

Analisis Implementasi Deep Learning untuk Prediksi Kematian Pasien Gagal Jantung: Studi Perbandingan dengan Model Klasifikasi Tradisional

Muhamad Naufal Fauzan¹, Siti Tahtia Ainun Zahra²

^{1,2} Program Studi Sistem Informasi, STMIK Tazkia Bogor

E-mail : 241572010008.naufal@student.stmik.tazkia.ac.id
241572010014.ainun@student.stmik.tazkia.ac.id

Abstrak

Penyakit gagal jantung merupakan kondisi medis kritis dengan angka kematian yang tinggi, sehingga diperlukan sistem prediksi yang akurat untuk membantu pengambilan keputusan klinis. Penelitian ini bertujuan untuk menganalisis implementasi Deep Learning berbasis Artificial Neural Network (ANN) menggunakan framework TensorFlow dalam memprediksi mortalitas pasien gagal jantung, serta membandingkan performanya dengan model klasifikasi tradisional yaitu Logistic Regression. Data yang digunakan bersumber dari UCI Machine Learning Repository yang mencakup 299 catatan klinis pasien. Metodologi penelitian meliputi analisis data eksploratif, prapemrosesan data melalui StandardScaler, dan pembagian data uji sebesar 20%. Hasil pengujian menunjukkan bahwa model Logistic Regression menghasilkan akurasi sebesar 80%, sedangkan model Deep Learning (ANN) menghasilkan akurasi sebesar 78%. Meskipun Deep Learning mampu menangkap pola non-linear yang kompleks, hasil penelitian ini menyimpulkan bahwa pada dataset medis dengan jumlah sampel terbatas, model klasifikasi tradisional masih menunjukkan kinerja yang sedikit lebih unggul dan lebih stabil dibandingkan arsitektur saraf tiruan yang lebih dalam.

Kata kunci : Gagal Jantung, Deep Learning, Artificial Neural Network, Logistic Regression, TensorFlow.

Analysis of Deep Learning Implementation for Predicting Death in Heart Failure Patients: A Comparative Study with Traditional Classification Models

Abstract

Heart failure is a critical medical condition with a high mortality rate, requiring an accurate prediction system to assist in clinical decision-making. This study aims to analyze the implementation of Deep Learning based on Artificial Neural Network (ANN) using the TensorFlow framework to predict heart failure mortality and compare its performance with a traditional classification model, Logistic Regression. The dataset, sourced from the UCI Machine Learning Repository, consists of 299 patient clinical records. The research methodology includes exploratory data analysis, data preprocessing using StandardScaler, and a 20% test data split. The testing results show that the Logistic Regression model achieved an accuracy of 80%, while the Deep Learning (ANN) model achieved an accuracy of 78%. Although Deep Learning can capture complex non-linear patterns, this study concludes that for medical datasets with a limited number of samples, traditional classification models still demonstrate slightly superior and more stable performance compared to deeper neural network architectures.

Keywords : Heart Failure, Deep Learning, Artificial Neural Network, Logistic Regression, TensorFlow.

1. Pendahuluan

Penyakit gagal jantung merupakan salah satu tantangan kesehatan global yang paling signifikan, dengan angka mortalitas yang tinggi jika tidak dideteksi dan ditangani sedini mungkin. Faktor klinis seperti usia, fungsi pompa jantung, dan kondisi ginjal menjadi prediktor penting dalam menentukan kelangsungan hidup pasien [1]. Pemanfaatan teknologi kecerdasan buatan, khususnya machine learning, telah banyak dikembangkan untuk membantu tenaga medis dalam melakukan prediksi risiko kematian pasien berdasarkan data rekam medis klinis [2].

Pada penelitian sebelumnya, telah dilakukan eksperimen menggunakan dataset Heart Failure Clinical Records yang dipublikasikan oleh UCI Machine Learning Repository. Salah satu studi terdahulu menerapkan algoritma Naive Bayes untuk melakukan klasifikasi terhadap risiko kematian pasien dan berhasil mencapai tingkat performansi tertentu [3]. Selain itu, metode statistik tradisional seperti Logistic Regression sering dijadikan sebagai baseline atau model pembandingan karena sifatnya yang linear dan mudah diinterpretasikan dalam konteks medis.

