



IMPLEMENTASI METODE SVM UNTUK MEMPREDIKSI RISIKO PENYAKIT JANTUNG PADA LANSIA

Fadli Husein Wattiheluw¹⁾

¹⁾Fakultas Teknik Informatika, Institut Teknologi Informasi dan Bisnis Stikom Ambon

Corresponding Author: ¹ fadliwattiheluw1994@gmail.com

Article Info

Article history:

Received: 25/09/2025

Revised: 1/10/2025

Accepted: 10/10/2025

Published: 15/10/2025

Keywords:

Heart disease, elderly, prediction, Support Vector Machine, classification

ABSTRACT

Heart disease is one of the main causes of death in elderly people throughout the world. Early detection of the risk of heart disease is very important to prevent more severe complications. This research aims to develop a heart disease risk prediction system in the elderly using the Support Vector Machine (SVM) method. The SVM method was chosen because of its ability to classify non-linear data effectively and accurately. The dataset used includes elderly patient data with various attributes such as age, blood pressure, cholesterol, heart rate and disease history. After going through the data pre-processing process, the model is trained and tested using the SVM algorithm. The test results show that the model is able to achieve an accuracy level of 93.55% with a precision value of 0.94, recall of 0.94, and f1-score of 0.93 (weighted average). With these data results, this system has the potential to become a tool in the initial screening process for the risk of heart disease in the elderly by medical personnel.

1. INTRODUCTION

Penyakit jantung merupakan salah satu penyebab utama kematian di dunia, terutama pada kelompok usia lanjut [1]. Berdasarkan data dari WHO, lebih dari 17 juta kematian setiap tahun disebabkan oleh penyakit kardiovaskular, dan hampir 75% terjadi di negara berkembang [1]. Di Indonesia sendiri, prevalensi penyakit jantung pada usia di atas 55 tahun cenderung meningkat seiring bertambahnya usia [2].

Seiring dengan meningkatnya jumlah lansia, diperkirakan risiko penyakit jantung juga akan semakin tinggi. Lansia memiliki risiko yang lebih besar karena adanya penurunan fungsi organ, sistem imun yang melemah, serta komorbiditas seperti hipertensi dan diabetes [3]. Oleh karena itu, diperlukan upaya preventif melalui deteksi dini risiko penyakit jantung untuk meminimalisir angka kematian dan biaya perawatan jangka panjang [4].

Permasalahan yang muncul adalah bagaimana cara mendeteksi risiko penyakit jantung pada lansia secara cepat dan akurat. Sistem konvensional dalam mendeteksi penyakit jantung umumnya memerlukan pemeriksaan medis lengkap dan proses diagnosa yang memakan waktu, tenaga, dan biaya yang tidak sedikit [5]. Oleh karena itu, perlu dikembangkan sistem berbasis teknologi yang mampu melakukan prediksi secara cepat dan efisien.

Salah satu solusi yang dapat diterapkan adalah penggunaan algoritma Machine Learning, khususnya metode Support Vector Machine (SVM). SVM merupakan metode klasifikasi yang cukup efektif dalam mengolah data kesehatan dan memiliki tingkat

akurasi tinggi dalam memprediksi penyakit kronis seperti jantung [6], [7]. Metode ini bekerja dengan memisahkan data ke dalam dua kelas menggunakan hyperplane terbaik berdasarkan karakteristik input, seperti tekanan darah, kadar kolesterol, usia, dan sebagainya [7].

Beberapa penelitian sebelumnya telah menguji efektivitas metode SVM dalam prediksi penyakit jantung. Misalnya, penelitian oleh Dey et al. menunjukkan bahwa SVM menghasilkan akurasi sebesar 88% dalam klasifikasi data penyakit jantung, lebih tinggi dibandingkan dengan metode lain seperti Decision Tree dan Naive Bayes [8]. Studi lain oleh Ramesh et al. juga menyimpulkan bahwa SVM mampu mengidentifikasi faktor risiko dengan sensitivitas dan spesifisitas yang baik [9].

Berdasarkan uraian tersebut, penelitian ini dilakukan dengan tujuan untuk mengimplementasikan metode SVM dalam memprediksi risiko penyakit jantung pada lansia. Penelitian ini bertujuan untuk mengembangkan sebuah sistem prediksi yang dapat membantu tenaga medis dan masyarakat umum dalam mendeteksi potensi risiko penyakit jantung sejak dini. Dengan adanya sistem prediksi ini, diharapkan dapat meningkatkan kesadaran dan pencegahan terhadap penyakit jantung pada lansia secara lebih efektif dan efisien [4].

