

**ANALISIS SENTIMEN DATA *TWITTER* TERHADAP
BAKAL CALON PRESIDEN GANJAR PRANOWO
TAHUN 2024 MENGGUNAKAN METODE
*NAIVE BAYES CLASSIFIER***

TUGAS AKHIR

Diajukan sebagai salah satu persyaratan guna memperoleh gelar Ahli Madya



**Oleh
TASYA MUTIA
NIM. 20037062**

**PROGRAM STUDI DIPLOMA III STATISTIKA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI PADANG**

2023

SURAT PERNYATAAN TIDAK PLAGIAT

Saya yang bertanda tangan di bawah ini:

Nama : Tasya Mutia
NIM : 20037062
Program Studi : DIII Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Dengan ini menyatakan, bahwa skripsi saya dengan judul "**Analisis Sentimen Data Twitter Terhadap Bakal Calon Presiden Ganjar Pranowo Tahun 2024 Menggunakan Metode Naive Bayes Classifier**" adalah benar merupakan hasil karya saya dan bukan merupakan plagiat dari karya orang lain atau pengutipan dengan cara-cara yang tidak sesuai dengan etika yang berlaku dalam tradisi keilmuan. Apabila suatu saat terbukti saya melakukan plagiat maka saya bersedia diproses dan menerima sanksi akademis maupun hukum sesuai dengan hukum dan ketentuan yang berlaku, baik di institusi UNP maupun di masyarakat dan negara.

Demikianlah pernyataan ini saya buat dengan penuh kesadaran dan rasa tanggung jawab sebagai anggota masyarakat ilmiah.

Diketahui Oleh,
Ketua Departemen Statistika



Dr. Yenni Kurniawati, S.Si, M.Si
NIP. 198402232010122005

Saya yang menyatakan,



Tasya Mutia
NIM. 20037062

HALAMAN PENGESAHAN LULUS UJIAN AKHIR

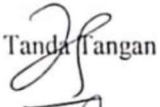


Nama : Tasya Mutia
NIM : 20037062
Program Studi : DIII Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

ANALISIS SENTIMEN DATA *TWITTER* TERHADAP BAKAL CALON PRESIDEN GANJAR PRANOWO TAHUN 2024 MENGUNAKAN METODE *NAIVE BAYES CLASSIFIER*

Dinyatakan lulus setelah dipertahankan di depan Tim Penguji Skripsi Departemen
Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Negeri Padang

Padang, 9 November 2023

Tim Penguji

	Nama	Tanda Tangan
Ketua	: Dra. Nonong Amalita, M.Si	
Anggota	: Dr. Dony Permana, M.Si	
Anggota	: Zamahsary Martha, S.Si.,M.Si	

PERSETUJUAN TUGAS AKHIR

**ANALISIS SENTIMEN DATA *TWITTER* TERHADAP BAKAL
CALON PRESIDEN GANJAR PRANOWO TAHUN 2024
MENGUNAKAN METODE *NAIVE BAYES CLASSIFIER***

Nama : Tasya Mutia
NIM : 20037062
Program Studi : DIII Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Padang, 3 November 2023

Mengetahui,
Ketua Departemen Statistika



Dr. Yenni Kurniawati, S.Si, M.Si
NIP. 198402232010122005

Disetujui Oleh:
Pembimbing



Dra. Nonong Amalita, M.Si
NIP. 196906151993032001

Analisis Sentimen Data *Twitter* Terhadap Bakal Calon Presiden Ganjar Pranowo Tahun 2024 Menggunakan Metode *Naive Bayes Classifier*

Tasya Mutia

ABSTRAK

Presiden Indonesia saat ini telah memegang jabatan selama 2 periode berturut-turut, berdasarkan peraturan tidak bisa dicalonkan lagi pada periode berikutnya. Beberapa lembaga survei telah mengeluarkan hasil surveinya terhadap beberapa bakal calon presiden tahun 2024. Ganjar Pranowo memiliki elektabilitas yang baik untuk menjadi calon Presiden Indonesia pada pemilihan presiden tahun 2024. Tujuan dari penelitian ini adalah untuk mengetahui klasifikasi dan tingkat akurasi terhadap bakal calon presiden Ganjar Pranowo Tahun 2024 menggunakan metode *Naive Bayes Classifier*.

