

Perbandingan Algoritma Klasifikasi Data Mining Dalam Diagnosa Penyakit Arteri Koroner

Bernadus Gunawan Sudarsono^{1*}, Edy Winarno²

¹ Fakultas Ilmu Komputer, Program Studi Sistem Informasi, Universitas Bung Karno, Jakarta, Indonesia

² Fakultas Teknik dan Ilmu Komputer, Program Studi Teknologi Informasi, Universitas Muhammadiyah Semarang, Semarang, Indonesia

Email: ^{1*}gunawanbernadus@ubk.ac.id, ²edywin@unimus.ac.id

Email Penulis Korespondensi: gunawanbernadus@ubk.ac.id

Submitted: 27/09/2024; Accepted: 01/12/2024; Published: 17/12/2024

Abstrak—Penyakit arteri koroner merupakan salah satu penyakit yang sering menyerang manusia. Penyebab terjadinya penyakit ini, karena adanya penyempitan atau penyumbatan pembuluh darah koroner yang memasok darah menuju jantung. Diagnosa penyakit arteri koroner oleh pihak medis selama ini terkendala keterbatasan jumlah dokter, dari segi jumlah dokter dan waktu, karena jumlah dokter spesialis terbatas. Keterbatasan jumlah dokter menimbulkan beberapa kesulitan bagi pihak medis yang mendiagnosa penyakit pasien dan seiring waktu bisa saja menjadi masalah yang serius. Teknologi informasi yang dapat membantu pihak medis adalah dengan menerapkan teknik data mining yang merupakan suatu teknik untuk membantu mendiagnosa penyakit arteri koroner. Data mining dapat mengidentifikasi pola atau hubungan antara gejala penyakit dan hasil diagnosa, sehingga dapat diketahui pasien dengan risiko tinggi terkena penyakit tersebut. Algoritma Naïve Bayes merupakan salah satu algoritma dari teknik klasifikasi Data Mining, yang berdasarkan pada teorema Bayes. Algoritma C4.5 merupakan salah satu algoritma dari teknik klasifikasi Data Mining, yang menggunakan pohon keputusan dalam mengklasifikasi data. Perbandingan algoritma dilakukan agar mendapatkan algoritma yang sesuai atau yang terbaik untuk digunakan dalam mendiagnosa suatu penyakit. Proses perbandingan algoritma Naïve Bayes dan algoritma C.45 dalam diagnosa penyakit arteri koroner, mendapatkan hasil algoritma terbaik berdasarkan nilai persentase terbesar yaitu algoritma C4.5, dengan nilai 46,9%.

Kata Kunci : Penyakit Arteri Koroner; Algoritma C4.5; Algoritma Naive Bayes; Data Mining; Perbandingan

Abstract—Coronary artery disease is one of the diseases that often attacks humans. The cause of this disease is due to narrowing or blockage of the coronary blood vessels that supply blood to the heart. The diagnosis of coronary artery disease by medical personnel has so far been constrained by the limited number of doctors, in terms of the number of doctors and time, because the number of specialist doctors is limited. The limited number of doctors causes several difficulties for medical personnel who diagnose the patient's disease and over time can become a serious problem. Information technology that can help medical personnel is by applying data mining techniques which are techniques to help diagnose coronary artery disease. Data mining can identify patterns or relationships between disease symptoms and diagnostic results, so that patients with a high risk of developing the disease can be identified. The Naïve Bayes algorithm is one of the algorithms of the Data Mining classification technique, which is based on Bayes' theorem. The C4.5 algorithm is one of the algorithms of the Data Mining classification technique, which uses decision trees in classifying data. Algorithm comparisons are carried out in order to obtain the appropriate or best algorithm for use in diagnosing a disease. The comparison process of the Naïve Bayes algorithm and the C.45 algorithm in diagnosing coronary artery disease, obtained the best algorithm results based on the largest percentage value, namely the C4.5 algorithm, with a value of 46.9%.

Keywords: Coronary Artery Disease; C4.5 Algorithm; Naive Bayes Algorithm; Data Mining; Comparison

1. PENDAHULUAN

Penyakit arteri koroner merupakan salah satu penyakit yang sering menyerang manusia. Penyebab terjadinya penyakit ini, karena adanya penyempitan atau penyumbatan pembuluh darah koroner yang memasok darah menuju jantung. Apabila penanganan penyakit ini tidak dilakukan secara serius, maka bisa saja berakibat fatal. Oleh sebab itu, pihak medis membutuhkan teknologi yang dapat membantu dalam mengidentifikasi pasien berisiko tinggi, sehingga bisa mendapat perawatan medis secara cepat dan tepat [1], [2].

Diagnosa penyakit arteri koroner oleh pihak medis selama ini terkendala keterbatasan jumlah dokter, dari segi jumlah dokter dan waktu, karena jumlah dokter spesialis terbatas. Keterbatasan jumlah dokter menimbulkan beberapa kesulitan bagi pihak medis yang mendiagnosa penyakit pasien dan seiring waktu bisa saja menjadi masalah yang serius. Misalnya saja, apabila ada seorang pasien yang hendak berkonsultasi mengenai penyakitnya, maka hasil diagnosa akan memakan waktu yang lumayan lama karena harus menunggu jadwal hadir dokter [3], [4].

Teknologi informasi yang dapat membantu pihak medis adalah dengan menerapkan teknik data mining yang merupakan suatu teknik untuk membantu mendiagnosa penyakit arteri koroner. Data mining dapat mengidentifikasi pola atau hubungan antara gejala penyakit dan hasil diagnosa, sehingga dapat diketahui pasien dengan risiko tinggi terkena penyakit tersebut. Penerapan data mining dalam mendiagnosa penyakit dapat dilakukan dengan menggunakan algoritma data mining seperti algoritma Naïve Bayes dan algoritma C.45 [5], [6].

Algoritma Naïve Bayes merupakan salah satu algoritma dari teknik klasifikasi Data Mining, yang berdasarkan pada teorema Bayes. Umumnya algoritma ini digunakan diberbagai bidang, salah satunya ialah bidang kesehatan, digunakan untuk mendiagnosa suatu penyakit tertentu. Kelebihan dari algoritma ini yaitu hanya menggunakan sejumlah kecil data pelatihan (training data) untuk memperkirakan parameter dalam klasifikasi. Sementara untuk

kelemahannya yaitu algoritma ini hanya digunakan untuk persoalan klasifikasi dengan supervised learning dan data-data kategorial[7], [8].

