Jurnal Kreativitas Mahasiswa Informatika **E-ISSN: 3164-1989**
Penerbit: Fakultas Ilmu Komputer Universitas Pamulang **P-ISSN: 2797-6327**
Vol. 7, No. 2, Agustus 2026 (478-491) **https://doi.org/10.32493/jatimika.v7i1**

Optimasi Algoritma Support Vector Machine Menggunakan Random Search Pada Sistem Diagnosa Depresi

Eva Nauli[1*], Nurjaya[2]

[1]Teknik Informatika, Universitas Pamulang, Jl. Raya Puspiptek No. 46, Buaran, Kec. Pamulang, Kota Tangerang Selatan, Banten, Indonesia
e-mail: [1]evanauli191933@gmail.com, [2]dosen00370@unpam.ac.id

Submitted Date: January 18, 2026 Reviewed Date: April 01, 2026
Revised Date: April 24, 2026 Accepted Date: May 24, 2026

Abstract

Depression is a serious mental disorder characterized by persistent mood disturbances and is a major contributor to the global burden of disability. This study aims to improve the accuracy of depression diagnosis using the Support Vector Machine (SVM) algorithm optimized with the Random Search technique. The dataset used was obtained from the Kaggle platform, focusing on classifying respondents into depressed and non-depressed categories. The methodology includes data cleaning, exploratory data analysis (EDA), feature engineering, data splitting, and model evaluation with and without Random Search optimization. The results show that the standard SVM model achieved an accuracy of 95.70% with an F1-score of 95.72%, while the optimized SVM using Random Search improved performance to an accuracy of 97.85% with an F1-score of 97.86% and an AUC of 99.66%. The optimized model demonstrated better evaluation results compared to the standard (default) SVM model. This study contributes to improving early detection of depression symptoms and demonstrates that combining machine learning techniques with hyperparameter optimization can produce prediction models that are more accurate, stable, and reliable.

Keywords: Depression, Support Vector Machine (SVM), Random Search, model optimization

Abstrak

Depresi merupakan gangguan mental serius yang ditandai dengan gangguan mood persisten dan menjadi kontributor utama beban disabilitas global. Penelitian ini bertujuan untuk meningkatkan akurasi diagnosa penyakit depresi menggunakan algoritma Support Vector Machine (SVM) yang dioptimalkan dengan teknik Random Search. Dataset yang digunakan diperoleh dari platform Kaggle, dengan fokus pada klasifikasi responden menjadi kategori depresi atau tidak depresi. Metodologi yang diterapkan melibatkan pembersihan data, exploratory data analysis (EDA), feature engineering, pembagian data, serta pengujian model dengan dan tanpa optimasi Random Search. Hasil penelitian menunjukkan bahwa model SVM standar menghasilkan akurasi sebesar 95,70% dengan F1-Score 95,72%, sedangkan setelah dilakukan optimasi menggunakan Random Search, performa model meningkat menjadi akurasi 97,85% dengan F1-Score 97,86% serta AUC sebesar 99,66%. Model yang dioptimalkan terbukti memberikan hasil evaluasi yang lebih baik dibandingkan dengan model SVM standar (default). Penelitian ini memberikan kontribusi terhadap peningkatan deteksi dini gejala depresi dan menunjukkan bahwa penggabungan teknik machine learning dengan optimasi hiperparameter mampu menghasilkan model prediksi yang lebih akurat, stabil, dan andal

Kata kunci: Depresi, Support Vector Machine (SVM), Random Search, optimasi model

E-ISSN: 3164-1989
P-ISSN: 2797-6327
https://doi.org/10.32493/jatimika.v7i1

1. Pendahuluan

Depresi (*depressive disorder*) telah berevolusi menjadi krisis kesehatan global yang multidimensi, melampaui sekadar fluktuasi emosi biasa menjadi gangguan mood persisten yang mendegradasi kualitas hidup, kesehatan fisik, dan produktivitas sosial individu. Badan Kesehatan Dunia (WHO) mengidentifikasi depresi sebagai kontributor utama beban disabilitas global, yang mempengaruhi sekitar 3,8% populasi dunia atau setara dengan 280 juta jiwa (WHO, 2023). Prevalensi ini tidak memandang usia, dengan dampak signifikan terlihat pada 5% populasi dewasa dan 5,7% pada lansia di atas 60 tahun. Lebih mengkhawatirkan lagi, depresi memiliki korelasi erat dengan peningkatan risiko mortalitas melalui bunuh diri. Data menunjukkan lebih dari 700.000 kasus bunuh diri setiap tahunnya terkait dengan depresi yang tidak terdiagnosis atau tidak tertangani dengan baik, menempatkan kelompok usia produktif remaja dan dewasa muda (15–29 tahun) sebagai korban utama (WHO, 2023). Fenomena ini menempatkan urgensi tinggi pada pengembangan metode deteksi dini yang efektif untuk memitigasi dampak fatal dari gangguan ini. Tantangan utama dalam penanganan depresi terletak pada kompleksitas diagnosisnya. Di negara-negara berkembang, diperkirakan 75% penderita tidak mendapatkan penanganan medis yang memadai (WHO, 2023).

Seiring dengan perkembangan teknologi kecerdasan buatan, berbagai pendekatan *machine learning* telah dikembangkan untuk meningkatkan akurasi diagnosis depresi, termasuk penggunaan algoritma Support Vector Machine (SVM) yang dikenal memiliki performa baik pada data berdimensi tinggi. Namun demikian, performa SVM sangat bergantung pada pemilihan hiperparameter yang optimal. Berbagai metode optimasi seperti Grid Search, Bayesian Optimization, dan Genetic Algorithm telah digunakan dalam penelitian sebelumnya. Bayesian Optimization dikenal mampu menemukan parameter optimal secara adaptif, namun memiliki kompleksitas implementasi yang lebih tinggi dan membutuhkan waktu komputasi yang relatif besar pada beberapa kasus. Sementara itu, Genetic Algorithm menawarkan kemampuan eksplorasi global yang baik, tetapi memerlukan pengaturan parameter tambahan serta proses iteratif yang lebih kompleks.

Dalam konteks penelitian ini, Random Search dipilih sebagai metode optimasi karena memiliki keseimbangan antara efisiensi komputasi dan kemampuan eksplorasi ruang parameter yang luas. Dibandingkan Grid Search yang bersifat ekshaustif, Random Search mampu menemukan kombinasi hiperparameter yang optimal dengan jumlah iterasi yang lebih sedikit. Selain itu, berdasarkan penelitian Mantovani et al. (2015), Random Search terbukti efektif dalam berbagai kasus optimasi SVM, khususnya ketika hanya sebagian kecil parameter yang berpengaruh signifikan terhadap performa model. Oleh karena itu, metode ini dianggap lebih praktis dan efisien untuk diterapkan pada dataset yang digunakan dalam penelitian ini.

