

## **ANALISIS SENTIMEN MASYARAKAT INDONESIA TERHADAP KASUS PERSETERUAN YAI MIM DAN SAHARA DI TIKTOK MENGGUNAKAN ALGORITMA NAIVE BAYES**

Darryl Yanuar Ar-rafi<sup>1</sup>, Asep Arwan Sulaeman<sup>2</sup>, Handala Simetris Harahap<sup>3</sup>

<sup>1,2,3</sup> Fakultas Teknik, Universitas Pelita Bangsa

[darryl.312210065@mhs.pelitabangsa.ac.id](mailto:darryl.312210065@mhs.pelitabangsa.ac.id), [aseparwan@pelitabangsa.ac.id](mailto:aseparwan@pelitabangsa.ac.id),

[handala.harahap@pelitabangsa.aci.d](mailto:handala.harahap@pelitabangsa.aci.d)

### **ABSTRACT**

*This study aims to analyze public sentiment toward the conflict case between Yai Mim and Sahara, which went viral on the TikTok platform. The data used in this study were TikTok user comments collected using Apify Instant Data Scraper, with a total of 10,504 comments. The research stages included data preprocessing (cleaning, normalization, tokenization, stopword removal, and stemming), sentiment labeling using a lexicon-based approach, feature weighting using the Term Frequency–Inverse Document Frequency (TF-IDF) method, and classification using the Naïve Bayes algorithm. As a comparison model, this study also implemented the Neural Network algorithm to compare classification performance. Model testing was conducted using four data split scenarios: 90:10, 80:20, 70:30, and 60:40 for training and testing data. The results showed that the Naïve Bayes model achieved the highest accuracy of 94.85% in the 90:10 scenario. Meanwhile, the Neural Network model demonstrated better performance with the highest accuracy of 96.49% in the 80:20 scenario. Based on these results, the 80:20 scenario was selected as the main reference because it provides a better balance in model evaluation. Overall, the combination of TF-IDF, Naïve Bayes, and Neural Network methods proved effective in classifying Indonesian sentiment comments on TikTok social media, with Neural Network showing more optimal performance compared to Naïve Bayes.*

*Keywords: Sentiment Analysis, TikTok, Naïve Bayes, Neural Network, TF-IDF, Text Mining*

### **ABSTRAK**

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap kasus perseteruan Yai Mim dan Sahara yang viral di platform TikTok. Data yang digunakan berupa komentar pengguna TikTok yang dikumpulkan menggunakan *Apify Instant Data Scraper* dengan total sebanyak 10.504 data. Tahapan penelitian meliputi preprocessing data (cleaning, normalisasi, tokenisasi, *stopword removal*, dan *stemming*), pelabelan sentimen menggunakan pendekatan lexicon-based, pembobotan fitur menggunakan metode *Term Frequency–Inverse Document Frequency (TF-IDF)*, serta klasifikasi menggunakan algoritma *Naïve Bayes*. Sebagai model pembandingan, penelitian ini juga menerapkan algoritma *Neural Network* untuk membandingkan performa klasifikasi. Pengujian model dilakukan dengan empat skenario pembagian data, yaitu 90:10, 80:20, 70:30, dan 60:40 antara data latih dan data uji. Hasil penelitian menunjukkan bahwa model *Naïve Bayes* menghasilkan akurasi tertinggi sebesar 94,85% pada skenario 90:10.

Sementara itu, model *Neural Network* menunjukkan performa yang lebih baik dengan akurasi tertinggi sebesar 96,49% pada skenario 80:20. Berdasarkan hasil tersebut, skenario 80:20 dipilih sebagai acuan utama karena memberikan keseimbangan yang lebih baik dalam evaluasi model. Secara keseluruhan, kombinasi metode TF-IDF, *Naïve Bayes*, dan *Neural Network* terbukti efektif dalam mengklasifikasikan sentimen komentar berbahasa Indonesia pada media sosial TikTok, dengan *Neural Network* menunjukkan performa yang lebih optimal dibandingkan *Naïve Bayes*.

