

## Klasifikasi Sentimen Presepsi Masyarakat di Instagram Terhadap Paslon Pilpres 2024 Menggunakan Naïve Bayes Classifier (NBC)

Lionita Asa Akbar<sup>1</sup>, Elin Haerani<sup>2\*</sup>, Fadhilah Syafria<sup>3</sup>, Alwis Nazir<sup>4</sup>, Elvia Budianita<sup>5</sup>  
<sup>1,2,3,4,5</sup>Teknik Informatika / Sains dan Teknologi, Universitas Sultan Syarif Kasim Riau  
\*email: [elin.haerani@uin-suska.ac.id](mailto:elin.haerani@uin-suska.ac.id)

DOI: <https://doi.org/10.31603/komtika.v8i1.11293>

Received: 26-04-2024, Revised: 21-06-2024, Accepted: 24-06-2024

### ABSTRACT

*The 2024 presidential election has attracted considerable attention as it has become a controversial issue among the public. Various positive and negative opinions generated can potentially turn into rumors. One of the means used by the public to express their opinions is the social media platform Instagram. Data on public opinions on Instagram can be processed into valuable information through sentiment classification. This research conducted sentiment classification on public perceptions towards the 2024 presidential candidates using a naïve Bayes classifier. The study utilized a dataset consisting of 1000 comments. These comments were collected from several posts on the social media platform Instagram discussing the presidential and vice-presidential candidates. The comments were manually labeled by an expert who is a lecturer in the Indonesian language. Classification was carried out after preprocessing and weighting TF-IDF stages. Based on the research findings, the naïve Bayes classifier method showed an accuracy of 82% and an F1-Score of 83.93% obtained from a 90%:10% split of training and testing data. These results indicate that the naïve Bayes classifier method is effective in classifying the sentiments of the public on Instagram towards the 2024 presidential candidates.*

**Keywords:** sentiment classification, presidential election, Instagram, naïve Bayes.

### ABSTRAK

Pilpres 2024 cukup menarik perhatian karena menjadi isu kontroversi di kalangan masyarakat. Berbagai opini positif dan negatif yang dihasilkan dapat menjadi sebuah rumor. Salah satu sarana yang digunakan masyarakat untuk menyampaikan opininya adalah media sosial Instagram. Data opini masyarakat di Instagram dapat diolah menjadi suatu informasi bermanfaat melalui klasifikasi sentiment. Pada penelitian ini dilakukan klasifikasi sentiment presepsi masyarakat terhadap paslon pilpres 2024 menggunakan *naïve bayes classifier*. Penelitian ini menggunakan dataset yang terdiri dari 1000 komentar. Komentar tersebut dikumpulkan dari beberapa postingan di media sosial Instagram yang membahas tentang pasangan calon presiden dan wakil presiden. Data komentar dilabeli secara manual oleh seorang pakar yang berprofesi sebagai dosen Bahasa Indonesia. Klasifikasi dilakukan setelah tahap *preprocessing* dan pembobotan TF-IDF. Berdasarkan hasil penelitian, metode *naïve bayes classifier* menunjukkan akurasi sebesar 82% dan *F1-Score* 83.93% yang diperoleh dari pembagian data latih dan uji sebesar 90%:10%. Hasil tersebut menunjukkan bahwa metode *naïve bayes classifier* efektif dalam mengklasifikasikan sentimen masyarakat di Instagram terhadap pasangan calon presiden 2024.

**Keywords:** klasifikasi sentimen, pilpres, instagram, *naïve bayes*.

### PENDAHULUAN

Pemilihan Umum merupakan suatu tradisi bagi negara yang menganut demokrasi yang digunakan untuk sarana rakyat dalam memberikan kedaulatannya untuk negara. Tahun 2024 akan menjadi tahun untuk diadakannya pemilu serentak pemilihan presiden dan wakil

presiden yang sering disebut Pilpres. Presiden merupakan kepala negara yang akan memimpin suatu negara yang dipilih oleh rakyat. Hasil yang di dapat dari pilpres tersebut yang akan menentukan arah kebijakan dan kemajuan suatu negara nantinya. Oleh karena itu, memahami persepsi, pandangan, sikap, dan reaksi masyarakat terkait dengan calon presiden dan wakil presiden menjadi sangat esensial [1]. Di era digital ini, sudah banyak media sosial yang dapat diakses oleh Masyarakat untuk melihat pemberitaan mengenai pilpres serta memberi komentar dan pendapat mereka terkait bakal calon presiden 2024.

