



Klasifikasi Partisipasi Pemilih pada Pemilihan Walikota Bima Tahun 2024 Menggunakan Metode *Naive Bayes Classifier*

Anggreni¹, Akrimul Hakim¹, Arif Rahman¹, Muhammad Imam Dinata¹

¹. Program Studi Sistem dan Teknologi Informasi, Universitas Muhammadiyah Mataram, Indonesia.

ABSTRACT

Demographic factors that influence voter participation rates, such as age, gender, and education level, are important so that the results can serve as a basis for strategies to increase voter participation in elections. The data used in this study comes from a survey of 800 respondents in Bima City, which includes variables such as age, gender, education, and level of participation in elections. This study aims to analyze and classify voter participation in the 2024 Bima Mayor Election using the Naïve Bayes Classifier method. The Naïve Bayes Classifier method is chosen because of its ability to effectively classify categorical data based on the principle of conditional probability. The process begins with a data preprocessing stage to clean and transform raw data into numerical data ready to be processed by the algorithm, followed by splitting the data into training and testing sets. Classification was carried out by building a model to predict voter participation status as "Yes" or "No". The developed model was able to perform effective classification by grouping voter data into the categories of "Yes" and "No" participation based on age, gender, and education variables, with an overall accuracy of 91%.

Keywords: Demographic Data, classification, Naïve Bayes Classifier, voters, Mayor

ABSTRAK

Faktor-faktor demografis yang mempengaruhi tingkat partisipasi pemilih, seperti umur, jenis kelamin, dan tingkat Pendidikan adalah hal yang penting, agar hasilnya dapat menjadi dasar bagi strategi meningkatkan partisipasi pemilih dalam pemilihan umum. Data yang digunakan dalam penelitian ini berasal dari survei terhadap 800 responden di Kota Bima, yang mencakup variabel umur, jenis kelamin, pendidikan, dan tingkat partisipasi dalam pemilu. Penelitian ini bertujuan untuk menganalisis dan mengklasifikasikan partisipasi pemilih dalam Pemilihan Wali Kota Bima tahun 2024 dengan memanfaatkan metode *Naïve Bayes Classifier*. Metode *Naïve Bayes Classifier* dipilih karena kemampuannya dalam klasifikasi data kategorikal secara efektif dengan prinsip probabilitas kondisional. Proses dimulai dengan tahapan preprocessing data untuk membersihkan dan mengubah data mentah menjadi data numerik yang siap diproses algoritma, diikuti dengan pembagian data menjadi data pelatihan dan pengujian. Klasifikasi dilakukan dengan membangun model untuk memprediksi status partisipasi pemilih sebagai "Ya" atau "Tidak". Model yang dikembangkan berhasil melakukan proses klasifikasi yang efektif dengan mengelompokkan data pemilih ke dalam kategori partisipasi "Ya" dan "Tidak" berdasarkan variabel usia, jenis kelamin, dan pendidikan, dengan akurasi keseluruhan mencapai 91%.

Keywords: Data Demografis, klasifikasi, *Naïve Bayes Classifier*, pemilih, Walikota

DOI: <https://doi.org/10.35746/jsn.v4i1.898>



1. Pendahuluan

Pilkada memberikan wewenang bagi masyarakat untuk memilih pemimpin yang mewakili kepentingan dan aspirasi lokal dan masyarakat semakin menyadari pentingnya peran mereka dalam menentukan arah pemerintahan daerah (Purba, 2023). Masyarakat luas dapat menyuarakan aspirasi melalui Pilkada yang menjadi sebuah pintu gerbang utama untuk mewujudkan suatu pemerintahan yang mungkin tidak dapat tercapai dari periode sebelumnya (Rotua Tinambunan & Oktafian Abrianto, 2021). Pemilihan kepala daerah dan wakil kepala daerah dimaksudkan untuk mencerminkan prinsip demokrasi, di mana kedaulatan berada di tangan rakyat dan dilaksanakan secara langsung melalui pemilihan umum (Maruci, 2025). Pemilihan Kepala Daerah (Pilkada) Kota Bima tahun 2024 menjadi momentum penting dalam perkembangan demokrasi lokal. Tingkat partisipasi pemilih tercatat sekitar 75% (KPU Kota Bima, 2024). Salah satu tantangan utama terlihat pada kelompok pemilih muda jumlah pemilih muda pada Pemilu 2024 sebanyak 18.430.639 pemilih atau 9 persen pemilih keseluruhan (BPS, 2023). Menjelang Pilkada November 2024, beberapa masalah meresahkan publik. Salah satunya adalah kandidat yang akan maju dalam pemilihan, munculnya politik dinasti, politik identitas, dan kemungkinan adalah masalah bagi kandidat yang akan dalam Pilkada 2024 (Romaito,dkk 2021).

Naïve Bayes adalah teknik klasifikasi probabilitas sederhana, menerapkan teorema Bayes, yang mampu menangani data kuantitatif dan diskrit untuk menghitung perkiraan probabilitas yang diperlukan untuk jenis analisis (Aeni,2020). Metode Naïve Bayes mengasumsikan bahwa atribut-atribut dalam data adalah independent satu sama lain. Naïve Bayes mengelompokan data dengan menghitung probabilitas kemunculan kelas target berdasarkan atribut-atribut yang terkait dengan memprediksi peluang dimasa yang akan datang berdasarkan pengalaman dimasa lampau (Putra, 2023). Naïve Bayes tidak bisa mendeteksi gambar, tetapi hanya bisa mendeteksi teks dan numerik. Untuk menghitung probabilitas dalam metode ini menggunakan pendekatan teorema bayes (Thwe, 2021). Naïve Bayes sangat sederhana dan efisien dalam hal waktu komputasi, sehingga dapat menangani dataset yang sangat besar tanpa memerlukan sumber daya komputasi yang besar selain itu metode ini sering kali memberikan hasil yang sangat baik bahkan dengan jumlah data latih yang terbatas (Wijaya, 2022). Analisis sentimen dari berbagai aspek menggunakan metode Naïve Bayes Classifier dalam konteks Pilkada 2024 di Indonesia menawarkan potensi besar untuk memahami opini publik dan meningkatkan proses demokrasi digital Dimana hasil analisis ini dapat menjadi alat yang bermanfaat bagi berbagai pemangku kepentingan, mulai dari pemerintah, lembaga pemilu, hingga tim kampanye politik (Ardiyani, 2025).

