

## PREDICTION OF MILENIAL AND POST MILENIAL VOTERS IN ELECTIONS USING THE NAIVE BAYES ALGORITHM

### PREDIKSI PEMILIH MILENIAL DAN PASCA MILENIAL DI PEMILU MENGGUNAKAN ALGORITMA NAIVE BAYES

Sista Faizzah Rohma<sup>1</sup>, Arif Senja Fitriani<sup>2</sup>

[faizzahsista@gmail.com](mailto:faizzahsista@gmail.com)<sup>1</sup>, [asfjim@umsida.ac.id](mailto:asfjim@umsida.ac.id)<sup>2</sup>

Program Studi Teknik Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia<sup>1,2</sup>

**Abstract.** The main problem in efforts to predict the current election is related to the fact that many millennials abstain from all over Indonesia, especially in the Wonokasian District, as evidenced by the high abstentions between regions. This research will test the classification based on election population data obtained from Wonokasian District in 2019 using the Naïve Bayes method. The variables used are the number of households, population origin, category, status (married/not married), address, RT, TPS, location, present (family) and present. The results of this study used Weka to make it easier for the village community to find out millennial data between those who abstain and those who do not abstain. From the results of the classification using the Naïve Bayes method to classify the presence status of 100 datasets, the highest results were obtained using manual calculations with a truth level of 76.67% and an error rate of 23.33%. As for 10% of the 100 datasets, the highest results were obtained using Weka with a truth level of 100% and an error rate of 0%.

**Keywords** - data mining, naïve bayes, sidoarjo

**Abstrak.** Masalah utama dalam upaya memprediksi pemilu saat ini terkait adanya fakta bahwa banyaknya milenial yang golput di seluruh wilayah Indonesia, khususnya di Kecamatan Wonokasian, dibuktikan dengan tingginya golput antar daerah. Penelitian akan menguji klasifikasi berdasarkan data penduduk pemilu yang diperoleh dari Kecamatan Wonokasian tahun 2019 menggunakan metode Naïve Bayes. Variabel yang digunakan adalah Jumlah KK, Asal Penduduk, Kategori, Status (Kawin/Belum Kawin), Alamat, RT, TPS, Lokasi, Hadir (keluarga) dan Hadir. Hasil dari penelitian ini menggunakan weka sehingga memudahkan bagi masyarakat desa mengetahui data milenial antara yang golput maupun tidak golput. Dari hasil klasifikasi pada metode Naïve Bayes untuk mengklasifikasi status kehadiran dari 100 dataset, diperoleh hasil tertinggi menggunakan perhitungan manual dengan perolehan tingkat kebenaran sebesar 76,67% dan tingkat kesalahan sebesar 23,33%. Adapun dari 10% dari 100 dataset, diperoleh hasil tertinggi menggunakan weka dengan perolehan tingkat kebenaran sebesar 100% dan tingkat kesalahan sebesar 0%

**Kata Kunci** - data mining, naïve bayes, sidoarjo

## I. PENDAHULUAN

Ciri-ciri negara demokratis ialah negara yang melibatkan masyarakat pada perencanaan maupun pelaksanaan pemilihan umum. Partisipasi masyarakat merupakan suatu aspek penting dalam tatanan negara demokrasi. Partisipasi politik dalam Pemilu berpengaruh terhadap legitimasi masyarakat kepada calon atau pasangan calon yang terpilih. Setiap masyarakat mempunyai preferensi dan kepentingan masing-masing untuk menentukan pilihan dalam pemilu. [1]

Masalah utama dalam upaya memprediksi pemilu di masa kini terkait munculnya fakta bahwa banyaknya milenial yang golput di seluruh wilayah Indonesia, khususnya di Kecamatan Wonokasian, dibuktikan dengan tingginya golput antar daerah. Golput juga merupakan hubungan dari sebab akibat) artinya tingkat kepedulian yang tinggi terjadi karena rendahnya pengetahuan dalam pemilu, rendahnya pengetahuan dalam pemilu terjadi karena kepedulian masyarakat akan kemajuan daerah yang juga rendah.

