
Analisis Sentimen Pengguna Twitter Terhadap Pemilu 2024 Berbasis Model XLM-T

Mochamad Rafli Ghufro^{1*}, Muhammad Farrih Mahabbataka Arsyada², Muhammad Rizano Lukman³, Yudhistira Azhar Haryono Putra⁴, Nur Aini Rakhmawati⁵

^{1,2,3,4,5}*Institut Teknologi Sepuluh Nopember, Teknologi Elektro dan Informatika Cerdas, Sistem Informasi, Surabaya, Indonesia*

Informasi Artikel

Diterima: 21-11-2023

Direvisi: 05-12-2023

Diterbitkan: 24-12-2023

Kata Kunci

Tweet; Sentimen; Pemilu; XLM-T

*Email Korespondensi:

5026211094@student.its.ac.id

Abstrak

Di era digital seperti saat ini, media sosial, khususnya Twitter, menjadi platform penting bagi masyarakat untuk berbagi pendapat, terutama seputar isu-isu politik seperti Pemilihan Presiden (Pemilu) 2024 di Indonesia. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat mengenai Pemilu 2024 berdasarkan data *tweet* yang dikumpulkan. Dengan menggunakan model *machine learning* berbasis XLM-T, penelitian ini berhasil mengklasifikasikan *tweet-tweet* ke dalam tiga kategori sentimen: positif, negatif, dan netral dengan besar tingkat akurasi model yang sebesar 68%. Hasil menunjukkan bahwa *tweet* dengan sentimen positif dan negatif lebih sering mendapatkan interaksi dari masyarakat dibandingkan dengan *tweet* bersentimen netral, mengindikasikan kecenderungan masyarakat untuk lebih aktif berinteraksi dengan pendapat yang memiliki posisi atau pendirian tertentu mengenai sebuah isu. Kesimpulannya, analisis sentimen dapat memberikan wawasan mendalam tentang pandangan masyarakat terhadap Pemilu 2024, yang bisa dimanfaatkan oleh pemangku kepentingan politik dalam merancang strategi kampanye mereka.

Abstract

In the current digital era, social media, especially Twitter, has become an important platform for people to share opinions, especially regarding political issues such as the 2024 Presidential Election (Pemilu) in Indonesia. This research aims to analyze public sentiment regarding the 2024 Election based on collected tweet data. By using an XLM-T based machine learning model, this research succeeded in classifying tweets into three sentiment categories: positive, negative and neutral with a model accuracy rate of 68%. The results show that tweets with positive and negative sentiments receive more interaction from the public compared to tweets with neutral sentiments, indicating the public's tendency to more actively interact with opinions that have a certain position or stance on an issue. In conclusion, sentiment analysis can provide deep insight into the public's views on the 2024 Election, which political stakeholders can utilize in designing their campaign strategies.

1. Pendahuluan

Perkembangan teknologi dan internet saat ini mempermudah manusia untuk saling berkomunikasi satu sama lain melalui platform yang biasa kita sebut dengan sosial media. Menurut Website DataIndonesia.id (2023), jumlah aktif pengguna media sosial di Indonesia per Januari 2023 sebanyak 167 juta orang. Data ini sangat besar jika dibandingkan dengan pengguna internet di Indonesia per Januari 2023, yaitu 212,9 juta. Kumar, R. & Akram, W. (2018) memiliki pendapat bahwa media sosial adalah *platform online* yang digunakan orang untuk membangun jaringan sosial dan hubungan sosial dengan orang lain yang memiliki minat, aktivitas, minat pribadi atau profesional, latar belakang, atau koneksi dunia nyata yang sama.

Sebelum masuk ke inti penelitian, kita akan membahas sosial media yang sangat populer untuk adu gagasan maupun pendapat, yaitu Twitter. Aplikasi Twitter merupakan aplikasi yang di dalamnya kita bisa menuliskan pendapat kita tentang apapun. Selain itu, kita juga bisa beradu gagasan dan saling membalas *tweet* dengan pengguna lain. Yang marak di Indonesia sekarang ini, Twitter sering digunakan masyarakat untuk berdebat masalah politik, apalagi fase terparah adalah fase mendekati Pemilihan Presiden. Pemilihan Presiden Indonesia 2024 sudah tinggal menghitung bulan. Pemilihan Presiden merupakan momen terbesar yang akan mengundang banyak sekali komentar, reaksi, maupun pendapat masyarakat Indonesia. Apalagi dengan media Twitter, pasti akan dipenuhi oleh berbagai opini-opini masyarakat Indonesia, entah itu bersifat positif, negatif, maupun netral. Maka dari itu, peneliti akan mengumpulkan data-data mengenai opini masyarakat di Twitter untuk dianalisis. Opini-opini ini yang akan peneliti gunakan untuk melihat sentimen masyarakat terhadap Pemilihan Presiden khususnya pada tahun 2024.

