kata_tidak_baku_hash`

case_folding	normalisasi
takut menyerah khawatir ya bagaimana ya	takut menyerah khawatir ya bagaimana ya
bantu mentalku beneran kayak gini tadi siang a...	bantu mentalku beneran kayak begini tadi siang...
kok banyak banget yang kirim pesan minta bikin...	kok banyak banget yang kirim pesan meminta bik...
menari di tepi kegelisahan	menari di tepi kegelisahan
kalau kamu lagi sendirian di rumah kamu suka g...	kalau kamu lagi sendirian di rumah kamu suka g...

Gambar 6. Hasil Case Folding dan Normalisasi

#### c. Tokenisasi

Proses ini memecah teks menjadi potongan-potongan kata (*tokens*) menggunakan metode `split()`

normalisasi	tokenize
takut menyerah khawatir ya bagaimana ya	[takut, menyerah, khawatir, ya, bagaimana, ya]
bantu mentalku beneran kayak begini tadi siang...	[bantu, mentalku, beneran, kayak, begini, tadi...
kok banyak banget yang kirim pesan meminta bik...	[kok, banyak, banget, yang, kirim, pesan, memi...
menari di tepi kegelisahan	[menari, di, tepi, kegelisahan]
kalau kamu lagi sendirian di rumah kamu suka g...	[kalau, kamu, lagi, sendirian, di, rumah, kamu...

Gambar 7. Hasil Tokenisasi

#### d. Stemming

Proses ini diawali dengan pembuatan objek stemmer menggunakan `StemmerFactory()` dari library Sastrawi, lalu dipanggil dengan `factory.create_stemmer()`. Selanjutnya, fungsi `stem_text(text)` dibuat untuk melakukan perulangan pada setiap kata dalam list token yang tersimpan di kolom stopword removal. Di dalam fungsi tersebut, setiap kata diproses dengan perintah `stemmer.stem(word)` yang akan menghilangkan awalan, sisipan, maupun akhiran sesuai kaidah bahasa Indonesia. Proses ini memastikan bahwa kata-kata seperti “berbicara”, “pembicaraan”, dan “berbicaralah” dikonversi menjadi bentuk dasar “bicara”, sehingga memperkuat konsistensi data sebelum tahap analisis sentimen.

stopword removal	stemming_data
[takut, menyerah, khawatir, ya, ya]	takut serah khawatir ya ya
[bantu, mentalku, benaran, kayak, siang, bilan...]	bantu mental benar kayak siang bilang diri heb...
[banget, kirim, pesan, bikin, menangis, dijela...]	banget kirim pesan bikin menang dijelaskan tagn...

Gambar 8. Hasil Stemming

### 2. Labelling

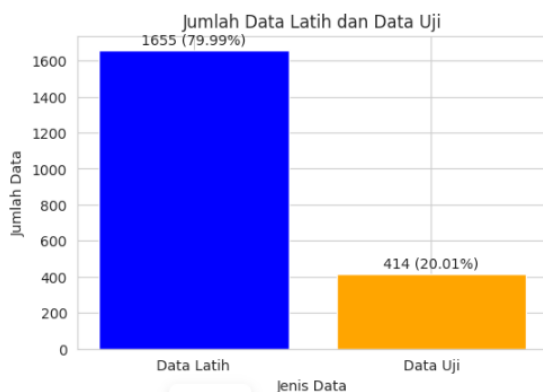
Tahap labelling menggunakan pendekatan lexicon-based dengan kamus sentimen InSet yang berisi daftar kata positif dan negatif. Berdasarkan fungsi `determine_sentiment()` menghitung jumlah kata positif (`positive_count`) dan negatif (`negative_count`) dalam teks. Nilai skor sentimen diperoleh dari selisih keduanya. Jika hasilnya  $> 0$  maka positif,  $< 0$  maka negatif, dan  $= 0$  maka netral.

