



Perbandingan Algoritma *Naive Bayes* Berbasis *Feature Selection Gain Ratio* dengan *Naive Bayes Konvensional* dalam Prediksi Komplikasi Hipertensi

I Made Arya Adinata Dwija Putra¹, I Made Gede Sunarya² dan I Gede Aris Gunadi³

1. Universitas Pendidikan Ganesha; arya.adinata@undiksha.ac.id
 2. Universitas Pendidikan Ganesha; sunarya@undiksha.ac.id
 3. Universitas Pendidikan Ganesha; igedearisgunadi@undiksha.ac.id
- * Korespondensi : igedearisgunadi@undiksha.ac.id

Sitasi: Putra, I. M. A. A. D.; Sunarya, I. M. G.; Gunadi, I. G. A. (2024). Perbandingan Algoritma *Naive Bayes* Berbasis *Feature Selection Gain Rasio* dengan *Naive Bayes Konvensional* dalam Memprediksi Komplikasi Hipertensi. *JTIM: Jurnal Teknologi Informasi Dan Multimedia*, 6(1), 37-49. <https://doi.org/10.35746/jtim.v6i1.488>

Diterima: 01-02-2024

Direvisi: 20-04-2024

Disetujui: 26-04-2024



Copyright: © 2024 oleh para penulis. Karya ini dilisensikan di bawah Creative Commons Attribution-ShareAlike 4.0 International License. (<https://creativecommons.org/licenses/by-sa/4.0/>).

Abstract: High blood pressure is a significant public health problem with a high prevalence in the Indonesian population. 2018 Riset Kesehatan Dasar (Riskesdas) data shows a prevalence of hypertension of 34.1% in individuals aged 18 years and over, with the highest figure in South Kalimantan and the lowest in Papua. Complications arising from hypertension can have serious impacts on organs such as the brain, eyes, heart and kidneys. The *Naive Bayes* algorithm is generally used in disease prediction, but the *Naive Bayes* algorithm has problems when selecting attributes, because *Naive Bayes* itself is a statistical classification method that is only based on Bayes' Theorem so it can only be used with the aim of predicting the probability of membership in a group or class. So attribute weighting is needed to increase accuracy more effectively. This research introduces *Gain Ratio* as an attribute weighting method to increase the accuracy of *Naive Bayes*. The aim of this study was to compare conventional *Naive Bayes* with *Naive Bayes Gain Ratio* in predicting complications of high blood pressure. The research results show that feature selection with gain ratio weighting can increase the accuracy of naive Bayes classification, with an average increase in accuracy of 20% compared to naive Bayes without feature selection. The precision value increased by 21% in the naive Bayes gain ratio algorithm for the kidney failure class, an increase of 3% in the heart class, and an increase of 31% in the stroke class, for the recall value the naive Bayes gain ratio increased by 35% in the heart class while in the kidney failure and stroke classes did not increase the recall value.

Keywords: naive bayes, gain ratio, high blood pressure, high blood pressure complications, data mining

Abstrak: Tekanan darah tinggi merupakan masalah kesehatan masyarakat yang signifikan dengan prevalensi yang tinggi pada penduduk Indonesia. Data Riset Kesehatan Dasar (Riskesdas) tahun 2018 menunjukkan prevalensi hipertensi sebesar 34,1% pada individu berusia 18 tahun ke atas, dengan angka tertinggi di Kalimantan Selatan dan terendah di Papua. Komplikasi yang timbul akibat hipertensi dapat berdampak serius pada organ tubuh seperti otak, mata, jantung, dan ginjal. Algoritma *naive bayes* pada umumnya digunakan dalam prediksi penyakit, namun algoritma *naive bayes* mempunyai kendala pada saat pemilihan atribut, karena *naive bayes* sendiri merupakan metode klasifikasi statistik berdasarkan *teorema bayes* sehingga hanya dapat digunakan dengan tujuan untuk memprediksi kemungkinan keanggotaan dalam suatu kelompok atau kelas. Maka diperlukan pembobotan atribut agar dapat meningkatkan akurasi dengan lebih efektif. Penelitian ini memperkenalkan *gain ratio* sebagai metode pembobotan atribut untuk meningkatkan akurasi *naive bayes*. Tujuan penelitian ini adalah membandingkan *naive bayes* konvensional dengan *naive*

bayes gain ratio dalam memprediksi komplikasi tekanan darah tinggi. Hasil penelitian ini menunjukkan bahwa *feature selection* dengan pembobotan *gain ratio* dapat meningkatkan akurasi klasifikasi *naive bayes*, dengan rata-rata peningkatan akurasi sebesar 20% dibandingkan dengan *naive bayes* tanpa *feature selection*. Nilai *precision* meningkat sebesar 21% pada algoritma *naive bayes gain ratio* untuk kelas gagal ginjal, peningkatan sebesar 3% pada kelas jantung, dan peningkatan sebesar 31% pada kelas stroke, untuk nilai *recall naive bayes gain ratio* meningkat sebesar 35% pada kelas jantung sedangkan pada kelas gagal ginjal dan stroke tidak terjadi peningkatan nilai *recall*.

