

# Perancangan Prototipe Aplikasi Prediksi Kematian Akibat Gagal Jantung Menggunakan Metode *Machine Learning* Berdasarkan Data *Heart Failure Clinical Records*

<sup>1</sup>Jumardin, <sup>2</sup>Handrie Noprisson

<sup>1</sup>Fakultas Teknik dan Informatika, Universitas Dian Nusantara, Indonesia

<sup>1</sup>411202009@mahasiswa.undira.ac.id; <sup>2</sup>handrie.noprisson@dosen.undira.ac.id

## Article Info

### Article history:

Received, 2024-10-22

Revised, 2024-11-25

Accepted, 2024-11-30

### Kata Kunci:

Gagal jantung

Random forest

Logistic regression

K-Nearest neighbors

Prediksi

### Keywords:

Heart failure

Random forest

Logistic regression

K-nearest neighbors

Prediction

## ABSTRAK

Penelitian ini bertujuan untuk mengembangkan prototipe Aplikasi Prediksi Kematian Akibat Gagal Jantung menggunakan metode machine learning berbasis data klinis *heart failure clinical records*. Aplikasi ini memanfaatkan data klinis pasien, seperti usia, tekanan darah, fraksi ejeksi, kadar kreatinin, dan atribut lainnya, untuk membangun model prediksi risiko kematian. Dalam pengembangan aplikasi, digunakan beberapa algoritma machine learning, yaitu Random Forest, Logistic Regression, dan K-Nearest Neighbors (KNN), untuk memodelkan dan menganalisis data. Dataset yang digunakan dalam penelitian ini mencakup 299 catatan klinis pasien dengan 13 kolom atribut. Atribut target adalah Death Event/Kematian, sedangkan atribut lainnya, seperti usia, jenis kelamin, riwayat penyakit (anemia, diabetes, tekanan darah tinggi), serta hasil tes laboratorium (kreatinin, sodium, dan fraksi ejeksi), digunakan sebagai prediktor. Aplikasi ini dilengkapi dengan beberapa menu utama untuk mendukung fungsionalitasnya, seperti dashboard, yang menyajikan ringkasan informasi statistik prediksi dan laporan terkait, serta blog / berita untuk edukasi kesehatan jantung. Menu data master memungkinkan pengelolaan data pendukung, sedangkan menu diagnosa digunakan untuk melakukan prediksi berdasarkan input data pasien. Menu riwayat diagnosa menyimpan hasil prediksi sebelumnya, sementara data pasien dan data admin memfasilitasi pengelolaan informasi pengguna.

## ABSTRACT

*his research aims to develop a prototype of the Heart Failure Death Prediction Application using machine learning methods based on clinical data from the Heart Failure Clinical Records. The application utilizes clinical patient data, such as age, blood pressure, ejection fraction, creatinine levels, and other attributes, to build a predictive model for mortality risk. Several machine learning algorithms, including Random Forest, Logistic Regression, and K-Nearest Neighbors (KNN), were employed to model and analyze the data. The dataset used in this study consists of 299 clinical records with 13 attribute columns. The target attribute is Death Event, while other attributes, such as age, gender, medical history (anemia, diabetes, high blood pressure), and laboratory test results (creatinine, sodium, and ejection fraction), were used as predictors. The application is equipped with several main menus to support its functionality, such as the Dashboard, which provides a summary of statistical prediction information and related reports, and Blog/News, which offers heart health education. The Data Master menu allows for the management of supporting data, while the Diagnosis menu is used to perform predictions based on patient input data. The Diagnosis History menu stores previous prediction results, while the Patient Data menu facilitates the management of patient information.*

