


Implementasi *Least Squares Support Vector Machine* dan SMOTE untuk Klasifikasi Kesehatan Mental

DOI: <http://dx.doi.org/10.35889/jutisi.v15i2.3613>

Creative Commons License 4.0 (CC BY – NC) 

Marcelino^{1*}, Muhammad Rizky Pribadi²

Informatika, Universitas Multi Data Palembang, Palembang, Indonesia

*e-mail *Corresponding Author*: marcel1212gt@gmail.com

Abstract

Mental health disorders such as depression, anxiety, and stress are global problems that require accurate early detection. This study proposes a mental health classification model using machine learning algorithms based on data from the Depression Anxiety Stress Scales (DASS-42) questionnaire and respondent demographic features. The main method used is Least Squares Support Vector Machine (LSSVM) combined with Synthetic Minority Oversampling Technique (SMOTE) and Backward Elimination feature selection. From testing on 39,775 respondent data, Backward Elimination successfully reduced more than 40% of the feature dimensions by selecting the most statistically significant attributes (p -value < 0.05). Oversampling with SMOTE proved successful in overcoming class imbalance in minority labels. Performance evaluation showed that LSSVM using the Radial Basis Function (RBF) kernel provided the most optimal results compared to the Linear and Polynomial kernels, with an F1-Score of 83.96% for Depression, 77.30% for Anxiety, and 82.00% for Stress. This proposed model contributes to the development of a more computational, efficient, and accurate mental health screening system.

Keywords: *Backward Elimination; DASS-42; LSSVM; Mental Health; SMOTE*

Abstrak

Gangguan kesehatan mental seperti depresi, kecemasan, dan stres merupakan masalah global yang memerlukan deteksi dini yang akurat. Penelitian ini mengusulkan model klasifikasi tingkat kesehatan mental menggunakan algoritma machine learning berdasarkan data kuesioner *Depression Anxiety Stress Scales* (DASS-42) dan fitur demografis responden. Metode utama yang digunakan adalah *Least Squares Support Vector Machine* (LSSVM) yang dikombinasikan dengan *Synthetic Minority Oversampling Technique* (SMOTE) dan seleksi fitur *Backward Elimination*. Dari pengujian terhadap 39.775 data responden, *Backward Elimination* berhasil mereduksi lebih dari 40% dimensi fitur dengan menyeleksi atribut yang paling signifikan secara statistik (p -value < 0.05). Oversampling dengan SMOTE terbukti berhasil mengatasi ketidakseimbangan kelas pada label minoritas. Evaluasi kinerja menunjukkan bahwa LSSVM menggunakan kernel *Radial Basis Function* (RBF) memberikan hasil paling optimal dibandingkan kernel Linear dan Polynomial, dengan pencapaian F1-Score sebesar 83.96% untuk Depresi, 77.30% untuk Kecemasan, dan 82.00% untuk Stres. Model yang diusulkan ini berkontribusi dalam pengembangan sistem screening kesehatan mental yang lebih komputasional, efisien, dan akurat.

Kata kunci: *Backward Elimination; DASS-42; Kesehatan Mental; LSSVM; SMOTE*

1. Pendahuluan

Kesehatan mental merupakan kondisi fundamental yang memengaruhi pikiran, emosi, dan perilaku seseorang yang variasinya merentang dari ringan hingga berat [1]. Di Indonesia, tiga jenis gangguan psikologis yang paling umum ditemukan dan sering kali terjadi secara tumpang tindih adalah depresi, kecemasan, dan stres [2]. Deteksi dini terhadap kondisi tersebut menjadi sangat krusial, salah satunya dengan memanfaatkan instrumen psikometri seperti *Depression Anxiety Stress Scales* (DASS-42) yang terdiri dari 42 item kuesioner [3]. Pengembangan sistem *screening* kesehatan mental berbasis komputasi diharapkan dapat

memberikan dukungan diagnosis yang lebih objektif dan efisien bagi masyarakat luas. Meskipun instrumen DASS-42 sangat efektif, penggunaan data berskala ini menghasilkan dataset yang bersifat multidimensi dan sangat kompleks, sehingga berisiko menurunkan efisiensi pelatihan model *machine learning* dan mengurangi akurasi. Selain masalah dimensi fitur yang besar, tantangan krusial lainnya adalah ketidakseimbangan kelas (*class imbalance*) yang signifikan di dalam *dataset*. Distribusi data sering kali didominasi oleh kelas *Extremely Severe* pada label kecemasan dan depresi, serta kelas Normal pada stres, sementara kelas *Mild* menjadi minoritas ekstrem. Jika tidak ditangani dengan tepat, ketidakseimbangan ini menyebabkan algoritma kurang peka dalam mengenali kelas minoritas dan cenderung memprediksi secara bias ke arah kelas mayoritas [4].

Beberapa penelitian terdahulu telah berupaya mengatasi permasalahan klasifikasi kesehatan mental menggunakan berbagai algoritma. Penelitian yang dilakukan oleh [5] menemukan bahwa algoritma *Support Vector Machine* (SVM) memberikan kinerja superior dalam memprediksi tingkat depresi, kecemasan, dan stres dibandingkan *K-Nearest Neighbor* (KNN). Dalam aspek optimasi fitur, menerapkan metode Backward Elimination pada Random Forest dan berhasil meningkatkan akurasi prediksi stres akademik dari 86% menjadi 88% [6]. Sementara itu, [7] membuktikan bahwa eliminasi fitur mundur pada komparasi SVM dan C4.5 mampu mereduksi kompleksitas dimensi hingga 75% tanpa mendegradasi performa. Namun, terdapat gap yang tersisa di mana SVM standar masih memiliki beban komputasi tinggi karena menyelesaikan masalah optimasi melalui *Quadratic Programming Problem* (QPP) yang sangat kompleks, menjadikannya lambat dan memakan banyak memori untuk dataset berskala besar [8].

