

Evaluasi Kinerja Kernel *Linear*, RBF, dan *Polynomial* pada Model *Support Vector Machine* untuk Prediksi Risiko Hipertensi

Saifuddin^{1*}, Lukman Azhari², Erni Widarti³, Wartono⁴

^{1,3,4}Program Studi Sistem Informasi Kota Cerdas, Fakultas Teknik, Universitas Tunas Pembangunan Surakarta

²Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Tangerang

Email: ¹saifuddin@lecture.utp.ac.id, ²lukman.azhari@ft-umt.ac.id, ³erni.widarti@lecture.utp.ac.id,

⁴wartono@lecture.utp.ac.id

Penulis Korespondensi:^{*}

(received: 10-11-25, revised: 19-11-25, accepted: 26-11-25)

Abstrak

Hipertensi adalah penyakit tidak menular dengan angka kematian tinggi dan sering disebut “silent killer” karena gejalanya kerap tidak terlihat pada awalnya. Deteksi dini diperlukan untuk mencegah komplikasi seperti penyakit jantung dan stroke. Penelitian ini bertujuan mengevaluasi kinerja tiga kernel pada algoritma *Support Vector Machine* (SVM): *Linear*, *Radial Basis Function* (RBF), dan *Polynomial*. Ketiga kernel tersebut dipilih untuk mewakili karakteristik pemetaan data yang berbeda: kernel *linear* berguna untuk memisahkan data secara linier, kernel RBF dapat menangani pola nonlinier yang kompleks, dan kernel *polynomial* untuk memodelkan interaksi antar fitur dengan tingkat kedalaman tertentu. Dataset berasal dari Kaggle dan mencakup 4.240 entri dengan 12 fitur prediktor yang merepresentasikan kondisi klinis dan perilaku pasien. Hasil eksperimen menunjukkan bahwa SVM-*Linear* mencapai *accuracy* 87,26% dengan ROC-AUC 0,9518; SVM-RBF memperoleh *accuracy* 87,38% dengan ROC-AUC 0,9450; dan SVM-*Polynomial* menghasilkan *accuracy* 86,56% dengan ROC-AUC 0,9422. SVM-*Linear* merupakan model paling optimal berdasarkan kombinasi *F1-score* dan ROC-AUC karena memberikan keseimbangan terbaik antara ketepatan dan sensitivitas. Sementara itu, meskipun SVM-RBF mencatat *accuracy* tertinggi, peningkatannya tidak konsisten pada metrik lain, dan SVM-*Polynomial* menunjukkan performa sedikit lebih rendah karena kompleksitas model yang kurang sesuai dengan karakteristik data. Penelitian ini berkontribusi dengan memberikan analisis komparatif terkait performa berbagai kernel SVM dalam klasifikasi risiko hipertensi.

Kata Kunci: Hipertensi, Klasifikasi, Kernel SVM, Prediksi Risiko, *Support Vector Machine*

Abstract

Hypertension is a non-communicable disease with a high mortality rate and is often referred to as a “silent killer” because its symptoms frequently do not appear in the early stages. Early detection is necessary to prevent complications such as heart disease and stroke. This study aims to evaluate the performance of three kernels in the Support Vector Machine (SVM) algorithm: Linear, Radial Basis Function (RBF), and Polynomial. These kernels were selected to represent different data mapping characteristics: the linear kernel is useful for linearly separable data, the RBF kernel can handle complex nonlinear patterns, and the polynomial kernel is used to model interactions between features at varying levels of depth. The dataset was obtained from Kaggle and consists of 4,240 entries with 12 predictor features representing patients' clinical conditions and behavioral characteristics. The experimental results show that SVM-Linear achieves an accuracy of 87.26% with a ROC-AUC of 0.9518; SVM-RBF reaches an accuracy of 87.38% with a ROC-AUC of 0.9450; and SVM-Polynomial produces an accuracy of 86.56% with a ROC-AUC of 0.9422. SVM-Linear is the most optimal model based on the combination of F1-score and ROC-AUC, as it provides the best balance between precision and sensitivity. Meanwhile, although SVM-RBF records the highest accuracy, the improvement is not consistent across other metrics, and SVM-Polynomial shows slightly lower performance due to its model complexity being less aligned with the characteristics of the data. This study contributes by providing a comparative analysis of the performance of various SVM kernels in classifying hypertension risk.

Keywords: Hypertension, Classification, SVM Kernel, Risk Prediction, *Support Vector Machine*

1. PENDAHULUAN

Hipertensi merupakan salah satu penyakit tidak menular yang paling banyak menyebabkan kematian di dunia [1]. Kondisi ini dikenal sebagai *silent killer* karena sering kali tidak menimbulkan gejala pada tahap awal,

tetapi dapat memicu komplikasi serius seperti stroke, gagal ginjal, dan penyakit jantung koroner [2]. Data dari *World Health Organization* (WHO) menunjukkan bahwa lebih dari 1,3 miliar orang di dunia mengalami tekanan darah tinggi, dengan tingkat kesadaran dan pengendalian yang masih rendah terutama di negara berkembang [3]. Dengan demikian, mengenali potensi hipertensi sejak tahap awal merupakan langkah yang sangat penting dalam mengurangi kemungkinan munculnya gangguan kardiovaskular.

Proses pemeriksaan medis konvensional dalam mendeteksi hipertensi umumnya membutuhkan sumber daya yang besar, meliputi waktu, biaya, dan tenaga medis yang terampil. Keterbatasan tersebut mendorong pengembangan pendekatan berbasis *machine learning* yang mampu mengenali pola kompleks pada data klinis secara otomatis. Pendekatan *machine learning* telah banyak digunakan dalam bidang kesehatan karena kemampuannya mengidentifikasi hubungan nonlinier antar variabel dan menghasilkan prediksi yang lebih cepat serta akurat dibandingkan metode statistik tradisional [4].

Salah satu metode pembelajaran mesin yang banyak digunakan dan terbukti handal dalam tugas klasifikasi pada bidang kesehatan adalah Support Vector Machine (SVM). Algoritma ini bekerja dengan memaksimalkan *margin* antar kelas untuk menghasilkan batas keputusan yang optimal, serta terbukti tahan terhadap *overfitting* pada data berdimensi tinggi [5]. Berbagai penelitian telah membuktikan efektivitas SVM dalam mendeteksi penyakit. Misalnya, penggunaan SVM untuk klasifikasi penderita *Gastroesophageal Reflux Disease* (GERD) berhasil mencapai akurasi 82,5% dengan *F1-score* 58,3% [6]. Dalam studi terkait deteksi diabetes, penerapan SVM dengan kernel *Radial Basis Function* (RBF) dilaporkan mampu memperoleh tingkat akurasi mencapai sekitar 91%. [7], sedangkan penelitian lain pada klasifikasi gagal jantung menggunakan kernel *linear* menghasilkan akurasi 89,13%, presisi 86,21%, *recall* 96,15%, dan *F1-score* 90,91% [8].