Kata kunci: ; naive bayes, gain ratio, tekanan darah tinggi, komplikasi tekanan darah tinggi, data mining

1. Pendahuluan

Penyakit tekanan darah tinggi atau yang biasa disebut dengan hipertensi merupakan salah satu penyakit yang umum diderita oleh masyarakat di Indonesia. Mengacu pada data Riset Kesehatan Dasar (Riskesdas) milik Kementerian Kesehatan Republik Indonesia tahun 2018 angka kejadian penyakit hipertensi berdasarkan hasil pengukuran pada penduduk usia 18 tahun sebesar 34,1%, kasus hipertensi tertinggi terjadi di Kalimantan Selatan sebesar 44,1%, sedangkan kasus terendah terjadi di Papua sebesar 22,2%. Hipertensi umumnya terjadi pada kelompok usia 31 hingga 44 tahun dengan kasus sebanyak 31,6%, usia 45 hingga 54 tahun sebesar 45,3%, dan usia 55 hingga 64 tahun sebesar 55,2% [1].

Penyakit tekanan darah tinggi yang tidak dilakukan pencegahan atau penanganan sedini mungkin dapat menyebabkan penyakit komplikasi yang lebih buruk, organ vital pada tubuh yang berpotensi mengalami komplikasi dari penyakit hipertensi antara lain otak, mata, jantung, ginjal, hingga dapat menyebabkan pecahnya pembuluh darah arteri perifer, kerusakan organ tubuh akibat komplikasi dari penyakit hipertensi bergantung kepada seberapa besar peningkatan tekanan darah dan durasi kondisi tekanan darah yang tidak terdeteksi dan tidak ditangani dengan tepat [2].

Mengacu pada penelitian yang dilakukan oleh Arya Adinata (2019) menunjukkan bahwa 72,9% responden dalam penelitian tersebut mengungkapkan bahwa mereka tidak memahami cara melakukan deteksi dini atau pencegahan penyakit hipertensi hingga komplikasinya, sehingga dibutuhkan teknologi untuk mengatasi masalah tersebut [3]. Aplikasi kecerdasan buatan yang dapat menggunakan metode data *mining*, sistem pakar, atau metode sistem pendukung keputusan dapat digunakan untuk membantu mendiagnosis suatu penyakit di dalam bidang kedokteran [4]. Teknologi dapat membantu pelaksanaan *surveilans* epidemiologi atau pengamatan kejadian penyakit dari hari ke hari dengan cara menganalisis data gejala dan melakukan prediksi, sehingga kejadian yang membahayakan dari suatu penyakit dapat dengan cepat diantisipasi dan diatasi. Untuk kasus komplikasi hipertensi maka dapat dilakukan prediksi penyakit komplikasi yang akan dialami oleh seorang pasien berdasarkan gejala-gejala yang dialami saat ini ketika mengalami tekanan darah tinggi dengan menerapkan teknik data *mining*.

Untuk melakukan prediksi suatu penyakit menggunakan teknik data *mining* maka dapat menggunakan metode klasifikasi yaitu *naive bayes*, *decision tree*, atau *neural network* [5][6][7]. Algoritma *naive bayes* dalam bidang kedokteran dapat dimanfaatkan untuk membantu prediksi penyakit dengan diterapkan pada suatu sistem pakar atau sistem pendukung keputusan [4]. Mengacu pada beberapa literatur, algoritma *naive bayes* memiliki kelemahan yang cukup signifikan saat melakukan pemilihan atribut, karena *naive bayes* merupakan metode klasifikasi data *mining* yang menggunakan konsep statistik yang berdasarkan pada *Teorema Bayes* sehingga hanya dapat digunakan untuk memprediksi peluang keanggotaan pada suatu grup atau kelas, untuk mengatasi hal tersebut dibutuhkan pembobotan atribut agar dapat meningkatkan akurasi *naive bayes*

yang lebih baik [8]. Proses pembobotan atribut dapat menggunakan *gain ratio* untuk meningkatkan akurasi algoritma *naive bayes* dengan peningkatan akurasi metode *naive bayes* yang cukup baik berkisar antara 0,03% hingga 16%, algoritma *naive bayes* dengan pemilihan atribut berdasarkan bobot *gain ratio* dipilih karena algoritma *naive bayes* memiliki keunggulan dari segi kemudahan proses perhitungan dan penggunaan sumber daya untuk proses *mining* atau proses komputasi termasuk paling rendah, kelemahan algoritma *naive bayes* dalam melakukan seleksi atribut dapat diperbaiki dengan pemilihan atribut menggunakan metode *gain ratio* [9][10][11].

Penelitian ini bertujuan untuk melakukan analisis klasifikasi algoritma *naive bayes* dengan pemilihan atribut (*feature selection*) berdasarkan bobot *gain ratio* yang dalam penelitian ini akan disebut *Naive Bayes Gain Rasio* (NBGR) hasil klasifikasi dari metode tersebut akan dibandingkan dengan metode *naive bayes* konvensional dalam memprediksi komplikasi hipertensi.

2. Metode

2.1. Kajian Literatur

Berikut beberapa teori yang digunakan dalam penelitian ini :

1. Hipertensi

Tekanan darah tinggi, juga dikenal sebagai hipertensi, adalah tekanan darah yang secara konsisten lebih tinggi dari batas normal. Ada 2 jenis pengukuran tekanan darah: sistolik dan diastolik. Darah sistolik tekanan adalah tekanan di arteri saat jantung berdetak, sedangkan tekanan diastolik adalah tekanan pada arteri saat jantung beristirahat. Jika tidak dikendalikan bisa dengan baik maka dapat meningkatkan risiko penyakit jantung, stroke, dan penyakit lainnya [12].

