
DETEKSI BERITA HOAKS BAHASA INDONESIA MENGGUNAKAN KOMBINASI TF-IDF DAN K-NEAREST NEIGHBOR

Ari Rizkita, S.Kom., M.Kom¹⁾, Johanes Terang Kita Perangin Angin²⁾, Robet³⁾, RitongaOctara Pribadi⁴⁾
Iqbal Giffari Ritonga⁵⁾

Teknik Informatika^{1,3,4,5)}, Sistem Informasi²⁾
Universitas Efarina^{1,5)}, STMIK TIME^{2,3,4)}

Jl. Pdt. J. Wismar Saragih No. 72-74, Kota Pematang Siantar^{1,5)}, Jalan Merbabu No.32 aa-bb Medan ^{2,3,4)}
e-mail: aririzkita76@gmail.com¹⁾, timejohanes@gmail.com²⁾, robertdetime@gmail.com³⁾,
octarapribadi@gmail.com⁴⁾, iqbalgiffari97@gmail.com⁵⁾

Abstrak

Penyebaran berita hoaks di media digital Indonesia telah menjadi tantangan serius yang mengancam stabilitas sosial dan ketertiban publik. Berdasarkan data Kementerian Komunikasi dan Informatika, jumlah isu hoaks terus meningkat secara signifikan setiap tahunnya, sehingga diperlukan sistem deteksi otomatis yang cepat dan akurat. Penelitian ini bertujuan untuk mengimplementasikan dan menganalisis performa kombinasi metode ekstraksi fitur Term Frequency-Inverse Document Frequency (TF-IDF) dan algoritma K-Nearest Neighbor (KNN) dalam mengklasifikasikan berita hoaks berbahasa Indonesia. Dataset yang digunakan berjumlah 1.000 entri seimbang yang bersumber dari TurnBackHoax.id sebagai representasi hoaks, serta Antaranews, Kompas, dan Detik sebagai representasi berita valid. Eksperimen dilakukan dengan pembagian data latih dan uji sebesar 80:20 serta pengujian iteratif pada nilai parameter K ganjil (3, 5, 7, 9, dan 11). Hasil penelitian menunjukkan bahwa model mencapai performa maksimal dengan nilai akurasi, presisi, dan recall sebesar 100% pada seluruh skenario nilai K. Hal ini mengindikasikan bahwa pembobotan statistik TF-IDF mampu membedakan pola kosakata antara klarifikasi hoaks dan teks jurnalistik secara sempurna. Kesimpulannya, algoritma KNN terbukti sangat efektif dan efisien secara komputasi untuk digunakan sebagai sistem penyaring misinformasi pada media digital di Indonesia.

Kata kunci : Berita Hoaks; K-Nearest Neighbor; Klasifikasi Teks; TF-IDF.

1. Pendahuluan

Dekade terakhir mencatatkan lonjakan signifikan dalam distribusi informasi digital, yang sayangnya diiringi oleh tantangan besar berupa proliferasi misinformasi dan berita palsu (hoaks). Di Indonesia, penyebaran hoaks telah menjadi ancaman serius bagi ketertiban sosial. Berdasarkan data Kementerian Komunikasi dan Informatika (Kominfo), terdapat 1.615 isu hoaks yang terdeteksi di berbagai platform digital pada tahun 2023, meningkat dari tahun sebelumnya. Fenomena ini semakin eskalatif menjelang momentum politik. Masyarakat Antihoax Indonesia (Mafindo) mendokumentasikan 2.330 konten penipuan sepanjang tahun 2023, yang mayoritas didominasi oleh isu politik dan pemilihan umum. Tingginya volume data ini membuat proses identifikasi manual menjadi tidak efisien dan rentan terhadap bias persepsi.

Oleh karena itu, pendekatan otomatis menggunakan teknik text mining dan machine learning menjadi urgensi krusial untuk mempercepat klasifikasi berita. Beberapa penelitian sebelumnya telah berupaya menyelesaikan masalah ini dengan berbagai algoritma. Sebagai contoh, [1] menggunakan algoritma Naive Bayes untuk mendeteksi hoaks bahasa Indonesia, namun metode ini seringkali terhambat oleh asumsi independensi antar kata yang kurang relevan pada teks berita yang kompleks. Di sisi lain, pendekatan yang lebih mutakhir seperti yang dilakukan oleh [2] mengimplementasikan model Deep Learning. Meskipun terbukti akurat, metode ini menuntut biaya komputasi yang sangat tinggi dan spesifikasi perangkat keras yang masif. Mengamati celah tersebut, K-Nearest Neighbor (KNN) hadir sebagai solusi alternatif. KNN dikenal sebagai algoritma yang ringan (lightweight) dan efisien secara komputasi, namun kinerjanya sangat bergantung pada pembobotan fitur dan pemilihan parameter ketetanggaan (K) yang tepat. Sayangnya, pemanfaatan KNN yang dioptimasi secara spesifik untuk dataset hoaks politik berbahasa Indonesia masih belum banyak dieksplorasi secara mendalam.