*This is an open access article under the CC BY-SA license.*



### Penulis Korespondensi:

Handrie Noprisson,

Fakultas Teknik dan Informatika,

Universitas Dian Nusantara, Indonesia

Email: handrie.noprisson@dosen.undira.ac.id

## 1. PENDAHULUAN

Gagal jantung adalah kondisi di mana jantung tidak dapat memompa darah dengan cukup untuk memenuhi kebutuhan tubuh. Penyakit ini masih menjadi ancaman global dan menjadi penyebab utama kematian nomor satu di dunia. Menurut data dari Organisasi Kesehatan Dunia (WHO), lebih dari 17 juta kematian di dunia disebabkan oleh penyakit jantung [1], [2]. Oleh karena itu, upaya untuk memprediksi risiko kematian akibat gagal jantung menjadi langkah awal yang penting dalam mencegah komplikasi lebih lanjut, meningkatkan perawatan, dan mengoptimalkan penanganan pasien. Prediksi risiko ini membantu dokter dalam merancang rencana perawatan yang tepat waktu, sehingga dapat mengurangi risiko komplikasi berbahaya. Dalam bidang kesehatan, keakuratan prediksi penyakit sangatlah krusial, karena membutuhkan analisis yang cermat dan keputusan yang tepat untuk menentukan diagnosis serta perawatan pasien [3]–[5].

Teknologi berperan untuk kemajuan bidang, termasuk bidang kesehatan dan pendidikan [6]–[12]. Saat ini, proses diagnosis dan prediksi risiko kematian gagal jantung masih terbatas pada penilaian klinis yang tergantung pada pengalaman dokter dan alat evaluasi tradisional seperti skor risiko. Oleh karena itu, penggunaan *machine learning* untuk menganalisis data klinis dapat memberikan sebuah informasi yang lebih mendalam dan prediksi yang lebih tepat serta akurat.

Model yang digunakan yaitu Logistic Regression, K-Nearest Neighbors (KNN), dan Random Forest. Model Logistic Regression dipilih karena memiliki nilai akurasi tertinggi pada beberapa penelitian yang ada Prinsip kerja KNN yaitu untuk mencari jarak terpendek antara k dengan yang terdekat pada data evaluasi dan data pelatihan. Random Forest menggabungkan beberapa klasifikasi yang dibuat oleh decision tree, serta melatih model untuk menentukan dimensi ruang fitur acak, tetapi juga jumlah prediksi yang akan dibuat berdasarkan beberapa keputusan yang dihasilkan oleh data [5], [13]–[15].

Penelitian ini memanfaatkan data klinis, seperti heart failure clinical records, yang mencakup informasi penting tentang profil pasien, riwayat medis, hasil tes diagnostik, dan tindakan medis yang telah dilakukan sebelumnya. Data yang telah diprediksi akan diproses dalam bentuk aplikasi prediksi kematian diakibatkan gagal jantung.

## 2. METODE PENELITIAN

Penelitian ini menggunakan data dari Heart Failure Clinical Records sebagai studi kasus untuk menguji efektivitas model prediktif yang dikembangkan. Dataset yang digunakan adalah *Heart Failure Prediction* dari rumah sakit yang terlibat meliputi Hungarian Institute of Cardiology, University Hospital Zurich, University Hospital Basel, V.A. Medical Center Long Beach, dan Cleveland Clinic Foundation [16]. Dataset ini berisi berbagai faktor yang berkontribusi terhadap kematian akibat gagal jantung, dengan 13 kolom atribut dan 299 baris data. Atribut yang menjadi target adalah Death Event/Kematian, sementara atribut lainnya akan digunakan sebagai prediktor. Dataset ini telah banyak digunakan para peneliti. Dataset berisi catatan klinis dari 299 pasien mencakup informasi seperti usia, jenis kelamin, riwayat penyakit (seperti anemia, diabetes, tekanan darah tinggi), hasil tes laboratorium (seperti kreatinin, sodium, dan fraksi ejeksi), dan status kematian seperti yang terlihat pada **Tabel 1**.

