



KLASIFIKASI PENYAKIT MIGRAIN MENGGUNAKAN METODE SUPPORT VECTOR MACHINE

Rizky Wahyudi¹, Kristin Impana Manik², Muhammad Alfin³, John Bush Henrydunan⁴, Muhammad Haikal Al Majid⁵, Kana Saputra⁶

¹²³⁴⁵⁶Program Studi Ilmu Komputer, FMIPA, Universitas Negeri Medan

Jl. William Iskandar Ps. V, Kenangan Baru, Kec. Percut Sei Tuan, Kabupaten Deli Serdang, Sumatera Utara 20221

¹rizkyw23@mhs.unimed.ac.id, ²kristinmanik.4233250033@mhs.unimed.ac.id,

³mhdalfinaja@mhs.unimed.ac.id, ⁴johnbushsimarmata@mhs.unimed.ac.id,

⁵muhhammadhaikalalmajid.4232250004@mhs.unimed.ac.id, ⁶kanasaputras@unimed.ac.id

Abstract

Migraine is a neurological condition that can significantly reduce a person's quality of life. While migraines are typically diagnosed clinically based on patient medical history, this approach is often subjective and prone to classification errors. This research aims to develop a more precise migraine classification model using Support Vector Machines (SVM). The study began with collecting migraine datasets from Kaggle, followed by descriptive analysis to understand the data characteristics. The preprocessing phase involved handling missing values, identifying and removing outliers using Winsorization and Interquartile Range (IQR), and normalizing numerical features with StandardScaler. Class imbalance in the dataset was addressed using the Synthetic Minority Over-sampling Technique (SMOTE). The SVM model was optimized using GridSearchCV with RBF kernel, C, and Gamma parameters. Model evaluation was conducted across three data-splitting scenarios (70:30, 80:20, and 90:10) using accuracy, F1-score, and ROC-AUC metrics. Optimal results were achieved with the 70:30 ratio, yielding an accuracy of 75.00%, an F1-score of 70.03%, and a ROC-AUC of 84.28%. These findings demonstrate the SVM model's effectiveness in classifying migraines. This study advances the application of machine learning for more accurate and reliable migraine diagnosis.

Keywords : support vector machine, migraine classification, machine learning, IQR, GridSearchCV

Abstrak

Migrain adalah gangguan neurologis yang dapat menurunkan kualitas hidup penderitanya. Diagnosis migrain umumnya dilakukan secara klinis berdasarkan riwayat medis pasien, namun metode ini sering bersifat subjektif dan berisiko mengalami kesalahan klasifikasi. Tujuan dari penelitian ini adalah untuk membuat model klasifikasi migrain berbasis Support Vector Machine (SVM) untuk diagnosis yang lebih tepat. Tahapan penelitian dimulai dengan pengumpulan dataset migrain dari Kaggle. Selanjutnya dilakukan analisis deskriptif untuk memahami karakteristik data. Pada tahap preprocessing, nilai yang hilang ditangani, outlier dideteksi dan dihapus menggunakan Interquartile Range (IQR) dan Winsorization, serta fitur numerik dinormalisasi dengan StandardScaler. Ketidakseimbangan kelas dalam dataset diatasi menggunakan Synthetic Minority Over-sampling Technique (SMOTE). Model SVM dioptimalkan menggunakan GridSearchCV dengan parameter C, Gamma, dan kernel RBF. Tiga skema pembagian data yaitu 70:30, 80:20, dan 90:10 digunakan untuk mengevaluasi model menggunakan Akurasi, F1-Score, dan ROC-AUC. Dengan akurasi sebesar 75,00%, F1-Score 70,03%, dan ROC-AUC 84,28%, hasil optimal dicapai pada rasio 70:30. Temuan ini menunjukkan seberapa baik model SVM dapat mengkategorikan migrain. Penelitian ini memajukan penggunaan pembelajaran mesin untuk diagnosis migrain yang lebih tepat dan dapat dipercaya.

Kata kunci : support vector machine, klasifikasi migrain, machine learning, IQR, GridSearchCV



1. PENDAHULUAN

Serangan sakit kepala berulang adalah ciri khas migrain, suatu kondisi neurologis yang sering kali disertai dengan keluhan sensorik lainnya, seperti sensitivitas terhadap cahaya dan suara, serta mual[1]. Menurut World Health Organization (WHO), migrain termasuk dalam 10 penyakit yang paling menyebabkan kecacatan di dunia, sehingga memengaruhi produktivitas individu dan kualitas hidup mereka secara signifikan. Di Indonesia sendiri, data Riskesdas tahun 2018 menunjukkan bahwa gangguan saraf, termasuk migrain, menjadi salah satu masalah kesehatan yang banyak dikeluhkan oleh masyarakat.

Kondisi neurologis yang dikenal sebagai migrain ditandai dengan nyeri sakit kepala berdenyut yang dapat berlangsung selama empat hingga tujuh puluh dua jam. Aktivitas fisik dapat memperburuk rasa sakit, yang sering kali bersifat unilateral. Migrain lebih sering terjadi pada wanita daripada pria[2]. Perubahan hormon yang terjadi selama siklus menstruasi dapat menjadi penyebabnya. Selain itu, sebuah penelitian oleh [3], mengidentifikasi korelasi yang kuat antara stres akademik dan peningkatan keluhan migrain di kalangan mahasiswa, yang mengindikasikan bahwa stres merupakan kontributor yang signifikan terhadap episode migrain.

Dalam dunia medis, diagnosis penyakit, termasuk migrain, masih banyak bergantung pada wawancara klinis dan riwayat kesehatan pasien, yang sering kali subjektif dan dipengaruhi oleh pengalaman dokter, sehingga berisiko menimbulkan ketidaktepatan diagnosis [4]. Kemajuan kecerdasan buatan (AI) dan pembelajaran mesin (Machine Learning) membuka peluang untuk meningkatkan akurasi serta efisiensi diagnosis melalui metode otomatis.

Sebuah subbidang kecerdasan buatan yang disebut pembelajaran mesin (machine learning/ML) memungkinkan sistem komputer untuk secara otomatis belajar dari kinerja masa lalu dan meningkatkannya tanpa pemrograman eksplisit. Dalam pembelajaran mesin, data dianalisis, pola diidentifikasi, dan prediksi atau penilaian dibuat dengan menggunakan model dan algoritme statistik[5].

Salah satu teknik yang banyak digunakan adalah Support Vector Machine (SVM), yang dikenal memiliki kinerja unggul dalam klasifikasi data dengan mencari hyperplane optimal untuk memisahkan kelas dengan margin maksimum.

