3rd Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)
30 Agustus 2023 – Jakarta, Indonesia
Volume 2, Nomor 2, September 2023 - ISSN 2962-8628 (*online*)

PENERAPAN *NAÏVE BAYES* UNTUK MENGANALISIS SENTIMEN PENGGUNA *TWITTER* TERHADAP PENETAPAN CALON PRESIDEN 2024 PDIP

Sulthan Laksono Ramadhan[1*], Windarto[2]

[1,2]Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta Selatan, Indonesia

Email: [1*]1911500641@student.budiluhur.ac.id, [2]windarto@budiluhur.ac.id
(* : corresponding author)

Abstrak-Pemilu di Indonesia adalah proses demokratisasi penting untuk menentukan wakil rakyat dan pemimpin negara. Indonesia menerapkan sistem demokrasi langsung dengan hak pilih rakyat dalam memilih perwakilan di lembaga politik. Ketua Umum Partai Demokrasi Indonesia (PDI) Perjuangan, Megawati Soekarnoputri, menunjuk Ganjar Pranowo sebagai calon presiden (capres) PDIP pada 21 April 2023. Meski Ganjar Pranowo memiliki prestasi yang mengesankan sebagai calon presiden pemilu 2024, tetap ada opini negatif di media sosial seperti Instagram, Twitter, dan Facebook. Penelitian ini menganalisis sentimen masyarakat terhadap Ganjar Pranowo sebagai calon presiden di Twitter menggunakan Algoritma Naïve Bayes. Dataset diperoleh melalui Aplikasi RapidMiner dalam rentang 10 Juni hingga 22 Juni 2023, mengumpulkan 10.346 data dengan kata kunci "ganjar capres" dan "ganjar presiden". Analisis sentimen menggunakan pembobotan kata TF-IDF dan Algoritma Naive Bayes, dengan hasil pengujian sebagai berikut: akurasi 69%, presisi 91%, dan recall 72%. Sentimen positif mencapai 84.18%, sementara negatif 15.82%. Sebagian besar masyarakat Indonesia memberikan pandangan positif terhadap Ganjar Pranowo sebagai calon presiden pada rentang waktu tersebut. Penelitian ini memberikan wawasan tentang respons masyarakat terhadap Ganjar Pranowo sebagai calon presiden. Hasilnya menunjukkan mayoritas sentimen positif, meski terdapat opini negatif di media sosial. Ini memberikan gambaran tentang persepsi masyarakat terhadap Ganjar Pranowo dan potensinya sebagai calon presiden.

Kata Kunci: analisis sentimen, ganjar pranowo, *Naïve Bayes*

IMPLEMENTATION OF NAIVE BAYES TO ANALYZE TWITTER USERS' SENTIMENTS FOR NOMINATION OF THE PDIP PRESIDENTIAL CANDIDATE IN 2024

Abstract-*Elections in Indonesia are an important democratization process to determine the people's representatives and state leaders. Indonesia implements a direct democracy system with the people's right to vote in electing representatives in political institutions. Chairperson of Partai Demokrasi Indonesia Perjuangan, Megawati Soekarnoputri, appointed Ganjar Pranowo as the PDIP presidential candidate (candidate) on April 21, 2023. Even though Ganjar Pranowo has impressive achievements as a presidential candidate for the 2024 elections, there is still negative opinion on social media such as Instagram, Twitter, and Facebook. This study analyzes public sentiment towards Ganjar Pranowo as a presidential candidate on Twitter using the Naïve Bayes Algorithm. The dataset was obtained through the RapidMiner Application in the range June 10 to June 22 2023, collecting 10,346 data with the keywords "reward for presidential candidates" and "reward for president". Sentiment analysis uses TF-IDF word weighting and the Naive Bayes Algorithm, with the following test results: 69% accuracy, 91% precision, and 72% recall. Positive sentiment reached 84.18%, while negative 15.82%. Most Indonesian people gave a positive view of Ganjar Pranowo as a presidential candidate during that period. This research provides insight into the public's response to Ganjar Pranowo as a presidential candidate. The results show the majority of positive sentiments, although there are negative opinions on social media. This provides an overview of the public's perception of Ganjar Pranowo and his potential as a presidential candidate.*

Keywords*: sentiment analysis, ganjar pranowo, Naïve Bayes*

1. PENDAHULUAN

Pemilu di Indonesia adalah proses demokratisasi yang penting dalam menentukan wakil rakyat dan pemimpin negara. Negara Indonesia menerapkan sistem demokrasi langsung dengan menggunakan hak pilih rakyat dalam memilih para perwakilan mereka di lembaga-lembaga politik. Sejarah pemilu di Indonesia dimulai pada tahun 1955 dengan pemilu pertama setelah kemerdekaan dari penjajahan Belanda. Pemilu ini merupakan tonggak penting dalam proses pembentukan negara demokratis di Indonesia. Namun, setelah itu, Indonesia mengalami masa-masa otoriterisme, di mana pemilu tidak berlangsung secara bebas dan adil. Setelah Reformasi pada tahun

1998, sistem politik Indonesia mengalami perubahan signifikan. Pemilu diadakan secara teratur setiap lima tahun dan melibatkan partai politik yang berkompetisi untuk mendapatkan kursi di parlemen dan kepemimpinan nasional. Pemilu di Indonesia melibatkan sejumlah partai politik yang bermacam-macam ideologi dan platform.

