



## **Analisis sentimen publik terhadap calon presiden 2024 di instagram menggunakan metode naïve bayes**

**Tamara Cindy Samsita Rani<sup>1</sup>, Eka Sahputra<sup>2</sup>**

Email: <sup>1</sup>tamaracindy539@gmail.com, <sup>2</sup>ekasahputra@umb.ac.id

<sup>1</sup>Universitas Muhammadiyah Bengkulu, Bengkulu 38119, Indonesia

<sup>2</sup>Universitas Muhammadiyah Bengkulu, Bengkulu 38119, Indonesia

Diterima: 30 November 2024 | Direvisi: 18 Desember 2024 | Disetujui: 27 Desember 2024

©2020 Program Studi Teknik Informatika Fakultas Ilmu Komputer,  
Universitas Muhammadiyah Riau, Indonesia

### **Abstrak**

Dalam penelitian ini, analisis sentimen publik terhadap pasangan capres di Indonesia pada tahun 2024 dilakukan melalui media sosial Instagram. Indonesia sendiri merupakan salah satu negara dengan jumlah pengguna Instagram terbanyak. Salah satu postingan Instagram yang saat ini menjadi sorotan publik ditemukan di akun kompascom, najwashihab, dan detikcom yang memposting berita tentang pasangan capres 2024. Dalam komentar tersebut terdapat berbagai macam komentar mulai dari hal-hal positif seperti dukungan hingga hal-hal negatif seperti mengomentari kekurangan masing-masing pasangan calon, untuk itu dilakukan klasifikasi teks untuk mengetahui opini publik tentang pasangan calon presiden. Proses klasifikasi komentar dalam penelitian ini menggunakan algoritma Naïve Bayes untuk menentukan nilai positif, netral dan negatif dari ribuan komentar. Metode yang akan diterapkan menggunakan bahasa pemrograman Python dengan pengujian matriks kebingungan untuk menentukan tingkat akurasi dalam model. Berdasarkan hasil pengujian, dapat disimpulkan bahwa penggunaan algoritma Naïve Bayes yang digunakan sebagai metode klasifikasi dalam analisis sentimen berbasis komentar di Instagram memiliki tingkat akurasi yang relatif baik dengan akurasi rata-rata lebih dari 60%.

**Kata kunci:** Klasifikasi, Instagram, naïve bayes, presiden, analisis sentimen.

### ***Analysis of public sentiment towards the 2024 presidential candidates on instagram using the naïve bayes method***

#### **Abstract**

*In this research, an analysis of public sentiment towards presidential candidate pairs in Indonesia in 2024 was carried out via the social media Instagram. Indonesia itself is one of the countries with the highest number of Instagram users. One of the Instagram posts that is currently in the public spotlight found on the kompascom, najwashihab, and detikcom accounts which post news about the 2024 presidential candidate pairs. In these comments there are various kinds of comments ranging from positive things such as support to negative things such as commenting on the shortcomings of each candidate pair, for this reason text classification is carried out to find out public opinion about the presidential candidate pair. The comment classification process in this research uses the Naïve Bayes algorithm to determine positive, neutral and negative values from thousands of comments. The method that will be applied uses the Python programming language with confusion matrix testing to determine the level of accuracy in the model. Based on the test results, it can be concluded that the use of the Naïve Bayes algorithm used as a classification method in comment-based sentiment analysis on Instagram has a relatively good accuracy rate with an average accuracy of more than 60%.*

**Keywords:** Classification, Instagram, naïve bayes, president, sentiment analysis.

## 1. PENDAHULUAN

Pemilu 2024 mendatang di Indonesia menandai momen penting dalam sejarah bangsa. Ini adalah pertama kalinya bangsa Indonesia mengikuti pemilu serentak yang tidak hanya mencakup pemilihan presiden, tetapi juga pemilihan perwakilan DPR RI, DPD RI, DPRD provinsi, dan DPRD kabupaten/kota, serta pemilihan gubernur dan bupati/walikota di seluruh Indonesia. Keseruan pemilu ini, khususnya Pilpres sudah mulai terasa sejak awal tahun 2022 [1]. Aturan pemilihan presiden di Indonesia dituangkan dalam Undang-Undang Nomor 42 Tahun 2008. Pemilihan ini diadakan setiap 5 tahun, dan jika seorang presiden telah menyelesaikan 2 masa jabatan berturut-turut, maka dia tidak dapat dicalonkan lagi sebagai presiden. Akibat aturan ini, presiden yang telah menjabat selama 2 masa jabatan berturut-turut tidak dapat mengikuti Pilpres 2024 [2].

Media sosial merupakan platform kampanye yang efektif dan efisien dibandingkan dengan media kampanye konvensional dengan atribut atau alat untuk pidato publik dan ruang terbuka [3]. Dalam riset analisis sentimen ini, Instagram dipilih karena merupakan platform media sosial yang mengutamakan berbagi foto dan video melalui berbagai fitur yang disediakan oleh aplikasi. Pengguna dapat berinteraksi di platform dengan saling menyukai, menulis komentar, berbagi dan menyimpan, serta berinteraksi secara pribadi melalui pesan pribadi di setiap akun[4].

Proses analisis sentimen melibatkan analisis, ekstraksi, dan pemahaman informasi tekstual secara otomatis untuk mendapatkan data sentimen dari kalimat opini atau komentar [5]. Data ini mengungkapkan sikap dan emosi seseorang terhadap entitas tertentu, misalnya seseorang, peristiwa, atau topik. Tujuan utamanya adalah untuk memberikan data berharga kepada individu dari kumpulan data tidak terstruktur [6]. Cara kerja analisis sentimen adalah dengan menghasilkan variasi dari sekumpulan besar data yang tersedia, kemudian mencoba menemukan pola yang sesuai dari sekumpulan data teks yang ada [7].

