



# Implementasi Prediksi Penyakit Jantung Menggunakan Data *Mining* Untuk Dunia Kesehatan

Mikhael Chandra Arta<sup>1</sup>, Nur Anwar<sup>1</sup>, Yulia Aneke Putri<sup>1</sup>, Suharjito<sup>1</sup>, Muhammad Asroll\*

<sup>1</sup>Industrial Engineering Department, BINUS Graduate Program – Master of Industrial Engineering, Bina Nusantara University, Jakarta, 11480, Indonesia

\*Corresponding author: [muhammad.asrol@binus.ac.id](mailto:muhammad.asrol@binus.ac.id)

## ARTICLE INFO

Received: 07-02-2024  
Revision: 22-03-2024  
Accepted: 07-04-2024

### Keywords:

*Data Mining*  
*Decision Tree*  
*Gradient Boosted*  
*Random Forest*  
*Recall*

## ABSTRACT

Jantung merupakan organ vital manusia yang sering menjadi penyebab kematian tertinggi. Penyakit jantung dapat diketahui dengan cara pemeriksaan dokter atau sejumlah tes kesehatan. Saat ini, perusahaan perlu memprediksi karyawannya yang kemungkinan memiliki riwayat atau calon pengidap penyakit jantung untuk mengurangi risiko kematian. Tindakan yang dilakukan dapat menggunakan pembelajaran mesin. Pembelajaran mesin memang dapat membantu dalam identifikasi awal penyakit dan meningkatkan hasil pengobatan. Sistem ini mampu memprediksi yang dapat membantu prediksi diagnosis penyakit jantung secara cepat dan akurat. Penelitian ini bertujuan untuk prediksi diagnosis penyakit jantung secara cepat dan akurat dengan menggunakan algoritma terbaik. Algoritma yang digunakan untuk melakukan prediksi yaitu *Decision Tree*, *Gradient Boosted*, dan *Random Forest*. Untuk prediksi, atribut yang digunakan adalah usia, jenis kelamin, tekanan darah, kolesterol, gula darah, detak jantung, jenis sakit dada dan tambahan adalah hasil pemeriksaan fisik lainnya. Dari hasil yang diperoleh, *Gradient Boosted* adalah algoritma yang memiliki AUC, presisi dan *recall* tertinggi dengan 86.6%.

## 1. PENDAHULUAN

Perkembangan ini, terdapat banyak berita yang membahas kesehatan. Baru-baru ini pemerintah Indonesia pun membuat “*Press Conference*” bersamaan dengan Hari Jantung Sedunia 2023. Pentingnya diketahui oleh masyarakat Indonesia tentunya untuk meningkatkan pemahaman terhadap kesehatan khususnya usia yang sudah tidak lagi muda. Jantung merupakan salah satu organ manusia yang vital selain dari otak. Jantung berfungsi sebagai pemompa darah untuk memenuhi kebutuhan oksigen dan nutrisi ke seluruh tubuh [1], [2]. Apabila jantung mengalami gangguan, peredaran darah dalam tubuh pun dapat terganggu. Ada sejumlah faktor yang meningkatkan risiko penyakit jantung. Beberapa diantaranya adalah sebagai berikut riwayat penyakit jantung dalam keluarga antara lain merokok, kolesterol, tekanan darah tinggi, obesitas dan kurangnya latihan fisik [3].

Pada bulan November 2023, Siloam Hospital membuat artikel mengenai penyebab kematian tertinggi di dunia, berdasarkan data *World Health Organization* (WHO). Sebanyak 85% kematian di dunia disebabkan oleh stroke dan serangan jantung yang rentan terjadi pada laki-laki usia lebih dari 45 tahun dan wanita lebih dari 50 tahun. Penyakit jantung yang hanya menyerang usia tua pun perlu dilakukan penelitian untuk membuktikan kebenarannya. Prediksi harus dilakukan untuk mengurangi risiko penyakit jantung. Diagnosis biasanya didasarkan pada tanda, gejala dan pemeriksaan fisik pasien. Memprediksi penyakit jantung mungkin saja menimbulkan praduga yang salah [4], [5]. Perusahaan menggunakan jasa rumah sakit untuk melakukan pemeriksaan fisik karyawan melalui program *medical checkup*. Data pemeriksaan biasanya menjadi keputusan perusahaan untuk penghematan biaya [6]. Kecerdasan manusia saja tidak cukup untuk diagnosis yang tepat, seperti hasil yang kurang akurat, pengalaman lebih sedikit, dan kinerja yang

bergantung pada waktu. Untuk meningkatkan akurasi prediksi kemungkinan manusia terserang penyakit jantung, pengembangan sistem pendukung keputusan menggunakan data *mining* akan menjadi topik utama penelitian ini [7].

