

KLASIFIKASI TINGKAT KEPOSITIFAN PENGIDAP PENYAKIT JANTUNG DENGAN PENDEKATAN ALGORITMA SVM BERBASIS BOTH KERNEL (LINEAR DAN POLYNOMIAL)

Classification Of Positivity Level Of Heart Diseases With Both Kernel Based Svm Algorithm Approach (Linear And Polynomial)

Prasetyo Wicaksono Aji^{*1}, Muhammad Fatchan², Aswan Supriyadi Sunge³

^{1,2,3} Program Studi Informatika, Fakultas Teknik,
Universitas Pelita Bangsa Cikarang,
Indonesia

e-mail. ^{*1}prasetyotio775@gmail.com, ²fatchan@pelitabangsa.ac.id, ³Aswan@pelitabangsa.ac.id

Abstrak

Penyakit jantung adalah kondisi dengan angka kematian yang signifikan, menyebabkan sekitar 12 juta kematian setiap tahun di seluruh dunia. Fakta ini menunjukkan pentingnya mendeteksi penyakit jantung pada tahap awal. Namun, diagnosa penyakit ini menjadi sebuah tantangan yang berat karena adanya hubungan yang rumit antara karakteristik penyakit jantung. Oleh karena itu, pemahaman mengenai atribut kunci yang digunakan dalam proses pengambilan keputusan atau klasifikasi penyakit jantung memiliki makna yang sangat penting. Dalam penelitian ini, Mesin Vektor Dukungan (SVM) digunakan sebagai alat Penelitian dengan berpusat pada penilaian penyakit jantung berdasarkan kondisi kesehatan pasien. Data medis digunakan sebagai variabel prediksi utama dalam penelitian ini. Hasil prediksi akan menetapkan penilaian 1 jika terdapat indikasi penyakit jantung pada pasien, dan sebaliknya, akan memberikan penilaian 0 jika tidak ada penyakit jantung yang terdeteksi. Proses pelatihan model SVM ini dilakukan dengan menggunakan bahasa pemrograman *Python*.

Kata Kunci : Prediksi Penyakit Jantung, Pembelajaran Mesin, Mesin Vektor Dukungan, SVM.

Informasi Artikel:

Submitted: November 2023, Accepted: November 2023, Published: November 2023
ISSN: 2685-4902 (media online), Website: <http://jurnal.umus.ac.id/index.php/intech>

Abstract

Heart disease, a condition associated with a significant global mortality rate, is responsible for an estimated 12 million deaths annually. This underscores the critical importance of early detection of heart disease. Nevertheless, diagnosing this ailment is a formidable task due to the intricate nature of its characteristics. Consequently, gaining insights into the essential factors utilized in the decision-making and classification of heart disease holds substantial significance. In this study, Support Vector Machines (SVM) are employed as a tool Research centered on the assessment of heart disease based on the patient's health condition. Medical data is used as the main predictive variable in this study. The prediction results will assign a rating of 1 if there is an indication of heart disease in the patient, and conversely, will give a rating of 0 if no heart disease is detected process is executed utilizing the Python programming language.

Keywords: Heart Disease Prediction, Machine Learning, Support Vector Machines, SVM.

PENDAHULUAN

Penyakit jantung merupakan kondisi mengancam nyawa yang perlu ditangani secepatnya karena bisa datang tiba-tiba bagi penderitanya. Faktor risiko penyakit jantung yang diidentifikasi berdasarkan kondisi fisik orang yang terkena harus diketahui sejak dini untuk meminimalkan risiko serangan mendadak atau dapat diperbaiki dengan berbagai cara seperti pola hidup sehat dan solusi alternatif. Olahraga dapat mengatur kesehatan jantung dalam tubuh. [1]

