

(Jurnal Edukasi dan Penelitian Informatika)

ISSN(e): 2548-9364 / ISSN(p): 2460-0741

Vol. 7 No. 3 Desember 2021

# Komparasi Support Vector Machine, Logistic Regression Dan Artificial Neural Network dalam Prediksi Penyakit Jantung

Fitri Handayani<sup>#1</sup>, Kartika Sari Kusuma<sup>#2</sup>, Hedy Leoni Asbudi<sup>#3</sup>, Rona Guines Purnasiwi<sup>#4</sup>, Resti Kusuma<sup>#5</sup>, Andi Sunyoto<sup>#6</sup>, Windha Mega Pradnya<sup>#7</sup>,

#Magister Teknik Informatika, Universitas AMIKOM Yogyakarta
II. Ring Road Utara, Condong Catur, Sleman, Yogyakarta

¹fitrihandayani@students.amikom.ac.id
²kartikasarikusuma@students.amikom.ac.id
³hedyleoni@students.amikom.ac.id
⁴ronaguines@students.amikom.ac.id
⁵restikusuma@students.amikom.ac.id
²andi@amikom.ac.id
²windha.m@amikom.ac.id

Abstrak— Penyakit jantung adalah salah satu penyakit yang menyebabkan resiko kematian cukup tinggi di dunia. Kolesterol, diabetes, tekanan darah tinggi merupakan faktor-faktor pemicu terjadinya penyakit jantung. Perlu deteksi sejak ini mengenai prediksi penyakit jantung pada setiap individu agar pencegahan dan pengobatan dapat segera dilakukan demi tingkat kesehatan yang lebih baik. Berbagai metode dapat dilakukan untuk melakukan deteksi penyakit jantung, baik dengan metode tradisional dan metode yang memanfaatkan teknologi. Saat ini mulai banyak bermunculan system pendeteksi penyakit jantung dengan memanfaatkan algoritma machine learning. Algoritma machine learning dianggap mudah untuk diaplikasikan untuk mengklasifikasikan apakah seseorang terkena penyakit jantung. Penelitian ini mencoba membandingkan hasil klasifikasi penyakit jantung menggunakan dataset public dari UCI menggunakan tiga algorima machine learning, yaitu Support Vector Machine (SVM), Logistic Regression (LR) dan Artificial Neural (ANN). Ketiga algorima tersebut Network menggunakan empat skenario pembagian data training dan testing yang berbeda, yaitu 90:10, 80:20, 70:40 dan 60:40 untuk mengetahui hasil terbaik dari tiga algoritma tersebut. Dari hasil eksperimen didapatkan hasil akurasi tertinggi pada metode Logistic Regression sebesar 86% menggunakan skenario pembagian data 80:20.

Kata kunci— Artificial Neural Network, Penyakit jantung, Logistic Regression, Support Vector Machine

# I. PENDAHULUAN

Kematian yang terjadi diseluruh dunia saat ini didominasi oleh penyakit jantung [1]. Menurut WHO, kematian global yang disebabkan penyakit jantung memiliki persentase 31%.[2]. Beberapa faktor resiko penyakit jantung yaitu diabetes, kolesterol, tekanan darah tinggi dan lainnya merupakan faktor yang sulit untuk diidentifikasi [3].

Adapun diagnosa penyakit jantung yang akurat dan tepat waktu penting dilakukan untuk pencegahan dan pengobatan gagal jantung [4]. Analisis data dalam jumlah besar memberikan peluang besar dan mendapatkan hasil terbaik[1]. Analisis data tersebut dapat dilakukan dalam prediksi penyakit jantung menggunakan teknik pembelajaran mesin yang sangat efisien bagi tenaga medis[2]. Oleh karena itu, penggunaan machine learning (ML) menjadi solusi untuk mengurangi dan memahami gejala yang berhubungan dengan penyakit jantung dan membantu praktisi membuat keputusan yang lebih akurat mengenai kesehatan pasien [5].

