

Analisis Sentimen Calon Gubernur Jawa Tengah 2024 Menggunakan Metode Naïve Bayes

Martin Nuh Hanan¹, Hamid Muhammad Jumasa², Ike Yunia Pasa³
^{1,2,3}Teknologi Informasi, Universitas Muhammadiyah Purworejo, Jawa Tengah
¹nuhhanan823@gmail.com*, ²hamidjumasa@umpwr.ac.id, ³ikeypasa@umpwr.ac.id

Abstract

Social media X (formerly Twitter) has become a digital public space where people freely express opinions, including those related to regional head elections. These opinions can be processed into valuable information to understand public perception of political candidates. This study aims to analyze public sentiment toward candidates for Governor of Central Java for the 2024–2029 period using the Naïve Bayes method. Data were collected through a crawling process on social media X using Tweet-Harvest with relevant keywords. The collected data were then processed through several preprocessing stages, including cleaning, case folding, normalization, stopword removal, tokenization, and stemming. Sentiment labeling was conducted automatically using the TextBlob library, classifying tweets into positive, negative, and neutral categories. The Naïve Bayes method was chosen due to its effectiveness and efficiency in text classification tasks. The experimental results show that the proposed model achieved an accuracy of 90.28% for Andika Perkasa and 84.51% for Ahmad Luthfi based on a 90:10 training and testing data split. Out of 452 analyzed tweets, Andika Perkasa received 350 positive sentiments, slightly higher than Ahmad Luthfi, who obtained 336 positive sentiments. These findings indicate that public perception of both candidates tends to be positive, with a marginal advantage for Andika Perkasa.

Keywords: sentiment analysis, gubernatorial election, social media x, naïve bayes, tweet-harvest

Abstrak

Media sosial X (sebelumnya *Twitter*) telah berkembang menjadi ruang publik digital yang memungkinkan masyarakat menyampaikan opini secara bebas, termasuk terkait pemilihan kepala daerah. Opini tersebut dapat dimanfaatkan sebagai sumber informasi untuk memahami persepsi publik terhadap calon pemimpin. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap calon Gubernur Jawa Tengah periode 2024–2029 menggunakan metode *Naïve Bayes*. Data dikumpulkan melalui proses *crawling* di media sosial X menggunakan *Tweet-Harvest* dengan kata kunci yang relevan. Data mentah selanjutnya melalui tahap *preprocessing* yang meliputi *cleaning*, *case folding*, normalisasi, *stopword removal*, tokenisasi, dan *stemming*. Pelabelan sentimen dilakukan secara otomatis menggunakan pustaka *TextBlob* untuk mengklasifikasikan *tweet* ke dalam kategori positif, negatif, dan netral. Metode *Naïve Bayes* dipilih karena kemampuannya dalam mengklasifikasikan teks secara efisien dan akurat. Hasil pengujian menunjukkan bahwa model menghasilkan akurasi sebesar 90,28% pada data sentimen Andika Perkasa dan 84,51% pada data sentimen Ahmad Luthfi dengan pembagian data latih dan data uji sebesar 90:10. Dari total 452 data yang dianalisis, Andika Perkasa memperoleh 350 sentimen positif, sedangkan Ahmad Luthfi memperoleh 336 sentimen positif. Hasil penelitian ini menunjukkan bahwa persepsi masyarakat terhadap kedua calon cenderung positif, dengan tingkat sentimen positif yang sedikit lebih tinggi pada Andika Perkasa.

Kata kunci: analisis sentimen, pemilihan gubernur, media sosial x, naïve bayes, tweet-harvest

©This work is licensed under a Creative Commons Attribution - ShareAlike 4.0 International License

1. Pendahuluan

Indonesia, salah satu negara yang penganut paham demokrasi, di mana pemilihan umum diselenggarakan secara periodik setiap lima tahun sekali. Salah satu agenda penting dalam sistem demokrasi Indonesia adalah Pemilihan Kepala Daerah atau disebut juga Pilkada yang dilakukan untuk memilih gubernur, bupati, dan wali kota beserta wakilnya [1]. Pada tahun 2024, Indonesia akan kembali menggelar pesta demokrasi, termasuk pemilihan gubernur di Jawa Tengah.

Dalam proses politik, opini publik menjadi faktor sentral yang memengaruhi tingkat popularitas dan elektabilitas seorang calon kepala daerah. Seorang tokoh politik yang ingin maju tentu akan mempertimbangkan pandangan masyarakat terhadap dirinya. Survei opini publik atau *polling* merupakan

salah satu instrumen yang kerap digunakan untuk memprediksi opini masyarakat terhadap isu atau tokoh tertentu [2]. Namun, pada era digital, opini masyarakat semakin mudah ditemukan di media daring, khususnya media sosial.

Platform X (sebelumnya *Twitter*) dijadikan salah satu sarana masyarakat untuk menyampaikan pandangan, berupa dukungan, kritik, ataupun sindiran politik [3]. Setiap hari, ribuan cuitan yang membicarakan calon kepala daerah bermunculan, sehingga menjadi sumber data berharga untuk memahami preferensi publik. Jumlah data yang besar ini membuat analisis manual menjadi tidak efektif. Oleh karena itu, diperlukan metode otomatis yang dapat mengolah data teks dalam jumlah besar secara cepat dan akurat. Salah satu teknik yang digunakan adalah *web scraping*, yaitu proses

memperoleh data dari situs *web* secara *auto* tanpa perlu menyalin secara *manual* [4].

Dalam penelitian ini, *web scraping* digunakan untuk mengumpulkan data dari *X* terkait dua kandidat calon gubernur Jawa Tengah periode 2024–2029, yaitu Andika Perkasa dan Ahmad Luthfi. Data tersebut kemudian dianalisis menggunakan metode *Naïve Bayes*. Metode ini dipilih karena memiliki kemampuan yang mumpuni dalam mengklasifikasikan teks ke dalam kategori positif, negatif, dan netral secara sederhana namun efektif [5]. Analisis sentimen dengan *Naïve Bayes* akan membantu memetakan bagaimana persepsi publik terbentuk di media sosial dan bagaimana sentimen masyarakat terbagi terhadap kedua kandidat.

Masyarakat sering menghadapi kesulitan dalam memilih calon pemimpin yang tepat, karena opini publik di media sosial kerap bercampur dengan propaganda politik, bias, bahkan berita palsu [6]. Selain itu, calon kepala daerah sering kali kesulitan merancang strategi kampanye yang tepat sasaran karena kurang memanfaatkan teknologi analisis sentimen. Melalui analisis sentimen, calon gubernur dapat memperoleh gambaran yang lebih objektif mengenai pandangan publik, sekaligus menyusun kebijakan yang lebih sesuai dengan kebutuhan masyarakat.

