ANALISIS SENTIMEN ISU KECURANGAN PEMILU 2024 BERDASARKAN OPINI PADA MEDIA SOSIAL *TWITTER* MENGGUNAKAN METODE CRISP-DM DENGAN ALGORITMA *NAÏVE BAYES CLASSIFIER*

Aab Abdullah Muttaqin, Syariful Alam, Mutiara Andayani Komara
Teknik Informatika, STT Wastukancana Purwakarta
Jl. Cikopak No.53, Sadang Purwakarta Jawa Barat – Indonesia
Abdullahaab796@gmail.com

ABSTRAK

Pemilu yang berlangsung pada tanggal 14 Februari 2024 di Indonesia telah menimbulkan berbagai pendapat tentang dugaan kecurangan pemilu di *platform* media sosial, terutama Twitter. Hal ini menjadi perhatian penting karena dapat mengancam stabilitas demokrasi di Indonesia, yang diatur dalam UU No 7 Tahun 2017 tentang Pemilihan Umum. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap dugaan isu kecurangan pemilu yang diungkapkan melalui *Twitter*, yang kemudian diklasifikasikan menjadi sentimen positif, negatif, dan netral. Dari 1857 *tweet* yang dianalisis setelah proses *text preprocessing*, terdapat 1810 *tweet* yang berhasil diidentifikasi: 821 *tweet* sentimen positif dengan 422 sentimen negatif, dan 567 *tweet* dengan sentimen netral. Metode klasifikasi yang digunakan adalah algoritma *Naïve Bayes*, yang menghasilkan tingkat akurasi sebesar 66.85%, Selain itu, nilai *Precision* 68.05%, *Recall* 56.11%, dan *F1-Score* 61,56%.

Kata kunci: *Analisis sentimen, Naïve Bayes,Pemilu curang*

1. PENDAHULUAN

Pemilihan umum di Indonesia telah dilaksanakan pada 14 Februari 2024. *Twitter* menjadi sumber informasi yang penting bagi masyarakat dalam menyuarakan pendapat mengenai pemilu, meskipun terkadang cuitan-cuitan tersebut tidak jelas atau salah. Hasil studi pemilu 2024 menemukan berbagai sentimen positif dan negatif yang dapat menjadi acuan untuk pemilu di masa depan. Hasil penelitian ini mencakup pemilihan presiden, gubernur, dan anggota legislatif. Agar opini yang disampaikan lebih bermanfaat dan informatif [1].

Indonesia merupakan salah satu negara dengan kemajuan yang sangat pesat dalam bidang teknologi informasi, terlihat dari jumlah pengguna sosial media yang semakin bertambah disetiap harinya, baik itu Facebook, Instagram, Twitter, dan sosial media lainnya. Pengguna Twitter di Indonesia mencapai angka 19,5 juta pada tahun dan menjadikan yang terbanyak ke empat di dunia [2].

Tiga faktor utama yang menyebabkan malapraktik pemilu di Indonesia dengan kuatnya hubungan patronase antara penyelenggara pemilu, calon legislatif (caleg), dan pemilih. Patronase politik mencakup *vote buying* sebagai salah satu bentuknya, selain bentuk-bentuk lain seperti *pork barrel projects, club goods* dan *gifts.* patronase itu sendiri merujuk pada "materi atau keuntungan lain yang diberikan oleh politisi kepada pemilih atau pendukung [3].

Beredarnya isu dan berita terkait kecurangan pemilu tidak harus menjadi perhatian Masyarakat fenomena ini sudah dialami oleh Indonesia sejak awal gerakan pemilu beberapa tahun yang lalu. sebelumnya [4]. Pada pemilu 2024, dugaan kecurangan dianggap lebih parah, dengan laporan indikasi perbedaan jumlah suara di formulir C1 dan hasil rekapitulasi di tingkat kecamatan, menimbulkan keraguan dan spekulasi publik

Dalam pemprosesan analisis sentimen terdapat beberapa metode yang bisa digunakan, salah satunya metode *Naïve Bayes*. Metode *Naïve Bayes ini* merupakan salah satu algoritma dari *machine learning*. Dalam perkembangan *database*, *Naïve Bayes Term*asuk *supervised learning* yaitu jenis *machine learning* yang membutuhkan sampel sebagai data latih yang memiliki label [5].

Analisis sentimen dengan algoritma *Naïve Bayes Classifier* bertujuan memahami opini publik tentang kecurangan pemilu 2024 di Twitter. Menggunakan metode CRISP-DM, penelitian ini fokus pada identifikasi dan analisis sentimen dalam *tweet* terkait isu tersebut. Algoritma *Naïve Bayes* dipilih karena kemampuannya mengklasifikasikan teks berdasarkan probabilitas fitur. Penelitian ini diharapkan memberikan wawasan tentang pandangan masyarakat mengenai kecurangan pemilu 2024. Pada penelitian A. Perdana, A. Hermawan, dan D. Avianto sebelum nya mengenai tentang pemilu menunjukan hasil dari penelitianya memberikan jumlah akhir Dalam penelitian sebelumnya, model *Naïve Bayes Classifier* menunjukkan tingkat akurasi yang tinggi [6]. kemudian dalam Penelitian S. Puad dan A. Susilo Yuda Irawan sebelumnya menunjukkan bahwa analisis sentimen media sosial dapat memberikan wawasan berharga tentang pandangan masyarakat terhadap Pemilihan Umum 2024 [7].

