

## **Sistem Diagnosis Penyakit Jantung Koroner Dengan Menggunakan Algoritma C4.5 Berbasis Website (Studi Kasus: RSUD Dr. Soedarso Pontianak)**

*Siti Riska Julianti Ihnur Alham<sup>1</sup>; Efy Yosrita<sup>2</sup>; Rizqia Cahyaningtyas<sup>3</sup>*

<sup>1, 2, 3</sup>Institut Teknologi PLN

<sup>1</sup> si1731151@itpln.ac.id

<sup>2</sup>efy.yosrita@itpln.ac.id

<sup>3</sup>rizqia@itpln.ac.id

### **ABSTRACT**

*Heart disease is also called coronary heart disease. Saat this method used by the hospital to give clinical decisions made based on the science and experience of doctors in treating coronary heart disease patients. by utilizing patient data that has been stored in the database, then created a pattern of determining coronary heart disease with intelligent computational techniques so that uncertainty of diagnosis results can be avoided. The authors conducted a study to diagnose coronary heart disease using the decision tree method with the C4.5 algorithm. The advantages of this algorithm are easily understood, and can be described in the form of a decision tree based on the risk factors experienced by the patient. Risk factors include age, gender, hereditary factors, systolic blood pressure, diastolic blood pressure, total cholesterol levels, HDL (High Density Lipoprotein) levels, LDL (Low Density Lipoprotein) levels and triglyceride levels. Based on testing using confusion matrix with accuracy tingat parameter obtained value of 94.4%*

**Keywords:** *heart coroner, C4.5, decision tree*

### **ABSTRAK**

*Penyakit jantung disebut juga dengan penyakit jantung koroner. Saat ini metode yang digunakan oleh pihak rumah sakit untuk memberi keputusan klinis dibuat berdasarkan ilmu pengetahuan dan pengalaman dari dokter dalam menangani pasien-pasien penyakit jantung koroner. dengan memanfaatkan data-data pasien yang telah tersimpan dalam basis data, kemudian dibuat suatu pola penentuan penyakit jantung koroner dengan teknik komputasi cerdas sehingga ketidak pastian hasil diagnosis dapat dihindari. Penulis melakukan penelitian untuk mendiagnosis penyakit jantung koroner menggunakan metode decision tree dengan algoritma C4.5. Kelebihan dari algoritma ini adalah dapat dipahami dengan mudah, dan dapat digambarkan dalam bentuk pohon keputusan berdasarkan faktor resiko yang dialami pasien. Faktor resikonya antara lain umur, jenis kelamin, faktor keturunan, tekanan darah sistolik, tekanan darah diastolik, kadar kolesterol total, kadar HDL (High Density Lipoprotein), kadar LDL (Low Density Lipoprotein) dan kadar trigliserida.Berdasarkan pengujian menggunakan confusion matrix dengan parameter tingat akurasi diperoleh nilai sebesar 94,4%*

**Kata kunci:** *jantung coroner, C4.5, pohon keputusan*

## 1. PENDAHULUAN

Penyakit jantung koroner (PJK) adalah penyakit yang disebabkan adanya plak yang menumpuk di dalam arteri koroner yang mensuplai oksigen ke otot jantung. Menurut statistik dunia, ada 9,4 juta kematian setiap tahun[1][2][3] yang disebabkan oleh penyakit kardiovaskuler dan 45% kematian tersebut disebabkan oleh penyakit jantung koroner, diperkirakan angka tersebut akan meningkat hingga 23,3 juta pada tahun 2030.[4]Penyakit jantung koroner (PJK) adalah penyakit jantung akibat penyempitan atau penyumbatan pembuluh nadi koroner. Penyempitan atau penyumbatan ini dapat menghentikan aliran darah ke otot jantung yang sering ditandai dengan rasa nyeri. Dalam kondisi lebih parah kemampuan jantung dalam memompa darah dapat hilang

