JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER

p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 3 Februari 2026
Hal: 1539-1548
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i3.631

Klasifikasi Sentiment Analysis Terhadap Usulan KB Vasektomi Syarat Penerima BANSOS dengan Metode Naive Bayes

Rahma Lubis[1], Sita Anggraeni[2*]

[1,2]Fakultas Teknologi Informasi, Program Studi Informatika, Universitas Nusa Mandiri, Jakarta, Indonesia

Email: [1]rahmaa1306@gmail.com , [2*]sita.sia@nusamandiri.ac.id
(* Email Corresponding Author: sita.sia@nusamandiri.ac.id)
Received: 1 Oktober 2025 | Revision: 27 Oktober 2025 | Accepted: 27 Oktober 2025

Abstrak

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap usulan kebijakan vasektomi sebagai syarat penerima bantuan sosial (BANSOS) yang dikemukakan oleh Gubernur Jawa Barat. Data penelitian diperoleh dari platform media sosial *X* pada periode 29 April - 31 Juli 2025 sebanyak 848 *tweet*. Metode yang digunakan meliputi tahapan preprocessing data (*case folding, cleaning*, normalisasi, tokenisasi, *stopword removal, stemming*), pelabelan sentimen, pembobotan dengan TF-IDF, serta *feature selection* menggunakan uji *Chi-Square*. Untuk mengatasi ketidakseimbangan kelas, dilakukan oversampling dengan metode *SMOTE*. Algoritma yang digunakan dalam klasifikasi adalah *Multinomial Naïve Bayes*. Hasil penelitian menunjukkan bahwa mayoritas sentimen masyarakat bersifat positif dengan persentase 76,03%, sedangkan sentimen negatif sebesar 23,97%. Model *Naïve Bayes* menghasilkan akurasi sebesar 83%, presisi 0,83 untuk kelas negatif dan 0,85 untuk kelas positif, *recall* 0,85 untuk kelas negatif dan 0,82 untuk kelas positif, serta F1-score masing-masing 0,84 dan 0,83. Temuan ini membuktikan bahwa algoritma *Naïve Bayes* efektif digunakan untuk analisis sentimen terkait kebijakan publik di media sosial.

Kata Kunci: Analisis Sentimen, Bantuan Sosial, *Naïve Bayes*, Vasektomi, *X*

Abstract

This study aims to analyze public sentiment towards the proposed vasectomy policy as a requirement for receiving social assistance (BANSOS) proposed by the Governor of West Java. The research data was obtained from social media platform X during the period of April 29 - July 31, 2025, totaling 848 tweets. The methods used include data preprocessing stages (case folding, cleaning, normalization, tokenization, stopword removal, stemming), sentiment labeling, weighting with TF-IDF, and feature selection using the Chi-Square test. To overcome class imbalance, oversampling was carried out using the SMOTE method. The algorithm used in the classification is Multinomial Naïve Bayes. The results of the study show that the majority of public sentiment is positive, with a percentage of 76.03%, while negative sentiment is 23.97%. The Naïve Bayes model produced an accuracy of 83%, a precision of 0.83 for the negative class and 0.85 for the positive class, a recall of 0.85 for the negative class and 0.82 for the positive class, and an F1-score of 0.84 and 0.83, respectively. These findings demonstrate that the Naïve Bayes algorithm is effective for analyzing sentiment related to public policy on social media.

Keywords*: Sentiment Analysis, Social Assistance, Naïve Bayes, Vasectomy, X*

1. PENDAHULUAN

Indonesia merupakan negara kepulauan terpadat keempat di dunia. Pada tahun 2024, populasi Indonesia diperkirakan tumbuh hingga mencapai 281.603.900 jiwa (*Goodstats*, 2024). Berdasarkan data tersebut, Jawa Barat tercatat sebagai provinsi dengan jumlah penduduk terbesar, yakni 50.345.200 jiwa atau sekitar 17,88% dari total penduduk nasional. Hal ini menunjukkan bahwa sebagian besar penduduk Indonesia masih terpusat di Pulau Jawa, terutama di wilayah baratnya.

Dalam upaya pengendalian laju pertumbuhan penduduk yang cepat, pemerintah melalui Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN) telah melaksanakan berbagai program, antara lain peningkatan pelayanan keluarga berencana (KB) dan kesehatan reproduksi yang mudah diakses, bermutu, dan efektif. Program tersebut bertujuan untuk membentuk keluarga kecil yang berkualitas, sebagaimana diatur dalam Undang-Undang Nomor 52 Tahun 2014 tentang Kependudukan dan Pembangunan Keluarga (BKKBN, 2014). Alat kontrasepsi digunakan oleh Program Keluarga Berencana. Alat kontrasepsi sendiri terdiri dari berbagai jenis, termasuk *Intra Uterine Device* (IUD), metode operasi wanita (MOW), implan, suntik, pil, kondom, dan metode operasi pria (MOP).

Menurut data Badan Pusat Statistik (2022), mencatat 55,36% pasangan usia subur (PUS) di Indonesia sedang menggunakan alat kontrasepsi. Program ini dirancang bagi pasangan usia subur dengan tujuan utama menurunkan angka kelahiran melalui penggunaan alat kontrasepsi secara teratur, sehingga peran serta masyarakat sangat penting [1]. Berdasarkan data tersebut, menunjukkan bahwa angka penggunaan alat kontrasepsi oleh pria masih sangat rendah, yaitu kondom sebesar 2,06% dan angka metode operasi pria (MOP) atau vasektomi sebesar 0,24%. Vasektomi/MOP, merupakan salah satu metode kontrasepsi permanen [2], [3]. Angka ini menunjukkan bahwa kontribusi pria dalam menggunakan alat kontrasepsi masih sangat kecil, berbanding terbalik dengan penggunaan alat kontrasepsi pada wanita, dimana KB suntik merupakan metode paling banyak digunakan yakni, mencapai 56,1%.

