**JURNAL ILMIAH FIFO** P-ISSN 2085-4315
Vol. 17 No. 1 (2025) 1-6 E-ISSN 2502-8332
DOI: 10.22441/fifo.2025.v17i1.001

# Klasifikasi Penyakit Kardiovaskular Menggunakan Algoritma Decision Tree C4.5

**Yuwan Jumaryadi[1]***

[1]Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Mercu Buana
Email: yuwan.jumaryadi@mercubuana.ac.id
Penulis Korespondensi*

(***received*****:** 07-05-24, ***revised*****:** 22-05-24, ***accepted*****:** 08-05-25)

**Abstrak**

Berdasarkan data yang diperoleh dari Kemenkes RI, kardiovaskular merupakan salah satu penyebab kematian tertinggi di Indonesia. Teknik Data Mining telah digunakan dalam beberapa penelitian di bidang akademik. Penelitian ini bertujuan untuk mengkategorikan penyakit kardiovaskular dan menganalisis hasil akurasi dari algoritman decision tree C4.5. Dataset penyakit kardiovaskular yang digunakan terdiri atas 2 kategori yaitu ada atau tidaknya penyakit kardiovaskular. Setelah data terkumpul, dilakukan preprocessing. Tahapan selanjutnya yaitu memisahkan data training dan data testing. Adapun algoritma yang digunakan untuk klasifikasi adalah decision tree. Setelah itu akan dilakukan evaluasi terhadap hasil klasifikasi untuk mendapatkan nilai akurasi, dan hasil tersebut akan dianalisis untuk menentukan kelayakan model klasifikasi. Berdasarkan hasil penelitian, algoritma decision tree dapat mengklasifikasi penyakit kardiovaskular dengan baik dan memperoleh akurasi tertinggi sebesar 63,62% pada kombinasi data testing dan training sebasar 70:30.

**Kata Kunci**: Decision Tree C4.5, Data Mining, Kardiovaskular, Klasifikasi

***Abstract***

*Based on data obtained from the Indonesian Ministry of Health, cardiovascular is one of the leading causes of death in Indonesia. Data Mining techniques have been used in several studies in the academic field. This study aims to categorize cardiovascular diseases and analyze the accuracy results of the C4.5 decision tree algorithm. The cardiovascular disease dataset used consists of 2 categories, namely the presence or absence of cardiovascular disease. After the data is collected, preprocessing is carried out. The next stage is to separate the training data and testing data. The algorithm used for classification is a decision tree. After that, an evaluation of the classification results will be carried out to obtain an accuracy value, and the results will be analyzed to determine the feasibility of the classification model. Based on the results of the study, the decision tree algorithm can classify cardiovascular diseases well and obtain the highest accuracy of 63.62% in a combination of testing and training data of 70:30.*

***Keywords****: Decision Tree C4.5, Data Mining, Cardiovascular, Classification*

## 1. PENDAHULUAN

Kemajuan teknologi informasi saat ini sangat bermanfaat bagi semua lapisan masyarakat [1]. Dalam kehidupanya sehari-hari banyak orang yang memanfaatkan perkembangan Teknologi Informasi [2]. Dengan menggunakan Teknologi informasi seseorang dapat mengerjakan pekerjaannya dengan lebih mudah. Penggunaan Teknologi informasi dapat diimplementasikan dalam berbagai bidang, salah satunya dalam bidang kesehatan. Kesehatan merupakan aspek terpenting didalam setiap kehidupan [3].

Permasalahan yang penting bagi instansi medis seperti rumah sakit adalah memberikan pelayanan yang baik kepada pasien dengan harga yang wajar sehingga diagnosis pasien yang tepat dan pilihan yang sesuai dapat dibuat untuk mencegah konsekuensi serius yang sangat tidak dapat ditoleransi [4]. Kardiovaskuler merupakan istilah gangguan jantung dan pembuluh darah. Salah satu penyakit kardiovaskular adalah penyakit jantung koroner, dimana penyakit ini merupakan penyebab tingginya tingkat kematian di dunia. Berdasarkan data yang diperoleh dari badan Kesehatan dunia WHO tahun 2012, factor penyebab kematian nomor satu didunia adalah kardiovaskuler [5].

