

## **BAB V**

### **ANALISIS DAN INTERPRETASI**

#### **5.1. ANALISIS DATA**

Pada sub bab ini dipaparkan analisis data menggunakan *software datamining* dan interpretasi hasil analisis persiapan proses data.

##### **5.1.1 Data Penelitian**

Sebelum proses analisis menggunakan *software datamining*, pertama-tama dilakukan pra-proses data. Pra-proses data pada penelitian ini meliputi seleksi data dan transformasi data excel ke dalam bentuk .CSV ,Data penelitian yang digunakan dalam penelitian ini merupakan semua data pasien yang telah direkap berdasarkan data yang ada pada UPTD Puskesmas Pakuan Baru, Adapun tampilan data tersebut dapat dilihat pada tabel 5.1 :

Tabel 5.1 Contoh Data Penelitian

No	Nama Pasien	JK	Alamat	K U	Tanggal Kunjungan	Demam	Muntah	Ruam Merah	Sakit Kepala	Musim	Jentik	Demam Berdarah
1	Abdullah	Lk	Wijaya Pura	Balita	09/01/2018	iya	tidak	iya	tidak	hujan	positif	positif
2	Adeva	Pr	Wijaya Pura	Balita	09/01/2018	iya	tidak	iya	tidak	hujan	positif	positif
3	Afif	Lk	Wijaya Pura	Balita	26/02/2016	iya	tidak	iya	tidak	hujan	positif	positif
4	Afifah	Pr	Wijaya Pura	Balita	26/02/2016	iya	tidak	iya	tidak	hujan	positif	positif
5	Agil Panca Purawan	Lk	Tambak Sari	Remaja	28/02/2019	iya	tidak	iya	tidak	hujan	positif	positif
...	...	...	...	...	...	...	...	...	...	...	...	...
168	Yusika Dwi Andra	Pr	Pakuan Baru	Balita	21/01/2015	iya	tidak	iya	tidak	hujan	negatif	positif
169	Zafira Hadijah Azzahra	Pr	Tambak Sari	Balita	08/03/2019	iya	tidak	iya	tidak	hujan	positif	positif
170	Zahwit	Lk	Wijaya Pura	Kanak-Kanak	26/02/2016	iya	tidak	iya	tidak	hujan	negatif	positif
171	Zaki Verdia	Lk	Tambak Sari	Kanak-Kanak	03/02/2016	iya	tidak	iya	tidak	hujan	negatif	positif
172	Zifaha Hadijah Azzahra	Pr	Tambak Sari	Balita	11/03/2019	iya	tidak	iya	tidak	hujan	positif	positif

### 5.1.2 Seleksi Data Penelitian

Seleksi data penelitian merupakan tahapan penyeleksian data. Pada tahap ini data yang tidak lengkap, atau data yang sama dan atribut yang tidak digunakan akan dihapus sehingga didapatkan data yang telah difokuskan hanya pada atribut data yang akan digunakan, atribut yang akan digunakan dalam perhitungan *Naive Bayes* adalah Jenis Kelamin, Alamat, Kategori Usia, Demam, Muntah, Ruam Merah, Sakit Kepala, Musim, Jentik, dan Demam Berdarah. Adapun data tersebut dapat dilihat pada tabel 5.2 :

**Tabel 5.2 Contoh Data yang Telah Diseleksi**

JK	Alamat	K U	Demam	Muntah	Ruam Merah	Sakit Kepala	Musim	Jentik	Demam Berdarah
Pr	Pakuan Baru	Remaja	iya	tidak	iya	tidak	hujan	positif	positif
Lk	Wijaya Pura	Dewasa	iya	tidak	iya	tidak	hujan	positif	positif
Lk	Wijaya Pura	Balita	iya	tidak	iya	tidak	hujan	negatif	positif
Pr	Wijaya Pura	Dewasa	tidak	tidak	tidak	iya	kemarau	positif	negatif
Lk	Tambak Sari	Remaja	iya	tidak	iya	tidak	kemarau	positif	positif

### 5.1.3 Konversi Hasil Data

Data yang ada pada *excel* kemudian dikonversi kedalam bentuk data *.csv* ini dilakukan agar data dapat diolah dengan menggunakan *tools weka*, adapun format data *.csv* tersebut sebagai berikut :

```
JK;Alamat;KU;Demam;Muntah;Ruam Merah;Sakit Kepala;Musim;
Jentik;Demam Berdarah
Pr;PakuanBaru;Remaja;iya;tidak;iya;tidak;hujan;positif;positif
```

Lk;WijayaPura;Dewasa;iya;tidak;iya;tidak;hujan;positif;positif

Lk;WijayaPura;Balita;iya;tidak;iya;tidak;hujan;negatif;positif

Pr;WijayaPura;Dewasa;tidak;tidak;tidak;iya;kemarau;positif;negatif

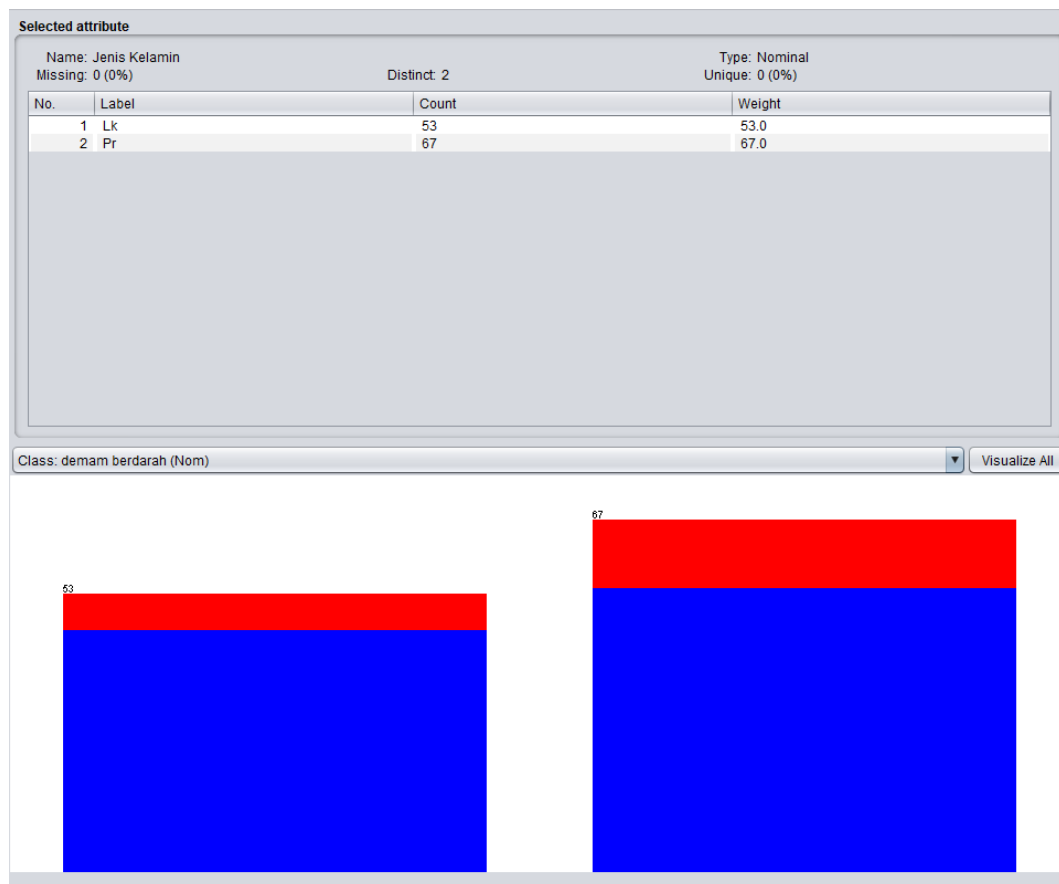
Lk;TambakSari;Remaja;iya;tidak;iya;tidak;kemarau;positif;positif

#### **5.1.4 Hasil Visualisasi Atribut Data *Training* dengan Menggunakan WEKA**

Berikut merupakan visualisasi atribut yang diambil dari data *training* dengan menggunakan tools WEKA :

1. Visualisasi Atribut Jenis Kelamin

Visualisasi untuk melihat grafik Jenis Kelamin ditampilkan pada gambar 5.1 yang menggunakan *tools* weka 3.8



**Gambar 5.1 Visualisasi Atribut Jenis Kelamin (Data Training)**

Gambar 5.1 adalah visualisasi dari atribut Jenis Kelamin yang mana dari 120 data pada *column selected Attribute* terdapat 0 (0%) *missing*. *Distinct 2 Attribute* terdapat *unique 0 (0%)* dengan jumlah data dari 53 berjenis kelamin laki-laki dan 67 berjenis kelamin perempuan.

## 2. Visualisasi Atribut Alamat

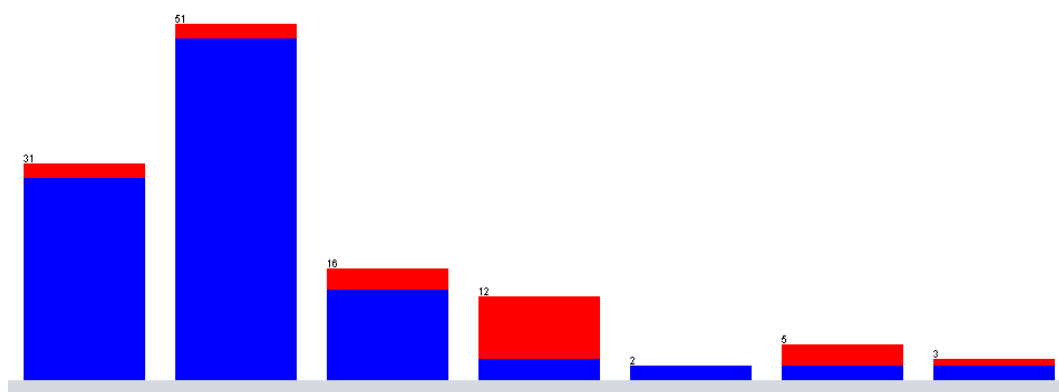
Visualisasi untuk melihat grafik alamat di tampilkan pada gambar 5.2 yang menggunakan tools weka 3.8

**Selected attribute**

Name: Alamat  
Missing: 0 (0%)  
Distinct: 7  
Type: Nominal  
Unique: 0 (0%)

No.	Label	Count	Weight
1	Wijaya Pura	31	31.0
2	Tambak Sari	51	51.0
3	Pakuan Baru	16	16.0
4	Thehok	12	12.0
5	Beringin	2	2.0
6	Sungai Asam	5	5.0
7	Talang Banjar	3	3.0

Class: demam berdarah (Nom) Visualize All



**Gambar 5.2 Visualisasi Atribut Alamat(Data Training)**

Gambar 5.2 adalah visualisasi atribut alamat, diketahui dari 120 data pada *column selected Attribute* terdapat 0 (0%) *missing*. *Distinct 7 Attribute* terdapat *unique* 0 (0%). Data pasien yang beralamat di Wijaya Pura sebanyak 31 data, Tambak Sari sebanyak 51 data, Pakuan Baru sebanyak 16 data, Thehok sebanyak 12 data, Beringin sebanyak 2 data, Sungai Asam sebanyak 5 data, Talang Banjar sebanyak 3 data.