Metode yang digunakan ialah metode klasifikasi *Naïve Bayes Classifier*, yaitu salah satu teknik pengklasifikasian pada data mining. Yang mana akan dilakukan analisis untuk memperoleh sebuah informasi terhadap tingkat kehadiran dalam pemilu. Diharapkan dari penelitian terhadap sampel data penduduk pemilu ini dapat diperoleh sebuah informasi yang dapat membantu pihak kecamatan dalam merancang suatu strategi untuk meningkatkan kepedulian masyarakat.

Peneliti akan menguji klasifikasi kehadiran berdasarkan data penduduk pemilu dari Kecamatan Wonokasian tahun 2019 menggunakan metode *Naive Bayes*. Adapun Variabel yang digunakan pada klasifikasi pemilu tahun 2019 antara lain Jumlah KK, Asal Penduduk, Kategori, Status (Kawin/Belum Kawin), Alamat, RT, TPS, Lokasi, Hadir (keluarga), adapun hasil klasifikasinya yaitu menentukan tingkat kehadiran seperti : Tidak Hadir Semua, Tidak Semua Hadir, Hadir Semua dan Hadir.

Berdasarkan permasalahan di atas peneliti bermaksud menguji perhitungan weka tentang pemilu yang di khususkan untuk pemilu milenial dan pasca milenial di desa Wonokasian, dengan judul "PREDIKSI PEMILIH MILENIAL DAN PASCA MILENIAL DI PEMILU MENGGUNAKAN ALGORITMA NAVIE BAYES". Hasil dari penelitian ini berupa weka perhitungan sehingga memudahkan bagi masyarakat desa mengetahui data milenial antara yang golput maupun tidak golput.

#### **Data Mining**

Data Mining ialah sebuah proses pengumpulan data yang akan diolah menggunakan berbagai macam metode yang sesuai dengan data. Istilah lain dari data mining adalah *knowledge-discovery in database (KDD)* (Fayyad et al. 1996). Pengetahuan yang dapat diperoleh berupa sebuah pola data atau relasi tabel yang valid dan tidak diketahui sebelumnya (Suyanto, 2017). Tujuan dari penggunaan data mining adalah guna memanfaatkan data yang diperoleh dan mengolahnya hingga mendapatkan informasi yang baru dan berguna. [6]

#### **Klasifikasi**

Klasifikasi adalah proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui (Han, 2006). [6]

#### **Metode Naive Bayes**

*Bayesian classification* ialah suatu pengklasifikasian statistik yang digunakan untuk memprediksi probabilitas keanggotaan suatu class. *Bayesian classification* didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan neural network dan decision tree.

*Bayesian classification* terbukti memiliki akurasi dan kecepatan yang tinggi ketika diaplikasikan pada database dengan data yang besar. Teorema Bayes memiliki bentuk umum sebagai berikut : [8]

#### **Weka**

WEKA adalah paket tools machine learning yang praktis. "WEKA" ialah singkatan dari "Waikato Environment for Knowledge Analysis", yang dibuat di Universitas Waikato, New Zealand yang digunakan untuk pendidikan, penelitian dan bermacam-macam aplikasi. WEKA terdapat tools untuk preprocessing data, clustering, aturan asosiasi, klasifikasi, visualisasi dan regresi. [10]

## **II. METODE**

### **2.1 Teknik Pengumpulan Data**

Dalam memperoleh data yang dapat menunjang penelitian, maka peneliti melakukan pencarian dan pengumpulan data. Beberapa metode yang dibutuhkan dalam pengumpulan data ialah sebagai berikut:

#### **1. Studi Pustaka (Literatur)**

Dalam metode ini menggunakan literatur yang digunakan dalam pencarian data yang berhubungan dengan data pemilu pada desa Wonokasian Kabupaten Sidoarjo. Metode pencarian data ini diperoleh melalui jurnal, buku dan browsing di internet. Literatur yang didapatkan dalam memperoleh informasi ialah berupa teori dasar yang berkaitan dengan penelitian ini.

#### **2. Observasi**

Metode pengumpulan data berupa observasi yang dilaksanakan dengan pengamatan secara langsung di desa Wonokasian Kabupaten Sidoarjo berdasarkan objek penelitian dengan pengambilan data yang akan digunakan. Sehingga akan didapatkan data yang cukup relevan dan informasi yang lebih jelas. Data yang didapat oleh peneliti ialah data pemilu.

