



Komparasi Ekstraksi Fitur TF-IDF dan Word2Vec pada Naïve Bayes untuk analisis Sentimen Pembangunan IKN di YouTube

Mu. Aldi Rahmad Fahrozi, Taghfirul Azhima Yoga Siswa*, Naufal Azmi Verdikha

Sains dan Teknologi, Teknik Informatika, Universitas Muhammadiyah Kalimantan Timur, Samarinda
Jl. Ir. H. Juanda No.15, Sidodadi, Kec. Samarinda Ulu, Kota Samarinda, Kalimantan Timur, Indonesia

Email: ¹2111102441019@umkt.ac.id, ^{2,*}tay758@umkt.ac.id, ³nav651@umkt.ac.id

Email Author Korespondensi: tay758@umkt.ac.id

Submitted: 13/01/2026; Accepted: 31/01/2026; Published: 31/01/2026

Abstrak—Pembangunan Ibu Kota Negara (IKN) memicu beragam respons publik di media sosial, khususnya YouTube, sehingga analisis sentimen diperlukan untuk memetakan persepsi masyarakat. Penelitian terdahulu menunjukkan akurasi yang relatif rendah, yaitu sebesar 60%, sehingga diperlukan pendekatan yang lebih efektif untuk meningkatkan kinerja klasifikasi. Penelitian ini bertujuan membandingkan kinerja Naïve Bayes dalam mengklasifikasikan sentimen publik terhadap pembangunan IKN menggunakan dua metode ekstraksi fitur, yakni TF-IDF dan Word2Vec. Data diperoleh dari komentar YouTube yang melalui tahap pre-processing, pelabelan oleh pakar, serta pengujian menggunakan 10-Fold Cross Validation. Hasil penelitian menunjukkan bahwa Multinomial Naïve Bayes berbasis TF-IDF memberikan kinerja terbaik dengan akurasi 83%, recall positif 82%, dan F1-score negatif 85%, dibandingkan Gaussian Naïve Bayes berbasis Word2Vec yang memperoleh akurasi 82% dengan recall positif 76%. Temuan ini menegaskan bahwa TF-IDF lebih efektif dan stabil dalam menangani karakteristik komentar singkat dibandingkan Word2Vec yang memerlukan korpus data lebih besar.

Kata Kunci: Analisis Sentimen; IKN; Ibu Kota Negara; Nusantara; Naïve Bayes; TF-IDF; Word2Vec

Abstract—The development of Indonesia's New Capital City (IKN) has generated diverse public responses on social media, particularly YouTube, making sentiment analysis necessary to map public perceptions. Previous studies have reported relatively low classification accuracy, reaching only 60%, indicating the need for more effective approaches to improve performance. This study aims to compare the performance of the Naïve Bayes algorithm in classifying public sentiment toward the IKN development using two feature extraction methods, namely TF-IDF and Word2Vec. The data were collected from YouTube comments and processed through preprocessing stages, expert-based labeling, and evaluation using 10-Fold Cross Validation. The results show that the TF-IDF-based Multinomial Naïve Bayes model achieves the best performance with an accuracy of 83%, a positive recall of 82%, and a negative F1-score of 85%, outperforming the Word2Vec-based Gaussian Naïve Bayes model, which attains an accuracy of 82% with a lower positive recall of 76%. These findings confirm that TF-IDF is more effective and stable in handling short-text comment characteristics than Word2Vec, which requires a larger corpus for optimal semantic representation.

Keywords: Sentiment Analysis; IKN; The Nation's Capital; Nusantara; Naïve Bayes; TF-IDF; Word2Vec

1. PENDAHULUAN

Pemindahan IKN ke Kalimantan Timur didorong oleh beban populasi dan masalah lingkungan Jakarta [1]. Meski diresmikan melalui UU No. 3 Tahun 2022 dengan pertimbangan lokasi strategis [2], temuan BPK mengenai kendala perencanaan dan lahan. Memicu berbagai respons publik terkait dampak sosial dan pembiayaan [3]. Analisis sentimen merupakan metode yang dapat digunakan untuk memahami berbagai respons serta mengukur opini masyarakat terhadap suatu topik [4]. Melalui pemanfaatan teks digital, analisis sentimen membantu bagaimana masyarakat mengekspresikan emosi, pandangan, dan tanggapan terhadap suatu peristiwa, produk, atau kebijakan tertentu [5]. Dalam pembangunan IKN, data sentimen banyak diperoleh dari platform media sosial yang menjadi wadah utama komentar masyarakat salah satunya YouTube [6].

YouTube, sebagai platform video dengan basis pengguna aktif terbanyak di dunia, menjadi arena diskusi di mana pengguna berinteraksi melalui fitur like, dislike, subscribe, dan kolom komentar [7]. Komentar pengguna yang diunggah pada platform YouTube mencerminkan berbagai bentuk sentimen publik, baik positif maupun negatif, sehingga relevan untuk dijadikan sumber data dalam analisis sentimen, selain itu dominasi YouTube sebagai platform berbasis video terbesar di dunia juga diperkuat oleh keberadaan YouTube Application Programming Interface (API) yang memfasilitasi proses pengambilan data sentimen [8]. Proyek IKN memicu beragam sentimen di kalangan masyarakat, yang terbagi menjadi sentimen positif, berfokus pada harapan pengembangan ekonomi lokal dan perbaikan infrastruktur, serta sentimen negatif yang mencerminkan kekhawatiran tentang penggusuran dan kerusakan lingkungan [9]. Penelitian terdahulu membahas analisis sentimen pembangunan IKN pada platform YouTube menunjukkan dominasi sentimen positif sebesar 14,30%, sementara sentimen negatif 5,60% dari 3.000 komentar [10]. Namun, penelitian pada platform Twitter menunjukkan dominasi sentimen negatif, 712 komentar, sedangkan sentimen positif 56 komentar [1]. Kedua penelitian tersebut memperkuat bahwa persepsi masyarakat terhadap pembangunan IKN di media sosial masih bersifat kontroversial.

