

SISTEM DETEKSI HOAKS PADA BERITA *ONLINE* MENGGUNAKAN METODE *NAIVE BAYES*

Muhammad Habibi Abdillah¹, Diah Rahmawati², Nia Komalasari³

Program Studi Teknik Informatika Fakultas Teknik Universitas Islam Syekh-Yusuf Tangerang^{1,2,3}

Email: 2104030055@students.unis.ac.id

Informasi	Abstract
Volume : 2	<p><i>The massive spread of hoax news online, particularly on the topic of natural disasters, has become a significant problem causing public unrest and anxiety. This research aims to design and build a system capable of automatically detecting hoax news to help improve digital literacy. The methodology involved collecting news data from online portals using web scraping techniques, followed by data preprocessing which included case folding, tokenization, stopwords removal, and stemming. The Term Frequency-Inverse Document Frequency (TF-IDF) method was used for feature representation. A classification model was built using the Multinomial Naïve Bayes algorithm and subsequently implemented as a functional web application using the Django framework. The model's performance evaluation demonstrated excellent results, achieving an Accuracy of 97.26%, with balanced Precision, Recall, and F1-Score values of 0.97 for both hoax and valid classes. This study concludes that the Naïve Bayes-based hoax detection system, implemented as a web application, is an effective and viable solution for assisting users in the initial verification of online news.</i></p> <p>Keyword: Django, hoax detection, Naïve Bayes, online news, text classification, TF-IDF.</p>
Nomor : 8	
Bulan : Agustus	
Tahun : 2025	
E-ISSN : 3062-9624	

Abstrak

Penyebaran berita hoaks di media online, khususnya pada topik bencana alam, telah menjadi masalah signifikan yang dapat menimbulkan keresahan di masyarakat. Penelitian ini bertujuan untuk merancang dan membangun sebuah sistem yang dapat mendeteksi berita hoaks secara otomatis guna membantu meningkatkan literasi digital. Metodologi yang digunakan dimulai dari pengumpulan data berita dari portal online melalui teknik web scraping, dilanjutkan dengan data preprocessing yang terdiri dari case folding, tokenization, stopwords removal, dan stemming. Untuk representasi fitur, digunakan metode pembobotan kata Term Frequency-Inverse Document Frequency (TF-IDF). Model klasifikasi dibangun menggunakan algoritma Naïve Bayes varian Multinomial dan diimplementasikan ke dalam aplikasi web fungsional menggunakan framework Django. Hasil pengujian kinerja model menunjukkan performa yang sangat baik, dengan tingkat Akurasi mencapai 97.26%, serta nilai Precision, Recall, dan F1-Score yang seimbang, yaitu 0.97 untuk kedua kelas (hoaks dan valid). Penelitian ini menyimpulkan bahwa sistem deteksi hoaks menggunakan metode Naïve Bayes yang diimplementasikan pada aplikasi web merupakan solusi yang efektif dan layak untuk membantu pengguna melakukan verifikasi awal terhadap berita online.

Kata Kunci: Berita online, deteksi hoaks, django, klasifikasi teks, Naïve Bayes, TF-IDF.

A. PENDAHULUAN

Era digital yang berkembang pesat membawa dampak besar di berbagai aspek kehidupan, termasuk kesehatan, komunikasi, dan pendidikan. Perkembangan teknologi informasi dan komunikasi memungkinkan masyarakat mengakses informasi dengan cepat melalui berbagai platform digital. Namun, kemajuan ini juga memunculkan tantangan serius, seperti maraknya penyebaran berita hoaks yang berpotensi merusak kualitas diskusi publik dan memicu polarisasi sosial. Fenomena ini terjadi pula di Indonesia, di mana berita hoaks tentang politik, kesehatan, hingga bencana alam menyebar dengan cepat di media sosial, sering kali menyebabkan kepanikan masyarakat. Salah satu contoh nyata adalah hoaks gempa bumi besar yang dikabarkan akan melanda Jakarta pada akhir 2019, yang meskipun dimaksudkan sebagai edukasi oleh media, tetap menimbulkan keresahan luas.

Kurangnya kesadaran masyarakat untuk memverifikasi kebenaran informasi menjadi faktor utama penyebaran hoaks. Banyak orang kesulitan membedakan berita valid dan berita palsu, sehingga disinformasi mudah menyebar dan memicu keresahan sosial. Dalam konteks bencana alam, hoaks dapat memperburuk situasi karena informasi keliru mengenai lokasi, jumlah korban, atau penanganan bencana dapat menghambat respon yang tepat. Oleh sebab itu, dibutuhkan solusi yang mampu mengidentifikasi dan memblokir berita palsu sebelum menyebar luas.

Machine learning, khususnya algoritma Naïve Bayes, menjadi salah satu solusi efektif untuk mendeteksi berita hoaks. Naïve Bayes merupakan metode klasifikasi yang dapat mengenali pola pada data teks dan memproses informasi dalam jumlah besar dengan cepat. Penelitian sebelumnya menunjukkan bahwa metode ini memiliki tingkat akurasi tinggi dalam klasifikasi teks, termasuk deteksi berita palsu. Dengan penerapan Naïve Bayes, diharapkan sistem deteksi hoaks dapat memberikan peringatan dini kepada pengguna serta meningkatkan kesadaran untuk memverifikasi informasi sebelum membagikannya.

Penelitian ini membatasi fokus pada berita kategori bencana alam, baik yang valid maupun yang mengandung hoaks, dari situs Detik, Kompas, CNN Indonesia, dan Turnback Hoax MAFINDO. Data dikumpulkan menggunakan teknik web scraping dari sumber resmi dan dataset publik. Pembatasan ini bertujuan agar penelitian lebih terfokus serta mampu memberikan kontribusi nyata dalam meningkatkan literasi digital masyarakat khususnya terkait berita bencana alam yang sensitif dan rawan menimbulkan kepanikan.

Dalam implementasinya, penelitian ini memanfaatkan Django sebagai framework berbasis Python untuk membangun aplikasi web pendeteksi hoaks. Django dipilih karena

kemampuannya menangani data besar, memproses teks dengan cepat, serta menyediakan struktur pengembangan yang aman dan efisien. Integrasi antara Django dan model Naïve Bayes memungkinkan pembuatan antarmuka yang mudah diakses oleh masyarakat sehingga mereka dapat mengecek validitas berita secara langsung melalui browser.