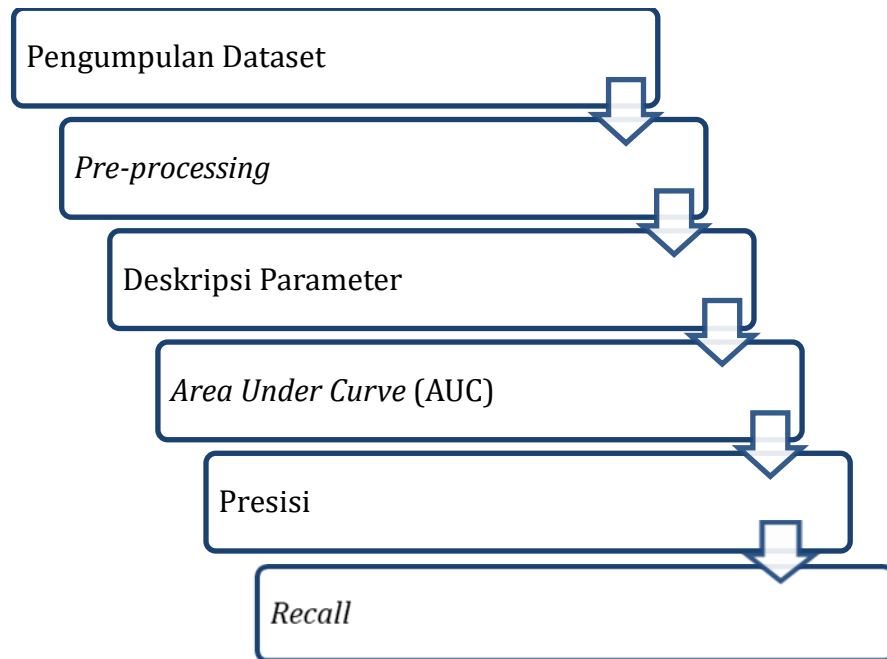
Beberapa penelitian telah dilakukan untuk mendiagnosis penyakit jantung. Mereka menggunakan berbagai macam teknik penambangan data (*data mining*) untuk mendiagnosis dan mencapai hasil yang berbeda untuk metode yang berbeda. Menurut Dangare et al, [8] mengembangkan sistem prediksi penyakit jantung (HDPS) menggunakan jaringan Neural. Untuk prediksi, sistem menggunakan jenis kelamin, tekanan darah, kolesterol, serta menambahkan dua parameter yaitu obesitas dan merokok untuk akurasi yang lebih baik. Hasil terlihat bahwa jaringan Neural dapat memprediksi penyakit jantung dengan akurasi hampir 100%. Penelitian oleh Palaniappan et al., [9] mengembangkan prototipe *Intelligent Heart Disease Prediction System* (IHDPS) dengan menggunakan *Decision Tree*, *Naïve Bayes*, dan *Neural Network*. Menggunakan profil medis seperti usia, jenis kelamin, tekanan darah dan gula darah dapat memprediksi kemungkinan pasien terkena penyakit jantung.

Beberapa penelitian lain oleh Haryadi et al., [10] mengembangkan sistem prediksi penyakit jantung menggunakan *K-Mean* klastering dengan tools rapid minner. Dengan menggunakan Rapdi minner penyakit jantung dapat dikelompokkan menjadi tiga kluster. *K-Mean* klastering juga pernah dilakukan untuk memetakan penyakit jantung di Rumah Sakit pemerintahan di Jakarta [11]. Selain itu, hasil penelitian Jindal et al., [12] mengembangkan sistem pembelajaran mesin untuk membuat keputusan yang andal dalam prediksi kardiovaskular. Algoritma ini untuk memprediksi terjadinya gagal jantung. Algoritma ini menggunakan *Naïve Bayes*, *Random Forest*, *Decision Trees*, *Logistic Regression*, *Support Vector Machines*, *K-Nearest Neighbors*, dan *Gradient Boosting*.

