

ANALISIS SENTIMEN PEMILIHAN PRESIDEN DAN WAKIL PRESIDEN TAHUN 2024 DI TWITTER MENGGUNAKAN METODE KLASIFIKASI *NAÏVE BAYES*

Bhilton M. Obidje¹⁾, Magdalena A. Ineke Pakereng²⁾

1. Teknik Informatika, Universitas Kristen Satya Wacana, Salatiga
2. Teknik Informatika, Universitas Kristen Satya Wacana, Salatiga

Article Info

Kata Kunci: Analisis Sentimen, *Naïve Bayes*, Positif, Negatif, Netral.

Keywords: *Sentiment Analysis, Naïve Bayes, Positive, Negative, Netral*

Article history:

Received 21 October 2024
Revised 12 November 2024
Accepted 13 December 2025
Available online 1 March 2025

DOI :

<https://doi.org/10.29100/jupi.v10i1.5836>

* Corresponding author.

Bhilton M. Obidje

E-mail address:

bhiltonian@gmail.com

ABSTRAK

Pemilihan Umum (PEMILU) dalam sejarah negara Indonesia telah dilaksanakan beberapa kali, namun pemilihan umum yang dilakukan secara langsung oleh masyarakat untuk memilih presiden di Indonesia baru pertama kali dimulai pada tahun 2004. PEMILU yang telah diselenggarakan pada tahun 2024 adalah proses untuk mewujudkan demokrasi di Indonesia. Oleh karena itu penelitian ini bertujuan untuk menganalisis sentimen masyarakat Indonesia tentang calon presiden yang dipilih pada PEMILU tahun 2024. Untuk menganalisis sentimen masyarakat pada penelitian ini akan menggunakan metode klasifikasi *Naive Bayes* untuk dikelompokkan menjadi sentimen positif, netral dan negatif. Oleh karena itu hasil penelitian ini adalah pengelompokan sentimen masyarakat tentang PEMILU 2024 dan menarik kesimpulan dari pengelompokan tersebut. Analisis sentimen yang didapatkan menggunakan metode *Naive Bayes* pada Calon Presiden (Capres) Anies Baswedan mendapatkan sentimen positif 58, netral 20, negatif 19 dengan *accuracy* = 0,75 dan *missclass* = 0,25, sedangkan pada capres Prabowo Subianto mendapatkan sentimen positif 53, netral 24, negatif 21 dengan *accuracy* = 0,5 dan *missclass* = 0,5 dan pada capres Ganjar Pranowo mendapatkan sentimen positif 77, netral 14, negatif 9 dengan *accuracy* = 0,8 dan *missclass* = 0,2.

ABSTRACT

General elections (PEMILU) in the history of the Indonesian state have been held several times, but the first general election carried out directly by the Indonesian people only started in 2004. THE ELECTION which will be held in 2024 is a process to realize democracy in Indonesia. Therefore, this research aims to analyze the sentiments of the Indonesian people regarding the presidential candidates who will be elected in the 2024 ELECTION. To analyze public sentiment in this research, the Naive Bayes classification method will be used to group them into positive, neutral and negative sentiments. Therefore, the results of this research are a grouping of public sentiment regarding the 2024 ELECTION and drawing conclusions from this grouping. Sentiment analysis obtained using the Naive Bayes method for presidential candidate Anies Baswedan got positive sentiment 58, neutral 20, negative 19 with accuracy = 0.75 and missclass = 0.25, while presidential candidate Prabowo Subianto got positive sentiment 53, neutral 24, negative 21 with accuracy = 0.5 and missclass = 0.5 and for the presidential candidate Ganjar Pranowo got a positive sentiment of 77, neutral 14, negative 9 with accuracy = 0.8 and missclass = 0.2.

I. PENDAHULUAN

PEMILIHAN Umum telah dilaksanakan beberapa kali di Indonesia sejak tahun 1955. Pemilihan umum presiden dan wakil presiden yang dilakukan secara langsung oleh rakyat Indonesia, pertama kali di laksanakan paska pada era reformasi tahun 2004 [17]. Pemilihan umum presiden dan wakil presiden yang akan diselenggarakan pada tahun 2024, merupakan Pemilu ke lima paska reformasi. Pemilu merupakan momen penting untuk mewujudkan demokrasi di negara ini. Media sosial merupakan salah satu pilihan calon presiden dan wakil

untuk menyampaikan pesan kampanye seperti visi dan misi. Banyak media sosial yang digunakan untuk menyampaikan pesan kampanye seperti twitter, facebook, tiktok dan lainnya. Berdasarkan data, pemilih PEMILU 2024 didominasi oleh generasi *milenial* dan generasi z [1]. Oleh karena itu kekuatan media sosial untuk PEMILU kali ini tidak bisa dianggap remeh.

Salah satu media sosial yang paling banyak digunakan di Indonesia yaitu Twitter, sehingga menjadi salah satu media pilihan calon presiden dan wakil presiden dalam melakukan kampanye yang bertujuan untuk menunjukkan dan menyampaikan citra positif mereka. Bulan April 2023, *We Are Social* melaporkan terdapat 372,9 juta pengguna twitter di seluruh dunia. Indonesia merupakan peringkat ke-6 dengan 14,75 juta pengguna [2].

Di twitter terdapat banyak pendapat dan komentar masyarakat baik positif, netral maupun negatif sebelum pemilu diselenggarakan. Twitter menyediakan tempat bagi para penggunanya untuk menyampaikan perasaannya tentang apapun. Salah satu contoh sumber data real time dunia nyata yaitu data *tweet* [18]. Dalam menganalisis opini yang ada di twitter biasanya menggunakan pendekatan sentimen yang dikategorikan dalam dalam tiga sentimen yaitu positif, netral dan negatif. Tiga kategori ini digunakan untuk menganalisis PEMILU 2024.

Dalam menganalisis sentimen akan menggunakan metode klasifikasi *Naive Bayes* menggunakan sentimen publik pada twitter. Tujuan Analisis sentimen untuk memahami cara masyarakat mengungkapkan dan opini mereka tentang PEMILU 2024 baik itu positif, netral maupun negatif. Sentimen analisis (*opinion mining*) merupakan salah satu cabang dari *text classification* pada bidang yang luas dari pengolahan bahasa alami. Tujuannya untuk menganalisis pendapat, sentimen, evaluasi, sikap, penilaian dan emosi seseorang pembicara atau penulis terkait topik, produk, layanan, individu dan lainnya [19].

