Analisis Sentimen Komentar YouTube Pernyataan Hilirisasi oleh Gibran Rakabuming Raka Menggunakan Algoritma Naïve Bayes [Sentiment Analysis of YouTube Comments on Gibran Rakabuming Raka's Statement on Downstream Industrialization Using the Naïve Bayes Algorithm]

Arif Rahman Hakim[1)], Sumarno[*,2)]

[1)]Program Studi Teknik Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia
[2)] Program Studi Teknik Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia
*Email Penulis Korespondensi: sumarno@umsida.ac.id

Abstract. *Social media has become an important platform for expressing public opinion, making YouTube comments a valuable source for sentiment analysis. This study analyzes public sentiment toward the downstream industry (hilirisasi) policy statements associated with Gibran Rakabuming Raka by comparing Multinomial, Bernoulli, and Gaussian Naive Bayes classifiers. Comments were collected via the YouTube Data API v3 from nine videos on Gibran's official channel and national news channels, filtered for topical relevance using zero-shot classification, and labeled using a context-conditioned IndoBERT sentiment model, yielding 1,080 labeled comments (665 positive, 415 negative). Preprocessing included case folding, cleaning, tokenization, normalization, stopword removal, stemming, and TF-IDF weighting, followed by twelve experimental scenarios combining two train-test split ratios with and without SMOTE. Multinomial Naive Bayes achieved the highest accuracy and F1-score, 82.87% and 87.11% respectively, on an 80:20 split with SMOTE, outperforming the Bernoulli and Gaussian variants.*
Keywords - *sentiment analysis; Bernoulli Naive Bayes; Multinomial Naive Bayes; Gaussian Naive Bayes; SMOTE*

Abstrak. *Media sosial menjadi sarana penting bagi masyarakat untuk menyampaikan opini, sehingga komentar YouTube merupakan sumber data yang potensial untuk analisis sentimen. Penelitian ini menganalisis sentimen komentar terhadap pernyataan kebijakan hilirisasi yang berkaitan dengan Gibran Rakabuming Raka dengan membandingkan kinerja Multinomial, Bernoulli, dan Gaussian Naive Bayes. Komentar dikumpulkan menggunakan YouTube Data API v3 dari sembilan video pada kanal resmi Gibran Rakabuming Raka dan kanal berita nasional, disaring berdasarkan relevansi topik menggunakan zero-shot classification, serta diberi label menggunakan model sentimen IndoBERT context-conditioned, sehingga diperoleh 1.080 komentar berlabel (665 positif, 415 negatif). Tahap prapemrosesan meliputi case folding, cleaning, tokenization, normalization, stopword removal, stemming, dan pembobotan TF-IDF, dilanjutkan dengan 12 skenario eksperimen yang mengombinasikan dua rasio pembagian data dengan dan tanpa SMOTE. Multinomial Naive Bayes memperoleh accuracy dan F1-score tertinggi, masing-masing sebesar 82,87% dan 87,11%, pada rasio 80:20 dengan SMOTE, mengungguli varian Bernoulli dan Gaussian.*
Kata Kunci - *analisis sentimen; Bernoulli Naive Bayes; Multinomial Naive Bayes; Gaussian Naive Bayes; SMOTE*

I. PENDAHULUAN

Perkembangan teknologi informasi pada era Society 5.0 telah mengubah pola komunikasi masyarakat melalui pemanfaatan media sosial sebagai sarana penyebaran informasi sekaligus ruang interaksi publik. Platform media sosial tidak lagi hanya digunakan untuk berbagi konten, tetapi juga menjadi wadah bagi masyarakat dalam menyampaikan opini, kritik, maupun dukungan terhadap berbagai isu sosial, ekonomi, dan politik [1]. Salah satu platform yang memiliki karakteristik tersebut adalah YouTube karena menyediakan fitur komentar yang memungkinkan pengguna memberikan tanggapan secara langsung terhadap suatu konten [2], [3]. Besarnya volume komentar yang dihasilkan menjadikan YouTube sebagai sumber data yang relevan untuk menggambarkan persepsi publik, meskipun karakteristik teksnya cenderung tidak terstruktur, mengandung bahasa informal, singkatan, dan variasi kosakata yang menyulitkan proses analisis secara manual [4]. Oleh karena itu, analisis sentimen sebagai bagian dari *Natural Language Processing* (NLP) banyak dimanfaatkan untuk mengidentifikasi kecenderungan opini masyarakat secara otomatis [5], [6].

Hilirisasi merupakan salah satu kebijakan strategis dalam proses industrialisasi yang bertujuan meningkatkan nilai tambah sumber daya alam melalui pengolahan sebelum dipasarkan sehingga diharapkan mampu memperkuat daya saing industri nasional dan mendorong pertumbuhan ekonomi [7], [8]. Isu tersebut disampaikan secara aktif oleh Gibran Rakabuming Raka melalui kanal YouTube resminya pada sektor umum, kelautan, pertanian, dan digital, serta turut diberitakan dan direspons oleh sejumlah kanal berita nasional, termasuk tanggapan yang bersifat kritis terhadap kebijakan tersebut. Beragam tanggapan yang disampaikan masyarakat melalui kolom komentar mencerminkan

Copyright © Universitas Muhammadiyah Sidoarjo. This preprint is protected by copyright held by Universitas Muhammadiyah Sidoarjo and is distributed under the Creative Commons Attribution License (CC BY). Users may share, distribute, or reproduce the work as long as the original author(s) and copyright holder are credited, and the preprint server is cited per academic standards.
Authors retain the right to publish their work in academic journals where copyright remains with them. Any use, distribution, or reproduction that does not comply with these terms is not permitted.

persepsi publik terhadap kebijakan tersebut. Volume komentar yang besar serta keragaman opini yang dihasilkan menjadikan komentar YouTube sebagai sumber data yang relevan untuk mengevaluasi respons publik secara sistematis melalui pendekatan analisis sentimen.

Berbagai penelitian terdahulu menunjukkan bahwa algoritma Naive Bayes merupakan salah satu metode klasifikasi yang efektif untuk analisis sentimen berbahasa Indonesia karena memiliki proses komputasi yang sederhana, efisien, dan mampu menghasilkan performa yang baik pada data teks [9], [10], [11]. Pada berbagai kasus analisis ulasan maupun opini di media sosial, Multinomial Naive Bayes sering dilaporkan memberikan akurasi yang tinggi karena memanfaatkan frekuensi kemunculan kata sebagai dasar klasifikasi [12]. Di sisi lain, Bernoulli Naive Bayes lebih berfokus pada keberadaan suatu fitur kata, sedangkan Gaussian Naive Bayes menggunakan asumsi distribusi normal pada data numerik sehingga ketiga varian tersebut memiliki karakteristik yang berbeda dalam memproses data teks [13], [14], [15].

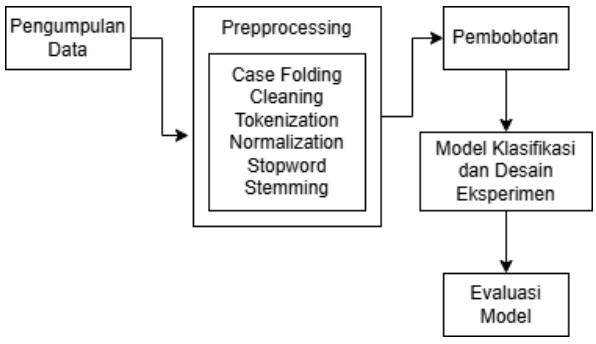
Hasil penelitian komparatif juga memperlihatkan bahwa performa setiap varian Naive Bayes sangat dipengaruhi oleh karakteristik dataset yang digunakan. Multinomial Naive Bayes umumnya menunjukkan hasil terbaik pada representasi teks berbasis frekuensi kata, sedangkan Bernoulli Naive Bayes dapat memberikan keseimbangan kinerja yang baik pada representasi fitur biner [12], [13]. Sementara itu, Gaussian Naive Bayes cenderung menghasilkan performa yang lebih rendah ketika diterapkan pada data teks dengan representasi TF-IDF yang bersifat *sparse* karena asumsi distribusi normal sulit dipenuhi [15]. Selain pemilihan algoritma, beberapa penelitian juga menunjukkan bahwa teknik *Synthetic Minority Oversampling Technique* (SMOTE) mampu meningkatkan kemampuan model dalam mengenali kelas minoritas pada kondisi dataset yang tidak seimbang [16], [17].

