

ANALISIS SENTIMEN PENGGUNA TIKTOK TERHADAP PILKADA SITUBONDO TAHUN 2024 MENGGUNAKAN *NAÏVE BAYES*

Ma'ruf Ubaidillah¹, Nur Azise², Achmad Baijuri³

¹ Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Ibrahimy, Indonesia

^{1,2} Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Ibrahimy, Indonesia

¹mrfubaidillah@gmail.com, ²nuriza3010@gmail.com, ³bayubai@gmail.com

Abstrak

Media sosial TikTok menjadi *platform* populer di kalangan masyarakat Situbondo, dan hampir semua lapisan masyarakat menggunakan *platform* karena kemudahannya dalam membuat dan membagikan video pendek dengan latar musik sebagai sarana hiburan, ekspresi diri, berinteraksi. Pada pemilihan kepala daerah tahun 2024, TikTok tidak hanya digunakan sebagai sarana hiburan, akan tetapi sebagai sarana menyampaikan opini, baik positif, netral, maupun negatif terhadap kandidat. Namun, banyaknya komentar tanpa penyaringan menimbulkan beragam persepsi yang sulit dipetakan. Kondisi ini mendorong penelitian untuk menganalisis sentimen masyarakat melalui pendekatan *data mining* menggunakan Algoritma *Naive Bayes* terhadap opini masyarakat melalui *platform* TikTok, sehingga opini publik dapat dikaji secara lebih objektif dan bermanfaat bagi strategi komunikasi politik. Penelitian ini menggunakan metode kuantitatif dengan pendekatan eksperimental, dimana pada penelitian sebelumnya membahas isu nasional pada *platform* X dan *Instagram*. Pengumpulan data penelitian ini melalui teknik *web scraping* dari 494 komentar TikTok yang membahas PILKADA Situbondo. Data yang terkumpul diproses melalui tahapan *text preprocessing*, kemudian diklasifikasikan menggunakan algoritma *Multinomial Naive Bayes*. Hasil penelitian menunjukkan bahwa model klasifikasi berbasis *Naive Bayes* mampu mengklasifikasikan sentimen dengan akurasi sebesar 70,59%. Kinerja model lebih baik dalam mengidentifikasi sentimen positif dibandingkan netral, yang masih sulit diklasifikasikan secara tepat. Temuan ini membuktikan bahwa metode *Naive Bayes* cukup efektif dalam pemetaan opini publik berbasis media sosial TikTok, meskipun masih terdapat ruang pengembangan melalui optimalisasi algoritma atau penambahan variasi dataset. Temuan ini dapat dimanfaatkan oleh kandidat, tim kampanye, maupun penyelenggara pemilu dalam merancang strategi komunikasi politik yang lebih tepat sasaran.

Kata kunci: analisis sentimen, pilkada, TikTok, *data mining*, *naive bayes*

1. Pendahuluan

Keterlibatan aktif warga negara dalam partisipasi politik dan pemilihan kepala daerah (PILKADA) merupakan pondasi utama pada sebuah sistem demokrasi untuk menentukan arah kebijakan politik (Heniasari et al., 2025). Perkembangan dan kemajuan dalam bidang teknologi informasi dan komunikasi, khususnya penggunaan sosial media memiliki pengaruh dan peranan penting di tengah masyarakat, termasuk dalam konteks politik (Simanullang et al., 2025). PILKADA di Kabupaten Situbondo tahun 2024 menjadi salah satu masa krusial ketika masyarakat akan memilih pemimpin yang dianggap dapat melanjutkan kepemimpinan daerah atau membawa perubahan yang signifikan. Media sosial telah muncul sebagai tempat utama di era digital saat ini bagi individu untuk menyuarakan pemikiran *warganet*, terlibat dalam wacana politik, dan mengkomunikasikan tujuan *warganet* dengan mengacu pada politisi saingan.

Salah satu media sosial yang saat ini sangat populer dan banyak digunakan oleh masyarakat Indonesia pada umumnya dan masyarakat Situbondo khususnya adalah TikTok. TikTok menjadi aplikasi sosial media yang memungkinkan pengguna membuat dan berbagi konten berupa video pendek yang disertai dengan latar dan musik yang menarik (Apriliani et al., 2023). Selain menjadi sarana ekspresi diri, hiburan, dan interaksi sosial, TikTok juga menjadi sarana penyampaian visi misi, sarana komunikasi politik dan membangun personal branding hingga membentuk suatu propaganda politik untuk membentuk dinamika politik (Simanullang et al., 2025) (Bhilton M. Obidje, 2025).

Banyaknya video, komentar, dan tagar TikTok yang berkaitan dengan PEMILU menimbulkan berbagai wacana publik yang hidup dan mengungkapkan opini tentang kandidat atau topik tertentu, sehingga dibutuhkan analisis sentiment terhadap data TikTok untuk memeriksa teks digital, mengidentifikasi sentimen atau emosi

yang ada di dalamnya, dan memastikan komentar warganet positif, netral, atau negatif (Khusnul Khotimah, 2024). Terdapat beberapa metode Machine Learning untuk melakukan klasifikasi analisis sentiment terhadap opini publik seperti *Naïve Bayes Clasifier*, *K-Nearest Neighbor*, *Decision Tree* untuk mendapatkan hasil terbaik (Cahya Kamilla et al., 2024). Penelitian ini menggunakan metode *Naïve Bayes* karena dikenal dengan kecepatan mengolah dalam jumlah besar dan menghasilkan akurasi yang tinggi (Siddiq et al., 2025) (Putri Nurrahmah Wear et al., 2024).