Rendahnya angka partisipasi pria mendorong Gubernur Jawa Barat mengusulkan kebijakan kontroversial yaitu menjadikan vasektomi sebagai syarat penerima bantuan sosial (BANSOS). Usulan ini menimbulkan perdebatan publik karena menyangkut hak reproduksi

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 3 Februari 2026
Hal: 1539-1548
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i3.631

sekaligus akses terhadap bantuan sosial. Dalam konteks kebijakan publik, media sosial berperan penting sebagai ruang diskusi masyarakat. Pada titik inilah analisis sentimen menjadi relevan. Analisis sentimen merupakan adalah metode untuk memahami, mengekstraksi, dan mengelola data teks secara otomatis untuk mengumpulkan data informasi tentang persepsi seseorang melihat kalimat tertentu [4], [5]. Prinsip dasarnya adalah pengelompokan berdasarkan polaritas pada teks digunakan dalam analisis sentimen, yang kemudian dibagi menjadi kategori positif dan kategori negatif. Ini juga dapat dimanfaatkan untuk melihat pendapat terhadap sebuah masalah baru, atau juga dapat digunakan untuk menemukan kecendurungan dalam pembahasan suatu topik [6]. Ada banyak algoritma yang dapat melakukan analisis sentimen. Salah satu metode populer dalam analisis sentimen adalah *Naive Bayes Classifier*, algoritma berbasis *Teorema Bayes* yang efisien, membutuhkan sedikit data latih, dan terbukti efektif untuk klasifikasi teks [7]. Dan untuk menilai kinerja algoritma klasifikasi dan membandingkan hasil prediksi model dengan data sebenarnya digunakan *confusion matrix*. Selain itu, *confusion matrix* juga dapat menghitung tingkat akurasi model [8].

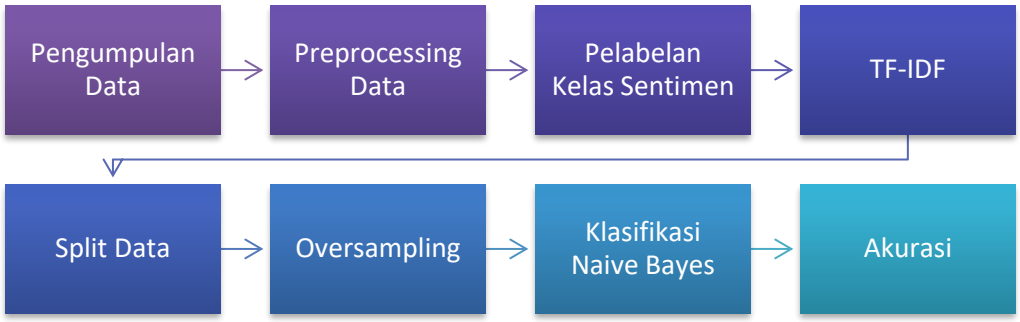
Beberapa penelitian terdahulu menunjukkan keberhasilan penggunaan *Naive Bayes* dalam berbagai isu sosial. Emil Salim & Achmad Solichin (2022) menggunakan *Naive Bayes* untuk menganalisis sentimen terhadap pelayanan Disdukcapil, dengan hasil 70,20% positif dan 29,80% negatif [9]. Fuad Amirullah, Syariful Alam, M. Imam Sulistyo (2023) menggunakan *Naive Bayes* untuk menganalisis sentimen masyarakat terhadap kinerja KPU menjelang pemilu 2024 berdasarkan opini *X* dengan hasil akurasi 74,5% [10]. Al Munawaroh, Reno Ridhoi, Rudiman (2024) menggunakan *Naive Bayes* untuk menganalisis sentimen masyarakat terhadap risiko pembangunan IKN berbasis Orange dengan hasil akurasi sebesar 87% [11]. Penelitian lain oleh Anastasya Nadia Puspitasari, Yulian Findawati, Yunianita Rahmawati (2024) terkait *e-commerce* menunjukkan akurasi 92% , presisi 90,35% dan *recall* 100% [12]. Riska Kurnia Septiani, Sita Anggraeni, Sandra Dewi Saraswati (2021) juga menggunakan Naive Bayes untuk klasifikasi sentimen IKN dengan akurasi hingga 100% pada split data [13].

Berdasarkan uraian yang tertera di atas maka peneliti tertarik untuk melakukan penelitian yang berjudul "Analisis Sentimen Usulan KB Vasektomi Syarat Penerima Bantuan Sosial (BANSOS) Jawa Barat Dengan *Naive Bayes*", judul ini dipilih sangat penting untuk mengetahui bagaimana masyarakat melihat kebijakan publik yang kontroversial, terutama yang berkaitan dengan hak reproduksi dan akses ke bantuan sosial. Penelitian ini bertujuan untuk mengetahui respons masyarakat menggunakan analisis sentimen di sosial media *X*. Dengan demikian, hasil penelitian tidak dapat digeneralisasikan pada media sosial atau platform lain karena adanya kemungkinan perbedaan karakteristik data dan pengguna. Algoritma *Naive Bayes* dipilih karena cukup akurat untuk mengklasifikasikan teks.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Pada tahap ini penulis menjabarkan tahapan penelitian yang akan dilalui. Tahapan penelitian yang dilakukan yaitu diilustrasikan pada Gambar 1.



Gambar 1. Tahapan Penelitian

a. *Crawling* atau Pengumpulan Dataset

Data dalam penelitian ini diambil dengan bantuan seorang kreator *YouTube* Bernama "*Helmi Satria*". Ia menyediakan akses ke *Google Colab* berisi kode untuk mengambil data dari media sosial *X*. Proses pengumpulan data diawali dengan mengakses tautan *Google Colab* berikut: https://colab.research.google.com/drive/1RyMMMOl4e3*X*ni*X*_fILfIoqnytbuB-4m. Setelah membuka *Google Colab*, langkah pertama yang perlu dilakukan adalah mengganti *X Auth Token* dengan token milik pengguna dari akun *X*. Jika *X Auth Token* yang sudah diganti dapat digunakan, maka proses pengumpulan data dapat dimulai. Selanjutnya, pengguna dapat memasukkan nama file (*filename)* yang akan digunakan untuk menyimpan hasil *crawling* data dalam format excel (.xls*x*) setelah data dapat dikumpulkan.

b. Pre-processing

Tahap *pre–processing* adalah proses mengolah data mentah menjadi kumpulan set data siap pakai. Ini dilakukan untuk memilih data mana dari dokumen yang akan diproses. Adapun beberapa tahapan yang dilakukan untuk memaksimalkan hasil dari *pre-processing,* diantaranya yaitu [14].