Untuk mengolah data opini yang tidak terstruktur tersebut, Natural Language Processing (NLP) diterapkan sebagai metode interaksi antara bahasa manusia dan mesin, mengubah teks mentah menjadi data terstruktur yang siap diklasifikasikan [11][12][13]. Dalam penerapannya, NLP terhubung dengan berbagai algoritma klasifikasi

seperti Naive Bayes, SVM, dan Random Forest, yang berperan dalam mengidentifikasi serta mengelompokkan sentimen berdasarkan pola bahasa yang dipelajari dari data teks [14]. Naive Bayes sering menjadi pilihan utama karena kecepatan pelatihan dan prediksinya yang unggul, serta variasi tipenya seperti Multinomial, Gaussian, Bernoulli, dan Complement yang dapat disesuaikan dengan karakteristik data [15][16]. Dalam analisis sentimen terkait pembangunan IKN, penelitian pada platform X menunjukkan akurasi 84% menggunakan Naive Bayes. Namun, penelitian terdahulu pada platform YouTube menggunakan Naive Bayes hanya memperoleh akurasi 60% karena ketimpangan terjadi data [17]. Hal tersebut menunjukkan perlunya peningkatan akurasi melalui kualitas data, tahapan pre-processing, dan teknik ekstraksi fitur yang optimal [2].

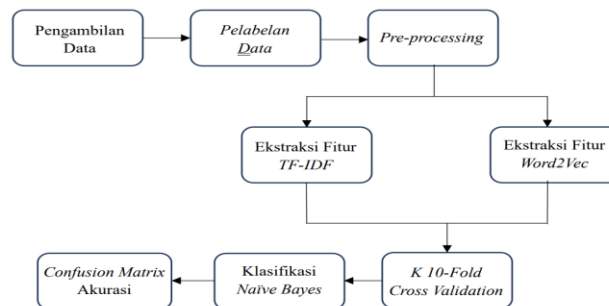
Ekstraksi fitur adalah bagian proses dalam NLP yang bertugas mengubah data mentah menjadi fitur numerik yang dapat diproses oleh algoritma klasifikasi [18]. Salah satu metode ekstraksi fitur adalah Term Frequency-Inverse Document Frequency (TF-IDF), sebuah metode yang bekerja dengan memberikan bobot tinggi pada kata-kata yang dianggap sering muncul dalam sebuah dokumen [19]. Riset [20] membuktikan efektivitasnya mencapai akurasi 92,5% saat dipadukan dengan Naive Bayes. Word2Vec merupakan metode yang merepresentasikan kata ke dalam bentuk vektor numerik untuk menangkap kesamaan makna antar kata dan keterkaitan kata dalam kalimat [21]. Penelitian terdahulu membuktikan bahwa Word2Vec mampu memberikan akurasi 81% pada klasifikasi sentimen menggunakan Naive Bayes [16]. Oleh karena itu, penelitian ini membandingkan TF-IDF dengan Word2Vec untuk menguji efektivitas ekstraksi fitur tersebut. Secara teoretis, Word2Vec dinilai lebih unggul karena menggunakan pendekatan word embedding yang mampu menangkap konteks semantik dan hubungan makna antar kata, berbeda dengan TF-IDF yang hanya berfokus pada statistik frekuensi kemunculan kata semata.

Terdapat sejumlah penelitian terdahulu yang berupaya menganalisis sentimen publik terkait IKN. Penelitian oleh Laia & Barus dan Saputri & Alita berfokus pada polarisasi sentimen di media sosial namun belum mengeksplorasi perbandingan efektivitas fitur ekstraksi secara mendalam [10][1]. Di sisi lain pada penelitian Mola yang menerapkan Naive Bayes pada data YouTube dengan akurasi yang masih terbatas di angka 60% akibat ketimpangan data [17]. Sementara itu, penelitian Islamanda & Sibaroni menguji Word2Vec namun menemukan bahwa metode ini memerlukan korpus besar untuk hasil optimal [16]. Berdasarkan tinjauan terhadap empat penelitian tersebut, ditemukan research gap berupa belum adanya studi komparatif yang secara spesifik menguji stabilitas kinerja antara ekstraksi fitur TF-IDF dan semantik Word2Vec pada algoritma Naive Bayes untuk data komentar singkat IKN yang cenderung tidak seimbang. Oleh karena itu, penelitian ini berkontribusi untuk mengisi celah tersebut dengan mengkomparasikan kinerja TF-IDF dan Word2Vec guna menentukan metode ekstraksi fitur yang paling optimal dalam meningkatkan akurasi klasifikasi sentimen pembangunan IKN.

2. METODOLOGI PENELITIAN

2.1 Metode Penelitian

Studi ini menguji sentimen IKN dari YouTube menggunakan pendekatan kuantitatif. Fokus utamanya adalah komparasi performa Naive Bayes berbasis TF-IDF vs Word2Vec yang diukur melalui metrik akurasi dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

Gambar 1, menggambarkan proses dalam klasifikasi teks, dimulai dari Pengambilan Data, dilanjutkan dengan Preprocessing dan Pelabelan Data, kemudian Ekstraksi Fitur, lalu divalidasi dengan K-Fold Cross Validation, diterapkan pada Model Naive Bayes, dan diakhiri dengan confusion matrix.

a. Pengambilan Data

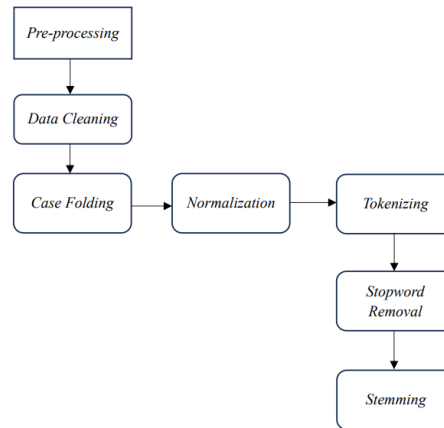
Pengambilan komentar dari 5 kanal YouTube bertopik IKN yang diunggah pada agustus 2025. Dataset yang dihasilkan mencakup atribut data_id dan isi komentar.

b. Pelabelan Data

Data hasil preprocessing diklasifikasikan ke dalam kategori sentimen positif dan negatif sebagai acuan pelatihan model [22]. Proses pelabelan dilakukan secara manual dan diverifikasi oleh satu orang pakar bahasa untuk menjamin validitas [23].

c. Preprocessing

Tahap Preprocessing merupakan tahap untuk mentransformasi data sesuai dengan format [24]. Rangkaian tahapan Preprocessing yang dilakukan ditampilkan pada Gambar 2.



