

## **Analisis Sentimen Terhadap Identitas Kependudukan Digital Menggunakan Algoritma Multinomial Naïve Bayes**

### ***Sentiment Analysis Of Identity Digital Population Using The Naïve Bayes Multinomial Algorithm***

**Alevia Mentari Putri<sup>1</sup>, Rosita Yanuarti\*<sup>2</sup>, Moh Dasuki<sup>3</sup>, Agus Milu Susetyo<sup>4</sup>**

<sup>1,2,3</sup>Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember

<sup>4</sup>Program Studi Pendidikan Bahasa dan Sastra Indonesia, FKIP, Universitas Muhammadiyah Jember

Email: <sup>1</sup>[aleviamp@gmail.com](mailto:aleviamp@gmail.com), <sup>2</sup>[rosita.yanuarti@unmuhjember.ac.id](mailto:rosita.yanuarti@unmuhjember.ac.id), <sup>3</sup>[moh.dasuki22@unmuhjember.ac.id](mailto:moh.dasuki22@unmuhjember.ac.id),

<sup>4</sup>[agusmilus@unmuhjember.ac.id](mailto:agusmilus@unmuhjember.ac.id)

\*Penulis Koresponden

Diterima: 27 Mei 2025

Direvisi: 29 Juli 2025

Disetujui: 26 Agustus 2025



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).  
Copyright (c) 2025 JUSTINDO

#### **ABSTRAK**

Inovasi baru dalam perkembangan teknologi kini pemerintah telah meluncurkan inovasi baru yaitu Identitas Kependudukan Digital berbasis digital melalui aplikasi yang dijadikan sebagai tujuan pemerintah dalam mengurangi cetak fisik KTP dan blangko sehingga lebih efisien dan mudah dalam mengakses identitas masyarakat dalam smartphone masing-masing dan digunakan untuk mengelola data identitas masyarakat. Aplikasi ini telah banyak menimbulkan pro dan kontra dari masyarakat yang menilai sebuah aplikasi IKD ini dalam mendukung digitalisasi, sehingga tujuan dari penelitian ini diperlukan analisis sentimen untuk mengetahui penilaian serta ulasan masyarakat terhadap adanya IKD tersebut sehingga dapat dijadikan sebagai bahan evaluasi oleh pemerintah. Media sosial banyak yang digunakan dalam menyampaikan pendapat serta opini masyarakat tentang IKD ini terutama pada platform X atau Twitter, dengan pengambilan data dengan teknik *crawling*. Analisis pada platform X ini sangat penting karena menjadi sumber pengumpulan data yang cepat dan luas dari penyampaian masyarakat tentang IKD tersebut. Pendekatan penelitian ini menggunakan algoritma Multinomial Naive Bayes untuk mengklasifikasi analisis sentimen masyarakat. Dari hasil pengujian menggunakan *K-Fold Cross Validation*, didapatkan hasil tertinggi pada 10 Fold uji ke-8 dengan nilai akurasi tertinggi 95%, presisi 94% dan recall 100%, sehingga algoritma Multinomial Naive Bayes ini cocok untuk dijadikan sebagai metode analisis data.

**Kata kunci:** IKD, analisis sentimen, multinomial naive bayes, x

#### **ABSTRACT**

*New innovations in technological development, now the government has launched a new innovation, namely digital-based Digital Population Identity through an application which is used as the government's goal of reducing the physical printing of KTPs and blanks so that it is more efficient and easier to access people's identities on their respective smartphones and is used to manage data. community identity. This application has given rise to many pros and cons, so the aim of this research requires sentiment analysis to find out the public's assessment and reviews of the existence of the IKD so that it can be used as evaluation material by the government. Social media is widely used to convey public opinion and opinions about IKD, especially on platform X or Twitter, by collecting data using crawling techniques. This research approach uses the Multinomial Naive Bayes algorithm to classify public sentiment analysis. From the test results using K-Fold Cross Validation, the highest results were obtained in the 8th 10 Fold test with the highest accuracy value of 95%, precision 94% and recall 100%, so the Multinomial Naive Bayes algorithm is suitable to be used as a data analysis method.*

**Keywords:** IKD, sentiment analysis, multinomial naive bayes, x

## 1. Pendahuluan

Perkembangan teknologi memunculkan sebuah inovasi baru yang canggih, dengan berkembangnya teknologi ini hamper setiap orang terkoneksi dengan internet yang dapat digunakan untuk mengakses segala informasi dengan mudah dan cepat. Perkembangan ini menjadikan perubahan komunikasi dari era industrialisasi menuju era informasi sehingga tidak ada lagi hambatan untuk saling interaksi (Munti dkk., 2020). Dengan adanya kecanggihan teknologi banyak bermunculan sebuah media sosial yang dapat digunakan dalam berbagi dan menciptakan jejaring sosial yang mengubah komunikasi menjadi dialog interaktif (Liedfray dkk., 2022).

Pemerintah telah meluncurkan sebuah inovasi baru dalam bidang administrasi kependudukan yaitu Identitas Kependudukan Digital berbasis digital yang digunakan untuk menyajikan dan menampilkan data penduduk melalui smartphone (Mirlana, et al., 2024). Dengan adanya IKD tersebut masyarakat akan lebih mudah dalam membawa KTP dan dapat lebih mudah dalam pemrosesan data, tetapi banyak perbincangan dalam masyarakat terutama pada media social X telah banyak komentar terhadap aplikasi tersebut, hal ini banyak menimbulkan pro dan kontra dari masyarakat, terutama di daerah pelosok karena keterbatasan akses internet (Hidayat, Rahman, Perdana, & Arbansyah, 2024). IKD ini sangat relevan ntuk dianalisis karena terdapat pendapat dan penilaian masyarakat yang dapat dijadikan sebagai bahan penilaian terhadap aplikasi.

