JURNAL FASILKOM
Volume 15 No. 1 | April 2025: 84-92

P-ISSN : 2089-3353
E-ISSN : 2808-9162

Deteksi Stres dan Depresi Unggahan Media Sosial dengan Machine Learning

Yefta Tolla[1], Kusrini[2]
[1,2]Informatika, Ilmu Komputer, Universitas Amikom Yogyakarta
[1]yeftatolla@students.amikom.ac.id*, [2]kusrini@amikom.ac.id

Abstract

This study presents a machine learning model that integrates Support Vector Machine (SVM), Natural Language Processing (NLP), and SHapley Additive exPlanations (SHAP) to detect stress and depression from Twitter posts. The dataset comprises posts collected over the past year, labeled into stress, depression, and neutral categories. NLP techniques—tokenization, stopword removal, stemming, and Term Frequency-Inverse Document Frequency (TF-IDF)—were applied to clean the text and extract relevant features. These steps ensure only meaningful words influence classification, improving predictive performance. SVM serves as the main classifier due to its effectiveness in high-dimensional text data. SHAP enhances interpretability by highlighting influential features that drive predictions, increasing transparency. Results show that words like "stress," "exhausted," and "anxious" strongly indicate stress, while "depression," "disappointed," and "giving up" signal depressive tendencies. Phrases like "mentally exhausted" further support the identification process. The model achieved 96.44% accuracy, along with high precision, recall, and F1-scores (averaging 96%). A confusion matrix confirmed the model's effectiveness in classifying the three categories with minimal error. The integration of NLP, SVM, and SHAP not only improves classification accuracy but also enhances explainability, making the model a promising tool for early detection and understanding of mental health conditions through social media analysis.

Keywords: *depression, stress, SHAP, SVM, NLP*

Abstrak

Penelitian ini menyajikan sebuah model pembelajaran mesin yang mengintegrasikan *Support Vector Machine* (SVM), *Natural Language Processing* (NLP), dan *SHapley Additive exPlanations* (SHAP) untuk mendeteksi stres dan depresi dari unggahan di Twitter. *Dataset* ini terdiri dari postingan yang dikumpulkan selama satu tahun terakhir, yang dilabeli ke dalam kategori stres, depresi, dan netral. Teknik *NLP-tokenization*, penghilangan *stopword*, *stemming*, dan *Term Frequency-Inverse Document Frequency* (TF-IDF)-diterapkan untuk membersihkan teks dan mengekstrak fitur-fitur yang relevan. Langkah-langkah ini memastikan hanya kata-kata yang bermakna yang memengaruhi klasifikasi, sehingga meningkatkan kinerja prediktif. SVM berfungsi sebagai pengklasifikasi utama karena keefektifannya dalam data teks berdimensi tinggi. SHAP meningkatkan kemampuan interpretasi dengan menyoroti fitur-fitur berpengaruh yang mendorong prediksi, sehingga meningkatkan transparansi. Hasil penelitian menunjukkan bahwa kata-kata seperti "stres", "lelah", dan "cemas" sangat mengindikasikan stres, sedangkan "depresi", "kecewa", dan "menyerah" menandakan kecenderungan depresi. Frasa seperti "lelah secara mental" semakin mendukung proses identifikasi. Model ini mencapai akurasi 96,44%, dengan presisi, recall, dan skor F1 yang tinggi (rata-rata 96%). Matriks kebingungan mengkonfirmasi keefektifan model dalam mengklasifikasikan tiga kategori dengan kesalahan minimal. Integrasi NLP, SVM, dan SHAP tidak hanya meningkatkan akurasi klasifikasi, tetapi juga meningkatkan penjelasan, menjadikan model ini alat yang menjanjikan untuk deteksi dini dan pemahaman kondisi kesehatan mental melalui analisis media sosial.

Kata kunci: depresi, stress, SHAP, SVM, NLP

1. Pendahuluan

Di Indonesia, masalah kesehatan mental seperti stres dan depresi telah menjadi semakin memprihatinkan dan menjadi salah satu masalah yang serius. Menurut laporan Survei Kesehatan Indonesia (SKI) oleh Kementerian Kesehatan, prevalensi depresi di Indonesia adalah 1,4% pada tahun 2023. Jika dilihat berdasarkan kelompok usia, prevalensi depresi tertinggi dialami oleh mereka yang berusia 15-24 tahun, atau Generasi Z, yaitu sebesar 2%. Angka prevalensi tinggi lainnya ditemukan pada kelompok lansia berusia 75 tahun ke atas sebesar 1,9%, diikuti kelompok usia 65-74 tahun sebesar 1,6%, 25-34 tahun sebesar 1,3%, 55-64 tahun sebesar 1,2%, dan 45-54 tahun sebesar 1,1%. Kelompok usia dengan prevalensi depresi nasional terendah adalah 35-44 tahun, yaitu 1%. Meskipun generasi muda memiliki prevalensi depresi tertinggi, hanya 10,4% yang mencari pengobatan. Kementerian Kesehatan (Kemenkes) menyatakan bahwa depresi yang tidak diobati pada Gen Z akan menyebabkan masalah sosial yang lebih tinggi, seperti memburuknya penyakit, bunuh diri, penyalahgunaan obat-obatan, dan lain-lain. Stres dan depresi yang tidak terdeteksi dan tidak mendapatkan penanganan yang tepat sering kali disebabkan oleh kurangnya kesadaran dan stigma negatif dalam masyarakat. Seperti yang terlihat pada upaya pencegahan yang berhasil dilakukan pada penelitian [1], hasil deteksi yang dilakukan pada mahasiswa