Penelitian ini menggunakan metode *Text Mining*, algoritma *Naive Bayes*. Metode ini bertujuan untuk kategorisasi sentimen dari hasil *crawling* data twitter. Algoritma *Naive Bayes* merupakan salah satu algoritma yang digunakan untuk klasifikasi dengan memanfaatkan statistik dan probabilitas untuk mengklasifikasi data. Probabilitas untuk menemukan peluang terbesar dari kemungkinan klasifikasi, dengan melihat frekuensi masing – masing klasifikasi dalam data pelatihan [20]. Terdapat dua tahapan dalam algoritma *Naive Bayes*. Pertama akan dilakukan proses analisis sampel dokumen seperti penyeleksian kosa kata, kemudian dilakukan proses untuk menentukan probabilitas setiap kategori berdasarkan dokumen sampel. Kedua akan dilakukan klasifikasi berdasarkan kosa kata yang muncul dalam dokumen klasifikasi.

Penelitian Dianati Duei Putri, dkk tentang Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter, menggunakan metode *Naive Bayes Classifier*. Penelitian ini memperlihatkan bahwa algoritma *Naive Bayes* mendapatkan *accuracy score* sebesar 0.8 atau 80%. Artinya sistem mampu memprediksi 80% secara akurat dari total data testing sebesar 20%. Hasil analisis dapat digunakan untuk memprediksi dataset baru tanpa harus dilakukan *labeling* terlebih dahulu. Klasifikasi *tweet* dari twitter mengenai kinerja DPR menunjukkan, 95 positif, 693 netral dan 758 negatif dari data hasil *crawling* sebanyak 1546. [3]

Penelitian Imam Kurniawan dan Ajib Susanto (2019) berjudul Implementasi Metode *K-Means* dan *Naive Bayes Classifier* untuk Analisis Sentimen Pemilihan Presiden (Pilpres). *K-Means* digunakan sebagai metode untuk analisis sentimen *clustering* pada data latih dan menghasilkan bobot positif atau negatif pada tiap dokumen latih. Klasifikasi dokumen uji menggunakan *Naive Bayes*, dilakukan percobaan selama dua kali menggunakan *confusion matrix*. Dari hasil pengujian 100 dan 150 data uji tersebut didapatkan akurasi rata – rata sebesar 93,35% dan *error rate* rata – rata sebesar 6.66%. [4]

Penelitian Samsi, dkk (2021) tentang Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19, menggunakan metode *Naive Bayes*. Penelitian menerapkan algoritma *Naive Bayes* untuk klasifikasi sentimen menggunakan kata kunci 'pembelajaran daring', 'kuliah', 'belajar', 'online' dan tagar #BelajarDariRumah yang difilterkan dengan kata kunci 'online' dan 'rumah'. Hasil penelitian memperlihatkan 30% sentimen positif, 69% sentimen negatif dan 1% netral. Penelitian ini menyimpulkan pembelajaran daring belum maksimal diterapkan di Indonesia pada masa pandemi dimana kekecewaan publik sangat tinggi pada awal November 2020. [5]

Penelitian dan aplikasi berbasis sentimen berkembang pesat contohnya di Amerika terdapat sekitar 20 – 30 perusahaan menggunakan ini. Analisis ini penting karena membantu perusahaan untuk memahami pendapat orang terhadap entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa dan topik [7]. Sentimen Analisis merupakan salah satu cabang penelitian *Text Mining*, *Opinion Mining* adalah riset komputasional dari opini, sentimen dan emosi yang diekspresikan secara tekstual [7] [19].

Ramesh Sharda, dkk (2018;278) mengemukakan bahwa pengembangan data teks (*text mining*) dalam database teks adalah proses semi otomatis ekstraksi pola (informasi dan pengetahuan) dari sumber data yang tidak terstruktur dalam jumlah besar [8]. *Text Mining* merupakan bagian dari *data mining*, nilai komersial lebih tinggi dibandingkan *data mining*. Umumnya 80% dokumen informasi perusahaan dalam bentuk teks. Namun demikian penggalian dan pencarian data teks lebih kompleks dengan pola *text* tidak terstruktur selalu menyulitkan. *Text*

Mining merupakan lingkup penelitian yang komperhensif, yang masuk hampir disetiap lini kehidupan kita [9]. *Text Mining* merupakan bagian dari penambangan data (*data mining*). Klasifikasi dokumen tekstual, dimana dokumen akan diklasifikasikan menurut topik dokumennya, biasanya penerapan fungsi *text mining*. Kata – kata dalam sebuah artikel dapat digunakan untuk menentukan jenis kategori artikel dengan bantuan *text mining*. Oleh karena itu *text mining* dapat dengan cepat membantu dalam pengelompokkan dokumen [10].

Pengklasifikasi *Naive Bayes* adalah metode klasifikasi berdasarkan teorema *Bayes*. Metode klasifikasi ini menggunakan metode probabilistik dan statistik, yang memprediksi peluang masa depan berdasarkan pengalaman masa lalu, pertama kali dikemukakan oleh ilmuwan Inggris Thomas Bayes. Oleh karena itu, metode ini dikenal dengan teorema Bayes. Ciri utama dari *Naive Bayes classifier* adalah adanya asumsi yang sangat kuat mengenai independensi setiap kondisi atau peristiwa. Dalam bukunya, Olson dan Delen menjelaskan bahwa *Naive Bayes* menghitung probabilitas setiap kelas keputusan dengan syarat kelas keputusan tersebut benar. Algoritma ini didasarkan pada asumsi bahwa atribut suatu objek bersifat independen. Probabilitas prediksi akhir dihitung sebagai jumlah frekuensi dari tabel keputusan "master" [11].

II. METODE PENELITIAN

Tahapan penelitian ini, akan dimulai dari studi literatur, pengumpulan data / *crawl data*, *cleansing data*, *pre-processing data*, *translate*, klasifikasi menggunakan metode *Naive Bayes* dan simpulan. Penelitian ini menggunakan bahasa pemrograman *python* mulai dari proses *crawling* data sampai proses pengklasifikasian.

