



ANALISIS SENTIMEN MASYARAKAT PADA TEKS POSTINGAN DAN KOMENTAR DI PLATFORM X DALAM PEMILIHAN PRESIDEN 2024 DI INDONESIA MELALUI PEMROGRAMAN PYTHON DAN METODE NAIVE BAYES

Rizki Amanda¹, Adelia Yuandhika², Nezalfa Sabrina³,
Jasmine Aulia⁴, Cahya Eka Melati⁵, Endang Sholihatin⁶

^{1,2,3,4,5} Program Studi Sains Data, ⁶ Program Studi Linguistik Indonesia,
Universitas Pembangunan Nasional "Veteran" Jawa Timur
endang.sholihatin.ak@upnjatim.ac.id

ABSTRACT

In 2024, an election for president and vice president will be held. A political figure who wants to run for president will certainly consider their popularity based on public opinion. In this case, analyzing public sentiment towards the text of posts and comments on platform X becomes very relevant and essential. This research aims to: 1) find out the patterns and trends in public opinion related to the 2024 presidential election through the text of posts and comments uploaded on platform X, 2) evaluate the accuracy of the Naive Bayes algorithm for sentiment analysis on platform X, and 3) identify positive and negative sentiments directly related to the 2024 presidential election. The method used in this research is quantitative method with Naive Bayes model for classification. The results showed that: 1) Patterns and trends in public opinion related to the 2024 presidential election through the text of posts and comments uploaded on platform X lead to the figure "Ganjar" with the assumption that it is the most talked about, 2) The accuracy of the Naive Bayes algorithm for sentiment analysis on platform X is in the low category, which is 62.5%, and 3) The identification of positive and negative sentiments directly related to the 2024 Presidential Election shows a very low positive sentiment accuracy of 43.75% and a negative sentiment accuracy that is in the medium category, which is 81.25%.

Keywords: *Sentiment Analysis, Naive Bayes, Classification, Presidential Election*

PENDAHULUAN

Indonesia adalah salah satu negara yang menerapkan sistem demokrasi, yang ditandai dengan pelaksanaan pemilihan umum untuk memilih presiden dan wakil presiden. Di negara demokrasi, pemilihan umum biasanya dilakukan secara rutin. Pemilihan umum presiden dan wakil presiden dijadwalkan pada tahun 2024. Seorang tokoh politik yang berkeinginan mencalonkan diri sebagai presiden tentu akan mempertimbangkan popularitas

mereka berdasarkan opini masyarakat (LD Mahbubah & E Zuliarso, 2019). Dahulu, masyarakat dapat menyampaikan pendapat, kritik, dan saran melalui media cetak. Namun, dengan kemajuan teknologi di zaman sekarang telah mengubah cara masyarakat dalam mengungkapkan pendapat mereka, tetapi juga menjadi arena dimana opini dan pandangan dapat menyebar dan mempengaruhi opini publik secara luas melalui media sosial, platform media sosial yang



diminati dengan kehadiran jutaan pengguna yang aktif setiap hari adalah platform aplikasi X.

Dalam hal ini analisis sentimen masyarakat terhadap teks postingan dan komentar pada platform aplikasi X menjadi sangat relevan dan esensial. Analisis sentimen, sebagai salah satu teknik pengolahan bahasa alami yang bertujuan untuk memahami dan mengevaluasi opini, pandangan ataupun perasaan yang terkandung dalam teks. Dengan menggunakan metode analisis sentimen, kita dapat mengidentifikasi pola-pola opini yang muncul dari teks-teks tersebut, termasuk dukungan maupun kritik terhadap para calon presiden

Salah satu metode yang populer dan terbukti efektif untuk analisis sentimen adalah *Naive Bayes*. *Naive Bayes* adalah teknik klasifikasi probabilistik sederhana yang menghitung probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang ada (Gustientiedina, dkk 2019). Meskipun asumsi ini terlihat sederhana, namun dalam praktiknya, *Naive Bayes* telah terbukti memberikan hasil yang memuaskan dalam berbagai konteks analisis sentimen. Dalam hal analisis sentimen, *Naive Bayes* dapat digunakan untuk mengklasifikasikan teks-teks menjadi kategori sentimen, seperti positif dan negatif.