P-ISSN : 2089-3353
E-ISSN : 2808-9162

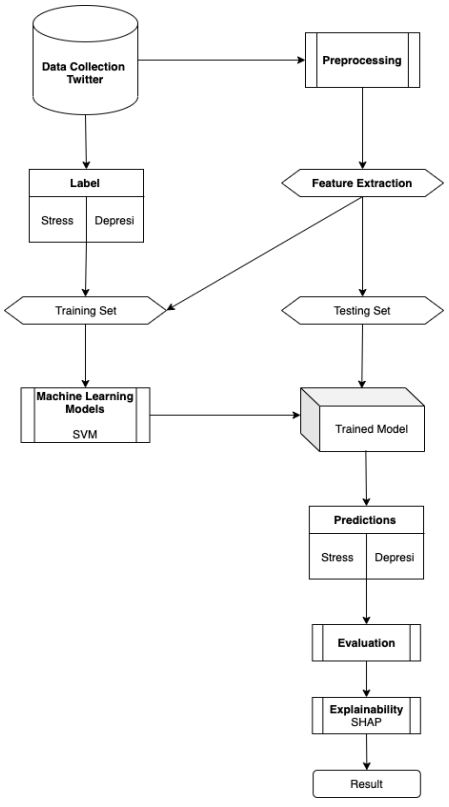
digunakan oleh institusi untuk merancang program-program yang mendukung kesehatan mental mahasiswa, serta mengembangkan strategi pencegahan dan intervensi yang efektif. Penelitian tersebut secara efektif mengenali gejala yang dirasakan atau terlihat dalam upaya mencegah depresi pada ibu hamil, sehingga mereka dapat mengambil langkah tindak lanjut yang tepat dan menerima perawatan jika ditemukan depresi [2]. Postingan media sosial dapat dianalisis menggunakan machine learning dan mendeteksi pola perilaku atau kata-kata yang mengindikasikan kemungkinan pengguna mengalami stres atau depresi [3]. Ada banyak faktor yang dapat menyebabkan seseorang menghadapi masalah kesehatan mental, seringkali karena tekanan pendidikan, pekerjaan, trauma akibat kehilangan atau perceraian, atau bahkan bencana [4,5,6]. Dengan memanfaatkan data yang dikumpulkan dari aktivitas online, data tersebut dapat digunakan untuk menganalisis sentimen dalam postingan atau komentar di media sosial. Hal ini membantu dalam memahami perasaan dan emosi yang mungkin terkait dengan masalah kesehatan mental [7]. Penggunaan teknologi bukan untuk menggantikan peran tenaga profesional kesehatan mental, tetapi merupakan alat yang dapat meningkatkan akses, daya tanggap, dan efektivitas layanan yang tersedia.

Penggunaan pembelajaran mesin untuk mengidentifikasi stres dan depresi dari unggahan media sosial telah menjadi subjek dari berbagai penelitian. Penelitian [8] dalam penelitianya tidak menyebutkan apakah model yang digunakan bisa memberikan penjelasan mengenai factor yang dapat membuat *tweet* tertentu diklasifikasikan sebagai depresi atau cemas, sulit untuk memahami apakah hasil ini akurat atau hanya karena pola tertentu dalam data tanpa adanya interpretasi yang jelas. Penelitian [9] menunjukkan bahwa perlu menambahkan parameter yang lebih rinci dalam melakukan anlisis kemungkinan depresi pengguna dengan informasi lebih dalam dari tenaga ahli. Penelitian lain [10] menggunakan model berbasis transformer, seperti BERT dan MentalBERT, yang diperkaya dengan fitur linguistik (LIWC). Meskipun model ini meningkatkan kinerja klasifikasi, interpretasi hasil tetap menjadi tantangan karena penelitian ini tidak memberikan banyak informasi mengenai fitur apa yang paling berpengaruh dalam klasifikasi, sebagaimana transformer base model seperti BERT dikenal sebagai model yang sulit diinterpretasikan. Penelitian [11] mengimplementasikan berbagai jenis klasifikasi mesin, namun menemukan bahwa menggunakan algoritma tersebut secara individual cenderung mengarah pada masalah *overfitting*. Penelitian [12] membandingkan *Neural Network* dan SVM dalam mendeteksi depresi dari Twitter. Hasil *Neural Network* menunjukkan hasil yang baik dalam *precision* dan *recall* menunjukkan bahwa model ini masih memiliki ruang untuk perbaikan. *Recall* yang rendah menunjukkan bahwa model mungkin keliru dan melewatkan banyak kasus depresi. Penelitian [13] menguji teknik NLP, seperti ELMo *embeddings* dan BERT *tokenizer*, dalam klasifikasi tekanan mental pada data *Reddit*, dengan F1-score sebesar 0.76. Penelitian ini tidak menjelaskan tentang fitur spesifik yang digunakan dalam model dan bagaimana kontribusi fitur terhadap klasifikasi.

Dari penelitian-penelitian tersebut, dapat dilihat bahwa sebagian besar masih berfokus pada peningkatan akurasi model tanpa memberikan perhatian yang cukup terhadap interpretabilitas hasil klasifikasi. Selain itu, belum banyak studi yang menggabungkan pendekatan *explainable AI* seperti SHAP (Shapley Additive exPlanations) untuk memahami kontribusi fitur terhadap hasil klasifikasi, terutama dalam konteks deteksi stres dan depresi. Oleh karena itu, penelitian ini mengisi celah tersebut dengan menerapkan teknik SHAP pada model SVM, sehingga tidak hanya menghasilkan klasifikasi yang akurat, tetapi juga dapat dijelaskan secara transparan. Kebaruan dari penelitian ini terletak pada integrasi teknik SHAP dalam model SVM serta fokus pada interpretabilitas, peningkatan recall, dan generalisasi model untuk deteksi stres dan depresi yang lebih akurat dan aplikatif.

2. Metode Penelitian

Bagian ini menguraikan langkah-langkah yang dilakukan secara terstruktur dan sistematis untuk mencapai tujuan yang telah ditetapkan. Penelitian ini dirancang untuk menganalisis dan mengembangkan model pendeteksi stres dan depresi dengan menggunakan algoritma SVM [14]. Dengan alur penelitian sebagai berikut :



Gambar 1. Alur Penelitian

P-ISSN : 2089-3353
E-ISSN : 2808-9162

2.1. Pengolahan Data

a. *Tools* dan *Library*
Penelitian ini menggunakan bahasa pemrograman Python dengan beberapa library utama, Tweepy untuk pengambilan data dari Twitter, *Scikit-learn* (sklearn) untuk implementasi model dan evaluasi performa, SHAP untuk interpretabilitas model, *NumPy* dan Pandas untuk manipulasi dan pengolahan data, Matplotlib untuk visualisasi grafik.

b. Pengumpulan Data
Penelitian ini mengumpulkan data dari *platform* media sosial X (Twitter), yang dipilih karena X menyediakan API (*Application Programming Interface*) yang memungkinkan peneliti untuk mengumpulkan data publik dan legal. Pengguna Twitter umumnya menulis menggunakan bahasa informal yang mencerminkan emosi dan kondisi mental secara spontan. Ini membantu dalam menangkap sinyal stres atau depresi secara lebih otentik. Twitter termasuk salah satu media sosial yang cukup populer di Indonesia, sehingga data yang diperoleh akan cukup representatif terhadap kondisi populasi yang diteliti. Penelitian yang dilakukan oleh Situmorang dan Purba [15] menemukan bahwa salah satu ciri tweet yang dapat mengindikasikan stres atau depresi adalah mengandung kata-kata negatif, yang dapat menunjukkan bahwa pengguna sedang mengalami emosi negatif. Untuk memastikan data yang dikumpulkan lebih relevan, proses pengumpulan kata kunci juga melibatkan psikolog untuk menentukan kata-kata yang biasa digunakan oleh individu yang mungkin mengalami stres atau depresi. Daftar kata kunci dapat dilihat pada Tabel 1.