Berdasarkan perumusan masalah, penelitian ini bertujuan untuk: (1) merancang sistem deteksi berita hoaks berbasis web menggunakan Django, (2) mengidentifikasi langkah-langkah klasifikasi data untuk mengecek akurasi deteksi hoaks, dan (3) mengukur tingkat akurasi metode Naïve Bayes dalam mendeteksi berita palsu. Manfaat yang diharapkan adalah terciptanya sistem yang membantu masyarakat dan pihak berwenang dalam mengidentifikasi hoaks, mengurangi penyebaran informasi palsu, serta meningkatkan kesadaran publik akan pentingnya verifikasi informasi.

B. METODE PENELITIAN

Penelitian ini menggunakan teknik web scraping untuk mengumpulkan data dari berbagai situs berita, seperti Detik, Kompas, dan CNN Indonesia, serta situs pemeriksa fakta seperti TurnBackHoax dan CekFakta. Data yang diambil meliputi judul, ringkasan isi berita, dan tanggal publikasi, dengan fokus pada berita yang berkaitan dengan bencana alam yang sering dikaitkan dengan hoaks atau misinformasi. Setelah pengumpulan data, tahap selanjutnya adalah proses pembersihan data (data preprocessing) yang mencakup case folding, tokenisasi, penghapusan stopwords, serta stemming. Pembobotan kata dilakukan menggunakan metode TF-IDF untuk mengukur relevansi kata-kata dalam setiap dokumen. Model klasifikasi dibangun menggunakan algoritma Naïve Bayes, yang digunakan untuk mengklasifikasikan berita menjadi hoaks atau valid berdasarkan data yang telah diproses.

Alur penelitian dilakukan secara sistematis dengan pengumpulan data, pembersihan data, pembobotan kata, serta klasifikasi menggunakan Naïve Bayes. Setelah model dibangun dan dievaluasi, sistem ini diimplementasikan ke dalam aplikasi web berbasis Django untuk mendeteksi berita hoaks secara otomatis. Pengujian dilakukan dengan metode Blackbox untuk menguji fungsionalitas sistem tanpa mengetahui struktur internalnya, memastikan bahwa input berita dapat diproses dengan tepat. Penelitian ini juga mencakup pengujian kinerja model menggunakan Confusion Matrix untuk menghitung akurasi, presisi, dan recall, serta pembagian data pelatihan dan pengujian dengan rasio 80:20. Jadwal penelitian mencakup tahap-tahap mulai dari pencarian dan pengajuan judul hingga pelaksanaan sidang skripsi pada Agustus 2025.

C. HASIL DAN PEMBAHASAN

Hasil Penelitian

Pada bagian ini akan menjelaskan hasil penelitian yang berupa implementasi dari rancangan yang telah dibuat pada bab sebelumnya. Implementasi pada penelitian ini dimulai dari pengumpulan data, *data preprocessing*, pembobotan kata menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF), pembuatan model menggunakan *Naïve Bayes*, dan pengujian model yang telah dibangun menggunakan *Confusion Matrix*.

Pengumpulan Data

Pengumpulan data dengan mengambil *dataset* yang didapatkan melalui *scraping* dan situs resmi kaggle. *Dataset* tersebut berisi kumpulan data berita yang masing-masing data memiliki sebuah label hoaks dan valid. Selanjutnya *dataset* tersebut akan dimuat kedalam program untuk dilakukan proses selanjutnya. Proses pengumpulan data merupakan tahap fundamental dalam penelitian ini. Data berita diambil menggunakan skrip *web scraping* yang dirancang khusus untuk penelitian ini dan dijalankan pada lingkungan *Google Colab*. Logika utama dari skrip tersebut, yang diimplementasikan untuk portal berita kompas, detik, cnn dan *turnbackhoaks* dijelaskan secara rinci pada tabel di bawah ini. Prinsip yang sama juga diterapkan dengan penyesuaian tag HTML untuk sumber berita lainnya.

Data Preprocessing

Data teks yang diperoleh dari hasil *web scraping* masih bersifat mentah dan tidak terstruktur. Data ini mengandung banyak "noise" seperti huruf kapital, tanda baca, angka, dan kata-kata umum yang tidak memiliki makna signifikan (*stopwords*) yang dapat mengganggu proses klasifikasi. Oleh karena itu, diperlukan tahapan *data preprocessing* untuk membersihkan dan mengubah data teks menjadi format yang lebih terstruktur dan siap untuk dianalisis oleh model *Naïve Bayes*.

Proses *preprocessing* dalam penelitian ini terdiri dari lima tahapan utama yang dijalankan secara berurutan: *Case Folding*, *Punctuation & Symbol Removal*, *Tokenization*, *Stopwords Removal*, dan *Stemming*.

Pembobotan Kata

Setelah data teks dibersihkan melalui tahapan *preprocessing*, setiap berita kini menjadi kumpulan token atau kata dasar. Namun, model *machine learning* tidak dapat memproses teks secara langsung; data tersebut harus diubah ke dalam bentuk representasi numerik. Dalam penelitian ini, metode yang digunakan untuk mengubah teks menjadi vektor numerik adalah TF-IDF (*Term Frequency-Inverse Document Frequency*). Berikut adalah cuplikan kode utama

yang digunakan untuk menginisialisasi dan menerapkan *TfidfVectorizer* pada data teks yang telah melalui tahap *preprocessing*.

Pemodelan Sistem dengan *Naïve Bayes*

Setelah tahap pembobotan kata, *dataset* telah siap dalam format numerik (matriks TF-IDF) yang dapat diproses oleh algoritma *machine learning*. Tahap selanjutnya adalah membangun model klasifikasi untuk memprediksi apakah sebuah berita termasuk dalam kategori hoaks atau valid. Sesuai dengan rancangan penelitian, algoritma yang digunakan adalah *Naïve Bayes*.

Dalam penelitian ini, varian *Naïve Bayes* yang diimplementasikan adalah *Multinomial Naïve Bayes*. Varian ini secara khusus sangat efektif untuk tugas klasifikasi teks di mana fitur-fitur data merepresentasikan frekuensi kemunculan kata (seperti pada TF-IDF).

