

Analisis Sentimen Pelantikan Presiden Indonesia 2024 Menggunakan Model Klasifikasi dan Algoritma *Naive Bayes*

Andika Prayoga Siswono*¹, Syahrul Fauzi², Aliefino Zalva Surya Hermawan³, Arif Riyandi⁴

Program Studi Sistem Informasi, Fakultas Rekayasa Industri, Telkom University
Jl. DI Panjaitan No128, Karangreja, Purwokerto Kidul, Kec. Purwokerto Selatan, Kab. Banyumas, Jawa Tengah, Indonesia 35147

¹ andikaps@student.telkomuniversity.ac.id

² syahrulfauzi@student.telkomuniversity.ac.id

³ aliefinozalvash@student.telkomuniversity.ac.id

⁴ arifriyandi@telkomuniversity.ac.id

Dikirim pada 16-11-2024, Direvisi pada 22-11-2024, Diterima pada 29-11-2024

Abstrak

Indonesia merupakan negara yang berbentuk republik yang dipimpin oleh seorang presiden. Indonesia telah menyelesaikan pemilihan presiden untuk periode 2024 sampai 2029. Pelantikan presiden disiarkan melalui media nasional seperti TV, Radio, *YouTube*, dan media lainnya. Berdasarkan pelantikan tersebut mengundang berbagai respon dari masyarakat, baik berupa ucapan selamat, harapan, maupun kritik. Namun, belum diketahui bagaimana sentimen masyarakat terhadap pelantikan tersebut, apakah didominasi oleh sentimen positif, negatif, atau netral. Penelitian ini bertujuan untuk menganalisis sentimen publik terkait pelantikan Presiden dan Wakil Presiden Indonesia periode 2024-2029 pada saluran *YouTube* milik Sekretariat Kepresidenan menggunakan algoritma *Naive Bayes* untuk mengelompokkan komentar ke dalam tiga kategori yaitu positif, negatif, netral. Data dikumpulkan melalui proses *crawling* di *Google Colab*, Menghasilkan 1.371 data komentar dari video pelantikan. Data tersebut diolah melalui tahapan *preprocessing*, yang termasuk *cleansing*, *case folding*, *tokenize*, *stopword removal*, *stemming* dan TF-IDF. Ketidakseimbangan data, diatasi menggunakan metode SMOTE. Hasil klasifikasi menunjukkan mayoritas komentar positif dan netral, mencerminkan penerimaan publik yang baik terhadap kepemimpinan baru. Model *Naive Bayes* ini menghasilkan akurasi sebesar 92% setelah data diseimbangkan akurasi ini mengalami peningkatan, sebelum diseimbangkan akurasi sebesar 86%, hal ini menunjukkan keefektifan dan ketepatan model dalam mengklasifikasikan sentimen. Hasil ini menegaskan bahwa metode *Naive Bayes* dapat diandalkan untuk analisis sentimen publik.

Kata Kunci: Analisis sentimen, pelantikan presiden, *Naive Bayes*

Ini adalah artikel akses terbuka di bawah lisensi [CC BY-SA](#).



Penulis Koresponden:

Andika Prayoga Siswono

Telkom University

Jl. DI Panjaitan No128, Karangreja, Purwokerto Kidul, Kec. Purwokerto Selatan, Kab. Banyumas, Jawa Tengah, Indonesia 35147

Email: andikaps@student.telkomuniversity.ac.id

I. PENDAHULUAN

Indonesia merupakan negara berbentuk republik yang dipimpin oleh seorang presiden. Sejak Indonesia berdiri dari tahun 1945 sampai sekarang, Indonesia telah dipimpin oleh delapan orang presiden. Presiden di Indonesia memiliki masa jabatan lima tahun dan dipilih oleh masyarakat melalui proses pemilihan presiden atau yang sering disebut pilpres[1]. Indonesia telah menyelesaikan proses pemilihan presiden untuk periode 2024 sampai 2029 yang dilakukan pada tanggal 14 Februari 2024. Presiden terpilih yaitu Prabowo Subianto dan wakil presiden terpilih Gibran Rakabuming Raka dilantik menjadi presiden dan

wakil presiden pada 20 Oktober 2024 yang disiarkan melalui berbagai media nasional seperti TV, Radio, *YouTube* dan berbagai media lainnya.

Berdasarkan dari pelantikan Presiden dan Wakil Presiden yang disiarkan melalui saluran *YouTube* sekretariat kepresidenan, mengundang berbagai respons dari masyarakat. baik berupa ucapan selamat, harapan, maupun kritik terhadap kebijakan dan visi, misi yang akan diterapkan. Respon dari masyarakat ini mencerminkan pandangan dan harapan terhadap masa depan Indonesia di bawah kepemimpinan yang baru. Penelitian ini juga didasari dari respon masyarakat tersebut yang bertebaran dalam kolom komentar selama proses pelantikan Presiden dan Wakil Presiden yang baru, pada saluran *YouTube* milik sekretariat kepresidenan.