Meskipun demikian, masih terdapat beberapa celah penelitian yang perlu dikaji lebih lanjut. Pertama, penelitian analisis sentimen pada komentar YouTube berbahasa Indonesia yang membandingkan secara langsung ketiga varian Naive Bayes, yaitu Multinomial, Bernoulli, dan Gaussian, masih relatif terbatas, terutama pada data komentar yang didominasi bahasa informal, singkatan, dan kosakata *slang*. Kedua, evaluasi pengaruh SMOTE umumnya hanya dilakukan pada satu varian algoritma atau satu skenario pengujian sehingga belum memberikan gambaran yang komprehensif mengenai efektivitas teknik tersebut pada ketiga varian Naive Bayes secara bersamaan. Kondisi tersebut menunjukkan perlunya penelitian yang mengevaluasi kinerja masing-masing varian algoritma melalui beberapa konfigurasi eksperimen pada dataset komentar YouTube berbahasa Indonesia.

Berdasarkan latar belakang dan celah penelitian tersebut, penelitian ini bertujuan menganalisis sentimen komentar YouTube terhadap pernyataan hilirisasi yang disampaikan oleh Gibran Rakabuming Raka dengan membandingkan performa algoritma Multinomial Naive Bayes, Bernoulli Naive Bayes, dan Gaussian Naive Bayes. Penelitian menggunakan data komentar yang diperoleh melalui YouTube Data API v3, diproses menggunakan tahapan *preprocessing* dan pembobotan TF-IDF, kemudian dievaluasi melalui 12 skenario eksperimen yang mengombinasikan tiga varian Naive Bayes, dua rasio pembagian data, serta perlakuan dengan dan tanpa SMOTE. Hasil penelitian diharapkan dapat mengidentifikasi model yang memberikan performa terbaik sekaligus memberikan gambaran mengenai pengaruh karakteristik algoritma dan penerapan SMOTE terhadap klasifikasi sentimen pada komentar YouTube berbahasa Indonesia.

II. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk membandingkan performa tiga varian algoritma Naive Bayes dalam klasifikasi sentimen komentar YouTube terhadap pernyataan hilirisasi oleh Gibran Rakabuming Raka. Tahapan penelitian meliputi pengumpulan data, *preprocessing*, pembobotan fitur menggunakan TF-IDF, pembentukan model klasifikasi, pengujian melalui 12 skenario eksperimen, dan evaluasi performa model.



Gambar 1. Tahapan Penelitian

Copyright © Universitas Muhammadiyah Sidoarjo. This preprint is protected by copyright held by Universitas Muhammadiyah Sidoarjo and is distributed under the Creative Commons Attribution License (CC BY). Users may share, distribute, or reproduce the work as long as the original author(s) and copyright holder are credited, and the preprint server is cited per academic standards.
Authors retain the right to publish their work in academic journals where copyright remains with them. Any use, distribution, or reproduction that does not comply with these terms is not permitted.

A. Pengumpulan Data

Data diperoleh menggunakan YouTube Data API v3 dari sembilan video yang membahas pernyataan kebijakan hilirisasi oleh Gibran Rakabuming Raka, terdiri atas empat video dari kanal resmi Gibran Rakabuming Raka (mencakup topik hilirisasi umum, kelautan, pertanian, dan digital) serta lima video dari empat kanal berita nasional (Kompas.com, KOMPASTV JAWA TENGAH, KOMPASTV, dan METRO TV) yang memberitakan maupun menanggapi pernyataan tersebut. Rincian video sumber disajikan pada Tabel 1.

Tabel 1. Daftar Video Sumber Data

No	Video ID	Judul	Kanal
1	ZhL1ABVBqB8	Hilirisasi dan Masa Depan Indonesia	Gibran Rakabuming Raka
2	ULJErSG7F8Y	Dari Pesisir ke Dunia: Saatnya Produk Laut Indonesia Naik Kelas	Gibran Rakabuming Raka
3	72Ko6qkf9TQ	Hilirisasi Pertanian: Swasembada Pangan dan Petani Naik Kelas	Gibran Rakabuming Raka
4	9gTdNSPimhs	Bukan Lagi Hanya Pasar Digital, Indonesia Harus Jadi Produsen Digital	Gibran Rakabuming Raka
5	Q3VeyP2DOUU	Genjot Hilirisasi Digital, Gibran Libatkan Anak Muda yang Ahli Crypto	Kompas.com
6	yeSb-DV8i_4	Gibran Bingung Ada Pihak Anti Hilirisasi	KOMPASTV JAWA TENGAH
7	1D6EwnDnnD8	Cerita Wapres Gibran Banyak Ditertawakan soal Hilirisasi Kemenyan	KOMPASTV
8	XpkxdvrmbJ4	Gus Imin Kritik Keras Hilirisasi Ugal-ugalan Gibran	METRO TV
9	OTNUEqh6wo8	Gibran Merasa Disepelekan Setiap Keluarkan Kata "Hilirisasi"	Kompas.com

Sebelum tahap pelabelan sentimen, seluruh komentar terlebih dahulu disaring berdasarkan relevansi topik menggunakan model zero-shot classification ModernBERT-large-zeroshot-v2.0. Setiap komentar diklasifikasikan ke dalam kategori "membahas kebijakan hilirisasi" atau "tidak membahas kebijakan hilirisasi", dan komentar dinyatakan relevan apabila memperoleh skor keyakinan ≥ 0,75 pada kategori tersebut. Tahap ini menghasilkan 1.337 komentar yang relevan dengan topik hilirisasi dari total 26.891 komentar mentah.

Komentar yang relevan selanjutnya diberi label sentimen secara otomatis menggunakan model IndoBERT-Sentiment yang bersifat context-conditioned (apriandito/indobert-sentiment-classifier). Berbeda dengan model sentimen umum, model ini menerima pasangan konteks dan teks, sehingga sentimen yang diprediksi merupakan sentimen komentar secara spesifik terhadap konteks "kebijakan hilirisasi", bukan sentimen umum komentar. Proses pelabelan menghasilkan distribusi awal sebanyak 672 komentar positif, 430 komentar negatif, dan 235 komentar netral. Karena tahap klasifikasi Naive Bayes pada penelitian ini bersifat biner, komentar berlabel netral dibuang, dan komentar duplikat turut dihapus, sehingga diperoleh 1.080 komentar akhir yang terdiri atas 665 komentar positif dan 415 komentar negatif.

B. Preprocessing Data

Komentar yang telah dikumpulkan diproses melalui enam tahapan *preprocessing* agar dapat digunakan sebagai masukan algoritma klasifikasi, yaitu:

1. Case folding, mengubah seluruh huruf menjadi huruf kecil.
2. Cleaning, menghapus URL, *mention*, *hashtag*, angka, emoji, dan tanda baca.
3. Tokenization, memisahkan kalimat menjadi token kata.
4. Normalization, mengubah kata tidak baku menggunakan kamus kustom sebanyak 1.305 entri.
5. Stopword removal, menghapus kata yang tidak berkontribusi terhadap sentimen menggunakan kamus kustom sebanyak 770 kata.
6. Stemming, mengubah kata menjadi bentuk dasar menggunakan pustaka Sastrawi.

C. Pembobotan Fitur

Data hasil *preprocessing* direpresentasikan ke dalam bentuk numerik menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF). Metode TF-IDF digunakan untuk memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam suatu dokumen dan tingkat kepentingannya terhadap keseluruhan dokumen, sehingga kata yang lebih representatif memiliki bobot yang lebih tinggi. Representasi numerik ini selanjutnya digunakan sebagai masukan pada proses klasifikasi menggunakan algoritma Naive Bayes.