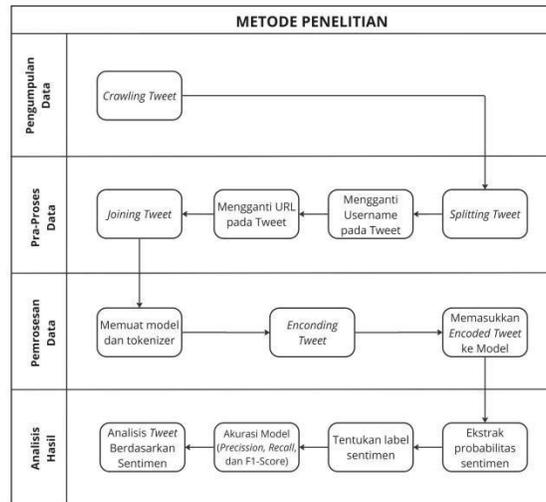
Untuk mengolah opini pada setiap *tweet*, kami menggunakan model *machine learning* bernama XLM-T, yaitu sebuah model berbasis XLM-roBERTa multibahasa yang telah dilatih dengan ~198 juta *tweet* dan disesuaikan untuk analisis sentimen (Barbieri et al., 2022, pp. 258–266). Alasan utama kami menggunakan model XLM-T karena fleksibilitasnya. XLM-T adalah model yang telah dilatih dengan beberapa macam *dataset* dari berbagai bahasa, termasuk Bahasa Indonesia sehingga tidak diperlukan konfigurasi yang rumit untuk menggunakannya. Selain itu, model XLM-T juga memiliki akurasi yang cukup tinggi dikarenakan model ini telah dilatih pada kumpulan data teks dan kode dalam jumlah yang sangat besar. Hal ini memungkinkannya untuk mencapai akurasi yang tinggi dalam mengidentifikasi sentimen *tweet*.

Sebelum penelitian ini, telah dilakukan beberapa penelitian sebelumnya yang akan menjadi rujukan peneliti. Penelitian sebelumnya dilakukan oleh Nurhuda, F., Sihwi, S. W., & Doewes, A. (2013) yang menganalisis sentimen masyarakat mendekati Pemilihan Presiden 2014 di Twitter dengan Metode *Naive Bayes Classifier*. Dengan menggunakan metode ini, pengklasifikasian data dapat dilakukan dengan cepat, akurat, dan masif. Penggunaan Metode *Naive Bayes Classifier* berguna untuk mendapatkan hasil akhir berupa persebaran *tweet* terhadap akun ketiga calon presiden yang diolah lagi menjadi persentase *tweet* positif, negatif, dan netral. Selain itu, terdapat penelitian lain juga yang juga meneliti tentang sentimen opini masyarakat di Twitter terhadap Pemilihan Presiden 2019. Penelitian ini dilakukan oleh Rozi, I. F., Pramitarini, Y., & Puspitasari, N. (2020) dengan menggunakan metode *backpropagation*. Metode *Backpropagation* adalah sebuah metode yang memaksimalkan hasil akhir dengan mengubah bobot tergantung pada kesalahan keluaran yang terjadi dan meminimalkan kesalahan yang terjadi (Munir, M. M., Fauzi, M. A., & Perdana, R. S., 2018). Data *tweet* yang terkumpul diklasifikasikan secara manual dan diukur keakuratannya dengan *backpropagation*. Namun, pada penelitian analisis sentimen opini mendekati Pemilihan Presiden 2019 di Twitter, peneliti akan menggunakan metode *web scraping* untuk metode pengumpulan datanya. Dengan *web scraping*, peneliti dapat mengumpulkan banyak data dari situs-situs web dengan waktu yang cepat.

Penelitian ini dilaksanakan dengan tujuan untuk mendapatkan hasil analisis sentimen masyarakat mendekati Pemilihan Presiden 2024 dan memberikan wawasan serta pemahaman lebih terhadap pandangan dan sikap masyarakat terkait Pemilihan Presiden 2024. Selain tujuan, peneliti juga mengharapkan agar hasil penelitian ini bermanfaat untuk praktisi politisi, kandidat presiden, maupun partai politik agar dapat merangkai strategi kampanyenya yang lebih baik dan tidak menimbulkan kegaduhan di kalangan masyarakat.

