

Analisis Sentimen Publik terhadap Purbaya Yudhi Sadewa di X menggunakan Metode Naive Bayes

Rian Nurdiana dan Alam Rahmatulloh*)

Program Studi Informatika Universitas Siliwangi, Jl. Siliwangi No. 24, Tasikmalaya, Jawa Barat, 46115, Indonesia.

*) Surel korespondensi : alam@unsil.ac.id

Abstract. *The rapid growth of the social media platform X (Twitter) has made it an essential medium for observing how the public responds to influential figures and economic policy issues. This study investigates public sentiment toward Purbaya Yudhi Sadewa by applying the Naive Bayes Classifier to tweet data collected from X. A total of 500 tweets containing the keyword "Purbaya Yudhi Sadewa" were gathered through API crawling between January and November 2025, and 451 tweets were retained after filtering irrelevant and duplicate entries. The workflow of this research consists of text preprocessing steps – such as case folding, tokenization, stopword removal, and stemming – followed by TF-IDF-based feature extraction, model training using Naive Bayes, and performance evaluation. The classifier produced an accuracy of 91.1%, with precision and recall values of 0.92 and 0.97 for negative sentiment, and 0.60 and 0.33 for positive sentiment. Further analysis showed that negative opinions dominated the dataset (427 tweets), while positive sentiments appeared far less frequently (32 tweets), primarily criticizing economic performance, inflation, and the stability of the rupiah. These findings highlight the capability of the Naive Bayes algorithm in effectively categorizing sentiment toward economic figures in Indonesian-language social media content.*

Keywords: *classifier, classification, inflation, opinion, TF-IDF*



This work is licensed under Attribution-NonCommercial-ShareAlike 4.0 International. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>

Diterbitkan oleh LPPM Institut Bisnis dan Informatika Kwik Kian Gie. Jl. Yos Sudarso Kav 87, Sunter Jakarta 14350, Indonesia. DOI : <https://doi.org/10.46806/jib.v15i1.1654>

1. Pendahuluan

Perkembangan teknologi informasi, khususnya media sosial, telah mengubah cara masyarakat dalam penyampaian pendapat, berdiskusi, dan memberikan tanggapan terhadap isu-isu publik (Kaplan, 2017). Salah satu platform yang paling aktif digunakan di Indonesia adalah X (sebelumnya Twitter). Melalui platform ini, pengguna dapat menyampaikan opini secara cepat dan langsung terkait berbagai persoalan sosial, politik, dan ekonomi. (Anton Purwadi, 2022). Arus komunikasi dua arah yang cepat menjadikan X sebagai sumber utama untuk memahami persepsi publik terhadap figur publik, kebijakan, dan lembaga pemerintahan (Puluhulawa, 2022).

Dalam konteks ekonomi Indonesia, tokoh publik seperti Purbaya Yudhi Sadewa memiliki peran penting dalam perumusan kebijakan dan pengambilan keputusan strategis. Reaksi publik terhadap pernyataan maupun kebijakan yang diasosiasikan dengan beliau dapat mencerminkan tingkat kepercayaan masyarakat terhadap kebijakan ekonomi nasional (Sualman, 2023). Namun, opini masyarakat di dunia maya bersifat beragam dan sulit dipetakan tanpa bantuan teknologi *text mining*. Di

sinilah analisis sentimen berperan dalam mengklasifikasikan opini publik ke dalam kategori positif, negatif, atau netral secara otomatis.

Penelitian ini mengusulkan penerapan metode *Naive Bayes* untuk mengklasifikasikan sentimen publik terhadap Purbaya Yudhi Sadewa di platform X, dengan tahapan mencakup *data crawling*, *text preprocessing*, *feature extraction* menggunakan *TF-IDF*, pelatihan model, dan evaluasi kinerja menggunakan metrik akurasi, presisi, *recall*, dan *F1-score* (Andrian et al., 2023). Inovasi utama penelitian ini terletak pada penerapan model *Naive Bayes* dalam konteks figur publik ekonomi Indonesia, dengan fokus pada opini berbahasa Indonesia yang mencerminkan persepsi masyarakat secara aktual.

2. Tinjauan Pustaka

2.1. Text Mining

Text mining merupakan proses pengolahan teks tidak terstruktur untuk mengekstraksi informasi dan pola yang bermakna melalui tahapan seperti pembersihan teks, normalisasi, tokenisasi, dan pemodelan fitur. Dalam konteks media sosial, *text mining* menjadi krusial karena teks yang dihasilkan bersifat singkat, informal, dan dinamis.

