

## Klasifikasi Berita Hoaks Pasca Pemilihan Umum Presiden Dan Wakil Presiden Republik Indonesia Tahun 2024 Menggunakan Algoritma *K-Nearest Neighbour*

Dennis Fitri Salsabilla Arianti<sup>1</sup>, Ulfi Saidata Aesy<sup>2\*</sup>, Arif Himawan<sup>3</sup>

<sup>1,2,3</sup>Program Studi Sistem Informasi, Fakultas Teknik dan Teknologi Informasi, Universitas Jenderal Achmad Yani Yogyakarta

### Article Info

#### Article history:

Received September 25, 2024

Accepted February 18, 2025

Published May 20, 2025

#### Keywords:

*K-Nearest Neighbour*

*TF-IDF*

*Classification News*

*Election 2024*

### ABSTRACT

*In the 2024 election, widely celebrated as a festival of democracy, the spread of hoax news through social media has become a significant concern. A total of 203 hoax issues were identified, resulting in the spread of 2,882 hoax contents across various websites. A survey by Katadata Insight Center (KIC) revealed that between 30% to 60% of Indonesians were exposed to such hoaxes, with only 21% to 36% of them able to accurately discern whether the news was factual or a hoax. This research aims to develop a model using the K-Nearest Neighbour (KNN) algorithm to evaluate the accuracy and reliability of news from various Indonesian news websites through machine learning-based classification techniques. By categorizing news into potentially factual and potentially hoax categories, this study seeks to provide insights into which news pages are more likely to present factual information. The method employed involves Term Frequency-Inverse Document Frequency (TF-IDF) word weighting combined with KNN modeling, using Euclidean distance for calculation. The KNN model achieved a training accuracy of 0.90 and a testing accuracy of 0.88. The findings indicate that news sites such as detik.com, okezone.com, liputan6.com, and cnnindonesia.com demonstrate higher accuracy in the news they present. The use of the KNN method successfully identified probabilities of factual and hoax potential across news websites, including kompas.com, liputan6.com, detiknews.com, antaranews.com, cnnindonesia.com, okezone.com, sindonews.com, kumparan.com, pikiranrakyat.com, and wartatransparansi.com.*



### Corresponding Author:

Ulfi Saidata Aesy,

Program Studi Sistem Informasi, Fakultas Teknik dan Teknologi Informasi

Universitas Jenderal Achmad Yani Yogyakarta,

Jl. Siliwangi, Ringroad Barat, Banyuraden, Gamping, Sleman, Daerah Istimewa Yogyakarta 55293

Email: \*ulfiaesy@gmail.com

## 1. PENGANTAR

Indonesia adalah salah satu negara demokrasi yang mana setiap warga negara memiliki hak yang sama untuk mengambil keputusan dalam pemerintahan. Salah satu cara Indonesia menerapkan sistem demokrasi yaitu dengan mengadakan Pemilihan Umum[1]. Seiring dengan berakhirnya masa jabatan Presiden dan Wakil Presiden Republik Indonesia yang ketujuh maka dilaksanakanlah pemilu yang rutin dilakukan setiap 5 tahun sekali untuk menentukan Presiden dan Wakil Presiden selanjutnya. Pada pemilu tahun 2024 ini dikenal sebagai pesta demokrasi sehingga terjadi banyak penyebaran berita hoaks di media sosial[2].

Berdasarkan data yang diperoleh dari Google Trends, pencarian kata terkait pemilu masih meningkat pesat hingga 3000% hal ini menunjukkan bahwa masyarakat masih antusias mencari informasi mengenai

pemilu melalui website[3]. Sebagai media penyampaian berita ternyata terdapat terdapat 203 isu hoaks dengan total sebaran 2.882 konten di website[4]. Berdasarkan survei yang telah dilakukan Katadata Insight Center (KIC) yang berkerja sama dengan Kementerian Komunikasi dan Informatika, hasilnya sebanyak 30% hingga 60% orang Indonesia yang terpapar hoaks. Namun hanya 21% sampai 36% saja yang mampu membedakan apakah berita tersebut hoaks atau tidak [5].

