

## **BAB III**

### **METODELOGI PENELITIAN**

#### **3.1 Perancangan Penelitian**

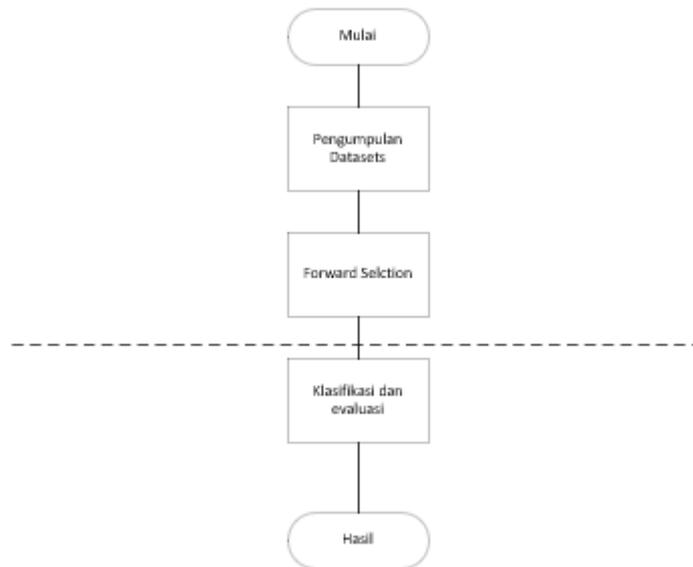
Adapun perancangan pada penelitian telah disusun sehingga alur dari proses penelitian ini sesuai dengan yang telah ditentukan.

Metode penelitian kuantitatif digunakan untuk meneliti pada populasi atau sample tertentu, pengumpulan data menggunakan instrumen penelitian, analisis data bersifat kuantitatif/statistik dengan tujuan untuk menguji hipotesis yang telah ditentukan. Metode penelitian kuantitatif disebut juga dengan metode discovery karena dengan metode ini dapat ditemukan dan dikembangkan berbagai hal baru sehingga memunculkan ilmu pengetahuan baru [2]

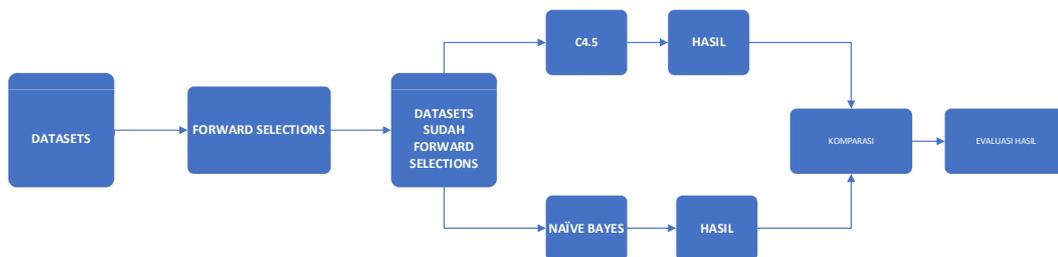
Metode penelitian kuantitatif dapat dibagi menjadi tiga sub kategori yaitu inferentia, experimental, dan simulasi. Jenis penelitian yang dilakukan dalam penelitian ini adalah jenis penelitian eksperimen. Metode ini menguji kebenaran sebuah hipotesis dengan statistik dan menghubungkannya dengan masalah penelitian. Metode penelitian eksperimen dapat diartikan sebagai metode penelitian yang digunakan untuk mencari pengaruh perlakuan tertentu terhadap yang lain dalam kondisi yang terkendalikan [18]

Jenis penelitian eksperimen dibagi menjadi dua kategori, yaitu eksperimen absolut dan eksperimen komparatif. Eksperimen absolut mengarah kepada dampak yang dihasilkan dari eksperimen, misalnya pengaruh honor dosen terhadap kinerja. Sedangkan eksperimen komparatif yaitu membandingkan dua objek yang berbeda, misalnya membandingkan dua algoritma yang berbeda dengan melihat hasil statistik masing-masing mana yang lebih baik. Pada penelitian ini, jenis penelitian yang diambil adalah Eksperimen Komparatif.