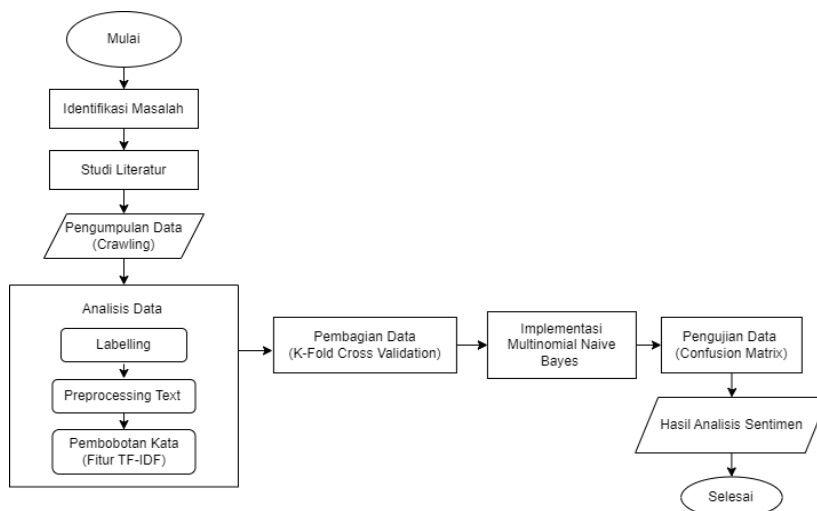
Pada analisis informasi masyarakat tentang Identitas Kependudukan Digital ini dilakukan dengan cara analisis sentimen. Analisis sentimen atau *Opinion Mining* yaitu salah satu cabang klasifikasi dalam mengekstrak data teks pada bidang pengolahan Bahasa alami atau *Natural Language* (NLP), *text mining* dengan tujuan untuk menganalisis sentiment, opini, komentar, ulasan dari sikap seseorang yang berkaitan dengan suatu produk atau fenomena tertentu (Mailoa 2021). Dalam analisis sentiment banyak terdapat algoritma yang digunakan untuk mengklasifikasikan dokumen atau teks, salah satu algoritma yang digunakan yaitu Multinomial Naïve Bayes.

Algoritma Multinomial Naïve Bayes adalah salah satu variasi dari Naïve Bayes Classifier, algoritma klasifikasi yang banyak digunakan dalam text mining serta analisis sentimen. Perbedaan algoritma Naïve Bayes Classifier dengan Multinomial Naïve Bayes ini dalam mengasumsikan tipe data berupa data diskrit seperti data teks yang terdiri dari kata-kata yang sering muncul dalam dokumen dan Tingkat akurasi yang lebih spesifik (Sabrani, Wedashwara and Bimantoro, 2020).

Tujuan dari penelitian ini adalah memahami sentimen masyarakat terhadap adanya Identitas Kependudukan Digital yang memiliki pro dan kontra pada masyarakat sehingga cocok digunakan dalam analisis sentimen berupa data teks dengan pengambilan data pada X. Hasil dari penelitian ini dapat dijadikan sebagai bahan penilaian dalam pengembangan lebih lanjut mengenai pengelolaan aplikasi tersebut.

## 2. Metode Penelitian

Metode penelitian merupakan proses atau langkah-langkah dalam pemecahan masalah sehingga memperoleh hasil yang diharapkan.



Gambar 1. Metode Penelitian

### 2.1. Identifikasi Masalah

Pada tahapan identifikasi masalah ini yaitu menganalisis permasalahan dalam menemukan sentimen pada data X (Twitter) tentang pandangan masyarakat terhadap adanya Identitas Kependudukan Digital dengan menggunakan algoritma multinomial naïve bayes.

### 2.2. Studi Literatur

Tahapan studi literatur ini melakukan, mencari serta mempelajari referensi terkait dari berbagai sumber berupa artikel atau jurnal yang berkaitan dengan penelitian yang dapat digunakan sebagai acuan.

### 2.3. Pengumpulan Data

Pengumpulan data penelitian yang digunakan yaitu data komentar pada X (Twitter), data dikumpulkan dengan Teknik *Crawling* data dengan mencari kata kunci terkait dan menggunakan Google Colab sebagai tools, pada bulan Oktober 2023 – April 2024 sebanyak 800 data dan telah dilakukan pemilahan data sehingga digunakan 620 data.

### 2.4. Labelling

Pelabelan data dilakukan secara manual oleh penulis dan divalidasi oleh Bapak Agus Milu Susetyo, M.Pd selaku dosen ahli Bahasa, data yang dilabeli yaitu berupa sentimen positif dan sentimen negatif.

### 2.5. Preprocessing Text

Tahapan preprocessing ini data yang telah dilakukan pelabelan akan dijadikan menjadi data terstruktur melalui beberapa tahapan sebagai berikut:

#### a. Cleansing

Tahapan untuk menghilangkan review dari kata-kata berupa karakter yang tidak diperlukan seperti, angka, URL, emoji, hastag, username dan lainnya.

#### b. Case folding

Pada tahap ini mengubah semua huruf menjadi huruf kecil (lowercase).

#### c. Tokenizing

Tahapan ini kalimat akan dipotong menjadi per kata.

#### d. Normalisasi

Tahapan ini mengembalikan kata atau membenarkan kata yang mengalami kesalahan penulisan kedalam bentuk asli.