2. Metode Penelitian

2.1 XLM-T Model



Gambar 1. Alur Metode Penelitian

Analisis sentimen merupakan metode yang dilakukan untuk mengetahui opini masyarakat menjelang masa pemilu 2024, dimana tahapan-tahapan yang harus dilakukan pada metode analisis sentimen adalah pengumpulan data, pra-proses data, pemrosesan data, dan analisis data yang detailnya terdapat pada Gambar 1 di bawah ini. Setelah melakukan pengumpulan data menggunakan metode *crawling tweet*, data baru akan diproses dan dimasukkan ke dalam model XLM-T, yaitu sebuah model berbasis XLM-roBERTa multibahasa yang telah dilatih dengan ~198 juta tweet dan disesuaikan untuk analisis sentimen (Barbieri et al., 2022, pp. 258–266).

2.2 Pengumpulan Data

Untuk mengumpulkan *tweet* dengan kata kunci 'Pemilu 2024', penulis menggunakan *library* pada bahasa pemrograman Python yang bernama *tweet-harvest*. *Tweet-harvest* merupakan *library open-source* yang memungkinkan pengguna untuk mengumpulkan *tweet* dengan kata kunci tertentu dan batas jumlah *tweet* yang ditentukan oleh pengguna. Data *tweet* yang dihasilkan kemudian akan disimpan ke dalam bentuk format *file csv*.

Sebelumnya, pembuat *library* telah melakukan perbaikan pada *tweet-harvest*. Satria (2023) mengatakan bahwa dirinya telah memodifikasi dan memperbarui *library* yang ia buat sebelumnya bernama *tweet-harvest* untuk membantu teman-temannya, khususnya yang mengerjakan makalah dan penelitian.

Menggunakan *library* *tweet-harvest*, penulis berhasil mengumpulkan sebanyak total 4372 baris data *tweet* yang dikumpulkan dalam jangka waktu 10 hari, yaitu mulai dari tanggal 1 September 2023, hingga tanggal 10 September 2023 (Lukman et al., 2023). Adapun selain *tweet* pengguna, atribut lain yang berhasil dikumpulkan adalah *created_at*, *id tweet*, *quote_count*, *reply_count*, *retweet_count*, *favourite_count*, *user id*, *username*, dan *tweet_url*. Atribut-atribut tersebut nantinya kembali digunakan untuk langkah lebih lanjut yaitu analisis pada bagian hasil dan pembahasan. Pada Tabel 1 ditunjukkan sebagian atribut juga contoh sampel data yang didapatkan.

Tabel 1. Contoh Atribut beserta Sampel Data

created_at	Fri Sep 01 13:29:35 +0000 2023
full_text	@ch_chotimah2 @ganjarpranowo Apakah Ternyata permainan istana juga,kalo benar adanya pemilu terburuk sepanjang sejarah Indonesia merdeka ya di 2024,penguasa cawe-cawe seolah nggak ingin kekuasaannya lepas. Yg jadi pertanyaan nya ada apa dengan penguasa????
quote_count	0
reply_count	1
retweet_count	0
favorite_count	2
username	Zanzia6
tweet_url	https://twitter.com/Zanzia6/status/1697602605541929152

2.3 Pra-Proses Data

Pra-proses data adalah salah satu bagian dari *data mining* dan isinya meliputi persiapan, kemudian transformasi data ke sebuah bentuk yang sesuai dengan prosedur *mining*. Pra-proses data memiliki tujuan untuk mengurangi ukuran data, mengidentifikasi hubungan antar data, menormalkan data, menghilangkan *outlier* dan mengekstrak fitur untuk data (Alasadi & Bhaya, 2017, p. 4102). Berikut adalah beberapa tahapan pra-proses data yang dilakukan yang detail setiap tahapannya dapat dilihat pada Tabel 2. Tahapan Pra Proses Data:

1. *Splitting Tweet*
2. Mengganti Username pada *Tweet*
3. Mengganti URL pada *Tweet*
4. *Joining Tweet*

