

**IMPLEMENTASI DETEKSI BERITA HOAKS DAN FAKTA
MENGUNAKAN METODE SUPPORT VECTOR MACHINE
(SVM) PADA TURNBACKHOAX DAN CNN INDONESIA
BERBASIS WEBSITE**

SKRIPSI

DISUSUN OLEH

DINA AULIA

NPM. 2209010105



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN
2026**

**IMPLEMENTASI DETEKSI BERITA HOAKS DAN FAKTA
MENGUNAKAN METODE SUPPORT VECTOR MACHINE
(SVM) PADA TURNBACKHOAX DAN CNN INDONESIA
BERBASIS WEBSITE**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer
(S.Kom) dalam Program Studi Sistem Informasi pada Fakultas Ilmu Komputer
dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara**

**DINA AULIA
NPM. 2209010105**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN
2026**

LEMBAR PENGESAHAN

Judul Skripsi : IMPLEMENTASI DETEKSI BERITA HOAKS DAN
FAKTA MENGGUNAKAN METODE SUPPORT
VECTOR MACHINE (SVM) PADA
TURNBACKHOAX DAN CNN INDONESIA
BERBASIS WEBSITE
Nama Mahasiswa : DINA AULIA
NPM : 2209010105
Program Studi : SISTEM INFORMASI

Menyetujui
Komisi Pembimbing



(Mahardika Abdi Prawira, S.Kom., M.Kom.)
NIDN. 0117088902

Ketua Program Studi



(Mahardika Abdi Prawira, S.Kom., M.Kom.)
NIDN. 0117088902

Dekan



(Dr. Al-Khwarizmi, S.Kom., M.Kom.)
NIDN. 0127099201

PERNYATAAN ORISINALITAS

**IMPLEMENTASI DETEKSI BERITA HOAKS DAN FAKTA
MENGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM)
PADA TURNBACKHOAX DAN CNN INDONESIA BERBASIS WEBSITE**

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, April 2026

Yang membuat pernyataan



Dina Aulia

NPM. 2209010105

**PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN
AKADEMIS**

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama : Dina Aulia
NPM : 2209010105
Program Studi : Sistem Informasi
Karya Ilmiah : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bedas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

**IMPLEMENTASI DETEKSI BERITA HOAKS DAN FAKTA
MENGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM)
PADA TURNBACKHOAX DAN CNN INDONESIA BERBASIS WEBSITE**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, April 2026

Yang membuat pernyataan



Dina Aulia

NPM. 2209010105

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap : Dina Aulia
Tempat dan Tanggal Lahir : Medan, 30 November 2003
Alamat Rumah : Jl. Sederhana No.72 Tembung Pasar 7
Telepon/Faks/HP : 081262419151
E-mail : dinaaulia3011@gmail.com
Instansi Tempat Kerja : -
Alamat Kantor : -

DATA PENDIDIKAN

SD : RIAD MADANI TAMAT: 2016
SMP : SMP N 29 MEDAN TAMAT: 2019
SMA : SMKN 1 PERCUT SEI TUAN TAMAT: 2022

KATA PENGANTAR



Alhamdulillah rabbil'alamin, puji dan syukur penulis panjatkan ke hadirat Allah SWT atas rahmat, kesehatan, dan kekuatan sehingga skripsi ini dapat diselesaikan dengan baik. Skripsi ini tidak hanya sebagai syarat memperoleh gelar Sarjana, tetapi juga merupakan hasil dari proses panjang yang penuh pembelajaran dan perjuangan. Penulis menyadari bahwa penyusunan skripsi ini tidak terlepas dari bantuan, dukungan, doa, serta bimbingan berbagai pihak.

Oleh karena itu, dengan penuh rasa hormat dan ketulusan hati, penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Allah SWT, atas segala rencana terbaik-Nya. Di saat penulis merasa lelah, ragu, dan hampir menyerah, pertolongan-Nya selalu datang dengan cara yang tidak pernah disangka.
2. Bapak Prof. Dr. Akrim, M.Pd, selaku Rektor Universitas Muhammadiyah Sumatera Utara (UMSU), yang telah memberikan kesempatan dan fasilitas bagi penulis untuk menempuh pendidikan di lingkungan akademik yang berkualitas.
3. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom., selaku Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU, yang telah mendukung terselenggaranya proses pendidikan dengan baik.
4. Ibu Dr. Firahti Rizky, S.Kom., M.Kom., selaku Wakil Dekan I Program Studi Sistem Informasi, yang telah memberikan arahan dan dukungan dalam proses akademik penulis.
5. Bapak Mahardika Abdi Prawira, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi serta Dosen Pembimbing yang dengan penuh kesabaran telah meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan, arahan, koreksi, serta motivasi kepada penulis. Setiap masukan

yang diberikan menjadi pembelajaran berharga dalam menyempurnakan skripsi ini.

6. Ibu Yoshida Sary, S.E., S.Kom, M.Kom., selaku Dosen Pembahas I yang telah memberikan kritik, saran, dan masukan yang sangat membangun demi penyempurnaan skripsi ini.
7. Bapak Martiano S.Pd, S.Kom., M.Kom., selaku Dosen Pembahas II yang telah memberikan kritik, saran, dan masukan yang sangat membangun demi penyempurnaan skripsi ini.
8. Ibu Indah Purnama Sari, S.T., M.Kom selaku dosen pembimbing akademik yang telah memberikan arahan, motivasi, serta bimbingan kepada penulis selama menjalani perkuliahan di Program Studi Sistem Informasi Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara.
9. Ayah tercinta, yang telah menjadi sosok panutan, pelindung, dan sumber kekuatan dalam hidup penulis. Terima kasih atas setiap tetes keringat, kerja keras, dan pengorbanan yang mungkin tidak pernah Ayah ceritakan. Nasihat yang sederhana namun penuh makna, serta kepercayaan yang Ayah berikan menjadi motivasi terbesar bagi penulis untuk terus melangkah dan tidak menyerah hingga skripsi ini dapat terselesaikan.
10. Ibu tercinta, yang doa-doanya tidak pernah putus mengiringi setiap langkah penulis. Terima kasih atas kasih sayang yang tulus, kesabaran yang tanpa batas, serta pelukan hangat yang selalu menjadi tempat kembali saat lelah dan ragu. Dukungan, perhatian, dan pengorbanan Ibu adalah kekuatan utama yang membuat penulis mampu bertahan melewati setiap proses dan tantangan dalam menyelesaikan skripsi ini.
11. Abang dan adik tersayang, yang selalu menjadi penghibur di tengah rasa lelah dan penat saat menghadapi skripsi. Terima kasih atas candaan sederhana yang mampu mengusir pusing, perhatian kecil yang terasa begitu berarti, serta kalimat-kalimat penuh keyakinan yang terus menguatkan bahwa semua ini pasti akan selesai dan menjadi yang terbaik. Kehadiran

kalian menjadi pengingat bahwa di balik setiap perjuangan, selalu ada keluarga yang percaya dan bangga tanpa syarat.

12. Teman-teman seperjuangan, Fikri Firmansyah , Dwi Shintya, Nabila Yusra, Rika dan Aulia Siddiq Asy Syifa yang bersama-sama tumbuh, belajar, dan berproses dalam dinamika dunia perkuliahan. Setiap diskusi, perdebatan, kerja kelompok, hingga perjuangan menyelesaikan skripsi menjadi pengalaman yang membentuk kedewasaan dan ketahanan penulis.
13. Aulia Raudatul Zannah, teman yang telah menemani penulis sejak bangku SMP hingga penulis menyelesaikan pendidikan S1, serta selalu memberikan dukungan, semangat, dan kebersamaan dalam setiap proses yang dilalui.
14. Penulis juga mengucapkan terima kasih yang sebesar-besarnya kepada teman-teman yang telah meluangkan waktu untuk hadir dan memberikan dukungan secara langsung pada saat sidang skripsi. Kehadiran, doa, serta semangat yang diberikan menjadi motivasi tersendiri bagi penulis dalam menghadapi dan menyelesaikan proses sidang dengan baik.
15. Diri sendiri, terima kasih sudah bertahan dan tidak menyerah sampai di titik ini. Terima kasih sudah kuat menghadapi rasa lelah, revisi yang tak ada habisnya, dan tekanan yang datang silih berganti. Kamu tetap memilih untuk bangkit, belajar, dan mencoba lagi. Penulis bangga pada diri sendiri karena sudah berani memulai, berani berproses, dan berani menyelesaikan apa yang telah dimulai. Skripsi ini adalah bukti nyata bahwa kamu mampu, kamu kuat, dan kamu layak untuk merasa bangga. Proud of you! kamu benar-benar berhasil melewati semuanya dengan versi terbaik dirimu.
16. Semua pihak yang terlibat langsung ataupun tidak langsung yang tidak dapat penulis ucapkan satu-persatu yang telah membantu penyelesaian skripsi ini.

**IMPLEMENTASI DETEKSI BERITA HOAKS DAN FAKTA
MENGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM)
PADA TURNBACKHOAX DAN CNN INDONESIA BERBASIS WEBSITE**

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 untuk 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, kemudian dilakukan pembobotan fitur menggunakan metode 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 menunjukkan performa klasifikasi yang sangat baik dalam membedakan berita hoaks dan fakta. Dengan demikian, metode ini berpotensi mendukung deteksi dini penyebaran disinformasi di Indonesia.

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

**IMPLEMENTATION OF HOAX AND FACT DETECTION USING THE
SUPPORT VECTOR MACHINE (SVM) METHOD ON TURNBACKHOAX
AND CNN INDONESIA WEB-BASED PLATFORMS**

ABSTRACT

The rapid advancement of information and communication technology has accelerated digital news distribution, yet it has simultaneously heightened the risk of spreading misinformation and fake news. Such disinformation poses a threat to public opinion and national stability. This study aims to classify news as either a "hoax" or a "fact" by employing the Support Vector Machine (SVM) algorithm. The dataset for hoaks was sourced from TurnBackHoax, while factual data was extracted from the CNN Indonesia online news portal. The data underwent rigorous pre-processing followed by feature weighting using the Term Frequency–Inverse Document Frequency (TF-IDF) method. Model performance was evaluated through a confusion matrix, measuring accuracy, precision, recall, and F1-score. The results demonstrate that the SVM model achieved an accuracy of 96.43%, indicating exceptional performance in distinguishing between fraudulent and credible news. Consequently, this approach offers significant potential for supporting early detection systems to combat the spread of disinformation in Indonesia.

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

DAFTAR ISI

ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitan.....	6
1.5 Manfaat Penelitian.....	7
BAB II LANDASAN TEORI	9
2.1 Berita Hoaks	9
2.2 Berita Fakta	10
2.3 Natural Language Processing (NLP).....	10
2.4 Dataset Penelitian	11
2.4.1 Struktur Dataset	12
2.5 Support Vector Machine (SVM)	13
2.5.1 Pengertian Support Vector Machine (SVM)	13
2.5.2 Jenis-jenis Support Vector Machine (SVM).....	14
2.5.3 Cara Kerja Support Vector Machine (SVM) Secara Umum	17
2.6 Penelitian Terdahulu.....	20
BAB III ANALISA DAN PERANCANGAN SISTEM.....	23
3.1 Metodologi Penelitian	23
3.2 Analisa Sistem.....	23
3.3 Analisa Kebutuhan Sistem	24
3.3.1 Kebutuhan Fungsional	24
3.3.2 Kebutuhan Non-Fungsional.....	26
3.4 Kerangka Berfikir Penelitian.....	27
3.5 Preprocessing Data	29
3.5.1 Case Folding	29
3.5.2 Penghapusan Angka, Tanda Baca, dan Simbol	29

3.5.3 Stopword Removal	30
3.5.4 Stemming / Tokenizing.....	31
3.5.5 Normalisasi Teks	31
3.6 Ekstraksi Fitur Teks	32
3.6.1 Term Frequency (TF).....	32
3.6.2 Inverse Document Frequency (IDF).....	33
3.6.3 TF-IDF	34
3.7 Model Support Vector Machine (SVM).....	35
3.7.1 Label Hoaks	36
3.7.2 Hyperplane SVM	36
3.7.3 Fungsi Klasifikasi	37
3.7.4 Optimasi (Soft Margin SVM).....	37
3.8 Entity Relationship Diagram	38
3.9 Analisa Alur Data Sistem	40
3.10 Perancangan Arsitektur Sistem.....	41
3.11 Rancangan Flowchart Sistem	42
3.12 Use Case Diagram	44
3.13 Perancangan Antarmuka Pengguna (User Interface Design).....	46
3.13.1 Dashboard	46
3.13.2 Halaman Klasifikasi.....	46
3.13.3 Hasil Klasifikasi.....	47
3.13.4 History	48
BAB IV IMPLEMENTASI DAN PENGUJIAN SISTEM.....	49
4.1 Kebutuhan Sistem.....	49
4.1.1 Kebutuhan Perangkat Keras (Hardware)	49
4.1.2 Kebutuhan Perangkat Lunak (Software).....	49
4.2 Implementasi Sistem	51
4.2.1 Implementasi Dataset.....	51
4.2.2 Karakteristik Kata pada Berita Hoaks dan Fakta Menggunakan WordCloud.....	51
4.2.3 Implementasi Preprocessing	53
4.2.4 Implementasi TF-IDF	55
4.2.5 Analisis Keseimbangan Distribusi Label Dataset.....	56

4.2.6 Split Data	57
4.2.7 Implementasi Model SVM.....	58
4.2.8 Implementasi Antarmuka Sistem.....	61
4.3 Pengujian Sistem	69
4.3.1 Pengujian Fungsional (Black Box Testing).....	70
4.4 Hasil Uji Coba	74
4.5 Kelebihan Aplikasi	75
4.6 Kekurangan Aplikasi	77
BAB V PENUTUP	78
5.1 Kesimpulan.....	78
5.2 Saran.....	79
DAFTAR PUSTAKA	81

DAFTAR TABEL

Tabel 2. 1 Struktur Dataset.....	12
Tabel 2. 2 Penjelasan cara kerja Support Vector Machine (SVM)	18
Tabel 2. 3 Tabel Penelitian Terdahulu	20
Tabel 4. 1 Blackbox Testing Fitur Register	70
Tabel 4. 2 Blackbox Testing Fitur Login	70
Tabel 4. 3 Blackbox Testing Halaman Dashboard.....	71
Tabel 4. 4 Blackbox Testing Fitur Klasifikasi Berita SVM.....	71
Tabel 4. 5 Blackbox Testing Halaman History	72
Tabel 4. 6 Blackbox Testing Halaman Profile	72
Tabel 4. 7 Blackbox Testing Penyimpanan ke Database	73

DAFTAR GAMBAR

Gambar 2. 1 Gambar Hyperplane Memisahkan Data dari Dua Kelas	14
Gambar 2. 2 Alur Kerja Support Vector Machine (SVM) Secara Umum	17
Gambar 3. 1 Kerangka Berfikir Penelitian.....	27
Gambar 3. 2 Entity Relationship Diagram	38
Gambar 3. 3 Alur Data Sistem	40
Gambar 3. 4 Flowchart Sistem.....	42
Gambar 3. 5 Use Case Diagram Sistem	44
Gambar 3. 6 Dashboard.....	46
Gambar 3. 7 Klasifikasi.....	47
Gambar 3. 8 Hasil Klasifikasi	47
Gambar 3. 9 History	48
Gambar 4. 1 Wordcloud Hoax	52
Gambar 4. 2 WordCloud Fakta	53
Gambar 4. 3 Implementasi TF-IDF.....	55
Gambar 4. 4 Distribusi Label	57
Gambar 4. 5 Implementasi Split Data	58
Gambar 4. 6 Confusion Matrix	61
Gambar 4. 7 Form Register.....	62
Gambar 4. 8 Form Login.....	63
Gambar 4. 9 Form Home	64
Gambar 4. 10 Form Classification	66
Gambar 4. 11 Deteksi Hoax	66
Gambar 4. 12 Deteksi Fakta.....	66

Gambar 4. 13 Form History	67
Gambar 4. 14 Profile	69

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan teknologi informasi dan komunikasi telah membawa perubahan besar dalam berbagai aspek kehidupan, termasuk dalam cara individu maupun organisasi berinteraksi dan bertukar informasi (Purnama Sari et al., 2025). Informasi dapat dibuat, dihilangkan, dan diubah dengan mudah. Teknologi juga memfasilitasi persebaran dan pertukaran informasi. Informasi yang telah diproses dan disebar ini menjadi salah satu sumber pengetahuan (knowledge) bagi manusia. Pengetahuan inilah yang kemudian mempengaruhi keputusan yang diambil di dalam masyarakat, termasuk keputusan mereka untuk beropini (Rahmawati et al., 2023). Namun, di balik kemajuan ini, terdapat ancaman serius yang dapat mengganggu stabilitas nasional, terutama dalam bentuk penyebaran hoaks dan disinformasi. Hoaks dan disinformasi adalah bagian dari tantangan era digital yang berdampak pada berbagai aspek kehidupan, termasuk politik, ekonomi, sosial budaya, dan keamanan negara. Di Indonesia, fenomena ini menjadi ancaman nyata bagi ketahanan nasional, mengingat penetrasi internet yang tinggi dan meningkatnya penggunaan media sosial sebagai sumber informasi utama Masyarakat (Sarjito, 2021).

Sebagai respons terhadap dinamika penyebaran informasi di era digital, berbagai media berupaya memperkuat perannya dalam menyajikan berita yang akurat dan kredibel, Salah satunya adalah CNN Indonesia. Pada tahun 2015, Transmedia bekerjasama dengan Turner Internasional untuk membentuk CNN di Indonesia yang dinamakan CNN Indonesia. CNN Indonesia adalah bagian dari

CNN International Commercial's Content Sales and Partnerships Group. Secara umum, CNN menjadi media terkemuka karena pemberitaan informasinya dianggap lebih akurat, memiliki branding yang kuat, up to date dalam menyebarkan informasi, faktual, berita yang disajikan rinci, inspiratif, dan lebih inovatif (Satrivi & Purnama, 2021).

Hoax menjadi masalah yang sangat signifikan karena dampaknya meluas dan berpengaruh besar terhadap masyarakat. Hoax dirancang untuk menarik perhatian pembaca, seringkali dengan menggunakan judul yang sensasional dan provokatif. Tingginya paparan masyarakat terhadap konten ini menjadi perhatian serius; diperkirakan 30% hingga hampir 60% masyarakat Indonesia mengalami paparan hoaks saat berinteraksi online, namun ironisnya, hanya 21% hingga 36% yang mampu mengenali atau mendeteksinya (FEBRIYANTY, 2023). Menanggapi meluasnya penyebaran informasi palsu ini, (Kementerian Komunikasi dan Digital, 2025) telah mengambil tindakan, berhasil mengidentifikasi dan mengklarifikasi sebanyak 1.923 konten hoaks, berita bohong, dan informasi palsu sepanjang tahun 2024. Hoax merupakan efek samping dari era keterbukaan, yang memiliki peluang untuk menciptakan perpecahan dan permusuhan karena dapat membuat masyarakat bingung akan sebuah kebenaran informasi (Rahmadhany et al., 2021). Hoaks dapat menimbulkan kepanikan, kebingungan, serta memicu konflik di tengah masyarakat. Bahkan lebih dari itu, informasi palsu juga berpotensi memengaruhi proses pengambilan keputusan, baik di tingkat individu maupun dalam perumusan kebijakan publik (Yopita Desriana Butar, 2024). Kondisi ini diperparah oleh cepatnya penyebaran informasi digital yang sering melampaui proses verifikasi.

Akibatnya, hoaks kerap tersebar luas sebelum klarifikasi resmi muncul, sehingga upaya penanggulangan misinformasi menjadi kurang efektif.

Untuk menjawab permasalahan ini, pendekatan yang dapat dilakukan adalah dengan mengevaluasi performa algoritma Support Vector Machine (SVM) dalam mengklasifikasikan berita menjadi dua kategori : Hoaks dan Fakta. Support Vector Machine atau SVM adalah algoritme pembelajaran mesin yang diawasi yang dapat digunakan untuk klasifikasi dan regresi. Cara kerja SVM didasarkan pada SRM atau Structural Risk Minimization yang dirancang untuk mengolah data menjadi Hyperplane yang mengklasifikasikan ruang input menjadi dua kelas. Teori SVM diawali dengan pengelompokan kasus-kasus linier yang dapat dipisahkan dengan hyperplane dan dibagi menurut kelasnya. Konsep SVM diawali dengan masalah klasifikasi dua kelas sehingga membutuhkan set pelatihan positif dan negatif. SVM akan berusaha mendapatkan hyperplane (pemisah) sebaik mungkin untuk memisahkan kedua kelas dan memaksimalkan margin kedua kelas tersebut (Binus University, 2022).

Pemilihan algoritma Support Vector Machine (SVM) didasari oleh banyak penelitian sebelumnya, algoritma ini menunjukkan performa yang tinggi dalam klasifikasi teks, termasuk dalam konteks deteksi berita hoaks. Berdasarkan hasil penelitian yang dilakukan oleh (Ropikoh et al., 2021) Algoritma Support Vector Machine (SVM) dengan kernel linear memiliki hasil prediksi yang bagus pada skenario 3 (80:20) karena model sanggup dalam mengklasifikasikan berita hoax dan bukan hoax covid-19. Hasil yang didapat yaitu ada 111 data hoax yang diprediksi hoax, ada 61 data hoax yang diprediksi bukan hoax. Sedangkan data bukan hoax yang diprediksi hoax ada 55 dan data bukan hoax diprediksi bukan hoax

ada 1408. Selain itu akurasi yang didapat pada skenario 80:20 juga memiliki nilai akurasi tertinggi yaitu sebesar 92,90%.

Dalam mengimplementasikan sistem ini, langkah-langkah yang dilakukan meliputi pengumpulan dan pengolahan data, preprocessing teks, ekstraksi fitur menggunakan metode TF-IDF, pelatihan model SVM, hingga evaluasi performa model menggunakan metrik seperti akurasi, presisi, recall, dan F1-score. Setelah model menunjukkan performa yang baik, sistem akan diintegrasikan dengan antarmuka website berbasis Python. Website ini menyediakan kolom input bagi pengguna untuk menempelkan teks berita, dan sistem akan menampilkan hasil klasifikasi apakah berita tersebut termasuk hoaks atau fakta, lengkap dengan skor kepercayaannya. Penelitian ini diharapkan dapat membantu masyarakat dalam mengenali berita palsu dengan lebih cepat dan efisien, sekaligus menjadi kontribusi dalam upaya mengurangi penyebaran hoaks di Indonesia.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini :

1. Bagaimana menganalisis karakteristik berita hoaks dan fakta pada dataset TurnBackHoax dan CNN Indonesia agar dapat digunakan dalam proses klasifikasi?
2. Bagaimana menguji tingkat akurasi dan performa metode *Support Vector Machine (SVM)* dalam mengklasifikasikan berita menjadi kategori hoaks dan fakta?

