Implementasi Text Mining untuk Analisis Opini Publik terhadap Calon Presiden Indonesia 2024

Gina Nur Fatimah*, Ghea Putri Amalia P, Fikran Muhammad Ikmal, Jasmine Azalia Pangesti, Ilham Faishal Mahdy

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

*10060120093@unisba.ac.id, 10060120080@unisba.ac.id, 10060120094@unisba.ac.id, 10060120098@unisba.ac.id, ilham.faishal@unisba.ac.id

Abstract. The holding of the Indonesian presidential general election in 2024, which began with the announcement of three pairs of presidential and vice-presidential candidates, was widely discussed on social media. Therefore, sentiment analysis was conducted on the three presidential candidates to determine their initial electability. Data was collected from the Twitter/X application on each candidate pair name with a total of 4.415 tweets. After going through a series of stages in data preprocessing, the number was reduced to 3.384 tweets. By conducting sentiment analysis using the VADER algorithm, it was found that the highest positive sentiment value with a result of 92.2% was on the presidential candidate pair Ganjar and Mahfud. Meanwhile, the presidential candidate pair that received the highest negative sentiment value with a result of 12.7% was Prabowo and Gibran

Keywords: *Presidentian candidate, sentiment analysis, social media, twitter/X*

Abstrak. Diselenggarakannya pemilihan umum presiden Indonesia pada tahun 2024 yang diawali dengan ditetapkannya tiga pasangan calon presiden dan calon wakil presiden, ramai diperbincangkan di media sosial. Oleh karena itu, dilakukan analisis sentiment terhadap ketiga calon presiden untuk mengetahui elektabilitas awal mereka. Data dikumpulkan dari aplikasi twitter/X pada masing-masing nama pasangan calon dengan jumlah data yang didapat sebanyak 4.415 cuitan. Setelah melalui serangkaian tahapan dalam *preprocessing data*, jumlahnya berkurang menjadi sebanyak 3.384 cuitan. Dengan dilakukan analisis sentimen menggunakan algoritma VADER, didapatkan bahwa hasil nilai sentimen positif tertinggi dengan hasil 92,2% yaitu pada calon pasangan presiden Ganjar dan Mahfud. Sementara itu, calon pasangan presiden yang mendapat nilai sentimen negatif tertinggi dengan hasil 12,7% yaitu pada pasangan Prabowo dan Gibran.

Kata Kunci: *Calon presiden, analisis sentiment, media sosial, twitter/X*

Corresponding Author
Email: ilham.faishal@unisba.ac.id

A. Pendahuluan

Indonesia akan menyelenggarakan pemilihan umum presiden dan wakil presiden untuk masa jabatan 2024-2029 pada 14 Februari 2024. Presiden adalah kepala negara sekaligus kepala pemerintahan Indonesia dengan tugas dan wewenang yang telah diatur dalam Undang-Undang Dasar 1945. Presiden dan wakil presiden Indonesia dipilih secara langsung melalui pemilihan umum yang diselenggarakan secara demokratis dan berdasarkan asas langsung, umum, bebas, rahasia, jujur, dan adil dengan masa jabatan selama lima tahun. Penetapan tiga pasangan calon presiden dan calon wakil presiden ini tentunya menimbulkan banyak reaksi dari masyarakat, terutama di media sosial. Media sosial sering kali menjadi tempat untuk menyuarakan pendapat, mengingat banyaknya masyarakat di berbagai kalangan usia yang saat ini menggunakan media sosial. Berdasarkan laporan *We Are Social* seperti yang dilaporkan oleh Databoks, pada bulan Januari 2023 jumlah pengguna internet di Indonesia telah mencapai 213 juta orang. Angka ini setara dengan 77% dari total populasi penduduk Indonesia yang berjumlah 276,4 juta [1].

