

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 Rohma¹⁾, Arif Senja Fitriani²⁾, Mochamad Alfian Rosid³⁾, Ade Eviyanti⁴⁾

¹⁾ Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

²⁾ Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

³⁾ Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

⁴⁾ Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

*Email Penulis Korespondensi: asfjim@umsida.ac.id

Abstract. *The main problem in efforts to predict the current elections is related to the fact that many millennials are abstaining throughout 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 method Naïve Bayes. 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 classification results on the Naïve Bayes method for classifying attendance status from 100 datasets, the highest result was 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 negara demokratis ialah negara yang melibatkan masyarakat dalam perencanaan maupun pelaksanaan pemilihan umum. Partisipasi masyarakat (pemilih) merupakan aspek penting dalam tatanan negara demokrasi. Misalnya dalam Pemilu partisipasi politik berpengaruh terhadap legitimasi masyarakat kepada calon atau pasangan calon yang terpilih. Setiap masyarakat memiliki preferensi dan kepentingan masing-masing untuk menentukan pilihan mereka dalam pemilu. [1]

Tujuan penulis pada bagian ini adalah berusaha menempatkan hakekat masyarakat (pemilih) dalam sistem pemerintahan yang demokratis. Selanjutnya akan diungkap pada langkah berikutnya bagaimana metode riset dan alur analisis yang digunakan dalam riset ini. Masalah utama dalam upaya memprediksi pemilu saat ini terkait adanya fakta bahwa banyaknya milenial yang golput di seluruh wilayah Indonesia, khususnya di Kecamatan Wonokasian, ini dibuktikan dengan tingginya golput antar daerah. Selain itu golput juga merupakan sebuah hubungan sebab akibat (kausalitas melingkar) 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 akan digunakan adalah metode *Naïve Bayes Classifier*, yang merupakan salah satu teknik pengklasifikasian dalam data mining. Dimana akan dilakukan analisis untuk memperoleh informasi terhadap data tingkat kehadiran dalam pemilu. Diharapkan dari penelitian yang dilakukan terhadap sampel data penduduk pemilu tersebut dapat diperoleh suatu informasi yang bisa membantu pihak kecamatan untuk merancang strategi dalam meningkatkan kepedulian masyarakat.

Penelitian akan menguji klasifikasi berdasarkan data penduduk pemilu yang diperoleh dari Kecamatan Wonokasian tahun 2019 dengan menggunakan metode *Naïve Bayes*. Variabel yang akan digunakan dalam melakukan klasifikasi pemilu tahun 2019 adalah Jumlah KK, Asal Penduduk, Kategori, Status (Kawin/Belum Kawin), Alamat, RT, TPS, Lokasi, Hadir (keluarga), Hadir sesuai data yang telah diambil dan sesuai dengan variabel yang akan diinputkan, maka hasil klasifikasinya nanti akan menentukan tingkat kehadiran seperti : Tidak Hadir Semua, Tidak Semua Hadir, Hadir Semua dan Hadir. Hasil dari penelitian ini berupa weka perhitungan sehingga memudahkan bagi masyarakat desa mengetahui data milenial antara yang golput maupun tidak golput.

A. Generasi Millennial

Generasi milenial merupakan generasi yang bertumbuh dan berkembang pada masa internet booming (Lyson, 2004; Putra, 2019). Generasi milenial sangat familiar dengan penggunaan teknologi, seperti handphone, komputer, laptop, maupun internet. pada tahun 2019 jumlah penduduk di Indonesia telah mencapai sekitar 271 juta jiwa (Worldometers, 2019). Sebanyak 31% dari jumlah penduduk di Indonesia yang merupakan generasi milenial (Worldometers, 2019). Angka ini menunjukkan bahwa generasi milenial mendominasi jumlah penduduk di Indonesia. [2]

B. Pemilihan Umum

Sesuai Undang-undang Nomor 8 Tahun 2012 Tentang Pemilihan Umum Anggota Dewan Perwakilan Rakyat, Dewan Perwakilan Daerah dan Dewan Perwakilan Rakyat Daerah pengertian pemilihan umum diuraikan secara detail. Pemilu ialah sarana pelaksanaan kedaulatan rakyat yang dilaksanakan secara langsung, umum, bebas, rahasia, jujur, dan adil dalam Negara Kesatuan Indonesia berdasarkan Pancasila dan Undang-undang Dasar Negara Republik Indonesia 1945. Dengan kata lain, pemilu merupakan sarana bagi rakyat untuk menjalankan kedaulatan dan merupakan lembaga demokrasi. [3]

C. Data Mining

Data Mining merupakan proses pengumpulan data yang diolah dengan berbagai macam metode yang sesuai. 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 data mining yaitu untuk memanfaatkan data dan mengolahnya hingga mendapatkan sebuah informasi yang baru dan berguna. [4]

D. Flowchart

Flowchart adalah sebuah teknik analitis yang digunakan untuk menjelaskan aspek-aspek sistem informasi dengan jelas, tepat dan logis (Krismiaji, 2005). [5] Secara umum, pengertian flowchart merupakan gambaran dari langkah-langkah dan urutan prosedur sebuah program secara grafik. Flowchart digunakan untuk analisis dan programmer dalam menyelesaikan masalah dengan membaginya ke dalam segmen-segmen yang lebih kecil dan digunakan untuk menganalisis alternatif-alternatif lain dalam pengoperasian.

II. PENELITIAN TERDAHULU

Nama Penulis	Judul	Keterangan
(Ratih Yulia Hayuningtyas, 2019) [6]	PENERAPAN ALGORITMA NAÏVE BAYES UNTUK REKOMENDASI PAKAIAN WANITA	Memberikan rekomendasi keputusan dalam menentukan pakaian wanita
(Haditsah Annur, 2018) [7]	KLASIFIKASI MASYARAKAT MISKIN MENGGUNAKAN METODE NAÏVE BAYES	Memprediksi dalam memberikan bantuan masyarakat miskin

(Arif Senja Fitriani, 2019) [8]	PENERAPAN DATA MINING MENGUNAKAN METODE KLASIFIKASI NAÏVE BAYES UNTUK MEMPREDIKSI PARTISIPASI PEMILIHAN GUBERNUR	Memprediksi hasil partisipasi pemilu dilihat dari prediksi kebenaran dan prediksi kesalahannya.
-------------------------------------	---	---

Penelitian ini memiliki kesamaan dengan penelitian yang sebelumnya terletak pada metode penelitian yang digunakan. Dari ke tiga penelitian yang sudah ada menggunakan metode naïve bayes, yang merupakan pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Sedangkan tujuan penelitian ini yaitu bagaimana dapat mengetahui jumlah warga yang tidak hadir maupun yang hadir dengan data yang sudah ada, sehingga dapat memprediksi hadir atau tidaknya warga saat pemilu berlangsung menggunakan metode *naïve bayes*.

