

Analisis Sentimen Publik di Twitter Pasca Debat Kelima Pilpres 2024 dengan Naive Bayes

Anjana Haya Atha Zharifa ^{1,*}, Erik Iman Heri Ujianto ²

¹ Program Studi Informatika, Universitas Teknologi Yogyakarta, Indonesia

² Program Studi Magister Teknologi Informasi, Universitas Teknologi Yogyakarta, Indonesia

* Correspondence: anjanahz25@gmail.com

Copyright: © 2024 by the authors

Received: 4 November 2024 | Revised: 11 November 2024 | Accepted: 1 Desember 2024 | Published: 19 Desember 2024

Abstrak

Pemilihan presiden di Indonesia menjadi topik yang sering dibahas di media sosial, khususnya *Twitter*. Platform ini menyediakan ruang bagi masyarakat untuk mengekspresikan pandangan terkait calon presiden dan isu-isu pemilihan, sehingga cocok sebagai sumber data untuk penelitian ini. Penelitian ini bertujuan menganalisis sentimen masyarakat terhadap berita pemilihan presiden di *twitter* menggunakan metode *naïve bayes classifier*. Data diambil dari *twitter* periode 5–13 Februari 2024 dengan total 2.561 komentar. Proses penelitian mencakup pengumpulan data, *preprocessing*, pelabelan data, serta pelatihan dan pengujian model. *Naïve Bayes* dipilih karena efisien dalam klasifikasi teks dan memiliki beberapa varian untuk eksperimen model. Sentimen diklasifikasikan menjadi tiga kategori utama yaitu positif, negatif, dan netral. Hasil penelitian menunjukkan komentar negatif mendominasi (41%), diikuti positif (37,3%) dan netral (21,7%). Model *Multi Naïve Bayes Classifier* memberikan akurasi tertinggi (81%), disusul *Bernoulli Naïve Bayes* (80%) dan *Gaussian Naïve Bayes* (76%). Perbedaan akurasi ini dipengaruhi sensitivitas model terhadap karakteristik data, seperti jumlah fitur dan distribusi sentimen. Penelitian ini berpotensi membantu tim kampanye memahami isu yang memicu respons negatif dan mendukung pembuat kebijakan merancang strategi komunikasi politik lebih efektif.

Kata kunci: analisis sentimen; *naïve bayes classifier*; opini publik; pemilihan presiden; *twitter*

Abstract

The presidential election in Indonesia is a frequently discussed topic on social media, especially Twitter. This platform provides a space for people to express their views on presidential candidates and election issues, making it suitable as a data source for this study. This study aims to analyze public sentiment towards presidential election news on Twitter using the Naïve Bayes Classifier method. Data was taken from Twitter for the period 5–13 February 2024 with a total of 2,561 comments. The research process includes data collection, preprocessing, data labeling, and model training and testing. Naïve Bayes was chosen because it is efficient in text classification and has several variants for model experiments. Sentiment is classified into three main categories, namely positive, negative, and neutral. The results showed that negative comments dominated (41%), followed by positive (37.3%) and neutral (21.7%). The Multi Naïve Bayes Classifier model provided the highest accuracy (81%), followed by Bernoulli Naïve Bayes (80%) and Gaussian Naïve Bayes (76%). This difference in accuracy is influenced by the model's sensitivity to data characteristics, such as the number of features and sentiment distribution. This research has the potential to help campaign teams understand the issues that trigger negative responses and support policy makers in designing more effective political communication strategies.

Keywords: sentiment analysis; *naïve bayes classifier*; public opinion; presidential election; *twitter*



PENDAHULUAN

Pemilihan presiden menjadi perhatian publik karena intensitas diskusi yang tinggi, terutama melalui media sosial seperti Twitter yang memengaruhi dan membentuk opini publik terkait para kandidat (Noorikhshan et al., 2023). Seperti yang dilansir dalam berita online, pembahasan terkait calon presiden dan calon wakil presiden memunculkan berbagai isu-isu positif dan negatif (Makarawung et al., 2024). Dalam era digital, platform media sosial seperti Twitter menjadi sumber utama perdebatan dan berbagi informasi seputar pemilihan presiden (Laorensa et al., 2024). Twitter sebagai platform media sosial yang sangat populer, memberikan wadah bagi masyarakat untuk berbagi pandangan, dukungan, atau kritik terhadap calon presiden, partai politik, dan isu-isu pemilihan presiden (Solihin et al., 2021). Analisis sentimen di Twitter dapat membantu memahami bagaimana masyarakat merespons peristiwa-peristiwa terkait pemilihan presiden dengan lebih mendalam (Sunata et al., 2024).

