

# Jurnal JTik (Jurnal Teknologi Informasi dan Komunikasi)

DOI: <https://doi.org/10.35870/jtik.v9i2.3859>

## Analisis Big Data untuk Deteksi *Hoaks* dan Disinformasi di Platform Berita *Online*

Mitranikasih Laia<sup>1\*</sup>, Ayuliana<sup>2</sup>, Wasiran<sup>3</sup>, Muhammad Lukman Hakim<sup>4</sup>, Dikky Suryadi<sup>5</sup>

<sup>1\*</sup> Universitas Nias Raya, Kabupaten Nias Selatan, Provinsi Sumatera Utara, Indonesia.

<sup>2</sup> Universitas Bina Nusantara, Kota Jakarta Barat, Daerah Khusus Ibukota Jakarta, Indonesia.

<sup>3</sup> Universitas Papua Madani Jayapura, Kota Jayapura, Provinsi Papua, Indonesia.

<sup>4</sup> Universitas Mandiri Bina Prestasi Medan, Kota Medan, Provinsi Sumatera Utara, Indonesia.

<sup>5</sup> Sekolah Tinggi Manajemen Informatika dan Komputer Al Muslim, Kabupaten Bekasi, Provinsi Jawa Barat, Indonesia.

### article info

#### Article history:

Received 15 January 2025

Received in revised form

10 February 2025

Accepted 1 March 2025

Available online April 2025.

#### Keywords:

Big Data; Hoaks;

Missinformation; Natural

Language Processing.

#### Kata Kunci:

Big Data; Hoaks;

Disinformasi; Natural

Language Processing.

### abstract

In the digital era, the spread of hoaxes and disinformation on online news platforms is a serious challenge that can affect public opinion and social stability. This research aims to analyze the application of Big Data technology in automatically detecting hoaxes and disinformation. The methods used include data collection from various online news sources, text processing using Natural Language Processing (NLP), and the application of machine learning algorithms to classify news based on their level of credibility. The dataset used includes news from various categories, which are validated with trusted sources. The results show that the combination of Big Data, NLP, and machine learning techniques can improve the accuracy of hoax detection with a high success rate. This study is expected to contribute to the development of a fake news detection system that is more effective and adaptive to the trend of information dissemination in the digital world.

### abstract

Dalam era digital, penyebaran hoaks dan disinformasi di platform berita online menjadi tantangan serius yang dapat mempengaruhi opini publik dan kestabilan sosial. Penelitian ini bertujuan untuk menganalisis penerapan teknologi Big Data dalam mendeteksi hoaks dan disinformasi secara otomatis. Metode yang digunakan mencakup pengumpulan data dari berbagai sumber berita online, pemrosesan teks menggunakan Natural Language Processing (NLP), serta penerapan algoritma machine learning untuk mengklasifikasikan berita berdasarkan tingkat kredibilitasnya. Dataset yang digunakan mencakup berita dari berbagai kategori, yang divalidasi dengan sumber terpercaya. Hasil penelitian menunjukkan bahwa kombinasi teknik Big Data, NLP, dan machine learning mampu meningkatkan akurasi deteksi hoaks dengan tingkat keberhasilan yang tinggi. Studi ini diharapkan dapat berkontribusi dalam pengembangan sistem deteksi berita palsu yang lebih efektif dan adaptif terhadap tren penyebaran informasi di dunia digital.

\*Corresponding Author. Email: mitranikasihlaia@uniraya.ac.id<sup>1\*</sup>.