### 3. Visualisasi Atribut Kategori Usia

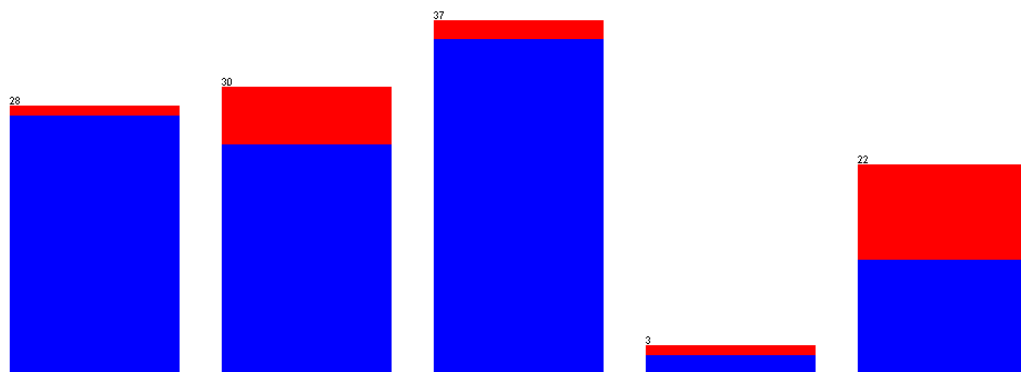
Visualisasi untuk melihat grafik kategori usia di tampilkan pada gambar 5.3 yang menggunakan tools weka 3.8

**Selected attribute**

Name: Kategori Usia  
Missing: 0 (0%)  
Distinct: 5  
Type: Nominal  
Unique: 0 (0%)

No.	Label	Count	Weight
1	Balita	28	28.0
2	Remaja	30	30.0
3	Kanak-Kanak	37	37.0
4	Lansia	3	3.0
5	Dewasa	22	22.0

Class: demam berdarah (Nom) Visualize All

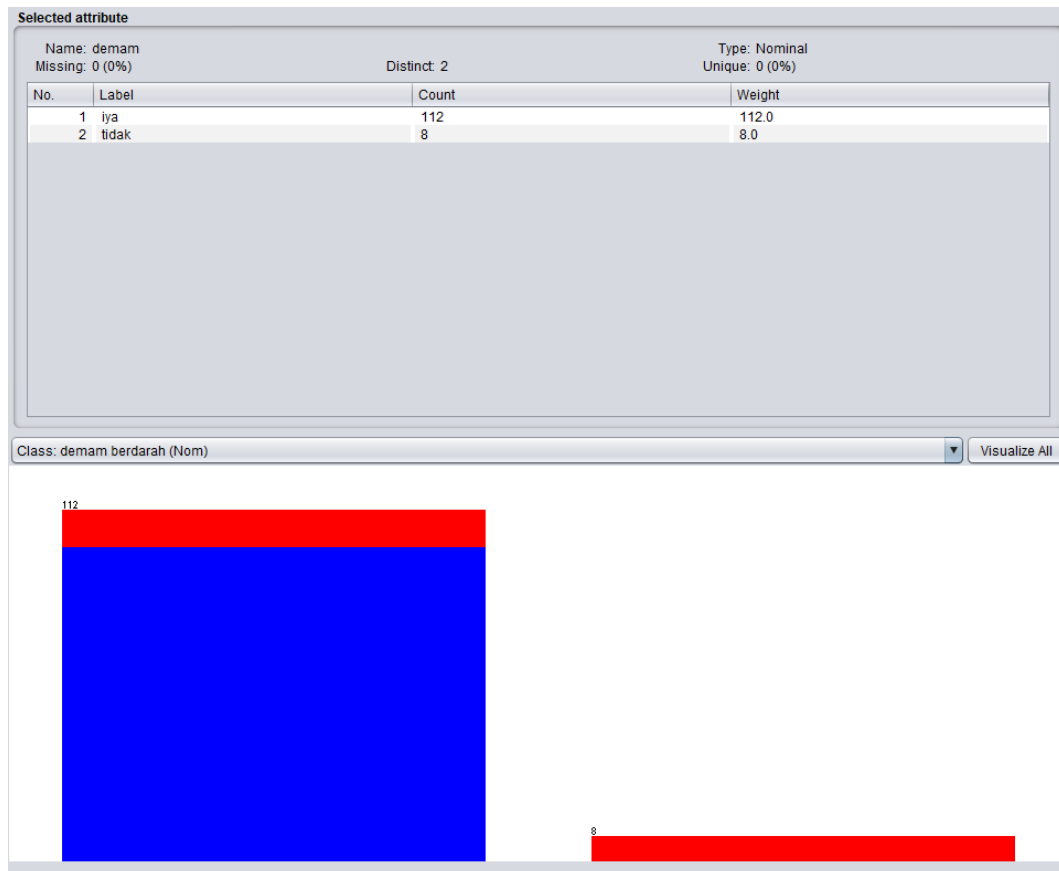


**Gambar 5.3 Visualisasi Atribut Kategori Usia(Data Training)**

Gambar 5.3 adalah visualisasi atribut kategori usia, diketahui dari 120 data pada *column selected Attribute* terdapat 0 (0%) *missing*. *Distinct 5 Attribute* terdapat *unique* 0 (0%). Data pasien yang berada di kategori umur balita (0-5thn) adalah sebanyak 28 data, Remaja(>12-25thn) sebanyak 30 data, Kanak-Kanak(>5-12thn) sebanyak 37 data, Lansia(>45thn) sebanyak 3 data, dan Dewasa(>25-45thn) sebanyak 22 data.

#### 4. Visualisasi Atribut Demam

Visualisasi untuk melihat grafik demam di tampilkan pada gambar 5.4 yang menggunakan tools weka 3.8



**Gambar 5.4 Visualisasi Atribut Demam(Data Training)**

Gambar 5.4 adalah visualisasi atribut demam, diketahui dari 120 data pada *column selected Attribute* terdapat 0 (0%) *missing*. *Distinct 2 Attribute* terdapat *unique* 0 (0%). Data pasien yang mengalami demam yaitu sebanyak 112 data dan yang tidak mengalami demam sebanyak 8 data.

##### 5. Visualisasi Atribut Muntah

Visualisasi untuk melihat grafik muntah di ditampilkan pada gambar 5.5 yang menggunakan tools weka 3.8

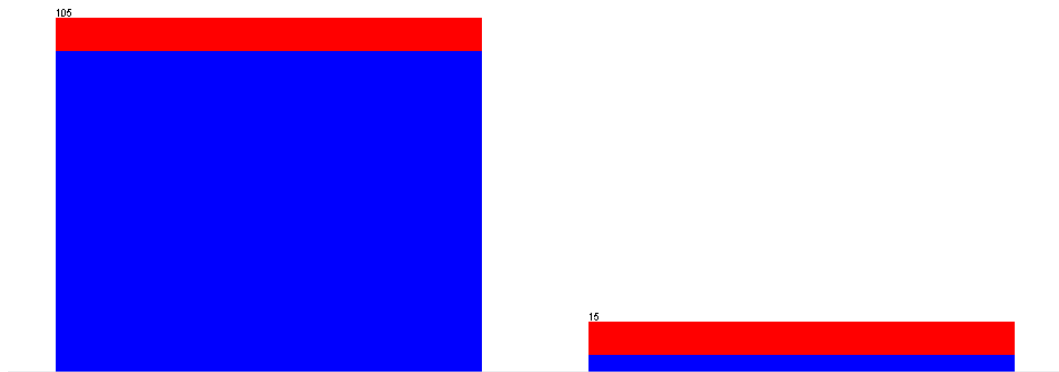


Selected attribute

Name: muntah  
Missing: 0 (0%)  
Distinct: 2  
Type: Nominal  
Unique: 0 (0%)

No.	Label	Count	Weight
1	tidak	105	105.0
2	iya	15	15.0

Class: demam berdarah (Nom) Visualize All

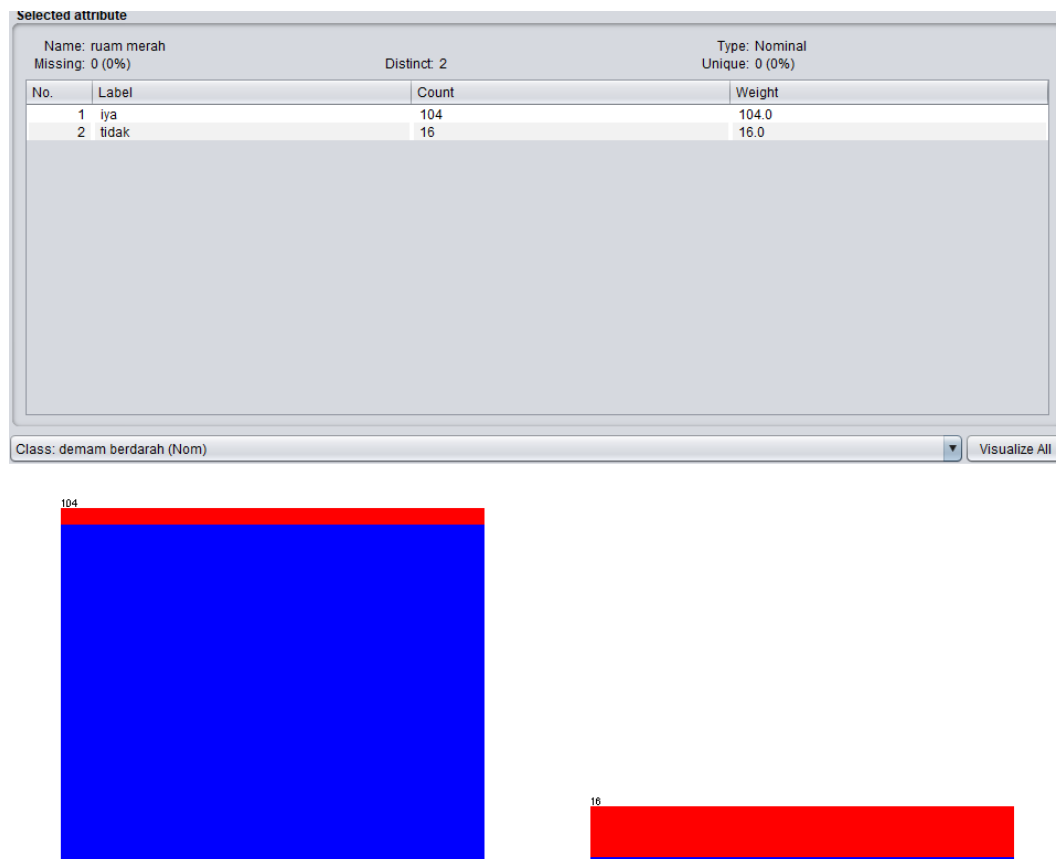


**Gambar 5.5 Visualisasi Atribut Muntah(Data Training)**

Gambar 5.5 adalah visualisasi atribut muntah, diketahui dari 120 data pada *column selected Attribute* terdapat 0 (0%) *missing*. *Distinct 2 Attribute* terdapat *unique* 0 (0%). Data pasien yang mengalami gejala muntah yaitu sebanyak 15 data dan yang tidak mengalami gejala muntah sebanyak 105 data.

#### 6. Visualisasi Atribut Ruam Merah

Visualisasi untuk melihat grafik ruam merah di tampilkan pada gambar 5.6 yang menggunakan tools weka 3.8

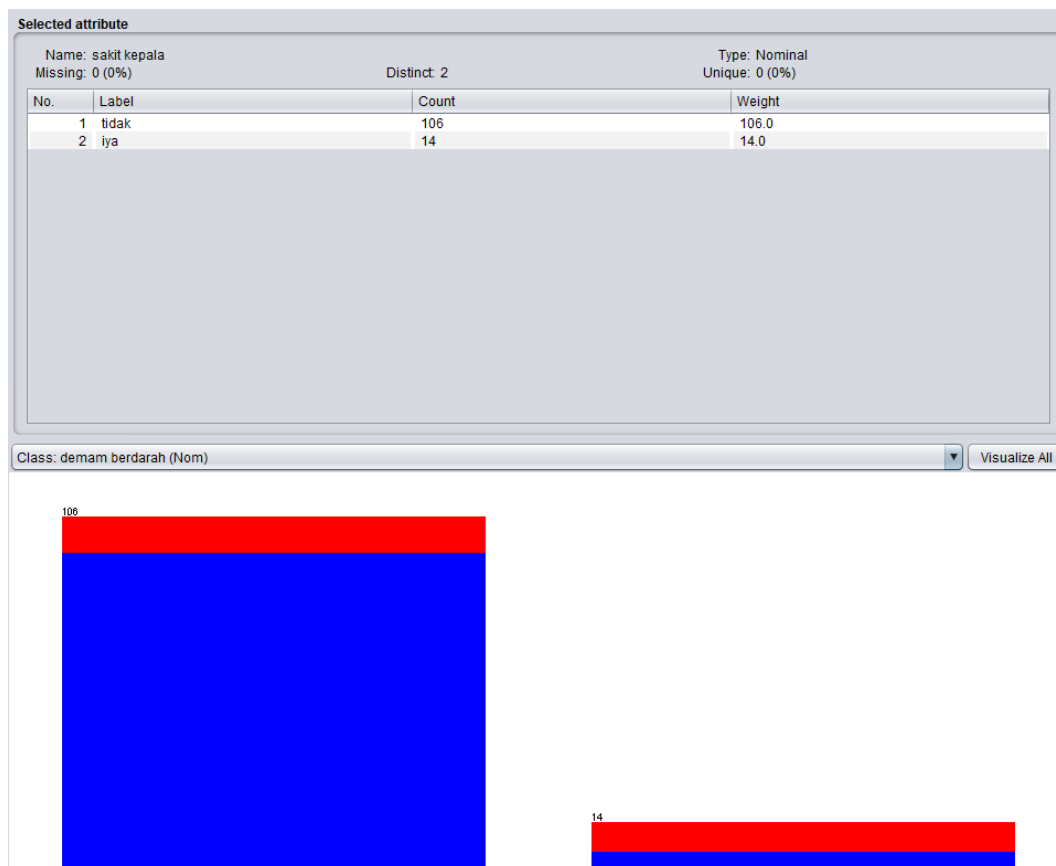


**Gambar 5.6 Visualisasi Atribut Ruam Merah(Data Training)**

Gambar 5.6 adalah visualisasi atribut ruam merah, diketahui dari 120 data pada *column selected Attribute* terdapat 0 (0%) *missing*. *Distinct 2 Attribute* terdapat *unique* 0 (0%). Data pasien yang mengalami ruam merah yaitu sebanyak 104 data dan yang tidak mengalami ruam merah sebanyak 16 data.