YouTube merupakan salah satu platform video *streaming* dengan pengguna terbanyak di Indonesia, dilansir dari situs datareportal dengan judul artikel Digital 2024: Indonesia, jumlah pengguna *YouTube* di Indonesia mencapai 139 juta pengguna di awal tahun 2024 yang mencakup 75 persen pengguna internet Indonesia[2]. *YouTube* menyediakan fitur bagi pengguna untuk merespon video dengan menulis opini di kolom komentar. Kolom komentar ini dapat dimanfaatkan sebagai sumber data dalam analisis sentimen yang memungkinkan pengelompokan pendapat menjadi opini positif, negatif, atau netral[3].

Analisis sentimen adalah metode untuk memperoleh informasi mengenai pandangan seseorang terhadap suatu peristiwa atau isu melalui data teks. Metode ini membantu dalam memahami sikap umum publik dalam menanggapi sebuah kejadian dari sudut pandang positif, negatif, maupun netral [4]. Data dari penelitian diperoleh dari proses *crawling* pada kolom komentar *YouTube* yang berhubungan dengan pelantikan presiden 2024.

Penelitian serupa juga dilakukan oleh Ilham Antasa Achmadi dkk. Fokus penelitian ini untuk menganalisis sentimen masyarakat indonesia terkait hasil pemilihan presiden 2024 melalui media sosial twitter atau sekarang lebih dikenal dengan X, memanfaatkan algoritma *Naive Bayes*. Hasil penelitian menunjukkan akurasi sebesar 57% untuk data satu bulan dan 63% untuk data lima bulan dengan mayoritas hasil positif di kedua data tersebut di angka 66% untuk data satu bulan dan 67% untuk data lima bulan[5].

Penelitian terkait dengan penggunaan metode *Naive Bayes* juga dilakukan oleh Kurnia Ardiansyah Lubis dkk. Penelitian ini menganalisis sentimen masyarakat indonesia terkait dengan pemindahan ibu kota ke kalimantan melalui platform twitter. Data yang diambil pada twitter menghasilkan 4.533 data mentah yang kemudian dilakukan proses *preprocessing* menjadi 1.449 data. Hasil dari penelitian ini memiliki sentimen positif sebesar 74% dan negatif sebesar 26% memiliki akurasi sebesar 76,30% pada perbandingan data latih 83% dan 17% untuk data uji[6].

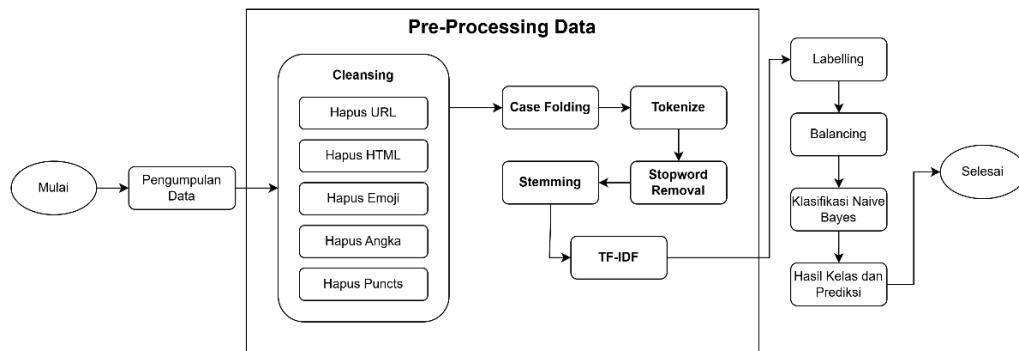
Penelitian yang dilakukan oleh Fahri Setiawan dan Pipin Farida Ariyani, memiliki tujuan untuk menganalisis sentimen publik terhadap putusan Mahkamah Konstitusi (MK) terkait sengketa pilpres 2024 melalui komentar *YouTube* pada saluran KompasTV dan CNN Indonesia dengan menggunakan metode *Naive Bayes*. Hasil penelitian ini menunjukkan mayoritas komentar positif terhadap putusan MK sebesar 63,8%, sentimen negatif 22,8% dan netral 13,3% [7].

Penelitian ini bertujuan untuk menganalisis sentimen publik mengenai pelantikan Presiden dan Wakil Presiden indonesia periode 2024-2029, melalui komentar pada saluran *YouTube* milik Sekretariat Kepresidenan. Analisis sentimen pada penelitian ini akan menggunakan metode *Naive Bayes* yang berfungsi untuk mengelompokkan komentar tersebut ke dalam tiga kategori yaitu positif, negatif, netral. Metode *Naive Bayes* dipilih karena beberapa keunggulan seperti tidak memerlukan banyak data latih untuk proses klasifikasi[8].

II. METODE PENELITIAN

A. Tahapan Penelitian

Penelitian ini akan dilakukan analisis sentimen dengan menggunakan metode *Naive Bayes*. Data yang telah diperoleh tidak bisa langsung diklasifikasikan. data harus melalui proses *preprocessing*, dilanjutkan dengan *labelling*, *balancing*, klasifikasi dengan *Naive Bayes*, dan diakhiri dengan hasil kelas dan prediksi.