Hasil penelitian (Nugroho et al., 2024) pada Analisis Sentimen Dugaan Pelanggaran PEMILU 2024 dengan menganalisis komentar negatif dan positif berdasarkan *Tweet* menggunakan algoritma *Naïve Bayes* menghasilkan nilai akurasi sebesar 84,13%, nilai *precision* sebesar 77,36%, dan nilai *recall* 84,13%. penelitian (Hidayatullah et al., 2026) Hasil analisis menggunakan algoritma *Naïve Bayes* dengan representasi TF-IDF menunjukkan hasil akurasi lebih tinggi dari dua metode lainnya, yaitu pada Paslon 1 sebesar 86,76%, Paslon 2 sebesar 76,96%, dan Paslon 3 sebesar 72,79%. Analisis menggunakan metode KNN dengan TF-IDF menghasilkan nilai akurasi sebesar 67,40% (paslon 1), Paslon 2 sebesar 71,32%, sedangkan algoritma Decision tree dengan representasi transformer menghasilkan nilai akurasi sebesar 72,55%. Hasil Penelitian pada analisis sentimen aplikasi mitra darat juga menunjukkan hasil akurasi algoritma *Naïve Bayes* lebih tinggi dari algoritma KNN (Wijaya et al., 2024). Selanjutnya pada penelitian (Pratama et al., 2026) analisis sentiment menggunakan *Naïve Bayes* dengan pembobotan TF-IDF menghasilkan hasil akurasi sebesar 90%.

Dalam konteks PILKADA di Kabupaten Situbondo, penggunaan *Naive Bayes* untuk menganalisis sentimen di TikTok diharapkan dapat membantu mengidentifikasi tren opini publik, memberikan wawasan kepada para kandidat mengenai berbagai persepsi *warganet* dengan memetakan komentar yang positif, netral, dan negatif. Tujuan penelitian ini untuk menguji hasil performa model algoritma *Naive Bayes* tentang opini publik masyarakat terhadap PILKADA Kabupaten Situbondo pada *platform* TikTok, serta memberikan gambaran tentang bagaimana *warganet* memandang PILKADA dalam rangka memahami opini publik dengan lebih baik dan meningkatkan taktik komunikasi politik para kandidat, staf kampanye, dan penyelenggara PILKADA dapat mengambil manfaat besar dari temuan analisis sentimen. Oleh karena itu, penelitian ini sangat relevan dengan PILKADA di Kabupaten Situbondo. Selain membantu pengelolaan data opini publik, analisis sentiment dapat membantu mengatasi pendapat *warganet* yang kontroversial dengan lebih baik dan dapat diterapkan dalam konteks sosial politik di masa depan.

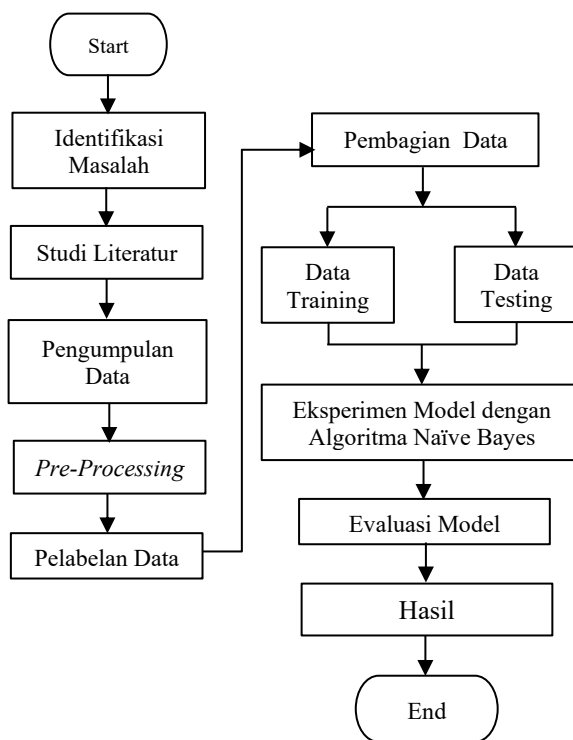
2. Metode

2.1. Jenis penelitian

Penelitian ini merupakan penelitian kuantitatif menggunakan metode eksperimen yang menghubungkan sebab-akibat antara video dan komentar *warganet* pada *platform* TikTok, untuk menentukan dan mengetahui hasil terbaik klasifikasi sentiment masyarakat terhadap PILKADA Kabupaten Situbondo berupa komentar positif, negatif, dan netral menggunakan algoritma *Naive Bayes* (Mursyid et al., 2024) (Sri Yani Kusumastuti et al., 2024).

2.2. Tahapan Penelitian

Metode penelitian ini dilakukan melalui beberapa tahapan sebagaimana ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Berdasarkan tahapan-tahapan penelitian yang ditunjukkan pada gambar di atas, maka penjelasan masing-masing tahapan tersebut sebagai berikut:

1. Identifikasi Masalah

Identifikasi masalah merupakan tahap awal penelitian, dimana penulis menemukan banyak komentar *warganet* berupa komentar negatif, positif, dan netral pada *platform* TikTok tentang PILKADA Kabupaten Situbondo yang tidak difilter, sehingga menimbulkan opini publik baik tentang calon kandidat, maupun program yang akan dilaksanakan oleh kandidat terpilih.

2. Studi literatur

Tahap ini dilakukan dengan membaca, merangkum buku, jurnal yang relevan dengan topik penelitian untuk memilih metode terbaik

terkait analisis sentiment PILKADA Kabupaten Situbondo (Hidayah et al., 2025) (Sri Yani Kusumastuti et al., 2024).

3. Pengumpulan Data Set

Pengumpulan data set dilakukan dengan melakukan *scrapping* terhadap komentar warganet pada Video di Platform TikTok terkait PILKADA Kabupaten Situbondo dan disimpan dalam bentuk file CSV dengan memanfaatkan perangkat lunak data miner (Salsabilla et al., 2025).