Tabel 1.Tabel Kata Kunci

	Kata Kunci
Berpotensi Stress	Stres, Capek mental, Cemas, Panik, Khawatir berlebihan, Paranoid, Nggak fokus, Deadlines, Exhausted / Kelelahan
Berpotensi Depresi	Depresi, Bersalah, Diabaikan, Frustasi, Gagal, Galau, Gelisah, Hilang kendali, Hampa, Kecewa, Kesepian, Menyerah, Menyesal, Pasrah, Putus asa, Sedih, Sendirian, Tak berharga, Terasing,, Terisolasi, Terluka, Terpuruk, Tertekan, Tersisih, Tidak berarti, Tidak berdaya, Tidak berguna, Tidak didengar, Tidak mampu
Normal	Antusias, Bahagia, Berharga, Berhasil, Bersama, Ceria, Dekat, Dihargai, Dipercaya, Diterima, Gembira, Mampu, Optimis, Positif, Riang, Semangat, Senang, Sukses, Tenang

Pada tabel 1 kata kunci untuk Stres sebanyak 9, kata kunci untuk Depresi sebanyak 29, dan kata kunci untuk normal sebanyak 19 sehingga total data sebanyak 57.

2.2. Preprocessing

a. Natural Language Processing
Natural Language Processing (NLP) menganalisis, memahami, dan menghasilkan bahasa seperti yang dilakukan manusia. NLP mengajarkan mesin atau komputer untuk mempelajari bagaimana manusia menggunakan bahasa [16]. Dalam penelitian ini, NLP memungkinkan model untuk bekerja dengan fitur teks yang lebih bermakna daripada menggunakan teks mentah dari unggahan media sosial. Hal ini dapat meningkatkan akurasi dan relevansi model dalam mendeteksi stres dan depresi karena fitur-fitur yang diekstrak dari NLP lebih mewakili pola emosional atau psikologis yang sebenarnya [17]. Preprocessing dilakukan untuk menstrukturkan data dan memudahkan proses pemodelan [18]. Langkah-langkah yang dilakukan meliputi lima tahap yaitu text cleaning untuk menghilangkan angka, pemisah seperti koma, titik, dan tanda baca lainnya, case folding untuk mengubah teks ke dalam format standar atau huruf kecil, stopword removal untuk menghilangkan kata-kata yang tidak relevan atau tidak bermakna, tokenizing untuk membagi string input dengan spasi, untuk menemukan bentuk dasar kata setelah stopword removal [19].

b. Pemisahan dan Pelabelan Data
Pemisahan data dilakukan untuk memisahkan dataset menjadi data uji dan data latih, dengan pembagian 80:20, dan diberi label stres, depresi, dan normal. Proses pelabelan dalam penelitian ini dilakukan dengan bantuan pakar atau psikolog yang menentukan kata kunci yang berpotensi menunjukkan stres, depresi, atau kondisi normal. Kata kunci ini kemudian digunakan untuk mengelompokkan data dari unggahan Twitter secara otomatis.

2.3. Ekstraksi Fitur

Ekstraksi fitur adalah langkah penting dalam klasifikasi teks. Metode yang sadar konteks, seperti perluasan tweet, telah terbukti meningkatkan kinerja analisis sentimen dengan memperkaya teks dengan frasa yang relevan [20].

2.4. Model Pembelajaran Mesin

a. *Support Vector Machine*
Model algoritma pembelajaran yang digunakan untuk mendeteksi stres dan depresi menggunakan teknik Support Vector Machine, yang dilatih menggunakan kernel linear. Parameter utama yang digunakan adalah: C =1.0, parameter regularisasi yang mengontrol trade-off antara kesalahan klasifikasi dan margin. Kernel "linear", menentukan jenis kernel yang digunakan untuk pemisahan data. Kernel linear dipilih karena cocok untuk data yang dapat dipisahkan secara linier.

Support Vector Machine (SVM) beroperasi dengan menggunakan prinsip *Structural Risk Minimization* (SRM). Metode pembelajaran mesin ini bertujuan untuk menemukan hyperplane terbaik yang dapat memisahkan dua kelas yang berbeda [21]. Algoritma klasifikasi ini digunakan untuk klasifikasi dan regresi dalam penelitian untuk mendeteksi stres dan depresi dari unggahan media sosial. Algoritma SVM diterapkan karena dapat bekerja secara efektif dalam mengklasifikasikan data dengan jumlah fitur yang banyak [22], seperti data teks dari postingan media sosial.

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i \; K(x_i, x) + b \qquad (1)$$

P-ISSN : 2089-3353
E-ISSN : 2808-9162

Ket :
α_i : bobot yang dihasilkan selama pelatihan
K (x_i, x) : fungsi kernel yang mengukur kesamaan antara data baru x dan data pelatihan x_i.
Prediksi akhir tergantung pada tanda dari fungsi :
Jika $f(x) > 0$, maka data baru diklasifikasikan sebagai salah satu kelas (misalnya depresi)
Jika $f(x) < 0$, maka data baru diklasifikasikan sebagai kelas lain (misalnya stres).

c. *Training* Model
Model ini mengalami proses pelatihan di mana ia diajarkan untuk mengenali pola dan membuat prediksi, sehingga memungkinkan sistem untuk membuat keputusan otomatis berdasarkan pola yang dipelajari dari data.

