

Systematic Literature Review (SLR): Pendekatan Natural Language Processing Dengan SVM Dan Naïve Bayes Untuk Deteksi Berita Palsu Bahasa Indonesia

Latar Belakang: Penyebaran berita palsu (hoaks) telah menjadi tantangan global. Pendekatan manual tidak lagi memadai untuk menangani volume informasi yang masif, sehingga deteksi otomatis menggunakan Machine Learning (ML) dan Natural Language Processing (NLP) menjadi kritis.

Tujuan SLR ini adalah menganalisis penelitian terbaru dalam deteksi berita palsu melalui lensa **Analisis SWOT** dan **Gap Analysis** untuk mengidentifikasi peluang pengembangan penelitian berbahasa Indonesia.

1. Analisis SWOT Terintegrasi

Faktor Internal	Faktor Eksternal
Kekuatan (Strengths): <ul style="list-style-type: none">• Fokus spesifik pada bahasa Indonesia (Studi 5)• Metode SVM dan Naïve Bayes terbukti efektif untuk klasifikasi teks• Pemanfaatan dataset lokal (turnbackhoax.id) meningkatkan relevansi kontekstual• Pencapaian metrik tinggi (precision 0.91, recall 1.00) pada studi bahasa Indonesia	Peluang (Opportunities): <ul style="list-style-type: none">• Potensi pengembangan model deep learning untuk bahasa Indonesia (Studi 2)• Integrasi analisis jejaring sosial dengan NLP (Studi 3)• Aplikasi pada situasi kritis seperti pandemi (Studi 4)• Pengembangan dashboard interaktif untuk visualisasi hasil
Kelemahan (Weaknesses): <ul style="list-style-type: none">• Ketergantungan pada dataset terbatas (Studi 5)• Model tradisional (SVM, NB) belum mengeksplorasi deep learning• Generalisasi terbatas antar domain dan topik• Kurangnya analisis bias dataset	Ancaman (Threats): <ul style="list-style-type: none">• Evolusi teknik hoaks berbasis AI generatif (Studi 3)• Dinamika media sosial yang cepat• Risiko kesalahan klasifikasi pada sistem otomatis• Kompleksitas linguistik bahasa Indonesia

2. Gap Analysis Berdasarkan Artikel

No	Judul Artikel	Summary Temuan Utama	Gap yang Diidentifikasi	Keterkaitan dengan Proposal Penelitian

1	Content-Based Fake News Detection With Machine and Deep Learning: a Systematic Review (https://doi.org/10.1016/j.neucom.2023.02.005)	Review pendekatan berbasis konten dan hybrid, menekankan keefektifan NLP	Kesenjangan Bahasa: Fokus utama pada bahasa Inggris	Menutup Gap: Penelitian fokus pada bahasa Indonesia sebagai kontribusi utama
2	Detection of fake news and hate speech for Ethiopian languages: a systematic review of the approaches (https://doi.org/10.1186/s40537-022-00619-x)	Soroti kurangnya penelitian untuk bahasa low-resource, rekomendasikan DL	Sumber Daya Bahasa Lokal: Kebutuhan model untuk bahasa non-Inggris	Relevansi Tinggi: Menjawab kebutuhan penelitian bahasa Indonesia sebagai "medium-resource language"
3	Systematic Review of Fake News, Propaganda, and Disinformation: Examining Authors, Content, and Social Impact Through Machine Learning (https://doi.org/10.1109/ACCESS.2025.3530688)	Analisis tiga domain: pembuat, konten, dan dampak sosial	Integrasi Multi-Domain: Perluanya gabungan analisis konten dan konteks sosial	Arah Pengembangan: RQ5 menyentuh aspek pengguna sebagai langkah awal integrasi perspektif sosial
4	Applications of machine learning for COVID-19 misinformation: a systematic review (https://doi.org/10.1007/s13278-022-00921-9)	Deteksi misinformasi COVID-19 dengan berbagai teknik ML	Spesialisasi Topik: Kebutuhan adaptasi model untuk topik spesifik	Potensi Ekspansi: Model dapat dikembangkan untuk topik spesifik Indonesia

5	An Indonesian Hoax News Detection System Using Reader Feedback and Naïve Bayes Algorithm (https://doi.org/10.2478/cait-2020-0006)	Sistem deteksi bahasa Indonesia dengan Naïve Bayes dan reader feedback	Model Tradisional vs Modern: Kesenjangan antara model sederhana dan pendekatan terkini	Peningkatan Metodologi: Eksplorasi SVM sebagai comparative method dan potensi DL
----------	--	--	---	---

3. Sintesis dan Rekomendasi untuk Penelitian

Berdasarkan analisis di atas, berikut rekomendasi untuk penelitian yang diusulkan:

1. Penargetan Gap Utama:

- Penelitian secara langsung menutup **kesenjangan bahasa Indonesia** yang diidentifikasi dalam Studi 1 dan 2
- **RQ5** tentang metrik evaluasi dari perspektif pengguna menjawab kebutuhan praktis yang belum dieksplorasi studi sebelumnya

2. Pengembangan Metodologi:

- Kombinasi **SVM** dan **Naïve Bayes** memberikan baseline yang kuat
- **Mixed approach** memungkinkan eksplorasi berbagai teknik NLP
- Rekomendasi: Pertimbangkan **word embeddings** dan feature engineering lanjutan untuk meningkatkan performa

3. Implementasi Dashboard:

- **RQ5** tentang KPI metrik untuk dashboard merupakan kontribusi orisinal
- Rekomendasi: Fokus pada **F1-score** sebagai balance antara precision dan recall untuk konteks deteksi berita palsu

4. Ekspansi Masa Depan:

- Integrasi dengan **analisis jejaring sosial** (mengacu Studi 3)
- Eksplorasi **deep learning** untuk peningkatan akurasi (mengacu Studi 2)
- Pengembangan model adaptif untuk **topik spesifik** Indonesia (mengacu Studi 4)

4. Kesimpulan

Kelima artikel review membangun argumen kuat bahwa penelitian deteksi berita palsu perlu:

1. **Lokal kontekstual** - bahasa dan isu spesifik
2. **Pendekatan hybrid** - kombinasi berbagai teknik
3. **Evaluasi komprehensif** - metrik yang relevan dengan kebutuhan pengguna

Penelitian yang diusulkan secara tepat menargetkan gap bahasa Indonesia dengan metodologi sistematis (PICO, RQ), sekaligus membuka jalan untuk pengembangan yang lebih advanced di masa depan melalui pendekatan mixed methods dan fokus pada implementasi dashboard.