Sebagai solusi atas permasalahan tersebut, penelitian ini mengusulkan implementasi algoritma *Least Squares Support Vector Machine* (LSSVM) [9] yang dikombinasikan dengan teknik SMOTE [10] dan seleksi fitur *Backward Elimination* [6]. Penggunaan LSSVM didasarkan pada kemampuannya menyelesaikan masalah optimasi melalui persamaan linier biasa, sehingga proses perhitungan menjadi jauh lebih efisien dan cepat [9]. Penelitian yang dilakukan oleh [11] telah membuktikan Efektivitas LSSVM dalam menangani data tabular dengan masalah ketidakseimbangan kelas juga telah dibuktikan pada studi prediksi cacat perangkat lunak (*software defect prediction*). Penelitian tersebut menerapkan LSSVM yang dikombinasikan dengan teknik SMOTE pada *dataset* prediksi cacat perangkat lunak, dan berhasil mencapai akurasi rata-rata tertinggi sebesar 93,87% serta menunjukkan peningkatan nilai AUC pada data yang tidak seimbang. Kebaruan dari riset ini terletak pada penggabungan varian kernel LSSVM dengan reduksi noise secara matematis melalui algoritma *Ordinary Least Squares* (OLS) *Backward Elimination* dan keseimbangan representasi kelas secara penuh menggunakan SMOTE pada dataset DASS-42. Pendekatan terpadu ini diharapkan mampu menghasilkan model klasifikasi kesehatan mental yang memiliki kecepatan komputasi optimal namun tetap mempertahankan tingkat presisi dan sensitivitas yang tinggi pada seluruh kategori gangguan.

2. Metodologi

Metodologi dalam penelitian ini difokuskan pada pemrosesan sistematis data mentah menjadi model klasifikasi yang dievaluasi performanya secara terstruktur.

2.1 Studi Literatur

Tahapan ini dimulai dengan melakukan pembelajaran literatur dari jurnal dan referensi terkait topik penelitian, yaitu klasifikasi tingkat kesehatan mental menggunakan data DASS-42. Studi difokuskan pada pemahaman metode *Least Squares Support Vector Machine* (LSSVM), penanganan *class imbalance* menggunakan SMOTE, dan penerapan *feature selection* dengan *Backward Elimination*.

2.2 Pengumpulan Dataset

Dataset yang digunakan diperoleh dari platform Kaggle dengan nama DASS-42 *Responses*. *Dataset* ini berisi 39.775 data responden dengan 172 atribut, yang mencakup skor kuesioner DASS-42, data demografis, dan *metadata* lainnya. Rincian pertanyaan DASS-42 dan variabel demografis dapat dilihat pada Tabel 1 dan Tabel 2.

Tabel 1. Daftar Atribut Pertanyaan DASS-42 c

No	Pertanyaan
1.	Menjadi marah karena hal-hal kecil/sepele
2.	Mulut terasa kering
3.	Tidak dapat melihat hal yang positif dari suatu kejadian
4.	Merasakan gangguan dalam bernapas (napas cepat, sulit bernapas)
5.	Merasa sepertinya tidak kuat lagi untuk melakukan suatu kegiatan
6.	Cenderung bereaksi berlebihan pada situasi
7.	Kelemahan pada anggota tubuh
8.	Kesulitan untuk relaksasi/bersantai
9.	Cemas yang berlebihan dalam suatu situasi namun bisa lega jika hal/situasi itu berakhir
10.	Pesimis
11.	Mudah merasa kesal
12.	Merasa banyak menghabiskan energi karena cemas
13.	Merasa sedih dan depresi
14.	Tidak sabaran
15.	Kelelahan
16.	Kehilangan minat pada banyak hal
17.	Merasa diri tidak layak
18.	Mudah tersinggung
19.	Berkeringat tanpa stimulasi oleh cuaca maupun latihan fisik
20.	Ketakutan tanpa alasan yang jelas
21.	Merasa hidup tidak berharga
22.	Sulit untuk beristirahat
23.	Kesulitan dalam menelan
24.	Tidak dapat menikmati hal-hal yang saya lakukan
25.	Perubahan kegiatan jantung dan denyut nadi tanpa stimulasi latihan fisik
26.	Merasa hilang harapan dan putus asa
27.	Mudah marah
28.	Mudah panik
29.	Kesulitan untuk tenang setelah sesuatu yang mengganggu
30.	Takut diri terhambat oleh tugas-tugas yang tidak biasa dilakukan
31.	Sulit untuk antusias pada banyak hal
32.	Sulit mentoleransi gangguan-gangguan terhadap hal yang sedang dilakukan
33.	Berada pada keadaan tegang
34.	Merasa tidak berharga
35.	Tidak dapat memaklumi hal apapun yang menghalangi Anda untuk menyelesaikan hal yang sedang Anda lakukan
36.	Ketakutan
37.	Tidak ada harapan untuk masa depan
38.	Merasa hidup tidak berarti
39.	Mudah gelisah
40.	Khawatir dengan situasi saat diri (menjadi panik dan mempermalukan diri sendiri)
41.	Gemetar
42.	Sulit untuk meningkatkan inisiatif dalam melakukan sesuatu

