

**Klasifikasi Penyakit Kardiovaskular Menggunakan Algoritma  
*Logistic Regression Dengan Optimasi Grid Search CV***

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



Namirah Rahmah Putri  
(202010370311204)

**Bidang Minat**  
*Data Science*

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

**2024**

**LEMBAR PERSETUJUAN**

**Klasifikasi Penyakit Kardiovaskular Menggunakan Algoritma  
Logistic Regression Dengan Optimasi Grid Search CV**

**TUGAS AKHIR**

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



Menyetujui,  
Malang, 21 Oktober 2024

Dosen Pembimbing 1



**Ir. Yufis Azhar S.Kom., M.Kom.**

**NIP. 10814100544PNS.**

## LEMBAR PENGESAHAN

### **Klasifikasi Penyakit Kardiovaskular Menggunakan Algoritma Logistic Regression Dengan Optimasi Grid Search CV TUGAS AKHIR**

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

Disusun Oleh :

**NAMIRAH RAHMAH PUTRI**

**202010370311204**

Tugas Akhir ini telah diuji dan dinyatakan lulus melalui sidang majelis pengujian  
pada tanggal 21 Oktober 2024

Menyetujui,

Dosen Penguji 1



**Christian Sri Kusuma Aditya S.Kom.,**

**M.Kom**

**NIP. 180327021991PNS.**

Dosen Penguji 2



**Setio Basuki MT., Ph.D.**

**NIP. 10809070477PNS.**



Mengetahui,  
Ketua Jurusan Informatika



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

**NIP. 10814100541PNS.**

## LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini :

**NAMA** : NAMIRAH RAHMAH PUTRI

**NIM** : 202010370311204

**FAK./JUR.** : Informatika

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul “Klasifikasi Penyakit Kardiovaskular Menggunakan Algoritma Logistic Regression Dengan Optimasi Grid Search CV” 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



Ir. Yufis Azhar S.Kom., M.Kom.

Malang, 21 Oktober 2024  
Yang Membuat Pernyataan



NAMIRAH RAHMAH PUTRI

## ABSTRAK

Penyakit Kardiovaskular merupakan salah satu penyebab utama kematian nomer satu di dunia, termasuk Indonesia. Berdasarkan data dari WHO dan RISKESDAS, angka kejadian penyakit ini terus mengalami peningkatan di tiap tahunnya sehingga memberikan dampak signifikan terhadap ekonomi dan sosial. Teknologi kecerdasan buatan khususnya machine learning, telah berperan penting dalam mendukung upaya diagnosis dan pencegahan penyakit kardiovaskular. Salah satu algoritma yang sering digunakan untuk mengklasifikasikan penyakit adalah Logistic Regression. Penelitian ini bertujuan untuk mengembangkan model Logistic Regression yang nantinya akan dioptimalkan melalui metode hyperparameter tuning dengan Grid Search CV guna meningkatkan performa model klasifikasi penyakit kardiovaskular. Dataset yang digunakan pada penelitian ini berjumlah 70.000 sampel dengan pembagian data 70% untuk data pelatihan dan 30% untuk data pengujian. Hasil penelitian menunjukkan bahwa optimasi menggunakan Grid Search CV mampu meningkatkan kinerja model dengan akurasi 73.08%, recall 72.50%, dan AUC – Score 79.42%. Dengan demikian, model Logistic Regression yang telah dioptimalkan ini berpotensi untuk memberikan kontribusi signifikan dalam proses klasifikasi dan pencegahan penyakit kardiovaskular secara lebih akurat.

**Kata Kunci :** Penyakit Kardiovaskular, Logistic Regression, Grid search CV, Klasifikasi

## ABSTRACT

Cardiovascular disease is one of the leading causes of death in the world, including Indonesia. Based on data from WHO and RISKESDAS, the incidence of this disease continues to increase every year, which has a significant economic and social impact. Artificial intelligence technology especially machine learning, has played an important role in supporting efforts to diagnose and prevent cardiovascular disease. One of the algorithms that is often used to classify diseases is Logistic Regression. This research aims to develop a Logistic Regression model that will be optimized through the hyperparameter tuning method with Grid Search CV to improve the performance of the cardiovascular disease classification model. The dataset used in this research amounted to 70,000 samples with 70% data division for training data and 30% for testing data. The results showed that optimization using Grid Search CV was able to improve model performance with 73.08% accuracy, 72.50% recall, and 79.42% AUC – Score. Therefore, this optimized Logistic Regression model has potential to make a significant contribution to the classification and prevention of cardiovascular disease more accurately.