1) *Cleansing* adalah proses pembersihan data *tweet* dari kata-kata yang tidak memberikan penjelasan apa pun.
2) *Case Folding* merupakan tahapan proses teks dimana semua teks diubah ke dalam *lower case* atau huruf kecil. Dalam proses ini hanya menerima huruf 'a' sampai 'z', selain huruf tersebut akan dihilangkan.
3) Normalisasi adalah proses mengubah kalimat menjadi kalimat normal, seperti mengubah kata-kata yang disingkat menjadi kata yang seharusnya dan mengubah kata-kata yang tidak baku menjadi kata-kata baku.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 3 Februari 2026
Hal: 1539-1548
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i3.631

4) *Tokenizing* atau proses tokenisasi adalah proses data, di mana kalimat dibagi menjadi kata perkata.
5) *Stopword Removal* adalah proses meninjau setiap kata pada teks untuk membuang kata-kata yang dianggap tidak penting.
6) *Stemming* adalah proses mengubah kata berimbuhan ke bentuk dasarnya (*root word*) dengan menggunakan *library* Sastrawi.

c. Pelabelan Kelas Sentimen

Pelabelan data pada analisis sentiment adalah proses memberikan label atau kategori pada setiap data teks untuk mengindikasikan sentiment atau perasaan yang terkandung dalam teks tersebut. Ini adalah langkah penting dalam melatih model analisis sentiment untuk memahami dan mengklasifikasikan apakah suatu teks mengandung sentiment positif, negatif, atau netral [15].

d. Pembobotan TF-IDF

Mengubah teks menjadi representasi numerik menggunakan teknik TF-IDF (*Term Frequency – Inverse Document Frequency*) dan melakukan *feature selection* (seleksi fitur) dengan metode uji *Chi-Square.*

e. Split Data

Membagi dataset menjadi dua bagian yaitu traning data dan testing data.

f. Oversampling Data Menggunakan SMOTE

SMOTE (Synthetic Minority Over-sampling Technique) adalah teknik untuk mengatasi masalah ketidakseimbangan kelas dalam data. SMOTE menghasilkan sampel sintetis baru dari kelas minoritas untuk mencapai keseimbangan dengan kelas mayoritas

g. Pembentukan Probabilitas (Training *Naïve Bayes*)

Pada fase ini, setiap *tweet* data training atau data testing yang telah dikumpulkan akan diubah menjadi sekumpulan kata yang mempunyai nilai probabilitas terhadap teks kelas yang ditetapkan.

h. Klasifikasi

Pada fase ini, *tweet* akan dikategorikan ke dalam sentimen positif dan sentimen negatif dengan algoritma klasifikasi yang dipakai yaitu *Naïve Bayes Classifier*.

i. Akurasi

Setelah klasifikasi selesai, kualitas model yang dibuat akan diuji dengan menghitung nilai akurasi.

2. HASIL DAN PEMBAHASAN

3.1 Crawling dan Pengumpulan Data

Pengumpulan data diambil dari platform media sosial *X* dengan *query* vasektomi bansos sebagai topik yang akan diteliti pada periode unggahan 29 April hingga 31 Juli 2025. Data yang digunakan pada penelitian ini sebanyak 848 data *tweet* dengan pencarian vasektomi bansos seperti pada Gambar 2. Cara yang dilakukan untuk pengambilan data dengan bantuan seorang Youtuber bernama "Helmi Satria'' dengan masuk ke link *google colab* yang sudah diberikan.

Gambar 2. Pengisian *X Auth* Token untuk pengambilan data

Dengan mengisi X Auth Token, *Google Colab* dapat digunakan untuk pengambilan data melalui akun media sosial *X* pengguna. Setelah proses autentikasi berhasil dan pencarian data dapat dilakukan. Proses pengumpulan data diawali dengan mengisi nama file atau *filename* yang akan digunakan untuk menyimpan hasil *crawling* atau pengumpulan data pada Gambar 3, yaitu dengan memasukkan *filename* "data_crawling.csv". Selanjutnya, penulis memasukkan kata kunci atau *search keyword* yang akan dicari yaitu, "vasektomi bansos", serta menentukan rentang tanggal pencarian yang dimulai dari tanggal 29 April hingga 31 Juli 2025 dengan pencarian hanya dilakukan pada unggahan berbahasa Indonesia. Setelah itu, menetapkan limit atau batas data yang akan diambil, yaitu sebanyak 800 data *tweet*.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 3 Februari 2026
Hal: 1539-1548
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i3.631

	conversation_id_str	created_at	favorite_count	full_text	id_str	image_url	in_reply_to_screen_name	lang	location	quote_cour
0	1941105276322816018	Fri Jul 04 23:24:24 +0000 2025	0	@cholilnafis Fakta di lapangan vasektomi itu u...	1941276939215569384	NaN	cholilnafis	in	NaN	
1	1941105276322816018	Fri Jul 04 15:35:59 +0000 2025	0	@cholilnafis Oh si baleguk anti islam itu ya y...	1941159060688245077	NaN	cholilnafis	in	NaN	
2	1941105276322816018	Fri Jul 04 12:02:16 +0000 2025	605	Masih mau menyaratkan vasektomi utk mandapat b...	1941105276322816018	https://pbs.twimg.com/media/GvAw881W8AA0-jD.jpg	NaN	in	NaN	
3	1939544829194915848	Wed Jul 02 14:43:30 +0000 2025	0	@MariaAlkaff_ Kang Dedi bagaimana: - Kasus pen...	1940421075344687144	NaN	MariaAlkaff_	in	NaN	
4	1940259136073068937	Wed Jul 02 10:37:34 +0000 2025	8	@ARSIPAJA Tp vasektomi dilarang dalam islam.. ...	1940359187072245915	NaN	Radenilham40726	in	NaN	
...	...	...	...	...	...	...	...	...	...	...
652	1919000135989661935	Sun May 04 12:08:22 +0000 2025	0	Dan kek HEH initu VASEKTOMI bukan KEBIRI lagia...	1919001150088716322	NaN	jenangcandill	in	NaN	
653	1919000135989661935	Sun May 04 12:04:20 +0000 2025	0	Sedang membaca huru hara MUI dan PBNU haramkan	1919000135989661935	NaN	NaN	in	NaN	

Gambar 3. Hasil *Crawling* atau Pengumpulan Data

Hasil *crawling* atau pengumpulan data menghasilkan sebanyak 848 data *tweet*, namun data masih mengandung beberapa variabel yang tidak diperlukan seperti, *conversation_id_str, created_id, favorite_count, id_str, image_url, in_reply_to_screen_name, lang, location, quote_count, reply_count, retweet_count, tweet_url, user_id_str, dan username*. Oleh karena itu, diperlukan tahapan *pre-processing* agar mendapatkan teks komentar yang diinginkan tanpa *noise*.