Berbagai penelitian telah memanfaatkan *text mining* untuk memahami opini dan perilaku publik. Menurut Purwadi et al. (2022) menggunakan *text mining* untuk menganalisis keterlibatan warga pada akun Twitter pemerintah daerah dan menemukan bahwa pola interaksi publik dapat diidentifikasi melalui analisis kata kunci dan struktur percakapan. Menurut Penelitian Puluhulawa (2022) menunjukkan bahwa *text mining* efektif untuk mengelompokkan tema diskusi publik selama masa *new normal*. Kedua studi tersebut menegaskan bahwa *text mining* merupakan fondasi penting dalam analisis sentimen di platform seperti X.

2.2. Sentiment Analysis

Sentiment analysis merupakan teknik untuk menentukan arah emosi atau pendapat yang terkandung dalam sebuah teks dengan mengklasifikasikannya ke dalam kategori positif, negatif, atau netral. Metode ini banyak dimanfaatkan untuk menelaah bagaimana masyarakat memandang suatu kebijakan, figur publik, maupun isu sosial dan ekonomi. Tahapannya mencakup pembersihan teks (*text preprocessing*), pengubahan teks menjadi fitur numerik, serta proses klasifikasi menggunakan algoritma machine learning.

Sejumlah penelitian sebelumnya menunjukkan bahwa sentiment analysis mampu memberikan gambaran yang akurat mengenai opini publik. Sari dan Putra (2024), misalnya, menerapkan algoritma Support Vector Machine dan Naive Bayes untuk mengevaluasi persepsi publik terhadap tokoh politik, dan memperoleh tingkat akurasi yang konsisten serta tinggi. Susanto dan Atho (2021) juga melakukan perbandingan antara algoritma Naive Bayes dan K-Means pada data Twitter

berbahasa Indonesia, dan hasilnya menunjukkan bahwa Naive Bayes memiliki performa yang lebih baik dengan akurasi melebihi 80%. Temuan-temuan ini memperkuat dasar penggunaan sentiment analysis dalam penelitian ini, khususnya untuk memahami pandangan masyarakat terhadap tokoh ekonomi nasional.

2.3. Naive Bayes Classifier

Naive Bayes Classifier adalah algoritma klasifikasi berbasis probabilistik yang memanfaatkan fitur (kata) dalam dokumen. Dalam analisis sentimen, metode ini menentukan peluang sebuah teks berada pada kelas tertentu dengan melihat pola kemunculan kata-katanya. Kelebihan Naive Bayes terletak pada kemampuannya mengolah teks pendek dan tidak baku serta efisiensi komputasinya yang tinggi.

Sejumlah penelitian telah membuktikan efektivitas Naive Bayes. Menurut Classifier et al. (2022) menganalisis opini mengenai vaksin COVID-19 dan mencapai akurasi di atas 80%. Andra et al. (2024) menerapkan metode ini pada analisis sentimen terkait Ibu Kota Nusantara (IKN) dan menemukan bahwa sentimen netral mendominasi diskusi publik. Menurut Mario et al. (2024) menggunakan Naive Bayes pada big data untuk menilai persepsi terhadap tokoh nasional, sementara Menurut Cahyono (2019) menunjukkan efektivitasnya pada teks berbahasa daerah. Konsistensi performa pada berbagai konteks ini menjadi dasar kuat penggunaan Naive Bayes pada analisis sentimen terhadap Purbaya Yudhi Sadewa.

2.3.1. Formulasi Matematis Naive Bayes

Sebagai dasar teoritis bagi penerapan *Naive Bayes Classifier* dalam penelitian ini, bagian ini memaparkan formulasi matematis serta penurunan yang mendasari aturan keputusan pada model. Penjelasan matematis membantu menegaskan alasan penggunaan model probabilistik sederhana ini untuk klasifikasi teks seperti tweet.

Teorema Bayes. Untuk sebuah dokumen (vektor fitur) X dan kelas C , Teorema Bayes menyatakan:

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)}.$$

Dalam tugas klasifikasi kita mencari kelas \hat{C} yang memaksimalkan probabilitas posterior:

$$\hat{C} = \underset{C}{\operatorname{argmax}} P(C|X).$$

Karena $P(X)$ adalah konstanta terhadap pilihan kelas, pembandingan dapat disederhanakan menjadi:

$$\hat{C} = \underset{C}{\operatorname{argmax}} P(X|C)P(C).$$

Asumsi *naive* (independensi kondisional fitur).

