



## **Implementasi Algoritma Naïve Bayes Dalam Mendiagnosa Penyakit Angin Duduk**

*Ahmad Turmudi Zy, Lutfi Adji Ardiansyah, Donny Maulana*

**Program Studi Teknik Informatika, Universitas Pelita Bangsa Korespondensi email:**

[turmudi@pelitabangsa.ac.id](mailto:turmudi@pelitabangsa.ac.id)

---

### **Abstrak**

Angina or wind sitting disease is chest pain caused by reduced blood flow to the heart, making it a severe pain in the left chest area and can radiate to the left shoulder and followed by breathless. The risk of a person experiencing sitting winds includes entering old age, having a family history of heart disease, hypertension, smoking and other medical conditions. Sitting wind disease can potentially lead to heart attacks if not treated properly. This study aims to determine the level of accuracy and the effect of the Naive Bayes algorithm on the sitting wind data used in this study. As well as getting information about accuracy, precision and recall that can be obtained when testing data using the Naïve Bayes algorithm. Data processing using the RapidMiner tool, the dataset used in this study is divided into 80% training data and 20% testing data. The results of this study stated that the accuracy rate is 87.50%, precision is 94.12%, and the recall is 80%. Based on the research conducted, it can be concluded that the process of determining the status of sitting wind patients using the Naïve Bayes algorithm has good accuracy in this study.

---

### **Informasi Artikel**

Diterima: 20-03-2021

Direvisi: 04-04-2021

Dipublikasikan: 30-04-2021

---

### **Keywords**

Sitting Wind, Naïve Bayes, Data Mining, RapidMiner

## Introduction

Penyakit Angin Duduk sudah cukup lama dikenal oleh masyarakat, penyakit ini sering dianggap masyarakat awam sebagai lanjutan dari masuk angin yang berbahaya dan akibatnya dapat menimbulkan

kematian yang mendadak bagi penderitanya. Menurut ilmu kesehatan atau ilmu medis tidak ada istilah ataupun penyakit angin duduk melainkan di dalam ilmu kesehatan penyakit ini termasuk dalam golongan penyakit tertentu yang bisa dikenali dari keluhan pasien, hasil tes

laboratorium ataupun ditelusuri dari riwayat terdahulu penyakit yang diderita oleh pasien maupun keluarganya tersebut.

Masalah penyakit angin duduk menurut penjelasan ilmu kesehatan disebut juga dengan istilah Angina Pektoris. Penyakit ini disebabkan karena adanya penyumbatan di pembuluh jantung yang mengakibatkan kurangnya asupan oksigen ke jantung yang menjadikannya nyeri hebat seperti terasa tertusuk di area dada sebelah kiri kemudian dapat juga menjalar sampai ke baru sebelah kiri maupun bagian punggung sebelah kiri dan diikuti sesak nafas. Biasanya jika serangan ini menyerang tidak cepat tertolong maka akibatnya pasien akan meninggal dunia.

Data kasus penyakit angina di Indonesia masih sangat terbatas, pada tahun 2009 jumlah kasus angina di provinsi Jawa Tengah sebesar 16.632 kasus, dari total kasus penyakit jantung koroner sebesar 24.031 kasus dan angka kematian akibat angin duduk

sebesar 89 kasus di Jawa Tengah (Sabransyah, et al., 2017). Dari sumber data Riset Kesehatan Dasar (Riskesdas), tingkat angka penyakit pembuluh darah dan jantung meningkat dari tahun ke tahun.

Berdasarkan penjelasan di atas, pada penelitian ini akan dilakukan proses implementasi metode Naïve Bayes untuk klasifikasi penyakit angin duduk. Pemilihan penggunaan algoritma tersebut lebih dikarenakan merupakan algoritma yang sangat populer dan banyak digunakan secara praktis. Untuk melakukan pengolahan data penyakit angin duduk dengan

menerapkan data mining. Penulis melakukan teknik klasifikasi menggunakan algoritma Naïve Bayes untuk melakukan pengolahan datanya menggunakan aplikasi RapidMiner sehingga akan didapatkan hasil akurasi, precision, dan recall dari algoritma Naïve Bayes terhadap penyakit angin duduk.

Adapun Tujuan yang akan dicapai dalam penelitian ini, yaitu:

1. Untuk mengimplementasikan metode klasifikasi data mining menggunakan algoritma Naïve Bayes untuk menentukan status angin duduk menggunakan RapidMiner
2. Untuk mendapatkan informasi tentang tingkat accuracy, precision dan recall yang di dapat dari hasil pengujian data menggunakan algoritma Naïve Bayes

## I. Landasan Teori

### 2.1 Definisi Angin Duduk

Penyakit Angin Duduk sudah cukup lama dikenal oleh masyarakat,

penyakit ini sering dianggap masyarakat awam sebagai lanjutan dari masuk angin yang berbahaya dan akibatnya dapat menimbulkan kematian yang mendadak bagi penderitanya. Menurut ilmu kesehatan atau ilmu medis tidak ada istilah ataupun penyakit angin duduk melainkan di dalam ilmu kesehatan penyakit ini termasuk dalam golongan penyakit tertentu yang bisa dikenali dari keluhan pasien, hasil tes laboratorium ataupun ditelusuri dari riwayat terdahulu penyakit yang diderita oleh pasien maupun keluarganya tersebut.