Salah satu bakal calon presiden Indonesia yang mencalonkan diri adalah pasangan Prabowo-Gibran. Prabowo sebelumnya sudah pernah mencalonkan diri sebagai presiden sebanyak tiga kali, dari hal tersebut banyak pemberitaan mengenai Prabowo yang mencalonkan diri lagi menjadi presiden di media sosial, ditambah dengan keputusan Prabowo yang memilih Gibran sebagai wakilnya juga menuai banyak beragam reaksi masyarakat. Ada beberapa artikel media berita yang melampirkan hasil wawancara langsung kepada masyarakat terkait respon terhadap Gibran yang mencalonkan diri sebagai wakil presiden. Artikel dari suara.com menyebutkan masyarakat yang kurang setuju terhadap Gibran yang menurut masyarakat terlalu muda dan kurangnya pengalaman untuk mencalonkan diri sebagai wakil presiden dan juga unsur yang bertentangan dengan prinsip demokrasi dimana Gibran yang merupakan anak presiden[2]. Lalu artikel dari KompasTV yang menyebutkan mengenai putusan MK terkait usia capres dan cawapres terus menuai dukungan dan kritikan [3] .

Instagram adalah salah satu jejaring sosial yang paling banyak digunakan oleh orang-orang untuk mengungkapkan pendapat dan pemikiran mereka tentang isu-isu. Instagram merupakan tempat untuk mengakses sebuah berita tentang politik, entertainment, olahraga, ataupun pengalaman diri sendiri. Pada laman dari akun media berita, banyak yang membahas terkait Prabowo-Gibran sebagai calon kandidat presiden dan wakil presiden Indonesia. Dalam postingan berita di Instagram tersebut, pada fitur kolom komentar banyak Masyarakat yang memberikan opini dan presepsinya terkait berita tersebut. Opini dan presepsi yang terdapat pada komentar di Instagram tersebut menarik untuk diklasifikasikan dalam sentimen positif atau sentimen negatif.

Pada penelitian metode yang digunakan untuk klasifikasi sentimen ialah *Naïve Bayes Classifier*. Ini merupakan satu teknik untuk mengkategorikan data dengan menggunakan prinsip probabilitas sederhana yang menerapkan teorema bayes dengan asumsi independensi yang signifikan[4]. Metode ini cocok untuk berbagai jenis kumpulan data dengan kemampuan klasifikasi yang cepat dan tingkat akurasi yang tinggi[5].

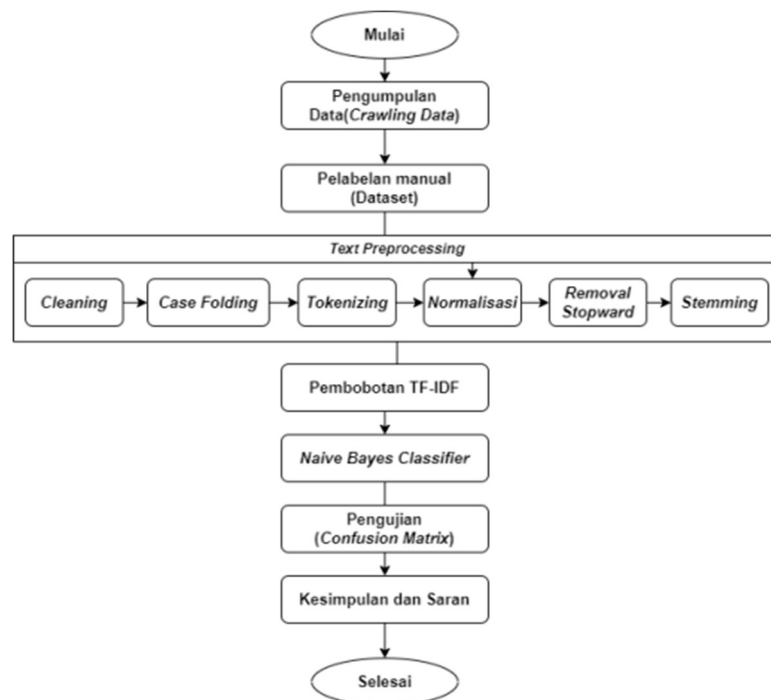
Terdapat beberapa penelitian sebelumnya tentang klasifikasi sentimen menggunakan metode *Naïve Bayes Classifier* dengan masalah dan kasus yang sama dan berbeda. Seperti penelitian dengan kasus terkait opini publik pada Prabowo Subianto bakal calon presiden 2024 di Twitter dengan data yang terkumpul 1050 data positif dan 1050 negatif mencapai hasil akurasi tertinggi 89% dengan nilai *precision* 89,7%, 88,6% *recall*, dan 88,9% *f1-score* pada uji percobaan 90:10[6]. Selanjutnya terkait kasus sentimen publik terhadap pemilu 2019 di twitter mendapatkan hasil akurasi mencapai 62% dengan nilai *recall* sebesar 45% dan presisi mencapai 41%. Selanjutnya terkait kasus bakal calon presiden 2024 *dataset* pertama 73,68 akurasi, *dataset* kedua 71,43 , *dataset* ketiga 60% , dan *dataset* terakhir 62,5%[7]. Selanjutnya terkait kasus berbeda tetapi dengan menggunakan metode *naïve bayes* dan KNN