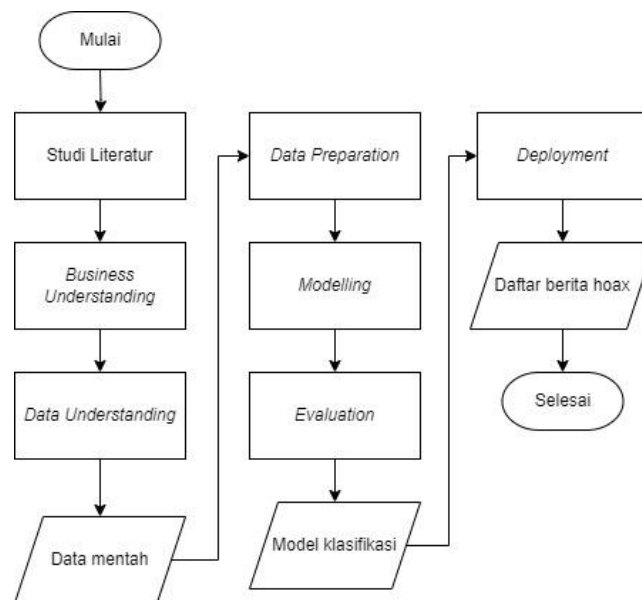
Berkaitan dengan hal tersebut, diperlukan model machine learning untuk melakukan klasifikasi berita hoaks pasca pemilu 2024 dengan bantuan algoritma K-Nearest Neighbour (KNN). Prinsip dari metode klasifikasi menggunakan KNN adalah prinsip perbandingan jarak antara data yang belum diketahui dengan data pelatihan yang telah ada [6]. KNN memiliki keuntungan dalam text mining, seperti mampu menangani dataset besar, dikombinasikan dengan teknik lain, dan memiliki potensi untuk menangani data noise, dan mengekstraksi informasi yang lebih komprehensif [7].

*Term Frequency-Inverse Document Frequency* (TF-IDF) akan penulis terapkan untuk meningkatkan akurasi dengan mengukur seberapa penting suatu kata dalam sebuah dokumen guna mengetahui potensi hoaks dari suatu berita sehingga mampu meningkatkan kemampuan model dalam memahami konteks dan pola tertentu [8]. Kombinasi KNN dan TF-IDF ini memungkinkan model untuk lebih akurat dalam mengidentifikasi berita hoaks dengan mempertimbangkan pola kata yang signifikan dalam teks. Pengambilan data dalam penelitian ini dengan melakukan scraping website turnbackhoax.id, kominfo.go.id, infopublik.id, dan rri.co.id sebagai data yang menjadi acuan pembelajaran mesin serta melakukan scraping website kompas.com, liputan6.com, detiknews.com, antaranews.com, cnnindonesia.com, okezone.com, sindonews.com, kumparan.com, pikiranrakyat.com, dan wartatransparansi.com. Pengembangan model ini meliputi akuisisi data, preprocessing, pemilihan algoritma, pelatihan dan evaluasi model [9]. Hasil dari penelitian memberikan informasi untuk mengetahui daftar berita yang memiliki potensi hoaks selama pasca pemilu 2024.

Hasil penelitian ini diharapkan dapat mengevaluasi situs berita yang memiliki potensi fakta yang besar. Penggunaan algoritma KNN dan pembobotan kata menggunakan TF-IDF, penelitian ini diharapkan dapat mengklasifikasikan berita yang berpotensi hoaks untuk menghadapi tantangan informasi yang tidak akurat.

## 2. METODE PENELITIAN

Penelitian ini menggunakan data dari hasil *scraping* dari sepuluh media berita dengan judul berita yang berkaitan dengan pemilu. Tahapan penelitian dapat dilihat pada gambar 1.



