

Analisis Sentimen Komentar Pengguna Instagram Mengenai Pelaksanaan Pemilu 2024 dengan Naïve Bayes dan Lexicon-Based

Cahyani Rahma Dewi, Agus Iskandar*

Fakultas Teknologi Komunikasi dan Informatika, Sistem Informasi, Universitas Nasional, Jakarta, Indonesia

Email: ¹cahyaniarahmadewi@gmail.com, ^{2,*}agus.iskandar@civitas.unas.ac.id

Email Penulis Korespondensi: agus.iskandar@civitas.unas.ac.id

Submitted: 14/08/2024; Accepted: 30/11/2024; Published: 30/11/2024

Abstrak—Perdebatan seputar pelaksanaan Pemilihan Umum 2024 yang menjadi sorotan utama di Indonesia, terutama dalam platform media sosial yang diminati oleh masyarakat. Perubahan pemimpin di Indonesia dan perbedaan emosional yang muncul dalam masyarakat menjadi perhatian penting. Pencarian figur kepemimpinan memunculkan berbagai perspektif teoritis, konseptual, dan budaya yang kompleks. Penulisan ini bertujuan untuk menganalisis sentiment Masyarakat terkait dengan pelaksanaan pemilihan umum 2024 dengan mengklasifikasikan sentimen sebagai positif, negatif, atau netral, membantu pemahaman tentang persepsi masyarakat terhadap para kandidat, isu-isu politik yang relevan, serta pola perilaku pemilih. Metodologi yang digunakan melibatkan pengumpulan data dengan teknik *Scraping* dari platform media sosial, yaitu Instagram dengan menggunakan kombinasi dari kedua algoritma, yaitu Naïve Bayes Classifier dan pelabelan Lexicon-Based. Kedua metode ini digunakan untuk melakukan analisis sentimen terhadap pemilihan umum dalam penelitian ini. Analisis sentimen terhadap Pemilu 2024 menggunakan model Naive Bayes dan InSet Lexicon menunjukkan hasil yang baik dengan akurasi 72% (presisi negatif 74%, netral 54%, positif 70%; recall positif 62%, netral 22%, negatif 87%). Penelitian ini berhasil melampaui akurasi model sebelumnya (72% akurasi, 70% presisi, 72% recall) dan mengungkapkan bahwa sentimen negatif lebih banyak ditemukan dalam opini masyarakat terhadap Pemilu 2024. Hal ini mengindikasikan adanya ketidakpuasan dan keraguan masyarakat terkait pelaksanaan Pemilu, yang diduga dipicu oleh masalah teknis dan ketidakpastian politik.

Kata Kunci: Analisis Sentimen; Pemilihan Umum; Media Sosial; Naïve Bayes; Lexicon-Based

Abstract—The debate surrounding the implementation of the 2024 General Election has taken centre stage in Indonesia, especially on social media platforms that are favoured by the public. The change of leaders in Indonesia and the emotional differences that emerge in society are of significant concern. The search for leadership figures brings up various complex theoretical, conceptual, and cultural perspectives. This paper aims to analyse people's sentiment related to the 2024 general election by classifying sentiment as positive, negative, or neutral, aiding understanding of people's perceptions of candidates, relevant political issues, and voter behaviour patterns. The methodology involved collecting data using scraping techniques from the social media platform Instagram using a combination of both Naïve Bayes Classifier and Lexicon-Based labelling algorithms. These two methods were used to conduct sentiment analysis towards the general election in this study. Sentiment analysis of the 2024 General Election using the Naive Bayes and InSet Lexicon models showed good results with an accuracy of 72% (precision negative 74%, neutral 54%, positive 70%; recall positive 62%, neutral 22%, negative 87%). This study successfully surpassed the accuracy of the previous model (72% accuracy, 70% precision, 72% recall) and revealed that negative sentiments were more prevalent in public opinion towards the 2024 General Election. This indicates that there is public dissatisfaction and doubt regarding the implementation of the election, which is thought to be triggered by technical problems and political uncertainty.

Keywords: Sentiment Analysis; General Election; Social Media; Naïve Bayes; Lexicon-Based

1. PENDAHULUAN

Perdebatan terkait Pemilihan Umum 2024 telah menjadi topik hangat di kalangan masyarakat Indonesia dan mendapatkan perhatian luas di berbagai platform media sosial. Media sosial, sebagai salah satu sumber informasi yang paling diminati, memainkan peran penting dalam membahas dinamika perubahan kepemimpinan di Indonesia[1]. Ketertarikan masyarakat terhadap sosok calon pemimpin sering kali menciptakan berbagai perspektif, yang membuat topik ini semakin relevan dan diperhatikan secara serius oleh berbagai segmen masyarakat. Ketidakpastian dan perbedaan emosional yang mencolok di masyarakat Indonesia menambah kompleksitas diskusi ini[2].

Media sosial menyediakan saluran informasi yang penting, terutama selama masa pemilihan umum, dengan menawarkan berbagai perspektif dan berita terbaru kepada public[3]. Komisi Pemilihan Umum (KPU) Indonesia, sebagai lembaga yang bertanggung jawab atas penyelenggaraan pemilu[4], aktif dalam menyebarkan informasi melalui akun Instagram resmi mereka (@KPU_RI) yang memiliki 624.000 pengikut dan 7.861 postingan. Akun ini secara rutin mengunggah informasi terbaru, dan pembaruan yang sering kali menarik perhatian banyak komentar dari masyarakat[5]. Komentar-komentar ini, baik yang positif, negatif, maupun netral, mencerminkan tingkat antusiasme masyarakat terhadap berita pemilu[6].

Analisis sentimen terhadap komentar-komentar ini melibatkan proses sistematis yang mencakup pengumpulan, pemrosesan, dan analisis opini serta sikap yang tercermin dari berbagai sumber, seperti media sosial dan komentar publik terkait kandidat, partai politik, serta isu kampanye[2]. Teknik-teknik analisis teks dan sentiment analysis digunakan untuk mengidentifikasi dan mengklasifikasikan sentimen sebagai positif, negatif,

atau netral[7]. Pendekatan ini membantu memahami persepsi masyarakat terhadap para kandidat dan isu-isu politik, serta pola perilaku pemilih[3]. Informasi yang diperoleh dari analisis sentimen ini memberikan pandangan mendalam yang dapat digunakan untuk strategi kampanye, respons terhadap opini publik, dan pemahaman lebih baik mengenai persepsi masyarakat terhadap pemilihan umum[8].