**Keywords :** Cardiovascular Disease, Logistic Regression, Grid Search CV, Classification

## LEMBAR PERSEMBAHAN

Puji Syukur kehadirat Allah SWT, karena berkat rahmat dan karunia-Nya sehingga penulis bisa menyelesaikan skripsi ini dengan baik. Penulis juga ingin mengucapkan terima kasih kepada :

1. Allah SWT, Tuhan Yang Maha Esa. Terima kasih yang tak terhingga penulis ucapkan karena telah memudahkan setiap kesulitan, dan selalu memberikan rejeki yang datangnya tak terduga.
2. Nenek dan tante yang selalu mengusahakan keinginan penulis. Tanpa mereka, penulis tidak akan bisa sampai di titik sekarang.
3. Almarhumah Ibu dan Almarhumah Ayah. Meski fisiknya tidak lagi di sisi, penulis percaya bahwa ibu dan ayah akan selalu mendukung dan mendoakan penulis kapan pun dan dimana pun. Selamat beristirahat dengan tenang Bu, Pak. Maaf kalau belum bisa membahagiakan.
4. Saudara perempuan, pakde dan bude saya serta seluruh anggota keluarga yang lain. Terima kasih karena selalu memberikan dukungan kepada penulis. Selalu bersedia untuk direpotkan, kapan pun dan di mana pun.
5. Bapak Ir. Yufis Azhar, S. Kom, M. Kom selaku Dosen Pembimbing. Terima kasih Pak karena selalu sabar dan tenang dalam memberikan respon serta bimbingan yang baik.
6. Ibu Vinna Rahmayanti Setyaning Nastiti, S. Si, M.Si yang telah memberikan motivasi untuk menjalani kehidupan dengan kuat.
7. Bapak dan Ibu dosen, serta civitas akademika dari Informatika UMM yang telah memberikan ilmu dan pengetahuan kepada penulis serta membantu penulis selama berada di masa perkuliahan .
8. Sahabat saya. Dari grup “Jogja Gas!” dan “spill the tea”. Terima kasih atas motivasi yang selalu diberikan ketika penulis merasa di titik terendah.
9. Teman sekaligus sahabat saya di masa perkuliahan. Leta, irham, elna, maudy, aqif, akram, faruq dan semua yang tidak bisa disebutkan satu persatu. Terima kasih yang tak terhingga atas dukungan dan bantuan yang selalu diberikan.
10. Tidak lain tidak bukan, diri saya sendiri. Terima kasih untuk selalu kuat di segala kondisi. Terima kasih telah bertahan sampai sejauh ini.

## KATA PENGANTAR

Puji Syukur kehadiran Allah SWT, karena berkat rahmat dan karunia-Nya sehingga penulis bisa menyelesaikan skripsi yang berjudul :

**“Klasifikasi Penyakit Kardiovaskular Menggunakan Algoritma *Logistic Regression* Dengan Optimasi *Grid Search CV*”.**

Skripsi ini disusun sebagai salah satu syarat untuk meraih gelar Sarjana Strata 1 (S1) pada Program Studi Informatika, Fakultas Teknik, Universitas Muhammadiyah Malang. Penulis menyadari bahwa masih terdapat banyak kekurangan dan jauh dari kata sempurna. Oleh karena itu, kritik dan saran sangat diperlukan guna perbaikan di masa yang akan datang.

Akhir kata, semoga skripsi ini sedikit banyak dapat memberikan informasi dan manfaat kepada para pembaca. Terlebih bisa berkontribusi terhadap ilmu pengetahuan khususnya dalam bidang teknologi dan medis.