Tabel 2. Tahapan Pra-Proses Data

Tahapan Pra Proses Data	
Tahapan	Tweet
Raw Data	Milenial juga kedudukannya sama apalagi jumlah pemilihannya paling besar untuk pemilu 2024 @amaruf_official https://t.co/3iHxts1Cb2
Splitting Tweet	'Milenial', 'juga', 'kedudukannya', 'sama', 'apalagi', 'jumlah', 'pemilihannya', 'paling', 'besar', 'untuk', 'pemilu', '2024', '@amaruf_official', 'https://t.co/3iHxts1Cb2'
Mengganti Username pada Tweet	'Milenial', 'juga', 'kedudukannya', 'sama', 'apalagi', 'jumlah', 'pemilihannya', 'paling', 'besar', 'untuk', 'pemilu', '2024', '@user', 'https://t.co/3iHxts1Cb2'
Mengganti URL pada Tweet	'Milenial', 'juga', 'kedudukannya', 'sama', 'apalagi', 'jumlah', 'pemilihannya', 'paling', 'besar', 'untuk', 'pemilu', '2024', '@user', 'http'
Joining Tweet	Milenial juga kedudukannya sama apalagi jumlah pemilihannya paling besar untuk pemilu 2024 @user http

2.4 Pemrosesan Data

Pemrosesan data yang dilakukan melibatkan beberapa langkah sesuai dengan algoritma yang telah dibuat dalam bahasa pemrograman Python. Adapun langkah-langkah tersebut adalah

1. Memuat model dan tokenizer
2. *Encoding Tweet*
3. Memasukkan *Encoded Tweet* ke Model

Pada algoritma yang dibuat oleh penulis, pada Gambar 2 merupakan potongan kode yang digunakan untuk memuat model dan tokenizer.

```
# load model and tokenizer
roberta = "cardiffnlp/twitter-xlm-roberta-base-sentiment"

model = AutoModelForSequenceClassification.from_pretrained(roberta)
tokenizer = AutoTokenizer.from_pretrained(roberta)
```

Gambar 2. Potongan Kode untuk Memuat Model dan Tokenizer

Kemudian pada tahapan selanjutnya, tweet yang telah melalui tahapan pra-proses data akan di *encoding*. *Encoding* memiliki tujuan untuk mengubah bentuk atau format dari sebuah data (Afrianto & Taliasih, 2020). Tabel 3 berikut merupakan contoh input dan *output* dari proses *encoding*.

Tabel 3. Proses Encoding Tweet

Proses Encoding	
Input	Milenial juga kedudukannya sama apalagi jumlah pemilihannya paling besar untuk pemilu 2024 @user http
Output	[0, 52999, 93, 289, 1220, 73726, 273, 1847, 94676, 19836, 119252, 273, 7392, 4693, 482, 32490, 12175, 387, 2357, 1374, 65918, 1621, 2]

Tahapan terakhir dalam pemrosesan data adalah memasukkan hasil *encoding tweet* tersebut ke dalam model yang sebelumnya telah dimuat, sehingga model dapat melakukan analisis sentimen terhadap input masukan.

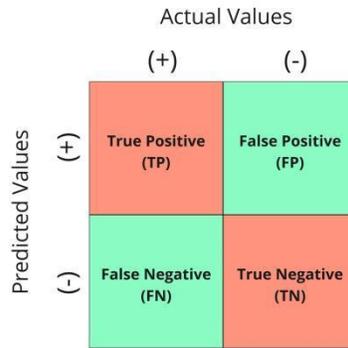
2.5 Analisis Hasil

Output yang dihasilkan oleh tahapan pemrosesan data adalah jumlah persentase sentimen *tweet* yang telah dimasukkan ke model sebelumnya. Persentase tersebut kemudian terbagi menjadi 3 kelas, yaitu sentimen positif, negatif, dan netral. Penentuan label pada setiap tweet berdasarkan persentase terbesar dari tiga kelas tersebut. Adapun contoh hasil penentuan label sentimen ditunjukkan pada Tabel 4.

Tabel 4. Hasil Analisis Sentimen Tweet

Tweet	Persentase			Label
	Positif	Negatif	Netral	
Kalau setiap pemilu yang dibahas suara di Jawa doang, ini indikasi harus ada perubahan di mekanisme perhitungan suara. Kasian kan yang diberitain Jawa Tengah Jawa Timur doang. ô¥ÿ, #Pemilu2024	0,019555	0,876151	0,104294	Negatif
Milenial juga kedudukannya sama apalagi jumlah pemilihannya paling besar untuk pemilu 2024 @amaruf_official https://t.co/3iHxts1Cb2	0,214214	0,100841	0,684945	Netral
Pemilu 2024 sukses dan lancar !! ayo Pilih ya #dukungpemilu2024 #SukseskanPemilu2024 #Pemilu2024kondusif https://t.co/mYOadHgt70	0,76006	0,049047	0,190893	Positif