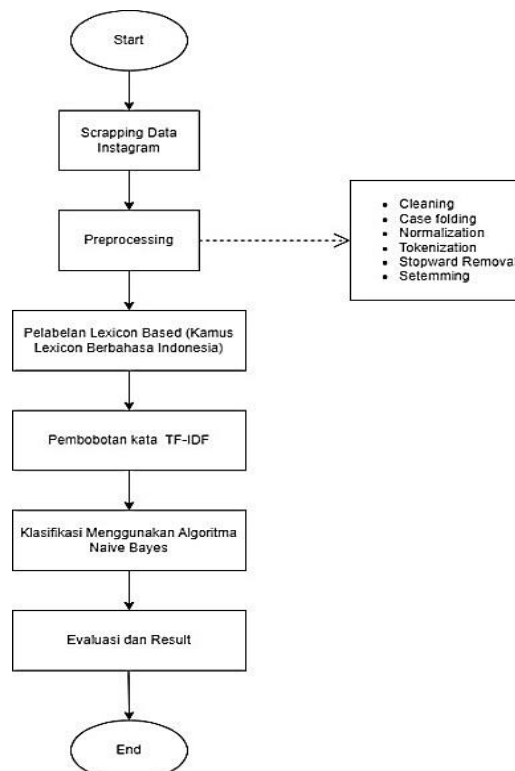
Dalam konteks ini, metode Naive Bayes Classifier dan pelabelan berbasis leksikon merupakan dua pendekatan utama yang dapat digunakan. Naive Bayes Classifier, sebagai metode pembelajaran mesin, memanfaatkan probabilitas dan statistik untuk mengkategorikan teks ke dalam sentimen yang berbeda[9]. Metode ini dikenal karena akurasi yang baik dalam klasifikasi teks, seperti yang dibuktikan oleh penelitian sebelumnya, termasuk studi oleh[5], yang melaporkan akurasi sebesar 73% dalam mengklasifikasikan tweet menggunakan metode Boyer-Moore. Di sisi lain, metode pelabelan berbasis leksikon melibatkan penggunaan daftar kata-kata yang telah diberi nilai sentimen untuk menilai sentimen dalam teks. Metode ini bekerja dengan mencocokkan kata-kata dalam teks dengan entri di dalam leksikon dan menjumlahkan nilai-nilai sentimen dari kata-kata tersebut[10][11]. Penelitian sejenis lainnya juga dilakukan pada tahun 2025 dengan tujuan mengetahui bagaimana sentimen masyarakat terhadap Pekan Olahraga Nasional (PON) XII di Aceh-Sumatera Utara. Data dari media sosial X diolah menggunakan algoritma Multinomial Naive Bayes setelah melalui serangkaian tahap pre-processing seperti pembersihan data, normalisasi kata, dan penghapusan stopword. Hasil akhir dari penelitian ini menunjukkan bahwa algoritma Multinomial Naive Bayes mampu menganalisis sentimen masyarakat dengan akurasi 79,77% pada data uji. Ini mengindikasikan bahwa sebagian besar masyarakat memiliki sentimen positif terhadap penyelenggaraan PON XII[12].

Dalam penelitian ini, kedua pendekatan tersebut diterapkan secara terpisah maupun bersamaan untuk menganalisis sentimen masyarakat Indonesia menjelang Pemilihan Umum 2024. Penggunaan kombinasi algoritma Naive Bayes Classifier dan pelabelan berbasis leksikon bertujuan untuk memberikan pemahaman yang lebih komprehensif tentang sentimen publik yang tercermin dalam komentar di platform media sosial, khususnya Instagram .

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini bertujuan untuk mengeksplorasi analisis sentiment pengguna Instagram terhadap pelaksanaan pemilu 2024 dengan pendekatan yang sistematis dan terstruktur, untuk menilai keberhasilan dan mendapatkan laporan akurat melalui penggunaan labelling data berbasis Bahasa Indonesia yang diharapkan dapat meningkatkan nilai akurasi hasil analisis. Untuk mencapai tujuan tersebut, penelitian ini mengikuti alur Gambar 1 sebagai berikut:



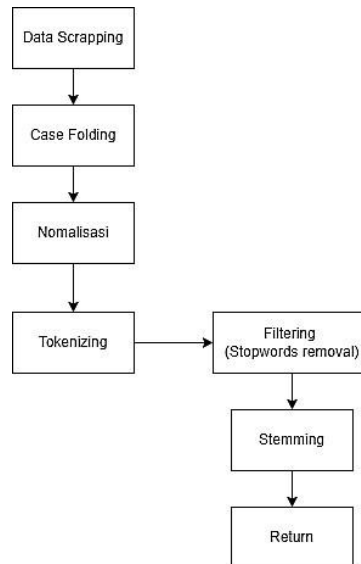
Gambar 1. Alur Penelitian

1. *Scrapping Data Instagram*

Scraping data ini diambil dari komentar pengguna Instagram dalam postingan akun KPU menggunakan *website scraper*. Data yang diambil merupakan data murni komentar dari Masyarakat pengguna Instagram dengan jumlah data yang diperoleh sebanyak kurang lebih 980 data.

2. *Preprocessing*

Pre-processing data merupakan tahapan yang diawali setelah pengambilan dataset atau *web scraping* dan pada tahap ini dilanjutkan dengan beberapa tahapan yaitu, Casefolding, Normalisasi, Tokenizing, Filtering (Stop word Removal), dan Stemming[13][14]. Berikut Gambar 1 menyajikan tahapan *pre-processing* yang dilakukan.



Gambar 2. Tahapan *Pre-Processing*

a. *Case Folding*

Pada tahapan ini data mengubah kalimat menjadi huruf kecil, memeriksa apakah kalimat adalah *string*, menghapus angka dari kalimat, menghapus spasi awal dan akhir kalimat, menghapus https dan http, menghilangkan tanda baca, mengganti karakter HTML dengan tanda petik, mengganti line baru dengan spasi, menghapus single baru, memisahkan dan menggabungkan kata, serta membersihkan nilai Na Ndari Data Frame. Kemudian data disimpan dalam nama column “text_clean”[8]. Berikut Tabel 1 yang berisi perbandingan sebelum dan sesudah dilakukan *Casefolding*.

Tabel 1. *Casefolding*

Sebelum	Sesudah
Sirekap gimana min, eror terus	sirekap gimana min eror terus
Min, gimana nih, istri sy sudah dari 2014 KTP bantu jogja, koq TPSnya daerah asal?Sudah DM koq tidak respon?#kpumelayani#pemiluserentak2024	min gimana nih istri sy sudah dari ktp bantu jogja kok tpsnya daerah asalsudah dm kok tidak respon kpu melayani pemilus erentak
Min masih ada si rekap yang belum dapat notif, pengaduannya kemana ya?	min masih ada si rekap yang belum dapat notif pengaduannya kemana ya
Emang gk ada undangan nya nih? Apa gk pake undangan tahun ini?	emang gk ada undangan nya nih apa gk pake undangan tahun ini
min sirekap nyaa tulung	min sirekap nyaa tulung
Yuk adek adek siswa siswi SMU / SMK/ sederajat yang sudah memenuhi syarat untuk datang ke TPS, gunakan Hak Pilihmu	yuk adek adek siswa siswi smu smk sederajat yang sudah memenuhi syarat untuk datang ke tps gunakan hak pilihmu

b. *Normalisasi*

Normalization, perbaikan kata baku dengan berpacu pada menggunakan kamus atau metode lainnya untuk melakukan perbaikan pada kata-kata yang tidak formal atau memiliki kesalahan ejaan[10]. Normalisasi kita menggunakan kamusalay dan di dalam kamus alay terdapat kata kata yang alay/gaul/singkatan dengan cara menormalkan kata-kata yang alay/gaul/singkatan. Lalu kemudian disimpan dengan nama column “normal”. Berikut Tabel 2 berisi perbandingan sebelum dan sesudah dilakukan normalisasi.

Tabel 2. Normalisasi

Sebelum	Sesudah
sirekap gimana min eror terus	sirekap gimana admin error terus
min gimana nih istri sy sudah dari ktp bantu jogja	admin gimana nih istri saya sudah dari ktp bantu jogja
koq tpsnya daerah asal sudah dm koq tidak respon	kok tpsnya daerah asal sudah dm kok tidak respon
kpu melayani pemilu serentak	kpu melayani pemilu serentak
min masih ada si rekap yang belum dapat notif	admin masih ada si rekap yang belum dapat notif
pengaduannya kemana ya	pengaduannya kemana ya
emang gk ada undangan nya nih apa gk pake	emang tidak ada undangan nya nih apa tidak pake
undangan tahun ini	undangan tahun ini
min sirekap nyaa tolong	admin sirekap nyaa tolong
yuk adek adek siswa siswi smu smk sederajat yang	yuk adek adek siswa siswi smu smk sederajat yang
sudah memenuhi syarat untuk datang ke tps	sudah memenuhi syarat untuk datang ke tps
gunakan hak pilihmu	gunakan hak pilihmu

c. *Tokenizing*

Pada tahap ini, data akan diubah atau diproses dengan menghilangkan tanda baca yang tidak diperlukan[15], sehingga memunculkan kalimat yang berdiri sendiri. Berikut Tabel 3 berisi perbandingan sebelum dan sesudah *Tokenizing*.

