

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Dalam era digital yang semakin berkembang, informasi dapat dengan mudah tersebar melalui berbagai platform media sosial dan situs berita *online* [1]. Namun, kemudahan akses terhadap informasi ini juga menimbulkan tantangan baru, yaitu penyebaran berita hoaks [2]. Hoaks merupakan informasi yang direkayasa untuk menutupi informasi sebenarnya. Dengan kata lain hoaks juga bisa diartikan sebagai upaya pemutarbalikan fakta menggunakan informasi yang seolah-olah meyakinkan tetapi tidak dapat diverifikasi kebenarannya [3]. Berita hoaks memiliki dampak terhadap masyarakat meliputi kebingungan, konflik sosial, dan penurunan kepercayaan publik [4]. Oleh karena itu, diperlukan suatu metode yang efektif dalam mengklasifikasikan berita hoaks secara akurat agar dapat digunakan sebagai pencegahan terhadap penyebaran informasi yang menyesatkan di masyarakat.

Dalam konteks ini, teknik *text mining* menjadi salah satu pendekatan yang dapat digunakan untuk menganalisis dan mendeteksi berita hoaks. Penelitian terdahulu yang menggunakan teknik *text mining* dalam klasifikasi berita hoaks pernah dilakukan sebelumnya oleh Ferguson dan Istiono yang membahas mengenai deteksi berita hoaks pada berita politik di Indonesia dengan metode *Naïve Bayes Classifier* (NBC). Akurasi dari metode yang diusulkan mencapai 88,9%, dengan presisi sebesar 93,33%, *recall* sebesar 84%, dan skor F1 sebesar 88,4% [5]. Selanjutnya, penelitian terbaru oleh Murti, et al., (2025) yang membahas mengenai perancangan model cerdas untuk deteksi berita hoaks menggunakan metode *K-Nearest Neighbor* (KNN). Hasil evaluasi model menunjukkan bahwa KNN menghasilkan *Mean Absolute Error* (MAE) sebesar 0.011 dan *Root Mean Squared Error* (RMSE) sebesar 0.077, yang lebih baik dibandingkan dengan kinerja metode lain seperti SVM dan *Neural Network* [6].

Berdasarkan uraian penelitian terdahulu, maka penelitian ini memilih KNN dan NBC sebagai objek analisis komparatif untuk mengevaluasi efektivitas kedua metode tersebut serta mengeksplorasi potensi peningkatan kinerja dalam klasifikasi berita hoaks. Meskipun KNN menunjukkan performa yang baik dalam beberapa penelitian dengan nilai MAE dan RMSE yang rendah, metode ini memiliki kelemahan dalam hal waktu komputasi yang tinggi pada dataset besar [7]. Di sisi lain, NBC memiliki keunggulan dalam kecepatan klasifikasi, tetapi akurasi

dapat menurun ketika menghadapi data dengan korelasi antar kata yang kompleks [8]. Oleh karena itu, tantangan utama yang masih dihadapi adalah tingkat akurasi dan efisiensi dalam mendeteksi hoaks secara tepat sehingga diperlukan pendekatan yang dapat meningkatkan kinerja metode-metode ini dalam klasifikasi berita hoaks.

Sebagai solusi, penelitian ini mengusulkan pendekatan berbasis optimasi *Randomized Search* untuk meningkatkan akurasi klasifikasi berita hoaks dengan mengoptimalkan performa KNN dan NBC. *Randomized Search* bekerja dengan mengeksplorasi berbagai kombinasi hyperparameter secara acak dalam rentang yang telah ditentukan, untuk mencari parameter yang dapat meningkatkan kinerja model dalam melakukan klasifikasi dengan lebih optimal. Dengan melakukan analisis komparatif antara KNN dan NBC yang diperkuat dengan *Randomized Search*, penelitian ini bertujuan untuk menemukan metode terbaik dalam klasifikasi berita hoaks, sehingga dapat membantu mengurangi penyebaran informasi palsu secara lebih efektif.

1.2. Rumusan Masalah

Rumusan masalah pada penelitian ini yaitu:

1. Bagaimana cara mengimplementasikan algoritma *K-Nearest Neighbor* dan *Naive Bayes* dengan pendekatan berbasis optimasi *Randomized Search* dalam klasifikasi berita hoaks?
2. Bagaimana kinerja akurasi dari algoritma *K-Nearest Neighbor* dan *Naive Bayes* dengan pendekatan berbasis optimasi *Randomized Search* serta manakah yang lebih unggul dalam melakukan klasifikasi berita hoaks?