$X = (x_1, x_2, \dots, x_n)$ adalah

Misalkan

sekumpulan fitur (mis. kata atau token). Asumsi *naive* menyatakan fitur-fitur tersebut bersifat independen bersyarat terhadap kelas C , sehingga:

$$P(X|C) = \prod_{i=1}^n P(x_i|C).$$

Dengan substitusi diperoleh aturan keputusan *naive*:

$$\hat{C} = \underset{C}{\operatorname{argmax}} P(C) \prod_{i=1}^n P(x_i|C).$$

Formulasi Multinomial untuk teks. Untuk representasi *bag-of-words* dengan vocabularium berukuran V dan count vector $w = (w_1, \dots, w_V)$ (di mana w_j adalah frekuensi kata ke- j pada dokumen), likelihood menurut model multinomial adalah:

$$P(w|C) = \frac{N!}{\prod_{j=1}^V w_j!} \prod_{j=1}^V \theta_{j|C}^{w_j}$$

dengan $N = \sum_{j=1}^V w_j$ dan $\theta_{j|C} = P(\text{kata } j|C)$. Karena faktor multinomial $\frac{N!}{\prod w_j!}$ sama untuk semua kelas pada dokumen yang sama, faktor ini diabaikan dalam argmax sehingga:

$$\hat{C} = \underset{C}{\operatorname{argmax}} P(C) \prod_{j=1}^V \theta_{j|C}^{w_j}$$

Estimasi parameter dan *Laplace smoothing*. Parameter $\theta_{j|C}$ diestimasi dari data pelatihan sebagai proporsi kemunculan kata pada kelas C :

$$\hat{\theta}_{j|C} = \frac{\text{count}(j,C)}{\sum_{k=1}^V \text{count}(k,C)}.$$

Untuk mencegah masalah nol (kata yang tidak muncul pada kelas tertentu), digunakan *Laplace smoothing* (add- α , biasanya $\alpha = 1$):

$$\hat{\theta}_{j|C} = \frac{\text{count}(j,C) + \alpha}{\sum_{k=1}^V \text{count}(k,C) + \alpha V}.$$

Transformasi log untuk kestabilan numerik. Mengambil log pada ekspresi di atas mengubah produk menjadi jumlah dan mencegah underflow numerik:

$$\hat{C} = \underset{C}{\operatorname{argmax}} (\log P(C) + \sum_{j=1}^V w_j \log \hat{\theta}_{j|C}).$$

Karena fungsi log monoton, transformasi ini tidak mengubah urutan argmax. Bentuk ini menunjukkan bahwa penilaian kelas adalah penjumlahan bobot kata (linear terhadap feature counts) ditambah log prior kelas.

Kaitan dengan TF-IDF. Model Multinomial secara teoritis konsisten dengan *count features*. Jika fitur yang dipakai adalah skor TF-IDF x_j alih-alih count w_j , aturan keputusan dapat ditulis secara analog:

$$\hat{C} = \underset{C}{\operatorname{argmax}} (\log P(C) + \sum_{j=1}^V x_j \log \tilde{\theta}_{j|C}),$$

tetapi perlu dicatat bahwa penggunaan TF-IDF mengubah interpretasi probabilistik murni karena x_j bukanlah count diskrit; dalam praktik, kombinasi TF-IDF dengan Naive Bayes sering meningkatkan kinerja meskipun interpretasi probabilitas menjadi heuristik.

Mengapa $P(X)$ dapat diabaikan pada argmax — bukti singkat. Untuk dua kelas arbitrer C_1 dan C_2 :

$$P(C_1 | X) \geq P(C_2 | X) \Leftrightarrow \frac{P(X|C_1)P(C_1)}{P(X)} \geq \frac{P(X|C_2)P(C_2)}{P(X)}.$$

Karena $P(X) > 0$, pembagi sama sehingga:

$$\Leftrightarrow P(X|C_1)P(C_1) \geq P(X|C_2)P(C_2),$$

yang menunjukkan penghapusan $P(X)$ tidak mengubah urutan perbandingan.

Keterbatasan asumsi independensi. Meskipun asumsi independensi jarang terpenuhi sepenuhnya dalam teks kata-kata sering saling bergantung (mis. frasa, negasi, sarkasme) Naive Bayes tetap memberikan performa kompetitif pada banyak tugas teks karena sinyal kata individu sering cukup informatif. Kelemahan ini perlu dicatat ketika menginterpretasikan hasil, khususnya untuk mendeteksi sentimen yang bersifat kontekstual atau ironis.

2.4. TF-IDF

TF-IDF (*Term Frequency–Inverse Document Frequency*) adalah teknik pembobotan kata yang bertujuan menonjolkan kata penting dalam dokumen dengan menghitung frekuensi kemunculan kata serta relevansinya terhadap seluruh korpus. Dalam analisis sentimen, TF-IDF memetakan teks ke dalam vektor numerik sehingga dapat digunakan sebagai input bagi algoritma klasifikasi seperti Naive Bayes.