Tabel 2. Daftar Variabel Demografis

No	Variabel demografis	Keterangan
1.	<i>education</i>	pendidikan
2.	<i>urban</i>	jenis tempat tinggal
3.	<i>gender</i>	jenis kelamin
4.	<i>engnat</i>	kewarganegaraan inggris
5.	<i>age</i>	umur
6.	<i>screensize</i>	ukuran layar
7.	<i>uniquenetworklocation</i>	lokasi jaringan
8.	<i>hand</i>	tangan yang sering digunakan
9.	<i>religion</i>	agama
10.	<i>orientation</i>	ketertarikan
11.	<i>race</i>	ras
12.	<i>voted</i>	pilihan
13.	<i>married</i>	status pernikahan
14.	<i>familysize</i>	jumlah keluarga
15.	<i>major</i>	jurusan

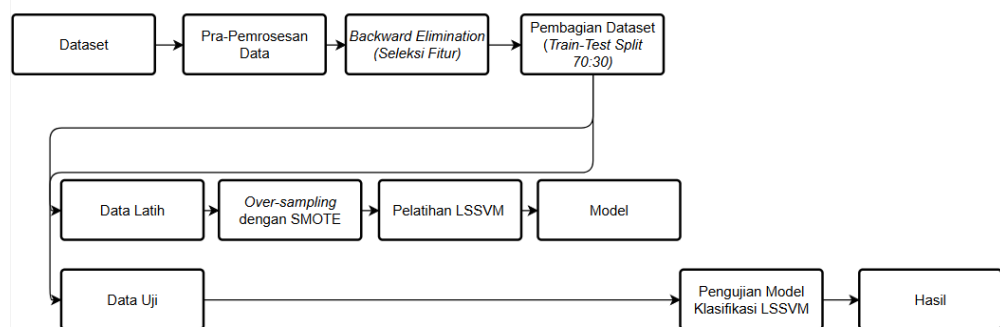
2.3 Perancangan

Pada tahapan ini dilakukan perancangan sistem, diawali dengan pra-pemrosesan data. Proses ini mencakup normalisasi skor jawaban dari skala 1-4 menjadi 0-3, agregasi skor untuk setiap subskala (Depresi, Kecemasan, Stres), dan pelabelan data ke dalam lima kelas standar DASS: *Normal, Mild, Moderate, Severe, dan Extremely Severe*. Mengingat algoritma LSSVM beroperasi menggunakan matriks numerik, seluruh variabel demografis kategorikal ditransformasikan menjadi representasi angka bulat menggunakan teknik Label Encoding [12].

Selanjutnya, seleksi fitur *Backward Elimination* diterapkan memanfaatkan algoritma *Ordinary Least Squares (OLS)* [13], di mana atribut dengan tingkat signifikansi *p-value > 0.05* dieliminasi dari sistem secara bertahap [14]. *Dataset* yang telah direduksi dimensinya kemudian dibagi menggunakan rasio 70% data latih dan 30% data uji. Untuk mengatasi *class imbalance* pada data latih, diterapkan teknik SMOTE yang mensintesis sampel kelas minoritas melalui interpolasi linier dengan tetangga terdekatnya [15], menggunakan persamaan 1

$$x_{synthetic} = x^i + \lambda \cdot (x_j - x_i) \tag{1}$$

di mana $x_{synthetic}$ adalah data sintetis baru, x_i adalah data asli kelas minoritas, x_j adalah tetangga terdekat, dan lambda adalah angka acak antara 0 dan 1. Skema perancangan dapat dilihat pada gambar 1.



Gambar 1. Skema Perancangan LSSVM

2.4 Pelatihan

Tahap pelatihan (*training*) bertujuan agar model dapat mempelajari pola dari data latih yang telah diseimbangkan dan terseleksi fiturnya. LSSVM memecahkan masalah optimasi melalui sistem persamaan linier dengan meminimalkan fungsi objektif pada persamaan 2.

$$\min \frac{1}{2} \|w\|^2 + \frac{c}{2} \sum_{i=1}^n \xi_i^2 \quad (2)$$

Model akan diuji pada tiga jenis kernel, yaitu *Linear*, *Polynomial*, dan *Radial Basis Function* (RBF). Pada pemodelan LSSVM, parameter utama yang menentukan keseimbangan antara margin maksimal dan toleransi kesalahan klasifikasi adalah parameter penalti (C). Selain itu, setiap fungsi kernel memiliki parameter spesifik yang harus dioptimasi. Pada kernel *Linear*, optimasi hanya bergantung pada parameter C. Pada kernel *Polynomial*, digunakan parameter derajat *polinomial* (d) dan parameter konstanta (c). Sedangkan pada kernel RBF, optimalisasi bergantung pada nilai C serta parameter gamma (γ) yang menentukan lebar pita (*bandwidth*) dari fungsi eksponensial *Gaussian*. yang dirumuskan pada persamaan 3.

$$K(x, y) = \exp(-\gamma(x - y)^2) \quad (3)$$

2.5 Pengujian

Sistem yang telah dilatih kemudian diuji menggunakan 30% data uji (*testing*). Evaluasi tingkat keberhasilan klasifikasi diukur menggunakan *Confusion Matrix* yang menghitung probabilitas performa berdasarkan *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Perhitungan metrik evaluasinya didasarkan pada persamaan 4, 5, 6, dan 7.

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

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

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

3. Hasil dan Pembahasan

3.1 Hasil Pra-Pemrosesan Data

Tahapan awal penelitian mengeksekusi pra-pemrosesan terhadap dataset mentah DASS-42 yang melibatkan 39.775 responden. Untuk memberikan gambaran terhadap data yang digunakan dalam penelitian, sebagian sampel dataset hasil pra-pemrosesan disajikan pada Tabel 3. Data tersebut merupakan hasil transformasi atribut kategorikal menggunakan teknik Label Encoding serta normalisasi skor DASS-42 ke dalam format numerik.

