JITET (Jurnal Informatika dan Teknik Elektro Terapan)
Vol. xx No. xx, xxxxxx xxx (xxx-xxx) https://doi.org/10.23960/jitet

KLASIFIKASI KESEHATAN MENTAL PADA USIA REMAJA MENGGUNAKAN METODE SVM

Elsa Eka Pratiwi[1*], Alaysha Rihadatul Aisy[2], Rahmaddeni[3], Nita Ananta[4]

[1,2,3,4]Teknik Informatika, Universitas Sains dan Teknologi Indonesia; Jl. Purwodadi Indah No.KM. 10, Pekanbaru, Riau

Received: xxxx-xx-xx
Accepted: xx-xx-xx

Keywords:
Kesehatan mental;
support vector machine;
SMOTE.

Corespondent Email:
rahmaddeni@usti.ac.id

Abstrak. *Kesehatan mental merupakan faktor krusial dalam kehidupan individu, khususnya pada remaja yang rentan mengalami gangguan mental akibat tekanan hidup. Penelitian ini bertujuan untuk mengklasifikasikan kondisi kesehatan mental pada remaja dengan menerapkan metode SVM. Data yang digunakan dalam penelitian ini bersumber dari dataset kesehatan mental. Proses analisis mencakup tahap preprocessing data, eksplorasi data, penerapan teknik oversampling menggunakan SMOTE, serta optimasi model SVM melalui Grid Search Cross Validation. Hasil penelitian menunjukkan bahwa model SVM memberikan performa optimal dalam mengklasifikasikan kesehatan mental. Dengan pembagian data sebesar 60:40, model memperoleh akurasi sebesar 79%. Precision untuk kelas 0 mencapai 0.79, sementara kelas 1 sebesar 0.80, menunjukkan tingkat ketepatan model yang cukup baik dalam mengidentifikasi setiap kategori. Selain itu, recall untuk kelas 0 sebesar 0.83, mengindikasikan bahwa model mampu mendeteksi sebagian besar data yang benar-benar termasuk dalam kategori tersebut. Penelitian ini membuktikan bahwa metode SVM efektif dalam mengidentifikasi kondisi kesehatan mental pada remaja dan berpotensi menjadi alat pendukung dalam sistem deteksi dini gangguan mental.*

Abstract. *Mental health is a crucial factor in an individual's life, especially for adolescents who are vulnerable to mental disorders due to life pressures. This study aims to classify adolescent mental health conditions by applying the SVM method. The data used in this study is sourced from a mental health dataset. The analysis process includes data preprocessing, data exploration, the application of oversampling techniques using SMOTE, and SVM model optimization through Grid Search Cross Validation. The research results indicate that the SVM model delivers optimal performance in classifying mental health. With a data split of 60:40, the model achieves an accuracy of 79%. The precision for class 0 reaches 0.79, while class 1 achieves 0.80, demonstrating the model's relatively high accuracy in identifying each category. Additionally, the recall for class 0 is 0.83, indicating that the model effectively detects the majority of data that truly belong to this category. This study proves that the SVM method is effective in identifying adolescent mental health conditions and has the potential to serve as a supporting tool in early detection systems for mental disorders.*

1. PENDAHULUAN

Kesehatan mental memiliki peran yang sangat penting dalam kehidupan individu. Dengan kondisi mental yang baik, seseorang dapat menjalankan aktivitasnya sebagai makhluk hidup. Keadaan mental yang sehat juga berkontribusi dalam perkembangan individu menuju arah yang lebih positif di masa depan [1]. Kesehatan mental merujuk pada kondisi di mana seseorang mampu mengenali potensi dirinya, menghadapi tekanan hidup yang wajar, bekerja dengan produktif, serta

memberikan kontribusi terhadap lingkungan sekitarnya [2]. Sementara itu, gangguan kesehatan mental dimaknai sebagai ketidak mampuan individu dalam menyesuaikan diri terhadap tuntutan serta kondisi lingkungan, yang pada akhirnya menyebabkan hambatan tertentu. Salah satu permasalahan kesehatan mental yang sering dialami oleh remaja seperti persoalan dalam pertemanan.

Sejalan dengan Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2020-2024, Pemerintah Indonesia melalui Kementerian Kesehatan Republik Indonesia (Kemenkes RI) menjadikan kesehatan mental sebagai salah satu fokus utama dalam program kesehatan nasional. Kesehatan mental memiliki peran penting dalam mendukung pembangunan negara dan pengembangan sumber daya manusia, terutama generasi muda yang akan menjadi penentu masa depan Indonesia serta berkontribusi dalam mewujudkan Visi Indonesia Emas 2045. Berdasarkan laporan I-NAMHS, sejumlah besar remaja mengalami permasalahan kesehatan mental, dengan sekitar satu dari tiga remaja (34,9%) mengalami gangguan tersebut dalam 12 bulan terakhir [3]. Proses klasifikasi kondisi kesehatan mental pada individu dewasa muda bersifat kompleks dan membutuhkan metode yang sistematis serta presisi tinggi. Pendekatan tradisional dalam proses pengkategorian sering kali mengandalkan penilaian subjektif dan keahlian klinis, yang rentan terhadap bias serta ketidakakuratan. Dalam beberapa tahun terakhir, kemajuan teknologi komputasi serta penerapan metode pembelajaran mesin semakin menarik perhatian dalam penelitian mengenai kesehatan mental.

