

Dana Sanjaya

Perancangan Sistem Informasi Analisis Sentimen Opini Publik Terhadap Tren (Ai) Menggunakan Metode Random Forest Dan...

 Quick Submit Quick Submit Universitas 17 Agustus 1945 Semarang

Document Details

Submission ID

trn:oid:::1:3614972705

Submission Date

Jul 21, 2026, 8:44 PM GMT+7

Download Date

Jul 21, 2026, 8:50 PM GMT+7

File Name

TI_Jurnal_Ilmiyah_Sains_Teknologi_dan_Informasi_Dana_Sanjaya.docx

File Size

92.8 KB

15 Pages

4,318 Words

29,724 Characters

22% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Small Matches (less than 14 words)

Top Sources

- 22%  Internet sources
- 0%  Publications
- 0%  Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Top Sources

- 22% Internet sources
- 0% Publications
- 0% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	journal.irpi.or.id	2%
2	Internet	jurnal.alimspublishing.co.id	1%
3	Internet	jurnal.uns.ac.id	1%
4	Internet	jurnalsyntaxadmiration.com	<1%
5	Internet	learning-gate.com	<1%
6	Internet	openjournal.unpam.ac.id	<1%
7	Internet	urj.uin-malang.ac.id	<1%
8	Internet	estudogeral.sib.uc.pt	<1%
9	Internet	repository.unsri.ac.id	<1%
10	Internet	www.jurnal.polgan.ac.id	<1%
11	Internet	journals.aiac.org.au	<1%

12	Internet	jurnal-itsi.org	<1%
13	Internet	journal.staiyipiqaubau.ac.id	<1%
14	Internet	eresources.thamrin.ac.id	<1%
15	Internet	cdn.juris.id	<1%
16	Internet	ejurnal.methodist.ac.id	<1%
17	Internet	journal.lppmunindra.ac.id	<1%
18	Internet	ceur-ws.org	<1%
19	Internet	www.ijcesen.com	<1%
20	Internet	ojs.inlic.org	<1%
21	Internet	scholar.hit.edu.cn	<1%
22	Internet	ejournal.iainpalopo.ac.id	<1%
23	Internet	ijins.umsida.ac.id	<1%
24	Internet	prosiding.unipma.ac.id	<1%
25	Internet	journal.eng.unila.ac.id	<1%

26	Internet	repository.unhas.ac.id	<1%
27	Internet	journal.uc.ac.id	<1%
28	Internet	repo.undiksha.ac.id	<1%
29	Internet	ejournal.unikama.ac.id	<1%
30	Internet	repo.itera.ac.id	<1%
31	Internet	e-journal.hamzanwadi.ac.id	<1%
32	Internet	journal.unram.ac.id	<1%
33	Internet	pusdikra-publishing.com	<1%
34	Internet	repository.itpln.ac.id	<1%
35	Internet	sistemasi.ftik.unisi.ac.id	<1%
36	Internet	computa.ipbcirebon.ac.id	<1%
37	Internet	ejournal.uinbukittinggi.ac.id	<1%
38	Internet	ejurnal.provisi.ac.id	<1%
39	Internet	repository.upnvj.ac.id	<1%

40	Internet	e-journal.stmik-bnj.ac.id	<1%
41	Internet	mohulilabshor.wordpress.com	<1%
42	Internet	www.zonagadget.co.id	<1%



Perancangan Sistem Informasi Analisis Sentimen Opini Publik Terhadap Tren (Ai) Menggunakan Metode *Random Forest* Dan *Naïve Bayes*

Dana Sanjaya^{1*}, Lela Nurlaela², Tuhfatul Habibah Hasibuan³

¹⁻³Program Studi Sistem Informasi, Institut Teknologi dan Bisnis Swadharma, Indonesia

Email: danasanjayas@gmail.com*, lela@swadharma.ac.id, tuhfatulhabibah@swadharma.ac.id

Abstract. The rapid advancement of Artificial Intelligence (AI) technology has generated diverse public opinions on social media, ranging from positive, negative, to neutral. Sentiment analysis is therefore needed to identify the tendency of public opinion toward this technological development. This study aims to design a Public Opinion Sentiment Analysis Information System on AI Technology Trends using Random Forest and Naïve Bayes algorithms, and to compare their performance to determine the best classification model. This research applies a quantitative approach using the Cross Industry Standard Process for Data Mining (CRISP-DM) method, supported by questionnaire data from 327 respondents using a Likert Scale, along with 4,104 public opinions scraped from social media (X Twitter). The data were processed through cleaning, case folding, normalization, tokenization, stopword removal, stemming, and feature extraction using Term Frequency–Inverse Document Frequency (TF-IDF), then classified using Random Forest and Naïve Bayes. Model performance was evaluated using accuracy, precision, recall, F1-score, Confusion Matrix, and ROC-AUC. Results show that Random Forest achieved higher accuracy than Naïve Bayes, with positive sentiment dominating, followed by negative and neutral sentiment. The findings were implemented into a web-based system using Streamlit to help users automatically analyze public opinion and support decision-making on AI technology adoption..

