



JUTEKOM

Jurnal Teknologi dan Ilmu Komputer

<https://jurnal.nolsatu.co.id/jutekom>

2024 Vol.1 No.3 / Hal.107-112

Klasifikasi Sentimen Masyarakat Terhadap Aplikasi Tiktok Menggunakan Algoritma Naive Bayes

Muhammad Nahdhudin¹, Nur Cahyo Hendro Wibowo², Maya Rini Handayani³, Khotibul Umam⁴^{1,2,4} Fakultas Sains dan Teknologi, UIN Walisongo, Kota Semarang, 50111, Indonesia³ Fakultas Komunikasi dan Penyiaran Islam, UIN Walisongo, Kota Semarang, 50111, Indonesianahdhudin345@gmail.com

Abstract

The rapid development of information and communication technology has created a new digital space for people to express their opinions, views, and preferences for various social phenomena, including digital applications. One platform that has emerged in recent years is TikTok, a short video-based app that has seen a surge in popularity globally. Although TikTok has millions of active users and is considered a medium of entertainment and digital creativity, the platform has also received various criticisms. Especially related to content, privacy, and its impact on social behavior. This study aims to classify the sentiment of the Indonesian people towards the TikTok application by utilizing a machine learning approach, especially the Multinomial Naïve Bayes algorithm. The dataset used consisted of 100,000 TikTok user reviews taken from the Google Play Store using web scraping techniques. The analysis process includes several main stages, namely pre-processing of text (such as tokenization, stopword removal, and stemming), sentiment labeling based on user ratings, feature extraction using the TF-IDF (Term Frequency Inverse Document Frequency) method, and classification stage using the Multinomial Naïve Bayes algorithm. The results of the evaluation showed that the model had an accuracy of 88.2%, accuracy of 89.1%, recall of 87.4% and an F1-score of 88.2%. The majority of user reviews are positive. These results show that Naive Bayes is effective in classifying text-based opinions in Indonesian. This research can be used as a reference in understanding public perception and the application of machine learning in the analysis of public opinion

Keywords: Analisis sentiment, Tiktok, Naïve Bayes, Text Mining, Natural Language Processing, Machine Learning

Abstrak

Perkembangan pesat teknologi informasi dan komunikasi telah menciptakan ruang digital baru bagi Masyarakat dalam mengekspresikan opini, pandangan, dan preferensi terhadap berbagai fenomena sosial, termasuk aplikasi digital. Salah satu platform yang mencuat dalam beberapa tahun terakhir adalah TikTok, sebuah aplikasi berbasis video pendek yang mengalami lonjakan popularitas secara global. Meskipun TikTok memperoleh jutaan pengguna aktif dan dianggap sebagai media hiburan serta kreativitas digital, platform ini juga menuai berbagai kritik. Terutama terkait konten, privasi, dan dampaknya terhadap perilaku sosial. Penelitian ini bertujuan untuk mengklasifikasikan sentimen Masyarakat Indonesia terhadap aplikasi TikTok dengan memanfaatkan pendekatan *machine learning*, khususnya algoritma Multinomial Naïve Bayes. Dataset yang digunakan terdiri dari 100.000 ulasan pengguna TikTok yang diambil dari Google Play Store menggunakan Teknik *web scraping*. Proses analisis meliputi beberapa tahapan utama, yaitu pra-pemrosesan teks (seperti *tokenisasi*, *stopword removal*, dan *stemming*), pelabelan sentimen berdasarkan rating pengguna, ekstraksi fitur menggunakan metode TF-IDF (Term Frequency Inverse Document Frequency), dan tahap klasifikasi menggunakan algoritma Multinomial Naïve Bayes. Hasil evaluasi menunjukkan bahwa model memiliki akurasi sebesar 88,2% precision 89,1%, recall 87,4% dan F1-score 88,2%. Mayoritas ulasan pengguna bersentimen positif. Hasil ini menunjukkan bahwa Naive Bayes efektif dalam mengklasifikasi opini berbasis teks dalalam bahasa Indonesia. Penelitian ini dapat dijadikan acuan dalam memahami persepsi publik serta penerapan *machine learning* dalam analisis opini masyarakat.