Nilai Term Frequency (TF) dihitung menggunakan Persamaan (1).

$$TF(t,d) = \frac{f_{t,d}}{\sum_k \ f_{k,d}} \qquad (1)$$

Copyright © Universitas Muhammadiyah Sidoarjo. This preprint is protected by copyright held by Universitas Muhammadiyah Sidoarjo and is distributed under the Creative Commons Attribution License (CC BY). Users may share, distribute, or reproduce the work as long as the original author(s) and copyright holder are credited, and the preprint server is cited per academic standards.
Authors retain the right to publish their work in academic journals where copyright remains with them. Any use, distribution, or reproduction that does not comply with these terms is not permitted.

dengan $f_{t,d}$merupakan frekuensi kemunculan term $t$pada dokumen d, sedangkan $\sum_k f_{k,d}$adalah jumlah seluruh term pada dokumen tersebut.

Nilai Inverse Document Frequency (IDF) dihitung menggunakan Persamaan (2).

$$IDF(t) = \log \frac{N}{df_t} \quad (2)$$

dengan Nmerupakan jumlah seluruh dokumen dan df_tadalah jumlah dokumen yang mengandung term t.

Bobot akhir diperoleh melalui perkalian nilai TF dan IDF sebagaimana ditunjukkan pada Persamaan (3).

$$TF\text{-}IDF(t,d) = TF(t,d) \times IDF(t) \quad (3)$$

Semakin tinggi nilai TF-IDF, semakin besar kontribusi suatu term dalam merepresentasikan isi dokumen sehingga fitur tersebut memiliki pengaruh yang lebih besar pada proses klasifikasi.

D. Model Klasifikasi dan Desain Eksperimen

Penelitian membandingkan tiga varian algoritma Naive Bayes, yaitu Multinomial Naive Bayes (MNB), Gaussian Naive Bayes (GNB), dan Bernoulli Naive Bayes (BNB). Secara umum, klasifikasi pada algoritma Naive Bayes didasarkan pada Teorema Bayes yang menghitung probabilitas suatu dokumen termasuk ke dalam kelas tertentu berdasarkan fitur yang dimiliki. Probabilitas posterior dihitung menggunakan Persamaan (4).

$$P(C_i \mid x) = \frac{P(x|C_i) \cdot P(C_i)}{P(x)} \quad (4)$$

dengan $P(C \mid X)$adalah probabilitas dokumen Xtermasuk ke dalam kelas C, $P(X \mid C)$adalah probabilitas kemunculan dokumen pada kelas C, $P(C)$merupakan probabilitas awal (prior) kelas, dan $P(X)$adalah probabilitas kemunculan dokumen.

Multinomial Naive Bayes mengklasifikasikan dokumen berdasarkan frekuensi kemunculan kata sehingga sesuai untuk representasi TF-IDF. Probabilitas Multinomial Naive Bayes dihitung menggunakan Persamaan (5).

$$P(x \mid C_i) = \prod_{k=1}^{n} P(x_k \mid C_i)^{f_i} \quad (5)$$

dengan $w_i$merupakan kata ke-i dan $f_i$adalah frekuensi kemunculan kata pada dokumen.

Bernoulli Naive Bayes menggunakan representasi biner yang hanya memperhatikan keberadaan atau ketidakberadaan suatu kata pada dokumen. Probabilitas Bernoulli Naive Bayes dihitung menggunakan Persamaan (6).

$$P(x \mid C_i) = \prod_{k=1}^{n} P(x_k \mid C_i)^{b_k}\ 1 - (P(x_k \mid C_i))^{(1-b_k)} \quad (6)$$

dengan $x_i = 1$apabila kata muncul pada dokumen dan $x_i = 0$apabila kata tidak muncul.

Sementara itu, Gaussian Naive Bayes mengasumsikan bahwa nilai fitur mengikuti distribusi normal sehingga turut dievaluasi untuk mengetahui performanya pada data teks hasil pembobotan TF-IDF. Probabilitas Gaussian Naive Bayes dihitung menggunakan Persamaan (7).

$$P(x_k \mid C_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x_k-\mu_i)^2}{2\sigma_i^2} \right) \quad (7)$$

dengan $\mu_C$merupakan nilai rata-rata fitur pada kelas C, sedangkan $\sigma_C$merupakan simpangan baku fitur pada kelas tersebut.

Untuk mengetahui pengaruh jenis algoritma, rasio pembagian data, dan penyeimbangan kelas, dilakukan pengujian terhadap 12 kombinasi eksperimen yang terdiri atas tiga algoritma, dua rasio *train-test split* (90:10 dan 80:20), serta dua perlakuan, yaitu tanpa SMOTE dan dengan SMOTE. Teknik *Synthetic Minority Oversampling Technique* (SMOTE) diterapkan hanya pada data latih untuk menyeimbangkan distribusi kelas pada skenario yang menggunakan SMOTE sehingga proses evaluasi pada data uji tetap mencerminkan kondisi data yang sebenarnya.

E. Evaluasi Model

Kinerja setiap model dievaluasi menggunakan confusion matrix yang terdiri atas empat komponen, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Berdasarkan confusion matrix tersebut, performa model diukur menggunakan empat metrik evaluasi, yaitu accuracy, precision, recall, dan F1-score. Keempat metrik dipilih karena mampu memberikan gambaran menyeluruh mengenai kemampuan model dalam mengklasifikasikan sentimen, terutama pada data yang memiliki distribusi kelas tidak seimbang.

Nilai accuracy dihitung menggunakan Persamaan (8).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

Accuracy menunjukkan proporsi keseluruhan prediksi yang berhasil diklasifikasikan dengan benar oleh model.

Nilai precision dihitung menggunakan Persamaan (9).

$$Precision = \frac{TP}{TP+FP} \quad (9)$$

Copyright © Universitas Muhammadiyah Sidoarjo. This preprint is protected by copyright held by Universitas Muhammadiyah Sidoarjo and is distributed under the Creative Commons Attribution License (CC BY). Users may share, distribute, or reproduce the work as long as the original author(s) and copyright holder are credited, and the preprint server is cited per academic standards.
Authors retain the right to publish their work in academic journals where copyright remains with them. Any use, distribution, or reproduction that does not comply with these terms is not permitted.

Precision menunjukkan kemampuan model dalam menghasilkan prediksi positif yang benar dari seluruh prediksi positif yang dihasilkan.

Nilai recall dihitung menggunakan Persamaan (10).

$$Recall = \frac{TP}{TP+FN} \quad (10)$$

Recall menunjukkan kemampuan model dalam mengidentifikasi seluruh data positif yang sebenarnya.

Nilai F1-score dihitung menggunakan Persamaan (11).

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (11)$$

F1-score merupakan rata-rata harmonis antara precision dan recall sehingga memberikan ukuran performa yang lebih seimbang, terutama pada kondisi distribusi kelas yang tidak seimbang.

Keempat metrik tersebut digunakan untuk membandingkan performa seluruh kombinasi eksperimen sehingga dapat diidentifikasi algoritma dan konfigurasi yang memberikan hasil klasifikasi terbaik berdasarkan nilai accuracy, precision, recall, dan F1-score.

III. HASIL DAN PEMBAHASAN

A. Statistik Dataset

Penelitian ini menggunakan data komentar yang diperoleh melalui YouTube Data API v3 dari sembilan video yang membahas pernyataan kebijakan hilirisasi oleh Gibran Rakabuming Raka, sebagaimana disajikan pada Tabel 1. Proses crawling dilakukan pada 22 Juli 2026 dan menghasilkan 26.891 komentar mentah.

Sebelum tahap pelabelan sentimen, seluruh komentar mentah disaring berdasarkan relevansi topiknya menggunakan model zero-shot classification. Dari 26.891 komentar mentah, hanya 1.337 komentar yang dinyatakan relevan dengan topik kebijakan hilirisasi berdasarkan ambang skor keyakinan sebesar 0,75. Tahap ini penting dilakukan mengingat komentar YouTube pada video bertopik figur publik cenderung memuat banyak komentar di luar substansi kebijakan yang dibahas, sehingga penyaringan relevansi topik diperlukan agar data yang dianalisis benar-benar mencerminkan opini terhadap kebijakan hilirisasi, bukan opini umum terhadap video maupun figur publik yang menyampaikannya.

