

Penerapan Algoritma Naïve Bayes Untuk Mengetahui Pasien Penyakit Gagal Jantung

Rdm Fiqriansyah¹, Farchan Akbar², Virginia Casanova Andiko³, Khalil Gibran Ahmad⁴, Errissya Rasywir⁵,
Despita Meisak⁶, Yovi Pratama⁷, Ayu Feranika⁸

^{1,2} Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dinamika Bangsa, Kota Jambi, Indonesia

Email : ¹fiqribae01@gmail.com, ²farchanakbar80@gmail.com, ³virginianova97@gmail.com,
⁴khalilgibranaahmad30@gmail.com, ⁵errissya.jurnal@gmail.com, ⁶despitam88@gmail.com, ⁷yovi.pratama@gmail.com,
⁸ayu.feranika2@gmail.com

Email Penulis Korespondensi : ayu.feranika2@gmail.com

Abstrak – Setiap tahunnya lebih dari 36 juta orang meninggal karena Penyakit Tidak Menular (PTM) (63% dari seluruh kematian). Lebih dari 9 juta kematian yang disebabkan oleh penyakit tidak menular terjadi sebelum usia 60 tahun, dan 90% dari kematian “dini” tersebut terjadi di negara berpenghasilan rendah dan menengah. Secara Global PTM penyebab kematian nomor satu setiap tahunnya adalah penyakit kardiovaskuler. Penyakit Kardiovaskuler adalah penyakit yang disebabkan gangguan fungsi jantung dan pembuluh darah, seperti : penyakit coroner, penyakit gagal jantung atau payuh jantung, hipertensi dan stroke. Diperkirakan sebanyak 17,3 juta kematian disebabkan oleh penyakit kardiovaskuler.

Kata Kunci : data Pasien; algoritma naïve bayes; WEKA; klasifikasi; akurasi

Abstract - Every year more than 36 million people die from non-communicable diseases (NCD) (63% of all deaths). More than 9 million deaths are caused by non-communicable diseases before the age of 60, and 90% of these “early” deaths occur in low- and middle-income countries. Globally, PTM is the number one cause of death every year is cardiovascular disease. Cardiovascular disease is a disease caused by impaired function of the heart and blood vessels, such as coronary disease, heart failure or heart failure, hypertension and stroke. It is estimated that 17.3 million deaths are caused by cardiovascular disease.

Keywords: patient data; naïve Bayes algorithm; WEKA; classification; accuracy

1. PENDAHULUAN

Kesehatan merupakan faktor terpenting yang harus dijaga oleh setiap individu, agar dalam menjalani aktivitas sehari-hari menjadi lebih produktif, tidak mudah lelah dan lebih focus dalam menyelesaikan suatu *project*. Namun, saat ini PTM menjadi penyebab utama kematian setiap tahunnya terdapat lebih dari 36 juta orang yang meninggal disebabkan PTM atau setara 63% dari total seluruh kematian. Secara terperinci PTM terjadi dibawah usia 60 tahun, dan mayoritas sampai 90% dari kematian tersebut terjadi di negara dengan tingkat perekonomian rendah. Salah satu PTM yang rentan terjadi terutama saat seorang individu berada pada usia produktif yaitu penyakit jantung (*Heart Disease*). Tingginya faktor kematian dari penyakit jantung karena kurangnya pengetahuan masyarakat terhadap gejala atau tanda-tanda saat seseorang tersebut mengidap penyakit ini. Penyakit jantung merupakan salah satu penyakit yang cukup berbahaya ketika menyerang seseorang, dimana penyebab utama penyakit jantung yaitu berasal dari pola hidup individu yang kurang sehat, mengkonsumsi makanan berkolesterol tinggi, penggunaan alkohol, tembakau, diet yang ekstrem serta penyebab lainnya. Penyakit jantung lebih rentan diderita oleh laki-laki, dimana perbandingannya sekitar satu dari tiga kemungkinan mengidap penyakit jantung sebelum usia 60 tahun. Pada wanita perbandingannya sekitar satu dari sepuluh kemungkinan yang mengidap penyakit jantung. Rasio yang cukup tinggi terkait penyakit jantung ini, menjadikannya sebagai salah satu penyakit yang akan menghasilkan sejumlah besar data pasien pengidap penyakit jantung. Bahkan industri perawatan kesehatan saat ini mampu memberikan data kompleks tentang pasien, sumber daya rumah sakit, diagnosa penyakit, catatan pasien elektronik, peralatan medis, dll. Penambahan data menjadi kian disoroti terutama pada berbagai layanan kesehatan. Besarnya jumlah data merupakan sumber daya utama yang akan diproses dan dianalisis guna diekstraksi pengetahuan yang memungkinkan dukungan untuk penghematan biaya dan pengambilan keputusan. Data mining merupakan proses untuk menemukan pola dan tren yang sebelumnya tidak diketahui dalam database serta menggunakan informasi tersebut dalam membangun model prediksi. Penambahan data menyediakan seperangkat alat dan teknik yang dapat diterapkan pada data yang diproses ini untuk menemukan pola tersembunyi