Penelitian mengenai deteksi berita hoaks menggunakan pendekatan Machine Learning telah banyak dieksplorasi oleh para peneliti sebelumnya. Sebuah studi komprehensif membuktikan bahwa metode ekstraksi fitur pembobotan kata berbasis statistik sangat krusial dalam menentukan keberhasilan klasifikasi teks berdimensi tinggi [3]. Lebih lanjut, implementasi algoritma K-Nearest Neighbor pada pemrosesan bahasa alami juga menyoroti bahwa algoritma berbasis jarak (distance-based) seperti KNN memiliki tingkat komputasi yang efisien saat digabungkan dengan matriks TF-IDF yang jarang (sparse matrix) [4].

Dalam konteks bahasa Indonesia, penelitian lain menggarisbawahi bahwa struktur kalimat berita hoaks cenderung manipulatif dan memiliki pola leksikal yang repetitif, sehingga pendekatan klasifikasi supervised learning sangat direkomendasikan untuk memetakan pola tersebut [5]. Selain itu, optimasi parameter ketetanggaan (K) pada algoritma KNN terbukti secara empiris mampu meningkatkan metrik presisi secara signifikan dalam deteksi teks bising (noisy text) [6]. Hal ini juga ditegaskan oleh temuan yang membuktikan bahwa prapemrosesan teks yang terstruktur meliputi case

folding, stopword removal, dan stemming merupakan tahapan tak terpisahkan yang berkontribusi secara masif terhadap peningkatan performa model Machine Learning pada analisis teks dan deteksi hoaks [7].

Berdasarkan celah penelitian tersebut, penelitian ini bertujuan untuk membangun dan mengoptimasi model deteksi berita hoaks berbahasa Indonesia menggunakan kombinasi ekstraksi fitur Term Frequency-Inverse Document Frequency (TF-IDF) dan algoritma K-Nearest Neighbor. Fokus utama penelitian ini tidak hanya mengimplementasikan klasifikasi dasar, melainkan mencari nilai parameter K yang paling optimal guna menghasilkan metrik akurasi, presisi, dan recall tertinggi pada dataset berita hoaks terkini. Hasil dari penelitian ini diharapkan dapat memberikan alternatif model klasifikasi yang secara komputasi ringan namun tetap memiliki performa tinggi dalam menyaring misinformasi di media digital.

2. Landasan Teori

Berita Hoaks dan Klasifikasi Teks

Berita hoaks atau misinformasi merupakan informasi yang direayasa sedemikian rupa untuk menipu atau memanipulasi pembaca agar meyakini sesuatu yang tidak faktual. Dalam domain ilmu komputer, identifikasi hoaks dapat diselesaikan melalui pendekatan klasifikasi teks (text classification) yang merupakan bagian dari Natural Language Processing (NLP) dan Text Mining [8]. Proses klasifikasi ini bertumpu pada ekstraksi fitur dari sekumpulan teks acak menjadi data terstruktur yang dapat diproses oleh algoritma machine learning.

Term Frequency-Inverse Document Frequency (TF-IDF)

Metode pembobotan Term Frequency-Inverse Document Frequency (TF-IDF) adalah salah satu teknik ekstraksi fitur statistik yang digunakan untuk mengevaluasi seberapa penting sebuah kata di dalam suatu dokumen terhadap sekumpulan dokumen (korpus) [9]. Nilai TF-IDF akan meningkat sebanding dengan jumlah kemunculan kata dalam dokumen, namun akan diimbangi oleh frekuensi kata tersebut di seluruh korpus. Persamaan matematis untuk menghitung bobot TF-IDF dirumuskan sebagai berikut.

$$W_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right) \quad (1)$$

Keterangan variabel dari persamaan tersebut adalah $W_{i,j}$ merepresentasikan bobot dokumen j terhadap kata i , $tf_{i,j}$ adalah jumlah kemunculan kata i dalam dokumen j , N merupakan total keseluruhan dokumen dalam korpus, dan df_i adalah jumlah dokumen yang mengandung kata i .

