

SENTIMEN ANALISIS PADA TWITTER  
MENGUNAKAN NAÏVE BAYES DAN REGRESI  
LOGISTIK UNTUK PEMILU PRESIDEN 2024

SKRIPSI SARJANA KOMPUTER

Oleh

Alisya Mutia Mantika

207064516054



PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNOLOGI KOMUNIKASI DAN  
INFORMATIKA

UNIVERSITAS NASIONAL

2023/2024

**SENTIMEN ANALISIS PADA TWITTER  
MENGUNAKAN NAÏVE BAYES DAN REGRESI  
LOGISTIK UNTUK PEMILU PRESIDEN 2024**

**SKRIPSI SARJANA**

Skripsi sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer dari  
Fakultas Teknologi Komunikasi dan Informatika

Oleh

**Alisya Mutia Mantika**

**207064516054**



**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNOLOGI KOMUNIKASI DAN  
INFORMATIKA**

**UNIVERSITAS NASIONAL**

**2023/2024**

## HALAMAN PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya penulis sendiri, dan semua sumber baik yang dikutip maupun dirujuk telah penulis nyatakan dengan benar. Bilamana di kemudian hari ditemukan bahwa karya tulis ini menyalahi peraturan yang ada berkaitan etika dan kaidah penulisan karya ilmiah yang berlaku, maka penulis bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku

Yang menyatakan,

Nama : Alisya Mutia Mantika

NIM : 207064516054

Tanda Tangan :



Tanggal : 26 Februari 2024

Mengetahui

Pembimbing I : Dr. Agung Triayudi, S.Kom., M.Kom. (.....)

Pembimbing II : Rima Tamara Aldisa, S.Kom., M.Kom. (.....)

# HALAMAN PENGESAHAN

## SENTIMEN ANALISIS PADA TWITTER MENGGUNAKAN NAÏVE BAYES DAN REGRESI LOGISTIK UNTUK PEMILU PRESIDEN 2024

Oleh

Alisya Mutia Mantika

207064516054

Program Studi Informatika

Fakultas Teknologi Komunikasi dan Informatika

Universitas Nasional

Disetujui pada Tanggal: 23 Februari 2024

Pembimbing I



(Dr. Agung Triayudi, S.Kom.,  
M.Kom.)

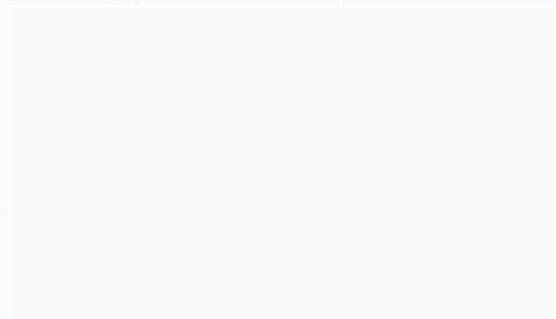
NIDN. 0107019009

Pembimbing II



(Rima Tamara Aldisa, S.Kom.,  
M.Kom.)

NIDN. 0318019401



## **HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Program Studi Informatika, Universitas Nasional, Fakultas Teknologi Komunikasi dan Informatika, saya yang bertanda tangan di bawah ini:

Nama : Alisya Mutia Mantika

NIM : 207064516054

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Nasional, Fakultas Teknologi Komunikasi dan Informatika, Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalti Free Right*) atas karya ilmiah saya yang berjudul:

### **SENTIMEN ANALISIS PADA TWITTER MENGGUNAKAN NAÏVE BAYES DAN REGRESI LOGISTIK UNTUK PEMILU PRESIDEN 2024**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak ini Universitas Nasional berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada tanggal : 23 Februari 2024

Yang menyatakan



(Alisya Mutia Mantika)

## KATA PENGANTAR

Puja, puji dan Syukur penulis panjatkan kepada Allah SWT yang senantiasa memberikan Rahmat dan karunia-Nya sehingga penulis mampu menyelesaikan skripsi dengan judul “SENTIMEN ANALISIS PADA TWITTER MENGGUNAKAN NAÏVE BAYES DAN REGRESI LOGISTIC UNTUK PEMILU PRESIDEN 2024” sebagai salah satu syarat kelulusan Program Studi Sarjana Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional.

Penelitian dan penulisan skripsi ini tidak terhindar dari bantuan berbagai pihak, oleh karena itu penulis menyampaikan banyak apresiasi dan terima kasih terutama kepada dosen pembimbing Tugas Akhir, Bapak Dr. Agung Triayudi, S.Kom., M.Kom. dan Bu Rima Tamara Aldisa, S.Kom., M.Kom. yang telah banyak membantu, membimbing, mengarahkan serta meluangkan banyak waktu, tenaga, pikiran, dan motivasi selama penelitian tugas akhir dan penyusunan skripsi ini. Penulis juga menyampaikan banyak terima kasih untuk:

1. Kedua orang tua penulis terus-menerus memberi motivasi dan menyemangati penulis.
2. Segenap staff dan dosen pengajar di Program Studi Informatika yang memberikan bantuan dan juga ilmu yang bermanfaat .
3. Teman-teman seangkatan dan sehimpuan yang berbalas-balasan dalam menyemangati dan membantu dalam pembuatan skripsi ini.
4. Seluruh keluarga besar penulis yang tidak bisa disebutkan semuanya tetapi tidak mengurangi rasa Terimakasih dan Syukur penulis.

Semoga Allah SWT membalas kebaikan dan bantuannya dengan hal yang sepadan. Dalam penulisan skripsi ini penulis paham dan sadar akan berbagai kesalahan maupun kekeliruan yang secara sengaja maupun tidak sengaja dilakukan, kritik dan sarannya akan sangat membantu penulis. Semoga skripsi ini memberikan manfaat di bidang Informatika.

Jakarta, Februari 2024



Penulis

## ABSTRAK

Sesuai dengan paham demokrasi yang menjadi dasar negara, Pemilihan umum akan dilaksanakan pada tahun 2024. Dalam pelaksanaan Pemilihan Umum terdapat kampanye untuk mengiring suara masyarakat untuk memilih kandidat yang terbaik menurut opini masyarakat. Media sosial Twitter menjadi salah satu media untuk menyuarakan opini juga berbagi informasi menjadi salah satu wadah berkampanye secara tidak langsung. Sosial media juga tidak luput dari isu-isu negatif, desas-desus masyarakat, bahkan jejak digital kandidat calon presiden yang dapat menjadi pertimbangan yang sangat penting dalam berkampanye. Penelitian ini bertujuan untuk melihat respon masyarakat terhadap kandidat calon presiden 2024. Penelitian ini dilakukan berdasarkan opini masyarakat terhadap kandidat calon presiden yang kemudian data opini masyarakat yang diambil dari media sosial Twitter akan melalui proses pre-processing untuk membersihkan data sebelum data diklasifikasikan menjadi permodelan Naive Bayes dan Regresi Linier. Kedua model klasifikasi tersebut kemudian dicari nilai akurasi kinerja tertinggi dan confusion matriks dengan data splitting 80:20. Hasil Penelitian menunjukkan bahwa model klasifikasi Naive Bayes memiliki nilai akurasi yang lebih tinggi dibandingkan model klasifikasi Regresi Logistik yaitu sebesar 63% untuk kandidat Anies Baswedan, 77% untuk kandidat Ganjar Pranowo, dan 44% untuk Prabowo Subianto. Nilai akurasi tertinggi diperoleh data sentimen kandidat calon presiden 2024 Ganjar Pranowo yaitu sebesar 77%.

Kata Kunci: Sentimen Analisis, Pemilu Presiden 2024, Regresi Logistik, Naive Bayes, Twitter



## ABSTRACT

In accordance with the notion of democracy which is the basis of the state of Indonesia, general elections will be held in 2024. In the implementation of the General Election there is a campaign to lead the public vote to choose the best candidate according to public opinion. Twitter social media is one of the media to voice opinions as well as share information to become one of the indirect campaigning platforms. Social media also does not escape negative issues, community rumors, and even the digital footprint of presidential candidates which can be a very important consideration in campaigning. This research aims to see the public's response to the 2024 presidential candidates. This research is conducted based on public opinion on presidential candidates, then public opinion data taken from Twitter social media will go through a pre-processing process to clean the data before the data is classified into Naive Bayes and Linear Regression modeling. The two classification models are then sought for the highest performance accuracy value and confusion matrix with 80:20 splitting data. The results showed that the Naive Bayes classification model had a higher accuracy value than the Logistic Regression classification model, which was 63% for Anies Baswedan candidate, 77% for Ganjar Pranowo candidate, and 44% for Prabowo Subianto. The highest accuracy value was obtained by the sentiment data of 2024 presidential candidate Ganjar Pranowo, which was 77%.

Keywords: Sentiment Analysis, 2024 Presidential Election, Logistic Regression, Naive Bayes, Twitter



## DAFTAR ISI

KATA PENGANTAR .....	vi
ABSTRAK .....	vii
ABSTRACT .....	viii
DAFTAR ISI .....	ix
DAFTAR GAMBAR .....	xi
DAFTAR TABEL .....	xii
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Identifikasi Masalah .....	2
1.3 Tujuan Penelitian .....	3
1.4 Manfaat Penelitian .....	3
1.5 Batasan Masalah .....	3
BAB II TINJAUAN PUSTAKA .....	5
2.1 Kajian Teori .....	5
2.1.1 Sentimen Analisis .....	5
2.1.2 Naïve Bayes .....	5
2.1.3 Regresi Logistik .....	6
2.1.4 Twitter Harvest .....	6
2.1.5 Confusion Matrix .....	7
2.1.6 Crawling Data .....	7
2.1.7 Data Preprocessing .....	8
2.2 Studi Literatur .....	9
BAB III METODE PENELITIAN .....	11
3.1 Waktu Penelitian .....	11
3.2 Alat dan Bahan .....	11
3.3 Fokus Penelitian .....	11
3.4 Sumber Data .....	12
3.5 Desain Penelitian .....	12
3.5.1 Tahapan Penelitian .....	12
BAB IV HASIL DAN PEMBAHASAN .....	14
4.1 Hasil Penelitian .....	14
4.1.1 Data Crawling .....	14

4.1.2	Data Pre-processing .....	16
4.2	Analisis menggunakan Algoritma Naïve Bayes dan Regresi Logistik .	22
4.3	Interface Dashboard Streamlit.....	30
4.4	Hasil Akhir .....	32
BAB VI KESIMPULAN DAN SARAN .....		33
5.1	Kesimpulan .....	33
5.2	Saran.....	33
DAFTAR PUSTAKA .....		34
LAMPIRAN.....		37



## DAFTAR GAMBAR

Gambar 2.1 Confusion Matriks Multiclass .....	7
Gambar 3.1 Tabel Data Twitter .....	12
Gambar 3.2 Tahapan Penelitian .....	13
Gambar 4.1 Skema alur data crawling dengan Twitter Harvest .....	15
Gambar 4.2 Alur Preprocessing Data.....	16
Gambar 4.3 Data sebelum dipreprocessing.....	17
Gambar 4.4 Data setelah filtering .....	17
Gambar 4.5 Data setelah dibersihkan.....	18
Gambar 4.6 Data setelah di stemming .....	18
Gambar 4.7 Data setelah tokenizing .....	19
Gambar 4.8 Hasil Labeling menggunakan model BERT.....	19
Gambar 4.9 Mengubah Labeling menjadi nilai sentimen .....	19
Gambar 4.9 Hasil Data Labeling.....	20
Gambar 4.10 Diagram Pie Data Sentimen Anies Baswedan .....	20
Gambar 4.11 Banyaknya Data Sentimen Anies Baswedan .....	20
Gambar 4.12 Diagram Pie Data Sentimen Ganjar Pranowo .....	21
Gambar 4.13 Banyaknya Data Sentimen Ganjar Pranowo .....	21
Gambar 4.14 Diagram Pie Data Sentimen Prabowo Subianto.....	21
Gambar 4.15 Banyaknya Data Sentimen Prabowo Subianto.....	22
Gambar 4.16 Data Splitting.....	22
Gambar 4.18 Hasil Klasifikasi Regresi Logistik Data Sentimen Anies B. ....	24
Gambar 4.19 Confusion Matriks Naïve Bayes Anies Baswedan.....	24
Gambar 4.20 Confusion Matriks Regresi Logistik Anies Baswedan .....	24
Gambar 4.21 Hasil Klasifikasi Naïve Bayes Data Sentimen Ganjar Pranowo .....	26
Gambar 4.22 Hasil Klasifikasi Regresi Logistik Data Sentimen Ganjar Pranowo.....	26
Gambar 4.23 Confusion Matrix Naïve Bayes Ganjar Pranowo .....	27
Gambar 4.24 Confusion Matrix .....	27
Gambar 4.25 Hasil Klasifikasi Naïve Bayes Data Sentimen Prabowo Subianto..	28
Gambar 4.26 Hasil Klasifikasi Regresi Logistik Data Sentimen Prabowo S .....	29
Gambar 4.27 Confusion Matrix Naïve Bayes Prabowo Subianto.....	29
Gambar 4.28 Confusion Matrix Regresi Logistik Prabowo Subianto .....	29
Gambar 4.29 Tampilan Halaman Hasil crawling data.....	30
Gambar 4.30 Tampilan Halaman Hasil Preprocessing data.....	31
Gambar 4.31 Tampilan hasil Klasifikasi Algoritma .....	31
Gambar 4.32 Tampilan Hasil Akhir.....	32

## DAFTAR TABEL

Tabel 3.1 Waktu Penelitian.....	11
Tabel 4.2 Laporan Klasifikasi Data Sentimen Ganjar Pranowo .....	25
Tabel 4.3. Laporan Klasifikasi Data Sentimen Prabowo Subianto.....	27
Tabel 4.4 Nilai Akurasi dari permodelan Naïve Bayes dan Regresi Logistik .....	32



## BAB I PENDAHULUAN

### 1.1 Latar Belakang

Indonesia sebagai negara yang berpaham demokrasi sesuai dengan Pasal 1 ayat 2 UUD 1945 yang menyatakan bahwa “Kedaulatan berada di tangan rakyat dan dilaksanakan menurut Undang-Undang Dasar”. Pemilihan Presiden dan wakil presiden yang dilangsungkan setiap 5 tahun sekali merupakan salah satu contoh diselenggarakannya demokrasi lewat suara atau pilihan rakyat yang menjadi penentu lewat pemungutan suara yang dilaksanakan secara simultan (Delvika et al., 2023).

