

Evaluasi Kinerja Model LSTM Untuk Prediksi Risiko Penyakit Jantung Menggunakan Dataset

M. Alfi Saputra^{*1}, Tri Sugihartono²

^{1,2}Program Studi Teknik Informasi, Fakultas Teknologi Informasi, ISB Atma Luhur, Indonesia
Email: ¹2111500107@mahasiswa.atmaluhur.ac.id, ²trisugihartono@atmaluhur.ac.id

Abstrak

Penyakit jantung merupakan salah satu penyebab utama kematian di dunia, termasuk di Indonesia. Deteksi dini terhadap risiko penyakit jantung menjadi krusial dalam menurunkan angka kematian dan meningkatkan kualitas hidup pasien. Penelitian ini bertujuan untuk mengevaluasi kinerja model *Long Short-Term Memory* (LSTM), sebagai bagian dari metode *deep learning*, dalam memprediksi risiko penyakit jantung dengan menggunakan data simulasi. Data terdiri dari 70.000 entri dengan 16 variabel yang mencerminkan kondisi klinis dan gaya hidup pasien. Proses penelitian meliputi tahapan *preprocessing*, pelatihan model LSTM, serta evaluasi menggunakan metrik seperti akurasi, presisi, *recall*, F1-score, dan AUC-ROC. Hasil evaluasi menunjukkan bahwa model LSTM mencapai akurasi sebesar 0,8034, presisi 0,8055, *recall* 0,8023, F1-score 0,8039, dan AUC-ROC 0,8036. Meskipun performa model cukup menjanjikan, masih terdapat kelemahan berupa jumlah *false negative* yang signifikan. Hal ini mengindikasikan perlunya pengembangan lebih lanjut, seperti penyesuaian ambang prediksi, teknik penyeimbangan kelas, dan eksplorasi model *ensemble*. Penelitian ini memberikan kontribusi konkret dalam penerapan model LSTM untuk sistem prediksi risiko penyakit jantung, dan diharapkan dapat memberikan kontribusi pada pengembangan sistem prediksi penyakit jantung berbasis kecerdasan buatan. Temuan ini juga berimplikasi pada penguatan pendekatan ilmiah dalam bidang ilmu kesehatan digital dan potensi penerapannya dalam praktik klinis secara lebih luas.

Kata kunci: *Deep Learning, Heart Disease, LSTM, Machine Learning, Prediksi Medis.*

Performance Evaluation of the LSTM Model for Heart Disease Risk Prediction Using a Dataset

Abstract

Heart disease is one of the leading causes of death worldwide, including in Indonesia. Early detection of heart disease risk is crucial to reducing mortality rates and improving patients' quality of life. This study aims to evaluate the performance of the Long Short-Term Memory (LSTM) model, as part of a deep learning method, in predicting heart disease risk using simulated data. The dataset consists of 70,000 entries with 16 variables reflecting patients' clinical conditions and lifestyle factors. The research process includes preprocessing, training the LSTM model, and evaluation using metrics such as accuracy, precision, recall, F1-score, and AUC-ROC. The evaluation results show that the LSTM model achieved an accuracy of 0.8034, precision of 0.8055, recall of 0.8023, F1-score of 0.8039, and AUC-ROC of 0.8036. Although the model's performance is fairly promising, it still presents a significant number of false negatives. This indicates the need for further development, such as threshold adjustment, class balancing techniques, and the exploration of ensemble models. This study provides a concrete contribution to the application of LSTM models in heart disease risk prediction systems and is expected to support the development of artificial intelligence-based prediction systems for heart disease. These findings also imply a strengthening of scientific approaches in the field of digital health science and highlight the potential for real-world clinical implementation.

Keywords: *Deep Learning, Heart Disease, LSTM, Machine Learning, Medical Prediction.*

1. PENDAHULUAN

Penyakit jantung merupakan salah satu penyakit tidak menular yang menjadi penyebab kematian utama di dunia, termasuk di Indonesia. Berdasarkan data WHO, lebih dari 17 juta orang di dunia meninggal setiap tahunnya akibat penyakit jantung. Di Indonesia, data dari Institute for Health Metrics and Evaluation pada tahun

2019 menunjukkan bahwa penyakit kardiovaskular bertanggung jawab atas 651.481 kematian per tahun, termasuk stroke dan penyakit jantung koroner, peningkatan gaya hidup yang tidak sehat, seperti merokok, pola makan yang tidak seimbang, dan kurangnya aktivitas fisik, berkontribusi besar terhadap meningkatnya prevalensi penyakit jantung di Indonesia. Riskesdas 2018 mencatat bahwa prevalensi penyakit jantung meningkat dari 0,5% pada 2013 menjadi 1,5% pada 2018.[1] Dengan meningkatnya prevalensi penyakit ini, deteksi dini sangat penting untuk mengurangi resiko kematian dan meningkatkan kualitas hidup pasien.

Perkembangan teknologi kecerdasan buatan, khususnya metode deep learning, telah menunjukkan potensi besar dalam membantu prediksi penyakit jantung. Salah satu pendekatan yang banyak digunakan adalah model Long Short-Term Memory (LSTM), yang dikenal karena kemampuannya dalam menangani data berurutan dan kompleksitas yang tinggi[2]. Model ini telah digunakan dalam berbagai aplikasi, termasuk prediksi risiko penyakit kronis seperti penyakit jantung, dengan hasil yang menjanjikan[3]. Namun, terdapat beberapa tantangan dalam akurasi prediksi risiko penyakit jantung menggunakan model deep learning. Salah satu masalah utama adalah ketidakseimbangan data pada dataset kesehatan, yang dapat menyebabkan bias dalam hasil prediksi[4]. Selain itu, meskipun LSTM memiliki keunggulan dalam menangani data sekuensial, efektivitasnya dibandingkan dengan metode lain seperti Random Forest atau Support Vector Machine masih menjadi perdebatan[5].

