

## **BAB I PENDAHULUAN**

### **1.1 Latar Belakang**

Penyebaran berita palsu (hoax) merupakan salah satu isu krusial di ruang digital Indonesia yang berpotensi mengancam harmoni sosial. Berbagai informasi yang tidak terverifikasi dan bersifat provokatif terbukti dapat meresahkan masyarakat serta berpotensi mengganggu kesatuan dan persatuan bangsa [1]. Arus informasi yang cepat di platform media sosial seringkali menjadi faktor utama yang mempercepat penyebaran disinformasi ini.

Skala masalah ini terkonfirmasi oleh data resmi pemerintah. Kementerian Komunikasi dan Digital (Komdigi) mencatat telah mengidentifikasi dan mengklarifikasi 1.923 konten hoaks sepanjang tahun 2024 saja. Temuan ini menunjukkan bahwa penyebaran hoaks terjadi secara masif dan konstan, dengan kategori penipuan (890 konten) serta isu politik (237 konten) menjadi jenis hoaks yang paling banyak beredar. Besarnya volume disinformasi ini membuktikan adanya tantangan serius dalam menjaga ruang siber yang sehat di Indonesia [2].

Dampak dari ribuan sebaran hoaks ini bersifat multidimensional dan sangat merugikan. Penyebaran informasi palsu dapat menggerus kepercayaan masyarakat terhadap institusi resmi, menciptakan kebingungan, dan memicu ketegangan sosial politik. Selain itu, hoaks dalam kategori penipuan dapat mengakibatkan kerugian finansial yang signifikan. Dalam jangka panjang, fenomena ini berpotensi merusak stabilitas sosial dan menghambat upaya penyelesaian masalah publik yang lebih mendesak [3].

Untuk mengatasi keterbatasan verifikasi manual di tengah volume berita yang besar, diperlukan pendekatan teknologi otomatis berbasis Artificial Intelligence, khususnya Machine Learning. Dalam domain klasifikasi teks, terdapat berbagai algoritma yang umum digunakan, seperti Naïve Bayes, Decision Tree, dan Support Vector Machine (SVM). Penelitian terdahulu seringkali melakukan perbandingan antar algoritma ini untuk menemukan performa terbaik.

Sebagai contoh, Imran, dkk. [4] melakukan klasifikasi berita hoax dengan membandingkan algoritma Naïve Bayes dan SVM. Hasil penelitian mereka menunjukkan bahwa SVM (akurasi 93,33%) lebih unggul dibandingkan Naïve Bayes (88,33%). Temuan serupa juga dilaporkan oleh Febriyanty, dkk. [5] yang menyimpulkan bahwa SVM lebih efektif dalam mendeteksi pola berita palsu dibandingkan metode probabilistik lainnya. Hal ini mengindikasikan bahwa SVM memiliki keandalan yang tinggi dalam menangani data teks berdimensi tinggi.

Meskipun efektivitas SVM telah banyak dibuktikan, mayoritas penelitian sebelumnya cenderung berfokus pada perbandingan kinerja antar-algoritma (misalnya SVM vs Naïve Bayes) atau optimasi parameter eksternal. Masih terdapat kesenjangan penelitian (research gap) di mana analisis mendalam mengenai pengaruh penggunaan jenis kernel internal SVM (seperti Linear, RBF, dan Polynomial) secara spesifik terhadap klasifikasi headline berita hoax berbahasa Indonesia masih jarang dilakukan. Padahal, pemilihan kernel yang tepat sangat krusial karena menentukan bagaimana data dipetakan dalam ruang fitur. Oleh karena itu, penelitian ini bertujuan untuk mengisi celah tersebut dengan membandingkan kinerja ketiga kernel SVM guna menemukan konfigurasi model yang paling optimal dan efisien untuk kasus ini.

Berdasarkan permasalahan tersebut, penelitian ini tidak hanya berfokus pada analisis performa algoritma di lingkungan eksperimen (Google Colab), tetapi juga ditekankan pada aspek implementasi. Hal ini krusial karena model klasifikasi yang memiliki akurasi tinggi tidak akan memberikan manfaat fungsional jika tidak diwujudkan dalam bentuk perangkat lunak yang dapat digunakan oleh masyarakat luas. Oleh karena itu, penelitian ini mengusulkan sebuah implementasi sistem deteksi hoaks berbasis web menggunakan framework React.js yang mengintegrasikan model SVM terbaik secara standalone. Dengan adanya implementasi ini, diharapkan hasil riset dapat langsung diterapkan untuk memvalidasi judul berita secara real-time oleh pengguna.



## 1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, maka permasalahan utama dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana tingkat akurasi perbandingan kernel Linear, RBF, dan Polynomial pada algoritma Support Vector Machine (SVM) untuk klasifikasi judul berita?
2. Bagaimana pengaruh pembobotan fitur TF-IDF terhadap kemampuan model dalam membedakan berita hoaks dan faktual?
3. Bagaimana mengimplementasikan model SVM terbaik ke dalam sebuah aplikasi berbasis web yang dapat melakukan prediksi secara instan dan mandiri (standalone)?