Keluhan dari angin duduk sebetulnya bersifat relatif sama, gangguan yang dirasakan yaitu mengeluh pada bagian dada sebelah kiri yang terasa berat, dan juga sakit yang dirasakan menjalar hingga ke punggung, leher, rahang, dan lengan kanan. Namun umumnya pasien merasakan tidak enak badan seolah-olah seperti ada angin yang memenuhi area dada dan perut tetapi sulit untuk dihilangkan. Penyebab dari angin duduk adalah adanya Sumbatan yang menutup di pembuluh darah jantung. Dengan adanya sumbatan itu maka darah yang harusnya berjalan mulus ke jantung malah tidak bisa masuk ke jantung akibatnya jantung akan kekurangan oksigen. Di tambah lagi dengan adanya aktifitas fisik seperti beratnya pekerjaan fisik dan emosi menjadikan lebih berat dan oleh karena itu menyebabkan meningkatnya kebutuhan akan oksigen ke jantung.

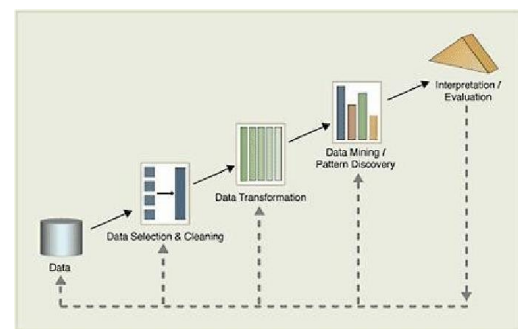
## 2.2 Data Mining

Data mining ialah salah satu ilmu komputer yang merupakan campuran sebuah metode analisis data yang menggunakan algoritma untuk memproses data berukuran besar atau dataset. Data mining sudah banyak

digunakan di banyak bidang, salah satunya dalam bidang kedokteran dan bisnis.

Data mining biasa disebut juga sebagai penemuan atau penambangan informasi yang sifatnya baru dengan mencari pola atau aturan tertentu dari sumber data (dataset) yang diharapkan dapat mengatasi kondisi tersebut, sehingga di dapatkan pengetahuan baru yang bisa digunakan nantinya.

Data mining memiliki tahapan- tahapan dalam pemrosesannya. Berikut adalah tahapan-tahapan dalam data mining.



Gambar 1 Tahapan Data Mining

### 1. Data Selection

Proses pemilihan dari sekumpulan data operasional yang ada sebelum masuk ke dalam tahap data mining. Pada proses ini dilakukan pemilihan data atribut atau variable yang relevan, nantinya data tersebut akan digunakan dalam penelitian.

### 2. Data Transformation

Setelah data telah ditentukan dan juga sudah dilakukan proses pembersihan data, selanjutnya ialah mengubah format data agar bisa diproses untuk data mining. Tahap ini juga merupakan proses transformasi pada data yang telah dipilih, sehingga

data tersebut sesuai untuk proses data mining

### 3. Knowledge Extraction

Suatu proses mencari pola dari data yang telah dipilih dengan menggunakan metode atau teknik yang bervariasi, pemilihan metode yang tepat akan berpengaruh pada tujuan penelitian. Intinya pola yang ingin ditampilkan dari metode maupun teknik yang dipilih sebelumnya misalnya Selection, klasifikasi, clustering, transforming, dan lain sebagainya.

### 4. Interpretation/Evaluation

Tahap ini merupakan tahap paling akhir dari beberapa tahapan data mining. Pada tahap ini memiliki tujuan untuk menampilkan pola atau aturan yang telah dihasilkan dari proses data mining lalu ditampilkan dengan bentuk sederhana agar bisa dimengerti oleh pihak yang berkepentingan.

#### 2.3 Definisi Algoritma Naive Bayes

Naïve Bayes adalah sebuah penerapan teorema Bayesian. Algoritma Naïve Bayes didasarkan pada suatu asumsi yang digunakan untuk menyederhanakan atribut dengan mengkondisikan setiap atribut tidak terkait satu sama lain Naïve Bayes juga bisa digolongkan pengklasifikasian menggunakan aturan statistika yang dapat bermanfaat bagi prediksi probabilitas dalam anggota kelas. Naïve bayes memiliki tingkat akurasi yang tergolong tinggi dan kecepatan komputasi yang relative cepat saat digunakan data berdimensi besar. Dalam Naïve Bayes diartikan bahwa tidak ada keterkaitan antar fitur.

Kelebihan dari Naïve Bayes adalah pada metode ini hanya memerlukan sedikit jumlah data latih untuk dapat menentukan parameter yang diperlukan untuk pengklasifikasian. Naive Bayes sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan.

$$P(x|H) = \frac{P(H|x)P(x)}{P(H)}$$

Dimana :

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan class spesifik

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

$P(H)$  : Probabilitas hipotesis H (prior probabilitas)

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

$P(X)$  : Probabilitas X

Dapat dijelaskan dimana  $P(H|X)$  itu merupakan probabilitas dari H yang di dalam X atau bahasa lain dari  $P(H|X)$  merupakan presentase dengan banyaknya H yang di dalam X,  $P(X|H)$  merupakan probabilitas X di dalam H,  $P(H)$  merupakan probabilitas dari H dan  $P(X)$  merupakan probabilitas prior dari

X. Persamaan tersebut yang kemudian menjadi dasar untuk metode Naive Bayes, karena setiap atribut diasumsikan tidak saling terkait antara satu dan lainnya (conditionally independent) maka persamaan dapat dinyatakan sebagai berikut.

