

# KLASIFIKASI *MACHINE LEARNING* UNTUK MENDETEKSI PENYAKIT JANTUNG DENGAN *REGRESI LOGISTIK* DAN *RANDOM FOREST*

Teni Deinarosa Hermalia, Mulyati Eka Saputri, Indri Nur Sukmawati

Program Studi Sains Data  
Fakultas Sains dan Teknologi  
Universitas Koperasi Indonesia

Email: [tenidhermalia.xmma@gmail.com](mailto:tenidhermalia.xmma@gmail.com), [mulyatiekasaputri@gmail.com](mailto:mulyatiekasaputri@gmail.com),  
[sukmawatiindri6@gmail.com](mailto:sukmawatiindri6@gmail.com)

## ABSTRAK

Penelitian ini mengkaji penerapan algoritma *machine learning*, yaitu *Regresi Logistik* dan *Random Forest*, untuk klasifikasi dan deteksi penyakit jantung menggunakan dataset yang diperoleh dari *Kaggle*. Dataset ini terdiri dari 14 atribut, termasuk informasi pasien seperti usia, jenis kelamin, tekanan darah, kadar kolesterol, dan indikator medis lainnya. *Regresi Logistik* dikenal dengan kesederhanaannya dan asumsi linier, sementara *Random Forest* merupakan metode *ensemble* yang mampu menangani hubungan data yang kompleks dan non-linier. Hasil penelitian menunjukkan bahwa *Random Forest* memiliki performa lebih baik dibandingkan *Regresi Logistik*, dengan akurasi sebesar 98,54% dibandingkan 79,5%, presisi 97% dibandingkan 85%, dan *recall* 100% dibandingkan 72%. Temuan ini menegaskan keunggulan *Random Forest* dalam menangkap pola data yang rumit dan memberikan prediksi yang sangat akurat, sehingga menjadi algoritma yang direkomendasikan untuk deteksi penyakit jantung. Penelitian ini menekankan pentingnya pemilihan metode *machine learning* yang tepat untuk diagnosis yang efektif di bidang kesehatan.

**Kata Kunci:** *Machine Learning*, Penyakit Jantung, *Regresi Logistik*, *Random Forest*, Klasifikasi, Diagnosis Medis.

## 1. PENDAHULUAN

Jantung adalah organ utama yang berperan dalam memompa darah guna memastikan oksigen dan nutrisi terdistribusi ke seluruh tubuh. Apabila jantung mengalami gangguan, peredaran darah dalam tubuh dapat terganggu sehingga menjaga kesehatan jantung sangatlah penting agar terhindar dari berbagai jenis

penyakit jantung. Penyakit jantung merupakan salah satu penyakit pembunuh nomor satu di dunia dan di Indonesia. Dalam data yang dikeluarkan oleh WHO pada tahun 2021, kematian akibat penyakit jantung mencapai angka 17,8 juta kematian atau satu dari tiga kematian di dunia setiap tahun disebabkan oleh penyakit jantung.

Deteksi dini penyakit jantung sangat penting karena dapat meningkatkan peluang kesembuhan, mencegah komplikasi serius, dan menekan angka kematian. Menurut penelitian yang dipublikasikan di Jurnal Kesehatan Komunitas Indonesia, deteksi dini melalui pemeriksaan tekanan darah dan penilaian risiko menggunakan Jakarta Cardiovascular Score efektif dalam mengidentifikasi individu dengan risiko tinggi penyakit jantung.

Penyebab penyakit jantung pada umumnya terdapat dua faktor risiko yaitu faktor risiko yang tidak dapat diubah dan dapat diubah. Faktor yang tidak dapat diubah meliputi genetika, gender, dan usia. Sedangkan faktor yang dapat diubah yaitu merokok, kolesterol tinggi, hipertensi, diabetes mellitus, kegemukan, stres, dan kurang berolahraga. (Erlin Kurnia, Bambang Prayogi, 2015). Terdapat empat gejala umum penyakit jantung atau kardiovaskular, yaitu sering mengalami nyeri dada terutama dada sebelah kiri, mengalami irama atau detak jantung yang tidak teratur, detak jantung yang sangat cepat dan kadang-kadang malah tidak terdeteksi denyut sama sekali pada nadi dan mengalami sesak napas (Putu and Dewi, 2016). Salah satu subset dari penyakit jantung yang paling mematikan di dunia adalah serangan jantung, yang menyumbang angka kematian signifikan di antara semua kasus penyakit jantung dengan angka kematian mencapai 12,90%.

