

Perbandingan Akurasi Metode Analisis Sentimen Untuk Evaluasi Opini Pengguna Pada Platform Media Sosial (Studi Kasus: Twitter)

Rabhita Mursyid¹, Aries Dwi Indriyanti²

^{1,2} Jurusan Teknik Informatika Fakultas Teknik Universitas Negeri Surabaya

¹rabhita.17051204074@mhs.unesa.ac.id

²ariesdwi@unesa.ac.id

Abstrak— Dalam era digitalisasi, media sosial telah menjadi platform utama bagi individu untuk berbagi opini dan berinteraksi dengan pengguna lainnya. Khususnya di Twitter, yang memiliki miliaran pengguna aktif, aliran informasi dan volume data yang besar memerlukan pendekatan yang efisien dalam menganalisis sentimen.

Penelitian ini bertujuan untuk menganalisis sentimen opini masyarakat di media sosial Twitter setelah Pemilihan Presiden 2024 di Indonesia. Dengan menggunakan teknik *pre-processing* dan tiga algoritma klasifikasi, yaitu *Naïve Bayes*, *Random Forest*, dan *Support Vector Machine* (SVM). Penelitian ini mengevaluasi performa metode analisis sentimen dalam mengidentifikasi dan mengklasifikasikan sentimen terkait pemilu.

Data yang digunakan berasal dari Twitter, dengan total 806 data. Proses melibatkan *crawling* data, pembersihan data, pelabelan data, dan pelatihan data dengan metode TF-IDF sebelum dilakukan klasifikasi sentimen.

Hasil eksperimen menunjukkan bahwa algoritma *Support Vector Machine* (SVM) memperoleh hasil terbaik dengan akurasi sebesar 74.79%, *recall* 74.79%, dan *precision* 76.04%. Penelitian ini menemukan bahwa SVM memiliki kinerja terbaik dalam mengidentifikasi dan mengklasifikasikan sentimen. Hasil analisis sentimen ini dapat memberikan wawasan berharga bagi pengambil keputusan dan *stakeholders* dalam pemilu 2024 di Indonesia, memfasilitasi pemahaman yang lebih mendalam tentang sentimen masyarakat.

Kata Kunci— Perbandingan, *Naïve Bayes*, *Random Forest*, SVM (*Support Vector Machine*), Pemilu 2024, Sentimen Publik.

I. PENDAHULUAN

Dalam era digitalisasi yang terus berkembang, media sosial telah menjadi sarana utama bagi individu untuk berbagi informasi, berinteraksi, dan menyampaikan pendapat mereka secara daring. Perkembangan teknologi saat ini membantu banyak orang dengan mudah mengutarakan pendapatnya di media sosial *online* sebagai bentuk *feedback* dan reaksi terhadap suatu hal. Fenomena ini memberikan dampak signifikan terhadap cara kita memahami dan mengukur sentimen pengguna terhadap berbagai topik, produk, atau layanan. Oleh karena itu, penting untuk memiliki alat analisis sentimen yang dapat menggali dan mengevaluasi opini pengguna secara efektif [1].

Referensi [2] mengemukakan bahwa menggunakan media sosial untuk mendapatkan opini masyarakat dapat dianggap lebih dapat diandalkan daripada survei, karena adanya risiko pernyataan pendapat yang tidak akurat. Analisis sentimen masyarakat sangat penting dalam mengidentifikasi masalah

yang berkaitan dengan kebijakan pemerintah. Melalui pendekatan ini, masyarakat dapat aktif terlibat dalam proses pengambilan keputusan untuk memajukan pertumbuhan dan inovasi yang bertujuan meningkatkan kualitas hidup mereka.

Salah satu platform yang paling banyak digunakan untuk berbagi opini dan pengalaman adalah media sosial. Miliaran pengguna aktif setiap harinya menghasilkan sejumlah besar data berupa teks, gambar, dan video. Analisis sentimen pada platform media sosial dapat memberikan wawasan yang berharga bagi berbagai kepentingan, seperti pemahaman persepsi pengguna terhadap merek, produk, atau kejadian tertentu [3].

Dari sekian banyak sosial media platform yang tersedia, Twitter menjadi salah satu yang favorit untuk digunakan. Sepanjang tahun 2018, jumlah aktif pengguna Twitter pada tiap harinya berhasil mencapai angka 126 juta atau naik 9% dari tahun sebelumnya [1]. Twitter sendiri merupakan media sosial real-time yang memungkinkan penggunanya berdiskusi berbagai topik dan mengungkapkan pemikiran secara spontan. Informasi yang dapat dibagikan di Twitter disebut tweet. Tweet memiliki batas yang dulunya 140 karakter menjadi 280 karakter. Tweet dapat berupa teks, foto, video, dan sebagainya, melalui tweet pengguna dapat berinteraksi satu sama lain. Pengguna dapat membagikan informasi, memberikan pendapat, menjelaskan suatu kejadian dan sebagainya [4].

Opini pengguna di media sosial memiliki nilai besar karena dapat memberikan gambaran tentang persepsi masyarakat terhadap suatu hal. Oleh karena itu, penting bagi pihak-pihak yang terkait, seperti perusahaan, penyelenggara acara, atau pemerintah, untuk dapat memahami dan mengevaluasi opini pengguna secara efisien. Namun, dengan volume data yang besar dan cepatnya aliran informasi di media sosial, analisis manual menjadi tidak memungkinkan dan tidak efisien [5].

Salah satu pendekatan yang dapat digunakan untuk mengevaluasi opini pengguna secara otomatis adalah melalui analisis sentimen. Analisis sentimen merupakan teknik yang memungkinkan sistem untuk mengidentifikasi dan mengekstrak sentimen atau perasaan yang terkandung dalam teks, baik itu positif, negatif, atau netral. Dengan menggunakan algoritma analisis sentimen, kita dapat mengotomatiskan proses evaluasi opini pengguna di media sosial, sehingga informasi yang berharga dapat diperoleh secara cepat dan efisien [6].

Analisis sentimen merupakan salah satu cara untuk mengetahui pendapat orang banyak terhadap perilaku sesuatu

seperti layanan publik, isu, kinerja pemerintahan atau hal lainnya [12]. Analisis sentimen ini dilakukan untuk melihat kecenderungan opini masyarakat terhadap isu-isu yang berkembang terutama di media sosial. Analisis sentimen ini merupakan suatu pengklasifikasian dengan mengekstraksi pendapat, emosi, dan evaluasi seseorang yang tertulis dalam sebuah pembicaraan mengenai topik tertentu dengan memanfaatkan *Natural Language Processing* (Aziz, 2022). Menurut Zulfa & Winarko, Analisis sentimen adalah sebuah riset komputasional dari opini, sentimen dan emosi yang diekspresikan secara tekstual (R. Sari, 2020). Oleh karena itu, analisis sentimen merupakan salah satu cabang dalam penelitian *text mining* yang melakukan klasifikasi dalam sebuah dokumen teks (Cahyaningtyas et al., 2021).

Analisis sentimen atau *opinion mining* adalah pembelajaran komputasi untuk mengidentifikasi dan mengekstrak juga mempelajari suatu opini, sentimen, emosi, penilaian dan pandangan dalam bentuk teks (Liu B. , 2012). Analisis Sentimen dapat mendeteksi apakah suatu teks mengandung opini positif, negatif maupun netral.

