Jurnal Manajemen Informatika Jayakarta
Volume 5, Nomor 4, Desember 2025: halaman: 311-322

P-ISSN: 2746-5985
e-ISSN: 2797-0930

DOI: https://doi.org/10.52362/jmijayakarta.v5i4.2102

PREDIKSI RISIKO PENYAKIT JANTUNG SEDERHANA MENGGUNAKAN ALGORITMA RANDOM FOREST CLASSIFIER DENGAN DATA GAYA HIDUP SISWA

[1.] Muji Santoso*, [2.] Muhammad Zydan K., [3.] Iwan, [4.] Mudrika

[1, 2, 3, 4] Program Studi Teknik Informatika, Fakultas Teknologi Industri
Universitas Gunadarma
Jalan Margonda Raya No. 100, Depok, Indonesia

E-Mail : * muji_santoso@staff.gunadarma.ac.id [1], mzydank01@gmail.com [2], iwan27skommt@gmail.com [3], mudrika@staff.gunadarma.ac.id [4]

Abstrak

Penyakit jantung merupakan salah satu penyebab utama kematian di dunia yang risikonya dapat dipengaruhi oleh gaya hidup tidak sehat. Penelitian ini bertujuan mengembangkan model prediksi sederhana risiko penyakit jantung pada siswa menggunakan algoritma Random Forest Classifier dengan data gaya hidup yang diperoleh dari Kaggle. Dataset mencakup jam belajar, jam ekstrakurikuler, jam tidur, jam bersosialisasi, aktivitas fisik, tingkat stres, jenis kelamin, dan nilai akademik. Proses penelitian meliputi pra-pemrosesan data, pelatihan model, evaluasi, serta pengujian pada data uji dan kasus individual. Hasil penelitian menunjukkan model mampu mencapai akurasi, precision, recall, dan f1-score sempurna (1.00). Analisis feature importance menegaskan bahwa tingkat stres, aktivitas fisik, dan jam belajar merupakan faktor paling berpengaruh terhadap risiko. Temuan ini menunjukkan bahwa model dapat digunakan sebagai alat deteksi dini sederhana, meskipun masih terbatas karena label risiko ditentukan dengan aturan sederhana, bukan data medis aktual.

Kata kunci : prediksi, penyakit, jantung, gaya hidup siswa, Random Forest

Abstract

Heart disease is one of the leading causes of death worldwide, and its risk can be influenced by unhealthy lifestyles. This study aims to develop a simple heart disease risk prediction model for students using the Random Forest Classifier algorithm with lifestyle data obtained from Kaggle. The dataset includes study hours, extracurricular hours, sleep duration, social interaction, physical activity, stress level, gender, and

This work is licensed under a Creative Commons Attribution 4.0 International License.
http://journal.stmikjayakarta.ac.id/index.php/JMIJayakarta

Jurnal Manajemen Informatika Jayakarta
Volume 5, Nomor 4, Desember 2025: halaman: 311-322
P-ISSN: 2746-5985
e-ISSN: 2797-0930

DOI: https://doi.org/10.52362/jmijayakarta.v5i4.2102

academic performance. The research process involved data preprocessing, model training, evaluation, and testing on both test data and individual cases. Results showed that the model achieved perfect accuracy, precision, recall, and f1-score (1.00). Feature importance analysis highlighted that stress level, physical activity, and study hours were the most influential factors in risk prediction. These findings suggest that the model can serve as a simple early detection tool, although it remains limited since the risk labels were based on heuristic rules rather than real medical data

Keywords *: prediction, heart disease, student lifestyle, Random Forest*

1. PENDAHULUAN *(or Introduction)*

Penyakit jantung merupakan salah satu penyebab utama kematian di dunia, termasuk di kalangan remaja dan dewasa. Menurut *World Health Organization* (WHO), penyakit kardiovaskular menyumbang sekitar 17,9 juta kematian setiap tahunnya. Faktor risiko penyakit jantung tidak hanya terkait dengan usia lanjut, tetapi juga dipengaruhi oleh gaya hidup tidak sehat seperti kurang tidur, stres berlebihan, kurang aktivitas fisik, dan pola makan tidak seimbang (WHO, 2021) **[1]**