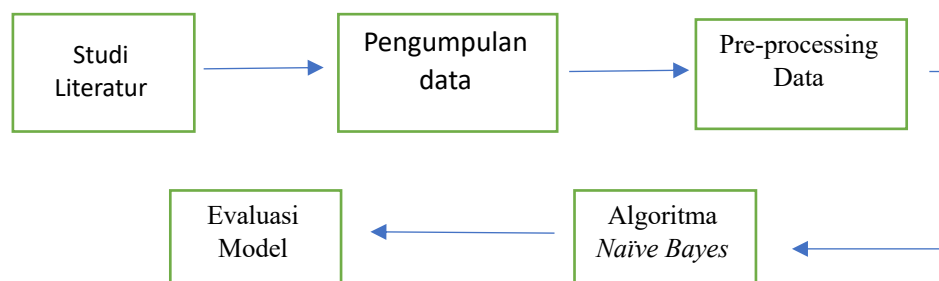
Analisis sentimen publik menggunakan metode *Naïve Bayes* dan *K-Nearest Neighbor terbukti efektif dalam mengklasifikasikan teks ke dalam kategori tertentu, seperti positif, negative* (Winarti, 2021). Penggunaan algoritma Naive Bayes menunjukkan akurasi tinggi dengan keunggulan dalam kecepatan dan pengolahan data berukuran besar mencapai 85,00%. Naïve Bayes memberikan pandangan yang komprehensif dalam memahami perilaku pemilih di Jakarta (Umat,dkk 2024). Metode Naïve Bayes Classifier terbukti efektif dalam mengklasifikasikan sentimen Masyarakat dalam pemilihan kepala daerah Tangerang Selatan 2024, data diambil dari aplikasi 'X' dan TikTok. Model yang dibangun mampu mengidentifikasi sentimen dengan tingkat akurasi yang cukup tinggi, yaitu 81,97% untuk data dari aplikasi 'X' dan 84,8% untuk TikTok (Prayoga, 2025). Berdasarkan hasil penelitian analisis sentimen terhadap calon kepala daerah 2024 menggunakan metode Naive Bayes dengan total data tweet 6465, terdapat 4169 tweet positif dan 2296 tweet dapat disimpulkan bahwa metode ini efektif dalam menganalisis sentiment secara otomatis, memberikan hasil cukup baik. Dari evaluasi model Naïve Bayes tersebut, MultinomialNB adalah yang terbaik dengan akurasi 75% (Fitriani, 2025). Mola (2025) dalam penelitiannya yang

berjudul Analisis Sentimen Terhadap Data Komentar Publik Mengenai Isu UU Pilkada 2024 Menggunakan Metode Naïve Bayes menyatakan bahwa metode ini lebih sensitif terhadap sentimen negatif, namun memiliki kelemahan dalam mengenali sentimen netral dan positif secara akurat. Hasil prediksi berdasarkan atribut kehadiran menunjukkan bahwa dari 105 data Testing, tingkat akurasi prediksi mencapai 97%, sedangkan tingkat kesalahan prediksi adalah 3%.

Penelitian diharapkan dapat menjadi inspirasi dan memahami dinamika partisipasi pemilih di Kota Bima, terutama dalam konteks pemilihan walikota yang akan datang. Dengan menggunakan metode *Naïve Bayes Classifier* yang telah terbukti efektif dalam konteks pemilu nasional dan media sosial, diharapkan penelitian ini dapat memberikan pengetahuan baru mengenai faktor-faktor yang mempengaruhi partisipasi pemilih serta menghasilkan model prediksi yang akurat dan implementatif. Selain itu, penelitian ini juga berkontribusi pada pengembangan ilmu pengetahuan dalam bidang analisis data dan politik, serta memberikan manfaat praktis bagi penyelenggara pemilu dalam upaya meningkatkan partisipasi masyarakat dan memberikan kontribusi yang jelas dalam memahami dan menganalisis partisipasi pemilih pada Pemilihan Walikota Bima tahun 2024.

2. Metodeologi Penelitian

Terdapat beberapa langkah yang harus dilakukan untuk dapat mengklasifikasikan partisipasi pemilih Pilkada Kota Bima 2024 dengan menggunakan algoritma *Naïve Bayes Classifier*. Berikut adalah langkah-langkah yang dilakukan:



Gambar 1. Tahapan Penelitian Klasifikasi Partisipasi Pemilih

Keterangan Gambar 1 :
Tahapan Penelitian

1. Studi Literatur

Dilakukan untuk memahami penelitian terdahulu terkait klasifikasi partisipasi pemilih, pemanfaatan algoritma *Naïve Bayes*, dan keterkaitannya dengan faktor demografis. Literatur dikaji dari jurnal nasional maupun internasional untuk memperoleh dasar teori yang kuat.

2. Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan untuk memperoleh informasi yang relevan terkait partisipasi masyarakat dalam Pemilihan Walikota Bima Tahun 2024. Proses pengumpulan data dilaksanakan melalui metode survei menggunakan kuesioner sebagai instrumen utama. Kuesioner disusun berdasarkan indikator partisipasi pemilih dan disebarkan secara langsung kepada masyarakat di seluruh wilayah Kota Bima. Teknik pengambilan data dilakukan secara random sampling, yaitu pemilihan responden secara acak tanpa memandang latar belakang sosial, pendidikan, atau usia, guna menjaga

objektivitas dan representasi populasi. Jumlah total responden yang berhasil dikumpulkan sebanyak 800 responden, dapat dilihat pada Tabel 1.

Tabel 1. Hasil Kuisisioner

No	Nama	Umur	Jenis Kelamin	Pendidikan	Status	Pekerjaan	Partisipasi
1	Sahrul ramadhan	37	laki-laki	SMA	Menikah	Formal	ya
2	Linda Andriani	27	Perempuan	SD	Menikah	Formal	tidak
3	Puspa Winarti	34	Perempuan	SMP	Cerai	Informal	ya
4	Jumrati Sani	26	Perempuan	SD	Belum Menikah	Formal	tidak
5	Jaenab	51	Perempuan	S1	Menikah	Formal	ya
6	Harijah	48	Perempuan	S2	Menikah	Informal	tidak
7	Alimuddin	58	laki-laki	SMA	Menikah	Formal	ya
8	Nur Hasanah	31	Perempuan	SMP	Menikah	Informal	tidak
9	Hamzah	25	laki-laki	SD	Menikah	Formal	ya
10	Nurhayati	26	Perempuan	SD	Menikah	formal	tidak

3. *Pre-processing* Data

Tahap ini meliputi pembersihan data, penghapusan data tidak lengkap, dan pengubahan data kategorikal ke bentuk numerik melalui *label encoding*. Sebagai contoh, variabel jenis kelamin dikodekan menjadi laki-laki = 1, perempuan = 2, sedangkan tingkat pendidikan dikodekan SD = 1, SMP = 2, SMA = 3, dan S1 = 4. Encoding dan One-Hot Encoding. Selain *label encoding*, dilakukan juga *one-hot encoding* pada variabel kategorikal agar lebih kompatibel dengan algoritma klasifikasi. Misalnya, tingkat pendidikan SMA akan diubah menjadi format biner [0,0,1,0]. Split Data Dataset dibagi menjadi dua bagian, yaitu 80% data latih (640 data) dan 20% data uji (160 data). Teknik ini memastikan bahwa model dilatih pada sebagian besar data, namun tetap diuji menggunakan data yang belum pernah dilihat.

Tabel 2. Pre Prosessing

Umur	Jenis Kelamin	Pendidikan	Partisipasi
37	1	3	1
27	2	1	2
34	2	2	1
26	2	1	2
51	2	4	1
48	2	4	2
58	1	3	1
31	2	2	2
25	1	1	1
26	1	1	2

4. Penerapan Algoritma *Naïve Bayes*

Algoritma yang digunakan adalah *Gaussian Naïve Bayes* karena sesuai untuk data dengan distribusi campuran (numerik dan kategorikal). Model dilatih menggunakan data latih, kemudian diuji dengan data uji untuk memprediksi apakah seorang responden akan berpartisipasi atau tidak.

5. Evaluasi Model

Performa model diukur menggunakan beberapa metrik:

- a. *Confusion matrix* untuk melihat distribusi prediksi benar dan salah.
- b. Akurasi, precision, recall, dan f1-score untuk menilai kualitas prediksi secara menyeluruh. Evaluasi dilakukan tidak hanya pada keseluruhan data, tetapi juga per variabel (umur, jenis kelamin, pendidikan).

3. Hasil dan Pembahasan

Pada proses klasifikasi menggunakan algoritma Naïve Bayes, data yang digunakan merupakan data yang telah melalui tahapan pre-processing. Data ini telah berbentuk numerik sehingga siap untuk proses selanjutnya, yaitu pengacakan data atau shuffling. Pengacakan data bertujuan untuk menghilangkan pengaruh urutan asli data sehingga model dapat belajar pola tanpa bias urutan. Berikut merupakan tabel data hasil dari tahapan pre-processing sebelum dilakukan pengacakan.

Tabel 3. Data Penelitian Sebelum Di *Shuffling*

No	Umur	Jenis Kelamin	Pendidikan	Partisipasi
1.	37	1	3	1
2.	27	2	1	2
3.	34	2	2	1
4.	26	2	1	2
5.	51	2	4	1
6.	48	2	4	2
7.	58	1	3	1
8.	31	2	2	2
9.	25	1	1	1
10.	26	2	1	2

Data pada Tabel 3 telah berbentuk numerik sehingga dapat langsung diproses pada tahapan selanjutnya. Setelah itu data dilakukan pengacakan urutan (*shuffle*) untuk memastikan distribusi data acak dan menghindari bias yang muncul dari urutan data awal. Berikut merupakan tabel hasil pengacakan data:

Tabel 4. Data Penelitian Setelah di *Shuffling*

No	Umur	Jenis Kelamin	Pendidikan	Partisipasi
1.	25	1	1	1
2.	27	2	1	2
3.	48	2	4	2
4.	37	1	3	1
5.	31	2	2	2
6.	34	2	2	1
7.	26	2	1	2
8.	51	2	4	1
9.	26	2	1	2
10.	58	1	3	1