Analisis sentimen adalah contoh bidang pemrosesan bahasa alami (NLP) yang paling populer. Pemrosesan bahasa alami (NLP) adalah bidang ilmiah yang menjelaskan bagaimana komputer dapat bertindak dan berpikir seperti manusia. Pemrosesan bahasa alami (NLP) adalah bagian dari kecerdasan buatan. Dalam perkembangan *data mining*, kecerdasan buatan (AI) merupakan salah satu dari empat cabang ilmu data mining, yaitu statistika, *database*, dan *information retrieval*. Dalam penerapannya, kecerdasan buatan (AI) juga membutuhkan pembelajaran mesin sebagai algoritma solusinya. Keberadaan *machine learning* dimanfaatkan untuk menggantikan manusia dalam pengambilan keputusan. Pembelajaran mesin tidak terasa seperti manusia sehingga keputusan dibuat berdasarkan data yang diproses.

Menurut Cutlip dan Center, opini adalah pernyataan tentang sikap terhadap masalah kontroversial (dalam Olii dan Erlita, 2011: 39). Opini muncul dari diskusi mengenai isu kontroversial yang menghasilkan berbagai pandangan berbeda. Istilah "opini" berasal dari bahasa asing "*opinion*," yang merupakan tanggapan atau jawaban terhadap suatu persoalan yang dinyatakan dalam kata-kata, baik secara tertulis maupun lisan. Opini juga dapat mencakup perilaku, sikap, pandangan, dan tanggapan.

Opini bisa diungkapkan secara aktif atau pasif, verbal (lisan), dan baik secara terbuka melalui kata-kata yang jelas, maupun melalui pilihan kata yang halus atau tersirat, serta dapat diartikan secara konotatif atau perseptual (personal). Opini juga bisa diekspresikan melalui perilaku, sikap, mimik wajah, bahasa tubuh, atau simbol-simbol tertulis [7].

Salah satu media yang banyak digunakan adalah twitter. Twitter merupakan media sosial untuk bertukar pikiran dan pendapat. Pengguna twitter dapat mengirim dan menerima pesan tweet berupa teks, gambar, ataupun video. Perbedaan dengan media sosial lain, di twitter karakter untuk menulis pesan dibatasi sampai 280 karakter, sedangkan media sosial

lainnya tidak dibatasi. Twitter bersifat publik sehingga status yang dibagikan dapat dilihat oleh orang lain meskipun bukan pengikutnya. Namun, pengiriman tweet juga dapat dibagikan hanya kepada temannya saja atau *followers*. Twitter mempunyai kelebihan yaitu jangkauan yang luas, dapat menjangkau publik figur, media promosi lebih luas, banyak jaringan, dan lebih mudah diukur kemampuannya.

Kerangka konseptual penelitian ini didasarkan pada integrasi antara pemahaman teoritis analisis sentimen, prinsip-prinsip pemrosesan bahasa alami, dan teknik-teknik pembelajaran mesin. Pertama, penelitian ini membangun fondasi pada konsep analisis sentimen, yang melibatkan identifikasi dan klasifikasi opini pengguna menjadi kategori positif, negatif, atau netral. Pemahaman mendalam terhadap teori analisis sentimen ini menjadi dasar dalam merancang algoritma yang dapat menangkap nuansa kompleks dari berbagai jenis bahasa dan ekspresi pengguna di platform Twitter.

Selanjutnya, kerangka konseptual ini mencakup penerapan prinsip-prinsip pemrosesan bahasa alami (NLP) untuk meningkatkan kecerdasan algoritma analisis sentimen. Aspek-aspek NLP, seperti tokenisasi, *stemming*, dan entitas pengenalan, digunakan untuk memproses dan mengurai teks pengguna sehingga algoritma dapat mengenali konteks dan makna yang lebih dalam dari setiap opini. Dengan memadukan prinsip NLP, penelitian ini bertujuan untuk mengetahui perbandingan akurasi metode analisis sentimen terhadap konten yang ada di Twitter.

Terakhir, kerangka konseptual ini mencakup pemanfaatan teknik-teknik pembelajaran mesin, seperti pembelajaran terawasi dan tidak terawasi, untuk melatih dan mengoptimalkan kinerja algoritma analisis sentimen. Penggunaan dataset yang memadai dan representatif menjadi kunci dalam mengembangkan model pembelajaran mesin yang dapat mengenali pola sentimen dengan tingkat akurasi yang tinggi. Integrasi ini mencerminkan pendekatan holistik dalam pengembangan algoritma analisis sentimen yang dapat memahami kompleksitas bahasa dan respons pengguna di platform media sosial Twitter. Dengan menggabungkan aspek-aspek ini, diharapkan penelitian ini dapat memberikan kontribusi signifikan dalam pengembangan algoritma analisis sentimen yang efektif dan relevan untuk evaluasi opini pengguna di dunia media sosial.

II. METODOLOGI PENELITIAN

A. Jenis/Pendekatan Penelitian

Pada penelitian ini, penulis menggunakan metode eksperimen/percobaan. Metode eksperimen adalah metode sistematis yang menghubungkan sebab-akibat. Penelitian ini tergolong penelitian kuantitatif yang di dalamnya mengandung syarat; *control*, *manipulate* dan *observation* (Nazir, 2017). Eksperimen dilakukan untuk mendapatkan model terbaik dengan mengimplementasikan tweets dengan analisis sentimen.

B. Rancangan Penelitian

Rancangan penelitian untuk "Perbandingan Akurasi Metode Analisis Sentimen Untuk Evaluasi Opini Pengguna Pada Platform Media Sosial (Studi Kasus: Twitter)" mencakup langkah-langkah yang sistematis dan terstruktur. Berikut adalah rancangan penelitian yang dapat diikuti:

1. Identifikasi Masalah dan Perumusan Tujuan
 - Identifikasi masalah terkait dengan analisis sentimen pada Twitter.
 - Perumusan tujuan penelitian untuk perbandingan algoritma analisis sentimen yang efektif.
2. Studi Literatur
 - Tinjau literatur terkait analisis sentimen, pemrosesan bahasa alami, dan pembelajaran mesin.
 - Identifikasi metode-metode terkini dan temuan-temuan terkait.
3. Pengembangan Kerangka Konseptual
 - Perancangan kerangka konseptual berdasarkan teori-teori yang muncul dari studi literatur
4. Pemilihan Metode Analisis Sentimen
 - Pemilihan metode analisis sentimen yang sesuai dengan tujuan penelitian
 - Perencanaan pendekatan analisis, apakah berbasis aturan, pembelajaran mesin, atau kombinasi keduanya
5. Desain Algoritma
 - Perbandingan akurasi metode analisis sentimen berdasarkan kerangka konseptual
 - Perancangan langkah-langkah pemrosesan data dan teknik analisis yang akan diterapkan
6. Perbandingan akurasi Perangkat Lunak
 - Implementasi algoritma dalam bentuk perangkat lunak atau *script*, menggunakan *Python* atau bahasa pemrograman lainnya
 - Integrasi teknik *crawling* data Twitter untuk pengumpulan *dataset*.
7. Pengumpulan Data
 - Studi literatur: Pengumpulan informasi dari buku, jurnal, dan artikel terkait
 - Observasi: Analisis konten dari literatur dan pemahaman karakteristik opini di Twitter
 - *Crawling* data Twitter: Pengumpulan *dataset* melalui teknik *crawling* untuk mendukung analisis sentimen
8. Validasi Algoritma
 - Pengujian dan validasi algoritma menggunakan *dataset* yang telah dikumpulkan
 - Evaluasi kinerja algoritma melalui metrik metrik seperti akurasi, presisi, dan *recall*
9. Analisis Data
 - Analisis hasil dari pengujian algoritma sentimen terhadap *dataset* Twitter
 - Penarikan kesimpulan terkait perbandingan akurasi algoritma metode analisis sentimen