Di kalangan pelajar, gaya hidup yang tidak sehat semakin meningkat seiring dengan tuntutan akademik yang tinggi. Kebiasaan seperti kurang tidur, belajar berlebihan, minim aktivitas fisik, dan manajemen stres yang buruk dapat meningkatkan risiko gangguan kesehatan jantung di masa depan. Sebuah penelitian oleh Alotaibi dkk. (2020) menunjukkan bahwa tingkat stres yang tinggi pada siswa berkorelasi dengan peningkatan tekanan darah dan detak jantung, yang merupakan indikator awal masalah kardiovaskular (Alotaibi dkk. 2020) **[2]**

Namun, tantangan utamanya berada pada strategi pengobatan dan intervensi yang digunakan untuk penyakit jantung dimulai terlambat karena beberapa alasan, seperti kurangnya kesadaran, gejala, motivasi, atau kesalahpahaman (Salah & Srinivas 2022) **[3]** Oleh karena itu, diperlukan pendekatan gaya hidup siswa seperti jam belajar, jam tidur, aktivitas fisik, tingkat stres, dan nilai akademik untuk memprediksi risiko penyakit jantung secara sederhana

Salah satu algoritma yang menunjukkan kinerja tinggi dalam prediksi penyakit jantung adalah Random Forest Classifier. Penelitian oleh Abdulraheem menunjukkan bahwa algoritma ini mencapai akurasi hingga 98% dalam memprediksi penyakit jantung koroner (Abdulraheem dkk. 2021) **[4]**

2. TINJAUAN LITERATUR *(or Literature Review)*

This work is licensed under a Creative Commons Attribution 4.0 International License.
http://journal.stmikjayakarta.ac.id/index.php/JMIJayakarta

Jurnal Manajemen Informatika Jayakarta
Volume 5, Nomor 4, Desember 2025: halaman: 311-322
P-ISSN: 2746-5985
e-ISSN: 2797-0930

DOI: https://doi.org/10.52362/jmijayakarta.v5i4.2102

2.1. Penyakit Jantung dan Faktor Risikonya

Penyakit jantung merupakan istilah umum untuk berbagai kondisi yang mempengaruhi jantung dan pembuluh darah. Ini termasuk penyakit arteri koroner (penyempitan pembuluh darah yang memasok darah ke jantung), stroke, gagal jantung, dan aritmia (gangguan irama jantung) (Benjamin et al., 2023) **[5].** Beban global penyakit jantung sangat besar, menjadi penyebab utama kematian dan disabilitas di seluruh dunia (Roth et al., 2020) **[6]** Di Indonesia, prevalensi penyakit jantung juga terus meningkat, menunjukkan tantangan kesehatan masyarakat yang signifikan (Kementerian Kesehatan RI, 2018) **[7]**

Penyakit jantung yang timbul seringkali multifaktorial, melibatkan interaksi kompleks antara faktor genetik dan lingkungan. Namun, faktor risiko gaya hidup memainkan peran krusial dalam perkembangan dan progresivitas penyakit jantung. Faktor-faktor risiko utama yang dapat dimodifikasi meliputi pola makan tidak sehat yang tinggi lemak jenuh, kolesterol, dan garam; kurangnya aktivitas fisik atau gaya hidup sedenter; kebiasaan merokok; dan konsumsi alkohol berlebihan (Malhotra dkk. 2021) **[8]**. Selain itu, kondisi medis seperti hipertensi (tekanan darah tinggi), diabetes mellitus, dan dislipidemia (kadar lipid abnormal dalam darah) seringkali terkait erat dengan gaya hidup dan secara signifikan meningkatkan risiko penyakit jantung (Arnett dkk. 2019) **[9]**. Identifikasi dan pengelolaan faktor-faktor risiko ini menjadi landasan penting dalam strategi pencegahan primer dan sekunder penyakit jantung.

2.2. Prediksi Risiko Penyakit Jantung

Prediksi risiko penyakit jantung merupakan langkah penting dalam upaya deteksi dini dan pencegahan penyakit kardiovaskular. Dalam beberapa tahun terakhir, metode pembelajaran mesin telah banyak diterapkan untuk meningkatkan akurasi prediksi penyakit jantung (Rahim dkk. 2023) **[10]**

Deteksi dini penyakit jantung sangat penting untuk mencegah komplikasi serius seperti gagal jantung, stroke, atau kematian mendadak. Namun, proses diagnosis konvensional biasanya memerlukan prosedur yang kompleks dan memakan waktu cukup lama, serta sangat bergantung pada keterampilan tenaga medis. Seiring dengan kemajuan teknologi informasi, khususnya di bidang kecerdasan buatan, pendekatan berbasis pembelajaran mesin kini banyak diterapkan dalam bidang medis sebagai solusi pendukung diagnosis (Handayani dkk. 2025) **[11]**

2.3. Machine Learning

Kemajuan teknologi machine learning (ML) memberikan pendekatan baru yang efektif dalam memprediksi risiko penyakit jantung. ML mampu memproses data medis kompleks dan mencari pola tersembunyi yang sulit dideteksi secara manual. Salah satu metode yang efektif adalah algoritma Random Forest yang digabungkan dengan teknik SMOTEENN untuk

This work is licensed under a Creative Commons Attribution 4.0 International License.