Analisis sentimen memiliki peran penting dalam menyediakan wawasan bagi pembuat kebijakan atau tim kampanye untuk menilai respons dan opini masyarakat terhadap isu pemilihan (Rahayu et al., 2024). Metode *naïve bayes classifier* dipilih karena kemampuannya yang tinggi dalam klasifikasi teks, efisiensi dalam pengolahan dataset besar (Ardiansyah et al., 2021; Sihombing et al., 2021), dan kemudahan implementasinya yang cocok untuk analisis sentimen pada teks singkat di media sosial (Salsabila et al., 2022). Metode *naïve bayes classifier* yang menggunakan analisis prinsip-prinsip probabilitas telah terbukti efektif dalam mengklasifikasikan teks ke dalam kategori sentimen yang berbeda, seperti positif, negatif, atau netral (Sarimole & Kudrat, 2024).

Platform Twitter sebagai media sosial populer, menjadi ruang utama bagi masyarakat untuk mendiskusikan dan menyebarkan opini politik, sehingga relevan dalam memahami persepsi publik terhadap isu pemilihan presiden. Fitur-fitur seperti *retweet*, *polling*, dan tagar memungkinkan diskusi berlangsung secara *real-time*, tetapi juga menghadirkan tantangan seperti polarisasi opini, penyebaran misinformasi, dan peningkatan *echo chamber* (Setiyawati & Cahyono, 2023). Polarisasi disebabkan oleh algoritma platform yang memprioritaskan konten dengan keterlibatan tinggi, sering kali berupa opini ekstrem atau kontroversial. Selain itu, sifat anonimitas Twitter memungkinkan terjadinya ujaran kebencian yang memperburuk kualitas diskusi (Verawati & Audit, 2022). Untuk memahami dinamika opini ini, analisis sentimen menggunakan metode seperti *Naïve Bayes* dapat menjadi langkah awal dalam memetakan persepsi publik terhadap isu-isu politik yang sedang berlangsung.

Naïve bayes adalah metode pengklasifikasi yang berbasis pada probabilitas sederhana dan dirancang untuk bekerja dengan asumsi bahwa variabel-variabel saling bebas (*independen*) (Retnosari, 2021). Algoritma ini menekankan pembelajaran pada probabilitas (Singgalen, 2022). Salah satu keuntungan utama dari algoritma *naïve bayes* adalah bahwa tingkat kesalahan yang ditemukan cenderung lebih rendah ketika digunakan pada dataset yang besar (Perdana et al., 2022). Selain itu, akurasi *naïve bayes* dan kecepatannya juga cenderung lebih tinggi ketika diterapkan pada dataset yang memiliki volume data yang lebih besar (Afriansyah et al., 2024). Metode *naïve bayes* juga sangat efisien dalam hal waktu komputasi dan penggunaan memori, sehingga sangat cocok untuk aplikasi real time (Fudholi et al., 2024). Selain itu, algoritma ini mudah diimplementasikan dan diinterpretasikan, sehingga banyak digunakan dalam berbagai bidang, termasuk pemrosesan bahasa alami, bioinformatika, dan sistem rekomendasi (Hakim & Sugiyono, 2024). Sederhananya, meskipun *naïve bayes* berdasarkan asumsi yang kuat dan terkadang tidak realistis tentang independensi fitur, ia tetap menjadi salah satu metode klasifikasi yang paling efektif dan efisien yang tersedia (Haq et al., 2024).

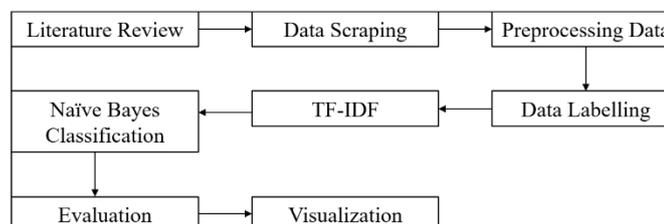
Berdasarkan hasil penelitian yang dilakukan oleh Febriyani & Februariyanti (2023), analisis *sentimen* terhadap Program Kampus Merdeka di Twitter menggunakan algoritma *naïve bayes classifier* menunjukkan 272 opini positif dan 229 opini negatif dengan akurasi rata-rata 60%, presisi 64%, *recall* 58%, dan *f1-score* 58%. Penelitian mendemonstrasikan kemampuan

naïve bayes dalam mengklasifikasikan sentimen publik dengan akurasi yang dapat diterima. Penelitian lain oleh mengevaluasi sentimen masyarakat terhadap kinerja Dewan Perwakilan Rakyat (DPR) di platform Twitter, dengan hasil 95 tweet bersentimen positif, 693 netral, dan 758 negatif. Studi ini menekankan efektivitas Naïve Bayes dalam mengidentifikasi sentimen publik terkait kinerja lembaga pemerintah, relevan untuk konteks pemantauan opini publik di media sosial.