Meskipun hasil-hasil tersebut menunjukkan potensi SVM dalam diagnosis penyakit, kinerjanya sangat bergantung pada jenis fungsi kernel yang digunakan. Kernel bekerja dengan mentransformasikan data ke dalam ruang fitur yang berdimensi lebih tinggi sehingga pola yang semula sulit dipisahkan dapat menjadi terpisahkan secara linier [9]. Kernel *linear* umumnya efektif untuk data dengan hubungan sederhana antar fitur [10], kernel RBF lebih unggul dalam menangani pola nonlinier kompleks [11], sedangkan kernel *polynomial* mampu memodelkan interaksi antar fitur dengan derajat tertentu [12]. Pemilihan kernel yang tepat menjadi kunci agar model SVM dapat mencapai keseimbangan antara akurasi dan kemampuan generalisasi [13].

Namun, sebagian besar penelitian terdahulu hanya berfokus pada penerapan satu jenis kernel untuk kasus penyakit tertentu tanpa melakukan perbandingan yang menyeluruh. Selain itu, banyak studi belum menilai konsistensi performa model melalui pendekatan validasi ganda yang dapat memastikan reliabilitas hasil. Kondisi ini menunjukkan adanya kesenjangan penelitian dalam hal pemahaman empiris mengenai pengaruh jenis kernel terhadap performa SVM, khususnya dalam konteks prediksi risiko hipertensi menggunakan data klinis pasien.

Berdasarkan permasalahan tersebut, tujuannya penelitian ini dilakukan adalah untuk mengevaluasi kinerja tiga tipe kernel diantaranya *Linear*, *Radial Basis Function* (RBF), dan *Polynomial* pada model SVM untuk prediksi risiko hipertensi. Pemilihan ketiga kernel tersebut mempertimbangkan karakteristik pemetaan data: kernel *Linear* sesuai untuk hubungan antar fitur yang cenderung linier, kernel RBF efektif menangani pola yang nonlinier dan kompleks, sedangkan kernel *Polynomial* memungkinkan penangkapan interaksi fitur melalui fungsi polinomial dengan tingkat kedalaman tertentu. Evaluasi dilakukan menggunakan metrik akurasi, *precision*, *recall*, *F1-score*, dan ROC-AUC dengan dua pendekatan validasi, yaitu *holdout validation* dan *k-fold cross-validation*. Kontribusi utama penelitian ini adalah penyajian analisis komparatif yang komprehensif mengenai performa kernel SVM dalam klasifikasi medis, yang diharapkan dapat menjadi referensi bagi pengembangan sistem prediksi risiko hipertensi berbasis *machine learning* yang lebih andal dan efisien.

2. METODE PENELITIAN

Penelitian ini menerapkan pendekatan kuantitatif berbasis eksperimen untuk mengembangkan serta menilai kinerja model Support Vector Machine (SVM) dengan tiga variasi kernel, yaitu *Linear*, *Radial Basis Function* (RBF), dan *Polynomial*, dalam tugas prediksi risiko hipertensi. Proses penelitian dilakukan melalui beberapa tahap sistematis yang meliputi pengumpulan data, praproses data, eksplorasi dan analisis deskriptif, implementasi model klasifikasi, serta evaluasi kinerja model. Rangkaian tahapan penelitian yang dilakukan ditunjukkan pada Gambar 1.



Gambar 1. Alur Tahapan Penelitian

2.1. Pengumpulan Data

Sumber data penelitian ini diambil dari repositori Kaggle, dengan dataset yang diberi nama *Hypertension Risk Model* [14]. Dataset ini berisi 4.240 data rekam medis dengan 13 atribut yang terdiri atas variabel klinis dan perilaku pasien. Fitur yang tersedia antara lain *age* (usia), *sysBP* (tekanan darah sistolik), *diaBP* (tekanan darah diastolik), *BMI* (indeks massa tubuh), *totChol* (kolesterol total), *glucose* (kadar glukosa darah), *cigsPerDay* (jumlah rokok per hari), serta variabel biner seperti *male*, *diabetes*, *BPMeds*, dan *currentSmoker*. Label target adalah variabel *Risk* yang menunjukkan apakah seseorang berisiko hipertensi (1) atau tidak (0). Dataset ini dipilih karena mencakup faktor-faktor fisiologis dan kebiasaan gaya hidup yang secara medis relevan untuk mendeteksi potensi hipertensi pada individu.

2.2. Praproses Data

Tahapan praproses dilakukan untuk menyiapkan data agar dapat digunakan dalam pelatihan model dengan kualitas yang optimal [15]. Langkah pertama adalah melakukan penanganan terhadap *missing value* menggunakan metode imputasi berbasis median untuk fitur numerik dan modus untuk fitur kategorikal, guna menghindari bias akibat penghapusan data. Selanjutnya, variabel kategorikal seperti *male* dan *currentSmoker* dikonversi ke bentuk numerik menggunakan *label encoding*.

Proses berikutnya adalah normalisasi data menggunakan *StandardScaler* untuk menyetarakan skala antar fitur, karena algoritma SVM sensitif terhadap perbedaan skala data. Setelah proses normalisasi, dilakukan penyeimbangan kelas menggunakan pendekatan SMOTE (*Synthetic Minority Oversampling Technique*) guna menyelesaikan ketimpangan distribusi pada variabel target *Risk*, di mana kelas *No Risk* mendominasi sekitar 69% data. SMOTE menghasilkan sampel sintetis pada kelas minoritas dengan memanfaatkan jarak antar data tetangga terdekat dalam ruang fitur, sehingga distribusi kelas menjadi seimbang tanpa menghapus data asli [16].

Langkah terakhir pada tahap praproses adalah pembagian dataset menjadi dua bagian menggunakan teknik *stratified sampling*, dengan 80% data sebagai data latih dan 20% sebagai data uji. Pemisahan ini menjaga proporsi kelas agar tetap representatif pada kedua subset data, sehingga evaluasi model menjadi lebih adil dan mencerminkan kinerja sebenarnya dari sistem prediksi risiko hipertensi.