Tabel 3. *Tokenizing*

Sebelum	Sesudah
sirekap gimana admin error terus	['sirekap', 'gimana', 'admin', 'error', 'terus']
admin gimana nih istri saya sudah dari ktp bantu jogja kok tpsnya daerah asal sudah dm kok tidak respon kpu melayani pemilu serentak	['admin', 'gimana', 'nih', 'istri', 'saya', 'sudah', 'dari', 'ktp', 'bantu', 'jogja', 'kok', 'tpsnya', 'daerah', 'asal', 'sudah', 'dm', 'kok', 'tidak', 'respon', 'kpu', 'melayani', 'pemilu', 'serentak']
admin masih ada si rekap yang belum dapat notif pengaduannya kemana ya	['admin', 'masih', 'ada', 'si', 'rekap', 'yang', 'belum', 'dapat', 'notif', 'pengaduannya', 'kemana', 'ya']
emang tidak ada undangan nya nih apa tidak pake undangan tahun ini	['emang', 'tidak', 'ada', 'undangan', 'nya', 'nih', 'apa', 'tidak', 'pake', 'undangan', 'tahun', 'ini']
admin sirekap nyaa tolong	['admin', 'sirekap', 'nyaa', 'tolong']
yuk adek adek siswa siswi smu smk sederajat yang sudah memenuhi syarat untuk datang ke tps gunakan hak pilihmu	['yuk', 'adek', 'adek', 'siswa', 'siswi', 'smu', 'smk', 'sederajat', 'yang', 'sudah', 'memenuhi', 'syarat', 'untuk', 'datang', 'ke', 'tps', 'gunakan', 'hak', 'pilihmu']

d. *Filtering*

Filtering (stop word removal), hapus kata-kata umum yang tidak memberikan makna khusus pada teks (stop words)[16], seperti "diri", "sih", "dih", "si". Berikut Tabel 4 berisi perbandingan sebelum dan sesudah *filtering*.

Tabel 4. *Filtering*

Sebelum	Sesudah
['sirekap', 'gimana', 'admin', 'error', 'terus']	['sirekap', 'gimana', 'admin', 'error', 'terus']
['admin', 'gimana', 'nih', 'istri', 'saya', 'sudah', 'dari', 'ktp', 'bantu', 'jogja', 'kok', 'tpsnya', 'daerah', 'asal', 'sudah', 'dm', 'kok', 'tidak', 'respon', 'kpu', 'melayani', 'pemilu', 'serentak']	['admin', 'gimana', 'nih', 'istri', 'ktp', 'bantu', 'jogja', 'kok', 'tpsnya', 'daerah', 'asal', 'dm', 'kok', 'respon', 'kpu', 'melayani', 'pemilu', 'serentak']
['admin', 'masih', 'ada', 'si', 'rekap', 'yang', 'belum', 'dapat', 'notif', 'pengaduannya', 'kemana', 'ya']	['admin', 'rekap', 'notif', 'pengaduannya']
['emang', 'tidak', 'ada', 'undangan', 'nya', 'nih', 'apa', 'tidak', 'pake', 'undangan', 'tahun', 'ini']	['emang', 'undangan', 'nya', 'nih', 'apa', 'pake', 'undangan', 'tahun']
['admin', 'sirekap', 'nyaa', 'tolong']	['admin', 'sirekap', 'nyaa', 'tulung']
['yuk', 'adek', 'adek', 'siswa', 'siswi', 'smu', 'smk', 'sederajat', 'yang', 'sudah', 'memenuhi', 'syarat', 'untuk', 'datang', 'ke', 'tps', 'gunakan', 'hak', 'pilihmu']	['yuk', 'adek', 'adek', 'siswa', 'siswi', 'smu', 'smk', 'sederajat', 'memenuhi', 'syarat', 'datang', 'tps', 'gunakan', 'hak', 'pilihmu']

e. *Stemming*

Dalam proses ini teks kembali ke dalam bentuk dasarnya dengan menggunakan library sastrawi[17]. Berikut Tabel 5 berisi perbandingan sebelum dan sesudah *Stemming*.

Tabel 5. Stemming

Sebelum	Sesudah
['sirekap', 'gimana', 'admin', 'error', 'terus']	sirekap gimana admin error terus
['admin', 'gimana', 'nih', 'istri', 'ktp', 'bantul', 'jogja', 'kok', 'tpsnya', 'daerah', 'asal', 'dm', 'kok', 'respon', 'kpu', 'melayani', 'pemilu', 'serentak']	admin gimana nih istri ktp 460entim jogja kok tpsnya daerah asal dm kok respon kpu layan milu serentak
['admin', 'rekap', 'notif', 'pengaduannya']	admin rekap notif adu
['emang', 'undangan', 'nya', 'nih', 'apa', 'pake', 'undangan', 'tahun']	emang undangan nya nih apa pake undang tahun
['admin', 'sirekap', 'nyaa']	admin sirekap nyaa
['yuk', 'adek', 'adek', 'siswa', 'siswi', 'smu', 'smk', 'sederajat', 'memenuhi', 'syarat', 'datang', 'tps', 'gunakan', 'hak', 'pilihmu']	yuk adek adek siswa siswi smu smk derajat penuh syarat 460entim tps guna hak pilih

3. Labelling

Pelabelan dalam penelitian menggunakan pelabelan dengan menggunakan Lexicon-Based approach berbahasa Indonesia. Dengan berpacu pada kamus negative dan kamus positive pada InSet Lexicon. InSet Lexicon adalah sentimen kata dalam sentim Indonesia yang memiliki nilai sentiment positif yang dapat dilihat pada Tbael 6 dan sentimen negatif pada Tabel 7, beserta dengan bobot atau nilai dari setiap kata tersebut . Setiap kata memiliki bobot yang berkisar antara -5 hingga +5. Nilai sentimen negatif menunjukkan bahwa kata memiliki sentiment negatif, sedangkan nilai positif menunjukkan bahwa kata memiliki sentiment positif.

Tabel 6. Kamus Lexicon Positif

Kamus	Bobot
hai	3
merekam	2
ekstensif	3
paripurna	1
detail	2

Tabel 7. Kamus Lexicon Negatif

Kamus	Bobot
putus tali gantung	-2
gelebah	-2
gobar hati	-2
tersentuh (perasaan)	-1
isak	-5

Berikut Tabel 8 menunjukkan hasil dari pelabelan dengan menggunakan *Inset Lexicon* yang sudah dikerjakan berdasarkan score-nya.

Tabel 8. Labelling Lexicon

Teks Stemming	Score	Sentimen
sirekap gimana admin error terus	-5	negative
admin gimana nih istri ktp bantul jogja kok tpsnya daerah asal dm kok respon kpu layan milu serentak	0	neutral
admin rekap notif adu	-3	negative
emang undang nih apa pake undang tahun	1	positive
admin sirekap	0	neutral
yuk adek adek siswa siswi smu smk derajat penuh syarat datang tps guna hak pilih	7	positive

4. Pembobotan TF-IDF

Langkah berikutnya setelah tahap pre-processing dan pelabelan data menggunakan lexicon-based adalah pembobotan TF-IDF. Pembobotan kata ini dilakukan secara otomatis dengan menggunakan bahasa pemrograman Python dan TfidfVectorizer dari sklearn[18], berdasarkan hasil stemming data sampel. Berikut Tabel 9 adalah dokumen TF-IDF. Berikut Tabel 9 berisi perbandingan sebelum dan sesudah pembobotan TF-IDF.