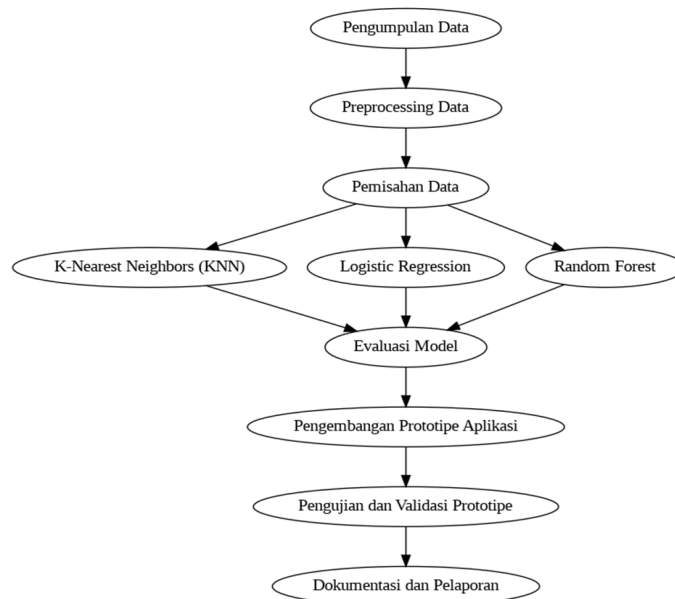
Tabel 1 Dataset Penelitian

| No  | Age | Anemia | Creatinine Phosphokinase | Diabetes | Ejection Fraction | ... | Sex | Smoking | Time | Death Event |
|-----|-----|--------|--------------------------|----------|-------------------|-----|-----|---------|------|-------------|
| 0   | 75  | 0      | 582                      | 0        | 20                | ... | 1   | 0       | 4    | 1           |
| 1   | 55  | 0      | 7861                     | 0        | 38                | ..  | 1   | 0       | 6    | 1           |
| ... | ... | ...    | ...                      | ...      | ...               | ..  | ... | ...     | ...  | ...         |
| 297 | 45  | 0      | 2413                     | 0        | 38                | ..  | 1   | 1       | 280  | 0           |
| 298 | 50  | 0      | 196                      | 0        | 45                | ..  | 1   | 1       | 285  | 0           |

Tahapan selanjutnya adalah validasi data yang bertujuan memastikan data yang digunakan dalam analisis valid, lengkap, konsisten, dan dapat diandalkan. Proses ini mencakup pemeriksaan atribut penting seperti usia, kondisi medis (anemia, diabetes, hipertensi), dan gaya hidup (merokok). Tahap *preprocessing data* dimulai dengan mengumpulkan data relevan dari berbagai sumber, memilih atribut penting, serta melakukan pembersihan data untuk mengatasi nilai yang hilang dan inkonsistensi. Pemodelan data dilakukan untuk memprediksi kematian akibat gagal jantung berdasarkan *heart failure clinical records* menggunakan algoritma seperti K-Nearest Neighbors (KNN), Logistic Regression, dan Random Forest. Model ini membantu mengidentifikasi faktor risiko utama yang berkontribusi terhadap kematian.

Setelah tahapan validasi, preprocessing, pemodelan, dan evaluasi data, langkah selanjutnya adalah perancangan prototipe aplikasi berbasis machine learning untuk prediksi kematian akibat gagal jantung. Prototipe ini menggunakan model terbaik dari evaluasi untuk memprediksi risiko secara real-time dengan

memanfaatkan data klinis seperti usia, kondisi medis, dan gaya hidup pasien. Prototipe dirancang untuk memberikan informasi prediktif yang akurat dan mudah digunakan oleh tenaga medis sebagai alat bantu pengambilan keputusan. Adapun tahapan lengkap dari penelitian ini dapat dilihat pada **Gambar 1**.