III. METODE

A. 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). [9] Klasifikasi data terdiri dari 2 langkah proses. Pertama adalah *learning* (fase *training*), dimana algoritma klasifikasi dibuat untuk menganalisa data *training* lalu direpresentasikan dalam bentuk *rule* klasifikasi. Proses kedua adalah klasifikasi, dimana data tes digunakan untuk memperkirakan akurasi dari *rule* klasifikasi.

B. Metode Naïve Bayes

Bayesian classification ialah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. *Bayesian classification* didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan decision tree dan neural network. *Bayesian classification* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar. Metode Bayes merupakan pendekatan statistic untuk melakukan inferensi induksi pada persoalan klasifikasi. Teorema Bayes memiliki bentuk umum sebagai berikut : [10]

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

Keterangan :

X = Data dengan class yang belum diketahui

H = Hipotesis data X merupakan suatu class spesifik

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

P(H) = Probabilitas hipotesis H (prior prob.)

P(X|H) = Probabilitas X berdasarkan kondisi tersebut

P(X) = Probabilitas dari X

C. Weka

WEKA ialah sebuah paket tools machine learning praktis. “WEKA” merupakan singkatan dari “Waikato Environment for Knowledge Analysis”, yang dibuat di Universitas Waikato, New Zealand untuk penelitian, pendidikan dan berbagai aplikasi. WEKA mengandung tools untuk preprocessing data, klasifikasi, regresi, clustering, aturan asosiasi, dan visualisasi. [11] Contoh penggunaan WEKA ialah dengan menerapkan sebuah metode pembelajaran ke dataset dan menganalisis hasilnya untuk memperoleh informasi tentang data atau menerapkan beberapa metode dan membandingkan performanya untuk dipilih.

D. Lokasi Dan Waktu Penelitian

Penelitian ini dilaksanakan pada bulan Oktober hingga Desember 2020 dengan melakukan pengumpulan data pada desa Wonokasian di Kabupaten Sidoarjo. Penelitian dilakukan setelah mendapat persetujuan dari dosen pembimbing. Data yang digunakan sebagai penelitian ialah data pemilu di desa Wonokasian Kabupaten Sidoarjo.

E. 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)

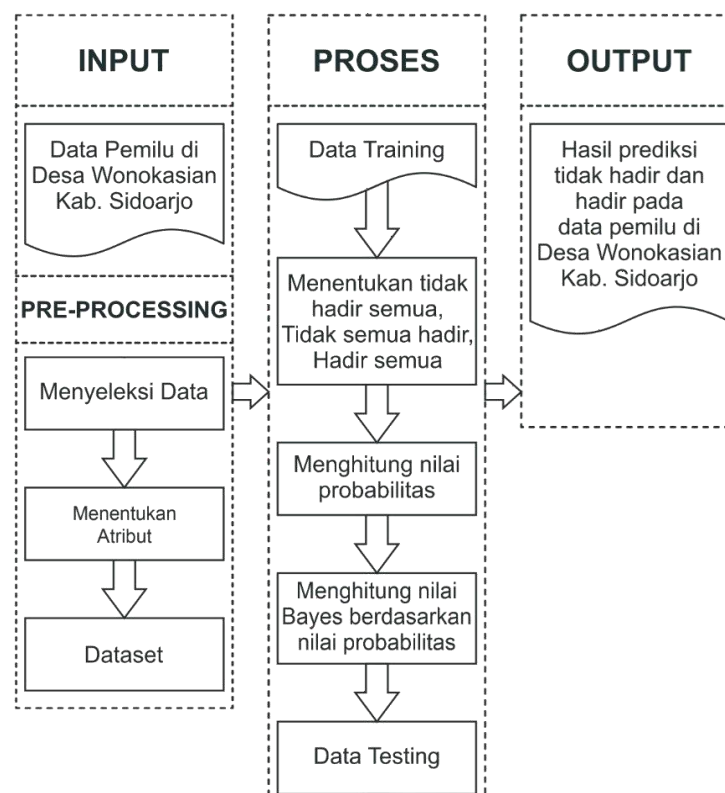
Dalam metode ini menggunakan literatur yang digunakan dalam pencarian data yang berhubungan dengan data pemilu pada desa Wonokasian Kabupaten Sidoarjo. Metode pencarian data ini diperoleh melalui jurnal, buku dan browsing di internet. Literatur yang didapatkan dalam memperoleh informasi ialah berupa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

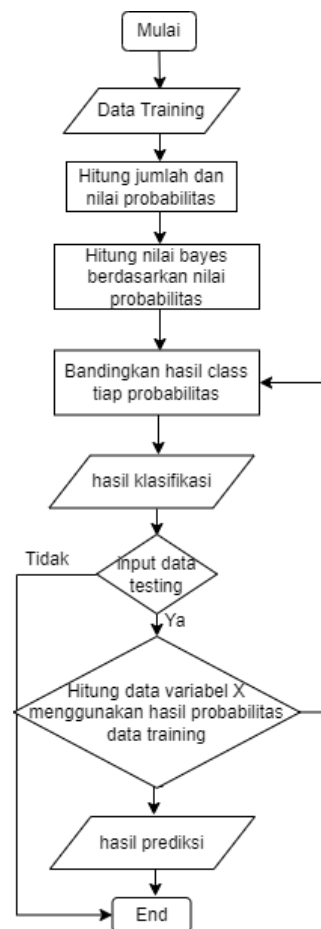
F. Diagram Klasifikasi

Perancangan diagram klasifikasi dengan menggunakan algoritma naïve bayes yaitu sebagai berikut:



Gambar 1. Diagram Klasifikasi

G. Flowchart



Gambar 2. Flowchart

IV. HASIL DAN PEMBAHASAN

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 data 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 data untuk mengklasifikasi Tidak Semua Hadir.