Ketua Umum Partai Demokrasi Indonesia (PDI) Perjuangan, Megawati Soekarnoputri secara resmi menunjuk Ganjar Pranowo sebagai calon presiden (capres) dari PDIP pada Jumat siang, 21 April 2023. Ada beberapa alasan mengapa Ketua umum PDIP memilih Ganjar Pranowo sebagai calon presiden. PDIP Hasto Kristiyanto di Jiexpo, Kemayoran, Jakarta Pusat mengatakan bahwa : "Yang dicari oleh Bu Mega dan dipersiapkan oleh Bu Mega adalah pemimpin yang kokoh secara ideologi, pemimpin yang visioner, pemimpin yang mempuni, pemimpin yang punya kemampuan profesional, tetapi sekaligus memahami kehendak rakyat" [1]. Demikian alasan mengapa Ketua Umum PDIP, Megawati Soekarnoputri memilih Ganjar Pranowo sebagai calon presiden.

Walaupun banyak prestasi yang dicapai oleh Ganjar Pranowo untuk menjadi kekuatan dalam menjadi kandidat calon presiden pemilu 2024, bukan berarti tidak banyak opini yang bersifat Positif dan Negatif dari masyarakat tentang Ganjar Pranowo. Opini-opini tersebut dapat dilihat di platform sosial media seperti *Instagram*, *Twitter*, *Facebook*, dll. Dari banyaknya opini tersebut, masyarakat sulit untuk mengetahui apakah Ganjar Pranowo selalu mendapatkan opini yang selalu negatif atau positif. Maka dari itu, Analisis Sentimen bisa dilakukan untuk menentukan apakah banyak dari opini tersebut bersifat negatif atau positif.

Penelitian Shima Fanissa, M. Ali Fauzi, dan Sigit Adinugroho [2] yang berjudul Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode *Naive Bayes* dan Seleksi Fitur *Query Expansion Ranking*, mendapatkan hasil seleksi fitur 75% memiliki akurasi terbaik sebesar 86.6%. Penelitian lain oleh Dwi Normawati dan Surya Allit Prayogi [3] yang berjudul Implementasi *Naïve Bayes Classifier* Dan *Confusion Matrix* Pada Analisis Sentimen Berbasis Teks Pada *Twitter*, penelitian ini menghasilkan pemaparan yang terstruktur pada proses dan hasil implementasi *Naïve Bayes Classifier* dan pengujian performa menggunakan *Confusion Matrix* yang didapatkan akurasi sebesar 82%, presisi 93%, dan *Recall* sebesar 52%. Penelitian lain oleh Muhammad Raihan Fais Sya'bani, Ultach Enri, dan Tesa Nur Padilah [4] yang berjudul Analisis Sentimen Terhadap Bakal Calon Presiden 2024 dengan Algoritma Naïve Bayes, hasil penelitian menyimpulkan bahwa warganet memiliki pandangan positif terhadap setiap calon presiden yang akan datang. Selanjutnya, evaluasi algoritme naïve bayes pada dataset pertama menunjukkan akurasi sebesar 73,68% dan AUC sebesar 0,74 pada lipatan (fold) ke-7. Pada dataset kedua, akurasi mencapai 71,43% dan AUC mencapai 1,0 pada lipatan ke-5. Untuk dataset ketiga, akurasi mencapai 60% dan AUC mencapai 0,92 pada lipatan ke-1. Sedangkan pada dataset terakhir, akurasi diperoleh sebesar 62,5% dengan AUC 0,65 pada lipatan ke-3.

Pembahasan Penelitian ini akan menggunakan Algoritme *Naïve Bayes*. Penelitian ini diharapkan dapat memberikan informasi opini publik terkait Ganjar Pranowo sebagai kandidat calon presiden pemilihan umum 2024. Berdasarkan penelitian sebelumnya, metode ini sangat cocok untuk berbagai jenis *dataset*, karena memiliki kecepatan dalam mengklasifikasikan data dan memberikan akurasi yang tinggi dalam prosesnya.

Analisis sentimen adalah bidang ilmu dalam data mining yang memiliki tujuan untuk menganalisis, memproses, dan mengekstraksi informasi dari data teks yang terkait dengan entitas tertentu, seperti layanan, produk, individu, organisasi, peristiwa, atau masalah dan topik tertentu. Metode ini berguna untuk memahami sentimen, opini, atau sikap yang terkandung dalam teks tersebut [5].

Naïve Bayes merupakan sebuah metode pengklasifikasian dengan menggunakan probabilitas sederhana yang berakar pada *Teorema Bayes* dan memiliki asumsi ketidaktergantungan (*independent*) yang tinggi dari masing – masing kondisi atau kejadian [6].

2. METODE PENELITIAN

2.1 Pengumpulan Data

Pada tahap pengumpulan data, dilakukan proses *Crawling* yang melibatkan beberapa langkah. Langkah-langkah tersebut meliputi membuat koneksi yang menghubungkan *RapidMiner* dengan akun *Twitter*. Data *Tweet* yang berhasil dikumpulkan akan disimpan ke alamat yang telah diatur. Rincian tentang tahapan pengumpulan data dapat dilihat pada Gambar 1 berikut ini.