Keywords: Artificial Intelligence; Information System; Naïve Bayes; Random Forest; Sentiment Analysis

Abstrak. Perkembangan teknologi *Artificial Intelligence* (AI) berlangsung sangat pesat dan menimbulkan beragam opini publik di media sosial, baik positif, negatif, maupun netral. Oleh karena itu, diperlukan analisis sentimen untuk mengetahui kecenderungan opini masyarakat terhadap perkembangan teknologi tersebut. Penelitian ini bertujuan merancang Sistem Informasi Analisis Sentimen Opini Publik terhadap Tren Teknologi *Artificial Intelligence* (AI) menggunakan algoritma *Random Forest* dan *Naïve Bayes*, serta membandingkan performa kedua algoritma untuk menentukan model klasifikasi terbaik. Penelitian ini menggunakan pendekatan kuantitatif dengan metode *Cross Industry Standard Process for Data Mining* (CRISP-DM), didukung data kuesioner dari 327 responden menggunakan Skala *Likert*, serta 4.104 opini publik hasil *scraping* dari media sosial X (*Twitter*). Data diproses melalui tahapan *cleaning*, *case folding*, normalisasi, tokenisasi, *stopword removal*, *stemming*, dan ekstraksi fitur menggunakan *Term Frequency–Inverse Document Frequency* (TF-IDF), kemudian diklasifikasikan menggunakan algoritma *Random Forest* dan *Naïve Bayes*. Performa model dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, *Confusion Matrix*, dan *Receiver Operating Characteristic–Area Under Curve* (ROC-AUC). Hasil penelitian menunjukkan bahwa algoritma *Random Forest* memperoleh akurasi lebih tinggi dibandingkan *Naïve Bayes* sehingga dipilih sebagai model terbaik. Hasil analisis sentimen menunjukkan bahwa sentimen positif mendominasi opini publik, diikuti sentimen negatif dan netral, yang menunjukkan bahwa masyarakat cenderung memberikan tanggapan positif terhadap perkembangan teknologi AI. Hasil tersebut diimplementasikan ke dalam sistem berbasis *web* menggunakan *Streamlit* untuk membantu pengguna menganalisis opini publik secara otomatis serta mendukung pengambilan keputusan terkait pemanfaatan teknologi AI.

Kata kunci: Analisis Sentimen; *Artificial Intelligence*; *Naïve Bayes*; *Random Forest*; Sistem Informasi

1. LATAR BELAKANG

Sulit dipungkiri bahwa perkembangan teknologi *Artificial Intelligence* (AI) telah membawa perubahan besar dalam berbagai bidang, termasuk dunia bisnis dan industri,

Naskah Masuk: 12 Juni 2024; Revisi: 12 Juni 2024; Diterima: 12 Juni 2024; Tersedia: 12 Juni 2024; Terbit: 12 Juni 2024;

PERANCANGAN SISTEM INFORMASI ANALISIS SENTIMEN OPINI PUBLIK TERHADAP TREN AI
MENGUNAKAN RANDOM FOREST DAN NAÏVE BAYES

mulai dari peningkatan efisiensi kerja, pemahaman perilaku pelanggan, hingga pengambilan keputusan berbasis data. Di sisi lain, opini publik terhadap AI juga semakin berkembang melalui media sosial dan platform digital, di mana persepsi positif dapat meningkatkan kepercayaan masyarakat terhadap perusahaan yang mengadopsi AI, sementara opini negatif dapat memengaruhi citra perusahaan terkait isu privasi, keamanan data, dan dampak terhadap tenaga kerja. Kondisi ini sejalan dengan temuan (Vasiljeva et al., 2021) yang menunjukkan bahwa sikap masyarakat terhadap AI bervariasi antarsektor industri dan sangat dipengaruhi oleh tingkat literasi serta pengalaman langsung terhadap teknologi tersebut. (Razaqi et al., 2024) juga menegaskan bahwa optimalisasi bisnis melalui AI perlu diiringi analisis terhadap peluang, tantangan, dan dampaknya di berbagai sektor agar perusahaan dapat merumuskan strategi yang tepat dan sesuai dengan kebutuhan publik.

Penelitian mengenai analisis sentimen dengan membandingkan algoritma klasifikasi telah banyak dilakukan pada berbagai domain. (Ammar & Wiranata, 2025) membandingkan *Naïve Bayes*, *Decision Tree*, dan *Random Forest* pada analisis sentimen pengguna aplikasi Samsat Digital Nasional, sementara (Wati et al., 2025) melakukan perbandingan serupa pada ulasan pengguna aplikasi Deepseek. *Naïve Bayes* juga digunakan oleh (Oktaviani et al., 2025) untuk menganalisis sentimen publik terkait insiden peretasan Pusat Data Nasional, serta oleh (Rudy Seftiawan et al., 2022) pada analisis sentimen aplikasi *e-Government*. *Random Forest* dikenal stabil dan tahan terhadap *overfitting* melalui kombinasi banyak pohon keputusan, sedangkan *Naïve Bayes* merupakan metode probabilistik yang sederhana, cepat, dan efektif untuk klasifikasi teks, sehingga kedua algoritma ini kerap dijadikan pembanding utama dalam studi analisis sentimen berbasis teks.

Meskipun demikian, penelitian-penelitian tersebut umumnya berfokus pada objek yang spesifik dan tunggal, seperti satu aplikasi atau satu peristiwa tertentu, serta hanya mengandalkan satu sumber data (media sosial atau ulasan aplikasi) tanpa validasi dari sumber data lain. Selain itu, sebagian besar penelitian berhenti pada tahap perbandingan performa algoritma tanpa mengimplementasikan hasilnya ke dalam sebuah sistem yang dapat digunakan secara berkelanjutan oleh pengguna nonteknis. Di sisi lain, penerapan *Streamlit* sebagai media penyajian hasil analisis data telah terbukti efektif pada domain lain, seperti prediksi harga saham (Dewi et al., 2025) dan analisis publikasi ilmiah bidang

kesehatan (Parker et al., 2021), namun penerapannya untuk memantau opini publik terhadap tren teknologi AI secara umum masih jarang ditemukan. Kebaruan penelitian ini terletak pada penggabungan dua sumber data yang saling melengkapi, yaitu data kuesioner berskala *Likert* dan data hasil *scraping* media sosial X (*Twitter*), untuk mengukur opini publik terhadap tren teknologi AI secara lebih komprehensif, serta implementasi langsung hasil analisis ke dalam sistem informasi berbasis web menggunakan *Streamlit* agar dapat diakses dan dimanfaatkan secara praktis oleh pengguna dalam pengambilan keputusan.