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen digunakan untuk Analisis sentimen digunakan untuk mengidentifikasi apakah sentimen yang diungkapkan dalam teks bersifat

positif, negatif, atau netral. Metode ini sangat populer dan memiliki dampak signifikan di berbagai bidang kehidupan.[8]. Penelitian analisis bertujuan untuk memahami informasi akurat tentang peristiwa dengan mengeksplorasi asal-usul, penyebab, dan alasan. Analisis sentimen, sebagai bagian dari *text mining*, berfokus pada pengolahan data dari aktivitas tertentu [9].

2.2. Kecurangan

Kecurangan disebut sebagai pelanggaran, kecurangan diartikan *Fraud* Menurut Kamus Besar Bahasa Indonesia (KBBI), istilah "curang" mengacu pada sesuatu yang tidak jujur atau tidak adil. atau menipu. Menurut ACFE *(Association of Certified Fraud Examiners)*, didefinisikan sebagai ketidak jujuran yang dilakukan dengan tujuan untuk mendapatkan keuntungan finansial dengan dengan menggunakan penipuan sebagai modus operasi utamanya [10].

2.3. *Naïve Bayes Classifier*

Klasifikasi *Naïve Bayes* adalah supervised learning karena melibatkan manusia yang mengklasifikasikan data untuk pelatihan. Algoritma ini memiliki waktu klasifikasi singkat, mempercepat analisis sentiment [11]. Algoritma *Naïve Bayes* juga sering digunakan untuk klasifikasi secara probabilistik, menghitung probabilitas setiap pada setiap data untuk menghasilkan klasifikasi sentimen positif, negatif, dan netral, dengan pembobotan TF-IDF [12].

2.4. *Twitter*

Twitter adalah *platform* media sosial di mana orang-orang dari seluruh dunia dapat mengekspresikan pendapat mereka. Data yang diperoleh dari *Twitter* memiliki potensi besar untuk dianalisis menjadi informasi berharga melalui analisis sentimen,Pendapat menganai berita, peluncuran produk, atau peristiwaa terkini dapat didiskusikan secara efektif Twitter [13].

2.5. *Confusion Matrix*

Confusion Matrix adalah alat yang berguna untuk mengevaluasi kinerja model prediksi. Dengan menggunakan metode ini, kita dapat membandingkan nilai aktual dengan nilai prediksi, dan menghasilkan ukuran seperti akurasi, presisi, dan *Recall*. *True Positive* (TP) mengacu pada data yang diprediksi positif dan benar-benar positif, sedangkan *True Negative* (TN) merujuk pada data yang diprediksi negatif dan benar-benar negatif. *False Positive* (FP) adalah data yang sebenarnya negatif tetapi diprediksi positif, dan *False Negative* (FN) adalah data yang sebenarnya positif tetapi diprediksi negatif [14].

Tabel 1. Struktur *confusion matrix*

Confusion Matrix		***Actual Value***		
		Positive	***Negative***	***Neutral***
Predicted Value	***Positive***	*(TP)*	*(FNe)*	*(FNt)*
	Negative	*(FP)*	*(TNe)*	*(FNt)*
	Neutral	*(FP)*	*(FNe)*	*(TNt)*

Tabel 1 menunjukkan struktur *Confusion Matrix* untuk membandingkan hasil klasifikasi sistem dengan klasifikasi sebenarnya, serta menghitung akurasi. *True Positive (TP)* adalah data positif yang dapat diverifikasi, dan *True Neutral* (TNt) adalah data negatif yang dapat diverifikasi. *True Negative* (TNe) mengacu pada fakta negatif yang dapat diverifikasi. Data yang merupakan *False Positive* (FP) adalah positif yang salah, *False Neutral* (FNt) adalah data netral yang salah, dan *False Negative* (FNe) adalah data negatif yang salah diklasifikasikan [15].

Metode *Confusion Matrix* memiliki beberapa parameter evaluasi yang berkaitan dengan model klasifikasi,yaitu sebagai berikut:

2.5.1. *Precision*

Merupakan parameter penilaian yang menghitung nilai rata-rata *precission* dari data hasil klasifikasi yaitu jumlah data yang benar antara nilai sebenarnya dengan hasil prediksi model yang dapat dihitung dengan rumus seperti pada persamaan berikut:

$$Precission = \frac{\sum_i^n \frac{TPi}{TPi + FPi}}{n} \quad (1)$$

Rumus 1 *Precision*

2.5.2. *Recall*

Parameter penilaian yang didapat dari jumlah data verifikasi memlalui rumus pada persamaan, serta data yang keluar dalam hasil klasifikasi.

$$Recall = \frac{\sum_i^n \frac{TPi}{TPi + FNi}}{n} \quad (2)$$

Rumus 2 *Recall*

2.5.3. *F1-Score*

F1-Score adalah matrix evaluasi yang digunaka untuk mengurangi Tingkat partisipasi angka kerja dari suatu model klasifikasi. [16] .

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

Rumus 3 *F1- Score*

2.5.4. *Accuracy*

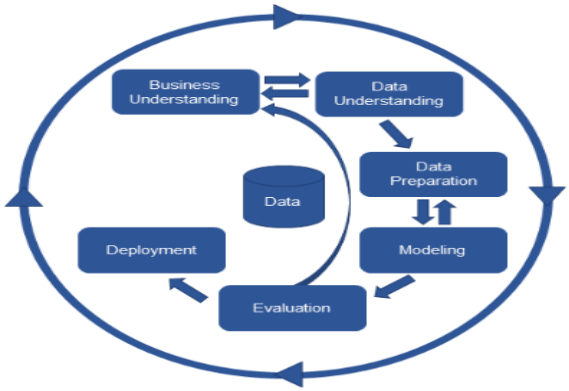
Parameter penilaian untuk menyesuaikan akurasi model dan Tingkat kepercayaan saat melakukan klasfikasi.

$$Accuracy = \frac{\sum_i^n \frac{TPi + TNi}{TPi + FPi + TNi + FNi}}{n} \quad (4)$$

Rumus 4 *Accuracy*

2.6. Crisp Dm

Penelitian ini menggunakan metode *Naïve Bayes Classifier* untuk klasifikasi. opini masyarakat dalam text mining, Bersama dengan model *Cross-Industry Standard Process for Data Mining* (CRISP-DM) yang disesuaikan dengan kebutuhan penelitian [17].



