



Analisis Sentimen Pengguna X Terhadap Pemasangan Alat Peraga Kampanye (Apk) Pemilu 2024 Menggunakan Algoritma Naïve Bayes

Adhi Jadmiko¹, Sri Lestanti², Sabitul Kirom³

Fakultas Teknologi Informasi, Universitas Islam Balitar^{1,2,3}

		Abstract
Received:	20 April 2025	Pemilihan umum tahun 2024 di Indonesia merupakan suatu agenda politik sebagai pemilihan calon kepemimpinan selanjutnya. Pemilihan calon presiden atau wakil presiden serta calon legislatif lainnya tentunya memakai alat peraga kampanye sebagai bentuk komunikasi dalam mengenalkan visi, misi, serta citra para calon pemimpin. Pada pemasangan APK tersebut menimbulkan berbagai sentimen masyarakat yang dituangkan pada tweets di platform X . Hal ini dipilih karena kehandalannya dalam klasifikasi teks pendek seperti tweets, serta kemampuannya dalam menangani analisis sentimen. Sumber data yang digunakan melalui <i>crawling</i> dari media sosial X berupa teks atau komentar. Teknik pengumpulan data dilakukan dengan cara pengamatan secara langsung terhadap topik permasalahan yaitu mencari hashtag #alatperagakampanye dan #apkpemilu2024. Hasil penelitian pada sentimen terhadap pemasangan APK pada pemilu 2024 pada proses labeling data menggunakan kamus <i>Lexicon Based</i> yang telah melewati tahap <i>text preprocessing</i> menunjukkan hasil sentimen positif sebesar 36,3%, hasil sentimen negatif sebesar 51,2 %, dan hasil sentimen netral hanya sebesar 12,3%. Selanjutnya proses pembobotan kata dengan TF-IDF serta tahap akhir dilakukan klasifikasi dengan algoritma naïve bayes yang menghasilkan nilai akurasi model sebesar 58% dari ketiga kategori tersebut.
Revised:	27 April 2025	
Accepted:	01 Mei 2025	
Keywords:		Analisis Sentimen, <i>Naïve Bayes</i> , Alat Peraga Kampanye, Pemilu Indonesia 2024
(*) Corresponding Author:		jadmikoa7@gmail.com , lestanti85@gmail.com , sabitulkirom@gmail.com
<p>How to Cite: Jadmiko, A., Lestanti, S., & Kirom, S. (2025). Analisis Sentimen Pengguna X Terhadap Pemasangan Alat Peraga Kampanye (Apk) Pemilu 2024 Menggunakan Algoritma Naïve Bayes. <i>Jurnal Ilmiah Wahana Pendidikan</i>, 11(5.B), 24-37. Retrieved from https://jurnal.peneliti.net/index.php/JIWP/article/view/11911.</p>		

PENDAHULUAN

Pemilihan umum di Indonesia merupakan salah satu momen penting dalam agenda politik negara. Sebagai negara demokratis, pemilihan umum menjadi ajang bagi masyarakat untuk berpartisipasi secara langsung dalam menentukan arah kepemimpinan negara. Dalam pemilu 2024 merupakan agenda pemilihan yang meliputi pilihan presiden dan wakil presiden, pilihan calon legislatif yang meliputi anggota DPR, DPD, serta DPRD tingkat provinsi maupun tingkat kabupaten maupun kota.

Kampanye politik menjadi salah satu aspek krusial dalam memperkenalkan visi, program, dan figur yang akan memimpin. Kegiatan ini merupakan salah satu bentuk komunikasi terencana guna mempengaruhi banyak khalayak untuk mencapai tujuan tertentu (Anam, 2018). Dalam konteks ini, pemasangan alat peraga kampanye telah menjadi suatu agenda wajib bagi para calon pemimpin sebagai bentuk strategi yang paling terlihat secara visual di ruang publik (Agustina 2020).

Dalam beberapa tahun terakhir, Indonesia telah menyaksikan perubahan dramatis dalam perilaku politik dan pola komunikasi masyarakat. Revolusi digital telah merubah lanskap politik secara signifikan, mempengaruhi cara politisi berinteraksi dengan pemilih dan cara pemilih mengakses informasi politik. Salah satunya yaitu melalui media sosial yang tidak hanya menjadi wadah untuk mengakses informasi, namun juga sebagai wadah untuk berbagi informasi serta bertukar pendapat. Kajian terbaru dari Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) menunjukkan bahwa pengguna internet Indonesia akan meningkat setiap tahun sebanyak 5,44%, mencapai sekitar 213 juta pengguna pada tahun 2023 (Bastoni, 2024). Hal ini menunjukkan bahwa media sosial sangat berperan andil dalam penyampaian dan penerimaan informasi secara masif.

Termasuk platform seperti X (yang sebelumnya bernama Twitter) yang telah menjadi kanal utama di mana diskusi politik, pertukaran opini, dan aktivitas kampanye berlangsung (Arianto 2021). X sebagai salah satu situs web media sosial terkenal di Indonesia yang menyediakan ruang bagi warganet untuk berdiskusi tentang berbagai isu termasuk politik.