Kemampuannya dalam menangani data berdimensi tinggi dan menghasilkan model yang akurat menjadikan SVM sebagai solusi potensial dalam penerapan medis, termasuk klasifikasi dan diagnosis penyakit [6].

Support Vector Machines (SVM) pertama kali dibuat oleh Vapnik pada tahun 1992 sebagai salah satu metode pembelajaran mesin yang bekerja dengan prinsip Structural Risk Minimization/SRM [7]. Dalam proses pengenalan pola, SVM menggunakan pendekatan statistik untuk menemukan support vector terbaik dan menggunakan margin terbesar untuk memisahkan dua kelas yang berbeda [8]. Dalam teori pembelajaran mesin, SVM adalah komponen penting. Dalam banyak aplikasi sains dan teknik, SVM sangat efektif terutama dalam masalah klasifikasi (pengenalan pola) [9]. Support Vector Machine (SVM) merupakan salah satu teknik yang sering digunakan dalam klasifikasi medis. SVM memiliki keunggulan dalam mengelola dataset berdimensi tinggi dan dapat berfungsi secara efektif meskipun dengan jumlah data pelatihan yang sedikit [10].

Mengingat masalah ini, tujuan dari penelitian ini adalah untuk mengklasifikasikan migrain dengan menggunakan pendekatan Support Vector Machine. Diharapkan bahwa metode ini akan meningkatkan akurasi diagnosis migrain dan membantu para profesional kesehatan untuk membuat keputusan yang lebih cepat dan akurat.

2. TINJAUAN PUSTAKA

Klasifikasi penyakit neurologis, termasuk migrain, telah menjadi subjek dari berbagai penelitian sebelumnya yang menggunakan berbagai metodologi pembelajaran mesin.

Penelitian sebelumnya dilakukan oleh [11] yang berjudul "Klasifikasi Penyakit Cacar Monyet Menggunakan Metode Support Vector Machine". Dalam penelitian ini, kernel Radial Basis Function (RBF) digunakan untuk pengujian, dan confusion matrix digunakan untuk membandingkan rasio data training dan testing dengan rasio 70:30, 80:20, dan 90:10. Menurut temuan pengujian, akurasi optimal 65% dicapai dengan parameter SVM, khususnya $C = 10$ dan $\gamma = 1$, ketika rasio data latih dan data uji adalah 90:10. Temuan ini menunjukkan bahwa pendekatan SVM berkinerja baik dalam klasifikasi penyakit cacar monyet.



Efektivitas Support Vector Machines (SVM) dalam mengklasifikasikan berbagai penyakit juga telah ditunjukkan oleh penelitian lain. Sebagai contoh, sebuah penelitian yang dilakukan oleh [12] menunjukkan bahwa dengan menggunakan kriteria penilaian Precision, Recall, dan F1-Score. SVM dapat mengklasifikasikan hipertensi dengan tingkat akurasi 97,15%. Selanjutnya, penelitian yang dilakukan oleh [6] menunjukkan bahwa dengan menggunakan langkah-langkah evaluasi yang sama. SVM dapat digunakan untuk mendiagnosis masalah kesehatan manusia dengan akurasi 99% melalui aplikasi web berbasis kecerdasan buatan. Namun, saat ini hanya ada sedikit penggunaan SVM dalam klasifikasi penyakit migrain, sehingga diperlukan lebih banyak penelitian untuk memahami potensinya di bidang ini.

Penelitian oleh [13] di bidang kategorisasi berbasis pembelajaran mesin menunjukkan bahwa metode seperti support vector machines (SVM), k-nearest neighbors (KNN), pohon keputusan, naïve bayes, dan KNN memiliki tingkat kemampuan yang berbeda dalam mengklasifikasikan data medis. Ketidakseimbangan data merupakan kendala utama saat menggunakan teknik-teknik ini, dan teknik resampling seperti oversampling dan undersampling sering digunakan untuk mengatasi masalah ini. Selain itu, kemampuan model untuk mengklasifikasikan penyakit secara signifikan dipengaruhi oleh pemilihan fitur yang tepat.

Penggunaan SVM dalam gangguan neurologis telah memberikan hasil yang menggembirakan. Dalam penelitian tentang klasifikasi penyakit stroke, [14] menggunakan SVM dan Particle Swarm Optimization (PSO) dan menemukan bahwa kombinasi ini meningkatkan akurasi model. Lebih lanjut, penelitian [15] menunjukkan bahwa metode ekstraksi fitur seperti Pitch Period Entropy (PPE) dapat meningkatkan sensitivitas SVM dalam klasifikasi penyakit Parkinson. Setelah modifikasi parameter dan pemilihan fitur yang sesuai, model klasifikasi yang diusulkan juga menunjukkan kinerja yang lebih baik.

Menurut penelitian yang dilakukan oleh [16] mengenai deteksi sinusitis dapat dilakukan menggunakan Support Vector Machine (SVM) dengan fitur suhu tubuh dan warna ingus. Data dari sensor MLX90614 dan TCS3200 diproses menggunakan Arduino Uno sebelum diklasifikasikan oleh SVM menjadi normal atau sinusitis. Dengan 40 data latih dan 20 data uji,

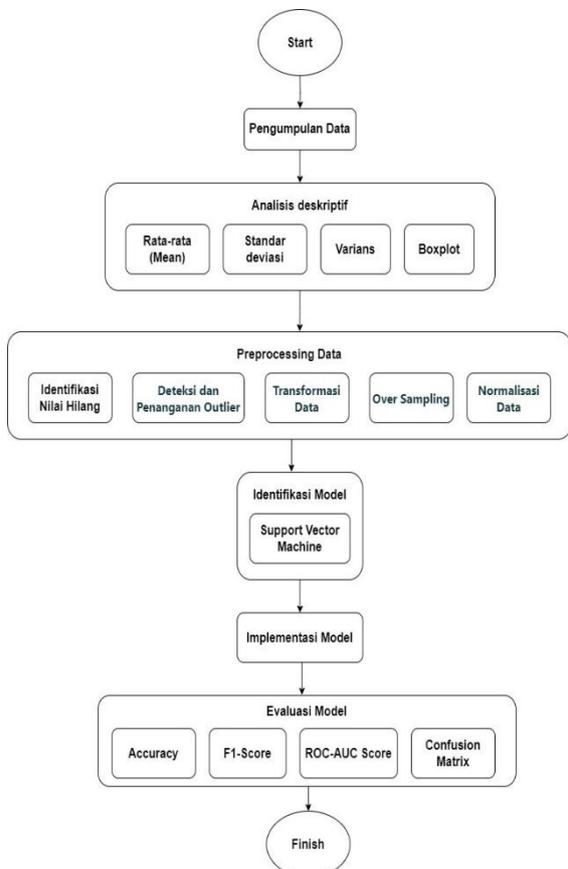
model ini mencapai akurasi 85% dengan waktu komputasi 42 ms. Hasil ini menunjukkan bahwa kombinasi sensor dan SVM dapat menjadi alternatif deteksi dini sinusitis yang lebih cepat dan murah dibandingkan MRI.