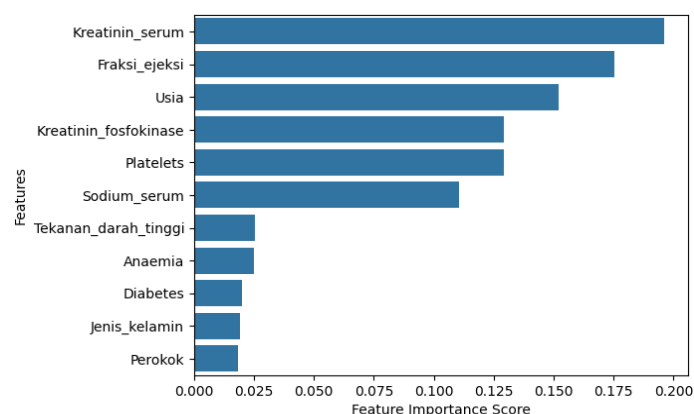


Gambar 1 Tahapan Penelitian

### 3. HASIL DAN ANALISIS

Dataset ini berisi faktor penyebab kematian akibat gagal jantung, dengan 13 kolom atribut dan 299 data. Atribut yang targetnya adalah “death event/kematian”, akan tetapi atribut lain digunakan sebagai predictor. Dataset ini berisi variabel-variabel terkait kondisi medis dan gaya hidup, seperti usia, anemia, kadar kreatinin fosfokinase, diabetes, fraksi ejeksi, tekanan darah tinggi, jumlah trombosit, kadar kreatinin serum, natrium serum, jenis kelamin, kebiasaan merokok, waktu, dan kejadian kematian. Dari dataset yang sudah dikumpulkan berdasarkan atribut yang sudah dijelaskan adalah penyakit jantung dengan nilai 1, dan tidak terdiagnosis penyakit jantung dengan nilai 0.

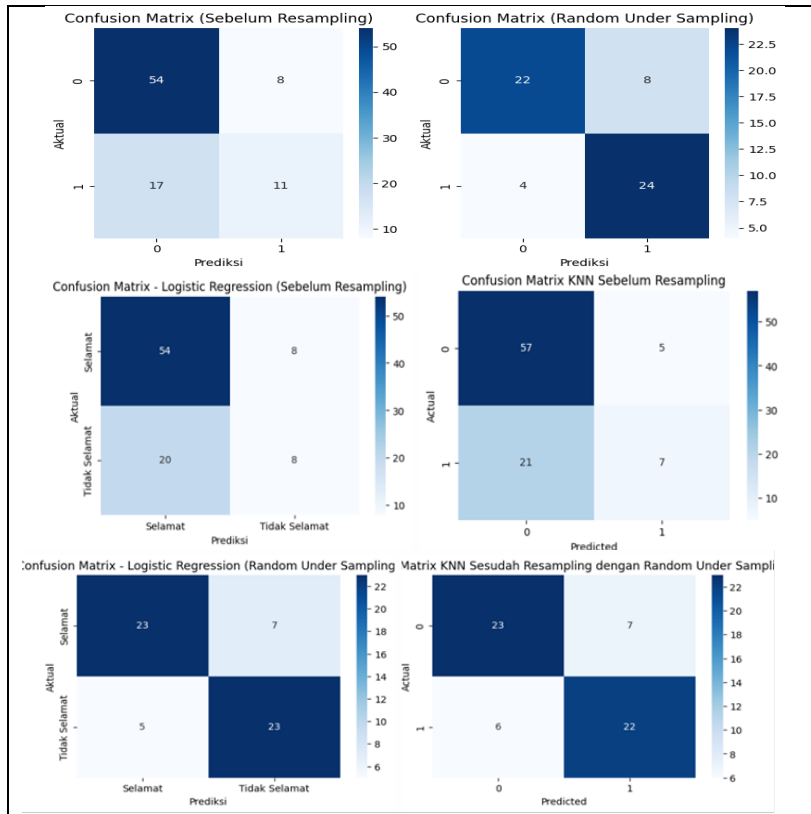
Proses preprocessing data yg tepat dan teliti sangat penting untuk membangun contoh machine learning yang akurat salah satunya *feature importance* digunakan untuk mengevaluasi kontribusi masing-masing fitur terhadap hasil prediksi dalam aplikasi prediksi kematian akibat gagal jantung, terutama pada model Logistic Regression, KNN, dan Random Forest. Analisis ini membantu memahami sejauh mana fitur tertentu memengaruhi performa model. Sebaran dari skor *feature importance* dapat dilihat pada **Gambar 2**.



