JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 5 Tahun 2026
ISSN 3025-0919 (media online)
Hal 1439-1443

Implementasi Metode Naive Bayes untuk Klasifikasi Berita Palsu dan Asli pada Platform Media Online

Alpian Maulana[1], Arjuna Ranma Putra[2], Diski Bima Pradana[3], Firmansyah Surya Kusuma[4], Muhammad Ali Sodikin[5], Rahmawati[6]

[1,2,3,4,5,6]Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia
Email: [1]alpianmaulana1305@gmail.com, [2]arjunaputra1507@gmail.com, [3]diskipradana@gmail.com, [4]firmansyahsuryakusuma@gmail.com, [5]mhmmdali2811@gmail.com, [6]dosen02394@unpam.ac.id

Abstrak−Maraknya penyebaran berita hoaks di ruang digital menimbulkan dampak sosial yang serius bagi khalayak luas. Penelitian ini merancang dan menguji sistem klasifikasi berita berbasis Multinomial Naive Bayes dengan memanfaatkan dataset Fake and Real News yang bersumber dari Kaggle, mencakup 44.919 artikel yang terbagi atas 23.502 berita palsu dan 21.417 berita terverifikasi. Proses pengolahan data meliputi serangkaian tahapan, yaitu penyeragaman huruf (*case folding*), pemecahan teks menjadi token (tokenisasi), eliminasi kata umum (*stopword removal*), dan normalisasi kata ke bentuk dasar (*lemmatization*). Representasi fitur dilakukan melalui pembobotan TF-IDF, sedangkan pembagian data menggunakan rasio 80:20 antara data latih dan data uji. Evaluasi menunjukkan performa model yang tinggi, dengan akurasi 93,52%, presisi 94%, recall 93%, dan F1-Score 94%, membuktikan bahwa pendekatan Naive Bayes layak diterapkan sebagai solusi deteksi berita palsu berbasis kecerdasan buatan.

Kata Kunci: Fake News; Naive Bayes; TF-IDF; Klasifikasi Teks; Natural Language Processing

Abstract−*The growing prevalence of misinformation across digital media platforms poses significant societal challenges. This study develops and evaluates a news classification system using the Multinomial Naive Bayes algorithm, leveraging the Fake and Real News dataset from Kaggle, which comprises 44,919 articles divided into 23,502 fabricated news items and 21,417 verified reports. The data preparation pipeline includes case folding, tokenization, stopword removal, and lemmatization, while TF-IDF weighting is used for feature representation. Data was partitioned at an 80:20 train-test ratio. Experimental outcomes demonstrate strong classifier performance, achieving 93.52% accuracy, 94% precision, 93% recall, and an F1-Score of 94%, confirming that Naive Bayes constitutes an efficient and practical baseline for automated fake news detection.*

Keywords: *Fake News; Naive Bayes; TF-IDF; Text Classification; Natural Language Processing*

1. PENDAHULUAN

Kemajuan teknologi informasi yang pesat telah merevolusi cara masyarakat mengakses dan berbagi informasi melalui berbagai kanal digital. Di balik kemudahan tersebut, muncul permasalahan serius berupa merebaknya konten berita yang tidak terverifikasi atau sengaja direkayasa untuk menyesatkan publik, yang lazim disebut sebagai fake news. Fenomena ini bukan sekadar gangguan informasi, melainkan ancaman nyata bagi stabilitas sosial, proses demokrasi, dan kepercayaan masyarakat terhadap media.

Data dari Reuters Institute Digital News Report 2023 mengungkapkan bahwa lebih dari 60% pengguna internet di seluruh dunia merasa kesulitan membedakan antara berita yang sahih dan yang palsu. Di tingkat nasional, Kementerian Komunikasi dan Informatika mencatat ribuan konten hoaks setiap tahun, dengan lonjakan signifikan terutama pada periode pemilu dan masa pandemi. Kondisi ini mendorong urgensi pengembangan sistem deteksi otomatis yang andal dan efisien.