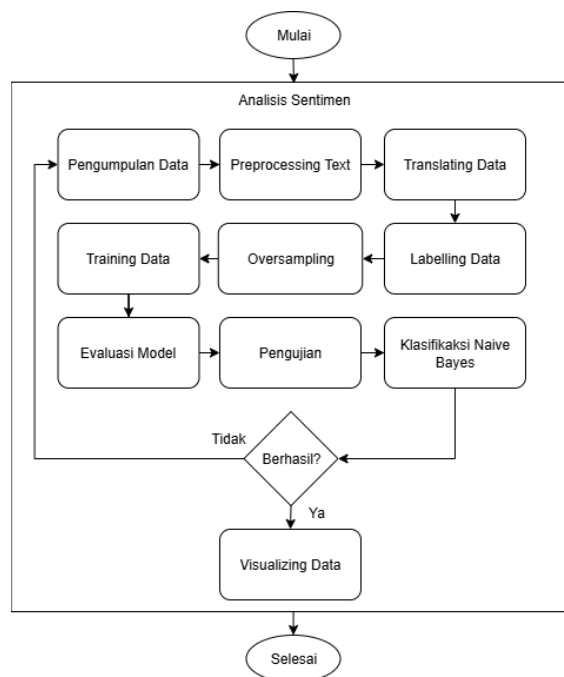
Berdasarkan latar belakang yang telah diuraikan, penelitian ini berfokus pada analisis sentimen masyarakat terhadap calon Gubernur Jawa Tengah periode 2024–2029 menggunakan metode *Naïve Bayes*. Penelitian ini dirancang untuk menjawab permasalahan terkait bagaimana opini masyarakat di media sosial *X* dapat dianalisis dan diklasifikasikan secara otomatis ke dalam kategori sentimen positif, negatif, dan netral.

Selain itu, penelitian ini juga mengkaji bagaimana kinerja metode *Naïve Bayes* dalam mengklasifikasikan sentimen opini masyarakat terhadap masing-masing calon gubernur. Adapun tujuan dari penelitian ini adalah untuk mengumpulkan data opini masyarakat melalui media sosial *X* menggunakan teknik *web scraping*, menganalisis serta mengklasifikasikan sentimen masyarakat terhadap calon Gubernur Jawa Tengah dengan metode *Naïve Bayes*, mengukur kinerja metode tersebut dalam klasifikasi sentimen berdasarkan tingkat akurasi, serta menyajikan hasil analisis yang dapat memberikan gambaran umum mengenai persepsi masyarakat terhadap calon Gubernur Jawa Tengah periode 2024–2029.

2. Metode Penelitian

Langkah-langkah penelitian ini disusun untuk memperoleh hasil analisis sentimen yang akurat dan dapat dipertanggungjawabkan. Secara umum, alur penelitian mencakup tahap identifikasi masalah, pengumpulan data, *preprocessing* teks, pelabelan data,

pelatihan model, evaluasi kinerja, klasifikasi sentimen, hingga visualisasi hasil analisis.



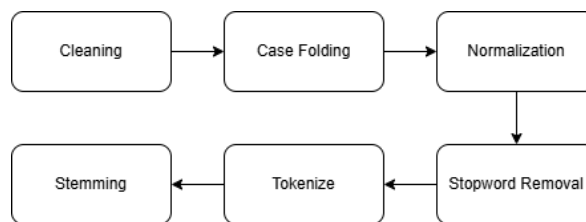
Gambar 1. Alur Penelitian

2.1 Pengumpulan Data

Data dalam penelitian ini diperoleh dari media sosial *X*. Platform ini dipilih karena bersifat terbuka, dinamis, dan banyak digunakan masyarakat untuk mengekspresikan opini politik. Proses pengumpulan data dilakukan dengan teknik *crawling* menggunakan *X API*, yaitu *interface* aplikasi yang memungkinkan peneliti membuka data secara legal dan terstruktur. Pengambilan data dilakukan dengan alat bernama *Tweet-Harvest*. Tools ini dapat dipakai dengan menentukan parameter tertentu, seperti kata kunci nama kandidat, tagar populer yang berkaitan dengan Pemilihan Gubernur Jawa Tengah 2024, rentang waktu, serta jumlah maksimum *tweet*. Data hasil *crawling* akan disimpan di dalam format terstruktur untuk dapat dianalisis lebih jauh.

2.2 Preprocessing Data

Sebelum dianalisis, data mentah melalui tahap *preprocessing*. Tahapan ini adalah tahapan awal yang paling krusial dalam melakukan analisis sentimen terutama menggunakan data *X*, karena dapat mempengaruhi hasil dari klasifikasi [7]. Tahap ini mencakup beberapa langkah penting seperti yang dapat disimak pada Gambar 2.2.



Gambar 2. Alur Preprocessing

2.2.1 Cleaning Data

Cleaning data bertujuan untuk menghilangkan unsur-unsur yang dapat mengganggu proses klasifikasi sentimen, seperti *tag* (#), *mention* (@), *URL*, *emoji*, dan *symbol* lainnya. Proses ini diharapkan untuk mengurangi *noise* seperti *html*, koneksi, *hashtag*, *username*, dan konten [8].

2.2.2 Case Folding

Case folding dilakukan untuk mengganti semua huruf yang ada dalam *tweet* menjadi huruf kecil. Tujuannya adalah untuk menyamakan bentuk kata, sehingga kata “Gubernur”, “gubernur”, dan “GUBERNUR” dianggap sama. Karakter dari setiap kata dalam data perlu distandarisasi [9].

2.2.3 Normalization

Normalization ialah tahap mengganti *slang word* atau variasi penulisan menjadi bentuk asli atau standar. Contohnya: “nggak” menjadi “tidak”, “bgt” menjadi “banget”, atau “gk” menjadi “tidak”. Tahapan ini penting karena banyak teks dari media sosial, seperti *X*, menggunakan bahasa *informal* atau *slang* yang bisa memengaruhi akurasi model jika tidak diseragamkan. Proses ini dinilai meningkatkan akurasi pada analisis sentimen [10].

2.2.4 Stopword Removal

Stopword removal berfungsi untuk menghilangkan kata yang muncul berulang namun tidak memiliki nilai penting untuk dianalisis, seperti “dan”, “di”, “yang”, “adalah”, dan lain-lain. Kata berulang yang tidak signifikan dan tidak relevan juga dihapus, seperti konjungsi dan kata ganti kepemilikan dan pribadi [11]. Hal ini membantu mengurangi *noise* dalam data.

2.2.5 Tokenizing

Tokenizing ialah proses pemecahan kalimat. Proses ini dilakukan untuk memecah kalimat menjadi kata-kata atau token menggunakan spasi atau tanda baca [12]. Tahapan ini sangat penting untuk memisahkan kata-kata dalam *text* sehingga lebih mudah dianalisis secara individu. Misalnya, kalimat “Saya memilih calon gubernur terbaik” akan dipecah menjadi [“Saya”, “memilih”, “calon”, “gubernur”, “terbaik”].

2.2.6 Stemming

Stemming diperlukan untuk mengganti kata ke bentuk aslinya. Tahap ini dilakukan dengan menghapus kata imbuhan yaitu awalan dan akhiran [13]. Misalnya, kata “berlari”, “pelari”, dan “lari-lari” disederhanakan sebagai “lari”. Tujuannya adalah agar kata-kata dengan makna dasar yang sama tidak dianggap berbeda oleh sistem.

