

Implementasi Deteksi Berita Hoaks dan Fakta Menggunakan Support Vector Machine Berbasis Website

Implementation of Hoax and Fact News Detection Using Website-Based Support Vector Machine

Dina Aulia¹, Mahardika Abdi Prawira²

^{1,2}Sistem Informasi, Universitas Muhammadiyah Sumatera Utara

E-mail: dinaaulia3011@gmail.com, mahardikaabdiprawira@umsu.ac.id²

Abstrak

Perkembangan teknologi informasi dan komunikasi mempercepat penyebaran informasi melalui media digital, namun juga meningkatkan risiko penyebaran berita palsu dan disinformasi yang dapat memengaruhi opini publik serta stabilitas nasional. Penelitian ini bertujuan mengklasifikasikan berita hoaks dan fakta menggunakan metode Support Vector Machine (SVM). Data hoaks diperoleh dari dataset TurnBackHoax, sedangkan data fakta diambil dari portal berita daring CNN Indonesia. Data diproses melalui tahapan pra-pemrosesan teks, kemudian dilakukan pembobotan fitur menggunakan Term Frequency–Inverse Document Frequency (TF-IDF). Model dievaluasi menggunakan confusion matrix dengan metrik akurasi, presisi, recall, dan F1-score. Hasil pengujian menunjukkan bahwa model SVM memperoleh akurasi sebesar 96,43%, yang mengindikasikan performa klasifikasi sangat baik dalam membedakan berita hoaks dan fakta. Metode ini terbukti efektif dan berpotensi mendukung deteksi dini penyebaran disinformasi di Indonesia..

Kata kunci: klasifikasi berita; hoaks; Support Vector Machine; TF-IDF; deteksi disinformasi

Abstract

The development of information and communication technology has accelerated the spread of information through digital media, yet has also increased the risk of fake news and disinformation that can influence public opinion and national stability. This study aims to classify hoax and factual news using the Support Vector Machine (SVM) method. Hoax data were obtained from the TurnBackHoax dataset, while factual data were collected from the CNN Indonesia online news portal. The data were processed through text preprocessing stages, followed by feature weighting using Term Frequency–Inverse Document Frequency (TF-IDF). The model was evaluated using a confusion matrix with accuracy, precision, recall, and F1-score metrics. The results show that the SVM model achieved an accuracy of 96.43%, indicating excellent classification performance in distinguishing hoax from factual news. This method proves effective and has the potential to support early detection of disinformation in Indonesia.

Keywords: news classification; hoax; Support Vector Machine; TF-IDF; disinformation detection

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah membawa perubahan besar dalam berbagai aspek kehidupan, termasuk dalam cara individu maupun organisasi berinteraksi



dan bertukar informasi [1]. Di Indonesia, penetrasi internet yang tinggi menjadikan platform digital sebagai sumber informasi utama masyarakat. Namun, kondisi ini turut memicu meluasnya penyebaran hoaks dan disinformasi yang berdampak serius terhadap stabilitas sosial, politik, ekonomi, dan keamanan nasional. Kementerian Komunikasi dan Digital mencatat bahwa sepanjang tahun 2024 telah teridentifikasi sebanyak 1.923 konten hoaks dan informasi palsu yang beredar di Indonesia [2]. Kondisi ini diperparah oleh cepatnya penyebaran informasi digital yang kerap melampaui proses verifikasi, sehingga hoaks tersebar luas sebelum klarifikasi resmi dapat menjangkau masyarakat [3].

Dampak hoaks tidak dapat dipandang sepele. Diperkirakan 30% hingga hampir 60% masyarakat Indonesia terpapar hoaks saat berinteraksi secara daring, namun hanya 21% hingga 36% yang mampu mengenalinya [4]. Hoaks dirancang dengan judul sensasional dan provokatif untuk menarik perhatian, sehingga berpotensi menimbulkan kepanikan, kebingungan, bahkan konflik di tengah masyarakat [5]. Lebih jauh, informasi palsu dapat memengaruhi pengambilan keputusan di tingkat individu maupun kebijakan publik [3]. Oleh karena itu, dibutuhkan pendekatan otomatis yang mampu mendeteksi dan mengklasifikasikan berita hoaks secara cepat dan akurat.

