

Perbandingan Performa Algoritma *VADER* dan *RoBERTa* pada *Twitter*

Comparison of VADER and RoBERTa Algorithm Performance on Twitter

¹Dita Nurmadewi*, ²Zakiul Fahmi Jailani, ³Ni Kadek Srimanik
^{1,2,3}Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Bakrie
Jl.H.R. Rasuna Said Kav C-22, Kuningan, Jakarta Selatan, Indonesia
*e-mail: dita.nurmadewi@bakrie.ac.id

(received: 31 May 2024, revised: 24 June 2024, accepted: 19 July 2024)

Abstrak

Penelitian ini membandingkan performa dua algoritma analisis sentimen, yaitu *VADER* (*Valence Aware Dictionary and sEntiment Reasoner*) dan *RoBERTa* (*Robustly Optimized BERT Pretraining Approach*), menggunakan dataset opini masyarakat terkait perubahan iklim di twitter. Analisis dilakukan untuk menentukan distribusi sentimen dari tweet yang dianalisis, apakah bersifat positif, negatif, atau netral. Selain itu, penelitian ini mengidentifikasi kata kunci yang paling sering muncul dari kumpulan tweet yang telah dianalisis. Analisis time series juga dilakukan untuk melihat distribusi sentimen selama 12 bulan. Keterhubungan antara kedua model tersebut dievaluasi menggunakan scatter plot matrix untuk tweet per dua bulan, guna menilai korelasi dan konsistensi hasil analisis sentimen antara *VADER* dan *RoBERTa*. Hasil penelitian menunjukkan bahwa *VADER* lebih efektif dalam situasi yang memerlukan tanggapan cepat terhadap perubahan sentimen publik, sementara *RoBERTa* lebih unggul dalam analisis mendalam pada konten yang lebih kompleks dan ambigu.

Kata kunci: algoritma, performa, *VADER*, *RoBERTa*, twitter

Abstract

This research compares the performance of two sentiment analysis algorithms, namely VADER (Valence Aware Dictionary and sEntiment Reasoner) and RoBERTa (Robustly Optimized BERT Pretraining Approach), using a dataset of public opinions regarding climate change on twitter. Analysis is carried out to determine the sentiment distribution of the tweets described, whether they are positive, negative or neutral. In addition, this research identifies the keywords that appear most frequently from the collection of tweets that have been analyzed. Time series analysis was also carried out to see the distribution of sentiment over 12 months. The relationship between the two models was evaluated using matrix scatter plot analysis for tweets per two months, to assess the correlation and consistency of sentiment results between VADER and RoBERTa. The results show that VADER is more effective in situations that require rapid responses to changes in public sentiment, while RoBERTa is superior in in-depth analysis of more complex and ambiguous content.

Keywords: algorithm, performance, *VADER*, *RoBERTa*, twitter

1 Pendahuluan

Dalam era kemajuan teknologi saat ini, terdapat banyak data yang bisa diperoleh, yang menjadi tantangan dan juga peluang bagi institusi. Data dapat menjadi masalah jika sebuah institusi tidak mampu menyimpan, mengelola, dan memprosesnya dengan baik [1]. Sebaliknya, data menjadi peluang jika dimanfaatkan untuk mengidentifikasi tren atau pola yang bisa digunakan untuk mendapatkan informasi di masa depan [2]. Data mining adalah sebuah teknik untuk menelusuri data guna membangun model, yang kemudian digunakan untuk mengenali pola data lain yang tidak terdapat dalam basis data yang tersimpan [3]. Data mining digunakan sebagai bidang studi yang bertujuan untuk menemukan, menggali, atau menambang pengetahuan dari data yang ada, sehingga sering disebut sebagai bagian dari *Knowledge Discovery in Database* (KDD) [4]. KDD mencakup beberapa tahapan, yaitu

pembersihan data, integrasi data, seleksi data, data mining, evaluasi model, dan penyajian pengetahuan menggunakan teknik visualisasi [5]. Proses kerja dalam data mining terbagi menjadi empat kategori, yaitu pemodelan prediksi, analisis klaster, analisis asosiasi, dan deteksi anomali [6]. Dalam model prediksi, terdapat dua jenis model, yaitu klasifikasi dan regresi. Klasifikasi merupakan teknik dalam data mining yang digunakan untuk membuat model dari data sampel yang belum diklasifikasikan, dengan tujuan mengelompokkan data sampel baru ke dalam kategori yang sesuai [7]. Klasifikasi termasuk dalam *supervised learning* karena model dikembangkan berdasarkan kumpulan data yang dianalisis terlebih dahulu. Pola yang diperoleh dari analisis ini kemudian diterapkan untuk mengklasifikasikan data testing. *Data training* dianalisis terlebih dahulu menggunakan algoritma klasifikasi, sementara *data testing* digunakan untuk mengevaluasi tingkat akurasi dari aturan klasifikasi yang diterapkan [8]. Terdapat beberapa metode klasifikasi, seperti VADER dan RoBERTa.

VADER (*Valence Aware Dictionary and sEntiment Reasoner*) adalah salah satu algoritma yang dirancang khusus untuk analisis sentimen teks pendek seperti tweet. VADER dikenal karena kemampuannya memberikan hasil yang cepat dan cukup akurat tanpa memerlukan pelatihan yang kompleks. Algoritma ini berbasis pada kamus kata yang telah diberi nilai sentimen dan aturan linguistik untuk menginterpretasikan teks. Oleh karena itu, VADER sangat efektif dalam situasi yang membutuhkan respon cepat terhadap perubahan sentimen publik [9]. Di sisi lain, RoBERTa (*Robustly Optimized BERT Pretraining Approach*) adalah model yang juga canggih dan kuat, hasil dari pengembangan lebih lanjut dari model BERT. RoBERTa mampu menangani konteks yang lebih kompleks dan memberikan analisis yang lebih mendalam melalui teknik *pretraining*. Model ini menggunakan jaringan saraf tiruan yang telah dilatih secara ekstensif pada berbagai jenis teks untuk memahami konteks dan nuansa bahasa dengan lebih baik [10].