Diagnosis penyakit jantung koroner dapat dilakukan menggunakan metode *decision tree* dengan algoritma C4.5. Kelebihan dari algoritma ini adalah dapat dipahami dengan mudah, dan dapat digambarkan dalam bentuk pohon keputusan berdasarkan faktor resiko yang dialami pasien. Faktor resikonya antara lain umur, jenis kelamin, faktor keturunan, tekanan darah sistolik, tekanan darah diastolik, kadar kolesterol total, kadar HDL (*High Density Lipoprotein*), kadar LDL (*Low Density Lipoprotein*) dan kadar trigliserid[5][6]. Penelitian ini menggunakan proses data mining. Data mining adalah sebuah proses, sehingga dalam melakukan prosesnya harus sesuai prosedur yaitu proses CRISP-DM (*Cross-Industry Standard Process for Data Mining*) [7][8][9] [10][11].

1. Business Understanding
2. Data Understanding
3. Data Preparation
4. Modelling
5. Evaluation
6. Deployment

Metode decision tree merupakan salah satu metode yang ada pada teknik klasifikasi dalam *data mining*. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan.[12] [13]Pohon keputusan juga berguna untuk mengekplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target.Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain ID3, C4.5, CART [13].

Algoritma C4.5 adalah algoritma klasifikasi data dengan teknik pohon keputusan yang memiliki kelebihan-kelebihan. Kelebihan ini misalnya dapat mengolah data numerik dan diskret, dapat menangani nilai atribut yang hilang, menghasilkan aturan - aturan yang mudah diinterpretasikan dan tercepat di antara algoritma algoritma yang lain [14] untuk membangun pohon keputusan pertama yang akan dilakukan yaitu memilih atribut sebagai akar. Kemudian membuat cabang untuk setiap nilai didalam akar tersebut. Langkah selanjutnya dengan membagi kasus dalam cabang. Kemudian ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama [14] [15][16]. Perhitungan nilai *entropy* yang dapat dilihat pada persamaan di bawah ini :

$$\text{Entropy}(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (1)$$

Keterangan:

- S = Himpunan Kasus  
n = Jumlah Partisi S  
Pi = Porsi dari Si terhadap S

*Gain* adalah salah satu *attribute selection measure* yang digunakan untuk memilih test atribut tiap node pada tree. Atribut dengan information *gain* tertinggi dipilih sebagai test atribut dari suatu *node* [14][15][16].

$$Gain(S, A) = Entropy(S) - \sum_{t=1}^n \frac{|S_t|}{|S|} * Entropy(S_t) \quad (2)$$

Keterangan:

S = Himpunan kasus

A = Atribut

n = Jumlah partisi atribut A

|S<sub>i</sub>| = Jumlah kasus pada partisi ke -i

## 2. METODE/PERANCANGAN PENELITIAN

### 2.1. Business Understanding

Pada tahap pertama berfokus pada kebutuhan berdasarkan penilaian bisnis, tujuan bisnis berdasarkan pengolahan data untuk mengambil keputusan mengenai diagnosis jantung koroner. Saat ini metode yang digunakan oleh pihak rumah sakit untuk memberi keputusan klinis dibuat berdasarkan ilmu dan pengetahuan dari pasien-pasien penyakit jantung koroner.

### 2.2. Data Understanding

Tahap data *understanding* penulis melakukan pemahaman data dari faktor risiko penyakit jantung koroner yang terdiri dari umur, jenis kelamin, faktor keturunan, tekanan darah sistolik, tekanan darah diastolik, kadar kolesterol total, kadar HDL, kadar LDL dan kadar trigliserida. Data yang digunakan tahun 2017 sampai 2021.