2. MATERIALS AND METHODS

Penelitian ini terdiri dari empat tahapan utama yang dijalankan secara berurutan, yaitu: pengambilan dataset, pembangunan model pembelajaran, pengujian

dan evaluasi model, serta perancangan antarmuka sistem prediksi. Rangkaian proses ini divisualisasikan pada Gambar 1, di mana setiap tahapan memberikan keluaran yang menjadi dasar bagi tahap berikutnya. Proses penelitian dimulai dari tahap pengumpulan data hingga tahap akhir berupa penyajian hasil prediksi melalui antarmuka berbasis web. Penjelasan lebih lanjut mengenai masing-masing tahapan akan diuraikan pada subbab-subbab berikutnya.

Gambar 1. Alur metode penelitian



2.1 Pengumpulan Dataset

Dataset yang digunakan dalam penelitian ini diperoleh dari situs *Kaggle*, dengan nama *Heart Disease Dataset*, yang merupakan hasil kompilasi data dari empat sumber utama, yaitu Cleveland, Hungary, Switzerland, dan Long Beach V, yang dikumpulkan sejak tahun 1988 [10]. Dataset ini tersedia secara publik dan diakses melalui situs *Kaggle*,

<https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>

Dataset ini terdiri atas 14 fitur yang mencakup atribut penting seperti usia, jenis kelamin, tekanan darah saat istirahat, kadar kolesterol, serta status penyakit jantung sebagai variabel target. Format file yang digunakan adalah CSV (Comma Separated Values), yang merupakan format populer dalam analisis data karena kemudahannya dalam integrasi dengan berbagai pustaka *machine learning*, termasuk *Scikit-learn* [11].

Penjelasan dari masing-masing atribut dapat dilihat pada Tabel 1. Dataset ini juga telah dipublikasikan ulang dalam platform berbagi data ilmiah seperti *Zenodo*, sehingga dapat digunakan secara terbuka oleh peneliti dan praktisi dalam pengembangan sistem prediksi penyakit jantung [10].

Tabel 1. Atribut Dataset

No	Atribut	Keterangan
1	age	usia (dalam tahun)
2	sex	jenis kelamin (1 = laki-laki, 0 = perempuan)
3	cp (chest pain type)	jenis nyeri dada (4 tipe)
4	trestbps (resting blood pressure)	tekanan darah saat istirahat (dalam mm Hg saat masuk rumah sakit)
5	chol (serum cholestoral)	kolesterol serum (mg/dl)
6	fbs (fasting blood sugar)	gula darah puasa (120 mg/dl)
7	restecg (resting electrocardiographic results)	hasil elektrokardiografi saat istirahat (nilai 0,1,2)

8	thalach (maximum heart rate achieved)	denyut jantung maksimum yang dicapai
9	exang (exercise induced angina)	angina yang diinduksi oleh latihan
10	oldpeak (ST depression induced by exercise relative to rest)	depresi ST yang diinduksi oleh latihan relatif terhadap istirahat
11	slope (the slope of the peak exercise ST segment)	kemiringan segmen ST saat latihan puncak
12	ca (number of major vessels)	jumlah pembuluh darah utama (0-3)
13	thal	0 = normal; 1 = cacat tetap; 2 = cacat yang dapat dipulihkan
14	target	0 = tidak penyakit jantung, 1 = penyakit jantung

2.2 Pra-pemrosesan data dan pembentukan model learning

Pra-pemrosesan data merupakan tahap penting dalam *pipeline* pembelajaran mesin (*machine learning*) untuk memastikan bahwa data yang digunakan bersih, konsisten, dan dalam format yang sesuai agar dapat diolah secara optimal oleh algoritma [12]. Proses ini mencakup beberapa langkah utama, di antaranya:

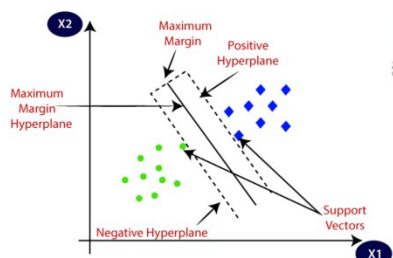
- **Pembersihan Data (*Data Cleaning*)**
Tahap ini melibatkan penghapusan atau penanganan nilai yang hilang (*missing values*), data duplikat, dan data yang tidak relevan agar tidak mengganggu proses pelatihan model.
- **Normalisasi Data (*Feature Scaling*)**
Normalisasi dilakukan untuk menyamakan skala antar fitur numerik, seperti usia dan tekanan darah, agar tidak terjadi dominasi oleh fitur tertentu. Salah satu metode normalisasi yang umum digunakan adalah *Standardization* [11].
- **Encoding Variabel Kategorikal**
Variabel non-numerik seperti jenis kelamin atau tipe nyeri dada dikonversi menjadi bentuk numerik menggunakan teknik seperti *Label Encoding* [13].
- **Split Dataset**
Data dibagi menjadi dua subset utama, yaitu data latih (*training data*) dan data uji (*testing data*), umumnya dengan rasio 80:20 atau 70:30. Pemisahan ini dilakukan untuk memastikan bahwa model dievaluasi secara adil pada data yang belum pernah dilihat selama proses pelatihan.

Setelah proses pra-pemrosesan selesai, model pembelajaran dikembangkan menggunakan algoritma *Support Vector Machine (SVM)*. *SVM* merupakan salah satu metode klasifikasi yang telah digunakan secara luas dalam berbagai bidang, seperti klasifikasi citra, teks, serta prediksi data medis. Sejak dikembangkan oleh Cortes dan Vapnik, *SVM* dikenal sebagai algoritma yang mampu memisahkan dua kelas

dengan menggunakan hyperplane optimal, yaitu garis atau bidang pemisah yang memberikan margin maksimum antara dua kelas [7].

Dalam konteks penelitian ini, dua kelas yang dimaksud adalah 1 = berisiko penyakit jantung dan 0 = tidak berisiko. Pembangunan model dilakukan menggunakan pustaka Scikit-learn dalam bahasa pemrograman Python. Proses klasifikasi diawali dengan pelatihan model pada data latih (training data), lalu menguji performanya menggunakan data uji (testing data). Model kemudian divalidasi melalui sejumlah metrik evaluasi performa, yaitu akurasi, presisi, recall, dan F1-score [14].

Sebagai ilustrasi, Gambar 2 memperlihatkan dua kelas data (kelas 1 dan kelas 2) yang direpresentasikan dalam ruang input dua dimensi (sumbu x dan y). Di antara dua kelas tersebut, SVM menghasilkan hyperplane linier yang memisahkan keduanya dengan bantuan dari tiga support vector utama. Ilustrasi ini menunjukkan bahwa SVM tidak hanya bekerja secara efektif dalam pemisahan data sederhana, tetapi juga mampu menangani masalah klasifikasi kompleks secara efisien [15].



Gambar 2. Klasifikasi SVM

Penelitian oleh Bansal et al. juga menunjukkan bahwa algoritma SVM mampu mengklasifikasikan objek seperti gambar atau teks dengan akurasi tinggi, berkat pendekatannya yang efisien dalam menemukan margin optimal [16]. Hal ini membuktikan bahwa SVM layak digunakan dalam tugas klasifikasi yang sensitif seperti prediksi penyakit jantung.

2.3 Evaluasi Model

Evaluasi model dilakukan untuk mengukur performa algoritma Support Vector Machine (SVM) dalam mengklasifikasikan risiko penyakit jantung. Dalam penelitian ini, evaluasi dilakukan menggunakan empat metrik utama yang sering digunakan dalam klasifikasi data, yaitu: *accuracy*, *presisi* *precision*, *recall*, dan *F1-score*. Keempat metrik ini mampu memberikan gambaran menyeluruh terhadap kualitas prediksi model, terutama dalam kasus klasifikasi biner seperti identifikasi pasien berisiko dan tidak berisiko terhadap penyakit jantung.

Penjelasan masing-masing metrik evaluasi adalah sebagai berikut:

- **Accuracy**
proporsi prediksi yang benar dari total keseluruhan data uji.

- **Precision**
Rasio antara prediksi positif yang benar terhadap seluruh prediksi positif yang dihasilkan model.
- **Recall**
Rasio antara prediksi positif yang benar terhadap seluruh kasus positif aktual.
- **F1-Score**
Rata-rata harmonis dari presisi dan recall, berguna untuk menyeimbangkan trade-off antara keduanya, khususnya saat data tidak seimbang.

Evaluasi ini sangat penting karena akurasi saja tidak cukup untuk menggambarkan performa model, terutama pada dataset yang kelasnya tidak seimbang. Oleh karena itu, F1-score, precision, dan recall menjadi pelengkap penting dalam mengukur keandalan model [17].