Tabel 9. Dokumen TF-IDF

Nomor Dokumen	Teks Stemming	Token Stemming
D1	sirekap gimana admin error terus	['sirekap', 'gimana', 'admin', 'error', 'terus']

Nomor Dokumen	Teks Stemming	Token Stemming
D2	admin rekap notif adu	['admin', 'rekap', 'notif', 'adu']
D3	emang undang nih apa pake undang tahun	['emang', 'undang', 'nih', 'apa', 'pake', 'undang', 'tahun']
D4	admin sirekap	['admin', 'sirekap']
D5	yuk adek adek siswa siswi smu smk derajat penuh syarat datang tps guna hak pilih	['yuk', 'adek', 'adek', 'siswa', 'siswi', 'smu', 'smk', 'derajat', 'penuh', 'syarat', 'datang', 'tps', 'guna', 'hak', 'pilih']

Pada Tabel 9. Hanya merupakan 5 sample yang diambil dari hasil pre-processing data untuk diproses dalam perhitungan manual dalam pembobotan kata TF-IDF. Kemudian dokumen akan dipecah menjadi term/perkata agar dapat di proses lebih lanjut pada Tabel 10.

Tabel 10. Perhitungan IDF

Token	D1	D2	D3	D4	D5	DF	IDF
Sirekap	1	0	0	1	0	2	$\log(5/2) = 0,3939$
Gimana	1	0	0	0	0	1	$\log(5/1) = 0,6989$
Admin	1	1	0	1	0	3	$\log(5/3) = 0,2118$
Error	1	0	0	0	0	1	$\log(5/1) = 0,6989$
Terus	1	0	0	0	0	1	$\log(5/1) = 0,6989$
Rekap	0	1	0	0	0	1	$\log(5/1) = 0,6989$
Notif	0	1	0	0	0	1	$\log(5/1) = 0,6989$
Adu	0	1	0	0	0	1	$\log(5/1) = 0,6989$
Yuk	0	0	0	0	1	1	$\log(5/1) = 0,6989$
Emang	0	0	1	0	0	1	$\log(5/1) = 0,6989$
Undang	0	0	2	0	0	2	$\log(5/2) = 0,3939$
Apa	0	0	1	0	0	1	$\log(5/1) = 0,6989$
Adek	0	0	0	0	2	2	$\log(5/2) = 0,3939$
Siswa	0	0	0	0	1	1	$\log(5/1) = 0,6989$
Pakai	0	0	1	0	0	1	$\log(5/1) = 0,6989$
Tahun	0	0	1	0	0	1	$\log(5/1) = 0,6989$
Siswi	0	0	0	0	1	1	$\log(5/1) = 0,6989$
Smu	0	0	0	0	1	1	$\log(5/1) = 0,6989$
Smk	0	0	0	0	1	1	$\log(5/1) = 0,6989$
Sederajat	0	0	0	0	1	1	$\log(5/1) = 0,6989$
Penuh	0	0	0	0	1	1	$\log(5/1) = 0,6989$
Syarat	0	0	0	0	1	1	$\log(5/1) = 0,6989$
Datang	0	0	0	0	1	1	$\log(5/1) = 0,6989$
Tps	0	0	0	0	1	1	$\log(5/1) = 0,6989$
Guna	0	0	0	0	1	1	$\log(5/1) = 0,6989$
Hak	0	0	0	0	1	1	$\log(5/1) = 0,6989$
Pilih	0	0	0	0	1	1	$\log(5/1) = 0,6989$

Setelah dihitung nilai IDF pada Tabel 10, maka berikut Tabel 11 merupakan perhitungan nilai WF dengan rumus $WF = TF \times IDF$

Tabel 11. Perhitugan WF

Token	$WF = TF \times IDF$				
	D1	D2	D3	D4	D5
Sirekap	0,3939	0	0	0,3939	0
Gimana	0,6989	0	0	0	0
Admin	0,2118	0,2118	0	0,2118	0
Error	0,6989	0	0	0	0
Terus	0,6989	0	0	0	0
Rekap	0	0,6989	0	0	0
Notif	0	0,6989	0	0	0
Adu	0	0,6989	0	0	0
Yuk	0	0	0	0,6989	0
Emang	0	0	0,6989	0	0
Undang	0	0,7878	0	0	0

Token	$WF = TF \times IDF$				
	D1	D2	D3	D4	D5
Apa	0	0,6989	0	0	0
Adek	0	0	0	0,7878	0
Siswa	0	0	0	0	0,6989
Pakai	0	0	0,6989	0	0
Tahun	0	0	0,7	0	0
Siswi	0	0	0	0	0,6989
Smu	0	0	0	0	0,6989
Smk	0	0	0	0	0,6989
Sederajat	0	0	0	0	0,6989
Penuh	0	0	0	0	0,6989
Syarat	0	0	0	0	0,6989
Datang	0	0	0	0	0,6989
Tps	0	0	0	0	0,6989
Guna	0	0	0	0	0,6989
Hak	0	0	0	0	0,6989
Pilih	0	0	0	0	0,6989

5. *Klasifikasi Naïve Bayes*

Setelah data dibersihkan dan diberi label, data dibagi menjadi dua bagian: data latih (70%) dan data uji (30%). Pembagian ini dilakukan secara acak. Sebelum melatih model Naive Bayes, kita hitung dulu frekuensi kemunculan setiap kata pada data latih menggunakan metode TF-IDF[19]. Berikut Tabel 12 berisi Data klasifikasi menggunakan metode Naive Bayes.

Tabel 12. Data Klasifikasi Naïve Bayes

Token	Sentiment	Jumlah Kosakata
['sirekap', 'gimana', 'admin', 'error', 'terus']	negative	5
['admin', 'rekap', 'notif', 'adu']	negative	4
['emang', 'undang', 'nih', 'apa', 'pake', 'undang', 'tahun']	positive	7
['admin', 'sirekap']	neutral	2
['yuk', 'adek', 'adek', 'siswa', 'siswi', 'smu', 'smk', 'derajat', 'penuh', 'syarat', 'datang', 'tps', 'guna', 'hak', 'pilih']	positive	15

Berdasarkan data pada Tabel 12. Dapat disimpulkan bahwa kosakata Positive berjumlah 22, Kosakata Negative berjumlah 9 dan Kosakata Neutral berjumlah 2. Sehingga total kosakata keseluruhan berjumlah 33 yang diperoleh dari perhitungan berikut.

$$n(\text{kosakata}) = |\text{positif}| + |\text{negative}| + |\text{netral}| = 22 + 9 + 2 = 33$$

a. Menghitung *probabilities* pada data uji dengan menggunakan rumus persamaan 1 berikut:

$$p(w_i|w_j) = \frac{1+n_i}{n+|\text{kosakata}|}$$

1) Term “pilih”

$$p(\text{'pilih'}|\text{positive}) = \frac{1+1}{22+33} = 0,036$$

$$p(\text{'pilih'}|\text{negative}) = \frac{1+0}{9+33} = 0,023$$

$$p(\text{'pilih'}|\text{neutral}) = \frac{1+0}{2+33} = 0,28$$

2) Term “adu”

$$p(\text{'adu'}|\text{positive}) = \frac{1+0}{22+33} = 0,018$$

$$p(\text{'adu'}|\text{negative}) = \frac{1+1}{9+33} = 0,048$$

$$p(\text{'adu'}|\text{neutral}) = \frac{1+0}{2+33} = 0,028$$

3) Term “admin”

$$p(\text{'admin'}|\text{positive}) = \frac{1+0}{22+33} = 0,018$$

$$p('admin'|negative) = \frac{1+2}{9+33} = 0,071$$

$$p('admin'|neutral) = \frac{1+1}{2+33} = 0,057$$

4) Term “error”

$$p('error'|positive) = \frac{1+0}{22+33} = 0,018$$

$$p('error'|negative) = \frac{1+1}{9+33} = 0,048$$

$$p('error'|neutral) = \frac{1+0}{2+33} = 0,028$$

5) Term “guna”