2.3 Translating Data

Translating data dilakukan untuk menerjemahkan teks berbahasa Indonesia ke bahasa Inggris. Langkah ini dipertimbangkan karena banyaknya penggunaan *library* analisis sentimen berbasis bahasa Inggris.

Dalam penelitian ini, *TextBlob* lah yang digunakan untuk proses pelabelan. *Library* ini ialah salah satu contoh *library* analisis sentimen yang dapat bekerja lebih baik apabila data set menggunakan bahasa Inggris. Oleh sebab itu, tahap ini mampu meningkatkan akurasi jika data tersedia dalam bahasa yang sesuai.

2.4 Labelling Data

Setelah proses penerjemahan, data kemudian diberi label untuk menentukan kelas sentimen positif, negatif, atau netral pada setiap *tweet*. Penelitian ini menggunakan *TextBlob*, sebuah pustaka *Python* yang menyediakan analisis sentimen berbasis *lexicon*. *TextBlob* menghasilkan dua nilai utama, yaitu *polarity* dan *subjectivity*. Nilai *polarity* berada pada rentang -1 hingga 1, angka sama dengan atau kurang dari -1 menunjukkan sentimen negatif, angka sama dengan atau lebih dari 1 menunjukkan sentimen positif, dan angka mendekati nol mengindikasikan sentimen netral. Sedangkan *subjectivity* menggambarkan sejauh mana opini bersifat subjektif atau faktual.

$$S = \frac{\sum_{i=1}^n \text{score}(w_i)}{n} \quad (2.1)$$

dengan *P* adalah nilai *polarity* keseluruhan dari suatu teks. Variabel w_i merupakan kata ke-*i* dalam teks yang terdeteksi dalam kamus sentimen (*lexicon*). Variabel $\text{score}(w_i)$ menunjukkan skor sentimen dari kata w_i , dengan rentang nilai antara -1 yang merepresentasikan sentimen sangat negatif hingga +1 yang merepresentasikan sentimen sangat positif. Variabel *n* adalah jumlah kata dalam teks yang memiliki skor sentimen.

Interpretasi dari nilai *polarity* *P* adalah sebagai berikut: apabila $P > 0$, maka teks dikategorikan sebagai sentimen positif. Apabila $P < 0$, maka teks dikategorikan sebagai sentimen negatif. Sedangkan apabila $P = 0$, maka teks dikategorikan sebagai sentimen netral.

2.5 Oversampling Data

Data yang tidak seimbang dapat memengaruhi performa model. Oleh karena itu, dilakukan proses *oversampling* untuk menambah data minoritas supaya sebanding dengan data mayoritas. Langkah ini membantu model mengenali pola pada setiap kelas secara proporsional dan mengurangi bias.

Oversampling memiliki beberapa macam cara salah satu contohnya ialah augmentasi berbasis *synonym replacement* dengan *WordNet*. Teknik ini sejalan dengan praktik yang dianjurkan oleh konsep *SMOTE* dan *sampling* seimbang dalam literatur statistik, di mana *oversampling* terbukti mampu memperbaiki bias model terhadap kelas mayoritas [14].

2.6 Pelatihan Model

Tahap berikutnya adalah pelatihan (*training*) data. Pada tahap ini, data hasil dari *preprocessing* dan pelabelan digunakan untuk melatih model klasifikasi

menggunakan algoritma *Naïve Bayes*. Rumus *Naïve Bayes* dijabarkan sebagai berikut ini:

$$P(C_k | X) = \frac{P(X | C_k) \cdot P(C_k)}{P(X)} \quad (2.2)$$

dengan $P(C_k | X)$ sebagai probabilitas dokumen X termasuk ke dalam kelas C_k (misalnya positif, negatif, atau netral). Variabel $P(X | C_k)$ adalah probabilitas kemunculan dokumen X jika diketahui kelasnya adalah C_k . $P(C_k)$ adalah probabilitas awal (*prior probability*) sebuah kelas C_k . dan $P(X)$ sebagai probabilitas total dari dokumen X .

Karena dokumen teks terdiri dari kata-kata, maka digunakan asumsi independensi (Naïf):

$$P(C_k | X) \propto P(C_k) \prod_{i=1}^n P(w_i | C_k) \quad (2.3)$$

dengan n adalah jumlah kata dalam dokumen X . Variabel w_i sebagai kata ke- i dalam dokumen. Dan $P(w_i | C_k)$ sebagai probabilitas kata w_i muncul pada kelas C_k .

Proses pelatihan bertujuan untuk mengajarkan model mengenali pola-pola tertentu dalam data sehingga mampu mengklasifikasikan sentimen *tweet* sesuai label yang telah diberikan.

2.7 Evaluasi Model

Setelah model selesai dilatih, dilakukan evaluasi untuk menilai seberapa baik kinerjanya pada data uji. Pada tahap ini, prediksi dari model yang telah dilatih sebelumnya timbang dengan label sebenarnya menggunakan *confusion matrix*. Metrik yang digunakan antara lain *accuracy*, *precision*, *recall*, dan *F1-score*. Evaluasi dapat memberikan gambaran secara penuh tentang kemampuan model mengidentifikasi sentimen secara tepat, terutama apabila distribusi kelas tidak seimbang.

Untuk mendapatkan hasil evaluasi yang optimal, pembagian *dataset* dilakukan dengan beberapa skenario rasio data latih (*train set*) dan data uji (*test set*), misalnya 90:10, 80:20, 70:30, 60:40, hingga 50:50. Dengan berbagai skenario ini, dapat dianalisis pengaruh proporsi data latih dan uji terhadap performa model serta kestabilan hasil evaluasi.

2.8 Pengujian Model

Tahap selanjutnya yaitu pengujian (*testing*) model. Pada tahap ini, *tweet* baru dimasukkan ke dalam model yang telah dilatih untuk memeriksa apakah analisis sentimen berjalan dengan benar. Pengujian bertujuan untuk mengevaluasi kemampuan model saat melakukan klasifikasi sentimen pada data yang benar-benar belum pernah dilihat sebelumnya. Dengan demikian, dapat dipastikan bahwa model mampu memberikan prediksi yang akurat dan konsisten sesuai dengan pola yang telah dipelajari selama proses pelatihan.

2.9 Klasifikasi Naïve Bayes

Setelah model diuji dan dianggap sesuai, tahap berikutnya adalah klasifikasi sentimen. *Dataset* berlabel dimasukkan ke algoritma *Naïve Bayes* untuk menghasilkan prediksi sentimen pada *tweet* baru. Penerapan metode *ensemble*, terutama *majority-voting* dari beberapa model pralatih seperti *multilingual BERT* dan *XLM-RoBERTa*, dapat meningkatkan akurasi klasifikasi sentimen pada korpus multibahasa (termasuk *dataset tweet* berbahasa Arab dan Inggris) hingga 3–5 % dibandingkan dengan model tunggal [15]. Temuan ini menegaskan bahwa tahap klasifikasi dengan pendekatan *ensemble* memiliki peran krusial dalam keberhasilan keseluruhan *pipeline* analisis sentimen.

