

SENTIMENT ANALYSIS OF POTENTIAL PRESIDENTIAL CANDIDATES FOR 2024 ON TWITTER USING THE MULTINOMIAL NAIVE BAYES METHOD

Muslikhah^{1*}, Muhammad Fairuzabadi², Wibawa³

^{1,2,3}Informatika Universitas PGRI Yogyakarta, Indonesia

muslikhahhh@gmail.com^{1}, fairuz@upy.ac.id², wibawa@upy.ac.id³*

**Corresponding author*

Manuscript Received August 17, 2023; Revised February 05, 2024; Accepted February 05, 2024;
Published April 26, 2024

ABSTRACT

This research aims to conduct sentiment analysis regarding potential presidential candidates in 2024, so that we can identify candidates who have positive, neutral and negative images in the view of the public on Twitter social media. This sentiment analysis helps candidates understand People's aspirations and adapt their communications accordingly. This research uses the naïve Bayes multinomial method and utilizes crawling technology on Twitter social media, data is collected and analyzed efficiently. The results of this research obtained 6000 comment data with each candidate having 2000 comments. Ganjar Pranowo had the highest positive sentiment (39,9%), followed by Anies Baswedan (36,8%) and Prabowo Subianto (25,9%). Ganjar also leads in neutral sentiment (38.8%). The highest number of negative sentiments was held by Prabowo Subianto (41%), Anies Baswedan (31.4%), Ganjar Pranowo (21.3%). So from these results, Ganjar Pranowo has the best electability based on public comments on Twitter social media.

Keywords: Twitter, sentiment analysis, potential presidential candidates 2024

ABSTRAK

Penelitian ini bertujuan untuk melakukan analisis sentimen mengenai bakal calon potensial presiden 2024, sehingga dapat diketahui calon kadidat yang memiliki image positif, netral, dan negatif dalam pandangan Masyarakat di media sosial twitter. Analisis sentimen ini membantu kandidat memahami aspirasi Masyarakat dan menyesuaikan komunikasi mereka. Penelitian ini menggunakan metode naïve bayes multinomial dan memanfaatkan teknologi crawling pada media sosial Twitter, data dikumpulkan dan dianalisis secara efisien. Hasil dari penelitian ini mendapatkan 6000 data komentar dengan masing-masing kandidat memiliki 2000 komentar. Ganjar Pranowo memiliki sentimen positif tertinggi (39,9%), disusul Anies Baswedan (36,8%) dan Prabowo Subianto (25,9%). Ganjar juga memimpin dalam sentimen netral (38,8%). Untuk jumlah sentimen negatif tertinggi dipegang oleh Prabowo subianto (41 %), Anies Baswedan (31,4%), Ganjar Pranowo (21,3%). Sehingga Dari hasil tersebut, Ganjar Pranowo memiliki elektabilitas terbaik berdasarkan komentar Masyarakat di media sosial Twitter.

Kata Kunci: Twitter, analisis sentimen, calon potensial presiden 2024

PENDAHULUAN

Indonesia merupakan sebuah negara yang mengadopsi sistem demokrasi. Salah satu manifestasi dari penerapan demokrasi di Indonesia adalah keberadaan pemilihan umum yang diadakan secara berkala (Andika et al., 2019). Pemilihan umum yang juga dikenal sebagai pemilu ini diselenggarakan setiap lima tahun sekali. Pemilihan Presiden merupakan momen penting yang ditunggu seluruh masyarakat Indonesia. Pada tahun 2024 yang akan datang merupakan tahun dalam pesta demokrasi terbesar bagi seluruh warga negara Indonesia, beberapa calon potensial dari partai mulai bermunculan. Berdasarkan beberapa survei, termasuk yang dilakukan oleh *Central for Political Communication Studies* (CPCS) di Jakarta pada hari Jum'at, 4 Februari 2022, terdapat tiga kandidat terkuat dengan tingkat simpati publik tertinggi, yaitu Anies Baswedan (14,8%), Prabowo Subianto (18,3%), dan Ganjar Pranowo (14,5%) (Aryadillah & Fitriansyah, 2022). Menjelang pemilihan Presiden, para kandidat sering memanfaatkan media sosial untuk meningkatkan popularitas mereka. Masyarakat menggunakan media sosial sebagai alat komunikasi. Menurut survei APJII tahun 2017, dari total populasi Indonesia yang berjumlah 262 juta jiwa, terdapat 143,26 juta orang yang memanfaatkan internet. Dari jumlah tersebut, 97,4% atau sekitar 129,2 juta pengguna mengakses konten media sosial, dan 56,01% di antaranya mengakses informasi terkait politik. (Yasya et al., 2019). Beberapa platform media sosial yang sering digunakan di Indonesia, salah satunya adalah Twitter / X.