### 3.2 Tahapan Penelitian



Gambar 3.1 Tahapan Penelitian



Gambar 3.2 Tahapan Pengolahan Data [17].

#### 1. Pengumpulan Data

Banyak penelitian yang telah dilakukan untuk memprediksi penyakit gagal jantung, namun belum diketahui algoritma mana yang paling akurat. Oleh karena itu dalam penelitian ini akan dilakukan komparasi algoritma Naive Bayes dan C4.5 untuk mengetahui algoritma mana yang lebih akurat dalam memprediksi penyakit gagal jantung. Sampel Penelitian Sampel dari Penelitian ini adalah data

profil Prediksi penyakit gagal jantung, data tersebut yang bersifat publik yang didapatkan dari Kaggle.com(<https://www.kaggle.com/datasets/fedesoriano/heart-failureprediction>) Data tersebut berisi 918 Observasi dengan 12 Atribut.

## **2. Pengolahan Awal Data**

Pada tahap ini data yang masih berupa nilai numerik dan juga continue ditransformasikan kedalam bentuk kategorikal dan dibuat skala atau interval sehingga menghasilkan range yang lebih kecil sebagai bahan pembelajaran algoritma Naive Bayes dan C4.5 akan lebih mudah di klasifikasikan dengan menggunakan Rapid Miner sehingga memudahkan dalam memproses klasifikasi pada dataset yang telah tersedia sebelumnya. Ada 12 atribut yang digunakan dalam mendiagnosa penyakit jantung, yaitu: *Age*, *Sex*, *Chest paint type*, *resting blood pressure*, *Serum cholesterol* dalam mg/dl, *fasting blood sugar*>120 mg/dl, *resting electrocardiographic result*, *Maximum heart rate achieved*, *Exercise induced angina*, *oldpeak*, *The slope of the peak exercise ST segment*, *Number of major vessels*, dan *Thal*. Penjelasan mengenai atribut tersebut diantaranya adalah sebagai berikut:

### **1. Usia/Age**

Penuaan dapat menyebabkan penurunan elastisitas dan pemenuhan aorta serta kekakuan arteri besar. Hal ini menyebabkan tekanan sistolik lebih tinggi dan meningkatkan impedansi penyebaran darah dari ventrikel kiri melalui sistem arteri dan pengiriman darah ke organ dan jaringan. Pembesaran ringan ventrikel kiri juga terjadi [8]. Menurut Dra. Ny. Jos Masdani seorang psikolog dari universitas Indonesia mengatakan bahwa usia lansia merupakan kelanjutan dari usia dewasa. [19]. Kedewasaan dapat dibagi menjadi empat bagian, yaitu:

- 1) Fase Iuventus (20 Tahun – 40 Tahun)
- 2) Fase Verilitas (40 Tahun-50 Tahun)
- 3) Fase Prasenium (55 Tahun-65 Tahun)
- 4) Fase Senium (Usia 65 Tahun hingga tutup usia)

## 2. Sex

Sebagian besar wanita terlindung dari resiko serangan jantung hingga usia 48 tahun karena status hormonal, kecuali yang memiliki kolesterol darah atau menderita diabetes. Gejala serangan jantung pada wanita dan pria tidak berbeda secara signifikan sehingga dimungkinkan lelaki ataupun perempuan dapat menderita penyakit jantung [19].