Algoritma C.45 merupakan salah satu algoritma dari teknik klasifikasi Data Mining, yang menggunakan pohon keputusan dalam mengklasifikasi data. Umumnya algoritma ini digunakan diberbagai bidang, salah satunya ialah bidang kesehatan, digunakan untuk mendiagnosa suatu penyakit tertentu. Kelebihan dari algoritma ini yaitu dapat mengatasi atribut numerik, atribut yang hilang, serta mengatasi masalah dengan dimensi yang tinggi. Sementara untuk kelemahannya yaitu pembentukan pohon keputusan yang kompleks dapat membuat overfitting, dimana model telah sesuai dengan data pelatihan, namun tidak dapat mengeneralisasi data baru dengan baik[9], [10].

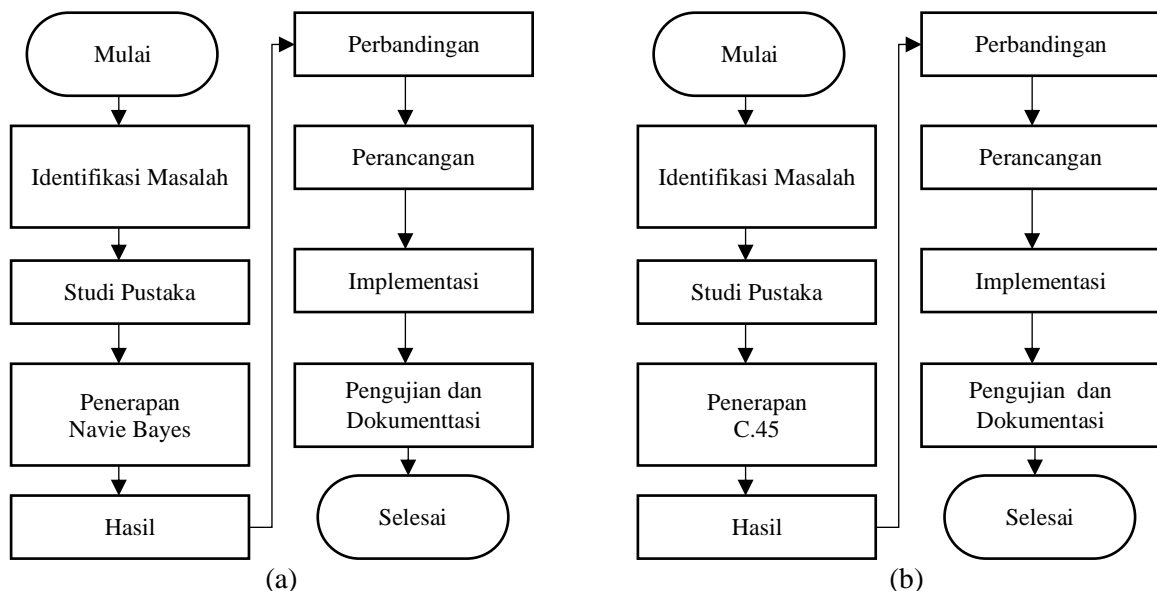
Banyaknya algoritma Data Mining yang bisa digunakan dalam mendiagnosa suatu penyakit, maka perlu dilakukan perbandingan algoritma. Perbandingan algoritma dilakukan agar mendapatkan algoritma yang sesuai atau yang terbaik untuk digunakan dalam mendiagnosa suatu penyakit. Setiap algoritma memiliki kelebihan maupun kekurangan tersendiri, juga terdapat perbedaan-perbedaan dalam teknik pengklasifikasiannya. Oleh sebab itu, dengan melakukan perbandingan algoritma, maka dapat diketahui mana algoritma yang paling sesuai untuk digunakan.

Berdasarkan penelitian sebelumnya yang dilakukan oleh Syamsul Bahri pada tahun 2019 tentang perbandingan algoritma Naïve Bayes dan C.45 untuk klasifikasi penyakit anak, mengatakan bahwa algoritma C.45 merupakan algoritma terbaik dalam melakukan klasifikasi dengan tingkat akurasi sebesar 90%[11]. Berdasarkan penelitian sebelumnya yang dilakukan oleh Yogie Indra Kurniawan tentang perbandingan algoritma Naïve Bayes dan C.45 dalam klasifikasi data mining, mengatakan bahwa hasil klasifikasi pada algoritma Naïve Bayes dan C.45 tidak dapat memberikan nilai mutlak, sehingga harus menggunakan kriteria dan variabel suatu kasus[12]. Berdasarkan penelitian sebelumnya yang dilakukan oleh Bayu Sugara tentang perbandingan akurasi algoritma Naïve Bayes dan C.45 untuk deteksi dini gangguan autisme pada anak, mengatakan bahwa algoritma Naïve Bayes menunjukkan nilai akurasi yang lebih besar dan lebih baik yaitu sebesar 73,33%[13].

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Metodologi penelitian adalah langkah-langkah yang harus ditempuh dalam melakukan suatu penelitian. Tujuan dari metodologi penelitian ini adalah untuk membuat penelitian menjadi terstruktur dan memiliki konsep yang jelas, sehingga hasilnya akurat dan tidak dipertanyakan kebenarannya. Perbandingan Algoritma Naïve Bayes Dan C.45 Dalam Klasifikasi Diagnosa Penyakit Arteri Koroner memiliki metode penelitian sebagai berikut:



Gambar 1. (a) Kerangka Metode Penelitian Naive Bayes, (b) Kerangka Metode Penelitian C.45

2.2 Data Mining

Data mining merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi maupun mengidentifikasi informasi yang bermanfaat dari berbagai database besar. Data mining menemukan hubungan yang berarti, pola kecenderungan dengan memeriksa sekumpulan besar data yang tersimpan dalam penyimpanan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika. Data mining melakukan ekstraksi untuk mendapatkan informasi penting yang sifatnya implisit dan sebelumnya tidak diketahui dari suatu data. Proses ini otomatis atau (biasanya) semi-otomatis. Penggunaan teknologi penalaran pola serta teknik - teknik statistik matematika, data mining dapat dikatakan sebagai proses dalam memperoleh informasi baru dari data - data[14]-[16].