Gambar 1. Tahapan Penelitian

B. Pengumpulan Data

Proses pengumpulan data dilakukan melalui cara *crawling* pada *YouTube* menggunakan *Google Collab* dan *library Google Api* untuk mengakses data komentar. Untuk melakukan *crawling* pada saluran *YouTube* sekretariat kepresidenan dibutuhkan video id agar dapat mengakses video yang ingin diambil datanya. Data komentar yang berhasil dikumpulkan melalui proses *crawling* berjumlah 1.371 komentar. Data yang berhasil dikumpulkan akan melalui tahap *preprocessing* untuk membersihkan data yang tidak diperlukan.

C. Preprocessing

Preprocessing merupakan salah satu tahapan terpenting yang bertujuan membuat data mentah menjadi data yang siap diolah pada tahapan berikutnya. Dalam penelitian ini tahap *preprocessing* memiliki beberapa langkah sebagai berikut :

- a. *Cleansing*
Cleansing merupakan proses untuk menghapus karakter yang tidak diperlukan dari teks. Tujuannya adalah untuk membersihkan noise yang mengganggu dalam data set[9].
- b. *Case Folding*
Case Folding merupakan proses mengganti setiap huruf dalam teks menjadi huruf kecil[9].
- c. *Tokenize*
Tokenize berfungsi untuk memecah sebuah kalimat menjadi kata-kata terpisah, sehingga kata-kata tersebut dapat dianalisis lebih lanjut[9].
- d. *Stopword Removal*
Stopword Removal adalah proses menghapus kata yang kurang bermakna berdasarkan daftar *stopword* yang telah ditentukan[10].
- e. *Stemming*
Stemming adalah proses untuk menghilangkan imbuhan di awal dan di akhir kata, dengan memperhatikan prefiks dan sufiks umum pada kata dasar[10].
- f. *Term Frequency – Inverse Document Frequency (TF-IDF)*
TF-IDF adalah proses pembobotan kata, di mana setiap kata diberi bobot untuk mengoptimalkan kemampuan dalam analisis sentimen[11].

D. Labelling Data

Tahapan selanjut setelah proses *preprocessing* adalah *labelling data*, yaitu proses memberi label atau kategori pada data. Pada penelitian ini data akan dibagi menjadi tiga bagian sentimen seperti positif, negatif, dan netral[12].

E. Balancing Data

Balancing data diperlukan ketika terjadi kasus ketidakseimbangan data, situasi di mana satu kelas memiliki jumlah data yang lebih besar dibandingkan kelas lainnya[13]. Untuk mengatasi ketidakseimbangan data bisa dilakukan penyeimbangan data dengan menerapkan metode *Synthetic Minority Over-Sampling technique (SMOTE)*. SMOTE adalah salah satu teknik yang sering digunakan

untuk memperbaiki ketidakseimbangan data, SMOTE menyintesis sampel baru dari kelas minoritas yang kemudian akan menyeimbangkan dataset, proses ini mencakup pembuatan *instance* baru dari kelas minoritas, dengan tujuan untuk memperbaiki kinerja model[11],[14].

F. Naive Bayes

Algoritma *Naive Bayes* adalah salah satu metode yang sering digunakan dalam data *mining* karena kemudahan dalam penggunaannya, kecepatan pemrosesan yang tinggi, kemudahan implementasi berkat struktur yang sederhana, dan memiliki tingkat akurasi yang lebih baik dibanding dengan metode lainnya. *Naive Bayes* merupakan sebuah teori yang berkaitan dengan konsep probabilitas bersyarat.[15],[16]. Secara umum rumus *Teorama Bayes* adalah sebagai berikut :

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

Rumus di atas menggambarkan konsep bersyarat dalam *teorama Bayes*. A adalah kelas yang sangat spesifik, sedangkan B menunjukkan data yang bervariasi di dalam kelas yang belum diklasifikasikan, serta $P(A|B)$ merujuk pada hasil probabilitas hipotesis berdasarkan kondisi tertentu. Representasi probabilitas awal dari suatu kelas berdasarkan kondisi hipotesis ditulis sebagai $P(A)$ dan $P(B|A)$. $P(B)$ mewakili probabilitas dari A[17].

III. HASIL DAN PEMBAHASAN

A. Pengolahan data

Proses *crawling* data komentar *YouTube* menggunakan *Google Colab* dengan menggunakan bahasa *Phyton*. Data yang dihasilkan dari proses *crawling* sebanyak 1.371 data, berikut adalah gambar data yang telah dihasilkan :

	author	updated_at	like_count	text	public
0	@MortalAja	2024-11-09T08:46:50Z	0	SELAMAT PAK PRABOWO	True
1	@MuhammadRizal-ck4he	2024-11-09T02:50:23Z	0	apa kabar bu Titiek	True
2	@gemanusantarachanel3008	2024-11-08T15:36:44Z	0	Balek kampong lek	True
3	@Akungachagenshin1	2024-11-07T17:30:59Z	0	5:24:50 kasian yg pegang pedang sebelah kiri	True
4	@andreygunawan7469	2024-11-07T12:23:12Z	0	Pak TOLONG naikan HARGA DIRI \nKAWASAKI NINJ...	True
...
1367	@BenHart-w9n	2024-10-20T08:00:11Z	2	Oke	True
1368	@Behemoth666-t1c	2024-10-20T08:00:10Z	2	IDID	True
1369	@rokim92	2024-10-20T08:00:07Z	3	Pindah	True
1370	@AhmadDaffaAthaHidayat-j6k	2024-10-20T07:59:54Z	3	Saya berdoa semoga kota Mekkah dan Madinah bis...	True
1371	@qsd-	2024-10-20T07:59:30Z	1	🔥	True