Berikut adalah rumus dari masing-masing metrik evaluasi:

- $$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$
- $$\text{Precision} = \frac{TP}{TP+FP}$$
- $$\text{Recall} = \frac{TP}{TP+FN}$$
- $$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

2.4 Pembuatan Interface Prediksi

Setelah model klasifikasi selesai dibangun dan dievaluasi, tahap selanjutnya adalah mengembangkan antarmuka pengguna (*user interface*) untuk memudahkan proses input data kesehatan oleh pengguna. Antarmuka ini dirancang agar pengguna dapat memasukkan informasi medis seperti usia, tekanan darah, kadar kolesterol, dan data lainnya, yang kemudian diproses oleh model untuk menghasilkan prediksi, apakah individu tersebut berisiko terkena penyakit jantung atau tidak.

Pembuatan antarmuka dilakukan menggunakan library Streamlit, yang memungkinkan pengembangan aplikasi web berbasis Python secara interaktif dan ringan [18]. Aplikasi ini dapat dijalankan secara lokal maupun diunggah ke layanan hosting web agar dapat diakses melalui peramban (*browser*) secara online.

3. RESULTS AND DISCUSSION

3.1 Dataset

Dataset ini terdiri dari 1.025 data entri, di mana 526 baris menunjukkan pasien dengan penyakit jantung dan 499 baris lainnya menunjukkan pasien tanpa penyakit jantung. Dataset ini memuat sejumlah atribut penting seperti: usia (*age*), jenis kelamin (*sex*), tipe nyeri dada (*chest pain type*), tekanan darah saat

istirahat (*resting blood pressure*), kolesterol dalam darah (*serum cholestoral*), kadar gula darah puasa (*fasting blood sugar*), hasil EKG saat istirahat (*resting electrocardiographic results*), denyut jantung maksimum (*maximum heart rate achieved*), nyeri dada akibat latihan (*exercise induced angina*), *oldpeak*, kemiringan segmen ST (*slope*), jumlah pembuluh darah utama (*number of major vessels*), dan kondisi thalium (*thal*). Beberapa contoh isi data dapat dilihat pada Tabel 2.

Tabel 2. Sample Dataset

No	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
1	64	1	0	128	263	0	1	105	1	0.2	1	1	3	1
2	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0
3	58	1	0	114	318	0	2	140	0	4.4	0	3	1	0
4	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0

3.2 Pembentukan Learning Dan Evaluasi Model

Tahap ini merupakan proses pembentukan model klasifikasi menggunakan algoritma Support Vector Machine (SVM). Dataset yang telah melalui tahap pra-pemrosesan kemudian dibagi menjadi data latih dan data uji berdasarkan kolom 'target', yang berisi label diagnosis: nilai 1 menunjukkan adanya indikasi penyakit jantung, sedangkan nilai 0 menunjukkan tidak adanya indikasi.

Model SVM dibangun menggunakan kernel linear dengan parameter `probability=True`, yang memungkinkan model memisahkan data menggunakan garis lurus serta memberikan hasil prediksi dalam bentuk probabilitas. Probabilitas ini dapat dimanfaatkan untuk menampilkan tingkat keyakinan model terhadap hasil klasifikasi, misalnya saat diintegrasikan ke dalam antarmuka prediksi.

Proses implementasi pelatihan model tersebut ditunjukkan pada Gambar 3 berikut ini:

```
# Training model SVM
model = SVC(kernel='linear', probability=True)
model.fit(X_train_scaled, y_train)
```

Gambar 3. Kode model SVM

Setelah proses pelatihan model selesai dilakukan, langkah selanjutnya adalah menyimpan model yang telah terbentuk dalam format file .pkl dengan memanfaatkan library pickle. Penyimpanan ini bertujuan agar model dapat digunakan kembali tanpa perlu melakukan pelatihan ulang. Selanjutnya, dilakukan evaluasi performa model guna mengukur seberapa baik model bekerja dalam melakukan klasifikasi. Evaluasi ini mencakup pengukuran terhadap beberapa metrik utama, yaitu: akurasi, presisi, recall, dan F1-score. Proses evaluasi dilakukan dengan menjalankan kode yang ditampilkan pada Gambar 4.

```
=== Evaluasi Model ===
y_pred = model.predict(X_test_scaled)

# Tampilkan classification report
print("=== Classification Report ===")
print(classification_report(y_test, y_pred))
```

Gambar 4. Kode Skor Model

Hasil dari evaluasi model kemudian dirangkum dalam Gambar 5, yang menunjukkan bahwa model berhasil mencapai presisi sebesar 95% untuk kelas 0 dan 77% untuk kelas 1, dengan recall sebesar 85% (kelas 0) dan 92% (kelas 1). Nilai F1-score masing-masing adalah 0.90 untuk kelas 0 dan 0.84 untuk kelas 1. Sementara itu, akurasi keseluruhan model mencapai 87% dari total 110 data uji.

