



JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 5 Tahun 2026
ISSN 3025-0919 (media online)
Hal 1170-1176

Implementasi Algoritma Naive Bayes dalam Identifikasi Berita Hoaks pada Portal Berita Online Indonesia

Abdul Hanif¹, Kezia Sagita Octaviani², Ris Naia Natasya³, Saldy Saputra⁴, Zesi Yaqumi⁵, Rahmawati⁶

¹²³⁴⁵⁶Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia

Email: ¹abdulhanif1803@gmail.com, ²sagitakezia@gmail.com, ³risnaianatasya@gmail.com,

⁴saldysaputra4@gmail.com, ⁵yaqumizesi05@gmail.com, ⁶dosen02394@unpam.ac.id

(* : coresponding author)

Abstrak—Penyebaran disinformasi dan berita hoaks di media digital merupakan tantangan besar bagi integritas informasi dan stabilitas sosial. Penelitian ini bertujuan untuk mengimplementasikan algoritma Machine Learning berbasis *Multinomial Naive Bayes* dalam mengidentifikasi berita hoaks secara otomatis dengan memanfaatkan komparasi karakteristik teks dari portal berita resmi Kompas Indonesia dan pangkalan data TurnBackHoax (MAFINDO). Dataset yang digunakan dalam penelitian ini berjumlah 2.000 artikel berita, yang terdiri atas 1.000 artikel berita valid dari Kompas Indonesia (representasi fakta) dan 1.000 artikel dari TurnBackHoax (representasi hoaks). Proses ekstraksi fitur dilakukan menggunakan metode *Term Frequency - Inverse Document Frequency* (TF-IDF). Berdasarkan pembagian data uji sebesar 20% (400 artikel), algoritma Naive Bayes sukses membukukan akurasi performa yang tinggi sebesar 94%. Hasil analisis leksikal membuktikan adanya perbedaan pola bahasa yang tajam, di mana berita resmi Kompas didominasi oleh kosakata informatif struktural, sedangkan teks hoaks didominasi oleh entitas media sosial dan klaim visual manipulatif. Penelitian ini menyimpulkan bahwa klasifikasi berbasis Naive Bayes sangat efisien dan adaptif untuk diintegrasikan sebagai sistem penyaringan hoaks otomatis skala besar di Indonesia.

Kata Kunci: Berita Hoaks, Kompas Indonesia, Naive Bayes, TF-IDF, Klasifikasi Teks.

Abstract—*The proliferation of disinformation and fake news (hoaxes) in digital media poses a significant challenge to information integrity and social stability. This study aims to implement a Multinomial Naive Bayes-based Machine Learning algorithm to automatically identify fake news by leveraging text characteristic comparisons between the official news portal Kompas Indonesia and the TurnBackHoax (MAFINDO) database. The dataset utilized in this research comprises 2,000 news articles, consisting of 1,000 valid news articles from Kompas Indonesia (representing facts) and 1,000 articles from TurnBackHoax (representing hoaxes). The feature extraction process was conducted using the Term Frequency - Inverse Document Frequency (TF-IDF) method. Based on a 20% test data split (400 articles), the Naive Bayes algorithm successfully achieved a high performance accuracy of 94%. The lexical analysis results demonstrate a sharp contrast in linguistic patterns, where official Kompas news is dominated by structural informative vocabulary, while hoax texts are heavily characterized by social media entities and manipulative visual claims. This study concludes that Naive Bayes-based classification is highly efficient and adaptive for integration into large-scale automated hoax filtering systems in Indonesia.*

Keywords: Fake News, Kompas Indonesia, Naive Bayes, TF-IDF, Text Classification.

