

# Sentiment Analysis of Using the YouTube Application Using the Naïve Bayes Method

Triana Putri, Siti Nurhaliza, Dodi Vionanda\*

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

\*Corresponding author: dodi\_vionanda@fmipa.unp.ac.id

Submitted : 10 Februari 2025

Revised : 17 Februari 2025

Accepted : 28 Februari 2025

## ABSTRACT

*The purpose of this study is to analyze user sentiment towards the YouTube application using the Naive Bayes method. With the increasing number of YouTube users worldwide, it is very important to understand user preferences and experiences. Sentiment analysis, the process of processing or extracting textual data to obtain information by classifying sentiment as positive or negative. Due to its efficiency and effectiveness, the Naive Bayes algorithm is a statistical method commonly used in sentiment analysis and natural language analysis. The data collection process in this study involved web scraping followed by various procedures such as cleaning, case folding, tokenization, stopword removal, and stemming. Feature selection was done using TF-IDF (Term Frequency-Inverse Document Frequency) to apply weights to each word according to its level of importance. The Naïve Bayes classifier was then applied to the previously collected data, and its accuracy was assessed using the accuracy, precision, recall, and F1-score metrics. The results showed that the Naïve Bayes method was effective in sentiment analysis for the YouTube application, with an accuracy of 82%, a precision of 83%, a recall of 98%, and an F1-score of 89%. This research provides valuable insights into YouTube user sentiment, enabling content creation and analysis to improve user experience and marketing strategies.*

**Keywords:** *YouTube, Sentiment Analysis, Naive Bayes, User Reviews*



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

## I. PENDAHULUAN

YouTube adalah platform berbagi video daring yang memungkinkan pengguna untuk mengunggah, menonton, dan berinteraksi dengan berbagai jenis video, mulai dari musik, vlog pribadi, tutorial, materi edukasi hingga hiburan (Febrialdi et al., 2024). Pada tahun 2022, YouTube tercatat memiliki 2,56 miliar pengguna di seluruh dunia, dan diprediksi jumlah tersebut akan mencapai 2,85 miliar pengguna pada tahun 2025 (Arrahman et al., 2024). Pada awal tahun 2022, jumlah pengguna YouTube di Indonesia mencapai 139 juta orang. Namun, sejalan dengan popularitasnya penting untuk memahami bagaimana pengguna merespons dan mengevaluasi pengalaman mereka dengan menggunakan aplikasi YouTube ini.

Salah satu cara untuk menganalisis respons pengguna adalah dengan melakukan analisis sentimen. Proses ini melibatkan pengolahan atau ekstraksi data tekstual guna menggali informasi yang terdapat dalam sebuah kalimat. Analisis sentimen bertujuan untuk memperoleh opini dari pengguna pada media sosial atau media lainnya yang ada di internet. Melalui analisis sentimen, teks dari komentar, produk, atau topik dapat diklasifikasikan ke dalam kategori sentimen positif atau negatif (Parulian et al., 2021). Dalam konteks ini, kita akan mengkaji bagaimana pengguna di Indonesia memberikan rating serta umpan balik terhadap pengalaman mereka menggunakan aplikasi YouTube.

Adapun kajian ini menggunakan pendekatan *Naive Bayes*, sebuah pendekatan statistik yang umum diterapkan dalam analisis sentimen dan pemrosesan bahasa alami. Keistimewaan utama dari algoritma *naive bayes* adalah anggapan kuat (naif) mengenai independensi antar kondisi atau suatu kejadian yang ada (Widianto et al., 2019). Metode *naive bayes* terbukti efektif dalam mengklasifikasikan teks berdasarkan sentimen, dengan memanfaatkan probabilitas dalam memprediksi kategori sentimen suatu teks. Kekuatan *Naive Bayes* adalah kemampuannya untuk mengklasifikasikan dengan jumlah data pelatihan yang kecil. Metode ini dapat menghitung parameter secara efisien, sehingga mencapai akurasi maksimum dalam waktu yang lebih singkat (Dalilah et al., 2023). Metode ini dinilai cocok untuk analisis sentimen karena berfungsi sebagai algoritma klasifikasi yang dapat mambagi data ke dalam kategori sentimen positif dan negatif secara efektif.