Terlepas dari penggunaan media sosial tersebut, platform X telah mengalami sejarah pergantian logo dari 2005 hingga 2023. Dilansir dari laman JawaPos.com bahwa logo Twitter yang identik dengan burung kolibri berwarna biru mengalami pergantian logo X dengan background hitam pada tahun 2023. Pergantian logo tersebut juga diiringi dengan bergantinya pemilik platform tersebut. Elon Musk sebagai pemilik akun X berhasil mengganti domain serta profil Twitter menjadi @X dengan adanya tujuan, yakni rebranding untuk perbaruan citra brand serta strategi bisnis pemasaran (Asy'Ari, 2023).

Pergantian logo dari X tidak serta merta menyurutkan pengguna X (yang sebelumnya bernama Twitter) setelah terjadinya penyesuaian terhadap para penggunanya karena terdapat kesamaan dari internal platform. Berdasarkan data yang dilansir oleh statista.com, pengguna X mencapai 187 juta pengguna aktif harian dengan jumlah pengguna sebanyak 15,1 juta per April 2021. Menurut Tankovka (2021) Indonesia menduduki peringkat ke-6 terbanyak berdasarkan pengguna X (Wahyudi, 2021). Dengan jutaan pengguna aktif tersebut, Twitter telah menjadi tempat di mana masyarakat berbagi pandangan, menyampaikan aspirasi, dan mengkritik berbagai isu di sekitarnya. Mulai dari isu sosial, pendidikan, ekonomi sampai politik. Selain itu, platform X termasuk salah satu platform yang dapat menyebarkan informasi secara cepat, sehingga suatu topik penting dan populer serta tren masa kini juga dapat diakses pada X.

Platform X ini juga telah menjadi salah satu kebutuhan pengguna dalam menyebarkan pendapat serta informasi dalam segala hal. Terutama pada saat berlangsungnya kampanye pemilu 2024 yang menjadi topik pembicaraan yang sangat populer di Indonesia. Salah satunya yaitu isu politik mengenai pemasangan alat peraga kampanye (APK). Hal ini dibuktikan melalui observasi pada platform X bahwa hashtag #alatperagakampanye #apkpemilu2024 memunculkan banyak statement dengan komentar positif, negatif, maupun komentar netral terhadap pemasangan APK tersebut. Situasi ini didukung dengan realita bahwa banyak APK yang dipasang dengan melanggar ketentuan. Seperti halnya, pemasangan APK tidak sesuai tempatnya, mengotori estetika kota, hingga merugikan pihak lain, sehingga berdampak terhadap aktivitas masyarakat.

Dalam menganalisa permasalahan ini, peneliti memilih platform X sebagai pengambilan data. Selain itu, X memuat banyak cuitan dengan sentimen yang beragam, sehingga memudahkan peneliti untuk menganalisis opini masyarakat secara efektif. Metode analisis sentimen menggunakan algoritma *Naïve Bayes* telah terbukti menjadi

alat yang efektif dalam memahami opini dan perasaan yang terkandung dalam teks. Algoritma ini dapat memberikan kategori dalam teks menjadi beberapa sentimen meliputi sentimen positif, negatif, atau netral yang memiliki tingkat akurasi baik.

Penelitian ini dapat memberikan informasi untuk mengetahui tingkat presentase komentar atau tweet dari platform X terhadap APK. Dimana dengan persentase dan respon tersebut bisa menjadi masukan dan evaluasi kepada kandidat politik untuk mengetahui bagaimana publik memandang calon kandidat serta kampanye nya. Selain itu, masyarakat umum bisa mendapatkan informasi tentang bagaimana orang lain memandang berbagai isu politik.

Meskipun banyak penelitian telah dilakukan tentang analisis sentimen di media sosial, masih sedikit penelitian yang secara khusus fokus pada analisis sentimen terhadap pemasangan alat peraga kampanye politik di Indonesia, terutama menggunakan data dari X. Oleh karena itu, berdasarkan latar belakang yang telah dipaparkan, peneliti mengangkat penelitian yang berjudul “Analisis Sentimen Pengguna X Terhadap Pemasangan Alat Peraga Kampanye (APK) Pemilu 2024 menggunakan Algoritma *Naïve Bayes*”.

METODE PENELITIAN

Waktu dan lokasi penelitian dengan judul Analisis Sentimen Pengguna X Terhadap Pemasangan Alat Peraga Kampanye (APK) Pemilu 2024 menggunakan Algoritma *Naïve Bayes* yang akan dilakukan pada bulan November 2023 sampai bulan Juli 2024. Untuk lokasi penelitian ini dilakukan di Universitas Islam Balitar Blitar (UNISBA) beralamat di Jalan Majapahit No. 04, Sananwetan, Kecamatan Sananwetan, Kota Blitar, Jawa Timur. Metode pengumpulan data dalam penelitian ini yaitu pengamatan langsung, menggunakan *API X* dan *Crawling*.

HASIL DAN PEMBAHASAN

1. Hasil Pengumpulan Data

Dalam penelitian analisis sentimen terhadap pemasangan alat peraga kampanye dalam pemilu 2024 ini telah mengumpulkan data berjumlah 2240 baris yang merupakan data mentah. Data diperoleh dari *web scraping* dengan menggunakan *playwright* dan *node.js*. *Node.js* digunakan untuk menjalankan *tweet-harvest* yang merupakan sebuah perangkat alat lunak untuk mengumpulkan data dari X. Kemudian untuk mendapatkan akses ke sumber data maka akan dilakukan input autentikasi berupa auth token yang dimiliki oleh peneliti. Selanjutnya akan dilakukan *crawling* data dengan memasukkan kata kunci untuk mendapatkan data.