Melalui media sosial, masyarakat dapat dengan bebas menyuarakan pendapat, pandangan, dan emosi mereka tentang berbagai peristiwa yang terjadi, baik itu isu politik, sosial, budaya, dan lainnya. Twitter/X menjadi salah satu platform media sosial yang sering digunakan oleh masyarakat untuk menyuarakan pendapat mereka, sekaligus merupakan platform yang sangat populer dan memiliki pengguna aktif di seluruh dunia. Indonesia menempati peringkat ke-6 di dunia dalam penggunaan twitter, dengan 63,3% pengguna berusia 16-64 tahun, yang sesuai dengan usia yang cukup untuk menjadi pemilih dan pemilu 2024 [2]. Twitter juga sering menjadi sumber data dalam *text mining* karena platform ini tidak hanya menghasilkan pencarian berdasarkan nama pengguna, tetapi juga berdasarkan teks dalam *tweet*. Selain itu, twitter menyediakan layanan *API (Application Programming Interface)* yang memungkinkan pengguna untuk mengambil data, dengan setiap tweet yang berhasil diambil akan dianggap sebagai satu data [3]. Dengan melalui twitter inilah, kita dapat melihat bagaimana pendapat masyarakat Indonesia terhadap calon presiden Indonesia.

Berdasarkan pendapat yang disampaikan oleh masyarakat, hal ini menjadi salah satu prediksi awal untuk elektabilitas presiden pada periode mendatang. Oleh karena itu, diperlukan mesin yang dapat menganalisis tweet, salah satunya melalui proses *text mining* menggunakan algoritma *Valence Aware Dictionary and Sentiment Reasoner* (VADER). VADER berfungsi sebagai model untuk menganalisis sentimen dan dapat mengidentifikasi variasi dalam data dengan mengukur intensitas kekuatan emosional yang terkandung dalam teks, sesuai dengan informasi yang terdapat dalam kamus data Lexicon yang digunakan [4].

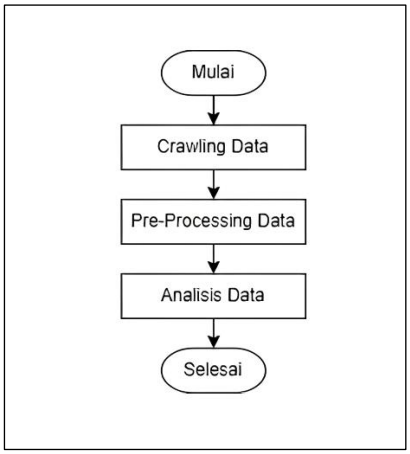
Kamus lexicon ini digunakan untuk mengevaluasi sentimen frasa dan kalimat tanpa memerlukan referensi tambahan. Sentimen dapat dikelompokkan dalam kategori seperti {negatif, netral, positif}, atau dapat diungkapkan secara numerik melalui rentang intensitas atau skor. Pendekatan leksikal melibatkan penilaian kategori atau skor sentimen untuk setiap kata dalam suatu kalimat, yang kemudian menghasilkan penilaian akhir terhadap kategori atau skor sentimen untuk seluruh kalimat tersebut [5]. Kelebihan dari penerapan deteksi polaritas VADER adalah adanya kamus yang mencakup nilai untuk setiap kata. Penentuan polaritas kalimat dilakukan dengan menggabungkan atribut "compound" dari setiap kata yang ada dalam teks. [6]. Dalam penelitian ini, analisis sentimen dilakukan berdasarkan opini masyarakat di media sosial twitter/X dengan tujuan untuk mengidentifikasi apakah sentimen tersebut positif, negatif, atau netral terhadap calon presiden Indonesia.

B. Metodologi Penelitian

Analisis sentimen adalah teknik pemprosesan teks yang digunakan untuk memahami perasaan atau opini seseorang terhadap topik tertentu. Teknik ini juga dikenal sebagai penambangan opini terhadap suatu entitas yang diungkapakan dalam bentuk teks. Jika disediakan sekelompok dokumen teks yang mencakup sentiment terkait suatu objek, opinion mining bertujuan untuk menarik informasi terkait atribut dan elemen dari objek yang diulas dalam setiap dokumen [3]. Tujuan utamanya adalah untuk menentukan polaritas dari data teks, apakah sentimen tersebut positif, negatif, atau netral. Analisis sentimen berusaha untuk mengidentifikasi sikap atau pandangan yang diungkapkan oleh seorang penulis terhadap suatu topik tertentu yang dapat

mencerminkan alasan, opini, penilaian, atau upaya penulis untuk mempengaruhi pembaca melalui kecenderungan tertentu.

