

OPTIMASI ALGORITMA NAIVE BAYES DENGAN PARTICLE SWARM OPTIMIZATION (PSO) UNTUK KLASIFIKASI PENYAKIT JANTUNG

OPTIMIZATION OF THE NAIVE BAYES ALGORITHM WITH PARTICLE SWARM OPTIMIZATION (PSO) FOR HEART DISEASE CLASSIFICATION

Ayu' Novita Sari^{1*}, Zainal Abidin², Bijanto³

Prodi Informatika Sekolah Tinggi Teknik Pati, Indonesia^{1,2}

Prodi Teknik Elektro Sekolah Tinggi Teknik Pati, Indonesia³

e-mail: ayunovita091001@gmail.com, zainal.fsr@gmail.com, bijantokakoi@sttp.ac.id

Abstract This research aims to optimize algorithm performance Naive Bayes in the classification of heart disease using the method Particle Swarm Optimization (PSO). Heart disease is a heart disease that has a high mortality rate throughout the world. The use of technology in the health sector, such as data mining, can help in early detection and treatment of heart disease. This research uses the PSO method to optimize the parameters of the algorithm Naive Bayes, thus increasing accuracy classification of heart disease. The data used in this research has a total of 303 records consisting of 13 attributes and 1 label containing identification of heart disease. The results showed that the heart disease classification model obtained value accuracy of 96.67%, precision amounting to 93.75% as well as recall by 100%. These results indicate that the algorithm interacts Naive Bayes with PSO resulted in a significant improvement in cardiac disease classification compared with use Naive Bayes conventionally. PSO optimization is able to improve the performance of the classification model, so that it can provide more optimal classification results.

Keywords: Algorithm Naive Bayes, method Particle Swarm Optimization (PSO), Heart disease.

Abstrak Penelitian ini bertujuan untuk mengoptimalkan kinerja algoritma Naive Bayes dalam klasifikasi penyakit jantung menggunakan metode Particle Swarm Optimization (PSO). Penyakit jantung merupakan salah satu penyakit yang memiliki tingkat kematian tinggi. Penggunaan teknologi dalam bidang kesehatan, seperti data mining, dapat membantu dalam deteksi dini serta penanganan penyakit jantung. Penelitian ini menggunakan metode PSO untuk mengoptimalkan parameter pada algoritma Naive Bayes, sehingga meningkatkan accuracy klasifikasi penyakit jantung. Data yang digunakan dalam penelitian ini memiliki jumlah 303 record terdiri dari 13 atribut serta 1 label yang berisi identifikasi penyakit jantung. Hasil penelitian menunjukkan bahwa model klasifikasi penyakit jantung didapatkan nilai accuracy sebesar 96,67%, precision sebesar 93,75% serta recall sebesar 100%. Hasil ini menunjukkan peningkatan yang signifikan dibandingkan dengan implementasi algoritma Naive Bayes tanpa di optimasi Particle Swarm Optimization (PSO). Optimasi Particle Swarm Optimization (PSO) mampu meningkatkan kinerja model klasifikasi, sehingga dapat memberikan hasil klasifikasi yang lebih optimal. Potensi menggunakan algoritma Naive Bayes dengan optimasi Particle Swarm Optimization (PSO) dalam klasifikasi penyakit jantung, yang dapat menjadi kontribusi penting dalam pengembangan sistem pendukung keputusan.

Kata Kunci: Algoritma Naive Bayes, metode Particle Swarm Optimization (PSO), Penyakit jantung.