dan juga memberikan sumber pengetahuan tambahan kepada profesional kesehatan untuk membuat keputusan yang lebih akurat.[1]

Algoritma *Naive Bayes* merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris *Thomas Bayes*, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai *Teorema Bayes*. *Teorema* tersebut dikombinasikan dengan *Naive* dimana diasumsikan kondisi antar atribut saling bebas.[2]

2. METODOLOGI PENELITIAN

Metodologi penelitian yang digunakan dalam pembuatan Penelitian ini adalah dengan Algoritma *Naive Bayes*. Algoritma tersebut akan melakukan klasifikasi terhadap dataset pasien penyakit gagal jantung.

2.1 Tahapan Penelitian



2.1.1 Data Pasien

Data yang di dapat Data Riset Kesehatan Dasar 2013, Badan Litbangkes Kementerian Kesehatan RI dan Data Penduduk Sasaran, Pusdatin Kementerian Kesehatan RI. (<https://m5.gs/bE42Nm>). atribut yang digunakan dengan 6 atribut seperti : Nama, Sesak nafas, Denyut nadi, Mual muntah, gelisah, gagal jantung.

2.1.2 Format Arff

WEKA mendukung beberapa format file dalam inputnya diantaranya adalah format arff, Data Pasien tersebut di buat menggunakan NotePad++ dengan format arff, karena WEKA merupakan tipe file text yang berisi berbagai instance data yang berhubungan dengan suatu set attribute data. Kemudia data-data ini disusun dengan menggunakan tanda koma (,) sebagai pemisah data per- atributnya.

2.1.3 Training

Pada tahap ini dilakukan proses training oleh algoritma klasifikasi naïve bayes untuk membentuk sebuah model classifier pada data pasien yang sudah ada. Model ini merupakan representasi pengetahuan yang akan digunakan untuk prediksi kelas data baru yang belum ada.

2.1.4 Naïve Bayes klasifikasi

Proses pengklasifikasi data menggunakan tools WEKA terhadap data pasien yang sudah di testing menggunakan model classifier yang sudah dibentuk.

2.1.5 Akurasi Hasil

menunjukkan kedekatan hasil pengukuran dengan nilai sesungguhnya.

2.1 Dataset

Dataset yang digunakan yaitu Data Pasien Penyakit Jantung yang berasal dari Data Riset Kesehatan Dasar 2013, Badan Litbangkes Kementerian Kesehatan RI dan Data Penduduk Sasaran, Pusdatin Kementerian Kesehatan RI. (<https://m5.gs/bE42Nm>). atribut yang digunakan dengan 6 atribut seperti : Nama, Sesak nafas, Denyut nadi, Mual muntah, gelisah, gagal jantung.[1]

2.2 Naïve Bayes

Naïve Bayes merupakan algoritma yang terdapat pada Teknik klasifikasi. Naïve Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu prediksi peluang masa depan berdasarkan pengalaman sebelumnya yang dikenal dengan Teorema Bayes.[3]

2.3 WEKA

Merupakan rangkaian perangkat lunak pembelajaran mesin yang ditulis dalam Bahasa Java, dikembangkan di Universitas Waikato, Selandia Baru. Perangkat lunak ini memiliki banyak algoritma *machine learning* untuk keperluan data mining. Juga memiliki banyak tools untuk mengelola data, mulai dari *pre-processing*, *classification rules*, dan *visualization*. [4]