Dalam penelitian ini, model yang digunakan berasal dari skenario terbaik, yaitu model dengan rasio pembagian data latih dan uji yang memberikan nilai evaluasi tertinggi dari lima skenario rasio yang diuji. Dengan demikian, model yang dipilih diharapkan mampu menghasilkan prediksi sentimen yang paling akurat dan andal dalam menganalisis opini publik terhadap calon gubernur Jawa Tengah.

2.10 Visualisasi Data

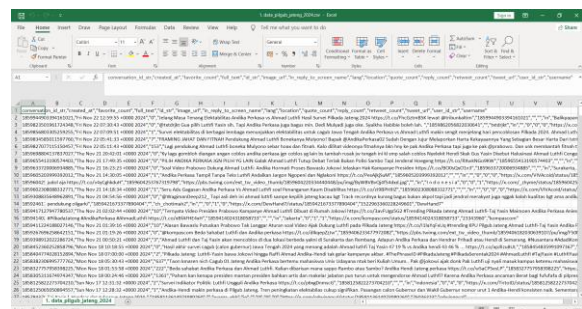
Data yang telah diklasifikasikan kemudian divisualisasi, yaitu menampilkan hasil klasifikasi dalam bentuk grafik batang dan *word cloud* untuk memudahkan pemahaman distribusi sentimen. Visualisasi *real-time* dapat memberikan pemahaman yang lebih baik terhadap pandangan masyarakat yang berkembang di media sosial [16].

Visualisasi ini berfungsi untuk menyajikan gambaran yang lebih intuitif mengenai persebaran opini, baik sentimen positif, negatif, maupun netral, sehingga hasil analisis tidak hanya bersifat numerik tetapi juga mudah dipahami oleh pembaca maupun pengambil keputusan.

3. Hasil dan Pembahasan

3.1. Pengumpulan Data

Pengumpulan data yang dilakukan di penelitian analisis sentimen Calon Gubernur dilakukan menggunakan suatu alat yang bernama *Tweet-Harvest*. Alat ini berbasis *Playwright* yang bekerja dengan cara mengumpulkan data dari hasil pencarian X .



Gambar 3. Carwing Data dalam Bentuk file CSV

Pada Gambar 3.1 adalah hasil data *tweet* yang diperoleh melalui teknik *scrapping/crawling* data. Data hasil *scrapping/crawling* berjumlah 462 data. Data ini lalu disimpan di sebuah file berbentuk *CSV*.

3.2. Preprocessing Data

Tahap *preprocessing* ini digunakan kepada *dataset* yang telah disimpan dalam bentuk file CSV yaitu pada kolom "*full_text*". Data yang ada di kolom "*full_text*" dilakukan berbagai macam tahap *preprocessing* supaya bisa dilakukan analisis sentimen dengan lebih akurat.

3.2.1. Cleaning Data

Pada tahap ini bagian-bagian yang tidak penting seperti tagar (#), *mention* (@), *URL*, *emoji*, serta simbol lainnya dihapus.

Tabel 1.Cleaning Data

Sebelum	Sesudah
Jelang Masa Tenang Elektabilitas Andika Perkasa vs Ahmad Luthfi Hasil Survei Pilkada Jateng 2024 https://t.co/FhcGzmb5K lewat @tribunkaltim	Jelang Masa Tenang Elektabilitas Andika Perkasa vs Ahmad Luthfi Hasil Survei Pilkada Jateng 2024 lewat

3.2.2. Case Folding

Pada tahap ini semua teks yang ada dalam kolom "*full_text*" diganti ke huruf kecil supaya variasi penulisan kata dapat disamakan.

Tabel 2.Case Folding

Sebelum	Sesudah
Jelang Masa Tenang Elektabilitas Andika Perkasa vs Ahmad Luthfi Hasil Survei Pilkada Jateng 2024 lewat	jelang masa tenang elektabilitas andika perkasa vs ahmad luthfi hasil survei pilkada jateng 2024 lewat

3.2.3. Normalization

Pada tahap ini kata tidak baku atau singkatan dikembalikan ke bentuk baku, misalnya "gk" menjadi "tidak".

Tabel 3.Normalization

Sebelum	Sesudah
jelang masa tenang elektabilitas andika perkasa vs ahmad luthfi hasil survei pilkada jateng 2024 lewat	jelang masa tenang elektabilitas andika perkasa lawan ahmad luthfi hasil survei pilkada jateng 2024 lampau

3.2.4. Stopword Removal

Pada tahap ini kata umum yang tidak memiliki nilai penting seperti "dan", "di", atau "yang" dihilangkan.

Tabel 4.Stopword Removal

Sebelum	Sesudah
jelang masa tenang elektabilitas andika perkasa lawan ahmad luthfi hasil	jelang masa tenang elektabilitas andika perkasa lawan ahmad luthfi hasil

survei pilkada jateng 2024 lampau	survei pilkada jateng 2024 lampau
--------------------------------------	--------------------------------------

3.2.5. Tokenizing

Pada tahap ini *tweet* pada kolom "*full_text*" dipecah jadi kata tunggal untuk mempermudah analisis.

Tabel 5.Tokenizing

Sebelum	Sesudah
jelang masa tenang elektabilitas andika perkasa lawan ahmad luthfi hasil survei pilkada jateng 2024 lampau	['jelang', 'masa', 'tenang', 'electabilitas', 'andika', 'perkasa', 'lawan', 'ahmad', 'luthfi', 'hasil', 'survei', 'pilkada', 'jateng', '2024', 'lampau']

3.2.6. Stemming

Pada tahap ini kata diubah menjadi bentuk dasar, misalnya "berlari" menjadi "lari".

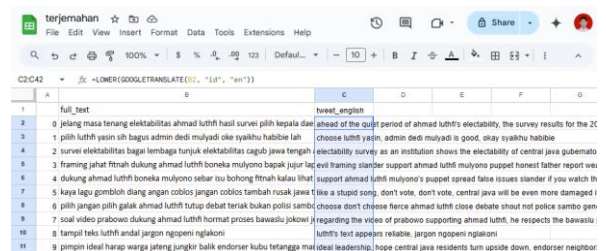
Tabel 6.Stemming

Sebelum	Sesudah
jelang masa tenang elektabilitas andika perkasa lawan ahmad luthfi hasil survei pilkada jateng 2024 lampau	jelang masa tenang elektabilitas andika perkasa lawan ahmad luthfi hasil survei pilkada jateng 2024 lampau

3.3. Translating Data

Data yang telah melewati tahapan stemming kemudian akan melalui tahap translating data. Tahapan translating merupakan proses di mana data teks berbahasa Indonesia diterjemahkan dalam bahasa Inggris. Hal ini diperlukan karena *tools* analisis sentimen pada penelitian ini adalah *TextBlob*, secara default hanya mendukung pengolahan teks berbahasa Inggris. Dengan melakukan proses penerjemahan, data menjadi kompatibel dengan sistem analisis *TextBlob* sehingga hasil klasifikasi sentimen dapat diperoleh secara akurat.