Gambar 2.Tahap Preprocessing

Gambar 2, merincikan langkah Preprocessing teks komentar, adapun urutan Preprocessing dimulai dari data cleaning, case folding, Normalization, tokenizing, stopwords removal, dan stemming.

- 1) Data cleaning berperan penting dalam memastikan kualitas data dengan mengidentifikasi dan menangani data yang hilang, duplikat, tidak konsisten, serta tidak relevan [12].
- 2) Case folding merupakan proses mengubah seluruh huruf dalam teks menjadi huruf kecil guna mempermudah analisis teks [25].
- 3) Normalization adalah proses mengonversi kata tidak baku atau salah ejaan menjadi kata baku sesuai kaidah bahasa Indonesia [26].
- 4) Tokenizing adalah proses memecah teks menjadi kata-kata agar setiap kata dapat dianalisis secara terstruktur [27].
- 5) Stopword removal adalah proses menghilangkan kata-kata yang tidak penting agar model dapat lebih fokus pada kata penting [28].
- 6) Stemming adalah proses mengubah kata berimbuhan menjadi kata dasar dengan cara menghilangkan imbuhan pada kata tersebut, sehingga diperoleh bentuk kata dasar [29].

d. Ekstraksi Fitur

Penelitian ini menerapkan dua metode sebagai yaitu, TF-IDF [30], yang hasilnya menjadi input bagi Multinomial Naive Bayes. Word2Vec yang digunakan sebagai input untuk Gaussian Naive Bayes. Metode ini memiliki dua arsitektur utama, yaitu Continuous Bag of Words (CBOW) dan Skip-gram [31]. Pada penelitian ini menggunakan Word2Vec Skip-gram. Metode perhitungan TF-IDF dirumuskan sebagai berikut [26]:

$$TF = \frac{\text{jumlah kemunculan term dalam dokumen}}{\text{total jumlah kata dalam dokumen}} \tag{1}$$

$$IDF = \log\left(\frac{\text{jumlah dokumen dalam korpus}}{\text{jumlah dokumen yang mengandung term}}\right) \tag{2}$$

$$W_{i,j} = tf_{i,j} \times \log\left(\frac{n}{df_j}\right) \tag{3}$$

Keterangan dari rumus tersebut adalah $W_{i,j}$ yang mempresentasikan bobot term, sementara $tf_{i,j}$ adalah frekuensi kemunculan term, yang kemudian dikalikan dengan logaritma dari total dokumen n dibagi frekuensi dokumen df_j .

e. Pembagian data

Evaluasi kinerja menggunakan Stratified 10-Fold Cross Validation untuk memastikan distribusi kelas seimbang. Proses dilakukan iteratif (9 latih, 1 uji) dengan ekstraksi fitur terpisah di setiap fold guna mencegah data leakage. Akurasi akhir diperoleh dari rata-rata seluruh iterasi.

f. Penerapan Naive Bayes

Naive Bayes mengklasifikasikan probabilitas kelas dengan asumsi independensi fitur [32]. Multinomial Naive Bayes diterapkan pada data diskret TF-IDF dengan parameter Alpha untuk mencegah zero probability [20], sedangkan Gaussian Naive Bayes digunakan untuk fitur kontinu Word2Vec guna memenuhi distribusi normal tanpa mengganggu nilai vektor semantik [16].

$$P(H | X) = \frac{P(X | H) \cdot P(H)}{P(X)} \tag{4}$$

Dalam persamaan tersebut, X merepresentasikan data yang kelasnya belum diketahui, sedangkan H adalah hipotesis kelas target. Notasi $P(H | X)$ menyatakan probabilitas posterior bahwa dokumen X termasuk dalam kelas H. Nilai ini ditentukan oleh $P(X | H)$ yang merupakan likelihood atau probabilitas kemunculan fitur X dalam kelas H, serta $P(H)$ sebagai probabilitas awal (prior). Adapun $P(X)$ adalah probabilitas data yang bernilai konstan bagi semua kelas sehingga dapat diabaikan dalam proses klasifikasi.

g. Evaluasi Model

Evaluasi kinerja dilakukan untuk membandingkan efektivitas model Naive Bayes dengan Ekstraksi fitur dalam mengklasifikasikan sentimen positif dan negatif. Berdasarkan Confusion Matrix, kinerja model diukur menggunakan empat metrik utama sebagaimana dirumuskan sebagai berikut [33].

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{5}$$

$$\text{Precision} = \frac{TP}{FP+TP} \times 100\% \tag{6}$$

$$\text{Recall} = \frac{TP}{FN+FP} \times 100\% \tag{7}$$

$$F1 - \text{score} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{8}$$

Adapun pengertian True Positive (TP) dan True Negative (TN) merepresentasikan jumlah data positif dan negatif yang terklasifikasi dengan benar. Sebaliknya, kesalahan klasifikasi terdiri dari False Positive (FP) atau prediksi positif pada data negatif, serta False Negative (FN) atau prediksi negatif pada data positif.