Menurut Andrian et al. (2023) menggabungkan TF-IDF dengan Naive Bayes dalam analisis minat bayar pajak berdasarkan tweet, dan memperoleh akurasi tinggi serta model yang stabil. Sementara itu, Menurut Suryono dan Setiawan (2024) membandingkan performa TF-IDF dengan metode pembobotan lain dan menemukan bahwa TF-IDF memberikan hasil paling konsisten untuk teks bahasa Indonesia di media sosial. Studi-studi tersebut mendukung relevansi TF-IDF sebagai metode representasi fitur dalam penelitian ini.'

2.4.1 formulasi Matematis TF-IDF

1) Term Frequency (TF)

TF mengukur seberapa sering kata muncul dalam dokumen. Untuk kata t pada dokumen d , definisinya:

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

di mana:

- $f_{t,d}$ = jumlah kemunculan kata t di dokumen d
- $\sum_k f_{k,d}$ = total seluruh kata dalam dokumen

Bentuk ini adalah *normalized TF*, digunakan agar dokumen panjang tidak menghasilkan bobot lebih besar secara otomatis.

2) Inverse Document Frequency (IDF)

IDF berfungsi merendahkan bobot kata yang terlalu umum muncul pada hampir semua dokumen.

Versi paling umum dan standar dalam IR adalah:

$$IDF(t) = \log\left(\frac{N}{df(t)}\right)$$

dengan:

- N = jumlah dokumen dalam korpus
- $df(t)$ = jumlah dokumen yang mengandung kata t

Untuk menghindari pembagian nol (bila kata muncul pada 0 dokumen lain), digunakan *smoothing*:

$$IDF(t) = \log\left(\frac{1+N}{1+df(t)}\right)$$

Tanpa penambahan konstanta "+1" lainnya — ini adalah bentuk *smoothed IDF* yang benar dan paling sering digunakan di penelitian akademik (termasuk pada *scikit-learn*).

3) TF-IDF

Bobot akhir TF-IDF untuk kata t pada dokumen d adalah:

$$TF - IDF(t, d) = TF(t, d) \times IDF(t)$$

Nilai TF-IDF besar apabila suatu kata:

1. sering muncul dalam dokumen (TF tinggi), dan
2. jarang muncul pada dokumen lain (IDF tinggi).

Dengan demikian kata yang benar-benar mewakili isi dokumen akan memiliki bobot terbesar.

4) Representasi Vektor Dokumen

Dokumen diubah ke bentuk vektor:

$$x_d = (TF - IDF(t_1, d), TF - IDF(t_2, d), \dots, TF - IDF(t_v, d))$$

di mana V adalah ukuran vocabulary.

Vektor inilah yang digunakan sebagai input pada algoritma *Naive Bayes*.

5) Hubungan TF-IDF dengan Naive Bayes

Ketika TF-IDF digunakan sebagai fitur dalam Naive Bayes, aturan klasifikasi menjadi:

$$\hat{C} = \underset{C}{\operatorname{argmax}} (\log P(C) + \sum_{j=1}^V x_j \log \theta_{j|C})$$

dengan:

- x_j = bobot TF-IDF kata ke- j
- $\theta_{j|C}$ = estimasi probabilitas kata ke- j pada kelas C

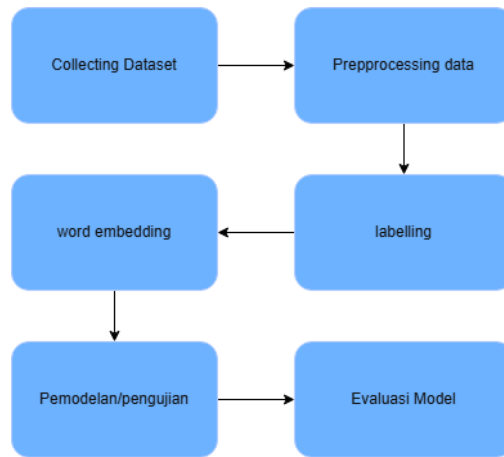
Meskipun TF-IDF tidak mewakili *count* kata secara murni, kombinasi ini secara empiris terbukti memberikan performa tinggi pada data teks pendek seperti tweet.