#### 7. Visualisasi Atribut Sakit Kepala

Visualisasi untuk melihat grafik ruam merah di tampilkan pada gambar 5.7 yang menggunakan tools weka 3.8

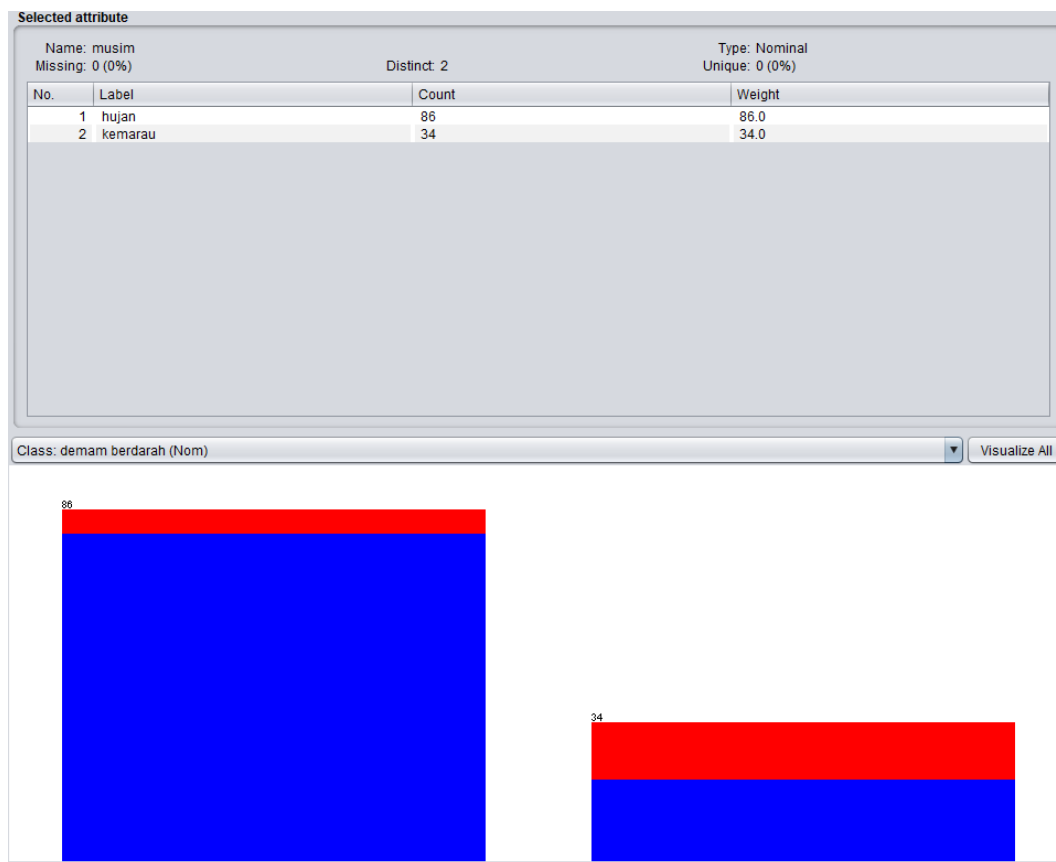


**Gambar 5.7 Visualisasi Atribut Sakit Kepala(Data Training)**

Gambar 5.7 adalah visualisasi atribut sakit kepala, diketahui dari 120 data pada *column selected Attribute* terdapat 0 (0%) *missing*. *Distinct 2 Attribute* terdapat *unique* 0 (0%). Data pasien yang mengalami gejala sakit kepala yaitu sebanyak 14 data dan yang tidak mengalami sebanyak 106 data.

#### 8. Visualisasi Atribut Musim

Visualisasi untuk melihat grafik musim di tampilkan pada gambar 5.8 yang menggunakan tools weka 3.8

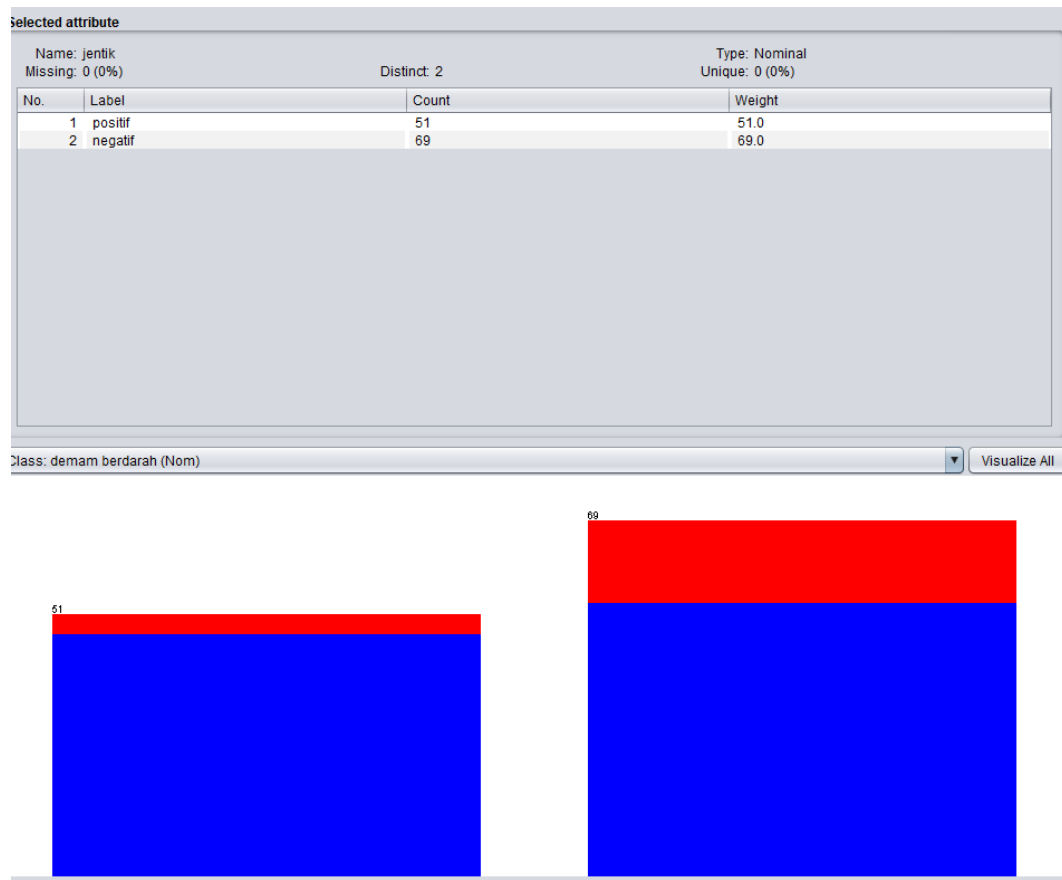


**Gambar 5.8 Visualisasi Atribut Musim(Data Training)**

Gambar 5.8 adalah visualisasi atribut musim, diketahui dari 120 data pada *column selected Attribute* terdapat 0 (0%) *missing*. *Distinct 2 Attribute* terdapat *unique* 0 (0%). Data pasien yang datang pada musim hujan yaitu sebanyak 86 data dan yang datang pada musim kemarau sebanyak 34 data.

#### 9. Visualisasi Atribut Jentik

Visualisasi untuk melihat grafik jentik di tampilkan pada gambar 5.9 yang menggunakan tools weka 3.8



**Gambar 5.9 Visualisasi Atribut Jentik(Data Training)**

Gambar 5.9 adalah visualisasi atribut jentik, diketahui dari 120 data pada *column selected Attribute* terdapat 0 (0%) *missing*. *Distinct 2 Attribute* terdapat *unique* 0 (0%). Data pasien yang positif terdapat jentik pada lingkungan rumahnya yaitu sebanyak 51 data dan yang negatif sebanyak 69 data.

#### 10. Visualisasi Atribut Demam Berdarah

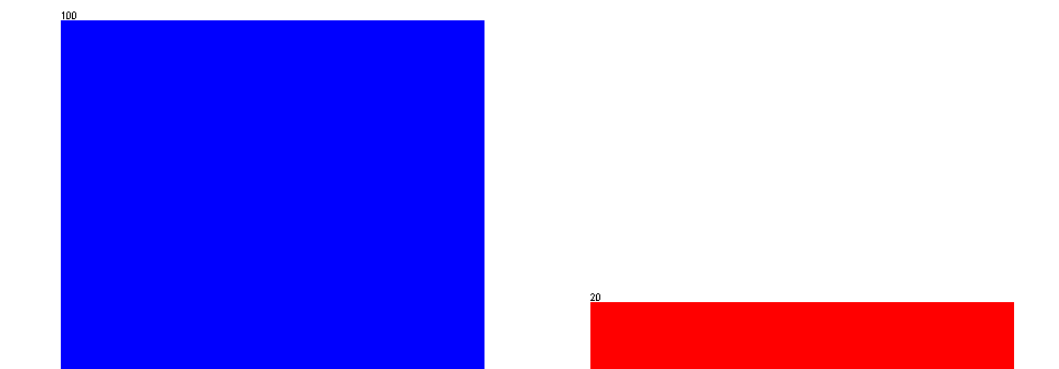
Visualisasi untuk melihat grafik demam berdarah di tampilkan pada gambar 5.10 yang menggunakan tools weka 3.8

**Selected attribute**

Name: demam berdarah  
Missing: 0 (0%)  
Distinct: 2  
Type: Nominal  
Unique: 0 (0%)

No.	Label	Count	Weight
1	positif	100	100.0
2	negatif	20	20.0

Class: demam berdarah (Nom) Visualize All



**Gambar 5.10 Visualisasi Atribut Demam Berdarah(Data Training)**

Gambar 5.10 adalah visualisasi atribut demam berdarah, diketahui dari 120 data pada *column selected Attribute* terdapat 0 (0%) *missing*. *Distinct 2 Attribute* terdapat *unique* 0 (0%). Data pasien yang positif mengidap demam berdarah yaitu sebanyak 100 data dan yang negatif sebanyak 20 data.

