SKANIKA: Sistem Komputer dan Teknik Informatika
Volume 8, Nomor 2, Juli 2025, Halaman 293-304 E-ISSN: 2721-4788

Komparasi Algoritma Klasifikasi *Machine Learning* dan Penerapan Metode *Ensemble Stacking* untuk Menganalisis Sentimen Kesehatan Mental

Annisa Maulana Majid[1*], Karina Imelda[2], Ismasari Nawangsih[3]
[1,2,3]Fakultas Teknik, Teknik Informatika, Universitas Pelita Bangsa, Bekasi, Indonesia
E-mail: [1*]annisa.maulanamajid@pelitabangsa, [2]karina.imelda@pelitabangsa.ac.id,
[3]ismasari.n@pelitabangsa.ac.id
(* : corresponding author)

Abstrak

Kesehatan mental sering tidak terdeteksi karena tidak terlihat secara fisik. Hal tersebut mengakibatkan terhambatnya proses penanganan secara cepat dan tepat. Sebagian individu memilih untuk berekspresi di media sosial dibanding akses jasa layanan profesional, namun penggunaan media sosial dapat memperburuk kondisi mental bahkan berdampak pada kondisi fisik, untuk itu perlu adanya analisa atau deteksi dini terhadap kesehatan mental dengan pendekatan teknologi *machine learning* untuk analisis data digital atau unggahan di media sosial. Penelitian sebelumnya terkait analisa sentimen kesehatan mental sudah dilakukan menggunakan algoritma klasifikasi *machine learning* namun perlu adanya peningkatan hasil akurasi. Penelitian ini membandingkan algoritma klasifikasi tunggal dan menerapkan metode ensemble *Stacking* dengan menggabungkan algoritma klasifikasi menjadi *base learner* dan *meta learner*. Hasil penelitian menunjukkan adanya peningkatan akurasi menggunakan metode *stacking* yaitu sebesar 88,13%.

Kata kunci: Sentimen, Kesehatan Mental, Klasifikasi, *Stacking*

Abstract

Mental health often goes undetected due to the absence of physical symptoms, which hinders timely and appropriate intervention. Many individuals choose to express their emotions on social media rather than access professional services. However, the use of social media can potentially worsen mental health conditions and even impact physical well-being. Therefore, early detection through the analysis of digital data, particularly social media posts, using machine learning approaches is essential. Previous research on mental health sentiment analysis has utilized classification algorithms, but accuracy improvement remains necessary. This study compares single classification algorithms and applies an ensemble stacking method that combines multiple classifiers as base learners and a meta-learner. The application of the stacking method resulted in an improved accuracy score of 88.13%.

Keywords: *Sentiment, Mental Health, Classification, Stacking*

1. PENDAHULUAN

Kesehatan mental merupakan suatu aspek penting bagi setiap individu karena dapat menunjang kesejahteraan individu secara menyeluruh. Berbeda dengan gangguan kesehatan fisik yang umumnya dapat dilihat, gangguan kesehatan mental sering kali tidak tampak secara fisik sehingga tidak terdeteksi pada tahap awal. Kondisi tersebut mengakibatkan keterlambatan proses penanganan secara tepat waktu. Gangguan mental seperti depresi atau cemas saat ini semakin banyak ditemukan, namun tingkat kepedulian masyarakat terkait isu penyakit kesehatan mental tergolong masih rendah, sehingga sebagian individu memilih mengekspresikan melalui media sosial dari pada mengakses langsung jasa layanan kesehatan profesional. Media sosial merupakan ruang untuk berekspresi namun dalam hal kesehatan mental dapat memiliki potensi risiko, terutama individu menerima respons negatif atau mengalami tindakan perundungan yang dapat memperburuk kondisi mentalnya. Gangguan kesehatan mental akan berdampak memperburuk kondisi fisik jika tidak ditindak sejak awal, seperti kelelahan, penurunan tingkat konsentrasi, sakit pada kepala, gangguan pencernaan, dan peningkatan risiko perilaku berbahaya, seperti penyalahgunaan zat adiktif, melukai diri, hingga upaya bunuh diri. Masalah kesehatan mental dapat berdampak luas, mulai dari individu hingga masyarakat dan perekonomian negara, karena

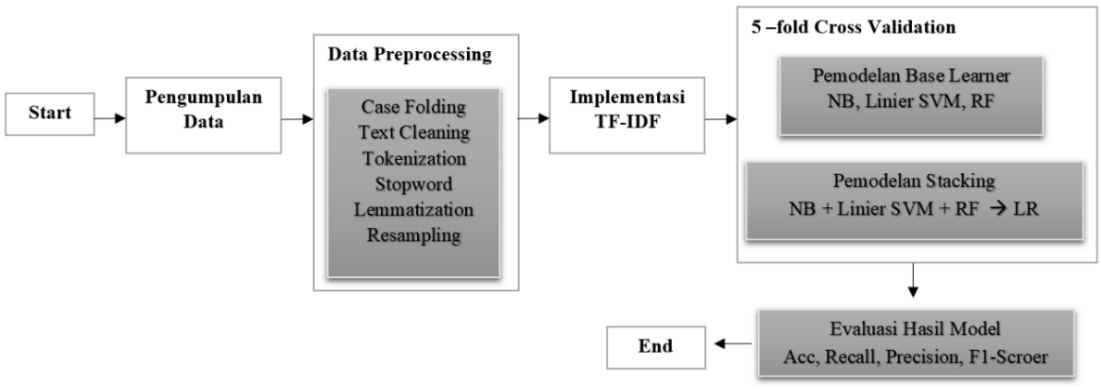
kesehatan mental berperan penting dalam menciptakan sumber daya manusia yang unggul.