2.3. Eksplorasi dan Analisis Deskriptif

Analisis deskriptif dilakukan untuk memahami karakteristik dasar data sebelum proses pelatihan model. Tahap ini mencakup pemeriksaan distribusi kelas pada variabel target untuk memastikan keseimbangan antara data berisiko hipertensi dan tidak berisiko. Selain itu, dilakukan visualisasi histogram pada variabel numerik seperti *age*, *BMI*, dan *sysBP* guna mengamati sebaran data serta mengidentifikasi adanya potensi *outlier*.

Selanjutnya, diterapkan analisis korelasi antar fitur melalui koefisien Pearson untuk mengetahui hubungan antar variabel numerik. Hasil analisis ini memberikan wawasan awal mengenai fitur-fitur yang memiliki kontribusi dominan terhadap risiko hipertensi dan menjadi dasar dalam menafsirkan hasil akhir model.

2.4. Implementasi Model Klasifikasi

Tahap implementasi bertujuan untuk membangun model klasifikasi dengan menerapkan pendekatan *Support Vector Machine* (SVM) melalui tiga jenis kernel, diantaranya *Linear*, *RBF (Radial Basis Function)*, dan *Polynomial*. SVM menentukan batas keputusan berupa sebuah *hyperplane* yang memisahkan dua kelas dalam ruang fitur dengan jarak pemisah yang dibuat sebesar mungkin [15]. Konsep dasar SVM adalah menemukan bidang keputusan yang memaksimalkan jarak antara dua kelas data, sehingga model memiliki kemampuan generalisasi yang lebih baik [17].

Dalam kasus di mana data tidak dapat dipisahkan secara linier pada ruang aslinya, SVM menggunakan fungsi kernel untuk memetakan data tersebut ke ruang berdimensi lebih tinggi (*feature space*) agar pemisahan linier menjadi mungkin. Dengan demikian, kernel berperan penting dalam menentukan bentuk *decision boundary* serta kompleksitas model [18]. Secara matematis, fungsi keputusan SVM dinyatakan melalui persamaan (1).

$$f(x) = \text{sign} \left(\sum_{i=1}^n \alpha_i y_i K(x_i, x) + b \right) \quad (1)$$

dimana x adalah vektor fitur dari sampel uji, x_i adalah vektor fitur dari sampel pelatihan ke-i, $y_i \in \{-1, 1\}$ adalah label kelas dari sampel ke-i, α_i merupakan koefisien Lagrange yang diperoleh dari proses optimasi, $K(x_i, x)$ adalah fungsi kernel yang mengukur kesamaan antara dua sampel, dan b adalah *bias* model yang menggeser posisi *hyperplane*.

Kernel *linear* digunakan ketika hubungan antar fitur bersifat linier, atau ketika data dapat dipisahkan oleh *hyperplane* tanpa perlu pemetaan ke ruang berdimensi tinggi [19]. Kernel ini memiliki bentuk matematis seperti pada persamaan (2).

$$K(x_i, x_j) = x_i^T x_j \quad (2)$$

di mana $x_i^T x_j$ menyatakan hasil perkalian titik (*dot product*) antara dua vektor fitur x_i dan x_j . Semakin besar nilai hasil perkalian titik ini, semakin tinggi tingkat kemiripan antara dua sampel data. Kernel *linear* umumnya digunakan ketika jumlah fitur lebih besar dari jumlah sampel, serta memberikan hasil yang efisien dan mudah diinterpretasikan pada data dengan distribusi sederhana.

Kernel RBF digunakan untuk menangani data yang memiliki pola nonlinier dengan distribusi kompleks [20]. Fungsi kernel ini memetakan data ke ruang fitur berdimensi tinggi menggunakan fungsi Gaussian sebagai ukuran kedekatan antar titik data [21]. Secara matematis kernel ini menggunakan persamaan (3).

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (3)$$

dimana $\|x_i - x_j\|^2$ adalah jarak *Euclidean* kuadrat antara dua vektor fitur, γ adalah parameter kernel yang mengatur tingkat kelengkungan permukaan keputusan.

Nilai γ yang kecil menyebabkan model memiliki *decision boundary* yang lebih halus (cenderung linier), sedangkan nilai γ yang besar menghasilkan batas keputusan yang lebih kompleks dan sensitif terhadap variasi lokal data. Kernel RBF sering menjadi pilihan utama dalam klasifikasi medis karena mampu menangkap hubungan nonlinier antar variabel fisiologis yang sulit diidentifikasi oleh metode linier.

Kernel *polynomial* digunakan untuk memodelkan interaksi antar fitur dengan tingkat kompleksitas yang dapat diatur melalui parameter derajat (*degree*) [22]. Kernel ini menghitung kesamaan antara dua sampel tidak hanya berdasarkan nilai fitur tunggal, tetapi juga kombinasi antar fitur hingga derajat tertentu [23]. Bentuk matematisnya ditunjukkan pada persamaan (4).

$$K(x_i, x_j) = (\gamma x_i^T x_j + r)^d \quad (4)$$

di mana γ adalah koefisien skala yang mengontrol kontribusi dari hasil perkalian titik, r adalah konstanta *bias term* yang menggeser fungsi polinomial, dan d adalah derajat polinomial yang menentukan tingkat kompleksitas hubungan antar fitur.

Setiap model SVM pada penelitian ini dilatih menggunakan parameter dasar dengan pengaturan $C = 1,0$ sebagai faktor regularisasi yang menyeimbangkan antara margin maksimum dan tingkat kesalahan klasifikasi, serta $\gamma = \text{scale}$ untuk kernel nonlinier (RBF dan *polynomial*). Proses pelatihan dilakukan menggunakan data latih yang telah dinormalisasi, kemudian model disimpan untuk tahap evaluasi kinerja.