### 5.1.5 Hasil Visualisasi Atribut Data Testing dengan Menggunakan WEKA

Berikut merupakan visualisasi atribut yang diambil dari data *testing* dengan menggunakan tools WEKA :

1. Visualisasi Atribut Jenis Kelamin

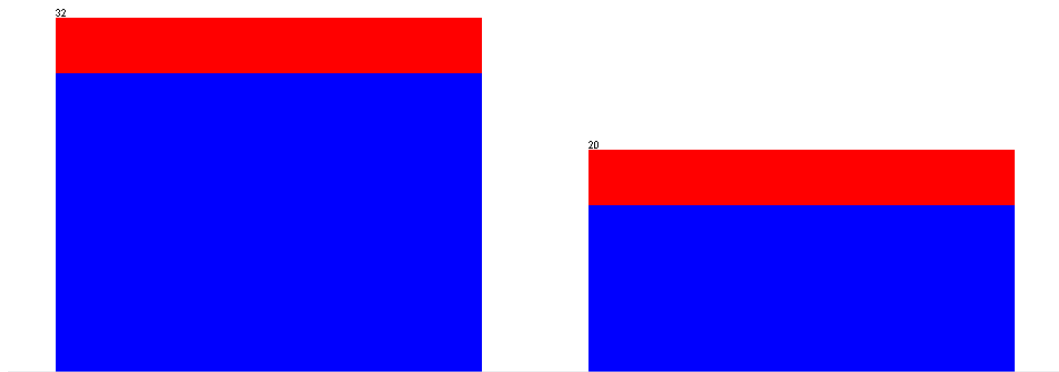
Visualisasi untuk melihat grafik Jenis Kelamin ditampilkan pada gambar 5.11 yang menggunakan tools weka 3.8

**Selected attribute**

Name: Jenis Kelamin  
Missing: 0 (0%)  
Distinct: 2  
Type: Nominal  
Unique: 0 (0%)

No.	Label	Count	Weight
1	Lk	32	32.0
2	Pr	20	20.0

Class: demam berdarah (Nom) Visualize All

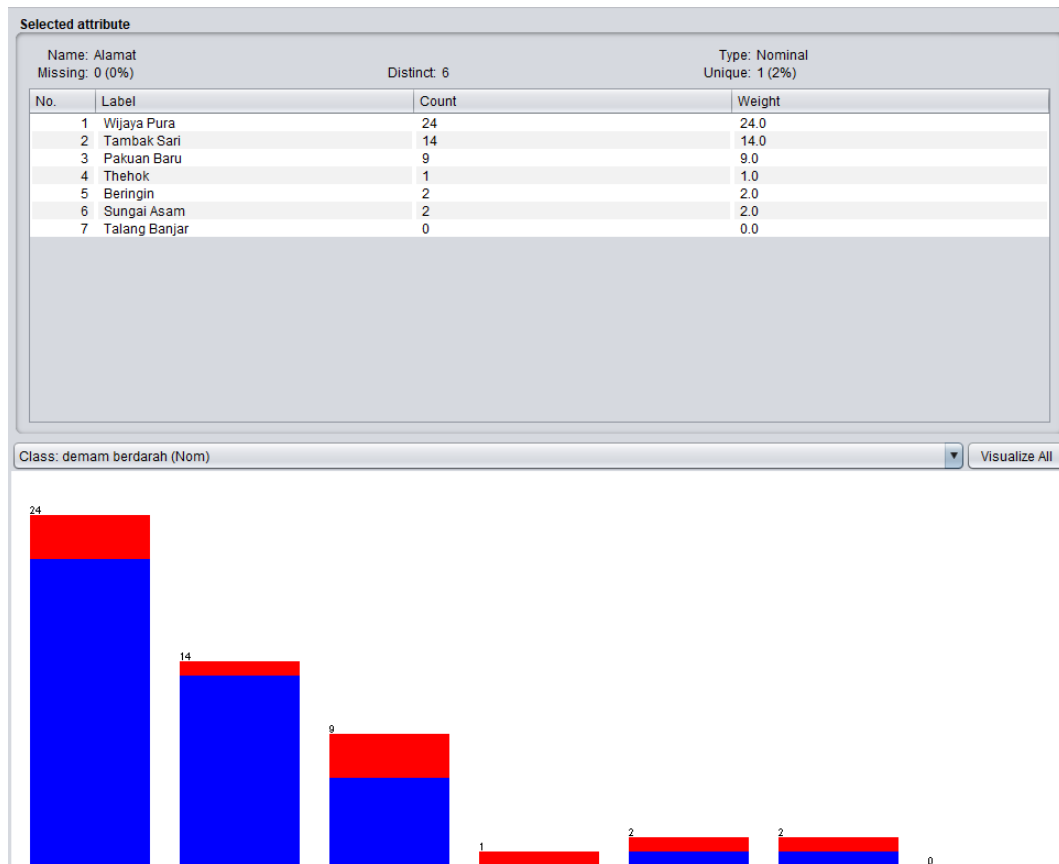


**Gambar 5.11 Visualisasi Atribut Jenis Kelamin(Data Testing)**

Gambar 5.11 adalah visualisasi dari atribut Jenis Kelamin yang mana dari 52 data pada *column selected Attribute* terdapat 0 (0%) *missing*. *Distinct 2 Attribute* terdapat *unique 0 (0%)* dengan jumlah data dari 32 berjenis kelamin laki-laki dan 20 berjenis kelamin perempuan.

## 2. Visualisasi Atribut Alamat

Visualisasi untuk melihat grafik alamat di tampilkan pada gambar 5.12 yang menggunakan tools weka 3.8



**Gambar 5.12 Visualisasi Atribut Alamat(Data Testing)**

Gambar 5.12 adalah visualisasi atribut alamat, diketahui dari 52 data pada *column selected Attribute* terdapat 0 (0%) *missing*. *Distinct 7 Attribute* terdapat *unique* 0 (0%). Data pasien yang beralamat di Wijaya Pura sebanyak 24 data, Tambak Sari sebanyak 14 data, Pakuan Baru sebanyak 9 data, Thehok sebanyak 1 data, Beringin sebanyak 2 data, Sungai Asam sebanyak 2 data, Talang Banjar sebanyak 0 data.

### 3. Visualisasi Atribut Kategori Usia

Visualisasi untuk melihat grafik kategori usia di tampilkan pada gambar 5.13 yang menggunakan tools weka 3.8.

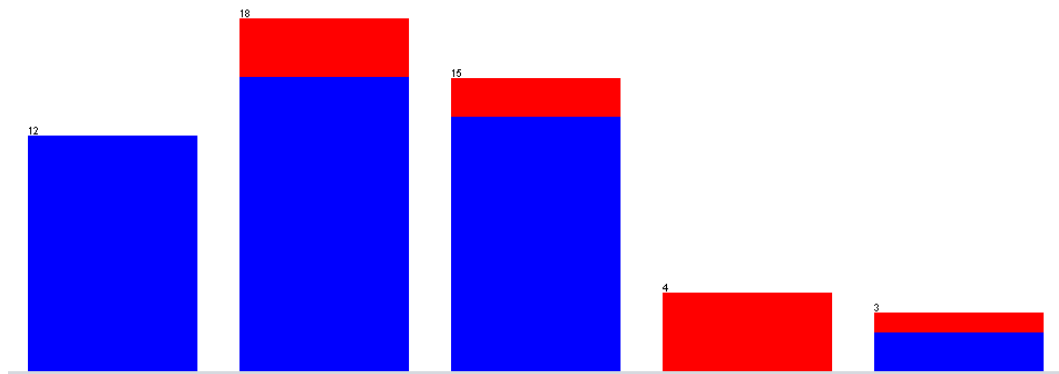


**Selected attribute**

Name: Kategori Usia  
Missing: 0 (0%)  
Distinct: 5  
Type: Nominal  
Unique: 0 (0%)

No.	Label	Count	Weight
1	Balita	12	12.0
2	Remaja	18	18.0
3	Kanak-Kanak	15	15.0
4	Lansia	4	4.0
5	Dewasa	3	3.0

Class: demam berdarah (Nom) Visualize All



**Gambar 5.13 Visualisasi Atribut Kategori Usia(Data Testing)**

Gambar 5.13 adalah visualisasi atribut kategori usia, diketahui dari 52 data pada *column selected Attribute* terdapat 0 (0%) *missing*. *Distinct 5 Attribute* terdapat *unique* 0 (0%). Data pasien yang berada di kategori umur balita (0-5thn) adalah sebanyak 12 data, Remaja(>12-25thn) sebanyak 18 data, Kanak-Kanak(>5-12thn) sebanyak 15 data, Lansia(>45thn) sebanyak 4 data, dan Dewasa(>25-45thn) sebanyak 3 data.

#### 4. Visualisasi Atribut Demam

Visualisasi untuk melihat grafik demam di ditampilkan pada gambar 5.14 yang menggunakan tools weka 3.8

**Selected attribute**

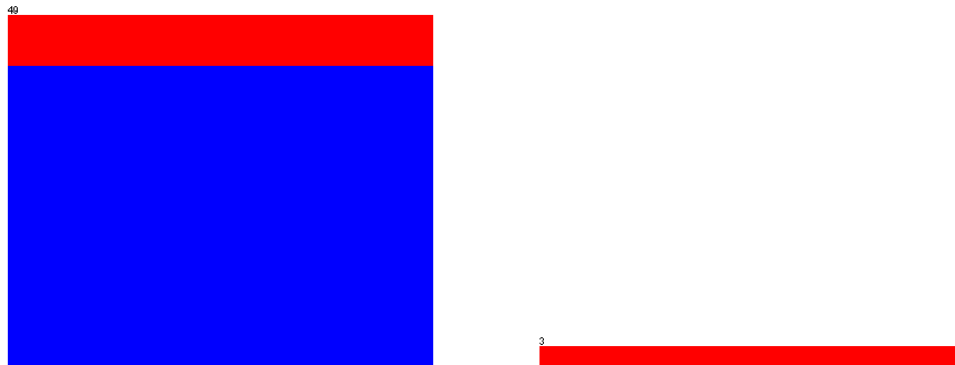
Name: demam  
Missing: 0 (0%)

Distinct: 2

Type: Nominal  
Unique: 0 (0%)

No.	Label	Count	Weight
1	iya	49	49.0
2	tidak	3	3.0

Class: demam berdarah (Nom) Visualize All



**Gambar 5.14 Visualisasi Atribut Demam(Data Testing)**

Gambar 5.14 adalah visualisasi atribut demam, diketahui dari 52 data pada *column selected Attribute* terdapat 0 (0%) *missing*. *Distinct 2 Attribute* terdapat *unique 0 (0%)*. Data pasien yang mengalami demam yaitu sebanyak 49 data dan yang tidak mengalami demam sebanyak 3 data.

#### 5. Visualisasi Atribut Muntah

Visualisasi untuk melihat grafik muntah di tampilkan pada gambar 5.15 yang menggunakan tools weka 3.8

Selected attribute

Name: muntah  
Missing: 0 (0%)

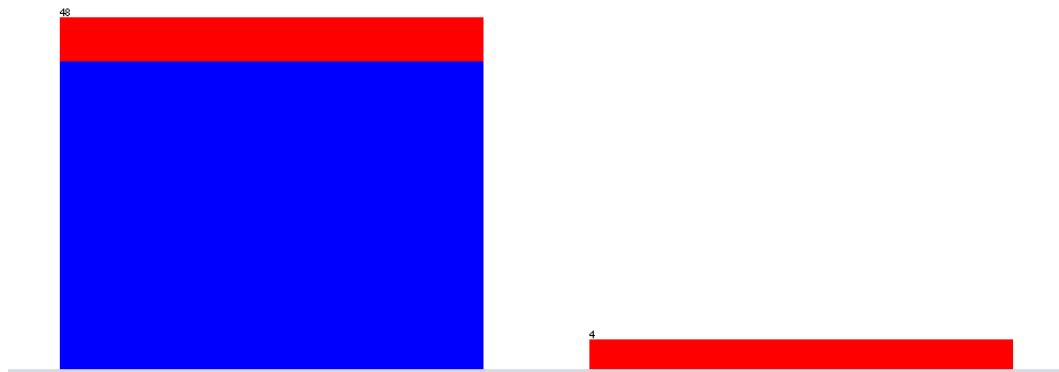
Distinct: 2

Type: Nominal  
Unique: 0 (0%)

No.	Label	Count	Weight
1	tidak	48	48.0
2	iya	4	4.0

Class: demam berdarah (Nom)