Riset Kesehatan Dasar (Riskesdas) di Indonesia telah mencatat adanya kenaikan hasil prevalensi terkait gangguan jiwa dari 1,7% yang terjadi pada tahun 2013 menjadi 6,7% pada tahun 2018 menggunakan alat skrining *Self Reporting Questionnaire*-20 (SRQ-20) [1]. Sedangkan Survei Kesehatan Indonesia (SKI) tahun 2023 memfokuskan pada gangguan jiwa berat, seperti gejala skizofrenia/psikosis, melalui pendekatan wawancara terstruktur. Hasil SKI menunjukkan bahwa 4,0% rumah tangga nasional memiliki anggota keluarga dengan gejala gangguan jiwa berat [2]. Perbedaan metode dan indikator membuat data ini tidak dapat dibandingkan secara langsung, namun tetap menunjukkan bahwa masalah kesehatan jiwa di Indonesia masih sangat signifikan dan membutuhkan intervensi yang berkelanjutan. Hasil riset ini menunjukkan urgensi untuk memperkuat upaya pencegahan dalam bidang kesehatan mental, termasuk melalui pemanfaatan teknologi. Salah satu pendekatan yang dapat dilakukan dengan penerapan algoritma *Machine learning* untuk analisis dan deteksi dini terhadap potensi gangguan kesehatan mental melalui data digital unggahan di media sosial. Penelitian sebelumnya tentang analisis sentimen kesehatan mental sudah pernah dilakukan. Penelitian [3] tentang analisis sentimen komentar isu kesehatan mental menunjukkan hasil akurasi sebesar 68,01% pada algoritma *Naive Bayes* dan 48,97% pada algoritma KNN. Penelitian [4] tentang klasifikasi stres pada media sosial twitter menunjukkan hasil akurasi sebesar 59,11% pada algoritma SVM. Penelitian [5] tentang analisis sentimen *mental illness* menunjukkan hasil akurasi sebesar 67,8% pada algoritma KNN, 80,6% pada algoritma *Random Forest*, dan 79 % pada algoritma *Neural Network*. Penelitian [6] tentang analisis sentimen kesehatan mental pada twitter menunjukkan hasil akurasi sebesar 58,39% pada algoritma KNN, 49,69% pada algoritma *Decision tree*, dan 56,83% pada algoritma SVM. Penelitian sebelumnya menunjukkan hasil akurasi algoritma klasifikasi belum maksimal sehingga perlu adanya peningkatan performa. Oleh karena itu, penelitian ini menggunakan metode Ensemble *Stacking* untuk peningkatan algoritma klasifikasi. Metode ini dipilih karena mampu menggabungkan kelebihan dari beberapa algoritma klasifikasi yang berbeda, dan menghasilkan model akhir yang lebih stabil dan akurat dibandingkan dengan penggunaan satu algoritma tunggal.

Pada penelitian ini membandingkan algoritma klasifikasi *Machine learning* diantaranya algoritma *Logistic Regression, Naïve Bayes, Linier SVM,* dan Random Forest dengan penerapan metode *Ensemble Stacking* melalui penggabungan algoritma klasifikasi menjadi *base learner* dan *meta learner* untuk meningkatkan akurasi terhadap analisis kesehatan mental.

2. METODE PENELITIAN

Penelitian ini akan membandingkan algoritma klasifikasi *Machine learning* yaitu *Logistic Regression, Naive Bayes, Linier Support Vector Machine* (SVM), dan *Random Forest* tunggal dengan penerapan metode *Ensemble Stacking* untuk menganalisis sentimen terhadap kesehatan mental. Proses penelitian menggunakan python dengan Google Colab. Gambar 1 menunjukkan tahapan pada penelitian.



Gambar 1. Tahapan pada Penelitian

Tahap awal pengumpulan data dilakukan dengan data publik dan mengambil data dari Twitter untuk pengujian. Data dilakukan *data pre-processing* untuk membersihkan dan menyiapkan data teks sebelum masuk ke tahap pemodelan. Tahapan yang dilakukan antara lain:

a. *Case Folding* untuk mengubah semua huruf menjadi huruf kecil.
   Contoh: Saya Suka menulis → saya suka menulis
b. *Text Cleaning* untuk menghapus tanda baca, angka, emotikon, dan simbol yang tidak diperlukan.
   Contoh: "Belajar!! #seru ☺" → "belajar seru"
c. *Tokenization* untuk memisahkan kalimat menjadi kata-kata.
   Contoh: "belajar data seru" → ["belajar", "data", "seru"]
d. *Stopword Removal* untuk menghapus kata-kata umum yang tidak terlalu penting.
   Contoh: "yuk", "dan", "yang"
e. *Lemmatization* untuk Mengubah kata ke bentuk dasarnya.
   Contoh: "berlari", "berlarian" → "lari"
f. *Resampling* untuk menyeimbangkan jumlah data tiap kelas jika jumlahnya tidak seimbang, agar model tidak bias terhadap salah satu kelas.

Setelah *data pre-processing* dilakukan lanjut pada tahap implementasi TF-IDF untuk merubah teks menjadi numerik, selanjutnya pada tahap pemodelan *stacking* menggunakan algoritma *Naïve Bayes*, *Linier SVM*, dan *Random Forest* sebagai *base leaner* dan *Logistic Regression* sebagai *meta leaner*. *Logistic Regression* dipilih karena mampu memproses hasil dari beberapa model dengan cepat dan efisien, struktur yang sederhana dan minim risiko *overfitting*, model ini cocok digunakan sebagai penggabung dalam sistem *stacking* tanpa menambah beban komputasi.