#### e. Stopword Removal

Pada tahapan ini mengambil kata-kata penting serta meningkatkan fokus analisis pada kata.

#### f. Stemming

Tahapan ini melakukan penghapusan pada kata awalan, sisipan dan berimbuhan pada kalimat menjadi kata dasar untuk mengurangi variasi kata yang memiliki makna yang sama.

### 2.6. Pembobotan TF-IDF

Pembobotan TF-IDF ini digunakan dalam menghitung bobot setiap kata dan menentukan nilai frekuensi sebuah kata dalam dokumen, sehingga seberapa sering kata muncul dalam dokumen. Perhitungan TF-IDF ini menunjukkan seberapa relevan kata yang terdapat pada dokumen. Dengan melakukan ekstraksi fitur pada setiap term pada dokumen. TF yaitu menghitung jumlah term (token) pada dataset, IDF adalah perhitungan bobot yang seberapa penting ditemukannya kata pada dokumen yang berbeda, jika semakin tinggi suatu kata yang muncul maka akan diberikan bobot rendah (Santoso, Armansyah, & Desliani 2022). Adapun perhitungan TF sebagai berikut:

$$tf_{t,d} = \frac{n_{t,d}}{N}. \quad (1)$$

Keterangan:

$tf_{t,d}$  = frekuensi kemunculan kata pada sebuah dokumen

$n_{t,d}$  = nilai istilah yang muncul

$N$  = jumlah semua term dalam dokumen

Setelah menghitung TF, selanjutnya yaitu menghitung nilai IDF dengan rumus yang digunakan sebagai berikut:

$$idf_d = \log \frac{N}{df} \quad (2)$$

Keterangan:

$idf_d$  = jumlah kemunculan suatu kata dalam dokumen

$N$  = total dokumen

$df$  = banyak dokumen yang mengandung term

Setelah berhasil menghitung nilai IDF, selanjutnya menentukan nilai TF-IDF, hasil pembobotan kata menggunakan TF-IDF ini merupakan hasil perkalian dari TF dengan IDF, ditunjukkan dengan rumus sebagai berikut:

$$tf\ idf_{t,d} = tf_{t,d} \times idf_d \quad (3)$$

Keterangan:

$tf\ idf_{t,d}$  = *Term Frequency-Inverse Document Frequency*

$tf_{t,d}$  = nilai TF

$idf_d$  = nilai IDF

## 2.7. K-Fold Cross Validation

K-Fold Cross Validation adalah Teknik validasi yang digunakan untuk mengurangi bias pada data dokumen, Teknik ini digunakan dalam penerapan perulangan model pada setiap data menjadi data latih dan data uji sehingga teruji validitasnya. K-Fold Cross Validation ini disebut juga dengan validasi silang yang menerapkan pemecahan data kedalam k-sub-set dengan data yang seimbang, data uji dan latih dilakukan secara iterative sebanyak k kali (Zhafira, Rahayudi, and Indriati 2021).

## 2.8. Multinomial Naïve Bayes

Multinomial Naïve Bayes ini termasuk bagian dari Naïve Bayes yang digunakan dalam untuk klasifikasi sederhana. Multinomial Naïve Bayes adalah pembelajaran supervised dengan menggunakan model probabilistic yang dipengaruhi term yang diperhitungkan, dengan tahapan memasukkan data untuk dilakukan perhitungan peluang muncul kelas pada data latih (Java and Syafrullah 2024). Tahapan metode Multinomial Naïve Bayes ini melalui perhitungan probabilitas prior, probabilitas kata-kata dan probabilitas dokumen. Perhitungan untuk menentukan peluang awal (probability prior) dengan rumus sebagai berikut:

$$P(c) = \frac{N_c}{N} \quad (4)$$

Keterangan:

$P(C)$  = kategori atau kelas

$N_c$  = banyaknya kategori c pada dokumen latih

$N$  = banyaknya keseluruhan dokumen latih yang digunakan

Untuk perhitungan probabilitas kata ke-n yang digunakan dalam pembobotan kata TF-IDF dengan rumus sebagai berikut:

$$P(t|c) = \frac{W_{ct}+1}{(\sum_{w' \in v} W_{ct'})+B'} \quad (5)$$

Keterangan:

$W_{ct}$  = nilai bobot TF-IDF atau nilai W dari term pada dokumen kategori c

$\sum_{w' \in v} W_{ct'}$  = total semua bobot TF-IDF atau semua W pada seluruh term pada kategori c

$B'$  = total W atau nilai IDF seluruh term pada semua dokumen

Selanjutnya yaitu melakukan tahap klasifikasi data berdasarkan hasil perhitungan sebagai berikut:

$$P(c|term\ dok\ d) = P(c) \times P(t_1|c) \times P(t_2|c) \times P(t_n|c) \quad (6)$$

Keterangan:

- $P(c|term\ dok\ d)$  = probabilitas suatu dokumen termasuk dalam kategori c
- $P(c)$  = probabilitas prior dari kategori c
- $P(t_n|c)$  = probabilitas kata ke-n pada kategori c

2.9. Confusion Matrix

Confusion Matrix adalah pengukuran untuk pengklasifikasian dalam data dengan tujuan untuk menghitung Tingkat nilai akurasi, presisi dan recall. Confusion Matrix disebut juga metode klasifikasi dan prediksi untuk kinerja Tingkat keberhasilan atau akurasi dengan jumlah baris sesuai dengan jumlah data testing (A. K. Santoso 2022). Perhitungan pengujian nilai pada confusion matrix ditunjukkan sebagai berikut:

Tabel 1. Confusion Matrix

Nilai Prediksi	Nilai Sebenarnya	
	Positif	Negatif
Positif	TP	FP
Negatif	FN	TN

Menurut Han et al. (2012) menyebutkan ada 4 *building blocks* yang digunakan dalam pengukuran evaluasi confusion matrix, sebagai berikut:

- 1). True Positive (TP) adalah banyaknya data prediksi positif yang berhasil diklasifikasi dengan benar.
- 2). True Negative (TN) adalah banyaknya data prediksi negatif yang berhasil diklasifikasi dengan benar.
- 3). False Positive (FP) adalah banyaknya data prediksi negatif yang salah diklasifikasikan sebagai data prediksi positif.
- 4). False Negative (FN) adalah banyaknya data prediksi positif yang salah diklasifikasikan sebagai data prediksi negatif.