A. Studi Literatur

Persiapan penelitian ini melalui pembelajaran terhadap metode, langkah – langkahnya seperti pada gambar 1



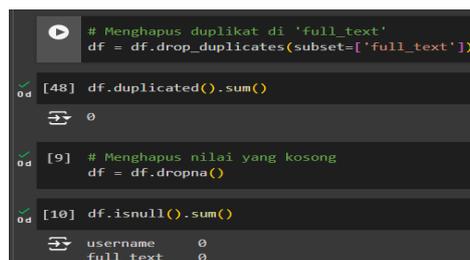
Gambar 1. Langkah- Langkah penelitian

B. Pengumpulan Data

Pengumpulan data dengan cara *mengcrawl* data twitter menggunakan API (Application Programming Interface) twitter. Kata kunci dalam pengumpulan data (*nama capres*) *Pemilu*. Data yang diambil dan digunakan dimulai dari tanggal 7 Januari 2024 (setelah debat ke-3 Presiden dan Wakil Presiden) - 14 Februari 2024 (PEMILU dilaksanakan). Data kemudian diubah dalam bentuk csv dan diproses menggunakan bahasa pemrograman *python*. Setelah itu atribut – atribut pada data yang tidak digunakan akan dihilangkan, Atribut yang digunakan untuk proses selanjutnya adalah atribut *full text* dan *username*.

C. Cleansing Data

Pada saat ini, nilai atau atribut yang identik akan dihapus dan nilai yang kosong akan dihapus. Selanjutnya, hilangkan *mention*, *hashtag*, *retweet*, *URL*, dan spasi yang berlebihan. Data *tweet* yang telah menjalani pembersihan data masih merupakan data mentah dan belum siap diolah untuk klasifikasi. Oleh karena itu dilakukan tahap pembersihan data untuk memperoleh data yang dapat diolah pada tahap selanjutnya



```
# Menghapus duplikat di 'full_text'
df = df.drop_duplicates(subset=['full_text'])

[48] df.duplicated().sum()
0

# Menghapus nilai yang kosong
df = df.dropna()

[10] df.isnull().sum()
username      0
full_text     0
```

Gambar 2. Menghapus duplikat dan menghapus nilai kosong

Gambar 2 menunjukkan proses penghapusan duplikat dan menghapus nilai yang kosong. Pada proses penghapusan duplikat menggunakan perintah di gambar. Perintah *subset* di gambar digunakan untuk menghapus nilai di

atribut *full text*. Pada perintah *df.duplicated().sum()* menunjukkan apakah masih ada duplikasi pada data atau tidak. Kemudian perintah *df = df.dropna()* bertujuan untuk menghapus nilai yang kosong dan perintah *df.isnull().sum()* menunjukkan hasil dari proses penghapusan nilai. Proses menghapus nilai yang sama dan kosong dilakukan berdasarkan data yang digunakan.

```
def bersihkan_data(text):  
    # Menghapus mention  
    text = re.sub(r'@[A-Za-z0-9_]+', '', text)  
    # Menghapus hashtag  
    text = re.sub(r'#\w+', '', text)  
    # Menghapus retweet  
    text = re.sub(r'RT[\s]+', '', text)  
    # Menghapus URL  
    text = re.sub(r'https?://\S+', '', text)  
    # Menghapus mention  
    text = re.sub(r'@[A-Za-z0-9 ]', '', text)  
    # Menghapus spasi yang berlebihan  
    text = re.sub(r'\s+', ' ', text).strip()  
  
    return text  
  
df['full_text'] = df['full_text'].apply(bersihkan_data)
```

Gambar 3. Menghapus *mention*, *hashtag*, *retweet*, *URL* dan spasi yang berlebihan

Pada Gambar 3 menunjukkan proses untuk menghapus *mention*, *hashtag*, *retweet*, *URL* dan spasi yang berlebihan. Data yang digunakan masih mentah, maka proses untuk membersihkan data untuk digunakan dalam proses selanjutnya.

D. Pre-processing Data

- Normalisasi : Normalisasi data yang dilakukan pada tahap ini merupakan perbaikan kata yang typo menjadi kata yang benar
- Stop Word : *Stop Word* merupakan langkah untuk menghapus kata yang tidak ingin digunakan dan diperlukan.
- Tokenizing : yaitu pemanggalan kalimat menjadi kata per kata.
- Stemming : Pada proses ini akan menghilangkan imbuhan seperti (pem, meng dan lainnya).

E. Translate

Pada tahap ini akan dilakukan *translate* dari bahasa Inggris ke bahasa Indonesia. *Package* yang digunakan adalah *translate-3.6.1* dan *libretranslatepy-2.1.1*. Proses *translate* ini diperlukan untuk mengklasifikasi data menjadi positif, netral dan negatif dengan metode *textblob*. Metode *textblob* lebih efektif dalam mengklasifikasi dengan data yang berbahasa Inggris.

```
# Translate bahasa Indonesia ke bahasa Inggris  
  
from translate import Translator  
  
def convert_eng(tweet):  
    translator = Translator(to_lang="en", from_lang="id")  
    translation = translator.translate(tweet)  
    return translation
```

Gambar 4. Translate dari bahasa Indonesia ke bahasa Inggris

Pada Gambar 4 menunjukkan perintah untuk mengubah data bahasa Indonesia ke dalam bahasa Inggris agar dapat diklasifikasikan menjadi sentimen positif, netral dan negatif. *Package* yang digunakan dalam proses ini adalah *package translate*.

F. Klasifikasi dengan Naive Bayes

Tahapan ini diklasifikasikan menggunakan metode Naive Bayes dengan data bersih. Proses klasifikasi menggunakan bahasa pemrograman Python untuk data pengujian dan pelatihan. Proses pemisahan data pengujian dan pelatihan menggunakan paket *train_test_split*. Data tersebut kemudian dianalisis dan dikategorikan menjadi sentimen positif, netral, dan negatif. Confusion matrix / matriks konfusi digunakan dalam proses evaluasi dan validasi hasil. Matriks konfusi merangkum hasil prediksi dari masalah klasifikasi. Matriks konfusi diterapkan dengan tujuan menganalisis seberapa baik model klasifikasi yang digunakan dapat membedakan kelas data yang berbeda.[12]

```
[40] x_train, x_test, y_train, y_test = train_test_split(x,y,test_size=0.2, random_state=42)
```

Gambar 5. Pembagian data latihan dan uji

Pada Gambar 5 menunjukkan perintah yang digunakan di *python* untuk pembagian data latih dan data uji. Variabel yang digunakan untuk membuat model dan memisahkan data adalah *train_test_split*, dimana pembagiannya adalah 80% data training dan 20% data uji sehingga *test_size* adalah 0,2.