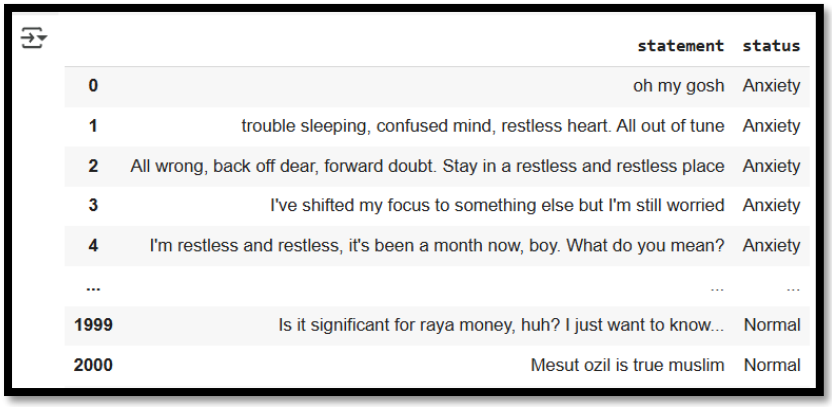
2.1. Pengumpulan Data

Penelitian ini memanfaatkan *dataset* yang diperoleh dari *Dataset Sentiment Analysis for Mental Health* yang bersumber dari data publik pada laman *Kaggle.com https://www.kaggle.com/datasets/suchintikasarkar/sentiment-analysis-for-mental-health* untuk pemodelan dan menggunakan *dataset* yang berasal dari API twitter sebanyak 100 data untuk *data testing*. *Dataset Sentiment Analysis for Mental Health* berjumlah 52.681 data yang terdiri dari tiga atribut yaitu unique_id, *statement*, dan status. *Dataset* merupakan *text* dengan atribut *statement* berupa pernyataan dan status sebagai label berupa status penyakit mental yang berasal dari tag tiap *entri dataset*. Tabel 1 menunjukkan Kategori/Label status pada *dataset* yang terdiri dari 7 kategori penyakit mental sebagai berikut:

Tabel 1. Kategori/Label pada *Dataset*

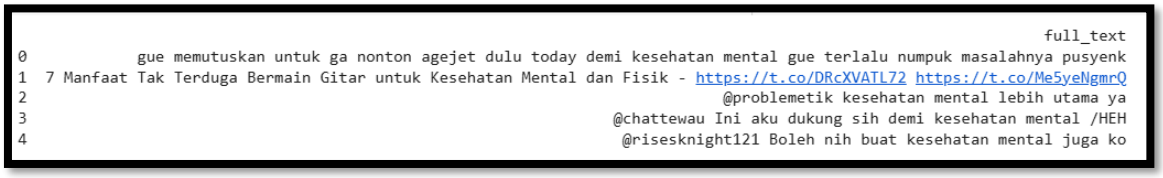
No.	Status	Jumlah
1	Normal	16.343
2	*Depression*	15.404
3	*Suicidal*	10.652
4	*Anxiety*	3.841
5	*Stress*	2.777
6	*Bi-Polar*	2.587
7	*Personality Disorder*	1.077
Total		52.681

Pada Gambar 2 menampilkan *Dataset Sentiment Analysis for Mental Health* dan Gambar 3 menampilkan *Dataset dari API Twitter*.



	statement	status
0	oh my gosh	Anxiety
1	trouble sleeping, confused mind, restless heart. All out of tune	Anxiety
2	All wrong, back off dear, forward doubt. Stay in a restless and restless place	Anxiety
3	I've shifted my focus to something else but I'm still worried	Anxiety
4	I'm restless and restless, it's been a month now, boy. What do you mean?	Anxiety
...	...	...
1999	Is it significant for raya money, huh? I just want to know...	Normal
2000	Mesut ozil is true muslim	Normal

Gambar 2. Tampilan *Dataset Sentiment Analysis for Mental Health*



	full_text
0	gue memutuskan untuk ga nonton agejet dulu today demi kesehatan mental gue terlalu numpuk masalahnya pusyenk
1	7 Manfaat Tak Terduga Bermain Gitar untuk Kesehatan Mental dan Fisik - https://t.co/DRcXVATL72 https://t.co/Me5yeNgmrQ
2	@problemetik kesehatan mental lebih utama ya
3	@chattewau Ini aku dukung sih demi kesehatan mental /HEH
4	@risesknight121 Boleh nih buat kesehatan mental juga ko

Gambar 3. Tampilan *Dataset* dari Twitter

2.2. *Data Preprocessing*

Dalam penelitian ini dilakukan beberapa tahap *data preprocessing* menggunakan bahasa python pada Google Colab yaitu tahap *case folding*, *text cleaning*, *tokenization*, *Stopwords*, *Lemmatization*, dan *Resampling*. *Case folding* merupakan suatu tahap yang digunakan untuk merubah semua teks pada *dataset* menjadi standar berupa huruf kecil (*lowercase*) [7]. *Text cleaning* merupakan tahapan pembersihan data dengan cara menghapus karakter simbol-simbol yang kurang tepat atau tidak relevan seperti menghapus simbol baca, simbol khusus, dan angka [8]. *Tokenization* merupakan tahap yang dilakukan untuk memecah teks kalimat ke dalam bentuk kata [9]. *Stopwords* merupakan kata yang muncul namun tidak memberikan informasi yang penting, *stopwords* digunakan untuk mengurangi informasi tidak penting sehingga memperkecil ukuran data untuk meningkatkan hasil akurasi data [10]. *Lematisation* merupakan teknik mengubah kata menjadi normalisasi bentuk dasar seperti bentuk kamus untuk meningkatkan kinerja dan menjaga fokus makna [11]. Tabel 2 menunjukkan sample proses dari *data preprocessing:*

