

## Analisis Sentimen Terhadap Calon Wakil Presiden Gibran Rakabuming Raka Menggunakan Algoritma Naive Bayes

**Muhammad Khumaidi Nursyarif<sup>1</sup>, Muhamad Wahyu Tirta<sup>2</sup>, Muhammad Rahman Hidayat<sup>3</sup>, Rudiman<sup>4</sup>**

Fakultas Sains dan Teknologi, Universitas Muhammadiyah Kalimantan Timur  
E-mail Korespondensi : <sup>1)</sup>[2011102441085@umkt.ac.id](mailto:2011102441085@umkt.ac.id), <sup>2)</sup>[2011102441014@umkt.ac.id](mailto:2011102441014@umkt.ac.id), <sup>3)</sup>[2011102441012@umkt.ac.id](mailto:2011102441012@umkt.ac.id),  
<sup>4)</sup>[rud959@umkt.ac.id](mailto:rud959@umkt.ac.id)

History Artikel

Diterima : 23 November 2023 Disetujui : 07 Januari 2024 Dipublikasikan : 26 April 2024

---

### **Abstract**

*The turnover of the president and vice president in Indonesia occurs every 5 years. In the year 2024, there will be an election, and one of the vice presidential candidates, listed as candidate number 2, is Gibran Rakabuming Raka, who is the son of Mr. Jokowi, currently serving as the 7th President. Many opinions have been expressed by the public regarding Mas Gibran, especially considering his age of 36 years, which is perceived as relatively young to lead the nation of Indonesia. Therefore, we intend to conduct research with the aim of identifying practical implications related to public perceptions of the potential vice presidential candidate. Data from comments on the YouTube video titled "[FULL] Gibran in Between Ganjar and Prabowo, Which One to Choose? | ROSI" underwent a classification process using the Naive Bayes algorithm for sentiment analysis. The accuracy obtained is 92.5%, with an f1 Score of 92.4%, Precision of 93.5%, and Recall of 92.5%.*

**Keywords:** *Gibran, Analisis Sentimen, Naive Bayes*

## ***Abstrak***

Pergantian presiden dan wakil presiden di Indonesia terjadi setiap 5 tahun sekali. Di tahun 2024 nanti akan diadakan pemilu dengan salah satu calon wakil presiden paslon nomor 2 adalah Gibran Rakabuming Raka yang merupakan anak dari Bapak Jokowi yang saat ini menjabat sebagai Presiden ke-7. Banyak opini yang diutarakan masyarakat terkait Mas Gibran apalagi di usia 36 Tahun yang dirasa masih terbilang cukup muda bila harus memimpin bangsa Indonesia. Dari hal tersebut kami ingin melakukan penelitian dengan tujuan mengidentifikasi implikasi praktis terkait persepsi masyarakat terhadap kandidat wakil presiden potensial. Data dari komentar pada video Youtube dengan judul “[FULL] Gibran di Antara Ganjar dan Prabowo, Pilih Mana? | ROSI” kami lakukan proses Klasifikasi menggunakan algoritma Naive Bayes untuk analisis sentimen-nya. Hasil akurasi yang didapatkan sebesar 92,5%, f1 Score 92,4%, Precision 93,5% dan Recall 92,5%.

**Kata Kunci:** *Gibran, Analisis Sentimen, Naive Bayes*

**How to Cite:** Nursyarif, Muhammad Khumaidi, dkk (2024). Analisis Sentimen Terhadap Calon Wakil Presiden Gibran Rakabuming Raka Menggunakan Algoritma Naive Bayes. *KOMPUTEK : Jurnal Teknik Universitas Muhammadiyah Ponorogo*, Vol 8 (1): Halaman 1-7

© 2024 Universitas Muhammadiyah Ponorogo. All rights reserved

---

ISSN 2614-0985 (Print)

ISSN 2614-0977 (Online)

## PENDAHULUAN

Di Indonesia, demokrasi terwujud melalui pemilihan umum yang diadakan setiap 5 tahun sekali untuk menentukan pemimpin negara (Hidayatillah et al., 2019) yang menganut sistem demokrasi dan diatur sesuai dengan Pasal 1 ayat 2 Undang-Undang Dasar 1945 (Joshua Lefrandt Thanos et al., 2020). Pada tahun 2024, akan dilaksanakan pemilihan umum Presiden dan Wakil Presiden periode 2024-2029. Sebagai calon presiden, seorang politikus tentu akan memperhitungkan seberapa populer mereka di mata masyarakat (Fatima Barata & Simanjuntak, 2019).