PENDAHULUAN

Kesehatan ialah bagian penting untuk manusia karena dengan sehat kita dapat melakukan berbagai kegiatan dan berfikir baik (J. Dian, 2021). Jantung ialah salah satu organ kesehatan yang harus kita jaga, jantung memiliki fungsi vital pada tubuh manusia jika terdapat kelainan kecil pada jantung maka akan berpengaruh besar pada tubuh manusia (J. Dian, 2021). Penyakit jantung serta pembuluh darah ialah penyebab utama kematian di negara berkembang (B. Prajapati, 2022). Penyakit jantung merupakan salah satu jenis penyakit yang memiliki tingkat kematian tinggi dan memerlukan penanganan segera karena seringkali muncul secara tiba-tiba pada penderitanya (H. M. Nawawi, 2019). Penyakit atau kendala jantung sendiri memiliki banyak tipe serta berbagai macam nama penyakitnya seperti kardiovaskuler, jantung koroner serta serangan jantung (Mesran, 2020). Bersumber pada informasi statistik dunia, terdapat 9,4 juta kematian tiap tahun yang diakibatkan oleh penyakit kardiovaskuler. Menurut organisasi kesehatan dunia WHO 45% kematian yang diakibatkan oleh penyakit jantung, angka tersebut akan bertambah sampai 23,3 juta pada 2030 mendatang (N. Rahmad, 2024). Jumlah Penyakit Jantung berdasarkan diagnosis dokter di Indonesia sebesar 1,5%, dengan peringkat jumlah tertinggi di setiap Provinsi yang ada di Indonesia, hasil Risesdas tahun 2018 juga menunjukkan jumlah penyakit jantung berdasarkan diagnosis dokter pada kelompok usia 35-44 tahun sebanyak (1,3%), usia 45-54 tahun sebanyak (2,4%), serta usia 55-64 tahun sebanyak (3,9%) (N. Rahmad, 2024). Meningkatnya angka kematian yang disebabkan oleh penyakit jantung, maka perlu mengetahui faktor apa saja yang mempengaruhi adanya resiko penyakit jantung agar kedepannya penyakit ini bisa ditangani sejak dini dengan langkah cepat dan tepat.

Kemajuan teknologi sudah membuat pertumbuhan dalam ilmu pengetahuan serta teknologi data sudah sangat pesat apalagi memasuki banyak bidang serta salah satu manfaatnya merupakan terciptanya sistem klasifikasi (Septiani, 2017) (Lasena, 2020). Dunia kedokteran salah satu yang menemukan manfaat pertumbuhan teknologi dalam pelaksanaan sistem klasifikasi (Budi, 2018). Banyak metode klasifikasi telah digunakan untuk membuat prediksi penyakit yang akurat (M. B. Priyanto, 2020). Pertumbuhan teknologi yang sangat pesat, memudahkan untuk menentukan penyakit tersebut membutuhkan data mining.

Data mining ialah bidang yang berpotensi menjadi solusi pendeteksi penyakit jantung sejak dini. Data mining dapat didefinisikan sebagai kategori luas pembelajaran berbasis pengetahuan yang mencakup sistem data, statistik, pembelajaran mesin, visualisasi, dan informasi berbasis pengetahuan (R. Hasan, 2020). Data mining memiliki beberapa metode antara lain *clustering*, asosiasi, prediksi serta klasifikasi (Pristyanto Y., 2019). Teknik klasifikasi beroperasi dengan mengenali pola atau model dari sebuah dataset, khususnya dataset penyakit jantung yang sudah ada di *repository*. Tujuannya menggunakan model ini untuk dapat menentukan atau mengklasifikasikan apakah seseorang terkena penyakit jantung atau tidak (Pristyanto A. N., 2021). Teknik klasifikasi memudahkan untuk menentukan adanya penyakit jantung pada tubuh manusia atau tidak, sedangkan teknik klasifikasi sendiri memiliki beberapa algoritma di dalamnya.

Klasifikasi merupakan proses pengelompokan objek ke dalam kelas-kelas berdasarkan ciri-ciri yang dimilikinya, yang dikenal sebagai kelas (*class*) (H. F. Putro, 2020). Salah satu metode klasifikasi yang umum digunakan merupakan metode *Naive Bayes*, penelitian sebelumnya telah menunjukkan bahwa metode *Naive Bayes* memiliki beberapa keunggulan, seperti kecepatan komputasi yang tinggi, algoritma yang sederhana, dan tingkat akurasi yang tinggi (H. F. Putro, 2020). Penelitian analisis penyakit jantung, maka pengolahan data akan dilakukan dengan mengklasifikasikan atribut menjadi dua kelas, yaitu "YA" dan "TIDAK", sehingga teknik klasifikasi paling tepat digunakan dalam penelitian ini.