Pemilihan tokoh calon presiden dan wakil presiden pastinya akan mempertimbangkan opini dan respon dari masyarakat terhadap setiap aspek seperti gagasan, visi misi, program yang akan dilaksanakan, masalah yang akan dipecahkan, dan lainnya. Dari opini tersebut akan tercipta survei lewat popularitas atau kecenderungan masyarakat sehingga tercipta pro dan kontra terkait calon presiden dan wakil presiden yang menjadi acuan dalam memilih kandidat yang tepat.

Pada era digital dimana teknologi informasi dan media sosial sudah sangat berkembang dan menjadi hal yang penting dalam sehari-hari masyarakat Indonesia juga dimanfaatkan dalam penyelenggaraan kampanye dan beropini (Syafii, 2023). Masyarakat Indonesia memanfaatkan media sosial sebagai media mendapatkan informasi, berkampanye, dan beropini secara publik.

*Twitter* menjadi salah satu media sosial yang kerap digunakan untuk mengemukakan pendapat, berkampanye, mendapatkan informasi terkini, membahas topik yang sedang hangat, dan berdiskusi secara objektif karena *Twitter* membatasi penulisan *tweets* hanya 280 kata (Vindua & Zailani, 2023). Indonesia berada di peringkat ke 5 pengguna media sosial *Twitter* terbanyak di dunia, atau sekitar 14,8 Juta pengguna. Data di media sosial *Twitter* berupa teks bisa diolah menjadi informasi yang berguna (Wenando et al., 2020).

Salah satu topik yang saat ini kerap dibahas adalah masalah pemilihan Presiden, baik tentang politik maupun aktifitas kandidatnya. Media sosial *Twitter* sebagai penyedia data opini universal dan juga menjadi sarana untuk ikut terlibat aktif dalam demokrasi serta mendukung calon presiden pilihan (Zain et al., 2023). Sebab *Twitter* merupakan media yang menampung aspirasi atau opini masyarakat tak jarang tweets atau konten yang ada bersifat negatif atau berisi cacian, makian maupun kata kata kasar.

Sentimen analisis adalah proses klasifikasi atau analisa opini, sentiment dan emosi individu yang dituangkan kedalam bentuk teks. Sentimen analisis dilakukan untuk menentukan apakah teks tersebut cenderung berkonotasi negative atau positif (Asmara et al., 2020). Lazimnya sentimen analisis dimanfaatkan untuk kepentingan politik, pemerintahan, Pendidikan, bisnis dan lainnya (Muzaki & Witanti, 2021).

Algoritma yang digunakan adalah algoritma *Naïve bayes* dan *Regresi Logistik*. Algoritma Naïve Bayes berfungsi untuk menghitung probabilitas suatu teks atau dokumen dalam setiap kategori sentimen dengan asumsi kata-kata dalam teks yang diproses bersifat independen, sedangkan Algoritma regresi logistik berfungsi untuk memprediksi probabilitas teks terhadap kelas sentimennya (Averina et al., 2022).

Regresi Logistik adalah salah satu algoritma klasifikasi di pembelajaran mesin digunakan untuk memprediksi probabilitas variabel dependen kategori. Metode ini merupakan bentuk umum regresi linier yang digunakan untuk mempelajari hubungan beberapa variabel dengan variabel biner atau probabilitistik (Bahtiar et al., 2023). Pada penelitian ini, Regresi logistik mengklasifikasikan data menjadi 3 kelas yaitu positif, netral dan negatif menggunakan regresi logistik multinomial.

## 1.2 Identifikasi Masalah

Berdasarkan latar belakang tersebut terdapat identifikasi masalah yang akan di bahas pada penelitian ini, ialah:



- Kegiatan berkampanye masih sering menggunakan kata-kata kasar dan saling menjatuhkan kandidat lainnya.
- Media social sering digunakan untuk mencaci, memaki, dan menjatuhkan satu sama lain.
- Masih banyak masyarakat yang pasif dalam demokrasi
- Semakin dekat dengan hari pemilu semakin gencar juga para pendukung kandidat tertentu sehingga mencoba segala cara untuk menjatuhkan atau membuat rumor tentang kandidat lain.

### 1.3 Tujuan Penelitian

Tujuan Penelitian ini ialah:

1. Menampilkan sentimen masyarakat terhadap pemilu presiden
2. Melihat respon masyarakat terhadap setiap kandidat calon presiden 2024
3. Menghasilkan klasifikasi terbaik dalam menganalisa sentiment Masyarakat pada pemilu 2024

### 1.4 Manfaat Penelitian

Manfaat dari penelitian ini antara lain:

1. Menghasilkan data negatif, positif dan netral dari opini masyarakat sehingga bisa dijadikan acuan dalam mengadakan kampanye atau pemilihan umum tanpa menggunjing atau mengadu kandidat presiden 2024.
2. Untuk memberikan data dari kampanye atau kandidat yang lebih unggul
3. Untuk melihat partisipasi masyarakat pada pemilu 2024

### 1.5 Batasan Masalah

Penelitian ini dibatasi pada:

1. Penelitian ini menggunakan data pada *Twitter*
2. Data yang diambil dari *Twitter* berupa *tweets*/konten yang berkaitan dengan calon presiden atau dengan hashtag yang berkaitan.



3. Penelitian ini menggunakan algoritma *Naïve Bayes* dan *Regresi Logistik* untuk mengolah data.
4. Penelitian ini menggunakan bahasa pemrograman *Python* dan menggunakan perangkat lunak Google colab



## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Kajian Teori

##### 2.1.1 Sentimen Analisis

Sentimen analisis adalah proses klasifikasi atau analisa opini, sentiment dan emosi individu yang dituangkan kedalam bentuk teks. Sentimen analisis dilakukan untuk menentukan apakah teks tersebut cenderung berkonotasi negatif atau positif (Asmara et al., 2020). Lazimnya sentimen analisis dimanfaatkan untuk kepentingan politik, pemerintahan, pendidikan, bisnis dan lainnya (Muzaki & Witanti, 2021).

Sentimen analisis dipergunakan untuk mengakumulasi opini masyarakat akan suatu isu sehingga data opini masyarakat tersebut dapat diklasifikasikan dan dimanfaatkan sesuai dengan kebutuhan. Sentimen analisis juga dimanfaatkan untuk memantau pandangan masyarakat terhadap aktivitas politik seperti mengamati sikap masyarakat terhadap kandidat calon presiden. Sentimen analisis mengklasifikasikan data dalam beberapa bagian menurut kategori positif, negatif atau netral (Buntoro et al., 2021).

##### 2.1.2 Naïve Bayes

Algoritma *Naïve Bayes* merupakan algoritma klasifikasi berbentuk probabilistik dan statistik yang berasumsi setiap atribut bersifat bebas (*independence*) atau ciri dari sebuah kelas tidak ada hubungannya dengan kelas lainnya (Fikri et al., 2020).

Berikut persamaan dari teorema Bayes (Fikri et al., 2020):

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \quad (1)$$

A, B = events

P(A|B) = probabilitas A jika B adalah True

P(B|A) = probabilitas B jika A adalah True

$P(A), P(B)$  = probabilitas kejadian A dan B

### 2.1.3 Regresi Logistik

Regresi Logistik adalah model statistik yang digunakan untuk menentukan apakah sebuah independent variable memiliki pengaruh terhadap sebuah binary dependent variable. Dengan menggunakan sigmoid function, Regresi Logistik menghasilkan output sebuah probabilitas antara angka 0 dan 1 (Husen et al., 2023).

Regresi Logistik mampu mengklasifikasikan sentimen menjadi dua kelas dengan label positif dan negatif atau beberapa kelas menggunakan regresi logistik multinomial. Dalam penelitian ini, regresi logistik berperan melatih dan menguji data yang sudah diklasifikasikan menjadi 3 kelas (positif, netral dan negatif) (Imamah & Rachman, 2020).

Persamaan Regresi Logistik dinyatakan sebagai berikut (Prabowo et al., 2023):

$$g_j(x) = \beta_{j0} + \beta_{j1}x_1 + \beta_{j2}x_2 + \dots + \beta_{jp}x_p \quad (2)$$

Keterangan:

$g(x)$  = Fungsi logit untuk kelas j

$\beta_{j0}$  = intercept untuk kelas j

$\beta_{j1}, \beta_{j2}, \dots, \beta_{jp}$  = koefisien untuk masing-masing variabel prediktor ( $x_1, x_2, \dots, x_p$ ) pada kelas j

### 2.1.4 Twitter Harvest

*Tweet Harvest* merupakan alat baris perintah (*command-line tool*) yang menggunakan *Playwright* untuk mengikis (*Scrape*) tweet dari hasil pencarian Twitter berdasarkan kata kunci dan rentang tanggal tertentu. Tweet yang diambil akan disimpan ke dalam file CSV. Skrip *Twitter harvest* hanya untuk tujuan Pendidikan. Twitter melarang

pengguna yang tidak diautentikasi melakukan pencarian atau pencarian lanjutan.

Untuk menggunakan skrip ini, Anda harus memiliki akun Twitter yang valid dan mendapatkan Token Akses, yang dapat diperoleh dengan masuk ke Twitter di browser Anda dan mengekstrak cookie `auth_token`. Tweet Harvest akan membuka browser Chromium dan menavigasi ke halaman pencarian Twitter.

Crawl data Twitter dengan Tweet harvest akan menimbulkan error jika data yang diambil lebih dari rate limit perharinya (Helmi Satria, 2023).

#### 2.1.5 Confusion Matrix

*Confusion Matrix* merupakan pengukuran performa algoritma klasifikasi machine learning yang akan mengukur Precision, Precision, Accuracy dan F1 Score. Evaluasi performa dilakukan dengan confusion matrix yang mengukur tingkat True Positive (TP rate), tingkat True Negative (TN rate), tingkat False Positive (FP rate), dan tingkat False Negative (FN rate).

	<b>Predicted Negatif</b>	<b>Predicted Positif</b>	<b>Predicted Netral</b>
<b>Aktual Negatif</b>	True Negatif(TN)	False Positif (FP)	False Positif (FP)
<b>Aktual Positif</b>	True Negatif (TN)	True Positif(TP)	False Negatif (TN)
<b>Aktual Netral</b>	False Negatif (FN)	False Positif (FP)	True Negatif (TN)

Gambar 2.1 Confusion Matriks Multiclass

#### 2.1.6 Crawling Data

Dalam pengambilan atau crawling data twitter, penelitian ini menggunakan library *tweet harvest* lewat *Google Colaboratory* sebagai platform *cloud computing* yang mirip seperti *Jupyter Notebook*. Untuk

*crawling* data pada twitter juga diperlukan token autentikasi (*authentication token*). Token autentikasi berguna untuk mengakses twitter yang kita sehingga *tweet harvest* dapat melakukan *crawling* data lewat akun *twitter* pribadi. Pada penelitian ini keyword yang digunakan untuk mengambil data ialah Ganjar Pranowo, Anies Baswedan dan Prabowo Subianto.

#### 2.1.7 Data Preprocessing

Data Preprocessing ialah proses mengolah data asli (*Raw Data*) menjadi data yang mudah dipahami atau mudah diklasifikasikan (Fathir et al., 2023). Teks pre-processing diperlukan untuk analisis teks mining melalui beberapa tahapan yaitu: *casefolding*, *tokenizing*, *filtering*, *stemming*, *tagging*, dan *analysis* (Samsir et al., 2021). Berikut penjelasan tahap-tahap pre-processing (Ramadhani & Wahyudin, 2022):

- a. *Filtering* adalah tahapan pre-processing yang menyaring elemen tertentu yang tidak relevan, seperti url, tanda baca, emoji, dan stopwords (kata yang sering muncul yang tidak memiliki nilai informasi yang signifikan).
- b. *Casefolding* adalah tahapan pre-processing yang mengubah huruf kapital pada teks menjadi huruf kecil.
- c. *Stemming* adalah tahapan pre-processing yang menghilangkan imbuhan atau afiks (akhiran atau awalan) dari sebuah kata sehingga kata tersebut menjadi kata dasarnya.
- d. *Penghapusan stopwords* adalah proses menghilangkan kata yang tidak memiliki nilai sentimen.
- e. *Tokenizing* adalah tahapan pre-processing pemecahan data teks.
- f. *Data Labeling* adalah proses memberikan kategori kepada teks sentimen sehingga permodelan dapat belajar dari data tersebut.

g. *Data Splitting* adalah proses membagi dataset untuk keperluan pelatihan dan pengujian model algoritma.

## 2.2 Studi Literatur

Pada Penelitian (Husen et al., 2023), Penelitian ini menggunakan 3 Algoritma komparasi yaitu SVM, Regresi Logistik dan Naïve bayes. Dari beberapa splitting data akurasi tertinggi dicapai dengan splitting data 90:10, Algoritma SVM memiliki akurasi tertinggi sebesar 88% sedangkan Algoritma Regresi Logistik memiliki akurasi sebesar 86% dan Algoritma Naïve Bayes memiliki akurasi sebesar 76%. Sehingga Algoritma Regresi Logistik memiliki akurasi tertinggi dibandingkan Naïve Bayes.

Pada Penelitian (Imamah & Rachman, 2020), Penelitian ini menggunakan Algoritma Regresi Logistik dan TF-IDF dalam memproses datanya memiliki tingkat keakuratan 94.71%.

Pada Penelitian (Assiroj et al., 2023), Penelitian ini menggunakan Algoritma Naïve Bayes, Support Vector Machine dan Regresi Logistik dalam menganalisa akurasi data. Regresi logistik memiliki akurasi model lebih tinggi daripada algoritma lainnya yaitu sebesar 77% sedangkan Algoritma Support Vector Machine memiliki akurasi sebesar 76% dan Algoritma Naïve Bayes memiliki akurasi sebesar 70%. Kesimpulannya pada penelitian ini Regresi Logistik lebih unggul dibandingkan Algoritma SVM dan Naïve Bayes.