3. Bagaimana merancang sistem deteksi berita hoaks dan fakta menggunakan metode *Support Vector Machine (SVM)* yang sesuai untuk teks berbahasa Indonesia?
4. Bagaimana mengimplementasikan hasil model deteksi berita hoaks & fakta tersebut ke dalam bentuk sistem berbasis website yang dapat digunakan oleh pengguna secara interaktif?

1.3 Batasan Masalah

Untuk menghindari pembahasan yang terlalu luas, terdapat beberapa batasan permasalahan, yaitu :

1. Penelitian ini hanya berfokus pada deteksi berita hoaks dan fakta dalam konteks berita online berbahasa Indonesia.
2. Dataset yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari repository Kaggle, dengan sumber dan batasan sebagai berikut :
 - a) Data berita faktual berasal dari **CNN Indonesia** yang mencakup periode publikasi tahun 2021–2023.
 - b) Data berita hoaks berasal dari **TurnBackHoax.id** yang mencakup periode publikasi tahun 2015–2023.
3. Metode yang digunakan dalam penelitian ini hanya berfokus pada algoritma Support Vector Machine (SVM) tanpa melakukan perbandingan langsung dengan algoritma lain.
4. Proses pengolahan data teks dibatasi pada tahap preprocessing teks seperti pembersihan data (cleaning), tokenisasi, stopword removal, dan stemming menggunakan pustaka bahasa Indonesia.

5. Implementasi sistem hanya dilakukan dalam bentuk website berbasis lokal (localhost) sebagai media uji coba, belum sampai tahap implementasi pada server publik atau integrasi dengan platform eksternal.
6. Evaluasi performa model dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score tanpa melibatkan pengujian real-time terhadap berita baru yang terus berkembang di internet.

1.4 Tujuan Penelitian

Adapun beberapa hal yang menjadi tujuan dari penelitian ini adalah sebagai berikut :

1. Untuk menganalisis karakteristik berita hoaks dan fakta yang terdapat pada dataset *TurnBackHoax* dan *CNN Indonesia* sehingga dapat digunakan sebagai dasar dalam proses klasifikasi berita menggunakan metode pembelajaran mesin. Analisis ini dilakukan dengan memahami struktur teks, pola bahasa, serta kata-kata kunci yang membedakan berita hoaks dari berita fakta.
2. Untuk merancang sistem deteksi berita hoaks dan fakta berbasis teks berbahasa Indonesia dengan menerapkan metode SVM. Perancangan mencakup tahapan preprocessing data, pelatihan model, serta pengujian terhadap data uji agar sistem yang dibangun dapat berfungsi secara optimal dan sesuai dengan karakteristik bahasa Indonesia.
3. Untuk mengimplementasikan hasil model deteksi ke dalam sistem berbasis website yang bersifat interaktif dan mudah digunakan oleh pengguna. Melalui sistem ini, pengguna dapat memasukkan teks berita atau link berita untuk kemudian mendapatkan hasil klasifikasi apakah berita tersebut tergolong hoaks atau fakta berdasarkan model SVM yang telah dilatih.

4. Untuk mengukur tingkat akurasi dan performa metode Support Vector Machine (SVM) dalam melakukan klasifikasi berita ke dalam dua kategori, yaitu hoaks dan fakta. Evaluasi dilakukan menggunakan metrik pengujian seperti akurasi, presisi, recall, dan F1-score guna mengetahui seberapa baik model mampu mengenali dan membedakan jenis berita.

1.5 Manfaat Penelitian

Adapun manfaat penelitian yang bisa didapat dari penulisan penelitian ini adalah :

1. Bagi Kampus

Penelitian ini dapat menjadi referensi ilmiah dan tambahan literatur dalam bidang kecerdasan buatan (Artificial Intelligence) dan pemrosesan bahasa alami (Natural Language Processing), khususnya terkait penerapan algoritma *Support Vector Machine (SVM)* untuk klasifikasi teks berbahasa Indonesia. Selain itu, hasil penelitian ini dapat digunakan sebagai contoh penerapan nyata konsep machine learning dalam mendukung pencegahan penyebaran berita hoaks di era digital, sehingga memperkaya sumber penelitian di lingkungan akademik.

2. Bagi Penulis (Mahasiswa)

Melalui penelitian ini, penulis mendapatkan pemahaman mendalam tentang tahapan pengembangan sistem berbasis machine learning, mulai dari analisis data, preprocessing teks, pelatihan model, evaluasi performa, hingga implementasi sistem berbasis website. Penelitian ini juga membantu penulis dalam mengasah kemampuan riset, logika pemrograman, dan pengolahan data teks, yang berguna untuk pengembangan diri serta karier di bidang teknologi informasi dan data science.

3. Bagi Masyarakat

Penelitian ini diharapkan dapat memberikan kontribusi nyata dalam membantu masyarakat Indonesia mengenali berita hoaks secara lebih mudah dan cepat melalui sistem deteksi berbasis website yang interaktif. Dengan adanya sistem ini, masyarakat dapat memverifikasi kebenaran suatu berita sebelum menyebarkannya, sehingga mampu menekan laju penyebaran hoaks dan meningkatkan literasi digital di lingkungan publik.

BAB II

LANDASAN TEORI

2.1 Berita Hoaks

Hoax adalah informasi yang sengaja dimanipulasi untuk menyembunyikan fakta sebenarnya. Dengan kata lain, hoax merupakan usaha untuk memutarbalikkan fakta dengan menggunakan informasi yang terlihat meyakinkan tetapi tidak dapat diverifikasi kebenarannya. Hoax juga dapat diartikan sebagai upaya mengaburkan informasi yang sebenarnya, dengan cara membanjiri media dengan pesan yang salah untuk menyembunyikan pesan yang benar. Tujuan dari penyebaran hoax yang disengaja adalah menciptakan ketidakamanan, ketidaknyamanan, dan kebingungan di kalangan masyarakat. Dalam keadaan bingung tersebut, Masyarakat cenderung mengambil keputusan yang lemah, tidak meyakinkan, bahkan salah. Awalnya, perkembangan hoax di media sosial dimulai sebagai sarana untuk perisakan (Asriadi & Hasyim, 2024).

Penyebaran berita palsu semakin banyak terjadi, terutama melalui media online, yang menunjukkan bahwa literasi digital masyarakat masih rendah. Informasi yang belum dicek kebenarannya dengan mudah dipercaya dan dibagikan lagi tanpa memikirkan dampak yang mungkin terjadi. Hal ini semakin parah karena adanya pihak-pihak tertentu yang sengaja membuat berita palsu untuk kepentingan politik, ekonomi, atau ideologi. Oleh karena itu, penting bagi masyarakat untuk lebih waspada dan kritis dalam menerima serta menyaring informasi, agar bisa memutuskan mata rantai penyebaran hoaks dan menjaga kestabilan sosial.

2.2 Berita Fakta

Fakta adalah informasi yang kebenarannya dapat dibuktikan dan diverifikasi. Dalam konteks ilmiah, fakta merupakan hasil pengamatan objektif yang dapat diuji kembali oleh siapa pun dengan metode yang sama. Fakta bersifat konsisten, memiliki dasar data yang jelas, serta dapat dipertanggungjawabkan kebenarannya. Sebaliknya, informasi yang tidak dapat dibuktikan atau telah dimanipulasi untuk memengaruhi opini publik tergolong sebagai non-fakta, seperti opini, spekulasi, atau hoaks (Khasanah et al., 2024). Penelitian lain yang dilakukan oleh (Aulia Fanny & Kaswadi Kaswadi, 2024) menegaskan bahwa fakta dapat dikenali melalui unsur ADIKSIMBA, yaitu *Apa, Di mana, Kapan, Siapa, Mengapa, dan Bagaimana*. Unsur-unsur tersebut membantu memastikan bahwa informasi yang disampaikan benar-benar menggambarkan suatu peristiwa nyata secara lengkap dan dapat diverifikasi. Berdasarkan kedua pandangan tersebut, berita fakta dapat dipahami sebagai laporan peristiwa yang disusun berdasarkan data nyata, melalui proses verifikasi yang jelas, serta memenuhi unsur-unsur informasi dasar. Dalam konteks penelitian ini, berita fakta yang digunakan berasal dari CNN Indonesia sebagai media kredibel yang menerapkan standar jurnalistik profesional, sehingga setiap informasinya telah melalui proses pemeriksaan dan dapat dipertanggungjawabkan kebenarannya.

2.3 Natural Language Processing (NLP)

Natural Language Processing (NLP) adalah bagian dari kecerdasan buatan yang bertujuan untuk memfasilitasi komunikasi antara komputer dan bahasa manusia. Tujuan utamanya adalah agar mesin bisa membaca, memahami, dan memproses bahasa manusia secara otomatis (Mubarok & Abdi, 2024). Dalam

penelitian "Implementasi Deteksi Berita Hoaks dan Fakta Menggunakan Metode Support Vector Machine (SVM) pada TurnBackHoax dan CNN Indonesia Berbasis Website", NLP berperan penting sebagai tahap awal sebelum data berita diolah lebih lanjut. Proses NLP meliputi beberapa langkah seperti tokenisasi, pembersihan teks, normalisasi, dan ekstraksi fitur, yang bertujuan untuk mengubah teks berita yang masih mentah menjadi bentuk numerik agar bisa dipahami oleh algoritma SVM. Dengan demikian, kualitas proses NLP sangat memengaruhi tingkat akurasi model dalam membedakan berita hoaks dan berita yang benar pada dataset TurnBackHoax dan CNN Indonesia.

2.4 Dataset Penelitian

Dataset adalah suatu database didalam memori (in-memory). Dataset memiliki semua karakteristik, fitur dan fungsi dari database biasa. Dataset dapat memiliki banyak tabel, dan tabel- tabel dapat memiliki hubungan (relationship). Tabel-tabel pada suatu dataset dapat memiliki foreign key dan integritas referensial. Dataset adalah objek yang merepresentasikan data dan relasinya di memory. Strukturnya mirip dengan data yang ada di database. Dataset berisi koleksi dari data tabel dan data (Yahya & Mahpuz, 2019).

Dataset yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari repository Kaggle, hasil penggabungan dua sumber data utama, yaitu TurnBackHoax dan CNN Indonesia, sehingga membentuk satu dataset terpadu. Penggunaan data sekunder ini dilakukan tanpa proses pengumpulan data secara langsung dari responden. Setelah proses penggabungan dan penyesuaian struktur data, diperoleh total 20.013 record berita dengan rentang waktu publikasi tahun 2015 hingga 2023.

Dataset tersebut terdiri dari berita hoaks yang bersumber dari **TurnBackHoax** pada periode **2015–2023** dan berita faktual yang bersumber dari **CNN Indonesia** pada periode **2021–2023**. Seluruh dataset diunduh melalui repository Kaggle dalam format Microsoft Excel (.xlsx), sehingga memudahkan proses pengolahan awal, seperti penggabungan data, pembersihan data (*data cleaning*), serta transformasi teks sebelum digunakan pada tahap pemodelan dan klasifikasi.

2.4.1 Struktur Dataset

Dataset gabungan yang digunakan dalam penelitian ini terdiri dari beberapa kolom utama yang berfungsi sebagai elemen penting dalam proses pengolahan dan analisis data. Setiap kolom dirancang untuk merepresentasikan informasi spesifik dari masing-masing entri berita, sehingga model dapat memahami konteks, isi, serta karakteristik data secara lebih komprehensif. Adapun kolom-kolom utama tersebut meliputi :

Tabel 2. 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 fokus 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 <i>statement</i> 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)

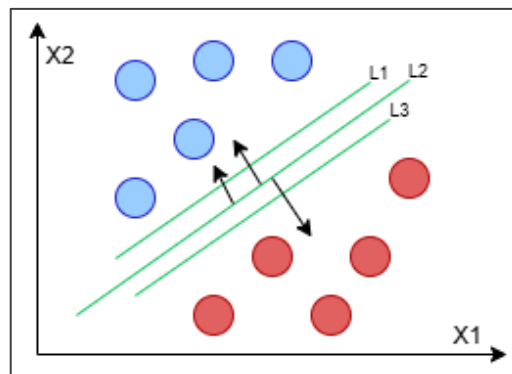
2.5 Support Vector Machine (SVM)

2.5.1 Pengertian Support Vector Machine (SVM)

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 (Dongoran & Sari, 2025). Sebagai algoritma pembelajaran terawasi (*supervised learning*), SVM bekerja berdasarkan konsep *Structural Risk Minimization* (SRM), yaitu pendekatan yang bertujuan meminimalkan risiko kesalahan dengan membentuk sebuah *hyperplane* sebagai

pemisah antar kelas pada ruang input. Teori SVM berawal dari permasalahan klasifikasi dua kelas, yang memerlukan data latih berisi kelas positif dan negatif. Pada prosesnya, SVM akan mencari *hyperplane* terbaik yang mampu memisahkan kedua kelas tersebut secara optimal dengan cara memaksimalkan margin di antara keduanya (Binus University, 2022) .

Berdasarkan penjelasan dari (geeksforgeeks, 2025) Sasaran utama SVM adalah memaksimalkan margin antara dua kelas. Semakin besar margin, semakin baik kinerja model pada data baru dan yang belum pernah dilihat sebelumnya.



Gambar 2. 1 Gambar Hyperplane Memisahkan Data dari Dua Kelas

2.5.2 Jenis-jenis Support Vector Machine (SVM)

Berdasarkan penjelasan dari (geeksforgeeks, 2025), Support Vector Machines (SVM) dapat diklasifikasikan ke dalam dua bagian utama berdasarkan sifat batas keputusan yang dihasilkan. Sejalan dengan itu, menurut (Andi Zulherry et al., 2024), algoritma SVM umumnya dikelompokkan ke dalam dua kategori utama, yaitu SVM linier dan SVM non-linier, yang masing-masing digunakan sesuai dengan pola dan kompleksitas data yang dianalisis, sehingga diperlukan penjelasan lebih lanjut mengenai kedua jenis SVM tersebut. yaitu :

1. SVM Linier

SVM Linier menggunakan batas keputusan linier untuk memisahkan titik data dari kelas yang berbeda. Jika data dapat dipisahkan secara linier dengan tepat, SVM linier sangat cocok. Ini berarti bahwa satu garis lurus (dalam 2D) atau bidang hiper (dalam dimensi yang lebih tinggi) dapat sepenuhnya membagi titik data ke dalam kelasnya masing-masing. Bidang hiper yang memaksimalkan margin antara kelas adalah batas keputusan.

Rumus dasarnya adalah :

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + \mathbf{b}$$

Penjelasan komponen sebagai berikut :

- a) $f(\mathbf{x})$ sebagai Decision Function merupakan fungsi keputusan yang menghasilkan nilai scalar untuk menentukan klasifikasi suatu data.
- b) \mathbf{x} sebagai Input Vector merupakan representasi data masukan dalam bentuk vector fitur.
- c) \mathbf{w} sebagai Weight Vector atau Bobot merupakan yang menentukan orientasi atau kemiringan dari hyperplane.
- d) \mathbf{w}^T sebagai Transpose of \mathbf{w} yang digunakan secara sistematis untuk memungkinkan operasi perkalian titik (*dot product*) dengan vector input \mathbf{x}
- e) \mathbf{b} sebagai Bias merupakan sebuah konstanta yang berfungsi untuk mengatur jarak atau pergeseran *hyperplane* dari titik pusat (*origin*).

2. SVM Non-Linear

SVM Non-Linear dapat digunakan untuk mengklasifikasikan data jika data tersebut tidak dapat dipisahkan menjadi dua kelas dengan garis lurus (dalam kasus 2D). Dengan menggunakan fungsi kernel, SVM non linier dapat menangani data yang dapat dipisahkan secara non-linier. Data masukan asli diubah oleh fungsi kernel ini menjadi ruang fitur berdimensi lebih tinggi tempat titik data dapat dipisahkan secara linier. SVM linier digunakan untuk menemukan batas keputusan non-linier dalam ruang yang dimodifikasi ini.

Rumus dasarnya adalah :

$$f(\mathbf{x}) = \sum_{i=1}^n \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b$$

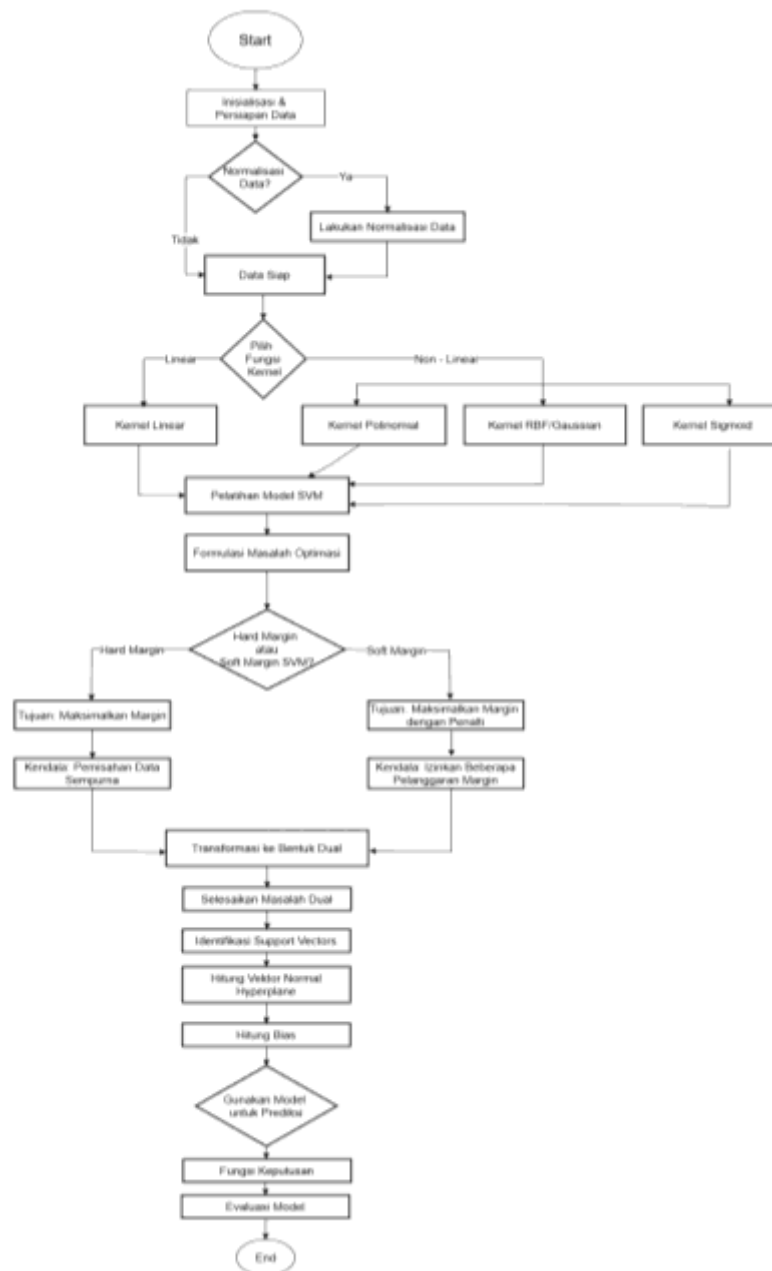
Penjelasan komponen sebagai berikut :

- a) $\sum_{i=1}^n$ sebagai Summation merupakan operator penjumlahan yang mempersentasikan akumulasi pengaruh dari titik data latih (n) yang berkontribusi dalam pembentukan model klasifikasi.
- b) α_i sebagai Lagrange Multiplier merupakan koefisien yang mempersentasikan bobot pengaruh dari setiap titik data \mathbf{x}_i terhadap penentuan hyperplane.
- c) y_i sebagai Label merupakan nilai target atau kategori data latih, yang umumnya direpresentasikan dengan nilai +1 untuk kelas positif dan -1 untuk kelas negative.
- d) $K(\mathbf{x}_i, \mathbf{x})$ sebagai Kernel Function merupakan fungsi pemetaan matematis yang memungkinkan SVM beroperasi pada ruang

dimensi tinggi tanpa harus menghitung koordinat data pada ruang tersebut secara eksplisit (*Kernel Trick*)



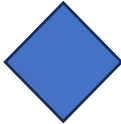


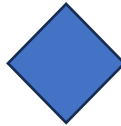



- e) x_i sebagai Support Vectors merupakan subset dari data latih yang terletak paling dekat dengan *hyperplane* atau berada di dalam area margin




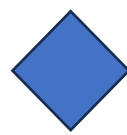






2.5.3 Cara Kerja Support Vector Machine (SVM) Secara Umum










Gambar 2. 2 Alur Kerja Support Vector Machine (SVM) Secara Umum

Tabel 2. 2 Penjelasan cara kerja Support Vector Machine (SVM)

No	Tahapan dalam Flowchart	Simbol	Penjelasan Detail
1	Start		Awal proses pembangunan model SVM dimulai.
2	Inisialisasi & Persiapan Data		Melakukan import dataset, membersihkan data (cleaning), dan preprocessing teks seperti tokenizing, stopword removal, stemming, serta ekstraksi fitur (TF-IDF atau BoW).
3	Cek Normalisasi Data		Mengecek apakah fitur numerik perlu dinormalisasi. Normalisasi penting untuk kernel non-linear.
4	Normalisasi Data (Jika Ya)		Melakukan normalisasi menggunakan teknik seperti Min-Max Scaling atau Standardization.
5	Data Siap		Data telah dibersihkan, diproses, dan siap digunakan untuk pelatihan SVM.
6	Pilih Fungsi Kernel		Menentukan jenis kernel SVM yang digunakan berdasarkan karakteristik data.
7	Kernel Linear		Dipilih jika data linear separable atau data teks berskala besar seperti TF-IDF. Kernel ini paling umum dipakai untuk klasifikasi teks.
8	Kernel Polynomial		Kernel non-linear untuk pola non-separable dengan kompleksitas polynomial.
9	Kernel RBF/Gaussian		Kernel non-linear paling populer untuk data dengan pola kompleks. Cocok untuk hubungan non-linear.