Gambar 1. Tahapan metode crisp-dm

2.7. *Bussines Understanding*

Pemahaman bisnis merupakan kajian tentang topik penelitian yang sedang dilakukan saat ini, penelitian tentang topik ini sedang dilakukan dengan pemanfaatan informasi dari twitter. [17].

2.8. *Data Understanding*

Persiapan data adalah proses menyiapkan data agar menjadi akurat, bersih, dan siap digunakan dalam penelitian.[17]

2.9. *Data Prepation*

Data Preparation adalah proses menyiapkan data untuk mendapatkan informasi yang akurat, menyeluruh, dan dapat diandalkan yang dapat digunakan [17]

2.10. *Modeling*

Pemilihan teknik pemodelan adalah langkah pertama yang akan diambil dan diikuti untuk pembuatan skenario pengujian dan memvalidasi kualitas sebuah model. Berdasarkan hal tersebut, metode yang diusulkan pada penelitian ini akan menggunakan pendekatan metode klasifikasi dengan algoritma *Naïve Bayes* [17].

2.11. *Evaluasi*

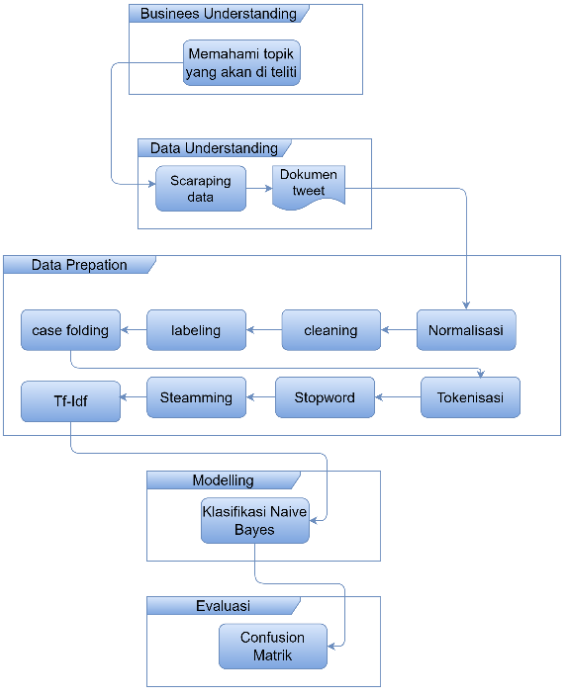
Penelitian ini akan menggunakan teknik *Cross Validation* dengan menggunakan *Naïve Bayes Classifier* untuk melakukan pengujian model [17].

2.12. *Devloyment*

Tahapan pengujian melalui eksperimen dan evaluasi dari model yang akan digunakan, dengan menghasilkan data yang dapat digunakan sebagai masukan dalam pembuatan model implementasi [17].

3. METODE PENELITIAN

The Cross-Industry Standard Process for Data Mining atau CRISP-DM, adalah, adalah model proses yang digunakan secara luas dalam penggalian data dan tidak bergantung pada industri tertentu [18]. Pada penelitian ini penulis menggunakan tahapan-tahapan pada metode Crisp-Dm seperti yang terlihat pada Gambar 2 di bawah ini.



Gambar 2. Kerangka Penelitian.

a. Pada bagian *Business Understanding* Peneliti memahami topik permasalahan untuk di teliti dimana topik yang di teliti merupakan analisis sentiment isu kecurangan pemilu 2024.
b. Data *Understanding* Peneliti mengambil data pada media sosial *Twitter*, Dimana cara mengambilan data dengan cara *scarping data*, kemudian menghasilkan dokumen *Tweet.*
c. Data *Prepation* Peneliti melakukan *Text Preprocessing* dimana tahapan ini untuk memebersihkan data dengan cara normalisasi,*Cleaning,labelling,Case folding,Tokenisasi,Stopword removeval.Steaming* kemudian melakukan pembobotan kata menggunakan tf-idf.
d. Tahapan modelling menggunakan *Naïve Bayes Classifier*.
e. Peneliti memberikan hasil evaluasi pada penilitian ini dengan menggunakan *Confusion Matrix.*

3.1. *Bussines Understanding*

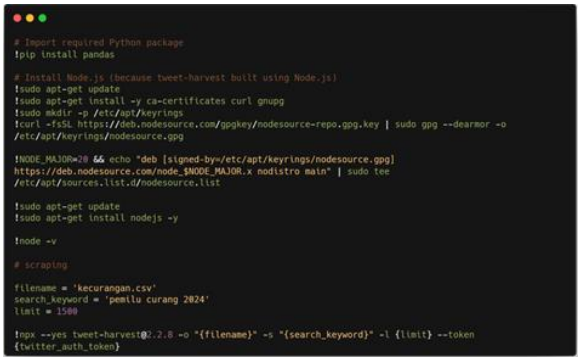
Pemahaman bisnis merupakan kajian tentang topik penelitian yang sedang dilakukan. Pada saat ini, Memahami subjek penelitian dilakukan dengan memanfaatkan informasi dari Twitter. Dimana topik saat ini mengenai isu kecurangan pemilu 2024 [17].

