

BAB I

PENDAHULUAN

1.1. LATAR BELAKANG

Hoax merupakan berita yang menyesatkan dan bisa saja berbeda dari pemahaman seseorang karena menyebarkan informasi palsu seolah-olah benar. Keberadaannya dapat merusak reputasi, mengurangi kepercayaan, dan memengaruhi banyak orang. *Hoax* sering kali sulit dikenali karena tidak memiliki sumber yang jelas dan dapat dipercaya. Biasanya, penyebaran *hoax* telah dirancang oleh pihak tertentu demi mencapai kepentingan pribadi [1].

Penyebaran *hoax* semakin mudah terutama di era digital saat ini, dengan jumlah pengguna internet di Indonesia yang mencapai 215,63 juta orang pada 2022-2023. Berdasarkan survei APJII, jumlah ini naik 2,67% atau 5,6 juta pengguna dari periode sebelumnya. Angka tersebut mencakup 78,19% dari total populasi Indonesia yang mencapai 275,77 juta jiwa. Tingkat penetrasi internet pun meningkat 1,17% dari 77,02% pada 2021-2022 [2].

Kemajuan teknologi komunikasi yang modern memungkinkan banyak hal menjadi lebih praktis. Internet misalnya, kini dapat menyediakan berbagai jenis informasi seperti politik, ekonomi, sosial, dan budaya hanya dengan bantuan mesin pencari. Interaksi dan transaksi informasi kini dapat dilakukan tanpa harus meninggalkan rumah, yang tentunya menghemat waktu, tenaga, serta biaya [3].

Namun, kemudahan ini membawa risiko, terutama dalam hal keaslian informasi. Beberapa informasi yang tersebar di internet tidak selalu mencerminkan kebenaran, sering kali berupa *hoax* yang dapat menyebabkan kebingungan di masyarakat dan merusak reputasi individu atau kelompok [4].

Dalam lima tahun terakhir, dari Agustus 2018 hingga akhir 2023, Kementerian Komunikasi dan Informatika (Kominfo) telah mengidentifikasi sebanyak 12.547 isu *hoax*. Tim AIS Kominfo mendapati bahwa 1.628 dari isu-isu tersebut terkait dengan kategori politik. Konten *hoax* ini didominasi oleh informasi yang berkaitan dengan partai politik, kandidat, dan proses pemilihan umum [5].

Penelitian oleh Timothy M. dkk. [6] mengatakan bahwa ada sekitar 800.000 situs di Indonesia yang diduga menjadi penyebar *hoax*. Dengan angka yang begitu besar, maka semakin sulit menyaring informasi yang benar di internet yang mana hal ini menunjukkan *urgensi* dari penelitian yang bertujuan untuk mendeteksi *hoax* sebelum informasi palsu tersebar beredar secara luas.

Dalam beberapa tahun belakangan, telah dilakukan penelitian untuk menilai kinerja metode klasifikasi pada berita *hoax* dengan menerapkan pendekatan *Machine Learning*. Hal ini dilakukan untuk mencari solusi yang efektif dalam mendeteksi dan mengatasi berita *hoax*. Sejumlah penelitian sebelumnya [7],[8],[9] telah mengadopsi pendekatan *Machine Learning* untuk mengevaluasi sejauh mana algoritma *Machine Learning* dapat mendeteksi berita *hoax*.

Beberapa penelitian sebelumnya yaitu penelitian oleh Fatahillah Arsyad, dkk [7], yang berjudul “Deteksi Konten *Hoax* Pada Media Berita Indonesia Menggunakan

Multinomial Naïve Bayes". Pada penelitian ini, algoritma Gaussian Naive Bayes (GNB) akan dibandingkan dengan *Multinomial Naïve Bayes (MNB)* kemudian hasil yang diperoleh menunjukkan *MNB* menunjukkan akurasi yang lebih baik dibanding dengan *GNB* dengan akurasi *MNB* mencapai 96%.

Kemudian penelitian yang dilakukan oleh Karima, dkk [8] dengan judul "*Comparison of Classification Algorithms for Predicting Indonesian Fake News using Balanced and Imbalanced Datasets*", menggunakan empat algoritma, yaitu *Logistic Regression*, *Naïve Bayes*, *Random Forest*, dan *Support Vector Machine (SVM)* dengan hasil penelitian ini yang menunjukkan bahwa algoritma *SVM* memberikan performa terbaik dalam mendeteksi berita *hoax* dibandingkan metode lainnya, dengan akurasi mencapai 85,47%.

Selanjutnya, penelitian oleh Liliek Triyono, dkk [9] yang berjudul "*Indonesian Fake News Detection Using Various Machine Learning Techniques*", dengan melakukan perbandingan dengan beberapa algoritma *machine learning*. Dalam penelitian ini lima algoritma diuji, yaitu *SVM*, *Logistic Regression*, *Decision Tree*, *Gradient Boosting*, dan *Random Forest*. Hasilnya menunjukkan bahwa *SVM* memiliki akurasi tertinggi sebesar 83,55%, sedangkan *Decision Tree* terendah di 75,33%.

Berdasarkan hasil penelitian sebelumnya, penelitian ini akan melakukan perbandingan kinerja antara algoritma *Multinomial Naive Bayes* dan *Support Vector Machine* dalam mengklasifikasi berita *hoax* karena kedua algoritma ini terbukti memiliki akurasi yang baik dalam beberapa klasifikasi, termasuk klasifikasi berita *hoax*. penelitian ini juga akan menggunakan teknik pembobotan *TF-IDF* karena

mampu meningkatkan *classification report* (*precision* dan *recall*) dengan baik dan dapat mengurangi ukuran *term* pada dataset yang digunakan. Kemudian kinerja kedua algoritma ini akan dievaluasi menggunakan *Confusion Matrix* dengan menghitung akurasi, *recall*, presisi, dan *f1-score*.