Penelitian ini dilakukan untuk mengatasi kekurangan penelitian sebelumnya yang memiliki sejumlah keterbatasan. Misalnya, penelitian oleh Febriyani & Februariyanti (2023) menggunakan data dari Twitter untuk menganalisis sentimen terhadap Program Kampus Merdeka, tetapi data yang diambil tidak relevan dengan momentum penting, sehingga hasilnya kurang mencerminkan dinamika opini masyarakat. Selain itu, penelitian Putri et al. (2022) terhadap kinerja DPR juga menunjukkan sampel kecil dan cakupan terbatas, yang kurang representatif dalam menggambarkan opini publik secara luas. Metode analisis yang digunakan pada penelitian terdahulu, seperti pada penelitian Sarimole & Kudrat (2024), juga terkadang kurang optimal dalam menangani kompleksitas teks media sosial, seperti slang, hashtag, atau emoji, yang dapat memengaruhi akurasi klasifikasi sentimen. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap berita pemilihan presiden di Twitter dengan metode *naïve bayes classifier*, menggunakan data dari periode yang relevan dan sampel yang lebih luas. Temuan penelitian ini diharapkan memberikan wawasan bagi tim kampanye dan pembuat kebijakan dalam memahami opini publik serta menyusun strategi komunikasi politik yang lebih efektif.

METODE

Pada gambar 1 menunjukkan bahwa penelitian ini menggunakan metode *naïve bayes* dengan delapan tahapan utama, yaitu *literature review*, *data scraping*, *data labelling*, *data preprocessing*, transformasi data menggunakan TF-IDF, klasifikasi sentimen, evaluasi model, dan visualisasi hasil, seperti yang terlihat Pada Gambar 1 dibawah ini. Tahap pertama, *literature review*, bertujuan memahami teori dasar dan metode penelitian terdahulu, membantu menentukan pendekatan terbaik. Selanjutnya, data dikumpulkan melalui *data scraping* menggunakan Twitter API dengan parameter seperti kata kunci "pemilihan presiden 2024" dan tagar populer. Data yang dikumpulkan berjumlah 2.651 *tweet*, meliputi variabel seperti sentimen (positif, negatif, atau netral), waktu posting, jumlah *likes*, dan *retweets*. Proses ini memastikan ketersediaan data yang relevan untuk analisis sentimen.



Gambar 1. Tahapan penelitian

Pada tahap *data labelling*, *tweet* diberi kategori sentimen menggunakan pendekatan semi-otomatis. *Tweet* dengan kata kunci tertentu dilabeli secara otomatis, sedangkan lainnya diperiksa manual untuk meningkatkan akurasi. Label sentimen terdiri atas tiga kategori, yaitu positif, negatif, dan netral. Tahap ini penting untuk memastikan kualitas dataset yang digunakan dalam pelatihan model. Proses *data preprocessing* dilakukan untuk membersihkan teks dengan menghapus elemen tidak relevan, seperti URL, simbol, dan angka. Langkah ini mencakup tokenisasi, normalisasi, penghapusan *stopwords*, dan *stemming*, menggunakan pustaka Python seperti NLTK dan spaCy.

Setelah *preprocessing*, data ditransformasikan menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF). Teknik ini menghitung bobot kata berdasarkan frekuensinya dalam dokumen tertentu dan keseluruhan dataset. TF-IDF dengan pengaturan *ngram* (1,2) digunakan untuk menangkap konteks kata yang lebih kaya, sehingga meningkatkan akurasi analisis sentimen. Pada tahap klasifikasi, algoritma *naïve bayes* (Multinomial, Bernoulli, dan Gaussian) diterapkan karena efisiensinya dalam menangani data teks besar. Model diuji untuk menentukan varian algoritma terbaik berdasarkan karakteristik dataset.

Evaluasi dilakukan menggunakan matriks kebingungan (*confusion matrix*), mencakup metrik seperti *precision*, *recall*, dan *f1-score*. Analisis hasil model memberikan wawasan tentang performa dalam mengklasifikasikan sentimen. Tahap akhir adalah visualisasi hasil menggunakan pustaka Matplotlib dan Seaborn. Diagram pie digunakan untuk menunjukkan distribusi sentimen, sementara *heatmap* memvisualisasikan performa klasifikasi. Tahapan sistematis ini memastikan penelitian dapat memberikan analisis sentimen yang akurat, relevan, dan bermanfaat dalam memahami opini publik terkait pemilihan presiden 2024 di *Twitter*.

HASIL DAN PEMBAHASAN

Hasil

Penelitian ini mengkaji distribusi sentimen pada tweet terkait debat capres 2024, dengan temuan bahwa sentimen negatif mendominasi diskusi publik. Akurasi model Naïve Bayes yang digunakan untuk analisis sentimen bervariasi, di mana beberapa model menunjukkan performa yang baik, sementara yang lain mengalami kesulitan dalam mengklasifikasikan kategori tertentu. Sebanyak 2561 komentar *twitter* dikumpulkan pada pemilihan presiden pasca debat ke-5 menggunakan metode analisis *naïve bayes classifier*. Data diperoleh dengan bantuan alat bernama *tweet-harvest*, yang dirancang untuk mempermudah akses dan pengunduhan tweet dalam jumlah besar serta memfilter data secara spesifik sesuai tema penelitian. *Tweet-harvest* juga menggunakan token autentikasi untuk memastikan pengumpulan data yang sah, sehingga data yang diperoleh lebih relevan dan terfokus. Analisis ini memberikan wawasan penting mengenai opini publik pasca debat serta efektivitas model dalam mengidentifikasi sentimen secara akurat.