Visualize All

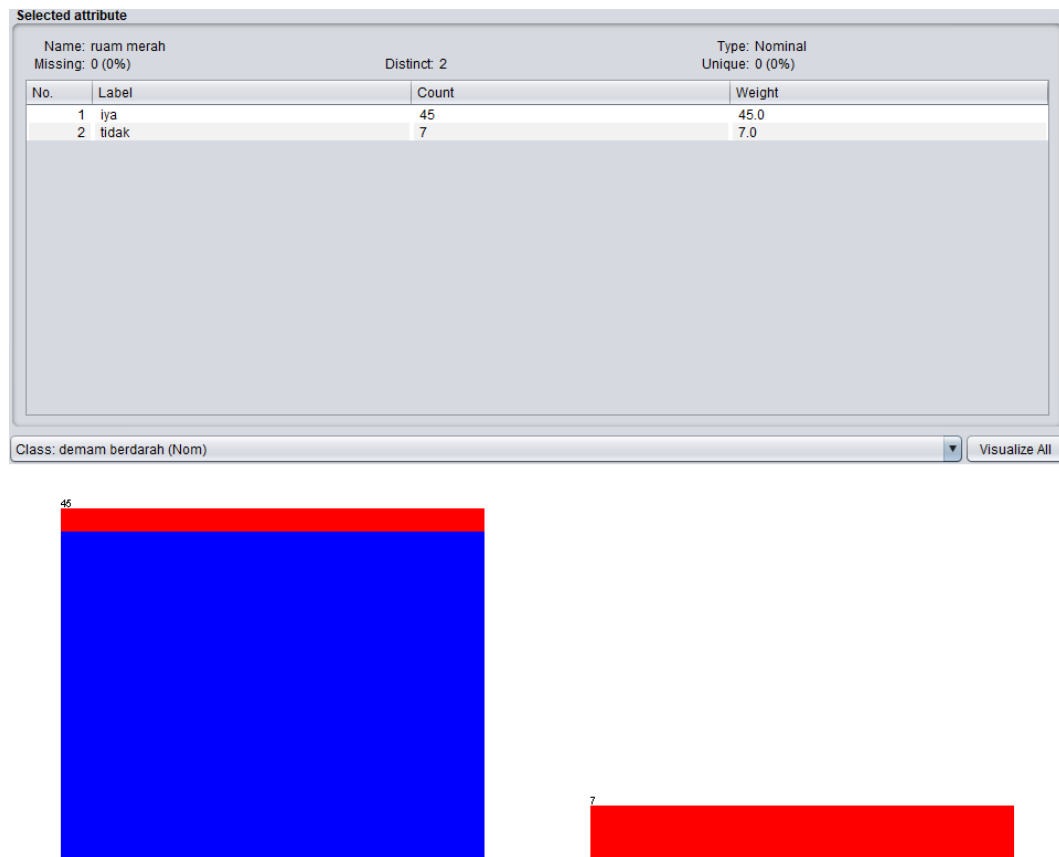


**Gambar 5.15 Visualisasi Atribut Muntah(Data Testing)**

Gambar 5.15 adalah visualisasi atribut muntah, diketahui dari 52 data pada *column selected Attribute* terdapat 0 (0%) *missing*. *Distinct 2 Attribute* terdapat *unique* 0 (0%). Data pasien yang mengalami gejala muntah yaitu sebanyak 4 data dan yang tidak mengalami gejala muntah sebanyak 48 data.

#### 6. Visualisasi Atribut Ruam Merah

Visualisasi untuk melihat grafik ruam merah di tampilkan pada gambar 5.16 yang menggunakan tools weka 3.8

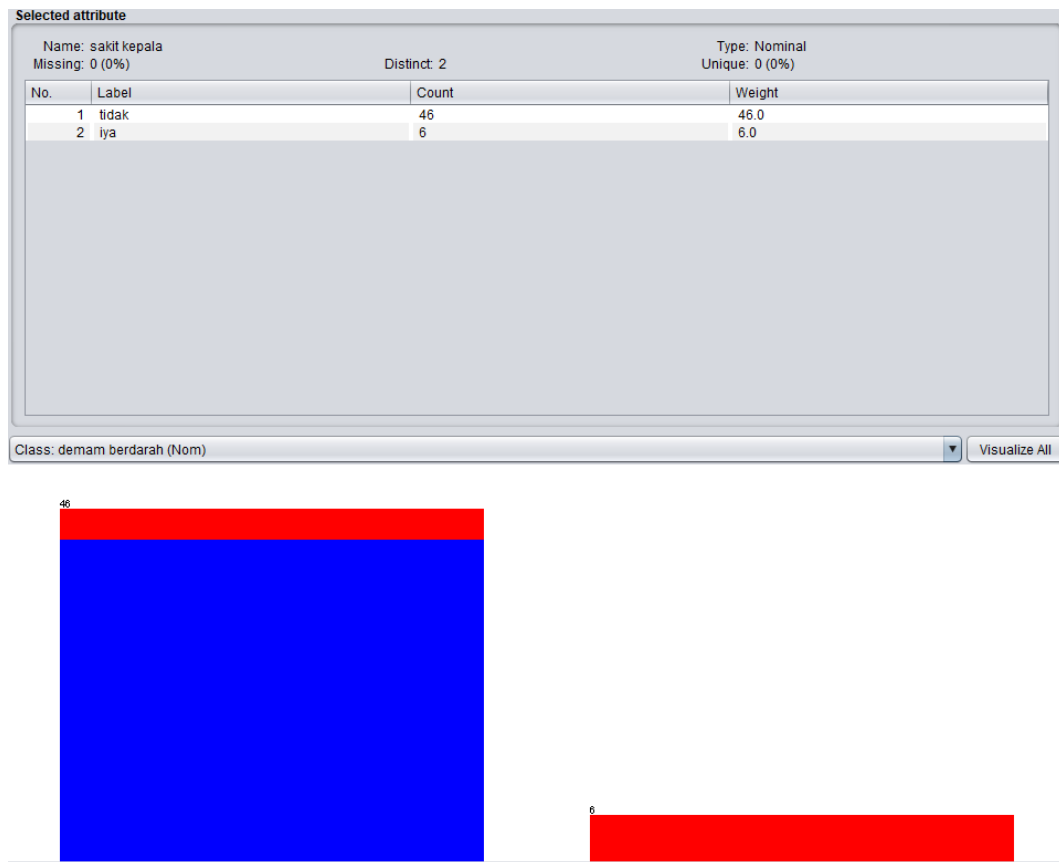


**Gambar 5.16 Visualisasi Atribut Ruam Merah(Data Testing)**

Gambar 5.16 adalah visualisasi atribut ruam merah, diketahui dari 52 data pada *column selected Attribute* terdapat 0 (0%) *missing*. *Distinct 2 Attribute* terdapat *unique* 0 (0%). Data pasien yang mengalami ruam merah yaitu sebanyak 45 data dan yang tidak mengalami ruam merah sebanyak 7 data.

#### 7. Visualisasi Atribut Sakit Kepala

Visualisasi untuk melihat grafik ruam merah di tampilkan pada gambar 5.17 yang menggunakan tools weka 3.8



**Gambar 5.17 Visualisasi Atribut Sakit Kepala(Data Testing)**

Gambar 5.17 adalah visualisasi atribut sakit kepala, diketahui dari 52 data pada *column selected Attribute* terdapat 0 (0%) *missing*. *Distinct 2 Attribute* terdapat *unique* 0 (0%). Data pasien yang mengalami gejala sakit kepala yaitu sebanyak 6 data dan yang tidak mengalami sebanyak 46 data.

#### 8. Visualisasi Atribut Musim

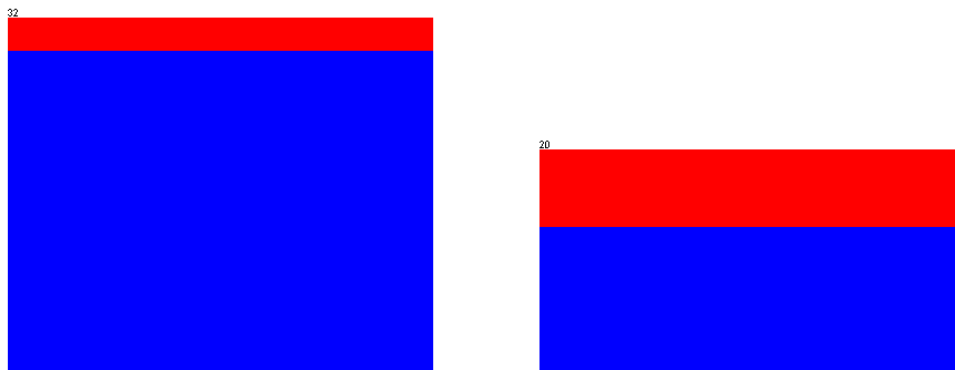
Visualisasi untuk melihat grafik musim di tampilkan pada gambar 5.18 yang menggunakan tools weka 3.8

**Selected attribute**

Name: musim  
Missing: 0 (0%)  
Distinct: 2  
Type: Nominal  
Unique: 0 (0%)

No.	Label	Count	Weight
1	hujan	32	32.0
2	kemarau	20	20.0

Class: demam berdarah (Nom) Visualize All

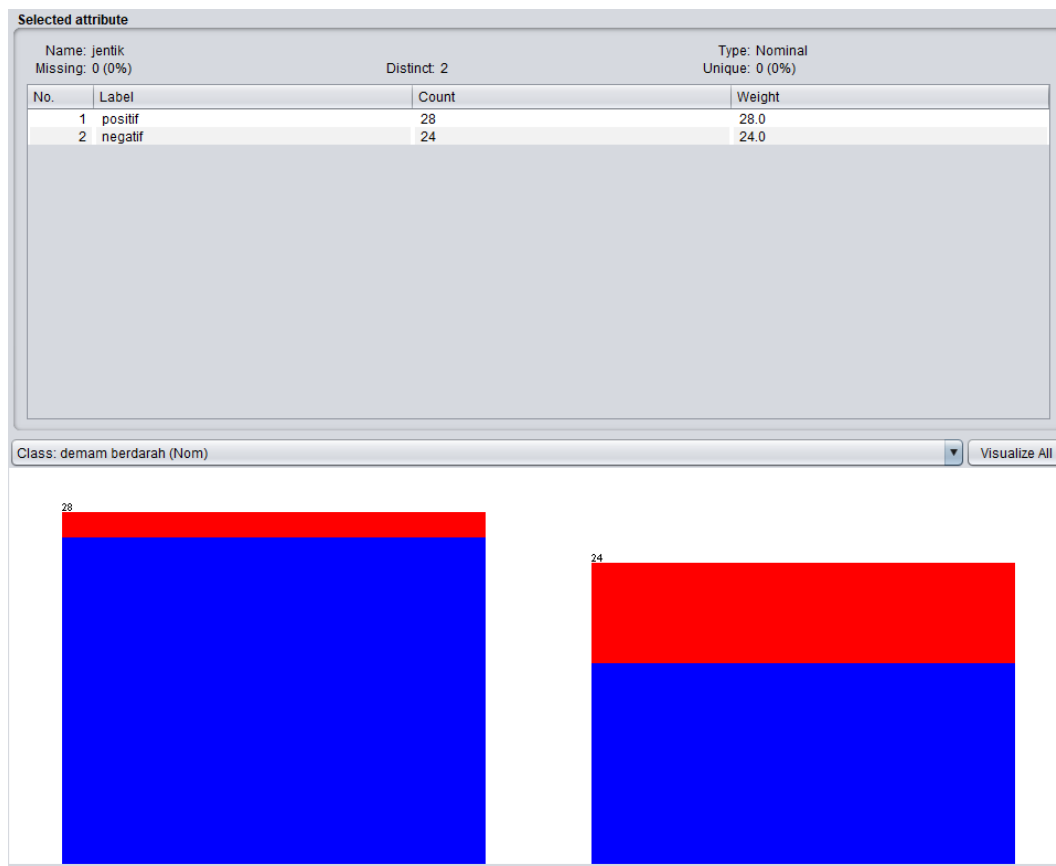


**Gambar 5.18 Visualisasi Atribut Musim(Data Testing)**

Gambar 5.18 adalah visualisasi atribut musim, diketahui dari 52 data pada *column selected Attribute* terdapat 0 (0%) *missing*. *Distinct 2 Attribute* terdapat *unique* 0 (0%). Data pasien yang datang pada musim hujan yaitu sebanyak 32 data dan yang datang pada musim kemarau sebanyak 20 data.

#### 9. Visualisasi Atribut Jentik

Visualisasi untuk melihat grafik jentik di tampilkan pada gambar 5.19 yang menggunakan tools weka 3.8



**Gambar 5.19 Visualisasi Atribut Jentik(Data Testing)**

Gambar 5.19 adalah visualisasi atribut jentik, diketahui dari 52 data pada *column selected Attribute* terdapat 0 (0%) *missing*. *Distinct 2 Attribute* terdapat *unique* 0 (0%). Data pasien yang positif terdapat jentik pada lingkungan rumahnya yaitu sebanyak 28 data dan yang negatif sebanyak 24 data.