Kesenjangan ini disebabkan oleh defisit sumber daya manusia profesional, keterbatasan anggaran kesehatan mental, serta stigma sosial yang masih melekat kuat di masyarakat. Metode diagnosis konvensional yang bergantung pada wawancara klinis dan kuesioner mandiri (*self-report*) seringkali mengandung bias subjektivitas, baik dari sisi pasien yang enggan terbuka maupun interpretasi klinisi. Variabilitas gejala antar individu—seperti perbedaan manifestasi pada remaja dibandingkan lansia—semakin mempersulit identifikasi yang presisi dan konsisten (Aziz & Abasa, 2025). Pasien seringkali tidak menyadari bahwa gejala fisik seperti gangguan tidur atau perubahan n afsu makan adalah manifestasi dari gangguan psikologis, sehingga mereka tidak mencari bantuan profesional tepat waktu. Oleh karena itu, kebutuhan akan metode diagnosis yang lebih objektif, terukur, dan dapat diakses secara luas menjadi sangat mendesak.

Perkembangan pesat teknologi komputasi, khususnya dalam domain kecerdasan buatan (*Artificial Intelligence*) dan pembelajaran mesin (*Machine Learning*), menawarkan paradigma baru dalam deteksi dini gangguan kesehatan mental. *Data mining*, sebagai proses ekstraksi informasi bernilai dari kumpulan data besar, telah menjadi fondasi dalam pengembangan model prediktif di bidang medis (Wu et al., 2021). Berbagai penelitian telah mengembangkan model *machine learning* untuk deteksi depresi dengan tingkat akurasi yang tinggi. Studi oleh Abdussamad et al. (2025) menunjukkan bahwa integrasi SVM dengan teknologi pengenalan wajah mampu mencapai akurasi 90,6%, sementara Rahayu et al. (2023) menemukan bahwa Random Forest memberikan performa terbaik dengan akurasi 95,7% dibandingkan algoritma lain. Selain itu, Aleem et al. (2022) melaporkan bahwa SVM, khususnya varian multikernel, dapat mencapai akurasi hingga 97,54%, menunjukkan potensi tinggi algoritma ini dalam klasifikasi data klinis.

Meskipun demikian, masih terdapat kesenjangan dalam penelitian terkait optimasi model, khususnya pada algoritma SVM yang sangat sensitif terhadap pemilihan hiperparameter seperti kernel, parameter regularisasi (C), dan gamma. Sebagian penelitian masih menggunakan parameter default atau pendekatan optimasi yang kurang efisien, sehingga performa model belum optimal secara konsisten.

Untuk mengatasi permasalahan tersebut, penelitian ini mengusulkan penerapan teknik Random Search untuk mengoptimasi hiperparameter SVM. Metode ini dipilih karena lebih efisien dalam mengeksplorasi ruang parameter dibandingkan Grid Search, serta memiliki kemampuan menemukan kombinasi parameter yang optimal dengan biaya komputasi yang lebih rendah (Mantovani et al., 2015).

Dengan demikian, penelitian ini bertujuan untuk meningkatkan kinerja model SVM dalam diagnosis depresi melalui optimasi hiperparameter menggunakan Random Search, sehingga dihasilkan model prediksi yang lebih akurat, stabil, dan memiliki kemampuan generalisasi yang lebih baik.

2. Metodologi Penelitian

Penelitian ini mengadopsi pendekatan kuantitatif berbasis *text mining* melalui kerangka kerja *Knowledge Discovery in Databases* (KDD). Kerangka kerja ini dipilih karena menyediakan alur yang sistematis dan terstruktur dalam mengekstraksi informasi berharga dari kumpulan data yang besar dan kompleks. Proses ini mencakup serangkaian tahapan yang dimulai dari pengumpulan data, pra-pemrosesan, transformasi, klasifikasi, hingga evaluasi hasil.

2.1. Analisis Kebutuhan Sistem

Penelitian ini dirancang menggunakan pendekatan kuantitatif eksperimental dalam bidang *data mining* dan *machine learning*. Metodologi penelitian mengikuti tahapan sistematis *Knowledge Discovery in Databases* (KDD) yang meliputi pengumpulan data, pra-pemrosesan data, transformasi, *data mining* (pemodelan), dan evaluasi/interpretasi. Seluruh proses komputasi dilakukan menggunakan bahasa pemrograman Python dengan dukungan pustaka-pustaka ilmiah standar.

Tabel 1. Analisa Kebutuhan Perangkat Keras dan Perangkat Lunak

Perangkat Keras	**Spesifikasi**
Processor	AMD A9-0945 Radeon R5
Penyimpanan	Minimum 1 TB SSD
RAM	Minimum 8GB
Koneksi Internet	Stabil
Perangkat Lunak	**Spesifikasi**
Sistem Operasi	Windows 10
Aplikasi Simulator	Jupyter
Bahasa Pemrograman	Python 3.x
Library Python	Pandas
	Numpy
	Matplotlib
	Seaborn

Scikit-learn

2.2. Analisis Data Eksploratif (Exploratory Data Analysis

Tahap ini bertujuan untuk memahami karakteristik dataset sebelum pemodelan. Hasil analisis menunjukkan adanya ketidakseimbangan kelas pada variabel target serta hubungan yang signifikan antara beberapa fitur numerik dengan kondisi depresi. Visualisasi dan analisis korelasi digunakan untuk mengidentifikasi fitur yang paling relevan. Secara praktis, temuan ini menjadi dasar dalam pemilihan fitur dan proses pemodelan, sehingga model yang dibangun lebih mampu menangkap pola yang berpengaruh terhadap prediksi depresi.