Data yang disimpan memiliki beberapa kolom pada data tabel tersebut yaitu kolom *conversation_id_str*, *created_at*, *favorite_count*, *id_str*, *image_url*, *in_reply_to_screen_name*, *lang*, *location*, *quote_count*, *reply_count*, *retweet_count*, *tweet_url*, *user_id_str*, *username*, dan *text*. Dari kelima belas kolom tersebut akan dilakukan penghapusan dan menyisakan kolom *text* saja. Hal ini dikarenakan kolom lainnya tidak relevan dengan penelitian ini. Dengan kata lain penelitian ini hanya berfokus pada pendapat atau komentar dari pengguna X terhadap pemasangan alat peraga kampanye pemilu 2024. Hasil pengumpulan dari (*crawling*) dapat dilihat pada tabel 1.

Tabel 1 Hasil Pengumpulan Data

No	Text
----	------

1	@pandji Sumpah bang, tolong berhenti. Setiap lihat baliho caleg selalu keinget wajah abang. 🙄
2	Baliho Arif Fathoni Bertebaran di Kota Surabaya Serius Adu Elektabilitas Dengan Eri Cahyadi #partaigolkar #golkarindonesia #golkarpedia #airlanggahartarto #golkaronsosmed https://t.co/4ZfWeC6DxG https://t.co/3TbuthbrvG
3	we info percetakan spanduk yg cepat n harga terjangkau
4	@CNNIndonesia Baliho caleg no brp nih ?
5	@DedynurPalakka @psi_id Bukan dengan dana besar ??? Spanduk baliho dimana-mana laporan nya 0 https://t.co/3epcEEMe8k

2. Hasil Text Preprocessing

Pada tahap *text preprocessing* tahap pertama meng install modul pada *python*. Data yang telah diperoleh melalui *crawling* kemudian dilakukan proses *preprocessing* tahapanya adalah *text cleaning*, *case folding*, normalisasi, *tokenziing*, *stopword removal* dan *steaming*. Berikut diuraikan tahap-tahap *preprocessing*.

3. Hasil Text Cleaning

Tahap *text cleaning* dilakukan dengan proses menghapus tag HTML, tag username, trim text, penghapusan tanda baca, hastag, url, username, dan menghapus spasi ganda. Hasil *text cleaning* dapat dilihat pada tabel 2.

Tabel 2 Hasil *Text Cleaning*

No	Sebelum <i>text cleaning</i>	Setelah <i>text cleaning</i>
1	@pandji Sumpah bang, tolong berhenti. Setiap lihat baliho caleg selalu keinget wajah abang. 🙄	Sumpah bang tolong berhenti setiap lihat baliho caleg selalu keinget wajah abang
2	Baliho Arif Fathoni Bertebaran di Kota Surabaya Serius Adu Elektabilitas Dengan Eri Cahyadi #partaigolkar #golkarindonesia #golkarpedia #airlanggahartarto #golkaronsosmed https://t.co/4ZfWeC6DxG https://t.co/3TbuthbrvG	Baliho Arif Fathoni Bertebaran di Kota Surabaya Serius Adu Elektabilitas Dengan Eri Cahyadi
3	we info percetakan spanduk yg cepat n harga terjangkau	we info percetakan spanduk yg cepat harga terjangkau
4	@CNNIndonesia Baliho caleg no brp nih ?	Baliho caleg no brp nih
5	@DedynurPalakka @psi_id Bukan dengan dana besar ??? Spanduk baliho dimana-mana laporan nya 0 https://t.co/3epcEEMe8k	Bukan dengan dana besar Spanduk baliho dimana mana laporan nya

Tahap selanjutnya akan dilakukan penghapusan *missing values*. Dalam data dapat diketahui bahwa terdapat 0 baris *missing values* yang dapat dilihat pada gambar 1.

```

Unnamed: 0      0
text            0
Gambar 1      processed_text      0  'alues
dtype: int64
    
```

Kemudian akan dilakukan penghapusan terhadap duplikasi baris pada data. Pada data terdapat 28 baris yang terduplikasi. Jumlah duplikasi data dapat dilihat pada gambar 2 dan gambar 3.

↔ Jumlah baris terduplikasi: 28

Unnamed: 0	text	processed_text
719	Solidaritas Pemuda Bali-Lombok melakukan dekla...	solidaritas pemuda bali lombok melakukan dekla...
921	Kegiatan Pemasangan balihobanner dalam rangka ...	kegiatan pemasangan balihobanner dalam rangka ...
1008	Muncul Foto Wajah Marshal Widiyanto di Baliho B...	muncul foto wajah marshal widiyanto di baliho b...
1119	Kegiatan Pemasangan balihobanner dalam rangka ...	kegiatan pemasangan balihobanner dalam rangka ...
1185	Balihan Dukungan Ahmad Luthifi Maju Pilgub Jat...	balihan dukungan ahmad luthifi maju pilgub jat...
1229	Balihan Arif Fathoni Bertebaran di Kota Surabay...	balihan arif fathoni bertebaran di kota surabay...
1232	UPP Kab Bantaeng melaksanakan giat silaturahmi...	upp kab bantaeng melaksanakan giat silaturahmi...
1243	Balihan Calon Peserta Pilkada yang Menyalahi Pe...	balihan calon peserta pilkada yang menyalahi pe...
1352	Jokowi Bagi-bagi Bansos di Dekat Spanduk Prabo...	jokowi bagi bagi bansos di dekat spanduk prabo...
1423	Video : Tentrem Semarang Punya Videotron 3D ht...	video tentrem semarang punya videotron
1522	Kegiatan Pemasangan balihobanner dalam rangka k...	kegiatan pemasangan balihobanner dalam rangka k...
1558	Pemasangan spanduk balihobanner mengajak masyarakat ...	pemasangan spanduk balihobanner mengajak masyarakat ...
1622	Desain Balihobanner https://t.co/zzU9aTXi6P	desain balihobanner

Gambar 3 Contoh Sampel Duplikasi Data

Data yang terduplikasi akan dihapus dan menyisakan 2212 baris data. Kemudian data akan diolah ke tahap selanjutnya.