Berbagai pendekatan komputasional telah dikembangkan untuk mengotomasi deteksi berita palsu, mulai dari Support Vector Machine (SVM), Decision Tree, Random Forest, Logistic Regression, hingga arsitektur deep learning seperti LSTM dan BERT. Meskipun model berbasis jaringan saraf mendalam menawarkan akurasi yang lebih tinggi, kompleksitas dan kebutuhan sumber daya komputasinya yang besar menjadi hambatan tersendiri untuk implementasi skala kecil. Algoritma Naive Bayes, sebaliknya, menawarkan keunggulan berupa komputasi ringan, implementasi sederhana, serta performa kompetitif pada data teks berdimensi tinggi.

Penelitian ini menggunakan dataset Fake and Real News yang dikurasi oleh Clément Bisaillon dan tersedia publik di platform Kaggle. Dataset ini mencakup ribuan artikel berbahasa Inggris yang telah diberi label secara terpercaya sebagai berita palsu maupun berita asli,

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 5 Tahun 2026
ISSN 3025-0919 (media online)
Hal 1439-1443

menjadikannya bahan uji yang ideal untuk keperluan klasifikasi teks.

Tujuan penelitian ini mencakup tiga hal utama: (1) merancang model klasifikasi berita palsu berbasis Multinomial Naive Bayes; (2) mengevaluasi kinerjanya menggunakan metrik akurasi, presisi, recall, dan F1-Score; serta (3) menganalisis kontribusi tahapan preprocessing dan pembobotan TF-IDF terhadap kualitas prediksi model.

2. METODE

2.1 Tinjauan Pustaka

Fake news dapat didefinisikan sebagai informasi yang sengaja direkayasa untuk menipu dan menyesatkan pembaca (Allcott & Gentzkow, 2017). Jenisnya beragam, mulai dari parodi, konten manipulatif, hingga informasi yang benar secara fakta namun disajikan dalam konteks yang keliru. Penelitian (Vosoughi et al., 2018) yang terbit di jurnal *Science* menunjukkan bahwa informasi palsu menyebar enam kali lebih cepat dibandingkan berita faktual di media sosial, menggambarkan betapa pentingnya intervensi teknologi dalam penanggulangan fenomena ini.

Dalam ranah pemrosesan bahasa alami (NLP), preprocessing teks merupakan tahap awal yang krusial untuk memastikan kualitas data sebelum analisis. Tahapan umumnya meliputi: (1) Case Folding, yakni konversi seluruh huruf ke format kecil guna menghindari duplikasi representasi kata; (2) Tokenisasi, pemisahan teks menjadi unit-unit kata; (3) Stopword Removal, penghapusan kata-kata fungsional yang tidak berkontribusi pada makna semantis; dan (4) Lemmatization, penyederhanaan kata ke bentuk dasarnya (lemma) untuk mereduksi variasi morfologis.

TF-IDF (Term Frequency-Inverse Document Frequency) merupakan metode pembobotan fitur teks yang mengkuantifikasi tingkat kepentingan suatu kata dalam sebuah dokumen secara relatif terhadap keseluruhan korpus. Semakin sering sebuah kata muncul dalam dokumen tetapi jarang ditemui di dokumen lain, semakin tinggi bobotnya. Formula TF-IDF dinyatakan sebagai:

$$\mathbf{TF\text{-}IDF(t, d, D) = TF(t, d) \times IDF(t, D)}$$

$$IDF(t, D) = \log(|D| / |\{d \in D : t \in d\}|)$$

Naive Bayes adalah pengklasifikasi probabilistik yang berlandaskan pada teorema Bayes dengan asumsi independensi antarfitur. Untuk klasifikasi dokumen teks, varian Multinomial Naive Bayes digunakan karena sesuai dengan representasi distribusi frekuensi kata. Formulasi teorema Bayes untuk klasifikasi adalah:

$$\mathbf{P(C \mid X) = P(X \mid C) \times P(C) / P(X)}$$

Beberapa studi sebelumnya mengkonfirmasi relevansi pendekatan ini. (Ahmed et al., 2018) melaporkan bahwa Naive Bayes mencapai akurasi 88,33% dalam klasifikasi berita, sementara (Abdullah All Tanvir et al., 2019) berhasil meningkatkan performa hingga 94,1% dengan metode ensemble. (Shu et al., 2017) mengamati bahwa model deep learning unggul dalam akurasi namun memerlukan sumber daya komputasi jauh lebih besar. (Granik & Mesyura, 2017) sendiri membuktikan efektivitas Naive Bayes dengan akurasi 74% pada dataset media sosial berskala kecil dengan preprocessing minimal.