Tabel 2. *Sample* Proses *Data Preprocessing*

Tahapan	Hasil
Kalimat Asli	I'm restless and restless, it's been a month now, boy. What do you mean?
Case Folding	i'm restless and restless, it's been a month now, boy. what do you mean?
Text Cleaning	im restless and restless it’s been a month now boy what do you mean
Tokenization	['im', 'restless', 'and', 'restless', 'its', 'been', 'a', 'month', 'now', 'boy', 'what', 'do', 'you', 'mean']
Stopwords Removal	['im', 'restless', 'restless', 'month', 'boy', 'mean']
Lemmatization	['im', 'restless', 'restless', 'month', 'boy', 'mean'] (tidak berubah karena semua kata sudah bentuk dasar)
Hasil Akhir	im restless restless month boy mean

Gambar 4 menunjukkan *dataset* sebelum tahap *data preprocessing* dan Gambar 5 menunjukkan *dataset* setelah tahap *data preprocessing*:

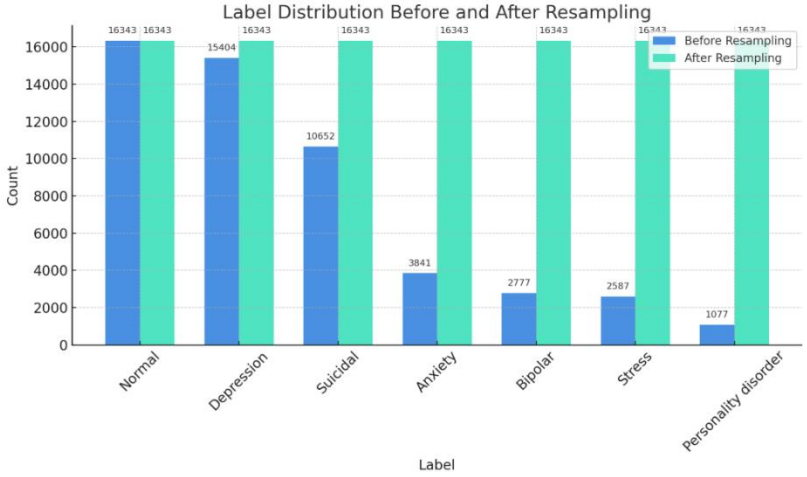
```
Sample data:
                                                                text \
0                                                           oh my gosh
1                trouble sleeping, confused mind, restless heart. All out of tune
2  All wrong, back off dear, forward doubt. Stay in a restless and restless place
3                   I've shifted my focus to something else but I'm still worried
4         I'm restless and restless, it's been a month now, boy. What do you mean?

     label
0  Anxiety
1  Anxiety
2  Anxiety
3  Anxiety
4  Anxiety
Applying text preprocessing...
```

Gambar 4. Tampilan Sebelum Tahap *Data Preprocessing*

```
                                                clean_text
0                                                  oh gosh
1           trouble sleeping confused mind restless heart tune
2  wrong back dear forward doubt stay restless restless place
3            ive shifted focus something else im still worried
4                          im restless restless month boy mean
```

Gambar 5. Tampilan Setelah Tahap *Data Preprocessing*

Dataset Sentiment Analysis for Mental Health memiliki data yang tidak seimbang, sehingga pada tahap *preprocessing* dilakukan tahap *resampling. Resampling* merupakan tahap yang dilakukan untuk menangani masalah ketidakseimbangan data pada kategori sentimen yang tidak merata karena ketidakseimbangan akan mempengaruhi hasil performa dan akurasi suatu model yang digunakan [12]. Gambar 6 menunjukkan grafik tampilan *dataset* sebelum dan sesudah tahap *resampling*.



Gambar 6. Hasil *Resampling* Data

Data uji atau *data testing* dilakukan *translate* bahasa ke bahasa Inggris kemudian dilakukan *pre-processing* data sehingga menghasilkan data sesuai Gambar 7 menunjukkan hasil *preprocessing*:

```
Data Uji:
                                                                clean_text
0                          decided watch agejet first today mental health piled problem pusyenk
1  unexpected benefit playing guitar mental physical health httpstcodrcxvatl httpstcomeyengmrq
2                                                  problemetic mental health important
3                                                  chattewau support mental health heh
4                                                       risesknight make mental health
```

Gambar 7. Hasil *Preprocessing Data Testing*

2.3. Implentasi TF-IDF

Pengertian dari Teknik *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) merupakan proses ekstraksi berguna dalam merubah suatu text dari dokumen ke vektor numerik agar dapat digunakan untuk pemodelan algoritma. Penerapan teknik ini sebelum tahap pemodelan dapat meningkatkan akurasi model klasifikasi. Penggunaan kata-kata untuk dapat mewakili atau dapat mengekstraksi makna dari sebuah teks yang besar disebut dengan Vektorisasi balik [13]. TF-IDF merupakan besaran standar statistik menunjukkan tingkat kontribusi sebuah kata pada sekumpulan dokumen. *Term Frequency* yaitu sebuah intensitas munculnya kata pada sebuah kumpulan dokumen dalam arti bahwa semakin tinggi intensitas suatu kata terdeteksi, maka jumlah nilai TF semakin bertambah besar, sedangkan *Inverse Document Frequency* (IDF) yaitu standar rendahnya intensitas suatu kata pada sebuah kumpulan dokumen dalam arti bahwa semakin rendah intensitas suatu kata terdeteksi, maka jumlah nilai IDF semakin bertambah besar [14].