2.3. Pra-pemrosesan Data (Data Preprocessing)

Data Kualitas adalah faktor kunci dalam keberhasilan model *machine learning*. Tahapan pra-pemrosesan yang dilakukan meliputi:

1. Pembersihan Data (Data Cleaning)
    a. Penghapusan Atribut Irrelevan:
    Atribut 'Name' dihapus dari dataset karena merupakan *identifier* unik yang tidak memiliki nilai prediktif terhadap kondisi depresi dan hanya berfungsi sebagai label identitas.
    b. Penanganan *Missing Value*:
    Dilakukan pengecekan terhadap nilai yang hilang. Jika ditemukan pada atribut numerik (seperti Umur atau CGPA), nilai kosong diimputasi menggunakan nilai rata-rata (*mean*). Jika pada atribut kategorikal, digunakan nilai modus (nilai yang paling sering muncul). Langkah ini penting untuk menjaga integritas jumlah sampel tanpa harus membuang data.
    c. Pendektesian Duplikasi:
    Memastikan tidak ada baris data ganda yang dapat membiaskan hasil pelatihan model.
2. Penanganan Outlier:

Analisis outlier dilakukan menggunakan metode *Interquartile Range* (IQR). Data yang berada di luar rentang Q1−1.5×IQR Q1 - 1.5 \times IQRQ1−1.5×IQR dan Q3+1.5×IQR Q3 + 1.5 \times IQR Q3+1.5×IQR diidentifikasi sebagai outlier. Pada penelitian ini, beberapa outlier terdeteksi pada fitur seperti *Umur* dan *Tekanan Akademik*. Alih-alih langsung menghapus data tersebut, dilakukan pengujian awal dengan membandingkan performa model SVM pada dua skenario, yaitu dengan penghapusan outlier dan tanpa penghapusan outlier. Hasil pengujian menunjukkan bahwa tidak terdapat peningkatan performa yang signifikan pada model setelah outlier dihapus.

Keputusan untuk mempertahankan outlier juga didukung oleh karakteristik algoritma SVM yang berfokus pada *support vectors*, sehingga tidak seluruh data memiliki pengaruh yang sama terhadap pembentukan *decision boundary*. Selain itu, dalam konteks data kesehatan mental, nilai ekstrem dapat merepresentasikan kondisi nyata yang relevan secara klinis, seperti tingkat tekanan yang sangat tinggi atau kondisi individu dengan karakteristik khusus. Oleh karena itu, mempertahankan outlier dianggap lebih tepat untuk menjaga representasi data dan meningkatkan kemampuan generalisasi model.

Pada tahap normalisasi data, penelitian ini secara eksplisit menggunakan metode StandardScaler untuk menstandarkan fitur numerik ke dalam distribusi dengan rata-rata 0 dan standar deviasi 1. Pemilihan metode ini didasarkan pada karakteristik SVM yang sensitif terhadap skala fitur serta kemampuannya dalam menangani distribusi data yang tidak seragam. Dengan demikian, proses normalisasi dapat direplikasi secara konsisten dalam penelitian selanjutnya.

3. Feature Engineering dan Transformasi Data:

Algoritma SVM bekerja secara matematis dengan input numerik, sehingga seluruh data kategorikal harus dikonversi.

    a. Label Encoding & Binary Encoding:
    Variabel biner seperti 'Gender' (Female=0, Male=1), 'Working Professional or Student' (Student=0, Professional=1), 'Have you ever had suicidal thoughts?' (No=0, Yes=1), 'Family History of Mental Illness', dan target 'Depression' diubah menjadi format numerik biner 0 dan 1.

b. Ordinal Encoding:
Variabel yang memiliki tingkatan logis, seperti 'Dietary Habits' (Unhealthy=0, Moderate=1, Healthy=2) dan 'Sleep Duration' (dikonversi menjadi urutan 0 hingga 3 berdasarkan durasi), diubah menggunakan Ordinal Encoding untuk mempertahankan informasi urutan tingkatannya

c. Normalisasi Data:
Fitur-fitur numerik dalam dataset memiliki rentang skala yang sangat berbeda (misalnya, Umur 18-60, CGPA 5-10, Tekanan Akademik 1-5). SVM sangat sensitif terhadap skala fitur karena menggunakan perhitungan jarak Euclidean untuk menentukan *hyperplane*. Jika tidak dinormalisasi, fitur dengan skala besar akan mendominasi fungsi objektif model. Oleh karena itu, teknik normalisasi (seperti *Min-Max Scaling* atau *StandardScaler*) diterapkan untuk menormalkan seluruh fitur numerik ke dalam rentang yang seragam, sehingga setiap fitur memiliki kontribusi yang seimbang dalam pembentukan model.

4. Pembagian Data (*Data Splitting*):
Dataset dibagi menjadi dua subset independen: data latih (*training set*) dan data uji (*testing set*). Proporsi pembagian yang digunakan adalah 80% untuk data latih (2.044 data) dan 20% untuk data uji (512 data). Pembagian ini bertujuan untuk melatih model pada satu himpunan data dan memvalidasi performanya pada himpunan data lain yang belum pernah dilihat sebelumnya (*unseen data*), guna mengukur kemampuan generalisasi model secara objektif.

2.4. Konsep Algoritma dan Optimasi

Support Vector Machine (SVM) *Support Vector Machine* (SVM) adalah algoritma *supervised learning* yang bekerja berdasarkan prinsip *Structural Risk Minimization* (SRM). Tujuan utama SVM adalah menemukan *hyperplane* pemisah terbaik yang memisahkan dua kelas data (Depresi vs Tidak Depresi) dengan *margin* maksimal. Margin didefinisikan sebagai jarak antara *hyperplane* dengan *support vectors*, yaitu titik-titik data terdekat dari masing-masing kelas (Aprilla et al., 2018). Fungsi keputusan SVM linear dapat dinyatakan sebagai:

Dimana w adalah vektor bobot yang menentukan orientasi *hyperplane*, x adalah vektor input, dan b adalah bias. Jika f(x) \geq 0, data diklasifikasikan ke kelas positif, dan sebaliknya. Dalam banyak kasus, data tidak dapat dipisahkan secara linear di ruang input asli. Untuk mengatasi hal ini, SVM menggunakan teknik *Kernel Trick*. Fungsi kernel memetakan data input ke dalam ruang fitur berdimensi lebih tinggi di mana data tersebut dapat dipisahkan secara linear. Kernel yang umum digunakan meliputi Linear, Polinomial, dan *Radial Basis Function* (RBF).

Optimasi *Random Search* Performa SVM sangat bergantung pada hiperparameter. Parameter utama yang perlu dioptimasi adalah:

a. C (Regularisasi):
Mengontrol *trade-off* antara margin yang lebar dan klasifikasi yang benar dari data latih. Nilai C yang besar memprioritaskan klasifikasi benar (risiko *overfitting*), sedangkan C kecil memprioritaskan margin lebar (risiko *underfitting*).

b. Kernel: Menentukan jenis fungsi pemetaan (Linear, RBF, Poly).

c. Gamma: Koefisien kernel untuk RBF/Poly, menentukan seberapa jauh pengaruh satu sampel pelatihan tunggal.