Implementasi pada penelitian ini akan banyak memanfaatkan penggunaan *library* dari *scikit-learn* pada bahasa pemrograman *python* yang menyediakan berbagai algoritma *machine learning*. Berdasarkan pemaparan diatas, maka peneliti akan melakukan penelitian dengan judul” **PERBANDINGAN METODE MULTINOMIAL NAÏVE BAYES DAN SUPPORT VECTOR MACHINE DALAM MENGLASIFIKASI BERITA HOAX POLITIK INDONESIA**”.

1.2. RUMUSAN MASALAH

Adapun rumusan masalah dalam penelitian berdasarkan uraian latar belakang adalah:

1. Bagaimana kinerja antara metode *Multinomial Naïve Bayes* dengan *Support Vector Machine* dalam mengklasifikasi berita *hoax* menggunakan ekstraksi fitur *TF-IDF* ?
2. Bagaimana kinerja antara kedua metode dalam hal akurasi, *recall*, *precision*, dan *F1-score* dalam mengklasifikasi berita *hoax*?

1.3. BATASAN MASALAH

Agar tidak meluas permasalahan yang ada dan lebih terarah maka dilakukan pembatasan masalah sebagai berikut:

1. Evaluasi dilakukan dengan mengukur kinerja model berdasarkan metrik akurasi, *precision*, *recall*, dan *F1-score*.
2. Evaluasi performa menggunakan *confusion matrix*
3. Alat analisis yang digunakan meliputi *Python*, *Scikit-learn*, *Pandas*, *NumPy*, serta penggunaan fitur ekstraksi seperti *TF-IDF*.
4. Ekstraksi fitur dilakukan menggunakan *Term Frequency-Inverse Document Frequency (TF-IDF)*.
5. Fokus hanya pada dua algoritma klasifikasi, yaitu *Multinomial Naïve Bayes* dan *Support Vector Machine*, dan tidak membahas metode klasifikasi lainnya.
6. Penelitian ini akan menggunakan *library scikit-learn* dalam pemrograman *Python* untuk implementasi algoritma dan teknik pembobotan.
7. Data *preprocessing* terbatas pada *case folding*, *cleansing*, *tokenizing*, *filtering*, *normalization* dan *stemming*.

1.4. TUJUAN DAN MANFAAT PENELITIAN

Adapun tujuan dan manfaat dari penelitian tugas akhir yang dilakukan akan dijelaskan sebagai berikut:

1.4.1. Tujuan Penelitian

Adapun tujuan yang ingin dicapai pada penelitian tugas akhir ini adalah :

1. Memahami karakteristik dataset *Indonesian Fact and Hoax Political News* yang diperoleh dari *website Kaggle*.

2. Menguji performa model *Multinomial Naive Bayes* dan *Support Vector Machine* dalam mengklasifikasi berita *hoax* dengan menerapkan berbagai teknik *text preprocessing* dan ekstraksi fitur.
3. Mengevaluasi kinerja kedua model berdasarkan empat skenario pembagian data menggunakan metrik *accuracy, precision, recall, F1-score*
4. Menentukan model yang lebih unggul dalam mengklasifikasi berita *hoax* berdasarkan 4 skenario pembagian data.

1.4.2. Manfaat Penelitian

Manfaat yang didapat dari penelitian ini, yaitu :

1. Memberikan wawasan tentang struktur dan karakteristik dataset *Indonesian Fact and Hoax Political News* yang diperoleh dari *website Kaggle*.
2. Memaparkan tahapan *Prerprocessing* teks dan ekstraksi fitur sebelum melakukan proses klasifikasi.
3. Menunjukkan perbandingan performa *MNB* dan *SVM* dalam berbagai skenario pembagian data.
4. Memberikan rekomendasi model yang paling unggul dari ke 4 skenario pembagian data.

1.5. SISTEMATIKA PENULISAN

Agar lebih mudah dalam melakukan pemahaman pada Laporan Tugas Akhir, maka terbentuklah sistematika penulisan yang akan diterapkan meliputi :

BAB I : PENDAHULUAN

Bab ini membahas tentang hal umum dalam penelitian, seperti latar belakang, penelitian, latar belakang, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, dan sistematika penulisan yang menjadi dasar dilaksanakannya penelitian.

BAB II : LANDASAN TEORI

Bab landasan teori ini berisi teori-teori yang digunakan juga beberapa sumber yang berkaitan langsung dengan isu yang akan diatasi. Sementara itu, teori yang akan dibahas meliputi *machine learning*, klasifikasi, algoritma *Naïve Bayes* dan *Support Vector Machine*, berita hoax, *NLP*, *Google Colab*, *Python*, *TF-IDF*, *Scikit-Learn*, Evaluasi model dengan *Confusion Matrix* serta penelitian sejenis.

BAB III : METODOLOGI PENELITIAN

Pada bab ini menjelaskan tentang kerangka kerja penelitian, metode pengumpulan data, metode klasifikasi, serta alat bantu yang digunakan pada penelitian ini.

BAB IV : ANALISIS DAN HASIL

Pada bab ini membahas proses perhitungan dan analisis menggunakan bahasa pemrograman *python* dengan Algoritma *Naïve*

Bayes dan *Support Vector Machine* terhadap dataset ‘berita hoax politik indonesia’. Selanjutnya, melakukan pengujian akurasi, *precision*, *recall* dan *f1-score* untuk mengevaluasi peforma kedua algortima.

BAB V : PENUTUP

Bab ini berisikan kesimpulan dari apa yang telah dilakukan oleh peneliti dan saran dari pembaca untuk pengembangan penelitian.