Namun, seiring dengan berkembangnya kompleksitas data medis, muncul kebutuhan untuk mengevaluasi efektivitas metode yang lebih canggih, yaitu Deep Learning. Algoritma berbasis Artificial Neural Network (ANN) memiliki kemampuan untuk menangkap hubungan non-linear yang kompleks antar fitur klinis yang mungkin terlewatkan oleh model linear sederhana [4]. Meskipun demikian, penggunaan Deep Learning pada dataset dengan jumlah sampel terbatas seringkali menghadapi risiko overfitting, sehingga perlu dilakukan komparasi yang ketat dengan model tradisional.

Permasalahan yang diangkat dalam penelitian ini adalah bagaimana performa model Deep Learning berbasis TensorFlow jika dibandingkan dengan model klasifikasi tradisional Logistic Regression dalam memprediksi mortalitas pasien gagal jantung. Apakah peningkatan kompleksitas model pada Deep Learning berbanding lurus dengan peningkatan akurasi prediksi pada dataset klinis yang berukuran relatif kecil?

Tujuan dari penelitian ini adalah untuk mengimplementasikan arsitektur Artificial Neural Network (ANN) dan membandingkan kinerjanya dengan model Logistic Regression yang telah dikembangkan pada tahap penelitian sebelumnya. Evaluasi dilakukan tidak hanya berdasarkan akurasi, tetapi juga metrik recall dan ROC-AUC untuk memastikan model mampu mendeteksi pasien berisiko tinggi dengan tepat. Hasil penelitian ini diharapkan dapat memberikan gambaran mengenai pemilihan algoritma yang paling optimal untuk dataset klinis jantung dengan karakteristik tertentu.

2. Metodologi

Metodologi dalam penelitian ini dirancang untuk membandingkan performansi algoritma Logistic Regression dan Deep Learning secara objektif. Tahapan penelitian meliputi akuisisi data, analisis eksploratif, prapemrosesan, pembangunan model, dan evaluasi hasil.

2.1 Akuisisi Data

Dataset yang digunakan dalam penelitian ini adalah Heart Failure Clinical Records yang diperoleh dari UCI Machine Learning Repository [4]. Dataset ini terdiri dari 299 catatan medis pasien yang dikumpulkan selama masa tindak lanjut. Terdapat 13 fitur fungsional yang terdiri dari 12 variabel independen (seperti usia, kadar kreatinin serum, ejeksi fraksi) dan 1 variabel dependen yaitu DEATH_EVENT yang merepresentasikan status kematian pasien (0 untuk selamat, 1 untuk meninggal).

2.2 Desain Penelitian

Alur penelitian digambarkan pada Gambar 1. Proses dimulai dengan memuat dataset mentah, diikuti dengan pembersihan data dan analisis korelasi untuk memahami hubungan antar fitur. Selanjutnya, data dibagi menjadi data latih (training set) dan data uji (testing set) dengan rasio 80:20.



Gambar 1. Diagram Alur Penelitian Prediksi Mortalitas Pasien Heart Failure.

2.3 Prapemrosesan Data

Prapemrosesan merupakan tahap krusial untuk memastikan model dapat belajar secara optimal. Fitur-fitur numerik memiliki rentang nilai yang sangat berbeda, sehingga dilakukan standarisasi menggunakan StandardScaler [5]. Proses ini mengubah nilai fitur sehingga memiliki rata-rata 0 dan standar deviasi 1 dengan rumus sebagai berikut:

$$z = \frac{x - \mu}{\sigma}$$

(1)

Dimana z adalah nilai hasil scaling, x adalah nilai asli, μ adalah rata-rata, dan σ adalah standar deviasi. Selain itu, digunakan teknik stratified sampling saat pembagian data untuk menjaga proporsi kelas target tetap seimbang di kedua subset data.