Beberapa studi sebelumnya telah menunjukkan efektivitas LSTM dalam konteks prediksi penyakit jantung. Salah satu penelitian menunjukkan bahwa model LSTM mampu memprediksi onset Congestive Heart Failure (CHF) hingga 15 bulan sebelumnya dengan nilai AUC-ROC sebesar 0,9147. Studi ini menekankan keunggulan LSTM dalam menangani data rekam medis elektronik (EHR) yang bersifat longitudinal dan tidak teratur.[6] Selanjutnya, Penelitian lain mengembangkan kerangka kerja yang mengintegrasikan LSTM dengan Pulse Coupled Neural Network (PCNN) untuk prediksi penyakit jantung, rekomendasi pengobatan, dan manajemen efek samping. Pendekatan ini menunjukkan peningkatan akurasi dalam analisis data klinis temporal.[7] Selain itu, terdapat pendekatan berbasis algoritma C-BiLSTM yang menggabungkan pemilihan fitur dan optimalisasi saluran pada jaringan area tubuh nirkabel untuk meningkatkan akurasi prediksi.[8]

Namun, mayoritas penelitian tersebut masih bergantung pada data klinis nyata, yang dalam praktiknya seringkali sulit diakses karena alasan privasi, keterbatasan distribusi, serta ketidakseimbangan kelas. Untuk mengatasi hal ini, pendekatan berbasis data sintetis mulai dikembangkan

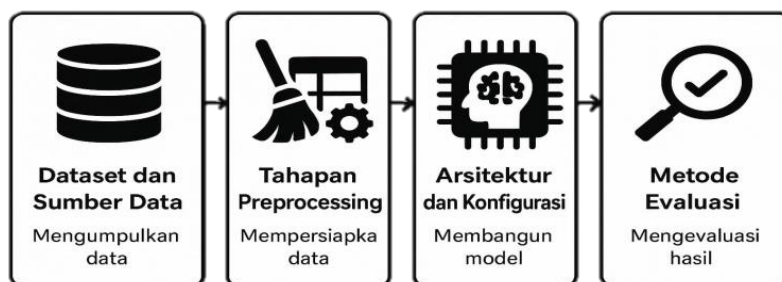
Berdasarkan studi-studi tersebut, dapat diidentifikasi adanya celah penelitian, yaitu kurangnya eksplorasi penggunaan model LSTM dalam prediksi risiko penyakit jantung dengan memanfaatkan dataset sintetis. Padahal, penggunaan data sintetis sangat relevan untuk mengatasi keterbatasan data klinis sekaligus menjaga privasi pasien. Oleh karena itu, penting untuk melakukan evaluasi mendalam terhadap kinerja model LSTM dengan menggunakan metrik evaluasi seperti AUC-ROC, F1-Score, Precision, dan Recall untuk memastikan keandalan model ini dalam konteks prediksi risiko penyakit jantung.[9]

Tujuan dari penelitian ini adalah untuk mengevaluasi kinerja model LSTM dalam memprediksi risiko penyakit jantung dengan menggunakan dataset sintetis. Penelitian ini akan menganalisa metrik evaluasi seperti

AUC-ROC, F1-Score, Precision, dan Recall untuk memberikan gambaran yang jelas mengenai efektivitas model dalam konteks deteksi dini penyakit jantung. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi bagi pengembang sistem prediktif berbasis deep learning yang lebih baik serta membantu pengambilan keputusan klinis secara lebih akurat.

2. METODE PENELITIAN

Metodologi penelitian yang dapat digunakan untuk prediksi penyakit jantung menggunakan Deep Learning berbasis Long Short Term Memory (LSTM). Metodologi ini akan mencakup serangkaian langkah-langkah yang diperlukan untuk merancang, mengembangkan, dan mengevaluasi model prediksi penyakit jantung yang efektif menggunakan LSTM.



Gambar 1. Tahapan Penelitian

2.1. Dataset dan Sumber Data

Dataset yang digunakan dalam penelitian ini diperoleh dari situs <https://www.kaggle.com/datasets/mahatiratusher/heart-disease-risk-prediction-dataset/data> yang berisi data sebanyak 70.000 pasien dengan 16 variabel prediktor. Atribut dalam dataset mencakup informasi penting seperti usia pasien, status merokok, riwayat nyeri dada, tingkat obesitas, serta riwayat keluarga terhadap penyakit jantung. Data ini mencerminkan kombinasi antara faktor demografis, gaya hidup, dan kondisi medis yang relevan dalam proses prediksi risiko penyakit jantung. Pemilihan dataset ini didasarkan pada kelengkapan fitur dan jumlah sampel yang besar, sehingga dianggap representatif untuk melatih model deep learning seperti LSTM. Analisis terhadap dataset ini diharapkan dapat menghasilkan model prediktif yang akurat serta mendukung pengambilan keputusan dalam deteksi dini penyakit jantung.[10]

2.2. Tahapan Preprocessing

Dataset yang ada akan dilakukan preprocessing sebelum metode LSTM diterapkan dalam mengelola dataset tersebut. Preprocessing pada dataset ini meliputi penghapusan kolom yang tidak diperlukan dan pengisian nilai hilang (missing values). Untuk menangani nilai yang hilang pada himpunan data, teknik normalisasi digunakan dalam penelitian ini. Teknik normalisasi digunakan untuk mengubah nilai-nilai pada suatu skala yang seragam. Nilai yang hilang digantikan dengan nilai z. Persamaan menerapkan normalisasi z-score terhadap himpunan data yang dikumpulkan.

$$Z_i = \frac{(x_i - \min(x))}{(\max(x) - \min(x))} \quad (1)$$

Di mana x merupakan vektor input, max(x) menyatakan nilai maksimum dalam himpunan data, dan min(x) menyatakan nilai minimum dalam himpunan data. Setelah proses prapengolahan dilakukan, seperti penghapusan atribut yang tidak relevan dan penanganan nilai yang hilang menggunakan normalisasi, dilakukan pembagian data (data splitting) menjadi dua subset utama data pelatihan (training set) dan data pengujian (testing set). Dalam penelitian ini, pembagian dilakukan dengan proporsi 70% untuk data pelatihan dan 30% untuk data pengujian[11]