Berbagai penelitian telah dilakukan untuk mengatasi permasalahan deteksi hoaks menggunakan pendekatan machine learning. Support Vector Machine (SVM) adalah metode supervised learning yang digunakan untuk menganalisis data dan mengenali pola, serta dapat diterapkan baik pada tugas regresi maupun klasifikasi [1]. Algoritma Support Vector Machine (SVM) terbukti menunjukkan performa tinggi dalam klasifikasi teks, termasuk deteksi berita hoaks. Ropikoh et al. membuktikan bahwa SVM dengan kernel linear mampu mencapai akurasi sebesar 92,90% dalam mengklasifikasikan berita hoaks terkait COVID-19 pada skenario pembagian data 80:20 [6]. Hasil serupa juga ditunjukkan oleh Wayan Sumartini Saraswati et al. (2023) dengan akurasi 99,4% pada dataset umum [7], serta FEBRIYANTY (2023) yang mencapai akurasi 100% dalam mendeteksi hoaks dari media online Indonesia menggunakan SVM [4]. Selain itu, Al-Khowarizmi et al. (2024) membuktikan bahwa SVM dengan kernel linear juga unggul dalam klasifikasi teks pada domain lain, yakni deteksi cyberbullying, dengan akurasi 92,3% yang melampaui kernel kubik sebesar 90% [8]. Metode ekstraksi fitur TF-IDF juga banyak digunakan dalam penelitian serupa karena kemampuannya merepresentasikan bobot kata secara efektif dalam dokumen teks [6]. Penelitian-penelitian tersebut menjadi landasan kuat untuk mengembangkan sistem klasifikasi hoaks yang lebih spesifik berdasarkan kategori konten.

Penelitian ini bertujuan mengevaluasi performa algoritma SVM dalam mengklasifikasikan berita hoaks dari dataset CNN Indonesia ke dalam dua kategori, yaitu politik dan non-politik, dengan memanfaatkan ekstraksi fitur TF-IDF. Proses yang dilakukan meliputi preprocessing teks, pelatihan model SVM, dan evaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil model kemudian diintegrasikan ke dalam antarmuka website berbasis Python yang memungkinkan pengguna mendeteksi hoaks secara langsung beserta skor kepercayaan. Penelitian ini diharapkan berkontribusi dalam upaya penanggulangan penyebaran hoaks di Indonesia secara lebih efisien dan terukur.

2. METODE PENELITIAN

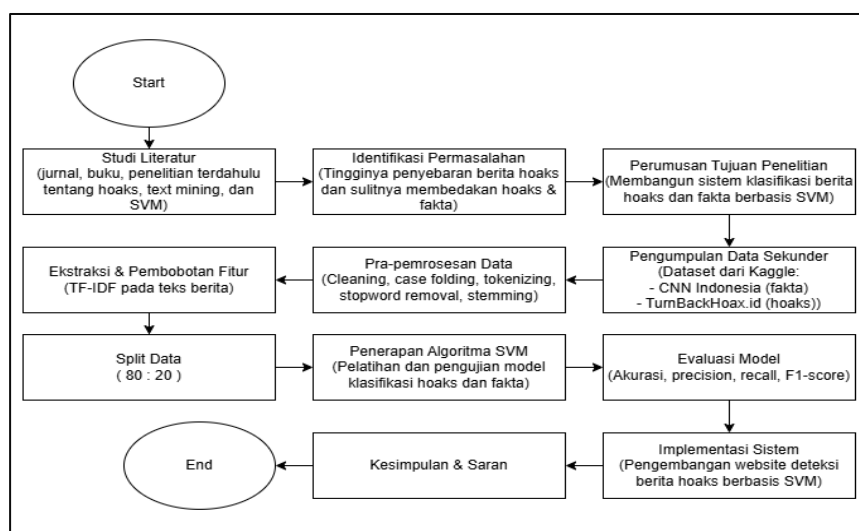
Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan metode komputasional untuk mengklasifikasikan berita hoaks menggunakan algoritma Support Vector Machine (SVM). Data dianalisis secara numerik dan dievaluasi melalui metrik akurasi, presisi, recall, dan F1-score.



2.1 Kerangka Berfikir Penelitian

Diagram kerangka pemikiran menggambarkan alur penelitian yang dimulai dari studi literatur mengenai berita hoaks, text mining, dan algoritma SVM, dilanjutkan dengan identifikasi masalah terkait tingginya penyebaran hoaks. Tujuan penelitian adalah membangun sistem klasifikasi berita hoaks dan fakta berbasis SVM.

Data sekunder diperoleh dari Kaggle (CNN Indonesia dan TurnBackHoax.id), lalu diproses melalui tahap pra-pemrosesan (cleaning, case folding, tokenizing, stopword removal, stemming) dan ekstraksi fitur TF-IDF. Dataset dibagi 80:20 untuk pelatihan dan pengujian model SVM, yang dievaluasi menggunakan akurasi, precision, recall, dan F1 score. Hasil akhirnya diimplementasikan sebagai sistem deteksi hoaks berbasis website, diakhiri dengan penarikan kesimpulan dan saran.