Jurnal Manajemen Informatika Jayakarta
Volume 5, Nomor 4, Desember 2025: halaman: 311-322
P-ISSN: 2746-5985
e-ISSN: 2797-0930

DOI: https://doi.org/10.52362/jmijayakarta.v5i4.2102

mengatasi ketidakseimbangan data, sehingga dapat meningkatkan akurasi prediksi penyakit jantung hingga 94% (Rahmada & Susanto, 2025) **[12].**

Selain itu, algoritma Support Vector Machine (SVM) juga digunakan dalam sistem prediksi penyakit jantung berbasis daring dengan akurasi yang cukup tinggi, sekitar 85% (Hidayat dkk. 2024) **[13**]. Implementasi machine learning dalam bidang kardiologi memberikan harapan untuk deteksi dini dan pengelolaan penyakit jantung yang lebih baik serta efisien secara klinis.

2.4. Algoritma Random Forest Classifier

Random Forest (RF) adalah algoritma ensemble learning yang kuat dan fleksibel untuk tugas klasifikasi dan regresi. RF bekerja dengan membangun sejumlah besar pohon keputusan (decision trees) selama proses pelatihan. Untuk tugas klasifikasi, output dari model RF adalah kelas yang menjadi mode dari kelas yang diprediksi oleh masing-masing pohon. Beberapa keunggulan algoritma RF meliputi ketahanannya terhadap overfitting, kemampuannya untuk menangani data dengan dimensi tinggi dan variabel yang berkorelasi, serta memberikan ukuran pentingnya fitur (*feature importance*). Algoritma RF telah berhasil diterapkan dalam berbagai domain, termasuk prediksi penyakit jantung, dengan menunjukkan kinerja yang kompetitif atau bahkan lebih baik dibandingkan dengan algoritma ML lainnya (Hossain dkk, 2023) **[14]**

2.5. Python

Python adalah bahasa pemrograman tingkat tinggi yang banyak digunakan dalam ilmu data karena sintaksnya yang sederhana dan mudah dipahami. Python didukung oleh ekosistem pustaka yang kuat seperti NumPy, Pandas, dan Scikit-Learn. Dalam konteks analisis sentimen, bahasa ini menyediakan berbagai modul dan fungsi yang memfasilitasi seluruh proses mulai dari pembersihan data teks, pembuatan representasi vektor menggunakan TF-IDF, hingga tahap pembangunan dan evaluasi model klasifikasi. Fleksibilitas yang dimiliki Python memungkinkan peneliti menulis kode secara efisien dan mengelola data dalam jumlah besar dengan kompleksitas yang tetap terkontrol.

Python memungkinkan pengembangan program pengolahan data secara cepat dan efektif, terutama dengan dukungan pustaka numerik seperti NumPy dan struktur data tabular lewat Pandas. Kombinasi ini menyediakan fondasi teknis yang kokoh untuk mengelola dataset ulasan pengguna secara terstruktur dan berulang, serta mempermudah proses eksperimen dalam membangun model prediktif yang bisa menyesuaikan dengan karakteristik data teks yang dianalisis (McKinney, 2022) **[15]**

This work is licensed under a Creative Commons Attribution 4.0 International License.

Jurnal Manajemen Informatika Jayakarta
Volume 5, Nomor 4, Desember 2025: halaman: 311-322
P-ISSN: 2746-5985
e-ISSN: 2797-0930

DOI: https://doi.org/10.52362/jmijayakarta.v5i4.2102

3. METODE PENELITIAN *(or Research Method)*

Metode yang digunakan adalah pendekatan kuantitatif dengan pendekatan *machine learning*. Data gaya hidup dan status penyakit jantung akan dikumpulkan dari sampel populasi yang relevan. Algoritma Random Forest Classifier akan diimplementasikan menggunakan perangkat lunak statistik dan machine learning seperti Python dengan library Scikit-learn. Proses pemodelan akan melibatkan tahapan pengumpulan data, prapemrosesan data, pembuatan model dan pengujian serta evaluasi kinerja model.