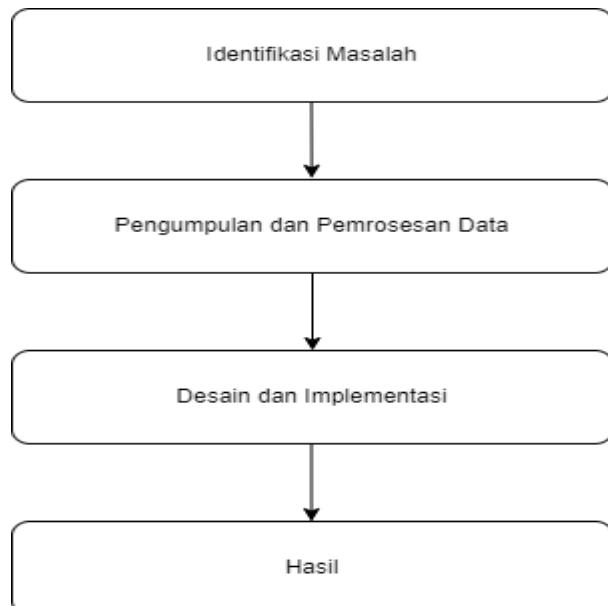
Gagal jantung, hipertensi, dan stroke merupakan penyakit yang berdampak signifikan. Berdasarkan data yang dilaporkan oleh Kementerian Pada tahun 2013 di Indonesia, terdokumentasikan sekitar 12,3 juta kasus kematian akibat penyakit kardiovaskular. Lebih dari 3 juta kematian terjadi sebelum individu mencapai usia 60 tahun, dan sebenarnya, ini bisa dihindari. Kelompok usia antara 20 hingga 40 tahun memiliki risiko penyakit kardiovaskular sekitar 37%, sementara bagi mereka yang berusia antara 40 hingga 60 tahun, risikonya meningkat hingga mencapai 71%. Penyakit ini menjadi penyebab utama kematian pada individu yang berusia di atas 40 tahun. Walaupun aterosklerosis, yang merupakan penyebab utama penyakit ini, dipengaruhi oleh berbagai faktor yang kompleks, termasuk faktor genetik/riwayat keluarga, penyakit jantung koroner, stroke, penyakit kardiovaskular, tekanan darah tinggi, usia, jenis kelamin (terutama pada pria), dan kebiasaan merokok masalah kadar lemak darah (*dislipidemia*), obesitas, diabetes, tingkat aktivitas fisik, dan masa *menopause*[2].

Penyakit jantung memiliki beragam jenis, termasuk penyakit *arteri koroner*, *aterosklerosis*, *aritmia*, *infark miokard* akut (kerusakan dan kematian *miokard*), gagal jantung, *kardiomiopati* (*infark miokard*), gagal jantung lemah dan *hipertrofik*, gangguan katup jantung, penyakit rematik, penyakit jantung bawaan, penyakit *serebrovaskular*, dan *fibrilasi atrium*. Faktor-faktor risiko yang dapat memicu penyakit jantung bisa ditekan dengan mengadopsi gaya hidup sehat, seperti menjalani aktivitas fisik secara teratur, menghindari merokok, serta mengurangi asupan makanan tinggi lemak dan menjaga pola makan yang seimbang dengan asupan serat yang cukup. Selain itu, penyakit jantung juga dipengaruhi oleh faktor risiko yang

tidak dapat diubah, seperti predisposisi genetik, usia, atau adanya penyakit jantung bawaan dalam riwayat keluarga. [3] Beberapa faktor risiko penyakit jantung diketahui, termasuk tekanan darah tinggi, *dislipidemia*, *hiperglikemia*, obesitas, merokok, tidak aktif, pola makan yang buruk, konsumsi alkohol berlebihan, dan faktor genetik kesehatan yang lemah. Namun, data negara untuk seluruh wilayah Indonesia belum tersedia. [3] Dalam konteks penambangan data, terdapat beragam fungsi yang berbeda, dan salah satunya adalah fungsi peringkat. Salah satu teknik yang dapat digunakan untuk melakukan klasifikasi adalah melalui menerapkan algoritma Mesin Vektor Dukungan (SVM). SVM merupakan pendekatan yang digunakan untuk memprediksi dan mengelompokkan data dengan menggunakan SVM. Telah dilakukan berbagai penelitian untuk mengklasifikasikan data pegawai dengan memanfaatkan algoritma SVM.