tentang opini masyarakat terhadap vaksin Covid-19 di twitter mendapatkan hasil bahwa metode *naïve bayes* lebih unggul dari pada KNN dengan nilai akurasi sebesar 90%[8].

Berdasarkan pertimbangan permasalahan yang telah diuraikan, pendekatan yang akan digunakan dalam penelitian ini menggunakan metode *Naïve Bayes Classifier* untuk mengelompokkan persepsi masyarakat terhadap pasangan bakal calon presiden dan wakil presiden Prabowo-Gibran dalam Pemilihan Presiden 2024. Penelitian ini akan memanfaatkan komentar yang disampaikan oleh masyarakat dalam kolom komentar Instagram, dengan tujuan mengklasifikasikan mereka ke dalam kategori pro (positif) atau kontra (negatif).

## METODE

Penelitian ini mencakup enam tahap yang disajikan pada gambar 1 yaitu tahap pengumpulan data, pelabelan manual *dataset*, *text preprocessing*, pembobotan kata, klasifikasi *Naïve Bayes Classifier*, dan pengujian. Alur penelitian disajikan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Penelitian studi Penelitian ini dimulai dengan melaksanakan pengumpulan dataset dari komentar instagram yang diambil menggunakan *web scraper* di google chrome yang bernama *Data Miner*. Data yang digunakan sebanyak 1000 data berbahasa Indonesia pada postingan instagram yang membahas tentang calon presiden dan wakil presiden Prabowo-Gibran. Tahap selanjutnya adalah pemberian label pada data yang telah didapat dengan menetapkan kategori label sentimen positif dan negatif. Tahap pemberian label ini dibuat secara non-otomatis dengan membaca satu-persatu data komentar untuk diberi label positif atau negatif oleh seorang pakar yang berprofesi sebagai dosen Bahasa Indonesia.

Langkah berikutnya adalah tahap *Pre-processing*, yang melibatkan pemilihan data untuk memastikan bahwa data yang digunakan menjadi lebih terstruktur. *Text preprocessing* berfungsi untuk mengurangi *noise* pada *dataset*[9]. *Pre-processing* terdiri dari enam tahap sebagai berikut:

1. *Data Cleaning*, adalah kalimat-kalimat dalam *dataset* dibersihkan dari segala sesuatu yang dapat mempengaruhi hasil analisis, seperti kata-kata yang memiliki dua karakter berulang atau lebih, *link*, nama pengguna (@nama pengguna), *hashtag*(#), angka, simbol [10] dll.
2. *Case folding*, langkah agar mengubah semua huruf dalam set data komentar untuk menjadi huruf kecil [11].
3. *Tokenization*, berguna membagi semua kata menjadi sebuah token[12].
4. Normalisasi merupakan tahap mengubah kata-kata yang dalam ejaan tidak baku menjadi ejaan yang benar seperti “dmn” menjadi “dimana”.
5. *Removal Stopward*, merupakan tahap penghapusan kata yang tidak penting yang sering muncul pada dokumen, misal “dan”, “atau”, “sebuah”.
6. *Stemming*, adalah langkah mengalihkan kata-kata yang berasal dari *removal stopward* ke bentuk dasarnya dengan cara menghilangkan imbuhan pada kata-kata dalam dokumen, misal berteman menjadi teman[4].