Tabel 1. Dataset 100

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
..

Tabel 2. Data Training 70

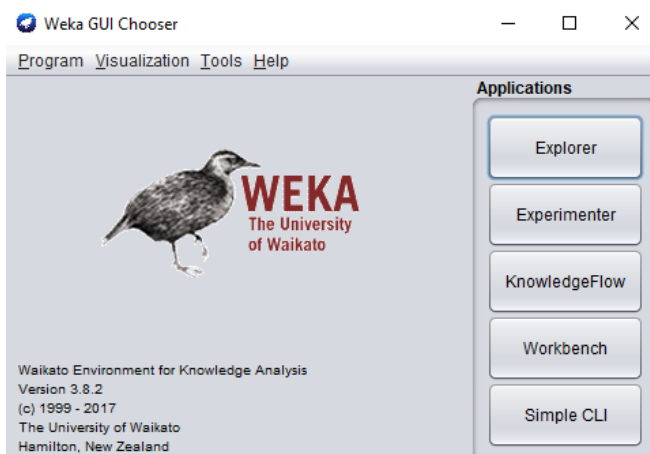
jml KK	Asal Penduduk	Kategori Masy kawin	jenis_kelamin	alamat	rt	tps	lokasi	hadir kel	hadir
tiga	SIDOARJO	PascaMilennial	Sudah Laki-Laki	WONOKASIAN	rt01	TPS-1	dalam	Tidak Hadir Semua	Tidak Hadir
tiga	SIDOARJO	Milennial	Sudah Perempuan	WONOKASIAN	rt01	TPS-1	dalam	Tidak Hadir Semua	Tidak Hadir
tiga	SIDOARJO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

Tabel 3. Data Testing 30

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

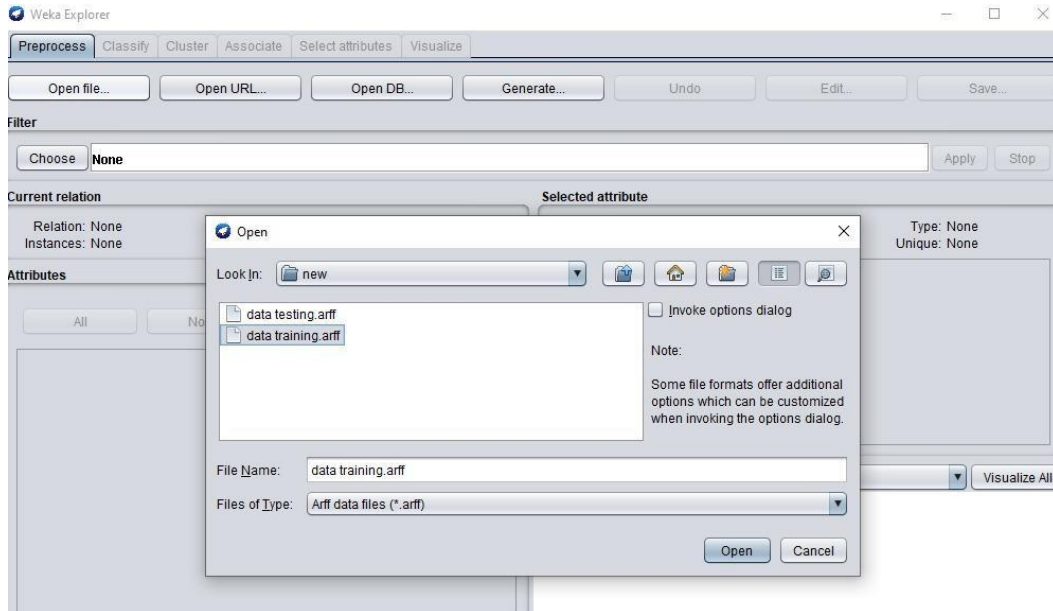
Tabel 4. Data Training 5 Dataset**Tabel 5.** Data Testing 5 Dataset

A. Halaman Awal Weka

**Gambar 2.** Halaman Awal Weka

Halaman ini ialah halaman saat membuka weka. Terdapat 5 menu pada weka diantaranya ialah Explorer, Experimenter, KnowledgeFlow, Workbench dan Simple CLI. Data tersebut akan diuji pada menu Explorer.