3. Metode

Penelitian ini melibatkan serangkaian tahapan yang sistematis untuk memastikan proses analisis sentimen publik terhadap Purbaya Yudhi Sadewa di media sosial X dapat dilakukan secara akurat dan terukur. Tahapan penelitian terdiri dari pengumpulan data, pra-pemrosesan teks (text preprocessing), pelabelan data, ekstraksi fitur, pembentukan model Naive Bayes, pengujian model, dan evaluasi kinerja model. Setiap tahap dilakukan secara berurutan guna menghasilkan model klasifikasi sentimen yang optimal. Diagram alir penelitian disajikan pada Gambar 1 berikut, yang menggambarkan keseluruhan alur proses analisis dari data mentah hingga hasil evaluasi model.

Penelitian ini memanfaatkan data ulasan yang diperoleh melalui pemrograman Python. Setelah proses pengumpulan selesai, data tersebut menjalani tahap pembersihan (preprocessing) dan pelabelan. Ketika data dianggap sudah layak, dilakukan pembagian dataset dengan proporsi seimbang antara sentimen positif dan negatif agar kualitas data yang digunakan untuk pelatihan model tetap optimal. Selanjutnya, fitur teks diekstraksi menggunakan metode TF-IDF untuk mengubah setiap kata dalam ulasan menjadi representasi numerik. Model klasifikasi sentimen kemudian dibangun menggunakan algoritma Naive Bayes,

yang dirancang untuk memproses data teks secara efisien. Tahap akhir adalah evaluasi performa model dengan menghitung metrik akurasi, precision, recall, dan F1-score.



Gambar 1. Diagram Alir

4. Hasil dan Pembahasan

4.1. Hasil Pengumpulan Data

Data penelitian ini diperoleh dari *media sosial X (Twitter)* menggunakan API publik dengan kata kunci “Purbaya Yudhi Sadewa” dan filter bahasa Indonesia (*lang:id*). Periode pengambilan data dilakukan selama bulan Januari hingga November 2025, menghasilkan sebanyak 500 tweet. Setelah dilakukan proses *cleaning* terhadap data duplikat, *spam*, dan konten non-relevan, sebanyak 451 tweet layak digunakan sebagai data utama.

conversation_id_str	created_at	favorite_count	full_text	id_str	image_url	in_reply_to_screen_name	lang	location	quote_count	reply_count	retweet_count	
0	198750443337704564	Sun Nov 09 15:54:21 +0000 2025	0	@direktoridosen Mulai dibegoni Purbaya. Dari awal juga jgn terlalu percaya si Purbaya ini bagus.	1987549326726602785	NaN	direktoridosen	in	NaN	0	0	0
1	19874450586889653	Sun Nov 09 15:53:42 +0000 2025	0	@ARSIPAJA Yg pada Terbus Gimenc2 Purbaya kmn	1987549166017585470	NaN	ARSIPAJA	in	NaN	0	0	0
2	1986701911698084230	Sun Nov 09 15:45:29 +0000 2025	0	@andybalala19 Dari awal pemerintahan beliau sudah banyak mengungkap kasus korupsi di tambah kehadiran pak menteri Purbaya membuat pemerintahan pak Prabowo semakin di cintai rakyat.	1987547095361007735	NaN	andybalala19	in	NaN	0	0	0
3	1987452143594942670	Sun Nov 09 15:36:43 +0000 2025	0	@enasdhuri @BosPurwa Aku tau ntah ya alergi bngt sm pejabat yg marah2 sambil dikeam kadang saya mkr kok jamannya seotah yg kerja bhr kudu dikeam mnyra pas purbaya rame pkrku heleh apalagi sm si maean ini. Ritosa aja wllkota ku dulu saya ketawain pas marah2 diporogatan lampur merah	1987544891392667952	NaN	enasdhuri	in	NaN	0	0	0
4	1987542948335022544	Sun Nov 09 15:29:00 +0000 2025	1	Bonus 10.000 saldo kalo install apk ShopeePay. Caranya Upgrade ke Spay plus login dg akun shopee &amp; masukin kode referral FFSJWV5WP diskon promo freelance job reseh zonasaang	1987542948335022544	NaN	NaN	in	NaN	0	0	0

Gambar 2. Hasil Pengumpulan Data

4.2. Hasil Preprocessing dan Ekstraksi Fitur

Tahap preprocessing dilakukan untuk memastikan kualitas data teks lebih baik sebelum masuk ke proses analisis. Langkah-langkahnya mencakup case folding, tokenisasi, penghapusan stopwords, stemming menggunakan algoritma Sastrawi, serta normalisasi kata-kata yang tidak baku. Setelah proses ini selesai, teks diubah menjadi bentuk numerik melalui teknik TF-IDF (Term Frequency-Inverse Document Frequency), yang berfungsi memberikan bobot lebih pada kata-kata penting yang mampu membedakan sentimen di antara tweet.