Gambar 2. Hasil *Crawling*

Dari hasil *crawling* tersebut akan dilanjutkan ke tahapan *preprocessing*. *Preprocessing* pada penelitian ini terdiri dari enam langkah, berikut adalah langkah-langkah dalam tahapan *preprocessing* :

a. *Cleansing*

Cleansing merupakan langkah awal dimana data akan dibersihkan dari atribut yang tidak digunakan selama proses analisis. Contoh atribut yang tidak digunakan seperti angka, *url*, emoji, dan simbol. Atribut tersebut nantinya akan dibersihkan dari dalam data.

	text	cleansing
0	5:24:50 kasian yg pegang pedang sebelah kiri	kasian yg pegang pedang sebelah kiri
1	Pak TOLONG naikan HARGA DIRI \nKAWASAKI NINJ...	Pak TOLONG naikan HARGA DIRI \nKAWASAKI NINJA
3	pak presidenku tolong bikin pabrik buat para p...	pak presidenku tolong bikin pabrik buat para p...
4	Sudah 2 periode kita tidak di pimpin presiden ...	Sudah periode kita tidak di pimpin presiden da...
5	Judul instrumen 5:25:43 apa ya?	Judul instrumen apa ya
...
1364	MAKASIH PAKKK	MAKASIH PAKKK
1365	6 jam nonton pak jokowi dan pak prabowo	jam nonton pak jokowi dan pak prabowo
1366	Oke	Oke
1368	Pindah	Pindah
1369	Saya berdoa semoga kota Mekkah dan Madinah bis...	Saya berdoa semoga kota Mekkah dan Madinah bis...

Gambar 3. Hasil Cleansing

b. *Case Folding*

Tahap ini merupakan tahapan dimana semua huruf akan diubah menjadi huruf kecil semua.

	text	cleansing	case_folding
0	5:24:50 kasian yg pegang pedang sebelah kiri	kasian yg pegang pedang sebelah kiri	kasian yg pegang pedang sebelah kiri
1	Pak TOLONG naikan HARGA DIRI \nKAWASAKI NINJ...	Pak TOLONG naikan HARGA DIRI \nKAWASAKI NINJA	pak tolong naikan harga diri \nkawasaki ninja
3	pak presidenku tolong bikin pabrik buat para p...	pak presidenku tolong bikin pabrik buat para p...	pak presidenku tolong bikin pabrik buat para p...
4	Sudah 2 periode kita tidak di pimpin presiden ...	Sudah periode kita tidak di pimpin presiden da...	sudah periode kita tidak di pimpin presiden da...
5	Judul instrumen 5:25:43 apa ya?	Judul instrumen apa ya	judul instrumen apa ya
...
1364	MAKASIH PAKKK	MAKASIH PAKKK	makasih pakkk
1365	6 jam nonton pak jokowi dan pak prabowo	jam nonton pak jokowi dan pak prabowo	jam nonton pak jokowi dan pak prabowo
1366	Oke	Oke	oke
1368	Pindah	Pindah	pindah
1369	Saya berdoa semoga kota Mekkah dan Madinah bis...	Saya berdoa semoga kota Mekkah dan Madinah bis...	saya berdoa semoga kota mekkah dan madinah bis...

Gambar 4. Hasil Case Folding

c. *Tokenize*

Tokenize adalah tahapan memecah kalimat dalam data menjadi unit-unit kecil seperti per kata yang disebut dengan token.

case_folding	tokenize
kasian yg pegang pedang sebelah kiri	[kasian, yg, pegang, pedang, sebelah, kiri]
pak tolong naikan harga diri \nkawasaki ninja	[pak, tolong, naikan, harga, diri, kawasaki, n...
pak presidenku tolong bikin pabrik buat para p...	[pak, presidenku, tolong, bikin, pabrik, buat, ...
sudah periode kita tidak di pimpin presiden da...	[sudah, periode, kita, tidak, di, pimpin, pres...
judul instrumen apa ya	[judul, instrumen, apa, ya]
...	...
makasih pakkk	[makasih, pakkk]
jam nonton pak jokowi dan pak prabowo	[jam, nonton, pak, jokowi, dan, pak, prabowo]
oke	[oke]
pindah	[pindah]
saya berdoa semoga kota mekkah dan madinah bis...	[saya, berdoa, semoga, kota, mekkah, dan, madi...

