**Makalah Tugas Akhir Mata Kuliah Teknologi Kecerdasan Buatan**

**“Prediksi Serangan Jantung Menggunakan Algoritma *Random Forest* berbasis *Website*: Sebuah Pendekatan Machine Learning untuk Kesehatan Jantung”**



**Disusun Oleh: Kelompok 17**

|  |  |
| --- | --- |
| **NIM** | **Nama** |
| 11321016 | Albert Arta Manik |
| 11321049 | Boy Tri Anugrah |
| 11321062 | Olyvia Siahaan |
| 11321072 | Herlina Nikita Purba |

FAKULTAS VOKASI  
PROGRAM STUDI DIPLOMA III TEKNOLOGI INFORMASI  
INSTITUT TEKNOLOGI DEL  
2023/2024

# DAFTAR ISI

[DAFTAR ISI 2](#_Toc152667239)

[DAFTAR TABEL 3](#_Toc152667240)

[DAFTAR GAMBAR 4](#_Toc152667241)

[BAB I PENDAHULUAN 5](#_Toc152667242)

[1.1 Latar Belakang 5](#_Toc152667243)

[1.2 Rumusan Masalah 6](#_Toc152667244)

[1.3 Tujuan 6](#_Toc152667245)

[1.4 Lingkup 6](#_Toc152667246)

[BAB II METODE PENELITIAN 7](#_Toc152667247)

[2.1 Metode Penelitian 7](#_Toc152667248)

[2.1.1 Pengumpulan Data 7](#_Toc152667249)

[2.1.2 Pengolahan Data 8](#_Toc152667250)

[2.2 Implementasi *Random Forest* 10](#_Toc152667251)

[2.3 Analisis 11](#_Toc152667252)

[2.4 Evaluasi 12](#_Toc152667253)

[2.5 Deployment 12](#_Toc152667254)

[BAB III HASIL DAN KESIMPULAN 16](#_Toc152667255)

[3.1 Hasil 16](#_Toc152667256)

[3.2 Kesimpulan 18](#_Toc152667257)

[REFERENSI 19](#_Toc152667258)

# DAFTAR TABEL

[Tabel 1 Parameter yang digunakan 8](#_Toc152668070)

[Tabel 2 Evaluasi *Random Forest* 12](#_Toc152668071)

# DAFTAR GAMBAR

[Gambar 1 Flowchart Metodologi Penelitian 8](#_Toc152668091)

[Gambar 2 Tampilan Data yang Diunduh 8](#_Toc152668092)

[Gambar 3 Jumlah Missing Value 10](#_Toc152668093)

[Gambar 4 Implementasi *Random Forest* 11](#_Toc152668094)

[Gambar 5 Confusion Matrix 12](#_Toc152668095)

[Gambar 6 Tampilan Halaman Form Inputan Iser 15](#_Toc152668096)

[Gambar 7 Implementasi Inputan Form User 18](#_Toc152668097)

# BAB I PENDAHULUAN

## Latar Belakang

Penyakit jantung, atau yang dikenal juga sebagai penyakit kardiovaskular, merupakan permasalahan kesehatan yang signifikan baik di negara-negara maju maupun berkembang. Penyakit ini menduduki peringkat teratas sebagai penyebab kematian global setiap tahunnya. Sejak tahun 2008, diperkirakan bahwa sekitar 17,3 juta kematian terkait dengan kondisi kardiovaskular [1]. Penyakit jantung adalah kondisi di mana kinerja jantung terhambat, menyebabkan ketidakmampuan jantung untuk efektif memompa darah dan oksigen ke seluruh tubuh. Hal ini mengakibatkan gangguan dalam fungsi dasar jantung sebagai motor utama sirkulasi darah [2]. Tingginya tingkat kematian yang disebabkan oleh penyakit jantung menekankan pentingnya melakukan prediksi penyakit ini sebagai langkah untuk mengurangi angka kematian akibat kondisi tersebut. Penyakit jantung disebabkan oleh berbagai faktor predisposisi, termasuk kebiasaan pribadi dan faktor genetik. Faktor-faktor risiko dari kebiasaan seperti merokok, konsumsi alkohol dan kafein yang berlebihan, tingkat stres, serta tingkat aktivitas fisik, bersama dengan faktor fisiologis lainnya seperti obesitas, hipertensi, kolesterol tinggi, tekanan darah tinggi, dan kondisi jantung pre-existing, semuanya merupakan elemen-elemen yang berkontribusi sebagai faktor predisposisi terhadap penyakit jantung [3].

