

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi dan komunikasi telah mengubah cara masyarakat mengakses dan membagikan berita. Platform media sosial dan situs berita online menjadi sumber utama bagi banyak orang untuk mendapatkan informasi dengan cepat dan mudah. Namun, pesatnya perkembangan ini juga membawa dampak negatif, yaitu maraknya penyebaran berita hoaks. Berita hoaks yang tidak terverifikasi kebenarannya dapat menyebabkan keresahan, menyebarkan disinformasi, dan memengaruhi opini publik secara signifikan. Menurut Kominfo (2023), lebih dari 12.000 konten hoaks terdeteksi dalam setahun, mayoritas di platform berita dan media sosial [1]. Hal ini menjadi tantangan besar, khususnya di Indonesia, yang memiliki pengguna internet aktif yang sangat besar.

Dalam konteks ini, masalah yang dihadapi adalah kesulitan dalam membedakan antara berita yang sahih dan yang tidak dapat dipertanggungjawabkan kebenarannya. Meskipun berbagai upaya telah dilakukan untuk menangani masalah, seperti pemblokiran situs hoaks dan penyuluhan kepada masyarakat, penyebaran berita hoaks tetap sulit dikendalikan. Penyebaran yang cepat melalui media sosial dan kurangnya kesadaran masyarakat menjadi faktor utama yang memperburuk keadaan. Oleh karena itu, deteksi otomatis terhadap berita hoaks menjadi hal yang sangat penting untuk diterapkan.

Penelitian ini bertujuan untuk mengembangkan sistem deteksi berita hoaks dengan menggunakan algoritma pembelajaran mesin (*machine learning*). Sistem diharapkan dapat membantu meningkatkan keakuratan dalam memverifikasi berita yang beredar di media sosial dan situs berita online. Beberapa algoritma yang digunakan dalam penelitian ini, seperti *Support Vector Machine* (SVM) dan *XGBoost*, telah terbukti efektif dalam tugas-tugas klasifikasi teks dan dapat digunakan untuk membedakan antara berita hoaks dan berita yang sahih. Melalui pendekatan ini, diharapkan dapat ditemukan solusi yang lebih efisien untuk

memerangi penyebaran hoaks di Indonesia.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, rumusan masalah yang akan dibahas dalam penelitian ini adalah sebagai berikut:

1. Bagaimana efektivitas penggunaan algoritma pembelajaran mesin dalam mendeteksi berita hoaks pada media sosial dan situs berita online Indonesia?
2. Seberapa besar pengaruh teknik preprocessing data, seperti tokenisasi dan penghapusan stopword, terhadap akurasi sistem deteksi hoaks menggunakan algoritma *machine learning*?
3. Algoritma pembelajaran mesin mana yang memiliki tingkat akurasi tertinggi dalam klasifikasi berita hoaks dan berita sah di Indonesia?
4. Seberapa besar pengaruh tuning hyperparameter menggunakan *GridSearchCV* terhadap peningkatan akurasi dalam deteksi berita hoaks menggunakan algoritma *machine learning*?

1.3 Batasan Masalah

Dalam penelitian ini, terdapat beberapa batasan masalah yang perlu diperhatikan agar fokus penelitian tetap terjaga dan tidak meluas ke hal-hal yang tidak relevan. Adapun batasan masalah yang ditetapkan dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini hanya akan berfokus pada deteksi berita hoaks berbentuk teks yang tersebar di media sosial dan situs berita online dalam bahasa Indonesia, dengan dataset yang diperoleh dari platform Kaggle.
2. Algoritma pembelajaran mesin yang digunakan untuk deteksi hoaks terbatas pada lima algoritma, yaitu *K-Nearest Neighbors* (KNN), *Naïve Bayes*, *Random Forest*, *Support Vector Machine* (SVM), dan *XGBoost*.
3. Teknik preprocessing data yang digunakan meliputi data cleaning, tokenisasi, penghapusan stopword, dan ekstraksi fitur menggunakan *TF-IDF*. Teknik lainnya, seperti stemming atau lemmatization, tidak akan

diterapkan dalam penelitian ini.

4. Penelitian ini hanya akan mengevaluasi performa model berdasarkan akurasi, precision, recall, F1-score, dan pengaruh tuning hyperparameter menggunakan *GridSearchCV* terhadap hasil klasifikasi berita hoaks.

1.4 Tujuan Penelitian

Tujuan yang akan dicapai oleh peneliti dalam penelitian ini adalah menghasilkan sistem klasifikasi berita hoaks berbasis algoritma pembelajaran mesin yang mampu mendeteksi berita tidak valid secara akurat dan efisien. Penelitian ini juga bertujuan untuk:

1. Menganalisis efektivitas beberapa algoritma pembelajaran mesin, yaitu *K-Nearest Neighbors* (KNN), *Naïve Bayes*, *Random Forest*, *Support Vector Machine* (SVM), dan *XGBoost* dalam klasifikasi berita hoaks berbahasa Indonesia.
2. Mengukur pengaruh teknik preprocessing data seperti data cleaning, tokenisasi, penghapusan stopword, dan ekstraksi fitur *TF-IDF* terhadap peningkatan akurasi sistem deteksi hoaks.
3. Mengetahui tingkat peningkatan kinerja model klasifikasi setelah dilakukan tuning hyperparameter menggunakan metode *GridSearchCV*.
4. Menghasilkan model deteksi berita hoaks yang memiliki tingkat akurasi tinggi serta dapat diimplementasikan ke dalam sistem berbasis web untuk mempermudah proses verifikasi berita secara otomatis.