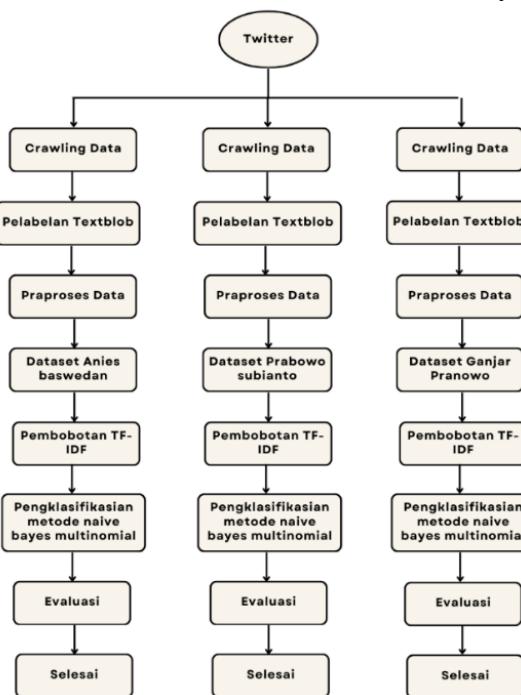
Twitter merupakan salah satu platform media sosial yang sangat populer di seluruh dunia, terbukti dengan peningkatan penggunaannya dari tahun ke tahun. Berdasarkan survei yang dilakukan oleh We Are Social dan Hootsuite, hingga awal tahun 2023, Twitter di Indonesia berada di peringkat kelima secara global dengan jumlah pengguna mencapai 24 juta orang. (We Are Social & Hootsuite, 2023). Selain itu, Twitter menyediakan fitur yang memungkinkan pengguna untuk mengirim dan membaca Tweet yang berisi teks, video, atau gambar. Penggunaan tagar atau hashtag (#) memudahkan pembahasan topik tertentu. Twitter juga dapat digunakan untuk mendapatkan informasi, seperti menganalisis sentimen masyarakat terhadap berbagai isu, termasuk calon presiden potensial 2024.

Ada berbagai isu terkait calon kandidat presiden yang memunculkan beragam opini, baik yang positif maupun negatif. Oleh karena itu, Twitter dianggap lebih diminati oleh masyarakat Indonesia karena dirasa lebih mudah dan sederhana dalam menyampaikan opininya. Setiap pernyataan yang dikeluarkan oleh calon kandidat presiden akan menimbulkan pro dan kontra di kalangan masyarakat. Ada yang merespons secara positif, negatif, atau netral. Dalam Konteks ini, penting untuk menganalisis komentar masyarakat sehingga opini publik terhadap tokoh tersebut dapat dikelompokkan berdasarkan label tersebut dan dapat diketahui seberapa besar dukungan atau penolakan yang diterima.

Analisis Sentimen merupakan bidang studi yang mengkaji opini, perasaan, penilaian, evaluasi, sikap, dan emosi individu terhadap suatu topik, layanan, produk, tokoh, organisasi, atau kegiatan tertentu (Nurjanah et al., 2017). Analisis sentimen bertujuan untuk menilai opini publik terhadap berbagai entitas, baik yang bersifat benda hidup maupun benda tak hidup. Untuk melakukan analisis ini, dapat dimanfaatkan data berupa tanggapan masyarakat di platform media sosial atau tulisan yang tersedia di berbagai situs web (Adhi et al., 2019). Sehingga pada penelitian ini penulis akan melakukan “Analisis Sentimen Terhadap Calon Potensial Presiden 2024 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Multinomial”.

METODE PENELITIAN

Metode yang dilakukan pada penelitian ini dengan mengambil dataset mengenai calon presiden 2024 yang ada pada Twitter dari tanggal 20 November 2023 hingga 20 Maret 2024. Penelitian ini menganalisis sentimen dengan menggunakan metode Naive Bayes Multinomial untuk menentukan akurasi dan sentimennya.