3.2. Data *Understanding*

Mengumpulkan data dari berbagai sumber, memeriksa dan mengkarakterisasi data, dan menganalisisnya kualitasnya adalah tugas-tugas penting dalam fase ini[18]. Data dikumpulkan dari *Twitter*, berfokus pada opini publik mengenai kecurangan pemilu 2024. *Scraping* dilakukan menggunakan VSCode, Python, dan Node.js dengan kata kunci "Pemilu curang 2024".

3.3. *Scraping* Data

Penelitian ini mengumpulkan data dari Twitter dengan menggunakan kata kunci. "pemilu curang 2024" menggunakan *Python*, VSCode, dan Node.js. Data dikumpulkan pada dua periode: pertama, dari 19 Februari hingga 9 Maret 2024 sebanyak 950 *tweet*; kedua, dari 15 Maret hingga 3 Juni 2024 sebanyak 907 *tweet*. Totalnya ada 1857 *tweet* yang siap dianalisis untuk memahami opini publik terkait kecurangan pemilu 2024. Hasilnya, *tweet* yang membahas topik ini berhasil dikumpulkan untuk analisis lebih lanjut.



Gambar 3. Proses *scraping* data

Gambar 3 adalah kode program untuk scraping data dengan menggunakan python dan Nodejs.

3.4. Data *Prepation*

Seleksi data perlu dilakukan dengan menetapkan kriteria inklusi dan eksklusi. Data yang berkualitas buruk bisa diperbaiki melalui proses pembersihan data [18]. Proses ini mencakup langkah-langkah seperti pemilihan atribut, normalisasi, pembersihan data, *Case Folding*, *Tokenisasi*, penghapusan kata umum, dan *Stemming*. Setelah preprocessing teks, dilakukan pembobotan menggunakn Istilah *frekuensi-invers* frekuensi dokumen, atau TF-IDF, digunakan. untuk mengidentifikasi Tingkat kepentingan setiap istilah dalam sebuah *tweet*.

3.5. *Modelling*

Pemodelan data melibatkan pemilihan teknik pemodelan, membangun kasus uji, dan membuat model. menjelaskan alasan pemilihannya. Parameter spesifik harus ditetapkan untuk membangun mode.[18]. Dalam pemodelan dengan *Naïve Bayes* Classifier, kumpulan fitur dan label diolah dengan penentuan parameter seperti smoothing dan distribusi prior untuk mengoptimalkan kinerja model.

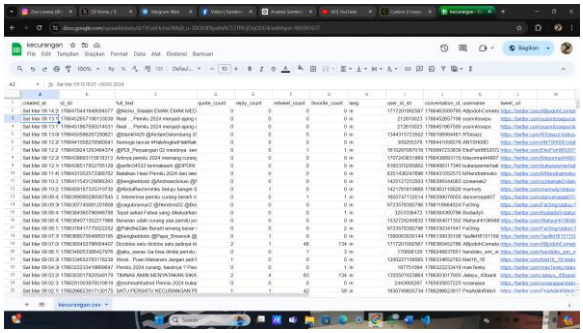
3.6. Evaluasi

Pada tahap evaluasi, hasil dievaluasi berdasarkan tujuan penelitian yang telah ditetapkan. Oleh karena itu, hasil tersebut harus ditafsirkan dan tindakan selanjutnya perlu ditentukan[18]. Tujuan untuk membandingkan nilai aktual dengan hasil prediksi model yang telah dikembangkan. *Matrix* seperti akurasi, presisi,recall dan *F1-Score* dapat digunakan untuk menentukan partisipasi kerja dalam model. model. *True Positive* merujuk pada data yang diprediksi positif dan benar.

4. HASIL DAN PEMBAHASAN

4.1. Hasil *Scraping* Data

Setelah mendapatkan data dari hasil *Scraping* maka data tersebut tersimpan di file berupa *csv*. Tahap selanjutnya akan di lakukan *Text Preprocessing*. *Dapat di lihat* pada gambar 3 di bawah ini.

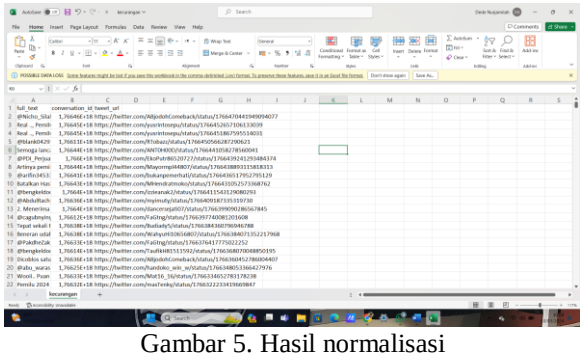


Gambar 4. Dokumen Hasil *Scraping*

Pada gambar 4 Merupakan hasil dari *scraping* data dengan total data keseluruhan 1857 data *tweet*.

4.2. Hasil *Preprocesing Text*

Text preprocessing adalah tahap data preparation dalam metode CRISP-DM untuk mempersiapkan dokumen *Tweet* yang belum terstruktur agar siap untuk analisis teks. Proses ini mencakup pemilihan atribut, normalisasi, *cleansing*, *labeling*, *case folding*, *tokenisasi*, *stopword removal*, dan *stemming*. Data hasil *scraping* akan diolah dengan menghapus atribut yang tidak digunakan, seperti *created_at*, *id_str*, *quote_count*, dan *reply_count*, dan hanya mengambil atribut *text* dari *tweet* pengguna Twitter.