$n$

$$P(X | Ci) = \prod_{k=1} P(Xk | Ci)$$

$k=1$

Berdasarkan persamaan tersebut

maka klas (label) dari data sampel X adalah kelas label yang memiliki, nilai maksimum

$$P(X | Ci) \times P(Ci)$$

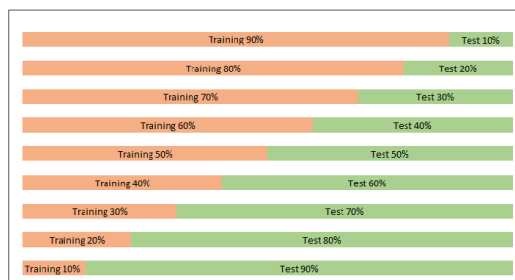
## 2.4 Klasifikasi

Klasifikasi adalah sebuah proses untuk mencari model atau fungsi yang menjelaskan dan membedakan kelas atau konsep dari data, dengan tujuan untuk menggunakan model dan melakukan prediksi dari kelas suatu objek dimana tidak diketahui label dari kelas tersebut.

## 2.5 Split Validation

Split Validation adalah teknik validasi yang membagi data menjadi dua bagian secara acak, bagian pertama digunakan sebagai data training dan bagian kedua digunakan sebagai data testing

Data training adalah data yang akan dipakai dalam melakukan pembelajaran sedangkan data testing adalah data yang belum pernah dipakai sebagai pembelajaran dan akan berfungsi sebagai data pengujian.



Gambar 2 Ilustrasi Split Validation

## 2.6 Confusion Matrix

Confusion matrix merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode kalsifikasi. Pada dasarnya Confusion matrix mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya Pada pengukuran kinerja menggunakan Confusion matrix, terdapat 4 (empat) istilah repersentasi hasil proses klasifikasi. Keempat istilah tersebut adalah True Positive (TP), True Negative (TN), False Positive (FP), dan Flase Negative (FN).

Tabel 1 Confusion Matrix

Nilai Prediksi	Positive	Negative
Positive	TP = True Positive	FN = False Negative
Negative	FP = False Positive	TN = True Negative

## 2.7 Akurasi

Akurasi diperlukan untuk evaluasi dan mengukur keakuratan dari hasil klasifikasi, semakin besar nilai akurasi maka semakin baik tingkat klasifikasinya. Rumus perhitungan nilai akurasi, sebagai berikut

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

Precision atau dikenal dengan nama confidence adalah sebuah model perhitungan untuk mencari hasil dari proporsi jumlah kasus dengan hasil diagnosa positif. Rumus perhitungan nilai precision, sebagai berikut

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

Recall atau sensitivity adalah sebuah model perhitungan untuk mencari hasil proporsi jumlah kasus positif yang diidentifikasi dengan benar. Rumus perhitungan nilai recall, sebagai berikut

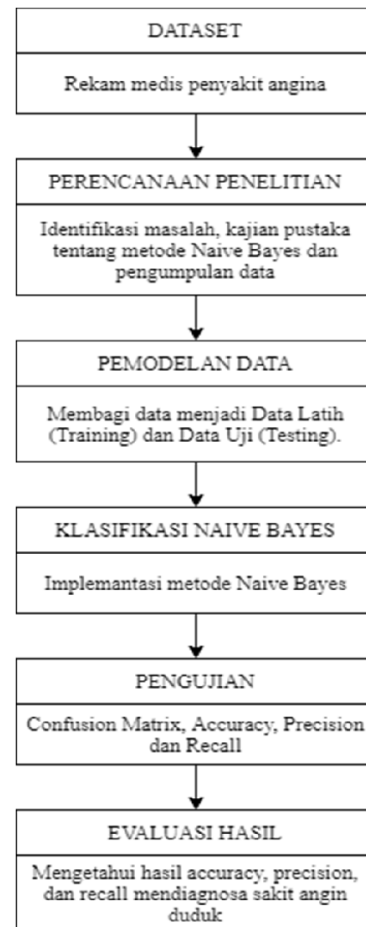
$$Recall = \frac{TP}{TP + FN} \times 100\%$$

## 2.6 Aplikasi RapidMiner

RapidMiner merupakan perangkat lunak untuk pengolahan data. Dengan menggunakan prinsip dan algoritma data mining, RapidMiner mengekstrak pola-pola dari data set yang besar dengan mengkombinasikan metode statistika, kecerdasan buatan dan database.

## 2.7 Kerangka Pemikiran

Kerangka pemikiran yakni alur pikiran penulis yang dijadikan sebagai skema atau dasar-dasar pemikiran untuk memperkuat indikator yang melatar belakangi penelitian ini. Dalam kerangka pemikiran ini peneliti akan mencoba menjelaskan masalah pokok penelitian. Penjelasan yang disusun akan menggabungkan antara teori dengan masalah yang diangkat dalam penelitian ini



Permasalahan pada penelitian ini ialah belum adanya metode yang dapat digunakan untuk memprediksi penyakit angina pektoris, kemudian dilakukan studi literatur sebagai referensi untuk menyelesaikan permasalahan penelitian ini. Dataset yang di dapatkan akan di olah dengan menggunakan algoritma Naive Bayes, dalam pengujian ini metode yang telah di tetapkan yakni menggunakan Confusion Matrix. Dan tools development yang digunakan untuk penelitian ini menggunakan RapidMiner.