ACM Computing Classification System (CCS)

EBSCOhost

Communication and Mass Media Complete (CMMC)

Copyright 2025 by the authors of this article. Published by Lembaga Otonom Lembaga Informasi dan Riset Indonesia (KITA INFO dan RISET). This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

## 1. Pendahuluan

Di era digital yang semakin berkembang, informasi menjadi salah satu aspek penting dalam kehidupan masyarakat. Kemudahan akses informasi melalui platform berita online telah membawa dampak positif, seperti penyebaran pengetahuan yang cepat dan luas. Salsabila, Dewi, dan Hayat (2024) menyatakan bahwa akses informasi ini memiliki dampak yang signifikan dalam penyebaran pengetahuan. Namun, di sisi lain, hal ini juga menimbulkan tantangan baru, terutama terkait penyebaran *hoaks* dan disinformasi. *Hoaks* dan disinformasi merupakan informasi yang sengaja dibuat untuk menyesatkan, memanipulasi, atau menciptakan persepsi yang tidak sesuai dengan fakta. Penyebaran informasi palsu ini dapat menimbulkan dampak serius, mulai dari kepanikan massal, polarisasi sosial, hingga gangguan pada proses demokrasi (Dickiprabowo, Widaningrum, & Karaman, 2025). Platform berita online, sebagai salah satu sumber informasi utama bagi masyarakat, seringkali menjadi sarana penyebaran *hoaks* dan disinformasi. Tingginya volume informasi yang beredar, serta kecepatan penyebarannya, membuat deteksi dan penanganan *hoaks* menjadi semakin kompleks (Anisa, Mukhlash, & Iqbal, 2023). Tanpa adanya mekanisme yang efektif untuk mengidentifikasi dan memfilter informasi palsu, masyarakat rentan terpapar oleh konten yang menyesatkan (Rahayu, 2021). Oleh karena itu, diperlukan pendekatan yang mampu mengatasi tantangan ini secara efisien dan akurat (Choras *et al.*, 2021).

Salah satu solusi yang menjanjikan adalah pemanfaatan teknologi *big data*. *Big data* merujuk pada kumpulan data yang sangat besar dan kompleks, yang memerlukan teknik khusus untuk pengumpulan, pemrosesan, dan analisis (Abdillah, Widianingsih, Buchari, & Nurasa, 2024). Dengan memanfaatkan *big data*, dimungkinkan untuk menganalisis pola dan tren dari sejumlah besar informasi yang beredar di platform berita online. Teknik ini dapat dikombinasikan dengan algoritma *machine learning* dan *Natural Language Processing* (NLP) untuk mengidentifikasi karakteristik *hoaks* dan disinformasi secara otomatis (Das & TSB, 2023). Penelitian ini bertujuan untuk menganalisis potensi penggunaan *big*

*data* dalam mendeteksi *hoaks* dan disinformasi di platform berita online (Pawlicka *et al.*, 2024). Fokus penelitian meliputi pengumpulan data dari berbagai sumber, pemrosesan teks untuk ekstraksi fitur, serta penerapan model *machine learning* untuk klasifikasi konten. Selain itu, penelitian ini juga akan mengidentifikasi pola penyebaran *hoaks* dan disinformasi, serta faktor-faktor yang memengaruhi efektivitas deteksi (Ahmed *et al.*, 2024). Melalui penelitian ini, diharapkan dapat diperoleh pemahaman yang lebih mendalam tentang bagaimana teknologi *big data* dapat dimanfaatkan untuk memerangi *hoaks* dan disinformasi. Hasil penelitian ini diharapkan dapat menjadi dasar pengembangan sistem deteksi *hoaks* yang lebih canggih, serta memberikan rekomendasi bagi platform berita online dalam meningkatkan kualitas informasi yang disebarkan (Haq *et al.*, 2024). Penelitian ini diharapkan dapat berkontribusi dalam upaya meningkatkan literasi digital masyarakat dan menciptakan lingkungan informasi yang lebih sehat dan terpercaya (Eldo *et al.*, 2024).