Kesehatan adalah keadaan yang diinginkan oleh setiap individu. Namun, kesejahteraan bukan sekadar memiliki kondisi fisik yang baik, melainkan juga mencakup kesejahteraan mental dan sosial, termasuk dalam konteks keluarga, yang merupakan komponen integral dari penilaian rumah sehat dan memiliki signifikansi yang setara dengan faktor-faktor lainnya. [4] Ini termasuk tidak hanya tidak adanya penyakit tetapi juga kesehatan, memahami kondisi mental seseorang. Tanpa kesehatan, orang akan mengalami penyakit fisik. Kesehatan juga merupakan keadaan dimana manusia mengalami keseimbangan yang unik, dipengaruhi oleh genetika, teknologi dan gaya hidup sehari-hari seperti makan, minum, bekerja, istirahat untuk menghadapi kehidupan secara mendalam. Kesehatan memiliki faktor penting seperti menjaga pola makan yang sehat, memperbanyak asupan air putih setiap hari, tidur yang cukup minimal 8 jam sehari.

METODOLOGI PENELITIAN



Gambar 2. 1 Flowmap SVM

- Identifikasi Masalah** Tujuan utama dari penelitian ini adalah untuk mengantisipasi apakah pasien mengalami penyakit jantung dengan mempertimbangkan jenis kelaminnya.
- Pengumpulan dan pemrosesan awal data.** Langkah Dalam langkah ini, dilakukan pengenalan sumber data sebagai data sekunder yang diperoleh dari dataset di kaggle.com

yang memuat 13 atribut (dari sejumlah besar atribut yang tersedia). Dataset ini berfokus pada penyakit jantung dan sudah tersedia dalam format saat ini (dataset penyakit jantung), meskipun dengan sedikit perbedaan format. Data ini telah banyak digunakan sebagai referensi dalam penelitian sebelumnya. Penggunaan data ini melibatkan proses Pemisahan data dilakukan menjadi dua bagian, yaitu dataset pelatihan yang digunakan untuk melatih model SVM dan atribut yang digunakan dalam proses ini, dan dataset pengujian, berperan sebagai prediktor dalam analisis ini. Hasil yang diharapkan adalah kelas target, dengan nilai 1 yang menunjukkan bahwa pasien menderita penyakit jantung, sementara nilai 0 mengindikasikan bahwa pasien tidak menderita penyakit jantung. Proses pelatihan model dilakukan dengan menggunakan bahasa pemrograman Python, dengan data X_training sekitar 80% dan X_test sekitar 20%. Selain itu, y_train sekitar 80% dan y_test sekitar 20% digunakan untuk pembagian data.

- c) **Desain dan pembuatan** Desain data, pelatihan dan pengujian dilakukan dengan bahasa pemrograman python dengan dataset penyakit jantung.
- d) **Hasil** Hasilnya adalah koefisien tertimbang untuk setiap atribut yang diprediksi.

Tabel 2. 1 Dataset Penyakit Kardiovaskular

NO	ATRIBUT	TIPE DATA
1	<i>Age</i>	<i>Integer</i>
2	<i>Sex</i>	<i>Integer</i>
3	<i>chest pain type</i>	<i>Integer</i>
4	<i>resting blood pressure</i>	<i>integer</i>
5	<i>Serum cholestorol in mg/dl</i>	<i>integer</i>
6	<i>Fasting blood sugar</i>	<i>integer</i>
7	<i>Resting electrocardiographic result (values 0,1,2)</i>	<i>integer</i>
8	<i>Maximum heart rate achieved</i>	<i>integer</i>
9	<i>Exercise induced angina</i>	<i>integer</i>
10	<i>Oldpeak = ST depression induced by exercise relative to rest</i>	<i>float</i>
11	<i>The slope of the peak exercise ST segment</i>	<i>integer</i>
12	<i>Number of major vessels (0-3) colored by flourosopy</i>	<i>integer</i>
13	<i>Thal: 3 = normal: 6 = fixed defect 7 = reversable defect</i>	<i>integer</i>
y	<i>Target : 1 = Penyakit jantung 0 = bukan penyakit jantung</i>	<i>integer</i>

2.2.1. Rumus Linear

Data yang telah diolah perlu dibagi Dipisah menjadi dua bagian, yakni data pelatihan yang berperan dalam melatih model SVM, sementara data pengujian digunakan untuk menguji sejauh mana model yang telah dilatih mampu membuat prediksi yang akurat menggunakan data yang belum pernah dilihat sebelumnya. Selama proses ini, rumus kernel linier juga diterapkan.

$$K(x, y) = x \cdot y$$

Di sini:

$K(x, y)$ adalah nilai yang dikalikan antara Kedua vektor, x dan y , merupakan vektor input

"." adalah operasi dot product (produk dot).