Komentar yang relevan selanjutnya diberi label sentimen menggunakan model IndoBERT-Sentiment context-conditioned. Hasil pelabelan menunjukkan distribusi awal sebanyak 672 komentar positif, 430 komentar negatif, dan 235 komentar netral. Karena skema klasifikasi Naive Bayes pada penelitian ini bersifat biner, komentar berlabel netral tidak diikutsertakan pada tahap selanjutnya. Setelah komentar duplikat turut dihapus, diperoleh 1.080 komentar sebagai dataset akhir yang terdiri atas 665 komentar (61,6%) berlabel positif dan 415 komentar (38,4%) berlabel negatif.

Distribusi kelas pada dataset akhir menunjukkan adanya ketidakseimbangan data (*imbalanced dataset*) dengan kecenderungan yang didominasi oleh sentimen positif. Kondisi ini mengindikasikan bahwa di antara komentar yang secara spesifik membahas kebijakan hilirisasi, opini masyarakat cenderung lebih banyak memberikan dukungan atau tanggapan positif dibandingkan kritik atau penolakan. Meskipun demikian, tingkat ketidakseimbangan pada dataset ini (rasio sekitar 1,6:1) tergolong lebih ringan dibandingkan ketidakseimbangan yang umum dijumpai pada data komentar media sosial, sehingga penerapan SMOTE pada penelitian ini tetap dievaluasi untuk mengetahui pengaruhnya terhadap performa model meskipun tingkat ketidakseimbangan kelasnya tidak ekstrem.

B. Hasil Preprocessing Data

Sebanyak 1.080 komentar hasil filter biner diproses melalui enam tahapan *preprocessing*. Tahapan tersebut bertujuan menyeragamkan format teks, menghilangkan elemen yang tidak informatif, serta mengubah kata menjadi bentuk dasarnya sehingga representasi teks yang dihasilkan lebih konsisten untuk proses pembobotan TF-IDF pada tahap berikutnya. Contoh penerapan tahapan *preprocessing* terhadap salah satu komentar disajikan pada Tabel 2.

Tabel 2. Contoh Tahapan Preprocessing Komentar

Tahap	Hasil
Komentar asli	Programnya bagus bgt, semangat pak gibran!!
Case Folding	programnya bagus bgt, semangat pak gibran!!
Cleaning	programnya bagus bgt semangat pak gibran
Tokenization	['programnya', 'bagus', 'bgt', 'semangat', 'pak', 'gibran']
Normalization	['programnya', 'bagus', 'banget', 'semangat', 'pak', 'gibran']
Stopword Removal	['programnya', 'bagus', 'banget', 'semangat', 'gibran']
Stemming	program bagus banget semangat gibran

Berdasarkan Tabel 2, terlihat bahwa proses *cleaning* menghilangkan tanda baca, angka, serta karakter non-alfabetis dari komentar, sedangkan tahap *normalization* dan *stopword removal* mengubah kata tidak baku menjadi kata baku

Copyright © Universitas Muhammadiyah Sidoarjo. This preprint is protected by copyright held by Universitas Muhammadiyah Sidoarjo and is distributed under the Creative Commons Attribution License (CC BY). Users may share, distribute, or reproduce the work as long as the original author(s) and copyright holder are credited, and the preprint server is cited per academic standards.
Authors retain the right to publish their work in academic journals where copyright remains with them. Any use, distribution, or reproduction that does not comply with these terms is not permitted.

sekaligus membuang kata yang tidak berkontribusi terhadap sentimen. Tahap *stemming* pada akhirnya menyederhanakan setiap kata menjadi bentuk dasarnya menggunakan pustaka Sastrawi, sehingga variasi bentuk kata yang merujuk pada makna yang sama dapat direpresentasikan secara seragam pada tahap pembobotan fitur.

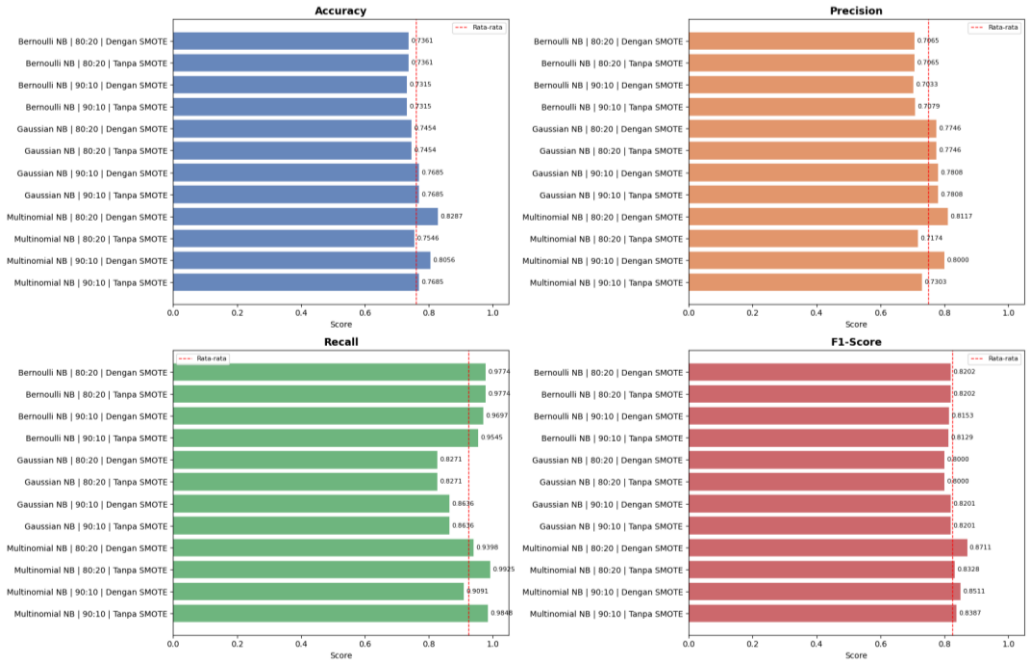
C. Hasil Evaluasi Model

Evaluasi dilakukan terhadap 12 kombinasi eksperimen yang terdiri atas tiga varian Naive Bayes, dua rasio pembagian data (*train-test split*), serta dua perlakuan, yaitu tanpa SMOTE dan dengan SMOTE. Setiap kombinasi dievaluasi menggunakan empat metrik, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil pengujian secara lengkap disajikan pada Tabel 3.

Tabel 3. Hasil Evaluasi Seluruh Kombinasi Eksperimen

No	Model	Split	SMOTE	Accuracy	Precision	Recall	F1-Score
1	Multinomial NB	90:10	Tanpa	76,85%	73,03%	98,48%	83,87%
2	Multinomial NB	90:10	Dengan	80,56%	80,00%	90,91%	85,11%
3	Multinomial NB	80:20	Tanpa	75,46%	71,74%	99,25%	83,28%
4	Multinomial NB	80:20	Dengan	82,87%	81,17%	93,98%	87,11%
5	Gaussian NB	90:10	Tanpa	76,85%	78,08%	86,36%	82,01%
6	Gaussian NB	90:10	Dengan	76,85%	78,08%	86,36%	82,01%
7	Gaussian NB	80:20	Tanpa	74,54%	77,46%	82,71%	80,00%
8	Gaussian NB	80:20	Dengan	74,54%	77,46%	82,71%	80,00%
9	Bernoulli NB	90:10	Tanpa	73,15%	70,79%	95,45%	81,29%
10	Bernoulli NB	90:10	Dengan	73,15%	70,33%	96,97%	81,53%
11	Bernoulli NB	80:20	Tanpa	73,61%	70,65%	97,74%	82,02%
12	Bernoulli NB	80:20	Dengan	73,61%	70,65%	97,74%	82,02%

Berdasarkan Tabel 3 terlihat bahwa setiap algoritma memberikan karakteristik performa yang berbeda pada setiap skenario pengujian, meskipun secara umum ketiga varian menghasilkan performa yang relatif berdekatan dibandingkan dengan pola yang lazim ditemukan pada dataset berukuran besar. Multinomial Naive Bayes memperoleh accuracy tertinggi sebesar 82,87% pada rasio pembagian data 80:20 dengan SMOTE, dan konfigurasi yang sama juga menghasilkan F1-score tertinggi sebesar 87,11%. Dengan demikian, satu konfigurasi yang sama unggul pada kedua metrik evaluasi utama, berbeda dengan pola umum pada dataset berskala besar yang kerap menghasilkan model accuracy terbaik dan model F1-score terbaik yang berbeda. Sementara itu, Gaussian Naive Bayes menghasilkan accuracy pada kisaran 74,54%–76,85%, dan Bernoulli Naive Bayes pada kisaran 73,15%–73,61%, sehingga selisih performa antaralgoritma pada penelitian ini tergolong tidak terlalu besar dibandingkan hasil pengujian pada dataset dengan jumlah data yang jauh lebih banyak. Untuk mempermudah interpretasi hasil pengujian, perbandingan performa seluruh kombinasi eksperimen divisualisasikan menggunakan diagram batang pada Gambar 2.



