

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam era informasi digital yang semakin berkembang, akses terhadap berita dan informasi menjadi lebih mudah melalui platform online, situs berita, dan media sosial. Namun, bersamaan dengan akses ini, kita juga dihadapkan pada masalah serius berupa penyebaran berita *hoax* atau palsu. Berita *hoax* adalah informasi yang tidak akurat atau tidak benar yang disebar dengan maksud menyesatkan atau mempengaruhi opini masyarakat (Rahmadhany, Aldila Safitri, and Irwansyah 2021). Fenomena ini memiliki dampak negatif yang signifikan, termasuk menimbulkan kebingungan, memperdalam polarisasi masyarakat, serta dapat merugikan individu dan institusi yang menjadi sasaran berita *hoax* (Rahutomo, Pratiwi, and Ramadhani 2019).

Dalam konteks inilah pentingnya pengembangan sistem yang mampu mengidentifikasi berita *hoax* secara efisien dan akurat menjadi semakin nyata. Dengan banyaknya jumlah berita yang dipublikasikan setiap hari, tugas ini menjadi semakin kompleks dan menantang. Inilah tempat potensi penerapan metode *machine learning* dalam analisis data berita.

Metode *machine learning* telah terbukti berhasil analisis teks, klasifikasi teks. Dengan memanfaatkan teknik-teknik ini, menganalisis berita dengan cepat, pola-pola yang mengindikasikan adanya berita *hoax*, dan membedakannya dari berita yang sah.

Melalui penelitian ini, tujuan utama penelitian ini adalah mengembangkan model *machine learning* yang dapat mengidentifikasi berita *hoax* dengan tingkat akurasi yang tinggi. Dalam rangka mencapai tujuan ini, penelitian ini akan melakukan analisis mendalam terhadap data berita yang mencakup berbagai karakteristik, seperti struktur teks, penggunaan kata kunci tertentu, sentimen yang terkandung dalam teks, dan faktor-faktor lain yang mungkin berkontribusi terhadap deteksi berita *hoax*.

Dengan menerapkan teknik-teknik *machine learning* seperti klasifikasi dan ekstraksi fitur, penelitian ini akan mengembangkan model yang dapat membedakan berita *hoax* dan berita sah. Selain itu, penelitian ini juga akan melakukan evaluasi mendalam terhadap kinerja model penelitian ini dengan menggunakan metrik evaluasi standar seperti akurasi, presisi, *recall*, dan f1-score (Tambunan, Nataliani, and Lestari 2021).

Hasil dari penelitian ini diharapkan dapat memberikan panduan dan wawasan bagi pihak-pihak yang terlibat dalam penyediaan dan konsumsi informasi, termasuk media, platform berita, dan masyarakat umum. Dengan pengembangan sistem yang efektif dalam mengidentifikasi berita *hoax*, dapat membantu melindungi integritas informasi, mencegah penyebaran berita palsu, dan membangun lingkungan informasi yang lebih dapat dipercaya di era digital yang penuh dengan tantangan ini.

1.2 Identifikasi Masalah

Terdapat perolehan identifikasi masalah dari latar belakang yang sudah di paparkan diatas, maka identifikasi masalah dalam penelitian ini yaitu:

1. Penyebaran Berita *Hoax*: Penyebaran berita *hoax* di media sosial dan platform berita daring semakin meresahkan masyarakat. Dampak dari berita *hoax* dapat berupa kebingungan publik, kerugian finansial, dan bahkan ancaman terhadap stabilitas sosial.
2. Keterbatasan Deteksi Manusia: Identifikasi berita *hoax* oleh manusia memerlukan waktu dan upaya yang signifikan. Keterbatasan ini membuat proses identifikasi lambat dan sulit dilakukan secara efisien, terutama dengan volume besar berita yang dipublikasikan setiap hari.
3. Diversifikasi Taktik *Hoax*: Pelaku yang menyebarkan berita *hoax* semakin canggih dalam merancang konten yang meyakinkan. Taktik diversifikasi dan modifikasi konten *hoax* membuatnya sulit dikenali oleh metode analisis tradisional.
4. Kompleksitas Analisis *Teks*: Analisis berita teks memerlukan pemahaman yang mendalam tentang bahasa dan konteks. Identifikasi berita *hoax* berdasarkan analisis teks memerlukan pendekatan yang canggih untuk mengenali pola dan kecenderungan *subtil* yang mungkin mengindikasikan adanya berita *hoax*.
5. Ketergantungan pada Data Pelatihan: Kualitas dan kuantitas data pelatihan berpengaruh besar pada performa model *machine learning*. Keterbatasan data pelatihan yang tersedia dapat membatasi kemampuan model dalam

mengidentifikasi berita *hoax*.

6. Pergeseran Pola *Hoax*: Pola penyebaran berita *hoax* dapat berubah seiring waktu. Model *machine learning* perlu mampu mengenali pergeseran pola ini agar tetap efektif dalam mengidentifikasi berita *hoax*.
7. Tingkat Akurasi yang Diinginkan: Meskipun model *machine learning* dapat membantu mengidentifikasi berita *hoax*, tingkat akurasi yang tinggi diperlukan agar model tersebut benar-benar efektif dalam mencegah penyebaran berita *hoax*.