Naïve Bayes adalah metode klasifikasi yang biasa digunakan dalam pembelajaran mesin dan statistik untuk mengkategorikan data ke dalam kelas atau kategori yang telah ditentukan sebelumnya [8]. Metode ini dipilih karena berakar pada teorema Bayes yang diusulkan oleh Thomas Bayes, yang menggunakan probabilitas dan statistik untuk memprediksi peristiwa masa depan berdasarkan data masa lalu. Pengklasifikasian dalam Teorema Bayes adalah dengan cara memproses klasifikasi dengan melatih data training [9]. Metode yang dinamai Teorema Bayes sangat efisien dalam mengkategorikan data, sehingga banyak peneliti sebelumnya yang memilihnya dalam studi pembelajaran mesin dan penambangan data [10]. Naïve Bayes menawarkan klasifikasi yang efisien, mewujudkan esensi teorema Bayes dalam membuat prediksi yang tepat [11].

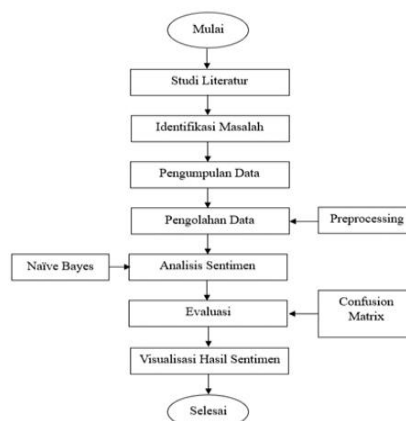
Metode Naïve Bayes sering digunakan dalam berbagai penelitian tentang analisis dokumen teks dan merupakan algoritma yang populer karena kemudahannya dan telah terbukti memberikan hasil yang memuaskan di berbagai domain penelitian [12]. Terdapat beberapa penelitian terdahulu yang menggunakan metode Naïve Bayes untuk analisis sentimen yakni penelitian dari Misrun et al pada tahun 2023 dan Simatupang et al tahun 2024 keduanya mendapatkan hasil yang cukup baik dengan akurasi diangka 78% dan 81% [13], [14].

Dalam konteks Pilpres 2024, di mana calon presiden baru akan bermunculan, analisis sentimen di Instagram dapat membantu dalam memahami opini publik, mengukur popularitas calon presiden, dan mendukung pengambilan keputusan politik yang lebih baik [15]. Tujuan dari penelitian ini adalah untuk mengklasifikasikan sentimen publik terhadap calon presiden berdasarkan komentar di Instagram, karena media sosial memposting ratusan bahkan ribuan komentar dalam waktu singkat, komentar publik harus diklasifikasikan dengan menggunakan metode Naïve Bayes. Karena jika Anda mengklasifikasikannya secara manual, akan memakan waktu lama dan sulit bagi manusia.

Hasilnya adalah pemahaman yang lebih dalam tentang pandangan publik terhadap pasangan kandidat dan memperoleh informasi tentang apakah sentimen publik cenderung positif, netral atau negatif terhadap pasangan kandidat berdasarkan postingan di Instagram. Ada beberapa tahapan dalam penelitian ini, yaitu pengumpulan data, pengolahan data, analisis sentimen, evaluasi, dan visualisasi hasil sentimen

## 2. METODE PENELITIAN

Klasifikasi Naïve Bayes melibatkan beberapa tahap, seperti yang digambarkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

### 2.1 Studi sastra

Merupakan proses pengumpulan, analisis, dan interpretasi informasi dari berbagai sumber sastra yang relevan dengan topik penelitian seperti membaca buku, artikel, jurnal, atau materi online dalam konteks yang sama tetapi dengan metode penyelesaiannya yang berbeda.

### 2.2 Identifikasi masalah

Identifikasi masalah dalam penelitian ini terkait dengan evaluasi dan klasifikasi teks atau data berbasis sentimen menggunakan metode machine learning algoritma Naïve Bayes.

### 2.3 Pengumpulan data

Dalam penelitian ini, proses pengumpulan data dilakukan dengan cara mengikis. Proses scraping dilakukan secara otomatis menggunakan ekstensi IG Comment Exporter di Google Chrome yang hasilnya adalah komentar di platform Instagram terutama dari akun Najwa Shihab, Kompascom dan Dertikcom. Berisi informasi tentang pasangan capres di tahun 2024.

### 2.4 Pemrosesan data

Data yang telah terkumpul kemudian memasuki tahap pra-pemrosesan atau persiapan data. Prapemrosesan adalah tahap penting Untuk mengurangi masalah yang mungkin mengganggu hasil pemrosesan data [16]. Tahap preprocessing bertujuan untuk membersihkan, memformat, dan menyiapkan data sehingga siap digunakan untuk analisis sentimen. Proses ini melibatkan langkah-langkah seperti di bawah ini:

1. Case Folding  
Case Folding mengubah karakter huruf besar menjadi huruf kecil yang seragam, sehingga semua huruf menjadi huruf kecil dan juga menghilangkan tanda baca dan angka [17].
2. Cleaning  
Pembersihan adalah menghapus karakter untuk mengurangi kebisingan dalam proses klasifikasi. Kata-kata yang harus dihapus termasuk "nama pengguna", "sebutan", dan sebagainya [18].
3. Tokenisasi  
Tokenisasi memisahkan teks menjadi kata-kata yang disebut token. Tokenisasi didasarkan pada pemisahan angka dan huruf, termasuk tanda baca dan spasi [19].
4. Normalisasi  
Normalisasi adalah proses mengembalikan kata-kata non-standar dalam Kamus Bahasa Indonesia ke kata-kata standar [20].
5. Stop Word  
Stop word adalah kata-kata yang sering digunakan yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki arti atau makna [21].
6. Stemming  
Stemming melibatkan standarisasi kata-kata dengan mengubahnya ke bentuk dasarnya melalui penghapusan afiks atau akhiran kata, sehingga mengatasi berbagai variasi morfologis dalam teks [22].
7. Pelabelan Data  
Proses pelabelan data dibagi menjadi tiga kelas, yaitu positif, netral dan negatif. Para peneliti menggunakan pustaka transformator yang tersedia dalam Python untuk membantu proses pelabelan.
8. Pembobotan TF-IDF  
TF-IDF, kependekan dari Term Frequency-Inverse Document Frequency, adalah teknik yang digunakan untuk mengubah kata-kata menjadi nilai numerik, sering disebut sebagai pembobotan, untuk mengukur signifikansinya. Tujuan TF-IDF adalah untuk menilai pentingnya sebuah kata dalam dokumen dengan memberikan bobot untuk setiap kata berdasarkan frekuensi dan kemunculannya di beberapa dokumen [23]. Proses ini melibatkan penentuan seberapa sering sebuah kata muncul dalam dokumen (frekuensi istilah) dan menimbangkannya terhadap frekuensinya di seluruh korpus dokumen (frekuensi dokumen terbalik) [24].