## **III. HASIL DAN PEMBAHASAN**

### **3.1 Flowchart Sistem Aplikasi**

Di bawah ini adalah gambaran dari flowchart metode *naive bayes* pada data testing dan data training untuk mengklasifikasi kehadiran masyarakat di desa Wonokasian Kabupaten Sidoarjo yaitu sebagai berikut:



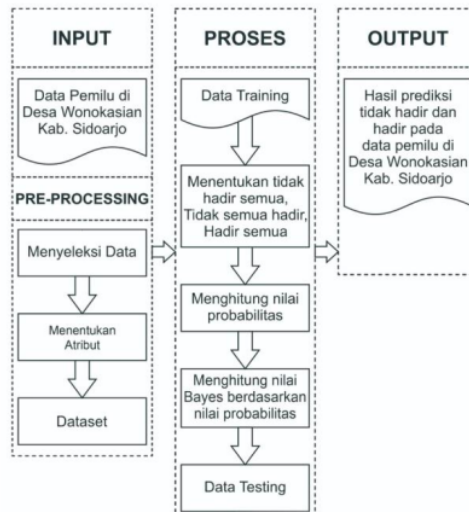
Gambar 3.1 Flowchart Proses *Naive Bayes*

Berdasarkan gambaran flowchart diatas, secara umum tahapan dari klasifikasi algoritma naïve bayes pada data training ialah sebagai berikut :

1. Membaca data training.
2. Menghitung jumlah dan probabilitas.
3. Menghitung nilai bayes berdasarkan probabilitas.
4. Membandingkan hasil perhitungan class pada tiap probabilitas
5. Mendapatkan output dari hasil klasifikasi
6. Input data testing
7. Menghitung data pada variabel X menggunakan hasil perhitungan probabilitas pada data training
8. Mendapatkan hasil output prediksi dari hasil klasifikasi.

### 3.2 Diagram Klasifikasi

Berikut ini Perancangan dari diagram klasifikasi menggunakan algoritma naïve bayes:



Gambar 3.2 Diagram Klasifikasi

### 3.7 Input

Data yang digunakan sebagai penelitian ialah data pemilu di desa Wonokasian Kabupaten Sidoarjo. Digunakan untuk memprediksi Tidak Hadir Semua, Hadir Semua, Tidak Hadir, dan Hadir di desa Wonokasian saat pemilu ialah sebanyak 4.216 data yang diperoleh pada tanggal 17 April 2019. Penggunaan dataset sejumlah 100 dataset yang mana akan diambil 70% untuk dijadikan sebagai data training dan 30% untuk dijadikan sebagai data testing. Data ini akan diolah menggunakan metode klasifikasi Naïve Bayes. Selain itu pada tahap ini akan dilakukan pengujian terhadap 10% dari 100 dataset untuk mengklasifikasi Tidak Semua Hadir.

**Tabel 1. Data Pemilu**

No	Jml KK	Asal Penduduk	Kategori Masy	Kawin	jenis_kelamin	alamat	rt	Tps	loka si	hadir kel	hadir
1	Dua	SIDOARJO	Milenial	S	P	WONOKASIAN	rt01	TPS-1	dalam	TSH	TH
2	Dua	SIDOARJO	Milenial	S	L	WONOKASIAN	rt01	TPS-1	dalam	TSH	H
3	Tiga	LUAR	Milenial	B	L	WONOKASIAN	rt01	TPS-1	dalam	THS	TH
4	Tiga	LUAR	PascaMilenial	S	P	WONOKASIAN	rt01	TPS-1	dalam	THS	TH
5	Tiga	LUAR	PascaMilenial	S	L	WONOKASIAN	rt01	TPS-1	dalam	THS	H
6	Dua	SIDOARJO	Milenial	S	L	WONOKASIAN	rt01	TPS-1	dalam	HS	H
7	Dua	SIDOARJO	Milenial	S	P	WONOKASIAN	rt01	TPS-1	dalam	HS	H
8	Dua	SIDOARJO	Milenial	S	L	WONOKASIAN	rt01	TPS-1	dalam	HS	H
9	Dua	SIDOARJO	Milenial	S	P	WONOKASIAN	rt01	TPS-1	dalam	HS	H
10	Dua	SIDOARJO	Milenial	S	L	WONOKASIAN	rt01	TPS-1	dalam	HS	H
..	..	..	..	..	..	..	..	..	..	..	..