## 3. Nyeri Dada/ *Chest Pain Type (CPT)*

*Chest pain* atau angina atau disebut juga *angine pectoris* yaitu nyeri dada atau ketidaknyamanan yang terjadi ketika jantung tidak mendapatkan darah dan oksigen yang dibutuhkan pada angina, kebutuhan akan peningkatan darah tidak terpenuhi dalam jangka waktu singkat. Angina dibagi menjadi empat, yaitu: *Asymptomatic*, *atypical angina*, *non anginal pain*, dan *typical angina*. *Asymptomatic* berarti pasien tanpa ada gejala nyeri dada, pada *typical angina* rasa tidak nyaman atau sakit terletak *subternally* dan sering menjalar yang dipicu oleh aktivitas atau emosi dan dapat hilang dengan beristirahat. *Atypical angina* terletak di dada kiri seperti ditusuk-tusuk, kadang-kadang terbatas pada daerah kecil dan terjadi berulang-ulang baik dalam jangka waktu yang singkat maupun berjam-jam [20].

## 4. Tekanan Darah/*Blood Pressure*

Tekanan darah atau "*Blood Pressure*" adalah kekuatan darah mendorong dinding arteri saat jantung memompa darah, jika tekanan darah terus meningkat dari waktu ke waktu dapat merusak tubuh dalam banyak hal. Tekanan darah dinyatakan dengan ukuran sistolik dan diastolik, tekanan sistolik mengacu pada tekanan darah ketika jantung berdetak atau pada saat terjadi kontraksi otot jantung. Sedangkan tekanan diastolik yaitu tekanan darah ketika jantung beristirahat atau tidak sedang berkontraksi [19]. Klasifikasi tekanan darah untuk usia 18 tahun lebih seperti yang diberikan oleh *the Seventh Report of the Joint National Committee on Prevention Detection, Evaluation, and Treatment of High Blood Pressure* adalah sebagai berikut [9].

1) Normal: Tekanan darah systolic <120; Diastolic < 80 mmHG

- 2) Prehipertensi: Tekanan darah systolic 120-139; Diastolic 80-89
- 3) Hipertensi Level 1: Tekanan darah systolic 140-159; diastolic 90-99
- 4) Hipertensi level 2: Tekanan darah systolic >160; diastolic > 100

### **5. Kolesterol/Serum cholesterol**

Kolesterol adalah lemak atau yang seperti zat lemak yang larut dalam air. Kolesterol diserap oleh usus atau dilepaskan dari hati ke aliran darah. Kolesterol tidak bergerak secara bebas dalam larutan, tetapi melekat ke pembawa protein yaitu lipoprotein. Kepadatan dan ukuran lipoprotein bervariasi, semakin kecil ukurannya semakin tinggi kepadatannya. Kolesterol dapat diangkut dalam lipoprotein kepadatan rendah, sehingga ada istilah kolesterol lipoprotein berkepadatan rendah (LDL). Ada juga kolesterol lipoprotein berkepadatan tinggi (HDL). LDL kolesterol disebut juga lemak jahat, semakin tinggi LDL kolesterol dalam darah, semakin tinggi resiko penyakit jantung koroner. Pengukuran kolesterol darah memberikan total kolesterol darah yaitu kolesterol LDL ditambah kolesterol HDL, Total kolesterol yang diperlukan adalah <200, sedangkan batas normal kolesterol adalah 200-239. Jika total kolesterol melebihi 239 maka dikatakan kolesterol tinggi [2].

### **6. Gula Darah/Fasting blood sugar >120 mg/dl**

Fasting Blood Sugar atau gula darah puasa yaitu level glukosa atau kadar glukosa dalam darah setelah berpuasa sekitar 8-10 jam yang dimaksudkan agar hasilnya maksimal dimana perut dalam keadaan kosong, jadi hasil pemeriksaan tidak terpengaruhi oleh zat-zat yang dimakan. Jika hasil pemeriksaan menunjukkan kadar glukosa setelah puasa 126 mg/dl (mmol/l) maka dikategorikan sebagai diabetes melitus, sebaliknya jika hasil pemeriksaan menunjukkan kadar glukosa puasa 110 mg/dl (mmol/l) maka dianggap normal, dan akan dianggap terganggu apabila berkisar antara 110-125 mg/dl [2].