Beberapa penelitian yang relevan dengan penelitian ini, termasuk jurnal Analisa Sentimen Twitter pada Pilpres 2019 Menggunakan Algoritma Naive Bayes oleh Mahbubah dan Zuliarso (2019). Hasil penelitian tersebut menunjukkan akurasi sekitar 73%. Terdapat akurasi 78% untuk umpan balik negatif dan 66% untuk umpan balik positif. Penelitian kedua, berjudul "Analisis Ragam Bahasa Willy The Kid," dilakukan pada

Platform X oleh Hermanto, dkk (2024). Temuan penelitian ini menunjukkan bahwa dalam komunikasinya, Willy The Kid menggunakan alat kriminal seperti hasutan, ujaran kebencian, dan penghinaan. Penelitian ketiga dengan judul "Analisis Fonologis Penggunaan Bahasa Prancis dalam Aplikasi Media Sosial" dilakukan oleh Febrianty, dkk pada tahun 2023. Temuan penelitian tersebut menunjukkan bahwa banyak orang yang menggunakan bahasa Galia ketika memposting di aplikasi "X".

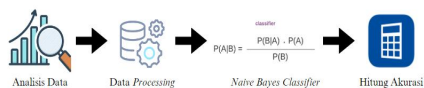
Penelitian ini bertujuan 1) untuk mengetahui pola dan tren dalam opini publik terkait Pilpres 2024 melalui teks postingan dan komentar yang diunggah pada platform aplikasi X; 2) Melakukan evaluasi terhadap akurasi algoritma *Naive Bayes* untuk analisis sentimen di platform tersebut; dan 3) Untuk mengidentifikasi sentimen positif dan negatif yang terkait langsung dengan Pilpres 2024. Dengan menggunakan pendekatan *Naive Bayes*, penelitian ini akan mencoba untuk mengidentifikasi dan menganalisis pola-pola sentimen yang dominan muncul dari teks-teks tersebut. Hasil ini diharapkan dapat memberikan wawasan yang penting mengenai pandangan masyarakat terhadap para calon presiden.

METODE PENELITIAN

Dalam studi ini, kami menerapkan metode kuantitatif dengan menggunakan data yang telah dikumpulkan dan diproses dari platform aplikasi X yang berisi postingan masyarakat terkait pemilihan presiden 2024. Sebanyak 156 dataset postingan dengan kata kunci "Pilpres", "Anies", "Prabowo", dan "Ganjar" telah kami himpun dari pengguna selama masa kampanye. Pemilihan pendekatan kuantitatif dipilih karena memungkinkan eksplorasi



rasi mendalam terhadap konten postingan, pemahaman konteks, serta ekstraksi informasi implisit yang mungkin tidak terungkap dengan pendekatan kuantitatif. Berikut adalah tahapan penelitian yang kami lakukan:



Gambar 1. Diagram Alir Sistem Analisis Sentimen

1. Analisis Data

Analisis data adalah suatu proses mengelola dan memahami data melalui langkah-langkah seperti organisasi, pengurutan, dan pembuatan satuan uraian dasar (Moleong, L. J, 2012). Kami melakukan analisis data dari platform X dengan kata kunci “Pilpres”, “Anies”, “Prabowo”, dan “Ganjar” terdiri dari 156 tweet, dimana sentimen setiap tweet ditentukan secara manual. Hasilnya menunjukkan bahwa terdapat 75 tweet dengan sentimen positif dan 81 tweet dengan sentimen negatif.

2. Data Processing

Data Processing pada analisis sentimen merupakan tahap untuk membersihkan (menghapus teks yang tidak relevan, tokenisasi, penghapusan kata-kata umum, penyesuaian huruf), ekstraksi fitur (mengubah kata-kata menjadi representasi numerik), serta penyiapan teks mentah hasil pengumpulan data dari web agar siap untuk dilakukan analisis sentimen (Landers, R. N.& Brusso, R. C, 2016).

3. Naive Bayes Classifier

Naive Bayes merupakan salah satu *algoritma machine learning* yang menggunakan Teorema Bayes untuk menghitung probabilitas dan melakukan klasifikasi.