2. Data Mining

Data *mining* merupakan kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Hasil dari data *mining* ini bisa digunakan untuk memperbaiki pengambilan keputusan di masa depan. Secara sederhana data *mining* adalah langkah-langkah dalam mendapatkan atau menemukan pengetahuan [13].

3. Naive Bayes

Algoritma *naive bayes* merupakan sebuah metode klasifikasi yang didasarkan pada probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes. Algoritma *naive bayes* memprediksi peluang dimasa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai *teorema bayes*. Ciri utama dari *naive bayes classifier* ini adalah asumsi yang sangat kuat (naif) akan independensi dari masing-masing kondisi atau kejadian. Algoritma *naive bayes* dapat dirumuskan menggunakan persamaan (1).

$$P(H|X) = \frac{P(X|H)}{P(X)} P(H) \quad (1)$$

Keterangan :

X : Data dengan kelas yang belum diketahui

H : Hipotesis data

P(H|X) : Probabilitas hipotesis H berdasarkan kondisi X (posteriori probabilitas)

P(H) : Probabilitas Hipotesis

P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H

P(X) : Probabilitas X

4. Gain Rasio

Pembobotan atribut kelas pada *naive bayes* dapat meningkatkan pengaruh prediksi. Dengan memperhitungkan bobot atribut terhadap kelas, maka yang menjadi dasar

ketepatan klasifikasi bukan hanya probabilitas melainkan juga dari bobot setiap atribut terhadap kelas.

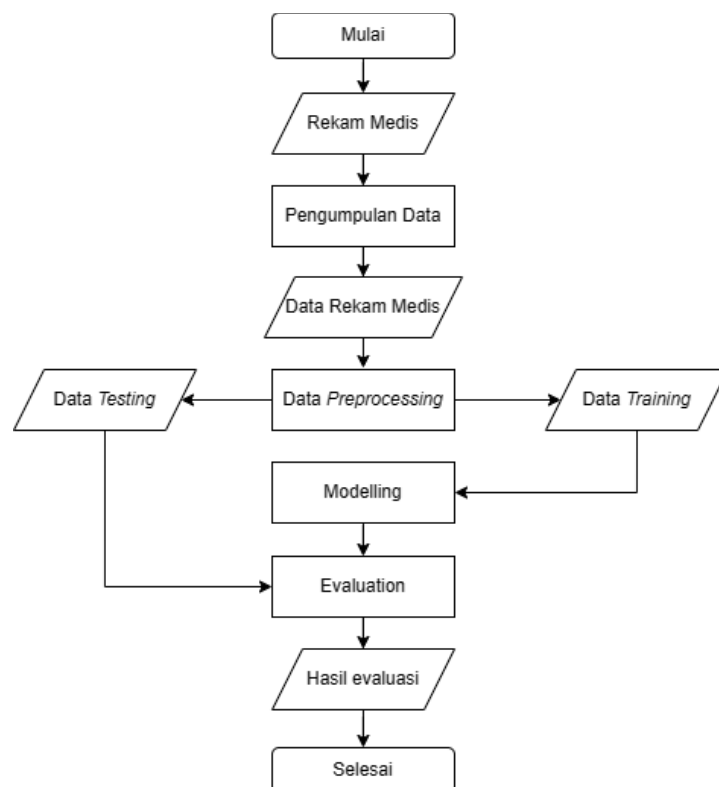
Perhitungan bobot setiap atribut dapat dilakukan menggunakan *gain ratio* [6]. *Gain ratio* pada persamaan (2) melakukan normalisasi pada *information gain* dengan menggunakan *split information (splitinfo)* [14]. *Splitinfo* merepresentasikan *entropy* atau informasi potensial yang dapat dilihat pada persamaan (3).

$$\text{gain ratio}(a) = \frac{\text{gain}(a)}{\text{splitinfo}(a)} \quad (2)$$

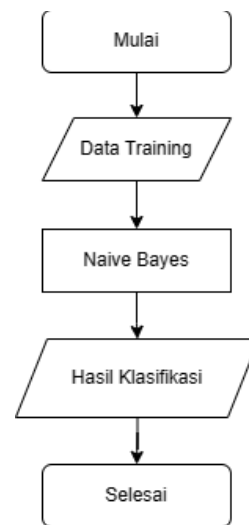
$$\text{splitinfo} = - \sum_{i=1}^n \frac{S_i}{S} \log_2 \frac{S_i}{S} \quad (3)$$

2.2. Metode Penelitian

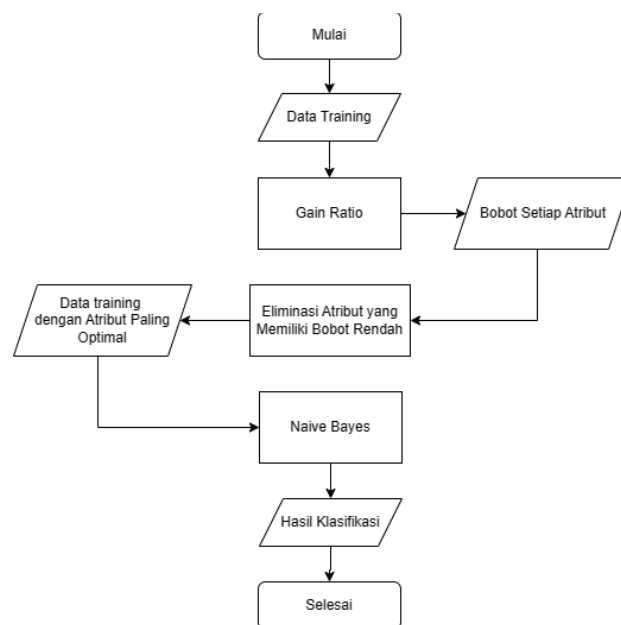
Penelitian ini terdiri dari empat proses yaitu pengumpulan data, data *preprocessing*, data *mining*, dan evaluasi. Alur penelitian ini dapat dilihat pada *flow chart* Gambar 1. Algoritma *naive bayes* konvensional dapat dilihat pada Gambar 2 dan Algoritma *Naive Bayes Gain Rasio* dapat dilihat pada Gambar 3.