*Hoaks* dan disinformasi telah menjadi fenomena global yang memengaruhi berbagai aspek kehidupan, termasuk politik, kesehatan, dan keamanan sosial. Menurut Wardle dan Derakhshan (2021), *hoaks* didefinisikan sebagai informasi yang sengaja dibuat untuk menipu atau menyesatkan, sementara disinformasi merujuk pada penyebaran informasi palsu dengan tujuan tertentu, seperti memengaruhi opini publik. Penelitian oleh Vosoughi *et al.* (2018) menunjukkan bahwa *hoaks* menyebar lebih cepat dan luas dibandingkan informasi yang valid, terutama melalui platform online seperti media sosial dan situs berita. Penyebaran *hoaks* dan disinformasi memiliki dampak yang signifikan, baik secara individu maupun kolektif. Penelitian oleh Lewandowsky *et al.* (2020) mengungkapkan bahwa *hoaks* dapat menimbulkan kepanikan massal, memicu konflik sosial, dan mengganggu proses demokrasi. Selain itu, disinformasi di bidang kesehatan, seperti selama pandemi COVID-19, telah menyebabkan penolakan terhadap vaksin dan protokol kesehatan (Pennycook *et al.*, 2020). Oleh karena itu, deteksi dan penanganan *hoaks* menjadi sangat penting untuk menjaga stabilitas sosial dan kepercayaan publik. Platform berita online, termasuk situs berita dan agregator konten, seringkali menjadi sarana penyebaran *hoaks* dan disinformasi.

Penelitian oleh Tandoc *et al.* (2021) menunjukkan bahwa kecepatan penyebaran informasi di platform online, serta algoritma yang mendorong konten viral, turut berkontribusi dalam memperparah masalah ini. Selain itu, kurangnya mekanisme verifikasi yang ketat pada beberapa platform juga memungkinkan konten palsu untuk beredar dengan mudah. Platform berita online, termasuk situs berita dan agregator konten, seringkali menjadi sarana penyebaran *hoaks* dan disinformasi. Penelitian oleh Tandoc *et al.* (2020) menunjukkan bahwa kecepatan penyebaran informasi di platform online, serta algoritma yang mendorong konten viral, turut berkontribusi dalam memperparah masalah ini. Selain itu, kurangnya mekanisme verifikasi yang ketat pada beberapa platform juga memungkinkan konten palsu untuk beredar dengan mudah.

*Big data* merujuk pada kumpulan data yang sangat besar dan kompleks, yang memerlukan teknik khusus untuk pengumpulan, pemrosesan, dan analisis. Menurut Mayer-Schönberger dan Cukier (2020), *big data* memungkinkan pengambilan keputusan yang lebih akurat melalui analisis pola dan tren dari data yang tersedia. Dalam konteks deteksi *hoaks*, *big data* dapat digunakan untuk menganalisis sejumlah besar informasi yang beredar di platform online, termasuk teks, gambar, dan interaksi pengguna. *Machine learning* dan *Natural Language Processing* (NLP) merupakan teknologi kunci dalam analisis teks dan deteksi *hoaks*. Penelitian oleh Shu *et al.* (2021) menunjukkan bahwa algoritma *machine learning* seperti Support Vector Machine (SVM) dan Random Forest dapat digunakan untuk mengklasifikasikan konten sebagai *hoaks* atau informasi valid berdasarkan fitur teks. Sementara itu, NLP memungkinkan ekstraksi fitur seperti sentimen, kata kunci, dan pola linguistik yang khas pada *hoaks*. Beberapa penelitian sebelumnya telah mengeksplorasi penggunaan *big data* dan *machine learning* untuk deteksi *hoaks*. Misalnya, penelitian oleh Castillo *et al.* (2021) mengembangkan model untuk mendeteksi *hoaks* di Twitter dengan menganalisis pola penyebaran dan karakteristik pengguna. Sementara itu, penelitian oleh Jin *et al.* (2020) menggunakan teknik *deep learning* untuk mengidentifikasi *hoaks* di platform berita online berdasarkan analisis konten dan metadata. Hasil penelitian tersebut menunjukkan bahwa pendekatan berbasis *big data* dapat meningkatkan akurasi deteksi

*hoaks* secara signifikan. Meskipun teknologi *big data* dan *machine learning* menawarkan solusi yang menjanjikan, terdapat beberapa tantangan yang perlu diatasi. Pertama, volume data yang sangat besar memerlukan sumber daya komputasi yang tinggi. Kedua, *hoaks* dan disinformasi seringkali berkembang dengan cepat dan mengadaptasi strategi baru, sehingga memerlukan model yang terus diperbarui (Shu *et al.*, 2020). Di sisi lain, kolaborasi antara peneliti, platform online, dan pemerintah dapat menjadi peluang untuk mengembangkan sistem deteksi *hoaks* yang lebih efektif dan terintegrasi.