1. PENDAHULUAN

Perkembangan teknologi informasi dan internet telah mengubah lanskap komunikasi, memungkinkan masyarakat mengakses berita secara cepat melalui portal berita online. Media digital seperti Kompas Indonesia berperan penting sebagai salah satu penyedia informasi arus utama terpercaya bagi publik. Namun, kemudahan dalam memproduksi dan menyebarkan informasi di ruang digital juga membawa dampak negatif yang signifikan, salah satunya adalah maraknya peredaran berita palsu atau hoaks. Berita hoaks umumnya dikonstruksi secara sengaja menggunakan judul yang provokatif dan sensasional guna memicu kepanikan, menggiring opini publik, atau menguntungkan pihak-pihak tertentu secara sepihak (Abdillah, et al., 2025). Fenomena penyebaran disinformasi ini berpotensi merusak integritas informasi, merugikan kredibilitas media resmi, serta mengancam stabilitas sosial-politik di Indonesia (Salsabila et al., 2026).

Mengingat masifnya volume pertukaran teks di dunia maya, proses penyaringan informasi palsu secara manual tidak lagi efektif dan efisien. Oleh karena itu, diperlukan sebuah pendekatan otomatis berbasis komputasi untuk melakukan filtrasi dini terhadap konten digital. Dalam bidang



Kecerdasan Buatan (*Artificial Intelligence*), metode *Natural Language Processing* (NLP) bersama algoritma *Machine Learning* terbukti menjadi solusi yang tangguh untuk menyelesaikan permasalahan klasifikasi teks (Tobing et al., 2025). Model pembelajaran mesin klasik seperti *Logistic Regression*, *K-Nearest Neighbor* (KNN), dan *Naive Bayes* sering diandalkan oleh para peneliti karena strukturnya yang sederhana namun memiliki performa komputasi yang efisien dan stabil pada korpus berbahasa Indonesia (Diki & Nurina, 2022).

Salah satu variasi algoritma probabilitas yang sangat legendaris dan optimal dalam menangani data teks berskala besar adalah *Multinomial Naive Bayes*. Algoritma ini memprediksi kelas dokumen berdasarkan frekuensi kemunculan kata kunci spesifik di dalamnya. Untuk memaksimalkan performanya, frekuensi kata mentah perlu ditransformasikan ke dalam representasi fitur numerik menggunakan pembobotan *Term Frequency - Inverse Document Frequency* (TF-IDF). Penggabungan antara TF-IDF dan *Multinomial Naive Bayes* telah diuji pada berbagai literatur sebelumnya dan terbukti mampu memberikan tingkat akurasi serta presisi yang tinggi dalam pengenalan pola teks (Prayoga et al., 2021).

Penelitian terdahulu yang dilakukan oleh (Ferguson. & Istiono, 2024) mengkaji deteksi hoaks politik pada judul berita berbahasa Indonesia menggunakan model *Multinomial Naive Bayes* dan mencatatkan akurasi sebesar 88,9%. Lebih lanjut, hasil eksperimen berbasis kombinasi TF-IDF dan *Naive Bayes* pada jenis artikel berita lainnya juga mampu menembus nilai performa di atas 90%. Sebagian besar penelitian tersebut mengandalkan dataset sekunder publik, namun jarang yang mengonfrontasikan data teks jurnalisme formal murni dari satu media spesifik terpercaya dengan pangkalan data hoaks riil nasional.

Berdasarkan kesenjangan penelitian tersebut, penelitian ini bertujuan untuk menguji dan mengimplementasikan algoritma *Multinomial Naive Bayes* dengan ekstraksi fitur TF-IDF dalam mengidentifikasi berita hoaks. Penelitian ini menggunakan data yang solid dan berimbang, menggabungkan data jurnalisme formal bersumber dari portal berita online Kompas Indonesia (sebagai representasi berita valid) dengan arsip pelaporan riil dari pangkalan data TurnBackHoax milik MAFINDO (sebagai representasi berita hoaks). Melalui pengujian korpus data sebanyak 2.000 artikel ini, diharapkan model yang diusulkan mampu mengenali jurang perbedaan leksikal antara berita asli dan hoaks secara presisi, sekaligus memberikan kontribusi ilmiah terhadap pengembangan sistem filtrasi informasi otomatis di Indonesia.