2.3 Klasifikasi

Artificial Intelligence (AI) mempunyai sistem yang dikembangkan pada pertengahan 1960 yang sering disebut dengan sistem klasifikasi. Salah satu pengembangan sistem klasifikasi dalam bidang Kesehatan untuk mendiagnosis berbagai penyakit, baik penyakit pada hewan, tumbuhan maupun manusia. Penelitian tersebut menjelaskan tentang sistem klasifikasi yang dapat mendeteksi penyebab, fakta atau kondisi yang dijadikan aturan dan menghasilkan informasi. [5]

2.4 Akurasi

Ukuran yang menentukan tingkat kemiripan antara hasil pengukuran dengan nilai yang sebenarnya diukur. dibutuhkan data atau referensi untuk membandingkan hasil klasifikasi yang dihasilkan. [6]

3. HASIL DAN PEMBAHASAN

3.1 DATA PASIEN

DATA PENYAKIT GAGAL JANTUNG					
Nama	Sesak Nafas	Denyut_Nadi	Mual_Muntah	Gelisah	Gagal Jantung
A	Lambat	Lemah	Ya	Ya	Sedang
B	Lambat	Normal	Ya	Ya	Sedang
C	Lambat	Kuat	Ya	Ya	Sedang
D	Normal	Lemah	Ya	Ya	Sedang
E	Normal	Normal	Ya	Ya	Parah
F	Normal	Kuat	Ya	Ya	Parah
G	Cepat	Lemah	Ya	Ya	Parah
H	Cepat	Normal	Ya	Ya	Parah
I	Cepat	Kuat	Ya	Ya	Parah
J	Lambat	Kuat	Tidak	Tidak	Sedang

Gambar 1. Data Pasien

Data kasus pasien yang diambil hanya 10 data yang terdiri dari 6 atribut antara lain :

1. nama,
2. sesak nafas,
3. denyut nadi,
4. mual muntah,
5. gelisah,
6. gagal jantung

3.2 DATA ARFF

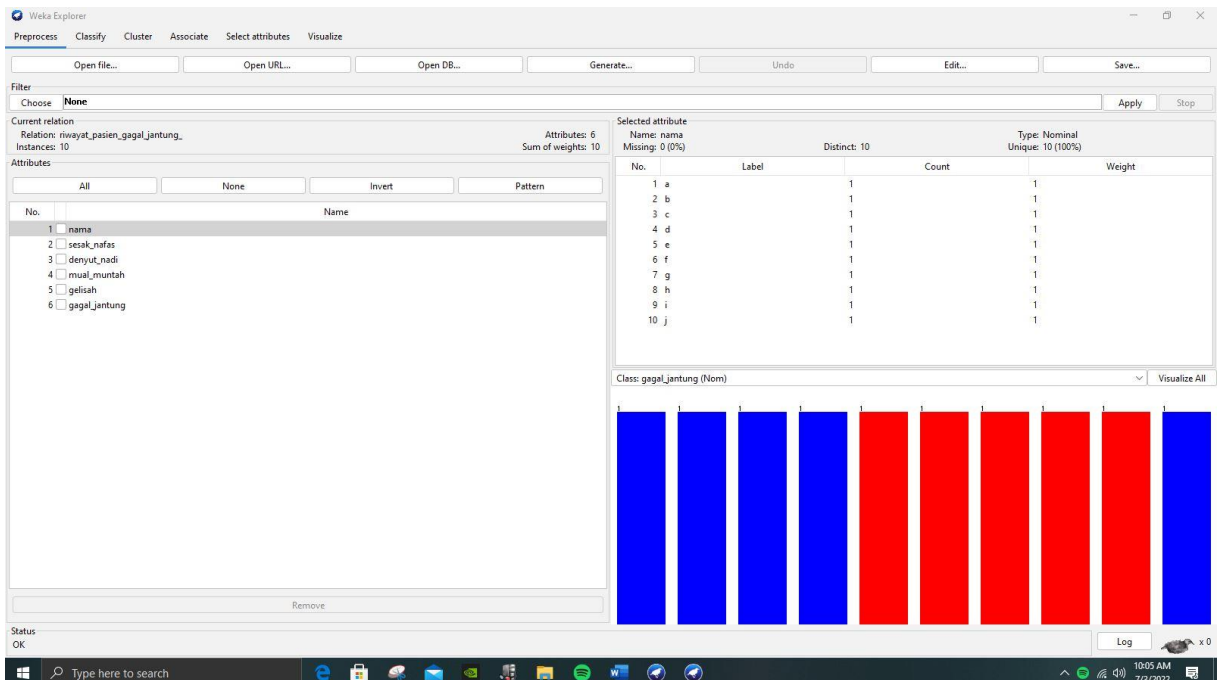
```

2 @relation riwayat_pasien_gagal_jantung_
3
4 @attribute nama {a, b, c, d, e, f, g, h, i, j}
5 @attribute sesak_nafas {lambat,normal,cepat}
6 @attribute denyut_nadi {lemah,normal,kuat}
7 @attribute mual_muntah {ya,tidak}
8 @attribute gelisah {ya,tidak}
9 @attribute gagal_jantung {sedang,parah}
10
11 @data
12 a lambat lemah ya ya sedang
13 b lambat normal ya ya sedang
14 c lambat kuat ya ya sedang
15 d normal lemah ya ya sedang
16 e normal normal ya ya parah
17 f normal kuat ya ya parah
18 g cepat lemah ya ya parah
19 h cepat normal ya ya parah
20 i cepat kuat ya ya parah
21 j lambat kuat tidak tidak sedang
    
```