Setelah melakukan tahap *Pre-processing*, kemudian masuk ke tahap *Term Frequency-Inverse Document Frequency*. TF-IDF adalah pendekatan untuk menyediakan nilai bobot pada keterhubungan antara sebuah kata (*term*) dengan sebuah dokumen [13]. Metode ini mengintegrasikan dua prinsip perhitungan bobot, yakni berapa kali sebuah kata ditemukan dalam dokumen tertentu dan seberapa jarang kata itu muncul di semua dokumen. Frekuensi kemunculan kata dalam dokumen tersebut mencerminkan signifikansinya dalam dokumen tersebut[14]. *Term Frequency* (TF) mengindikasikan berapa kali sebuah kata muncul dalam dokumen tertentu. Meskipun begitu, nilai TF tidak mengungkapkan seberapa keberartian kata tersebut dalam konteks dokumen tersebut. Oleh karena itu, untuk memberi penekanan pada kata yang jarang muncul di seluruh dokumen, digunakan pendekatan lain yang disebut *Inverse Document Frequency* (IDF) yang memberikan bobot pada kata yang jarang muncul di seluruh dokumen[15]. Berikut ini rumus untuk menghitung TF-IDF:

$$TF = \frac{\text{Jumlah kemunculan kata dalam dokumen}}{\text{Total jumlah kata dalam dokumen}} \quad (1)$$

$$IDF = \log\left(\frac{\text{Count}(docs)}{\sum_{i=0}^n \text{count}(word, docs)}\right) \quad (2)$$

$$TF - IDF = TF(word) * IDF(word) \quad (3)$$

Keterangan:

$\text{Count}(docs)$  = Jumlah semua dokumen  
 $\sum_{i=0}^n \text{count}(word, docs)$  = Jumlah dari dokumen yang memiliki kata tersebut atau DF  
 $TF(word)$  = Jumlah *term frequency*  
 $IDF(word)$  = Jumlah *inverse document frequency*

Tahap selanjutnya adalah klasifikasi *Naïve Bayes Classifier* (NBC). NBC merupakan algoritma klasifikasi analisis yang dapat menyajikan estimasi probabilitas bahwa data tertentu akan termasuk dalam kelas tertentu, berdasarkan perhitungan probabilitasnya[16]. Secara sederhana, *Naïve Bayes Classifier* beranggapan bahwa kehadiran fitur khusus di suatu kelas

tidak terhubung dengan keberadaan fitur lain yang secara inheren saling bergantung satu sama[17]. Dalam penelitian ini, digunakan Pengklasifikasi *Naive Bayes* tipe *Multinomial*, yang beroperasi berdasarkan konsep frekuensi kata, yaitu seberapa sering sebuah kata muncul dalam dokumen. Model *MultinomialNB* punya kapasitas untuk mengelompokkan data yang tidak dapat diwakili secara numerik[18]. Berikut ini rumus persamaan *Naive Bayes Classifier*:

$$P(A|B) = \frac{P(B|A)*P(A)}{P(B)} \quad (4)$$

*Naive Bayes Classifier* dapat disebut sebagai *naive bayes multinomial* yang merupakan model yang tepat dalam mengklasifikasi teks atau dokumen dapat dilihat pada persamaan (5).

$$V_{MAP} = \text{argmax } P(v_j | a_1, a_2, a_3 \dots a_n) \quad (5)$$

Rumus pengujian sentiment dapat dilihat pada persamaan (6).

$$V_{MAP} \text{ argmax}_{v_j \in V} P(a_i v_j) \times \prod P(v_j) \quad (6)$$

Rumus hitung data uji dapat dilihat pada persamaan (7) dan (8).

$$P(v_j) = \frac{|Dok_i|}{\text{training}} \quad (7)$$

$$P(a_i|v_j) = \frac{n_i+1}{n+|\text{kosakata}|} \quad (8)$$

Keterangan:

$P(v_j)$  = probabilitas masing-masing dokumen dalam sebuah himpunan dokumen

$P(a_i|v_j)$  = probabilitas kenampakan kata  $a_i$  dalam suatu dokumen dengan kategori kelas  $v_j$

Setelah melakukan pengklasifikasian, tahap selanjutnya adalah melakukan pengujian. Data akan terbagi menjadi data untuk pelatihan (*training data*) dan untuk pengujian (*testing data*) agar melatih algoritma NBC untuk membangun model klasifikasi. Evaluasi performa klasifikasi dilakukan menggunakan *confusion matrix* yang dapat dilihat pada Tabel 1.