Gambar 2. **Perbandingan 12 Kombinasi**

Copyright © Universitas Muhammadiyah Sidoarjo. This preprint is protected by copyright held by Universitas Muhammadiyah Sidoarjo and is distributed under the Creative Commons Attribution License (CC BY). Users may share, distribute, or reproduce the work as long as the original author(s) and copyright holder are credited, and the preprint server is cited per academic standards.
Authors retain the right to publish their work in academic journals where copyright remains with them. Any use, distribution, or reproduction that does not comply with these terms is not permitted.

Berdasarkan Gambar 2 terlihat bahwa ketiga varian Naive Bayes tidak membentuk kelompok performa yang terpisah secara tegas sebagaimana lazim ditemukan pada dataset berskala besar, melainkan saling berdekatan pada kisaran *accuracy* 73%–83%. Pola ini mengindikasikan bahwa pada dataset berukuran relatif kecil dan dengan tingkat ketidakseimbangan kelas yang tidak ekstrem, perbedaan karakteristik representasi fitur pada setiap varian Naive Bayes belum tentu menghasilkan kesenjangan performa yang besar. Pembahasan lebih rinci mengenai penyebab perbedaan performa antaralgoritma disajikan pada subbab berikutnya.

D. Perbandingan Kinerja Tiga Varian Naive Bayes

Hasil pengujian pada Tabel 3 dan Gambar 2 menunjukkan bahwa Multinomial Naive Bayes (MNB) secara konsisten memberikan keseimbangan terbaik antara *precision* dan *recall* dibandingkan dua varian lainnya. MNB memperoleh *accuracy* pada kisaran 75,46%–82,87% dengan *accuracy* tertinggi pada rasio 80:20 dengan SMOTE, sekaligus menghasilkan F1-score tertinggi sebesar 87,11% pada konfigurasi yang sama. Keunggulan ini tidak terlepas dari kesesuaian MNB dengan representasi TF-IDF, di mana bobot kemunculan kata dimanfaatkan secara langsung sebagai dasar perhitungan probabilitas, sehingga algoritma ini mampu membedakan kontribusi setiap istilah terhadap kelas sentimen secara lebih proporsional dibandingkan pendekatan berbasis keberadaan kata semata. Temuan ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa Multinomial Naive Bayes merupakan salah satu algoritma yang paling sesuai untuk klasifikasi dokumen teks dengan representasi berbasis frekuensi kata [12], [10].

Berbeda dengan pola yang umum dilaporkan pada penelitian dengan volume data besar, Bernoulli Naive Bayes (BNB) pada penelitian ini justru menghasilkan *accuracy* terendah di antara ketiga varian, yaitu pada kisaran 73,15%–73,61%. Nilai *recall* BNB tercatat sangat tinggi, mencapai 95,45%–97,74% pada seluruh skenario, namun diikuti oleh *precision* yang rendah, yaitu hanya 70,33%–70,79%. Kondisi ini menunjukkan bahwa representasi biner pada BNB membuat model sangat sensitif dalam mendeteksi kelas positif, tetapi kurang mampu membedakan komentar negatif yang memiliki kosakata serupa dengan komentar positif sebagaimana terlihat pada hasil analisis *word cloud* yang menunjukkan tumpang tindih kata dominan antara kedua kelas sentimen, seperti *hilir*, *wakil*, *presiden*, dan *gibran*. Akibatnya, banyak komentar negatif yang justru diprediksi sebagai positif sehingga menurunkan nilai *precision* dan *accuracy* secara keseluruhan meskipun kemampuan mendeteksi kelas positif tetap tinggi.

Sementara itu, Gaussian Naive Bayes (GNB) menghasilkan performa pada kisaran *accuracy* 74,54%–76,85%, berada di antara MNB dan BNB. Secara teoritis, GNB mengasumsikan bahwa setiap fitur mengikuti distribusi normal, sebuah asumsi yang kurang sesuai dengan karakteristik data hasil pembobotan TF-IDF yang bersifat berdimensi tinggi dan *sparse* [15], [13]. Meskipun demikian, pada penelitian ini kesenjangan performa GNB terhadap MNB maupun BNB tidak sebesar temuan yang umum dilaporkan pada dataset berskala besar. Hal ini kemungkinan berkaitan dengan ukuran dataset yang jauh lebih kecil (1.080 komentar) serta tingkat ketidakseimbangan kelas yang relatif ringan pada penelitian ini, sehingga pelanggaran asumsi distribusi normal pada representasi TF-IDF tidak berdampak sebesar pada dataset yang lebih besar dan lebih timpang.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa pemilihan varian Naive Bayes tetap berpengaruh terhadap performa klasifikasi, namun besarnya pengaruh tersebut turut dipengaruhi oleh karakteristik dataset, khususnya ukuran data dan tingkat ketidakseimbangan kelas. MNB tetap menjadi pilihan paling sesuai apabila prioritas utama adalah memperoleh keseimbangan *accuracy* dan F1-score yang tinggi, sedangkan BNB perlu digunakan dengan pertimbangan lebih hati-hati pada dataset dengan kosakata yang tumpang tindih antarkelas, karena kecenderungannya menghasilkan banyak prediksi positif yang keliru.

E. Pengaruh Penerapan SMOTE

Selain membandingkan ketiga varian Naive Bayes, penelitian ini juga mengevaluasi pengaruh penerapan *Synthetic Minority Oversampling Technique* (SMOTE) terhadap performa klasifikasi. Penerapan SMOTE dilakukan mengingat distribusi dataset tetap menunjukkan ketidakseimbangan kelas, yaitu 61,6% komentar positif dan 38,4% komentar negatif, meskipun tingkat ketidakseimbangan tersebut tergolong lebih ringan dibandingkan kondisi umum pada data komentar media sosial.

Berdasarkan Tabel 3, penerapan SMOTE memberikan pengaruh yang berbeda pada tiap varian algoritma. Pada Multinomial Naive Bayes (MNB), SMOTE justru meningkatkan nilai *accuracy* dan *precision*, tetapi diikuti oleh penurunan nilai *recall*. Pada rasio 90:10, *precision* meningkat dari 73,03% menjadi 80,00%, sedangkan *recall* menurun dari 98,48% menjadi 90,91%. Pola serupa terlihat pada rasio 80:20, dengan *precision* meningkat dari 71,74% menjadi 81,17%, sementara *recall* menurun dari 99,25% menjadi 93,98%. Pola ini berbeda dengan *trade-off* yang umum dilaporkan pada penerapan SMOTE, yang biasanya meningkatkan *recall* kelas minoritas dengan konsekuensi penurunan *precision*. Perbedaan arah ini terjadi karena kelas mayoritas pada dataset penelitian ini adalah kelas positif, bukan kelas negatif seperti pada penelitian dengan skema pelabelan sebelumnya. Sebelum penerapan SMOTE, model cenderung bias memprediksi komentar sebagai positif akibat proporsi data latih yang lebih banyak pada kelas tersebut, sehingga *recall* kelas positif tinggi namun banyak komentar negatif yang turut terprediksi positif (*precision* rendah).

Copyright © Universitas Muhammadiyah Sidoarjo. This preprint is protected by copyright held by Universitas Muhammadiyah Sidoarjo and is distributed under the Creative Commons Attribution License (CC BY). Users may share, distribute, or reproduce the work as long as the original author(s) and copyright holder are credited, and the preprint server is cited per academic standards.
Authors retain the right to publish their work in academic journals where copyright remains with them. Any use, distribution, or reproduction that does not comply with these terms is not permitted.

