

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Hoaks adalah kebenaran yang tidak relevan atau kabar bohong yang disebarkan secara sengaja untuk mempengaruhi banyak orang agar dapat menodai suatu citra dan kredibilitas. Hoaks bertujuan mempengaruhi pembaca untuk mengambil tindakan sesuai dengan informasi palsu yang didapat. Informasi palsu dan menyesatkan ini berbahaya karena dapat menyesatkan persepsi manusia dengan menyampaikan informasi yang tidak benar sebagai sebuah kebenaran sehingga sangat memungkinkan untuk membawa dampak pada pemikiran manusia. Dengan begitu pesat dan cepatnya perkembangan informasi yang berkembang di dunia digital, maka berita hoax juga akan bertambah dengan sangat signifikan. Oleh karena itu diperlukan sistem yang dapat melakukan klasifikasi berita hoaks secara otomatis. Beberapa tahun belakangan ini, untuk membantu pengguna *online* mengidentifikasi informasi yang memiliki nilai dan manfaat, telah ada penelitian yang luas tentang pembentukan sebuah *framework* yang otomatis dan efektif dalam mendeteksi berita hoaks *online* [1].

Saat ini ada beberapa perkembangan penelitian tentang pengklasifikasian berita hoaks. Pada penelitian Azizah Zain tentang identifikasi hoaks menggunakan *Naïve Bayes Classifier* (NBC) menyimpulkan bahwa NBC dengan tambahan seleksi fitur Mutual Information (MI) dan Information Gain (IG) dapat menghasilkan nilai *F1-score* sebesar 0.9433 sedangkan NBC tanpa seleksi fitur menghasilkan *F1-score* sebesar 0.8026, dengan kata lain seleksi fitur dapat menambah nilai *F1-score* sebesar 0.1408[1]. Pada penelitian Hanifan Aulia tentang perbandingan metode pada klasifikasi berita hoaks tentang *COVID-19* didapatkan nilai akurasi tertinggi yaitu 0,87198. Pada penelitian ini juga mengatakan data yang berbentuk teks menyebabkan data memiliki dimensi yang tinggi dan fitur yang sangat banyak. Fitur yang banyak akan memakan waktu yang lama serta tidak efisien[2].

Dalam penelitian Albert Bill Alroy yang mengklasifikasi hoaks menggunakan metode seleksi fitur berdasarkan *Information Gain*, disimpulkan bahwa seleksi fitur berdasarkan *Information Gain* memiliki pengaruh dalam klasifikasi hoaks dengan menerapkan *Maximum Entropy*. Hal ini dibuktikan dengan meningkatnya hasil akurasi klasifikasi hoaks dengan akurasi sebesar 0,8 dibandingkan jika tidak menggunakan seleksi fitur menghasilkan akurasi sebesar 0,78[3]. Dengan begitu seleksi fitur perlu dilakukan untuk dapat menghilangkan fitur-fitur yang tidak relevan yang akan menjadi gangguan apabila disimpan. Dalam penelitian Inoshika dilrukshi mengatakan dengan menghapus kata-kata yang tidak relevan dari ruang fitur, dimensinya dapat dikurangi dan karenanya, jaranganya dapat dikurangi. Kata-kata yang tidak relevan dapat didefinisikan sebagai kata-kata umum (kata-kata yang sering muncul) dan kata-kata noise (kata-kata yang jarang muncul). Meskipun dengan menghapus kata-kata yang tidak relevan ini, masih mungkin mengandung beberapa kata yang tidak relevan. Oleh karena itu, diperlukan teknik seleksi fitur untuk memilih set fitur terbaik[4].

Pada penelitian oleh Galuh Fadillah Grandis tentang Seleksi Fitur *Gain Ratio* pada Analisis Sentimen Kebijakan Pemerintah didapatkan hasil pengujian *f-measure* mendapatkan nilai 0,74 sedangkan hasil dari pengujian menggunakan seleksi fitur *Information Gain* mendapatkan nilai *f-measure* paling tinggi berada di angka 0,77 akan tetapi, dari beberapa percobaan *gain ratio* menghasilkan nilai *F-measure* yang lebih stabil dibandingkan dengan *information gain*[5]. Salah satu masalah utama dalam klasifikasi teks adalah tingginya dimensi ruang fitur. Sebagian besar fitur tidak relevan, berisik, dan berlebihan, yang dapat menurunkan akurasi sistem oleh karena itu, seleksi fitur diperlukan [6]. Dalam penelitian yang bertujuan untuk meneliti penggunaan pemilihan fitur menggunakan *gain ratio* pada kasus pemilihan teks untuk mengetahui hasil presentase akurasi yang didapatkan dalam mendeteksi kebenaran pada suatu berita ini, akan digunakan gabungan judul dan isi berita ini sebagai penambah informasi pada model. Penggunaan judul beserta isi artikel akan menyebabkan besarnya dimensi dari dataset, oleh sebab itu dilakukan seleksi fitur untuk mengurangi dimensi dari *input* yang dilatih tanpa mengurangi dan bahkan meningkatkan performa model[7].