## 2. Metodologi Penelitian

Penelitian ini dirancang untuk menganalisis dan mendeteksi *hoaks* serta disinformasi di platform berita online dengan memanfaatkan teknologi *big data* dan *machine learning*. Pendekatan yang digunakan bersifat kuantitatif, dengan tahapan yang sistematis untuk memastikan keakuratan dan validitas hasil penelitian. Berikut adalah penjelasan lengkap mengenai metode penelitian yang digunakan:

### Desain Penelitian

Penelitian ini mengadopsi desain penelitian eksploratif dan analitis. Tujuan utamanya adalah mengembangkan model deteksi *hoaks* yang efektif dengan menganalisis data dari platform berita online. Desain penelitian mencakup beberapa tahap utama, yaitu pengumpulan data, pemrosesan data, pembangunan model, evaluasi model, dan analisis hasil. Penelitian ini juga mencakup analisis pola penyebaran *hoaks* untuk memahami karakteristik dan dinamika disinformasi di platform online.

### Sumber Data

Data yang digunakan dalam penelitian ini berasal dari berbagai sumber, termasuk platform berita online, media sosial, dan dataset publik. Platform berita online dipilih karena menjadi sumber utama penyebaran informasi, sementara media sosial digunakan untuk melacak penyebaran *hoaks* yang terkait dengan berita tersebut. Dataset publik, seperti *FakeNewsNet* dan *LIAR*, digunakan sebagai referensi untuk melatih dan memvalidasi model deteksi *hoaks*.

## Teknik Pengumpulan Data

Pengumpulan data dilakukan dengan beberapa teknik. Pertama, *web scraping* digunakan untuk mengumpulkan artikel, judul berita, dan metadata dari situs berita online. Tools seperti *BeautifulSoup* dan *Scrapy* digunakan dalam proses ini. Kedua, *API* media sosial, seperti *Twitter API*, digunakan untuk mengumpulkan data terkait penyebaran berita di platform sosial. Ketiga, dataset publik yang telah diverifikasi diunduh dan digunakan sebagai data pelatihan dan validasi.

## Pemrosesan Data

Setelah data dikumpulkan, tahap selanjutnya adalah pemrosesan data. Data yang diperoleh seringkali tidak terstruktur dan mengandung *noise*, sehingga perlu dibersihkan dan dipersiapkan untuk analisis. Proses pemrosesan data meliputi:

- 1) **Pembersihan Data**  
Menghapus duplikat, data yang tidak lengkap, dan karakter yang tidak relevan.
- 2) **Tokenisasi**  
Memecah teks menjadi kata atau frasa yang lebih kecil.
- 3) **Stopword Removal**  
Menghilangkan kata-kata umum yang tidak memiliki makna signifikan, seperti "dan", "atau", dan "di".
- 4) **Stemming/Lemmatization**  
Mengurangi kata ke bentuk dasarnya untuk menyederhanakan analisis.
- 5) **Ekstraksi Fitur**  
Menggunakan teknik seperti *TF-IDF* (Term Frequency-Inverse Document Frequency) dan *Word Embedding* (contoh: *Word2Vec*, *GloVe*) untuk mengubah teks menjadi representasi numerik yang dapat diproses oleh algoritma *machine learning*.