Tabel 3. Sampel Dataset Hasil Pra-Pemrosesan

Gender	Age	Urban	Q3	Q5	Q13	Q21	Depression
1	21	2	2	3	3	2	Severe
0	19	1	0	1	1	0	Normal
1	25	3	3	2	3	3	Extremely Severe
0	22	2	1	2	1	1	Moderate

Q3, Q5, Q13, dan Q21 merupakan sebagian atribut pertanyaan DASS-42 yang digunakan dalam proses klasifikasi kesehatan mental. Skor asli kuesioner berskala 1 hingga 4 ditransformasikan secara sistematis menjadi rentang 0 hingga 3. Selanjutnya, perhitungan agregasi dilakukan pada 14 item pertanyaan spesifik untuk masing-masing subskala (Depresi, Kecemasan, dan Stres) dalam melabeli setiap responden ke dalam lima tingkat keparahan (*Normal*, *Mild*, *Moderate*, *Severe*, *Extremely Severe*) berdasarkan ambang batas DASS [3]. Atribut demografis yang masih berbentuk kategorikal teks dikonversi menjadi data numerik menggunakan teknik *Label Encoding* agar sepenuhnya kompatibel dengan perhitungan matriks matematis. Sebagai bentuk optimasi memori untuk dataset berskala besar, seluruh tipe data dikonversi ke dalam format *float32*, yang secara praktis mereduksi beban penggunaan RAM komputasi hingga 50%. Hasil dari fase ini adalah matriks data numerik yang bersih, yang juga menyingkap adanya fenomena

ketidakseimbangan kelas (*class imbalance*) ekstrem, di mana label *Extremely Severe* sangat mendominasi dibandingkan kelas minoritas *Mild*.

3.2 Hasil Seleksi Fitur (*Backward Elimination*)

Matriks data yang telah siap kemudian dioptimasi melalui fase seleksi fitur menggunakan metode *Backward Elimination*. Proses reduksi ini memanfaatkan algoritma *Ordinary Least Squares* (OLS) untuk menguji signifikansi statistik seluruh 27 fitur awal (gabungan respons kuesioner dan variabel demografis). Atribut yang memiliki nilai *p-value* > 0.05 dieliminasi secara iteratif [6], [7]. Pada skala Depresi, sistem mengeliminasi 13 fitur dan menyisakan 14 fitur; skala Kecemasan menyisakan 15 fitur; dan skala Stres menyisakan 12 fitur. Secara statistik, fitur demografis seperti usia, edukasi, dan agama tereliminasi karena tidak memiliki korelasi linier yang kuat, dan hanya menyisakan variabel demografis tunggal yang relevan seperti *gender*, *married*, dan *urban*. Sebagai ilustrasi hasil seleksi fitur, sebagian atribut yang dipertahankan dan dieliminasi oleh metode *Backward Elimination* dapat dilihat pada Tabel 4.

Tabel 4. Contoh Hasil Seleksi Fitur *Backward Elimination*

Atribut	<i>p-value</i>	Status
<i>gender</i>	0.001	Dipertahankan
<i>married</i>	0.013	Dipertahankan
<i>urban</i>	0.004	Dipertahankan
<i>religion</i>	0.421	Dieliminasi
<i>education</i>	0.318	Dieliminasi

3.3 Hasil Pembagian dan Penyeimbangan Data (SMOTE)

Dataset dengan subset fitur optimal selanjutnya didistribusikan menggunakan skema pembagian 70% untuk data latih (*training*) dan 30% untuk data uji (*testing*). Untuk menanggulangi bias algoritma terhadap kelas mayoritas yang ditemukan pada fase pertama, teknik *Synthetic Minority Oversampling Technique* (SMOTE) diterapkan secara eksklusif pada data latih. Sintesis sampel kelas minoritas dilakukan melalui interpolasi linier dengan mengatur parameter *k-neighbors* secara dinamis [15]. Selanjutnya, augmentasi kelas minoritas melalui SMOTE membuahakan penyeimbangan yang komprehensif, yang dirangkum pada Tabel 5.

Tabel 5. Perbandingan Jumlah Data Latih Sebelum dan Sesudah SMOTE

Skala Psikologi	Jumlah Data Latih	Jumlah Data Latih (Setelah SMOTE)
<i>Depression</i>	19.845	30.095
<i>Anxiety</i>	19.845	32.470
<i>Stress</i>	19.845	32.125

3.4 Hasil Pelatihan dan Pengujian Model LSSVM

Hasil dari fase pelatihan dan pengujian menunjukkan bahwa model LSSVM mampu melakukan klasifikasi tingkat kesehatan mental terhadap data uji berdasarkan pola yang dipelajari dari data latih. Sistem menghasilkan prediksi kelas kesehatan mental pada masing-masing responden sesuai dengan tingkat keparahan DASS-42. Contoh hasil klasifikasi model terhadap data uji dapat dilihat pada Tabel 6.