Media sosial merupakan media berbasis internet yang memiliki peran penting yaitu sebagai sarana efektif dalam memenuhi kebutuhan informasi yang bersifat publik. Pengguna media sosial juga dapat memposting status dan komentar dalam bentuk *text*, *audio*, foto, dan lain-lain [11]. Pemanfaatan media sosial sebagai sumber informasi dapat memengaruhi sikap atau pandangan seseorang. Informasi yang diperoleh dari media sosial dapat meningkatkan pengetahuan dan pemahaman individu terhadap suatu isu, baik itu setuju maupun tidak setuju [12]. Salah satu media sosial yang banyak diakses oleh masyarakat adalah twitter, yang berganti nama menjadi X pada akhir Juli 2023. Menurut data dan laporan dari We Are Social, pengguna twitter pada tahun 2022 mencapai lebih dari 18 juta orang. Hal ini menunjukkan bahwa twitter masih banyak digunakan oleh masyarakat. Twitter adalah jejaring sosial atau media sosial yang dapat menghubungkan antar pengguna untuk dapat saling berkomunikasi atau dapat digunakan untuk penyebaran informasi [13]. Twitter telah menjadi salah satu platform media sosial yang populer dan memiliki potensi besar untuk digunakan dalam pengumpulan data mengenai isu atau opini masyarakat [14], [15]. Dengan demikian, data twitter dapat digunakan untuk menganalisis pandangan dan opini masyarakat mengenai berbagai isu, termasuk perubahan iklim. Isu perubahan iklim akan menjadi fokus data set dalam penelitian ini, yang merupakan hasil *crawling* di twitter selama 12 bulan.

Penelitian ini akan menggunakan dataset opini masyarakat terkait perubahan iklim di Twitter. Perbandingan performa dua algoritma analisis sentimen, yaitu VADER dan RoBERTa, akan dilakukan. Selain menentukan distribusi sentimen dari *tweet* yang dianalisis, penelitian ini juga akan mengidentifikasi kata kunci yang paling sering muncul dan melakukan analisis *time series* untuk melihat distribusi sentimen selama 12 bulan. Evaluasi keterhubungan antara kedua model tersebut akan dilakukan menggunakan *scatter plot matrix* untuk *tweet*, guna menilai korelasi dan konsistensi hasil analisis sentimen antara VADER dan RoBERTa.

2 Tinjauan Literatur

Bagian ini mengulas kajian pustaka, teori dasar, dan penelitian sebelumnya yang berkaitan dengan masalah penelitian. Kajian pustaka tersebut akan menjadi dasar dalam pelaksanaan penelitian ini. Bagian ini juga mencakup penjelasan tentang konsep media sosial twitter, serta algoritma VADER dan RoBERTa.

2.1 Media Sosial Twitter

Penelitian terdahulu menunjukkan bahwa media sosial, termasuk twitter memiliki pengaruh yang signifikan dalam membentuk opini masyarakat. Banyak pengguna twitter yang aktif memberikan pendapat atau opini mereka tentang isu-isu publik seperti politik, sosial, dan lingkungan. *Tweet-tweet* ini dapat mempengaruhi pandangan dan persepsi pengguna twitter lainnya tentang suatu isu tertentu, dan dapat membentuk opini publik yang lebih luas [16]. Penelitian lain juga menemukan bahwa media sosial dapat membentuk opini publik, memperkuat opini publik yang sudah ada, atau bahkan memperlemah opini publik [17]. Hal ini dikarenakan twitter dapat mempengaruhi norma sosial dan identitas kelompok melalui penggunaan tagar (*hashtag*) dan grup (*group*) tertentu yang digunakan oleh pengguna untuk memperkuat pandangan dan opini mereka. Selain itu, *tweet* yang dikirim oleh pengguna dapat mempengaruhi persepsi dan emosi pengguna twitter lainnya. Twitter adalah salah satu *platform* media sosial yang digunakan secara luas dan cepat, sehingga dapat memengaruhi budaya politik dan opini publik tentang berbagai isu. Faktor lingkungan seperti media massa dapat mempengaruhi opini publik di twitter, terutama melalui penggunaan link dan sumber berita yang dibagikan oleh pengguna twitter [18].

2.2 Algoritma VADER (*Valence Aware Dictionary and sEntiment Reasoner*)

Pada penelitian terdahulu, VADER digunakan untuk menganalisis sentimen publik pada *tweet* tentang PeduliLindungi. Penelitian ini menunjukkan bagaimana VADER dapat efektif digunakan untuk memperoleh wawasan yang lebih mendalam mengenai opini publik tentang topik medis dan kesehatan, serta menekankan pentingnya analisis sentimen dalam bidang kesehatan [19]. Penelitian lain bertujuan untuk mengetahui persepsi pelanggan terhadap produk dan layanan yang memberikan kontribusi signifikan, dimana algoritma VADER mencapai tingkat akurasi 88,57% untuk menghasilkan kinerja analisis opini terbaik [20]. Terdapat juga penelitian yang membandingkan kinerja klasifikasi proses anotasi leksikon VADER dengan anotasi manual pada kumpulan data dari Amazon untuk menentukan tingkat orientasi pelanggan, dimana penggunaan VADER dalam proses anotasi menghasilkan nilai akurasi yang unggul jika di banding manual [21].