Perkalian linier bekerja dengan melakukan perkalian titik antara dua vektor masukan x dan y .

Jika $x = [x_1, x_2, \dots, x_n]$ dan $y = [y_1, y_2, \dots, y_n]$, maka hasil perkalian liniernya adalah:

$$K(x, y) = x_1 * y_1 + x_2 * y_2 + \dots + x_n * y_n$$

2.2.2. Rumus Polynomial

Kernel polinomial merupakan bentuk khusus dari fungsi kernel yang umumnya diterapkan dalam algoritme Mesin Vektor Dukungan (SVM) untuk mengubah cara data direpresentasikan menjadi bentuk yang tidak bersifat linier. Penggunaan fungsi kernel polinomial memungkinkan SVM untuk menggambarkan data dalam ruang dimensi yang lebih tinggi dengan menerapkan fungsi polinomial. Di bawah ini, ada rumus yang secara tradisional digunakan untuk mengilustrasikan ide perkalian polinomial dalam konteks ini.

$$K(x, y) = (x \cdot y + c)^d$$

Di sini:

$K(x, y)$ adalah nilai kernel antara dua vektor masukan x dan y .

x dan y adalah vektor masukan.

"." adalah operasi perkalian titik.

c adalah konstanta (deviasi).

d adalah eksponen polinomial yang digunakan dalam pemetaan.

Hasil perkalian polinomial adalah pangkat polinomial perkalian titik antara vektor input x dan y , ditambah konstanta c .

Contoh kernel polinomial dengan pangkat $d = 2$:

$$K(x, y) = (x \cdot y + c)^2$$

Misalnya, jika $x = [x_1, x_2]$ dan $y = [y_1, y_2]$, maka hasil dari kernel polinomial akan menjadi:

$$K(x, y) = (x_1 * y_1 + x_2 * y_2 + c)^2$$

HASIL DAN PEMBAHASAN

Dalam penelitian ini, pengujian dilakukan terhadap Algoritma *Support Vector Machine* (SVM) dengan menerapkan kernel *Linear* dan *Polynomial*. Evaluasi kinerja algoritma klasifikasi dilakukan menggunakan kumpulan data yang dapat diakses secara publik melalui repositori Kaggle.com, yang berfokus pada data mengenai Penyakit Jantung. Dataset yang

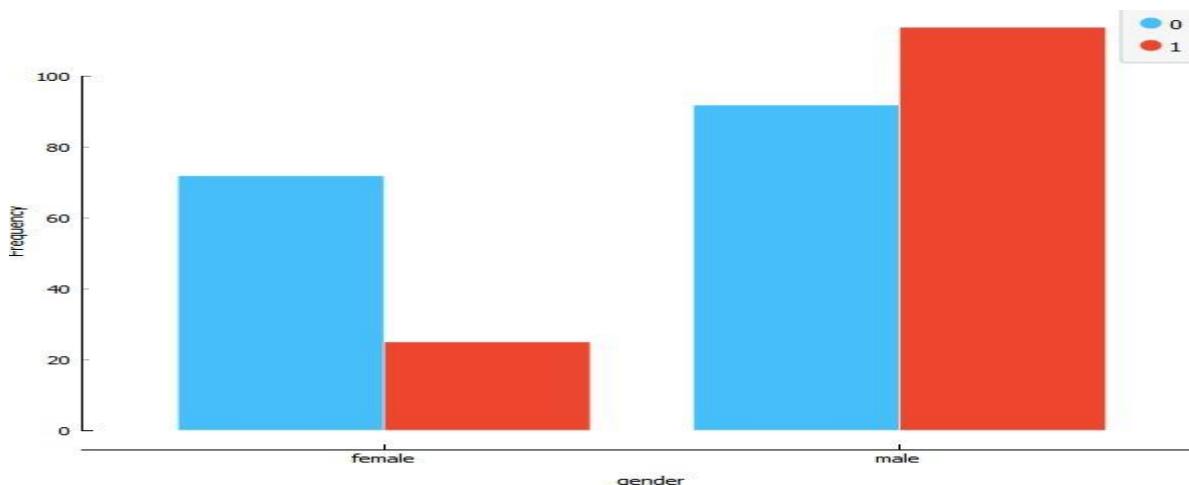
digunakan terdiri dari 14 atribut dan mencakup 303 entri data yang menggabungkan tipe data *nominal* dan *numerik*. Penelitian ini menitikberatkan pada analisis data yang berhubungan dengan pasien yang mengalami penyakit jantung. Selanjutnya, dilakukan tahap validasi untuk mengukur kinerja kedua algoritma dengan menggunakan berbagai metrik, seperti *AUC*, Akurasi (*Accuracy*), *F1 Score*, Akurasi, dan *Callback*. Tujuan utama dari pengujian ini adalah untuk mengidentifikasi algoritma yang dapat memberikan tingkat akurasi tertinggi. Seluruh proses pengujian dilakukan dengan memanfaatkan aplikasi *Orange Python*, yang memiliki kemampuan untuk mengelola dataset yang terdiri dari 14 baris dengan berbagai jenis data, baik *numerik* maupun *nominal*.

```
[ ] df.info
<bound method DataFrame.info of
   age  sex  cp  trestbps  chol  fbs  restecg  thalach  exang  oldpeak
0    63  1  3    145  233  1     0    150     0    2.3
1    37  1  2    130  250  0     1    187     0    3.5
2    41  0  1    130  204  0     0    172     0    1.4
3    56  1  1    120  236  0     1    178     0    0.8
4    57  0  0    120  354  0     1    163     1    0.6
...
298  57  0  0    140  241  0     1    123     1    0.2
299  45  1  3    110  264  0     1    132     0    1.2
300  68  1  0    144  193  1     1    141     0    3.4
301  57  1  0    130  131  0     1    115     1    1.2
302  57  0  1    130  236  0     0    174     0    0.0

  slope  ca  thal  target
0      0  0    1      1
1      0  0    2      1
2      2  0    2      1
3      2  0    2      1
4      2  0    2      1
...
298    1  0    3      0
299    1  0    3      0
300    1  2    3      0
301    1  1    3      0
302    1  1    2      0

[303 rows x 14 columns]>
```