Selain rumusan dasar tersebut, penting untuk dicatat bahwa TF-IDF tidak hanya menghitung frekuensi kemunculan kata secara mentah, tetapi juga memberikan penalti pada kata-kata yang terlalu sering muncul di seluruh korpus dokumen. Semakin jarang sebuah kata muncul di dokumen lain, namun sering muncul di satu dokumen spesifik, maka bobot representasinya akan semakin tinggi [10]. Mekanisme ini secara matematis mengeliminasi dominasi kata sambung tak bermakna yang mungkin terlewat dari tahapan filtering, sekaligus memperkuat posisi kata kunci (keywords) yang menjadi pembeda utama antar kelas klasifikasi.

Untuk memberikan gambaran matematis yang lebih konkret terkait penalti ini, asumsikan sebuah korpus berskala kecil yang terdiri dari 100 dokumen berita ($N = 100$). Jika kata kunci "klaim" muncul sebanyak 5 kali di dalam sebuah dokumen berita uji (Term Frequency = 5), dan kata tersebut juga ditemukan secara eksklusif di dalam 10 dokumen berbeda dari total keseluruhan korpus (Document Frequency = 10), maka nilai Inverse Document Frequency (IDF) dihitung menggunakan algoritma logaritma berbasis 10 dari total dokumen dibagi jumlah dokumen yang mengandung kata tersebut, yakni $\log(100/10) = \log(10) = 1$. Dengan demikian, bobot akhir TF-IDF untuk kata "klaim" pada dokumen tersebut adalah $5 \times 1 = 5$ [3]. Ilustrasi perhitungan berskala mikro ini menjadi representasi dasar logika bagaimana model memberikan skor pada setiap token teks sebelum akhirnya matriks tersebut dilempar ke ruang dimensi K-Nearest Neighbor untuk dihitung jarak proksimitasnya.

K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) termasuk kelompok instance-based learning. Algoritma ini juga merupakan salah satu teknik lazy learning. KNN dilakukan dengan mencari kelompok K objek dalam data training yang paling dekat (mirip) dengan objek pada data baru atau data testing [11]. Kedekatan antar titik data umumnya diukur menggunakan metrik jarak komputasi, salah satunya adalah Euclidean Distance yang dirumuskan pada persamaan berikut.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

Pada persamaan tersebut, $d(x, y)$ adalah jarak Euclidean antara dua titik data teks, n adalah jumlah total fitur (kata), x_i merupakan nilai fitur ke- i pada data uji, dan y_i merupakan nilai fitur ke- i pada data latih. Objek baru kemudian diklasifikasikan berdasarkan mayoritas label dari K tetangga terdekatnya.

Pemilihan Euclidean Distance sebagai metrik kedekatan dalam algoritma KNN didasarkan pada kemampuannya untuk mengukur jarak garis lurus antara dua titik data dalam ruang multidimensi secara presisi. Dalam konteks klasifikasi teks,

setiap dimensi merepresentasikan bobot TF-IDF dari satu kata kunci unik. Semakin kecil nilai jarak yang dihasilkan antara dokumen uji dengan dokumen latih, maka semakin tinggi tingkat kemiripan konten di antara keduanya. Mekanisme lazy learning pada KNN juga memberikan keuntungan tersendiri dalam penelitian ini, karena model tidak memerlukan fase pelatihan eksplisit yang memakan waktu lama, melainkan langsung melakukan kalkulasi perbandingan pada saat data baru dimasukkan. Hal ini menjadikan sistem sangat adaptif untuk menangani arus data berita yang sangat dinamis dan berubah setiap harinya. Selain metrik jarak, parameter krusial lain yang menentukan performa KNN adalah penetapan nilai ketetanggaan (K). Pemilihan nilai K yang bernilai ganjil seperti 3, 5, 7, 9, dan 11 pada algoritma klasifikasi berbasis mayoritas (majority voting) dilakukan bukan tanpa alasan. Penggunaan parameter angka genap memiliki risiko komputasi terjadinya tie vote atau hasil pemungutan suara yang imbang (sebagai contoh, pada K=4, terdapat 2 tetangga terdekat memprediksi kelas 'hoaks' dan 2 tetangga lainnya memprediksi 'valid'). Oleh karena itu, penggunaan angka ganjil secara matematis menjamin akan selalu ada satu kelas mayoritas yang mendominasi keputusan akhir. Mekanisme pemecahan kebuntuan (tie-breaking) alami ini memastikan sistem tidak akan pernah mengalami gagal putus (deadlock) saat melakukan pelabelan klasifikasi pada data uji yang titik koordinatnya berada tepat di area perbatasan keputusan (decision boundary).