Beberapa penelitian terdahulu telah melakukan penelitian tentang prediksi penyakit jantung dengan dataset publik yang bersumber dari UCI Machine Learning dengan berbagai metode prediksi. Penelitian [6] menggunakan metode Logistic Regression, Random Forest, Stochastic Gradient Boosting dan Support Vector Machine dengan pembagian data 70%:30%. Berdasarkan hasil penelitian tersebut performa terbaik diperoleh ketika

menggunakan metode Logistic Regression dengan ACU sebesar 91% dan akurasi 86%.

Penelitian lain oleh [4] menggunakan metode Logistic regression, k-nearest neighbor, ANN, SVM, decision tree, Naive Bayes, and random forest dengan pembagian data 90%:10% dan akurasi terbaik sebesar 86% didapatkan ketika menggunakan metode SVM. Penelitian selanjutnya oleh [1] menggunakan metode Naive Bayes dengan pembagian data 60%:40%. Beberapa penelitian tersebut menunjukkan bahwa penelitian pembelajaran mesin melakukan pembagian data latih dan data uji bervariasi. Penelitian [2] menghasilkan akurasi menggunakan metode SVM sebesar 92.30%, Logistic Regression sebesar 93.40%, dan ANN sebesar 92,30% dapat disimpulkan hasil akurasi dari ketiga metode tersebut tidak terlalu signifikan dengan pengolahan pembagian data latih dan uji tidak bervariasi. Berdasarkan permasalahan tersebut, penelitian selanjutnya akan melakukan eksperimen untuk membandingkan metode SVM, Logistic Regression, dan ANN berdasarkan pembagian data testing dan data training bervariasi bertujuan untuk mendapatkan tingkat akurasi pembagian data yang lebih baik dengan menggunakan tools pengujian yang berbeda dari penelitian sebelumnya. Pemilihan ketiga metode bertujuan untuk mengetahui metode terbaik dari ketiga metode tersebut.

# II. LITERATUR REVIEW

Beberapa penelitian sebelumnya banyak membahas klasifikasi penyakit jantung menggunakan berbagai metode. Pada penelitian [7] melakukan survey penelitian mengenai metode yang banyak dipakai dalam melakukan klasifikasi penyakit jantung. Hasilnya didapatkan metode klasifikasi yang digunakan adalah Support Vector Machines (SVM), K-Nearest Neighbour (KNN), NaïveBayes, Decision Trees (DT) dan Random Forest (RF) dengan nilai akurasi yang cukup bervariasi.

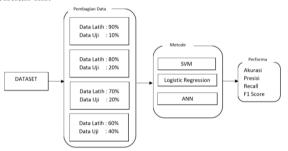
Pada penelitian [8] melakukan perbandingan enam metode yaitu SVM, Random Forest, Logistic Regression, KNN, ANN dan Naïve Bayes menggunakan enam tools yang berbeda yaitu orange, RapidMiner, Weka, MAtlab, Knime dan Sci-Kit learn. Masing-masing metode diuji menggunakan keenam tools tersebut dan dari hasil eksperimen diketahui bahwa metode Naïve Bayes dan Logistic Regression memiliki nilai akurasi yang cukup tinggi dan stabil. Penelitian lain yang menggunakan metode Logistic Regression adalah penelitian [6] dan [4]. Penelitian [6] menggunakan Linear Regression untuk dikombinasikan menggunakan ROC untuk meningkatkan akurasi, sedangkan penelitian [4] mengkombinasikan Linear Regresion dan ANN dengan algoritma pemilihan fitur Relief, mRMR dan Lasso. Selain itu, penelitian sebelumnya [9] membuat sebuah MDSS (Medical Decision Support System) untuk membantu memprediksi jenis penyakit jantung ada menggunakan ANN dan KNN.

Penelitian [10] melakukan identifikasi fitur yang paling signifikan dan teknik data mining yang dapat meningkatkan akurasi prediksi penyakit jantung dengan menggunakan k-NN, Decision Tree, Naive Bayes, Logistic Regression (LR), Support Vector Machine (SVM), Neural Network dan Vote (kombinasi Naïve Bayes dan Logistic Regression).