3.2 Tahapan Preprocessing

Pada tahapan ini data akan melewati beberapa proses, yaitu proses *case folding, cleaning, tokenization, stopwords, stemming,* dan normalisasi kata.

a. Case Folding

Case folding adalah proses mengubah semua karakter huruf dalam data menjadi huruf kecil atau *lower case*. Proses ini dilakukan agar kata yang sama tetapi ditulis dengan huruf yang berbeda tetap terbaca sebagai arti yang sama. Penerapan *case folding* menggunakan fungsi *lower()* dalam *Python* seperti pada Tabel 1. Setelah semua proses *case folding* selesai dilakukan, data *tweet* akan mengalami perubahan. Berikut perbandingan data *tweet* sebelum dan sesudah proses *case folding.*

Tabel 1. Sebelum dan Sesudah Case Folding

full_text	case_folding
Dua program utama Dedi Mulyadi yang menyita perhatian media asing adalah terkait pengiriman siswa ke barak militer dan wacana vasektomi sebagai syarat warga miskin terima bansos. ~MD #DediMulyadi #Vasektomi #BarakMiliter https://t.co/H6yv1KduVl	dua program utama dedi mulyadi yang menyita perhatian media asing adalah terkait pengiriman siswa ke barak militer dan wacana vasektomi sebagai syarat warga miskin terima bansos. ~md #dedimulyadi #vasektomi #barakmiliter https://t.co/h6yv1kduvl

b. Cleaning

Proses *cleaning* adalah proses pembersihan, proses ini bertujuan untuk mengurangi *noise* pada *dataset*. Karakter yang dihilangkan pada proses ini *RT, Link/URL* yang terletak baik di depan, tengah, maupun akhir, *hastag* (#), *mention* atau *user*, angka, dan simbol – simbol lainnya. Berikut adalah tahapan dari cleaning data tampak pada Tabel 2. Setelah atribut yang tidak diperlukan pada proses cleaning sudah dihilangkan. Ini akan mempermudah klasifikasi dan pemrosesan data sebelum masuk ke tahapan analisa. Berikut perbandingan data *tweet* sebelum dan sesudah proses cleaning.

Tabel 2. Hasil Cleaning Pada Case Folding

case_folding	cleaning
dua program utama dedi mulyadi yang menyita perhatian media asing adalah terkait pengiriman siswa ke barak militer dan wacana vasektomi sebagai syarat warga miskin terima bansos. ~md #dedimulyadi #vasektomi #barakmiliter https://t.co/h6yv1kduvl	dua program utama dedi mulyadi yang menyita perhatian media asing adalah terkait pengiriman siswa ke barak militer dan wacana vasektomi sebagai syarat warga miskin terima bansos md dedimulyadi vasektomi barakmiliter

c. Tokenization

Tokenization adalah sebuah proses pemotongan data yang sebelumnya kalimat kemudian dipecah menjadi potongan kata tunggal. Hasil deduksi ini disebut token. Setelah semua proses *tokenize* selesai dilakukan, data *tweet* akan mengalami perubahan. Berikut perbandingan data *tweet* sebelum dan sesudah proses *tokenize* sesuai pada tampilan Tabel 3.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 3 Februari 2026
Hal: 1539-1548
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i3.631

Tabel 3. Hasil Tokenize Pada Cleaning

Cleaning	tokenize
dua program utama dedi mulyadi yang menyita perhatian media asing adalah terkait pengiriman siswa ke barak militer dan wacana vasektomi sebagai syarat warga miskin terima bansos md dedimulyadi vasektomi barakmiliter	['dua', 'program', 'utama', 'dedi', 'mulyadi', 'yang', 'menyita', 'perhatian', 'media', 'asing', 'adalah', 'terkait', 'pengiriman', 'siswa', 'ke', 'barak', 'militer', 'dan', 'wacana', 'vasektomi', 'sebagai', 'syarat', 'warga', 'miskin', 'terima', 'bansos', 'md', 'dedimulyadi', 'vasektomi', 'barakmiliter']

d. Normalisasi Kata

Pada tahap ini, normalisasi kata dilakukan untuk mengubah kata-kata yang tidak baku, singkatan, atau penulisan yang tidak sesuai menjadi bentuk kata baku atau standar yang sesuai dengan kamus yang telah ditentukan. Proses diawali dengan mengimpor file yang akan digunakan sebagai kamus pembantu dalam proses normalisasi kata, yaitu file bernama "kamuskatabaku.xlsx". Setelah file berhasil diimpor, isi dari file tersebut dibaca dan dikonversi menjadi sebuah kamus dengan pasangan, kata tidak baku sebagai kunci dan kata baku sebagai nilai. Setelah itu, setiap kata dalam data yang telah melalui tahap tokenisasi dicocokkan dengan kamus tersebut, sehingga kata tidak baku akan digantikan dengan kata baku yang sesuai. Berikut perbandingan data *tweet* sebelum dan sesudah normalisasi kata seperti Tabel 4 dan perbaikan singkatan pada Tabel 5.