Tabel 1. Sampel data ulasan

Tanggal	Tweet
2/12/2024	Yah menang di arab saudi. Karena emang Anis keturunan orang sana. Cuma arab saudi saja yg menang.
2/12/2024	Sedih campur kesel bgt temen2 gua mayoritas pilih 02 dan yang bikin kesel mereka tuh bener2 ganonton debat gatau visi misi capresnya cuma liat potongan video ditiktok dan bilang anis gabisa kerja lah pendukung 01 dibayar lah Azzzzajwkyahazkzkzk

Tabel 2. Hasil *case folding* sampai *Stemming*

Ulasan Awal	@maspandam @jokowi Lah sama kaya anis yang bilang saya tidak ingin menjadi penghianat untuk bersaing dengan pak prabowo itu adalah hal yang tidak mungkin saya lakukan
<i>Case folding</i>	kaya anis bilang penghianat bersaing prabowo lakukan
<i>Tokenize</i>	['kaya', 'anis', 'bilang', 'penghianat', 'bersaing', 'prabowo', 'lakukan']
<i>Remove Stopwords</i>	kaya anis bilang penghianat bersaing prabowo lakukan
<i>Stemming</i>	kaya anis bilang penghianat saing prabowo laku

Tabel 1 adalah hasil ulasan dari 2.651 *tweet*. Hasil tersebut didapatkan dari kode yang dijalankan untuk mengambil data dari Twitter dengan mengatur hasil pencarian data yaitu debat capres 2024 dari tanggal 11-13-2023 sampai 02-13-2024, serta tidak lupa untuk mengatur limitasi pengambilan data kurang lebih 500 data. Setelah berhasil mendapat seluruh data dan terpenuhi segala parameter yang sudah dimasukkan di dalam *tweet-harvest*, hasil akan disimpan dalam bentuk file *Comma-Separated Values* (CSV) dengan nama file yaitu *pilpres.csv*.

Hasil pada tahapan *preprocessing* data adalah kompleksitas teks dan meningkatkan akurasi model dalam mendeteksi sentimen seperti pada tabel 2. Pada tahap pembersihan data, elemen-elemen seperti URL, mention, hashtag, simbol, dan angka dihapus untuk mempertahankan kata-kata yang bermakna. Selanjutnya, dilakukan penghapusan stopwords menggunakan library NLTK dan daftar *stopwords* bahasa Indonesia, memastikan hanya kata-kata kunci relevan yang dipertahankan. Proses *tokenizing* memecah teks menjadi token individu untuk analisis lanjutan dalam pembangunan model klasifikasi sentimen menggunakan *naïve bayes*. Tahap akhir adalah *stemming*, mengubah kata-kata menjadi bentuk dasarnya untuk mengurangi variasi kata yang memiliki akar sama, menghasilkan teks yang lebih sederhana dan siap untuk analisis sentimen pada komentar pemilihan presiden 2024 di Twitter.

Hasil pengambilan data menggunakan *tweet-harvest* dilakukan kurang lebih selama 7 hari menghasilkan total 2.651 *tweet* terkumpul. Berdasarkan data yang sudah terkumpul tersebut akan data akan dilakukan pemisahan data untuk proses pembelajaran dan pengujian dengan pembagian yaitu 2120 data *tweet* sebagai data latih dan 531 sebagai data uji seperti pada tabel 3, dari hasil pemisahan tersebut akan dilakukan hasil uji dan analisa menggunakan metode *naïve bayes*. Pemisahan data dengan rasio 80:20 dilakukan untuk membantu menjaga keseimbangan antara pelatihan model yang efektif dan evaluasi yang objektif.

Tabel 3. Jumlah Dataset

Data	Jumlah
Pelatihan	2.120
Pengujian	531

Kemudian dilakukan pelabelan data menggunakan *lexicon*, *lexicon* digunakan untuk mempercepat proses pelabelan awal dan memastikan data diklasifikasikan sesuai dengan kata-kata kunci sentimen. Tujuan dari proses ini adalah untuk mengkategorikan teks *tweet* ke dalam kategori sentimen positif, negatif, atau netral berdasarkan keberadaan kata-kata sentimen dalam teks tersebut. Hasil dari proses ini adalah penentuan sentimen untuk setiap teks dalam dataset. Terlihat bahwa setelah hasil pelabelan menunjukkan 575 data bersifat netral, 990 data bersifat positif, dan 1086 bersifat negatif. Hasil tersebut akan dijadikan sebagai dataset untuk melakukan proses pelatihan dan pengujian data.