4. Hasil Case Folding

Tahap *case folding* dilakukan untuk menyamakan karakter huruf atau data. Proses merubah semua karakter huruf menjadi huruf kecil. Contoh hasil *case folding* seperti pada tabel 3.

Tabel 3 Hasil Case Folding

No	Sebelum Case Folding	Setelah Case Folding
1	Sumpah bang tolong berhenti setiap lihat balihobanner selalu keinget wajah abang	sumpah bang tolong berhenti setiap lihat balihobanner selalu keinget wajah abang
2	Balihan Arif Fathoni Bertebaran di Kota Surabaya Serius Adu Elektabilitas Dengan Eri Cahyadi	balihan arif fathoni bertebaran di kota surabaya serius adu elektabilitas dengan eri cahyadi
3	we info percetakan spanduk yg cepat harga terjangkau	we info percetakan spanduk yg cepat harga terjangkau
4	Balihan caleg no brp nih	balihan caleg no brp nih
5	Bukan dengan dana besar Spanduk balihobanner dimana mana laporannya	bukan dengan dana besar spanduk balihobanner dimana mana laporannya

5. Hasil Normalisasi

Setelah proses *case folding* akan dilanjutkan dengan normalisasi singkatan. Normalisasi singkatan ini menggunakan database yang dibuat oleh peneliti sendiri. Tabel 4 merupakan beberapa contoh normalisasi singkatan yang dibuat oleh peneliti. Dan tabel 5 hasil teks yang telah dinormalisasi.

Tabel 4 Normalisasi Singkatan

Kata awal	Setelah dinormalisasi	Kata awal	Setelah dinormalisasi
Kyk	Kaya	Tdk	Tidak
Btw	By the way	Jd	Jadi

Yg	Yang	Brp	Berapa
Lg	Lagi	Spt	Seperti
Km	Kamu	Blm	Belum

Tabel 5 Hasil Normalisasi Singkatan

Text Normalisasi
sumpah bang tolong berhenti setiap lihat baliho caleg selalu keinget wajah abang
baliho arif fathoni bertebaran di kota surabaya serius adu elektabilitas dengan eri cahyadi
we info percetakan spanduk yang cepat harga terjangkau
baliho caleg no berapa nih
bukan dengan dana besar spanduk baliho dimana mana laporan nya co epeeme

6. Hasil *Tokenization*

Tahap *tokenization* dilakukan untuk memisahkan teks menjadi token atau kata. Hasil dari proses ini dapat ditunjukkan pada tabel 6. Pada tahap ini akan ditambahkan kolom baru berupa `text_token` yang berisi token dari teks pada data.

Tabel 6 Hasil *Tokenizing*

No	Sebelum <i>Tokenizing</i>	Setelah <i>Tokenizing</i>
1	sumpah bang tolong berhenti setiap lihat baliho caleg selalu keinget wajah abang	sumpah, bang, tolong, berhenti, setiap, lihat, baliho, caleg, selalu, keinget, wajah, abang
2	baliho arif fathoni bertebaran di kota surabaya serius adu elektabilitas dengan eri cahyadi	baliho, arif, fathoni, bertebaran, di, kota, surabaya, serius adu, elektabilitas, dengan, eri, cahyadi
3	we info percetakan spanduk yang cepat harga terjangkau	we, info, percetakan, spanduk, yang, cepat, harga, terjangkau
4	baliho caleg no berapa nih	baliho, caleg, no, berapa, nih
5	bukan dengan dana besar spanduk baliho dimana mana laporan nya	bukan, dengan, dana, besar, spanduk, baliho, dimana, mana, laporan, nya

7. Hasil *Stopword Removal*

Tahap selanjutnya adalah *stopword removal*. Tahap ini akan dilakukan penghapusan kata atau token yang sering muncul dan tidak memiliki makna penting dalam teks. Stopword akan dihapus dari teks untuk mengurangi dimensi dan meningkatkan efisiensi analisis sentimen. Pada penelitian ini stopwords yang digunakan adalah stopwords bahasa Indonesia. Terdapat beberapa penambahan kata pada stopwords seperti di, ada, ya, yang, sama, itu, saja, dan sebagainya. Hasil dari *stopword removal* dapat dilihat pada tabel 7.