2.3 Algoritma Naïve Bayes

Naive bayes merupakan pengklasifikasi probabilitas sederhana berdasarkan pada teorema bayes. Keuntungan dari klasifikasi adalah bahwa ia hanya membutuhkan sejumlah kecil data pelatihan untuk memperkirakan parameter (sarana dan varians dari variabel) yang diperlukan untuk klasifikasi. Karena variabel independen diasumsikan, hanya variasi dari variabel untuk masing-masing kelas harus ditentukan, bukan seluruh matriks kovarians. Dalam prosesnya, naive bayes mengasumsikan bahwa asal atau tidaknya suatu fitur pada suatu kelas tidak berhubungan dengan ada atau tidaknya fitur lain dikelas yang sama. Rumus perhitungan yang digunakan pada algoritma naive bayes adalah sebagai berikut[17], [18]:

$$P(B|A) = \frac{P(A|B)*P(B)}{P(A)} \quad (1)$$

Dimana :

- P(B|A) : Peluang B jika diketahui keadaan jenis penyakit A
- P(A|B) : Peluang evidence A jika diketahui hipotesis B
- P(B) : Probabilitas hipotesis B tanpa memandang evidence apapun
- P(A) : Peluang evidence penyakit A

Pada penerapannya sebagai pengklasifikasi teks atau dokumen dapat menggunakan persamaan berikut :

$$V_{MAP} \underset{V_j \in V}{\operatorname{argmax}} = \frac{P(a_1, a_2, \dots, a_n | V_j) P(V_j)}{P(a_1, a_2, \dots, a_n)} \quad (2)$$

Dimana :

- V_{MAP} : Probabilitas tinggi
- P(V_j) : Peluang jenis penyakit ke
- P($a_1, a_2, \dots, a_n | V_j$): Peluang atribut (inputan) jika diketahui keadaan V_j
- P(a_1, a_2, \dots, a_n): Peluang atribut-atribut utang

Menghitung P($a_1, a_2, \dots, a_n | V_j$) P(V_j) semakin sulit karena jumlah gejala p ($a_1 a_2 \dots a_n | V_j$) p (v_j) bisa jadi sangat besar. Hal ini disebabkan jumlah gejala tersebut sama dengan jumlah kombinasi gejala dikali dengan jumlah kategori yang ada, Sehingga dapat diasumsikan dengan rumus berikut :

$$P(a_i | V_j) = \frac{nc + mp}{n + m} \quad (3)$$

Dimana :

- nc : jumlah record pada data learning $V = V_j$ dan $a = a_i$
- mp : jumlah parameter dari banyaknya jenis class/penyakit
- n : jumlah record pada data learning $V = V_j$ tiap class/penyakit
- m : jumlah parameter

Pengklasifikasian dengan menggunakan algoritma naïve bayes dapat dilakukan dengan langkah – langkah sebagai berikut :

- a. Menentukan nilai nc untuk setiap class.
- b. Menghitung nilai P($a_i | V_j$) dan menghitung nilai P(V_j)
- c. Menghitung nilai P($a_i | V_j$) * P(V_j) untuk setiap class.
- d. Menentukan hasil klasifikasi yaitu class dengan hasil perkalian yang terbesar.

2.4 Algoritma C4.5

Algoritma C.45 (Classifier version 4.5) merupakan algoritma yang digunakan untuk membangun sebuah model pohon keputusan dapat digunakan untuk melakukan proses klasifikasi atau regresi. Algoritma ini dikembangkan oleh Ross Quinlan pada tahun 1993 sebagai pengembangan dari algoritma sebelumnya yang dikenal sebagai ID3 (Iterative Dichotomiser 3). Algoritma C.45 adalah salah satu algoritma yang populer dan banyak digunakan dalam bidang data mining dan pembelajaran mesin. Tujuan utama dari algoritma C.45 adalah membangun pohon keputusan yang optimal berdasarkan data pelatihan yang diberikan. Pohon keputusan ini dapat digunakan untuk mengklasifikasikan contoh-contoh baru berdasarkan atribut-atribut yang diamati. Adapun langkah-langkah dari penerapan algoritma ini sebagai berikut[19]–[21]:

- a. Pilih atribut yang paling baik sebagai atribut pengelompokan pada setiap level pohon keputusan. Atribut yang paling baik dipilih berdasarkan metrik pengukuran seperti gain informasi atau rasio gain informasi.
- b. Bagi dataset berdasarkan nilai-nilai atribut yang dipilih pada langkah pertama. Setiap nilai atribut akan menjadi cabang dari node saat ini.
- c. Jika semua contoh dalam satu cabang memiliki label kelas yang sama, maka cabang tersebut menjadi daun (leaf) pada pohon keputusan.
- d. Jika terdapat campuran label kelas pada satu cabang, maka ulangi langkah 1-3 untuk cabang tersebut.
- e. Ulangi langkah-langkah 1-4 untuk setiap cabang hingga seluruh dataset terbagi secara homogen atau mencapai kondisi berhenti yang ditentukan (misalnya, mencapai tingkat kedalaman maksimum).
- f. Setelah pohon keputusan terbentuk, pruning (pemangkasan) dapat dilakukan untuk mengurangi overfitting dan meningkatkan generalisasi model.

Algoritma C4.5 memiliki keunggulan, seperti mampu menangani atribut numerik, mampu menangani dataset dengan missing values, dan dapat menghasilkan model yang lebih kompak. Namun, algoritma ini memiliki beberapa kelemahan, seperti cenderung memilih atribut dengan banyak nilai dan memerlukan waktu komputasi yang relatif tinggi untuk dataset yang besar.

2.5 Penyakit Arteri Koroner

Penyakit arteri koroner, juga dikenal sebagai penyakit jantung koroner (PJK), adalah kondisi dimana arteri koroner yang memasok darah ke otot jantung mengalami penyempitan atau penyumbatan. Arteri koroner adalah pembuluh darah yang penting karena menyediakan oksigen dan nutrisi yang dibutuhkan oleh otot jantung untuk berfungsi dengan baik. Penyakit arteri koroner terjadi ketika plak, yang terdiri dari lemak, kolesterol, dan zat lainnya, menumpuk di dinding arteri koroner, proses ini disebut aterosklerosis. Plak yang menumpuk seiring waktu dapat menyempitkan lumen arteri dan mengurangi aliran darah ke jantung. Jika plak pecah atau terbentuk bekuan darah di sekitar plak, dapat terjadi penyumbatan total atau sebagian arteri koroner, yang dapat menyebabkan serangan jantung atau angina (nyeri dada). Gejala penyakit arteri koroner dapat bervariasi, tergantung pada tingkat penyempitan arteri koroner dan seberapa banyak aliran darah yang terganggu. Adapun beberapa gejala yang umum terjadi meliputi [22], [23]:

- Angina: Nyeri, tekanan, atau ketidaknyamanan pada dada yang sering terjadi saat aktivitas fisik atau stres. Nyeri ini dapat menjalar ke lengan, bahu, rahang, atau punggung.
- Serangan jantung (infark miokard): Terjadi ketika aliran darah ke bagian otot jantung terhalangi sepenuhnya, menyebabkan kerusakan permanen pada jaringan jantung. Gejalanya meliputi nyeri dada yang parah dan berkepanjangan, sesak napas, mual, muntah, dan berkeringat dingin.
- Dispnea (sesak napas): Kesulitan bernapas atau merasa tidak dapat bernapas dengan mudah, terutama saat melakukan aktivitas fisik.
- Kelelahan: Merasa lelah atau lemah secara berlebihan, bahkan dengan aktivitas ringan.
- Palpitasi: Sensasi tidak teratur atau tidak biasa dalam detak jantung.