10	Kernel Sigmoid		Kernel dengan karakteristik mirip neural network, digunakan jika data sesuai namun jarang dipakai untuk teks.
11	Pelatihan Model SVM		Data dilatih menggunakan kernel terpilih. Proses mencari pemisah terbaik (hyperplane).
12	Formulasi Masalah Optimasi		SVM menyusun masalah matematis untuk memaksimalkan margin sambil meminimalkan error.
13	Hard Margin atau Soft Margin		Menentukan pendekatan SVM berdasarkan kualitas data: hard margin untuk data bersih dan separable, soft margin untuk data nyata yang mengandung noise.
14	Hard Margin – Tujuan		Mencapai pemisahan sempurna antar kelas tanpa kesalahan.
15	Hard Margin – Kendala		Mengharuskan data benar-benar separable, sulit untuk kasus real.
16	Soft Margin – Tujuan		Mencapai margin besar dengan beberapa pelanggaran (menggunakan parameter C).
17	Soft Margin – Kendala		Mengizinkan beberapa data salah klasifikasi demi generalisasi lebih baik.
18	Transformasi ke Bentuk Dual		Persamaan SVM diubah ke bentuk dual agar kernel dapat diterapkan dan proses komputasi lebih efisien.
19	Selesaikan Masalah Dual		Menyelesaikan Quadratic Programming untuk mendapatkan nilai α (Lagrange multipliers).

20	Identifikasi Support Vectors		Menentukan titik-titik data yang paling mempengaruhi pembentukan hyperplane.
21	Hitung Vektor Normal Hyperplane (w)		Menghitung arah pemisah antar kelas berdasarkan support vectors.
22	Hitung Bias (b)		Menghitung posisi konkrit hyperplane dalam ruang fitur.
23	Gunakan Model untuk Prediksi		Model digunakan untuk memprediksi apakah berita baru termasuk hoaks atau fakta.
24	Fungsi Keputusan		Menghitung skor jarak data dari hyperplane untuk menentukan kelas prediksi.
25	Evaluasi Model		Mengukur performa model menggunakan akurasi, precision, recall, F1-score, dan confusion matrix.
26	End		Proses pembangunan model selesai dan siap diterapkan.

2.6 Penelitian Terdahulu

Tabel 2. 3 Tabel Penelitian Terdahulu

No	Peneliti (Tahun)	Judul Penelitian	Hasil Penelitian
1	(Rahmawati, 2022)	Deteksi Berita Hoaks di Website TurnbackHoax dengan Machine Learning	Akurasi SVM 83% untuk klasifikasi hoaks
2	(Wayan Sumartini Saraswati et al., 2023)	Support Vector Machine For Hoax Detection	Akurasi 99,4% pada dataset umum

3	(FEBRIYANTY, 2023)	Deteksi Berita Hoax Dari Media Online Indonesia Menggunakan Algoritma Naive Bayes Dan Support Vector Machine	Akurasi SVM mencapai 100% pada dataset berita online Indonesia
4	(Fernandes & Tahara Shita, 2024)	Penerapan Metode SVM dan Random Forest untuk Mendeteksi Berita Hoaks pada PT. Global Arrow	Meskipun algoritma seperti SVM (Support Vector Machine) dan klasifikasi RandomForest mampu mencapai skor prediksi di atas 90%, masih terdapat sekitar 10% hasil prediksi yang mungkin salah.
5	(Anisya, 2024)	Implementasi Metode Support Vector Machine (Svm) Dalam Klasifikasi Judul Berita Hoaks Politik	Akurasi terbaik 92% dengan kernel RBF pada klasifikasi judul berita hoaks politik
6	(Al-Khowarizmi et al., 2024)	Optimization of support vector machine with cubic kernel function to detect cyberbullying in social networks	Model Support Vector Machine (SVM) dengan linear kernel terbukti lebih unggul dalam mendeteksi cyberbullying pada tweet kebijakan "cipta kerja" dengan mencapai akurasi 92,3%, dibandingkan dengan model SVM yang menggunakan cubic kernel yang menghasilkan akurasi 90%.
7	(Kurnia et al., 2024)	Perbandingan Naïve Bayes dan CNN yang Dioptimasi PSO pada Identifikasi Berita Hoax Politik Indonesia	Penelitian ini membuktikan bahwa Naïve Bayes lebih efektif daripada CNN dalam mendeteksi berita hoaks politik berbahasa Indonesia, sehingga layak dijadikan pilihan utama.
8	(Kamal et al., 2025)	Identifikasi Berita Palsu di Portal Media Online Menggunakan Model IndoBERT dan LSTM	Model ensemble IndoBERT–LSTM berhasil mendeteksi hoaks politik dengan akurasi tinggi (F1 0,9544; AUC 0,9936) dan mengungguli model tunggal. Meski masih ada error 4,59%, performanya dapat ditingkatkan dengan fitur kontekstual dan dataset lebih beragam. Model ini berpotensi

			diterapkan pada platform verifikasi seperti TurnBackHoax.id.
9	(Jocelynnne et al., 2025)	Detection of Political Hoax News Using Fine-Tuning IndoBERT	Penelitian ini menunjukkan bahwa IndoBERT yang di-fine-tune mampu mendeteksi hoaks politik dengan akurasi tinggi (95% dan AUC 0.946). Model ini efektif memahami konteks bahasa Indonesia, namun masih terbatas oleh bias dan ketidakseimbangan data, serta hanya optimal untuk domain berita politik.
10	(Azis et al., 2025)	Rancangan Sistem Deteksi Berita Hoaks dengan IndoBERT Berbasis Dataset Scraping Seimbang (2020–2025)	Penelitian ini berhasil merancang sistem deteksi hoaks berbasis IndoBERT yang di-fine-tune, menggunakan dataset hasil scraping dari CNN Indonesia, Detik.com, Kompas.com, dan TurnBackHoax.id periode 2020–2025.

BAB III

ANALISA DAN PERANCANGAN SISTEM

3.1 Metodologi Penelitian

Jenis penelitian yang digunakan dalam penelitian ini adalah penelitian kuantitatif eksperimental dengan pendekatan komputasional. Penelitian ini bersifat kuantitatif karena data yang digunakan dianalisis secara numerik menggunakan metode klasifikasi dan dievaluasi melalui metrik seperti akurasi, presisi, recall, dan F1-score.

Pendekatan eksperimental digunakan karena peneliti menguji dan membandingkan performa algoritma Support Vector Machine (SVM) dalam mengklasifikasikan berita Hoaks dan Fakta. Proses penelitian dilakukan secara sistematis dengan menggunakan dataset teks yang telah dibersihkan dan melalui tahapan pra-pemrosesan hingga evaluasi model.

3.2 Analisa Sistem

Penyebaran berita hoaks, khususnya pada topik politik, semakin meningkat akibat cepatnya arus informasi di media digital. Masyarakat sering kesulitan membedakan antara berita benar dan palsu karena keterbatasan waktu, pengetahuan, dan akses untuk melakukan verifikasi. Proses pengecekan secara manual juga lambat dan tidak efisien, sehingga hoaks terlanjur menyebar luas sebelum ada klarifikasi resmi.

Jumlah berita yang besar setiap hari membuat verifikasi manual tidak lagi memadai. Karena itu, diperlukan sistem deteksi otomatis yang mampu mengklasifikasikan berita secara cepat dan objektif. Pendekatan berbasis machine learning dapat membantu mengidentifikasi pola teks yang menunjukkan indikasi

hoaks, sehingga mendukung literasi digital dan mengurangi dampak penyebaran informasi palsu

3.3 Analisa Kebutuhan Sistem

Analisis kebutuhan sistem dilakukan untuk memastikan bahwa aplikasi deteksi berita hoaks yang dikembangkan mampu berjalan sesuai tujuan penelitian, yaitu memberikan hasil klasifikasi teks berita secara akurat, cepat, dan mudah dipahami oleh pengguna. Pada tahap ini, kebutuhan sistem dibagi menjadi dua kelompok utama, yaitu kebutuhan fungsional dan kebutuhan non-fungsional.

3.3.1 Kebutuhan Fungsional

Kebutuhan fungsional menggambarkan fungsi inti yang harus disediakan oleh sistem agar proses klasifikasi berita dapat berjalan dengan baik. Sistem deteksi berita hoaks berbasis machine learning yang dikembangkan dalam penelitian ini dirancang untuk menerima input teks dari pengguna, memproses teks, menjalankan model SVM, serta menampilkan hasil klasifikasi. Adapun kebutuhan fungsional sistem adalah sebagai berikut:

a) Sistem dapat memuat dan menyimpan dataset internal sebagai basis model.

Sistem harus memiliki mekanisme untuk menyimpan dataset internal yang telah dipersiapkan sebelumnya sebagai data latih untuk model SVM. Dataset ini digunakan secara konsisten dan tidak dapat diubah oleh pengguna untuk menjaga stabilitas dan akurasi model.

b) Sistem dapat menerima input teks berita dari pengguna.

Sistem menyediakan form atau kolom input yang memungkinkan pengguna memasukkan teks berita secara manual untuk dianalisis.

c) Sistem dapat melakukan preprocessing teks secara otomatis.

Sistem harus mampu menjalankan tahapan pembersihan teks seperti case folding, penghapusan tanda baca, stopwords removal, serta stemming/tokenizing secara otomatis sebelum memasuki tahap ekstraksi fitur.

d) Proses Sistem mampu melakukan ekstraksi fitur menggunakan TF-IDF.

Sistem harus mengekstraksi fitur teks menggunakan metode TF-IDF atau teknik representasi teks lain sesuai model yang digunakan, sehingga teks dapat diubah menjadi bentuk numerik yang dapat diproses oleh algoritma.

e) Menampilkan Sistem dapat menjalankan algoritma Support Vector Machine (SVM) untuk proses klasifikasi.

Sistem harus mampu mengolah teks yang telah diekstraksi untuk menentukan apakah informasi tersebut termasuk kategori hoaks (1) atau fakta (0).

f) Sistem mampu menampilkan hasil klasifikasi secara jelas kepada pengguna.

Output sistem harus menampilkan label klasifikasi (hoaks atau fakta) dan dapat dilengkapi dengan skor tingkat keyakinan (confidence score) apabila diterapkan.

g) Sistem menyediakan antarmuka berbasis web yang mudah digunakan.

Sistem harus memiliki tampilan antarmuka web yang sederhana, responsif, dan mudah dipahami, sehingga pengguna dapat melakukan input teks dan melihat hasil klasifikasi tanpa hambatan.

3.3.2 Kebutuhan Non-Fungsional

Kebutuhan non-fungsional menggambarkan kualitas dan batasan sistem agar dapat berjalan dengan baik.

a) Kemudahan Penggunaan (Usability)

Antarmuka sistem harus sederhana sehingga pengguna hanya perlu memasukkan teks berita dan langsung mendapatkan hasil klasifikasi.

b) Kinerja (Performance)

Proses klasifikasi harus berjalan cepat dan menghasilkan output dalam waktu singkat meskipun dilakukan secara lokal.

c) Ketersediaan Sistem Lokal (Portability)

Sistem dapat dijalankan pada komputer lokal tanpa membutuhkan koneksi internet atau server eksternal.

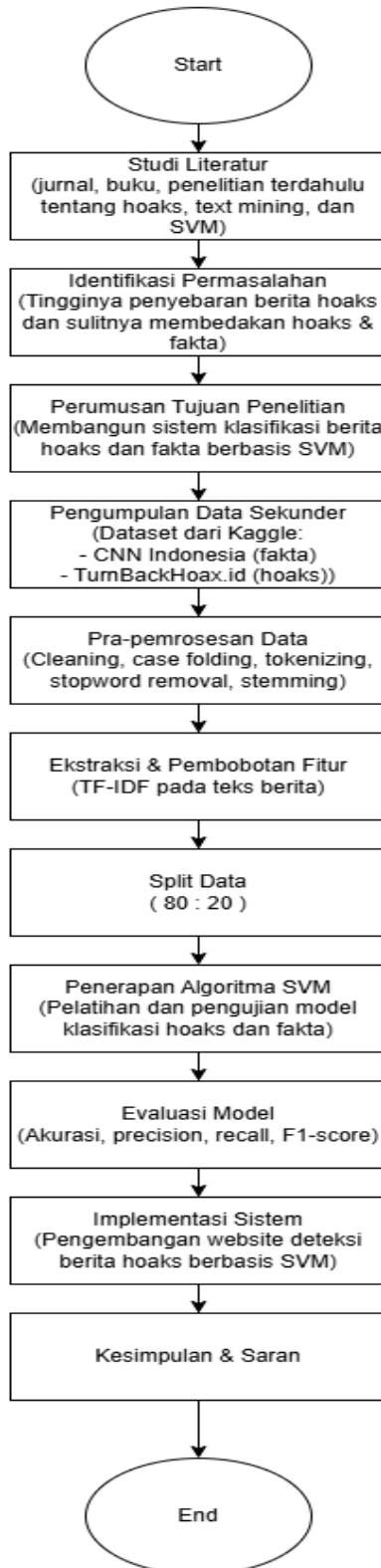
d) Scalability (Kemampuan Menangani Variasi Input)

Meski dibuat untuk kebutuhan penelitian, sistem harus mampu menangani variasi panjang teks dan kompleksitas struktur kalimat tanpa mengalami error atau penurunan performa.

e) Security (Keamanan Sistem)

Sistem harus memastikan keamanan data dengan mencegah modifikasi tidak sah terhadap dataset internal maupun file model. Selain itu, sistem harus memvalidasi seluruh input pengguna untuk menghindari kesalahan format, teks berbahaya, atau potensi eksploitasi yang tidak diinginkan.

3.4 Kerangka Berfikir Penelitian



Gambar 3. 1 Kerangka Berfikir Penelitian

Diagram kerangka pemikiran diatas menggambarkan alur penelitian secara sistematis yang dimulai dari studi literatur untuk memperoleh landasan teori dan pemahaman konsep terkait berita hoaks, text mining, dan algoritma Support Vector Machine (SVM). Berdasarkan studi literatur, dilakukan identifikasi permasalahan, yaitu tingginya penyebaran berita hoaks serta kesulitan masyarakat dalam membedakan berita hoaks dan fakta. Selanjutnya, peneliti merumuskan tujuan penelitian, yakni membangun sistem klasifikasi berita hoaks dan fakta berbasis SVM. Tahap berikutnya adalah pengumpulan data sekunder yang diperoleh dari repository Kaggle, dengan data berita faktual dari CNN Indonesia dan berita hoaks dari TurnBackHoax.id.

Data yang terkumpul kemudian melalui proses pra-pemrosesan, meliputi cleaning, case folding, tokenizing, stopword removal, dan stemming agar data siap dianalisis. Setelah itu dilakukan ekstraksi dan pembobotan fitur menggunakan metode TF-IDF untuk merepresentasikan teks berita dalam bentuk numerik. Selanjutnya, dataset dibagi menjadi data latih dan data uji dengan perbandingan 80:20, di mana 80% data digunakan untuk proses pelatihan model dan 20% sisanya digunakan untuk pengujian. Tahap berikutnya adalah penerapan algoritma Support Vector Machine (SVM) untuk melatih dan menguji model klasifikasi, yang kemudian dievaluasi menggunakan metrik akurasi, precision, recall, dan F1-score. Hasil dari model tersebut diimplementasikan dalam bentuk sistem berbasis website untuk deteksi berita hoaks, dan diakhiri dengan penarikan kesimpulan serta saran berdasarkan hasil penelitian yang diperoleh.

3.5 Preprocessing Data

Pada penelitian ini, preprocessing dilakukan terhadap kolom title, karena judul berita mengandung ringkasan informasi utama yang sering menjadi pemicu persepsi hoaks atau fakta di masyarakat. Judul dipilih karena bersifat ringkas, provokatif, dan representatif terhadap isi berita.

Secara formal :

$$t_i' = f(t_i)$$

dengan:

- t_i = judul asli
- t_i' = judul setelah preprocessing

3.5.1 Case Folding

Case folding mengubah seluruh huruf pada judul menjadi huruf kecil (*lowercase*). Tujuannya untuk menghilangkan perbedaan makna akibat kapitalisasi dan menyeragamkan teks.

Representasi Formal

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

Contoh :

Judul asli	Setelah case folding
PUNYA HUTANG 50 MILIAR, SANDIAGA UNO GUGAT ANIES BASWEDAN	punya hutang 50 miliar sandiaga uno gugat anies baswedan

3.5.2 Penghapusan Angka, Tanda Baca, dan Simbol

Pada tahap ini dilakukan penghapusan:

- Angka (misalnya 50)
- Tanda baca (koma)
- Simbol non-alfabet

Angka dihapus karena tidak selalu relevan secara semantik dalam klasifikasi hoaks berbasis teks judul.

Representasi Formal :

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

Contoh :

Sebelum	Sesudah
punya hutang 50 miliar sandiaga uno gugat anies baswedan	punya hutang miliar sandiaga uno gugat anies baswedan

3.5.3 Stopword Removal

Stopword adalah kata umum yang sering muncul tetapi tidak memiliki bobot informasi tinggi dalam proses klasifikasi.

Representasi Formal :

$$Tokens' = Tokens - Stopwords$$

Contoh Stopword Bahasa Indonesia :

“yang”, “dan”, “di”, “ke”, “dari”

Pada contoh judul ini, tidak banyak stopwords, sehingga sebagian besar token tetap dipertahankan.

Hasil
["punya", "hutang", "miliar", "sandiaga", "uno", "gugat", "anies", "baswedan"]

3.5.4 Stemming / Tokenizing

a) Stemming

Stemming mengubah kata berimbuhan menjadi bentuk dasar.

Contoh

- “memiliki” → “milik”
- “menggugat” → “gugat”

Pada judul ini, kata “**gugat**” sudah berupa kata dasar sehingga tidak berubah.

Representasi Formal :

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

Hasil
["punya", "hutang", "miliar", "sandiaga", "uno", "gugat", "anies", "baswedan"]

b) Tokenizing

Tokenizing memecah judul menjadi kata-kata penyusunnya.

Representasi Formal

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

Hasil
["punya", "hutang", "miliar", "sandiaga", "uno", "gugat", "anies", "baswedan"]

3.5.5 Normalisasi Teks

Normalisasi bertujuan menyamakan variasi penulisan kata agar konsisten di seluruh dataset.

Contoh :

- “hutang” → “utang”
- Penulisan nama tetap dipertahankan sebagai entitas teks

Hasil Akhir Preprocessing
["punya", "utang", "miliar", "sandiaga", "uno", "gugat", "anies", "baswedan"]

3.6 Ekstraksi Fitur Teks

Setelah seluruh judul berita pada kolom title melalui tahap preprocessing, data teks yang dihasilkan masih berupa kumpulan kata dan belum dapat diproses secara langsung oleh algoritma machine learning. Oleh karena itu, diperlukan tahap ekstraksi fitur untuk mengubah data teks menjadi representasi numerik. Pada penelitian ini, metode ekstraksi fitur yang digunakan adalah Term Frequency–Inverse Document Frequency (TF-IDF).

TF-IDF merupakan metode pembobotan kata yang bertujuan untuk mengukur tingkat kepentingan suatu kata dalam sebuah dokumen relatif terhadap seluruh kumpulan dokumen. Metode ini mempertimbangkan dua aspek utama, yaitu frekuensi kemunculan kata dalam satu dokumen dan tingkat penyebaran kata tersebut di seluruh dokumen. Dengan demikian, TF-IDF mampu memberikan bobot tinggi pada kata-kata yang bersifat spesifik dan informatif, serta bobot rendah pada kata-kata yang umum.

3.6.1 Term Frequency (TF)

Term Frequency (TF) digunakan untuk mengukur seberapa sering suatu kata (term) muncul dalam sebuah dokumen. Semakin sering suatu kata muncul

dalam satu judul berita, maka kata tersebut dianggap semakin penting dalam merepresentasikan isi judul tersebut.

Rumus Term Frequency :

$$TF(t, d) = \frac{f_{t,d}}{\sum_k f_{k,d}}$$

dengan keterangan:

- t = kata (term) tertentu,
- d = dokumen atau judul berita,
- $f_{t,d}$ = jumlah kemunculan kata t dalam dokumen d ,
- $\sum_k f_{k,d}$ = total seluruh kata dalam dokumen d .

Rumus ini menghasilkan nilai TF dalam rentang 0 hingga 1, sehingga setiap dokumen memiliki skala bobot yang seimbang meskipun panjang judul berbeda.

Contoh Perhitungan TF :

Misalkan sebuah judul berita setelah preprocessing memiliki **8 kata**, dan kata “**gugat**” muncul sebanyak **1 kali**, maka nilai Term Frequency dihitung sebagai berikut:

$$TF(\text{"gugat"}) = \frac{1}{8}$$

Nilai tersebut menunjukkan bahwa kata “*gugat*” menyusun satu per delapan dari keseluruhan kata dalam judul berita tersebut.

3.6.2 Inverse Document Frequency (IDF)

Inverse Document Frequency (IDF) digunakan untuk mengukur tingkat kepentingan suatu kata dengan memperhatikan seberapa sering kata tersebut muncul di seluruh dokumen. Kata yang muncul di banyak dokumen dianggap

kurang informatif, sedangkan kata yang jarang muncul memiliki nilai IDF yang lebih tinggi.

Rumus Inverse Document Frequency

$$IDF(t) = \log \left(\frac{N}{df_t} \right)$$

dengan keterangan:

- N = jumlah seluruh dokumen atau judul berita,
- df_t = jumlah dokumen yang mengandung kata t .

Penggunaan fungsi logaritma bertujuan untuk menstabilkan nilai IDF agar tidak terlalu besar dan tetap proporsional.

Contoh Perhitungan IDF

Misalkan:

- Jumlah seluruh judul berita $N = 2.000$,
- Kata “**gugat**” muncul pada **40 judul berita**,

maka nilai IDF dihitung sebagai berikut:

$$IDF(\text{"gugat"}) = \log \left(\frac{2000}{40} \right)$$

Nilai IDF yang dihasilkan menunjukkan bahwa kata “*gugat*” tidak terlalu umum dan memiliki tingkat kekhususan tertentu dalam kumpulan judul berita.