Penelitian Patrik Galih Adi Prananda, dkk (2025) tentang Penerapan Algoritma Naive Bayes untuk Klasifikasi Sentimen Ulasan Pengguna Aplikasi Dramabox dari Ulasan Play Store. Penelitian ini mengklasifikasi ulasan menjadi dua kategori sentimen, yaitu positif dan negatif. Dengan pembagian data latih 80% dan data uji 20%, dan memperoleh akurasi 74.72%, hal ini menunjukkan performa cukup baik dalam memprediksi sentimen secara keseluruhan. Model ini juga memiliki kemampuan yang luar biasa dalam mendeteksi sentimen negatif dengan *recall* 99.15% dan *F1-Score* 83.87% (Patrik Galih Adi Prananda et al., 2025).

Penelitian Riska Astuti dan Angel Tiarma Sipahutar (2025) tentang Perbandingan Klasifikasi Berita Hoax Politik pada Media Sosial X Menggunakan Algoritma Naive Bayes dan Random Forest. Hasilnya, *random forest* memiliki performa yang lebih unggul dengan nilai akurasi sebesar 89%, presisi sebesar 83%, *recall* sebesar 89%, dan *F1-Score* sebesar 89%, dibandingkan *naive bayes* yang menghasilkan nilai akurasi sebesar 77%, presisi sebesar 78%, *recall* sebesar 77% dan *F1-Score* sebesar 77%. Penelitian ini juga menekankan peran penting teknologi dalam mendeteksi hoax, menunjukkan bahwa algoritma *machine learning* dapat digunakan secara efektif untuk mengurangi penyebaran informasi yang salah di media sosial (Astuti & Sipahutar, 2025).

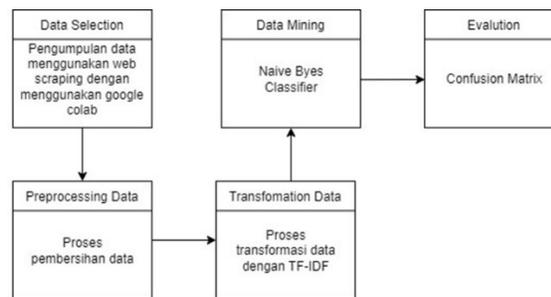
Penelitian Nabil Safiq Ramadan dan Dedi Darwis (2025) tentang Perbandingan Metode *Naive Bayes* dan *SVM* untuk Sentimen Analisis Masyarakat Terhadap Serangan *Ransomware* pada Data KIP-K. Pada penelitian ini data dikumpulkan melalui proses *crawling* menggunakan API dan diproses melalui tahapan *labelling*, *preprocessing*, dan klasifikasi. Dari 2.648 data mentah, diperoleh 1.738 data yang bersih. Hasil klasifikasi menunjukkan bahwa *SVM* memiliki akurasi yang lebih tinggi (88%) dibandingkan *Naive Bayes Classifier* (70%). Mayoritas sentimen yang ditemukan bersifat negatif (49.6%), menunjukkan kekhawatiran pengguna terhadap insiden tersebut serta harapan agar pemerintah meningkatkan keamanan data (Ramadan & Darwis, 2025).

Analisis sentimen penggunaan aplikasi YouTube memiliki signifikansi yang besar dalam memahami preferensi, tren, dan pandangan pengguna terhadap platform ini. Hasil dari analisis ini dapat menyajikan wawasan yang bernilai bagi *software developer*, pemasar konten, dan peneliti untuk mengembangkan pengalaman pengguna dan strategi pemasaran di berbagai wilayah, serta meningkatkan kepuasan dan loyalitas pengguna yang mana bertujuan untuk memberikan pemahaman mendalam tentang sentimen pengguna YouTube menggunakan metode *naive bayes* sebagai kerangka analisis utama.

## II. METODE PENELITIAN

### 2.1 Pengumpulan Data

Kumpulan data yang digunakan dalam kajian ini ialah ulasan pengguna aplikasi *YouTube* yang diambil dengan bantuan *google colab* yang berjumlah 995 ulasan dengan kategori ulasan yang paling relevan. Data tersebut diambil dengan menggunakan *web scraping* yang diperoleh dari link website *google playstore* pada aplikasi *YouTube*. Penelitian ini menggunakan *tools google colab* dengan bahasa pemrograman *python* untuk membantu dalam pengambilan dan pengolahan data. Adapun kajian ini menggunakan teknik *Knowledge Discovery in Database* (KDD). Prosedur pada KDD yaitu *Data Selction*, *Preprocessing Data*, *Transformation*, *Data Mining*, dan *Evaluation*.