Tabel 1. *Confusion Matrix*

	<i>Actual Positive</i>	<i>Actual Negative</i>
<i>Predicted Positive</i>	TP	FP
<i>Predicted Negative</i>	FN	TN

*Confusion matrik* adalah cara pengkajian model klasifikasi dalam pengujian untuk mengestimasi objek yang benar dan salah[19] terkait berapa jumlah *true positive*(TP), *false positive*(FP), *true negative*(TN), *false negative*(FN). Berdasarkan pengujian yang dilakukan, akan diketahui berapa jumlah *accuracy*, *precision*, *recall*, dan *f1-score* yang lebih tinggi dari pembagian data. Rumus untuk menghitung *accuracy*, *precision*, *recall*, dan *f1-score* adalah sebagai berikut:

$$accuracy = \frac{TP + TN}{TP + PN + FP + FN} * 100\% \quad (9)$$

$$recall = \frac{TP}{TP + FN} * 100\% \quad (10)$$

$$precision = \frac{TP}{TP + FP} * 100\% \quad (11)$$

$$F1 - Score = \frac{recall * precision}{recall + precision} \quad (12)$$

## HASIL DAN PEMBAHASAN

### Pengumpulan data

Tahap pengumpulan data penelitian ini dilakukan menggunakan alat *Dataminer*. Data dikumpulkan sebanyak 1000 data komentar yang didapatkan dari beberapa akun instagram, yaitu @indozone.id, @tvonenews, @folkative, @kompas.com, @golkar.indonesia. komentar-komentar tersebut dikumpulkan dari postingan yang membahas tentang bakal calon presiden dan wakil presiden Prabowo-Gibran. Data disimpan menggunakan format excel lalu untuk proses selanjutnya disimpan dengan format csv.

### Pelabelan Manual Dataset

Tahap pelabelan diproses secara manual dan divalidasi oleh seorang pakar yang berprofesi sebagai dosen bahasa Indonesia. 1000 data tersebut akan diberi label positif dan negatif. Hasil pelabelan data disimpan dalam file berformat csv yang dapat disajikan seperti pada Gambar 2.

	Username IG	Komentar	LABEL
0	danielssyah	Kalo mau anak Presiden knapa ga milih AHY aja ...	negatif
1	__eleanor__979	Pak kayak gk ada yg lain wakilny,.. padahal bp...	negatif
2	malik_bonjovi	Dengan begini saya yakin 1000000000000% bahwa ...	negatif
3	fitzrzkyr	Kayanya gagal lagi ketiga kalinya pak Prabowo 😞😞	negatif
4	hery.bian	KALIAN PADA GA SADAR APA YAA MASIH MAU AJA DIB...	negatif
...	...	...	...
995	elodiestella4\r	Prabowo pilihan terbaik untuk rakyat R I	positif
996	vernanda_della	Pak, saya tadi sudah ke TPS buat nyoblos bapak...	positif
997	muhammad_wawan_derzet	Dari awal pak prabowo mencalonkan presiden say...	positif
998	lindaaldino78	Masya Allah...sehat selalu bapak prabowo,, bny...	positif
999	ellisusanti77	Yakin pak Prabowo menang d pemilu kali ini,buk...	positif

1000 rows x 4 columns

Gambar 2. Hasil Pelabelan Manual

Pada gambar 2 menunjukkan data komentar yang telah diberi label positif dan negatif dan telah di validasi oleh seorang pakar yaitu dosen bahasa Indonesia.

### Text Preprocessing

Tahap *text preprocessing* adalah tahap pembersihan data yang mencakup *cleaning*, *case folding*, *tokenizing*, normalisasi, *removal stopward*, *stemming*. Tabel menunjukkan hasil *text preprocessing*.