Gambar 1. Kerangka Berfikir Penelitian

2.2 Dataset

Dataset adalah suatu database didalam memori (in-memory). Dataset memiliki semua karakteristik, fitur dan fungsi dari database biasa [9]. Dataset yang digunakan merupakan data sekunder yang diperoleh dari repository Kaggle, hasil penggabungan dua sumber utama yaitu TurnBackHoax.id sebagai sumber berita hoaks dan CNN Indonesia sebagai sumber berita faktual. Dataset mencakup rentang waktu publikasi tahun 2015 hingga 2023 dengan total 20.013 record berita. Dataset diunduh dalam format Microsoft Excel (.xlsx) untuk memudahkan proses pengolahan awal sebelum tahap pemodelan.

Dataset terdiri dari dua label klasifikasi, yaitu hoaks dan fakta. Hoaks merupakan informasi yang sengaja dimanipulasi untuk memutarbalikkan fakta, bersifat meyakinkan namun tidak dapat diverifikasi kebenarannya, dan bertujuan menciptakan kebingungan serta ketidakpercayaan di masyarakat [10]. Sebaliknya, fakta adalah informasi yang kebenarannya dapat dibuktikan secara objektif, konsisten, serta memiliki dasar data yang jelas dan dapat dipertanggungjawabkan [11]. Penelitian lain yang dilakukan oleh [12], menegaskan bahwa fakta dapat dikenali melalui unsur

ADIKSIMBA, yaitu Apa, Di mana, Kapan, Siapa, Mengapa, dan Bagaimana. Berikut struktur dataset yang digunakan:

Tabel 1. Struktur Dataset

No	Nama Kolom	Deskripsi
1	no	Nomor urut data sebagai identitas setiap baris.
2	title	Judul berita baik dari CNN Indonesia maupun TurnBackHoax.id. Judul membantu model mengenali konteks awal konten berita
3	timestamp	Waktu publikasi berita. Informasi ini diambil dari masing-masing sumber sesuai tanggal rilis kontennya.
4	fulltext	Isi lengkap dari berita. Kolom ini berisi teks utama yang menjadi focus analisis klasifikasi.
5	tags	Label atau kategori berita dari sumber asli, seperti kategori politik, nasional, atau tag khusus tertentu.
6	author	Nama penulis atau pihak yang mengunggah konten. Pada CNN Indonesia biasanya berupa nama jurnalis. Sedangkan pada TurnBackHoax merupakan pencatat verifikasi.
7	url	Alamat tautan berita asli untuk keperluan pelacakan dan verifikasi.
8	Text_new	Cuplikan teks spesifik yang diambil dari bagian narasi yang beredar, biasanya berupa statement pendek yang menjadi inti klaim hoaks atau opini provokatif.
9	hoax	Label kelas untuk klasifikasi berita : 1. 1 = Hoaks (data dari TurnBackHoax.id) 2. 0 = Fakta (data dari CNN Indonesia)

2.3 Pra-pemrosesan Data

Sebelum data digunakan untuk pelatihan model, dilakukan tahap pra-pemrosesan teks yang meliputi empat langkah utama, yaitu cleaning (pembersihan karakter tidak relevan seperti simbol dan angka), case folding (konversi seluruh teks menjadi huruf kecil), tokenizing (pemecahan teks menjadi satuan kata), stopword removal (penghapusan kata umum yang tidak bermakna), dan stemming (pengubahan kata ke bentuk dasarnya). Tahapan ini bertujuan mengurangi noise pada data teks sehingga model dapat belajar dari fitur yang lebih representatif. Tahapan meliputi :

2.3.1 Case Folding

Rumus:

$$t_{lower} = lowercase(t_i)$$

Fungsi	Sebelum	Sesudah
Mengubah semua huruf menjadi kecil untuk menyeragamkan teks.	PUNYA HUTANG 50 MILIAR, SANDIAGA UNO GUGAT ANIES BASWEDAN	punya hutang 50 miliar sandiaga uno gugat anies baswedan

2.3.2 Penghapusan Angka, Tanda Baca, dan Simbol

Rumus :



$$t_{clean} = \text{remove_nonalphabet}(t_{lower})$$

Fungsi	Sebelum	Sesudah
Menghapus karakter non-alfabet yang tidak relevan.	punya hutang 50 miliar sandiaga uno gugat anies baswedan	punya hutang miliar sandiaga uno gugat anies baswedan

2.3.3 Stopword Removal

Rumus :

$$Tokens' = Tokens - Stopwords$$

Fungsi	Sebelum	Sesudah
Menghapus kata umum yang tidak memiliki makna penting.	punya hutang miliar sandiaga uno gugat anies baswedan	["punya", "hutang", "miliar", "sandiaga", "uno", "gugat", "anies", "baswedan"]

