

ANALISIS SENTIMEN PENGGUNA PLATFORM X TERHADAP KEBIJAKAN BBM E10 MENGGUNAKAN ALGORITMA NAIVE BAYES

Suci Intania Indah¹, Hannie², Ahmad Khusaeri³

^{1,2,3}Universitas Singaperbangsa Karawang; Jl. HS.Ronggo Waluyo, Puseurjaya, Kec. Telukjambe Tim., Kabupaten Karawang, Jawa Barat; 41361

Keywords:

Analisis Sentimen, Naïve Bayes, KDD, BBM E10, Media Sosial X.

Correspondent Email:

2210631250100@student.unsika.ac.id

Abstrak. Kebijakan mandatori campuran etanol 10% pada bahan bakar minyak (BBM E10) yang diumumkan pada Oktober 2025 memicu beragam respons dari masyarakat Indonesia, khususnya di platform media sosial X. Penelitian ini bertujuan menganalisis sentimen pengguna Platform X terhadap kebijakan BBM E10 dan mengevaluasi performa model klasifikasi menggunakan algoritma Multinomial Naïve Bayes. Metode yang diterapkan adalah Knowledge Discovery in Database (KDD) dengan tahapan data selection, preprocessing, transformation, data mining, dan evaluasi. Data dikumpulkan melalui crawling menggunakan API Twitter dengan kata kunci "etanol" dan "BBM E10" pada periode September 2025 hingga Januari 2026, menghasilkan 4.847 tweet. Setelah penghapusan duplikat, diperoleh 2.385 data yang dilabeli secara hybrid—kombinasi pelabelan otomatis berbasis leksikon dan verifikasi manual oleh ahli bahasa. Transformasi fitur dilakukan menggunakan TF-IDF. Klasifikasi dilakukan dengan tiga skenario pembagian data: 70:30, 80:20, dan 90:10. Hasil analisis menunjukkan sentimen positif mendominasi dengan 1.034 tweet (43,35%), diikuti negatif 989 tweet (41,47%), dan netral 362 tweet (15,18%). Skenario 90:10 menghasilkan performa terbaik dengan akurasi 70,7%, presisi 69%, recall 57%, dan F1-Score 54%. Penelitian ini membuktikan bahwa pendekatan text mining berbasis Multinomial Naïve Bayes mampu mengklasifikasikan sentimen publik terhadap kebijakan BBM E10 secara otomatis dengan tingkat akurasi yang memadai.



Copyright © 2026 The Author(s), Published by Jurnal Informatika dan Teknik Elektro Terapan. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Abstract. The mandatory policy of 10% ethanol blending in fuel oil (BBM E10), announced in October 2025, triggered a wide range of responses from the Indonesian public, particularly on the X social media platform. This study aims to analyze the sentiment of Platform X users towards the BBM E10 policy and evaluate the performance of a classification model built using the Multinomial Naïve Bayes algorithm. The method applied is Knowledge Discovery in Database (KDD), encompassing data selection, preprocessing, transformation, data mining, and evaluation stages. Data was collected through crawling using the Twitter API with keywords "ethanol" and "BBM E10" during September 2025 to January 2026, yielding 4,847 tweets. After removing duplicates, 2,385 records were labeled using a hybrid approach—combining lexicon-based automatic labeling with manual verification by a language expert. Feature transformation was performed using TF-IDF. Classification was conducted under three data splitting scenarios: 70:30, 80:20, and 90:10. Results showed that positive sentiment dominated with 1,034 tweets (43.35%), followed by negative with 989 tweets (41.47%), and neutral with 362 tweets (15.18%). The 90:10 scenario produced the best performance with an accuracy of 70.7%, precision of 69%, recall of 57%, and

F1-Score of 54%. This study demonstrates that the Multinomial Naïve Bayes-based text mining approach is capable of automatically classifying public sentiment towards the BBM E10 policy with adequate accuracy.

1. PENDAHULUAN

Kehidupan manusia saat ini sangat bergantung pada dua aspek utama, yaitu energi terutama bahan bakar minyak (BBM) dan teknologi digital[1]. Kebijakan tentang BBM seringkali menjadi topik yang sensitif di kalangan masyarakat[2]. Di sisi lain, platform media sosial seperti X menjadi wadah di mana suara publik disuarakan dengan cepat dan masif. Perpaduan antara kebijakan energi dan ekspresi digital inilah yang melatarbelakangi penelitian ini.