2.2 Dataset

Dataset yang digunakan adalah Fake and Real News Dataset yang dikurasi oleh Clément Bisaillon, tersedia secara terbuka melalui Kaggle. Dataset ini terdiri atas dua berkas CSV: Fake.csv berisi 23.502 artikel berita palsu yang dikumpulkan dari sumber-sumber yang telah diidentifikasi oleh Politifact.com dan Wikipedia, sedangkan True.csv berisi 21.417 artikel berita terverifikasi dari Reuters.com. Setiap entri memuat empat atribut: judul (title), isi artikel (text), kategori berita (subject), dan tanggal publikasi (date). Distribusi kelas dataset disajikan pada Tabel 1.

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 5 Tahun 2026
ISSN 3025-0919 (media online)
Hal 1439-1443

Tabel 1. Distribusi Dataset Fake and Real News

Kelas	Jumlah Artikel	Persentase (%)
Fake	23.502	52,35%
Real	21.417	47,65%
Total	**44.919**	**100%**

2.3 Preprocessing Teks

Seluruh proses pembersihan data dilaksanakan secara berurutan. Pertama, kolom judul dan isi berita digabungkan untuk memperoleh representasi teks yang komprehensif. Selanjutnya dilakukan konversi huruf ke format kecil menggunakan fungsi str.lower() pada Python. Karakter spesial, tanda baca, angka, serta URL kemudian dihapus menggunakan ekspresi reguler (regex). Proses tokenisasi memanfaatkan library NLTK, diikuti penghapusan 179 stopword bahasa Inggris bawaan NLTK. Tahap akhir adalah lemmatization menggunakan WordNetLemmatizer dari NLTK untuk mengonversi setiap kata ke bentuk dasarnya.

2.4 Ekstraksi Fitur dan Pembagian Data

Teks yang telah diproses diubah menjadi vektor numerik menggunakan TF-IDF Vectorizer dari library Scikit-learn, dengan konfigurasi: max_features = 50.000 (membatasi kosakata pada 50.000 term tertinggi), ngram_range = (1, 2) untuk menyertakan unigram dan bigram, serta min_df = 2 untuk mengabaikan term yang muncul hanya pada satu dokumen. Dataset kemudian dibagi dengan rasio 80:20, menghasilkan 35.939 data latih dan 8.980 data uji. Teknik 10-fold cross-validation juga diterapkan untuk memvalidasi stabilitas model.

2.5 Model dan Evaluasi

Model Multinomial Naive Bayes diimplementasikan menggunakan Scikit-learn dengan parameter alpha = 1,0 (Laplace smoothing) guna menangani masalah zero probability. Evaluasi kinerja model dilakukan menggunakan confusion matrix yang menghasilkan empat komponen: True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Metrik yang dihitung meliputi akurasi, presisi, recall, dan F1-Score.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pengujian Model

Pengujian dijalankan terhadap 8.980 sampel data uji menggunakan model Multinomial Naive Bayes yang dilatih pada data yang telah melewati seluruh tahapan preprocessing. Model berhasil mencapai tingkat akurasi keseluruhan sebesar 93,52%, mengindikasikan bahwa lebih dari sembilan dari setiap sepuluh artikel berhasil diklasifikasikan dengan tepat. Rekapitulasi metrik evaluasi disajikan pada Tabel 2.

Tabel 2. Rekapitulasi Hasil Evaluasi Model

Metrik	Nilai
Accuracy	93,52%
Precision	94%
Recall	93%
F1-Score	94%

Perincian kinerja per kelas disajikan dalam classification report pada Tabel 3, yang memperlihatkan keseimbangan performa model dalam mengidentifikasi kedua kategori berita.

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 5 Tahun 2026
ISSN 3025-0919 (media online)
Hal 1439-1443

Tabel 3. Classification Report per Kelas

Kelas	Precision	Recall	F1-Score	Support
Real (0)	0,94	0,93	0,93	4.330
Fake (1)	0,93	0,94	0,94	4.650
Accuracy			**0,94**	**8.980**
Macro Avg	0,94	0,93	0,94	8.980
Weighted Avg	0,94	0,94	0,94	8.980

3.2 Pembahasan

Capaian akurasi 93,52% membuktikan bahwa Multinomial Naive Bayes mampu beroperasi secara efektif sebagai sistem klasifikasi berita palsu. Nilai presisi 94% menunjukkan bahwa artikel yang diprediksi sebagai kelas tertentu memang benar termasuk ke dalam kelas tersebut, dengan tingkat false positive yang terjaga rendah. Recall 93% mengindikasikan kemampuan model dalam mengenali sebagian besar artikel pada masing-masing kelas, sementara F1-Score 94% mencerminkan keseimbangan optimal antara presisi dan recall.