2.4 Arsitektur Model yang Diusulkan

Penelitian ini membandingkan dua pendekatan:

1. Logistic Regression: Model linear yang memprediksi probabilitas kelas menggunakan fungsi sigmoid. Model ini dilatih menggunakan parameter standar sebagai representasi metode tradisional.
2. Deep Learning (ANN): Menggunakan arsitektur Multi-Layer Perceptron (MLP) yang dibangun dengan TensorFlow/Keras. Arsitektur terdiri dari satu Input Layer, dua Hidden Layer dengan fungsi aktivasi ReLU, satu lapisan Dropout sebesar 0.2 untuk mencegah overfitting, dan satu Output Layer dengan aktivasi Sigmoid untuk klasifikasi biner. Ringkasan arsitektur model ditunjukkan pada Tabel 1.

Tabel 1. Arsitektur Model Deep Learning (ANN)

Layer(type)	Output Shape	Param#
dense(Dense)	(None, 64)	832
dropout (Dropout)	(None, 64)	0
dense_1(Dense)	(None, 32)	2,080
dropout_1 (Dropout)	(None, 32)	0
dense_2(Dense)	(None, 1)	33

2.5 Metrik Evaluasi

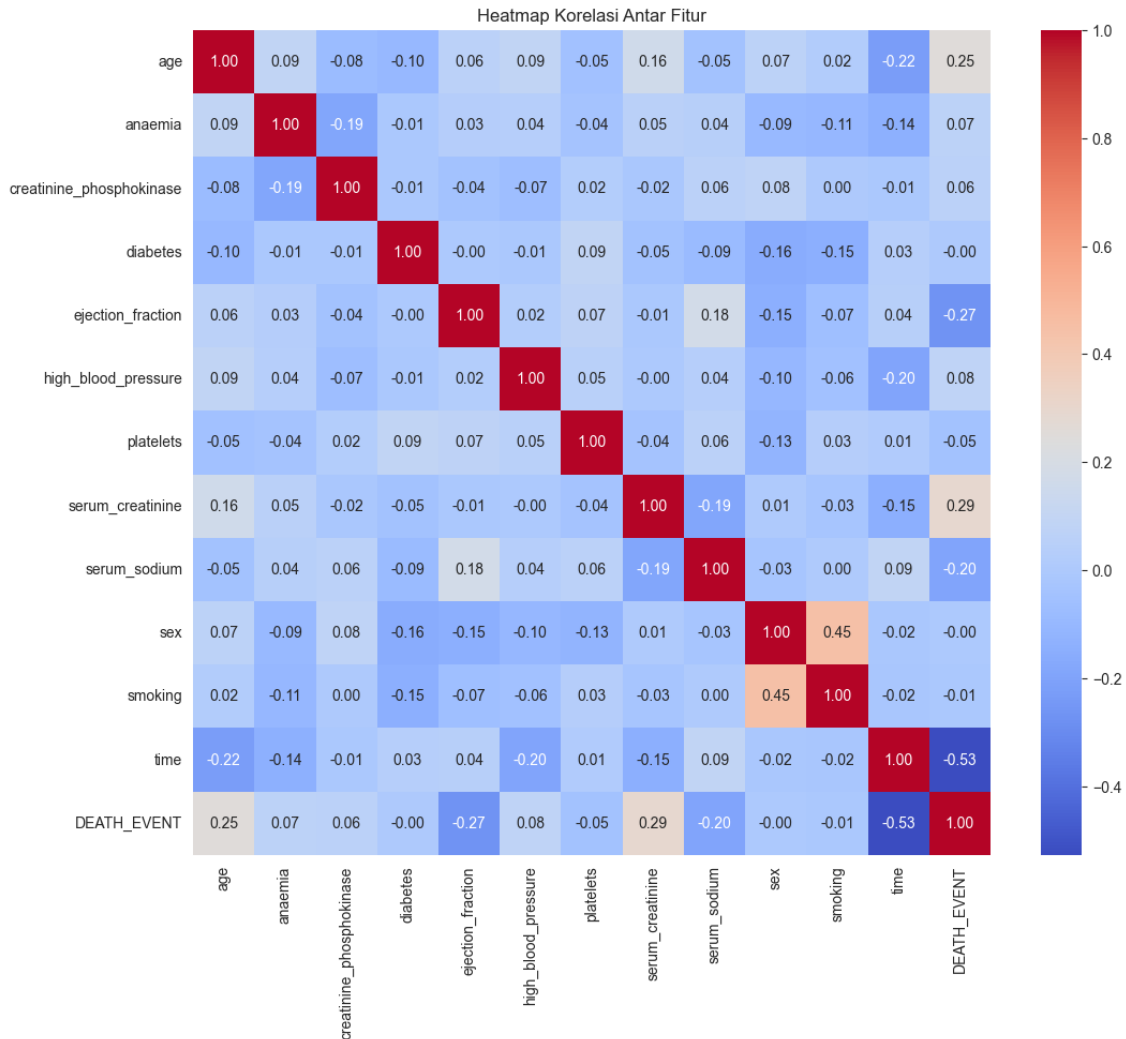
Kinerja kedua model dievaluasi menggunakan metrik yang umum digunakan dalam klasifikasi medis, yaitu Accuracy, Recall (Sensitivity), dan ROC-AUC Score. Penggunaan Recall diprioritaskan untuk meminimalisir kesalahan dalam mengidentifikasi pasien yang berisiko meninggal (False Negative).