Gambar 1. Alur Penelitian

### 2.2 Web Scraping

*Web scraping* merupakan proses yang dilakukan untuk mengumpulkan informasi dari sebuah situs web. Ini dilakukan dengan menggunakan sebuah program untuk mengekstrak konten dari situs web dan menyimpannya dalam format yang lebih terstruktur, seperti spreadsheet.

### 2.3 Preprocessing Data

*Preprocessing data* adalah serangkaian langkah yang dilakukan membersihkan data agar lebih terstruktur dan relevan sebelum dianalisis. *Cleaning* data mencakup penghapusan URL, *mention*, *hashtag*, serta karakter lain selain huruf dan spasi. Selanjutnya, dilakukan tahap *case folding*, *tokenizing*, *stopword*, dan *stemming* untuk meningkatkan kualitas data yang akan digunakan.

### 2.4 Transformation Data

Transformasi dilakukan menggunakan fitur TF-IDF untuk menerapkan bobot pada setiap kata berdasarkan frekuensinya menggunakan skema pembobotan TF-IDF. Metode ini berfungsi untuk mengurangi bobot kata-kata yang umum digunakan, meningkatkan bobot kata-kata yang jarang muncul namun memiliki nilai penting dalam menentukan isi data, serta membantu dalam pencarian informasi yang lebih mendetail dan meningkatkan kinerja model.

### 2.5 Naïve Bayes Classifier

*Naïve bayes* adalah algoritma pada *data mining* yang digunakan untuk mengklasifikasikan objek berdasarkan prinsip probabilitas dan statistik, serta mampu memprediksi kemungkinan suatu kejadian dengan menganalisis frekuensi atau jumlah kemunculan setiap kategori pada data pelatihan. Proses klasifikasi ada dua tahap, yaitu pelatihan data (*training*) dan pengujian data (*testing*) pada model yang telah dibangun. Dataset dibagi menggunakan metode split memakai perbandingan 8:2, di mana 80% data digunakan untuk pelatihan dan 20% untuk pengujian. Pemisahan dataset dilakukan secara acak dari keseluruhan data yang tersedia (Syamsul et al., 2022). Data pelatihan berguna untuk mengembangkan model klasifikasi, sedangkan data pengujian berguna untuk menilai kinerja model. Adapun rumus teorema bayes dijelaskan pada Persamaan (1).

$$P(H|X) = \frac{p(X|H) \cdot P(H)}{P(X)} \tag{1}$$

Keterangan :

- X : Data yang tidak diketahui kelasnya
- H : Hipotesis dari data X yang merupakan kelas spesifik
- P(H|X) : Probabilitas hipotesis H berdasarkan kondisi X
- P(H) : Probabilitas hipotesis H
- P(X|H) : Probabilitas X berdasarkan kondisi dari hipotesis H
- P(X) : Probabilitas dari X

### 2.6 Evaluation

Proses ini dilakukan dengan menggunakan *Confusion Matrix*, hal ini guna untuk pertimbangan kineja model Naïve Bayes untuk memvisualisasikan kinerja prediksi model dengan membandingkan nilai prediksi dengan nilai yang sesungguhnya. Matriks ini mencakup nilai akurasi, nilai akurasi, *recall*, dan *F1-score*. Matriks ini dapat mencerminkan seberapa baik model dapat mengklasifikasikan sentiment ulasan dengan benar.

- 1) *Accuracy* : mengukur tingkat kesesuaian antara nilai prediksi yang dihasilkan model dengan nilai sesungguhnya.
- 2) *Precision* : menunjukkan seberapa ketepatan model dalam mengklasifikasikan data sebagai positif. *Precision* menjadi metrik penting ketika tingkat *False Positive* dalam model cukup tinggi.
- 3) *Recall* : menghitung proporsi data positif aktual yang berhasil diidentifikasi dengan benar sebagai *True Positive*. *Recall* menjadi metrik yang penting saat *False Negative* memiliki dampak signifikan.
- 4) *F1-score* : rata-rata nilai antara *precision* dan *recall* untuk memberikan cerminan yang setara mengenai kinerja model.