3. HASIL DAN PEMBAHASAN

3.1 Pengambilan Data

Penelitian ini menerapkan metode scraping untuk mengambil data pada platform YouTube menggunakan YouTube Data API v3. Adapun data yang diperoleh ditunjukkan pada Gambar 3.

video_id	komentar
YvpdP_44gPA	Jika dilakukan perekayasaan terhadap siklus hu...
YvpdP_44gPA	Hujan itu justru bagus untuk kawasan hutan... ..
YvpdP_44gPA	Rekayasa hujan memang IKN gak tapi dipindah Ba...
YvpdP_44gPA	Hukum direkayasa,ljazah direkayasa,data direkay...

Gambar 3.Data Komentar YouTube

Proses scraping menghasilkan 4.093 data komentar unik melalui metode scraping pada Gambar 3.

3.2 Pelabelan Data

Setelah data berhasil diambil, data diklasifikasikan ke dalam sentimen positif dan negatif. Untuk menjamin validitas konteks, pelabelan ini divalidasi oleh pakar bahasa, Ibu Purwanti, M.Hum. (Universitas Mulawarman). Hasil pelabelan data ditampilkan pada Gambar 4.

video_id	komentar	sentimen
dywMpVoqE-w	Gpp yg penting IKN 100% pindah , asal jangan d...	negatif
dywMpVoqE-w	IKN bakar duit negara	negatif
dywMpVoqE-w	Kalao sampe sekelas presiden ngomongim ginian....	negatif

Gambar 4. Pelabelan Data

Berdasarkan Gambar 4, Pelabelan bersama pakar bahasa terhadap 4.093 data valid menunjukkan sentimen positif 1.937, negatif 1.504 dan, netral 652 komentar terkait pembangunan IKN. Berdasarkan alur klasifikasi tersebut, dalam penelitian ini, pelabelan data difokuskan hanya pada dua kelas sentimen, yaitu positif dan negatif yang menyisakan 3.441 komentar.

3.3 Preprocessing

Tahap pre-processing terdiri dari, case folding, normalisasi 2.780 kata tidak baku, tokenizing, stopword removal pada 1.974 komentar, serta stemming untuk menyederhanakan kata menjadi bentuk dasar. Dari keseluruhan 3.441

komentar, seluruh teks berhasil diproses dan direpresentasikan kembali dalam bentuk token yang bersih. Hasil Salah satu sampel data dari sebelum dan sesudah pre-processing pada Gambar 5 sebagai berikut.

text_stopwords_removed	text_stemmed
[tidak, apa, apa, penting, ikn, pindah, asal, ...	[tidak, apa, apa, penting, ikn, pindah, asal, ...
[ikn, bakar, duit, negara]	[ikn, bakar, duit, negara]
[kalau, sekelas, presiden, berbicara, beginian...	[kalau, kelas, presiden, bicara, begini, ada, ...

Gambar 5. Hasil Preprocessing

Gambar 5, menunjukkan hasil akhir preprocessing yang mana merubah dari kalimat “gpp yang penting ikn pindah” Menjadi “tidak apa penting ikn pindah”, proses ini memastikan simplifikasi bentuk kata sebelum masuk ke tahap ekstraksi fitur.

3.4 Ekstraksi Fitur

3.4.1 TF-IDF

Hasil ekstraksi fitur TF-IDF dapat dilihat pada Gambar 6 dan Gambar 7.

	apa	gegabah	ikn	investor	iya	kabarnya	lebih
sekarang							
apa	1						
kabarnya ikn			1				
menghilang				1			
gegabah		1					
urusi ikn			1				
lebih							1
penting							
omnibuslaw							
melancarkan				1	1		
investor ikn			1				
iya							

Gambar 6. Dimensi TF-IDF

Representasi biner pada Gambar 6 memperlihatkan struktur data input biner 1 dan 0, berdasarkan kosa kata yang ditentukan merepresentasikan matriks kemunculan kata sebagai basis perhitungan bobot TF-IDF. Adapun hasil perhitungan TF-IDF sebagaimana yang ditunjukkan gambar 7.

Term	DF	IDF	TF-IDF
ikn	3107	1.092	1.267
tidak	863	2.372	0.861
indonesia	793	2.457	0.783

Gambar 7. Hasil pembobotan TF-IDF

Gambar 7 tersebut merangkum hasil perhitungan bobot kata menggunakan metode TF-IDF, yang menampilkan Term mulai dari "ikn" hingga "orang" beserta statistik frekuensi kemunculannya dalam seluruh dokumen (DF), nilai keunikannya (IDF), dan skor akhir relevansinya (TF-IDF).

3.4.2 Word2Vec

Hasil vektorisasi Word2Vec pada Gambar 8, memperlihatkan kluster kata-kata dengan konteks serupa yang berada dalam kedekatan ruang vektor. Visualisasi hasil Word2Vec dapat dilihat pada Gambar 8.

Kata	Dimensi-1	Dimensi-2	Dimensi-3	Dimensi-4
ikn	-0.120317	0.111968	-0.292131	0.112619
tidak	-0.240942	0.258747	-0.118282	0.262553
indonesia	-0.117730	0.152435	-0.270065	-0.185346

Gambar 8. Word2Vec

Gambar 8, menampilkan dimensi vektor Word2Vec kata “ikn”, “tidak”, dan “indonesia”, yang masing-masing direpresentasikan sebagai matrix dengan empat dimensi tampilan awal dari 3441 x 100, Nilai vektor di rata-ratakan untuk membentuk representasi dokumen sebagai masukan pada proses klasifikasi sentimen.

3.5 Pembagian Data

Pada skema 10-fold cross validation dengan teknik stratified, sebanyak 3.441 komentar dibagi ke dalam sepuluh fold dengan proporsi kelas yang tetap terjaga. Setiap fold terdiri atas 3.065 data latih dan 341 data uji, dengan distribusi kelas positif sebanyak 1.722 data latih dan 191 data uji, serta kelas negatif 1.343 data latih dan 150 data uji. Pendekatan ini memastikan keseimbangan kelas pada setiap fold sehingga evaluasi kinerja model dapat dilakukan secara valid dan konsisten. Adapun visualisasi nya pada Gambar 9.