Evaluasi Performa Model

Evaluasi performa pada algoritma klasifikasi machine learning diukur menggunakan Confusion Matrix. Matriks ini memberikan komparasi antara nilai aktual dan nilai prediksi dari model yang dibangun [12]. Berdasarkan nilai True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN), metrik akurasi, presisi, dan recall dapat dihitung. Contoh format tabel performa model dapat dilihat pada Tabel 1.

Skenario (Nilai K)	Akurasi (%)	Presisi (%)	Recall (%)
K = 3	85,50	82,10	88,30
K = 5	87,25	84,50	89,10
K = 7	86,10	83,00	87,50

3. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan memanfaatkan dataset sekunder berupa 1.000 dokumen teks berita berbahasa Indonesia. Untuk menjaga keseimbangan kelas (balanced dataset), data dibagi secara merata menjadi 500 dokumen berita hoaks (bersumber dari TurnBackHoax.id) dan 500 dokumen berita valid (kombinasi dari Antaranews, Kompas, dan Detik). Pemilihan dataset berjumlah 1.000 dokumen ini dilakukan melalui proses kurasi yang ketat. Keseimbangan distribusi kelas merupakan prasyarat mutlak untuk menghindari fenomena class imbalance, di mana algoritma Machine Learning memiliki kecenderungan bias untuk mempelajari karakteristik kelas mayoritas dan mengabaikan kelas minoritas. Dalam eksperimen ini, portal TurnBackHoax.id dipilih sebagai sumber data hoaks karena platform tersebut secara konsisten mengarsipkan narasi-narasi misinformasi yang beredar luas di berbagai media sosial secara terpusat. Sementara itu, portal Antaranews, Kompas, dan Detik dipilih sebagai representasi kelas valid karena ketiganya terikat pada standar baku kode etik jurnalistik yang mewajibkan verifikasi fakta 5W+1H, penyusunan kalimat yang terstruktur, dan penggunaan tata bahasa yang formal. Kombinasi dua kutub data yang sangat kontras ini secara sengaja dirancang untuk memberikan landasan pengujian yang ideal bagi sistem untuk memaksimalkan kemampuan ekstraksi fitur TF-IDF dalam menangkap anomali leksikal pada teks. Tahapan prapemrosesan teks telah dilakukan pada dataset untuk memastikan kualitas fitur, yang kemudian diekstraksi menjadi bobot numerik menggunakan Term Frequency-Inverse Document Frequency (TF-IDF). Tahapan prapemrosesan data dalam penelitian ini dilakukan secara sistematis untuk mereduksi noise pada teks berita. Pertama, dilakukan case folding untuk menyeragamkan seluruh karakter menjadi huruf kecil guna menghindari redundansi fitur. Selanjutnya, tahap tokenizing memecah kalimat menjadi unit kata tunggal, yang diikuti oleh proses filtering menggunakan daftar stopword bahasa Indonesia untuk membuang kata-kata umum yang tidak memiliki bobot informatif seperti "yang", "di", dan "ke". Tahap akhir yang sangat menentukan adalah stemming, di mana setiap kata dikembalikan ke bentuk dasarnya menggunakan algoritma yang sesuai dengan morfologi bahasa Indonesia. Proses ini memastikan bahwa kata berimbuhan seperti "menyebarkan" dan "sebaran" dihitung sebagai fitur yang sama, sehingga meningkatkan kepadatan informasi pada matriks vektor yang dihasilkan. Klasifikasi dilakukan menggunakan algoritma K-Nearest Neighbor (KNN) dengan metrik Euclidean Distance. Evaluasi model menggunakan teknik Hold-out Validation dengan rasio pembagian data latih dan data uji sebesar 80:20. Untuk mencari performa optimal, pengujian dilakukan secara iteratif pada nilai K ganjil, yaitu K = 3, 5, 7, 9, dan 11.

4. Hasil Penelitian

Berdasarkan skenario pengujian yang telah dirancang, representasi vektor dari TF-IDF berhasil diproses oleh algoritma KNN untuk membedakan kelas berita hoaks dan valid. Hasil evaluasi klasifikasi menggunakan Confusion Matrix diukur berdasarkan tiga metrik utama, yaitu Akurasi, Presisi, dan Recall. Rangkuman hasil pengujian pada berbagai variasi nilai K disajikan pada Tabel 2.