Random Search adalah teknik optimasi yang memilih kombinasi parameter secara acak dari distribusi statistik yang ditentukan. Berbeda dengan *Grid Search* yang mencoba semua kombinasi secara "brute force", *Random Search* mengeksplorasi ruang parameter secara lebih efisien. Penelitian menunjukkan bahwa *Random Search* memiliki probabilitas tinggi untuk menemukan set parameter yang optimal dengan jumlah iterasi yang jauh lebih sedikit, terutama ketika hanya sebagian parameter yang benar-benar berpengaruh signifikan terhadap performa model (Mantovani et al., 2015). Dalam penelitian ini, *Random Search* digunakan dengan skema validasi silang (*cross-validation*) untuk memastikan parameter terpilih valid secara statistik.

2.5. Evaluasi Mode

Evaluasi kinerja model dilakukan menggunakan *Confusion Matrix*, sebuah tabel yang membandingkan prediksi model dengan label sebenarnya. Dari matriks ini, dihitung metrik-metrik evaluasi sebagai berikut:

1. Akurasi (*Accuracy*):
   Proporsi total prediksi yang benar.
2. Presisi (*Precision*):
   Tingkat ketepatan prediksi positif (seberapa banyak yang diprediksi depresi benar-benar depresi).
3. *Recall* (Sensitivitas):
   Kemampuan model mendeteksi seluruh kasus positif (seberapa banyak penderita depresi yang berhasil dideteksi). Metrik ini sangat krusial dalam diagnosis medis untuk menghindari *False Negative.*
4. *F1-Score*:
   Rata-rata harmonik antara presisi dan recall, memberikan gambaran performa yang seimbang terutama pada data tidak seimbang.
5. *ROC-AUC*:
   Area Under the Receiver Operating Characteristic Curve digunakan untuk mengukur kemampuan diskriminasi model pada berbagai ambang batas. Nilai AUC mendekati 1 menunjukkan performa klasifikasi yang sempurna.

3. Hasil dan Diskusi

Penelitian ini bertujuan untuk mengevaluasi kinerja algoritma Support Vector Machine (SVM) sebelum dan sesudah optimasi menggunakan Random Search dalam klasifikasi depresi. Hasil pengujian menunjukkan bahwa model SVM standar telah memiliki performa yang baik dengan akurasi sebesar 95,70%, F1-score 95,72%, dan AUC 99,02%. Hal ini menunjukkan bahwa SVM mampu memisahkan kelas depresi dan tidak depresi dengan cukup efektif. Setelah dilakukan optimasi menggunakan Random Search, performa model meningkat menjadi akurasi 97,85%, F1-score 97,86%, dan AUC 99,66%. Peningkatan sebesar 2,15% ini menunjukkan bahwa pemilihan hiperparameter yang tepat berkontribusi terhadap peningkatan kemampuan model dalam mengklasifikasikan data. Meskipun peningkatan terlihat relatif kecil, hasil ini signifikan dalam konteks diagnosis medis karena berkaitan dengan penurunan kesalahan klasifikasi, khususnya false negative, sehingga lebih banyak kasus depresi dapat terdeteksi.

Berdasarkan confusion matrix, model hasil optimasi menunjukkan penurunan jumlah kesalahan prediksi dibandingkan model standar. Hal ini mengindikasikan bahwa Random Search berhasil menemukan kombinasi parameter optimal yang meningkatkan sensitivitas model terhadap kelas depresi. Dengan demikian, model menjadi lebih andal dalam mendukung deteksi dini. Jika dibandingkan dengan penelitian sebelumnya, hasil penelitian ini menunjukkan performa yang lebih tinggi. Namun, perbandingan tersebut tidak dapat dilakukan secara langsung karena perbedaan dataset, jenis fitur, dan karakteristik data yang digunakan. Dataset pada penelitian ini berbasis survei dengan variabel psikososial yang terstruktur, sehingga memiliki karakteristik yang berbeda dibandingkan dataset lain seperti data teks atau data klinis.

Selain itu, hasil analisis faktor seperti tekanan akademik, durasi tidur, dan riwayat pikiran bunuh diri berperan sebagai fitur dalam model SVM. Fitur-fitur tersebut telah melalui proses feature engineering dan normalisasi sebelum digunakan dalam pelatihan model. Dalam mekanisme SVM, kontribusi fitur tidak dinyatakan secara eksplisit dalam bentuk bobot, tetapi melalui pembentukan support vectors yang menentukan decision boundary. Oleh karena itu, fitur dengan relevansi tinggi secara tidak langsung meningkatkan kemampuan model dalam memisahkan kelas. Secara keseluruhan, hasil penelitian menunjukkan bahwa optimasi hiperparameter menggunakan Random Search efektif dalam meningkatkan kinerja SVM. Model yang dihasilkan tidak hanya memiliki akurasi tinggi, tetapi juga lebih sensitif dalam mendeteksi kasus depresi, sehingga berpotensi digunakan sebagai alat bantu dalam sistem deteksi dini gangguan kesehatan mental.

3.1 Evaluasi Model

Analisis awal terhadap distribusi data mengungkapkan karakteristik penting dari populasi sampel yang dapat mempengaruhi risiko depresi. Dari total 2.556 responden, distribusi kelas target menunjukkan ketidakseimbangan: 455 responden (17,8%) teridentifikasi mengalami depresi (Kelas 'Yes'), sedangkan 2.101 responden (82,2%) tidak mengalami depresi (Kelas 'No'). Ketidakseimbangan ini menegaskan pentingnya penggunaan metrik evaluasi seperti F1-Score dan AUC, bukan hanya akurasi, karena model yang bias ke kelas mayoritas bisa saja memiliki akurasi tinggi namun gagal mendeteksi depresi.