Dengan kemajuan teknologi, berbagai upaya dilakukan untuk memberikan kenyamanan kepada manusia, salah satunya adalah perkembangan dalam ranah keilmuan Kecerdasan Buatan (*Artificial Intelligence*/AI). Maka dalam penelitian ini, akan dijelaskan mengenai Artificial Intelligence, khususnya dalam ranah *Machine Learning* (ML), yang memiliki kemampuan untuk melakukan prediksi terhadap potensi penyakit jantung berdasarkan data terkait. Salah satu teknik *Machine Learning* yang diterapkan untuk memprediksi penyakit jantung adalah menggunakan metode *Random Forest*. Dalam penelitian yang berjudul “*Prediction of Cardiovascular Disease using Machine Learning*” yang dilakukan oleh M. Balakrishnan et al., bertujuan mengembangkan sistem diagnosis otomatis untuk memprediksi penyakit jantung menggunakan berbagai model analisis yang memiliki dataset dengan 13 atribut dan 303, kemudian data tersebut diproses setelah dikumpulkan. Algoritma klasifikasi, seperti algoritma *Random Forest* memprediksi fitur untuk memberikan akurasi dan presisi tinggi. Hasil dataset berisi nilai diagnosis, yaitu 0 dan 1, dimana 0 menunjukkan ketiadaan penyakit jantung dan 1 menunjukkan kehadiran penyakit jantung. Algoritma *Random Forest* berhasil memprediksi dengan akurasi yang tinggi [4]. Oleh karena itu, berdasarkan pemaparan diatas, tugas akhir ini menggunakan algoritma *Random Forest* dalam memprediksi serangan jantung.

## Rumusan Masalah

Bagaimana penerapan algoritma *Random Forest* untuk prediksi serangan jantung?

## Tujuan

Mengetahui cara penerapan algoritma *Random Forest* untuk prediksi serangan jantung.