Untuk rumus Confusion Matrix mempresentasikan tingkat akurasi dari klasifikasi data pada dokumen Dimana Tingkat akurasi ini menunjukkan jumlah prediksi benar.

$$\text{Akurasi} : \frac{TP+TN}{TP+TN+FP+FN} \tag{7}$$

$$\text{Presisi} : \frac{TP}{TP+FP} \tag{8}$$

$$\text{Recall} : \frac{TP}{TP+FN} \tag{9}$$

3. Hasil dan Pembahasan

3.1. Pengumpulan Data

Pengumpulan data pada penelitian ini diambil menggunakan Google Colab melalui proses crawling dengan cara menggunakan library yang disediakan oleh bahasa pemrograman Python yang tersambung pada platform X (Twitter) dengan kata pencarian Identitas Kependudukan Digital dengan jumlah data keseluruhan yang dicari pada periode bulan Oktober 2023 – April 2024 sebanyak 800 data, Dimana data tersebut telah dilakukan pemilahan sehingga data yang digunakan dalam penelitian ini sebanyak 620 data.

Gambar 2. Hasil Pengumpulan Data

### 3.2. Labelling Data

Proses pelabelan data ini terdapat dua jenis sentimen yaitu sentimen positif dan sentimen negative, pada pelabelan ini dilakukan secara manual oleh penulis dan dosen ahli Bahasa. Hasil dari pelabelan data ditunjukkan dengan angka 0 dan 1, Dimana angka 0 menunjukkan komentar negative dan 1 menunjukkan komentar positif.

	full_text	KLASIFIKASI
0	@bitxt @_salmonsashimi_ No 2 nya Identitas Kependudukan Digital Yang E-ktp aja e nya ga kepake ...	0
1	Mulai tahun 2024 pemerintah akan menggantikan KTP dengan identitas kependudukan digital (IKD) at...	1
2	Halo Warga Jakarta Dihimbau untuk seluruh Warga Jakarta AYO SEGERA DOWNLOAD APLIKASI IKD (Identi...	1
3	@niridns maksudnya Identitas Kependudukan Digital kak? penerapannya kakak harus datang langsung ...	0
4	Beneran urus perpanjangan paspor pede aja karena udah pake app Identitas Kependudukan Digital yang...	0
...	...	...
615	Yuk aktivasi Identitas Kependudukan Digital (IKD) sekarang juga! Berikut tata caranya ya	1
616	Pemerintah Kota (Pemkot) Tangerang melalui Dinas Kependudukan dan Catatan Sipil (Disdukcapil) Ko...	1
617	Pelaksanaan Aktivasi Identitas Kependudukan Digital dalam Kegiatan Rumah UKM Kebumen #DukcapilGo...	1
618	Dukcapil Klaten hadir kembali di CarFreeDay hari Minggu tgl 1 Oktober 2023 jam 06.30-09.00 lokas...	1
619	#TemanPemilih Ketua KPU Kab. Pinrang Muh. Ali Jodding menghadiri undangan Pj. Gubernur Sulawesi ...	1

Gambar 3. Hasil Labelling Data

### 3.3. Preprocessing Text

Data yang telah dilakukan pelabelan selanjutnya akan dilakukan preprocessing text sehingga menjadi data yang terstruktur dengan tujuan untuk menghilangkan noise, membersihkan dan menyederhanakan data teks menjadi makna yang jelas (Aditama, Pratama, Wiwaha, & Rakhmawati, 2020) beberapa tahapan sebagai berikut:

#### 1). Cleansing

Tahapan cleansing ini untuk menghilangkan sebuah karakter berupa angka, URL, hastag, emoji, username, tanda baca dan lain-lain untuk menghilangkan noise. Adapun hasil dari cleansing ditunjukkan pada gambar 4 dibawah ini

	full_text	KLASIFIKASI	cleansing
0	@bitxt @_salmonsashimi_ No 2 nya Identitas Kependudukan Digital Yang E-ktp aja e nya ga kepake ...	0	No nya Identitas Kependudukan Digital Yang Ektp aja e nya ga kepake apalagi yg aplikasi blass or...
1	Mulai tahun 2024 pemerintah akan menggantikan KTP dengan identitas kependudukan digital (IKD). B...	1	Mulai tahun pemerintah akan menggantikan KTP dengan identitas kependudukan digital IKD Berikut i...