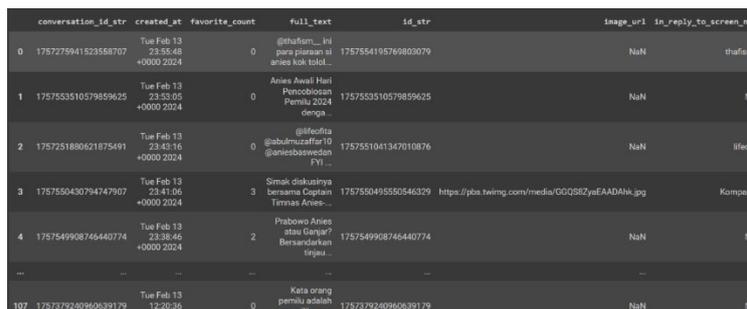
G. Simpulan

Dari data yang sudah diklasifikasikan akan dilakukan perbandingan sentimen masyarakat kepada tiap calon presiden dan penarikan kesimpulan

III. HASIL DAN PEMBAHASAN

Hasil dan pembahasan sebagai berikut : Pengumpulan data dilakukan menggunakan *crawling* data di twitter menggunakan API (*Application Programming Interface*). Pengumpulan data ini menggunakan *access token* dari twitter. *Access token* merupakan gabungan angka dan huruf, nanti *access token* ini akan menjadi alat untuk dimasukkan ke API untuk proses pengumpulan data [16]. Pengumpulan data twitter menggunakan kata kunci "*Anies pemilu 2024 lang:id until:2024-02-14 since:2024-01-07*", "*Prabowo pemilu 2024 lang:id until:2024-02-14 since:2024-01-07*" , "*Ganjar pemilu 2024 lang:id until:2024-02-14 since:2024-01-07*". Data yang di *crawling* berjumlah kurang lebih 300. Data dalam bahasa Indonesia dari tanggal 7 Januari 2024 – 14 – Februari 2024.

Tahapan ini dilakukan pembersihan data, seperti menghilangkan atribut yang tidak digunakan, duplikat pada atribut dan nilai, menghapus nilai yang kosong, menghapus *mention*, *hashtag*, *retweet*, *URL*, *mention*, spasi yang berlebihan. Atribut yang tidak diperlukan untuk proses pengklasifikasian ini dihapus agar lebih mudah dan jelas saat proses pengklasifikasian. Dikarenakan data yang digunakan masih data mentah maka tidak dipungkiri masih ada pendobelan (data yang sama) dan data kosong. Penghapusan nilai yang kosong dan pendobelan dilihat dari isi data di atribut *full_text*, jika ada data yang kosong atau pendobelan maka akan dilakukan penghapusan pada data tersebut.



	conversation_id_str	created_at	favorite_count	full_text	id_str	image_url	in_reply_to_screen_name
0	1757275941523558707	Tue Feb 13 23:55:48 +0000 2024	0	@thafam... ini para piraan si anies kok tolol...	1757554195769803079	NaN	thafam...
1	1757553510579859625	Tue Feb 13 23:53:05 +0000 2024	0	Anies Awal Hari Pencoblosan Pemilu 2024 denga...	1757553510579859625	NaN	NaN
2	1757251880621875491	Tue Feb 13 23:45:16 +0000 2024	0	@lifecofita @gabumozaffar10 @anisabawendon FYI ...	1757551041347010876	NaN	lifecof...
3	1757550430794747907	Tue Feb 13 23:41:06 +0000 2024	3	Simak diskusinya bersama Captain Timnas Anies...	1757550495550546329	https://pbs.twimg.com/media/GGQ28ZyeEAAADAhk.jpg	Kompass...
4	1757549908746440774	Tue Feb 13 23:38:46 +0000 2024	2	Prabowo Anies atau Ganjar? Berdasarkan tinjau...	1757549908746440774	NaN	NaN
...
107	1757379240960639179	Tue Feb 13 12:20:36	0	Kata orang pemilu adalah...	1757379240960639179	NaN	NaN

Gambar 5. Data mentah dari hasil *crawling data*

Pada Gambar 5 menunjukkan data mentah yang didapatkan dari *crawling data* dengan kata kunci "*Anies pemilu 2024 lang:id until:2024-02-14 since:2024-01-07*". Data yang diambil hanya dari tanggal 7 Januari 2024 – 14 Februari 2024. Pada data ini masih memiliki atribut lengkap seperti *conversation_id_str*, *created_at*, *favourite_count*, *full_text*, *ide_str*, *image_url*, dan lainnya.



	username	full_text
0	iamwerewolf29	ini para piraan si anies kok tolol tolol amat...
1	IDNTimes	anies awal hari pencoblosan pemilu 2024 denga...
2	dewi_nanay	fyi jg bang masa tenang itu berlaku utk pelaks...
3	KompasTV	simak diskusinya bersama captain timnas aniesm...
4	SinarOnline	prabowo anies atau ganjar bersandarkan tinjaua...
...
95	musniumar_	anies tidak pernah politisasi dirty vote semua...
96	keynara181	dan inshaallah besok ustadz akan dpt hadiah t...
97	dandelonjkey	jejak besok pemilu pak anies menang yuk bisa yukk
98	StanKumisDowoon	siapa big wkwk anies lg brhenti diselidiki krn...
99	zahidahaisyh	kampanye pak anies bisa dibilang cukup unik da...

Gambar 6. Data yang sudah dilakukan proses *cleansing*

Pada Gambar 6 menunjukkan data yang sudah dilakukan proses *cleansing*, seperti menghilangkan *mention*, *hashtag*, *retweet*, dan lainnya. Kemudian pada Gambar 6 juga menunjukkan bahwa hanya 2 atribut yang digunakan untuk proses yaitu *username* dan *full_text*. Semua huruf besar pada data sudah diubah menjadi huruf kecil.

