

PENERAPAN ALGORITMA NAIVE BAYES UNTUK MEMPREDIKSI KEHADIRAN MASYARAKAT DALAM PEMILIHAN KEPALA DAERAH STUDI KASUS DESA MATABESI

Kristoforus Boik Tanii¹, Yoseph Pius Kurniawan Kelen², Krisantus Jumarto Tey Seran³
^{1,2,3}Universitas Timor

(Program Studi Teknologi Informasi Fakultas Pertanian Sains dan Kesehatan Universitas Timor)
(KM 9. Kelurahan Sasi, Kefamenanu, Timor Tengah Utara,
Nusa Tenggara Timur, telp. 0812 3956 4700)

e-mail: ¹kristoforustanii18@gmail.com, ²yosepkelen@unimor.ac.id, ³krisantusteyseran@unimor.ac.id,

Abstrak

Partisipasi masyarakat dalam pemilihan kepala daerah (Pilkada) merupakan indikator penting dalam keberhasilan demokrasi lokal. Penelitian ini bertujuan untuk memprediksi tingkat kehadiran masyarakat dalam Pilkada dengan menggunakan algoritma Naive Bayes. Data yang digunakan berasal dari Desa Matabesi, Kecamatan Biboki Moenleu, Kabupaten Timor Tengah Utara, yang mencakup 316 data pemilih tetap dengan atribut seperti usia, jenis kelamin, lokasi tempat tinggal, aksesibilitas ke TPS, pendapatan bulanan, pendidikan, pekerjaan, dan kepedulian politik. Proses klasifikasi dilakukan dalam beberapa tahap: pengumpulan data, preprocessing, pemodelan, pengujian, dan evaluasi menggunakan RapidMiner. Model dibangun dengan membagi data menjadi 70% data pelatihan dan 30% data pengujian. Hasil evaluasi menunjukkan bahwa model memiliki tingkat akurasi, presisi, dan recall sebesar 100%, yang menandakan bahwa algoritma Naive Bayes mampu memprediksi kehadiran dengan sangat baik. Faktor-faktor seperti kedekatan dengan TPS dan kepedulian politik tinggi menjadi penentu utama kehadiran. Temuan ini memberikan implikasi positif dalam membantu penyelenggara pemilu untuk menyusun strategi peningkatan partisipasi pemilih berbasis data.

Kata kunci: Naive Bayes, Pilkada, Data Mining, Prediksi Kehadiran, RapidMiner.

Abstract

Public participation in regional head elections (Pilkada) is a key indicator of successful local democracy. This study aims to predict voter attendance using the Naive Bayes algorithm. The dataset was collected from Matabesi Village, Biboki Moenleu District, North Central Timor Regency, consisting of 316 registered voters with attributes such as age, gender, residence location, accessibility to polling stations, monthly income, education, occupation, and political awareness. The classification process followed several stages: data collection, preprocessing, modeling, testing, and evaluation using RapidMiner. The dataset was split into 70% training and 30% testing. Evaluation results showed that the model achieved 100% accuracy, precision, and recall, indicating the high effectiveness of the Naive Bayes algorithm in predicting attendance. Key influencing factors included proximity to polling stations and high political awareness. These findings offer valuable insights for election organizers to develop data-driven strategies to increase voter participation.

Keywords: Naive Bayes, Regional Election, Data Mining, Attendance Prediction, RapidMiner.

1. PENDAHULUAN

Perkembangan ilmu pengetahuan dan teknologi telah memberikan dampak signifikan terhadap hampir seluruh aspek kehidupan manusia, menjadikannya lebih efisien dan praktis [10]. Salah satu wujud nyata dari kemajuan tersebut adalah komputer, yang kini telah menjadi alat bantu utama dalam berbagai aktivitas sehari-hari. Peran komputer semakin strategis dalam era Revolusi Industri 4.0 dan masyarakat 5.0, termasuk dalam bidang politik. Dalam konteks ini, teknologi komputer

dapat dimanfaatkan untuk mendukung proses demokrasi, seperti memprediksi partisipasi masyarakat dalam pemilihan umum serta membantu perhitungan dan rekapitulasi suara.

Sebagai negara demokratis, Indonesia mengandalkan pemilihan umum, termasuk Pemilihan Kepala Daerah (Pilkada), sebagai sarana untuk menentukan pemimpin. Tingkat kehadiran masyarakat dalam Pilkada menjadi faktor penting dalam menentukan legitimasi dan kualitas pemerintahan yang terbentuk. Rendahnya partisipasi pemilih sering kali disebabkan oleh berbagai kendala, seperti minimnya akses informasi, ketidakpuasan terhadap proses pemilu, hambatan logistik, kesenjangan sosial, hingga kondisi ekonomi yang tidak mendukung. Hal ini berdampak pada menurunnya keikutsertaan masyarakat dalam proses demokrasi, khususnya dalam pemilihan kepala daerah.

Pentingnya partisipasi aktif masyarakat dalam pemilu merupakan salah satu indikator keberhasilan demokrasi di tingkat lokal. Dalam upaya mengatasi tantangan tersebut, pemanfaatan teknologi informasi, khususnya metode dalam ilmu komputer, dapat menjadi solusi yang efektif. Salah satu pendekatan yang dapat digunakan adalah data mining, yaitu proses penemuan pola-pola tersembunyi dari data dalam skala besar menggunakan teknik statistik, kecerdasan buatan, dan pembelajaran mesin [1]. Data mining memiliki dua pendekatan utama: klasifikasi dan clustering. Klasifikasi berfokus pada pembentukan model untuk memetakan data ke dalam kelas-kelas tertentu [8] sedangkan clustering mengelompokkan data berdasarkan tingkat kemiripan antar objek [5].