Proses penerjemahan ini dilakukan secara otomatis menggunakan *Google Sheets*. Aplikasi *spreadsheet* ini berbasis *cloud* yang dikembangkan oleh *Google* sebagai bagian dari *Google Workspace* yang tersedia dan dapat diakses secara gratis di *Google*. Aplikasi ini dipilih karena menyediakan *tools* atau layanan penerjemahan gratis sampai 10.000 baris.



Gambar 3.Translating Data

3.4. Labelling Data

Tahapan *labelling* merupakan proses di mana label sentimen diberikan pada data yang telah dibagi dua yaitu data milik Andika Perkasa dan Ahmad Lutfi. Kedua data teks ini akan diberikan label berdasarkan hasil analisis menggunakan *TextBlob*. Setelah data text diterjemahkan ke dalam bahasa Inggris, *TextBlob* digunakan untuk menghitung nilai polaritas dari masing-masing teks.

Proses diawali dengan mengambil semua teks *tweet* dalam bahasa Inggris yang disimpan dalam kolom *tweet_english* dari *dataset*, lalu dikonversi menjadi *list* dengan nama *data_tweet*. Selanjutnya, dilakukan inisialisasi nilai polaritas awal sebesar nol, serta beberapa variabel penghitung untuk mencatat jumlah *tweet* dengan sentimen Positif, Negatif, dan Netral.

Setiap *tweet* dalam *list data_tweet* kemudian dianalisis satu per satu menggunakan *TextBlob*. Proses ini menghasilkan nilai polaritas, yaitu angka desimal antara -1 hingga 1 yang menggambarkan kecenderungan sentimen suatu teks. Nilai polaritas positif menunjukkan bahwa *tweet* mengandung sentimen positif, nilai nol menandakan netral, sedangkan nilai negatif menunjukkan adanya sentimen negatif. Berdasarkan nilai ini, setiap *tweet* akan melewati proses klasifikasi ke dalam tiga kategori sentimen tersebut. Jumlah total untuk masing-masing kategori diperbarui setiap kali klasifikasi dilakukan, dan status sentimen untuk tiap *tweet* disimpan dalam list bernama *status*.

	full_text	tweet_english	klasifikasi
0	jelang masa tenang elektabilitas andika perkasa...	Ahead of the quiet period, Andika Perkasa's el...	Netral
1	gua pilih lutfi yasin sih tapi andika perkasa...	I chose lutfi yasin but Andika Perkasa is als...	Positif
2	survei elektabilitas di bagai lembaga tunjuk e...	The electability survey as an institution dete...	Netral
3	framing jahat dan fitnah dukung boneka mulyono...	evil and slanderous framing supports mulyono d...	Negatif
4	lagi dukung boneka mulyono sebar hoax dan fitn...	I'm still supporting Mulyono's puppets by spre...	Positif
...
447	pdip yakin andika perkasa bisa lebih unggul la...	Pdip believes Andika Perkasa can outperform il...	Netral
448	gw juga yakin jika andika perkasa jadi cagub j...	I'm also sure that if Andika Perkasa becomes C...	Positif
449	pdip yakin andika perkasa bisa lebih unggul la...	Pdip believes Andika Perkasa can outperform il...	Netral
450	pdip yakin andika perkasa bisa lebih unggul la...	Pdip believes Andika Perkasa can outperform il...	Netral
451	aku dukung jend purn andika yg punya darah mag...	I support Gen. Purn Andika who has Magelang bl...	Netral

Gambar 5. Labelling Data Andika Perkasa

	full_text	tweet_english	klasifikasi
0	jelang masa tenang elektabilitas vs ahmad luth...	ahead of the quiet period electability vs Ahma...	Netral
1	gua pilih lutfi yasin sih bagus min dedi mul...	I choose Lutfi Yasin, he's good, Min Dedi Mul...	Positif
2	survei elektabilitas bagai lembaga tunjuk elek...	Electability survey as an institution to deter...	Positif
3	framing jahat fitnah dukung ahmad lutfi bonek...	evil framing, slander, support Ahmad Lutfi, M...	Positif
4	dukung ahmad lutfi boneka mulyono sebar hoax ...	support Ahmad Lutfi, Mulyono doll spreads sia...	Positif
...
447	pdip yakin lebih unggul lawan ahmad lutfi pil...	Pdip believes it is superior to Ahmad Lutfi L...	Positif
448	gw yakin jadi cagub jateng bellau unggul ahmad...	I'm sure that as Central Java gubernatorial ca...	Positif
449	pdip yakin lebih unggul lawan ahmad lutfi pil...	Pdip believes it is superior to Ahmad Lutfi L...	Positif
450	pdip yakin lebih unggul lawan ahmad lutfi pil...	Pdip believes it is superior to Ahmad Lutfi L...	Positif
451	aku dukung jend purn andika yg punya darah mag...	I support Gen. Purn Andika who has Magelang bl...	Netral

Gambar 6. Labelling Data Ahmad Lutfi

3.5. Oversampling

Pada tahap ini, peneliti melakukan augmentasi data sentimen dengan pendekatan sederhana namun efektif, yakni menggunakan pergantian kata dengan sinonim. Tujuannya adalah untuk menyeimbangkan jumlah data dalam masing-masing kategori sentimen, terutama jika terdapat ketimpangan antara kelas positif, netral, dan negatif.

Pendekatan ini perlu dilakukan pada penelitian ini karena jumlah data pelatihan tidak seimbang yaitu data positif sekitar 60%, data netral sekitar 30%, dan data negatif hanya sekitar 10%. Hal ini dapat mempengaruhi kinerja model pembelajaran mesin di mana mesin akan lebih menilai data sebagai sentimen yang lebih mayoritas yaitu sentimen positif.

3.6. Pelatihan Model *Naïve Bayes*

Pelatihan model dilakukan dengan algoritma *Naïve Bayes* terhadap data teks yang bersumber dari media sosial *X*, dengan objek utama dua tokoh publik yaitu Andika Perkasa dan Ahmad Luthfi. Kategori sentimen yang digunakan mencakup positif, negatif, dan netral. Proses pelatihan model merupakan tahap di mana algoritma *Naïve Bayes* dibekali dengan data terlabel yang telah melalui proses pra-proses dan vektorisasi sehingga dapat mengenali pola dalam data teks yang berbeda.

Dalam penelitian ini, pelatihan model dilakukan dengan lima rasio pembagian data latih dan data uji, yaitu 90:10, 80:20, 70:30, 60:40, dan 50:50. Variasi rasio ini digunakan untuk mengamati bagaimana ukuran dari data latih dan data uji dapat berperan dalam performa model dalam mengklasifikasikan sentimen.