2.4. Teknik Pemodelan

Pemodelan dalam penelitian ini mengimplementasikan algoritma metode klasifikasi yaitu *Logistic Regression, Naïve Bayes, Linier SVM, Random Forest*, yang dikombinasikan menggunakan metode *Ensemble Stacking*.

a. *Logistic Regression*

Logistic Regression termasuk pada kelompok algoritma klasifikasi dengan metode berupa statistik yang biasanya digunakan untuk keperluan analisis data biner dengan cara menghubungkan model atribut variabel yang independen dengan probabilitas pada sebuah kejadian [15]. *Logistic Regression* merupakan pendekatan analisis statistika yang dapat dimanfaatkan dalam mengatasi suatu masalah pemisahan kelompok dengan cara menghitung seluruh kemungkinan suatu data yang termasuk ke salah satu kelas. Model *Logistic Regression* dikembangkan dari metode *Regresi Linier* dimana *Logistic* memiliki fitur bobot dan nilai bias [16].

b. *Naïve Bayes*

Naïve Bayes merupakan kelompok dari algoritma klasifikasi berasal pada *Teorema Bayes* yang secara sederhana berguna menghitung seluruh probabilitas dengan menggabungkan kombinasi nilai frekuensi dari *dataset*. Metode ini dilakukan dengan memprediksi suatu kasus yang berasal dari klasifikasi yang sebelumnya sudah diperoleh untuk digunakan dalam pengambilan keputusan [17]. *Naïve Bayes* termasuk algoritma sederhana dari kelompok klasifikasi yang setiap fiturnya bersifat independen dan sangat efektif dalam menyelesaikan tugas klasifikasi berupa teks karena menggunakan pendekatan probabilitas berdasarkan frekuensi kemunculan kata [18].

c. *Linier Support Vector Machine (SVM)*

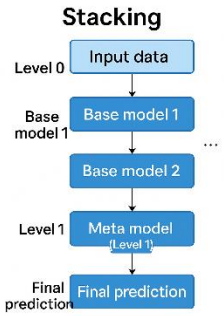
SVM termasuk metode klasifikasi yang optimal dan efisien pada pengelolaan suatu prediksi, algoritma ini memiliki kernel baik *linier* maupun *non-linier*. Linier SVM dapat menghasilkan klasifikasi akurat pada data yang terpisah secara *linier* dengan memisahkan dua kelas dalam fitur dimensi yang tinggi. *Linier* SVM dapat mengatasi pemisahan data yang dilakukan secara *linier* pada variabel bernilai tinggi namun tidak banyak melakukan perubahan [19]. Cara kerja SVM yaitu mencari garis batas atau disebut sebagai *hyperplane* atau yang mampu mewakili nilai selisih dari suatu kumpulan data [20].

d. *Random Forest*

Random Forest termasuk pada algoritma metode klasifikasi menggambarkan *tree* dengan nodenya digunakan untuk meminimalisir adanya *squared-error loss* algoritma ini menggunakan metode CART dengan cara memilih atribut secara acak membentuk pohon keputusan yang kemudian akan membentuk sekumpulan pohon sebagai *forest* [21]. *Random Forest* merupakan metode yang dilakukan dengan cara menggabungkan kumpulan pohon keputusan secara acak untuk membentuk pohon-pohon berupa *forest* untuk meningkatkan kinerja serta keakuratan nilai prediksi [22].

e. *Stacking*

Stacking merupakan salah satu dari bagian metode *Ensemble* yang bekerja dengan cara menggabungkan hasil prediksi dari berbagai model *Machine learning* atau metode klasifikasi untuk menghasilkan model dengan kinerja yang kuat [23]. *Stacking* merupakan bagian dari metode *Ensemble* yang memanfaatkan kombinasi algoritma *Machine learning* berguna menghasilkan prediksi kuat, terdiri dari dua level (level 0 dan 1). Level 0 merupakan *base learner* atau disebut sebagai *single model* dan level 1 merupakan *meta learner*. Pada level 0 merupakan hasil model dari *base learner* yang akan digunakan untuk input data *meta learner* di level 1 [24]. Metode ini digunakan untuk meningkatkan akurasi prediksi melalui penggabungan berbagai model pembelajaran tingkat rendah (*low level learners*), yang dikombinasikan menggunakan algoritma *meta learner* tingkat tinggi (*base metalearner*). *Meta learner* ini berperan dalam menggabungkan seluruh prediksi dari model-model pembelajar tingkat rendah menjadi satu prediksi akhir yang lebih kuat [25]. Berikut Gambar 8. menunjukan gambaran proses metode *Stacking*.



Gambar 8. Gambaran Proses Metode *Stacking*

f. Validasi dan Evaluasi

Penelitian sentiment analisis ini mengimplementasikan proses validasi menggunakan *K-fold cross validation*. *K-fold cross validation* adalah pendekatan informasi data yang dimanfaatkan dalam menguji efektifitas dari suatu model yang diimplementasikan. *K-fold cross validation* akan memisahkan *dataset* terbentuk menjadi set *training* dan set *testing*. Mekanisme validasi mengimplementasikan *K-fold cross validation* yaitu melakukan pelatihan model atau algoritma menggunakan data latih, kemudian melakukan pengujian menggunakan data uji dengan pembagian data sebanyak *k-fold* yang ditentukan [26]. *Cross Validation* dapat mengevaluasi hasil analisis untuk meningkatkan efisiensi komputasi dalam proses pembangunan model algoritma, sekaligus menjaga ketepatan hasil estimasi yang dihasilkan. *K-fold cross validation* data dibentuk menjadi K kelompok data secara seimbang [27]. Penelitian ini menggunakan K sebanyak *5fold* dengan *4fold* sebagai *data training* dan *1fold* sebagai *data testing*. K sebanyak *5fold* merupakan pendekatan yang umum direkomendasikan untuk mengevaluasi kemampuan model dalam melakukan generalisasi terhadap data. Selesai proses pelatihan dan pengujian model, selanjutnya yaitu tahap evaluasi menggunakan tabel *Confusion Matrix*. *Confusion Matrix* termasuk teknik evaluasi dalam algoritma klasifikasi yang disajikan bentuk tabel berguna dalam melakukan perbandingan hasil perkiraan prediksi algoritma dengan nilai fakta pada suatu data. *Confusion Matrix* akan menunjukkan secara terperinci tentang perkiraan data benar atau tidaknya tiap label atau kategori, sehingga memudahkan dalam menilai performa model secara menyeluruh [28]. *Confusion Matrix* akan menampilkan nilai Akurasi, Presisi, Recall, dan *F1-Score*. Tabel 3 menunjukkan *Confusion Matrix* sebagai berikut [29].