10. Penulisan Laporan Penelitian

Penulisan laporan penelitian yang mencakup semua tahap penelitian, temuan, dan implikasi praktis dan teoritis

11. Diseminasi Hasil

- Presentasi hasil penelitian dalam konferensi ilmiah atau seminar
- Publikasi dalam jurnal artikel ilmiah



Gbr. 1 Flowchart

C. Populasi Dan Sampel

Populasi merujuk pada seluruh kumpulan data yang relevan dengan penelitian yang dilakukan. Misalnya, jika peneliti ingin menganalisis sentimen publik terhadap suatu merek di media sosial, maka populasi dapat mencakup semua

postingan, komentar, atau ulasan yang menyebut merek tersebut di berbagai platform media sosial. Mengingat populasi ini bisa sangat besar dan sulit untuk diakses sepenuhnya, peneliti sering kali menggunakan teknik sampling untuk mengurangi jumlah data yang harus dianalisis tanpa mengorbankan akurasi hasil.

Sampel adalah subset dari populasi yang dipilih untuk dianalisis. Pemilihan sampel yang representatif sangat penting untuk memastikan bahwa hasil analisis dapat digeneralisasi ke seluruh populasi. Dalam analisis sentimen, sampel dapat dipilih secara acak atau dengan menggunakan metode sampling lain seperti stratified sampling atau systematic sampling. Teknik-teknik ini membantu peneliti dalam mengatasi kendala waktu dan sumber daya, sambil tetap memberikan gambaran yang akurat tentang sentimen keseluruhan populasi.

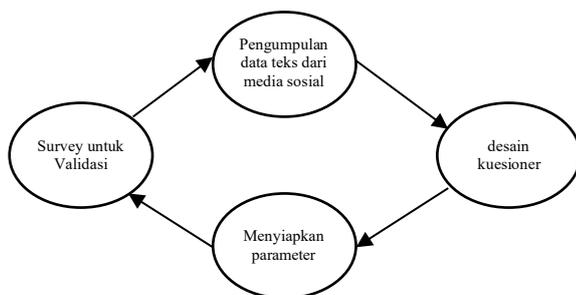
D. Sumber Data Penelitian

Sumber data mengacu pada segala hal yang dapat memberikan informasi mengenai data. Berdasarkan asalnya, data dapat dibagi menjadi dua jenis, yaitu data primer dan data sekunder.

1. Data primer adalah data yang dikumpulkan langsung oleh peneliti dari sumber asli untuk menyelesaikan masalah yang sedang diteliti.
2. Data sekunder adalah data yang sudah dikumpulkan sebelumnya untuk tujuan lain dan dapat diakses dengan cepat, seperti literatur, artikel, jurnal, dan situs web terkait penelitian.

Selain data primer, sumber data yang dipakai peneliti adalah sumber data sekunder, data sekunder didapat melalui berbagai sumber yaitu literatur artikel, serta situs di internet yang berkenaan dengan penelitian yang dilakukan.

E. Instrumen Penelitian



Gbr. 2 Instrumen Penelitian

F. Teknik Pengumpulan Data

Penelitian ini melibatkan sejumlah metode pengumpulan data yang dirancang untuk mendukung pemahaman dan analisis terhadap topik "Perbandingan Akurasi Metode Analisis Sentimen Untuk Evaluasi Opini Pengguna Pada Platform Media Sosial (Studi Kasus: Twitter)." Metode pengumpulan data yang digunakan melibatkan studi literatur, observasi, dan teknik pengambilan data dari platform Twitter.

Studi literatur menjadi langkah awal dalam merinci kerangka konseptual dan memahami konsep-konsep dasar terkait analisis sentimen, pemrosesan bahasa alami, dan pembelajaran mesin. Pada tahap ini, peneliti menyelidiki berbagai buku, jurnal, dan artikel terkait untuk memperoleh dasar teoritis yang kokoh untuk pengembangan algoritma. Observasi menjadi metode kedua, yang melibatkan analisis konten dari literatur-literatur yang telah dipelajari dan juga pemahaman terhadap karakteristik opini pengguna di platform Twitter. Observasi ini mendukung pemilihan metode analisis sentimen yang sesuai dan relevan dengan konteks media sosial tersebut. Selain itu, penelitian ini juga memanfaatkan teknik *crawling* data di Twitter menggunakan *Python 3.10*. *Crawling* data memungkinkan peneliti untuk mengumpulkan *dataset* yang representatif dari opini pengguna di Twitter. Proses ini dapat mencakup pengambilan tweet-tweet tertentu, pengumpulan *metadata*, dan informasi lain yang relevan untuk mendukung analisis sentimen.

G. Teknik Analisis Data

Text Mining

Text Mining adalah proses menemukan informasi dan pengetahuan dari teks untuk memperoleh pola spesifik. Teknik ini sering digunakan untuk analisis informasi dan pengambilan keputusan dengan data teks dalam jumlah besar. Tahap awal melibatkan data *pre-processing*, yang mencakup metode untuk membersihkan dan menstandarisasi data teks. Beberapa metode tersebut meliputi *cleansing*, *case folding*, *tokenizing*, *filtering*, dan *stemming*, yang dijelaskan lebih lanjut sebagai berikut:

1. *Cleansing* : Menghapus komponen tidak perlu seperti HTML, *emoticon*, *hashtag*, *username*, *retweet*, URL, dan alamat *website*.
2. *Case Folding* : Mengubah teks menjadi huruf kecil dan menghapus karakter non-huruf seperti angka dan tanda baca.
3. *Tokenizing* : Memisahkan kalimat menjadi kata tunggal atau token.
4. *Filtering* : Menghapus *stopwords* atau kata-kata yang tidak penting untuk analisis sentimen menggunakan *stopword list*.
5. *Stemming* : proses mengubah kata berimbuhan menjadi bentuk kata dasar sesuai struktur morfologinya, sehingga kata yang telah diubah memiliki arti yang sama dengan kata dasarnya.

III. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Dalam langkah-langkah pengumpulan data peneliti menggunakan bahasa pemrograman *Python* untuk mengambil data dari platform Twitter dengan fokus pada topik hangat yaitu pemilu 2024 di Indonesia. Peneliti memanfaatkan kata kunci "#pemilu2024" dalam bahasa Indonesia (lang:id) sebagai kriteria pencarian. Data yang diambil mencakup

tweet-tweet yang mengandung kata kunci tersebut. Selanjutnya hasil pengambilan data disimpan dalam sebuah file dengan nama "pemilu.csv," yang memiliki format berkas CSV. Untuk mengontrol jumlah data yang diambil maka peneliti membatasi hingga 1000 tweet terbaru yang memenuhi kriteria pencarian. Pengambilan data dilakukan dengan menggunakan perintah "!npx --yes tweet-harvest@latest," yang merupakan instruksi untuk menjalankan perangkat lunak tweet-harvest. Dengan melalui langkah-langkah ini maka peneliti berhasil mengumpulkan data yang diperlukan untuk analisis terkait dengan pemilu 2024.