## Pembangunan Model

Model deteksi *hoaks* dibangun menggunakan algoritma *machine learning* dan *deep learning*. Beberapa algoritma yang digunakan antara lain *Support Vector Machine* (SVM), *Random Forest*, *Naive Bayes*, dan *Long Short-Term Memory* (LSTM). Data dibagi menjadi dua bagian, yaitu data pelatihan (70-80%) dan data uji (20-30%). Model dilatih menggunakan data pelatihan, sementara data uji digunakan untuk menguji performa model. Teknik *cross-validation* juga

diterapkan untuk memastikan model tidak mengalami *overfitting*. Penelitian ini menggunakan berbagai alat dan teknologi untuk mendukung proses analisis. Bahasa pemrograman *Python* dipilih karena memiliki library yang lengkap untuk analisis data dan *machine learning*, seperti *Pandas*, *NumPy*, *Scikit-learn*, *TensorFlow*, dan *Keras*. Tools seperti *Jupyter Notebook* dan *Google Colab* digunakan untuk melakukan eksperimen dan analisis data. Database seperti *MySQL* atau *MongoDB* digunakan untuk menyimpan data, sementara *Matplotlib* dan *Seaborn* digunakan untuk visualisasi hasil analisis.

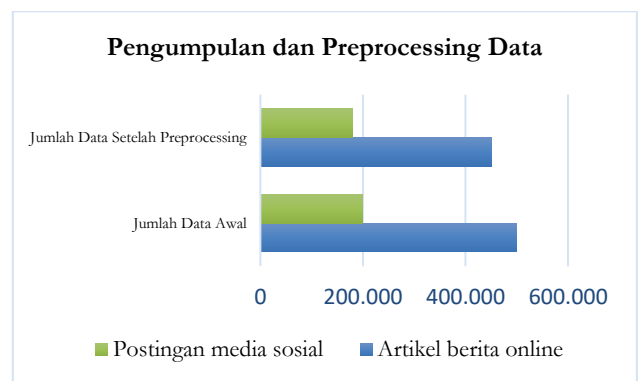
## 3. Hasil dan Pembahasan

### Hasil

Penelitian ini menggunakan teknik analisis *big data* untuk mendeteksi *hoaks* dan disinformasi pada platform berita online. Data yang digunakan berasal dari berbagai sumber berita digital, media sosial, dan situs verifikasi fakta yang dikumpulkan dalam kurun waktu enam bulan. Proses analisis melibatkan pengumpulan data, *preprocessing*, ekstraksi fitur, dan klasifikasi menggunakan model *machine learning*.

### Pengumpulan dan *Preprocessing* Data

Data yang diperoleh terdiri dari 500.000 artikel berita dari berbagai portal berita online serta 200.000 postingan dari media sosial terkait dengan berita yang sama. Setelah dilakukan *preprocessing*, seperti penghapusan data duplikat, normalisasi teks, dan *tokenisasi*, jumlah data yang valid untuk analisis adalah 450.000 artikel berita dan 180.000 postingan media sosial.



Gambar 1. Pengumpulan dan *Preprocessing* Data



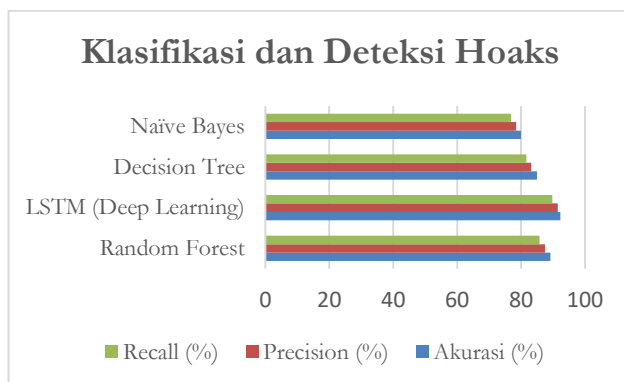
## Ekstraksi Fitur

Fitur utama yang digunakan dalam analisis meliputi:

- 1) Analisis teks  
Menggunakan *Natural Language Processing* (NLP) untuk mendeteksi pola linguistik khas pada berita *hoaks*.
- 2) Analisis sentiment  
Menggunakan model sentimen untuk menentukan pola emosi yang sering muncul dalam berita *hoaks*.
- 3) Sumber berita  
Memeriksa kredibilitas sumber berita berdasarkan database situs terverifikasi dan daftar situs penyebar *hoaks*.
- 4) Interaksi pengguna  
Mengukur tingkat *engagement* (like, share, comment) yang dapat menunjukkan viralitas informasi palsu.