Pemerintah Indonesia melalui Menteri Energi dan Sumber Daya Mineral (ESDM) Bahlil Lahadalia mengumumkan kebijakan mandatori campuran etanol 10% pada bahan bakar minyak (BBM E10) pada 7 Oktober 2025 (Nugraha Perdana, 2025). Berdasarkan laporan We Are Social pada Juli 2025, Indonesia termasuk negara dengan pengguna X terbesar keempat di dunia dengan sekitar 23,76 juta pengguna, sehingga platform X menjadi wadah yang aktif untuk menyuarakan opini. Banyak pengguna merespons kebijakan ini secara positif, mengaitkannya dengan pengurangan emisi karbon, namun sebagian lain mengungkapkan kekhawatiran terkait dampak etanol terhadap komponen mesin kendaraan.

Kondisi polarisasi opini ini menunjukkan perlunya pendekatan sistematis untuk mengukur dan menganalisis sentimen digital tersebut. Teknik text mining dengan fokus analisis sentimen terbukti efektif untuk memahami respons publik terhadap kebijakan [3]. Algoritma Naïve Bayes dipilih karena dikenal sederhana, cepat, dan memiliki akurasi yang baik pada penelitian-penelitian terkait kebijakan publik, seperti analisis sentimen kenaikan harga BBM oleh Desiani (2022) [4].

Penelitian terkait sebelumnya oleh Chidayat et al. (2025) menerapkan algoritma Naïve Bayes untuk menganalisis sentimen terkait skandal BBM oplosan di platform X, dengan nilai recall 75% meskipun akurasi keseluruhan terbatas pada 57%[5]. Terdapat kesenjangan penelitian (research gap) berupa minimnya studi yang secara spesifik meneliti respons publik terhadap kebijakan BBM E10, yang

memiliki karakteristik dan dampak berbeda dari isu BBM lainnya.

Tujuan penelitian ini adalah: (1) mengetahui sentimen pengguna media sosial X terhadap kebijakan BBM E10 melalui klasifikasi tweet menjadi kategori positif, negatif, dan netral menggunakan algoritma Naïve Bayes; dan (2) mengevaluasi performa model berdasarkan tiga skenario pembagian data dengan metodologi KDD.

2. TINJAUAN PUSTAKA

2.1. Kebijakan BBM E10

Kebijakan adalah kegiatan yang ditetapkan oleh pemerintah dengan tujuan tertentu[6]. BBM E10 adalah jenis bahan bakar yang terdiri dari campuran bensin murni dengan 10% etanol atau alkohol yang dihasilkan dari bahan tanaman seperti tebu, singkong, atau jagung.

2.2. Media Sosial X

X, sebelumnya dikenal sebagai Twitter, adalah platform media sosial yang memungkinkan pengguna menyampaikan pendapat secara bebas [7]. Pertumbuhan X terus meningkat setiap waktu dan dimanfaatkan pengguna untuk menyampaikan berbagai opini, komentar, serta kritik terhadap isu yang sedang berkembang[8].

2.3. Analisis Sentimen

Analisis sentimen adalah teknik untuk mengidentifikasi perasaan yang diekspresikan dalam teks dan mengelompokkannya ke dalam kategori sentimen positif, negatif, atau netral[9]. Analisis sentimen merupakan aktivitas mencari opini masyarakat tentang suatu objek yang ingin diketahui [10].

2.4. Text Mining

Text mining adalah tahap analisis teks yang dilakukan secara otomatis oleh komputer untuk menggali informasi berkualitas dari suatu rangkaian teks. Tujuan text mining adalah menemukan pola pada data agar dapat digunakan untuk mencapai tujuan tertentu[11].

2.5. Knowledge Discovery in Database (KDD)

KDD adalah proses mencari informasi yang lebih berharga, mudah dipahami, dan baru dari penyimpanan data yang besar dan kompleks [12]. Tahapan KDD meliputi data selection, preprocessing, data transformation, data mining, dan evaluation[13].

2.6. Algoritma Naïve Bayes

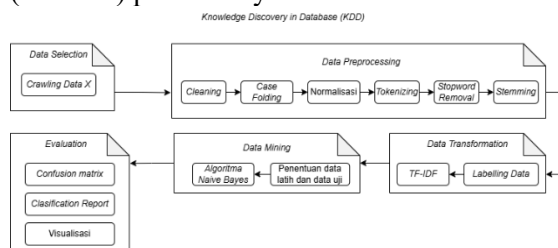
Naïve Bayes adalah algoritma klasifikasi yang menggunakan probabilitas sederhana berdasarkan Teorema Bayes dengan asumsi bahwa setiap kondisi bersifat saling tidak bergantung[2]. Kelebihan algoritma ini adalah sederhana namun mampu menghasilkan akurasi yang tinggi, bahkan dengan jumlah data yang relatif sedikit [14].

2.7. Confusion Matrix

Confusion matrix adalah tabel yang menunjukkan jumlah data uji yang diklasifikasikan dengan benar dan salah, digunakan untuk menghitung akurasi, presisi, recall, dan F1-Score[1].