Gambar 2. Data Arff

Data kasus pasien diatas yang telah diubah dan juga data-data tersebut disusun dengan menggunakan tanda (,) sebagai pemisah data dan attribute lalu data tersebut di ubah ke format arff proses tersebut dilakukan dengan menggunakan Notepad++ agar data tersebut dapat di proses dan di klasifikasikan dengan menggunakan tools WEKA.

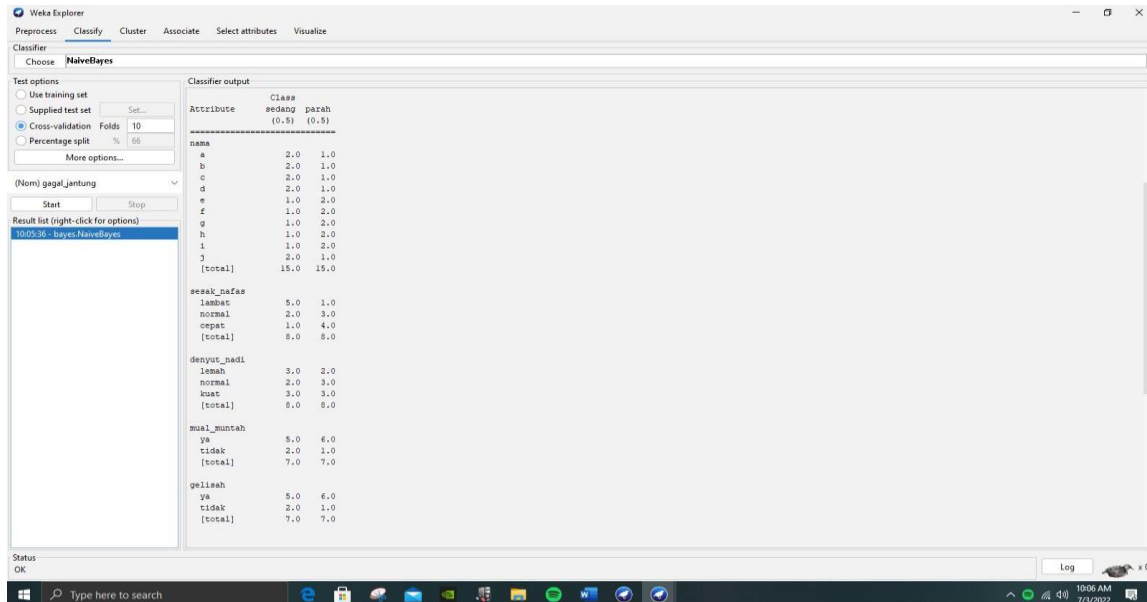
3.3 PEMROSESAN



Gambar 3. Pemrosesan

Tampilan pemrosesan data seperti normalisasi, standarisasi, downsampling untuk class tertentu pada data pasien riwayat penyakit gagal jantung.