Terakhir, pada penelitian [2] melakukan studi perbandingan dan analisis algoritma Machine Learning untuk memprediksi penyakit jantung. Algoritma yang dibandingkan antara lain, Logistic Regression, K-nearest Neighbour (KNN), Support Vector Machine (SVM), Naïve Bayes, Decision Trees, Random Forest, Artificial Neural Network (ANN), Multi-Layer Perceptron (MLP), Deep Neural Network (DNN).

#### III. METODE PENELITIAN

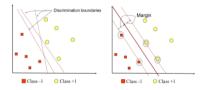
Penelitian ini menggunakan tiga metode pembelajaran mesin dalam prediksi penyakit jantung yang terdiri dari Support Vector Machine (SVM), Logistic Regression dan Artificial Neural Network (ANN). Dataset yang telah diperoleh dibagi sesuai dengan kriteria kemudian dilakukan pengujian menggunakan tiga metode tersebut untuk mendapatkan performa dari masing-masing metode yang digunkan. Gambar 1. Menggambarkan alur pada penelitian ini.



Gambar 1 Alur penelitian

#### A. Support Vector Machine (SVM)

Support Vector Machine yaitu merupakan merode pembelajaran terbimbing dalam melakukan klasifikasi. SVM dapat digunakan pada klasifikasi linier atau non linier [2]. Hyperplane terbaik dapat ditemuka dengan mengukur margin hyperplane dan mencari titik maksimalnya. Pemisahan kelas-kelas oleh SVM pada Gambar 2.



Gambar 2 Pemisahan dua kelas -1 dan +1 dengan SVM (Nugroho, Witarto, & Handoko, 2003)

Ada beberapa pattern yang terkelompok dalam 2 kelas -1 (merah) dan +1 (kuning). Terdapat 2 kasus dalam memisahkan kelas oleh Hyperplane yaitu kelas yang dapat dipisahkan secara sempurna disebut dengan SVM linier dan kelas yang yang tidak dapat dipisahkan dengan sempurna disebut SVM non linier.

Pada dasarnya SVM non linier merupakan solusi dari masalah SVM linier dengan melakukan fungsi kernel pada ruang fitur dimensi tinggi. Definisi persamaan SVM linier dan non linear dapat dilihat Tabel 1.

TABEL I Defenisi Persamaan SVM Linier dan Non Linier

Sifat SVM	Jenis Kernel	Definisi Rumus
SVM	Linier	K(x,y) = x.y
Linier		
SVM Non	Polynomial	$K(x,y) = (x,y + 1)^p$
Linier	Gaussian	$K(x,y) = e^{-\ x-y\ ^2/2\sigma^2}$
	RBF	
	Sigmoid	$(x, y) = \tanh(Kx. y - \delta)$

#### Keterangan:

p : pangkat (degree of)

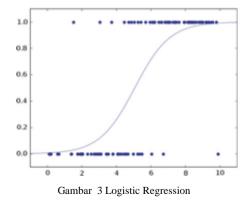
σ : nilai sigma δ : nilai delta

# B. Logistic Regression

pembelajaran Logistic Regression merupakan terbimbing pada masalah regresi dan klasifikasi. Logistic Regression menggunakan probabilitas untuk memprediksi klasifikasi data kategorikal. Nilai input dapat digabungkan secara linier menggunakan fungsi sigmoid atau logistik dan nilai koefisien untuk memprediksi hasil. Estimasi kemungkinan maksimum dianggap menggunakan fungsi sigmoid untuk mengasumsikan data yang paling mungkin, dan probabilitas diberikan antara 0 sampai 1 mengatakan apakah suatu peristiwa akan terjadi atau tidak. Ketika ambang keputusan digunakan, maka itu menjadi masalah klasifikasi. Ini bisa menjadi jenis yang berbeda seperti Biner (0 atau 1), Multinomial (tiga atau lebih klasifikasi tanpa pemesanan), Ordinal (tiga atau lebih klasifikasi dengan pemesanan). Ini adalah model yang sederhana untuk diterapkan dan dapat memberikan prediksi yang baik [2].

$$p = {1 \atop 1} + e - (a + bX)$$
 (1)

dimana dalam Persamaan (1) p adalah probabilitas dan a, b akan menjadi parameter model, X adalah faktor.