Tabel 3. *Confusionn Matrix*

Kelas Prediksi	**Kelas Sebenarnya**	
	Positif	Negatif
Positif	*True Positives (TP)*	*False Positives (FP)*
Negatif	*False Negatives (FN)*	*True Negatives (TN)*

Persamaan *Confusion Matrix* sebagai berikut [29]:

1) Akurasi merupakan perbandingan jumlah seluruh prediksi data yang bernilai benar dengan jumlah total keseluruhan data yang bernilai benar dan bernilai salah.

$$\text{Akurasi: } \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

2) Presisi merupakan perbandingan data yang bernilai benar dengan jumlah total data yang sebenarnya. Presisi menunjukkan keadaan antara prediksi positif dibandingkan dengan kondisi aktual positif berdasarkan data yang telah ditetapkan.

$$\text{Presisi: } \frac{TP}{TP + FP} \quad (2)$$

3) *Recall* merupakan perbandingan nilai data benar dengan total keseluruhan yang bernilai benar positif. *Recall* akan menunjukkan hasil relevan dari kata kunci yang ditetapkan.

$$\text{Recall: } \frac{TP}{TP + FN} \quad (3)$$

4) *F1-Score* merupakan nilai rata rata yang dibandingkan dari nilai *precision* dan nilai *recall* yang sudah diberikan nilai skor. Hasil *F1-Score* lebih bernilai rendah dari pada nilai hasil akurasi.

$$\text{F1-Score: } 2\frac{(\text{Pr}esisi \times \text{Re}call)}{(\text{Pr}esisi + \text{Re}call)} \quad (4)$$

3. HASIL DAN PEMBAHASAN

Penelitian ini menguji beberapa algoritma klasifikasi menggunakan *python* dengan Google Colab. Pengujian pertama mengimplementasikan algoritma tunggal *Logistic Regression*. Gambar 9 menunjukkan hasil dari algoritma *Logistic Regression.*

```
🔎 Cross-validation Results (5-Fold):

🔹 Logistic Regression
Accuracy  : 0.8597
Precision : 0.8569
Recall    : 0.8597
F1        : 0.8574
```

Gambar 9. Hasil Algoritma *Logistic Regression*

Pengujian kedua mengimplementasikan algoritma tunggal *Naïve Bayes*, Gambar 10 menunjukkan hasil dari Algoritma *Naïve Bayes*.

```
🔹 Naive Bayes
Accuracy  : 0.7316
Precision : 0.7306
Recall    : 0.7316
F1        : 0.7270
```

Gambar 10. Hasil Algoritma *Naïve Bayes*

Pengujian ketiga mengimplementasikan algoritma tunggal *Linier* SVM, Gambar 11 menunjukkan hasil dari Algoritma *Linier* SVM.

```
🔹 Linear SVM
Accuracy  : 0.8754
Precision : 0.8712
Recall    : 0.8754
F1        : 0.8720
```

Gambar 11. Hasil Algoritma *Linier SVM*

Pengujian keempat mengimplementasikan algoritma tunggal *Random Forest*. Gambar 12 menunjukkan hasil dari Algoritma *Random Forest*.

```
🔹 Random Forest
Accuracy  : 0.6485
Precision : 0.7022
Recall    : 0.6485
F1        : 0.6490
```

Gambar 12. Hasil Algoritma *Random Forest*

Pengujian kelima menggabungkan semua algoritma klasifikasi dengan menerapkan metode *Ensemble Stacking*. *Base learner* mengimplementasikan algoritma *Naïve Bayes, Linier SVM*, dan *Random Forest* sedangkan *meta learner* menggunakan *Logistic Regression*. Gambar 13 menunjukkan hasil dari penerapan metode *Stacking*.

```
🔹 Stacking Ensemble (NB + SVM + RF → Logistic Regression)
Accuracy  : 0.8813
Precision : 0.8788
Recall    : 0.8813
F1        : 0.8797
```

Gambar 13. Hasil Penerapan *Stacking*

Penelitian dilakukan sebanyak lima kali dengan menguji *Dataset Sentiment Analysis for Mental Health*. Tabel 4 menunjukkan hasil perbandingan dari pengujian yang diimplementasikan.

Tabel 4. Hasil Perbandingan Pengujian

Algoritma	***Accuracy***	***Precision***	***Recall***	***F1-Score***
Logistic Regression	85.97%	85.69%	85.97%	85.74%
Naïve Bayes	73.16%	73.06%	73.16%	72.70%
Linier SVM	87.54%	87.12%	87.54%	87.20%
Random Forest	64.85%	70.22%	64.85%	64.90%
Stacking	88.13%	87.88%	88.13%	87.97%

Pengujian keenam dilakukan dengan menguji model menggunakan *data testing* yang diambil dari twitter, sebanyak 100 data. Gambar 14 menunjukkan hasil uji dari *data testing* dan hasil pengujian menggunakan *data testing* dari Twitter ditunjukkan pada Tabel 5.

```
🚀 Evaluating on Test Set: Stacking Ensemble
                    precision    recall  f1-score   support

           Anxiety       1.00      1.00      1.00         8
        Depression       1.00      1.00      1.00         4
            Normal       1.00      1.00      1.00        80
Personality disorder       1.00      1.00      1.00         1
            Stress       1.00      1.00      1.00         6
          Suicidal       1.00      1.00      1.00         1

          accuracy                           1.00       100
         macro avg       1.00      1.00      1.00       100
      weighted avg       1.00      1.00      1.00       100
```