Data yang digunakan pada penelitian ini adalah data primer yang diperoleh dari aplikasi twitter/X. Twitter/X adalah platform komunikasi online yang menawarkan kemungkinan bagi pengguna untuk berbagi pendapat mereka secara langsung melalui teks singkat. Format singkat ini memungkinkan esensi pesan disampaikan secara jelas dan padat, dan dapat dipublikasikan untuk dilihat oleh semua orang atau kelompok tertentu. Penelitian ini melalui beberapa tahapan analisis, berikut adalah tahapan analisisnya:



Gambar 1. Tahapan Analisis

Data *Crawling*

Proses *crawling* data adalah teknik untuk mengumpulkan data secara otomatis dari berbagai sumber di internet menggunakan perangkat lunak. Sumber data yang dapat di *crawl* bisa diambil dari halaman *web*, *blog*, dan media sosial. Pada penelitian ini, sumber data yang digunakan ialah melalui media sosial twitter/X. *Tweet-harvest* merupakan sebuah alat yang dirancang untuk mengambil atau merangkum data dari Twitter/X. Dengan menggunakan alat ini, kita dapat memperoleh informasi yang lebih komprehensif mengenai topik yang sedang dibicarakan, baik dalam bentuk *tweet*, *retweet*, serta balasan dengan kata kunci dan jumlah data yang diinginkan.

Data *Preprocessing*

Setelah proses *crawling*, langkah selanjutnya adalah *preprocessing*. *Pre-processing* data melibatkan serangkaian langkah untuk membersihkan dan menyederhanakan teks agar dapat diolah lebih lanjut [3]. Berikut adalah tahapan dalam proses ini:

1. *Case Folding* melibatkan konversi semua huruf kapital dalam ulasan pada dokumen menjadi huruf kecil.
2. *Cleaning* adalah tahapan di mana dokumen dibersihkan dari kata-kata yang tidak relevan seperti entitas *tweet* seperti *mention*, *retweet*, dan *hastag*.
3. *Tokenization* adalah proses membagi kata-kata berdasarkan unit-unit individu, memecahnya menjadi potongan-potongan tunggal.
4. Normalisasi melibatkan koreksi kata-kata yang salah dalam teks berdasarkan korpus yang telah dibuat.
5. *Stopword* merupakan langkah untuk menghapus kata-kata yang tidak relevan dengan topik dokumen, karena kata-kata tersebut tidak memiliki kontribusi yang signifikan terhadap akurasi dalam klasifikasi sentimen.
6. *Stemming* adalah suatu proses yang terintegrasi dalam sistem IR (*Information Retrieval*) yang mengubah kata-kata dalam suatu dokumen ke bentuk kata akarnya dengan melewati serangkaian tahap tertentu sehingga tersisa kata dasarnya saja.

Analisis Data

Dalam menganalisis sentimen, metode yang digunakan adalah *lexicon based* menggunakan *library Vader sentiment. Lexicon based* adalah sebuah metode yang melibatkan pemilahan kata-kata kunci dalam dokumen berdasarkan suatu kamus atau lexikon yang telah ada. Dalam penerapannya, terdapat dua kamus yang digunakan sebagai *wordlist*, yakni kamus berisi kumpulan kata dengan sentimen positif dan kamus yang berisi kumpulan kata dengan sentimen negatif [7].

Valance Aware Dictionaty and Sentiment Reasoner (VADER) adalah suatu metode analisis berbasis leksikon yang menganalisis teks dengan memanfaatkan leksikon untuk menghasilkan kelompok sentimen positif, negatif, dan netral beserta skor total atau skor komposit [8]. Analisis sentimen dengan Vader-Lexicon didukung oleh *database* yang berisi teks dengan klasifikasi sentiment positif, negatif, dan netral. Sebelum melakukan analisis sentimen, data hasil *preprocessing* diterjemahkan dahulu ke dalam bahasa inggris karena database yang digunakan berbahasa inggris, kemudian di terjemahkan kembali ke dalam bahasa Indonesia. Nilai compound dihitung dengan menjumlahkan skor valensi setiap kata dalam leksikon, disesuaikan dengan aturan Vader dan kemudian di normalisasi menjadi rentang antara -1 dan +1. Skor compound ini berfungsi sebagai parameter standar untuk mengklasifikasikan kalimat, dimana nilai ambang yang biasanya disarankan adalah sebagai berikut [9]:

1. Sentimen Positif:

   Skor compound $\geq 0{,}5$
2. Sentimen Netral

   Skor compound $>$ -0,5 dan skor compound $< 0{,}5$
3. Sentimen Negatif

   Skor compound $\leq 0{,}5$

Hasil dari nilai compound akan mengungkapkan opini publik terhadap calon presiden Indonesia 2024. Dengan demikian, analisis sentimen ini dapat memberikan wawasan tentang pandangan masyarakat terhadap calon presiden yang bersangkutan. Pengambilan data dilakukan setelah pengumuman calon presiden dan wakil presiden, serta sebelum para kandidat melakukan debat. Oleh karena itu, opini yang terdapat dalam analisis ini mungkin dapat mengalami perubahan seiring berjalannya waktu dan perkembangan kampanye, terutama setelah perdebatan antara para calon presiden.

C. Hasil Penelitian dan Pembahasan

Pengumpulan Data

Pengumpulan data tanggapan dan pandangan masyarakat Indonesia terkait Calon Presiden 2024 dilakukan menggunakan teknik *crawling* data pada media sosial Twitter/X, khususnya dalam bentuk cuitan. Teknik *crawling* data dilakukan dengan menggunakan Google Collab dengan *library tweet-harvest*. *Library* ini memiliki kemampuan untuk mengakses data publik dari platform media sosial X. Dengan menggunakan *library* ini, peneliti dapat mengambil *tweet*, *retweet*, dan balasan dengan kata kunci tertentu serta menentukan jumlah data yang diinginkan.

Crawling data dilakukan pada tanggal 16 November 2023 dengan kata kunci yang dipilih untuk mengambil data dari cuitan di Twitter/X mencakup Anies dan Anies Muhaimin sebagai calon presiden pertama, Prabowo dan Prabowo Gibran sebagai calon presiden kedua, serta Ganjar dan Ganjar Mahfud sebagai calon presiden ketiga. Jenis cuitan yang diambil dari kata kunci tersebut sesuai dengan hasil pencarian yang ditemukan pada kolom *search* aplikasi Twitter/X menggunakan kata kunci yang sama. Twitter/X menyediakan *Application Programming Interface* (API), memungkinkan siapa pun untuk mengakses informasi web dari aplikasi tersebut. Walaupun demikian, API Twitter/X memiliki batasan penggunaan, terutama dalam mengakses data cuitan dengan jumlah tertentu. Setelah melalui proses *crawling* dari pengambilan kata kunci, diperoleh dataset yang dijelaskan lebih lanjut pada Tabel 1.

Tabel 1. Hasil Pengumpulan Data

Pasangan Calon	Jumlah Data
1 (Anies-Muhaimin)	1025
2 (Prabowo-Gibran)	1497
3 (Ganjar-Mahfud)	1893
Total	4415

Dalam tahap pengumpulan data, tidak hanya atribut cuitan yang diperoleh, melainkan juga beberapa atribut yang dapat menjadi sumber informasi tambahan dalam penelitian ini. Berikut adalah rincian mengenai jenis atribut yang diperoleh selama proses crawling data, sebagaimana dijelaskan dalam Tabel 2.

Tabel 2. Atribut Data Cuitan Twitter

Nama Atribut	Penjelasan Atribut	Tipe Data Atribut
created_at	Waktu dan tanggal dari pembuatan cuitan tersebut	*Date*
Tweet	Cuitan	*Object*
reply_count	Banyaknya balasan dari cuitan	*Object*
retweet_count	Banyaknya cuitan yang dibagikan	*Object*
favorite_count	Banyaknya yang menyukai cuitan	*Object*
lang	Bahasa yang digunakan	*Object*
username	Nama dari akun pembuatan cuitan tersebut	*Object*
tweet_url	Tautan dari cuitan tersebut	*Object*

Preprocessing Data

Pada proses ini dilakukan tahapan *preprocessing* data untuk memastikan data yang digunakan dalam analisis memiliki kualitas yang baik. Tahapan *preprocessing* ini melibatkan serangkaian langkah pembersihan, sesuai dengan metodologi penelitian yang telah ditetapkan. Beberapa tahapan pembersihan yang akan dilakukan mencakup langkah-langkah seperti case folding, cleaning, tokenizing, normalisasi, removal stopwords, stemming, serta penghapusan *missing value* dan data duplikat.