$$tf * idf_{i,j} = tf_{i,j} \times idf_i \quad (1)$$

- Bobot maksimum jika kata t sering muncul di beberapa dokumen.
- Jika kata t sering muncul di banyak dokumen, atau jika kata t jarang muncul di beberapa dokumen, maka bobotnya akan relatif rendah.
- Bobot terendah jika kata t ada di semua dokumen.

### 2.5 Analisis Sentimen Menggunakan Metode Naïve Bayes

Naïve Bayes adalah teknik pembelajaran mesin yang mengandalkan perhitungan probabilitas. Salah satu kekuatannya terletak pada kemampuannya untuk membuat prediksi yang akurat dengan data pelatihan minimal, karena memperkirakan parameter (seperti rata-rata dan varians variabel) yang diperlukan untuk klasifikasi hanya menggunakan kumpulan data kecil [25].

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (2)$$

Informasi:

X = kategori data tidak dikenal

H = data hipotesis X termasuk dalam kategori tertentu

P(H|X) = probabilitas hipotesis H berdasarkan kondisi X (probabilitas posterior)

P(H) = probabilitas hipotesis H (probabilitas sebelumnya)

P(X|H) = probabilitas X berdasarkan kondisi ini

Probabilitas P(X) = X

### 2.6 Evaluasi

Para peneliti melakukan evaluasi kinerja menggunakan matriks kebingungan untuk menilai efektivitas model klasifikasi Naïve Bayes, menghasilkan nilai akurasi, presisi, dan penarikan [26].

- Akurasi menunjukkan proporsi informasi yang dijawab dengan benar oleh sistem dibandingkan dengan total informasi.
- Presisi mengukur tingkat akurasi terkait permintaan pengguna dan respons sistem.
- Recall mengukur akurasi antara informasi yang ditarik dan informasi yang diambil sebelumnya.

Rumus untuk matriks kebingungan, seperti yang diuraikan dalam [27], adalah sebagai berikut: Akurasi (A) dihitung dengan membagi total nilai positif sejati dan negatif sejati dengan jumlah keseluruhan data.

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

Presisi (P) adalah persentase nilai Positif sejati dari semua nilai Positif yang diprediksi.

$$P = \frac{TP}{(TP + FP)} \times 100\% (3)$$

Recall (R) adalah persentase prediksi Positif dibandingkan dengan Positif sejati.

$$R = \frac{TP}{(TP + FN)} \times 100\% (4)$$

### 2.7 Visualization of Sentient Results

Visualization is a process or technique for creating graphical representations of data. The goal is to visualize data in a more intuitive and easy to understand way.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Hasil Pengikisan Data

Pengikisan dilakukan secara otomatis menggunakan Ekstensi Eksporir Komentar IG di Google Chrome. Sebelum menggunakan Ekstensi IG, instal terlebih dahulu dan aktifkan di Google Chrome. Tahap awal yang dilakukan sebelum proses scraping adalah menentukan dari alamat URL Instagram mana komentar akan diambil. Peneliti mengikis salah satu postingan di akun Najwashihab, Kompascom, dan Detikcom sebagai dataset..