2.5. Evaluasi

a. Pengukuran Kinerja Model
Evaluasi model sangat penting untuk menilai performa dan interpretabilitas dalam aplikasi *machine learning*, terutama jika diterapkan pada NLP [23]. Dalam penelitian ini, evaluasi performa model dilakukan dengan menggunakan beberapa metrik utama, yaitu akurasi, *precision*, *recall*, dan f1-*score*. Mengingat bahwa penelitian ini berfokus pada klasifikasi tiga kelas (stres, depresi, dan normal), maka penting untuk menggunakan metrik yang dapat mengevaluasi performa model secara menyeluruh di setiap kelas. Salah satu metrik utama yang digunakan adalah Macro-F1 Score. Alasan pemilihannya adalah karena Macro-F1 menghitung rata-rata F1-*score* dari masing-masing kelas secara setara tanpa memperhatikan proporsi jumlah data di setiap kelas. Juga dilakukan perbandingan dengan model SVM yang menggunakan SHAP dan tanpa SHAP. Model akan dievaluasi untuk mengukur akurasi dan interpretabilitas untuk meningkatkan prediksi dan memahami fitur yang paling berpengaruh dalam deteksi stres dan depresi dengan SVM.

b. Validasi Model
Setelah model dibuat, akan dilakukan evaluasi untuk melihat kinerja model dengan *Cross Validation*. Ini membantu menilai seberapa baik model dapat digeneralisasi ke data yang belum pernah ada sebelumnya. *Cross Validation* juga digunakan untuk memastikan bahwa model tidak *overfitting* dan dapat bekerja dengan baik pada dataset yang lebih luas dan beragam.

Juga digunakan *Confusion matrix* dalam penelitian ini bertujuan untuk mengevaluasi kinerja model klasifikasi dalam mendeteksi stres, depresi, dan kondisi normal dari unggahan media sosial. *Confusion matrix* memberikan gambaran menyeluruh mengenai jumlah prediksi yang benar dan salah untuk setiap kelas. Dengan demikian, *confusion matrix* menjadi alat yang penting dalam mengukur efektivitas dan ketepatan model dalam klasifikasi multi-kelas.

2.6. *SHapley Additive exPlanations*

Menerapkan teknik SHAP (*SHapley Additive exPlanations*) pada model untuk mendapatkan interpretabilitas dari setiap prediksi yang dibuat [24]. Untuk memahami bagaimana setiap atribut berkontribusi pada pembuatan prediksi, *SHapley Additive exPlanations* (SHAP) digunakan. SHAP dapat mengidentifikasi fitur mana yang paling penting untuk prediksi yang dibuat oleh model SVM dan seberapa besar dampaknya terhadap keluaran model pembelajaran mesin [24,25].

$$\phi_i(f) = \sum_{S \subseteq N \setminus \{i\}} \left(\frac{|S|! \,.(|N|-|S|-1)!}{|N|!} \right) \left(f(S \cup \{i\}) - f(S) \right) \quad (2)$$

$\phi_i(f)$: Nilai SHAP untuk fitur i
S : Subset dari fitur yang tidak termasuk fitru i
N : Semua fitur
$f(S)$: Prediksi model berdasarkan subset fitur S

3. Hasil dan Pembahasan

3.1. *Dataset*

Data dikumpulkan dengan menggunakan kata kunci yang telah ditentukan, dengan postingan yang dibuat dari Januari 2024 hingga Desember 2024. Dari proses perayapan data, 3.935 *tweet* diperoleh dan digunakan untuk penelitian ini. Sampel data ditunjukkan pada Tabel 2.

Tabel 2. Data

No	Tweet
1	baru kali ini gw ngerasain yg namanya anxiety perkara kurang tidur. rasanya cemas bingung ngerasa bersalah dan sedih. semoga tidak berkepanjangan
2	sekarang aku lagi di fase yang lucas rasakan yaitu Depresi. Rasanya hampa banget gak ada semangat
3	Aku tak tau harus apa dan bagaimana. Aku senang melihat mu baik saja dan bahagia. Dgn mu atau tdk nyatanya hidup ku masih baik saja. Yang jelas untuk saat ini sukses dlm karir dan menjadi wanita yg lebih baik lagi itu tujuan ku. Perihal tentang kita ku pasrahkan dgn Tuhanku.
...	……
3.935	kesepian

3.2. Data *Processing*

Pembersihan data dilakukan pada teks mentah yang diambil dari media sosial. Untuk memastikan data benar-benar siap untuk digunakan, prosedur ini memerlukan beberapa proses, seperti pembersihan teks, *case folding*, penghilangan *stopword*, dan *tokenizing*. Proses ini memastikan teks yang digunakan lebih bersih, terstruktur, dan siap untuk dianalisis oleh model machine learning, yaitu model SVM yang telah dibangun, yang nantinya juga akan diuji interpretabilitasnya dengan teknik SHAP. Berikut tahapan pemrosesan data yang dilakukan :

a. Text Cleaning
Dataset yang sebelumnya diperoleh kemudian dibersihkan dengan menghapus elemen yang tidak diperlukan dalam analisis teks. Dilakukan untuk

P-ISSN : 2089-3353
E-ISSN : 2808-9162

meningkatkan akurasi model dengan hanya menyisahkan teks yang relevan. Langkah-langkah yang dilakukan :

1) Menghapus tanda baca (.,!?@#$%^&*())
2) Menghapus angka (0,1,2,3,4,5,6,7,8,9)
3) Menghapus emoji (😳 😊 🔥)
4) Menghapus URL atau tautan (https://contoh.com)
5) Menghapus mention (@username) dan hastag (#trending)

Sampel proses Text Cleaning :

Sebelum Text Cleaning: "drama 2024 : marahan sama temen sekelas h-1 uprak gagal 568902836× nangis karena ngerasa bakal gagal pas utbk (ternyata iyah wkwkwk) dapet pressure dari orang-orang yang bikin aku depresi dan hampir bundir hampir kena tipu hp rusak dan semua memori ilang kesleo https://t.co/29YxZ4ieR4"

Sesudah Text Cleaning: "drama marahan sama temen sekelas h uprak gagal nangis karena ngerasa bakal gagal pas utbk ternyata iyah wkwkwk dapet pressure dari orangorang yang bikin aku depresi dan hampir bundir hampir kena tipu hp rusak dan semua memori ilang kesleo"

b. Case Folding

Case Folding dilakukan untuk mengubah semua huruf kapital dalam teks menjadi huruf kecil (A-Z menjadi a – z). Ini dilakukan untuk menghilangkan perbedaan antara huruf besar dan kecil sehingga analisis teks bisa menjadi lebih konsisten. Tidak ada perubahan lain pada teks seperti spasi dan kata tetap sama, proses ini diterapkan karena model machine learning dan NLP membedakan huruf besar dan kecil. *Case Folding* membantu agar "Stres" dan "stress" dianggap sebagai kata yang sama, bukan entitas yang berbeda.