2.3.4 Stemming

Rumus :

$$Stem(w) = root(w)$$

Fungsi	Sebelum	Sesudah
Mengubah kata menjadi bentuk dasarnya.	["punya", "hutang", "miliar", "sandiaga", "uno", "gugat", "anies", "baswedan"]	["punya", "hutang", "miliar", "sandiaga", "uno", "gugat", "anies", "baswedan"]

2.3.5 Tokenizing

Rumus :

$$Tokens = \{w_1, w_2, \dots, w_n\}$$

Fungsi	Sebelum	Sesudah
Memecah teks menjadi kata-kata sebagai unit analisis.	["punya", "hutang", "miliar", "sandiaga", "uno", "gugat", "anies", "baswedan"]	["punya", "hutang", "miliar", "sandiaga", "uno", "gugat", "anies", "baswedan"]

2.3.6 Normalisasi Teks

Fungsi	Sebelum	Sesudah
Menyeragamkan variasi penulisan kata agar konsisten.	["punya", "hutang", "miliar", "sandiaga", "uno", "gugat", "anies", "baswedan"]	["punya", "utang", "miliar", "sandiaga", "uno", "gugat", "anies", "baswedan"]



2.4 Ekstraksi Fitur TF-IDF

Ekstraksi fitur dilakukan menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) untuk mengubah data teks menjadi representasi numerik. TF-IDF mengukur seberapa penting suatu kata dalam sebuah dokumen relatif terhadap seluruh koleksi dokumen. Pembagian data dilakukan dengan rasio 80:20, di mana 80% data digunakan sebagai data latih (training) dan 20% sebagai data uji (testing).

Setelah data bersih, dilakukan Ekstraksi Fitur menggunakan TF-IDF (Term Frequency-Inverse Document Frequency), yaitu metode yang memberikan bobot numerik pada setiap kata berdasarkan frekuensi kemunculannya dalam dokumen dibandingkan sebarannya di seluruh dataset. Hasilnya berupa matriks fitur yang merepresentasikan karakteristik unik setiap berita, dengan kata seperti "jokowi", "anies", dan "video" memiliki nilai TF-IDF tertinggi. Matriks ini selanjutnya digunakan sebagai input model SVM untuk menentukan *hyperplane* optimal dalam klasifikasi data.

```

10 Fitur Pertama:
['aa' 'aam' 'abad' 'abadi' 'abai' 'abang' 'abbas' 'abdul' 'abdullah' 'abe']

==== NILAI TF-IDF DOKUMEN PERTAMA ====
Kata: aa | TF-IDF: 0.000000
Kata: aam | TF-IDF: 0.000000
Kata: abad | TF-IDF: 0.000000
Kata: abadi | TF-IDF: 0.000000
Kata: abai | TF-IDF: 0.000000
Kata: abang | TF-IDF: 0.000000
Kata: abbas | TF-IDF: 0.000000
Kata: abdul | TF-IDF: 0.000000
Kata: abdullah | TF-IDF: 0.000000
Kata: abe | TF-IDF: 0.000000

==== 10 KATA DENGAN TOTAL TF-IDF TERTINGGI (GLOBAL) ====
Kata: jokowi | Total TF-IDF: 420.733141
Kata: soal | Total TF-IDF: 305.679179
Kata: anies | Total TF-IDF: 266.771397
Kata: jadi | Total TF-IDF: 266.434894
Kata: video | Total TF-IDF: 264.579532
Kata: tak | Total TF-IDF: 260.461521
Kata: pdip | Total TF-IDF: 243.822993
Kata: dpr | Total TF-IDF: 241.650229
Kata: milu | Total TF-IDF: 231.725656
Kata: foto | Total TF-IDF: 224.765186
    
```

Gambar 2. Implementasi TF_IDF

2.5 Penerapan Algoritma SVM

Algoritma Support Vector Machine (SVM) digunakan untuk membangun model klasifikasi berita hoaks dan fakta. Berdasarkan penjelasan dari [13], sasaran utama SVM adalah memaksimalkan margin antara dua kelas, di mana semakin besar margin yang dihasilkan, semakin baik kinerja model dalam menangani data baru yang belum pernah dilihat sebelumnya. Dalam penelitian ini, SVM bekerja dengan mencari hyperplane optimal yang memisahkan kelas hoaks dan fakta dengan margin terbesar. Model dilatih menggunakan data latih hasil ekstraksi TF-IDF, kemudian diuji pada data uji untuk menghasilkan prediksi klasifikasi.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Evaluasi Model SVM