3.6.3 TF-IDF

TF-IDF merupakan hasil perkalian antara nilai Term Frequency dan Inverse Document Frequency. Nilai ini menggambarkan tingkat kepentingan akhir suatu kata dalam sebuah dokumen secara keseluruhan.

Rumus TF-IDF :

$$TF-IDF(t, d) = TF(t, d) \times IDF(t)$$

Dengan mengalikan kedua nilai tersebut, TF-IDF mampu menyeimbangkan antara kata yang sering muncul dalam satu dokumen dan kata yang jarang muncul dalam seluruh dokumen.

Interpretasi TF-IDF

Kata-kata seperti:

- “**utang**”
- “**gugat**”
- “**miliar**”

cenderung memiliki nilai TF-IDF yang tinggi karena:

1. Kata tersebut sering muncul dalam judul berita yang mengandung klaim, konflik, atau tuduhan, dan
2. Kata tersebut relatif jarang muncul di judul berita umum atau non-kontroversial.

Dengan demikian, nilai TF-IDF yang tinggi menunjukkan bahwa kata tersebut memiliki kontribusi besar dalam membedakan antara judul berita hoaks dan fakta.

Seluruh nilai TF-IDF ini selanjutnya digunakan sebagai vektor fitur yang menjadi input bagi model klasifikasi Support Vector Machine (SVM).

3.7 Model Support Vector Machine (SVM)

Pada penelitian ini, SVM digunakan untuk mengklasifikasikan judul berita ke dalam dua kelas, yaitu hoaks dan fakta, berdasarkan fitur numerik yang dihasilkan dari proses ekstraksi fitur menggunakan metode TF-IDF.

SVM bekerja dengan mencari sebuah fungsi pemisah (decision boundary) berupa garis atau bidang pemisah yang disebut hyperplane. Hyperplane ini dirancang sedemikian rupa agar dapat memisahkan data dari dua kelas dengan jarak

maksimum (margin maksimum). Pendekatan ini menjadikan SVM sangat efektif untuk menangani data teks yang berdimensi tinggi, seperti vektor TF-IDF.

3.7.1 Label Hoaks

Pada penelitian ini, setiap judul berita diberikan label kelas berdasarkan sumber datanya. Label ini digunakan sebagai nilai target (*output*) dalam proses pelatihan dan pengujian model SVM.

$$y = \begin{cases} 1, & \text{Hoaks} \\ 0, & \text{Fakta} \end{cases}$$

dengan keterangan:

- **1 (Hoaks)** menunjukkan judul berita yang berasal dari sumber TurnBackHoax,
- **0 (Fakta)** menunjukkan judul berita yang berasal dari CNN Indonesia.

Pemberian label ini memungkinkan model SVM untuk mempelajari pola perbedaan antara judul berita hoaks dan fakta berdasarkan karakteristik kata yang terkandung di dalamnya.

3.7.2 Hyperplane SVM

Setelah judul berita diubah menjadi vektor numerik TF-IDF, setiap judul direpresentasikan sebagai sebuah titik dalam ruang berdimensi tinggi. SVM kemudian berusaha mencari sebuah hyperplane yang dapat memisahkan titik-titik tersebut berdasarkan kelasnya.

Secara matematis, hyperplane dapat dinyatakan dengan persamaan:

$$w \cdot x + b = 0$$

dengan keterangan:

- x merupakan vektor fitur TF-IDF dari judul berita,
- w adalah vektor bobot yang menunjukkan tingkat kepentingan setiap fitur,
- b adalah bias atau konstanta pemindah hyperplane.

Hyperplane ini berfungsi sebagai batas keputusan (*decision boundary*) yang membagi ruang fitur menjadi dua bagian, masing-masing merepresentasikan kelas hoaks dan fakta.

3.7.3 Fungsi Klasifikasi

Setelah hyperplane diperoleh, proses klasifikasi dilakukan dengan menghitung nilai fungsi keputusan terhadap setiap data uji. Fungsi klasifikasi pada SVM dirumuskan sebagai berikut:

$$f(x) = \text{sign}(w \cdot x + b)$$

Nilai fungsi tersebut menentukan kelas suatu judul berita, dengan ketentuan:

- $f(x) = +1$ menunjukkan bahwa judul berita diklasifikasikan sebagai **hoaks**,
- $f(x) = -1$ menunjukkan bahwa judul berita diklasifikasikan sebagai **fakta**.

Dalam implementasi penelitian ini, hasil klasifikasi tersebut kemudian disesuaikan dengan label kelas yang digunakan, yaitu:

- $+1 \rightarrow 1$ (Hoaks),
- $-1 \rightarrow 0$ (Fakta).

3.7.4 Optimasi (Soft Margin SVM)

Pada praktiknya, data judul berita hoaks dan fakta tidak selalu dapat dipisahkan secara sempurna oleh satu hyperplane. Oleh karena itu, penelitian ini menggunakan pendekatan Soft Margin SVM, yang mengizinkan adanya beberapa kesalahan klasifikasi dengan tujuan memperoleh model yang lebih fleksibel dan mampu melakukan generalisasi dengan baik.

Fungsi objektif Soft Margin SVM dirumuskan sebagai berikut:

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i$$

dengan batasan:

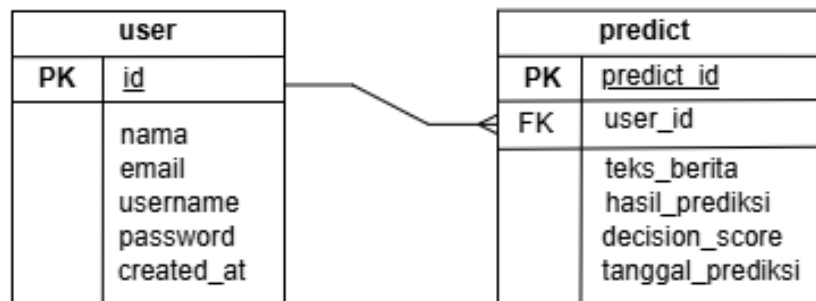
$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i$$

dengan keterangan:

- $\|\mathbf{w}\|^2$ berfungsi untuk memaksimalkan margin,
- ξ_i adalah *slack variable* yang merepresentasikan tingkat kesalahan klasifikasi,
- C adalah parameter regularisasi yang mengatur keseimbangan antara margin maksimum dan jumlah kesalahan,
- y_i adalah label kelas dari data ke- i .

Nilai parameter C yang lebih besar akan memberikan penalti lebih besar terhadap kesalahan klasifikasi, sedangkan nilai C yang lebih kecil akan menghasilkan margin yang lebih lebar namun mengizinkan lebih banyak kesalahan.

3.8 Entity Relationship Diagram



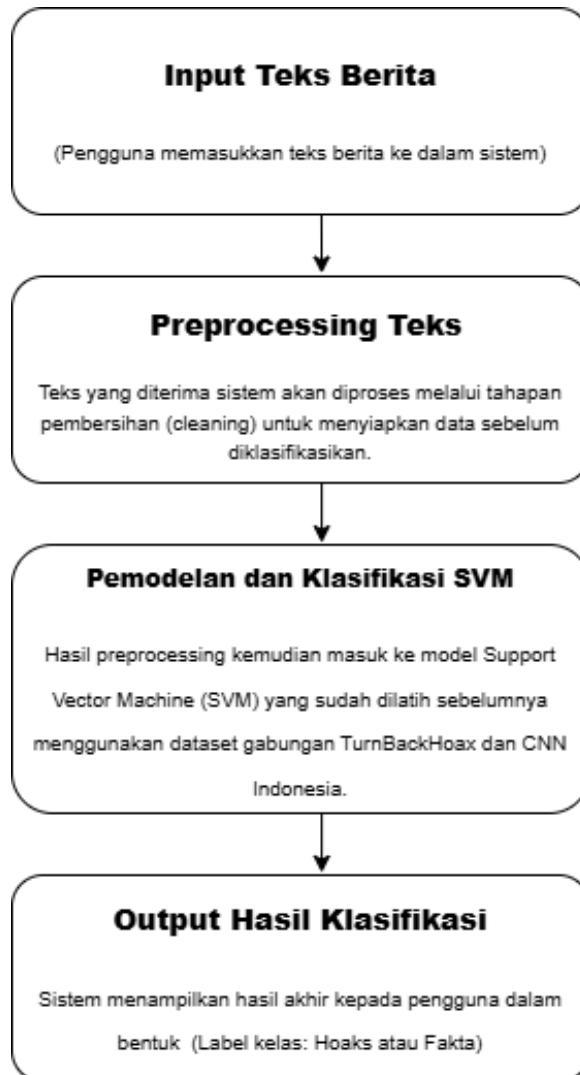
Gambar 3. 2 Entity Relationship Diagram

ERD tersebut menggambarkan struktur basis data pada sistem deteksi berita hoaks berbasis Support Vector Machine (SVM) yang terdiri dari dua entitas utama, yaitu user dan predict. Entitas user berfungsi untuk menyimpan data pengguna sistem, dengan atribut id sebagai primary key yang digunakan untuk mengidentifikasi setiap pengguna secara unik, serta atribut pendukung seperti nama, email, username, password, dan created_at untuk kebutuhan autentikasi dan manajemen akun. Entitas predict digunakan untuk menyimpan data hasil prediksi

berita yang dilakukan oleh pengguna, dengan `predict_id` sebagai primary key dan `user_id` sebagai foreign key yang menghubungkan data prediksi dengan pengguna yang melakukan proses deteksi.

Atribut `teks_berita` menyimpan isi berita yang dianalisis, `hasil_prediksi` menunjukkan hasil klasifikasi model SVM (hoaks atau fakta), `decision_score` merepresentasikan tingkat keyakinan model terhadap hasil prediksi, dan `tanggal_prediksi` mencatat waktu dilakukannya proses prediksi. Relasi antara entitas `user` dan `predict` bersifat one-to-many, di mana satu pengguna dapat melakukan banyak prediksi berita, sedangkan setiap data prediksi hanya terkait dengan satu pengguna. Struktur ERD ini menunjukkan bahwa database difokuskan pada penyimpanan data pengguna dan hasil prediksi, sementara proses pengolahan teks dan klasifikasi dilakukan di sisi aplikasi.

3.9 Analisa Alur Data Sistem



Gambar 3. 3 Alur Data Sistem

Gambar 3.3 menggambarkan aliran data dalam sistem deteksi berita hoaks berbasis SVM. Proses dimulai ketika pengguna memasukkan teks berita ke dalam sistem. Data teks tersebut kemudian melalui tahap *preprocessing*, yaitu pembersihan dan penyiapan teks seperti menghapus karakter khusus, melakukan tokenisasi, dan menghilangkan stopwords. Setelah data bersih, teks tersebut diubah menjadi representasi fitur dan dimasukkan ke model Support Vector Machine (SVM) yang telah dilatih menggunakan dataset gabungan TurnBackHoax dan CNN

Indonesia. Model kemudian melakukan proses klasifikasi untuk menentukan apakah teks tersebut termasuk hoaks atau fakta. Hasil akhir berupa label “Hoaks” atau “Fakta” ditampilkan kembali kepada pengguna sebagai output sistem.

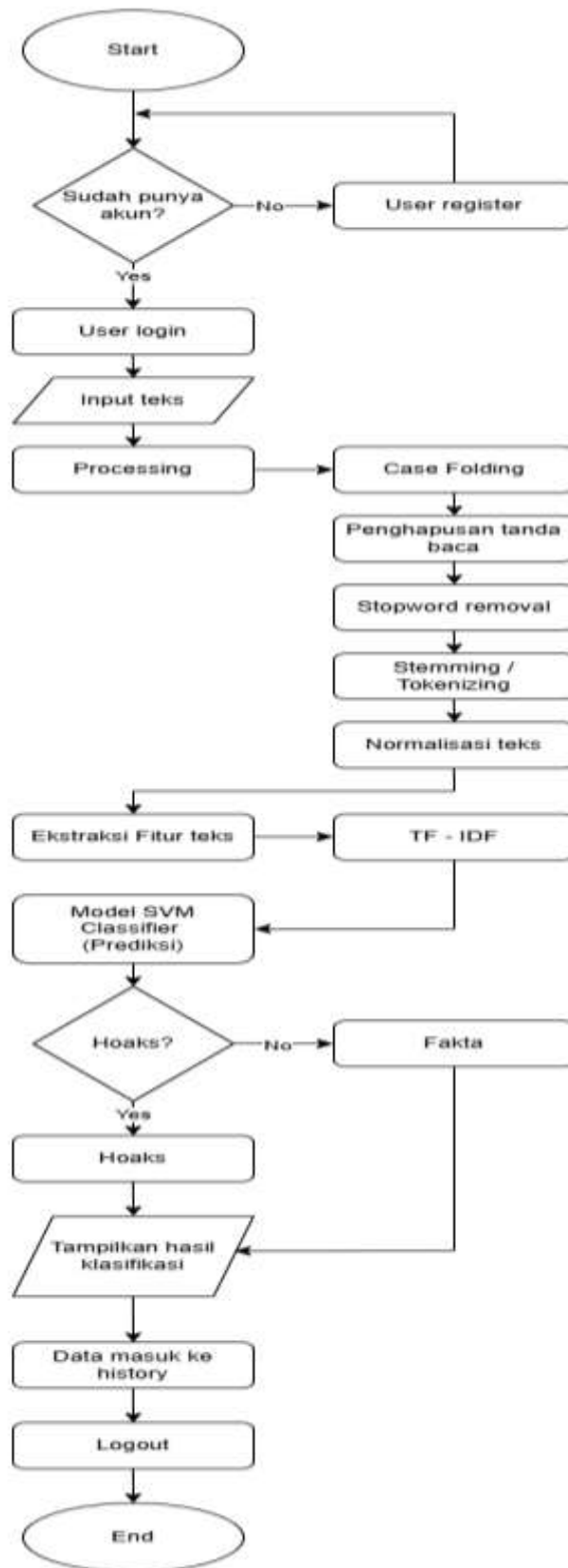
3.10 Perancangan Arsitektur Sistem

Karena sistem hanya berjalan secara lokal, arsitektur dibuat sederhana agar mudah diimplementasikan dan digunakan. Sistem ini dibangun dengan pendekatan website lokal berbasis Python Flask dan model machine learning untuk klasifikasi teks berita.

Komponen Utama :

1. Frontend Website (HTML/CSS/PHP)
 - a. Form untuk input teks berita.
 - b. Tombol “Klasifikasi” untuk memproses teks.
 - c. Area untuk menampilkan hasil klasifikasi kepada pengguna.
2. Backend (Python Flask)
 - a. File preprocessing untuk membersihkan dan menyiapkan teks.
 - b. File TF-IDF vectorizer (.pkl) untuk mengubah teks menjadi representasi numerik.
 - c. File model SVM (.pkl) untuk melakukan klasifikasi.
 - d. Endpoint Flask yang menerima input teks dan mengembalikan hasil klasifikasi.
3. Model Machine Learning
 - a. Dibuat menggunakan Python di Jupyter Notebook.
 - b. Model disimpan dalam format .pkl agar dapat di-load oleh Flask saat sistem dijalankan.

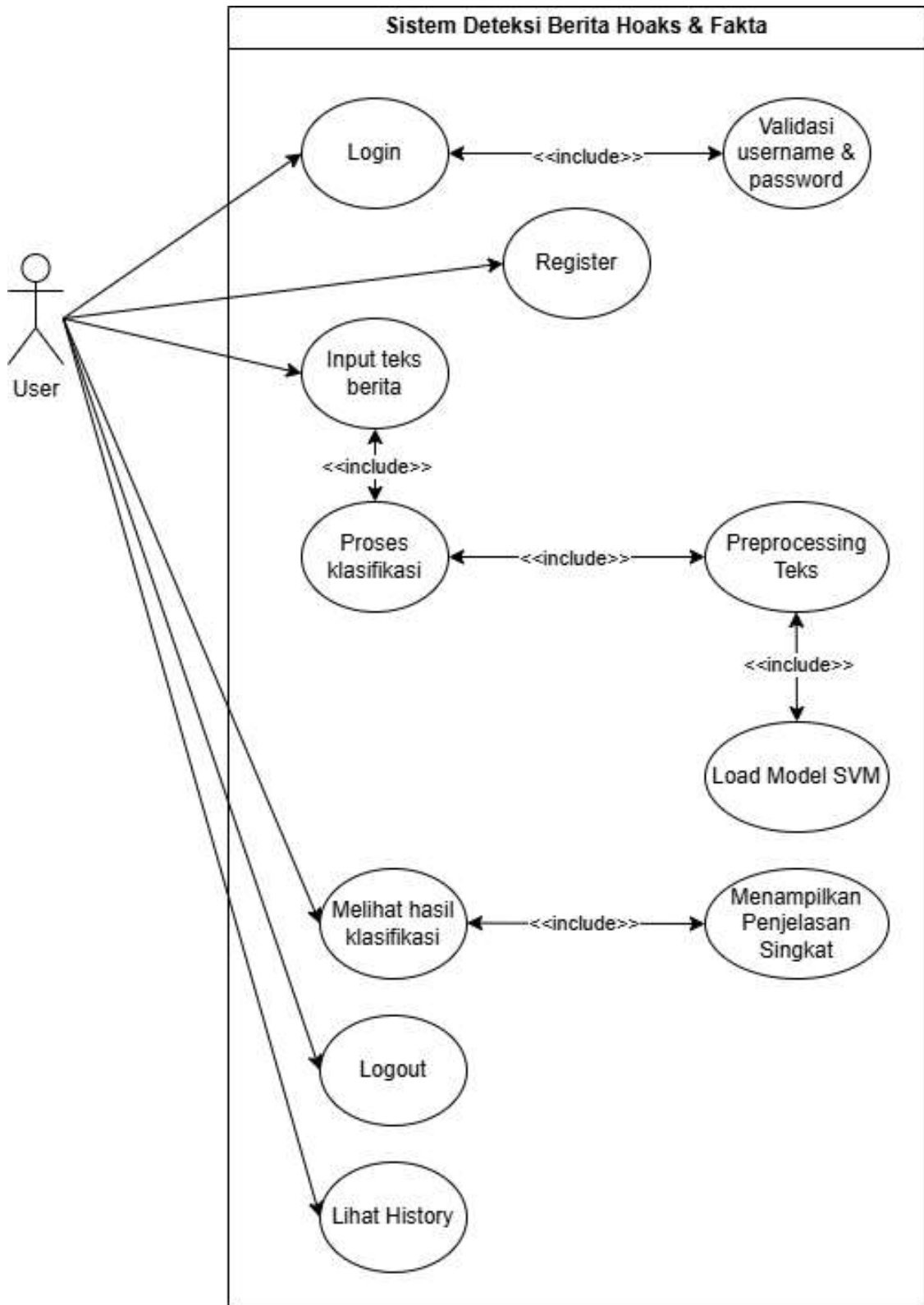
3.11 Rancangan Flowchart Sistem



Gambar 3. 4 Flowchart Sistem

Gambar 3.4 menggambarkan alur kerja sistem klasifikasi teks berbasis SVM. Flowchart sistem dimulai ketika pengguna membuka website lalu melalui cek kepemilikan akun. Jika belum memiliki akun, pengguna harus melakukan registrasi, sedangkan jika sudah memiliki akun dapat langsung login. Setelah berhasil masuk, pengguna memasukkan teks yang ingin diperiksa. Teks tersebut kemudian diproses melalui tahap *preprocessing* meliputi *case folding*, penghapusan tanda baca, *stopword removal*, *tokenizing/stemming*, dan normalisasi teks. Hasil *preprocessing* diekstraksi menggunakan TF-IDF untuk menghasilkan fitur numerik yang kemudian dikirim ke model SVM guna menentukan apakah teks termasuk hoaks atau fakta. Hasil klasifikasi ditampilkan kepada pengguna dan disimpan ke dalam riwayat pengecekan. Proses diakhiri dengan pengguna melakukan logout dari sistem.

3.12 Use Case Diagram



Gambar 3. 5 Use Case Diagram Sistem

Sistem Deteksi Berita Hoaks & Fakta ini dirancang untuk membantu pengguna dalam melakukan verifikasi informasi secara cepat dan akurat. Pada tahap awal, sistem menyediakan mekanisme Login, di mana pengguna harus memasukkan *username* dan *password* yang kemudian diverifikasi oleh sistem untuk memastikan keamanan dan otentikasi. Bagi pengguna baru yang belum memiliki akun, tersedia fitur Register yang memungkinkan pembuatan akun dengan mengisi data yang diperlukan. Setelah selesai menggunakan sistem, pengguna dapat keluar melalui fitur Logout untuk menjaga kerahasiaan data dan riwayat penggunaan.

Fungsi utama dari sistem ini adalah Input teks berita, yaitu tahap ketika pengguna memasukkan potongan berita atau informasi yang ingin diperiksa kebenarannya. Setelah teks dimasukkan, sistem akan secara otomatis menjalankan rangkaian proses klasifikasi. Proses ini dimulai dari Preprocessing Teks, seperti pembersihan teks, normalisasi, dan transformasi data agar siap dianalisis oleh model. Setelah preprocessing selesai, sistem akan menjalankan Load Model SVM untuk melakukan analisis dan menentukan apakah teks tersebut termasuk kategori Hoaks atau Fakta.

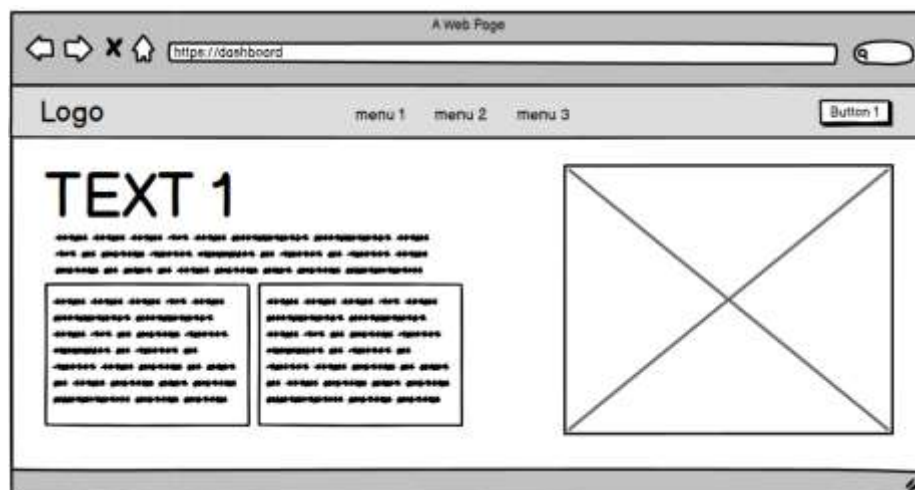
Setelah proses klasifikasi berhasil dilakukan, sistem akan menampilkan hasil kepada pengguna melalui fitur Melihat Hasil Klasifikasi, yang tidak hanya memberikan label akhir namun juga menyertakan Penjelasan Singkat sebagai detail pendukung agar pengguna lebih memahami alasan dari hasil tersebut. Selain itu, sistem juga menyediakan fitur Lihat History yang menyimpan seluruh teks berita yang pernah diperiksa beserta hasil klasifikasinya. Fitur ini membantu pengguna meninjau kembali riwayat pengecekan informasi kapan pun diperlukan.