3. METODE PENELITIAN

Metodologi Knowledge Discovery in Database (KDD) digunakan sebagai metode penelitian. Objek penelitian adalah komentar masyarakat seputar kebijakan BBM E10 di platform X yang dikumpulkan melalui crawling menggunakan Python Google Colab dengan kata kunci "etanol" dan "BBM E10" pada rentang waktu September 2025 hingga Januari 2026. Penggunaan metode KDD pada penelitian ini dikerjakan tanpa iterasi kembali ke fase pertama, sehingga setelah fase kelima (evaluasi) proses dinyatakan selesai.



Gambar 3. 1 Proses KDD

3.1. Data Selection

Tahap pertama adalah pengumpulan data relevan dari media sosial X melalui teknik crawling menggunakan API Twitter. Data yang

dikumpulkan adalah tweet berbahasa Indonesia yang mengandung kata kunci "E10" dan "etanol" dalam rentang waktu 1 September 2025 hingga 30 Januari 2026.

3.2. Preprocessing

Preprocessing bertujuan untuk mentransformasi data mentah menjadi struktur yang optimal untuk analisis. Tahapan yang dilakukan secara berurutan adalah *cleansing*, *case folding*, normalisasi, *Tokenizing*, *Stopword removal*.

3.3. Transformation

Tahap transformasi meliputi dua proses. Pertama, pelabelan data secara hybrid, kombinasi pelabelan otomatis berbasis leksikon dan verifikasi manual oleh Ahli Bahasa Indonesia untuk mengklasifikasikan tweet ke dalam kelas positif, negatif, dan netral. Kedua, TF-IDF (*Term Frequency-Inverse Document Frequency*) untuk mengubah data yang telah dipraproses menjadi representasi numerik.

3.4. Data Mining

Pada tahap ini data yang telah ditransformasi diklasifikasikan menggunakan algoritma Multinomial Naïve Bayes. Dipilih tiga skenario pembagian data latih dan data uji, yaitu 70:30, 80:20, dan 90:10, untuk membandingkan konsistensi dan performa model pada kondisi yang berbeda.

3.5. Evaluasi

Evaluasi dilakukan menggunakan confusion matrix dan classification report untuk mengukur performa klasifikasi per kelas sentimen. Metrik yang dihitung meliputi akurasi, presisi, recall, dan F1-Score. Hasil klasifikasi juga divisualisasikan dalam bentuk word cloud untuk menampilkan kata-kata yang paling sering muncul pada masing-masing kelas sentimen.

4. HASIL DAN PEMBAHASAN

4.1. Data Selection

Proses crawling menggunakan API Twitter menghasilkan 4.847 data tweet dalam rentang waktu 1 September 2025 hingga 28 Januari 2026.

Hasil *Crawling* data

id	user	time	text	location
1	@rpurwa	2020-03-01 14:45:02	Berita bahwa kata-kata ini akan dicampur etanol broo lu kenapa dah kaya anak kecil pemikirannya	
2	@innovacommunity	2020-03-01 14:45:02	Semua BBM di beredar di dunia pada akhirnya akan dicampur etanol broo lu kenapa dah kaya anak kecil pemikirannya	
3	@innovacommunity	2020-03-01 14:45:02	Semua BBM di beredar di dunia pada akhirnya akan dicampur etanol broo lu kenapa dah kaya anak kecil pemikirannya	
4	@innovacommunity	2020-03-01 14:45:02	Semua BBM di beredar di dunia pada akhirnya akan dicampur etanol broo lu kenapa dah kaya anak kecil pemikirannya	

Gambar 4.1 Hasil Crawling data

Setelah pengecekan dan penghapusan data duplikat menggunakan fungsi `df.drop_duplicates()`, jumlah data berkurang menjadi 2.385 entri. Dari keseluruhan atribut dataset, hanya atribut `full_text` yang digunakan dalam analisis sentimen.

4.2. Preprocessing

Selanjutnya dilakukan tahap preprocessing data yang bertujuan untuk membersihkan dan mempersiapkan data sebelum proses analisis. Tahap ini dilakukan secara bertahap, dimulai dari cleaning data, hingga stemming untuk mengembalikan setiap kata ke bentuk dasarnya. Proses preprocessing bertujuan meningkatkan kualitas data sehingga hasil analisis yang diperoleh menjadi lebih akurat dan optimal.

4.2.1. Cleaning

Pada proses cleaning, seluruh kolom dataset tidak memiliki nilai kosong (missing values), dan setelah penghapusan duplikat data berkurang dari 4.847 menjadi 2.385 entri. Dilanjut penghapusan symbol atau emotikon yang tidak relevan.