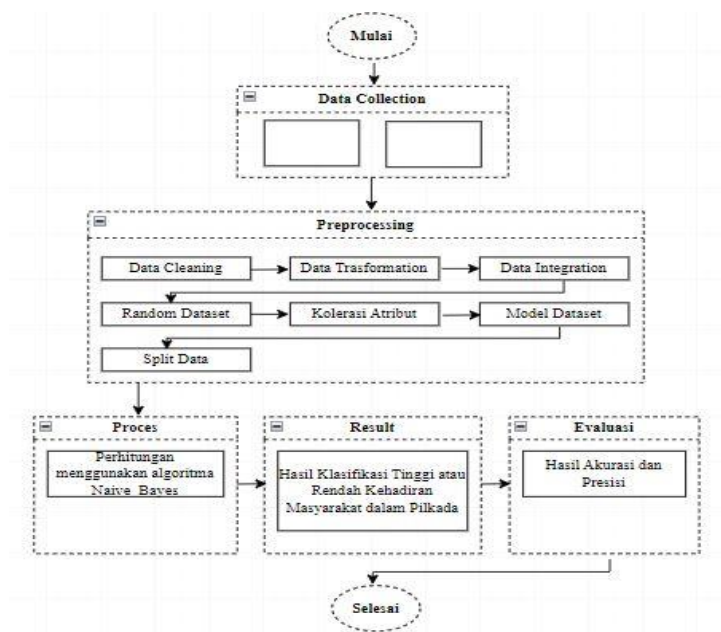
Metode klasifikasi seperti *Naïve Bayes* merupakan salah satu teknik yang efektif untuk memprediksi kemungkinan partisipasi masyarakat dalam pemilu. Algoritma ini menggunakan pendekatan probabilistik yang dikembangkan oleh Thomas Bayes, untuk memprediksi kejadian berdasarkan data historis.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk menganalisis dan memprediksi tingkat partisipasi masyarakat dalam Pilkada menggunakan algoritma *Naïve Bayes*. Penelitian dilakukan di Desa Matabesi, Kecamatan Biboki Moenleu, Kabupaten Timor Tengah Utara, Provinsi Nusa Tenggara Timur.

2. METODE PENELITIAN

2.1. Tahapan Penelitian

Agar penelitian ini dapat berjalan secara sistematis dan mencapai tujuan yang diharapkan, diperlukan penerapan metode yang terstruktur dan terencana. Oleh karena itu, rangkaian tahapan yang dilaksanakan dalam penelitian ini adalah sebagai berikut:



Gambar 1. Tahapan penelitian

1. *Data Collection*

Proses klasifikasi ini dimulai dari tahap pengumpulan data (*Data Collection*). Pada tahap ini, peneliti mengumpulkan data dari sumber yang valid, yaitu daftar pemilih tetap tahun 2024 dan data kehadiran pemilu tahun 2023 yang diperoleh dari Komisi Pemilihan Umum (KPU) Kabupaten Timor Tengah Utara. Data yang dikumpulkan mencakup atribut-atribut penting seperti usia, jenis kelamin, lokasi tempat tinggal, jarak ke TPS, pendapatan bulanan, pendidikan terakhir, jenis pekerjaan, dan tingkat kepedulian politik. Tujuannya adalah membangun dataset yang representatif terhadap kondisi sosial masyarakat untuk dianalisis lebih lanjut.

2. *Preprocessing*

Selanjutnya, data yang sudah dikumpulkan tidak langsung digunakan, melainkan melalui tahap pra-pemrosesan (*Preprocessing*) untuk memastikan bahwa data tersebut bersih, konsisten, dan siap dianalisis. Tahapan *preprocessing* terdiri dari data *cleaning*, data *integration*, data *transformation*, *random dataset*, korelasi atribut, model dataset dan *split* data.

a. *Data Cleaning*

Di tahap ini, data diperiksa dari duplikasi, nilai kosong, dan data yang tidak valid. Kesalahan-kesalahan ini diperbaiki, misalnya mengisi data kosong dengan nilai modus atau menghapus baris yang rusak..

b. *Data Transformation*

Pada proses ini data yang telah diperiksa akan diubah dari data kategori menjadi data numerik agar bisa diproses oleh algoritma. Misalnya, nilai "Tinggi" dikonversi menjadi 3, "Sedang" menjadi 2, dan "Rendah" menjadi 1.

c. *Data Integration*

Selanjutnya, data telah ubah Menggabungkan data dari dua sumber berbeda (daftar pemilih tetap 2024 dan data rekapitulasi pilkada tahun 2019) menjadi satu dataset utuh dan seragam.

d. *Random Dataset*

Setelah data terintegrasi, data kemudian diacak (*random shuffle*). Proses ini penting untuk menghindari bias urutan data yang bisa memengaruhi hasil pelatihan model. Dengan randomisasi, distribusi data akan lebih representatif sehingga model dapat belajar dengan lebih adil terhadap berbagai pola yang ada.

e. *Model Dataset*

Dataset yang telah diacak dan siap pakai akan digunakan sebagai model dataset. Dalam penelitian ini, model dataset mencakup seluruh atribut yang tersedia seperti umur, pekerjaan, pendidikan, status kehadiran sebelumnya, dan sebagainya. Semua atribut ini akan dipertimbangkan oleh algoritma klasifikasi dalam menentukan prediksi kehadiran.

f. *Split Data*

Tahap terakhir dalam proses *preprocessing* adalah melakukan pembagian dataset menjadi dua bagian utama, yaitu data *training* dan data *testing*. Data *training* mencakup 70% dari keseluruhan dataset dan berfungsi untuk melatih algoritma klasifikasi, seperti algoritma Naïve Bayes, agar dapat mengenali pola-pola yang terdapat dalam data. Sementara itu, data *testing* mencakup 30% dari dataset dan digunakan untuk mengukur sejauh mana model yang telah dilatih mampu melakukan klasifikasi dengan akurat terhadap data baru yang belum dikenalnya. Pembagian dengan rasio 70:30 ini bertujuan untuk memastikan bahwa model memiliki cukup data untuk belajar serta cukup data untuk dilakukan pengujian performa secara objektif.

3. *Process*

Masuk ke tahap process, yaitu inti dari penelitian ini. Di sini dilakukan pengujian menggunakan algoritma *Naïve Bayes*. Dataset training dimasukkan ke dalam *RapidMiner*, kemudian model dibangun berdasarkan prinsip probabilistik. Setiap atribut pada data diuji untuk menghitung peluang kehadiran (Hadir atau Tidak Hadir) berdasarkan data historis. Pengujian dilakukan terhadap data testing, dan hasil prediksi dibandingkan dengan nilai aktualnya.

4. *Result*

Setelah proses klasifikasi selesai, dilanjutkan ke tahap hasil (*Result*). Pada tahap ini, ditampilkan prediksi dari data testing, apakah masing-masing individu diprediksi hadir atau tidak hadir dalam Pilkada. Hasil ini akan memberikan gambaran umum tingkat kehadiran masyarakat berdasarkan

kombinasi atribut-atribut yang dimiliki.

5. Evaluasi

Tahap terakhir adalah evaluasi (*Evaluation*). Di tahap ini, model dievaluasi menggunakan *confusion matrix* yang menghasilkan metrik seperti akurasi, precision, dan *recall*. Evaluasi ini bertujuan untuk mengetahui seberapa akurat model dalam memprediksi kehadiran masyarakat. Dalam penelitian ini, model berhasil mencapai akurasi 100% yang berarti tidak ada kesalahan klasifikasi pada data uji.