Tabel 4. Contoh Bahasa Gaul dan Perbaikannya

Bahasa Gaul	Perbaikan
gue, gw, gua	aku
lu, loe, lo	kamu

Tabel 5. Contoh Kata Singkatan dan Perbaikannya

Singkatan	Perbaikan
g, ga, gk, tdk	tidak
kl, klo, klu, kalo	kalau
mksd, mksud	maksud

e. Stopword Removal

Proses *stopword removal* bertujuan meyederhanakan data dan meningkatkan akurasi analisis sentiment dengan menghapus kata-kata yang dianggap tidak penting, seperti kata penghubung, kata ganti, atau kata yang tidak ada kaitannya dengan analisis sentiment akan dihapus. Proses *stopword removal* dimulai dengan menggunakan *library* NLTK (*Natural Language Toolkit*) sesuai dengan Tabel 6 untuk mengunduh daftar *stopwords* Bahasa Indonesia. Selanjutnya, fungsi *remove_stopwords()* dijalankan untuk menghilangkan kata-kata yang ada dalam daftar *stopwords* dari setiap token yang dibuat selama tahap tokenisasi. Berikut perbandingan data *tweet* sebelum dan sesudah *stopword removal.*

Tabel 6. Hasil *Stopword Removal* Pada Normalisasi

Normalisasi	stopword removal
['dua', 'program', 'utama', 'dedi', 'mulyadi', 'yang', 'menyita', 'perhatian', 'media', 'asing', 'adalah', 'terkait', 'pengiriman', 'siswa', 'ke', 'barak', 'militer', 'dan', 'wacana', 'vasektomi', 'sebagai', 'syarat', 'warga', 'miskin', 'terima', 'bansos', 'md', 'dedimulyadi', 'vasektomi', 'barakmiliter']	['program', 'utama', 'dedi', 'mulyadi', 'menyita', 'perhatian', 'media', 'asing', 'terkait', 'pengiriman', 'siswa', 'barak', 'militer', 'wacana', 'vasektomi', 'syarat', 'warga', 'miskin', 'terima', 'bantuansosial', 'md', 'dedimulyadi', 'vasektomi', 'barakmiliter']

f. Stemming Data

Tahap terakhir pada tahapan *preprocessing* adalah *stemming*. *Stemming* dilakukan untuk mengubah kata menjadi bentuk dasarnya. Pada langkah ini, dimulai dengan mengunduh dan membaca daftar kata baku dari file kata_kbb.tXt sebagai referensi. Selanjutnya, dilakukan instalasi stemmer dari Sastrawi. Fungsi stemming seperti Tabel 7 dibuat dengan cara menerapkan metode stem() dari Sastrawi pada setiap token yang telah melalui proses stopword removal sebelumnya. Berikut perbandingan data tweet sebelum dan sesudah stemming data.

Tabel 7. Hasil *Stemming* Data Pada *Stopword Removal*

Stopword removal	stemming
['program', 'utama', 'dedi', 'mulyadi', 'menyita', 'perhatian', 'media', 'asing', 'terkait', 'pengiriman', 'siswa', 'barak', 'militer', 'wacana', 'vasektomi', 'syarat', 'warga', 'miskin', 'terima', 'bantuansosial', 'md', 'dedimulyadi', 'vasektomi', 'barakmiliter']	['program', 'utama', 'dedi', 'mulyadi', 'sita', 'perhati', 'media', 'asing', 'kait', 'kirim', 'siswa', 'barak', 'militer', 'wacana', 'vasektomi', 'syarat', 'warga', 'miskin', 'terima', 'bantuansosial', 'md', 'dedimulyadi', 'vasektomi', 'barakmiliter']

3.3 Pelabelan Kelas Sentimen

Setelah proses *preprocessing* selesai dilakukan, tahapan selanjutnya adalah pelabelan kelas sentiment. Pada tahap ini, hasil pelabelan akan menjadi dasar dalam menentukan analisis penelitian sesuai dengan representasi data yang dihasilkan. Pelabelan sentiment dilakukan menggunakan *library TeXtBlob* dengan hasil pada Tabel 8. Proses dimulai dengan menerjemahkan data *tweet* ke dalam Bahasa Inggris, kemudian dilakukan proses klasifikasi sentiment. Sentiment yang dihasilkan terbagi ke dalam dua kelas, yaitu sentiment positif dan sentiment negatif. Langkah diawali dengan penginstallan *library TeXtBlob* dan *deep-translator* yang akan digunakan dalam proses *translate* dan *labeling* data. Tahap selanjutnya yaitu proses translate. Setelah data tweet berhasil di *translate*, pelabelan sentimen diproses

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 3 Februari 2026
Hal: 1539-1548
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i3.631

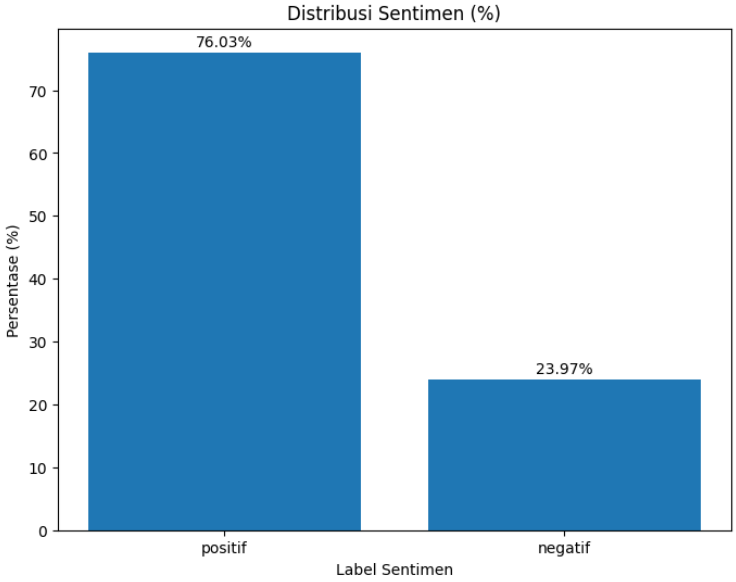
secara otomatis dengan melakukan skor pada masing-masing kata yang sudah diterjemahkan sebelumnya. Jika skor < 0.0 maka dikategorikan sebagai sentimen negatif selain skor itu maka dikategorikan sentimen negatif.