### **7. Resting Electrocardiographic Result**

Electrocardiographic atau EKG merupakan tes yang paling sering digunakan untuk mendiagnosa angina yang dapat mendiagnosa infark miokard lebih dini. EKG akan menunjukkan nilai abnormal sebelum kerusakan pada otot jantung

ditunjukkan oleh tes enzim jantung (toponin dan CK-MB). Elektrokardiogram mendapatkan impuls listrik jantung melalui kulit dada, diagnosis EKG pada pasien dengan nyeri dada membedakan pasien menjadi dua kelompok populasi besar yaitu, elevasi segmen ST dan elevasi segmen non ST. Tes EKG sangat penting karena merupakan petunjuk adanya hipertrofi atau pembesaran atrium kiri dan peningkatan atau tekanan volume atrium kiri. Pembesaran pada jantung, terutama otot ventrikel kiri mudah terdeteksi oleh EKG, pasien dengan jantung yang disebabkan oleh hipertrofi kardiomiopati hipertrofik biasanya memiliki EKG normal [2].

### **8. Maximum Heart Rate Achieved**

Denyut jantung dikendalikan oleh otak dan saraf, saraf yang merangsang jantung untuk berdetak lebih cepat disebut saraf simpatik, saraf tersebut dirangsang oleh otak. Sedangkan saraf yang menyebabkan jantung berdetak lebih lambat dinamakan saraf vagus. Denyut jantung maksimal diperoleh dengan mengurangi usia individu dengan 220, rata-rata denyut jantung normal untuk usia 70 tahun yang sedang beristirahat 60 kali per menit, denyut jantung maksimal adalah 150 denyut per menit. Pada usia 20 denyut jantung tertinggi untuk jantung normal dapat mencapai antara 200-220 denyut permenit yang disebut denyut jantung maksimum yang dicapai. Dokter menyarankan bahwa denyut jantung tidak boleh melebihi 85% dari denyut jantung maksimal, yaitu sekitar 170 denyut permenit. [2]

### **9. Exercise Induced Angina**

Sebuah latihan uji tekanan menggunakan treadmill atau sepeda dibawah pengawasan dokter, biasanya level latihan yang aman adalah yang latihan yang hanya menyebabkan ketidaknyamanan atau sesak nafas ringan. Sebuah uji olahraga treadmill menggunakan protokol bruce digunakan diseluruh dunia untuk mendeteksi iskemia miokard yang dipicu oleh olahraga [2]

### **10. Oldpeak**

Oldpeak atau disebut juga depresi yang disebabkan oleh latihan relatif, depresi atau stres dapat menyebabkan nyeri dan kerusakan pada jantung. Ketika arteri

koroner menyempit karena plak maka akan menyebabkan terjadinya nyeri dada, dan nyeri dada diperparah oleh aktivitas seperti berjalan mendaki bukit, jika penderita menghadapi gangguan emosi mendadak maka nyeri dada pada saat tidak beraktivitas/istirahat dapat terjadi. Stress yang berhubungan dengan kegiatan sederhana sehari-hari dapat menurunkan suplai darah ke otot jantung pada pasien dengan penyakit jantung koroner, begitu juga dengan olahraga yang berlebihan dapat merusak sistem kardiovaskular. Pemantauan denyut jantung selama periode pemulihan secara substansial menambah nilai pengujian latihan stress [2]

### 11. The slope of the peak exercise ST segment

Tanda dari elektrokardiografi latihan miokard iskemia diinduksi adalah depresi dari segment ST. Sebuah depresi segment ST horizontal atau downsloping sama atau lebih besar dari 1 mm dalam dua atau lebih memicu diagnostik. Poin diagnostik mencakup depresi segment ST lebih besar dari 1 mm atau lebih dan menunda memperlambat denyut jantung [8]

### 12. HeartDisease

Yang dihasilkan dari pembobotan klasifikasi dari algoritma naïve bayes dan algoritma C4.5