#### 10. Visualisasi Atribut Demam Berdarah

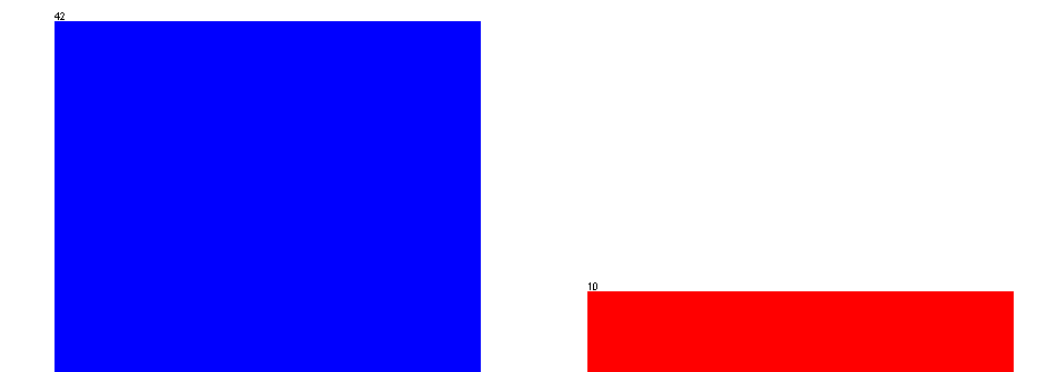
Visualisasi untuk melihat grafik demam berdarah di tampilkan pada gambar 5.20 yang menggunakan tools weka 3.8

**Selected attribute**

Name: demam berdarah  
Missing: 0 (0%)  
Distinct: 2  
Type: Nominal  
Unique: 0 (0%)

No.	Label	Count	Weight
1	positif	42	42.0
2	negatif	10	10.0

Class: demam berdarah (Nom) Visualize All



**Gambar 5.20 Visualisasi Atribut Demam Berdarah(Data *Testing*)**

Gambar 5.20 adalah visualisasi atribut demam berdarah, diketahui dari 52 data pada *column selected Attribute* terdapat 0 (0%) *missing*. *Distinct 2 Attribute* terdapat *unique* 0 (0%). Data pasien yang positif mengidap demam berdarah yaitu sebanyak 42 data dan yang negatif sebanyak 10 data.

## 5.2 Hasil Klasifikasi *Naive Bayes* Menggunakan Tool *WEKA*

Hasil klasifikasi akan dibagi menjadi 2 yaitu hasil klasifikasi dengan menggunakan data *training* dan hasil klasifikasi dengan menggunakan data *testing*:

### 5.2.1 Hasil Klasifikasi *Naive Bayes* dengan Menggunakan Data *Training*



Hasil klasifikasi data *training* dengan menggunakan algoritma *naive bayes* dilakukan dengan 5 test yaitu menggunakan *Use Training Set*, *5 Cross-Validation*, *10 Cross-Validation*, *60% Percentage Split*, dan *80% Percentage Split*. Berikut merupakan hasil klasifikasi data *training* dengan metode *naive bayes* menggunakan tool *WEKA*:

#### 1. Hasil Klasifikasi *Naive Bayes* Menggunakan Tool *WEKA* (*Use Training Set*)

Test *Use Training Set* melakukan pengetesan data menggunakan data *Training* itu sendiri.

The screenshot shows the Weka Explorer interface with the Naive Bayes classifier selected. The 'Test options' section is set to 'Use training set'. The 'Classifier output' pane displays the following results:

```

=== Evaluation on training set ===
Time taken to test model on training data: 0.02 seconds
=== Summary ===
Correctly Classified Instances      112           93.3333 %
Incorrectly Classified Instances     8             6.6667 %
Kappa statistic                     0.76
Mean absolute error                  0.0657
Root mean squared error              0.2131
Relative absolute error              23.3554 %
Root relative squared error          57.1848 %
Total Number of Instances           120

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
               0,960   0,200   0,960     0,960   0,960     0,760  0,971    0,993    positif
               0,800   0,040   0,800     0,800   0,800     0,760  0,971    0,926    negatif
Weighted Avg.  0,933   0,173   0,933     0,933   0,933     0,760  0,971    0,982

=== Confusion Matrix ===
 a  b  <-- classified as
96  4  | a = positif
 4 16 | b = negatif
  
```

**Gambar 5.21** Klasifikasi *Naive Bayes* Data *Training* (*Use Training Set*)

Gambar 5.21 merupakan hasil klasifikasi *naive bayes* pada tools *WEKA* dengan menggunakan *use training set* yang menunjukkan hasil 112 prediksi benar dengan akurasi sebesar 93,33 % dan 8 prediksi salah dengan persentasi 6,67 % dengan waktu klasifikasi selama 0,02 detik.

Perhitungan manualnya *Confusion Matrix* :

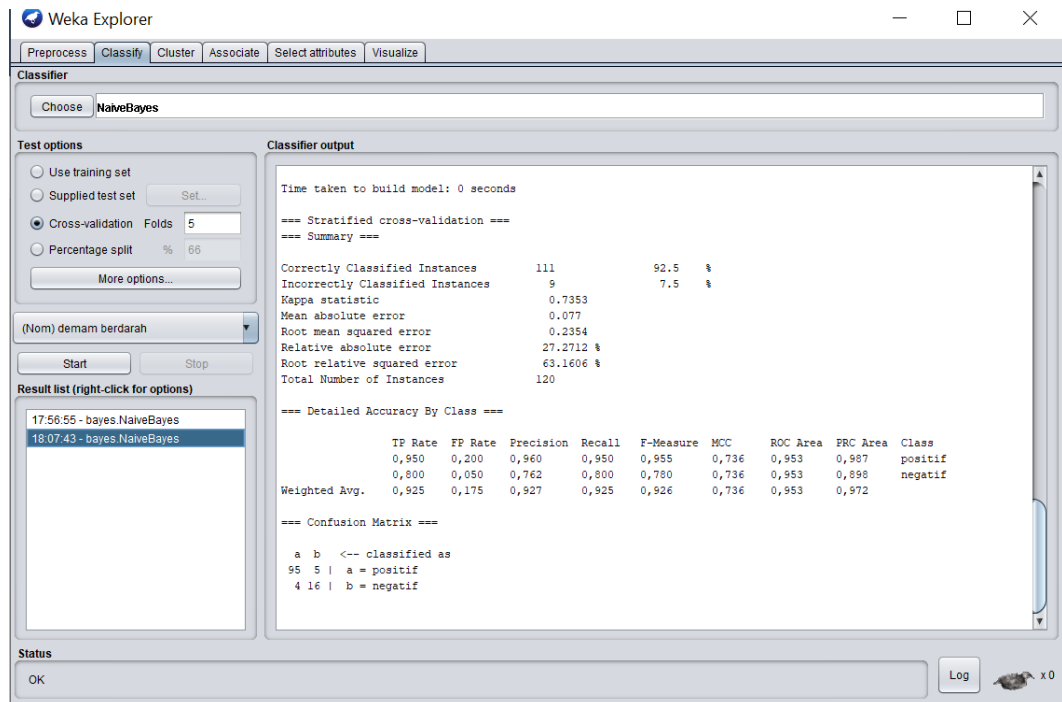
TP	FP
FN	TN

96	4
4	16

- a. Accuracy =  $((TP + TN) / (TP + TN + FP + FN)) * 100\%$   
=  $((96 + 16) / (96 + 16 + 4 + 4)) * 100\%$   
= 0,933
- b. Precision =  $((TP / (TP + FP)) * 100\%$   
=  $((96 / (96 + 4)) * 100\%$   
= 0,960
- c. Recall =  $((TP / (TP + FN)) * 100\%$   
=  $((96 / (96 + 4)) * 100\%$   
= 0,960

2. Hasil Klasifikasi *Naive Bayes* Menggunakan *Tool WEKA (5 Cross-Validation)*

Test *5 Cross-Validation* melakukan pengetesan data dimana data *training* dibagi menjadi *k* buah *subset* (subhimpunan). Dimana *k* adalah nilai dari *fold*. Pada pengetesan ini nilai *fold* adalah 5. Selanjutnya untuk tiap dari *subset*, akan dijadikan data tes dari hasil klasifikasi yang dihasilkan dari *k-1 subset* lainnya. Jadi, akan ada 5 kali tes. Dimana setiap data akan menjadi data tes sebanyak 1 kali dan menjadi data training sebanyak *k-1* kali.



**Gambar 5.22 Klasifikasi *Naive Bayes* Data Training (5 *Cross-Validation*)**

Gambar 5.22 merupakan hasil klasifikasi *naive bayes* pada tools WEKA dengan menggunakan 5 *Cross-Validation* yang menunjukkan hasil 111 prediksi benar dengan akurasi sebesar 92,5% dan 9 prediksi salah dengan persentasi 7,5% dengan waktu klasifikasi selama 0 detik.

Perhitungan manualnya *Confusion Matrix* :

TP	FP
FN	TN

95	5
4	16

$$\begin{aligned}
 \text{a. Accuracy} &= ((\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})) * 100\% \\
 &= ((95 + 16) / (95 + 16 + 5 + 4)) * 100\% \\
 &= 0,925
 \end{aligned}$$

$$b. \text{ Precision} = ((TP / (TP + FP)) * 100\%)$$

$$= ((96 / (96 + 5)) * 100\%)$$

$$= 0,950$$

$$c. \text{ Recall} = ((TP / (TP + FN)) * 100\%)$$

$$= ((96 / (96 + 4)) * 100\%)$$

$$= 0,960$$

### 3. Hasil Klasifikasi *Naive Bayes* Menggunakan *Tool WEKA* (10 *Cross-Validation*)

Test 10 *Cross-Validation* melakukan pengtesan data dimana nilai *fold* adalah 10. Selanjutnya untuk tiap dari *subset*, akan dijadikan data tes dari hasil klasifikasi yang dihasilkan dari k-1 *subset* lainnya. Jadi, akan ada 10 kali tes. Dimana setiap data akan menjadi data tes sebanyak 1 kali dan menjadi data *Training* Sebanyak k-1 kali

The screenshot shows the WEKA software interface with the Naive Bayes classifier selected. The 'Classifier output' pane displays the following results:

```

Time taken to build model: 0 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      112           93.3333 %
Incorrectly Classified Instances     8             6.6667 %
Kappa statistic                     0.76
Mean absolute error                 0.0778
Root mean squared error             0.2414
Relative absolute error             27.6231 %
Root relative squared error        64.7642 %
Total Number of Instances          120

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0,960  0,200  0,960   0,960  0,960     0,760  0,944   0,982  positif
0,800  0,040  0,800   0,800  0,800     0,760  0,944   0,884  negatif
Weighted Avg.   0,933  0,173  0,933   0,933  0,933     0,760  0,944   0,966

=== Confusion Matrix ===
 a b  <-- classified as
96 4 | a = positif
 4 16 | b = negatif

```

**Gambar 5.23** Klasifikasi Naive Bayes Data *Training* (10 *Cross-Validation*)

Gambar 5.23 merupakan hasil klasifikasi *naive bayes* pada *tools* WEKA dengan menggunakan 10 *Cross-Validation* yang menunjukkan hasil 112 prediksi benar dengan akurasi sebesar 93,33% dan 8 prediksi salah dengan persentasi 6,67% dengan waktu klasifikasi selama 0 detik.