Beberapa penelitian mengenai klasifikasi penyakit jantung telah banyak dilakukan antara lain, penelitian yang dilakukan oleh Donny Maulana dan Rezaldi Yahya pada tahun 2019 berjudul Implementasi Algoritma *Naive Bayes* Untuk Klasifikasi Penderita Penyakit Jantung Di Indonesia Menggunakan Rapid Miner. Metode klasifikasi dengan penerapan algoritma *Naive Bayes* bisa digunakan untuk melaksanakan prediksi terdapat penyakit jantung ataupun tidak. Hasil dari penelitian data menyatakan tingkat *Accuracy*, *Precision* serta *Recall* yang telah diputuskan dalam klasifikasi *Naive Bayes* (N. Gligorijevic, 2019). Penelitian lain yang dilakukan oleh Arni Sepharni,

Irwansyah Edo Hendrawan, dan Chaerur Rozikin dengan objek yang sama, yaitu klasifikasi penyakit jantung, namun dengan menerapkan metode yang berbeda, yaitu Algoritma C.45. Dalam penelitian ini, Algoritma C.45 digunakan untuk melakukan klasifikasi dan prediksi berdasarkan data historis pasien yang akan diperiksa. Metode ini memungkinkan penentuan prediksi penyakit jantung berdasarkan aturan yang dihasilkan dari analisis data pasien (A. Sepharni, 2022). Beberapa penelitian yang sebelumnya telah dilakukan terdapat perbedaan dalam penggunaan metode dan data set dari setiap peneliti. Masih belum adanya optimasi yang dilakukan untuk meningkatkan kinerja algoritma, sehingga penelitian ini akan menggunakan algoritma *Naive Bayes* sebagai algoritma utama yang akan di optimasi dengan menggunakan PSO. Algoritma *Naive Bayes* ialah algoritma salah satu metode yang digunakan untuk mengidentifikasi dan memprediksi penyakit (Z. Maisat, 2023). Dengan diimplementasikannya optimasi dari algoritma tersebut akan mendapatkan hasil yang lebih tinggi dari peneliti sebelumnya.

Berdasarkan latar belakang masalah yang sudah disampaikan diatas, maka penelitian dengan judul "optimasi Algoritma *Naive Bayes* dengan *Particle Swarm Optimization* (PSO) untuk Klasifikasi Penyakit Jantung" untuk mengetahui kinerja dari *Naive Bayes* yang di optimasi dengan Optimasi *Particle Swarm Optimization* (PSO) dalam klasifikasi penyakit jantung. Untuk mengetahui hasil dari nilai *Accuracy*, *Precision*, dan *Recall*.

METODE PENELITIAN

Dataset yang digunakan penelitian ini menggunakan data penyakit jantung yang berjudul *Heart Failure Prediction* yang berasal dari *website* data internasional yaitu *website The Kaggle Learning Repository*, data yang digunakan bersifat publik data ini sudah banyak digunakan oleh beberapa peneliti di seluruh dunia. *Keggel* ialah *platform* yang menampung arsip data internasional untuk penelitian. *Dataset* yang digunakan dalam penelitian ini terdiri dari 13 atribut yang bersifat independen serta 1 label yang bersifat dependen, dengan jumlah 303 *records*. Metode yang akan diterapkan dalam penelitian ini yaitu Algoritma *Naive Bayes* yang dioptimasi menggunakan *Particle Swarm Optimization* (PSO).

A. Optimasi Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) merupakan sebuah algoritma optimisasi yang diperkenalkan pada tahun 1995 oleh Dr. James Kennedy dan Dr. Russell C Eberhart. Algoritma ini terinspirasi oleh pola perilaku populasi burung dan lebah. Salah satu keunggulan utama dari metode ini merupakan kesederhanaan konsepnya, kemudahan implementasinya dan efisiensi dalam perhitungan, jika dibandingkan dengan berbagai algoritma matematika dan teknik *heuristik* lainnya (A. Herliana, 2019). Setiap partikel diperlakukan seperti sebuah titik pada suatu dimensi ruang tertentu. Kemudian terdapat dua faktor yang memberikan karakter terhadap status partikel pada ruangan pencarian yaitu posisi partikel dan kecepatan partikel.

Berikut ini merupakan formula matematika yang menggambarkan posisi dan kecepatan partikel pada dimensi ruang tertentu, Berikut ini merupakan model matematika yang menggambarkan mekanisme updating status partikel (A. Herliana, 2019). Berikut ini rumus yang menggambarkan updating status partikel menurut Kennedy and Eberhart [1995]:

$$V_i(t) = V_i(t-1) + c_1r_1 (X_i^L - X_i(t-1)) + c_2r_2 (X_i^G - X_i(t-1)) \quad (1)$$

$$X_i(t) = V_i(t) + X_i(t-1) \quad (2)$$

Dimana :