Tabel 3. 1 Keterangan Atribut Dataset

Atribut	Nilai	Kategori
<i>Age</i>	<40	Inventus
	40-54	Verilitas
	55-64	Prasenum
	>=65	Senium
<i>Sex</i>	1	Laki-Laki
	0	Perempuan
	1	<i>Typical Angina</i>

<i>Chest Paint Type</i>	2	<i>Atypical Angina</i>
	3	<i>Non Angina Pain</i>
	4	<i>Asympitomatic</i>
	<120	Normal
<i>Resting Blood Pressure</i>	120-139	Prehipertensi
	140-159	Hipertensi Level 1
	>159	Hipertensi Level 2
<i>Serum cholestoral dalam mg/dl</i>	<200	Normal
	200-239	Batas normal tinggi
	>239	Tinggi
<i>Fasting blood sugar &gt;120 mg/dl</i>	0	Tidak
	1	Ya
<i>Resting electrocardiographic result</i>	0	Normal
	1	having ST-T wave abnormality
	2	left ventricular hypertrophy
<i>HeartDisease</i>	0	Tidak
	1	Ya
<i>Excercise Induced</i>	0	Tidak

<i>Angina</i>	1	Ya
<i>Maximum Heart Rate Achieved (Thalac)</i>	#NAME?	Normal
	<>220-usia	Tidak Normal
<i>Old Peak</i>	<1	0
	>=1 dan <2	1
	>=2 dan <3	2
	>=3 dan <4	3
	>=4	4
<i>The Slope of the peak exercise ST segment</i>	1	<i>Unsloping</i>
	2	<i>Flat</i>
	3	<i>Downsloping</i>

### 3.3 Forward Selection

merupakan salah satu metode permodelan (pembangunan model linier) untuk menemukan kombinasi peubah yang “terbaik” dari suatu gugus peubah. Dalam prosedur forward Selection, sekiranya variabel masuk kedalam persamaan maka tidak bisa dihilangkan. Selain itu, forward selection dapat berarti memasukan variabel bebas yang memiliki korelasi yang paling erat dengan variabel tak bebasnya (variabel yang paling potensial untuk memiliki hubungan linier dengan Y)[21] F-measure dihitung dengan menggabungkan precision dan recall menjadi satu nilai dengan menggunakan rumus berikut :

$$F\text{-measure} = 2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall})$$

precision adalah rasio dari jumlah prediksi yang benar dalam kelas positif terhadap jumlah total prediksi dalam kelas positif.

recall adalah rasio dari jumlah prediksi yang benar dalam kelas positif terhadap jumlah total item dalam kelas positif.

Nilai F-measure berkisar antara 0 dan 1, di mana nilai yang lebih tinggi menunjukkan kinerja yang lebih baik dari sistem klasifikasi.

### 3.4 Pengujian Algoritma Decision Tree C4.5

Pengujian data yang digunakan dalam algoritma C4.5 adalah dataset prediksi gagal jantung yang didapatkan dari Kaggle. Data yang digunakan dalam pengujian ini dapat dilihat pada tabel berikut:

**Tabel 3.2** Dataset Prediksi Gagal Jantung

Row No.	Weight	predicti...	confide...	confide...	Se...	Se...	...	ChestP...	ChestP...	ChestP...	Fasting...	Fasting...	Resting...	Resting...	Res...	E...	Ex...	S...	S
1	0	0	0.962	0.038	1	0	1	0	0	0	1	0	1	0	0	1	0	1	0
2	1	1	0	1	0	1	0	1	0	0	1	0	1	0	0	1	0	0	1
3	0	0	0.962	0.038	1	0	1	0	0	0	1	0	0	1	0	1	0	1	0
4	1	1	0.250	0.750	0	1	0	0	1	0	1	0	1	0	0	0	1	0	1
5	0	0	0.962	0.038	1	0	0	1	0	0	1	0	1	0	0	1	0	1	0
6	0	0	0.962	0.038	1	0	0	1	0	0	1	0	1	0	0	1	0	1	0
7	0	0	0.962	0.038	0	1	1	0	0	0	1	0	1	0	0	1	0	1	0
8	0	0	0.962	0.038	1	0	1	0	0	0	1	0	1	0	0	1	0	1	0
9	1	1	0.073	0.927	1	0	0	0	1	0	1	0	1	0	0	0	1	0	1
10	0	0	0.962	0.038	0	1	1	0	0	0	1	0	1	0	0	1	0	1	0
11	0	0	0.962	0.038	0	1	0	1	0	0	1	0	1	0	0	1	0	1	0
12	1	1	0.073	0.927	1	0	1	0	0	0	1	0	0	1	0	0	1	0	1
13	0	0	0.962	0.038	1	0	1	0	0	0	1	0	1	0	0	1	0	1	0