Gambar 1. Alur Penelitian



Gambar 2. Algoritma Naive Bayes Konvensional



Gambar 3. Algoritma Naive Bayes Gain Ratio

1. Pengumpulan Data

Pada tahap ini penulis melakukan pengumpulan data yang digunakan sebagai data *training* dan *testing* pada proses *mining pada* unit Pusat Pelaporan dan Informasi, Rumah Sakit Umum Daerah Kabupaten Buleleng, data yang dikumpulkan merupakan rekam medis pasien yang mengalami penyakit hipertensi hingga menderita komplikasi akibat dari hipertensi tersebut pada tahun 2019. Rekam medis yang dikumpulkan akan melalui tahap seleksi terlebih dahulu sesuai dengan kebutuhan sehingga didapatkan data *training* yang sesuai dengan kriteria yaitu data yang berasal dari rekam medis pasien yang mengalami hipertensi dan kemudian mengalami penyakit komplikasi yang diakibatkan oleh hipertensi.

2. Data Preprocessing

Tahap data *preprocessing* dilakukan dengan data *transformation* untuk mentransformasi bentuk data agar lebih mudah melakukan *modelling*. Data yang telah melalui proses tersebut kemudian dibagi 80% sebagai data *training* dan 20% sebagai data *testing*. Data *testing* masuk ke proses pemodelan hingga didapatkan hasil prediksi komplikasi

hipertensi kemudian data *testing* akan digunakan untuk mengevaluasi hasil prediksi masing-masing metode.

3. Modelling

Pada tahap ini data yang sudah melalui tahap *preprocessing* mulai dilakukan proses *mining* menggunakan metode pertama yaitu *naive bayes* dengan pembobotan *gain ratio* dan *naive bayes* konvensional. *Output* dari tahap ini yaitu masing-masing metode akan menghasilkan prediksi penyakit komplikasi hipertensi berdasarkan data training yang digunakan. Hasil prediksi dari algoritma *naive bayes gain ratio* dan *naive bayes* konvensional dievaluasi untuk melihat perbandingan kedua metode.

4. Evaluasi

Tahap evaluasi digunakan untuk mengukur performa dari setiap metode yang digunakan dalam proses data *mining*. Hasil dari pengujian atau evaluasi akan dipetakan dalam bentuk *confusion matrix*. *Pemetaan confusion matrix* dapat dilihat pada Tabel 1. Untuk menghitung akurasi dalam *confusion matrix* dapat dihitung menggunakan persamaan 4, untuk menghitung *precision* menggunakan persamaan 5, dan menghitung *recall* menggunakan persamaan 6.

Tabel 1. Confusion Matrix

	Hasil Prediksi		
	+	-	
Kenyataan	+	True Positive (TP)	False Positive (FP)
	-	False Negative (FN)	True Negative (TN)

$$Akurasi = \frac{TP + TN}{TP + FP + FN + TN} \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

Tahap evaluasi ini juga membandingkan performa masing-masing metode dari segi akurasi, *precision*, dan *recall*. Hasil perbandingan tersebut memberikan pengetahuan apakah dengan menambahkan proses pembobotan menggunakan *gain ratio* pada algoritma *naive bayes* dapat meningkatkan performa dari algoritma *naive bayes* untuk kasus prediksi komplikasi hipertensi.

3. Hasil

3.1. Pengumpulan Data

Hasil pengumpulan data yang dilakukan dengan cara mengambil gejala dan penyakit yang tertulis pada rekam medis pasien yang memiliki riwayat hipertensi, didapatkan berupa 14 data gejala pada Tabel 2 serta 3 klasifikasi penyakit dengan banyaknya data yaitu 50 pasien pada Tabel 3. Gejala yang dialami pasien terdiri dari pusing, mual atau muntah, nyeri dada, nyeri kepala, lemas, kebas, gemetar, mati rasa pada bagian tubuh, penurunan kesadaran, sesak, jantung berdebar, nyeri ulu hati, penurunan berat badan, kehilangan nafsu makan, berbicara pelo atau susah berbicara, merasa lelah, dan edema. Penyakit komplikasi yang berhasil didapatkan terdiri dari jantung, *stroke*, dan gagal ginjal.