	full_text	created_at	c.cleaned_text
0	mulai dibegoin purbaya purbaya purbaya yudhi sadewa purbaya purbaya yudhi sadewa sadewa dari awal juga jgn terlalu percaya si purbaya purbaya purbaya yudhi sadewa purbaya purbaya yudhi sadewa sadewa ini bagus	Sun Nov 09 15:54:21 +0000 2025	mulai dibegoin purbaya purbaya purbaya yudhi sadewa purbaya purbaya yudhi sadewa sadewa dari awal juga jgn terlalu percaya si purbaya purbaya purbaya yudhi sadewa purbaya purbaya yudhi sadewa sadewa ini bagus
1	yg pada terbius gimmic2 purbaya purbaya purbaya yudhi sadewa purbaya purbaya yudhi sadewa sadewa kmrn	Sun Nov 09 15:53:42 +0000 2025	yg pada terbius gimmic2 purbaya purbaya purbaya yudhi sadewa purbaya purbaya yudhi sadewa sadewa kmrn
2	dari awal pemerintahan beliau sudah banyak mengungkap kasus korupsi di tambah kehadiran pak menteri purbaya purbaya purbaya yudhi sadewa purbaya purbaya yudhi sadewa sadewa membuat pemerintahan pak prabowo semakin di cintai rakyat	Sun Nov 09 15:45:29 +0000 2025	dari awal pemerintahan beliau sudah banyak mengungkap kasus korupsi di tambah kehadiran pak menteri purbaya purbaya purbaya yudhi sadewa purbaya purbaya yudhi sadewa sadewa membuat pemerintahan pak prabowo semakin di cintai rakyat
3	aku ntah ya alergi bngt sama pejabat yang marah2 sambil direkam kadang saya mkr kok jamannya seolah yang kerja bnr harus direkam mknya pas purbaya purbaya purbaya yudhi sadewa purbaya purbaya yudhi sadewa sadewa rame pkrku heleh apalagi sama si maman ini risma saja walkota ku dulu saya ketawain pas marah2 diperapatan lumpur merah	Sun Nov 09 15:36:43 +0000 2025	aku ntah ya alergi bngt sama pejabat yang marah2 sambil direkam kadang saya mkr kok jamannya seolah yang kerja bnr harus direkam mknya pas purbaya purbaya purbaya yudhi sadewa purbaya purbaya yudhi sadewa sadewa rame pkrku heleh apalagi sama si maman ini risma saja walkota ku dulu saya ketawain pas marah2 diperapatan lumpur merah
5	purbaya style	Sun Nov 09 15:25:54 +0000 2025	purbaya style

Gambar 3. Hasil Pre Processing

Dari hasil TF-IDF, kata-kata yang paling berkontribusi pada sentimen positif adalah “percaya”, “stabil”, dan “optimis”, sedangkan pada sentimen negatif dominan muncul kata “melemah”, “gagal”, dan “turun”. Kata “rupiah” dan “ekonomi” menjadi kata umum (netral) dengan *weight* tertinggi di semua kelas.

4.3. Kinerja Model Naïve Bayes

Model Multinomial Naive Bayes dipilih dalam penelitian ini karena dikenal mampu mengklasifikasikan teks dengan baik berdasarkan asumsi bahwa setiap kata pada dokumen tidak saling bergantung satu sama lain. Dataset kemudian dipisahkan menjadi dua bagian, yaitu 80% untuk proses pelatihan dan 20% untuk pengujian. Tahap pelatihan dilakukan menggunakan mekanisme fit-transform pada fitur yang telah direpresentasikan dalam bentuk TF-IDF, sementara tahap pengujian bertujuan mengevaluasi sejauh mana model dapat mengenali pola sentimen pada data baru yang belum pernah diproses sebelumnya. Ringkasan hasil evaluasi model dapat dilihat pada Tabel 1.

Tabel 1. Pengujian Model

support	precision	recall	f1-score	support	precision	recall	f1-score		
	negatif	0.92	0.97	0.94		negatif	0.92	0.97	0.94
69	positif	0.60	0.33	0.43	9	positif	0.60	0.33	0.43