Pada Penelitian (Romadhon & Kurniawan, 2021), Penelitian ini menggunakan Algoritma KNN, Regresi Logistik dan Naïve Bayes dalam proses pengolahan data dimana KNN memiliki akurasi tertinggi dengan nilai 0.750 sedangkan Regresi logistik dan Naïve bayes memiliki nilai yang sama yaitu 0.708. Berdasarkan Nilai Presisinya, KNN memiliki presisi tertinggi dengan nilai 0.750 sedangkan Regresi logistik dan Naïve bayes memiliki nilai yang sama yaitu 0.700. Kesimpulannya pada penelitian ini



Algoritma KNN memiliki nilai akurasi Tertinggi dibandingkan Regresi Logistik dan Naïve Bayes.

Pada Penelitian (Utami et al., 2021), Penelitian ini menggunakan Algoritma Neural Network, Naïve Bayes dan Regresi logistik dalam mencari akurasi tertinggi dan AUC dari data diabetes. Hasil dari pengujian model adalah Regresi Logistik memiliki nilai akurasi dan AUC tertinggi yaitu 75.78% dan AUC 0.801 sedangkan Naïve Bayes memiliki nilai akurasi 74.87% serta AUC 0.799 dan Neural Network memiliki nilai akurasi 69.27% serta AUC 0.736.

Pada Penelitian (Das et al., 2023), Penelitian ini menganalisa performa algoritma Regresi Logistik, Naïve Bayes, KNN, Decision Tree, Random Forest, dan SVM dalam mendeteksi Hate Speech pada Twitter. Dari algoritma yang digunakan pada penelitian ini, Algoritma Random Forest, Decision Tree dan Support Vector Machine memiliki akurasi tertinggi yaitu sebesar 98.2%, 96.2%, dan 95.5%. Sedangkan Naïve Bayes dan Regresi Logistik memiliki nilai akurasi terendah yaitu sebesar 87.7% dan 71.7%.



## BAB III METODE PENELITIAN

### 3.1 Waktu Penelitian

*Tabel 3.1 Waktu Penelitian*

No	Kegiatan	Waktu Pelaksanaan																			
		Oktober				November				Desember				Januari				Februari			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Penentuan Topik																				
2	Studi Literatur																				
3	Pengumpulan Data																				
4	Perancangan Algoritma																				
6	Seminar Proposal																				
7	Revisi																				
8	Perancangan Sistem Lanjutan																				
9	Pengujian Sistem																				
10	Penyusunan Laporan																				
11	Sidang Akhir																				

### 3.2 Alat dan Bahan

Laptop Merk Acer Swift SF114-34

Spesifikasi :

- Sistem Operasi Windows 11 64 bit
- Processor Intel® Pentium® Silver N6000 @ 1.10GHz 1.11 GHz
- Ram 4GB
- Microsoft Excel 2023
- Google Colaboratory Python
- Data set Twitter

### 3.3 Fokus Penelitian

Penelitian ini berfokus pada menentukan algoritma mana yang lebih unggul diantara algoritma Naïve Bayes dan Regresi Logistik dengan menganalisis sentiment Masyarakat Indonesia dengan menganalisa karakteristik (tingkat akurasi, presisi dan recall) data tweets masyarakat Indonesia tentang kandidat calon presiden Indonesia sehingga dapat

membantu atau bermanfaat pada pelaksanaan kampanye dan penyelenggaraan pemilu 2024.

### 3.4 Sumber Data

Data yang didapatkan diambil dengan library tweet harvest yang mengambil 512 data tweets berkaitan dengan Ganjar Pranowo, 509 data tweets berkaitan dengan Anies Baswedan dan 552 data tweets berkaitan dengan Prabowo Subianto. Data Tweets tersebut diambil pada bulan November 2023 dengan 12 Atribut yang meliputi created\_at, id\_str, full\_text, quote\_count, reply\_count, retweet\_count, favorite\_count, lang, user\_id, conversation\_id\_str, username, dan tweet\_url.

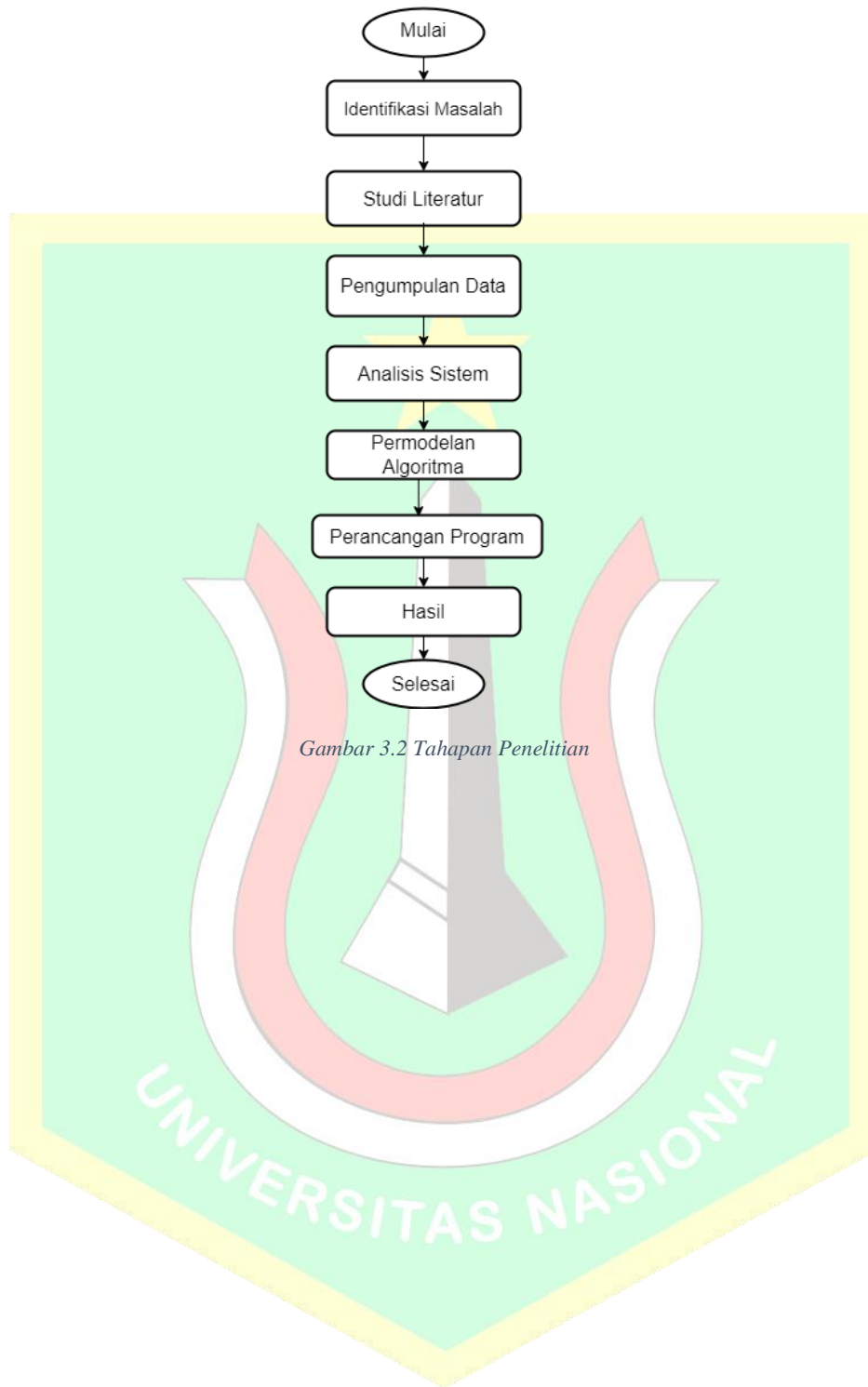
Created_at	id_str	full_text	Quote_count	Reply_count	retweet_count	Favorite_count	lang	user_id	Conversation_id_str	username	tweet_url
Tue Nov 07 05:04:48 +0000 2023	1,72E+18	@Koriandri @aniesbaswedan @mohmahfudmd Beneran blum tau? Kalo Anies bukankah petugas partai Nasdem? : siapa presidennya? Anies... apa partainya? Nasdem. Ya, kaleee. Kalo Mahfud MD sih, non partai tapi sebelum lahir pun udah Nahdliyyin. Kalo blum tau, rajin! baca bro <a href="https://t.co/UzEgZVO9v">https://t.co/UzEgZVO9v</a>	0	0	0	0	in	1,23E+18	1,72E+18	Bentir16	<a href="https://twitter.com/Bentir16/status/1721755558393557487">https://twitter.com/Bentir16/status/1721755558393557487</a>
Tue Nov 07 05:02:38 +0000 2023	1,72E+18	Ini membahas lapangan hasil karya anak bangsa yg dibangun masia kepemimpinan mas @aniesbaswedan di DKI, Buka babas @OldTraffordNews ,	0	0	0	0	in	1,44E+18	1,72E+18	choymarkochoy	<a href="https://twitter.com/choymarkochoy/status/1721755013746471308">https://twitter.com/choymarkochoy/status/1721755013746471308</a>
Tue Nov 07 05:02:43 +0000 2023	1,72E+18	HANYA GANJAR YANG KONSISTEN TOLAK ISRAEL DATANG KE INDONESIA #ganjarmahfud2024 #GanjarMahfud #sagademokrasi #paktelur #anakpresiden #ganjarpranowo #aniesbaswedan #cakimin #prabowo <a href="https://t.co/Zw4GQSsamm">https://t.co/Zw4GQSsamm</a>	0	0	0	0	in	1,54E+18	1,72E+18	Ketik_salah24	<a href="https://twitter.com/Ketik_salah24/status/1721755034290106622">https://twitter.com/Ketik_salah24/status/1721755034290106622</a>

Gambar 3.1 Tabel Data Twitter

### 3.5 Desain Penelitian

#### 3.5.1 Tahapan Penelitian

Penelitian ini dilakukan dengan tahapan *flowchart* berikut:



*Gambar 3.2 Tahapan Penelitian*

## BAB IV HASIL DAN PEMBAHASAN

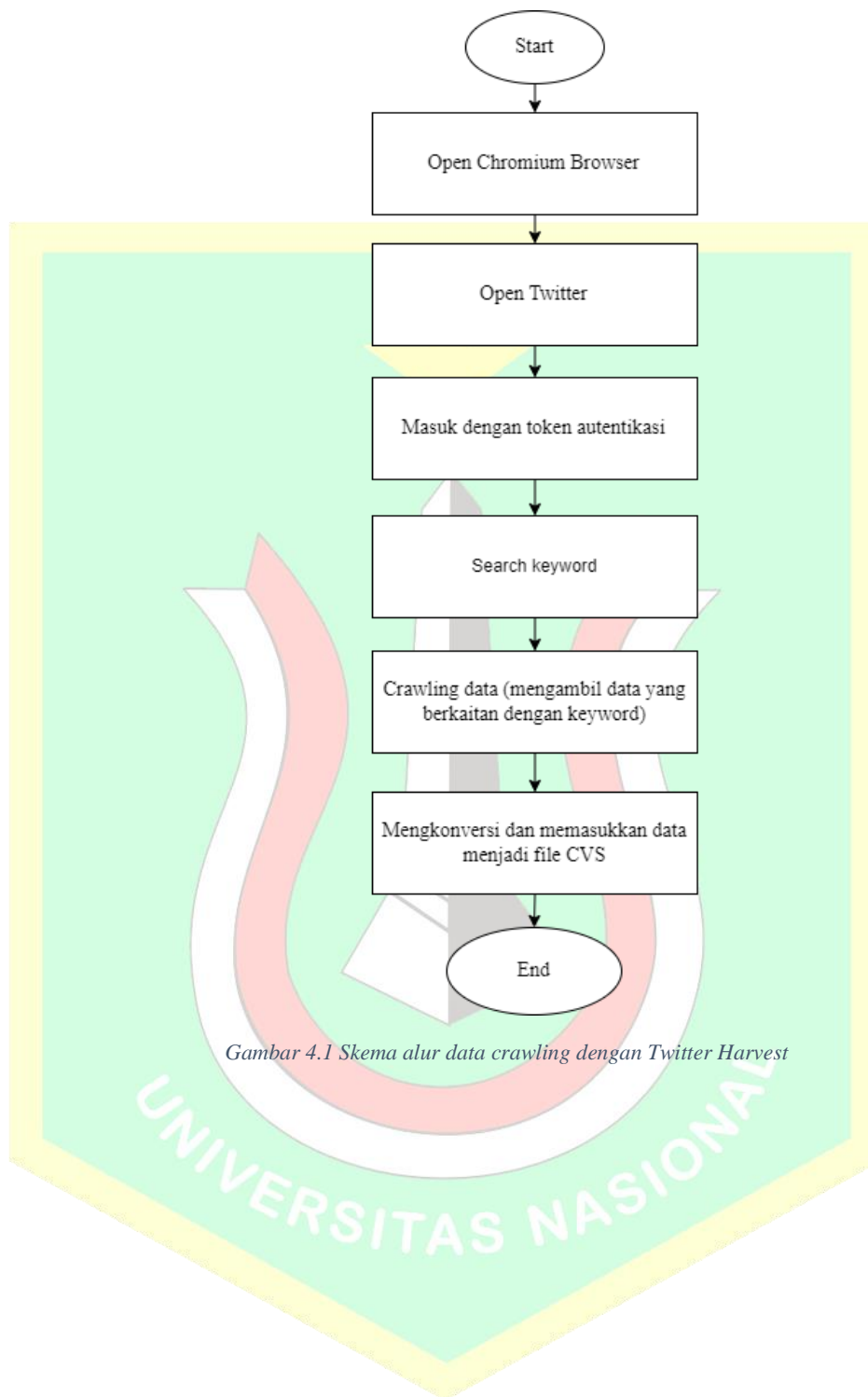
### 4.1 Hasil Penelitian

#### 4.1.1 Data Crawling

Penelitian ini menggunakan *Twitter Harvest* untuk mengumpulkan data dari Twitter. Data yang didapat berjumlah 1.575 data dan memiliki 12 Atribut yang meliputi *created\_at*, *id\_str*, *full\_text*, *quote\_count*, *reply\_count*, *retweet\_count*, *favorite\_count*, *lang*, *user\_id*, *conversation\_id\_str*, *username*, dan *tweet\_url* yang diperoleh pada bulan November 2023. Data yang didapatkan terbagi menjadi 3 sesuai dengan nama kandidat calon presiden.