Gambar 2 Skor Feature Importance

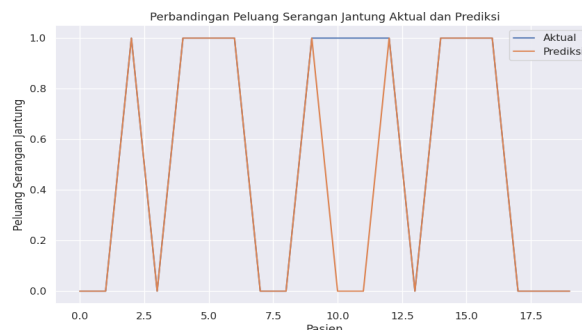
Confusion matrix digunakan dalam Aplikasi Prediksi Kematian Akibat Gagal Jantung untuk mengevaluasi kinerja model dengan membandingkan hasil prediksi dengan data sebenarnya. Visualisasi confusion matrix memungkinkan analisis mendalam terhadap performa model sebelum dan sesudah penerapan teknik Random Under Sampling. Teknik ini bertujuan untuk mengatasi masalah ketidakseimbangan data (imbalanced data), di mana jumlah pasien yang selamat dan tidak selamat mungkin sangat berbeda. Dengan

mengurangi data pada kelas mayoritas, model diharapkan dapat lebih akurat dalam memprediksi kelas minoritas, seperti pasien yang tidak selamat. Confusion matrix digunakan untuk membandingkan performa model Logistic Regression, KNN, dan Random Forest sebelum dan setelah resampling seperti yang terlihat pada **Gambar 3**.



Gambar 3 Confusion Matrix Logistic Regression, KNN, dan Random Forest

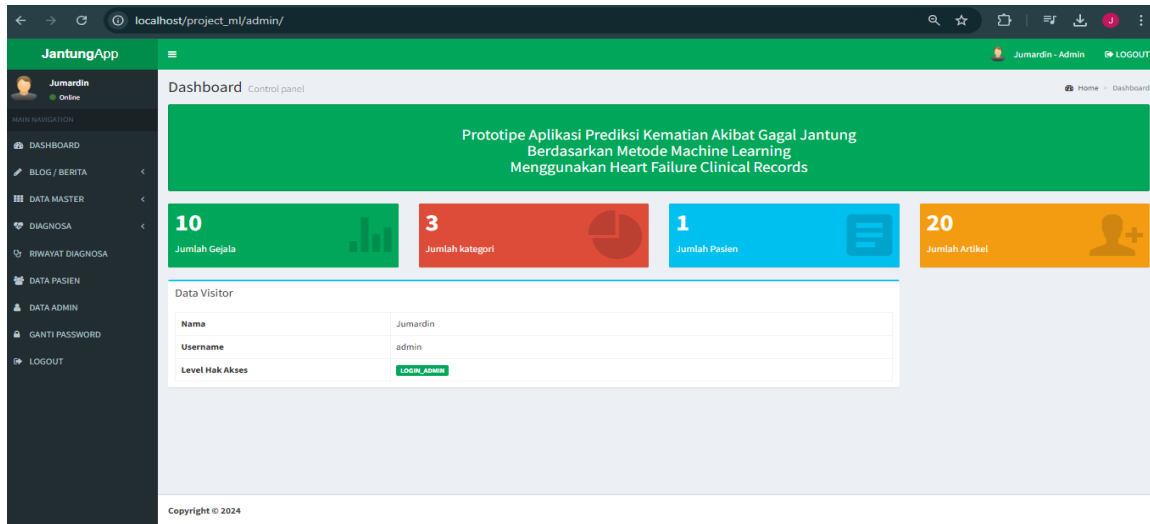
Evaluasi dilakukan untuk menentukan model terbaik pada aplikasi prediksi kematian akibat gagal jantung, menggunakan algoritma Logistic Regression, KNN, dan Random Forest. Grafik perbandingan menunjukkan peluang aktual dan hasil prediksi untuk 20 pasien pertama berdasarkan data yang tersedia. Logistic Regression memprediksi kemungkinan kematian akibat gagal jantung berdasarkan hubungan linier antara fitur-fitur utama dan probabilitas hasil, yang divisualisasikan melalui garis prediksi. Model memberikan prediksi berdasarkan jarak antara data pasien dalam ruang fitur, menggambarkan hubungan kedekatan dengan nilai actual seperti yang ada pada **Gambar 4**.