Selain itu, penting juga untuk melakukan evaluasi performa terhadap hasil analisis sentimen yang dilakukan. Metode yang sering digunakan untuk melakukan validasi terhadap hasil analisis sentimen salah satunya adalah alat evaluasi metrik untuk *machine learning* yang bernama *F1-score*. Metrik evaluasi yang efektif sangat penting dalam menilai kinerja model *machine learning*. Salah satu metrik tersebut adalah *F1-Score*, yang banyak digunakan untuk masalah klasifikasi, pengambilan informasi, dan tugas-tugas NLP (Logunova, 2023). Pada evaluasi performa ini, metrik-metrik yang kemudian digunakan disini adalah *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Untuk menghitung metrik-metrik tersebut diperlukan beberapa komponen pada *Confusion Matrix* yang ditunjukkan pada Gambar 3, yaitu TP, FP, FN dan TN.



Gambar 3. Confusion Matrix

Untuk mengukur keakuratan suatu model, pengukuran dilakukan dengan cara membandingkan data yang diklasifikasikan dengan data secara keseluruhan (Rivaldi et al., 2022).

$$Accuracy = \frac{\text{correct classification}}{\text{all classification}} \quad (1)$$

Untuk mengukur *F1-Score*, dibutuhkan 2 metrik lainnya, yaitu *Precision* dan *Recall*. Adapun *Precision* adalah pengukuran proporsi prediksi "positif" yang benar yang dibuat oleh model (Logunova, 2023).

$$Precision = \frac{TP}{(TP + FP)} \quad (2)$$

Sedangkan *Recall* mengukur proporsi sampel kelas positif yang diidentifikasi dengan benar oleh model (Logunova, 2023).

$$Recall = \frac{TP}{(TP + FN)} \quad (3)$$

Apabila kedua metrik kedua metrik sudah didapatkan, selanjutnya *F1-Score* dapat dihitung.

$$F1 - Score = \frac{2 \cdot Precision \cdot Recall}{(Precision + Recall)} \quad (4)$$

F1-Score mempunyai rentang nilai 0 hingga 1. Semakin tinggi *F1-Score* maka semakin baik pula kualitas pengklasifikasi, dan sebaliknya.

3. Hasil dan Pembahasan

Pada proses pengumpulan data, didapatkan total jumlah sebanyak 4372 *tweet* dengan persebaran dapat dilihat pada Tabel 5. Selanjutnya, diambil sampel dari hasil analisis sentimen sebanyak 251 *tweet* yang diambil dengan cara *random sampling* untuk dilakukan pelabelan sentimen secara manual. Sebanyak 251 *tweet* tersebut digunakan sebagai perbandingan dengan hasil analisis sentimen oleh model XLM-T. Hasil perbandingan tersebut yang nantinya digunakan sebagai ukuran performa model yang hasilnya ditunjukkan pada Tabel 6.

Tabel 5. Jumlah Tweet Beserta Label

Label	Netral	Positif	Negatif
Jumlah Tweet	1701	2097	574

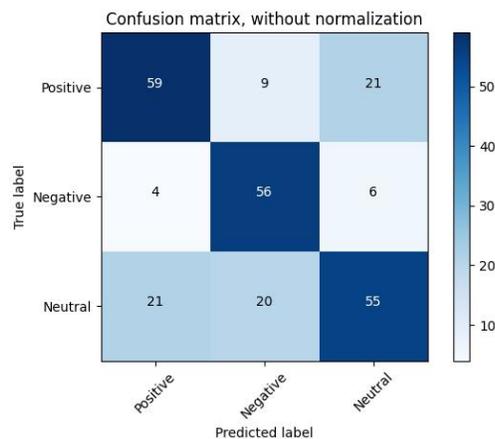
Berdasarkan perhitungan yang dilakukan, didapatkan akurasi model XLM-T yang digunakan untuk analisis sentimen pada kasus ini adalah sebesar 68%. Pada tabel 6 juga dapat dilihat performa untuk setiap sentimen. Adapun dari 251 *tweet* yang digunakan sebagai perbandingan, didapatkan bahwa sentimen yang paling baik

adalah sentimen negatif dengan *F1-Score* sebesar 0.74. Hal tersebut juga terlihat dari metrik lainnya, yaitu *Precision* dan *Recall* yang cenderung tinggi dibanding dengan sentimen lainnya.