Pada tahapan *preprocessing*, informasi mengenai perubahan dan pembersihan data akan disajikan secara terstruktur dalam bentuk tabel. Rincian tahapan pembersihan data dari awal hingga pembersihan perbaikan ejaan akan tercatat dan dapat diakses pada Tabel 3. Tabel ini akan memberikan gambaran yang jelas mengenai perubahan dan pemrosesan yang telah dilakukan pada data sebelum dilakukan analisis lebih lanjut, berikut merupakan salah satu *tweet* yang sudah melewati serangkaian tahapan dalam proses *preprocessing*.

Tabel 3. *Preprocessing* Data

Tahapan	Sebelum	Sesudah
Case Folding	2024, saatnya kita mendukung pemimpin yang gak pernah berkhianat pada rakyat. Anies-Muhaimin, pilihan cerdas! #AniesHarmonis	2024, pilihan kita pasti jatuh ke anies-muhaimin. mereka emang punya rekam jejak bersih, bro! #aniesharmonis
Cleaning	2024, pilihan kita pasti jatuh ke anies-muhaimin. mereka emang punya rekam jejak bersih, bro! #aniesharmonis	pilihan kita pasti jatuh ke anies muhaimin mereka emang punya rekam jejak bersih bro

Tahapan	Sebelum	Sesudah
Tokenization	pilihan kita pasti jatuh ke anies muhaimin mereka emang punya rekam jejak bersih bro	['pilihan', 'kita', 'pasti', 'jatuh', 'ke', 'anies', 'muhaimin', 'mereka', 'emang', 'punya', 'rekam', 'jejak', 'bersih', 'bro']
Normalisasi	['pilihan', 'kita', 'pasti', 'jatuh', 'ke', 'anies', 'muhaimin', 'mereka', 'emang', 'punya', 'rekam', 'jejak', 'bersih', 'bro']	['pilihan', 'kita', 'pasti', 'jatuh', 'ke', 'anies', 'muhaimin', 'mereka', 'memang', 'punya', 'rekam', 'jejak', 'bersih', 'bro']
Stopword Removal	['pilihan', 'kita', 'pasti', 'jatuh', 'ke', 'anies', 'muhaimin', 'mereka', 'memang', 'punya', 'rekam', 'jejak', 'bersih', 'bro']	['pilihan', 'jatuh', 'anies', 'muhaimin', 'rekam', 'jejak', 'bersih', 'bro']
Stemming	['pilihan', 'jatuh', 'anies', 'muhaimin', 'rekam', 'jejak', 'bersih', 'bro']	['pilihan', 'jatuh', 'anies', 'muhaimin', 'rekam', 'jejak', 'bersih', 'bro']

Pada tahapan selanjutnya, dilakukan proses penghapusan data yang hilang (*missing value*) dan data yang duplikat pada dataset cuitan calon presiden. Langkah-langkah ini dilakukan untuk meningkatkan kualitas dan keakuratan data yang akan digunakan dalam analisis. Data yang hilang atau duplikat dapat memengaruhi hasil analisis dan interpretasi. Setelah dilakukan penghapusan, perubahan pada jumlah data secara kuantitatif dijelaskan dalam Tabel 4.

Tabel 4. Hasil Pengumpulan Data

Pasangan Calon	Missing Value			Duplikat		
	Sebelum	Sesudah	Selisih	Sebelum	Sesudah	Selisih
1 (Anies-Muhaimin)	1025	1014	11	1014	530	484
2 (Prabowo-Gibran)	1497	1490	7	1490	1006	484
3 (Ganjar-Mahfud)	1893	1892	1	1893	1848	45

Setelah melalui serangkaian tahapan dalam proses *preprocessing*, termasuk penanganan *missing value* dan data duplikat, terjadi pengurangan jumlah data dari jumlah awal sebanyak 4.415 tweet menjadi 3.384 tweet. Meskipun mengalami pengurangan, proses ini memastikan bahwa dataset yang digunakan untuk penelitian memiliki kualitas yang lebih baik dan lebih bersih. Lebih lanjut, implementasi *preprocessing* juga mempersiapkan data dengan lebih baik, menjadikannya siap untuk analisis lebih lanjut. Hasil dari proses ini memberikan dasar yang kuat untuk pemahaman yang lebih mendalam terhadap data cuitan mengenai calon presiden, memungkinkan penelitian berlanjut dengan lebih akurat dan informatif.