	stemming_data	Score	Sentiment
0	takut serah khawatir ya ya	2	Positif
1	bantu mental benar kayak siang bilang diri heb...	-2	Negatif
2	banget kirim pesan bikin menang dijelaskan tagn...	1	Positif
3	tari tepi gelisah	-1	Negatif
4	rumah suka gelisah taksi	0	Netral

Gambar 9. Hasil Labelling

### 3. Split Data

Tahap split data dilakukan untuk membagi dataset menjadi data latih (80%) dan data uji (20%) menggunakan fungsi `train_test_split()` dari `scikit-learn`. Variabel X berisi teks hasil stemming dari kolom `stemming_data`, sedangkan y berisi label sentimen dari kolom `Sentiment`. Parameter `random_state=42` digunakan agar pembagian data konsisten.

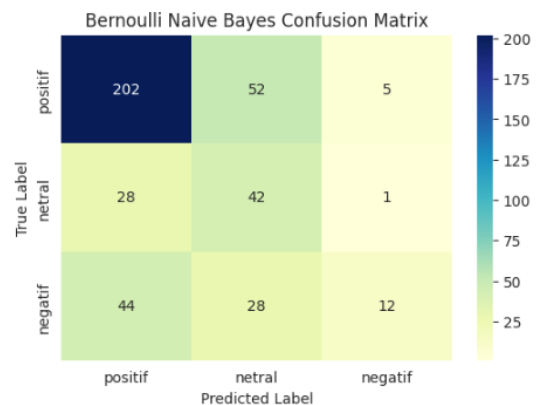


Gambar 10. Uji Data Training dan Testing

### 3.4 Modelling

#### 1. Bernouli Naie Bayes

Berdasarkan hasil *classification report* dari model Bernouli Naive Bayes, terlihat bahwa model ini memiliki akurasi keseluruhan sebesar 62%, yang berarti dari seluruh data uji, sekitar 62% prediksi sentimen berhasil diklasifikasikan dengan benar. Kelas Negatif memiliki kinerja terbaik dengan *precision* 0.74, *recall* 0.78, dan *f1-score* 0.76, menunjukkan bahwa model cukup andal dalam mengenali sentimen negatif. Sebaliknya, kelas Netral memiliki *precision* rendah (0.34) namun *recall* cukup tinggi (0.59), menandakan bahwa model sering salah mengenali kelas lain sebagai netral, tetapi relatif mampu menangkap sebagian besar data netral yang sebenarnya. Untuk kelas Positif, performanya masih rendah dengan *recall* hanya 0.14 dan *f1-score* 0.24, menandakan model kesulitan membedakan sentimen positif dari yang lain. Nilai *macro average* yang berada di kisaran 0.48 menunjukkan ketidakseimbangan performa antar kelas, sedangkan *weighted average* 0.60 menegaskan bahwa performa model masih didominasi oleh kelas dengan jumlah data terbanyak, yaitu kelas negatif.

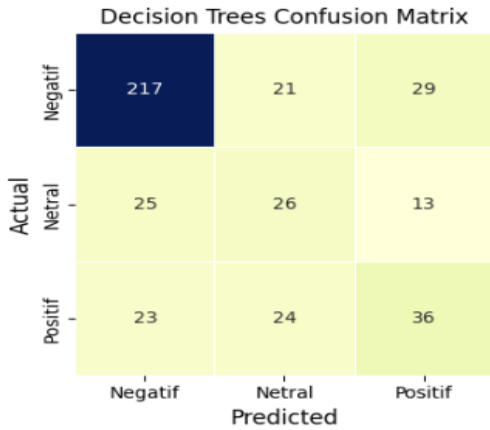


Gambar 10. BNN Matrix

#### 2. Decision Tree

Berdasarkan hasil *classification report* untuk model Decision Tree, diperoleh akurasi sebesar 67,4%, menunjukkan bahwa model ini mampu mengklasifikasikan data dengan tingkat ketepatan yang cukup baik dibandingkan beberapa model lainnya. Kelas Negatif kembali menjadi yang paling akurat dengan *precision* 0.819, *recall* 0.813, dan *f1-score* 0.816, menandakan bahwa model sangat efektif dalam mengenali sentimen negatif. Namun, performa menurun signifikan pada kelas Netral dengan *precision* 0.366 dan *f1-score* 0.385, serta pada kelas Positif dengan *f1-score* 0.447, yang mengindikasikan bahwa model masih kesulitan membedakan kedua kelas tersebut secara konsisten. Nilai *macro average* 0.549 mencerminkan adanya ketimpangan performa