Gambar 4. Hasil Tahapan Cleansing

#### 2). Case Folding

Pada tahapan case folding ini semua huruf akan diubah menjadi huruf kecil (lowercase). Hasil case folding ditunjukkan pada gambar 5 dibawah ini

	full_text	KLASIFIKASI	cleansing	case_folding
0	@bitxt @_salmonsashimi_ No 2 nya Identitas Kependudukan Digital Yang E-ktp aja e nya ga kepake ...	0	No nya Identitas Kependudukan Digital Yang Ektp aja e nya ga kepake apalagi yg aplikasi blass or...	no nya identitas kependudukan digital yang ektp aja e nya ga kepake apalagi yg aplikasi blass or...
1	Mulai tahun 2024 pemerintah akan menggantikan KTP dengan identitas kependudukan digital (IKD). B...	1	Mulai tahun pemerintah akan menggantikan KTP dengan identitas kependudukan digital IKD Berikut i...	mulai tahun pemerintah akan menggantikan ktp dengan identitas kependudukan digital ikd berikut i...

Gambar 5. Hasil Tahapan Case Folding

#### 3). Tokenizing

Setelah proses cleansing dan case folding, pada tokenizing ini dokumen akan dipotong dari kalimat menjadi per kata. Hasil dari tokenizing ditunjukkan pada gambar 6 dibawah ini

	full_text	KLASIFIKASI	cleansing	case_folding	tokenize
0	@bitxt @_salmonsashimi_ No 2 nya Identitas Kependudukan Digital Yang E-ktp aja e nya ga kepake ...	0	No nya Identitas Kependudukan Digital Yang Ektp aja e nya ga kepake apalagi yg aplikasi blass or...	no nya identitas kependudukan digital yang ektp aja e nya ga kepake apalagi yg aplikasi blass or...	[no, nya, identitas, kependudukan, digital, yang, ektp, aja, e, nya, ga, kepake, apalagi, yg, ap...
1	Mulai tahun 2024 pemerintah akan menggantikan KTP dengan identitas kependudukan digital (IKD). B...	1	Mulai tahun pemerintah akan menggantikan KTP dengan identitas kependudukan digital IKD Berikut i...	mulai tahun pemerintah akan menggantikan ktp dengan identitas kependudukan digital ikd berikut i...	[mulai, tahun, pemerintah, akan, menggantikan, ktp, dengan, identitas, kependudukan, digital, ik...

Gambar 6. Hasil Tahapan Tokenizing

4). Normalisasi

Tahapan normalisasi yaitu mengembalikan kata atau membenarkan kesalahan penulisan kedalam bentuk asli. Hasil normalisasi ditunjukkan pada gambar 7 dibawah ini

	full_text	KLASIFIKASI	cleansing	case_folding	tokenize	normalisasi
0	@_salmonsashimi_ No 2 nya Identitas Kependudukan Digital Yang E-ktp aja e nya ga kepake ...	0	No nya Identitas Kependudukan Digital Yang Ektp aja e nya ga kepake apalagi yg aplikasi blass or...	no nya identitas kependudukan digital yang ektp aja e nya ga kepake apalagi yg aplikasi blass or...	[no, nya, identitas, kependudukan, digital, yang, ektp, aja, e, nya, ga, kepake, apalagi, yg, ap...	[no, nya, identitas, kependudukan, digital, yang, ektp, saja, e, nya, tidak, pakai, apalagi, ...
1	Mulai tahun 2024 pemerintah akan menggantikan KTP dengan identitas kependudukan digital (IKD). B...	1	Mulai tahun pemerintah akan menggantikan KTP dengan identitas kependudukan digital IKD Berikut i...	mulai tahun pemerintah akan menggantikan ktp dengan identitas kependudukan digital ikd berikut i...	[mulai, tahun, pemerintah, akan, menggantikan, ktp, dengan, identitas, kependudukan, digital, ik...	[mulai, tahun, pemerintah, akan, menggantikan, ktp, dengan, identitas, kependudukan, digital, ik...

Gambar 7. Hasil Tahapan Normalisasi

5). Stopword Removal

Proses stopwords removal ini mengambil kata-kata penting untuk meningkatkan fokus analisis. Hasil stopwords removal pada gambar 8 dibawah ini

	full_text	KLASIFIKASI	cleansing	case_folding	tokenize	normalisasi	stopword removal
0	@_salmonsashimi_ No 2 nya Identitas Kependudukan Digital Yang E-ktp aja e nya ga kepake ...	0	No nya Identitas Kependudukan Digital Yang Ektp aja e nya ga kepake apalagi yg aplikasi blass or...	no nya identitas kependudukan digital yang ektp aja e nya ga kepake apalagi yg aplikasi blass or...	[no, nya, identitas, kependudukan, digital, yang, ektp, aja, e, nya, ga, kepake, apalagi, yg, ap...	[no, nya, identitas, kependudukan, digital, yang, ektp, saja, e, nya, tidak, pakai, apalagi, ...	[identitas, kependudukan, digital, ektp, saja, tidak, yang, aplikasi, tidak, buat.]
1	Mulai tahun 2024 pemerintah akan menggantikan KTP dengan identitas kependudukan digital (IKD). B...	1	Mulai tahun pemerintah akan menggantikan KTP dengan identitas kependudukan digital IKD Berikut i...	mulai tahun pemerintah akan menggantikan ktp dengan identitas kependudukan digital ikd berikut i...	[mulai, tahun, pemerintah, akan, menggantikan, ktp, dengan, identitas, kependudukan, digital, ik...	[mulai, tahun, pemerintah, akan, menggantikan, ktp, dengan, identitas, kependudukan, digital, ik...	[pemerintah, menggantikan, ktp, identitas, kependudukan, digital, ikd, informasi, terkait, ikd]

