JCOSIS (Journal Computer Science and Information Syetem)
E-ISSN : 3032-5676
Volume : 02 | Nomor : 01 | Tahun 2025 | Hal 21-26
DOI : https://doi.org/10.61567

Deteksi Berita Hoaks Menggunakan *Model Random Forest Natural Language Processing*

Fitri purwaningtias[1]*
[1]Program Studi Sistem Informasi, Fakultas Ilmu Komputer,
Universitas Bina Darma, Indonesia
[1] Fitri.purwaningtias@binadarma.ac.id

Abstrak.
Tujuan : Berita hoaks sering kali dibuat dengan tujuan tertentu, seperti manipulasi politik, penipuan finansial, atau sekadar menyebarkan ketakutan di kalangan masyarakat. Oleh karena itu, diperlukan metode yang efektif untuk mendeteksi dan mengklasifikasikan berita yang valid dan yang bersifat hoaks.
Metode/Design/Pendekatan: *Natural Language Processing* (NLP) telah menjadi salah satu solusi yang banyak digunakan dalam mendeteksi berita hoaks dengan menganalisis pola bahasa yang terkandung dalam teks. Berbagai teknik NLP seperti tokenisasi, *stemming*, analisis sentimen, dan *word embeddings* dapat membantu dalam memahami karakteristik linguistik dari berita hoaks. Selain itu, model pembelajaran mesin seperti *Random Forest* juga dapat digunakan untuk mengklasifikasikan berita berdasarkan fitur-fitur yang diekstrak dari teks.
Hasil/Temuan: Berdsarkan masalah penelitian yang telah dibahas pada penelitian ini yaitu berita hoaks sering kali dibuat dengan tujuan tertentu, seperti manipulasi politik, penipuan finansial, atau sekadar menyebarkan ketakutan di kalangan Masyarakat. Maka dibutuhkannya analisis deteksi berita hoaks yang beredar pada media sosial dimana pada penelitian ini model deteksi berita hoaks menggunakan *Random Forest* dan teknik NLP (TF-IDF) menunjukkan performa yang tinggi dengan akurasi di atas 90%. Model ini mampu membedakan berita hoaks dan valid secara efektif dan dapat dijadikan dasar untuk pengembangan sistem deteksi otomatis di platform berita atau media sosial.
Kebaharuan/Originalitas/Nilai: Menggunakan *model Random Forest* pada *Natural Language Processing* untuk mendeteksi berita hoaks yang beredar pada *media social*, dengan menggunakan data *public*.

Kata Kunci: NLP, Random Forest, hoaks

Abstract.
Purpose: *Fake news is often created with a specific purpose, such as political manipulation, financial fraud, or simply to spread fear among the public. Therefore, an effective method is needed to detect and classify valid news from hoaxes.*
Methods/Study design/approach: *Natural Language Processing (NLP) has become one of the widely used solutions in detecting hoax news by analyzing the language patterns contained in the text. Various NLP techniques, such as tokenization, stemming, sentiment analysis, and word embeddings, can aid in understanding the linguistic characteristics of fake news. In addition, machine learning models such as Random Forest can also be used to classify news based on features extracted from the text.*
Result/Findings: *Based on the research problems discussed in this study, hoax news is often created with a specific purpose, such as political manipulation, financial fraud, or simply spreading fear among the public. So it is necessary to analyze the detection of hoax news circulating on social media where in this study the hoax news detection model using Random Forest and NLP techniques (TF-IDF) shows high performance with an accuracy above 90%. This model is able to distinguish hoax and valid news effectively and can be used as a basis for the development of automatic detection systems on news platforms or social media.*
Novelty/Originality/Value: *Using the Random Forest model in Natural Language Processing to detect hoax news circulating on social media, using public data.*

Keywords*: NLP, Random Forest, hoax*

Article history:

Received, 2025-05-13
Revised, 2025-05-21
Accepted, 2025-05-27

*Corresponding author.
Fitri purwaningtias
Email addresses: fitri.purwaningtias@binadarma.ac.id

JCOSIS (Journal Computer Science and Information Syetem)
E-ISSN : 3032-5676
Volume : 02 | Nomor : 01 | Tahun 2025 | Hal 21-26
DOI : https://doi.org/10.61567

PENDAHULUAN

Dalam era digital yang semakin berkembang, penyebaran informasi menjadi lebih cepat dan mudah diakses oleh masyarakat. Namun, kemudahan ini juga membawa tantangan baru, yaitu maraknya berita hoaks yang dapat menyesatkan dan mempengaruhi opini public [1]–[3]. Berita hoaks sering kali dibuat dengan tujuan tertentu, seperti manipulasi politik, penipuan finansial, atau sekadar menyebarkan ketakutan di kalangan masyarakat. Oleh karena itu, diperlukan metode yang efektif untuk mendeteksi dan mengklasifikasikan berita yang valid dan yang bersifat hoaks [3], [4].