**Tabel 1.** Data Mentah

No	Umur	Jenis kelamin	Tekanan darah sistolik	Tekanan darah diastolic	kadar kolesterol total	HDL	LDL	Trigliserida	Faktor keturunan	Target
1	52	0	180	98	147	47,4	76	263	1	1
2	48	0	127	75	172	22	100	252	1	1
3	64	0	122	85	177	138,2	25,1	102	0	0
4	38	1	153	98	156	30	100	228	1	1
5	64	1	128	82	163	75	37,6	145	0	0
6	40	0	118	84	164	100	33	144	0	0
7	75	1	130	70	272	45	177	248	0	0
8	64	1	144	91	247	59,5	161	131	1	1
9	46	0	127	81	128	65	49,9	144	0	0
10	66	0	132	53	146	31,2	97	151	1	1
11	64	0	120	70	134	22	106	112	0	1
12	70	1	131	83	210	63	182	220	1	1
13	67	1	200	120	121	226	24,8	151	1	0
14	37	1	118	77	143	84,8	30,3	117	0	0
15	62	0	128	83	108	70,9	54	142	0	0
16	51	1	110	70	191	60,7	144	82	0	0
17	8	1	127	75	22	150	240	210	1	1
18	59	0	113	68	202	31,2	160	18	1	1
19	46	1	153	79	186	61	131	78	0	1
20	64	0	126	83	113	164	32,1	105	0	0

### 2.3. Data Preparation

Tahapan ini seluruh data dilakukan klasifikasi ke dalam kelas-kelas tertentu pada setiap atribut agar dapat diolah dengan algoritma C4.5.

**Tabel 2.** Data Klasifikasi

No	Umur	Jeniskelamin	Tekanan darah sistolik	Tekanan darah diastolik	kadar kolesterol total	HDL	LDL	Triglicerida	Faktor keturunan	Target
1	3	P	HB	HR	N	S	MO	T	1	YA
2	2	P	NT	N	N	R	MO	T	1	YA
3	3	P	NT	NT	N	T	O	N	0	TIDAK
4	2	L	HR	HR	N	R	MO	T	1	YA
5	3	L	NT	NT	N	T	O	N	0	TIDAK
6	2	P	N	NT	N	T	O	N	0	TIDAK
7	5	L	NT	N	T	S	T	T	0	TIDAK
8	3	L	HR	HR	T	S	O	N	1	YA
9	2	P	NT	NT	N	T	O	N	0	TIDAK
10	4	P	NT	N	N	R	O	BT	1	YA
11	4	P	NT	N	N	S	O	T	1	YA
12	4	L	NT	NT	BT	T	T	T	1	YA
13	4	L	HB	HB	N	T	O	BT	1	TIDAK
14	2	L	N	N	N	T	O	N	0	TIDAK
15	3	P	NT	NT	N	T	O	N	0	TIDAK
16	3	L	N	N	N	T	BT	N	0	TIDAK
17	1	L	NT	N	N	T	ST	T	1	YA
18	2	P	HR	NT	N	S	O	T	1	YA
19	2	L	HR	N	N	T	BT	N	0	YA
20	3	P	NT	NT	N	T	O	N	0	TIDAK

### 2.4. Modeling

Pada tahap pemodelan data yang sudah di klasifikasi, kemudian diproses menggunakan metode data mining yaitu algoritma C4.5. Berikut langkah-langkah perhitungan *entropy* dan *gain* pada setiap atribut :

#### 1. Perhitungan nilai entropy

Langkah pertama dalam perhitungan algoritma C4.5 yaitu mencari nilai *entropy*. Rumus mencari *entropy* sebagai berikut:

$$\text{Entropy}(S) = \sum_{i=1}^n - p_i * \log_2 p_i$$

Entropy total (S) =  $-(\text{jumlah target ya/total target}) * \log_2(\text{jumlah target ya/total target}) + (-\text{jumlah target tidak/total target}) * \log_2(\text{jumlah target tidak/total target})$