Ekspor Model untuk Implementasi Sistem

Setelah model *Naïve Bayes* berhasil dilatih dan divalidasi, langkah terakhir sebelum implementasi pada sistem web adalah mengeksport model tersebut. Proses ini bertujuan untuk menyimpan "kecerdasan" yang telah dipelajari oleh model ke dalam sebuah file, sehingga aplikasi web dapat menggunakannya untuk melakukan prediksi secara *real-time* tanpa perlu mengulang proses pelatihan yang memakan waktu.

Dalam penelitian ini, terdapat dua komponen penting yang diekspor: (1) Objek *TF-IDF Vectorizer*: Komponen ini berisi seluruh kosakata (vocabulary) dan bobot IDF yang telah dipelajari dari data latih. Ini sangat penting agar teks baru yang dimasukkan oleh pengguna dapat diubah menjadi vektor numerik dengan cara yang sama persis. (2) Objek Model *Naïve Bayes*: Komponen ini adalah model yang telah dilatih dan berisi semua probabilitas yang dibutuhkan untuk melakukan klasifikasi. Kedua komponen ini disimpan (di-serialisasi) ke dalam file biner menggunakan pustaka *pickle* dari Python.

Implementasi Sistem Web dengan Django

Setelah model *machine learning* berhasil dibangun dan diekspor, tahap selanjutnya adalah mengimplementasikannya ke dalam sebuah aplikasi web yang dapat diakses oleh pengguna. Sesuai dengan rancangan penelitian, aplikasi ini dibangun menggunakan *framework Django* dengan bahasa pemrograman Python.

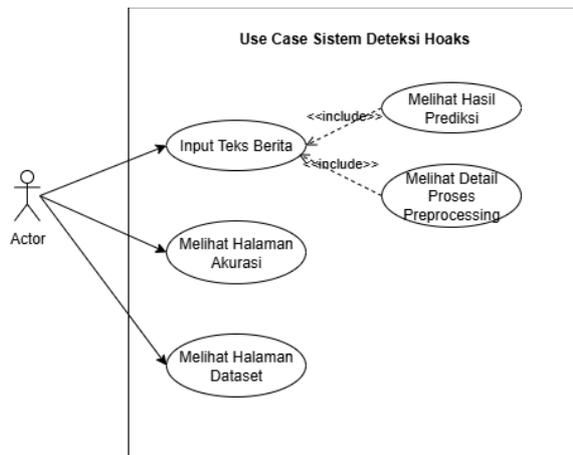
Perancangan Sistem

Sebelum masuk ke detail implementasi teknis, bagian ini menyajikan perancangan fungsional dan alur kerja sistem yang menjadi acuan dalam pengembangan aplikasi.

Perancangan ini divisualisasikan menggunakan *Use Case Diagram* untuk menggambarkan fungsionalitas dan *Activity Diagram* untuk menggambarkan alur kerja.

Use Case Diagram

Use Case Diagram digunakan untuk menggambarkan interaksi fungsional antara pengguna (*actor*) dengan fitur-fitur utama yang disediakan oleh sistem deteksi hoaks. Diagram ini menunjukkan lingkup kegunaan dari sistem yang dibangun.

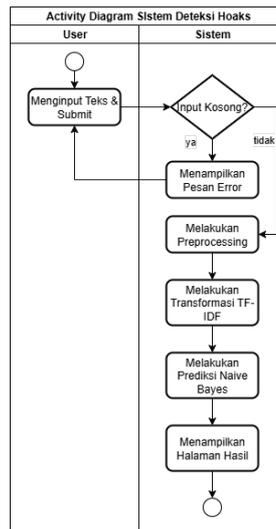


Gambar 7. Use Case Diagram Sistem Deteksi Hoaks

Berdasarkan Gambar diatas, dapat dilihat bahwa terdapat satu Aktor yaitu Pengguna yang dapat berinteraksi dengan sistem. Pengguna dapat melakukan tiga fungsi utama: menginput teks berita untuk klasifikasi, melihat halaman *dataset*, dan melihat halaman akurasi. Use case "Input Teks Berita" secara otomatis akan menyertakan (<<include>>) dua *use case* lain, yaitu "Melihat Hasil Prediksi" dan "Melihat Detail Proses Preprocessing" yang hasilnya akan ditampilkan setelah pengguna mengirimkan teks.

Activity Diagram

Activity Diagram digunakan untuk memodelkan alur kerja (*workflow*) sistem secara lebih rinci. Diagram ini menggambarkan langkah-langkah yang terjadi secara berurutan, mulai dari saat pengguna memasukkan teks hingga sistem menampilkan hasil akhir dari proses klasifikasi.



Gambar 8. Activity Diagram Sistem Deteksi Hoaks

Activity Diagram pada Gambar diatas memodelkan alur kerja sistem dari awal hingga akhir. Alur dimulai saat pengguna menginput teks, kemudian sistem melakukan validasi untuk memastikan input tidak kosong. Jika valid, sistem menjalankan serangkaian proses (*preprocessing*, *transformasi TF-IDF*, *prediksi Naive Bayes*) hingga hasil akhir ditampilkan kepada pengguna.

Arsitektur dan Alur Kerja Aplikasi

Aplikasi ini dikembangkan mengikuti pola arsitektur *MVT (Model-View-Template)* yang menjadi standar pada Django. Alur kerja sistem dari sudut pandang pengguna hingga pemrosesan di *backend* adalah sebagai berikut: (1) *Permintaan Halaman oleh Pengguna*: Pengguna mengakses halaman utama aplikasi. *urls.py* pada Django akan mengarahkan permintaan ini ke fungsi view yang sesuai. (2) *Tampilan Halaman Input*: Fungsi *view* kemudian me-render sebuah *template HTML (index.html)* yang berisi antarmuka utama, termasuk formulir dengan area teks untuk input berita.

Proses Prediksi di Backend

Proses prediksi di backend dimulai ketika pengguna memasukkan teks dan menekan tombol "Submit", lalu data dikirim ke file *views.py*. Di dalamnya, sistem memuat file *tfidf_vectorizer.pkl* dan *naive_bayes_model.pkl* yang telah dilatih sebelumnya. Teks masukan kemudian diproses melalui tahapan *preprocessing* yang sama seperti pada data latih, termasuk *case folding*, *tokenisasi*, dan langkah lainnya, sebelum diubah menjadi vektor numerik menggunakan *vectorizer* yang dimuat. Vektor tersebut kemudian diproses oleh model untuk menghasilkan prediksi kelas, apakah berita termasuk hoaks atau valid, sekaligus memberikan tingkat keyakinan berupa nilai probabilitas.