Tabel 6. Performa Model

Performa Model				
Label	Precision	Recall	F1-Score	
Netral		0.67	0.57	0.61
Positif		0.70	0.66	0.68
Negatif		0.65	0.84	0.74

Pada gambar 4 dapat dilihat bahwa *tweet* dengan label sentimen negatif paling sedikit mengalami kesalahan klasifikasi oleh model XLM-T, yaitu hanya 10 *tweet*. Sedangkan *tweet* bersentimen netral adalah *tweet* yang paling banyak salah diklasifikasikan oleh model, kemudian disusul oleh *tweet* yang memiliki sentimen positif. *Tweet* yang bersentimen positif, model salah mengklasifikasikan 30 *tweet* dengan 9 *tweet* diklasifikasikan ke dalam label negatif dan 21 *tweet* ke dalam label netral. Adapun sebanyak 41 *tweet* bersentimen netral yang salah diklasifikasikan oleh model, sebanyak 21 *tweet* diklasifikasikan ke dalam label positif dan 20 *tweet* ke dalam label negatif.



Gambar 4. Confusion Matrix

Hasil dari perhitungan didapatkan dari data *tweet* yang dikumpulkan dalam jangka waktu 10 hari, yaitu mulai dari tanggal 1 September 2023, hingga tanggal 10 September 2023. 4 atribut yang menjadi fokus disini adalah total *quote*, total *reply*, total *retweet*, dan total *like*.

Atribut total *quote* akan menunjukkan seberapa berdampak sebuah *tweet* sehingga dibagikan kembali dan ditambah dengan tanggapan baru dari akun yang melakukan *quote* *tweet*. Dapat dilihat *tweet* dengan sentimen positif memiliki total *quote* paling banyak yaitu 400 dengan rata-rata terbesar juga yaitu 0,190748689 yang ditunjukkan pada tabel 8. Sedangkan untuk sentimen negatif memiliki total *quote* paling sedikit dengan 101 *quote*, tetapi rata-ratanya lebih besar daripada sentimen netral. Sentimen netral pada tabel 9 memiliki total *quote* sebesar 263 dengan rata-rata 0,154614932, sedangkan sentimen negatif memiliki rata-rata 0,175958188. Hal ini menunjukkan masyarakat lebih mudah untuk melakukan *quote tweet* pada *tweet* yang bersentimen negatif, dibuktikan dengan meskipun jumlah *tweet* sentimen negatif paling sedikit namun rata-rata *quote tweet* nya lebih besar daripada sentimen netral dan hanya berjarak sedikit dengan sentimen positif.

Atribut total *reply* menunjukkan seberapa banyak *tweet* itu dibalas oleh orang lain. Dengan kata lain, atribut ini juga menunjukkan seberapa *relate tweet* itu dengan kondisi lingkungan sekitar, maupun kondisi para pengguna Twitter. Statistik *tweet* bersentimen positif memiliki total *reply* terbanyak sebanyak 2.112 *reply*, dengan rata-rata 1,007153076. Selanjutnya, total *reply* dari *tweet* bersentimen netral yang lebih sedikit dari

data sebelumnya, yaitu 1183 *reply*, dengan rata-rata 0,695473251. Terakhir, total *reply* yang paling sedikit adalah dari *tweet* yang bersentimen negatif, yaitu sebanyak 513 *reply*, dengan rata-rata 0,893728223 yang terlihat pada tabel 7. Jika membandingkan melalui rata-rata total *reply* dari masing-masing *tweet* yang memiliki sentimen positif, negatif, dan netral, perbandingan tertinggi berada di sentiment positif. Ini berarti, bahwa *tweet* yang bersentimen positif lebih menarik orang untuk membalas dan berdiskusi lebih lanjut terkait *tweet* tersebut.

Atribut total *retweet* menunjukkan seberapa pentingnya sebuah *tweet* untuk dibagikan kembali tanpa harus ditambahkan apapun. Tujuan pembagian kembali ini bisa menjadi tanda bahwa setuju dengan *tweet* tersebut atau merasa bahwa harus lebih banyak orang yang membaca *tweet* tersebut. Bisa dilihat pada tabel 8, *tweet* dengan sentimen positif memiliki total *retweet* terbanyak dengan 5.517 dan rata-rata terbesar juga dengan 2,630901288. Rata-rata terbesar kedua adalah *tweet* dengan sentimen negatif, meskipun memiliki jumlah *retweet* lebih kecil daripada sentimen netral. Dengan data tersebut menunjukkan masyarakat cenderung sering membagikan kembali *tweet* dengan sentimen positif atau negatif daripada netral. Hal ini akan menunjukkan keberpihakan masyarakat juga.