User Id	User Name	Comment Id	Comment Text
36241010308	baangkitts	18014805668844930	@nadirahayu koe aa sang maju isone ko ngalak ngalak
5743498612	nedirahayu	1787788150482278	@baangkitts ndelek rekam jejak'e.
36241010308	baangkitts	17982041276614029	@nadirahayu isone kok ngalak ngalak
5743498612	nedirahayu	17912677923505952	@baangkitts alu itu nyampene laita, manglana koe itu ndelok'o rekam jejak'e, ben eroh bosik'e Prabowo - Gibran.
36241010308	baangkitts	18011896187121553	@nadirahayu hoo nak bosok ra dadi merhan mbt walok
36241010308	baangkitts	1804917749556057	@nadirahayu ojo gara gara bedo pilihan sampean nginiki
5743498612	nedirahayu	18227027859300112	@baangkitts Prabowo pelaku penculikan aktivis pro demokrasi tahun 1997-1998 [berdasarkan surat keputusan dewan k
36241010308	baangkitts	17956316557087663	@nadirahayu fenomena 5 tahun sekali
5743498612	nedirahayu	18054075096002775	@baangkitts menjatikan lawan politik berdasarkan fakta rekam jejaknya itu diabolikan dalam pemilu, yang tidak boleh
5743498612	nedirahayu	18000146528464131	@baangkitts sebelum Prabowo tidak diadili atas pelanggaran HAM, maka nya itu akan melekat kepada Prabowo.
18275188993	keosvll	18039503983615813	@nadirahayu rnr 1 parah
45094373961	febriant0618	1822143371249102	Bair saja survei bayaran, nanti juga ganjar mahfud yang di lantik betuk gak bang @najibamullah
13223705274	abdulmuni384	1787732756994652	AMIN...
1918439869	rntn_mamand	1827622149708630	
7941430997	yoelgaring	1829300516448069	@ikanda no more Indonesia forever. No AMIN
11563857819	verren_is	1796655327664981	@dani_saiyan89 keren kan
11563857819	verren_is	17913645593852377	@amandayllana60 iya presiden nya rohinnya
22905614157	dheco_yeremia	1788515663095119	tiap saat ada survey
8537528404	affhn_	17992641706592741	@amandayllana60 BOT di pake buat nyaspe
552701046	andriakambej	18004716362343228	@erfan_maullana berarti iya yg ga bisa
552701046	andriakambej	17855050626067235	@erfan_maullana wkwkwkema ngakak nih somus survei malahan 2014 2019 wooo yg menang ..coba iya cari lagi somu
552701046	andriakambej	18037401010579042	@erfan_maullana masuk iya lupa jgga quick count aja wooo bisa menang sampek sujud syukur
552701046	andriakambej	1790485406906659	@erfan_maullana kok balas lsi ...bukannya yg di bahr: d atas ltbang kompas
4928336230	erlanggyrtho55	18010365778979068	All in prabowo
4840949846	cidngarilurua	18218919535267972	@dijamrudny amang iya lah bang
67582425133	nicahozozhva18	17870182107055164	@muhammadalfal5h kata nispie gua mlilih ya
1764408860	niss_kasuar	18018094957965487	Di tempat lain 02 di 53%
608276794	robby.suhendra	17956979893088579	Lagi nyari buzzer 02 yg nyerang paslon 1 & 3
57642150630	granatgadin	18006746111329252	suave maramput..mambuat opini seolah menang..politik basi...bellimbing sayur ttp yungsap
57901163424	sofin003	17910490914796981	SIWA GRATIS TUNJANGAN BUMIL 67, jgn lupa ya mas anis... HUTANGNYA dibalik dulu, Semflaga uya menunggu loh
3403990143	nlarohkaroh	18214584849282338	@didi_hy_chuan panas rosoooo
3403990143	nlarohkaroh	17897019413936581	@nadirahayu kau aj, lmi ttp Prabowo
3403990143	nlarohkaroh	17995124096218039	Prabowo

Gambar 2. Hasil Pengikisan Data

Total dataset berjumlah 39.015 komentar dengan 28.407 dari akun Najwashihab, 5.279 dari akun Kompascom, dan 5.329 dari akun Detikcom. Perbandingan yang digunakan untuk setiap dataset adalah 80% data pelatihan dan 20% data pengujian. Di bawah ini adalah himpunan data terperinci untuk setiap akun:

Tabel 1. Detail Kumpulan Data Najwashihab

<b>Instagram Komentar Dataset</b>	<b>Data Pelatihan</b>	<b>22,726 (80%)</b>		
	<b>Data Pengujian</b>	<b>5,681 (20%)</b>		
		<b>Pelatihan</b>	<b>Pengujian</b>	<b>Jumlah</b>
<b>Kelas</b>	Positif	7,576	1,895	9,471
	Netral	7,575	1,893	9,468
	Negatif	7,575	1,893	9,468
	<b>Jumlah Komentar</b>	<b>22,726</b>	<b>5,681</b>	<b>28,407</b>

Tabel 2. Detail Dataset Kompascom

<b>Instagram Komentar Dataset</b>	<b>Data Pelatihan</b>	<b>22,726 (80%)</b>		
	<b>Data Pengujian</b>	<b>5,681 (20%)</b>		
		<b>Pelatihan</b>	<b>Pengujian</b>	<b>Jumlah</b>
<b>Kelas</b>	Positif	1,409	352	1,761
	Netral	1,407	352	1,759
	Negatif	1,407	352	1,759
	<b>Jumlah Komentar</b>	<b>4,223</b>	<b>1,000</b>	<b>5,279</b>

Tabel 3. Detail Dataset Detikcom

<b>Instagram Komentar Dataset</b>	<b>Data Pelatihan</b>	<b>22,726 (80%)</b>		
	<b>Data Pengujian</b>	<b>5,681 (20%)</b>		
		<b>Pelatihan</b>	<b>Pengujian</b>	<b>Jumlah</b>
<b>Kelas</b>	Positif	1,409	352	1,761
	Netral	1,407	352	1,759
	Negatif	1,407	352	1,759
	<b>Jumlah Komentar</b>	<b>4,223</b>	<b>1,000</b>	<b>5,279</b>

### 3.2 Hasil Uji Model Dataset Kompascom

Para peneliti melakukan pengujian model menggunakan pustaka Scikit-learn dalam bahasa pemrograman Python. Di bawah ini adalah ilustrasi yang menampilkan hasil pengujian model pada dataset dari Kompascom, Najwashihab, dan Detikcom. Hasilnya termasuk nilai akurasi matriks kebingungan untuk setiap himpunan data.

Tabel 4. Hasil Uji Model Dataset Kompascom

Confusion Matrix:			
	Negative	Neutral	Positive
negative	247	152	1
neutral	60	432	3
positive	17	116	28
Accuracy Score:		0.6695075757575758	

Akurasi dataset Kompascom adalah 0,6695 atau persentase 66,95%.

Tabel 5. Hasil Uji Model Dataset Najwashihab

Confusion Matrix:			
	Negative	Neutral	Positive
Negative	1435	582	5
Neutral	412	2304	30
Positive	142	512	260
Accuracy Score:		0.7038014783526927	

Akurasi dataset Najwashihab adalah 0,7038 atau jika persentasenya 70,38%.

Tabel 6. Hasil Uji Model Dataset Detikcom

Confusion Matrix:			
	Negative	Neutral	Positive
Negative	354	96	0
Neutral	142	345	2
Positive	46	69	12
Accuracy Score:		0.6669793621013134	

Akurasi dataset Detikcom adalah 0,6669 atau persentase 66,69%. Nilai akurasi yang diperoleh dari ketiga kumpulan data dihitung dengan membagi jumlah nilai matriks kebingungan diagonal dengan total data.