Gambar 4 Grafik Aktual dan Prediksi

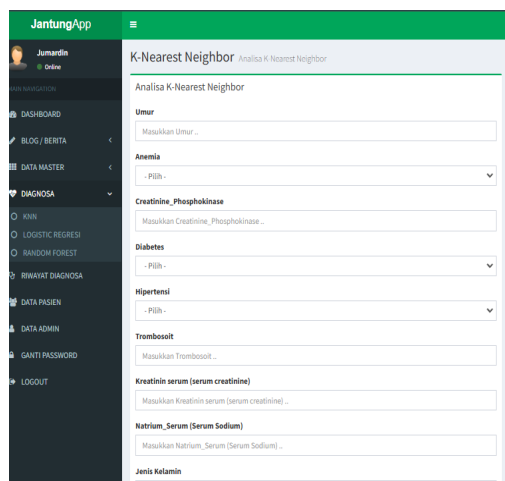
Perancangan prototipe aplikasi prediksi kematian akibat gagal jantung menggunakan metode machine learning berdasarkan data *heart failure clinical records* dilengkapi dengan beberapa menu utama untuk mendukung fungsionalitasnya. Menu Dashboard memberikan ringkasan informasi penting, seperti statistik prediksi dan laporan terkait. Menu Blog / Berita berisi informasi terkini mengenai kesehatan jantung dan edukasi bagi pengguna. Data Master digunakan untuk mengelola data pendukung, seperti parameter medis dan kategori risiko. Menu Diagnosa memungkinkan admin atau pengguna untuk melakukan prediksi berdasarkan input data

pasien, sedangkan Riwayat Diagnosa menampilkan hasil prediksi sebelumnya yang telah tersimpan dalam sistem. Selain itu, terdapat menu Data Pasien untuk manajemen informasi pasien, Data Admin untuk pengelolaan pengguna sistem, serta fitur Ganti Password dan Logout untuk keamanan akses aplikasi seperti yang terlihat pada **Gambar 5**.



Gambar 5 Tampilan Dashboard Aplikasi

Dalam aplikasi ini, model terbaik yang dipilih dari hasil evaluasi di antara algoritma Logistic Regression, K-Nearest Neighbors (KNN), dan Random Forest diimplementasikan untuk mendukung proses prediksi. Pemilihan model dilakukan berdasarkan kinerja yang diukur menggunakan metrik seperti akurasi, precision, recall, dan F1-score. Model terbaik diintegrasikan dalam menu Diagnosa, di mana pengguna dapat memasukkan data pasien seperti usia, kondisi medis, dan hasil tes klinis untuk mendapatkan hasil prediksi risiko kematian akibat gagal jantung. Implementasi model ini dapat dilihat pada **Gambar 6**.



Gambar 6 Antarmuka diagnosa model Logistic Regression, KNN, dan Random Forest

#### 4. KESIMPULAN

Penelitian ini berhasil mengembangkan prototipe aplikasi prediksi kematian akibat gagal jantung menggunakan metode machine learning berdasarkan data heart failure clinical records. Dataset yang digunakan mencakup 299 catatan klinis pasien dengan 13 kolom atribut, di mana atribut target adalah Death Event/Kematian dan atribut lainnya, seperti usia, jenis kelamin, riwayat penyakit (anemia, diabetes, tekanan darah tinggi), hasil tes laboratorium (kreatinin, sodium, dan fraksi ejeksi), digunakan sebagai prediktor. Model Random Forest menunjukkan performa terbaik dengan akurasi tertinggi, diikuti oleh K-Nearest Neighbors (KNN) dan Logistic Regression, yang diimplementasikan dalam aplikasi untuk memprediksi risiko kematian secara akurat. Aplikasi ini dilengkapi dengan menu fungsional seperti dashboard, blog / berita, data master, diagnosa, dan riwayat diagnosa, yang mendukung pengguna, terutama tenaga medis, dalam pengambilan keputusan klinis dan pengelolaan data pasien.

## UCAPAN TERIMA KASIH

Terima kasih kepada Fakultas Teknik dan Informatika dan Lembaga Riset dan Pengabdian Kepada Masyarakat Universitas Dian Nusantara yang telah mendukung penelitian ini.