Gambar 1. Pengumpulan Data

2.2 *Preprocessing*

Preprocessing adalah tahap penting dalam analisis data yang dilakukan sebelum proses klasifikasi [7]. Tahap *Preprocessing* antara lain: *Case folding*, *Cleansing*, mengubah *Slang Word*, menghapus *Stop Word*, dan *Stemming*.

2.2.1 *Case Folding*

Case folding adalah proses konversi semua huruf dalam dokumen atau kalimat menjadi huruf kecil [8]. Dalam proses *Case folding*, teks disamakan menjadi huruf kecil (*lowercase*) secara keseluruhan. Contohnya, kata 'Presiden' atau ‘PRESIDEN' akan diubah menjadi ‘presiden’.

2.2.2 *Cleansing*

Tahapan ini, semua karakter didalam teks yang bukan alfabet dihapus sehingga dapat mengurangi karakter yang tidak dikehendaki dan tidak memiliki arti dalam analisis sentimen [9]. Dalam proses pembersihan (*Cleansing*), teks akan disaring dan dihapus. Proses pembersihan terdiri dari beberapa tahap, seperti hapus *Mention*, hapus *hashtag*, hapus *Link*, hapus selain angka dan huruf, hapus spasi berlebih, dan hapus kata yang hanya memiliki 1 atau 2 huruf.

a. Hapus *Mention*: Proses ini akan menghapus *mention* pada teks, sebagai contoh kata ‘@akuKamu’ akan di hapus dari kalimat.
b. Hapus *Hashtag*: Proses hapus *hashtag* dilakukan untuk menghilangkan kata *hashtag*. Contoh kata ‘#pemilu’ akan di hapus dari kalimat.
c. Hapus *Link*: Proses hapus *link* dilakukan untuk menghilangkan *link* pada teks. Contoh kata ‘http://t.co/iGkdfkj’ akan dihapus dari kalimat.
d. Hapus selain angka dan huruf: Proses hapus selain angka dan huruf dilakukan untuk menghilangkan karakter selain huruf dan angka yang ada pada teks. Contoh ‘pemilihan presiden 2024!’, maka akan diubah menjadi ‘pemilihan presiden 2024’.
e. Hapus spasi berlebih: Proses hapus spasi berlebih dilakukan untuk menghilangkan whitespace yang berlebihan pada teks. Contoh ‘pemilihan presiden 2024’, maka akan diubah menjadi ‘pemilihan presiden 2024’.
f. Hapus kata yang hanya memiliki 1 atau 2 huruf: Proses hapus kata yang hanya memiliki 1 atau 2 huruf dilakukan untuk menghilangkan kata yang hanya memiliki 1 atau 2 huruf saja pada teks. Contoh kata ‘a’ dan ‘di’ akan di hapus dari teks.

2.2.3 Mengubah *Slang Word*

Proses pengubahan *Slang Word* melibatkan penggantian setiap kata gaul, kata singkatan, atau kata tidak baku menjadi bentuk standarnya. Misalnya, kata 'utk' menjadi 'untuk', 'yng' menjadi 'yang', dan 'apotik' menjadi 'apotek'.

2.2.4 Menghapus *Stop Word*

Proses penghapusan *Stop Word* melibatkan penghilangan kata-kata yang memiliki sedikit makna namun sering muncul dalam sebuah teks. Contohnya, kata-kata seperti 'untuk', 'yang', dan 'apa'.

2.2.5 *Stemming*

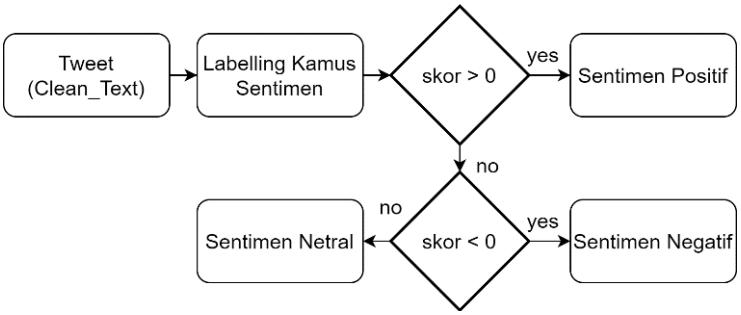
Proses *Stemming* melibatkan pengubahan kata-kata yang memiliki imbuhan menjadi bentuk dasarnya, dengan menggunakan pustaka Sastrawi. Contohnya, kata 'menambah' diubah menjadi 'tambah', dan kata 'menyiapkan' diubah menjadi 'siap'.