Oleh karena itu dalam penelitian ini dilakukan Seleksi fitur menggunakan gain ratio yang dimana metode ini akan menghilangkan informasi yang dianggap tidak relevan. Tetapi tidak menutup kemungkinan metode ini juga akan menghilangkan informasi yang relevan. Oleh sebab itu perlu mengetahui batas yang tepat agar informasi yang relevan hilang secara efektif tanpa menghilangkan informasi yang relevan.

1.2 Rumusan Masalah

Berdasarkan permasalahan yang telah dijelaskan pada latar belakang maka Rumusan masalah dari penelitian ini adalah

- 1) Bagaimana cara mendeteksi berita hoaks secara otomatis ?
- 2) Bagaimana pengaruh seleksi fitur berdasarkan Gain Ratio terhadap akurasi dalam klasifikasi berita hoaks?

1.3 Maksud dan Tujuan

Adapun maksud dari penelitian ini adalah untuk mengimplementasikan pemilihan fitur dengan pembobotan *Gain Ratio* dalam mendeteksi kebenaran suatu berita. Penelitian ini memiliki tujuan sebagai berikut

- 1) Membuat model prototype Klasifikasi Berita Hoaks melalui Seleksi Fitur Menggunakan *Gain Ratio*
- 2) Mengukur penggunaan pemilihan fitur dengan *Gain Ratio* dalam mengukur fitur yang relevan untuk mengklasifikasi sebuah berita,

1.4 Batasan Masalah

Berdasarkan latar belakang dan rumusan masalah maka batasan masalah dalam penelitian ini yaitu:

1. *Dataset* yang digunakan berupa data berita Politik, ekonomi, kesehatan
2. Pengujian sistem pendeteksi dilakukan dengan *offline*.
3. *Dataset* yang dipakai menggunakan Bahasa Indonesia
4. Metode klasifikasi yang digunakan adalah *Naïve Bayes*

1.5 Metode Penelitian

Tahapan-tahapan yang dilakukan dalam pelaksanaan penelitian ini adalah sebagai berikut:

1. Studi Literatur

Penelitian ini dimulai dengan melakukan studi pustaka yaitu dengan cara mengumpulkan bahan-bahan referensi dari berbagai sumber baik berupa buku, artikel, jurnal serta jurnal mengenai pemilihan fitur menggunakan Gain Ratio untuk prediksi hoaks pada berita

2. Identifikasi

Tahap identifikasi melibatkan pengumpulan data berita yang akan digunakan dalam penelitian ini. Data berita harus dipilih dengan cermat untuk mencakup berbagai jenis hoaks dan berita sah, serta harus mencakup fitur-fitur yang relevan untuk analisis menggunakan metode *Gain Ratio*.

3. Analisis Masalah

Pada tahap ini dilakukan analisis permasalahan yang ada berupa analisis kebutuhan implementasi algoritma

4. Perancangan dan implementasi

Pada tahap ini mulai dilakukan perancangan dan pembuatan model deteksi berita hoaks dan mengimplementasikan model menjadi sebuah prototype yang akan digunakan dalam proses pengujian

5. Pengujian

Pada tahap ini akan dilakukan serangkaian pengujian dari model yang sudah dibuat dengan menggunakan data uji untuk mengetahui akurasi dan performa model. Pengujian dilakukan untuk mengevaluasi kinerja metode pemilihan fitur menggunakan Gain Ratio dalam mendeteksi hoaks pada berita. Evaluasi dilakukan menggunakan metrik-metrik evaluasi yang relevan seperti, *presicion*, *recall*, dan *f1-score*.

1.6 Sistematika Penulisan

Sistematika penulisan penelitian skripsi yang akan dilakukan terdiri dari beberapa bagian yaitu sebagai berikut :

BAB 1 PENDAHULUAN

Pada bab ini membahas uraian mengenai latar belakang masalah penelitian, Rumusan masalah, maksud dan tujuan, batasan masalah, metodologi penelitian, dan sistematika penulisan.

BAB 2 TINJAUAN PUSTAKA

Pada bab ini membahas mengenai teori-teori pendukung yang menjadi landasan penelitian ini. Adapun teori yang dibahas terkait Berita Hoaks, sistem Pendeteksi, *Text Preprocessing*, klasifikasi berita hoaks dan non hoaks, dan perangkat lunak yang digunakan.

BAB 3 ANALISIS KEBUTUHAN IMPLEMENTASI ALGORITMA

Pada bab ini berisi deskripsi umum masalah, data set berita, Rancangan Model deteksi untuk menghasilkan model prototype.

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Pada bab ini dilakukan implementasi sistem berdasarkan hasil analisis dan perancangan, kedalam bahasa pemrograman dan *tools* yang telah ditetapkan

BAB 5 KESIMPULAN DAN SARAN

Bab ini membahas tentang kesimpulan yang ditarik dari hasil pengujian yang telah dilakukan serta saran pengembangan ilmu untuk penelitian di masa yang akan datang.