2.3 Algoritma RoBERTa (*Robustly Optimized BERT Pretraining Approach*)

Dalam konteks aplikasi kesehatan mental, khususnya untuk deteksi perasaan depresi sementara, RoBERTa, telah dimanfaatkan untuk menganalisis data dari twitter. Model ini efektif dalam mengidentifikasi nuansa emosional dan kontekstual dari teks yang berkaitan dengan indikasi depresi. Kinerja RoBERTa dalam memahami nuansa bahasa alami menunjukkan kapasitasnya yang signifikan dalam memberikan wawasan yang mendalam tentang kondisi emosional pengguna, yang tercermin dari interaksi mereka di media sosial [22]. Selanjutnya, RoBERTa juga diterapkan dalam penelitian analisis sentimen terhadap pandangan publik mengenai profesi PNS (Pegawai Negeri Sipil) berdasarkan data *tweet*. Penelitian ini mengeksplorasi bagaimana RoBERTa dapat membantu memahami perspektif masyarakat tentang profesi PNS di twitter, dengan melalui proses *scraping*, *pre-processing* teks, hingga pelabelan dalam tiga kategori berbeda. Hasil penelitian menunjukkan bahwa algoritma RoBERTa memiliki akurasi yang sudah baik, dengan rata-rata akurasi keseluruhan mencapai 90% [23].

3 Metode Penelitian

Penelitian ini bersifat analitis dan bertujuan untuk memberikan pemahaman yang lebih mendalam mengenai kinerja algoritma VADER dan RoBERTa dalam menganalisis data opini publik tentang perubahan iklim di twitter. Data yang digunakan dalam penelitian adalah *tweet* yang mengandung kata kunci terkait perubahan iklim dan SDGs 13 hanya pada tahun 2022 saja. Tahapan penelitian ini ditunjukkan pada gambar 1.



Gambar 1. Tahapan penelitian

Pada gambar 1, tahapan dimulai dari pengumpulan data. Pengumpulan data dari twitter menggunakan twitter API (*Application Programming Interface*) dan penyaringan *tweet* berdasarkan

kata kunci terkait perubahan iklim dan SDGs 13. Data yang diperoleh kemudian disimpan dalam format CSV(*Comma-Separated Value*). Data digunakan sebagai data set untuk menguji performa algoritma VADER dan RoBERTa. Langkah berikutnya yaitu pra pemrosesan data, di mana data *tweet* dibersihkan, termasuk penghapusan tanda baca dan karakter yang tidak relevan, serta transformasi semua teks menjadi huruf besar. Pada tahap ini juga dilakukan *stemming* dan *lemmatization*. Untuk membantu mengurangi variasi bentuk kata yang berbeda-beda namun memiliki akar kata yang sama, sehingga memudahkan analisis teks dan meningkatkan efisiensi model dalam memproses data. Berikutnya dilakukan analisis frekuensi kata yang paling banyak untuk mengetahui bahwa data yang di ambil sudah relevan dengan topik atau isu yang dibutuhkan.

Langkah berikutnya dilakukan penerapan model analisis sentimen RoBERTa dan VADER untuk mengklasifikasikan sentimen dari *tweet-tweet* yang terkumpul. Model ini menghasilkan tiga kategori sentimen, yaitu negatif, netral, dan positif, berdasarkan evaluasi konten teks dari masing-masing *tweet*. Langkah berikutnya yaitu mengembangkan visualisasi grafis untuk menampilkan distribusi sentimen positif, negatif, dan netral sepanjang waktu tertentu, memberikan gambaran jelas tentang dinamika emosional dalam diskursus publik tentang perubahan iklim. Pendekatan ini tidak hanya memperkuat analisis kuantitatif dengan visualisasi yang menarik, tetapi juga memudahkan pemahaman yang kompleks terhadap interaksi dalam media sosial terkait isu perubahan iklim. Langkah terakhir yaitu analisis komparatif untuk menilai efektivitas dari dua algoritma analisis sentimen dengan melihat hasil visualisasi menggunakan *scatter plot*, yang digunakan untuk memproses data *tweet* per dua bulan. Analisis ini bertujuan untuk mengidentifikasi seberapa konsisten dan berbedanya hasil yang dihasilkan oleh kedua algoritma tersebut dalam mengukur sentimen.

4 Hasil dan Pembahasan

4.1 Hasil Pengumpulan Data dan Pra-Pemrosesan Data

Pengumpulan data dilakukan dengan fokus kepada kata kunci terkait *climate change* atau perubahan iklim. Kata kunci yang dimaksud adalah *global*, *climate*, *change*, dan *earth*. Pengumpulan data dari twitter menggunakan twitter API (*Application Programming Interface*) dan penyaringan *tweet* berdasarkan kata kunci tersebut. Pengumpulan data dimulai dari penyiapan akun dan akses API kemudian mengatur kriteria pencarian seperti menentukan kata kunci yang relevan. Proses selanjutnya membangun skrip pengambilan data menggunakan Python yang ditunjukkan pada gambar 2 dan dilanjutkan penyaringan serta penyimpanan data dalam format CSV. Data yang berhasil dikumpulkan sebanyak 30.000 data. Pada pra-pemrosesan data menjadi sebanyak 11.979 data.

```
import tweepy
import csv

# Konfigurasi API key dan token
consumer_key = "YOUR_CONSUMER_KEY"
consumer_secret = "YOUR_CONSUMER_SECRET"
access_token = "YOUR_ACCESS_TOKEN"
access_token_secret = "YOUR_ACCESS_TOKEN_SECRET"

# Autentikasi ke Twitter API
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth, wait_on_rate_limit=True)

# Kata kunci pencarian dan periode waktu
search_words = "#climatechange OR #globalwarming OR #sustainability OR #SDG13"
```