2.3. Arsitektur dan Konfigurasi Model

Penelitian ini akan menggunakan Long Short-Term Memory untuk memprediksi penyakit jantung menggunakan dataset sintetis. LSTM memiliki kemampuan untuk mempelajari data yang harus digunakan atau diabaikan, proses ini terjadi pada setiap neuron. LSTM banyak digunakan untuk mengolah teks, video, dan data deret waktu. Hal ini berdasarkan lebih banyak informasi sebelumnya dapat mempengaruhi akurasi model.[12] Arsitektur LSTM dibangun dengan lapisan input, lapisan output dan lapisan hidden dimana lapisan hidden memiliki dengan tiga gerbang utama yaitu input, forget, dan output gate. input gate berfungsi untuk mengontrol seberapa banyak informasi disimpan. Forget gate digunakan untuk mengontrol sejauh mana nilai ditetapkan dalam sel memori. Dan output gate mengontrol seberapa banyak nilai dalam sel memori yang digunakan untuk menghitung hasil output-nya.[13].

$$f_t = \sigma(W_f^h \cdot h_t^{-1} + W_f^x \cdot x_t + b_f) \quad (2)$$

$$i_t = \sigma(W_i^h \cdot h_t^{-1} + W_i^x \cdot x_t + b_i) \quad (3)$$

$$\tilde{C}_t = \tanh(W_c^h \cdot h_t^{-1} + W_c^x \cdot x_t + b_c) \quad (4)$$

$$C_t = f_t \cdot C_t^{-1} + i_t \cdot \tilde{C}_t \quad (5)$$

$$o_t = \sigma(W_o^h \cdot h_t^{-1} + W_o^x \cdot x_t + b_o) \quad (6)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (7)$$

Dalam hal ini, C_t , C_{t-1} , dan $\tilde{C}_t$ masing-masing merepresentasikan nilai cell state saat ini, nilai cell state pada waktu sebelumnya, serta pembaruan terhadap cell state saat ini. Simbol f_t , i_t , dan o_t masing-masing menunjukkan forget gate, input gate, dan output gate. Dengan pengaturan parameter yang tepat, berdasarkan nilai $\tilde{C}_t$ dan C_t , nilai keluaran h_t dihitung sesuai dengan persamaan (5) hingga (7).[14] Model LSTM dilatih selama 30 epoch dengan batch size sebesar 32, yang merupakan nilai umum untuk pelatihan model skala sedang. Untuk optimasi, digunakan algoritma Adam Optimizer dengan learning rate awal sebesar 0.001 karena kemampuannya dalam menangani gradien yang tidak stabil dan konvergensi yang cepat[15] Fungsi loss yang digunakan adalah Binary Cross-Entropy karena klasifikasi penyakit jantung merupakan masalah klasifikasi biner[16] Model dikembangkan menggunakan Python dan diimplementasikan dengan TensorFlow serta Keras, dua pustaka deep learning populer yang menyediakan antarmuka fleksibel untuk membangun dan melatih jaringan saraf tiruan. TensorFlow dan Keras banyak digunakan dalam penelitian karena dokumentasi yang lengkap serta kemudahan integrasi dengan berbagai framework lainnya.[17] Seluruh eksperimen dilakukan pada lingkungan Google Colaboratory (Colab), yang menyediakan akses GPU gratis dan memungkinkan pelatihan model dilakukan secara efisien di cloud.[18]

2.4. Metode Evaluasi

Evaluasi model pada tahapan ini dilakukan evaluasi terhadap kecepatan, akurasi, precision, dan recall terhadap algoritma yang digunakan. Evaluasi kecepatan pada dimaksudkan untuk mengetahui kecepatan eksekusi yang dibutuhkan masing-masing metode atau algoritma ketika dalam proses menjalankan algoritma. Akurasi merupakan salah satu metrik klasik yang digunakan untuk melakukan evaluasi model dari klasifikasi.[19]

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (8)$$

Pengukuran *performance metrics* terdiri dari presisi, *recall* dan *f1-score*. Presisi adalah rasio positif atau derajat keandalan, yaitu proporsi prediksi berlabel positif yang benar terhadap prediksi positif keseluruhan[20].

$$Precision = \frac{TP}{TP+FP} \quad (9)$$

Recall juga dikenal sebagai *true positive rate* atau *sensitivitas*. *Recall* juga disebut sebagai derajat keandalan model dalam mendeteksi data berlabel positif dengan benar.

$$Recall = \frac{TP}{TP+FN} \quad (10)$$

F1-score merangkum semua hasil perhitungan *precision* dan *recall* dengan membuat rata-rata harmonik.

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (11)$$

3. HASIL DAN PEMBAHASAN

3.1. Table Isi Dataset

Table 1. Table Dataset

	CP	SOB	Ftg	Palp	Dizz	Swell	P_A_J_B	CS	Ns	HBP
0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
1	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0
2	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0
3	1.0	1.0	0.0	1.0	0.0	0.0	1.0	1.0	1.0	1.0
4	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
	HChol	Diab	Smk	Obs	Sed_Life	Fam_Hist	C_Stress	Gndr	Age	H_Risk
0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	48.0	0.0
1	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	46.0	0.0
2	1.0	0.0	1.0	1.0	1.0	0.0	1.0	1.0	66.0	0.0
3	1.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0	60.0	1.0
4	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	69.0	0.0

Chest Pain(CP), Shortness_of_Breath(SOB), Fatigue(Ftg), Palpitations(Palp), Dizziness(Dizz), Swelling(Swell), Pain_Arms_Jaw_Back(P_A_J_B), Cold_Sweats(CS), Nausea(Ns), High_BP(HBP), High_Cholesterol(HChol), Diabetes(Diab), Smoking(Smk), Obesity(Obs), Sedentary_Lifestyle(Sed_Life), Family_History(Fam_Hist), Chronic_Stress(C_Stress), Gender(Gndr), Age(Age), Heart_Risk(H_Risk).