Gambar 8. Hasil Tahapan Stopword Removal

6). Stemming

Pada tahapan stemming ini dilakukan penghapusan pada kata awalan, sisipan dan berimbuhan pada kalimat menjadi kata dasar. Hasil stemming pada gambar 9 dibawah ini

	full_text	KLASIFIKASI	cleansing	case_folding	tokenize	normalisasi	stopword removal	stemming
0	@_salmonsashimi_ No 2 nya Identitas Kependudukan Digital Yang E-ktp aja e nya ga kepake ...	0	No nya Identitas Kependudukan Digital Yang Ektp aja e nya ga kepake apalagi yg aplikasi blass or...	no nya identitas kependudukan digital yang ektp aja e nya ga kepake apalagi yg aplikasi blass or...	[no, nya, identitas, kependudukan, digital, yang, ektp, aja, e, nya, ga, kepake, apalagi, yg, ap...	[no, nya, identitas, kependudukan, digital, yang, ektp, saja, e, nya, tidak, pakai, apalagi, ...	[identitas, kependudukan, digital, ektp, saja, tidak, yang, aplikasi, tidak, buat.]	[identitas, duduk, digital, ektp, saja, tidak, pakai, yang, aplikasi, tidak, buat]
1	Mulai tahun 2024 pemerintah akan menggantikan KTP dengan identitas kependudukan digital (IKD). B...	1	Mulai tahun pemerintah akan menggantikan KTP dengan identitas kependudukan digital IKD Berikut i...	mulai tahun pemerintah akan menggantikan ktp dengan identitas kependudukan digital ikd berikut i...	[mulai, tahun, pemerintah, akan, menggantikan, ktp, dengan, identitas, kependudukan, digital, ik...	[mulai, tahun, pemerintah, akan, menggantikan, ktp, dengan, identitas, kependudukan, digital, ik...	[pemerintah, menggantikan, ktp, identitas, kependudukan, digital, ikd, informasi, terkait, ikd]	[perintah, ganti, ktp, identitas, duduk, digital, ikd, informasi, kait, ikd]

Gambar 9. Hasil Tahapan Stemming

### 3.4. Pembobotan TF-IDF

Setelah tahapan preprocessing text, Term Frequency (TF) untuk menghitung jumlah token (term) pada dataset, Inverse Document Frequency (IDF) untuk menghitung bobot seberapa sering ditemukannya kata (term) dalam dokumen (Santoso, Armansyah, & Desliani, 2022), tahap selanjutnya pembobotan TF-IDF dengan melakukan ekstraksi fitur pada setiap term pada komentar dan akan menghasilkan nilai TF-IDF sebagai berikut:

Tabel 2. Hasil Pembobotan TF-IDF

No	Text	TF-IDF
1.	identitas duduk digital ektp saja tidak pakai yang aplikasi tidak buat	['0,0888', '0,0883', '0,0888', '0,2572', '0,3502', '0,5830', '0,3502', '0,3471', '0,2376', '0,3845']
2.	perintah ganti ktp identitas duduk digital ikd informasi kait ikd	['0,3386', '0,3415', '0,2719', '0,1136', '0,1129', '0,1136', '0,3234', '0,5044', '0,5450']

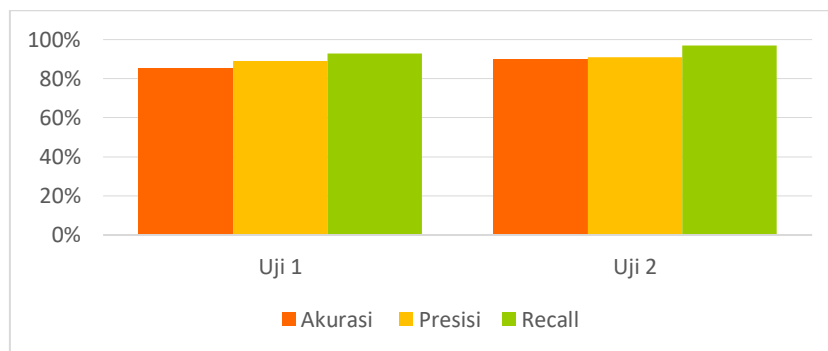
### 3.5. Pembagian dan Pengujian (K-Fold Cross Validation)

#### 1). 2-Fold Cross Validation

Pada 2-fold cross validation data akan dibagi menjadi 2 bagian langkah uji yang sama banyak sehingga masing-masing berjumlah 310 data. Pengujian 2-fold cross validation ini dibagi menjadi 2 langkah uji untuk menghitung nilai akurasi pada setiap langkah uji. Pada uji langkah 2 menghasilkan nilai akurasi yang tertinggi yaitu 90%, presisi 91% dan recall 97%.

Tabel 3. Hasil Pengujian 2 Fold

2-Fold	Multinomial naïve bayes		
	Akurasi	Presisi	Recall
Uji 1	85%	89%	93%
Uji 2	90%	91%	97%



Gambar 10. Hasil 2-Fold Cross Validation

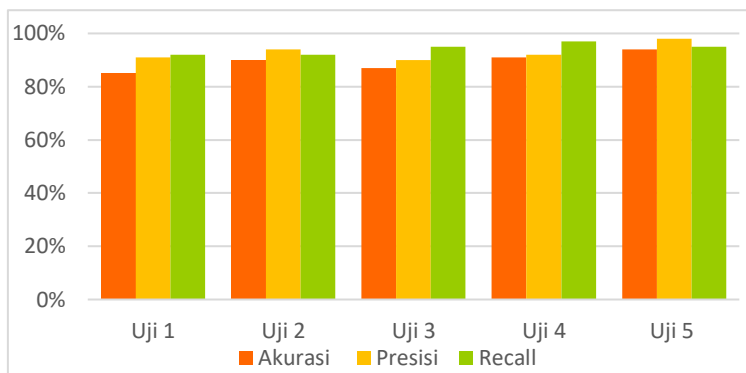
#### 2). 5-Fold Cross Validation

Pada 5-fold cross validation data akan dibagi 5 bagian langkah uji yang sama banyak sehingga masing-masing berjumlah 124 data. Pengujian 5-fold cross validation ini dibagi menjadi 5 langkah uji untuk menghitung nilai akurasi pada setiap langkah uji. Uji Langkah 5 menghasilkan nilai akurasi yang tertinggi yaitu 94%, presisi 98% dan recall 95%.