Gambar 2. Contoh skrip python

4.2 Hasil Analisa Frekuensi Kata

Analisis frekuensi kata menunjukkan bahwa istilah yang paling sering muncul dalam *tweet* tentang perubahan iklim adalah "*climate change*," yang paling dominan, diikuti oleh "*global heat*". Visualisasinya dapat dilihat pada gambar 3 berjudul *word cloud*. Kemudian dari 30 kata yang paling umum yang ditunjukkan pada gambar 4, ditemukan dalam *tweet*, kata "*climate*" muncul sebanyak 12,361 kali, "*change*" muncul 10,803 kali, "*global*" muncul 3,892 kali, dan "*heat*" muncul 2,520 kali. Hasil ini divisualisasikan dalam bentuk *word cloud*, di mana tulisan "*climate change*" ditampilkan paling besar, menandakan frekuensi kemunculannya yang tinggi dalam diskusi terkait perubahan iklim di twitter.



Gambar 3. Word cloud

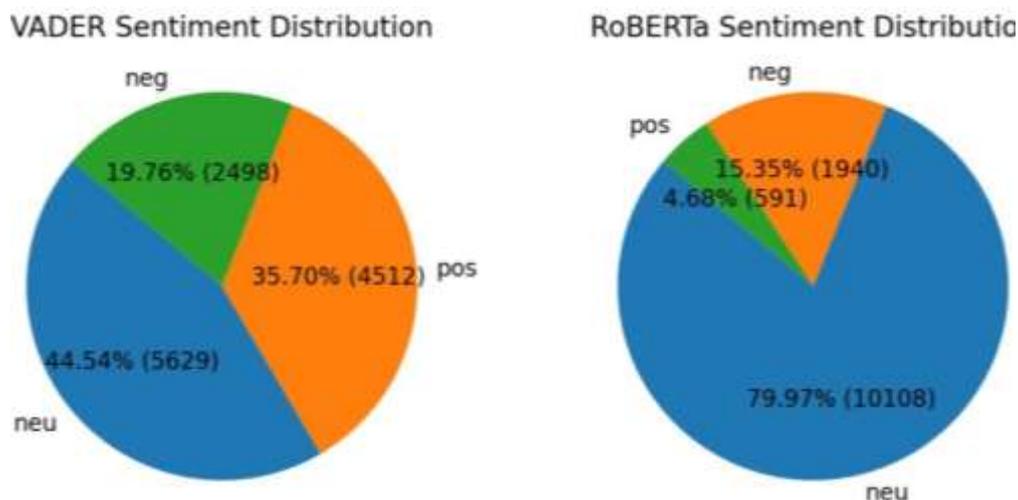
Top 30 Most Common Words in Tweets



Gambar 4. Kata kunci

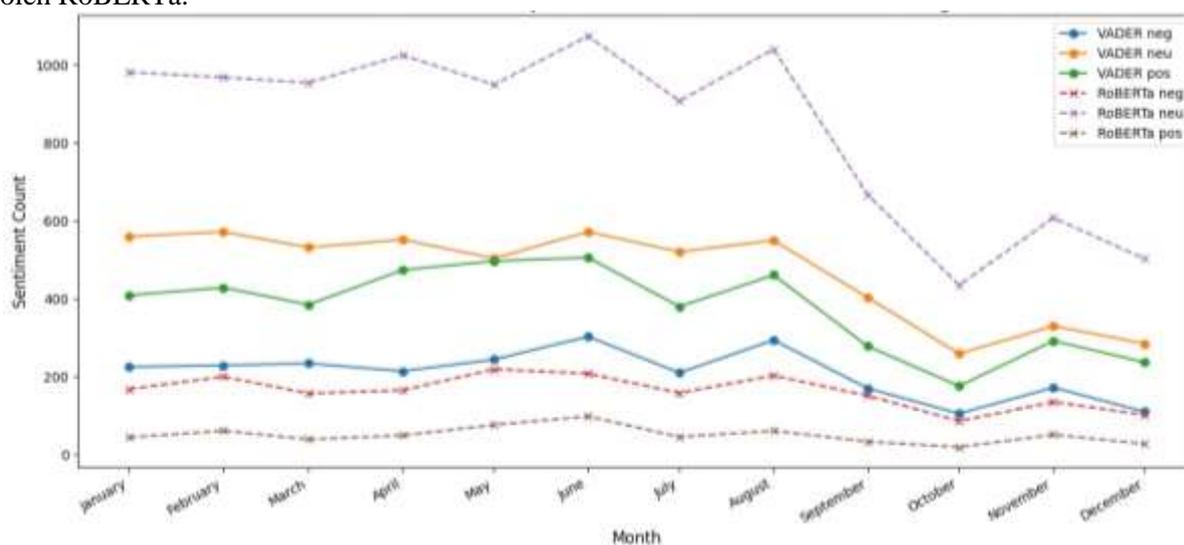
4.3 Hasil Perbandingan Sentimen Algoritma VADER dan RoBERTa

Setelah proses pembersihan data (*data cleaning*), semua data *tweet* yang ada ditetapkan warna sentimennya menggunakan dua algoritma, yaitu VADER dan RoBERTa. Algoritma VADER dan RoBERTa digunakan untuk mengklasifikasikan setiap *tweet* ke dalam tiga kategori sentimen: positif, negatif, atau netral. VADER, yang merupakan algoritma analisis sentimen berbasis leksikon, mengevaluasi sentimen berdasarkan intensitas kata. Sementara, RoBERTa merupakan model transformasi berbasis *deep learning*, memberikan penilaian sentimen dengan memahami konteks yang lebih dalam dari teks *tweet*. Dengan menggunakan kedua algoritma ini, setiap *tweet* dapat diberi label sentimen yang akurat, membantu dalam memahami persepsi publik terhadap perubahan iklim.