Gambar 1. Flowchart Metode Penelitian

Pengumpulan Data

Data dikumpulkan dari komentar Twitter mulai tanggal 20 November 2023 sampai 20 Maret 2024 yang menggunakan keywords “anies baswedan”, “prabowo subianto” dan “ganjar pranowo”. Pengambilan data memanfaatkan twitter API (Application Programming Interface) menggunakan Teknik crawling. Banyak data yang pada masing-masing tokoh yaitu 2000 data tweet sehingga banyak data yang dikumpulkan sebanyak 6000 data dan kemudian disimpan dalam tabel dengan format .csv.



Gambar 2. Contoh Data di X (twitter)

Pelabelan

Pada tahap pelabelan peneliti melakukan pelabelan data secara otomatis menggunakan Textblob. Textblob adalah salah satu library yang terdapat pada python yang berupa kamus khusus untuk menentukan nilai sentimen dari teks yang terdiri dari kata-kata dengan nilai sentiment positif, negatif, dan netral. Yang Dimana untuk melakukan proses natural language processing pada dataset (Reynaldhi & Sibaroni, 2021).

Praproses Data

Tahapan Praproses Data dalam analisis sentimen tujuannya adalah untuk membersihkan data dengan menghapus tweet yang berulang dan menstandarkan kalimat yang tidak baku agar diperoleh data yang bersih (Putranti & Winarko, 2014). Berikut beberapa tahap praproses data yang dapat dilihat pada gambar 3.3.



Gambar 3. Gambar tahapan praproses Data

1. *Case Folding*, Proses mengonversi seluruh karakter alfabet dalam naskah komentar menjadi huruf kecil (Yoga Pratama et al., 2021).
2. *Text Cleaning*, Tahapan yang bertujuan untuk menghilangkan atau menyaring elemen-elemen yang tidak diperlukan, seperti tanda baca, angka, simbol, sebutan pengguna, tagar, tautan, dan sebagainya (Putra et al., 2021).

3. *Tokenisasi*, proses memisahkan kalimat ke dalam kata-kata atau istilah, setiap kata yang terpisah disebut sebagai token. Tujuan dari tokenisasi ini adalah untuk mempermudah analisis atas makna dari setiap kata (Buntoro, 2017).
4. *Remove Stopword*, proses penghapusan semua kata yang ada pada data komentar yang tidak memiliki makna. Contohnya kata-kata penghubung seperti "dan", "atau", "tetapi", "akan", serta kata-kata sejenis lainnya (Pradana & Hayaty, 2019).
5. *Stemming*, proses yang dilakukan yaitu proses mengubah kata ke bentuk dasarnya dengan menghilangkan imbuhan baik di awal maupun di akhir kata. Saat melakukan stemming digunakan library python yang disebut sastrawi (Juanita, 2020).

Metode Naive Bayes Multinomial

Pada tahap ini data-data yang telah diproses pada tahap sebelumnya akan diproses menggunakan algoritma naive bayes multinomial (Fairuzabadi et al., 2024). Rumus perhitungan Naive Bayes Multinomial dapat dilihat persamaan berikut:

Langkah awal dalam metode naïve bayes multinomial adalah menghitung probabilitas untuk setiap kelas berdasarkan keseluruhan data latih, yang disimbolkan dengan P. Probabilitas prior dihitung untuk masing-masing kategori sentimen "positif", "netral", dan "negatif". dapat dilihat pada persamaan berikut:

$$P(\text{Kelas} | D) = \frac{P(\text{Kelas})}{P(D)} \quad (1)$$

Dimana:

- $P(\text{Kelas} | D)$ adalah probabilitas Kelas Sentimen (kelas sentimen : positif, netral, negatif).
- $(\text{Kelas} | D)$ adalah jumlah data latih kelas sentimen.
- $P(D)$ adalah jumlah total data latih

Dalam tahap pelatihan, dilakukan analisis terhadap contoh data yang merepresentasikan dokumen. Selanjutnya, dilakukan perhitungan perbandingan antara istilah pada data uji dengan setiap kategori kelas yang ada dengan dihitung menggunakan persamaan berikut:

$$P(a_j | v_j) = \frac{nc + mp}{n+m} \quad (2)$$

Dimana:

- nc : jumlah kalimat sentimen pada data latih
- p : probabilitas setiap kelas pada data latih
- m : jumlah term pada data uji
- n : jumlah total sentimen pada data latih

Perhitungan untuk menentukan klasifikasi pada data uji dilakukan dengan persamaan berikut:

$$V_{nb} = \operatorname{argmax}_{v_j \in V} P(v_j) \prod P(a_i | v_j) \quad (3)$$

Dimana:

- ai= atribut atau fitur ke-i
- vj= kelas ke-j (positif, negative atau netral)

- $V =$ himpunan kelas target
- $V_{map} =$ kelas sentimen suatu tweet

HASIL DAN PEMBAHASAN

Berdasarkan hasil penelitian yang telah dilaksanakan, dapat disimpulkan bahwa kinerja sistem berfungsi dengan baik sesuai dengan yang diharapkan.