Tabel 4. Hasil Pengujian 5 Fold

5-Fold	Multinomial naïve bayes		
	Akurasi	Presisi	Recall
Uji 1	86%	91%	92%
Uji 2	90%	94%	92%
Uji 3	87%	90%	95%
Uji 4	91%	92%	97%
Uji 5	94%	98%	95%





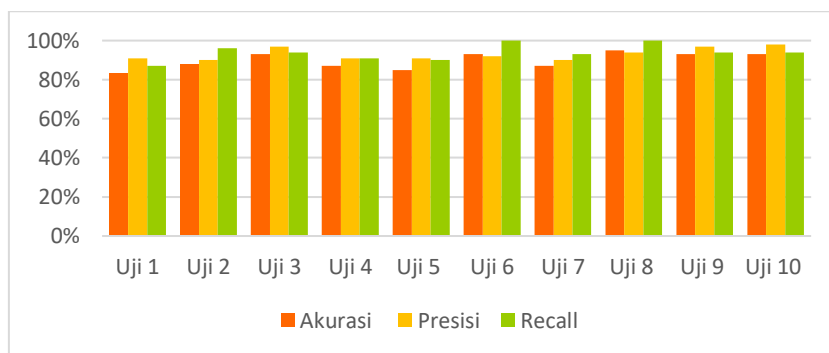
Gambar 11. Hasil 5-Fold Cross Validation

3). 10-Fold Cross Validation

Pada 10-fold cross validation ini data akan dibagi menjadi 10 bagian yang sama sehingga masing-masing berjumlah 62 data. Pengujian 10-fold cross validation ini dibagi menjadi 10 langkah uji untuk menghitung nilai akurasi pada setiap langkah uji. Uji Langkah 8 menghasilkan nilai akurasi yang tertinggi yaitu 95%, presisi 94% dan recall 100%.

Tabel 5. Hasil Pengujian 10 Fold

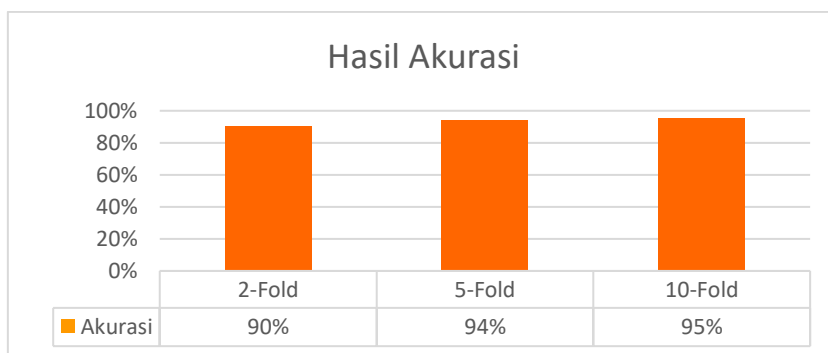
10-Fold	Multinomial naïve bayes		
	Akurasi	Presisi	Recall
Uji 1	83%	91%	87%
Uji 2	88%	90%	96%
Uji 3	93%	97%	94%
Uji 4	87%	91%	91%
Uji 5	85%	91%	90%
Uji 6	93%	92%	100%
Uji 7	87%	90%	93%
Uji 8	95%	94%	100%
Uji 9	93%	97%	94%
Uji 10	93%	98%	94%



Gambar 12. Hasil 10-Fold Cross Validation

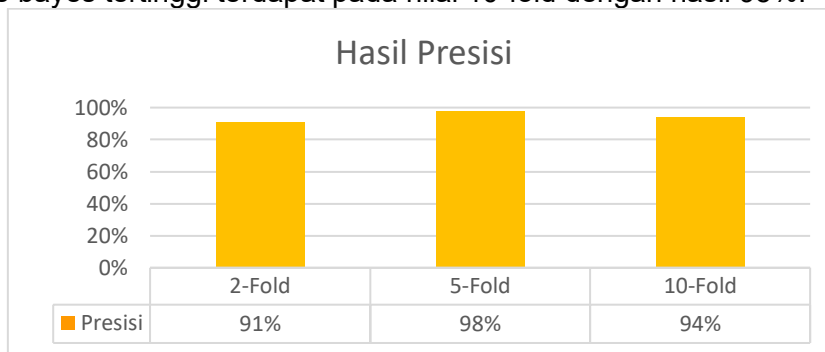
3.6. Hasil Analisis

Dari hasil pengujian k-fold cross validation 2, 5 dan 10 menggunakan multinomial naïve bayes diperoleh hasil nilai keseluruhan dari akurasi, presisi, dan recall.