## III. HASIL DAN PEMBAHASAN

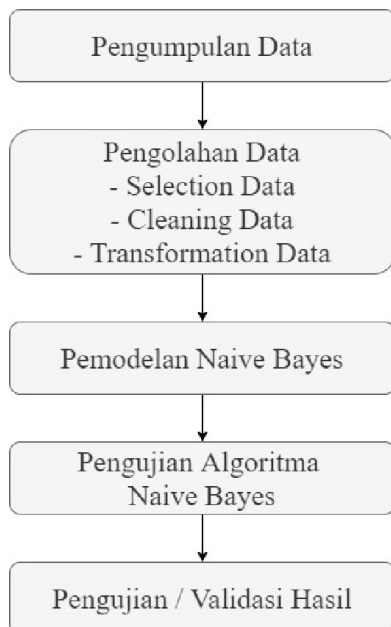
### 3.1 Objek Penelitian

Penelitian yang dilakukan ini adalah penelitian deskriptif dengan metode pendekatan kuantitatif dan

analisis kausalitas. Penelitian deskriptif digunakan untuk mendeskripsikan suatu keadaan atau fenomena apa adanya. Pada penelitian ini, data yang digunakan adalah data yang diambil dari sumber Kaggle.com. Data tersebut berisikan Atribut-atribut yang berkaitan dengan angin duduk dan akan diolah menggunakan beberapa metode data mining sehingga diperoleh satu metode yang paling akurat dan dapat digunakan sebagai bahan acuan dalam mendiagnosa penyakit angin duduk.

### 3.2 Metode Penelitian

Data yang digunakan adalah data yang diambil dari sumber Kaggle.com. Data tersebut berisikan Atribut-atribut yang berkaitan dengan angin duduk dan akan diolah menggunakan beberapa metode data mining sehingga diperoleh satu metode yang paling akurat dan dapat digunakan sebagai bahan acuan dalam mendiagnosa penyakit angin duduk.



**Gambar 3** Tahap Penelitian

### 3.3 Pengolahan Data

Analisa data merupakan proses awal yang harus dilaksanakan untuk menentukan permasalahan yang sedang dihadapi. Data yang akan dijadikan dataset dalam penelitian ini adalah data pasien yang memiliki Penyakit angin duduk, data tersebut akan dibagi menjadi 80% data training dan 20% data testing. Memisahkan data menjadi training dan testing set dimaksudkan agar model yang diperoleh nantinya memiliki kemampuan generalisasi yang baik dalam melakukan klasifikasi data.

#### A. Cleaning Data

Pada tahap data cleaning merupakan proses pembersihan data mana saja yang akan dipakai untuk penghapusan data dengan menghapus missing value, duplikasi data, dan memeriksa inkonsistensi data dan memperbaiki kesalahan pada data.

#### B. Selection Data

Pada proses ini dilakukan pemilihan data atribut atau variable yang relevan, karena tidak semua data yang terdapat data mentah akan digunakan, nantinya data tersebut akan digunakan dalam penelitian.

**Tabel 2** Seleksi Data

Atribut sebelum Selection		Atribut sesudah Selection	
No	Atribut	No	Atribut
1	Status	1	Usia
2	Usia	2	Tipe Perokok
3	Tipe Perokok	3	Hipertensi
4	Rokok perbatang	4	Riw. Keluarga Angina

5	Hipertensi	5	Riw. Keluarga Serangan Jantung
6	Riw. Keluarga Angina	6	Diabetes
7	Riw. Keluarga Serangan Jantung	7	Status
8	Riw. Keluarga Stroke	8	-
9	Diabetes	9	-

### C. Transformation Data

Setelah data telah ditentukan dan juga sudah dilakukan proses pembersihan data, selanjutnya ialah mengubah format data agar bisa diproses untuk data mining. dalam pengolahan dan dapat dihasilkannya hasil akurasi yang baik maka data harus di transforming kedalam bentuk data yang mudah di pahami

### 3.4 Dataset

Dataset merupakan Sebuah himpunan data yang berasal dari informasi masa- masa lampau dan dikelola menjadi sebuah informasi untuk melakukan teknik dari ilmu data mining. Pada penelitian ini jumlah dataset yang digunakan sebanyak 200 data, yang terdiri dari atribut dan variabel yang sudah dijelaskan sebelumnya, Data yang sudah disiapkan untuk klasifikasi dibagi menjadi dua, untuk data training (80%) dan data testing (20%).