Gambar 5. Hasil normalisasi

Gambar 5 Adalah hasil normalisasi dimana proses ini menghapus kolom yang tidak terpakai. Tahapan selanjutnya adalah *Cleaning* merupan tahapan untuk membersihkan atau menghilangkan komponen-komponen atau filter yang terdapat.Pada hasil *Cleansing* dapat dilihat pada Tabel 2 di bawah ini.

Tabel 2. Proses *Cleansing*

No	Sebelum	Sesudah
1	@abu_waras Ga bisa dinilai pemilu 2024....brutal, curang, culas dan ga ada malu. Mau dinilai minus juga ga pantas	Ga bisa dinilai pemilu 2024 brutal, curang, culas dan ga ada malu. Mau dinilai minus juga ga pantas

Pada Tabel 2 proses *cleansing* merupakan perbersihan url ,@,#,*,(),/,+, emoji dan lainya.

Case Folding adalah proses untuk mempertahankan penyamaan huruf pada *Tweet* dengan mengurangi semua huruf pada *Tweet* menjadi huruf kecil atau *lowercase.h* Hasil casefolding bisa dilihat pada Tabel 3.

Tabel 3. Proses *case folding*

No	Sebelum	Sesudah
1	Ga bisa dinilai pemilu 2024 brutal, curang, culas dan ga ada malu. Mau dinilai minus juga ga pantas	ga bisa dinilai pemilu 2024 brutal, curang, culas dan ga ada malu. mau dinilai minus juga ga pantas

Tabel 3 merubah data *tweet* menjadi huruf kecil semua atau *lowercase.*

Tokennisasi merupakan tahapan proses untuk memisahkan data teks menjadi token. Tahapan ini di lakukan untuk memecah karakter dalam suatu teks menjadi suatu kata. Berikut contoh dari hasil *Tokenisasi* sebagai berikut. bisa di lihat pada tabel 3.

Tabel 4. Proses tokenisasi

ga	bisa	dinilai	Pemilu	2024
brutal	curang	culas	dan	ga
ada	malu	mau	dinilai	minus
juga	ga	pantas		

Pada Tabel 4 proses tokenisasi membuat kalimat data menjadi perkata.

Stopword Removal merupalam tahapan mengahpus kata-kata yang sering muncul dan kata-kata yang tidak ada hubunganya dengan analisi sentimen ini. Pada kata yang tidak di pakai seperti ‘saya’,‘yang’,’ini’ dan masih banyak yang lainya.Hasil *Stopword* bisa dilihat pada Tabel 4.

Tabel 5. Proses *stopword removal*

dinilai	pemilu	2024	brutal	curang
culas	malu	mau	dinilai	minus
Pantas				

Tabel 4 Proses pembersihan kata yang tidak memiliki arti.

Tahap berikutnya adalah *Stemming*, data yang seletah diproses sebelumnya akan diubah dari kata berimbuhan menjadi kata yang tidak memiliki arrti. Proses *Stemming* ini menggunakan *library Sastrawi* untuk melanjutkan ke tahap berikutnya. Hasil dari steaming bisa di lihat pada Tabel 6.

Tabel 6. Hasil proses *stemming*

Sebelum	Sesudah
dinilai pemilu 2024 brutal curang culas malu mau dinilai minus pantas	nilai pemilu 2024 Brutal curang culai malum au nilai minus pantas

Tabel 6 merupakan merubah kata menjadi kata dasar seperti contoh pada tabel di atas.

4.3. TF-IDF

Setelah tahap *text preprocessing*, dilanjutkan dengan TF-IDF untuk memberikan bobot untuk setiap periode dalam *tweet*, mencerminkan pentingnya term tersebut. Bobot dihitung untuk setiap *term* di setiap dokumen. Contoh sampel data yang digunakan sebagai data training mencakup 1 sentimen positif, 1 sentimen negatif, dan 1 sentimen netral, seperti terlihat pada tabel 7.

Tabel 7. Sampel dokumen

No	Dokumen	Label
1	nampaknya pemilu 2024 juga menunjukkan ciri yg sama, hanya legitimasi kekuasaan orang curang.	Negatif
2	ya alloh laknatlah siapapun orang yang berbuat curang di pemilu 2024 ini *term*asuk buzzer bayarannya. aaminn.	Netral
3	daripada cuap curang pemilu 2024 tapi pas di tawarin mentri langsung nyungsep	Positif

Setelah memperoleh sampel data. Kemudian selanjutnya menghitung nilai bobot kata atau *term*, angka 1 untuk kata yang muncul dan angka 0 untuk kata yang tidak muncul pada setiap dokumen, seperti yang ditampilkan pada tabel 8.

Tabel 8. Nilai *term frequency*

Dokumen	TF		
	D1	D2	D3
nilai	1	0	0
pemilu	1	1	1
laknat	0	0	1
curang	1	1	1
malu	1	0	0
minus	1	0	0
tunjuk	0	1	0
legitimasi	0	1	0
kuasa	0	1	0
brutal	1	0	0
buat	0	0	1
buzzer	0	1	0

Setelah memperoleh frekuensi kata, langkah selanjutnya adalah menghitung frekuensi dokumen berdasarkan jumlah kata setiap dokumen. Kemudian,

total semua dokumen dibagi dengan frekuensi dokumen dengan mendapatkan hasil nilai D/df, seperti pada tabel 9.