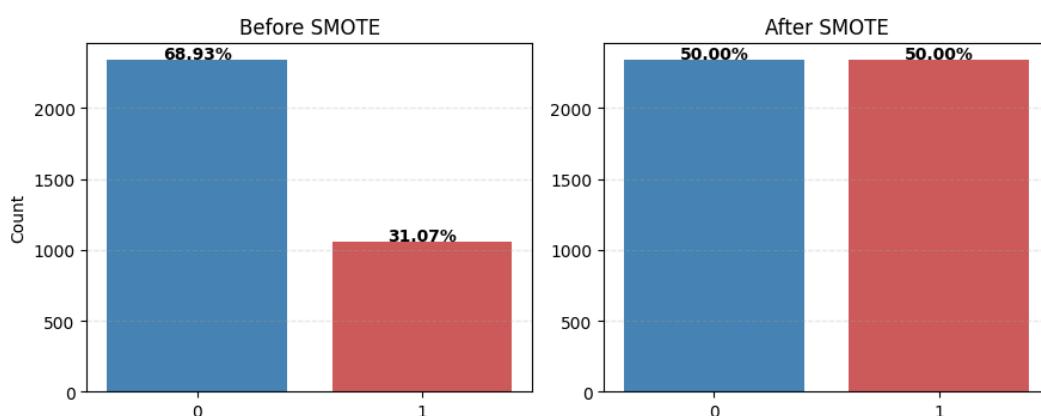
2.5. Evaluasi Kinerja Model

Tahap terakhir penelitian ini adalah evaluasi performa model menggunakan data uji. Evaluasi dilakukan melalui beberapa metrik, yaitu akurasi, presisi, *recall*, *F1-score*, dan ROC-AUC, yang diturunkan dari *confusion matrix*. Selain evaluasi dengan *holdout validation*, dilakukan pula *5-fold cross-validation* untuk memastikan konsistensi hasil pada pembagian data yang berbeda. Visualisasi *Receiver Operating Characteristic (ROC) Curve* digunakan untuk menggambarkan kemampuan model dalam membedakan antara pasien berisiko hipertensi dan tidak berisiko. Nilai *Area Under the Curve* (AUC) digunakan sebagai indikator kuantitatif terhadap performa keseluruhan model [24]. Hasil dari ketiga kernel kemudian dibandingkan untuk menentukan kernel dengan kinerja terbaik dalam memprediksi risiko hipertensi berdasarkan parameter evaluasi yang telah ditetapkan.

3. HASIL DAN PEMBAHASAN

Data penelitian ini bersumber dari *platform Kaggle* dan berisi 4.240 sampel dengan 12 variabel prediktor yang menggambarkan kondisi klinis serta karakteristik perilaku pasien, serta satu variabel target bernama *Risk* yang menunjukkan status risiko hipertensi (0 = tidak berisiko, 1 = berisiko) [14]. Fitur yang digunakan meliputi atribut seperti *age*, *BMI*, *sysBP*, *diaBP*, *totChol*, *heartRate*, serta faktor perilaku seperti *currentSmoker* dan *cigsPerDay*. Sebelum dilakukan pelatihan model, dataset melalui tahap praproses data yang diawali dengan penanganan nilai hilang (*missing value*) menggunakan metode imputasi berbasis median untuk fitur numerik dan modus untuk fitur kategorikal, guna menghindari bias akibat penghapusan data. Selanjutnya, variabel kategorikal seperti *male* dan *currentSmoker* dikonversi ke bentuk numerik menggunakan label encoding.

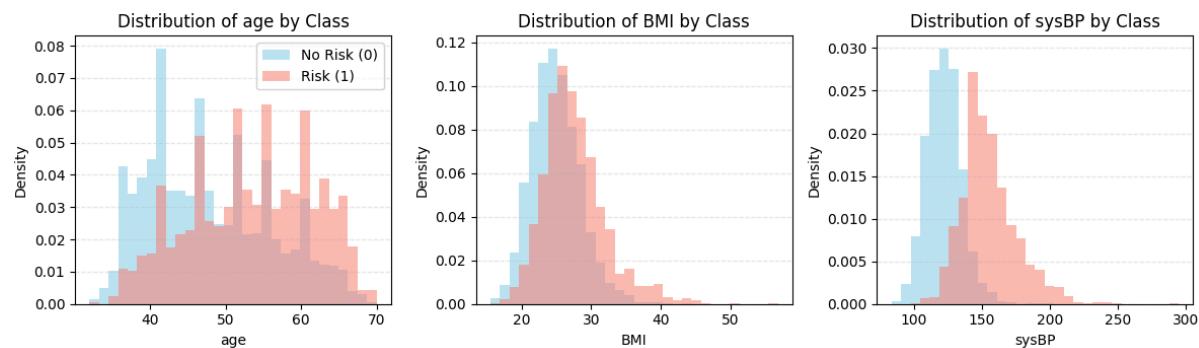
Langkah berikutnya adalah normalisasi data menggunakan *StandardScaler* untuk menyetarakan skala antar fitur, karena algoritma SVM sensitif terhadap perbedaan skala nilai. Setelah data ternormalisasi, dilakukan penyeimbangan kelas menggunakan metode SMOTE untuk menangani ketimpangan distribusi pada variabel target *Risk*. Untuk memahami proporsi awal antara kelompok *No Risk* dan *Risk*, dilakukan visualisasi distribusi kelas sebelum dan sesudah penerapan SMOTE. Visualisasi ini bertujuan mendeteksi ketidakseimbangan data yang dapat memengaruhi performa model klasifikasi, terutama pada algoritma yang peka terhadap proporsi kelas seperti SVM. Hasil distribusi kelas sebelum dan setelah diterapkan SMOTE ditampilkan pada Gambar 2.



Gambar 2. Distribusi Kelas Sebelum dan Sesudah Penerapan

Berdasarkan Gambar 2, analisis menunjukkan bahwa dataset mengalami ketidakseimbangan kelas. Kelas *No Risk* (0) mendominasi sebesar 68,93%, sedangkan kelas *Risk* (1) hanya mencakup 31,07% dari total data. Kondisi ini berpotensi menyebabkan bias pada model klasifikasi, di mana model cenderung memprediksi kelas mayoritas. Setelah penerapan SMOTE, proporsi kelas menjadi seimbang yaitu 50% untuk masing-masing kategori, sehingga model dapat mempelajari karakteristik kedua kelas secara proporsional dan menghasilkan evaluasi yang lebih representatif.