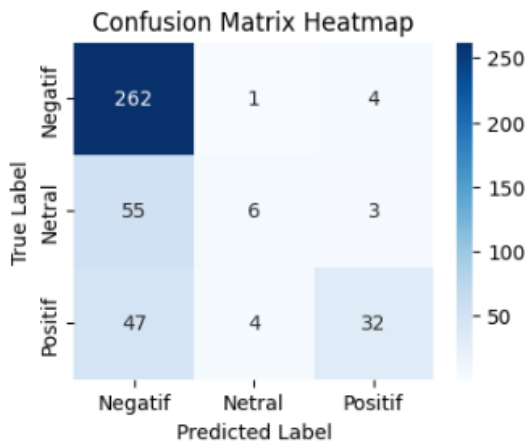
antar kelas, sementara *weighted average* 0.675 menunjukkan bahwa hasil keseluruhan model dipengaruhi lebih besar oleh kelas dengan jumlah data yang dominan, yaitu kelas negatif.



Gambar 11. Decision Tree Matrix

3. Logistic Regression

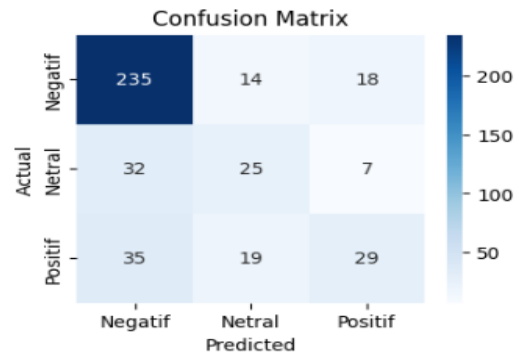
Model Logistic Regression menunjukkan akurasi 72%, dengan performa terbaik pada kelas Negatif (*f1-score* 0.83) yang berhasil dikenali hampir sempurna (*recall* 0.98). Namun, model masih lemah dalam membedakan kelas Netral (*f1-score* 0.16) dan cukup moderat pada kelas Positif (*f1-score* 0.52). Secara keseluruhan, performa model masih dipengaruhi dominasi kelas negatif, terlihat dari *weighted average* sebesar 0.67.



Gambar 12. Logistic Regression Matrix

4. XG Boost

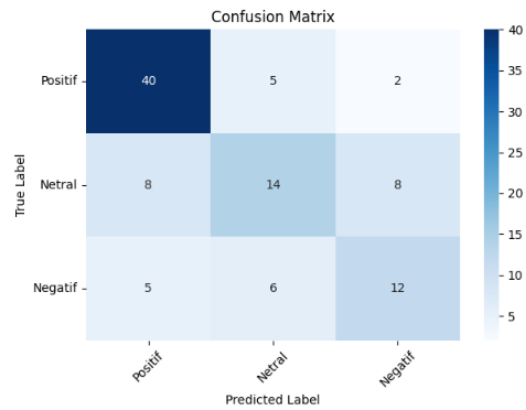
Model XGBoost mencapai akurasi 69,81%, menunjukkan performa yang cukup stabil dalam klasifikasi sentimen. Kelas Negatif memiliki hasil terbaik dengan *f1-score* 0.83 dan *recall* 0.88, menandakan kemampuan tinggi dalam mengenali data negatif. Sementara itu, kelas Netral dan Positif masih menunjukkan performa sedang dengan *f1-score* masing-masing 0.41 dan 0.42.