Penelitian ini menambahkan sesuatu yang baru pada optimasi model dan pemrosesan data. Tiga rasio data pelatihan dan pengujian (70:30, 80:20, dan 90:10) dibandingkan untuk mengidentifikasi desain yang optimal, dan outlier ditangani dengan menggunakan pemangkasan IQR dan winsorisasi (10%). Untuk meningkatkan akurasi model, kombinasi parameter $C = [0.1, 1, 10]$ dan $\text{Gamma} = [0.1, 0.01, 0.001]$, dan Kernel (RBF) yang terbaik dijamin melalui optimasi hiperparameter menggunakan GridSearchCV. Untuk analisis tambahan, evaluasi dilakukan dengan menggunakan F1-score, ROC AUC Score, dan Confusion Matrix dalam bentuk heatmap. Untuk meningkatkan ketergantungan model klasifikasi migrain, stabilitas distribusi fitur setelah prapemrosesan juga diperhitungkan. Penelitian ini menawarkan wawasan tentang optimasi model dan membantu meningkatkan akurasi kategorisasi migrain berbasis machine learning melalui metodologi yang metodis dan menyeluruh.

3. METODOLOGI PENELITIAN

3.1. Tahapan Penelitian

Penelitian ini menggunakan performa algoritma Support Vector Machine dalam klasifikasi penyakit migrain menggunakan dataset migraine dari situs Kaggle. Diagram alir pada Gambar 1 memberikan gambaran umum tentang keseluruhan alur kerja penelitian.



Gambar 1. Diagram Alir Penelitian

3.2. Pengumpulan Data

Penelitian ini menggunakan dataset migraine yaitu "migraine_data.csv" dari Kaggle yang mencakup 400 data pasien yang terkena migrain. Dataset yang digunakan ini adalah dataset medis yang berisi informasi mengenai karakteristik penderita migrain dan klasifikasi jenis migrain yang dialami. Karakteristik dataset ini adalah multivariat dengan tipe data kategorikal dan numerik (integer). Ada satu kelas target dan dua puluh tiga fitur dalam dataset ini. Berbagai jenis migrain termasuk dalam dataset ini, termasuk "Migrain without aura", "Typical aura with migrain", "Basilar-type aura", "Migrain hemiplegia sporadis", "Migrain hemiplegia familial", "Typical aura without migrain", dan "Other".

Tabel 1. Deskripsi Fitur Pada Dataset Migraine

No	Fitur	Deskripsi	Tipe Data
1	Age	Usia pasien	Int64

2	Duration	Durasi serangan migrain	Int64
3	Frequency	Frekuensi serangan migrain	Int64
4	Location	Lokasi nyeri	Int64
5	Character	Karakteristik nyeri	Int64
6	Intensity	Intensitas nyeri	Int64
7	Nausea	Ada tidaknya mual	Int64
8	Vomit	Ada tidaknya muntah	Int64
9	Phonophobia	Sensitivitas terhadap suara	Int64
10	Photophobia	Sensitivitas terhadap cahaya	Int64
11	Visual	Gangguan penglihatan	Int64
12	Sensory	Gangguan sensorik	Int64
13	Dysphasia	Gangguan berbicara	Int64
14	Dysarthria	Kesulitan artikulasi	Int64
15	Vertigo	Pusing berputar	Int64
16	Tinnitus	Telinga berdenging	Int64
17	Hypoacusis	Gangguan pendengaran	Int64
18	Diplopia	Penglihatan ganda	Int64
19	Defect	Defisit neurologis	Int64
20	Ataxia	Gangguan koordinasi	Int64
21	Conscience	Kesadaran	Int64
22	Paresthesia	Kesemutan	Int64
23	DPF	Riwayat keluarga	Int64
24	Type	Jenis migrain (target)	object

3.3. Analisis Deskriptif

Analisis deskriptif dilakukan untuk memahami karakteristik dataset migrain, termasuk pemeriksaan apakah terdapat skala data yang tidak seragam. Parameter yang digunakan dalam analisis ini adalah statistik deskriptif seperti rata-rata (mean), standar deviasi, nilai minimum dan maksimum untuk



melihat rentang nilai dari setiap fitur. Analisis juga menggunakan boxplot untuk memeriksa distribusi data serta mendeteksi adanya pencilan (outlier) pada fitur-fitur seperti Age dan Duration yang memiliki rentang nilai lebih besar dibandingkan fitur lainnya. Selain itu, countplot digunakan untuk melihat distribusi kelas target, yang menunjukkan ketidakseimbangan pada dataset dengan kelas "Typical aura with migraine" mendominasi jumlah sampel dibandingkan kelas lainnya seperti "Sporadic hemiplegic migraine" dan "Familial hemiplegic migraine". Analisis ini penting dilakukan sebelum proses normalisasi dan penanganan ketidakseimbangan kelas untuk memastikan model SVM dapat bekerja optimal dalam mengklasifikasikan jenis-jenis migrain.

3.4. Preprocessing Data

Berikut ini adalah bagian dari tahap preprocessing penelitian ini. Pada awalnya, 400 titik data pasien dalam dataset migrain diperiksa untuk mencari nilai yang hilang. Boxplot kemudian digunakan untuk mengidentifikasi pencilan, terutama pada karakteristik Usia dan Durasi, yang lebih bervariasi daripada fitur lainnya. Teknik Interquartile Range (IQR) dan winsorization (10%) digunakan untuk menangani pencilan untuk mengurangi nilai ekstrem tanpa menghapus data. StandardScaler digunakan untuk menstandarkan semua fitur numerik selama tahap normalisasi data. Hal ini memastikan bahwa setiap fitur memiliki skala yang seragam dengan rata-rata 0 dan standar deviasi 1, yang meningkatkan kinerja model SVM. Terakhir, untuk mengatasi ketidakseimbangan kelas pada dataset migrain, dilakukan oversampling menggunakan SMOTE (Synthetic Minority Over-sampling Technique) agar model dapat belajar dengan lebih adil terhadap semua kategori migrain seperti "Typical aura with migraine", "Migraine without aura", "Basilar-type aura", "Sporadic hemiplegic migraine", "Familial hemiplegic migraine", "Typical aura without migraine" dan "Other".