Gangguan pada jantung dan pembuluh darah dapat menyebabkan penyakit kardiovaskuler. Menurut data dari Kemenkes RI, angka kematian yang disebabkan penyakit kardiovaskular di Indonesia sebanyak 650.000 per tahun [6]. Penyebab utamanya adalah penyakit jantung koroner dan penyakit stroke [7]. Ketersediaan dan distribusi dokter spesialis merupakan hal yang mempengaruhi kualitas pelayanan terhadap penyakit

P-ISSN 2085-4315
E-ISSN 2502-8332

kardiovaskuler. Jika dibandingkan dengan kebutuhan penduduk di Indonesia, maka saat ini Indonesia memiliki doktes spesialis penyakit jantung dan pembuluh darah yang sedikit sehingga hal ini akan mempengaruhi kualitas pelayanan terhadap penyakit kardiovaskuler [8].

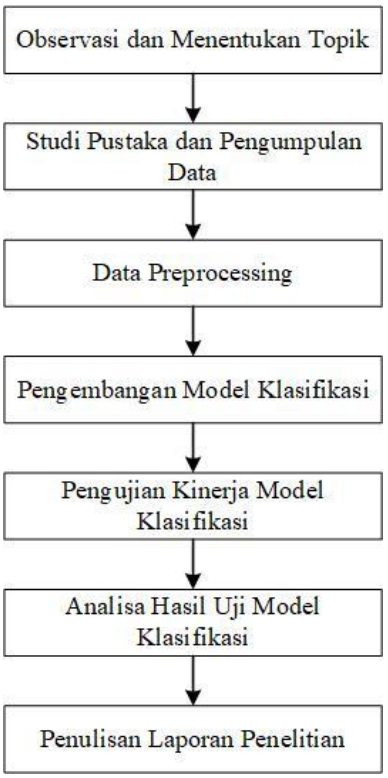
Dalam beberapa tahun ini jumlah data di bidang kesehatan telah meningkat. Data-data tersebut dapat digunakan untuk meningkatkan pelayanan di bidang kesehatan dengan menggunakan teknik Data Mining [9]. Model prediktif untuk peramalan di bidang kesehatan dapat menjadi alat yang berguna dalam pelayanan yang ada [10]. Metode data mining dapat digunakan untuk mengidentifikasi model yang tepat untuk memperbaiki setiap permasalahan dengan baik [11][12]. Pada penelitian yang dilakukan oleh Samodro, dkk (2023) menggunakan Naïve Bayes dan Decision Tree C4.5 untuk memprediksi penyakit diabetes. Hasil penelitian menunjukkan bahwa decision tree memiliki tingkat akurasi yang lebih tinggi dalam memprediksi diabetes dibandingkan dengan naïve bayes dengan dengan nilai 96,36% dan 90,45% [13].

*Decision tree* merupakan algoritma klasifikasi yang sederhana namun efektif [14][15][16][17]. Salah satu jenis algoritma decision tree adalah C4.5, dimana algoritma decision tree C4.5 efisien dalam menangani atribut bertipe diskret dan numerik [18]. Decision tree dapat memecah data kompleks menjadi beberapa aturan yang dapat dimengerti sehingga dapat digunakan untuk analitik data [19]. Penelitian ini bertujuan untuk mengetahui tingkat akurasi dari algoritma decision tree dengan menggunakan data penyakit kardiovaskular.

## 2. METODE PENELITIAN

### 2.1. Tahapan Penelitian

Pada Gambar 1 merupakan tahapan penelitian yang dilakukan dalam penelitian ini.




Gambar 1. Tahapan Penelitian

Berikut ini merupakan penjelasan tahapan penelitian pada Gambar 1.

1. Observasi dan Menentukan topik

Tahap awal dalam penelitian ini dengan melakukan observasi terlebih dahulu terhadap obyek penelitian. Setelah melakukan observasi terhadap obyek penelitian, kemudian peneliti mencoba menganalisa terhadap permasalahan yang dihadapi pada obyek penelitian. Dari permasalahan yang ditemukan pada obyek penelitian tersebut kemudian akan ditentukan topik yang akan diangkat untuk penelitian yang akan dilakukan.

2. Studi pustaka dan pengumpulan data

Dalam merumuskan masalah penelitian, penulis mengumpulkan beberapa bahan referensi yang berasal dari beberapa sumber antara lain buku, dan karya ilmiah online serta beberapa literatur dari internet. Hal ini

P-ISSN 2085-4315
E-ISSN 2502-8332

dilakukan untuk mencari solusi terhadap permasalahan yang dihadapi oleh peneliti didalam menyusun laporan tesis.