Tabel 9. Hasil *inverse document frequency*

Dokumen	TF			DF	D/DF	IDF Log (D/DF	1+DF
	D1	D2	D3				
nilai	1	0	0	1	3	0,4771	1,4771
pemilu	1	1	1	3	1	0,0000	1,0000
laknat	0	0	1	1	3	0,4771	1,4771
curang	1	1	1	3	1	0,0000	1,0000
malu	1	0	0	1	3	0,4771	1,4771
minus	1	0	0	1	3	0,4771	1,4771
tunjuk	0	1	0	1	3	0,4771	1,4771
legitimasi	0	1	0	1	3	0,4771	1,4771
kuasa	0	1	0	1	3	0,4771	1,4771
brutal	1	0	0	1	3	0,4771	1,4771
buat	0	0	1	1	3	0,4771	1,4771
buzzer	0	1	0	1	3	0,4771	1,4771
nilai	1	0	0	1	3	0,4771	1,4771
pemilu	1	1	1	3	1	0,0000	1,0000

Keterangan: **df:** Banyaknya kemunculan kata di masing-masing dokumen. **D/df:** Total jumlah dokumen dibagi dengan banyaknya kemunculan kata tersebut. **Log idf:** Hasil logaritma dari D/df. **1+df:** Log idf ditambah 1, di mana angka 1 berperan sebagai nilai *inverse*.

Tabel 10. Hasil perhitungan tf-idf

Dokumen	W=TF(IDF+1)		
	D1	D2	D3
nilai	1,4771	0	0
pemilu	1,0000	1,0000	1,0000
laknat	0	0	1,4771
curang	1,0000	1,0000	1,0000
malu	1,4771	0	0
minus	1,4771	0	0
tunjuk	0	1,4771	0
legitimasi	0	1,4771	0
kuasa	0	1,4771	0
brutal	1,4771	0	0
buat	0	0	1,4771
buzzer	0	1,4771	0
J Rangking	7,91	8	5

Pada tabel diatas merupakan *Term Frequency-Inverse Document Frequency* (TF-IDF) digunakan untuk mengektrak nilai tf-idf dengan menganalisis setiap kata dari semua dokumen. Perhitungan W=TF(IDF+1) dilakukan dengan mengalikan frekuensi kemunculan kata dengan nilai 1+df.

4.4. Hasil Klasifikasi *Naïve Bayes*

Setelah proses tf-idf selesai, data siap diolah untuk pengujian model *Naïve Bayes* di *Python*. Data ini kemudian dibagi menjadi dua bagian 1629 data untuk pelatihan (*training*) dan 181 data untuk pengujian (testing). Dengan total 1810 data, pembagian ini dilakukan dengan proporsi 90% untuk pelatihan dan 10% untuk pengujian. untuk pengujian model dapat dilihat pada gambar 6.

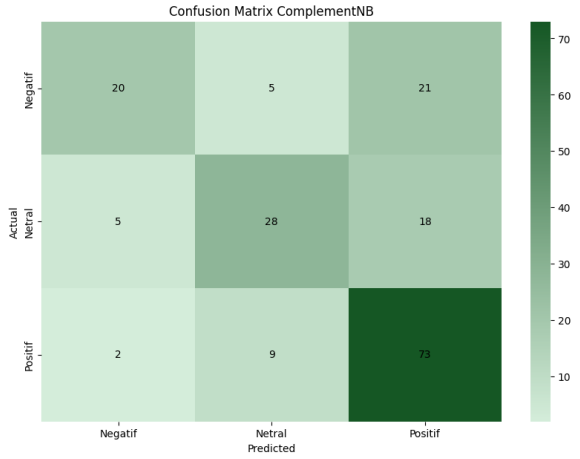
```
# Inisialisasi model MultinomialNB
MNB = MultinomialNB()
MNB.fit(x_train_tfidf, y_train)# Tidak perlu menggunakan .toarray() untuk
MultinomialNB
# Prediksi
predicted = MNB.predict(x_test_dtm)
accuracy = accuracy_score(predicted, y_test)

# Print akurasi
print('MultinomialNB model accuracy is', str('{:04.2f}'.format(accuracy * 100)) + '%')
print('-------------------------------------------------')

Output:

ComplementNB model accuracy is 66.85
-------------------------------------------------
```

Gambar 6. Pengujian model

Gambar 6 Memproses data *training* dengan hasil akurasi sebesar 66.85%.

4.5. Hasil *Confusion Matrix*

Tahap berikutnya adalah evaluasi menggunakan *Confusion Matrix* untuk membandingkan prediksi model dengan nilai aktual. *Confusion Matrix* membantu menghitung metrik seperti akurasi, presisi, *Recall*, dan *F1-Score* untuk menilai kinerja model. Misalnya, *True Positive* menunjukkan prediksi positif yang benar, sementara metrik lainnya memberikan informasi tambahan tentang keakuratan dan konsistensi model.



Gambar 7. Hasil *confusion matrix*

Setelah mendapatkan nilai untuk *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN), Langkah selanjutnya adalah melakukan percobaan di *Python* menggunakan rumus yang telah ditentukan sebelumnya. Kemudia didapatkan hasil dari *confusion matrix* sebagai berikut.