Sebelum Case Folding: "Berani banget gw mau UAS tapi modul gak kesentuh sama sekali karena gw dah burnout. Capek mental capek fisik"

Sesudah Case Folding: "berani banget gw mau uas tapi modul gak kesentuh sama sekali karena gw dah burnout capek mental capek fisik"

c. Tokenizing

Tahapan tokenizing diperlukan karena model yang digunakan (TF-IDF dan SVM) bekerja dengan representasi kata sebagai fitur numerik. Tanpa tokenizing, teks tidak dapat diolah dengan baik oleh algoritma *machine learning.*

Langkah-langkah yang dilakukan :

1) Menggunakan word_tokenize() dari NLTK untuk memecah teks menjadi kata-kata
2) Setiap kata dipisahkan berdasarkan spasi, sehingga menjadi unit yang dapat dianalisis.
3) Teks yang sudah dipisah menjadi daftar kata-kata (tokens).

Sebelum *Tokenizing*: "tau gasi rasanya stres banget sampe mau nangis aja gabisa "

Sesudah Tokenizing: "[tau, gasi, rasanya, stres, banget, sampe, mau, nangis, aja, gabisa]"

d. Stopword Removal

Tahapan ini digunakan agar TF-IDF dan model yang dibangun dapat bekerja lebih baik dengan kata-kata yang memiliki makna penting. Stopword dilakukan untuk mengurangi kata yang tidak relevan dalam mendeteksi stress dan depresi sehingga bisa dihapus tanpa kehilangan informasi penting.

Langkah-langkah yang dilakukan :

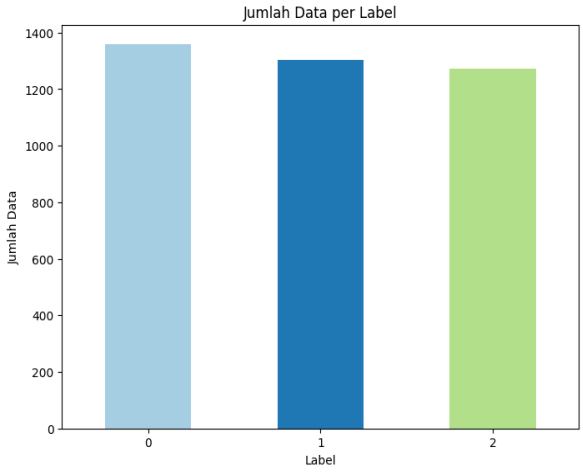
1) Melanjutkan hasil *tokenizing* berupa kata-kata
2) Menggunakan daftar *stopword* dari Sastrawi
3) Setiap kata dalam teks dibandingkan dengan daftar stopword
4) Jika termasuk stopword, maka akan dihapus dari daftar tokens.

Sebelum Stopword Removal: "namun, memang, di, situasi, ekonomi, kayak, gini, bisa, nggak, sih, isi, chat, tuh, diusahain, selain, capek, capek, dan, capek, aku, capek, mengeluh, dan, dikeluhin, capek, everyone, works, everyone, is, exhausted, everyone, is, stressed, out, kerja, udah, capek, tolong, chatroom, isinya, jangan, capek, juga"

Sesudah Stopword Removal: "memang, situasi, ekonomi, kayak, gini, sih, isi, chat, tuh, diusahain, capek, capek, capek, aku, capek, mengeluh, dikeluhin, capek, everyone, works, everyone, is, exhausted, everyone, is, stressed, out, kerja, udah, capek, chatroom, isinya, jangan, capek"

3.3. Pelabelan Data

Data diberi label sebagai stres (0), depresi (1), dan normal (2). Jumlah data pada masing-masing label dapat dilihat pada Gambar 2.



Gambar 2. Grafik jumlah data perlabel

Gambar 2 menunjukkan data pada label 0 (1360) label 1 (1302) label 2 (1273), dari grafik terlihat dataset relatif seimbang untuk masing-masing label. Hal ini penting untuk menjaga performa model mendeteksi setiap kategori tanpa bias terhadap salah satu label.

3.3. *Splitting* Data

Splitting data dilakukan untuk memisahkan dataset menjadi data uji dan data latih, dengan pembagian 80:20 merupakan langkah penting dalam pengembangan model yang efektif [27]. Sebanyak

P-ISSN : 2089-3353
E-ISSN : 2808-9162

3148 data digunakan untuk melatih model Support Vector Machine (SVM), sedangkan 787 sisanya digunakan untuk menguji performa model setelah proses pelatihan selesai.

3.4. Ekstraksi Fitur

Mengekstrak fitur yang relevan dari data teks untuk diproses dalam machine learning menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency). Frekuensi setiap kata dalam satu teks dan dalam semua dokumen digunakan untuk menentukan bobot kata [28], [29].

3.5. *Model Machine Learning*

Untuk mengembangkan model prediktif untuk mendeteksi stres dan depresi dari postingan media sosial, Support Vector Machine (SVM) akan digunakan sebagai salah satu algoritma machine learning.

a. Prediksi
Model ini diuji dengan mengambil data acak untuk secara otomatis mendeteksi apakah teks tersebut mengindikasikan stres, depresi, atau netral. Hasilnya dapat dilihat pada Tabel 3.

Tabel 3. Sample Hasil Prediksi

Text	Label	Prediksi
cari dimana hadirmu disyukuri biarpun sederhana usahamu dihargai kabarmu dinanti	2	netral
drawing is so tiring capek otak capek mental capek badan capek pokoknya jd cepet capek laper pusing pegel	0	stres
putus asa gini keadilan dunia musti dituntut keadilan akhirat	1	depresi
hati tenang jantung berdebar cepat badan lemes tempelin tangan dada ngerasain detak jantungnya rasain gangguan kecemasan gangguan panik	0	stress
smpsma pertumbuhanku macet dibilang tingginya gak nambah stres ekskul minatin syarat minimal nangis gagal lakuin biar renang minum suplemen gak ngaruh	0	stress
tau wanita menyerah hubungan komunikasi perhatian terabaikan kesalahan diulangulang hati disepelekan kesepian wanita mengungkapkannya	1	depresi
capek mental	0	stres

Pada Tabel 3 dapat dilihat bahwa setiap prediksi yang dilakukan berhasil menghasilkan keputusan yang tepat sesuai dengan label masing-masing teks. Hal ini menunjukkan efektivitas model SVM dalam membedakan antara teks yang mengandung indikasi stres, depresi, dan netral secara tepat.

b. Evaluasi
Model SVM yang telah dilatih kemudian digunakan untuk memprediksi label dari data pengujian yang baru.