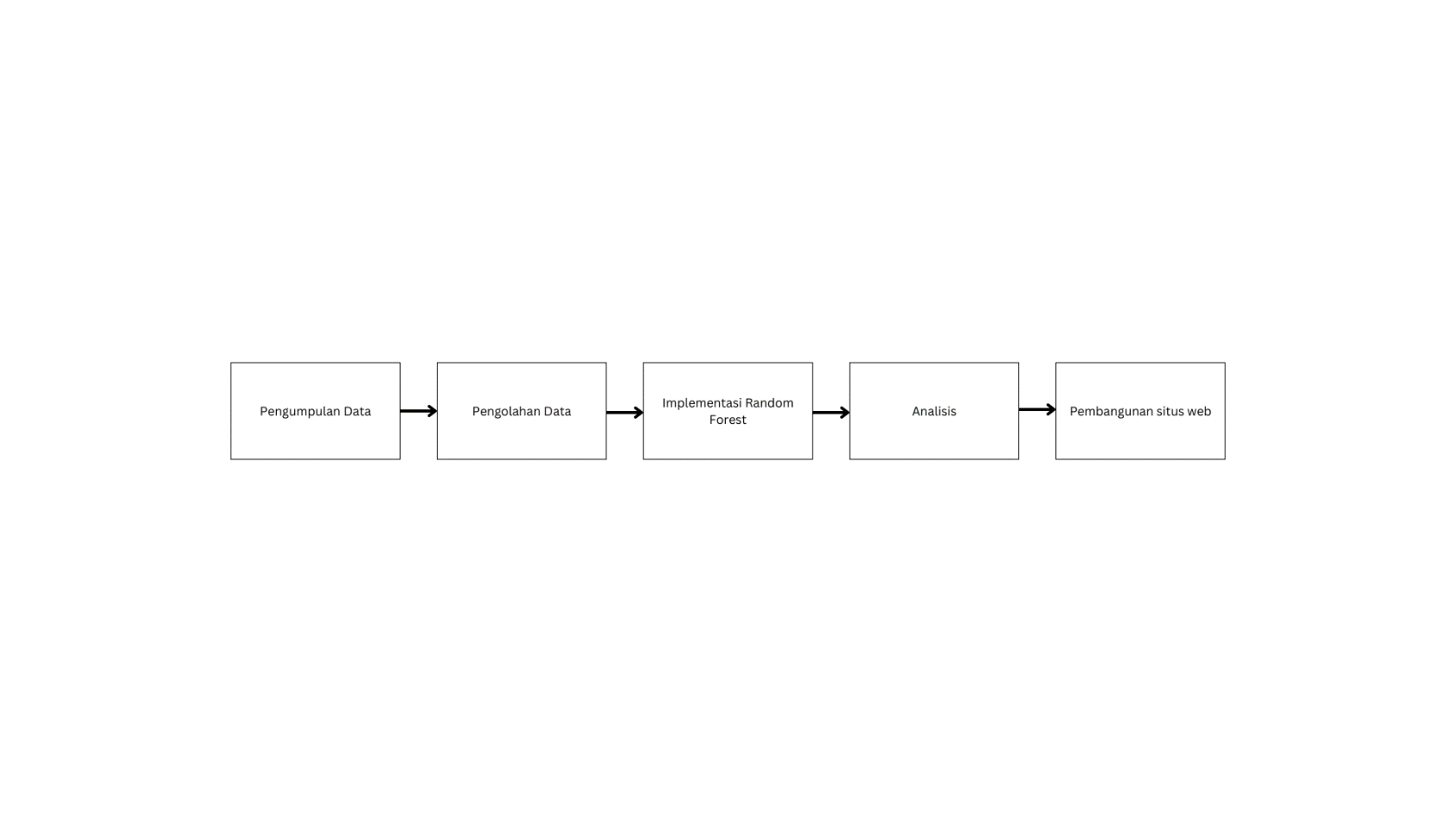
## Lingkup

1. Dataset berasal dari Kaggle yang berjumlah 303 data dengan 13 atribut.
2. Algoritma yang digunakan adalah *Random Forest*.
3. Bahasa pemrograman yang digunakan adalah bahasa Python.

# BAB II METODE PENELITIAN

## Metode Penelitian

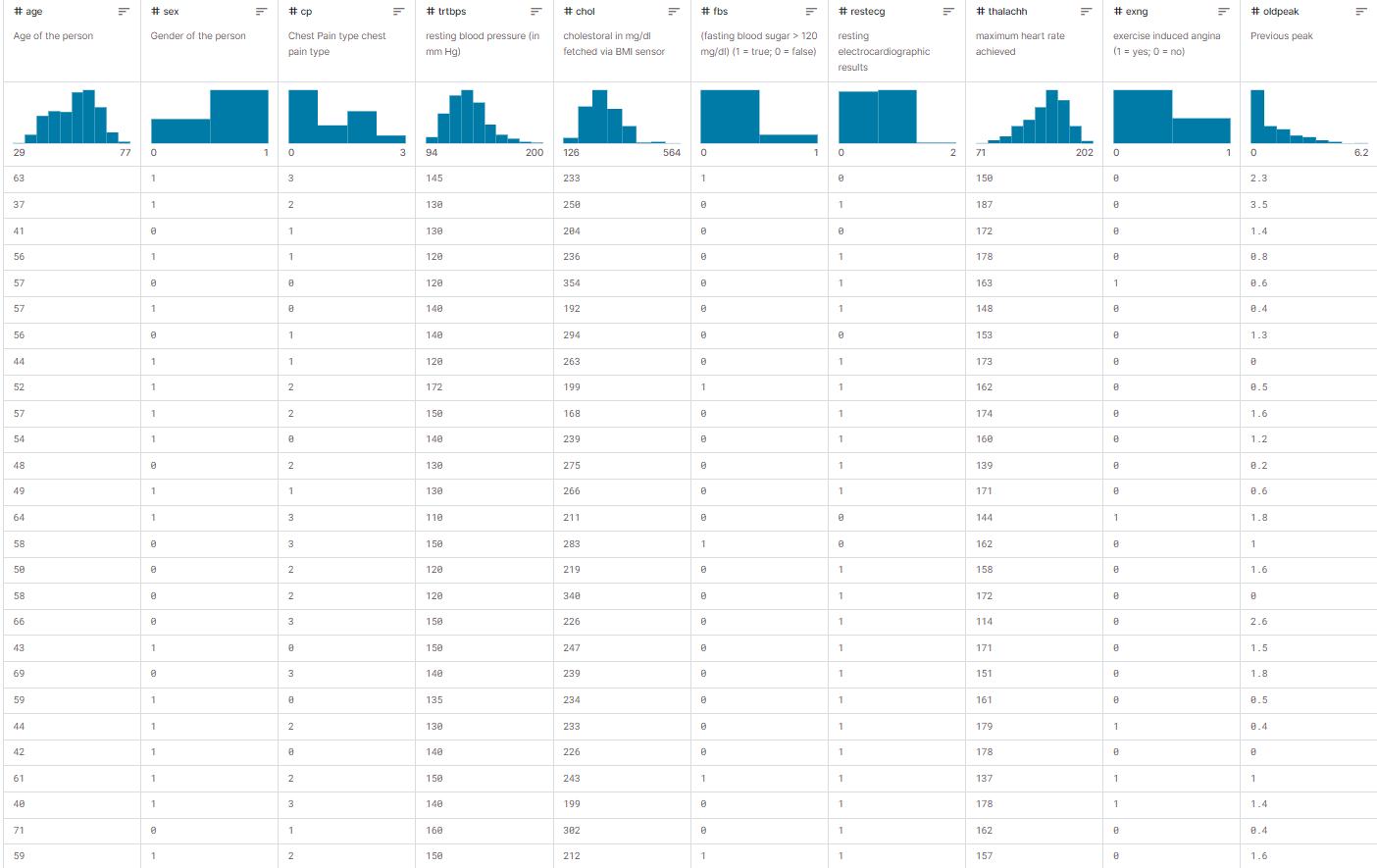
Penelitian mengenai prediksi serangan jantung dengan memanfaatkan algoritma *Random Forest* melibatkan lima langkah penelitian. Langkah-langkah tersebut mencakup: (1) pengumpulan data; (2) proses pengolahan data; (3) implementasi algoritma *Random Forest*; (4) analisis; dan (5) pembangunan situs *web*. Metode penelitian yang dilakukan pada penelitian ini dapat dilihat pada Gambar 1.