Gambar 2. 2 Dataset Penelitian



Gambar 2. 3 Dataset Gender

Algoritma SVM beroperasi menggunakan serangkaian fungsi matematika yang memiliki definisi yang sangat penting. Kernel berperan dalam mengambil data input dan mengubahnya menjadi ruang multidimensi yang dikenal sebagai kernel. Dalam penelitian ini, dataset dibagi

menjadi dua bagian, yaitu dataset pelatihan dan dataset pengujian, dengan maksud untuk melakukan evaluasi terhadap performa model klasifikasi, metrik yang digunakan berdasarkan pada area di bawah kurva (*AUC*). Algoritma SVM memiliki beragam jenis fungsi kernel, termasuk yang *linier*, *non-linier*, *polinomial*, fungsi basis radial (RBF), dan sigmoid. Dalam penelitian ini, peneliti menerapkan dua jenis kernel, yaitu kernel Linier dan Polynomial.

1) Linear

$$K(x, y) = x \cdot y$$

2) Polynomial

$$K(x, y) = (x \cdot y + c)^d$$

Peneliti mengevaluasi kinerja model dengan menggunakan tiga pengukuran, yakni akurasi klasifikasi, dan gain,. Akurasi mencerminkan perbandingan antara jumlah prediksi yang tepat dengan total prediksi yang dilakukan. Selain itu, dalam laporan juga terdapat nilai sensitivitas atau recall.

Hasil Negative Benar (TN) Merupakan jumlah individu yang diidentifikasi sebagai tidak menderita penyakit jantung, dan memang benar bahwa mereka tidak mengalami penyakit jantung.

Hasil Positif Benar (TP) merujuk pada data pasien yang telah diprediksi oleh model SVM dengan benar bahwa mereka akan mengalami penyakit jantung.

Hasil Negative Salah (FN) Ini adalah data yang sebenarnya mengidap penyakit, tetapi telah keliru diklasifikasikan sebagai yang tidak menderita penyakit.