Gambar 14. Hasil Uji *Data testing*

Tabel 5. Hasil Perbandingan Pengujian Menggunakan *Data testing Twitter*

Algoritma	***Accuracy***	***Precision***	***Recall***	***F1-Score***
Logistic Regression	82%	91%	82%	85%

Naïve Bayes	24%	70%	24%	28%
Linier SVM	84%	90%	84%	87%
Random Forest	81%	68%	81%	74%
Stacking	100%	100%	100%	100%

Penelitian menggunakan *data testing* dari Twitter menghasilkan nilai akurasi 100 % pada metode *stacking*. Namun, perlu dicermati bahwa jumlah data uji yang digunakan relatif kecil hanya 100 data dan memiliki karakteristik yang hampir sama, sehingga model lebih mudah mengenali pola dalam data dan menghasilkan performa tinggi. Oleh sebab itu, uji lanjutan dengan *dataset* yang lebih besar dan beragam sangat diperlukan guna menguji konsistensi dan kemampuan generalisasi model secara menyeluruh.

Berdasarkan dari pengujian pada penelitian ini menunjukkan bahwa terbukti terjadi suatu peningkatan menggunakan penggabungan algoritma klasifikasi dengan penerapan metode *Ensemble Stacking*. Penelitian ini memperoleh hasil terbaik dengan akurasi sebesar 88.13%. Hasil tersebut diperoleh dari penerapan *stacking* yang menggabungkan seluruh algoritma klasifikasi. Hasil membuktikan adanya peningkatan dibandingkan menggunakan algoritma tunggal.

4. KESIMPULAN DAN SARAN

Penelitian ini mendapatkan hasil terbaik dengan penerapan metode *Ensemble Stacking* menghasilkan hasil akurasi terbaik yaitu sebesar 88.13%. Hasil tersebut meningkat dibanding menggunakan algoritma tunggal. Model diuji menggunakan *data testing* twitter menghasilkan akurasi baik mencapai 100%. Namun demikian, hasil akurasi 100% pada data Twitter perlu ditinjau secara kritis karena jumlah data yang terbatas dan kemungkinan karakteristik data yang homogen. Implikasi dari peningkatan akurasi ini menunjukkan potensi sistem untuk mendukung deteksi dini kesehatan mental secara lebih akurat dalam praktik nyata. Pengujian yang akan datang dapat menggunakan dataset lain atau metode algoritma klasifikasi lain untuk perbandingan keakurasian data. Hasil penelitian dapat diimplementasikan menjadi sistem aplikasi diagnosa kesehatan mental untuk deteksi dini kesehatan mental. Pengembangan teknologi juga perlu dilakukan untuk menggunakan metode yang menghasilkan akurasi paling terbaik.

UCAPAN TERIMA KASIH

Kami mengucapkan terima kasih kepada bagian DPPM Universitas Pelita Bangsa yang telah memberikan bantuan pendanaan dengan demikian penelitian dapat terwujud secara optimal. Penghargaan yang sama kami sampaikan juga kepada berbagai pihak yang turut berkontribusi dalam bentuk dukungan secara eksplisit maupun implisit, dalam proses pelaksanaan hingga selesainya penelitian ini.

DAFTAR PUSTAKA

[1] Kementrian Kesehatan Republik Indonesia, "Hasil Riskesdas 2013," *Expert Opin Investig Drugs*, vol. 7, no. 5, pp. 803–809, 2013, doi: 10.1517/13543784.7.5.803.

[2] Kemenkes, "Survei Kesehatan Indonesia 2023 (SKI)," *Kemenkes*, p. 235, 2023.

[3] N. Nurhayati *et al.*, "Analisis Sentimen Komentar Youtube Terhadap Isu Kesehatan Mental Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbor (KNN)," *Syntax: Computer Science and Information Technology*, no. 6, pp. 8–16, 2025, doi: DOI:10.46576/syntax.v6i1.6060.

[4] J. R. Wiyani, I. Indriati, and S. Sutrisno, "Klasifikasi Stres berdasarkan Unggahan pada Media Sosial Twitter menggunakan Metode Support Vector Machine dan Seleksi Fitur Information Gain," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 12, pp. 6003–6009, 2022.

[5] K. D. Odja *et al.*, "Mental illness detection using sentiment analysis in social media," in *Procedia Computer Science*, Elsevier B.V., 2024, pp. 971–978. doi: 10.1016/j.procs.2024.10.325.