Penyakit arteri koroner merupakan salah satu penyebab utama kematian di banyak negara. Faktor risiko yang dapat meningkatkan kemungkinan seseorang mengembangkan penyakit arteri koroner meliputi usia, riwayat keluarga/keturunan, merokok, hipertensi (tekanan darah tinggi), kolesterol tinggi, obesitas, diabetes, gaya hidup tidak sehat, dan kurangnya aktivitas fisik. Menghindari penyakit arteri koroner melibatkan perubahan gaya hidup, seperti menjaga pola makan sehat, berhenti merokok, menghindari stres, dan rutin berolahraga. Terkadang, pengobatan medis seperti obat-obatan untuk mengendalikan tekanan darah, kolesterol, atau mengurangi risiko pembekuan darah juga dapat diberikan.

3. HASIL DAN PEMBAHASAN

Analisa merupakan proses menguraikan suatu permasalahan menjadi beberapa elemen terkecil, guna mempermudah dalam menyelesaikan permasalahan yang ada. Tahap analisa memerlukan suatu pendekatan analisa guna menghindari kesalahan yang mungkin saja muncul pada tahap berikutnya. Tahap analisa ini merupakan tahapan yang sangat penting, pendekatan yang dilakukan adalah mendefinisikan masalah pada sistem yang sedang berjalan dan sekaligus melakukan evaluasi setiap cara kerja sistem yang sedang berjalan berdasarkan.

Permasalahan yang terjadi untuk mendiagnosa penyakit arteri koroner yaitu membandingkan algoritma Naïve Bayes dan algoritma C.45, untuk menentukan algoritma yang terbaik. Adanya permasalahan tersebut, maka dibutuhkan suatu teknik yang menjadi solusi untuk menentukan algoritma terbaik dengan tingkat persentase validasi tertinggi. Teknik yang akan digunakan yaitu dengan teknik Data Mining, menggunakan algoritma Naïve Bayes dan algoritma C.45 untuk mendiagnosa penyakit arteri koroner.

Proses diagnosa penyakit arteri koroner dimulai dari menentukan jumlah gejala yang akan di terapkan dalam perbandingan algoritma Naïve Bayes dan algoritma C.45. Kemudian gejala yang sudah ditentukan, akan di klasifikasi berdasarkan cara kerja kedua algoritma, sehingga didapatkan hasil diagnosa dari kedua algoritma tersebut. Apabila sudah didapatkan hasil diagnosa dengan menggunakan kedua algoritma yang telah di tentukan, maka dilakukan perbandingan dari hasil diagnosa kedua algoritma tersebut. Algoritma yang memiliki persentase validasi tertinggi, maka ditetapkan sebagai algoritma yang terbaik dalam mendiagnosa penyakit arteri koroner. Hasil dari diagnosa penyakit dengan algoritma terbaik yang telah ditetapkan maka bisa di update pada sistem, lalu sistem menyimpan data yang telah di update.

3.1 Penerapan Algoritma Naïve Bayes

Penyelesaian yang dilakukan yaitu dengan menggunakan algoritma Naïve Bayes dapat dilakukan seperti pada langkah berikut :

- Menghitung nilai probabilitas gejala penyakit P(A) dengan menggunakan rumus (1), dengan menggunakan rumus berikut :

$$P(A1) = \frac{5}{5} = 1$$

$$P(A2) = \frac{2}{5} = 0,4$$



$$P(A3) = \frac{3}{5} = 0,6$$

$$P(A4) = \frac{2}{5} = 0,4$$

$$P(A5) = \frac{1}{5} = 0,2$$

Sehingga akan didapatkan nilai probabilitas P(A) seperti yang bisa diketahui pada Tabel 1 berikut :

Tabel 1. Nilai Probabilitas P(A)

No.	Pasien	Gejala (A)					Nilai Probabilitas P(A)
		A1	A2	A3	A4	A5	
1	Pasien 1	Ya			Ya		1
2	Pasien 2	Ya	Ya				0,4
3	Pasien 3	Ya	Ya	Ya			0,6
4	Pasien 4	Ya		Ya			0,4
5	Pasien 5	Ya		Ya	Ya	Ya	0,2

b. Menghitung nilai peluang gejala penyakit P(B) menggunakan rumus (2), dengan menggunakan rumus berikut ini:

$$P(B1) = \frac{(B1)}{B1+B2+B3+B4+B5}$$

$$= \frac{1}{1+0,4+0,6+0,4+0,2}$$

$$= \frac{1}{2,6}$$

$$= 0,38$$

$$P(B2) = \frac{(B2)}{B1+B2+B3+B4+B5}$$

$$= \frac{0,4}{1+0,4+0,6+0,4+0,2}$$

$$= \frac{0,4}{2,6}$$

$$= 0,15$$

$$P(B3) = \frac{(B3)}{B1+B2+B3+B4+B5}$$

$$= \frac{0,6}{1+0,4+0,6+0,4+0,2}$$

$$= \frac{0,6}{2,6}$$

$$= 0,23$$

$$P(B4) = \frac{(B4)}{B1+B2+B3+B4+B5}$$

$$= \frac{0,4}{1+0,4+0,6+0,4+0,2}$$

$$= \frac{0,4}{2,6}$$

$$= 0,15$$

$$P(B5) = \frac{(B5)}{B1+B2+B3+B4+B5}$$

$$= \frac{0,2}{1+0,4+0,6+0,4+0,2}$$

$$= \frac{0,2}{2,6}$$

$$= 0,07$$

Sehingga akan didapatkan nilai peluang P(B) seperti yang bisa diketahui pada tabel berikut :

Tabel 2. Nilai Peluang P(B)