Gambar 5. Hasil Tokenize

d. *Stopword Removal*

Tapan ini merupakan proses penghapusan kata-kata yang dianggap tidak penting selama proses analisis.

case_folding	tokenize	stopword
kasian yg pegang pedang sebelah kiri	[kasian, yg, pegang, pedang, sebelah, kiri]	[kasian, pegang, pedang, sebelah, kiri]
pak tolong naikan harga diri \nkawasaki ninja	[pak, tolong, naikan, harga, diri, kawasaki, n...]	[tolong, naikan, harga, kawasaki, ninja]
pak presidenku tolong bikin pabrik buat para p...	[pak, presidenku, tolong, bikin, pabrik, buat,...]	[presidenku, tolong, bikin, pabrik, pengaggur,...]
sudah periode kita tidak di pimpin presiden da...	[sudah, periode, kita, tidak, di, pimpin, pres...]	[periode, pimpin, presiden, kemiliterang, gag...]
judul instrumen apa ya	[judul, instrumen, apa, ya]	[judul, instrumen]
...
makasih pakkk	[makasih, pakkk]	[makasih, pakkk]
jam nonton pak jokowi dan pak prabowo	[jam, nonton, pak, jokowi, dan, pak, prabowo]	[jam, nonton, jokowi, prabowo]
oke	[oke]	[]
pindah	[pindah]	[pindah]
saya berdoa semoga kota mekkah dan madinah bis...	[saya, berdoa, semoga, kota, mekkah, dan, madi...]	[berdoa, semoga, kota, mekkah, madinah, dikuas...]

Gambar 6. Hasil Stopword Removal

e. *Stemming*

Stemming adalah proses mengubah kata menjadi bentuk dasarnya dengan menghapus akhiran atau awalan.

tokenize	stopword	stemming
[kasian, yg, pegang, pedang, sebelah, kiri]	[kasian, pegang, pedang, sebelah, kiri]	[kasi, pegang, pedang, belah, kiri]
[pak, tolong, naikan, harga, diri, kawasaki, n...]	[tolong, naikan, harga, kawasaki, ninja]	[tolong, naik, harga, kawasaki, ninja]
[pak, presidenku, tolong, bikin, pabrik, buat,...]	[presidenku, tolong, bikin, pabrik, pengaggur,...]	[presiden, tolong, bikin, pabrik, pengaggur, n...]
[sudah, periode, kita, tidak, di, pimpin, pres...]	[periode, pimpin, presiden, kemiliterang, gag...]	[periode, pimpin, presiden, kemiliterang, gag...]
[judul, instrumen, apa, ya]	[judul, instrumen]	[judul, instrumen]
...
[makasih, pakkk]	[makasih, pakkk]	[makasih, pakkk]
[jam, nonton, pak, jokowi, dan, pak, prabowo]	[jam, nonton, jokowi, prabowo]	[jam, nonton, jokowi, prabowo]
[oke]	[]	[]
[pindah]	[pindah]	[pindah]
[saya, berdoa, semoga, kota, mekkah, dan, madi...]	[berdoa, semoga, kota, mekkah, madinah, dikuas...]	[doa, moga, kota, mek, madinah, asai, israel, ...]

Gambar 7. Hasil Stemming

f. TF-IDF

TF-IDF proses memberi bobot pada kata berdasarkan frekuensi kemunculannya dalam data set. Berikut adalah hasil dari proses TF-IDF

term	idf	tf0	tf1	tf2	tf3	tf4	tf5	tf6	tf7	...	tf1263	tf1264	tf1265	tf1266	tf1267	tf1268	tf1269	tf1270	tf1271	tf1272	
0	aaamiin	6.455984	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0000
1	aahpokoknya	6.455984	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0000
2	aalamin	6.455984	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0000
3	aallah	6.455984	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0000
4	aamiin	5.762837	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.279822	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0000
...
3715	yya	6.455984	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0000
3716	zallm	6.455984	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0705
3717	zaman	6.455984	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0000
3718	zealand	6.455984	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0000
3719	zero	6.455984	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0000

Gambar 8. Hasil TF-IDF

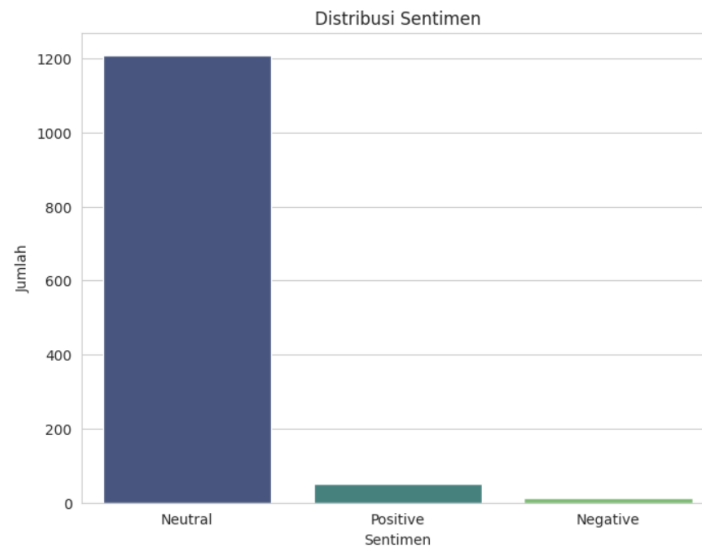
B. Labeling Data

Tahapan selanjutnya setelah proses *preprocessing* adalah labeling data. Data set akan diberikan kategori atau label berdasarkan sentimennya seperti positif, negatif atau netral. Berikut adalah gambar hasil dari proses labeling.