3.5. Identifikasi Model

Model yang digunakan dalam penelitian ini adalah algoritma Support Vector Machine (SVM). Metode Support Vector Machine (SVM) berlandaskan pada konsep minimisasi risiko struktural (SRM). SVM bertujuan untuk menentukan hyperplane optimal dalam ruang input yang dapat membedakan dua kelas. Hyperplane ini ditentukan dengan mengukur margin antara kedua kelas dan mencari titik maksimumnya [17]. Dalam klasifikasi penyakit migrain, SVM dipilih karena kemampuannya dalam menangani data berdimensi tinggi dan menghasilkan batas keputusan yang jelas. Algoritma ini bekerja dengan memetakan data ke ruang dimensi yang lebih tinggi menggunakan fungsi kernel, sehingga data yang tidak dapat dipisahkan secara linear pada ruang asalnya dapat dipisahkan pada ruang baru tersebut.

Kernel Radial Basis Function (RBF) digunakan dalam penelitian ini karena kemampuannya dalam mengelola interaksi fitur non-linear. Parameter C, yang mengatur pertukaran antara margin dan kesalahan klasifikasi, dan parameter Gamma, yang menentukan tingkat pengaruh satu sampel data terhadap sampel lainnya, adalah parameter utama yang telah dioptimalkan. Nilai C yang rendah menghasilkan margin yang lebar tetapi toleransi kesalahan klasifikasi yang lebih tinggi, sedangkan nilai C yang tinggi akan menghasilkan margin yang kecil dengan kesalahan klasifikasi yang rendah pada data pelatihan tetapi berisiko overfitting. $C = [0.1, 1, 10]$ dan $\text{Gamma} = [0.1, 0.01, 0.001]$ digabungkan dalam GridSearchCV untuk optimasi parameter untuk mengidentifikasi konfigurasi terbaik yang menghasilkan akurasi klasifikasi maksimum untuk tujuh kategori migrain dalam set data.

3.6. Implementasi Model

Proses implementasi Support Vector Machine (SVM) dimulai dengan inisialisasi model menggunakan parameter optimal yang diperoleh melalui GridSearchCV. Parameter yang dioptimasi meliputi nilai C yang mengontrol trade-off antara margin dan kesalahan klasifikasi, nilai gamma



yang mengatur kompleksitas batas keputusan, serta pemilihan kernel RBF yang mampu menangani data non-linear. Selanjutnya, model dilatih menggunakan dataset migrain yang telah melalui tahap preprocessing (penanganan outlier, normalisasi, dan SMOTE), dengan tiga skenario pembagian data yaitu rasio 70:30, 80:20, dan 90:10 antara training dan testing set. Pembagian data dilakukan dengan metode stratified split untuk memastikan proporsi kelas tetap seimbang di kedua set data. Proses training dilakukan menggunakan metode fit() dengan data training, kemudian model diuji pada data testing untuk mendapatkan prediksi kelas migrain.

3.7. Evaluasi Model

Evaluasi model dilakukan menggunakan Accuracy, F1-Score, dan ROC-AUC Score untuk menilai kehandalan SVM dalam mengklasifikasikan penyakit migrain. Confusion Matrix sendiri digunakan sebagai alat ukur kinerja dalam pemrosesan data mining dengan model klasifikasi, di mana pengukuran ini memberikan informasi perbandingan antara hasil klasifikasi algoritma yang digunakan dengan hasil klasifikasi sebenarnya [18]. Confusion matrix disajikan dalam bentuk tabel matriks, di mana jika dataset terdiri dari dua kelas, satu kelas dianggap sebagai positif, sementara kelas lainnya dianggap sebagai negatif [19].

Tabel 2. Confussion Matrix

	ACTUAL: TRUE	ACTUAL: FALSE
PREDICT: TRUE	TP (True Positive)	FP (False Positive)
PREDICT: FALSE	TN (True Negative)	FN (False Negative)

Pada Tabel 2 di atas, terdapat empat istilah dalam confusion matrix yang menggambarkan hasil klasifikasi, yaitu True Positive (TP) yang menunjukkan data positif yang diprediksi benar, True Negative (TN) untuk data negatif yang diprediksi benar, False Positive (FP) sebagai error tipe 1 di mana data negatif diprediksi sebagai positif, dan False Negative (FN)

sebagai error tipe 2 di mana data positif diprediksi sebagai negatif.

Dalam penerapannya, confusion matrix dimanfaatkan untuk menghitung nilai akurasi, presisi, recall, dan F1 score. Akurasi menunjukkan tingkat keberhasilan model dalam mengklasifikasikan data secara tepat, dan dihitung menggunakan rumus:

$$\text{Akurasi} = \frac{TN + TP}{TN + TP + FP + FN} \quad (1)$$

(Sumber: [18])

Presisi mengukur akurasi antara data yang diminta dan hasil prediksi model, dengan rumus:

$$\text{Presisi} = \frac{TP}{TP + FP} \quad (2)$$

(Sumber: [18])

Recall atau sensitivity menggambarkan sejauh mana model berhasil menemukan informasi yang benar, dihitung dengan rumus:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

(Sumber: [18])

Sedangkan F1 score merupakan perbandingan rata-rata presisi dan recall yang dibobotkan, dihitung dengan rumus:

$$\text{F1 Score} = 2 \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \quad (4)$$

(Sumber: [18])

Alat visualisasi untuk mengevaluasi kinerja model klasifikasi adalah kurva Receiver Operating Characteristic (ROC), yang membandingkan tingkat positif dan negatif yang sebenarnya. Kemampuan model untuk membedakan antara kelas positif dan negatif diukur dengan area di bawah kurva ROC, atau AUC (Area Under the Curve), yang berkisar dari 0 hingga 1. Nilai AUC yang mendekati 1 menunjukkan kinerja yang optimal; semakin besar AUC, semakin baik model tersebut dalam melakukan klasifikasi. [20], [21].

$$\text{AUC} = \frac{1}{2} \sum_{i=0}^n (x_{i+1} - x_i)(y_{i+1} - y_i) \quad (5)$$



Dimana:

x_i dan y_i adalah titik-titik pada kurva ROC

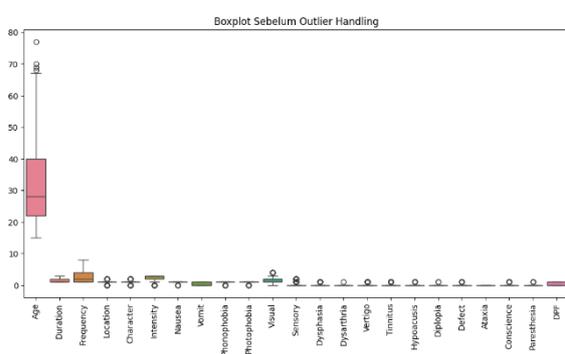
x_i adalah False Positive Rate (FPR) dan y_i adalah

True Positive Rate (TPR) pada threshold tertentu [22].