2.2. Data Mining

Data mining merupakan proses penambangan data yang memanfaatkan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk menggali informasi dan pengetahuan yang tersembunyi dari kumpulan data berskala besar. Teknik yang umum digunakan meliputi pengenalan pola, pengelompokan, asosiasi, prediksi, dan klasifikasi. Tujuan utamanya adalah menyusun informasi menjadi bentuk yang dapat dipahami dan dimanfaatkan dalam pengambilan keputusan.

2.3. Naive Bayes

Naïve Bayes adalah metode klasifikasi berbasis probabilistik yang menggunakan Teorema Bayes untuk menghitung kemungkinan suatu data termasuk dalam kelas tertentu. Metode ini mengasumsikan bahwa setiap atribut bersifat independen terhadap atribut lainnya. Keunggulannya terletak pada kesederhanaan dan efisiensi, serta kemampuannya memberikan hasil akurat meskipun digunakan pada data yang kompleks. Naïve Bayes hanya memerlukan data latih dan data uji, dan sering digunakan karena performanya yang baik dalam praktik. Persamaan dasar dari Teorema Bayes yang digunakan dalam klasifikasi Naïve Bayes ditunjukkan pada Persamaan (1):

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} = \frac{P(X|H)P(H)}{\sum_{i=1}^n P(X|H_i)P(H_i)} \quad (1)$$

Dimana :

X = Data kelas yang belum diketahui

H = Hipotesis data X adalah kelas spesifik

P(H|X) = Kemungkinan Hipotesa H berdasarkan keadaan X (posterior probability)

P(H) = Kemungkinan Hipotesa H (prior probability)

P(X|H) = Kemungkinan hipotesa X berdasarkan keadaan hipotesa H

P(X) = Kemungkinan hipotesa dari X

2.4. RapidMiner

RapidMiner merupakan perangkat lunak berbasis Java yang digunakan untuk proses analisis data dan data mining, baik dalam bidang pendidikan, penelitian, maupun industri. Aplikasi ini menyediakan antarmuka pengguna (UI) yang memungkinkan perancangan proses analisis melalui *drag-and-drop*, dan menghasilkan file XML untuk menjalankan alur kerja secara otomatis. RapidMiner mendukung berbagai format file seperti .xls, .csv, dan lainnya, serta menawarkan berbagai algoritma untuk klasifikasi, pengelompokan, dan regresi. Selain itu, platform ini dapat diintegrasikan dengan berbagai bahasa pemrograman secara fleksibel.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Pengumpulan Data

Penelitian ini melakukan wawancara dan observasi pada Desa Matabesi yang dimana masyarakat Matabesi dijadikan sebagai sampel dalam rangka memprediksi kehadiran mereka dalam pemilihan

kepala daerah. Data yang diambil sebanyak 316 yang kemudian data ini akan dibagi menjadi data training yakni 221 data dan data testing yakni 95 data.

Tabel 1. Data Training

No	Nama	Usia	Jenis Kelamin	Lokasi Tempat Tinggal	Aksesibilitas ke TPS	Pendapatan Bulanan	Pendidikan	Pekerjaan	Kepedulian Politik	Kehadiran
1	P1	Tua	P	Desa	Dekat	Rendah	Tidak Sekolah	Wiraswasta	Tinggi	Hadir
2	P2	Tua	P	Desa	Jauh	Rendah	Tidak Sekolah	Wiraswasta	Rendah	Tidak Hadir
3	P3	Dewasa	L	Desa	Dekat	Menengah	SMP	Wiraswasta	Tinggi	Hadir
4	P4	Dewasa	P	Desa	Dekat	Rendah	SMP	Wiraswasta	Tinggi	Hadir
5	P5	Tua	L	Desa	Dekat	Rendah	SD	Wiraswasta	Tinggi	Hadir
6	P6	Remaja	P	Desa	Dekat	Rendah	SMA	Wiraswasta	Tinggi	Hadir
7	P7	Dewasa	L	Desa	Jauh	Rendah	Tidak Sekolah	Wiraswasta	Rendah	Tidak Hadir
8	P8	Dewasa	L	Desa	Jauh	Rendah	Tidak Sekolah	Wiraswasta	Rendah	Tidak Hadir
9	P9	Dewasa	L	Desa	Dekat	Menengah	SMA	Wiraswasta	Tinggi	Hadir
10	P10	Remaja	L	Desa	Dekat	Menengah	SMA	Wiraswasta	Tinggi	Hadir
11	P11	Dewasa	L	Desa	Dekat	Menengah	SMA	Wiraswasta	Tinggi	Hadir
12	P12	Dewasa	L	Desa	Dekat	Rendah	Tidak Sekolah	Wiraswasta	Tinggi	Hadir
13	P13	Dewasa	L	Desa	Dekat	Menengah	Tidak Sekolah	Wiraswasta	Tinggi	Hadir
14	P14	Tua	P	Desa	Dekat	Rendah	Tidak Sekolah	Wiraswasta	Rendah	Hadir
...
221	P221	Dewasa	P	Desa	Dekat	Rendah	SMA	Wiraswasta	Tinggi	Hadir

Sebanyak 221 data training digunakan untuk melatih model prediksi dengan algoritma Naïve Bayes. Data ini berisi informasi historis terkait partisipasi masyarakat dalam pemilu sebelumnya maupun hasil survei, sehingga memungkinkan model mempelajari pola dari atribut-atribut yang relevan. Atribut seperti akses ke TPS, pekerjaan, dan kepedulian politik berperan penting dalam membentuk hasil prediksi. Sementara itu, data testing yang digunakan untuk menguji model ditampilkan pada Tabel 4.2.

Tabel 2. Data Testing

NO	Nama	Usia	Jenis Kelamin	Lokasi Tempat Tinggal	Aksesibilitas ke TPS	Pendapatan Bulanan	Pendidikan	Pekerjaan	Kepedulian Politik	Kehadiran
1	P1	Remaja	P	Kota	Jauh	Rendah	SMA	Pengangguran	Rendah	?
2	P2	Dewasa	P	Kota	Jauh	Rendah	SMA	Wiraswasta	Rendah	?
3	P3	Remaja	L	Desa	Dekat	Rendah	Tidak Sekolah	Pengangguran	Tinggi	?
4	P4	Remaja	L	Desa	Dekat	Rendah	Tidak Sekolah	Wiraswasta	Rendah	?
5	P5	Remaja	L	Kota	Jauh	Menengah	SMA	Wiraswasta	Rendah	?
6	P6	Dewasa	L	Kota	Jauh	Menengah	SMP	Wiraswasta	Rendah	?
7	P7	Dewasa	L	Desa	Jauh	Menengah	SMA	Wiraswasta	Rendah	?
8	P8	Tua	L	Desa	Jauh	Rendah	Tidak Sekolah	Wiraswasta	Tinggi	?
...
95	P95	Remaja	L	Desa	Dekat	Rendah	SMA	Wiraswasta	Tinggi	?