Performa tinggi ini tidak terlepas dari beberapa faktor penentu. Pertama, pipeline preprocessing yang komprehensif — meliputi case folding, pembersihan karakter, stopword removal, dan lemmatization — berperan vital dalam mereduksi noise dan variabilitas teks. Kedua, representasi fitur berbasis TF-IDF mampu memberikan bobot diskriminatif yang tinggi pada kata-kata yang paling relevan untuk membedakan berita palsu dari berita asli. Ketiga, karakteristik Multinomial Naive Bayes yang bekerja berdasarkan distribusi frekuensi kata sangat sesuai dengan nature data teks berdimensi tinggi seperti yang digunakan dalam penelitian ini.

Temuan ini sejalan dengan hasil (Mccallum & Nigam, 1998) yang menyimpulkan bahwa Multinomial Naive Bayes merupakan salah satu algoritma paling efektif untuk klasifikasi dokumen teks. Dibandingkan dengan penelitian (Granik & Mesyura, 2017) yang memperoleh akurasi 74% pada dataset media sosial, penelitian ini mencapai hasil yang lebih baik, yang dapat dikaitkan dengan penggunaan dataset yang lebih besar, pipeline preprocessing lebih lengkap, dan pemanfaatan TF-IDF sebagai representasi fitur. Dengan demikian, kombinasi ketiga komponen — preprocessing, TF-IDF, dan Multinomial Naive Bayes — terbukti menjadi pendekatan yang seimbang antara efisiensi komputasi dan akurasi prediksi.

4. KESIMPULAN

Penelitian ini berhasil merancang dan menguji sistem klasifikasi berita berbasis Multinomial Naive Bayes dengan mengintegrasikan pipeline preprocessing teks yang terstruktur dan pembobotan fitur TF-IDF. Pengujian terhadap 8.980 sampel data uji menghasilkan akurasi 93,52%, presisi 94%, recall 93%, dan F1-Score 94%, menunjukkan bahwa model memiliki kemampuan klasifikasi yang tinggi dan seimbang antar kedua kelas.

Dari temuan ini disimpulkan bahwa Multinomial Naive Bayes merupakan alternatif yang efektif dan efisien untuk membangun sistem deteksi berita palsu, khususnya pada teks berbahasa Inggris berskala besar. Untuk pengembangan ke depan, disarankan agar penelitian berikutnya membandingkan Naive Bayes dengan algoritma lain seperti SVM dan BERT, memperluas cakupan pada dataset berbahasa Indonesia, mengintegrasikan fitur non-tekstual (sumber berita, metadata), serta mengembangkan sistem menjadi aplikasi berbasis web atau mobile yang mampu mendeteksi berita palsu secara real-time.

REFERENCES

Abdullah All Tanvir, Mahir, E. M., Akhter, S., & Huq, M. R. (2019). Detecting Fake News using Machine Learning and Deep Learning Algorithms. *2019 7th International Conference on Smart Computing & Communications (ICSCC)*, 1–5. https://doi.org/10.1109/ICSCC.2019.8843612

Ahmed, H., Traore, I., & Saad, S. (2018). Detecting opinion spams and fake news using text classification. *SECURITY AND PRIVACY*, *1*(1). https://doi.org/10.1002/spy2.9

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 5 Tahun 2026
ISSN 3025-0919 (media online)
Hal 1439-1443

Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, *31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Bisaillon, C. (2020). Fake and Real News Dataset. Kaggle. https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset

Granik, M., & Mesyura, V. (2017). Fake news detection using naive Bayes classifier. *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, 900–903. https://doi.org/10.1109/UKRCON.2017.8100379

Mccallum, A., & Nigam, K. (1998). *A Comparison of Event Models for Naive Bayes Text Classification*. www.aaai.org

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media. *ACM SIGKDD Explorations Newsletter*, *19*(1), 22–36. https://doi.org/10.1145/3137597.3137600

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559