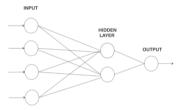
Algoritma Logistic Regression adalah metode regresi dan klasifikasi untuk mengevaluasi dataset yang berisi

satu atau lebih variabel independen yang menyimpulkan suatu hasil. Hasil diukur dengan variabel yang dibagi (dimana dapat menjadi dua kemungkinan hasil) [6]. Gambar 3 menunjukkan dua kemungkinan hasil dari Logistic Regression.

## C. Artificial Neural Network (ANN)

Artificial neural yaitu bagian dari pembelajaran mesin yang menyerupai seperti cara kerja otak manusia dalam menangkap informasi dan memberikan respon pada saat yang bersamaan [2]. Kata 'neural' diambil dari unit fungsional dasar neuron atau sel saraf dari sistem saraf manusia yang terdapat di otak dan bagian lain dari tubuh manusia. Target utama dari pendekatan jaringan saraf tiruan adalah untuk menyelesaikan masalah sebagai otak manusia lakukan [11].

ANN memiliki kemampuan memperlajari dan memodelkan hubugan non-linier dan kompleks, yang mana hal ini sangat penting secara *real timei* karena banyak hubungan antara input dan output tidak linier kan kompleks. Setelah input awal dan *relationship* dipelajari, model dapat menyimpulkan *relationship* yang tidak terlihat pada data yang tidak terlihat secara baik untuk dapat dilakukan prediksi [12].



Gambar 4 Struktur Artificial Neural Network

Bagian input bertugas untuk menerima data lalu meneruskannya ke hiden layer. Pada hiden layer terjadi proses identifikasi pola dengan cara melakukan beberapa perhitungan pada data.setelah proses klasifikasi di hidden layer selesai hasil klasifikasi tersebut lalu diteruskan ke output layer [2]. Proses training pada ANN dimulai saat i yang merupakan neuron bersama dengan m input. ketika neuron i berada di hidden layer pertama maka seluruh input seharusnya terkoneksi ke hidden layer berikutnya. jika tidaka maka input dapat terhubung dengan neuron output yang lain. Node y merupakan konsep yang sangat penting di neural network. Node y merupakan neuron output namun dapat juga di representasikan dengan y<sub>i,j</sub> untuk mendefinisikan i<sup>th</sup> yang merupakan input dari neuron j di hidden layer berikutnya. input dari neuron i dihitung menggunakan rumus fungsi aktivasi:

dihitung menggunakan rumus fungsi aktivasi :
$$f(x) = \frac{1}{1+e^{-x}}$$

$$y_i = f_i(Net_i)$$
(3)

 $Net_i$  merupakan keseluruhan dari node input multiplied berdasarkan bobot dari neuron i yang telah dihitung menggunakan Persamaan (4).

$$Net_{i} = \sum_{j=1}^{m} w_{i,j} \times y_{i,j} + b_{i,0}$$
 (4)

Korespondensi: Fitri Handayani 331

Ketika  $w_{i,j}$  merupakan bobot dari input j pada neuron i. Error yang dialami oleh input  $q^{th}$  dapat didambarkan oleh rumus :

$$e_{q} = t_{q-a_{q}}^{m} \tag{5}$$

Setelah menerapkan semua rumus diatas maka presentasi error akan diminimalisir [13]. ANN biasanya digunakan untuk memprediksi masalah-malah yang nonlinear seperti bidang komunikasi, industri elektronik, industri kimia, industri produksi energi dan industri obatobatan [9].

# D. Dataset Penyakit Jantung

Dataset prediksi penyakit jantung yang akan digunakan pada penelitian ini bersumber dari UCI Machine Learning [14]. Dataset dibagi menjadi data latih dan data uji. Kemudian diberikan label yang diprediksi menderita penyakit jantung atau tidak menggunakan metode pembelajaran mesin yang telah ditentukan.