3.4 HASIL DATA TRAINING



Gambar 4. Hasil Data Training

Berikut adalah hasil proses training oleh algoritma klasifikasi naïve bayes yang sudah membentuk sebuah model classifier pada data pasien yang sudah ada, model ini sebagai representasi pengetahuan yang akan digunakan untuk prediksi kelas data baru yang belum ada.

=== Classifier model (full training set) ===

Naive Bayes Classifier

Attribute	Class	
	sedang (0.5)	parah (0.5)

Nama		
A	2.0	1.0
B	2.0	1.0
C	2.0	1.0
D	2.0	1.0
E	1.0	2.0
F	1.0	2.0
G	1.0	2.0
H	1.0	2.0
I	1.0	2.0
J	2.0	1.0
Total	15.0	15.0

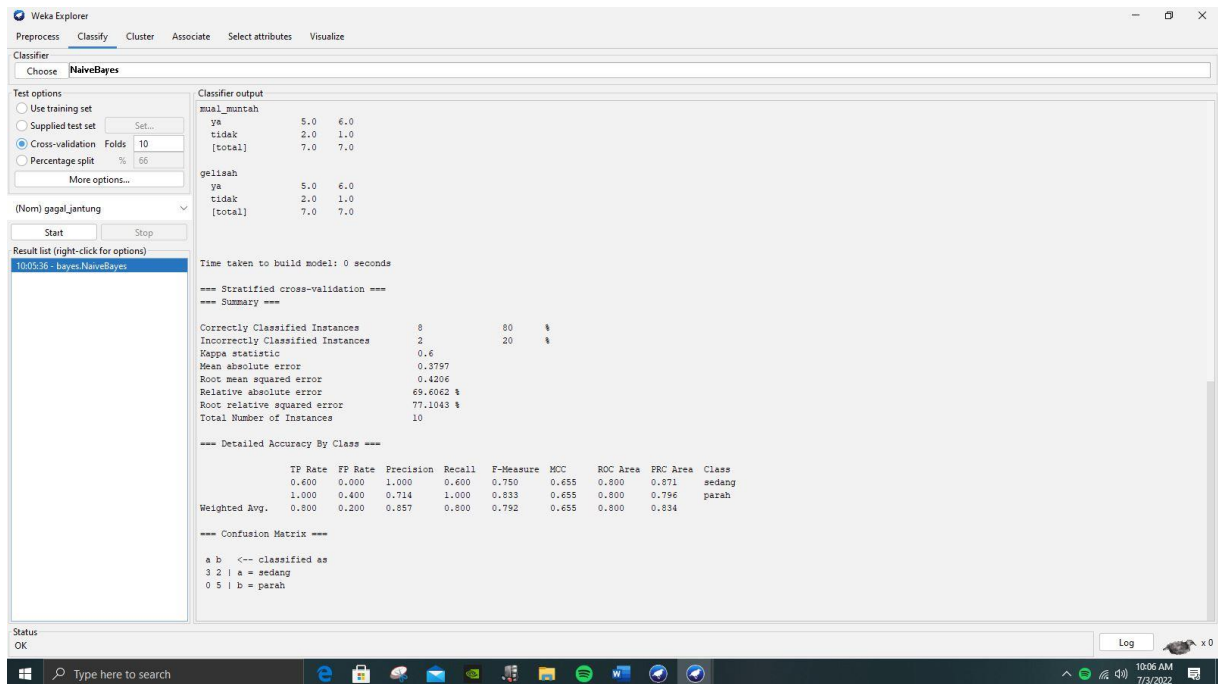
Sesak_Nafas		
Lambat	5.0	1.0
Normal	2.0	3.0
Cepat	1.0	4.0
Total	8.0	8.0

Denyut_Nadi		
Lemah	3.0	2.0
Normal	2.0	3.0
Kuat	3.0	3.0
Total	8.0	8.0

Mual_muntah		
Ya	5.0	6.0
Tidak	2.0	1.0
Total	7.0	7.0

Gelisah		
Ya	5.0	6.0
Tidak	2.0	1.0
Total	7.0	7.0

3.5 HASIL AKURASI



Gambar 5. Hasil Akurasi

Tampilan diatas merupakan cross-validation dimana sudah didapatkan instance yang diklasifikasikan dengan benar 80% dan instance salah diklasifikasikan 20%