Penelitian ini merupakan penelitian terapan. Metode yang diterapkan yaitu *Naive Bayes Classifier* dalam melihat tingkat akurasi terhadap bakal calon presiden Ganjar Pranowo Tahun 2024. Data yang diambil berdasarkan hasil dari *crawling* data dengan kata kunci “capres granjar pranowo” sebanyak 473 dengan histori *tweet* pada tanggal 21 April 2023 sampai dengan 31 Juli 2023.

Hasil dari penelitian ini menunjukkan bahwa masyarakat memiliki sentimen positif terhadap bakal calon presiden Ganjar Pranowo. Metode *Naive Bayes Classifier* memiliki tingkat akurasi yang tinggi dengan nilai *accuracy* sebesar 82% dan *recall* positif sebesar 83% serta *recall* negatif sebesar 75%. Adapun nilai *recall* untuk kelas sentimen positif lebih tinggi dibanding sentimen negatif. Hal ini disebabkan karena jumlah data untuk kelas positif lebih banyak dibandingkan kelas negatif.

Kata Kunci : Analisis Sentimen, Capres, Ganjar Pranowo, *Naive Bayes Classifier*.

***Sentiment Analysis of Twitter Data on Presidential Candidate Ganjar Pranowo
in 2024 Using the Naive Bayes Classifier Method***

Tasya Mutia

ABSTRACT

The current President of Indonesia has held office for 2 consecutive terms, based on regulations he cannot be nominated again in the next period. Several survey institutions have released the results of their surveys on several presidential candidates in 2024. Ganjar Pranowo has good electability to become a candidate for President of Indonesia in the 2024 presidential election. The aim of this research is to determine the classification and level of accuracy of presidential candidate Ganjar Pranowo in 2024. 2024 using the Naive Bayes Classifier method.

This research is applied research. The method applied is the Naive Bayes Classifier to see the level of accuracy of the 2024 presidential candidate Ganjar Pranowo. The data taken is based on the results of crawling data with the keyword "presidential candidate Granjar Pranowo" as many as 473 with a history of tweets from April 21 2023 to July 31 2023.

This research is applied research. The method applied is the Naive Bayes Classifier to see the level of accuracy of the 2024 presidential candidate Ganjar Pranowo. The data taken is based on the results of crawling data with the keyword "presidential candidate Granjar Pranowo" as many as 473 with a history of tweets from April 21 2023 to July 31 2023.

Keywords: *Sentiment Analysis, Presidential Candidates, Ganjar Pranowo, Naive Bayes Classifier.*

KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Assalamu'alaikum Warahmatullahi Wabaratauh

Alhamdulillah, segala puji dan syukur bagi Allah SWT atas berkah, rahmat, nikmat, kesehatan serta kelancaran yang diberikan kepada hamba-Nya dalam menyelesaikan Tugas Akhir yang berjudul “**Analisis Sentimen Data *Twitter* Terhadap Bakal Calon Presiden Ganjar Pranowo Tahun 2024 Menggunakan Metode *Naive Bayes Classifier***”. Shalawat beserta salam selalu turunkan kepada Nabi Muhammad SAW yang telah mengajarkan kita untuk menjadi manusia yang beradab dan berilmu pengetahuan yang seperti dirasakan pada saat ini.

Tugas Akhir ini disusun sebagai salah satu persyaratan untuk mendapatkan gelar Ahli Madya pada Program Studi Diploma III Statistika, Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang. Dalam penyusunan Tugas Akhir ini tidak terlepas dari bantuan, bimbingan, arahan serta dukungan dari berbagai pihak. Oleh karena itu, ucapan terimakasih di sampaikan kepada :

1. Ibu Dra. Nonong Amalita, M.Si, selaku dosen pembimbing dan pembimbing akademik yang telah memberikan arahan dan bimbingan dalam penyelesaian Tugas Akhir.
2. Bapak Dr. Dony Permana, M.Si, selaku dosen penguji yang telah memberi arahan dalam penulisan Tugas Akhir.