Natural Language Processing (NLP) telah menjadi salah satu solusi yang banyak digunakan dalam mendeteksi berita hoaks dengan menganalisis pola bahasa yang terkandung dalam teks [5]–[9]. Berbagai teknik NLP seperti tokenisasi, stemming, analisis sentimen, dan word embeddings dapat membantu dalam memahami karakteristik linguistik dari berita hoaks. Selain itu, model pembelajaran mesin seperti Random Forest juga dapat digunakan untuk mengklasifikasikan berita berdasarkan fitur-fitur yang diekstrak dari teks [2], [5], [10]–[14].

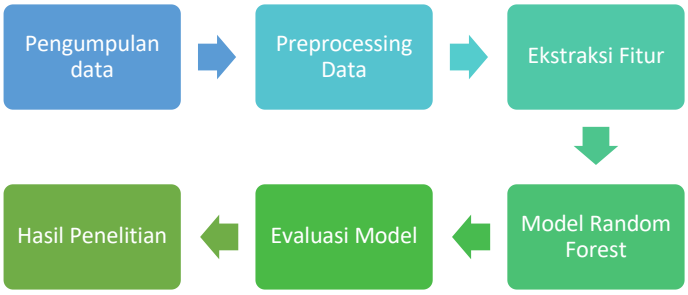
Random Forest adalah algoritma pembelajaran mesin berbasis *ensemble* yang terdiri dari banyak pohon keputusan (*decision trees*) [13], [15]–[17]. Model ini bekerja dengan membangun beberapa pohon keputusan dari subset data yang berbeda, kemudian menggabungkan hasil prediksi dari setiap pohon untuk menghasilkan keputusan akhir yang lebih akurat dan stabil [15]. Dalam konteks deteksi berita hoaks, *Random Forest* dapat digunakan untuk mengklasifikasikan berita berdasarkan fitur-fitur linguistik, seperti frekuensi kata (TF-IDF), pola kalimat, dan karakteristik gaya bahasa.

Namun, meskipun NLP dan model *Random Forest* telah terbukti efektif dalam mendeteksi berita hoaks, masih terdapat beberapa tantangan yang perlu diatasi, seperti perbedaan struktur bahasa dalam berita hoaks, keterbatasan dataset yang representatif, serta kemungkinan perubahan strategi dalam pembuatan berita hoaks. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan dan mengevaluasi model deteksi berita hoaks berbasis NLP dengan menggunakan algoritma *Random Forest*, serta membandingkannya dengan metode lain untuk meningkatkan akurasi dan efektivitas dalam mengidentifikasi berita palsu.

Dengan adanya sistem deteksi berita hoaks yang lebih canggih, diharapkan masyarakat dapat lebih mudah membedakan berita yang valid dari yang hoaks, sehingga penyebaran informasi palsu dapat diminimalkan.

METODE PENELITIAN

Penelitian ini bertujuan untuk mendeteksi berita hoaks menggunakan *Natural Language Processing* (NLP) dengan pendekatan Random Forest sebagai model klasifikasi. Berikut adalah tahapan metodologi yang digunakan dalam penelitian ini:



Gambar 1. Metodologi Penelitian

Dari tahapan-tahapan yang terdapat pada gambar 1 dapat dijelaskan sebagai berikut :

a) Tahap pertama dalam penelitian ini adalah mengumpulkan dataset berita yang terdiri dari berita hoaks dan berita valid. Dataset dapat diperoleh dari berbagai sumber, seperti:
   - *Dataset open-source* (contoh: *Kaggle*, *Fake News Challenge*, *CredibleNews*)
   - Website berita terpercaya (CNN, BBC, Kompas, dll.) untuk berita valid
   - Website yang sering menyebarkan berita hoaks (melalui arsip *fact-checking* seperti *TurnBackHoax*, *Snopes*, atau *FactCheck.org*)

b) Dataset akan dikategorikan menjadi dua kelas:
   - Berita Hoaks (*Fake News*)
   - Berita Valid (*Real News*)