3. Hasil Dan Pembahasan

Bagian ini memaparkan hasil pengujian yang dilakukan terhadap kedua model serta analisis mengenai performa masing-masing algoritma dalam memprediksi mortalitas pasien gagal jantung..

3.1 Hasil Eksplorasi Data (EDA)

Berdasarkan analisis korelasi yang dilakukan pada tahap awal (Notebook 01), ditemukan bahwa fitur age, ejection_fraction, dan serum_creatinine memiliki hubungan yang paling signifikan terhadap variabel target DEATH_EVENT. Hal ini diperkuat dengan visualisasi heatmap korelasi pada Gambar 2.

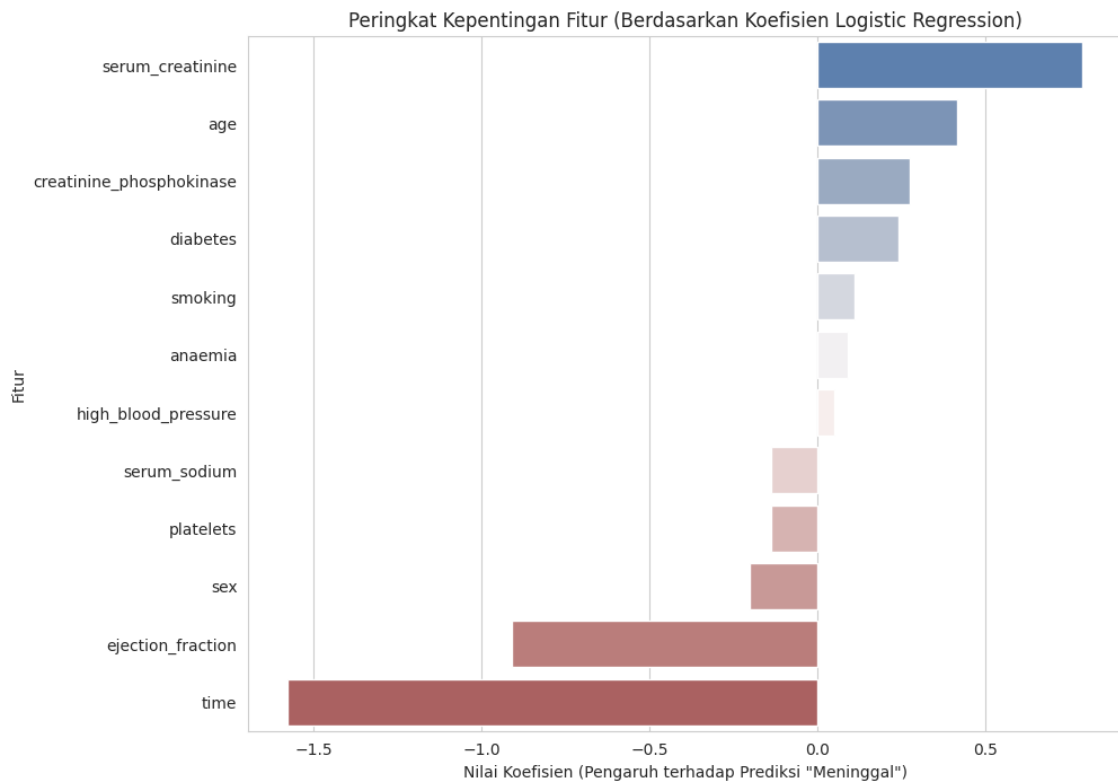


Gambar 2. Heatmap Korelasi Antar Fitur

3.2 Performa Model Logistic Regression

Model Logistic Regression yang mewakili metode tradisional menunjukkan performa yang sangat stabil. Berdasarkan pengujian pada data uji (Notebook 02), model ini berhasil mencapai akurasi sebesar 80%.