Gambar 5. Perbandingan distribusi sentimen hasil algoritma VADER dan RoBERTa

Pada gambar 5, terlihat bahwa algoritma RoBERTa cenderung mengklasifikasikan tweet sebagai netral dalam jumlah yang jauh lebih tinggi dibandingkan dengan VADER, dengan persentase mencapai 79,9% dari total tweet yang dianalisis. Sebaliknya, VADER hanya mencatat 44,54% tweet sebagai netral. Tingginya angka netral dari RoBERTa menunjukkan bahwa algoritma ini lebih konservatif dalam pengklasifikasian sentimen, mungkin karena lebih berhati-hati dalam menangani ambiguitas dalam teks tweet. RoBERTa cenderung tidak mengklasifikasikan tweet sebagai positif atau negatif kecuali sentimennya sangat jelas. Di sisi lain, VADER menunjukkan distribusi yang lebih seimbang antara sentimen positif dan negatif. Hal ini menandakan bahwa VADER lebih peka terhadap variasi emosi dalam tweet dan lebih cepat menangkap nuansa emosional yang mungkin tidak terdeteksi oleh RoBERTa.

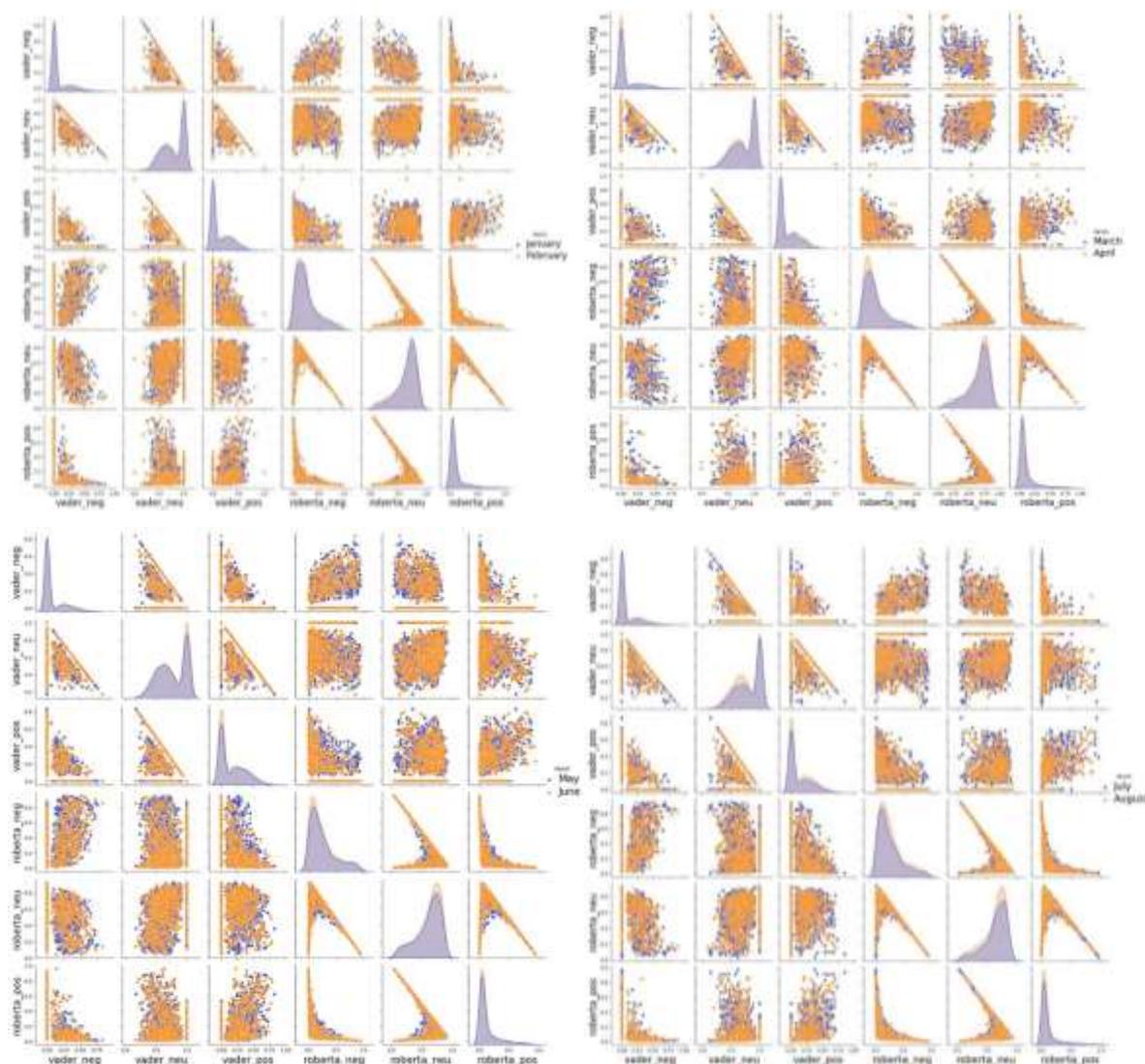


Gambar 6. Perbandingan sentimen algoritma VADER dan RoBERTa secara temporal

Selain melihat perbandingan distribusi sentimen hasil dari algoritma VADER dan RoBERTa, dilakukan juga analisis temporal terhadap distribusi sentimen yang dihasilkan oleh algoritma VADER dan RoBERTa yang ditunjukkan pada gambar 6. Analisis temporal menawarkan wawasan penting mengenai bagaimana sentimen publik berfluktuasi sepanjang tahun. Dalam periode Januari hingga April, data menunjukkan bahwa jumlah sentimen negatif yang dikategorikan oleh VADER cenderung stabil atau stagnan. Ini berbeda dengan hasil dari RoBERTa, yang menunjukkan fluktuasi yang lebih mencolok dalam periode yang sama. Setelah periode tersebut, dinamika fluktuasi jumlah sentimen—baik negatif, positif, maupun netral—antara kedua algoritma cenderung serupa di bulan-bulan berikutnya.