Penerapan teknologi *machine learning* dalam dunia medis telah terbukti meningkatkan akurasi deteksi penyakit jantung. Penelitian yang diterbitkan dalam Jurnal Teknologi Sistem Informasi dan Aplikasi mengembangkan model *Support Vector Machine (SVM)* untuk klasifikasi penyakit jantung, yang berhasil meningkatkan akurasi diagnosis hingga 96,56%. Selain *Support Vector Machine (SVM)*, algoritma lain seperti *k-Nearest Neighbors (k-NN)* dan *Decision Tree* telah dibandingkan dalam penelitian untuk mendeteksi penyakit jantung. Berdasarkan hasil penelitian yang didapat bahwa pada *k-NN* didapatkan hasil akurasi sebesar 62%, kemudian untuk *Random Forest* didapat hasil akurasi sebesar 87% dan *Decision Tree* didapat nilai akurasi sebesar 90%. Menunjukkan bahwa pemilihan algoritma yang tepat dapat berpengaruh signifikan terhadap akurasi prediksi.

Identifikasi risiko penyakit jantung diperlukan untuk merencanakan pencegahan yang efektif. *Machine learning* digunakan karena mampu menganalisis data dengan banyak variabel, memberikan hasil yang lebih akurat, dan mempercepat

proses klasifikasi. Penelitian ini membandingkan algoritma *Regresi Logistik* dan *Random Forest* berdasarkan parameter dalam dataset. Perbandingan keduanya memungkinkan evaluasi efektivitas algoritma sederhana dan kompleks dalam mengklasifikasi risiko penyakit jantung.

## 2. TINJAUAN PUSTAKA

### 2.1 Penyakit Jantung

Jantung berfungsi memompa darah ke seluruh tubuh. Untuk itu otot jantung memerlukan oksigen dan nutrisi yang cukup. Oksigen dan nutrisi diangkut oleh darah melalui pembuluh darah khusus yang disebut arteri koroner.

Penyakit jantung biasanya terjadi karena kerusakan sel otot-otot jantung dalam memompa aliran darah keseluruh tubuh, yang disebabkan kekurangan oksigen yang dibawa darah ke pembuluh darah di jantung. Atau juga karena terjadi kejang pada otot jantung yang menyebabkan kegagalan organ jantung dalam memompa darah, sehingga menyebabkan kondisi jantung tidak dapat melaksanakan fungsinya dengan baik. (Eka Wahyudi, Sri Hartati, 2017).

Pada penelitian ini proses diagnosis klasifikasi penyakit jantung dilakukan dengan algoritma *machine learning* menggunakan *Regresi Logistik* dan *Random Forest*. Data yang digunakan bersumber dari Kaggle, dengan 14 atribut dimana variabel bebas nya adalah atribut informasi pasien dan variabel terikat nya adalah apakah mereka memiliki penyakit jantung atau tidak. Atribut pada dataset meliputi *age*, *sex*, *cp* (*chest pain type*), *trestbps*, *chol*, *fbs*, *restecg*, *thalach*, *exang*, *oldpeak*, *slope*, *ca* (jumlah pembuluh darah), *thal*, dan *target* (hasil akhir memiliki penyakit jantung atau tidak).

Dengan mengandalkan data yang komprehensif, seperti usia, jenis kelamin, dan hasil medis lainnya, teknologi ini dapat memberikan prediksi yang lebih cepat dan akurat, sehingga dapat membantu dalam penanganan dan pencegahan penyakit jantung secara lebih efektif.

### 2.2 Machine Learning

*Machine learning* adalah pembelajaran mesin yang sangat membantu dalam menyelesaikan masalah, membuat mudah dalam mengerjakan sesuatu. Di bidang rumah sakit atau bidang kesehatan, machine learning memudahkan dalam mengerjakan sesuatu, contohnya dokter bisa mendiagnosa penyakit jantung dalam waktu cepat tanpa memakan waktu yang lama. Bukan hanya pekerjaan dokter saja yang dibuat mudah, para medis juga diberi keuntungan yang sangat besar sehingga dalam melakukan pekerjaan mereka semakin mudah. *Machine learning* membuat pekerjaan dalam mengklasifikasi penyakit dibidang kesehatan menjadi mudah seperti,

mengetahui jenis penyakit dan memberikan hasil dalam bentuk gambar yang lebih optimal.