Kata kunci: Analisis sentiment, Tiktok, Naïve Bayes, Text Mining, Natural Language Processing, Machine Learning

Jutekom is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.



1. Pendahuluan

Perkembangan teknologi digital telah merevolusi cara manusia berkomunikasi, mencari informasi, dan mengekspresikan diri. Media sosial menjadi salah satu produk dari era digital yang paling berpengaruh, dengan pertumbuhan pengguna yang terus meningkat setiap tahunnya. Aplikasi media sosial ini tidak hanya menjadi alat komunikasi, tetapi juga menjadi ruang publik tempat lahirnya opini, tren, dan fenomena sosial baru.

Salah satu platform media sosial yang mengalami lonjakan pengguna secara signifikan adalah TikTok, sebuah aplikasi berbasis video pendek yang diluncurkan oleh perusahaan asal Tiongkok, ByteDance. Sejak dirilis secara global pada tahun 2018, TikTok telah diunduh lebih dari 2 miliar kali ini diseluruh dunia dan menjadi salah satu aplikasi yang populer di berbagai negara. Aplikasi ini bisa memungkinkan pengguna membuat video berdurasi pendek dengan musik, efek visual, dan filter kreatif, yang kemudian dapat dibagikan secara luas.[1]

Meski mendapatkan popularitas luar biasa, TikTok juga menghadapi berbagai kontroversi. Beberapa pihak menganggap platform ini sebagai sarana ekspresi yang positif dan kreatif, sementara yang lain menilai bahwa TikTok menyebarkan konten yang tidak pantas, adiktif, dan berpotensi merusak nilai-nilai sosial. situasi ini menjadikan TikTok sebagai objek yang menarik untuk dianalisis, khususnya dalam hal persepsi dan sentimen masyarakat terhadapnya.[1]

Analisis sentimen adalah satu teknik dalam text Mining dan Natural Language Processing (NLP) yang bertujuan untuk mengetahui opini publik berdasarkan analisis teks, seperti komentar, ulasan, atau postingan media sosial. Analisis ini dapat mengidentifikasi kecenderungan emosi atau sikap seseorang terhadap suatu subjek tertentu, yang klasifikasi ke dalam kategori positif, negatif, dan netral.[3]

Dalam penelitian ini, algoritma Naive Bayes Classifier digunakan sebagai metode utama untuk klasifikasi sentimen. Naive Bayes merupakan algoritma klasifikasi berbasis probabilitas yang sederhana namun efektif, terutama dalam menangani data teks. [4] Keunggulan algoritma ini terletak pada efisiensi komputasi serta kemampuannya dalam memberikan hasil yang

cukup akurat meskipun asumsi independensi antar fitur seringkali tidak sepenuhnya terpenuhi.[5]

Penelitian ini bertujuan untuk mengkaji opini Masyarakat Indonesia terhadap TikTok berdasarkan data ulasan pengguna dari Google Play store. Dengan menggunakan Teknik web scraping, text preprocessing, ekstraksi fitur dengan metode TF-IDF, dan klasifikasi menggunakan Multinomial Naïve Bayes.[6]

2. Metodologi Penelitian

2.1. Data Penelitian

Data yang digunakan dalam penelitian ini diperoleh melalui proses web scraping terhadap ulasan aplikasi TikTok yang terdapat di Google Play Store. Pengambilan data dilakukan menggunakan library Python seperti selenium dan pandas. Dataset akhir berisi sekitar 100.000 baris ulasan.