Tabel 7 Hasil *Stopword Removal*

No	Sebelum <i>Stopword Removal</i>	Setelah <i>Stopword Removal</i>
1	sumpah bang tolong berhenti setiap lihat baliho caleg selalu keinget wajah abang	sumpah, bang, tolong, berhenti, lihat, baliho, caleg, keinget, wajah, abang

2	baliho arif fathoni bertebaran di kota surabaya serius adu elektabilitas dengan eri cahyadi	baliho, arif, fathoni, bertebaran, kota, surabaya, serius, adu, elektabilitas, eri, cahyadi
3	we info percetakan spanduk yang cepat harga terjangkau	we, info, percetakan, spanduk, cepat, harga, terjangkau
4	baliho caleg no berapa nih	baliho, caleg, no, ini
5	bukan dengan dana besar spanduk baliho dimana mana laporan nya	dana, spanduk, baliho, dimana, laporan

8. Hasil Stemming

Setelah itu akan dilakukan tahap *stemming* dimana akan mengubah kata dalam bentuk dasarnya saja. *Stemming* dilakukan pada kolom `text_token` akan dilakukan *stemming* sehingga menghasilkan kata-kata dalam bentuk dasarnya saja. Hasil dari *stemming* dapat dilihat pada tabel 8.

Tabel 8 Hasil *Steaming*

No	Sebelum <i>Steaming</i>	Setelah <i>Steaming</i>
1	sumpah bang tolong berhenti setiap lihat baliho caleg selalu keinget wajah	sumpah, bang, tolong, henti, lihat, baliho, caleg, keinget, wajah, abang
2	baliho arif fathoni bertebaran di kota surabaya serius adu elektabilitas dengan eri cahyadi	baliho, arif, fathoni, tebar, kota, surabaya, serius, adu, elektabilitas, eri, cahyadi
3	we info percetakan spanduk yang harga terjangkau cepat	we, info, cetak, spanduk, cepat, harga, terjangkau
4	baliho caleg no berapa nih	baliho, caleg, no, nih
5	bukan dengan dana besar spanduk baliho dimana mana laporan	dana, spanduk, baliho, dimana, laporan

9. Hasil Pelabelan Data dengan *Lexicon Based*

Setelah melakukan tahap preprocessing data akan dilabeli dengan menggunakan kamus *Lexicon Based*. Pada kamus kamus *Lexicon* yang digunakan merupakan kamus inset terdiri dari 3609 kata positif dan 6609 kata negatif. Setiap kata yang terdapat pada kamus *lexicon* memiliki nilai bobot dari -5 hingga +5. Data akan diberi label sebagai sentimen positif, netral, atau negatif sesuai dengan skor yang diperoleh. Berikut merupakan contoh daftar kata yang memiliki bobot dan dikategorikan sebagai sentimen negatif dan sentimen positif. Contoh daftar kata pada kamus *Lexicon Based* dapat dilihat pada tabel 9.

Tabel 9 Contoh Daftar Kata Pada Kamus *Lexicon Based*

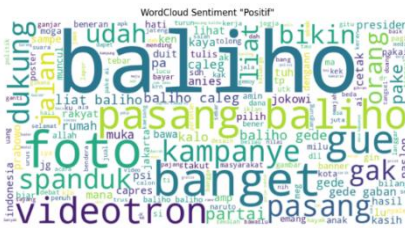
No.	Sentimen positif		Sentimen negative	
	Kata	Bobot	Kata	Bobot
1	Merekam	2	Pemarah	-4
2	Paripurna	1	Nyata	-2
3	Perubahan	4	Bersekutu	-5
4	Coba	2	Sia-Sia	-4
5	Prihatin	2	Mati	-5

6	Pesat	3	Hati-Hati	-3
7	Tinggal	1	Merebut	-3
8	Terjamin	5	Kalau	-1
9	Pilihan	5	Cabut	-5

Daftar kata di kamus *Lexicon Based* yang memiliki nilai positif dan negatif di gabungkan menjadi satu data. Kemudian data hasil *text preprocessing* yang berjumlah 2212 di input untuk dilakukan pelabelan sentimen. Data akan dilakukan pembobotan untuk menentukan sentimen dari setiap data yang ada. Kemudian membuat variabel label sentimen untuk menentukan data positif, negatif atau netral. Tabel 10 merupakan hasil pelabelan sentimen menggunakan kamus *lexicon based*.

Tabel 10 Hasil Pelabelan Sentimen

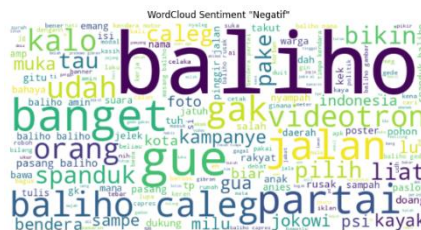
No	Text	Sentimen
1	sumpah bang tolong berhenti setiap lihat baliho caleg selalu keinget wajah abang	Negatif
2	baliho arif fathoni bertebaran di kota surabaya serius adu elektabilitas dengan eri cahyadi	Negatif
3	we info percetakan spanduk yang harga terjangkau cepat	Positif
4	Baliho caleg no berapa nih	Netral
5	bukan dengan dana besar spanduk baliho dimana mana laporan	Positif



Gambar 5 Visualisasi Wordcloud Sentimen Positif

Pada sentimen negatif, visualisasi wordcloud menunjukkan sebagai berikut :

1. Dominasi Kata "baliho": Kata "baliho" jelas mendominasi dengan frekuensi yang sangat tinggi, mencapai 1163, yang menunjukkan bahwa kata ini sangat relevan dan sering muncul dalam konteks negatif dalam dataset. Ini bisa menandakan bahwa baliho (sebagai media iklan) mungkin sering dikaitkan dengan pandangan negatif, terutama jika dataset berkaitan dengan iklan atau kampanye politik.
2. Frekuensi Kata Lain: Kata-kata seperti "caleg" dan "jalan" juga memiliki frekuensi yang signifikan, masing-masing 149 dan 126. Hal ini bisa menunjukkan topik lain yang sering mendapat respons negatif, seperti politik dan sistem peradilan.
3. Kata-kata Umum dalam Sentimen Negatif: Kata-kata seperti "gile", "sial", dan "parah" juga muncul dengan frekuensi yang cukup tinggi, mengindikasikan bahwa ini adalah ungkapan yang umum digunakan untuk mengekspresikan sentimen negatif.