3. Bapak Zamahsary Martha, S.Si., M.Si, selaku dosen penguji yang telah memberi arahan dalam penulisan Tugas Akhir.
4. Ibu Dr. Yenni Kurniawati, S.Si., M.Si., selaku Kepala Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang.
5. Bapak Dodi Vionanda, M.Si., Ph.D., Koordinator Program Studi Diploma III Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang.
6. Bapak dan Ibu dosen serta seluruh Tenaga Kependidikan Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang.
7. Kedua orang tua tersayang, serta keluarga besar Muchlis yang selalu memberikan kasih sayang dan *support* dalam penyelesaian Tugas Akhir.
8. Teman-teman yang telah memberikan dukungan dalam penyelesaian Tugas Akhir.

Penulisan Tugas Akhir ini masih banyak kesalahan dan kekurangan. Oleh karena itu, kritik dan saran yang bersifat membangun sangat diharapkan dalam penyempurnaan penulisan Tugas Akhir. Semoga Tugas Akhir ini dapat bermanfaat bagi seluruh pihak yang membacanya.

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Padang, 3 November 2023

Tasya Mutia

DAFTAR ISI

ABSTRAK	i
KATA PENGANTAR	iii
DAFTAR GAMBAR	vi
DAFTAR TABEL	vii
DAFTAR LAMPIRAN	vii
BAB I PENDAHULUAN	1
A. Latar Belakang Masalah	1
B. Batasan Masalah	6
C. Rumusan Masalah	6
D. Tujuan Masalah	6
E. Manfaat Penelitian	7
BAB II KAJIAN TEORI	8
BAB III METODOLOGI PENELITIAN	18
A. Jenis Penelitian	18
B. Data dan Sumber Data	18
C. Variabel Penelitian	18
D. Teknik Pengumpulan Data	19
E. Teknik Analisis Data	20
F. Diagram Alur	21
BAB IV HASIL DAN PEMBAHASAN	22
A. Analisis Sentimen	22
B. Deskripsi Data	32
C. Analisis Data	36
D. Pembahasan	43
BAB V PENUTUP	45
A. Kesimpulan	45
B. Saran	45
DAFTAR PUSTAKA	47
LAMPIRAN	50

DAFTAR GAMBAR

Gambar	Halaman
1. Persentase Hasil Survei Terhadap Bakal Calon Presiden Tahun 2024.....	3
2. Perbandingan Jumlah Penduduk dan Jumlah Pengguna <i>Twitter</i> di Indonesia....	4
3. Contoh <i>Wordcloud</i>	15
4. <i>Syntax</i> Pengumpulan Data	19
5. Menyimpan Data.....	20
6. Diagram Alur	21
7. Jumlah Sentimen Positif dan Sentimen Negatif.....	33
8. <i>Wordcloud</i> Hasil <i>Preprocessing</i>	34
9. <i>Wordcloud</i> Hasil <i>Preprocessing</i> Sentimen Positif.....	35
10. <i>Wordcloud</i> Hasil <i>Preprocessing</i> Sentimen Negatif	36

DAFTAR TABEL

Tabel	Halaman
1. Variabel Data	19
2. Hasil <i>Crawling</i> Data.....	22
3. Contoh Hasil <i>Preprocessing Remove Mention</i>	23
4. Contoh Hasil <i>Preprocessing Remove Url</i>	25
5. Contoh Hasil <i>Preprocessing Remove Hashtag</i>	26
6. Contoh Hasil <i>Preprocessing Case Folding</i>	27
7. Contoh Hasil <i>Preprocessing Stopword</i>	28
8. Contoh Hasil <i>Preprocessing Tokenizing</i>	29
9. Contoh Hasil <i>Preprocessing Steaming</i>	30
10. Perhitungan TF-IDF dengan Manual	31
11. Hasil Pelabelan <i>Tweet</i>	32
12. Hasil Pembagian Data <i>Traning</i> dan Data <i>Testing</i>	37
13. Frekuensi Kemunculan <i>Term</i> untuk Contoh	38
14. <i>Confusion Matriks</i>	41
15. Hasil <i>Confusion Matriks</i>	42