Tabel 4. 1 Hasil Cleaning

Sebelum Cleaning	Sesudah Cleaning
@rpurwa @innovacommunity Semua BBM di beredar di dunia pada akhirnya akan dicampur etanol broo lu kenapa dah kaya anak kecil pemikirannya	Semua BBM di beredar di dunia pada akhirnya akan dicampur etanol broo lu kenapa dah kaya anak kecil pemikirannya

4.2.2. Case Folding

Seluruh teks kemudian diubah menjadi huruf kecil pada tahap case folding.

Tabel 4. 2 Hasil Case Folding

Sebelum Case Folding	Sesudah Case Folding
Semua BBM di beredar di dunia pada akhirnya akan dicampur etanol broo lu kenapa dah	semua BBM di beredar di dunia pada akhirnya akan dicampur etanol broo lu kenapa dah

kaya	anak	kecil	kaya	anak	kecil
pemikirannya			pemikirannya		

4.2.3. Normalisasi

Normalisasi dilakukan menggunakan kamus yang memuat pasangan kata baku dan tidak baku.

Tabel 4. 3 Hasil Normalisasi

Sebelum Normalisasi	Sesudah Normalisasi
semua BBM di beredar di dunia pada akhirnya akan dicampur etanol broo lu kenapa dah kaya anak kecil pemikirannya	semua BBM di beredar di dunia pada akhirnya akan dicampur etanol broo kamu kenapa sudah kayak anak kecil pemikirannya

4.2.4. Tokenizing

Kalimat tweet dipecah menjadi token pada tahap tokenizing.

Tabel 4. 4 Hasil Tokenizing

Sebelum Tokenizing	Sesudah Tokenizing
semua BBM di beredar di dunia pada akhirnya akan dicampur etanol broo kamu kenapa sudah kayak anak kecil pemikirannya	['semua', 'BBM', 'di', 'beredar', 'di', 'dunia', 'akan', 'dicampur', 'etanol', 'broo', 'kamu', 'kenapa', 'sudah', 'kayak', 'anak', 'kecil', 'pemikirannya']

4.2.5. Stopword Removal

Pada tahap ini, kata yang dianggap tidak relevan akan dihapus.

Tabel 4. 5 Hasil Stopword Removal

Sebelum Stopword Removal	Sesudah Stopword Removal
['semua', 'BBM', 'di', 'beredar', 'di', 'dunia', 'akan', 'dicampur', 'etanol', 'broo', 'kamu', 'kenapa', 'sudah', 'kayak', 'anak', 'kecil', 'pemikirannya']	['BBM', 'beredar', 'dunia', 'dicampur', 'etanol', 'broo', 'kayak', 'anak', 'pemikirannya']

4.2.6. Stemming

dan terakhir kata-kata direduksi ke bentuk dasarnya melalui stemming menggunakan library Sastrawi.

Tabel 4. 6 Hasil Stemming

Sebelum Stemming	Sesudah Stemming
['BBM', 'beredar', 'dunia', 'dicampur', 'etanol', 'broo', 'kayak', 'anak', 'pemikirannya']	BBM edar dunia campur etanol broo kayak anak pikir

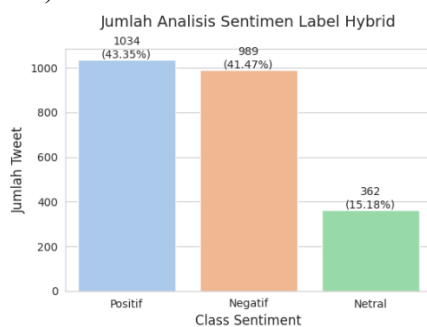
Setelah melewati seluruh tahapan preprocessing, dihasilkan dataset yang lebih terstruktur dan siap untuk dianalisis