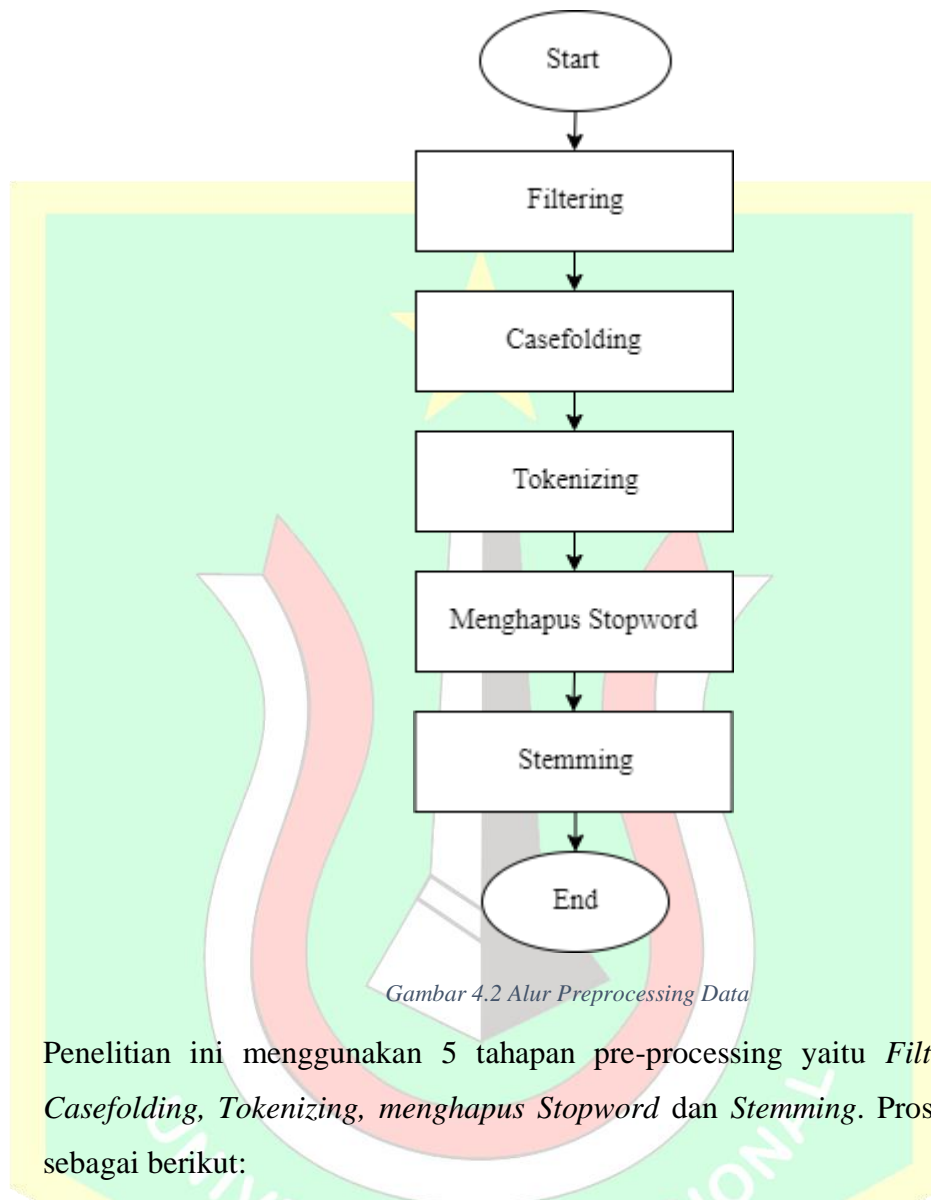
Dengan keyword Anies Baswedan, data yang didapat berjumlah 509 data. Dengan keyword Ganjar Pranowo, data yang didapat berjumlah 512 data. Dengan keyword Prabowo, data yang didapat berjumlah 552 data.

Dalam proses data crawling ini di perlukan token autentikasi (*authentication token*) pada akun kita untuk dilakukan proses crawling. Twitter Harvest juga memiliki rate limit untuk membatasi pengambilan data dan menghindari penyalahgunaan.



*Gambar 4.1 Skema alur data crawling dengan Twitter Harvest*

#### 4.1.2 Data Pre-processing



Penelitian ini menggunakan 5 tahapan pre-processing yaitu *Filtering*, *Casefolding*, *Tokenizing*, *menghapus Stopword* dan *Stemming*. Prosesnya sebagai berikut:

1. *Filtering Data*

Pada tahap *filtering data* teks yang belum bersih dari karakter yang tidak digunakan seperti *mention*, *hashtag*, emoji, *url*, dan tanda baca dihapuskan sehingga teks yang diperoleh hanya teks sentimennya saja. Data yang asli atau data mentah seperti di Gambar 4.3.

index	full_text
0	@Koriandri @aniesbaswedan @mohmahfudmd Beneran blum tau? Kalo Anies bukankah petugas partai Nasdem? : siapa presidennya? Anies...apa partainya? Nasdem. Ya, kaleee. Kalo Mahfud MD sih, non partai tapi sebelum lahir pun sudah Nahdliyin. Kalo blum tau, rajin2 baca bro. <a href="https://t.co/UzrEgZVO9v">https://t.co/UzrEgZVO9v</a>
1	Ini membahas lapangan hasil karya anak bangsa yg dibangun masa kepemimpinan mas @aniesbaswedan di DKI, Buka bahas @OldTraffordNews ,
2	HANYA GANJAR YANG KONSISTEN TOLAK ISRAEL DATANG KE INDONESIA #ganjarmahfud2024 #GanjarMahfud #jagademokrasi #paktelur #anakpresiden #ganjarpranowo #aniesbaswedan #cakimin #prabowo <a href="https://t.co/Zw4GQSSamu">https://t.co/Zw4GQSSamu</a>
3	@Tita83079013 Tapi aku ragu ketiga calon bahas ekonomi syariah melulu @ganjarpranowo @aniesbaswedan @prabowo tidak ada ekonomi kerakyatan Pancasila
4	@DPP_PKB @aniesbaswedan @cakimiNOW Bbm gratis jadi ??? Taik

Gambar 4.3 Data sebelum dipreprocessing

Kemudian *filtering data* dilakukan kepada atribut full\_text.

index	full_text	clean_text
0	@Koriandri @aniesbaswedan @mohmahfudmd Beneran blum tau? Kalo Anies bukankah petugas partai Nasdem? : siapa presidennya? Anies...apa partainya? Nasdem. Ya, kaleee. Kalo Mahfud MD sih, non partai tapi sebelum lahir pun sudah Nahdliyin. Kalo blum tau, rajin2 baca bro. <a href="https://t.co/UzrEgZVO9v">https://t.co/UzrEgZVO9v</a>	Beneran blum tau? Kalo Anies bukankah petugas partai Nasdem? : siapa presidennya? Anies...apa partainya? Nasdem. Ya, kaleee. Kalo Mahfud MD sih, non partai tapi sebelum lahir pun sudah Nahdliyin. Kalo blum tau, rajin2 baca bro.
1	Ini membahas lapangan hasil karya anak bangsa yg dibangun masa kepemimpinan mas @aniesbaswedan di DKI, Buka bahas @OldTraffordNews ,	Ini membahas lapangan hasil karya anak bangsa yg dibangun masa kepemimpinan mas di DKI, Buka bahas ,
2	HANYA GANJAR YANG KONSISTEN TOLAK ISRAEL DATANG KE INDONESIA #ganjarmahfud2024 #GanjarMahfud #jagademokrasi #paktelur #anakpresiden #ganjarpranowo #aniesbaswedan #cakimin #prabowo <a href="https://t.co/Zw4GQSSamu">https://t.co/Zw4GQSSamu</a>	HANYA GANJAR YANG KONSISTEN TOLAK ISRAEL DATANG KE INDONESIA
3	@Tita83079013 Tapi aku ragu ketiga calon bahas ekonomi syariah melulu @ganjarpranowo @aniesbaswedan @prabowo tidak ada ekonomi kerakyatan Pancasila	Tapi aku ragu ketiga calon bahas ekonomi syariah melulu tidak ada ekonomi kerakyatan Pancasila
4	@DPP_PKB @aniesbaswedan @cakimiNOW Bbm gratis jadi ??? Taik	Bbm gratis jadi ??? Taik

Gambar 4.4 Data setelah filtering

Pada Gambar 4.3, Data teks masih memiliki karakter yang membuat teks sulit di proses. Pada Gambar 4.4, Data teks sudah bersih dari *mention*, *hashtag*, *url* dan karakter lain yang tidak digunakan. Pada proses filtering juga terdapat penghapusan data nan atau data kosong.

## 2. Casefolding



index	full_text	clean_text
0	@Koriandri @aniesbaswedan @mohmahfudmd Beneran blum tau? Kalo Anies bukannya petugas partai Nasdem? : siapa presidennya? Anies...apa partainya? Nasdem. Ya, kaleeee. Kalo Mahfud MD sih, non partai tapi sebelum lahir pun sudah Nahdliyin. Kalo blum tau, rajin2 baca bro. <a href="https://t.co/UzrEgZVO9v">https://t.co/UzrEgZVO9v</a>	beneran blum tau kalo anies bukannya petugas partai nasdem siapa presidennya anies apa partainya nasdem ya kaleeee kalo mahfud md sih non partai tapi sebelum lahir pun sudah nahdliyin kalo blum tau rajin baca bro
1	Ini membahas lapangan hasil karya anak bangsa yg dibangun masa kepemimpinan mas @aniesbaswedan di DKI, Buka bahas @OldTraffordNews ,	ini membahas lapangan hasil karya anak bangsa yg dibangun masa kepemimpinan mas di dki buka bahas
2	HANYA GANJAR YANG KONSISTEN TOLAK ISRAEL DATANG KE INDONESIA #ganjarmahfud2024 #GanjarMahfud #jagademokrasi #paktelur #anakpresiden #ganjarpranowo #aniesbaswedan #cakimin #prabowo <a href="https://t.co/Zw4GQSSamu">https://t.co/Zw4GQSSamu</a>	hanya ganjar yang konsisten tolak israel datang ke indonesia
3	@Tita83079013 Tapi aku ragu ketiga calon bahas ekonomi syariah melulu @ganjarpranowo @aniesbaswedan @prabowo tidak ada ekonomi kerakyatan Pancasila	tapi aku ragu ketiga calon bahas ekonomi syariah melulu tidak ada ekonomi kerakyatan pancasila
4	@DPP_PKB @aniesbaswedan @cakimiNOW Bbm gratis jadi ??? Taik	bbm gratis jadi taik

Gambar 4.5 Data setelah dibersihkan

Pada Gambar 4.5, data yang sudah *dicleaning data* kemudian dibersihkan dari tanda baca, digit, ruang kosong (*extra whitespace*) dan huruf kapital diubah menjadi huruf kecil.

### 3. Stemming

Stemming mengubah kata-kata yang memiliki imbuhan menjadi kata dasarnya.

clean_text	stemword
beneran blum tau kalo anies bukannya petugas partai nasdem siapa presidennya anies apa partainya nasdem ya kaleeee kalo mahfud md sih non partai tapi sebelum lahir pun sudah nahdliyin kalo blum tau rajin baca bro	beneran blum tau kalo anies bukannya tugas partai nasdem siapa presiden anies apa partai nasdem ya kaleeee kalo mahfud md sih non partai tapi belum lahir pun sudah nahdliyin kalo blum tau rajin baca bro
ini membahas lapangan hasil karya anak bangsa yg dibangun masa kepemimpinan mas di dki buka bahas	ini bahas lapang hasil karya anak bangsa yg bangun masa pimpin mas di dki buka bahas
hanya ganjar yang konsisten tolak israel datang ke indonesia	hanya ganjar yang konsisten tolak israel datang ke indonesia
tapi aku ragu ketiga calon bahas ekonomi syariah melulu tidak ada ekonomi kerakyatan pancasila	tapi aku ragu tiga calon bahas ekonomi syariah melulu tidak ada ekonomi rakyat pancasila
bbm gratis jadi taik	bbm gratis jadi taik

Gambar 4.6 Data setelah di stemming

### 4. Tokenizing dan menghapus *stopword*

*Tokenizing* memisahkan sebuah kalimat menjadi penggalan kata. Sedangkan penghapusan *stopword* berguna untuk menghilangkan kata kata yang tidak memiliki arti sentimen atau kata-kata dasar seperti kata dari, ke, ini, dan, dalam, yang, dan lain sebagainya.

stemword	Tokenized_and_No_Stopwords
beneran blum tau kalo anies bukankah tugas par...	[beneran, blum, tau, kalo, anies, tugas, parta...
ini bahas lapang hasil karya anak bangsa yg ba...	[bahas, lapang, hasil, karya, anak, bangsa, yg...
hanya ganjar yang konsisten tolak israel datan...	[ganjar, konsisten, tolak, israel, indonesia]
tapi aku ragu tiga calon bahas ekonomi syariah...	[ragu, calon, bahas, ekonomi, syariah, melulu...
bbm gratis jadi taik	[bbm, gratis, taik]

Gambar 4.7 Data setelah tokenizing

## 5. Data Labeling

Data labeling adalah proses untuk mengkategorikan teks menjadi nilai sentimennya. Penelitian ini menggunakan model BERT untuk labeling data berbahasa Indonesia. Model BERT akan memproses data teks kemudian memberikan label atau klasifikasinya dengan kategori 1 – 5 star sesuai dengan nilai sentimen teks, contohnya pada Gambar 4.8.

	clean_text	Sentimen_Label	Sentimen_Score
0	beneran blum tau kalo anies bukankah petugas p...	1 star	0.514378
1	ini membahas lapangan hasil karya anak bangsa ...	5 stars	0.423526
2	hanya ganjar yang konsisten tolak israel datan...	1 star	0.330613
3	tapi aku ragu ketiga calon bahas ekonomi syari...	2 stars	0.350466
4	bbm gratis jadi taik	1 star	0.285715

Gambar 4.8 Hasil Labeling menggunakan model BERT

Kemudian mengubah labeling dengan format 1-5 star menjadi labeling sentimen Negative, Neutral, dan Positive dimana 1-2 star merupakan sentimen Negative, 3 stars merupakan sentimen Neutral dan 4-5 stars merupakan sentimen Positive seperti pada Gambar 4.9.