Jumlah data  $N = 200$

Testing =  $20\% \times 200$

$N = 40$

**Tabel 3** Dataset Yang Digunakan

Dataset	Data Training	Data Testing
---------	---------------	--------------

200	Angina	Tidak Angina	Angina	Tidak Angina
	80	80	20	20

### 3.5 Tahap Pengujian

Pada tahap pengujian ini, dilakukan 2 jenis tahapan yakni pengujian dilakukan dengan menggunakan perhitungan manual dan proses RapidMiner

### 3.6 Perhitungan Manual Naïve Bayes

**Tabel 4** Tabel Data Testing

USIA	MERO KOK	HIPER TENSI	RW KEL. ANGINA	RW KEL. JANTUNG	DIABETES	STAT
Dewasa	Tinggi	Normal	Tidak	Tidak Ada	Tidak Diabet	Tidak Ar
Lanjut Usia	Rendah	Normal	Tidak	Tidak Ada	Tidak Diabet	Tidak Ar
Lanjut Usia	Rendah	Normal	Tidak	Ada	Tidak Diabet	Angin
Lanjut Usia	Rendah	Normal	Tidak	Tidak Ada	Tidak Diabet	Tidak Ar
Lanjut Usia	Tinggi	Ringan	Tidak	Ada	Tidak Diabet	Angin
Lanjut Usia	Tinggi	Normal	Tidak	Tidak Ada	Tidak Diabet	Angin
Lanjut Usia	Rendah	Normal	Tidak	Tidak Ada	Tidak Diabet	Tidak Ar
Lanjut Usia	Tinggi	Normal	Tidak	Tidak Ada	Tidak Diabet	Angin
Lanjut Usia	Sedang	Normal	Tidak	Ada	Tidak Diabet	Angin
Lanjut Usia	Sedang	Normal	Tidak	Ada	Tidak Diabet	Angin

#### A. Perhitungan Probabilitas Prior

Dari 40 data testing, diketahui bahwa pada label status Angina ( $C_0$ ) sebanyak 20/40 dan label status Tidak angina ( $C_1$ ) sebesar 20/40 dengan hasil yang sama yaitu 0,50 atau 50%.

**Tabel 5** Hitung Probabilitas Prior

$P(C_i)$	JUML AH	HAS IL
ANGINA $P(C_0)$	$\frac{100}{200}$	0,5
TIDAK ANGINA $P(C_1)$	$\frac{100}{200}$	0,5

#### B. Perhitungan Probabilitas Posterior X Bersyarat



**P(X|Ci)** Perhitungan probabilitas posterior, dilakukan pada data testing sebanyak 40 data dengan menggunakan X sebagai vector dalam diagnosa penyakit angin duduk.

### Probabilitas Usia

Pada atribut Usia, dibagi menjadi 2 kategori yaitu, Usia 21 – 40 th adalah dewasa, sedangkan 41 – 80 th adalah lanjut usia, Nilai probabilitas  $P(X|Ci)$ ,

Lanjut Usia: Angina  $90/178 = 0,5056$ ,  
Tidak Angina  $88/178 = 0,4944$

Dewasa: Angina  $10/22 = 0,4545$ , Tidak Angina  $12/22 = 0,4944$

**Tabel 6** Probabilitas Usia

X1	Usia	Jumlah data	Jumlah dipilih		Probabilitas	
			Angina	Tidak angina	Angina	Tidak angina
	Dewasa	22	10	12	0,4545	0,5455
	Lanjut Usia	178	90	88	0,5056	0,4944

### Probabilitas Merokok

Pada atribut ini, dibagi menjadi 3 tingkat yaitu apakah pasien saat ini aktif merokok, sudah tidak merokok, dan tidak pernah merokok, Nilai probabilitas  $P(X|Ci)$

Tinggi: Angina  $61/82 = 0,7439$ , Tidak Angina  $21/82 = 0,2561$

Sedang: Angina  $26/41 = 0,6341$ , Tidak Angina  $15/41 = 0,3659$

Rendah: Angina  $13/77 = 0,1688$ , Tidak Angina  $64/77 = 0,8312$

**Tabel 7** Probabilitas Merokok

X2	Merokok	Jumlah data	Jumlah dipilih		Probabilitas	
			Angina	Tidak angina	Angina	Tidak angina
	Tinggi	82	61	21	0,7439	0,2561
	Sedang	41	26	15	0,6341	0,3659
	Rendah	77	13	64	0,1688	0,8312

### Probabilitas Hipertensi

Pada atribut ini, tingkat hipertensi dibagi menjadi 3 tingkat, menurut Sumber: *2003 Guidelines of ESH*. Klasifikasi Hipertensi terdiri dari:

**Tabel 8** Klasifikasi Hipertensi oleh *Guidelines of ESH 2003*

Kategori	Sistolik (mmHg)	Distolik (mmHg)
Optimal	< 120	< 80
Normal	120 - 129	80 - 84
Normal Tinggi	130 - 139	85 - 89
Tingkat 1 (Ringan)	140 - 159	90 - 99
Tingkat 2 (Sedang)	160 - 179	100 - 109
Tingkat 3 (Berat)	$\geq 180$	$\geq 110$
Hipertensi Tekanan Sistolik Terisolasi	> 140	< 90

Dari tabel di atas, maka atribut hipertensi di bagi menjadi 3 tingkat, dari normal, ringan dan sedang,