Tabel 4. Jumlah sentimen

Data	Jumlah
Netral	575
Positif	990
Negatif	1086

Tabel 4 menunjukkan jumlah data sentimen, dengan 575 data bersentimen netral, 990 data bersentimen positif, dan 1086 data bersentimen negatif. Sentimen negatif memiliki jumlah tertinggi, diikuti oleh positif, sementara sentimen netral memiliki jumlah terendah. Tahapan selanjutnya yaitu Pelatihan dan Pengujian Data, pada tahap ini terdapat tahap pemisahan data (*data splitting*) dan Proses pelatihan (*training*). Pada tahapan pemisahan data, dilakukan untuk memastikan bahwa model pembelajaran mesin yang akan dibangun dapat diuji secara adil pada

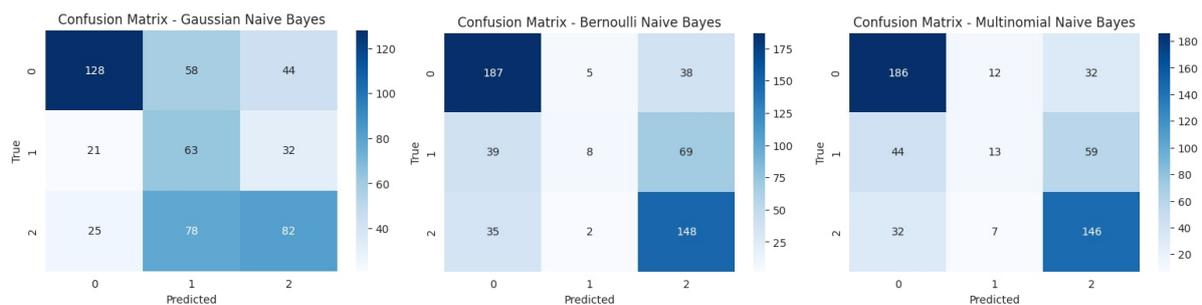
data yang belum pernah diuji sebelumnya. Hasil ini penting untuk mempersiapkan langkah-langkah selanjutnya dalam analisis sentimen menggunakan metode *naïve bayes* pada komentar pemilihan presiden 2024 di *Twitter*.

Tabel 5. Hasil akurasi dari pengujian

<i>Tweet</i>	Tanggal	Waktu	Sentimen	Akurasi
berani nggak polling capres liat sih presiden pilih rakyat tantang nih biar rakyat presiden pilih bpk kuat sih	02-07-2024	10.00	Negatif	98,57%
prabowo sorot satu capres temu walhi ada sembunyi	02-07-2024	10.00	Positif	95,57%
momentum pilpres pikir omong beliau paslon sempurna ada lesser evil and greater evil sana tinggal sikap pilih lantas faktor pilih paslon	02-07-2024	10.50	Negative	80,67%
...		
semarang spirit utama juang ganjar mahfud pilpres ganjar pranowo baik rakyat	02-10-2024	13.00	Netral	90,34%

Proses selanjutnya adalah proses pengujian yang bertujuan untuk memahami akurasi yang terjadi ketika proses pelatihan terjadi, sehingga diharapkan dari hasil pengujian dapat di dapatkan nilai akurasi yang baik untuk proses penelitian analisis sentimen yang dilakukan. Setelah melakukan proses pendefinisian lapisan latihan, data akan memasukan proses pelatihan dengan menggunakan *naïve bayes*. Hasil dari proses ini yaitu diharapkan mesin dapat belajar dalam melakukan klasifikasi data, sehingga dapat melakukan analisa sentimen.

Tabel 5 menunjukkan hasil pengujian analisis sentimen pada sejumlah *tweet* menggunakan metode *naïve bayes*, dengan klasifikasi sentimen negatif, positif, atau netral. Secara keseluruhan, model ini memiliki rata-rata akurasi sebesar 88,27%, dengan performa terbaik pada sentimen negatif yang mencapai akurasi hingga 98,57%, diikuti sentimen positif dengan akurasi hingga 95,57%. Meskipun demikian, model menghadapi tantangan dalam beberapa kasus dengan akurasi lebih rendah, seperti 80,67% pada *tweet* berisi sentimen negatif. Hal ini menunjukkan bahwa model mengalami kesulitan pada teks dengan struktur yang dapat mempengaruhi ketepatan klasifikasi sentimen.

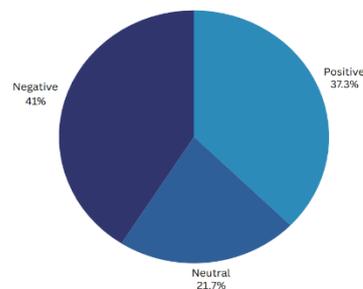


Gambar 2. Hasil evaluasi *naïve bayes*

Pada gambar 2 divisualisasikan hasil evaluasi menggunakan *naïve bayes*. Model *Gaussian Naive Bayes* memiliki akurasi tinggi dalam memprediksi kelas 0 dengan *precision*

0,79 dan *recall* 0,84, namun kesulitan membedakan kelas 1 dan 2, terbukti dari *precision* dan *recall* rendah pada kedua kelas tersebut. Hal ini menunjukkan bahwa meskipun cukup baik dalam mengidentifikasi sentimen tertentu, model ini sering salah mengklasifikasikan kelas 1 dan 2, terutama saat distribusi data tidak normal. Model *Bernoulli Naïve Bayes* menunjukkan performa lebih baik dalam memprediksi kelas 0 dan 2 dengan *precision* dan *recall* tinggi, namun memiliki *precision* rendah pada kelas 1. *Multinomial Naïve Bayes* memprediksi kelas 0 dengan benar sebanyak 186 kali dengan *precision* 0,73 dan *recall* 0,80, namun kesalahan signifikan terjadi pada kelas 1. F1-score menunjukkan Gaussian unggul dalam kelas 0, sedangkan Bernoulli dan Multinomial lebih andal dalam klasifikasi keseluruhan tetapi lemah pada kelas tertentu, disebabkan oleh karakteristik masing-masing model.

