

BAB 5

SIMPULAN SARAN

5.1 Simpulan

Penelitian ini menunjukkan bahwa model Naive Bayes memiliki kinerja yang cukup baik dalam mendeteksi berita hoax di Indonesia, terutama ketika menggunakan teknik ekstraksi fitur CountVectorizer dengan konfigurasi unigram (1,1). Teknik ini menghasilkan akurasi tertinggi sebesar 93% pada data uji, mengindikasikan bahwa unigram mampu menangkap informasi yang paling relevan dan informatif untuk model. Penggunaan TfidfVectorizer juga memberikan hasil yang kompetitif dengan akurasi tertinggi sebesar 90.6% pada konfigurasi unigram.

Konfigurasi unigram (1,1) dan kombinasi unigram bigram (1,2) memberikan hasil akurasi yang sama karena bigram pada konfigurasi (1,2) tidak memberikan informasi tambahan yang signifikan. Ini disebabkan oleh pola yang relevan dalam dataset lebih banyak ditangkap oleh unigram. Sebaliknya, bigram (2,2) dan kombinasi bigram trigram (2,3) menunjukkan hasil yang serupa karena trigram dalam konfigurasi (2,3) cenderung menambahkan fitur yang bersifat noise tanpa meningkatkan representasi data. Sementara itu, konfigurasi trigram (3,3) menghasilkan akurasi yang lebih rendah, yaitu 58.2% pada CountVectorizer dan 59.5% pada TfidfVectorizer, karena model kesulitan menangani kompleksitas fitur yang berlebihan dan cenderung overfitting pada data latih.

Secara keseluruhan, hasil penelitian ini menegaskan pentingnya pemilihan teknik ekstraksi fitur dan konfigurasi n-gram dalam pengembangan sistem deteksi berita hoax. CountVectorizer dengan unigram terbukti lebih efektif dalam menangkap pola-pola sederhana yang relevan untuk klasifikasi, sedangkan TfidfVectorizer lebih cocok untuk tugas yang membutuhkan pembobotan kata berdasarkan distribusinya. Penelitian ini juga menunjukkan bahwa penambahan kompleksitas melalui bigram dan trigram tidak selalu meningkatkan kinerja, terutama ketika data yang digunakan tidak memiliki pola berulang yang cukup kompleks untuk mendukung n-gram yang lebih tinggi.

Hasil penelitian ini diharapkan dapat menjadi panduan dalam pengembangan lebih lanjut model deteksi hoax yang tidak hanya akurat tetapi juga efisien. Model ini juga memiliki potensi untuk diintegrasikan ke dalam sistem keamanan siber guna memitigasi dampak berita palsu yang semakin meluas di era digital.

5.2 Saran

Berdasarkan hasil penelitian ini, beberapa saran diajukan untuk pengembangan lebih lanjut. Penelitian di masa depan disarankan untuk mengeksplorasi teknik ekstraksi fitur yang lebih canggih, seperti Word2Vec atau embedding berbasis transformer, yang dapat menangkap konteks kata secara lebih baik. Selain itu, penggunaan dataset yang lebih beragam, perlu dipertimbangkan agar model lebih adaptif terhadap berbagai jenis hoax.

Pengembangan model juga dapat diperluas melalui integrasi dengan aplikasi berbasis web atau mobile untuk menguji efektivitasnya dalam skenario real-time. Pendekatan yang menggabungkan teks dengan elemen lain, seperti gambar atau metadata, juga dapat memberikan hasil yang lebih menyeluruh. Dengan langkah-langkah ini, penelitian di bidang deteksi hoax diharapkan dapat memberikan kontribusi yang lebih besar dalam menangkal penyebaran informasi palsu.