Contohnya bisa berupa ulasan yang menyuarakan ketidakpuasan terhadap layanan pelanggan atau kualitas produk. Sentimen negatif berkaitan dengan pengalaman yang kurang memuaskan atau kekecewaan. Adapun untuk mengatasi *cross labelling* yang mengacu pada situasi di mana dua atau lebih label diberikan pada satu data latih atau sampel teks yang dapat menyebabkan ketidakjelasan terkait dengan label yang sesuai untuk suatu teks. Maka peneliti membuat aturan khusus untuk mengatasi hal tersebut dengan mengklasifikasikan teks tersebut ke dalam sentimen netral dengan ketentuan aturan pelabelan sentimen sebagai berikut:

```
import pandas as pd

# Tentukan nilai pemilu sesuai nama file yang ingin Anda baca
pemilu = "pemilu.csv"

# Spesifikasikan path ke file CSV Anda
file_path = f"tweets-data/pemilu.csv"

# Baca file CSV ke dalam DataFrame pandas
df = pd.read_csv(file_path,
delimitter=";", on_bad_lines="skip")

# Tampilkan DataFrame
display(df)
```

Gbr. 3 Pengumpulan Data

Berikut adalah hasil dari *crawling data tweet* :

full_text,Sentimen
@Fahrihamzah Bung, kalao sekelas rakyat dpt segini, kira kira influencer / artis dpt SENNA ZALVA REXONA BABY SUJI JOIN GRUP : https://t.co/bMloYzyWvw tamara
Mau dikasih fakta dilapangan klo kampanye 02 sepi, para kang survey gak peduli, k @Wajidan_Channel Kampanye 01 Rame full militan prabowo 2014 & 2019 Ka SENNA ZALVA REXONA BABY SUJI JOIN GRUP : https://t.co/bMloYzyWvw tamara
02 di GBK Kampanye 02 Cak Imin JIS VS GBK Cak Imin,netral
gue ke solo kirain bakal rame kampanye 02, ternyata jalan sampe macet pada kam @kafiradikalis Kampanye 01 Rame full militan prabowo 2014 & 2019 Kampan @dinikopi13 @itsindahg Kalau Bali ya nggak heran. Dari dulu mostly all in PDIP. Ga Hormat utk pendukung Prabowo Gibran yg mengerti dan tdk memaksakan masuk ke Sorry yee, kandang @Masinton 02 di GBK Kampanye 02 #PrabowoGibran https://
Ew Palmerah macet gara-gara kampanye 02.,negatif
@KopiKoetjoer @kafiradikalis Ini kampanye 02 dibanding!,netral

Gbr. 4 Hasil Pengumpulan Data

B. Pelabelan Dataset

Langkah berikutnya adalah melakukan proses pelabelan data pada tweet yang akan dilakukan secara manual menggunakan aplikasi Microsoft Excel. Proses ini memungkinkan peneliti untuk mendapatkan label atau kategori tertentu pada setiap tweet berdasarkan isinya yaitu negatif, positif dan netral. Sentimen positif adalah perasaan atau ekspresi yang mendukung atau menyukai suatu hal. Misalnya jika seseorang memberikan ulasan yang memuji kualitas suatu produk maka dianggap sentimen positif. Hal ini sering terkait dengan pengalaman menyenangkan atau kepuasan. Sentimen negatif sebaliknya adalah perasaan atau ekspresi yang mengecam atau tidak menyukai suatu hal.

TABEL I
 KETENTUAN LABELLING

Dimensi	Sentimen Positif	Sentimen Negatif	Sentimen Netral
Ciri	Ekspresi kepuasan, pujian, atau dukungan positif.	Kritik, ketidakpuasan, atau ekspresi negatif.	Tidak ada ekspresi yang jelas positif atau negatif.
Contoh	Ayo kita sukseskan pemilu yang adil dan transparan	Apaan deh tiap kampanye cuman joget joget aja!	Saya masih bingung untuk menentukan pilihan kemana

Dengan melakukan pelabelan ini maka penelitian ini dapat memperoleh pemahaman yang lebih mendalam tentang sentimen dan pandangan masyarakat terkait pemilu 2024. Pelabelan data ini merupakan langkah penting untuk memvalidasi dan mengklasifikasikan data *tweet* sehingga analisis selanjutnya dapat dilakukan dengan lebih akurat dan informatif.

full_text,"Sentimen"
kalao kelas rakyat artis waras kampanye gbkvsjjs,"negatif"
senna zalva rexona baby suji join grup tamara amin ahok libur fedi nuril dante kapa kasih fakta lapang kampanye sepi kang survey peduli karna ikat kontrak giring opini channel kampanye rame full militan prabowo kampanye rame full militan jokowi kan senna zalva rexona baby suji join grup tamara amin abah ahok libur fedi nuril dante k kampanye imin imin,"netral"
solo kirain rame kampanye jalan sampe macet kampanye,"positif"
kampanye rame full militan prabowo kampanye rame full militan jokowi kampanye s nggak heran mostly pdip gatahu kampanye kemarin hasil sulit cari bilang anies rame hormat pendukung prabowo gibran erti paksa masuk stadion batas stadion penuh kan sorry kandang kampanye,"positif"
palmerah macet garagara kampanye,"negatif"
kampanye bandung,"netral"

Gbr. 5 Hasil Pelabelan

Setelah dilakukan proses pelabelan data dan pembersihan data duplikat diperoleh hasil keseluruhan total data sebanyak 806 data dengan jumlah kategori tweet terbanyak adalah data tweet berkategori negatif sebanyak 383 data. Kemudian jumlah data tweet berkategori positif sebanyak 362 data dan kategori netral sebanyak 61 data.

C. Text Pre-processing

Pre-processing adalah tahapan awal dalam pengolahan data di mana data mentah atau tidak terstruktur diubah menjadi data yang lebih terstruktur dan siap digunakan untuk analisis lebih lanjut (Hermawan & Bellanar Ismiati, 2020). Pada penelitian ini dilakukan enam tahapan Pre-processing menggunakan bahasa pemrograman python yaitu sebagai berikut :

1. *Cleansing dan Transform Case*

Pada tahapan ini dilakukan proses membersihkan dokumen dari kata-kata yang tidak diperlukan atau disebut juga sebagai tahapan *cleaning*. Tujuannya agar dokumen yang akan diolah menjadi lebih bersih dan relevan. Salah satu cara untuk membersihkan dokumen adalah dengan menghapus kata yang tidak berkontribusi dalam analisis teks (Syarifuddin, 2020). Lalu dilakukan juga *Transform case* untuk mengubah huruf pada kata-kata dalam dokumen menjadi huruf kecil atau huruf besar. Pada penelitian ini akan diubah menjadi huruf kecil. Adapun proses *Cleansing dan Transform case* menggunakan bahasa pemrograman python untuk melakukan proses permbersihan data sentimen. Adapun proses dalam *Cleansing dan Transform case* ini dapat ditunjukkan pada Gambar 6 :

```
# Menyimpan ke file CSV
data_clean.to_csv("/content/Casefolding_B
areksa.csv", index=False)
```

Gbr. 6 Kode Program *Cleansing dan Transform Case*

Hasil dari proses *cleaning* dapat dilihat pada tabel hasil data setelah dan sesudah *Cleansing dan transform case* yang ditunjukkan pada Tabel II.