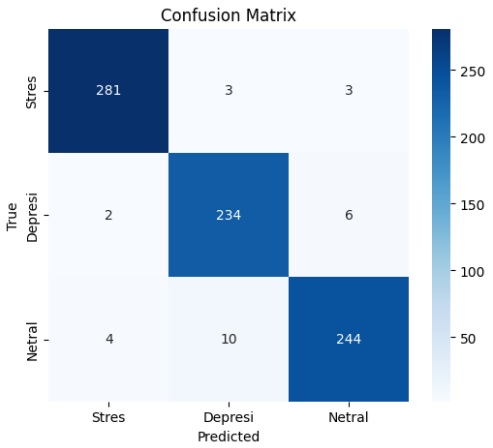
1) Akurasi Kinerja Model
Untuk menilai performa dari metode ini, model harus dievaluasi. Hasil evaluasi dapat dilihat pada Gambar 3.

```
Accuracy: 0.9644218551461246
              precision    recall  f1-score   support

           0       0.98      0.98      0.98       287
           1       0.95      0.97      0.96       242
           2       0.96      0.95      0.95       258

    accuracy                           0.96       787
   macro avg       0.96      0.96      0.96       787
weighted avg       0.96      0.96      0.96       787
```

Gambar 3. Hasil Akurasi

Dari Gambar 3, hasil evaluasi model klasifikasi untuk mendeteksi stres (0), depresi (1), dan netral (2) ditunjukkan. Model ini memiliki akurasi keseluruhan sebesar 96,44%, dengan kinerja yang sangat baik di semua kelas, seperti yang ditunjukkan oleh nilai presisi, recall, dan F1-score yang tinggi secara konsisten (rata-rata 0,96). Hasil ini menunjukkan bahwa model ini mampu memprediksi ketiga label secara akurat, baik dalam mengidentifikasi dengan benar data yang termasuk dalam kelas tertentu dan memastikan bahwa prediksinya tepat. Dukungan sampel untuk setiap label juga cukup seimbang: 287 untuk stres, 242 untuk depresi, dan 258 untuk netral.

2) Confusion Matrix
Hasil *confusion matrix* dari model terbaik dapat dilihat pada Gambar 4.



Gambar 4. Confusion Matrix

Gambar 4 menunjukkan bahwa model berhasil mengidentifikasi 281 sampel stres dengan benar, 234 sampel depresi dengan benar, dan 244 sampel netral dengan benar. Hasil ini menunjukkan bahwa model memiliki akurasi yang tinggi dengan kesalahan yang relatif rendah dalam mendeteksi potensi masing-masing label.

P-ISSN : 2089-3353
E-ISSN : 2808-9162

3) Cross Validation

Setelah model dibuat, evaluasi akan dilakukan untuk menilai performa model dengan menggunakan Cross Validation. Cross Validation juga digunakan untuk memastikan bahwa model tidak overfitting dan dapat bekerja dengan baik pada dataset yang lebih luas dan beragam. Hal ini dapat dilihat pada Gambar 5.

	Fold	Cross-Validation Score
0	Fold 1	0.960610
1	Fold 2	0.963151
2	Fold 3	0.954257
3	Fold 4	0.949174
4	Fold 5	0.925032
5	Mean	0.950445

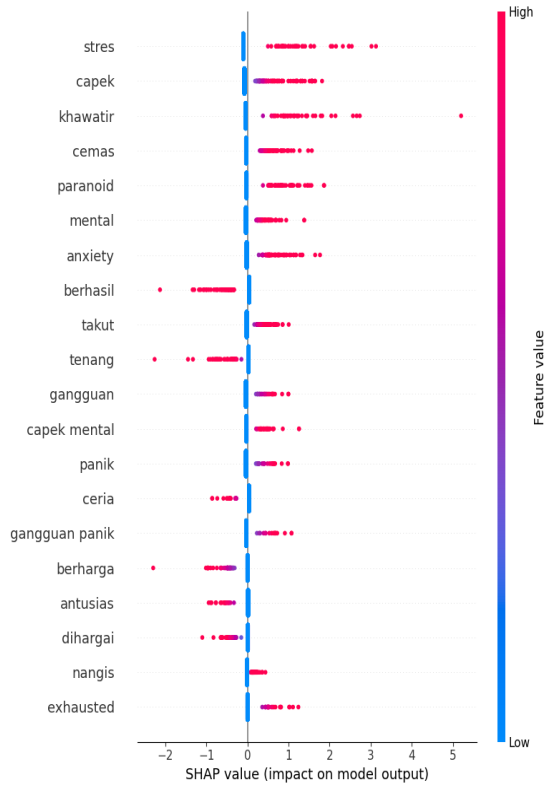
Gambar 5. Cross Validation

Hasil validasi silang menunjukkan nilai akurasi model SVM pada 5 fold data, dengan nilai tertinggi pada Fold 2 (96.31%) dan terendah pada Fold 5 (92.50%), sehingga menghasilkan rata-rata akurasi sebesar 95.04%. Hal ini mengindikasikan bahwa model memiliki performa yang konsisten dan cukup baik, meskipun terdapat sedikit variasi antar lipatan karena perbedaan karakteristik data. Secara keseluruhan, dapat dilihat bahwa model mampu memprediksi dengan akurasi yang tinggi pada dataset ini.