1.3 Batasan Masalah

Untuk Menghindari meluasnya permasalahan dalam penelitian ini, hingga batas permasalahan dalam permasalahan dalam penelitian ini antara lain:

1. Tipe Berita: Penelitian akan membatasi analisis pada berita teks yang berasal dari sumber berita terpercaya, dengan fokus utama pada berita detik.com dan *turnbackhoax.id*.
2. Konteks Berita *Hoax*: Penelitian ini akan berfokus pada mengidentifikasi berita *hoax* yang terkait dengan isu berita, khususnya yang terdapat dalam sumber berita yang telah disebutkan.
3. Algoritma *Machine learning*: Penelitian akan menerapkan algoritma klasifikasi teks, seperti *Logistic Regression*, *Decision Tree*, *Gradient Boosting*, *Random Forest* untuk mengidentifikasi berita *hoax* dalam konteks berita detik.com dan *turnbackhoax.id*.
4. Pengujian Model: Pengujian performa model akan dilakukan dengan

menggunakan metrik akurasi, presisi, *recall*, dan *f1-score*. Evaluasi model ini akan merujuk pada data manusia yang telah terlatih, dengan fokus pada sumber berita yang dijelaskan sebelumnya.

1.4 Rumusan Masalah

Berdasarkan latar belakang tersebut peneliti merumuskan pertanyaan yaitu:

1. Efektivitas Teknik *Crawling*

Sejauh mana teknik *crawling* efektif dalam mengambil data dari berbagai sumber, termasuk situs web, media sosial, atau sumber lainnya, untuk analisis berita *hoax*?

2. Kualitas *Preprocessing* Data

Bagaimana kualitas *preprocessing* data memengaruhi pembersihan data dari noise dan normalisasi data untuk mempersiapkan data guna digunakan dalam model *machine learning* identifikasi berita *hoax*?

3. Optimalitas Pelabelan Data

Bagaimana pelabelan data dapat dioptimalkan untuk kategori atau klasifikasi yang sesuai dengan tujuan penelitian, khususnya dalam konteks mengidentifikasi berita *hoax*? Apakah pembagian data sebagai *training* atau testing dilakukan secara optimal?

4. Kinerja Model *Machine learning*

Seberapa baik model *machine learning*, seperti *Logistic Regression*, *Decision Tree*, *Gradient Boosting*, dan *Random Forest*, dapat mempelajari pola dari data *training* dan mengklasifikasikan data testing dalam mengenali berita *hoax*?

5. Evaluasi Akurasi Model

Bagaimana hasil evaluasi akurasi model setelah dilatih dan diuji dengan data testing, dengan menggunakan metrik seperti akurasi, presisi, *recall*, dan *f1-score*?

1.5 Tujuan Penelitian

Adapun penelitian dengan tujuan yang akan dilakukan adalah:

1. Mengembangkan Model Deteksi Berita *Hoax*: Tujuan utama penelitian ini adalah mengembangkan model deteksi berita *hoax* berbasis *machine learning* yang mampu mengidentifikasi berita *hoax* dengan tingkat akurasi yang tinggi.
2. Menganalisis Fitur *Teks*: Penelitian ini bertujuan untuk menganalisis berbagai fitur teks yang dapat digunakan sebagai penanda berita *hoax*, termasuk kata kunci, entitas, struktur kalimat, dan sentimen.
3. Mengukur Kinerja Model: Penelitian ini bertujuan untuk mengukur dan menganalisis kinerja model *machine learning* dalam mengklasifikasikan berita *hoax*, dengan menggunakan metrik akurasi, presisi, *recall*, dan *f1-score*.
4. Mengidentifikasi Pengaruh Data Pelatihan: Tujuan ini adalah untuk memahami bagaimana jumlah dan kualitas data pelatihan memengaruhi kinerja model dalam mengenali berita *hoax*.
5. Mengenali Pergeseran Pola *Hoax*: Penelitian ini bertujuan untuk mengidentifikasi sejauh mana model mampu mengenali perubahan pola penyebaran berita *hoax* seiring waktu.

1.6 Manfaat Penelitian

Manfaat yang bisa di peroleh dari penelitian ini ialah:

1. Peningkatan Kepercayaan Informasi, Dengan mengembangkan model yang efektif dalam mengenali berita *hoax*, penelitian ini dapat membantu meningkatkan kepercayaan masyarakat terhadap informasi yang mereka terima dari berbagai sumber.
2. Pengembangan Teknik Analisis Teks, Penelitian ini dapat memajukan pengembangan teknik analisis teks dan penerapan *machine learning* dalam konteks analisis berita, memberikan pandangan baru dalam mengatasi tantangan berita *hoax*.
3. Kontribusi terhadap Literatur Penelitian, Penelitian ini dapat menjadi tambahan berharga dalam literatur ilmiah tentang deteksi berita *hoax* dan penerapan *machine learning* dalam analisis teks.
4. Pedoman Bagi Platform Berita dan Media Sosial, Hasil penelitian ini dapat digunakan sebagai pedoman bagi platform berita dan media sosial untuk mengembangkan algoritma deteksi berita *hoax* guna menjaga integritas informasi.