4.4 Hasil Komparatif Algoritma VADER dan RoBERTa

Penelitian ini menggunakan *scatter plot matrix* untuk memvisualisasikan hubungan antara algoritma VADER dan RoBERTa yang ditunjukkan pada gambar 7. Grafik ini memungkinkan peneliti untuk melihat keterkaitan antara kedua algoritma. Dengan memvisualisasikan hasil dari dua algoritma yang berbeda, peneliti dapat mengevaluasi konsistensi antar algoritma tersebut.



Gambar 7. Scatter plot matrix hubungan antara VADER dan RoBERTa

Pada gambar 7, dapat dilihat bahwa warna biru mewakili data dari bulan januari, maret, mei, dan juli. Sementara warna oranye mewakili data dari bulan february, april, juni dan agustus. Selama bulan januari dan february, grafik menunjukkan bahwa distribusi skor sentimen negatif, netral, dan positif dari algoritma VADER dan RoBERTa cukup konsisten. Scatter plot tersebut menunjukkan adanya kesesuaian dalam distribusi skor sentimen, meskipun ada sedikit perbedaan dalam intensitas sentimen. Scatter plot memperlihatkan bahwa sentimen netral lebih dominan dibandingkan sentimen negatif dan positif. Density plot di sisi kiri atas menunjukkan bahwa VADER cenderung mendeteksi sentimen negatif dengan nilai yang lebih tinggi dibandingkan RoBERTa, yang menunjukkan distribusi lebih merata dan cenderung ke nilai yang lebih rendah. Hal ini menunjukkan bahwa VADER mungkin lebih sensitif terhadap deteksi sentimen negatif, sementara RoBERTa lebih konservatif dalam memberikan skor sentimen negatif. Hasil serupa juga terlihat pada scatter plot matrix untuk bulan Maret-April, Mei-Juni, serta Juli-Agustus. Kedua algoritma menunjukkan konsistensi dalam

mendeteksi sentimen netral, sementara terdapat perbedaan dalam deteksi sentimen negatif dan positif. Ini menunjukkan bahwa pemilihan algoritma dapat mempengaruhi hasil analisis sentimen.

5 Kesimpulan

Penelitian ini membandingkan dua algoritma analisis sentimen, VADER dan RoBERTa, untuk mengklasifikasikan tweet ke dalam kategori sentimen positif, negatif, atau netral. Hasil analisis data menunjukkan bahwa kata kunci "*climate change*" adalah yang paling dominan dalam percakapan mengenai perubahan iklim di twitter, dengan "*climate*" dan "*change*" sebagai kata-kata yang paling sering muncul. Analisis sentimen menggunakan VADER dan RoBERTa menghasilkan temuan yang menarik. VADER cenderung lebih seimbang dalam distribusi sentimen positif dan negatif, sementara RoBERTa lebih konservatif dengan mayoritas *tweet* diklasifikasikan sebagai netral. Analisis temporal menunjukkan fluktuasi yang berbeda antara kedua algoritma selama tahun, dengan RoBERTa menunjukkan fluktuasi yang lebih signifikan dalam klasifikasi sentimen. Penelitian ini juga menggunakan *scatter plot matrix* untuk mengevaluasi keterkaitan antara hasil analisis sentimen dari kedua algoritma. Hasilnya menunjukkan bahwa meskipun ada konsistensi dalam mendeteksi sentimen netral, terdapat perbedaan dalam deteksi sentimen negatif dan positif antara VADER dan RoBERTa, yang mengindikasikan bahwa pilihan algoritma dapat mempengaruhi hasil akhir analisis sentimen.

Secara keseluruhan, penelitian ini memberikan pemahaman yang mendalam tentang perbandingan performa algoritma VADER dan RoBERTa pada opini terkait *climate change* dimana masing-masing algoritma memiliki keunggulan. Berdasarkan hasil, algoritma VADER lebih efektif dalam situasi yang membutuhkan tanggapan cepat terhadap perubahan sentimen publik, sedangkan RoBERTa lebih cocok untuk analisis mendalam pada konten yang lebih kompleks dan ambigu. Misalnya, jika tujuan utama adalah respons cepat terhadap perubahan sentimen, VADER mungkin lebih cocok. Namun, untuk analisis yang lebih mendalam, RoBERTa bisa memberikan wawasan yang lebih kompleks dan berlapis sehingga memungkinkan analisis yang lebih akurat mengenai sentimen yang sebenarnya dirasakan oleh pengguna twitter.

Referensi

- [1] A. Susilo dan A. Wahab, "Penerapan Metode Clustering Dalam Pengelompokkan Data Curah Hujan Dengan Algoritma K-Means," *Jurnal Teknologi Pelita Bangsa*, vol. 14, no. 4, pp. 212-215, 2023.
- [2] K. Setiawan dan F. Ramdhani, "Application of Data Mining for Customer Prediction at PT. XYZ using Linear Regression Algorithm," *Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 490-497, 2024.
- [3] N. M. Zawad, "Application of Data Mining in Healthcare of Bangladesh," *International Journal of Information System and Computer Science*, vol. 7, no. 2, pp. 89-103, 2023.
- [4] Sumiyatun, Y. Cahyadi dan E. Iskandar, "Data Mining untuk Memprediksi Status Kelulusan Mahasiswa," *Jurnal Teknologi dan Ilmu Komputer*, vol. 21, no. 3, pp. 11-19, 2023.
- [5] I. K. J. Arta, G. Indrawan dan G. R. Dantes, "Data Mining Rekomendasi Calon Mahasiswa Berprestasi di STMIK Denpasar Menggunakan Metode *Technique for Others Reference by Similarity to Ideal Solution*," *Jurnal Ilmu Komputer Indonesia*, vol. 4, no. 1, pp. 11-21, 2019.
- [6] Z. Aziz dan R. Bestak, "Insight into Anomaly Detection and Prediction and Mobile Network Security Enhancement Leveraging K-Means Clustering on Call Detail Records," *Sensors*, vol. 24, no. 6, pp. 1-18, 2024.
- [7] R. Hakiki, A. Pambudi dan Asriyanik, "Classification of Public Sentiment Toward 2024 Presidential Candidates on Social Media Platform X using Naïve Bayes Algorithm," *Journal of Artificial Intelligence and Engineering Applications*, vol. 3, no. 2, pp. 551-556, 2024.