**Kata Kunci:** Analisis Sentimen, TikTok, *Naïve Bayes*, *Neural Network*, TF-IDF, *Text Mining*

### **A. Pendahuluan**

Media sosial memiliki peran penting dalam pembentukan opini publik di era digital. Salah satu *platform* yang paling berpengaruh adalah TikTok, aplikasi berbagi video pendek yang dikembangkan oleh ByteDance. Aplikasi ini memungkinkan pengguna membuat, mengedit, dan membagikan video berdurasi singkat yang dilengkapi efek musik serta filter visual menarik [1]. Di Indonesia, TikTok telah menjadi bagian dari gaya hidup digital masyarakat karena kemampuannya menyebarkan informasi dan hiburan secara cepat dan luas.

Menurut laporan KompasTekno, Indonesia menempati posisi pertama sebagai negara dengan jumlah pengguna TikTok terbanyak di dunia, yaitu mencapai 157 juta pengguna aktif [1]. Data dari GoodStats juga menunjukkan bahwa masyarakat Indonesia menghabiskan rata-rata 45

jam per bulan menggunakan TikTok, jauh di atas rata-rata global [2]. Angka tersebut menunjukkan bahwa TikTok bukan hanya media hiburan, melainkan juga ruang sosial tempat masyarakat mengekspresikan pendapat dan membentuk persepsi terhadap isu-isu sosial yang sedang berkembang di masyarakat digital.

Fenomena viral di TikTok tidak hanya mencakup konten hiburan, tetapi juga isu sosial yang memicu berbagai opini publik. Salah satu yang cukup mencuri perhatian adalah kasus perseteruan antara Yai Mim dan Sahara, dua figur publik yang saling menanggapi melalui konten TikTok hingga memicu gelombang komentar masyarakat. Reaksi publik terhadap kasus ini beragam, mulai dari dukungan, kritik, hingga ujaran kebencian. Fenomena ini memperlihatkan bagaimana media sosial dapat memperkuat arus opini

dan menciptakan polarisasi di ruang digital [3].

Penelitian terdahulu menunjukkan bahwa analisis sentimen merupakan pendekatan yang efektif untuk memahami opini masyarakat terhadap isu sosial di media sosial. Apriliani et al. [4] menemukan bahwa algoritma *Naïve Bayes* mampu mencapai akurasi sekitar 80% dalam analisis sentimen ulasan TikTok. Arvyantomo dan Ratama [3] juga memperoleh hasil serupa ketika menerapkan algoritma yang sama pada data komentar *Facebook*, menegaskan efektivitasnya dalam mengklasifikasikan sentimen teks informal. Selanjutnya, Rahmadani et al. [5] menjelaskan bahwa kombinasi metode *Term Frequency–Inverse Document Frequency* (TF-IDF) dan *Naïve Bayes* dapat meningkatkan akurasi analisis terhadap komentar pengguna media sosial.

Namun, sebagian besar penelitian terdahulu berfokus pada opini umum atau ulasan produk, bukan pada isu sosial viral di TikTok seperti kasus Yai Mim dan Sahara. Selain itu, sebagian penelitian belum mengoptimalkan tahap *preprocessing teks* untuk bahasa Indonesia yang kaya akan slang, emotikon, dan negasi, serta

jarang membahas perbedaan karakteristik antarplatform seperti *Twitter* dan TikTok. Dengan demikian, dibutuhkan penelitian yang berfokus pada analisis sentimen masyarakat Indonesia terhadap kasus viral di TikTok dengan menerapkan kombinasi TF-IDF dan *Naïve Bayes* sebagai metode utama [4], [5], [6], [7], [8].

Selain itu, penelitian ini juga mengupayakan peningkatan efisiensi pengolahan data dengan mengintegrasikan penggunaan *Apify Instant Data Scraper* untuk pengumpulan komentar dan *Google Colab* sebagai lingkungan komputasi berbasis *cloud* untuk pelatihan model [7]. Melalui pendekatan tersebut, penelitian ini diharapkan dapat memberikan hasil analisis yang lebih cepat, akurat, serta berkontribusi terhadap pengembangan riset analisis sentimen berbahasa Indonesia dalam memahami opini publik di platform TikTok.