1. Status Pekerjaan:
   Terdapat disparitas ekstrem dalam prevalensi depresi berdasarkan status. Kelompok pelajar/mahasiswa (Student) menunjukkan tingkat depresi sebesar 50,2%, angka yang sangat mengkhawatirkan jika dibandingkan dengan kelompok profesional bekerja (Working Professional) yang hanya 9,9%. Temuan ini sejalan dengan penelitian Abdussamad et al. (2025) yang menyoroti kerentanan mahasiswa terhadap gangguan mental akibat tekanan akademik dan transisi kehidupan.
2. Tekanan Akademik (*Academic Pressure*):
   Data menunjukkan korelasi linear positif yang kuat antara tingkat tekanan akademik dan risiko depresi. Pada responden yang melaporkan tekanan akademik "Sangat Rendah" (skala 1), tingkat depresi hanya 17,2%. Angka ini melonjak drastis seiring peningkatan tekanan, mencapai 84,7% pada responden dengan tekanan akademik "Sangat Tinggi" (skala 5). Ini memvalidasi hipotesis bahwa beban studi adalah prediktor kuat depresi pada populasi muda.
3. Durasi Tidur (*Sleep Duration*):
   Pola tidur menunjukkan hubungan "U-shaped" dengan kesehatan mental. Responden dengan durasi tidur kurang dari 5 jam memiliki risiko depresi tinggi (25,4%), begitu pula mereka yang tidur berlebihan lebih dari 8 jam (24,3%). Responden dengan tidur normal (7-8 jam) memiliki risiko terendah. Ini mengkonfirmasi literatur medis bahwa gangguan tidur (insomnia atau hipersomnia) adalah gejala kardinal sekaligus pemicu depresi.
4. Riwayat Bunuh Diri (*Suicidal Thoughts*)
   Variabel ini menjadi indikator paling kritis. Responden yang mengaku pernah memiliki pikiran bunuh diri memiliki prevalensi depresi 28,8%, jauh lebih tinggi dibandingkan 7,3% pada mereka yang tidak. Dalam konteks *data mining*, variabel ini memiliki "information gain" yang tinggi untuk klasifikasi
5. Kepuasan Kerja/Belajar:
   Terdapat korelasi negatif yang konsisten; semakin tinggi tingkat kepuasan kerja (*Job Satisfaction*) atau kepuasan belajar (*Study Satisfaction*), semakin rendah tingkat depresi. Pada tingkat kepuasan kerja maksimal (skala 5), tingkat depresi turun hingga 5,3%.

Analisis korelasi (*heatmap*) antar fitur numerik memperlihatkan bahwa fitur-fitur seperti *Academic Pressure*, *Financial Stress*, dan *Work Pressure* memiliki korelasi positif yang signifikan terhadap label target *Depression*. Ini menunjukkan bahwa dataset yang digunakan memiliki kualitas fitur yang baik dan relevan secara psikologis untuk tugas prediksi.

3.2 Hasil Pelatihan Model SVM Standar

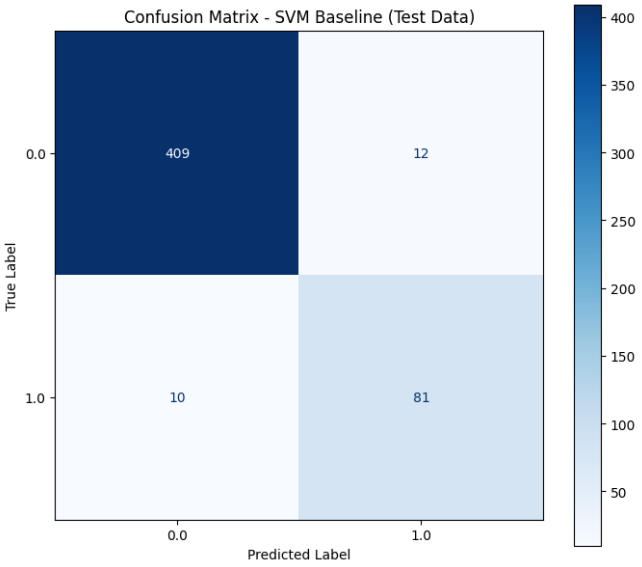
Pada tahap eksperimen pertama, model SVM dilatih menggunakan parameter standar (*default*) dari pustaka Scikit-learn (Kernel Linear, C=1.0, Gamma='scale'). Model dilatih pada 2.044 data latih dan diuji pada 512 data uji.

Hasil evaluasi pada data uji adalah sebagai berikut:

Tabel 2. Hasil Uji Tanpa Random Search (Standar)

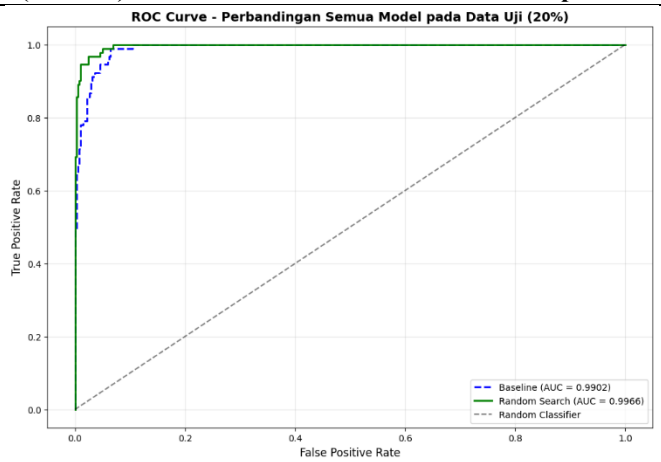
Model	Hasil
Accuracy	0.9570
Precision	0.9574
Recall	0.9570
F1-Score	0.9572

Berdasarkan *Confusion Matrix* model standar:



1. Gambar 1. Confusion Matrix SVM Standar

Gambar Confusion Matrix menunjukkan bahwa model SVM standar mampu mengklasifikasikan sebagian besar data dengan benar, yaitu 409 *True Negative* dan 81 *True Positive*. Namun, masih terdapat 12 *False Positive* dan 10 *False Negative*. Kesalahan *false negative* menjadi perhatian penting karena menunjukkan adanya kasus depresi yang tidak terdeteksi. Oleh karena itu, diperlukan optimasi model untuk meningkatkan sensitivitas dan mengurangi kesalahan klasifikasi.