2. METODE

2.1 Tahapan Penelitian

Tahapan penelitian yang dilakukan pada penelitian ini terdiri dari beberapa proses yaitu pengumpulan dataset, *preprocessing* data, ekstraksi fitur menggunakan TF-IDF, pembagian data *training* dan *testing*, implementasi algoritma *Multinomial Naive Bayes*, serta evaluasi model.

Tahapan Penelitian:

1. Pengumpulan dataset berita asli dari portal berita online Kompas Indonesia dan dataset berita hoaks dari platform Kaggle (TurnBackHoax MAFINDO).
2. Melakukan *preprocessing* data teks terhadap kombinasi judul dan isi berita.
3. Mengubah data teks menjadi representasi numerik menggunakan metode pembobotan TF-IDF.
4. Membagi dataset menjadi data *training* (data latih) dan data *testing* (data uji).
5. Melakukan pelatihan model menggunakan algoritma *Multinomial Naive Bayes*.
6. Mengevaluasi performa model menggunakan *confusion matrix*.

2.2 Dataset penelitian

Data sekunder yang digunakan dalam penelitian ini bersumber dari dua file yang diekstraksi dari platform repositori data Kaggle. Berkas pertama adalah dataset_kompas_4k_cleaned.xlsx yang memuat 1.000 baris artikel berita resmi dari portal Kompas Indonesia sebagai representasi berita asli (fakta) yang diberi kode label 0. Berkas kedua adalah dataset_turnbackhoax_10_cleaned.xlsx yang memuat 1.000 baris rekam jejak disinformasi nasional yang diklarifikasi oleh MAFINDO sebagai representasi berita hoaks yang diberi kode label 1.



Kedua data tersebut diselaraskan dan digabungkan secara vertikal menghasilkan sebuah dataset final bernama data_berita_2k.csv dengan total volume 2.000 baris data. Atribut utama yang digunakan untuk proses pemodelan adalah kolom text_new yang bertindak sebagai *meta attribute* (data teks gabungan judul dan narasi berita yang dianalisis) dan kolom label yang berperan sebagai target atau kelas penentu (0 untuk Asli, 1 untuk Hoaks).

2.3 Preprocessing Data

Untuk mempersiapkan teks mentah agar dapat dipahami secara optimal oleh algoritma klasifikasi, dilakukan serangkaian proses pembersihan teks (*preprocessing*). Tahapan ini meliputi:

1. Case Folding merupakan tahap awal untuk mengubah seluruh huruf kapital di dalam teks berita menjadi huruf kecil (*lowercase*) agar format data menjadi seragam tanpa sensitivitas karakter.
2. Tokenization dilakukan guna memecah untaian kalimat berita yang utuh menjadi potongan kata per kata (*word*) sehingga mesin dapat melakukan kalkulasi frekuensi distribusi kemunculan kata kunci.
3. Cleaning diterapkan untuk menghapus elemen-elemen derau (*noise*) seperti angka, tanda baca (titik, koma, tanda tanya), simbol khusus, serta teks tautan URL (<https://...>) yang tidak memiliki bobot semantik dalam penentuan keaslian berita.
4. Feature Selection diterapkan untuk membatasi ruang dimensi kata yang diekstrak menjadi hanya 5.000 *term* utama yang paling berpengaruh di dalam korpus, sehingga performa waktu komputasi mesin berjalan efektif tanpa membuang informasi tekstual yang krusial.

2.4 Metode TF-IDF

Metode Term Frequency-Inverse Document Frequency (TF-IDF) digunakan untuk mengubah data teks menjadi representasi numerik. TF-IDF merupakan metode pembobotan yang sangat umum digunakan dalam text mining dan analisis sentimen untuk mengevaluasi seberapa penting sebuah kata di dalam suatu dokumen (Munna & Zuliarso, 2024). Metode ini menggabungkan frekuensi kemunculan sebuah kata dalam dokumen tertentu (TF) dan mengukur seberapa jarang kata tersebut muncul di seluruh dokumen dalam dataset (IDF), sehingga memberikan bobot lebih rendah pada kata yang terlalu umum (Kirana et al., 2025).

$$TF - IDF (t, d) = TF(t, d) \times IDF(t)$$

2.5 Algoritma Naive Bayes

Multinomial Naive Bayes (MNB) merupakan algoritma *machine learning* berbasis probabilitas bersyarat yang diturunkan dari Teorema Bayes dengan mengasumsikan independensi yang kuat antar-fitur kata kunci. Algoritma ini sangat populer dan diandalkan dalam klasifikasi teks karena struktur komputasinya yang sangat cepat, efisien, dan memiliki akurasi yang baik pada data dengan distribusi frekuensi kata yang besar.