Full_Text	cleaning	tokenizing	normalisasi	tokenisasi	stopword removal	stemming_data
Baru nalar kaku di-ku-Indonesia grup BSM	Baru nalar kaku di-ku-Indonesia grup BSM	Baru nalar kaku di-ku-Indonesia grup BSM	Baru nalar kaku di-ku-Indonesia grup BSM	[Baru, nalar, kaku, di, ku, Indonesia, grup, BSM]	[Baru, nalar, kaku, di, ku, Indonesia, grup, BSM]	Baru nalar kaku di-ku-Indonesia grup BSM
Tanah kaku di-ku-Indonesia grup BSM	Tanah kaku di-ku-Indonesia grup BSM	Tanah kaku di-ku-Indonesia grup BSM	Tanah kaku di-ku-Indonesia grup BSM	[Tanah, kaku, di, ku, Indonesia, grup, BSM]	[Tanah, kaku, di, ku, Indonesia, grup, BSM]	Tanah kaku di-ku-Indonesia grup BSM
Baru nalar kaku di-ku-Indonesia grup BSM	Baru nalar kaku di-ku-Indonesia grup BSM	Baru nalar kaku di-ku-Indonesia grup BSM	Baru nalar kaku di-ku-Indonesia grup BSM	[Baru, nalar, kaku, di, ku, Indonesia, grup, BSM]	[Baru, nalar, kaku, di, ku, Indonesia, grup, BSM]	Baru nalar kaku di-ku-Indonesia grup BSM
Baru nalar kaku di-ku-Indonesia grup BSM	Baru nalar kaku di-ku-Indonesia grup BSM	Baru nalar kaku di-ku-Indonesia grup BSM	Baru nalar kaku di-ku-Indonesia grup BSM	[Baru, nalar, kaku, di, ku, Indonesia, grup, BSM]	[Baru, nalar, kaku, di, ku, Indonesia, grup, BSM]	Baru nalar kaku di-ku-Indonesia grup BSM
Baru nalar kaku di-ku-Indonesia grup BSM	Baru nalar kaku di-ku-Indonesia grup BSM	Baru nalar kaku di-ku-Indonesia grup BSM	Baru nalar kaku di-ku-Indonesia grup BSM	[Baru, nalar, kaku, di, ku, Indonesia, grup, BSM]	[Baru, nalar, kaku, di, ku, Indonesia, grup, BSM]	Baru nalar kaku di-ku-Indonesia grup BSM
Baru nalar kaku di-ku-Indonesia grup BSM	Baru nalar kaku di-ku-Indonesia grup BSM	Baru nalar kaku di-ku-Indonesia grup BSM	Baru nalar kaku di-ku-Indonesia grup BSM	[Baru, nalar, kaku, di, ku, Indonesia, grup, BSM]	[Baru, nalar, kaku, di, ku, Indonesia, grup, BSM]	Baru nalar kaku di-ku-Indonesia grup BSM
Baru nalar kaku di-ku-Indonesia grup BSM	Baru nalar kaku di-ku-Indonesia grup BSM	Baru nalar kaku di-ku-Indonesia grup BSM	Baru nalar kaku di-ku-Indonesia grup BSM	[Baru, nalar, kaku, di, ku, Indonesia, grup, BSM]	[Baru, nalar, kaku, di, ku, Indonesia, grup, BSM]	Baru nalar kaku di-ku-Indonesia grup BSM
Baru nalar kaku di-ku-Indonesia grup BSM	Baru nalar kaku di-ku-Indonesia grup BSM	Baru nalar kaku di-ku-Indonesia grup BSM	Baru nalar kaku di-ku-Indonesia grup BSM	[Baru, nalar, kaku, di, ku, Indonesia, grup, BSM]	[Baru, nalar, kaku, di, ku, Indonesia, grup, BSM]	Baru nalar kaku di-ku-Indonesia grup BSM
Baru nalar kaku di-ku-Indonesia grup BSM	Baru nalar kaku di-ku-Indonesia grup BSM	Baru nalar kaku di-ku-Indonesia grup BSM	Baru nalar kaku di-ku-Indonesia grup BSM	[Baru, nalar, kaku, di, ku, Indonesia, grup, BSM]	[Baru, nalar, kaku, di, ku, Indonesia, grup, BSM]	Baru nalar kaku di-ku-Indonesia grup BSM

Gambar 4. 1 Hasil Preprocessing

4.3. Data Transformation

Pelabelan dilakukan secara hybrid. Pelabelan otomatis berbasis leksikon menghasilkan distribusi awal: 1.187 positif, 742 negatif, dan 456 netral. Selanjutnya verifikasi manual oleh Ahli Bahasa Indonesia untuk mengoreksi ketidaktepatan label akibat penggunaan bahasa gaul, sarkasme, maupun konteks kalimat. Hasil akhir setelah pelabelan hybrid menghasilkan distribusi: 1.034 positif (43,35%), 989 negatif (41,47%), dan 362 netral (15,18%).



Gambar 4. 2 Hasil Pelabelan

Transformasi fitur menggunakan TfidfVectorizer menghasilkan matriks berukuran 2.385 baris $\times$ 4.634 kolom, yang berarti terdapat 2.385 data tweet dan 4.634 kata unik sebagai fitur. Sepuluh kata dengan skor TF-IDF tertinggi adalah etanol (162,909), bbm (155,607), e (128,462), campur (118,756), kendaraan (71,183), Pertamina (61,723), mesin (59,008), bahlil (58,029), kurang (56,846), dan bensin (56,791).