Gambar 13. XG Boost Matrix

5. IndoBert

Model IndoBERT memperoleh akurasi 66%, menunjukkan performa yang cukup baik untuk analisis sentimen berbasis bahasa Indonesia. Kelas Positif menjadi yang paling unggul dengan *f1-score* 0.80 dan *recall* 0.85, menandakan model sangat efektif mengenali sentimen positif. Sebaliknya, performa pada kelas Netral (*f1-score* 0.51) dan Negatif (*f1-score* 0.53) masih perlu ditingkatkan karena model cenderung kesulitan membedakan kedua kelas tersebut. Nilai *macro average* dan *weighted average* yang sama-sama sekitar 0.61-0.65 menunjukkan bahwa meskipun model stabil secara keseluruhan, hasilnya masih didominasi oleh keberhasilan pada kelas positif.



Gambar 14. IndoBert Matrix

3.5 Evaluation

Dari hasil modelling, terlihat bahwa Logistic Regression memiliki akurasi tertinggi sebesar 72%, menunjukkan kemampuan terbaik dalam klasifikasi keseluruhan. Namun, XGBoost menunjukkan keseimbangan performa yang baik dengan *weighted F1-score* tertinggi (0.68) dan hasil yang cukup stabil di tiap kelas. Decision Tree juga tampil solid dengan akurasi 67,4% dan *macro F1* 0.55, menandakan model cukup andal pada semua kategori, meski masih sedikit tertinggal dari XGBoost dalam hal konsistensi. IndoBERT, meskipun akurasinya 66%, unggul pada kelas Positif (*f1-score* 0.80),

memperlihatkan kemampuannya memahami konteks bahasa alami yang lebih baik. Sementara itu, Bernoulli Naive Bayes memiliki akurasi terendah (62%) dan kinerja tidak seimbang antar kelas. Secara keseluruhan, XGBoost dan Logistic Regression menjadi model paling kompetitif, dengan XGBoost unggul dalam keseimbangan dan stabilitas antar kelas, sedangkan Logistic Regression lebih optimal pada data yang didominasi oleh kelas negatif. Berikut tabel perbandingan :

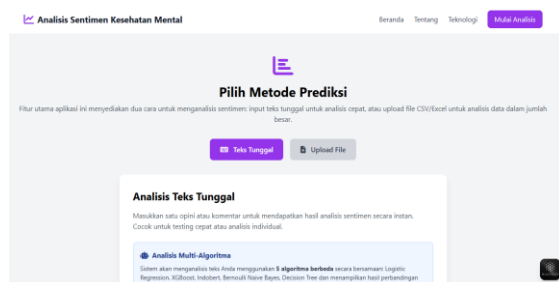
**Tabel 1 Perbandingan Algoritma Machine Learning**

Model	Accuracy	F1 Negatif	F1 Netral	F1 Positif	Macro Avg F1	Weighted Avg F1
Bernoulli Naive Bayes	0.62	0.76	0.44	0.24	0.48	0.60
Decision Tree	0.674	0.816	0.385	0.447	0.549	0.675
Logistic Regression	0.72	0.83	0.16	0.52	0.51	0.67
XGBoost	0.6981	0.83	0.41	0.42	0.55	0.68
IndoBERT	0.66	0.53	0.51	0.80	0.61	0.65

Model Logistic Regression memiliki akurasi tertinggi sebesar 0.72, menunjukkan performa paling baik secara keseluruhan, terutama dalam mengenali kelas negatif. XGBoost menempati posisi kedua dengan hasil yang seimbang antar kelas dan *weighted F1-score* tertinggi (0.68), menandakan kestabilan performa di berbagai kategori sentimen. Decision Tree menunjukkan kinerja yang cukup stabil, namun masih kurang akurat dalam mengidentifikasi kelas netral dan positif. Sementara itu, IndoBERT unggul pada kelas positif dengan *F1-score* 0.80, yang mencerminkan kemampuannya memahami konteks bahasa alami dengan baik. Di sisi lain, Bernoulli Naive Bayes memiliki performa terendah, terutama pada kelas positif, sehingga kurang optimal digunakan untuk data dengan konteks bahasa yang kompleks.