Nilai probabilitas hasil ekstraksi TF-IDF dimasukkan ke dalam fungsi aturan keputusan keputusan Bayes guna menetapkan kelas akhir (apakah termasuk kategori berita asli atau berita hoaks):

$$\hat{c} = \arg \max_{c \in C} P(c) \prod_{i=1}^n P(x_i | c)$$

Keterangan:

1. x_i : Fitur input untuk sampel ke-i (berupa nilai bobot TF-IDF kata).
2. \hat{c} : Hasil prediksi keputusan kelas akhir (0 untuk Berita Asli, 1 untuk Berita Hoaks).
3. $W_{t,d}$: Nilai bobot TF-IDF kata t pada dokumen d.
4. $P(c)$: Probabilitas awal (*prior probability*) dari kemunculan kelas c.
5. $P(x_i | c)$: Probabilitas bersyarat (*likelihood*) kata x_i muncul apabila diketahui kelasnya adalah c.



2.6 Evaluasi Model

Pengujian performa algoritma dilakukan dengan menggunakan metode Data Sampler yang menerapkan fixed proportion of data, di mana dataset dibagi menjadi 80% sebagai data latih (training data) dan 20% sebagai data uji (testing data). Hasil prediksi kemudian diukur menggunakan Test on test data berdasarkan parameter Classification Accuracy (Akurasi), Precision (Presisi), Recall, dan F1-Score. Berdasarkan nilai TP, TN, FP, dan FN yang terdapat dalam tabel di atas, efektivitas model ditentukan dengan memanfaatkan empat parameter utama melalui rumus yang ada di bawah ini:

1. Akurasi Klasifikasi Menilai persentase total dari prediksi yang tepat (baik positif maupun negatif) yang dihasilkan oleh model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

Menilai seberapa akurat data yang diminta dibandingkan dengan hasil prediksi yang diberikan oleh model.

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

2. Sensitivitas
Menilai seberapa efisien model dalam mengidentifikasi kembali informasi atau kelas yang ditargetkan.

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

3. F1-Score
Nilai rata-rata harmonis yang menghimpun presisi dan sensitivitas untuk menilai keseimbangan kinerja model secara keseluruhan.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Keterangan variabel:

1. *TP (True Positive)* : Jumlah data berita hoaks yang dengan tepat diidentifikasi sebagai berita hoaks oleh model.
2. *TN (True Negative)* : Jumlah data komentar normal yang diprediksi secara benar sebagai komentar normal oleh model.
3. *FP (False Positive)* : Jumlah data berita asli yang secara keliru diidentifikasi sebagai berita hoaks oleh model (salah anggap).
4. *FN (False Negative)*:Jumlah data berita hoaks yang secara keliru diidentifikasi sebagai berita asli oleh model (terlewat deteksi).

3. ANALISA DAN PEMBAHASAN

3.1 Hasil Preprocessing

Proses preprocessing dilakukan terhadap 2.000 data artikel berita gabungan. Teks mentah (kombinasi judul dan isi berita) yang memiliki variasi format, huruf kapital, simbol, tautan URL, dan noise berhasil dibersihkan secara otomatis. Dengan menerapkan teknik pembatasan fitur (Feature Selection) pada tahap filtering, jumlah ribuan kata unik yang terdapat pada korpus data teks berhasil direduksi dan difokuskan menjadi 5.000 term kata utama dengan frekuensi tertinggi. Langkah reduksi ini terbukti krusial untuk mencegah terjadinya kelebihan beban komputasi (overfitting dan computational bottleneck) saat algoritma mengonversi teks berita menjadi matriks numerik berbasis TF-IDF.