$$p('guna'|positive) = \frac{1+1}{22+33} = 0,036$$

$$p('guna'|negative) = \frac{1+0}{9+33} = 0,023$$

$$p('guna'|neutral) = \frac{1+0}{2+33} = 0,028$$

6) Term “sirekap”

$$p('sirekap'|positive) = \frac{1+0}{22+33} = 0,018$$

$$p('sirekap'|negative) = \frac{1+1}{9+33} = 0,048$$

$$p('sirekap'|neutral) = \frac{1+1}{2+33} = 0,057$$

7) Term “hak”

$$p('hak'|positive) = \frac{1+1}{22+33} = 0,036$$

$$p('hak'|negative) = \frac{1+0}{9+33} = 0,023$$

$$p('hak'|neutral) = \frac{1+0}{2+33} = 0,028$$

8) Term “adek”

$$p('adek'|positive) = \frac{1+1}{22+33} = 0,036$$

$$p('adek'|negative) = \frac{1+0}{9+33} = 0,023$$

$$p('adek'|neutral) = \frac{1+0}{2+33} = 0,028$$

9) Term “syarat”

$$p('syarat'|positive) = \frac{1+1}{22+33} = 0,036$$

$$p('syarat'|negative) = \frac{1+0}{9+33} = 0,023$$

$$p('syarat'|neutral) = \frac{1+0}{2+33} = 0,028$$

10) Term “notif”

$$p('notif'|positive) = \frac{1+0}{22+33} = 0,018$$

$$p('notif'|negative) = \frac{1+1}{9+33} = 0,048$$

$$p('notif'|neutral) = \frac{1+0}{2+33} = 0,028$$

Selanjutnya dilakukannya perkalian terhadap setiap *probabilities* data uji dari setiap kelas masing-masing

$$p(D - U)|positive = 0,036 \times 0,018 \times 0,018 \times 0,018 \times 0,036 \times 0,018 \times 0,036 \times 0,036 \times 0,036 \times 0,018 = 4,28 \times 10^{-13}$$

$$p(D - U)|negative = 0,023 \times 0,048 \times 0,071 \times 0,048 \times 0,023 \times 0,048 \times 0,023 \times 0,023 \times 0,023 \times 0,048 = 1,2 \times 10^{-10}$$

$$p(D - U)|neutral = 0,028 \times 0,028 \times 0,057 \times 0,028 \times 0,028 \times 0,057 \times 0,028 \times 0,028 \times 0,028 \times 0,028 = 1,21 \times 10^{-11}$$

b. Menghitung *prior probabilities*

$$p(positive) = \frac{2}{5} = 0,4$$

$$p(negative) = \frac{2}{5} = 0,4$$

$$p(neutral) = \frac{1}{5} = 0,2$$

c. Menghitung *posterior probabilities*

$$p(D - U)|positive = 0,4 \times 4,28 \times 10^{-13} = 1,71 \times 10^{-13}$$

$$p(D - U)|negative = 0,4 \times 1,2 \times 10^{-10} = 4,8 \times 10^{-9}$$

$$p(D - U)|neutral = 0,2 \times 1,21 \times 10^{-11} = 2,4 \times 10^{-10}$$

3. HASIL DAN PEMBAHASAN

3.1 Hasil

Hasil ini merupakan hasil yang dihasilkan penulis dalam melakukan penelitiannya source code phyton dengan menggunakan alat google colab yang digunakan penulis dalam melakukan penelitiannya dengan menjabarkan secara rinci hasil akurasi yang didapatkan pada penelitian ini serta jumlah sentiment yang terjadi di Masyarakat pada pengguna komentar Instagram dan juga kata yang paling banyak Masyarakat ungkapkan pada setiap ulasan sentiment yang dihasilkan.

3.1.1 Dataset

Proses pengumpulan data pada penelitian dilakukan menggunakan *web scrapper*. *Web scrapper* yang penulis gunakan sangat amat membantu penulis dalam melakukan scrapping data yang diperlukan. Dengan itu penulis mendapatkan jumlah data sekitar 900 data. Ini merupakan hasil data yang sudah penulis berhasil lakukan *scraping* dengan jumlah data yang didapatka sekitar 900 data. Dapat dilihat pada Gambar 3 berikut ini.

A	B	C
4	mlyh_23	10/2/2024 0:23 Sirekap gimana min, eror terus
5	pungkasdi	10/2/2024 0:25 Min, gimana nih, istri sy sudah dari 2014 KTP bantu! jogja, koq TPSnya daerah asal? Sudah DM koq tidak respon? #kpu melayani #pemilu serentak 2024
6	bt70016	10/2/2024 0:26 Min masih ada si rekap yang belum dapat notif, pengaduanmya kemana ya?
7	boltrainer	10/2/2024 0:30 Emang gk ada undangan nya nih? Apa gk pake undangan tahun ini?
8	bnan_jh	10/2/2024 0:50 min sirekap myaa tolong
9	kpkotam	10/2/2024 0:50 Keren??
0	sang_alang	10/2/2024 1:02 Seandainya gk bisa login sirekap dan ada yang belum dapat notif dari kpu bagaimana solusinya??
1	kpu4lawar	10/2/2024 1:35 Yuk adek adek siswa siswi SMU / SMK / sederajat yang sudah memenuhi syarat untuk datang ke TPS, gunakan Hak Pilihmu
2	kpuprovrb	10/2/2024 1:50 Ke apotik beli obat, berharap sakit cepat pulih. Pemilu sudah semakin dekat, saatnya bersiap-siap untuk memilih ??????????
3	itsmerika	10/2/2024 1:51 aplikasi sirekap di suruh update gilirannya dah update malah gabisa login gimana nih min? huftt
4	kpu_kab_r	10/2/2024 2:01 Keren?????????
5	kadangsuk	10/2/2024 2:32 Sirekap gimana ni
6	nannaas	10/2/2024 2:47 Sirekap ini gimana woy?
7	indahsora	10/2/2024 2:51 Min kenapa per tps cuma dapat zjt doang tapi katanya seharusnya dapat lebih dari situ katanya dari 4.2 apa bener min dari KPU cuma zjt @kpu_ri
8	salvian63	10/2/2024 2:53 Min, sirekap gk bisa kirim foto form c hasil min itu gimana yaa?
9	kpkabbai	10/2/2024 2:57 Ayoi! Teman pilih jangan lupa datang ke tps tanggal 14 Februari 2024 dan gunakan hak pilihmu?????
0	then_bags	10/2/2024 3:00 haloooo KPU...sampaikan saat ini sy belum mendapatkan chatbot dr kpu utk sirekap...minta solusi bgmn klo sampai hari H tdk mendapatkan chatbot
1	kpkabbai	10/2/2024 3:13 Kerenmm bettt dah??
2	kpkabpal	10/2/2024 3:21 14 Februari ke TPS???????
3	kpu_ri	10/2/2024 3:24 @kpubanyumas repost ya bestie
4	kpu_ri	10/2/2024 3:27 @kadangkukaheran sabar ya kak, kami sdg mengusahakan prosesnya??
5	mutiarama	10/2/2024 3:43 @kpu_ri Hai min, data saksi di sirekap kok gk mau ke hapus ya, saya sudah coba clear data dan clear chace tapi data saksiya masi ada, mohon solusinya min
6	amir_mah	10/2/2024 3:44 Apakah honor kpps yg non ASN/PNS, tp tidak memiliki NPWP, tetap di kenakan pajak?
7	rihinra	10/2/2024 4:40 Ka dana di desa BOP saya ga sesuai kak
8	eni_j24	10/2/2024 5:41 Adain KPU online dong, biar anak* Rantau yg blm bisa pulang bisa milih.
9	kpusergai	10/2/2024 5:50 Gk sabar nunggu 4 Hari lagi
0	kyagillya	10/2/2024 6:28 Sedih bgt gag bs myoblos tahun ini... km udh telat ngurusin surat pindah memilih...??
1	yefidar	10/2/2024 6:57 @kyagillyaalea_85 bisa kok oke ktp aja jd DPK...aku pun gtu
2	yefidar	10/2/2024 6:58 @eni_j24 bisa tdi DPK myoblosnya iam 12-13 stana masuk DPK suratnya ktn

Gambar 3. Data Komentar Instagram

Pada Gambar 4 menjelaskan bahwa dari beberapa postingan yang penulis berhasil mengambil data komentarnya, dapat dilihat bahwa penulis tidak mengambil semua postingan yang ada di akun KPU tapi hanya postingan yang berkaitan dengan jelang pelaksanaan pemilu 2024 dengan data yang diambil hanya dalam periode bulan September 2023 hingga Februari 2024.