2.3 *Labelling*

Pelabelan adalah tahapan untuk memberikan kelas pada suatu dokumen atau teks berdasarkan karakteristik atau cirinya [10]. Untuk melakukan perhitungan skor sentimen digunakan pendekatan kamus sentimen. Kamus tersebut berisikan kata sentimen positif dan negatif. Skor suatu kata akan bernilai +1 jika kata tersebut adalah kata opini positif, dan bernilai -1 jika kata tersebut adalah kata opini negatif [11]. Berikut persamaan 1 mengenai proses perhitungan skor nilai.

$$\text{skor} = \left(\sum Kata\ Positif\right) - \left(\sum Kata\ Negatif\right) \quad (1)$$

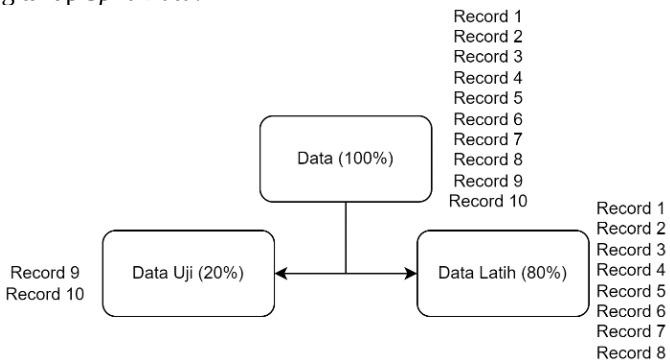
Proses *Labelling* dengan kamus sentimen melibatkan kamus kata positif dan kamus kata negatif yang diperoleh dari penelitian sebelumnya yang dipublikasikan di laman *GitHub* (https://github.com/masdevid/ID-OpinionWords). Proses utama *Labelling* dilakukan dengan menggunakan metode *Labelling* dengan kamus sentimen. Gambar 2 menunjukkan ilustrasi tahap *Labelling.*



Gambar 2. Tahap *Labelling*

2.4 *Split Data*

Pada tahap *Split Data*, *Tweet* yang telah diberi label akan dibagi menjadi dua bagian, yaitu data uji dan data latih. Proses *Split Data* dilakukan dengan membagi *dataset* menjadi 80% data latih dan 20% data uji. Gambar 3 memberikan ilustrasi tentang tahap *Split Data*.



Gambar 3. Tahap *Split Data*

2.5 *TF-IDF*

TF-IDF adalah metode yang digunakan untuk menghitung bobot kata-kata dalam dokumen kunci di setiap kategori. Metode ini juga digunakan untuk mencari kata-kata kunci yang memiliki kemiripan dengan kategori yang tersedia [3]. *TF-IDF* digunakan untuk mengukur frekuensi relatif dari istilah tertentu dalam kumpulan dokumen dan untuk menilai sejauh mana sebuah kata umum atau jarang digunakan di antara korpus teks yang terstruktur. Metode ini memberikan bobot pada setiap kata berdasarkan perhitungan *TF-IDF* [7]. Berikut ini persamaan 2 adalah rumus untuk menghitung TF.

$$\text{TF(t)} = f_{t,d} \,/\, \sum t, d \quad (2)$$

Dalam rumus ini, ft,d mewakili frekuensi sebuah kata dalam dokumen, dan ∑ t,d mewakili total kata yang terdapat dalam dokumen tersebut. Selanjutnya, untuk menghitung IDF (*Inverse Document Frequency*) dari dokumentasi tersebut, kita dapat menggunakan persamaan 3 berikut:

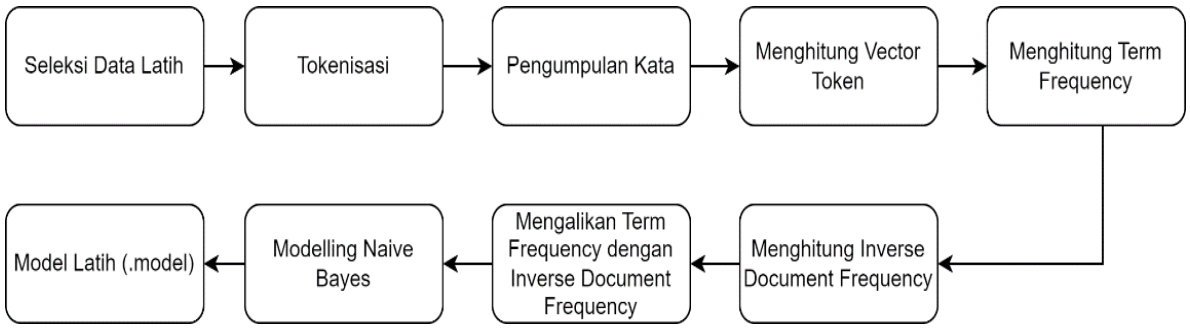
$$\text{IDF(t)} = \log\left(|\text{D}| \,/\, \text{f}_{\text{t,D}}\right) \quad (3)$$

|D| mengacu pada jumlah dokumen dalam koleksi, sementara ft,D mengacu pada jumlah dokumen di mana kata t muncul dalam koleksi tersebut. Setelah mendapatkan nilai-nilai TF dan IDF, langkah selanjutnya adalah menghitung nilai *TF-IDF* menggunakan persamaan 4 berikut:

$$\mathrm{TF} - \mathrm{IDF} = \mathrm{TF(t)} * \mathrm{IDF(t)} \quad (4)$$

2.6 *Modelling*

Pada tahap ini, terdapat delapan proses utama yang harus dilakukan untuk menghasilkan sebuah model latih. Delapan proses tersebut meliputi seleksi data latih, tokensasi, pengumpulan kata, menghitung *vector* token, menghitung *term frequency*, menghitung *Inverse Document Frequency*, mengalikan *term frequency* dengan *Inverse Document Frequency*, dan *Modelling Naïve Bayes*. Ilustrasi dari proses-proses tersebut dapat ditemukan pada Gambar 4 berikut ini.