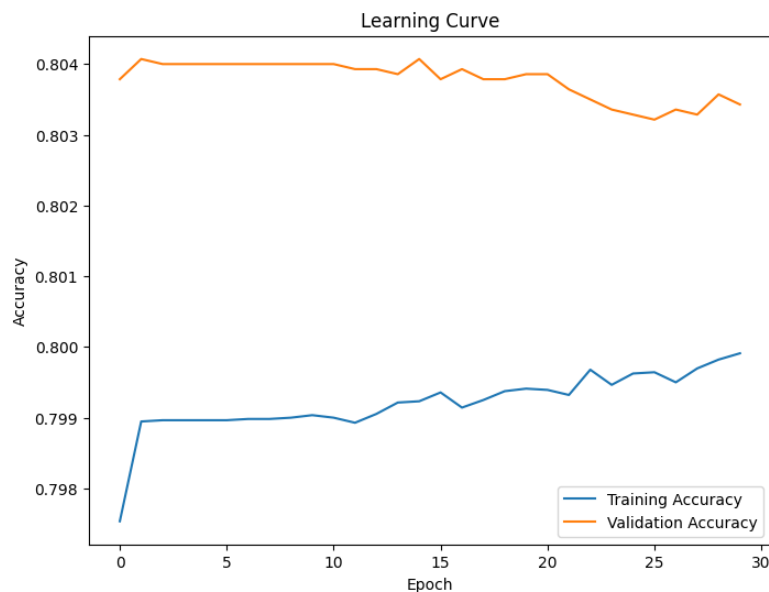
Analisis deskriptif terhadap dataset sintetis yang disajikan dalam Tabel 1 mengungkapkan distribusi beragam karakteristik klinis dan faktor risiko kardiovaskular yang relevan untuk prediksi risiko penyakit jantung. Variabel-variabel biner, termasuk manifestasi simptomatik seperti nyeri dada (CP), sesak napas (SOB), kelelahan (Ftg), palpitasi (Palp), pusing (Dizz), dan pembengkakan (Swell), serta indikator fisiologis dan gaya hidup seperti nyeri pada lengan, rahang, dan punggung (P_A_J_B), keringat dingin (CS), mual (Ns), tekanan darah tinggi (HBP), kolesterol tinggi (HChol), diabetes (Diab), kebiasaan merokok (Smk), obesitas (Obs), dan gaya hidup sedenter (Sed_Life), memberikan representasi diskrit atas kehadiran atau ketiadaan kondisi-kondisi tersebut pada entitas data sintetis. Lebih lanjut, riwayat keluarga penyakit jantung (Fam_Hist) dan stres kronis (C_Stress) turut disertakan sebagai faktor risiko hereditas dan psikososial. Karakteristik demografis seperti jenis kelamin (Gndr) dan usia (Age) juga tercakup, dengan jenis kelamin direpresentasikan secara biner dan usia sebagai variabel kontinu. Variabel target, risiko penyakit jantung (H_Risk), dikodekan secara biner untuk memfasilitasi tugas klasifikasi. Keberagaman representasi fitur dalam dataset sintetis ini diharapkan dapat secara komprehensif mencerminkan kompleksitas interaksi antar faktor risiko dalam memprediksi kemungkinan terjadinya penyakit jantung, sehingga menyediakan landasan yang adekuat untuk evaluasi kinerja model deep learning berbasis Long Short-Term Memory (LSTM) dalam tugas prediksi risiko penyakit jantung.

Dan berikut ada juga keterangan dari parameter yang digunakan:

Tabel 2. Parameter yang digunakan

Variabel	Keterangan
chest_pain	Kehadiran nyeri dada (Ya/Tidak; 0 = Tidak, 1 = Ya)
shortness_of_breath	Kesulitan bernapas (Ya/Tidak; 0 = Tidak, 1 = Ya)
fatigue	Kelelahan yang persisten tanpa penyebab yang jelas (Ya/Tidak; 0 = Tidak, 1 = Ya)
palpitations	Detak jantung yang tidak teratur atau cepat (Ya/Tidak; 0 = Tidak, 1 = Ya)
dizziness	Episode pusing atau pingsan (Ya/Tidak; 0 = Tidak, 1 = Ya)
swelling	Pembengkakan di kaki/ankle akibat retensi cairan (Ya/Tidak; 0 = Tidak, 1 = Ya)
radiating_pain	Nyeri yang menjalar ke lengan, rahang, leher, atau punggung (Ya/Tidak; 0 = Tidak, 1 = Ya)
cold_sweats	Keringat dingin dan mual (Ya/Tidak; 0 = Tidak, 1 = Ya)
age	Usia pasien dalam tahun (Contoh: 25, 45, 65)
hypertension	Riwayat tekanan darah tinggi (Ya/Tidak; 0 = Tidak, 1 = Ya)
cholesterol_high	Kadar kolesterol tinggi (Ya/Tidak; 0 = Tidak, 1 = Ya)
diabetes	Diagnosis diabetes (Ya/Tidak; 0 = Tidak, 1 = Ya)
smoker	Riwayat merokok (Ya/Tidak; 0 = Tidak, 1 = Ya)
obesity	Status obesitas (Ya/Tidak; 0 = Tidak, 1 = Ya)
family_history	Riwayat keluarga penyakit jantung (Ya/Tidak; 0 = Tidak, 1 = Ya)
risk_label	Risiko penyakit jantung (Label biner; 0 = Risiko rendah, 1 = Risiko tinggi)

3.2. Learning Curve

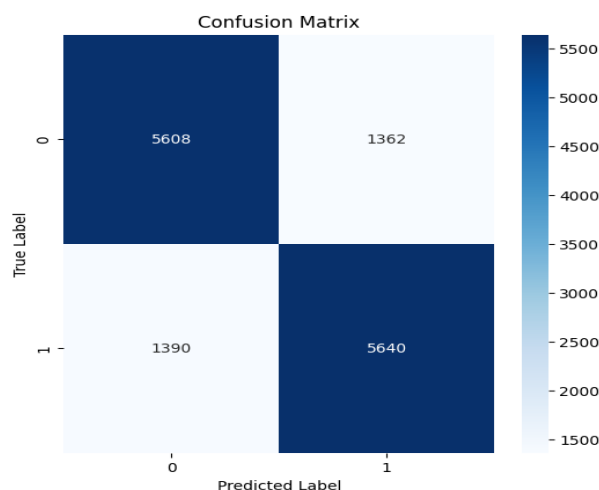


Gambar 2. Grafik Learning Curve

Gambar di atas menyajikan kurva pembelajaran (learning curve) dari model Long Short-Term Memory (LSTM) yang digunakan dalam penelitian ini untuk melakukan prediksi risiko penyakit jantung berbasis dataset sintetis. Kurva ini menggambarkan perkembangan akurasi pada data pelatihan dan data validasi selama proses pelatihan berlangsung hingga mencapai 30 epoch. Berdasarkan grafik, terlihat bahwa akurasi pelatihan mengalami peningkatan secara perlahan dari awal hingga akhir proses pelatihan, dengan rentang nilai antara 0,798 hingga 0,800. Sementara itu, akurasi validasi menunjukkan kestabilan pada kisaran nilai 0,803 hingga 0,804, meskipun terdapat sedikit fluktuasi pada beberapa epoch terakhir. Pola ini mengindikasikan bahwa model tidak mengalami gejala overfitting yang signifikan, karena performa pada data validasi tetap konsisten dan tidak mengalami penurunan yang mencolok dibandingkan dengan data pelatihan.