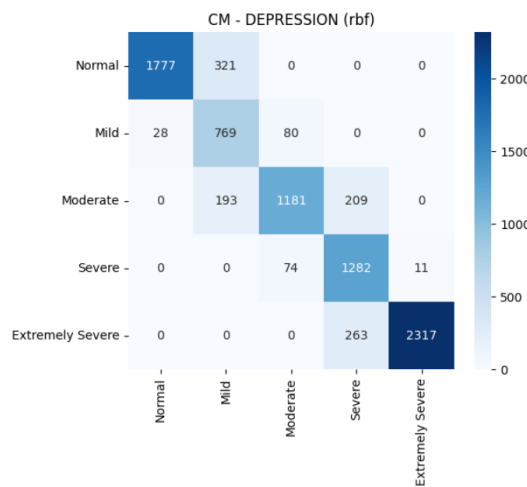
Tabel 6. Contoh Hasil Prediksi Model LSSVM

Data Uji	Label Aktual	Hasil Prediksi
1	<i>Normal</i>	<i>Normal</i>
2	<i>Mild</i>	<i>Mild</i>
3	<i>Extremely Severe</i>	<i>Extremely Severe</i>
4	<i>Normal</i>	<i>Normal</i>
5	<i>Extremely Severe</i>	<i>Extremely Severe</i>
6	<i>Moderate</i>	<i>Moderate</i>
7	<i>Severe</i>	<i>Severe</i>
8	<i>Moderate</i>	<i>Moderate</i>
9	<i>Severe</i>	<i>Severe</i>
10	<i>Moderate</i>	<i>Moderate</i>

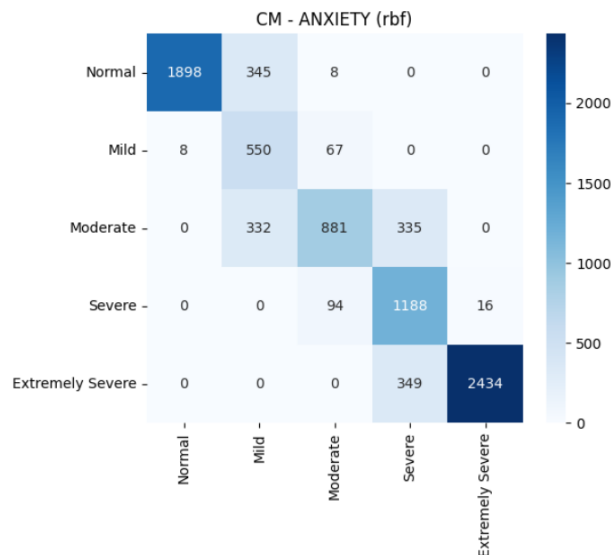
Berdasarkan hasil pengujian tersebut, model mampu mengenali sebagian besar data uji dengan baik. Hal ini menunjukkan bahwa proses pelatihan menggunakan kombinasi LSSVM, SMOTE, dan *Backward Elimination* berhasil mempelajari pola klasifikasi kesehatan mental secara efektif.

Pengujian dilakukan terhadap masing-masing skala psikologis menggunakan tiga varian kernel komputasi LSSVM (*Linear*, *Polynomial*, dan *RBF*). Pada penelitian ini, berdasarkan hasil pencarian parameter optimal (*hyper-parameter tuning*), nilai parameter yang diterapkan pada model untuk kernel *Linear* adalah $C = 10$. Pada kernel *Polynomial*, parameter yang digunakan adalah $C = 10$, $d = 3$, dan $c = 1$. Sementara itu, model terbaik pada kernel *RBF* dicapai dengan menggunakan kombinasi parameter $C = 10$ dan $\gamma = 0.1$

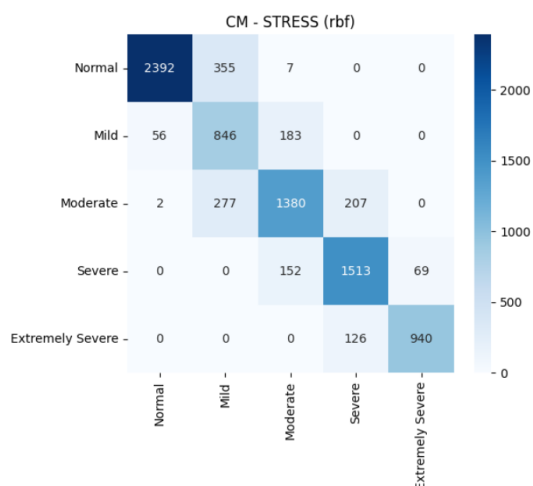
Untuk memvalidasi akurasi prediksi dari model terbaik, evaluasi klasifikasi divisualisasikan menggunakan *Confusion Matrix*. Matriks ini memetakan proporsi tebakan yang benar (*True Positive/True Negative*) berbanding tebakan yang salah (*False Positive/False Negative*) dari sistem terhadap data uji. Hasil pemetaan *Confusion Matrix* dari implementasi LSSVM pada pengukuran tingkat Depresi, Kecemasan, dan Stres secara berturut-turut dapat dilihat pada gambar 2, gambar 3, dan gambar 4.



Gambar 2. Visualisasi *Confusion Matrix* pada skala Depresi



Gambar 3. Visualisasi *Confusion Matrix* pada skala Kecemasan



Gambar 4. Visualisasi Confusion Matrix pada skala Stres

Meskipun model dengan kernel RBF memberikan performa terbaik dibandingkan kernel lainnya, masih terdapat beberapa data yang gagal dikenali dengan benar oleh sistem. Kesalahan klasifikasi umumnya terjadi pada kelas yang memiliki karakteristik psikologis yang saling berdekatan, seperti kategori *Mild* dan *Moderate*, atau *Normal* dan *Mild*. Fenomena tersebut menunjukkan bahwa respons psikologis manusia memiliki pola yang kompleks dan saling tumpang tindih antar tingkat keparahan. Selain itu, beberapa responden memiliki skor DASS-42 yang berada di ambang batas antar kategori sehingga menyulitkan model dalam menentukan kelas yang paling tepat. Contoh data yang mengalami kesalahan klasifikasi dapat dilihat pada Tabel 7.