Setelah pengacakan data selesai, langkah berikutnya adalah memisahkan data menjadi dua bagian utama yaitu fitur (umur, jenis kelamin, pendidikan) dikodekan dengan (x) dan target partisipasi dikodekan dengan (y). fitur merupakan variable input yang digunakan model untuk belajar, sedangkan target adalah variable output yang ingin diprediksi. Tabel 5 merupakan pemisahan fitur (x) yang terdiri dari umur, jenis kelamin, dan Pendidikan:

Tabel 5 Pemisahan Fitur Umur, Jenis Kelamin, Pendidikan (X)

No	X		
	Umur	Jenis Kelamin	Pendidikan
1.	25	1	1
2.	27	2	1
3.	48	2	4

No	X		
	Umur	Jenis Kelamin	Pendidikan
4.	37	1	3
5.	31	2	2
6.	34	2	2
7.	26	2	1
8.	51	2	4
9.	26	2	1
10.	58	1	3

Sedangkan target partisipasi (y) yang berisi nilai partisipasi adalah sebagai berikut:

Tabel 6. Pemisahan Target Partisipasi (Y)

No	Y
	Partisipasi
1.	1
2.	2
3.	2
4.	1
5.	2
6.	1
7.	2
8.	1
9.	2
10.	1

Tahapan ini penting agar model dapat dikenal pada variabel input dan output secara terpisah sehingga proses training dapat berjalan dengan baik.

Tahapan selanjutnya merupakan *categorical encoding* menggunakan *one-hot encoding* pada fitur kategori yaitu jenis kelamin dan pendidikan, data di ubah ke bentuk numerik biner. Tahapan ini mengubah setiap kategori menjadi kolom baru yang berisi nilai 0 atau 1. Tabel 7 adalah hasil data fitur setelah dilakukan tahapan one-hot encoding:

Tabel 7. Hasil Tahapan *One-Hot Encoding*

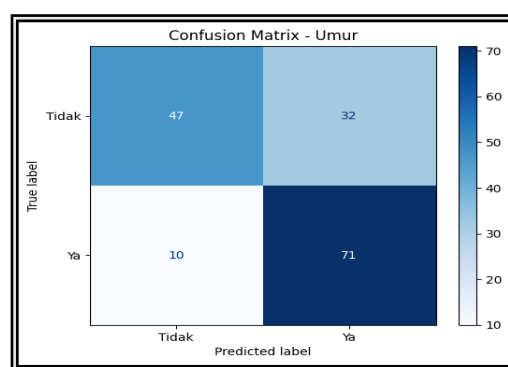
No	Umur	Jk_1	Jk_2	Pndk_1	Pndk_2	Pndk_3	Pndk_4
1.	25	1	0	1	0	0	0
2.	27	0	1	1	0	0	0
3.	48	0	1	0	0	0	1
4.	37	1	0	0	0	1	0
5.	31	0	1	0	1	0	0
6.	34	0	1	0	1	0	0
7.	26	0	1	1	0	0	0
8.	51	0	1	0	0	0	1
9.	26	0	1	1	0	0	0
10.	58	1	0	0	0	1	0

Dengan *one-hot encoding*, setiap kategori menjadi angka biner terpisah agar informasi kategori tetap terjaga namun bentuknya sesuai dengan kebutuhan input model numerik.

Pada tahap persiapan fitur (*feature preparation*), dilakukan seleksi dan penggabungan fitur yang akan digunakan dalam model. Fitur numerik seperti umur langsung dipertahankan, sedangkan fitur kategori yang telah *diencode* disiapkan sebagai subset yang terpisah. Kemudian, seluruh fitur tersebut digabungkan menjadi satu set data lengkap yang siap untuk proses klasifikasi. Setelah itu, data dibagi dalam dua kelompok yakni data training dan data testing dengan rasio 80:20, yang mana data training digunakan untuk membangun model dan data testing digunakan untuk evaluasi performa model secara independen.

Model klasifikasi *Gaussian Naïve Bayes* kemudian dibuat dan ditraining menggunakan data training yang telah disiapkan. Setelah proses training selesai, diterapkan evaluasi terhadap model yang telah dibangun dengan mengukur berbagai metrik kinerja, yakni *accuracy*, *precision*, *recall*, dan *f1-score* untuk memberikan gambaran menyeluruh mengenai kemampuan model dalam mengenali dan mengklasifikasikan data dengan benar. Selain itu, *confusion matrix* juga dianalisis guna melihat distribusi klasifikasi prediksi yang benar dan salah, sehingga evaluasi terhadap kekuatan dan kelemahan model dapat dilakukan secara lebih mendalam. Dengan rangkaian proses ini, didapatkan sebuah model *klasifikasi Naïve Bayes* yang tidak hanya terlatih secara efektif tetapi juga telah melalui pengujian kinerja yang baik, sehingga hasil klasifikasi yang diperoleh dapat diandalkan dan memberikan nilai interpretatif yang kuat terhadap data yang dianalisis.

Hasil *confusion matrix* dari variabel umur, merupakan hasil dari *confusion matrix* untuk rasio data training dan data testing 80:20 yang dapat dilihat pada gambar dibawah ini:



Gambar 2. Hasil *Confusion Matrix* Fitur Umur

Gambar *confusion matrix* berdasarkan fitur **Umur**, model memprediksi sebanyak 47 data “Tidak” dengan benar (True Negative), namun terdapat 32 data yang salah diklasifikasikan sebagai “Ya” (False Positive). Untuk kelas “Ya”, sebanyak 10 data salah diprediksi menjadi “Tidak” (False Negative), dan 71 data berhasil diprediksi dengan benar (True Positive). Hal ini menunjukkan bahwa meskipun model cukup sensitif dalam mengenali kelas “Ya”, masih ada kesalahan cukup signifikan dalam membedakan kelas “Tidak” berdasarkan fitur umur.