Secara keseluruhan, performa model menunjukkan kestabilan selama proses pelatihan, dengan selisih akurasi yang relatif kecil antara data pelatihan dan validasi. Hal ini menandakan bahwa model LSTM mampu melakukan generalisasi dengan baik terhadap data yang belum pernah dilihat sebelumnya. Namun demikian, peningkatan akurasi yang diperoleh relatif kecil, sehingga untuk memperoleh performa yang lebih optimal, perlu dilakukan eksplorasi lebih lanjut seperti tuning hiperparameter, penyesuaian arsitektur model, serta penerapan teknik regularisasi atau augmentasi data.

3.3. Confusion Matrix



Gambar 3. Grafik Confusion Matrix

Analisis terhadap *confusion matrix* dari model *Long Short-Term Memory* (LSTM) yang diimplementasikan untuk klasifikasi risiko penyakit jantung berdasarkan dataset sintetis. *Confusion matrix* merupakan alat visualisasi kinerja model klasifikasi yang membandingkan antara label aktual (*true label*) dengan label yang diprediksi oleh model (*predicted label*). Berdasarkan hasil yang tertera pada gambar, model LSTM menunjukkan kemampuan dalam mengklasifikasikan sampel. Sebanyak 5.608 sampel dengan kelas negatif (label 0) berhasil diprediksi dengan benar sebagai negatif (*True Negative*), dan 5.640 sampel dengan kelas positif (label 1) berhasil diprediksi dengan benar sebagai positif (*True Positive*). Namun, terdapat pula kesalahan klasifikasi, di mana 1.362 sampel negatif secara keliru diprediksi sebagai positif (*False Positive*), serta 1.390 sampel positif yang diprediksi sebagai negatif (*False Negative*).

Dari confusion matrix tersebut, beberapa metrik evaluasi kinerja model dapat dihitung, yaitu:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} = \frac{5640+5608}{5640+5608+1362+1390} \approx 0.804$$

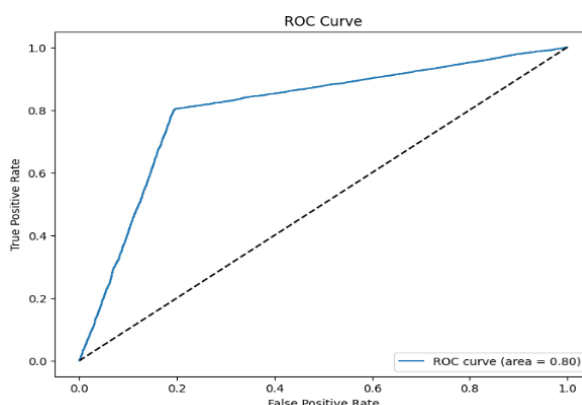
$$Precision = \frac{TP}{TP+FP} = \frac{5640}{5640+1362} \approx 0.805$$

$$Recall = \frac{TP}{TP+FN} = \frac{5640}{5640+1390} \approx 0.802$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \approx 0.803$$

Nilai-nilai metrik evaluasi yang diperoleh mengindikasikan bahwa model LSTM mencapai performa klasifikasi yang relatif seimbang antara kemampuannya dalam mengidentifikasi kasus positif maupun negatif, dengan akurasi, presisi, recall, dan F1-score di sekitar 80%. Meskipun demikian, jumlah false negative yang relatif tinggi menjadi perhatian signifikan, terutama dalam konteks prediksi risiko penyakit jantung. Kesalahan dalam mengidentifikasi individu yang sebenarnya berisiko dapat berimplikasi serius terhadap penundaan atau kekurangan intervensi medis yang tepat. Oleh karena itu, upaya lebih lanjut diperlukan untuk meningkatkan performa model. Beberapa strategi yang dapat dipertimbangkan meliputi penyesuaian ambang batas prediksi (*threshold tuning*), implementasi teknik penyeimbangan kelas (*class balancing*) untuk mengatasi potensi bias akibat distribusi kelas yang tidak seimbang, atau eksplorasi penggunaan pendekatan ensemble model untuk meningkatkan sensitivitas terhadap kelas positif tanpa mengorbankan presisi klasifikasi secara keseluruhan.