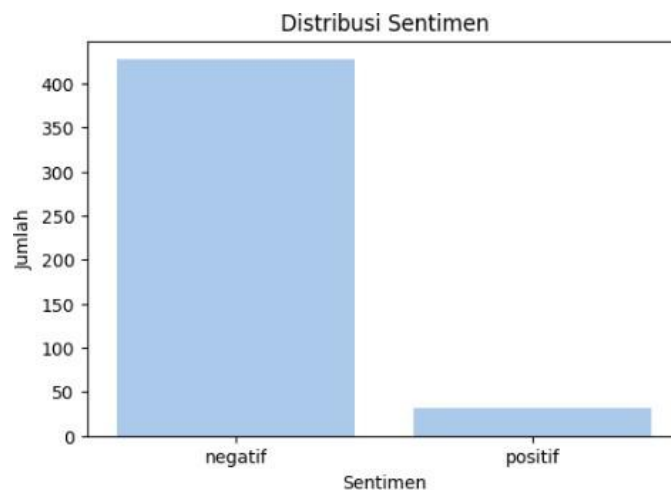
Nilai akurasi keseluruhan model mencapai 91%, yang menunjukkan bahwa algoritma Naive Bayes memiliki performa yang sangat baik dalam mengklasifikasikan opini publik terhadap isu-isu yang berkaitan dengan Purbaya Yudhi Sadewa, khususnya dalam konteks kebijakan ekonomi dan stabilitas makro.

Kinerja tinggi pada kelas *negatif* menunjukkan bahwa model mampu dengan baik mengenali tweet bernada kritik atau ketidakpuasan terhadap kebijakan ekonomi. Sebaliknya, nilai *recall* yang rendah pada kelas *positif* menandakan bahwa masih ada keterbatasan model dalam mendeteksi opini positif yang cenderung menggunakan ekspresi halus atau sarkasme.

Temuan penelitian ini konsisten dengan studi yang dilakukan oleh Ramadhani et al., yang mengungkapkan bahwa algoritma Naive Bayes mampu mengklasifikasikan opini publik mengenai tokoh di media sosial dengan akurasi antara 83–90%. Metode ini unggul karena proses komputasinya yang ringan serta kemampuan generalisasinya yang cukup baik, meskipun masih memiliki keterbatasan ketika harus memahami konteks bahasa yang bersifat ambigu atau mengandung unsur ironi.

4.4. Analisis Sentimen Publik

Tahapan Hasil klasifikasi akhir menunjukkan dominasi sentimen negatif 427, diikuti dan positif 32. Analisis lebih lanjut terhadap kata kunci negatif menunjukkan bahwa sebagian besar kritik publik berkaitan dengan isu kinerja ekonomi nasional, inflasi, dan stabilitas mata uang rupiah, yang dianggap belum optimal.



Gambar 4 Diagram Batang

Sementara itu, sentimen positif umumnya muncul dari pengguna yang mendukung pendekatan analitis dan kebijakan berbasis data yang dilakukan oleh Purbaya Yudhi Sadewa selaku ekonomi pemerintah. Pola ini sejalan dengan hasil penelitian sebelumnya, bahwa persepsi publik terhadap kebijakan ekonomi nasional sangat dipengaruhi oleh kejelasan komunikasi

dan hasil nyata di lapangan. Visualisasi hasil sentimen ditunjukkan pada Gambar 4, di mana diagram batang memperlihatkan distribusi proporsi sentimen pada keseluruhan data.

4.5. Perbandingan Dengan Penelitian Sebelumnya

Visualisasi Tabel 2 memperlihatkan perbandingan performa penelitian ini dengan beberapa studi terdahulu di ranah sosial media menggunakan naive bayes.

Tabel 2. Perbandingan Penelitian Sebelumnya

Peneliti	Objek	Metode	Akurasi	Kelebihan	Kekurangan
Ramadhani et al. (2023) [4]	Tokoh publik nasional	Naive Bayes	83.4%	Data besar, hasil stabil	Tidak fokus pada isu ekonomi
Sari & Putra (2024) [7]	Tokoh politik	SVM	85.2%	Fitur kuat	Kompleksitas tinggi
Penelitian ini (2025)	Purbaya Yudhi Sadewa	Naive Bayes	91.1%	Akurasi tinggi, efisien	Belum mempertimbangkan <i>sentiment context</i>

Dari Tabel 2, terlihat bahwa pendekatan Naive Bayes memberikan hasil kompetitif dengan kompleksitas rendah, sehingga cocok untuk pemantauan opini publik secara real-time. Namun, kelemahannya terletak pada sensitivitas terhadap konteks bahasa sarkasme dan ironi yang masih sulit diidentifikasi oleh model berbasis probabilitas sederhana.