Dalam mengekspresikan opini, masyarakat menggunakan jejaring sosial media yang terkenal saat ini yaitu Youtube. Secara umum, YouTube sering digunakan untuk mengunggah video yang berisi tentang kehidupan pribadi pengguna, berbagi informasi, serta menyampaikan berita (Ihendra & Laugu, 2020). Isi dari konten tersebut bisa mencerminkan perasaan yang dirasakan oleh pengguna (Furqon, 2021). Oleh karena itu, diperlukan analisis yang tepat untuk mengolah data teks agar dapat menyajikan informasi berharga tentang pendapat seseorang terhadap tokoh politik atau figur public (Zaini et al., 2021). Komentar-komentar dalam bentuk teks tersebut dapat dikumpulkan dan dianalisis menggunakan teknik analisis sentimen, yang bertujuan untuk memahami opini atau pandangan yang mengungkapkan persepsi dengan nuansa positif atau negative (Santoso & Nugroho, 2019). Pendapat yang dijadikan sumber analisis berasal dari komentar di halaman fanpage yang berjudul "Gibran di Antara Ganjar dan Prabowo, Pilih Mana? | ROSI" di platform komentar YouTube.

Dalam penelitian ini, data yang dikumpulkan akan diarahkan ke proses klasifikasi menggunakan metode Naïve Bayes. Metode ini merupakan teknik klasifikasi yang bergantung pada nilai probabilitas statistik sederhana dan mengasumsikan independensi yang kuat, sejalan dengan aturan Bayesian (Duei Putri et al., 2022). Pendekatan analisis Bayesian ini menggunakan informasi sebelumnya dan informasi dari sampel data untuk menghasilkan probabilitas posterior (Ruhana, 2019). Penggunaan metode Klasifikasi Naïve Bayes melibatkan pengujian data testing dengan memanfaatkan data training (Farah Zhafira et al., 2021).

Pemilihan Metode Naïve Bayes Classifier dipilih karena kesederhanaan metodenya, kecepatan prosesnya, dan tingkat kepercayaan yang tinggi, seperti yang dijelaskan oleh (Anggraeni, 2022). Harapannya, penelitian ini mampu mengklasifikasikan teks dengan akurasi yang baik, memungkinkan ekstraksi informasi

yang terkandung, serta memberikan observasi yang berguna dari data yang disajikan bagi pihak yang membutuhkan.

## METODE PENELITIAN

Dalam studi ini terdapat serangkaian langkah-langkah untuk menganalisis sentimen di platform media sosial Youtube. Penelitian ini memanfaatkan algoritma Naive Bayes sebagai model untuk mengklasifikasikan sentimen, dengan rencana melakukan proses validasi serta evaluasi terhadap data yang terkumpul.

Data yang digunakan berupa hasil proses crawling data di media sosial Youtube di channel KOMPASTV dengan judul "[FULL] Gibran di Antara Ganjar dan Prabowo, Pilih Mana? | ROSI". Video tersebut dinarasikan oleh Rosi sebagai pembicara dan Gibran Rakabuming Raka sebagai narasumber terkait dukungan yang ia berikan terhadap Prabowo atau Ganjar. Data diambil dari komentar pada video youtube dengan total 5.580 komentar terkait pendapat masyarakat terhadap Gibran Rakabuming Raka.

Terdapat berbagai komentar positif dan negatif terkait Gibran yang maju sebagai calon Wakil Presiden digandeng Bapak Prabowo Subianto paslon nomor 2. Setelah proses crawling, data kemudian dilakukan pembersihan di Python dan dilakukan pelabelan menggunakan metode vader\_lexicon, selanjutnya Data selanjutnya dilakukan proses analisis sentimen menggunakan software Orange Sentimen Analisis

Analisis sentimen merupakan sebuah metode dalam melakukan ekstraksi data opini, merujuk pada proses memahami dan mengelola data secara otomatis untuk melihat sentimen yang terdapat dalam sebuah opini (Sari & Wibowo, 2019). Dalam ranah yang lebih luas dalam pemrosesan bahasa alami, linguistik komputasional, dan penambangan teks yang bertujuan untuk menganalisis pendapat, sentimen, dan emosi seseorang. Sentimen yang diambil sendiri, terkait dengan suatu topik, produk, layanan, organisasi, individu, atau kegiatan tertentu baik itu pembicara atau penulis.