Gambar 5 Visualisasi Wordcloud Sentimen Negatif

Data akan dibagi menjadi data training dan data testing dengan perbandingan 70:30. Perbandingan ini memiliki arti bahwa sebanyak 70% data akan menjadi data training, dan 30% lainnya sebagai data testing. Jumlah data training diperoleh sebesar 1548 data sedangkan pada data testing diperoleh 664 data.

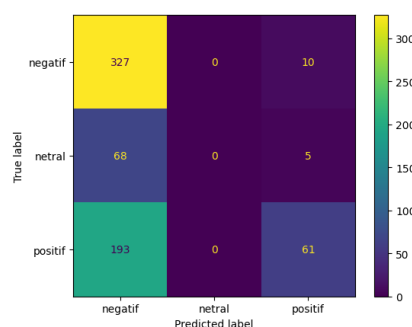
Proses klasifikasi dilakukan dengan multinomial *naïve bayes* dengan nilai alpha 1. Parameter ini digunakan untuk membantu mengatasi masalah zero probability yang mungkin terjadi pada data. Kemudian akan dilakukan pelatihan model dengan menggunakan data training (1548 data) yang disimpan dalam X_train. X_train merupakan fitur yang merepresentasikan teks ke dalam bentuk numerik (dalam penelitian menggunakan TF-IDF) dan y_train adalah label kelas setiap teks. Selama proses pelatihan, model akan menghitung probabilitas kemunculan setiap kata dalam setiap kelas (probabilitas likelihood). Kemudian akan dilakukan proses klasifikasi dengan data testing (664 data). Data testing akan diubah menjadi vector menggunakan cara yang sama saat melatih model. Lalu model akan menghitung probabilitas posterior dari setiap kelas untuk data testing. Terakhir, akan dilakukan pengambilan keputusan dimana probabilitas posterior tertinggi akan dipilih sebagai prediksi kelas pada data testing.

Untuk mengetahui performa model sudah baik maka akan dilakukan demo dengan menggunakan data yang tidak ada pada model untuk mengetahui prediksi dari sentimen pada opini yang di inputkan peneliti. Demo akan dilakukan dengan menggunakan *library joblib* yang memungkinkan memuat kembali model *naïve bayes* yang telah dibuat sebelumnya. `joblib.dump()` akan digunakan untuk menyimpan model *Naïve Bayes* dan *vectorizer* yang digunakan untuk mengubah teks menjadi representasi numerik. Kemudian file yang telah disimpan dalam *joblib* akan dimuat dan digunakan sebagai demo. Langkah selanjutnya menginputkan teks yang akan diklasifikasikan. Kemudian teks akan diubah kedalam vector numerik menggunakan *vectorizer* yang telah disimpan sebelumnya. Terakhir sistem akan melakukan prediksi untuk memprediksi probabilitas kelas teks input. Hasil yang ditampilkan berupa prediksi sentimen beserta nilai probabilitasnya.

- Negatif (0.538392): Hasil analisis menunjukkan bahwa kalimat tersebut memiliki probabilitas tertinggi untuk dikategorikan sebagai sentimen negatif. Sistem yang digunakan untuk menganalisis kalimat ini meyakini bahwa kalimat tersebut memang mengekspresikan ketidaksenangan atau gangguan.
- Positif (0.290584): Probabilitas untuk kategori positif juga muncul, namun dengan nilai yang lebih rendah. Ini bisa terjadi karena beberapa alasan, seperti adanya kata-kata yang memiliki konotasi positif dalam kalimat, meskipun secara keseluruhan kalimatnya negatif.
- Netral (0.171025): Probabilitas untuk kategori netral juga ada, namun dengan nilai yang paling rendah. Ini menunjukkan bahwa sistem kurang yakin bahwa kalimat tersebut tidak memiliki sentimen yang jelas, baik positif maupun negatif.

12. Analisis Hasil Evaluasi

Berdasarkan hasil klasifikasi sentimen menggunakan metode *naive bayes* diperoleh hasil yang ditunjukkan oleh gambar label *confussion matrix* 8.