Berdasarkan beberapa studi pendahuluan maka penelitian ini bertujuan untuk prediksi diagnosis penyakit jantung secara cepat dan akurat dengan menggunakan algoritma terbaik. Kompleksitas masalah dapat dianalisis menggunakan pertimbangan jumlah data, waktu eksekusi, dan ruang yang dibutuhkan. Algoritma yang digunakan untuk melakukan prediksi yaitu *Decision Tree*, *Gradient Boosted*, dan *Random Forest*.

## 2. METODE PENELITIAN

Bagian ini berisi uraian jelas tentang metodologi yang dipergunakan dalam penelitian. Objek Penelitian ini menggunakan data yang diperoleh dari World Health Organization (WHO) yaitu berbentuk laporan penyakit jantung selama tahun 2023. Penelitian ini menggunakan data mining dimana akan menganalisis tiga metode yaitu *decision tree*, *gradient boosted tree* dan *random forest*. Tahapan penelitian ini dapat dilihat pada Gambar 1.



**Gambar 1.** Tahapan penelitian data *mining*

Penelitian ini menggunakan data mining dimana akan menganalisis tiga metode yaitu *decision tree*, *gradient boosted tree* dan *random forest*. Berikut penjelasan dari masing-masing algoritma:

### 2.1 Decision Tree

*Decision tree* ialah suatu teknik dalam data mining yang dipakai untuk mengklasifikasikan dan meramalkan dengan memanfaatkan peraturan-peraturan yang terwujud dalam format pohon keputusan [13]–[15]. Struktur pohon

keputusan terdiri dari simpul-simpul yang mewakili keputusan atau prediksi, beserta cabang-cabang yang mencerminkan potensi hasil atau atribut yang dipakai dalam proses pengambilan keputusan [16].

### 2.2 Gradient Boosted Tree

*Gradient Boosted Trees* adalah metode dalam pembelajaran ensemble yang menggabungkan beberapa pohon keputusan secara berurutan dengan fokus pada perbaikan kelemahan model sebelumnya. Pendekatan ini melibatkan pembangunan pohon keputusan satu per satu, di mana setiap pohon berusaha untuk mengoreksi kesalahan prediksi yang dibuat oleh model sebelumnya [17]. Prosesnya dimulai dengan membuat pohon keputusan pertama, yang mungkin belum optimal dalam performa. Kesalahan prediksi dari pohon pertama diidentifikasi, dan pohon keputusan kedua dibuat untuk memperbaiki kesalahan tersebut. Proses ini berlanjut dengan pembuatan pohon-pohon berikutnya, yang masing-masing fokus pada mengatasi kesalahan prediksi yang masih tersisa. Konsep "gradient" dalam *Gradient Boosted Trees* menunjukkan pendekatan penyesuaian bobot untuk meningkatkan optimasi kesalahan [18]. Bobot diberikan pada setiap pohon baru berdasarkan seberapa baik atau buruk pohon sebelumnya dalam menangani kesalahan tersebut. Dengan demikian, setiap pohon memberikan kontribusi untuk memperbaiki prediksi secara keseluruhan. Salah satu implementasi terkemuka dari *Gradient Boosted Trees* adalah *XGBoost (eXtreme Gradient Boosting)*, yang telah berhasil dalam berbagai tugas seperti klasifikasi dan regresi pada data terstruktur dan tidak terstruktur.

### 2.3 Random Forest

Metode klasifikasi *ensemble* merupakan algoritma pembelajaran yang membangun serangkaian pengklasifikasi daripada hanya satu pengklasifikasi, dan kemudian mengklasifikasikan titik data baru dengan mengambil suara dari prediksi mereka [19]–[21]. Pengklasifikasi *ensemble* yang paling umum digunakan melibatkan *Bagging*, *Boosting*, dan RF [22]. Prosesnya dapat dijelaskan sebagai berikut:

1. Pemilihan Sampel *Bootstrap (Bagging)*: Untuk setiap pohon keputusan, dataset asli diambil secara acak dengan penggantian, menghasilkan sampel bootstrapped.
2. Pembentukan Pohon Keputusan antara lain:
  - Setiap pohon keputusan dibangun dengan menggunakan sebagian dari sampel *bootstrapped*.
  - Saat membangun setiap simpul pada pohon, sejumlah fitur dipilih secara acak sebagai kandidat untuk pembagian (split).
  - Pemilihan fitur ini bertujuan untuk meningkatkan variasi di antara pohon-pohon dalam *ensemble*.
3. Voting (Suara): Ketika observasi baru harus diprediksi, setiap pohon memberikan suaranya, dan kelas yang paling sering muncul dianggap sebagai prediksi akhir.
4. Agregasi Hasil:
  - Untuk klasifikasi, hasil dari setiap pohon dijumlahkan atau dihitung untuk menentukan kelas mayoritas.
  - Untuk regresi, hasil prediksi dari setiap pohon diambil rata-rata untuk memberikan prediksi akhir.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Pengumpulan Data

Dataset yang digunakan dalam artikel ini diambil dari Kaggle.com. Dataset ini terdiri dari 12 atribut dan 918 data, seperti terlihat pada Tabel 1.

**Tabel 1.** Dataset

Age	Sex	Chest Pain Type	Resting BP	Cholesterol	Fasting BS	Resting ECG	Max HR	Exercise Angina	Oldpeak	ST_Slope	Heart Disease
40	M	ATA	140	289	0	Normal	172	N	0	Up	0
49	F	NAP	160	180	0	Normal	156	N	1	Flat	1
37	M	ATA	130	283	0	ST	98	N	0	Up	0
48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	1
54	M	NAP	150	195	0	Normal	122	N	0	Up	0

Atribut-atribut yang terdapat dalam dataset ini adalah sebagai berikut:

- a. Age: usia pasien (tahun)
- b. Sex: jenis kelamin (M: male, F: female)
- c. ChestPainType : jenis sakit dada (TA: Typical Angina, ATA: Atypical Angina, NAP: Non-Anginal Pain, ASY: Asymptomatic)
- d. RestingBP: tekanan darah saat relax (mmHg)
- e. Cholesterol: kadar kolesterol (mm/dl)
- f. FastingBP: gula darah saat puasa (1: jika fastingBS > 120 mg/dl, 0: selain dari itu)

- g. RestingECG: hasil electrocardiogram saat relax (Normal: normal, ST: terdapat ST-T wave anomaly, LVH: terdapat kondisi ventricular hypertrophy)
- h. MaxHR: maximum detak jantung (antara 60 – 202 bpm)
- i. ExerciseAngina: exercise-induced angina (Y: yes, N: no)
- j. Oldpeak: oldpeak = ST
- k. ST\_Slope: slope dari segmen ST (Up: upsloping, Flat: flat, Down: downsloping)
- l. HeartDisease: output (1: heart disease, 0: normal)

3.2 Pre-Processing

Untuk memastikan akurasi dan potensi performa tertinggi dari algoritma pembelajaran mesin, pra pemrosesan data sangat penting untuk analisa dan penelitian data. Untuk meningkatkan nilainya bagi penelitian, maka dilakukan pembersihan data. Pembersihan data (*Cleaning Data*) adalah mengoreksi nomor yang hilang, menghapus duplikat, dan menyelesaikan konflik, prosedur ini diperlukan untuk menyediakan kumpulan data yang bersih dan andal yang berfungsi sebagai dasar analisa yang benar dan hasil model yang dapat diprediksi. Rangkuman data pada Tabel 2 tidak ditemukan data hilang.

