

Model Klasifikasi Mental Siswa Menggunakan Algoritma *Support Vector Machine*

DOI: <http://dx.doi.org/10.35889/progresif.v21i2.2813>

Creative Commons License 4.0 (CC BY – NC)



Charitas Fibriani^{1*}, Dian Novita Kristiyani²

¹Sistem Informasi, Universitas Kristen Satya Wacana, Salatiga, Indonesia

²Hubungan Masyarakat, Universitas Kristen Satya Wacana, Salatiga, Indonesia

*e-mail *Corresponding Author*: charitas.fibriani@uksw.edu

Abstract

Student mental health plays a vital role in academic performance and social well-being. This study aims to build a classification model using the Support Vector Machine (SVM) algorithm, based on 15 features covering demographic, academic, and behavioral aspects. The dataset, obtained from Kaggle, contains 426 records of junior and senior high school students. Key preprocessing steps include one-hot encoding, feature standardization, train-test splitting (80:20), and handling class imbalance with SMOTE. The model was trained using the Radial Basis Function (RBF) kernel and optimized using Grid Search CV to find the best parameters. Evaluation results show 65% accuracy, with better performance in predicting students without mental health issues (Absence). However, low recall for the Presence class indicates a need for improved strategies to handle data imbalance. This study highlights the potential of machine learning, particularly SVM, as a tool for early mental health detection in students, provided that effective data preprocessing is applied.

Keywords: *Student mental health; classification; Support Vector Machine; SMOTE; machine learning*

Abstrak

Kesehatan mental siswa berpengaruh besar terhadap prestasi akademik dan kesejahteraan sosial. Penelitian ini bertujuan membangun model klasifikasi kondisi mental siswa menggunakan algoritma *Support Vector Machine* (SVM) berbasis 15 fitur demografis, akademik, dan perilaku. Dataset yang digunakan berasal dari Kaggle, terdiri atas 426 data siswa SMP dan SMA. Tahapan penelitian meliputi preprocessing dengan *one-hot encoding*, standarisasi numerik, pembagian data (80:20), serta penanganan ketidakseimbangan data menggunakan SMOTE. Model dilatih menggunakan kernel *Radial Basis Function* (RBF) dan dioptimasi dengan *Grid Search CV*. Hasil evaluasi menunjukkan akurasi sebesar 65%, dengan kinerja lebih baik dalam mengenali siswa tanpa gangguan mental (*absence*) dibandingkan siswa dengan gangguan mental (*presence*). Rendahnya *recall* pada kelas *Presence* mengindikasikan perlunya strategi lanjutan terhadap ketidakseimbangan data. Penelitian ini menunjukkan bahwa machine learning, khususnya SVM, berpotensi sebagai alat bantu dalam deteksi awal kesehatan mental siswa jika disertai pengolahan data yang tepat.

Kata kunci: *Kesehatan mental siswa; Klasifikasi; Support Vector Machine; SMOTE; Machine learning*

1. Pendahuluan

Kesehatan mental siswa merupakan aspek penting yang berdampak langsung terhadap prestasi akademik, perkembangan sosial, dan kesejahteraan jangka panjang mereka. Gangguan mental yang tidak terdeteksi dan tidak ditangani sejak dini dapat menghambat potensi siswa, mempengaruhi hubungan sosial, serta meningkatkan risiko masalah psikologis jangka panjang. Pendekatan sistematis dan berbasis data untuk memantau kondisi mental siswa secara lebih obyektif dan akurat merupakan salah satu usaha untuk meningkatkan kesadaran akan pentingnya kesehatan mental dalam konteks Pendidikan [1].

Saat ini, gangguan mental pada generasi muda terjadi hampir di seluruh dunia, bahkan 30% gangguan mental ini terjadi pada siswa [2]. Jenis gangguan mental yang terjadi pada para siswa adalah depresi dengan gejala-gejala yang dapat mengganggu aktifitas siswa sehari-hari. Siswa menjadi kehilangan minat untuk melakukan aktifitasnya baik di lingkungan sekolah maupun di lingkungan keluarganya [3]. Peningkatan tekanan akademik, perubahan lingkungan sosial, serta dinamika keluarga turut memperparah kondisi ini. Namun, belum banyak institusi pendidikan yang memiliki sistem deteksi dini berbasis data untuk mengenali potensi gangguan mental siswa secara sistematis dan terukur.

Sebagai solusi, metode yang dapat digunakan dalam mendeteksi kondisi kesehatan mental adalah dengan memanfaatkan teknik klasifikasi dalam pembelajaran mesin (*machine learning*). Teknik klasifikasi termasuk dalam jenis *supervised learning*, teknik ini hanya dapat dilakukan pada dataset yang memiliki label item [4]. Metode klasifikasi memiliki banyak algoritma, salah satunya adalah *Support Vector Machine* (SVM), metode ini dikenal unggul dalam mengklasifikasikan data berdimensi tinggi dan menangani masalah klasifikasi biner maupun multi-kelas dengan akurasi yang baik [5]. Pada penelitian Raahmatulloh et al. 2024 [6] dan Ruliah 2015 [7], SVM mampu melakukan prediksi terhadap data dimensi tinggi dengan mengoptimalkan kernel SVM. Penelitian Shaik et al. 2024 [8] dan Alruwais et al. 2023 [9] menggunakan SVM untuk analisis kesehatan mental pada pengguna sosial media, performa klasifikasi yang dihasilkan cukup baik meskipun data yang digunakan tidak linier. Berdasarkan pertimbangan penelitian-penelitian tersebut, SVM terbukti mampu melakukan prediksi terhadap data dengan dimensi tinggi dan tidak linier, oleh karena itu penelitian ini mengusulkan pembangunan model klasifikasi berbasis algoritma *Support Vector Machine* (SVM) untuk mendeteksi kondisi kesehatan mental siswa. Model dikembangkan dengan memanfaatkan 15 fitur yang mencakup aspek demografis, akademik, dan perilaku siswa, seperti performa akademik, absensi, usia, jenis kelamin, kebiasaan tidur, aktivitas fisik, dan partisipasi dalam kegiatan ekstrakurikuler.