Salah satu algoritma yang umum digunakan dalam pembelajaran mesin adalah *Support Vector Machine* (SVM). Algoritma ini memiliki keunggulan dalam mengolah data berukuran besar, mengenali pola *non-linear*, serta menghasilkan klasifikasi yang lebih akurat [4]. Penelitian ini bertujuan untuk mengembangkan model klasifikasi kesehatan mental pada kelompok dewasa muda dengan memanfaatkan teknik SVM. Diharapkan model ini dapat membantu tenaga medis serta peneliti dalam mengidentifikasi individu yang berisiko mengalami gangguan kesehatan mental, sekaligus merancang strategi pengobatan yang lebih efektif serta tepat sasaran.

Penelitian sebelumnya oleh [4], yang berjudul "Aplikasi Penyaringan Spam e-Mail Menggunakan Teknik *Machine Learning* dengan Metode *Support Vector Machines*" menunjukkan bahwa penggunaan *Support Vector Machine* mampu mencapai akurasi sebesar 95%. Sementara itu, studi berikutnya oleh [5], dalam peneIitian berjudul "Prediksi Analisis Sentimen Data Debat Pemilihan Presiden 2024 Menggunakan *Support Vector Machine* (SVM)" memperoleh tingkat akurasi pengujian sebesar 94%. Penelitian Iain yang dilakukan oleh [6], dengan judul "Penerapan Algoritma *Support Vector Machine* (SVM) dengan TF-IDF N-Gram untuk *Text Classification"* membagi data *train-test split* ke dalam empat skenario, yakni 60:40, 70:30, 80:20, dan 90:10. Hasil terbaik yang diperoleh adalah akurasi sebesar 70% pada skenario 90:10. Studi lainnya yang dilakukan oleh [7], berjudul "*Support Vector Machine* untuk Pengenalan Bentuk Manusia Menggunakan Kumpulan Fitur yang Dioptimalkan" mencatat hasil akurasi klasifikasi terbaik sebesar 85,8%. Berdasarkan permasalahan tersebut, penelitian ini akan mencoba menerapkan algoritma *Support Vector Machine* (SVM) dalam klasifikasi kesehatan mental pada kelompok usia remaja.

2. TINJAUAN PUSTAKA

2.1. Kesehatan Mental

Secara umum, kesehatan mental mengacu pada kematangan seseorang dalam aspek emosional dan sosial, yang memungkinkannya untuk beradaptasi dengan dirinya sendiri dan lingkungan sekitarnya, serta kemampuan untuk menanggung tanggung jawab dalam hidup dan siap menghadapi berbagai masalah yang muncul.

Kesehatan mental manusia dipengaruhi oleh dua faktor utama, yaitu faktor internal dan eksternal. Faktor internal berasal dari dalam diri individu, seperti sifat, bakat, dan keturunan. Sementara itu, faktor eksternal berasal dari luar diri seseorang, seperti lingkungan, keluarga, interaksi sosial, budaya, dan agama. Faktor eksternal yang positif dapat membantu menjaga kesehatan mental seseorang, sedangkan faktor eksternal yang negatif dapat menyebabkan gangguan mental. Namun, banyak orang cenderung lebih mudah terpengaruh oleh faktor internal dan eksternal yang buruk, yang pada

akhirnya dapat mengakibatkan kondisi mental yang tidak sehat [8].

2.2. *Machine Learning*

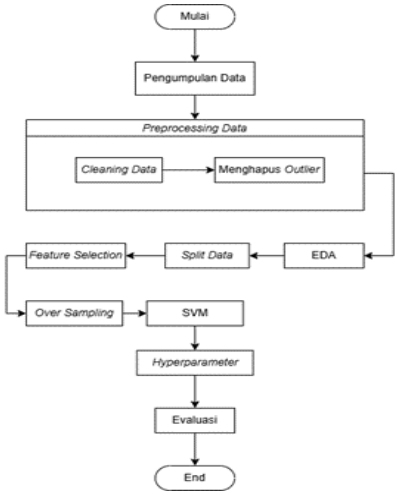
Machine Learning adalah bidang ilmu yang mempelajari cara membuat kompute mampu belajar dan meningkatkan kemampuannya dengan meniru atau bahkan melampaui kemampuan belajar manusia [9]. Teknik pembelajaran mesin otomatis dapat mempermudah proses ini dengan secara otomatis memilih model, mengoptimalkan *hyperparameter*, dan mengurangi ketergantungan pada intervensi manusia. Pendekatan ini juga dapat meringankan beban kerja ahli radiologi dan patologi, serta menghemat waktu dan tenaga dalam proses diagnosis data pencitraan medis [10].