$X_i(t)$ = posisi partikel i saat iterasi t

$V_i(t)$ = kecepatan partikel i saat iterasi t

i = indeks partikel

t = iterasi ke- t

N = ukuran dimensi ruang

c_1 c_2 = Suatu konstanta yang bernilai positif yang biasanya disebut dengan

learning factor

- r_1, r_2 = Suatu bilangan random yang bernilai antara 0 sampai 1
- $X_{pbest\ i}$ = Posisi terbaik partikel i
- $X_{gbest\ t}$ = Posisi terbaik global
- X_i^L = $X_{i1}^L, X_{i2}^L, \dots, X_{iN}^L$ Mempresentasikan local best dari partikel ke- i
- X^G = $X_{i1}^G, X_{i2}^G, \dots, X_{iN}^G$ Mempersentasikan global best dari seluruh kawan

B. Klasifikasi menggunakan *Naive Bayes*

Klasifikasi Bayesian juga dikenal sebagai klasifikasi Naive Bayes merupakan metode klasifikasi statistik yang mengandalkan teorema bayes untuk memprediksi probabilitas suatu objek masuk ke dalam kelas tertentu. Ketika diterapkan pada basis data besar, klasifikasi bayesian telah terbukti memiliki tingkat akurasi dan kecepatan yang tinggi[30][37]. Persamaan dari teorema bayes sebagai berikut:

$$P(H|X) = \frac{P(H)P(X|H)}{P(X)} \quad (3)$$

Keterangan :

- X = Data dengan *class* yang belum diketahui
- H = *Hipotesis* Data X merupakan suatu *class spesifik*
- $P(H|X)$ = *Probabilitas hipotesis* H berdasar kondisi X (*posterior probability*)
- $P(H)$ = *Probabilitas hipotesis* H (*prior probability*)
- $P(X|H)$ = *Probabilitas hipotesis* H berdasar kondisi pada *hipotesis* H
- $P(X)$ = *Probabilitas* dari X

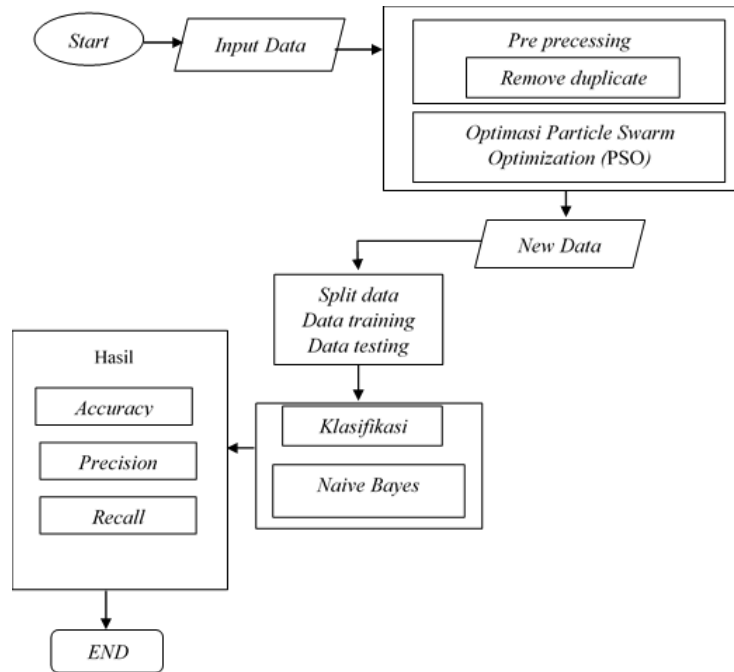
Naive bayes merupakan penyederhanaan metode *bayes*. *Teorema bayes* setelah disederhanakan menjadi:

$$P(H|X) = \frac{P(H)P(X|H)}{P(X)} \quad (4)$$

Dalam *bayes rule* diterapkan untuk menghitung *probabilitas* akhir atau klasifikasi diperoleh dengan menggunakan dua sumber informasi yaitu informasi awal (*prior*) dan informasi yang diperbarui (*posterior*) menggunakan aturan *bayes* untuk menghitung *bayes*.