## REFERENSI

- [1] N. L. K. A. Arsani, N. P. D. S. Wahyuni, N. N. M. Agustini, and M. Budiawan, "Deteksi Dini dan Pencegahan Penyakit Kardiovaskuler," *Proceeding Senadimas Undiksha*, p. 663, 2022.
- [2] B. F. Sitanggang and P. Sitompul, "Deteksi Awal Kelangsungan Hidup Pasien Gagal Jantung Menggunakan Machine Learning Metode Random Forest," *Innov. J. Soc. Sci. Res.*, vol. 4, no. 2, pp. 3347–3357, 2024.
- [3] D. P. Sari, M. Mustain, and M. Maksun, "Gambaran Pengelolaan Hipervolemia pada Gagal Jantung Kongestif di Rumah Sakit," *J. Keperawatan Berbudaya Sehat*, vol. 1, no. 1, pp. 9–15, 2023.
- [4] C. Risa, "Hubungan Pengetahuan, Sikap Dan Beban Kerja Terhadap Perilaku Perawat Kritis Dalam Memberikan Mobilisasi Dini Pada Pasien Gagal Jantung," *Altra J. Keperawatan Holistik*, vol. 1, no. 2, pp. 63–70, 2024.
- [5] A. Homaidi and Z. Fatah, "Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung," *G-Tech J. Teknol. Terap.*, vol. 8, no. 3, pp. 1720–1728, 2024.
- [6] D. Ramayanti, S. D. Asri, and L. Lionie, "Implementasi Model Arsitektur VGG16 dan MobileNetV2 Untuk Klasifikasi Citra Kupu-Kupu," *JSAI (Journal Sci. Appl. Informatics)*, vol. 5, no. 3, pp. 182–187, 2022.
- [7] S. D. Asri, I. Jaya, A. Buono, and S. H. Wijaya, "Fish Detection in Seagrass Ecosystem using Masked-Otsu in HSV Color Space," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 12, 2022.
- [8] H. Noprisson and Budiyarti, "Aplikasi Manajemen Pemeliharaan Produk Perangkat Lunak," *J. Sci. Appl. Informatics*, vol. 1, no. 2, pp. 41–45, 2018.
- [9] V. Ayumi, "Studi Pendahuluan: Pengembangan Aplikasi m-BCARE Untuk Pasien Penderita Kanker Payudara," *JUSIBI (Jurnal Sist. Inf. dan E-Bisnis)*, vol. 3, no. 1, pp. 26–33, 2021.
- [10] A. Ratnasari, D. Fitriana, and W. H. Haji, "BPTrends Redesign Methodology (BPRM) for the Development Disaster Management Prevention Information System," in *Proceedings of the 2020 2nd Asia Pacific Information Technology Conference, 2020*, pp. 113–117.
- [11] I. Nurhaida, V. Ayumi, H. Noprisson, A. Ratnasari, M. Utami, and E. D. Putra, "Web Development Using WISDM and RAD," in *2020 International Conference on Information Technology Systems and Innovation (ICITSI)*, 2020, pp. 57–61.
- [12] A. Edwita, D. I. Sensuse, and H. Noprisson, "Critical success factors of information system development projects," vol. 2017, pp. 285–290, 2017.
- [13] S. Adi and A. Wintarti, "Komparasi metode support vector machine (SVM), K-Nearest Neighbors (KNN), Dan Random Forest (RF) untuk prediksi penyakit gagal jantung," *MATHunesa J. Ilm. Mat.*, vol. 10, no. 2, pp. 258–268, 2022.
- [14] M. A. Abubakar, M. Muliadi, A. Farmadi, R. Herteno, and R. Ramadhani, "Random Forest Dengan Random Search Terhadap Ketidakseimbangan Kelas Pada Prediksi Gagal Jantung," *J. Inform.*, vol. 10, no. 1, pp. 13–18, 2023.
- [15] U. Sunarya and T. Haryanti, "Perbandingan Kinerja Algoritma Optimasi pada Metode Random Forest untuk Deteksi Kegagalan Jantung," *J. Rekayasa Elektr.*, vol. 18, no. 4, 2022.
- [16] Fedesoriano, "Heart Failure Prediction Dataset," *Kaggle*, Sep-2021. [Online]. Available: <https://www.kaggle.com/fedesoriano/heart-failure-prediction>.