### 3.3 Evaluasi Model

Para peneliti menggunakan pustaka Scikit-learn dalam bahasa pemrograman Python untuk proses evaluasi model. Evaluasi model melibatkan penilaian dan pengukuran kualitas prediksi atau model klasifikasi berdasarkan data yang digunakan. Hal ini dicapai dengan membandingkan hasil prediksi model dengan nilai data yang sebenarnya. Selanjutnya, kinerja model klasifikasi untuk setiap kelas dapat ditentukan dari nilai presisi, recall, dan skor F1. Nilai presisi, keingatan, dan skor F1 yang lebih tinggi menunjukkan performa model yang lebih baik. Metrik ini digambarkan pada Gambar 6 untuk analisis lebih lanjut.

Tabel 7. Hasil Kinerja Model Kompascom

Data Kompas				
Classification Report:				
	Precision	Recall	F1-Score	Support
Negative	0.76	0.62	0.68	400
Neutral	0.62	0.87	0.72	495
Positive	0.88	0.17	0.29	161
Accuracy			0.67	1056
Macro Avg	0.75	0.55	0.57	1056
Weighted Avg	0.71	0.67	0.64	1056

Hasil evaluasi model dataset Kompascom dapat dilihat bahwa nilai presisi untuk mengukur seberapa akurat sistem mengklasifikasikan data pada kelas negatif adalah 76%, netral 62%, dan positif adalah 88%, yang berarti semakin tinggi nilai presisi, semakin akurat sistem dalam mengklasifikasikan data dalam kategori tersebut. Pada dataset presisi Kompascom, nilai tertinggi berada di kelas positif, yaitu 88%. Sementara itu, penarikan mengukur seberapa baik sistem menemukan data dalam dokumen yang benar-benar termasuk dalam kategori kelas negatif pada 62%, netral pada 87%, dan positif pada 17%, yang berarti bahwa dalam kumpulan data penarikan nilai tertinggi jatuh ke dalam kelas netral pada 87%. F1-Score, metrik yang berasal dari rata-rata harmonik presisi dan keingatan, berfungsi sebagai ukuran komprehensif kinerja sistem. Ini mencerminkan presisi, yang mengukur keakuratan prediksi positif, dan recall, yang mengukur kemampuan sistem untuk mengidentifikasi semua contoh yang relevan. F1-Score yang lebih tinggi menunjukkan kinerja sistem yang unggul, menandakan keseimbangan antara presisi dan pemingatan. Dalam dataset Kompascom, nilai skor f1 tertinggi termasuk dalam kategori kelas netral sebesar 72%.

Selanjutnya, kami mengukur kualitas kinerja model prediksi untuk kumpulan data Najwashihab. Berikut ini dapat dilihat pada Tabel 8.

Tabel 8. Hasil Kinerja Model Najwashihab

Data Najwa				
Classification Report:	Precision	Recall	F1-Score	Support
Negative	0.72	0.71	0.72	2022
Neutral	0.68	0.84	0.75	2746
Positive	0.88	0.28	0.43	914
Accuracy			0.70	5682
Macro Avg	0.76	0.61	0.63	5682
Weighted Avg	0.73	0.70	0.69	5682

Hasil evaluasi model data Najwashihab dapat dilihat bahwa nilai presisi untuk mengukur seberapa akurat sistem mengklasifikasikan data kelas negatif adalah 72%, netral adalah 68%, dan positif adalah 88%. Artinya, pada dataset presisi Kompascom nilai tertinggi berada di kelas positif, yaitu 88%. Sementara itu, penarikan mengukur seberapa baik sistem menemukan data dalam dokumen yang benar-benar termasuk dalam kategori kelas negatif pada 71%, netral pada 84%, dan positif pada 28%, yang berarti bahwa dalam kumpulan data penarikan nilai tertinggi jatuh ke dalam kelas netral pada 84%. F1-Score adalah kombinasi dari presisi dan penarikan. Semakin tinggi nilai skor f1, semakin baik kinerja sistem. Dalam dataset Kompascom, nilai skor f1 tertinggi termasuk dalam kategori kelas netral sebesar 75%.

Terakhir, mengukur kualitas kinerja model prediksi dataset Detikcom. Berikut ini dapat dilihat pada Tabel 9.

Tabel 9. Hasil Kinerja Model Detikcom

Data Detik				
Classification Report:	Precision	Recall	F1-Score	Support
Negative	0.65	0.79	0.71	450
Neutral	0.68	0.71	0.69	489
Positive	0.86	0.09	0.17	127
Accuracy			0.67	1066
Macro Avg	0.73	0.53	0.52	1066
Weighted Avg	0.69	0.67	0.64	1066

Hasil evaluasi model data Najwashihab dapat dilihat bahwa nilai presisi untuk mengukur seberapa akurat sistem mengklasifikasikan data kelas negatif adalah 65%, netral 68%, dan positif 86%. Artinya, pada dataset presisi Kompascom nilai tertinggi berada di kelas positif, yaitu 88%. Sementara itu, penarikan mengukur seberapa baik sistem menemukan data dalam dokumen yang benar-benar termasuk dalam kategori kelas negatif pada 79%, netral pada 71%, dan positif pada 9%, yang berarti bahwa dalam kumpulan data penarikan nilai tertinggi jatuh ke dalam kelas negatif pada 79%. F1-Score adalah kombinasi dari presisi dan penarikan. Semakin tinggi nilai skor f1, semakin baik kinerja sistem. Dalam dataset Kompascom, nilai skor f1 tertinggi termasuk dalam kategori kelas negatif sebesar 71%.