2.3. *Support Vector Machine* (SVM)

Support Vector Machine (SVM) adalah algoritma yang bertujuan untuk menemukan *hyperplane* yang dapat memisahkan titik data dari kelas-kelas yang berbeda dengan jelas [6]. Metode ini merupakan sebuah *framework* penguat gradien yang memanfaatkan algoritma pembelajaran berbasis pohon (*tree-based learning*). *Framework* ini dirancang untuk bersifat terdistribusi dan efisien, dengan keunggulan seperti kecepatan pelatihan yang lebih tinggi, penggunaan memori yang lebih rendah, akurasi yang lebih baik, dukungan untuk pembelajaran paralel dan terdistribusi, kompatibilitas dengan GPU, serta kemampuan untuk menangani data dalam skala besar [11].

3. METODE PENELITIAN

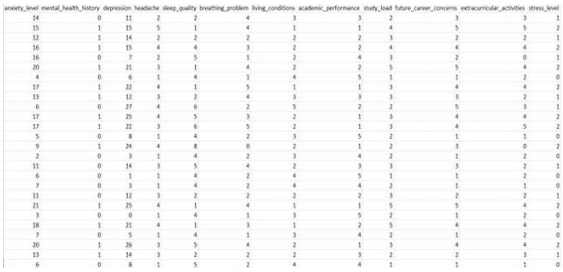
Penelitian ini memanfaatkan kerangka pembelajaran mesin guna mengklasifikasikan kesehatan mental pada pasien dewasa muda menggunakan teknik *Support Vector Machine* (SVM). Rincian tahap-tahap penelitian disajikan seperti yang ditampilkan pada Gambar 1.



Gambar 1 Metodologi penelitian

3.1. Pengumpulan Data

Tahap awal dalam penelitian ini dimulai dengan proses pengumpulan data, di mana langkah ini dilakukan dengan mengumpulkan informasi yang berkaitan dengan kesehatan mental. Dalam penelitian ini, dataset yang digunakan berasal dari *platform* GitHub dengan total 1.100 baris data dan 12 kolom.



Gambar 2 Dataset kesehatan mental

3.2. *Preprocessing* Data

Preprocessing data merupakan tahap penting dalam *pipeline data science* yang memiliki dampak besar terhadap hasil akhir analisis [12]. Langkah ini bertujuan untuk memastikan bahwa data yang digunakan bersih, konsisten, serta siap dianalisis atau dimodelkan. Proses *preprocessing data* mencakup berbagai tahapan yang dirancang untuk menangani permasalahan seperti data yang hilang, *noise*, inkonsistensi, atau format yang tidak sesuai.

3.2.1. *Cleaning Data*

Cleaning data merupakan tahap krusial yang harus diperhatikan dalam proses analisis data serta pemodelan *machine learning*. Dengan menghilangkan nilai yang hilang, duplikasi, dan

pencilan, dataset menjadi lebih rapi, konsisten, serta siap untuk dianalisis lebih lanjut. Akibatnya, analisis statistik menjadi lebih akurat, dan model *machine learning* dapat bekerja dengan lebih andal [13]. Pada atribut yang telah dipilih, tidak ditemukan nilai kosong, seperti pada Gambar 3.

```
[3] df.isnull().sum()
```

	0
anxiety_level	0
mental_health_history	0
depression	0
headache	0
sleep_quality	0
breathing_problem	0
living_conditions	0
academic_performance	0
study_load	0
future_career_concerns	0
extracurricular_activities	0
stress_level	0