2.2. Perancangan

Perancangan penelitian ini disusun sebagai rangkaian proses sistematis yang bertujuan untuk mengklasifikasikan sentimen masyarakat terhadap aplikasi TikTok. Proses ini terbagi dalam beberapa tahapan, pengumpulan data yang menggunakan teknik scraping untuk mengambil ulasan aplikasi TikTok dari Google Play Store, Pra-pemrosesan teks yang membersihkan dan menyiapkan data ulasan untuk dianalisis, Labeling sentimen yang memberikan label sentimen pada data berdasarkan nilai rating, Ekstraksi fitur yang dimana mengubah teks ulasan menjadi representasi numerik, Pelatihan model dengan menggunakan algoritma Multinomial Naive Bayes untuk melakukan pelatihan model berdasarkan data latih, Pengujian dan Evaluasi yang menguji performa model dengan data uji dan mengukur akurasi serta metrik lainnya.[7]

2.3. Analisis Sentimen

Analisis Sentimen adalah proses komputasional untuk mengidentifikasi dan mengkategorikan opini yang dinyatakan dalam suatu teks, terutama untuk menentukan apakah sikap penulis terhadap suatu topik bersifat positif, negatif, atau netral. Teknik ini banyak digunakan dalam berbagai bidang seperti pemasaran, politik, ekonomi, dan kajian sosial, terutama ketika opini publik

terhadap suatu topik ingin dianalisis secara otomatis dalam skala besar.[8]

Sumber data untuk analisis sentimen biasanya berasal dari media sosial, ulasan pengguna (user reviews), komentar publik, dan forum diskusi. Dalam konteks ini, analisis sentimen terhadap Tiktok dapat memberikan gambaran umum mengenai persepsi masyarakat terhadap aplikasi tersebut.[9]

2.4. Text Mining dan Natural Language Processing (NLP)

Text Mining adalah proses untuk mengekstraksi informasi berguna dari kumpulan teks tidak terstruktur melalui teknik statistik, pembelajaran mesin, dan linguistik komputasional. Sedangkan Natural Language Processing (NLP) merupakan cabang dari kecerdasan buatan (AI) yang berkaitan dengan interaksi antara komputer dan bahasa manusia.[10] Dalam analisis sentimen, NLP juga digunakan untuk memproses teks agar bisa dianalisis secara otomatis oleh mesin, termasuk dalam proses tokenisasi, stemming, stopword removal, dan konversi ke bentuk fitur numerik.[11]

2.5. Naive Bayes Classifier

Naive Bayes Classifier merupakan algoritma pembelajaran mesin berbasis probabilistik yang digunakan untuk klasifikasi data. algoritma ini didasarkan pada Toerema Bayes dengan asumsi bahwa setiap fitur bersifat independen satu sama lain. Dalam konteks analisis sentimen, fitur-fitur tersebut biasanya berupa kata-kata dalam dokumen teks.[12]

Rumus dasar Toerem Bayes dinyatakan sebagai berikut:

$$P(C|X) = P(X|C)P(C)P(X)$$

Keterangan:

$P(C|X)$ adalah probabilitas hipotesis C (kelas tertentu, misal positif) terhadap data X (fitur input)

$P(X|C)$ adalah probabilitas data X diberikan kelas C

$P(C)$ adalah probabilitas awal dari kelas C

$P(X)$ adalah probabilitas awal dari data X

Naive Bayes memiliki keunggulan dalam hal efisiensi, kecepatan, dan performa yang cukup baik meskipun dengan asumsi yang sederhana. Karena itu, algoritma ini sering digunakan untuk klasifikasi teks, termasuk klasifikasi sentiment. [13]

2.6. Pra-pemrosesan Teks

Pra-pemrosesan teks (text preprocessing) adalah tahap awal yang sangat penting dalam text mining dari NLP. khususnya untuk memastikan bahwa data teks dapat dianalisis secara akurat, Proses ini meliputi beberapa tahapan, antara lain:

- Tokenisasi: memecah kalimat menjadi kata-kata atau token
- Stopword Removal: Menghapus kata-kata umum yang tidak memiliki makna penting (seperti "dan", "yang", "adalah").
- Stemming: Mengubah kata ke bentuk dasar (misalnya, "bermain", "bermainan", menjadi "main").
- Lowercasing: Mengubah semua huruf menjadi huruf kecil agar tidak dianggap berbeda.