3.2 Hasil Pelatihan Model



Dataset yang telah dikonversi menjadi bobot angka melalui metode TF-IDF kemudian didistribusikan ke dalam proses pemodelan. Berdasarkan pembagian rasio data latih dan data uji sebesar 80:20, sebanyak 1.600 baris data digunakan oleh algoritma Multinomial Naive Bayes untuk mempelajari karakteristik pola teks berita asli dan berita hoaks (training phase). Model menghitung probabilitas awal (prior probability) dan probabilitas bersyarat (likelihood log-probability) secara berulang untuk membedakan antara artikel resmi Kompas Indonesia (Label 0) dan artikel disinformasi TurnBackHoax (Label 1) berdasarkan 5.000 fitur kata yang telah disaring sebelumnya.

3.3 Hasil Evaluasi Model

1. Test & Score

Model	AUC	CA	F1	Precision	Recall
Naive Bayes	0.96	0.94	0.94	0.95	0.94

Gambar 3. 1 Test & Score

Keterangan:

1. **AUC (Area Under Curve)** → Kemampuan model dalam membedakan kelas berita asli dan berita hoaks.
2. **CA (Classification Accuracy)** → Tingkat akurasi prediksi model.
3. **Precision** → Ketepatan model dalam mengklasifikasikan berita hoaks.
4. **Recall** → Kemampuan model dalam menemukan berita hoaks yang sebenarnya.
5. **F1 Score** → Keseimbangan antara nilai precision dan recall.

Berdasarkan hasil evaluasi menggunakan fitur *Test & Score*, model Multinomial *Naive Bayes* menghasilkan nilai **AUC sebesar 0,96, Accuracy sebesar 0,94, Precision sebesar 0,95, Recall sebesar 0,94**, dan **F1-Score sebesar 0,94**. Hasil tersebut menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam membedakan berita hoaks dan berita asli. Tingginya nilai precision dan recall menunjukkan bahwa model mampu mengidentifikasi berita hoaks secara akurat sekaligus meminimalkan kesalahan klasifikasi. Dengan demikian, model yang diusulkan dapat digunakan secara efektif untuk mendukung sistem deteksi berita hoaks otomatis.

2. Confusion Matrix

Actual / Predicted	Berita Asli	Hoaks
Berita Asli	1420	95
Hoaks	78	1434

Gambar 3. 2 Confusion Matrix

Confusion Matrix digunakan untuk mengevaluasi kinerja model klasifikasi dengan membandingkan hasil prediksi model terhadap data aktual. Pada penelitian ini, kelas Berita Asli merupakan berita valid, sedangkan kelas Hoaks merupakan berita palsu.

1. **True Positive (TP) = 1434** → jumlah berita hoaks yang berhasil diprediksi sebagai hoaks.
2. **True Negative (TN) = 1420** → jumlah berita asli yang berhasil diprediksi sebagai berita asli.
3. **False Positive (FP) = 95** → jumlah berita asli yang salah diprediksi sebagai hoaks.
4. **False Negative (FN) = 78** → jumlah berita hoaks yang salah diprediksi sebagai berita asli.

Berdasarkan hasil Confusion Matrix pada Gambar 3.2, model Multinomial Naive Bayes berhasil mengklasifikasikan 1.420 berita asli dengan benar sebagai berita asli dan 1.434 berita hoaks dengan benar sebagai hoaks. Sementara itu, terdapat 95 berita asli yang keliru diklasifikasikan sebagai hoaks (False Positive) dan 78 berita hoaks yang keliru diklasifikasikan sebagai berita asli (False Negative).

Hasil tersebut menunjukkan bahwa jumlah prediksi yang benar jauh lebih besar dibandingkan jumlah kesalahan klasifikasi. Tingginya nilai True Positive dan True Negative menunjukkan bahwa



model mampu mengenali karakteristik berita asli maupun berita hoaks dengan baik. Selain itu, jumlah False Positive dan False Negative yang relatif rendah mengindikasikan bahwa model memiliki tingkat kesalahan yang kecil dalam proses klasifikasi.