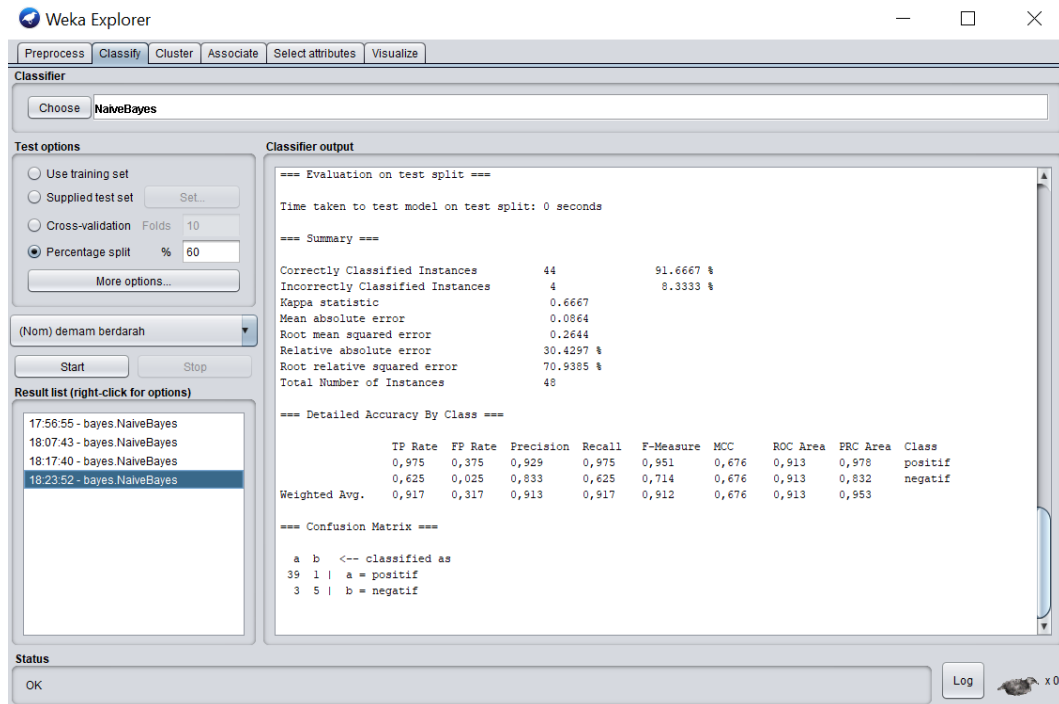
Perhitungan manualnya *Confusion Matrix* :

TP	FP	96	4
FN	TN	4	16

- a. Accuracy =  $((TP + TN) / (TP + TN + FP + FN)) * 100\%$   
=  $((96 + 16) / (96 + 16 + 4 + 4)) * 100\%$   
= 0,933
- b. Precision =  $((TP / (TP + FP)) * 100\%$   
=  $((96 / (96 + 4)) * 100\%$   
= 0,960
- c. Recall =  $((TP / (TP + FN)) * 100\%$   
=  $((96 / (96 + 4)) * 100\%$   
= 0,960

#### 4. Hasil Klasifikasi *Naive Bayes* Menggunakan *Tool* WEKA (60% *Percentage Split*)

Tes *Percentage Split* hasil klasifikasi akan dites dengan menggunakan k% dari data tersebut. Pada tes ini akan digunakan 60% *Percentage Split* dari data.



**Gambar 5.24** Klasifikasi *Naive Bayes* Data Training (60% Percentage Split)

Gambar 5.24 merupakan hasil klasifikasi *naive bayes* pada tools WEKA dengan menggunakan 60% *Percentage Split* yang menunjukkan hasil 44 prediksi benar dengan akurasi sebesar 91,67% dan 4 prediksi salah dengan persentasi 8,33% dengan waktu klasifikasi selama 0 detik.

Perhitungan manualnya *Confusion Matrix* :

TP	FP
FN	TN

39	1
3	5

- a. Accuracy =  $((TP + TN) / (TP + TN + FP + FN)) * 100\%$   
=  $((39 + 5) / (39 + 5 + 1 + 3)) * 100\%$   
= 0,917
- b. Precision =  $((TP / (TP + FP)) * 100\%$

$$= ((39 / (39 + 1)) * 100\%)$$

$$= 0,975$$

c. Recall  $= ((TP / (TP + FN)) * 100\%)$

$$= ((39 / (39 + 3)) * 100\%)$$

$$= 0,929$$

## 5. Hasil Klasifikasi *Naive Bayes* Menggunakan *Tool WEKA* (80% *Percentage Split*)

Tes *Percentage Split* hasil klasifikasi akan dites dengan menggunakan k% dari data tersebut. Pada tes ini akan digunakan 80% *Percentage Split* dari data.

The screenshot shows the WEKA software interface. The 'Classifier' tab is active, and 'NaiveBayes' is selected. The 'Test options' section shows 'Percentage split' set to 80%. The 'Classifier output' window displays the following results:

```

=== Evaluation on test split ===
Time taken to test model on test split: 0 seconds

=== Summary ===
Correctly Classified Instances      22          91.6667 %
Incorrectly Classified Instances    2           8.3333 %
Kappa statistic                    0.7778
Mean absolute error                 0.116
Root mean squared error            0.2868
Relative absolute error             35.5164 %
Root relative squared error        64.6391 %
Total Number of Instances          24

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall  F-Measure  MOC      ROC Area  PRC Area  Class
               0,944   0,167   0,944     0,944   0,944     0,778   0,954   0,985   positif
               0,833   0,056   0,833     0,833   0,833     0,778   0,954   0,906   negatif
Weighted Avg.   0,917   0,139   0,917     0,917   0,917     0,778   0,954   0,965

=== Confusion Matrix ===
 a b  <-- classified as
17 1 | a = positif
 1 5 | b = negatif

```

**Gambar 5.25** Klasifikasi *Naive Bayes* Data Training (80% *Percentage Split*)

Gambar 5.25 merupakan hasil klasifikasi *naive bayes* pada *tools WEKA* dengan menggunakan 80% *Percentage Split* yang menunjukkan hasil 20 prediksi benar dengan akurasi sebesar 91,67% dan 2 prediksi salah dengan persentasi 8,33% dengan waktu klasifikasi selama 0 detik.

Perhitungan manualnya *Confusion Matrix* :

TP	FP
FN	TN

17	1
1	5

- a. Accuracy =  $((TP + TN) / (TP + TN + FP + FN)) * 100\%$   
=  $((17 + 5) / (17 + 5 + 1 + 1)) * 100\%$   
= 0,917
- b. Precision =  $((TP / (TP + FP)) * 100\%$   
=  $((17 / (17 + 1)) * 100\%$   
= 0,944
- c. Recall =  $((TP / (TP + FN)) * 100\%$   
=  $((17 / (17 + 1)) * 100\%$   
= 0,944

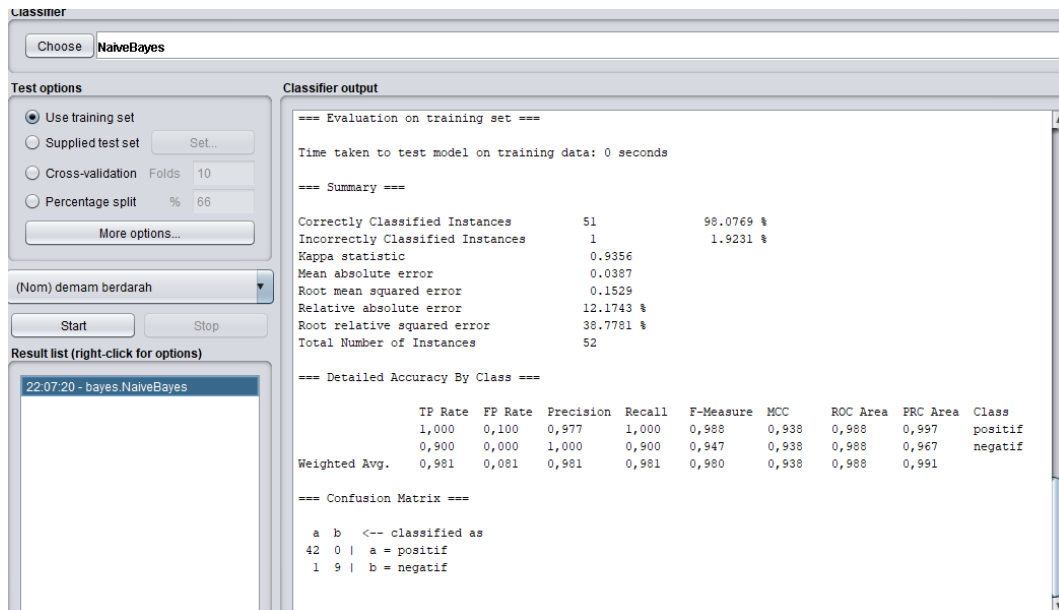
### 5.2.2 Hasil Klasifikasi *Naïve Bayes* dengan Menggunakan Data *Testing*

Hasil klasifikasi data *testing* dengan menggunakan algoritma *naive bayes* dilakukan dengan 5 test yaitu menggunakan *Use Training Set*, *5 Cross-Validation*, *10 Cross-Validation*, *60% Percentage Split*, dan *80% Percentage Split*. Berikut merupakan hasil klasifikasi data *testing* dengan metode *naive bayes* menggunakan *tool WEKA*:

#### 1. Hasil Klasifikasi *Naive Bayes* Menggunakan *Tool WEKA (Use Training Set)*

Test *Use Training Set* melakukan pengetestan data menggunakan data *Training* itu sendiri.





**Gambar 5.26 Klasifikasi Naive Bayes Data Testing(Use Training Set)**

Gambar 5.26 merupakan hasil klasifikasi *naive bayes* pada *tools* WEKA dengan menggunakan *use training set* yang menunjukkan hasil 51 prediksi benar dengan akurasi sebesar 98,08 % dan 1 prediksi salah dengan persentasi 1,92 % dengan waktu klasifikasi selama 0 detik.

Perhitungan manualnya *Confusion Matrix* :

TP	FP
FN	TN

42	0
1	9

$$\begin{aligned}
 \text{a. Accuracy} &= ((\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})) * 100\% \\
 &= ((42 + 9) / (42 + 9 + 0 + 1)) * 100\% \\
 &= 0,981
 \end{aligned}$$

$$\begin{aligned}
 \text{b. Precision} &= ((\text{TP} / (\text{TP} + \text{FP})) * 100\% \\
 &= ((42 / (42 + 0)) * 100\% \\
 &= 1
 \end{aligned}$$

$$\begin{aligned}
 \text{c. Recall} &= ((\text{TP} / (\text{TP} + \text{FN})) * 100\%) \\
 &= ((42 / (42 + 1)) * 100\%) \\
 &= 0,977
 \end{aligned}$$

## 2. Hasil Klasifikasi *Naive Bayes* Menggunakan *Tool WEKA* (5 *Cross-Validation*)

Test 5 *Cross-Validation* melakukan pengetestan data dimana data dibagi menjadi k buah *subset* (subhimpunan). Dimana k adalah nilai dari *fold*. Pada pengetestan ini nilai *fold* adalah 5. Selanjutnya untuk tiap dari *subset*, akan dijadikan data tes dari hasil klasifikasi yang dihasilkan dari k-1 *subset* lainnya. Jadi, akan ada 5 kali tes. Dimana setiap data akan menjadi data tes sebanyak 1 kali.

The screenshot shows the WEKA Classifier window with the following details:

- Classifier:** NaiveBayes
- Test options:**
  - Use training set:
  - Supplied test set:  (Set...)
  - Cross-validation:**  Folds: 5
  - Percentage split:  % 66
- Classifier output:**

```

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      50           96.1538 %
Incorrectly Classified Instances    2            3.8462 %
Kappa statistic                    0.866
Mean absolute error                 0.0618
Root mean squared error             0.1889
Relative absolute error             19.3499 %
Root relative squared error         47.8848 %
Total Number of Instances          52

=== Detailed Accuracy By Class ===
                TP Rate  FP Rate  Precision  Recall  F-Measure  MOC      ROC Area  PRC Area  Class
Weighted Avg.   0,962   0,162   0,963     0,962   0,960     0,874   0,945   0,972
                1,000   0,200   0,955     1,000   0,977     0,874   0,945   0,982   positif
                0,800   0,000   1,000     0,800   0,889     0,874   0,945   0,930   negatif

=== Confusion Matrix ===
 a  b  <-- classified as
42  0  | a = positif
 2  8  | b = negatif

```
- Result list (right-click for options):**
  - 22:07:20 - bayes.NaiveBayes
  - 22:17:23 - bayes.NaiveBayes**

**Gambar 5.27** Klasifikasi *Naive Bayes* Data Testing (5 *Cross-Validation*)

Gambar 5.27 merupakan hasil klasifikasi *naive bayes* pada *tools WEKA* dengan menggunakan 5 *Cross-Validation* yang menunjukkan hasil 50 prediksi

benar dengan akurasi sebesar 96,15% dan 2 prediksi salah dengan persentasi 3,85% dengan waktu klasifikasi selama 0 detik.