**Gambar 1 Flowchart Metodologi Penelitian**

### Pengumpulan Data

Set data yang digunakan diambil dari situs *web* Kaggle <https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset/>. Terdiri dari 303 pasien dengan 10 variabel.



**Gambar 2 Tampilan Data yang Diunduh**

Dan berikut ada juga keterangan dari parameter yang digunakan:

**Tabel 1 Parameter yang digunakan**

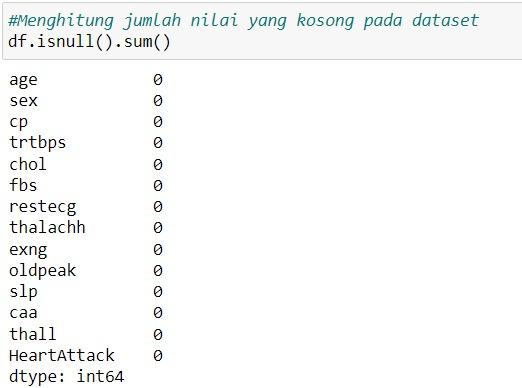
|  |  |  |  |
| --- | --- | --- | --- |
| **Variabel** | | Keterangan | |
| **age** | | Usia pasien yang bersangkutan (29 – 77) | |
| **sex** | | Jenis kelamin pasien (0 perempuan, 1 laki-laki) | |
| **cp** | | Jumlah pembuluh darah utama (0-3) | |
| **trtbps** | Tekanan darah saat istirahat (dalam mmHg) | |
| **chol** | Kolestrol dalam mg/dl yang diambil melalui sensor BMI | |
| **fbs** | Gula darah puasa > 120 mg/dl (1 = benar; 0 = salah) | |
| **restecg** | Hasil elektrokardiografi saat istirahat | |
| **thalachh** | Detak jantung maksimum yang dicapai | |
| **exng** | Angina yang diinduksi oleh olahraga (1 = ya; 0 = tidak) | |
| **oldpeak** | Puncak sebelumnya | |

### Pengolahan Data

Terlebih dahulu, melakukan pengolahan data dengan prapemrosesan terhadap dataset yang telah dikumpulkan untuk mengurangi kompleksitas dan meningkatkan keterjangkauan pengguna. Pemprosesan data (*data preprocessing*) adalah serangkaian teknik yang diterapkan pada dataset untuk mengatasi ketidakpastian dan ketidak konsistenan, sehingga mengakibatkan penghapusan noise, nilai yang hilang, dan data yang tidak konsisten dalam database [5]. Dengan memanfaatkan dataset dari *Kaggle* yang terdiri dari 13 atribut yang ditetapkan, di mana faktor-faktor tersebut memiliki peran yang signifikan dalam penyakit jantung.