Setelah data latih diseimbangkan melalui SMOTE, kecenderungan bias tersebut berkurang, sehingga model menjadi lebih selektif dalam memprediksi kelas positif tercermin dari naiknya *precision* sekaligus turunnya *recall*.

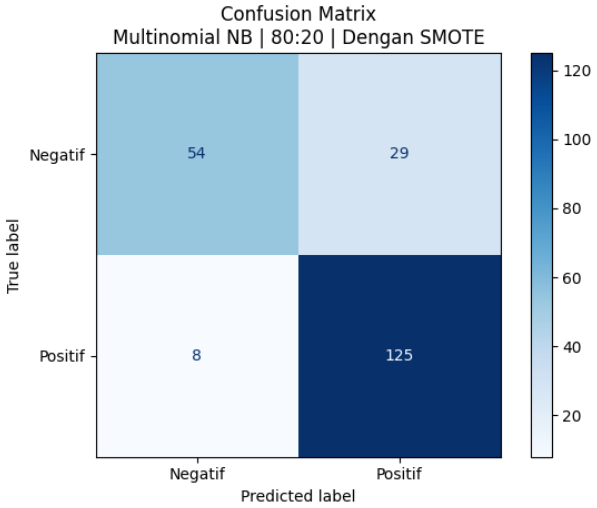
Berbeda dengan MNB, penerapan SMOTE pada Gaussian Naive Bayes (GNB) tidak menunjukkan perubahan nilai pada seluruh metrik evaluasi, baik pada rasio 90:10 maupun 80:20. Kondisi ini mengindikasikan bahwa penambahan sampel sintetis tidak memberikan pengaruh berarti terhadap estimasi rata-rata dan simpangan baku fitur pada masing-masing kelas, yang menjadi dasar perhitungan probabilitas pada GNB. Hal ini konsisten dengan karakteristik GNB yang mengasumsikan distribusi normal pada data kontinu; penambahan sampel sintetis dalam jumlah terbatas pada ruang fitur TF-IDF yang berdimensi tinggi belum cukup mengubah parameter distribusi tersebut secara signifikan.

Pada Bernoulli Naive Bayes (BNB), pengaruh SMOTE juga tergolong sangat kecil. Pada rasio 80:20, seluruh metrik evaluasi tidak menunjukkan perubahan sama sekali, sedangkan pada rasio 90:10 hanya terjadi perubahan minor, yaitu *precision* menurun dari 70,79% menjadi 70,33% dan *recall* meningkat dari 95,45% menjadi 96,97%. Minimnya pengaruh SMOTE pada BNB berkaitan dengan representasi fitur yang bersifat biner (keberadaan atau ketidakberadaan kata), sehingga penambahan sampel sintetis pada ruang fitur kontinu TF-IDF yang kemudian dibinerisasi cenderung menghasilkan pola fitur yang serupa dengan data asli, dan tidak banyak mengubah batas keputusan model.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa efektivitas SMOTE sangat bergantung pada karakteristik algoritma dan tingkat ketidakseimbangan data. Pada dataset dengan ketidakseimbangan kelas yang relatif ringan seperti pada penelitian ini, SMOTE hanya memberikan pengaruh nyata pada algoritma yang secara langsung memanfaatkan proporsi frekuensi kata dalam perhitungan probabilitasnya, yaitu Multinomial Naive Bayes, sedangkan pengaruhnya terbatas pada algoritma yang berbasis distribusi kontinu (Gaussian Naive Bayes) maupun representasi biner (Bernoulli Naive Bayes).

F. Analisis Confusion Matrix

Berdasarkan hasil evaluasi pada Tabel 3, konfigurasi Multinomial Naive Bayes dengan rasio pembagian data 80:20 dan penerapan SMOTE dipilih sebagai model terbaik karena memperoleh *accuracy* tertinggi sekaligus F1-score tertinggi, yaitu masing-masing sebesar 82,87% dan 87,11%. Untuk mengetahui kemampuan klasifikasi model tersebut secara lebih rinci, dilakukan analisis menggunakan *confusion matrix* sebagaimana disajikan pada Gambar 3.



Gambar 3. Confusion Matrix Eksperimen Terbaik

Berdasarkan Gambar 3, dari 83 komentar yang termasuk kelas negatif, sebanyak 54 komentar berhasil diklasifikasikan dengan benar sebagai negatif, sedangkan 29 komentar diprediksi sebagai positif (*False Positive*). Pada kelas positif, dari 133 komentar, sebanyak 125 komentar berhasil dikenali sebagai positif, sedangkan hanya 8 komentar yang diprediksi sebagai negatif (*False Negative*). Hasil tersebut menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengenali sentimen positif, namun masih mengalami kesulitan dalam membedakan sebagian komentar negatif dari komentar positif.

Jumlah *False Positive* yang lebih tinggi dibandingkan *False Negative* menunjukkan bahwa model cenderung lebih sering salah mengklasifikasikan komentar negatif sebagai positif dibandingkan sebaliknya. Kondisi ini konsisten dengan hasil analisis *word cloud* pada subbab berikutnya, yang memperlihatkan tumpang tindih kata dominan antara kedua kelas sentimen, seperti *hilir*, *wakil*, *presiden*, *indonesia*, *mas*, dan *gibran*. Tumpang tindih kosakata tersebut menyebabkan model kesulitan membedakan konteks sentimen ketika suatu komentar negatif menggunakan kata-kata

Copyright © Universitas Muhammadiyah Sidoarjo. This preprint is protected by copyright held by Universitas Muhammadiyah Sidoarjo and is distributed under the Creative Commons Attribution License (CC BY). Users may share, distribute, or reproduce the work as long as the original author(s) and copyright holder are credited, and the preprint server is cited per academic standards.
Authors retain the right to publish their work in academic journals where copyright remains with them. Any use, distribution, or reproduction that does not comply with these terms is not permitted.

yang secara umum juga sering muncul pada komentar positif, misalnya ketika kritik disampaikan menggunakan istilah atau nama figur publik yang sama dengan konteks komentar dukungan. Karakteristik ini umum dijumpai pada data komentar media sosial, terutama pada topik yang bersifat politis, karena makna sentimen sering kali bergantung pada konteks kalimat secara utuh dan bukan semata-mata pada kemunculan kata tertentu.

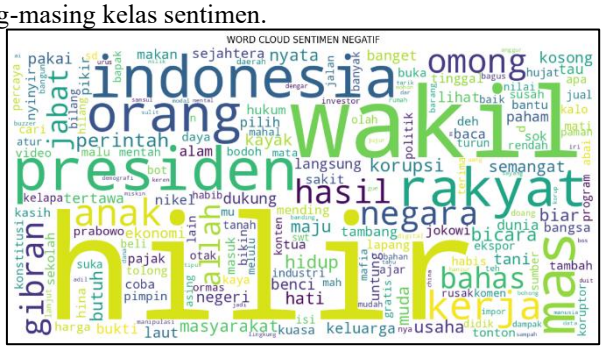
Meskipun demikian, jumlah prediksi yang benar tetap jauh lebih besar dibandingkan kesalahan klasifikasi pada kedua kelas, dengan total 179 dari 216 data uji (82,87%) diklasifikasikan dengan tepat. Hasil ini memperkuat temuan sebelumnya bahwa Multinomial Naive Bayes mampu memanfaatkan representasi TF-IDF secara efektif, sehingga tetap menghasilkan tingkat klasifikasi yang cukup stabil pada data komentar YouTube berbahasa Indonesia meskipun ukuran dataset relatif kecil.

G. Analisis Word Cloud

Selain evaluasi kuantitatif, penelitian ini juga melakukan analisis secara visual menggunakan word cloud untuk mengidentifikasi kata-kata yang paling dominan pada masing-masing kelas sentimen.