Atribut total *like* menunjukkan seberapa banyak sebuah *tweet* disukai oleh masyarakat luas. Hasilnya tidak jauh berbeda dengan atribut lain, *tweet* sentimen positif memiliki total *like* terbesar dengan 13.437 dan rata-rata sebesar 6,407725322. *Tweet* sentimen negatif pada tabel 7 meskipun memiliki total *like* terendah namun memiliki rata-rata lebih besar dari *tweet* sentimen netral.

Index Sentimen	Total Quote			Total Reply			Total Retweet			Total Like		
	Negatif	Positif	Netral	Negatif	Positif	Netral	Negatif	Positif	Netral	Negatif	Positif	Netral
Mean	0.175958188	0.190748689	0.154614932	0.893728223	1.007153076	0.695473251	1.210801394	2.630901288	1.038800705	3.716027875	6.407725322	3.577307466
Standard Error	0.044223111	0.042136207	0.045339327	0.33739424	0.262811154	0.178867899	0.426633569	0.445696995	0.164719867	0.961833	1.134523781	0.665442149
Median	0	0	0	0	0	0	0	0	0	0	0	0
Mode	0	0	0	0	0	0	0	0	0	0	0	0
Standard Deviation	1.059510426	1.929543874	1.869938073	8.083391438	12.03491447	7.377081168	10.22141381	20.40980808	6.793571319	23.04388077	51.95326178	27.44495115
Sample Variance	1.122562344	3.723139563	3.496668396	65.34121714	144.8391664	54.42132656	104.4773002	416.560266	46.15261127	531.0204407	2699.14141	753.2253436
Kurtosis	56.58467384	902.0611768	1028.478728	275.5224079	483.2872449	644.3359134	412.4445141	264.9142955	170.6891629	322.7680275	306.1669413	399.3587779
Skewness	7.409937951	26.6097132	29.41088111	16.46313204	21.1233288	23.97177949	19.19872251	14.18852663	12.03628535	16.35727045	16.08680505	17.9784372
Range	9	71	68	142	314	221	226	517	117	480	1264	749
Minimum	0	0	0	0	0	0	0	0	0	0	0	0
Maximum	9	71	68	142	314	221	226	517	117	480	1264	749
Sum	101	400	263	513	2112	1183	695	5517	1767	2133	13437	6085
Count	574	2097	1701	574	2097	1701	574	2097	1701	574	2097	1701
Largest(1)	9	71	68	142	314	221	226	517	117	480	1264	749
Smallest(1)	0	0	0	0	0	0	0	0	0	0	0	0

Gambar 5. Statistik Deskriptif Sentimen Negatif, Positif, Netral

4. Kesimpulan

Penelitian ini menganalisis karakteristik *tweet* di platform Twitter berdasarkan sentimen *tweet* yang berkaitan dengan Pemilu 2024. Data yang diperoleh dari *tweet* masyarakat Indonesia yang berkaitan dengan Pemilu 2024 kemudian dianalisis untuk menentukan sentimen dari setiap *tweet* yang diperoleh. Selain itu, kami membuat algoritma pada bahasa pemrograman Python dan menggunakan model *machine learning* XLM-T untuk melakukan proses klasifikasi sentimen pada *tweet*. Menggunakan metode *web scraping*, penelitian ini berhasil mendapatkan total sebanyak 4372 *tweet*. Berdasarkan analisis sentimen yang dilakukan, didapatkan bahwa *tweet* yang bersentimen netral adalah sebanyak 39%, positif 48%, dan negatif 13% sehingga dapat disimpulkan bahwa kecenderungan sentimen pengguna Twitter terhadap Pemilu 2024 adalah positif. Adapun evaluasi performa model yang dilakukan terhadap hasil analisis sentimen menunjukkan bahwa model memiliki tingkat akurasi sebesar 68% dengan performa klasifikasi terbaik adalah sentimen negatif berdasarkan *F1-Score* yang didapatkan.