Tabel 2. Text Preprocessing

Proses	Hasil
Data Mentah	Ijin Pak, saya mengundurkan diri untuk memilih Njenengan @prabowo
Cleaning	Ijin Pak saya mengundurkan diri untuk memilih Njenengan
Case Folding	ijin pak saya mengundurkan diri untuk memilih njenengan
Tokenizing	[izin,pak,saya,mengundurkan,diri,untuk,memilih,njenengan]
Normalisasi	[izin,pak,saya,mengundurkan,diri,untuk,memilih,kamu]
Removal Stopward	[izin,mengundurkan,memilih]
Stemming	[izin,undur,pilih]

Tahap *text preprocessing* menghasilkan data *text* berdasarkan frekuensi kata. Kata yang berfrekuensi tinggi akan divisualisasikan ke dalam bentuk *wordcloud* positif dan negatif seperti pada Gambar 3.



Gambar 3 wordcloud Negatif dan Positif

Pada gambar 3 menunjukkan wordcloud positif dan negatif, wordcloud tersebut menampilkan visual kata yang sering muncul pada setiap sentiment setelah melakukan semua tahap *text preprocessing*. *Text preprocessing* sangat penting digunakan untuk memastikan data yang dianalisis bersih dan konsisten. tahap *cleaning* yang merupakan tahap pembersihan data pada kalimat yang terdapat simbol, angka, nama pengguna (@nama pengguna), tahap *case folding* mengubah kata menjadi huruf kecil untuk menjaga konsistensi data, tahap *tokenization* memecah teks menjadi token, tahap *normalisasi* mengubah kata menjadi ejaan yang lebih benar, tahap *removal stopward* menghapus kata yang tidak memiliki nilai informasi signifikan seperti “di”, “dan”, “yang” untuk fokus pada kata-kata yang lebih penting, dan terakhir tahap *stemming* mengubah kata-kata menjadi bentuk dasar untuk menggabungkan variasi kata yang memiliki makna sama. Gambar 3 wordcloud negatif memperlihatkan kata-kata seperti "prabowo," "pilih," "gibran," "presiden," dan "negara" yang sering muncul dalam konteks sentimen negatif. Ini menunjukkan bahwa dalam teks yang dianalisis, topik-topik ini sering dikaitkan dengan perasaan negatif. Sebaliknya, pada wordcloud positif, kata-kata yang sama juga muncul namun dalam konteks yang lebih mendukung atau optimis.

### Pembobotan TF-IDF

Tahap pembobotan TF-IDF dilakukan menggunakan alat *TfidfVectorizer* di library python *scikit-learn*. Gambar 4 menunjukkan hasil pembobotan TF-IDF untuk masing-masing kata dalam teks.





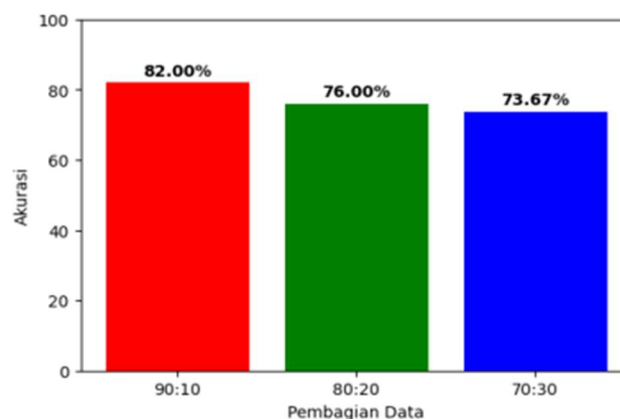


47 komentar *sentiment* yang diprediksi label juga positif sehingga menghasilkan TP (*True Positive*), pada *true label* 9 komentar *sentiment* yang diprediksi label positif sehingga menjadi FP(*False Positive*), dan pada *true label* 35 komentar *sentiment* yang diprediksi label negatif sehingga menghasilkan TN(*True Negative*). Rasio 80:20 memperlihatkan peningkatan jumlah data uji yang menghasilkan lebih banyak TP (*True Positive*) dan TN (*True Negatives*), tetapi juga menunjukkan peningkatan jumlah kesalahan prediksi. Akhirnya, rasio 70:30 menampilkan jumlah data uji yang lebih besar, memberikan lebih banyak wawasan tentang kemampuan generalisasi model, meskipun dengan peningkatan dalam jumlah kesalahan prediksi. Evaluasi menggunakan *confusion matrix* membantu dalam memahami kekuatan dan kelemahan model, serta bagaimana perubahan dalam pembagian data dapat mempengaruhi performa secara keseluruhan. Hasil skenario pembagian data dengan menggunakan *confusion matrix* dapat dilihat dari Tabel 3.