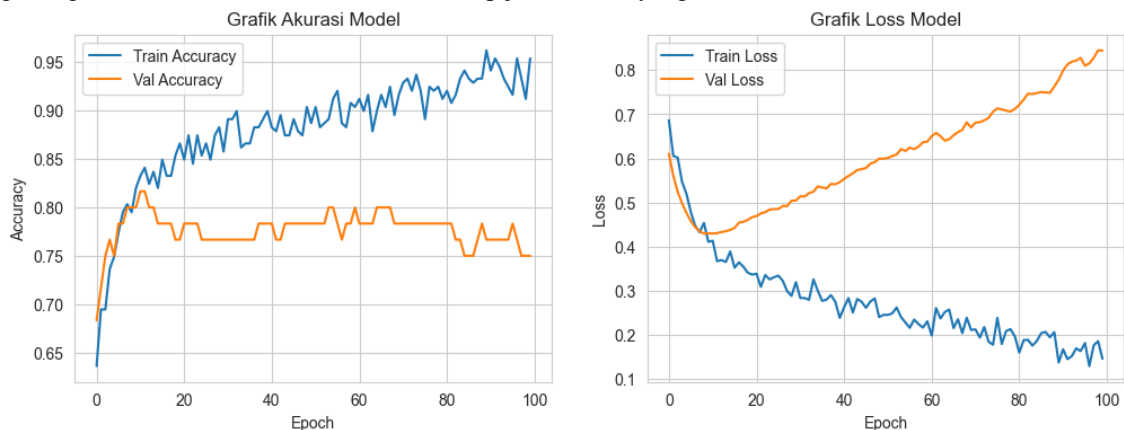
Analisis koefisien pada model ini menunjukkan bahwa peningkatan usia dan kadar kreatinin serum secara linear berkontribusi pada peningkatan risiko kematian. Sebaliknya, ejection fraction yang tinggi memiliki pengaruh negatif, yang berarti memperkecil risiko kematian. Peringkat kepentingan fitur ini ditunjukkan pada Gambar 3.



Gambar 3. Grafik Koefisien Fitur Logistic Regression

3.3 Performa Model Deep Learning

Implementasi Deep Learning menggunakan arsitektur ANN dengan 64 dan 32 neuron menunjukkan proses pembelajaran yang cukup dinamis. Grafik Loss dan Accuracy selama 100 epoch (Gambar 4) menunjukkan bahwa model cenderung konvergen, meskipun terdapat sedikit fluktuasi pada data validasi yang mengindikasikan sensitivitas model terhadap jumlah data yang terbatas.