	clean_text	Sentimen_value	Labeling	Sentimen_score
0	beneran blum tau kalo anies bukankah petugas p...	Negative	1 star	0.514378
1	ini membahas lapangan hasil karya anak bangsa ...	Positive	5 stars	0.423526
2	hanya ganjar yang konsisten tolak israel datan...	Negative	1 star	0.330613
3	tapi aku ragu ketiga calon bahas ekonomi syari...	Negative	2 stars	0.350466
4	bbm gratis jadi taik	Negative	1 star	0.285715

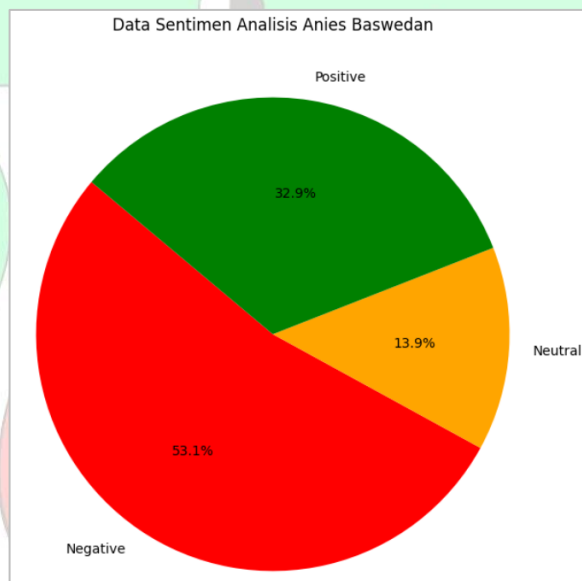
Gambar 4.9 Mengubah Labeling menjadi nilai sentimen

Kemudian menambahkan nilai dari sentimen. Contohnya Sentimen Negative yaitu label 0, Sentimen Neutral yaitu label 1, dan Sentimen Positive yaitu label 2 seperti pada Gambar 4.9.

	clean_text	Sentimen_value	label
0	beneran blum tau kalo anies bukankah petugas p...	Negative	0
1	ini membahas lapangan hasil karya anak bangsa ...	Positive	2
2	hanya ganjar yang konsisten tolak israel datan...	Negative	0
3	tapi aku ragu ketiga calon bahas ekonomi syari...	Negative	0
4	bbm gratis jadi taik	Negative	0

Gambar 4.9 Hasil Data Labeling

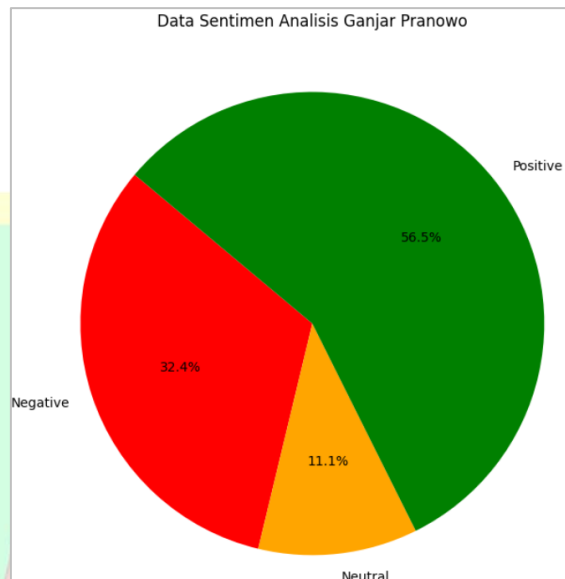
Setelah proses labeling maka data calon presiden Indonesia sudah bisa diketahui, yaitu sebagai berikut:



Gambar 4.10 Diagram Pie Data Sentimen Anies Baswedan

```
1 senti_anies['Sentimen_value'].value_counts()
Negative      271
Positive      168
Neutral        71
Name: Sentimen_value, dtype: int64
```

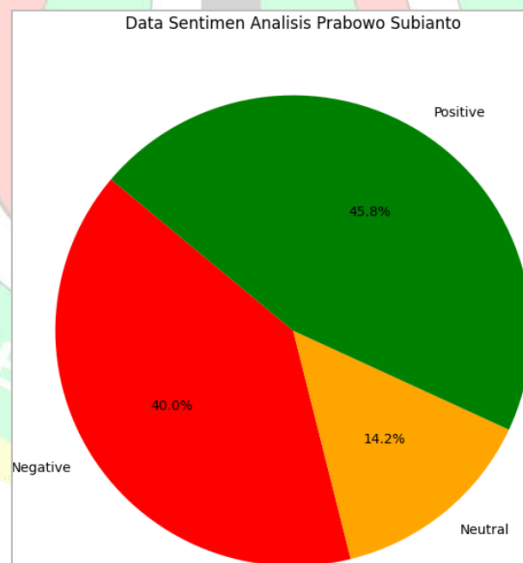
Gambar 4.11 Banyaknya Data Sentimen Anies Baswedan



Gambar 4.12 Diagram Pie Data Sentimen Ganjar Pranowo

```
1 senti_Ganjar['Sentimen_value'].value_counts()
Positive      290
Negative      166
Neutral        57
Name: Sentimen_value, dtype: int64
```

Gambar 4.13 Banyaknya Data Sentimen Ganjar Pranowo



Gambar 4.14 Diagram Pie Data Sentimen Prabowo Subianto

```

1 senti_Prabowo['Sentimen_value'].value_counts()

Positive    229
Negative    200
Neutral      71
Name: Sentimen_value, dtype: int64

```

Gambar 4.15 – Banyaknya Data Sentimen Prabowo Subianto

## 6. Splitting data

Splitting data adalah proses membagi data menjadi 2 kategori yaitu data train dan data testing. Penelitian ini membagi dataset menjadi 80:20. Data training 80% untuk melatih evaluasi model sedangkan data testing 20% untuk menguji performa model.

```

1 from sklearn.model_selection import train_test_split
2
3 # Misalnya, X adalah fitur dan y adalah label
4 X_train, X_test, y_train, y_test = train_test_split(senti_anies['clean_text'],
5                                                    senti_anies['label'], test_size=0.2, random_state=42)
6

```

Gambar 4.16 Data Splitting

## 4.2 Analisis menggunakan Algoritma Naïve Bayes dan Regresi Logistik

Penelitian ini menggunakan algoritma Naive Bayes dan Regresi Logistik dalam mengklasifikasikan data sentimen analisis dengan data yang dibagi menjadi 80% data train dan 20% data test. Kemudian masing-masing data kandidat calon presiden diproses menggunakan Permodelan Naive Bayes dan Regresi Logistik. Hasil dari setiap model terdapat pada Table 4.1, Gambar 4.17 dan Gambar 4.18 untuk hasil klasifikasi data sentimen kandidat calon presiden Anies Baswedan, Table 4.2.2, Gambar 4.2.5 dan gambar 4.2.6 untuk hasil klasifikasi data sentimen kandidat calon presiden Ganjar Pranowo, dan Table 4.2.3, Gambar 4.2.9 dan Gambar 4.2.10 untuk klasifikasi data sentimen kandidat calon presiden Prabowo Subianto.

Pada Table 4.1 menunjukkan bahwa nilai akurasi permodelan Naive Bayes adalah sebesar 63% sedangkan nilai presisinya adalah 64%, 75% dan 54%. Sedangkan, Nilai akurasi permodelan Regresi Logistik adalah sebesar 61% sedangkan nilai presisinya adalah 74%, 50%, dan 44%. Sehingga berdasarkan penilaian Tingkat akurasi permodelan pada data sentimen Anies Baswedan, algoritma yang nilai akurasinya lebih unggul adalah Nilai

akurasi Naive bayes yaitu sebesar 63% sedangkan nilai akurasi Regresi Logistik sebesar 61%. Confusion matrix data sentimen kandidat calon presiden 2024 Anies Baswedan digambarkan pada Gambar 4.19 dan Gambar 4.20.

Tabel 4.1 Laporan Klasifikasi Data Sentimen Anies Baswedan					
Hasil Evaluasi Model Data sentimen Anies Baswedan					
Algoritma	Sentimen	Presisi	Recall	F1-socre	Akurasi
Naive Bayes	Negative	64%	92%	75%	63%
	Neutral	75%	19%	30%	
	Positive	54%	26%	35%	
Regresi Logistik	Negative	74%	71%	72%	61%
	Neutral	50%	6%	11%	
	Positive	44%	70%	54%	

```

Accuracy of Naive Bayes Multinomial classifier on test set: 0.63

Classification Report:
              precision    recall  f1-score   support

     0       0.64         0.92         0.75         59
     1       0.75         0.19         0.30         16
     2       0.54         0.26         0.35         27

   accuracy          0.63
  macro avg       0.64         0.45         0.47
 weighted avg       0.63         0.63         0.57

Confusion Matrix:
[[54  1  4]
 [11  3  2]
 [20  0  7]]

```

Gambar 4.17 Hasil Klasifikasi Naive Bayes Data Sentimen Anies baswedan

Accuracy of Logistic Regression classifier on test set: 0.61

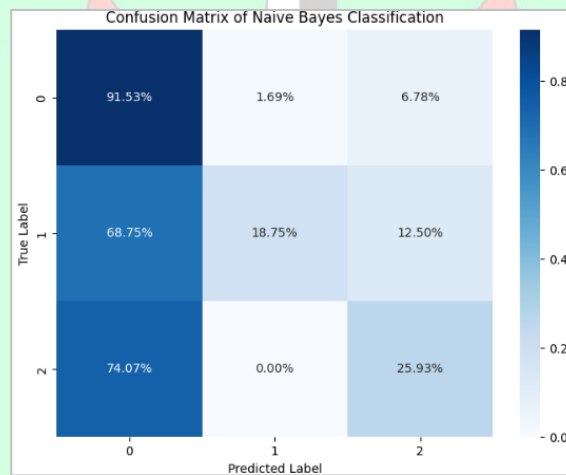
Laporan Klasifikasi:

	precision	recall	f1-score	support
0	0.74	0.71	0.72	59
1	0.50	0.06	0.11	16
2	0.44	0.70	0.54	27
accuracy			0.61	102
macro avg	0.56	0.49	0.46	102
weighted avg	0.62	0.61	0.58	102

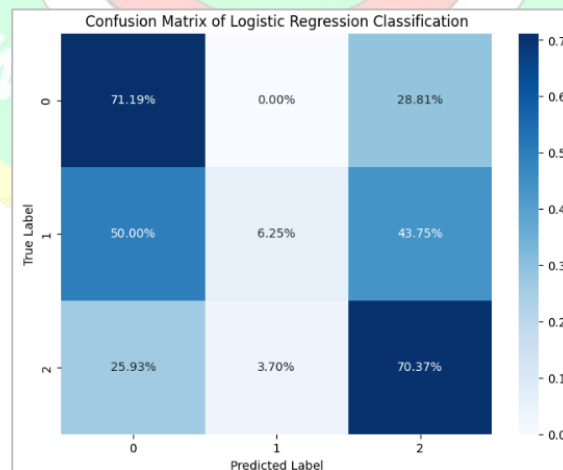
Confusion Matrix:

```
[[42  0 17]
 [ 8  1  7]
 [ 7  1 19]]
```

Gambar 4.18 Hasil Klasifikasi Regresi Logistik Data Sentimen Anies Baswedan



Gambar 4.19 Confusion Matriks Naïve Bayes Anies Baswedan



Gambar 4.20 Confusion Matriks Regresi Logistik Anies Baswedan



Pada Table 4.2 dan Gambar 4.21 menunjukkan bahwa nilai akurasi permodelan Naive Bayes adalah sebesar 77% sedangkan nilai presisinya adalah 72%, 33% dan 81%. Sedangkan, Nilai akurasi permodelan Regresi Logistik ada pada Gambar 4.22 yaitu sebesar 77% sedangkan nilai presisinya adalah 84%, 0%, dan 76%. Sehingga berdasarkan penilaian Tingkat akurasi permodelan pada data sentimen Ganjar Pranowo, kedua algoritma memiliki nilai akurasi yang sama yaitu sebesar 77%, yang menjadikan kedua permodelan seimbang. Confusion matrix data sentimen kandidat calon presiden 2024 Ganjar Pranowo digambarkan pada Gambar 4.23 dan Gambar 4.24.

Tabel 4.1 Laporan Klasifikasi Data Sentimen Ganjar Pranowo

Hasil Evaluasi Model Data sentimen Ganjar Pranowo					
Algoritma	Sentimen	Presisi	Recall	F1-score	Akurasi
Naive Bayes	Negative	72%	77%	74%	77%
	Neutral	33%	8%	13%	
	Positive	81%	90%	85%	
Regresi Logistik	Negative	84%	70%	76%	77%
	Neutral	0%	0%	0%	
	Positive	76%	95%	85%	

```

Accuracy of Naive Bayes Multinomial classifier on test set: 0.77

Classification Report:
              precision    recall  f1-score   support

     0       0.72         0.77         0.74         30
     1       0.33         0.08         0.13         12
     2       0.81         0.90         0.85         61

 accuracy          0.77         0.77         0.77         103
  macro avg       0.62         0.58         0.58         103
 weighted avg     0.73         0.77         0.74         103

Confusion Matrix:
[[23  2  5]
 [ 3  1  8]
 [ 6  0 55]]