Tabel 7. Performa Model Andika Perkasa

Pembagian Rasio	Accuracy	Precision	Recall	F1-score
Rasio 90:10	90.28%	93%	81%	82%
Rasio 80:20	82.64%	83%	71%	71%
Rasio 70:30	83.33%	82%	72%	73%
Rasio 60:40	81.94%	86%	72%	72%
Rasio 50:50	80.22%	77%	69%	69%

Berdasarkan hasil pengujian pada Tabel 7, performa model *Naïve Bayes* dalam mengklasifikasikan sentimen terhadap calon Gubernur Jawa Tengah Andika Perkasa menunjukkan hasil yang cukup baik pada berbagai skema pembagian data latih dan data uji. Pembagian data dengan rasio 90:10 menghasilkan nilai akurasi tertinggi sebesar 90,28%, dengan *precision* sebesar 93%, *recall* 81%, dan *f1-score* 82%. Tingginya nilai akurasi dan *precision* pada rasio ini menunjukkan bahwa model mampu mengklasifikasikan sentimen dengan tingkat kesalahan yang relatif rendah ketika jumlah data latih lebih dominan.

Tabel 8. Performa Model Ahmad Luthfi

Pembagian Rasio	Accuracy	Precision	Recall	F1-score
Rasio 90:10	84.51%	83%	75%	75%
Rasio 80:20	80.85%	82%	74%	74%
Rasio 70:30	79.15%	81%	71%	72%
Rasio 60:40	75.80%	76%	68%	68%
Rasio 50:50	75.78%	75%	67%	67%

Hasil pengujian performa model *Naïve Bayes* terhadap data sentimen calon Gubernur Jawa Tengah Ahmad Luthfi yang ditampilkan pada Tabel 8 menunjukkan

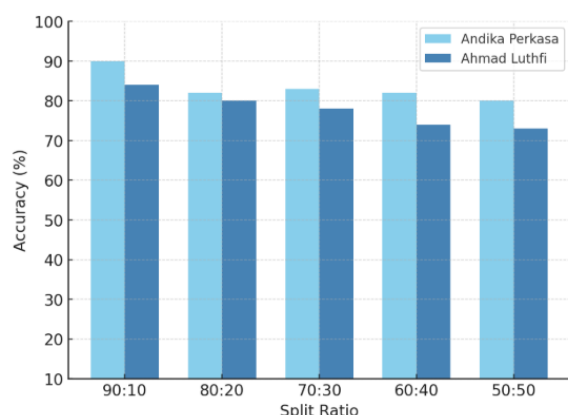
pola yang relatif serupa dengan hasil pada Tabel 7. Pembagian data dengan rasio 90:10 menghasilkan performa terbaik dengan nilai akurasi sebesar 84,51%, *precision* 83%, *recall* 75%, dan *f1-score* 75%. Hasil ini menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam mengklasifikasikan sentimen masyarakat ketika didukung oleh jumlah data latih yang besar.

Namun, dibandingkan dengan hasil pada data Andika Perkasa, nilai akurasi dan metrik evaluasi lainnya pada data Ahmad Luthfi cenderung lebih rendah pada seluruh rasio pembagian data. Perbedaan ini dapat disebabkan oleh karakteristik data sentimen yang lebih beragam atau tingkat ambiguitas opini yang lebih tinggi. Penurunan performa model juga terlihat ketika proporsi data latih dikurangi, yang menunjukkan bahwa ukuran data latih memiliki pengaruh signifikan terhadap kinerja model dalam mengklasifikasikan sentimen secara akurat.

3.7. Evaluasi Model *Naïve Bayes*

Evaluasi performa model dilakukan dengan menilai hasil klasifikasi pada tiap skenario pembagian data latih dan data uji. Evaluasi yang dipakai mencakup *accuracy*, *precision*, *recall*, dan *F1-score* yang dihitung dengan *classification_report* dari *library Scikit-learn* [17], [18]. Visualisasi hasil evaluasi ditampilkan dalam bentuk grafik batang menggunakan *Seaborn* dan *Matplotlib* untuk menunjukkan perbandingan akurasi pada tiap skenario.

Berdasarkan hasil evaluasi, rasio 90:10 menghasilkan akurasi tertinggi baik pada klasifikasi sentimen untuk Andika Perkasa maupun Ahmad Luthfi. Temuan ini menunjukkan bahwa model mampu belajar lebih baik ketika diberi lebih banyak data latih, serta tetap mempertahankan kinerja saat diuji pada data uji yang lebih sedikit. Oleh karena itu, skenario 90:10 dipilih sebagai rasio optimal yang digunakan pada tahap pengujian dan klasifikasi lanjutan.



Gambar 6. Perbandingan Akurasi Model

3.8. Pengujian Model pada Tweet Baru

Pada tahap pengujian, model *Naïve Bayes* yang telah dilatih dengan akurasi tertinggi yaitu skenario 90:10

digunakan untuk memprediksi sentimen dari beberapa *tweet* baru yang tidak terdapat dalam data latih. Contoh *tweet* seperti “Andika tidak bisa kerja” dan “Rambo hebat!” diubah terlebih dahulu menjadi vektor numerik menggunakan *CountVectorizer*, kemudian diprediksi oleh model.

Pengujian ini dilakukan untuk memastikan bahwa model bekerja dengan semestinya. Hasil menunjukkan jika model dari pengujian mampu mengenali pola bahasa dengan cukup baik dan memberikan prediksi sentimen yang sesuai dengan konteks *tweet*.

```
# Cetak hasil
for tweet, pred in zip(new_tweets, new_pred):
    print(f"Tweet: '{tweet}' -> Sentimen: {pred}")
```

Tweet: 'Andika kerja nyata' -> Sentimen: Positif
 Tweet: 'Rambo hebat!' -> Sentimen: Negatif
 Tweet: 'Andika bersama rakyat' -> Sentimen: Positif

Gambar 7. Pengujian Model Andika Perkasa

```
# Cetak hasil
for tweet, pred in zip(new_tweets, new_pred):
    print(f"Tweet: '{tweet}' -> Sentimen: {pred}")
```

Tweet: 'Lutfi kerja nyata' -> Sentimen: Positif
 Tweet: 'Ladusing hebat!' -> Sentimen: Positif
 Tweet: 'Lutfi bersama rakyat' -> Sentimen: Positif

Gambar 8. Pengujian Model Ahmad Luthfi

3.9. Klasifikasi Sentimen Calon Gubernur

Model dengan skenario 90:10 kemudian digunakan untuk melakukan klasifikasi terhadap *dataset*. Pada tahap ini, setiap *tweet* dalam *dataset* diberi label sentimen baru menggunakan metode *Naïve Bayes*. Hasil dari proses ini adalah *dataset* yang telah terklasifikasi ke dalam tiga kategori sentimen, yaitu positif, netral, dan negatif. Jumlah masing-masing sentimen kemudian dihitung secara otomatis.