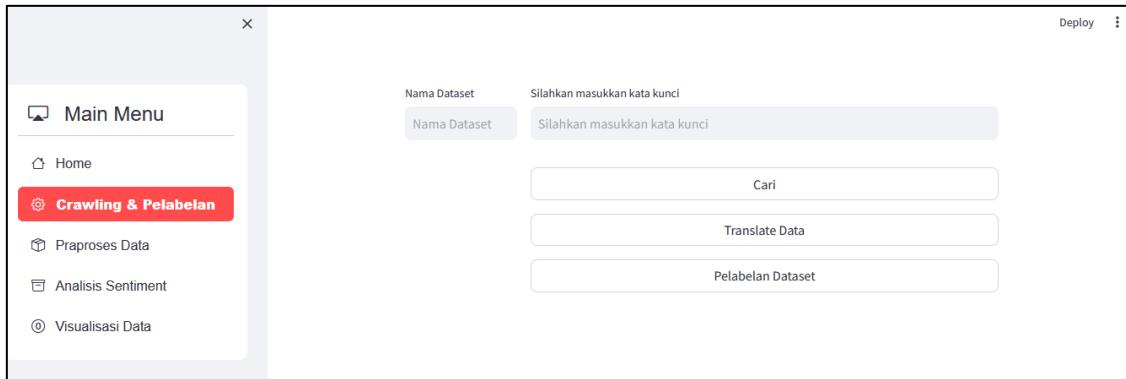
Implementasi Sistem

Implementasi sistem analisis sentimen calon potensial presiden 2024 pada media sosial Twitter menggunakan metode Naïve Bayes Multinomial ini merupakan tindakan yang diambil setelah tahap perancangan sistem, dengan tujuan untuk merealisasikan sistem yang disusulkan dan mengevaluasi kelayakan pembuatan sistem tersebut. Halaman Home merupakan tampilan utama dalam sistem ini. Tampilan halaman utama dapat dilihat pada ilustrasi yang disajikan dalam Gambar 4.



Gambar 4. Tampilan Halaman Utama

Halaman Crawling Data & Pelabelan merupakan halaman kedua yang diakses dalam sistem ini. Halaman ini menunjukkan 2 tahapan, yaitu proses crawling data dan pelabelan data. Proses Crawling Data dimulai dengan pengambilan data dari Twitter, di mana pengguna dapat memasukkan kata kunci di kolom pencarian yang tersedia. Halaman Crawling data ini digunakan khusus untuk pengambilan data secara real-time, yang kemudian ditampilkan dalam tabel di bawahnya sebelum melanjutkan ke proses pelabelan data. Tampilan proses crawling Data & Pelabelan dapat dilihat pada Gambar 5.



Gambar 5. Tampilan Halaman Crawling Data & Pelabelan

Halaman ini menyajikan proses Praproses Data dari Komentar yang sudah terkumpul dengan cara memilih File Data komentar yang sudah kita Crawling Dari twitter, kemudian sistem akan secara otomatis mengolah data tersebut. Hasil praproses akan disimpan dalam file excel dan hasil data yang sudah melalui tahap praproses bisa dilihat pada table dibawahnya. Tampilan Praproses Data bisa dilihat pada Gambar 6.

	Tweet	Translated_Text
0	_Dwi	_Dwi
1	_Lbs Manusia paling durhaka di alam semesta ?????	_LBS HUMAN MOST DUR
2	Kalau mengatakan amin dianggap mendukung dia orang ? Pola berpikir yg miskin	If you say Amin is consider
3	Capres wowo lol oon penghianat negara buka rekam jejak orang tuanya smpai wowo	Candidate Wowo Lol Oor
4	pilih PERUBAHAN by erros djarot	Select changes by erros l
5	_TV AMIN diatas angin	_TV Amin on the wind
6	pilih PERUBAHAN by erros djarot	Select changes by erros l
7	pendukung pa ini adalah yg jg yg prnah menunggangi pa ...mereka ramai di medsos,	PA supporters are the on

Gambar 6. Tampilan Halaman Praposes Data

Halaman Analisis Sentimen adalah halaman dimana kita akan menganalisis Komentar yang akan kita analisis sentimennya dan juga menampilkan hasil sentimen dari skomentar tersebut. Seperti contoh pada tampilan dibawah sistem menganalisis sebuah kalimat komentar baru yang berbunyi “ayo semangat kalian hebat” secara otomatis sistem yang sudah dilatih menggunakan metode *naïve bayes multinomial* akan menganalisis bahwa komentar tersebut mengandung sentimen positif. Tampilan Halaman Analisis Sentimen dapat dilihat pada Gambar 7.