4. Pre-processing

Pre-processing data merupakan tahap membersihkan data yang diperoleh dari proses *scrapping* untuk menghilangkan data yang bermasalah atau tidak konsisten, sehingga data-data tersebut dapat dipahami (Cahya Kamilla et al., 2024) (Salsabilla et al., 2025). *Pre-processing* terdiri dari beberapa tahapan (Cahya Kamilla et al., 2024) antara lain :

a. *Cleaning*

Cleansing adalah proses menghilangkan data atau teks yang kosong atau tidak perlu untuk memverifikasi bahwa data analitik sudah akurat dan bersih.

b. *Case Folding*

Case folding adalah Teknik mengubah huruf kapital dalam kumpulan data menjadi huruf kecil dikenal untuk menyamakan format huruf untuk analisis yang lebih sederhana.

c. *Tokenizing*

Frasa akan dibagi menjadi unit kata, atau token, selama proses *Tokenizing*, dan ini akan dipisahkan dengan koma atau tanda baca lainnya.

d. *Remove Stopword*

Remove Stopword merupakan proses di mana kata-kata penghubung yang dianggap tidak perlu dan tidak mempengaruhi emosi kalimat akan dihilangkan dari analisis.

e. *Steaming*

Proses merubah kata menjadi bentuk dasar dengan menghilangkan imbuhan, baik awalan, sisipan, maupun akhiran.

5. Pelabelan Data

Proses pelabelan dilakukan dengan mengumpulkan data set, kemudian mengelompokkannya pada bentuk sentiment negatif, positif, dan netral (Mario et al., 2025).

6. Split Data

Tahap ini merupakan tahap pembagian data menjadi dua, yaitu data *training* untuk melatih model klasifikasi, dan data *testing* untuk melakukan pengujian model dan mendapatkan hasil yang akurat dengan perbandingan 80% data training dan 20% data testing (Sulistiowati & Santoso, 2025).

7. Eksperimen Model

Setelah dilakukan pembagian data, tahap selanjutnya melakukan eksperimen

menggunakan algoritma *Naive Bayes* yang merupakan metode klasifikasi dengan teorema Bayes untuk menentukan kelas pada suatu item berdasarkan fitur yang menghasilkan performa akurasi tinggi menggunakan perhitungan probabilitas pada Persamaan (1). (Apriyani et al., 2024) (Arifqi et al., 2024).

$$P(X|H) = \frac{P(X|H)P(H)}{P(X)} \tag{1}$$

Keterangan:

X: data sampel dengan kelas yang tidak diketahui

H: hipotesis data X yang merupakan data dengan kelas spesifik

P(H|X): probabilitas hipotesis H berdasarkan kondisi X

P(H): probabilitas dari hipotesa H

P(H|X): probabilitas hipotesis X berdasarkan kondisi H

P(X): probabilitas X

Ekperimen yang dilakukan pada penelitian ini menggunakan Bahasa pemrograman Python untuk menganalisis penggunaan algoritma Naive Bayes pada data set yang telah didapatkan.

8. Evaluasi Model

Setelah dilakukan eksperimen data dengan Algoritma *Naive Bayes*, tahap selanjutnya adalah melakukan evaluasi model menggunakan Teknik *confusion matrix* untuk mengetahui komentar warganet pada video terkait PILKADA di platform TikTok berdasarkan 3 kategori, yaitu positif, negatif, dan netral dengan mencari nilai *True Positive* (TP), *False Positive* (FP), *True Negative* (TN), dan *False Negative* (FN) seperti pada Tabel 1 (Arifqi et al., 2024).

Tabel 1 *Confusion matrix*

<i>Confusion matrix</i>		Prediksi		
		Negatif	Netral	Positif
Kelas	Negatif	TN	FNt	FP
	Netral	FN	TNt	FP
	Positif	FN	FNt	TP

Keterangan:

a. *True Positive* (TP), adalah kelas yang diprediksi mempunyai nilai positif dan faktanya positif.

b. *False Positive* (FP), adalah nilai kelas yang diprediksi mempunyai nilai positif dan faktanya negatif atau netral.

c. *True Negative* (TN), adalah kelas yang diprediksi mempunyai nilai negatif dan faktanya negatif.

d. *False Negatif* (FN), adalah kelas yang diprediksi mempunyai nilai negatif dan faktanya positif atau netral

e. *True Neutral* (TNt), adalah kelas yang diprediksi mempunyai nilai netral dan faktanya netral.

f. *False Neutral* (FNt), adalah kelas yang diprediksi.

Penghitungan *confusion matrix* menghasilkan empat perhitungan:

a. Akurasi

Nilai akurasi diperoleh berdasarkan hasil dari perhitungan presentse suatu kelas terprediksi oleh model yang sudah dibuat dengan Persamaan (2) berikut.

$$Accuracy = \frac{TP+TN}{(TP+TN+FP+FN)} \times 100\% \quad (2)$$

b. Precision

Presisi merupakan hasil penilaian yang sesuai antara permintaan pengguna dengan jawaban yang diberikan oleh sistem menggunakan Persamaan (3) berikut.

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

c. Recall

Untuk menghitung *recall* menggunakan Persamaan (4) berikut.

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

d. F1-score

F1-score merupakan nilai yang dihasilkan dari perhitungan kombinasi nilai *precision* dan *recall* dengan Persamaan (5) berikut.

$$F1 - Score = 2x \frac{precision \times recall}{precision + recall} \quad (5)$$

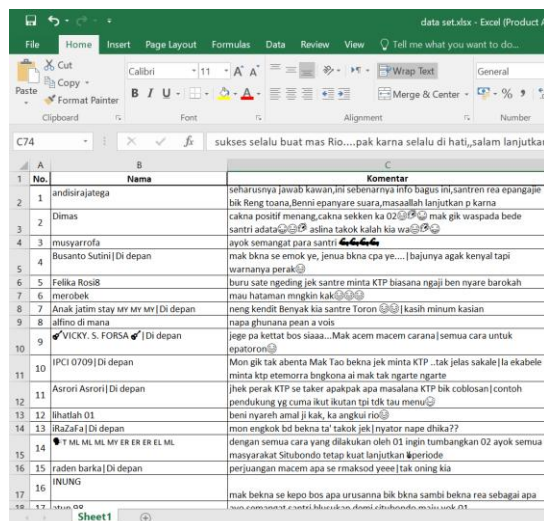
3. Hasil dan Pembahasan

3.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini menggunakan data sekunder berupa komentar di platform TikTok, yang berkaitan dengan PILKADA di Kabupaten Situbondo. Data tersebut diperoleh dengan cara melakukan *scrapping* pada komentar terkait PILKADA di Kabupaten Situbondo pada platform TikTok, memanfaatkan perangkat lunak *Data Miner*. *Scrapping data* dimulai dengan memasukkan kata kunci yang berhubungan dengan penelitian tersebut di aplikasi TikTok, kemudian mencari video TikTok yang memiliki banyak respon dari *warganet*. Data yang telah dikumpulkan melalui *scrapping* kemudian diolah menggunakan *text mining* untuk membersihkan dan mempersiapkan pada analisis lebih lanjut. Data yang berhasil didapatkan dalam penelitian ini sebanyak 494 data ditampilkan pada aplikasi Microsoft Office *excel*. Pada Gambar 2 berikut merupakan hasil *scrapping* yang disimpan pada aplikasi *Microsoft Office Excel*.