## III. HASIL DAN PEMBAHASAN

Kajian ini menerapkan teknik *web scraping* serta memakai library “google-play-scraper” dengan bahasa pemrograman Python di *Google Colab*. Data ini dikumpulkan mencakup berbagai informasi, seperti nama pengguna, rating, waktu ulasan, isi ulasan, serta label sentimen (positif atau negatif).

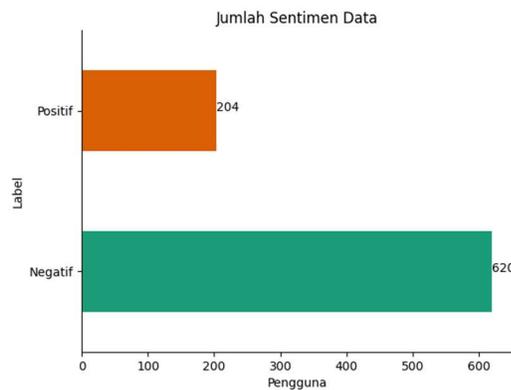
Tabel 1. Hasil Scraping Data

UserName	score	at	content	Label
Roma Dhoni	3	27/05/2024 12:37	1. <input checked="" type="checkbox"/> Saya mohon. Tolong tambah-kan tombol ter- lama di-youtube short-s. Se-cepat-nya. Terimakasih... 2. [...] Saya mohon. Tolong di setiap semua channel ada	

UserName	score	at	content	Label
			status aktif atau tidak dalam YouTube di semua channel yang ada tersebut. Terimakasih...	
Mickiliem Bun	1	27/05/2024 12:23	gausah di update versi baru youtube sekarang sudah di predikat aplikasi tidak berguna untuk nonton tidak percaya bahwa updatean bisa berguna tambah rusak sama sekali youtube adalah aplikasi tidak berguna sama sekali dunia	Negatif
Muhar Alfian	1	27/05/2024 12:16	Knp skrng jadi jelek gini sih youtube, looks nya kek aplikasi jadul, setiap mau full layar pasti gak bisa malah ngestuck, balikin lagi lah ke youtube yg dulu	Negatif

**A. Labeling**

Selama proses *scraping* data ini, pelabelan sentimen dilakukan berdasarkan nilai di kolom “score”. Data dengan skor kurang dari tiga sebagai “Negatif”, dan data yang memiliki skor empat dan lima sebagai “Positif”. Berikut ini adalah jumlah hasil pelabelan yang telah dilakukan:



**Gambar 2.** Visualisasi untuk Data Hasil Pelabelan

Diagram diatas menunjukkan jumlah sentimen “Positif” dan sentiment “Negatif”, dimana terlihat bahwa pada ulasan youtube penggunaanya sentimen “Negatif” berjumlah 620 pengguna, sentimen “Positif” berjumlah 204 pengguna dan terdapat 171 ulasan yang tidak tergolong sentimen “Positif” dan sentimen “Negatif”.

**B.Data Cleaning**

Tahap pertama dalam pemrosesan data meliputi menghilangkan URL, *mention*, *hashtag*, dan karakter selain huruf dan spasi. Langkah ini berguna untuk memastikan bahwa analisis dibatasi pada teks yang memiliki struktur yang jelas. Berikut ini merupakan hasil dari tahap *cleaning data* yang telah dilakukan:

**Tabel 2.** Hasil Data yang Telah di *Cleaning*

Input Proses	Output proses
Halo teman", 😊 aplikasi ini sangat bagus,, video dan short nya semua bagus",, saya beri rating 5 karna,, aplikasi ini bisa melatih kekreatifan,, :)	halo teman aplikasi ini sangat bagus video dan short nya semua bagus saya beri rating karna aplikasi ini bisa melatih kekreatifan
YouTube, knp hari ini subtitle CC-nya gk keluar 🙄. Aku sampe panik Loh, nyari nyari tutorial cara memunculkan subtitle cc, eh liat komen ternyata banyak yg gk muncul juga hari ini. Tolong perbaiki yaa	youtube knp hari ini subtitle ccnya gk keluar aku sampe panik loh nyari nyari tutorial cara memunculkan subtitle cc eh liat komen ternyata banyak yg gk muncul juga hari ini tolong perbaiki yaa

### C. Case Folding

*Case folding* ialah teknik mengubah huruf menjadi bentuk yang unik, ini dilakukan dengan mengubahnya ke dalam huruf kecil (*lowercase*). Berikut ini adalah hasil dari *case folding* yang telah dilakukan:

**Tabel 3.** Hasil dari Proses *Case Folding*

Input Proses	Output proses
Tidak bisa menonton secara landscape. Sudah di landscape tapi bentuk video tetap potrait.	tidak bisa menonton secara landscape sudah di landscape tapi bentuk video tetap potrait
Memutar video tanpa bisa memperbesar layar, dan selalu close jika diperbesar.	memutar video tanpa bisa memperbesar layar dan selalu close jika diperbesar

### D. Tokenizing

Tokenisasi merupakan pemisahan teks menjadi bagian - bagian yang lebih kecil. Ini dapat berbentuk kata, frasa, atau simbol tertentu pada suatu kalimat. Berikut merupakan hasil dari proses *tokenizing* yang telah dilakukan :

**Tabel 4.** Hasil dari Proses *Tokenizing*

Input Proses	Output proses
youtube sekarang ini sering sekali subtitle tidak aktif alias tidak bekerja padahal ada pilihan untuk subtitle tapi kenapa tidak bekerja dgn baik	['youtube', 'subtitle', 'aktif', 'alias', 'pilihan', 'subtitle', 'dgn']
subtitle yang tersedia di setiap video tidak bisa dipakaisubtitle tidak muncul tolong diperbaiki	['subtitle', 'tersedia', 'video', 'dipakaisubtitle', 'muncul', 'tolong', 'diperbaiki']

### E. Stopword Removal

*Stopword removal* ialah jenis analisis teks yang bertujuan untuk menghilangkan kata-kata umum yang biasanya tidak memberikan kontribusi berarti pada teks. Berikut adalah hasil *stopword removal* yang telah dilakukan :

**Tabel 3.** Hasil dari Proses *Stopword Removal*

Input Proses	Output Proses
makin sering di update makin tambah rusak ni youtube contoh keluar sendiri tanpa alasan tolong di perbaiki bug nya	update rusak ni youtube contoh alasan tolong perbaiki bug nya
aplikasi bodoh banyak bug dan sering error video tidak bisa diputar padahal pake wifi pantes yt sepi	aplikasi bodoh bug error video diputar pake wifi pantes yt sepi

### F. Stemming

*Stemming* bertujuan untuk memangkas atau menghapus akhiran pada kata dan diperoleh bentuk kata awal atau kata yang lebih umum. Berikut adalah hasil *stemming* yang telah dilakukan:

**Tabel 4.** Hasil dari Proses *Stemming*

Input Proses	Output Proses
['nonton', 'vidio', 'bug', 'nya', 'tolong', 'diperbaiki', 'ya', 'bug', 'nya']	nonton vidio bug nya tolong baik ya bug nya
['youtube', 'tv', 'google', 'tv', 'loading', 'thumbnail', 'cepat', 'tolong', 'diperbaiki']	youtube tv google tv loading thumbnail cepat tolong baik

### G. Splitting Data

Penelitian ini dilakukan dengan membagi dataset dengan ketentuan rasio 0.2, ini dipilih agar 80% data digunakan sebagai data latih, sementara 20% sisanya dialokasikan untuk data uji.

### H. Evaluation

Selanjutnya tahap klasifikasi, proses ini guna untuk membangun model *machine learning* untuk *data training* dan *data testing* yang dipilih dengan acak. *Cross-validation* diterapkan untuk mengevaluasi performa model, sehingga