Opini publik mengenai AI di media sosial seperti *Twitter/X*, *TikTok*, dan *Instagram* terus meningkat setiap harinya, mengandung informasi penting berupa sentimen positif, negatif, maupun netral terhadap perkembangan teknologi tersebut. Namun, analisis data opini secara manual dinilai tidak efektif mengingat volume data yang sangat besar, sehingga pihak yang berkepentingan kesulitan memahami opini publik secara cepat dan terstruktur. Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk merancang sistem informasi analisis sentimen opini publik terhadap tren teknologi AI menggunakan metode *Random Forest* dan *Naïve Bayes*, membandingkan performa kedua algoritma berdasarkan nilai *accuracy*, *precision*, *recall*, dan *F1-score* untuk menentukan metode terbaik, serta mengimplementasikan hasilnya ke dalam sistem berbasis web yang dapat membantu pengguna memahami persepsi masyarakat secara lebih cepat, akurat, dan terstruktur.

2. KAJIAN TEORITIS

Sistem merupakan kesatuan yang terdiri atas berbagai komponen yang saling berhubungan dan bekerja secara terintegrasi untuk mencapai tujuan tertentu (Satria et al., 2021; Syaifunazhirin, 2023). Sementara itu, informasi adalah data yang telah diolah dan diorganisasikan sehingga memiliki nilai tambah dan dapat digunakan sebagai dasar pengambilan keputusan (Rahman et al., 2022; Supriyati et al., 2022). Kombinasi keduanya membentuk sistem informasi, yaitu perpaduan antara teknologi informasi, aktivitas pengguna, dan prosedur organisasi yang berfungsi mendukung operasi, manajemen, serta pengambilan keputusan dalam suatu organisasi (Agustika et al., 2023; H. Arifin et al., 2024). Perancangan sistem informasi sendiri merupakan proses sistematis

PERANCANGAN SISTEM INFORMASI ANALISIS SENTIMEN OPINI PUBLIK TERHADAP TREND AI
MENGUNAKAN RANDOM FOREST DAN NAÏVE BAYES

mulai dari analisis kebutuhan, penggambaran, hingga pengujian untuk mempermudah pengelolaan data secara efektif dan efisien (Agung et al., 2022; M.Ulil Absor, 2024).

Analisis adalah proses menguraikan suatu permasalahan atau data menjadi bagian yang lebih kecil agar mudah dipahami dan dapat digunakan dalam pengambilan keputusan (Nugrahanti & Sussolaikah, 2021; Putra et al., 2025). Ketika diterapkan pada data teks, proses ini dikenal sebagai analisis sentimen, yaitu teknik dalam *Natural Language Processing* (NLP) yang mengklasifikasikan opini atau emosi pengguna—positif, negatif, atau netral—terhadap suatu objek menggunakan pendekatan *machine learning* maupun *deep learning* (Nawulansih et al., 2025; Yanuargi et al., 2025).

Opini publik terhadap *Artificial Intelligence* (AI) terbukti memengaruhi kecepatan adopsi teknologi, reputasi, serta strategi bisnis perusahaan. Dukungan publik yang positif mempercepat integrasi AI ke dalam proses bisnis, sedangkan resistensi publik dapat menghambat inovasi dan menurunkan kepercayaan konsumen (Vasiljeva et al., 2021). Di sisi lain, penerapan AI juga membantu perusahaan mengotomatisasi proses bisnis dan mengolah data dalam jumlah besar secara cepat untuk mendukung pengambilan keputusan yang lebih tepat (Razaqi et al., 2024).

Pengolahan data dalam jumlah besar tersebut dilakukan melalui *data mining*, yaitu proses sistematis untuk mengekstraksi pola dan informasi penting dari data menggunakan teknik statistik, *machine learning*, dan kecerdasan buatan (Gheisari, 2023; Srirahayu et al., 2023). Salah satu algoritma *data mining* yang umum digunakan untuk klasifikasi adalah *Random Forest*, yaitu metode *ensemble learning* yang menggabungkan banyak *decision tree* untuk menghasilkan prediksi yang lebih akurat, stabil, dan tahan terhadap *overfitting* (Fitriani Noer Jamilah, 2026; Khoeruddin et al., 2023). Algoritma lain yang banyak digunakan dalam analisis sentimen adalah *Naïve Bayes*, yaitu metode klasifikasi berbasis probabilitas yang menerapkan Teorema Bayes untuk menentukan kemungkinan suatu data masuk ke kelas tertentu berdasarkan fitur yang dimiliki; algoritma ini dikenal sederhana, cepat, dan efisien dalam mengolah data teks (Oktaviani et al., 2025; Rudy Seftiawan et al., 2022).

Hasil analisis kedua algoritma tersebut diimplementasikan melalui *Streamlit*, yaitu *framework* berbasis Python yang digunakan untuk membangun aplikasi web interaktif secara cepat, khususnya untuk kebutuhan visualisasi data dan pengembangan aplikasi *machine learning* tanpa memerlukan kemampuan *front-end* yang kompleks (Dewi et al., 2025; Parker et al., 2021). Proses pelatihan model dilakukan menggunakan *Google Colab*, yaitu platform komputasi berbasis *cloud* yang menyediakan lingkungan *Jupyter Notebook* untuk menjalankan kode *Python* secara daring tanpa instalasi lokal, serta mendukung penggunaan GPU/TPU untuk mempercepat proses eksperimen data (O. Arifin et al., 2025; Yakut, 2023).