Proses ini penting agar model Naive Bayes dapat mengenali pola kata dengan lebih akurat dalam proses klasifikasi.[14]

3. Hasil dan Pembahasan

Tabel 1. Hasil Pra-pemrosesan Teks

No	Review Asli	Setelah Proprocessing
1	"Bagus aku suka banget"	"bagus suka banget"
2	"Bisa mendownload tapi tidak bisa masuk..."	"download tidak bisa masuk"
3	"Tiktok sekarang udah jelek kebanyakan update"	"tiktok jelek update"

a) Distribusi Data Sentimen

Data telah diberi label sentimen menjadi tiga kelas yaitu, Positif, Negatif, dan Netral. Berikut adalah distribusinya:

Tabel 2. Distribusi Data Sentimen

Sentimen	Jumlah Data	Presentase
Positif	62.312	62.31%
Negatif	28.504	28.50%
Netral	9.184	9.18%

Dominan sentiment positif menunjukkan bahwa mayoritas pengguna Tiktok sebagai platform yang menyenangkan dan inovatif. Banyak ulasan positif menyebutkan kesenangan dalam menonton video lucu, kemudahan pengguna aplikasi, serta kreativitas konten. Hal ini menyatakan bahwa platform dengan konten berbasis hiburan cenderung memperoleh respons positif dari pengguna. Namun, ulasan negatif umumnya mencerminkan kekhawatiran pengguna terkait privasi data, ketergantungan, serta konten yang tidak sesuai.

Pembagian Data : Training dan Testing

Untuk membangun dan mengevaluasi model, dataset dibagi menjadi data training dan testing dengan perbandingan 80:20:

- 80% data (80.000 data) digunakan untuk training model, yaitu proses belajar model dari data yang sudah dilabeli
- 20% data (20.000 data) digunakan untuk testing, yaitu menguji performa model terhadap data yang belum pernah dilihat sebelumnya.

b) Evaluasi Kinerja Model

Pengguna algoritma Multinomial Naïve Bayes menunjukan hasil Evaluasi sebagai berikut:

- Akurasi: 88.2%
- Precision: 89.1%
- Recall: 87.4%
- F1-Score: 88.2%

Nilai- nilai ini menunjukkan performa model yang sangat baik dalam mengklasifikasikan data ulasan pengguna. Terutama pada data yang telah

diolah dengan TF-IDF, model mampu mengenali pola-pola kata yang mempresentasikan sentiment dengan cukup akurat.

Beberapa kesalahan klasifikasi ditemukan pada kelas negatif maupun positif. Hal ini lazim terjadi pada teks pendek yang mengandung ambiguitas linguistic.

Namun secara keseluruhan, performa model masih sangat kompetitif dan cukup optimal untuk klasifikasi teks berbahasa Indonesia, terutama mengingat volume data yang besar dan kompleksitas ragam Bahasa yang digunakan pengguna media sosial. Dengan pelatihan dan validasi yang cukup, algoritma ini dapat dijadikan acuan awal dalam sistem pengambilan Keputusan berbasis opini public, baik untuk keperluan riset, evaluasi produk, maupun pengawasan konten digital.

c) Analisis Mendalam terhadap Respons Pengguna

Hasil analisis terhadap 100.000 ulasan pengguna TikTok menunjukkan bahwa sentimen yang terbentuk tidak hanya terbatas pada emosi dasar positif atau negative, tetapi juga mencerminkan persepsi public terhadap aspek-aspek fungsional, sosial, dan emosional dari penggunaan aplikasi. Berdasarkan pengelompokan tematik dari data ulasan, terdapat tiga kategori utama persepsi pengguna yang menonjol, yaitu:

- Aspek hiburan dan kreativitas: Konten Video pendek yang menghibur menjadi faktor utama daya tarik Tiktok, terutama dikalangan usia muda.
- Kekhawatiran terhadap konten dan dampak sosial: Beberapa pengguna menyampaikan kritik terhadap konten yang dianggap tidak layak, misalnya vulgaritas, challenge berbahaya, atau ujaran kebencian.
- Kepuasan dan kekhawatiran teknis: Komentar seperti lag, update bermasalah, atau login error banyak ditemukan, namun umumnya tidak mengarah ke

penilaian sangat negatif. Ini dapat menjadi masukan penting bagi pengembang aplikasi.

- d) Hasil Klasifikasi Naive Bayes
Setelah data dikonversi ke format numerik menggunakan metode TF-IDF, algoritma Multinomial Naive Bayes digunakan melakukan klasifikasi.