### 3.7.2 Data Training

Data sebanyak 100 *record* dataset berikut akan dilakukan proses perhitungan secara manual dengan menggunakan data sebanyak 70 *record* dataset menggunakan metode klasifikasi dengan algoritma naïve bayes. Berikut adalah pengambilan data training untuk perhitungan manual.

**Tabel 2. Data Training 70**

jml KK	Asal Penduduk	Kategori Masy	kawin	jenis_kelamin	alamat	rt	tps	lokasi	hadir kel	hadir	
tiga	SIDOARJO	PascaMilenial	Sudah	Laki-Laki	WONOKASIAN	rt01	TPS-1	dalam	Tidak Hadir	Semua Tidak Hadir	
tiga	SIDOARJO	Milenial	Sudah	Perempuan	WONOKASIAN	rt01	TPS-1	dalam	Tidak Hadir	Semua Tidak Hadir	
tiga	SIDOARJO	Pemula	Belum	Perempuan	WONOKASIAN	rt01	TPS-1	dalam	Tidak Hadir	Semua Tidak Hadir	
dua	SIDOARJO	Milenial	Sudah	Perempuan	WONOKASIAN	rt01	TPS-1	dalam	Hadir	Semua Hadir	
dua	SIDOARJO	PascaMilenial	Sudah	Laki-Laki	WONOKASIAN	rt01	TPS-1	dalam	Hadir	Semua Hadir	
tiga	SIDOARJO	Milenial	Belum	Laki-Laki	WONOKASIAN	rt01	TPS-1	dalam	Hadir	Semua Hadir	
tiga	SIDOARJO	Milenial	Sudah	Perempuan	WONOKASIAN	rt01	TPS-1	dalam	Hadir	Semua Hadir	
tiga	SIDOARJO	PascaMilenial	Sudah	Laki-Laki	WONOKASIAN	rt01	TPS-1	dalam	Hadir	Semua Hadir	
dua	SIDOARJO	PascaMilenial	Sudah	Perempuan	WONOKASIAN	rt01	TPS-1	dalam	Tidak	Semua Hadir	Hadir
dua	SIDOARJO	PascaMilenial	Sudah	Laki-Laki	WONOKASIAN	rt01	TPS-1	dalam	Tidak	Semua Hadir	Tidak Hadir
..	..	..	..	..	..	..	..	..	..	..	

### 3.7.3 Data Testing

Pada data testing yang diambil dari data pemilu di desa Wonokasian, sebanyak 30 *record* data akan dilakukan proses prediksi dan presentase benar maupun salahnya hasil dari prediksi yang diperoleh dari proses perhitungan manual pada data training sebanyak 70 *record* data pemilu. Berikut ini ialah contoh data testing pada tabel 3.

**Tabel 3. Data Testing 30**

jml KK	Asal Penduduk	Kategori Masy	kawin	jenis_kelamin	alamat	rt	tps	lokasi	hadir kel	hadir
tiga	SIDOARJO	PascaMilenial	Sudah	Laki-Laki	WONOKASIAN	rt01	TPS-1	dalam	Tidak Hadir	Semua ?
tiga	SIDOARJO	Milenial	Sudah	Perempuan	WONOKASIAN	rt01	TPS-1	dalam	Tidak Hadir	Semua ?
tiga	SIDOARJO	Pemula	Belum	Perempuan	WONOKASIAN	rt01	TPS-1	dalam	Tidak Hadir	Semua ?
dua	SIDOARJO	Milenial	Sudah	Perempuan	WONOKASIAN	rt01	TPS-1	dalam	Hadir	Semua ?
dua	SIDOARJO	PascaMilenial	Sudah	Laki-Laki	WONOKASIAN	rt01	TPS-1	dalam	Hadir	Semua ?
tiga	SIDOARJO	Milenial	Belum	Laki-Laki	WONOKASIAN	rt01	TPS-1	dalam	Hadir	Semua ?
tiga	SIDOARJO	Milenial	Sudah	Perempuan	WONOKASIAN	rt01	TPS-1	dalam	Hadir	Semua ?
tiga	SIDOARJO	PascaMilenial	Sudah	Laki-Laki	WONOKASIAN	rt01	TPS-1	dalam	Hadir	Semua ?
dua	SIDOARJO	PascaMilenial	Sudah	Perempuan	WONOKASIAN	rt01	TPS-1	dalam	Tidak	Semua Hadir ?
dua	SIDOARJO	PascaMilenial	Sudah	Laki-Laki	WONOKASIAN	rt01	TPS-1	dalam	Tidak	Semua Hadir ?
..	..	..	..	..	..	..	..	..	..	..