$P(X|Ci)$

Normal: Angina  $67/150 = 0,4467$ , Tidak Angina  $83/150 = 0,5533$

Ringan: Angina  $23/37 = 0,6216$ , Tidak Angina  $14/37 = 0,3784$

Sedang: Angina  $10/13 = 0,7692$ , Tidak Angina  $3/13 = 0,2308$

**Tabel 9** Probabilitas Hipertensi

X3	Hipertensi	Jumlah data	Jumlah dipilih		Probabilitas	
			Angina	Tidak angina	Angina	Tidak angina
	Normal	150	67	83	0,4467	0,5533
	Ringan	37	23	14	0,6216	0,3784
	Sedang	13	10	3	0,7692	0,2308

### Probabilitas Riwayat Kel. Angina

Pada attribut ini, apakah pasien memiliki riwayat keluarga penderita angin duduk, maka atribut ini dibagi menjadi 2 tingkat,  
 $P(X|Ci)$

Iya: Angina  $15/21 = 0,7143$ , Tidak Angina  $6/21 = 0,2857$

Tidak: Angina  $85/179 = 0,4749$ , Tidak Angina  $94/179 = 0,5251$

**Tabel 10** Probabilitas Rw Kel. Angina

X4	Rw kel. Angina	Jumlah data	Jumlah dipilih		Probabilitas	
			Angina	Tidak Angina	Angina	Tidak Angina
	Iya	21	15	6	0,7143	0,2857
	Tidak	179	85	94	0,4749	0,5251

### Probabilitas Riwayat Kel. Jantung

Pada attribut ini, apakah pasien memiliki riwayat keluarga penderita serangan jantung, maka attribut ini dibagi menjadi 2 tingkat  
 $P(X|Ci)$

Ada: Angina  $53/65 = 0,8154$ , Tidak Angina  $12/65 = 0,1846$

Tidak Ada: Angina  $47/135 = 0,3481$ , Tidak Angina  $88/135 = 0,6519$

**Tabel 11** Probabilitas Rw Kel Jantung

X5	Rw kel. Jantung	Jumlah data	Jumlah dipilih		Probabilitas	
			Angina	Tidak Angina	Angina	Tidak Angina
	Ada	65	53	12	0,8154	0,1846
	Tidak Ada	135	47	88	0,3481	0,6519

### Probabilitas Diabetes

Pembuatan tingkat pada variable Diabetes sesuai dengan keadaan pasien, maka dibagi menjadi 2 tingkat  
 $P(X|Ci)$

Diabet: Angina  $5/7 = 0,7143$ , Tidak Angina  $2/7 = 0,2857$

Tidak Diabet: Angina  $95/193 = 0,4922$ , Tidak Angina  $98/193 = 0,5078$

**Tabel 12** Probabilitas Diabetes

X6	Diabetes	Jumlah data	Jumlah dipilih		Probabilitas	
			Angina	Tidak Angina	Angina	Tidak Angina
	Diabet	7	5	2	0,7143	0,2857
	Tidak Diabet	193	95	98	0,4922	0,5078

## 3.7 Perhitungan Manual

Berikut ini adalah perhitungan manual dengan data testing menggunakan RapidMiner.

### A. Pendefinisian Variabel

Berdasarkan data testing di atas dapat didefinisikan data testing X sebagai berikut: X = (Xusia = "Lanjut Usia", Xmerokok = "Rendah", Xhipertensi = "Normal", Xrw kel. Angina = "Tidak", Xrw kel. Jantung = "Tidak Ada", Xdiabetes = "Tidak Diabet")

### B. Pendefinisian Probabilitas Prior P(Ci)

Hasil dari probabilitas prior berdasarkan persamaan diatas, untuk menghasilkan nilai pada label Angina (C0) sebesar 0,50 dan untuk menghasilkan nilai pada label Tidak Angina (C1) yaitu sebesar 0,50

### C. Perhitungan Probabilitas Data Testing

Berdasarkan data testing diatas dapat dilakukan sebuah klasifikasi Angina (C0) dengan ketentuan nilai masing-masing kriteria yaitu: 0,5056, 0,1688, 0,4467, 0,4749, 0,3481, 0,4922.

Kemudian nilai dari masing-masing kriteria tersebut dikali dengan  $P(X|C0) = 0,5056 \times 0,1688 \times 0,4467 \times 0,4749 \times 0,3481 \times 0,4922 = 0,0031$

Untuk menghitung klasifikasi Tidak Angina (C1) dengan ketentuan nilai masing-masing kriteria yaitu:  $0,4944, 0,8312, 0,5533, 0,5251, 0,6519, 0,5078$ .

Kemudian nilai dari masing-masing kriteria tersebut dikali dengan  $P(X|Ci) = 0,4944 \times 0,8312 \times 0,5533 \times 0,5251 \times 0,6519 \times 0,5078 = 0,0395$

#### D. Perhitungan $P(X|Ci)$

Perhitungan lanjutan untuk klasifikasi Angina (C0) adalah dengan cara mengalikan  $P(X|C0)$  dengan  $P(C0)$ :

$$\begin{aligned} P(C0|X) &= P(X|C0) \times P(C0) \\ &= 0,0031 \times 0,50 \\ &= 0,0016 \times 100\% \\ &= 0,16\% \end{aligned}$$

Kemudian untuk klasifikasi Tidak Angina (C1) adalah dengan cara mengalikan  $P(X|C1)$  dengan  $P(C1)$ :