C. Tahapan Penelitian

Tahapan pada penelitian ini disajikan dalam bentuk *flowchart* alur perancangan sistem yang ditujukan pada [gambar 1](#), sebagai berikut:



(Sumber: Peneliti)

HASIL DAN PEMBAHASAN

1. Persiapan Data

Pada langkah persiapan data ini, *dataset* yang digunakan penelitian ini menggunakan data berjudul “*Heart Failure Prediction*” yang berasal dari Kaggel. Data tersebut terdiri dari 303 *records*, dengan 13 *atribut* serta 1 label. Mengacu pada data tersebut, *atribut output* pada data tersebut mempunyai 2 nilai yaitu 0 dan 1. Data yang digunakan dalam penelitian ini dapat dilihat pada [tabel 4.1](#) sebagai berikut:

Tabel 1. Sampel data yang digunakan penelitian

id	age	sex	cp	trebps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
1	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
2	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
3	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
...														
14	64	1	3	110	211	0	0	144	1	1.8	1	0	2	1
15	58	0	3	150	283	1	0	162	0	1	2	0	2	1
16	50	0	2	120	219	0	1	158	0	1.6	1	0	2	1
...														
123	41	0	2	112	268	0	0	172	1	0	2	0	2	1
124	54	0	2	108	267	0	0	167	0	0	2	0	2	1
125	39	0	2	94	199	0	1	179	0	0	2	0	2	1
...														
200	65	1	0	110	248	0	0	158	0	0.6	2	2	1	0
201	44	1	0	110	197	0	0	177	0	0	2	1	2	0
202	60	1	0	125	258	0	0	141	1	2.8	1	1	3	0
...														
300	45	1	3	110	264	0	1	132	0	1.2	1	0	3	0
302	57	1	0	130	131	0	1	115	1	1.2	1	1	3	0
303	57	0	1	130	236	0	0	174	0	0	1	1	2	0

(Sumber: Kaggel 2021)

Berdasarkan [tabel 4.1](#) dapat dijelaskan bahwa atribut “*output*” merupakan variabel target dimana terdapat 2 variabel target yaitu 0 dan 1. Kategori 0 terdapat sebanyak 138 data dan kategori 1 terdapat sebanyak 165 data. Berdasarkan pembagian dari keseluruhan kategori tersebut, maka terdapat total keseluruhan data sebanyak 303 data. Pada atribut “*id*” digunakan sebagai *index* atau penomoran data, untuk atribut yang lain data dijelaskan pada [tabel 4.2](#) sebagai berikut (A. Samosir, 2021):

Tabel 2. Penjelasan atribut data

No	Atribut	Keterangan
1	Age	: Umur pasien
2	sex	: Jenis kelamin pasien, atribut ini memiliki 2 nilai, yakni nilai 1 untuk laki-laki dan nilai 0 untuk perempuan
3	Cp	: Tipe nyeri dada yang diderita pasien. Atribut ini memiliki 4 nilai, yaitu : Nilai 0: asymptomatic Nilai 1: atypical angina Nilai 2: non anginal pain Nilai 3: typical angina
4	Trestbps	: Resting blood pressure yaitu tekanan darah pasien ketika dalam keadaan istirahat. Satuan yang dipakai adalah mm Hg.
5	Chol	: Kolesterol yaitu kadar kolesterol dalam darah pasien, dengan satuan mg/dl.
6	Fbs	: Fasting blood sugar yaitu kadar gula darah pasien, atribut fbs ini hanya memiliki 2 nilai yaitu 1 jika kadar gula darah pasien lebih dari 120 mg/dl, dan 0 jika kadar gula darah pasien kurang dari sama dengan 120 mg/dl
7	Restecg	: Resting electrocardiographic yaitu kondisi ECG pasien ketika dalam keadaan istirahat. Atribut ini memiliki 3 nilai yaitu nilai 1 untuk keadaan normal, nilai 2 untuk keadaan ST-T wave abnormality yaitu keadaan dimana gelombang inversions T dan atau ST meningkat maupun menurun lebih dari 0,5 mV dan nilai 3 untuk keadaan dimana ventricular kiri mengalami hipertropi
8	Thalach	: Rata-rata detak jantung pasien dalam satu menit
9	Exang	: Keadaan dimana pasien akan mengalami nyeri dada apabila berolah raga, 0 jika tidak nyeri, dan 1 jika menyebabkan nyeri
10	Oldpeak	: Penurunan ST akibat olahraga.
11	Slope	: Slope dari puncak ST setelah berolah raga. Atribut ini memiliki 3 nilai yaitu 0 untuk downsloping, 1 untuk flat, dan 2 untuk upsloping
12	Ca	: Banyaknya pembuluh darah yang terdeteksi melalui proses pewarnaan flourosopy
13	Thal	: Detak jantung pasien. Atribut ini memiliki 3 nilai yaitu 1 untuk fixed defect, 2 untuk normal dan 3 untuk reversal defect.