Gambar 3 menunjukkan distribusi sentimen dari dataset dalam bentuk diagram pie, dengan kategori negatif, positif, dan netral. Sentimen negatif mendominasi sebesar 41%, yang mungkin dipengaruhi oleh polarisasi opini publik terkait topik debat atau sikap terhadap calon presiden, mencerminkan ketidakpuasan atau kritik yang kuat dalam diskusi ini. Di sisi lain, sentimen positif mencakup 37,3%, menunjukkan bahwa terdapat dukungan atau pandangan optimis yang cukup besar, meskipun masih lebih rendah dibandingkan sentimen negatif. Sentimen netral berada di angka 21,7%, mengindikasikan bahwa sebagian kecil dari diskusi tetap objektif atau tidak condong ke salah satu arah tertentu. Implikasi dari distribusi ini menunjukkan bahwa diskusi publik di media sosial cenderung kritis, yang mungkin berpengaruh pada persepsi umum terkait calon atau kebijakan pemilihan. Hasil ini penting untuk memahami opini publik yang terpolarisasi dan menjadi landasan dalam mengevaluasi strategi komunikasi politik bagi para calon.



Gambar 3. Hasil distribusi sentimen

Pembahasan

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terkait pemilihan presiden melalui komentar di Twitter setelah debat kelima, menggunakan metode *naïve bayes classifier*. Hasil analisis menunjukkan dominasi sentimen negatif sebesar 41%, yang mencerminkan kecenderungan pengguna *twitter* untuk bersikap kritis atau menunjukkan ketidakpuasan, kemungkinan sebagai refleksi dari frustrasi publik terhadap isu politik yang muncul dalam debat. Fenomena polarisasi dalam komentar ini mengindikasikan bahwa isu politik sering kali memunculkan perbedaan pandangan tajam, di mana *Twitter* yang mendorong kebebasan berekspresi berperan dalam memperkuat opini ekstrem dan subjektif. Karakteristik ini mungkin berkontribusi pada tingginya bias sentimen negatif, karena pengguna merasa lebih bebas menyuarakan ketidakpuasan.

Sentimen positif mencapai 37,3%, yang menandakan adanya segmen masyarakat yang masih optimis atau memberikan dukungan kepada calon presiden. Hal ini mungkin muncul dari pendukung calon tertentu yang merasa puas dengan performa kandidatnya, terutama setelah debat kelima. Sementara itu, sentimen netral hanya sebesar 21,7%, yang menunjukkan minimnya pandangan objektif di tengah diskusi terkait pemilihan presiden. Sifat pemilu yang cenderung memicu opini kuat bisa menjadi alasan rendahnya sentimen netral. Penelitian lain

menunjukkan bahwa peristiwa politik intens, seperti debat, sering kali mengurangi netralitas publik karena masyarakat cenderung memilih posisi tegas.

Implementasi metode *naïve bayes classifier* berhasil menunjukkan akurasi yang cukup tinggi, yaitu 81%, dalam mengklasifikasikan sentimen komentar. Angka ini menunjukkan keefektifan *naïve bayes* dalam menangani dataset ini, namun juga mencerminkan adanya tantangan dalam mengatasi teks yang ambigu atau memiliki konteks tersirat. Model Multinomial Naïve Bayes terbukti lebih cocok untuk data ini karena mampu menangani frekuensi kata dengan lebih efektif pada teks media sosial, yang cenderung menggunakan variasi kata tinggi. Meskipun perbedaan akurasi antara model Multinomial, Bernoulli, dan *Gaussian Naïve Bayes* tidak signifikan, model Multinomial unggul dalam penelitian ini. Hal ini kemungkinan besar disebabkan oleh karakteristik data *multiclass* yang membutuhkan pendekatan kompleks untuk menangani variasi komentar.

Beberapa faktor sosial atau politik mungkin turut memengaruhi dominasi sentimen negatif ini, seperti ketidakpuasan terhadap isu-isu spesifik yang muncul selama debat. Di media sosial, fenomena polarisasi sering kali diperparah oleh kemampuan *twitter* untuk menyebarkan opini secara cepat dan luas, yang pada akhirnya mendorong publik untuk menunjukkan sikap yang kuat. Sentimen netral yang rendah juga mengindikasikan bahwa diskusi politik di media sosial cenderung menarik opini yang tegas, baik pro maupun kontra.