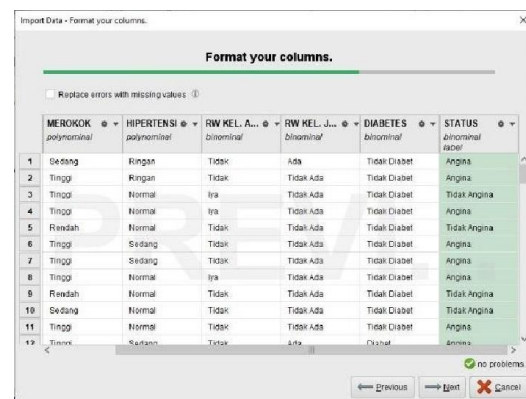
$$\begin{aligned} P(C1|X) &= P(X|C1) \times P(C1) \\ &= 0,0395 \times 0,50 \\ &= 0,0198 \times 100\% \\ &= 1,98\% \end{aligned}$$

Dari perhitungan diatas dapat dihasilkan nilai  $P(C0|X) = 0,16\%$  sedangkan nilai  $P(C1|X) = 1,98\%$ . Berdasarkan nilai tersebut dapat diambil kesimpulan bahwa  $P(C0|X) < P(C1|X)$ . maka data testing tersebut di klasifikasikan kedalam label Tidak Angina.

### 3.8 Implementasi Klasifikasi Naïve Bayes dengan RapidMiner

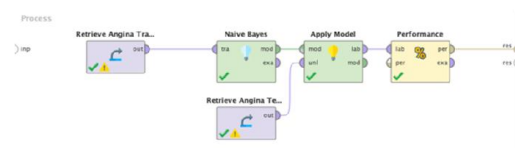
Pada pengujian ini dimaksudkan untuk mengetahui hasil prediksi dari RapidMiner, apakah hasil dari perhitungan manual sama atau tidak dengan RapidMiner.

#### A. Tahapan Pengujian



Gambar 4 Import Dataset

#### B. Proses Validasi



Gambar 5 Tahap Pengujian RapidMiner

Tahap ini dilakukan implementasi algoritma Naïve Bayes dengan tools RapidMiner, yaitu menentukan accuracy, precision recall, dan menentukan performance.

**PerformanceVector**

```

PerformanceVector:
accuracy: 87.50%
ConfusionMatrix:
True:  Tidak Angina  Angina
Tidak Angina:  16      1
Angina:  4      19
precision: 94.12% (positive class: Tidak Angina)
ConfusionMatrix:
True:  Angina  Tidak Angina
Angina:  19      4
Tidak Angina:  1      16
recall: 80.00% (positive class: Tidak Angina)
ConfusionMatrix:
True:  Angina  Tidak Angina
Angina:  19      4
Tidak Angina:  1      16
AUC (optimistic): 0.945 (positive class: Tidak Angina)
AUC: 0.935 (positive class: Tidak Angina)
AUC (pessimistic): 0.925 (positive class: Tidak Angina)
    
```

**Gambar 6** Hasil Performance Vector

**3.9 Prediksi dari accuracy Split Validation**

	true Tidak Angina	true Angina	class precision
pred. Tidak Angina	16	1	94.12%
pred. Angina	4	19	82.61%
class recall	80.00%	95.00%	

**Gambar 7** Nilai accuracy 87.50%

Nilai precision dihitung dengan cara membagi data TP (True Positive) dengan jumlah data TP (True Positive) di tambah dengan FP (False Positive).

	true Angina	true Tidak Angina	class precision
pred. Angina	19	4	82.61%
pred. Tidak Angina	1	16	94.12%
class recall	95.00%	80.00%	

**Gambar 8** Nilai Precision 94.12%

Hasil pengukuran data precision yang diperoleh mencapai 94,12%, jumlah pred Tidak Angina yang diklasifikasikan sebagai true Tidak Angina oleh classifier sebesar 16 data.

	true Angina	true Tidak Angina	class precision
pred. Angina	19	4	82.61%
pred. Tidak Angina	1	16	94.12%
class recall	95.00%	80.00%	

**Gambar 9** Nilai Recall 80.00%

Hasil pengukuran data recall dengan true Tidak Angina mencapai 80%, sedangkan untuk data recall dengan true Angina mencapai 95%

Dari hasil perhitungan sample testing sebanyak 40 data, kemudian hasil dari sistem RapidMiner didapatkan hasil prediksi nilai accuracy, precision dan recall, adalah sebagai berikut.

**Tabel 12** Jumlah data dan hasil performance

Data Training	Data Testing	Accu racy	Pre cision	Recall
80%	20%	87,50%	94,12%	80%

**IV. KESIMPULAN DAN SARAN**

**4.1 Kesimpulan**

Berdasarkan pembahasan dan pegujian menggunakan Algoritma Naïve Bayes yang dilakukan penulis. Dapat disimpulkan bahwa.

1. Teknik klasifikasi dengan metode Naïve Bayes dapat digunakan untuk penentuan status pasien apakah mengidap penyakit angin duduk atau tidak. Metode Naïve Bayes mampu mengklasifikasikandata penyakit angin duduk berdasarkan umur, merokok, riwayat penyakit keluarga, diabetes dan faktor lain yang mempengaruhi penyakit angin duduk. Perhitungan dalam Klasifikasi penyakit angin duduk Menggunakan Tools Rapidminer dengan Metode klasifikasi Naïve Bayes yang di Testing menggunakan Tools

Rapid Minner mampu memberikan hasil keputusan yang realtime. Sehingga perhitungan hasil klasifikasi dapat dengan jelas membedakan antara pasien tersebut memiliki penyakit angin duduk atau tidak.