Sebanyak 95 data testing digunakan untuk menguji dan memvalidasi model yang telah dilatih. Karena data ini belum pernah digunakan dalam proses pelatihan, pengujian ini penting untuk menilai kemampuan model dalam memprediksi data baru serta menghindari overfitting, yaitu kondisi di mana model terlalu bergantung pada data training dan gagal beradaptasi dengan data yang berbeda.

3.2. Hasil Penyajian Data

Penelitian ini mengklasifikasikan data berdasarkan variabel yang mencerminkan karakteristik masyarakat Desa Matabesi, seperti usia, jenis kelamin, lokasi tempat tinggal, jarak ke TPS, pendapatan, pendidikan, pekerjaan, dan kepedulian politik menggunakan metode Naïve Bayes. Analisis frekuensi menunjukkan distribusi kehadiran dalam Pilkada berdasarkan tiap kategori variabel. Misalnya, usia terbagi menjadi remaja, dewasa, dan tua; lokasi ke dalam desa dan kota; serta pendapatan dalam kategori rendah, menengah, dan tinggi.

Selanjutnya, dihitung nilai *likelihood* untuk mengetahui probabilitas kehadiran berdasarkan karakteristik tertentu. Beberapa variabel seperti usia, lokasi, dan kepedulian politik menunjukkan pengaruh signifikan terhadap kemungkinan hadir atau tidak hadir. Nilai-nilai ini menjadi dasar dalam klasifikasi menggunakan algoritma Naïve Bayes, karena menentukan hasil prediksi pada data testing. Dengan demikian, analisis frekuensi dan probabilitas tidak hanya memberikan gambaran deskriptif, tetapi juga menjadi landasan dalam membangun model prediksi yang akurat.

Tabel 3. Frekuensi setiap Variabel

Tabel Frekuensi		Kehadiran		Total	Kehadiran		Total
		Hadir	Tidak Hadir		Hadir	Tidak Hadir	
Usia	Remaja	24	5	29			
	Dewasa	87	3	90			
	Tua	101	1	102			
	Total	212	9	221			
Jenis Kelamin	Laki – laki	106	4	110			
	Perempuan	106	5	111			
	Total	212	9	221			
Lokasi	Desa	210	5	215			
	Kota	2	2	6			
	Total	212	9	221			
Jarak TPS	Dekat	208	1	209			
	Jauh	4	8	12			
	Total	212	9	221			
Pendapatan	Rendah	177	7	184			
	Menengah	30	2	32			
	Tinggi	5	0	5			
	total	212	9	212			
Pendidikan	Tidak sekolah	96	3	99			
	SD	38	2	40			
	SMP	32	1	33			
	SMA	32	3	35			
	Sarjana	14	0	14			
Pekerjaan	Total	212	9	221			
	Pengangguran	8	3	11			
	Wiraswasta	197	6	203			
	PNS	7	0	7			
Kepedulian politik	Total	212	9	221			
	Rendah	2	9	11			
	Tinggi	210	0	210			
	total	212	9	221			

3.3. Implementasi RapidMiner

RapidMiner digunakan dalam penelitian ini sebagai alat analisis untuk membangun dan menguji model prediksi kehadiran dalam Pilkada. Dengan antarmuka visual yang intuitif, RapidMiner mempermudah proses pemodelan mulai dari pemuatan data, pembersihan, hingga penerapan algoritma. Dataset dibagi menjadi data training dan testing, lalu dianalisis menggunakan algoritma Naive Bayes untuk klasifikasi.

1. Import Data

Mengimpor data adalah langkah pertama yang penting dalam analisis data. *Rapidminer* menyediakan berbagai cara untuk mengimpor data dari berbagai sumber seperti file lokal (csv, excel), database, atau sumber lain seperti api. Ada dua jenis data yang di *import* yaitu data *training*

dan data *testing*. Data *training* adalah dataset pelatihan yang berisi fitur dan label yang akan digunakan untuk melatih model. Data ini penting karena memberikan informasi yang diperlukan oleh model untuk belajar mengenali pola dan hubungan antara variabel. Data *training* memiliki total data sebanyak 221 data.