1. **Pengumpulan Data**
   Tahap awal dalam penelitian ini melibatkan pengumpulan data gaya hidup siswa yang relevan untuk prediksi risiko penyakit jantung. Data ini diperoleh dari berkas student_lifestyle_dataset..csv yang telah disediakan dari situs kaggle.

2. **Prapemrosesan Data**
   Setelah data dikumpulkan, tahap selanjutnya adalah pra-pemrosesan data untuk memastikan kualitas dan kesiapan data sebelum digunakan dalam pemodelan. Pada tahap ini, dilakukan pengecekan terhadap nilai yang hilang (missing values) di seluruh dataset, dan hasilnya menunjukkan bahwa tidak terdapat nilai yang hilang sehingga data dapat digunakan secara utuh.

3. **Pembuatan Model dan Pengujian**
   Tahap inti dari analisis ini adalah pembuatan dan pengujian model prediktif. Algoritma yang digunakan adalah Random Forest Classifier, yaitu metode ensemble yang menggabungkan banyak decision trees untuk meningkatkan akurasi dan mengurangi risiko overfitting.

4. **Evaluasi Model**
   Evaluasi dilakukan untuk menilai performa model dalam mengklasifikasikan risiko penyakit jantung. Berdasarkan hasil pengujian, model menunjukkan akurasi yang sangat tinggi yaitu 1.0000 (100%).

4. HASIL DAN PEMBAHASAN *(or Results and Analisys)*

4.1. Tahap Pengumpulan Data

Tahap awal dalam penelitian ini melibatkan pengumpulan data gaya hidup siswa yang relevan untuk prediksi risiko penyakit jantung. Data ini diperoleh dari berkas student_lifestyle_dataset..csv yang telah disediakan dari situs kaggle dengan link tautan https://www.kaggle.com/datasets/charlottebennett1234/lifestyle-factors-and- their-impact-on-students?resource=download yang mencakup berbagai atribut seperti durasi jam belajar,

This work is licensed under a Creative Commons Attribution 4.0 International License.

Jurnal Manajemen Informatika Jayakarta
Volume 5, Nomor 4, Desember 2025: halaman: 311-322
P-ISSN: 2746-5985
e-ISSN: 2797-0930

DOI: https://doi.org/10.52362/jmijayakarta.v5i4.2102

jam ekstrakurikuler, jam tidur, jam bersosialisasi, jam aktivitas fisik per hari, tingkat stres, jenis kelamin, dan nilai akademis siswa. Data ini menjadi pondasi utama untuk membangun dan melatih model machine learning yang akan memprediksi risiko penyakit jantung sederhana, seperti yang ditunjukkan pada Gambar 1

```
Cek Missing Values:
Student_ID                          0
Study_Hours_Per_Day                 0
Extracurricular_Hours_Per_Day       0
Sleep_Hours_Per_Day                 0
Social_Hours_Per_Day                0
Physical_Activity_Hours_Per_Day     0
Stress_Level                        0
Gender                              0
Grades                              0
Heart_Disease_Risk                  0
dtype: int64
```

Gambar 1. Data Awal Siswa

4.2. Tahap Prapemrosesan Data

Setelah data dikumpulkan, tahap selanjutnya adalah pra-pemrosesan data untuk memastikan kualitas dan kesiapan data sebelum digunakan dalam pemodelan, lihat pada Gambar 2. Pada tahap ini, dilakukan pengecekan terhadap nilai yang hilang (missing values) di seluruh dataset, dan hasilnya menunjukkan bahwa tidak terdapat nilai yang hilang sehingga data dapat digunakan secara utuh. Selanjutnya dilakukan pemilihan fitur dan label: kolom seperti Study_Hours_Per_Day, Extracurricular_Hours_Per_Day, Sleep_Hours_Per_Day, Social_Hours_Per_Day, Physical_Activity_Hours_Per_Day, Stress_Level, Gender, dan Grades dipilih sebagai fitur (X). Karena dataset asli tidak menyediakan label risiko penyakit jantung, dibuatlah kolom baru Heart_Disease_Risk
(y) berdasarkan aturan sederhana, yaitu siswa dikategorikan berisiko (1) jika memiliki tingkat stres tinggi (High), jam tidur kurang dari 6 jam per hari, atau aktivitas fisik kurang dari 2 jam per hari; selain kondisi tersebut dikategorikan tidak berisiko (0). Kolom Student_ID dihilangkan karena tidak relevan dengan pemodelan. Fitur kategorikal seperti Stress_Level dan Gender diubah menjadi representasi numerik menggunakan teknik one-hot encoding agar dapat diproses oleh algoritma. Setelah itu, dataset dibagi menjadi data pelatihan dan data pengujian dengan rasio 80:20, yaitu 1600 sampel untuk pelatihan dan 400 sampel untuk pengujian.