*Machine learning* bermula saat manusia memikirkan bagaimana cara agar komputer dapat belajar dari pengalaman atau dapat mengingat apa saja yang barusan dioperasikan di komputer tersebut. Hal tersebut terbukti pada tahun 1952, Arthur Samuel menciptakan sebuah program *game of checkers*, pada sebuah komputer IBM. Program tersebut dapat mempelajari gerakan untuk memenangkan permainan *checkers* dan menyimpan gerakan tersebut ke dalam memorinya. *Machine learning* dibuat untuk membantu manusia dalam menyelesaikan masalah serta tidak juga merepotkan dalam hal penggunaan *machine learning* karena dia bisa berjalan sendiri tanpa berulang ulang di install.

### 2.3 Algoritma Regresi Logistik

*Regresi logistik* merupakan algoritma pembelajaran mesin yang sangat terkenal setelah *regresi linier*. Keduanya memiliki banyak kesamaan, tetapi perbedaan utama terletak pada penggunaannya. *Regresi linier* digunakan untuk memprediksi atau memperkirakan nilai, sedangkan *regresi logistik* digunakan untuk menyelesaikan tugas klasifikasi. *Regresi logistik* berupaya mengidentifikasi hubungan antara variabel dependen (Y), yang bersifat dikotomis (misalnya, ya atau tidak), dengan satu atau lebih variabel independen yang dapat berupa kategori atau kontinu. Di pusat analisis *regresi logistik* adalah tugas memperkirakan peluang log dari suatu peristiwa melalui fungsi *regresi linier berganda* (Gunawan et al., 2020).

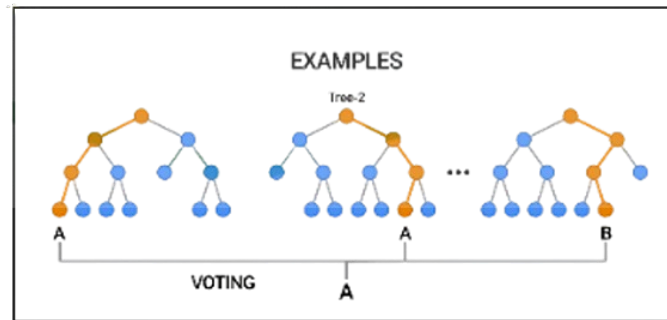
*Regresi Logistik* digunakan ketika variabel target yang akan diestimasi adalah biner atau kategorikal. Algoritma ini memodelkan probabilitas kategori yang berbeda berdasarkan kombinasi nilai-nilai prediktor, tujuannya adalah untuk memprediksi kategori atau kelas dari variabel target.

### 2.4 Algoritma Random Forest

*Random forest* adalah salah satu klasifikasi *machine learning* yang memiliki kinerja yang sangat tinggi, skabilitas yang baik serta mudah untuk digunakan. Model dari *random forest* sendiri adalah model ensemble atau model yang terdiri dari beberapa model *decision tree* dimana algoritma tersebut menggunakan subset *bootstrap* yang berasal dari sampel pemilihan yang acak model ensemble memiliki ide dasar dengan kombinasi model pembelajaran yang lemah dengan model pembelajaran yang kuat.

Lebih tepatnya *random forest* adalah penggabungan kombinasi algoritma dengan serangkaian pengklasifikasi *tree*, dengan setiap *tree* memberikan suara untuk kelas yang paling populer sehingga mendapatkan hasil pengurutan akhir. *Random forest* juga biasanya dilatih dengan menggunakan metode '*bagging*' yang merupakan

kumpulan dari beberapa *decision tree* yang memiliki kombinasi model pembelajaran untuk dapat meningkatkan hasil keseluruhan data.