#### Missing Feature Values

Dalam konteks DataFrame yang diberikan, ditegaskan bahwa tidak terdapat data yang kosong atau hilang (*null*) dalam setiap kolom. Jumlah *non-null count* sebanyak 303 pada setiap kolom menandakan bahwa setiap entri atau elemen dataset telah terisi dengan nilai yang ditentukan [6]. Dengan total jumlah baris atau entri sebanyak 303, kesimpulan yang dapat diambil adalah dataset ini lengkap dan tidak memiliki informasi yang tidak terisi atau dilupakan. Dengan kata lain, setiap kolom, yang mencakup parameter-parameter terkait penyakit jantung, memiliki data yang lengkap dan tidak mengandung nilai yang kosong atau hilang. Keberadaan dataset yang lengkap ini menjadi aspek penting untuk keakuratan dan efektivitas analisis serta pemodelan yang akan dilakukan terkait deteksi penyakit jantung. Gambar 3 merupakan penjumlahan dari data yang hilang (*missing value*)



**Gambar 3 Jumlah Missing Value**

#### Scaling Features

Dalam langkah-langkah pra-pemrosesan data untuk persiapan model *machine learning*, potongan kode tersebut mengimplementasikan *Scaling Features* dengan menggunakan *Standard Scaler* dari pustaka *scikit-learn*. Dalam proses ini, kolom-kolom tertentu yang dipilih dari DataFrame, yaitu *'age', 'trtbps', 'chol', 'thalachh', dan 'oldpeak'*, telah diubah skala nilainya. Dengan menerapkan penskalaan fitur ini, data menjadi lebih siap untuk digunakan dalam tahap pelatihan model dengan hasil yang lebih konsisten dan akurat.

#### Splitting Data

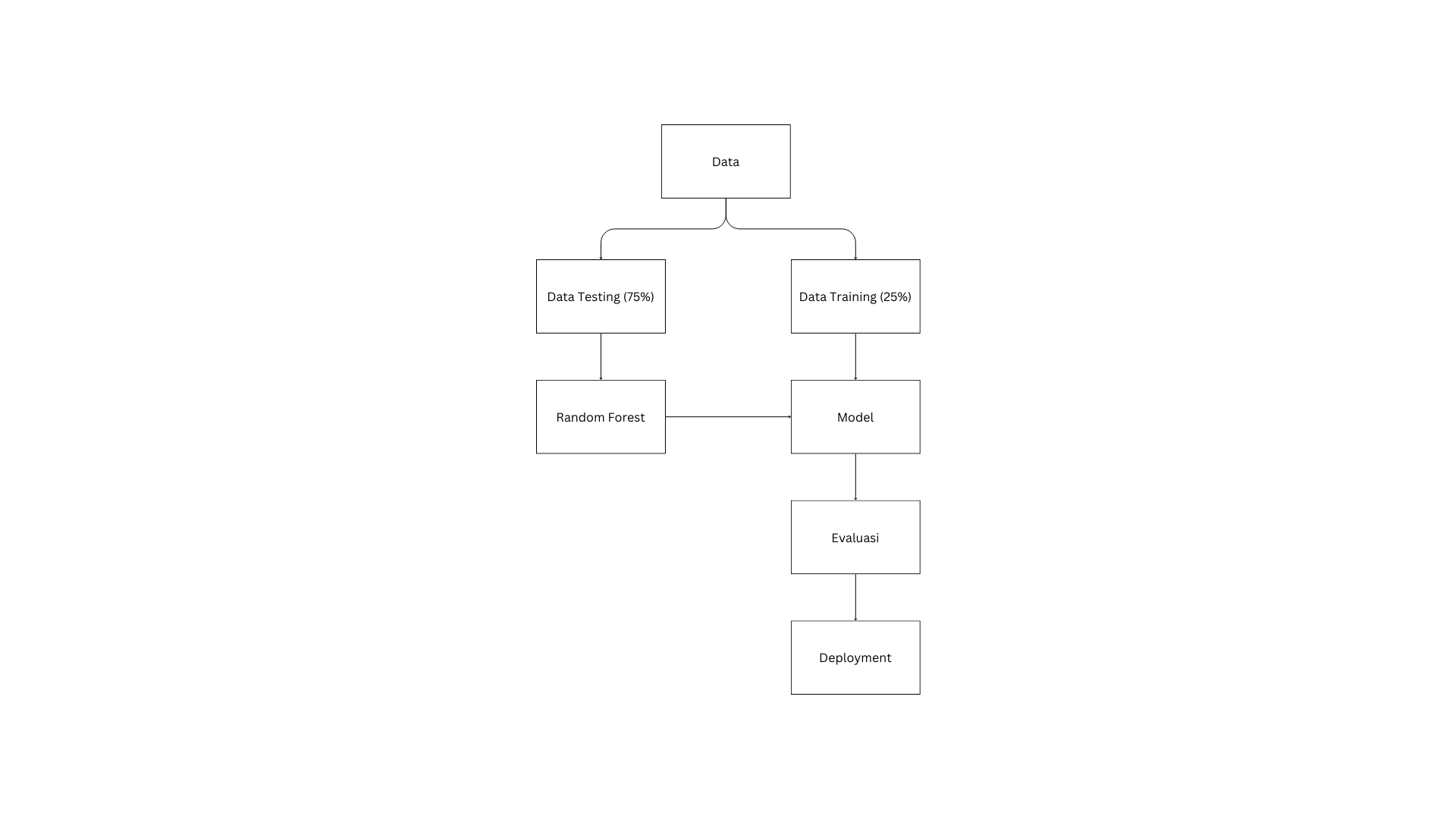
Pemisahan dataset adalah proses membagi dataset menjadi dua bagian atau lebih, yang umumnya digunakan untuk melatih dan menguji model machine learning. Dalam hal ini, langkah-langkah tersebut secara khusus memisahkan variabel target (y) dari variabel fitur (X), sehingga dataset dapat disiapkan untuk tahap pelatihan dan pengujian model.