**Tabel 2** Pra Pemrosesan Dataset Penyakit Jantung

Heart Disease	Age	Sex	ChestPain Type	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	Exercise Angina	Oldpeak	ST_Slope
0	40	M	ATA	140	289	0	Normal	172	N	0.0	Up
1	49	F	NAP	160	180	0	Normal	156	N	1.0	Flat
0	37	M	ATA	130	283	0	ST	98	N	0.0	Up
1	48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat
0	54	M	NAP	150	195	0	Normal	122	N	0.0	Up
0	39	M	NAP	120	339	0	Normal	170	N	0.0	Up
0	45	F	ATA	130	237	0	Normal	170	N	0.0	Up
0	54	M	ATA	110	208	0	Normal	142	N	0.0	Up
1	37	M	ASY	140	207	0	Normal	130	Y	1.5	Flat
0	48	F	ATA	120	284	0	Normal	120	N	0.0	Up
0	37	F	NAP	130	211	0	Normal	142	N	0.0	Up
1	58	M	ATA	136	164	0	ST	99	Y	2.0	Flat
0	39	M	ATA	120	204	0	Normal	145	N	0.0	Up
1	49	M	ASY	140	234	0	Normal	140	Y	1.0	Flat
0	42	F	NAP	115	211	0	ST	137	N	0.0	Up
0	54	F	ATA	120	273	0	Normal	150	N	1.5	Flat
1	38	M	ASY	110	196	0	Normal	166	N	0.0	Flat
0	43	F	ATA	120	201	0	Normal	165	N	0.0	Up
1	60	M	ASY	100	248	0	Normal	125	N	1.0	Flat
1	36	M	ATA	120	267	0	Normal	160	N	3.0	Flat
0	43	F	TA	100	223	0	Normal	142	N	0.0	Up
0	44	M	ATA	120	184	0	Normal	142	N	1.0	Flat
0	49	F	ATA	124	201	0	Normal	164	N	0.0	Up
1	44	M	ATA	150	288	0	Normal	150	Y	3.0	Flat
0	40	M	NAP	130	215	0	Normal	138	N	0.0	Up
0	36	M	NAP	130	209	0	Normal	178	N	0.0	Up
0	53	M	ASY	124	260	0	ST	112	Y	3.0	Flat

Bagian utama ini adalah evaluasi hasil dan tingkat efisiensi yang dihasilkan oleh berbagai metode klasifikasi. Metrik ini memberikan rincian penting mengenai tingkat keberhasilan dan keakuratan pendekatan klasifikasi yang digunakan.

3.3. Deskripsi Parameter

Langkah pertama adalah melakukan analisa deskripsi terhadap data yang digunakan. Dapat dilihat pada Tabel 3 untuk analisa deskripsi parameter yang digunakan.

**Tabel 3** Deskripsi Parameter

Atribut	Jenis Atribut	Keterangan
Age	Numerical	Min: 28 Max: 77 Avg: 53.511 St Dev: 9.433
Chest Pain	Categorical	ASY: 496 (54.03%) NAP: 203 (22.11%) ATA: 173 (18.85%) TA: 46 (5.01%)

Cholesterol	Numerical	Min: 0 Max: 603 Avg: 198.8 St Dev: 109.4
Exercise Angina	Categorical	N: 547 (59.59%) Y: 371 (40.41%)
Fasting BS	Numerical	Min: 0 Max: 1 Avg: 0.223 St Dev: 0.423
Heart Disease	Categorical	Range 2 (No): 508 (55.34%) Range 1 (Yes): 410 (44.66%)
Max HR	Numerical	Min: 60 Max: 202 Avg: 136.809 St Dev: 25.460
Old Peak	Numerical	Min: -2.6 Max: 6.2 Avg: 0.887 St Dev: 1.067
Resting BP	Numerical	Min: -2.6 Max: 6.2 Avg: 0.887 St Dev: 1.067
Resting ECG	Categorical	Normal: 552 (60.13%) LVH: 188 (20.48%) ST: 178 (19.39%)
ST Slope	Categorical	Flat: 460 (50.11%) Up: 395 (43.03%) Down: 63 (6.86%)
Sex	Categorical	Male: 725 (78.98%) Female: 193 (21.02%)

### 3.4. Area Under Curve (AUC)

Tahap pembelajaran mesin, *Area Under Curve* AUC menentukan rasio proyeksi yang tepat terhadap semua perkiraan untuk mengevaluasi kompetensi model. AUC memiliki range nilai 0 sampai dengan 1. Dalam hal UAC, *Gradient Boosted* berkinerja lebih baik dibandingkan algoritma lainnya. Karena algoritma *Gradient Boosted* dapat mengelola interaksi yang rumit dan meminimalkan *overfitting*. Hasil AUC yang diperoleh adalah 0.923. Gambar 2 merupakan skor akurasi dari algoritma yang digunakan.

Scores						
Model	AUC	CA	F1	Prec	Recall	MCC
Tree	0.778	0.798	0.798	0.799	0.798	0.592
Gradient Boosting	0.923	0.866	0.866	0.866	0.866	0.728
Random Forest	0.917	0.854	0.854	0.854	0.854	0.705

**Gambar 2.** Skor Model Klasifikasi

### 3.5. Presisi

Ketepatan suatu klasifikasi terkadang disebut sebagai nilai prediksi positif, mengukur seberapa efektif pengklasifikasian tersebut memperkirakan kelas positif. Dengan metode *Gradient Boosted* menghasilkan skor presisi yang lebih besar. Dapat dilihat pada Gambar 2 skor presisi adalah 86,6%.