Gambar 7. Tampilan Halaman Analisis Sentimen

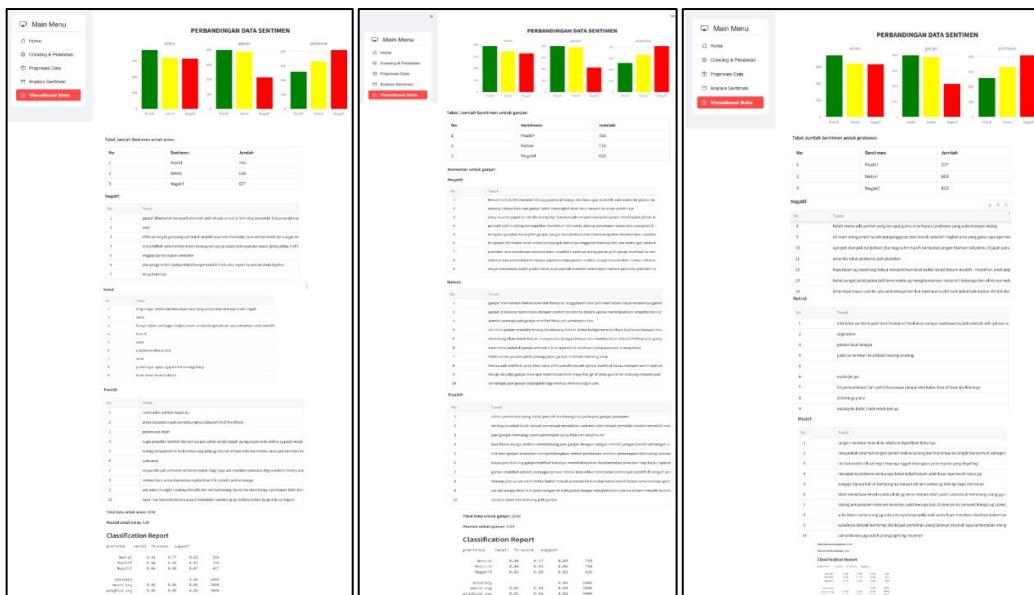
Halaman terakhir dari semua proses ini adalah halaman visualisasi data, pada penelitian ini penulis membuat 3 grafik visualisasi, yaitu grafik analisis sentimen dataset Anies Baswedan, Prabowo Subianto, dan Ganjar Pranowo. Selain grafik batang, pada menu ini ada beberapa bagian lain yaitu tabel yang menunjukkan jumlah sentimen perkandidat, tabel komentar perkategori sentimen, dan *classification report*.

Tabel Perbandingan Hasil dari dataset Anies Baswedan, Prabowo Subianto, dan Ganjar Pranowo dapat dilihat pada Tabel 1.

Tabel 1. Perbandingan Dataset Para Kandidat

Nama	Jumlah Komentar	Sentimen Positif	Sentimen Netral	Sentimen Negatif	Akurasi Metode (%)
Anies Baswedan	2000	735	638	627	86
Prabowo Subianto	2000	517	663	820	84
Ganjar Pranowo	2000	798	776	426	84

Tampilan Halaman Visualisasi Data dapat dilihat pada Gambar 8.