Menampilkan Hasil: Hasil prediksi dan tingkat keyakinan dikirim kembali ke sebuah *template HTML* untuk ditampilkan secara jelas kepada pengguna.

Tampilan Antarmuka Sistem Hasil Implementasi

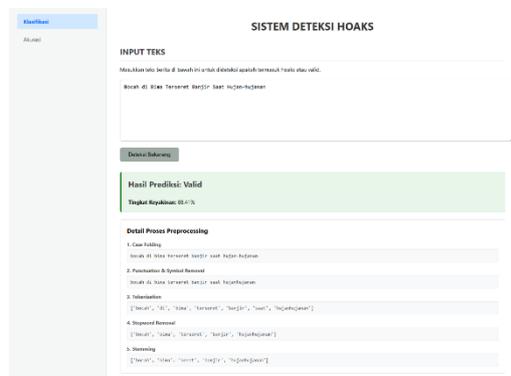
Berikut adalah tampilan antarmuka dari aplikasi web Sistem Deteksi Hoaks yang telah berhasil dibangun.

Halaman Klasifikasi



Gambar 9. Halaman utama klasifikasi teks

Gambar diatas menunjukkan halaman utama sistem, yang merupakan implementasi dari perancangan antarmuka pada Gambar diatas. Pada halaman ini, pengguna dapat memasukkan teks berita yang ingin diperiksa.



Gambar 10. Halaman Hasil Klasifikasi

Setelah pengguna memasukkan teks dan menekan tombol submit, sistem akan menampilkan halaman hasil seperti pada Gambar diatas. Halaman ini menampilkan label prediksi beserta tingkat keyakinan model terhadap prediksi tersebut, memberikan umpan balik langsung kepada pengguna.

Halaman Dataset

Klasifikasi

Dataset

Data Latih

Data Uji

Akurasi

SISTEM DETEKSI HOAKS

Dataset Latih

	Judul	Paragraf Utama	Label
0	Akses Terdampak Bencana, Kunjungan Wisata Pantai Selatan Turun 80%	Kunjungan wisatawan ke pantai selatan Cianjur turun drastis. Pasalnya wisatawan enggan untuk berlibur ke destinasi pantai selatan lantaran dampak bencana alam longsor dan pergerakan tanah di jalur menuju selatan Cianjur beberapa waktu lalu.	0
1	Tanggul Kembali Jebol, Jalur Pantura Demak-Kudus Lumpuh	Tanggul Sungai Wulan Di Demak, Jawa Tengah, yang baru dua pekan selesai diperbaiki, kembali jebol pada Minggu dini hari. Air yang tumpah merendam ribuan rumah serta memutus jalan nasional Demak menuju Kudus.	0
2	Pakar Ingatkan Peningkatan Suhu Global Bisa Perburuk Intensitas Bencana Alam di Indonesia	KOMPAS.com - Guru Besar Departemen Kesehatan Lingkungan Fakultas Kesehatan Masyarakat (FKM) Universitas Indonesia Prof Bambang Wispriono mengatakan, Indonesia berada di Cincin Api Pasifik (Ring of Fire).	0
3	Banjir Besar Landa Jepang, Satu Juta Warga Diminta Mengungsi	Lebih dari satu juta orang harus mengungsi ketika hujan deras memicu banjir dan longsor di Jepang bagian barat pada Sabtu (14/8).	0
4	Update Longsor Sumedang: 13 Orang Meninggal Dunia	Upaya pencarian korban longsor di Desa Cihanjuang, Cimanggung, Sumedang terus dilakukan tim SAR gabungan. Hingga minggu kemarin total ada 13 korban meninggal.	0
5	10 Link Twibbon Hari Konservasi Alam Nasional 2024, Cek di Sini	Tanggal 10 Agustus adalah Hari Konservasi Alam Nasional (HKAN). Peringatan ini ditetapkan oleh Presiden RI keenam, Susilo Bambang Yudhoyono melalui Keputusan Presiden Republik Indonesia Nomor 22 Tahun 2009.	0
6	Deret Bencana Iklim Dunia dengan Kerugian Terbesar Sepanjang 2023	Bencana alam imbasa krisis iklim melanda sejumlah negara di dunia sepanjang 2023. Beberapa di antaranya memicu kerugianA masif. Simak daftarnya berikut.	0

Gambar 11. Implementasi halaman data latih

Klasifikasi

Dataset

Data Latih

Data Uji

Akurasi

SISTEM DETEKSI HOAKS

Dataset Uji

	Judul	Paragraf Utama
2682	https://turbachoaax.id/2024/05/08/-gelombang-tinggi-20-meter-sapu-habis-laut-indonesia-dan-ribuan-warga/	Hasil periksa fakta Amanda Rahma Faktanya, tidak benar gelombang tinggi 20 meter menyapu habis laut Indonesia dan ribuan warga. N menyampaikan
2683	[SALAH] Foto Banjir di Kalimantan Selatan	Hasil Periksa Fakta Natalia Kristian (Anggota Komisiarlat MAFINDO Universitas Indonesia). Informasi yang salah. Faktanya, foto tersebut at 2012. = = = = KATEGORI: Konten yang Salah/False Context = = = = SUMBER: Twitterhttps://archive.vn/79iK2 = = = = NARASI: A& Semoga pemerintah terutama pak@jokowibisa memperhatikan dan ikut membantu warga di Kalsel, Indonesia bukan hanya pulau Jawa, l cepat surut, dan semua selamat, aamiin#KalselUgalIndonesia#PrayforKalselA&A = = = = PENJELASAN: Beredar sebuah postingan c beberapa foto bencana banjir yang diklaim adalah foto banjir di Kalimantan Selatan. Postingan ini disukai sebanyak 1,5 ribu kali, dikomer kembali 534 kali. Setelah melakukan penelusuran, foto yang digunakan ternyata adalah foto banjir yang terjadi di Krai Krasnodar, Rusia p menghancurkan hampir 13.000 rumah dan 171 orang meninggal, banjir tersebut terjadi setelah badai yang menghantam Krasnodar serta banjir yang terjadi di Kalimantan Selatan terdata per 14 Januari, sebanyak 67.842 orang dan 19.452 unit rumah terdampak banjir tersebut banjir di Kalimantan Selatan adalah tidak benar sehingga termasuk dalam kategori Konten yang Salah/False Context. = = = = REFEREN https://id.wikipedia.org/wiki/Banjir_Krai_Krasnodar_2012 https://www.rt.com/news/toll-dead-region-floods-633/ https://news.detik.com/i tevas-akibat-banjir-di-rusia https://kaltim.suara.com/read/2021/01/15/113301/banjir-awal-tahun-di-kalsel-jadi-yang-terburuk-dalam-10 KristianEditor: Bentang Febrilyan
		Hasil Periksa Fakta Indri Pramesti Widyaningrum (Anggota Komisiarlat MAFINDO Universitas Gunadarma) Bukan pertemuan untuk memin melawan wabah virus Corona, tetapi pertemuan antar-agama dan multikultural untuk mengenang para korban COVID-19 yang diadakan Napoli, Italia. = = = = Kategori: Manipulated Content/Konten Yang Dimanipulasi = = = = Sumber: Twitter (https://archive.st/vkdd) Laman