Dalam proses klasifikasi, pengelompokan dilakukan berdasarkan probabilitas maksimum. Metode *Naive Bayes* sering dipilih oleh peneliti karena dianggap efektif dalam menghasilkan hasil yang akurat dengan tingkat akurasi yang tinggi (Kurniawan, 2019). Berikut adalah rumus untuk melakukan perhitungan *Naive Bayes*.

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad (1)$$

Keterangan:

- *A* : Hipotesis data dengan suatu *class* tertentu
- *B* : Data yang belum memiliki *class*
- *P(A|B)*: Probabilitas *posterior* yaitu peluang *class* A berdasarkan kondisi B
- *P(B|A)*: Nilai probabilitas *Likelihood* yaitu peluang *class* B berdasarkan kondisi pada hipotesis A
- *P(A)*: Probabilitas *prior* yaitu peluang awal munculnya kategori A
- *P(B)*: Probabilitas *Evidence* yaitu peluang munculnya kelas B

4. Hitung Akurasi

Pengujian akurasi dengan metode *Naive Bayes* dilakukan untuk menentukan seberapa akurat metode ini. Dalam pengujian ini, semua istilah yang diperoleh dari proses *pre-processing* akan digunakan tanpa seleksi fitur. Akurasi metode *Naive Bayes* diuji menggunakan *Confusion Matrix* dengan rumus berikut:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

Keterangan:

- *TP* : *True Positif*
- *TN* : *True Negatif*
- *FP* : *False Positif*
- *FN* : *False Negatif*

Rentang nilai hasil akurasi:

- Tinggi : Akurasi diatas 95%
- Sedang : Akurasi antara 80% - 95%



- Rendah : Akurasi antara 60% - 80%
- Sangat Rendah: Akurasi dibawah 60%

HASIL DAN PEMBAHASAN

3.1 Pola dan Tren Opini Terkait Pilpres 2024 di Platform X

Pembahasan mengenai pola dan tren opini terkait Pilpres 2024 didasarkan pada analisis data tweet yang dikumpulkan dari platform X menggunakan kata kunci "Pilpres", "Anies", "Prabowo", dan "Ganjar". Kata kunci ini dipilih untuk memastikan bahwa data yang terkumpul secara spesifik membahas Pemilihan Presiden Republik Indonesia tahun 2024. Data tersebut dikumpulkan secara acak berdasarkan kata kunci yang telah ditentukan, dengan tujuan mengidentifikasi sentimen publik terhadap calon-calon presiden dan proses pemilihan.

Tabel 1. Pembagian Data

Jenis Sentimen	Positif	Negatif
Latih	60	64
Uji	15	17

Berdasarkan tabel 1 diatas data yang dikumpulkan terdiri dari 156 tweet dalam bahasa Indonesia yang diunduh dari platform X. Data yang berhasil terkumpul diidentifikasi berdasarkan jenisnya, yaitu positif dan negatif. Ini dilakukan untuk melakukan analisis sentimen terhadap opini publik. Setelah itu, data itu dipisahkan menjadi dua bagian: data latih dan data uji, dengan pembagian 80% untuk data latih dan 20% untuk data uji.

Tabel 2. Data Latih

	Sentimen	Username	Postingan
39	Positif	@Gojeck7	Anies Mewakili rakyat lah bukan parpol,mosok g...
15	Positif	@Wong_Loman	Entah kenapa dimataku terlihat keren ya Anies ...
59	Positif	@DRoesyadi	Pengadilan mliter yang harusnya mengadili Prab...
114	Positif	@s14rNew	Pantesan kemarin pada minta2 Ganjar kesana, ru...
135	Positif	@adios7777173781	Tanpa KAMERA GANJAR bukan Apa2
...
50	Negatif	@RachlanNashidik	Orang cuma menggugat Prabowo menculik. Jangan ...
118	Negatif	@wahyu_bewe	Ditambah strategi pemenangan ganjar gak berhas...
141	Negatif	@jamesbakara	Otakmu sama dengan ganjar, Prabowo dulu mengaj...
21	Negatif	@NinaOsie	Kalau pak @prabowo idealis man utk bangsa ini,...
103	Negatif	@dhemit_is_back	Ngeyel karena Bingung Balikkan Modal Dari Big ...