Gambar 4. Postingan Instagram

3.1.2 Pre-Processing

Hasil yang di peroleh dari proses ini melibatkan beberapa tahap yaitu, case folding, normalisasi, tokenizing, filtering, dan terakhir stemming. Dalam proses ini penelitian melakukan pembersihan kata sebelum dilakukannya ke langkag selanjutnya dapat dilihat pada Gambar 5 yang merupakan hasil dari pre-processing yang penelitian ini lakukan. Dilihat juga dalam gambar hasil penelitian ini melakukan pemisahan column dalam menyimpan setiap kata yang sudah di proses hingga menjadi tahap akhir yaitu dalam bentuk text. Pada tahap ini pula banyak sekali data yang terkadang tidak terbaca ataupun susah untuk dilakukan proses pre-processing seperti hal nya pada proses normalisasi, pada penelitian ini menggunakan kamus Bahasa gaul yang berguna untuk membantu tahap pre-processing normalisasi dapat berjalan dengan efisien dan efektif. Pengguna Bahasa gaul dalam penelitian tidak seakan akan Bahasa yang hanya dikeluarkan menurut kita saja, tapi peneliti juga membaca data set terlebih dahulu agar dapat memahami Bahasa yang digunakan oleh masyarakat pengguna Instagram sebelum dibuatnya kamus Bahasa gaul.

```
[ ] df['stemmed'] = df['stop'].apply(stemming_text)
df.head()
```

Name	Comment	text_clean	normal	token	stop	stemmed
aurora2k19	Min gimana nih SIREKAP masih banyak yg blm bis...	min gimana nih sirekap masih banyak yg blm bis...	admin gimana nih sirekap masih banyak yang bel...	[admin, gimana, nih, sirekap, masih, banyak, y...	[admin, gimana, nih, sirekap, banyak, padahal...	[admin, gimana, nih, sirekap, banyak, padahal...
lutikhaerudin	H-4 aplikasi masih error terus gimana hari H d...	aplikasi masih error terus gimana hari diakses...	aplikasi masih error terus gimana hari diakses...	[aplikasi, masih, error, terus, gimana, hari, diakses...	[aplikasi, error, terus, gimana, hari, diakses...	[aplikasi, error, terus, gimana, hari, akses...
niyh_23	Sirekap gimana min, eror terus	sirekap gimana min eror terus	sirekap gimana admin eror terus	[sirekap, gimana, admin, eror, terus]	[sirekap, gimana, admin, eror, terus]	[sirekap, gimana, admin, eror, terus]
pungkasdenie	Min, gimana nih, istri sy sudah dari 2014 KTP ...	min gimana nih istri sy sudah dari ktp bantu...	admin gimana nih istri saya sudah dari ktp ban...	[admin, gimana, nih, istri, saya, sudah, dari...	[admin, gimana, nih, istri, ktp, bantu, joga...	[admin, gimana, nih, istri, ktp, bantu, joga...
bt70016	Min masih ada si rekap yang belum dapat notif...	min masih ada si rekap yang belum dapat notif...	admin masih ada si rekap yang belum dapat noti...	[admin, masih, ada, si, rekap, yang, belum, da...	[admin, rekap, notif, pengaduannya]	[admin, rekap, notif, adu]

```
#mengubah kebentuk teks kembali
data['teks'] = data['stemmed'].apply(lambda tokens: ' '.join(tokens))
data.head()
```

Name	Comment	text_clean	normal	token	stop	stemmed	teks
aurora2k19	Min gimana nih SIREKAP masih banyak yg blm bis...	min gimana nih sirekap masih banyak yg blm bis...	admin gimana nih sirekap masih banyak yang bel...	[admin, gimana, nih, sirekap, masih, banyak, y...	[admin, gimana, nih, sirekap, banyak, padahal...	[admin, gimana, nih, sirekap, banyak, padahal...	admin gimana nih sirekap banyak padahal tangga...
lutikhaerudin	H-4 aplikasi masih error terus gimana hari H d...	aplikasi masih error terus gimana hari diakses...	aplikasi masih error terus gimana hari diakses...	[aplikasi, masih, error, terus, gimana, hari, diakses...	[aplikasi, error, terus, gimana, hari, diakses...	[aplikasi, error, terus, gimana, hari, diakses...	aplikasi error terus gimana hari akses, uta or...
niyh_23	Sirekap gimana min, eror terus	sirekap gimana min eror terus	sirekap gimana admin eror terus	[sirekap, gimana, admin, eror, terus]	[sirekap, gimana, admin, eror, terus]	[sirekap, gimana, admin, eror, terus]	sirekap gimana admin eror terus
pungkasdenie	Min, gimana nih, istri sy sudah dari 2014 KTP ...	min gimana nih istri sy sudah dari ktp bantu...	admin gimana nih istri saya sudah dari ktp ban...	[admin, gimana, nih, istri, saya, sudah, dari...	[admin, gimana, nih, istri, ktp, bantu, joga...	[admin, gimana, nih, istri, ktp, bantu, joga...	admin gimana nih istri ktp bantu

Gambar 5. Output Pre-Processing

3.1.3 Labelling lexicon

Dalam tahap ini, penelitian memanfaatkan InSet Lexicon, sebuah kamus kata yang telah dilabel dengan sentimen positif, negatif, atau netral. Setiap kata dalam sebuah ulasan akan dicocokkan dengan kata-kata dalam leksikon, dan kemudian skor sentimen dari masing-masing kata akan dijumlahkan untuk mendapatkan skor sentimen keseluruhan ulasan. Skor sentimen inilah yang menjadi indikasi awal mengenai sentimen yang terkandung dalam ulasan tersebut. Visualisasi dari hasil pelabelan awal menggunakan InSet Lexicon dapat dilihat pada Gambar 6, yang merupakan hasil dari gambaran distribusi skor sentimen pada seluruh dataset yang dilakukan dalam penelitian ini.

	stemmed	polarity_score	sentiment
0	['admin', 'gimana', 'nih', 'sirekap', 'banyak']	-2	negative
1	['aplikasi', 'error', 'terus', 'gimana', 'hari']	-3	negative
2	['sirekap', 'gimana', 'admin', 'error', 'terus']	-5	negative
3	['admin', 'gimana', 'nih', 'istri', 'ktp', 'ba']	5	positive
4	['admin', 'rekap', 'notif', 'adu']	-3	negative

Gambar 6. Output Labelling

3.1.4 Ekstraksi fitur TF-IDF

```
TfidfTransformer()
(0, 2893) 0.3629780012910555
(0, 2629) 0.3656547718714476
(0, 2039) 0.3838709146236218
(0, 1925) 0.4073556071821076
(0, 985) 0.34423687802481234
(0, 919) 0.3290050849374216
(0, 271) 0.3104422996745937
(0, 35) 0.3132575506759746
(1, 2862) 0.2684905006391598
(1, 2020) 0.23124430503344592
(1, 1226) 0.347547926153404
(1, 985) 0.24695498019238796
(1, 919) 0.23602771643792178
(1, 789) 0.3328298889110735
(1, 327) 0.27788084520225626
(1, 157) 0.3565756427956594
(1, 74) 0.42091503106756406
```