(Sumber: Kaggel 2021)

Data ini merupakan data publik yang dapat diakses oleh siapa saja, dari data diatas setiap atribut sudah di jelaskan dengan keterangan yang jelas agar memudahkan dalam proses penelitian nantinya.

Penelitian ini peneliti dapat menjadikan sebagai objek penelitian untuk klasifikasi penyakit jantung. Peneliti sudah menjelaskan lengkap darimana data yang peneliti dapatkan dan jumlah atribut dan atribut *output* sebagai label yang mempunyai 2 nilai yaitu 0 untuk terdeteksi negatif terkena penyakit jantung sedangkan 1 untuk positif terkena penyakit jantung.

Banyaknya algoritma yang ada tetapi peneliti menggunakan *Naive Bayes* memiliki keunggulan seperti kecepatan komputasi yang tinggi, algoritma yang sederhana dan akurasi yang

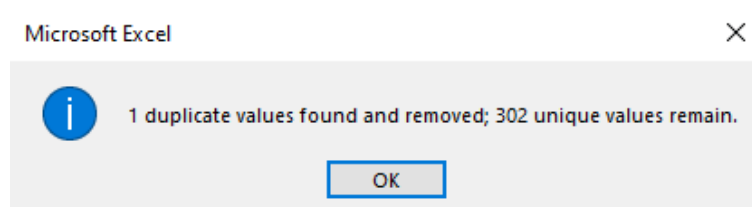
tinggi (H. F. Putro, 2020). Menggunakan *Naive Bayes* memiliki keuntungan yaitu membutuhkan jumlah data pelatihan yang relatif kecil untuk mengestimasi parameter yang diperlukan dalam proses pengklasifikasian (E. Manalu, 2017). Situasi dunia nyata yang kompleks, *Naive Bayes* seringkali memberikan hasil yang lebih baik dari pada yang diharapkan.

2. Preprocessing

Sebelum data diolah menggunakan algoritma, data perlu melalui tahap preprocessing terlebih dahulu. Penelitian ini, pada tahapan *preprocessing* meliputi *remove duplicate*. Berikut merupakan penjabaran dari proses *preprocessing* data yang meliputi *remove duplicate*:

a. Remove duplicates

Pada tahap *remove duplicates* data akan diperiksa secara menyeluruh dan jika terdapat data yang ganda, maka data ganda tersebut akan dihapus. Apabila tidak ada data yang ganda, maka keseluruhan data yang didapatkan oleh peneliti dapat diolah tanpa menghilangkan salah satu *record*. Penelitian ini dalam memeriksa data yang ganda menggunakan bantuan *microsoft excel* dapat dilihat pada [gambar 1](#) sebagai berikut:



Gambar 1: *Removed duplicate*

Selanjutnya yaitu proses dari *remove duplicates* menggunakan bantuan *microsoft excel*. [Gambar 4.1](#) terdapat hasil "1 duplicate values found and removed; 302 unique values remain" yang berarti terdapat 1 nilai duplikat ditemukan dengan ada data yang duplikat, data duplikat tersebut akan dihapus supaya dalam pengolahan data mendapatkan hasil yang lebih maksimal saat proses klasifikasi.

3. Split Data

Split data merupakan proses pembagian data menjadi dua bagian yaitu data *training* dan data *testing*. Rasio dalam menentukan data *training* dan data *testing* dalam penelitian ini menggunakan rasio 90%:10% karena dianggap sebagai rasio paling baik untuk pelatihan (*training*) dan data sampling (*testing*) (Adinugroho, 2022) (L. Efrizoni, 2022)t.

Split data dalam penelitian ini bertujuan untuk meningkatkan efisiensi sistem karena sistem tidak perlu mengolah keseluruhan data, akan tetapi sistem mengolah data sampling (data *testing*). Penentuan data *testing* dalam penelitian ini ditentukan secara acak (random) sejumlah 10% dari total keseluruhan *record* data yaitu 30 data dari total 302 data.

4. Klasifikasi menggunakan Algoritma *Naive Bayes*

Tahap klasifikasi menggunakan algoritma *Naive Bayes* dalam penelitian ini menggunakan tahap terakhir dan tahap utama dalam mencari akurasi data, *precision* dan *recall*. Proses perhitungan *Naive Bayes* merupakan mode klasifikasi statistik yang mengandalkan *teorema bayes* untuk memprediksi *probabilitas* suatu objek masuk ke dalam kelas tertentu.