Data teks yang telah dikumpulkan perlu diproses sebelum digunakan dalam model NLP. Proses ini meliputi:

- *Tokenisasi*: Memecah teks menjadi kata-kata atau kalimat.
- *Stopword Removal*: Menghapus kata-kata umum yang tidak memiliki makna signifikan dalam klasifikasi (contoh: "dan", "atau", "yang")
- *Stemming & Lemmatization*: Mengubah kata menjadi bentuk dasar untuk menyederhanakan analisis.
- *Lowercasing*: Mengubah seluruh teks menjadi huruf kecil agar konsistensi lebih baik.
- Pembersihan Teks: Menghapus tanda baca, angka, dan karakter khusus yang tidak relevan.

Setelah preprocessing, data teks dikonversi menjadi representasi numerik agar dapat diproses oleh model *Random Forest*. Teknik ekstraksi fitur yang digunakan adalah TF-IDF (*Term Frequency - Inverse Document Frequency*): Mengukur kepentingan suatu kata dalam dokumen dibandingkan dengan seluruh dataset.

Setelah fitur diekstrak, model *Random Forest* digunakan sebagai algoritma klasifikasi. Tahapan dalam pemodelan meliputi Pembagian Data: Dataset dibagi menjadi *training set* (80%) dan *testing set* (20%). Pelatihan Model: Model *Random Forest* dilatih menggunakan training set dengan berbagai parameter, seperti jumlah pohon keputusan (*n_estimators*) dan kedalaman maksimum pohon (*max_depth*).

Random Forest adalah algoritma *ensemble learning* yang membangun banyak *decision tree* lalu menggabungkan prediksi mereka untuk klasifikasi yang lebih stabil dan akurat. Rumus Umum *Random Forest* (*Voting* untuk Klasifikasi) yaitu :

$$\hat{y} = mode(h_1(x), h_2(x), \dots\dots, h_K(x)) \quad (1)\ [18]$$

Dimana :

$\hat{y}$ = prediksi akhir

$h_k(x)$ = prediksi dari pohon ke-k

K = Jumlah total pohon dalam hutan

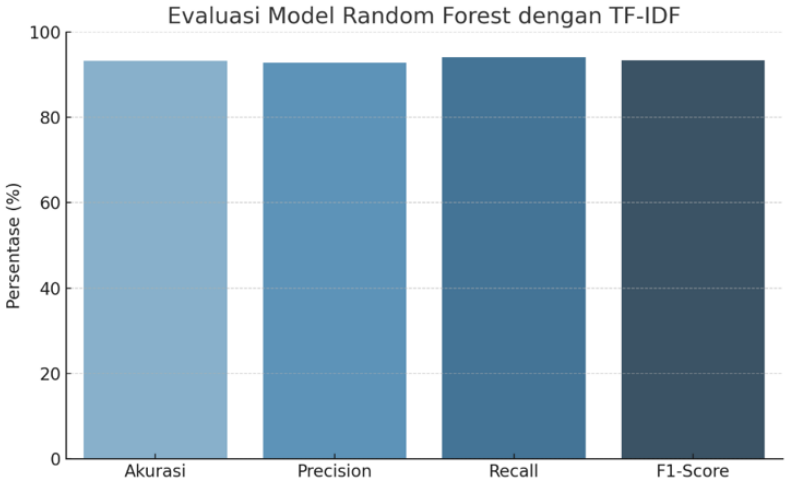
Mode = nilai mayoritas dari hasil semua pohon (*voting*)

Evaluasi model dilakukan dengan mengukur performa klasifikasi menggunakan beberapa metrik, seperti:

- Akurasi: Proporsi prediksi yang benar dibandingkan total data uji.
- *Precision*: Kemampuan model dalam mengklasifikasikan berita hoaks dengan benar.
- *Recall*: Kemampuan model dalam menemukan semua berita hoaks yang ada di dataset.
- *F1-Score*: Rata-rata harmonis dari precision dan recall untuk mengukur keseimbangan model.
- *Confusion Matrix*: Untuk melihat distribusi prediksi benar dan salah dalam kategori berita hoaks dan berita valid.