Gambar 12. Implementasi halaman data uji

Halaman Akurasi

Klasifikasi

Akurasi

SISTEM DETEKSI HOAKS

Akurasi Model

Akurasi Keseluruhan

Akurasi mengukur seberapa sering model membuat prediksi yang benar secara keseluruhan. Nilai ini didapat dari total prediksi benar dibagi dengan total seluruh data uji.

97.26%

Confusion Matrix

Confusion Matrix memberikan rincian visual dari performa prediksi model. Matriks ini membandingkan antara label data yang sebenarnya (Aktual) dengan hasil tebakan model (Prediksi).

		Prediksi Model	
		Valid	Hoaks
Aktual	Valid	266	8
	Hoaks	7	266

Dari matriks di atas, dapat dilihat bahwa terdapat 266 berita valid yang berhasil diprediksi dengan benar (True Negative) dan 266 berita hoaks yang juga berhasil diidentifikasi (True Positive). Namun, terdapat kesalahan prediksi, yaitu 8 berita valid yang salah dianggap hoaks (False Positive) dan 7 berita hoaks yang terlewat atau dianggap valid (False Negative).

Laporan Rinci

Laporan ini meneliti performa untuk setiap kelas. Precision mengukur tingkat presisi dari prediksi yang dibuat, misalnya, dari semua berita yang dilabel "hoaks" oleh model, berapa persen yang benar-benar hoaks. Recall mengukur kemampuan model untuk menemukan semua kasus yang relevan, misalnya, dari seluruh berita hoaks yang ada, berapa persen yang berhasil terdeteksi. F1-Score adalah nilai tunggal yang menyeimbangkan antara Precision dan Recall.

```

precision recall f1-score support
valid (0) 0.97 0.97 0.97 274
hoaks (1) 0.97 0.97 0.97 273
accuracy avg 0.97 0.97 0.97 547
weighted avg 0.97 0.97 0.97 547

precision recall f1-score support
valid (0) 0.97 0.97 0.97 274
hoaks (1) 0.97 0.97 0.97 273
accuracy avg 0.97 0.97 0.97 547
weighted avg 0.97 0.97 0.97 547

```

Dataset Penelitian

Model ini dilatih dan diuji menggunakan dataset berita yang telah dikumpulkan. Bagi peneliti yang ingin memili lebih lanjut atau memverifikasi data, dataset yang digunakan dapat diakses melalui badan berikut:

[Lihat Data Latih](#) | [Lihat Data Uji](#)

Gambar 13. Implementasi halaman akurasi

Untuk memberikan gambaran visual mengenai kinerja model, disediakan halaman Akurasi yang menampilkan hasil evaluasi seperti *Confusion Matrix*.

Hasil Pengujian

Pada tahap ini, disajikan hasil pengujian dari dua aspek utama: kinerja model klasifikasi dan fungsionalitas sistem web. Untuk kinerja model, disajikan hasil dari skenario input Gabungan (Judul + Paragraf) yang merupakan skenario dengan performa terbaik dan diimplementasikan pada sistem.

Hasil Pengujian Kinerja Model

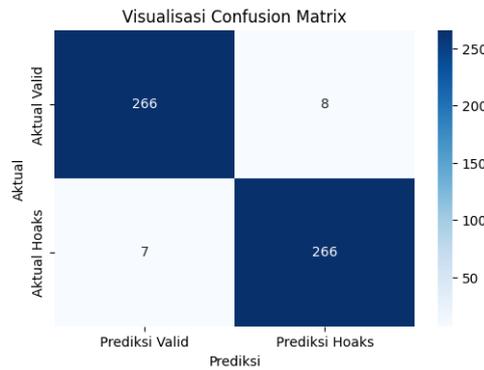
Pengujian kinerja model dilakukan untuk mengevaluasi secara kuantitatif seberapa baik algoritma Naïve Bayes dalam mengklasifikasikan berita hoaks dan valid menggunakan skenario input gabungan. Hasil pengujian pada data uji disajikan pada tabel di bawah ini.

Tabel 3. Hasil Kinerja Model

Skenario Input	Akurasi (%)	Presisi (rata-rata)	Recall (rata-rata)
Judul Saja	93.78%	0.94	0.94
Paragraf Utama Saja	97.07%	0.97	0.97
Gabungan	97.26%	0.97	0.97

Berdasarkan hasil rekapitulasi pada Tabel diatas, dapat dilihat perbandingan kinerja model pada ketiga skenario pengujian. Skenario input Gabungan (Judul + Paragraf) menunjukkan performa tertinggi dengan tingkat Akurasi 97.26%. Terdapat peningkatan akurasi yang signifikan sekitar 3.5% dari skenario Judul Saja (93.78%) ke skenario Paragraf Utama Saja (97.07%). Hal ini mengindikasikan bahwa paragraf utama berita mengandung fitur-fitur (kata-kata) yang jauh lebih informatif bagi model untuk membuat prediksi dibandingkan hanya judulnya saja.