Beberapa faktor sosial atau politik mungkin turut memengaruhi dominasi sentimen negatif ini, seperti ketidakpuasan terhadap isu-isu spesifik yang muncul selama debat. Penelitian sebelumnya oleh Febriyani & Februariyanti (2023) menunjukkan bahwa sentimen publik pada Program Kampus Merdeka di *Twitter* didominasi oleh opini positif (54%), berbeda dengan hasil penelitian ini yang menunjukkan dominasi sentimen negatif (41%). Perbedaan ini dapat disebabkan oleh sifat isu politik yang lebih cenderung memicu perdebatan dan polarisasi, dibandingkan isu pendidikan yang bersifat lebih umum dan netral. Selain itu, hasil penelitian Putri et al. (2022) yang menganalisis sentimen terhadap kinerja DPR menunjukkan dominasi sentimen negatif (48%), mendekati hasil penelitian ini. Namun, penelitian mereka memiliki skala data yang lebih kecil dan cakupan yang terbatas pada komentar tertentu, sedangkan penelitian kami mencakup periode yang lebih spesifik (pasca debat kelima) dengan sampel yang lebih luas. Hal ini menunjukkan bahwa penelitian ini lebih representatif dalam menggambarkan dinamika opini publik pada momen politik penting seperti debat pemilihan presiden..

Akurasi *naïve bayes* dalam penelitian ini mencapai 81%, yang lebih tinggi dibandingkan penelitian sebelumnya oleh Febriyani & Februariyanti (2023) dengan akurasi rata-rata 60%. Peningkatan ini dapat dikaitkan dengan penggunaan metode *preprocessing* yang lebih lengkap, seperti tokenisasi, normalisasi, dan stemming, serta penerapan TF-IDF dengan *ngram range* (1,2) untuk menangkap konteks kata yang lebih kompleks. Selain itu, hasil ini juga lebih unggul dibandingkan penelitian Putri et al. (2022), yang menggunakan *naïve bayes* dengan data sentimen kinerja DPR, dengan akurasi 68%. Dengan akurasi yang lebih tinggi, penelitian ini membuktikan keunggulan model *Multinomial Naïve Bayes* dalam menangani data teks media sosial yang kompleks dan dinamis. Penelitian ini tidak hanya menegaskan efektivitas *naïve bayes* dalam analisis sentimen, tetapi juga menunjukkan bahwa dalam pemanfaatan data dari periode tertentu yang relevan secara politik, memberikan hasil yang lebih mendalam dan kontekstual.

SIMPULAN

Analisis sentimen terhadap komentar debat presiden 2024 di platform *Twitter* menggunakan metode *naïve bayes*, mengidentifikasi dominasi sentimen negatif sebesar 41%, diikuti oleh sentimen positif 37,3% dan netral 21,7%. Dengan metode *multi naïve bayes classifier* mencapai akurasi tertinggi 81%, model ini terbukti lebih unggul karena dapat

menangani data yang beragam sehingga sesuai untuk analisis sentimen media sosial. Temuan ini dapat membantu calon presiden dan tim kampanye dalam menyusun strategi komunikasi di media sosial, karena tim dapat memahami persepsi publik dengan lebih baik dan mengidentifikasi area yang memerlukan perbaikan. Secara akademis, penelitian ini turut memberikan kontribusi pada pemahaman tentang opini publik di media sosial, khususnya dalam konteks peristiwa politik penting seperti debat presiden, serta menyoroti peran media sosial dalam membentuk dinamika komunikasi politik di era digital.