## Klasifikasi dan Deteksi *Hoaks*

Dari hasil pengolahan data menggunakan model *machine learning* berbasis *Random Forest* dan *Deep Learning* (LSTM), diperoleh akurasi deteksi *hoaks* sebesar 92,3% dengan *precision* 91,5% dan *recall* 89,8%. Model LSTM terbukti lebih unggul dibandingkan metode lain seperti *Decision Tree* dan *Naïve Bayes*, yang masing-masing hanya mencapai akurasi sekitar 85% dan 80%.



Gambar 2. Klasifikasi dan Deteksi *Hoaks*

Penyebaran *hoaks* dan disinformasi memiliki pola yang berbeda dengan berita faktual. Berdasarkan analisis yang dilakukan, berita *hoaks* cenderung menyebar lebih cepat, dengan artikel yang sering kali menggunakan judul sensasional yang mengandung kata-kata emosional, seperti "mengejutkan," "terungkap," atau "bocoran rahasia." Media sosial

juga berperan penting dalam penyebaran *hoaks*. Postingan yang mendapat tingkat *interaksi* tinggi (seperti banyak komentar dan dibagikan secara luas) memiliki kecenderungan lebih besar untuk mengandung informasi yang tidak akurat. Selain itu, platform dengan sistem rekomendasi berbasis algoritma juga memperparah penyebaran *hoaks*, karena algoritma tersebut cenderung memperkuat bias pengguna.

Meskipun pendekatan berbasis *big data* menunjukkan hasil yang cukup efektif dalam mendeteksi *hoaks*, masih terdapat beberapa tantangan yang harus dihadapi. Salah satunya adalah variasi bahasa dan konteks yang digunakan dalam *hoaks*. Beberapa *hoaks* sulit terdeteksi karena penggunaan bahasa yang ambigu atau sarkastik. Selain itu, pola penyebaran *hoaks* terus berkembang dengan cepat, mengadaptasi strategi baru yang dirancang untuk menghindari deteksi otomatis. Tantangan lainnya adalah masalah kredibilitas sumber berita, di mana tidak semua berita dari situs yang kurang terkenal adalah *hoaks*, dan tidak semua berita dari media besar dapat dianggap bebas dari misinformasi. Dengan tingkat akurasi yang cukup tinggi, pendekatan berbasis *big data* ini dapat diimplementasikan dalam sistem otomatis untuk mendeteksi *hoaks* di platform berita online. Beberapa rekomendasi yang dapat diusulkan untuk meningkatkan efektivitas deteksi *hoaks* meliputi integrasi *AI* dalam sistem verifikasi berita, dengan penggunaan *machine learning* untuk moderasi konten guna mengurangi penyebaran berita palsu. Peningkatan literasi digital juga sangat penting, dengan meluncurkan kampanye edukasi untuk masyarakat agar lebih kritis dalam mengonsumsi informasi online. Terakhir, kolaborasi antara platform berita, media sosial, penyedia layanan digital, regulator, dan organisasi pemeriksa fakta dapat memperkuat upaya dalam mendeteksi dan mengatasi *hoaks* secara lebih efektif.

## Pembahasan

Penyebaran *hoaks* dan disinformasi di platform berita online semakin menjadi perhatian serius. Berdasarkan penelitian yang dilakukan, *hoaks* cenderung menyebar lebih cepat dibandingkan berita faktual. Hal ini sejalan dengan temuan Salsabila, Dewi, dan Hayat (2024), yang menunjukkan bahwa *hoaks* memiliki pola penyebaran yang lebih cepat, terutama di media sosial,