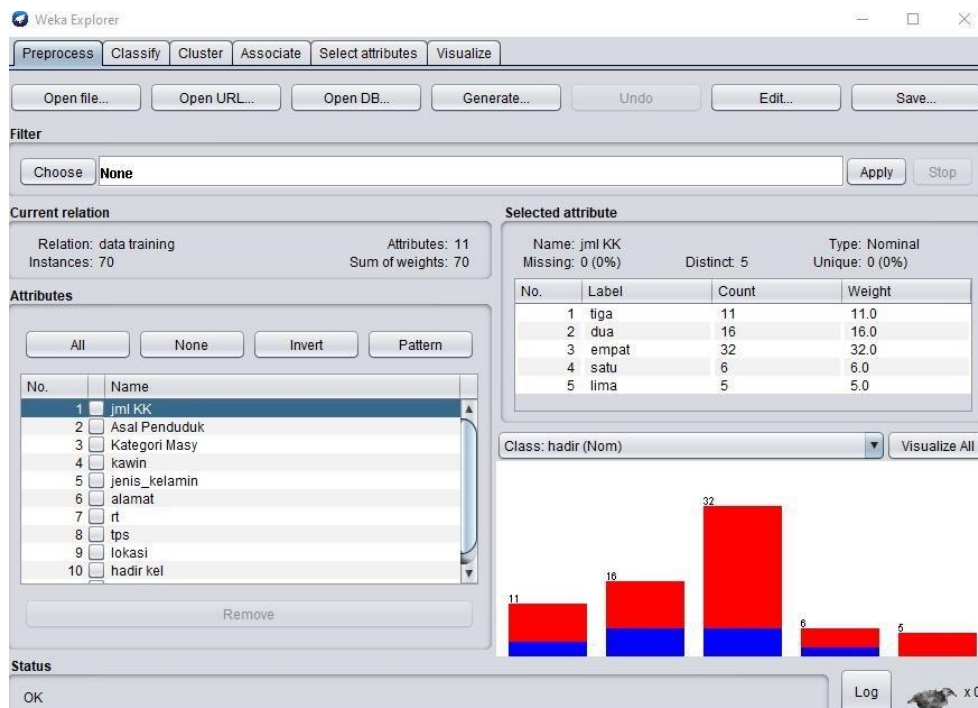
B. Halaman Menu Explorer Data Training



Gambar 3. Halaman Menu Explorer Data Training

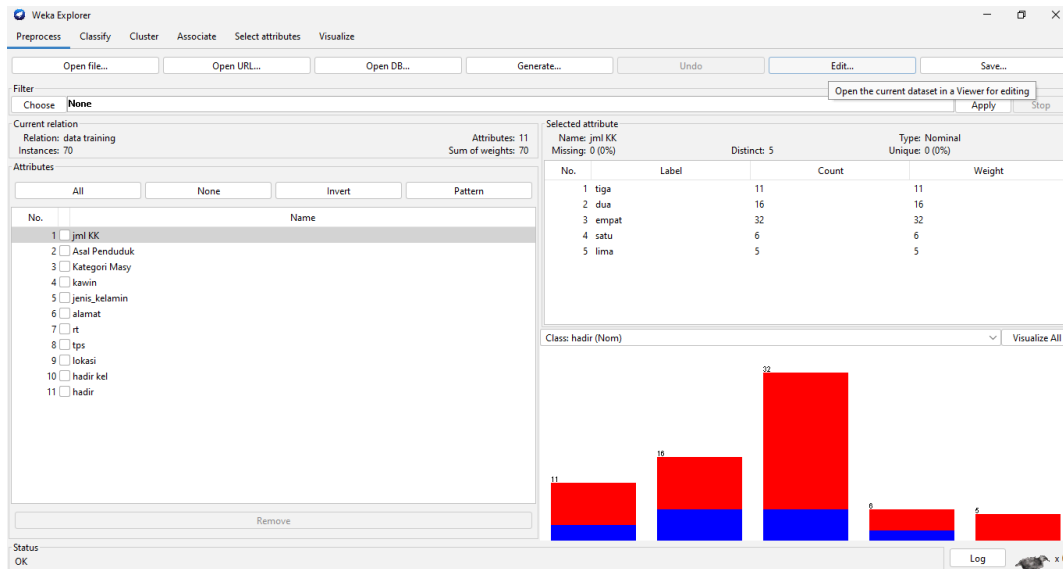
Pada halaman menu explorer terdapat 5 menu pilihan, diantaranya yaitu Preprocess, Classify, Cluster, Associate, Select attribute dan Visualize. Pada tahap pengujian ini akan dilakukan pada menu Preprocess dengan membuka file pada Open File, lalu masukkan data yang akan diuji yaitu data training.

C. Halaman Atribut Data Training



Gambar 4. Halaman Atribut Data Training

Setelah memasukkan file, akan muncul atribut yang terdapat dalam data pemilu. Pada halaman atribut terdapat 11 atribut yang digunakan, diantaranya ialah Jumlah KK, Asal Penduduk, Kategori Masyarakat, Kawin, Jenis Kelamin, Alamat, RT, TPS, Lokasi, Hadir Kel, Hadir. Class yang digunakan untuk pengujian yaitu Jumlah kehadiran setiap KK dengan nilai seperti yang ditampilkan di gambar 4.3 dari 70 data pemilu.



Gambar 5. Halaman Edit Viewer

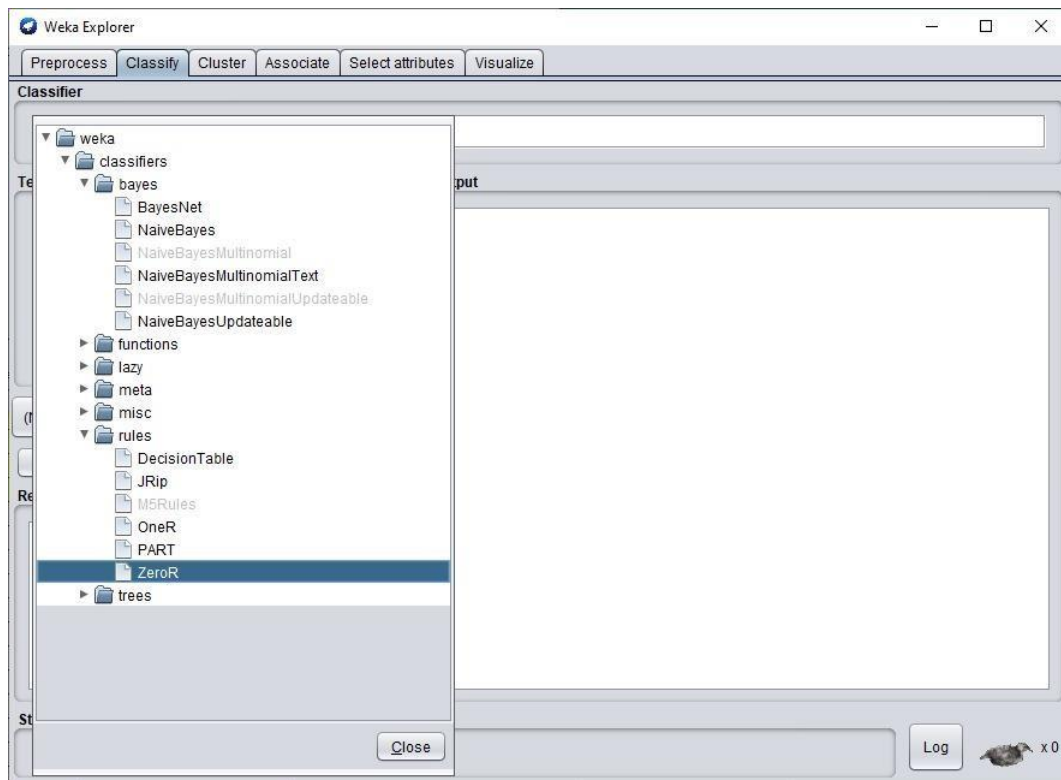
Pada gambar 5. untuk melihat atribut data training yang akan diuji, masuk ke menu edit lalu klik akan muncul atribut viewer pada atribut data training.