Gambar 4. Tahap *Modelling*

2.7 *Naïve Bayes*

Naive Bayes didasarkan pada *Teorema Bayes*, yang menyediakan cara untuk menghitung probabilitas suatu kejadian berdasarkan probabilitas kondisional dari kejadian tersebut dan probabilitas dari kejadian yang terkait. *Teorema Bayes* dinyatakan secara matematis dalam persamaan 5 berikut:

$$P(H|E) = \frac{P(E|H).P(H)}{P(E)} \quad (5)$$

Keterangan:
E = Data yang kelasnya belum diketahui.
H = Spesifikasi kelas hipotesis untuk data E.
P(H|E) = Jika memiliki informasi tentang E, maka dapat menghitung probabilitas bahwa H akan terjadi.
P(E|H) = Jika mengetahui H, maka kemungkinan E akan terjadi adalah probabilitas posterior E dengan H.
P(H) = Jika kita melihat probabilitas kejadian H sebelum ada informasi lain maka disebut sebagai probabilitas prior H.
P(E) = probabilitas prior, probabilitas kejadian E.

Selanjutnya adalah penentuan probabilitas prior bagi tiap kategori berdasarkan sampel dokumen. Pada tahap klasifikasi ditentukan nilai kategori dari suatu dokumen berdasarkan *term* yang muncul dalam dokumen yang diklasifikasi [12].

2.8 *Confusion Matrix*

Confusion Matrix adalah suatu metode yang umumnya digunakan untuk melakukan perhitungan tingkat akurasi pada *text mining* [8]. Contoh *Confusion Matrix* seperti pada Tabel 1.

Tabel 1. *Confusion Matrix*

		Kelas Aktual	
		Positif	Negatif
Kelas Prediksi	Positif	TP	FP
	Negatif	FN	TN

Keterangan:
TP (*True Positive*) = jumlah dokumen dari kelas Positif yang benar diklasifikasikan sebagai kelas Positif
TN (*True Negative*) = jumlah dokumen dari kelas Negatif yang benar diklasifikasikan sebagai kelas Negatif
FP (*False Positive*) = jumlah dokumen dari kelas Negatif yang salah diklasifikasikan sebagai kelas Positif
FN (*False Negative*) = jumlah dokumen dari kelas Positif yang salah diklasifikasikan sebagai kelas Negatif

Terdapat tiga parameter yang akan dihitung, yaitu accuracy, precision, dan *Recall*. Rumus *Confusion Matrix* untuk menghitung accuracy (Persamaan 6), precision (Persamaan 7), dan *Recall* (Persamaan 8) seperti berikut.

$$accuracy = \frac{TP + TN}{TOTAL} \quad (6)$$

$$precision = \frac{TP}{TP + FP} \quad (7)$$

$$recall = \frac{TP}{TP + FN} \quad (8)$$

2.9 Pengujian

Untuk mengungkapkan sentimen dari cuitan di *Twitter* yang membahas tentang Ganjar Pranowo sebagai kandidat calon presiden pada pemilu 2024, dilakukan pengujian metode yang melibatkan implementasi dalam sebuah aplikasi *web*. Dari tahapan tersebut diharapkan menghasilkan evaluasi perhitungan klasifikasi yang direpresentasikan dalam bentuk *Confusion Matrix*. Hasil dari klasifikasi *Confusion Matrix* akan dikategorikan ke dalam empat kelompok yaitu positif benar (*True Positive*), positif salah (*False Positive*), negatif benar (*True Negative*), dan negatif salah (*False Negative*). Dari informasi ini dapat memberikan wawasan tentang efektivitas metode yang digunakan dalam menganalisis sentimen dari cuitan-cuitan di *Twitter* terkait Ganjar Pranowo sebagai kandidat calon presiden.

3. HASIL DAN PEMBAHASAN

3.1 Tahap Pengumpulan Data

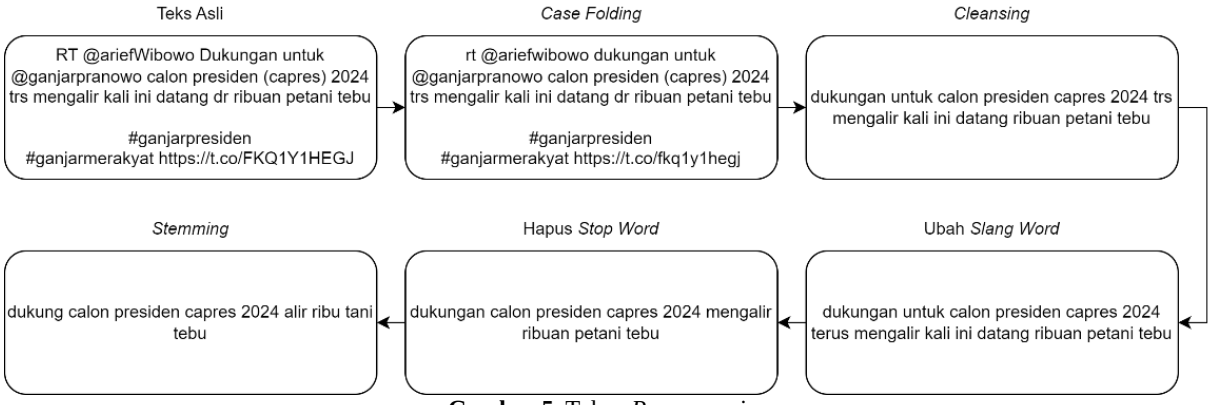
Data penelitian ini berasal dari platform media sosial *Twitter* dalam bentuk *Tweet*. *Dataset* ini diperoleh menggunakan Aplikasi *RapidMiner*, dengan rentang 10 Juni 2023 hingga 22 Juni 2023, dan berhasil mengumpulkan sebanyak 10.346 data. Kata kunci yang digunakan adalah "ganjar capres" dan "ganjar presiden". Selanjutnya data akan dilakukan *Preprocessing*. *Sample* Data Penelitian bisa dilihat pada tabel 2.