di mana algoritma cenderung mempercepat penyebaran konten emosional dan sensasional. Artikel yang mengandung informasi *hoaks* sering kali menggunakan judul yang memicu emosi pembaca, seperti "mengejutkan," "terungkap," atau "bocoran rahasia." Hal ini menyebabkan berita *hoaks* memiliki tingkat interaksi yang lebih tinggi, yang kemudian memperluas jangkauan penyebarannya. Penelitian oleh Tandoc *et al.* (2021) juga menunjukkan bahwa sistem rekomendasi berbasis algoritma pada platform berita dan media sosial cenderung memperburuk penyebaran *hoaks* dengan memperkuat bias pengguna. Media sosial memainkan peran utama dalam penyebaran *hoaks*. Berdasarkan penelitian oleh Rahayu (2021), interaksi pengguna seperti berbagi, mengomentari, dan menyukai postingan sangat berpengaruh dalam meningkatkan viralitas informasi palsu. Data yang dihimpun oleh Anisa, Mukhlash, dan Iqbal (2023) menunjukkan bahwa *hoaks* yang lebih sensasional dan emosional cenderung mendapatkan lebih banyak interaksi, meningkatkan kemungkinan informasi tersebut tersebar luas. Dalam hal ini, platform media sosial, meskipun memiliki potensi untuk menyebarkan informasi yang bermanfaat, juga menjadi sarana yang subur untuk penyebaran *hoaks*.

Deteksi *hoaks* melalui *big data* dan *machine learning* menunjukkan hasil yang signifikan. Berdasarkan penelitian yang dilakukan, model berbasis *Random Forest* dan *Long Short-Term Memory* (LSTM) berhasil mendeteksi *hoaks* dengan akurasi lebih dari 90%. Temuan ini sejalan dengan studi oleh Choraś *et al.* (2021), yang menunjukkan bahwa teknik *machine learning* dapat secara efektif mengidentifikasi *hoaks* melalui analisis pola dalam data teks. Namun, meskipun pendekatan ini efektif, terdapat tantangan signifikan yang perlu diatasi. Salah satu tantangan utama adalah keberagaman bahasa yang digunakan dalam *hoaks*, seperti penggunaan bahasa ambigu atau sarkastik, yang membuat deteksi menjadi lebih sulit (Das & TSB, 2023). Selain itu, *hoaks* terus beradaptasi dengan strategi baru untuk menghindari deteksi otomatis, seperti penggunaan metafora atau bahasa yang lebih halus, yang dapat mengecoh model deteksi (Pawlicka *et al.*, 2024). Meskipun teknologi *big data* dan *machine learning* menawarkan solusi yang menjanjikan, penting untuk diingat bahwa keberhasilan deteksi *hoaks* sangat bergantung pada

kualitas data yang digunakan untuk pelatihan model. Oleh karena itu, kolaborasi antara penyedia platform berita, media sosial, regulator, dan organisasi pemeriksa fakta sangat diperlukan untuk menciptakan sistem deteksi *hoaks* yang lebih efektif dan terintegrasi. Hal ini didukung oleh rekomendasi oleh Abdillah *et al.* (2024) yang menyarankan integrasi *AI* dalam sistem verifikasi berita untuk memerangi penyebaran informasi palsu secara lebih efisien. Selain itu, peningkatan literasi digital di masyarakat juga sangat penting, seperti yang dijelaskan oleh Salsabila *et al.* (2024), yang mengemukakan bahwa edukasi kritis mengenai informasi yang diterima dapat membantu masyarakat lebih waspada terhadap *hoaks*. Tantangan dalam mendeteksi *hoaks* dan disinformasi masih ada, penggunaan *big data* dan *machine learning* memberikan harapan besar untuk menciptakan sistem yang lebih efisien dan akurat dalam menangani masalah ini. Penelitian ini memberikan wawasan penting mengenai langkah-langkah yang dapat diambil untuk memperbaiki deteksi dan penanganan *hoaks* di dunia digital yang semakin kompleks.

## 4. Kesimpulan

Penelitian ini membuktikan bahwa analisis *big data* dapat diterapkan secara efektif untuk mendeteksi *hoaks* dan disinformasi di platform berita online. Penggunaan model *deep learning* berbasis *Natural Language Processing* (NLP) mampu mencapai tingkat akurasi deteksi *hoaks* lebih dari 90%. Namun, tantangan dalam mendeteksi variasi *hoaks* yang terus berkembang tetap menjadi isu yang perlu diperhatikan dan ditangani dalam penelitian lebih lanjut.