Meskipun skenario "Gabungan" adalah yang terbaik, peningkatannya tidak terlalu signifikan dibandingkan dengan skenario "Paragraf Utama Saja". Ini menunjukkan bahwa paragraf utama sudah hampir mencukupi bagi model untuk mencapai performa puncaknya. Distribusi hasil prediksi Gabungan yang lebih rinci untuk skenario ini dapat dilihat pada *Confusion Matrix* berikut:



Gambar 14. Confusion Matrix

Dari matriks di atas, dapat dilihat bahwa terdapat 266 berita valid yang berhasil diprediksi dengan benar (*True Negative*) dan 266 berita hoaks yang juga berhasil diidentifikasi (*True Positive*). Namun, terdapat kesalahan prediksi, yaitu 8 berita valid yang salah dianggap hoaks (*False Positive*) dan 7 berita hoaks yang terlewat atau dianggap valid (*False Negative*).

Hasil Pengujian Blackbox

Pengujian sistem secara *blackbox* dilakukan untuk mengevaluasi respons dan tingkat keyakinan sistem terhadap berbagai jenis input. Pengujian ini menggunakan contoh berita hoaks dan valid untuk melihat bagaimana sistem bereaksi pada skenario input yang berbeda-beda: hanya judul, hanya paragraf utama, dan gabungan keduanya. Tujuannya adalah untuk mengetahui pada kondisi input mana sistem memberikan tingkat keyakinan prediksi yang paling tinggi. Hasil dari pengujian studi kasus ini disajikan pada tabel di bawah ini.

Tabel 4. Pengujian Blackbox

Jenis Berita	Skenario Input	Contoh Teks (Diringkas)	Output Sistem (Prediksi & Tingkat Keyakinan)
Valid	Judul Saja	Angin Puting Beliung Terjang Jakut, Sejumlah Bangunan dan Kendaraan Rusak	(Valid) 90.06%
Valid	Paragraf Utama Saja	JAKARTA, Kompas.com - Angin puting beliung menerjang kawasan Bendungan Melayu, Kelurahan Tugu Selatan, Koja, Jakarta Utara, Sabtu (22/3/2025) sekitar pukul 03.14 WIB.	(Valid) 95.20%
		Angin Puting Beliung Terjang Jakut, Sejumlah Bangunan dan Kendaraan	(Valid)

Valid	Gabungan	Rusak JAKARTA, Kompas.com - Angin puting beliung menerjang kawasan...	95.97%
Hoaks	Judul Saja	Gelombang tinggi 20 meter sapu habis laut indonesia dan ribuan warga	(Hoaks) 76.61%
Hoaks	Paragraf Utama Saja	Hasil periksa fakta Amanda Rahma Faktanya, tidak benar gelombang tinggi 20 meter menyapu habis laut Indonesia dan ribuan warga. Narator dalam video ini hanya menyampaikan	(Hoaks) 98.44%
Hoaks	Gabungan	Gelombang tinggi 20 meter sapu habis laut indonesia dan ribuan warga Hasil periksa fakta Amanda Rahma Faktanya, tidak benar gelombang tinggi 20 meter menyapu habis laut Indonesia dan ribuan warga. Narator dalam video ini hanya menyampaikan	(Hoaks) 95.53%

Berdasarkan hasil pengujian pada Tabel diatas, dapat ditarik beberapa analisis kunci mengenai perilaku sistem. Pada kasus berita valid ("Angin Puting Beliung"), sistem secara konsisten berhasil mengklasifikasikan berita sebagai Valid. Tingkat keyakinan model terus meningkat seiring dengan kelengkapan informasi, dari 90.06% pada input judul saja, hingga mencapai 95.97% pada input gabungan. Hal ini menunjukkan bahwa untuk berita faktual, penambahan detail dari paragraf utama memperkuat keyakinan model terhadap prediksi yang benar.

Sementara itu, pada kasus berita hoaks ("Gelombang Tinggi 20 Meter"), saat hanya diberikan judul yang bombastis, sistem berhasil memprediksi sebagai Hoaks namun dengan tingkat keyakinan yang relatif rendah (76.61%). Keyakinan ini melonjak drastis menjadi 98.44% ketika hanya diberikan paragraf utama yang berisi frasa kunci seperti "Hasil periksa fakta", yang membuktikan bahwa model sangat efektif dalam mengidentifikasi kata-kata sanggahan. Menariknya, pada input gabungan, tingkat keyakinan (95.53%) sedikit lebih rendah dibandingkan hanya dari paragraf, yang mengindikasikan bahwa judul sensasional masih sedikit memengaruhi perhitungan probabilitas, walaupun prediksi akhir tetap benar.

Secara keseluruhan, pengujian ini menunjukkan bahwa sistem paling optimal dan memiliki keyakinan tertinggi ketika diberikan konteks yang lengkap (gabungan judul dan paragraf).

Pembahasan

Pada sub-bab ini dibahas secara mendalam hasil pengujian yang telah dilakukan, mencakup analisis kinerja model, analisis kesalahan, perbandingan dengan penelitian terdahulu, serta implikasi dan keterbatasan penelitian. Berdasarkan hasil pengujian kuantitatif, model mencapai akurasi tertinggi sebesar 97,26% dengan nilai presisi dan recall yang seimbang, masing-masing sebesar 0,97 pada kedua kelas. Temuan ini diperkuat oleh hasil pada Tabel 4.2 yang menunjukkan bahwa tingkat keyakinan model cenderung meningkat seiring dengan kelengkapan informasi yang diberikan. Sebagai contoh, pada pengujian berita valid berjudul “Angin Puting Beliung”, tingkat keyakinan meningkat dari 90,06% (judul saja) menjadi 95,97% ketika menggunakan skenario gabungan. Hal ini membuktikan bahwa semakin lengkap konteks yang tersedia, semakin akurat dan yakin model Naïve Bayes dalam melakukan klasifikasi.

Meskipun kinerjanya tergolong tinggi, analisis kesalahan pada berita hoaks di Tabel 4.2 mengungkap fenomena menarik. Pada kasus judul bombastis “Gelombang tinggi 20 meter sapu habis”, sistem mampu memprediksi sebagai hoaks namun hanya dengan keyakinan 76,61%. Keyakinan meningkat drastis menjadi 98,44% saat hanya paragraf utama yang diproses, karena mengandung frasa kunci seperti “Hasil periksa fakta” dan “tidak benar”. Menariknya, tingkat keyakinan sedikit menurun menjadi 95,53% ketika judul dan paragraf digabungkan. Hal ini menunjukkan bahwa meskipun indikator fact-check kuat, keberadaan kata-kata pemicu hoaks pada judul tetap memberi pengaruh negatif pada keyakinan model, meski hasil klasifikasi akhirnya tetap benar. Temuan ini menyoroti sensitivitas model terhadap komposisi kata pada input.