Pada tahap *pre-processing* data akan dilakukan normalisasi *stopword*, *tokenizing* dan *stemming*. Normalisasi dilakukan untuk membenarkan kata yang *typo*, penghapusan nama capres lainnya dan lainnya. Contohnya, jika data yang digunakan adalah data Anies, maka akan dihilangkan kata Prabowo dan Ganjar, begitupun sebaliknya.

	username	full_text
0	iamwerewolf29	ini para piaraan si anies kok tolol tolol amat...
1	IDNTimes	anies awali hari pencoblosan pemilu 2024 denga...
2	dewi_nanay	fyi jg bang masa tenang itu berlaku utk pelaks...
3	KompasTV	simak diskusinya bersama captain timnas aniesm...
4	SinarOnline	prabowo anies atau ganjar bersandarkan tinjaua...
...
95	musniumar_	anies tidak pernah politisasi dirty vote semua...
96	keynara181	dan inshaaallah besok ustadz akan dpt hadiah t...
97	dandelionjkey	jejak besok pemilu pak anies menang yuk bisa yukk
98	StanKumisDwoon	siapa blg wkwk anies lg brhenti diselidiki krn...
99	zahidahaisyh	kampanye pak anies bisa dibilang cukup unik da...

100 rows x 2 columns

Gambar 7. Data yang belum dilakukan normalisasi

Pada Gambar 7 menunjukkan data yang belum dilakukan proses normalisasi, dimana pada Gambar 7 masih terdapat kata "Prabowo", "Ganjar". Pada tahap normalisasi nama – nama capres lainnya akan dihapus dan akan dilakukan pembenaran kata – kata typo.

	username	full_text
0	iamwerewolf29	peliharaan si anies kok tolol sih podcast ttng...
1	IDNTimes	anies awali hari pencoblosan pemilu 2024 salat...
2	dewi_nanay	informasi pak masa tenang berlaku utk pelaksan...
3	KompasTV	simak diskusinya bersama captain timnas aniesm...
4	SinarOnline	anies bersandarkan tinjauan pendapat minggu l...
...
95	musniumar_	anies pernah politisasi dirty vote semua dimua...
96	keynara181	inshaaallah besok ustadz dpt hadiah terindah l...
97	dandelionjkey	jejak besok pemilu pak anies menang yuk yukk
98	StanKumisDwoon	siapa blg wkwk anies lg brhenti diselidiki krn...
99	zahidahaisyh	kampanye pak anies dibilang cukup unik menjadi...

100 rows x 2 columns

Gambar 8. Data yang sudah dilakukan normalisasi

Pada Gambar 8 menunjukkan data yang sudah dilakukan normalisasi seperti hilangnya kata "Prabowo", "Ganjar". Kemudian pembenaran kata – kata yang *typo*.

Pada tahapan *Stop Word* dilakukan untuk menghapus kata – kata yang tidak diperlukan atau kurang penting seperti "yang" dan lainnya. Kemudian pada tahap *tokenizing* akan mengubah kalimat menjadi kata per kata.

```

0 [para, peliharaan, si, anies, kok, tolol, tolo...
1 [anies, awali, hari, pencoblosan, pemilu, 2024...
2 [informasi, juga, pak, masa, tenang, berlaku, ...
3 [simak, diskusinya, bersama, captain, timnas, ...
4 [anies, bersandarkan, tinjauan, pendapat, ming...
...
95 [anies, pernah, politisasi, dirty, vote, semua...
96 [inshaaallah, besok, ustadz, dpt, hadiah, teri...
97 [jejak, besok, pemilu, pak, anies, menang, yuk...
98 [siapa, blg, wkwk, anies, lg, brhenti, diselid...
99 [kampanye, pak, anies, dibilang, cukup, unik, ...
Name: full_text, Length: 100, dtype: object
    
```

Gambar 9. Data yang sudah dilakukan *tokenizing*

Pada Gambar 9 menunjukkan data yang sudah dilakukan *tokenizing* dimana kalimat dalam data diubah menjadi kata yang akan digunakan untuk proses selanjutnya. Kemudian pada proses *stemming* dihilangkan imbuhan

seperti "meng", "an", "pem" dan lainnya. Contoh perubahan kata menjadi kata dasar seperti dibilang, menjadi bilang dan lainnya.

	username	full_text
0	iamwerewolf29	peliharaan si anies kok tolol sih podcast ttng...
1	IDNTimes	anies awali hari pencoblosan pemilu 2024 salat...
2	dewi_nanay	informasi pak masa tenang berlaku utk pelaksan...
3	KompasTV	simak diskusinya bersama captain timnas aniesm...
4	SinarOnline	anies bersandarkan tinjauan pendapat minggu l...
...
95	musniumar_	anies pernah politisasi dirty vote semua dimua...
96	keynara181	inshaaallah besok ustadz dpt hadiah terindah l...
97	dandelionjkey	jejak besok pemilu pak anies menang yuk yukk
98	StanKumisDowoon	siapa blg wkwk anies lg brhenti diselidiki krn...
99	zahidahaisyh	kampanye pak anies dibilang cukup unik menjadi...

Gambar 10. Data yang belum dilakukan proses *stemming*

Pada Gambar 10 menunjukkan data yang belum dilakukan proses *stemming*, dimana data – data tersebut masih menggunakan imbuhan seperti "diskusinya", "bersandarkan", dan kata lainnya.

	full_text
0	para piara si anies kok tolol tolol sih podcas...
1	anies awal hari coblos milu 2024 salat subuh j...
2	fyi bang masa tenang laku laksana serta tim ka...
3	simak diskusi sama captain timnas aniesmuhaimi...
4	prabowo anies ganjar sandar tinjau dapat mingg...
...	...
92	anies pernah politisasi dirty vote semua muat ...
93	inshaaallah besok ustadz dpt hadiah indah yait...
94	jejak besok milu pak anies menang yuk yukk
95	siapa blg anies brhenti selidik lagi nyapres k...
96	kampanye pak anies bilang cukup unik jadi angi...