***	10 Kata Teratas Berdasarkan Skor TF-IDF:
etanol	162.909267
bbm	155.606941
e	128.461566
campur	118.756252
kendaraan	71.182833
pertamina	61.723177
mesin	59.008449
bahlil	58.029197
kurang	56.845987
bensin	56.790677
dtype:	float64

Gambar 4. 3 kata dengan skor TF-IDF tertinggi

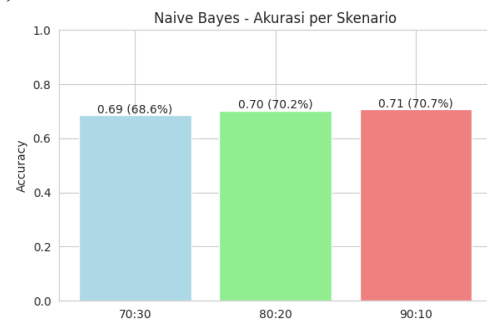
4.4. Data Mining

Pembagian data pada tiga skenario disajikan pada Tabel 4.7.

Tabel 4. 7 Pembagian Data Tiga Skenario

Skenario	Pembagian Data	Data Latih	Data Uji
Skenario 1	70:30	1669	716
Skenario 2	80:20	1908	477
Skenario 3	90:10	2146	239

Model klasifikasi Multinomial Naïve Bayes diimplementasikan menggunakan Scikit-Learn. Hasil akurasi awal dari ketiga skenario menunjukkan tren peningkatan: skenario 70:30 menghasilkan akurasi 68,6%, skenario 80:20 sebesar 70,2%, dan skenario 90:10 sebesar 70,7%.



Gambar 4. 4 Hasil akurasi

4.5. Evaluasi

Evaluasi model dilakukan menggunakan confusion matrix untuk mengukur performa klasifikasi secara rinci per kelas sentimen. Metrik yang dihitung meliputi akurasi, presisi, recall, dan F1-Score.

4.5.1. Skenario 70:30

		Confusion Matrix - Skenario 70:30		
Actual	Negatif	237	9	51
	Netral	40	5	64
	Positif	60	1	249
		Predicted		
		Negatif	Netral	Positif

Gambar 4. 5 Confusion Matrix Skenario 70:30

Model diuji terhadap 716 data uji. Kelas negatif berhasil diprediksi benar sebanyak 237 data, namun salah mengklasifikasikan 9 data sebagai netral dan 51 data sebagai positif. Kelas netral hanya diprediksi benar 5 data, sementara

40 data diprediksi sebagai negatif dan 64 data sebagai positif. Kelas positif berhasil diprediksi benar sebanyak 249 data, dengan 60 data salah diklasifikasikan sebagai negatif. Hasil perhitungan metrik disajikan pada Tabel 4.8.

Tabel 4. 8 Metrik Evaluasi Skenario 70:30

Kelas	TP	FP	FN	Precision	Recall	F1-Score
Negtif	237	100	60	0,703	0,798	0,748
Netral	5	10	104	0,333	0,046	0,081
Positif	249	115	61	0,684	0,803	0,739
Accuracy						0,686

4.5.2. Skenario 80:20

Confusion Matrix - Skenario 80:20

Actual	Negtif	162	3	33
	Netral	28	4	40
	Positif	37	1	169
		Negtif	Netral Predicted	Positif

Gambar 4. 6 Confusion Matrix Skenario 80:20

Model diuji terhadap 477 data uji. Kelas negatif berhasil diprediksi benar sebanyak 162 data, kelas netral hanya 4 data, dan kelas positif 169 data. Terjadi peningkatan performa pada kelas negatif dan positif dibandingkan skenario sebelumnya, sebagaimana ditunjukkan Tabel 4.9.

Tabel 4. 9 Metrik Evaluasi Skenario 80:20

Kelas	TP	FP	FN	Precision	Recall	F1-Score
Negtif	162	65	36	0,713	0,816	0,762
Netral	4	4	68	0,500	0,056	0,100
Positif	169	73	38	0,697	0,811	0,753
Accuracy						0,702

4.5.3. Skenario 90:10

Confusion Matrix - Skenario 90:10

Actual	Negtif	83	1	15
	Netral	14	2	20
	Positif	20	0	84
		Negtif	Netral Predicted	Positif

Gambar 4. 7 Confusion Matrix Skenario 90:10

Model diuji terhadap 239 data uji. Kelas negatif berhasil diprediksi benar sebanyak 83 data, kelas netral hanya 2 data, dan kelas positif 84 data. Skenario ini menghasilkan performa terbaik sebagaimana ditunjukkan Tabel 4.10.