Jika dibandingkan dengan penelitian sebelumnya oleh Nova Agustina (2021) yang juga menggunakan Naïve Bayes untuk deteksi berita palsu di media sosial, kinerja model dalam penelitian ini tergolong lebih unggul. Penelitian tersebut memperoleh akurasi 81% dan F1-Score 87%, sedangkan penelitian ini berhasil mencapai akurasi 97,26% dengan keseimbangan presisi dan recall yang lebih baik. Perbedaan ini kemungkinan dipengaruhi oleh variasi dataset—penelitian ini menggunakan sumber berita yang lebih beragam dan kredibel, serta difokuskan pada topik bencana alam, sehingga pola kata kunci antara berita hoaks dan valid lebih jelas dan mudah dikenali.

Secara praktis, hasil penelitian ini menunjukkan bahwa sistem deteksi hoaks yang dikembangkan berpotensi menjadi alat bantu efektif bagi masyarakat dalam memfilter informasi, khususnya pada topik bencana alam. Namun, terdapat beberapa keterbatasan yang perlu diakui, antara lain: (1) ruang lingkup data yang terbatas hanya pada topik bencana alam sehingga kinerja pada topik lain belum teruji, (2) sistem hanya mampu memproses teks dan belum dapat memverifikasi informasi berbasis gambar, infografis, atau video, dan (3) kesulitan mendeteksi hoaks dengan bahasa sarkastik atau satir yang minim kata kunci eksplisit. Keterbatasan ini menjadi peluang bagi penelitian selanjutnya untuk mengembangkan model yang lebih adaptif dan komprehensif.

D. KESIMPULAN

Berdasarkan hasil pengujian, dapat disimpulkan bahwa hasil penelitian dan implementasi sistem yang telah dilakukan, maka dapat ditarik beberapa kesimpulan sebagai berikut: (1) Penelitian ini telah berhasil merancang dan membangun sebuah sistem deteksi hoaks dalam aplikasi berbasis web menggunakan framework Django. Perancangan sistem ini menerapkan arsitektur MVT (*Model-View-Template*), di mana model *Naïve Bayes* yang telah dilatih melalui *google colab* di diekspor ke dalam format *.pkl* berhasil diintegrasikan ke dalam *view* untuk memproses input teks dari pengguna dan menampilkan hasil prediksi secara *real-time* pada *template*. (2) Langkah-langkah klasifikasi data yang diperlukan untuk mendeteksi hoaks dalam penelitian ini meliputi beberapa tahapan utama. Proses dimulai dari pengumpulan data berita melalui *web scraping*, dilanjutkan dengan *data preprocessing* (mencakup *case folding*, *tokenization*, *stopwords removal*, dan *stemming*), kemudian pembobotan kata menggunakan metode TF-IDF untuk transformasi data teks menjadi vektor numerik, hingga akhirnya dilakukan pelatihan model klasifikasi menggunakan algoritma *Naïve Bayes*. (3) Tingkat akurasi yang dicapai oleh sistem deteksi hoaks ini menunjukkan bahwa metode *Naïve Bayes* efektif untuk diimplementasikan. Berdasarkan pengujian, sistem mampu mencapai tingkat akurasi tertinggi sebesar 97.26%. Hasil ini secara signifikan lebih unggul dibandingkan penelitian sejenis sebelumnya yang juga menggunakan *Naïve Bayes*, seperti penelitian oleh Nova Agustina (2021) yang memperoleh akurasi 81%.

E. DAFTAR PUSTAKA

Aghnia. (2022). *Filosofi Teorema Bayes terhadap Cara Pandang Kehidupan*.
<https://www.its.ac.id/news/2024/06/25/filosofi-teorema-bayes-terhadap-cara->

pandang-kehidupan/

- Alamsyah, I. L., Aulya, N., & Satriya, S. H. (2024). Transformasi Media Dan Dinamika Komunikasi Dalam Era Digital : Tantangan Dan Peluang Ilmu Komunikasi. *Jurnal Ilmiah Research Student*, 1(3), 168–181. <https://doi.org/10.61722/jirs.v1i3.554>
- Bahrudin, R. M., Ridwan, M., & Darmojo, H. S. (2019). Penerapan Helpdesk Ticketing System Dalam Penanganan Keluhan Penggunaan Sistem Informasi Berbasis Web. *JUTIS*, 7(1).
- Balai Monitor Spektrum Frekuensi Radio Kelas 1 Semarang. (2023). Kenali Cirinya Agar Tidak Mudah Termakan Hoaks. <https://balmonsemarang.postel.go.id/kenali-cirinya-agar-tidak-mudah-termakan-hoaks/>
- Box, C. (2023). Jurnal Widya Perancangan Aplikasi Sistem Informasi Penjualan Makanan Cepat Saji Berbasis Web Studi Kasus Kedai Jurnal Widya. 4(April), 117–139.
- Darmawan, I. A., Randy, M. F., Yuniato, I., Mutoffar, M. M., & Salis, M. T. P. (2022). Penerapan Data Mining Menggunakan Algoritma Apriori Untuk Menentukan Pola Golongan Penyandang Masalah Kesejahteraan Sosial. *Sebatik*, 26(1), 223–230. <https://doi.org/10.46984/sebatik.v26i1.1622>
- Diskominfo. (2022). Pengertian Hoax dan Cara Menangkalnya. <https://diskominfo.badungkab.go.id/artikel/42985-pengertian-hoax-dan-cara-menangkalnya>
- Dwi Normawati, S. A. P. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. 5(September), 697–711.
- Farid, M., Mursyidin, I. H., Rahmawati, D., Informatika, T., & Syekh-yusuf, U. I. (2025). Perancangan Sistem Informasi Poin Pelanggaran Siswa / I. 9(1), 195–200.
- Gani, A. G., Dewi, P. F., Sugiharto, A., Caringin, D., & Bandung, T. (2023). Sistem Informasi Point of Sale Berbasis Web Pada Dapur Caringin Tilu Bandung. *Jurnal Sistem Informasi Universitas Suryadarma*, 10(2). <https://doi.org/10.35968/jsi.v10i2.1072>
- Gat. (2023). Pemanfaatan Python dan Framework Django Sebagai Dashboard Sistem Informasi Pengelolaan Skripsi Pada STMIK Pontianak. *Seminar Nasional Corisindo*, 128–133. <https://stmikpontianak.org/ojs/index.php/corisindo/article/download/154/24>
- Gelar Guntara, R. (2023). Visualisasi Data Laporan Penjualan Toko Online Melalui Pendekatan Data Science Menggunakan Google Colab. *Jurnal Ilmiah Multidisiplin*, 2(6), 1–10. <https://journal-nusantara.com/index.php/JIM/article/view/1578>
- Giacomelli. (2023). Django REST Framework Basics. <https://testdriven.io/blog/drfs-basics/?#django-rest-framework>