Penelitian Terdahulu. Wati et al. (2025) membandingkan *Random Forest* dan *Naïve Bayes* pada analisis sentimen ulasan pengguna aplikasi Deepseek dan menemukan bahwa *Random Forest* unggul dengan akurasi 96,38%, dibandingkan *Naïve Bayes* sebesar 72,96%, karena *Random Forest* lebih mampu mengenali pola kata dan struktur kalimat yang kompleks. Sejalan dengan itu, Ammar & Wiranata, (2025) membandingkan *Naïve Bayes*, *Decision Tree*, dan *Random Forest* pada analisis sentimen pengguna aplikasi Samsat Digital Nasional (SIGNAL), dan menyimpulkan bahwa *Random Forest* menjadi algoritma paling efektif dengan akurasi 92,9%, diikuti *Naïve Bayes* (89,4%) dan *Decision Tree* (86,8%). Kedua penelitian tersebut menjadi acuan bahwa *Random Forest* secara konsisten menunjukkan performa lebih unggul dibandingkan *Naïve Bayes* dalam berbagai konteks analisis sentimen berbasis teks, sehingga memperkuat dasar pemilihan kedua algoritma ini sebagai objek perbandingan dalam penelitian ini.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan objek berupa opini publik terhadap tren teknologi *Artificial Intelligence* (AI) di media sosial Twitter/X. Disain penelitian mengacu pada metode *Cross Industry Standard Process for Data Mining* (CRISP-DM), yang terdiri atas enam tahap: *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*.

Populasi penelitian adalah opini publik berbahasa Indonesia mengenai AI di Twitter/X, dengan sampel berupa 4.104 data opini hasil *scraping* menggunakan kata kunci terkait AI. Sebagai data pendukung, disebarkan kuesioner daring melalui

PERANCANGAN SISTEM INFORMASI ANALISIS SENTIMEN OPINI PUBLIK TERHADAP TREN AI
MENGUNAKAN RANDOM FOREST DAN NAÏVE BAYES

Instagram, TikTok, dan Twitter/X menggunakan *Google Form*, dan diperoleh 327 responden yang menjawab 20 pertanyaan (19 pilihan ganda menggunakan Skala *Likert* dan 1 esai) terkait persepsi terhadap AI. Data kuesioner digunakan untuk mendukung analisis kebutuhan pengguna, bukan sebagai data latih model klasifikasi.

Instrumen kuesioner disusun menggunakan Skala *Likert* lima poin, dan hasil pengujian menunjukkan seluruh butir pertanyaan valid dan reliabel, sehingga layak digunakan untuk mengukur persepsi masyarakat terhadap perkembangan teknologi AI.

38 Data teks hasil *scraping* diproses melalui tahapan *cleaning*, *case folding*, normalisasi kata tidak baku, tokenisasi, *stopword removal*, dan *stemming* menggunakan *library Sastrawi* untuk Bahasa Indonesia — mengikuti prosedur baku *text preprocessing* pada penelitian sejenis (Oktaviani et al., 2025; Rudy Seftiawan et al., 2022) sehingga tidak dirinci ulang di sini. Data yang telah bersih selanjutnya diekstraksi menjadi representasi numerik menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF), lalu dibagi menjadi data latih dan data uji sebelum masuk ke tahap pemodelan.

39 Alat analisis data yang digunakan meliputi *Google Colab* sebagai lingkungan komputasi berbasis *cloud* untuk pelatihan model, *library Python* (*Pandas*, *NumPy*, *Scikit-learn*) untuk pengolahan data dan pemodelan, serta *Streamlit* untuk implementasi sistem berbasis web.

Model penelitian menggunakan dua algoritma klasifikasi. Algoritma pertama, *Random Forest*, bekerja dengan membentuk banyak pohon keputusan secara acak melalui teknik *bootstrap sampling* dan pemilihan fitur acak pada setiap simpul pohon, kemudian menggabungkan seluruh hasil prediksi pohon melalui mekanisme *voting* mayoritas untuk menentukan kelas sentimen akhir; pemilihan pemisah pada setiap pohon didasarkan pada tingkat kemurnian kelas (*Gini Index*) dan besarnya penurunan ketidakpastian data setelah pemisahan (*Information Gain*), mengikuti formulasi baku algoritma ini (Fitriani Noer Jamilah, 2026; Khoeruddin et al., 2023). Algoritma kedua, *Naïve Bayes*, menentukan kelas sentimen suatu opini dengan menghitung peluang kemunculan setiap kata pada masing-masing kelas berdasarkan Teorema Bayes, kemudian memilih kelas dengan nilai

e-ISSN: 2964-3104; p-ISSN: 2964-3090, Hal 00-00

peluang gabungan tertinggi sebagai hasil klasifikasi, mengikuti formulasi baku pada penelitian sejenis (Oktaviani et al., 2025)

Kinerja kedua model dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, *confusion matrix*, dan kurva *Receiver Operating Characteristic–Area Under Curve* (ROC-AUC), dengan kriteria kelayakan model berupa selisih akurasi data latih dan data uji tidak lebih dari 10% guna menghindari *overfitting*. Model dengan performa terbaik selanjutnya diimplementasikan ke dalam sistem informasi berbasis web menggunakan *Streamlit*, yang menyajikan distribusi sentimen, *word cloud*, tabel ringkasan klasifikasi, dan laporan performa model secara interaktif.

4. HASIL DAN PEMBAHASAN

4.1 Deskripsi Data Penelitian

Data utama penelitian diperoleh melalui proses *scraping* pada media sosial X (Twitter) menggunakan kata kunci terkait *Artificial Intelligence* (AI), menghasilkan 4.104 opini publik berbahasa Indonesia. Dari 15 atribut yang tersedia pada hasil *scraping* (di antaranya *full_text*, *created_at*, *username*, *retweet_count*, *favorite_count*), hanya atribut *full_text* yang digunakan sebagai data analisis karena memuat isi opini pengguna, sedangkan atribut lain berfungsi sebagai metadata pendukung. Sebagai data pelengkap, disebarkan kuesioner daring melalui Instagram, TikTok, dan X (Twitter) yang berhasil menjangkau 327 responden.