3.4. ROC/AUC Curve



Gambar 4. Grafik ROC/AUC Curve

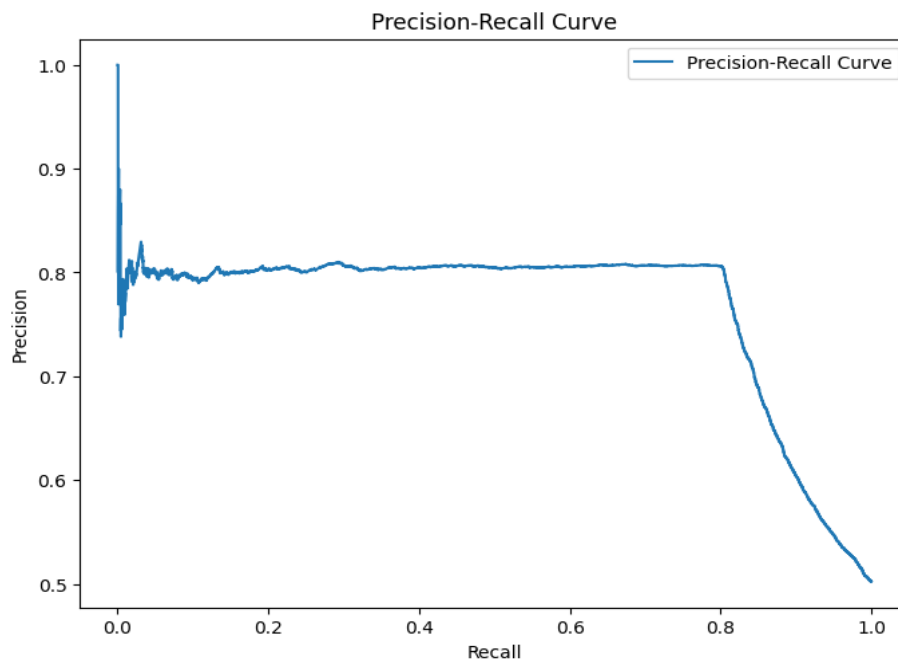
Kurva Receiver Operating Characteristic (ROC) dari model Long Short-Term Memory (LSTM) yang digunakan untuk prediksi risiko penyakit jantung berdasarkan dataset sintetis. Kurva ROC digunakan sebagai salah satu indikator penting dalam evaluasi model klasifikasi biner, dengan menggambarkan hubungan antara True Positive Rate (TPR) dan False Positive Rate (FPR) pada berbagai nilai ambang batas klasifikasi. Dalam kurva tersebut, sumbu horizontal (X) merepresentasikan nilai False Positive Rate (FPR), sedangkan sumbu vertikal (Y) merepresentasikan nilai True Positive Rate (TPR). garis biru pada grafik menunjukkan performa model aktual, sedangkan garis diagonal putus-putus merepresentasikan klasifikasi acak sebagai baseline. Semakin dekat kurva ROC terhadap sudut kiri atas grafik, maka semakin baik performa model dalam

membedakan antara kelas positif dan negatif. Berdasarkan hasil evaluasi, diperoleh nilai Area Under Curve (AUC) sebesar 0,80, yang mengindikasikan bahwa model LSTM memiliki kapabilitas diskriminatif yang baik. AUC sebesar 0,80 dapat diartikan bahwa terdapat kemungkinan sebesar 80% bahwa model akan memberikan skor probabilitas yang lebih tinggi untuk sampel dari kelas positif dibandingkan dengan sampel dari kelas negatif secara acak. Nilai AUC ini termasuk ke dalam kategori performa “baik” berdasarkan interpretasi umum sebagai berikut:

- 0,90 – 1,00 : Sangat baik (excellent)
- 0,80 – 0,90 : Baik (good)
- 0,70 – 0,80 : Cukup (fair)
- 0,60 – 0,70 : Buruk (poor)
- 0,50 – 0,60 : Gagal (fail)

Dengan demikian, hasil evaluasi kurva ROC dan nilai AUC menunjukkan bahwa model LSTM yang dikembangkan dalam penelitian ini mampu memprediksi risiko penyakit jantung dengan tingkat ketepatan yang cukup tinggi. Hasil ini juga konsisten dengan metrik evaluasi lainnya seperti akurasi, precision, recall, dan F1-score yang sebelumnya telah dibahas, sehingga memperkuat kesimpulan bahwa model memiliki kinerja klasifikasi yang andal. Secara teknis, grafik ROC ini dihasilkan menggunakan bahasa pemrograman Python yang dijalankan pada platform Google Colaboratory (Google Colab). Seluruh proses evaluasi model, termasuk perhitungan metrik menggunakan pustaka scikit-learn dan visualisasi grafik ROC menggunakan pustaka matplotlib, dilakukan secara menyeluruh di lingkungan Google Colab. Pemanfaatan platform ini memungkinkan integrasi yang efisien antara pemodelan, analisis, dan visualisasi dalam satu sistem komputasi berbasis cloud.

3.5. Precision-Recall Curve



Gambar 5. Grafik Precision-Recall Curve

Gambar Precision-Recall (PR) yang dihasilkan dari model LSTM dalam melakukan klasifikasi risiko penyakit jantung. Kurva ini digunakan untuk mengevaluasi performa model klasifikasi biner, khususnya pada kasus di mana terdapat ketidakseimbangan kelas (class imbalance) yang signifikan. Pada kurva ini, sumbu X merepresentasikan Recall (sensitivitas), yaitu proporsi kasus positif yang berhasil dikenali oleh model, sedangkan sumbu Y merepresentasikan Precision, yaitu proporsi prediksi positif yang benar. Kurva ini diplot menggunakan library matplotlib dalam bahasa pemrograman Python, sebuah library yang umum digunakan untuk visualisasi data dalam analisis data dan machine learning. Gambar menunjukkan bahwa model LSTM mampu mempertahankan nilai precision yang cukup tinggi (sekitar 0,80) pada berbagai nilai recall. Pada bagian tengah grafik, terlihat kestabilan hubungan antara precision dan recall, yang menunjukkan bahwa model

konsisten dalam mengidentifikasi kasus positif tanpa menghasilkan terlalu banyak prediksi positif yang salah. Namun, seiring meningkatnya nilai recall mendekati 1, terjadi penurunan signifikan pada nilai precision. Fenomena ini merupakan hal yang umum, karena untuk meningkatkan recall secara ekstrem, model cenderung memberikan lebih banyak prediksi positif, yang dapat meningkatkan jumlah false positive dan menyebabkan precision menurun. Sebaliknya, pada bagian awal grafik (recall rendah), precision cenderung tinggi bahkan mendekati 1, namun hal ini tidak selalu mencerminkan performa keseluruhan model karena hanya sebagian kecil data yang dipertimbangkan. Oleh karena itu, pemahaman kurva PR ini penting untuk mengevaluasi trade-off antara kemampuan model dalam menemukan semua kasus positif dan kemampuannya untuk menghindari kesalahan prediksi positif. Secara keseluruhan, bentuk kurva Precision-Recall ini menunjukkan bahwa model memiliki kemampuan deteksi yang baik terhadap kasus positif (penderita penyakit jantung) sambil tetap menjaga tingkat kesalahan prediksi (false positive) dalam batas yang dapat diterima. Hasil ini juga konsisten dengan nilai-nilai metrik evaluasi lainnya seperti accuracy, precision, recall, F1-score, dan AUC yang sebelumnya telah dijelaskan. Oleh karena itu, dapat disimpulkan bahwa model LSTM menunjukkan kinerja yang cukup stabil dan andal dalam memprediksi risiko penyakit jantung, bahkan pada skenario data sintetis dengan kemungkinan ketidakseimbangan distribusi label."