Gambar 3 Atribut *missing values*

3.2.2. Menghapus *Outlier*

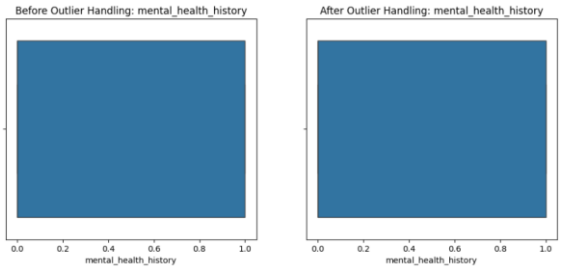
Tahap berikutnya adalah pemeriksaan *outlier* dalam dataset. Fungsi dari tahap ini adalah untuk meminimalkan pengaruh *outlier* ekstrem yang berpotensi menyebabkan bias pada model, menjaga keseimbangan distribusi data, serta meningkatkan kinerja model prediktif [14].

```
# 2. Preprocessing Data
## Cleaning Data: Drop Missing Values & Duplicates
df = df.dropna()
df = df.drop_duplicates()

## Handling Outliers using IQR
for col in df.select_dtypes(include=[np.float64, np.int64]).columns:
    Q1 = df[col].quantile(0.25)
    Q3 = df[col].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    df[col] = np.where(df[col] < lower_bound, lower_bound, df[col])
    df[col] = np.where(df[col] > upper_bound, upper_bound, df[col])

# Visualisasi distribusi data sebelum dan sesudah mengatasi outlier
for col in df.select_dtypes(include=[np.float64, np.int64]).columns:
    plt.figure(figsize=(12,5))
    plt.subplot(1,2,1)
    sns.boxplot(x=df[col])
    plt.title(f'Before Outlier Handling: {col}')

    plt.subplot(1,2,2)
    sns.boxplot(x=df[col])
    plt.title(f'After Outlier Handling: {col}')
    plt.show()
```

Gambar 4 Mengatasi *outlier*

3.3. *Exploratory Data Analysis* (EDA)

Exploratory Data Analysis (EDA) merupakan proses mengeksplorasi dan menganalisis dataset bertujuan untuk memahami karakteristik, pola, serta hubungan antar variabel baik secara visual maupun statistik. EDA memiliki tujuan utama untuk memperoleh wawasan (*insight*) dari data, mengidentifikasi permasalahan seperti *missing values* atau *outlier*, serta mempersiapkan data untuk analisis atau pemodelan lebih lanjut [15]. Penanganan *outlier* merupakan tahap krusial dalam *preprocessing data* guna memastikan bahwa analisis statistik dan model *machine learning* didasarkan pada data yang bersih serta representatif. Dengan menghapus *outlier*, distribusi data menjadi lebih stabil, sehingga meningkatkan performa model serta akurasi dalam analisis statistik.



Gambar 5 Proses setelah menghapus *outlier*

3.4. *Splitting Data*

Pemisahan data (*data splitting*) merupakan salah satu tahap krusial dalam analisis data serta pengembangan model *machine learning*. Tujuan dari proses ini adalah membagi dataset menjadi beberapa subset yang dimanfaatkan untuk pelatihan, validasi, serta pengujian model. Pemisahan data atau *data splitting* dilakukan dengan pembagian dataset menjadi dua bagian, yakni data latih dan data uji [16]. Data latih berperan dalam proses pelatihan model, sedangkan data uji digunakan untuk mengevaluasi performa model yang telah dilatih. Proses pemisahan data ini dilakukan dengan berbagai proporsi, seperti 90%:10%, 80%:20%, 70%:30%, dan 60%:40%, yang bertujuan untuk membandingkan hasil akurasi terbaik.

3.5. *Feature Selection*

Setelah tahap *preprocessing data* selesai, proses selanjutnya adalah melakukan *label encoding* atau mengubah data kategori menjadi format numerik. *Label encoding* ialah metode dalam pemrosesan data yang berfungsi untuk mengonversi nilai-nilai dalam satu kolom kategori ke dalam bentuk numerik atau label.

Semua teks yang berbentuk label dikonversi menjadi nilai numerik [17]. Sementara itu, atribut pada peneIitian ini dapat dilihat pada Tabel 1.

Tabel 1 Atribut yang digunakan

Variabel X	Variebel Y
'stress_level', 'anxiety_level', 'depression', 'headache', 'future_career_concerns', 'extracurricular_activities', 'study_load', 'breathing_problem', 'living_conditions', 'sleep_quality', 'academic_performance'	'mental_health _history'

Pada atribut "mental_health_history" terdapat 2 label yaitu 0 dan 1 yang artinya label 0 yaitu "sehat mental" dan 1 yaitu "gangguan mental".

3.6. *Oversampling*

Oversampling dengan SMOTE (*Synthetic Minority Over-sampling Technique*) ialah metode yang diterapkan untuk mengatasi ketidak seimbangan kelas (*imbalanced dataset*) dalam machine learning [18]. Metode ini berfungsi menambahkan data sintetis pada kelas minoritas guna menyeimbangkan distribusi kelas, sehingga model machine learning dapat belajar lebih optimal tanpa cenderung bias terhadap kelas mayoritas.

```
# 7. Over Sampling menggunakan SMOTE
smote = SMOTE(random_state=42)
X_train_resampled, y_train_resampled = smote.fit_resample(X_train, y_train)

#Sebelum menggunakan SMOTE
print(y_train.value_counts(), "\n")

#Setelah menggunakan SMOTE
print(pd.Series(y_train_resampled).value_counts())
```

```
mental_health_history
0    395
1    315
Name: count, dtype: int64

mental_health_history
1    395
0    395
Name: count, dtype: int64
```

Gambar 6 *Oversampling* menggunakan SMOTE

Kelas 0 terdiri dari 395 sampel, sedangkan Kelas 1 memiliki 315 sampel. Hal ini menunjukkan bahwa dataset tidak seimbang karena Kelas 0 lebih dominan dibandingkan Kelas 1. Setelah menerapkan SMOTE, Kelas 1, yang sebelumnya merupakan kelas minoritas, telah di *over-sampling* hingga mencapai 395 sampel, sementara Kelas 0 tetap memiliki 395 sampel. Dengan demikian, dataset kini menjadi seimbang, karena kedua keIas memiliki jumlah sampel yang sama.