Tabel 3. *Sampel data testing*

id	age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
1	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
9	52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
13	49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
15	58	0	3	150	21	1	0	162	0	1	2	0	2	1
18	66	0	3	150	226	0	1	114	0	2.6	0	0	2	1

id	age	sex	cp	trtbps	chol	fbs	restecg	lhalachh	exng	oldpeak	slp	caa	thall	output
143	42	0	2	120	209	0	1	173	0	0	1	0	2	1
144	67	0	0	106	223	0	1	142	0	0.3	2	2	2	1
145	76	0	2	140	197	0	2	116	0	1.1	1	0	2	1
146	70	1	1	156	245	0	0	143	0	0	2	0	2	1
147	44	0	2	118	242	0	1	149	0	0.3	1	1	2	1
....														
295	63	1	0	140	187	0	0	144	1	4	2	2	3	0
296	63	0	0	124	197	0	1	136	1	0	1	0	2	0
297	59	1	0	164	176	1	0	90	0	1	1	2	1	0
298	57	0	0	140	241	0	1	123	1	0.2	1	0	3	0
302	57	0	1	130	236	0	0	174	0	0	1	1	2	0

(Sumber: Peneliti)

Pada [tabel 3](#) merupakan sampel data yang akan digunakan untuk perhitungan *Naive Bayes*, tetapi sebelum masuk ketahap perhitungan *Naive Bayes* data tersebut akan di konversi dahulu agar memudahkan proses perhitungan. Berdasarkan hasil dari sampel perhitungan perbandingan antara probabilitas ya dan tidak pada id ke 1 didapatkan bahwa nilai terbesar yaitu terdapat pada klasifikasi “Ya”, oleh karena itu id ke 1 hasil klasifikasi final menggunakan *Naive Bayes* dikategorikan sebagai “ 1”. Apabila dibandingkan dengan data awal sebelum proses klasifikasi menggunakan *Naive Bayes*, hasil klasifikasi menggunakan *Naive Bayes* diklasifikasi tepat seperti pada data awal. Artinya, id ke 1 termasuk dalam salah satu data yang akan digunakan sebagai acuan dalam menghitung nilai dari akurasi data.

5. Optimasi Algoritma *Naive Bayes* menggunakan PSO

Hasil penelitian ini berupa nilai Accuracy, Precision dan recall yang digunakan untuk menentukan tingkat klasifikasi. Hasil penelitian ini ditunjukkan pada tabel berikut:

Tabel 4. Hasil penelitian

Parameter	Hasil
Accuracy	95,77%
Precision	93,75%
Recall	100%

Berdasarkan [tabel 4](#), peneliti dalam hal ini berhasil melakukan peningkatan optimasi pada algoritma *Naive Bayes* dengan *Particle Swarm Optimization* (PSO) untuk klasifikasi penyakit jantung, didapatkan nilai *accuracy* sebesar **96,67%**, *precision* sebesar **93,75%** serta *Recall* sebesar **100%**.

KESIMPULAN

Berdasarkan penelitian yang dilakukan mengenai optimasi algoritma *Naive Bayes* dengan optimasi *Particle Swarm Optimization* (PSO) untuk klasifikasi penyakit jantung, maka kesimpulan yang dapat diambil sebagai berikut:

1. Penelitian ini menerapkan teknik optimasi *Particle Swarm Optimization* (PSO) pada algoritma *Naive Bayes* menghasilkan tingkat akurasi yang lebih baik dibandingkan menggunakan dengan menggunakan algoritma individual pada penelitian ini menggunakan algoritma *Naive Bayes* dengan optimasi *Particle Swarm Optimization* (PSO) untuk meningkatkan nilai hasil klasifikasi. Cara kerja *Particle Swarm Optimization* (PSO) dapat meningkatkan performa klasifikasi *Naive Bayes* agar lebih akurat dengan cara mengoptimalkan aspek tertentu seperti bobot fitur, *feature selection* maupun *hyperparameter*. Tujuan utama yaitu untuk meningkatkan kinerja klasifikasi dari model *Naive Bayes* dengan mengoptimalkan parameter-parameter yang relevan.

2. Penelitian ini berhasil mencapai tujuannya dengan menerapkan metode *Particle Swarm Optimization* (PSO) pada algoritma Naive Bayes untuk meningkatkan akurasi dalam klasifikasi penyakit jantung. Hasilnya menunjukkan peningkatan signifikan, dengan *accuracy* mencapai 96,67%, *precision* sebesar 93,75% dan *recall* sebesar 100%.