- Google Cloud. (2025). Kelemahan machine learning. <https://cloud.google.com/learn/what-is-machine-learning?hl=id#disadvantages-of-machine-learning>
- Haana Udtari Anjani, Vitriani Vitriani, & Mulya Hastuti. (2024). Pemanfaatan Media Google Colaboratory Pada Mata Pelajaran Informatika di SMA Negeri 5 Pekanbaru. *SOKO GURU: Jurnal Ilmu Pendidikan*, 4(1), 101–108. <https://doi.org/10.55606/sokoguru.v4i1.3613>
- Hidayat, T., Siddiq, M. J., Jayasri, S., Suhendi, A., & Rizky, R. (2025). Analisis Sentimen Opini Masyarakat Terhadap Pilkada 2024 Di Media Sosial Twitter Menggunakan Algoritma Naïve Bayes. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(2), 609–622. <https://doi.org/10.23960/jitet.v13i2.6280>
- Jaga Fakta. (2023). Hati - hati! Kenali Bentuk Hoax yang Beredar. <https://jagafakta.semarangkota.go.id/2023/09/14/hati-hati-kenali-bentuk-hoax-yang-beredar/>
- KBBI (Kamus Besar Bahasa Indonesia). (n.d.-a). Pengertian deteksi. Retrieved January 15, 2025, from <https://kbbi.web.id/deteksi>
- KBBI (Kamus Besar Bahasa Indonesia). (n.d.-b). Pengertian Hoaks KBBI. <https://kbbi.kemdikbud.go.id/>.
- Munawar. (2019). Hibah Internal Sistem Pendeteksi Berita Palsu (Fake News) Di Media Sosial.
- Nur Prasetyo, D., & Rahmawati, D. (2023). Perancangan Sistem Pemilihan Ketua Osis Dan Ketua Kelas Di Smk Yuppentek 2 Kabupaten Tangerang Berbasis Web Menggunakan Unified Modelling Language (Uml). In *Jurnal Teknik Informatika Unis* (Vol. 11, Issue 2).
- Polresta Yogyakarta. (2024). Waspada Terhadap Hoax: Hindari Dampak Negatif di Era Digital. <https://jogja.polri.go.id/yogyakarta/tribrata-news/online/detail/waspada-terhadap-hoax--hindari-dampak-negatif-di-era-digital.html>
- Prastyo, D., Mursyidin, I. H., & Irawan, D. (2024). Klasifikasi Sentimen Komentar YouTube dengan NLP pada Debat Pilkada Banten. 7(2). <https://doi.org/10.32877/bt.v7i2.1833>
- Pricillia, T., & Zulfachmi. (2021). Perbandingan Metode Pengembangan Perangkat Lunak (Waterfall, Prototype, RAD). *Jurnal Bangkit Indonesia*, 10(1), 6–12. <https://doi.org/10.52771/bangkitindonesia.v10i1.153>
- Putra, A. P., Andriyanto, F., Karisman, K., Harti, T. D. M., & Sari, W. P. (2020). Pengujian Aplikasi Point of Sale Menggunakan Blackbox Testing. *Jurnal Bina Komputer*, 2(1), 74–78. <https://doi.org/10.33557/binakomputer.v2i1.757>
- Qodo Nathaniel Kenvel. (2023). Django Templates: Best Practices for Web Development. <https://www.qodo.ai/blog/django-templates-best-practices-for-web-development/>

- R. Destriana, s.m. Husain, N. Handayani, and A. T. P. S. (2021). Diagram UML Dalam Membuat Aplikasi.
- Sabita, H., Herwanto, R., Syafitri, Y., & Prasetyo, B. D. (2022). Pengembangan Aplikasi Akreditasi Program Studi Berbasis Framework Django. *Jurnal Informatika*, 22(1), 33–37. <https://doi.org/10.30873/ji.v22i1.3143>
- Sujana, D., Juhara, S., Rahayu, M., Mahmudin, M., & Fany Kurnia Pratama, M. (2023). Deep Learning Untuk Mendeteksi Kelelahan Pekerja. *Prosiding TAU SNARS-TEK Seminar Nasional Rekayasa Dan Teknologi*, 2(1), 11–17. <https://doi.org/10.47970/snarstek.v2i1.496>
- Tabitha, A., & Panjaitan, B. (2021). Deteksi Hoaks Pada Berita Berbahasa Indonesia Seputar COVID-19. 10, 76–85.
- UNMAHA. (2024). Django Rest Framework: Membangun API Restful dengan Mudah. <https://blog.unmaha.ac.id/django-rest-framework-membangun-api-restful-dengan-mudah/>
- Utami, M. C., Sunardi, S., & Iksari, I. H. (2024). Analisis Penerapan Machine Learning dalam Deteksi Objek, Klasifikasi, dan Optimasi Robotika. 2(4), 805–808.
- Vosoughi, S., Deb, R., & Aral, S. (2018). False News is Big News. *MIT Initiative on the Digital Economy Research Brief*, 359(6380), 1146–1151. http://ide.mit.edu/sites/default/files/publications/2017_IDE_Research_Brief_False_News.pdf
- Wijoyo, A., Saputra, A. Y., Ristanti, S., Rafly, S., Ban, S., Komputer, F. I., Informatika, T., Pamulang, U., Raya, J., & No, P. (2024). Pembelajaran Machine Learning. 3(2), 375–380.
- Yani, C. (2020). Pencegahan Hoax Di Media Sosial Guna Memelihara Harmoni Sosial. *Jurnal Lemhannas RI*, 7(4), 15–21. <https://doi.org/10.55960/jlri.v7i4.107>
- Zhou, X., & Zafarani, R. (2018). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. <https://doi.org/10.1145/339504>