No.	1: jml KK	2: Asal Penduduk	3: Kategori Masy	4: kawin	5: jenis_kelamin	6: alamat	7: rt	8: tps	9: lokasi	10: hadir_kel	11: hadir
1	tiga	SIDOARIO	PascaMilennial	Sudah	Laki-Laki	WONO...	rt01	TPS-1	dalam	Tidak Hadi...	Tidak H...
2	tiga	SIDOARIO	Milennial	Sudah	Perempuan	WONO...	rt01	TPS-1	dalam	Tidak Hadi...	Tidak H...
3	tiga	SIDOARIO	Pemula	Belum	Perempuan	WONO...	rt01	TPS-1	dalam	Tidak Hadi...	Tidak H...
4	dua	SIDOARIO	Milennial	Sudah	Perempuan	WONO...	rt01	TPS-1	dalam	Hadir Semua	Hadir
5	dua	SIDOARIO	PascaMilennial	Sudah	Laki-Laki	WONO...	rt01	TPS-1	dalam	Hadir Semua	Hadir
6	tiga	SIDOARIO	Milennial	Belum	Laki-Laki	WONO...	rt01	TPS-1	dalam	Hadir Semua	Hadir
7	tiga	SIDOARIO	Milennial	Sudah	Perempuan	WONO...	rt01	TPS-1	dalam	Hadir Semua	Hadir
8	tiga	SIDOARIO	PascaMilennial	Sudah	Laki-Laki	WONO...	rt01	TPS-1	dalam	Hadir Semua	Hadir
9	dua	SIDOARIO	PascaMilennial	Sudah	Perempuan	WONO...	rt01	TPS-1	dalam	Tidak Sem...	Hadir
10	dua	SIDOARIO	PascaMilennial	Sudah	Laki-Laki	WONO...	rt01	TPS-1	dalam	Tidak Sem...	Tidak H...
11	empat	SIDOARIO	Milennial	Belum	Laki-Laki	WONO...	rt01	TPS-1	dalam	Hadir Semua	Hadir
12	empat	SIDOARIO	Milennial	Sudah	Laki-Laki	WONO...	rt01	TPS-1	dalam	Hadir Semua	Hadir
13	empat	SIDOARIO	PascaMilennial	Sudah	Laki-Laki	WONO...	rt01	TPS-1	dalam	Hadir Semua	Hadir
14	empat	SIDOARIO	PascaMilennial	Sudah	Perempuan	WONO...	rt01	TPS-1	dalam	Hadir Semua	Hadir
15	tiga	SIDOARIO	PascaMilennial	Sudah	Laki-Laki	WONO...	rt01	TPS-1	dalam	Hadir Semua	Hadir
16	tiga	SIDOARIO	Milennial	Belum	Laki-Laki	WONO...	rt01	TPS-1	dalam	Hadir Semua	Hadir
17	tiga	SIDOARIO	PascaMilennial	Sudah	Perempuan	WONO...	rt01	TPS-1	dalam	Hadir Semua	Hadir
18	empat	SIDOARIO	Milennial	Sudah	Laki-Laki	WONO...	rt01	TPS-1	dalam	Hadir Semua	Hadir
19	empat	SIDOARIO	PascaMilennial	Sudah	Perempuan	WONO...	rt01	TPS-1	dalam	Hadir Semua	Hadir
20	empat	LUAR	PascaMilennial	Sudah	Laki-Laki	WONO...	rt01	TPS-1	dalam	Hadir Semua	Hadir
21	empat	SIDOARIO	Pemula	Belum	Perempuan	WONO...	rt01	TPS-1	dalam	Hadir Semua	Hadir
22	dua	SIDOARIO	PascaMilennial	Sudah	Perempuan	WONO...	rt04	TPS-2	dalam	Tidak Sem...	Hadir
23	empat	SIDOARIO	PascaMilennial	Pernah	Perempuan	WONO...	rt04	TPS-2	dalam	Hadir Semua	Hadir
24	empat	SIDOARIO	Milennial	Belum	Perempuan	WONO...	rt04	TPS-2	dalam	Hadir Semua	Hadir
25	empat	SIDOARIO	Milennial	Belum	Laki-Laki	WONO...	rt04	TPS-2	dalam	Hadir Semua	Hadir
26	empat	SIDOARIO	Pemula	Belum	Laki-Laki	WONO...	rt04	TPS-2	dalam	Hadir Semua	Hadir
27	satu	LUAR PROV	PascaMilennial	Sudah	Laki-Laki	WONO...	rt04	TPS-2	dalam	Hadir Semua	Hadir
28	satu	SIDOARIO	PascaMilennial	Sudah	Laki-Laki	WONO...	rt04	TPS-2	dalam	Hadir Semua	Hadir
29	satu	SIDOARIO	Milennial	Sudah	Perempuan	WONO...	rt04	TPS-2	dalam	Tidak Hadi...	Tidak H...
30	satu	LUAR	PascaMilennial	Pernah	Perempuan	WONO...	rt04	TPS-2	dalam	Tidak Sem...	Tidak H...

Gambar 6. Halaman Viewer Atribut Data Training

Pada gambar 6. menampilkan halaman Viewer yang digunakan untuk melihat atribut data training yang telah diubah pada format arff. Pada halaman tersebut terdapat atribut data training yang digunakan yaitu sebanyak 70 data training.