Row No.	Kehadiran	No	Nama	Usia	Jenis Kelamin	Lokasi Tem...	Aksesibilita...	Pendapatan ...	Pendidikan	Pekerjaan	Kepedulian ...
1	Hadir	1	P1	Tua	P	Desa	Dekat	Rendah	Tidak Sekolah	Wiraswasta	Tinggi
2	Tidak Hadir	2	P2	Tua	P	Desa	Jauh	Rendah	Tidak Sekolah	Wiraswasta	Rendah
3	Hadir	3	P3	Dewasa	L	Desa	Dekat	Menengah	SMP	Wiraswasta	Tinggi
4	Hadir	4	P4	Dewasa	P	Desa	Dekat	Rendah	SMP	Wiraswasta	Tinggi
5	Hadir	5	P5	Tua	L	Desa	Dekat	Rendah	SD	Wiraswasta	Tinggi
6	Hadir	6	P6	Remaja	P	Desa	Dekat	Rendah	SMA	Wiraswasta	Tinggi
7	Tidak Hadir	7	P7	Dewasa	L	Desa	Jauh	Rendah	Tidak Sekolah	Wiraswasta	Rendah
8	Tidak Hadir	8	P8	Dewasa	L	Desa	Jauh	Rendah	Tidak Sekolah	Wiraswasta	Rendah
9	Hadir	9	P9	Dewasa	L	Desa	dekat	Menengah	SMA	Wiraswasta	Tinggi
10	Hadir	10	P10	Remaja	L	Desa	dekat	Menengah	SMA	Wiraswasta	Tinggi
11	Hadir	11	P11	Dewasa	L	Desa	Dekat	Menengah	SMA	Wiraswasta	Tinggi
12	Hadir	12	P12	Dewasa	L	Desa	Dekat	Rendah	Tidak Sekolah	Wiraswasta	Tinggi
13	Hadir	13	P13	Dewasa	L	Desa	Dekat	Menengah	Tidak Sekolah	Wiraswasta	Tinggi
14	Hadir	14	P14	Tua	P	Desa	Dekat	Rendah	Tidak Sekolah	Wiraswasta	Rendah
Row No.	Kehadiran	No	Nama	Usia	Jenis Kelamin	Lokasi Tem...	Aksesibilita...	Pendapatan ...	Pendidikan	Pekerjaan	Kepedulian ...
208	Hadir	208	P208	Dewasa	L	Desa	Dekat	Rendah	SMA	Wiraswasta	Tinggi
209	Hadir	209	P209	Tua	L	Desa	Dekat	Rendah	Tidak Sekolah	Wiraswasta	Tinggi
210	Hadir	210	P210	Tua	L	Desa	Dekat	Rendah	SMP	Wiraswasta	Tinggi
211	Tidak Hadir	211	P211	Remaja	L	Kota	Jauh	Menengah	SMA	Wiraswasta	Rendah
212	Hadir	212	P212	Dewasa	L	Desa	Dekat	Rendah	SD	Wiraswasta	Tinggi
213	Hadir	213	P213	Dewasa	P	Desa	Dekat	Rendah	SD	Wiraswasta	Tinggi
214	Hadir	214	P214	Tua	P	Desa	Dekat	Menengah	SMP	Wiraswasta	Tinggi
215	Tidak Hadir	215	P215	Remaja	P	Kota	Jauh	Rendah	SMA	Pengangguran	Rendah
216	Hadir	216	P216	Tua	L	Desa	Dekat	Rendah	Tidak Sekolah	Wiraswasta	Tinggi
217	Hadir	217	P217	Tua	L	Desa	Dekat	Rendah	SD	Wiraswasta	Tinggi
218	Hadir	218	P218	Tua	L	Desa	Dekat	Rendah	Tidak Sekolah	Wiraswasta	Tinggi
219	Hadir	219	P219	Remaja	P	Desa	Dekat	Rendah	SMA	Wiraswasta	Tinggi
220	Hadir	220	P220	Dewasa	L	Desa	Dekat	Rendah	SD	Wiraswasta	Tinggi
221	Hadir	221	P221	Dewasa	P	Desa	Dekat	Rendah	SMA	Wiraswasta	Tinggi

Gambar 2. Data Training

Data *testing* adalah dataset yang akan digunakan untuk menguji kinerja model setelah dilatih. Penting untuk memastikan bahwa data pengujian terpisah dari data pelatihan agar model dapat dievaluasi secara objektif tanpa risiko *overfitting*, yaitu ketika model terlalu spesifik terhadap data pelatihan dan tidak dapat menggeneralisasi pada data baru. Data *testing* memiliki total data sebanyak 95 data.

No	Nama	Usia	Jenis Kelamin	Lokasi Tem...	Aksesibilita...	Pendapatan ...	Pendidikan	Pekerjaan	Kepedulian ...
1	P1	Remaja	P	Kota	Jauh	Rendah	SMA	Pengangguran	Rendah
2	P2	Dewasa	P	Kota	Jauh	Rendah	SMA	Wiraswasta	Rendah
3	P3	Remaja	L	Desa	Dekat	Rendah	Tidak Sekolah	Pengangguran	Tinggi
4	P4	Remaja	L	Desa	Dekat	Rendah	Tidak Sekolah	Wiraswasta	Rendah
5	P5	Remaja	L	Kota	Jauh	Menengah	SMA	Wiraswasta	Rendah
6	P6	Dewasa	L	Kota	Jauh	Menengah	SMP	Wiraswasta	Rendah
7	P7	Dewasa	L	Desa	Jauh	Menengah	SMA	Wiraswasta	Rendah
8	P8	Tua	L	Desa	Jauh	Rendah	Tidak Sekolah	Wiraswasta	Tinggi
9	P9	Remaja	L	Kota	Jauh	Menengah	SMA	Wiraswasta	Tinggi
10	P10	Remaja	L	Desa	Dekat	Rendah	SMA	Pengangguran	Tinggi
11	P11	Remaja	P	Kota	Jauh	Menengah	SMA	Wiraswasta	Rendah
12	P12	Remaja	L	Desa	Dekat	Rendah	Tidak Sekolah	Wiraswasta	Rendah
13	P13	Remaja	P	Desa	Dekat	Rendah	SMP	Wiraswasta	Tinggi
14	P14	Remaja	P	Desa	Dekat	Rendah	SMP	Pengangguran	Tinggi
15	P15	Remaja	P	Desa	Dekat	Rendah	Tidak Sekolah	Pengangguran	Tinggi
No	Nama	Usia	Jenis Kelamin	Lokasi Tem...	Aksesibilita...	Pendapatan ...	Pendidikan	Pekerjaan	Kepedulian ...
81	P81	Dewasa	L	Kota	Jauh	Menengah	SMP	Wiraswasta	Tinggi
82	P82	Tua	L	Desa	Dekat	Rendah	Tidak Sekolah	Wiraswasta	Tinggi
83	P83	Tua	L	Desa	Dekat	Rendah	Tidak Sekolah	Wiraswasta	Tinggi
84	P84	Remaja	L	Desa	Dekat	Rendah	SMA	Pengangguran	Tinggi
85	P85	Remaja	P	Desa	Dekat	Rendah	SMA	Pengangguran	Tinggi
86	P86	Remaja	L	Kota	Jauh	Menengah	SMA	Wiraswasta	Rendah
87	P87	Dewasa	L	Desa	Dekat	Rendah	Tidak Sekolah	Wiraswasta	Tinggi
88	P88	Tua	L	Desa	Dekat	Rendah	Tidak Sekolah	Wiraswasta	Tinggi
89	P89	Dewasa	P	Desa	Dekat	Menengah	Tidak Sekolah	Wiraswasta	Tinggi
90	P90	Remaja	P	Kota	Jauh	Menengah	SMA	Wiraswasta	Tinggi
91	P91	Remaja	L	Desa	Dekat	Rendah	SMP	Wiraswasta	Tinggi
92	P92	Dewasa	P	Kota	Jauh	Menengah	SMA	Wiraswasta	Rendah
93	P93	Remaja	L	Desa	Dekat	Rendah	SMA	Pengangguran	Tinggi
94	P94	Tua	L	Desa	Dekat	Rendah	Tidak Sekolah	Wiraswasta	Tinggi
95	P95	Remaja	L	Desa	Dekat	Rendah	SMA	Wiraswasta	Tinggi