TABEL III
 HASIL SESUDAH *CLEANSING*

Sebelum	Sesudah	Kategori
Suasana Kampanye, 02 Di GBK , Prabowo-Gibran menuju kemenangan. #KamuKITA02BIRUinGbk https://t.co/8JEFgtZPNr	suasana kampanye di gbk prabowogibran menuju kemenangan	Positif
Pemberangkatan Distribusi Kotak dan Bilik Suara Pemilu tahun 2024 Kapolres Madiun AKBP M Ridwan, S.I.K., https://t.co/UtfHVJk60n . bersama Forkopimda Kab. Madiun dan KPU melaksanakan pemberangkatan distribusi kotak dan bilik suara Pemilu tahun 2024 bertempat di Gedung Korpi alun https://t.co/gEPLXNEj3r	pemberangkatan distribusi kotak dan bilik suara pemilu tahun kapolres madiun akbp m ridwan sik bersama forkopimda kab madiun dan kpu melaksanakan pemberangkatan distribusi kotak dan bilik suara pemilu tahun bertempat di gedung korpi alun	Netral
FAKTA ² LAPANGAN KAMPANYE 02 Inilah hasil 'survey fakta' yg membuat Jokowi turun gunung, dan Prabowo MARAH BESAR. Padahal sudah digelontorkan dana besar, tapi tetap sepi pengunjung dan sepi. Hal ini yg bikin tetangga sebelah PANIK, GUSAR dan BINGUNG https://t.co/COjaed58mI	fakta lapangan kampanye inilah hasil survey fakta yg membuat jokowi turun gunung dan prabowo marah besar padahal sudah digelontorkan dana besar tapi tetap sepi pengunjung dan sepi hal ini yg bikin tetangga sebelah panik gusar dan bingung	Negatif

Tabel II menunjukkan perbedaan sebelum dan sesudah dilakukan proses *Cleansing* di mana hasilnya membuat teks menjadi lebih terstruktur dan tidak terdapat *symbol* dan kata – kata yang tidak relevan seperti

```
import pandas as pd
import re

def clean_text(df, text_field,
new_text_field_name):
    df[new_text_field_name] =
df[text_field].str.lower()
    df[new_text_field_name] =
df[new_text_field_name].apply(lambda
elem: re.sub(r"(@[A-Za-z0-9]+)|([\^0-9A-
Za-z \t#])|(\w+:\/\/\/\S+)|^rt|http.+?",
"", elem))
    df[new_text_field_name] =
df[new_text_field_name].apply(lambda
elem: re.sub(r"\d+", "", elem))
    df[new_text_field_name] =
df[new_text_field_name].apply(lambda
elem: re.sub(r"#\S+", "", elem)) #
Menghapus hashtag dan teksnya
    return df

# Membaca data
file_path = "/content/oldat (6).csv"
df = pd.read_csv(file_path)

# Menjalankan fungsi clean_text
df['text_clean'] =
df['full_text'].str.lower()
data_clean = clean_text(df, 'full_text',
'text_clean')
```

https://t.co/UtfHVJk60n. dan #KamuKITA02BIRUinGbk. Lalu tabel ini menunjukkan perbedaan sebelum dan sesudah dilakukan proses *Transform case* di mana hasilnya membuat teks berhuruf kecil.

2. Filter Stopword

Filter stopwords adalah proses menghilangkan kata-kata stop (*stop words*) dari teks. Kata-kata stop adalah kata-kata yang dianggap tidak signifikan dalam pemrosesan bahasa alami karena mereka tidak memberikan nilai tambah dalam memilih konten yang relevan. Contoh kata-kata stop meliputi “dan,” “atau,” “di,” dan sejenisnya.

turun gunung dan prabowo marah besar padahal sudah digelontorkan dana besar tapi tetap sepi pengunjung dan sepi hal ini yg bikin tetangga sebelah panik gusar dan bingung	prabowo marah digelontorkan dana sepi pengunjung sepi yg bikin tetangga sebelah panik gusar bingung	
---	---	--

Tabel III menunjukkan perbedaan sebelum dan sesudah dilakukan proses *Filter stopwords* di mana hasilnya membuat kata kata penghuhung seperti dan, pada, itu dan lain- lain menjadi hilang.

3. Tokenisasi

Tokenisasi adalah proses mengubah dokumen teks menjadi serangkaian token atau unit-token, seperti kata-kata atau frasa-frasa, yang lebih mudah diolah oleh komputer (Rofiqi et al., 2019). Pada proses *Tokenize* penelitian ini setiap kata dalam dokumen dirubah menjadi berbentuk token atau kata – kata.

```
import nltk
nltk.download('punkt') #memanggil fungsi punkt
from nltk.tokenize import sent_tokenize, word_tokenize
data_clean['text_tokens'] = data_clean['text_StopWord'].apply(lambda x: word_tokenize(x))
data_clean.head()
```

Gbr. 7 Kode Program Tokenisasi

Berikut adalah tabel dari hasil dari *Tokenize* yaitu :

TABEL IVV
SEBAGIAN HASIL DATA *FILTER STOPWORDS*

Sebelum	Sesudah	Kategori
suasana kampanye gbk prabowogibran kemenangan	['suasana', 'kampanye', 'gbk', 'prabowogibran', 'kemenangan']	<i>Positif</i>
pemberangkatan distribusi kotak bilik suara pemilu kapolres madiun akbp m ridwan sik bersama forkopimda kab madiun kpu melaksanakan pemberangkatan distribusi kotak bilik suara pemilu	['pemberangkatan', 'distribusi', 'kotak', 'bilik', 'suara', 'pemilu', 'kapolres', 'madiun', 'akbp', 'm', 'ridwan', 'sik', 'forkopimda', 'kab', 'madiun', 'kpu', 'melaksanakan', 'pemberangkatan', 'distribusi', 'kotak', 'bilik', 'suara', 'pemilu', 'bertempat', 'gedung', 'korpi', 'alun']	<i>Netral</i>
fakta lapangan kampanye inilah hasil survey fakta yg membuat jokowi	['fakta', 'lapangan',	<i>Negatif</i>

```
#Install modul nltk
!pip install nltk

#Import library nltk corpus
import nltk.corpus
nltk.download('stopwords')
from nltk.corpus import stopwords
stop = stopwords.words('indonesian')
data_clean['text_StopWord'] = data_clean['text_clean'].apply(lambda x: '.join([word for word in x.split() if word not in (stop)])')
data_clean.head(10)
```

Gbr. 6 Kode Program *Filter Stopwords*

Berikut adalah tabel dari hasil dari *Filter stopwords* yaitu :

TABEL IIIII
SEBAGIAN HASIL DATA *FILTER STOPWORDS*

Sebelum	Sesudah	Kategori
suasana kampanye di gbk prabowogibran menuju kemenangan	suasana kampanye gbk prabowogibran kemenangan	<i>Positif</i>
pemberangkatan distribusi kotak bilik suara pemilu tahun kapolres madiun akbp m ridwan sik bersama forkopimda kab madiun dan kpu melaksanakan pemberangkatan distribusi kotak bilik suara pemilu tahun bertempat di gedung korpi alun	pemberangkatan distribusi kotak bilik suara pemilu kapolres madiun akbp m ridwan sik forkopimda kab madiun kpu melaksanakan pemberangkatan distribusi kotak bilik suara pemilu	<i>Netral</i>
fakta lapangan kampanye inilah hasil survey fakta yg membuat jokowi	fakta lapangan kampanye hasil survey fakta yg jokowi turun gunung	<i>Negatif</i>