# E. Confussion Matrix

Confusion matrix yaitu merupakan pengujian untuk mengetahui kinerja algoritma pembelajaran mesin[9]. Pengujian ini berdasarkan tabel confusion matrik (lihat Tabel 2) [4][15].

TABEL II CONFUSION MATRIX

	Prediksi		
Aktual	Sakit Jantung	Tidak	Sakit
		Jantung	
Sakit Jantung	TP	FN	
Tidak Sakit Jantung	FP	TN	

Berdasarkan Tabel 2, TP yaitu jumlah data bernilai positif dan hasil prediksi positif. TN yaitu jumlah data bernilai positif dan hasil prediksi negatif. FP yaitu jumlah data bernilai negatif dan hasil prediksi positif. Dan TN yaitu jumlah data bernilai negatif dan hasil prediksi negati [3]. Confusion matrix terdiri dari beberapa perhitungan [16]:

#### 1. Presisi

Presisi merupakan perhitungan untuk mengetahui jumlah data yang benar positif dari semua hasil prediksi benar positif. Presisi dapat dilakukan menggunakan Persamaan (5) [15][3][16].

Persamaan (5) [15][3][16].  
Presisi = 
$$\frac{TP}{TP+FP}$$
 (5)

### 2. Recall

Recall merupakan perhitungan untuk mengetahui jumlah data yang prediksi benar positif dari hasil seluruh benar positif. Recall menggunakan Persamaan (6) [15][3][9].

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

### 3. F1 Score

F1 Score merupakan perhitungan untuk mengetahui rata-rata dari perbandingan presisi maupun recall. F1

score dapat dilakukan perhitungan menggunakan Persamaan (7) [15][3] [16].

$$F1 \ Score = 2 \times \frac{Recall \times Presisi}{Recall + Presisi}$$
 (7)

#### 4. Akurasi

Akurasi merupakan perhitungan untuk mengetahui keakuratan model dalam klasifikasi yang benar. Akurasi dapat dilakukan dengan Persamaan (8) [15][4][3].

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

#### IV. HASIL DAN DISKUSI

Penelitian ini mengguanakan metode prediksi terdiri dari SVM, Logistic Regression dan ANN. Dataset yang digunakan berjumlah 303 dengan jumlah atribut 13 (lihat Tabel 3).

TABEL III
ATRIBUT PREDIKSI PENYAKIT JANTUNG

N o	Atribut	Kategori	Keterangan
1	Umur (Age)	-	menampilkan usia individu.
2	Jenis Kelamin (Sex)	1 = laki-laki 0 = perempuan	menampilkan jenis kelamin individu
3	Jenis nyeri dada (chest pain type/cp)	1 = angina tipikal 2 = angina atipikal 3 = bukan nyeri angina 4 = asimtotik	menampilkan jenis nyeri dada yang dialami oleh individu
4	Tekanan Darah (Resting Blood Presure/trestb ps)	-	menampilkan nilai tekanan darah dalam satuan mmHg.
5	Kolesterol Serum (serum Cholesterol/ch ol)	-	menampilkan kolesterol serum dalam mg/dl (unit)
6	Gula Darah Puasa (fasting blood sugar/fbs)	1 = Jika gula darah puasa > 120mg/dl 0 = lainnya	membandingkan nilai gula darah puasa seseorang dengan 120mg/dl
7	ECG (Resting ECG/restecg)	0 = biasa 1 = memiliki kelainan gelombang ST-T 2 = hipertrofi ventrikel kiri	menampilkan hasil elektrokardiografi
8	Detak jantung maks tercapai (Max heart rate achieved/thala ch)	-	menampilkan detak jantung maksimal yang dicapai oleh seorang individu.
9	Angina yang diinduksi oleh olahraga (Exercise	1 = ya 0 = tidak	-

N	Atribut	Kategori	Keterangan
0			
	induced angina/exang)		
10	ST depression induced by exercise relative to rest (oldpeak)	-	menampilkan nilai yang merupakan bilangan integer atau float
11	Peak exercise ST segment (slope)	1 = menanjak 2 = datar 3 = miring ke bawah	-
12	Jumlah pembuluh besar (0–3) yang diwarnai oleh flourosopy(ca )	-	menampilkan nilai nteger atau float.
13	Thal	3 = biasa 6 = cacat tetap 7 = cacat reversibel	-