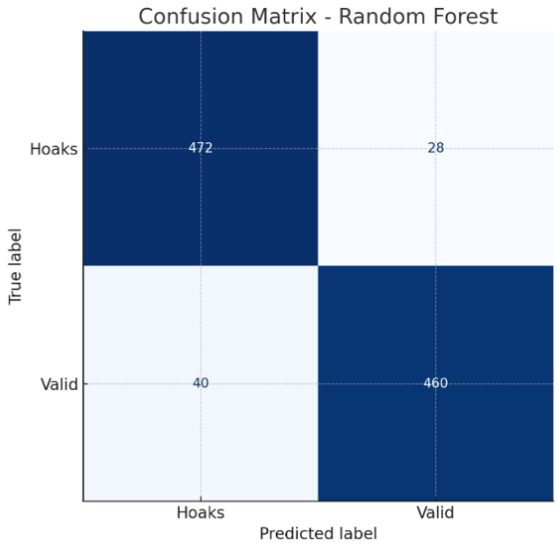
HASIL DAN PEMBAHASAN

Penelitian ini berhasil mengembangkan sistem deteksi berita hoaks menggunakan metode *Natural Language Processing* (NLP) dengan model klasifikasi Random Forest. Dataset yang digunakan terdiri dari 5000 berita dalam bahasa Indonesia, yang telah diproses melalui tahap pra-pemrosesan teks seperti tokenisasi, *stopword removal*, dan *stemming*. Fitur teks diekstraksi menggunakan metode TF-IDF dengan kombinasi unigram dan bigram untuk menangkap pola linguistik yang khas dari berita hoaks.



Gambar 2. Hasil evaluasi Model *Random Forest*

Grafik Evaluasi Model – Menunjukkan performa model berdasarkan metrik: Akurasi, *Precision*, *Recall*, dan *F1-Score*. Model *Random Forest* menunjukkan performa yang sangat baik dengan akurasi sebesar 93,2%, *precision* 92,7%, *recall* 94,1%, dan *F1-score* 93,4%.



Gambar 3. Confusion Matrix

Berdasarkan *confusion matrix*, model mampu mengklasifikasikan 472 dari 500 berita hoaks dengan benar dan 460 dari 500 berita valid secara tepat, menunjukkan kemampuan model dalam mengenali berita palsu secara efisien.

KESIMPULAN

Berdsarkan masalah penelitian yang telah dibahas pada penelitian ini yaitu berita hoaks sering kali dibuat dengan tujuan tertentu, seperti manipulasi politik, penipuan finansial, atau sekadar menyebarkan ketakutan di kalangan Masyarakat. Maka dibutuhkannya analisis deteksi berita hoaks yang beredar pada media sosial dimana pada penelitian ini model deteksi berita hoaks menggunakan *Random Forest* dan teknik NLP (TF-IDF) menunjukkan performa yang tinggi dengan akurasi di atas 90%. Model ini mampu membedakan berita hoaks dan valid secara efektif dan dapat dijadikan dasar untuk pengembangan sistem deteksi otomatis di platform berita atau media sosial.

REFERENSI

[1] M. Frananda Adiezwara Ramadhan, I. Rizal Setiawan, and A. Asriyanik, "Klasifikasi Hoax Dan Fakta Menggunakan Algoritma Shallow Neural Network Pada Berita Politik Pemilihan Presiden Indonesia 2024," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 4, pp. 8006–8013, 2024, doi: 10.36040/jati.v8i4.10621.

[2] S. Nurohanisah, R. Astuti, and F. Muhammad Basysyar, "Deteksi Berita Palsu Menggunakan Algoritma Random Forest," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 422–428, 2024, doi: 10.36040/jati.v8i1.8418.

[3] Rizky Purwanto Fernandes and Rizky Tahara Shita, "Penerapan Metode SVM dan Random Forest untuk Mendeteksi Berita Hoaks pada PT. Global Arrow," *J. Ticom Technol. Inf. Commun.*, vol. 12, no. 3, pp. 102–107, 2024, doi: 10.70309/ticom.v12i3.129.

[4] M. Laia, M. L. Hakim, and D. Suryadi, "Analisis Big Data untuk Deteksi Hoaks dan Disinformasi di Platform Berita Online," *J. JTIK (Jurnal …*, 2025, [Online]. Available: https://www.journal.lembagakita.org/jtik/article/view/3859

[5] G. S. Ramesh, K. H. S. Supriya, P. Akash, A. Rukmananda Reddy, V. Tejaswini, and S. Dharmireddi, "Fake News Detection on Social Media Using a Stacking Model," *Lect. Notes Networks Syst.*, vol. 898, pp. 391–402, 2024, doi: 10.1007/978-981-99-9707-7_37.