Pada Tabel 8 Hasil uji model pada fitur umur, merupakan hasil dari uji model dengan mengukur, *precision*, *recall*, *accuracy*, dan *f1-score* yang dapat dilihat pada tabel dibawah ini:

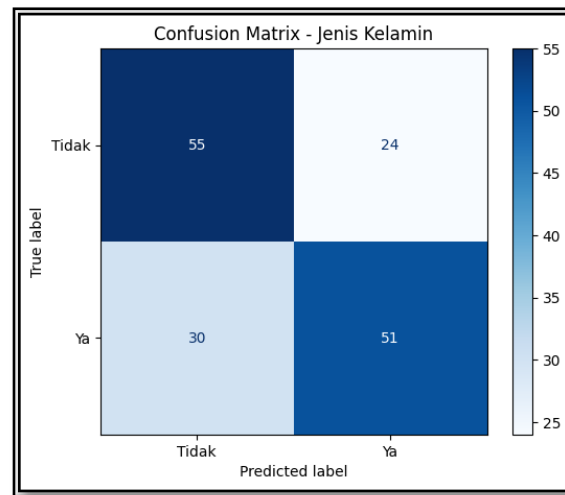
Tabel 8. Hasil Metrik Fitur Umur

Metrik	Tidak	Ya
Precision	0.82	0.69
Recall	0.59	0.88
F1-Score	0.69	0.77
Support	79	81
Accuracy	0.74	

Model klasifikasi yang menggunakan fitur umur menampilkan performa yang cukup memadai. Untuk kelas “Tidak”, nilai precision sebesar 0.82 mencerminkan bahwa 82% data yang diprediksi sebagai “Tidak” memang benar berasal dari kelas tersebut, menunjukkan kemampuan model dalam menghindari prediksi salah pada kelas ini. Namun, recall sebesar 0.59 mengindikasikan bahwa model hanya mampu mendeteksi sekitar 59% dari total data aktual “Tidak”, sehingga terdapat sejumlah prediksi yang melewatkan kelas ini. Pada sisi kelas “Ya”, recall yang tinggi sebesar 0.88 menandakan bahwa model mampu mengenali dengan sangat baik data positif, meskipun precision-nya sedikit lebih rendah pada angka 0.69, yang berarti ada 31% prediksi “Ya” yang merupakan false alarm. F1-Score sebagai harmonisasi antara precision dan

recall berada di angka 0.69 untuk kelas “Tidak” dan 0.77 untuk kelas “Ya.” Secara keseluruhan, akurasi keseluruhan model mencapai 74%, menandakan kemampuan model yang moderat dalam klasifikasi berdasarkan fitur umur tetapi dengan kecenderungan untuk lebih sensitif dalam mendeteksi kelas “Ya.”

Hasil *confusion matrix* dari variabel jenis kelamin, merupakan hasil dari *confusion matrix* untuk rasio data training dan data testing 80:20 yang dapat dilihat pada Gambar 3 ini:



Gambar 3. Hasil *Confusion Matrix* Fitur Jenis Kelamin

Gambar *confusion matrix* untuk fitur Jenis Kelamin, model menghasilkan nilai prediksi dengan jumlah True Negative (TN) sebanyak 55, yang berarti sebanyak 55 data aktual berlabel “Tidak” berhasil diprediksi dengan benar sebagai “Tidak”. Namun, sebanyak 24 data lainnya salah diklasifikasikan menjadi “Ya” (False Positive/FP). Sementara itu, model juga salah memprediksi 30 data berlabel “Ya” menjadi “Tidak” (False Negative/FN), dan hanya 51 data yang berhasil diprediksi dengan benar sebagai “Ya” (True Positive/TP). Hasil ini menunjukkan bahwa model memiliki performa yang cukup baik dalam mengenali kelas “Tidak”, meskipun terdapat kesalahan yang cukup pada pengidentifikasian kelas “Ya”.

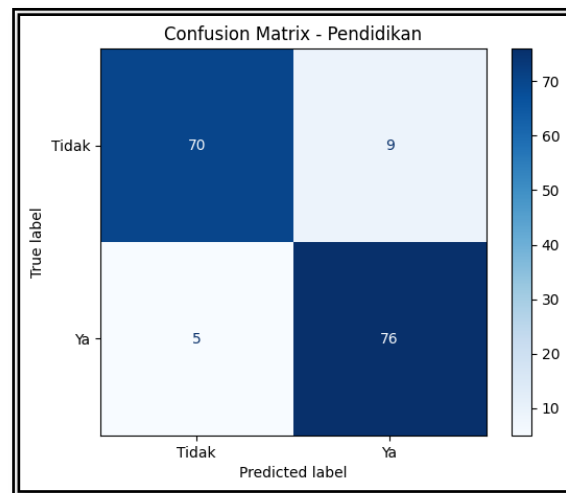
Pada Tabel 9 Hasil uji model variabel jenis kelamin, merupakan hasil dari uji model dengan mengukur accuracy, precision, recall, dan f1-score yang dapat dilihat pada Tabel 9.