Berdasarkan hasil evaluasi yang ditampilkan pada Gambar 3, model SVM berhasil mencapai akurasi sebesar 96,43% pada data uji. Nilai akurasi ini menunjukkan bahwa model mampu mengklasifikasikan berita dengan sangat baik dari keseluruhan data yang diujikan.



```

Evaluasi Model...

===== HASIL EVALUASI =====
Accuracy : 96.43 %

===== CONFUSION MATRIX =====
[[1865  61]
 [  82 1994]]

===== CLASSIFICATION REPORT =====
      precision    recall  f1-score   support

     0       0.96     0.97     0.96     1926
     1       0.97     0.96     0.97     2076

 accuracy          0.96          0.96          0.96     4002
  macro avg       0.96          0.96          0.96     4002
 weighted avg     0.96          0.96          0.96     4002
    
```

Gambar 3. Hasil Evaluasi Model SVM

3.1.1 Confusion Matrix

Berdasarkan Tabel dibawah ini, dari total 4.002 data uji, model berhasil memprediksi 1.865 data kelas negatif (berita fakta) dengan benar (True Negative/TN) dan 1.994 data kelas positif (berita hoaks) dengan benar (True Positive/TP). Model salah memprediksi 61 data fakta sebagai hoaks (False Positive/FP) dan 82 data hoaks sebagai fakta (False Negative/FN). Hasil ini menunjukkan bahwa model memiliki performa yang seimbang dalam mendeteksi kedua kelas.

Tabel 2. Confusion Matrix

	Prediksi Negatif (0)	Prediksi Positif (1)
Aktual Negatif (0)	1865 (TN)	61 (FP)
Aktual Positif (1)	82 (FN)	1994 (TP)

3.1.2 Classification Report

Berdasarkan Tabel 3.2, kelas fakta (0) memiliki precision 0,96, recall 0,97, dan F1-score 0,96, sedangkan kelas hoaks (1) memiliki precision 0,97, recall 0,96, dan F1-score 0,97. Nilai macro dan weighted average sebesar 0,96 menunjukkan model bekerja seimbang dan andal. Tingginya precision dan recall menandakan model mampu meminimalkan kesalahan prediksi serta mendeteksi data hoaks dengan baik.

Tabel 3. Classification Report

Kelas	Precision	Recall	F1-Score	Support
0 (Fakta)	0.96	0.97	0.96	1926
1 (Hoaks)	0.97	0.96	0.97	2076
Accuracy	-	-	0.96	4002
Macro Avg	0.96	0.96	0.96	4002
Weighted Avg	0.96	0.96	0.96	4002

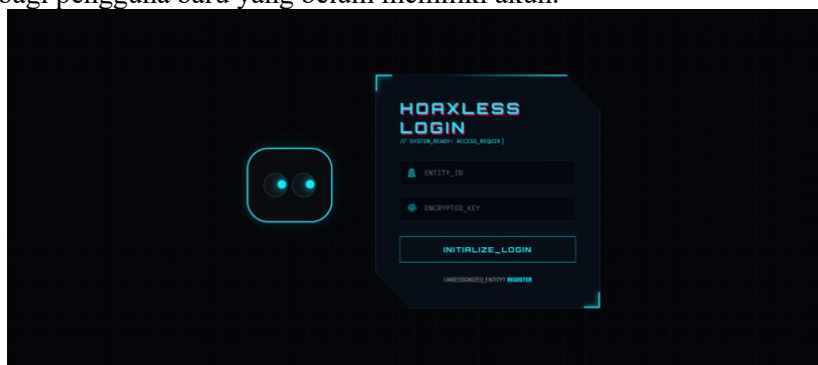
3.2 Hasil Implementasi Sistem Web HOAXLESS



Sistem deteksi hoaks HOAXLESS merupakan aplikasi web bertema cyberpunk yang menyediakan fitur autentikasi dan klasifikasi berita real-time menggunakan model SVM. Berikut penjelasan tiap halaman pada sistem.