**Gambar 2.4** *Random Forest*

Algoritma *random forest* diringkas dengan beberapa urutan sebagai berikut :

1. Sampel bootstrap pada ukuran  $n$  secara acak dilakukan dengan teknik sampling dari data input
2. Pada setiap *node* dilakukan hal dengan fitur sebanyak  $d$  tanpa dilakukan replacement, pecahan node digunakan pada fitur untuk memberikan split yang baik berdasarkan pada fungsi objektif salah satunya adalah gain.
3. Mengulangi langkah (1) dan (2)
4. Hasil prediksi dari *decision tree* dapat menjadi keputusan mengenai kelas data.

## 2.5 Perbandingan Regresi Logistik dan Random Forest

Algoritma *regresi logistik* dan *random forest*, keduanya memiliki keunggulan dan kekurangan yang berbeda sesuai dengan karakteristik data yang digunakan. *Regresi logistik*, yang dikenal karena kesederhanaannya, cocok untuk aplikasi dengan data yang memiliki hubungan linier antara fitur dan target variabel. Keunggulan utama dari *regresi logistik* adalah kemudahan dalam pemahaman dan interpretasi model, sehingga sering dipilih ketika transparansi hasil diperlukan. Namun, *regresi logistik* kurang efektif jika data memiliki hubungan yang kompleks dan *non-linear* antara variabel independen dan dependen. Di sisi lain, *random forest*, sebagai algoritma *ensemble*, lebih unggul dalam menangani data yang lebih besar, kompleks, dan *non-linear*. *Random forest* dapat mengatasi *noise* dan interaksi antar fitur dengan lebih baik, namun kelemahan utamanya terletak pada kurangnya interpretabilitas, karena keputusan dibuat berdasarkan banyaknya pohon keputusan yang terlibat dalam model. Secara umum, *regresi logistik* lebih cocok untuk data yang sederhana dan hubungan linier, sementara *random forest* lebih unggul pada data dengan interaksi fitur yang rumit dan kompleksitas yang lebih tinggi (Breiman, 2001; Herwanto, 2019; Andriani & Aulia, 2020).

### 3. METODE PENELITIAN

#### 3.1 Jenis Penelitian

Penelitian ini menggunakan algoritma *machine learning* (*Regresi Logistik* dan *Random Forest*) dalam mendeteksi penyakit jantung, sehingga hasilnya dapat digunakan untuk aplikasi praktis di bidang kesehatan.

#### 3.2 Dataset

- Sumber dataset : Dataset yang digunakan dalam penelitian ini bersumber dari [www.kaggle.com](http://www.kaggle.com)
- Deskripsi Dataset  
Dataset berisi data pasien dengan atribut sebagai berikut:
  - Variabel Independen : Usia, jenis kelamin, tekanan darah, kadar kolesterol, detak jantung maksimal, dll.
  - Variabel Dependen : Diagnosis penyakit jantung (1 = ada penyakit jantung , 0 = tidak ada penyakit jantung).
- Persiapan Dataset : Persiapan data dilakukan dengan memeriksa kelengkapan dataset, termasuk menangani nilai yang hilang (*missing values*) dan memastikan format data sesuai untuk analisis. Dataset kemudian dibagi menjadi data latih (*training set*) dan data uji (*testing set*), menggunakan rasio 80:20, untuk memastikan validitas hasil model.

#### 3.3 Tahap Penelitian

Tahap penelitian ini dilakukan secara bertahap dengan gambaran diagram alir seperti dalam gambar berikut :



Penelitian ini dimulai dengan studi literatur untuk memahami konsep penyakit jantung dan algoritma klasifikasi *machine learning*, seperti *regresi logistik* dan *random forest*. Selanjutnya, dilakukan pembuatan program yang mengimplementasikan algoritma tersebut menggunakan bahasa pemrograman tertentu. Pengambilan Dataset ini dari sumber terpercaya dan berisi parameter

kesehatan pasien yang akan digunakan sebagai input untuk model. Program yang dibuat diuji dengan dataset tersebut melalui tahap *preprocessing*, pelatihan, validasi, dan pengujian. Hasil akhirnya berupa evaluasi performa model, seperti akurasi dan sensitivitas, untuk menentukan algoritma terbaik dalam mendeteksi penyakit jantung.