2. Gambar 2. *ROC Curve* - Evaluasi Akhir (Data Uji 20%) Tanpa Random Search

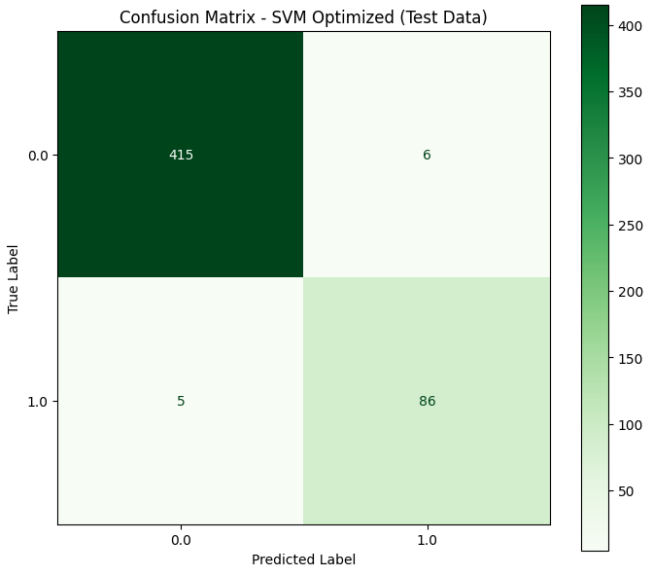
Model standar menunjukkan performa yang sangat baik dalam melakukan klasifikasi, dengan nilai ***Accuracy*** **sebesar 95,70%,** yang mengindikasikan tingkat ketepatan prediksi yang tinggi. Nilai ***Precision*** **sebesar 95,74%** dan ***Recall*** **sebesar 95,70%** menunjukkan bahwa model mampu mengidentifikasi kelas positif secara akurat dan konsisten. Hal ini diperkuat oleh **F1-Score sebesar 95,72%,** yang mencerminkan keseimbangan yang baik antara precision dan recall. Selain itu, nilai ***AUC*** **sebesar 99,02%** menandakan kemampuan model yang sangat kuat dalam membedakan kelas positif dan negatif, sehingga model standar dapat dikatakan memiliki kinerja yang andal.

3.3 Hasil Pelatihan Model SVM dengan Optimasi Random Search

Proses *Random Search* melakukan iterasi acak untuk menemukan kombinasi yang memaksimalkan skor akurasi pada validasi silang. Model terbaik yang dihasilkan dari proses ini kemudian dievaluasi kembali menggunakan data uji yang sama.
Hasil evaluasi Model SVM Teroptimasi (*Random Search SVM*):

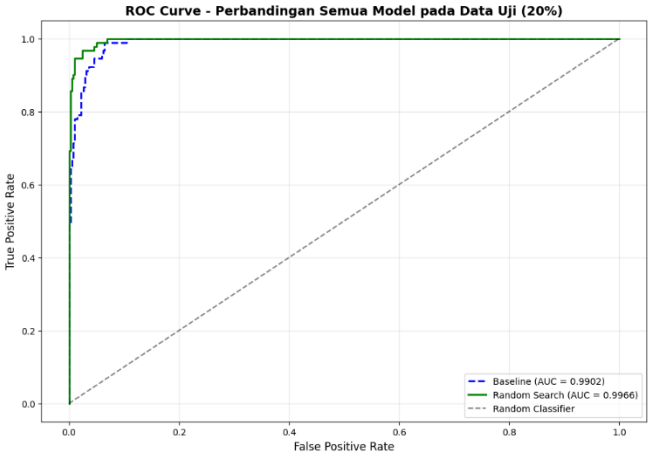
Tabel 3. Hasil Data Uji Dengan Random Search

Model	**Hasil**
Accuracy	0.9785
Precision	0.9786
Recall	0.9785
F1-Score	0.9786



Gambar 3. Confusion Matrix SVM Dengan Random Search

Gambar Confusion Matrix menunjukkan peningkatan kinerja model setelah optimasi Random Search, dengan 415 *True Negative* dan 86 *True Positive*. Kesalahan klasifikasi menurun menjadi 6 *False Positive* dan 5 *False Negative*. Penurunan *false negative* menunjukkan bahwa lebih banyak kasus depresi berhasil terdeteksi, yang sangat penting dalam konteks diagnosis dini. Secara praktis, model yang dioptimasi lebih andal dan aman digunakan karena mampu mengurangi risiko kesalahan deteksi.



Gambar 4. *ROC Curve* - Evaluasi Akhir (Data Uji 20%) Dengan Random Search

Model Model yang dioptimasi menggunakan ***Random Search*** menunjukkan peningkatan performa yang signifikan dibandingkan model standar. Model ini mencapai ***Accuracy*** **sebesar 97,85%,** ***Precision*** **97,86%,** ***Recall*** **97,85%,** dan ***F1-Score*** **97,86%,** yang menandakan peningkatan kemampuan prediksi dan konsistensi klasifikasi. Selain itu, nilai ***AUC*** **sebesar 99,66%** menunjukkan kemampuan yang sangat kuat dalam membedakan kelas positif dan negatif. Jika dibandingkan dengan model standar, terjadi peningkatan ***Accuracy*** **sebesar 2,15%,** ***Precision*** **sebesar 2,12%,** ***Recall*** **sebesar 2,15%,** ***F1-Score*** **sebesar 2,14%,** serta ***AUC*** **sebesar 0,64%.** Hasil ini membuktikan bahwa penerapan Random Search efektif dalam meningkatkan kinerja model secara menyeluruh pada seluruh metrik evaluasi.

Tabel 4. Perbandingan Performa Model SVM Standar vs SVM Random Search

Metric	Standard SVM	SVM *Random Search*	Peningkatan
Accuracy	95,70%	97,85%	+2,15%
Precision	95,74%	97,86%	+2,12%
Recall	95,70%	97,85%	+2,15%
F1-Score	95,72%	97,86%	+2,14%
AUC	99,02%	99,66%	+0,64%

Peningkatan performa model SVM setelah dioptimasi menggunakan Random Search terlihat secara konsisten pada seluruh metrik evaluasi, khususnya pada ***Accuracy* dan *AUC*.** Nilai ***AUC* sebesar 0,9966** menunjukkan bahwa model teroptimasi memiliki kemampuan diskriminasi yang hampir sempurna, dengan probabilitas **99,66%** dalam membedakan secara benar antara responden yang mengalami depresi dan yang tidak mengalami depresi. Hasil ini mengindikasikan bahwa optimasi hiperparameter berhasil meningkatkan kualitas batas keputusan (decision boundary) SVM secara signifikan.

Analisis ***Confusion Matrix*** menunjukkan bahwa penerapan optimasi Random Search pada algoritma SVM menghasilkan peningkatan kinerja klasifikasi yang jelas dibandingkan model standar. Pada **SVM standar**, jumlah ***True Negative* (TN)** tercatat sebanyak **409, *False Positive* (FP)** sebanyak **12, *False Negative* (FN)** sebanyak **10**, dan ***True Positive* (TP)** sebanyak **81**. Hasil ini menunjukkan bahwa meskipun model sudah cukup baik, masih terdapat sejumlah kasus depresi yang tidak terdeteksi (FN) serta kesalahan klasifikasi pada kelas negatif.