Tabel 2. Data Penelitian

Created_at	*User*	*Text*
2023-06-01 21:59:16	Mohamad Guntur Romli	Sindir Capres yg Cemas Dijegal: Kalau Sudah Mau Nyalon Jangan Takut Isu Apapun https://t.co/RAnhCzaK6L
2023-06-02 05:01:00	CNN Indonesia	Ganjar Klaim Tak Pakai APBD untuk Safari Politik Capres https://t.co/GXRWEd8UHM

3.2 Tahap *Preprocessing*

Tahap *Preprocessing* adalah langkah yang hanya dapat dilakukan setelah memiliki satu atau lebih *dataset* dalam basis data yang telah dikumpulkan. Tahap ini terdiri dari lima (5) proses utama, yaitu: *Case folding*, *Cleansing*, merubah *Slang Word*, menghapus *Stop Word*, dan *Stemming*. Berikut ilustrasi tahap *Preprocessing* pada gambar 5.



Gambar 5. Tahap *Preprocessing*

Dalam penelitian ini, tahap *Preprocessing* akan mengolah dan menghasilkan 10.224 data *Tweet* yang telah diubah menjadi lebih terstruktur atau yang disebut sebagai *clean text*. Setelahnya data yang sudah di *Preprocessing* akan disimpan dalam basis data untuk tahap selanjutnya, yaitu proses *Labelling*. Namun, ada beberapa data yang tidak memiliki *clean text* dikarenakan setelah dilakukan *Preprocessing* tidak ada kata yang memiliki arti, data tersebut tidak akan dimasukkan ke dalam *table data_Preprocessing* di *database*.

3.3 Tahap *Labelling*

Tahap *Labelling*, dilakukan setelah tersedianya satu atau lebih data *clean text* pada basis data (*database*) hasil *Preprocessing*. Tahapan ini melibatkan penggunaan kamus sentimen untuk menentukan kelas sentimen dengan menghitung skor sentimen. Detail proses perhitungan skor sentimen pada sebuah *Tweet* dapat ditemukan dalam tabel 3.

Tabel 3. Tahap *Labelling*

Clean text	Kata Positif	Kata Negatif
dukung calon presiden alir ribu tani tebu	dukung	
Jumlah	1	0

Dari tabel 3, terlihat bahwa terdapat 1 kata positif dalam *clean text*, yang ditemukan berdasarkan frekuensi kemunculan kata positif. Sementara itu, tidak terdapat kata negatif dalam *clean text*, yang ditemukan berdasarkan frekuensi kemunculan kata negatif. Dengan menggunakan persamaan (1), skor untuk *Tweet* "dukung calon presiden alir ribu tani tebu" dapat dihitung sebagai berikut.

Skor = (jumlah kata positif) – (jumlah kata negatif) = 1-0 = 1

Setelah mendapatkan nilai skor, langkah berikutnya adalah memberikan kelas sentimen berdasarkan aturan yang dijelaskan dalam gambar 2.

If skor > 0 : Kelas = ‘positif’;
Else if skor < 0 : Kelas = ‘negatif’;
Else : Kelas = ‘netral’;

Dengan demikian, dapat disimpulkan bahwa *Tweet* ‘dukung calon presiden alir ribu tani tebu' akan diklasifikasikan sebagai kelas positif, karena nilai skornya lebih besar dari 0. Dalam penelitian ini, terdapat total 8.445 *Tweet* yang telah diberi label. Ada 1.779 *Tweet* yang berkelas netral, *Tweet* yang berkelas netral tidak akan digunakan pada tahap selanjutnya. Hanya *Tweet* yang berkelas positif dan negatif yang akan digunakan pada tahap selanjutnya.

3.4 Tahap *Split Data*

Tahap *Split Data* dilakukan setelah mendapatkan satu atau lebih data yang telah diberi label pada basis data hasil *Labelling*. *Tweet* yang telah diberi label akan dibagi menjadi dua bagian, yaitu data latih dan data uji. *Split Data* dilakukan dengan menggunakan rasio 80:20, atau 80% data latih dan 20% data uji.