### 3.8 Proses

Pada tahap ini, proses dilakukan dengan cara sebagai berikut:

1. Menghitung data training, Melakukan perhitungan manual pada data training.
2. Menentukan kehadiran pada data pemilu. Data yang nantinya diolah digunakan untuk menentukan hadir maupun tidaknya pelayanan pada data pemilu
3. Menghitung nilai probabilitas  
 Menghitung nilai probabilitas pada setiap class dan seluruh variabel
4. Menghitung nilai bayes berdasarkan nilai probabilitas  
 Menghitung nilai bayes dari setiap hasil perhitungan dari nilai probabilitas pada seluruh variable
5. Menghitung prediksi menggunakan data testing  
 Menghitung prediksi hadir maupun tidak hadir kehadiran pada data pemilu menggunakan data testing.

### 3.9 Output

Output ialah hasil akhir yang didapatkan setelah melakukan tahap proses. Dalam penelitian ini memiliki output berupa mengetahui prediksi hadir maupun tidaknya peserta pada data pemilu di desa Wonokasian. Berikut ini ialah hasil perbandingan dari prediksi hadir dan tidak hadir pada data pemilu.

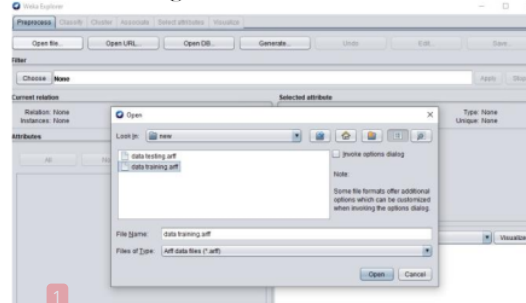
### 3.10 Halaman Awal Weka



Gambar 3.3 Halaman Awal Weka

Halaman berikut merupakan halaman awal weka. Terdapat lima menu pada tools weka yaitu diantaranya adalah Explorer, Experimenter, KnowledgeFlow, Workbench dan Simple CLI. Data tersebut nantinya diuji dalam menu Explorer.

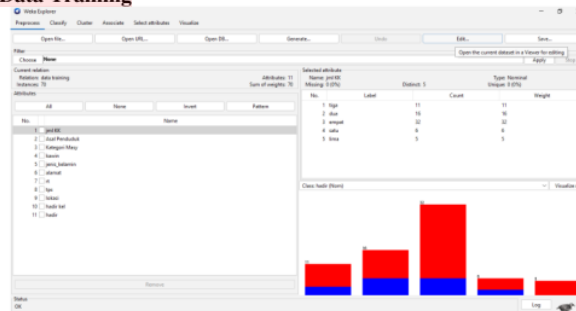
### 3.12 Halaman Menu Explorer Data Training



Gambar 3.4 Halaman Menu Explorer Data Training

Di halaman menu explorer terdapat lima menu, diantaranya ialah Preprocess, Classify, Cluster, Associate, Select attribute dan Visualize. Dalam tahap pengujian data dilakukan dalam menu Preprocess yaitu dengan cara membuka file di Open File, kemudian masukkan data training.