### 3.6 Deployment

Pada tahap deployment, dikembangkan sebuah website analisis sentimen kesehatan mental yang dirancang untuk membandingkan kinerja lima algoritma machine learning, yaitu Logistic Regression, XGBoost, IndoBERT, Bernoulli Naive Bayes, dan Decision Tree.

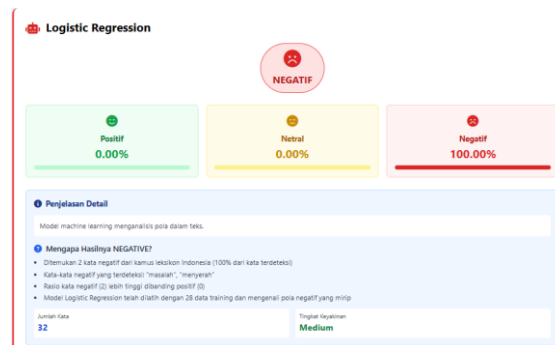


Gambar 15. Analisis Teks

Gambar 15 merupakan tampilan dari sisi user yang di mana pengguna dapat memasukkan opini mereka ke dalam kolom yang disediakan. Kemudian sistem secara otomatis menampilkan hasil analisis sentimen beserta perbandingan keluaran dari masing-masing algoritma tersebut sesuai gambar 16.

Algoritma	Hasil Sentimen	Skor Positif	Skor Netral	Skor Negatif	Akurasi
Logistic Regression	NEGATIF	0.00%	0.00%	100.00%	High
XGBoost	NEGATIF	0.00%	0.00%	100.00%	Very High
IndoBERT	NEGATIF	0.00%	0.00%	100.00%	Very High
Bernoulli Naive Bayes	NEGATIF	0.00%	0.00%	100.00%	High
Decision Tree	NEGATIF	0.00%	0.00%	100.00%	High

Gambar 16. Perbandingan Hasil Algoritma



Gambar 17. Hasil Algoritma

Gambar 17 menampilkan hasil analisis sentimen menggunakan algoritma Logistic Regression dengan prediksi akhir NEGATIF dan tingkat keyakinan Medium. Model menunjukkan skor Positif 0.00%, Netral 0.00%, dan Negatif 100.00%, menandakan dominasi sentimen negatif. Sistem mendeteksi kata negatif seperti “masalah” dan “menyerah” dari leksikon Indonesia, sehingga rasio negatif lebih tinggi.

## IV. KESIMPULAN

Penelitian ini menyimpulkan bahwa analisis sentimen terhadap isu kesehatan mental pada forum daring menunjukkan dominasi ekspresi negatif seperti depresi, kecemasan, dan pikiran bunuh diri, sehingga memerlukan pendekatan machine learning yang mampu mengolah teks secara efektif. Setelah melalui tahapan CRISP-DM, hasil evaluasi lima model menunjukkan bahwa Logistic Regression memberikan akurasi tertinggi sebesar 72% dan menjadi model paling optimal secara keseluruhan, sementara XGBoost menunjukkan kestabilan performa terbaik melalui *weighted F1-score* tertinggi, dan IndoBERT unggul dalam mengenali sentimen positif berkat kemampuan pemahaman konteks bahasa alami yang lebih baik. Perbandingan ini menegaskan

bahwa tidak ada satu model yang unggul secara absolut untuk seluruh kategori sentimen, namun kombinasi hasil menunjukkan bahwa Logistic Regression dan XGBoost merupakan kandidat terbaik untuk implementasi sistem analisis sentimen kesehatan mental, yang berpotensi mendukung monitoring emosional masyarakat digital secara lebih akurat dan responsif.

Ke depan, penelitian ini dapat ditingkatkan dengan memperkaya dan menyeimbangkan data agar setiap jenis sentimen dapat terbaca lebih akurat oleh model. Melibatkan psikolog atau psikiater dalam proses pelabelan data secara manual juga akan sangat membantu agar makna emosional dalam setiap tulisan pengguna benar-benar dipahami sesuai konteks kesehatan mental. Selain itu, pengembangan model berbasis Transformer serta perluasan sumber data dari berbagai platform dapat membuat hasil analisis lebih menyeluruh. Dengan langkah-langkah ini, sistem analisis sentimen yang dibangun diharapkan dapat menjadi alat bantu yang lebih manusiawi dan bermanfaat dalam memahami kondisi emosional masyarakat secara lebih mendalam.