Tabel 3. *Accuracy, Precision, Recall, dan F1-Score* berdasarkan *Confusion Matrix*

Split data	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
90%:10%	82.00%	83.93%	83.93%	83.93%
80%:20%	76.00%	73.50%	83.50%	78.18%
70%:30%	73.67%	69.84%	85.71%	76.97%

Pada tabel 3 berdasarkan *confusion matrix* hasil tertinggi terdapat pada pembagian data 90%:10% yang mana pada model mencapai *accuracy* 82.00% yang menunjukkan bahwa 82% dari semua prediksi adalah benar. *Precision* sebesar 83.93% yang menunjukkan semua prediksi positif, 83.93% adalah benar. *Recall* sebesar 83.93% menunjukkan bahwa semua *instance* positif yang sebenarnya, 83.93% berhasil diklasifikasikan dengan benar oleh model. Nilai *F1-Score* sebesar 83.93% memberikan keseimbangan antara *precision* dan *recall*. Grafik hasil perbandingan data dan akurasi disajikan seperti pada Gambar 6.



Gambar 6 Grafik Hasil Perbandingan Data

Pada gambar 6 menunjukkan model dilatih dengan 90% data dan diuji dengan 10% data mendapatkan akurasi yang dicapai adalah 82.00%, model dilatih dengan 80% data dan diuji dengan 20% data akurasi yang dicapai adalah 76.00%, dan model dilatih dengan 70% data dan diuji dengan 30% data akurasi yang dicapai adalah 73.67%. Grafik ini menunjukkan bahwa semakin banyak data yang digunakan untuk pelatihan (90:10), semakin tinggi akurasi

model. Sebaliknya, ketika lebih banyak data digunakan untuk pengujian (70:30), akurasi model menurun.

## KESIMPULAN

Penelitian ini menggunakan *dataset* yang terdiri dari 1000 komentar. Komentar tersebut dikumpulkan dari beberapa postingan di media sosial Instagram yang membahas tentang pasangan calon presiden dan wakil presiden Prabowo-Gibran. Data komentar dilabeli secara manual oleh seorang pakar yang berprofesi sebagai dosen Bahasa Indonesia. Klasifikasi dilakukan setelah tahap preprocessing dan TF-IDF. Berdasarkan hasil penelitian, metode *naïve bayes classifier* menunjukkan akurasi sebesar 82% dan *F1-Score* 83.93% yang diperoleh dari pembagian data latih dan uji sebesar 90%:10%. Hasil tersebut mengindikasikan bahwa metode *naïve bayes classifier* efektif dalam mengklasifikasikan sentimen masyarakat di Instagram terhadap pasangan calon presiden 2024 Prabowo-Gibran.

## DAFTAR PUSTAKA

- [1] O. Manullang and C. Prianto, “Analisis Sentimen dalam Memprediksi Hasil Pemilu Presiden dan Wakil Presiden : Systematic Literature Review,” vol. 04, no. 02, pp. 104–113, 2023.
- [2] M. I. Baktora, “Pro dan Kontra Gibran Rakabuming Raka jadi Cawapres di Pemilu 2024, Bagaimana Pandangan Kaum Muda, Yakin Banyak Dipilih?” Accessed: Mar. 04, 2024. [Online]. Available: <https://www.suara.com/kotaksuara/2023/11/24/170000/pro-dan-kontra-gibran-rakabuming-raka-jadi-cawapres-di-pemilu-2024-bagaimana-pandangan-kaum-muda-yakin-banyak-dipilih>
- [3] K. Makasar, “Pro Kontra Putusan MK, Gerindra: Ada Operasi Senyap Jegal Gibran Maju di Pilpres,” KompasTV. Accessed: Dec. 07, 2023. [Online]. Available: <https://www.kompas.tv/regional/458338/pro-kontra-putusan-mk-gerindra-ada-operasi-senyap-jegal-gibran-maju-di-pilpres>
- [4] D. Normawati and S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, 2021.
- [5] S. Juanita, “Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes,” *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 552, 2020, doi: 10.30865/mib.v4i3.2140.
- [6] A. Halim, Y. Yusra, M. Fikry, M. Irsyad, and E. Budianita, “Klasifikasi Sentimen Masyarakat Di Twitter Terhadap Prabowo Subianto Sebagai Bakal Calon Presiden 2024 Menggunakan M-KNN,” *J. Inf. Syst. Res.*, vol. 5, no. 1, pp. 202–212, 2023, doi: 10.47065/josh.v5i1.4054.
- [7] M. R. Fais Sya' bani, U. Enri, and T. N. Padilah, “Analisis Sentimen Terhadap Bakal Calon Presiden 2024 Dengan Algoritme Naïve Bayes,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 2, p. 265, 2022, doi: 10.30865/jurikom.v9i2.3989.
- [8] M. S. Sulasno and H. Amalia, “Analisis Sentimen Opini Masyarakat Indonesia Terhadap Vaksin Covid-19 Pada Sosial Media Twitter Menggunakan Metode Naïve