4.2 Hasil Pra-pemrosesan dan Pelabelan Data

Data teks diproses melalui tahapan *cleaning*, *case folding*, normalisasi, tokenisasi, *stopword removal*, dan *stemming* menggunakan *library Sastrawi*, kemudian diberi label sentimen secara manual ke dalam tiga kategori. Tabel 1 menunjukkan distribusi label sentimen pada dataset.

Tabel 1. Distribusi Label Sentimen Dataset

Kelas Sentimen	Jumlah Data	Persentase
Positif	1.382	33,7%

**PERANCANGAN SISTEM INFORMASI ANALISIS SENTIMEN OPINI PUBLIK TERHADAP TREN AI
MENGUNAKAN RANDOM FOREST DAN NAÏVE BAYES**

Netral	1.384	33,7%
Negatif	1.338	32,6%
Total	4.104	100%

Distribusi kelas yang relatif seimbang ini mendukung proses pelatihan model tanpa bias signifikan terhadap salah satu kelas — sejalan dengan prinsip bahwa keseimbangan kelas pada data latih penting untuk performa klasifikasi yang stabil (Srirahayu dkk., 2023). Dataset kemudian dibagi menjadi 2.547 data latih dan 637 data uji, dan diekstraksi menggunakan metode TF-IDF dengan ukuran kosakata (*vocabulary*) sebanyak 10.000 fitur.

4.3 Hasil Pemodelan dan Evaluasi

4.3.1 Random Forest

Model *Random Forest* dibangun dengan membentuk banyak pohon keputusan melalui proses *bootstrap sampling* dan pemilihan fitur acak pada setiap simpul. Proses *split* pada pohon keputusan ditentukan berdasarkan nilai *Gini Index* — pada *root node*, diperoleh nilai *Gini* sebesar 0,6666 yang mengindikasikan ketiga kelas sentimen masih terdistribusi relatif seimbang sebelum proses pemisahan dilakukan. Setelah *split* pertama, nilai *Information Gain* sebesar 0,0266 menunjukkan bahwa atribut yang dipilih berhasil menurunkan tingkat ketidakmurnian data, sehingga menghasilkan simpul anak yang lebih homogen dan mendukung akurasi klasifikasi yang lebih baik. Hasil evaluasi model *Random Forest* pada 637 data uji ditunjukkan pada Tabel berikut.

Tabel 2. *Classification Report Random Forest*

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Negatif	0,82	0,84	0,83
Netral	0,71	0,75	0,73
Positif	0,86	0,82	0,84
<i>Accuracy</i>			79,91%

Confusion matrix menunjukkan model berhasil mengklasifikasikan 170 dari 211 data negatif, 143 dari sekitar 194 data netral, dan 196 dari 232 data positif secara benar. Kurva ROC-AUC memperoleh nilai 0,94 untuk kelas negatif, 0,88 untuk kelas netral, dan

0,94 untuk kelas positif — menunjukkan kemampuan diskriminasi model yang sangat baik pada seluruh kelas.

4.3.2 Naïve Bayes

Model *Naïve Bayes* dibangun menggunakan algoritma *Multinomial Naïve Bayes* dengan parameter *smoothing* $\alpha = 0,1$ hasil *hyperparameter tuning* melalui *GridSearchCV*. Nilai probabilitas awal (*prior probability*) masing-masing kelas dihitung dari proporsi data latih: kelas negatif 0,3318, netral 0,2972, dan positif 0,3710 — menunjukkan kelas positif memiliki peluang awal tertinggi sebelum mempertimbangkan fitur kata pada dokumen. Sebagai ilustrasi perhitungan probabilitas fitur, kata "bagus" memperoleh bobot TF-IDF tertinggi pada kelas positif (48,70), kata "buruk" pada kelas negatif (40,30), dan kata "biasa" pada kelas netral (35,60), yang masing-masing menjadi penciri (*indicator*) kuat bagi kelasnya. Hasil evaluasi model *Naïve Bayes* pada data uji yang sama ditunjukkan pada Tabel berikut.

Tabel 3. *Classification Report Naïve Bayes*

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Negatif	0,73	0,79	0,76
Netral	0,66	0,48	0,55
Positif	0,71	0,81	0,76
<i>Accuracy</i>			70,17%

Model *Naïve Bayes* memperlihatkan kelemahan signifikan pada kelas netral, dengan *recall* hanya 0,48 karena sering tertukar dengan kelas negatif (54 kesalahan) dan positif (45 kesalahan). Nilai AUC untuk kelas netral juga paling rendah (0,77) dibandingkan kelas negatif (0,89) dan positif (0,90), yang mengindikasikan bahwa asumsi independensi antarfitur pada *Naïve Bayes* kurang mampu menangkap konteks kalimat yang ambigu atau tumpang tindih antarkelas.

4.3.3 Perbandingan Model

Perbandingan performa kedua algoritma disajikan pada Tabel berikut.