#### REFERENSI

- (WHO), W. H. O. (2021). *Comprehensive Mental Health Action Plan 2013-2030*. <https://www.who.int/publications/i/item/9789240031029>
- Gao, Y., Qian, S., Li, Z., Wang, P., Wang, F., & He, Q. (2021). Digital twin and its application in transportation infrastructure. *Proceedings 2021 IEEE 1st International Conference on Digital Twins and Parallel Intelligence, DTPI 2021*, 298-301. <https://doi.org/10.1109/DTPI52967.2021.9540108>
- Ifdil, I., Situmorang, D. D. B., Firman, F., Zola, N., Rangka, I. B., & Fadli, R. P. (2023). Virtual reality in Metaverse for future mental health-helping profession: an alternative solution to the mental health challenges of the COVID-19 pandemic. *Journal of Public Health (United Kingdom)*, 45(1), E142-E143. <https://doi.org/10.1093/pubmed/fdac049>
- Lase, Y. Y., Lubis, A. R., Elyza, F., & Syaflia, S. A. (2023). Mental Health Sentiment Analysis on Social Media TikTok with the Naïve Bayes Algorithm. *Proceedings - 2023 6th International Conference on Computer and Informatics Engineering: AI Trust, Risk and Security Management (AI Trism), IC2IE 2023*, 186-191. <https://doi.org/10.1109/IC2IE60547.2023.10331126>
- Lu, W., Jiang, J., Shi, Y., Zhong, X., Gu, J., Huangfu, L., & Gong, M. (2023). Application of Entity-BERT model based on neuroscience and brain-like cognition in electronic medical record entity recognition. *Frontiers in Neuroscience*, 17. <https://doi.org/10.3389/fnins.2023.1259652>
- Md Suhaimin, M. S., Ahmad Hijazi, M. H., Moug, E. G., Nohuddin, P. N. E., Chua, S., & Coenen, F. (2023). Social media sentiment analysis and opinion mining in public security: Taxonomy, trend analysis, issues and future directions. In *Journal of King Saud University - Computer and Information Sciences* (Vol. 35, Issue 9). <https://doi.org/10.1016/j.jksuci.2023.101776>
- Oyebode, O., Alqahtani, F., & Orji, R. (2020). Using Machine Learning and Thematic Analysis Methods to Evaluate Mental Health Apps Based on User Reviews. *IEEE Access*, 8, 111141-111158. <https://doi.org/10.1109/ACCESS.2020.3002176>
- Putler, D. S., & Krider, R. E. (2020). A Process Model for Data Mining—CRISP-DM. In *Customer and Business Analytics* (pp. 43-56). <https://doi.org/10.1201/b12040-8>
- Rohan Dixit, Geetika Chawla, & Ishaan Bajaj. (2022). Mental Health Monitoring using Sentiment Analysis. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 324-330. <https://doi.org/10.32628/cseit228443>
- Usmani, S. S., Sharath, M., & Mehendale, M. (2022). Future of mental health in the metaverse. In *General Psychiatry* (Vol. 35, Issue 4). <https://doi.org/10.1136/gpsych-2022-100825>
- Xie, Y., Lian, K., Liu, Q., Zhang, C., & Liu, H. (2021). Digital twin for cutting tool: Modeling, application and service strategy. *Journal of Manufacturing Systems*, 58, 305-312. <https://doi.org/10.1016/j.jmsy.2020.08.007>
- Yuda, T. K., & Kühner, S. (2023). Bringing Indonesia into the global welfare regime debate: A literature review and future research agenda. *Asian Social Work and Policy Review*, 17(2), 103-114. <https://doi.org/10.1111/aswp.12275>