Gambar 4. Word Cloud Sentimen Positif



Gambar 5. Word Cloud Sentimen Negatif

Berdasarkan hasil *word cloud* sentimen positif, kata yang paling dominan muncul adalah *hilir*, *indonesia*, *presiden*, *mas*, dan *wakil*, diikuti oleh kata-kata seperti *moga*, *muda*, *semangat*, *gibran*, dan *maju*. Dominasi kata-kata tersebut menunjukkan bahwa komentar positif umumnya berisi dukungan terhadap figur publik, harapan terhadap kemajuan bangsa, serta apresiasi terhadap gagasan hilirisasi yang disampaikan. Karakteristik ini mengindikasikan bahwa sebagian pengguna YouTube memberikan respons yang optimistis terhadap materi yang disampaikan dalam video.

Pada *word cloud* sentimen negatif, kata yang paling dominan adalah *hilir*, *wakil*, *presiden*, *indonesia*, dan *mas*, diikuti oleh kata-kata seperti *rakyat*, *orang*, *kerja*, *anak*, dan *omong*. Berbeda dengan komentar positif, kata-kata pada kelas negatif seperti *omong*, *bahas*, *jabat*, *perintah*, dan *korupsi* menunjukkan bahwa komentar negatif lebih banyak memuat kritik terhadap kredibilitas maupun kinerja pihak yang menyampaikan kebijakan, keraguan terhadap implementasi program, serta sentimen politik yang berkembang di ruang diskusi publik, dan bukan semata-mata kritik terhadap substansi kebijakan hilirisasi itu sendiri.

Perbandingan kedua *word cloud* tersebut memperlihatkan adanya tumpang tindih kata dominan yang cukup signifikan antara kelas positif dan negatif, khususnya pada kata *hilir*, *wakil*, *presiden*, *indonesia*, *mas*, dan *gibran*. Kata-kata tersebut merupakan istilah dan nama figur publik yang secara alami sering muncul pada kedua kelas sentimen karena berkaitan langsung dengan topik dan objek pembahasan, sehingga kata-kata tersebut cenderung kurang diskriminatif sebagai pembeda sentimen. Sebaliknya, kata-kata yang lebih spesifik pada masing-masing kelas, seperti *semangat*, *maju*, dan *dukung* pada kelas positif, serta *omong*, *bahas*, dan *korupsi* pada kelas negatif, menjadi pembeda yang lebih bermakna antara kedua kelas. Kondisi tumpang tindih kosakata ini turut menjelaskan temuan pada analisis *confusion matrix* sebelumnya, yaitu kecenderungan model untuk salah mengklasifikasikan sebagian komentar negatif sebagai positif akibat kemiripan kata-kata umum yang digunakan pada kedua kelas.

Secara keseluruhan, hasil analisis *word cloud* menunjukkan bahwa proses *preprocessing* dan pembobotan TF-IDF berhasil mempertahankan kata-kata yang relevan dengan topik hilirisasi, namun tingginya frekuensi kata-kata umum yang tidak spesifik terhadap sentimen turut menjadi salah satu faktor yang memengaruhi performa klasifikasi model.

H. Pembahasan Umum

Berdasarkan keseluruhan hasil penelitian, terdapat empat temuan utama yang dapat disimpulkan. Pertama, Multinomial Naive Bayes merupakan algoritma dengan performa terbaik pada penelitian ini, dengan satu konfigurasi yang sama (rasio 80:20 dengan SMOTE) menghasilkan *accuracy* tertinggi sekaligus F1-score tertinggi. Kondisi ini berbeda dengan pola yang umum dilaporkan pada penelitian berskala data besar, yang biasanya menghasilkan konfigurasi terbaik yang berbeda antara *accuracy* dan F1-score. Sementara itu, Bernoulli Naive Bayes dan Gaussian Naive Bayes menghasilkan performa yang relatif berdekatan dan berada di bawah Multinomial Naive Bayes, dengan selisih *accuracy* antaralgoritma yang tidak sebesar temuan pada penelitian dengan volume data yang jauh lebih besar.

Kedua, pengaruh penerapan SMOTE terbukti bersifat spesifik terhadap karakteristik masing-masing algoritma, bukan bersifat seragam pada seluruh model. Pada Multinomial Naive Bayes, SMOTE meningkatkan *precision* namun

Copyright © Universitas Muhammadiyah Sidoarjo. This preprint is protected by copyright held by Universitas Muhammadiyah Sidoarjo and is distributed under the Creative Commons Attribution License (CC BY). Users may share, distribute, or reproduce the work as long as the original author(s) and copyright holder are credited, and the preprint server is cited per academic standards.
Authors retain the right to publish their work in academic journals where copyright remains with them. Any use, distribution, or reproduction that does not comply with these terms is not permitted.

menurunkan *recall*, arah *trade-off* yang berbeda dengan pola umum pada literatur, sebagai konsekuensi dari kelas mayoritas pada dataset penelitian ini yang merupakan kelas positif. Sebaliknya, pada Gaussian Naive Bayes dan Bernoulli Naive Bayes, SMOTE hanya memberikan pengaruh yang sangat kecil atau bahkan tidak berpengaruh sama sekali terhadap seluruh metrik evaluasi. Temuan ini menunjukkan bahwa efektivitas SMOTE tidak dapat digeneralisasi lintas algoritma, melainkan perlu dievaluasi secara khusus sesuai mekanisme perhitungan probabilitas masing-masing varian Naive Bayes.

Ketiga, distribusi dataset yang menunjukkan 61,6% komentar bersentimen positif dan 38,4% komentar bersentimen negatif mengindikasikan bahwa di antara komentar yang secara spesifik relevan dengan topik kebijakan hilirisasi, opini masyarakat cenderung lebih banyak memberikan dukungan dibandingkan kritik. Meskipun demikian, hasil analisis *word cloud* pada kelas negatif tetap menunjukkan adanya kritik terhadap kredibilitas dan implementasi kebijakan, sehingga opini publik pada topik ini tetap bersifat heterogen dan tidak sepenuhnya seragam.

Keempat, hasil analisis *confusion matrix* dan *word cloud* secara bersama-sama menunjukkan bahwa kesalahan klasifikasi model terutama disebabkan oleh tumpang tindih kosakata umum antara kelas positif dan negatif, seperti nama figur publik dan istilah kebijakan yang muncul pada kedua kelas. Temuan ini menegaskan bahwa keterbatasan performa model pada penelitian ini tidak semata-mata disebabkan oleh pemilihan algoritma atau teknik penyeimbangan data, tetapi juga oleh karakteristik bahasa pada komentar media sosial yang maknanya sering kali bergantung pada konteks, bukan hanya pada kemunculan kata secara individual.

Secara keseluruhan, penelitian ini menunjukkan bahwa pemilihan varian Naive Bayes, karakteristik ukuran dan distribusi dataset, serta strategi penanganan ketidakseimbangan data merupakan faktor yang saling berkaitan dalam menentukan keberhasilan klasifikasi sentimen pada komentar YouTube berbahasa Indonesia, khususnya pada dataset berskala kecil yang telah difilter berdasarkan relevansi topik.

IV. KESIMPULAN

Penelitian ini berhasil menganalisis sentimen komentar YouTube terhadap pernyataan kebijakan hilirisasi yang berkaitan dengan Gibran Rakabuming Raka menggunakan tiga varian algoritma Naive Bayes melalui 12 skenario pengujian, dengan data yang telah disaring berdasarkan relevansi topik dan diberi label secara otomatis menggunakan model IndoBERT-Sentiment *context-conditioned*, sehingga diperoleh 1.080 komentar akhir yang didominasi oleh sentimen positif (61,6%) dibandingkan negatif (38,4%). Hasil evaluasi menunjukkan bahwa Multinomial Naive Bayes pada rasio pembagian data 80:20 dengan SMOTE menjadi model dengan performa terbaik, dengan *accuracy* sebesar 82,87% sekaligus F1-score tertinggi sebesar 87,11%, mengungguli Bernoulli Naive Bayes (73,15%–73,61%) dan Gaussian Naive Bayes (74,54%–76,85%) yang menghasilkan performa relatif berdekatan; penerapan SMOTE sendiri terbukti hanya berpengaruh signifikan pada Multinomial Naive Bayes, sementara hasil analisis *confusion matrix* dan *word cloud* menunjukkan bahwa kesalahan klasifikasi model terutama disebabkan oleh tumpang tindih kosakata umum antara kedua kelas sentimen. Berdasarkan temuan tersebut, penelitian selanjutnya disarankan untuk menyertakan kelas netral atau menerapkan *aspect-based sentiment analysis*, membandingkan pendekatan Naive Bayes dengan metode *deep learning* seperti IndoBERT atau LSTM, serta memperluas cakupan sumber data agar diperoleh gambaran opini publik yang lebih komprehensif.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada Universitas Muhammadiyah Sidoarjo yang telah memberikan dukungan terhadap pelaksanaan penelitian ini. Apresiasi juga disampaikan kepada seluruh pihak yang telah memberikan bantuan, fasilitas, serta masukan selama proses pengumpulan data, pengolahan data, dan penyusunan artikel ini sehingga penelitian dapat diselesaikan dengan baik.

