

BAB I

PENDAHULUAN

Bab pendahuluan ini berisikan latar belakang masalah, rumusan masalah, tujuan, manfaat, ruang lingkup, serta sistematika penulisan pada skripsi mengenai klasifikasi berita hoaks Indonesia menggunakan pendekatan BoW, TF-IDF, dan *Support Vector Machine* (SVM).

1.1 Latar Belakang

Kemajuan teknologi informasi dan luasnya penetrasi internet telah memudahkan individu untuk membuat, membagikan, dan menyebarkan informasi, baik yang valid maupun tidak. Informasi ini dapat tersebar dengan cepat dan mudah dikonsumsi oleh masyarakat. Negara Indonesia sendiri merupakan salah satu negara dengan jumlah pengguna internet yang tinggi, terutama di kalangan pemuda. Menurut survei dari Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), jumlah pengguna internet di Indonesia pada tahun 2024 mencapai 221 juta orang, naik sebesar 2,8% dari tahun sebelumnya yang tercatat sebanyak 215 juta pengguna internet. Dengan total populasi 278,7 juta jiwa, persentase pengguna internet di Indonesia mencapai 79,5% (Rizaty, 2024). Peningkatan pengguna internet di Indonesia ini membuat ekosistem penyebaran informasi menjadi lebih cepat dan dinamis, yang memungkinkan distribusi berita penting dapat diterima secara instan kepada sejumlah besar masyarakat. Di sisi lain, penyebaran berita yang cepat ini tidak menjamin informasi dan berita yang diperoleh dapat divalidasi dari sumber yang valid, sehingga banyak bermunculan berita palsu dari sumber yang tidak terpercaya di internet. Beberapa tahun belakang, dunia informasi di berbagai negara termasuk Indonesia sedang diserang dengan maraknya berita-berita palsu atau hoaks.

Hoaks biasanya diartikan sebagai informasi yang belum dapat dipastikan kebenarannya (Syaifullah & Anggraini, 2022). Hoaks juga merupakan bentuk manipulasi informasi dengan tujuan menyembunyikan kebenaran yang ada. Dengan kata lain, hoaks adalah upaya untuk menyembunyikan fakta dengan membanjiri media dengan pesan yang menyesatkan untuk menutupi informasi yang akurat (Listuti, 2023). Dampak dari penyebaran berita hoaks bagi pemerintah adalah merosotnya kepercayaan masyarakat terhadap lembaga pemerintahan, serta dapat menimbulkan keraguan, kecemasan, rasa

takut, dan kecurigaan terhadap kinerja dan kebijakan pemerintah. Selain itu, berita hoaks dapat memicu ujaran kebencian, diskriminasi, dan permusuhan terhadap kelompok tertentu yang berbeda suku, agama, ras, atau etnis, sehingga dapat menimbulkan gesekan dan kekerasan sosial (Subekti et al., 2020). Penyebaran berita hoaks juga memiliki konsekuensi dan merugikan banyak pihak, di mana berita hoaks juga dapat merugikan berbagai aspek, diantaranya waktu, ekonomi, hubungan sosial, dan lain sebagainya (Munirul et al., 2020). Untuk menghindari dampak-dampak buruk tersebut, maka dibutuhkan sistem untuk membantu mengklasifikasikan berita hoaks, agar masyarakat dapat mengetahui informasi yang mereka dapatkan tersebut tergolong sebagai berita hoaks atau valid.

Berdasarkan penelitian sebelumnya, klasifikasi berita hoaks yang telah dilakukan oleh Fauzy dan Erwin pada tahun 2023. Mereka mengembangkan sistem untuk mendeteksi berita palsu di media sosial Twitter menggunakan pendekatan *deep learning* dengan metode *Convolutional Neural Network* (CNN), *Term Frequency-Inverse Document Frequency* (TF-IDF), *Bidirectional Encoder Representations from Transformers* (BERT), dan *Global Vector* (GLOVE). Sistem yang mereka kembangkan menghasilkan akurasi sebesar 98,57% (Fauzy & Setiawan, 2023). Pada tahun 2022, penelitian serupa juga dilakukan oleh Rafique, dkk. Mereka melakukan penelitian terkait deteksi berita palsu Urdu menggunakan ekstraksi fitur BoW dengan berbagai klasifikasi *machine learning*, seperti *Naive Bayes*, *Support Vector Machine*, *Random Forest*, *Logistic Regression*, *Gradient Boost*, dan LSTM, serta didapatkan akurasi tertinggi pada model *Random Forest* sebesar 92% (Rafique et al., 2022).