Gambar 8. Tampilan Visualisasi Data

KESIMPULAN

Berdasarkan hasil Analisis Sentimen terhadap calon potensial presiden 2024 pada media sosial Twitter menggunakan metode *Naïve Bayes Multinomial*, beberapa hasil yang diperoleh adalah sebagai berikut:

1. Dari penelitian yang telah dilakukan sistem analisis sentimen dapat digunakan untuk mengetahui opini masyarakat ditwitter terhadap calon potensial presiden Indonesia tahun 2024.
2. Metode Naïve Bayes terbukti dapat digunakan untuk analisis sentimen data komentar secara otomatis dengan menggunakan sistem.
3. Hasil dari penelitian ini mendapatkan 6000 data komentar Twitter dari tanggal 20 November 2023 sampai 20 Maret 2024 dengan masing-masing kandidat memiliki 2000 komentar. Ganjar Pranowo memiliki sentimen positif tertinggi (39,9%), disusul Anies Baswedan (36,8%) dan Prabowo Subianto (25,9%). Ganjar juga memimpin dalam sentimen netral (38,8%). Untuk jumlah sentimen negatif tertinggi dipegang oleh Prabowo subianto (41%), Anies Baswedan (31,4%), Ganjar Pranowo (21,3%). Sehingga dari hasil tersebut, Ganjar Pranowo memiliki elektabilitas terbaik berdasarkan komentar masyarakat di media sosial Twitter.

DAFTAR PUSTAKA

- Adhi, M. S., Nafan, M. Z., & Usada, E. (2019). Pengaruh Semantic Expansion pada Naïve Bayes Classifier untuk Analisis Sentimen Tokoh Masyarakat. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 3(2), 141–147. <https://doi.org/10.29207/resti.v3i2.901>
- Andika, L. A., Azizah, P. A. N., & Respatiwulan, R. (2019). Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier. In *Indonesian Journal of Applied Statistics* (Vol. 2, Issue 1, p. 34). <https://doi.org/10.13057/ijas.v2i1.29998>
- Aryadillah, A., & Fitriansyah, F. (2022). Strategi Kampanye Politik Anies Baswedan dalam Membangun Citra Politik Pada Pemilihan Presiden Tahun 2024. *Jurnal Public Relations (J-PR)*, 3(1), 87–92. <https://doi.org/10.31294/jpr.v3i1.1150>
- Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *Integer Journal*, 2(1), 32–41.
- Fairuzabadi, M., Sinambela, M., Taju, S. W., Syahrani, A., Arni, S., Yuliansyah, H., Saputra, A. I. H., Stephane, I., Pakpahan, A., Lubis, M., & Liem, A. T. (2024). *Data Science: Sebuah Pengantar untuk Pemula*. Yayasan Kita Menulis.
- Juanita, S. (2020). Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes. *Jurnal Media Informatika Budidarma*, 4, 552–558.
- Nurjanah, W. E., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet. In *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN* (Vol. 2548, p. 964X).

- Pradana, A. W., & Hayaty, M. (2019). The Effect of Stemming and Removal of Stopwords on the Accuracy of Sentiment Analysis on Indonesian-language Texts. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 375–380. <https://doi.org/10.22219/kinetik.v4i4.912>
- Putra, A., Haeirudin, D., Khairunnisa, H., & Latifah, R. (2021). Analisis Sentimen Masyarakat Terhadap Kebijakan PPKM Pada Media Sosial Twitter Menggunakan Algoritma SVM. In *Seminar Nasional Sains dan Teknologi 2021* (Issue November, pp. 1–6).
- Putranti, N. D., & Winarko, E. (2014). Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine. In *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)* (Vol. 8, Issue 1, p. 91). <https://doi.org/10.22146/ijccs.3499>
- Reynaldhi, M. A. R., & Sibaroni, Y. (2021). Analisis Sentimen Review Film pada Twitter menggunakan Metode Klasifikasi Hybrid Naïve Bayes dan Decision Tree. *E-Proceeding of Engineering*, 8(5), 10127–10137.
- We Are Social & Hootsuite. (2023). *Digital 2023: Indonesia. We Are Social*. <https://wearesocial.com/id/blog/2023/01/digital-2023/>
- Yasya, W., Muljono, P., Seminar, K. B., & Hardinsyah, H. (2019). Pengaruh Penggunaan Media Sosial Facebook Dan Dukungan Sosial Online Terhadap Perilaku Pemberian Air Susu Ibu. *Jurnal Studi Komunikasi Dan Media*, 23(1), 71. <https://doi.org/10.31445/jskm.2019.1942>
- Yoga Pratama, A., Umaidah, Y., Voutama, A., Informatika, T., Ilmu Komputer, F., Singaperbangsa Karawang Ds Paseurjaya, U., Telukjambe Timur, K., Karawang, K., & Barat, J. (2021). Analisis Sentimen Media Sosial Twitter Dengan Algoritma K-Nearest Neighbor Dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja). *Jurnal Sains Komputer & Informatika (J-SAKTI*, 5(2), 897–910.