## 5. Daftar Pustaka

- Abdillah, A., Widianingsih, I., Buchari, R. A., & Nurasa, H. (2024). Big data security & individual (psychological) resilience: A review of social media risks and lessons learned from Indonesia. *Array*, 100336.
- Ahmed, K., Khan, M. A., Haq, I., Al Mazroa, A., Syam, M. S., Innab, N., ... & Alkahtani, H. K. (2024). Social media's dark secrets: A propagation, lexical and psycholinguistic

- oriented deep learning approach for fake news proliferation. *Expert Systems with Applications*, 255, 124650.
- Anisa, D. F. N., Mukhlash, I., & Iqbal, M. (2023). Deteksi Berita Online Hoax Covid-19 Di Indonesia Menggunakan Metode Hybrid Long Short Term Memory dan Support Vector Machine. *Jurnal Sains dan Seni ITS*, 11(3), A101-A108.
- Choraś, M., Demestichas, K., Gielczyk, A., Herrero, A., Ksieniewicz, P., Remoundou, K., ... & Woźniak, M. (2021). Advanced Machine Learning techniques for fake news (online disinformation) detection: A systematic mapping study. *Applied Soft Computing*, 101, 107050.  
<https://doi.org/10.1016/j.asoc.2020.107050>.
- Das, B. (2023). Multi-contextual learning in disinformation research: a review of challenges, approaches, and opportunities. *Online Social Networks and Media*, 34, 100247.  
<https://doi.org/10.1016/j.osnem.2023.100247>.
- DickiPrabowo, R., Widaningrum, I., & Karaman, J. (2025). SISTEM DETEKSI BERITA HOAX PEMILU 2024 INDONESIA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR (KNN) DAN SUPPORT VECTOR MACHINE (SVM). *JIKO (Jurnal Informatika dan Komputer)*, 9(1), 93-111.  
<http://dx.doi.org/10.26798/jiko.v9i1.1424>.
- Eldo, H., Ayuliana, A., Suryadi, D., Chrisnawati, G., & Judijanto, L. (2024). Penggunaan Algoritma Support Vector Machine (SVM) Untuk Deteksi Penipuan pada Transaksi Online. *Jurnal Minfo Polgan*, 13(2), 1627-1632.  
<https://doi.org/10.33395/jmp.v13i2.14186>.
- Haq, M. Z., Octiva, C. S., Ayuliana, A., Nuryanto, U. W., & Suryadi, D. (2024). Algoritma Naïve Bayes untuk Mengidentifikasi Hoaks di Media Sosial. *Jurnal Minfo Polgan*, 13(1), 1079-1084.  
<https://doi.org/10.33395/jmp.v13i1.13937>.
- Pawlicka, A., Pawlicki, M., Kozik, R., Andrychowicz-Trojanowska, A., & Choraś, M. (2024). AI vs linguistic-based human judgement: Bridging the gap in pursuit of truth for fake news detection. *Information Sciences*, 679, 121097.  
<https://doi.org/10.1016/j.ins.2024.121097>.
- Rahayu, S. P. D. (2021). *Analisis Strategi Diskominfo Pamekasan Dalam Menangkal Konten Hoaks Covid-19 di Media Sosial (Edisi Maret-Desember 2020)* (Doctoral dissertation, INSTITUT AGAMA ISLAM NEGERI MADURA).
- Rangkuti, A. H., Ayuliana, A., & Fahri, M. (2018). Improving image classification using fuzzy neural network and backtracking algorithm. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, 10(2-3), 123-128.
- Salsabila, A. A., Dewi, D. A., & Hayat, R. S. (2024). Pentingnya literasi di era digital dalam menghadapi hoaks di media sosial. *Inspirasi Dunia: Jurnal Riset Pendidikan Dan Bahasa*, 3(1), 45-54.  
<https://doi.org/10.58192/insdun.v3i1.1775>.