full_text	tweet_english	klasifikasi	klasifikasi_naive_bayes
jejang masa tenang elektabilitas andika perkasa...	ahead of the quiet period of electability, and...	Positif	Positif
pilih sih andika perkasa bagus admin dedi mulyad...	fd choose andika perkasa, good admin dedi mu...	Positif	Positif
survei elektabilitas bagi lembaga tunjuck elek...	electability survey as an institution shows th...	Positif	Positif
framing jahat fitnah dukung boneka mulyono bap...	evil framing slander support mulyono's puppet...	Positif	Positif
dukung boneka mulyono sebar isu bohong fitnah ...	support mulyono's puppet spread false issues a...	Positif	Positif

Gambar 9. Klasifikasi *Naïve Bayes* Andika Perkasa

Setelah proses klasifikasi sentimen dilakukan, langkah selanjutnya adalah menghitung jumlah data pada masing-masing kategori sentimen untuk mengetahui distribusi sentimen masyarakat terhadap setiap calon gubernur.

full_text	tweet_english	klasifikasi	klasifikasi_naive_bayes
jetang masa tenang elektabilitas ahmad luthfi ...	ahead of the quiet period of ahmad luthfi's et...	Netral	Positif
pilih luthfi yasin sih bagus admin dedi mulyad...	choose luthfi yasin, admin dedi mulyadi is goo...	Positif	Positif
survei elektabilitas bagi lembaga tunjuck elek...	electability survey as an institution shows th...	Positif	Positif
framing jahat fitnah dukung ahmad luthfi bonek...	evil framing slander support ahmad luthfi muly...	Positif	Positif
dukung ahmad luthfi boneka mulyono sebar isu b...	support ahmad luthfi mulyono's puppet spread f...	Negatif	Negatif

Gambar 10. Jumlah *Naïve Bayes* Ahmad Luthfi

Gambar 10 menunjukkan hasil perhitungan jumlah sentimen terhadap calon Gubernur Jawa Tengah Andika Perkasa berdasarkan hasil klasifikasi *Naïve Bayes*. Dari hasil tersebut dapat diketahui bahwa sentimen positif mendominasi dibandingkan dengan

sentimen netral dan negatif. Dominasi sentimen positif ini mengindikasikan bahwa mayoritas opini masyarakat di media sosial X terhadap Andika Perkasa cenderung bersifat mendukung, meskipun masih terdapat sejumlah opini netral dan negatif yang mencerminkan adanya pandangan yang beragam dari masyarakat.

Selain menganalisis distribusi sentimen pada Andika Perkasa, penelitian ini juga melakukan proses klasifikasi dan perhitungan sentimen terhadap calon gubernur lainnya, yaitu Ahmad Luthfi, untuk memperoleh gambaran perbandingan persepsi publik.

```
# 6. Hitung jumlah tiap kategori sentimen
sentiment_counts = data_new['klasifikasi_naive_bayes'].value_counts()
print(sentiment_counts)

klasifikasi_naive_bayes
Positif    332
Netral     78
Negatif    42
Name: count, dtype: int64
```

Gambar 11. Jumlah Sentimen Andika Perkasa

Gambar 11 menampilkan hasil perhitungan jumlah sentimen terhadap calon Gubernur Jawa Tengah Ahmad Luthfi berdasarkan klasifikasi *Naïve Bayes*. Hasil yang ditampilkan menunjukkan bahwa sentimen positif juga mendominasi opini masyarakat, meskipun jumlahnya lebih rendah dibandingkan dengan sentimen positif pada Andika Perkasa. Keberadaan sentimen netral dan negatif pada data ini menunjukkan bahwa persepsi publik terhadap Ahmad Luthfi bersifat beragam, dengan adanya dukungan, sikap netral, serta kritik dari masyarakat.

Untuk memperjelas distribusi sentimen tersebut, dilakukan perhitungan akhir jumlah sentimen pada masing-masing kategori yang kemudian disajikan secara rinci.

```
# 6. Hitung jumlah tiap kategori sentimen
sentiment_counts = data_new['klasifikasi_naive_bayes'].value_counts()
print(sentiment_counts)

klasifikasi_naive_bayes
Positif    314
Netral     85
Negatif    53
Name: count, dtype: int64
```

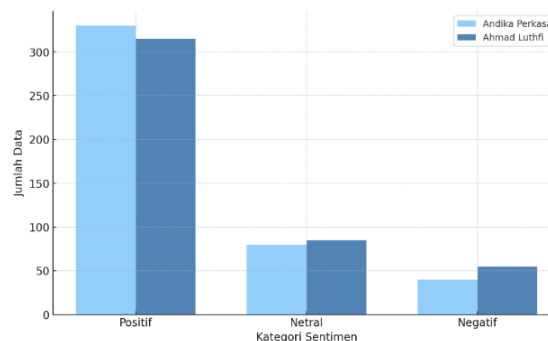
Gambar 12. Perhitungan Sentimen Ahmad Luthfi

Gambar 12 menunjukkan hasil akhir perhitungan jumlah sentimen positif, netral, dan negatif terhadap calon Gubernur Jawa Tengah Ahmad Luthfi. Data ini merupakan rekapitulasi dari proses klasifikasi yang telah dilakukan sebelumnya. Hasil perhitungan ini memberikan gambaran kuantitatif mengenai persebaran sentimen masyarakat, yang dapat digunakan sebagai dasar untuk membandingkan tingkat dukungan dan persepsi publik terhadap kedua calon gubernur secara objektif.

3.10. Visualisasi Sentimen Calon Gubernur

Setelah proses klasifikasi selesai, hasil prediksi divisualisasikan untuk memunculkan hasil berupa gambar mengenai distribusi opini publik. Visualisasi dibuat dengan bentuk diagram batang yang

menampilkan jumlah *tweet* pada masing-masing kategori sentimen, yaitu positif, negatif, dan netral. Warna yang berbeda dipakai untuk membedakan kategori, serta jumlah *tweet* ditampilkan di atas batang untuk mempermudah interpretasi. Grafik ini berfungsi untuk menunjukkan secara kuantitatif persebaran sentimen publik terhadap masing-masing tokoh.



Gambar 13. Visualisasi Distribusi Sentimen

Sebagai pelengkap, penelitian ini juga memanfaatkan Wordcloud sebagai visualisasi kata dengan kuantitas paling sering muncul dalam *tweet* [19]. *Wordcloud* dibentuk dengan menggabungkan seluruh teks, kemudian menghapus kata-kata umum yang tidak bermakna (*stopwords*). Hasilnya, kata-kata dominan seperti “dukong”, “pilih”, “tolak”, atau yang lainnya mencerminkan isu atau emosi yang sering dibahas publik.



Gambar 14. Wordcloud Andika Perkasa



Gambar 15. Wordcloud Ahmad Luthfi

Visualisasi ini penting karena dapat membantu masyarakat maupun calon gubernur memahami narasi dominan yang berkembang, serta memberikan wawasan praktis sesuai dengan manfaat penelitian.