1.3. Tujuan dan Manfaat Penelitian

1.3.1. Tujuan Penelitian

Tujuan yang diharapkan dari penelitian ini yaitu:

1. Untuk mengimplementasikan algoritma *K-Nearest Neighbor* dan *Naive Bayes* dengan pendekatan berbasis optimasi *Randomized Search* dalam klasifikasi berita hoaks.
2. Untuk menganalisis kinerja akurasi dari algoritma *K-Nearest Neighbor* dan *Naive Bayes* dengan pendekatan berbasis optimasi *Randomized Search* serta manakah yang lebih unggul dalam melakukan klasifikasi berita hoaks.

1.3.2. Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini yaitu:

1. Dengan menerapkan optimasi *Randomized Search* pada algoritma *K-Nearest Neighbor* (KNN) dan *Naïve Bayes Classifier* (NBC), penelitian ini diharapkan dapat meningkatkan akurasi klasifikasi berita hoaks. Model yang lebih akurat dapat membantu dalam mengidentifikasi berita palsu dengan lebih efektif dan meminimalkan kesalahan dalam deteksi.
2. Penelitian ini dapat memberikan wawasan kepada akademisi dan praktisi tentang keunggulan serta kelemahan dari metode KNN dan NBC dalam klasifikasi berita hoaks. Dengan adanya analisis komparatif, hasil penelitian ini dapat menjadi referensi dalam memilih metode yang paling sesuai untuk tugas klasifikasi serupa di masa depan.
3. Dengan hasil penelitian ini, diharapkan dapat memberikan kontribusi bagi pengembang sistem deteksi berita hoaks, khususnya dalam mengembangkan model klasifikasi yang lebih andal. Hal ini dapat membantu masyarakat dalam memverifikasi informasi sebelum menyebarkannya di media sosial, sehingga mengurangi dampak negatif dari berita hoaks.

1.4. Batasan Masalah

Batasan masalah pada penelitian ini yaitu:

1. Dataset yang digunakan pada penelitian ini diambil dari Kaggle yaitu Indonesia *False News (Hoax) Dataset*.
2. Atribut *input* pada penelitian ini berupa judul berita dan narasi berita.
3. Atribut *output* pada penelitian ini berupa pelabelan hoaks dalam bentuk hoaks (1) dan fakta (0).

1.5. Keterbaruan Penelitian

Berikut ini akan diuraikan keterbaruan dari penelitian ini melalui *literature review* penelitian terdahulu yang berkaitan dengan topik penelitian antara lain:

1. Menurut Ferguson dan Istiono (2024) dalam penelitian yang berjudul “*Hoax News Detection in Indonesian Political Headlines Using Multinomial Naive Bayes*” didapatkan hasil evaluasi kinerja algoritma *Naive Bayes* dalam klasifikasi berita hoaks dengan *Confusion Matrix* memperoleh akurasi mencapai 88,9%, dengan presisi sebesar 93,33%, *recall* sebesar 84%, dan skor F1 sebesar 88,4% [5].

2. Menurut Kurnia, et al., (2024) dalam penelitian yang berjudul "Perbandingan *Naïve Bayes* dan CNN yang Dioptimasi PSO pada Identifikasi Berita *Hoax* Politik Indonesia" didapatkan hasil evaluasi model *Naïve Bayes* secara konsisten mampu mengungguli kinerja CNN, baik dengan atau tanpa penerapan PSO. Akurasi model *Naïve Bayes* mencapai 90,71%, sementara CNN mencapai 80,86% tanpa PSO, 79,68% dengan PSO, dan *Naïve Bayes* dengan PSO mencapai 90,25% [9].
3. Menurut Murti, et al., (2025) dalam penelitian yang berjudul "*Design of Intelligent Model for Text-Based Fake News Detection Using K-Nearest Neighbor Method*" didapatkan hasil evaluasi model menunjukkan bahwa KNN menghasilkan *Mean Absolute Error* (MAE) sebesar 0.011 dan *Root Mean Squared Error* (RMSE) sebesar 0.077, yang lebih baik dibandingkan dengan kinerja metode lain seperti SVM dan *Neural Network* [6].

Keterbaruan dari penelitian ini yaitu menggabungkan dua metode klasifikasi yang telah terbukti efektif, yaitu K-Nearest Neighbor (KNN) dan Naïve Bayes Classifier (NBC), dengan pendekatan optimasi Randomized Search untuk meningkatkan akurasi dalam deteksi berita hoaks. Berbeda dengan penelitian sebelumnya yang hanya membandingkan metode KNN dan NBC secara terpisah, penelitian ini melakukan analisis komparatif dengan menerapkan teknik boosting untuk mengoptimalkan performa kedua algoritma tersebut. Dengan adanya pendekatan optimasi Randomized Search, penelitian ini diharapkan dapat mengatasi kelemahan dari masing-masing algoritma, seperti tingginya waktu komputasi pada KNN dan sensitivitas NBC terhadap fitur dengan korelasi tinggi.