```

*Gambar 4.21 Hasil Klasifikasi Naïve Bayes Data Sentimen Ganjar Pranowo*

```

Accuracy of Logistic Regression classifier on test set: 0.77

Laporan Klasifikasi:
              precision    recall  f1-score   support

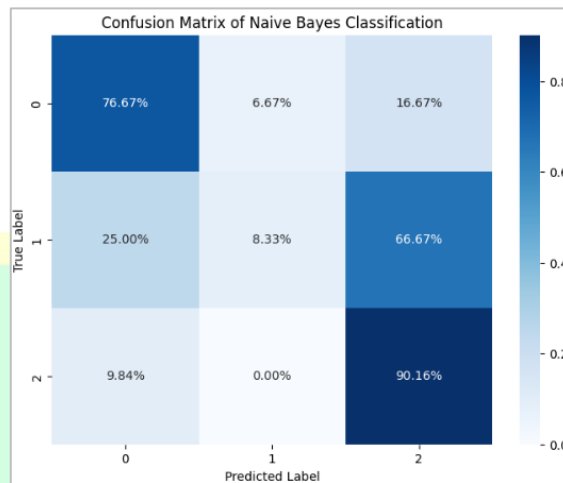
     0       0.84         0.70         0.76         30
     1       0.00         0.00         0.00         12
     2       0.76         0.95         0.85         61

 accuracy          0.77         0.77         0.77         103
  macro avg       0.53         0.55         0.54         103
 weighted avg     0.70         0.77         0.72         103

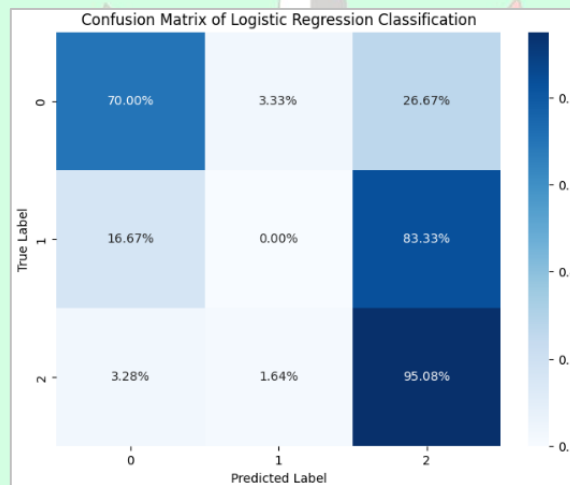
Confusion Matrix:
[[21  1  8]
 [ 2  0 10]
 [ 2  1 58]]

```

*Gambar 4.22 Hasil Klasifikasi Regresi Logistik Data Sentimen Ganjar Pranowo*



Gambar 4.23 Confusion Matrix Naïve Bayes Ganjar Pranowo



Gambar 4.24 Confusion Matrix

Pada Table 4.3 dan Gambar 4.25 menunjukkan bahwa nilai akurasi permodelan Naive Bayes adalah sebesar 44% sedangkan nilai presisinya adalah 33%, 50% dan 52%. Sedangkan, Nilai akurasi permodelan Regresi Logistik ada pada Gambar 4.26 sebesar 41% sedangkan nilai presisinya adalah 33%, 50%, dan 46%. Sehingga berdasarkan penilaian Tingkat akurasi permodelan pada data sentimen Prabowo Subianto, algoritma yang nilai akurasinya lebih unggul adalah Nilai akurasi Naive bayes yaitu sebesar 44% sedangkan nilai akurasi Regresi Logistik sebesar 41%. Confusion matrix data sentimen kandidat calon presiden 2024 Prabowo Subianto digambarkan pada Gambar 4.27 dan Gambar 4.28.

Tabel 4.2. Laporan Klasifikasi Data Sentimen Prabowo Subianto

### Hasil Evaluasi Model Data sentimen Prabowo Subianto

Algoritma	Sentimen	Presisi	Recall	F1-score	Akurasi
Naive Bayes	Negative	33%	42%	37%	44%
	Neutral	50%	14%	22%	
	Positive	52%	53%	52%	
Regresi Logistik	Negative	33%	45%	38%	41%
	Neutral	50%	7%	12%	
	Positive	46%	47%	47%	

Accuracy of Naive Bayes Multinomial classifier on test set: 0.44

#### Classification Report:

	precision	recall	f1-score	support
0	0.33	0.42	0.37	33
1	0.50	0.14	0.22	14
2	0.52	0.53	0.52	53
accuracy			0.44	100
macro avg	0.45	0.37	0.37	100
weighted avg	0.45	0.44	0.43	100

#### Confusion Matrix:

```
[[14  0 19]
 [ 5  2  7]
 [23  2 28]]
```

Gambar 4.25 Hasil Klasifikasi Naïve Bayes Data Sentimen Prabowo Subianto

Accuracy of Logistic Regression classifier on test set: 0.41

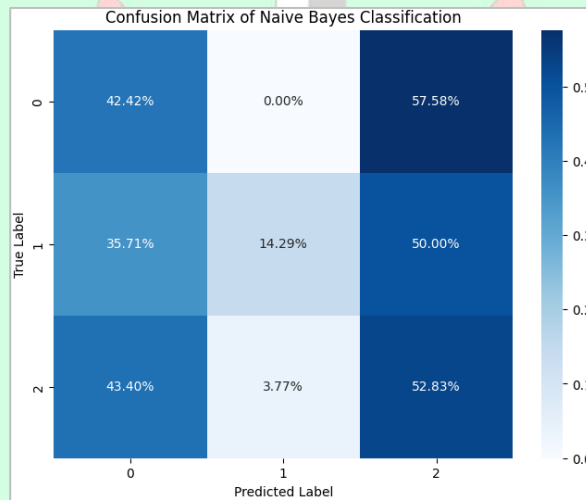
Laporan Klasifikasi:

	precision	recall	f1-score	support
0	0.33	0.45	0.38	33
1	0.50	0.07	0.12	14
2	0.47	0.47	0.47	53
accuracy			0.41	100
macro avg	0.44	0.33	0.33	100
weighted avg	0.43	0.41	0.39	100

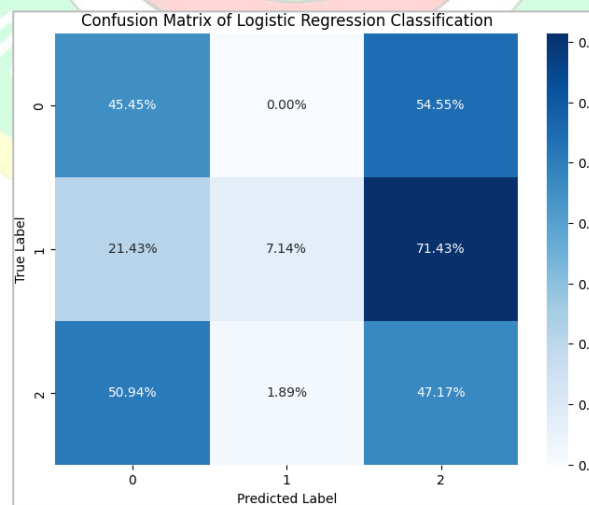
Matriks Kebingungan:

```
[[15  0 18]
 [ 3  1 10]
 [27  1 25]]
```

Gambar 4.26 Hasil Klasifikasi Regresi Logistik Data Sentimen Prabowo Subianto



Gambar 4.27 Confusion Matrix Naïve Bayes Prabowo Subianto



Gambar 4.28 Confusion Matrix Regresi Logistik Prabowo Subianto

### 4.3 Interface Dashboard Streamlit

Pada penelitian ini menggunakan streamlit untuk menampilkan hasil dari proses sentimen analisis menggunakan algoritma Naïve Bayes dan Regresi Logistik terhadap data teks Twitter untuk melihat atau menganalisa sentimen masyarakat terhadap kandidat calon presiden pada pemilu 2024.

**SENTIMEN ANALISIS PADA TWITTER MENGGUNAKAN NAIVE BAYES DAN REGRESI LOGISTIK UNTUK PEMILU PRESIDEN 2024**

Hasil Crawling data

1. Hasil Crawling Data Anies Baswedan

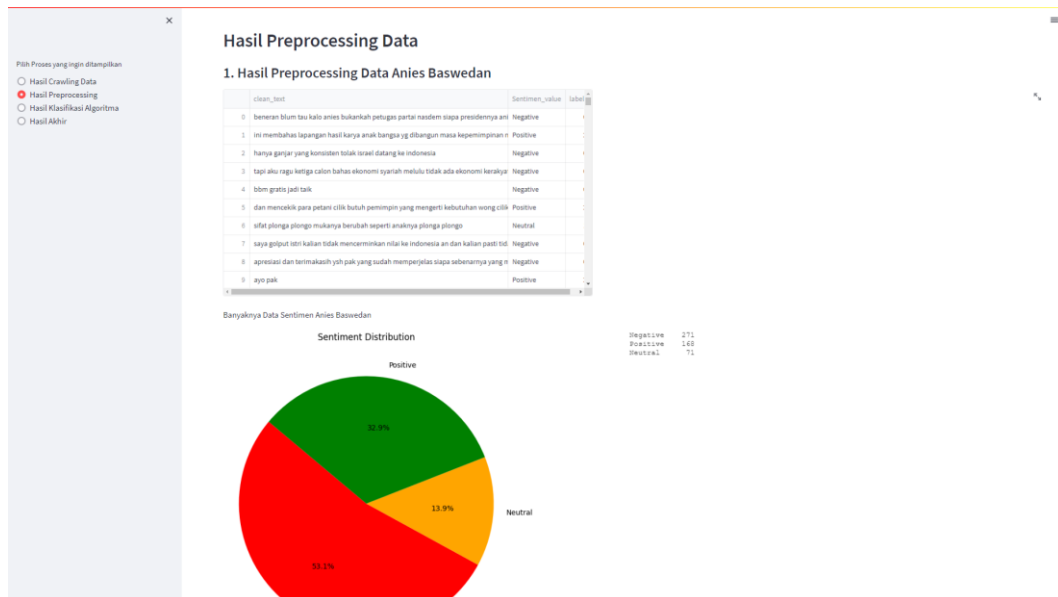
	created_at	id_str	full_text	quote_count	reply_count	retweet_count	favorite_count	lang	user_id_str
0	Tue Nov 07 05:04:48 +0000 2023	1721760000000000000	@Korandi @aniesbaswedan @mohamudmd Benaran blum tau? Kalo Anies bukan	0	0	0	0	in	1225000000000000000
1	Tue Nov 07 05:02:38 +0000 2023	1721760000000000000	ini membahas lapangan hasil karya anak bangsa yg dibangun masa kepemimpinan r	0	0	0	0	in	1444320000000000000
2	Tue Nov 07 05:02:43 +0000 2023	1721760000000000000	KANNA GANJAR WANGI KCHSISTEN TOLAK ISRAEL DATANG KE INDONESIA @ganjarpr	0	0	0	0	in	1539600000000000000
3	Tue Nov 07 05:03:52 +0000 2023	1721760000000000000	@Tfak307013 Tapi aku ragu ketiga calon bahas ekonomi syarah melulu @ganjarpr	0	0	0	0	in	1579420000000000000
4	Tue Nov 07 05:02:37 +0000 2023	1721760000000000000	@PPP_PKB @aniesbaswedan @cakimNOW Blm gratis jad TTT Tak	0	0	0	0	in	7943750000000000000
5	Tue Nov 07 05:02:05 +0000 2023	1721750000000000000	dan menasih para petani cilik. Butuh pemimpin yang mengerti kebutuhan wong cil	0	0	0	0	in	504205380
6	Tue Nov 07 05:02:25 +0000 2023	1721750000000000000	@STrenowat102 @aniesbaswedan @ch_hotmail2 @Mduak17 @Demyongar7 @e	0	0	0	0	in	1574280000000000000
7	Tue Nov 07 05:02:02 +0000 2023	1721750000000000000	@Tfak307013 @ganjarpranowo @aniesbaswedan saya gigitu nih kalian tidak mer	0	0	0	0	in	1579420000000000000
8	Tue Nov 07 04:59:44 +0000 2023	1721750000000000000	Apresiasi dan terimakasih yoh pak @aniesbaswedan yang sudah memperjalan staga	0	0	0	1	in	1265280000000000000
9	Tue Nov 07 04:59:23 +0000 2023	1721750000000000000	@sobot_anies @tousipudjastuti @aniesbaswedan @cakimNOW Ayo pak	0	0	0	0	in	1104400000000000000

2. Hasil Crawling Data Ganjar Pranowo

	created_at	id_str	full_text	quote_count	reply_count	retweet_count	favorite_count	lang	user_id_str
0	Tue Nov 07 04:56:57 +0000 2023	1721750000000000000	Report dari Bapak @hary_tanoe <a href="https://t.co/APgfuupgMk">https://t.co/APgfuupgMk</a> @GanjarPranowo mihak	0	0	0	0	in	84393847
1	Tue Nov 07 04:56:58 +0000 2023	1721750000000000000	Gapol datang ke Peta Rakyat Ganjar Pranowo yang berlangsung di Cilacap 15 Bertan	0	0	0	0	in	2505761813
2	Tue Nov 07 04:56:58 +0000 2023	1721750000000000000	Ratusan Pekerja Migran Indonesia (PMI) di kota Riyadh Saudi Arabia deklarasi dukun	0	0	0	0	in	27321805898
3	Tue Nov 07 04:56:51 +0000 2023	1721750000000000000	Pak Ganjar Pranowo memperlihatkan bahwa beliau memiliki masa depan milenia	0	0	0	0	in	1586570000000000000
4	Tue Nov 07 04:56:57 +0000 2023	1721750000000000000	Benarlah...??? NguAA, kita buktikan 2024 Ganjar Presiden. Mau list sebagai dukun	0	0	0	0	in	1505180000000000000
5	Tue Nov 07 04:55:08 +0000 2023	1721750000000000000	Report dari Bapak @hary_tanoe <a href="https://t.co/CUdUEYdE">https://t.co/CUdUEYdE</a> @GanjarPranowo mihakful	0	0	0	0	in	231355136
6	Tue Nov 07 04:55:08 +0000 2023	1721750000000000000	@PartaSoemed @ganjarpranowo @tousipudjastuti semoga membawa oke!2 bantua	0	0	0	0	in	97897218
7	Tue Nov 07 04:55:08 +0000 2023	1721750000000000000	Ganjar Pranowo menjadi orang paling depan dalam memajukan Produk-produk asli	0	0	0	0	in	1288960000000000000
8	Tue Nov 07 04:55:01 +0000 2023	1721750000000000000	Ganjar Pranowo menjadi orang paling depan dalam memajukan Produk-produk asli	0	0	0	1	in	10923968340
9	Tue Nov 07 04:54:56 +0000 2023	1721750000000000000	@dunga_kadun @IndeSul9009850 @muchi_ar @ganjarpranowo @mohamfulu	0	0	0	0	in	1543137185

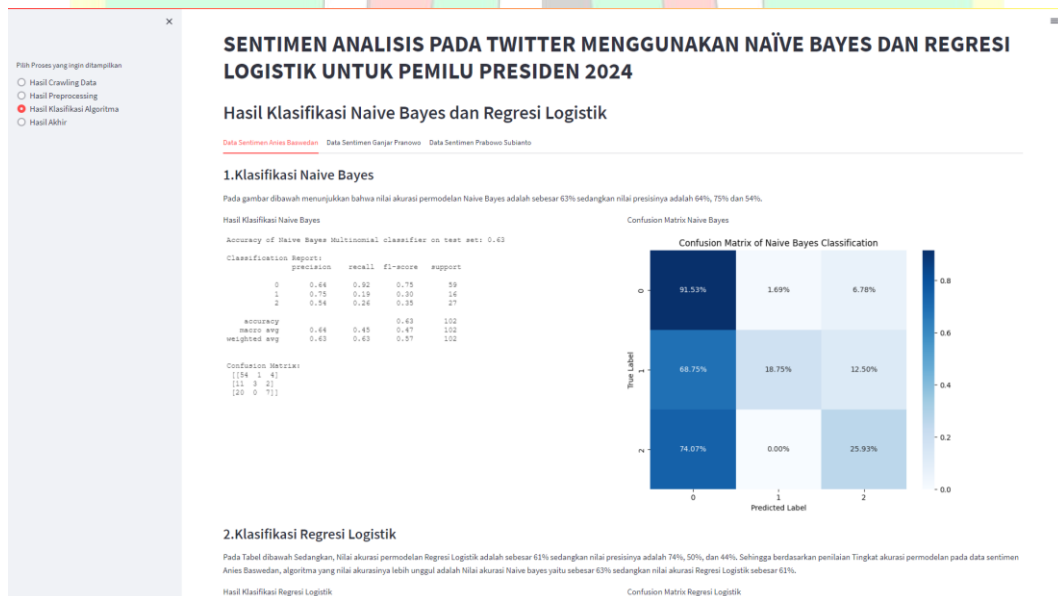
Gambar 4.29 Tampilan Halaman Hasil crawling data

Gambar 4.29 menampilkan Hasil Crawling data. Data twitter terkait Anies Baswedan sebanyak 509 data, Data twitter terkait Ganjar Pranowo sebanyak 512 data dan data twitter terkait Prabowo Subianto sebanyak 552 data.