Pada tahun 2021, Penelitian yang dilakukan oleh Prayoga, dkk berfokus pada deteksi hoaks menggunakan model *Bernoulli Naive Bayes* dengan ekstraksi fitur yang digunakan adalah TF-IDF. Penelitian yang dilakukan ini berhasil mencapai akurasi 98,5% dalam mendeteksi berita hoaks (Prayoga et al., 2021). Selain itu, Rahmawati juga melakukan penelitian yang sama tentang klasifikasi berita hoaks. Beliau mengombinasikan TF-IDF dengan membandingkan tiga model klasifikasi (*Support Vector Machine*, *Random Forest*, dan Regresi Logistik). Dari hasil penelitian beliau, model SVM memiliki akurasi terbaik sebesar 83% (Rahmawati, 2022).

Berdasarkan penelitian terdahulu, ekstraksi fitur menggunakan BoW dan TF-IDF sangat mempengaruhi keakuratan suatu model. TF-IDF digunakan untuk mengukur seberapa penting suatu kata dalam sebuah dokumen, sedangkan BoW untuk menghitung frekuensi kemunculan kata-kata dalam sebuah dokumen. Selain itu, BoW dan TF-IDF menawarkan pendekatan pemrosesan teks yang sederhana namun efektif. Hal ini membuat TF-IDF dan BoW mudah untuk diimplementasikan dan diterapkan ke dalam berbagai model pemrosesan bahasa alami tanpa memerlukan kompleksitas yang tinggi. Maka pada penelitian ini, peneliti mencoba mengombinasikan BoW dan TF-IDF dengan model *Support Vector Machine* (SVM).

1.2 Rumusan Masalah

Berdasarkan latar belakang penelitian, rumusan masalah pada penelitian ini adalah bagaimana mengklasifikasikan berita hoaks di Indonesia menggunakan ekstraksi fitur BoW dan TF-IDF dengan model *Support Vector Machine*.

1.3 Tujuan dan Manfaat

Tujuan umum dari penelitian ini untuk mengevaluasi bagaimana penggunaan kombinasi BoW dan TF-IDF mempengaruhi kinerja model *Support Vector Machine* (SVM) dalam melakukan klasifikasi berita hoaks Indonesia.

Manfaat dari penelitian ini, diharapkan dapat memberikan kontribusi dalam pembuatan sistem yang mampu mengidentifikasi tingkat hoaks suatu berita secara efektif dan akurat, sehingga masyarakat masyarakat dapat mengetahui informasi yang mereka dapatkan tersebut tergolong sebagai berita hoaks atau valid.

1.4 Ruang Lingkup

Dalam skripsi ini perlu adanya batasan-batasan agar pembahasan lebih terarah dan tidak melebihi target yang diteliti. Batasan-batasan tersebut adalah :

1. Data yang digunakan berasal dari dua sumber, yang pertama melalui penelitian (Rahutomo et al, 2019) dengan total 600 data yang terdiri dari 228 berita hoaks dan 372 berita valid. Kedua berasal dari repositori GitHub bernama Pierobeat, dengan total 500 data yang terdiri dari 250 berita hoaks dan 250 berita valid. Kedua sumber

ini hasil dari *scrapping* situs berita resmi Indonesia (TurnBackHoax, Liputan6, Kompas, Detik, dan CNN Indonesia).

2. Data yang diklasifikasikan merupakan data berbahasa Indonesia.
3. Metode klasifikasi *machine learning* yang digunakan dalam penelitian ini adalah *Support Vector Machine*.

1.5 Sistematika Penulisan

Untuk memberikan suatu gambaran yang urut dan jelas, maka skripsi ini terbagi menjadi beberapa pokok bahasan, yaitu :

BAB I PENDAHULUAN

Bab ini membahas latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika dalam penelitian mengenai klasifikasi berita hoaks Indonesia menggunakan pendekatan BoW dan TF-IDF dengan model *Support Vector Machine*.

BAB II LANDASAN TEORI

Bab ini membahas latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika dalam penelitian mengenai klasifikasi berita hoaks Indonesia menggunakan pendekatan BoW dan TF-IDF dengan model *Support Vector Machine*.

BAB III METODOLOGI PENELITIAN

Bab ini membahas tentang metodologi penelitian yang digunakan dalam penelitian mengenai klasifikasi berita hoaks Indonesia menggunakan pendekatan BoW dan TF-IDF dengan model *Support Vector Machine*.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan mengenai analisis dan hasil penelitian mengenai klasifikasi berita hoaks Indonesia menggunakan pendekatan BoW dan TF-IDF dengan model *Support Vector Machine*.

BAB V PENUTUP

Bab ini merupakan penutup yang berisi kesimpulan dari uraian yang telah dijelaskan sebelumnya dan saran untuk pengembangan penelitian lebih lanjut.