Hasil evaluasi dari model menunjukkan:

Tabel 4. Hasil Klasifikasi

Metrik	Nilai
Akurasi	88.2%
Precision	89.1%
Recall	87.4%
F1_Score	88.2%

- e) Berikut adalah ringkasan confusion matrix dari hasil klasifikasi:

Tabel 5. Hasil Klasifikasi

	Pred: Positif	Pred: Negatif	Pred: Netral
Aktual Positif	12.345	1.233	765
Aktual Negatif	1.102	9.845	756
Aktual Netral	532	489	8.163

Sebagian besar sentimen berhasil diklasifikasikan dengan benar, meskipun terdapat kebingungan antara kelas netral dan negatif karena kata-kata bernuansa campuran.

- f) Efektivitas Naive Bayes dalam Klasifikasi Teks
Naive Bayes terbukti unggul dalam menangani data teks besar berbahasa Indonesia, Meskipun algoritma ini mengasumsikan independensi antar fitur,

hasil tetap kompetitif dibanding metode klasifikasi lainnya. Hal ini menunjukkan bahwa dengan preprocessing yang baik dan representasi fitur yang tepat, model sederhana seperti Naive Bayes dapat memberikan hasil optimal untuk klasifikasi opini Masyarakat. Dalam komentar aplikasi mobile berbahasa Indonesia menunjukkan bahwa Naive Bayes unggul dalam hal kecepatan pelatihan, konsumsi memori rendah, dan hasil klasifikasi yang stabil. Efektivitas ini menjadikan Naive Bayes sebagai Solusi praktis untuk analisis opini dalam skala besar, terutama dalam konteks penggunaan Bahasa Indonesia yang relatif belum sebanyak Bahasa Inggris dalam riset NLP.

4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan mengenai klasifikasi sentimen masyarakat terhadap aplikasi TikTok menggunakan algoritma Naive Bayes, dapat disimpulkan bahwa algoritma ini mampu mengklasifikasikan data teks dengan tingkat akurasi yang cukup baik, Proses analisis sentimen masyarakat dapat diidentifikasi sebagai positif, netral, atau negatif dengan tingkat keakuratan yang dapat diterima.

Penggunaan Naive Bayes sebagai metode klasifikasi terbukti efektif karena kesederhanaannya, kecepatan dalam proses komputasi, serta performa yang cukup andal pada data teks yang bersifat linier. Selain itu, hasil klasifikasi ini juga dapat menjadi acuan bagi pengembang atau pihak terkait untuk memahami persepsi publik terhadap aplikasi TikTok dan juga meningkatkan kualitas layanan berdasarkan masukan dari pengguna.

Daftar Rujukan

- [1] A. Bahri, C. M. Damayanti, Y. Helena Sirait, and F. Alfariy, "Aplikasi Tiktok Sebagai Media Pembelajaran Bahasa Inggris Di Indonesia," *J. Indones. Sos. Sains*, vol. 3, no. 1, pp. 120–130, Jan. 2022, doi: 10.36418/jiss.v3i1.525.
- [2] Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, "Analisis sentimen aplikasi tiktok menggunakan