Tabel 4. Perbandingan Performa *Random Forest* dan *Naïve Bayes*

Metrik	<i>Random Forest</i>	<i>Naïve Bayes</i>	Selisih
--------	----------------------	--------------------	---------

**PERANCANGAN SISTEM INFORMASI ANALISIS SENTIMEN OPINI PUBLIK TERHADAP TREND AI
MENGUNAKAN RANDOM FOREST DAN NAÏVE BAYES**

<i>Accuracy</i>	79,91%	70,17%	9,74%
<i>F1-Score (Positif)</i>	0,84	0,78	0,06
<i>F1-Score (Netral)</i>	0,73	0,55	0,18
<i>F1-Score (Negatif)</i>	0,83	0,76	0,07

Random Forest unggul di seluruh metrik, dengan selisih paling besar pada kelas netral. Hasil ini sejalan dengan konsep *ensemble learning* yang mendasari *Random Forest*, di mana kombinasi banyak pohon keputusan menghasilkan prediksi yang lebih stabil dan tahan terhadap *overfitting* dibandingkan model probabilistik tunggal seperti *Naïve Bayes* (Fitriani Noer Jamilah, 2026; Khoeruddin dkk., 2023). Berdasarkan hasil tersebut, *Random Forest* dipilih sebagai model utama untuk diimplementasikan pada sistem.

Temuan ini konsisten dengan penelitian Wati dkk. (2025) yang membandingkan kedua algoritma pada ulasan aplikasi Deepseek, di mana *Random Forest* juga unggul dengan selisih akurasi yang besar terhadap *Naïve Bayes*. Hal serupa ditemukan oleh Ammar dan Wiranata (2025) pada analisis sentimen aplikasi SIGNAL, di mana *Random Forest* konsisten menjadi algoritma dengan performa terbaik dibandingkan *Naïve Bayes* dan *Decision Tree*. Ketiga penelitian — termasuk penelitian ini — menguatkan pola bahwa *Random Forest* secara umum lebih unggul untuk tugas klasifikasi sentimen berbasis teks berbahasa Indonesia, meskipun selisih akurasi bervariasi bergantung pada karakteristik dan volume data pada masing-masing domain.

4.4 Hasil Analisis Sentimen dan Persepsi Publik

Berdasarkan hasil klasifikasi model terbaik, distribusi sentimen publik terhadap tren AI menunjukkan sentimen positif mendominasi sebesar 36,5% (1.181 data), diikuti sentimen negatif 32,7% (1.056 data), dan sentimen netral 30,8% (996 data). Temuan ini menunjukkan kecenderungan masyarakat memberikan tanggapan positif terhadap perkembangan AI, meski proporsi sentimen negatif yang cukup tinggi mengindikasikan adanya kekhawatiran nyata terkait dampak teknologi tersebut.

Hasil ini diperkuat oleh data kuesioner, yang menunjukkan rata-rata persentase persetujuan responden sebesar 86,50% (kategori Sangat Baik) terhadap manfaat AI dalam berbagai aspek kehidupan. Pada pertanyaan terbuka, mayoritas responden (27,52%)

e-ISSN: 2964-3104; p-ISSN: 2964-3090, Hal 00-00

menyatakan AI membantu mempercepat penyelesaian pekerjaan, diikuti kemudahan aktivitas sehari-hari (22,02%) dan dukungan proses pembelajaran (14,68%). Namun demikian, sebagian responden juga menyoroti perlunya kebijaksanaan dalam penggunaan AI (12,23%) mengingat potensi dampak negatif seperti penyebaran informasi tidak benar (9,48%), pelanggaran privasi (4,28%), dan berkurangnya kebutuhan tenaga kerja pada bidang tertentu (3,67%). Konsistensi antara hasil analisis sentimen media sosial dan hasil kuesioner menunjukkan bahwa kedua sumber data saling menguatkan validitas temuan penelitian ini, sekaligus mengonfirmasi bahwa opini publik memiliki peran penting dalam membentuk penerimaan masyarakat terhadap teknologi AI (Vasiljeva et al., 2021).

4.5 Implementasi Sistem

Model *Random Forest* yang terpilih diimplementasikan ke dalam sistem informasi berbasis web menggunakan *framework Streamlit*, dengan empat menu utama: *Dashboard* (ringkasan distribusi sentimen), *Analisis Opini* (klasifikasi teks tunggal beserta *confidence score*), *Analisis Dataset* (klasifikasi *batch* dari berkas .csv/.xlsx dengan visualisasi *pie chart*, *bar chart*, dan *word cloud*), dan *Evaluasi Model* (perbandingan performa kedua algoritma). Pengujian fungsional menggunakan metode *Black Box Testing* terhadap 15 skenario menunjukkan seluruh fitur sistem berjalan sesuai spesifikasi yang dirancang, termasuk validasi masukan pada kondisi teks kosong dan berkas dengan format tidak sesuai.

4.6 Implikasi Penelitian

Secara teoritis, penelitian ini memperkuat bukti empiris bahwa *Random Forest* lebih unggul dibandingkan *Naïve Bayes* untuk klasifikasi sentimen teks berbahasa Indonesia, khususnya dalam menangani kelas netral yang cenderung ambigu. Secara terapan, sistem yang dihasilkan dapat dimanfaatkan oleh perusahaan, pengembang teknologi, maupun pembuat kebijakan untuk memantau opini publik terhadap AI secara cepat dan terstruktur, sehingga dapat menjadi dasar dalam perumusan strategi komunikasi maupun kebijakan pengembangan teknologi yang lebih responsif terhadap persepsi masyarakat.

5. KESIMPULAN DAN SARAN

PERANCANGAN SISTEM INFORMASI ANALISIS SENTIMEN OPINI PUBLIK TERHADAP TREND AI MENGGUNAKAN RANDOM FOREST DAN NAÏVE BAYES

Penelitian ini berhasil merancang Sistem Informasi Analisis Sentimen Opini Publik terhadap Tren Teknologi *Artificial Intelligence* (AI) menggunakan algoritma *Random Forest* dan *Naïve Bayes*. Hasil pengujian menunjukkan *Random Forest* memperoleh akurasi 79,91%, lebih unggul dibandingkan *Naïve Bayes* sebesar 70,17%, sehingga dipilih sebagai model utama pada sistem berbasis *Streamlit*. Hasil analisis terhadap 4.104 opini publik dari X (Twitter) menunjukkan sentimen positif mendominasi sebesar 36,5%, diikuti negatif 32,7% dan netral 30,8%, yang mengindikasikan masyarakat cenderung menanggapi perkembangan AI secara positif — temuan yang diperkuat oleh hasil kuesioner terhadap 327 responden dengan tingkat persetujuan rata-rata 86,50%. Berdasarkan hasil tersebut, sistem informasi analisis sentimen yang dikembangkan telah mampu menjalankan fungsinya sesuai dengan tujuan penelitian.