3.2.1 Login

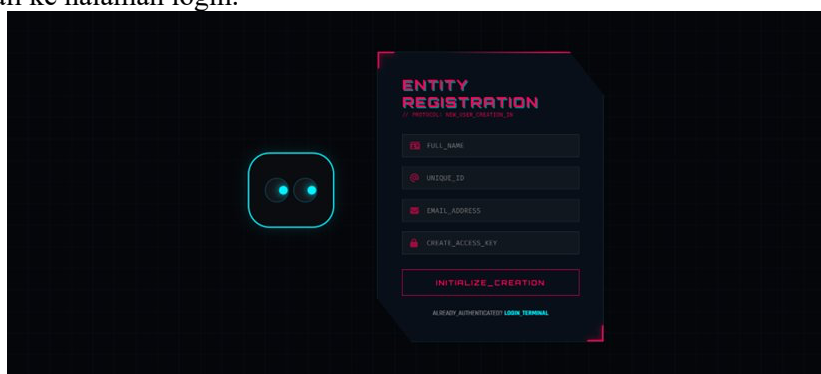
Berdasarkan Gambar 4, halaman login menampilkan form autentikasi dengan dua field input, yaitu Entity ID untuk memasukkan username dan Encrypted Key untuk memasukkan kata sandi. Terdapat tombol INITIALIZE_LOGIN untuk memproses autentikasi, serta tautan REGISTER bagi pengguna baru yang belum memiliki akun.



Gambar 4. Login Page

3.2.2 Register

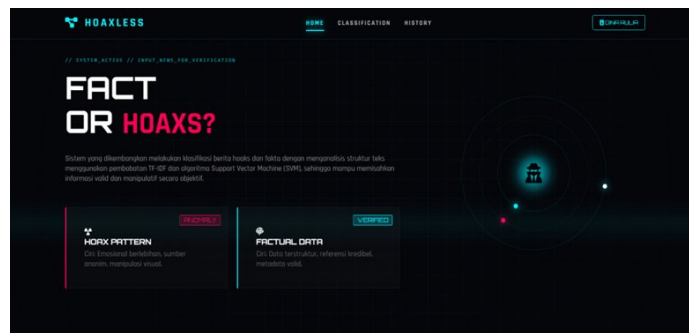
Berdasarkan Gambar 5, halaman registrasi atau Entity Registration menampilkan formulir pendaftaran yang terdiri dari empat field input, yaitu Full Name (nama lengkap), Unique ID (username), Email Address, dan Create Access Key (kata sandi). Setelah mengisi seluruh field, pengguna dapat menekan tombol INITIALIZE_CREATION untuk menyelesaikan proses pendaftaran. Bagi pengguna yang sudah memiliki akun, tersedia tautan LOGIN TERMINAL untuk kembali ke halaman login.



Gambar 5. Register Page

3.2.3 Home

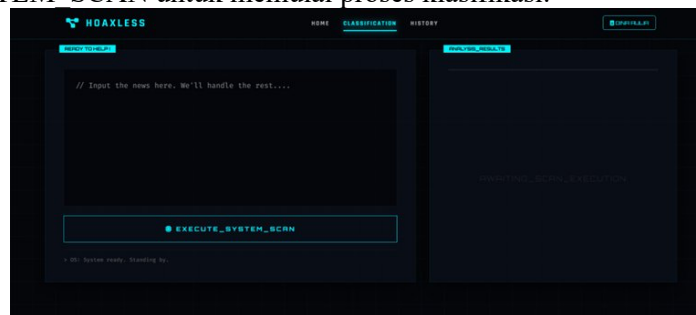
Berdasarkan Gambar 6, halaman home menampilkan tagline "FACT OR HOAXS?", deskripsi sistem (TF-IDF & SVM), kartu Hoax Pattern (ANOMALY) dan Factual Data (VERIFIED), serta menu Classification dan History.



Gambar 6. Home Page

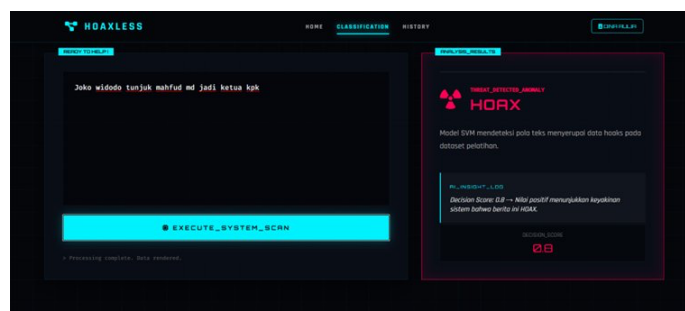
3.2.4 Classification

Berdasarkan Gambar 7, halaman classification menyediakan area input teks di sisi kiri dengan instruksi "Input the news here. We'll handle the rest...." dan panel ANALYSIS_RESULTS di sisi kanan yang menampilkan status "AWAITING_SCAN_EXECUTION" sebelum klasifikasi dijalankan. Pengguna dapat memasukkan teks berita kemudian menekan tombol EXECUTE_SYSTEM_SCAN untuk memulai proses klasifikasi.