X\_train, X\_valid, y\_train, y\_valid = train\_test\_split(X, y, random\_state = 101, stratify=y, test\_size=0.25)

Dalam hal ini, 25% dari data akan menjadi subset pengujian, sementara 75% akan menjadi subset pelatihan.

## Implementasi *Random Forest*

Dalam metode yang diajukan, deteksi penyakit jantung dapat dilakukan secara lebih efisien dan ekonomis dalam waktu singkat. Pendekatan ini menggunakan data yang sudah diproses sebelumnya untuk melatih dan menguji menggunakan algoritma *machine learning*. Pada tahap awal, data yang telah diproses dibagi menjadi dua bagian, dengan sebagian besar digunakan dalam tahap pelatihan (75%) dan sisanya (25%) untuk tahap pengujian. Selama tahap pelatihan dan pengujian, sistem yang diusulkan melibatkan penggunaan algoritma *Random Forest* untuk melatih dataset. Menggunakan platform Jupyter, kami telah melatih dan akhirnya memprediksi hasil untuk seorang pasien. Seperti yang ditunjukkan dalam Gambar 4, sistem yang telah terlatih dapat mengindikasikan adanya penyakit kardiovaskular pada pasien. Tujuan utama dari pengembangan sistem ini adalah untuk mendeteksi penyakit jantung dengan lebih efisien dan akurat.



**Gambar 4 Implementasi Random Forest**

Evaluasi kinerja *Random Forest* dilakukan dengan menggunakan beberapa parameter pengukuran, seperti akurasi, presisi, *recall*, dan *F1-Score*. Akurasi, sebagai parameter umum dan sederhana dalam mengevaluasi kinerja algoritma klasifikasi, mencerminkan tingkat kebenaran prediksi dalam bentuk persentase. Presisi pada *Random Forest* mengukur seberapa akurat model dalam memprediksi positif. Dalam kasus ini, positif dapat merujuk pada kelas tertentu yang ingin diprediksi. *Recall* pada *Random Forest* mengukur seberapa baik model dalam menemukan semua kasus positif yang sebenarnya di dataset. *F1-Score* adalah suatu metrik yang menggabungkan presisi dan *recall* untuk memberikan gambaran keseluruhan kinerja model. *F1-Score* baik digunakan ketika kita ingin mencapai keseimbangan antara presisi dan recall [7]*.*

## Analisis

Pada tahap analisis, kami melakukan evaluasi mendalam terhadap model yang dihasilkan untuk prediksi serangan jantung. Berikut adalah matriks kebingungan yang mencerminkan hasil pengujian model ditunjukkan pada Gambar 5.