Setelah dilakukan optimasi menggunakan Random Search, performa model mengalami perbaikan signifikan. Jumlah *True Negative* (TN) meningkat menjadi 415, sementara *False Positive* (FP) berkurang menjadi 6. Selain itu, jumlah *False Negative* (FN) berhasil ditekan hingga 5 kasus, dan *True Positive* (TP) meningkat menjadi 86. Penurunan jumlah *False Negative* dari 10 menjadi 5 merupakan capaian yang sangat penting dalam konteks diagnosis depresi, karena mengurangi risiko individu dengan depresi yang tidak terdeteksi. Peningkatan *True Positive* juga menunjukkan bahwa model teroptimasi memiliki sensitivitas yang lebih baik dalam mengidentifikasi kasus depresi yang sebenarnya, sehingga lebih andal dan aman untuk digunakan sebagai alat bantu skrining dini.

3.4 Pembahasan

Hasil penelitian ini menunjukkan bahwa penerapan Random Search mampu meningkatkan kinerja algoritma SVM dalam diagnosis depresi, dengan peningkatan akurasi sebesar 2,15%. Meskipun peningkatan ini terlihat relatif kecil, dampaknya cukup signifikan dalam konteks klasifikasi medis, terutama karena disertai dengan penurunan jumlah *false negative*, sehingga lebih banyak kasus depresi dapat terdeteksi. Hal ini menunjukkan bahwa optimasi hiperparameter berperan penting dalam meningkatkan sensitivitas dan keandalan model. Jika dibandingkan dengan penelitian terdahulu, akurasi model sebesar 97,85% secara numerik lebih tinggi dibandingkan Aprilla et al. (2018) dan Rahayu et al. (2023). Namun, perbandingan ini perlu dilakukan secara hati-hati karena perbedaan dataset, jenis fitur, dan karakteristik data yang digunakan. Dataset pada penelitian ini berbasis survei dengan variabel psikososial yang terstruktur, sehingga tidak sepenuhnya sebanding dengan dataset lain seperti data teks atau data klinis.

Selain itu, hasil analisis faktor risiko seperti tekanan akademik, durasi tidur, dan riwayat pikiran bunuh diri tidak hanya berfungsi sebagai temuan deskriptif, tetapi juga digunakan sebagai fitur dalam model SVM. Dalam mekanisme SVM, kontribusi fitur tidak dinyatakan secara eksplisit dalam bentuk bobot yang mudah diinterpretasikan, melainkan melalui pembentukan *support vectors* dan *decision boundary*. Fitur-fitur yang memiliki daya diskriminatif tinggi akan lebih berpengaruh dalam menentukan batas pemisah antar kelas, sehingga secara tidak langsung berkontribusi terhadap peningkatan performa model.

4. Kesimpulan

Penelitian ini berhasil mengembangkan dan mengevaluasi sistem diagnosa penyakit depresi menggunakan algoritma *Support Vector Machine* (SVM) yang dioptimalkan dengan metode *Random Search*. Berdasarkan analisis hasil eksperimen, dapat ditarik beberapa kesimpulan utama:

1. Integrasi teknik optimasi *Random Search* terbukti sangat efektif dalam meningkatkan kinerja diagnostik algoritma SVM. Akurasi model meningkat secara signifikan dari 95,70% pada model standar menjadi 97,85% pada model teroptimasi. Peningkatan ini juga diikuti oleh kenaikan nilai F1-Score dan AUC yang mencapai 0,996, mendekati performa klasifikasi sempurna
2. Analisis data mengidentifikasi faktor-faktor risiko utama depresi, yaitu status sebagai pelajar/mahasiswa, tingkat tekanan akademik yang tinggi, riwayat pikiran bunuh diri, dan durasi tidur yang tidak normal. Temuan ini memberikan wawasan berharga bagi praktisi kesehatan mental untuk memfokuskan intervensi pada kelompok rentan ini.

a. Saran

Berdasarkan hasil tersebut, penelitian ini menyarankan pengembangan lebih lanjut ke arah implementasi sistem nyata. Model ini dapat dikemas menjadi aplikasi berbasis web atau *mobile* yang *user-friendly* untuk memfasilitasi skrining massal. Selain itu, penelitian masa depan disarankan untuk mengeksplorasi penggunaan dataset yang lebih besar dan beragam dari berbagai latar belakang demografis negara untuk menguji ketahanan model secara global, serta membandingkan hasil optimasi ini dengan metode *Deep Learning* terbaru untuk melihat potensi peningkatan akurasi lebih lanjut.

Referensi

Benedikt, L., Joshi, C., Nolan, L., De Wolf, N., & Schouten, B. (2020). *Optical Character Recognition and Machine Learning Classification of Shopping Receipts*. Abdussamad, S. N., Doholio, N. P., Lasaleng, W. P., Usia, P. A. I. N., Rahman, M. I., & Adam, D. P. J. 2025. Klasifikasi Tingkat Depresi Mahasiswa Menggunakan Image Recognition dengan Support Vector Machine. *Research in the Mathematical and Natural Sciences*, 4(1), 30–36.

Abimanyu, S., Bahtiar, N., & Sarwoko, E. A. 2023. Implementasi Metode Support Vector Machine (SVM) dan t-Distributed Stochastic Neighbor Embedding (t-SNE) untuk Klasifikasi Depresi. *Jurnal Masyarakat Informatika*, 14(2), 146–158.

Aleem, S., Huda, N., Amin, R., Khalid, S., Alshamrani, S. S., & Alshehri, A. 2022. Machine Learning Algorithms for Depression: Diagnosis, Insights, and Research Directions. *Electronics*, 11(7), 1111.

Aziz, F., & Abasa, S. 2025. Pengembangan dan Validasi Model Hybrid Machine Learning untuk Diagnosis Awal Depresi. *Journal Pharmacy and Application of Computer Sciences*, 3(1), 8–15.

Aziz, F., Ishak, P., & Abasa, S. 2024. Klasifikasi Depresi Menggunakan Support Vector Machine: Pendekatan Berbasis Data Text Mining. *Journal Pharmacy and Application of Computer Sciences*, 2(2), 33–38.

Chen, E., & Asta, M. 2022. Using Jupyter Tools to Design an Interactive Textbook to Guide Undergraduate Research in Materials Informatics. *Journal of Chemical Education*, 99(10), 3601–3606.

Dharmawan, W. S. 2021. Komparasi Algoritma Klasifikasi SVM-PSO dan C4.5-PSO dalam Prediksi Penyakit Jantung. *Jurnal Informatika, Manajemen dan Komputer*, 13(2), 31–41.

Garg, K., Thoma, A., Avramovic, G., Gilbert, L., Shawky, M., Ray, M. R., & Lambert, J. S. 2024. Biomarker-Based Analysis of Pain in Patients with Tick-Borne Infections before and after Antibiotic Treatment. *Antibiotics*, 13(8), 693.