This work is licensed under a Creative Commons Attribution 4.0 International License.

Jurnal Manajemen Informatika Jayakarta
Volume 5, Nomor 4, Desember 2025: halaman: 311-322
P-ISSN: 2746-5985
e-ISSN: 2797-0930

DOI: https://doi.org/10.52362/jmijayakarta.v5i4.2102

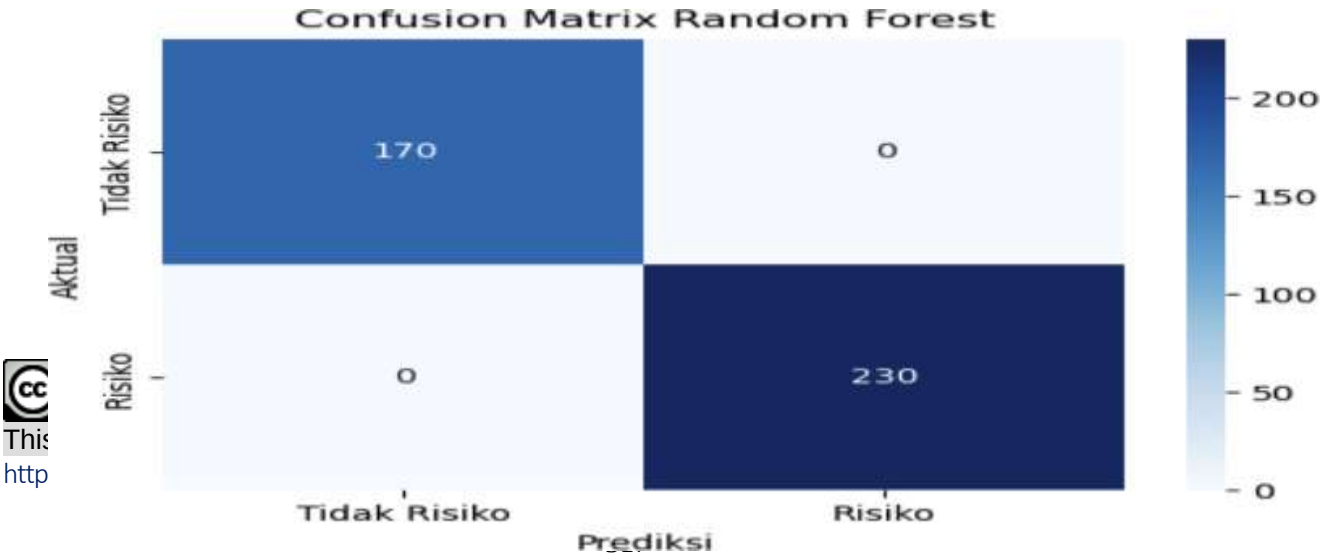
Data Awal dengan Label:

	Student_ID	Study_Hours_Per_Day	Extracurricular_Hours_Per_Day	Sleep_Hours_Per_Day	Social_Hours_Per_Day	Physical_Activity_Hours_Per_Day	Stress_Level	Gender	Grades	Heart_Disease_Risk
0	1	6.9	3.8	8.7	2.8	1.8	Moderate	Male	7.48	1
1	2	5.3	3.5	8.0	4.2	3.0	Low	Female	6.88	0
2	3	5.1	3.9	9.2	1.2	4.6	Low	Male	6.68	0
3	4	6.5	2.1	7.2	1.7	6.5	Moderate	Male	7.20	0
4	5	8.1	0.6	6.5	2.2	6.6	High	Male	8.78	1

Gambar 2. Pengelompokan Data Siswa Berdasarkan Aktivitas Harian

4.3. Tahap Pembuatan Model dan Pengujian

Tahap inti dari analisis ini adalah pembuatan dan pengujian model prediktif seperti dtunjukkan pada Gambar 3 Algoritma yang digunakan adalah Random Forest Classifier, yaitu metode ensemble yang menggabungkan banyak decision trees untuk meningkatkan akurasi dan mengurangi risiko overfitting. Model diinisialisasi dengan 100 trees (n_estimators=100) dan dilatih menggunakan data pelatihan (X_train dan y_train) yang telah melalui tahapan preprocessing. Proses pelatihan ini memungkinkan model untuk mempelajari pola hubungan antara atribut gaya hidup siswa (seperti jam tidur, aktivitas fisik, tingkat stres, dan nilai akademik) dengan label risiko penyakit jantung yang sudah ditentukan. Setelah pelatihan, model diuji pada data pengujian (X_test), dan hasil prediksi (y_pred) kemudian dibandingkan dengan label aktual (y_test) untuk menilai sejauh mana model mampu mengenali pola yang benar.