Gambar 4.30 Tampilan Halaman Hasil Preprocessing data

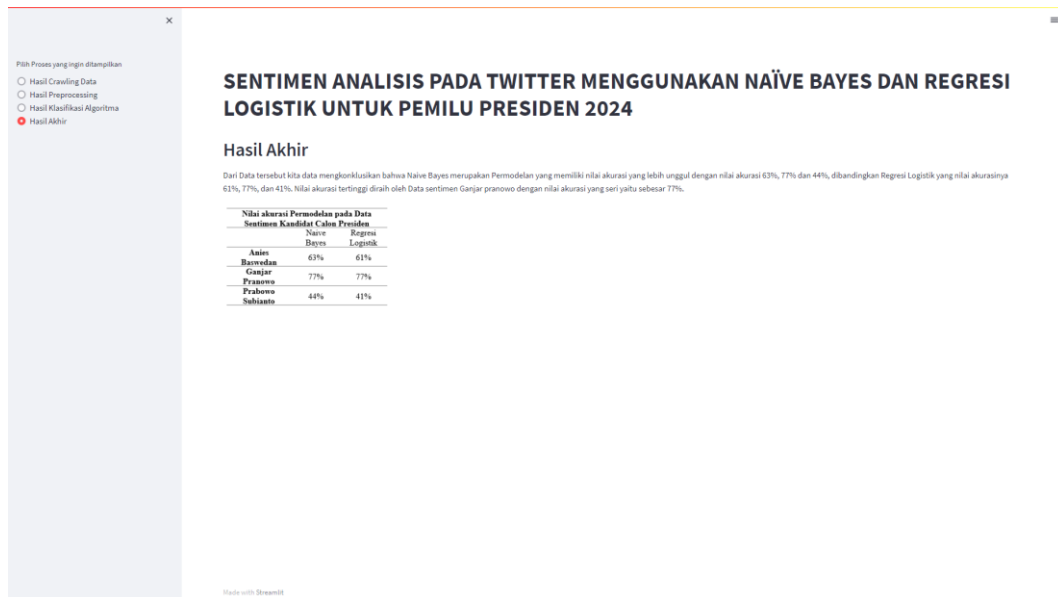
Gambar 4.30 menampilkan Hasil Preprocessing data dan seberapa banyak data sentimen sesuai dengan labelingnya dengan diagram pie.



Gambar 4.31 Tampilan hasil Klasifikasi Algoritma

Gambar 4.3.3 menampilkan hasil klasifikasi menggunakan algoritma Naïve Bayes dan Regresi Logistik dari setiap data kandidat calon presiden. Hasil klasifikasi meliputi akurasi, presisi, recall, f1-score dan Confusion matrix.





Gambar 4.32 Tampilan Hasil Akhir

Gambar 4.32 menampilkan hasil akhir dari sentimen analisis, dimana algoritma Naïve Bayes lebih unggul dibandingkan Algoritma Regresi Logistik.

#### 4.4 Hasil Akhir

Pada Table 4 menunjukkan nilai akurasi dari setiap kandidat calon presiden dan juga permodelan Algoritma yang digunakan. Dari table tersebut kita data mengkonklusikan bahwa Naive Bayes merupakan Permodelan yang memiliki nilai akurasi yang lebih unggul dengan nilai akurasi 63%, 77% dan 44%, dibandingkan Regresi Logistik yang nilai akurasinya 61%, 77%, dan 41%. Nilai akurasi tertinggi diraih oleh Data sentimen Ganjar pranowo dengan nilai akurasi yang seri yaitu sebesar 77%.

Tabel 3.4 Nilai Akurasi dari permodelan Naïve Bayes dan Regresi Logistik

Nilai akurasi Permodelan pada Data Sentimen Kandidat Calon Presiden		
	Naive Bayes	Regresi Logistik
Anies Baswedan	63%	61%
Ganjar Pranowo	77%	77%
Prabowo Subianto	44%	41%

## BAB VI

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Penelitian ini menggunakan Algoritma Naïve Bayes dan Regresi Logistik dalam permodelan klasifikasi sentimen data twitter terkait dengan kandidat calon presiden pada pemilu 2024. Data yang didapatkan pada bulan November 2023 yaitu sebanyak 1.573 data yang terdiri dari 512 data tweets yang berkaitan dengan Ganjar Pranowo, 509 data tweets berkaitan dengan Anies Baswedan dan 552 data tweets yang berkaitan dengan Prabowo Subianto pada bulan November 2023. Didapatkn hasil bahwa pada tahap labeling terdapat 3 kategori label yaitu negatif, netral dan positif. Dimana klasifikasi yang unggul dalam permodelan yang dilakukan adalah Naïve Bayes yang bernilai akurasi 63%, 77% dan 44%, dibandingkan Regresi Logistik yang nilai akurasinya 61%, 77%, dan 41%. Nilai akurasi tertinggi dari naive bayes diraih oleh Data sentimen Ganjar pranowo dengan nilai akurasi yang seri yaitu sebesar 77%.

#### 5.2 Saran

Saran peneliti yaitu untuk penelitian selanjutnya untuk menggunakan data yang lebih relevan dan yang diperoleh selama masa kampanye sehingga data sentimen lebih banyak dan juga beragam, menambahkan pembobotan kata, algoritma yang lebih bervariasi, dan untuk tampilan website yang lebih baik dan lebih user-friendly

## DAFTAR PUSTAKA

- Asmara, R., Ardiansyah, M. F., & Anshori, M. (2020). Analisa Sentiment Masyarakat terhadap Pemilu 2019 berdasarkan Opini di Twitter menggunakan Metode Naive Bayes Classifier. *INOVTEK Polbeng - Seri Informatika*, 5(2), 193. <https://doi.org/10.35314/isi.v5i2.1095>
- Assiroj, P., Kurnia, A., & Alam, S. (2023). The performance of Naïve Bayes, support vector machine, and logistic regression on Indonesia immigration sentiment analysis. *Bulletin of Electrical Engineering and Informatics*, 12(6), 3843–3852. <https://doi.org/10.11591/eei.v12i6.5688>
- Averina, A., Hadi, H., & Siswanto, J. (2022). Analisis Sentimen Multi-Kelas Untuk Film Berbasis Teks Ulasan Menggunakan Model Regresi Logistik. *Teknika*, 11(2), 123–128. <https://doi.org/10.34148/teknika.v11i2.461>
- Bahtiar, S. A. H., Dewa, C. K., & Luthfi, A. (2023). Comparison of Naïve Bayes and Logistic Regression in Sentiment Analysis on Marketplace Reviews Using Rating-Based Labeling. *Journal of Information Systems and Informatics*, 5(3), 915–927. <https://doi.org/10.51519/journalisi.v5i3.539>
- Buntoro, G., Arifin, R., Syaifuddiin, G., Selamat, A., Krejcar, O., & Hamido, F. (2021). THE IMPLEMENTATION OF THE MACHINE LEARNING ALGORITHM FOR THE SENTIMENT ANALYSIS OF INDONESIA'S 2019 PRESIDENTIAL ELECTION. *IJUM Engineering Journal*, 22(1), 78–92. <https://doi.org/10.31436/iiumej.v22i1.1532>
- Das, S., Bhattacharyya, K., & Sarkar, S. (2023). Performance Analysis of Logistic Regression, Naive Bayes, KNN, Decision Tree, Random Forest and SVM on Hate Speech Detection from Twitter. *International Research Journal of Innovations in Engineering and Technology*, 07(03), 07–03. <https://doi.org/10.47001/IRJIET/2023.703004>
- Delvika, B., Apriana, Abror, N., & Gurning, U. R. (2023). Perbandingan Algoritma NBC dan C4.5 Dalam Analisa Sentimen Pemilihan Presiden 2024 Pada Twitter. *SENTIMAS: Seminar Nasional Penelitian Dan Pengabdian Masyarakat*, 1(1), 41–48.
- Fathir, Hariyadi, M. A., & Miftachul A, Y. (2023). ANALISIS SENTIMEN ARTIKEL BERITA PEMILU BERBASIS METODE KLASIFIKASI. *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi*, 4(2), 485–493. <https://doi.org/10.35870/jimik.v4i2.220>
- Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter.

- SMATIKA JURNAL*, 10(02), 71–76.  
<https://doi.org/10.32664/smatika.v10i02.455>
- Helmi Satria. (2023, October 18). *Cara crawl data twitter lebih dari 1500 tweets - 18 Oct*. Youtube. <https://youtu.be/1uOm95nMcuk>
- Husen, R. A., Astuti, R., Marlia, L., Rahmadden, R., & Efrizoni, L. (2023). Analisis Sentimen Opini Publik pada Twitter Terhadap Bank BSI Menggunakan Algoritma Machine Learning. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(2), 211–218.  
<https://doi.org/10.57152/malcom.v3i2.901>
- Imamah, & Rachman, F. H. (2020). Twitter Sentiment Analysis of Covid-19 Using Term Weighting TF-IDF And Logistic Regresion. *2020 6th Information Technology International Seminar (ITIS)*, 238–242.  
<https://doi.org/10.1109/ITIS50118.2020.9320958>
- Muzaki, A., & Witanti, A. (2021). SENTIMENT ANALYSIS OF THE COMMUNITY IN THE TWITTER TO THE 2020 ELECTION IN PANDEMIC COVID-19 BY METHOD NAIVE BAYES CLASSIFIER. *Jurnal Teknik Informatika (Jutif)*, 2(2), 101–107.  
<https://doi.org/10.20884/1.jutif.2021.2.2.51>
- Prabowo, R., Sujaini, H., & Rismawan, T. (2023). Analisis Sentimen Pengguna Twitter Terhadap Kasus COVID-19 di Indonesia Menggunakan Metode Regresi Logistik Multinomial. *Jurnal Sistem Dan Teknologi Informasi (JustIN)*, 11(1), 85. <https://doi.org/10.26418/justin.v11i1.57450>
- Ramadhani, S. H., & Wahyudin, M. I. (2022). Analisis Sentimen Terhadap Vaksinasi Astra Zeneca pada Twitter Menggunakan Metode Naïve Bayes dan K-NN. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, 6(4), 526–534. <https://doi.org/10.35870/jtik.v6i4.530>
- Romadhon, M. R., & Kurniawan, F. (2021). A Comparison of Naive Bayes Methods, Logistic Regression and KNN for Predicting Healing of Covid-19 Patients in Indonesia. *2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT)*, 41–44.  
<https://doi.org/10.1109/EIConCIT50028.2021.9431845>
- Samsir, Irmayani, D., Edi, F., Harahap, J. M., Jupriaman, Rangkuti, R. K., Ulya, B., & Watrianthos, R. (2021). Naives Bayes Algorithm for Twitter Sentiment Analysis. *Journal of Physics: Conference Series*, 1933(1), 012019.  
<https://doi.org/10.1088/1742-6596/1933/1/012019>
- Syafii, M. I. (2023). SENTIMEN ANALISIS PADA MEDIA SOSIAL MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER (NBC). *Jurnal Teknologi Pintar*, 3(2).

- Utami, D. Y., Nurlelah, E., & Hasan, F. N. (2021). Comparison of Neural Network Algorithms, Naive Bayes and Logistic Regression to predict diabetes. *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, 5(1), 53–64. <https://doi.org/10.31289/jite.v5i1.5201>
- Vindua, R., & Zailani, A. U. (2023). Analisis Sentimen Pemilu Indonesia Tahun 2024 Dari Media Sosial Twitter Menggunakan Python. *JURIKOM (Jurnal Riset Komputer)*, 10(2), 479. <https://doi.org/10.30865/jurikom.v10i2.5945>
- Wenando, F. A., Hayami, R., & Anggrawan, A. J. (2020). ANALISIS SENTIMEN PADA PEMERINTAHAN TERPILIH PADA PILPRES 2019 DITWITTER MENGGUNAKAN ALGORITME NAÏVEBAYES. *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)*, 7(1), 101–106. <https://doi.org/10.33330/jurteks.v7i1.851>
- Zain, H. H., Awannga, R. M., & Rahayu, W. I. (2023). Perbandingan Model Svm, Knn Dan Naïve Bayes Untuk Analisis Sentiment Pada Data Twitter: Studi Kasus Calon Presiden 2024. *JIMPS: Jurnal Ilmiah Mahasiswa Pendidikan Sejarah*, 8(3).