3.2. Pre-processing Data

Sebelum masuk pada tahap *preprocessing* data, dilakukan proses menterjemah komentar *warganet* yang berbahasa Madura ke dalam bahasa Indonesia. Beberapa tahapan yang dilakukan pada *preprocessing* antara lain:



Gambar 2. Hasil *Scrapping* Data

1. Cleaning

Cleaning merupakan tahap awal prosedur *preprocessing* dimulai. Untuk meminimalkan gangguan atau kesalahan selama pemrosesan data, karakter non alfabet dibersihkan atau dihilangkan pada langkah ini. *whitespace*, *emoticon*, *link double*, *URL*, *mention*, duplikasi teks dan angka duplikat semuanya dihapus pada tahap ini. Pada prosedur ini juga menggunakan Fungsi pustaka NLTK (*Natural Language Toolkit*) dari bahasa *Python* Pada tabel 2 menunjukkan teks sebelum dan sesudah proses *cleaning*.

Tabel 2. Proses *Cleaning*

Komentar	Cleaning
seharusnya jawab kawan, ini sebenarnya info bagus ini, santren rea epangajie bik Reng toana, Benni epanyare suara, masaallah lanjutkan p karna	seharusnya jawab kawan ini sebenarnya info bagus ini Santri ini disuruh ngaji oleh orang tuanya bukan disuruh cari suara masaallah lanjutkan p karna
cakna positif menang, cakna sekken ka 02 😂😂😂 mak gik waspada bede santri adata 😂😂😂 aslina takok kalah kia wa 😂😂😂	katanya positif menangkatanya sudah mantap ke kosong dua kok masih khawatir ada santri men dataaslinya takut kalah juga ya
Ayok semangat para santri 🤔🤔🤔🤔	Ayo semangat para santri
mak bkna se emok ye, jenua bkna cpa ye.... buru satea ngeding jek santre minta KTP biasana ngaji ben nyare barokah	kok kamu yang repot ya emangnya kamu siapa baru sekarang dengar santri minta KTP biasanya ngaji dan cari barokah

2. Case Folding

Case folding adalah prosedur *preprocessing data* yang berusaha mengubah semua karakter menjadi huruf kecil. Pada tahap ini dengan memanfaatkan fungsi *.lower()* dari pustaka *pandas* teks akan tampil dalam format huruf kecil yang seragam setelah proses ini dijalankan. Pada

tabel 3 merupakan proses sebelum dan sesudah *case folding*.

Tabel 3. Proses *Case Folding*

Cleaning	Case Folding
seharusnya jawab kawan ini sebenarnya info bagus ini Santri ini disuruh ngaji oleh orang tuanya bukan disuruh cari suara masaallah lanjutkan p karna	seharusnya jawab kawan ini sebenarnya info bagus ini santri ini disuruh ngaji oleh orang tuanya bukan disuruh cari suara masaallah lanjutkan p karna
katanya positif menangkatanya sudah mantap ke kosong dua kok masih khawatir ada santri men dataaslinya takut kalah juga ya	katanya positif menangkatanya sudah mantap ke kosong dua kok masih khawatir ada santri men dataaslinya takut kalah juga ya
ayo semangat para santri	ayo semangat para santri
kok kamu yang repot ya emangnya kamu siapa baru sekarang dengar santri minta KTP biasanya ngaji dan cari barokah	kok kamu yang repot ya emangnya kamu siapa baru sekarang dengar santri minta ktp biasanya ngaji dan cari barokah

3. *Tokenizing*

Tokenizing. Untuk memudahkan analisis pada langkah pemrosesan berikutnya, tujuan utamanya adalah memisahkan setiap kata yang membentuk sebuah frasa. Dalam proses ini memanfaatkan fungsi dari `.split()` atau NLTK dan perpustakaan NLP lainnya. Pada tabel 4 merupakan proses sebelum dan sesudah *Tokenizing*.

Tabel 4 Proses Sebelum dan Sesudah *Tokenizing*

<i>Case Folding</i>	<i>Tokenizing</i>
seharusnya jawab kawan ini sebenarnya info bagus ini santri ini disuruh ngaji oleh orang tuanya bukan disuruh cari suara masaallah lanjutkan p karna	['seharusnya', 'jawab', 'kawan', 'ini', 'sebenarnya', 'info', 'bagus', 'ini', 'santri', 'ini', 'disuruh', 'ngaji', 'oleh', 'orang', 'tuanya', 'bukan', 'disuruh', 'cari', 'suara', 'masaallah', 'lanjutkan', 'p', 'karna']
katanya positif menangkatanya sudah mantap ke kosong dua kok masih khawatir ada santri men dataaslinya takut kalah juga ya	['katanya', 'positif', 'menangkatanya', 'sudah', 'mantap', 'ke', 'kosong', 'dua', 'kok', 'masih', 'khawatir', 'ada', 'santri', 'men', 'dataaslinya', 'takut', 'kalah', 'juga', 'ya']
ayo semangat para santri	['ayo', 'semangat', 'para', 'santri']
kok kamu yang repot ya emangnya kamu siapa baru sekarang dengar santri minta ktp biasanya ngaji dan cari barokah	['kok', 'kamu', 'yang', 'repot', 'ya', 'emangnya', 'kamu', 'siapa', 'baru', 'sekarang', 'dengar', 'santri', 'minta', 'ktp', 'biasanya', 'ngaji', 'dan', 'cari', 'barokah']

4. *Remove Stopword*

Pada tahap ini melibatkan penghilangan kata-kata yang dianggap secara signifikan berkontribusi pada makna dokumen. Kata-kata yang bersifat umum dan tidak menyampaikan informasi yang tepat, seperti “ini,” “kok,” “masih,” “ke,” “atau,” dan sebagainya, biasanya dihilangkan. Untuk menyaring hanya istilah-istilah yang relevan. Pada tabel 5 berikut merupakan proses sebelum dan sesudah *remove stopwords*.

