

**Klasifikasi Risiko Penyakit Jantung Menggunakan *Random Forest*: Pengembangan Model Prediktif Melalui Hyperparameter Tuning**

**Laporan Tugas Akhir**

Diajukan Untuk Memenuhi  
Persyaratan Guna Meraih Gelar Sarjana  
Informatika Universitas Muhammadiyah Malang



Fildzah Lathifah  
202010370311206

**Bidang Minat**  
Data Sains

**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS MUHAMMADIYAH MALANG**

**2024**

## LEMBAR PENGESAHAN

**Cardiac Disease Risk Classification Using a Random Forest:  
Predictive Model Development Through Hyperparameter Tuning**

### TUGAS AKHIR

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1  
Informatika Universitas Muhammadiyah Malang

Disusun Oleh :

**Fildzah Lathifah**

**202010370311206**

Tugas Akhir ini telah diuji dan dinyatakan lulus melalui sidang majelis penguji  
pada tanggal 26 Maret 2024.

Menyetujui,

Dosen Penguji 1



**Briansyah Setio Wivono S.Kom.,**

**M.Kom**

**NIP. 190913071987PNS.**

Dosen Penguji 2



**Ali Sofyan Kholimi S.Kom., M.Kom.**

**NIP. 10814100562PNS.**

Mengetahui,  
Jurusan Informatika



**Ir. Galih Wasis Wicaksono S.kom, M.Cs.**

**NIP. 10814100541PNS.**

## LEMBAR PERSETUJUAN

### **Cardiac Disease Risk Classification Using a Random Forest: Predictive Model Development Through Hyperparameter Tuning**

#### **TUGAS AKHIR**

**Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1  
Informatika Universitas Muhammadiyah Malang**

Menyetujui,

Malang, 26 Maret 2024

Dosen Pembimbing 1

Dosen Pembimbing 2



**Vinna Rahmayanti S S.Si., M.Si**

**NIP. 180306071990PNS.**

**Christian Sri Kusuma Aditya**

**S.Kom., M.Kom**

**NIP. 180327021991PNS.**

## LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini :

**NAMA** : Fildzah Lathifah

**NIM** : 202010370311206

**FAK./JUR.** : Informatika

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul **“Cardiac Disease Risk Classification Using a Random Forest: Predictive Model Development Through Hyperparameter Tuning”** beserta seluruh isinya adalah karya saya sendiri dan bukan merupakan karya tulis orang lain, baik sebagian maupun seluruhnya, kecuali dalam bentuk kutipan yang telah disebutkan sumbernya.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya ini, atau ada klaim dari pihak lain terhadap keaslian karya saya ini maka saya siap menanggung segala bentuk resiko/sanksi yang berlaku.