- Bayes dan k- Nearest Neighbors,” vol. 1, no. 1, 2022.
- [9] P. Pandunata, C. K. Ananta, and Y. Nurdiansyah, “Analisis Sentimen Opini Publik Terhadap Pekan Olahraga Nasional Pada Instagram Menggunakan Metode Naïve Bayes Classifier,” *INFORMAL Informatics J.*, vol. 7, no. 2, p. 146, 2022, doi: 10.19184/isj.v7i2.33928.
- [10] W. Bourequat and H. Mourad, “Sentiment Analysis Approach for Analyzing iPhone Release using Support Vector Machine,” *Int. J. Adv. Data Inf. Syst.*, vol. 2, no. 1, pp. 36–44, 2021, doi: 10.25008/ijadis.v2i1.1216.
- [11] 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.
- [12] A. Rahman, E. Utami, and S. Sudarmawan, “Sentimen Analisis Terhadap Aplikasi pada Google Playstore Menggunakan Algoritma Naïve Bayes dan Algoritma Genetika,” *J. Komtika (Komputasi dan Inform.)*, vol. 5, no. 1, pp. 60–71, 2021, doi: 10.31603/komtika.v5i1.5188.
- [13] R. L. Musyarofah, E. U. Utami, and S. R. Raharjo, “Analisis Komentar Potensial pada Social Commerce Instagram Menggunakan TF-IDF,” *J. Eksplora Inform.*, vol. 9, no. 2, pp. 130–139, 2020, doi: 10.30864/eksplora.v9i2.360.
- [14] Sintia, S. Defit, and G. W. Nurcahyo, “Product Codefication Accuracy With Cosine Similarity and Weighted Term Frequency and Inverse Document Frequency (Tf-Idf),” *J. Appl. Eng. Technol. Sci.*, vol. 2, no. 2, pp. 14–21, 2021, doi: 10.37385/jaets.v2i2.210.
- [15] R. Wati, S. Ernawati, and H. Rachmi, “Pembobotan TF-IDF Menggunakan Naïve Bayes pada Sentimen Masyarakat Mengenai Isu Kenaikan BIPIH,” *J. Manaj. Inform.*, vol. 13, no. 1, pp. 84–93, 2023, doi: 10.34010/jamika.v13i1.9424.
- [16] “Implementasi Data Mining Menggunakan Metode Naïve Bayes Classifier Untuk Data Kenaikan Pangkat Dinas Ketenagakerjaan Kota Medan,” *J. Sci. Soc. Res.*, vol. 5, no. 1, p. 85, 2022, doi: 10.54314/jssr.v5i1.804.
- [17] M. Ismali, N. Hassan, and S. S. Bafjaish, “Comparative Analysis of Naive Bayesian Techniques in Health-Related for Classification Task,” *J. Soft Comput. Data Min.*, vol. 1, no. 2, pp. 1–10, 2020.
- [18] E. Hasibuan and E. A. Heriyanto, “Analisis Sentimen Pada Ulasan Aplikasi Amazon Shopping Di Google Play Store Menggunakan Naive Bayes Classifier,” *J. Tek. dan Sci.*, vol. 1, no. 3, pp. 13–24, 2022, doi: 10.56127/jts.v1i3.434.
- [19] Y. A. Singgalen, “Penerapan Metode CRISP-DM dalam Klasifikasi Data Ulasan Pengunjung Destinasi Danau Toba Menggunakan Algoritma Naïve Bayes Classifier (NBC) dan Decision Tree (DT),” *J. Media Inform. Budidarma*, vol. 7, no. 3, pp. 1551–1562, 2023, doi: 10.30865/mib.v7i3.6461.