Malang, 25 September 2024



Namirah Rahmah Putri



## DAFTAR ISI

LEMBAR PERSETUJUAN.....	ii
LEMBAR PENGESAHAN .....	iii
LEMBAR PERNYATAAN .....	iv
ABSTRAK .....	v
ABSTRACT .....	vi
LEMBAR PERSEMBAHAN .....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR .....	xi
DAFTAR TABEL.....	xii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	3
1.3. Tujuan Penelitian.....	4
1.4. Batasan Masalah.....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Studi Literatur.....	5
2.2 Penyakit Kardiovaskular .....	7
2.3 Machine Learning.....	7
2.3.1 <i>Supervised Learning</i> .....	8
2.3.2 Unsupervised Learning .....	8
2.3.3 Semi-supervised .....	8
2.3.4 Reinforcement .....	8
2.4 Klasifikasi.....	9
2.4.1 Klasifikasi Biner.....	9
2.4.2 Klasifikasi Multiclasss.....	9
2.4.3 Klasifikasi Multi-Label .....	9
2.5 Logistic Regression .....	9
2.6 Hyperparameter Tuning .....	10
2.7 Grid Search CV .....	10
2.8 Evaluasi Model.....	11
2.8.1 Akurasi .....	12
2.8.2 Presisi .....	12

2.8.3	Recall.....	12
2.8.4	F1-Score .....	12
BAB III	METODOLOGI PENELITIAN .....	13
3.1	RANCANGAN PENELITIAN .....	13
3.2	Dataset .....	14
3.3	Pemrosesan Data .....	15
3.3.1	Feature engineering.....	16
3.3.2	Data dropping.....	18
3.3.3	Feature selection.....	18
3.4	Pembagian Data.....	19
3.5	Logistic Regression .....	19
3.6	<i>Hyperparameter Tuning Grid Search CV</i> .....	19
3.7	Evaluasi Model.....	21
BAB IV	HASIL DAN PEMBAHASAN .....	22
4.1	Pemrosesan Data .....	22
4.1.1	Feature Engineering .....	22
4.1.2	Data dropping.....	23
4.1.3	Feature selection.....	24
4.2	Data Splitting.....	24
4.3	Hyperparameter Tuning .....	25
4.4	Evaluasi Model.....	25
4.4.1	Hasil Sebelum Grid Search CV.....	26
4.4.2	Hasil Sesudah Grid Search CV .....	27
4.4.3	Perbandingan Hasil .....	28
4.5	Analisa Perbandingan.....	29
4.6	Mc.Nemar Test.....	30
BAB V	KESIMPULAN .....	31
5.1	Kesimpulan .....	31
5.2	Saran.....	31
DAFTAR PUSTAKA	.....	32
LAMPIRAN	.....	36