Tabel 5 Proses Sebelum dan Sesudah *Remove Stopword*

<i>Tokenizing</i>	<i>Remove Stopword</i>
['seharusnya', 'jawab', 'kawan', 'ini', 'sebenarnya', 'info', 'bagus', 'ini', 'santri', 'ini', 'disuruh', 'ngaji', 'oleh', 'orang', 'tuanya', 'bukan', 'disuruh', 'cari', 'suara', 'masaallah', 'lanjutkan', 'p', 'karna']	['kawan', 'info', 'bagus', 'santri', 'disuruh', 'ngaji', 'orang', 'tuanya', 'disuruh', 'cari', 'suara', 'masaallah', 'lanjutkan', 'p', 'karna']
['katanya', 'positif', 'menangkatanya', 'sudah', 'mantap', 'ke', 'kosong', 'dua', 'kok', 'masih', 'khawatir', 'ada', 'santri', 'men', 'dataaslinya', 'takut', 'kalah', 'juga', 'ya']	['positif', 'menangkatanya', 'mantap', 'kosong', 'khawatir', 'santri', 'men', 'dataaslinya', 'takut', 'kalah', 'ya']
['ayo', 'semangat', 'para', 'santri']	['ayo', 'semangat', 'santri']
['kok', 'kamu', 'yang', 'repot', 'ya', 'emangnya', 'kamu', 'siapa']	['repot', 'ya', 'emangnya']
['baru', 'sekarang', 'dengar', 'santri', 'minta', 'ktp', 'biasanya', 'ngaji', 'dan', 'cari', 'barokah']	['dengar', 'santri', 'ktp', 'ngaji', 'cari', 'barokah']

5. *Stemming Data*

Stemming merupakan tahap terakhir dalam prosedur persiapan teks. Tujuan utamanya adalah mengembalikan kata-kata yang terinfleksi ke bentuk dasarnya. Misalnya, kata-kata “bermain”, “memasak”, dan “terlambat” akan diubah menjadi “main”, “masak”, dan “lambat”. Prosedur ini dilakukan setelah tahap tokenisasi dan penghapusan kata *stopword*. Hasil *steaming* ditunjukkan pada tabel 6 berikut.

Tabel 6. *Stemming Data*

<i>Remove Stopword</i>	<i>Stemming Data</i>
['kawan', 'info', 'bagus', 'santri', 'disuruh', 'ngaji', 'orang', 'tuanya', 'disuruh', 'cari', 'suara', 'masaallah', 'lanjutkan', 'p', 'karna']	kawan info bagus santri suruh ngaji orang tua suruh cari suara masaallah lanjut p karna
['positif', 'menangkatanya', 'mantap', 'kosong', 'khawatir', 'santri', 'men', 'dataaslinya', 'takut', 'kalah', 'ya']	positif menangkatanya mantap kosong khawatir santri men dataaslinya takut kalah ya
['ayo', 'semangat', 'santri']	ayo semangat santri
['repot', 'ya', 'emangnya']	repot ya emang
['dengar', 'santri', 'ktp', 'ngaji', 'cari', 'barokah']	dengar santri ktp ngaji cari barokah

3.3. *Wordcloud*

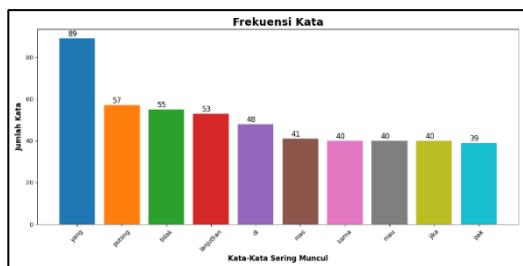
Wordcloud merupakan tahap untuk mendapatkan kata-kata yang sering ditemukan pada kolom komentar warganet (Irvandi et. al, 2023). Tahap ini terdiri dari dua proses, yaitu:

1. *Wordcloud* sebelum *Pre-processing*

Kata-kata yang sering ditemukan dan mendominasi pada tahap *wordcloud* sebelum *Pre-processing*, seperti “yang”, “potong”, “tidak”, dan “lanjutkan” ditunjukkan pada Gambar 3. Sedangkan frekuensi kata sebelum dilakukan *Pre-processing* ditunjukkan pada Gambar 4.



Gambar 3. Wordcloud Sebelum Pre-processing



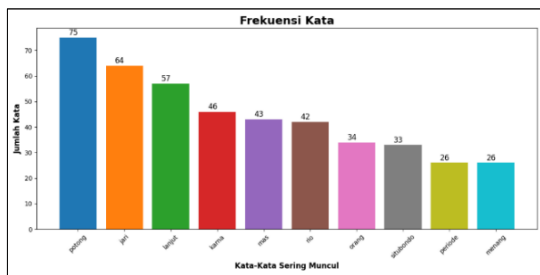
Gambar 4. Frekuensi Kata Sebelum Pre-processing

2. Wordcloud setelah Pre-processing
 Wordcloud setelah Pre-processing hasil penghapusan kata-kata yang tidak perlu dan pengurangan kata-kata ke bentuk dasarnya mewakili ide utama teks, seperti yang ditunjukkan pada Gambar 5.



Gambar 5. Wordcloud Setelah Pre-processing

Sedangkan frekuensi kata setelah Pre-processing ditunjukkan pada Gambar 6.