3.7. *Support Vector Machine* (SVM)

Implementasi algoritma *Support Vector Machine* (SVM) dalam penelitian ini menggunakan metode SVC (*Support Vector Classification*) dengan *kernel linear*. *Kernel linear* dipilih karena kemampuannya dalam menangani data yang dapat dipisahkan secara linier, sehingga cocok untuk kasus klasifikasi di mana hubungan antara fitur dan label dapat dijelaskan dengan garis lurus. Parameter `probability=True` diatur agar model dapat memperkirakan probabilitas kelas, yang berguna untuk evaluasi kinerja model melalui metrik seperti ROC-AUC.

Proses pelatihan model dilakukan dengan menggunakan data yang telah di-*resample* (`X_train_resampled` dan `y_train_resampled`) untuk mengatasi ketidakseimbangan kelas. *Resampling* ini memastikan bahwa model tidak bias terhadap kelas mayoritas, sehingga dapat meningkatkan generalisasi dan akurasi prediksi pada data yang tidak terlihat.

Dengan implementasi ini, diharapkan model SVM dapat memberikan performa yang baik dalam mengklasifikasikan data, sekaligus memberikan estimasi probabilitas yang akurat untuk setiap kelas. Hal ini penting untuk analisis lebih lanjut dan interpretasi hasil klasifikasi dalam konteks penelitian yang lebih luas.

```
# 8. Implementasi Algoritma SVM
svm = SVC(kernel='linear', probability=True)
svm.fit(X_train_resampled, y_train_resampled)
```

```
SVC
SVC(kernel='linear', probability=True)
```

Gambar 7 Implementasi algoritma SVM

3.8. *Hyperparameter*

Peran *Hyperparameter* tuning memiliki signifikansi yang besar. *Hyperparameter tuning* merupakan prosedur yang terstruktur untuk mengoptimalkan nilai-nilai *Hyperparameter* demi meningkatkan kinerja model. Proses ini mencakup eksplorasi menyeluruh terhadap ruang *Hyperparameter* yang luas guna

menemukan kombinasi parameter yang memberikan performa terbaik pada dataset yang digunakan [19].

Gambar 8 *Hyperparameter* dengan GridSearchCV

GridSearchCV merupakan metode yang dipakai untuk menemukan kombinasi *hyperparameter* optimal dalam model *machine learning.* Proses ini dilakukan dengan menguji seluruh kombinasi *hyperparameter* yang tersedia dan menentukan yang menghasilkan performa terbaik, biasanya berdasarkan skor validasi [20].

3.9. Evaluasi

Selanjutnya, model dinilai berdasarkan kinerjanya menggunakan metrik seperti akurasi, precision, recall, dan F1-score [21]. Model yang telah selesai dilatih akan diuji dengan data pengujian. Tahap evaluasi bertujuan untuk mengukur sejauh mana model mampu melakukan prediksi dengan baik. Berbagai metrik, termasuk akurasi, *precision*, *recall*, dan *F1-score*, yang memanfaatkan *confusion matrix*, digunakan untuk menilai performa prediksi. Selain itu, validasi silang dilakukan dalam tahap evaluasi ini guna memperjelas kinerja model serta memberikan gambaran yang lebih menyeluruh.

4. HASIL DAN PEMBAHASAN

4.1. Implementasi *Support Vector Machine*

Dalam penelitian ini, algoritma yang diterapkan adalah *Support Vector Machine* (SVM) untuk mengklasifikasikan kesehatan mental pada remaja. Proses klasifikasi dilakukan dengan berbagai skenario pembagian data, yaitu 90%:10%, 80%:20%, 70%:30%, dan 60%:40%, sebagaimana tercantum dalam Tabel 2 dan Tabel 3.

Selain itu, hasil klasifikasi menunjukkan jumlah individu dalam setiap kategori kesehatan mental yang diprediksi oleh model, seperti yang ditampilkan dalam Tabel 2.

Tabel 2. Hasil Klasifikasi Kesehatan Mental dengan SVM

Splitting Data	Sehat Mental (0)	Gangguan Mental (1)
90:10	135 individu	265 individu
80:20	154 individu	225 individu
70:30	173 individu	206 individu
60:40	143 individu	215 individu

Berdasarkan tabel di atas, model SVM mampu mengelompokkan individu ke dalam dua kategori utama, yaitu sehat mental (0) dan gangguan mental (1). Pola distribusi prediksi mengindikasikan bahwa penggunaan lebih banyak data latih meningkatkan kemampuan model dalam mengidentifikasi individu dengan gangguan mental.