Defect	0.015	0.12	0.0	1.0
Ataxia	0.0	0.0	0.0	0.0
Conscience	0.017	0.13	0.0	1.0
Paresthesia	0.07	0.08	0.0	1.0
DPF	0.41	0.49	0.0	1.0

4. HASIL DAN PEMBAHASAN

4.1. Analisis Data



Gambar 1. Boxplot Sebelum Outlier Handling

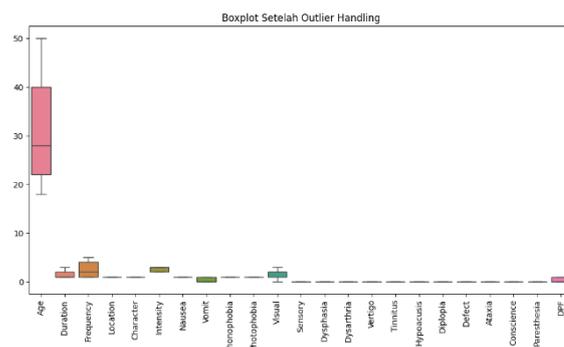
Berdasarkan Gambar 1, Boxplot menunjukkan adanya outlier signifikan terutama pada fitur *Age* dan *Duration*, sementara fitur lainnya memiliki skala kecil dengan beberapa outlier. Perbedaan skala antar fitur ini mengindikasikan perlunya normalisasi agar tidak mendominasi model. Outlier yang ada dapat mempengaruhi hasil analisis sehingga perlu ditangani menggunakan IQR.

Tabel 3. Statistik Deskriptif Fitur

Fitur	Mean	Std	Min	Max
Age	31.75	12.13	15.0	77.0
Duration	1.61	0.77	1.0	3.0
Frequency	2.37	1.67	1.0	8.0
Location	0.97	0.26	1.0	2.0
Character	0.97	0.27	1.0	2.0
Intensity	2.47	0.76	2.0	3.0
Nausea	0.99	0.11	0.0	1.0
Vomit	0.32	0.46	0.0	1.0
Phonophobia	0.98	0.14	0.0	1.0
Photophobia	0.98	0.14	0.0	1.0
Dysarthria	0.002	0.05	0.0	1.0
Vertigo	0.12	0.33	0.0	1.0
Tinnitus	0.06	0.23	0.0	1.0
Hypoacusis	0.01	0.12	0.0	1.0
Diplopia	0.005	0.07	0.0	1.0

Dari statistik deskriptif yang ditunjukkan pada Tabel 3, terlihat bahwa sebagian besar fitur bersifat biner (0 atau 1). Skewness pada beberapa fitur juga mengindikasikan perlunya normalisasi atau scaling sebelum diterapkan pada SVM, agar model dapat bekerja secara optimal. Sebelum proses thresholding dan segmentasi fitur, penanganan outlier dan balancing data perlu dilakukan untuk memastikan akurasi model tidak terpengaruh oleh distribusi yang tidak seimbang.

4.2. Preprocessing Data



Gambar 2. Boxplot setelah Outlier Handling

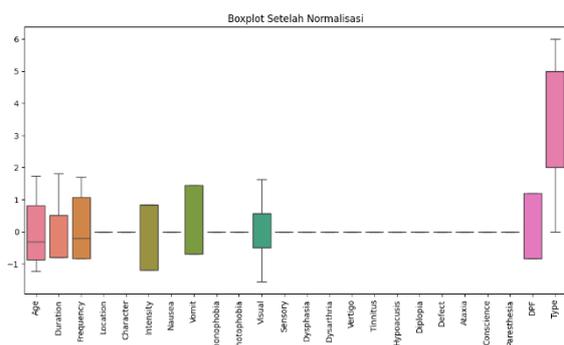
Setelah penanganan outlier, boxplot pada Gambar 4 menunjukkan bahwa distribusi data telah lebih terkontrol, dengan outlier yang berkurang secara signifikan dibandingkan sebelumnya. Variabel "Age" masih memiliki rentang nilai yang lebih besar dibandingkan fitur lainnya, tetapi pencilan yang ekstrem tampaknya telah dikurangi. Penanganan outlier menggunakan IQR dan winsorization (10%) telah diterapkan untuk meningkatkan kualitas data sebelum digunakan dalam SVM untuk klasifikasi penyakit migrain. Dengan data yang lebih bersih, model diharapkan dapat memiliki akurasi yang lebih baik dan menghindari dampak negatif dari nilai ekstrem.



#	Column	Non-Null Count	Dtype
0	Age	400 non-null	float64
1	Duration	400 non-null	float64
2	Frequency	400 non-null	float64
3	Location	400 non-null	float64
4	Character	400 non-null	float64
5	Intensity	400 non-null	float64
6	Nausea	400 non-null	float64
7	Vomit	400 non-null	float64
8	Phonophobia	400 non-null	float64
9	Photophobia	400 non-null	float64
10	Visual	400 non-null	float64
11	Sensory	400 non-null	float64
12	Dysphasia	400 non-null	float64
13	Dysarthria	400 non-null	float64
14	Vertigo	400 non-null	float64
15	Tinnitus	400 non-null	float64
16	Hypoacusis	400 non-null	float64
17	Diplopia	400 non-null	float64
18	Defect	400 non-null	float64
19	Ataxia	400 non-null	float64
20	Conscience	400 non-null	float64
21	Paresthesia	400 non-null	float64
22	DPF	400 non-null	float64
23	Type	400 non-null	int64

Gambar 3. Informasi Dataset Setelah Ditransformasi

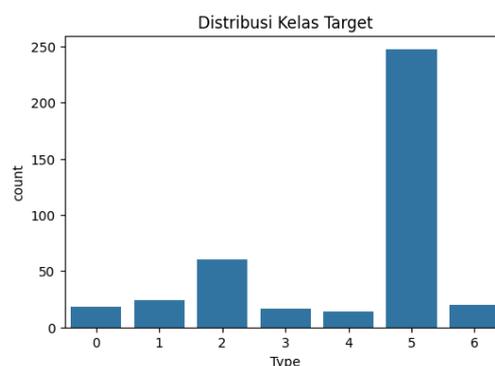
Hasil transformasi data pada Gambar 5 menunjukkan bahwa seluruh fitur numerik telah dikonversi ke float64 karena disebabkan oleh proses standarisasi menggunakan StandardScaler. Ini memastikan bahwa setiap fitur memiliki skala yang seragam, dengan mean 0 dan standar deviasi 1, sehingga meningkatkan kinerja model berbasis jarak seperti SVM. Sementara itu, label kelas "Type" tetap dalam format int64, menunjukkan bahwa label target tidak ikut terdampak oleh normalisasi dan tetap dalam bentuk kategorikal numerik yang sesuai untuk klasifikasi.