### 3.4 Visualisasi Hasil Analisis Sentimen

Sebelum memvisualisasikan hasil analisis sentimen, pertama-tama buat persentase untuk kelas positif, netral, dan negatif untuk setiap kandidat. Para peneliti menggunakan pustaka Pandas dalam bahasa pemrograman Python. Berikut ini adalah persentase hasil masing-masing kelas positif, netral, dan negatif dari analisis sentimen masing-masing kandidat, yaitu Anies, Prabowo, dan Ganjar pada akun Instagram Kompascom, Najwashihab, dan detikcom masing-masing. Di bawah ini adalah penjelasan tentang hasil sentimen untuk setiap akun:

#### 1. Hasil Sentimen Akun Kompascom

Tabel 10. Persentase Sentimen Dataset Kompascom

Distribusi Sentimen untuk Data Kompas.com			
	Ganjar	Prabowo	Anis
Positive	15.17%	9.44%	15.08%
Neutral	71.33%	60.67%	67.60%
Negative	13.51%	29.89%	17.32%

Hasil dari dataset KompascomAnies memperoleh sentimen positif sebesar 15,08%, netral 67,60%, dan negatif 17,32%. Sementara itu, Prabowo menerima sentimen positif sebesar 9,44%, netral 60,67%, dan sentimen negatif sebesar 29,89%. Dan Ganjar menerima sentimen positif sebesar 15,17%, netral 71,33%, dan negatif 13,51%.

#### 2. Hasil Sentimen Akun Najwashihab

Tabel 11. Persentase Sentimen Kumpulan Data Najwashihab

Distribusi Sentimen untuk Data Najwa Shihab			
	Ganjar	Prabowo	Anis
Positive	9.93%	10.98%	20.14%
Neutral	59.37%	57.88%	52.03%
Negative	30.71%	31.14%	27.82%

Hasil dari dataset NajwashihabAnies memperoleh sentimen positif sebesar 20,14%, netral 52,03%, dan negatif 27,82%. Sementara itu, Prabowo menerima sentimen positif sebesar 10,98%, netral 57,88%, dan sentimen negatif sebesar 31,14%. Dan Ganjar menerima sentimen positif sebesar 9,93%, netral 59,37%, dan negatif 30,71%.

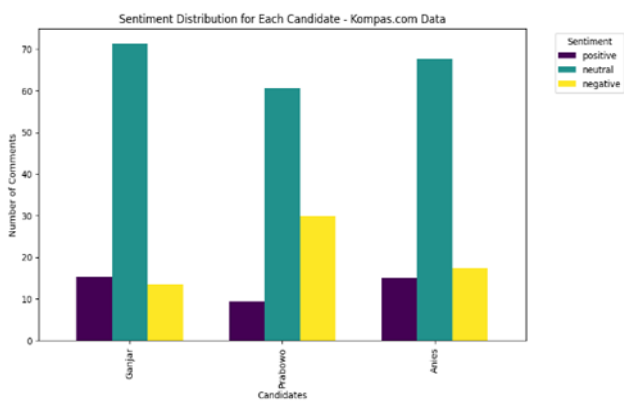
### 3. Hasil Sentimen Akun Detikcom

Tabel 12. Persentase Sentimen Dataset Detikcom

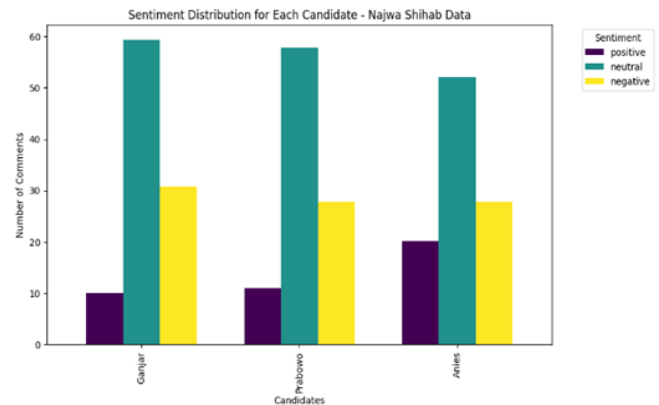
Distribusi Sentimen untuk Data Detik.com			
	Ganjar	Prabowo	Anis
Positive	16.45%	14.81%	7.00%
Neutral	65.79%	52.59%	53.67%
Negative	17.76%	32.59%	39.33%

Hasil dari dataset DetikcomAnies memperoleh sentimen positif sebesar 7%, netral 53,67%, dan negatif 39,33%. Sementara itu, Prabowo menerima sentimen positif sebesar 14,81%, netral 52,59%, dan sentimen negatif sebesar 32,59%. Dan Ganjar menerima sentimen positif sebesar 16,45%, netral 65,79%, dan negatif 17,76%.

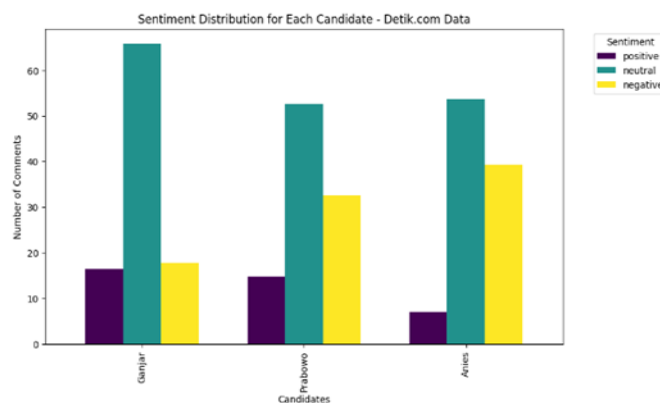
Setelah menyelesaikan perhitungan persentase masing-masing kandidat, kemudian visualisasikan hasil sentimen dalam bentuk diagram batang untuk membandingkan dan memberikan gambaran yang jelas tentang perbandingan sentimen dari ketiga akun tersebut. Di bawah ini adalah gambar visualisasi dalam bentuk diagram batang untuk setiap akun:



Gambar 3. Diagram Dataset Kompascom



Gambar 4. Diagram Kumpulan Data Najwashihab



Gambar 5. Diagram Dataset Detikcom

Penelitian ini berhasil dilakukan dengan cukup baik dengan menggunakan model klasifikasi analisis sentimen menggunakan algoritma Naïve Bayes. Proses pengikisan yang dilakukan berhasil mengumpulkan 39.015 dataset komentar tentang pasangan capres 2024 yang diperoleh dari tiga akun yang membahas pasangan capres yaitu akun Najwashihab, Detikcom, dan Kompascom sebagai bahan perbandingan. Hasil pada akun Najwa adalah 70,38% akurasi, pada akun Kompascom nilai akurasi yang dihasilkan adalah 66,95%, dan pada akun Detikcom nilai akurasi yang dihasilkan adalah 66,69%. Untuk itu, dapat disimpulkan bahwa nilai akurasi tertinggi diperoleh dari akun Najwashihab dengan akurasi 70,38%. Metode ini didukung oleh penelitian Imran *dkk.* [28], yang menggunakan metode Naïve Bayes dan SVM sebagai bahan perbandingan dan hasilnya menunjukkan bahwa kinerja yang

diperoleh dengan menggunakan metode Naïve Bayes terbukti lebih baik dalam klasifikasi teks dibandingkan dengan metode SVM. Penelitian lain dari Misrun dkk juga menunjukkan hasil yang cukup baik dalam penggunaan metode Naïve Bayes dimana hasil akurasi yang di dapat yakni sebesar 78% dengan data yang digunakan melalui komentar youtube.

#### 4. KESIMPULAN

Berdasarkan hasil pengujian, dapat disimpulkan bahwa menggunakan algoritma Naïve Bayes yang digunakan sebagai metode klasifikasi dalam analisis sentimen berbasis komentar di Instagram memiliki tingkat akurasi yang cukup baik dengan akurasi rata-rata lebih dari 60%. Dapat disimpulkan juga bahwa hasil sentimen netral lebih dominan daripada sentimen positif dan negatif karena jumlah komentar publik kurang signifikan apakah komentar tersebut termasuk dalam kategori positif atau negatif. Untuk penelitian lebih lanjut, ada beberapa komponen yang perlu ditingkatkan, seperti menambahkan lebih banyak dataset pelatihan untuk menyamakan jumlah data untuk setiap kelas.

#### DAFTAR PUSTAKA

- [1] A. D. Akmal, I. Permana, H. Fajri, and Y. Yuliarti, "Opini Masyarakat Twitter terhadap Kandidat Bakal Calon Presiden Republik Indonesia Tahun 2024," *J. Manaj. dan Ilmu Adm. Publik*, vol. 4, no. 4, pp. 292–300, 2022, doi: 10.24036/jmiap.v4i4.160.
- [2] H. Permana, Y. Herry Chrisnanto, and H. Ashaury, "Analisis Sentimen Terhadap Bakal Calon Presiden 2024 Dengan Algoritma Multinomial Naïve Bayes Dan Oversampling Smote," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 5, pp. 3257–3264, 2024, doi: 10.36040/jati.v7i5.7309.
- [3] I. P. D. W. Darmawan, G. A. Pradnyana, and I. B. N. Pascima, "Optimasi Parameter Support Vector Machine Dengan Algoritma Genetika Untuk Analisis Sentimen Pada Media Sosial Instagram," *SINTECH (Science Inf. Technol. J.)*, vol. 6, no. 1, pp. 58–67, 2023, doi: 10.31598/sintechjournal.v6i1.1245.
- [4] A. Hanana, "Trend Postingan Selebrasi sebagai Bentuk Eksistensi Diri Generasi Muda di Sosial Media Instagram," *AL MUNIR J. Komun. dan Penyiaran Islam*, vol. 13, no. 01, pp. 87–107, 2022.
- [5] S. A. Aaputra, Didi Rosiyadi, Windu Gata, and Syepny Maulana Husain, "Sentiment Analysis Analysis of E-Wallet Sentiments on Google Play Using the Naive Bayes Algorithm Based on Particle Swarm Optimization," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 3, pp. 377–382, 2019, doi: 10.29207/resti.v3i3.1118.
- [6] T. D. Putra, E. Utami, and M. P. Kurniawan, "Analisis Sentimen Pemilu 2024 dengan Naive Bayes Berbasis Particle Swarm Optimization (PSO)," *Explore*, vol. 13, no. 1, pp. 1–5, 2023, doi: 10.35200/ex.v11i2.13.
- [7] R. Safri Irawansyah, L. A. Syamsul Irfan A, and G. Wahyu Wiriasto, "Analisis Sentimen Terhadap Program Merdeka Belajar-Kampus Merdeka (Mbk) Pada Twitter Menggunakan Algoritma Naive Bayes Classifier (Nbc)," *J. Teknol. Informasi, Komput. dan Apl.*, vol. 5, no. 2, pp. 1–8, 2023.
- [8] Normah, B. Rifai, S. Vambudi, and R. Maulana, "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE," *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.
- [9] E. B. Susanto, P. A. Christianto, M. R. Maulana, and S. W. Binabar, "Analisis Kinerja Algoritma Naïve Bayes pada Dataset Sentimen Masyarakat Aplikasi NEWSAKPOLESamsat Jawa Tengah," *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 3, no. 3, pp. 234–241, 2022.
- [10] K. Anwar, "Analisa sentimen Pengguna Instagram Di Indonesia Pada Review Smartphone Menggunakan Naive Bayes," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 4, pp. 148–155, 2022, doi: 10.30865/klik.v2i4.315.
- [11] M. S. Adhi, M. Z. Nafan, and E. Usada, "Pengaruh Semantic Expansion pada Naïve Bayes Classifier untuk Analisis Sentimen Tokoh Masyarakat," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 2, pp. 141–147, 2019, doi: 10.29207/resti.v3i2.901.
- [12] R. D. Septiana, A. B. Susanto, and T. Tukiyyat, "Analisis Sentimen Vaksinasi Covid-19 Pada Twitter Menggunakan Naive Bayes Classifier Dengan Feature Selection Chi-Squared Statistic dan Particle Swarm Optimization," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 5, no. 1, pp. 49–56, 2021, doi: 10.47970/siskom-kb.v5i1.228.
- [13] Chely Aulia Misrun, E. Haerani, M. Fikry, and E. Budianita, "Analisis sentimen komentar youtube terhadap Anies Baswedan sebagai bakal calon presiden 2024 menggunakan metode naive bayes classifier," *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 4, no. 1, pp. 207–215, 2023, doi: 10.37859/coscitech.v4i1.4790.
- [14] D. S. Simatupang and S. Nursinta, "Analisis sentimen terhadap berita hoaks lowongan kerja dengan metode naive bayes," *J. Comput. Sci. Inf. Technol.*, vol. 5, no. 2, pp. 474–482, 2024.
- [15] R. N. Mauliza and Y. R. Sipayung, "Penerapan Text Mining Dalam Menganalisis Pendapat Masyarakat Terhadap Pemilu 2024 Pada Media Sosial X Menggunakan Metode Naive Bayes," vol. 9, no. 1, pp. 1–16, 2024.
- [16] B. S. Prakoso, D. Rosiyadi, H. S. Utama, and D. Aridarma, "Klasifikasi Berita Menggunakan Algoritma Naive Bayes Classifier Dengan Seleksi Fitur Dan Boosting," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 2, pp. 227–232, 2019, doi: 10.29207/resti.v3i2.1042.
- [17] S. M. Permataning Tyas, B. S. Rintyarna, and W. Suharso, "The Impact of Feature Extraction to Naïve Bayes Based Sentiment Analysis on Review Dataset of Indihome Services," *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 13, no. 1, pp. 1–10, 2022, doi: 10.31849/digitalzone.v13i1.9158.
- [18] Y. Oktaviani, I. G. Putu, W. Wedashwara, and A. Zubaidi, "Klasifikasi Teks Ulasan Pada Web Tripadvisor Tentang Wisata Alam Pulau Lombok Menggunakan Metode Naive Bayes Classifier," vol. 4, no. 2, pp. 253–262, 2022.
- [19] F. Pramono, Didi Rosiyadi, and Windu Gata, "Integrasi N-gram, Information Gain, Particle Swarm Optimization di Naïve Bayes untuk Optimasi Sentimen Google Classroom," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 3, pp. 383–388, 2019, doi: 10.29207/resti.v3i3.1119.
- [20] R. R. Juandri, R. W. P. Pamungkas, and A. Fathurrozi, "Rancangan Aplikasi Portal Media Sosial Sebagai Analisis Sentimen Publik Menggunakan Machine Learning Dengan Algoritma Naive Bayes," *J. Students' Res. Comput. Sci.*, vol. 3, no. 2, pp. 217–228, 2022, doi: 10.31599/jsrcs.v3i2.1760.
- [21] N. F. Hasan and H. Wonda, "Sentiment Analysis of Public Opinion Regarding Papuan Local Languages Condition Using Data Science Approach," *J. Teknol. Inf. dan Komun.*, vol. 13, no. 2, pp. 125–139, 2022.