Tabel 2. Gejala Komplikasi Hipertensi

Kode Gejala	Gejala
G1	Pusing
G2	Mual atau Muntah
G3	Nyeri Dada
G4	Nyeri Kepala
G5	Mati rasa pada bagian tubuh tertentu
G6	Penurunan Kesadaran
G7	Sesak
G8	Jantung Berdebar
G9	Nyeri Uluhati
G10	Penurunan Berat Badan
G11	Kehilangan Nafsu Makan
G12	Berbicara Pelo atau Susah Berbicara
G13	Merasa Lelah
G14	Edema

Tabel 3. Data Pasien Komplikasi Hipertensi

No	Pasien	Gejala					Diagnosa
		G1	G2	G3	...	G14	
1	pasien 1	Ya	Ya	Tidak	...	Tidak	jantung
2	pasien 2	Tidak	Ya	Ya	...	Tidak	jantung
3	pasien 3	Tidak	Tidak	Tidak	...	Tidak	Stroke
...
50	pasien 50	Tidak	Tidak	Tidak	...	Tidak	Stroke

Tabel 3 menunjukkan gejala yang dialami setiap pasien akan diisi dengan Ya sedangkan gejala yang tidak dialami oleh pasien akan diisi dengan Tidak, dan pada kolom terakhir menunjukkan hasil diagnosa yang ditulis pada rekam medis pasien.

3.2. Data Preprocessing

Berikutnya dilakukan transformasi data untuk memudahkan proses pemodelan dengan mengubah status Ya menjadi angka 1 dan status tidak menjadi 0, kemudian penyakit gagal ginjal diberikan kode G, jantung diberikan kode J, dan stroke diberikan kode S seperti ditunjukkan pada Tabel 4.

Tabel 4. Hasil Data Preprocessing

G1	G2	G3	G12	...	G14	P
0	1	1	0	...	1	G
0	1	0	0	...	0	G
1	0	0	0	...	0	S
...
0	0	0	1	...	0	S

3.3. Modelling

1. Naive Bayes Gain Ratio

Pada tahap pertama peneliti menghitung bobot setiap atribut gejala menggunakan *gain ratio* berdasarkan Persamaan (2) dan telah diurutkan dari bobot terbesar ke terkecil yang hasilnya dapat dilihat pada Tabel 5, hasil perhitungan bobot tersebut kemudian

digunakan untuk menentukan seleksi atribut yang masuk ke perhitungan *naive bayes* dan dilanjutkan dengan melakukan klasifikasi dengan algoritma *naive bayes*.

Tabel 5. Hasil Perhitungan *Gain Ratio*

No	Atribut	Gain Ratio
1	G11	0,525
2	G13	0,525
3	G14	0,525
4	G10	0,462
5	G6	0,433
6	G12	0,433
7	G2	0,316
8	G5	0,214
9	G3	0,211
10	G9	0,183
11	G8	0,164
12	G1	0,164
13	G7	0,110
14	G4	0,043

Berdasarkan pada Tabel 5 maka dapat dilihat bahwa atribut G11, G13, dan G14 memiliki bobot terbesar yang artinya atribut tersebut memiliki pengaruh yang cukup besar terhadap klasifikasi data dan sebaliknya atribut G4 memiliki bobot *gain ratio* terendah sehingga pengaruh atribut tersebut terhadap hasil klasifikasi juga rendah. Berdasarkan bobot *gain ratio* tersebut penulis menemukan bahwa atribut yang paling optimal untuk dieliminasi adalah atribut G4, G7, G1, dan G8.

Setelah didapatkan atribut yang digunakan untuk proses klasifikasi, atribut tersebut masuk ke perhitungan *naive bayes*. Berdasarkan hasil seleksi atribut menggunakan *gain ratio* maka atribut yang masuk ke proses perhitungan *naive bayes* adalah G11, G13, G14, G10, G6, G12, G2, G5, G3, G9. Langkah-langkah menghitung *naive bayes* dibantu dengan menggunakan bahasa pemrograman python dengan hasil yang menunjukkan bahwa model telah berhasil melakukan prediksi komplikasi hipertensi pada Gambar 4.

```
# Melakukan prediksi komplikasi hipertensi
prediksi = modelnb.predict(x_test)
prediksi

array(['J', 'J', 'S', 'J', 'S', 'S', 'G', 'J', 'G', 'J'], dtype='<U1')
```

Gambar 4. Hasil Prediksi Metode *Naive Bayes Gain Ratio*

2. *Naive Bayes* Konvensional

Berbeda dengan *Naive Bayes Gain Ratio*, algoritma *Naive Bayes* konvensional memasukkan seluruh atribut yang ada tanpa mengeliminasi atribut apapun seperti terlihat dari flowchart pada Gambar 2, sehingga atribut yang masuk pada proses ini adalah atribut G1 hingga atribut G14. Hasil prediksi menggunakan metode *Naive Bayes* konvensional dapat dilihat pada Gambar 5.


```

# Melakukan prediksi komplikasi hipertensi
prediksi = modelnb.predict(x_test)
prediksi

array(['J', 'J', 'G', 'J', 'J', 'J', 'J', 'J', 'J'], dtype='<U1')

```

Gambar 5. Hasil Prediksi Metode *Naive Bayes* Konvensional

3.4. Evaluasi

Setelah mendapatkan hasil prediksi komplikasi hipertensi dari setiap metode selanjutnya penulis melakukan evaluasi algoritma naive bayes gain ratio dengan menggunakan *confusion matrix* dan distribusi data *training* maupun *testing* menggunakan mekanisme *cross validation*. Data dibagi menjadi empat bagian sehingga terdapat empat kali pengujian dengan data *training* dan data *testing* yang berbeda-beda.

1. Evaluasi *Naive Bayes Gain Ratio*

Pengujian hasil klasifikasi naive bayes gain ratio dengan menggunakan *confusion matrix* dan mekanisme *cross validation* didapatkan hasil seperti pada Tabel 6. Berdasarkan hasil *confusion matrix* tersebut maka penulis dapat menghitung nilai *precision*, *recall*, dan akurasi dari algoritma naive bayes gain ratio yang dapat dilihat pada Tabel 7.