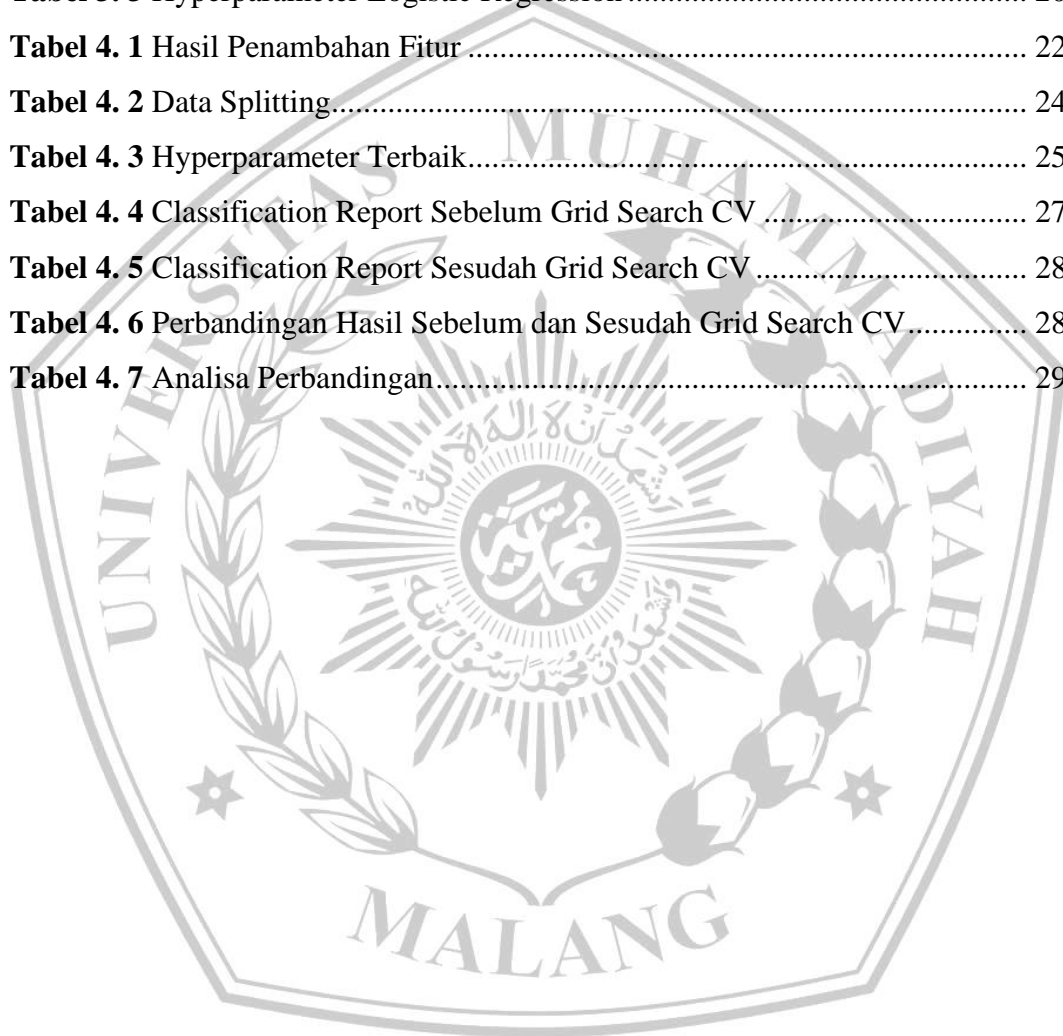
## DAFTAR GAMBAR

<b>Gambar 2. 1</b>	4 Tipe Machine Learning [19].....	8
<b>Gambar 2. 2</b>	Arsitektur Model Logistic Regression [21] .....	10
<b>Gambar 3. 1</b>	Rancangan Penelitian.....	13
<b>Gambar 4. 1</b>	Hyperparameter Tuning.....	25
<b>Gambar 4. 2</b>	Confusion Matrix Sebelum Grid Search CV .....	26
<b>Gambar 4. 3</b>	Confusion Matrix Sesudah Grid Search CV.....	27



## DAFTAR TABEL

<b>Tabel 2. 1</b> Studi Literature.....	5
<b>Tabel 2. 2</b> Confusion Matrix .....	11
<b>Tabel 3. 1</b> Informasi Detail Dataset.....	14
<b>Tabel 3. 2</b> Penambahan Fitur.....	16
<b>Tabel 3. 3</b> Hyperparameter Logistic Regression .....	20
<b>Tabel 4. 1</b> Hasil Penambahan Fitur .....	22
<b>Tabel 4. 2</b> Data Splitting.....	24
<b>Tabel 4. 3</b> Hyperparameter Terbaik.....	25
<b>Tabel 4. 4</b> Classification Report Sebelum Grid Search CV .....	27
<b>Tabel 4. 5</b> Classification Report Sesudah Grid Search CV .....	28
<b>Tabel 4. 6</b> Perbandingan Hasil Sebelum dan Sesudah Grid Search CV.....	28
<b>Tabel 4. 7</b> Analisa Perbandingan.....	29



## DAFTAR LAMPIRAN

Lampiran 1. Contoh Dataset .....	36
Lampiran 2. Dataset Setelah Pre Processing.....	37
Lampiran 3. Visualisasi Feature Selection.....	38