Gambar 7. Output TF-IDF

Gambar 7 menunjukkan proses perhitungan bobot pentingnya setiap kata dalam sebuah dokumen dan seluruh kumpulan dokumen. Dengan menggunakan library scikit-learn, matriks term-document diubah menjadi representasi vektor TF-IDF. Vektor ini kemudian digunakan sebagai input untuk model pembelajaran mesin, sehingga memungkinkan model untuk belajar pola dan hubungan antara kata-kata dalam dokumen, yang akan menghasilkan nilai TF-IDF yang tertera pada Gambar 7 yaitu hasil output TF-IDF pada penelitian ini

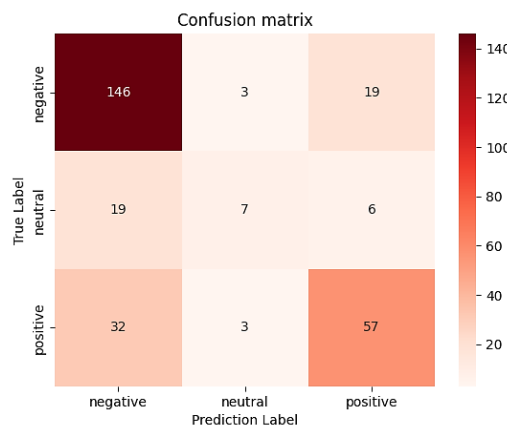
3.1.5 Evaluasi Model

```
[27] columns = ['negative', 'neutral', 'positive']
confm = confusion_matrix(y_test, y_pred)
df_cm = DataFrame(confm, index=columns, columns=columns)

ax = sn.heatmap(df_cm, cmap='Reds', annot=True, fmt='d')
ax.set_title('Confusion matrix')
ax.set_xlabel('Prediction Label')
ax.set_ylabel('True Label')
```

Gambar 8. Source Code Evaluasi Model

Pada Gambar 8 ditampilkan source code yang digunakan untuk membentuk sebuah confusion matrix. Confusion matrix merupakan alat penting dalam evaluasi kinerja model klasifikasi, yang memungkinkan kita untuk mengevaluasi seberapa baik model dalam mengklasifikasikan data. Confusion matrix menyajikan perbandingan antara label yang diprediksi oleh model dan label sebenarnya dari data. Dengan menggunakan confusion matrix dapat dengan jelas melihat model performa pada berbagai kategori dan mengidentifikasi area di mana model mungkin memerlukan perbaikan. Hasil dari confusion matrix, seperti yang ditunjukkan pada Gambar 9 di bawah ini, memberikan insight mendalam mengenai akurasi dan efisiensi model yang sedang dikembangkan dalam penelitian ini.



Gambar 9. Confusion Matrix

Penelitian ini melibatkan pengujian terhadap sejumlah 292 data yang dipilih secara acak menggunakan metode random sampling dengan menggunakan algoritma naive bayes. Dari total data yang tersedia, pembagian dilakukan dengan rasio 30:70, di mana 30% data digunakan sebagai data uji dan 70% sebagai data latih. Berikut cara menghitung nilai accuracy, precision, recall dan juga F1 Score dengan melihat confusion matrix yang sudah diperoleh dari hasil penelitian ini :

a. Accuracy

$$\frac{TP+TN+TNet}{TP+FN+FP+FNTNet+FNet} = \frac{57+146+7}{292} = 0,72$$

b. Precision

$$Negatif = \frac{Jumlah\ Prediksi\ benar\ satu\ kelas}{Total\ prediksi\ satu\ kelas} = \frac{146}{146+19+31} = 0,74$$

$$Positif = \frac{Jumlah\ Prediksi\ benar\ satu\ kelas}{Total\ prediksi\ satu\ kelas} = \frac{57}{19+6+57} = 0,70$$

$$Netral = \frac{Jumlah\ Prediksi\ benar\ satu\ kelas}{Total\ prediksi\ satu\ kelas} = \frac{7}{3+7+3} = 0,54$$

c. Recall

$$Negatif = \frac{Jumlah\ prediksi\ benar\ satu\ kelas}{Total\ jumlah\ sampel\ yang\ sebenarnya\ dalam\ kelas} = \frac{146}{146+3+19} = 0,87$$

$$Positif = \frac{Jumlah\ prediksi\ benar\ satu\ kelas}{Total\ jumlah\ sampel\ yang\ sebenarnya\ dalam\ kelas} = \frac{57}{32+3+57} = 0,62$$

$$Netral = \frac{Jumlah\ prediksi\ benar\ satu\ kelas}{Total\ jumlah\ sampel\ yang\ sebenarnya\ dalam\ kelas} = \frac{7}{19+7+6} = 0,22$$

d. F-1 Score

$$F - 1\ Score\ Negatif = \frac{2 \times (Presisi\ N \times Recall\ N)}{Presisi\ N + Recall\ N} = \frac{2 \times (0,74 \times 0,87)}{0,74 + 0,87} = 0,80$$

$$F - 1\ Score\ Positif = \frac{2 \times (Presisi\ P \times Recall\ P)}{Presisi\ P + Recall\ P} = \frac{2 \times (0,70 \times 0,62)}{0,70 + 0,62} = 0,66$$

$$F - 1\ Score\ Netral = \frac{2 \times (Presisi\ Net \times Recall\ Net)}{Presisi\ Net + Recall\ Net} = \frac{2 \times (0,54 \times 0,22)}{0,54 + 0,22} = 0,66$$

3.2 Pembahasan

Dalam proses penelitian ini mengenai Pelaksanaan Pemilu 2024 menggunakan Data Komentar Instagram dengan Metode Naïve Bayes serta Pelabelan menggunakan *InSet Lexicon*. Menggunakan data sebanyak 972 data. Data dikumpulkan dalam rentang waktu yang cukup panjang, yaitu dari September 2023 hingga Februari 2024. Pengambilan data pada rentan waktu ini dikarenakan pada waktu tersebut jelang pelaksanaan pemilu sangat ramai dibicarakan oleh masyarakat. Data komentar diambil menggunakan tools atau web scraper. penggunaan *web scraper* memudahkan pengumpulan data secara efisien, namun data yang diambil sering kali tidak terstruktur dan mengandung banyak noise, seperti emotikon, singkatan, atau kesalahan penulisan. Oleh karena itu, tahap pre-processing sangat penting dalam penelitian ini, disini data akan dioleh dan untuk membersihkan dan menyusun data mentah agar lebih siap untuk dianalisis. Data yang tidak terstruktur dan mengandung banyak noise akan sulit diproses oleh algoritma. *Pre-processing* pada penelitian ini melalui beberapa tahap yaitu, case folding, normalisasi, tokenizing, filtering, dan yang terakhir yaitu stemming. Setiap langkah pre-processing ini harus dilakukan dengan sangat hati-hati karena akan sangat mempengaruhi hasil akhir dari analisis. Jika data tidak diproses dengan benar, hasil dari algoritma Naïve Bayes atau pelabelan sentimen dapat menjadi tidak akurat, yang pada gilirannya dapat mempengaruhi kesimpulan penelitian.

Setelah melakukan tahap *pre-processing* dilakukannya pelabelan dengan menggunakan *Inset Lexicon* dapat dilihat hasil dari sentiment masyarakat yang dihasilkan pada pelabelan menggunakan *Inset Lexicon* pada Gambar 10.

```

results = data['teks'].apply(sentiment_analysis_lexicon_indonesia)
results = list(zip(*results))
data['polarity_score'] = results[0]
data['sentiment'] = results[1]
print(data['sentiment'].value_counts())
print(data.shape)

sentiment
negative    521
positive    332
neutral     119
Name: count, dtype: int64
(972, 5)
    
```

Gambar 10. Sentiment Masyarakat

meningkatkan kinerja model Naïve Bayes dan pelabelan Lexicon-Based secara keseluruhan. Dibandingkan dengan penelitian Prasetyo, Buntoro, dan Mustikasari (2020) yang mencapai akurasi 66,5%, penelitian ini menunjukkan bahwa perhatian yang lebih besar terhadap detail dalam pre-processing dapat menghasilkan model yang lebih akurat dan andal. Hasil ini tidak hanya menunjukkan efektivitas pendekatan yang digunakan dalam menangani komentar dalam bahasa tidak formal di media sosial tetapi juga memberikan kontribusi yang berarti bagi pengembangan metodologi dalam analisis sentimen berbasis media sosial.