Setelah memperoleh hasil klasifikasi, model dievaluasi menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score* untuk setiap skenario pembagian data. Tabel 3 menyajikan hasil evaluasi performa model berdasarkan metrik tersebut.

Tabel 3 Hasil evaluasi SVM dengan beberapa splitting data

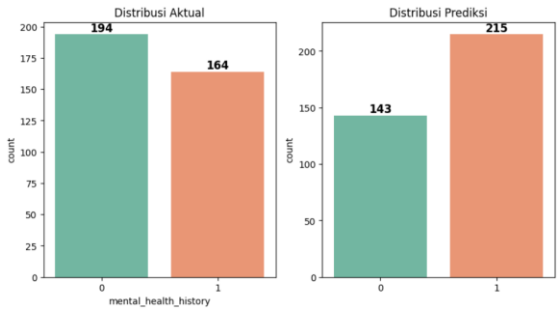
Splitting Data	Label	Precision	Recall	F1-Score	Akurasi
90:10	0	0,82	0,69	0,75	0,73
	1	0,64	0,78	0,70	
80:20	0	0,74	0,81	0,77	0,75
	1	0,77	0,69	0,72	
70:30	0	0,84	0,72	0,78	0,78
	1	0,74	0,85	0,79	
60:40	0	0,79	0,83	0,81	0,79
	1	0,80	0,74	0,77	

Akurasi model mengalami peningkatan seiring dengan semakin besarnya proporsi data latih (*training data*). *Splitting* 60:40 memberikan akurasi tertinggi sebesar 0,79, sedangkan splitting 90:10 menghasilkan akurasi terendah sebesar 0,73. Hal ini menunjukkan bahwa menambahkan lebih banyak data uji (*testing data*) tidak selalu meningkatkan

performa model. *Precision* dan *recall* pada kelas 0 (mental sehat) serta kelas 1 (gangguan mental) menunjukkan variasi yang berbeda di setiap proporsi *splitting data*. Pada *splitting* 60:40, nilai F1-Score tertinggi untuk kelas 0 dan 1 masing-masing adalah 0,81 dan 0,77, yang menunjukkan keseimbangan terbaik antara *Precision* dan *recall*. Namun, pada *splitting* 90:10, *recall* pada kelas 1 lebih tinggi (0,78) dibandingkan *splitting* lainnya, tetapi *precision* kelas 1 lebih rendah (0,64), yang berarti model lebih banyak menangkap kasus positif namun dengan tingkat kesalahan yang lebih tinggi. *Splitting* 70:30 memberikan keseimbangan yang baik antara *precision*, *recall*, dan F1-Score, dengan akurasi 0,78. *Splitting* 60:40 memiliki akurasi tertinggi (0,79), namun *recall* kelas 1 lebih rendah dibandingkan *splitting* 70:30, yang berarti model pada *splitting* 60:40 sedikit kurang optimal dalam menangkap semua kasus gangguan mental. *Splitting* 70:30 dapat dianggap sebagai pilihan yang paling optimal karena menghasilkan *recall* yang tinggi pada kelas 1 (0,85), sehingga model lebih baik dalam mendeteksi individu yang mengalami gangguan mental.

4.2. Evaluasi Model

Setelah dilakukan proses pemodelan algoritma SVM, tahap selanjutnya adalah evaluasi model. Tujuan dari tahapan ini adalah untuk memastikan bahwa model tersebut dapat bekerja dengan baik dan memenuhi tujuan penelitian. Evaluasi model membantu peneliti memahami sejauh mana model dapat digeneralisasi, serta mengidentifikasi kelebihan dan kekurangan model tersebut.

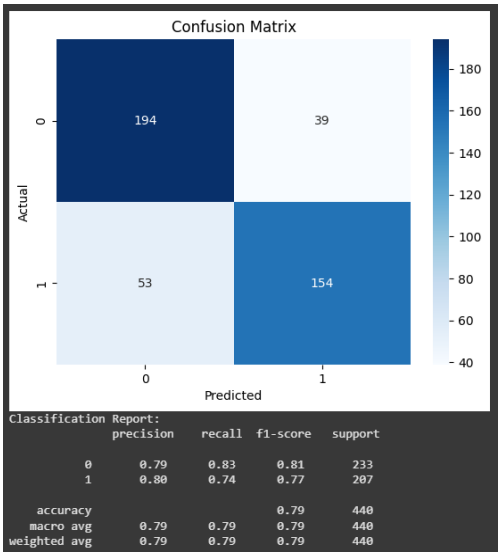


Gambar 9 Grafik distribusi kelas aktual dan prediksi model SVM

Berikut penjelasan dari hasil grafik seperti pada Gambar 9.