DAFTAR PUSTAKA

- A. Herliana, T. A. (2019). Feature Selection of Diabetic Retinopathy Disease Using Particle Swarm Optimization and Neural Network. *2018 6th Int. Conf. Cyber IT Serv. Manag. CITSM 2018*, 2016-2019.
- A. Samosir, M. S. (2021). Komparasi Algoritma Random Forest, Naïve Bayes dan K-Nearest Neighbor Dalam klasifikasi Data Penyakit Jantung. *Pros. Semin. Nas. Darmajaya*, 214-222.
- A. Sepharni, I. E. (2022). Klasifikasi Penyakit Jantung dengan Menggunakan Algoritma C4.5. *STRING (Satuan Tulisan Ris. dan Inov. Teknol)*, 117.
- Adinugroho, R. (2022). Perbandingan Rasio Split Data Training Dan Data Testing Menggunakan Metode Lstm Dalam Memprediksi Harga Indeks Saham Asia.
- B. Prajapati, S. P. (2022). An Implementation of IRTBS—for Observing Cardiac Patients. *in Smart Innovation, Systems and Technologies*.
- Budi, T. P. (2018). Sistem Prediksi Penyakit Kanker Serviks Menggunakan CART, Naive Bayes, dan k-NN. *Creat. Inf. Technol. J.*, 83.
- E. Manalu, F. A. (2017). Penerapan Algoritma Naive Bayes Untuk Memprediksi Jumlah Produksi Barang Berdasarkan Data Persediaan dan Jumlah Pemesanan Pada CV. Papadan Mama Pastries. *J. Mantik Penusa*, 16-21.
- H. F. Putro, R. T. (2020). Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan. *J. Teknol. Inf. dan Komun.*
- H. M. Nawawi, J. J. (2019). Komparasi Algoritma Neural Network Dan Naïve Bayes Untuk Memprediksi Penyakit Jantung. *J. Pilar Nusa Mandiri*, 189-194.
- J. Dian, F. D. (2021). Sistem Monitoring Detak Jantung Untuk Mendeteksi Tingkat Kesehatan Jantung Berbasis Internet Of Things Menggunakan Android. *JUPITER(Jurnal Penelit. Ilmu dan Teknol. Komputer)*, 69-75.
- L. Efrizoni, S. D. (2022). Komparasi Ekstraksi Fitur dalam Klasifikasi Teks Multilabel Menggunakan Algoritma Machine Learning. *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, 653-666.
- Lasena, H. H. (2020). Real Time Analisis Berbasis Internet Of Things Untuk Prediksi Iklim Lahan Pertanian. *J. Media Inform. Budidarma*, 834-840.
- M. B. Priyantono, A. A. (2020). Sistem Prediksi Gejala Virus Korona dengan Metode Forward Chaining. *JTERA (Jurnal Teknol. Rekayasa)*, 111.
- Mesran, D. P. (2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. *J. Media Inform. Budidarma*, 437.
- N. Gligorijevic, D. R. (2019). Повышенная Чувствительность Тромбоцитов К Действию Инсулиноподобного Фактора Роста 1 У Больных Сахарным Диабетом 2-Го Типа. *Биохимия*, 1511-1518.
- N. Rahmad, A. D. (2024). Pengaruh Latihan Aktifitas Rehabilitasi Jantung Fase II Terhadap Tingkat Kebugaran dan Daya Tahan Pasien Penyakit Jantung Koroner. *J. Penelit. Kesehat. Suara Forikes*, 1057-1062.
- Pristyanto, A. N. (2021). Algoritme Stacking Untuk Klasifikasi Penyakit Jantung Pada Dataset Imbalanced Class. *Pseudocode*, 21-26.
- Pristyanto, Y. (2019). Penerapan Metode Ensemble Untuk Meningkatkan Kinerja Algoritme Klasifikasi Pada Imbalanced Dataset. *J. Teknoinfo*, 11.
- R. Hasan, S. P. (2020). R. HModelling and predicting student's academic performance

- using classification data mining techniques. *Int. J. Bus. Inf. Syst.*, 403-422.
- Septiani, W. D. (2017). Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis. *None*, 76-84.
- Z. Maisat, E. D. (2023). Implementasi Optimasi Hyperparameter GridSearchCV Pada Sistem Prediksi Serangan Jantung Menggunakan SVM Implementation of GridSearchCV Hyperparameter Optimization in Heart Attack Prediction System Using SVM. 8-15.