Adapun kelas keluaran terdiri dari dua kelas yaitu pasien yang menderita penyakit jantung dan pasien yang tidak menderita penyakit jantung. Pengujian dilakukan berdasarkan pembagian data latih dan data uji dengan metode pengujian confusion matrik berikut:

#### A. Data latih dan uji 90:10

TABEL IV PEMBAGIAN DATA 90%: 10%

Confusion Matrik	Precision	Recall	F1 Score	Accuracy
SVM	0.82	0.82	0.82	0.80
Logistic	0.82	0.82	0.82	0.80
Regression				
ANN	0.82	0.82	0.82	0.80

Berdasarkan Tabel 4, ketiga metode menghasilkan nilai presisi, recall dan f1 score sama yaitu 0.82. sedangkan nilai akurasi sebesar 0.80.

### B. Data latih dan uji 80:20

TABEL V PEMBAGIAN DATA 80%: 20%

Confusion Matrik	Precisi on	Recall	F1 Score	Accuracy
SVM	0.67	0.88	0.76	0.70
Logistic	0.88	0.88	0.88	0.86
Regression				
ANN	0.87	0.84	0.86	0.85

Berdasarkan Tabel 5, Logistic Regression presisi tertinggi 0.88, SVM dan Logistic Regression recall tertinggi 0.88, Logistic Regression f1 score tertinggi 0.88 dan Logistic Regression akurasi tertinggi 0.86.

# C. Data latih dan uji 70:30

TABEL VI PEMBAGIAN DATA 70%: 30%

Confusion	Precision	Recall	F1 Score	Accuracy
Matrik				
SVM	0.67	0.90	0.77	0.70
Logistic	0.82	0.84	0.83	0.81
Regression				
ANN	0.84	0.84	0.84	0.82

Berdasarkan Tabel 6, ANN presisi tertinggi 0.84, SVM recall tertinggi 0.90, ANN f1 score tertinggi 0.84 dan ANN akurasi tertinggi 0.82.

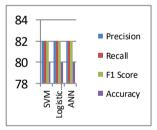
# D. Data latih dan uji 60:40

TABEL VII PEMBAGIAN DATA 60% : 40%

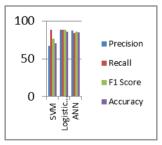
Confusion Matrik	Precision	Recall	F1 Score	Accuracy
SVM	0.70	0.89	0.78	0.72
Logistic	0.86	0.86	0.86	0.83
Regression				
ANN	0.86	0.81	0.84	0.82

Berdasarkan Tabel 7, Logistic Regression dan ANN presisi tertinggi 0.86, SVM recall tertinggi 0.89, Logistic Regression f1 score tertinggi 0.86 dan Logistic Regression akurasi tertinggi 0.83.

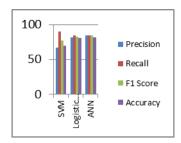
Perbandingan hasil pengujian pembagian data seperti pada gambar berikut.



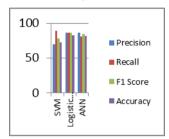
Gambar. 5 Pembagian data 90%: 10%



Gambar. 6 Pembagian data 80%: 20%



Gambar 7 Pembagian data70%: 30%



Gambar 8 Pembagian data 60%: 40%

Hasil pengujian perbandingan jumlah data latih dan data uji menggunakan metode SVM, Logistic Regression dan ANN menunjukkan bahwa pembagian data 90:10 (lihat Gambar 5) ketiga metode menghasilkan akurasi sama, pembagian data 80:20 (lihat Gambar 6) dan 60:40 (lihat Gambar 8) metode Logistic Regression menghasilkan akurasi tertinggi dan pembagian data 70:30 (lihat Gambar 7) menghasilkan akurasi tertinggi metode ANN.