Tabel 9. Hasil Metrik Fitur Jenis Kelamin

Metrik	Tidak	Ya
Precision	0.65	0.68
Recall	0.7	0.63
F1-Score	0.67	0.65
Support	79	81
Accuracy	0.66	

Penggunaan fitur jenis kelamin sebagai variabel prediktor menghasilkan performa yang relatif lebih rendah dibandingkan fitur umur. Nilai precision pada kelas “Tidak” dan “Ya” masing-masing sebesar 0.65 dan 0.68, yang menggambarkan bahwa hanya sekitar dua pertiga prediksi model yang tepat pada kedua kelas ini. Untuk recall, model mampu mengenali 70% contoh aktual kelas “Tidak” sementara hanya 63% untuk kelas “Ya,” menunjukkan kemampuan deteksi yang kurang seimbang dan kurang optimal. F1-Score yang relatif rendah yakni 0.67 untuk kelas “Tidak” dan 0.65 untuk kelas “Ya” menandakan adanya kebutuhan untuk perbaikan dalam hal keseimbangan antara precision dan recall. Dengan akurasi keseluruhan sebesar 66%, model menunjukkan keterbatasan jika hanya mengandalkan fitur jenis kelamin sebagai variabel prediktor, sehingga kurang dapat diandalkan sebagai satu-satunya fitur dalam proses klasifikasi.

Hasil *confusion matrix* dari variabel pendidikan, merupakan hasil dari *confusion matrix* untuk rasio data training dan data testing 80:20 yang dapat dilihat pada Gambar 4.



Gambar 4. Hasil *Confusion Matrix* Fitur Pendidikan

Gambar *confusion matrix* untuk fitur Pendidikan, model menghasilkan performa yang sangat mirip dengan gabungan fitur. True Negative (TN) mencapai 70, dengan False Positive (FP) sebanyak 9. Sedangkan untuk kelas “Ya”, False Negative (FN) berjumlah 5, dan True Positive (TP) mencapai 76. Hasil ini menggambarkan bahwa fitur pendidikan memberikan kontribusi yang kuat dan konsisten terhadap kinerja model dalam klasifikasi, dengan tingkat akurasi baik pada kedua kelas.

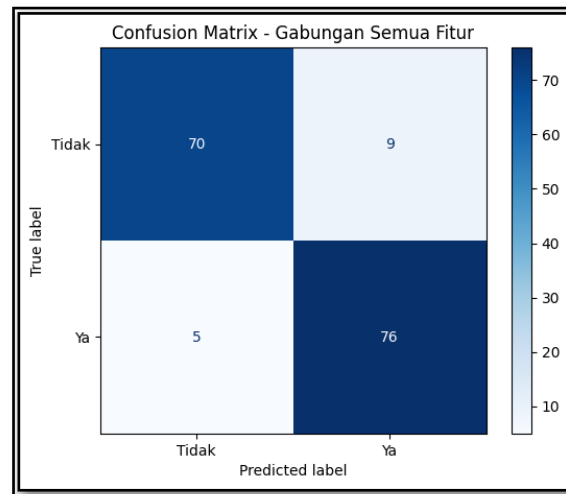
Pada Tabel 10 Hasil uji model variabel pendidikan, merupakan hasil dari uji model dengan mengukur accuracy, precision, recall, dan f1-score yang dapat dilihat pada tabel dibawah ini:

Tabel 10. Hasil Metrik Fitur Pendidikan

Metrik	Tidak	Ya
Precision	0.93	0.89
Recall	0.89	0.94
F1-Score	0.91	0.92
Support	79	81
Accuracy	0.91	

Pada penggunaan fitur pendidikan, model memperlihatkan performa yang sangat baik dan signifikan dibandingkan fitur tunggal lainnya. Precision yang tinggi, yakni 0.93 pada kelas “Tidak” dan 0.89 pada kelas “Ya,” menunjukkan kemampuan model untuk melakukan prediksi yang sangat akurat dan minim kesalahan false positive. Recall yang tinggi 0.89 untuk kelas “Tidak” dan 0.94 untuk kelas “Ya” menandakan kemampuan model dalam mendeteksi proporsi tinggi data yang benar-benar berasal dari masing-masing kelas. Nilai F1-Score yang mencapai 0.91 dan 0.92 untuk masing-masing kelas memperlihatkan keseimbangan yang optimal antara precision dan recall. Dengan akurasi keseluruhan mencapai nilai 91%, penggunaan fitur pendidikan memberikan kontribusi positif yang kuat terhadap kekuatan model klasifikasi ini, menjadikannya variabel yang sangat relevan dan informatif dalam konteks prediksi.

Gambar 5 Hasil *confusion matrix* gabungan semua variabel, merupakan hasil dari *confusion matrix* untuk rasio data training dan data testing 80:20



Gambar 5. Hasil *Confusion Matrix* Keseluruhan Fitur

Gambar *confusion matrix* yang menggunakan Gabungan Semua Fitur, model menunjukkan hasil yang lebih baik. Tingkat True Negative (TN) mencapai 70, yang berarti 70 data berlabel “Tidak” berhasil diklasifikasikan dengan benar. Kesalahan False Positive (FP) tercatat hanya 9. Di sisi lain, False Negative (FN) sebanyak 5 data berlabel “Ya” yang salah diprediksi sebagai “Tidak”, sementara True Positive (TP) mencapai 76 data. Hasil ini mengindikasikan bahwa dengan menggunakan gabungan semua fitur, model menjadi lebih akurat dan lebih efektif dalam membedakan kelas “Ya” maupun “Tidak”.