<p>kampanye hasil survey fakta yg jokowi turun gunung prabowo marah digelontorkan dana sepi pengunjung sepi yg bikin tetangga sebelah panik gusar bingung</p>	<p>'kampanye', 'hasil', 'survey', 'fakta', 'yg', 'jokowi', 'turun', 'gunung', 'prabowo', 'marah', 'digelontorkan', 'dana', 'sepi', 'pengunjung', 'sepi', 'yg', 'bikin', 'tetangga', 'sebelah', 'panik', 'gusar', 'bingung']</p>	<pre>print("-----") def get_stemmed_term(document): return [term_dict[term] for term in document] #memisahkan file eksekusinya setelah pembacaan term selesai data_clean['text_steamindo'] = data_clean['text_tokens'].apply(lambda x: ' '.join(get_stemmed_term(x))) data_clean.head(20)</pre>
---	---	---

Tabel IV menunjukkan perbedaan sebelum dan sesudah dilakukan proses Tokenize di mana hasilnya membuat teks menjadi kata-kata atau frasa-frasa, yang lebih mudah diolah oleh komputer

4. Stemming

Kemudian selanjutnya dilakukan proses *Stemming*. *Stemming* adalah teknik yang digunakan untuk mengekstrak bentuk dasar kata dengan menghilangkan imbuhan dari kata tersebut.

```
from Sastrawi.Stemmer.StemmerFactory
import StemmerFactory
factory = StemmerFactory()
stemmer = factory.create_stemmer()

#-----STEMMING -----
---
from Sastrawi.Stemmer.StemmerFactory
import StemmerFactory

# buat stemmer
factory = StemmerFactory()
stemmer = factory.create_stemmer()

# proses stemming
def stemmed_wrapper(term):
    return stemmer.stem(term)

term_dict = {}
hitung=0

for document in data_clean['text_tokens']:
    for term in document:
        if term not in term_dict:
            term_dict[term] = ' '
print(len(term_dict))
print("-----")
for term in term_dict:
    term_dict[term] = stemmed_wrapper(term)
    hitung+=1
    print(hitung, ":", term, ":", term_dict[term])

print(term dict)
```

Gbr. 8 Kode Program Stem

Berikut adalah tabel dari hasil dari *Stemming* yaitu :

TABEL V
SEBAGIAN HASIL DATA STEM

Sebelum	Sesudah	Kategori
['suasana', 'kampanye', 'gbk', 'prabowogibran', 'kemenangan']	suasana kampanye gbk prabowogibran menang	<i>Positif</i>
['pemberangkatan', 'distribusi', 'kotak', 'bilik', 'suara', 'pemilu', 'kapolres', 'madiun', 'akbp', 'm', 'ridwan', 'sik', 'forkopimda', 'kab', 'madiun', 'kpu', 'melaksanakan', 'pemberangkatan', 'distribusi', 'kotak', 'bilik', 'suara', 'pemilu', 'bertempat', 'gedung', 'korpi', 'alun']	berangkat distribusi kotak bilik suara milu kapolres madiun akbp m ridwan sik forkopimda kab madiun kpu laksana berangkat distribusi kotak bilik suara milu	<i>Netral</i>
['fakta', 'lapangan', 'kampanye', 'hasil', 'survey', 'fakta', 'yg', 'jokowi', 'turun', 'gunung', 'prabowo', 'marah', 'digelontorkan', 'dana', 'sepi', 'pengunjung', 'sepi', 'yg', 'bikin', 'tetangga', 'sebelah', 'panik', 'gusar', 'bingung']	fakta lapang kampanye hasil survey fakta yg jokowi turun gunung prabowo marah gelontor dana sepi unjung sepi yg bikin tetangga belah panik gusar bingung	<i>Negatif</i>

Tabel 4.5 menunjukkan perbedaan sebelum dan sesudah dilakukan proses *Stem* di mana hasilnya membuat kata kata yang bukan dasar menjadi kata dasar seperti skrg menjadi sekarang, blk menjadi balik dan lain – lain.

5. Filter token by length

Lalu proses terakhir pada proses *Preprocessing* yaitu *Filter tokens by length* mengacu pada proses

menghilangkan token (kata-kata) yang memiliki panjang karakter di luar batas tertentu.

```
# Fungsi untuk mendapatkan token
berdasarkan panjangnya
def filter_tokens_by_length(text,
min_length=4, max_length=8):
    tokens = text.split()
    filtered_tokens = [token for token in
tokens if min_length <= len(token) <=
max_length]
    return ' '.join(filtered_tokens)

# Menambahkan kolom baru dengan token
berdasarkan panjangnya
data_clean['text_filtered_tokens'] =
data_clean['text_steamindo'].apply(filter
_tokens_by_length)

# Menampilkan hasil 20 baris pertama
data_clean.head(20)
```

Gbr. 9 Kode Program Filter tokens by length

Berikut adalah beberapa hasil dari proses *Filter tokens by length* dengan parameter huruf 4-15 huruf per kata :

TABEL VI
 SEBAGIAN HASIL DATA FILTER TOKENS BY LENGTH

Sebelum	Sesudah	Kategori
suasana kampanye gbk prabowogibran menang	suasana kampanye prabowogibran menang	Positif
berangkat distribusi kotak bilik suara milu kapolres madiun akbp m ridwan sik forkopimda kab madiun kpu laksana berangkat distribusi kotak bilik suara milu	berangkat distribusi kotak bilik suara milu kapolres madiun akbp ridwan forkopimda madiun laksana berangkat distribusi kotak bilik suara milu	Netral
fakta lapang kampanye hasil survey fakta yg jokowi turun gunung prabowo marah gelontor dana sepi unjung sepi yg bikin tetangga belah panik gusar bingung	fakta lapang kampanye hasil survey fakta jokowi turun gunung prabowo marah gelontor dana sepi unjung sepi bikin tetangga belah panik gusar bingung	Negatif

Tabel VI menunjukkan perbedaan sebelum dan sesudah dilakukan proses *Filter tokens by length* di mana hasilnya membuat kata – kata yang jumlah hurufnya kurang dari 4 akan dihapus.

D. Pemodelan

Pada proses implementasi penelitian ini terdapat tiga proses utama yaitu pertama adalah proses pembobotan TF-IDF. Peneliti menggunakan proses ini untuk menghitung bobot kata dalam sebuah teks berdasarkan seberapa sering kata tersebut muncul dalam dokumen dan frekuensi kemunculan kata tersebut di semua dokumen. Proses visualisasi data kemudian dilakukan dan hasilnya ditampilkan dalam bentuk grafik, bagan, atau diagram, memberikan wawasan yang lebih intuitif dan mudah dipahami mengenai distribusi sentimen. Terakhir, proses klasifikasi dilakukan dengan menggunakan algoritma *Naive Bayes*, *Support Vector Machine*, dan *Random Forest*. Teks yang diproses dan diberi bobot digunakan untuk melatih model klasifikasi yang dapat memprediksi sentimen berdasarkan fitur-fitur teks. Berikut adalah penjelasan lebih lanjut mengenai ketiga proses tersebut:

1. Pembobotan TF-IDF

Pada langkah pertama proses implementasi yaitu dilakukan terlebih dahulu pembobotan TF-IDF untuk menentukan bobot kata pada dokumen. Berikut kode programnya:

```
from sklearn.feature_extraction.text
import TfidfVectorizer
tfidfconverter =
TfidfVectorizer(max_features=2000,
min_df=5, max_df=0.7,
stop_words=stopwords.words('english'), ngr
am_range=(1,3))
X1 =
tfidfconverter.fit_transform(processed_tw
eets).toarray()
```

Gbr. 10 Kode Program TF-IDF

Proses TF-IDF pada data pemilu 2024 melibatkan beberapa langkah. Pertama, menggunakan *TfidfVectorizer* dari *scikit-learn* untuk mengonversi teks menjadi matriks TF-IDF. Setelah menginisialisasi *TfidfVectorizer* maka metode *fit_transform* dipanggil untuk menyesuaikan model dan mentransformasikan data teks menjadi matriks TF-IDF.