DAFTAR LAMPIRAN

Lampiran	Halaman
1. Kode Program <i>Phyton</i> untuk <i>Crawling</i> Data	50
2. Kode Program <i>Phyton Preprocessing</i> Data	50
3. Kode Program <i>Phyton</i> Proses Pelabelan Data	55
4. Kode Program <i>Phyton Wordcloud</i>	55
5. Pembobotan TF-IDF	56
6. Pemisahan Sentimen Positif dan Sentimen Negatif.....	57
7. Jumlah Kata Sentimen Positif dan Kata Sentimen Negatif.....	57
8. Pembagian Data <i>Traning</i> dan Data <i>Testing</i>	58
9. Klasifikasi <i>Naive Bayes Classifier</i>	58
10. <i>Output</i> Akurasi	58
11. <i>Output Confusion Matriks</i>	59
12. Pelabelan Data.....	59
13. <i>Term</i> Sentimen Positif.....	60
14. <i>Term</i> Sentimen Negatif	72

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Sistem pemerintahan di Indonesia menurut Pasal 4 ayat 1 Undang-Undang Dasar (UUD) 1945 menganut sistem pemerintahan presidensial. Pemilihan umum yang akan dilaksanakan pada tahun 2024 telah mengatur prosedur dan persyaratan menjadi calon Presiden dan Wakil Presiden (Sukimin, 2020). Pasal 222 Undang-Undang Republik Indonesia Nomor 7 tahun 2017 menyatakan bahwa “Pasangan calon Presiden diusulkan oleh Partai Politik atau gabungan Partai Politik peserta pemilu yang memenuhi persyaratan perolehan kursi paling sedikit 20% dari jumlah kursi Dewan Perwakilan Rakyat (DPR) atau memperoleh 25% dari suara sah secara nasional pada Pemilihan Umum anggota DPR sebelumnya”.

Presiden merupakan kepala negara dan kepala pemerintah yang memegang kendali atas kebijakan dan tindakan pemerintah serta memiliki otoritas untuk mengeluarkan peraturan-peraturan yang berlaku di negara tersebut. Masa jabatan presiden di Indonesia dalam 1 periode selama 5 tahun, namun dapat diperpanjang hingga 2 periode atau selama 10 tahun secara berturut-turut (Sya'bani dkk, 2022). Menurut Undang-Undang No.42 Tahun 2008 pemilihan Presiden dan Wakil Presiden dilaksanakan secara efektif dan efisien berdasarkan asas langsung, umum, bebas, rahasia, jujur dan adil.

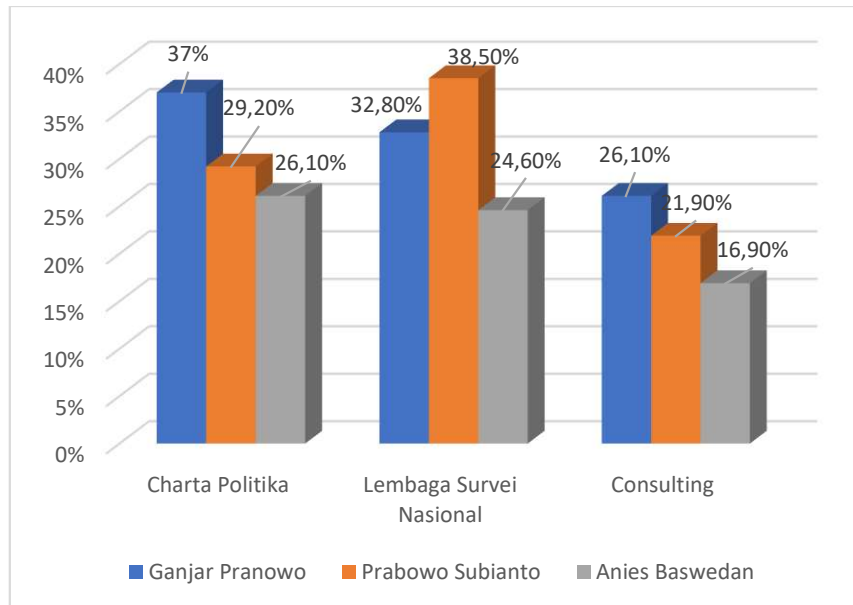
Indonesia telah dipimpin oleh beberapa orang presiden diantaranya, Dr. Ir. H. Soekarno (1945-1967), Jendral TNI H.M. Soeharto (1967-1998), Prof. Dr. Ing. B. J. Habibie (1998-1999), K. H. Abdurrahman Wahid (1999-2001), Prof. Dr. Hj. Diah

Megawati Setiawati Soekarnoputri (2001-2004), Jendral TNI Prof. Dr. Dr. H. Susilo Bambang Yudhoyono, M.A (2004-2009, 2009-2014), dan Ir. H. Joko Widodo (2014-2019, 2019-2024). Presiden Indonesia saat ini Joko Widodo memegang jabatan sebagai presiden selama 2 periode berturut-turut, berdasarkan peraturan tidak bisa dicalonkan lagi pada periode berikutnya.