Nilai-nilai ini menunjukkan bahwa model memiliki kinerja klasifikasi yang cukup baik, terutama dalam mengenali pasien yang terindikasi penyakit jantung (kelas 1) dengan recall yang tinggi, serta tetap menjaga presisi yang baik pada pasien yang tidak terindikasi penyakit jantung (kelas 0).

```
=== Classification Report ===
```

	precision	recall	f1-score	support
0	0.95	0.85	0.90	71
1	0.77	0.92	0.84	39
accuracy			0.87	110
macro avg	0.86	0.88	0.87	110
weighted avg	0.89	0.87	0.87	110

Gambar 3. Hasil Skor Evaluasi

3.3 Antarmuka Prediksi Berbasis Streamlit

Interface dilakukan dengan menggunakan aplikasi pihak ketiga yakni streamlit. Sebelum melakukan penerapan di aplikasi tersebut, model berekstensi .pkl yang sudah tersimpan dilakukan import dengan menggunakan library pickle. Data yang sudah tersimpan dalam bentuk pickle diproses dengan menggunakan library streamlit. Setelah library dijalankan akan membentuk file python dengan ekstensi .py yang berisikan aplikasi pendeteksi penyakit jantung. File python tersebut selanjutnya dilakukan proses unggah ke sistem streamlit di <https://streamlit.io/>. sistem prediksi pembuatan sistem prediksi jantung dilakukan pada situs tersebut dengan menghubungkan pada repository github. Pada akhirnya, jantung telah selesai dibangun.

Prediksi Risiko Penyakit Jantung Lansia

isi Formulir Pasien (Klik untuk membuka/tutupi)

Usia (x50 tahun)

65

Jenis Kelamin

Wanita (0)

Tipe Nyeri Dada

Typical Angina (0)

Tekanan Darah Saat Istirahat (mm Hg)

130

Kolesterol (mg/dl)

250

Gambar 5. Sample data input

Beberapa parameter yang menjadi input pada sistem prediksi jantung dapat dilihat pada Gambar 5. Pada gambar tersebut terdapat beberapa parameter yang menjadi contoh dalam melakukan input parameter. Pada parameter age dapat dimasukkan usia dengan usia berupa tahun, sex yang merupakan jenis kelamin dengan 1 merupakan laki laki sementara 0 adalah perempuan, 4 tipe nyeri dada, tekanan darah, dan kolesterol. Sementara itu, hasil prediksi penyakit jantung dapat dilihat pada Gambar 6. Hasil input parameter pada tahapan sebelumnya dapat muncul pada bagian Input Parameter. Sedangkan untuk hasil prediksi berada pada bagian dibawahnya. Berdasarkan input parameter pada gambar tersebut melihtakan hasil prediksi berupa pasien memiliki indikasi penyakit jantung.

Input Parameters

	usia	jenis_kelamin	tipe_nyeri_dada	tekanan_darah	kolesterol	gula_darah	hasil_ekg	detak
0	65	Pria (1)	Typical Angina (0)	130	250	Tidak (0)	Normal (0)	

Prediksi

Hasil Prediksi

⚠️ Pasien memiliki risiko penyakit jantung (80.99%)

Gambar 6. Hasil prediksi

4. CONCLUSION

Penelitian ini telah berhasil mengimplementasikan algoritma Support Vector Machine (SVM) untuk membangun sistem klasifikasi risiko penyakit jantung pada pasien berbasis data medis. Dataset yang digunakan terdiri dari 1.025 data entri dengan 14 atribut utama. Proses pembentukan model meliputi tahapan pra-pemrosesan, pelatihan model menggunakan kernel linear, evaluasi kinerja model, hingga pembuatan antarmuka berbasis web dengan Streamlit.