Mengetahui,  
Dosen Pembimbing

Malang, 26 Maret 2024  
Yang Membuat Pernyataan



  
Fildzah Lathifah

Vinna Rahmayanti S S.Si., M.Si

## ABSTRAK

Masalah kesehatan global yang serius dari penyakit jantung menempatkan beban berat pada sistem kesehatan dunia. Pencegahan dan manajemen yang efektif bergantung pada identifikasi dini dan penilaian risiko yang akurat. Pada penelitian ini, algoritma Random Forest dan tuning hyperparameter Grid Search CV digunakan untuk mengantisipasi risiko penyakit jantung. Sumber data utama adalah "Dataset Penyakit Jantung" yang cukup besar dari Kaggle. Tujuannya adalah untuk meningkatkan keandalan dan keakuratan model Random Forest, yang memungkinkan profesional perawatan kesehatan membuat keputusan yang dapat dipertahankan berdasarkan prediksi risiko yang konsisten. Untuk mengelola nilai yang hilang, menghapus kolom yang tidak digunakan, dan menstandarkan kumpulan data menggunakan penskalaan Min-Max, diperlukan persiapan data. Akurasi model awal adalah 98,5%, namun setelah penyesuaian hyperparameter menggunakan CV Grid Search yang divalidasi silang 5 kali lipat, akurasi ini meningkat secara nyata hingga 100%. Temuan ini menunjukkan peningkatan kemampuan pengambilan keputusan klinis model Random Forest dan peningkatan kemanjuran prediksi. Studi ini menggarisbawahi bagaimana algoritma Random Forest, bersama dengan penyesuaian hyperparameter, memiliki potensi untuk merevolusi industri perawatan kesehatan dan meningkatkan pemahaman kita tentang penilaian risiko penyakit jantung. Profesional perawatan kesehatan dapat mengidentifikasi orang yang berisiko tinggi dan melakukan tindakan pencegahan yang tepat dengan menyediakan model prediksi yang dapat dipercaya dan dapat dijelaskan. Tujuan akhir studi ini adalah untuk mengurangi beban penyakit jantung di seluruh dunia dan meningkatkan efek kesehatan secara keseluruhan bagi mereka yang terkena dampaknya.

**Kata kunci:** Penyakit jantung, Random Forest, Grid Search CV, Pembelajaran Mesin, Informasi klinis, Elemen risiko.

## ABSTRACT

The serious global health problem of heart disease places a heavy burden on the world's healthcare systems. Effective prevention and management depend on early identification and accurate risk assessment. In this study, the Random Forest algorithm and Grid Search CV hyperparameter tuning are used to anticipate the risk of developing heart disease. The main data source is the sizable "Heart Disease Dataset" from Kaggle. The goal is to increase the reliability and accuracy of the Random Forest model, enabling healthcare professionals to make defensible decisions based on consistent risk predictions. In order to manage missing values, remove unused columns, and standardize the dataset using Min-Max scaling, data preparation is necessary. The initial model accuracy is 98.5%, however following hyperparameter tuning using a 5-fold cross-validated Grid Search CV, this accuracy rises noticeably to an amazing 100%. These findings demonstrate the Random Forest model's improved clinical decision-making capabilities and enhanced prediction efficacy. The study underlines how the Random Forest algorithm, together with hyperparameter tuning, has the potential to revolutionize the healthcare industry and increases our understanding of heart disease risk appraisal. Healthcare professionals may identify high-risk people and put the right preventative measures in place by providing a trustworthy, explicable prediction model. The study's ultimate goal is to lessen the worldwide burden of heart disease and improve the overall health effects for those who are impacted.

**Keywords:** Heart disease, Random Forest, Grid Search CV, Machine Learning, Clinical information, Risk elements.

## LEMBAR PERSEMBAHAN

Puji syukur kepada Allah SWT atas rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir ini. Penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Ibu Vinna Rahmayanti S, S.Si., M.Si dan Bapak Christian Sri Kusuma Aditya S.Kom., M.Kom selaku pembimbing tugas akhir.
2. Bapak/Ibu Dekan Fakultas Teknik Universitas Muhammadiyah Malang.
3. Bapak/Ibu Ketua Jurusan Teknik Informatika Universitas Muhammadiyah Malang.
4. Kedua orang tua saya, yang telah memberikan dukungan moral dan spiritual yang tak terukur sepanjang hidup saya.
5. Rekan-rekan sejawat saya yang telah memberikan banyak masukan, saran, dan motivasi selama proses penulisan tugas akhir ini.
6. Sahabat-sahabat terbaik saya, yang selalu ada dalam suka maupun duka, memberikan dukungan tanpa syarat.
7. Dan semua pihak yang tidak dapat saya sebutkan satu persatu, yang telah memberikan dukungan dan doa kepada saya, sehingga tugas akhir ini dapat terselesaikan.