3. Data Preprocessing

Pada tahapan data processing perlu untuk mempersiapkan data untuk membangun model klasifikasi. Persiapan data diantaranya dengan melakukan pengecekan data terhadap kelengkapan data, penanganan terhadap hilangnya sebagian data dan inkonsistensi data. Pada tahapan ini juga ditentukan sampel yang akan digunakan dalam penelitian ini. Dataset yang digunakan dalam penelitian ini berasal dari Kaggle. Pada penelitian ini akan digunakan data testing dengan perbandingan 70:30, 80:20, dan 90:10. Pada penelitian ini akan dilihat data testing yang lebih akurat dengan komposisi yang digunakan.

4. Pengembangan Model Klasifikasi

Berdasarkan latar belakang dan tinjauan studi yang telah dilakukan pada tahap sebelumnya, maka dibuat hipotesis terhadap penelitian berdasarkan permasalahan yang ada. Selain itu juga dilakukan pengembangan model klasifikasi. Adapun Model yang akan dikembangkan yaitu Decision Tree C4.5.

5. Pengujian Kinerja Model Klasifikasi

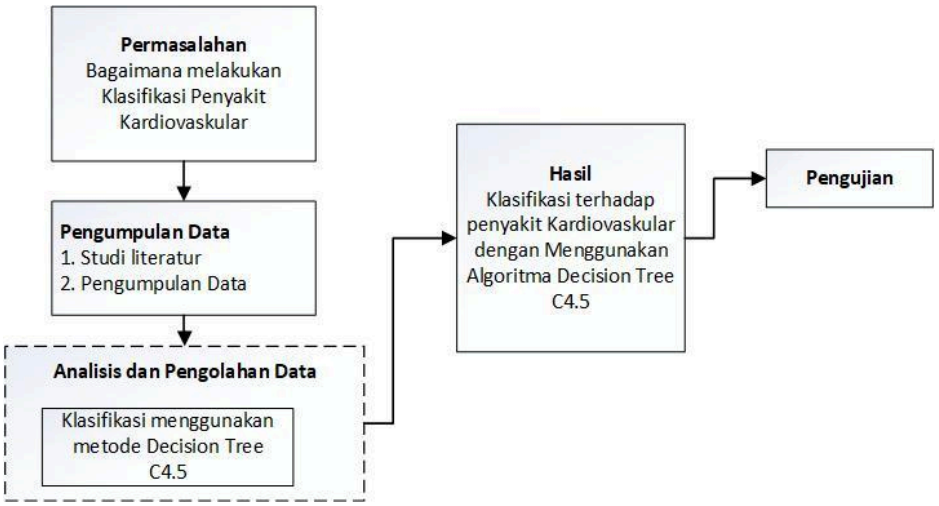
Tahapan berikut yang dilakukan penulis adalah menguji hasil kinerja model prediksi. Pengujian dilakukan dengan mengukur kinerja dari masing-masing model prediksi yang telah dibangun. Pengukuran kinerja model prediksi meliputi Accuracy, Recall, Precision, dan F1-Measure.

6. Penulisan Laporan Penelitian

Pada langkah terakhir dilakukan penyusunan laporan penelitian. Adapun penyusuan laporan penelitian dimulai dari awal penelitian sampai selesai.

## 3. HASIL DAN PEMBAHASAN

Pada bab ini akan dijelaskan mengenai analisis terhadap data yang akan diklasifikasi. Pada Gambar 2 merupakan kerangka konsep penelitian yang dilakukan.




Gambar 2. Kerangka Konsep Penelitian

### 3.1. Pengumpulan Data

Data yang didapatkan untuk penelitian ini berasal dari Kaggle [20]. Adapun total dataset yang didapatkan adalah 70000 record. Data yang didapatkan terdiri dari 11 atribut fitur dan 1 atribut target. Pada Tabel 1 merupakan detil atribut yang digunakan dalam penelitian.

Tabel 1. Atribut Data Penelitian

| Atribut | Tipe Data | Keterangan |
|---|---|---|
| Usia | Numerik | Usia (dalam hari) |
| Jenis Kelamin | Numerik | (1 = Laki-laki, 2 = Perempuan) |
| Tinggi | Numerik | int (cm) |
| Berat Badan | Numerik | float (kg) |

P-ISSN 2085-4315
E-ISSN 2502-8332

| | | |
|---|---|---|
| Tekanan darah sistolik | Numerik | int |
| Tekanan Darah Diastolik | Numerik | int |
| Kolesterol | Numerik | int (1: normal, 2: di atas normal, 3: jauh di atas normal) |
| Glukosa | Numerik | int (1: normal, 2: di atas normal, 3: jauh di atas normal) |
| Merokok | Numerik | binary |
| Asupan Beralkohol | Numerik | binary |
| Aktivitas fisik | Numerik | binary |
| Status Kardiovaskular | Numerik | binary |

Pada Tabel 2 merupakan rekapitulasi terhadap data yang digunakan dalam penelitian.