Tabel 7. Contoh Data Salah Prediksi

Data Uji	Label Aktual	Hasil Prediksi
27	<i>Severe</i>	<i>Moderate</i>
44	<i>Normal</i>	<i>Mild</i>
46	<i>Normal</i>	<i>Mild</i>
47	<i>Mild</i>	<i>Moderate</i>
52	<i>Normal</i>	<i>Mild</i>
53	<i>Moderate</i>	<i>Mild</i>
55	<i>Severe</i>	<i>Moderate</i>
69	<i>Normal</i>	<i>Mild</i>
70	<i>Moderate</i>	<i>Mild</i>
72	<i>Mild</i>	<i>Moderate</i>

Kesalahan prediksi tersebut menunjukkan bahwa beberapa kelas memiliki karakteristik yang sangat berdekatan sehingga batas keputusan antar kategori menjadi sulit dipisahkan secara sempurna oleh model klasifikasi. Kondisi ini umum terjadi pada data kesehatan mental karena respons psikologis manusia bersifat subjektif dan memiliki tingkat variasi yang tinggi. Berdasarkan nilai dari ketiga matriks pengujian di atas, performa metrik dari keseluruhan arsitektur model dan kernel yang dikembangkan kemudian dirangkum pada Tabel 7.

Tabel 7. Performa Klasifikasi LSSVM pada Kernel *Linear*, *Polynomial*, dan RBF

Skala Psikologi	Kernel	Accuracy	Precision	Recall	F1-Score
Depression	<i>Linear</i>	60.81%	57.59%	48.93%	43.80%
	<i>Polynomial</i>	80.41%	78.33%	79.89%	77.57%
	RBF	86.14%	83.89%	86.12%	83.96%
Anxiety	<i>Linear</i>	66.30%	61.59%	52.12%	49.80%
	<i>Polynomial</i>	80.91%	75.64%	78.66%	75.34%
	RBF	81.73%	78.22%	81.64%	77.30%
Stress	<i>Linear</i>	53.71%	53.24%	48.81%	40.18%
	<i>Polynomial</i>	77.47%	78.19%	78.27%	76.29%
	RBF	83.14%	82.03%	82.84%	82.00%

3.5 Pembahasan

Berdasarkan hasil pengujian secara keseluruhan yang dirangkum pada Tabel 4, arsitektur LSSVM menunjukkan variasi performa yang sangat dipengaruhi oleh jenis transformasi ruang dimensi (kernel). Kernel *Radial Basis Function* (RBF) terbukti secara konsisten mendominasi dan menghasilkan metrik kinerja tertinggi di seluruh kategori, dengan *capaian F1-Score* sebesar 83,96% untuk klasifikasi Depresi, 82,00% untuk Stres, dan 77,30% untuk Kecemasan. Sebaliknya, kernel Linear memberikan performa yang sangat tidak memadai, dengan *F1-Score* 40,18% pada skala Stres. Ketimpangan hasil ini mengkonfirmasi bahwa data yang bersumber dari instrumen psikologi klinis manusia seperti respons DASS-42 sangat rentan terhadap bias persepsi dan memiliki distribusi pola yang sangat non-linear serta saling tumpang tindih antar batas kelasnya. Hal ini sejalan dengan karakteristik instrumen psikometri klinis di Indonesia yang menunjukkan adanya tingkat variabilitas respons non-linear yang tinggi antar individu [16]. Kernel RBF membuktikan keandalannya secara matematis dalam memetakan kompleksitas dimensi ruang fitur kuesioner psikologis ini ke tingkat *hyperplane* yang lebih tinggi untuk mengisolasi dan memisahkan kelas-kelas keparahan secara presisi.

Temuan-temuan yang dihasilkan dalam penelitian ini memberikan penguatan (relevansi) yang sangat signifikan dalam upaya pengintegrasian keilmuan dengan riset-riset terdahulu yang sejenis. Keunggulan arsitektur berbasis *Support Vector* dalam menganalisis data kesehatan mental selaras dan secara langsung menguatkan hipotesis yang diajukan oleh [5] serta [17], maupun [18] yang menyimpulkan bahwa algoritma klasifikasi berbasis margin maksimal selalu memberikan kinerja superior dibandingkan algoritma *machine learning* lainnya. Pada aspek optimasi fitur, temuan riset ini mengintegrasikan dan memperluas keberhasilan metode eliminasi atribut. Penelitian yang dilakukan oleh [6] dan [7] membuktikan bahwa penerapan *Backward Elimination* pada data akademik mampu memangkas kompleksitas dimensi, riset ini memberikan penguatan pada dataset berskala masif. Eliminasi lebih dari 40% fitur terbukti berhasil menjaga *F1-Score* tetap tinggi. Lebih lanjut, pengintegrasian teknik SMOTE dalam riset ini mengkonfirmasi temuan teoritis dari penelitian sebelumnya [15] mengenai kelemahan algoritma terhadap bias kelas mayoritas. Pemulihan ekuivalensi data latih dari 19.845 menjadi rata-rata 32.000 data sintetik menguatkan teori bahwa penanganan class *imbalance* diperlukan pada rekam psikologis di mana anomali keparahan seringkali menjadi minoritas ekstrem, sebagaimana yang juga diteliti oleh [10] serta [19] yang membuktikan penyeimbangan data sintesis mampu meminimalisasi bias prediksi kelas mayoritas pada rekam psikologis mahasiswa.