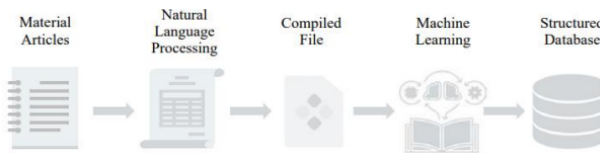
Tugas pokok dalam analisis sentimen untuk mengelompokkan teks yang terdapat dalam sebuah kalimat atau dokumen. Pendapat yang disampaikan dalam kalimat atau dokumen tersebut nantinya dilakukan pengelompokan yang bersifat positif, negatif, atau netral. Selain itu, analisis sentimen juga dapat mengungkapkan perasaan emosional seperti sedih, gembira, atau marah.

Text mining merupakan sebuah metode yang digunakan untuk menangani permasalahan terhadap Klasifikasi, Clustering, Information extraction, dan information retrieval (Afdal & Elita, 2022). Text mining bertujuan untuk menghasilkan informasi dari kumpulan dokumen. Kemampuan Text Mining terletak pada kemampuannya untuk menghasilkan informasi melalui,

Muhammad Khumaidi Nursyarif, dkk (2024). Analisis Sentimen Terhadap Calon Wakil Presiden Gibran Rakabuming Raka Menggunakan Algoritma Naive Bayes

proses, pengelompokan, dan analisis data yang bersifat tidak terstruktur dalam jumlah besar. Teknik ini digunakan untuk mendapatkan informasi berharga dari serangkaian dokumen yang bersumber dari teks dengan format yang tidak terstruktur.

Proses ekstraksi informasi dalam text mining dapat mencakup analisis sentimen, yang secara emosional mengidentifikasi apakah pernyataan bersifat positif atau negatif. Objek dari text mining umumnya berupa dokumen yang tidak terstruktur atau semi terstruktur. Secara efektif, text mining mampu mengekstrak informasi yang diperlukan dari sejumlah dokumen. Gambar 1 menggambarkan bagaimana teks yang tidak terstruktur diubah menjadi materi data terstruktur dalam text mining, yang kemudian disimpan dalam basis data yang terstruktur.



Gambar 1 Alur Text Mining

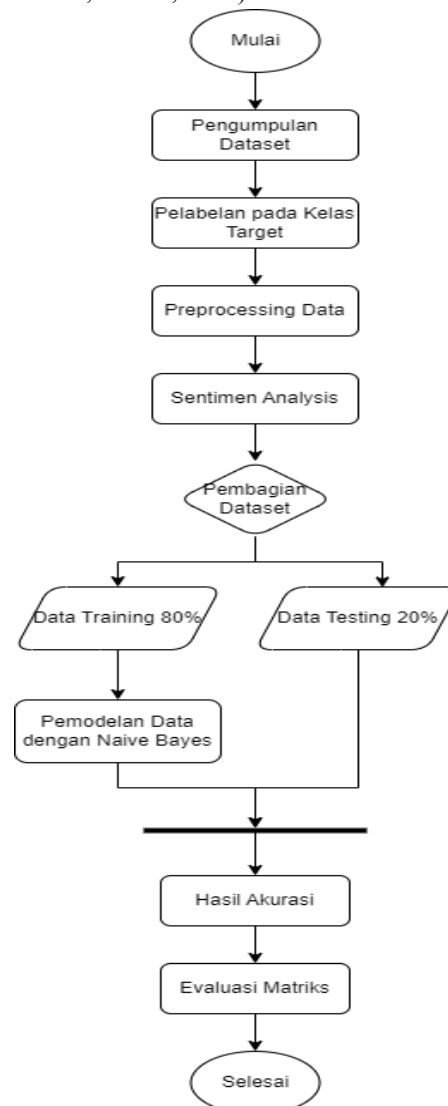
Youtube Application Programming Interface merupakan Public API yang dimiliki oleh Youtube, berfungsi untuk para pengembang dapat membuat sebuah aplikasi atau program yang memiliki fitur yang tersedia oleh Youtube (Alamri et al., n.d.). Application Programming Interface memberikan kemampuan kepada pengembang untuk menggabungkan dua bagian aplikasi atau bahkan mengintegrasikan aplikasi yang berbeda secara bersamaan. Pada penelitian yang dilakukan, API Youtube digunakan khususnya untuk melakukan pengambilan data komentar di salah satu video Youtube. Data yang diambil melalui proses scrapping yang kemudian dilakukan pengolahan dan dimodelkan untuk mendapatkan kesimpulan yang informatif.