Berdasarkan latar belakang yang telah dijelaskan, maka rumusan masalah dalam penelitian ini adalah bagaimana membangun model klasifikasi yang dapat memprediksi kondisi kesehatan mental siswa berdasarkan data demografis, akademik, dan perilaku. Penelitian ini bertujuan untuk mengembangkan model klasifikasi kesehatan mental siswa menggunakan algoritma SVM dan mengevaluasi kinerja model SVM dalam mengklasifikasikan kondisi kesehatan mental berdasarkan fitur-fitur seperti data demografis, akademik, dan perilaku. Manfaat dari penelitian ini adalah menyediakan alat bantu prediktif yang dapat digunakan oleh lembaga pendidikan untuk melakukan intervensi dini secara lebih tepat sasaran. Kebaruan dari penelitian ini terletak pada pengembangan model klasifikasi kesehatan mental bagi siswa usia remaja dengan pendekatan algoritma SVM yang memadukan dimensi akademik, demografis, dan perilaku secara komprehensif. Kebaharuan pada penelitian ini adalah model klasifikasi kesehatan mental pada siswa usia remaja menggunakan metode SVM.

2. Tinjauan Pustaka

Penelitian mengenai klasifikasi kesehatan mental berbasis *machine learning* telah banyak dilakukan dalam berbagai pendekatan data dan algoritma. Penelitian Hairani et al. 2021 [10] menggunakan sistem pakar berbasis metode *Dempster-Shafer* untuk melakukan klasifikasi terhadap pasien dengan gangguan skizofrenia. Penelitian ini memanfaatkan 43 parameter gejala berdasarkan panduan psikologi klinis dan menghasilkan sistem klasifikasi berbasis aturan, namun belum memanfaatkan pendekatan pembelajaran mesin modern.

Penelitian Rahmatulloh et al. 2024 [6] melakukan klasifikasi kesehatan mental pada karyawan perusahaan teknologi menggunakan algoritma *Support Vector Machine* (SVM). Penelitian ini menitikberatkan pada optimasi parameter kernel SVM, yaitu nilai C dan gamma, untuk memperoleh model klasifikasi yang optimal. Parameter input dalam penelitian ini meliputi data karyawan yang diperoleh melalui survei psikologis.

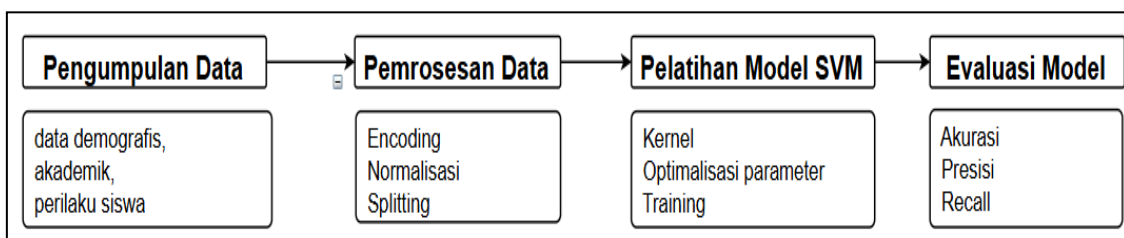
Penelitian Shaikh et al. 2024[8] menerapkan algoritma SVM yang dioptimasi dengan *Levy Flight Gannet Optimization* untuk memprediksi krisis kesehatan mental berdasarkan data teks dari media sosial. Sementara itu, Alruwais et al. 2023 [9] menggunakan pendekatan *deep learning dan Natural Language Processing* (NLP) untuk mendeteksi tingkat depresi juga dari data media sosial. Keduanya menggunakan data tidak terstruktur berupa teks, dan metode klasifikasi diarahkan untuk menganalisis emosi dan status psikologis berdasarkan pola bahasa pengguna.

Penelitian Ruliah 2015 [7] mengaplikasikan metode SVM pada data citra medis untuk klasifikasi stadium parasit malaria. Penelitian ini menunjukkan kekuatan SVM dalam menangani data berdimensi tinggi dan non-linier, meskipun konteksnya bukan kesehatan mental secara langsung namun penelitian ini membuktikan bahwa SVM mampu melakukan prediksi untuk data non linier.

Beberapa penelitian lain menggunakan SVM dalam ranah analisis sentimen sebagai studi pendukung efektivitas algoritma ini. Nugroho et al. 2024 [11] mengklasifikasikan sentimen pengguna Twitter terhadap COVID-19, sementara Nurkarifin et al. 2024 [12] serta Husnah et al. 2025 [13] menerapkan pendekatan serupa untuk menganalisis aplikasi pencari kerja dan isu kekerasan dalam rumah tangga. Sedangkan Pamungkas et al. [14]. Meskipun konteksnya berbeda, penelitian-penelitian ini membuktikan bahwa SVM efektif untuk data teks dengan dimensi tinggi dan data dengan distribusi kelas yang tidak seimbang.