Tabel 2 menunjukkan data latih digunakan untuk memberikan pelatihan pada model *Naive Bayes* dalam analisis sentimen terkait Pilpres 2024 pada platform aplikasi X, terdiri dari tiga kolom: Sentimen, Username, dan Postingan. Sentimen yang dilakukan terbagi menjadi dua kategori yaitu, "Positif" yang menggambarkan dukungan dan "Negatif" yang menggambarkan kritik terhadap subjek yang dibicarakan. Data ini akan digunakan untuk melatih model *Naive Bayes* yang berbasis probabilitas, yang kemudian akan mempelajari pola-pola dalam data untuk digunakan dalam memprediksi sentimen tweet baru, sehingga meningkatkan akurasi analisis sentimen terkait Pilpres 2024.

Tabel 3. Data Uji

	Sentimen	Username	Postingan
12	Positif	@Damzkiel	Njirrrrr keren banget emang Pak Anies, pokony...
127	Positif	@Ondoafi4	Karena dia tahu dalam hatinya bahwa ganjar ada...
25	Positif	@adarezekei	Jangan luoa itu efek Anies bro!!!
0	Positif	@CallMeAu	suka sama jawaban pak anies, yang gak terkecoh...
63	Positif	@topatiyo	Prabowo gelontorin dana 400M loh buat pak anie...
125	Positif	@jemmy18825	Loe sebaiknya beri saran sama Ganjar ... Udah ...
48	Positif	@AinurHamim	NasDem tinggalkan Anies stlh berhasil tambah k...
27	Positif	@MrsWuGrey	Amin. Semoga ibunda sehat selalu ya. Seumuran...
140	Positif	@gudel175959	Suruh buktiin aja,sampai saat ini Ganjar belum...
130	Positif	@GentongPisang	Nah ini bener. Jokowi ya gambling. Kalo di pdi...



Berdasarkan *Script 1*, proses pelatihan model Naive Bayes dimulai dengan membagi data menjadi dua bagian: data latih (training data) dan data uji (testing data). Langkah selanjutnya melibatkan pelatihan model *Naive Bayes* yang menyimpan seluruh probabilitas yang ada serta evaluasi model dimana metrik yang digunakan untuk melakukan evaluasi adalah metrik akurasi yang akan mengukur kemampuan model dalam memprediksi kelas yang benar.

Tabel 4. Model Confusion Matrix

		Nilai sebenarnya	
		TRUE	FALSE
Nilai prediksi	TRUE	TP (True Positive) <i>Correct result</i>	FP (False Positive) <i>Unexpected result</i>
	FALSE	FN (False Negative) <i>Missing result</i>	TN (True Negative) <i>Correct absence of result</i>

Model Confusion Matrix yang ditampilkan dalam Tabel 4 berisi informasi tentang bagaimana model klasifikasi *Naive Bayes* memprediksi data. Model ini terdiri dari empat elemen:

- *True Positive* (TP): jumlah data yang sebenarnya positif dan diprediksi sebagai positif oleh model.
- *True Negative* (TN): jumlah data yang sebenarnya negatif dan diprediksi sebagai negatif oleh model.
- *False Positive* (FP): jumlah data yang sebenarnya negatif tetapi diprediksi sebagai positif oleh model.
- *False Negative* (FN): jumlah data yang sebenarnya positif tetapi diprediksi sebagai negatif oleh model.

Accuracy: 0.625

Classification Report:				
	precision	recall	f1-score	support
Negatif	0.59	0.81	0.68	16
Positif	0.70	0.44	0.54	16
accuracy			0.62	32
macro avg	0.65	0.62	0.61	32
weighted avg	0.65	0.62	0.61	32

Gambar 4. Akurasi Klasifikasi Naive Bayes

Berdasarkan *output* akurasi klasifikasi *Naive Bayes* pada Gambar 7, menunjukkan tabel evaluasi untuk dua kelas, "Negatif" dan "Positif". *Precision* untuk negatif adalah 0.59 dan positif 0.70, sedangkan *recall* untuk negatif adalah 0.81 dan positif 0.44, dengan *F1-score* masing-masing 0.68 dan 0.54. Dari hasil diatas juga didapatkan akurasi keseluruhan model sebesar 62,5%. Hal ini berarti bahwa model mampu mengklasifikasikan data dengan benar sebanyak 62,5% dari total data. Dengan nilai akurasi yang berada pada rentang nilai 60%-80%, maka dapat dikatakan bahwa metode *Naive Bayes* kurang efektif dalam mengklasifikasikan sentimen positif dan negatif pada platform aplikasi X.