Naïve Bayes Classifier adalah salah satu algoritma untuk klasifikasi yang berdasarkan teorema Bayes (R. Sari & Hayuningtyas, 2019). Dalam perbandingan dengan model klasifikasi lainnya, seperti Decision Trees atau Neural Network, Naive Bayes Classifier menunjukkan kinerja yang sangat baik. Keunggulan penggunaan metode ini terletak pada kebutuhan data pelatihan (training data) yang relatif kecil untuk menetapkan estimasi parameter yang diperlukan dalam proses klasifikasi. Secara umum, Naïve Bayes Classifier memiliki formulasi standar seperti dijelaskan berikut:

$$P(A|B) = \frac{P(A|B)P(A)}{P(B)}$$

- A = hipotesis data B adalah kelas khusus
- B = data dengan kelas tidak dikenal
- $P(A|B)$  = probabilitas hipotesis A didasarkan pada kondisi B
- $P(B|A)$  = probabilitas hipotesis B didasarkan pada kondisi A
- $P(A)$  = probabilitas hipotesis A
- $P(B)$  = probabilitas B

Naive Bayes menerapkan fungsi statistik menggunakan teorema Bayes dengan asumsi keberadaan dari suatu fitur tertentu terhadap suatu kelas yang tidak berhubungan dengan fitur lainnya (Wilianto, Pudjiantoro, & Umbara, 2017). Naïve Bayes merupakan suatu metode klasifikasi yang menggunakan perhitungan probabilitas (Ling, Kencana, & Oka, 2014).



Gambar 2 Alur Perancangan Sistem

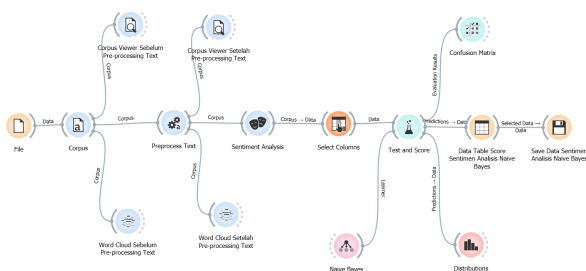
Berdasarkan Gambar 2 dilakukan proses awal sesuai gambar alur diatas dengan studi pustaka melalui beberapa jurnal, buku dan sumber lainnya. Dari hal tersebut menjadi

referensi dalam melaksanakan penelitian tentang Analisis Sentimen terhadap calon Wakil Presiden Gibran Rakabuming Raka Menggunakan Algoritma Naive Bayes. Tahap pengumpulan data diperoleh dari proses Crawling data komentar di Youtube dengan menggunakan Python. Langkah selanjutnya kami melakukan pelabelan menggunakan modul vader\_lexicon.

Data kemudian dilakukan preprocessing dan sentimen analisis menggunakan software Orange. Untuk Proses Split Data dibagi menjadi 80% data training dan 20% data testing, yang kemudian data training dilakukan pemodelan menggunakan algoritma Naive Bayes. Hasil pelatihan didapatkan nilai akurasi, presisi, recal, dan nilai fl dari model yang dibuat. Model kemudian divisualisasikan klasifikasi menggunakan Confusion Matrix untuk melihat penyebaran data.

**HASIL DAN PEMBAHASAN**

Hasil analisa dari penelitian yang dilakukan didapatkan analisis sentimen menggunakan Algoritma Naive Bayes melalui twitter. Dari proses pengumpulan data, selanjutnya yang kita lakukan adalah mengolah data ke dalam aplikasi Orange. Aplikasi ini merupakan sebuah aplikasi yang dirancang untuk bisa melakukan proses analytics dengan menerapkan konsep Visual Programing (Free Coding). Data diinput ke dalam Orange untuk melakukan pemodelan data menggunakan algoritma Naive Bayes dengan menggunakan beberapa widget yang sudah tersedia di dalam aplikasi. Berikut merupakan arsitektur perancangan alur proses penelitian.