1) Akurasi

$$\text{Akurasi} = \frac{\text{TP} + \text{TNe} + \text{TNt}}{\text{Total}} = \frac{20 + 29 + 73}{181}$$

$$= \frac{121}{181} = 0{,}6685$$

2) *Preccision*

$$\text{Precission Positif} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{73}{73 + 17 + 22}$$

$$= \frac{73}{112} = 0{,}6517$$

$$\text{Precission Negatif} = \frac{\text{TNe}}{\text{TNe} + \text{FNe}} = \frac{20}{20+5+2} = \frac{20}{27} = 0{,}7407$$

$$\text{Precission Netral} = \frac{\text{TNt}}{\text{Tnt} + \text{FNt}} = \frac{28}{28+4+9} = \frac{28}{42} = 0{,}6666$$

Total = 0,6517 + 0,7407 + 0,6666 x 100 = 68,05%

3) *Recall*

$$\text{Recall Positif} = \frac{\text{TP}}{\text{TP} + \text{FNe} + \text{FNt}} = \frac{73}{73+9+3} = \frac{73}{84} = 0{,}8690$$

$$\text{Recall Negatif} = \frac{\text{TNe}}{\text{TNe} + \text{FNt} + \text{FP}} = \frac{20}{20+4+21} = \frac{20}{45} = 0{,}4444$$

$$\text{Recall Netral} = \frac{\text{TNt}}{\text{TNt} + \text{FNe} + \text{FP}} = \frac{28}{28+5+17} = \frac{28}{51} = 0{,}5490$$

Total = 0,8690 + 0,4444 + 0,5490 x 100 = 56,11%

4) *F1-Score*

$$\text{F1(Score)} = 2 \times \frac{\text{Precision Positif x Recall positif}}{\text{precision positif} + \text{Recall Positif}} = 2 \times \frac{0{,}6517 \times 0{,}8690}{0{,}6517 + 0{,}8690} = 0{,}7448$$

$$\text{F1} - \text{(Score)} = 2 \times \frac{\text{Precision Negatif x Recall Negatif}}{\text{precision Negatif} + \text{Recall Negatif}} = 2 \times \frac{0{,}7472 \times 0{,}4444}{0{,}7472 + 0{,}4444} = 0{,}5573$$

$$\text{F1} - \text{(Score)} = 2 \times \frac{\text{Precision Netral x Recall Netral}}{\text{precision Netral} + \text{Recall Netral}} = 2 \times \frac{0{,}6666 \times 0{,}5490}{0{,}6666 + 0{,}5490} = 0{,}6026$$

Total = 0,7448 + 0,5573 + 0,6062 x 100 = 61,56%

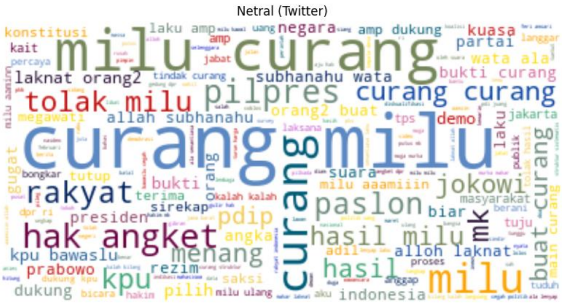
Dari hasil perhitungan evaluasi memahami kinerja model klasifikasi secara lebih efektif dengan menggunakan *confusion matrix* Jika nilai akurasi tinggi tetapi nilai recall rendah, mungkin ada kasus positif yang tidak terdeteksi oleh model. Sebaliknya, nilai presisi yang rendah bisa menunjukkan banyak prediksi positif yang salah.

4.6. Hasil *Wordcloud*

Langkah terakhir adalah memvisualisasikan data menggunakan *Wordcloud* untuk mengidentifikasi kata-kata yang sering muncul dalam setiap kalimat.Di bawah ini adalah hasil *Wordcloud* untuk sentimen positif, seperti yang terlihat pada gambar 8.

Gambar 8. Hasil *wordcloud* positif

Pada gambar 8, kata-kata yang sering muncul pada sentimen positif yaitu: “curang”, “pilpres”, “hak angkket”, “jokiwi”, dan lainnya. Sedangkang *wordcloud* sentimen negatif dapat dilihat pada gambar 9.



Gambar 9. Hasil *wordcloud* negatif

Pada gambar 9 dilihat kata yang sering muncul pada sentimen negatif yaitu: “curang”, “tolak”, “hasil pemilu”, “brutal”, dan lain lain. Untuk *word cloud* sentimen netral dapat dilihat pada gambar 10.

Gambar 10. Hasil *wordcloud* netral

Pada gambar 10 dilihat kata yang sering muncul pada sentimen netral yaitu: “curang”, “pilpres”, “tolak pemilu”, “paslon”,”rakyat”,”alloh laknat” dan seterusnya.

5. KESIMPULAN DAN SARAN

Penelitian ini menganalisis sentimen di Twitter mengenai opini tentang isu kecurangan pemilu 2024 Menggunakan algoritma *Naïve Bayes* untuk klasifikasi sentimen. Hasilnya menunjukkan bahwa 45.4% dari total *tweet* mengandung sentimen positif, 23.3% mengandung sentimen negatif, dan 31.3% mengandung sentimen netral. Evaluasi model menggunakan *Confusion Matrix* menunjukkan akurasi

66.85%, presisi 68.05%, *recall* 56.11%, dan *F1-Score* 61.56%. Model dapat memprediksi dengan benar sekitar 66.85% dari total *tweet* yang dianalisis, namun nilai recall yang lebih rendah dibandingkan dengan presisi menunjukkan ada sejumlah *tweet* positif yang tidak terdeteksi oleh model. Penelitian selanjutnya disarankan penggunaan *dataset* yang lebih banyak dan mencoba algoritma lain seperti *K-Nearest Neighbors, Support Vector Machine, Random Forest, dan Regresi Logistik* untuk meningkatkan kinerja model.

DAFTAR PUSTAKA

[1] I. Kurniawan and A. Susanto, "Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019," *Eksplora Inform.*, vol. 9, no. 1, pp. 1–10, Sep. 2019, doi: 10.30864/eksplora.v9i1.237.