## LAMPIRAN

**Link Github :** <https://github.com/Alisya0910/SkripsiAlisya.git>

### **Coding Google Colab:**

```
#mengimport library yang dibutuhkan
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import string
import re

Data = pd.read_excel('GanjarPresidentTweets.xlsx')

Data = Data[['full_text']]
Data['clean_text'] = Data.loc[:, 'full_text']

#Text-Cleaning (URLs, Mentions, etc)
#installing tweet-preprocessor
!pip install tweet-preprocessor
import preprocessor as p
def preprocess_tweet(row):
    text = row['clean_text']
    text = p.clean(text)
    return text

Data['clean_text'] = Data.apply(preprocess_tweet, axis=1)

#untuk menghapus data yang Nan atau data kosong setelah di filtering
```

```
#Data = Data[Data != '']
```

```
#Data = Data.dropna()
```

```
#menghapus extra white space, punctuation, dan menlowercase kan  
Data['clean_text'] = Data['clean_text'].str.lower().str.replace('[^\w\s]','  
').str.replace('\s\s+', ' ').str.replace('\d+', '')
```

```
pip install Sastrawi
```

```
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
```

```
import pandas as pd
```

```
# Membuat objek stemmer
```

```
factory = StemmerFactory()
```

```
stemmer = factory.create_stemmer()
```

```
# Membuat DataFrame contoh
```

```
Data['stemword'] = pd.DataFrame(Data['clean_text'])
```

```
# Mendefinisikan fungsi stemming
```

```
def stemming(text):
```

```
    return stemmer.stem(text)
```

```
# Menerapkan fungsi stemming ke kolom 'TextColumn'
```

```
Data['stemword'] = Data['clean_text'].apply(str).apply(stemming)
```

```
import pandas as pd
```

```
from nltk.tokenize import word_tokenize
```

```
from nltk.corpus import stopwords
```

```
import nltk
```



```

# Download kumpulan stopword untuk bahasa Indonesia
nltk.download('punkt')
nltk.download('stopwords')

# Contoh DataFrame
Data['stemword'] = pd.DataFrame(Data['stemword'])

# Fungsi untuk melakukan tokenisasi dan penghapusan stopwords
def tokenize_and_remove_stopwords(text):
    # Tokenisasi kata-kata
    words = word_tokenize(text)

    # Mengambil stopwords bahasa Indonesia
    stop_words = set(stopwords.words('indonesian'))

    # Menghapus stopwords dari kata-kata
    filtered_words = [word for word in words if word.lower() not in stop_words]

    return filtered_words

# Tokenisasi dan penghapusan stopwords pada setiap elemen dalam kolom 'Text'
Data['Tokenized_and_No_Stopwords'] =
Data['stemword'].apply(tokenize_and_remove_stopwords)

!pip install transformers
from transformers import pipeline
import pandas as pd

```

```

# Inisialisasi model BERT untuk analisis sentimen

classifier = pipeline('sentiment-analysis', model='nlptown/bert-base-multilingual-uncased-sentiment')

# Fungsi untuk prediksi sentimen pada teks
def predict_sentiment(text):
    result = classifier(text)[0]
    sentiment_label = result['label']
    sentiment_score = result['score']
    return sentiment_label, sentiment_score

# Prediksi sentimen untuk setiap teks dalam series
sentiment_predictions = Data['clean_text'].apply(predict_sentiment)

# Menyatukan hasil prediksi dengan dataset
result_df = pd.DataFrame(list(sentiment_predictions),
                           columns=['Sentimen_Label', 'Sentimen_Score'])
final_result = pd.concat([Data['clean_text'], result_df], axis=1)

print(final_result)

import pandas as pd

def Sentiment_startovalue(label):
    if label in ["1 star", "2 stars"]:
        return 'Negative'
    elif label == "3 stars":
        return 'Neutral'
    elif label in ["4 stars", "5 stars"]:

```

```

        return 'Positive'

    else:

        return 'Unknown'

# Apply the function to the entire 'Sentimen_Label' column
Data['Sentimen_value'] =
result_df['Sentimen_Label'].apply(Sentiment_startovalue)

Data['Labeling'] = result_df['Sentimen_Label']
Data['Sentimen_score'] = result_df['Sentimen_Score']

senti_Ganjar = Data[['clean_text','Sentimen_value']]

from sklearn.preprocessing import LabelEncoder
label_encoder = LabelEncoder()
senti_Ganjar['label'] =
label_encoder.fit_transform(senti_Ganjar['Sentimen_value'])

senti_Ganjar['Sentimen_value'].value_counts()

# Count the occurrences of each sentiment
sentiments = senti_Ganjar['Sentimen_value'].value_counts()

# Create a pie chart
labels = sentiment_counts.keys()
sizes = sentiment_counts.values()

plt.figure(figsize=(8, 8))

```

```
plt.pie(sizes, labels=labels, autopct='%1.1f%%', startangle=140, colors=['red',  
'orange', 'green'])
```

```
plt.title('Sentiment Distribution')
```

```
plt.show()
```

```
from sklearn.model_selection import train_test_split
```

```
# Misalnya, X adalah fitur dan y adalah label
```

```
X_train, X_test, y_train, y_test =  
train_test_split(senti_Ganjar['clean_text'], senti_Ganjar['label'], test_size=0.2,  
random_state=42)
```

```
from sklearn.feature_extraction.text import CountVectorizer
```

```
vectorizer = CountVectorizer()
```

```
X_train_vectorized = vectorizer.fit_transform(X_train)
```

```
X_test_vectorized = vectorizer.transform(X_test)
```

```
from sklearn.naive_bayes import MultinomialNB
```

```
from sklearn.metrics import accuracy_score, classification_report,  
confusion_matrix
```

```
import matplotlib.pyplot as plt
```

```
naive_bayes = MultinomialNB()
```

```
naive_bayes.fit(X_train_vectorized, y_train)
```

```
# Predictions on the test set
```

```
predictions = naive_bayes.predict(X_test_vectorized)
```

```
# Evaluate the model
```

```

accuracy_NB = accuracy_score(y_test, predictions)

print(f'Accuracy of Naive Bayes Multinomial classifier on test set:
{accuracy_NB:.2f}')

# Classification report
print('\nClassification Report:\n', classification_report(y_test, predictions))

# Confusion matrix
conf_matrix_NB = confusion_matrix(y_test, predictions)

#confusion Confusion_matrix percent
conf_matrix_NB_percent = conf_matrix_NB.astype('float')/
conf_matrix_NB.sum(axis=1)[:, np.newaxis]

print('\nConfusion Matrix:\n', conf_matrix_NB)

# Example: Confusion Matrix Heatmap
import seaborn as sns

plt.figure(figsize=(8, 6))
sns.heatmap(conf_matrix_NB_percent, annot=True, cmap='Blues',
xticklabels=naive_bayes.classes_, yticklabels=naive_bayes.classes_, fmt='.2%')
plt.title('Confusion Matrix of Naive Bayes Classification')
plt.xlabel('Predicted Label')
plt.ylabel('True Label')
plt.show()

```

```

from sklearn.linear_model import LogisticRegression

logistic_regression = LogisticRegression()

logistic_regression.fit(X_train_vectorized, y_train)

# Prediksi pada set pengujian
predictions = logistic_regression.predict(X_test_vectorized)

# Evaluasi model
accuracy_RL = accuracy_score(y_test, predictions)

print(f'Accuracy of Logistic Regression classifier on test set: {accuracy_RL:.2f}')

# Laporan klasifikasi
print("\nLaporan Klasifikasi:\n", classification_report(y_test, predictions))

# Matriks kebingungan
conf_matrix_LR = confusion_matrix(y_test, predictions)

conf_matrix_LR_percent = conf_matrix_LR.astype('float')/
conf_matrix_NB.sum(axis=1)[:, np.newaxis]

print("\nConfusion Matrix:\n", conf_matrix_LR)

# Example: Confusion Matrix Heatmap
import seaborn as sns

plt.figure(figsize=(8, 6))

sns.heatmap(conf_matrix_LR_percent, annot=True, cmap='Blues',
            xticklabels=naive_bayes.classes_, yticklabels=naive_bayes.classes_, fmt='.2%')

plt.title('Confusion Matrix of Logistic Regression Classification')

plt.xlabel('Predicted Label')

plt.ylabel('True Label')

plt.show()

```

```
print('Accuration of Naive Bayes Multinomial Classification :', accuracy_NB)
```

```
print('Accuration of Logistic Regression Classification :', accuracy_RL)
```

Coding Streamlit:

```
#mengimport library yang dibutuhkan
import streamlit as st
import pandas as pd

def main():
    st.title('Analisis sentimen menggunakan data twitter untuk Pemilu Presiden')

    menu = st.sidebar.radio("Pilih tahapan proses", ("Hasil Data Crawling", "Hasil Data preprocessing", "Hasil Klasifikasi Algoritma", "Hasil Akhir"))

    if menu == "Hasil Data Crawling":
        st.header("Hasil Data Crawling")
        st.subheader("1. Hasil data Crawling tentang Anies Baswedan")
        data = pd.read_excel('AniesPresidentTweets.xlsx')
        st.write(data)

        st.subheader("2. Hasil data Crawling tentang Ganjar Pranowo")
        data = pd.read_excel('GanjarPresidentTweets.xlsx')
        st.write(data)

        st.subheader("3. Hasil data Crawling tentang Prabowo Subianto")
        data = pd.read_excel('PrabowoPresidentTweets.xlsx')
        st.write(data)
```



```

if menu == "Hasil Data preprocessing":
    st.subheader("1. Hasil Data Preprocessing tentang Anies Baswedan")
    data = pd.read_excel('Preprocessed_Anies.xlsx')
    st.write(data)
    st.write('Banyaknya Data Sentimen Anies Baswedan')
    col1,col2 = st.columns(2)
    with col1:
        st.image('SS/SenDis_Anies.png')
    with col2:
        st.image('SS/SenDis_Anies(1).png')

    st.subheader("2. Hasil data Preprocessing tentang Ganjar Pranowo")
    data = pd.read_excel('Preprocessed_Ganjar.xlsx')
    st.write(data)
    st.write('Banyaknya Data Sentimen Ganjar Pranowo')
    col1,col2 = st.columns(2)
    with col1:
        st.image('SS/SenDis_Ganjar.png')
    with col2:
        st.image('SS/SenDis_Ganjar(1).png')

    st.subheader("3. Hasil data Preprocessing tentang Prabowo Subianto")
    data = pd.read_excel('Preprocessed_Prabowo.xlsx')
    st.write(data)
    st.write('Banyaknya Data Sentimen Prabowo Subianto')
    col1,col2 = st.columns(2)
    with col1:
        st.image('SS/SenDis_Prabowo.png')

```

```

with col2:

    st.image('SS/SenDis_Prabowo(1).png')

if menu == "Hasil Klasifikasi Algoritma":
    st.header('Hasil Klasifikasi Naive Bayes dan Regresi Logistik')
    tab1, tab2, tab3 = st.tabs(["Hasil Klasifikasi Data Anies Baswedan", "Hasil
    Klasifikasi Data Ganjar Pranowo", "Hasil Klasifikasi Data Prabowo Subianto"])

    with tab1:
        st.subheader("1. Klasifikasi Naive Bayes")
        st.write()

        col1,col2 = st.columns(2)
        with col1:
            st.image('SS/HasilNB_Anies.png')
        with col2:
            st.image('SS/NB_Anies.png')

        st.subheader("2. Klasifikasi Regresi Logistik")
        st.write()
        col1,col2 = st.columns(2)
        with col1:
            st.write()
            st.image('SS/HasilLR_Anies.png')
        with col2:
            st.write()
            st.image('SS/LR_Anies.png')

    with tab2:

```

```
st.subheader("1. Klasifikasi Naive Bayes")
```

```
st.write()
```

```
col1,col2 = st.columns(2)
```

```
with col1:
```

```
    st.image('SS/HasilNB_Ganjar.png')
```

```
with col2:
```

```
    st.image('SS/NB_Ganjar.png')
```

```
st.subheader("2. Klasifikasi Regresi Logistik")
```

```
st.write()
```

```
col1,col2 = st.columns(2)
```

```
with col1:
```

```
    st.write()
```

```
    st.image('SS/HasilLR_Ganjar.png')
```

```
with col2:
```

```
    st.write()
```

```
    st.image('SS/LR_Ganjar.png')
```

```
with tab3:
```

```
    st.subheader("1. Klasifikasi Naive Bayes")
```

```
    st.write()
```

```
    col1,col2 = st.columns(2)
```

```
    with col1:
```

```
        st.image('SS/HasilNB_Prabowo.png')
```

```
    with col2:
```

```
        st.image('SS/NB_Prabowo.png')
```

```

st.subheader("2. Klasifikasi Regresi Logistik")

st.write()

col1,col2 = st.columns(2)

with col1:
    st.write()
    st.image('SS/HasilLR_Prabowo.png')
with col2:
    st.write()
    st.image('SS/LR_Prabowo.png')

else:
    st.header("Hasil Akhir")

    st.write('Dari table tersebut kita data mengkonklusikan bahwa Naive Bayes
    merupakan Permodelan yang memiliki nilai akurasi yang lebih unggul dengan
    nilai akurasi 63%, 77% dan 44%, dibandingkan Regresi Logistik yang nilai
    akurasinya 61%, 77%, dan 41%.')

    st.image('SS/HasilAkhir.png')

# Run the app
if __name__ == '__main__':
    main()




```

**Turnitin**

Turnitin			
ORIGINALITY REPORT			
24%	22%	10%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS
PRIMARY SOURCES			
1	Submitted to University of North Carolina, Greensboro Student Paper	2%	
2	ejournal.itn.ac.id Internet Source	1%	
3	etheses.uin-malang.ac.id Internet Source	1%	
4	www.researchgate.net Internet Source	1%	
5	Submitted to Sriwijaya University Student Paper	1%	
6	Submitted to Universitas Tadulako Student Paper	1%	
7	repository.upbatam.ac.id Internet Source	1%	
8	eprints.ums.ac.id Internet Source	1%	
9	kc.umn.ac.id Internet Source	1%	

**Journal :**


<https://journal.mediadigitalpublikasi.com/index.php/sana/article/view/267>

 **SaNa: Journal of Blockchain, NFTs and Metaverse Technology**  


**Submissions**


**My Queue**

**Archives** 1

 **Help**

**Archived St**


 Search

 **Filters**

**New Submission**

267

**Alisya Mutia Mantika et al.**  
Sentiment Analysis on Twitter Using Naïve Bayes a...

 1

**Published**

**View**