Detail lain yang didapatkan adalah bahwa *tweet* dengan sentimen positif dan negatif memiliki kecenderungan lebih ramai ditanggapi oleh masyarakat Indonesia. Hal ini dibuktikan dari total jumlah dan rata-rata tiap atribut. Masyarakat lebih sering berinteraksi dengan *tweet* yang tidak bersentimen netral. Hal ini mungkin didasari dari masyarakat yang mempunyai pendapat atau pendirian sendiri mengenai sebuah topik. Penelitian ini diharapkan dapat memberikan landasan yang kuat untuk pemahaman lebih dalam terhadap

dinamika opini dan respons masyarakat Indonesia terhadap Pemilihan Presiden 2024 di platform media sosial, khususnya Twitter. Informasi yang diperoleh dari analisis sentimen ini dapat menjadi sumber berharga bagi pihak terkait, seperti pengambil kebijakan, kandidat, atau lembaga survei, untuk memahami lebih baik keinginan, kekhawatiran, serta aspirasi masyarakat dalam konteks perhelatan politik mendatang. Dengan demikian, penelitian ini berpotensi memberikan kontribusi signifikan dalam memandu upaya-upaya untuk membangun pemilu yang lebih inklusif dan responsif terhadap harapan serta keprihatinan masyarakat Indonesia.

5. Referensi

- Afrianto, I., & Taliasih, N. (2020): *Sistem Keamanan Basis Data Klien P.T. Infokes Menggunakan Kriptografi Kombinasi RC4 Dan Base64*. Jurnal Nasional Teknologi Dan Sistem Informasi, 6(1), 9–18. <https://doi.org/10.25077/teknosi.v6i1.2020.9-18>
- Ahmadi, A. (2015): *Analisis Konstruksi Hukum Konstitusionalitas Pemilu Serentak Pada Tahun 2019*. Al-'Adl, 8(1), 1-19
- Akram, W., & Kumar, R. (2017): *A Study on Positive and Negative Effects of Social Media on Society*. International Journal of Computer Sciences and Engineering, 5(10), 351–354. <https://doi.org/10.26438/ijcse/v5i10.351354>
- Alasadi, S. A., & Bhaya, W. S. (2017). Review of data preprocessing techniques in data mining. *Journal of Engineering and Applied Sciences*, 12(16), 4102-4107.
- Barbieri, F., Anke, L. E., & Camacho-Collados, J. (2022): *XLM-T: Multilingual Language Models in Twitter for Sentiment Analysis and Beyond* (pp. 258–266). European Language Resources Association. <https://aclanthology.org/2022.lrec-1.27/>
- Indonesia, Data. "Pengguna Media Sosial Di Indonesia Sebanyak 167 Juta Pada 2023." *DataIndonesia.id*, 3 Feb. 2023, dataIndonesia.id/internet/detail/pengguna-media-sosial-di-indonesia-sebanyak-167-juta-pada-2023. Diakses pada 19 September 2023, dari <https://dataIndonesia.id/internet/detail/pengguna-media-sosial-di-indonesia-sebanyak-167-juta-pada-2023>
- Logunova, I. (2023, July 11): *F1 Score in Machine Learning*. Serokell Software Development Company. Diakses pada 26 September 2023, dari <https://serokell.io/blog/a-guide-to-f1-score>
- Muhammad Rizano Lukman, Muhammad Farrih Mahabbataka Arsyada, Mochamad Rafli Ghufron, Yudhistira Azhar Haryono Putra, & Nur Aini Rakhmawati. (2023). *Dataset Analisis Sentimen Twitter Pada Pemilu 2024 (1.0) [Data set]*. Zenodo. <https://doi.org/10.5281/zenodo.8392888>
- Munir, M., Fauzi, M., & Perdana, R. (2018): *Implementasi Metode Backpropagation Neural Network berbasis Lexicon Based Features dan Bag of Words Untuk Identifikasi Ujaran Kebencian Pada Twitter*. Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer, 2(10), 3182–3191.
- Nurhuda, F., Widya Sihwi, S., & Doewes, A. (2016). *Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier*. Jurnal Teknologi & Informasi ITSsmart, 2(2), 35. <https://doi.org/10.20961/its.v2i2.630>
- Rivaldi, A. A., Azra, B., Ziaulhaq, Y. I., & Rakhmawati, N. A. (2022): *Analisis Karakteristik Akun Twitter Berdasarkan Sentimen Pendapat Terkait Undang-Undang PSE*. SATIN – Sains Dan Teknologi Informasi, 8(2). <https://doi.org/10.33372/stn.v8i2.876>
- Rozi, I. F., Pramitarini, Y., & Puspitasari, N. (2020): *Analisis Mengenai Calon Presiden Indonesia 2019 Di Twitter Menggunakan Metode Backpropagation*. Jurnal Informatika Polinema, 6(2).
- Vimal B.(2020). *Application of Logistic Regression in Natural Language Processing [Online]*. www.ijert.org .