Tabel 2. Rekap Data Kardiovaskular

| | | Status Kardiovaskular | |
|---|---|---|---|
| Jenis Kelamin | Jumlah | 0 | 1 |
| L | 45530 | 22914 | 22616 |
| P | 24470 | 12107 | 12363 |
| Kolesterol | | | |
| 1 | 52385 | 29330 | 23055 |
| 2 | 9549 | 3799 | 5750 |
| 3 | 8066 | 1892 | 6174 |
| Glukosa | | | |
| 1 | 59479 | 30894 | 28585 |
| 2 | 5190 | 2112 | 3078 |
| 3 | 5331 | 2015 | 3116 |
| Merokok | | | |
| 0 | 63831 | 31781 | 32050 |
| 1 | 6169 | 3240 | 2929 |
| Asupan Beralkohol | | | |
| 0 | 66236 | 33080 | 33156 |
| 1 | 3764 | 1941 | 1823 |
| Aktivitas fisik | | | |
| 0 | 13739 | 6378 | 7361 |
| 1 | 56261 | 28643 | 27618 |

### 3.2. Analisis Data

Berdasarkan data rekapitulasi pada Tabel 2, dapat diketahui bahwa pada penelitian yang dilakukan ada 45530 pasien atau 65,04% yang berjenis kelamin laki-laki dan 24470 pasien atau 34,96% yang berjenis kelamin perempuan. Selain itu dari data kolesterol dapat diketahui bahwa ada 52385 pasien atau 74,84% yang memiliki kolesterol normal, 9549 pasien atau 13,64% yang memiliki kolesterol diatas normal, dan 8066 pasien atau 11,52% yang memiliki kolesterol jauh diatas normal. Setelah itu berdasarkan data glukosa dapat diketahui bahwa 59479 atau 84,97% yang memiliki glukosa normal, 5190 atau 7,41% yang memiliki glukosa diatas normal, dan 5331 pasien atau 7,33% yang memiliki glukosa jauh diatas normal. Kemudian ada 63831 atau 91,19% pasien yang tidak merokok, dan 6169 atau 8,81% pasien yang merokok. Dari asupan alkohol, didapatkan 66236 atau 94,62% pasien yang tidak mengkonsumsi alkohol, dan 3764 atau 5,38% pasien yang mengkonsumsi alkohol. Dan dari aktivitas fisik diketahui bahwa 13739 atau 19,63% pasien yang tidak melakukan aktivitas fisik, dan 56261 atau 80,37% pasien yang melakukan aktivitas fisik.

### 3.3. Pengujian Kinerja Model Klasifikasi

Pada tahapan pengujian akan digunakan 2 algoritma untuk menguji tingkat akurasi yang lebih baik. Berikut ini merupakan langkah-langkah pengujian kinerja model klasifikasi:

1. Persiapan Data

Pada tahap awal dalam penelitian akan dilakukan pengumpulan data untuk proses klasifikasi. Dataset yang telah terkumpul akan dibersihkan, dan dibagi menjadi data fitur dan data label.

2. Pemilihan Algoritma

Setelah dataset diberikan label, maka langkah selanjutnya adalah memilih algoritma untuk mengklasifikasikan dataset penyakit kardiovaskular. Dalam penelitian ini, algoritma yang digunakan untuk melakukan klasifikasi adalah Decision Tree.

3. Pembagian Data

Setelah pemilihan algoritma, maka data akan dibagi data menjadi data training dan data testing. Pemodelan klasifikasi akan menggunakan data training, sedangkan untuk menguji kinerja model yang telah dibuat akan menggunakan data testing. Adapun kombinasi data training dan data testing yang digunakan dalam penelitian ini adalah 90:10, 80:20, dan 70:30.

4. Evaluasi Model

Tahapan selanjutnya adalah mengevaluasi model menggunakan kombinasi data training dan data testing yang telah ditentukan. Pada tahap evaluasi model dilakukan perbandingan hasil data uji menggunakan nilai akurasi, presisi, *recall*, dan *f1-score* dengam nenggunakan kombinasi dataset dan data training yang telah ditentukan. Adapan jumlah dataset yang digunakan sejumlah 70.000, kemudian dataset dipisah antara data traininig dan data testing untuk setiap algoritma.