This
http

Jurnal Manajemen Informatika Jayakarta
Volume 5, Nomor 4, Desember 2025: halaman: 311-322
P-ISSN: 2746-5985
e-ISSN: 2797-0930

DOI: https://doi.org/10.52362/jmijayakarta.v5i4.2102

Gambar 3. Kalasifikasi Random Forest

4.4. Tahap Evaluasi Model

Evaluasi dilakukan untuk menilai performa model dalam mengklasifikasikan risiko penyakit jantung. Berdasarkan hasil pengujian, model menunjukkan akurasi yang sangat tinggi yaitu 1.0000 (100%). Hal ini berarti seluruh sampel pada data pengujian berhasil diprediksi dengan benar tanpa ada kesalahan klasifikasi. Hasil confusion matrix mendukung temuan ini dengan memperlihatkan bahwa semua kasus risiko rendah dan risiko tinggi diklasifikasikan sesuai label aktual. Laporan klasifikasi (classification report) juga menunjukkan metrik yang sempurna, dengan nilai precision, recall, dan f1-score sebesar 1.00 untuk kedua kelas. Lebih lanjut, analisis feature importance dari Random Forest mengungkapkan bahwa fitur Grades (nilai akademik) dan Sleep_Hours_Per_Day (jam tidur per hari) merupakan variabel yang paling berkontribusi dalam prediksi risiko. Hasil sempurna ini kemungkinan besar dipengaruhi oleh kuatnya hubungan antara label risiko (yang didasarkan pada tingkat stres, jam tidur, dan aktivitas fisik) dengan variabel-variabel gaya hidup siswa, sehingga model dapat membedakan kedua kelas risiko dengan sangat jelas.

Untuk menguji kapabilitas model dalam skenario nyata, dilakukan prediksi pada para siswa secara individual dengan menampilkan seluruh atribut gaya hidup mereka secara vertikal beserta hasil prediksi risiko yang ditunjukkan pada Gambar 4. Dari uji coba ini, model mampu memberikan klasifikasi yang konsisten dengan pola data masing-masing siswa. Sebagai contoh, siswa dengan jam tidur rendah (<6 jam) dan tingkat stres tinggi diprediksi memiliki risiko penyakit jantung, sementara siswa dengan tidur cukup dan aktivitas fisik teratur diprediksi tidak berisiko. Pada Gambar 5, bahwa tingkat stres lebih dominan sebagai penyebab penyakit jantung pada anak.

This work is licensed under a Creative Commons Attribution 4.0 International License.

Jurnal Manajemen Informatika Jayakarta
Volume 5, Nomor 4, Desember 2025: halaman: 311-322
P-ISSN: 2746-5985
e-ISSN: 2797-0930