## **B. Metode Penelitian**

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen komputasional, di mana data berupa komentar pengguna TikTok dianalisis secara otomatis

menggunakan algoritma *machine learning*. Pendekatan kuantitatif dipilih karena fokus penelitian ini adalah mengukur performa model klasifikasi sentimen berdasarkan nilai numerik seperti akurasi, presisi, *recall*, dan *F1-score* [5].

Metode eksperimen komputasional memungkinkan proses analisis dilakukan secara terukur dan dapat direplikasi. Tahapan utama meliputi pengumpulan data komentar TikTok menggunakan *Apify Instant Data Scraper* [7], pra-pemrosesan teks (meliputi pembersihan data, tokenisasi, dan normalisasi), pembobotan fitur dengan *Term Frequency–Inverse Document Frequency* (TF-IDF) [17], serta pelatihan model klasifikasi menggunakan algoritma *Naïve Bayes* [16].

Pendekatan ini dipilih karena efisien dalam menganalisis data teks berskala besar, serta telah terbukti efektif dalam penelitian-penelitian terdahulu yang menggunakan kombinasi TF-IDF dan *Naïve Bayes* untuk analisis sentimen berbahasa Indonesia [5].

Data yang digunakan dalam penelitian ini berasal dari komentar pengguna TikTok yang bersifat publik

pada beberapa video yang membahas kasus perseteruan antara Yai Mim dan Sahara, yang sempat viral pada tahun 2025. Pemilihan data dilakukan pada periode ketika kasus tersebut berada pada puncak perhatian publik, sehingga komentar yang dikumpulkan mencerminkan respons dan opini masyarakat secara aktual.

Proses pengumpulan data dilakukan menggunakan *Apify Instant Data Scraper*, yaitu *platform* berbasis *cloud* yang mampu mengekstraksi komentar secara otomatis dari halaman TikTok [7]. Melalui alat ini, peneliti berhasil mengumpulkan lebih dari 10.000 komentar yang terdiri dari teks komentar, nama pengguna, dan waktu unggahan.

Seluruh data disimpan dalam format *CSV* (*Comma-Separated Values*) untuk memudahkan proses pengolahan pada tahap selanjutnya menggunakan *Google Colab*. Data yang dikumpulkan hanya mencakup informasi publik dan tidak melibatkan identitas pribadi pengguna, sehingga penelitian ini tetap mematuhi prinsip etika penelitian dan kebijakan privasi *platform* TikTok.

### **C. Hasil Penelitian dan Pembahasan**

Pada penelitian ini dilakukan pengumpulan data berupa komentar pengguna TikTok terkait kasus perseteruan antara Yai Mim dan Sahara yang viral pada tahun 2025. Proses pengambilan data dilakukan menggunakan teknik *scraping* dengan bantuan *Apify Instant Data Scraper* melalui *Google Chrome*. Data yang diperoleh disimpan dalam format CSV untuk memudahkan proses analisis.

Pemilihan metode *scraping* dilakukan karena TikTok tidak menyediakan akses API publik secara bebas untuk mengambil data komentar dalam jumlah besar. Selain itu, penggunaan API resmi umumnya memerlukan autentikasi, registrasi aplikasi, serta memiliki keterbatasan akses data, sehingga kurang fleksibel untuk kebutuhan penelitian ini. Oleh karena itu, *scraping* dipilih karena lebih praktis, efisien, dan mampu mengumpulkan data komentar publik secara otomatis.

Berdasarkan hasil *scraping*, diperoleh sebanyak 10.504 komentar dengan tiga atribut utama, yaitu *text*, *createTimeISO*, dan *uniqueId*. Atribut *text* berisi isi komentar, *createTimeISO* menunjukkan waktu

unggahan, dan *uniqueId* sebagai identitas pengguna.

Tahap pra-pemrosesan data telah dijelaskan pada Bab III. Pada bagian ini ditampilkan hasil dari proses tersebut sebagai bentuk implementasi terhadap data penelitian. Proses *preprocessing* menghasilkan data teks yang telah dibersihkan dari karakter yang tidak diperlukan, dinormalisasi ke bentuk baku, serta diubah menjadi kata dasar melalui tahapan *cleaning*, *case folding*, normalisasi, *tokenization*, *stopword removal*, dan *stemming*.