Gambar 11. Data yang sudah dilakukan proses *stemming*

Pada Gambar 11 menunjukkan data yang sudah dilakukan proses *stemming* seperti kata "diskusinya" menjadi "diskusi", "bersandarkan" menjadi "sandar" dan kata – kata lainnya. Kemudian proses *translate*. Pada tahap ini akan dilakukan *diterjemahkan* bahasa Indonesia ke bahasa Inggris. Terjemahaman karena metode yang digunakan adalah *textblob*, *textblob* lebih efektif saat melakukan analisis ketika data dalam bahasa Inggris. etode *Textblob* adalah perpustakaan untuk memproses data teks. *TextBlob* menyediakan API yang dapat digunakan untuk pemrosesan bahasa alami (NLP) seperti ekstraksi frasa kata benda, analisis sentimen, klasifikasi, dan terjemahan., dan lain – lain. *TextBlob* menganalisis data dengan menggunakan pola model *pre-trained* untuk menilai teks sebagai positif, negatif atau netral. *Textblob* menggunakan model berbasis metode *naive bayes yang dilatih pada dataset untuk mengklasifikasikan teks berdasarkan polaritas dan subjektivitas*. [13]

	full_text	tweet_english
0	para piara si anies kok tolol tolol sih podcas...	the petara si anies kok moron silly anyway pod...
1	anies awal hari coblos milu 2024 salat subuh j...	anies beginning of the day coblos milu 2024 co...
2	fyi bang masa tenang laku laksana serta tim ka...	fyi bang quiet period of behavior and the milu...
3	simak diskusi sama captain timnas aniesmuhaimi...	see the discussion with the captain of the nat...
4	prabowo anies ganjar sandar tinjau dapat mingg...	prabowo anies odd backup survey got last week ...
...
92	anies pernah politisasi dirty vote semua muat ...	anies ever politicized dirty vote all loaded t...
93	inshaaallah besok ustadz dpt hadiah indah yait...	inshaaallah tomorrow ustadz can get a beautifu...
94	jejak besok milu pak anies menang yuk yukk	trail tomorrow milu pak anies win yuk yukk
95	siapa blg anies brhenti selidik lagi nyapres k...	who is blg anies brhentearech again hispres if...
96	kampanye pak anies bilang cukup unik jadi angi...	the Pak Anies campaign said it was quite uniqu...

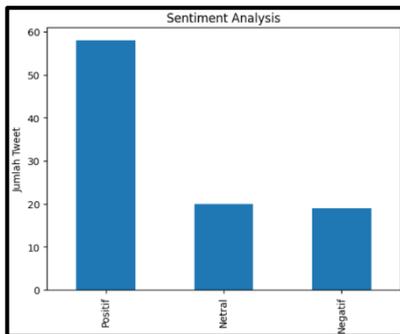
Gambar 12. Data yang sudah di *translate* ke bahasa Inggris

Pada Gambar 12 menunjukkan data yang sudah dilakukan proses *translate* dari bahasa Indonesia ke bahasa Inggris. Data *ditranslate* dikarenakan *package* yang digunakan hanya bisa mengklasifikasi data dalam bahasa Inggris. Kemudian dilakukan analisis dan pengklasifikasian untuk mendapatkan klasifikasi sentimen menjadi positif, negatif dan netral.

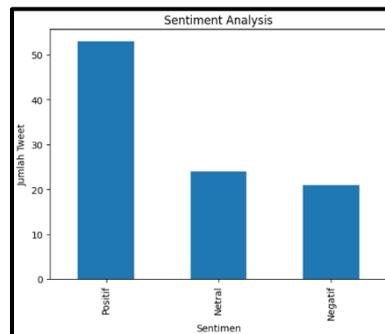
	full_text	tweet_english	Subjectivity	Polarity	Analysis
0	para piara si anies kok tolol tolol sih podcas...	the petara si anies kok moron silly anyway pod...	0.937500	-0.650000	Negatif
1	anies awal hari coblos milu 2024 salat subuh j...	anies beginning of the day coblos milu 2024 co...	0.000000	0.000000	Netral
2	fyi bang masa tenang laku laksana serta tim ka...	fyi bang quiet period of behavior and the milu...	0.333333	0.000000	Netral
3	simak diskusi sama captain timnas aniesmuhami...	see the discussion with the captain of the nat...	0.535714	0.203571	Positif
4	prabowo anies ganjar sandar tinjau dapat mingg...	prabowo anies odd backup survey got last week ...	0.362500	0.313889	Positif

Gambar 13. Data yang sudah diklasifikasikan

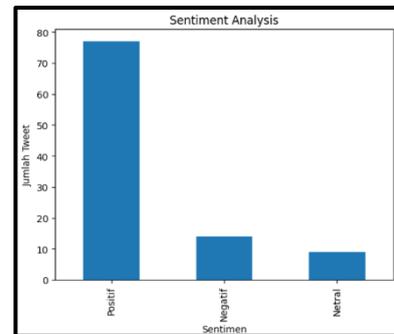
Pada Gambar 13 menunjukkan data yang sudah diklasifikasi menjadi 3 yaitu positif, negatif dan netral. Data menunjukkan atribut *subjectivity* dan *polarity*. Faktor yang menentukan data terbagi menjadi positif, netral dan negatif mengikuti cara kerja dari *textblob*. *Textblob* menggunakan pola model *pre-trained* untuk menilai teks sebagai positif, negatif atau netral. *Textblob* menggunakan model berbasis metode *naive bayes* yang dilatih pada dataset untuk mengklasifikasikan teks berdasarkan polaritas dan subjektivitas.[13] Dari data yang sudah diklasifikasi akan dilanjutkan ke tahap visualisasi data dalam bentuk diagram batang. Data yang didapatkan pada capres Anies adalah positif 58, netral 20 dan negatif 19. Pada capres Prabowo mendapatkan positif 53, netral 24 dan negatif 21. Sedangkan pada capres Ganjar mendapatkan sentimen positif 77, netral 9 dan negatif 14.



Gambar 14. Sentimen Anies



Gambar 15. Sentimen Prabowo



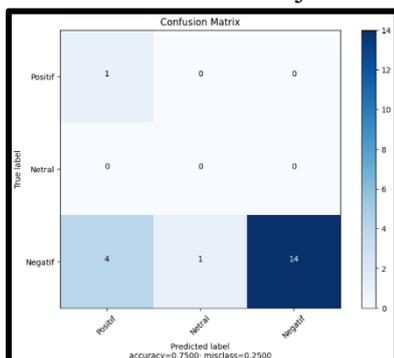
Gambar 16. Sentimen Ganjar

Pada Gambar 14 menunjukkan diagram batang hasil analisis sentimen capres Anies yang dimana hasil sentimen positif berjumlah 58, netral 20 dan negatif 19.

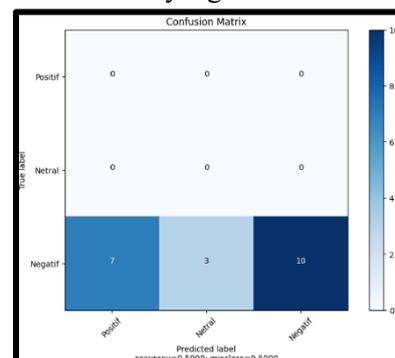
Pada Gambar 15 menunjukkan diagram batang hasil analisis sentimen capres Prabowo yang dimana hasil sentimen positif berjumlah 53, netral 24 dan negatif 21.