Gambar 3. Data Testing

2. Menganalisis Hasil

a. Model Sub Proses Pada *Cross Validation RapidMiner*

Sebelum menyajikan gambar model sub proses dalam *Cross Validation* pada RapidMiner, perlu dipahami bahwa *Cross Validation* merupakan teknik evaluasi model yang membagi data menjadi beberapa bagian (folds) untuk proses pelatihan dan pengujian secara berulang. Teknik ini digunakan untuk meningkatkan reliabilitas hasil evaluasi serta mengurangi bias akibat pembagian data yang tidak merata. Dalam model *sub process* pada RapidMiner, terdapat beberapa komponen penting, dimulai dari pengambilan data, pelatihan model, penerapan model, hingga evaluasi kinerja. Tahap pertama adalah pengambilan data, di mana data pelatihan diambil menggunakan operator *Retrieve* dari jalur //Local Repository/data/data training. Data ini berisi fitur dan label yang digunakan untuk melatih model agar dapat mengenali pola dari atribut-atribut yang tersedia. Selanjutnya, data pengujian juga diambil dari jalur repositori sebagai data yang belum dikenali oleh model. Data pengujian digunakan untuk menilai kinerja model secara objektif, guna menghindari overfitting, yaitu kondisi di mana model hanya bekerja baik pada data pelatihan namun gagal menggeneralisasi pada data baru.

b. Pelatihan model dengan naïve bayes

Setelah data pelatihan dimuat, proses selanjutnya adalah membangun model klasifikasi menggunakan algoritma Naive Bayes. Algoritma ini bekerja dengan prinsip probabilistik dan mengasumsikan bahwa setiap fitur bersifat independen terhadap fitur lainnya. Melalui data pelatihan, model belajar mengenali pola hubungan antara atribut dengan label kehadiran. Untuk meningkatkan keakuratan prediksi, digunakan juga *Laplace Correction*, yaitu parameter yang diaktifkan guna mengatasi masalah *zero probability*—yakni kondisi di mana suatu kategori tidak muncul dalam data pelatihan. Koreksi ini memungkinkan model untuk tetap memberikan estimasi probabilitas secara realistis, meskipun data kategori tersebut jarang atau tidak muncul selama proses pelatihan.

3. Penerapan Model

Setelah model Naive Bayes dilatih, langkah berikutnya adalah menerapkan model tersebut ke data pengujian menggunakan operator *Apply Model*. Proses ini memungkinkan model untuk menghasilkan prediksi berdasarkan pola yang telah dipelajari dari data pelatihan. Dengan menerapkan model ke dataset pengujian, kita dapat melihat seberapa baik model dapat melakukan prediksi pada data yang belum pernah dilihat sebelumnya.

4. Evluasi kinerja Moder

Langkah akhir dalam proses ini adalah evaluasi kinerja model menggunakan operator *Performance* di RapidMiner. Model yang telah diuji pada data pengujian kemudian dinilai berdasarkan ketepatan hasil prediksinya terhadap label sebenarnya. Evaluasi ini menggunakan beberapa metrik utama, yaitu akurasi (persentase prediksi yang benar), presisi (proporsi prediksi positif yang tepat), dan recall atau sensitivitas (proporsi data positif yang berhasil dikenali dengan benar). Metrik-metrik ini memberikan gambaran menyeluruh mengenai kemampuan model dalam melakukan klasifikasi secara akurat dan konsisten.

5. Hasil Perfomance Vector

Dalam analisis kinerja model, penting untuk mengevaluasi bagaimana algoritma bekerja dalam mengklasifikasikan data. Gambar 4 menunjukkan hasil *Performance Vector*, yang menggambarkan metrik kinerja seperti akurasi, *precision*, *recall*. Melalui gambar ini, kita dapat memahami sejauh mana model berhasil dalam mengidentifikasi dan membedakan kelas yang berbeda dalam dataset yang digunakan.

	true Tidak Hadir	true Hadir	class precision
pred Tidak Hadir	24	0	100.00%
pred Hadir	0	71	100.00%
class recall	100.00%	100.00%	

Gambar 4. Hasil Performance Vector

Gambar di atas menunjukkan hasil Performance Vector dari model yang telah diuji, memberikan gambaran tentang kinerja model dalam mengklasifikasikan data berdasarkan metrik evaluasi yang relevan.

- a. Precision: Mengukur seberapa banyak prediksi positif yang benar dibandingkan dengan total prediksi positif.

- 1) Untuk kelas "Tidak Hadir":

$$\text{Precision (Tidak Hadir)} = \frac{TP}{TP+FP} = \frac{24}{24+0} = 100\%$$

- 2) Untuk kelas "Hadir":

$$\text{Precision (Hadir)} = \frac{TP}{TP+FP} = \frac{71}{71+0} = 100\%$$

Model ini juga memiliki precision 100.00% untuk kedua kelas (Hadir dan Tidak Hadir). Ini berarti setiap prediksi positif yang dibuat oleh model benar-benar positif, tanpa ada kesalahan dalam klasifikasi.

- b. Recall: Mengukur seberapa banyak prediksi positif yang benar dibandingkan dengan total data positif yang sebenarnya.

- 1) Untuk kelas "Tidak Hadir":

$$\text{Recall (Tidak Hadir)} = \frac{TP}{TP+FN} = \frac{24}{24+0} = 100\%$$

- 2) Untuk kelas "Hadir":

$$\text{Recall (Hadir)} = \frac{TP}{TP+FN} = \frac{71}{71+0} = 100\%$$

Dengan recall 100.00%, model mampu mengenali semua instance yang sebenarnya positif tanpa ada yang terlewat. Hal ini menunjukkan bahwa model tidak melewatkan satupun data yang seharusnya diprediksi sebagai "Hadir" atau "Tidak Hadir".

- c. Akurasi : Mengukur seberapa banyak prediksi yang benar dibandingkan dengan total prediksi.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{24+71}{24+71+0+0} = \frac{95}{95} = 100\%$$

Model ini memiliki akurasi sebesar 100.00%, yang menunjukkan bahwa semua prediksi yang dihasilkan oleh model benar. Setiap instance yang diprediksi sebagai "Tidak Hadir" atau "Hadir" sesuai dengan label sebenarnya. Akurasi yang tinggi ini menunjukkan bahwa model bekerja dengan sangat baik pada dataset ini.