Tabel 8. Hasil Pelabelan Kelas Sentimen

Stemming	*translate*	label sentimen
['gubernur', 'jawa', 'barat', 'dedi', 'mulyadi', 'mengklarifikasi', 'nyata', 'kait', 'bijak', 'vasektomi', 'bijak', 'mensyaratkan', 'vasektomi', 'warga', 'terima', 'bantu', 'sosial', 'bantuansosial', 'bijak', 'vasektomi', 'bijak']	*['Governor', 'Javanese', 'Western', 'Dedi', 'MulyadiMengk clarification', 'real', 'hook', 'wise', 'vasectomy', 'wise', 'require vasectomy', 'citizens', 'accept', 'help', 'social', 'social assistance', 'wise', 'vasectomy', 'wise']*	positif
['beda', 'boleh', 'paksa', 'sulit', 'miskin', 'butuh', 'bantuansosial', 'hidup', 'kasih', 'syarat', 'kuasa', 'vasektomi', 'baca', 'uu', 'bahas', 'paksa', 'kontrasepsi', 'penyalahgunaan', 'kuasa', 'manfaat', 'kondisi', 'daya', 'pidana']	*['programs', 'main', 'Dedi', 'Mulyadi', 'Sita', 'Pay attention', 'media', 'foreign', 'hook', 'send', 'students', 'barracks', 'military', 'discourse', 'vasectomy', 'requirements', 'residents', 'poor', 'accept', 'aids social', 'MD', 'Dedimulyadi', 'Vasectomy', 'Barakmiliter']*	negatif

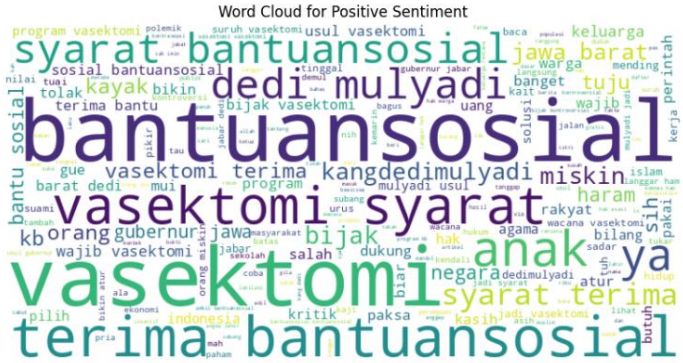
a. Visualisasi

Pada tahap ini peneliti melakukan visualisasi data yaitu proses representasi data dalam bentuk grafik atau gambar untuk menyampaikan informasi secara visual. Visualisasi membantu dalam memahami data yang lebih baik dengan menyoroti pola, tren, dan anomali yang mungkin tidak terlihat dalam bentuk table atau teks. Proses visualisasi data ini melibatkan pembuatan bar chart dan word cloud untuk memberikan gambaran visual sesuai Gambar 4 mengenai distribusi sentiment dan kata-kata yang sering muncul dalam dataset.



Gambar 4. Barchart Sentiment

Dapat kita lihat pada Gambar 5 bahwa dari sebanyak 848 data tweet yang ada tercipta persentase sentiment positif sebesar 76,03% dan sentiment negatif sebesar 23,97%.



Gambar 5. *Word Cloud* Sentimen Positif

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 3 Februari 2026
Hal: 1539-1548
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i3.631

Gambar 6. *Word Cloud* Sentimen Negatif

Ketika kata-kata disebutkan pada *wordcloud* lebih sering sesuai tampilan Gambar 6, *font*nya akan lebih besar. Sebaliknya, jika kata tersebut jarang disebutkan, *font*nya akan lebih kecil. Dalam gambar berikut, *Wordcloud* menampilkan kata-kata yang paling banyak digunakan dalam *tweet* dengan label negatif dan positif.

3.4 Pembobotan TF-IDF

TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah teknik yang digunakan untuk mengubah teks menjadi fitur numerik yang dapat digunakan oleh algoritma *machine learning*. TF-IDF pada tampilan Gambar 7 mengukur seberapa penting suatu kata dalam sebuah dokumen relatif terhadap *corpus*. Pada penelitian ini menggunakan *TfidfVectorizer* dari *sklearn* untuk mentransformasi data teks menjadi TF-IDF, dan melakukan seleksi fitur menggunakan *Chi-Square* untuk memilih 1000 kata yang paling relevan terhadap kelas sentiment yang bertujuan mengurangi *noise*, mempercepat komputasi, dan meningkatkan performa klasifikasi.

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.feature_selection import chi2, SelectKBest

vectorizer = TfidfVectorizer()
X_tfidf = vectorizer.fit_transform(data['stemming'].apply(lambda x: ' '.join(x)))
selector = SelectKBest(k=1000)
X_tfidf = selector.fit_transform(X_tfidf, data['label'])
```

Gambar 7. *Source Code TF-IDF*

3.5 Split Data

Split data adalah tahapan validasi yang melakukan validasi secara sederhana dengan membagi dataset secara acak menjadi dua data terpisah yaitu data *training* atau data latih dan data *testing* atau data uji sesuai Gambar 8. Pada validasi ini, dataset dibagi dengan rasio 70% untuk data *training* dan 30% untuk data *testing* dengan nilai *random state* sebesar 42 dan menggunakan *library Sklearn* untuk pembagian data. Pada proses ini perwakilan data akan dilabeli dengan variabel X dan Y. Hasil dari pembagian data didapati 592 data *training* dan 255 data *testing*, yang bisa dilihat pada gambar. 8

```
from sklearn.model_selection import train_test_split

# Join the stemmed tokens back into strings
X = data['stemming'].apply(lambda x: ' '.join(x))
y = data['label']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

print("Jumlah data training:", len(X_train))
print("Jumlah data testing:", len(X_test))
```