DOI: https://doi.org/10.52362/jmijayakarta.v5i4.2102

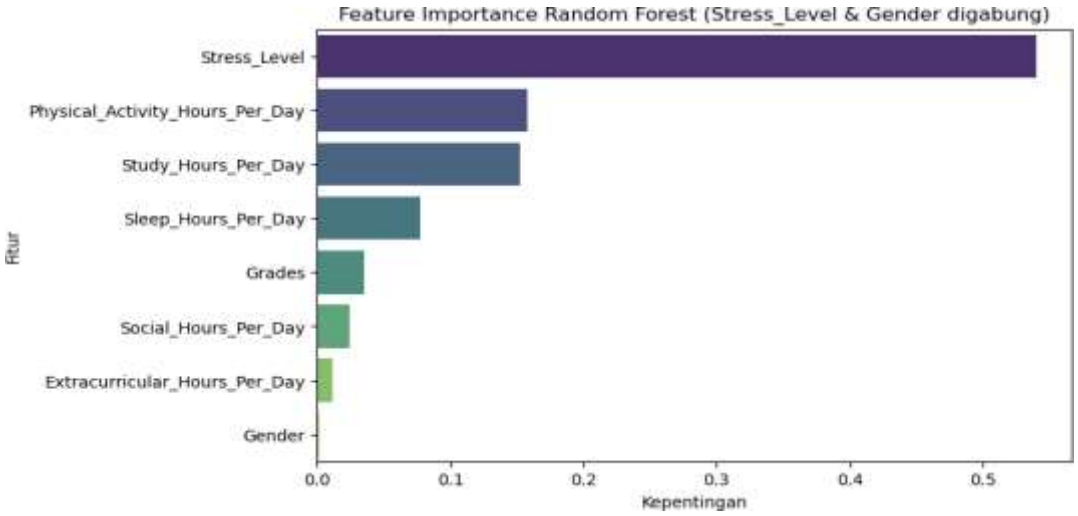
```
Prediksi untuk 5 siswa:

Data Siswa ke- 1
Study_Hours_Per_Day                   6.2
Extracurricular_Hours_Per_Day         4.0
Sleep_Hours_Per_Day                   6.0
Social_Hours_Per_Day                  3.5
Physical_Activity_Hours_Per_Day       4.3
Stress_Level                     Moderate
Gender                             Female
Grades                                7.2
Name: 1860, dtype: object
Prediksi Risiko Penyakit Jantung: Rendah

Data Siswa ke- 2
Study_Hours_Per_Day                   5.9
Extracurricular_Hours_Per_Day         2.2
Sleep_Hours_Per_Day                   9.9
Social_Hours_Per_Day                  5.6
Physical_Activity_Hours_Per_Day       0.4
Stress_Level                          Low
Gender                             Female
Grades                                7.8
Name: 353, dtype: object
Prediksi Risiko Penyakit Jantung: Tinggi

Data Siswa ke- 3
Study_Hours_Per_Day                   9.6
Extracurricular_Hours_Per_Day         2.3
Sleep_Hours_Per_Day                   8.3
Social_Hours_Per_Day                  3.5
Physical_Activity_Hours_Per_Day       0.3
Stress_Level                         High
Gender                             Female
Grades                               8.52
Name: 1333, dtype: object
Prediksi Risiko Penyakit Jantung: Tinggi

Data Siswa ke- 4
Study_Hours_Per_Day                   7.2
Extracurricular_Hours_Per_Day         0.5
Sleep_Hours_Per_Day                   7.6
Social_Hours_Per_Day                  2.1
Physical_Activity_Hours_Per_Day       6.6
Stress_Level                     Moderate
Gender                             Female
Grades                               7.78
Name: 905, dtype: object
Prediksi Risiko Penyakit Jantung: Rendah

Data Siswa ke- 5
Study_Hours_Per_Day                   9.0
Extracurricular_Hours_Per_Day         3.6
Sleep_Hours_Per_Day                   9.9
Social_Hours_Per_Day                  0.3
Physical_Activity_Hours_Per_Day       1.2
Stress_Level                         High
Gender                               Male
Grades                               7.98
Name: 1289, dtype: object
Prediksi Risiko Penyakit Jantung: Tinggi
```

Gambar 4. Uji Coba Pada Siswa

This work is licensed under a Creative Commons Attribution 4.0 International License.

Jurnal Manajemen Informatika Jayakarta
Volume 5, Nomor 4, Desember 2025: halaman: 311-322
P-ISSN: 2746-5985
e-ISSN: 2797-0930

DOI: https://doi.org/10.52362/jmijayakarta.v5i4.2102



Gambar 5. Tingkat Pengaruh

5. KESIMPULAN *(or Conclusion)*

Penelitian ini berhasil membangun model prediksi risiko penyakit jantung sederhana pada siswa menggunakan algoritma Random Forest dengan memanfaatkan data gaya hidup yang diperoleh dari Kaggle. Proses preprocessing berjalan dengan baik tanpa adanya data yang hilang, dilanjutkan dengan pemilihan fitur penting seperti jam tidur, aktivitas fisik, tingkat stres, jam belajar, nilai akademik, serta beberapa faktor lain. Hasil pengujian menunjukkan bahwa model mampu memberikan performa sempurna dengan akurasi, precision, recall, dan f1-score sebesar 1.00 pada data uji. Analisis *feature importance* mengungkapkan bahwa tingkat stres, aktivitas fisik, dan jam belajar merupakan faktor utama yang memengaruhi risiko penyakit jantung, diikuti oleh jam tidur, nilai akademik, dan faktor gaya hidup lainnya. Selain itu, uji coba pada data individual juga menunjukkan konsistensi hasil prediksi yang mudah dipahami, sehingga model ini berpotensi digunakan sebagai sistem deteksi dini berbasis data gaya hidup siswa.

