

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Berita hoaks telah menjadi permasalahan signifikan di Indonesia, terutama dalam konteks politik seperti pemilu. Penyebaran informasi palsu dipermudah oleh penggunaan media sosial dan internet, yang menjadi saluran utama diseminasi informasi. Kondisi ini mengancam integritas informasi serta kepercayaan publik terhadap proses politik. Menurut Masyarakat Telematika Indonesia (MASTEL), berita hoaks politik dapat merusak proses demokrasi, memecah belah masyarakat, dan mengubah persepsi publik terhadap institusi politik [1, 2, 3]. Hingga Mei 2023, Tim AIS Kementerian Komunikasi dan Informatika mencatat 11.642 konten hoaks sejak Agustus 2018, dengan kategori politik mencapai 1.373 konten [4]. Faktor sosial seperti pengaruh jaringan sosial dan karakteristik kepribadian, termasuk narsisme dan Machiavellianisme, berperan dalam penyebaran hoaks [5, 6]. Kombinasi faktor ini menjelaskan mengapa masyarakat sulit membedakan informasi asli dari hoaks, terutama dalam lingkungan digital yang sangat cepat berubah.

Melihat dampak negatif yang ditimbulkan oleh berita hoaks, terutama dalam konteks politik seperti pada Pemilu Presiden 2019, muncul kebutuhan untuk mengembangkan metode untuk mendeteksi penyebaran informasi palsu. Pada pemilu tersebut, penyebaran hoaks sangat masif, terutama melalui platform media sosial seperti Twitter, yang menyoroti pentingnya metode deteksi yang efisien dan cepat [7]. Dalam hal ini, pendekatan machine learning menawarkan solusi yang dapat diandalkan dengan kemampuan klasifikasi yang mampu mengidentifikasi berita palsu secara efektif. Penggunaan machine learning tidak hanya mempermudah proses identifikasi hoaks, tetapi juga memberikan skalabilitas yang tinggi, mengingat banyaknya informasi yang beredar di internet. Selain itu, metode manual untuk memverifikasi berita memerlukan waktu dan biaya yang besar, sehingga machine learning menjadi pilihan yang lebih efisien dalam mendeteksi hoaks [8]. Kampanye disinformasi juga diakui sebagai ancaman signifikan dalam konteks keamanan siber, karena dapat memanipulasi opini publik dan memicu ketidakstabilan sosial [9]. Penyebaran berita palsu juga dapat memfasilitasi serangan siber dengan menciptakan kebingungan dan memanfaatkan

kurangnya kepercayaan pada sumber informasi yang sah, sehingga mempermudah aktor jahat untuk menipu pengguna [10]. Oleh karena itu, penerapan machine learning dalam klasifikasi berita hoaks menjadi salah satu langkah untuk mengatasi penyebaran disinformasi di era digital ini.

Banyak upaya mendeteksi berita hoaks menggunakan machine learning di Indonesia, berbagai pendekatan telah dikembangkan dengan hasil yang beragam. Pendekatan tradisional, seperti yang diteliti oleh Nugraha [11], mengevaluasi kinerja beberapa algoritma klasifikasi menggunakan dataset berjumlah 600 sampel berita Indonesia. Studi ini menunjukkan bahwa algoritma K-Nearest Neighbors mencapai akurasi tertinggi sebesar 72,8% dan presisi 76,2%. Meskipun hasil ini cukup menjanjikan, keterbatasan pendekatan tradisional dalam menangani kompleksitas data yang terus berkembang mendorong penelitian ke arah metode yang lebih canggih. Sebagai alternatif yang menjanjikan, pendekatan ensemble learning telah menunjukkan potensi yang lebih besar. Penelitian oleh Haumahu [12] mengeksplorasi penggunaan model XGBoost menggunakan 500 sampel berita berbahasa Indonesia, menghasilkan peningkatan signifikan dengan akurasi mencapai 89%, presisi 90%, dan recall 80%. Temuan ini mengindikasikan bahwa metode ensemble learning mampu mengatasi keterbatasan model individual dan memberikan performa yang lebih baik dengan ukuran dataset serupa.

Dalam perkembangan teknik deteksi berita hoaks, LightGBM muncul sebagai alternatif yang menjanjikan di samping XGBoost dalam kategori ensemble learning. Sebagai implementasi yang efisien dari algoritma Gradient Boosting Decision Tree (GBDT), LightGBM didesain khusus untuk mengoptimalkan kecepatan pelatihan dan skalabilitas, terutama ketika berhadapan dengan dataset berukuran besar. Keunggulan LightGBM terletak pada dua inovasi teknisnya: Gradient-based One-Side Sampling (GOSS) dan Exclusive Feature Bundling (EFB) [13]. GOSS meningkatkan efisiensi dengan memfokuskan sampling pada data yang memiliki gradien tinggi, sementara EFB berperan dalam mereduksi dimensi fitur. Kombinasi kedua teknik ini menghasilkan peningkatan efisiensi komputasi yang signifikan tanpa mengorbankan akurasi model. Dibandingkan dengan XGBoost, LightGBM sering kali menjadi pilihan para praktisi machine learning karena keunggulannya dalam kecepatan pelatihan, meskipun dari segi akurasi, performanya tidak selalu mengungguli XGBoost [14, 15]. Salah satu fitur kunci LightGBM adalah strategi pertumbuhan pohon secara leaf-wise, yang memungkinkan pembentukan struktur pohon yang lebih dalam dan berpotensi meningkatkan akurasi, terutama pada dataset dengan karakteristik tertentu [16].