Gambar 7. Classification Page

Berdasarkan Gambar 8, ketika pengguna memasukkan teks "Joko widodo tunjuk mahfid md jadi ketua kpk", sistem menampilkan hasil klasifikasi HOAX pada panel ANALYSIS_RESULTS. Panel tersebut memberikan keterangan bahwa model SVM mendeteksi pola teks yang menyerupai data hoaks pada dataset pelatihan, disertai informasi Decision Score sebesar 0,8 yang bernilai positif, menunjukkan keyakinan sistem bahwa berita tersebut adalah hoaks.

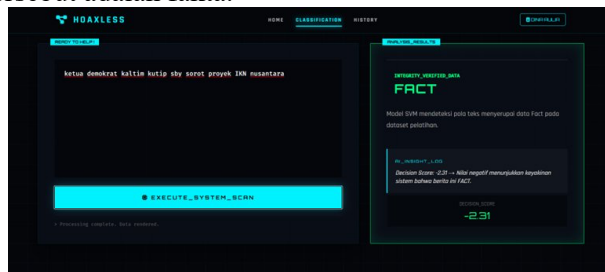


Gambar 8. Result Hoax

Berdasarkan Gambar 9, ketika pengguna memasukkan teks "ketua demokrat kaltim kutip sby sorot proyek IKN nusantara", sistem menampilkan hasil klasifikasi FACT pada panel ANALYSIS_RESULTS dengan label INTEGRITY_VERIFIED_DATA. Sistem memberikan keterangan bahwa model SVM mendeteksi pola teks yang menyerupai data Fact pada dataset



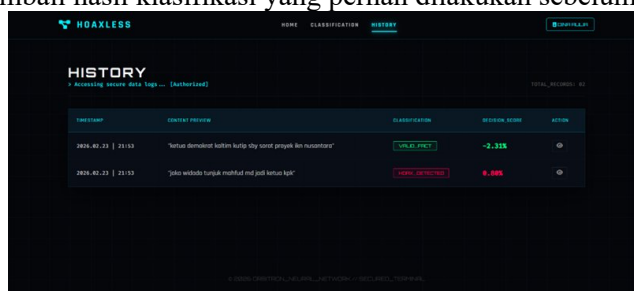
pelatihan, dengan Decision Score sebesar -2,31 yang bernilai negatif, menunjukkan keyakinan sistem bahwa berita tersebut adalah fakta.



Gambar 9. Result Fact

3.2.5 History

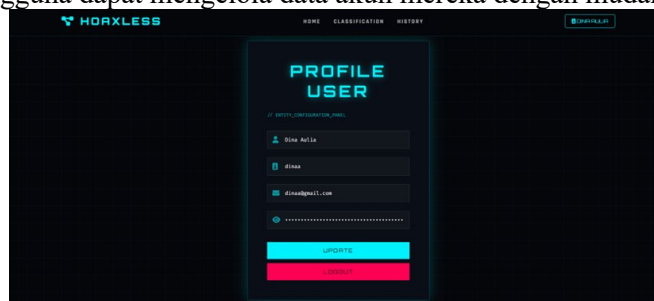
Berdasarkan Gambar 10, halaman history menampilkan daftar riwayat klasifikasi dalam bentuk tabel yang memuat kolom Timestamp (waktu klasifikasi), Content Preview (pratinjau teks berita yang diuji), Classification (hasil klasifikasi), Decision Score, dan Action. Setiap entri riwayat dilengkapi dengan label VALID_FACT berwarna cyan untuk berita fakta dan HOAX_DETECTED berwarna merah muda untuk berita hoaks. Fitur ini memudahkan pengguna untuk menelusuri kembali hasil klasifikasi yang pernah dilakukan sebelumnya.



Gambar 10. History Page

3.2.6 Profile

Berdasarkan Gambar 11, halaman Profile User menampilkan informasi akun pengguna yang meliputi nama lengkap, username, alamat email, dan kata sandi yang disamarkan. Terdapat dua tombol pada halaman ini, yaitu tombol UPDATE berwarna cyan untuk memperbarui data profil pengguna, dan tombol LOGOUT berwarna merah muda untuk keluar dari sistem. Halaman ini memastikan pengguna dapat mengelola data akun mereka dengan mudah dan aman.



Gambar 11. Profile Page

KESIMPULAN DAN SARAN

Penelitian ini berhasil mengklasifikasikan berita hoaks politik dan non-politik menggunakan metode Support Vector Machine (SVM) dengan akurasi tinggi sebesar 96,43%, yang didukung oleh tahapan preprocessing dan pembobotan TF-IDF yang efektif dalam meningkatkan kualitas data teks. Kelebihan penelitian ini terletak pada kemampuan SVM dalam menangani data berdimensi tinggi serta keberhasilan implementasi model ke dalam sistem berbasis web yang interaktif.