Hasil Positive Salah (FP) Ini merujuk pada data individu yang sebelumnya dianggap memiliki penyakit, tetapi pada kenyataannya Individu-individu tersebut tidak mengalami penyakit tersebut. Perhitungan tingkat presisi, akurasi, dan spesifikasinya dilakukan berikut.:

$$\text{Akurasi} = (TP+TN) / (TP+TN+FP+FN)$$

$$\text{Pengingat} = TP / (TP+FN)$$

$$\text{Akurasi} = TP / (TP+FP)$$

Tabel 2. 2 Kernel Linear

Model	AUC	CA	F1	Prec	Recall	MCC
SVM	0,809	0,743	0,743	0,745	0,743	0,486

Tabel 2. 3 Kernel Polynomial

Model	AUC	CA	F1	Prec	Recall	MCC
SVM	0,864	0,809	0,809	0,809	0,809	0,616

KESIMPULAN

Melalui studi pada dataset klasifikasi penyakit jantung dari Kaggle.com menggunakan SVM, ditemukan dua pendekatan yang berbeda, yaitu metode Linear dan Polynomial. Hasilnya adalah sebagai berikut: ketika data pelatihan dan data pengujian verifikasi Dalam pengaturan pembagian data 90:10, metode Polynomial memberikan tingkat keakuratan sekitar 80%, sedangkan metode Linear memberikan tingkat akurasi sekitar 80% juga. Namun, ketika data dibagi dengan rasio 80:20, metode Polynomial menghasilkan kinerja terbaik dengan tingkat akurasi sekitar 86%, sementara metode Linear memberikan tingkat akurasi sekitar 80%. SVM memiliki kapabilitas untuk mengelompokkan individu berdasarkan status penyakit jantung mereka, baik itu positif atau negatif. Prakondisi Kernel SVM, yang mencakup kernel linier dan kernel Polynomial, memungkinkan SVM untuk menangani masalah klasifikasi yang bersifat nonlinier dan dataset yang lebih kompleks.

DAFTAR PUSTAKA

- [1] Syamsiah and A. Darwaman, “Analisa Particle Swarm Optimization Terhadap Kepuasan,” *Semin. Nas. Ris. dan Teknol. (SEMNAS RISTEK)*, pp. 143–148, 2020.
- [2] H. Azis, P. Purnawansyah, F. Fattah, and I. P. Putri, “Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung,” *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 81–86, 2020, doi: 10.33096/ilkom.v12i2.507.81-86.
- [3] A. Arifullah, M. Risnawaty, and A. Riza, “Saat ini banyak makanan yang siap saji dijual di mana-mana seperti di cafe , restoran dan warung- warung makan pinggir jalan . Makanan tidak sehat tersebut memiliki kandungan garam , gula dan lemak berlebihan yang tidak baik untuk kesehatan . Dengan adany,” no. 80, pp. 1–10, 1945.
- [4] A. Achmad, A. Adnan, and M. Rijal, “Klasifikasi Penyakit Pernapasan berbasis Visualisasi Suara menggunakan Metode Support Vector Machine,” *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 6, no. 1, pp. 78–83, 2022, doi: 10.47970/siskom-kb.v6i1.330.
- [5] D. Rusdiaman and D. Rosiyadi, “Analisa Sentimen Terhadap Tokoh Publik Menggunakan Metode Naïve Bayes Classifier Dan Support Vector Machine,” *J. Comput. Eng. Syst. Sci.*, vol. 4, no. 2, pp. 2502–7131, 2019.
- [6] W. S. Dharmawan, “I N F O R M a T I K a Dalam Prediksi Penyakit Jantung,” *J. Inform. Manaj. dan Komput.*, vol. 13, no. 2, pp. 31–41, 2021.
- [7] R. Annisa, “Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung,” *J. Tek. Inform. Kaputama*, vol. 3, no. 1, pp. 22–28, 2019.
- [8] H. Amalia, “Perbandingan Metode Data Mining Svm Dan Nn Untuk Klasifikasi Penyakit Ginjal Kronis,” *Maret*, vol. 14, no. 1, p. 1, 2018.
- [9] A. Damayunita, R. S. Fuadi, and C. Juliane, “Comparative Analysis of Naive Bayes, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) Algorithms for Classification of Heart Disease Patients,” *J. Online Inform.*, vol. 7, no. 2, pp. 219–225, 2022, doi: 10.15575/join.v7i2.919.
- [10] F. Handayani, “Komparasi Support Vector Machine, Logistic Regression Dan Artificial Neural Network Dalam Prediksi Penyakit Jantung,” *J. Edukasi dan*

Penelit. Inform., vol. 7, no. 3, p. 329, 2021, doi: 10.26418/jp.v7i3.48053.