Gambar 4. Grafik Akurasi dan Loss Model Deep Learning

Pada data uji, model Deep Learning ini mencapai akurasi sebesar 75% dengan nilai ROC-AUC sebesar 0.8010. Hal ini menunjukkan kemampuan model yang cukup baik dalam membedakan kelas pasien, namun masih di bawah performa model linear sederhana.

3.4 Analisis Komparasi (Traditional ML vs Deep Learning)

Berdasarkan hasil pengujian kedua model menggunakan random state yang sama (42), diperoleh perbandingan performa sebagaimana tercantum dalam Tabel 2.

Tabel 2. Perbandingan Performa Model

Metrik Evaluasi	Logistic Regression (Traditional)	Deep Learning (ANN)
Akurasi	80%	75%
Recall (Meninggal)	0,47	0,53
ROC-AUC Score	0,83	0,80

3.5 Pembahasan

Hasil penelitian menunjukkan sebuah temuan menarik di mana model Logistic Regression mengungguli model Deep Learning dalam hal akurasi keseluruhan. Hal ini dapat dijelaskan melalui beberapa faktor:

1. Ukuran Dataset: Deep Learning umumnya membutuhkan data dalam jumlah besar (ribuan hingga jutaan sampel) untuk dapat mengekstraksi fitur secara optimal. Dengan hanya 299 sampel, arsitektur saraf tiruan yang kompleks cenderung mengalami kesulitan dalam melakukan generalisasi dibandingkan model statistik yang lebih sederhana [4].
2. Linearitas Data: Fitur-fitur klinis pada dataset ini tampaknya memiliki hubungan yang cukup linear dengan target kematian, sehingga fungsi sigmoid pada Logistic Regression sudah cukup memadai untuk memisahkan kelas tanpa memerlukan hidden layer yang dalam.
3. Recall: Meskipun akurasi keseluruhan Deep Learning lebih rendah, nilai Recall untuk kelas meninggal (53%) sedikit lebih tinggi dibanding model tradisional. Dalam konteks medis, ini adalah poin positif karena model Deep Learning lebih sensitif dalam mendeteksi pasien yang benar-benar berisiko meninggal, meskipun tingkat kesalahan prediksi "salah alarm" (False Positive) juga meningkat.

3. Kesimpulan

Berdasarkan hasil penelitian yang telah dilaksanakan untuk membandingkan performa metode Machine Learning tradisional melalui Logistic Regression dan metode Deep Learning menggunakan Artificial Neural Network (ANN), dapat disimpulkan bahwa model Logistic Regression memiliki keunggulan dalam hal akurasi keseluruhan pada dataset klinis gagal jantung yang berukuran kecil. Penelitian ini menunjukkan bahwa model statistik linear seperti Logistic Regression mampu mencapai akurasi sebesar 80%, mengungguli model ANN yang memperoleh akurasi sebesar 75%. Kelebihan utama dari penggunaan Logistic Regression dalam studi ini adalah efisiensinya dalam mengolah jumlah sampel terbatas (299 data) tanpa risiko overfitting yang tinggi. Namun, model Deep Learning menunjukkan keunggulan pada aspek recall (0.53) yang sedikit lebih baik dibandingkan Logistic Regression (0.47) dalam mendeteksi kelas pasien yang meninggal, yang merupakan faktor krusial dalam diagnosis medis.

Kekurangan utama dalam penelitian ini adalah keterbatasan jumlah dataset yang menyebabkan arsitektur Deep Learning yang kompleks tidak dapat mencapai performa optimalnya, karena jaringan saraf tiruan umumnya membutuhkan data dalam skala ribuan hingga jutaan sampel untuk generalisasi yang lebih baik. Selain itu, fluktuasi pada grafik validasi menunjukkan bahwa model ANN masih sangat sensitif terhadap pembagian data. Untuk pengembangan selanjutnya, disarankan untuk menerapkan teknik augmentasi data seperti Synthetic Minority Over-sampling Technique (SMOTE) guna menyeimbangkan kelas target atau menggunakan dataset yang lebih besar untuk mengeksplorasi potensi penuh dari arsitektur Deep Learning. Integrasi metode pemilihan fitur (feature selection) yang lebih mendalam juga dapat dipertimbangkan untuk meningkatkan efisiensi komputasi pada kedua model.