Berdasarkan penelitian-penelitian sebelumnya, penelitian ini memiliki beberapa *state of the art* yang menjadi nilai kebaruannya. Pertama, objek penelitian difokuskan pada siswa sekolah menengah (SMP dan SMA), berbeda dari sebagian besar penelitian terdahulu yang menggunakan data mahasiswa, karyawan, atau data pengguna media sosial. Hal ini menuntut pemilihan fitur yang lebih relevan dengan karakteristik remaja. Kedua, data yang digunakan terdiri dari 15 parameter yang mencakup aspek demografis (usia, jenis kelamin, status tempat tinggal), akademik (nilai, absensi), dan perilaku (aktivitas fisik, kebiasaan tidur, keikutsertaan dalam organisasi/ekstrakurikuler), sehingga lebih komprehensif daripada penelitian sebelumnya yang cenderung fokus pada satu jenis data seperti teks atau citra. Ketiga, metode klasifikasi yang digunakan tetap berbasis SVM, namun pendekatan optimasi dilakukan secara sistematis menggunakan *GridSearchCV* untuk menemukan kombinasi parameter terbaik, serta mengatasi ketidakseimbangan data dengan teknik *SMOTE*. Hal ini menjadikan model tidak hanya akurat, tetapi juga lebih adaptif terhadap kompleksitas data kelas minoritas, yaitu siswa dengan gangguan mental. Penelitian ini menawarkan kontribusi baru dalam pengembangan sistem deteksi dini kesehatan mental siswa sekolah menengah dengan pendekatan klasifikasi berbasis *machine learning* yang menyatukan fitur multidimensional dan teknik optimasi serta penanganan data yang tidak berimbang kelasnya.

3. Metodologi



Gambar 1. Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk membangun dan mengevaluasi model klasifikasi kesehatan mental siswa. Teknik *supervised learning* dengan algoritma *Support Vector Machine* (SVM) digunakan untuk membentuk model prediktif berdasarkan data fitur yang telah dikumpulkan. Tahapan penelitian dapat dilihat pada Gambar 1. Tahap pertama dari penelitian ini adalah pengumpulan data. Pada pengumpulan data, data yang digunakan adalah data sekunder yang diambil dari situs Kaggle. Data ini diunggah oleh Muhammad Habli Nasution dengan nama file "Student Mental Health Dataset.csv", Data ini berisi total 426 data dengan rincian 214 data kondisi mental siswa sekolah menengah dan 212 data kondisi mental siswa sekolah menengah atas. Data ini diambil dari sebuah sekolah swasta di Indonesia. Data yang diambil ini akan digunakan sebagai data training untuk membangun model klasifikasi dan data pengujian untuk evaluasi model klasifikasi yang dibangun. Data yang diambil ini berisi informasi terkait kondisi siswa beserta status mentalnya. Status mental pada data ini akan menjadi kelas label untuk model, sedangkan kondisi siswa akan menjadi fitur yang digunakan sebagai parameter pada model klasifikasi yang akan dibangun. Kelas label dari data ini dibagi menjadi 2 kelas yaitu *Absence* (Sehat) dan *Presence* (Terganggu). Kelas data ini

dapat dilihat pada kolom fitur “*Mental Health Status*”. Parameter pada data dapat dilihat pada Tabel 1.

Tabel 1. Parameter Kesehatan Mental Siswa

Fitur	Keterangan	Nilai	Tipe Data
<i>Academic Performance</i>	Nilai akademik siswa	A, B, C, D, E	String
<i>Absences</i>	Kehadiran siswa di kelas	Rasio	Numerik
<i>Age</i>	Usia siswa	Rasio	Numerik
<i>Gender</i>	Jenis kelamin siswa	<i>Female Student</i> <i>Male Student</i>	String
<i>Residence Status</i>	Siswa tinggal bersama siapa	<i>With mother and father</i> <i>With mother only</i> <i>With father only</i> <i>Other</i>	String
<i>Father Education Level</i>	Jenjang pendidikan ayah	<i>Associate Degree</i> <i>Senior High School</i> <i>Bachelor</i> <i>Master</i>	String
<i>Father Job Category</i>	Jenis pekerjaan ayah	<i>Private employee</i> <i>Self employed</i> <i>Government employes</i> <i>Lecturer/ teacher</i> <i>SOC Employed</i> <i>Other</i>	String
<i>Father Work Location</i>	Lokasi pekerjaan ayah	<i>Working out of City</i> <i>Working within the city</i> <i>Other</i>	String
<i>Mother Education Level</i>	Jenjang pendidikan ibu	<i>Associate Degree</i> <i>Senior High School</i> <i>Bachelor</i> <i>Master</i>	String
<i>Mother Job Category</i>	Jenis pekerjaan ibu	<i>Private employee</i> <i>Self employed</i> <i>Government employes</i> <i>Lecturer/ teacher</i> <i>SOC Employed</i> <i>Other</i>	String
<i>Mother Work Location</i>	Lokasi pekerjaan ibu	<i>Working out of City</i> <i>Working within the city</i> <i>Other</i>	String
<i>Sleeping Habit</i>	Kebiasaan tidur siswa	<i>< 8 hours</i> <i>Above 8 hours</i>	String
<i>Physical Activity</i>	Kegiatan fisik	<i>Yes, No</i>	String
<i>Organization Involvement</i>	Keterlibatan organisasi	<i>Yes, No</i>	String
<i>Extracurricular Involvement</i>	Keterlibatan siswa pada ekstrakurikuler	<i>Yes, No</i>	String

Tahap kedua dalah pemrosesan data. Sebelum dilakukan pemodelan data akan melalui beberapa tahap *preprocessing*. Ada beberapa *preprocessing*, yaitu *encoding*, normalisasi dan *splitting*. Proses pertama adalah *encoding*, proses ini melakukan *encoding* data dari parameter yang nilainya masih berbentuk *non numerik* seperti pada fitur seperti *gender*, *job category*, dan *residence status*. Proses *encoding* data kategorikal untuk kasus data *non numerik* menggunakan fungsi *one-hot encoding*. Proses *preprocessing* selanjutnya adalah normalisasi. Proses ini dilakukan pada fitur dengan nilai numerik seperti *absences* dan *age*. Proses normalisasi dilakukan untuk meningkatkan performa SVM. Proses *preprocessing* selanjutnya pada penelitian ini adalah *splitting*. Proses ini membagi data yang telah dikumpulkan pada tahap sebelumnya. Data akan dibagi menjadi 2, data untuk training model klasifikasi dan data untuk evaluasi model

klasifikasi. Komposisi pembagian adalah 80% untuk data training untuk model klasifikasi dan 20% untuk data evaluasi model klasifikasi. Proses *preprocessing* terakhir yang dilakukan adalah proses penanganan data yang tidak seimbang menggunakan SMOTE (*Synthetic Minority Over-sampling Technique*).