## DAFTAR PUSTAKA

- Adriani Nurian, B. N. (2023). Analisis Sentimen Ulasan Pengguna Aplikasi Google Play Menggunakan Naïve Bayes. *Jurnal Informatika dan Teknik Elektro Terapan (JITET)*, 11, Vol. 11, No. 3.
- Akmal, K. (2023). Perbandingan Metode Algoritma Naïve Bayes Dan K-Nearest Neighbors Untuk Klasifikasi Penyakit Stroke. *Jurnal Mahasiswa Teknik Informatika (JATI)*, 7, Vol. 7.
- Bobby Kurniadi Widodo, N. H. (2022). Penerapan Algoritma Naive Bayes Untuk Analisis. *Jurnal Teknologi Informasi*, 523-533.
- Dalilah Nurwahidah, G. D. (2023). Analisis Sentimen Data Ulasan Pengguna Aplikasi Google Kelas Pada Google Play Store Menggunakan Algoritma Naïve Bayes. *Jurnal Mahasiswa Teknik Informatika (JATI)*, 7, Vol. 7, No. 6.
- Fandhilah, A. A. (2023). Analisis Sentimen Aplikasi Quizizz pada Google PlayStore Menggunakan NaïveBayes. *Sains dan Teknologi Informasi (SATIN)*, 9, Vol. 9, No. 2.
- Ifran Ricky Afandi, F. N. (2022). Analisis Sentimen Opini Masyarakat Terkait Pelayanan Jasa Ekspedisi Anteraja Dengan Metode Naive Bayes. *Jurnal Linguistik Komputasional*, Vol. 5, No. 2.
- Kemp, S. (2022, Februari 15). *DIGITAL 2022 : INDONESIA*. Diambil kembali dari DATAREPORTAL: <https://datareportal.com/reports/digital-2022-indonesia>
- Mampe Parulian Munthe, A. S. (2021). Sentiment Analysis Comment on Indonesian Youtube Channel About Food Vlogger Using. *eProceedings og Engineering*, Vol. 8, No. 6.
- Mathoril Hudha, E. S. (2022). ANALISIS SENTIMEN PENGGUNA YOUTUBE TERHADAP TAYANGAN #MATANAJWAMENANTITERAWAN DENGAN METODE NAÏVE BAYES CLASSIFER. *JIKO (Jurnal Informatika dan Komputer)*, 1-6.
- Michael Suhendra, W. S. (2021). Analisis Sentimen Pada Ulasan Aplikasi Video Conference Menggunakan Naive Bayes. *Jurnal Ilmiah Sains & Teknologi (SINSBERTEK)*, 2, Vol. 2, No. 1.
- Mochammad Febrialdi Ansyah, A. G. (2024). Analisis Sentimen Komentar Youtube Terhadap Tayangan #Terbaru! Temuan dan Masalah Ahlak DiPonpes Al-zaytun Menggunakan Metode Naïve Bayes. *Jurnal Teknologi Terapan (G-Tech)*, Vol. 8, No. 2.
- Nabil Safiq Ramadan, D. D. (2025). PERBANDINGAN METODE NAÏVE BAYES DAN SVM UNTUK SENTIMENANALISISMASYARAKAT TERHADAP SERANGAN RANSOMWARE PADA DATA KIP-K. *Jurnal Sistem Informasi dan Informatika (Simika)*, Vol. 8, No. 1.
- Patrik Galih Adi Prananda, M. F. (2025). PENERAPAN ALGORITMA NAÏVE BAYES UNTUK KLASIFIKASI SENTIMEN ULASAN PENGGUNA APLIKASI DRAMABOX DARI ULASANPLAY STORE. *Journal of Data Analytics, Information, and Computer Science (JDAICS)*, Volume 2, No 1.
- Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap. *Jurnal Inovtek Polbeng*, Vo. 3, No. 1.
- Rika Astuti, A. T. (2025). PERBANDINGAN KLASIFIKASI BERITA HOAX POLITIK PADA MEDIA SOSIAL X MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN RANDOM FOREST. *Jurnal Ilmu Komputer, Sistem Informasi dan Teknologi Informasi (Innotech)*, Vol. 2, No. 1.
- Risav Arrahman Firsttama, A. A. (2024). Analisis Sentimen Komentar Youtube Konferensi Tingkat Tinggi G20 Menggunakan Metode Naive Bayes. *Jurnal Teknologi dan Sistem Informasi Bisnis (JTeksis)*, Vol. 6, No. 2.
- Syamsul Mujahidin, B. P. (2022). Implementasi Analisis Sentimen Masyarakat Mengenai Kenaikan Harga BBM Pada Komentar YoutubeDengan Metode Gaussian naïve bayes. *Jurnal Vocational Teknik Elektronika dan Informatika (VoteTEKNIKA)*, Vol. 10, No. 3.
- Widianto, M. H. (2019). "Algoritma Naive Bayes". BINUS UNIVERSITY BANDUNG - KAMPUS TEKNOLOGI KREATIF. Available : <https://binus.ac.id/bandung/2019/12/algoritma-naive-bayes/#:~:text=Ciri utama dari naive bayes ,benar%2C mengingat vektor informasi obyek.>