**Gambar 5 Confusion Matrix**

Matriks kebingungan memberikan gambaran yang lebih rinci tentang performa model, memperlihatkan seberapa baik model dapat membedakan antara kelas positif dan negatif. Dengan menggunakan matriks kebingungan, dapat menghitung parameter seperti akurasi, presisi, recall, dan F1-Score.

## Evaluasi

**Tabel 2 Evaluasi Random Forest**

|  |  |
| --- | --- |
| Accuracy | 100.0 |
| Precision | 1.0 |
| Recall | 1.0 |
| F1 Score | 1.0 |

Dalam hasil evaluasi model yang diberikan, semua metrik menunjukkan performa yang sangat baik. Akurasi sebesar 100.0% mengindikasikan bahwa model berhasil memprediksi seluruh sampel dengan benar tanpa adanya kesalahan. Presisi sebesar 1.0 menunjukkan bahwa setiap prediksi positif yang dilakukan oleh model benar, sehingga tidak ada false positive. Recall yang juga mencapai nilai 1.0 menunjukkan kemampuan model untuk menemukan semua instance positif yang sebenarnya dalam dataset. F1 Score sebesar 1.0 menggambarkan keseimbangan yang optimal antara presisi dan recall. Dengan angka-angka yang mendekati kesempurnaan ini, model tampaknya memiliki kemampuan prediksi yang sangat baik pada dataset yang dievaluasi.

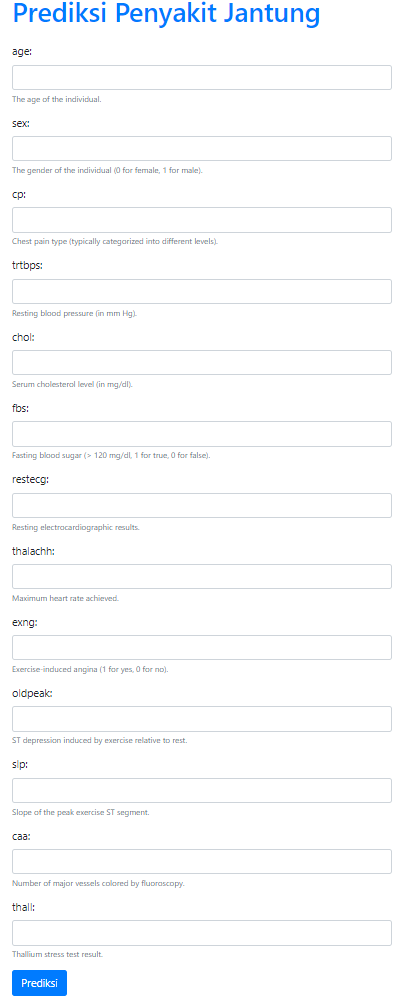
## Deployment

*Flask* adalah *framework* yang digunakan untuk membangun aplikasi *web* dengan menggunakan bahasa pemrograman Python*. Flask* adalah sebuah *web framework* yang ditulis dengan bahasa Python dan tergolong sebagai jenis *microframework*, karena tidak memerlukan *library* untuk menggunakannya. Dengan menggunakan *Flask* dan bahasa Python, pengembang dapat membuat sebuah *web* yang terstruktur dan dapat mengatur *behaviour* suatu *web* dengan lebih mudah. *Flask* digunakan sebagai kerangka kerja aplikasi dan tampilan suatu *web*site. Penggunaan *flask* dapat membuat pengembang lebih mudah dan terstruktur dalam membuat sebuah *web* [8]*.*

Keunggulan *flask* *python*:

* Dapat menangani fungsi HTTP *request* dengan mudah
* Ringan untuk dijalankan karena mempunyai *core* yang sederhana
* APl yang baik dan konsisten
* Mudah utnuk diperiksa secara menyeluruh
* Mudah untuk di *deploy*
* Fleksibilitas tinggi, dengan konfigurasi yang sangat mudah untuk diubah.

Tampilan halaman *web* dengan framework ini menampilkan 13 form inputan yang diminta dari user dan merupakan atribut yang digunakan pada dataset. Tampilan halaman form inputan user dapat dilihat pada Gambar 5 berikut.