## DAFTAR PUSTAKA

- [1] Prevention of cardiovascular disease: guidelines for assessment and management of cardiovascular risk. Geneva: World Health Organization; 2021.
- [2] Juslim, R. Rukma and Herawati, Fauna (2018) Penyakit Kardiovaskular. In: Penyakit Kardiovaskular. Seri Pengobatan Rasional . Graha Ilmu, Yogyakarta, pp. 1-14. ISBN 978-602-262-869-9; E-ISBN: 978-602-262-870-5
- [3] Nugroho, A., Gumelar, A. B., Sooi, A. G., Sarvasti, D., & Tahalele, P. L. (2020). Perbandingan Performansi Algoritma Pengklasifikasian Terpadu Untuk Kasus Penyakit Kardiovaskular. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 4(5), 998-1006.
- [4] Redaksi Sehat Negeriku. 2021. Penyakit Jantung Koroner Didominasi Masyarakat Kota. Diakses pada 24 Oktober 2023 dari <https://sehatnegeriku.kemkes.go.id/baca/umum/20210927/5638626/penyakit-jantung-koroner-didominasi-masyarakat-kota/>
- [5] Dritsas, E., Alexiou, S., & Moustakas, K. (2022, April). Cardiovascular Disease Risk Prediction with Supervised Machine Learning Techniques. In *ICT4AWE* (pp. 315-321).
- [6] Ulianova S. "Cardiovascular Disease Dataset". Kaggle. 2019. <https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset/data>. Diakses pada 10 Oktober 2023
- [7] Honestya, G., Sajida, M., & Ramadhanu, A. (2024). Klasifikasi Jenis Daun Herbal Klasifikasi Jenis Daun Herbal Menggunakan Metode Logistic Regression dan Decision Tree Classifier Berdasarkan Fitur (Warna dan Bentuk). *Journal of Information System and Education Development*, 2(1), 52-55.
- [8] Panda, N. R. (2022). A review on logistic regression in medical research. *National Journal of Community Medicine*, 13(04), 265-270.
- [9] Boateng, E. Y., & Abaye, D. A. (2019). A review of the logistic regression model with emphasis on medical research. *Journal of data analysis and information processing*, 7(04), 190.

- [10] Pratama, Y. A., Budiman, F., Winarno, S., & Kurniawan, D. (2023). Analisis Optimasi Algoritma Decision Tree, Logistic Regression dan SVM Menggunakan Soft Voting. *Jurnal Media Informatika Budidarma*, 7(4), 1908-1919.
- [11] Ciu, T., & Oetama, R. S. (2020). Logistic regression prediction model for cardiovascular disease. *IJNMT (International Journal of New Media Technology)*, 7(1), 33-38.
- [12] Andi, A., Thamrin, T., Susanto, A., Wijaya, E., & Djohan, D. (2023). Analysis of the random forest and grid search algorithms in early detection of diabetes mellitus disease. *Jurnal Mantik*, 7(2), 1117-1124.
- [13] Ahmed Arafa, A. H., Radad, M., Badawy, M. M., & El-Fishawy, N. (2022). Logistic regression hyperparameter optimization for cancer classification. *Menoufia Journal of Electronic Engineering Research*, 31(1), 1-8.
- [14] Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Perbandingan metode naïve bayes dan support vector machine pada analisis sentimen twitter. *SMATIKA Jurnal: STIKI Informatika Jurnal*, 10(02), 71-76.
- [15] Jayaditya, I. K. A., & Kadyanan, I. G. A. G. A. Implementasi Random Forest Pada Klasifikasi Penyakit Kardiovaskular dengan Hyperparameter Tuning Grid Search.
- [16] Gunawan, M. I., Sugiarto, D., & Mardianto, I. (2020). Peningkatan Kinerja Akurasi prediksi penyakit diabetes mellitus menggunakan metode grid Search Pada algoritma logistic regression. *JEPIN (Jurnal Edukasi Dan Penelitian Informatika)*, 6(3), 280-284.
- [17] Yan, L. D., Sufra, R., St Sauveur, R., Jean-Pierre, M. C., Apollon, A., Malebranche, R., ... & Safford, M. (2024). Spectrum of prevalent cardiovascular diseases in urban Port-au-Prince, Haiti: a population-based cross-sectional study. *The Lancet Regional Health–Americas*, 33.
- [18] Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR)*. [Internet], 9(1), 381-386.
- [19] Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN computer science*, 2(3), 160.