5. Kesimpulan

Penelitian ini berhasil mencapai tujuan yang telah ditetapkan pada bab pendahuluan, yaitu menerapkan metode Naive Bayes untuk mengklasifikasikan sentimen publik terhadap Purbaya Yudhi Sadewa di platform X dengan tahapan data crawling, text preprocessing, feature extraction menggunakan TF-IDF, pelatihan model, dan evaluasi kinerja. Sebagaimana diharapkan, inovasi penerapan model Naive Bayes dalam konteks figur publik ekonomi Indonesia dengan fokus pada opini berbahasa Indonesia telah menghasilkan klasifikasi sentimen yang akurat dengan tingkat akurasi 91,1%, precision 0,92, recall 0,97, dan F1-score 0,94 untuk kelas negatif.

Hasil dan analisis menunjukkan kesesuaian dengan gap penelitian yang diidentifikasi pada pendahuluan. Pertama, penelitian ini berhasil mengisi kekosongan objek penelitian dengan memfokuskan pada persepsi terhadap tokoh ekonomi tertentu, bukan hanya isu kebijakan makro. Kedua, konteks bahasa Indonesia informal di media sosial berhasil ditangani melalui proses preprocessing yang komprehensif. Ketiga, model kontekstual yang

menggabungkan pendekatan Naive Bayes dengan data mentah pengguna X terbukti mampu menghasilkan klasifikasi yang representatif terhadap realitas opini publik, dengan dominasi sentimen negatif (93%) yang mencerminkan ketidakpuasan masyarakat terhadap isu kinerja ekonomi nasional, inflasi, dan stabilitas rupiah.

Berdasarkan hasil dan pembahasan, prospek pengembangan penelitian ini dapat dilakukan melalui beberapa arah. Pertama, integrasi dengan metode deep learning seperti BERT atau LSTM untuk meningkatkan kemampuan deteksi sentimen positif dan menangani konteks bahasa kompleks seperti sarkasme dan ironi. Kedua, penerapan teknik ensemble learning dengan menggabungkan Naive Bayes dan SVM untuk meningkatkan performa klasifikasi pada kelas minoritas. Ketiga, pengembangan sistem monitoring sentimen real-time berbasis dashboard yang dapat membantu stakeholder memahami dinamika opini publik secara responsif terhadap kebijakan ekonomi. Keempat, perluasan studi lanjut dengan analisis komparatif sentimen terhadap beberapa tokoh ekonomi nasional untuk memahami pola persepsi publik secara lebih komprehensif. Penelitian ini membuka peluang penerapan praktis dalam mendukung pengambilan keputusan strategis berbasis analisis sentimen media sosial di sektor ekonomi dan pemerintahan.

Daftar Pustaka

- Andrian, B. W., Adline, F., & Tobing, T. (2023). Implementation Of Naïve Bayes Algorithm In Sentiment Analysis Of Twitter Social Media Users Regarding Their Interest To Pay The Tax. 1733–1742.
- Anton Purwadi, Kismartini Kismartini, Retno Sunu Astuti, & T. A. (2022). Keterlibatan Warga Terhadap Akun Twitter Pemerintah Provinsi Jawa Tengah, Jawa Timur, Jawa Barat : Analisa Data Mining. 11(1), 42–49. <https://doi.org/10.31289/perspektif.v11i1.5283>
- Cahyono, Y. (2019). Analisis Sentiment Tweets Berbahasa Sunda Menggunakan Naive Bayes Classifier dengan Seleksi Feature Chi Squared Statistic. 4(3), 87–94.
- Classifier, F., Kasus, S., & Sinovac, V. (2022). Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 di Twitter Menggunakan Metode Random Forest Classifier (Studi Kasus: Vaksin Sinovac). 8106, 64–72.
- Kaplan, A. M., & Haenlein, M. (2017). Users of the world , unite ! The challenges and opportunities of Social Media. February 2010. <https://doi.org/10.1016/j.bushor.2009.09.003>
- Mario, G., Paridy, C., Aristo, A., Sinlae, J., & Ngaga, E. (2024). Analisis Sentimen di Media Sosial X tentang IKN dengan Naïve Bayes. 417–426.
- No, V., Setiawan, A., & Suryono, R. R. (2024). Analisis Sentimen Ibu Kota Nusantara menggunakan Algoritma Support Vector Machine dan Naïve Bayes. 8(1), 183–192. <https://doi.org/10.29408/edumatic.v8i1.25667>
- Puluhulawa, F. Y. (2022). What Do Opinion Leaders Share from the New Normal on Twitter ? : A Qualitative Content Analysis. 2020, 183–196.

- Sualman, I. (2023). Social Media Usage and Interactions : An In-Depth Interview on BERSIH 2 . 0 Social Media Visual Framing. 39(1), 224-239.
- Susanto, A., & Atho, M. (2021). Sentiment Analysis on Indonesia Twitter Data Using Naïve Bayes and K-Means Method. 6(1), 40-45.