Perhitungan manualnya *Confusion Matrix* :

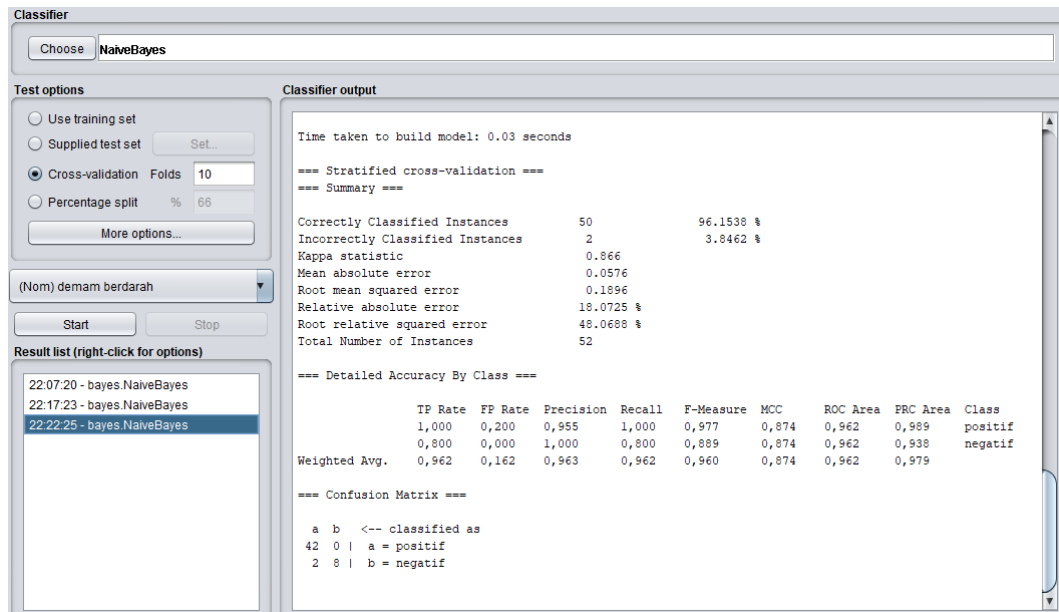
TP	FP
FN	TN

42	0
2	8

- a. Accuracy =  $((TP + TN) / (TP + TN + FP + FN)) * 100\%$   
 $= ((42 + 8) / (42 + 8 + 0 + 2)) * 100\%$   
 $= 0,962$
- b. Precision =  $((TP / (TP + FP)) * 100\%$   
 $= ((42 / (42 + 0)) * 100\%$   
 $= 1$
- c. Recall =  $((TP / (TP + FN)) * 100\%$   
 $= ((42 / (42 + 2)) * 100\%$   
 $= 0,955$

### 3. Hasil Klasifikasi *Naive Bayes* Menggunakan *Tool WEKA* (10 *Cross-Validation*)

Test 10 *Cross-Validation* melakukan pengetesan data dimana nilai *fold* adalah 10. Selanjutnya untuk tiap dari *subset*, akan dijadikan data tes dari hasil klasifikasi yang dihasilkan dari k-1 *subset* lainnya. Jadi, akan ada 10 kali tes. Dimana setiap data akan menjadi data tes sebanyak 1 kali dan menjadi data *Training* Sebanyak k-1 kali



**Gambar 5.28 Klasifikasi Naive Bayes Data Testing (10 Cross-Validation)**

Gambar 5.28 merupakan hasil klasifikasi *naive bayes* pada *tools* WEKA dengan menggunakan 10 *Cross-Validation* yang menunjukkan hasil 50 prediksi benar dengan akurasi sebesar 96,15% dan 2 prediksi salah dengan persentasi 3,85% dengan waktu klasifikasi selama 0,03 detik.

Perhitungan manualnya *Confusion Matrix* :

TP	FP
FN	TN

42	0
2	8

$$a. \text{ Accuracy} = ((TP + TN) / (TP + TN + FP + FN)) * 100\%$$

$$= ((42 + 8) / (42 + 8 + 0 + 2)) * 100\%$$

$$= 0,962$$

$$b. \text{ Precision} = ((TP / (TP + FP)) * 100\%$$

$$= ((42 / (42 + 0)) * 100\%$$

$$= 1$$

$$\begin{aligned}
 \text{c. Recall} &= ((\text{TP} / (\text{TP} + \text{FN})) * 100\%) \\
 &= ((42 / (42 + 2)) * 100\%) \\
 &= 0,955
 \end{aligned}$$

#### 4. Hasil Klasifikasi *Naive Bayes* Menggunakan *Tool WEKA* (60% *Percentage Split*)

Tes *Percentage Split* hasil klasifikasi akan dites dengan menggunakan k% dari data tersebut. Pada tes ini akan digunakan 60% *Percentage Split* dari data.

The screenshot shows the WEKA Classifier interface with the following details:

- Classifier:** NaiveBayes
- Test options:** Percentage split is selected at 60%.
- Classifier output:**
  - Time taken to test model on test split: 0 seconds
  - Summary:
 

Correctly Classified Instances	21	100 %
Incorrectly Classified Instances	0	0 %
Kappa statistic	1	
Mean absolute error	0.025	
Root mean squared error	0.0586	
Relative absolute error	7.9112 %	
Root relative squared error	16.1027 %	
Total Number of Instances	21	
  - Detailed Accuracy By Class:
 

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	positif
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	negatif
  - Confusion Matrix:
 

```

          a b <-- classified as
          18 0 | a = positif
           0 3 | b = negatif
          
```

**Gambar 5.29** Klasifikasi *Naive Bayes* Data *Testing* (60% *Percentage Split*)

Gambar 5.29 merupakan hasil klasifikasi *naive bayes* pada *tools WEKA* dengan menggunakan 60% *Percentage Split* yang menunjukkan hasil 21 prediksi benar dengan akurasi sebesar 100% dan 0 prediksi salah dengan persentasi 0% dengan waktu klasifikasi selama 0 detik.

Perhitungan manualnya *Confusion Matrix* :

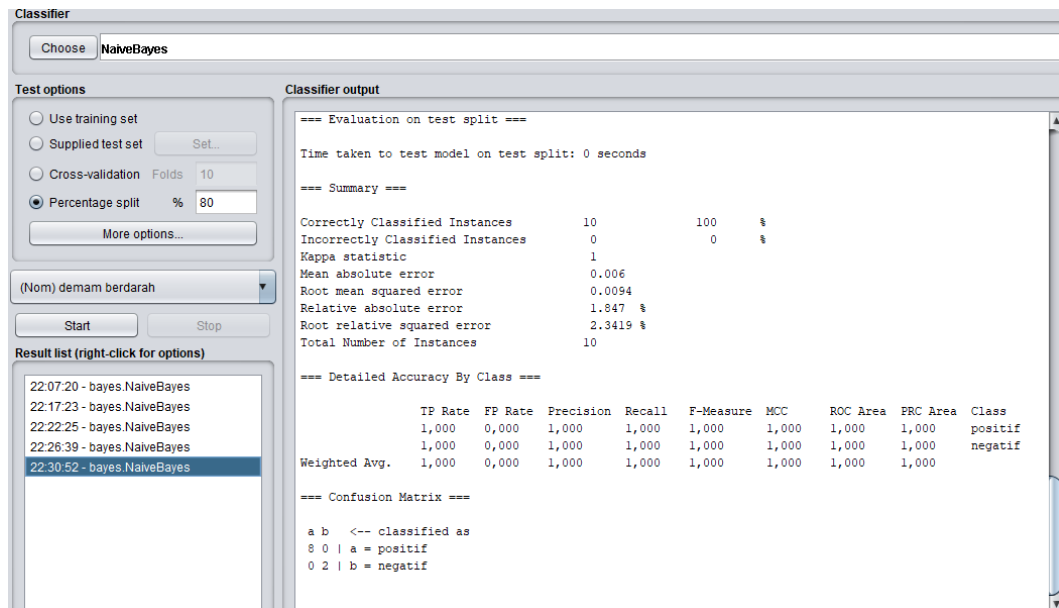
TP	FP
FN	TN

18	0
0	3

- a. Accuracy =  $((TP + TN) / (TP + TN + FP + FN)) * 100\%$   
=  $((18 + 3) / (18 + 3 + 0 + 0)) * 100\%$   
= 1
- b. Precision =  $((TP / (TP + FP)) * 100\%$   
=  $((18 / (18 + 0)) * 100\%$   
= 1
- c. Recall =  $((TP / (TP + FN)) * 100\%$   
=  $((18 / (18 + 0)) * 100\%$   
= 1

5. Hasil Klasifikasi *Naive Bayes* Menggunakan *Tool WEKA* (80% *Percentage Split*)

Tes *Percentage Split* hasil klasifikasi akan dites dengan menggunakan k% dari data tersebut. Pada tes ini akan digunakan 80% *Percentage Split* dari data.



**Gambar 5.30** Klasifikasi *Naive Bayes* Data Testing(80% Percentage Split)

Gambar 5.30 merupakan hasil klasifikasi *naive bayes* pada tools WEKA dengan menggunakan 80% *Percentage Split* yang menunjukkan hasil 10 prediksi benar dengan akurasi sebesar 100% dan 0 prediksi salah dengan persentasi 0% dengan waktu klasifikasi selama 0 detik.

Perhitungan manualnya *Confusion Matrix* :

TP	FP
FN	TN

8	0
0	2

a. Accuracy =  $((TP + TN) / (TP + TN + FP + FN)) * 100\%$

$$= ((8 + 2) / (8 + 2 + 0 + 0)) * 100\%$$

$$= 1$$

b. Precision =  $((TP / (TP + FP)) * 100\%$

$$= ((8 / (8 + 0)) * 100\%$$

$$= 1$$

$$\begin{aligned} \text{c. Recall} &= ((\text{TP} / (\text{TP} + \text{FN})) * 100\% \\ &= ((8 / (8 + 0)) * 100\% \\ &= 1 \end{aligned}$$

### 5.3 Hasil Evaluasi Rata-Rata Akurasi Dari 5 *Test Options*

Setelah dilakukan analisis klasifikasi *naive bayes* pada *tool WEKA* menggunakan *Use Training Set*, *5 Fold Cross Validation*, *10 Fold Cross Validation*, *60% Percentage Split*, dan *80% Percentage Split*, maka diperoleh akurasi tertinggi yaitu dengan menggunakan *Use Training Set* dan *10 Fold Cross Validation* dengan persentasi akurasi yaitu 93,33 % dan untuk *Correctly Classified Instances* adalah sebesar 6,67 %. Perbandingan hasil analisis dapat dilihat pada tabel 5.3.



Tabel 5.3 Perbandingan Evaluasi Akurasi WEKA

<b>Model Evaluasi</b>	<b>Akurasi</b>	<b>Jumlah Kelas Data <i>Training</i></b>	<b>Persentasi Data <i>Training</i></b>	<b>Jumlah Kelas Data <i>Testing</i></b>	<b>Persentasi Data <i>Testing</i></b>
<i>Use Training Set</i>	<i>Correctly Classified Instances</i>	112	93,33%	51	98,08%
	<i>Incorrectly Classified Instances</i>	8	6,67%	1	1,92%
<i>5 Fold Cross Validation</i>	<i>Correctly Classified Instances</i>	111	92,5%	50	96,15%
	<i>Incorrectly Classified Instances</i>	9	7,5%	2	3,85%
<i>10 Fold Cross Validation</i>	<i>Correctly Classified Instances</i>	112	93,33%	50	96,15%
	<i>Incorrectly Classified Instances</i>	8	6,67%	2	3,85%
<i>60% Percentage Split</i>	<i>Correctly Classified Instances</i>	44	91,67%	21	100%
	<i>Incorrectly Classified Instances</i>	4	8,33%	0	0%
<i>80% Percentage Split</i>	<i>Correctly Classified Instances</i>	22	91,67%	10	100%
	<i>Incorrectly Classified Instances</i>	2	8,33%	0	0%
<i>Average</i>	<i>Correctly Classified Instances</i>	80,2	92,5%	36,4	98,08%
	<i>Incorrectly Classified Instances</i>	6,2	7,5%	1	1,92%

Berdasarkan tabel diatas dapat disimpulkan bahwa rata-rata dari persentase Hasil Klasifikasi *Naive Bayes* dengan menggunakan 5 *test options*. Pada data *training* kelas rata-rata *Correctly Classified Instances* adalah 80,2 dan *Incorrectly Classified Instances* 6,2 serta persentase akurasi *Correctly Classified Instances* sebesar 92,5% dan *Incorrectly Classified Instances* 7,5%. Pada data *testing* kelas rata-rata *Correctly Classified Instances* adalah 36,4 dan *Incorrectly Classified Instances* 1 serta persentase akurasi *Correctly Classified Instances* sebesar 98,08% dan *Incorrectly Classified Instances* 1,92%.