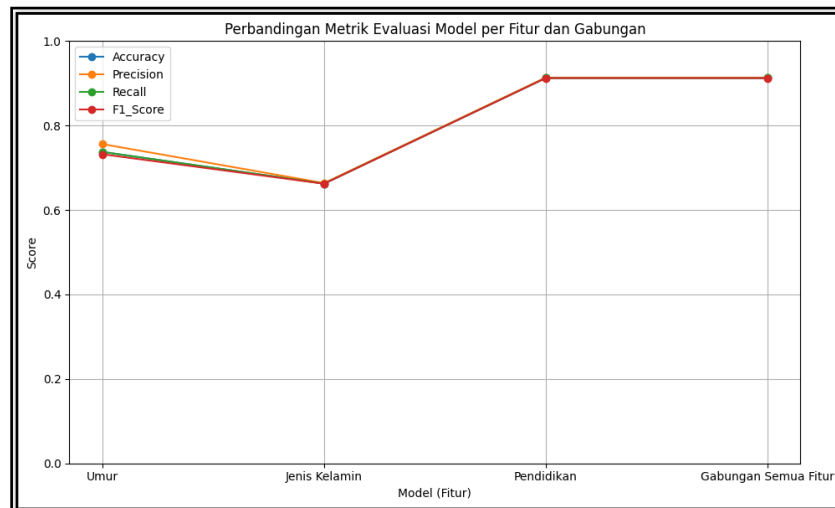
Pada Tabel 11 Hasil uji model gabungan semua variabel, merupakan hasil dari uji model dengan mengukur accuracy, precision, recall, dan f1-score yang dapat dilihat pada tabel dibawah ini:

Tabel 11. Hasil Metrik Keseluruhan Fitur

Metrik	Tidak	Ya
Precision	0.93	0.89
Recall	0.89	0.94
F1-Score	0.91	0.92
Support	79	81
Accuracy	0.91	

Menggunakan gabungan semua fitur yang tersedia, model mempertahankan performa yang sangat baik dan serupa dengan penggunaan fitur pendidikan tunggal. Precision kelas “Tidak” berada pada angka 0.93 dan kelas “Ya” sebesar 0.89, yang memperlihatkan prediksi yang sangat akurat di kedua kelas. Recall yang tetap tinggi di angka 0.89 untuk kelas “Tidak” dan 0.94 untuk kelas “Ya” mengindikasikan model sangat efektif dalam mendeteksi data sebenarnya dari kedua kelas secara seimbang. F1-Score mencapai 0.91 dan 0.92, memberikan gambaran keseimbangan optimal antara precision dan recall. Akurasi keseluruhan tetap sebesar 91%, mengindikasikan bahwa integrasi seluruh fitur memberikan hasil yang sangat signifikan, meningkatkan ketahanan dan ketepatan model dalam mengklasifikasikan data ke dalam kelas yang benar.

Untuk mengevaluasi hasil dari masing-masing fitur independen terhadap performa model *Naïve Bayes*, dilakukan visualisasi berupa grafik garis akurasi. Gambar 6 menampilkan perbandingan nilai akurasi model ketika dijalankan dengan satu fitur tunggal tunggal (umur, jenis kelamin, dan pendidikan) maupun dengan gabungan ketiga fitur.



Gambar 6. Grafik Line Keseruluhan Variabel

Berdasarkan Gambar 6 yang menampilkan grafik perbandingan metrik evaluasi model pada berbagai fitur, dapat diobservasi bahwa model dengan performa terbaik dalam hal akurasi dan metrik lainnya adalah model yang memanfaatkan fitur tingkat pendidikan serta gabungan semua fitur, dengan nilai akurasi mencapai sekitar 0.91. Sebaliknya, model yang menggunakan fitur jenis kelamin memiliki akurasi dan metrik evaluasi lainnya yang berada di posisi terendah, dengan akurasi sekitar 0.66. Sedangkan fitur umur menampilkan performa yang cukup baik, dengan akurasi sekitar 0.74.

Hasil ini menunjukkan bahwa dalam memprediksi partisipasi pemilih pada Pilkada Kota Bima Tahun 2024, fitur pendidikan memberikan kontribusi paling signifikan terhadap keakuratan model, baik klasifikasi secara individual maupun klasifikasi dengan kombinasi gabungan keseluruhan fitur. Hal ini sejalan dengan performa metrik precision, recall, dan F1-Score yang juga menunjukkan nilai tertinggi pada fitur ini dan gabungan semua fitur. Sebaliknya, fitur jenis kelamin memiliki pengaruh yang lebih rendah terhadap performa model, menandakan bahwa variabel ini kurang efektif sebagai prediktor tunggal untuk partisipasi pemilih dalam konteks data yang dianalisis.

Adapun fluktuasi metrik dalam grafik ini memperlihatkan bahwa meskipun penggunaan beberapa fitur secara bersama-sama dapat meningkatkan kestabilan dan keseimbangan performa, tidak semua fitur memberikan kontribusi yang setara. Temuan ini penting sebagai evaluasi kritis dalam menentukan variabel mana yang paling memengaruhi model klasifikasi partisipasi pemilih, serta memberikan arah untuk optimasi model berikutnya dengan fokus pada fitur edukasi dan kombinasi fitur yang tepat.

4. Pembahasan

Hasil penelitian ini memperlihatkan bahwa metode NBC sangat efektif untuk mengklasifikasikan partisipasi pemilih pada Pilkada lokal. Tingginya akurasi menunjukkan bahwa faktor demografi sederhana seperti umur, jenis kelamin, dan pendidikan dapat menjadi prediktor yang cukup kuat. Temuan menarik lain adalah kecenderungan partisipasi lebih tinggi pada kelompok perempuan dan berpendidikan SMA/S1. Hal ini memberikan implikasi praktis bagi KPU Kota Bima untuk memperkuat sosialisasi politik pada pemilih muda dan berpendidikan rendah yang partisipasinya masih relatif rendah.