No	Gejala (A)	Nilai Probabilitas P(A)	Nilai Peluang P(B)
1	A1	1	0,38
2	A2	0,4	0,15
3	A3	0,6	0,23
4	A4	0,4	0,15
5	A5	0,2	0,07

c. Mencari nilai evidence P(A|B) dengan cara menghitung nilai probabilitas P(A) terhadap nilai peluang P(B), menggunakan rumus berikut :

$$P(A|B) = \frac{P(A|B)}{\sum_{B=1}^5 P(A|B)_i * P(A)_i}$$

$$= \frac{P(A|B)}{P(A)_1 * P(B)_1 + P(A)_2 * P(B)_2 + \dots + P(A)_5 * P(B)_5}$$

$$P(A|B) = \frac{P(A|B)}{(1*0,38) + (0,4*0,15) + (0,6*0,23) + (0,4*0,15) + (0,2*0,07)}$$



$$= \frac{P(A|B)}{1,38+0,55+0,83+0,55+0,27}$$

$$= 3,58$$

- d. Mencari nilai probabilitas hipotesis $P(B|A)$, dengan cara menghitung nilai probabilitas $P(A)$ dan nilai peluang $P(B)$ terhadap nilai evidence $P(A|B)$, menggunakan rumus berikut :

$$P(B|A) = \frac{P(A) \cdot P(B)}{P(A|B)} \tag{4}$$

$$P(B|A)_1 = \frac{1 \cdot 0,38}{3,58} = 0,106$$

$$P(B|A)_2 = \frac{0,4 \cdot 0,15}{3,58} = 0,016$$

$$P(B|A)_3 = \frac{0,6 \cdot 0,23}{3,58} = 0,038$$

$$P(B|A)_4 = \frac{0,4 \cdot 0,15}{3,58} = 0,016$$

$$P(B|A)_5 = \frac{0,2 \cdot 0,07}{3,58} = 0,003$$

- e. Setelah diketahui seluruh nilai $P(B|A)$, maka jumlahkan seluruh nilainya untuk mendapatkan nilai bayes, menggunakan rumus berikut :

$$\sum_{k=1}^n \text{Bayes} = P(B|A)_1 \cdot P(A)_1 + P(B|A)_2 \cdot P(A)_2 + \dots + P(B|A)_5 \cdot P(A)_5 \tag{5}$$

$$= (0,106 \cdot 1) + (0,016 \cdot 0,4) + (0,038 \cdot 0,6) + (0,016 \cdot 0,4) + (0,003 \cdot 0,2)$$

$$= 0,106 + 0,006 + 0,022 + 0,006 + 0,001$$

$$= 0,141$$

- f. Menghitung persentase nilai bayes, dengan menggunakan rumus berikut ini :

$$\text{Bayes}_{\text{persentase}} = \text{Bayes} \cdot 100\% \tag{6}$$

$$\text{Bayes}_{\text{persentase}} = 0,141 \cdot 100\%$$

$$= 14,1\%$$

- g. Hasil diagnosa penyakit arteri koroner dengan menggunakan algoritma Naïve Bayes, dapat diketahui tingkat persentase akurasi sebesar 14,1%.

3.2 Penerapan Algoritma C4.5

Melakukan proses perbandingan selanjutnya, penyelesaian masalah untuk mendiganosa penyakit arteri koroner yang dilakukan dengan menggunakan algoritma C.45 dapat dilakukan seperti langkah – langkah berikut:

- a. Membuat tabel keputusan, berdasarkan sampel data yang telah diperoleh, seperti pada tabel berikut:

Tabel 3. Tabel Keputusan

Pasien	Gejala (A)					Tingkatan
	Nyeri Dada (A1)	Sesak Nafas (A2)	Keringat Dingin (A3)	Mual (A4)	Kelelahan (A5)	
Pasien 1	Ya	Tidak	Tidak	Tidak	Tidak	Ringan
Pasien 2	Ya	Ya	Tidak	Tidak	Tidak	Sedang
Pasien 3	Ya	Tidak	Tidak	Ya	Tidak	Sedang
Pasien 4	Ya	Tidak	Ya	Tidak	Tidak	Sedang
Pasien 5	Ya	Tidak	Ya	Ya	Ya	Berat

- b. Menghitung nilai entropy dari setiap gejala penyakit arteri koroner, dengan menggunakan rumus berikut :

$$\text{Entropy}(S) = \sum_{i=0}^n -P_i \cdot \log_2 P_i \tag{7}$$

$$\text{Entropy}(\text{Total}) = \left(-\frac{1}{5} \cdot \log_2 \left(\frac{1}{5}\right)\right) + \left(-\frac{3}{5} \cdot \log_2 \left(\frac{3}{5}\right)\right) + \left(-\frac{1}{5} \cdot \log_2 \left(\frac{1}{5}\right)\right)$$

$$= ((-0,2) \cdot (-2,321)) + ((-0,6) \cdot (-0,736)) + ((-0,2) \cdot (-2,321))$$

$$= (0,046 + 0,441 + 0,046)$$

$$= 0,533$$

$$\text{Entropy}(A1, \text{Ya}) = \left(-\frac{1}{5} \cdot \log_2 \left(\frac{1}{5}\right)\right) + \left(-\frac{3}{5} \cdot \log_2 \left(\frac{3}{5}\right)\right) + \left(-\frac{1}{5} \cdot \log_2 \left(\frac{1}{5}\right)\right)$$

$$= ((-0,2) \cdot (-2,321)) + ((-0,6) \cdot (-0,736)) + ((-0,2) \cdot (-2,321))$$

$$= (0,046 + 0,441 + 0,046)$$

$$= 0,533$$

$$\text{Entropy}(A1, \text{Tidak}) = \left(-\frac{0}{5} \cdot \log_2 \left(\frac{0}{5}\right)\right) + \left(-\frac{0}{5} \cdot \log_2 \left(\frac{0}{5}\right)\right) + \left(-\frac{0}{5} \cdot \log_2 \left(\frac{0}{5}\right)\right)$$

$$= (0 \cdot 0) + (0 \cdot 0) + (0 \cdot 0)$$

$$= (0 + 0 + 0)$$

$$= 0$$

$$\text{Entropy}(A2, \text{Ya}) = \left(-\frac{0}{5} \cdot \log_2 \left(\frac{0}{5}\right)\right) + \left(-\frac{1}{5} \cdot \log_2 \left(\frac{1}{5}\right)\right) + \left(-\frac{0}{5} \cdot \log_2 \left(\frac{0}{5}\right)\right)$$