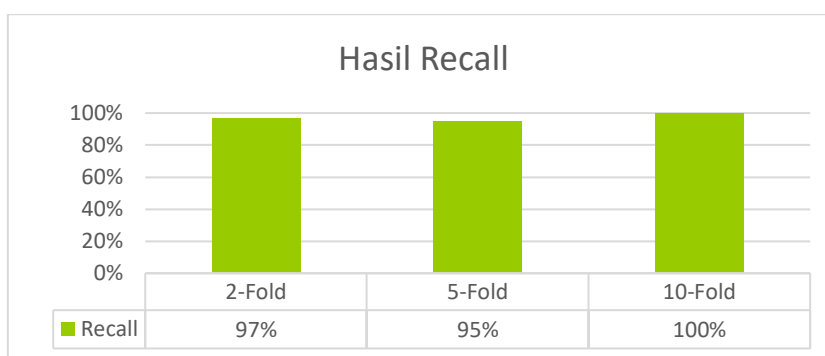
Gambar 13. Hasil Nilai Akurasi

Pada gambar diketahui bahwa perhitungan akurasi dengan k-fold cross validation menggunakan multinomial naïve bayes tertinggi terdapat pada nilai 10-fold dengan hasil 95%.



Gambar 14. Hasil Nilai Presisi

Diketahui pada gambar perhitungan presisi dengan k-fold cross validation menggunakan multinomial naïve bayes nilai presisi tertinggi terdapat pada nilai 10-fold dengan hasil 98%.



Gambar 15. Hasil Nilai Recall

Pada gambar perhitungan recall dengan k-fold cross validation menggunakan multinomial naïve bayes nilai recall tertinggi terdapat pada nilai 10-fold dengan hasil 100%.

#### 4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan dengan pengumpulan 800 data telah dilakukan pembersihan atau pemilahan pada data yang tidak relevan serta data duplikat sehingga data yang digunakan yaitu 620 data dengan 504 berupa sentimen positif dan 116 sentimen negatif. Pengujian k-fold cross validation 2, 5, dan 10 telah menghasilkan nilai tertinggi pada 10-fold langkah uji ke-8 yaitu dengan hasil nilai akurasi 95%, presisi 94% dan recall 100%. Berdasarkan metode multinomial naïve bayes, klasifikasi ini menghasilkan kinerja yang baik dan cocok digunakan dalam analisis data teks. Adapun saran untuk penelitian selanjutnya yaitu, menambahkan dataset yang lebih besar dari berbagai platform atau sumber.

#### Daftar Pustaka

- Aditama, M. I., Pratama, R. I., Wiwaha, K. H., & Rakhmawati, N. A. (2020). Analisis Klasifikasi Sentimen Pengguna Media Sosial Twitter Terhadap Pengadaan Vaksin COVID-19. *Journal Information Engineering and educational Technology* .
- Hidayat , R., Rahman, R. N., Perdana, M. R., & Arbansyah. (2024). Analisis Sentimen Aplikasi Identitas Kependudukan Digital (IKD) Menggunakan Metode Naive Bayes . *Jurnal Sistem Informasi dan Ilmu Komputer* , 129-140.
- Java, Muhammad Arya, and Mohammad Syafrullah. 2024. "Analisis Sentimen Ulasan Pengguna Aplikasi Threads Pada Google Play Store Menggunakan Multinomial Naive Bayes Dan Support Vector Machine." 12: 75–80.
- Liedfray, Tongkotow, Fonny J Waani, and Jouke J Lasut. 2022. "Peran Media Sosial Dalam Mempererat Interaksi Antar Keluarga Di Desa Esandom Kecamatan Tombatu Timur Kabupaten Tombatu Timur Kabupaten Minasa Tenggara." *Jurnal Ilmiah Society* 2(1): 2.

- Mailoa, Felix Fridom. 2021. "Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas Di Indonesia." *Journal of Information Systems for Public Health* 6(1): 44. doi:10.22146/jisph.44455.
- Mirlana, D. E., Sunarso, S., Halim, A., Putra, R., Widyaningsih, A. Z., & Khasanah, U. (2024). PEDAMPINGAN PENERBITAN IDENTITAS KEPENDUDUKAN DIGITAL. *Jurnal Abdimas FKIP UTP Surakarta*, 420-426.
- Sabrani, Alif, I W Gede Putu Wirarama Wedashwara, and Fitri Bimantoro. 2020. "Metode Multinomial Naive Bayes Untuk Klasifikasi Artikel Online Tentang Gempa Di Indonesia (Multinomial Naïve Bayes Method for Classification of Online Article About Earthquake in Indonesia)." *Jtika* 2(1): 91–92. <http://jtika.if.unram.ac.id/index.php/JTIKA/>.
- Santoso, Aloysius Kurniawan. 2022. "Analisis Sentimen Twitter Bahasa Indonesia Menggunakan Pendekatan Machine Learning." *Jurnal Informatika Kaputama (JIK)* 6(2): 129–36. doi:10.59697/jik.v6i2.111.
- Santoso, Heri, Armansyah Armansyah, and Dita Desliani. 2022. "Analisis Sentimen Mahasiswa Terkait Pembelajaran Tatap Muka Menggunakan Metode Naive Bayes Classifier." *Techno.Com* 21(3): 644–54. doi:10.33633/tc.v21i3.6262.
- Yona Sidratul Munti, Novi, and Dwi Asril Syaifuddin. 2020. "Analisa Dampak Perkembangan Teknologi Informasi Dan Komunikasi Dalam Bidang Pendidikan." *Jurnal Pendidikan Tambusai* 4(2): 1799–1805.
- Zhafira, Dhaifa Farah, Bayu Rahayudi, and Indriati Indriati. 2021. "Analisis Sentimen Kebijakan Kampus Merdeka Menggunakan Naive Bayes Dan Pembobotan TF-IDF Berdasarkan Komentar Pada Youtube." *Jurnal Sistem Informasi, Teknologi Informasi, dan Edukasi Sistem Informasi* 2(1): 55–63. doi:10.25126/justsi.v2i1.24.