Data yang telah melalui tahap *preprocessing* selanjutnya dilakukan proses pelabelan sentimen. Pada penelitian ini, pelabelan dilakukan menggunakan metode *lexicon-based*, yaitu dengan memanfaatkan kamus sentimen untuk mengidentifikasi polaritas suatu teks.

Kamus sentimen yang digunakan berasal dari dataset *Kaggle* yang terdiri dari kamus kata positif, kamus kata negatif, serta daftar kata netral. Setiap kata dalam teks akan dibandingkan dengan kamus tersebut untuk menentukan skor sentimen. Jika suatu kata termasuk dalam kamus positif maka skor akan bertambah (+1), sedangkan jika termasuk dalam

kamus negatif maka skor akan berkurang (-1).

Proses pelabelan dilakukan menggunakan bahasa pemrograman Python pada *Google Colab*. Setiap teks yang telah diproses akan dihitung skor sentimennya berdasarkan jumlah kemunculan kata positif dan negatif. Jika skor lebih dari 0 maka dikategorikan sebagai sentimen positif, jika kurang dari 0 sebagai sentimen negatif, dan jika sama dengan 0 maka dikategorikan sebagai netral.

Hasil pelabelan sangat bergantung pada kelengkapan kamus sentimen yang digunakan, sehingga terdapat kemungkinan beberapa teks tidak terklasifikasi secara tepat karena keterbatasan kosakata dalam kamus.

Kelas Sentimen	Kamus Kaggle
Netral	5.654
Negatif	2.487
Positif	2.363
Jumlah	10.504

Selanjutnya dilakukan penghilangan data dengan label netral. Hal ini dilakukan karena penelitian ini berfokus pada analisis sentimen positif dan negatif, sehingga data netral tidak digunakan dalam proses klasifikasi.

Kelas Sentimen	Kamus Kaggle
Negatif	2.487
Positif	2.363
Jumlah	4.850

Maka hasil akhir pelabelan yang didapatkan adalah sebesar 4.850 dari 10.504 ulasan dengan hasil kelas sentimen negatif sebesar 2.487 dan kelas positif sebesar 2.363.

Setelah melalui tahap preprocessing dan pelabelan, data kemudian diubah ke dalam bentuk numerik menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF). Pembobotan ini bertujuan untuk mengukur tingkat kepentingan suatu kata dalam keseluruhan dokumen. Proses pembobotan dilakukan menggunakan pustaka *scikit-learn* pada *Google Colab* dengan fungsi *TfidfVectorizer*. Hasil dari proses ini berupa representasi numerik dalam bentuk matriks fitur yang menunjukkan bobot masing-masing kata terhadap seluruh dokumen.

Selain itu, dilakukan analisis terhadap kata-kata dengan bobot tertinggi untuk mengetahui kata yang paling berpengaruh dalam dataset. Contoh hasil pembobotan TF-IDF ditunjukkan pada Tabel.

No	Jumlah Teks	Teks	Bobot
1	4138	Tidak	0.045912
2	2946	Orang	0.031671
3	4501	Yai	0.029835
4	3989	Tanah	0.024731
5	3493	Sahara	0.024230
6	3058	Parkir	0.022567
7	3394	Rental	0.020979
8	3617	Sehat	0.020398
9	1568	Jalan	0.019566
10	2476	Mim	0.019144

terlihat bahwa kata-kata seperti “tidak”, “orang”, “yai”, dan “sahara” memiliki bobot yang relatif tinggi. Hal ini menunjukkan bahwa kata-kata tersebut sering muncul dan memiliki peran penting dalam merepresentasikan isi komentar pada dataset.

### Implementasi *Naïve Bayes*

Setelah data melalui tahap preprocessing, pelabelan, dan pembobotan menggunakan TF-IDF, langkah selanjutnya adalah proses klasifikasi menggunakan algoritma *Naïve Bayes*. Pada penelitian ini digunakan metode *Multinomial Naïve Bayes* yang diimplementasikan melalui pustaka *scikit-learn* pada Google Colab.

