

BAB I

PENDAHULUAN

A. Latar Belakang

Penyebaran berita hoax memberikan dampak negatif yang serius bagi masyarakat, mulai dari kecemasan hingga gangguan terhadap stabilitas sosial dan politik. Hal ini diperparah oleh pesatnya perkembangan media sosial yang memungkinkan informasi menyebar secara masif dalam waktu singkat tanpa kontrol memadai. Berita hoax sering kali dirancang agar tampak meyakinkan, sehingga masyarakat mudah terpengaruh tanpa melakukan verifikasi terlebih dahulu. Dampaknya meluas, mulai dari penyebaran informasi yang salah hingga pembentukan opini keliru yang mempengaruhi kebijakan dan persepsi publik. Oleh karena itu, diperlukan sistem yang mampu mendeteksi keberadaan berita hoax secara otomatis untuk mencegah kerugian lebih lanjut.

Berbagai penelitian telah mengusulkan metode klasifikasi berita hoax dengan pendekatan yang beragam. Pada penelitian (Ropikoh dkk., 2021), diusulkan metode *Support Vector Machine* (SVM) untuk melakukan klasifikasi berita hoax terkait *COVID-19* menggunakan *cosine similarity*. Pada penelitian (Mahfudh, 2019) diusulkan metode *Naive Bayes* dalam mendeteksi berita hoax dengan teknik *preprocessing* komprehensif. Pada penelitian (Tambunan et al., 2023) diusulkan metode *Random Forest* dalam klasifikasi berita hoax. Pada penelitian (Ripa'i, Santoso, 2024) diusulkan metode *deep learning* dengan algoritma *Bert* untuk mendeteksi berita hoax. Meskipun setiap metode memiliki keunggulan seperti akurasi tinggi, masing-masing pendekatan juga memiliki keterbatasan

Metode-metode yang digunakan dalam klasifikasi berita hoax memiliki kelebihan dan kekurangan masing-masing. *Support Vector Machine* (SVM) cukup efektif untuk menangani data yang kompleks dan memberikan hasil akurasi tinggi, namun memerlukan tuning parameter yang intensif dan memiliki waktu komputasi tinggi untuk dataset besar (Ropikoh dkk., 2021). *Naive Bayes* unggul dalam kesederhanaan dan kecepatan komputasi, tetapi kurang optimal untuk dataset dengan fitur yang saling bergantung (Mahfudh, 2019). *Random Forest* memberikan keandalan dan akurasi tinggi melalui pendekatan *ensemble*, namun membutuhkan sumber daya komputasi besar (Tambunan dkk., 2023). Metode *deep learning* dengan algoritma *BERT* menawarkan representasi teks yang mendalam dan mampu memahami konteks secara *bidirectional*, tetapi membutuhkan sumber daya komputasi yang besar dan waktu pelatihan yang lama (Ripa'i, Santoso, 2024). Dengan optimalisasi yang tepat, metode-metode ini dapat disesuaikan untuk kebutuhan klasifikasi yang lebih spesifik dan efisien.

Metode *Support Vector Machine* (SVM) telah terbukti cukup akurat dalam melakukan klasifikasi berita hoax, namun metode ini memiliki beberapa kelemahan yang menjadi tantangan dalam implementasinya. Salah satu kelemahan utama SVM adalah memiliki waktu *training* yang lama. *Tuning* yang tidak optimal dapat menghasilkan performa model yang buruk, terutama pada dataset yang besar dan tidak seimbang. Selain itu, proses komputasi pada SVM membutuhkan waktu yang cukup lama, terutama ketika berhadapan dengan jumlah fitur yang tinggi. Hal ini disebabkan oleh sifat algoritma SVM yang bergantung pada perhitungan matrik kernel, yang kompleksitasnya meningkat seiring bertambahnya ukuran dataset.

Pada penelitian ini, diusulkan sebuah solusi untuk mengatasi kendala pada waktu pelatihan SVM dalam klasifikasi berita hoax. *Stochastic Gradient Descent* (SGD) diimplementasikan sebagai metode optimasi untuk mempercepat proses pelatihan pada model SVM dengan memanfaatkan subset acak data pada setiap iterasi. Kombinasi metode SVM dan SGD ini memungkinkan pemrosesan data berskala besar dengan efisien tanpa mengorbankan akurasi prediksi.

Tujuan penelitian ini adalah mengimplementasikan SDG untuk optimasi waktu training pada klasifikasi berita hoax menggunakan SVM. Pemanfaatan SGD diharapkan dapat mereduksi waktu komputasi metode SVM dalam klasifikasi berita hoax. Implementasi ini bertujuan mendukung proses deteksi berita hoax secara lebih cepat dan efisien, sehingga relevan untuk diterapkan pada skenario dunia nyata yang membutuhkan analisis data berskala besar dalam waktu singkat, seperti media sosial dan platform berita online.

B. Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan di atas , maka dapat dirumuskan beberapa permasalahan yang akan dikaji dalam penelitian ini, yaitu:

1. Seberapa akurat algoritma SVM dalam melakukan klasifikasi berita hoax?
2. Bagaimana pengaruh metode SGD untuk mengurangi waktu *training* pada algoritma SVM dalam klasifikasi berita hoax ?

C. Tujuan Penelitian

penelitian tentang implementasi metode SGD untuk optimasi waktu training pada klasifikasi berita hoax menggunakan SVM ini memiliki tujuan sebagai berikut:

1. Mengetahui tingkat keakuratan algoritma SVM dalam mengklasifikasi berita hoax.
2. Mengetahui tingkat pengaruh penerapan metode SGD untuk mengurangi waktu training pada algoritma svm dalam klasifikasi berita hoax.

D. Batasan Masalah

Untuk menghindari pembahasan yang terlalu luas dan memastikan penelitian ini lebih terarah, maka dalam penelitian ini ditetapkan batasan masalah sebagai berikut:

1. Klasifikasi Berita Hoax Merupakan proses pemisahan berita yang mengandung informasi palsu atau hoax dari berita valid menggunakan algoritma pembelajaran mesin.
2. *Support Vector Machine* (SVM) Merupakan algoritma pembelajaran mesin yang digunakan untuk klasifikasi dan regresi, yang berfungsi untuk mencari *hyperplane* optimal yang memisahkan data ke dalam berbagai kategori.
3. *Stochastic Gradient Descent* (SGD) adalah algoritma pengoptimalan yang digunakan dalam pembelajaran mesin untuk meminimalkan fungsi kerugian dengan memperbarui bobot model secara berulang.

4. Optimasi waktu *training* adalah upaya untuk meningkatkan efisiensi proses pelatihan model dalam machine learning agar dapat selesai dalam waktu yang lebih singkat tanpa mengorbankan akurasi atau kualitas hasil prediksi model.

E. Manfaat Penelitian

Adapun beberapa manfaat yang diharapkan dari penelitian adalah sebagai berikut:

1. Deteksi Hoax yang Cepat dan Akurat.

Implementasi algoritma SVM yang dioptimalkan menggunakan SGD) memungkinkan deteksi berita hoax secara cepat tanpa mengorbankan akurasi, bahkan pada dataset besar.

2. Penghematan Sumber Daya Komputasi.

Optimalisasi waktu pelatihan melalui SGD mengurangi kebutuhan sumber daya komputasi, menjadikan sistem deteksi hoax lebih efisien.