### 3.13 Halaman Atribut Data Training

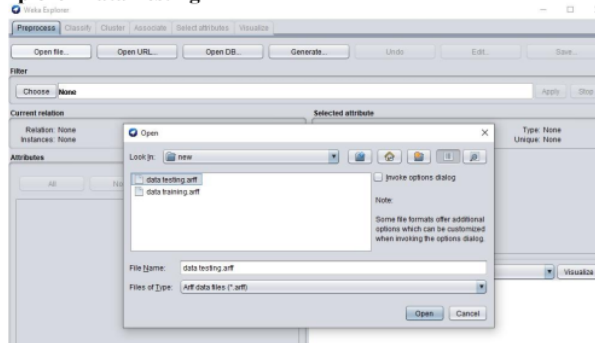


Gambar 3.5 Halaman Edit Viewer



Pada tahap berikut ini pilih Start untuk melihat hasil klasifikasi dari data 70 data training. Dari hasil pengujian diperoleh tingkat kebenaran sebesar 91,42% pada *Correctly Classified Instances* dan tingkat kesalahan sebesar 8,57% pada *Incorrect Classified Instances*.

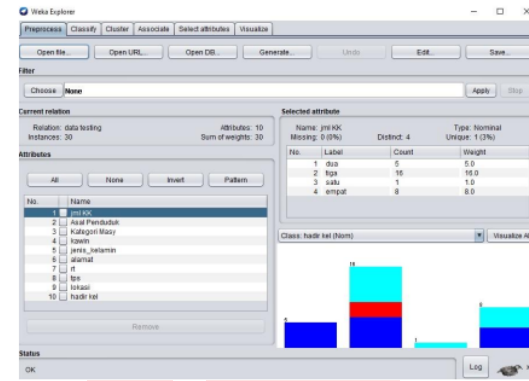
### 3.16 Halaman Menu Explorer Data Testing



Gambar 3.9 Halaman Explorer Data Testing

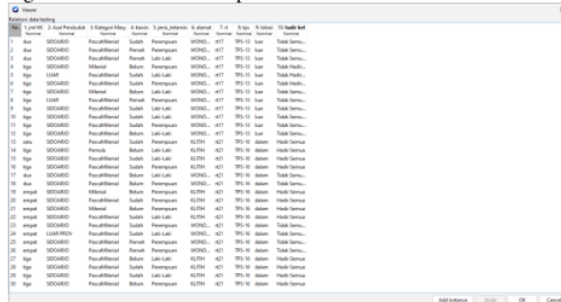
Dalam explorer terdapat lima menu, antara lain Preprocess, Classify, Cluster, Associate, Select attribute dan Visualize. Dalam tahap pengujian data dilakukan dalam menu Preprocess yaitu dengan cara membuka file di Open File, lalu masukkan data yang akan diuji yaitu data testing.

### 3.17 Halaman Atribut Data Testing



Gambar 3.10 Halaman Atribut Data Testing

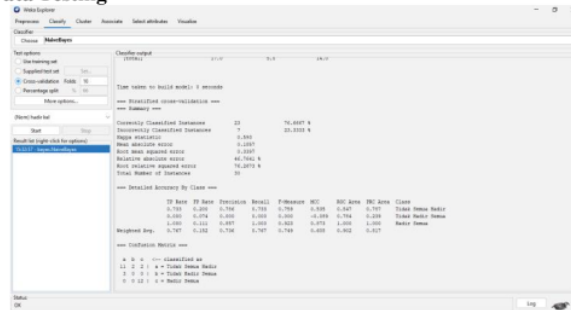
Setelah open file dari data testing, lalu tampil atribut yang ada pada data pemilu. Di halaman atribut terdapat 10 atribut, antara lain Jumlah KK, Asal Penduduk, Kategori Masyarakat, Kawin, Jenis Kelamin, Alamat, RT, TPS, Lokasi, Hadir Kel. Class yang digunakan untuk pengujian yaitu Jumlah kehadiran setiap KK dengan nilai seperti yang ditampilkan di gambar 4.9 dari 30 data pemilu.



Gambar 3.11 Viewer Atribut Data Testing

Pada gambar 3.11 berupa Viewer yang berfungsi melihat atribut dari data testing yang sudah diubah menjadi format arff. Di halaman ini menampilkan atribut dari data testing yaitu sebanyak 30 data.