Tahap ketiga adalah pelatihan model SVM klasifikasi menggunakan algoritma SVM. Data yang digunakan untuk melatih model adalah sejumlah 80% dari data yang ada. Fungsi optimasi model SVM dapat dilihat pada rumus (1) [7], fungsi optimasi ini dapat diselesaikan menggunakan fungsi *library Scikit-learn*. Penyelesaian fungsi optimasi ini menggunakan *trick kernelling radial basis function* (RBF) untuk mengubah data menjadi dimensi yang lebih tinggi supaya dapat dipisahkan, dimana data dapat dipisahkan secara *non-linier*. Rumus *kernelling* RBF dapat dilihat pada rumus (2) [7], Optimalisasi parameter c , γ dan kernel dilakukan menggunakan fungsi Grid Search untuk mendapatkan kombinasi parameter terbaik.

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (1)$$

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (2)$$

Keterangan:

α_i : bobot yang akan ditentukan oleh proses *training*

y_i : label kelas dari data *training*

Tahap terakhir dari penelitian adalah evaluasi model. Model dievaluasi menggunakan *confusion matrix* untuk menguji performa klasifikasi. Matrik evaluasi klasifikasi yang dipakai pada penelitian ini adalah *accuracy*, *precision*, *recall* dan *F1-Score*. Metrik ini digunakan untuk menilai kemampuan model dalam membedakan kategori kesehatan mental siswa dengan benar. *Accuracy* adalah pengukuran seberapa sering model membuat prediksi benar, rumus akurasi dapat dilihat pada rumus (3). *Precision* adalah pengukuran seberapa banyak prediksi positif benar-benar positif, rumus presisi dapat dilihat pada rumus (4). *Recall* adalah pengukuran seberapa baik model menangkap semua data positif, rumus *recall* dapat dilihat pada rumus (5). Sedangkan *F-Score* adalah rata-rata harmonis dari *precision* dan *recall*, rumusnya dapat dilihat pada rumus (6) [15].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$F - Score = (1 + \beta^2) \frac{PR}{R + \beta^2 P} \quad (6)$$

Keterangan:

TP : *True Positive*

FN : *False Negative*

FP : *False Positive*

TN : *True Negative*

P : *Precision*

R : *Recall*

β : parameter penimbang

4. Hasil dan Pembahasan

4.1. Data Penelitian

Tahap pertama pada penelitian ini adalah mengumpulkan data terkait kesehatan mental pada siswa. Pada Gambar 2 dapat dilihat contoh tampilan dataset student mental health yang diambil dari situs Kaggle. Pada dataset ini terdapat 18 kolom yaitu: *Date ID*, *Date Set*, *Academic*

Perfomance, Absence, Age, Gender, Residence Status, Father Education Level, Father Job Category, Father Work Location, Mother Education Level, Mother Job Category, Mother Work Location, Sleeping Habit, Physical Activity, Organization Involvement, Extracurricular Involvement sebagai kolom fitur klasifikasi kesehatan mental siswa, sedangkan kolom *Mental Status* sebagai fitur target pada kesehatan mental siswa.

Data ID	Data Set	Academic Performance	Absences	Age	Gender	Residence Status	Father Education Level	Father Job Category	Father Work Location	Mother Education Level	Mother Job Category	Mother Work Location	Sleeping Habit	Physical Activity	Organization Involvement	Extracurricular Involvement	Mental Health Status	
1	1 B		3	13	Male Students	With Father and Motjher	Bachelor	Private Employees	Working Out of City	Master	SOC Employed	Working Out of City	< 8 Hours	Yes	Yes	Yes	Absence	
2	1 A			9	13	Male Students	With Father and Motjher	Master	Private Employees	Working Out of City	Bachelor	Other	Working Within the City	< 8 Hours	Yes	Yes	No	Presence
3	1 B			4	13	Female Students	With Father and Motjher	Senior High School	Self Employed	Working Within the City	Senior High School	Other	Other	Above 8 Hours	No	Yes	Yes	Absence
4	1 B			12	14	Male Students	With Father and Motjher	Master	Self Employed	Working Out of City	Associate Degree	Private Employees	Other	< 8 Hours	Yes	Yes	No	Absence
5	1 B			5	14	Female Students	With Father and Motjher	Bachelor	Private Employees	Working Out of City	Bachelor	SOC Employed	Working Out of City	< 8 Hours	No	Yes	Yes	Presence