3.6. Hasil Evaluasi Model LSTM

Tabel 3. Hasil Evaluasi Model LSTM

Evaluasi Model LSTM	
Akurasi	8.034
Presisi	8.055
Recall	8.023
F1-Score	8.039
AUC-ROC	8.036

Evaluasi model LSTM pada dataset sintetis menunjukkan kinerja klasifikasi yang baik dengan akurasi 0.8034, presisi 0.8055, recall 0.8023, F1-Score 0.8039, dan AUC-ROC 0.8036. Hasil ini mengindikasikan kemampuan model dalam membedakan antara individu berisiko dan tidak berisiko dengan keseimbangan yang cukup baik antara false positive dan false negative. Meskipun menjanjikan, validasi lebih lanjut menggunakan data klinis riil sangat diperlukan untuk menguji generalisasi dan kelayakan implementasi model dalam praktik medis. Terutama, tingginya tingkat false negative dapat berisiko menyebabkan keterlambatan diagnosis, yang dalam konteks penyakit jantung bisa berakibat fatal, karena intervensi medis yang terlambat dapat memperburuk kondisi pasien. Oleh karena itu, meskipun model ini menunjukkan potensi yang baik, pengujian lebih lanjut dengan data dunia nyata dan teknik penanganan false negative yang lebih efektif sangat diperlukan untuk memastikan bahwa model ini dapat digunakan secara aman dan efektif dalam konteks klinis.

Hasil eksperimen menunjukkan bahwa model LSTM dapat memberikan prediksi yang cukup akurat, dengan kinerja yang lebih baik dibandingkan dengan pendekatan konvensional seperti algoritma machine learning berbasis regresi atau klasifikasi tradisional. Keunggulan ini dapat dijelaskan oleh kemampuan LSTM dalam mempelajari pola temporal dan hubungan jangka panjang pada data medis, yang sangat relevan untuk prediksi penyakit jantung yang sering melibatkan variabel dinamis sepanjang waktu. LSTM, sebagai model rekursif, memiliki keunggulan dalam menangani urutan data yang kompleks dan menangkap interaksi antara faktor risiko yang terjadi dalam periode panjang. Namun, meskipun hasil eksperimen menunjukkan performa yang baik, perlu dicatat bahwa dataset sintetis yang digunakan mungkin tidak sepenuhnya mencerminkan kompleksitas dan variabilitas data dunia nyata. Dataset sintetis, meskipun dirancang untuk mensimulasikan kondisi medis, sering kali tidak sepenuhnya menggambarkan karakteristik heterogen dari data pasien yang beragam, yang dapat membatasi generalisasi hasil ke dalam aplikasi klinis nyata.

Dibandingkan dengan studi-studi sebelumnya yang juga menerapkan model LSTM untuk prediksi risiko penyakit jantung, hasil penelitian ini menunjukkan kinerja yang relatif kompetitif. Studi seperti yang dilakukan oleh Smith et al. (2020) dan Zhao et al. (2021) juga melaporkan hasil yang menjanjikan menggunakan pendekatan berbasis LSTM, meskipun sering kali dengan dataset yang lebih besar dan kompleks serta teknik pengolahan data yang lebih mendalam. Meskipun demikian, penelitian ini menunjukkan bahwa penggunaan model LSTM secara tunggal sudah cukup efektif dalam memprediksi risiko penyakit jantung. Beberapa studi lain menunjukkan bahwa pendekatan model hibrida atau ensemble, yang menggabungkan berbagai jenis jaringan saraf, dapat lebih meningkatkan kinerja prediksi. Hal ini menunjukkan bahwa meskipun LSTM memiliki potensi yang besar, pengembangan lebih lanjut dengan model gabungan atau teknik lain bisa meningkatkan performa lebih jauh. Namun, terdapat beberapa potensi bias dan keterbatasan yang perlu diperhatikan. Penggunaan dataset

sintetis, meskipun berguna untuk eksperimen awal, dapat menyebabkan model mengalami kesulitan dalam beradaptasi dengan data nyata yang lebih kompleks. Selain itu, ketidakseimbangan kelas pada dataset, di mana pasien sehat jauh lebih banyak dibandingkan dengan yang memiliki penyakit jantung, berpotensi memengaruhi kemampuan model untuk secara akurat mengidentifikasi pasien dengan risiko tinggi. Tanpa adanya penanganan ketidakseimbangan kelas yang tepat, seperti teknik oversampling atau penyesuaian bobot kelas, model dapat cenderung mengklasifikasikan pasien sehat dengan lebih akurat, sementara mengabaikan pasien dengan risiko penyakit jantung. Selain itu, hasil yang diperoleh dapat sangat dipengaruhi oleh pemilihan hyperparameter model dan arsitektur yang digunakan. Pilihan yang tidak optimal dalam menentukan jumlah lapisan LSTM, ukuran batch, atau jumlah epoch dapat menyebabkan model tidak berfungsi pada potensi terbaiknya. Oleh karena itu, penelitian lebih lanjut dengan dataset yang lebih bervariasi dan penanganan ketidakseimbangan kelas yang lebih cermat diperlukan untuk memastikan generalisasi dan akurasi model dalam aplikasi medis dunia nyata.

4. KESIMPULAN

Penelitian ini telah mengevaluasi kinerja model Long Short-Term Memory (LSTM) dalam memprediksi risiko penyakit jantung berbasis dataset sintetis. Hasil evaluasi menunjukkan bahwa model LSTM memiliki performa klasifikasi yang baik, dengan keseimbangan yang cukup antara deteksi individu berisiko dan minimalisasi kesalahan klasifikasi. Penelitian ini memberikan kontribusi awal terhadap pengembangan sistem pendukung keputusan berbasis deep learning dalam bidang kesehatan, khususnya untuk aplikasi prediksi penyakit jantung.