Namun, penelitian ini masih memiliki kekurangan, yaitu keterbatasan jumlah dan keseimbangan dataset serta penggunaan metode representasi fitur yang masih sederhana. Oleh karena itu, penelitian ini masih berpotensi untuk dikembangkan lebih lanjut melalui penggunaan dataset yang lebih besar, metode yang lebih kompleks, serta pengembangan fitur sistem yang lebih lengkap.

DAFTAR PUSTAKA

- [1] D. Dongoran and I. P. Sari, "Implementasi Klasifikasi Data Tracer Study Pada Universitas Muhammadiyah Sumatera Utara Dengan Pemanfaatan Data Mining Menggunakan Kombinasi Algoritma Support Vector Machine dan Neural Network," *Hello World Jurnal Ilmu Komputer*, vol. 4, no. 1, pp. 12–24, Apr. 2025, doi: 10.56211/helloworld.v4i1.619.
- [2] Kementerian Komunikasi dan Digital, "Komdigi Identifikasi 1.923 Konten Hoaks Sepanjang Tahun 2024," Siaran Pers No. 08/HM-KKD/01/2025. Accessed: Nov. 25, 2025. [Online]. Available: <https://www.komdigi.go.id/berita/siaran-pers/detail/komdigi-identifikasi-1923-konten-hoaks-sepanjang-tahun-2024>
- [3] Yopita Desriana Butar, "Analisis Penyebaran Hoax Di Media Sosial Dan Dampaknya Terhadap Masyarakat," *Jurnal Pendidikan, Bahasa dan Budaya*, vol. 3, no. 2, pp. 252–258, May 2024, doi: 10.55606/jpbb.v3i2.3201.
- [4] N. E. FEBRIYANTY, "DETEKSI BERITA HOAX DARI MEDIA ONLINE INDONESIA MENGGUNAKAN ALGORITMA NAIVE BAYES DAN SUPPORT VECTOR MACHINE," Malang, 2023.
- [5] A. Rahmadhany, A. Aldila Safitri, and I. Irwansyah, "Fenomena Penyebaran Hoax dan Hate Speech pada Media Sosial," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 3, no. 1, pp. 30–43, Jan. 2021, doi: 10.47233/jteksis.v3i1.182.
- [6] I. A. Ropikoh, R. Abdulhakim, U. Enri, and N. Sulistiyowati, "Penerapan Algoritma Support Vector Machine (SVM) untuk Klasifikasi Berita Hoax Covid-19," 2021. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [7] N. Wayan Sumartini Saraswati, I. Putu Krisna Suarendra Putra, I. Dewa Made Krishna Muku, and G. Dana Pramitha, "Support Vector Machine For Hoax Detection," *SINTECH JOURNAL*, vol. 6, pp. 107–117, Aug. 2023, [Online]. Available: <https://doi.org/10.31598>
- [8] Al-Khowarizmi, I. P. Sari, and H. Maulana, "Optimization of support vector machine with cubic kernel function to detect cyberbullying in social networks," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 22, no. 2, pp. 329–339, Apr. 2024, doi: 10.12928/TELKOMNIKA.v22i2.25437.
- [9] Yahya and Mahpuz, "Penggunaan Algoritma K-Means Untuk Menganalisis Pelanggan Potensial Pada Dealer SPS Motor Honda Lombok Timur Nusa Tenggara Barat," *Jurnal Informatika dan Teknologi*, vol. 2, no. 2, pp. 109–118, 2019.
- [10] M. Asriadi and Q. Hasyim, "Pelatihan Literasi Media Sosial Bagi Pemilih Pemula Tentang Hoax, Hate Speech dan Negative Campaign," vol. 02, pp. 124–133, Jun. 2024.
- [11] I. N. Khasanah, N. Syelena Azzahra, L. Febrianto, and M. R. Saifudin, "Perbedaan Fakta dan Non Fakta di Media : Definisi, Jenis, dan Memeriksa Fakta Informasi Digital," 2024.



- [12] Aulia Fanny and Kaswadi Kaswadi, “Peningkatan Kemampuan Analisis Fakta dan Opini Peserta Didik Kelas XI TEK 2 SMK Negeri 5 Surabaya dengan Model Pembelajaran Kooperatif,” *Pragmatik : Jurnal Rumpun Ilmu Bahasa dan Pendidikan* , vol. 2, no. 4, pp. 213–222, Sep. 2024, doi: 10.61132/pragmatik.v2i4.1071.
- [13] geeksforgeeks, “Algoritma Support Vector Machine (SVM),” <https://www.geeksforgeeks.org/machine-learning/support-vector-machine-algorithm/>.