Dalam penelitian ini, 8.445 *Tweet* yang telah diberi label akan dibagi menggunakan rasio 80:20. Oleh karena itu, akan diperoleh 1.689 *Tweet* berlabel sebagai data uji dan 6.756 *Tweet* berlabel sebagai data latih. *Tweet* yang telah diberi label tersebut kemudian akan disimpan dalam basis data untuk tahap selanjutnya, yaitu *Modelling*.

3.5 Tahap *Modelling*

Tahap *Modelling* merupakan tahapan yang dilakukan setelah *Tweet* melalui proses *Preprocessing*, *Labelling*, dan pembagian data. Tahapan ini bertujuan untuk memperoleh model latih atau pengetahuan melalui data latih yang ada. Tahap ini terdapat delapan (8) proses utama antara lain: seleksi data latih, tokenisasi, pengumpulan kata, menghitung *vector* token, penghitungan *Term frequency*, penghitungan *Inverse Document Frequency*, mengalikan *Term frequency* dengan *Inverse Document Frequency*, *Modelling Naïve Bayes*. Tahap *Modelling* menghasilkan probabilitas Likelihood seperti pada Tabel 4.

Tabel 4. Probabilitas *Likelihood*

Kata	Probabilitas *Likelihood* di Positif	Probabilitas *Likelihood* di Negatif
ganjar	0.16	0.153
pasti	0.08	0.038
menang	0.08	0.038
pemilu	0.08	0.038
masyarakat	0.12	0.038
bogor	0.08	0.038
sambut	0.08	0.076
baik	0.08	0.038
indonesia	0.08	0.038
dukung	0.08	0.038
capres	0.08	0.076
tangerang	0.04	0.076
kota	0.04	0.076
tidak	0.04	0.076
asal	0.04	0.076
bangun	0.04	0.076
jalan	0.04	0.076
populer	0.04	0.076
video	0.04	0.076
porno	0.04	0.076

Jika sudah mendapatkan probabilitas *likelihood* kata, langkah selanjutnya adalah membuat file model. Langkah ini menjadikan data dari tabel diatas menjadi file model (.model). Dari hasil *Modelling* menghasilkan 6.756 data *Tweet* dengan jumlah label positif sebanyak 5.929 dan label negatif sebanyak 827.

3.6 Tahap Klasifikasi *Naïve Bayes*

Tahap klasifikasi menggunakan *Naive Bayes* adalah tahap yang dilakukan setelah tahap *Modelling* dalam pengolahan data menggunakan metode *Naive Bayes*. Tujuan dari tahap klasifikasi ini adalah untuk memprediksi label atau kelas dari data uji berdasarkan model yang telah dilatih pada tahap sebelumnya. Persiapan data melibatkan memilih model pelatihan. Model pelatihan yang terpilih akan digunakan sebagai dasar untuk melakukan klasifikasi pada data uji yang ada. Data uji yang akan digunakan adalah contoh data uji yang tercantum dalam tabel 5.

Tabel 5. Data Uji

Data Uji	Teks Bersih	Label Sentimen
Uji 1	video bogor sambut baik ganjar	Positif
Uji 2	populer masyarakat kota tidak dukung	Negatif

Untuk memprediksi kelas dari Data Uji tersebut, kita dapat mengestimasi probabilitas posterior untuk setiap kelas dengan menggunakan pendekatan *Naive Bayes*.

a. P (positif|Uji 1) = P(positif) * P(video|positif) * P(bogor|positif) * P(sambut|positif) * P(baik|positif) * P(ganjar|positif) = 0.5 * 0.04 * 0.08 * 0.08 * 0.08 * 0.16 = 0,0000016384

b. P (negatif|Uji 1) = P(negatif) * P(video|negatif) * P(bogor|negatif) * P(sambut|negatif) * P(baik|negatif) * P(ganjar|negatif) = 0.5 * 0.076 * 0.038 * 0.076 * 0.038 * 0.153 = 0,000000633881344
c. P (positif|Uji 2) = P(positif) * P(populer|positif) * P(masyarakat|positif) * P(kota|positif) * P(tidak|positif) * P(dukung|positif) = 0.5 * 0.04 * 0.12 * 0.04 * 0.04 * 0.08 = 0,0000003072
d. P (negatif|Uji 2) = P(negatif) * P(populer|negatif) * P(masyarakat|negatif) * P(kota|negatif) * P(tidak|negatif) * P(dukung|negatif) = 0.5 * 0.076 * 0.038 * 0.076 * 0.076 * 0.038 = 0,000000316940672

Tabel 6. Tabel hasil prediksi *Naïve Bayes*

Data Uji	Teks Bersih	Label Sentimen	Probabilitas	Label Prediksi
Uji 1	video bogor sambut baik ganjar	Positif	Positif = 0,0000016384 Negatif = 0,000000633881344	Positif
Uji 2	populer masyarakat kota tidak dukung	Negatif	Positif = 0,0000003072 Negatif = 0,000000316940672	Negatif

Tabel 6 merupakan hasil prediksi *Naïve Bayes*, dari tabel diatas dapat diketahui bahwa klasifikasi menggunakan *Naïve Bayes* dapat melakukan prediksi dengan benar.