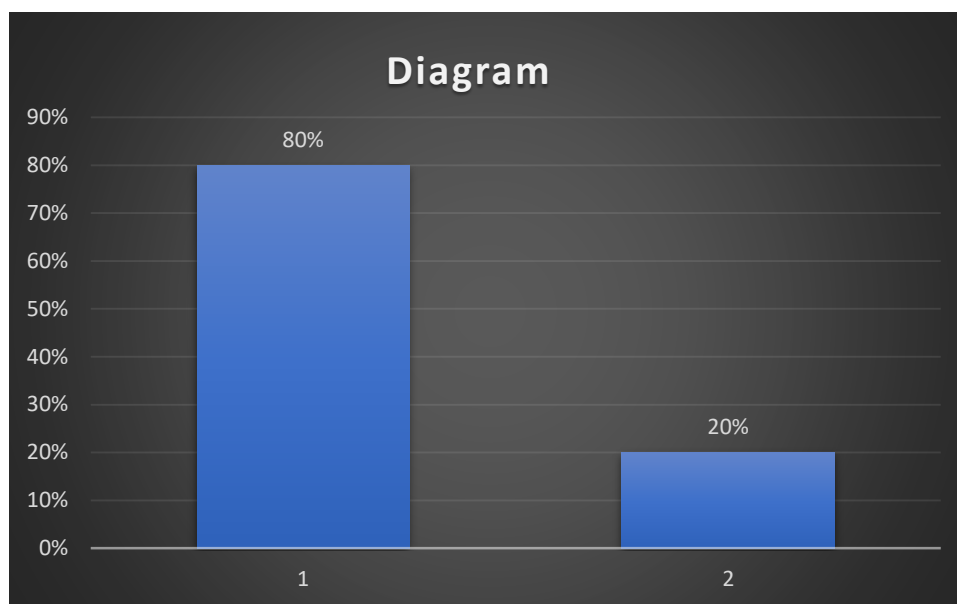
=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	8	80	%
Incorrectly Classified Instances	2	20	%
Kappa statistic	0.6		
Mean absolute error	0.3797		
Root mean squared error	0.4206		
Relative absolute error	69.6062 %		
Root relative squared error	77.1043 %		
Total Number of Instances	10		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.600	0.000	1.000	0.600	0.750	0.655	0.800	0.871	Sedang
	1.000	0.400	0.714	1.000	0.833	0.655	0.800	0.796	Parah
W Avg	0.800	0.200	0.857	0.800	0.792	0.655	0.800	0.834	



Gambar 6. Diagram

Jadi, Hasil training data dan testing pada Pasien penyakit gagal jantung dengan hasil akurasi 80% benar dan salah 20% dengan training data, serta metode dan parameter yang di pilih.

4. KESIMPULAN

Penyakit jantung merupakan salah satu dari jenis PTM yang rentan menyerang terutama pria dengan usia dibawah 60 tahun. Oleh sebab itu penelitian ini berfokus untuk menyelidiki suatu algoritma, apakah memiliki tingkat akurasi yang tinggi guna pendeteksi penyakit jantung melalui objek menggunakan dataset (*heart disease*). Setelah mengeksekusi dataset dengan algoritma yang dipilih didapatkan algoritma naïve bayes dengan hasil akurasi 80% Untuk pengembangannya, peningkatan akurasi bisa dilakukan dengan menghibridkan algoritma klasifikasi ini dengan beberapa algoritma lain, sehingga tidak menutup kemungkinan bisa menghasilkan kinerja yang berbeda dari algoritma tersebut.