3.13 Perancangan Antarmuka Pengguna (User Interface Design)

3.13.1 Dashboard

Gambar 3.6 Pada bagian paling atas terdapat header browser dengan address bar bertuliskan “https://dashboard”, yang menunjukkan ini adalah aplikasi berbasis web. Di bawahnya terdapat navbar yang berisi Logo di sisi kiri, menu navigasi (menu 1, menu 2, menu 3) di bagian tengah, serta sebuah tombol aksi (Button 1) di sisi kanan. Struktur ini menunjukkan sistem memiliki beberapa halaman yang dapat diakses melalui menu utama.

Halaman ini memperlihatkan halaman beranda (landing/dashboard awal). Di sisi kiri terdapat judul besar “TEXT 1” dengan deskripsi singkat di bawahnya, serta dua kotak informasi yang kemungkinan berisi ringkasan fitur atau penjelasan sistem. Di sisi kanan terdapat area besar berbentuk persegi dengan tanda silang (placeholder gambar/ilustrasi), yang biasanya digunakan untuk menampilkan ilustrasi, grafik, atau visual utama.

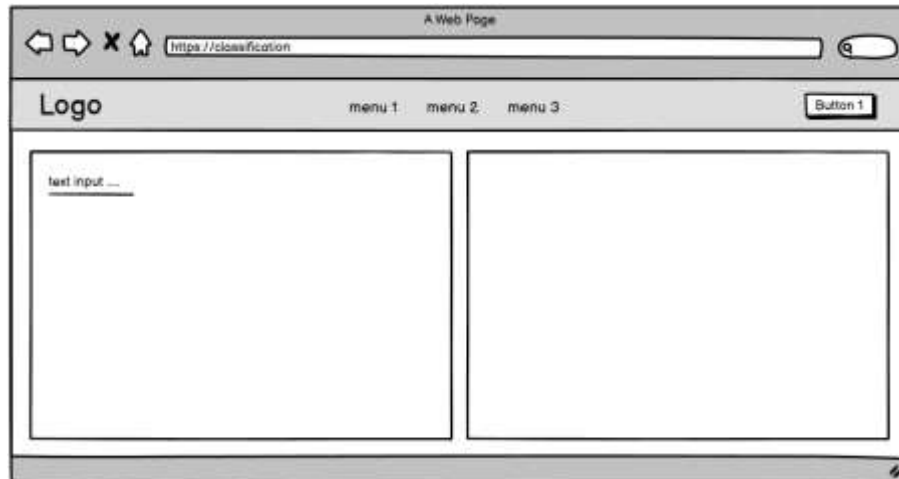


Gambar 3. 6 Dashboard

3.13.2 Halaman Klasifikasi

Gambar 3.7 menampilkan halaman dengan dua panel utama berdampingan. Panel kiri berisi area “text input”, yang menunjukkan adanya fitur input data atau form

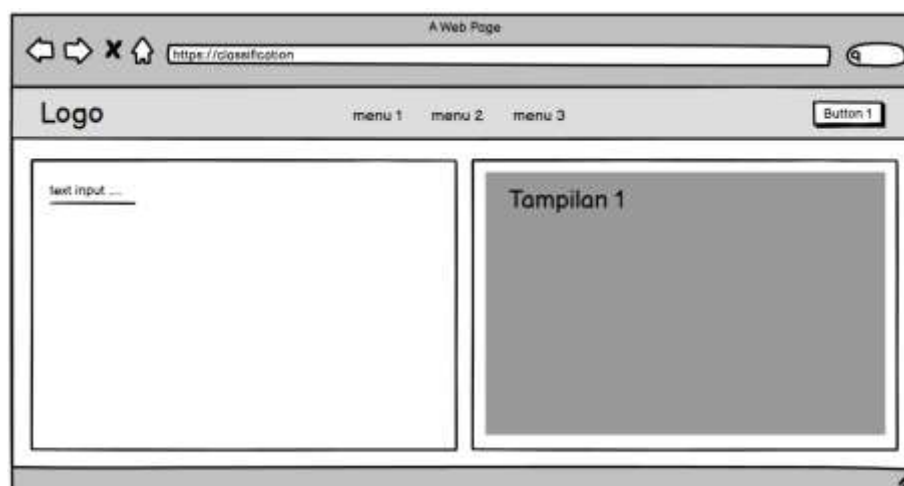
(misalnya untuk memasukkan teks berita). Panel kanan masih kosong, yang kemungkinan akan digunakan untuk menampilkan hasil proses atau output dari sistem.



Gambar 3. 7 Klasifikasi

3.13.3 Hasil Klasifikasi

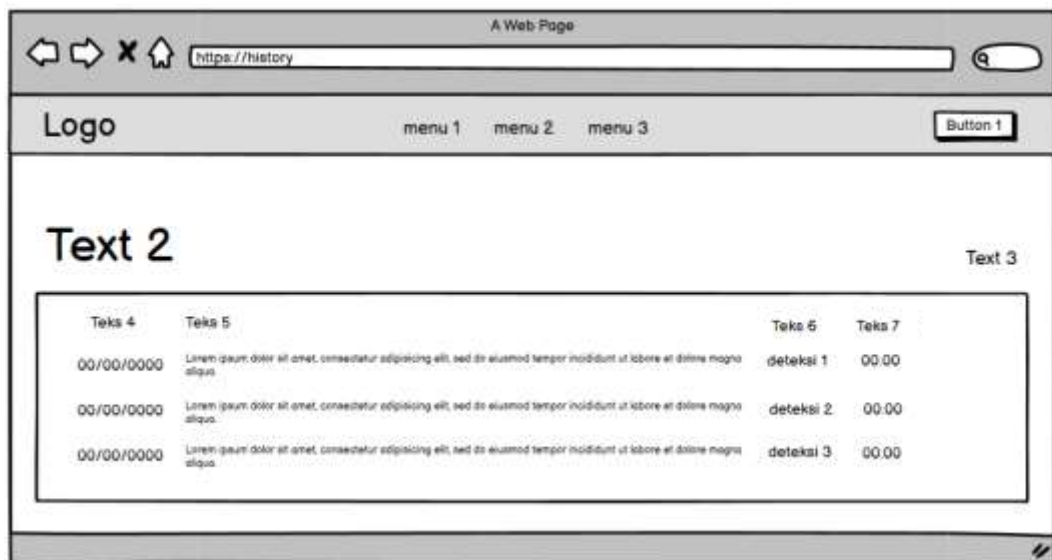
Gambar 3.8 merupakan pengembangan dari halaman sebelumnya. Panel kiri tetap berfungsi sebagai area input, sedangkan panel kanan kini menampilkan “Tampilan 1”, yang menggambarkan area hasil atau visualisasi output setelah sistem memproses data. Ini menunjukkan adanya alur kerja: input → proses → hasil.



Gambar 3. 8 Hasil Klasifikasi

3.13.4 History

Gambar 3.9 menampilkan halaman riwayat atau data log. Di bagian atas terdapat judul “Text 2” dan “Text 3” sebagai header halaman dan kemungkinan filter atau informasi tambahan. Di bawahnya terdapat tabel dengan beberapa kolom seperti tanggal (00/00/0000), deskripsi teks (Teks 5), hasil deteksi (deteksi 1, 2, 3), dan skor (00.00). Halaman ini berfungsi untuk menampilkan daftar data yang telah diproses sebelumnya dalam bentuk tabel terstruktur.



Gambar 3. 9 History

BAB IV

IMPLEMENTASI DAN PENGUJIAN SISTEM

4.1 Kebutuhan Sistem

Kebutuhan sistem merupakan spesifikasi perangkat keras dan perangkat lunak yang digunakan dalam proses pengembangan dan implementasi sistem klasifikasi berita hoaks dan fakta. Kebutuhan ini bertujuan untuk memastikan sistem dapat berjalan dengan optimal sesuai dengan perancangan yang telah dilakukan pada bab sebelumnya.

4.1.1 Kebutuhan Perangkat Keras (Hardware)

Perangkat keras yang digunakan dalam penelitian dan pengembangan sistem ini memiliki spesifikasi sebagai berikut:

- a) Processor : Intel Core i7-12650H
- b) RAM : 16 GB DDR5
- c) Storage : SSD 1 TB NVMe
- d) Sistem Operasi : Windows 11 64-bit
- e) Perangkat pendukung : Mouse dan keyboard

Spesifikasi tersebut sudah memadai untuk melakukan proses pengolahan data teks, preprocessing, pembobotan TF-IDF, training dan pengujian model Support Vector Machine (SVM), serta menjalankan sistem berbasis web dengan stabil dan optimal.

4.1.2 Kebutuhan Perangkat Lunak (Software)

Perangkat lunak yang digunakan dalam pengembangan sistem adalah sebagai berikut:

- a) Bahasa Pemrograman :

- Python (backend machine learning dan API Flask)
 - PHP (frontend dan manajemen session/login)
- b) Framework Backend :
- Flask (sebagai REST API untuk proses klasifikasi)
- c) Library Machine Learning :
- Scikit-learn (implementasi algoritma Support Vector Machine)
- d) Library Pengolahan Data :
- Pandas
 - NumPy
- e) Library Text Processing :
- Sastrawi (stemming Bahasa Indonesia)
- f) Metode Pembobotan Fitur :
- TF-IDF (Term Frequency – Inverse Document Frequency)
- g) IDE / Code Editor :
- Visual Studio Code
- h) Web Server :
- XAMPP (untuk menjalankan Apache dan MySQL sebagai server PHP)
 - Flask Development Server (untuk menjalankan backend machine learning)
- i) Database :
- MySQL
- j) Web Browser :
- Google Chrome

4.2 Implementasi Sistem

4.2.1 Implementasi Dataset

Dataset yang digunakan dalam penelitian ini adalah dataset Indonesian Fact and Hoax Political News yang diperoleh dari Kaggle dan bersumber dari organisasi pemeriksa fakta TurnBackHoax dan CNN Indonesia.

Dataset terdiri dari dua kategori utama, yaitu:

- a) Hoaks (label = 1)
- b) Fakta (label = 0)

Data yang digunakan berupa judul atau teks berita yang telah diklasifikasikan sebelumnya. Dataset kemudian dibagi menjadi dua bagian, yaitu:

- a) Data latih (80%)
- b) Data uji (20%)

Pembagian data ini bertujuan untuk melatih model SVM serta menguji performa model dalam mengklasifikasikan data baru.

4.2.2 Karakteristik Kata pada Berita Hoaks dan Fakta Menggunakan WordCloud

a) WordCloud Hoaks (1)

Wordcloud untuk hoax (Label 1) menunjukkan kata-kata yang sering muncul dalam berita atau konten yang dikategorikan sebagai hoax. Kata yang paling dominan adalah "video", diikuti oleh "foto", "hoax", dan "jokowi", menunjukkan bahwa konten hoax sering berupa media visual seperti video atau gambar yang dikaitkan dengan isu politik atau figur publik seperti Presiden Jokowi.

Kata lain seperti "indonesia", "klarifikasi", dan "covid" mengindikasikan bahwa hoax juga banyak terkait isu nasional dan pandemi,

sementara kata "benar" dan "akun" mungkin muncul karena narasi yang menyesatkan tentang kebenaran informasi atau penyebar informasi di media sosial. Hal ini menunjukkan bahwa hoax cenderung menekankan media yang bisa memicu emosi dan kontroversi serta membutuhkan klarifikasi resmi.



Gambar 4. 1 Wordcloud Hoax

b) WordCloud Fakta (0)

wordcloud untuk fakta (Label 0) didominasi oleh kata-kata seperti "jokowi", "soal", "dpr", dan "tak", yang menunjukkan konten fakta lebih fokus pada isu politik formal, pembahasan kebijakan, dan laporan berita yang jelas. Kata-kata lain seperti "pdip", "anies", "capres", dan "milu" mengacu pada nama partai dan calon politik, menandakan bahwa fakta cenderung memuat informasi objektif dan terkait proses demokrasi. Penggunaan kata "jadi" dan "minta" menunjukkan penyampaian informasi dalam bentuk laporan atau pernyataan resmi.

menjadi "alir") guna mengurangi dimensi fitur. Hasil akhir dari rangkaian proses ini adalah sekumpulan kata dasar bersih yang merepresentasikan konten esensial dari data uji maupun data latih. Berikut tahapannya :

a) Case Folding

```
Case Folding : [salah] anies baswedan dekat dengan aliran krsiten sesat alpha omega di papua
```

b) Remove Salah

```
Remove 'salah' : anies baswedan dekat dengan aliran krsiten sesat alpha omega di papua
```

c) Remove angka

```
Remove Angka : anies baswedan dekat dengan aliran krsiten sesat alpha omega di papua
```

d) Remove punctuation

```
Remove Punct : anies baswedan dekat dengan aliran krsiten sesat alpha omega di papua
```

e) After tokenizing

```
Tokenizing : ['anies', 'baswedan', 'dekat', 'dengan', 'aliran', 'krsiten', 'sesat', 'alpha', 'omega', 'di', 'papua']
```

f) After stopword removal

```
Stopword Removed: ['anies', 'baswedan', 'dekat', 'aliran', 'krsiten', 'sesat', 'alpha', 'omega', 'papua']
```

g) After stemming

```
Stemming : anies baswedan dekat alir krsiten sesat alpha omega papua
```

h) Final result

```
HASIL AKHIR : anies baswedan dekat alir krsiten sesat alpha omega papua
```

i) Reprocessing ke seluruh data

```
Reproses seluruh data... | 20011/20011 [12:19:00:00, 27.04it/s]
```

4.2.4 Implementasi TF-IDF

Setelah data bersih, dilakukan tahap Ekstraksi Fitur menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). Metode ini memberikan bobot numerik pada setiap kata berdasarkan frekuensi kemunculannya dalam sebuah dokumen (*Term Frequency*) dibandingkan dengan sebarannya di seluruh dataset (*Inverse Document Frequency*). Berdasarkan hasil pengolahan, diperoleh bobot fitur yang merepresentasikan tingkat kepentingan setiap kata dalam korpus, seperti kata "jokowi", "anies", dan "video" yang memiliki nilai TF-IDF global tertinggi. Transformasi ini menghasilkan matriks fitur yang merepresentasikan karakteristik unik dari setiap berita, yang selanjutnya akan digunakan sebagai input utama bagi model SVM untuk menentukan batas pemisah (*hyperplane*) yang optimal dalam mengklasifikasikan 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: 231.735656
kata: foto | Total TF-IDF: 224.765186
```

Gambar 4. 3 Implementasi TF-IDF

4.2.5 Analisis Keseimbangan Distribusi Label Dataset

Distribusi label dalam suatu dataset klasifikasi merupakan aspek yang sangat penting karena berpengaruh langsung terhadap performa dan objektivitas model machine learning yang dibangun. Berdasarkan hasil analisis distribusi label pada dataset penelitian ini, diperoleh bahwa kelas hoax (label 1) berjumlah 10.381 data atau sebesar 51,88%, sedangkan kelas non-hoax (label 0) berjumlah 9.630 data atau sebesar 48,12%.

Selisih proporsi antara kedua kelas tersebut hanya sebesar 3,75%, yang menunjukkan bahwa distribusi data berada dalam kondisi yang relatif seimbang (balanced). Secara umum, suatu dataset dikategorikan sebagai balanced apabila distribusi antar kelas tidak menunjukkan perbedaan proporsi yang signifikan atau ekstrem. Dalam konteks ini, komposisi yang mendekati rasio 50:50 menandakan bahwa masing-masing kelas memiliki representasi yang hampir setara dalam proses pelatihan model.

Keseimbangan dataset memiliki peran krusial dalam sistem klasifikasi, khususnya pada penelitian deteksi berita hoaks. Dataset yang tidak seimbang (imbalanced dataset) berpotensi menyebabkan model cenderung memprediksi kelas mayoritas secara dominan, sehingga menghasilkan bias klasifikasi. Kondisi tersebut dapat menurunkan kemampuan model dalam mengenali kelas minoritas, yang pada akhirnya berdampak pada rendahnya nilai recall atau F1-score pada kelas tertentu. Oleh karena itu, penggunaan dataset yang seimbang membantu memastikan bahwa model memiliki kesempatan yang setara untuk mempelajari karakteristik masing-masing kelas.

Selain itu, dengan kondisi dataset yang balanced, penelitian ini tidak memerlukan teknik penyeimbangan tambahan seperti oversampling (misalnya SMOTE) atau undersampling. Hal ini memperkuat validitas eksperimen karena hasil evaluasi model benar-benar mencerminkan kemampuan algoritma dalam mempelajari pola data secara alami tanpa intervensi manipulasi distribusi kelas.

Dapat disimpulkan bahwa dataset yang digunakan dalam penelitian ini telah memenuhi kriteria keseimbangan distribusi label, sehingga mendukung proses pelatihan dan evaluasi model klasifikasi secara adil, objektif, dan representatif terhadap kedua kategori, yaitu berita hoaks dan fakta.

```
===== DISTRIBUSI LABEL =====  
hoax  
1      10381  
0       9630  
Name: count, dtype: int64  
  
===== PERSENTASE (%) =====  
hoax  
1      51.876468  
0      48.123532
```

Gambar 4. 4 Distribusi Label

4.2.6 Split Data

Selanjutnya, dataset dibagi menjadi dua bagian utama, yaitu Data Train dan Data Test, dengan rasio pembagian sebesar 80:20. Pengalokasian ini menghasilkan 16.008 data latih yang digunakan oleh algoritma SVM untuk mempelajari pola klasifikasi hoaks, serta 4.002 data uji yang berfungsi untuk mengevaluasi performa dan generalisasi model terhadap data baru yang belum pernah dilihat sebelumnya. Pembagian ini dilakukan guna memastikan bahwa model yang dibangun memiliki tingkat akurasi yang objektif dan terhindar dari masalah *overfitting*.