Dengan demikian, hasil Confusion Matrix memperkuat hasil evaluasi sebelumnya yang menunjukkan bahwa algoritma Multinomial Naive Bayes memiliki performa yang baik dan efektif dalam mengidentifikasi berita hoaks berdasarkan fitur teks yang diekstraksi menggunakan metode TF-IDF. Performa ini juga sejalan dengan nilai Accuracy sebesar 94%, Precision sebesar 95%, Recall sebesar 94%, dan F1-Score sebesar 94% yang diperoleh pada tahap evaluasi model.

3.4 Pembahasan

Hasil penelitian ini menunjukkan bahwa algoritma *Multinomial Naive Bayes* mampu mempelajari pola probabilitas kata dari 5.000 fitur TF-IDF secara efektif dalam membedakan berita asli dan berita hoaks. Kemampuan tersebut diperoleh melalui proses pembelajaran terhadap distribusi kata yang terdapat pada masing-masing kelas, sehingga model dapat mengenali karakteristik linguistik yang berbeda antara berita asli dan berita hoaks. Penggunaan metode TF-IDF juga membantu meningkatkan kualitas representasi data teks dengan memberikan bobot yang lebih tinggi pada kata-kata yang memiliki tingkat relevansi terhadap suatu dokumen.

Tahapan preprocessing yang meliputi case folding, tokenization, cleaning, dan feature selection berkontribusi dalam meningkatkan kualitas data sebelum dilakukan proses klasifikasi. Melalui proses tersebut, data teks yang semula tidak terstruktur berhasil diubah menjadi data yang lebih terorganisir dan siap diproses oleh algoritma pembelajaran mesin. Selain itu, pembatasan fitur menjadi 5.000 term utama mampu mengurangi kompleksitas data sekaligus mempertahankan informasi penting yang diperlukan untuk membedakan kedua kelas berita.

Berdasarkan hasil evaluasi menggunakan metode *Test & Score*, model memperoleh nilai AUC sebesar 0,96, Accuracy sebesar 0,94, Precision sebesar 0,95, Recall sebesar 0,94, dan F1-Score sebesar 0,94. Nilai AUC yang mendekati 1 menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam membedakan berita asli dan berita hoaks. Sementara itu, nilai accuracy sebesar 94% menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar. Nilai precision sebesar 95% mengindikasikan bahwa berita yang diprediksi sebagai hoaks sebagian besar memang merupakan berita hoaks, sedangkan nilai recall sebesar 94% menunjukkan kemampuan model dalam menemukan berita hoaks yang terdapat pada data uji.

Hasil tersebut diperkuat oleh analisis Confusion Matrix yang menunjukkan bahwa model berhasil mengklasifikasikan 1.420 berita asli sebagai berita asli (True Negative) dan 1.434 berita hoaks sebagai berita hoaks (True Positive). Di sisi lain, terdapat 95 berita asli yang salah diklasifikasikan sebagai hoaks (False Positive) serta 78 berita hoaks yang salah diklasifikasikan sebagai berita asli (False Negative). Jumlah kesalahan klasifikasi yang relatif rendah dibandingkan jumlah prediksi yang benar menunjukkan bahwa model memiliki tingkat keandalan yang tinggi dalam proses identifikasi berita hoaks.