Fold	Train_Pos	Train_Neg	Train_Total	Test_Pos	Test_Neg	Test_Total
1	1722	1343	3065	191	150	341
2	1722	1343	3065	191	150	341
3	1722	1343	3065	191	150	341
4	1721	1344	3065	192	149	341
5	1721	1344	3065	192	149	341
6	1721	1344	3065	192	149	341
7	1722	1344	3066	191	149	340
8	1722	1344	3066	191	149	340
9	1722	1344	3066	191	149	340
10	1722	1344	3066	191	149	340

Gambar 9. 10-Fold Cross Validation

Merujuk Gambar 9, menunjukkan adanya perbedaan kecil jumlah data pada setiap fold akibat penerapan stratified 10-fold cross-validation yang harus mempertahankan proporsi kelas sekaligus membagi data ke dalam bilangan bulat. Dari total 1.504 data negatif dan 1.937 data positif, setiap fold uji berisi sekitar 150 data negatif dan 191 data positif, sehingga jumlah data uji berada pada kisaran 341 data, sedangkan data latih berjumlah sekitar 3.065 data. Variasi ini merupakan hal wajar dari pembagian stratified pada dataset dengan jumlah data yang tidak habis dibagi sepuluh, serta tidak memengaruhi validitas evaluasi model. stratified memungkinkan setiap fold merepresentasikan distribusi kelas yang serupa dengan dataset asli, sehingga hasil evaluasi model klasifikasi sentimen yang diperoleh lebih objektif dan dapat dipercaya.

3.6 Algoritma Naive Bayes

Setelah pembagian data 10-fold cross validation, dilakukan pelatihan dan pengujian pada setiap fold untuk membandingkan Multinomial Naive Bayes berbasis TF-IDF dan Gaussian Naive Bayes berbasis Word2Vec. Hasil komparasi kinerja kedua metode tersebut disajikan pada Tabel 1.

Tabel 1. Akurasi Metode Klasifikasi dan Ekstraksi

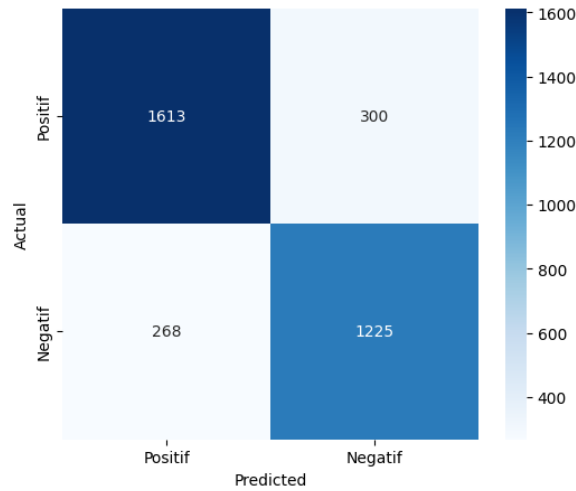
	Multinomial Naive Bayes + Tf-Idf	Gaussian Naive Bayes + Word2Vec
Fold 1	0,8328	0,8240
Fold 2	0,8563	0,8299
Fold 3	0,8446	0,8123
Fold 4	0,8827	0,8563
Fold 5	0,8182	0,8094
Fold 6	0,7977	0,7947
Fold 7	0,8265	0,8000
Fold 8	0,8206	0,8088
Fold 9	0,8441	0,8324
Fold 10	0,8088	0,7824
Rata-rata	0,8332	0,8150

Tabel 1, menyajikan hasil komparasi kinerja yang menunjukkan bahwa Multinomial Naive Bayes berbasis TF-IDF memperoleh rata-rata akurasi lebih tinggi sebesar 0,8332 dibandingkan Gaussian Naive Bayes berbasis Word2Vec sebesar 0,8150. Meskipun demikian, pada beberapa fold tertentu, Word2Vec menunjukkan performa yang kompetitif. Perbedaan kinerja ini mengindikasikan bahwa pendekatan berbasis bobot frekuensi kata cenderung lebih stabil pada karakteristik data yang digunakan, sementara representasi semantik Word2Vec lebih efektif pada kondisi data dengan variasi konteks yang lebih beragam. Mengingat distribusi kelas yang tidak

seimbang, evaluasi kinerja model tidak hanya bergantung pada nilai akurasi sehingga diperlukan analisis lanjutan menggunakan confusion matrix.

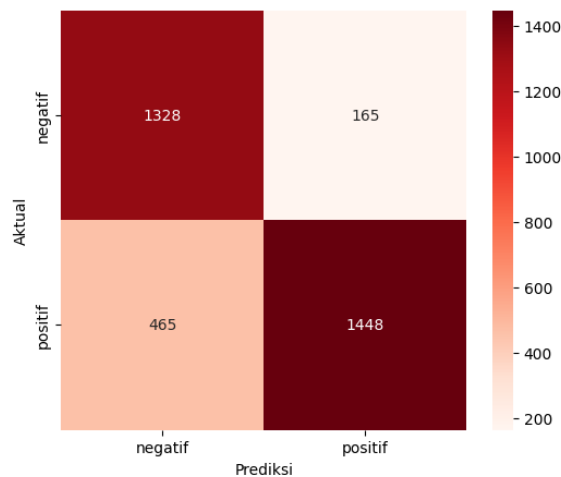
3.7 Evaluasi Model

Evaluasi performa model dilakukan menggunakan confusion matrix 2x2 untuk mengukur efektivitas kombinasi Naive Bayes dengan TF-IDF dan Word2Vec dalam mengklasifikasikan sentimen, sebagaimana ditunjukkan pada Gambar 10 dan 11:



Gambar 10. Confusion matrix (Naive Bayes + Tf-Idf)

Gambar 11, menunjukkan 1.225 data True Negative dan 1.613 data True Positive. Sebaliknya, kesalahan prediksi terjadi pada 268 data False Positive dan 300 data False Negative.