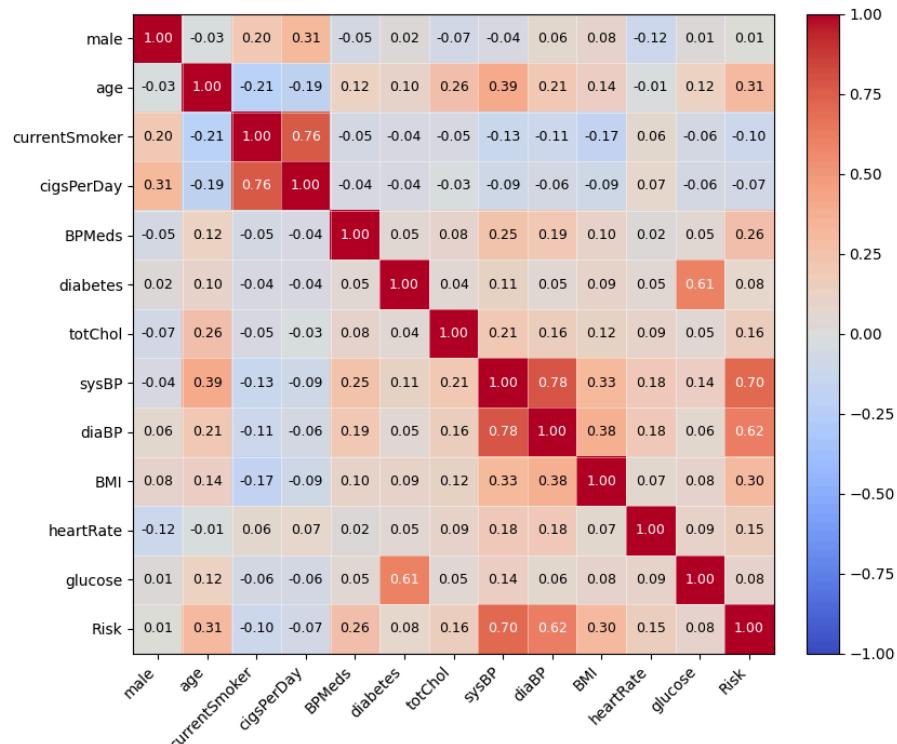
Proses selanjutnya adalah melakukan analisis distribusi variabel numerik terhadap label risiko untuk memahami perbedaan pola antara kelompok *No Risk* dan *Risk*. Analisis ini difokuskan pada variabel *age*, *BMI*, dan *sysBP*, yang dianggap berperan penting dalam menentukan tingkat risiko hipertensi. Visualisasi hasil analisis ditunjukkan pada Gambar 3.



Gambar 3. Distribusi Variabel Numerik Berdasarkan Label Risiko

Berdasarkan Gambar 3, fitur *age*, *BMI*, dan *sysBP* memperlihatkan pola yang berbeda antara kelompok *No Risk* dan *Risk*. Pasien dengan risiko hipertensi cenderung memiliki nilai *age* dan *sysBP* yang lebih tinggi, sedangkan nilai *BMI* menunjukkan sebaran yang relatif serupa namun dengan puncak distribusi yang bergeser ke kanan pada kelompok *Risk*. Perbedaan ini menunjukkan bahwa faktor usia dan tekanan darah sistolik memiliki kontribusi signifikan terhadap peningkatan risiko hipertensi.

Selanjutnya, dilakukan analisis korelasi antar fitur numerik menggunakan Pearson correlation untuk menilai sejauh mana hubungan antarvariabel. Hasilnya ditampilkan pada Gambar 4.



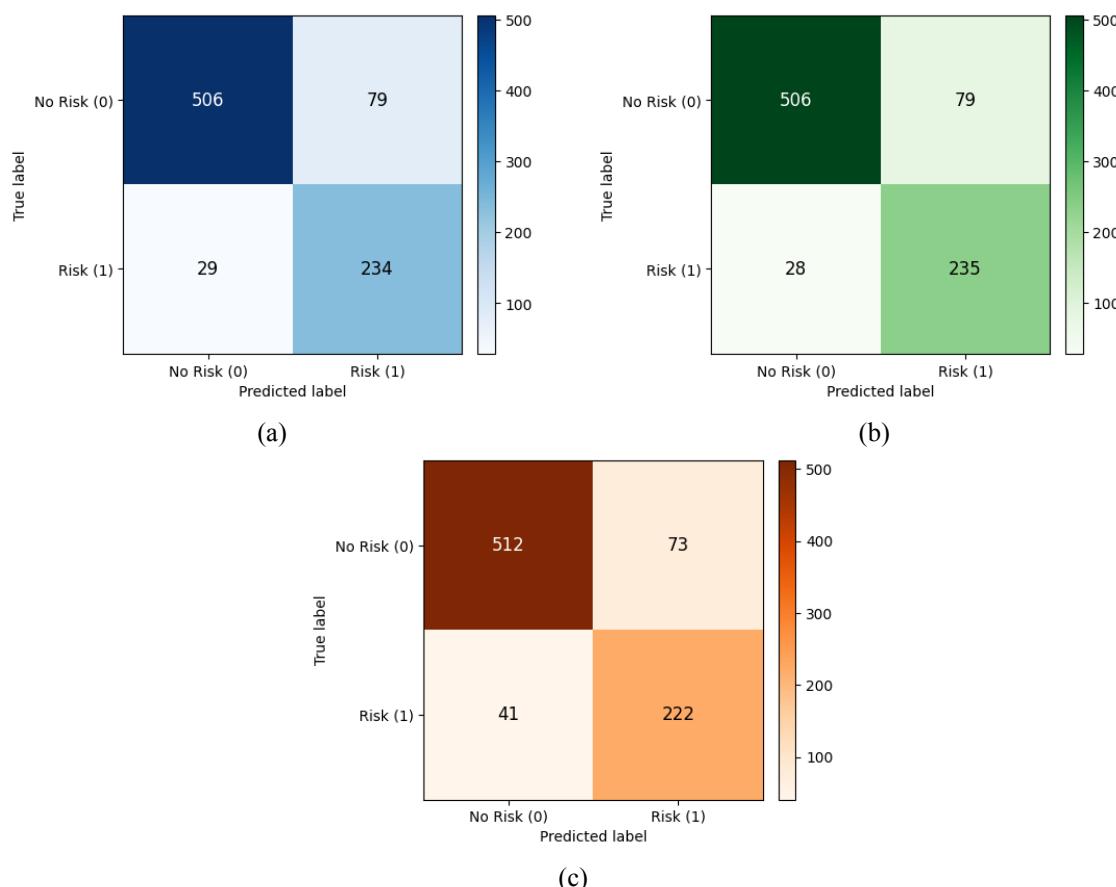
Gambar 4. Korelasi Antar Variabel Numerik

Gambar 4 menunjukkan bahwa terdapat korelasi positif kuat antara *sysBP* dan *diaBP* ($r = 0,78$) serta antara *sysBP* dan variabel target *Risk* ($r = 0,70$). Korelasi yang relatif tinggi juga ditemukan antara *diaBP* dan *Risk* ($r = 0,62$), mengindikasikan bahwa tekanan darah sistolik dan diastolik merupakan indikator paling signifikan

terhadap risiko hipertensi. Sebaliknya, korelasi antar variabel perilaku seperti *currentSmoker* dan *cigsPerDay* juga tinggi ($r = 0,76$), namun kontribusinya terhadap risiko hipertensi relatif lebih rendah dibandingkan parameter tekanan darah.