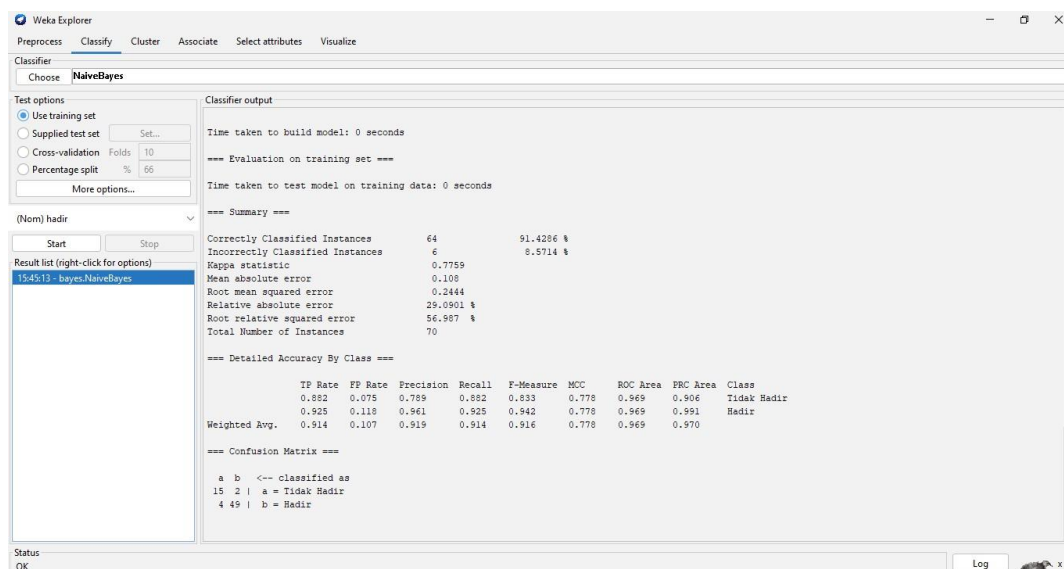
D. Halaman Menu Klasifikasi



Gambar 7. Halaman Menu Klasifikasi

Pada halaman menu klasifikasi, terdapat bermacam-macam metode klasifikasi data mining. Pengujian klasifikasi data pelanggan ini akan diuji menggunakan metode Naïve Bayes. Tahap berikutnya ialah memasukkan metode Naïve Bayes yang akan digunakan yaitu pada bagian Filter Choose pilih folder rules lalu pilih ZeroR. Berikut ini ialah rincian dari hasil klasifikasi dari 70 data training menggunakan Test Option berupa Use Training Set.

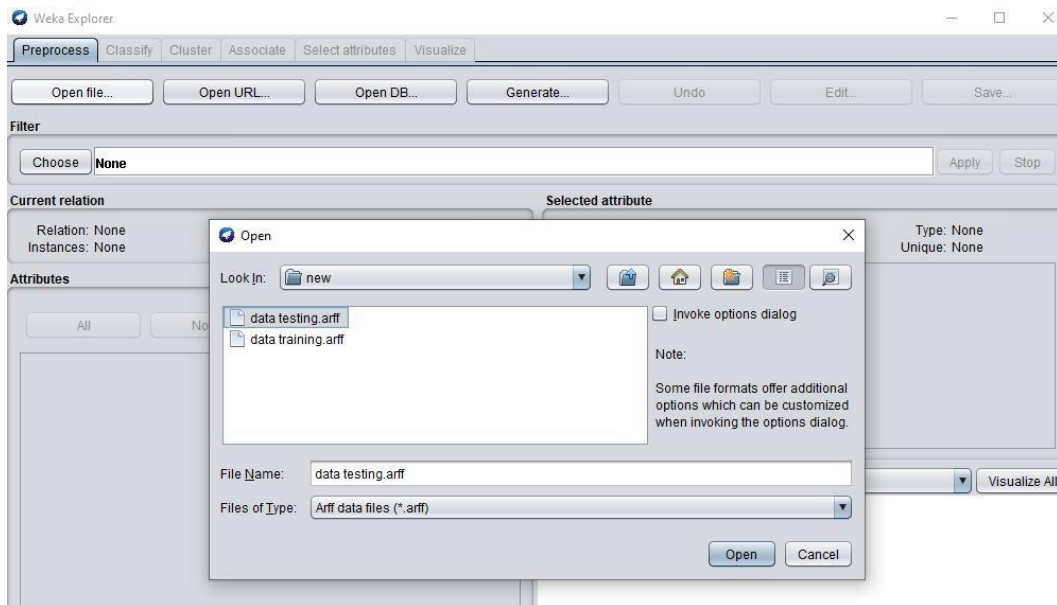
E. Hasil Klasifikasi Data Training 70 dataset



Gambar 8. Hasil Klasifikasi Data Training

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*.

F. Halaman Menu Explorer Data Testing



Gambar 9. Halaman Menu Explorer Data Testing

Pada halaman menu explorer terdapat 5 menu pilihan, diantaranya yaitu Preprocess, Classify, Cluster, Associate, Select attribute dan Visualize. Pada tahap pengujian ini akan dilakukan pada menu Preprocess dengan membuka file pada Open File, lalu masukkan data yang akan diuji yaitu data testing.

G. Halaman Atribut Data Testing

No.	Label	Count	Weight
1	dua	5	5.0
2	tiga	16	16.0
3	satu	1	1.0
4	empat	8	8.0

Gambar 10. Halaman Atribut Data Testing

Setelah memasukkan file, akan muncul atribut yang terdapat dalam data pemilu. Pada halaman atribut terdapat 10 atribut yang digunakan, diantaranya ialah 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.