3.7 Hasil Pengujian

Data uji yang telah dilakukan prediksi sebanyak 1.689 data, data akan direpresentasikan dengan *Confusion Matrix*. Tabel 7 berikut adalah tabel representasi *Confusion Matrix*:

Tabel 7. *Confusion Matrix*

		Nilai Aktual	
		Positif	Negatif
Nilai Prediksi	Positif	1081	416
	Negatif	99	93

Berdasarkan tabel 7, dapat digunakan persamaan (6), (7), dan (8) untuk menghitung nilai akurasi, presisi, dan *Recall*. Hasil Akurasi, Presisi, dan *Recall* dapat ditemukan dalam tabel 8 berikut.

Tabel 8. Hasil Akurasi, Presisi, dan *Recall*

Pengujian	
Akurasi	0.69 (69%)
Presisi	0.91 (91%)
Recall	0.72 (72%)

Berdasarkan tabel 8 di atas, dapat diketahui bahwa hasil pengujian menunjukkan bahwa Algoritme *Naïve Bayes* mampu memperoleh nilai akurasi 69%, presisi 91%, dan *Recall* 72%. Berdasarkan evaluasi sentimen terhadap 8.445 data yang mencakup data latih dan data uji, temuan menunjukkan bahwa mayoritas masyarakat Indonesia mengekspresikan sentimen positif sebesar 84.18% dalam jangka waktu 10 Juni hingga 22 Juni 2023. Sementara itu, sebesar 15.82% menyatakan sentimen negatif.

4. KESIMPULAN

Berdasarkan analisis terhadap 8.445 *Tweet* pada rentang waktu 10 Juni sampai 22 Juni 2023, Sentimen positif tersebut mencapai 84.18%, sedangkan sentiment negatif sebesar 15.82%. Dapat diketahui bahwa mayoritas masyarakat Indonesia memiliki pandangan positif terhadap Ganjar Pranowo Sebagai Kandidat Calon Presiden Pemilihan Umum 2024. Penggunaan pembobotan kata *TF-IDF* dan Algoritme *Naive Bayes* dalam melakukan analisis sentimen telah memberikan hasil nilai pengujian dan evaluasi sebagai berikut: akurasi sebesar 69%, presisi sebesar 91%, dan *Recall* sebesar 72%. Aplikasi Analisis Sentimen dimasa yang akan datang diharapkan, Mengajak kolaborasi dengan pihak-pihak terkait, seperti ahli bahasa atau praktisi industri, untuk mendapatkan wawasan dan masukan yang berharga dalam mengembangkan dan meningkatkan aplikasi ini.

DAFTAR PUSTAKA

[1] A. E. Prawira, "Alasan PDIP Jadikan Ganjar Pranowo Capres: Memahami Kehendak Rakyat," liputan6, 2023. https://www.liputan6.com/health/read/5267688/alasan-pdip-jadikan-ganjar-pranowo-capres-memahami-kehendak-rakyat (accessed Jun. 07, 2023).

[2] S. Fanissa, M. A. Fauzi, and S. Adinugroho, "Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 8, pp. 2766–2770, 2018, [Online]. Available: https://www.researchgate.net/publication/322959527

[3] 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, [Online]. Available: http://ejurnal.tunasbangsa.ac.id/index.php/jsakti/article/view/369

[4] M. R. F. 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.

[5] V. K. S. Que, A. Iriani, and H. D. Purnomo, "Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 9, no. 2, pp. 162–170, 2020, doi: 10.22146/jnteti.v9i2.102.

[6] F. V. Sari and A. Wibowo, "ANALISIS SENTIMEN PELANGGAN TOKO ONLINE JD.ID MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER BERBASIS KONVERSI IKON EMOSI," *J. SIMETRIS*, vol. 10, no. 2, pp. 681–686, 2019.

[7] I. P. Rahayu, A. Fauzi, and J. Indra, "Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Naive Bayes Dan Support Vector Machine," *J. Sist. Komput. dan Inform.*, vol. 4, no. 2, pp. 296–301, 2022, doi: 10.30865/json.v4i2.5381.

[8] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *JEPIN (Jurnal Edukasi dan Penelit. Inform.)*, vol. 4, no. 2, pp. 17–29, 2018, [Online]. Available: www.femaledaily.com

[9] H. Tuhuteru, "Analisis Sentimen Masyarakat Terhadap Pembatasan Sosial Berksala Besar Menggunakan Algoritma Support Vector Machine," *Inf. Syst. Dev.*, vol. 5, no. 2, pp. 7–13, 2020.

[10] M. P. Wibowo, S. Amini, Indra, and D. Kusumaningsih, "ANALISIS SENTIMEN MASYARAKAT INDONESIA PADA TWITTER TERHADAP ISU RESESI 2023 MENGGUNAKAN METODE NAIVE BAYES," Semin. Nas. Mhs. Fak. Teknol. Inf., vol. 2, no. 1, pp. 201–210, 2023.

[11] M. Priandi and Painem, "Analisis Sentimen Masyarakat Terhadap Pembelajaran Daring di Era Pandemi Covid-19 pada Media Sosial Twitter Menggunakan Ekstraksi Fitur Countvectorizer dan Algoritma K-Nearest Neighbor," Semin. Nas. Mhs. Ilmu Komput. dan Apl., pp. 311–319, 2021.

[12] D. D. Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, pp. 34–40, Jan. 2022, doi: 10.23960/jitet.v10i1.2262.