Proses berikutnya adalah pembangunan model klasifikasi risiko hipertensi menggunakan algoritma *Support Vector Machine* (SVM) dengan tiga jenis kernel, yaitu *Linear*, *Radial Basis Function* (RBF), dan *Polynomial*. Kernel *linear* digunakan untuk data dengan hubungan antar fitur yang cenderung linier, kernel RBF digunakan untuk menangani pola nonlinier yang kompleks, sedangkan kernel *polynomial* digunakan untuk memodelkan interaksi antar fitur dengan tingkat kedalaman tertentu. Dataset yang telah melalui tahap praproses, meliputi imputasi nilai hilang, normalisasi, dan penyeimbangan kelas menggunakan Synthetic Minority Oversampling Technique (SMOTE), dibagi menjadi 80% data latih dan 20% data uji melalui stratified split agar distribusi kelas tetap proporsional. Setiap model SVM dilatih dengan konfigurasi parameter yang sama, yaitu $C = 1,0$, $gamma = scale$, dan $degree = 3$, guna menjaga objektivitas perbandingan. Parameter C mengatur keseimbangan antara margin dan kesalahan klasifikasi untuk mencegah *overfitting*, sedangkan $gamma = scale$ memungkinkan kernel nonlinier beradaptasi terhadap variasi data secara stabil. Nilai $degree = 3$ pada kernel *polynomial* dipilih untuk memberikan kompleksitas yang moderat agar mampu menangkap interaksi antar fitur tanpa meningkatkan beban komputasi secara signifikan.

Setelah proses pelatihan selesai, model dievaluasi menggunakan confusion matrix untuk menggambarkan kemampuan dalam membedakan antara kelas No Risk (0) dan Risk (1) melalui empat komponen utama, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Hasil *confusion matrix* untuk ketiga kernel ditunjukkan pada Gambar 5.

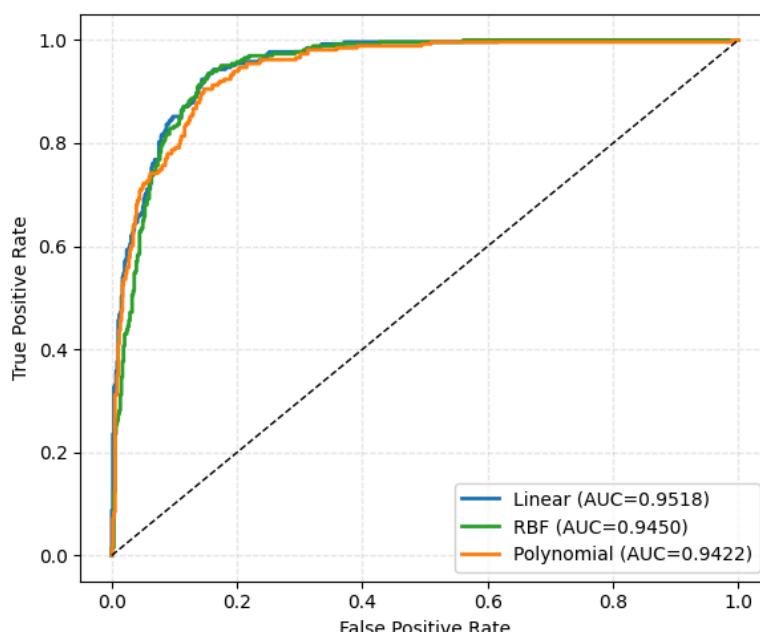


Gambar 5. Confusion Matrix: (a) SVM-*Linear*, (b) SVM-RBF, dan (c) SVM-*Polynomial*

Berdasarkan Gambar 5, terlihat bahwa ketiga model memiliki pola klasifikasi yang serupa dengan tingkat kesalahan yang relatif rendah. Model SVM-*Linear* (Gambar 5a) mampu mengklasifikasikan 506 data *No Risk* dan 234 data *Risk* secara benar, dengan jumlah *False Positive* dan *False Negative* masing-masing 79 dan 29. Model SVM-RBF (Gambar 5b) menunjukkan hasil yang hampir identik, hanya berbeda satu unit pada *False*

Negative. Sementara itu, SVM-Polynomial (Gambar 5c) sedikit menurun kinerjanya dengan jumlah *False Negative* yang lebih tinggi, yaitu 41 kasus salah klasifikasi pada kelas *Risk*.

Untuk mengevaluasi performa generalisasi model secara menyeluruh, digunakan analisis kurva ROC (*Receiver Operating Characteristic*). Kurva ROC menggambarkan perbandingan antara *True Positive Rate* dan *False Positive Rate* pada berbagai nilai *threshold* untuk setiap jenis kernel. Visualisasi kurva ROC tersebut disajikan pada Gambar 6.



Gambar 6. ROC Curve Untuk Tiga Kernel SVM

Gambar 6 memperlihatkan bahwa seluruh kernel menghasilkan area kurva yang luas dengan nilai AUC di atas 0,94, hal ini menunjukkan bahwa model mampu membedakan kategori *Risk* dan *No Risk* dengan tingkat pemisahan yang sangat baik. Kurva SVM-*Linear* tampak berada sedikit di atas dua kernel lainnya di sebagian besar titik, memperkuat hasil bahwa model *linear* lebih optimal untuk pola hubungan fitur yang cenderung proporsional dan tidak terlalu kompleks.

Performa masing-masing model kemudian dinilai secara kuantitatif menggunakan metrik *precision*, *recall*, *F1-score*, *accuracy*, dan ROC-AUC. Hasil lengkap evaluasi ditunjukkan pada Tabel 1.