REFERENSI *(Reference)*

[1] WHO. (2025). Cardiovascular diseases. World Health Organization. [Online] Available: https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1

[2] Alotaibi AD, Alosaimi FM, Alajlan AA, Bin Abdulrahman KA. The relationship between sleep quality, stress, and academic performance among medical students. J Family Community Med. 2020 Jan-Apr;27(1):23-28. doi: 10.4103/jfcm.JFCM_132_19. Epub 2020 Jan 13. PMID: 32030075; PMCID:

This work is licensed under a Creative Commons Attribution 4.0 International License.

DOI: https://doi.org/10.52362/jmijayakarta.v5i4.2102

PMC6984036.

[3] Salah, H., & Srinivas, S. (2022). Explainable machine learning framework for predicting long-term cardiovascular disease risk among adolescents. Scientific reports, 12(1), 21905. https://doi.org/10.1038/s41598-022-25933-5

[4] Abdulraheem, F., Mustafa, M., & Ahmed, M. F. (2021). Heart disease prediction using machine learning algorithms: A comparative study. International Journal of Advanced Computer Science and Applications, 12(7), 277–285.

[5] Benjamin, E. J., Muntner, P., Alonso, A., Bittencourt, M. S., Callaway, C. W., Carson, A. P., ... & Virani, S. S. (2023). Heart Disease and Stroke Statistics—2023 Update: A Report From the American Heart Association. Circulation, 147(8), e93–e621.

[6] Roth, G. A., Mensah, G. A., Johnson, C. O., Addolorato, G., Ammirati, E., Baddour, L. M.& Fuster, V. (2020). Global Burden of Cardiovascular Diseases and Risk Factors, 1990–2019: Update From the GBD 2019 Study. Journal of the American College of Cardiology (JACC), 76(25), 2982–3021.

[7] Kemenkes Ri. (2018). Hasil Riset Kesehatan Dasar Tahun 2018. Kementrian Kesehatan Ri, 53(9), 1689–1699.

[8] Malhotra, R. K., Ramar, K., Carden, K. A., et al. (2021). Sleep is essential to health: an American Academy of Sleep Medicine position statement. *Journal of Clinical Sleep Medicine*, 17(10), 2115-2119.

[9] Arnett, D. K., Blumenthal, R. S., Albert, M. A., et al. (2019). 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation, 140(11), e596-e646.

[10] Rahman, M. M., Karim, A., & Islam, S. (2022). Non-clinical predictors of cardiovascular risk in adolescents: A machine learning approach. BMC Public Health, 22(1), 1025. https://doi.org/10.1186/s12889-022-12845-2

[11] Handayani, A., Salsabila, S., Firdausiyah, A., Setiawan, A., & Nia Nesicha, Y. (2025). Predictive Analysis Heart Disease Based on Machine Learning Using the Random Forest Algorithm. Journal of Artificial Intelligence and Engineering Applications (JAIEA), 4(3), 1980–1986. https://doi.org/10.59934/jaiea.v4i3.1060

[12] Rahmada, A. ., & Susanto, E. R. (2025). Peningkatan Akurasi Prediksi Penyakit

This work is licensed under a Creative Commons Attribution 4.0 International License.

P-ISSN: 2746-5985
e-ISSN: 2797-0930

DOI: https://doi.org/10.52362/jmijayakarta.v5i4.2102

Jantung dengan Teknik SMOTEENN pada Algoritma Random Forest. Jurnal Pendidikan Dan Teknologi Indonesia, 4(12), 795-803. https://doi.org/10.52436/1.jpti.524

[13] Hidayat, R., Sy, Y. S., Sujana, T., Husnah, M., Saputra, H. T., & Okmayura, F. (2024). Implementasi Machine Learning Untuk Prediksi Penyakit Jantung Menggunakan Algoritma Support Vector Machine. BIOS : Jurnal Teknologi Informasi Dan Rekayasa Komputer, 5(2), 161-168. https://doi.org/10.37148/bios.v5i2.152

[14] Hossain, S., Hasan, M.K., Faruk, M.O. et al. Machine learning approach for predicting cardiovascular disease in Bangladesh: evidence from a cross-sectional study in 2023. BMC Cardiovasc Disord 24, 214 (2024). https://doi.org/10.1186/s12872-024-03883-2

[15] McKinney, W. (2022). Python for Data Analysis (3rd ed.). ISBN 9781098104030. O'Reilly Media. https://wesmckinney.com/book/

This work is licensed under a Creative Commons Attribution 4.0 International License.