**Gambar 3 Rangkaian Alur Orange**

Preprocess Text digunakan untuk melakukan proses cleaning data pada kata-kata yang dapat menyebabkan menurunnya akurasi atau berpengaruh pada data seperti tanda “@” dan “#” dengan bermacam fitur dari Processing, contohnya pada Tokenization, Transformation, Filtering, dan lain – lain, guna membantu membersihkan, dan mengekstraksi fitur. Beberapa proses yang digunakan pada penelitian ini pada tahapan Preprocess Text adalah ;

1. Transformation
  - a. Lower case : fitur ini digunakan untuk merubah data masukan dan mentransformasi data masukan tersebut menjadi huruf kecil.
  - b. Remove accents : fitur ini digunakan untuk menghapus semua aksen atau diacritics pada teks (naïve > naive).
  - c. Parse HTML : fitur ini digunakan untuk mendeteksi tag html dan merubahnya menjadi kalimat biasa.
  - d. Remove URLs : fitur ini digunakan untuk menghapus url.
2. Tokenization adalah metode yang digunakan untuk memecah teks menjadi komponen yang lebih kecil.
  - a. Regexp : fitur ini digunakan untuk memisahkan teks berdasarkan ekspresi regular yang ada pada aplikasi Orange Data mining.
3. Filtering
  - a. Stopwords : fitur ini digunakan untuk menghapus stopword dari teks. Pada aplikasi Orange data mining ini telah disediakan pula library stopwords untuk Bahasa Indonesia ataupun kita dapat menginput sendiri stopwords yang akan kita gunakan.
  - b. Regexp symbols: fitur ini digunakan untuk menghapus tanda baca.

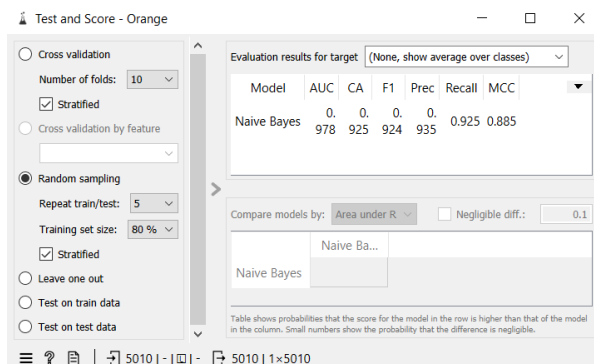


**Gambar 4 Word Cloud**

Algoritma Naïve Bayes dilakukan untuk penilaian dan matriks kebingungan dengan persyaratan bahwa dataset yang digunakan untuk pembelajaran terawasi harus memiliki variabel target atau label. Dalam penerapan algoritma data dibagi terlebih dahulu dengan pembagian data latihan sebesar 80% dan data testing sebesar 20%.

Hasil dari analisis sentimen pengguna Youtube terhadap calon Wakil Presiden Gibran Rakabuming Raka menggunakan Algoritma Naive Bayes mendapat nilai

Accuracy : 92%, Precision : 93%, Recall : 92%, dan F1 Score : 92%. Berikut adalah gambar Hasil Evaluasi Sentimen Analysis.



**Gambar 5 Akurasi Naive Bayes**

Data yang telah selesai dikelompokkan akan menghasilkan popularitas dengan bantuan confusion matrix seperti pada gambar berikut.

		Predicted			Σ
		negatif	netral	positif	
Actual	negatif	750	0	10	760
	netral	185	1678	177	2040
	positif	6	0	2204	2210
Σ		941	1678	2391	5010

**Gambar 6 Confusion Matrix**

**KESIMPULAN**

Penerapan algoritma Naive Bayes yang digunakan dalam melakukan klasifikasi analisis sentimen terhadap opini masyarakat yang mengandung sentimen negatif, positif dan netral menghasilkan nilai akurasi yang sangat baik. Akurasi yang didapatkan berdasarkan proses cleansing dan preprocessing yang dilakukan sehingga menghasilkan akurasi yang diharapkan. Dari total data 5010 data masuk diperoleh 2210 sentimen positif, 2040 sentimen netral dan 760 sentimen negatif. Hasil yang didapatkan dari penggunaan algoritma Naive Bayes terhadap sentimen analisis calon Wakil Presiden Gibran Rakabuming Raka menghasilkan nilai Akurasi sebesar 92,5%, f1 Score 92,4%, Precision 93,5% dan Recall 92,5%.

**REFERENSI**

Ihendra, J., & Laugu, N. (2020). Eksistensi Media Sosial, Youtube, Instagram dan Whatsapp Ditengah Pandemi Covid-19 Dikalangan Masyarakat Virtual Indonesia. Baitul Ulum: Jurnal Ilmu Perpustakaan Dan Informasi, 4(1).

<https://databooks.com>

Afdal, M., & Elita, L. R. (2022). PENERAPAN TEXT MINING PADA APLIKASI TOKOPEDIA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR. *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, 8(1), 78. <https://doi.org/10.24014/rmsi.v8i1.16595>

Alamri, S. D., Lumenta, A. S. M., & Rindengan, Y. D. Y. (n.d.). KKT Report Video Management App On Youtube.