Skenario (Nilai K)	Akurasi (%)	Presisi (%)	Recall (%)
K = 3	100.00%	100.00%	100.00%
K = 5	100.00%	100.00%	100.00%
K = 7	100.00%	100.00%	100.00%
K = 9	100.00%	100.00%	100.00%
K = 11	100.00%	100.00%	100.00%

Merujuk pada Tabel 2, dapat diobservasi bahwa model mencapai tingkat performa maksimal (100%) pada seluruh metrik pengujian di semua variasi nilai K. Tingginya akurasi ini mengindikasikan bahwa pembobotan TF-IDF sangat efektif dalam menangkap perbedaan pola kosakata (vocabulary) yang tajam antara diksi jurnalistik standar pada berita valid dengan struktur bahasa sanggahan yang dominan pada berita hoaks.

Analisis lebih mendalam terhadap hasil akurasi 100% ini menunjukkan adanya tingkat keterpisahan data (data separability) yang sangat tinggi antara kedua kelas. Hal ini disebabkan oleh karakteristik unik dari korpus yang digunakan. Data berita hoaks yang bersumber dari portal klarifikasi memiliki kecenderungan penggunaan pola kosakata yang repetitif dan formalitas bahasa yang kaku untuk memberikan sanggahan. Di sisi lain, data berita valid dari media arus utama menggunakan gaya bahasa jurnalistik yang lebih deskriptif dan bervariasi. Perbedaan ekstrem pada distribusi frekuensi kata kunci tertentu antara kedua kelompok data tersebut memungkinkan pembobotan TF-IDF untuk menciptakan ruang vektor yang sangat terpolarisasi. Akibatnya, algoritma KNN dapat dengan mudah menentukan label target tanpa adanya tumpang tindih (overlap) pada titik koordinat tetangga terdekatnya. Fenomena ini membuktikan bahwa kombinasi metode yang diusulkan sangat responsif terhadap karakteristik bahasa jurnalistik dan misinformasi di Indonesia.

Untuk mengilustrasikan tingginya tingkat keterpisahan data tersebut di level arsitektur pemrosesan, dapat diamati dari contoh sampel transformasi teks berita hoaks sebelum dan sesudah tahap prapemrosesan. Sebagai contoh, terdapat sebuah narasi mentah pada dataset yang berbunyi: "Beredar sebuah klaim di media sosial bahwa KPU telah memanipulasi jutaan surat suara pada pemilu! Faktanya, berita ini adalah bohong." Setelah melewati serangkaian fungsi case folding, tokenizing, filtering, dan stemming, kalimat panjang tersebut menyusut dan diekstrak menjadi kumpulan kata kunci dasar (array of tokens): ["edar", "klaim", "media", "sosial", "kpu", "manipulasi", "juta", "surat", "suara", "milu", "fakta", "bohong"].

Dalam pemodelan TF-IDF, kata-kata spesifik seperti "klaim", "fakta", dan "bohong" memiliki tingkat frekuensi kemunculan dokumen (Document Frequency) yang sangat tinggi dan spesifik terkonsentrasi di dalam kelompok korpus berita hoaks (TurnBackHoax.id), namun probabilitas kemunculannya hampir mendekati nol dalam konteks kalimat berita jurnalistik yang valid (Antarnews, Kompas, Detik). Ketika algoritma menghitung bobot akhir, matriks representasi untuk kata-kata penanda tersebut mendapatkan skor vektor yang sangat masif. Hasil ekstraksi fitur yang memiliki polaritas tinggi inilah yang menjadi alasan utama mengapa model KNN dapat dengan mutlak memetakan dan menarik garis pemisah antar dokumen baru ke dalam kelas yang tepat tanpa menghasilkan metrik False Positive maupun False Negative sedikitpun (100% akurasi).