Tabel 4. 10 Metrik Evaluasi Skenario 90:10

Kelas	TP	FP	FN	Precision	Recall	F1-Score
Negtif	83	34	16	0,714	0,836	0,769
Netral	2	1	34	0,667	0,056	0,103
Positif	84	35	20	0,720	0,808	0,753
Accuracy						0,707

4.5.4. Perbandingan Ketiga Skenario.

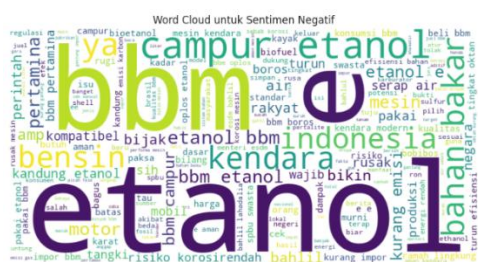
Ringkasan perbandingan performa ketiga skenario disajikan pada Tabel 4.11

Tabel 4. 11 Perbandingan Performa Ketiga Skenario

Skenario	Akurasi	Presisi	Recall	F1-Score
70:30	68,6%	57%	55%	52%
80:20	70,2%	64%	56%	54%
90:10	70,7%	69%	57%	54%

Skenario 90:10 menghasilkan performa terbaik pada seluruh metrik evaluasi. Tren peningkatan yang konsisten seiring bertambahnya proporsi data latih menunjukkan bahwa model Naïve Bayes bergantung pada ketersediaan data latih yang memadai.

4.5.5. Visualisasi Word Cloud.



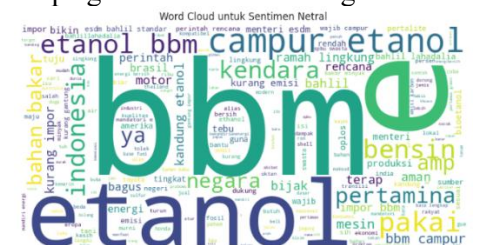
Gambar 4. 8 Word Cloud Sentimen Negatif

Pada kelas negatif, kata dominan adalah etanol, BBM, bahan bakar, korosi, rusak, karat, dan boros, mengindikasikan kekhawatiran terhadap dampak teknis etanol pada mesin kendaraan.



Gambar 4. 9 Word Cloud Sentimen Positif

Pada kelas positif, kata dominan adalah etanol, BBM, dukung, mandiri energi, ramah lingkungan, dan solusi, menunjukkan dukungan terhadap agenda ketahanan energi nasional.



Gambar 4. 10 Word Cloud Sentimen Netral

Pada kelas netral, kata dominan bersifat informatif seperti rencana, standar, wajib, dan esdm.

4.6. Pembahasan

Hasil analisis sentimen terhadap 2.385 tweet menunjukkan bahwa kelas positif mendominasi dengan 43,35%, diikuti kelas negatif 41,47%, dan netral 15,18%. Selisih kelas positif dan negatif yang hanya sekitar 1,88% mengonfirmasi adanya polarisasi opini yang hampir berimbang di kalangan pengguna X, yang mengindikasikan bahwa kebijakan BBM E10 merupakan isu kontroversial. Kondisi ini menjadi sinyal bagi pemangku kepentingan bahwa sosialisasi dan edukasi

publik mengenai manfaat serta keamanan BBM E10 perlu diperkuat.

Penerapan Multinomial Naïve Bayes terbukti cukup efektif mengklasifikasikan sentimen meskipun data tweet bersifat noisy dan menggunakan bahasa tidak baku. Tingginya presisi pada kelas negatif dan positif membuktikan bahwa pembobotan TF-IDF berhasil menonjolkan kata-kata kunci yang memiliki daya pembeda kuat antar kelas.

Performa rendah pada kelas netral (F1-Score tidak melebihi 0,103) di seluruh skenario disebabkan dua faktor utama: (1) ketidakseimbangan distribusi data, di mana kelas netral hanya mencakup 15,18% dari total data; dan (2) karakteristik linguistik komentar netral yang ambigu dan memiliki kemiripan leksikal dengan kedua kelas lainnya. Nilai F1-Score yang lebih rendah dibandingkan akurasi pada seluruh skenario mengonfirmasi adanya pengaruh class imbalance ini.