REFERENCES

- [1] B. L. Kementerian, “Info Datin Jantung.”
- [2] Yuli Mardi, “Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . Jurnal Edik Informatika,” *J. Edik Inform.*, vol. 2, no. 2, hal. 213–219, 2019.
- [3] R. I. Borman dan M. Wati, “Penerapan Data Mining Dalam Klasifikasi Data Anggota Kopdit Sejahtera Bandarlampung Dengan Algoritma Naïve Bayes,” *J. Ilm. Fak. Ilmu Komput.*, vol. 9, no. 1, hal. 25–34, 2020.
- [4] F. Wulandari, P. A. Jusia, dan J. Jasmir, “Klasifikasi Data Mining Untuk Mendiagnosa Penyakit ISPA Menggunakan Metode Naïve Bayes Pada Puskesmas Jambi Selatan,” *J. Manaj. Teknol. dan Sist. Inf.*, vol. 2, no. 3, hal. 214–227, 2020, [Daring]. Tersedia pada: <http://ejournal.stikom-db.ac.id/index.php/jimsi/article/view/895>.
- [5] Y. I. Kurniawan, “Perbandingan Algoritma Naive Bayes dan C.45 dalam Klasifikasi Data Mining,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, hal. 455, 2018, doi: 10.25126/jtiik.201854803.
- [6] C. Science, “Prediksi Penyakit Jantung dengan Algoritma Klasifikasi,” vol. 5, no. 1, hal. 978–979, 2019.

- [7] Adi Nugroho, Agustinus Bimo Gumelar, Adri Gabriel Sooai, Dyana Sarvasti, dan Paul L Tahalele, "Perbandingan Performansi Kinerja Algoritma Pengklasifikasian Terpandu Untuk Kasus Penyakit Kardiovaskular," *J. RESTI (Rekayasa Sist. dan Teknol. Informatika)*, vol. 4, no. 5, hal. 998–1006, 2020, doi: 10.29207/resti.v4i5.2316.
- [8] M. Rahmadi, F. Kaurie, dan T. Susanti, "Uji Akurasi Dataset Pasien Pasca Operasi Menggunakan Algoritma Naïve Bayes Menggunakan Weka Tools," *JURIKOM (Jurnal Ris. Komputer)*, vol. 7, no. 1, hal. 134, 2020, doi: 10.30865/jurikom.v7i1.1761.
- [9] M. Rahmadi, F. Kaurie, dan T. Susanti, "Uji Akurasi Dataset Pasien Pasca Operasi Menggunakan Algoritma Naïve Bayes Menggunakan Weka Tools," *JURIKOM (Jurnal Ris. Komputer)*, vol. 7, no. 1, hal. 134, 2020, doi: 10.30865/jurikom.v7i1.1761.
- [10] R. Annisa, "Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung," *J. Tek. Inform. Kaputama*, vol. 3, no. 1, hal. 22–28, 2019.
- [11] Q. Hasanah, H. Oktavianto, dan Y. D. Rahayu, "Jurnal Smart Teknologi Analisis Algoritma Gaussian Naïve Bayes Terhadap Klasifikasi Data Pasien Penderita Gagal Jantung Gaussian Naïve Bayes Algorithm Analysis Of Data Classification Of Heart Failure Patiens Jurnal Smart Teknologi," *J. Smart Teknol.*, vol. 3, no. 4, hal. 382–389, 2022.
- [12] F. Novaldy dan A. Herliana, "Penerapan Pso Pada Naïve Bayes Untuk Prediksi Harapan Hidup Pasien Gagal Jantung," *J. Responsif Ris. Sains dan Inform.*, vol. 3, no. 1, hal. 37–43, 2021, doi: 10.51977/jti.v3i1.396.
- [13] D. Cahya Putri Buani, "Penerapan Algoritma Naïve Bayes dengan Seleksi Fitur Algoritma Genetika Untuk Prediksi Gagal Jantung," *EVOLUSI J. Sains dan Manaj.*, vol. 9, no. 2, hal. 43–48, 2021, doi: 10.31294/evolusi.v9i2.11141.
- [14] M. KATANA, A. Utami, dan K. Miraswan, "Klasifikasi Penyakit Jantung Menggunakan Metode Random Forest," vol. 1, no. 1, hal. 22–26, 2021, doi: 10.12487/AMRI.v1i1.xxxxx.
- [15] D. P. Utomo dan M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *J. Media Inform. Budidarma*, vol. 4, no. 2, hal. 437, 2020, doi: 10.30865/mib.v4i2.2080.
- [16] H. M. Nawawi, J. J. Purnama, dan A. B. Hikmah, "Komparasi Algoritma Neural Network Dan Naïve Bayes Untuk Memprediksi Penyakit Jantung," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, hal. 189–194, 2019, doi: 10.33480/pilar.v15i2.669.
- [17] R. Annisa, "Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung," *J. Tek. Inform. Kaputama*, vol. 3, no. 1, hal. 22–28, 2019.