Menurut Lingkaran Survei Indonesia (LSI) banyaknya lembaga survei yang merilis opini publik terkait beberapa kandidat bakal calon Presiden 2024 diantaranya yaitu Anies Baswedan, Prabowo Subianto, Ganjar Pranowo, Ridwan Kamil, Agus H. Yudhoyono, Sandiaga Uno, Puan Maharani, Khofifah Indra Parawansa, Erick Thohir, Tri Rismaharini dan Airlangga Hartarto. Beberapa kandidat bakal calon Presiden di atas terdapat 3 tokoh yang memiliki elektabilitas yang baik untuk menjadi calon Presiden Indonesia pada pemilihan Presiden tahun 2024 yaitu Ganjar Pranowo, Prabowo Subianto dan Anies Baswedan (Akmal dkk, 2022)

Beberapa lembaga survei yang merilis opini publik terhadap bakal calon presiden 2024 melalui wawancara tatap muka sebanyak 900 hingga 2000 responden masyarakat Indonesia yang minimal telah berusia 17 Tahun dan memiliki KTP. Wawancara dilakukan pada tanggal 2 Mei 2023 hingga 2 Juni 2023, dengan pengambilan responden melalui teknik pengambilan sampel secara acak sistematis (*multistage random sampling*) dengan *klaim margin of error* 2,0% hingga 4,0% pada tingkat kepercayaan (*level of confidence*) 95% (Susetio, 2022). Persentase hasil survei yang dirilis oleh beberapa lembaga survei Charta Politika, Lembaga

Survei Nasional dan *Consulting* terhadap bakal calon presiden 2024 dapat dilihat pada Gambar 1.



Sumber: Lembaga Survei Indonesia

Gambar 1. Persentase Hasil Survei Terhadap Bakal Calon Presiden Tahun 2024

Berdasarkan hasil survei yang dirilis terdapat satu tokoh bakal calon presiden tertinggi yaitu Ganjar Pranowo. Menurut Rico (2020) faktor yang menunjukan Ganjar Pranowo dijadikan sebagai calon Presiden yaitu gaya komunikasi yang relatif mudah dicerna oleh publik, sehingga mudah dikenal luas oleh masyarakat.

Survei politik melalui wawancara tatap muka memerlukan biaya yang sangat besar. Oleh karena itu, upaya yang dilakukan untuk melihat opini publik terhadap bakal calon presiden tahun 2024 melalui media sosial *twitter*. Opini dan pendapat masyarakat pada *twitter* dapat dijadikan sebagai instrumen untuk mengukur tingkat kepuasan masyarakat mengenai bakal calon presiden 2024. Berdasarkan Badan Pusat Statistika (BPS) dan *We Are Sosial*, jumlah penduduk dan pengguna aktif *twitter* di Indonesia dapat dilihat pada Gambar 2.



Gambar 2. Perbandingan Jumlah Penduduk dan Jumlah Pengguna Twitter di Indonesia

Berdasarkan Gambar 2 meningkatnya jumlah penduduk di Indonesia mengakibatkan pengguna aktif *twitter* juga meningkat. Menurut *We Are Sosial* Indonesia menjadi salah satu negara dengan pengguna *twitter* terbesar di dunia. Posisi Indonesia berada dibawah Brasil yang memiliki 19,05 juta pengguna *twitter*. Sedangkan Amerika Serikat berada di posisi pertama dengan 76,9 juta pengguna *twitter*.

Twitter merupakan salah satu media sosial yang didirikan oleh Jack Dorsey yang umumnya berfungsi untuk mengirimkan pesan atau mendapatkan berbagai informasi (*tweet*) (Yan dkk, 2020). Pengguna *twitter* akan memberi kabar terbaru atau komentar tentang hal yang sedang menjadi topik utama. Oleh karena itu, hal yang menjadi topik utama akan sering dikomentari dan menjadi *tranding* terutama di media sosial *twitter*.