#### a) Entropy total

$$= (-10/20) * \log_2(10/20) + (-10/20) * \log_2(10/20) \\ = 1$$

#### b) Entropy umur

$$= (-4/7) * \log_2(4/7) + (-3/7) * \log_2(3/7) \\ = 0,985$$

2. Perhitungan nilai gain

Setelah selesai melakukan perhitungan entropy pada masing-masing subset atribut, dilakukannya perhitungan gain untuk memilih atribut yang digunakan sebagai akar dengan nilai gain tertinggi dengan rumus:

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

Gain = Entropy total - (target/ total target \*Entropy target)

a) Gain umur

$$\begin{aligned} &= 1 - (1/20*0) + (7/20*0,985) + (7/20*0,863) + (4/20*0,811) + (1/20*0) \\ &= 0,191 \end{aligned}$$

b) Gain Faktor keturunan

$$\begin{aligned} &= 1 - (10/20*0,469) + (10/20*0,469) \\ &= 0,531 \end{aligned}$$

3. Ulangi Langkah perhitungan *entropy* dan gain ke semua atribut hingga mendapatkan hasil seperti table 3 Dari hasil tabel maka didapatkan nilai gain tertinggi yaitu faktor keturunan sebagai akar pertama, untuk akar berikutnya maka dilakukan perhitungan untuk *entropy* dan gain pada atribut selanjutnya.

**Tabel 3.** Hasil Perhitungan Nilai Entropy dan Gain

		JUMLAH (S)	YA (Si)	TIDAK (Si)	ENTROPY	GAIN
TOTAL		20	10	10	1	
UMUR	1	1	1	0	0	0,191
	2	7	4	3	0,985	
	3	7	2	5	0,863	
	4	4	3	1	0,811	
	5	1	0	1	0	
JENIS KELAMIN	L	10	5	5	1	0
	P	10	5	5	1	
						0,353
TEKANAN DARAH SISTOLIK	N	3	0	3	0	
	NT	11	5	6	0,994	
	HR	4	4	0	0	
	HS	0	0	0	0	
	HB	2	1	1	1	
TEKANAN DARAH DIASTOLIK	N	8	5	3	0,954	0,294
	NT	8	2	6	0,811	
	HR	3	3	0	0	
	HS	0	0	0	0	
	HB	1	0	1	0	

KADAR KOLESTEROL TOTAL					0,052
	N	17	8	9	0,998
	BT	1	1	0	0
	T	2	1	1	1
HDL					0,333
	R	3	3	0	0
	S	5	4	1	0,722
	T	12	3	9	0,811
LDL					0,249
	O	12	4	8	0,918
	MO	3	3	0	0
	BT	2	1	1	1
	T	2	1	1	1
	ST	1	1	0	0
TRIGLISERIDA					0,322
	N	10	2	8	0,722
	BT	2	1	1	1
	T	8	7	1	0,544
	ST	0	0	0	0
FAKTOR KETURUNAN					0,531
	1	10	9	1	0,469
	0	10	1	9	0,469

## 2.5. Evaluation

Tahap evaluasi merupakan hasil dari data mining, dapat disimpulkan berdasarkan pohon keputusan yang dihasilkan. Pada pengujian hasil akurasi menggunakan metode confussion matrix.