3.6. *SHAP (SHapley Additive exPlanations)*

Teknik SHAP akan diterapkan untuk mendapatkan interpretabilitas untuk setiap prediksi yang dibuat.

a. Hasil visualisasi SHAP untuk label stres

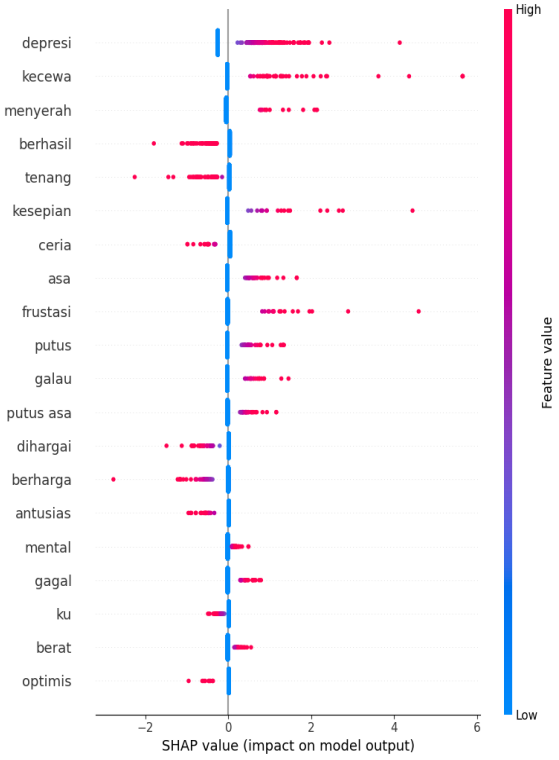


Gambar 6. Hasil Visualisasi SHAP untuk label stres

Grafik pada Gambar 6 menunjukkan hasil penerapan teknik SHAP pada model untuk mengetahui interpretabilitas dari setiap prediksi untuk label 'stress'. Pada grafik ini, kita dapat melihat bahwa fitur 'stres', 'capek', dan 'khawatir' memiliki titik merah yang jauh ke kanan artinya fitur ini sangat berkontribusi ke arah prediksi positif. Fitur berada pada level tertinggi ketika nilai SHAP-nya mendekati atau melebihi +1.5. Dalam konteks model ini, nilai SHAP > +1.0 sudah menunjukkan pengaruh yang sangat signifikan terhadap output model. Sebaliknya, nilai SHAP < -0.5 menunjukkan kontribusi negatif yang kuat. Interpretasi ini membantu menjelaskan bagaimana model menggunakan fitur-fitur tersebut untuk menghasilkan prediksi stres secara transparan.

b. Hasil visualisasi SHAP untuk label depresi

Bagan pada Gambar 7 menunjukkan hasil dari penerapan teknik SHAP pada model untuk memahami interpretasi setiap prediksi untuk label 'depresi'. Pada grafik ini, kita dapat melihat bahwa fitur 'depresi', 'kecewa', dan 'menyerah' memiliki titik merah yang jauh ke kanan artinya fitur ini sangat berkontribusi ke arah prediksi positif. Fitur berada pada level tertinggi ketika nilai SHAP-nya mendekati atau melebihi +1.5. Dalam konteks model ini, nilai SHAP > +1.0 sudah menunjukkan pengaruh yang sangat signifikan terhadap output model. Sebaliknya, nilai SHAP < -0.5 menunjukkan kontribusi negatif yang kuat. Interpretasi ini membantu menjelaskan bagaimana model menggunakan fitur-fitur tersebut untuk menghasilkan prediksi stres secara transparan.



Gambar 7. Hasil Visualisasi SHAP untuk label depresi

4. Kesimpulan

Model SVM yang dibangun dengan menerapkan teknik SHAP untuk mendeteksi stres dan depresi dari unggahan media sosial bertujuan untuk mengevaluasi keefektifan dan memahami interpretabilitas dari setiap prediksi. Hasil penelitian menunjukkan bahwa model SVM mencapai akurasi 96,44%, dengan kinerja yang sangat baik di semua kelas, yang ditunjukkan dengan nilai precision, recall, dan F1-score yang tinggi secara konsisten (rata-rata 0,96). Teknik SHAP yang diterapkan secara signifikan membantu dalam mengidentifikasi fitur-fitur penting, seperti fitur 'stres', 'capek', dan 'khawatir', yang berada di tingkat tertinggi dalam berkontribusi pada prediksi stres. Sementara itu, untuk label depresi, fitur 'depresi', 'kecewa', dan 'menyerah' berada di level tertinggi dalam memberikan kontribusi pada prediksi depresi, yang ditunjukkan oleh nilai SHAP yang tinggi. Dengan demikian, teknik ini dapat mendukung pengembangan alat prediksi berbasis AI yang lebih transparan dan dapat diandalkan. Namun, penelitian ini memiliki keterbatasan terkait ukuran dataset dan potensi bias dalam data. Penelitian selanjutnya diharapkan dapat mengatasi keterbatasan ini dengan memperluas dataset dan mengintegrasikan metode baru untuk meningkatkan kinerja model yang dibangun.

Daftar Rujukan

[1] C. V. Lotulung and I. G. Purnawinadi, "Deteksi Dini Depresi Mahasiswa Baru Jurusan Keperawatan," *Jurnal Ilmu Pendidikan dan Psikologi*, vol. 4, no. 2, 2024, doi: https://doi.org/10.51878/paedagogy.v4i2.3042.

[2] Y. Sulistyorini, Mahmudah, and N. Puspitasari, "Peningkatan Kemampuan Deteksi Dini Depresi pada Ibu Hamil di Kota Surabaya," *Jurnal Ilmiah Pengabdian kepada Masyarakat*, vol. 8, no. 3, pp. 469–476, 2023, doi: https://doi.org/10.33084/pengabdianmu.v8i3.4469.

[3] A. Ahmed *et al*., "Machine learning models to detect anxiety and depression through social media: A scoping review," *Computer Methods and Programs in Biomedicine Update*, vol. 2, 2022, doi: https://doi.org/10.1016/j.cmpbup.2022.100066.

[4] G. Gunasekar, S. G. Moorthy, K. Chakrapani, and G. Ponsam, "Early Detection of Depression from Social Media Data Using Machine Learning Algorithms," in *2020 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS)*, Chennai: IEEE, 2020, pp. 1–6. doi: 10.1109/ICPECTS49113.2020.9336974.

[5] E. Z. Heanoy and N. R. Brown, "Impact of Natural Disasters on Mental Health: Evidence and Implications," *Healthcare*, vol. 12, no. 18, 2024, doi: 10.3390/healthcare12181812.

[6] A. Sood, D. Sharma, M. Sharma, and R. Dey, "Prevalence and repercussions of stress and mental health issues on primary and middle school students: a bibliometric analysis," *Frontiers in Psychiatry*, vol. 15, 2024, doi: https://doi.org/10.3389/fpsyt.2024.1369605.

[7] S. Chancellor and M. De Choudhury, "Methods in Predictive Techniques for Mental Health Status on Social Media: A Critical Review," *NPJ Digital Medicine*, vol. 3, no. 43, 2020, [Online]. Available: https://www.nature.com/articles/s41746-020-0233-7

[8] K. Rahayu, V. Fitria, D. Septhya, Rahmaddeni, and L. Efrizoni, "Klasifikasi Teks untuk Mendeteksi Depresi dan Kecemasan Pada Pengguna Twitter Berbasis Machine Learning," *Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 108–114, 2023, doi: https://doi.org/10.57152/malcom.v3i2.780.