Anggraeni, D. T. (2022). Analisis Sentimen Vaksinasi Booster Covid-19 pada Platform Twitter Menggunakan Metode Naive Bayes. *EXPERT: Jurnal Manajemen Sistem Informasi Dan Teknologi*, 12(2), 113. <https://doi.org/10.36448/expert.v12i2.2812>

Duei Putri, D., Nama, G. F., & Sulistiono, W. E. (2022). Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier. *Jurnal Informatika Dan Teknik Elektro Terapan*, 10(1). <https://doi.org/10.23960/jitet.v10i1.2262>

Farah Zhafira, D., Rahayudi, B., & Korespondensi, P. (2021). ANALISIS SENTIMEN KEBIJAKAN KAMPUS MERDEKA MENGGUNAKAN NAIVE BAYES DAN PEMBOBOTAN TF-IDF BERDASARKAN KOMENTAR PADA YOUTUBE (Vol. 2, Issue 1).

Fatima Barata, M., & Simanjuntak, M. B. (2019). STRATEGI PUBLISITAS & PROPAGANDA POLITIK (Studi pada Kandidat Presiden & Wakil Presiden, Jokowi-KH. Ma'ruf Amin) PUBLICITY STRATEGIES & PROPAGANDA POLITICS (Study of Candidates for President & Vice President, Jokowi-KH. Ma'ruf Amin).

Furqon, M. (2021). UPAYA PERLINDUNGAN HUKUM TERHADAP KONTEN KREATOR PADA PLATFORM YOUTUBE Legal Protection Measures To Creators' Content on Youtube Platforms.

Hidayatillah, R., Mirwan, M., Hakam, M., & Nugroho, A. (2019). Levels of Political Participation Based on Naive Bayes Classifier. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 13(1), 73. <https://doi.org/10.22146/ijccs.42531>

Joshua Lefrandt Thanos, C., Rumokoy, D. A., & Pondaag, A. H. (2020). TINJAUAN YURIDIS MENGENAI AMBANG BATAS PENCALONAN PRESIDEN DAN WAKIL PRESIDEN MENURUT UNDANG-UNDANG NOMOR 7 TAHUN 2017 TENTANG PEMILIHAN UMUM 1 Oleh. In *Lex Administratum: Vol. VIII (Issue 3)*. <https://editorialindonesia.com/2017/07/21/ini-arti->

- Ling, J., Kencana, I. P. E. N., & Oka, T. B. (2014). ANALISIS SENTIMEN MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER DENGAN SELEKSI FITUR CHI SQUARE. *EJurnal Matematika*, 3(3), 92–99.
- Ruhyana, N. (2019). ANALISIS SENTIMEN TERHADAP PENERAPAN SISTEM PLAT NOMOR GANJIL/GENAP PADA TWITTER DENGAN METODE KLASIFIKASI NAIVE BAYES. [www.situs.com](http://www.situs.com)
- Santoso, E. B., & Nugroho, A. (2019). Analisis Sentimen Calon Presiden Indonesia 2019 Berdasarkan Komentar Publik Di Facebook. *Eksplora Informatika*, 9(1), 60–69. <https://doi.org/10.30864/eksplora.v9i1.254>
- Sari, F. V., & Wibowo, A. (2019). ANALISIS SENTIMEN PELANGGAN TOKO ONLINE JD.ID MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER BERBASIS KONVERSI IKON EMOSI. 10(2).
- Sari, R., & Hayuningtyas, R. Y. (2019). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Pada Wisata TMII Berbasis Website. *Indonesian Journal on Software Engineering (IJSE)*, 5(2), 51–60. <https://doi.org/10.31294/ijse.v5i2.6957>
- Wilianto, L., Pudjiantoro, T. H., & Umbara, F. R. (2017). Analisis Sentimen Terhadap Tempat Wisata Dari Komentar Pengunjung Dengan Menggunakan Metode Naive Bayes Classifier Studi Kasus Jawa Barat. In *SNATIF*.
- Zaini, F., #1, A., Fauzi, M., Arifandy, S., Rasyad, M., #3, C., Aini, N., & #4, R. (2021). Bagaimana Masyarakat Menyikapi Pembelajaran Tatap Muka: Analisis Komentar Masyarakat pada Media Sosial Youtube Menggunakan Algoritma Deep Learning Sekuensial dan LDA. In *JLK (Vol. 4)*.