REFERENSI

- Afriansyah, M., Saputra, J., Ardhana, V. Y. P., & Sa'adati, Y. (2024). Algoritma Naive Bayes Yang Efisien Untuk Klasifikasi Buah Pisang Raja Berdasarkan Fitur Warna. *Journal of Information Systems Management and Digital Business (JISMDB)*, 1(2), 236–248. <https://doi.org/10.59407/jismdb.v1i2.438>
- Ardiansyah, M., Sunyoto, A., & Luthfi, E. T. (2021). Analisis Perbandingan Akurasi Algoritma Naive Bayes dan C4.5 untuk Klasifikasi Diabetes. *Edumatic: Jurnal Pendidikan Informatika*, 5(2), 147-156. <https://doi.org/10.29408/edumatic.v5i2.3424>
- Febriyani, E., & Februariyanti, H. (2023). Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Algoritma Naive Bayes Classifier Di Twitter. *Jurnal Tekno Kompak*, 17(1), 25–38. <https://doi.org/10.33365/jtk.v17i1.2061>
- Fudholi, L. A., Rahaningsih, N., & Dana, R. D. (2024). Sentimen Analisis Perilaku Penggemar Coldplay Di Media Sosial Twitter Menggunakan Metode Naive Bayes. *Jurnal Mahasiswa Teknik Informatika*, 8(3), 4150–4159. <https://doi.org/10.36040/jati.v8i3.9827>
- Hakim, Z. R., & Sugiyono, S. (2024). Analisa Sentimen Terhadap Kereta Cepat Jakarta – Bandung Menggunakan Algoritma Naive Bayes Dan K-Nearest Neighbor. *Jurnal Sains Dan Teknologi*, 5(3), 939–945. <https://doi.org/10.55338/saintek.v5i3.1423>
- Haq, M. Z., Octiva, C. S., Ayuliana, A., Nuryanto, U. W., & Suryadi, D. (2024). Algoritma Naive Bayes untuk Mengidentifikasi Hoaks di Media Sosial. *Jurnal Minfo Polgan*, 13(1), 1079–1084. <https://doi.org/10.33395/jmp.v13i1.13937>
- Laorensa, E., Suri, E. W., & Dani, R. (2024). Peran Media Sosial Dalam Membentuk Persepsi Pemilih Pada Pemilu 2024 (Studi Di Kabupaten Bengkulu Tengah). *Jurnal Ilmiah Administrasi Publik*, 2(1), 1–16.
- Makarawung, Y. A., Wulandari, Y. F., & Himawan, S. (2024). Analisis Konten TikTok dalam Komunikasi Politik Capres- Cawapres di Pemilu 2024 untuk Generasi Z. 3(4), 320–336.
- Noorikhsan, F. F., Ramdhani, H., Sirait, B. C., & Khoerunisa, N. (2023). Dinamika Internet, Media Sosial, dan Politik di Era Kontemporer: Tinjauan Relasi Negara-Masyarakat. *Journal of Political Issues*, 5(1), 95–109. <https://doi.org/10.33019/jpi.v5i1.131>
- Perdana, A., Hermawan, A., & Avianto, D. (2022). Analisis Sentimen Terhadap Isu Penundaan Pemilu di Twitter Menggunakan Naive Bayes Classifier. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 11(2), 195–200. <https://doi.org/10.32736/sisfokom.v11i2.1412>
- Putri, D. D., Nama, G. F., & Sulistiono, W. E. (2022). Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier. *Jurnal Informatika Dan Teknik Elektro Terapan*, 10(1), 34–40. <https://doi.org/10.23960/jitet.v10i1.2262>
- Rahayu, D. D., Fatchan, M., & Ligouri, A. (2024). Analisis Sentimen Twitter Terpilihnya Prabowo - Gibran Menggunakan Metode Neural Network. *Jurnal Tematik*, 11(1), 85–91. <https://doi.org/10.38204/tematik.v11i1.1943>
- Retnosari, R. (2021). Analisis Kelayakan Kredit Usaha Mikro Berjalan Pada Perbankan Dengan Metode Naive Bayes. *Jurnal PROSISKO*, 8(1), 53–59.

- Salsabila, S. M., Murtopo, A. A., & Fadhilah, N. (2022). Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naive Bayes Classifier. *Jurnal Manajemen Informatika Politeknik Ganesha*, 11(2), 30–35. <https://doi.org/10.33395/jmp.v11i2.11640>
- Sarimole, F. M., & Kudrat. (2024). Analisis Sentimen Terhadap Aplikasi Satu Sehat Pada Twitter Menggunakan Algoritma Naive Bayes Dan Support Vector Machine. *Jurnal Sains Dan Teknologi*, 5(3), 783–790. <https://doi.org/10.55338/saintek.v5i1.2702>
- Setiyawati, D., & Cahyono, N. (2023). Analisa Sentimen Pengguna Sosial Media Twitter Terhadap Perokok di Indonesia. *Indonesian Journal of Computer Science Attribution*, 12(1), 262–272. <https://doi.org/10.33022/ijcs.v12i1.3154>
- Sihombing, L. O., Hannie, H., & Dermawan, B. A. (2021). Sentimen Analisis Customer Review Produk Shopee Indonesia Menggunakan Algoritma Naive Bayes Classifier. *Edumatic: Jurnal Pendidikan Informatika*, 5(2), 233-242. <https://doi.org/10.29408/edumatic.v5i2.4089>
- Singalen, Y. A. (2022). Analisis Sentimen Wisatawan Melalui Data Ulasan Candi Borobudur di Tripadvisor Menggunakan Algoritma Naive Bayes Classifier. *Building of Informatics, Technology and Science (BITS)*, 4(3), 1343–1352. <https://doi.org/10.47065/bits.v4i3.2486>
- Solihin, F., Awaliyah, S., & Shofa, A. M. A. (2021). Pemanfaatan Twitter Sebagai Media Penyebaran Informasi Oleh Dinas Komunikasi dan Informatika. *Jurnal Pendidikan Ilmu Pengetahuan Sosial (JPIS)*, 1(13), 52–58.
- Sunata, M. H. A., Irwiensyah, F., & Hasan, F. N. (2024). Analisis Sentimen Calon Presiden 2024 di Media Sosial X Menggunakan Naive Bayes dan SMOTE. *Jurnal Media Informatika Budidarma*, 8(3), 1313-1322. <https://doi.org/10.30865/mib.v8i3.7708>
- Verawati, I., & Audit, B. S. (2022). Algoritma Naive Bayes Classifier Untuk Analisis Sentiment Pengguna Twitter Terhadap Provider By. u. *Jurnal Media Informatika Budidarma*, 6(3), 1411-1417. <https://doi.org/10.30865/mib.v6i3.4132>