5. Interpretasi Hasil

Setelah mengevaluasi model yang digunakan, maka diperlukan interpretasi terhadap hasil yang didapatkan.

### 3.4. Pembahasan

Dalam tahap ini, sebuah model *machine learning* dibuat dari dataset yang ada. Hasil dari tahap ujicoba dua algoritma klasifikasi dapat dilihat pada Tabel 3.

Tabel 3. Perbandingan Akurasi Algoritma Decision Tree dan Naive Bayes

| Kombinasi Data Training: Testing | Akurasi | |
|---|---|---|
| | Decision Tree | Naïve Bayes |
| 90:10 | 62,63% | 56,49% |
| 80:20 | 61,04% | 57,14% |
| 70:30 | 63,62% | 56,91% |

Pada Tabel 3 menunjukan hasil accuracy dari Decision Tree 62,63%, sedangkan Naive Bayes 56,49% pada kombinasi data training dan data testing 90:10. Sedangkan pada kombinasi data training dan data testing 80:20 menunjukan hasil accuracy dari Decision Tree 61,04%, sedangkan Naive Bayes 57,14%. Kemudian pada kombinasi data training dan data testing 70:30 menunjukan hasil accuracy dari Decision Tree 63,02%, sedangkan Naive Bayes 56,91%. Sehingga didapatkan bahwa akurasi decision tree lebih tinggi pada ketiga kombinasi data training dan data testing dalam memprediksi penyakit kardiovaskular. Perbedaan tingkat akurasi dikarenakan karakteristik dari dataset dan parameter yang berbeda dari masing-masing algoritma. Dari kesesuaian karakteristik dataset, maka Decision Tree dapat bekerja lebih baik dibanding Naïve Bayes untuk mengklasifikan penyakit kardiovaskular. Hasil penelitian yang dilakukan sama dengan penelitian yang dilakukan oleh Samodro, dkk (2023) untuk memprediksi penyakit diabetes. Dimana hasil penelitian menunjukkan bahwa algoritma Decision Tree C4.5 memiliki tingkat akurasi yang lebih tinggi dalam memprediksi diabetes dibandingkan dengan naïve bayes [13].

## 4. KESIMPULAN

Penelitian ini menggunakan algoritma Decision Tree C4.5 untuk memprediksi penyakit kardiovaskular. Adapun Naive Bayes digunakan untuk perbandingan Tingkat akurasi algoritma. Berdasarkan penelitian yang dilakukan, Dua pengklasifikasi memprediksi label dalam dataset berdasarkan kategori yang ada. Berdasarkan penelitian yang dilakukan diketahui bahwa algoritma decision tree C4.5 memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan Naive Bayes pada 3 kombinasi data training dan data testing untuk untuk memprediksi penyakit kardiovaskular. Berdasarkan hasil penelitian, Nilai akurasi dari Decision Tree C4.5 lebih baik dalam memprediksi penyakit kardiovaskular karena dapat memberikan prediksi yang lebih akurat.

## DAFTAR PUSTAKA

[1] A. A. Mutezar and Umniy Salamah, “Pengembangan Sistem Manajemen Event Pameran Karya Mahasiswa Menggunakan Metode Extreme Programming,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, pp. 809–819, 2021, doi: 10.29207/resti.v5i4.3249.

[2] S. Rahayu, A. Nugroho, E. D. Putra, M. Purba, and H. Setiawan, “Komparasi Hasil Color Feature

P-ISSN 2085-4315
E-ISSN 2502-8332

Extraction HSV , LAB dan YCrCb pda Algoritma SVM untuk Klasifikasi Spesies Burung," *JSAI J. Sci. Appl. Informatics*, vol. 06, no. 03, pp. 482–487, 2023.

[3] A. Muzakir and R. A. Wulandari, "Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree," *Sci. J. Informatics*, vol. 3, no. 1, pp. 19–26, 2016.

[4] V. S. K. Reddy, P. Meghana, N. V. S. Reddy, and B. A. Rao, "Prediction on Cardiovascular disease using Decision tree and Naïve Bayes classifiers," in *1st International Conference on Artificial Intelligence, Computational Electronics and Communication System (AICECS 2021)*, Manipal, 2022, pp. 1–7. doi: 10.1088/1742-6596/2161/1/012015.

[5] C. J. McAloon *et al.*, "The changing face of cardiovascular disease 2000–2012: An analysis of the world health organisation global health estimates data," *Int. J. Cardiol.*, vol. 224, pp. 256–264, 2016, doi: 10.1016/j.ijcard.2016.09.026.