Visualisasi Data

Sebelum kita memasuki analisis lebih mendalam, mari kita pertama-tama menjelajahi hasil visualisasi dengan *wordcloud* yang mengungkapkan frekuensi kata terbanyak dalam *tweet* yang terkait dengan tiga calon presiden. *Wordcloud* memberikan visual dari teks di mana kata-kata yang paling sering muncul ditampilkan dengan ukuran teks yang lebih besar dan semakin menonjol.

Gambar 2. Visualisasi Wordcloud Capres 1

Gambar 3. Visualisasi Wordcloud Capres 2

Gambar 4. Visualisasi *Wordcloud* Capres 3

Berdasarkan visualisasi *wordcloud* di atas, dapat dilihat bahwa selain nama calon presiden dan wakil presiden, terdapat kata-kata yang sering keluar melalui *tweet* tentang pembahasan pada *keyword* yang diambil, dalam hal ini adalah nama calon presiden. Pada Gambar 2, kata yang dominan adalah "harmoni". Gambar 3 menampilkan kata "Indonesia" sebagai salah satu elemen yang sering. Sementara pada gambar 4, terdapat kata "menang" yang sering dilontarkan. Analisis lebih lanjut terhadap kata-kata ini dapat memberikan pemahaman mendalam mengenai isu-isu yang mendominasi dalam cuitan terkait tiga calon presiden.

Analisis Sentimen

Pada proses ini dilakukan pengkategorian data ke dalam kategori negatif, netral, dan positif berdasarkan pelabelan *package* vader_lexicon. Maka didapatkan hasil sebagai berikut:

Tabel 5. *Labelling* Data Capres 01

No	Label	Tweet
1	*Positive*	Pilihan jatuh anies muhaimin rekam jejak bersih bro
2	*Positive*	Tidak ragu anies Muhaimin tandem bawa perubahan positif negara
⋮	⋮	⋮
529	*Negative*	Anies Muhaimin justru capres cawapres penuh beban kelam
530	*Neutral*	Komisi pemilihan umum kpu menetapkan pasangan anies baswedan Muhaimin Iskandar ganjar pranowo Mahfud md Prabowo subianto

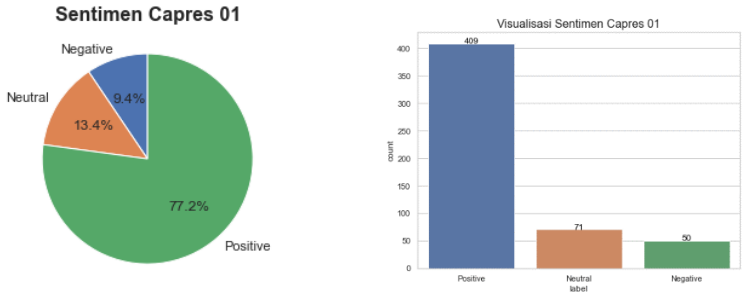
Tabel 6. *Labelling* Data Capres 02

No	Label	Tweet
1	*Positive*	Aceh Prabowo menang baik pilpres pilpres
2	*Positive*	Menyerukan negara tindakan membantu meringankan penderitaan rakyat palestina bertahan dekade prabowo prabowo subianto
⋮	⋮	⋮
1005	*Positive*	Jakarta paguyuban becak motor jawa mendukung calon presiden calon wakil presiden prabowo subianto gibran rakabuming raka pilpres prabowo subianto
1006	*Positive*	Keren habis prabowo gibran

Tabel 7. *Labelling* Data Capres 03

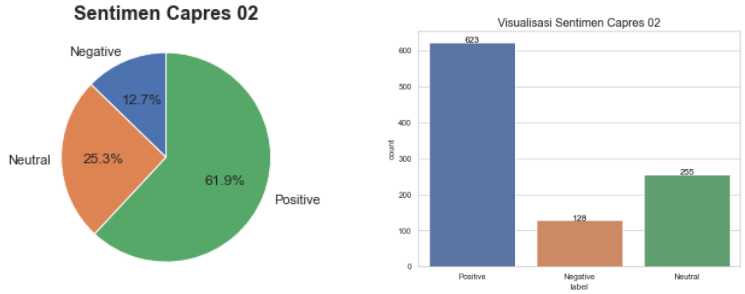
No	Label	Tweet
1	*Positive*	Mendukung penuh wacana ganjar demokrasi jujur berintegritas mari tegakkan keadilan ganjar Mahfud menang
2	*Positive*	Sangat mengapresiasi komitmen ganjar pranowo menjaga keadilan negara semoga teladan pemimpin
⋮	⋮	⋮
1847	*Positive*	Ganjar Mahfud nomor dukung ganjar Mahfud menang
1848	*Positive*	Reformasi pemimpin kuat ganjar pranowo memiliki komitmen ganjar Mahfud menang