Gambar 6. Frekuensi Kata setelah Pre-processing

3.4. Pelabelan Data

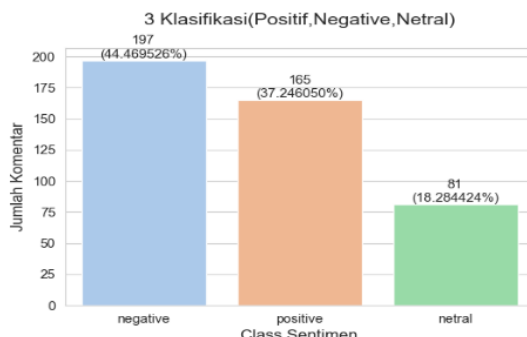
Proses pelabelan ini dilakukan secara manual dan setiap potongan data harus diperiksa secara terpisah untuk menentukan kategorisasi sentimennya. Konteks dan makna umum teks dipertimbangkan saat memberikan label. Sentimen positif, negatif, dan netral adalah tiga kategori utama

yang digunakan dalam membagi sentimen. Gambar 7 merupakan contoh dari data yang sudah diberi label.

stemming data	Sentimen
0 kawan info bagus santri suruh ngaji orang tua suruh cari suara masaallah lanjut p karna	negative
1 positif menangkarnya mantap kosong khawatir santri men dataasalnya takut kalah ya	negative
2 ayo semangat santri	positive
3 repot ya orang	negative
4 dengar santri ktp ngaji cari barokah	negative
5 hataman kak	positive
6 kendit santri turun	negative
7 guna pakai voice	negative
8 jaga ketat bos bermacam pakai	negative
9 ngomong ktp gak banget	negative
10 ktp sampek ngebacot ktp coblos	negative
11 cari amal kak rio	negative
12 kalok gak takut	negative
13 kosong tumbang kosong ayok masyarakat situbondo kuat lanjut periode	negative
14 juang ya	netral
15 kepo bos urus	negative
16 ayo semangat santri blusukan situbondo maju yok kosong	positive
17 terus juang ulama salah paham iya santri masuk paksa adukung calon	positive
18 santri serbaguna nama suruh aja	positive
19 lapor biar bukti	negative

Gambar 7. Hasil Pelabelan Data

Distribusi data berdasarkan hasil penandaan manual secara visual ditampilkan pada Gambar 8.



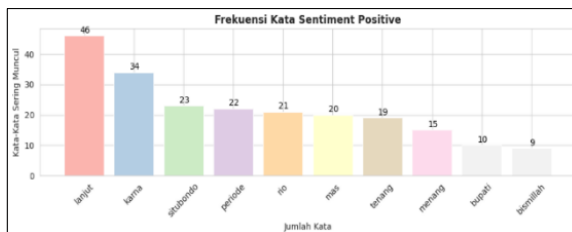
Gambar 8. Diagram Batang Klasifikasi Sentimen.

Berikut merupakan hasil wordcloud dari klasifikasi sentimen positif ditunjukkan pada Gambar 9.



Gambar 9. Wordcloud Sentimen Positif

Frekuensi kata merupakan frekuensi kata dari sentimen positif ditunjukkan pada Gambar 10.

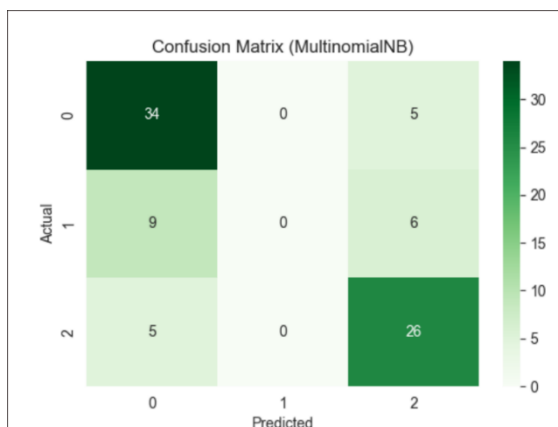


Gambar 10. Frekuensi Kata Sentimen Positif

Hasil wordcloud dari klasifikasi sentimen negatif ditunjukkan pada Gambar 11. di bawah ini.

3.7. Evaluasi Model

Tahap berikutnya adalah mengevaluasi kinerja model *Multinomial Naive Bayes* setelah dilatih dan menghasilkan prediksi. Untuk menentukan distribusi prediksi akurat dan tidak akurat di antara kelas-kelas, evaluasi dilakukan dengan menghitung *confusion matrix* yang ditunjukkan pada Gambar 17.



Gambar 17. *Confusion matrix*

Selain itu, matrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score* diperoleh dari setiap kelas menggunakan fungsi *classification report*. Dengan memanfaatkan fungsi *accuracy_score* digunakan untuk menentukan akurasi model dan menampilkan persentase total prediksi model yang sesuai dengan label aktual ditunjukkan pada Gambar 18.

Classification Report (MultinomialNB):				
	precision	recall	f1-score	support
negative	0.71	0.87	0.78	39
netral	0.00	0.00	0.00	15
positive	0.70	0.84	0.76	31
accuracy			0.71	85
macro avg	0.47	0.57	0.52	85
weighted avg	0.58	0.71	0.64	85

Accuracy (MultinomialNB): 0.7059

Gambar 18. Hasil *Classification Report*

Tahap selanjutnya merupakan perhitungan manual untuk menentukan akurasi model klasifikasi sentimen. Jumlah prediksi yang akurat dibandingkan dengan jumlah total data testing untuk menentukan akurasi. Pada hasil *confusion matrix* menunjukkan jumlah prediksi yang benar negatif 34, netral 0, dan positif 26 dengan total 60 data yang dikategorikan dengan akurat. Ada 85 data uji secara keseluruhan. Berikut merupakan hasil perhitungan manual untuk menentukan hasil akurasi model menggunakan Persamaan (2).

$$\text{akurasi} = \frac{34 + 0 + 26}{34 + 0 + 5 + 9 + 5 + 0 + 0 + 6 + 26} \times 100\%$$

$$\text{akurasi} = \frac{60}{85} \times 100\% = 0,7056 \text{ atau } 70,59\%$$

Berikut merupakan perhitungan manual *Classification Report* seperti *precision*, *recall*, dan *F1-score* pada tiap kelas.