- [20] Das, A. (2024). Logistic regression. In Encyclopedia of Quality of Life and Well-Being Research (pp. 3985-3986). Cham: Springer International Publishing.
- [21] Torres, R., Ohashi, O., & Pessin, G. (2019). A machine-learning approach to distinguish passengers and drivers reading while driving. *Sensors*, 19(14), 3174.
- [22] Owen, L. (2022). Hyperparameter Tuning with Python: Boost your machine learning model's performance via hyperparameter tuning. Packt Publishing Ltd.
- [23] Putri, T. A. E., Widiharih, T., & Santoso, R. (2023). Penerapan Tuning Hyperparameter Randomsearchcv Pada Adaptive Boosting Untuk Prediksi Kelangsungan Hidup Pasien Gagal Jantung. *Jurnal Gaussian*, 11(3), 397-406.
- [24] Elgeldawi, E., Sayed, A., Galal, A. R., & Zaki, A. M. (2021, November). Hyperparameter tuning for machine learning algorithms used for arabic sentiment analysis. In *Informatics* (Vol. 8, No. 4, p. 79). MDPI.
- [25] Anggoro, D. A., & Mukti, S. S. (2021). Performance Comparison of Grid Search and Random Search Methods for Hyperparameter Tuning in Extreme Gradient Boosting Algorithm to Predict Chronic Kidney Failure. *International Journal of Intelligent Engineering & Systems*, 14(6)
- [26] Tang, K. S., Medeiros, E. D., & Shah, A. D. (2020). Wide pulse pressure: a clinical review. *The Journal of Clinical Hypertension*, 22(11), 1960-1967.
- [27] Wu, Y., Li, D., & Vermund, S. H. (2024). Advantages and Limitations of the Body Mass Index (BMI) to Assess Adult Obesity. *International Journal of Environmental Research and Public Health*, 21(6), 757.
- [28] Yano, Y., Kim, H. C., Lee, H., Azahar, N., Ahmed, S., Kitaoka, K., ... & Viera, A. J. (2022). Isolated diastolic hypertension and risk of cardiovascular disease: controversies in hypertension-pro side of the argument. *Hypertension*, 79(8), 1563-1570.
- [29] Buyrukoğlu, S., & Akbaş, A. (2022). Machine learning based early prediction of type 2 diabetes: A new hybrid feature selection approach using correlation matrix with heatmap and SFS. *Balkan Journal of Electrical and Computer Engineering*, 10(2), 110-117.



- [30] Panda, N. R. (2022). A review on logistic regression in medical research. *National Journal of Community Medicine*, 13(04), 265-270.
- [31] Dinh, A., Miertschin, S., Young, A., & Mohanty, S. D. (2019). A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. *BMC medical informatics and decision making*, 19(1), 1-15.
- [32] Krittanawong, C., Virk, H. U. H., Bangalore, S., Wang, Z., Johnson, K. W., Pinotti, R., ... & Tang, W. W. (2020). Machine learning prediction in cardiovascular diseases: a meta-analysis. *Scientific reports*, 10(1), 16057.





UNIVERSITAS  
MUHAMMADIYAH  
MALANG



## FAKULTAS TEKNIK

### INFORMATIKA

informatika.umm.ac.id | informatika@umm.ac.id

### FORM CEK PLAGIARISME LAPORAN TUGAS AKHIR

Nama Mahasiswa : Namirah Rahmah Putri  
 NIM : 202010370311204  
 Judul TA : Klasifikasi Penyakit Kardiovaskular Menggunakan Algoritma  
*Logistic Regression Dengan Optimasi Grid Search CV*

#### Hasil Cek Plagiarisme dengan Turnitin

No.	Komponen Pengecekan	Nilai Maksimal Plagiarisme (%)	Hasil Cek Plagiarisme (%) *
1.	Bab 1 – Pendahuluan	10 %	5 %
2.	Bab 2 – Daftar Pustaka	25 %	8 %
3.	Bab 3 – Analisis dan Perancangan	25 %	2 %
4.	Bab 4 – Implementasi dan Pengujian	15 %	5 %
5.	Bab 5 – Kesimpulan dan Saran	5 %	0 %
6.	Makalah Tugas Akhir	20%	0 %

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

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

Mengetahui,

Pemeriksa (Staff TU)

(.....)



Kampus I  
 Jl. Bandung 1 Malang, Jawa Timur  
 P. +62 341 551 253 (Hunting)  
 F. +62 341 460 435

Kampus II  
 Jl. Bendungan Sutarni No. 155 Malang, Jawa Timur  
 P. +62 341 551 149 (Hunting)  
 F. +62 341 582 060

Kampus III  
 Jl. Raya Tlogomas No.248 Malang, Jawa Timur  
 P. +62 341 464 318 (Hunting)  
 F. +62 341 460 435  
 E: webmaster@umm.ac.id