1. Terdapat 194 individu yang dikategorikan sebagai "sehat mental" (label 0) berdasarkan data sebenarnya.
2. Sebanyak 164 individu dikategorikan sebagai "gangguan mental" (label 1).
3. Model SVM memprediksi 143 individu sebagai "sehat mental" (label 0).
4. Sebanyak 215 individu diprediksi mengalami "gangguan mental" (label 1).

Untuk memahami performa klasifikasi lebih lanjut, *confusion matrix* digunakan untuk menggambarkan jumlah prediksi yang benar dan salah pada masing-masing kategori.



Gambar 10. Evaluasi model dengan *confusion matrix*

Berikut hasil evaluasi seperti pada Gambar 10.

1. Model SVM dengan splitting data 60:40 menunjukkan performa yang baik dalam mengklasifikasikan kondisi kesehatan mental remaja dengan akurasi 79%.
2. *Precision* untuk kelas 0 sebesar 0.79 dan kelas 1 sebesar 0.80, yang berarti model memiliki tingkat ketepatan yang cukup baik dalam mengidentifikasi setiap kategori.
3. *Recall* untuk kelas 0 sebesar 0.83, menunjukkan bahwa model mampu mendeteksi mayoritas data yang benar-benar termasuk dalam kategori ini. Namun, recall untuk kelas 1 lebih rendah yaitu 0.74, yang berarti masih ada beberapa data yang tidak terdeteksi dengan benar sebagai kategori ini.

4. *F1-score* rata-rata sebesar 0.79, menunjukkan keseimbangan yang baik antara precision dan recall.

5. KESIMPULAN

Penelitian ini menunjukkan bahwa model *Support Vector Machine* (SVM) dengan *splitting data* 60:40 mampu mengklasifikasikan kondisi kesehatan mental usia remaja dengan tingkat akurasi tertinggi sebesar 79% dengan memprediksi 143 individu sebagai "sehat mental" (label 0), dan sebanyak 215 individu diprediksi mengalami "gangguan mental" (label 1). Model ini memiliki *precision* sebesar 0.79 untuk kelas 0 dan 0.80 untuk kelas 1, yang mengindikasikan tingkat ketepatan yang baik dalam mengidentifikasi masing-masing kategori. *Recall* untuk kelas 0 mencapai 0.83, menunjukkan bahwa model dapat mendeteksi sebagian besar data yang benar-benar termasuk dalam kategori ini. Namun, *recall* untuk kelas 1 lebih rendah, yaitu 0.74, yang mengindikasikan masih adanya data yang tidak terdeteksi dengan benar. Dengan nilai rata-rata *F1-score* sebesar 0.79, model ini menunjukkan keseimbangan yang baik antara *precision* dan *recall*, sehingga dapat digunakan sebagai alat bantu dalam mendukung sistem deteksi dini gangguan kesehatan mental pada remaja.

UCAPAN TERIMA KASIH

Kami mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah mendukung kelancaran penelitian ini, mulai dari tahap penulisan hingga proses penerbitan. Ucapan terima kasih kami tujukan kepada orang tua, rekan-rekan, penerbit, serta banyak pihak lain yang tidak dapat kami sebutkan satu per satu. Kami juga menyampaikan apresiasi yang tulus kepada tim Jurnal Informatika dan Teknik Elektro Terapan yang telah meluangkan waktu untuk membaca, mengevaluasi, dan memberikan kesempatan kepada kami untuk mempublikasikan karya ilmiah ini. Semoga hasil penelitian ini dapat menjadi referensi yang bermanfaat bagi para pembaca serta memberikan kontribusi positif bagi masyarakat secara luas.

DAFTAR PUSTAKA

[1] Larissa, V. (2020). Kesehatan mental pada anak dan remaja. Universitas Persada Indonesia. Fakultas Psikologi.

[2] WHO. (2022). Mental health. World Health Organization Regional Office for Europe. https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-ourresponse.

[3] Center for Reproductive Health, University of Queensland, & Johns Bloomberg Hopkins School of Public Health. (2022). National Adolescent Mental Health Survey (I-NAMHS). https://qcmhr.org/outputs/reports/12-i-namhs-report-bahasa-indonesia.

[4] Anggraini, D., & Sutabri, T. (2024). Aplikasi Penyaringan Spam e-Mail Menggunakan Teknik Machine Learning dengan Metode Support Vector Machines. IJM: Indonesian Journal of Multidisciplinary, 2(3). https://journal.csspublishing/index.php/ijm.

[5] Kusman, V., Metayani, V., & Karnalim, O. (2024). Prediksi Analisis Sentimen Data Debat Pemilihan Presiden 2024 Menggunakan Support Vector Machine (SVM). Jurnal Keilmuan Dan Teknik Informatika, 16(1). https://doi.org/10.35891/explorit.