- [22] S. Puad, G. Garno, and A. Susilo Yuda Irawan, "Analisis Sentimen Masyarakat Pada Twitter Terhadap Pemilihan Umum 2024 Menggunakan Algoritma Naïve Bayes," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 3, pp. 1560–1566, 2023, doi: 10.36040/jati.v7i3.6920.
- [23] M. Qorib, T. Oladunni, M. Denis, E. Ososanya, and P. Cotae, "Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset," *Expert Syst. Appl.*, vol. 212, no. August 2022, p. 118715, 2023, doi: 10.1016/j.eswa.2022.118715.
- [24] S. Hikmawan, A. Pardamean, and S. N. Khasanah, "Sentimen Analisis Publik Terhadap Joko Widodo terhadap wabah Covid-19 menggunakan Metode Machine Learning," *J. Kaji. Ilm.*, vol. 20, no. 2, pp. 167–176, 2020, doi: 10.31599/jki.v20i2.117.
- [25] O. P. Zusrotun, A. C. Murti, and R. Fiati, "Analisis Sentimen Terhadap Belajar Online pada Media Sosial Twitter Menggunakan Algoritma Naive Bayes," *J. Nas. Pendidik. Tek. Inform.*, vol. 11, no. 3, pp. 310–319, 2022, doi: 10.23887/janapati.v11i3.49160.
- [26] N. L. W. S. R. Ginantra, C. P. Yanti, G. D. Prasetya, I. B. G. Sarasvananda, and I. K. A. G. Wiguna, "Analisis Sentimen Ulasan Villa di Ubud Menggunakan Metode Naive Bayes, Decision Tree, dan K-NN," *J. Nas. Pendidik. Tek. Inform.*, vol. 11, no. 3, pp. 205–215, 2022, doi: 10.23887/janapati.v11i3.49450.
- [27] N. Y. Utama, K. M. Lhaksana, and I. Kurniawan, "Text Analysis of Applicants for Personality Classification Using Multinomial Naïve Bayes and Decision Tree," *J. Infotel*, vol. 12, no. 3, pp. 72–81, 2020, doi: 10.20895/infotel.v12i3.505.
- [28] B. Imran, M. N. Karim, and N. I. Ningsih, "Klasifikasi Berita Hoax Terkait Pemilihan Umum Presiden Republik Indonesia Tahun 2024 Menggunakan Naïve Bayes Dan Svm," *Din. Rekayasa*, vol. 20, no. 1, pp. 1–9, 2024, doi: 10.20884/1.dinarek.2024.20.1.27.