### 3.4 Preprocessing Data

Pada tahap preprocessing data, seluruh kolom kecuali kolom target digunakan sebagai fitur independen (X), sementara kolom target dijadikan variabel dependen (Y) dengan nilai 0 untuk tidak ada penyakit jantung dan 1 untuk ada penyakit jantung. Data kemudian dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian menggunakan fungsi train test split, dengan parameter random state = 42 untuk menjaga konsistensi pembagian. Selanjutnya, data diperiksa untuk memastikan tidak ada nilai hilang atau anomali, jika ditemukan anomali maka dilakukan pembersihan atau imputasi data. Fitur numerik dinormalisasi agar skala antarvariabel seragam, sementara fitur kategori di-*encode* agar dapat digunakan oleh algoritma *machine learning*. Proses ini memastikan data dalam kondisi optimal untuk pelatihan dan evaluasi model.

## 4. HASIL PENELITIAN DAN PEMBAHASAN

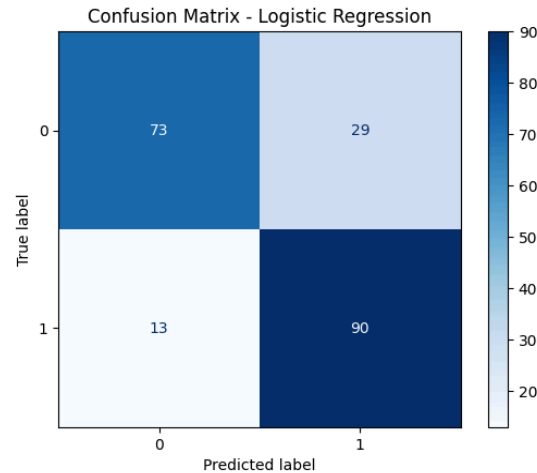
### 4.1. Karakteristik Akurasi *machine learning Regresi Logistik* dan *Random Forest*.

Penelitian ini memperoleh hasil untuk mengkararakteristik *machine learning* berdasarkan regresi logistik dan random forest. Parameter yang dihitung berupa akurasi, presisi dan *recall* untuk setiap *classifier*. Akurasi pada *machine learning* adalah rasio prediksi benar (positif dan negatif) dari keseluruhan data. Jika akurasi pada algoritma tersebut bagus maka data set yang memiliki jumlah data untuk *False Negative* dan *False Positive* mendekati simetris.

#### 4.1.1 Akurasi Regresi Logistik

Model *Logistic Regression* memiliki akurasi sebesar 79.5%, menunjukkan bahwa model mampu memprediksi dengan cukup baik.

Berdasarkan *confusion matrix*, terdapat 73 prediksi benar untuk kelas 0 dan 90 prediksi benar untuk kelas 1, dengan kesalahan berupa 29 *false positives* (salah prediksi ada penyakit jantung) dan 13 *false negatives* (salah prediksi tidak ada penyakit jantung). Secara keseluruhan, model cukup baik untuk mendeteksi penyakit jantung, terutama pada kelas positif (ada penyakit jantung).

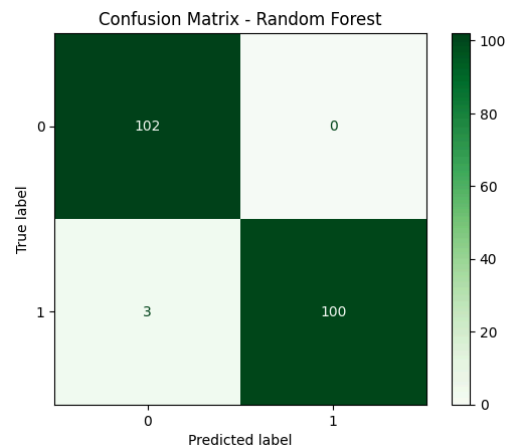


Gambar 4.1 *confussion matrix regresi logistic*

#### 4.1.2 Akurasi *Random Forest*

Model *Random Forest* memiliki akurasi 98.54%, menunjukkan bahwa model ini sangat baik dalam memprediksi data dengan tingkat kesalahan yang sangat kecil.

Dari *confusion matrix*, model menghasilkan 102 prediksi benar untuk kelas 0 dan 100 prediksi benar untuk kelas 1, dengan hanya 3 *false negatives* (salah memprediksi ada penyakit jantung sebagai tidak ada). Secara keseluruhan, model *Random Forest* menunjukkan performa yang sangat tinggi dan lebih unggul dibanding *Logistic Regression* pada dataset ini.