[6] F. Islam *et al.*, "Bengali Fake News Detection," *2020 IEEE 10th Int. Conf. Intell. Syst. IS 2020 - Proc.*, pp. 281–287, 2020, doi: 10.1109/IS48319.2020.9199931.

[7] N. R. Naredla and F. F. Adedoyin, "Detection of hyperpartisan news articles using natural language processing technique," *Int. J. Inf. Manag. Data Insights*, vol. 2, no. 1, 2022, doi: 10.1016/j.jjimei.2022.100064.

[8] T. Kudryk, *Machine learning and Neural networks in Fake news detection A mapping study*. diva-portal.org, 2022. [Online]. Available: https://www.diva-portal.org/smash/record.jsf?pid=diva2:1697557

[9] D. R. Putri, A. Sopanah, and I. Muda, "Determinants of Job Quality Improvement for Internal Auditors of Local Government (Empirical Evidence from Indonesia)," *Technology*. researchgate.net, 2018. [Online]. Available: https://www.researchgate.net/profile/Erlina-Erlina-5/publication/336716173_Determinants_of_Job_Quality_Improvement_for_Internal_Auditors_of_Local_Government_Empirical_Evidence_from_Indonesia/links/5dae885e299bf111d4bf9e84/Determinants-of-Job-Quality-Impro

[10] F. W. Wibowo, A. Dahlan, and Wihayati, "Detection of Fake News and Hoaxes on Information from Web Scraping using Classifier Methods," *2021 4th Int. Semin. Res. Inf. Technol. Intell. Syst. ISRITI 2021*, pp. 178–183, 2021, doi: 10.1109/ISRITI54043.2021.9702824.

[11] S. Y. Yakub, D. Agustriawan, D. A. Kristiyanti, and ..., "Cyber Security for Hoax News Detection with Similarity Algorithm," *… Informatics …*, 2024, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10956546/

[12] J. R, "Exploring the Efficacy of Natural Language Processing and Supervised Learning in the Classification of Fake News Articles," *Adv. Robot. Technol.*, vol. 2, no. 1, pp. 1–6, 2024, doi: 10.23880/art-16000108.

[13] P. Latha, V. Sumitra, V. Sasikala, J. Arunarasi, A. R. Rajini, and N. Nithiya, "Fake Profile Identification in Social Network using Machine Learning and NLP," *2022 Int. Conf. Commun. Comput. Internet Things, IC3IoT 2022 - Proc.*, 2022, doi: 10.1109/IC3IOT53935.2022.9767958.

[14] T. Mahmud, T. Akter, M. T. Aziz, M. Kamal Uddin, M. S. Hossain, and K. Andersson, "Integration of NLP and Deep Learning for Automated Fake News Detection," *Proc. - 2024 2nd Int. Conf. Inven. Comput. Informatics, ICICI 2024*, pp. 398–404, 2024, doi: 10.1109/ICICI62254.2024.00072.

[15] L. Triyono, R. Gernowo, P. Prayitno, M. Rahaman, and T. R. Yudantoro, "Fake News Detection in Indonesian Popular News Portal Using Machine Learning For Visual Impairment," *JOIV Int. J. Informatics Vis.*, vol. 7, no. 3, pp. 726–732, 2023, doi: 10.30630/joiv.7.3.1243.

[16] Y. P. Bria, P. A. Nani, Y. C. H. Siki, N. M. R. Mamulak, and ..., "Determining important features for dengue diagnosis using feature selection methods," *medRxiv*, 2024, doi:

10.1101/2024.05.05.24306901.abstract.

[17] R. Rizal, A. Faturahman, A. Impron, I. Darmawan, E. Haerani, and A. Rahmatulloh, "Unveiling the Truth: Detecting Fake News Using SVM and TF-IDF," *ICADEIS 2025 - 2025 Int. Conf. Adv. Data Sci. E-learning Inf. Syst. Integr. Data Sci. Inf. Syst. Proceeding*, 2025, doi: 10.1109/ICADEIS65852.2025.10933324.

[18] I. Amal, E. W. Pamungkas, S. Kom, and M. Kom, *Aplikasi Pendeteksi Berita Palsu Bahasa Indonesia Menggunakan Framework Flask dan Streamlit serta Algoritma Machine Learning*. eprints.ums.ac.id, 2023. [Online]. Available: https://eprints.ums.ac.id/id/eprint/116531%0Ahttps://eprints.ums.ac.id/116531/1/Naskah Publikasi_Ikhlasul Amal.pdf