Gambar 2. Dataset Student Mental Health

Setelah data diperoleh akan dilakukan *preprocessing*. Pada penelitian ini terdapat beberapa proses pada tahapan *preprocessing*. Pada Gambar 3 dapat dilihat kode program untuk melakukan teknik *encoding* menggunakan teknik *one-hot encoding* untuk data kategorikal seperti fitur-fitur *non-numerik* seperti *gender, job category, dan residence status*. Fungsi yang digunakan untuk proses ini adalah fungsi *dummies*. Pada Gambar 4 dapat dilihat kode program untuk melakukan normalisasi data fitur numerik seperti *absences* dan *age* untuk meningkatkan performa SVM menggunakan fungsi *standard scale*. Pada Gambar 5 dapat dilihat kode program untuk membagi dataset menjadi data training dan data pengujian. Pembagian data pengujian sebesar 0,2 atau sebesar 20% dari total dataset, artinya data yang digunakan sebagai data training sejumlah 80% dari dataset tersedia. Pada Gambar 6 dapat dilihat kode program untuk penanganan data yang tidak seimbang. Data yang tidak seimbang dapat mengakibatkan prediksi tidak muncul. Teknik *oversampling* yang digunakan untuk menangani ketidak seimbangan data adalah menggunakan fungsi *SMOTE (Synthetic Minority Over-sampling Technique)*. Fungsi ini digunakan untuk membuat sampel sintetis dari kelas minoritas.

```
# 4. One-hot encoding fitur kategorikal
X_encoded = pd.get_dummies(X)
```

Gambar 3. Kode Program Proses *Encoding*

```
# 5. Standarisasi fitur
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X_encoded)
```

Gambar 4. Kode Program Proses Normalisasi

```
# 6. Split data untuk training dan testing
X_train, X_test, y_train, y_test = train_test_split(
    X_scaled, y, test_size=0.2, random_state=42
)
```

Gambar 5. Kode Program Proses *Splitting*

```
# 7. Terapkan SMOTE hanya pada data training
smote = SMOTE(random_state=42)
X_train_resampled, y_train_resampled = smote.fit_resample(X_train, y_train)
```

Gambar 6. Kode Program Proses *SMOTE*

4.2. Pelatihan Model Klasifikasi Kesehatan Mental Siswa

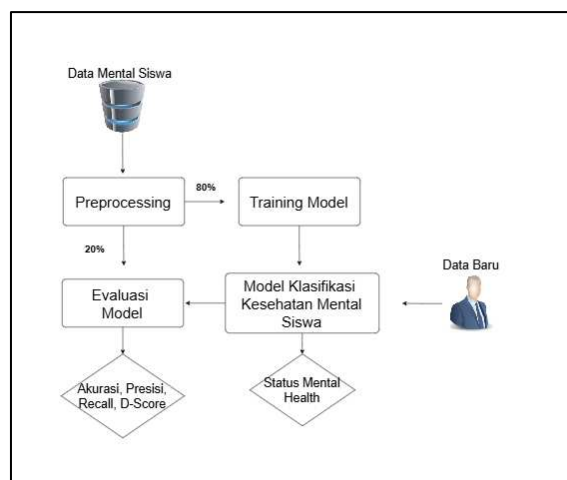
Setelah melalui proses preprocessing dan data siap digunakan, maka selanjutnya membangun model training klasifikasi kesehatan mental siswa menggunakan SVM. Model klasifikasi yang dibangun dan dilatih pada penelitian ini mempertimbangkan beberapa hal. Pertama, dataset pada penelitian tidak bisa dipisahkan secara linier, oleh karena itu data harus diubah menjadi dimensi yang lebih tinggi supaya dapat dipisahkan secara linier di ruang tersebut (*kernel trick*). Fungsi kernel trick yang digunakan pada penelitian ini adalah *Radial Basis Function* (RBF). Pemilihan RBF ini selain fungsi ini cocok untuk data non linier namun karena parameter yang digunakan pada fungsi ini lebih sederhana yaitu c dan γ . Kedua, model klasifikasi pada penelitian ini membutuhkan kombinasi parameter terbaik dan tidak perlu lagi teknik acak. Fungsi yang digunakan untuk teknik optimasi hyperparameter via pencarian grid untuk mendapatkan kombinasi parameter c , γ dan kernel terbaik pada penelitian ini adalah fungsi *Grid Search*. Kode program untuk model training data dapat dilihat pada Gambar 7. Pada Gambar 7 fungsi optimasi tujuan berdasarkan rumus (1) dapat dilihat pada baris *grid_search*. Pada baris ini dilakukan penyelesaian fungsi optimasi secara otomatis, dimana fungsi *GridSearchCV* mencari kombinasi terbaik dari C dan γ yang mempengaruhi bentuk *hyperplane*. Sedangkan fungsi kemiripan *non-linier kernel* RBF berdasarkan rumus (2) didefinisikan menggunakan fungsi *GridSearchCV* dengan mendefinisikan nilai γ untuk menentukan seberapa sempit atau lebarnya kurva kemiripan antara dua data.

```
# 8. GridSearchCV untuk optimasi parameter SVM
param_grid = {
    'C': [0.1, 1, 10],
    'gamma': ['scale', 0.01, 0.001],
    'kernel': ['rbf']
}

grid_search = GridSearchCV(SVC(probability=True, random_state=42), param_grid, cv=5)
grid_search.fit(X_train_resampled, y_train_resampled)
```

Gambar 7. Kode Program Model Training Klasifikasi Kesehatan Mental Siswa

Berdasarkan konsep yang telah dijabarkan maka untuk menggambarkan prinsip kerja model klasifikasi kesehatan mental siswa dapat dilihat pada Gambar 8. Data mental siswa yang dikumpulkan akan melalui proses *preprocessing* termasuk di dalamnya terdapat *encoding* data, normalisasi, SMOT dan *splitting* data. Setelah data siap digunakan, data akan dibagi komposisi jumlahnya, 80% akan digunakan sebagai data *training* untuk membangun model sedangkan 20% akan digunakan sebagai data pengujian untuk evaluasi model yang akan dibangun. Fungsi optimasi SVM akan diterapkan pada data *training*, menggunakan *kernel* RBF untuk mengatasi dimensi data yang tinggi. Hasil dari optimasi adalah model klasifikasi kesehatan mental siswa, model ini dapat digunakan untuk memrediksi status kesehatan mental dari siswa sebagai data baru.