3.3 Identifikasi Sentimen Positif dan Negatif Terkait Pilpres 2024 di Platform X.

Confusion Matrix:

```
[[13  3]
 [ 9  7]]
```

Gambar 5. Hasil Confusion Matrix

Berdasarkan hasil diagonal utama *confusion matrix*, didapatkan hasil *presicion* untuk setiap kelas dan *rate error* sistem sebagai berikut:



a. *Precision Sentimen Positif*

$$\frac{TP}{TP + FP} = \frac{7}{7 + 9} = 0,4375$$

Berdasarkan hasil klasifikasi, tingkat ketepatan klasifikasi sentimen positif adalah 0,4375 atau 43,75%. Hal ini berarti bahwa model mampu mengklasifikasikan data dengan sentimen positif dengan benar sebanyak 43,75% dari total data yang mengandung sentimen positif. Berdasarkan nilai tersebut dapat disimpulkan bahwa model sangat kurang dalam mengklasifikasikan sentimen positif. Melalui hal ini juga diketahui bahwa tingkat kesalahan klasifikasi sentimen positif adalah 56,25%. Hal ini berarti bahwa model salah mengklasifikasikan data dengan sentimen positif sebanyak 56,25% dari total data yang mengandung sentimen positif.

b. *Precision Sentimen Negatif*

$$\frac{TN}{TN + FN} = \frac{13}{13 + 3} = 0,8125$$

Berdasarkan hasil klasifikasi, tingkat ketepatan klasifikasi sentimen negatif adalah 0,8125 atau 81,25%. Hal ini berarti bahwa model mampu mengklasifikasikan data dengan sentimen negatif dengan benar sebanyak 81,25% dari total data yang mengandung sentimen negatif. Hasil tersebut menunjukkan bahwa model dapat mengklasifikasikan sentimen negatif dengan cukup tepat. Diketahui pula tingkat kesalahan klasifikasi sentimen negatif adalah 18,75%. Hal ini berarti bahwa model salah mengklasifikasikan data dengan sentimen negatif sebanyak 18,75% dari total data yang mengandung sentimen negatif.

c. *Error Rate System*

$$\frac{FP + FN}{TP + TN + FN + FP} = \frac{9 + 3}{7 + 13 + 3 + 9} = 0,375$$

Berdasarkan hasil perhitungan tersebut, *error rate system* sebesar 0,375 atau 37,5%. Hal ini mengindikasikan bahwa sekitar 37,5% dari seluruh data tidak terklasifikasikan secara benar oleh algoritma *Naive Bayes Classifier*. Meskipun akurasi secara umum mencapai 62,5%, *error rate system* yang relatif tinggi menunjukkan adanya ruang untuk perbaikan dalam proses klasifikasi sentimen.

KESIMPULAN

Berdasarkan penelitian yang telah dilakukan didapatkan kesimpulan bahwa dari total 156 tweet tentang Calon Presiden RI yang dikumpulkan secara acak dari platform aplikasi X. Topik yang paling banyak dibicarakan dan sedang tren dalam tweet dalam aplikasi X adalah calon presiden RI nomor urut 3, yaitu Ganjar Pranowo.

Didapatkan pula hasil klasifikasi sentimen menggunakan metode *Naive Bayes* menghasilkan nilai akurasi sebesar 62,5% atau 0,625. Nilai tersebut berada diantara 60% - 80% sehingga dapat dikatakan bahwa metode *Naive Bayes* kurang cukup efektif dalam mengklasifikasikan sentimen positif dan negatif pada tweet dalam platform aplikasi X.

Melalui analisis, didapatkan jumlah tweet positif sebanyak 75 tweet dan negatif sebanyak 81 tweet. *Naive Bayes* mampu mengklasifikasikan secara tepat sentimen positif sebesar 43,75% dan negatif sebesar 81,25%. Nilai sentimen positif yang kurang dari 60% menunjukkan bahwa keefektifan model bernilai sangat rendah. Sedangkan nilai sentimen negatif yang berada pada rentang nilai 80% - 95% menunjukkan keefektifan sedang. Sehingga metode ini mungkin dapat lebih ditingkatkan lagi untuk mendapatkan hasil yang lebih baik.