5. Kesimpulan dan Saran

Model yang dikembangkan berhasil melakukan proses klasifikasi yang efektif dengan mengelompokkan data pemilih ke dalam kategori partisipasi "Ya" dan "Tidak" berdasarkan variabel usia, jenis kelamin, dan pendidikan, dengan akurasi keseluruhan mencapai 91%. Disarankan menggunakan dataset yang lebih banyak dan beragam agar hasil lebih akurat dan bisa mewakili populasi yang lebih luas. Melengkapi data dengan fitur atau variabel tambahan yang dirasa dapat berpengaruh pada partisipasi pemilih. Mencoba metode klasifikasi atau algoritma lain untuk membandingkan kinerja dan menemukan model yang optimal dalam memprediksi partisipasi selain dengan metode naive bayes classifier.

Daftar Pustaka

- Ardiyani, B. (2025). ANALISIS KAMPANYE CALON WALIKOTA DALAM PILKADA KOTA BEKASI TAHUN 2024 MELALUI MEDIA. Bagus Surya Ardiyan. WALIKOTA_DALAM_PILKADA_KOTA_BEKASI_TAHUN_2024_MELALUI_MEDIA. https://www.academia.edu/127068692/ANALISIS_KAMPANYE_CALON_WALIKOTA_DALAM_PILKADA_KOTA_BEKASI_TAHUN_2024_MELALUI_MEDIA.
- E. S. Romaito, M. K. Anam, Rahmadden, Ulfah, and A. Noviciate, "Dalam Analisa Sentimen Pilkada Pada Twitter," CSRID (Computer Science Research and Its Development Journal), pp. 169–179, 2021, <https://doi.org/10.22303/csrid.13.3.2021.169-179>
- KPU Kota Bima (2024). Laporan Kinerja Komisi Pemilihan Umum. [LAKIP Kobi 2024 Ketua.pdf - Google Drive](#)
- K. Aeni and M. F. Asy'ari, "Prediksi Kepuasan Layanan Akademik Menggunakan Algoritma Naïve Bayes," JATISI (Jurnal Tek. Inform. dan Sist. Informasi), vol. 7, no. 3, pp. 601–609, 2020, <https://doi.org/10.35957/jatisi.v7i3.603>.
- Maruci, C. (2025). "The 2025 Regional Head Elections (Pilkada) Phenomenon: Preparation, Candidates, and Political Dynamics in the Election", Sultan Agung Notary Law Review, 7(1), 21. <https://doi.org/10.30659/SANLAR.7.1.21-40>
- Mola, S. A. S. (2025). "Analisis Sentimen Terhadap Data Komentar Publik Mengenai Isu UU Pilkada 2024 Menggunakan Metode Naïve Bayes dan K-Nearest Neighbor", Journal of Informatics Management and Information Technology. ISSN 2773-44744 (Media Online). 5(3), 337-343, <https://doi.org/10.47065/jimat.v5i3.514>
- Purba, A. R. (2023). Tinjauan Yuridis Pelaksanaan Pemilihan Kepala Daerah (Pilkada) Serentak. *Jurnal Ilmiah METADATA*, 5(3), 47-64. <https://doi.org/10.47652/metadata.v5i3.393>
- Prayoga, R. G. (2025). Analisis Sentimen Masyarakat Terhadap Pelaksanaan Pemilihan Kepala Daerah Tangerang Selatan 2024 Menggunakan Metode Naïve Bayes Clasifier. Universitas Islam Negeri Syarif Hidayatullah Jakarta. <https://repository.uinjkt.ac.id/dspace/handle/123456789/89016>
- Tinambunan, H. S. R., & Abrianto, B. O. (2021). Persoalan Dilematis Lembaga Perwakilan Daerah Dalam Sistem Ketatanegaraan Indonesia. *Jurnal Magister Hukum Udayana (Udayana Master Law Journal)*, 10(1), 64-78. <https://doi.org/10.24843/JMHU.2021.V10.I01.P06>
- Putra, R. F., Zebua, R. S. Y., Budiman, B., Rahayu, P. W., Bangsa, M. T. A., Zulfadhilah, M., ... & Andiyani, A. (2023). *Data Mining: Algoritma dan Penerapannya*. PT. Sonpedia Publishing Indonesia.
- Statistik Politik, Volume 11 ISSN 2303-2448 BPS,2023 [Statistik Politik 2023 - Badan Pusat Statistik Indonesia](#)
- Thwe, P. (2021). Naïve Bayes Classifier for Sentiment Analysis. <https://www.academia.edu/download/87172887/IJCIRAS1711.pdf>
- Umat, Y. N. K., Nafsyi, D. R., Kusumaningsih, D., & Hakim, L. (2024). Analisis Faktor Yang Mempengaruhi Pemilihan Gubernur Daerah Khusus Jakarta Menggunakan Algoritma

- Naive Bayes Dan Regresi Logistik. *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab*, 9(2), 211-224. <https://doi.org/10.36341/rabit.v9i2.4778>
- Wijaya, B. A., Kumar, V., Jhon Wau, B. F., Tanjung, J. P., & Dharshinni, N. P. (2022). Application of Data Mining using Naive Bayes for Student Success Rates in Learning. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 6(4), 1980. <https://doi.org/10.30865/MIB.V6I4.4639>
- Winarti, T., Indriyawati, H., Vydia, V., & Christanto, F. W. (2021). Performance comparison between naive bayes and k-nearest neighbor algorithm for the classification of Indonesian language articles. *IAES International Journal of Artificial Intelligence*, 10(2), 452. <https://doi.org/10.11591/ijai.v10.i2.pp452-457>