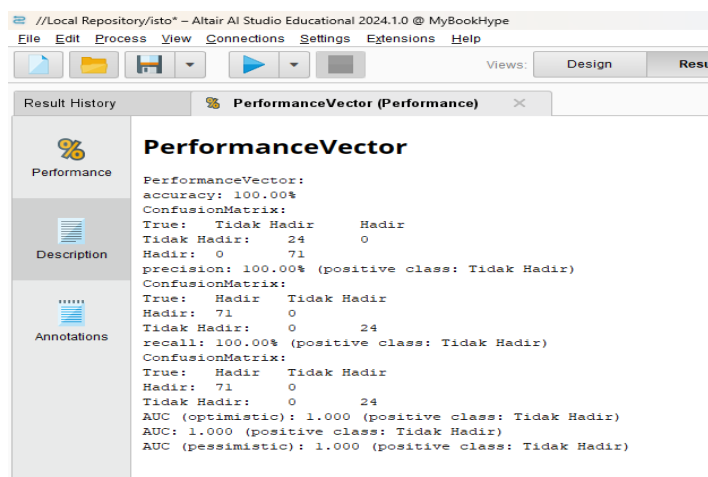
6. Performance Vektor

Hasil Performance Vector di atas menunjukkan bahwa model Naïve Bayes yang digunakan dalam pengujian memiliki performa yang sempurna dengan tingkat akurasi 100%. Berikut adalah rincian hasilnya:

- Akurasi (Accuracy): Model mencapai akurasi 100%, yang berarti semua prediksi yang dihasilkan oleh model sesuai dengan label sebenarnya tanpa adanya kesalahan.
- Confusion Matrix: Untuk kelas Tidak Hadir, model memprediksi 24 data dengan benar

- sebagai "Tidak Hadir" dan tidak ada kesalahan prediksi. Untuk kelas Hadir, model memprediksi 71 data dengan benar sebagai "Hadir" dan juga tidak ada kesalahan prediksi.
- c. Precision (Positive Class: Tidak Hadir): Precision 100% untuk kelas positif "Tidak Hadir" menunjukkan bahwa semua data yang diprediksi sebagai "Tidak Hadir" memang benar-benar berasal dari kelas tersebut. Tidak ada data yang salah diklasifikasikan sebagai "Tidak Hadir."
 - d. Recall (Positive Class: Tidak Hadir): Recall 100% menunjukkan bahwa model berhasil mengidentifikasi semua data yang sebenarnya masuk dalam kelas "Tidak Hadir" tanpa ada yang terlewatkan.

Secara keseluruhan, hasil ini menunjukkan bahwa model Naïve Bayes yang digunakan dalam pengujian memberikan klasifikasi yang sangat baik, dengan performa sempurna dalam memprediksi kedua kelas (Hadir dan Tidak Hadir). Data tersebut dapat dilihat pada gambar 5 dibawah ini:



Gambar 5. Perfoma Vector

Dari 95 data, tingkat kehadiran adalah 75% (71 orang) dan tingkat ketidakhadiran adalah 25% (24 orang). Sedangkan berdasarkan data reel dari total 324 data, persentase kehadiran adalah 58% (188 orang) dan persentase ketidakhadiran adalah 42% (136 orang). Sehingga dapat disimpulkan bahwa kedua data diatas memiliki tingkat kehadiran yang tinggi menunjukkan bahwa sebagian besar orang hadir dari pada presentase ketidakhadiran, sehingga tingkat ketidakhadiran yang signifikan mungkin memerlukan perhatian lebih lanjut.

3.4. Pembahasan

Naive Bayes dipilih karena kemampuannya mengolah data dengan banyak atribut serta efisien dalam menghasilkan klasifikasi berbasis probabilitas. Data penelitian terdiri dari beberapa atribut seperti usia, jenis kelamin, lokasi tempat tinggal, aksesibilitas ke TPS, pendapatan bulanan, pendidikan, pekerjaan, dan kepedulian politik. Masing-masing variabel diasumsikan memiliki pengaruh terhadap partisipasi masyarakat dalam pemilihan.

Proses analisis dilakukan menggunakan aplikasi RapidMiner yang mempermudah pemodelan prediktif dengan antarmuka visual. Data terlebih dahulu melalui tahap preprocessing untuk membersihkan data yang tidak konsisten, menghilangkan nilai kosong, dan memastikan semua data berada dalam format yang sesuai. Selanjutnya, data dilatih menggunakan algoritma Naive Bayes, kemudian diuji menggunakan data testing yang belum pernah dilihat oleh model.

Hasil dari penerapan model menunjukkan bahwa atribut seperti Aksesibilitas ke TPS dan Kepedulian Politik memiliki pengaruh dominan terhadap prediksi kehadiran. Individu yang tinggal dekat dengan TPS dan memiliki tingkat kepedulian politik tinggi lebih cenderung untuk hadir. Variabel lain yang juga memperkuat prediksi kehadiran adalah status pekerjaan (memiliki pekerjaan), tingkat pendidikan minimal SMA, usia tua, dan jenis kelamin laki-laki. Sebaliknya, masyarakat dengan akses TPS yang jauh, kepedulian politik rendah, penghasilan rendah, tidak bekerja, berpendidikan rendah, usia remaja, dan perempuan lebih banyak diprediksi tidak hadir.

Model prediksi yang dibangun dengan algoritma Naive Bayes berhasil memprediksi bahwa dari 95 data testing, sebanyak 71 orang (75%) diperkirakan hadir dalam pemilihan kepala daerah, dan 24 orang (25%) diprediksi tidak hadir. Sementara itu, berdasarkan data aktual dari total 324 data masyarakat, tercatat bahwa 188 orang (58%) benar-benar hadir dan 136 orang (42%) tidak hadir. Hal ini menunjukkan bahwa prediksi model cukup akurat, meskipun masih terdapat sedikit perbedaan dengan kenyataan, yang bisa disebabkan oleh variabel lain di luar model yang belum terobservasi atau oleh keterbatasan data.

Secara keseluruhan, hasil akhir ini menunjukkan bahwa model yang dibangun mampu mengidentifikasi pola-pola kehadiran secara efektif dan dapat digunakan sebagai alat bantu pengambilan keputusan. Hasil ini juga diharapkan menjadi acuan bagi pemerintah desa atau penyelenggara pemilu dalam merancang strategi peningkatan partisipasi masyarakat, misalnya melalui penyediaan akses TPS yang lebih mudah dan peningkatan sosialisasi politik di masyarakat yang kurang teredukasi. Dengan kata lain, algoritma Naive Bayes tidak hanya berfungsi sebagai alat klasifikasi, tetapi juga memberikan wawasan penting untuk strategi implementatif yang berdampak pada kebijakan partisipatif.