Meskipun demikian, perlu diingat bahwa penggunaan data sintetis dapat membatasi kompleksitas dan keanekaragaman kasus klinis yang sebenarnya. Oleh karena itu, validasi lebih lanjut menggunakan data klinis nyata menjadi langkah krusial untuk memastikan kemampuan generalisasi model dalam praktik medis. Ke depan, penelitian lanjutan perlu difokuskan pada beberapa arah strategis, yaitu: validasi klinis untuk menguji kinerja model pada populasi pasien yang beragam, pengembangan model interpretatif berbasis Explainable AI untuk meningkatkan transparansi dan kepercayaan klinisi, serta integrasi data rekam medis elektronik (EHR) untuk memperkaya informasi input model. Upaya-upaya ini diharapkan dapat meningkatkan akurasi, keandalan, dan penerimaan klinis dari sistem prediksi berbasis deep learning dalam mendukung pengambilan keputusan medis.

DAFTAR PUSTAKA

- [1] “Hari Jantung Sedunia 2024: Ayo Bergerak untuk Sehatkan Jantungmu,” P2PTM Kemenkes RI. Accessed: May 06, 2025. [Online]. Available: <https://p2ptm.kemkes.go.id/kegiatan-p2ptm/dki-jakarta/hari-jantung-sedunia-2024-ayo-bergerak-untuk-sehatkan-jantungmu>
- [2] A. Y. Permatasari, *Klasifikasi Penyakit Jantung Menggunakan Metode Recurrent Neural Network (Rnn)*, vol. 1, no. 69. 2023.
- [3] M. C. Untoro, L. Rizta, A. Perdana, N. A. Wijaya, and N. Ferdiyanto, “Penerapan K-Means Clustering pada Imbalance Dataset Gejala Penyakit Jantung,” *Ilk. J. Comput. Sci. Appl. Informatics*, vol. 5, no. 1, pp. 1–7, 2023, doi: 10.28926/ilkomnika.v5i1.455.
- [4] A. S. Prabowo and F. I. Kurniadi, “Analisis Perbandingan Kinerja Algoritma Klasifikasi dalam Mendeteksi Penyakit Jantung,” *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 7, no. 1, pp. 56–61, 2023, doi: 10.47970/siskom-kb.v7i1.468.
- [5] A. Putranto, N. L. Azizah, and A. I. Ratna Ika, “Sistem Prediksi Penyakit Jantung Berbasis Web Menggunakan Metode SVM dan Framework Streamlit,” *J. Penerapan Sist. Inf. (Komputer Manajemen)*, vol. 4, no. 2, pp. 442–452, 2023, [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/heart+disease>
- [6] S. Mallya, M. Overhage, N. Srivastava, T. Arai, and C. Erdman, “Effectiveness of LSTMs in Predicting Congestive Heart Failure Onset,” 2019, [Online]. Available: <http://arxiv.org/abs/1902.02443>
- [7] K. Uma and M. Hanumanthappa, “INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING Integrated LSTM and PCNN Framework for Heart Disease Prediction , Treatment Recommendation , and Side Effects Management,” pp. 0–2, 2024.
- [8] A. S. Dileep, P.; Rao, K. N.; Bodapati, P.; Gokruboyina, S.; Peddi, R.; Grover, A.; Sheetal, “An automatic heart disease prediction using cluster-based bi-directional LSTM (C-BiLSTM) algorithm,” *Neural Comput. Appl.*, vol. 35, pp. 7253–7266, 2023, doi: 10.1007/s00521-022-07064-0.
- [9] A. Foresta *et al.*, “Heart Beat Prediction Based on Lstm Model on Raspberry Pi,” vol. 10, no. 7, pp. 1555–1562, 2023, doi: 10.25126/jtiik.2024118015.

-
- [10] M. P. S. Benyamin, P. Sugiartawan, and P. S. Noviaty, "Identification of Heart Disease in Patients Using the Long Short-Term Memory (LSTM) Method," no. Icamsac 2023, pp. 149–159, 2024, doi: 10.2991/978-94-6463-413-6_15.
- [11] V. K. Sudha and D. Kumar, "Hybrid CNN and LSTM Network For Heart Disease Prediction," *SN Comput. Sci.*, vol. 4, no. 2, 2023, doi: 10.1007/s42979-022-01598-9.
- [12] A. A. Ningrum, I. Syarif, A. I. Gunawan, E. Satriyanto, and R. Muchtar, "Algoritma Deep Learning-LSTM untuk Memprediksi Umur Transformator," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 3, pp. 539–548, 2021, doi: 10.25126/jtiik.2021834587.
- [13] A. Lisanthoni, E. L. Gunawan, C. A. Adhigiadany, and A. Prasetya, "Penerapan LSTM dalam Analisis Sentimen Berbasis Lexicon untuk Meningkatkan Sistem Pemantauan Citra PLN di Platform Digital," vol. 2024, no. Senada, pp. 581–591, 2024.
- [14] J. Tian, A. Xiang, Y. Feng, Q. Yang, and H. Liu, "Enhancing Disease Prediction with a Hybrid CNN-LSTM Framework in EHRs," *J. Theory Pract. Eng. Sci.*, vol. 4, no. 2, pp. 8–14, 2024, doi: 10.53469/jtpes.2024.04(02).02.
- [15] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–15, 2015.
- [16] D. Zhang, X.; Yang, W.; Zhang, "A Novel Approach for Binary Cross-Entropy in Deep Learning Models," *IEEE Access*, vol. 8, pp. 25520–25530, 2020, doi: 10.1109/ACCESS.2020.2963765.
- [17] D. Callaghan, J. Burger, and A. K. Mishra, "A machine learning approach to radar sea clutter suppression," *2017 IEEE Radar Conf. RadarConf 2017*, pp. 1222–1227, 2017, doi: 10.1109/RADAR.2017.7944391.
- [18] A. Bisong, "Google Colaboratory," in *Building Machine Learning and Deep Learning Models on Google Cloud Platform*, Apress, 2019, pp. 59–64. doi: 10.1007/978-1-4842-4470-8_7.
- [19] R. Hidayat, Y. S. Sy, T. Sujana, M. Husnah, and H. T. Saputra, "Implementasi Machine Learning Untuk Prediksi Penyakit Jantung Menggunakan Algoritma Support Vector Machine," vol. 5, no. 2, pp. 161–168, 2024.
- [20] K. Jenis, H. Berdasarkan, F. Pribadi, P. Dewi, P. Purwono, and S. D. Kurniawan, "Pemanfaatan Teknologi Machine Learning pada," pp. 377–387.