Gambar 8 Label *Confussion Matrix*

Dari gambar 8 dapat dijabarkan bahwa warna yang lebih terang mewakili nilai yang lebih tinggi, sedangkan warna yang lebih gelap mewakili nilai yang lebih rendah. Nilai pada diagonal utama (dari kiri atas ke kanan bawah) menunjukkan jumlah prediksi benar (*true positive*) setiap sentimen. Untuk mempermudah mengetahui nilai dari *True Positive*, *True Negatif*, *False Positive*, dan *False Negatif* dari setiap sentimen, maka *confusion matrix* dari gambar 23 akan dibuat ke dalam model dua kelas. Hal ini dilakukan dengan membandingkan sentimen asli dan bukan sentimen asli, misalnya sentimen positif dengan sentimen bukan positif (sentimen negatif ditambah sentimen netral). *Confusion matrix* untuk sentimen positif, netral, maupun negatif dapat dijabarkan dengan tabel-tabel sebagai berikut:

Tabel 11 *Confusion Matrix* Sentimen Positif

	Positif	Bukan positif
Positif	61 (tp)	193 (fn)
Bukan positif	15 (fp)	395 (tn)

Tabel 12 *Confusion Matrix* Sentimen Netral

	Netral	Bukan netral
Netral	0 (tp)	73 (fn)
Bukan netral	0 (fp)	591 (tn)

Tabel 13 *Confusion Matrix* Sentimen Negatif

	Negatif	Bukan negative
Negatif	327 (tp)	10 (fn)
Bukan negative	261 (fp)	66 (tn)

Confusion matrix digunakan untuk menilai performa klasifikasi dengan membandingkan label yang diprediksi oleh model dengan label sebenarnya. Dari tabel di atas dapat dijabarkan bahwa:

1. *Confusion Matrix* untuk Sentimen Positif
 - a. *True Positive* (TP): 61 - Komentar positif yang benar diidentifikasi sebagai positif.
 - b. *False Negatif* (FN): 193 - Komentar positif yang salah diidentifikasi sebagai bukan positif.
 - c. *False Positive* (FP): 15 - Komentar bukan positif yang salah diidentifikasi sebagai positif.
 - d. *True Negatif* (TN): 395 - Komentar bukan positif yang benar diidentifikasi sebagai bukan positif.
2. *Confusion Matrix* untuk Sentimen Netral
 - a. *True Positive* (TP): 0 - Komentar netral yang benar diidentifikasi sebagai netral.
 - b. *False Negatif* (FN): 73 - Komentar netral yang salah diidentifikasi sebagai bukan netral.
 - c. *False Positive* (FP): 0 - Komentar bukan netral yang salah diidentifikasi sebagai netral.
 - d. *True Negatif* (TN): 591 - Komentar bukan netral yang benar diidentifikasi sebagai bukan netral.
3. *Confusion Matrix* untuk Sentimen Negatif
 - a. *True Positive* (TP): 327 - Komentar negatif yang benar diidentifikasi sebagai negatif.
 - b. *False Negatif* (FN): 10 - Komentar negatif yang salah diidentifikasi sebagai bukan negatif.

- c. *False Positive* (FP): 261 - Komentar bukan negatif yang salah diidentifikasi sebagai negatif.
- d. *True Negatif* (TN): 66 - Komentar bukan negatif yang benar diidentifikasi sebagai bukan negatif.

Dari hasil *confusion matrix* yang dijabarkan sebelumnya dapat diperoleh hasil *akurasi*, *presisi*, *recall*, dan *f1-score* yang dapat dilihat pada tabel dibawah ini.

Tabel 14 Hasil Klasifikasi Sentimen Dengan Metode *Naive Bayes*

Sentimen	Presisi	Recall	F1-score	Akurasi
Positif	80%	24%	37%	58%
Netral	0%	0%	0%	58%
Negatif	56%	97%	71%	58%

Dari tabel 14 dapat diketahui bahwa akurasi model dengan menggunakan metode *naive bayes* yaitu 58%. Pada sentimen positif memiliki nilai *presisi* sebesar 80%, nilai *recall* 24%, dan nilai *f1-score* 37%. Pada sentimen netral memiliki nilai *presisi* 0%, nilai *recall* 0% dan nilai *f1-score* 0%. Sedangkan pada sentimen negatif memiliki nilai *presisi* 56%, nilai *recall* 97%, dan diikuti nilai *f1-score* 71%.

Ditinjau dari metode yang digunakan dalam penelitian ini, *naive bayes* masih cukup baik dalam mengklasifikasikan sentimen positif dan negatif, dengan performa yang lebih baik untuk sentimen negatif. Hal ini dapat dilihat dari nilai *recall* yang lebih tinggi dari seluruh sentimen. Akan tetapi, metode ini memiliki kelemahan dalam mengklasifikasikan sentimen netral yang ditunjukkan dengan nilai *presisi*, *recall*, dan *f1-score* yang rendah. Hal ini dikarenakan ketidakseimbangan kelas yang menyebabkan data akan bias ke sentimen yang memiliki jumlah data yang besar. Keseimbangan antar kelas merupakan faktor penting yang mempengaruhi performa model *naive bayes*. Ketidakseimbangan kelas akan mengakibatkan model *naive bayes* lebih fokus pada pola data sentimen positif dan negatif sehingga kurang memperhatikan pola dari data pada sentimen netral. Namun disisi lain, berdasarkan persentase akurasi model *naive bayes* yang dihasilkan dalam penelitian ini sejalan dengan kajian yang dilakukan oleh Darwis, dkk (2020) yang menganalisis sentimen warga Twitter kepada pelayanan yang diberikan BMKG Nasional. Hasil penelitian menunjukkan bahwa hasil akurasi model mencapai 69,97% yang dapat diartikan bahwa model *naive bayes* cukup baik serta memberikan kemudahan dalam mengklasifikasikan sentimen positif, negatif, maupun netral.