Berikut adalah visualisasi sentimen berupa diagram lingkaran dan diagram batang dari ketiga Calon Presiden Indonesia 2024.



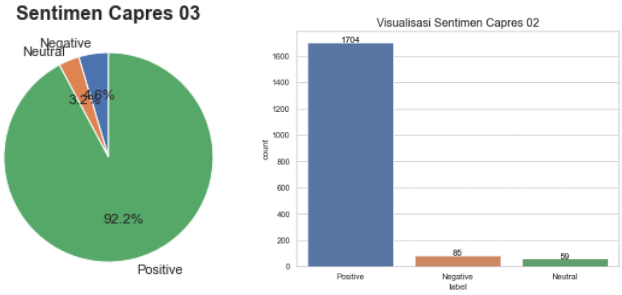
Gambar 5. Visualisasi Diagram Lingkaran dan Diagram Batang Sentimen Capres 01

Berdasarkan visualisasi diagram lingkaran dan diagram batang diatas, terlihat bahwa tanggapan dan pandangan masyarakat Indonesia terhadap Calon Presiden 2024 dengan nomor urut 01 cenderung lebih banyak pada sentimen positif. Dengan persentase positif sebesar 77,2%, atau setara dengan 409 responden memberikan tanggapan positif. Sementara itu, tanggapan netral mencapai 13,4% atau sebanyak oleh 71 responden, dan tanggapan negatif sebesar 9,4%, atau sebanyak 50 responden.



Gambar 6. Visualisasi Diagram Lingkaran dan Diagram Batang Sentimen Capres 02

Berdasarkan visualisasi diagram lingkaran dan diagram batang yang diatas, terlihat bahwa tanggapan dan pandangan masyarakat Indonesia terhadap Calon Presiden 2024 dengan nomor urut 02 cenderung lebih banyak pada sentimen positif. Dengan persentase positif mencapai 61,9% atau setara dengan 623 responden memberikan tanggapan positif. Sementara itu, tanggapan netral mencapai 25,3% atau sebanyak oleh 255 responden, dan tanggapan negatif sebesar 12,7% atau sebanyak 128 responden.



Gambar 7. Visualisasi Diagram Lingkaran dan Diagram Batang Sentimen Capres 03

Berdasarkan visualisasi diagram lingkaran dan diagram batang yang diatas, terlihat bahwa tanggapan dan pandangan masyarakat Indonesia terhadap Calon Presiden 2024 dengan nomor urut 03 lebih banyak pada sentimen positif. Dengan persentase positif mencapai 92,2%

atau setara dengan 1704 responden memberikan tanggapan positif, sementara tanggapan negatif sebesar 4,6% atau 85 responden, dan tanggapan netral sekitar 3,2% atau sebanyak oleh 59 responden.

D. Kesimpulan

Penelitian ini menggunakan uji sentiment analysis dengan algoritma VADER pada dataset yang diambil dari hasil crawling twitter dengan detail topik yang berkaitan dengan pandangan masyarakat terkait pasangan calon presiden dan wakilnya untuk pemilu yang akan datang. Dari hasil crawling pada tanggal 16 november 2023 dapat disimpulkan beberapa hal yaitu berdasarkan hasil analisis sentimen pada pasangan Anies dan Muaihmin mendapat 77,2% sentimen positif, 13,4% sentimen netral dan 9,4% sentimen negatif. Pasangan calon presiden Prabowo dan wakil presiden Gibran mendapat 61,9% sentimen positif, 25,3% sentimen netral dan 12,7% sentimen negatif. Pasangan calon presiden Ganjar dan wakil presiden Mahfud mendapat 92.2% sentimen positif, 3,2% sentimen netral dan 4,6% sentimen negatif.