```
✓ Data berhasil dimuat!  
Jumlah Data : 20010  
Jumlah Fitur : 5000  
  
🔄 Membagi data train & test...  
Jumlah Data Train : 16008  
Jumlah Data Test : 4002
```

Gambar 4. 5 Implementasi Split Data

4.2.7 Implementasi Model SVM

Hasil evaluasi menunjukkan bahwa algoritma Support Vector Machine (SVM) mampu memberikan performa yang sangat baik dalam mengklasifikasikan berita hoaks. Berdasarkan hasil pengujian pada data uji sebanyak 4.002 data, diperoleh nilai **akurasi sebesar 96,43%**.

Hasil tersebut diperoleh dari Confusion Matrix berikut:

$$\begin{bmatrix} 1865 & 61 \\ 82 & 1994 \end{bmatrix}$$

Keterangan:

1. True Negative (TN) = 1865
2. False Positive (FP) = 61
3. False Negative (FN) = 82
4. True Positive (TP) = 1994

A. Perhitungan Evaluasi Model

1. Akurasi (Accuracy)

Akurasi digunakan untuk mengukur proporsi prediksi yang benar terhadap seluruh data uji.

$$= \frac{1994 + 1865}{4002}$$

$$= \frac{3859}{4002} = 0,9643 = 96,43\%$$

Nilai ini menunjukkan bahwa model berhasil mengklasifikasikan 96,43% data secara benar dari total keseluruhan data uji.

2. Presisi (Precision)

Presisi mengukur tingkat ketepatan model dalam memprediksi suatu kelas.

Untuk kelas hoaks (kelas 1):

$$= \frac{1994}{1994 + 61}$$

$$= \frac{1994}{2055} = 0,97$$

Untuk kelas berita asli (kelas 0):

$$= \frac{1865}{1865 + 82}$$

$$= \frac{1865}{1947} = 0,96$$

Nilai presisi yang tinggi menunjukkan bahwa model jarang melakukan kesalahan tipe I (False Positive).

3. Recall

Recall mengukur kemampuan model dalam mendeteksi seluruh data yang benar-benar termasuk dalam suatu kelas.

Untuk kelas hoaks:

$$= \frac{1994}{1994 + 82}$$

$$= \frac{1994}{2076} = 0,96$$

Untuk kelas berita asli:

$$\begin{aligned} &= \frac{1865}{1865 + 61} \\ &= \frac{1865}{1926} = 0,97 \end{aligned}$$

Nilai recall yang tinggi menunjukkan bahwa model mampu meminimalisir kesalahan tipe II (False Negative).

4. F1-Score

F1-Score merupakan rata-rata harmonik antara precision dan recall, yang dirumuskan sebagai:

$$[F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}]$$

Berdasarkan perhitungan, diperoleh nilai F1-Score untuk kedua kelas berada pada kisaran 0,96 – 0,97, dengan rata-rata sebesar 0,96. Hal ini menunjukkan bahwa model memiliki keseimbangan yang optimal antara ketepatan (precision) dan sensitivitas (recall).

B. Interpretasi Hasil

1. Ketepatan Prediksi.

Model berhasil memprediksi 1.865 data kelas 0 (Berita Asli) dan 1.994 data kelas 1 (Berita Hoaks) secara benar, dengan total kesalahan klasifikasi hanya 143 data dari 4.002 data uji (3,57%).

2. Presisi dan Recall.

Nilai precision dan recall yang seimbang (0,96–0,97) menunjukkan bahwa model tidak hanya memiliki tingkat akurasi yang tinggi, tetapi juga stabil

dalam meminimalisir kesalahan tipe I (False Positive) maupun tipe II (False Negative).

3. F1-Score.

Dengan nilai F1-score rata-rata sebesar 0,96, dapat disimpulkan bahwa model memiliki keseimbangan performa yang sangat baik dalam mendeteksi berita hoaks secara konsisten, sehingga layak digunakan sebagai sistem pendukung dalam identifikasi penyebaran informasi palsu.

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

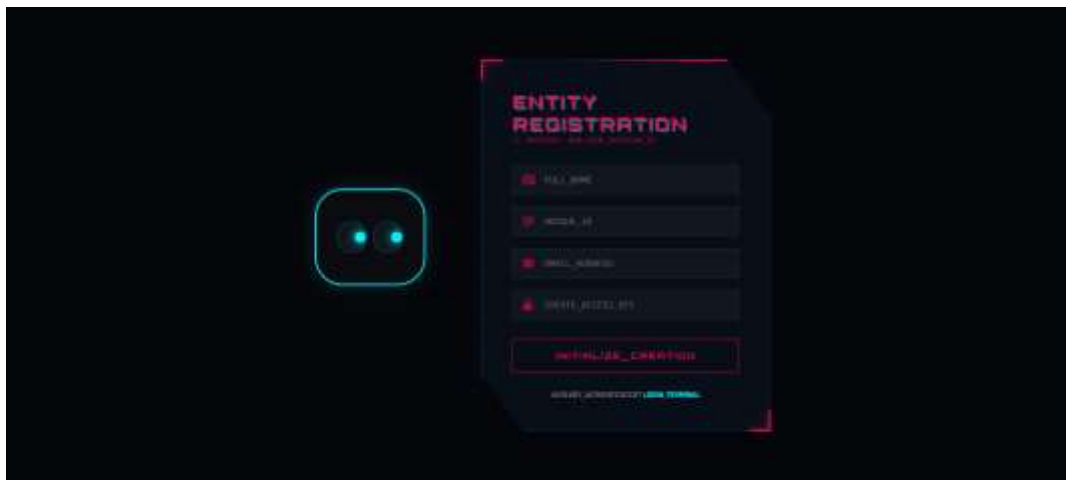
Gambar 4. 6 Confusion Matrix

4.2.8 Implementasi Antarmuka Sistem

a) Form Register

Form ini merupakan halaman registrasi akun (Entity Registration) dengan tampilan futuristik bernuansa gelap dan aksen warna neon merah muda. Di bagian atas terdapat judul “Entity Registration” yang menegaskan bahwa pengguna harus membuat identitas atau akun baru sebelum mengakses sistem.

Form ini menyediakan beberapa kolom input, yaitu Name, Username, Email Address, dan Create Access Key (kata sandi). Setiap kolom dilengkapi ikon untuk memperjelas fungsinya. Setelah seluruh data diisi, pengguna dapat menekan tombol “REGISTER” untuk menyelesaikan proses pendaftaran. Di bagian bawah terdapat opsi Login bagi pengguna yang sudah memiliki akun, sehingga mereka dapat langsung masuk tanpa perlu mendaftar ulang.



Gambar 4. 7 Form Register

b) Form Login

Form ini menampilkan halaman login (Access Terminal) dengan desain serupa namun menggunakan aksesoris warna biru neon. Judul “Access Terminal” menunjukkan bahwa halaman ini berfungsi sebagai gerbang autentikasi pengguna. Form ini memiliki dua kolom utama, yaitu Username/Email dan Password, yang digunakan untuk memverifikasi identitas pengguna yang telah terdaftar. Setelah memasukkan data yang benar, pengguna menekan tombol “LOGIN” untuk memulai sesi atau masuk ke dalam sistem. Jika belum memiliki akun, tersedia tautan Register di bagian bawah untuk mengarahkan pengguna ke halaman pendaftaran. Secara keseluruhan, kedua form ini dirancang sebagai

sistem autentikasi yang terdiri dari proses registrasi dan login dengan tampilan modern dan terstruktur.



Gambar 4. 8 Form Login

c) Dashboard (Home)

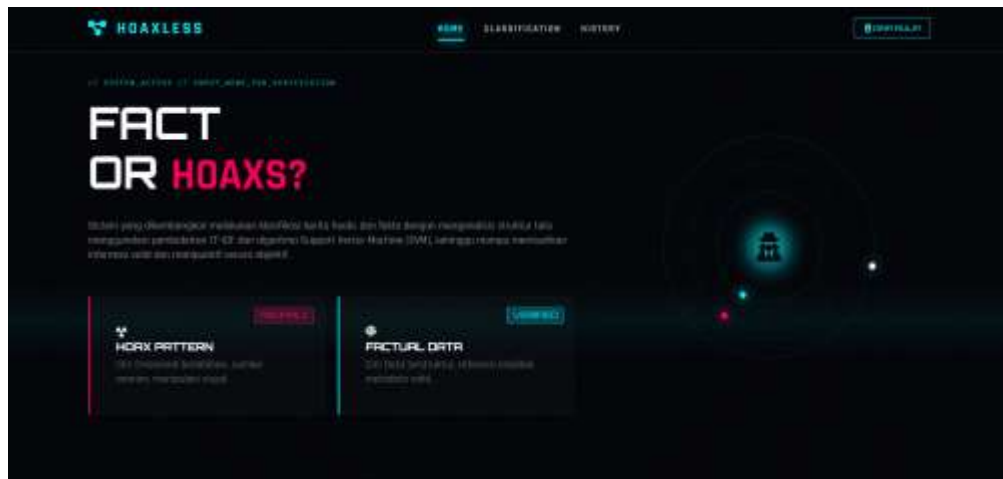
Tampilan pada gambar tersebut merupakan halaman beranda (Home) dari sistem bernama Hoaxless, yaitu sebuah platform klasifikasi berita yang berfokus pada pendeteksian fakta atau hoaks. Di bagian atas terdapat navbar dengan menu Home, Classification, dan History, serta tombol profil pengguna di sisi kanan, yang menunjukkan bahwa sistem ini memiliki fitur autentikasi dan penyimpanan riwayat analisis.

Pada bagian utama halaman, terdapat judul besar “FACT OR HOAX?” yang menjadi fokus utama antarmuka. Di bawahnya terdapat deskripsi singkat yang menjelaskan bahwa sistem ini melakukan klasifikasi berita menggunakan metode Support Vector Machine (SVM) dengan pendekatan TF-IDF, sehingga mampu menganalisis struktur teks dan membedakan informasi valid dari informasi manipulatif secara objektif.

Di bagian bawah terdapat dua panel informasi. Panel pertama bertema merah dengan label Anomaly dan judul Hoax Pattern, yang menjelaskan ciri-

ciri berita hoaks seperti emosional berlebihan, sumber anonim, dan manipulasi visual. Panel kedua bertema biru dengan label Verified dan judul Factual Data, yang menjelaskan karakteristik berita fakta seperti data terstruktur, referensi kredibel, dan metodologi yang jelas.

Secara keseluruhan, halaman ini dirancang dengan gaya futuristik bernuansa gelap dan grid digital, memberikan kesan sistem analisis berbasis teknologi canggih. Tampilan ini berfungsi sebagai pengantar sekaligus edukasi kepada pengguna mengenai perbedaan antara berita hoaks dan berita fakta sebelum melakukan proses klasifikasi.



Gambar 4. 9 Form Home

d) Classification

Gambar pertama menampilkan halaman Classification pada sistem Hoaxless. Halaman ini berfungsi sebagai tempat pengguna melakukan proses analisis berita. Di sisi kiri terdapat area input berupa kotak teks besar dengan instruksi “Input the news here. We’ll handle the rest...”, yang menunjukkan bahwa pengguna cukup memasukkan teks berita yang ingin diperiksa. Di bawahnya terdapat tombol “Execute_System_Scan” yang berfungsi untuk menjalankan proses klasifikasi. Tampilan ini dirancang menyerupai terminal

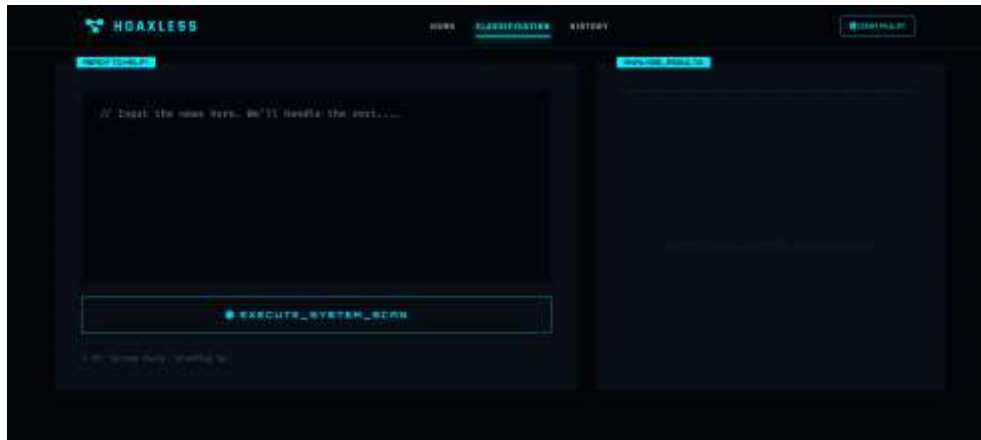
sistem dengan nuansa gelap dan aksen biru neon, sehingga memberi kesan proses analisis berbasis teknologi canggih. Di sisi kanan terdapat panel Analysis Results yang awalnya masih kosong dan akan menampilkan hasil setelah proses dijalankan.

Pada gambar kedua, terlihat contoh hasil analisis setelah pengguna memasukkan teks berita, yaitu “joko widodo tunjuk mahfud md jadi ketua kpk”. Setelah tombol dijalankan, panel Analysis Results menampilkan status deteksi dengan label merah HOAX disertai indikator Threat Detected: Anomaly. Sistem menjelaskan bahwa model SVM mendeteksi pola teks yang menyerupai data hoaks berdasarkan dataset pelatihan. Serta, terdapat informasi Decision Score (misalnya 0.8) yang menunjukkan tingkat keyakinan model terhadap hasil klasifikasi.

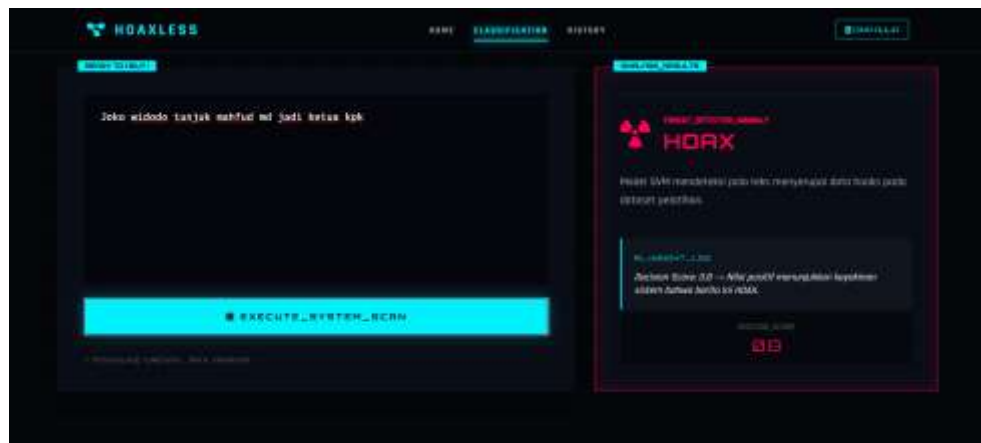
Pada gambar ketiga, terlihat contoh hasil analisis setelah pengguna memasukkan teks berita, yaitu “ketua demokrat kaltim kutip sby sorot proyek ikn nusantara”. Setelah tombol dijalankan, panel Analysis Results menampilkan status deteksi dengan label hijau FACT disertai indikator Integrity Verified Data. Sistem menjelaskan bahwa model SVM mendeteksi pola teks yang menyerupai data fakta berdasarkan dataset pelatihan. Serta, terdapat informasi Decision Score (misalnya -2.31) yang menunjukkan tingkat keyakinan model terhadap hasil klasifikasi.

Secara keseluruhan, halaman ini memperlihatkan alur kerja sistem klasifikasi teks berbasis Machine Learning, di mana pengguna memasukkan teks berita, sistem memproses menggunakan model SVM, lalu menampilkan

hasil analisis berupa label (Fakta/Hoaks), skor keputusan, dan tingkat ancaman secara visual dan informatif.



Gambar 4. 10 Form Classification



Gambar 4. 11 Deteksi Hoax

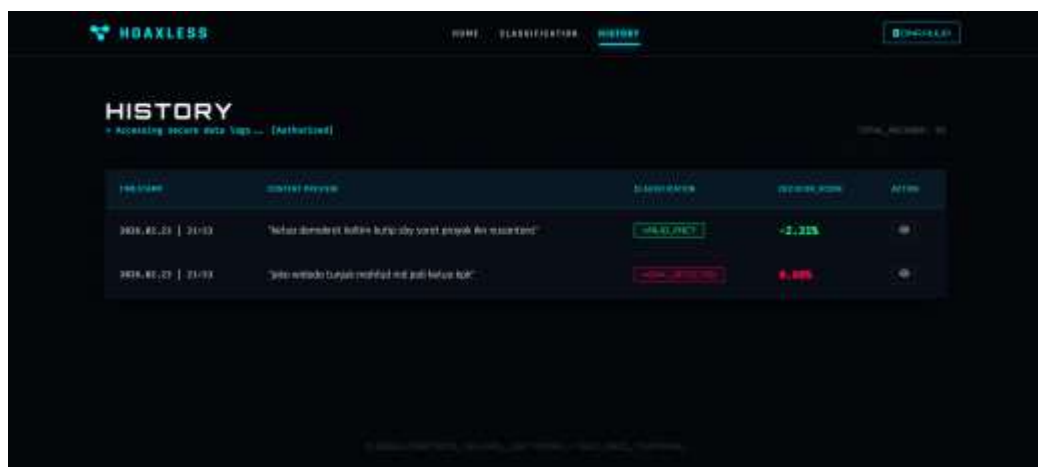


Gambar 4. 12 Deteksi Fakta

e) History

Gambar pertama menampilkan halaman History pada sistem Hoaxless. Halaman ini berfungsi sebagai riwayat hasil klasifikasi yang telah dilakukan oleh pengguna. Di bagian atas terdapat keterangan “Accessing secure data logs [Authorized]” yang menunjukkan bahwa data riwayat hanya dapat diakses oleh pengguna yang telah login. Data ditampilkan dalam bentuk tabel dengan beberapa kolom, yaitu Timestamp (tanggal dan waktu analisis), Content Preview (cuplikan teks berita yang dianalisis), Classification (hasil klasifikasi seperti Valid Fact atau Hoax Detected), Decision Score (nilai skor keputusan model), dan Action (ikon untuk melihat detail).

Pada contoh yang ditampilkan, terdapat dua riwayat analisis: satu diklasifikasikan sebagai Valid Fact dengan skor negatif (menunjukkan kecenderungan ke kelas fakta), dan satu lagi sebagai Hoax Detected dengan skor positif yang menunjukkan kecenderungan ke kelas hoaks. Warna hijau digunakan untuk fakta dan merah untuk hoaks agar perbedaan terlihat jelas secara visual. Halaman ini membantu pengguna memantau dan meninjau kembali hasil deteksi yang pernah dilakukan.



The screenshot shows the 'HISTORY' page of the Hoaxless system. At the top, there is a navigation bar with 'HOME', 'KLASIFIKATOR', and 'HISTORY' (highlighted). Below the navigation bar, the page title 'HISTORY' is displayed, followed by a status message: 'Accessing secure data logs ... [Authorized]'. The main content is a table with the following columns: 'TIMESTAMP', 'CONTENT PREVIEW', 'CLASSIFICATION', 'DECISION SCORE', and 'ACTION'. Two rows of data are visible:

TIMESTAMP	CONTENT PREVIEW	CLASSIFICATION	DECISION SCORE	ACTION
2023-02-23 21:03	"Beliau beranda di halaman kumpunya saat sedang beres-beres"	VALID FACT	-2.21%	🔍
2023-02-23 21:03	"Ada website yang terlihat tidak baik-baik saja"	HOAX DETECTED	6.88%	🔍

Gambar 4. 13 Form History

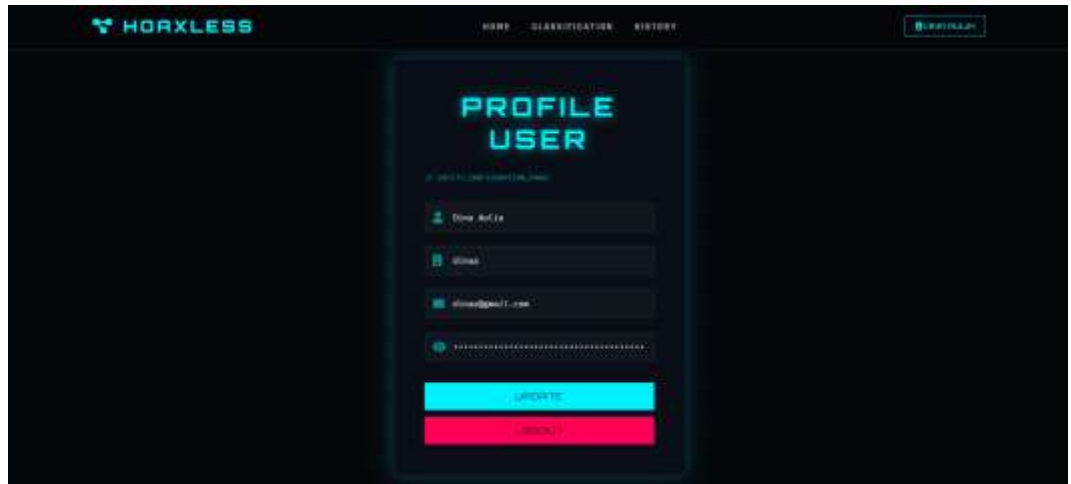
f) Profile

Gambar tersebut menampilkan halaman Profile User pada sistem Hoaxless. Halaman ini berfungsi sebagai panel pengelolaan dan konfigurasi akun pengguna setelah berhasil login. Di bagian tengah layar terdapat kartu profil dengan judul besar “PROFILE USER” yang menegaskan bahwa halaman ini digunakan untuk mengatur informasi identitas akun.

Di dalamnya terdapat beberapa field data, yaitu Nama lengkap, Username, Email, dan Password. Password ditampilkan dalam bentuk karakter tersembunyi (titik/asterisk) sebagai bentuk perlindungan keamanan data. Setiap kolom dilengkapi ikon untuk memperjelas jenis informasi yang diinput atau ditampilkan.

Di bagian bawah terdapat dua tombol utama, yaitu Update yang berfungsi untuk menyimpan perubahan data profil apabila pengguna melakukan pengeditan, serta Logout yang digunakan untuk keluar dari sistem dan mengakhiri sesi login.

Tampilan halaman ini tetap konsisten dengan desain keseluruhan sistem, menggunakan latar belakang gelap dengan aksen biru neon yang memberikan kesan futuristik dan profesional. Secara fungsional, halaman ini memastikan pengguna dapat mengelola informasi akunnya secara mandiri dan aman dalam sistem.



Gambar 4. 14 Profile

4.3 Pengujian Sistem

Pengujian sistem merupakan tahapan penting dalam pengembangan perangkat lunak yang bertujuan untuk memastikan bahwa sistem yang dibangun telah memenuhi kebutuhan fungsional yang telah ditetapkan. Pada penelitian ini, metode pengujian yang digunakan adalah Blackbox Testing, yaitu metode pengujian yang berfokus pada pengujian fungsi sistem tanpa menganalisis struktur internal atau kode program.

Pengujian dilakukan dengan memberikan berbagai skenario input pada setiap fitur utama aplikasi, kemudian mengamati kesesuaian output yang dihasilkan dengan hasil yang diharapkan. Fitur yang diuji meliputi halaman registrasi, login, dashboard, input teks berita, proses klasifikasi (deep scan), halaman riwayat (history), profile, serta fitur logout.

Berdasarkan hasil pengujian yang telah dilakukan, seluruh fitur sistem dapat berjalan sesuai dengan spesifikasi fungsional. Sistem mampu memproses input pengguna dengan baik, menampilkan hasil klasifikasi secara akurat sesuai dengan

model yang digunakan, serta memberikan respons yang tepat terhadap kondisi kesalahan (error handling).

Dengan demikian, dapat disimpulkan bahwa sistem yang dikembangkan telah memenuhi aspek fungsionalitas dan layak untuk digunakan sebagai aplikasi deteksi berita palsu berbasis machine learning.

4.3.1 Pengujian Fungsional (Black Box Testing)

Tabel 4. 1 Blackbox Testing Fitur Register

No	Fitur yang Diuji	Skenario Pengujian	Data Uji	Hasil yang diharapkan	Hasil pengujian	status
1	Register	Semua Field diisi dengan benar	Nama, username, email dan password valid	Sistem menyimpan data dan redirect ke login	Berhasil sesuai harapan	Valid
2	Register	Email sudah terdaftar	Email yang sama	Sistem menolak dan menampilkan pesan error	Berhasil menampilkan pesan error	Valid
3	Register	Field kosong	Salah satu field kosong	Sistem menampilkan peringatan	Muncul pesan peringatan	Valid

Tabel 4. 2 Blackbox Testing Fitur Login

No	Fitur yang Diuji	Skenario Pengujian	Data Uji	Hasil yang diharapkan	Hasil pengujian	status
1	Login	Username & password benar	Data user terdaftar	Masuk ke dashboard	Berhasil login	Valid
2	Login	Password salah	Password tidak sesuai	Sistem menolak login dan tampil pesan error	Pesan error muncul	Valid

3	Login	Field Kosong	Salah satu field kosong	System menampilkan peringatan	Peringatan muncul	Valid
---	-------	--------------	-------------------------	-------------------------------	-------------------	--------------

Tabel 4. 3 Blackbox Testing Halaman Dashboard

No	Fitur yang Diuji	Skenario Pengujian	Data Uji	Hasil yang diharapkan	Hasil pengujian	status
1	Akses halaman home	User mengakses halaman home	Klik menu Home	Halaman dashboard tampil dengan informasi sistem	Halaman tampil dengan baik	Valid
2	Navigasi menu	User memilih menu Classification	Klik menu Classification	Sistem mengarah ke halaman klasifikasi	Berhasil berpindah halaman	Valid
3	Navigasi menu	User memilih menu history	Klik menu History	Sistem mengarahkan ke halaman riwayat	Berhasil berpindah halaman	Valid

Tabel 4. 4 Blackbox Testing Fitur Klasifikasi Berita SVM

No	Fitur yang Diuji	Skenario Pengujian	Data Uji	Hasil yang diharapkan	Hasil pengujian	status
1	Input teks	User memasukkan teks berita	Teks berita valid	Sistem memproses teks menggunakan model SVM	Teks berhasil di proses	Valid
2	Klasifikasi Hoaks	Input berita hoaks	Sampel berita hoaks	Sistem menampilkan "HOAX"	Label tampil sesuai	Valid
3	Klasifikasi Fakta	Input berita fakta	Sampel berita fakta	Sistem menampilkan "FACT"	Label tampil sesuai	Valid
4	Input Kosong	User tidak memasukkan teks	-	Sistem menampilkan pesan peringatan	Pesan error muncul	Valid

5	Decision Score	Sistem menampilkan nilai decision score	Hasil klasifikasi	Nilai decision score tampil sesuai perhitungan mode;l	Nilai tampil dengan benar	Valid
---	----------------	---	-------------------	---	---------------------------	--------------

Tabel 4. 5 Blackbox Testing Halaman History

No	Fitur yang Diuji	Skenario Pengujian	Data Uji	Hasil yang diharapkan	Hasil pengujian	status
1	Akses history	User membuka halaman history	Klik menu history	Sistem menampilkan data Riwayat klasifikasi	Data tampil sesuai database	Valid
2	Tampilan data	Sistem menampilkan timestamp	Data tersimpan	Waktu klasifikasi tampil sesuai waktu input	Timestamp sesuai	Valid
3	Tampilan label	Sistem menampilkan hasil klasifikasi	Data hoaks dan fakta	Label tampilkan sesuaikan hasil sebelumnya	Label sesuai	Valid
4	Total records	Sistem menampilkan total data	Beberapa data yang tersimpan	Jumlah total records sesuai dengan jumlag di databases sesuai user id	Jumlah sesuai	Valid

Tabel 4. 6 Blackbox Testing Halaman Profile

No	Fitur yang Diuji	Skenario Pengujian	Data Uji	Hasil yang diharapkan	Hasil pengujian	status
1	Menampilkan data profile	User membuka halaman profile setelah login	Sesssion valid	Data nama, username, email tampil sesuai database	Data tampil sesuai	Valid

2	Update data profile	User mengubah nama / username / email lalu klik update	Nama : Dina Aulia	Data berhasil; terperbarui dan tersimpan di database	Data berhasil update	Valid