Gambar 8. Model Training Klasifikasi Kesehatan Mental Siswa

4.3. Contoh Kasus Prediksi menggunakan Model Klasifikasi Mental Siswa

Pada pembagian dataset, terdapat pembagian sebesar 20% dari dataset untuk digunakan sebagai pengujian. Data pengujian ini digunakan untuk mengevaluasi model klasifikasi kesehatan mental. Konsep evaluasi adalah membandingkan prediksi label item kelas menggunakan model dengan data label item pada data aslinya. Pada bagian ini akan ditampilkan satu contoh prediksi kasus menggunakan model klasifikasi kesehatan mental menggunakan SVM, akan diujikan dengan memasukkan data baru untuk didapatkan label item kelas, apakah "Presence" ataukah "Absence". Contoh data baru ini diambil dari 20% dari dataset yang digunakan untuk pengujian. Contoh data inputan nilai dari masing-masing fitur dapat dilihat pada Tabel 2. Kode program inputan data baru yang akan diprediksi dan hasil label item kelasnya dapat dilihat pada Gambar 9. Berdasarkan inputan nilai, label item untuk data baru tersebut adalah "Absence", artinya berdasarkan nilai fitur yang dimiliki oleh siswa tersebut maka kondisi mentalnya masuk ke kelas sehat yang artinya tidak mengalami gangguan mental.

Tabel 2. Nilai Inputan Siswa Baru

Fitur	Nilai
Academic Performance	B
Absences	5
Age	14
Gender	Male Students
Residence Status	With Mother Only
Father Education Level	Bachelor
Father Job Category	Private Employed
Father Work Location	Working Out of City
Mother Education Level	Bachelor
Mother Job Category	Self Employed
Mother Work Location	Working Out of City
Sleeping Habit	< 8 Hours'
Physical Activity	No
Organization Involvement	No
Extracurricular Involvement	Yes

```
# 10. Prediksi data baru
data_baru = pd.DataFrame([
    {
        'Academic Performance': 'B',
        'Absences': 5,
        'Age': 14,
        'Gender': 'Male Students',
        'Residence Status': 'With Mother Only',
        'Father Education Level': 'Bachelor',
        'Father Job Category': 'Private Employees',
        'Father Work Location': 'Working Out of City',
        'Mother Education Level': 'Bachelor',
        'Mother Job Category': 'Self Employed',
        'Mother Work Location': 'Working Out of City',
        'Sleeping Habit': '< 8 Hours',
        'Physical Activity': 'No',
        'Organization Involvement': 'No',
        'Extracurricular Involvement': 'Yes'
    }
])
Hasil Prediksi: Absence
```

Gambar 9. Kode Program Prediksi Data Baru dan Hasil Prediksi

4.4. Evaluasi Model Klasifikasi Mental Siswa

Tahap terakhir pada penelitian ini adalah pengujian model klasifikasi kesehatan mental siswa menggunakan confusion matrix. Evaluasi dilakukan pada 20% dari dataset. Pada tahap ini akan dilakukan pengujian antara hasil prediksi menggunakan model klasifikasi kesehatan mental menggunakan SVM yang dibangun dengan hasil label pada data asli di dataset. Contoh prediksi kasus dapat dilihat pada pembahasan sebelumnya. Matrik pengukuran yang akan dilikat adalah

accuracy, *precision*, *recall* dan *F1-Score*. Berdasarkan rumus (3) untuk mendapatkan *accuracy*, rumus (4) untuk *precision*, rumus (5) untuk *recall*, dan rumus (6) untuk *F-score* maka dapat dilihat pada Gambar 10 kode program untuk melakukan evaluasi model klasifikasi yang dibangun. Hasil evaluasi model klasifikasi dapat dilihat pada Gambar 11. Parameter terbaik menggunakan fungsi *Grid Search* adalah nilai $c=10$ yang menjadikan model lebih tegas dan memberi pinalti lebih besar terhadap kesalahan, nilai $\gamma='scale'$ yang artinya γ otomatis disesuaikan dengan data bukan nilai tetap, nilai $kernel='rbf'$ yang artinya menggunakan kernel rbf yang cocok untuk data non-linier. Pada Gambar 10 dapat dilihat dari kelas "Absence" dan "Presence". Pada kelas "Absence" dapat dilihat nilai *Precision* dari *absence* sebesar 0,68 yang artinya 68% dari semua prediksi "Absence" benar. Nilai *Recall* dari *absence* sebesar 0,87 yang artinya 87% dari seluruh kasus "Absence" nyata berhasil terdeteksi. Nilai *F1-Score* dari *absence* sebesar 0,76 artinya model cukup seimbang dalam mengenali kelas ini. Sedangkan pada kelas "Presence" dapat dilihat nilai *Precision* dari *presence* sebesar 0,53 yang artinya 53% dari semua prediksi "Presence" benar. Nilai *Recall* dari *presence* hanya sebesar 0,26 yang artinya hanya 26% dari seluruh kasus "Presence" yang berhasil dikenali. Nilai *F1-Score* dari *presence* sebesar 0,35 artinya model cukup buruk dalam mengenali kelas ini. Nilai akurasi sebesar untuk model ini sebesar 65%.

```
# 9. Evaluasi model terbaik
best_model = grid_search.best_estimator_
y_pred = best_model.predict(X_test)
print("Best Parameters:", grid_search.best_params_)
print(classification_report(y_test, y_pred, zero_division=0))
```

Gambar 10. Kode Program Evaluasi Model

	precision	recall	f1-score	support
Absence	0.68	0.87	0.76	55
Presence	0.53	0.26	0.35	31
accuracy			0.65	86
macro avg	0.60	0.57	0.55	86
weighted avg	0.62	0.65	0.61	86

Gambar 11. Hasil Evaluasi Model

4.5. Pembahasan

Penelitian ini bertujuan membangun model klasifikasi kesehatan mental siswa menggunakan algoritma Support Vector Machine (SVM). Hasil evaluasi model menunjukkan tingkat akurasi 65%, nilai yang cukup bagus untuk sebuah model. Namun saat diamati nilai *accuracy*, *precision*, *recall* dan *F-score* dari masing-masing kelas ada sedikit tidak keseimbangan. Nilai masing-masing matrik evaluasi untuk kelas "Absence" secara normatif sangat bagus, artinya model mengenali kelas tersebut dengan baik. Sedangkan nilai masing-masing matrik evaluasi untuk kelas "Presence" kurang memuaskan, bahkan nilai matrik F-score dibawah 50%, artinya model yang dibangun sangat buruk dalam mengenali kelas "Presence" tersebut.