$$= (0 \cdot 0) + ((-0,2) \cdot (-2,321)) + (0 \cdot 0)$$



$$= (0 + 0,046 + 0)$$

$$= 0,046$$

$$\text{Entropy}(A2, \text{Tidak}) = \left(-\frac{1}{5} * \log_2 \left(\frac{1}{5}\right)\right) + \left(-\frac{2}{5} * \log_2 \left(\frac{2}{5}\right)\right) + \left(-\frac{1}{5} * \log_2 \left(\frac{1}{5}\right)\right)$$

$$= ((-0,2) * (-2,321)) + ((-0,4) * (-1,321)) + ((-0,2) * (-2,321))$$

$$= (0,046 + 0,528 + 0,046)$$

$$= 0,620$$

$$\text{Entropy}(A3, \text{Ya}) = \left(-\frac{0}{5} * \log_2 \left(\frac{0}{5}\right)\right) + \left(-\frac{1}{5} * \log_2 \left(\frac{1}{5}\right)\right) + \left(-\frac{1}{5} * \log_2 \left(\frac{1}{5}\right)\right)$$

$$= (0 * 0) + ((-0,2) * (-2,321)) + ((-0,2) * (-2,321))$$

$$= (0 + 0,046 + 0,046)$$

$$= 0,092$$

$$\text{Entropy}(A3, \text{Tidak}) = \left(-\frac{1}{5} * \log_2 \left(\frac{1}{5}\right)\right) + \left(-\frac{2}{5} * \log_2 \left(\frac{2}{5}\right)\right) + \left(-\frac{0}{5} * \log_2 \left(\frac{0}{5}\right)\right)$$

$$= ((-0,2) * (-2,321)) + ((-0,4) * (-1,321)) + (0 * 0)$$

$$= (0,046 + 0,528 + 0)$$

$$= 0,574$$

$$\text{Entropy}(A4, \text{Ya}) = \left(-\frac{0}{5} * \log_2 \left(\frac{0}{5}\right)\right) + \left(-\frac{1}{5} * \log_2 \left(\frac{1}{5}\right)\right) + \left(-\frac{1}{5} * \log_2 \left(\frac{1}{5}\right)\right)$$

$$= (0 * 0) + ((-0,2) * (-2,321)) + ((-0,2) * (-2,321))$$

$$= (0 + 0,046 + 0,046)$$

$$= 0,092$$

$$\text{Entropy}(A4, \text{Tidak}) = \left(-\frac{1}{5} * \log_2 \left(\frac{1}{5}\right)\right) + \left(-\frac{2}{5} * \log_2 \left(\frac{2}{5}\right)\right) + \left(-\frac{0}{5} * \log_2 \left(\frac{0}{5}\right)\right)$$

$$= ((-0,2) * (-2,321)) + ((-0,4) * (-1,321)) + (0 * 0)$$

$$= (0,046 + 0,528 + 0)$$

$$= 0,574$$

$$\text{Entropy}(A5, \text{Ya}) = \left(-\frac{0}{5} * \log_2 \left(\frac{0}{5}\right)\right) + \left(-\frac{0}{5} * \log_2 \left(\frac{0}{5}\right)\right) + \left(-\frac{1}{5} * \log_2 \left(\frac{1}{5}\right)\right)$$

$$= (0 * 0) + (0 * 0) + ((-0,2) * (-2,321))$$

$$= (0 + 0 + 0,046)$$

$$= 0,046$$

$$\text{Entropy}(A5, \text{Tidak}) = \left(-\frac{1}{5} * \log_2 \left(\frac{1}{5}\right)\right) + \left(-\frac{3}{5} * \log_2 \left(\frac{3}{5}\right)\right) + \left(-\frac{0}{5} * \log_2 \left(\frac{0}{5}\right)\right)$$

$$= ((-0,2) * (-2,321)) + ((-0,6) * (-0,736)) + (0 * 0)$$

$$= (0,046 + 0,441 + 0)$$

$$= 0,487$$

Sehingga akan didapatkan nilai entropy dari setiap gejala penyakit arteri koroner seperti yang bisa diketahui pada tabel berikut :

Tabel 4. Nilai Entropy

Gejala (A)	Jumlah Pasien	Tingkatan			Nilai Entropy	
		Ringan	Sedang	Berat		
Total	5	1	3	1	0,533	
A1	Ya	5	1	3	1	0,533
	Tidak	0	0	0	0	0
A2	Ya	1	0	1	0	0,046
	Tidak	4	1	2	1	0,620
A3	Ya	2	0	1	1	0,092
	Tidak	3	1	2	0	0,574
A4	Ya	2	0	1	1	0,092
	Tidak	3	1	2	0	0,574
A5	Ya	1	0	0	1	0,046
	Tidak	4	1	3	0	0,487

- c. Menghitung nilai gain dari setiap gejala penyakit arteri koroner, dengan menggunakan rumus berikut :

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=0}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i)$$

(8)

$$\text{Gain}(\text{Total}, A1) = 0,533 - \left(\left(\frac{5}{5} * 0,533\right) + \left(\frac{0}{5} * 0\right)\right)$$

$$= (0,533 - (0,533 + 0))$$

$$= (0,533 - 0,533)$$

$$= 0$$

$$\begin{aligned} \text{Gain}(\text{Total}, A2) &= 0,533 - \left(\left(\frac{1}{5} * 0,046 \right) + \left(\frac{4}{5} * 0,620 \right) \right) \\ &= (0,533 - (0,009 + 0,496)) \\ &= (0,533 - 0,505) \\ &= 0,028 \end{aligned}$$

$$\begin{aligned} \text{Gain}(\text{Total}, A3) &= 0,533 - \left(\left(\frac{2}{5} * 0,092 \right) + \left(\frac{3}{5} * 0,574 \right) \right) \\ &= (0,533 - (0,036 + 0,344)) \\ &= (0,533 - 0,380) \\ &= 0,153 \end{aligned}$$

$$\begin{aligned} \text{Gain}(\text{Total}, A4) &= 0,533 - \left(\left(\frac{2}{5} * 0,092 \right) + \left(\frac{3}{5} * 0,574 \right) \right) \\ &= (0,533 - (0,036 + 0,344)) \\ &= (0,533 - 0,380) \\ &= 0,153 \end{aligned}$$