Gambar 4.2 *confussion matrix random forest*



## 4.2 Hasil karakteristik presisi machine learning *Logistic Regression* dan *Random Forest*.

*Precision* merupakan rasio prediksi benar positif dibandingkan keseluruhan data yang benar positif. Sebagai contoh, berapa prediksi pasien yang benar-benar sakit jantung dari total keseluruhan yang diprediksi positif.

### 4.2.1 *Precision Logistic Regression*

Untuk hasil *precision classification report* pada *Logistic Regression* di dapatkan untuk kelas 0 (tidak ada penyakit jantung) adalah 85%, sedangkan untuk kelas 1 (ada penyakit jantung) adalah 76%, artinya model lebih akurat dalam memprediksi pasien yang tidak memiliki penyakit jantung.

precision	
0	0.85
1	0.76

**Gambar 4.3** Hasil *Precision* pada *Logistic Regression*

### 4.2.2 *Precision Random Forest*

Untuk hasil *precision classification report* pada *Random Forest* untuk kelas 0 (tidak ada penyakit jantung) adalah 97%, sementara kelas 1 (ada penyakit jantung) mencapai 100%, yang berarti model sangat akurat dalam memprediksi pasien dengan penyakit jantung.

precision	
0	0.97
1	1.00

**Gambar 4.4** Hasil *Precision* pada *Random Forest*

## 4.3 Hasil karakteristik *recall* machine learning *Logistic Regression* dan *Random Forest*.

*Recall* merupakan rasio untuk prediksi benar positif yang dibandingkan dengan keseluruhan data yang benar positif. Sebagai contoh beberapa pasien yang sakit jantung dari keseluruhan pasien yang sebenarnya sakit jantung.

#### 4.3.1 Recall Logistic Regression

Untuk hasil *recall classification report* pada *Logistic Regression* untuk kelas 0 sebesar 72%, sementara kelas 1 mencapai 87%, menandakan model lebih baik dalam mendeteksi pasien dengan penyakit jantung.

	precision	recall
0	0.85	0.72
1	0.76	0.87

Gambar 4.5 Hasil Recall pada Logistic Regression

#### 4.3.2 Recall Random Forest

Untuk hasil *recall classification report* pada *Logistic Regression* untuk kelas 0 sebesar 100%, dan kelas 1 sebesar 97%, menandakan model hampir sempurna dalam mendeteksi pasien tanpa penyakit jantung dan sangat baik dalam mendeteksi pasien dengan penyakit jantung.

	precision	recall
0	0.97	1.00
1	1.00	0.97

Gambar 4.6 Hasil Recall pada Random Forest

#### 4.4 Perbandingan Akurasi, Precision, dan Recall

Perbandingan *Logistic Regression* dan *Random Forest* untuk nilai akurasi , *precision* dan *recall* diambil pada kelas 0 dari *classification report*. Didapatkan hasil perbandingan seperti pada tabel 4.1.

Tabel 4.1 Tabel Perbandingan *Logistic Regression* dan *Random Forest*

Model	Akurasi	Precision	Recall
Logistic Regression	79.5 %	0.85	0.72
Random Forest	98.54 %	0.97	1.00

## 5. KESIMPULAN

Berdasarkan hasil pengujian dan analisis yang telah dilakukan oleh peneliti, maka dapat diambil kesimpulan.

1. Nilai akurasi, precision dan *recall* pada *classification report* untuk model *Logistic Regression* dan *Random Forest* didapatkan akurasi pada *Logistic Regression* sebesar (79.5%), presisi sebesar (85 %), *recall* sebesar (72 %), *Random Forest* untuk akurasi sebesar (98.54%), presisi sebesar (97%), *recall* sebesar (100%).
2. Berdasarkan nilai akurasi menunjukkan bahwa model *Random Forest* memiliki akurasi 98.54%, yang jauh lebih tinggi dibandingkan model *Logistic Regression* dengan akurasi 79.51%. Hal ini menunjukkan bahwa *Random Forest* lebih baik dalam menangkap pola kompleks pada *dataset*, sehingga menghasilkan prediksi yang lebih akurat. *Logistic Regression* memiliki keterbatasan karena asumsinya linier, sedangkan *Random Forest*, dengan pendekatan *ensemble* dan banyak pohon keputusan, mampu menangani hubungan *non-linier* antara fitur dan target dengan lebih baik. Oleh karena itu, untuk *dataset* ini, *Random Forest* lebih direkomendasikan karena performanya yang jauh lebih unggul.