Gambar 11. Confusion matrix (Naive Bayes + Word2Vec)

Gambar 12, menunjukkan 1.326 data True Negative dan 1.448 data True Positive. Adapun kesalahan klasifikasi terdiri dari 465 data False Positive dan 165 data False Negative. Adapun hasil evaluasi menjadi nilai hitung untuk akurasi, precision, recall dan f1-Score pada Tabel 2.

Tabel 2. Perbandingan hasil antara Tf-Idf dan Word2Vec

		Precision	Recall	F1-Score
Multinomial Naive Bayes + Tf-Idf	Positif	80	72	81
	Negatif	86	84	85
	Akurasi			83
Gaussian Naive Bayes+ Word2Vec	Positif	90	76	82
	Negatif	74	89	81
	Akurasi			82

Berdasarkan hasil evaluasi pada Tabel 2, model Multinomial Naive Bayes yang dikombinasikan dengan Tf-Idf terbukti lebih konsisten dalam mengenali sentimen negatif. Hal ini ditunjukkan oleh nilai F1-Score negatif sebesar 85 yang lebih tinggi dibandingkan F1-Score positif sebesar 81. Kinerja pada kelas negatif terlihat lebih



stabil dengan Precision 86% dan Recall 84%, sedangkan kelas positif mencatat angka Precision 80% dan Recall 82%. Sementara itu, model Gaussian Naive Bayes yang dikombinasikan dengan Word2Vec menunjukkan karakteristik yang berbeda. Model ini sangat presisi pada kelas positif dengan nilai 90% namun kurang mampu menjangkau seluruh data positif karena nilai Recall hanya 76%. Pola sebaliknya terjadi pada kelas negatif di mana Recall sangat tinggi mencapai 89% namun memiliki Precision yang lebih rendah yaitu 74%. Walaupun terdapat ketimpangan tersebut, nilai akhir F1-Score kedua kelas relatif berimbang di angka 82 dan 81.

3.8 Pembahasan

Hasil penelitian ini menunjukkan bahwa model klasifikasi berbasis TF-IDF dengan Multinomial Naive Bayes memiliki performa yang lebih unggul dibandingkan Word2Vec dengan, dengan akurasi sebesar 83% untuk TF-IDF dan 82% untuk Word2Vec. Temuan ini sejalan dengan penelitian terdahulu yang menyatakan bahwa metode berbasis pembobotan statistik seperti TF-IDF cenderung lebih stabil pada dataset berskala kecil [34]. Rendahnya performa Word2Vec dalam penelitian ini dipengaruhi oleh keterbatasan jumlah data serta karakteristik komentar YouTube yang singkat dan juga ketimpangan data. Perbedaan ini sejalan dengan penelitian terdahulu yang menyebutkan bahwa model Word2Vec sangat bergantung pada korpus yang besar untuk mempelajari relasi semantik secara optimal [35]. Namun dari keterbatasan data tersebut, Word2vec Naive bayes memiliki akurasi 82% berbanding dengan penelitian terdahulu yang menerapkan metode yang sama tetapi akurasi yang diperoleh lebih rendah sebesar 77.18% dengan data 6.495 [16]. Meskipun demikian, capaian akurasi 83% untuk TF-IDF dan 82% pada Word2Vec pada penelitian ini tetap menunjukkan performa yang kompetitif dan mengalami peningkatan dibandingkan studi terdahulu mengenai analisis sentimen IKN menggunakan Naive Bayes yang hanya mencapai akurasi 60% pada kasus ketimpangan data di platform serupa [17]. Hal ini menunjukkan bahwa pendekatan yang digunakan dalam penelitian ini mampu meningkatkan kinerja analisis sentimen terhadap isu IKN dibandingkan penelitian terdahulu.

4. KESIMPULAN

Berdasarkan penelitian analisis sentimen IKN di YouTube, metode TF-IDF dengan Multinomial Naive Bayes terbukti lebih unggul dengan akurasi 83% dibandingkan Word2Vec dengan Gaussian Naive Bayes yang memperoleh 82%. TF-IDF menunjukkan kinerja yang lebih stabil dan seimbang, sedangkan Word2Vec belum bekerja secara optimal dalam mendeteksi keseluruhan sentimen positif. Hal ini terjadi karena algoritma Word2Vec membutuhkan korpus data yang jauh lebih besar untuk dapat membentuk representasi semantik yang akurat, berbeda dengan TF-IDF yang tetap efektif menangani frekuensi kata pada dataset berskala kecil.