Model klasifikasi untuk prediksi status kesehatan mental yang dibangun pada penelitian ini menggunakan obyek siswa pada sekolah menengah umum dan sekolah menengah atas. Hal ini berbeda dengan penelitian Chen 2024 [3] dan Li et al. [1] yang memilih mahasiswa sebagai obyek penelitiannya. Perbedaan pemilihan obyek akan mempengaruhi variabel yang digunakan karena secara normatif, variabel yang mempengaruhi mental remaja dan dewasa sedikit berbeda. Sebagai contoh pada usia remaja gaya hidup konsumsi alkohol maupun merokok tidak dianggap sebagai variabel yang mempengaruhi mental siswa. Asumsi ini diambil karena pada alkohol dan rokok belum legal dikonsumsi oleh siswa, walaupun terkadang terdapat kasus yang di luar kewajaran.

Hasil penelitian ini memperkuat penelitian Rahmatulloh et al. 2024 [6] yang menekankan pentingnya pemilihan parameter kernel c dan γ dalam penggunaan SVM. Hasil penelitian ini membuktikan relevansi terkait pengaturan parameter tersebut, karena dalam penelitian ini

fungsi *GridSearchCV* terbukti efektif dalam mengoptimasi SVM agar mampu menangani data berdimensi tinggi

Penelitian Shaikh et al. 2024 [8] dan Alruwais et al. 2023[9] menggunakan data teks dari media sosial untuk klasifikasi tingkat depresi menggunakan SVM. Meskipun jenis datanya berbeda, penelitian ini menguatkan bahwa SVM tetap relevan sebagai pendekatan andal di berbagai jenis data dalam konteks kesehatan mental, termasuk data kategorikal dan numerik yang digunakan dalam penelitian ini.

Sementara itu, penelitian Ruliah 2015 [7] membuktikan efektivitas SVM dalam klasifikasi gambar medis (parasit malaria), menunjukkan keunggulan SVM dalam menangani data kompleks dan tidak terpisah secara linier, seperti halnya dalam penelitian ini yang menggunakan *kernel Radial Basis Function* (RBF) untuk mengatasi kompleksitas dimensi data mental siswa.

Namun, temuan penelitian ini juga menunjukkan kelemahan serupa dengan beberapa penelitian terdahulu: rendahnya recall pada kelas minoritas (*Presence*). Ini memperkuat temuan bahwa SVM cenderung bias terhadap kelas mayoritas dalam kasus ketidakseimbangan data Hairani et al. 2021[10]. Pendekatan SMOTE yang diterapkan dalam penelitian ini menjadi kontribusi penting untuk meningkatkan kesetaraan deteksi antar kelas untuk mengatasi ketidakseimbangan data, meskipun hasilnya belum sepenuhnya optimal.

Secara keseluruhan, penelitian ini memperkuat literatur bahwa SVM efektif dalam klasifikasi kasus kesehatan mental, khususnya dengan kombinasi *preprocessing* yang tepat, optimasi parameter, dan penanganan ketidakseimbangan data. Pengintegrasian 15 fitur demografis, akademik, dan perilaku siswa, penelitian ini juga memperluas cakupan fitur dibandingkan studi sebelumnya, yang mayoritas hanya fokus pada satu jenis data seperti teks atau citra.

5. Simpulan

Model klasifikasi kesehatan mental siswa pada penelitian ini menggunakan 15 fitur yaitu: performa akademik siswa, jumlah ketidakhadiran di kelas, usia siswa, jenis kelamin siswa, status tempat tinggal siswa, tingkat pendidikan dan pekerjaan orang tua, lokasi kerja orang tua, kebiasaan tidur siswa, aktivitas fisik siswa, serta keterlibatan siswa dalam organisasi dan kegiatan ekstrakurikuler. Label target dari dataset ini adalah Mental Health Status yang dikategorikan menjadi *Absence* (Sehat) dan *Presence* (Terganggu). Model yang dibangun melewati *preprocessing* dengan teknik one-hot encoding, fungsi Standar Scale, teknik Splitting, dan fungsi SMOTE. Sedangkan untuk pelatihan model klasifikasi menggunakan fungsi *Grid Search* dan kernel RBF. Berdasarkan hasil evaluasi klasifikasi, model yang dibangun cukup baik mengenali kelas *Absence* namun sangat buruk dalam mengenali kelas *Presence* padahal justru kelas tersebut yang penting dalam konteks deteksi kesehatan mental. Hal tersebut disebabkan ketidakseimbangan data jumlah kelas *Presence* dan kelas *Absence* pada dataset. Meskipun nilai akurasi sebesar 65%, namun model hanya bagus pada satu kelas saja.

Saran untuk pengembangan penelitian ini dengan mencoba menyesuaikan threshold prediksi, hal ini sebagai alat bantu untuk dataset yang terlalu konservatif ke satu sisi. Saran pengembangan lain adalah dengan uji variasi c dan γ lebih luas dalam fitur Grid Search. Saran pengembangan terakhir adalah dengan eksperimen dengan model lain seperti random Forest yang lebih toleran terhadap data yang tidak seimbang kelasnya.