Penelitian ini menawarkan kontribusi yang fundamental bagi pengembangan konsep komputasi dan ilmu pengetahuan di bidang informatika medis. Secara algoritmik, kontribusi utama riset ini adalah pembuktian bahwa implementasi LSSVM mampu memecahkan kendala limitasi skalabilitas yang selama ini menjadi kelemahan SVM standar. Dengan memformulasikan masalah optimasi *Quadratic Programming Problem* (QPP) yang kompleks [8] menjadi sistem persamaan linier biasa [9], penelitian ini mengkonstruksi sebuah *framework* baru yang sangat efisien dan kebal terhadap kendala memori komputasi saat memproses hampir 40.000 data. Konsep ini sejalan dengan temuan [11] serta [20] mengenai keandalan kombinasi LSSVM dan SMOTE pada data tabular berdimensi tinggi. Secara interdisipliner, riset ini juga memberikan wawasan teoritis baru mengenai hierarki prediktor demografis. Implementasi OLS *Backward Elimination* mengungkap fenomena statistik di mana variabel demografis yang sering dianggap krusial seperti usia, edukasi, dan agama justru tereliminasi sepenuhnya. Sistem membuktikan bahwa pola respons murni dari kuesioner, dipadukan dengan variabel spesifik tunggal (*gender, married, urban*), memiliki bobot prediksi matematis yang jauh lebih absolut. Secara praktis, kontribusi arsitektur integratif ini meletakkan fondasi infrastruktur perangkat lunak yang sangat ideal untuk pengembangan aplikasi *self-screening* kesehatan mental berskala masif yang bekerja secara *real-time*, presisi, dan terhindar dari redundansi input demografi pengguna.

4. Simpulan

Penelitian ini berhasil membuktikan bahwa algoritma *Least Squares Support Vector Machine* (LSSVM) efektif dalam mengklasifikasikan tingkat keparahan kesehatan mental berdasarkan respons kuesioner DASS-42. Penggunaan kernel *Radial Basis Function* (RBF) terbukti menjadi fungsi paling optimal dalam menangani pola *non-linear* pada dataset respons psikologis, dengan *capaian F1-Score* sebesar 83,96% untuk klasifikasi Depresi, 82,00% untuk Stres, dan 77,30% untuk Kecemasan. Integrasi metode *Backward Elimination* berperan krusial dalam mereduksi dimensi atribut lebih dari 40%, meminimalisasi beban komputasi fitur *noise*

hingga menyisakan rata-rata 12 hingga 15 fitur esensial. Sementara itu, penerapan teknik SMOTE secara efektif memulihkan ekuivalensi data latih, memastikan model memiliki sensitivitas klasifikasi yang seimbang dan terhindar dari bias terhadap kelas mayoritas. Secara keseluruhan, pendekatan reduksi dimensi dan penyeimbangan data pada arsitektur LSSVM ini diharapkan mampu menjadi fondasi *artificial intelligence* yang tangguh bagi pengembangan instrumen *screening* kesehatan mental yang presisi dan efisien di masa depan.