Penelitian ini memiliki keterbatasan, yaitu pelabelan sentimen yang dilakukan secara manual berpotensi menimbulkan subjektivitas, data hanya bersumber dari satu platform media sosial, serta *Naïve Bayes* menunjukkan keterbatasan dalam menangani kalimat kompleks dan sarkasme sehingga generalisasi hasil perlu dilakukan secara hati-hati. Penelitian selanjutnya dapat mengembangkan sistem dengan menerapkan metode klasifikasi pada berbagai topik atau domain lain, serta memanfaatkan data opini publik yang berasal dari berbagai platform media sosial selain X (Twitter), seperti Instagram, Facebook, TikTok, atau platform lainnya, guna memperoleh representasi opini publik yang lebih menyeluruh.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada dosen pembimbing atas arahan dan masukan selama proses penelitian, serta kepada Institut Teknologi dan Bisnis Swadharma atas dukungan fasilitas yang diberikan. Artikel ini merupakan bagian dari skripsi penulis yang berjudul "Perancangan Sistem Informasi Analisis Sentimen Opini Publik terhadap Tren Teknologi *Artificial Intelligence* (AI) Menggunakan Metode *Random Forest* dan *Naïve Bayes*".

DAFTAR REFERENSI

Agung, F. N., Junaedi, I., & Yulianto, A. B. (2022). Perancangan Sistem Informasi Pelayanan Customer Dengan Platform Web. *Jurnal Manajemen Informatika Jakarta*, 2, 320–325. <https://doi.org/10.52362/jmijayakarta.v2i4.916>

- 14 Agustika, F., Siregar, S., Obara, D., & Paramarta, V. (2023). Telaah Teknologi Informasi Dan Sistem Informasi Dalam Organisasi Dengan Lingkungan. *Jurnal Bisnis Kolega (JBK)*, 9(1), 24–33. <https://doi.org/https://doi.org/10.57249/jbk.v9i1.104>
- 3 Ammar, H., & Wiranata, A. D. (2025). Perbandingan Tingkat Akurasi Algoritma Naïve Bayes , Decision Tree dan Random Forest dalam Analisis Sentimen Pengguna Aplikasi Samsat Digital Nasional. *Indonesian Journal of Applied Informatics(IJAI)*, 10(1), 59–73. <https://doi.org/10.20961/ijai.v10i1.111010>
- 6 Arifin, H., Ahsyar, T. K., Islam, U., Sultan, N., & Riau, K. (2024). Analisis Keamanan Sistem Informasi Menggunakan Cobit 2019 pada Sistem Sawit Rakyat Online (SRO) Studi Kasus PTPN V. *Jurnal Teknologi Sistem Informasi Dan Aplikasi*, 7(3), 903–911. <https://doi.org/10.32493/jtsi.v7i3.39650>
- 18 Arifin, O., Azim, F., Hartati, Y., Widyawati, D. K., & Ahmad, A. L. (2025). Comparative Study on the Efficiency of Deep Learning Model Training in Cloud Environments: Google Colab vs AWS. *Jurnal Pendidikan Dan Teknologi Informasi*, 5(2), 324–332. <https://doi.org/http://dx.doi.org/10.51454/decode.v5i2.1197>
- 5 Dewi, K., Susilowati, S., Kusmintarti, A., & Riwanjanti, N. I. (2025). Integrating Streamlit , Facebook prophet , and machine learning models to create an interactive stock price and risk prediction dashboard. *Edelweiss Applied Science and Technology*, 9(11), 1453–1469. <https://doi.org/10.55214/2576-8484.v9i11.11246>
- 25 Fitriani Noer Jamilah, D. M. I. (2026). Implementasi Data Mining Untuk Prediksi Dan Klasifikasi Tingkat Stres Menggunakan Algoritma Random Forest. *Jurnal Informatika Dan Teknik Elektro Terapan(JITET)*, 14(1), 905–913. <https://doi.org/https://doi.org/10.23960/jitet.v14i1.8838>
- 21 Gheisari, M. (2023). Data Mining Techniques for Web Mining : A Survey. *Artificial Intelligence and Applications*, 1(1), 3–10. <https://doi.org/10.47852/bonviewAIA2202290>
- 26 Khoeruddin, A., Sudrajat, F. A., & Purnama, G. (2023). Optimasi Fitur Seleksi Random Forest Menggunakan GA Dalam Klasifikasi Data Penyakit Gagal Jantung. *Jurnal Penelitian Teknologi Informasi Dan Sains(JPTIS)*, 1(2), 1–9. <https://doi.org/https://doi.org/10.54066/jptis.v1i2.323>
- 37 M.Ulil Absor, A. H. (2024). Gudang Jurnal Multidisiplin Ilmu Perancangan Sistem Informasi UKS SDN 1 Pakis Kabupaten Banyuwangi. *Gudang Jurnal Multidisiplin Ilmu*, 2(10), 247–256. <https://doi.org/https://doi.org/10.59435/gjmi.v2i10.991>
- 4 Nawulansih, D. F., Santi, N. C., Aristia, I., Informatika, T., Nahdlatul, U., Sunan, U., Bojonegoro, G., Informasi, S., Nahdlatul, U., Sunan, U., & Bojonegoro, G. (2025). Aplikasi DANA di Google Play Store: Penerapan Support Vector Machine dan Synthetic Minority Over-sampling Technique. *Jurnal Pendidikan Dan Teknologi Indonesia (JPTI)*, 5(9), 2660–2671. <https://doi.org/https://doi.org/10.52436/1.jpti.1053>
- 24 Nugrahanti, F., & Sussolaikah, K. (2021). Metode Importance Performa Analisis (IPA) sebagai Penentuan Kepuasan dan Kenyamanan Pengguna Website Universitas PGRI Madiun. *Jurnal Ilmiah Sistem Informasi, Teknologi Informasi Dan Sistem Komputer*, 16(2), 128–134.
- 28