2. Data Visualization

Setelah tahapan-tahapan sebelumnya selesai dilakukan langkah selanjutnya adalah melakukan visualisasi hasil data dengan wordcloud. Berikut kode programnya:

```
from wordcloud import WordCloud
import matplotlib.pyplot as plt

# Menggabungkan semua teks hasil TF-IDF
menjadi satu string
text_combined = '
'.join(processed_tweets)

# Membuat WordCloud
```



```

accuracy_score, recall_score,
precision_score, f1_score, roc_auc_score

# Random Forest
print("Random Forest")
print('Accuracy = ',
round(accuracy_score(y_test,
predictions_en)*100, 2), '%')
print('Recall = ',
round(recall_score(y_test,
predictions_en, average='weighted')*100,
2), '%')
print('Precision = ',
round(precision_score(y_test,
predictions_en, average='weighted')*100,
2), '%')
print("Confusion Matrix:")
print(confusion_matrix(y_test,
predictions_en))
print("")
print("Classification Report:")
print(classification_report(y_test,
predictions_en, target_names=['positif',
'negatif', 'netral']))
print("")

# Naive Bayes
print("Naive Bayes")
print('Accuracy = ',
round(accuracy_score(y_test,
predictions_nb)*100, 2), '%')
print('Recall = ',
round(recall_score(y_test,
predictions_nb, average='weighted')*100,
2), '%')
print('Precision = ',
round(precision_score(y_test,
predictions_nb, average='weighted')*100,
2), '%')
print('F1-Score = ',
round(f1_score(y_test, predictions_nb,
average='weighted')*100, 2), '%')
print("Confusion Matrix:")
print(confusion_matrix(y_test,
predictions_nb))
print("")
print("Classification Report:")
print(classification_report(y_test,
predictions_nb, target_names=['positif',
'negatif', 'netral']))
print("")

# Support Vector Machine
print("Support Vector Machine")
print('Accuracy = ',
round(accuracy_score(y_test,
predictions_svm)*100, 2), '%')
    
```

```

print('Recall = ',
round(recall_score(y_test,
predictions_svm, average='weighted')*100,
2), '%')
print('Precision = ',
round(precision_score(y_test,
predictions_svm, average='weighted')*100,
2), '%')
print('F1-Score = ',
round(f1_score(y_test, predictions_svm,
average='weighted')*100, 2), '%')
print("Confusion Matrix:")
print(confusion_matrix(y_test,
predictions_svm))
print("")
print("Classification Report:")
print(classification_report(y_test,
predictions_svm, target_names=['positif',
'negatif', 'netral']))
    
```

Gbr. 15 Kode Program Pemodelan

Berikut adalah hasil pengujian berdasarkan pembagian data uji dan latihan:

TABEL VII
 PERBANDINGAN EVALUASI BERDASARKAN PERBANDINGAN RASIO

	Random Forest	SVM	Naive Bayes
Perbandingan	Akurasi		
7 : 3	73.14 %	74.79 %	60.74 %
8 : 2	61.11 %	72.22 %	54.32 %
9 : 1	65.43 %	61.73 %	48.15 %

TABEL VIII
 PERBANDINGAN EVALUASI BERDASARKAN PERBANDINGAN RASIO

	Random Forest	SVM	Naive Bayes
Perbandingan	Recall		
7 : 3	73.14 %	74.79 %	60.74 %
8 : 2	61.11 %	72.22 %	54.32 %
9 : 1	65.43 %	61.73 %	48.15 %

TABEL IX
 PERBANDINGAN EVALUASI BERDASARKAN PERBANDINGAN RASIO

	Random Forest	SVM	Naive Bayes
Perbandingan	Precision		
7 : 3	74.54 %	76.04 %	65.46 %
8 : 2	63.7 %	73.41 %	57.58 %
9 : 1	67.2 %	63.7 %	57.47 %

Dalam perbandingan evaluasi berdasarkan rasio data pelatihan dan pengujian tiga algoritma klasifikasi yaitu *Random Forest*, *Support Vector Machine* (SVM), dan *Naive Bayes* telah dievaluasi untuk menilai kinerja algoritma tersebut dalam memprediksi sentimen pemilu 2024. *Support*

Vector Machine (SVM) memperoleh hasil terbaik dengan akurasi sebesar 74.79%, recall 74.79%, dan *precision* 76.04%. Hal ini menunjukkan bahwa SVM memiliki kemampuan yang lebih baik dalam mengidentifikasi dan mengklasifikasikan sentimen pada skenario ini dibandingkan dengan *Random Forest* dan *Naïve Bayes*. Meskipun *Random Forest* juga memberikan hasil yang cukup baik terlihat bahwa SVM memiliki keunggulan dalam memberikan prediksi yang lebih akurat dan konsisten berdasarkan nilai akurasi.

IV. KESIMPULAN

Penelitian ini berhasil mengatasi tantangan dalam mengembangkan algoritma analisis sentimen yang efektif dan akurat untuk mengevaluasi opini pengguna di platform media sosial Twitter, khususnya terkait pemilu 2024 di Indonesia. Dengan langkah-langkah yang komprehensif, mulai dari pengumpulan data hingga evaluasi model, penelitian ini membuktikan bahwa pendekatan yang terintegrasi dan berbasis pada pemrosesan teks yang cermat dapat meningkatkan kualitas analisis sentimen.

Pertama-tama, langkah pengumpulan data yang menggunakan bahasa pemrograman *Python* dan kata kunci yang relevan telah memastikan bahwa dataset yang digunakan mencakup informasi terbaru dan terkait dengan topik yang diteliti. Hal ini penting untuk memastikan representativitas data yang dianalisis. Selanjutnya, proses pelabelan dataset secara manual membantu memvalidasi dan mengklasifikasikan data dengan lebih akurat, serta mengatasi masalah *cross labelling* dengan menerapkan aturan khusus.

Selanjutnya, dengan menggunakan teknik preprocessing yang komprehensif, termasuk *cleansing*, *transform case*, *filter stopword*, *tokenisasi*, *stemming*, dan *filter token by length*, data mentah berhasil diubah menjadi format yang lebih terstruktur dan siap digunakan untuk analisis lebih lanjut. Proses ini membantu menghilangkan *noise* dan meningkatkan kualitas data, sehingga model klasifikasi dapat menghasilkan prediksi yang lebih akurat.

