

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Penyebaran berita hoaks di Indonesia telah menjadi masalah serius dalam beberapa tahun terakhir, terutama dengan semakin berkembangnya penggunaan media sosial. Informasi palsu dapat menyebar dengan cepat dan sulit untuk dikendalikan, mengakibatkan kebingungan dan disinformasi di kalangan masyarakat [1]. Banyak berita hoaks yang menyentuh isu-isu sensitif, seperti politik, kesehatan, atau keamanan, yang dapat memicu keresahan publik. Hal ini menimbulkan tantangan besar dalam hal penyaringan dan identifikasi berita yang valid.

Seiring dengan meningkatnya ancaman berita hoaks, teknologi mulai memainkan peran penting dalam membantu mengatasi masalah ini. Salah satu teknologi yang telah banyak digunakan untuk klasifikasi berita adalah *machine learning*. *Machine learning* memungkinkan komputer untuk belajar dari data, mengidentifikasi pola, dan membuat keputusan tanpa diprogram secara eksplisit. Dalam konteks berita hoaks, model *machine learning* dapat digunakan untuk mempelajari karakteristik teks berita yang mengandung informasi palsu dan membedakannya dari berita yang benar.

Salah satu pendekatan yang sederhana namun efektif dalam *machine learning* adalah model linear seperti *Logistic Regression* dan *Ridge Classifier*. *Logistic Regression* adalah model probabilistik yang digunakan untuk mengklasifikasikan data ke dalam dua kelas berdasarkan hubungan linear antara fitur masukan. Model ini telah terbukti efisien untuk tugas-tugas klasifikasi teks, termasuk deteksi berita hoaks, karena kemampuannya dalam menangkap pola sederhana dalam data yang tinggi dimensinya [2, 3].

Di sisi lain, *Ridge Classifier*, yang didasarkan pada prinsip regresi linear dengan regularisasi L_2 , memberikan keunggulan dalam menghadapi data yang mengalami multikolinearitas atau *overfitting*. Dengan menambahkan penalti pada koefisien model, *Ridge Classifier* mampu menghasilkan prediksi yang lebih stabil dan generalisasi yang lebih baik, terutama pada data teks yang kompleks seperti berita hoaks berbahasa Indonesia [4].

Pada penelitian ini, fokus utama adalah penerapan *Logistic Regression*

dan *Ridge Classifier* untuk klasifikasi berita hoaks berbahasa Indonesia. Berita hoaks dalam bahasa Indonesia memiliki karakteristik unik yang tidak selalu dapat ditangkap oleh model-model standar yang dirancang untuk bahasa lain. Oleh karena itu, penting untuk mengeksplorasi model linear yang tidak hanya efisien, tetapi juga dapat menangani struktur dan kompleksitas data bahasa Indonesia secara spesifik. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam upaya mitigasi penyebaran berita hoaks di Indonesia.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, dapat diidentifikasi beberapa rumusan masalah yang menjadi fokus dalam penelitian ini, yaitu sebagai berikut

1. Bagaimana mengembangkan dan mengimplementasikan model *Logistic Regression* dan *Ridge Classifier* yang efektif untuk mendeteksi dan mengklasifikasikan berita hoaks berbahasa Indonesia?
2. Bagaimana akurasi dan performa model *Logistic Regression* dan *Ridge Classifier* dalam mendeteksi berita hoaks berbahasa Indonesia?
3. Bagaimana menerapkan metodologi penelitian ilmiah secara sistematis dalam pengembangan sistem deteksi hoaks berbahasa Indonesia?

1.3 Tujuan Penelitian

Selama pelaksanaan penelitian diperlukan maksud dan tujuan dijabarkan sebagai berikut.

1. Untuk mengembangkan dan mengimplementasikan model *Logistic Regression* dan *Ridge Classifier* yang efektif dalam mendeteksi dan mengklasifikasikan berita hoaks berbahasa Indonesia.
2. Untuk Mengevaluasi performa model menggunakan berbagai metrik evaluasi seperti accuracy, precision, recall, dan F1-score.
3. Untuk mempelajari dan menerapkan metodologi penelitian ilmiah secara sistematis dalam konteks pengembangan sistem deteksi hoaks.

1.4 Batasan Masalah

Penelitian ini memiliki batasan-batasan yang ditetapkan untuk memastikan fokus yang jelas dan hasil yang dapat diukur. Adapun batasan-batasan yang diterapkan adalah sebagai berikut:

1. Penelitian ini hanya akan memfokuskan pada berita hoaks dalam bahasa Indonesia, sehingga tidak mencakup berita hoaks dalam bahasa lain atau isu-isu internasional.
2. Data yang digunakan dalam penelitian ini diperoleh dari Kompas, Tempo, dan Mafindo sebagai media berita dari Indonesia. Penelitian ini tidak mencakup berita dari media cetak atau sumber lain yang tidak tersedia secara daring.
3. Model yang digunakan akan dibatasi pada *Logistic Regression* dan *Ridge Classifier*.

1.5 Urgensi Penelitian

1. Penyebaran berita hoaks di Indonesia membutuhkan solusi teknologi yang efektif untuk mendeteksi dan mengklasifikasikan berita palsu secara akurat.
2. Kebutuhan akan sistem deteksi hoaks yang mampu memahami karakteristik unik berita berbahasa Indonesia.
3. Perlunya eksplorasi dan implementasi model *machine learning* yang dapat menangani kompleksitas data bahasa Indonesia secara spesifik untuk meningkatkan akurasi deteksi berita hoaks.

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

1.6 Manfaat Penelitian

1. Memberikan pengalaman praktis bagi mahasiswa dalam menerapkan konsep *machine learning* untuk menyelesaikan masalah nyata di masyarakat.
2. Menghasilkan model klasifikasi yang dapat digunakan untuk mendeteksi berita hoaks berbahasa Indonesia secara lebih efektif dan efisien.
3. Memberikan kontribusi pada pengembangan sistem deteksi hoaks yang dapat membantu masyarakat dalam mengidentifikasi berita palsu.
4. Memperkaya penelitian tentang penggunaan model *Logistic Regression* dan *Ridge Classifier* dalam konteks klasifikasi teks berbahasa Indonesia.
5. Menyediakan dasar pengetahuan dan metodologi yang dapat digunakan untuk pengembangan sistem deteksi hoaks yang lebih kompleks di masa depan.
6. Menghasilkan dokumentasi dan referensi yang dapat bermanfaat bagi penelitian selanjutnya dalam bidang deteksi hoaks dan pengolahan bahasa alami.