KESIMPULAN

1. Penerapan metode *naive bayes* terkait dengan analisis sentimen alat peraga kampanye pemilu 2024 menggunakan data dari X. Dari hasil analisis yang sudah dilakukan komentar negatif mendominasi sebesar 51,2 % diikuti komentar positif 36.3% dan netral 12.3 %. Pada tahap klasifikasi dengan algoritma *naive bayes* dilakukan pembagian data training dan data testing. Data training berjumlah 1548 dan data testing berjumlah 664.
2. Pengujian dengan menggunakan *confusion matrix* dan klasifikasi sentimen menggunakan algoritma *naive bayes* menunjukkan bahwa sentimen positif memiliki nilai *presisi* sebesar 80% *recall* 24% dan *F1-score* 37%. Pada sentimen netral memiliki nilai memiliki nilai *presisi* 0% *recall* 0% dan *F1-score* 0%. Pada sentimen negatif memiliki nilai *presisi* 56% *recall* 97% dan *F1-score* 71%. Sedangkan untuk nilai akurasi sama yaitu 58% untuk ketiga kategori sentimen positif, negatif dan netral.

DAFTAR PUSTAKA

- Agustina, Lidya. 2020. "Viralitas Konten Di Media Sosial." *Majalah Ilmiah Semi Populer Komunikasi Massa* 1(2):149–60.
- Anam, C. (2018). Pengaturan Alat Peraga Kampanye Dalam Pelaksanaan Pemilihan Kepala Daerah. *JKMP (Jurnal Kebijakan Dan Manajemen Publik)*, 6(2), 113–118. <https://doi.org/10.21070/jkmp.v6i2.3005>
- Alita, D., & Isnain, A. R. (2020). Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan *Random Forest Classifier*. *Jurnal Komputasi*, 8(2), 50-58.
- Asy'Ari, A. I. (2023). *Sejarah Logo Twitter, dari Balon Hijau Menjadi X*. JawaPos.Com. <https://www.jawapos.com/lifestyle/011813388/sejarah-logo-twitter-dari-balon-hijau-menjadi-x>
- Bastoni, R. P. (2024). *Analisis Sentimen Opini Publik Terhadap Calon Presiden Indonesia 2024 Menggunakan Metode Naive Bayes Dan Lexicon Based Pada Platform YouTube Skripsi Oleh*. UIN Syarif Hidayatullah.
- Bruns, A., dkk. (2021). *X and the transformation of journalism: A comparison of trends in five countries*. *Journalism Studies*, 22(4), 490-510.
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.
- Darwis, D., Siskawati, N., & Abidin, Z. (2020). Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional. *Jurnal TEKNO KOMPAK*, 15(1), 131–145.
- Géron, A. (2019). *Hands-On Machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, Inc.
- Hidayat, S., Risky, A., & Ode Muhamad Sulihin, L. (2024). *Penegakan Hukum Pidana Oleh Bawaslu Atas Pelanggaran Pemasangan Alat Peraga Kampanye Di Kota Kendari* (Vol. 3, Issue 1).
- Iryana, T. M., Indriati, I., & Adikara, P. P. (2021). Analisis Sentimen Masyarakat Terhadap Mass Rapid Transit Jakarta Menggunakan Metode *Naive Bayes* Dengan Normalisasi Kata. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5(6), 2753-2760.
- Nooryuda Prasetya, Y., & Winarso, D. (2021). Penerapan *Lexicon Based* Untuk Analisis Sentimen Pada *Twitter* Terhadap Isu *Covid-19*. *Jurnal Fasilkom*, 11(2), 97-103.
- Permadi, Vynska Amalia. 2020. "Analisis Sentimen Menggunakan Algoritma *Naive Bayes* Terhadap Review Restoran Di Singapura." *Jurnal Buana Informatika* 11(2):141–51. doi: 10.24002/jbi.v11i2.3769.
- Pratiwi, M. A. (2022). *Analisis Sentimen Untuk Mengetahui Perspektif Masyarakat Mengenai Vaksin Covid-19* [Universitas Atma Jaya Yogyakarta]. [http://repositori.unsil.ac.id/4671/6/9.BAB II.pdf](http://repositori.unsil.ac.id/4671/6/9.BAB%20II.pdf)
- Rahmi, F. (2020). *Analisis Sentimen Berdasarkan Aspek Pada Ulasan Tempat Wisata Menggunakan Modified K-Nearest Neighbor* [Universitas Komputer Indonesia]. [https://elibrary.unikom.ac.id/id/eprint/4043/%0Ahttps://elibrary.unikom.ac.id/id/eprint/4043/8/UNIKOM_Faisal Rahmi_BAB 2.pdf](https://elibrary.unikom.ac.id/id/eprint/4043/%0Ahttps://elibrary.unikom.ac.id/id/eprint/4043/8/UNIKOM_Faisal%20Rahmi_BAB%202.pdf)
- Salsabila, N. A. (2022). *Analisis Sentimen Pada Media Sosial X Terhadap Tokoh Gus Dur Menggunakan Metode Naive Bayes Dan Support Vector Machine (Svm)* (Bachelor's thesis, Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta).

- Wahyudi, M. M. (2021). *Analisis Sentimen Terhadap Opini Masyarakat Terkait Vaksinasi COVID-19 pada Twitter dengan Algoritma Random Forest Classifier dan Informatika Gain*. Universitas Siliwangi.
- Yutika, C. H., Adiwijaya, A., & Al Faraby, S. (2021). Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan *Naïve Bayes*. *Jurnal Media Informatika Budidarma*, 5(2), 422-430.
- Zhao, (2020). *Research on Web Crawling Technology Based on Multithreading*. In *2020 5th International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*, 433-437.