No.	1: jml KK	2: Asal Penduduk	3: Kategori Masy	4: kawin	5: jenis_kelamin	6: alamat	7: rt	8: tps	9: lokasi	10: hadir kel
1	dua	SIDOARIO	PascaMilienial	Sudah	Perempuan	WONO...	rt17	TPS-13	luar	Tidak Semu...
2	dua	SIDOARIO	PascaMilienial	Pernah	Perempuan	WONO...	rt17	TPS-13	luar	Tidak Semu...
3	dua	SIDOARIO	PascaMilienial	Pernah	Laki-Laki	WONO...	rt17	TPS-13	luar	Tidak Semu...
4	tiga	SIDOARIO	Milienial	Belum	Perempuan	WONO...	rt17	TPS-13	luar	Tidak Hadir...
5	tiga	LUAR	PascaMilienial	Sudah	Laki-Laki	WONO...	rt17	TPS-13	luar	Tidak Hadir...
6	tiga	SIDOARIO	PascaMilienial	Sudah	Perempuan	WONO...	rt17	TPS-13	luar	Tidak Hadir...
7	tiga	SIDOARIO	Milienial	Belum	Laki-Laki	WONO...	rt17	TPS-13	luar	Tidak Semu...
8	tiga	LUAR	PascaMilienial	Pernah	Perempuan	WONO...	rt17	TPS-13	luar	Tidak Semu...
9	tiga	SIDOARIO	PascaMilienial	Sudah	Laki-Laki	WONO...	rt17	TPS-13	luar	Tidak Semu...
10	tiga	SIDOARIO	PascaMilienial	Sudah	Laki-Laki	WONO...	rt17	TPS-13	luar	Tidak Semu...
11	tiga	SIDOARIO	PascaMilienial	Sudah	Perempuan	WONO...	rt17	TPS-13	luar	Tidak Semu...
12	tiga	SIDOARIO	PascaMilienial	Belum	Laki-Laki	WONO...	rt17	TPS-13	luar	Tidak Semu...
13	satu	SIDOARIO	PascaMilienial	Sudah	Perempuan	KLITH	rt21	TPS-16	dalam	Hadir Semua
14	tiga	SIDOARIO	Pernula	Belum	Laki-Laki	KLITH	rt21	TPS-16	dalam	Hadir Semua
15	tiga	SIDOARIO	PascaMilienial	Sudah	Laki-Laki	KLITH	rt21	TPS-16	dalam	Hadir Semua
16	tiga	SIDOARIO	PascaMilienial	Sudah	Perempuan	KLITH	rt21	TPS-16	dalam	Hadir Semua
17	dua	SIDOARIO	PascaMilienial	Belum	Laki-Laki	WONO...	rt21	TPS-16	dalam	Tidak Semu...
18	dua	SIDOARIO	PascaMilienial	Belum	Perempuan	WONO...	rt21	TPS-16	dalam	Tidak Semu...
19	empat	SIDOARIO	Milienial	Belum	Perempuan	KLITH	rt21	TPS-16	dalam	Hadir Semua
20	empat	SIDOARIO	PascaMilienial	Sudah	Perempuan	KLITH	rt21	TPS-16	dalam	Hadir Semua
21	empat	SIDOARIO	Milienial	Belum	Perempuan	KLITH	rt21	TPS-16	dalam	Hadir Semua
22	empat	SIDOARIO	PascaMilienial	Sudah	Laki-Laki	KLITH	rt21	TPS-16	dalam	Hadir Semua
23	empat	SIDOARIO	PascaMilienial	Sudah	Perempuan	WONO...	rt21	TPS-16	dalam	Tidak Semu...
24	empat	LUAR PROV	PascaMilienial	Sudah	Laki-Laki	WONO...	rt21	TPS-16	dalam	Tidak Semu...
25	empat	SIDOARIO	PascaMilienial	Pernah	Perempuan	WONO...	rt21	TPS-16	dalam	Tidak Semu...
26	empat	SIDOARIO	PascaMilienial	Pernah	Perempuan	WONO...	rt21	TPS-16	dalam	Tidak Semu...
27	tiga	SIDOARIO	PascaMilienial	Belum	Laki-Laki	KLITH	rt21	TPS-16	dalam	Hadir Semua
28	tiga	SIDOARIO	PascaMilienial	Sudah	Perempuan	KLITH	rt21	TPS-16	dalam	Hadir Semua
29	tiga	SIDOARIO	PascaMilienial	Sudah	Laki-Laki	KLITH	rt21	TPS-16	dalam	Hadir Semua
30	tiga	SIDOARIO	PascaMilienial	Belum	Perempuan	KLITH	rt21	TPS-16	dalam	Hadir Semua

Gambar 11. Halaman Viewer Atribut Data Testing

Pada gambar 4.10 menampilkan halaman Viewer yang digunakan untuk melihat atribut data training yang telah diubah pada format arff. Pada halaman tersebut terdapat atribut data training yang digunakan yaitu sebanyak 30 data testing.

H. Hasil Klasifikasi Data Testing

Classifier: NaiveBayes

Test options: Use training set, Supplied test set, Cross-validation (Folds: 10, Percentage split: 66), More options...

Classifier output (total: 17.0, 3.0, 14.0)

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	23	76.667 %
Incorrectly Classified Instances	7	23.333 %
Kappa statistic	0.593	
Mean absolute error	0.1957	
Root mean squared error	0.3397	
Relative absolute error	46.7641 %	
Root relative squared error	76.2673 %	
Total Number of Instances	30	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.733	0.200	0.786	0.733	0.759	0.535	0.847	0.797	Tidak Semua Hadir
	0.000	0.074	0.000	0.000	0.000	-0.059	0.784	0.239	Tidak Hadir Semua
	1.000	0.111	0.857	1.000	0.923	0.873	1.000	1.000	Hadir Semua
Weighted Avg.	0.767	0.152	0.736	0.767	0.748	0.608	0.902	0.817	

=== Confusion Matrix ===

a	b	c	-- classified as
11	2	1	a = Tidak Semua Hadir
3	0	1	b = Tidak Hadir Semua
0	0	12	c = Hadir Semua

Gambar 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*.