Hasil evaluasi menunjukkan bahwa model memiliki performa yang baik dengan nilai akurasi sebesar 87%, precision 95% untuk kelas 0 dan 77% untuk kelas 1, serta recall 85% untuk kelas 0 dan 92% untuk kelas 1. Nilai-nilai tersebut menunjukkan bahwa model mampu mengenali pasien yang terindikasi penyakit jantung dengan baik, serta mempertahankan ketepatan prediksi pada pasien yang tidak terindikasi.

Sistem prediksi yang dikembangkan juga telah berhasil diintegrasikan ke dalam antarmuka web berbasis Streamlit, yang memudahkan pengguna dalam melakukan input data dan memperoleh hasil prediksi secara langsung dan interaktif.

Dengan demikian, model ini berpotensi digunakan sebagai alat bantu skrining awal untuk mendeteksi risiko penyakit jantung, terutama bagi kalangan medis atau pengguna umum yang membutuhkan pemeriksaan cepat dan efisien. Penelitian ini dapat dikembangkan lebih lanjut dengan penggunaan dataset yang lebih besar serta eksplorasi algoritma lain untuk peningkatan akurasi dan generalisasi model.

REFERENCES

- [1] World Health Organization, "Cardiovascular diseases (CVDs)," 2021. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-cvds>
- [2] Kementerian Kesehatan RI, "Infodatin: Penyakit Jantung," Pusat Data dan Informasi, 2020. [Online]. Available: <https://pusdatin.kemkes.go.id>
- [3] I. Rahmawati dan H. A. Nugroho, "Risiko Penyakit Jantung pada Lansia: Faktor dan Pencegahan," Jurnal Kesehatan Masyarakat, vol. 8, no. 2, pp. 140–147, 2020.
- [4] A. Supriyanto, R. H. Putra, dan A. F. Harahap, "Sistem Prediksi Penyakit Jantung Menggunakan Metode Klasifikasi," Jurnal Teknologi Informasi dan Ilmu Komputer, vol. 5, no. 1, pp. 1–6, 2022.
- [5] S. Yadav, R. Patel, dan R. Sharma, "Heart Disease Prediction using Hybrid Model," Procedia Computer Science, vol. 167, pp. 1960–1968, 2020.
- [6] J. Han, M. Kamber, dan J. Pei, Data Mining: Concepts and Techniques, 3rd ed. Waltham, MA: Morgan Kaufmann, 2012.
- [7] C. Cortes dan V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995. [Online]. Available: <https://doi.org/10.1007/BF00994018>
- [8] R. Dey, S. Chakraborty, dan A. Mukherjee, "Heart Disease Prediction Model Using SVM and Logistic Regression," International Journal of Innovative Research in Computer and Communication Engineering, vol. 9, no. 4, pp. 2981–2988, 2021.
- [9] G. Ramesh dan B. V. Kavitha, "Prediction of Heart Disease Using Support Vector Machine with Optimization Techniques," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 5, no. 2, pp. 63–70, 2020.
- [10] F. Soriano, "Heart Failure Prediction Dataset," Kaggle, 2022. [Online]. Available: <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>
- [11] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow, 2nd ed. Sebastopol, CA: O'Reilly Media, 2019.

- [12] S. B. Kotsiantis, D. Kanellopoulos, dan P. E. Pintelas, "Data Preprocessing for Supervised Learning," *International Journal of Computer Science*, vol. 1, no. 2, pp. 111–117, 2006.
- [13] A. Géron, "Data Preprocessing – Label Encoding & One-Hot Encoding," *Hands-On Machine Learning*, 2nd ed., O'Reilly, 2019, ch. 2.
- [14] A. Géron, "Evaluating Classifiers: Accuracy, Precision, Recall," *Hands-On Machine Learning*, 2nd ed., O'Reilly, 2019, ch. 3.
- [15] R. Abdu, I. Kurniawan, dan A. Rahmadani, "Implementasi Algoritma Support Vector Machine (SVM) untuk Klasifikasi Citra," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 1, pp. 20–26, 2020.
- [16] M. Bansal, K. Kumari, dan R. K. Sehgal, "Comparative Analysis of SVM and CNN for Image Classification," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 11, no. 3, pp. 88–92, 2022.
- [17] D. M. W. Powers, "Evaluation: From Precision, Recall and F-measure to ROC, Informedness, Markedness & Correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.
- [18] Streamlit, "Streamlit Docs," *Streamlit.io*, 2023. [Online]. Available: <https://docs.streamlit.io>