- [8] L. A. Akbar, E. Haerani, F. Syafria, A. Nazir dan E. Budianita, "Klasifikasi Sentimen Presepsi Masyarakat di Instagram Terhadap Paslon Pilpres 2024 Menggunakan Naïve Bayes Classifier (NBC)," *Jurnal Komputasi dan Informatika*, vol. 8, no. 1, pp. 102-112, 2024.
- [9] K. Barik dan S. Misra, "Analysis of Customer Reviews with An Improved Vader Lexicon Classifier," *Journal of Big Data*, vol. 11, no. 10, pp. 1-29, 2024.
- [10] Y. Wiciaputra, J. Young dan A. Rusli, "Bilingual Text Classification in English and Indonesian via Transfer Learning using XLM-RoBERTa," *International Journal of Advances in Soft Computing and its Applications*, vol. 13, no. 3, pp. 73-87, 2021.
- [11] H. Dmour, R. Masa'deh, A. Salman dan M. Y. Abubhhashesh, "Influence of Social Media Platforms on Public Health Protection Against the COVID-19 Pandemic via the Mediating Effects of Public Health Awareness and Behavioral Changes: Integrated Model," *Journal of Medical Internet Research*, vol. 22, no. 8, pp. 1-15, 2020.
- [12] F. Solihin, S. Awaliyah dan A. M. A. Shofa, "Pemanfaatan Twitter Sebagai Media Penyebaran Informasi Oleh Dinas Komunikasi dan Informatika," *Jurnal Pendidikan Ilmu Pengetahuan Sosial*, vol. 13, no. 1, pp. 52-58, 2021.
- [13] D. Nurmadewi, S. Yuniati, M. Amaliah, H. Hanifah, U. B. Purwanti, M. S. Arum dan N. W. Kusuma, "Sentiment Analysis of Jokowi's Candidate Discourse in Three Periods using the Naïve Bayes Method," *Jurnal Sistem Informasi*, vol. 12, no. 11, pp. 166-177, 2023.
- [14] D. Nurmadewi, C. P. Gusti, E. C. Hernanto, F. Ananda, M. A. Hutagalung dan N. Hikmah, "Analisis Sentimen untuk Melihat Respon Masyarakat Terhadap Vaksin Pfizer," *Journal of Technology and Informatics*, vol. 4, no. 2, pp. 77-80, 2023.
- [15] L. Mutawalli, M. T. A. Zaen dan W. Bogy, "Klasifikasi Teks Sosial Media Twitter menggunakan Support Vector Machine (Studi Kasus Penusukan Wiranto)," *Jurnal Informatika dan Rekayasa Elektronika*, vol. 2, no. 2, pp. 43-51, 2019.
- [16] P. Bintoro, T. H. Andika, A. F. Yulia dan P. Widiandana, "Sentiment Analysis on Twitter using Machine Learning Approach," *International Journal of Software Engineering and Informatics*, vol. 1, no. 1, pp. 33-39, 2023.
- [17] Juleha, J. Yuniar dan N. R. Marsuki, "Peran Media Sosial Dalam Dinamika Opini Publik dan Partisipasi Politik Era Digital," *Journal of Social Humanities and Education*, vol. 3, no. 1, pp. 38-45, 2024.
- [18] F. Z. Abraham, F. Solihin, S. Awaliyah dan A. M. A. Shofa, "Pemanfaatan Twitter Sebagai Media Penyebaran Informasi Oleh Dinas Komunikasi dan Informatika," *Jurnal Pendidikan Ilmu Pengetahuan Sosial*, vol. 13, no. 1, pp. 52-58, 2021.
- [19] F. Illia, M. P. Eugenia dan S. A. Rutba, "Sentiment Analysis on PeduliLindungi Application using TextBlob and VADER Library," dalam *The 1st International Conference on Data Science and Official Statistics*, Jakarta, 2021.
- [20] V. Nurcahyawati dan Z. Mustaffa, "Vader Lexicon and Support Vector Machine Algorithm to Detect Customer Sentiment Orientation," *Journal of Information Systems Engineering and Business Intelligence*, vol. 9, no. 1, pp. 109-119, 2023.
- [21] V. Nurcahyawati dan Z. Mustaffa, "Vader Lexicon and Support Vector Machine Algorithm to Detect Customer Sentiment Orientation," *Journal of Information Systems Engineering and Business Intelligence*, vol. 9, no. 1, pp. 109-119, 2023.