di dalam kesimpulan dan intisari/abstrak. Contoh:

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat baik secara teknis maupun non-teknis, baik bagi objek penelitian, pengguna, maupun pengembangan ilmu pengetahuan di masa depan. Adapun manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut:

Manfaat Teoritis:

1. Penelitian ini dapat memperkaya khasanah teori dalam bidang pembelajaran mesin (*machine learning*), khususnya dalam penerapan algoritma-algoritma untuk deteksi berita hoaks. Hasil penelitian ini diharapkan dapat memberikan kontribusi pada pengembangan teori-teori dalam klasifikasi teks, serta cara-cara efektif untuk mengidentifikasi informasi yang tidak valid atau hoaks.
2. Penelitian ini dapat menjadi referensi tambahan bagi penelitian-penelitian sebelumnya yang membahas deteksi hoaks menggunakan algoritma *machine learning*. Secara khusus, ini dapat memperluas pemahaman tentang pengaruh preprocessing data dan tuning hyperparameter terhadap akurasi model.

Manfaat Praktis:

1. Penelitian ini dapat memberikan solusi praktis dalam menangani masalah penyebaran berita hoaks yang seringkali meresahkan masyarakat. Dengan adanya sistem deteksi berita hoaks yang akurat, masyarakat akan lebih mudah memperoleh informasi yang valid dan dapat meminimalkan dampak negatif dari berita palsu.
2. Sistem deteksi hoaks yang dihasilkan dalam penelitian ini dapat diimplementasikan di platform media sosial dan situs berita online untuk membantu memverifikasi berita secara otomatis. Hal ini dapat mempercepat proses deteksi dan memastikan berita yang tersebar lebih terpercaya.
3. Pihak berwenang, seperti lembaga pemerintah dan organisasi media, dapat memanfaatkan hasil penelitian ini untuk meningkatkan sistem verifikasi berita mereka. Dengan menggunakan algoritma *machine learning*, proses penyaringan informasi yang beredar dapat dilakukan dengan lebih efisien, membantu mencegah penyebaran hoaks secara lebih luas.
4. Penelitian ini dapat menjadi acuan bagi peneliti selanjutnya yang tertarik dalam bidang deteksi hoaks dan penerapan pembelajaran mesin pada teks.

Peneliti selanjutnya dapat mengembangkan atau memperbaiki model yang dihasilkan, serta mengadaptasi teknik yang digunakan pada berbagai bahasa atau dataset lainnya.

1.6 Sistematika Penulisan

Sistematika penulisan dalam skripsi disusun agar pembahasan lebih terarah, runtut, dan mudah dipahami. Adapun susunan sistematika penulisan skripsi terdiri dari lima bab utama, yaitu sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi uraian mengenai latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan. Bab ini memberikan gambaran umum tentang alasan penelitian dilakukan, ruang lingkup penelitian, serta arah dan tujuan yang ingin dicapai.

BAB II TINJAUAN PUSTAKA

Bab ini berisi kajian teori dan penelitian terdahulu yang relevan dengan topik penelitian. Di dalamnya dijelaskan teori-teori dasar terkait berita hoaks, konsep *machine learning*, algoritma yang digunakan (*K-Nearest Neighbors*, *Naïve Bayes*, *Random Forest*, *Support Vector Machine*, dan *XGBoost*), serta teknik preprocessing teks seperti tokenisasi, stopword removal, dan *TF-IDF*. Bab ini juga memuat kerangka berpikir yang menjadi dasar dalam pengembangan sistem deteksi hoaks.

BAB III METODE PENELITIAN

Bab ini menjelaskan metodologi yang digunakan dalam penelitian, mencakup pendekatan penelitian, tahapan penelitian, sumber dan jenis data, proses pengumpulan dan pengolahan data, serta algoritma pembelajaran mesin yang diterapkan. Bab ini juga membahas tahapan tuning hyperparameter menggunakan *GridSearchCV* serta metrik evaluasi yang digunakan, seperti akurasi, precision, recall, dan F1-score.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi hasil pengujian dan analisis terhadap model deteksi berita hoaks yang

dikembangkan. Di dalamnya dijelaskan hasil preprocessing data, performa masing-masing algoritma sebelum dan sesudah tuning hyperparameter, serta perbandingan akurasi antar model. Bab ini juga membahas implementasi model terbaik ke dalam aplikasi berbasis web sebagai sistem pendukung deteksi berita hoaks secara otomatis.

BAB V PENUTUP

Bab ini berisi kesimpulan dari hasil penelitian yang telah dilakukan serta saran yang dapat diberikan untuk penelitian selanjutnya. Kesimpulan disusun berdasarkan hasil analisis dan temuan yang telah dijabarkan pada bab sebelumnya, sedangkan saran berisi rekomendasi untuk pengembangan model atau penerapan sistem pada skala yang lebih luas di masa mendatang.