Saran untuk penelitian selanjutnya yaitu Melakukan crawling data terbaru dan lebih banyak agar hasil analisis lebih sesuai dengan perkembangan kampanye terutama setelah dilakukan perdebatan antara para calon presiden. Penelitian selanjutnya dapat menggunakan beberapa metode pengujian untuk memperoleh hasil yang lebih baik. Pada algoritma VADER data tidak dilatih, sehingga diharapkan penelitian selanjutnya dilakukan pelatihan data untuk memperoleh hasil yang baik.

Acknowledge

Penulis mengucapkan terima kasih kepada Bapak Ilham Faishal Mahdy, S.Stat., M.Stat., selaku dosen pembimbing yang telah meluangkan waktu, memberikan arahan, ilmu, serta motivasi selama proses penelitian. Penulis juga menyampaikan terima kasih kepada seluruh pihak dan teman-teman yang tidak dapat disebutkan satu persatu atas dukungan dan persahabatannya.

Daftar Pustaka

[1] C. M. Annur, “Jumlah Pengguna Internet di Indonesia (Januari 2013-Januari 2023),” Katadata.co.id. [Daring]. Tersedia pada: https://databoks.katadata.co.id/datapublish/2023/09/20/pengguna-internet-di-indonesia-tembus-213-juta-orang-hingga-awal-2023

[2] M. R. Fais Sya’ bani, U. Enri, dan T. N. Padilah, “Analisis Sentimen Terhadap Bakal Calon Presiden 2024 Dengan Algoritme Naïve Bayes,” *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 2, hlm. 265, 2022, doi: 10.30865/jurikom.v9i2.3989.

[3] S. N. J. Fitriyyah, N. Safriadi, dan E. E. Pratama, “Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes,” *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 5, no. 3, hlm. 279, 2019, doi: 10.26418/jp.v5i3.34368.

[4] S. Elbagir dan J. Yang, “Language Toolkit and VADER Sentiment,” *Proceedings of the International MultiConference of Engineers and Computer Scientists (IMECS)*, vol. 0958, hlm. 12–16, 2019.

[5] H. Saif, Y. He, M. Fernandez, dan H. Alani, “Contextual semantics for sentiment analysis of Twitter,” *Inf Process Manag*, vol. 52, no. 1, hlm. 5–19, 2016, doi: 10.1016/j.ipm.2015.01.005.

[6] M. Ghiassi dan S. Lee, “A domain transferable lexicon set for Twitter sentiment analysis using a supervised machine learning approach,” *Expert Syst Appl*, vol. 106, hlm. 197–216, 2018, doi: 10.1016/j.eswa.2018.04.006.

[7] P. Amira Sumitro *dkk.*, “Analisis Sentimen Terhadapat Vaksin Covid-19 di Indonesia pada Twitter Menggunakan Metode Lexicon Based”, [Daring]. Tersedia pada: https://developer.twitter.com

[8] N. Anggraini dan H. Suroyo, “Comparison of Sentiment Analysis against Digital Payment ‘T-cash and Go-pay’ in Social Media Using Orange Data Mining,” *Journal of*

Information Systems and Informatics, vol. 1, no. 2, hlm. 152–163, 2019, doi: 10.33557/journalisi.v1i2.21.

[9] A. D. Sofia dan A. Kudus, "Pengelompokan Kabupaten/Kota berdasarkan Indikator Indeks Pembangunan Manusia 2022 Menggunakan K-Harmonic Means Clustering," *Jurnal Riset Statistika*, vol. 3, no. 2, hlm. 163–172, Des 2023, doi: 10.29313/jrs.v3i2.3130.

[10] J. W. Arisa dan N. Hajarisman, "Determinasi Derajat Kelangsungan Hidup Anak Menggunakan Multigroup Structural Equation Modeling," *Jurnal Riset Statistika*, vol. 3, no. 2, hlm. 147–154, Des 2023, doi: 10.29313/jrs.v3i2.3047.

[11] T. Azzahra, "Pemodelan Regresi Hurdle Negative Binomial pada Jumlah Kasus Difteri Jawa Barat 2020," *Jurnal Riset Statistika*, vol. 3, no. 2, hlm. 125–130, Des 2023, doi: 10.29313/jrs.v3i2.3014.