3	Validasi input kosong	User mengosongkan salah satu field lalu klik update	Nama kosong	Sistem menolak dan menampilkan pesan error	Validasi berjalan	Valid
4	Validasi format email	User memasukkan email tidak valid	Dina.gmail.com	Sistem menolak dan menampilkan notifikasi kesalahan	Notifikasi muncul	Valid
5	logout	User menekan tombol logout	Klik logout	Sistem menghapus session dan kembali ke halaman dashboard (home)	Logout berhasil	Valid

Tabel 4. 7 Blackbox Testing Penyimpanan ke Database

No	Fitur yang Diuji	Skenario Pengujian	Data Uji	Hasil yang diharapkan	Hasil pengujian	status
1	Penyimpanan Hasil	User melakukan klasifikasi	Input teks berita	Sistem menyimpan teks, hasil klasifikasi, decision score, dan timestamp	Data tersimpan dengan benar	Valid
2	Autentikasi Akses	User belum login mencoba akses History	Akses langsung URL	Sistem menolak akses dan redirect ke login	Redirect berjalan	Valid

4.4 Hasil Uji Coba

Pengujian sistem dilakukan menggunakan metode Blackbox Testing, yaitu pengujian yang berfokus pada fungsi sistem tanpa melihat struktur kode program. Pengujian ini bertujuan untuk memastikan bahwa setiap fitur dalam aplikasi berjalan sesuai dengan kebutuhan fungsional yang telah dirancang.

Berdasarkan hasil pengujian yang telah dilakukan, diperoleh hasil sebagai berikut:

1. Halaman Home (Dashboard).

Halaman utama berhasil ditampilkan dengan baik dan seluruh menu navigasi (Home, Classification, dan History) dapat diakses tanpa kendala. Sistem mampu mengarahkan pengguna ke halaman yang sesuai ketika menu dipilih.

2. Fitur Klasifikasi Berita.

Sistem mampu menerima input teks berita dari pengguna dan memprosesnya menggunakan model Support Vector Machine (SVM).

- a) Ketika diberikan teks berita hoaks, sistem menampilkan label *“HOAX”*.
- b) Ketika diberikan teks berita fakta, sistem menampilkan label *“FACT”*.
- c) Sistem juga menampilkan nilai *decision score* sebagai hasil perhitungan model.
- d) Jika input kosong, sistem menampilkan pesan peringatan sehingga tidak terjadi error.

3. Fitur Penyimpanan Data.

Setiap hasil klasifikasi yang dilakukan pengguna berhasil disimpan ke dalam basis data, termasuk teks berita, hasil klasifikasi, nilai decision score, dan waktu (timestamp).

4. Halaman History (Riwayat Deteksi).

Sistem mampu menampilkan data riwayat klasifikasi sesuai dengan data yang tersimpan pada database. Informasi yang ditampilkan meliputi waktu deteksi, cuplikan teks, hasil klasifikasi, serta decision score. Jumlah total data (total records) juga sesuai dengan jumlah data pada database.

5. Keamanan Akses.

Sistem membatasi akses halaman tertentu (seperti History) hanya untuk pengguna yang telah login. Jika pengguna belum login, sistem secara otomatis mengarahkan ke halaman login.

Berdasarkan seluruh pengujian tersebut, dapat disimpulkan bahwa aplikasi telah berjalan sesuai dengan kebutuhan fungsional dan tidak ditemukan kesalahan sistem yang bersifat kritis.

4.5 Kelebihan Aplikasi

Berdasarkan hasil pengujian dan analisis sistem yang telah dilakukan, aplikasi ini memiliki beberapa kelebihan, antara lain:

1. **Klasifikasi Otomatis Berbasis Machine Learning.** Sistem menggunakan algoritma Support Vector Machine (SVM) yang telah terbukti efektif dalam tugas klasifikasi teks. Proses klasifikasi dilakukan secara otomatis sehingga pengguna tidak perlu memiliki keahlian teknis khusus untuk menggunakannya.

2. **Antarmuka yang Mudah Digunakan.** Aplikasi dirancang dengan tampilan yang sederhana dan intuitif. Pengguna cukup memasukkan teks berita dan sistem akan langsung memberikan hasil klasifikasi secara cepat, sehingga dapat digunakan oleh masyarakat umum tanpa hambatan teknis.
3. **Transparansi Hasil melalui Decision Score.** Selain menampilkan label klasifikasi (HOAX atau FACT), sistem juga menampilkan nilai *decision score* yang mencerminkan tingkat keyakinan model terhadap hasil prediksi. Hal ini meningkatkan transparansi dan kepercayaan pengguna terhadap sistem.
4. **Penyimpanan Riwayat Klasifikasi.** Setiap hasil klasifikasi tersimpan secara otomatis ke dalam basis data beserta informasi timestamp. Fitur ini memungkinkan pengguna untuk memantau dan meninjau kembali riwayat deteksi berita yang telah dilakukan sebelumnya.
5. **Validasi Input yang Baik.** Sistem dilengkapi dengan validasi input sehingga apabila pengguna tidak mengisi kolom teks, sistem akan menampilkan pesan peringatan tanpa menimbulkan error. Hal ini menjaga stabilitas dan keandalan aplikasi.
6. **Keamanan Akses Pengguna.** Sistem menerapkan mekanisme autentikasi yang membatasi akses ke fitur tertentu hanya bagi pengguna yang telah login. Ini menjaga keamanan data riwayat klasifikasi dari akses yang tidak sah.
7. **Respons Sistem yang Cepat.** Proses klasifikasi teks berlangsung dalam waktu singkat karena model SVM yang digunakan telah dilatih dan

disimpan sebelumnya (*pre-trained model*), sehingga tidak perlu proses pelatihan ulang setiap kali pengguna melakukan deteksi.

4.6 Kekurangan Aplikasi

Meskipun sistem telah berjalan dengan baik, terdapat beberapa keterbatasan yang masih dimiliki oleh aplikasi ini, antara lain:

1. Model Terbatas pada Dataset Tertentu.

Model klasifikasi dilatih menggunakan dataset tertentu (TurnBackHoax), sehingga performa sistem sangat bergantung pada kualitas dan cakupan dataset tersebut. Jika diberikan berita dengan topik yang sangat berbeda, kemungkinan akurasi dapat menurun.

2. Hanya Menggunakan Analisis Teks.

Sistem hanya menganalisis konten teks dan belum mampu mendeteksi hoaks berbasis gambar, video, atau multimedia lainnya.

3. Belum Mendukung Multi-User Role.

Sistem belum memiliki pembagian hak akses lanjutan seperti admin dan user dengan fitur manajemen data yang berbeda.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian mengenai *Klasifikasi Berita Hoaks dan Fakta Menggunakan Support Vector Machine (SVM) pada Dataset TurnBackHoax dan CNN Indonesia*, maka dapat diperoleh beberapa kesimpulan sebagai berikut:

1. Tahapan preprocessing yang meliputi *cleaning, case folding, tokenizing, stopword removal*, dan *stemming* terbukti mampu meningkatkan kualitas data teks dengan menghilangkan noise dan menyederhanakan struktur kata. Proses ini berperan penting dalam meningkatkan performa model klasifikasi.
2. Metode pembobotan fitur menggunakan TF-IDF (Term Frequency–Inverse Document Frequency) efektif dalam merepresentasikan teks berita ke dalam bentuk numerik sehingga dapat diproses oleh algoritma machine learning. Representasi ini membantu model dalam mengidentifikasi kata-kata yang memiliki kontribusi signifikan terhadap proses klasifikasi.
3. Algoritma Support Vector Machine (SVM) mampu membangun model klasifikasi yang optimal dalam memisahkan kelas berita hoaks dan fakta berdasarkan hyperplane terbaik. Penggunaan SVM menunjukkan kemampuan yang baik dalam menangani data berdimensi tinggi seperti data teks.
4. Berdasarkan hasil evaluasi menggunakan confusion matrix serta metrik accuracy, precision, recall, dan F1-score, model menunjukkan performa yang baik dengan tingkat akurasi sebesar 96.43%. Nilai tersebut

menunjukkan bahwa model memiliki kemampuan yang cukup tinggi dalam membedakan berita hoaks dan fakta secara tepat.

5. Implementasi model ke dalam sistem berbasis web berhasil dilakukan dengan mengintegrasikan model machine learning menggunakan Python (Flask) dan basis data untuk penyimpanan hasil klasifikasi. Sistem yang dibangun memungkinkan pengguna melakukan deteksi berita secara langsung melalui antarmuka yang interaktif.

Secara keseluruhan, penelitian ini membuktikan bahwa pendekatan Support Vector Machine (SVM) dengan pembobotan TF-IDF merupakan metode yang efektif dan relevan dalam melakukan klasifikasi berita hoaks pada media online Indonesia, khususnya dalam konteks kategori hoaks dan fakta.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran untuk pengembangan lebih lanjut, yaitu:

1. Penelitian selanjutnya dapat menggunakan dataset dengan jumlah data yang lebih besar dan distribusi kelas yang lebih seimbang agar model memiliki kemampuan generalisasi yang lebih baik terhadap data baru.
2. Dapat dilakukan perbandingan performa antara SVM dengan algoritma klasifikasi lainnya seperti Naïve Bayes, Random Forest, K-Nearest Neighbor, maupun pendekatan berbasis Deep Learning seperti LSTM atau BERT untuk memperoleh model dengan akurasi yang lebih optimal.
3. Penggunaan teknik representasi fitur yang lebih canggih seperti Word2Vec, FastText, atau contextual embedding (misalnya IndoBERT) dapat diteliti lebih lanjut untuk meningkatkan kualitas ekstraksi fitur.

4. Sistem dapat dikembangkan dengan menambahkan fitur analisis tambahan seperti tingkat kepercayaan prediksi (*confidence score*), visualisasi data, atau analisis tren penyebaran hoaks.
5. Pengembangan sistem ke dalam bentuk aplikasi mobile atau integrasi dengan platform media sosial dapat meningkatkan pemanfaatan sistem secara lebih luas dalam membantu masyarakat mendeteksi dan meminimalisir penyebaran berita hoaks.

DAFTAR PUSTAKA

- Al-Khowarizmi, Sari, I. P., & Maulana, H. (2024). Optimization of support vector machine with cubic kernel function to detect cyberbullying in social networks. *Telkomnika (Telecommunication Computing Electronics and Control)*, 22(2), 329–339. <https://doi.org/10.12928/TELKOMNIKA.v22i2.25437>
- Andi Zulherry, Fanny Ramadhani, & Andy Satria. (2024). Klasifikasi Data Tracer Study Dengan Pemanfaatan Data Mining Menggunakan Algoritma Support Vector Machine dan Neural Network. *Portal Riset Dan Inovasi Sistem Perangkat Lunak*, 2(1), 45–54. <https://doi.org/10.59696/prinsip.v2i1.73>
- Anisya, C. (2024). *IMPLEMENTASI METODE SUPPORT VECTOR MACHINE (SVM) DALAM KLASIFIKASI JUDUL BERITA HOAKS POLITIK*.
- Asriadi, M., & Hasyim, Q. (2024). *Pelatihan Literasi Media Sosial Bagi Pemilih Pemula Tentang Hoax, Hate Speech dan Negative Campaign*. 02, 124–133.
- Aulia Fanny, & Kaswadi Kaswadi. (2024). 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*, 2(4), 213–222. <https://doi.org/10.61132/pragmatik.v2i4.1071>
- Azis, M. A., Respati, A. A., & Daniati, E. (2025). *Rancangan Sistem Deteksi Berita Hoaks dengan IndoBERT Berbasis Dataset Scraping Seimbang (2020–2025)* (Vol. 9). Online.
- Binus University. (2022, February 14). *Support Vector Machine Algorithm*. <https://sis.binus.ac.id/> <https://sis.binus.ac.id/2022/02/14/support-vector-machine-algorithm/>
- Dongoran, D., & Sari, I. P. (2025). 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*, 4(1), 12–24. <https://doi.org/10.56211/helloworld.v4i1.619>
- FEBRIYANTY, N. E. (2023). *DETEKSI BERITA HOAX DARI MEDIA ONLINE INDONESIA MENGGUNAKAN ALGORITMA NAIVE BAYES DAN SUPPORT VECTOR MACHINE*.

- Fernandes, R. P., & Tahara Shita, R. (2024). Penerapan Metode SVM dan Random Forest untuk Mendeteksi Berita Hoaks pada PT. Global Arrow. *Jurnal TICOM: Technology of Information and Communication*, 12(3), 2024. <https://www.jpnn.com/tag/asli>
- geeksforgeeks. (2025, May 28). *Algoritma Support Vector Machine (SVM)*. <https://www.geeksforgeeks.org/machine-learning/support-vector-machine-algorithm/>.
- Jocelyne, C., Tobing, L., Lanang Wijayakusuma, I., Putu, L., & Harini, I. (2025). Detection of Political Hoax News Using Fine-Tuning IndoBERT. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 9, Issue 2). <http://jurnal.polibatam.ac.id/index.php/JAIC>
- Kamal, A. M., Chrisnanto, Y. H., & Yuniarti, R. (2025). Identifikasi Berita Palsu di Portal Media Online Menggunakan Model IndoBERT dan LSTM. *JURIKOM (Jurnal Riset Komputer)*, 12(3), 287–297. <https://doi.org/10.30865/jurikom.v12i3.8660>
- Kementerian Komunikasi dan Digital. (2025, January 8). *Komdigi Identifikasi 1.923 Konten Hoaks Sepanjang Tahun 2024*. Siaran Pers No. 08/HM-KKD/01/2025. <https://www.komdigi.go.id/berita/siaran-pers/detail/komdigi-identifikasi-1923-konten-hoaks-sepanjang-tahun-2024>
- Khasanah, I. N., Syelena Azzahra, N., Febrianto, L., & Saifudin, M. R. (2024). *Perbedaan Fakta dan Non Fakta di Media : Definisi, Jenis, dan Memeriksa Fakta Informasi Digital*.
- Kurnia, Y., Kusuma, E. D., Kusuma, L. W., Suwitno, & Apridius, W. (2024). Perbandingan Naïve Bayes dan CNN yang Dioptimasi PSO pada Identifikasi Berita Hoax Politik Indonesia. *Bit-Tech*, 6(3), 340–352. <https://doi.org/10.32877/bt.v6i3.1225>
- Mubarok, M. I., & Abdi, M. (2024). IMPLEMENTASI NATURAL LANGUAGE PROCESSING DALAM PERANCANGAN APLIKASI CHATBOT PADA FIKTI UMSU. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 6).
- Purnama Sari, I., Krianto Sulaiman, O., Apdilah, D., & Simanjuntak, P. (2025). Analisis Komparatif Support Vector Machine dan Random Forest untuk

- Deteksi Email Phishing. *AICOMS* 2024, 4(2), 18–27. <https://doi.org/10.58466/aicoms.v4i1.1806>
- Rahmadhany, A., Aldila Safitri, A., & Irwansyah, I. (2021). Fenomena Penyebaran Hoax dan Hate Speech pada Media Sosial. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 3(1), 30–43. <https://doi.org/10.47233/jteksis.v3i1.182>
- Rahmawati. (2022). *DETEKSI BERITA HOAX PADA WEBSITE TURNBACKHOAX DENGAN MENGGUNAKAN MACHINE LEARNING*.
- Rahmawati, D., Setyo Putro Robawa, R., Faiq Al Abiyyi, M., Daffa NRF, P., Ilman Nugraha, R., Puguh Margono, F., Praditya, Mi. A., & Sholihatin, E. (2023). *Analisis Hoaks dalam Konteks Digital: Implikasi dan Pencegahannya di Indonesia*. 3, 10819–10829.
- Ropikoh, I. A., Abdulhakim, R., Enri, U., & Sulistiyowati, N. (2021). Penerapan Algoritma Support Vector Machine (SVM) untuk Klasifikasi Berita Hoax Covid-19. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 5, Issue 1). <http://jurnal.polibatam.ac.id/index.php/JAIC>
- Sarjito, A. (2021). Hoaks, Disinformasi, dan Ketahanan Nasional: Ancaman Teknologi Informasi dalam Masyarakat Digital Indonesia. In *Journal of Governance and Local Politics* (Vol. 5, Issue 2).
- Satrivi, N., & Purnama, C. (2021). Pembentukan Opini Publik Indonesia oleh Cable News Network (CNN) Indonesia Berkenaan dengan Isu Sampah Plastik. *Padjadjaran Journal of International Relations*, 3(2), 228–241. <https://doi.org/10.24198/padjir.v3i2.31172.33515>
- Wayan Sumartini Saraswati, N., Putu Krisna Suarendra Putra, I., Dewa Made Krishna Muku, I., & Dana Pramitha, G. (2023). Support Vector Machine For Hoax Detection. *SINTECH JOURNAL*, 6, 107–117. <https://doi.org/10.31598>
- Yahya, & Mahpuz. (2019). Penggunaan Algoritma K-Means Untuk Menganalisis Pelanggan Potensial Pada Dealer SPS Motor Honda Lombok Timur Nusa Tenggara Barat. *Jurnal Informatika Dan Teknologi*, 2(2), 109–118.
- Yopita Desriana Butar. (2024). Analisis Penyebaran Hoax Di Media Sosial Dan Dampaknya Terhadap Masyarakat. *Jurnal Pendidikan, Bahasa Dan Budaya*, 3(2), 252–258. <https://doi.org/10.55606/jpbb.v3i2.3201>

Fix Skripsi 2026

ORIGINALITY REPORT

29%
SIMILARITY INDEX

27%
INTERNET SOURCES

15%
PUBLICATIONS

17%
STUDENT PAPERS

PRIMARY SOURCES

1	repository.umsu.ac.id Internet Source	4%
2	Submitted to Universitas Muhammadiyah Palembang Student Paper	1%
3	Submitted to Universitas Muhammadiyah Sumatera Utara Student Paper	1%
4	jurnal.polibatam.ac.id Internet Source	1%
5	journal.diginus.id Internet Source	1%
6	medium.com Internet Source	1%
7	repository.itpln.ac.id Internet Source	1%
8	Submitted to Universitas Putera Batam Student Paper	1%
9	Abidin Abidin. "ANALISIS SENTIMEN ULASAN APLIKASI SIREKAP MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE", Jurnal Informatika dan Teknik Elektro Terapan, 2025 Publication	1%
10	journal.unpacti.ac.id	

	Internet Source	1 %
11	Indra Purnama, Yeni Setiani, Farhan Ari Nur Wibisono. "Analisis Dan Visualisasi Data Menggunakan Looker Studio Pada Dataset New York City Property Sales", Jurnal Minfo Polgan, 2025 Publication	<1 %
12	journal.unpad.ac.id Internet Source	<1 %
13	Submitted to UPN Veteran Yogyakarta Students Paper	<1 %
14	ejurnal.seminar-id.com Internet Source	<1 %
15	Submitted to UNIVERSITAS BUDI LUHUR Students Paper	<1 %
16	etheses.uin-malang.ac.id Internet Source	<1 %
17	text-id.123dok.com Internet Source	<1 %
18	Submitted to Universitas Muria Kudus Students Paper	<1 %
19	ejurnal.stmik-budidarma.ac.id Internet Source	<1 %
20	Yonal Supit. "Penerapan Algoritma Support Vector Machine (SVM) untuk Deteksi Penyakit Mata Katarak", INDONESIA JOURNAL ON DATA SCIENCE, 2025 Publication	<1 %
21	Submitted to Universitas Muhammadiyah Purwokerto	<1 %

Student Paper

22	j-innovative.org Internet Source	<1 %
23	Submitted to Universitas Brawijaya Student Paper	<1 %
24	repository.uinsaizu.ac.id Internet Source	<1 %
25	jurnal-ticom.jakarta.aptikom.org Internet Source	<1 %
26	Submitted to Telkom University Student Paper	<1 %
27	Submitted to Universitas Musamus Merauke Student Paper	<1 %
28	Submitted to Universitas Sumatera Utara Student Paper	<1 %
29	Submitted to Universitas Muslim Indonesia Student Paper	<1 %
30	Submitted to Universitas Gadjah Mada Student Paper	<1 %
31	ejournal.antarbangsa.ac.id Internet Source	<1 %
32	repository.uigm.ac.id Internet Source	<1 %