### 3.18 Hasil Klasifikasi Data Testing



Gambar 3.12 Hasil Klasifikasi Data Testing

Pada tahap berikut ini pilih Start untuk melihat hasil klasifikasi dari data 30 data testing. Dari hasil pengujian diperoleh tingkat kebenaran sebesar 76,66% pada *Correctly Classified Instances* dan tingkat kesalahan sebesar 23,33% pada *Incorrect Classified Instances*. Pada tahap berikut ini pilih Start untuk melihat hasil klasifikasi dari data 5 data testing. Dari hasil pengujian diperoleh tingkat kebenaran sebesar 100% pada *Correctly Classified Instances* dan tingkat kesalahan sebesar 0% pada *Incorrect Classified Instances*.

### 3.19 Hasil Presentase

Dari hasil pengolahan data yang diperoleh dari data training sebanyak 70% dan data testing sebanyak 30% berdasarkan dataset sebanyak 100 data dengan menggunakan perhitungan manual dan weka ialah sebagai berikut:

Tabel 4. Hasil Persentase Pengujian 100 Dataset

Pengujian	Dataset	Training	Testing	Benar	% Benar	Salah	% Salah
Weka	100	70	30	23	76,67%	7	23,33%
Manual	100	70	30	24	80%	6	20%

Pada 100 dataset menggunakan weka diperoleh prediksi kebenaran sebesar 76,67% dan prediksi kesalahan sebesar 23,33%. Pada dataset 100 menggunakan perhitungan manual diperoleh prediksi kebenaran sebesar 80% dan prediksi kesalahan sebesar 20%.

Berdasarkan hasil klasifikasi menggunakan metode Naïve Bayes untuk mengklasifikasi status kehadiran pada 100 dataset, lalu diperoleh perhitungan hasil tertinggi menggunakan perhitungan manual dengan perolehan persentase sebesar 80%.

Dari hasil perbandingan yang diperoleh dari data training sebanyak 70% dan data testing sebanyak 30% berdasarkan dataset sebanyak 100 data dengan menggunakan perhitungan manual dan weka ialah sebagai berikut:

## VI. KESIMPULAN

Berdasarkan hasil dari klasifikasi menggunakan metode Naïve Bayes untuk mengklasifikasi status kehadiran dari 100 dataset, lalu diperoleh hasil tertinggi menggunakan perhitungan manual dengan perolehan tingkat kebenaran sebesar 76,67% dan tingkat kesalahan sebesar 23,33%.

## REFERENSI

- [1] W.-K. Chen, Linear Networks and Systems. Belmont, CA: Wadsworth, 1993, pp. 123-135.
- [2] R. Hayes, G. Pisano, D. Upton, and S. Wheelwright, *Operations, Strategy, and Technology: Pursuing the competitive edge*. Hoboken, NJ: Wiley, 2005.
- [3] The Oxford Dictionary of Computing, 5th ed. Oxford: Oxford University Press, 2003.
- [4] A. Rezi and M. Allam, "Techniques in array processing by means of transformations," in *Control and Dynamic Systems*, Vol. 69, Multidimensional Systems, C. T. Leondes, Ed. San Diego: Academic Press, 1995, pp. 133-180.
- [5] O. B. R. Strimpel, "Computer graphics," in *McGraw-Hill Encyclopedia of Science and Technology*, 8th ed., Vol. 4. New York: McGraw-Hill, 1997, pp. 279-283.

# CEK\_PLAGIASI\_JURNAL SISTA FAIZZAH SENASAINS 6

## ORIGINALITY REPORT

20%

SIMILARITY INDEX

15%

INTERNET SOURCES

11%

PUBLICATIONS

5%

STUDENT PAPERS

## PRIMARY SOURCES

1

[proceeding.unpkediri.ac.id](http://proceeding.unpkediri.ac.id)

Internet Source

10%

2

[www.neliti.com](http://www.neliti.com)

Internet Source

5%

3

A S Fitriani, M A Rosid, F Muharram, F L Kodriyah. "Attribute analysis with classification algorithm on election participation", IOP Conference Series: Materials Science and Engineering, 2020

Publication

5%

Exclude quotes Off

Exclude matches < 5%

Exclude bibliography Off