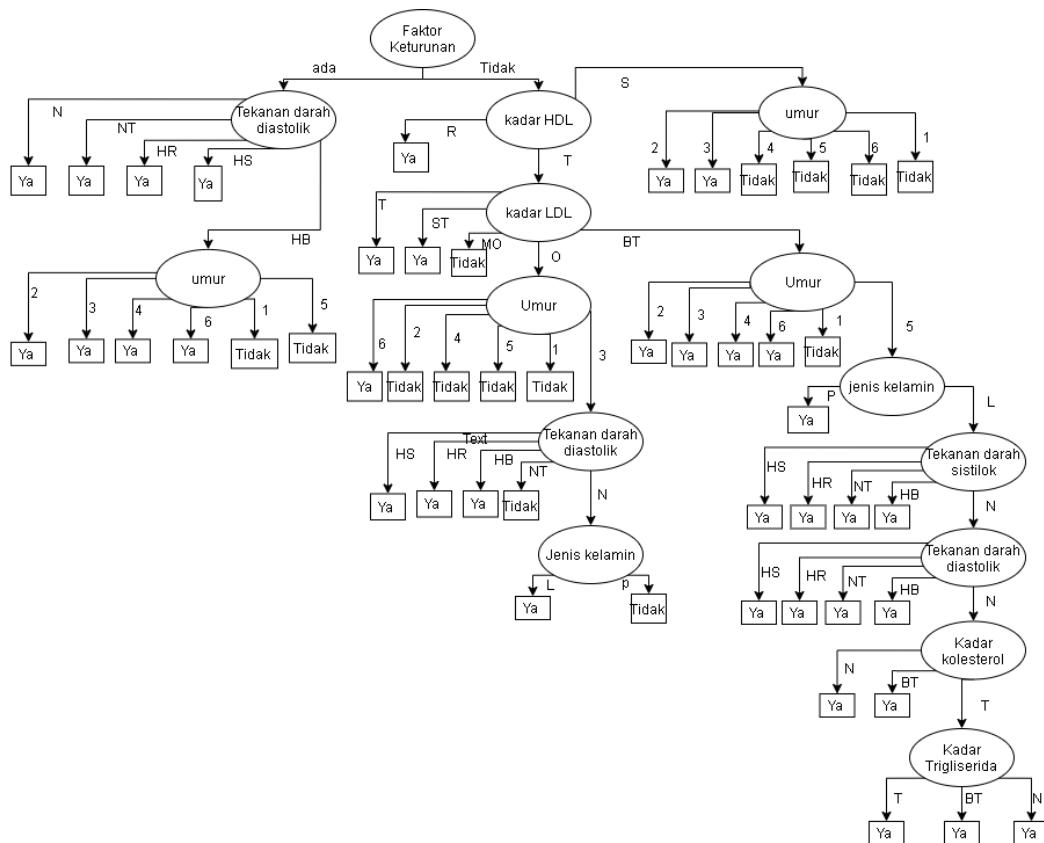
## 2.6. Deployment

Pada tahap ini penulis melakukan penyusunan laporan dari proses yang telah dilakukan, untuk penyajian hasil ditampilkan dengan tampilan yang mudah dipahami agar dapat diimplementasikan.

## 3. HASIL DAN PEMBAHASAN

Data yang digunakan 300 data berisi variabel umur, jenis kelamin, tekanan darah sistolik, tekanan darah diastolik, kadar kolesterol total, kadar HDL, kadar LDL, kadar trigliserida, dan faktor keturunan, Kemudian di implementasikan menggunakan metode *decision tree*. data akan diolah menjadi data pelatihan. dan data pengujian.

Data pelatihan digunakan untuk membuat pohon keputusan di dalam sistem. Pohon keputusan digunakan sebagai acuan untuk menentukan apakah seseorang berdasar data yang digunakan terdiagnosis jantung koroner atau tidak. Berdasarkan hasil 110 data latih yang digunakan maka menghasilkan sebuah pohon keputusan seperti gambar 1



Keterangan

1 : < 35  
2 : 35-44  
3 : 45-54  
4 : 55-64  
5 : 65-74  
6 : >=75

N : Normal  
NT : Normal Tinggi  
HR : Hipertensi Rendah  
HS : Hipertensi Sedang  
HB : Hipertensi Berat  
BT : Batas Tinggi  
T : Tinggi

ST : Sangat Tinggi  
R : Rendah  
S : Sedang  
O : Optimal  
MO : Mendekati Optimal  
L : Laki-Laki  
P : Perempuan

**Gambar 1.** Pohon Keputusan

Data pengujian digunakan untuk mengetahui keakuratan dari hasil klasifikasi diukur dengan menggunakan *confusion matrix*, [17] Dimana data hasil prediksi berada diantara dua kelas yaitu kelas positif dan kelas negatif. Pengukuran digunakan adalah pengukuran akurasi. Berikut tabel *confusion matrix* dengan jumlah data uji 90 data dari hasil data pengujian menghasilkan data 85 data teridentifikasi benar dan 5 data teridentifikasi salah.