5. KESIMPULAN

Berdasarkan hasil penelitian analisis sentimen pengguna Platform X terhadap kebijakan BBM E10 menggunakan algoritma Naïve Bayes dengan pendekatan KDD, dapat disimpulkan:

- a. Hasil analisis sentimen terhadap 2.385 tweet menunjukkan bahwa sentimen masyarakat didominasi kelas positif sebanyak 1.034 tweet (43,35%), diikuti kelas negatif 989 tweet (41,47%), dan netral 362 tweet (15,18%). Masyarakat secara umum memberikan respons yang beragam namun cenderung positif terhadap kebijakan campuran etanol 10% pada bahan bakar.
- b. Evaluasi performa model menggunakan tiga skenario pembagian data menunjukkan bahwa skenario 90:10 menghasilkan performa terbaik dengan akurasi 70,7%, presisi 69%, recall 57%, dan F1-Score 54%. Skenario 80:20 menghasilkan akurasi 70,2%, sedangkan skenario 70:30 menghasilkan akurasi terendah 68,6%. Semakin besar proporsi data latih, semakin baik kemampuan model dalam melakukan klasifikasi. Rendahnya F1-Score dibandingkan akurasi mengonfirmasi pengaruh ketidakseimbangan distribusi data, khususnya pada kelas netral.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan dan kontribusi terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] A. R. Hanum *et al.*, “MENDETEKSI BERITA HOAKS PERFORMANCE ANALYSIS OF THE BERT TEXT CLASSIFICATION ALGORITHM,” vol. 11, no. 3, pp. 537–546, 2024, doi: 10.25126/jtiik938093.
- [2] A. Pebdika *et al.*, “KLASIFIKASI MENGGUNAKAN METODE NAIVE BAYES UNTUK MENENTUKAN CALON PENERIMA PIP,” vol. 7, no. 1, pp. 452–458, 2023.
- [3] A. Laksono, F. A. Rahmadani, and T. Informatika, “Analisis Sentimen Opini Publik di platform X (Twitter) terhadap Bencana Banjir Sumatra,” vol. 5, pp. 689–698, 2026.
- [4] A. Desiani, “PERBANDINGAN IMPLEMENTASI ALGORITMA NAÏVE BAYES DAN K-NEAREST NEIGHBOR PADA KLASIFIKASI,” vol. 7, no. 2, pp. 104–110, 2022.
- [5] A. T. Chidayat, D. F. Shiddieq, and D. Nurhayati, “Public Sentiment Analysis on Platform X Following the Mixed Fuel Scandal Using the Naïve Bayes Algorithm Analisis Sentimen Publik di Platform X Pasca Skandal Bahan Bakar Minyak Oplosan Menggunakan Algoritma Naïve Bayes,” vol. 5, no. October, pp. 1220–1230, 2025.
- [6] T. Tuamnggor and H. Sazali, “Penggunaan Hashtag # KaburAjaDulu dalam Pembentukan Algoritma Kritik Publik di Media Baru Abstrak,” vol. 6, no. 2, pp. 1353–1362, 2025.
- [7] B. Atau, P. Yang, A. E. Wulanda, M. Nur, and A. Abdullah, “NUSANTARA : Jurnal Ilmu Pengetahuan Sosial,” vol. 12, no. 5, pp. 1847–1853, 2025.
- [8] P. A. Nugroho, I. Fenriana, R. Arijanto, and M. Kom, “IMPLEMENTASI DEEP LEARNING MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK (CNN) PADA EKSPRESI MANUSIA,” vol. 1, 2020.
- [9] A. Subhan, E. Sedyono, and O. Dwi, “Analisis Sentimen Berbasis Ontologi di Level Kalimat untuk Mengukur Persepsi Produk,” vol. 02, pp. 84–97, 2015.
- [10] D. Nurmawati, C. P. Gusti, E. C. Hernanto, F. Ananda, M. A. Hutagalung, and N. Hikmah, “Analisis Sentimen untuk Melihat Respon Masyarakat Terhadap Vaksin Pfizer,” vol. 4, no. 2, pp. 77–80, 2023, doi: 10.37802/joti.v4i2.322.
- [11] A. Aziza and S. Rani, “Klasifikasi Sentimen Radikalisme dalam Konten Dakwah Radikal Indonesia melalui Media Sosial Twitter dengan Menggunakan Analisis Sentimen dan Text Mining,” vol. 14, no. 3, 2023.
- [12] F. Ardiansyah and I. W. Siadi, “Klasifikasi Customer Relationship Management Menggunakan Dataset KDD Cup 2009 dengan Teknik Reduksi Dimensi Classification of Customer Relationship Management using KDD Cup 2009 Dataset with Dimension Reduction Technique,” vol. 11, no. 28, 2022, doi: 10.34010/komputika.v11i2.6498.
- [13] J. Sains, A. Galang, W. Maharani, and A. Wijaya, “Penerapan Data Mining untuk Memprediksi Kelulusan Mahasiswa Menggunakan Decision Tree Jurnal Sains dan Teknologi,” vol. 02, no. 02, pp. 91–96, 2025.
- [14] A. S. Putra, O. Marshella, and F. Agus, “Evaluasi Kinerja Algoritma Berbasis Decision Tree pada Dataset Skala Kecil : Studi Perbandingan Analisis Komparatif,” vol. 8, no. 1, pp. 6–11, 2026.