Model dilatih menggunakan data latih (*training data*) pada masing-masing skenario percobaan,

kemudian diuji menggunakan data uji (*testing data*) untuk mengukur performa klasifikasi. Pengujian dilakukan sebanyak empat kali dengan variasi pembagian data, yaitu 90:10, 80:20, 70:30, dan 60:40. Salah satu hasil skenario dari penerapan dengan model *Naïve Bayes* dapat dilihat pada Gambar dibawah.

```

Percobaan 90:10
-----
Akurasi: 0.9485

Classification Report:
      precision    recall  f1-score   support

negatif    0.95    0.95    0.95     249
positif    0.95    0.95    0.95     236

accuracy          0.95          0.95          0.95     485
macro avg    0.95    0.95    0.95     485
weighted avg  0.95    0.95    0.95     485
    
```

### Implementasi *Neural Network*

Sebagai model pembanding, penelitian ini juga menerapkan algoritma *Neural Network* untuk mengetahui performa klasifikasi yang lebih optimal. Model diimplementasikan menggunakan metode *Multi-Layer Perceptron (MLPClassifier)* dari pustaka *scikit-learn* pada Google Colab.

Data yang digunakan merupakan hasil preprocessing dan pembobotan TF-IDF yang sama dengan model *Naïve Bayes*, sehingga perbandingan kedua model dilakukan secara objektif. Pengujian *Neural Network* juga dilakukan sebanyak empat kali menggunakan pembagian data 90:10,

80:20, 70:30, dan 60:40. Salah satu hasil skenario dari penerapan dengan model *Neural Network* dapat dilihat pada Gambar dibawah.

```

=====
Percobaan 80:20
=====
Akurasi: 0.9649

Classification Report:
      precision    recall  f1-score   support

   negatif      0.96      0.97      0.97      497
   positif      0.97      0.96      0.96      473

 accuracy          0.96          0.96          0.96          970
 macro avg         0.97          0.96          0.96          970
 weighted avg     0.97          0.96          0.96          970
    
```

Evaluasi model dilakukan untuk mengetahui kinerja algoritma *Naïve Bayes* dan *Neural Network* dalam mengklasifikasikan sentimen komentar pengguna TikTok. Proses evaluasi menggunakan *confusion matrix* serta metrik akurasi, presisi, *recall*, dan *F1-score*. Pengujian dilakukan sebanyak empat kali percobaan dengan variasi pembagian data latih dan data uji, yaitu 90:10, 80:20, 70:30, dan 60:40.

Hasil evaluasi model *Naïve Bayes* disajikan pada Tabel sedangkan hasil evaluasi model *Neural Network* disajikan pada Tabel.

Ket.	Percobaan 1	Percobaan 2	Percobaan 3	Percobaan 4
Data Training	90%	80%	70%	60%
Data Testing	10%	20%	30%	40%
True Positive	224	445	664	875
True Negative	236	464	696	918
False Positive	13	33	50	77
False Negative	12	28	45	70
Precision	95%	94%	93%	92%
Recall	95%	94%	93%	92%
F1-Score	95%	94%	93%	92%
Accuracy	94,85%	93,71%	93,47%	92,42%

Tabel 4.7 Hasil Evaluasi Model *Neural Network*.

Ket.	Percobaan 1	Percobaan 2	Percobaan 3	Percobaan 4
Data Training	90%	80%	70%	60%
Data Testing	10%	20%	30%	40%
True Positive	225	452	671	899
True Negative	242	484	718	938
False Positive	7	13	28	56
False Negative	11	21	38	46
Precision	96%	97%	95%	95%
Recall	96%	96%	95%	95%
F1-Score	96%	96%	95%	95%

Berdasarkan Tabel 4.7, dapat diketahui bahwa model *Naïve Bayes* mampu memberikan performa yang sangat baik pada seluruh skenario pengujian. Nilai akurasi tertinggi diperoleh pada percobaan dengan pembagian data 90:10 sebesar 94,85%, sedangkan akurasi terendah terdapat pada pembagian 60:40 sebesar 92,42%. Perbedaan nilai akurasi antar percobaan relatif kecil, sehingga menunjukkan bahwa model memiliki performa yang stabil.