Pada Gambar 16 menunjukkan diagram batang hasil analisis sentimen capres Ganjar yang dimana hasil sentimen positif berjumlah 77, netral 14 dan negatif 9.

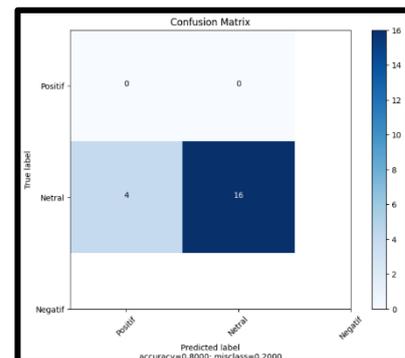
Kemudian data dibuat dalam bentuk *Confusion Matrix* dimana data pada capres Anies mendapatkan akurasi = 0,75 dan *misclass* = 0,25, sedangkan pada data capres Prabowo mendapatkan akurasi 0,5 dan *misclass* = 0,5 dan pada data capres Ganjar mendapatkan akurasi = 0,8 dan *misclass* = 0,2. Akurasi merupakan metode pengujian algoritma berdasarkan tingkat kedekatan antara nilai prediksi dengan nilai sebenarnya. *Misclass* menunjukkan seberapa sering model klasifikasi membuat prediksi yang salah [14]. Oleh karena itu *accuracy* menunjukkan akurasi dan *misclass* menunjukkan model klasifikasi yang salah.



Gambar 17. Confusion matrix Anies



Gambar 18. Confusion matrix Prabowo



Gambar 19. Confusion matrix Ganjar

Pada Gambar 17 menunjukkan *confusion matrix* dari hasil analisis sentimen dari capres Anies dimana mendapatkan *accuracy* = 0,75 dan *misclass* = 0,25. Hasil ini membuktikan bahwa analisis yang dilakukan terhadap capres Anies cukup baik / cukup akurasi karena hampir mencapai 80% atau 0,8 persen.

Pada Gambar 18 menunjukkan *confusion matrix* dari hasil analisis sentimen dari capres Prabowo dimana mendapatkan *accuracy* = 0,5 dan *misclass* = 0,5. Hasil ini menunjukkan bahwa analisis yang dilakukan terhadap capres Prabowo belum baik dikarenakan akurasi dan misclass / kesalahan prediksi memiliki nilai yang sama.

Pada Gambar 19 menunjukkan *confusion matrix* dari hasil analisis sentimen dari capres Ganjar dimana mendapatkan *accuracy* = 0,8 dan *misclass* = 0,2. Hasil ini menunjukkan bahwa analisis yang dilakukan terhadap capres Ganjar sangat baik karena nilai akurasi mendapatkan 0,8% atau 80% dan nilai dari misclass hanya 0,2% atau 20%

Setelah dilakukan proses *confussion matrix* maka data di *confusion matrix* akan dihitung *precision*, *recall*, *f1-score* dan *support*. Proses penghitungan ini untuk memvalidasi dan mengevaluasi data yang sudah diproses di *confusion matrix*. *Precision* merupakan perhitungan yang menentukan jumlah true positif dari seluruh hasil prediksi positif. Penarikan kembali adalah perhitungan yang menentukan jumlah data yang mewakili prediksi positif-benar dari semua hasil positif-benar. Skor F1 merupakan perhitungan yang rata-rata perbandingan antara presisi dan recall. [15].

Classification report :				
	precision	recall	f1-score	support
Negatif	1.00	0.20	0.33	5
Netral	0.00	0.00	0.00	1
Positif	0.74	1.00	0.85	14
accuracy			0.75	20
macro avg	0.58	0.40	0.39	20
weighted avg	0.77	0.75	0.68	20

Gambar 20. Classification report Anies

Classification report :				
	precision	recall	f1-score	support
Negatif	0.00	0.00	0.00	7
Netral	0.00	0.00	0.00	3
Positif	0.50	1.00	0.67	10
accuracy			0.50	20
macro avg	0.17	0.33	0.22	20
weighted avg	0.25	0.50	0.33	20

Gambar 21. Classification Report Prabowo

Classification report :				
	precision	recall	f1-score	support
Negatif	0.00	0.00	0.00	4
Positif	0.80	1.00	0.89	16
accuracy			0.80	20
macro avg	0.40	0.50	0.44	20
weighted avg	0.64	0.80	0.71	20

Gambar 22. Classification Report Ganjar

Pada Gambar 20 menunjukkan Classification Report capres Anies berdasarkan confusion matrix capres Anies. Perhitungan dari classification report capres Anies baik itu precision, recall f1-score dan lainnya mendapat akurasi 0,75%, dimana akurasi yang didapatkan cukup baik dalam mengklasifikasi.

Pada Gambar 21 menunjukkan Classification Report capres Prabowo berdasarkan confusion matrix capres Prabowo. Perhitungan dari classification report capres Prabowo baik itu precision, recall f1-score dan lainnya mendapat akurasi 0,5 %, dimana akurasi yang didapatkan belum baik dalam mengklasifikasi.

Pada Gambar 22 menunjukkan Classification Report capres Ganjar berdasarkan confusion matrix capres Ganjar. Perhitungan dari classification report capres Ganjar baik itu precision, recall f1-score dan lainnya mendapat akurasi 0,8 %, dimana akurasi yang didapatkan baik dalam mengklasifikasi.

Berdasarkan hasil pengujian sentimen masyarakat tentang calon presiden di tahun 2024 dengan pembagian data 80% data training dan 20% data uji maka mendapat hasil akurasi dari setiap capres yang berbeda baik itu hasil yang kurang baik, cukup baik dan baik. Hasil ini berbeda dengan jurnal yang ditulis oleh Jepi, Debby dan Auliya tentang "Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring" dimana pembagian data uji dan data latih dengan jumlah yang sama yaitu 80 % data training dan 20% data uji dan mendapatkan hasil akurasi sebesar 84,93%, dimana hasil ini adalah hasil yang baik. Perbandingan – perbandingan ini menjadi manfaat untuk penelitian yang akan dilakukan kedepan dengan membandingkan akurasi dari setiap metode yang digunakan baik itu metode naive bayes, k-nn dan lainnya [7].