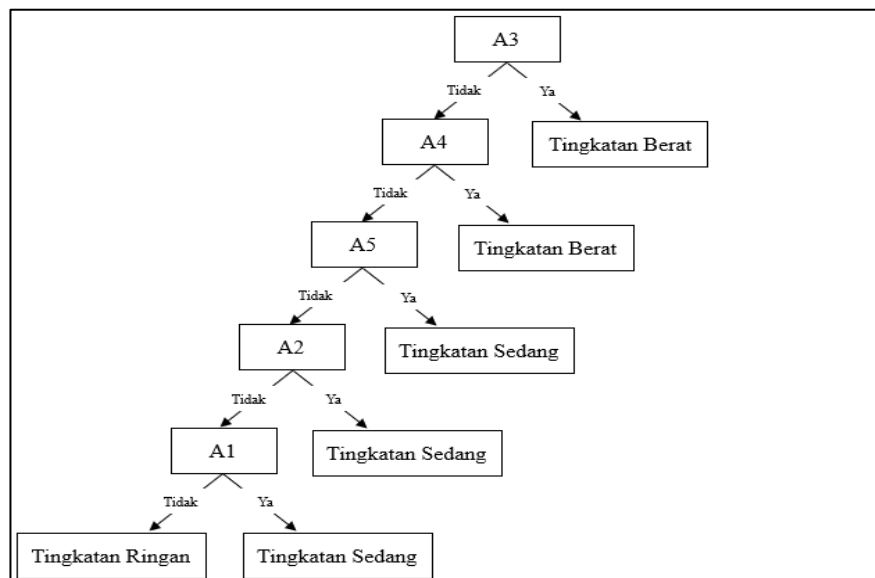
$$\begin{aligned} \text{Gain}(\text{Total}, A5) &= 0,533 - \left(\left(\frac{1}{5} * 0,046 \right) + \left(\frac{4}{5} * 0,487 \right) \right) \\ &= (0,533 - (0,009 + 0,389)) \\ &= (0,533 - 0,398) \\ &= 0,135 \end{aligned}$$

Sehingga akan didapatkan nilai gain dari setiap gejala penyakit arteri koroner seperti yang bisa diketahui pada tabel berikut :

Tabel 5. Nilai Gain

Gejala (A)	Nilai Gain
A1	0
A2	0,028
A3	0,153
A4	0,153
A5	0,135

- d. Membuat pohon keputusan berdasarkan nilai gain tertinggi dari setiap gejala penyakit arteri koroner, yang bisa dilihat pada Gambar 2 berikut:



Gambar 2. Pohon Keputusan

- e. Setelah mengetahui nilai gain dari setiap gejala penyakit arteri koroner dan telah membuat pohon keputusan, maka jumlahkan seluruh nilai gain untuk mengetahui tingkat persentase akurasi, menggunakan rumus berikut :

$$\begin{aligned} \sum_{k=1}^n \text{Gain} &= (A_1 + A_2 + A_3 + A_4 + A_5) * 100\% \\ &= (0 + 0,028 + 0,153 + 0,153 + 0,135) * 100\% \\ &= 0,469 * 100\% \\ &= 46,9\% \end{aligned}$$

Hasil diagnosa penyakit arteri koroner dengan menggunakan algoritma C.45, dapat diketahui tingkat persentase akurasi sebesar 46,9%.

3.3 Hasil Pengujian

Tahap pengujian dalam penelitian ini difokuskan pada perbandingan perhitungan dari kedua algoritma yang digunakan. Proses perbandingan ini dapat diketahui dari nilai persentase dari kedua algoritma. Nilai persentase dari kedua algoritma yang digunakan merupakan unsur yang menjadi tolak ukur berakhirnya perhitungan yang dilakukan. Apabila salah satu algoritma mendapatkan nilai persentase tertinggi, maka dapat ditentukan sebagai algoritma terbaik dalam mendiagnosa penyakit arteri koroner. Hasil diagnosa dari kedua algoritma yang digunakan dapat dilihat pada Tabel 6 berikut ini :

Tabel 6. Hasil Diagnosa Kedua Algoritma

No.	Algoritma	Nilai Persentase	Rangking
1	Naïve Bayes	14,1%	2
2	C.45	46,9%	1

Pada perhitungan tahap terakhir, yaitu pada perhitungan nilai persentase, maka perhitungan berhenti, karena hasil dari nilai persentase digunakan sebagai tolak ukur menentukan algoritma terbaik dalam mendiagnosa penyakit arteri koroner. Hasil perhitungan terakhir dari kedua algoritma tersebut adalah sebagai berikut :

a. Algoritma Naïve Bayes

$$Bayes_{persentase} = Bayes * 100\%$$

$$Bayes_{persentase} = 0,141 * 100\%$$

$$= 14,1\%$$

b. Algoritma C.45

$$\sum_{k=1}^n Gain = (A_1 + A_2 + A_3 + A_4 + A_5) * 100\%$$

$$= (0 + 0,028 + 0,153 + 0,153 + 0,135) * 100\%$$

$$= 0,469 * 100\%$$

$$= 46,9\%$$

Berdasarkan hasil dari penerapan perbandingan algoritma *naïve bayes* dan *c.45*, maka dapat diketahui bahwa, algoritma *c.45* menjadi algoritma terbaik yang paling cepat untuk mendiagnosa penyakit arteri koroner.

4. KESIMPULAN

Hasil dari penelitian merupakan sebuah penarikan kesimpulan, dimana kesimpulan menggambarkan hasil proses yang telah dilalui pada penelitian. Adapun hasil kesimpulan yang didapatkan dari penelitian bahwasannya dengan penerapan proses perbandingan algoritma dalam klasifikasi diagnosa penyakit arteri koroner, dilakukan dengan menggunakan dua algoritma yaitu algoritma Naïve Bayes dan algoritma C.45. Proses perbandingan algoritma Naïve Bayes dan algoritma C.45 dalam diagnosa penyakit arteri koroner, mendapatkan hasil algoritma terbaik berdasarkan nilai persentase terbesar yaitu algoritma C.45, dengan nilai 46,9%.