Hahn, A. W., Zsombor-Pindera, J., Kennepohl, P., & DeBeer, S. 2024. Introducing SpectraFit: An Open-Source Tool for Interactive Spectral Analysis. *ACS Omega*, 9(22), 23252–23265.

Haque, U. M., Kabir, E., & Khanam, R. 2021. Detection of child depression using machine learning methods. *PLoS ONE*, 16(12), e0261131.

Harahap, H. S., Indrayana, Y., Putra, H. S., & Retnowati, I. 2021. Deteksi Dini dan Edukasi Pentingnya Penatalaksanaan Depresi dan Faktor-Faktor Risikonya pada Pasien Stroke Iskemik di RSUD Provinsi Nusa Tenggara Barat. *Jurnal Pengabdian Masyarakat Sains Indonesia*, 3(1).

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D.,... & Oliphant, T. E. 2020. Array programming with NumPy. *Nature*, 585(7825), 357–362.

Indrika, S. N., & Chris, A. 2022. Hubungan tingkat depresi dengan kekuatan genggaman tangan pada mahasiswa fakultas kedokteran universitas tarumanagara. *Ebers Papyrus*, 28(1), 74–82.

Ismail, A., Eid, A., Mjlae, S. A., Yaseen, S., Ali, M., & Rababah, I. 2025. Comparative Analysis of Machine Learning Libraries for Neural Networks: A Benchmarking Study. *bioRxiv*, 2025.02.02.635632.

Ismiralda, J. F., Hayati, L. N., & Umar, F. 2022. Sistem Pakar Penentuan Tingkat Depresi Pada Ibu Hamil Menggunakan Certainty Factor Berbasis Web. *Buletin Sistem Informasi dan Teknologi Islam*, 3(4), 332–340.

Khoong, W. H. 2020. DEBoost: A Python Library for Weighted Distance Ensembling in Machine Learning. *Preprints*, 2020050354.

Lavanya, A., Sindhuja, S., Gaurav, L., & Ali, W. 2023. A Comprehensive Review of Data Visualization Tools: Features, Strengths, and Weaknesses. *International Journal of Computer Engineering in Research Trends*, 10(1), 10–20.

Lee, Y. E., & Biromo, A. R. 2024. Depresi dan kerentanan lansia di panti wreda wilayah Jabodetabek. *Tarumanagara Medical Journal*, 6(1), 30813.

Linge, S., & Langtangen, H. P. 2020. *Programming for Computations - Python: A Gentle Introduction to Numerical Simulations with Python 3.6*. Cham: Springer Nature.

Mantovani, R. G., Rossi, A. L. D., Vanschoren, J., Bischl, B., & De Carvalho, A. C. P. L. F. 2015. Effectiveness of Random Search in SVM hyper-parameter tuning. *Proceedings of the 2015 International Joint Conference on Neural Networks (IJCNN)*, 1–8.

Martins, S. G. 2022. Car Race - An Engaging Interdisciplinary Python Based Project Activity. *Journal of Engineering Education Transformations*, 35(3), 45–51.

Mehdizadeh, S. A., Noshad, M., Chaharlangi, M., & Ampatzidis, Y. 2023. Development of an Innovative Optoelectronic Nose for Detecting Adulteration in Quince Seed Oil. *Foods*, 12(23), 4350.

Nwulu, N. I., Damisa, U., & Gbadamosi, S. L. 2021. Students Perception about the Use of Jupyter Notebook in Power Systems Education. *International Journal of Engineering Pedagogy*, 11(1), 78–86.

Pablo, F., Lopez, A., Privada, U., & Privada, U. 2022. A Low-Cost Inclinometer with Data Acquisition Library Developed in Phyton, for In-Situ Displacement Measurements. *Proceedings of the COBRAE 2022*.

Pakan, R., & Mailoa, E. 2024. Komparasi Kinerja Algoritma Mechine Learning Untuk Mengklasifikasi Penyakit Kanker Payudara. *Progresif: Jurnal Ilmiah Komputer*, 20(1), 493–503.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O.,... & Duchesnay, E. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Rahayu, K., Fitria, V., Septhya, D., Rahmaddeni, R., & Efrizoni, L. 2023. Klasifikasi Teks untuk Mendeteksi Depresi dan Kecemasan pada Pengguna Twitter Berbasis Machine Learning. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(2), 108–114.

Ramadhan, A. G. 2023. Data Mining untuk Segmentasi Pelanggan dengan Algoritma K-Means: Studi Kasus pada Data Pelanggan di Toko Retail. *Syntax Literate: Jurnal Ilmiah Indonesia*, 8(10), 5698–5715.

Rifatama, M. I., Faisal, M. R., Herteno, R., Budiman, I., & Mazdadi, M. I. 2023. Optimasi Algoritma K-Nearest Neighbor Dengan Seleksi Fitur Menggunakan Xgboost. *Jurnal Informatika Dan Rekayasa Elektronik*, 6(1), 64–72.

Rizty, L. E., & Kusumiati, R. Y. E. 2020. Hubungan Dukungan Sosial (Suami) dengan Kecenderungan Depresi Postpartum. *Jurnal Ilmiah Bimbingan Konseling Undiksha*, 11(2), 112–118.

Santoso, M. B., Asiah, D. H. S., & Kirana, C. I. 2018. Bunuh Diri Dan Depresi Dalam Perspektif Pekerjaan Sosial. *Prosiding Penelitian Dan Pengabdian Kepada Masyarakat*, 4(3), 390.

Sharma, A., & Verbeke, W. J. M. I. 2020. Improving Diagnosis of Depression With XGBOOST Machine Learning Model and a Large Biomarkers Dutch Dataset. *Frontiers in Big Data*, 3, 15.

Sulistyorini, W., & Sabarisman, M. 2017. Depresi: Suatu Tinjauan Psikologis. *Sosio Informa*, 3(2), 153–164.

Talaei Khoei, T., & Kaabouch, N. 2023. Machine Learning: Models, Challenges, and Research Directions. *Future Internet*, 15(10), 332.

Waskom, M. L. 2021. seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60), 3021.

WHO. 2023. *Depressive disorder (depression)*. Diakses dari https://www.who.int/news-room/fact-sheets/detail/depression

Wu, W. T., Li, Y. J., Feng, A. Z., Li, L., Huang, T., Xu, A. D., & Lyu, J. 2021. Data mining in clinical big data: the frequently used databases, steps, and methodological models. *Military Medical Research*, 8(1), 44.

Wulandari, A. 2020. Aplikasi Support Vector Machine (SVM) untuk Pencarian Binding Site Protein-Ligan. *Mathunesa: Jurnal Ilmiah Matematika*, 8(2), 157–161.