REFERENCES

- [1] G. A. Saputri and D. Alita, "Analisis Sentimen Twitter Terhadap Pemindahan Ibu Kota Negara Menggunakan Support Vector Machine," *J. Inform. J. Pengemb. IT*, vol. 9, no. 3, pp. 213–223, 2024, doi: 10.30591/jpit.v9i3.6612.
- [2] F. Zamzami, R. Hidayat, and R. Fathonah, "Penerapan Algoritma Naive Bayes Classifier Untuk Analisis Sentimen Komentar Twitter Proyek Pembangunan Ikn," *Fakt. Exacta*, vol. 17, no. 1, pp. 47–57, 2024, doi: 10.30998/faktorexacta.v17i1.22265.
- [3] M. Dimas, R. Chasis, D. Azahari, and M. I. Sa'ad, "Analisis Sentimen Terhadap Kontroversi Pembangunan IKN di Media Sosial Twitter Menggunakan Metode Naive Bayes," *Bull. Inf. Technol.*, vol. 6, no. 2, pp. 97–108, 2025, doi: 10.47065/bit.v5i2.1993.
- [4] F. Fitriyadi and A. Astikasari, "Analisis Sentimen Masyarakat Terhadap Kebijakan Kenaikan UMK 6,5% Menggunakan Metode Naive Bayes," *J. Ris. Sist. dan Teknol. Inf.*, vol. 3, no. 1, pp. 26–35, 2024, doi: 10.30787.
- [5] A. W. Nyoman, N. M. A. E. D. Wirastuti, and I. B. G. Manuaba, "Analisis Sentimen Tanggapan Masyarakat Tentang Garuda IKN Menggunakan Metode Naive Bayes," *Decod. J. Pendidik. Teknol. Inf.*, vol. 5, no. 1, pp. 27–40, 2025, doi: 10.51454/decode.v5i1.860.
- [6] K. K. Safra and E. Zuliarso, "Analisis Sentimen Terhadap Pelaksanaan Pilkada 2024 Pada Media Sosial Youtube Menggunakan Metode Decision Tree," *J. Inform. Teknol. dan Sains*, vol. 7, no. 1, pp. 117–126, Feb. 2025, doi: 10.51401/jinteks.v7i1.5295.
- [7] C. A. Misrun, E. Haerani, M. Fikry, and E. Budianita, "Analisis sentimen komentar youtube terhadap Anies Baswedan sebagai bakal calon presiden 2024 menggunakan metode naive bayes classifier," *J. Comput. Sci. Inf. Technol.*, vol. 5, no. 2, pp. 358–366, 2024, doi: 10.37859.
- [8] A. A. M. Putra, Islamiyah, and L. J. Muhammad, "Analisis Sentimen Pengguna Youtube Terhadap Uang Baru Tahun Emisi 2022 Menggunakan Metode Naive Bayes Classifier," *Adopsi Teknol. dan Sist. Inf.*, vol. 3, no. 1, pp. 17–27, 2024, doi: 10.30872/atasi.v3i1.1177.
- [9] A. Yusuf, A. Rizani, R. Fitri, K. Nursyaiful Priyo Pamungkas, and W. Arifha Saputra, "Sentimen Positif Atau Negatif: Perspektif Masyarakat Terhadap Pemindahan Ibu Kota Negara Positive or Negative Sentiment: Public Perspectives on the Relocation of the National Capital," *J. Masy. Indones.*, vol. 50, no. 2, pp. 277–300, 2024, doi: 10.55981/jmi.2024.8842.
- [10] N. A. Laia and S. P. Barus, "Analisis Sentimen YouTube: 'Di Balik Ambisi Jokowi dalam IKN,'" *J. Pustaka AI (Pusat Akses Kaji. Teknol. Artif. Intell.*, vol. 5, no. 1, pp. 07–12, 2025, doi: 10.55382/jurnalpustakaai.v5i1.891.
- [11] N. Hadi and D. Sugiarto, "Analisis Sentimen Pembangunan IKN pada Media Sosial X Menggunakan Algoritma SVM, Logistic Regression dan Naive Bayes," *J. Inform. J. Pengemb. IT*, vol. 10, no. 1, pp. 37–49, 2025, doi:



- 10.30591/jpit.v10i1.7106.
- [12] A. Lia, A. Rahim, and T. A. Yoga Siswa, “Analisis Sentimen Aplikasi Mysiloam Menggunakan Metode Naïve Bayes,” *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 1, 2025, doi: 10.23960/jitet.v13i1.5997.
- [13] Nurfiyah and R. Nurfan, “Penerapan Metode Natural Language Processing (NLP) Pada Question Answering System Untuk Media Informasi Mahasiswa Universitas Bhayangkara Jakarta Raya,” *J. Inf. Inf. Secur.*, vol. 4, no. 2, pp. 175–186, 2023.
- [14] H. Wijaya and N. Hayati, “Natural Language Processing (NLP) Untuk Analisis Sentimen Ulasan Seblak Bandung Pedas Kudus,” *J. Bus. Audit Inf. Syst.*, vol. 8, no. 1, pp. 13–22, 2025, doi: 10.30813.
- [15] P. Octavia, “Implementasi Algoritma Naïve Bayes Dalam Klasifikasi Produk – Produk Terlaris,” *JSAI J. Sci. Appl. Informatics*, vol. 7, no. 2, pp. 411–414, 2024, doi: 10.36085.
- [16] M. D. Islamanda and Y. Sibaroni, “Whoosh User Sentiment Analysis on Social Media Using Word2Vec and the Best Naïve Bayes Probability Model,” *J. dan Penelit. Tek. Inform.*, vol. 8, no. 3, pp. 1558–1568, 2024, doi: 10.33395.
- [17] S. A. S. Mola, Iqbal Muhammad Iskandar, J. E. Pidu Dimu, and W. Y. Seran, “Analisis Sentimen Pembangunan Ibu Kota Negara Indonesia Menggunakan Metode Naïve Bayes, Dan K-Nearest Neighbor,” *HOAQ (High Educ. Organ. Arch. Qual. J. Teknol. Inf.)*, vol. 15, no. 2, pp. 151–157, 2024, doi: 10.52972/hoaq.vol15no2.p151-157.
- [18] D. Alwan and M. A. Ridla, “Averaged Word2vec sebagai Ekstraksi Fitur pada Analisis Sentimen Ulasan Film di IMDb menggunakan Artificial Neural Network (ANN),” *J. Sist. dan Teknol. Inf. Indones.*, vol. 9, no. 1, pp. 36–45, 2024, doi: doi.org/10.32528/justindo.v9i1.1204.
- [19] A. Firizkiansah, A. Muhammad, and I. R. Maulana, “Optimasi Klasifikasi Data Teks Menggunakan Algoritma Logistic Regression dengan TF-IDF dan SMOTE,” *JIKOMTI J. Ilm. Ilmu Komput. dan Teknol. Inf.*, vol. 2, no. 1, pp. 29–36, 2025, [Online]. Available: <https://ojs.sains.ac.id/index.php/Jikomti/article/view/97/119>
- [20] Ardiansyah and Kurniawan, “Optimasi Metode Naïve Bayes Classifier Menggunakan Pendekatan Term Frequency-Inverse Document Frequency (TF-IDF) Pada Analisis Sentimen,” *J. Sci. Appl. Informatics*, vol. 7, no. 3, pp. 458–464, 2024, doi: 10.36085.
- [21] D. N. Febianty and M. Rahardi, “A Sentiment Analysis of Public Perception Toward Pets in Public Spaces Using Logistic Regression and Word Embedding,” *J. Appl. Informatics Comput.*, vol. 9, no. 4, pp. 1846–1851, 2025, doi: 10.30871/jaic.v9i4.10245.
- [22] H. Firda et al., “Perbandingan Pelabelan Rating-based dan Inset Lexicon-based dalam Analisis Sentimen Menggunakan SVM (Studi Kasus: Ulasan Aplikasi GoBiz di Google Play Store),” *Sist. J. Sist. Inf.*, vol. 14, pp. 516–528, 2025, doi: <https://doi.org/10.32520/stmsi.v14i2.4795>.
- [23] M. R. Virgiansyah, S. Stephanie, and M. Rizky Pribadi, “Analisis Sentimen terhadap Jalan Rusak di Palembang Pada Media Sosial Menggunakan Algoritma Naïve Bayes,” *TIN Terap. Inform. Nusant.*, vol. 5, no. 1, pp. 1–10, 2024, doi: 10.47065/tin.v5i1.5239.
- [24] S. Shevira, I. M. A. D. Suarjaya, and P. W. Buana, “Pengaruh Kombinasi dan Urutan Pre-Processing pada Tweets Bahasa Indonesia,” *JITTER, J. Ilm. Teknol. dan Komput.*, vol. 3, no. 2, p. 1074, 2022, doi: 10.24843/jrti.2022.v03.i02.p06.
- [25] R. Rahman Salam, M. Fajri Jamil, and Y. Ibrahim, “Sentiment Analysis of Cash Direct Assistance Distribution for Fuel Oil Using Support Vector Machine Analisis Sentimen Terhadap Bantuan Langsung Tunai (BLT) Bahan Bakar Minyak (BBM) Menggunakan Support Vector Machine,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. April, pp. 27–35, 2023.
- [26] M. R. Syafapri, E. Haerani, I. Iskandar, and L. Afriyanti, “Sentiment classification of interfaith marriage ban using Naive Bayes Classifier method,” *J. Comput. Sci. Inf. Technol. (CoSciTech)*, vol. 5, no. 1, pp. 10–18, 2024, doi: 10.37859.
- [27] A. Putri, C. S. Hardiana, E. Novfuja, F. Try, and P. Siregar, “Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. April, pp. 20–26, 2023, doi: 10.57152.
- [28] F. A. Ryandi, D. Pratiwi, S. Sari, and J. Sains, “Analisis Sentimen Masyarakat Di Media Sosial X Terhadap Kemenkes Dengan Naive Bayes dan SVM,” *J. Sains dan Teknol.*, vol. 7, no. 1, pp. 1–6, 2025.
- [29] R. Kusumastuti, D. Oktafiani, and Y. Astica, “Optimasi Algoritma Stemming Porter untuk Pemrosesan Teks dalam Bahasa Indonesia,” vol. 6, no. 1, pp. 42–52, 2025, [Online]. Available: <https://jifsi.unisti.ac.id/index.php/JIFSI%0D>
- [30] Maulidya Prastita Syah, Ajeng Puspa Wardani, Mohammad Idhom, and Trimono, “Perbandingan Representasi Teks Tf-Idf Dan Bert Terhadap Akurasi Cosine Similarity Dalam Penilaian Otomatis Jawaban Berbasis Teks,” *Data Sci. Indones.*, vol. 5, no. 1, pp. 47–59, 2025, doi: 10.47709/dsi.v5i1.6021.
- [31] M. Permatasari and N. N. Pusparini, “Analisis Kinerja Sistem Informasi Pengiriman Surat Dengan Pendekatan UML Pada Perusahaan Ekspedisi memiliki dampak yang signifikan terhadap produktivitas dan efektivitas pengelolaan proses sebuah perusahaan ekspedisi dengan menggunakan pendekatan UML,” *Switch J. Sains dan Teknol. Inf.*, vol. 2, no. 6, 2024, doi: 10.62951.
- [32] Q. Hasanah, H. Oktavianto2, and Y. D. Rahayu, “Analisis Algoritma Gaussian Naive Bayes Terhadap Klasifikasi Data Pasien Penderita Gagal Jantung Gaussian Naive Bayes Algorithm Analysis Of Data Classification Of Heart Failure Patiens Jurnal Smart Teknologi,” *J. Smart Teknol.*, vol. 3, no. 4, pp. 382–389, 2022, [Online]. Available: <https://jurnal.unmuhjember.ac.id/index.php/JST/article/view/7597/3890>
- [33] P. Bintoro, T. H. Andika, A. F. Yulia, and P. Widiandana, “Analisis Sentimen di Twitter Menggunakan Pendekatan Machine Learning Sentiment Analysis on Twitter Using Machine Learning Approach,” vol. 1, no. 1, pp. 33–39, 2023, [Online]. Available: <https://share.google/WZmGhqKiyF7fITXOJ>
- [34] D. E. Cahyani and I. Patasik, “Performance comparison of TF-IDF and Word2Vec models for emotion text classification,” *Bull. Electr. Eng. Informatics*, vol. 10, no. 5, pp. 2780–2788, 2021, doi: 10.11591/eei.v10i5.3157.
- [35] Z. Zhan, “Comparative Analysis of TF-IDF and Word2Vec in Sentiment Analysis : A Case of Food Reviews,” *ITM Web Conf.*, vol. 02013, 2025, doi: doi.org/10.1051/itmconf/20257002013.