I. Hasil Klasifikasi Data Training 5 dataset

The screenshot shows the Weka Classifier interface with the NaiveBayes classifier selected. The 'Classifier output' pane displays the following results:

```

=== Evaluation on training set ===
Time taken to test model on training data: 0 seconds

=== Summary ===
Correctly Classified Instances      5      100 %
Incorrectly Classified Instances    0        0 %
Kappa statistic                     1
Mean absolute error                 0.0051
Root mean squared error             0.0078
Relative absolute error             1.0448 %
Root relative squared error         1.5987 %
Total Number of Instances          5

=== Detailed Accuracy By Class ===
          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
1.000  0.000  1.000  1.000  1.000  1.000  1.000  1.000  Hadir
1.000  0.000  1.000  1.000  1.000  1.000  1.000  1.000  Tidak Hadir
Weighted Avg.  1.000  0.000  1.000  1.000  1.000  1.000  1.000  1.000

=== Confusion Matrix ===
 a b  <-- classified as
 3 0 | a = Hadir
 0 2 | b = Tidak Hadir

```

Gambar 13. Hasil Klasifikasi Data Training 5 Dataset

Pada tahap berikut ini pilih Start untuk melihat hasil klasifikasi dari data 5 data training. Dari hasil pengujian diperoleh tingkat kebenaran sebesar 100% pada *Correctly Classified Instances* dan tingkat kesalahan sebesar 0% pada *Incorrect Classified Instances*.

J. Hasil Klasifikasi Data Testing 5 dataset

The screenshot shows the Weka Classifier interface with the NaiveBayes classifier selected. The 'Classifier output' pane displays the following results:

```

=== Evaluation on training set ===
Time taken to test model on training data: 0 seconds

=== Summary ===
Correctly Classified Instances      5      100 %
Incorrectly Classified Instances    0        0 %
Kappa statistic                     1
Mean absolute error                 0.0051
Root mean squared error             0.0078
Relative absolute error             1.0448 %
Root relative squared error         1.5987 %
Total Number of Instances          5

=== Detailed Accuracy By Class ===
          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
1.000  0.000  1.000  1.000  1.000  1.000  1.000  1.000  Hadir
1.000  0.000  1.000  1.000  1.000  1.000  1.000  1.000  Tidak Hadir
Weighted Avg.  1.000  0.000  1.000  1.000  1.000  1.000  1.000  1.000

=== Confusion Matrix ===
 a b  <-- classified as
 3 0 | a = Hadir
 0 2 | b = Tidak Hadir

```

Gambar 14. Hasil Klasifikasi Data Testing 5 Dataset

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*.

K. 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 6. Hasil Persentase Pengujian 100 Dataset

	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%.

Dari hasil klasifikasi menggunakan metode Naïve Bayes untuk mengklasifikasi status kehadiran pada 100 dataset, diperoleh 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:

Tabel 7. Hasil Persentase Pengujian 10% dari 100 Dataset

Perbandingan	Dataset	Training	Testing	Benar	% Benar	Salah	% Salah
Manual	100	5	5	2	60%	3	40%
Weka	100	5	5	5	100%	0	0%
Weka	4.249	-	-	3.846	90,51%	403	9,48%

Pada dataset 100 menggunakan weka diperoleh prediksi kebenaran sebesar 100% dan prediksi kesalahan sebesar 0%. Pada 10% dari 100 dataset menggunakan perhitungan manual diperoleh prediksi kebenaran sebesar 60% dan prediksi kesalahan sebesar 40%.

Dari hasil klasifikasi menggunakan metode Naïve Bayes untuk mengklasifikasi status kehadiran pada 10% dari 100 dataset, diperoleh hasil tertinggi menggunakan perhitungan weka dengan perolehan persentase sebesar 100%.

V. SIMPULAN

Berdasarkan 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%. Pada penelitian selanjutnya dapat menggunakan metode lain untuk digunakan sebagai perbandingan.

UCAPAN TERIMA KASIH

Dalam penyusunan penelitian ini, penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan dan bantuannya, sehingga penelitian ini dapat diselesaikan dengan baik dan sesuai harapan.

REFERENSI

- [1] D. M. Liando, "Pemilu dan Partisipasi Politik Masyarakat (Studi Pada Pemilihan Anggota Legislatif Dan Pemilihan Presiden Dan Calon Wakil Presiden Di Kabupaten Minahasa Tahun 2014)," *J. LPPM Bid. EkoSosBudKum*, vol. 3, no. 2, pp. 14–28, 2016.
- [2] Natalia, "Pengaruh Ulasan Pelanggan Online Terhadap Keputusan Generasi Milenial Dalam Memesan Kamar Hotel," pp. 5–14, 2017.
- [3] D. T. Wahyuni, T. Sutojo, and A. Luthfiarta, "Prediksi Hasil Pemilu Legislatif DKI Jakarta Menggunakan Naïve Bayes Dengan Algoritma Genetika Sebagai Fitur Seleksi," *Udinus*, 2004.
- [4] A. J. Nathan and A. Scobell, "Model Algoritma K-nearest Neighbor untuk memprediksi kelulusan mahasiswa," *Foreign Aff.*, vol. 91, no. 5, pp. 1–9, 2012.
- [5] Verawati and P. D. Liksha, "Aplikasi Akuntansi Pengolahan Data Jasa Service Pada Pt. Budi Berlian Motor Lampung," *J. Sist. Inf. Akunt.*, vol. 1, no. 1, pp. 1–14, 2018.
- [6] R. Y. Hayuningtyas, "Penerapan Algoritma Naïve Bayes untuk Rekomendasi Pakaian Wanita," *J. Inform.*, vol. 6, no. 1, pp. 18–22, 2019, doi: 10.31294/ji.v6i1.4685.
- [7] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2,

- pp. 160–165, 2018, doi: 10.33096/ilkom.v10i2.303.160-165.
- [8] A. S. Fitriani, T. Informatika, F. Teknik, and U. M. Sidoarjo, “Penerapan Data Mining Menggunakan Metode Klasifikasi Naïve Bayes untuk Memprediksi Partisipasi Pemilihan Gubernur,” vol. 3, no. 2, pp. 98–104, 2019.
- [9] S. Hendrian, “Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan,” *Fakt. Exacta*, vol. 11, no. 3, pp. 266–274, 2018, doi: 10.30998/faktorexacta.v11i3.2777.
- [10] A. Saleh, “Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga,” *Creat. Inf. Technol. J.*, 2015.
- [11] D. Purnamasari, J. Henharta, Y. P. Sasmita, F. Ihsani, and I. W. S. Wicaksana, “Machine Learning ‘Get Easy Using WEKA,’” *Dapur Buku*, pp. 1–40, 2013.

Conflict of Interest Statement:

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.