	stemming	score	sentiment
0	['kasi', 'pegang', 'pedang', 'belah', 'kiri']	0.0	Neutral
1	['tolong', 'naik', 'harga', 'kawasaki', 'ninja']	0.0	Neutral
2	['presiden', 'tolong', 'bikin', 'pabrik', 'pen...']	0.0	Neutral
3	['periode', 'pimpin', 'presiden', 'kemiliteran...']	0.0	Neutral
4	['judul', 'instrumen']	0.0	Neutral
...
1268	['makasih', 'pakkk']	0.0	Neutral
1269	['jam', 'nonton', 'jokowi', 'prabowo']	0.0	Neutral
1270	[]	0.0	Neutral
1271	['pindah']	0.0	Neutral
1272	['doa', 'moga', 'kota', 'mek', 'madinah', 'asa...']	0.0	Neutral

Gambar 9. Hasil Labeling

Setelah dilakukan proses labeling data terdapat ketidak seimbangan data dalam data set. Berikut adalah gambar ketidak seimbangan data dalam data set.



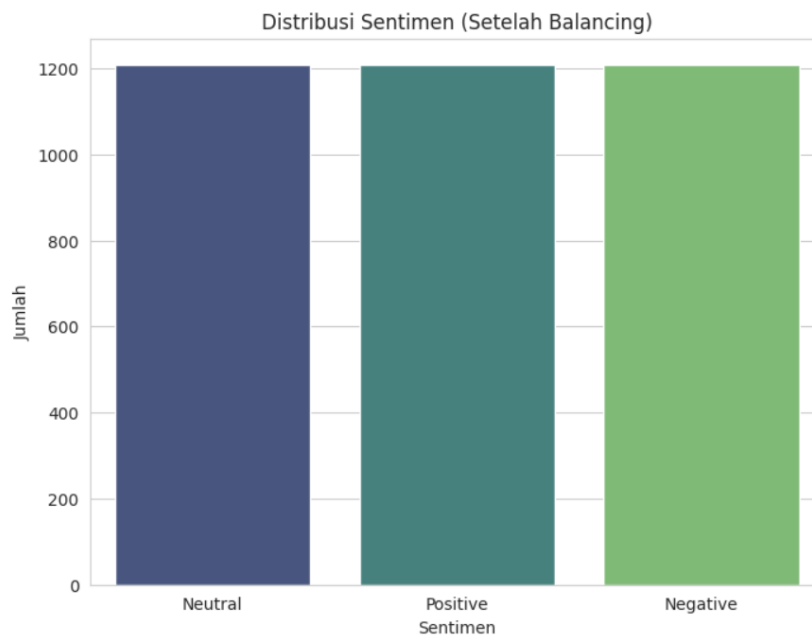
Gambar 10. Ketidakseimbangan Data

Gambar 10 merupakan gambar dari ketidakseimbangan data, di mana ada kelas yang mendominasi kelas lainnya. Hal ini perlu diseimbangkan agar model yang dibuat tidak bias dan dapat mengklasifikasikan data lebih akurat. Jika model yang terdapat ketidakseimbangan data tidak dilakukan *balancing* maka akan menurunkan tingkat akurasi dari model, seperti pada gambar 11.

	precision	recall	f1-score	support
0	1.00	0.86	0.92	254
1	0.07	1.00	0.13	1
3	0.00	0.00	0.00	0
8	0.00	0.00	0.00	0
9	0.00	0.00	0.00	0
13	0.00	0.00	0.00	0
15	0.00	0.00	0.00	0
18	0.00	0.00	0.00	0
20	0.00	0.00	0.00	0
23	0.00	0.00	0.00	0
28	0.00	0.00	0.00	0
accuracy			0.86	255
macro avg	0.10	0.17	0.10	255
weighted avg	1.00	0.86	0.92	255

Gambar 11. Hasil Klasifikasi sebelum Balancing

Pada gambar 11 merupakan hasil dari klasifikasi menggunakan *Naive Bayes* pada data yang terdapat ketidakseimbangan. Hasil dari klasifikasi tersebut menunjukkan hasil akurasi sebesar 86%. Oleh sebab itu, akan dilakukan proses *balancing* atau penyeimbangan data menggunakan metode SMOTE untuk meningkat akurasi dari model. Hasil penyeimbangan dengan SMOTE dapat dilihat pada gambar 12.



Gambar 12. Hasil Balancing

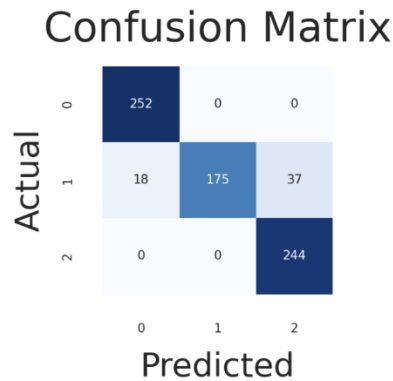
Gambar 12 merupakan hasil dari proses *balancing* menggunakan SMOTE, hasil dari penyeimbangan ini yang nantinya akan digunakan dalam proses klasifikasi *Naive Bayes*. Dengan menggunakan data yang telah diseimbangkan akan meningkatkan akurasi dari model yang dibuat.