1. Negatif

Pada hasil *confusion matrix* negatif menunjukkan jumlah TP untuk negatif=34, sedangkan FP untuk negatif 9+5=14. Oleh karena itu, persamaan (3) digunakan untuk menentukan *precision* model, menghasilkan nilai *precision* 0.71 atau sekitar 70.83%.

$$\text{Precision} = \frac{34}{34 + 14} = 0,7083$$

Selanjutnya menentukan *recall* dengan FN untuk negatif=5. Oleh karena itu, persamaan (4) digunakan untuk menentukan *recall* model, menghasilkan nilai *recall* 0.87 atau 87.18%.

$$\text{Recall} = \frac{34}{34 + 5} = 0.8718$$

Selanjutnya menentukan *F1-score* dengan menghitung hasil *precision* dan *recall* persamaan (5) digunakan untuk menentukan *F1-score* model, menghasilkan nilai *F1-score* 0.7816 atau 78.16%.

$$F1 = 2 \times \frac{0.7083 \times 0.8718}{0.7083 + 0.8718}$$

$$F1 = 2 \times \frac{0.6176}{1.5801} = 2 \times 0.3908 = 0.7816$$

2. Netral

Pada hasil *confusion matrix* netral menunjukkan jumlah TP untuk netral=0, sedangkan FP untuk netral=0. Oleh karena itu, rumus (3) digunakan untuk menentukan *precision* model, menghasilkan nilai *precision* 0.00 atau sekitar 0.00%.

$$\text{Precision} = \frac{0}{0 + 0} = 0.00$$

Selanjutnya menentukan *recall* dengan FN untuk negatif 9+6=15. Oleh karena itu, persamaan (5) digunakan untuk menentukan *recall* model, menghasilkan nilai *recall* 0.00 atau 0.00%.

$$\text{Recall} = \frac{0}{0 + 15} = 0.00$$

Selanjutnya menentukan *F1-score* dengan menghitung hasil *precision* dan *recall*. Persamaan (4) digunakan untuk menentukan *F1-score* model, menghasilkan nilai *F1-score* 0.00 atau 0.00%.

$$F1 - \text{score} = 2 \times \frac{0}{0} = 2 \times 0 = 0.00$$

3. Positif

Pada hasil *confusion matrix* positif menunjukkan jumlah TP untuk positif=26, sedangkan FP untuk positif 6+5=11. Oleh karena itu, persamaan (3) digunakan untuk menentukan *precision* model, menghasilkan nilai *precision* 0.70 atau sekitar 70.27%.

$$Precision = \frac{26}{26 + 11} = 0,7027$$

Selanjutnya menentukan *recall* dengan FN untuk negatif = 5. Oleh karena itu, persamaan (4) digunakan untuk menentukan *recall* model, menghasilkan nilai *recall* 0.84 atau 83.87%.

$$Recall = \frac{26}{26 + 5} = 0,8387$$

Selanjutnya menentukan *F1-score* dengan menghitung hasil *precision* dan *recall* rumus (5) digunakan untuk menentukan *F1-score* model, menghasilkan nilai *F1-score* 0.7647 atau 76.47 %.

$$Recall = 2x \frac{0,7027 \times 0,8387}{0,7027 \times 0,8387}$$

$$Recall = 2x \frac{0,5894}{1,5414} = 2 \times 0,3824 = 0,7647$$

Hasil penelitian menunjukkan Metode *Naive Bayes Classifier* efektif dalam melakukan analisis sentimen pada PILKADA di Kabupaten Situbondo tahun 2024 di platform TikTok. Data yang diperoleh melalui proses *scraping* menggunakan *Data Miner* berjumlah 494 komentar, yang kemudian data di *Preprocessing* meliputi *Cleansing*, *Case Folding*, *Tokenizing*, *Remove Stopword*, *Stemming Data* dan dikategorikan ke dalam tiga kategori sentimen *positive*, *negative*, dan *neutral*. Teknik *Multinomial Naive Bayes* digunakan untuk membuat model, dengan *split data* menjadi data *training* sebesar 79,95% dan *testing* 20,05%. Hasil evaluasi menunjukkan nilai akurasi model mencapai 70,59%, dan menghasilkan nilai presisi, *recall*, dan *F1-score* yang berbeda-beda tiap kategori sentiment.

Pada penelitian ini terdapat peningkatan nilai akurasi dari hasil penelitian terdahulu (Cindy et al., 2024) dengan rata-rata nilai akurasi lebih dari 60%, yaitu nilai akurasi pada akun Najwa sebesar 70,38%, pada akun Kompasiana.com sebesar 66,95%, dan akun Detik.com sebesar 66,69%. Hasil penelitian ini juga menunjukkan peningkatan yang signifikan dari hasil penelitian sebelumnya (Mario et al., 2025) dengan hasil akurasi 61%, *precision* 71%, *recall* 61%, dan *F1-score* 54% dengan analisis dan mengidentifikasi tiga isu dominan, yaitu dukungan pembangunan IKN, kritik terhadap anggaran, dan diskusi netral terkait proses konstruksi pada sentiment positif, negatif, dan netral.

4. Kesimpulan

Berdasarkan hasil eksperimen dan evaluasi model, hasil akurasi masih dapat ditingkatkan tidak hanya menggunakan metode *Naive Bayes*, diharapkan bisa membandingkan berbagai teknik klasifikasi lainnya termasuk *Random Forest*, *Support Vector Machine* (SVM), atau algoritma berbasis *deep learning*. Untuk membuat hasil penelitian lebih mencerminkan opini publik, juga dimungkinkan untuk memperluas jumlah dan variasi data dari platform media sosial lainnya. Disarankan pula untuk menerapkan strategi optimasi atau teknik *feature engineering* yang lebih canggih untuk

meningkatkan kemampuan model dalam mengenali sentimen, terutama pada kategori netral yang seringkali sulit dikategorikan. Sehingga para kandidat, staf kampanye, dan penyelenggara PILKADA dapat memanfaatkan temuan penelitian ini untuk mengidentifikasi isu-isu yang muncul, merancang strategi kampanye dalam segi komunikasi politik yang lebih tepat sasaran, dan memperoleh pemahaman yang lebih terukur tentang opini publik.