**Tabel 4.** Confusion Matrix

actual	Prediksi	
	kelas = ya	kelas = tidak
kelas = ya	TP (True Positive) 68	FP (False Positive) 3
kelas = tidak	FN (False Negative) 2	TN (True Negative) 17

Perhitungan akurasi dilakukan dengan membagi jumlah data yang diklasifikasikan secara benar dengan total data sampel yang diuji. Rumus dari akurasi yaitu:

$$akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Keterangan:

TP = 68

TN = 17

FP = 2

FN = 3

$$akurasi = \frac{68 + 17}{68 + 17 + 2 + 3}$$

$$akurasi = \frac{85}{90}$$

$$akurasi = 0,944 \times 100\%$$

$$akurasi = 94,4 \%$$

Berdasarkan hasil akurasi yang dilakukan didapatkan dengan hasil akurasi sebesar 94,4% dengan metode *decision tree*, maka penelitian ini menerapkan metode *decision tree* algoritma C4.5 pada kasus diagnosis jantung koroner.

#### **4. KESIMPULAN DAN SARAN**

##### **4.1. Kesimpulan**

Dari hasil penelitian yang sudah dilakukan, kesimpulan yang dapat diambil adalah Algoritma C4.5 untuk diagnosis penyakit jantung koroner dapat diimplementasikan dengan baik dengan pengujian perbandingan data latih dan data uji sebesar 70 :30 dan menghasilkan akurasi sebesar 94,4%

##### **4.2. Saran**

Adapun saran yang dibutuhkan agar penelitian ini dapat berkembang diantara lain:

1. Program dapat mengimport data selain data bertipe.csv.
2. Pada pengembangan penelitian selanjutnya dapat dikembangkan dengan menggunakan algoritma *elman recurrent neural network* (ERNN).
3. Menambah jumlah data untuk memperbaiki hasil dari proses pelatihan

#### **UCAPAN TERIMAKASIH**

Ucapan terima kasih sebesar-besarnya saya ucapan kepada RSUD Dr. Soedarso Pontianak dan Institut Teknologi PLN kesempatan untuk melakukan penelitian ini.

#### **DAFTAR PUSTAKA**

- [1] V. M. Tarawan, R. Lesmana, H. Gunawan, and J. windi Gunandi, “GAMBARAN PENGETAHUAN PENCEGAHAN PENYAKIT JANTUNG KORONER PADA WARGA DUSUN III DESA MEKARMANIK KECAMATAN CIMENYAN KABUPATEN BANDUNG,” vol. 4, no. 1, pp. 10–14, 2020.
- [2] W. G. Amisi, jeini E. Nelwan, and F. k Kolibu, “HUBUNGAN ANTARA HIPERTENSI DENGAN KEJADIAN PENYAKIT JANTUNG KORONER PADA PASIEN YANG BEROBAT DI RUMAH SAKIT UMUM PUSAT Prof. Dr. R. D. KANDOU MANADO Windy,” vol. 7.
- [3] K. Noviyanti and E. K. S. L, “Hubungan profil lipid dan C-reactive protein ( CRP ) dengan derajat stenosis koroner pada penyakit jantung koroner stabil,” vol. 10, no. 1, pp. 165–168,