## 1.3 Batasan Masalah

Untuk menjaga agar penelitian ini tetap fokus, spesifik, dan dapat diselesaikan dengan baik, maka ruang lingkup penelitian dibatasi pada beberapa hal berikut:

1. Data: Data yang digunakan adalah teks judul (headline) berita berbahasa Indonesia. Penelitian ini tidak menganalisis isi (body) artikel berita secara keseluruhan.
2. Sumber Data: Data berita hoax dikumpulkan dari situs portal pemeriksa fakta (seperti Turnbackhoax.id), sementara data berita faktual dikumpulkan dari portal berita nasional terpercaya (seperti Kompas.com, Detik.com, dll.).
3. Ekstraksi Fitur: Metode ekstraksi fitur yang digunakan hanya TF-IDF (Term Frequency-Inverse Document Frequency). Penelitian ini tidak membandingkan dengan metode lain seperti Bag-of-Words atau Word2Vec.
4. Algoritma: Algoritma klasifikasi yang digunakan secara eksklusif adalah Support Vector Machine (SVM), sesuai dengan judul penelitian.

5. Kernel SVM: Analisis perbandingan kernel yang diuji dibatasi hanya pada tiga jenis, yaitu: Linear, RBF (Radial Basis Function), dan Polynomial.
6. Implementasi sistem dibangun menggunakan framework React.js dengan arsitektur client-side processing.

#### **1.4 Tujuan Penelitian**

Selaras dengan rumusan masalah yang telah dijabarkan, maka tujuan dari dilaksanakannya penelitian ini adalah sebagai berikut:

1. Membangun dan menerapkan alur kerja (pipeline) model Machine Learning untuk klasifikasi judul berita, mulai dari tahap pra-pemrosesan teks hingga penerapan metode ekstraksi fitur TF-IDF.
2. Menerapkan algoritma Support Vector Machine (SVM) untuk melatih model yang mampu mengklasifikasikan judul berita (headline) ke dalam kategori hoax atau faktual.
3. Menganalisis dan menyajikan hasil perbandingan kinerja—berdasarkan metrik akurasi, presisi, recall, dan F1-score—dari penerapan tiga kernel SVM (yaitu Linear, RBF, dan Polynomial).
4. Mengidentifikasi dan menentukan kernel SVM yang memberikan performa paling optimal untuk studi kasus klasifikasi berita hoax dan faktual pada dataset yang digunakan.
5. Membangun dan mengimplementasikan sistem aplikasi deteksi hoaks yang mampu melakukan klasifikasi secara real-time berdasarkan parameter model SVM yang telah dilatih.

#### **1.5 Manfaat Penelitian**

##### **1.5.1 Manfaat Teoritis (Akademis)**

1. Memberikan kontribusi bagi pengembangan studi di bidang Natural

Language Processing (NLP) dan Text Mining, khususnya untuk penanganan klasifikasi teks (judul berita) berbahasa Indonesia.

2. Menyajikan data empiris mengenai perbandingan kinerja kernel (Linear, RBF, Polynomial) pada algoritma SVM untuk studi kasus klasifikasi berita hoax, yang dapat menjadi referensi dan bahan perbandingan bagi penelitian selanjutnya.

### **1.5.2 Bagi Masyarakat**

1. Membangun dan mengimplementasikan sistem aplikasi deteksi hoaks yang mampu melakukan klasifikasi secara real-time berdasarkan parameter model SVM yang telah dilatih.

## **1.6 Sistematika Penulisan**

Untuk memberikan gambaran yang jelas dan terstruktur mengenai keseluruhan isi laporan skripsi ini, berikut adalah sistematika penulisan yang dibagi ke dalam lima bab utama:

### **1. BAB 1: PENDAHULUAN**

Bab ini berisi pendahuluan umum tentang penelitian, mencakup Latar Belakang Masalah yang menjelaskan maraknya berita hoax dan potensi SVM, Rumusan Masalah, Batasan Masalah, Tujuan Penelitian, Manfaat Penelitian (teoritis dan praktis), serta diakhiri dengan Sistematika Penulisan.

### **2. BAB 2: LANDASAN TEORI**

Bab ini menguraikan berbagai teori dan konsep yang menjadi dasar penelitian. Materi yang dibahas meliputi tinjauan Penelitian Terdahulu (State of the Art), definisi Klasifikasi Teks, konsep Berita Hoax, tahapan Pra-pemrosesan Teks, metode ekstraksi fitur TF-IDF, teori dasar Support Vector Machine (SVM) dan konsep Kernel Trick, serta metrik Evaluasi Model yang digunakan.

### 3. BAB 3: METODOLOGI PENELITIAN

Bab ini menjelaskan secara rinci langkah-langkah teknis yang dilakukan selama penelitian. Pembahasan mencakup Kerangka Kerja Penelitian (diagram alir), metode Pengumpulan Data, Alat dan Bahan (perangkat keras dan lunak) yang digunakan, serta Skenario Pengujian untuk setiap kernel SVM.

### 4. BAB 4: HASIL DAN PEMBAHASAN

Bab ini merupakan inti dari penelitian yang menyajikan temuan dari eksperimen. Bab ini menampilkan Hasil Pengumpulan Data, hasil Pra-pemrosesan Teks, dan hasil pengujian model untuk setiap kernel (Linear, RBF, Polynomial). Bagian utama dari bab ini adalah Pembahasan (Analisis Hasil), yang mencakup Tabel Perbandingan Kinerja Kernel untuk menentukan model terbaik.

### 5. BAB 5: PENUTUP

Bab ini merupakan bab penutup yang berisi Kesimpulan dari seluruh hasil penelitian untuk menjawab rumusan masalah, serta Saran yang berisi masukan dan rekomendasi untuk penelitian selanjutnya di masa mendatang.