IV. SIMPULAN

Berdasarkan penelitian dan pengolahan data yang berjumlah kurang lebih 300, maka dapat ditarik kesimpulan bahwa dengan menggunakan metode *Naive Bayes* dapat melakukan klasifikasi terhadap sentimen masyarakat pada aplikasi twitter. Sentimen masyarakat di twitter rata – rata, merupakan sentimen positif yang artinya capres yang ikut dalam PEMILU ini merupakan calon presiden yang baik di mata masyarakat pengguna twitter. Penggunaan algoritma *Naive Bayes* dalam penelitian ini memiliki rata – rata akurasi 0,63% dari 3 data capres berjumlah kurang lebih 300 data secara total dan 100 pada masing – masing capres yang artinya proses klasifikasi yang dilakukan belum terlalu baik. Oleh karena itu saran untuk penelitian berikutnya dapat menggunakan algoritma klasifikasi lainnya untuk membandingkan keakuratan akurasi pada data yang digunakan.

DAFTAR PUSTAKA

- [1] Nabilah Muhamad (2023, 05 Juli). Jumlah Pemilu 2024 Berdasarkan Kelompok Usianya (Juli 2023). Diakses pada 24 oktober 2023, dari <https://databoks.katadata.co.id/datapublish/2023/07/05/kpu-pemilih-pemilu-2024-didominasi-oleh-kelompok-gen-z-dan-milenial>
- [2] Cindy Mutia Annur (2023, 31 Mei) 10 Negara dengan Jumlah Pengguna Twitter Terbanyak di Dunia (April 2023). Diakses pada 30 Oktober 2023, dari <https://databoks.katadata.co.id/datapublish/2023/05/31/jumlah-pengguna-twitter-di-indonesia-capai-1475-juta-per-april-2023-peringkat-keenam-dunia>
- [3] Putri, Duei Dianati; Nama, Gigih Forda; Sulistiono, Wahyu Eko, 2022. "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier". Jurnal Informatika dan Teknik Elektro Terapan (JITET). Vol 10 No. 1.
- [4] Kurniawan, Imam; Susanto, Ajib, 2019. "Implementasi Metode K-Means dan Naive Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019". Jurnal Eksplorasi Informatika.
- [5] Samsir; Ambiyar; Verawardina, Unung; Firman Edi; Watrianthos, Ronal. 2021. "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naive Bayes". Jurnal Media Informatika Budidarma. Vol 5 No. 1.
- [6] Buntoro, Ghulam Asrofi. 2017. "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter". *Integer Journal* Vol 2 No. 1
- [7] Supriyanto, Jepi; Alita, Debby; Isnain, Auliya Rahman. 2023. "Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring". Jurnal Informatika dan Rekayasa Perangkat Lunak (JATIKA). Vol 4, No. 1
- [8] Lutfiansyah, Raihanda; Wasito, Budi. 2023. "Analisis Sentimen Terhadap Para Kandidat Presiden 2024 Berdasarkan Netizen Pengguna Twitter Dengan Metode *Data Mining* dan *Text Mining*".
- [9] Firdaus, Ali; Firdaus, Wahyu Istalama. 2021. "*Text Mining* Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi. Jurnal Penelitian dan Ilmu Komputer. Vol 13 No. 1
- [10] Arsad, Asriyani; Mustamin, Hamid; Santosa, M. 2024. "Penerapan Teks Mining dan *Cosine Similarity* Untuk Menentukan Kesamaan Dokumen Skripsi". Indonesian Journal on Information System.
- [11] Mustafa, M. Syukri; Ramadhan, Muh Rizky; Thenata, Angelina P. 2017. "Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier". *Cite e Journal*. Vol 4 No. 2.
- [12] Farmati, Narti. 2022. "Perbandingan Algoritma C4.5 dan Naive Bayes dalam Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Daring". Jurnal Teknologi Informasi dan Multimedia. Vol 4 No. 1, hlm 1-12.
- [13] Parlita, Rizky; Pradika, Sunu Ilham; Hakim, Amir Muhammad, N.M, Kholilus Rachman. 2020. "Analisis Sentimen Twitter Terhadap Bitcoin dan Cryptocurrency Berbasis Python Textblob". Jurnal Ilmiah Teknologi Informasi dan Robotika, Vol 2 No. 2
- [14] Ayudhitama, Annisa Putri; Pujiyanto, Utomo. 2020. "Analisa 4 Algoritma Dalam Klasifikasi Penyakit Liver Menggunakan Rapidminer". Jurnal Informatika Polinema. Vol 6.
- [15] Handayani, Fitri; Kusuma, Kartika Sari; Asbudi, Hedy Leoni; Purnasiwi, Rona Guines; Kusuma, Resti; Sunyoto, Andi; Pradnya, Windha Mega. 2021. "Komparasi Support Vector Machine, Logistic Regression dan Artificial Neural Network dalam Prediksi Penyakit Jantung". Jurnal Edukasi dan Penelitian Informatika, Vol 7 No. 3.
- [16] Tim Hapus Tweet (2024, 11 Juni). Apa itu API Twitter? Semua Detail Tentang Antarmuka Ini. Diakses pada 22 Juni 2024, dari <https://tweetdelete.net/id/resources/what-is-twitter-api-all-the-details-about-this-interface/>.
- [17] Admin SMP (14 Februari 2024). Dibalik Sejarah Pemilihan Umum Republik Indonesia. Diakses pada 22 Juni 2024, dari <https://ditsmp.kemdikbud.go.id/dibalik-sejarah-pemilihan-umum-republik-indonesia/>.
- [18] Setiyawati, Dewi; Cahyono, Nuri. 2023. "Analisa Sentimen Pengguna Sosial Media Twitter Terhadap Perokok di Indonesia". Indonesia Journal of Computer Science, Vol 12. No. 1.
- [19] Marga, Nurman Satya; Isnain, Auliya Rahman; Alita, Debby. 2020. "Sentimen Analisis Tentang Kebijakan Pemerintah Terhadap Kasus Corona Menggunakan Metode Naive Byes". Jurnal Informatika dan Rekayasa Perangkat Lunak (JATIKA), Vol. 2 No. 4
- [20] Wibisono, Aria. 2023. "Filtering Spam Email Menggunakan Metode Naive Bayes. Jurnal Teknologi Pintar, Vol 3 No. 4