#### V. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan terhadap dataset UCI mengenai klasifikasi penyakit jantung menggunakan tiga metode yang berbeda, yaitu SVM. Logistic Regression dan ANN ditemukan hasil sebagai berikut, saat menggunakan pembagian data latih dan data uji dengan komposisi 90:10 ketika metode tersebut menghasilkan akurasi yang sama yaitu sebesar 80%., saat menggunakan komposisi 80:20 hasil terbaik ketika diperoleh menggunakan metode Logistic Regression sebesar 80%, kemudian ketika data dibagi dengan persentase performa terbaik 70:30 hasil didapatkan saat menggunakan metode ANN yaitu sebesar 82%, dan terakhir saat menggunakan perbandingan 60:40 performa terbaik diperoleh ketika menggunakan metode Logistic Regression dengan akurasi 83%.

#### REFERENSI

- [1] R. Venkatesh, C. Balasubramanian, dan M. Kaliappan, "Development of Big Data Predictive Analytics Model for Disease Prediction using Machine learning Technique," J. Med. Syst., 2019.
- [2] R. Katarya dan S. Kumar, "Machine Learning Techniques for Heart Disease Prediction: A Comparative Study and Analysis," *Health Technol. (Berl).*, no. 123456789, 2020.
- [3] S. Mohan, C. Thirumalai, dan G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," *IEEE Access*, vol. 7, hal. 81542–81554, 2019.
- [4] A. U. Haq, J. P. Li, M. H. Memon, S. Nazir, dan R. Sun, "A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms," *Hindawi*, vol. 2018, 2018.
- [5] A. Hajjam, E. Hassani, E. Andr, dan A. K. G, "Informatics in Medicine Unlocked Classification models for heart disease prediction using feature selection and PCA," vol. 19, 2020.
- [6] R. Kannan dan V. Vasanthi, "Machine Learning Algorithms with ROC Curve for Predicting and Diagnosing the Heart Disease," *Spinger*, 2019.
- [7] V. V Ramalingam, A. Dandapath, dan M. K. Raja, "Heart disease prediction using machine learning techniques: A survey," *Int. J. Eng. Technol.*, no. March 2018, 2019.
- [8] I. Tougui, A. Jilbab, dan J. El Mhamdi, "Heart disease classification using data mining tools and machine learning techniques," *Spinger*, 2020.
- [9] O. Terrada dan O. Bouattane, "Classification and Prediction of atherosclerosis diseases using machine learning algorithms," 2019 5th Int. Conf. Optim. Appl., hal. 1–5, 2019.
- [10] M. S. Amin, Y. K. Chiam, dan K. D. Varathan, "Abstract Cardiovascular disease is one of the biggest cause for morbidity and mortality among the," *Telemat. Informatics*, 2018.
- [11] M. Diwakar, A. Tripathi, K. Joshi, M. Memoria, dan P. Singh, "Materials Today: Proceedings Latest trends on heart disease prediction using machine learning and image fusion," *Mater. Today Proc.*, no. xxxx, 2020.
- [12] R. T. S. I. Muthulakshmi, "An optimal artificial neural network based big data application for heart disease diagnosis and classification model," J. Ambient Intell. Humaniz. Comput., no. 123456789, 2020.
- [13] C. B. C. Latha dan S. C. Jeeva, "IMPROVING THE ACCURACY OF PREDICTION OF HEART DISEASE RISK BASED ON ENSEMBLE CLASSIFICATION TECHNIQUES," *Informatics Med. Unlocked*, 2019.
- [14] A. Janosi, W. Steinbrunn, M. Pfisterer, dan R. Detrano, "UCI Machine Learning Repository," 1988. [Daring]. Tersedia pada: https://archive.ics.uci.edu/ml/datasets/Heart+Disease.
- [15] B. Santoso, "An Analysis of Spam Email Detection Performance Assessment Using Machine Learning," *JOIN (Jurnal Online Inform.*, vol. 4, no. 1, hal. 53–56, 2019.
- [16] Derisma, "Perbandingan Kinerja Algoritma untuk Prediksi Penyakit Jantung dengan Teknik Data Mining," J. Appl. Informatics Comput., vol. 4, no. 1, hal. 84–88, 2020.