### 3.6. Recall

*Recall* adalah proporsi sampel data dari kelas yang diminati, juga dikenal kelas positif. Skor *recall* yang lebih tinggi dapat dicapai dengan mengoptimalkan jumlah prediksi positif yang sebenarnya dengan *Gradient Boosted* menghasilkan skor lebih besar yaitu 86,6%.

## 4. KESIMPULAN

Berdasarkan analisis dapat diketahui bahwa dengan sistem prediksi pembelajaran mesin yang cepat dan tepat membantu identifikasi masalah penyakit jantung. Kumpulan data yang digunakan untuk menilai adalah algoritma

*decision trees, Gradient Boosted, dan Random Forest.* Studi ini menemukan bahwa masing-masing algoritma menghasilkan pendekatan berbeda-beda, tentunya memiliki kekurangan dan kelebihan masing-masing. Dari hasil penelitian ini dapat disimpulkan bahwa metode *Gradient Boosting* adalah metode yang memiliki tingkat akurasi prediksi yang paling tinggi. Penelitian masa depan, dapat dilakukan dengan mengkaji lebih banyak algoritma untuk meningkatkan kegunaan dalam analisa klinis yang lebih mudah dan akurat seperti Algoritma Apriori yang merupakan salah satu algoritma pada data mining untuk mencari *frequent item/itemset* pada transaksional database.