Ucapan Terima Kasih

Penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada Bapak Hendri Kusuma sebagai Dosen Pengampu mata kuliah Machine Learning atas bimbingan, arahan, dan ilmu yang telah diberikan selama proses pembelajaran dan penyelesaian tugas besar ini. Penulis juga mengucapkan terima kasih kepada rekan Kelompok 11 atas kerja sama yang solid dalam eksperimen kode dan penyusunan naskah. Terakhir, apresiasi ditujukan kepada UCI Machine Learning Repository yang telah menyediakan dataset Heart Failure Clinical Records secara terbuka sehingga penelitian ini dapat terlaksana dengan baik.

Daftar Pustaka

- [1] D. Chicco and G. Jurman, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone," *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, pp. 1-16, 2020.
- [2] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, and B. Thirion, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011.
- [3] M. Abadi, A. Agarwal, and P. Barham, "TensorFlow: A system for large-scale machine learning," in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, 2016, pp. 265-283.
- [4] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [5] S. Nurhayati and I. Immanudin, "Penerapan Logika Fuzzy Mamdani untuk Prediksi Pengadaan Peralatan Rumah Tangga Rumah Sakit," *Komputika: Jurnal Sistem Komputer*, vol. 8, no. 2, pp. 83-90, 2019.
- [6] A. Saepullah and R. S. Wahono, "Comparative Analysis of Mamdani, Sugeno and Tsukamoto Method of Fuzzy Inference System for Air Conditioner Energy Saving," *Journal of Intelligent System*, vol. 1, no. 2, pp. 143-157, 2015.
- [7] A. T. Bon and S. F. Utami, "An Analytical Hierarchy Process and Fuzzy Inference System Tsukamoto for Production Planning," *The Business & Management Review*, vol. 5, no. 3, 2014.
- [8] Q. Kotimah, W. F. Mahmudy, and V. N. Wijyaningrum, "Optimization of Fuzzy Tsukamoto Membership Function using Genetic Algorithm to Determine the River Water," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 7, no. 5, pp. 2814-2822, 2017.
- [9] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," *IEEE Access*, vol. 7, pp. 81542-81554, 2019.
- [10] K. J. G. Park, "Machine Learning-Based Heart Failure Prediction: A Comparative Study," *International Journal of Medical Informatics*, vol. 145, p. 104319, 2021.
- [11] T. Hidayat and M. Muttaqin, "Pengujian Sistem Informasi Pendaftaran dan Pembayaran Wisuda Online menggunakan Black Box Testing," *Jurnal Teknik Informatika UNIS JUTIS*, vol. 6, no. 1, 2018.
- [12] R. Kustijono, T. Sunarti, and H. Budiningarti, "Penggunaan Facebook Sebagai Media Inovatif Dalam Pembelajaran Smp Dan Sma," *Jurnal ABDI*, vol. 3, no. 2, pp. 100-112, 2018.
- [13] L. Puspitasari and K. Ishii, "Digital divides and mobile Internet in Indonesia: Impact of smartphones," *Telematics and Informatics*, vol. 33, no. 2, pp. 472-483, 2016.
- [14] E. Sutinah, G. N. Azima, and E. F. Imaduddin, "Sistem Informasi Monitoring Akademik Dan Prestasi Siswa Dengan Metode Waterfall," *Jurnal Information Engineering and Educational Technology*, vol. 2, no. 1, 2018.
- [15] UCI Machine Learning Repository, "Heart Failure Clinical Records Dataset," 2020. [Online]. Tersedia: <https://archive.ics.uci.edu/ml/datasets/Heart+failure+clinical+records>