- [22] A. N. Sutranggono, R. Sarno dan I. Ghozali, “Multi-Class Multi-Level Classification of Mental Health Disorders Based on Textual Data from Social Media,” *Journal of Information and Communication Technology*, vol. 23, no. 1, pp. 77-104, 2024.
- [23] A. Jaya, “Analisis Sentimen pandangan public terhadap profesi PNS (Pegawai Negeri Sipil) dari Twitter menerapkan Indonesian Roberta Base Sentiment Classifier,” *Indonesian Journal of Data and Science*, vol. 4, no. 1, pp. 38-44, 2023.
- [24] H. Tohari, Analisis Serta Perancangan Sistem Informasi Melalui Pendekatan UML, Yogyakarta: Penerbit Andi, 2014.
- [25] Kusri dan A. Kristanto, Tuntunan Praktis Membangun Sistem Informasi Akuntansi dengan Visual Basic dan SQL Server, Yogyakarta: Penerbit Andi, 2009.
- [26] A. Kadir, Pengenalan Sistem Informasi Edisi Revisi, Yogyakarta: Penerbit Andi, 2013.
- [27] F. Ilmijayanti dan D. I. K. Dewi, “Persepsi Pengguna Taman Tematik Kota Bandung Terhadap Aksesibilitas Dan Pemanfaatannya,” *RUANG, Volume 1, Nomor 1, 2015 ISSN: 2356-0088*, p. 23, 2015.
- [28] A. B. M. Alkatiri, Z. Nadiah dan A. N. S. Nasution, “Opini Publik Terhadap Penerapan New Normal di Media Sosial Twitter.,” *Journal of Strategic Communication*, vol. 11, no. 1, pp. 199-26, 2020.
- [29] A. A. Septiadi dan M. M. Yasa, “Kekuatan Mengikat Paris Agreement Kepada Negara-Negara Anggotanya,” *Jurnal Kertha Desa*, vol. 9, no. 8, pp. 13-21, 2021.
- [30] N. Puspita dan A. Hervino, “Implementasi Ratifikasi Paris Agreement oleh Indonesia dan Pengaruhnya Terhadap Kebijakan Perekonomian Indonesia.,” *Jurnal Komunikasi Hukum*, vol. 9, no. 1, pp. 704-728, 2023.
- [31] S. Manandhar, W. Pratoomchai, K. Ono, S. Kazama dan d. D. Komori, “Local People’s Perceptions of Climate Change and Related Hazards in Mountainous Areas of Northern Thailand,” 2018.
- [32] D. Nurhayati, Y. Dhokhikahb dan M. Mandala, “Persepsi dan Strategi Adaptasi Masyarakat Terhadap Perubahan Iklim di Kawasan Asia Tenggara.,” *Jurnal Proteksi: Jurnal Lingkungan Berkelanjutan*, vol. 1, no. 1, 2020.
- [33] L. Malihah, “Tantangan dalam Upaya Mengatasi Dampak Perubahan Iklim dan Mendukung Pembangunan Ekonomi Berkelanjutan: sebuah tinjauan.,” *Jurnal Kebijakan Pembangunan*, vol. 17, no. 2, pp. 219-232, 2022.
- [34] D. Nurhayati, Y. Dhokhikahb dan M. Mandala., “Persepsi dan Strategi Adaptasi Masyarakat Terhadap Perubahan Iklim di Kawasan Asia Tenggara,” *Jurnal Proteksi: Jurnal Lingkungan Berkelanjutan*, vol. 1, no. 1, 2020.
- [35] Bappenas, “Kaji Ulang Rencana Aksi Nasional Adaptasi Perubahan Iklim (RAN API),” *Kajian Basis Ilmiah Bahaya Perubahan Iklim.*, 2018.
- [36] IRID, “Public Opinion and Twentieth-Century Diplomacy: A Global Perspective,” *London: Bloomsbury Academic*, 2022.
- [37] L. Malihah, “Tantangan dalam Upaya Mengatasi Dampak Perubahan Iklim dan Mendukung Pembangunan Ekonomi Berkelanjutan: sebuah tinjauan,” *Jurnal Kebijakan Pembangunan*, vol. 17, no. 2, pp. 219-232, 2022.

- [38] N. Puspita dan A. Hervino, "Implementasi Ratifikasi Paris Agreement oleh Indonesia dan Pengaruhnya Terhadap Kebijakan Perekonomian Indonesia.," *Jurnal Komunikasi Hukum*, vol. 9, no. 1, pp. 704-728, 2023.
- [39] Nurbaya, Siti dan et.al., "Trilogi Indonesia Menghadapi Perubahan Iklim: Urgensi, Politik Dan Tata Kelola Perubahan Iklim.," *Jakarta : PT. Kompas Media Nusantara.*, 2019.
- [40] H. Tsaqib dan et.al., "Studi Komparatif Tingkat Pengetahuan Perubahan Iklim pada Mahasiswa Rumpun Sosial-Humaniora dan Rumpun Sains-Teknologi di Universitas Indonesia.," *Jurnal KSM Eka Prasetya UI*, vol. 2, no. 2, 2020.
- [41] M. Taddicken, A. Reif dan I. Hoppe, "What do People Know About Climate Change - and How Confident Are They? on Measurements and Analyses of Science Related Knowledge.," *Journal of Science Communication*, vol. 17, no. 3, pp. 1-26, 2018.
- [42] I. Syahputra, "Opini Publik," (*Bandung : Simbiosis Rekatama Media*), 2018.
- [43] D. Hucker, "Public Opinion and Twentieth-Century Diplomacy: A Global Perspective.," *London: Bloomsbury Academic*, 2020.
- [44] A. B. M. Alkatiri, Z. Nadiah dan A. N. S. Nasution, "Opini Publik Terhadap Penerapan New Normal di Media Sosial Twitter.," *Journal of Strategic Communication*, vol. 11, no. 1, pp. 199-26, 2020.
- [45] F. Sultan dan S. Purwanti, "Pembentukan opini publik pada akun twitter pribadi novel baswedan.," *eJournal Ilmu Komunikasi*, vol. 10, no. 4, pp. 155-164, 2022.