[6] M. Langgeng Wicaksono, R. Rusdah, and D. Apriana, "Analisis Sentimen Kesehatan Mental Menggunakan K-Nearest Neighbors pada Sosial Media Twitter," *Bit (Fakultas Teknologi Informasi Universitas Budi Luhur)*, vol. 19, no. 2, pp. 98–103, 2022, doi: 10.36080/bit.v19i2.2042.
[7] K. Rahayu *et al.*, "Klasifikasi Teks untuk Mendeteksi Depresi dan Kecemasan pada Pengguna Twitter Berbasis Machine Learning," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 108–114, 2023, doi: 10.57152/malcom.v3i2.780.
[8] A. S. D. P. Sinaga and A. S. Aji, "Analisis Sentimen Publik Terhadap Mayor Teddy Indra Wijaya dengan Pendekatan Logistic Regression," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. January, pp. 222–231, 2025, doi: https://doi.org/10.57152/malcom.v5i1.1752.
[9] D. R. F. Daud, B. Irawan, and A. Bahtiar, "Penerapan Metode Naive Bayes Pada Analisis Sentimen Aplikasi Mcdonalds Di Google Play Store," *JATI: Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 1, pp. 759–766, 2024, doi: 10.36040/jati.v8i1.8784.
[10] A. Karim *et al.*, "Anticipating impression using textual sentiment based on ensemble LRD model," *Expert Systems with Applications*, vol. 263, no. August 2023, pp. 1-15, 2025, doi: https://doi.org/10.1016/j.eswa.2024.125717.
[11] K. Aziz *et al.*, "Enhanced UrduAspectNet: Leveraging Biaffine Attention for superior Aspect-Based Sentiment Analysis," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 9, p. 102221, 2024, doi: 10.1016/j.jksuci.2024.102221.
[12] C. Chulyatunni'mah R. Kurniawan and S. Anwar, "Analisis Sentimen Penggemar Treasure Di Karnaval Mandiri Menggunakan Naïve Bayes," *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, vol. 4, pp. 203–213, 2025, doi: https://doi.org/10.53513/jursi.v4i1.10606.
[13] L. Afuan, M. Khanza, and A. Z. Hasyati, "Peningkatan Analisis Sentimen Pelantikan Presiden Ri Tahun 2024 Pada X Menggunakan Naive Bayes Classifier Yang Dioptimalkan SMOTE," *Jurnal Teknik Informatika (JUTIF)*, vol. 6, no. 1, pp. 325–333, 2025, DOI: https://doi.org/10.52436/1.jutif.2025.6.1.4290.
[14] S. Syafrizal, M. Afdal, and R. Novita, "Analisis Sentimen Ulasan Aplikasi PLN Mobile Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 10–19, 2023, doi: 10.57152/malcom.v4i1.983.
[15] Y. Handayani *et al.*, "Perbandingan Algoritma Logistic Regression Dan Naïve BAYES Classifier dalam Identifikasi Penyakit Liver," *Journal of Science and Social Research*, vol. 8, no. 2, pp. 1435–1440, 2025, https://doi.org/10.54314/jssr.v8i2.2892.
[16] V. Vajrobol, B. B. Gupta, and A. Gaurav, "Mutual information based logistic regression for phishing URL detection," *Cyber Security and Applications*, vol. 2, no. February, 2024, doi: 10.1016/j.csa.2024.100044.
[17] W. Ningsih *et al.*, "Perbandingan Algoritma SVM dan Naïve Bayes dalam Analisis Sentimen Twitter pada Penggunaan Mobil Listrik di Indonesia," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 556–562, 2024, doi: 10.57152/malcom.v4i2.1253.
[18] T. Anderson, S. Sarkar, and R. Kelley, "Analyzing public sentiment on sustainability: A comprehensive review and application of sentiment analysis techniques," *Natural Language Processing Journal*, vol. 8, no. June, p. 100097, 2024, doi: 10.1016/j.nlp.2024.100097.
[19] S. D. Wahyuni and R. H. Kusumodestoni, "Optimalisasi Algoritma Support Vector Machine (SVM) Dalam Klasifikasi Kejadian Data Stunting," *Bulletin of Information Technology (BIT)*, vol. 5, no. 2, pp. 56–64, 2024, doi: 10.47065/bit.v5i2.1247.
[20] S. Nauli *et al.*, "Klasifikasi Kalimat Perundungan pada Twitter Menggunakan Algoritma Support Vector Machine," *JIPI: Jurnal Ilmiah Penelitian dan Pembelajaran Informatika*, vol. 10, no. 1, pp. 107–122, 2025, https://doi.org/10.29100/jipi.v10i1.5749.

[21] P. K. Sari, R. R. Suryono. "Komparasi Algoritma Support Vector Machine dan Random Forest Untuk Analisis Sentimen Metaverse," *Jurnal MNEMONIC,* vol. 7, no. 1, pp. 31–39, 2024.
[22] S. Nurohanisah, R. Astuti, and F. Muhammad Basysyar, "Deteksi Berita Palsu Menggunakan Algoritma Random Forest," *JATI: Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 1, pp. 422–428, 2024, doi: 10.36040/jati.v8i1.8418.
[23] S. Sudarto, and K. Kusrini, "Klasifikasi Tsunami Gempa Bumi dengan Teknik Stacking Ensemble Machine Learning,"*Jurnal Informatika Polinema*, vol. 10. No.4 2024 doi:10.33795/jip.v10i4.5655.
[24] J. Prasetya, S. I. Fallo, and M. A. Aprihartha, "*Stacking* Machine Learning Model for Predict Hotel Booking Cancellations," *Jurnal Matematika, Statistika dan Komputasi*, vol. 20, no. 3, pp. 525–537, 2024, doi: 10.20956/j.v20i3.32619.
[25] A. Daza, *et al.*, "*Stacking* Ensemble Approach to Diagnosing the Disease of Diabetes," *Informatics in Medicine Unlocked*, vol. 44, no. December 2023, 2024, doi: 10.1016/j.imu.2023.101427.
[26] Y. N. Fuadah, *et al.*, "Optimasi Convolutional Neural Network dan K-Fold Cross Validation pada Sistem Klasifikasi Glaukoma," *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, vol. 10, no. 3, pp. 728-741, 2022, doi: 10.26760/elkomika.v10i3.728.
[27] W. A. Firmansyach, U. Hayati, and Y. Arie Wijaya, "Analisa Terjadinya Overfitting Dan Underfitting Pada Algoritma Naive Bayes Dan Decision Tree Dengan Teknik Cross Validation," *JATI: Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 1, pp. 262–269, 2023, doi: 10.36040/jati.v7i1.6329.
[28] F. R. Valerian *et al.*, "Klasifikasi Tingkat Obesitas Menggunakan Metode GBM dan Confusion Matrix," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 2, pp. 2242–2249, 2025.
[29] Y. N. Rumbia, "Perbandingan Metode KNN dan Naive Bayes untuk Klasifikasi Kelulusan Mahasiswa pada Mata Kuliah Probstat," *Jurnal PTI: Jurnal Pendidikan Teknologi Informasi*, vol. 12, pp. 7–13, 2025, doi: 10.35134/jpti.v12i1.228.