Tabel 1. Hasil Evaluasi Kinerja Model

Model	Kelas	Precision	Recall	F1-score	Accuracy	ROC-AUC
SVM-Linear	No Risk	0,9458	0,8650	0,9036	0,8726	0,9518
	Risk	0,7476	0,8897	0,8125		
SVM-RBF	No Risk	0,9476	0,8650	0,9044	0,8738	0,9450
	Risk	0,7484	0,8935	0,8146		
SVM-Polynomial	No Risk	0,9259	0,8752	0,8998	0,8656	0,9422
	Risk	0,7525	0,8441	0,7957		

Tabel 1 memperlihatkan kinerja tiga model *Support Vector Machine* (SVM) dengan kernel *Linear*, *Radial Basis Function* (RBF), dan *Polynomial* dalam klasifikasi risiko hipertensi. Model SVM-*Linear* memberikan kinerja paling seimbang dan stabil, dengan *accuracy* sebesar 0,8726 atau 87,26%, *F1-score* rata-rata tinggi, serta ROC-AUC tertinggi yaitu 0,9518. Nilai *precision* untuk kelas *No Risk* mencapai 0,9458, sedangkan *recall* untuk kelas *Risk* sebesar 0,8897, menunjukkan keseimbangan optimal antara deteksi pasien berisiko dan penghindaran kesalahan klasifikasi. Kinerja kuat kernel *linear* ini disebabkan oleh hubungan antar fitur klinis seperti *age*, *BMI*, *sysBP*, dan *totChol* yang cenderung linier, sehingga batas pemisah linier mampu bekerja efektif tanpa memerlukan transformasi nonlinier. Selain itu, kernel *linear* memiliki kompleksitas komputasi rendah dan

memberikan hasil yang efisien untuk data berskala besar atau aplikasi klinis yang memerlukan interpretabilitas tinggi.

Model SVM-RBF menghasilkan *accuracy* tertinggi sebesar 0,8738 atau 87,38%, sedikit lebih baik dari kernel *linear* (selisih 0,12%), menunjukkan kemampuannya menangkap hubungan nonlinier antar fitur. Kernel ini membentuk batas keputusan yang melengkung untuk menyesuaikan variasi data yang tidak sepenuhnya linier, terutama pada fitur seperti tekanan darah (*sysBP*, *diaBP*) dan kolesterol (*totChol*). Namun, peningkatan akurasi ini tidak diikuti oleh kenaikan berarti pada *F1-score* (0,8595) maupun ROC-AUC (0,9450). Hal tersebut menunjukkan bahwa fleksibilitas RBF tidak memberikan keuntungan signifikan pada data yang relatif linier dan telah dinormalisasi dengan baik.

Model SVM-*Polynomial* memiliki kinerja sedikit lebih rendah dengan *accuracy* 0,8656 atau 86,56%, *F1-score* 0,8477, dan ROC-AUC 0,9422. Kompleksitas fungsi polinomial derajat tiga tidak sepenuhnya sesuai dengan struktur data yang sederhana, sehingga cenderung menimbulkan *overfitting* ringan pada data latih. Akibatnya, kemampuan generalisasi terhadap data uji menurun. Selain itu, kebutuhan komputasi yang lebih tinggi membuat kernel ini kurang efisien dibandingkan kernel *linear* maupun RBF.

Secara keseluruhan, SVM-*Linear* terbukti paling optimal berdasarkan kombinasi nilai *F1-score* dan ROC-AUC tertinggi, dengan keseimbangan yang baik antara akurasi, sensitivitas, dan efisiensi komputasi. Kernel RBF unggul dalam fleksibilitas pemetaan nonlinier namun tidak memberikan peningkatan signifikan, sedangkan *Polynomial* menunjukkan performa terendah karena kompleksitasnya tidak sesuai dengan pola data yang cenderung linier.

4. KESIMPULAN

Penelitian ini berhasil mengevaluasi kinerja algoritma *Support Vector Machine* (SVM) dengan tiga jenis kernel, yaitu *Linear*, *Radial Basis Function* (RBF), dan *Polynomial*, untuk klasifikasi risiko hipertensi menggunakan dataset klinis yang telah melalui proses imputasi, normalisasi, serta penyeimbangan kelas dengan SMOTE. Hasil eksperimen menunjukkan bahwa SVM-*Linear* mencapai *accuracy* sebesar 87,26% dan ROC-AUC 0,9518, SVM-RBF memperoleh *accuracy* tertinggi 87,38% dengan ROC-AUC 0,9450, sedangkan SVM-*Polynomial* menghasilkan *accuracy* 86,56% dan ROC-AUC 0,9422. Model SVM-*Linear* memberikan hasil paling optimal berdasarkan kombinasi nilai *F1-score* dan ROC-AUC tertinggi, menunjukkan keseimbangan terbaik antara ketepatan dan sensitivitas. Kernel ini mampu merepresentasikan hubungan linier antar fitur klinis seperti *age*, *BMI*, *sysBP*, dan *totChol* secara efisien dengan kompleksitas komputasi yang rendah. Model SVM-RBF menunjukkan kemampuan adaptif dalam menangani pola nonlinier, meskipun peningkatan akurasi yang diperoleh tidak signifikan. Sementara itu, SVM-*Polynomial* memiliki performa sedikit lebih rendah karena kompleksitasnya tidak sebanding dengan struktur data yang relatif sederhana. Secara keseluruhan, kernel *linear* dinilai paling efektif dan efisien untuk prediksi risiko hipertensi pada dataset ini. Hasil ini mengindikasikan bahwa hubungan antar fitur klinis terhadap risiko hipertensi bersifat hampir linier, sehingga model dengan kompleksitas rendah mampu menghasilkan prediksi yang akurat. Untuk penelitian selanjutnya, disarankan menambahkan fitur klinis dan perilaku yang lebih beragam, menerapkan optimasi hiperparameter berbasis *Grid Search* atau *Bayesian Optimization*, serta membandingkan kinerja SVM dengan model pembelajaran mendalam untuk meningkatkan akurasi dan kemampuan generalisasi.