## REFERENCES

- [1] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [2] W. Sudrajat, I. Cholid, and J. Petrus, "Penerapan Algoritma K-Means Clustering untuk Pengelompokan UMKM Menggunakan Rapidminer," *J. JUPITER*, vol. 14, no. 1, pp. 27–36, 2022.
- [3] M. R. Nugroho, I. E. Hendrawan, and P. P. Purwantoro, "Penerapan Algoritma K-Means Untuk Klasterisasi Data Obat Pada Rumah Sakit ASRI," *Nuansa Inform.*, vol. 16, no. 1, pp. 125–133, 2022, doi: 10.25134/nuansa.v16i1.5294.
- [4] T. Taslim and F. Fajrizal, "Penerapan algoritma k-mean untuk clustering data obat pada puskesmas rumbai," *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 7, no. 2, pp. 108–114, 2016, doi: 10.31849/digitalzone.v7i2.602.
- [5] D. A. Ramadhanty, R. Syafitri, E. Raswir, and D. Meisak, "Implementasi Data Mining Untuk Menentukan Persediaan Stok Obat Jurnal Informatika Dan Rekayasa Komputer ( JAKAKOM )," *J. Inform. dan Rekayasa Komput.*, vol. 1, no. 2, pp. 155–160, 2022.
- [6] Normah, B. Rifai, S. Vambudi, and R. Maulana, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Fashion Hijab Banten," *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.
- [7] P. M. S. Tarigan, J. T. Hardinata, H. Qurniawan, M. Safii, and R. Winanjaya, "Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang (Studi Kasus: Toko Sinar Harahap)," *G-Tech J. Teknol. Terap.*, vol. 7, no. 1, pp. 119–126, 2023, doi: 10.33379/gtech.v7i1.1938.
- [8] M. Chaitrali, S. Dangare, M. Sulabha, and S. Apte, "a Data Mining Approach for Prediction of Heart Disease Using Neural Networks," *Int. J. Comput. Eng. Technol. (IJCET)*, vol. 3, no. 3, pp. 30–40, 2012, [Online]. Available: <http://ssrn.com/abstract=2175569>
- [9] S. Palaniappan and R. Awang, "Intelligent heart disease prediction system using data mining techniques," *AICCSA 08 - 6th IEEE/ACS Int. Conf. Comput. Syst. Appl.*, no. December, pp. 108–115, 2008, doi: 10.1109/AICCSA.2008.4493524.
- [10] D. Haryadi and D. M. U. Atmaja, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Tingkat Risiko Penyakit Jantung," *J. Informatics Commun. Technol.*, vol. 3, no. 2, pp. 51–66, 2021, doi: 10.52661/j\_ict.v3i2.85.
- [11] H. Manurung and D. Irawati Sukma, "Aplikasi K-means cluster untuk pengelompokan unit kerja yang beresiko Covid-19: Studi kasus pada industri layanan kesehatan," *Oper. Excell. J. Appl. Ind. Eng.*, vol. 12, no. 2, pp. 1–10, 2020.
- [12] H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, "Heart disease prediction using machine learning algorithms," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1022, no. 1, pp. 1–10, 2021, doi: 10.1088/1757-899X/1022/1/012072.
- [13] E. Fauziningrum and E. I. Sulistyaningsih, "Penerapan Data Mining Metode Decision Tree Untuk Mengukur Penguasaan Bahasa Inggris Maritim (Studi Kasus Di Universitas Maritim Amni)," *J. Sains Dan Teknol. Marit*, vol. 22, no. 1, p. 41, 2021, doi: 10.33556/jstm.v22i1.285.
- [14] Normah, B. Rifai, S. Vambudi, and R. Maulana, "Implementasi Metode Decision Tree Klasifikasi Data Mining Untuk Prediksi Peminatan Jurusan Robotika oleh Mahasiswa," *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.
- [15] 2016 Banu G. R., "A Role of decision Tree classification data Mining Technique in Diagnosing Thyroid disease A Role of decision Tree classification data Mining Technique in Diagnosing Thyroid disease," *Int. J. Comput. Sci. Eng.*, vol. 4, no. 11, pp. 111–115, 2017, [Online]. Available: [https://www.researchgate.net/profile/Gulmohamed\\_Banu/publication/312167378\\_A\\_Role\\_of\\_decision\\_Tree\\_classification\\_data\\_Mining\\_Technique\\_in\\_Diagnosing\\_Thyroid\\_disease/links/5874ff9108ae6eb871c9a470/A-Role-of-decision-Tree-classification-data-Mining-Techniq](https://www.researchgate.net/profile/Gulmohamed_Banu/publication/312167378_A_Role_of_decision_Tree_classification_data_Mining_Technique_in_Diagnosing_Thyroid_disease/links/5874ff9108ae6eb871c9a470/A-Role-of-decision-Tree-classification-data-Mining-Techniq)
- [16] H. H. Patel and P. Prajapati, "Study and Analysis of Decision Tree Based Classification Algorithms," *Int. J. Comput. Sci. Eng.*, vol. 6, no. 10, pp. 74–78, 2018, doi: 10.26438/ijcse/v6i10.7478.
- [17] E. Firasari, U. Khultsum, M. N. Winnarto, and R. Risnandar, "Kombinasi K-NN dan Gradient Boosted Trees untuk Klasifikasi Penerima Program Bantuan Sosial," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 6, p. 1231, 2020, doi: 10.25126/jtiik.0813087.
- [18] S. E. Suryana, B. Warsito, and S. Suparti, "Penerapan Gradient Boosting Dengan Hyperopt Untuk Memprediksi Keberhasilan Telemarketing Bank," *J. Gaussian*, vol. 10, no. 4, pp. 617–623, 2021, doi: 10.14710/j.gauss.v10i4.31335.

- [19] R. Supriyadi, W. Gata, N. Maulidah, and A. Fauzi, "Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah," *E-Bisnis J. Ilm. Ekon. dan Bisnis*, vol. 13, no. 2, pp. 67–75, 2020, doi: 10.51903/e-bisnis.v13i2.247.
- [20] Suci Amaliah, M. Nusrang, and A. Aswi, "Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijiwa Bantaeng," *VARIANSI J. Stat. Its Appl. Teach. Res.*, vol. 4, no. 3, pp. 121–127, 2022, doi: 10.35580/variansiunm31.
- [21] I. Kurniawan, D. C. P. Buani, A. Abdussomad, W. Apriliah, and R. A. Saputra, "Implementasi Algoritma Random Forest Untuk Menentukan Penerima Bantuan Raskin," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 2, pp. 421–428, 2023, doi: 10.25126/jtiik.20231026225.
- [22] Ö. Akar and O. Güngör, "Classification of multispectral images using Random Forest algorithm," *J. Geod. Geoinf.*, vol. 1, no. 2, pp. 105–112, 2012, doi: 10.9733/jgg.241212.1.