```
Jumlah data training: 592
Jumlah data testing: 255
```

Gambar 8. *Source Code Spliting* Data

3.6 Oversampling SMOTE

Dalam dataset yang diberi label kata kunci, ditemukan ketidakseimbangan kelas yang signifikan. Oleh karena itu, peneliti menggunakan metode *oversampling* yang disebut SMOTE (*Synthetic Minority Oversampling Technique*) untuk menyeimbangkan kembali data tersebut dalam Tabel 9. Ketidakseimbangan kelas terjadi ketika jumlah sampel dari satu kelas (kelas minoritas) secara signifikan lebih rendah daripada jumlah sampel dari kelas lainnya (kelas mayoritas). Hal ini dapat membuat model pembelajaran mesin condong ke arah kelas mayoritas. SMOTE mengatasi masalah ini dengan membuat sampel sintetis dari kelas minoritas, sehingga menyeimbangkan jumlah sampel dari kedua kelas. Metode ini tidak hanya mereplikasi data yang ada tetapi juga menghasilkan data baru yang sangat mirip dengan data asli, sehingga memungkinkan model untuk belajar lebih efektif dari kedua kelas.

Tabel 9. Perbandingan Sebelum dan Sesudah Oversampling

Sentimen	Sebelum Oversampling	Sesudah Oversampling
Negatif	203	644
Positif	644	644

3.7 Naïve Bayes

Algoritma *Multinomial Naïve bayes* digunakan untuk pemodelan klasifikasi pada penelitian ini dengan menggunakan *library sklearn* pada *python* untuk menerapkannya melalui *import MultinomialNB*. Algoritma *multinomial Naïve Bayes* digunakan untuk menangani data berbasis teks, seperti ulasan atau dokumen teks dan berfungsi untuk mengkategorikan kata-kata atau teks ke dalam berbagai kategori atau sentiment seperti positif dan negatif. Selanjutnya, evaluasi performa klasifikasi *Naïve Bayes* dalam Gambar 9 dilakukan dengan menghitung nilai akurasi, presisi, *recall*, dan *F1-score* menggunakan *Confusion Matrix*.

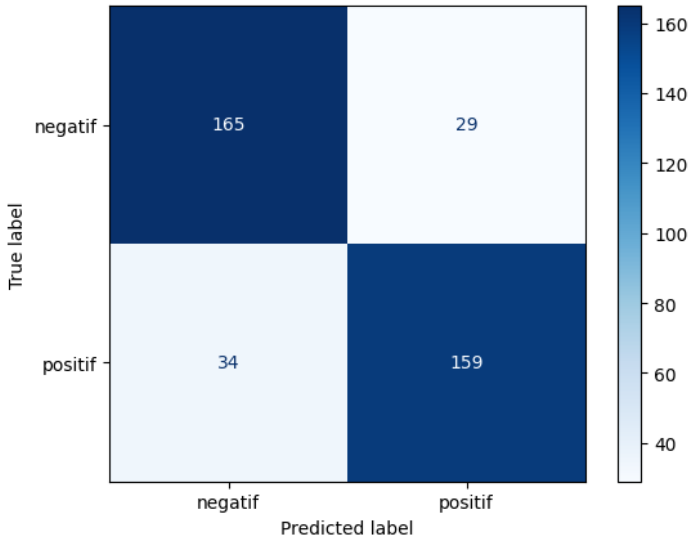
JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 3 Februari 2026
Hal: 1539-1548
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i3.631

```
Akurasi Naive Bayes: 0.8372093023255814
              precision    recall  f1-score   support

     negatif       0.83      0.85      0.84       194
     positif       0.85      0.82      0.83       193