[6] Humas Fakultas Kedokteran Universitas Brawijaya, "World Heart Day 2023: Use Heart Know Heart," Prasetya Online. Accessed: Sep. 08, 2024. [Online]. Available: https://prasetya.ub.ac.id/world-heart-day-2023-use-heart-know-heart/

[7] G. Setiawan and M. Christiany Halim, "Pengaruh Asam Lemak Omega-3 terhadap Penyakit Kardiovaskular," *Contin. Prof. Dev.*, vol. 49, no. 3, pp. 160–163, 2022.

[8] E. Faizal, "Case Based Reasoning Diagnosis Penyakit Cardiovascular Dengan Metode Simple Matching Coefficient Similarity," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 2, p. 83, 2014, doi: 10.25126/jtiik.201412116.

[9] Endang S Kresnawati, Yulia Resti, Bambang Suprihatin, M. Rendy Kurniawan, and Widya Ayu Amanda, "Coronary Artery Disease Prediction Using Decision Trees and Multinomial Naïve Bayes with k-Fold Cross Validation," *Inomatika*, vol. 3, no. 2, pp. 174–189, 2021, doi: 10.35438/inomatika.v3i2.266.

[10] P. Mallioris, E. Diamantis, C. Bialas, and D. Bechtsis, "Predictive maintenance framework for assessing health state of centrifugal pumps," *IAES Int. J. Artif. Intell.*, vol. 13, no. 1, pp. 850–862, 2024, doi: 10.11591/ijai.v13.i1.pp850-862.

[11] B. Priambodo *et al.*, "Predicting Employee Turnover in IT Industries using Correlation and Chi-Square Visualization," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 12, pp. 71–75, 2022.

[12] H. Li *et al.*, "Prediction of the Vanadium Content of Molten Iron in a Blast Furnace and the Optimization of Vanadium Extraction," *Separations*, vol. 10, no. 10, 2023, doi: 10.3390/separations10100521.

[13] M. M. J. Samodro, M. K. Biddinika, and A. Fadlil, "Klasifikasi Penyakit Diabetes dengan Algoritma Decision Tree dan Naïve Bayes," *Resist. (Elektronika Kendali Telekomun. Tenaga List. Komputer) Vol.*, vol. 6, no. 2, pp. 113–118, 2023.

[14] A. Arista, "Comparison Decision Tree and Logistic Regression Machine Learning Classification Algorithms to determine Covid-19," *Sinkron*, vol. 7, no. 1, pp. 59–65, 2022, doi: 10.33395/sinkron.v7i1.11243.

[15] A. Nugroho, H. L. H. S. Warnars, F. L. Gaol, and T. Matsuo, "Trend of Stunting Weight for Infants and Toddlers Using Decision Tree," *IAENG Int. J. Appl. Math.*, vol. 52, no. 1, 2022.

[16] J. J. Purnama, H. M. Nawawi, S. Rosyida, Ridwansyah, and Risandar, "Klasifikasi Mahasiswa HER Berbasis Algoritma SVM dan Decision Tree," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 6, pp. 1253–1260, 2020, doi: 10.25126/jtiik.202073080.

[17] T. Setiyorini and R. T. Asmono, "Komparasi Metode Decision Tree, Naive Bayes Dan K-Nearest Neighbor Pada Klasifikasi Kinerja Siswa," *J. Techno Nusa Mandiri*, vol. 15, no. 2, p. 85, 2018, doi: 10.33480/techno.v15i2.889.

[18] Nurfitrayani, Islamiyah, and A. P. A. Masa, "Penerapan Klasifikasi Algoritma C4.5 Dan Algoritma C5.0 Untuk Mengetahui Tingkat Kepuasan Mahasiswa Terhadap Website Sistem Informasi Terpadu Layanan Program Studi (SIPLO)," *J. Media Inform. Budidarma*, vol. 7, pp. 1877–1887, 2023, doi: 10.30865/mib.v7i4.6433.

[19] V. P. Prasetyo, M. F. A. Ulin Nuha, M. H. Hakiki, R. A. Vinarti, and A. Djunaidy, "Comparison of Data Mining Techniques on Stroke Clinical Dataset," *Procedia Comput. Sci.*, vol. 234, pp. 502–511, 2024, doi: 10.1016/j.procs.2024.03.033.

[20] S. Ulianova, "Cardiovascular Disease dataset," kaggle. Accessed: Dec. 01, 2023. [Online]. Available: https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset/data