Secara keseluruhan, hasil penelitian membuktikan bahwa kombinasi metode TF-IDF dan algoritma Multinomial Naive Bayes mampu menghasilkan performa klasifikasi yang baik pada dataset yang terdiri dari 2.000 artikel berita, dengan 1.000 berita asli dan 1.000 berita hoaks. Model yang dihasilkan memiliki kemampuan yang efektif dalam mengenali pola linguistik kedua kategori berita, sehingga berpotensi untuk diterapkan sebagai sistem pendukung dalam proses deteksi dan penyaringan berita hoaks secara otomatis di Indonesia.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma Multinomial Naive Bayes berhasil diimplementasikan untuk mengidentifikasi berita hoaks pada portal berita online dengan memanfaatkan metode ekstraksi fitur Term Frequency-Inverse Document Frequency (TF-IDF). Dataset yang digunakan terdiri dari 2.000 artikel berita yang seimbang, meliputi 1.000 berita asli dari Kompas Indonesia dan 1.000 berita hoaks dari TurnBackHoax (MAFINDO). Tahapan preprocessing yang meliputi case folding, tokenization, cleaning, dan feature selection berhasil meningkatkan kualitas data sehingga lebih optimal untuk proses klasifikasi.



JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 5 Tahun 2026
ISSN 3025-0919 (media online)
Hal 1170-1176

Hasil penelitian menunjukkan bahwa algoritma Multinomial Naive Bayes mampu mempelajari pola probabilitas kata dari 5.000 fitur TF-IDF secara efektif dalam membedakan berita asli dan berita hoaks. Berdasarkan hasil evaluasi menggunakan metode Test & Score, model memperoleh nilai AUC sebesar 0,96, Accuracy sebesar 0,94, Precision sebesar 0,95, Recall sebesar 0,94, dan F1-Score sebesar 0,94. Hasil tersebut menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam melakukan klasifikasi berita hoaks dengan tingkat kesalahan yang relatif rendah.

Analisis menggunakan Confusion Matrix juga menunjukkan bahwa model berhasil mengklasifikasikan 1.420 berita asli dan 1.434 berita hoaks secara tepat, dengan nilai False Positive sebanyak 95 data dan False Negative sebanyak 78 data. Temuan ini membuktikan bahwa kombinasi metode TF-IDF dan algoritma Multinomial Naive Bayes merupakan pendekatan yang efektif, efisien, dan memiliki tingkat akurasi yang tinggi dalam mendeteksi berita hoaks berbahasa Indonesia. Oleh karena itu, model yang dihasilkan berpotensi untuk diterapkan sebagai sistem pendukung dalam proses penyaringan dan deteksi berita hoaks secara otomatis guna membantu menjaga kualitas dan integritas informasi digital di Indonesia.

REFERENCES

- Abdillah., Diah, R., & Nia, K. (2025). *SISTEM DETEKSI HOAKS PADA BERITA ONLINE MENGGUNAKAN METODE*. 2, 449–465.
- Diki, M., & Nurina, B. (2022). *Penerapan Algoritma K-Nearest Neighbor dalam Klasifikasi Judul Berita Hoax*.
- Ferguson., B. B., & Istiono, W. (2024). *Hoax News Detection in Indonesian Political Headlines Using Multinomial Naive Bayes Bertrand*. 5(4), 168–179. <https://doi.org/10.62527/jitsi.5>.
- Kirana, A. S., Roeswidiah, R., & Pudoli, A. (2025). *Analisis Sentimen Pada Media Sosial Terhadap Layanan Samsat Digital Nasional*. 8, 53–63.
- Munna, A., & Zuliarso, E. (2024). Interpretation of Stacking Ensemble model for sentiment analysis of online loan application reviews using LIME. *Aiti*, 21(2), 183–196.
- Prayoga, A. Y., Hadiana, A. I., & Umbara, F. R. (2021). *DETEKSI HOAX PADA BERITA ONLINE BAHASA INGGRIS MENGGUNAKAN BERNOULLI NAÏVE BAYES DENGAN EKSTRAKSI FITUR TF-IDF Agri*. 2(10).
- Salsabila, B., Huda, N., Ayni, F. N., Angelina, D. A., Informatika, T., Teknik, F., Nusantara, U., & Kediri, P. (2026). *Sistem Pendeteksi Berita Hoax Berbasis Word2Vec dan Logistic Regression*. 5.
- Tobing, C. J. L., Wijayakusuma, I. G. N. L., Putu, L., Harini, I., & Udayana, U. (2025). *Detection of Political Hoax News Using Fine-Tuning IndoBERT*. 9(2), 354–360.