[2] P. Mimboro, "Sentimen Twitter terhadap PILKADA kota Medan menggunakan metode Naïve Bayes," *https://lenteradua.net/jurnal/index.php/jnanaloka/article/view/42*, 2022, doi: 10.36802/jnanaloka.v3-no1-27-32.

[3] Mudiyati Rahmatunnisa, "Menyoal Praktek Vote Buying Dan Implikasinya Terhadap Integritas Pemilu," *J. Keadilan Pemilu*, vol. 2, pp. 35–49, 2022, [Online]. Available: http://journal.bawaslu.go.id/index.php/JKP/article/download/170/124

[4] M. Nurkamiden, "SiRekap : Tantangan dan Potensi Kekeliruan Proses Rekapitulasi Pemilu Serentak di Indonesia SiRekap: Challenges and Potential Errors in the Recapitulation Process of Simultaneous Elections in Indonesia," *Sosiol. J. Penelit. dan Pengabdi. Kpd. Masy.*, vol. 1, no. 2, pp. 101–110, 2024, [Online]. Available: http://ejurnal.fis.ung.ac.id/index.php/sjppm/about%0ASiRekap

[5] Salsabilla Na, "Analisis sentimen pada media sosial twitter terhadap tokoh gus dur menggunakan metode naïve bayes dan support vector machine (svm)," Jakarta, 2022. Accessed: May 16, 2024. [Online]. Available: https://repository.uinjkt.ac.id/

[6] A. Perdana, A. Hermawan, and D. Avianto, "Analisis Sentimen Terhadap Isu Penundaan Pemilu di Twitter Menggunakan Naive Bayes Clasifier," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 11, no. 2, pp. 195–200, Jul. 2022, doi: 10.32736/sisfokom.v11i2.1412.

[7] S. Puad and A. Susilo Yuda Irawan, "ANALISIS SENTIMEN MASYARAKAT PADA TWITTER TERHADAP PEMILIHAN UMUM 2024 MENGGUNAKAN ALGORITMA NAÏVE BAYES," Malang, 2023. [Online]. Available: https://ejournal.itn.ac.id/index.php/jati/article/view/6920

[8] C. B. Saputra, A. Muzakir, and D. Udariansyah, "ANALISIS SENTIMEN MASYARAKAT TERHADAP #2019GANTIPRESIDEN BERDASARKAN OPINI DARI TWITTER MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER," *Bina Darma Conf. Comput. Sci.*, 2020, [Online]. Available: https://conference.binadarma.ac.id/index.php/BDCCS/article/view/155

[9] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional," *J. TEKNO KOMPAK*, vol. 15, no. 1, 2020, Accessed: Jul. 13, 2024. [Online]. Available: https://ejurnal.teknokrat.ac.id/

[10] S. Safuan and B. Budiandru, "Modus Kecurangan & Program Anti Kecurangan di Pelabuhan (Studi Kasus Pelabuhan di Jakarta)," *Owner*, vol. 3, no. 2, p. 54, Jul. 2019, doi: 10.33395/owner.v3i2.131.

[11] M. Dinda, M. Azzahra, T. Hafid, and S. Alam, "ANALISIS SENTIMEN ULASAN PRODUK SERUM WAJAH PADA BEAUTY BRAND SOMETHINC MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER," Malang, 2023. [Online]. Available: https://reviews.femaledaily.com/products/treatment/s

[12] G. Darmawan, S. Alam, and M. Imam Sulistyo, "ANALISIS SENTIMEN BERDASARKAN ULASAN PENGGUNA APLIKASI MYPERTAMINA PADA GOOGLE PLAYSTORE MENGGUNAKAN METODE NAÏVE BAYES," *literasisains.id*, vol. 2, no. 3, pp. 100–108, 2023, doi: 10.55123.

[13] S. Ayudya, A. Armand, M. Hafid, and M. R. Muttaqin, "ANALISIS SENTIMEN SISTEM E-TILANG PADA PLATFORM TWITTER MENGGUNAKAN METODE NAIVE BAYES," 2023.

[14] A. N. S. Rahayu, T. I. Hermanto, and I. M. Nugroho, "SENTIMENT ANALYSIS USING K-NEAREST NEIGHBOR BASED ON PARTICLE SWARM OPTIMIZATION ACCORDING TO SUNSCREEN'S REVIEWS," *J. Tek. Inform.*, vol. 3, no. 6, pp. 1639–1646, Dec. 2022, doi: 10.20884/1.jutif.2022.3.6.425.

[15] D. N. Fitriana and Y. Sibaroni, "ARJUNA) Managed by Ministry of Research, Technology, and Higher Education," *Accredit. by Natl. J. Accredit.*, vol. 4, no. 2, pp. 846–853, 2020, Accessed: Jun. 22, 2024. [Online]. Available: http://jurnal.iaii.or.id

[16] R. Merdiansah and A. Ali Ridha, "Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT," *J. Ilmu Komput. dan Sist. Inf. (JIKOMSI*, vol. 7, no. 1, pp. 221–228, 2024, [Online]. Available: https://ejournal.sisfokomtek.org/index.php/jikom/article/view/2895

[17] L. Hermawati, V. Berland, A. Rahmadiah, E. Hutabarat, and D. Dwi Saputra, "Jurnal Sistim Informasi dan Teknologi https://jsisfotek.org/index.php Komparasi Metode Text Mining Terhadap Masalah Pengklasifikasian Narasi Informative & Non Informative Pada twitter @PLN_123," vol. 5, no. 1, 2023, doi: 10.37034/jsisfotek.v4i2.191.

[18] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," *Procedia Comput. Sci.*, vol. 181, no. 2019, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.