### 3.5 Pengujian Algoritma Naïve Bayes

Pengujian data yang digunakan dalam algoritma Naïve Bayes adalah dataset prediksi gagal jantung yang didapatkan dari Kaggle. Data yang digunakan dalam pengujian ini dapat dilihat pada tabel berikut.

**Tabel 3.3** Dataset Prediksi Gagal Jantung

Row No.	Weight	predicti...	confide...	confide...	Se...	Se...	...	ChestP...	ChestP...	ChestP...	Fasting...	Fasting...	Resting...	Resting...	Res...	E...	Ex...	S...	S
1	0	0	0.962	0.038	1	0	1	0	0	0	1	0	1	0	0	1	0	1	0
2	1	1	0	1	0	1	0	1	0	0	1	0	1	0	0	1	0	0	1
3	0	0	0.962	0.038	1	0	1	0	0	0	1	0	0	1	0	1	0	1	0
4	1	1	0.250	0.750	0	1	0	0	1	0	1	0	1	0	0	0	1	0	1
5	0	0	0.962	0.038	1	0	0	1	0	0	1	0	1	0	0	1	0	1	0
6	0	0	0.962	0.038	1	0	0	1	0	0	1	0	1	0	0	1	0	1	0
7	0	0	0.962	0.038	0	1	1	0	0	0	1	0	1	0	0	1	0	1	0
8	0	0	0.962	0.038	1	0	1	0	0	0	1	0	1	0	0	1	0	1	0
9	1	1	0.073	0.927	1	0	0	0	1	0	1	0	1	0	0	0	1	0	1
10	0	0	0.962	0.038	0	1	1	0	0	0	1	0	1	0	0	1	0	1	0
11	0	0	0.962	0.038	0	1	0	1	0	0	1	0	1	0	0	1	0	1	0
12	1	1	0.073	0.927	1	0	1	0	0	0	1	0	0	1	0	0	1	0	1
13	0	0	0.962	0.038	1	0	1	0	0	0	1	0	1	0	0	1	0	1	0

Table 3.3 Perhitungan Entropy dan Gaint

			YA	TIDAK		
		<b>Jumlah (S)</b>	<b>A (Si)</b>	<b>B (Si)</b>	<b>Entropy</b>	<b>Gain</b>
<b>Total</b>		918	508	410	0,991763566	
<b>Age</b>						0,007268211
	Kurang dari Enam Puluh Lima Setengah	849	458	391	0,995502922	
	Lebih dari Enam Puluh Lima Setengah	69	50	19	0,849054424	
<b>Sex</b>						0,06849493
	Male	725	458	267	0,949339049	

	Female	193	50	143	0,825335739	
<b>Chest Pain Type</b>						0,225043645
	ASY	496	392	104	0,740865686	
	ATA	173	24	149	0,580898693	
	NAP	203	72	131	0,938178216	
	TA	46	20	26	0,987692509	
<b>Resting BP</b>						0,008799533
	Kurang dari 140	591	303	288	0,999535272	
	Lebih dari 140	327	205	122	0,953014178	
<b>Colesterol</b>						0,0000417
	Kurang dari 300	813	451	362	0,991338079	
	Lebih dari 300	105	57	48	0,994693795	