C. Klasifikasi *Naive Bayes*

Tahapan selanjut setelah proses *preprocessing* dan labeling adalah melakukan klasifikasi *Naive Bayes*. Data yang telah diolah kemudian akan diklasifikasikan, hasil dari klasifikasi *Naive Bayes* adalah sebagai berikut :

“Prabowo”, “Presiden”, “Rakyat”, “Selamat”, “Jokowi”, “Indonesia”, “Pimpin”, “Moga”, “Negara”. Kata yang sering muncul tersebut ditampilkan dengan ukuran huruf yang lebih besar menunjukkan bahwa kata tersebut merupakan topik utama dalam analisis sentimen. Kata lain yang ditampilkan lebih kecil menandakan bahwa frekuensi kemunculan kata tersebut rendah dalam analisis sentimen ini.

E. Confusion Matrix



Gambar 16. Confusion Matrix

Gambar 15 merupakan *confusion matrix* yang digunakan untuk mengevaluasi kinerja model klasifikasi *Naive Bayes*. Berikut ini adalah penjelasan dari *confusion matrix* tersebut :

- a. Kelas 0
Untuk kelas 0, model berhasil mengidentifikasi 252 *instance* dengan benar sebagai kelas 0, hal ini menunjukkan akurasi yang tinggi dalam mengenali data dari kelas ini. Dengan tidak adanya *instance* dari kelas lain yang diklasifikasikan sebagai kelas 0, maka prediksi untuk kelas ini sangat akurat tanpa kesalahan.
- b. Kelas 1
Pada kelas 1, model mengklasifikasikan dengan benar sebanyak 175 *instance* sebagai kelas 1. Namun terdapat 18 *instance* yang seharusnya masuk kelas 1, tetapi salah diprediksi sebagai kelas 0, dan 37 *instance* yang diprediksi sebagai kelas 2. Hal ini menunjukkan bahwa meskipun model memprediksi kelas 1 dengan baik, masih ada beberapa kesalahan prediksi yang perlu diperbaiki.
- c. Kelas 2
Untuk kelas 2, model berhasil memprediksi 244 *instance* dengan benar sebagai kelas 2. Tidak ada salah prediksi dari kelas lain yang diklasifikasikan sebagai kelas 2, hal ini menunjukkan akurasi yang tinggi dalam pengenalan data kelas 2.

Secara keseluruhan menunjukkan bahwa *confusion matrix* model *Naive Bayes* ini memiliki kinerja yang sangat baik dalam mengklasifikasikan tiap kelasnya. Hal ini ditunjukkan dari tingkat akurasi yang tinggi dengan berhasil mengidentifikasi *instance* tiap kelas dengan benar, seperti pada kelas 1 dan 2 dengan tidak adanya *instance* yang salah diprediksi.

II. KESIMPULAN

Berdasarkan hasil dari penelitian ini yaitu analisis sentimen pada pelantikan Presiden dan Wakil Presiden Indonesia 2024 menggunakan model dan klasifikasi *Naive Bayes*. Data yang digunakan dalam penelitian ini berjumlah 1.371 yang didapat dari hasil *crawling* pada kolom komentar saluran *YouTube* sekretariat kepresidenan. Data yang didapat kemudian diolah melalui proses *preprocessing* yang meliputi tahap *cleansing* untuk membersihkan data, *case folding* untuk mengubah semua huruf menjadi huruf kecil, *tokenize* untuk memecah kalimat menjadi per kata, *stopword removal* untuk menghapus kata yang kurang bermakna, *stemming* untuk mengubah kata menjadi bentuk dasarnya, dan TF-IDF untuk memberi bobot pada kata. Setelah *preprocessing* dilanjutkan dengan *labeling* untuk memberi label atau kategori pada data

berdasarkan sentimennya. Dikarenakan terdapat ketidak seimbangan dalam data maka dilakukan proses penyeimbangan data menggunakan metode SMOTE. Sebelum diseimbangkan hasil akurasi dari klasifikasi sebesar 86%, setelah data diseimbangkan menghasilkan akurasi sebesar 92%. Secara keseluruhan, penelitian ini menunjukkan bahwa Presiden dan Wakil Presiden Indonesia yang baru secara umum disambut dengan baik, hal ini ditunjukkan dengan mayoritas sentimen positif dan netral. Model ini berhasil memberikan hasil yang akurat dalam mengidentifikasi sentimen publik dengan presisi yang akurat, terutama dalam mengklasifikasikan sentimen positif dan netral. Dengan demikian, metode *Naive Bayes* terbukti mampu mengklasifikasikan sentimen secara akurat, sehingga dapat digunakan untuk analisis sentimen karena keandalannya dalam melakukan analisis pada media sosial.