REFERENSI

[1] A. Fandhy dkk., “Media Baru dalam Dinamika Komunikasi Politik pada Era Society 5.0,” Jurnal Indonesia : Manajemen Informatika dan Komunikasi (JIMIK), vol. 6, no. 1, Jan 2025, doi: 10.35870/jimik.v6i1.1280.

[2] N. T. Putri, J. Patricia, dan G. R. Qaira, “PENGELOLAAN MEDIA SOSIAL YOUTUBE SEBAGAI STRATEGI DIGITAL PUBLIC RELATIONS DI HARIAN BHIRAWA,” RELASI: Jurnal Penelitian Komunikasi, vol. 05, no. 04, hlm. 14–21, 2025, doi: 10.69957/relasi.v5i04.1873.

[3] S. De Putri dan A. Toni, “SOSIOLOGI: Jurnal Ilmiah Kajian Ilmu Sosial dan Budaya ANALISIS PERCAKAPAN NETIZEN PADA CHANNEL YOUTUBE @TotalPolitik SEBAGAI MEDIA RUANG PUBLIK KOMUNIKASI POLITIK,” SOSIOLOGI: Jurnal Ilmiah Kajian Ilmu Sosial dan Budaya, vol. 26, no. 2, hlm. 133–150, 2024, doi: 10.23960/sosiologi.v26i2.1389.

Copyright © Universitas Muhammadiyah Sidoarjo. This preprint is protected by copyright held by Universitas Muhammadiyah Sidoarjo and is distributed under the Creative Commons Attribution License (CC BY). Users may share, distribute, or reproduce the work as long as the original author(s) and copyright holder are credited, and the preprint server is cited per academic standards.
Authors retain the right to publish their work in academic journals where copyright remains with them. Any use, distribution, or reproduction that does not comply with these terms is not permitted.

[4] W. Jaya Kusoema dan I. Ibrahim, "Sistemasi: Jurnal Sistem Informasi Analisis Sentimen dalam Kasus Korupsi PT. Pertamina menggunakan Metode indoBERT dan RCNN Sentiment Analysis on the PT Pertamina Corruption Case using IndoBERT and RCNN Methods," Sistemasi: Jurnal Sistem Informasi, vol. 14, no. 5, hlm. 2246–2257, 2025, [Daring]. Tersedia pada: http://sistemasi.ftik.unisi.ac.id

[5] P. Arsi dan R. Waluyo, "ANALISIS SENTIMEN WACANA PEMINDAHAN IBU KOTA INDONESIA MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)," Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), vol. 8, no. 1, hlm. 147–156, Feb 2021, doi: 10.25126/jtiik.202183944.

[6] I. Hulu dan H. Budiati, "Implementasi Text Mining Pada Pengukuran Sentimen Opini Masyarakat Terhadap Universitas Kristen Immanuel," JURNAL SAINS DAN KOMPUTER, vol. 7, no. 02, hlm. 1–7, Jul 2023, doi: 10.61179/jurnalinfact.v7i02.445.

[7] M. Agung dan E. A. W. Adi, "Peningkatan Investasi Dan Hilirisasi Nikel Di Indonesia," Jurnal Ilmu Sosial dan Pendidikan (JISIP), vol. 6, no. 2, Mar 2022, doi: 10.36312/jisip.v6i1.3085.

[8] K. Alawiyah Matondang, A. Handayani, A. Parlindungan Hasibuan, dan J. Simanungkalit, "STRATEGI HILIRISASI DI INDONESIA DALAM MENGHADAPI KEBIJAKAN PROTEKSIONISME PADA KOMODITAS BAJA/BESI," Jurnal Masharif al-Syariah: Jurnal Ekonomi dan Perbankan Syariah, vol. 9, no. 5, hlm. 3371–3377, 2024, doi: 10.30651/jms.v9i5.24339.

[9] D. Pramudita, Y. Akbar, dan T. Wahyudi, "Analisis Sentimen Terhadap Program Kartu Indonesia Pintar Kuliah pada Media Sosial X Menggunakan Algoritma Naive Bayes," MALCOM: Indonesian Journal of Machine Learning and Computer Science, vol. 4, no. 4, hlm. 1420–1430, Agu 2024, doi: 10.57152/malcom.v4i4.1565.

[10] M. Indra, D. Wahiddin, dan S. A. P. Lestari, "Implementasi Algoritma Naïve Bayes Pada Analisis Sentiemen Terhadap Larangan Ekspor Nikel Oleh Pemerintah Indonesia," Scientific Student Journal for Information, Technology and Science, vol. 5, no. 2, hlm. 137–144, Jul 2024.

[11] F. Adelia Irawan, A. Rialdy Atmadja, A. Wahana, T. Informatika, dan U. Sunan Gunung Djati Bandung, "Analisis Sentimen Ulasan Aplikasi Bank Digital Menggunakan Algoritma Naive Bayes," Journal of Computer Science and Information Technology E-ISSN, vol. 4, no. 2, hlm. 60–68, Jul 2024, doi: 10.47065/explorer.v4i2.1181.

[12] C. Naulak, "A comparative study of Naive Bayes Classifiers with improved technique on Text Classification," Mei 2022. doi: 10.36227/techrxiv.19918360.v1.

[13] W. M. P. Dhuhita dan F. Zone, "Perbandingan Kinerja Algoritma Multinomial dan Bernoulli Naïve Bayes dalam Mengklasifikasikan Komentar Cyberbullying," Komputika : Jurnal Sistem Komputer, vol. 12, no. 2, hlm. 109–117, Okt 2023, doi: 10.34010/komputika.v12i2.9767.

[14] N. A'ayunnisa, Y. Salim, dan H. Azis, "Analisis performa metode Gaussian Naïve Bayes untuk klasifikasi citra tulisan tangan karakter arab," Indonesian Journal of Data and Science (IJODAS), vol. 3, no. 3, hlm. 115–121, Des 2022.

[15] I. B. Mahendra, I. M. G. Sunarya, dan I. M. A. Wirawan, "Comparison of Multinomial, Bernoulli, and Gaussian Naïve Bayes for Complaint Classification in Pro Denpasar Application," JUITA: Jurnal Informatika, vol. 13, no. 1, hlm. 77–86, Mar 2025.

[16] M. A. Hermawan, A. Faqih, dan G. Dwilestari, "IMPLEMENTASI AKURASI MODEL NAIVE BAYES MENGGUNAKAN SMOTE DALAM ANALISIS SENTIMEN PENGGUNA APLIKASI BRIMO," Jurnal Informatika dan Teknik Elektro Terapan, vol. 13, no. 1, Jan 2025, doi: 10.23960/jitet.v13i1.5748.

[17] S. F. Huwaida, R. Kusumawati, dan B. Isnaini, "Analisis Sentimen Komentar YouTube terhadap Pemindahan Ibu Kota Negara Menggunakan Metode Naïve Bayes," Jambura Journal of Informatics, vol. 6, no. 1, hlm. 26–39, Apr 2024, doi: 10.37905/jji.v6i1.24718.

Conflict of Interest Statement:
The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © Universitas Muhammadiyah Sidoarjo. This preprint is protected by copyright held by Universitas Muhammadiyah Sidoarjo and is distributed under the Creative Commons Attribution License (CC BY). Users may share, distribute, or reproduce the work as long as the original author(s) and copyright holder are credited, and the preprint server is cited per academic standards.
Authors retain the right to publish their work in academic journals where copyright remains with them. Any use, distribution, or reproduction that does not comply with these terms is not permitted.