Tabel 6. Hasil Confusion Matrix Naive Bayes Gain Ratio

Pengujian	Aktual	Prediksi		
		Gagal Ginjal	Jantung	Stroke
1	Gagal Ginjal	2	0	0
	Jantung	0	5	0
	Stroke	0	0	3
2	Gagal Ginjal	2	0	0
	Jantung	0	6	0
	Stroke	0	0	2
3	Gagal Ginjal	2	0	0
	Jantung	0	6	0
	Stroke	0	1	1
4	Gagal Ginjal	2	0	0
	Jantung	0	6	0
	Stroke	0	1	1

Tabel 7. Hasil Evaluasi Naive Bayes Gain Ratio

Pengujian	Class	Precision	Recall	Akurasi
1	Gagal Ginjal	1,00	1,00	100%
	Jantung	1,00	1,00	
	Stroke	1,00	1,00	
2	Gagal Ginjal	1,00	1,00	100%
	Jantung	1,00	1,00	
	Stroke	1,00	1,00	
3	Gagal Ginjal	1,00	1,00	90%
	Jantung	0,86	1,00	
	Stroke	1,00	0,50	
4	Gagal Ginjal	1,00	1,00	90%
	Jantung	0,86	1,00	
	Stroke	1,00	0,50	
Rata-rata	Gagal Ginjal	1,00	1,00	95%
	Jantung	0,93	1,00	
	Stroke	1,00	0,75	

Hasil evaluasi pada Tabel 7 menunjukkan bahwa algoritma *naive bayes* berbasis *feature selection* menggunakan *gain ratio* memberikan hasil prediksi komplikasi hipertensi yang baik dengan rata-rata akurasi sebesar 95%.

2. Evaluasi *Naive Bayes Konvensional*

Untuk melihat performa dari algoritma *naive bayes gain ratio*, maka penulis menghitung *naive bayes* konvensional tanpa adanya seleksi atribut menggunakan *gain ratio* dengan menggunakan data yang sama, kemudian dilakukan evaluasi metode *naive bayes* konvensional dan didapatkan hasil *confusion matrix* seperti pada Tabel 8, dan hasil evaluasi pada Tabel 9.

Tabel 8. Hasil *Confusion Matrix Naive Bayes Konvensional*

Pengujian	Aktual	Prediksi		
		Gagal Ginjal	Jantung	Stroke
1	Gagal Ginjal	2	0	0
	Jantung	2	3	0
	Stroke	0	0	3
2	Gagal Ginjal	2	0	0
	Jantung	1	3	2
	Stroke	0	0	2
3	Gagal Ginjal	2	0	0
	Jantung	0	3	3
	Stroke	0	1	1
4	Gagal Ginjal	2	0	0
	Jantung	0	6	0
	Stroke	0	1	1

Tabel 9. Hasil *Evaluasi Naive Bayes Konvensional*

Pengujian	Class	Precision	Recall	Akurasi
1	Gagal Ginjal	0,50	1,00	80%
	Jantung	1,00	0,60	
	Stroke	1,00	1,00	
2	Gagal Ginjal	0,67	1,00	70%
	Jantung	1,00	0,50	
	Stroke	0,50	1,00	
3	Gagal Ginjal	1,00	1,00	60%
	Jantung	0,75	0,50	
	Stroke	0,25	0,50	
4	Gagal Ginjal	1,00	1,00	90%
	Jantung	0,86	1,00	
	Stroke	1,00	0,50	
Rata-rata	Gagal Ginjal	0,79	1,00	75%
	Jantung	0,90	0,65	
	Stroke	0,69	0,75	

Hasil evaluasi pada Tabel 9 menunjukkan bahwa algoritma *naive bayes* berbasis *feature selection* menggunakan *gain ratio* memberikan hasil prediksi komplikasi hipertensi yang baik dengan rata-rata akurasi sebesar 75%. Sehingga dapat dilihat bahwa *naive bayes gain ratio* memberikan rata-rata akurasi yang lebih unggul 20%.

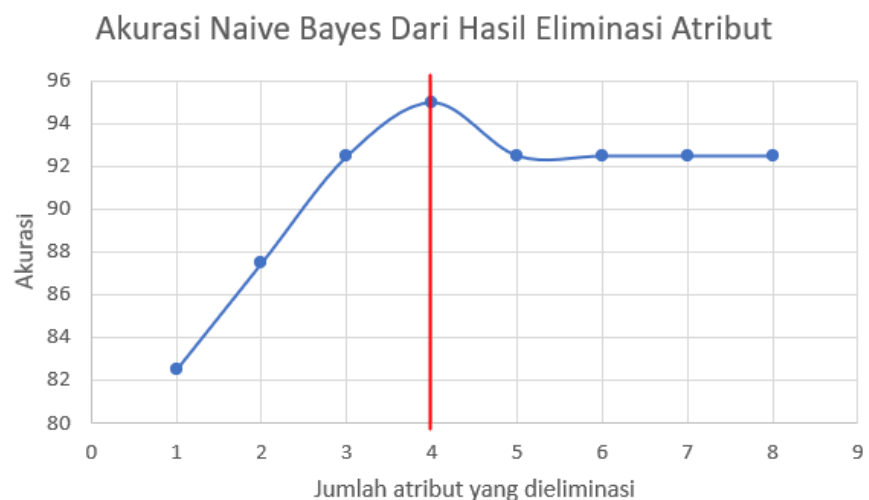
4. Pembahasan

Klasifikasi komplikasi hipertensi menggunakan algoritma *naive bayes* berbasis *feature selection gain rasio* dan *naive bayes* konvensional menghasilkan tingkat akurasi, *precision*, dan *recall* yang berbeda-beda. Perbedaan hasil tersebut terjadi karena setiap metode memberikan *treatment* yang berbeda kepada *dataset*. Pengaruh terbesar yang peneliti temukan dari tiga metode yang diajukan terletak pada proses perhitungan bobot setiap atribut.