Pada model *Neural Network*, hasil pengujian menunjukkan performa yang lebih tinggi dibandingkan *Naïve Bayes* pada seluruh skenario. Akurasi tertinggi diperoleh pada pembagian data 80:20 sebesar 96,49%, sedangkan akurasi terendah diperoleh pada pembagian 60:40 sebesar 94,74%.

Jika dilihat dari *confusion matrix* kedua model, jumlah prediksi benar (*true positive* dan *true negative*) lebih

dominan dibandingkan kesalahan prediksi (*false positive* dan *false negative*) pada setiap skenario. Hal ini menunjukkan bahwa kedua model memiliki kemampuan yang baik dalam membedakan sentimen positif dan negatif.

Selain itu, nilai *precision*, *recall*, dan *F1-score* pada masing-masing percobaan juga menunjukkan hasil yang seimbang. Kondisi ini mengindikasikan bahwa model tidak hanya memiliki tingkat akurasi tinggi, tetapi juga konsisten dalam melakukan klasifikasi tanpa bias signifikan terhadap salah satu kelas sentimen.

Berdasarkan keseluruhan hasil, skenario 80:20 dipilih sebagai acuan utama penelitian karena memberikan keseimbangan yang baik antara data latih dan data uji, serta menghasilkan akurasi tertinggi pada model *Neural Network* sebesar 96,49%.

Secara keseluruhan, hasil pengujian menunjukkan bahwa kombinasi pembobotan TF-IDF dengan algoritma klasifikasi mampu menghasilkan performa yang baik dalam analisis sentimen komentar TikTok.

Berdasarkan hasil pengujian, algoritma *Naïve Bayes* dan *Neural*

*Network* sama-sama mampu menghasilkan performa klasifikasi yang baik dengan akurasi di atas 92% pada seluruh skenario pengujian.

Model *Naïve Bayes* menunjukkan performa stabil dengan akurasi tertinggi sebesar 94,85% pada pembagian data 90:10. Hal ini menunjukkan bahwa semakin besar data latih, maka model semakin optimal dalam mempelajari pola data.

Namun, model *Neural Network* memberikan hasil yang lebih unggul dibandingkan *Naïve Bayes*. Akurasi tertinggi diperoleh pada pembagian data 80:20 sebesar 96,49%, yang menunjukkan bahwa *Neural Network* lebih mampu mengenali pola kompleks pada data teks hasil pembobotan TF-IDF.

Penurunan akurasi pada skenario 70:30 dan 60:40 menunjukkan bahwa berkurangnya jumlah data latih berdampak pada performa model, meskipun penurunannya tidak terlalu signifikan. Hal ini menunjukkan kedua model tetap stabil pada berbagai skenario pembagian data.

Keberhasilan model dipengaruhi oleh tahapan *preprocessing*, pelabelan menggunakan kamus sentimen, serta pembobotan TF-IDF yang membantu meningkatkan

kualitas representasi fitur teks. Namun, pendekatan *lexicon-based* masih memiliki keterbatasan dalam menangani konteks bahasa seperti sarkasme, ironi, atau bahasa tidak baku.

Secara keseluruhan, penelitian ini menunjukkan bahwa algoritma *Neural Network* lebih optimal dalam analisis sentimen komentar TikTok, sedangkan *Naïve Bayes* tetap efektif sebagai model baseline karena lebih sederhana dan efisien secara komputasi.

#### **D. Kesimpulan**

Berdasarkan hasil penelitian analisis sentimen komentar pengguna TikTok terhadap kasus perseteruan Yai Mim dan Sahara menggunakan metode TF-IDF, *Naïve Bayes*, dan *Neural Network*, maka dapat diperoleh kesimpulan sebagai berikut:

1. Penelitian berhasil mengumpulkan sebanyak 10.504 komentar pengguna TikTok melalui teknik *scraping* menggunakan *Apify Instant Data Scraper*. Data kemudian melalui tahapan *preprocessing* meliputi *cleaning*, *case folding*, normalisasi, *tokenization*, *stopword removal*, *stemming*, dan pelabelan sentimen.
2. Hasil pengujian menggunakan algoritma *Naïve Bayes* menunjukkan performa yang baik dan stabil pada seluruh skenario pembagian data. Akurasi tertinggi diperoleh pada pembagian data 90:10 dengan nilai 94,85%.
3. Hasil pengujian menggunakan algoritma *Neural Network* menunjukkan performa yang lebih baik dibandingkan *Naïve Bayes* pada seluruh skenario pengujian. Akurasi tertinggi diperoleh pada pembagian data 80:20 dengan nilai 96,49%.
4. Berdasarkan perbandingan kedua model, algoritma *Neural Network* dinilai lebih optimal dalam mengklasifikasikan sentimen komentar TikTok karena mampu menghasilkan akurasi lebih tinggi, sedangkan *Naïve Bayes* tetap efektif sebagai model baseline dengan proses komputasi yang lebih sederhana dan cepat.
5. Kombinasi *preprocessing*, pembobotan TF-IDF, serta

algoritma klasifikasi terbukti efektif dalam menganalisis sentimen komentar pengguna TikTok terkait isu viral di media sosial.

## DAFTAR PUSTAKA

- S. R. Department, "Jumlah Pengguna TikTok di Indonesia Capai 157,6 Juta pada 2024," 2024. [Online]. Available: <https://tekno.kompas.com/read/2024/10/25/15020057/indonesia-pengguna-tiktok-terbesar-di-dunia-tembus-157-juta-kalahkan-as>
- DataReportal, "Rata-Rata Waktu Penggunaan TikTok di Indonesia Mencapai 45 Jam per Bulan," 2025, *Data.AI via GoodStats*. [Online]. Available: <https://data.goodstats.id/statistika/publik-ri-habiskan-hampir-45-jam-sebulan-di-tiktok-kalahkan-rata-rata-global-OxYPF>
- M. Arvyantomo and N. Ratama, "Analisis Sentimen Masyarakat Indonesia Terhadap Invasi Rusia Di Ukraina Menggunakan Metode Naïve Bayes Pada Media Sosial Facebook," *J. Ilmu Komput. dan Pendidik.*, vol. 1, pp. 705–717, 2023, [Online]. Available: <https://journal.mediapublikasi.id/index.php/logic>
- N. Apriliani, N. Suarna, and W. Prihartono, "Analisis Sentimen Review Penggunaan Tiktok Melalui Pendekatan Algoritma Naïve Bayes," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 6, pp. 3725–3731, 2024, doi: 10.36040/jati.v7i6.8299.
- P. S. Rahmadani, F. C. Tampubolon, A. N. Jannah, N. L. H. Hutabarat, and A. M. Simarmata, "Tiktok Social Media Sentiment Analysis Using the Nave Bayes Classifier Algorithm," *Sinkron*, vol. 7, no. 3, pp. 995–999, 2022, doi: 10.33395/sinkron.v7i3.11579.
- A. Kusuma and A. Nugroho, "Analisa Sentimen Pada Twitter Terhadap Kenaikan Tarif Dasar Listrik Dengan Metode Naïve Bayes," *J. Ilm. Teknol. Inf. Asia*, vol. 15, no. 2, pp. 137–146, 2021, doi: 10.32815/jitika.v15i2.557.
- Farendika Rezzi, "Penerapan Algoritma Multinomial Naïve Bayes dengan Penyeimbangan Data SMOTE pada KI Asifikasi Sentimen Pengguna Shopee terhadap Produk Facial Wash Kahf," *Uranus J. Ilm. Tek. Elektro, Sains dan Inform.*, vol. 3, no. 3, pp. 01–12, 2025, doi: 10.61132/uranus.v3i3.1022.
- R. A. Puspala Sari and I. D. Jaya, "Penerapan Algoritma Naïve Bayes Untuk Analisis Sentimen Pengguna My JNE," *J. Multidisiplin West Sci.*, vol. 2, no. 03, pp. 215–221, 2023, doi: 10.58812/jmws.v2i03.269.
- A. Ashari, G. Dwi, F. Harahap, and R. Rosnelly, "ANALISIS SENTIMEN PADA ULASAN PENGGUNA TIKTOK DAN TOKOPEDIA MENGGUNAKAN MESIN LEARNING BERBASIS NAIVE BAYES CLASSIFIER," pp. 132–140, 2025, doi: <https://doi.org/10.46576/syntax.v6i1.6960>.
- J. Alfiah Zulqornain and P. Pandu Adikara, "Analisis Sentimen Tanggapan Masyarakat Aplikasi Tiktok Menggunakan Metode Naïve Bayes dan Categorical Proportional Difference (CPD),"