- algoritma naïve bayes dan support vector machine,” *TEKNOSAINS J. Sains, Teknol. dan Inform.*, vol. 10, no. 2, pp. 176–184, Jul. 2023, doi: 10.37373/tekno.v10i2.419.
- [3] L. Oktasari, Y. H. Chrisnanto, and R. Yuniarti, “TEXT MINING DALAM ANALISIS SENTIMEN ASURANSI MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER,” *Pros. Sains Nas. dan Teknol.*, vol. 1, no. 1, Sep. 2016, doi: 10.36499/psnst.v1i1.1506.
- [4] A. Turmudi Zy, A. Nugroho, A. Rivaldi, and I. Afriantoro, “Analisis Sentimen Terhadap Pembobolan Data pada Twitter dengan Algoritma Naive Bayes,” *J. Teknol. Inform. dan Komput.*, vol. 8, no. 2, pp. 202–213, Sep. 2022, doi: 10.37012/jtik.v8i2.1240.
- [5] Muflih Ihza Rifatama, Mohammad Reza Faisal, Rudy Herteno, Irwan Budiman, and Muhammad Itqan Mazdadi, “OPTIMASI ALGORITMA K-NEAREST NEIGHBOR DENGAN SELEKSI FITUR MENGGUNAKAN XGBOOST,” *J. Inform. dan Rekayasa Elektron.*, vol. 6, no. 1, pp. 64–72, Apr. 2023, doi: 10.36595/jire.v6i1.723.
- [6] A. Purnamawati, M. N. Winarto, and M. Mailasari, “Analisis Sentimen Aplikasi TikTok menggunakan Metode BM25 dan Improved K-NN Fitur Chi-Square,” *J. Komtika (Komputasi dan Inform.)*, vol. 7, no. 1, pp. 97–105, May 2023, doi: 10.31603/komtika.v7i1.8938.
- [7] I. R. Ainunnisa and S. Sulastri, “Analisis Sentimen Aplikasi Tiktok dengan Metode Support Vector Machine (SVM), Logistic Regression dan Naïve Bayes,” *J. Teknol. Sist. Inf. dan Apl.*, vol. 6, no. 3, pp. 423–430, Jul. 2023, doi: 10.32493/jtsi.v6i3.31076.
- [8] M. G. Pradana, A. C. Nurcahyo, and P. H. Saputro, “PENGARUH SENTIMEN DI SOSIAL MEDIA DENGAN HARGA SAHAM PERUSAHAAN,” *Eduatic - Sci. J. Informatics Educ.*, vol. 6, no. 2, May 2020, doi: 10.21107/edutic.v6i2.6992.
- [9] N. F. Hilmi and F. Irwiensyah, “Analisis Sentimen Terhadap Aplikasi Tiktok Dari Ulasan Pada Google Playstore Menggunakan Metode Naïve Bayes,” *SMATIKA J.*, vol. 14, no. 01, pp. 146–156, Jul. 2024, doi: 10.32664/smatika.v14i01.1210.
- [10] Runimeirati, Abdul Muis, and Figur Muhammad, “Pelatihan Text Mining Menggunakan Bahasa Pemrograman Python,” *Abdimas Langkanae*, vol. 3, no. 1, pp. 36–46, Jan. 2023, doi: 10.53769/abdimas.3.1.2023.83.
- [11] A. Hermawan, I. Jowensen, J. Junaedi, and Edy, “Implementasi Text-Mining untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine,” *JST (Jurnal Sains dan Teknol.)*, vol. 12, no. 1, pp. 129–137, Apr. 2023, doi: 10.23887/jstundiksha.v12i1.52358.
- [12] A. R. Harungguan, H. Napitupulu, and F. Firdaniza, “Analisis Sentimen dengan Metode Klasifikasi Naïve Bayes dan Seleksi Fitur Chi-Square,” *In Search*, vol. 22, no. 1, pp. 115–122, Jul. 2023, doi: 10.37278/insearch.v22i1.677.
- [13] P. H. Nadia, Siti Sri Wahyuni, and Anggraini Puspita Sari, “Penerapan Naive Bayes Classifier untuk Analisis sentimen Ulasan Pelanggan pada Frenz Accessories Handphone,” *J. Mhs. Tek. Inform.*, vol. 3, no. 2, pp. 19–29, Oct. 2024, doi: 10.35473/jamastika.v3i2.3245.
- [14] M. A. Rosid, A. S. Fitriani, I. R. I. Astutik, N. I. Mulloh, and H. A. Gozali, “Improving Text Preprocessing For Student Complaint Document Classification Using Sastrawi,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 874, no. 1, p. 012017, Jun. 2020, doi: 10.1088/1757-899X/874/1/012017.
- [15] B. Hakim, “Analisa Sentimen Data Text Preprocessing Pada Data Mining Dengan Menggunakan Machine Learning,” *JBASE - J. Bus. Audit Inf. Syst.*, vol. 4, no. 2, Aug. 2021, doi: 10.30813/jbase.v4i2.3000.