REFERENCES

- [1] Jimmy, E. H. Hermaliani, dan L. Kurniawati, “Analisis Klasifikasi Sentimen Pengguna Media Sosial Twitter Terhadap Penundaan Pemilu Presiden Tahun 2024,” *J. Indones. Manaj. Inform. dan Komun.*, vol. 4, no. 2, hal. 570–579, 2023, doi: <https://doi.org/10.35870/jimik.v4i2.243>.
- [2] A. S. Afif dan A. R. Pratama, “Analisis Sentimen Kebijakan Pendidikan di Masa Pandemi COVID-19 dengan CrowdTangle di Instagram,” *Automata*, vol. 2, no. 2, hal., 2021, [Daring]. Tersedia pada: <https://journal.uii.ac.id/AUTOMATA/article/view/19429>
- [3] T. D. Putra, E. Utami, dan M. P. Kurniawan, “Analisis Sentimen Pemilu 2024 dengan Naive Bayes Berbasis Particle Swarm Optimization (PSO),” *Explore*, vol. 13, no. 1, hal. 1–5, 2023, [Daring]. Tersedia pada: <https://www.journal.utmmataram.ac.id/index.php/explore/article/download/13/9>
- [4] D. P. Kussanti dan F. Azizi, “Company Profile Komisi Pemilihan Umum Republik Indonesia Sebagai Media Informasi Kepada Publik Eksternal,” *J. Public Relations*, vol. 2, no. 1, hal. 67–71, 2021, doi: <https://doi.org/10.31294/jpr.v2i1.513>.
- [5] A. R. Alifvia dan U. Saprudin, “Analisis Sentimen Review Data Twitter Komisi Pemilihan Umum (KPU) Menggunakan Metode Naïve Bayes,” *J. Inf. dan Komput.*, vol. 11, no. 01, hal. 81–84, 2023, doi: <https://doi.org/10.35959/jik.v11i01.407>.
- [6] R. Vindua dan A. U. Zailani, “Analisis Sentimen Pemilu Indonesia Tahun 2024 Dari Media Sosial Twitter Menggunakan Python,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 10, no. 2, hal. 479–487, 2024, doi: [10.30865/jurikom.v10i2.5945](https://doi.org/10.30865/jurikom.v10i2.5945).
- [7] F. A. Artanto, “Support Vector Machine Berbasis Particle Swarm Optimization Pada Analisis Sentimen Anggota KPPS,” *J. FASILKOM (teknologi Inf. dan Ilmu KOMputer)*, vol. 14, no. 1, hal. 75–79, 2024, doi: <https://doi.org/10.37859/jf.v14i1.6795>.
- [8] D. S. Nugroho, I. F. Hanif, M. A. Hasbi, F. Fredianto, A. M. Saputra, dan R. Zildjian, “Analisis Sentimen Dugaan Pelanggaran Pemilu 2024 Berdasarkan Tweet Menggunakan Algoritma Naïve Bayes Classifier,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 3, hal. 1169–1176, 2024, doi: <https://doi.org/10.57152/malcom.v4i3.1496>.
- [9] E. Hasibuan dan E. A. Heriyanto, “Analisis Sentimen Pada Ulasan Aplikasi Amazon Shopping Di Google Play Store Menggunakan Naive Bayes Classifier,” *J. Tek. Dan Sci.*, vol. 1, no. 3, hal. 13–24, 2022, doi: <https://doi.org/10.56127/jts.v1i3.434>.
- [10] O. Manullang, C. Prianto, dan N. H. Harani, “Analisis Sentimen Untuk Memprediksi Hasil Calon Pemilu Presiden Menggunakan Lexicon Based Dan Random Forest,” *J. Ilm. Inform.*, vol. 11, no. 02, hal. 159–169, 2023, doi: <https://doi.org/10.33884/jif.v11i02.7987>.
- [11] S. Ratnaswari, N. C. Wibowo, dan D. S. Y. Kartika, “Analisis Sentimen Menggunakan Metode Lexicon-Based Dan Support Vector Machine Pada Presiden Dan Wakil Presiden Indonesia Periode 2024–2029,” *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 1, hal. 362–368, 2025, doi: <http://dx.doi.org/10.23960/jitet.v13i1.5604>.
- [12] M. Saputra dan M. Iqbal, “Analisis Sentimen Masyarakat Terhadap PON XII Aceh-Sumut Menggunakan Algoritma Naïve Bayes,” *J. Sist. Inf. dan Sist. Komput.*, vol. 10, no. 1, hal. 39–48, 2025, doi: <https://doi.org/10.51717/simkom.v10i1.679>.
- [13] V. F. Lestari, P. Arsi, dan P. Subarkah, “Sentimen Analisis Evaluasi Pengguna Aplikasi Orbit Telkomsel Pada Ulasan Playstore Menggunakan Algoritma Naïve Bayes,” *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 9, no. 4, hal. 2244–2255, 2024, doi: <https://doi.org/10.29100/jupi.v9i4.5538>.
- [14] D. P. Yani, “Klasifikasi Sentimen Transformasi dan Reformasi Sepak Bola Indonesia Pada Twitter Menggunakan Algoritma Bernoulli Naïve Bayes,” *Univ. Islam Negeri Sultan Syarif Kasim Riau Repos.*, vol. 4, no. 3, hal. 452–458, 2023, [Daring]. Tersedia pada: <http://repository.uin-suska.ac.id/id/eprint/71577>
- [15] D. Abimanyu, “Analisis Sentimen Akun Twitter Apex Legends Menggunakan VADER,” *Univ. Islam Negeri Sultan Syarif Kasim Riau Repos.*, vol. 5, no. 03, hal. 423–431, 2022, [Daring]. Tersedia pada: <http://repository.uin-suska.ac.id/id/eprint/61480>
- [16] J. Supriyanto, D. Alita, dan A. R. Isnain, “Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring,” *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 4, no. 1, hal. 74–80, 2023, doi: <https://doi.org/10.33365/jatika.v4i1.2468>.
- [17] D. D. Putri, G. F. Nama, dan W. E. Sulistiono, “Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier,” *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, hal. 34–40, 2022, doi: <http://dx.doi.org/10.23960/jitet.v10i1.2262>.

- [18] M. I. Maulana, "Klasifikasi Sentiment Ulasan Aplikasi Sausage Man Menggunakan VADER Lexicon dan Naïve Bayes Classifier," *Univ. Islam Negeri Sultan Syarif Kasim Riau Repos.*, vol. 4, no. 3, hal. 485–492, 2023, [Daring]. Tersedia pada: <http://repository.uin-suska.ac.id/id/eprint/71622>
- [19] D. Darwis, N. Siskawati, dan Z. Abidin, "Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional," *J. Tekno Kompak*, vol. 15, no. 1, hal. 131–145, 2021, doi: 10.33365/jtk.v15i1.744.
- [20] H. Prasetyo, G. A. Buntoro, dan D. Mustikasari, "Analisis Sentimen Pada Channel Autonetmagz Terhadap Review Mobil Almaz 2019 Dengan Metode Naive Bayes Classifier Dan Lexicon Based," *Komputek J. Tek. Univ. Muhammadiyah Ponorogo*, vol. 4, no. 1, hal. 58–70, 2020, [Daring]. Tersedia pada: <https://studentjournal.umpo.ac.id/index.php/komputek>