PERANCANGAN SISTEM INFORMASI ANALISIS SENTIMEN OPINI PUBLIK TERHADAP TREN AI
MENGUNAKAN RANDOM FOREST DAN NAÏVE BAYES

<https://doi.org/https://doi.org/10.33998/processor.2021.16.2.996>

- Oktaviani, Kurniasari, E., Rosdiana, & Putra, F. K. (2025). Pemanfaatan Metode Naïve Bayes untuk Analisis Sentimen Terkait Insiden Server Down dan Peretasan pada Pusat Data Nasional. *Jurnal Minfo Polgan*, 14(1), 926–933. <https://doi.org/https://doi.org/10.33395/jmp.v14i1.14929>
- Parker, A., Heflin, A., & Jones, L. C. (2021). Analyzing University of Virginia Health publications using open data , Python , and Streamlit. *Journal of the Medical Library Association(JMLA)*, 109(4), 688–689. <https://doi.org/https://doi.org/10.5195/jmla.2021.1360>
- Putra, A., Mansyuri, U., & Aryono, G. D. P. (2025). Analisis Penggunaan Sistem Face Recognition dalam Mengelola Absensi Karyawan di PT Bintang Inspeksi Indonesia. *Jurnal Ilmiah Sistem Informasi Dan Ilmu Komputer*, 5(1), 1–8. <https://doi.org/https://doi.org/10.55606/juisik.v5i1.940>
- Rahman, A., Latifah, E. D., Fachrurrazi, S., Pertanian, A. F., Malikussaleh, U., Sistem, P., Fakultas, I., Malikussaleh, U., & Utara, A. (2022). Peranan Teknologi Informasi Dalam Meningkatkan Partisipasi Politik Warga Negara. *Jurnal Ilmiah Sistem Informasi*, 6, 11–23. <https://doi.org/https://doi.org/10.29103/sisfo.v6i1.7961>
- Razaqi, A. W., Loka, N. P., Alfian, M., & Yudha, H. (2024). Optimalisasi Bisnis Melalui Artificial Intelligence Dengan Analisis Peluang, Tantangan Dan Dampak Di Berbagai Sektor. *Jurnal Teknologi Informasi*, 5(3), 727–741. <https://doi.org/10.46576/djtechno>
- Rudy Seftiawan, Kurniawan, T. B., & Misinem, M. (2022). Analisis Sentimen pada e-Government Mobile Application Indonesia Menggunakan Algoritma Naïve Bayes. *Jurnal Ilmiah Komputasi*, 21(1), 43–50. <https://doi.org/https://doi.org/10.32409/jikstik.21.1.2811>
- Satria, W. I., Hermawan, Y., & Figo, M Akbar, H. (2021). Definisi Sistem Portal Pengetahuan Dalam Pemerintahan Berdasarkan Pendekatan Modifikasi Soft System Methodology dan Hermeneutic. *Jurnal Informatika Kesatuan (JIKES)*, 1(1), 61–70. <https://doi.org/10.37641/jikes.v1i1.518>
- Srirahayu, Agustina, Priadi, & Setya, L. (2023). Review Paper Data Mining Klasifikasi Data Mining. *Jurnal Ilmiah Informatika Global*, 14(1), 7–12. <https://doi.org/10.36982/jiig.v13i2.2307>
- Supriyati et al. (2022). Journal of Information System. *Jurnal Sistem Informasi (Journal of Information System)*, 18(2), 1–81. <https://doi.org/https://doi.org/10.21609/jsi.v18i2>
- Syaifunazhirin, F. (2023). Literatur system-of-system untuk desain dan fungsi sistem informasi teknologi. *Integrated Journal of Information Technology and Vocational Education*, 5(2), 96–104. <https://doi.org/10.17509/integrated.v5i2.64127>
- Vasiljeva, T., Kreituss, I., & Lulle, I. (2021). Artificial Intelligence : The Attitude of the Public and Representatives of Various Industries. *Journal of Risk and Financial Management*, 14(8), 1–17. <https://doi.org/https://doi.org/10.3390/jrfm14080339>
- Wati, F. F., Suleman, & Widodo, A. E. (2025). Analisis Sentimen Ulasan Pengguna

e-ISSN: 2964-3104; p-ISSN: 2964-3090, Hal 00-00

Aplikasi Deepseek Menggunakan Algoritma Random Forest Dan Naive Bayes. *Computer and Network Technology, Universitas BSI*, 5(1), 8–15. <https://doi.org/https://doi.org/10.31294/hqpha267>

Yakut, Ö. (2023). Diabetes Prediction Using Colab Notebook Based Machine Learning Methods. *International Journal of Computational and Experimental Science and ENgineering (IJCESN)*, 9(1), 36–41. <https://doi.org/10.22399/ijcesen.1185474>

Yanuargi, B., Ema Utami, K., & Parikesit, A. A. (2025). Deep Learning Neural Dalam Analisis Sentimen : Sebuah Studi Literatur. *Jurnal Ilmiah Teknologi Sistem Informasi*, 6(3), 321–328. <https://doi.org/10.62527/jitsi.6.3.499>