Daftar Referensi

- [1] B. Li and Y. Xu, "Research on evaluation model of College Students' mental health," in *Proc. Int. Conf. Health Big Data Smart Sports (HBDSS)*, IEEE, Oct. 2021, pp. 114–117, doi: 10.1109/HBDSS54392.2021.00031.
- [2] N. Dhariwal et al., "A pilot study on AI-driven approaches for classification of mental health disorders," *Front. Hum. Neurosci.*, vol. 18, Apr. 2024, doi: 10.3389/fnhum.2024.1376338.
- [3] J. Chen, "Convolutional Neural Network with Fast Exponential Linear Unit Based Psychological Mental State Identification of College Students," in *Proc. 2nd Int. Conf. Networks, Multimedia and Inf. Technol. (NMITCON)*, IEEE, Aug. 2024, pp. 1–4, doi: 10.1109/NMITCON62075.2024.10699189.
- [4] M. Al-Farouni, J. Jeyasudha, V. N. Sree, R. Venkatasubramanian, and A. Ameelia Roseline, "Predictive Analytics for Mental Health Crises using Social Media Data with Attention Mechanism based Support Vector Machine Classification," in *Proc. 1st Int. Conf., Syst. and Inf. Technol. (SSITCON)*, IEEE, Oct. 2024, pp. 1–5, doi: 10.1109/SSITCON62437.2024.10796269.

- [5] R. Resmiati and T. Arifin, "Klasifikasi Pasien Kanker Payudara Menggunakan Metode Support Vector Machine dengan Backward Elimination," *SISTEMASI: J. Sist. Informasi*, vol. 10, no. 2, pp. 381-392, May 2021, doi: 10.32520/stmsi.v10i2.1238.
- [6] A. Rahmatulloh, A. Taufiq, I. Darmawan, E. Haerani, and R. Rizal, "Support Vector Machine Parameters Optimization using Firefly Algorithm for Mental Health Prediction," in *Proc. 9th Int. Conf. Informatics Comput. (ICIC), IEEE*, Oct. 2024, pp. 1-6, doi: 10.1109/ICIC64337.2024.10956409.
- [7] R. Ruliah, "Klasifikasi Stadium Trophozoite, Schizonts, Gametocyter pada Sediaan Darah Plasmodium Falciparum dengan Pendekatan Support Vector Machine," *Progresif: J. Ilm. Komput.*, vol. 6, no. 2, Aug. 2015, doi: <http://dx.doi.org/10.35889/progresif.v6i2.78>.
- [8] A. A. Shaikh, M. Al-Farouni, B. M. Manjula, S. Sivagami, and S. Devi, "Levy Flight Based Gannet Optimization Algorithm with Euclidean Distance-Support Vector Machine for Predictive Analysis of Mental Health Crisis Using Social Media Data," in *Proc. Int. Conf. Integrated Intell. Commun. Syst. (ICIICS), IEEE*, Nov. 2024, pp. 1-5, doi: 10.1109/ICIICS63763.2024.10859744.
- [9] N. Alruwais, H. Alamro, M. M. Eltahir, A. S. Salama, M. Assiri, and N. A. Ahmed, "Modified arithmetic optimization algorithm with Deep Learning based data analytics for depression detection," *AIMS Math.*, vol. 8, no. 12, pp. 30335-30352, 2023, doi: 10.3934/math.20231549.
- [10] H. Hairani, K. Kurniawan, K. A. Latif, and M. Innuddin, "Metode Dempster-Shafer untuk Diagnosis Dini Jenis Penyakit Gangguan Jiwa Skizofrenia Berbasis Sistem Pakar," *SISTEMASI: J. Sist. Informasi*, vol. 10, no. 2, pp. 280-290, May 2021, doi: 10.32520/stmsi.v10i2.1195.
- [11] A. W. Nugroho and N. Norhikmah, "Sentiment Analysis using the Support Vector Machine Algorithm on Covid_19," *SISTEMASI: J. Sist. Informasi*, vol. 13, no. 4, p. 1758, Jul. 2024, doi: 10.32520/stmsi.v13i4.3778.
- [12] M. Nurkarifin, I. Hermanto, and M. G. Resmi, "Analisis Sentimen Terhadap Aplikasi Lowongan Kerja Menggunakan Lexicon Based Features Dan Support Vector Machine," *JUTISI: J. Ilm. Tek. Inform. dan Sist. Informasi*, vol. 13, no. 1, Apr. 2024, doi: <http://dx.doi.org/10.35889/jutisi.v13i1>.
- [13] M. Husnah and R. Hidayat, "Analysis of Public Perception on Domestic Violence Cases using Support Vector Machine Algorithm," *SISTEMASI: J. Sist. Informasi*, vol. 14, no. 1, pp. 117-126, Jan. 2025, doi: 10.32520/stmsi.v14i1.4724.
- [14] D. A. Pamungkas, A. Amali, and U. D. Soer, "Analisis Sentimen Publik Terhadap Polusi Udara di Kota Jakarta: Perbandingan Algoritma Support Vector Machine, Naive Bayes, dan Random Forest," *JUTISI: J. Ilm. Tek. Inform. dan Sist. Informasi*, vol. 13, no. 3, pp. 132-145, Dec. 2025, doi: <http://dx.doi.org/10.35889/jutisi.v13i3.2414>.
- [15] M. Mulyana and W. Utomo, "Optimized Weight Evolutionary-based Support Vector Machine (SVM) Optimization for Comment Sentiment," *SISTEMASI: J. Sist. Informasi*, vol. 14, no. 1, pp. 147-156, Jan. 2025, doi: 10.32520/stmsi.v14i1.4762.