4. Kesimpulan

Dalam penelitian mengenai analisis sentimen terhadap Calon Gubernur Jawa Tengah periode 2024–2029 menggunakan metode *Naïve Bayes*, diperoleh beberapa temuan penting. Berdasarkan hasil analisis sentimen, kedua calon gubernur, yaitu Andika Perkasa dan Ahmad Luthfi, mendapatkan persepsi yang cukup positif dari publik di media sosial *X* (dulu *Twitter*). Dari total 452 data, Andika Perkasa memperoleh 350 sentimen positif, 84 netral, dan 18 negatif, sedangkan Ahmad Luthfi memperoleh 336 sentimen positif, 98 netral, dan 18 negatif.

Evaluasi performa model klasifikasi sentimen yang telah dilakukan dengan algoritma *Naïve Bayes* menunjukkan hasil cukup baik. Pada rasio data latih dan uji 90:10, model menghasilkan akurasi tertinggi sebesar 90,28% untuk data Andika Perkasa dan 84,51% untuk data Ahmad Luthfi. Akurasi model cenderung menurun seiring dengan bertambahnya proporsi data uji pada rasio 80:20, 70:30, 60:40, maupun 50:50, meskipun secara keseluruhan model tetap menunjukkan performa yang baik dengan mempertahankan nilai *precision*, *recall*, dan *f1-score* pada tiap kelas sentimen.

Selain itu, melalui formula dukungan, diperoleh skor masing-masing calon sebesar 51,07% untuk Andika Perkasa dan 48,93% untuk Ahmad Luthfi. Namun demikian, hasil resmi Pilkada Jawa Tengah 2024 menunjukkan bahwa Ahmad Luthfi unggul dengan perolehan 59,14% suara, sedangkan Andika Perkasa hanya memperoleh 40,86% suara.

Hasil penelitian ini membuktikan bahwa media sosial *X*, apabila dikombinasikan dengan algoritma *Naïve Bayes Classifier*, dapat menjadi sarana yang efektif untuk menangkap dan mengolah opini masyarakat secara *real-time*. Meski begitu, hasil analisis sentimen tidak selalu sepenuhnya merepresentasikan preferensi masyarakat secara keseluruhan, sehingga tetap perlu dipadukan dengan sumber data lain untuk menghasilkan gambaran yang lebih komprehensif.

Daftar Rujukan

- [1] A. Rizal, "Pemilukada Serentak 2024: Dinamika, Urgensi dan Tantangan," *Proc. Ist Conf. Soc. Polit. Cult. 2022*, vol. 1, pp. 136–140, 2022.
- [2] D. Permatasari, K. A. Nasution, M. R. Khairi, and F. F. Tambunan, "Opini Publik dalam Sistem Demokrasi," vol. 7482, pp. 41–46, 2025.
- [3] M. Zaskya, A. Boham, and L. J. H. Lotulung, "Twitter Sebagai Media Mengungkapkan Diri Pada Kalangan Milenial Twitter as a Media for Self-Disclosure Among Millennials," pp. 1–8, 2020.
- [4] D. D. Ayani, H. S. Pratiwi, and H. Muhandi, "Implementasi Web Scraping untuk Pengambilan Data pada Situs

- Marketplace," vol. 7, no. 4, pp. 257–262, 2019.
- [5] F. Liantoni and H. Nugroho, "Klasifikasi Daun Herbal Menggunakan Metode *Naïve Bayes Classifier* dan *K- Nearest Neighbor*," vol. 5, no. 1, pp. 9–16, 2015.
- [6] Y. Sunardhi, "Propaganda di Media Sosial," pp. 1–11, 2025, [Online]. Available: <https://www.yosephsunardhi.com/2025/01/propaganda-di-media-sosial.html>
- [7] S. A. F. Syifa Khairunnisa, Adiwijaya, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi)," vol. 5, no. April, pp. 406–414, 2021, doi: 10.30865/mib.v5i2.2835.
- [8] R. Vindua and A. U. Zailani, "Analisis Sentimen Pemilu Indonesia Tahun 2024 Dari Media Sosial Twitter Menggunakan Python," *JURIKOM (Jurnal Ris. Komputer)*, vol. 10, no. 2, p. 479, 2023, doi: 10.30865/jurikom.v10i2.5945.
- [9] T. Ridwansyah, "Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan *Naïve Bayes Classifier*," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 5, pp. 178–185, 2022, doi: 10.30865/klik.v2i5.362.
- [10] N. Saputra, T. B. Adji, and A. E. Permanasari, "Analisis Sentimen Data Presiden Jokowi Dengan Preprocessing Normalisasi Dan Stemming Menggunakan Metode *Naïve Bayes* Dan *Svm*," *J. Din. Inform.*, vol. 5, no. November, pp. 1–13, 2015.
- [11] A. F. Hidayatullah, "PENGARUH STOPWORD TERHADAP PERFORMA KLASIFIKASI TWEET BERBAHASA INDONESIA," vol. 1, no. 1, pp. 1–4, 2016.
- [12] S. Ramadhani, D. Azzahra, U. I. Negeri, and S. S. Kasim, "Comparison of K-Means and K-Medoids Algorithms in Text Mining based on Davies Bouldin Index Testing for Classification of Student 's Thesis," vol. x, no. x, pp. 24–33, 2022.
- [13] M. U. Albab, Y. K. P, and M. N. Fawaiq, "Optimization of the Stemming Technique on Text preprocessing President 3 Periods Topic," vol. 20, no. 2, pp. 1–10, 2023.
- [14] M. Faruqziddan, E. H. S. Aulia, S. D. Azzahra, and P. B. Utomo, "Perbandingan Skenario Balancing Oversampling dan Undersampling dalam Klasifikasi Resiko Kambuh Kanker Tiroid menggunakan Algoritma SVM Linear," *JAMI J. Ahli Muda Indones.*, vol. 5, no. 2, pp. 172–182, 2024, doi: 10.46510/jami.v5i2.320.
- [15] M. A. Hasan, "Ensemble Language Models for Multilingual Sentiment Analysis," pp. 1–9, 2024, [Online]. Available: <http://arxiv.org/abs/2403.06060>
- [16] A. Lavanya, W. Ali, D. J. Lloret, S. D. Vidya Sagar, and C. Bharadwaj, "A Real-time Visualization of Global Sentiment Analysis on Declaration of Pandemic," *International Journal of Computer Engineering in Research Trends*, vol. 9, no. 6. pp. 104–113, 2022. doi: 10.22362/ijcert/2022/v9/i06/v9i0602.
- [17] S. Puad, G. Garno, and A. Susilo Yuda Irawan, "Analisis Sentimen Masyarakat Pada Twitter Terhadap Pemilihan Umum 2024 Menggunakan Algoritma *Naïve Bayes*," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 3, pp. 1560–1566, 2023, doi: 10.36040/jati.v7i3.6920.
- [18] M. K. Khatami, "Analisis Sentimen Twitter Menggunakan *Naïve Bayes* dan Support Vector Machine Terhadap KPU Pada Pemilihan Umum Presiden 2024," 2024.
- [19] M. Galih Pradana, "Penggunaan Fitur Wordcloud Dan Document Term Matrix Dalam Text Mining," *J. Ilm. Infomatika*, vol. 08, no. 01, pp. 38–43, 2020.