4. KESIMPULAN

Dari hasil penelitian mengenai kehadiran masyarakat dalam Pemilihan Kepala Daerah di Desa Matabesi menggunakan metode klasifikasi Naive Bayes, dapat disimpulkan bahwa proses prediksi dilakukan melalui beberapa tahapan utama. Tahapan tersebut meliputi pengumpulan data historis pemilih, pembersihan dan persiapan data, pemilihan fitur-fitur yang relevan, pembagian data menjadi data latih dan data uji, pembangunan model Naive Bayes, serta evaluasi model menggunakan Confusion Matrix. Model ini dirancang untuk memprediksi kemungkinan seorang pemilih akan hadir pada hari pemungutan suara berdasarkan karakteristik demografis, geografis, politik, dan sosial yang dimilikinya. Hasil prediksi ini diharapkan dapat membantu pihak penyelenggara pemilu dalam merencanakan strategi serta menetapkan kebijakan yang lebih tepat sasaran dalam meningkatkan partisipasi pemilih. Berdasarkan evaluasi Confusion Matrix, model prediksi yang dibangun mampu mengklasifikasikan data secara cukup akurat. Model berhasil memprediksi dengan benar sebanyak 71 orang yang hadir (prediksi positif yang benar), serta memprediksi dengan benar sebanyak 24 orang yang tidak hadir (prediksi negatif yang benar). Hasil ini menunjukkan bahwa algoritma Naive Bayes cukup efektif dalam mengidentifikasi pola kehadiran masyarakat dan dapat dijadikan sebagai acuan dalam pengembangan sistem pendukung keputusan di bidang kepemiluan.

Daftar Pustaka

- [1] M. N. Zarti, E. Sahputra, A. Sonita, and Y. Apridiansyah, "Application of Data Mining Using the Naïve Bayes Classification Method to Predict Public Interest Participation in the 2024 Elections," *Jurnal Komputer, Informasi dan Teknologi (JKOMITEK)*, vol. 3, no. 1, pp. 105–114, 2023. [Online]. Available: <https://doi.org/10.53697/jkomitek.v3i1.1192>
- [2] G. P. Kawani, "Implementasi Naive Bayes," *Journal of Informatics, Information System, Software Engineering and Applications (INISTA)*, vol. 1, no. 2, pp. 73–81, 2019. [Online]. Available: <https://doi.org/10.20895/inista.v1i2.73>
- [3] E. O. Manhitu, Y. P. K. Kelen, and D. Chrisinta, "Implementasi algoritma k-nearest neighbor untuk klasifikasi omset usaha mikro di kabupaten Timor Tengah Utara," *Zonasi Jurnal Sistem Informasi*, vol. 7, no. 1, pp. 304–316, 2025.
- [4] Y. R. Nasution, Suhardi, and I. H. Satrio, "Penerapan Algoritma Klasifikasi Naïve Bayes Untuk Analisis Sentimen Tentang Pemilu 2024," *JURNAL ELEKRONIKA DAN KOMPUTER*, vol. 17, no. 2, pp. 635–644, 2024. [Online]. Available: <http://jurnal.goretanpena.com/index.php/JSSR>
- [5] A. Pangestu, "Penerapan Data Mining Menggunakan Algoritma K-Means Pengelompokan Pelanggan Berdasarkan Kubikasi Air Terjual Menggunakan WEKA," *Jurnal Teknologi dan Sistem Komputer*, vol. 11, no. 3, pp. 67–71, 2021.
- [6] V. R. Prasetyo, H. Lazuardi, A. A. Mulyono, and C. Lauw, "Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Regresi Linier," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 7, no. 1, pp. 8–17, 2021. [Online]. Available:

- <https://doi.org/10.25077/teknosi.v7i1.2021.8-17>
- [7] R. Rachman and R. N. Handayani, "Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM," *Jurnal Informatika*, vol. 8, no. 2, pp. 111–122, 2021. [Online]. Available: <https://doi.org/10.31294/ji.v8i2.10494>
- [8] W. I. Rahayu, A. Anindita, and M. N. Fauzan, "Penentuan Validasi Data Pemilih dan Klasifikasi Hasil Pemilu DPRD Kab. Bone untuk Memprediksi Partai Pemenang Menggunakan Metode Naive Bayes," *Jurnal Teknik Informatika*, vol. 14, no. 1, pp. 32–39, 2022.
- [9] N. Rahmadani, R. Risnawati, and M. D. Sena, "Penerapan Algoritma Naïve Bayes Dalam Penentuan Kelayakan Penerima Bantuan Program Keluarga Harapan," *Jurnal Teknisi*, vol. 3, no. 2, p. 40, 2023. [Online]. Available: <https://doi.org/10.54314/teknisi.v3i2.1391>
- [10] M. Simanjuntak, N. Nurfalinda, et al., "Penerapan Metode Naive Bayes Untuk Memprediksi Status Kehadiran Masyarakat Dalam Pemilihan Gubernur," *Student Online Journal*, pp. 152–162, 2022. [Online]. Available: <https://soj.umrah.ac.id/index.php/SOJFT/article/view/1576>
- [11] A. S. Fitriani, "Penerapan Data Mining Menggunakan Metode Klasifikasi Naïve Bayes untuk Memprediksi Partisipasi Pemilihan Gubernur," *JTAM (Jurnal Teori dan Aplikasi Matematika)*, vol. 3, no. 2, pp. 98–104, 2019.
- [12] K. D. Usfinit, Y. P. K. Kelen, B. Baso, and H. H. Ullu, "Sistem Pakar Mendiagnosa Penyakit Pada Tanaman Jagung Menggunakan Metode Teorema Bayes Berbasis Website," *Jurnal Informatika Teknologi dan Sains (JINTEKS)*, vol. 7, pp. 326–335, 2025.
- [13] Dawis, A. M., Rahmayanti, D., Rachman, T., Impron, A., & Kelen, Y. P. K. (2025). *Pendekatan Modern Dalam Analisis Dan Desain Teknologi Informasi*. Get Press Indonesia. ISBN: 978-623-125-601-0.
- [14] Kelen, Yoseph PK, et al. "Decision support system for the selection of new prospective students using the simple additive weighted (SAW) method." *AIP Conference Proceedings*. Vol. 2798. No. 1. AIP Publishing, 2023. <https://doi.org/10.1063/5.0154676>.



ZONasi: Jurnal Sistem Informasi

Is licensed under a [Creative Commons Attribution International \(CC BY-SA 4.0\)](https://creativecommons.org/licenses/by-sa/4.0/)