Malang, 15 Januari 2024



Fildzah Lathifah

## KATA PENGANTAR

Dengan memanjatkan puji syukur kehadirat Allah SWT. Atas limpahan rahmat dan hidayah-NYA sehingga peneliti dapat menyelesaikan tugas akhir yang berjudul :

**“KLASIFIKASI RISIKO PENYAKIT JANTUNG MENGGUNAKAN  
*RANDOM FOREST*: PENGEMBANGAN MODEL PREDIKTIF MELALUI  
HYPERPARAMETER TUNING”**

Di dalam tulisan ini disajikan pokok-pokok bahasan yang meliputi algoritma dan deskripsi dataset yang digunakan, pra-pemrosesan yang dipilih, evaluasi dan hasil pengujian pada dataset yang telah ditentukan.

Peneliti menyadari sepenuhnya bahwa dalam penulisan tugas akhir ini masih banyak kekurangan dan keterbatasan. Oleh karena itu peneliti mengharapkan saran yang membangun agar tulisan ini bermanfaat bagi perkembangan ilmu pengetahuan.

Malang, 15 Januari 2024

  
Fildzah Lathifah



## DAFTAR ISI

<b>LEMBAR PENGESAHAN .....</b>	<b>ii</b>
<b>LEMBAR PERSETUJUAN .....</b>	<b>iii</b>
<b>LEMBAR PERNYATAAN .....</b>	<b>iv</b>
<b>ABSTRAK .....</b>	<b>v</b>
<b>ABSTRACT .....</b>	<b>vi</b>
<b>LEMBAR PERSEMBAHAN .....</b>	<b>vii</b>
<b>KATA PENGANTAR .....</b>	<b>viii</b>
<b>DAFTAR ISI .....</b>	<b>ix</b>
<b>DAFTAR GAMBAR .....</b>	<b>x</b>
<b>DAFTAR TABEL .....</b>	<b>xi</b>
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	4
1.3 Tujuan Penelitian .....	4
1.4 Batasan Penelitian .....	4
<b>BAB II TINJAUAN PUSTAKA .....</b>	<b>5</b>
2.1 Penelitian Terdahulu .....	5
2.2 Penyakit Jantung .....	6
2.3 Machine Learning .....	6
2. 3.1 Random Forest .....	7
2. 4 Hyperparameter Tuning dan GridSearchCV .....	7
<b>BAB III METODOLOGI PENELITIAN .....</b>	<b>8</b>
3.1 Algoritma Machine Learning .....	8
3.2 Dataset .....	8
3.3 Pra-Pemrosesan Data .....	10
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>14</b>
<b>BAB V ENUTUP .....</b>	<b>16</b>
5.1 Kesimpulan .....	16
5.2 Saran .....	16
<b>DAFTAR PUSTAKA .....</b>	<b>18</b>

## DAFTAR GAMBAR

<b>Gambar 3.3</b> Alur persiapan dataset.....	11
<b>Gambar 4</b> Confusion Matriks .....	14



## DAFTAR TABEL

<b>Tabel 2.1</b> Penelitian Terdahulu.....	5
<b>Tabel 3.2.1</b> Deskripsi Dataset.....	8
<b>Tabel 3.2.2</b> Nilai lima baris pertama kumpulan data.....	9
<b>Tabel 4.1</b> Hasil laporan klasifikasi sebelum dilakukan tuning.....	15
<b>Tabel 4.2</b> Hasil laporan klasifikasi setelah dilakukan tuning.....	15