Beberapa penelitian telah menunjukkan potensi besar LightGBM dalam tugas klasifikasi berita, di mana salah satu studi berhasil mencapai akurasi 86% dengan menggunakan confusion matrix untuk klasifikasi tipe berita [17].

Penelitian ini akan menggunakan dataset yang mencakup spektrum lebih luas dari konten hoaks, dengan penekanan khusus pada konteks politik dan sosial di Indonesia. Dataset ini dirancang untuk mewakili berbagai kategori hoaks, termasuk berita politik, ekonomi, hingga isu sosial, sehingga dapat meningkatkan generalisasi model yang dikembangkan. Model yang digunakan dalam penelitian ini, termasuk LightGBM, diharapkan mampu mengolah dataset berukuran besar dan beragam untuk menghasilkan sistem klasifikasi berita hoaks yang akurat. Dengan pendekatan ini, penelitian ini diharapkan dapat memberikan kontribusi signifikan terhadap pengembangan metode deteksi berita hoaks yang lebih efektif dan relevan dalam konteks Indonesia.

1.2 Rumusan Masalah

1. Bagaimana cara membuat model *machine learning* untuk klasifikasi berita hoaks dalam Bahasa Indonesia menggunakan algoritma *Light Gradient Boosting Machine*?
2. Berapa tingkat akurasi klasifikasi berita hoaks dalam Bahasa Indonesia yang dapat dicapai oleh algoritma *Light Gradient Boosting Machine*?

1.3 Tujuan Penelitian

1. Membuat model *machine learning* untuk klasifikasi berita hoaks dalam Bahasa Indonesia menggunakan algoritma *Light Gradient Boosting Machine*
2. Berapa tingkat akurasi klasifikasi berita hoaks dalam bahasa Indonesia yang dapat dicapai oleh algoritma *Light Gradient Boosting Machine*

1.4 Urgensi Penelitian

Dalam era digital yang berkembang pesat, penyebaran berita hoaks di Indonesia, terutama melalui media sosial, telah menjadi masalah serius yang mengancam proses demokrasi dan kepercayaan publik, khususnya dalam konteks politik dan pemilu. Saat ini, verifikasi kebenaran berita masih banyak dilakukan secara manual, yang membutuhkan waktu dan sumber daya yang

besar. Keterbatasan ini memperlambat proses deteksi dan penanganan berita palsu, sehingga memungkinkan informasi menyesatkan untuk terus menyebar luas. Oleh karena itu, penelitian ini penting untuk mengembangkan solusi berbasis teknologi yang dapat mengotomatisasi klasifikasi berita hoaks. Dengan menggunakan algoritma LightGBM, diharapkan deteksi berita hoaks dapat dilakukan lebih akurat dan efisien, mengurangi ketergantungan pada verifikasi manual, serta memberikan kontribusi signifikan dalam menangani penyebaran disinformasi di Indonesia.

1.5 Luaran Penelitian

Luaran utama dari penelitian ini adalah menghasilkan model klasifikasi yang dapat mendeteksi berita hoaks berbahasa Indonesia dengan akurasi tinggi menggunakan algoritma LightGBM. Model ini diharapkan dapat diimplementasikan pada platform digital untuk mendukung proses verifikasi otomatis dalam mendeteksi berita palsu secara cepat dan efisien. Selain itu, penelitian ini juga menargetkan luaran akademis berupa publikasi ilmiah di jurnal bereputasi, khususnya di bidang Natural Language Processing (NLP), keamanan informasi, atau pembelajaran mesin. Dengan demikian, penelitian ini tidak hanya memberikan dampak praktis dalam menangani penyebaran berita hoaks di Indonesia, tetapi juga berkontribusi pada perkembangan riset ilmiah di tingkat internasional.

1.6 Manfaat Penelitian

1. Meningkatkan efisiensi deteksi berita hoaks di Indonesia.
Dengan klasifikasi otomatis menggunakan LightGBM, proses identifikasi berita hoaks dapat dilakukan lebih cepat dan akurat, mengurangi ketergantungan pada verifikasi manual serta mempercepat respons terhadap penyebaran informasi palsu.
2. Mengurangi dampak negatif berita hoaks terhadap masyarakat.
Model ini membantu mencegah penyebaran informasi yang dapat merusak kepercayaan publik dan mengganggu stabilitas sosial, khususnya dalam konteks politik dan pemilu, dengan menyediakan deteksi hoaks yang andal dan tepat waktu.