- 2019, doi: 10.1556/ism.v10i1.332.
- [4] L. Ghani, M. D. Susilawati, and H. Novriani, “Faktor Risiko Dominan Penyakit Jantung Koroner di Indonesia,” *Bul. Penelit. Kesehat.*, vol. 44, no. 3, pp. 153–164, 2016, doi: 10.22435/bpk.v44i3.5436.153-164.
- [5] L. A. Larandipa, F. T. P. W. D. Triyanto, P. Jantung, F. Resiko, and J. S. Tiruan, “APLIKASI JARINGAN SYARAF TIRUAN UNTUK DIAGNOSA PENYAKIT JANTUNG KORONER ( PJK ) DENGAN METODE BACKPROPAGATION Skripsi ini dibahas aplikasi Jaringan Syaraf Tiruan ( JST ) untuk diagnosa Penyakit Jantung Koroner ( PJK ) dengan metode backpropagation . Aplik,” vol. 01, no. 1, 2013.
- [6] N. Effendy and A. Faisal, “PREDIKSI PENYAKIT JANTUNG KORONER ( PJK ) BERDASARKAN FAKTOR RISIKO MENGGUNAKAN JARINGAN SYARAF TIRUAN BACKPROPAGATION,” vol. 2008, no. Snati, 2008.
- [7] S. Masripah, “Komparasi Algoritma Klasifikasi Data Mining untuk Evaluasi Pemberian Kredit,” vol. 3, no. 1, pp. 187–193, 2016.
- [8] A. Asistyasari, “PENCARIAN PRODUK E-COMMERCE BERDASARKAN JUDUL DAN DESKRIPSI MENGGUNAKAN VECTOR SPACE MODEL,” pp. 53–58, 2020.
- [9] U. Pauziah, “ANALISIS PENENTUAN KARYAWAN TERBAIK MENGGUNAKAN METODE ALGORITMA NAIVE BAYES ( STUDI KASUS PT . XYZ ),” no. April, pp. 94–102, 2017.
- [10] N. Nuraeni, “Penentuan Kelayakan Kredit Dengan Algoritma Naïve Bayes Classifier : Studi Kasus Bank Mayapada Mitra Usaha Cabang PGC,” vol. III, no. 1, pp. 9–15, 2017.
- [11] F. Nurhassan, N. Hikmah, and D. Y. Utami, “Perbandingan algoritma c4.5, knn, dan naive bayes untuk penentuan model klasifikasi penanggung jawab bsi entrepreneur center,” vol. 14, no. 2, pp. 169–174, 2018.
- [12] S. Wahyuni, kana S. S, and M. I. Perangin-angin, “IMPLEMENTASI RAPIDMINER DALAM MENGANALISA DATA MAHASISWA DROP OUT,” vol. 10, pp. 2013–2016, 2017.
- [13] F. Dwi Meliani Achmad, Budanis, Slamat, “Klasifikasi Data Karyawan Untuk Menentukan Jadwal Kerja Menggunakan Metode Decision Tree,” *J. IPTEK*, vol. 16, no. 1, pp. 18–23, 2012, [Online]. Available: <http://jurnal.itats.ac.id/wp-content/uploads/2013/06/3.-BUDANIS-FINAL-hal-17-23.pdf>.
- [14] A. Sonita and R. Kundari, “APLIKASI SELEKSI CALON PENDONOR,” vol. VI, no. September, 2019.
- [15] L. Hermawanti, “Penerapan algoritma klasifikasi c4.5 untuk diagnosis penyakit kanker payudara,” vol. 7, no. 1, pp. 57–64, 2012.
- [16] B. Sugara, D. Widyatmoko, B. S. Prakoso, and D. M. Saputro, “Penerapan Algoritma C4.5 untuk Deteksi Dini Autisme Pada Anak,” *Semin. Nas. Teknol. Inf. dan Komun.*, vol. 2018, no. Sentika, pp. 87–96, 2018.
- [17] D. Hanifah, C. Prianto, and N. Riza, *Buku laporan rancang bangun aplikasi pengambilan keputusan dalam pemilihan karyawan pada kegiatan akademik perusahaan dengan menggunakan perbandingan metode topsis dan metode promethee*. 2020.