- J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 7, pp. 2886–2890, 2021, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/9396>
- M. Ziddan, H. F. Adiyatma, and A. P. Sari, "Analisis Sentimen Produk pada Tiktok Shop dengan Metode Naive Bayes," vol. 9, pp. 22098–22106, 2025.
- A. N. Elly Indrayuni, "ANALISIS SENTIMEN APLIKASI TIKTOK SHOP SELLER CENTER MENGGUNAKAN NAIVE BAYES, SVM DAN LOGISTIC REGRESSION," *LPPM Nusa Mandiri*, vol. 18, no. 1, pp. 84–92, 2023, doi: <https://doi.org/10.33480/inti.v20i1.6851>.
- J. C. Setiawan, K. M. Lhaksmana, and B. Bunyamin, "Sentiment Analysis of Indonesian Tik Tok Review Using LSTM and IndoBERTweet Algorithm," *JIPi (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 8, no. 3, pp. 774–780, 2023, doi: [10.29100/jipi.v8i3.3911](https://doi.org/10.29100/jipi.v8i3.3911).
- T. Ramadhani, Y. A. Sari, and E. Santoso, "Analisis Sentimen Masyarakat Indonesia terhadap Covid-19 pada Media Sosial Twitter menggunakan Metode Naive Bayes," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 12, pp. 5680–5686, 2021, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/10325>
- N. Hayati, S. Tri Nowo, B. Suhardi, and R. Rosnelly, "Analisis Sentimen Komentar Youtube Terhadap Isu Kesehatan Mental Menggunakan Algoritma Naive Bayes Dan K-Nearest Neighbor (Knn)," *Syntax J. Softw. Eng. Comput. Sci. Inf. Technol.*, vol. 6, no. 1, pp. 8–16, 2025, doi: [10.46576/syntax.v6i1.6060](https://doi.org/10.46576/syntax.v6i1.6060).
- Y. Yulistiani and S. Styawati, "Analisis Sentimen Terhadap Calon Presiden Indonesia 2024 dengan Metode Extreme Gradient Boosting (XGBOOST)," *J. Inform. J. Pengemb. IT*, vol. 9, no. 3, pp. 322–328, 2024, doi: [10.30591/jpit.v9i3.6127](https://doi.org/10.30591/jpit.v9i3.6127).
- D. Pramudita, Y. Akbar, and T. Wahyudi, "MALCOM: Indonesian Journal of Machine Learning and Computer Science Sentiment Analysis of the Indonesian Smart College Card Program on Social Media X Using the Naive Bayes Algorithm Analisis Sentimen Terhadap Program Kartu Indonesia Pintar Kuliah Pada Med," *Malcom*, vol. 4, no. October, pp. 1420–1430, 2024, doi: <https://doi.org/10.57152/malcom.v4i4.1565>.
- M. R. Maulana and R. A. Putri, "Sentiment Analysis of the 2024 General Election Commission in Indonesia through Twiter using the Support Vector Machine (SVM) Algorithm," *Sistemasi*, vol. 13, no. 3, p. 1226, 2024, doi: [10.32520/stmsi.v13i3.4127](https://doi.org/10.32520/stmsi.v13i3.4127).
- I. Syahroh, S. D. Saputra, R. W. Saputra, V. H. Pranatawijaya, and R. Priskila, "Perbandingan Analisis Sentimen Setelah Pilpres 2024 Di Twitter Menggunakan Algoritma Machine Learning," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, 2024, doi: [10.23960/jitet.v12i2.4249](https://doi.org/10.23960/jitet.v12i2.4249).