## 6. DAFTAR PUSTAKA

1. Pradana, D. G., Alghifari, M. L., Juna, M. F., & Palaguna, S. D. (2022). Klasifikasi penyakit jantung menggunakan metode artificial neural network. *Indonesian Journal of Data and Science (IJODAS)*, 3, 55–60.
2. Retnoningsih, E., & Pramudita, R. (2020). Mengenal machine learning dengan teknik supervised dan unsupervised learning menggunakan Python. *Bina Insani ICT Journal*, 7, 156–165.
3. Mardini, N. I., Marlana, L., & Azhar, E. (2019). Regresi logistik pada model problem-based learning berbantu software Cabri 3D. *Jurnal Mercumatika: Jurnal Penelitian Matematika dan Pendidikan Matematika*, 4, 64–70.
4. 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.
5. Heryadi, Y., & Wahyuno, T. (2020). *Machine learning: Konsep dan implementasi* (1st ed.). Yogyakarta: Penerbit Gava Media.
6. Fatmawati. (2017). *Identifikasi karakteristik penyakit jantung poliklinik RSU Koroner di Bahteramas Provinsi Sulawesi Utara*.
7. Telaumbanua, F. D., Hulu, P., Nadeak, T. Z., Lumbantong, R. R., & Dharma, A. (2020). Penggunaan machine learning di bidang kesehatan. *Jurnal Teknologi dan*

- Ilmu Komputer Prima (JUTIKOMP)*, 2(2), 391–399.  
<https://doi.org/10.34012/jutikomp.v2i2.657>
8. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.  
<https://doi.org/10.1023/A:1010933404324>
  9. Herwanto, E. (2019). Perbandingan kinerja regresi logistik dan random forest dalam klasifikasi data kesehatan. *Jurnal Teknologi dan Sistem Komputer*, 8(2), 100-109.  
Diakses dari <https://jtsiskom.undip.ac.id/index.php/jtsiskom/article/view/1257>
  10. Andriani, L., & Aulia, F. (2020). Perbandingan regresi logistik dan random forest dalam klasifikasi cuaca wilayah Jawa Tengah. *Jurnal Ilmiah Informatika*, 4(1), 45-51.  
Diakses dari <https://ejournal.undip.ac.id/index.php/jii/article/view/31718>
  11. Liu, Y., Wang, Y., & Zhang, J. (2012). New machine learning algorithm: Random forest. In *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 7473, pp. 246–252).  
[https://doi.org/10.1007/978-3-642-34062-8\\_32](https://doi.org/10.1007/978-3-642-34062-8_32)
  12. Alodokter. (n.d.). Penyakit jantung koroner - Gejala, penyebab, dan pengobatan.  
Retrieved from <https://www.alodokter.com>
  13. Universitas Teknokrat Indonesia. (n.d.). *Satu dari tiga kematian disebabkan oleh jantung: Ayo cegah serangan jantung*. Retrieved from  
<https://jim.teknokrat.ac.id/index.php/informatika/article/view/1866>
  14. DQLab. (2022). Studi kasus random forest machine learning untuk pemula data.  
Retrieved from <https://dqlab.id>
  15. Universitas Gadjah Mada. (n.d.). *Random forest untuk data mining*. Retrieved from  
<https://journal.ugm.ac.id/ijccs/article/view/15523/11717>
  16. Pencegahan dan penyembuhan penyakit jantung. (n.d.). Retrieved from  
[https://www.google.co.id/books/edition/Pencegahan\\_penyembuhan\\_penyakit\\_jantung/\\_eJcZg-gGLuWC?hl=en&gbpv=1&dq=jantung&pg=PA55&printsec=frontcover](https://www.google.co.id/books/edition/Pencegahan_penyembuhan_penyakit_jantung/_eJcZg-gGLuWC?hl=en&gbpv=1&dq=jantung&pg=PA55&printsec=frontcover)
  17. Google Colab. (n.d.). Retrieved from  
<https://colab.research.google.com/drive/1mIgwITpeBnQUqfUD5FFCkjCdqw58Pa5?usp=sharing#scrollTo=bLQ9UMH7cHO0>