DAFTAR PUSTAKA

- [1] M. R. Fais Sya' Bani, U. Enri, And T. N. Padilah, "Analisis Sentimen Terhadap Bakal Calon Presiden 2024 Dengan Algoritme Naïve Bayes," *Jurikom (Jurnal Ris. Komputer)*, Vol. 9, No. 2, P. 265, 2022.
- [2] S. Kemp, "Digital 2024: Indonesia," *Datareportal*. <https://Datareportal.Com/Reports/Digital-2024-Indonesia>
- [3] Chely Aulia Misrun, E. Haerani, M. Fikry, And E. Budianita, "Analisis Sentimen Komentar *YouTube* Terhadap Anies Baswedan Sebagai Bakal Calon Presiden 2024 Menggunakan Metode *Naive Bayes* Classifier," *J. Coscitech (Computer Sci. Inf. Technol.*, Vol. 4, No. 1, Pp. 207–215, 2023.
- [4] H. Setiawan And I. Zufria, "Analisis Sentimen Pembatalan Indonesia Sebagai Tuan Rumah Piala Dunia Fifa U-20 Menggunakan Naïve Bayes," *J. Media Inform. Budidarma*, Vol. 7, No. 3, Pp. 1003–1012, 2023.
- [5] I. L. Achmadi, H. Yuana, And M. F. Rahmat, "Analisis Sentimen Hasil Pemilihan Presiden Ri 2024 Pada Twitter Atau X Dengan Metode *Naive Bayes* Classifier," *J. Soc. Sci. Res.*, Vol. 4, No. 5, Pp. 9033–9048, 2024.
- [6] K. A. Lubis, M. T. A. Bangsa, And A. Yudertha, "Analisis Sentimen Opini Masyarakat Terhadap Pindahannya Ibu Kota Indonesia Dengan Menggunakan Klasifikasi Naïve Bayes," *J. Teknoinfo*, Vol. 18, No. 1, Pp. 226–238, 2024.
- [7] F. Setiawan And P. F. Ariyani, "Pada *YouTube* Berbasis Web Dengan Naïve Bayes Sentiment Analysis Of The Constitutional Court ' S Decision On The 2024 Presidential Election Dispute On Web-Based *YouTube* With *Naive Bayes*," *Semin. Nas. Mhs. Fak. Teknol. Inf.*, Vol. 3, No. 2, Pp. 769–778, 2024.
- [8] B. Ramadhani And R. R. Suryono, "Komparasi Algoritma Naïve Bayes Dan Logistic Regression Untuk Analisis Sentimen Metaverse," *J. Media Inform. Budidarma*, Vol. 8, No. 2, Pp. 714–725, 2024.
- [9] S. Fauzi, A. P. Siswono, And A. F. Prayitno, "Analisis Sentimen Jersey Erspo Timnas Indonesia Menggunakan Model Klasifikasi Dan Algoritma Naïve Bayes," *J. Tekno Kompak*, Vol. 19, No. 1, Pp. 25–36, 2024.
- [10] A. D. Dayani, Yuhandri, And G. W. Nurcahyo, "Analisis Sentimen Terhadap Opini Publik Pada Sosial Media Twitter Menggunakan Metode Support Vector Machine," *J. Komtekinfo*, Vol. 11, No. 1, Pp. 1–10, 2024.
- [11] T. Z. Fatikhah, M. Muhtarom, And I. Oktaviani, "Analisis Sentimen Ulasan Mie Gacoan Solo Veteran Di Google Maps Menggunakan," *Zo. J. Sist. Inf.*, Vol. 6, No. 3, Pp. 649–658, 2024.
- [12] A. O. Praneswara And N. Cahyono, "Analisis Sentimen Ulasan Aplikasi Tiktok Shop Seller Center Di Google Playstore Menggunakan Algoritma *Naive Bayes*," *Indones. J. Comput. Sci.*, Vol. 12, No. 1, Pp. 3925–3940, 2023.
- [13] K. Akbar And M. Hayaty, "Data Balancing Untuk Mengatasi Imbalance Dataset Pada Prediksi Produksi Padi," *J. Ilm. Intech Inf. Technol. J. Umus*, Vol. 2, No. 02, Pp. 1–14, 2020.
- [14] T. A. Dewi And E. Mailoa, "Jurnal Indonesia : Manajemen Informatika Dan Komunikasi Perbandingan Implementasi Metode Smote Pada Algoritma Support Vector Machine (Svm) Dalam Analisis Sentimen Opini Jurnal Indonesia : Manajemen Informatika Dan Komunikasi," *J. Indones. Manaj. Inform. Dan Komun.*, Vol. 4, No. 3, Pp. 849–855, 2023.
- [15] N. Muslimah, A. Sutikno, And S. Kom, "Analisis Sentimen Komentar Netizen Pada Brand Skincare The Originote Menggunakan Metode Naïve Bayes," *Proc. Natl. Conf. Electr.*, Vol. 3, No. 1, Pp. 826–835, 2023.

- [16] A. Safira, F. N. Hasan, K. J. Timur, And N. B. Classifier, "Paylater Menggunakan Metode *Naive Bayes*," *Zo. J. Sist. Inf.*, Vol. 5, No. 1, Pp. 59–70, 2023.
- [17] E. Magdalena And B. Hakim, "Sentimen Kebijakan Pemerintah Terhadap Kepercayaan Masyarakat Menggunakan Metode Nlp Rule Based Sentiment On Public Trust Using The Nlp Rule Based Method," *Justin (Jurnal Sist. Dan Teknol. Informasi)*, Vol. 12, No. 1, Pp. 175–182, 2024.