

1. Pendahuluan

Latar Belakang

Ditengah perkembangan era teknologi digital, media sosial menjadi sarana yang paling banyak digunakan untuk berbagi dan bertukar informasi. Berdasarkan survey yang dilakukan oleh DataIndonesia [1], pada tahun 2023 indonesia menduduki peringkat kelima sebagai negara dengan jumlah pengguna media sosial terbesar di dunia, dengan jumlah pengguna sebanyak 167 juta setara dengan 60.4% jumlah penduduk di Indonesia. Konsumsi media sosial yang tinggi tersebut dapat menyebabkan informasi menjadi semakin mudah tersebarluaskan, namun minimnya sikap kritis dalam melakukan filter informasi yang didapat bisa membuat berita *hoax* menjadi tersebar dengan mudah [2]. Mastel.id [3] telah melakukan survey mengenai penyebaran berita *hoax* di Indonesia dan ditemukan bahwa sosial media menjadi saluran penyebaran berita *hoax* tertinggi dengan persentase 80,5%.

Hoax merupakan berita yang sengaja dipalsukan lalu disebarakan tanpa dasar kebenaran, *hoax* digunakan untuk menipu publik agar mempercayai sesuatu yang tidak terjadi [4]. *Hoax* dapat menimbulkan adanya bias informasi di masyarakat, selain itu juga dapat memunculkan adanya kebencian, penggiringan opini dan permusuhan [5]. Sehingga diperlukan adanya upaya untuk mengurangi penyebaran *hoax*, salah satunya dengan membangun sistem yang dapat mendeteksi *hoax*.

Penelitian mengenai deteksi *hoax* sudah dilakukan dengan menggunakan berbagai metode. Pada penelitian ini akan dibangun sistem klasifikasi *hoax* pada twitter yang mengimplementasikan seleksi fitur *Information Gain* serta pembobotan TF-IDF pada metode klasifikasi *Naive Bayes Multinomial*. Tujuan dari penelitian ini adalah untuk mengetahui pengaruh penggunaan *Information Gain* terhadap performansi dari sistem yang dibangun. *Naive Bayes Multinomial* dipilih sebagai metode pada penelitian ini karena metode ini dapat mengklasifikasikan teks dengan baik [6], serta penggunaan pembobotan TF-IDF dalam klasifikasi dapat meningkatkan akurasi performansi menjadi lebih optimal [7]. Penggunaan fitur seleksi dapat meningkatkan nilai performansi dari sistem yang dibangun, *Information Gain* merupakan metode yang efisien digunakan untuk seleksi fitur karena mampu menentukan fitur yang paling merepresentasikan suatu dokumen dan juga dapat meningkatkan performa klasifikasi [8].

Topik dan Batasannya

Berdasarkan latar belakang diatas, maka topik yang dibahas dalam tugas akhir ini yaitu bagaimana menerapkan seleksi fitur *Information Gain* dengan model klasifikasi *Naive Bayes Multinomial* dan pembobotan TF-IDF pada identifikasi berita *hoax*.

Adapun Batasan masalah pada tugas akhir ini adalah sebagai berikut:

1. Data yang digunakan merupakan hasil *crawling* dari Twitter.
2. Data yang digunakan merupakan kategori sosial politik di Indonesia.
3. Data yang digunakan berjumlah 5047 tweet.
4. Pemberian label pada data dilakukan secara manual dengan dua kategori : *hoax* dan *nonhoax*.

Tujuan

Penelitian ini bertujuan untuk mengevaluasi pengaruh penggunaan seleksi fitur *Information Gain* pada klasifikasi berita *hoax* di twitter dengan menggunakan metode klasifikasi *Naive Bayes Multinomial* dan pembobotan TF-IDF.

Organisasi Tulisan

Urutan penulisan dalam penelitian ini adalah sebagai berikut: bagian satu berisi latar belakang penelitian, bagian dua menunjukkan studi literatur mengenai teori yang digunakan, bagian tiga berisi alur sistem yang dibangun, bagian empat menjelaskan mengenai hasil pengujian, dan bagian lima berisi kesimpulan dan saran untuk penelitian selanjutnya.