Secara keseluruhan peneliti terkendala ketika menerapkan pemisahan data *training* dan data testing karena jumlah data yang berhasil didapatkan hanya sedikit sehingga ketika dilakukan pemisahan maka ada kelas yang tidak diwakilkan dalam data testing dan ada kelas yang tidak muncul pada hasil prediksi setiap metode, untuk mengatasi hal tersebut peneliti mengacak posisi *record* pasien dan mengambil data yang digunakan sebagai *training* dan testing dengan mekanisme *cross validation* dan memastikan pada setiap data *training* dan testing setiap kelas yang ada sudah diwakilkan.

Algoritma *naive bayes* konvensional menghitung nilai probabilitas dari setiap atribut dan kelas tanpa memperhatikan apakah setiap atribut yang dihitung memiliki pengaruh terhadap hasil klasifikasi atau tidak [6]. Algoritma *naive bayes gain ratio* memiliki hasil rata-rata akurasi yang lebih unggul 20% dibanding *naive bayes* konvensional karena setiap atribut pada *dataset* dihitung bobotnya terlebih dahulu menggunakan metode *gain ratio* kemudian atribut yang memiliki bobot *gain ratio* yang rendah akan dieliminasi sehingga proses klasifikasi *naive bayes* menjadi lebih baik [7] [8].

Untuk membuktikan apakah metode *gain ratio* dapat digunakan dalam menghitung bobot setiap atribut yang memiliki pengaruh terhadap hasil klasifikasi *naive bayes* maka peneliti mencoba untuk mengeliminasi atribut-atribut berdasarkan bobot gain rasio pada **Tabel 5** satu persatu dimulai dari atribut dengan bobot terendah. Peneliti mengevaluasi dengan melihat tingkat akurasi setiap proses eliminasi atribut dan menyajikannya dalam bentuk grafik pada Gambar 6. Berdasarkan grafik tersebut penulis menemukan bahwa hasil akurasi terbaik sebesar 95% didapatkan dengan mengeliminasi empat atribut yang memiliki bobot gain rasio terendah yaitu G4, G7, G1, dan G8. Dari hasil percobaan yang peneliti lakukan dengan cara mengeliminasi atribut yang memiliki nilai *gain ratio* terendah membuktikan bahwa nilai *gain ratio* dapat digunakan untuk mengetahui seberapa besar pengaruh suatu atribut terhadap hasil klasifikasi *naive bayes* dalam memprediksi komplikasi hipertensi [7] [8].



Gambar 6. Grafik Akurasi Berdasarkan Eliminasi Atribut dengan Bobot Terendah

Untuk menentukan batas dalam mengeleminasi atribut berdasarkan bobot gain ratio selain dengan mekanisme yang penulis lakukan, juga dapat menggunakan nilai standar

deviasi sebagai batas atau *threshold* bobot eliminasi atribut[15]. Berdasarkan **Tabel 5** maka nilai standar deviasi dari bobot gain ratio setiap atribut adalah 0,1712 sehingga atribut yang dieliminasi berdasarkan batas nilai standar deviasi memiliki hasil yang sama dengan atribut yang dieliminasi berdasarkan mekanisme yang penulis lakukan. Secara lebih detail perbandingan naive bayes konvensional (NB) dengan naive bayes gain ratio (NBGR) dapat dilihat pada Tabel 10.

Tabel 10. Perbandingan NB dengan NBGR

Class	NB			NBGR		
	Precision	Recall	Akurasi	Precision	Recall	Akurasi
Gagal ginjal	0,79	1,00		1,00	1,00	
Jantung	0,90	0,65	75%	0,93	1,00	95%
Stroke	0,69	0,75		1,00	0,75	

Terjadi peningkatan nilai precision sebesar 21% pada algoritma *naive bayes gain ratio* untuk kelas gagal ginjal, peningkatan sebesar 3% pada kelas jantung, dan peningkatan sebesar 31% pada kelas stroke, untuk nilai *recall* terjadi peningkatan pada algoritma *naive bayes gain ratio* sebesar 35% pada kelas jantung sedangkan pada kelas gagal ginjal dan stroke tidak terjadi peningkatan nilai *recall*. Dari segi akurasi terjadi peningkatan sebesar 20% pada algoritma *naive bayes gain ratio*.