Gambar 4. Boxplot Setelah Normalisasi

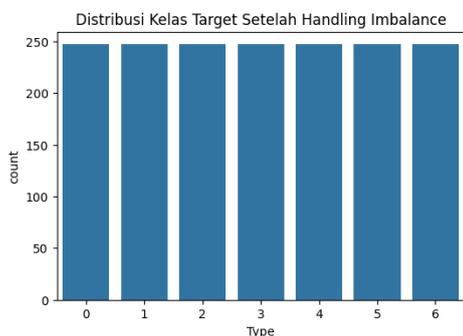
Setelah normalisasi, boxplot pada Gambar 5 menunjukkan bahwa skala semua fitur telah diseragamkan dengan nilai berada dalam rentang yang lebih terkontrol. Penerapan StandardScaler berhasil menstandarisasi seluruh

fitur ke dalam skala yang lebih seimbang, dengan mean 0 dan standar deviasi 1, sehingga model berbasis jarak seperti SVM dapat bekerja lebih optimal. Terlihat bahwa label kelas "Type" tidak mengalami normalisasi karena merupakan target pada dataset. Variabel yang sebelumnya memiliki perbedaan skala besar, seperti "Age" dan "DPF", kini telah berada dalam skala yang lebih seimbang dengan fitur lainnya. Hal ini penting dalam SVM karena algoritma ini sensitif terhadap skala fitur dan normalisasi akan membantu meningkatkan kinerja model dengan mencegah fitur dengan nilai besar mendominasi proses klasifikasi.



Gambar 5. Count Plot Sebelum Handling Imbalance

Distribusi kelas target pada Gambar 7 menunjukkan bahwa dataset tidak seimbang, dengan kelas 5 mendominasi jumlah sampel, sementara kelas lainnya memiliki jumlah data yang jauh lebih sedikit. Ketidakseimbangan ini bisa menyebabkan bias pada model SVM, karena model cenderung lebih akurat dalam memprediksi kelas mayoritas dibandingkan kelas minoritas. Untuk mengatasi masalah ini, teknik oversampling menggunakan SMOTE (Synthetic Minority Over-sampling Technique) diterapkan agar model tetap mampu mengenali semua kelas dengan baik.



Gambar 6. Count Plot Setelah Handling Imbalance

Setelah menerapkan SMOTE, distribusi kelas target menjadi seimbang dengan setiap kelas memiliki jumlah sampel yang kurang lebih sama seperti terlihat pada Gambar 8. Hal ini mengatasi masalah ketidakseimbangan data sebelumnya sehingga model SVM tidak akan condong ke kelas mayoritas dan dapat belajar dengan lebih adil terhadap semua kategori. Dengan distribusi yang merata ini, performa model diharapkan meningkat.

4.3. Pembagian Data

Setelah melalui tahap preprocessing dan penyeimbangan data menggunakan SMOTE, dataset dibagi menjadi data latih (DL) dan data uji (DU) dengan tiga skenario rasio pembagian, yaitu 70:30 (DL=280, DU=120), 80:20 (DL=320, DU=80), dan 90:10 (DL=360, DU=40) untuk mengevaluasi pengaruh jumlah data pelatihan terhadap performa model SVM. Pembagian dilakukan menggunakan metode stratified split, sehingga proporsi masing-masing kelas tetap seimbang dalam data latih dan uji.

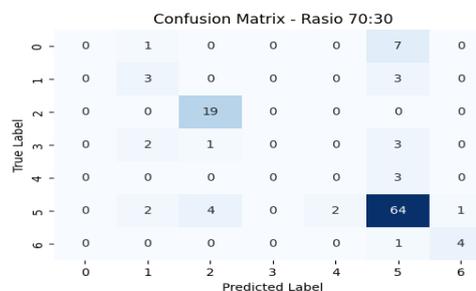
4.4. Evaluasi Model

Temuan evaluasi model yang terdiri dari akurasi, F1-score, ROC AUC, dan Confusion Matrix dengan metrik True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN), diperoleh dari hasil pengujian yang telah dilakukan.

Tabel 4. Hasil Evaluasi Model

Rasio	Accuracy	F1-Score	ROC-AUC
70:30	75,00%	70,03%	84,28%
80:20	70,00%	66,36%	81,92%
90:10	70,00%	67,22%	82,27%

Evaluasi model seperti pada Tabel 4 menunjukkan bahwa performa klasifikasi menggunakan SVM dengan kernel RBF cukup bervariasi di berbagai rasio pembagian data. Dengan rasio 70:30, model mencapai accuracy 75,00%, F1-score 70,03%, dan ROC-AUC 84,28%, yang menunjukkan keseimbangan yang cukup baik antara klasifikasi positif dan negatif. Pada rasio 80:20, terjadi sedikit penurunan pada accuracy menjadi 70,00%, F1-score turun menjadi 66,36%, dan ROC-AUC menurun ke 81,92%, menandakan bahwa model mungkin mengalami kesulitan dalam menangani data uji yang lebih sedikit dibandingkan sebelumnya. Namun, pada rasio 90:10, accuracy tetap 70,00%, dengan F1-score yang sedikit meningkat menjadi 67,22% dan ROC-AUC mencapai 82,27%. Hal ini menunjukkan bahwa model tetap mampu mempertahankan performa meskipun data uji semakin sedikit, tetapi tidak memberikan peningkatan yang signifikan. Parameter terbaik yang diperoleh dari GridSearchCV adalah C = 1, gamma = 0.1, dan kernel = 'rbf', yang menunjukkan bahwa model bekerja optimal dengan margin keputusan yang moderat dan tingkat kompleksitas sedang.



Gambar 7. Confusion Matrix Rasio 70:30



Confusion Matrix dengan rasio 70:30 seperti pada Gambar 9 menunjukkan performa model SVM dengan kernel RBF dalam mengklasifikasikan data uji ke dalam tujuh kelas. Model mampu mengklasifikasikan kelas 5 dengan baik, terbukti dari 64 prediksi benar, namun masih terjadi kesalahan klasifikasi pada kelas lainnya. Misalnya, kelas 0 sering diklasifikasikan sebagai kelas 5, sementara kelas 2 memiliki 19 prediksi benar tetapi juga mengalami beberapa kesalahan klasifikasi. Kesalahan umum terjadi pada kelas dengan jumlah sampel kecil, menunjukkan bahwa model mungkin mengalami kesulitan dalam menangani kelas minoritas.