[6] Arifin, N., Enri, U., & Sulistiyowati, N. (2021). Penerapan Algoritma Support Vector Machine (SVM) dengan TF-IDF N-Gram untuk Text Classification. STRING (Satuan Tulisan Riset Dan Inovasi Teknologi), 6(2).

[7] Wenda, A. (2022). Support Vector Machine untuk Pengenalan Bentuk Manusia Menggunakan Kumpulan Fitur yang Dioptimalkan. Jurnal Sains Dan Teknologi, 11(1), 77–84. https://doi.org/10.23887/jst-undiksha.v11i1.

[8] Hamdani, I. M., Nurhidayat, N., Karman, A., Adhalia H, N. F., & Julyaningsih, A. H. (2024). Edukasi dan Pelatihan Data Science dan Data Preprocessing. Intisari: Jurnal Inovasi Pengabdian Masyarakat, 2(1), 19–26. https://doi.org/10.58227/intisari.v2i1.125.

[9] M. M. Teye, "Understanding Of Machine Learning With Deep Learning: Architectures, Workflow, Applications And Future Directions," Mdpi, Vol. 12, No. 5, Apr. 2023, Doi:

Https://Doi.Org/10.3390/Computers12050091

[10] A. E. E. Rashed, A. M. Elmorsy, And A. E. M. Atwa, "Comparative Evaluation Of Automated Machine Learning Techniques For Breast Cancer Diagnosis," Biomed Signal Process Control, Vol. 86, 2023, Doi: Https://Doi.Org/10.1016/J.Bspc.2023.105016.

[11] M. Cakir, M. Yilmaz, M. A. Oral, H. O. Kazanci, And Okan Oral, "Accuracy Assessment Of Rferns, Nb, Svm, And KNN Machine Learning Classifiers In Aquaculture," Journal Of King Saud University-Science, Vol. 35, No. 6, Aug. 2023, Doi: Https://Doi.Org/10.1016/J.Jksus.2023.102754.

[12] Nisa, K. (2024). Klasifikasi Penyakit Gangguan Mental Dengan Algoritma LightGBM. Jurasik (Jurnal Riset Sistem Informasi Dan Teknik Informatika), 9(2), 1086-1094.

[13] Handoko, W., & Iqbal, M. (2021). Prediksi Peminatan Program Studi Pada Penerimaan Mahasiswa STMIK Royal Menggunakan Naïve Bayes. Journal of Science and Social Research, 2, 231–235. http://jurnal.goretanpena.com/index.php/JSSR.

[14] Nabila, N., & Effendie, A. (2023). Prediksi Harga Minyak Mentah WTI Menggunakan Gabungan Arsitektur GRU dengan Penanganan Outlier. Doctoral Dissertation, Universitas Gadjah Mada. http://etd.repository.ugm.ac.id/.

[15] Elfaladonna, F., Isa, I. G. T., Sartika, D., & Putra, A. M. (2024). Buku Ajar Dasar Exploratory Data Analysis (EDA) (Vol. 01). NEM. https://doi.org/10.30813/j-alu.v2i2.6030.

[16] Atmojo, F., & Nurlita, C. (2024). Analisis Pemanfaatan Machine Learning Guna Prediksi Indeks Pembangunan Manusia. Jurnal Sistem Informasi Dan Teknik Komputer, 9(2).

[17] LaRose, R., & Coyle, B. (2020). Robust data encodings for quantum classifiers. Physical Review A, 102(3), 89–100. https://doi.org/10.17933/jppi.2021.110106.

[18] Qadrini, L., Hikmah, H., & Megasari, M. (2022). Oversampling, Undersampling, Smote SVM dan Random Forest pada Klasifikasi Penerima Bidikmisi Sejawa Timur Tahun 2017. Journal of Computer System and Informatics (JoSYC), 3(4), 386–391. https://doi.org/10.47065/josyc.v3i4.2154

[19] Rahman Ramli, A., Salsabila, A., & Adiba, F. (2024). Analisis Performa Convolutional Neural Networkdengan Hyperparameter Tuning dalam Mendeteksi Gambar Deepfake. Jurnal INSYPRO : Information System and Processing, 9(2). http://journal.uin-alauddin.ac.id/index.php/insypro.

[20] Darmawan, Z., & Dianta, A. (2023). Implementasi Optimasi Hyperparameter GridSearchCV Pada Sistem Prediksi Serangan Jantung Menggunakan SVM. Teknologi: Jurnal Ilmiah Sistem Informasi, 13(1), 8–15. https://doi.org/10.26594/teknologi.v13i1.3098.

[21] Azis, H., Alisma, Purnawansyah, & Nirmala. (2024). Analisis Kinerja Algoritma Pembelajaran Mesin Ensembel Pada Dataset Multi Kelas Citra Jaffe. Jurnal Ilmiah NERO , 9(2).