Daftar Pustaka:

- Apriliani, N., Suarna, N., Prihartono, W., Informatika, T., Akuntansi, K., Cirebon, K., Sentimen, A., & Bayes, A. N. (2023). *ANALISIS SENTIMEN REVIEW PENGGUNAAN TIKTOK MELALUI PENDEKATAN ALGORITMA NAIVE BAYES*. 7(6), 3725–3731.
- Apriyani, M. E., Renaldi, R., Cinderatama, T. A., Informasi, J. T., Malang, P. N., Test, U. A., Penyelenggara, A., & Internet, J. (2024). *ANALISIS SENTIMEN BERITA HOAX MENGGUNAKAN NAIVE*. 1–6.
- Arifqi, T., Suarna, N., & Prihartono, W. (2024). Penggunaan Naive Bayes Dalam Menganalisis Sentimen Ulasan Aplikasi Mcdonald'S Di Indonesia. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 1949–1956. <https://doi.org/10.36040/jati.v8i2.8740>
- Bhilton M. Obidje, M. A. I. P. (2025). *Analisis sentimen pemilihan presiden dan wakil presiden tahun 2024 di twitter menggunakan metode klasifikasi*. 10(1), 424–433.
- Cahya Kamilla, A., Priyani, N., Priskila, R., & Handrianus Pranatawijaya, V. (2024). Analisis Sentimen Film Agak Laen Dengan Kecerdasan Buatan: Text Mining Metode Naive Bayes Classifier. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 2923–2928. <https://doi.org/10.36040/jati.v8i3.9587>
- Cindy, T., Rani, S., & Sahputra, E. (2024). *Jurnal Computer Science and Information Technology (CoSciTech) bayes*. 5(3), 720–729.
- Heniasari, R. D., Aliftania, D., Adzania, K. H., & Setiawan, D. (2025). *Partisipasi Pemilih Pemula Pada Pemilihan Kepala Daerah di Kota Magelang*. 3.
- Hidayah, A. K., Erwadi, Y., Handayani, S., Informatika, T., Informasi, S., Negara, I. K., & Forest, R. (2025). *IBU KOTA NEGARA DI TWITTER MENGGUNAKAN*. 9(5), 9109–9114.
- Hidayatullah, S., Waskita, A. A., Hindasyah, A., S-, T. I., Pamulang, U., & Selatan, T. (2026). *Analisis Sentimen Pengguna X Terhadap Pemilihan Gubernur Dki Jakarta Tahun 2024 Dengan Algoritma Naive Bayes , K- Nearest Neighbor Dan Decision Tree Pastikan Anda*

- menggunakan style*. 10(1), 203–214.
- Khusnul Khotimah, A. (2024). Analisis Sentimen Terhadap Kualitas Pelayanan. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 3044–3048. <https://doi.org/10.36040/jati.v8i3.9520>
- Mario, G., Paridy, C., Aristo, A., Sinlae, J., & Ngaga, E. (2025). *Analisis Sentimen di Media Sosial X tentang IKN dengan Naïve Bayes*. 417–426.
- Mursyid, R., Indriyanti, A. D., Teknik, J., Fakultas, I., Universitas, T., & Surabaya, N. (2024). *Perbandingan Akurasi Metode Analisis Sentimen Untuk Evaluasi Opini Pengguna Pada Platform Media Sosial (Studi Kasus : Twitter)*. 06, 371–383.
- Nugroho, D. S., Hanif, I. F., & Hasbi, M. A. (2024). *Sentiment Analysis of Alleged 2024 Election Fraud Based on Tweets Using the Naïve Bayes Classifier Algorithm Analisis Sentimen Dugaan Pelanggaran Pemilu 2024 Berdasarkan Tweet Menggunakan Algoritma Naïve Bayes Classifier*. 4(July), 1169–1176.
- Pratama, P. W., Pamungkas, E. W., Informatika, T., & Surakarta, U. M. (2026). *Analisis Sentimen Terhadap Komentar Pada Video Viral (Fyp) Tiktok Menggunakan Metode Naïve Bayes*. 9, 35–46.
- Putri Nurrahmah Wear, Idris, N., & Rosita, I. (2024). *ANALISIS SENTIMEN TERHADAP HASIL PILPRES 2024 PADA APLIKASI TIKTOK DAN X MENGGUNAKAN METODE NAÏVE*. 1.
- Salsabilla, F. N., Witanti, A., Informasi, F. T., Mercu, U., Yogyakarta, B., Bantul, K., & Yogyakarta, D. I. (2025). *PRESIDEN JOKOWI PADA MEDIA SOSIAL X*. 8, 106–115.
- Siddiq, M. J., Jayasri, S., Suhendi, A., Hidayat, T., Rizky, R., Islam, U., Yusuf, S., Raya, J., & Cikaliung, L. (2025). *MASYARAKAT TERHADAP PILKADA 2024 DI MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAÏVE*. 13(2), 609–622.
- Simanullang, A., Sosial, F. I., Medan, N., Medan, K., & Utara, S. (2025). *Peran Media Sosial Tiktok dalam Transformasi Politik Gen Z Pada Pemilihan Presiden Tahun 2024 di Kelurahan Tegal Rejo*. 6, 432–450. <https://doi.org/10.46730/japs.v6i3.327>
- Sri Yani Kusumastuti, Faisal, A., & Rahayu, D. H. (2024). *No Title*.
- Sulistiowati, Y., & Santoso, B. J. (2025). *ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI MOBILE SP4N-LAPOR ! DENGAN PENDEKATAN MACHINE LEARNING*. 283–290.
- Wijaya, A., Rivaldo, M., & Pribadi, M. R. (2024). *AICOMS Applied Information Technology and Computer Science Analisis Sentimen Terhadap Aplikasi Mitra Darat Menggunakan Algoritma Naive Bayes Classifier dan K-Nearest Neighbor*. 3(1), 9–14.