

BAB 1

PENDAHULUAN

1.1 LATAR BELAKANG

Indonesia adalah salah satu negara demokrasi yang mana setiap warga negara memiliki hak yang sama untuk mengambil keputusan dalam pemerintahan. Salah satu cara Indonesia menerapkan sistem demokrasi yaitu dengan mengadakan Pemilihan Umum (Nugroho & Sukmariningsih, 2020). Seiring dengan berakhirnya masa jabatan Presiden dan Wakil Presiden Republik Indonesia yang ketujuh maka dilaksanakanlah pemilu yang rutin dilakukan setiap 5 tahun sekali untuk menentukan Presiden dan Wakil Presiden selanjutnya. Pada pemilu tahun 2024 ini dikenal sebagai pesta demokrasi sehingga terjadi banyak penyebaran berita hoaks di media sosial. (Imran et al., 2024).

Berdasarkan data yang diperoleh dari *Google Trends*, pencarian kata terkait pemilu masih meningkat pesat hingga 3000% hal ini menunjukkan bahwa masyarakat masih antusias mencari informasi mengenai pemilu melalui *website* (Google Trends, 2024). Sebagai media penyampaian berita ternyata terdapat terdapat 203 isu hoaks dengan total sebaran 2.882 konten di *website* (Katadata, 2024). Berdasarkan survei yang telah dilakukan Katadata Insight Center (KIC) yang berkerja sama dengan Kementerian Komunikasi dan Informatika, hasilnya sebanyak 30% hingga 60% orang Indonesia yang terpapar hoaks. Namun hanya 21% sampai 36% saja yang mampu membedakan apakah berita tersebut hoaks atau tidak (Fajri, 2023).

Berkaitan dengan hal tersebut, diperlukan model *machine learning* untuk melakukan klasifikasi berita hoaks pasca pemilu 2024 dengan bantuan algoritma *K-Nearest Neighbour* (KNN). Prinsip dari metode klasifikasi menggunakan KNN adalah prinsip perbandingan jarak antara data yang belum diketahui dengan data pelatihan yang telah ada (Zhang & Li, 2023). KNN memiliki keuntungan dalam *text mining*, seperti mampu menangani dataset besar, dikombinasikan dengan teknik lain, dan memiliki potensi untuk menangani data *noise*, dan mengekstraksi

informasi yang lebih komprehensif (Taha et al., 2024).

Term *Frequency-Inverse Document Frequency* (TF-IDF) akan penulis terapkan untuk meningkatkan akurasi dengan mengukur seberapa penting suatu kata dalam sebuah dokumen guna mengetahui potensi hoaks dari suatu berita sehingga mampu meningkatkan kemampuan model dalam memahami konteks dan pola tertentu. Pengambilan data dalam penelitian ini dengan melakukan *scraping website* turnbackhoax.id, kominfo.go.id, infopublik.id, dan rri.co.id sebagai data yang menjadi acuan pembelajaran mesin serta melakukan *scraping website* kompas.com, liputan6.com, detiknews.com, antaranews.com, cnnindonesia.com, okezone.com, sindonews.com, kumparan.com, pikiranrakyat.com, dan wartatransparansi.com. Pengembangan model ini meliputi akuisisi data, *preprocessing*, pemilihan algoritma, pelatihan dan evaluasi model. (Villela et al., 2023). Hasil dari penelitian memberikan informasi untuk mengetahui daftar berita yang memiliki potensi hoaks selama pasca pemilu 2024.

Hasil penelitian ini diharapkan dapat mengevaluasi situs berita yang memiliki potensi fakta yang besar. Penggunaan algoritma KNN dan pembobotan kata menggunakan TF-IDF, penelitian ini diharapkan dapat mengklasifikasikan berita yang berpotensi hoaks untuk menghadapi tantangan informasi yang tidak akurat.

1.2 PERUMUSAN MASALAH

Berdasarkan latar belakang yang telah diuraikan diatas, berikut adalah rumusan masalah dalam penelitian ini:

1. Bagaimana algoritma KNN dapat melakukan klasifikasi berita hoaks?
2. Apakah model yang dibuat dapat membantu masyarakat dalam melawan disinformasi?

1.3 BATASAN MASALAH

Batasan masalah dari penelitian ini adalah sebagai berikut:

1. Data pada penelitian ini tidak *up-to-date* yaitu berita yang di *scraping* minimal dari tanggal 15 Februari dan maksimal 15 Mei.

2. Data berita yang dikumpulkan hanya yang berkaitan dengan pemilu dan pilpres 2024.
3. Data latih hanya berasal dari turnbackhoax.id, kominfo.go.id, infopublik.id, dan rri.co.id.
4. Implementasi algoritma KNN hanya pada *website* kompas.com, liputan6.com, detiknews.com, antaranews.com, cnnindonesia.com, okezone.com, sindonews.com, kumparan.com, pikiranrakyat.com, dan wartatransparansi.com.

1.4 PERTANYAAN PENELITIAN

Berdasarkan permasalahan yang diangkat terdapat beberapa pertanyaan penelitian sebagai berikut:

1. Bagaimana cara memperoleh data berita hoaks dan non-hoaks?
2. Apa saja atribut kolom yang dibutuhkan untuk mengkasifikasi potensi berita hoaks?
3. Bagaimana cara algoritma dalam mengklasifikasi berita yang berpotensi hoaks dan non-hoaks?

1.5 TUJUAN PENELITIAN

Tujuan dari penelitian ini adalah sebagai berikut:

1. Membangun sebuah pemodelan menggunakan algoritma *K-Nearest Neighbour* untuk mengevaluasi keakuratan dan keandalan berita dari berbagai situs web berita di Indonesia dengan menggunakan teknik klasifikasi berbasis *machine learning*.
2. Mengidentifikasi sejauh mana berita-berita yang disajikan oleh situs-situs tersebut dapat dikategorikan sebagai berpotensi fakta atau berpotensi hoaks. Dengan melakukan analisis terhadap berita dari situs-situs seperti Kompas, Liputan6, Detik, CNN Indonesia, Okezone, Sindo News, Kumparan, Pikiran Rakyat, dan Wartatransparansi.
3. Memberikan wawasan berupa laman berita yang memiliki informasi potensi fakta tinggi.

1.6 MANFAAT HASIL PENELITIAN

Adapun manfaat dari penelitian ini yaitu:

1. Memilah informasi yang benar dan yang salah terkait pemilu.
2. Mengidentifikasi portal berita mana yang memiliki tingkat akurasi tinggi dalam penyajian berita.
3. Menanggulangi penyebaran berita palsu pascapemilu 2024.
4. Kontribusi perkembangan pengetahuan dalam klasifikasi berita menggunakan algoritma KNN.