## DAFTAR PUSTAKA

- [1] C. A. ul Hassan et al., “Effectively Predicting the Presence of Coronary Heart Disease Using Machine Learning Classifiers,” *Sensors*, vol. 22, no. 19, 2022, doi: 10.3390/s22197227.
- [2] S. Diwan, G. S. Thakur, S. K. Sahu, M. Sahu, and N. K. Swamy, “Predicting Heart Diseases through Feature Selection and Ensemble Classifiers,” *J. Phys. Conf. Ser.*, vol. 2273, no. 1, 2022, doi: 10.1088/1742-6596/2273/1/012027.
- [3] V. S. K. Reddy, P. Meghana, N. V. S. Reddy, and B. A. Rao, “Prediction on Cardiovascular disease using Decision tree and Naïve Bayes classifiers,” *J. Phys. Conf. Ser.*, vol. 2161, no. 1, 2022, doi: 10.1088/1742-6596/2161/1/012015.
- [4] M. M. Saim and H. Ammor, “Comparative study of machine learning algorithms (SVM, Logistic Regression and KNN) to predict cardiovascular diseases,” *E3S Web Conf.*, vol. 351, 2022, doi: 10.1051/e3sconf/202235101037.
- [5] M. El-Hasnony, O. M. Elzeki, A. Alshehri, and H. Salem, “Multi-Label Active Learning-Based Machine Learning Model for Heart Disease Prediction,” *Sensors*, vol. 22, no. 3, 2022, doi: 10.3390/s22031184.
- [6] G. N. Ahmad, H. Fatima, Shafiullah, A. Salah Saidi, and Imdadullah, “Efficient Medical Diagnosis of Human Heart Diseases Using Machine Learning Techniques with and Without GridSearchCV,” *IEEE Access*, vol. 10, no. April, pp. 80151–80173, 2022, doi: 10.1109/ACCESS.2022.3165792.
- [7] N. Absar et al., “The Efficacy of Machine-Learning-Supported Smart System for Heart Disease Prediction,” *Healthc.*, vol. 10, no. 6, pp. 1–19, 2022, doi: 10.3390/healthcare10061137.
- [8] R. Ghorbani and R. Ghousi, “Predictive data mining approaches in medical diagnosis: A review of some diseases prediction,” *Int. J. Data Netw. Sci.*, vol. 3, no. 2, pp. 47–70, 2019, doi: 10.5267/j.ijdns.2019.1.003.
- [9] Nouman and S. Muneer, “A Systematic Literature Review on Heart Disease Prediction Using Block chain and Machine Learning Techniques,” vol. 4, pp. 1–6, 2022.
- [10] R. R. Sarra, A. M. Dinar, M. A. Mohammed, and K. H. Abdulkareem, “Enhanced Heart Disease Prediction Based on Machine Learning and  $\chi^2$  Statistical Optimal Feature Selection Model,” *Designs*, vol. 6, no. 5, 2022, doi: 10.3390/designs6050087.
- [11] O. S. Hasan and I. A. Saleh, “Development of Heart Attack Prediction Model Based on Ensemble Learning,” *Eastern-European J. Enterp. Technol.*, vol. 4, no. 2–112, pp. 26–34, 2021, doi: 10.15587/1729-4061.2021.238528.