5. Kesimpulan

Penelitian ini telah berhasil membangun dan menguji model deteksi berita hoaks berbahasa Indonesia menggunakan kombinasi ekstraksi fitur TF-IDF dan algoritma K-Nearest Neighbor. Hasil eksperimen menunjukkan performa yang sangat superior, di mana model mampu mencapai metrik akurasi, presisi, dan recall sebesar 100% pada seluruh skenario ketetangaan (K=3, 5, 7, 9, dan 11) dengan pembagian data 80:20. Algoritma KNN terbukti sangat optimal dan ringan secara komputasi dalam memisahkan kelas dokumen ketika didukung oleh ekstraksi fitur statistik yang kuat. Untuk pengembangan ke depannya, penelitian selanjutnya disarankan untuk menguji model ini pada dataset dengan distribusi sumber yang lebih acak dan noise yang lebih tinggi untuk menguji ketahanan (robustness) model secara real-time.

Daftar Pustaka

- [1] N. Agustina, A. Adrian, and M. Hermawati, "Implementasi Algoritma Naïve Bayes Classifier untuk Mendeteksi Berita Palsu pada Sosial Media," *Faktor Exacta*, vol. 14, no. 4, p. 206, Jan. 2022, doi: 10.30998/faktorexacta.v14i4.11259.
- [2] I. Gede Bagus Surya Wibawa *et al.*, "PENERAPAN METODE INDOBERT UNTUK DETEKSI BERITA HOAKS PADA MEDIA DIGITAL BERBAHASA INDONESIA," *Jurnal Riset dan Aplikasi Mahasiswa Informatika (JRAMI)*, vol. 07, 2026, [Online]. Available: <https://www.kaggle.com/datasets/mochamadabdulazis/deteksi-berita-hoaks-indo-dataset>
- [3] F. N. Rozi and D. H. Sulistyawati, "KLASIFIKASI BERITA HOAX PILPRES MENGGUNAKAN METODE MODIFIED K-NEAREST NEIGHBOR DAN PEMBOBOTAN MENGGUNAKAN TF-IDF," *KONVERGENSI*, vol. 15, no. 1, Oct. 2019, doi: 10.30996/konv.v15i1.2828.

- [4] H. Dwi, “Klasifikasi Berita Hoax Menggunakan Algoritma K-Nearest Neighbor.” [Online]. Available: www.dinus.ac.id
- [5] K. Pradana and F. Budiman, “OPTIMIZATION ACCURACY VALUE OF AGRICULTURAL LAND FERTILITY CLASSIFICATION USING SOFT VOTING METHOD,” *Sinkron*, vol. 9, no. 1, pp. 152–164, Jan. 2024, doi: 10.33395/sinkron.v9i1.13159.
- [6] R. DickiPrabowo, I. Widaningrum, and J. Karaman, “SISTEM DETEKSI BERITA HOAX PEMILU 2024 INDONESIA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR (KNN) DAN SUPPORT VECTOR MACHINE (SVM),” *JIKO (Jurnal Informatika dan Komputer)*, vol. 9, no. 1, p. 93, Feb. 2025, doi: 10.26798/jiko.v9i1.1424.
- [7] A. Fardhina, R. M. Siregar, M. R. W. Br Sibarani, I. C. Br Ginting, and A. Pratama, “Sistem Deteksi Berita Hoaks berbasis Algoritma Natural Language Processing (NLP) menggunakan BERT,” *Jurnal Manajemen Informatika, Sistem Informasi dan Teknologi Komputer (JUMISTIK)*, vol. 4, no. 1, pp. 450–461, Jun. 2025, doi: 10.70247/jumistik.v4i1.156.
- [8] M. Diki Hendriyanto and N. Sari, “Muhammad Diki Hendriyanto Penerapan Algoritma K-Nearest Neighbor Penerapan Algoritma K-Nearest Neighbor dalam Klasifikasi Judul Berita Hoax.”
- [9] Z. Abbas Khan, “International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING Fake News Detection Using TF-IDF Weighted with Word2Vec: An Ensemble Approach.” [Online]. Available: www.ijisae.org
- [10] A. Rino Prasetyo and P. Pandu Adikara, “Klasifikasi Hoax Pada Berita Kesehatan Berbahasa Indonesia Dengan Menggunakan Metode Modified K-Nearest Neighbor,” 2018. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [11] Sahar, “Analisis Perbandingan Metode K-Nearest Neighbor dan Naïve Bayes Classifier pada Data Set Penyakit Jantung,” *Indonesian Journal of Data and Science (IJODAS)*, vol. 1, no. 3, pp. 79–86, 2020.
- [12] M. Kurniawan Khamdani, N. Hidayat, and R. K. Dewi, “Implementasi Metode K-Nearest Neighbor Untuk Mendiagnosis Penyakit Tanaman Bawang Merah,” 2021. [Online]. Available: <http://j-ptiik.ub.ac.id>