Confusion Matrix - Rasio 80:20

0	0	1	0	0	0	5	0
1	0	2	0	0	0	1	0
2	0	0	13	0	0	0	0
3	0	1	1	0	0	2	0
4	0	0	0	0	0	2	0
5	1	1	2	0	2	39	4
6	0	0	0	0	0	1	2
	0	1	2	3	4	5	6

Gambar 8. Confusion Matrix Rasio 70:30

Confusion Matrix dengan rasio 80:20 seperti pada Gambar 10 menunjukkan bahwa model SVM dengan kernel RBF berhasil mengklasifikasikan sebagian besar sampel dengan baik, terutama pada kelas mayoritas seperti kelas 5 yang memiliki jumlah prediksi benar tertinggi (39). Selain itu, kelas 2 juga menunjukkan hasil yang cukup baik dengan 13 prediksi benar. Model mampu mengenali pola utama dalam data, meskipun masih terdapat beberapa variasi dalam klasifikasi. Distribusi prediksi ini mencerminkan kemampuan model dalam menggeneralisasi pola dari data latih ke data uji dengan rasio 80:20.

Confusion Matrix - Rasio 90:10

0	0	1	0	0	0	3	0
1	0	1	0	0	0	0	0
2	0	0	4	0	0	0	0
3	0	0	1	0	0	0	0
4	0	0	0	0	0	1	0
5	0	0	2	0	2	22	2
6	0	0	0	0	0	0	1
	0	1	2	3	4	5	6

Gambar 9. Confusion Matrix Rasio 90:10

Confusion Matrix dengan rasio 90:10 seperti pada Gambar 11 menunjukkan bahwa model SVM dengan kernel RBF tetap mampu mengenali pola klasifikasi dengan baik, terutama pada kelas 5 dengan prediksi benar tertinggi (22). Model juga dapat mengklasifikasikan beberapa kelas lain dengan cukup akurat, seperti kelas 2 dan kelas 0. Namun, terdapat sedikit kesalahan prediksi pada beberapa kelas, misalnya kelas 6 yang masih memiliki beberapa misclassifications. Meskipun demikian, dengan lebih banyak data latih, model tetap mempertahankan kemampuannya dalam mengenali pola utama dalam dataset.

4.5. Perbandingan dengan Studi Sebelumnya dan Analisis Keterbatasan

Model SVM pada penelitian ini memiliki kinerja yang lebih baik dengan akurasi 75,00% dibandingkan dengan model SVM yang digunakan pada penelitian sebelumnya oleh Anugrah dkk. [11], yang menggunakan SVM untuk klasifikasi penyakit cacar monyet dengan akurasi mencapai 65%. Penggunaan metode preprocessing yang lebih menyeluruh dalam penelitian ini, termasuk SMOTE untuk mengatasi ketidakseimbangan kelas dan penanganan outlier dengan pendekatan IQR dan winsorization, kemungkinan besar menjadi penyebab peningkatan akurasi.

Performa model SVM dalam penelitian ini juga konsisten dengan penelitian oleh Ramadhani dkk. [6] dan Putri & Fahmi [12], yang menunjukkan bahwa SVM dapat mengkategorikan penyakit secara akurat. Kompleksitas data migrain yang lebih tinggi, yang memiliki tujuh kelas, menjadi salah satu alasan penurunan akurasi dalam



penelitian ini (75% vs 97-99% dalam penelitian sebelumnya), karena penelitian sebelumnya biasanya menggunakan kategorisasi biner.

Meski demikian, penelitian ini memiliki sejumlah keterbatasan. Dataset yang digunakan berasal dari sumber sekunder (Kaggle), sehingga berpotensi mengandung bias seleksi karena tidak semua variabel klinis utama tercakup. Selain itu, teknik SMOTE yang digunakan untuk mengatasi ketidakseimbangan kelas dapat menghasilkan data sintetik yang tidak sepenuhnya merepresentasikan kondisi nyata, sehingga berisiko menyebabkan overfitting. Pemilihan fitur juga dilakukan tanpa konsultasi langsung dengan tenaga medis, sehingga beberapa karakteristik penting untuk diagnosis migrain mungkin tidak terakomodasi secara optimal.

5. KESIMPULAN DAN SARAN

Dengan akurasi 75,00%, F1-score 70,03%, dan ROC-AUC 84,28%, hasil penilaian model menunjukkan bahwa rasio 70:30 memberikan kinerja yang optimal, yang menunjukkan bahwa model tersebut dapat mengklasifikasikan data secara efektif dan seimbang dalam menangani kelas positif dan negatif. Meskipun rasio 80:20 dan 90:10 menunjukkan perbedaan kinerja yang kecil, variasi ini tidak signifikan secara statistik, yang menunjukkan bahwa model ini tetap cukup stabil dalam berbagai situasi pembagian data. Model ini dapat belajar dari distribusi data yang lebih merata berkat pendekatan SMOTE, yang membantu mengatasi ketidakseimbangan kelas. Model ini menunjukkan hasil klasifikasi yang menjanjikan dengan menggunakan hyperparameter optimal $C = 1$, $\gamma = 0.1$, dan kernel = rbf. Untuk penelitian selanjutnya, disarankan eksplorasi metode balancing lain seperti ADASYN untuk lebih meningkatkan akurasi model. Selain itu, perbandingan dengan metode lain seperti deep learning dapat dilakukan untuk melihat pendekatan yang lebih optimal dalam klasifikasi.

6. UCAPAN TERIMA KASIH

Peneliti mengucapkan terima kasih kepada situs Kaggle yang telah menyediakan dataset migrain yang menjadi dasar penelitian ini sehingga memungkinkan analisis klasifikasi penyakit migrain dengan metode Support Vector

Machine. Selain itu, peneliti juga menyampaikan apresiasi yang sebesar-besarnya kepada Bapak Kana Saputra, S.Pd., M.Kom., selaku dosen pembimbing yang telah memberikan arahan, bimbingan, serta masukan berharga dalam setiap tahap penelitian ini. Dukungan dari berbagai pihak ini sangat membantu dalam mencapai hasil penelitian yang lebih baik.