Analisis sentimen adalah studi komputasi dari opini-opini, serta emosi yang diekspresikan dalam teks (Liu, 2022). Tugas dasar dalam analisis sentimen adalah

mengembangkan sistem untuk menganalisis, mengidentifikasi dan mengungkapkan pendapat dalam bentuk teks dan juga merupakan proses yang berfungsi untuk mengidentifikasi berupa opini atau sentimen dari isi *dataset* (Fikri dkk, 2020). Analisis sentimen dapat membantu memperoleh gambaran umum persepsi masyarakat dengan cara mengelompokkan jenis opini menjadi kategori negatif dan positif terhadap bakal calon presiden tahun 2024 (Darwis dkk, 2021).

Teknik pembelajaran mesin untuk analisis sentimen adalah *Naive Bayes Classifier* (NBC), *Support Vector Mechine* (SVM) dan *K-nearest Neighbor* (KNN). Pada penelitian ini algoritma yang digunakan adalah *Naive Bayes Classifier* (NBC). NBC merupakan teknik pembelajaran mesin yang berbasis probabilistik. NBC adalah metode sederhana tetapi memiliki akurasi serta performansi yang tinggi dalam pengklasifikasian teks (Routray dkk, 2013). Analisis sentimen yang dibangun menggunakan NBC memperoleh akurasi sebesar 83% dan rata-rata harmonik sebesar 90,713% (Ling dkk, 2014).

Berdasarkan penelitian Singgalen (2021) yang berjudul “Pemilihan Metode dan Algoritma dalam Analisis Sentimen di Media Sosial, Sistematis *Literature Review*” menyimpulkan Algoritma *Naive Bayes Classifier* merupakan algoritma yang cukup dominan digunakan untuk penelitian analisis sentimen pada media sosial. Pada penelitian lainnya telah dilakukan oleh Sya'bani dkk (2022), mengenai Analisis Sentimen Terhadap Bakal Calon Presiden 2024 dengan Algoritma *Naive Bayes* menghasilkan evaluasi dari algoritma *Naive Bayes* yaitu sebesar 73,68%.

Berdasarkan uraian yang telah dipaparkan, maka dilakukan penelitian analisis sentimen untuk mengetahui opini masyarakat terhadap bakal calon presiden 2024

Ganjar Pranowo, dengan penelitian yang berjudul “**Analisis Sentimen Data Twitter Terhadap Bakal Calon Presiden Ganjar Pranowo Tahun 2024 Menggunakan Metode *Naive Bayes Classifier***”.

B. Batasan Masalah

Batasan masalah dalam penelitian ini membahas tentang bakal calon presiden Ganjar Pranowo tahun 2024 dengan menggunakan *tweet* berbahasa Indonesia serta data *tweet* yang diambil berdasarkan historis *tweet* dari tanggal 21 April 2023 sampai dengan 31 Juli 2023.

C. Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah sebagai berikut.

1. Bagaimana klasifikasi opini pengguna *twitter* terhadap bakal calon presiden Ganjar Pranowo 2024 menggunakan metode *Naive Bayes Classifier* ?
2. Bagaimana tingkat akurasi algoritma *Naive Bayes Classifier* terhadap bakal calon presiden Ganjar Pranowo 2024?

D. Tujuan Masalah

Tujuan dari penelitian ini adalah sebagai berikut.

1. Untuk mengetahui klasifikasi opini pengguna *twitter* terhadap bakal calon presiden Ganjar Pranowo 2024 menggunakan metode *Naive Bayes Classifier* .
2. Untuk mengetahui tingkat akurasi algoritma *Naive Bayes Classifier* dalam pengklasifikasian sentimen pengguna *twitter* terhadap bakal calon presiden Ganjar Pranowo 2024.

E. Manfaat Penelitian

Manfaat yang didapatkan dari penelitian ini adalah sebagai berikut.

1. Bagi peneliti, menambah wawasan dan pengetahuan mengenai analisis sentimen opini masyarakat terhadap bakal calon presiden 2024 pada data twitter menggunakan algoritma *Naive Bayes Classifier*.
2. Bagi masyarakat, hasil penelitian ini dapat dimanfaatkan sebagai bahan referensi dalam memilih calon presiden Indonesia periode 2024 sampai 2029.
3. Bagi mahasiswa, sebagai referensi atau bahan bacaan untuk melakukan penelitian selanjutnya.