[9] S. Mutmainah, "Kemungkinan Depresi dari Postingan pada Media Sosial," *Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi*, vol. 1, no. 2, 2022, doi: 10.20885/snati.v1i2.11.

[10] L. Ilias, S. Mouzakitis, and D. Askounis, "Calibration of Transformer-Based Models for Identifying Stress and Depression in Social Media," *IEEE Transactions On Computational Social Systems*, vol. 11, no. 2, pp. 1979–1990, 2024, doi: 10.1109/TCSS.2023.3283009.

[11] S. Dabhane and P. M. Chawan, "Depression Detection on Social Media using Machine Learning Techniques," *International Journal for Scientific Research & Development*, vol. 9, no. 4, pp. 291–294, 2021.

[12] F. A. Bakar, N. M. Nawi, and A. A. Hezam, "Predicting Depression Using Social Media Posts," *Journal of Soft Computing and Data Mining*, vol. 2, no. 1, pp. 39–48, 2021, doi: https://doi.org/10.30880/jscdm.2021.02.02.004.

[13] S. Inamdar, R. Chapekar, S. Gita, and B. Pradhan, "Machine Learning Driven Mental Stress Detection on Reddit Posts Using Natural Language Processing," *Springer Human-Centric Intelligent*, vol. 3, no. 2, pp. 80–91, 2023, doi: 10.1007/s44230-023-00020-8.

[14] N. Paul E and S. Juliet, "Comparative Analysis of Machine Learning Techniques for Mental Health Prediction," in *2023 8th International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore, India: IEEE, 2023, pp. 1–6. doi: 10.1109/ICCES57224.2023.10192763.

[15] G. F. Situmorang and R. Purba, "Deteksi Potensi Depresi dari Unggahan Media Sosial X Menggunakan Teknik NLP dan Model IndoBERT," *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 2, 2024, doi: https://doi.org/10.47065/bits.v6i2.5496.

[16] A. Le Glaz *et al*., "Machine Learning and Natural Language Processing in Mental Health: Systematic Review," *Journal of Medical Internet Research*, vol. 23, no. 5, 2021, doi: 10.2196/15708.

[17] T. Zhang, A. M. Schoene, Shaoxiong Ji, and S. Ananiadou, "Natural language processing applied to mental illness detection: a narrative review," *npj Digital Medicine*, vol. 5, no. 46, 2022, doi: https://doi.org/10.1038/s41746-022-00589-7.

[18] H. Duong and T. Nguyen-Thi, "A review: Preprocessing Techniques and Data Augmentation for Sentiment Analysis," *Computational Social Networks*, vol. 8, no. 1, 2021, doi: https://doi.org/10.1186/s40649-020-00080-x.

[19] W. L. L. Phyu, H. M. S. Naing, and W. P. Pa, "Improving the Performance of Low-resourced Speaker Identification with Data Preprocessing," *Journal of ICT Research and Applications*, vol. 17, no. 3, 2023, doi: https://doi.org/10.5614/itbj.ict.res.appl.2023.17.3.1.

[20] B. Tahayna, R. K. Ayyasamy, and R. Akbar, "Context-Aware Sentiment Analysis using Tweet Expansion Method," *Journal of ICT Research and Applications*, vol. 16, no. 2, 2022, doi: https://doi.org/10.5614/itbj.ict.res.appl.2022.16.2.3.

[21] S. A. Utiarahman and A. M. M. Pratama, "Penerapan Support Vector Machine dan Random Forest ClassifierUntuk Klasifikasi Tingkat Obesitas," *Jurnal FASILKOM*, vol. 14, no. 3, 2024, doi: https://doi.org/10.37859/jf.v14i3.8104.

[22] O. C. Atalaya *et al*., "Supervised learning using support vector machine applied to sentiment analysis of teacher performance satisfaction," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 28, no. 1, pp. 516–524, 2022, doi: http://doi.org/10.11591/ijeecs.v28.i1.pp516-524.

[23] V. Bidve *et al*., "Use of explainable AI to interpret the results of NLP models for sentimental analysis," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 35, no. 1, pp. 511–519, 2024, doi: http://doi.org/10.11591/ijeecs.v35.i1.pp511-519.

[24] N. Nordin, Z. Zainol, M. H. M. Noor, and L. F. Chan, "An explainable predictive model for suicide attempt risk using an ensemble learning and Shapley Additive Explanations (SHAP) approach," *Asian Journal of Psychiatry*, vol. 79, 2023, doi: https://doi.org/10.1016/j.ajp.2022.103316.

[25] Y. Nohara, K. Matsumoto, H. Soejima, and N. Nakashima, "Explanation of machine learning models using shapley additive explanation and application for real data in hospital," *Computer Methods and Programs in Biomedicine*, vol. 214, 2022, doi: https://doi.org/10.1016/j.cmpb.2021.106584.

P-ISSN : 2089-3353
E-ISSN : 2808-9162

[26] S. M. Lundberg *et al.*, "From local explanations to global understanding with explainable AI for trees," *Nature Machine Intelligence*, vol. 2, pp. 55–67, 2020, doi: https://doi.org/10.1038/s42256-019-0138-9.
[27] M. Haqqi, L. Rochmah, A. D. Safitri, R. A. Pratama, and Tarwoto, "ImplementationOf Machine Learning To IdentifyTypes Of Waste Using CNN Algorithm," *Jurnal FASILKOM*, vol. 14, no. 3, pp. 761–765, 2024, doi: https://doi.org/10.37859/jf.v14i3.8116.
[28] C. A. N. Agustina, R. Novita, Mustakim, and N. E. Rozanda, "The Implementation of TF-IDF and Word2Vec on Booster Vaccine Sentiment Analysis Using Support Vector Machine Algorithm," *Procedia Computer Science*, vol. 234, pp. 156–163, 2024, doi: https://doi.org/10.1016/j.procs.2024.02.162.
[29] M. Liang and T. Niu, "Research on Text Classification Techniques Based on Improved TF-IDF Algorithm and LSTM Inputs," *Procedia Computer Science*, vol. 208, pp. 460–470, 2022, doi: https://doi.org/10.1016/j.procs.2022.10.064.