DAFTAR PUSTAKA:

- [1] N. Meylakh and L. A. Henderson, "Exploring alterations in sensory pathways in migraine," *J. Headache Pain*, vol. 23, no. 1, pp. 1-10, 2022, doi: 10.1186/s10194-021-01371-y.
- [2] C. P. Azolia and R. Andriani, "Faktor-faktor yang berpengaruh terhadap migrain tanpa aura pada mahasiswa kedokteran Universitas Tarumanagara," *Tarumanagara Med. J.*, vol. 5, no. 2, pp. 236-243, 2023, doi: 10.24912/tmj.v5i2.24787.
- [3] J. T. Rumat and A. Rantepadang, "Hubungan Stres Akademik dengan Keluhan Migrain pada Mahasiswa," *Nutr. J.*, vol. 7, no. 2, p. 150, 2023, doi: 10.37771/nj.v7i2.943.
- [4] M. Yuichi and Y. A. Susetyo, "Klasifikasi Penyakit Migrain dengan Metode Naïve Bayes pada Dataset Kaggle," *J. Indones. Manaj. Inform. dan Komun.*, vol. 6, no. 1, pp. 139-151, Jan. 2025, doi: 10.35870/jimik.v6i1.1150.
- [5] C. A. Putra, R. Pratama, and T. Sutabri, "Analisis Manfaat Machine Learning Pada Next-Generation Firewall Sophos Xg 330 Dalam Mengatasi Serangan Sql Injection," *J. Manaj. Inform. dan Sist. Inf.*, vol. 6, no. 2, pp. 197-204, 2023, doi: 10.36595/misi.v6i2.886.
- [6] M. A. Ramadhani *et al.*, "IMPLEMENTASI ALGORITMA SUPPORT VECTOR MACHINE (SVM) UNTUK DIAGNOSIS KESEHATAN MANUSIA BERBASIS WEB," *J. Ners*, vol. 9, no. 1, pp. 896-902, 2025, doi: 10.31004/jn.v9i1.31481.
- [7] N. Maulidah, R. Supriyadi, D. Y. Utami, F. N. Hasan, A. Fauzi, and A. Christian, "Prediksi Penyakit Diabetes Melitus Menggunakan Metode Support Vector Machine dan Naive Bayes," *Indones. J. Softw. Eng.*, vol. 7, no. 1, pp. 63-68, 2021, doi:



- 10.31294/ijse.v7i1.10279.
- [8] R. S. Tantika and A. Kudus, "Penggunaan Metode Support Vector Machine Klasifikasi Multiclass pada Data Pasien Penyakit Tiroid," *Bandung Conf. Ser. Stat.*, vol. 2, no. 2, pp. 159–166, 2022, doi: 10.29313/bcss.v2i2.3590.
- [9] N. G. Ramadhan and A. Khoirunnisa, "Klasifikasi Data Malaria Menggunakan Metode Support Vector Machine," *J. Media Inform. Budidarma*, vol. 5, no. 4, p. 1580, 2021, doi: 10.30865/mib.v5i4.3347.
- [10] F. Abdusyukur, "Penerapan Algoritma Support Vector Machine (Svm) Untuk Klasifikasi Pencemaran Nama Baik Di Media Sosial Twitter," *Komputa J. Ilm. Komput. dan Inform.*, vol. 12, no. 1, pp. 73–82, 2023, doi: 10.34010/komputa.v12i1.9418.
- [11] W. Anugrah, E. Haerani, Y. Yusra, and L. Oktavia, "Klasifikasi Penyakit Cacar Monyet Menggunakan Metode Support Vector Machine," *J. Comput. Syst. Informatics*, vol. 5, no. 3, p. 45, 2024, doi: 10.47065/josyc.v5i3.5149.
- [12] E. Z. H. Putri and H. Fahmi, "Implementasi Metode Support Vector Machine pada Klasifikasi Diagnosis Penyakit Hipertensi," *J. Ris. Mhs. Mat.*, vol. 3, no. 5, pp. 241–250, Jun. 2024, doi: 10.18860/jrmm.v3i5.27312.
- [13] I. F. Yuliati and P. R. Sihombing, "Penerapan Metode Machine Learning dalam Klasifikasi Risiko Kejadian Berat Badan Lahir Rendah di Indonesia," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 20, no. 2, pp. 417–426, 2021, doi: 10.30812/matrik.v20i2.1174.
- [14] Y. Ayuningtyas and I. M. Suartana, "Klasifikasi Penyakit Stroke Menggunakan Support Vector Machine (SVM) dan Particle Swarm Optimization (PSO)," *JINACS (Journal Informatics Comput. Sci.*, vol. 04, no. 04, pp. 452–457, 2023.
- [15] P. K. Handayani, "Penerapan Algoritma Support Vector Machine (Svm) Untuk Analisis Pola Klasifikasi Pada Parkinson'S Dataset," *Indones. J. Technol. Informatics Sci.*, vol. 3, no. 1, pp. 31–35, 2021, doi: 10.24176/ijtis.v3i1.7530.
- [16] B. N. Rahman, R. Maulana, and F. Utamingrum, "Sistem Pendeteksi Penyakit Sinusitis berdasarkan Kondisi Ingus dan Suhu Tubuh menggunakan Support Vector Machine (SVM)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 2, pp. 545–551, 2022, [Online]. Available: <https://j-ptiik.uib.ac.id/index.php/j-ptiik/article/view/10555>
- [17] S. Rahayu and Y. Yamasari, "Klasifikasi Penyakit Stroke dengan Metode Support Vector Machine (SVM)," *J. Informatics Comput. Sci.*, vol. 5, no. 03, pp. 440–446, 2024, doi: 10.26740/jinacs.v5n03.p440-446.
- [18] P. Romadloni, B. Adhi Kusuma, and W. Maulana Baihaqi, "Komparasi Metode Pembelajaran Mesin Untuk Implementasi Pengambilan Keputusan Dalam Menentukan Promosi Jabatan Karyawan," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 2, pp. 622–628, 2022, doi: 10.36040/jati.v6i2.5238.
- [19] S. Saikin and K. Kusriani, "Model Data Mining Untuk Karakteristik Data Traveller Pada Perusahaan Tour and Travel," *J. Manaj. Inform. dan Sist. Inf.*, vol. 2, no. 2, p. 61, 2019, doi: 10.36595/misi.v2i2.105.
- [20] V. Arinal and E. Sentosa, "Klasifikasi Tingkat Kesejahteraan RW 006 Kelurahan Kalideres Jakarta Barat dengan Metode K-Nearest Neighbor," *J. Pendidik. dan Konseling*, vol. 4, no. 4, pp. 5621–5638, 2022, doi: 10.31004/jpdk.v4i4.6372.
- [21] D. T. Wilujeng, M. Fatekurohman, and I. M. Tirta, "Analisis Risiko Kredit Perbankan Menggunakan Algoritma K-Nearest Neighbor dan Nearest Weighted K-Nearest Neighbor," *Indones. J. Appl. Stat.*, vol. 5, no. 2, p. 142, 2023, doi: 10.13057/ijas.v5i2.58426.
- [22] Suwarno and A. Abdillah, "Penerapan Algoritma Bayesian Regularization Backpropagation Untuk Memprediksi Penyakit Diabetes," *J. MIPA*, vol. 39, no. 2, pp. 150–158, 2016.