**Gambar 6 Tampilan Halaman Form Inputan Iser**

Form ini digunakan oleh *user* (pengguna) untuk mengisi *age* (umur), *sex* (jenis kelamin), *cp* (tipe nyeri dada (biasanya dikategorikan ke dalam level yang berbeda rentang 0-3), *trtbps* (tekanan darah saat istirahat (dalam mm Hg), *chol* (tingkat kolesterol serum (dalam mg/dl), *fbs* (gula darah puasa (> 120 mg/dl, 1 untuk benar, 0 untuk salah), *restecg* (hasil elektrokardiografi saat istirahat), *thalachh* (denyut jantung maksimal yang dicapai), *exng* (angina yang dipicu olahraga (1 untuk ya, 0 untuk tidak)), *oldpeak* (depresi segmen ST yang dipicu oleh olahraga relatif terhadap istirahat), *slp* (kemiringan segmen ST puncak olahraga), *caa* (jumlah pembuluh besar yang terlihat berwarna pada pemeriksaan fluoroskopi), dan *thall* (hasil tes stres *thallium*).

# BAB III HASIL DAN KESIMPULAN

## Hasil

Hasil evaluasi model menunjukkan bahwa penerapan *Random Forest* dapat efektif untuk memprediksi serangan jantung, terbukti dari evaluasi akurasi, presisi, recall, dan f1 score. Akurasi model *Random Forest* mencapai 100%. Prediksi yang dihasilkan oleh model, setelah dilatih dan diuji dengan menggunakan seluruh data, menunjukkan kinerja yang sangat baik. Hasil implementasi dari pemanfaatan algoritma *Random Forest* dan Pembangunan pada halaman situs *web* dengan *form* dapat dilihat pada Gambar 6 berikut ini.



**Gambar 7 Implementasi Inputan Form User**

## Kesimpulan

Berdasarkan analisa yang dilakukan akan hasil penelitian ini, berikut adalah beberapa hal yang dapat disimpulkan:

1. Proses implementasi algoritma *Random Forest* yang digunakan sukses dengan meraih tingkat akurasi sebesar 100%.
2. Implementasi algoritma *Random Forest* di bangun ke dalam halaman situs *web* dengan prediksi hasil yakni *'More Chance of Heart Attack'* yang menunjukkan kena serangan jantung (1) dan *'Less Chance of Heart Attack'* yang menunjukkan tidak kena serangan jantung (0).

# REFERENSI

[1] D. Yana *et al.*, “Penerapan Metode Teorema Bayes Pada Sistem Pakar Untuk Mendiagnosa Gangguan Sistem Kardioveskular Pada Rumah Sakit Umum Pusat Haji Adam Malik,” *J. SI (Sistem Inf.*, vol. 2020, p. PP, 2020.

[2] Irpanudin, Reka, R. Nur Anggraeni, P. Pratama, A. Sujjada, and A. Fergina, “Prediksi Penyakit Jantung Menggunakan Metode Deep Neural Network dengan Memanfaatkan Internet of Things,” *J. Inf. dan Teknol.*, vol. 5, pp. 45–55, 2023.

[3] Dimsyiar M Al Hafiz, Khoirul Amaly, Javen Jonathan, M Teranggono Rachmatullah, and Rosidi, “Sistem Prediksi Penyakit Jantung Menggunakan Metode Naive Bayes,” *J. Rekayasa Elektro Sriwij.*, vol. 2, no. 2, pp. 151–157, 2021.

[4] M. Balakrishnan, A. B. Arockia Christopher, P. Ramprakash, and A. Logeswari, “Prediction of Cardiovascular Disease using Machine Learning,” *J. Phys. Conf. Ser.*, vol. 1767, no. 1, 2021.

[5] Oon Wira Yuda, Darmawan Tuti, Lim Sheih Yee, and Susanti, “Penerapan Penerapan Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode *Random Forest*,” *SATIN - Sains dan Teknol. Inf.*, vol. 8, no. 2, pp. 122–131, 2022.

[6] S. B. Kotsiantis and D. Kanellopoulos, “Data preprocessing for supervised leaning,” *Int. J. …*, vol. 1, no. 2, pp. 1–7, 2006.

[7] A. Primajaya and B. N. Sari, “*Random Forest* Algorithm for Prediction of Precipitation,” *Indones. J. Artif. Intell. Data Min.*, vol. 1, no. 1, p. 27, 2018.

[8] R. Somya, “Perancangan Aplikasi Chatting Berbasis *Web* di PT. Pura Barutama Kudus menggunakan Socket.IO dan Framework Foundation,” *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 4, no. 1, pp. 8–15, 2018.