Pada tahap pemodelan, penelitian ini mengimplementasikan tiga algoritma klasifikasi yang berbeda: *Naïve Bayes*, *Random Forest*, dan *Support Vector Machine* (SVM). Hasil evaluasi menunjukkan bahwa SVM memiliki kinerja terbaik dengan akurasi, *recall*, dan *precision* yang lebih tinggi dibandingkan dengan dua algoritma lainnya. Ini menunjukkan bahwa SVM adalah pilihan yang lebih efektif dalam mengidentifikasi dan mengklasifikasikan sentimen pada data tweet terkait pemilu 2024 di Twitter.

Dengan demikian, pendekatan yang digunakan dalam penelitian ini tidak hanya meningkatkan akurasi analisis sentimen, tetapi juga efektivitasnya dalam mengevaluasi opini pengguna di platform media sosial Twitter. Dengan pemahaman yang lebih mendalam tentang sentimen masyarakat, informasi yang diperoleh dari analisis ini dapat memberikan wawasan yang berharga bagi para pengambil keputusan dan stakeholders terkait pemilu 2024 di Indonesia.

V. SARAN

Berdasarkan hasil penelitian ini, beberapa saran yang dapat diberikan untuk pengembangan lebih lanjut adalah sebagai berikut:

1. Melakukan penelitian lanjutan untuk mengeksplorasi teknik-teknik pemrosesan teks yang lebih canggih dan algoritma-algoritma klasifikasi yang lebih kompleks. Hal ini dapat membantu meningkatkan kinerja dan akurasi analisis sentimen, terutama ketika menghadapi data yang lebih kompleks dan besar.
2. Memperluas cakupan data dengan mengumpulkan informasi dari berbagai sumber sosial media selain Twitter, seperti Facebook, Instagram, dan YouTube. Hal ini dapat memberikan gambaran yang lebih lengkap tentang sentimen masyarakat terkait topik tertentu.
3. Melakukan validasi eksternal terhadap model yang dikembangkan dengan menguji kinerjanya pada dataset yang berbeda atau di lingkungan yang berbeda. Hal ini penting untuk memastikan generalitas dan keandalan model dalam berbagai konteks.
4. Lebih mendalam dalam interpretasi hasil analisis sentimen dengan mempertimbangkan konteks sosial, politik, dan budaya yang relevan. Ini dapat membantu menghindari kesimpulan yang keliru atau bias yang mungkin timbul dari analisis secara teknis saja.
5. Menerapkan hasil penelitian ini dalam praktik nyata, misalnya dengan membangun sistem monitoring sentimen secara real-time untuk pemilu atau kegiatan politik lainnya. Ini dapat memberikan manfaat langsung bagi pemangku kepentingan dan masyarakat umum.

VI. UCAPAN TERIMA KASIH

Peneliti senantiasa dengan tulus berterima kasih kepada Tuhan Yang Maha Esa atas berkah dan bimbinganNya, yang memungkinkan peneliti menyelesaikan proyek dan artikel ilmiah ini dengan baik. Penulis juga mengucapkan terima kasih kepada Dosen Pembimbing yang memberikan arahan dan masukan berharga. Orang tua peneliti yang selalu mendoakan dan membantu juga patut disyukuri. Dukungan dari saudara-saudara di Jurusan Teknik Informatika, khususnya kelas TI 2017 B, sangat membantu peneliti menyelesaikan skripsi ini.

VII. REFERENSI

- [1]. Husada, H. C., & Paramita, A. S. (2021). Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM). *Teknika*, 10(1), 18–26. <https://doi.org/10.34148/teknika.v10i1.311>
- [2]. Corallo, A. et al. 2015. Sentiment Analysis for Government: An Optimized Approach, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9166(July), pp. 98– 112
- [3]. Razaq, E. R. M., Jacob, D. W., & Hamami, F. (2021). Analisis Sentimen Kepuasan Mahasiswa Terhadap Pembelajaran Online Selama Pandemi Covid-19 Pada Media Sosial Twitter Menggunakan Perbandingan Algoritma Klasifikasi. *E-Proceeding of Engineering*, 8(5), 9000.
- [4]. Fadiyah, B. T., Ratnawati, D. E., & Arwani, I. (2022). Analisis Sentimen Pengguna Twitter terhadap Pembayaran Cashless menggunakan ShopeePay dengan Algoritma Random Forest. *Jurnal*

- Pengembangan Teknologi Informasi Dan Ilmu Komputer, 6(3), 1426–1433. <http://j-ptiik.ub.ac.id>
- [5]. Febriyani, E., & Februriyanti, H. (2022). Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Naive Bayes Di Twitter. *Jurnal TEKNO KOMPAK*, 17(2), 25–38.
- [6]. Irawan, F. R., Jazuli, A., & Khotimah, T. (2022). Analisis Sentimen Terhadap Pengguna Gojek Menggunakan Metode K-Nearest Neighbors Sentiment Analysis of Gojek Users Using K-Nearest Neighbor. *JIKO (Jurnal Informatika Dan Komputer)*, 5(1), 62–68. <https://doi.org/10.33387/jiko>
- [7]. Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi. *Jurnal Teknoinfo*, 14(2), 115. <https://doi.org/10.33365/jti.v14i2.679>
- [8]. Hudha, M., Supriyati, E., & Listyorini, T. (2022). Analisis Sentimen Pengguna Youtube Terhadap Tayangan #Matanajwamenantiterawan Dengan Metode Naïve Bayes Classifier. *JIKO (Jurnal Informatika Dan Komputer)*, 5(1), 1–6. <https://doi.org/10.33387/jiko.v5i1.3376>
- [9]. Sandryan, M. K., Rahayudi, B., & Ratnawati, D. E. (2021). Analisis Sentimen Pada Media Sosial Twitter Terhadap Undang-Undang Cipta Kerja Menggunakan Algoritma Backpropagation dan Term Frequency-Inverse Document Frequency. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 5(12), 5349–5355.
- [10]. Nuraliza, H., Pratiwi, O. N., & Hamami, F. (2022). Analisis Sentimen IMDb Film Review Dataset Menggunakan Support Vector Machine (SVM) dan Seleksi Feature Importance. *Jurnal Mirai Manajemen*, 7(1), 1–17.
- [11]. Arsi, P., & Waluyo, R. (2021). Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 8(1), 147. <https://doi.org/10.25126/jtiik.0813944>
- [12]. Harpizon, H. A. R., Kurniawan, R., Iwan Iskandar, Salambue, R., Budianita, E., & Syafria, F. (2022). Analisis Sentimen Komentar Di YouTube Tentang Ceramah Ustadz Abdul Somad Menggunakan Algoritma Naïve Bayes. *JNKTI (Jurnal Nasional Komputasi Dan Teknologi Informasi)*, 5(1), 131–140. <http://repository.uin-suska.ac.id/59746/>
- [13]. Hasiholan, A., Cholissodin, I., & Yudistira, N. (2022). Analisis Sentimen Tweet Covid-19 Varian Omicron pada Platform Media Sosial Twitter menggunakan Metode LSTM berbasis Multi Fungsi Aktivasi dan GLOVE. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 6(10), 4653–4661. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/11648>
- [14]. Hakim, M. L. (2023). Terhadap Pengesahan Rkuhp Pada Twitter Menggunakan Algoritma Support Vector Machine (SVM) Dan Decision Tree Dengan Synthetic Minority Oversampling Technique (SMOTE) Skripsi.