33	repository.unissula.ac.id Internet Source	<1 %
34	digilib.uinsgd.ac.id Internet Source	<1 %
35	repository.unived.ac.id Internet Source	<1 %

36	repository.ub.ac.id Internet Source	<1 %
37	Ichsani Mursidah, Remi Sanjaya, Bambang Yulianto, Dhian Sweetania, Puji Sularsih. "Analisis Sentimen Pengguna pada Ulasan Game Honkai Star Rail Menggunakan Algoritma Support Vector Machine (SVM)", <i>Jurnal Minfo Polgan</i> , 2025 Publication	<1 %
38	Submitted to Universitas Islam Indonesia Student Paper	<1 %
39	shallbd.com Internet Source	<1 %
40	sriwijayamediapermata.id Internet Source	<1 %
41	repository.its.ac.id Internet Source	<1 %
42	Erma Puspita, Nuril Lutvi Azizah, Cindy Taurusta, Yunianita Rahmawati. "Integrated Population Administration System With Automated WhatsApp Notifications", <i>Academia Open</i> , 2025 Publication	<1 %
43	Junaidy, Muhammad Fauzan, Roberto Kaban. "Analisis Sentimen Publik Terhadap RUU KUHAP di Platform X Menggunakan Metode TF-IDF dan Naïve Bayes", <i>Jurnal Dinamika Informatika</i> , 2026 Publication	<1 %
44	repository.uin-suska.ac.id Internet Source	<1 %

45	digilib.uns.ac.id Internet Source	<1 %
46	eprints.amikom.ac.id Internet Source	<1 %
47	Submitted to Sriwijaya University Student Paper	<1 %
48	docplayer.info Internet Source	<1 %
49	www.scribd.com Internet Source	<1 %
50	id.mashable.com Internet Source	<1 %
51	Zahrotun Ni'mah -, Bambang Irawan -, Nur Ariesanto Ramdhan -. "ANALISIS SENTIMEN ULASAN PENGGUNA TERHADAP GAME ZENLESS ZONE ZERO MENGGUNAKAN METODE BI-DIRECTIONAL LSTM", Jurnal Informatika dan Teknik Elektro Terapan, 2026 Publication	<1 %
52	hostjournals.com Internet Source	<1 %
53	123dok.com Internet Source	<1 %
54	Maulana Ichsan Ichsan. "IMPLEMENTASI MACHINE LEARNING UNTUK DETEKSI PENYAKIT PADA KUCING MENGGUNAKAN RANDOM FOREST", Jurnal Informatika dan Teknik Elektro Terapan, 2025 Publication	<1 %
55	Submitted to University of Sussex Student Paper	<1 %

56	repository.usu.ac.id Internet Source	<1 %
57	M. Adreansyah Pratama Lubis, Ria Eka Sari. "Penerapan Metode Klasifikasi Support Vector Machine dalam Analisis Sentiment Ulasan Pengguna Layanan Goride Pada Aplikasi Gojek di Google Playstore", <i>DEVICE : JOURNAL OF INFORMATION SYSTEM, COMPUTER SCIENCE AND INFORMATION TECHNOLOGY</i> , 2025 Publication	<1 %
58	Submitted to Universitas Tarumanagara Student Paper	<1 %
59	jurnal.itg.ac.id Internet Source	<1 %
60	kc.umn.ac.id Internet Source	<1 %
61	repository.upr.ac.id Internet Source	<1 %
62	Lely Kurniawati, Dadang Priyanto, Neny Sulistia Ningsih, Moch Syahrir, Ria Rismayati. "Comparison of Decision Tree-Based Methods in Lung Disease Detection", <i>Jurnal Bumigora Information Technology (BITe)</i> , 2025 Publication	<1 %
63	Submitted to Universitas Pancasila Student Paper	<1 %
64	Submitted to University of Hertfordshire Student Paper	<1 %
65	eprints.amikompurwokerto.ac.id Internet Source	<1 %

66	jurnal.risetilmiah.ac.id Internet Source	<1 %
67	repository.teknokrat.ac.id Internet Source	<1 %
68	Submitted to Universitas 17 Agustus 1945 Semarang Student Paper	<1 %
69	joincs.umsida.ac.id Internet Source	<1 %
70	proceeding.unpkediri.ac.id Internet Source	<1 %
71	Submitted to Universitas Dinamika Student Paper	<1 %
72	Submitted to Universitas Islam Negeri Raden Fatah Student Paper	<1 %
73	Submitted to Universitas Negeri Semarang - iTh Student Paper	<1 %
74	Muhamad Azril, Giatika Crisnawati. "Klasifikasi Sentimen Komentar Pengguna Tiktok Mengenai Program Makan Bergizi Gratis (MBG) Dengan Support Vector Machine (SVM)", RIGGS: Journal of Artificial Intelligence and Digital Business, 2026 Publication	<1 %
75	Nyoman Dinda Indira Sudiasta Putri, I Made Dendi Maysanjaya, I Made Gede Sunarya. "Perbandingan Kinerja Algoritma Naive Bayes dan K-Nearest Neighbor dalam Menganalisis Sentimen Pengguna Game Free Fire", Pseudocode, 2025	<1 %

Publication

76	Submitted to Universitas Islam Malang Student Paper	<1 %
77	docs.google.com Internet Source	<1 %
78	eprints.umk.ac.id Internet Source	<1 %
79	repository.unisbablitar.ac.id Internet Source	<1 %
80	repository.usni.ac.id Internet Source	<1 %
81	sefidvash.net Internet Source	<1 %
82	Mubarak Mubarak, Lili Tanti, Rika Rosnelly. "Perbandingan Algoritma Decision Tree dan Naive Bayes Pada Analisis Sentimen Masyarakat Terhadap Pejabat Pertamina Pasca Kasus Pertamina Oplosan", Jurnal Minfo Polgan, 2026 Publication	<1 %
83	Submitted to UIN Sultan Syarif Kasim Riau Student Paper	<1 %
84	Submitted to UM Surabaya Student Paper	<1 %
85	ojs.uajy.ac.id Internet Source	<1 %
86	www.repository.umuslim.ac.id Internet Source	<1 %
87	Arkananta Emier, Daffa Arief, Aqila Adam, Jubel Hiero Oktovan Lumban Gaol et al.	<1 %

"Penerapan Aplikasi Kasir Berbasis Web Untuk Mengoptimalkan Layanan Transaksi Pada UMKM", RIGGS: Journal of Artificial Intelligence and Digital Business, 2026

Publication

88

Submitted to Universitas Muhammadiyah Prof. Dr. Hamka

Student Paper

<1 %

89

eprints.walisongo.ac.id

Internet Source

<1 %

90

repository.uinsu.ac.id

Internet Source

<1 %

91

Hafizha Nurul Qolby, Rangga Gelar Guntara, Syti Sarah Maesaroh. "Comparative Analysis of Machine Learning Models for BUMN Bank Stock Sentiment Classification During Danantara Formation Period", INOVTEK Polbeng - Seri Informatika, 2025

Publication

<1 %

92

Kiki Setiawan, Humam Mu'asyir. "Perbandingan Metode Naïve Bayes dan Support Vector Machine untuk Klasifikasi Sentimen Ulasan Wisatawan: Studi Kasus Mulia Resort Nusa Dua Bali", Jurnal JTİK (Jurnal Teknologi Informasi dan Komunikasi), 2026

Publication

<1 %

93

Muhammad Irvan Maulana, Savana Putra Aditama, Harun Al Rosyid. "Komparasi Algoritma Naive Bayes dan Support Vector Machine pada Analisis Sentimen Komentar Instagram Laga El Clásico Barcelona vs Real Madrid", Jurnal Dinamika Informatika, 2026

Publication

<1 %

94	Sebastian Ubaidillah Royan, Nana Suarna, Irfan Ali, Dodi Solihudin. "ANALISIS SENTIMEN ULASAN PRODUK SKINCARE DI SHOPEE UNTUK MENINGKATKAN KUALITAS PRODUK MENGGUNAKAN METODE SUPPORT VECTOR MACHINE", Jurnal Informasi dan Komputer, 2025 Publication	<1 %
95	Submitted to Universitas Islam Riau Student Paper	<1 %
96	acopen.umsida.ac.id Internet Source	<1 %
97	citarumpedia.id Internet Source	<1 %
98	doku.pub Internet Source	<1 %
99	ejournal.catursakti.ac.id Internet Source	<1 %
100	journalversa.com Internet Source	<1 %
101	jurnal.yudharta.ac.id Internet Source	<1 %
102	jurnalkwangsan.kemdikbud.go.id Internet Source	<1 %
103	penerbitadm.pubmedia.id Internet Source	<1 %
104	su-plus.strathmore.edu Internet Source	<1 %
105	www.researchgate.net Internet Source	<1 %

106	www.trivusi.web.id Internet Source	<1 %
107	Erika Sita Dewi, Ariella Chandra Naya, Yuyun Umaidah. "Analisis Sentimen Komentar Instagram terkait Isu Sosial pada Akun Detikcom", Jurnal Pendidikan Tambusai, 2026 Publication	<1 %
108	Irfandi Rusdiansyah, Ridwan Pangestu, Devina Azalia, Muhammad Faiz Zhafran, Ferdy Saputra, Fachri Amsury. "Integrasi Model Klasifikasi Tingkat Stress Mahasiswa Berbasis Natural Language Processing", RIGGS: Journal of Artificial Intelligence and Digital Business, 2025 Publication	<1 %
109	Muhamad Ali Zaenal Abidin. "Evaluasi Sentimen Ulasan Pengguna CGV Cinemas Indonesia Menggunakan Metode Naive Bayes dan Support Vector Machine", Indonesian Journal on Software Engineering (IJSE), 2025 Publication	<1 %
110	Submitted to Universitas Khairun Student Paper	<1 %
111	birdsofpeace.org Internet Source	<1 %
112	conference.ut.ac.id Internet Source	<1 %
113	core.ac.uk Internet Source	<1 %
114	dl.ucsc.cmb.ac.lk Internet Source	<1 %

115	ejournal.pppmitpa.or.id Internet Source	<1 %
116	ejournal.undiksha.ac.id Internet Source	<1 %
117	ejurnal.poliban.ac.id Internet Source	<1 %
118	id,123dok.com Internet Source	<1 %
119	idr.uin-antasari.ac.id Internet Source	<1 %
120	journal.ilmudata.co.id Internet Source	<1 %
121	pt.scribd.com Internet Source	<1 %
122	repository.uph.edu Internet Source	<1 %
123	repository.usu.ac.id Internet Source	<1 %
124	sipora.polije.ac.id Internet Source	<1 %
125	www.cnnindonesia.com Internet Source	<1 %
126	Fahmi, Moh Hisyam. "Pengaruh Employee Engagement Melalui Religiosity Terhadap Job Satisfaction dan Employee Performance Aparatur Sipil Negara Kabupaten Batang", Universitas Islam Sultan Agung (Indonesia), 2024 Publication	<1 %

127	Submitted to Hankuk University of Foreign Studies <small>Student Paper</small>	<1 %
128	Jeremi Azero Putra, Alexander Dharmawan, Jutono Gondohanindijo. "Sentimen Analisis Aplikasi Digitalent Mobile Menggunakan Naive Bayes Dan SVM Dengan Ekstraksi Fitur TT-IDF", INTECOMS: Journal of Information Technology and Computer Science, 2024 <small>Publication</small>	<1 %
129	M Choirul Amri, Muhammad Farros Atiqi, Friza Talita, Ilka Zufria. "Implementasi Algoritma Naive Bayes Untuk Klasifikasi Berita Online Terhadap Program Prioritas Pemerintah Kota Medan", DEVICE : JOURNAL OF INFORMATION SYSTEM, COMPUTER SCIENCE AND INFORMATION TECHNOLOGY, 2025 <small>Publication</small>	<1 %
130	Yudha Prastya. "HYBRID CNN-SVM UNTUK ANALISIS SENTIMEN KLUB TIM NASIONAL", Jurnal Informatika dan Teknik Elektro Terapan, 2025 <small>Publication</small>	<1 %
131	Yulia Fatma, Rahmad Gunawan, Nurkhairi Fitri, Rahmad Firdaus, Regiolina Hayami, Soni Soni. "Klasifikasi Algoritma Kriptografi pada Pesan Terenkripsi menggunakan Support Vector Machine (SVM)", JURNAL FASILKOM, 2026 <small>Publication</small>	<1 %
132	ejournal2.unud.ac.id <small>Internet Source</small>	<1 %

133	elibrary.undipa.ac.id Internet Source	<1 %
134	journal.lppmunindra.ac.id Internet Source	<1 %
135	jurnal.arkainstitute.co.id Internet Source	<1 %
136	jurnal.kdi.or.id Internet Source	<1 %
137	ojs.stmik-banjarbaru.ac.id Internet Source	<1 %
138	pika.ugm.ac.id Internet Source	<1 %
139	repo.itera.ac.id Internet Source	<1 %
140	repositori.usu.ac.id:8080 Internet Source	<1 %
141	repository.dinamika.ac.id Internet Source	<1 %
142	repository.ppns.ac.id Internet Source	<1 %
143	subset.id Internet Source	<1 %
144	tekno.kompas.com Internet Source	<1 %
145	turnbackhoax.id Internet Source	<1 %
146	www.e-journal.stmiklombok.ac.id Internet Source	<1 %
147	www.journal.sekawan-org.id	

Internet Source

<1 %

148

Hari Murti, Sulastris Sulastris, Dwi Budi Santoso, Dwi Agus Diartono, Kristiawan Nugroho. "Design of Intelligent Model for Text-Based Fake News Detection Using K-Nearest Neighbor Method", sinkron, 2025

Publication

<1 %

149

Ichwanulhadi, M.. "Pengembangan Alat Kalibrator Inkubator Bayi (Incubator Tester) Berbasis Internet of Things (IOT)", Universitas Islam Sultan Agung (Indonesia), 2024

Publication

<1 %

150

Windi Clarisha, A. Astri Merilsa Fani, Dewi Fatmarani Suriyanto, Nur Fadilah. "Sentiment Analysis of Local Sunscreen Skintific, Somethinc, and Avoskin with Naive Bayes and SVM", Jambura Journal of Electrical and Electronics Engineering, 2025

Publication

<1 %

151

Yusuf Kurnia, Ellysha Dwiyanthi Kusuma, Lianny Wydiastuty Kusuma, Suwitno, Welman Apridius. "Perbandingan Naive Bayes dan CNN yang Dioptimasi PSO pada Identifikasi Berita Hoax Politik Indonesia", bit-Tech, 2024

Publication

<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography On



Medan, 28 April 2026

No : 009/LoA/ALGORITMA-Vol.10 No.1/4/2026

Hal : Surat Keterangan Penerimaan Naskah Publikasi Artikel Ilmiah

Kepada Yth :

Bapak/ Ibu Penulis (*Author*)

Dina Aulia; Mahardika Abdi Prawira;

Di

Tempat

Assalamu'alaikum Warahmatullahi Wabarakatuh
Dengan Hormat,

Tim Redaksi Jurnal Ilmu Komputer dan Informatika "ALGORITMA" Prodi Ilmu Komputer Fakultas Sains dan Teknologi UIN Sumatera Utara Medan telah menerima paper dengan judul:

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

Menyatakan bahwa naskah tersebut telah diterima oleh tim redaksi Jurnal Ilmu Komputer dan Informatika ALGORITMA dan dinyatakan **LULUS** seleksi reviewer dan akan segera diterbitkan oleh Jurnal Ilmu Komputer dan Informatika "ALGORITMA" pada Volume 10 No.1, April 2026.

Demikian surat keterangan ini dibuat untuk dapat dipergunakan sebagaimana mestinya.

Pimpinan Dewan Redaksi

A circular blue stamp with the text "JURNAL ILMU KOMPUTER DAN INFORMATIKA" around the perimeter and "ALGORITMA" in the center. Overlaid on the stamp is a handwritten signature in black ink that reads "Furqan".

Dr. Mhd. Furqan, S.Si., M.Comp.Sc



Prodi Ilmu Komputer, Fakultas Sains dan Teknologi UIN Sumatera Utara Medan
Jl. Lap. Golf No.120, Kabupaten Deli Serdang, Sumatera Utara 20353, Indonesia
Email : algoritma@uinsu.ac.id

Accredited
SINTA 5

ALGORITMA

Jurnal Ilmu Komputer dan Informatika

E-ISSN : 2598-6341

[HOME](#) [ABOUT](#) [USER HOME](#) [SEARCH](#) [CURRENT](#) [ARCHIVES](#) [ANNOUNCEMENTS](#)

[Home](#) > [User](#) > [Author](#) > **Active Submissions**

Active Submissions

ACTIVE
ARCHIVE

ID	WAKTU SUBMIT	SEC	AUTHOR	TITLE	STATUS
29421	04-24	AHT	Aulia, Tanjung	IMPLEMENTASI DETEKSI BENYAK HONG SAY PAKSA MENGGUNAKAN...	Awaiting assignment

Start a New Submission
[CLICK HERE](#) to go to step one of the five step submission process.

Refbacks

ALL
NEW
PUBLISHED
HONORED

DATE ADDED	HTS	URL	ARTICLE	TITLE	STATUS	ACTION
There are currently no refbacks.						

- [Editorial Team](#)
- [Peer Reviewer](#)
- [Focus and Scope](#)
- [Publication Ethics](#)
- [Author Guideline](#)
- [Article Processing Charges](#)
- [Online Submissions](#)
- [Copyright Notice](#)
- [Contact Us](#)

OPEN JOURNAL SYSTEMS

USER

You are logged in as...
dinaaulia03

- [My Journals](#)
- [My Profile](#)



UMSU
Teguh | Cerdas | Terpercaya

MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PESAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 83/SK/BA-PT/Akred/PT/06/2019
Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20236 Telp. (061) 6622400 - 66234567 Fax. (061) 6625474 - 6631003

Berita Acara Pembimbingan Proposal

Nama Mahasiswa : Dina Aulia Program Studi : Sistem Informasi
NPM : 2209010105
Nama Dosen : Mahardika Abdi Prawira Judul Penelitian : Implementasi Deteksi Berita Hoaks dan Fakta Menggunakan Metode Support Vector Machine (SVM) pada TurnBackHoax dan CNN Indonesia Berbasis Website
Pembimbing : Tanjung, S.Kom, M.Kom

Tanggal Bimbingan	Hasil Evaluasi	Paraf Dosen
18 / 11 / 2025	Revisi Bab 1	
24 / 11 / 2025	Revisi Bab 1	
1 / 12 / 2025	Revisi Bab 2	
5 / 12 / 2025	Revisi Bab 3	
8 / 12 / 2025	Revisi Bab 3	
10 / 12 / 2025	Revisi Flowchart	
12 / 12 / 2025	Revisi Struktur Dataset	
15 / 12 / 2025	ACC Sempro	





MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 8816/KUBAN-PT/Akred/PT/IB/2019
Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003

www.umsu.ac.id info@umsu.ac.id [umsuamedan](https://www.facebook.com/umsuamedan) [umsuamedan](https://www.instagram.com/umsuamedan) [umsuamedan](https://www.youtube.com/umsuamedan) [umsuamedan](https://www.linkedin.com/umsuamedan)

24 / 02 / 2026	Preview Website	A
27 / 02 / 2026	Revisi BAB IV	H
02 / 03 / 2026	Revisi BAB V	H
04 / 03 / 2026	ACC Sidang	H





UMSU
Unggul | Cerdas | Terpercaya

MAJLIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 86/SK/AN-PT/Akred/P1/02/2018

Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6621003

info@fak.umsu.ac.id

itk@fak.umsu.ac.id

[umsumedan](https://www.facebook.com/umsumedan)

[umsumedan](https://www.instagram.com/umsumedan)

[umsumedan](https://www.linkedin.com/company/umsumedan)

[umsumedan](https://www.youtube.com/channel/UC...)

Medan, 04 Maret 2026

Diketahui oleh :

Ketua Program Studi

(Mahardika Abdi Prawira Tanjung, S.Kom., M.Kom)

Disetujui oleh :

Dosen Pembimbing

(Mahardika Abdi Prawira Tanjung, S.Kom., M.Kom)





UMSU
Unggul | Cerdas | Berprestasi

MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PEMERINAN PURAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UMSU Terakreditasi Unggul Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 114/UKIDAN-PT/2019/2024

Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax (061) 6625474 - 6631003

Website: www.umsu.ac.id

Email: info@umsu.ac.id

Facebook: [umsuamedan](https://www.facebook.com/umsuamedan)

Instagram: [umsuamedan](https://www.instagram.com/umsuamedan)

Twitter: [umsuamedan](https://twitter.com/umsuamedan)

YouTube: [umsuamedan](https://www.youtube.com/umsuamedan)

FORMULIR PERBAIKAN UJIAN SKRIPSI

Pada hari ini, Jumat 13 Maret 2026 telah dilaksanakan Ujian Skripsi bagi mahasiswa Fakultas Ilmu Komputer Dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara. Sbb:

Nama Mahasiswa : Dina Aulia
NPM : 2209010105
Program Studi : Sistem Informasi
Judul Proposal : Implementasi Deteksi Berita Hoaks Dan Fakta Menggunakan Metode Support Vector Machine (SVM) Pada Turnbackhoax Dan CNN Indonesia Berbasis Website

Materi/Point yang Diperbaiki :

		Paraf
Yoshida Sary, S.Kom., M.Kom	Indikator Hoak (fact berdasarkan paper Bab 2 paper, BAB 3 yg akan dipalcat Bab 4 implementasi sistem.	
Martiano, S.Kom., M.Kom.	Referensi Mas - Tujuan penelitian. BAB 3, Cara analisis	
Mahardika Prawira Tanjung, S.Kom.,M.Kom.	N/A	

Berita acara ini **ditandatangani** setelah skripsi diperbaiki sesuai petunjuk/arahan dari Pembimbing dan Penguji/Pembahas.