REFERENCES

- [1] A. F. Rianny and G. Testiana, "Penerapan Data Mining untuk Klasifikasi Penyakit Jantung Koroner Menggunakan Algoritma Naïve Bayes," *MDP Student Conf.*, vol. 2, no. 1, pp. 297–305, 2023, doi: 10.35957/mdp-sc.v2i1.4388.
- [2] S. Riska, J. I. Alham, E. Yosrita, and ; Rizqia Cahyaningtyas, "Sistem Diagnosis Penyakit Jantung Koroner Dengan Menggunakan Algoritma C4.5 Berbasis Website (Studi Kasus: RSUD Dr. Soedarso Pontianak)," *PETIR J. Pengkaj. dan Penerapan Tek. Inform.*, vol. 14, no. 2, 2021, doi: 10.33322/petir.v14i2.1388.
- [3] H. Azis, P. Purnawansyah, F. Fattah, and I. P. Putri, "Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung," *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 81–86, 2020, doi: 10.33096/ilkom.v12i2.507.81-86.
- [4] D. Derisma, "Perbandingan Kinerja Algoritma untuk Prediksi Penyakit Jantung dengan Teknik Data Mining," *J. Appl. Informatics Comput.*, vol. 4, no. 1, pp. 84–88, 2020, doi: 10.30871/jaic.v4i1.2152.
- [5] M. Dewi, T. H. Saragih, and R. Herteno, "Penerapan SMOTE-NCL untuk Mengatasi Ketidakseimbangan Kelas pada Klasifikasi Penyakit Jantung Koroner," *J. Inform. Polinema*, vol. 10, no. 1, pp. 27–34, 2023, doi: 10.33795/jip.v10i1.1394.
- [6] D. Sitanggang, N. Nicholas, V. Wilson, A. R. A. Sinaga, and A. D. Simanjuntak, "Implementasi Data Mining Untuk Memprediksi Penyakit Jantung Menggunakan Metode K-Nearest Neighbor Dan Logistic Regression," *J. Tek. Inf. dan Komput.*, vol. 5, no. 2, p. 493, 2022, doi: 10.37600/tekinkom.v5i2.698.
- [7] Alvina Felicia Watratan, Arwini Puspita. B, and Dikwan Moeis, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia," *J. Appl. Comput. Sci. Technol.*, vol. 1, no. 1, pp. 7–14, 2020, doi: 10.52158/jacost.v1i1.9.
- [8] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *J. KomtekInfo*, vol. 10, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.



- [9] S. Ucha Putri, E. Irawan, and F. Rizky, “Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan Algoritma C4.5,” *KESATRIA J. Penerapan Sist. Inf. (Komputer Manajemen)*, vol. 2, no. 1, pp. 39–46, 2021.
- [10] Y. S. Eirlangga and A. E. Syaputra, “Klasifikasi Penjurusan pada Sekolah Menengah Atas (SMA) dengan Metode Algoritma C4.5,” *J. Inf. dan Teknol.*, vol. 4, no. 3, pp. 160–165, 2022, doi: 10.37034/jidt.v4i3.235.
- [11] Farmawati and Narti, “Perbandingan Algoritma C4.5 dan Naive Bayes dalam Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Daring,” *JTIM J. Teknol. Inf. dan Multimed.*, vol. 4, no. 1, pp. 1–12, 2022.
- [12] U. K. Pemilihan Penggunaan Jenis Di Bidang Swasta HEnok Hayati Novia Cahya Utami, A. lia Hananto, and S. Shofiah Hilabi, “Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 Dan Naïve Bayes,” *J. Soc. Sci. Res.*, vol. 3, pp. 503–513, 2023.
- [13] N. Gusriani Fitri, S. Adilya, and F. Azizi, “Comparison of the Naive Bayes Classification System and C4.5 for the Diagnosis of Stroke Perbandingan Sistem Klasifikasi Naive Bayes dan C4.5 Untuk Diagnosa Penyakit stoke,” *SENTIMAS Semin. Nas. Penelit. dan Pengabd. Masy.*, pp. 49–55, 2023.
- [14] O. Prasetya and S. Machfud, “Analisis Prediksi Penyakit Demensia Alzheimer Menggunakan Metode Decision Tree C4.5 Dan Naive Bayes,” *J. Inform. Utama*, pp. 14–25, 2022.
- [15] M. F. Rizqullah, N. T. Raihana, and M. I. Jambak, “Komparasi Penerapan Algoritma C4.5, K-Nearest Neighbor, dan Naive Bayes untuk Keberlangsungan Pasien Gagal Jantung,” *Klik Kaji. Ilm. Inform. Dan Komput.*, vol. 4, no. 5, pp. 2580–2587, 2024, doi: 10.30865/klik.v4i5.1788.
- [16] A. Novriandy, “Implementasi Algoritma Naive Bayes dan Algoritma C4.5 dalam Klasifikasi Kelayakan Bantuan UMKM,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 1, pp. 208–217, 2023, doi: 10.30865/klik.v4i1.1099.
- [17] I. Nurjanah, J. Karaman, I. Widaningrum, D. Mustikasari, and S. Sucipto, “Penggunaan Algoritma Naive Bayes Untuk Menentukan Pemberian Kredit Pada Koperasi Desa,” *Explorer (Hayward)*, vol. 3, no. 2, pp. 77–87, 2023.
- [18] N. S. Fauziah and R. D. Dana, “Implementasi Algoritma Naive bayes dalam Klasifikasi Status Kesejahteraan Masyarakat Desa Gunungsari,” *Blend Sains J. Tek.*, vol. 1, no. 4, pp. 295–305, 2023, doi: 10.56211/blendsains.v1i4.234.
- [19] C. Nas, “Data Mining Prediksi Minat Calon Mahasiswa Memilih Perguruan Tinggi Menggunakan Algoritma C4.5,” *J. Manaj. Inform.*, vol. 11, no. 2, pp. 131–145, 2021, doi: 10.34010/jamika.v11i2.5506.
- [20] F. F. Nugraha, I. Sunandar, and C. Juliane, “Penerapan Data Mining Dengan Metode Klasifikasi Menggunakan Algoritma C4.5,” *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. March, pp. 10–20, 2022.
- [21] R. Girsang, E. F. Ginting, and M. Hutasuhut, “Penerapan Algoritma C4.5 Pada Penentuan Penerima Program Bantuan Pemerintah Daerah,” *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 4, p. 449, 2022, doi: 10.53513/jursi.v1i4.5727.
- [22] E. I. Scandea, M. Aqsha, R. Sugiarto, F. Lestari, and D. Hartanti, “Penerapan Data Mining Untuk Menganalisis Data Faktor Resiko Penyakit Jantung Menggunakan Metode Logistic Regression,” in *Prosiding Seminar Nasional Teknologi Informasi dan Bisnis (SENATIB) 2023*, 2023, pp. 683–688.
- [23] D. A. M. Reza, A. M. Siregar, and Rahmat, “Penerapan Algoritma K-Nearest Neighbord Untuk Prediksi Kematian Akibat Penyakit Gagal Jantung,” *Sci. Student J. Information, Technol. Sci.*, vol. III, no. 1, pp. 105–112, 2022.