5. Kesimpulan

Berdasarkan hasil dan pembahasan perbandingan antara *naive bayes* berbasis *feature selection* menggunakan *gain ratio* dengan *naive bayes* konvensional untuk klasifikasi komplikasi hipertensi, ditemukan beberapa simpulan penting. Pembobotan *gain ratio* mampu mengindikasikan sejauh mana pengaruh atribut terhadap hasil klasifikasi *naive bayes*. Hasil eksperimen menunjukkan bahwa akurasi optimal dapat dicapai dengan mengeliminasi empat atribut dengan bobot *gain ratio* terendah dalam kasus komplikasi hipertensi. *Feature selection* dengan pembobotan *gain ratio* dapat meningkatkan akurasi klasifikasi *naive bayes*, dengan rata-rata peningkatan akurasi sebesar 20% dibandingkan dengan *naive bayes* tanpa *feature selection*. Berdasarkan *confusion matrix* dapat dilihat bahwa nilai *precision* lebih unggul sebesar 21% pada algoritma *naive bayes gain ratio* untuk kelas gagal ginjal, peningkatan sebesar 3% pada kelas jantung, dan peningkatan sebesar 31% pada kelas stroke, untuk nilai *recall* terjadi peningkatan pada algoritma *naive bayes gain ratio* sebesar 35% pada kelas jantung sedangkan pada kelas gagal ginjal dan stroke tidak terjadi peningkatan nilai *recall*.

Ucapan Terima Kasih: Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa karena atas karunia-Nya penyusunan artikel ini dapat diselesaikan dengan baik, serta ucapan terima kasih penulis ucapkan kepada semua pihak yang telah membantu dalam penyusunan penelitian ini.

Referensi

- [1] Badan Penelitian dan Pengembangan Kesehatan, *Laporan Nasional Riskesdas 2018*. Jakarta: Badan Penelitian dan Pengembangan Kesehatan, 2020.
- [2] Rokom, "Hipertensi Penyakit Paling Banyak Diidap Masyarakat." Accessed: Jan. 31, 2024. [Online]. Available: <https://sehatnegeriku.kemkes.go.id/baca/umum/20190517/5130282/hipertensi-penyakit-paling-banyak-diidap-masyarakat/>
- [3] A. Adinata, "Pengembangan Tensimeter Digital Bayesian Berbasis Arduino Untuk Prediksi Komplikasi," Universitas Pendidikan Ganesha, Singaraja, 2019.

-
- [4] R. Rizky, S. Susilawati, Z. Hakim, and L. Sujai, "Sistem Pakar Deteksi Penyakit Hipertensi Dan Upaya Pencegahannya Menggunakan Metode Naive Bayes Pada RSUD Pandeglang Banten," *JURNAL TEKNIK INFORMATIKA UNIS*, vol. 7, no. 2, pp. 138–144, Jan. 2020, doi: 10.33592/jutis.v7i2.395.
- [5] M. M. Santoni, N. Chamidah, and N. Matondang, "Prediksi Hipertensi menggunakan Decision Tree, Naïve Bayes dan Artificial Neural Network pada software KNIME," *Techno.Com*, vol. 19, no. 4, pp. 353–363, Nov. 2020, doi: 10.33633/tc.v19i4.3872.
- [6] .. A. T., .. H., .. A. A., .. E.-S. M., and M. M. Eid, "A Comparative Analysis of Methods for Detecting and Diagnosing Breast Cancer Based on Data Mining," *Journal of Artificial Intelligence and Metaheuristics*, vol. 4, no. 2, pp. 08–17, 2023, doi: 10.54216/JAIM.040201.
- [7] S. Reddy, N. Sethi, R. Rajender, and G. Mahesh, "Forecasting Diabetes Correlated Non-alcoholic Fatty Liver Disease by Exploiting Naive Bayes Tree," *ICST Transactions on Scalable Information Systems*, p. 173975, Jul. 2018, doi: 10.4108/eai.29-4-2022.173975.
- [8] N. A. Siagian, "Pembobotan Fitur Dataset Menggunakan Gain Ratio Guna Meningkatkan Akurasi Metode Naïve Bayesian Classifier," Universitas Sumatera Utara, 2020.
- [9] F. Ghasemi, B. S. Neysiani, and N. Nematbakhsh, "Feature Selection in Pre-Diagnosis Heart Coronary Artery Disease Detection: A heuristic approach for feature selection based on Information Gain Ratio and Gini Index," in *2020 6th International Conference on Web Research (ICWR)*, IEEE, Apr. 2020, pp. 27–32. doi: 10.1109/ICWR49608.2020.9122285.
- [10] I. Santiko and I. Honggo, "Naive Bayes Algorithm Using Selection of Correlation Based Featured Selections Features for Chronic Diagnosis Disease," *IJIS: International Journal of Informatics and Information Systems*, vol. 2, no. 2, pp. 56–60, Sep. 2019, doi: 10.47738/ijis.v2i2.14.
- [11] F. M. Okikiola, O. S. Adewale, and O. O. Obe, "A DIABETES PREDICTION CLASSIFIER MODEL USING NAIVE BAYES ALGORITHM," *FUDMA JOURNAL OF SCIENCES*, vol. 7, no. 1, pp. 253–260, Feb. 2023, doi: 10.33003/fjs-2023-0701-1301.
- [12] A. N. Desai, "High Blood Pressure," *JAMA*, vol. 324, no. 12, p. 1254, Sep. 2020, doi: 10.1001/jama.2020.11289.
- [13] E. Prasetyowati, *Data mining : pengelompokan data untuk informasi dan evaluasi*. Pamekasan: Duta Media Publishing, 2017.
- [14] D. Jollyta, *Konsep data mining dan penerapan*. Deepublish, 2020.
- [15] M. I. Prasetyowati, N. U. Maulidevi, and K. Surendro, "Determining threshold value on information gain feature selection to increase speed and prediction accuracy of random forest," *J Big Data*, vol. 8, no. 1, p. 84, Dec. 2021, doi: 10.1186/s40537-021-00472-4.