DAFTAR PUSTAKA

- [1] F. Husaini and T. R. Fonna, "Hipertensi dan Komplikasi Yang Menyertai Hipertensi," *J. Med. Nusant.*, vol. 2, no. 3, pp. 1–13, 2024.
- [2] P. A. Jamaica, A. Y. Rindarwati, D. P. Destiani, and L. Salsabila, "Skrining dan Edukasi Sebagai Upaya Pengendalian Hipertensi Putri," *BATIK J. Pengemb. dan Pengabdi. Masy. Multikultural*, vol. 2, no. 3, pp. 104–110, 2024.
- [3] N. Fatimah and H. Hermawati, "Gambaran Upaya Pengendalian Pada Penderita Hipertensi di Puskesmas Tasikmadu," *IJOH Indones. J. Public Heal.*, vol. 2, no. 1, pp. 121–132, 2024.
- [4] M. Nur, R. Pahlevi, and T. Badriyah, "Implementasi dan Optimasi Hyperparameter pada Model Machine learning untuk Prediksi Diabetes dengan Integrasi Aplikasi Telemedicine," *JEPIN (Jurnal Edukasi dan Penelit. Inform.)*, vol. 11, no. 2, pp. 287–296, 2025.
- [5] M. F. R. Mahendra and N. L. A. dan Sumarno, "Implementasi Machine Learning Untuk Memprediksi Cuaca Menggunakan Support Vector Machine," *J. Ilm. KOMPUTASI*, vol. 23, no. 1, pp. 45–50, 2024.
- [6] T. F. Ramadhan, A. Asrianda, and R. Risawandi, "Penerapan Metode Algoritma SVM (Support Vector Machine) Untuk Klasifikasi Penderita Penyakit Gastroesophageal Reflux Disease," *RABIT J. Teknol. dan Sist. Inf. Univrab*, vol. 10, no. 2, pp. 1212–1219, 2025.
- [7] H. S. Wafa, A. I. Hadiana, and F. R. Umbara, "Prediksi Penyakit Diabetes Menggunakan Algoritma

- Support Vector Machine (SVM)," *Informatics Digit. Expert*, vol. 4, no. 1, pp. 40–45, 2022.
- [8] L. N. Farida and S. Bahri, "Klasifikasi Gagal Jantung Menggunakan Metode SVM (Support Vector Machine)," *Komputika J. Sist. Komput.*, vol. 13, no. 2, pp. 149–156, 2025, doi: 10.34010/komputika.v13i2.11330.
- [9] A. G. Sooai, P. A. Nani, N. M. R. Mamulak, C. O. Sianturi, S. C. Sianturi, and A. H. Mondolang, "Klasifikasi Citra Daun Anggur Menggunakan SVM Kernel Linear," *JOINTECS (Journal Inf. Technol. Comput. Sci.)*, vol. 7, no. 1, pp. 19–26, 2022, doi: 10.31328/jointecs.v8i1.4496.
- [10] E. R. M. Sholihah, I. G. S. M. Diyasa, and E. Y. Puspaningrum, "Perbandingan Kinerja Kernel Linear dan RBF Support Vector Machine Untuk Analisis Sentimen Ulasan Pengguna KAI Access Pada Google Play Store," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 728–733, 2024.
- [11] I. Putri, F. Fadlyyah, and A. Razi, "Implementasi Metode SVM RBF (Radial Basis Function) Kernel Untuk Klasifikasi Status Gizi Pada Balita," *J. Teknol. Terap. Sains 4.0*, vol. 5, no. 2, pp. 43–56, 2024.
- [12] M. R. Pradana, W. Witanti, and A. Komarudin, "Prediksi Tingkat Keparahan Diabetes Melitus Menggunakan Support Vector Machine (SVM) dengan Kernel Polinomial dan RBF," *J. LOCUS Penelit. Pengabdi.*, vol. 4, no. 8, pp. 7521–7533, 2025.
- [13] A. F. Rochim, K. Widyaningrum, and D. Eridani, "Performance Comparison of Support Vector Machine Kernel Functions in Classifying COVID-19 Sentiment," in *4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, IEEE, 2021, pp. 224–228. doi: 10.1109/ISRITI54043.2021.9702845.
- [14] R. Khan, "Hypertension Risk Model," Kaggle. [Online]. Available: <https://www.kaggle.com/datasets/khan1803115/hypertension-risk-model-main>
- [15] R. I. Borman and M. Wati, "Penerapan Data Mining Dalam Klasifikasi Data Anggota Kopdit Sejahtera Bandarlampung Dengan Algoritma Naïve Bayes," *J. Ilm. Fak. Ilmu Komput.*, vol. 9, no. 1, pp. 25–34, 2020.
- [16] A. Syukron, "Penerapan Metode SMOTE Untuk Mengatasi Ketidakseimbangan Kelas Pada Prediksi Gagal Jantung," *J. Teknol. Inf. dan Terap.*, vol. 10, no. 1, pp. 47–50, 2023, doi: 10.25047/jtit.v10i1.313.
- [17] R. Hoque, M. Billah, A. Debnath, S. M. S. Hossain, and N. Bin Sharif, "Heart Disease Prediction using SVM," *Int. J. Sci. Res. Arch.*, vol. 11, no. 2, pp. 412–420, 2024, doi: 10.30574/ijsra.2024.11.2.0435.
- [18] D. S. Pamungkas and L. Sihombing, "Penggunaan Kernel SVM untuk Klasifikasi Pergerakan Jari Menggunakan Sinyal EMG," *J. Elem.*, vol. 7, no. 2, pp. 1–6, 2021.
- [19] A. Zeputra and F. Utaminingrum, "Perbandingan Akurasi untuk Deteksi Pintu berbasis HOG dengan Klasifikasi SVM menggunakan Kernel Linear, Radial Basis Function dan Polinomial pada Raspberry Pi," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 11, pp. 4746–4757, 2021.
- [20] S. D. Wahyuni and R. H. Kusumodestoni, "Optimalisasi Algoritma Support Vector Machine (SVM) Dalam Klasifikasi Kejadian Data Stunting," *Bull. Inf. Technol.*, vol. 5, no. 2, pp. 56–64, 2024, doi: 10.47065/bit.v5i2.1247.
- [21] Z. Arifin, D. F. Rahman, B. S. Rintyarna, and D. Daryanto, "Penerapan Algoritma Support Vector Machine Berbasis Kernel Radial Basis Function dalam Klasifikasi Sel Kanker," *BIOS J. Teknol. Inf. dan Rekayasa Komput.*, vol. 4, no. 2, pp. 100–106, 2023.
- [22] R. Mukarramah, D. Atmajaya, and L. Budi, "Performance comparison of support vector machine (SVM) with linear kernel and polynomial kernel for multiclass sentiment analysis on twitter," *Ilk. J. Ilm.*, vol. 13, no. 2, pp. 168–174, 2021.
- [23] M. Mardewi and N. Yarkuran, "Klasifikasi Kategori Obat Menggunakan Algoritma Support Vector Machine," *J. Pharm. and Appl. Comput. Sci.*, vol. 1, no. 1, pp. 27–32, 2023.
- [24] D. T. Murdiansyah, "Prediksi Stroke Menggunakan Extreme Gradient Boosting," *JIKO (Jurnal Inform. dan Komputer)*, vol. 8, no. 2, p. 419, 2024, doi: 10.26798/jiko.v8i2.1295.