    accuracy                           0.84       387
   macro avg       0.84      0.84      0.84       387
weighted avg       0.84      0.84      0.84       387
```

Gambar 9. Hasil Akurasi, Presisi, *Recall, F1-score Naïve Bayes*



Gambar 10. *Confusion Matrix Naïve Bayes*

Pada Gambar. 10 menampilkan hasil dari *Confusion Matrix* pada metode *Naïve Bayes*. Berikut penjabaran manual dalam perhitungan gambar 9 berdasarkan *confusion matrix* dari gambar, perlu identifikasi jumlah prediksi benar dan salah untuk setiap kelas, kemudian menghitung akurasi, presisi, *recall*, dan *F1-score*.

A. Negatif
   a. Prediksi Negatif = 165
   b. Prediksi Positif = 29

B. Positif
   a. Prediksi Negatif = 34
   b. Prediksi Positif = 159

Perhitungan Manual:

1. $Akurasi = \frac{TP+TN}{TP+TN+FP+FN} = \frac{159+165}{159+165+29+34} = \frac{324}{387} = 0{,}8302$ .....[1]
2. Kelas Negatif
   a. Presisi Negatif
      $Presisi = \frac{TP}{TP+FP} = \frac{165}{165+34} = \frac{165}{193} = 0{,}83$ .....[2]
   b. Recall Negatif
      $Recall = \frac{TP}{TP+FN} = \frac{165}{165+29} = \frac{165}{194} = 0{,}85$ .....[3]
   c. F1-score Negatif
      $F1 = 2X\frac{presisi\ x\ recall}{presisi+recall} = 2x\frac{0.83\ x\ 0.85}{0.83+0.85} = 0.84$ .....[4]
3. Kelas Positif
   a. Presisi Positif
      $presisi = \frac{TP}{TP+FP} = \frac{159}{159+29} = \frac{159}{188} = 0{,}85$ .....[5]
   b. Recall Positif
      $Recall = \frac{TP}{TP+FN} = \frac{159}{159+34} = \frac{159}{193} = 0{,}82$ .....[6]
   c. F1-score Positif

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 3 Februari 2026
Hal: 1539-1548
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i3.631

$$F1 = 2X\frac{presisi\ x\ recall}{presisi+recall} = 2x\frac{0.85\ x\ 0.82}{0.85+0.82} = 0.83 \quad \text{.....[7]}$$

3.8 Evaluasi Model

Berdasarkan hasil pengujian menggunakan algoritma *Multinomial Naïve Bayes*, diperoleh akurasi sebesar 83%, yang artinya model mampu menunjukkan kemampuan cukup baik dalam melakukan klasifikasi sentiment. Dari hasil pengujian, diperoleh nilai presisi, recall, dan F1-score yang seimbang. Pada kelas negatif, model menghasilkan presisi 0,83, recall 0,85 dan F1-sore 0,84. Hal ini berarti sebagian besar data negatif dapat dikenali dengan baik, meskipun masih ada beberapa yang salah terklasifikasi sebagai positif. Pada kelas positif, model memperoleh presisi 0,85, recall 0,82, dan F1-score 0,83 yang menunjukkan model juga cukup baik dalam mengenali data dengan sentimen positif.

Hasil *confusion matrix* memperlihatkan bahwa terdapat 165 data negatif dan 159 data positif yang berhasil diprediksi dengan benar. Namun, terdapat kesalahan klasifikasi berupa 29 data negatif yang salah diprediksi sebagai positif (*False Positive*) serta 34 data positif yang salah diprediksi sebagai negatif (*False Negative*).

Secara keseluruhan, penerapan tahapan preprocessing, TF-IDF, serta feature selection berhasil membantu model menghasilkan performa yang stabil. Dengan nilai evaluasi yang relatif seimbang, *Naïve Bayes* dapat dinyatakan layak digunakan untuk analisis sentimen pada penelitian ini.

4. KESIMPULAN

Berdasarkan hasil pembahasan, penelitian ini berhasil menganalisis sentimen masyarakat yang disampaikan melalui platform *X* terhadap usulan kebijakan syarat penerima bantuan sosial (BANSOS). Dari data yang dikumpulkan dimulai pada 29 April hingga 31 Juli 2025 terkumpul sebanyak 848 data *tweet*, setelah melalui berbagai tahapan. Ditemukan bahwa mayoritas sentimen cenderung ke positif dengan persentase 76,03% sedangkan untuk sentimen negatif sebesar 23,97%. Penelitian ini juga membuktikan bahwa algoritma *Naive Bayes* cukup efektif dalam mengklasifikasikan sentimen data teks dari media sosial. Setelah melalui tahapan *preprocessing* data dan penyeimbangan kelas menggunakan *SMOTE*, model klasifikasi menghasilkan performa yang baik, dengan nilai untuk akurasi yang diperoleh mencapai 83%, dengan presisi untuk kelas negatif sebesar 0,83 dan kelas positif 0,85. Selain itu, *recall* untuk kelas negatif sebesar 0,85 dan kelas positif 0,82, sedangkan nilai *F1-score* masing-masing adalah 0,84 untuk kelas negatif dan 0,83 untuk kelas positif.

REFERENCES

[1] S. Yusran and Z. Larisu, “Jurnal Kendari Kesehatan Masyarakat (JKKM) Vol . 1 No . 2 Tahun 2022 Pengaruh Efektivitas Penyuluhan dan Pelayanan KB Gratis terhadap Penerimaan Kontrasepsi Vasektomi pada PUS di Kota Kendari Jurnal Kendari Kesehatan Masyarakat (JKKM) Vol . 1 No . 2 T,” vol. 1, no. 2, pp. 48–58, 2022.

[2] D. Rahmawati and R. E. Ariningtyas, “Faktor Yang Mempengaruhi Kesediaan Suami Sebagai Akseptor Vasektomi,” vol. 15, no. 1, pp. 26–30, 2025.

[3] Z. A. Kahfilani, M. Umar, A. Ghazali, and D. Muntazhar, “Penggunaan Kontrasepsi Vasektomi : Kesehatan , Agama , dan Keharmonisan Rumah Tangga,” vol. 1, no. 2.

[4] T. C. Herdiyani and A. U. Zailani, “Sentiment Analysis Terkait Pemindahan Ibu Kota Indonesia Menggunakan Metode Random Forest Berdasarkan Tweet Warga Negara Indonesia,” *J. Teknol. Sist. Inf.*, vol. 3, no. 2, pp. 154–165, 2022, doi: 10.35957/jtsi.v3i2.2920.

[5] F. V. Sari and A. Wibowo, “Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi,” *J. SIMETRIS*, vol. 10, no. 2, pp. 681–686, 2020.

[6] S. K. M.Kom, Ahmad Fauzi and S. K. M.Kom, Agus Heri Yunial, *ANALISIS SENTIMEN (Sentiment Analysis): Evaluasi Sentimen Layanan Dataset Twitter US Airline*. Sleman: CV Bintang Semesta Media, 2024. [Online]. Available: https://www.google.co.id/books/edition/Analisis_Sentimen_Sentiment_Analysis/dgoREQAAQBAJ?hl=id&gbpv=1&dq=sentimen&pg=PA12&printsec=frontcover

[7] A. Sitanggang, Y. Umaidah, Y. Umaidah, R. I. Adam, and R. I. Adam, “Analisis Sentimen Masyarakat Terhadap Program Makan Siang Gratis Pada Media Sosial X Menggunakan Algoritma Naïve Bayes,” *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4902.

[8] F. Romadoni, Y. Umaidah, and B. N. Sari, “Text Mining Untuk Analisis Sentimen Pelanggan Terhadap Layanan Uang Elektronik Menggunakan Algoritma Support Vector Machine,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 9, no. 2, pp. 247–253, 2020, doi: 10.32736/sisfokom.v9i2.903.

[9] E. Salim and A. Solichin, “Analisis Sentimen Pada Media Sosial Twitter Terhadap Pelayanan Dinas Kependudukan Dan Pencatatan Sipil Menggunakan Algoritma Naïve Bayes,” *IDEALIS Indones. J. Inf. Syst.*, vol. 5, no. 2, pp. 79–86, 2022, doi: 10.36080/idealis.v5i2.2961.

[10] Fuad Amirullah, Syariful Alam, and M.Imam Sulistyo S, “Analisis Sentimen Terhadap Kinerja KPU Menjelang Pemilu 2024 Berdasarkan Opini Twitter Menggunakan Naïve Bayes,” *STORAGE J. Ilm. Tek. dan Ilmu Komput.*, vol. 2, no. 3, pp. 69–76,

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 3 Februari 2026
Hal: 1539-1548
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i3.631

2023, doi: 10.55123/storage.v2i3.2293.

[11] A. Munawaroh, R. Ridhoi, and R. Rudiman, "Sentiment Analysis Dengan Naïve Bayes Berbasis Orange Terhadap Resiko Pembangunan Ikn," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 587–592, 2024, doi: 10.36040/jati.v8i1.8454.

[12] A. N. Puspitasari, Y. Findawati, and Y. Rahmawati, "ANALISIS SENTIMEN TWEET PENGGUNA E-COMMERCE DENGAN MENGGUNAKAN METODE KLASIFIKASI NAÏVE BAYES," vol. 9, no. 3, pp. 1123–1132, 2024.

[13] R. K. Septiani, S. Anggraeni, and S. D. Saraswati, "Klasifikasi Sentimen Terhadap Ibu Kota Nusantara (IKN) pada Media Sosial Menggunakan Naive Bayes," *Tek. J. Ilm. Bid. Ilmu Rekayasa*, vol. 16, no. 2, pp. 245–254, 2022, [Online]. Available: https://jurnal.polsri.ac.id/index.php/teknika/article/view/4875

[14] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *J. KomtekInfo*, vol. 10, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.

[15] S. . MMSI, Yessy Asri, D. D. D. K. M.Kom, S. K. M.Kom, Listra Frigia Missianes Horhoruw, and S. A. R. ST, *MACHINE LEARNING & DEEP LEARNING : Analisis Sentimen Menggunakan Ulasan Pengguna Aplikasi*. Ponorogo: Uwais Inspirasi Indonesia, 2024. [Online]. Available: https://www.google.co.id/books/edition/MACHINE_LEARNING_DEEP_LEARNING_Analisis/Yu7uEAAAQBAJ?hl=id&gbpv=1&dq=sentimen&pg=PA64&printsec=frontcover