Melalui analisis sentimen terhadap Calon Presiden RI ini dapat diketahui informasi tentang tren serta sentimen positif dan negatif dari Calon Presiden RI 2024. Analisis ini sangat berguna untuk mengetahui perkiraan dukungan publik, kecenderungan opini masyarakat, dan potensi dampaknya terhadap hasil pemilihan umum Presiden Republik Indonesia 2024.

DAFTAR PUSTAKA

- Mahbubah, L. D., & Zuliarso, E. (2019). *Analisa Sentimen Twitter Pada Pilpres 2019 Menggunakan Algoritma Naive Bayes*.
- Kurnila, M. T. I. N. A., Kanafi, F., Debora, S., Salsabila, D., Februani, A. F., Suprpto, G. P., ... & Sholihatin, E. (2023). *Analisis Kesantunan Berbahasa Warganet Twitter@ YUPIEN_FESS Dalam Menfess UPNSBY*. *Innovative: Journal Of Social Science Research*, 3(6), 2417-2429.
- Taufiq, M., Maliki, D. O., Maldini, A. S., Ekamartha, K. N., Saputra, K. N. C., Ahmad, S. H., ... & Sholihatin, E. (2023). *Pentingnya Etika Berbahasa Sebagai Upaya Pencegahan Kasus Kejahatan Berbahasa Di Media Digital*. *Bureaucracy Journal: Indonesia Journal of Law and Social-Political Governance*, 3(2), 2116-2125.
- Febrianty, T., Ramadhani, S., Putri, A. R. I., Octaviani, M. A., & Sholihatin, E. (2023). *PHONOLOGICAL ANALYSIS OF SLANG USE IN SOCIAL MEDIA APPLICATION "X"*. *JOURNAL OF HUMANITIES AND SOCIAL STUDIES*, 1(03), 1249-1256.
- Haq, M. F. T., Hendratno, A., Arfa, M. N., Amanullah, A. F., & Sholihatin, E. (2024). *ANALYSIS OF CHANGES IN THE FOREIGN LANGUAGE LEXICON USED BY THE MILLENNIAL GENERATION IN THE X APPLICATION IN INDONESIA*. *JOURNAL OF HUMANITIES AND SOCIAL STUDIES*, 2(01), 49-56.
- Soetanto, B. J., Akbar, D. A. H., Anindhyta, E. D. X., Fadlurahman, F., Nurunnisa, I. A., Paramita, M. D., ... & Sholihatin, E. (2023). *PENGUNAAN BAHASA TABU OLEH GENERASI Z KOTA SURABAYA DI MEDIA SOSIAL TIKTOK*. *Serunai: Jurnal Ilmiah Ilmu Pendidikan*, 9(2).
- Hermanto, A. A. P., Damayanti, A. D., Salsabilla, A. A., Terra, E. A., Sholihatin, E., Putri, S., & Marsanda, V. (2024). *Analisis Ragam Bahasa Willy The Kid di Platform X*. *EI-Mujtama: Jurnal Pengabdian Masyarakat*, 4(3), 207-213.
- Moleong, L. J. (2012). *Metodologi Penelitian Kualitatif*. Bandung: PT. Remaja Rosdakarya.
- Landers, R. N., Brusso, R. C., Cavanaugh, K. J., & Collmus, A. B. (2016). *A primer on theory-driven web scraping: Automatic extraction of big data from the Internet for use in psychological research*.
- Kurniawan, I. & Susanto, A., 2019. *Implementasi Metode K-Means dan Naive Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019*. *Jurnal Eksplora Informatika*, IX(1).
- Gustientiedina, G., Siddik, M., & Deselinta, Y. (2019). *Penerapan Naive Bayes untuk Memprediksi Tingkat Kepuasan Mahasiswa Terhadap Pelayanan Akademis*. *Jurnal Infomedia: Teknik Informatika, Multimedia, dan Jaringan*, 4(2), 89-93.