2. Dari percobaan yang telah dilakukan, klasifikasi dapat berjalan cukup baik pada data training sebesar 160 data dan data testing 40 data. Berdasarkan hasil perhitungan, diketahui bahwa dari hasil pengujian data, memperoleh hasil presentase accuracy sebesar 87,50%, precision 94,12%, dan recall 80%. Hasil yang diperoleh sudah menunjukkan sesuatu yang baik dalam proses mendiagnosa penyakit angin duduk dalam klasifikasinya. Dengan penghitungan metode Naive Bayes mampu memberikan penjabaran secara signifikan dengan nilai akurasi yang baik.

#### 4.2 Saran

Dari penelitian yang dilakukan, penulis dapat memberikan saran untuk kedepannya lebih baik maka yang dapat disarankan adalah sebagai berikut:

1. Perlu adanya penelitian lebih lanjut dengan melakukan pengujian dengan algoritma lain atau dilakukan komparasi algoritma seperti C4.5, K-NN, Neural Network dan sebagainya. Agar memperoleh perbandingan tingkat akurasi yang paling akurat dalam membuat keputusan prediksi diagnosa penyakit angin duduk.
2. Diharapkan juga untuk menambahkan atribut yang lebih spesifik dan lebih banyak dalam menentukan klasifikasi penyakit angin duduk

#### V. DAFTAR PUSTAKA

- [1] F. R. Tsani, "Hubungan Antara Faktor Lingkungan Dan Perilaku Dengan Kejadian Penyakit Jantung Koroner (Studi Kasus di Rumah Sakit X Kota Semarang)," *Unnes Journal of Public Health*, vol. 2, no. 3, 2013.
- [2] M. Sabransyah, Y. N. Nasution and F. D. T. Amijaya, "Aplikasi Metode Naive Bayes dalam Prediksi Risiko Penyakit Jantung," *Jurnal Eksponensial*, vol. 8, no. 2, pp. 111-118, 2017.
- [3] Fatmawati, "Perbandingan Algoritma Klasifikasi Data Mining Model C4.5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa," *Jurnal Techno Nusa Mandiri*, vol. 13, no. 1, pp. 50-59, 2016.
- [4] N. A. Windiastuti, S. Santosa and C. Supriyanto, "Algoritma Klasifikasi data mining NAIIVE BAYES berbasis Particle Swarm Optimization untuk deteksi penyakit jantung," *Pseudocode*, vol. 1, no. 1, pp. 11-14, 2014.
- [5] D. Novianti and C. Sitasi, "Implementasi Metode Naïve Bayes Pada Dataset Hepatitis Menggunakan Rapid Miner," *J. Komput dan Inform. Akad. Bina Saran Inform 21.1*, vol. 1, no. 1, pp. 49-54, 2019.

- [6] W. R. Dewi, "Tribun Jogja Health," *Jangan Anggap Sepele Sakit Angin Duduk*, p. 13, 26 April 2015.
- [7] A. W. Sudoyo, *Buku Ajar Ilmu Penyakit Dalam*, Jakarta: Interna Publishing, 2009
- [8] Brown CT, "Penyakit Aterosklerotik Koroner," in *Wilson LM Patofisiologi*, Jakarta, Penerbit Buku Kedokteran EGC, 2006, p. 583.
- [9] M. Arif, *Asuhan Keperawatan Klien dengan Gangguan Sistem Kardiovaskular dan Hematologi*, Jakarta: Salemba Medika, 2009.
- [10] A. Zumantara, "Makalah Angina Pektoris New," Scribd, 21 April 2014. [Online]. Available: <https://www.scribd.com/document/219349171/Makalah-Angina-Pektoris-New>. [Accessed 8 September 2020].
- [11] V. Vijayan and A. Ravikumar, "Study of Data Mining Algorithms for Prediction and Diagnosis of Diabetes Mellitus," *International journal of computer applications*, vol. 95, no. 17, pp. 12-16, 2014.
- [12] D. S. W, "Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 dan NAIVE BAYES untuk Prediksi Penyakit Hepatitis," *None*, vol. 13, no. 1, pp. 76-84, 2017.
- [13] Y. P. A and M. E. M, "Optimasi Stemming Porter KBBI dan Cross Validation Naïve Bayes untuk Klasifikasi Topik Soal UN Bahasa Indonesia," *J. Ilm. Komputasi*, vol. 17, no. 4, 2018.
- [14] R. Posumah, "Data WHO: Lebih dari 17 Juta Orang di Dunia Meninggal Karena Serangan Jantung, Kenali Gejala Awalnya," *Tribun Manado*, 18 Juli 2019. [Online]. Available: <https://manado.tribunnews.com/2019/07/18/data-who-lebih-dari-17-juta-orang-di-dunia-meninggal-karena-serangan-jantung-kenali-gejala-awalnya?page=2>. [Accessed 9 September 2020].
- [15] S. A. A, S. M and S. O, "Klasifikasi Stroke Berdasarkan Kelainan Patol Dengan Vector Quantization," 248-523-1-Pb, vol. 8, no. 2, pp. 117-122, 2014.