Daftar Referensi

- [1] K. Vitoasmara, F. V. Hidayah, N. I. Purnamasari, R. Y. Aprillia, and L. D. D. A, "Gangguan Mental (Mental Disorders)," *Student Res. J.*, vol. 2, no. 3, pp. 57–68, 2024.
- [2] M. A. Hakim and N. V. Aristawati, "Mengukur depresi, kecemasan, dan stres pada kelompok dewasa awal di Indonesia: Uji validitas dan reliabilitas konstruk DASS-21," *J. Psikol. Ulayat*, vol. 10, no. 2, pp. 232–250, 2023, doi: 10.24854/jpu553.
- [3] N. A. SHAYAN, A.-R. NIAZI, A. M. WASEQ, and H. ÖZCEBE, "Depression, Anxiety, and Stress Scales 42 (DASS-42) in Dari-Language: Validity and Reliability Study in Adults, Herat, Afghanistan," *Bezmialem Sci.*, vol. 9, no. 3, pp. 356–362, 2021, doi: 10.14235/bas.galenos.2020.4250.
- [4] V. M. A. Simalango and W. F. Senjaya, "Evaluasi Algoritma Decision Tree dan Random Forest serta Efektivitas Feature Selection dalam Memprediksi Kesehatan Mental," *J. Strateg. - J. Maranatha*, vol. 7, no. 1, pp. 201–215, 2025, [Online]. Available: <https://strategi.it.maranatha.edu/index.php/strategi/article/view/583>
- [5] S. Juwariyah, A. Hulvi, N. Riduan, and K. Kusri, "Mengukur Faktor Demografi Psikologis: Memprediksi Depresi, Kecemasan, dan Stres dengan menggunakan Machine Learning," *Komputika J. Sist. Komput.*, vol. 13, no. 2, pp. 149–156, 2024, doi: 10.34010/komputika.v13i2.11793.
- [6] S. D. Amalia, M. A. Barata, and P. E. Yuwita, "Optimization of Random Forest Algorithm with Backward Elimination Method in Classification of Academic Stress Levels," *J. Appl. Informatics Comput.*, vol. 9, no. 3, pp. 633–641, 2025, doi: 10.30871/jaic.v9i3.9280.
- [7] E. P. Efendi, M. A. Barata, and A. D. Ardianti, "Komparasi Algoritma Svm Dan C4 . 5 Dengan Backward Elimination Untuk Klasifikasi Tingkat Stres Mahasiswa Comparative Analysis Of Svm And C4 . 5 Algorithms With Backward," *JIKO (Jurnal Inform. dan Komputer)*, no. 3, pp. 539–549, 2024, doi: 10.26798/jiko.v9i3.2130.
- [8] Y. Xia, J. Zhao, Q. Ding, and A. Jiang, "Incipient Chiller Fault Diagnosis Using an Optimized Least Squares Support Vector Machine With Gravitational Search Algorithm," *Front. Energy Res.*, vol. 9, no. November, pp. 1–13, 2021, doi: 10.3389/fenrg.2021.755649.
- [9] S. Acharya, T. Kar, U. C. Samal, and P. K. Patra, "Performance Comparison between SVM and LS-SVM for Rice Leaf Disease detection," *ICST Trans. Scalable Inf. Syst.*, vol. 10, no. 6, pp. 1–7, 2023, doi: 10.4108/eetsis.3940.
- [10] F. M. Basysyar, G. Dwilestari, and A. I. Purnamasari, "Analysis Student Emotions and Mental Health on Cumulative Gpa Using Machine Learning and Smote," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 10, no. 2, pp. 361–368, 2024, doi: 10.33480/jitk.v10i2.5967.
- [11] M. R. Pribadi, H. D. Purnomo, and Hendry, "A three-step combination strategy for addressing outliers and class imbalance in software defect prediction," *IAES Int. J. Artif. Intell.*, vol. 13, no. 3, pp. 2987–2998, 2024, doi: 10.11591/ijai.v13.i3.pp2987-2998.
- [12] J. T. Hancock and T. M. Khoshgoftaar, "Survey on categorical data for neural networks," *J. Big Data*, 2020, doi: 10.1186/s40537-020-00305-w.
- [13] D. H. Maulud and A. M. Abdulazeez, "A Review on Linear Regression Comprehensive in Machine Learning," *J. Appl. Sci. Technol. Trends*, vol. 01, no. 04, pp. 140–147, 2020, doi: 10.38094/jastt1457.
- [14] W. Nugroho, "Optimasi Metode K-Nearest Neighbours dengan Backward Elimination Menggunakan Dataset Software Effort Estimation Bianglala Informatika," *J. Khatulistiwa Inform.*, vol. 8, no. 2, pp. 129–133, 2020, [Online]. Available: [https://download.garuda.kemdikbud.go.id/article.php?article=2191946&val=10501&title=Optimasi Metode K-Nearest Neighbours dengan Backward Elimination Menggunakan Dataset Software Effort Estimation](https://download.garuda.kemdikbud.go.id/article.php?article=2191946&val=10501&title=Optimasi%20Metode%20K-Nearest%20Neighbours%20dengan%20Backward%20Elimination%20Menggunakan%20Dataset%20Software%20Effort%20Estimation)
- [15] Y. Desnelita, N. Nasution, L. Suryati, and F. Zoromi, "Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang Impact of SMOTE on Random Forest Classifier Performance based on Imbalanced Data," *Matrik J. Manajemen, Tek.*

- Inform. dan Rekayasa Komput.*, vol. 21, no. 3, 2022, doi: 10.30812/matrik.v21i3.1726.
- [16] R. Widyana, Sumiharso, and R. M. Safitri, "Psychometric Properties of Internet-Administered Version of Depression , Anxiety and Stress Scales (DASS-42) in sample Indonesian Adult Psychometric Properties of Internet-Administered Version of Depression , Anxiety and Stress Scales (DASS-42) in sam," *Talent Dev Excell*, no. June, 2022, [Online]. Available: https://www.researchgate.net/profile/Rahma-Widyana/publication/361179345_Psychometric_Properties_of_Internet-Administered_Version_of_Depression_Anxiety_and_Stress_Scales_DASS-42_in_sample_Indonesian_Adult/links/62a13f1d6886635d5ccaf778/Psychometric-Properties-of-Internet-Administered-Version-of-Depression-Anxiety-and-Stress-Scales-DASS-42-in-sample-Indonesian-Adult.pdf?_sg%5B0%5D=started_experiment_milestone&origin=journalDetail
- [17] N. Fathirachman Mahing, A. Lazuardi Gunawan, A. Foresta Azhar Zen, F. Abdurrachman Bachtiar, and S. Agung Wicaksono, "Klasifikasi Tingkat Stress dari Data Berbentuk Teks dengan Menggunakan Algoritma Support Vector Machine (SVM) dan Random Forest," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 5, pp. 1067–1076, 2024, doi: 10.25126/jtiik.1078010.
- [18] Z. Fatah and U. Hasanah, "Prediksi Tingkat Stress dan Kesehatan Mental Mahasiswa Menggunakan Algoritma SVM," *JAMASTIKA (Jurnal Mhs. Tek. Inform.*, vol. 4, no. 2, pp. 200–207, 2025, [Online]. Available: <https://jurnal.unw.ac.id/index.php/jamastika/article/view/4542>
- [19] G. Samudra, A. W. Anggraeni, A. R. B. Jamroni, A. Sarif, and W. Wiyanto, "Efektivitas Teknik SMOTE Dalam Meningkatkan Performa Naïve Bayes Deteksi Gangguan Kecemasan Mahasiswa," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 12, no. 3, 2025, doi: 10.35957/jatisi.v12i3.12197.
- [20] T. F. Husin, M. R. Pribadi, and Yohannes, "Implementation of LSSVM in Classification of Software Defect Prediction Data with Feature Selection," in *2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, Jakarta, 2022, pp. 126–131. doi: 10.23919/EECSI56542.2022.9946611.