<b>Fasting BS</b>						0,054880174
	ya	214	170	44	0,733002265	
	tidak	704	338	366	0,998858621	
<b>RestingECG</b>						0,011699201
	LVH	187	101	86	0,995353665	
	Normal	552	286	266	0,999052844	
	ST	178	120	58	0,910622368	
<b>Max HR2</b>						0,028686481
	Kurang dari 130	361	243	118	0,911696251	
	Lebih dari 130	556	264	292	0,998169816	
<b>ExerciseAngina</b>						0,013870918
	Y	371	235	136	0,948007374	

	N	546	272	274	0,999990321	
<b>ST_Slope</b>						0,013681145
	Down	63	40	23	0,946818832	
	Flat	460	279	181	0,967007548	
	Up	394	188	206	0,998493919	

### 3.6. Perhitungan Algoritma C4.5

$$\text{Entropy Total} = \left(-\frac{508}{918} \log_2 \left(\frac{508}{918}\right)\right) + \left(-\frac{410}{918} \log_2 \left(\frac{410}{918}\right)\right) = 0.991763566$$

$$(">60") = \left(-\frac{50}{69} \log_2 \left(\frac{50}{69}\right)\right) + \left(-\frac{19}{69} \log_2 \left(\frac{19}{69}\right)\right) = 0.849054424$$

$$("<=60") = \left(-\frac{458}{849} \log_2 \left(\frac{458}{849}\right)\right) + \left(-\frac{391}{849} \log_2 \left(\frac{391}{849}\right)\right) = 0.99550922$$

### 3.7. Perhitungan Gain

$$= -\left(\frac{849}{918}\right) \times 0.99550922 + \left(\frac{69}{918}\right) \times 0.849054424 = 0.007268211$$

### 3.8. Perhitungan Split Info

$$\left(-\frac{69}{918} \log_2 \left(\frac{69}{918}\right)\right) + \left(-\frac{849}{918} \log_2 \left(\frac{849}{918}\right)\right) = 0.38378005$$

### 3.9. Perhitungan Gain Ratio

$$= \frac{0.007268211}{0.38378005} = 0.01893848$$

Keterangan:

Untuk variable yang lain dihitung seperti diatas

### 3.10. Perhitungan algoritma Nive Bayes

$$P(Y = \text{Positif}) = \frac{\text{jumlah positif}}{\text{Jumlah Total}} = \frac{447}{918} = 0.4869281046$$

$$P(Y = \text{negatif}) = \frac{\text{jumlah negatif}}{\text{Jumlah Total}} = \frac{344}{918} = 0.3747276688$$

Jenis Kelamin

$$P(Y = Male) = \frac{\text{jumlah positif}}{\text{Jumlah Total}} = \frac{0.902}{918} = 0.0009825708$$

$$P(Y = Male) = \frac{\text{jumlah negatif}}{\text{Jumlah Total}} = \frac{0.651}{918} = 0.0007091503$$

$$P(Y = Female) = \frac{\text{jumlah positif}}{\text{Jumlah Total}} = \frac{0.908}{918} = 0.0009891068$$

$$P(Y = Female) = \frac{\text{jumlah negatif}}{\text{Jumlah Total}} = \frac{0.349}{918} = 0.0003801743$$

= P(X| remark = Positif)

$$= 0.0009825708 \times 0.0009891068$$

$$= 0.00097186746$$

= P(X| remark = Negatif)

$$= 0.0007091503 \times 0.0003801743$$

$$= 0.00026960072$$

\***Keterangan:** Untuk variable yang lain dihitung seperti diatas

### 3.6 Penerapan Algoritma Naive Bayes dan C4.5

Pada tahap ini memilih dan menerapkan teknik yang tepat untuk mendapatkan hasil yang optimal, pada penelitian ini algoritma yang digunakan adalah algoritma Naive Bayes dan C4.5 dilakukan dengan beberapa percobaan dengan menggunakan operator split data dan tahap ini dilakukan dengan bantuan *tools software Rapid Miner*.