- [12] S. Begum, A. Siddique, and R. Tiwari, "A Study for Predicting Heart Disease using Machine Learning," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 10, pp. 4584–4592, 2021.
- [13] Anderies, J. A. R. W. Tchín, P. H. Putro, Y. P. Darmawan, and A. A. S. Gunawan, "Prediction of Heart Disease UCI Dataset Using Machine Learning Algorithms," *Eng. Math. Comput. Sci. J.*, vol. 4, no. 3, pp. 87–93, 2022, doi: 10.21512/emacsjournal.v4i3.8683.
- [14] P. Włodarczak, "Learning Techniques," *Mach. Learn. its Appl.*, pp. 149–160, 2019, doi: 10.1201/9780429448782-9.
- [15] Mehmood et al., "Prediction of Heart Disease Using Deep Convolutional Neural Networks," *Arab. J. Sci. Eng.*, vol. 46, no. 4, pp. 3409–3422, 2021, doi: 10.1007/s13369-020-05105-1.
- [16] K. M. Almustafa, "Prediction of heart disease and classifiers' sensitivity analysis," *BMC Bioinformatics*, vol. 21, no. 1, pp. 1–18, 2020, doi: 10.1186/s12859-020-03626-y.
- [17] Heart Disease Dataset. Available online: <https://www.kaggle.com/johnsmith88/heart-disease-dataset> (accessed on 30 June 2023).
- [18] E. H. Houssein, R. E. Mohamed, and A. A. Ali, "Heart disease risk factors detection from electronic health records using advanced NLP and deep learning techniques," *Sci. Rep.*, vol. 13, no. 1, pp. 1–19, 2023, doi: 10.1038/s41598-023-34294-6.
- [19] R. Katarya and P. Srinivas, "Predicting Heart Disease at Early Stages using Machine Learning: A Survey," *Proc. Int. Conf. Electron. Sustain. Commun. Syst. ICESC 2020*, no. Icesc, pp. 302–305, 2020, doi: 10.1109/ICESC48915.2020.9155586.
- [20] V. S. K. Reddy, P. Meghana, N. V. S. Reddy, and B. A. Rao, "Prediction on Cardiovascular disease using Decision tree and Naïve Bayes classifiers," *J. Phys. Conf. Ser.*, vol. 2161, no. 1, 2022, doi: 10.1088/1742-6596/2161/1/012015.
- [21] K. A. Enriko, "Comparative study of heart disease diagnosis using top ten data mining classification algorithms," *ACM Int. Conf. Proceeding Ser.*, pp. 159–164, 2019, doi: 10.1145/3338188.3338220.
- [22] Baccouche, B. Garcia-Zapirain, C. C. Olea, and A. Elmaghraby, "Ensemble deep learning models for heart disease classification: A case study from Mexico," *Inf.*, vol. 11, no. 4, pp. 1–28, 2020, doi: 10.3390/INFO11040207.
- [23] M. Pal and S. Parija, "Prediction of Heart Diseases using Random Forest," *J. Phys. Conf. Ser.*, vol. 1817, no. 1, 2021, doi: 10.1088/1742-6596/1817/1/012009.
- [24] M. Vj, "PREDICTION OF PRE-CARDIAC DISEASE USING ML & DL TECHNIQUES IN T-KINTER," vol. 10, pp. 1874–1883, 2023.

- [25] F. I. Alarsan and M. Younes, “Analysis and classification of heart diseases using heartbeat features and machine learning algorithms,” *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0244-x.
- [26] S. I. Ayon, M. M. Islam, and M. R. Hossain, “Coronary Artery Heart Disease Prediction: A Comparative Study of Computational Intelligence Techniques,” *IETE J. Res.*, vol. 68, no. 4, pp. 2488–2507, 2022, doi: 10.1080/03772063.2020.1713916.
- [27] N. Chandrasekhar and S. Peddakrishna, “Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization,” *Processes*, vol. 11, no. 4, 2023, doi: 10.3390/pr11041210.





### FORM CEK PLAGIARISME LAPORAN TUGAS AKHIR

Nama Mahasiswa : Fildzah Lathifah

NIM : 202010370311206

Judul TA : Klasifikasi Risiko Penyakit Jantung Menggunakan Random Forest: Pengembangan Model Prediktif Melalui Hyperparameter Tuning

#### Hasil Cek Plagiarisme dengan Turnitin

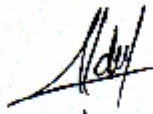
No.	Komponen Pengecekan	Nilai Maksimal Plagiarisme (%)	Hasil Cek Plagiarisme (%) *
1.	Bab 1 – Pendahuluan	10 %	4 %
2.	Bab 2 – Daftar Pustaka	25 %	9 %
3.	Bab 3 – Analisis dan Perancangan	25 %	6 %
4.	Bab 4 – Implementasi dan Pengujian	15 %	3 %
5.	Bab 5 – Kesimpulan dan Saran	5 %	4 %
6.	Makalah Tugas Akhir	20%	0 %

\*) Hasil cek plagiarism diisi oleh pemeriksa (staf TU)

\*) Maksimal 5 kali (4 Kali sebelum ujian, 1 kali sesudah ujian)

Mengetahui,

Pemeriksa (Staff TU)

  
dny

(.....)