Gambar 1. Tahapan penelitian

### a. *Business Understanding*

Pada tahapan pemahaman bisnis dilakukan penentuan tujuan proyek dan menentukan batasan penelitian [10]. Selain itu, menentukan faktor apa saja yang mempengaruhi klasifikasi berita hoaks pasca pemilu serta memeriksa apakah ada keterkaitan antar atribut. Hal ini untuk dijadikan acuan untuk mengklasifikasi berita hoaks pasca pemilu pada tahap-tahap selanjutnya.

*b. Data Understanding*

Pada tahap ini dilakukan pengumpulan data melalui *scraping website* turnbackhoax, Kominfo, infopublik (pemilu, politik, ekonomi, sosial dan budaya), RRI, kompas, liputan6, detiknews, antaranews, cnnindonesia, okezone, sindonews, kumparan, pikiranrakyat, dan wartatransparansi.com. Data berita yang ada pada *website* turnbackhoax.id dan kominfo.go.id sebagian besar digunakan untuk data label “berpotensi hoaks”, sedangkan data berita dari infopublik.id dan rri.co.id sebagian besar akan diberi label “berpotensi fakta”. Proses *scraping* menggunakan *library requests*, *BeautifulSoup*, dan *extension Instant Data Scraper* [11].

*c. Data Preparation*

Menyiapkan data untuk siap dilakukan pemodelan, yaitu dengan melakukan normalisasi, *text preprocessing* seperti *case folding*, *replace texts*, *cleaning*, *filtering*, *stemming*, dan *tokenizing* [12]. Pada tahap ini juga dilakukan pembobokan kata menggunakan TF-IDF.

*d. Modelling*

Hasil dari TF-IDF yang telah dibuat, selanjutnya diproses untuk pembuatan model menggunakan algoritma KNN yang akan digunakan dalam memprediksi berita yang berpotensi hoaks [13].

*e. Evaluation*

Selanjutnya mengevaluasi model yang sudah dibuat pada tahap pemodelan dengan menggunakan *confusion matrix* untuk menilai apakah *accuracy*, *precision*, *recall*, dan *F1-score* yang dicapai model sudah baik. Terdapat beberapa metrik evaluasi yang termasuk dalam *classification report* [14]:

- 1) *Precision*, mengukur tingkat keakuratan prediksi positif dari model.
- 2) *Recall*, mengukur seberapa baik model dapat menangkap instance positif.
- 3) *F1-score*, rata-rata harmonik dari *precision* dan *recall*. *F1-score* memberikan keseimbangan antara *precision* dan *recall*.
- 4) *Accuracy*, adalah rasio antara jumlah prediksi yang benar dengan total jumlah prediksi yang dibuat oleh model.

*f. Deployment*

Penerapan model pada data yang belum diketahui labelnya. Hasilnya berupa daftar berita terkait pemilu yang berpotensi fakta dan berpotensi hoaks. Pada tahap ini model diterapkan pada hasil *scraping* dari laman berita kompas.com, liputan6.com, detiknews.com, antaranews.com, cnnindonesia.com, okezone.com, sindonews.com, kumparan.com, pikiranrakyat.com, dan wartatransparansi.com.

**3. HASIL DAN ANALISIS**

Pada bagian ini adalah pembahasan hasil penelitian berdasarkan tahap penelitian yang ditunjukkan gambar 1. Pada *business understanding* dilakukan penentuan laman berita yang akan dijadikan *data training*. Laman yang dijadikan *data training* yaitu turnbackhoax.id, kominfo.go.id, infopublik.id, dan rri.co.id Alasan menggunakan laman berita turnbackhoax.id karena *website* ini telah diakui sebagai Komunitas Online Anti Hoax di Indonesia[15]. Selain itu alasan memilih *website* infopublik.id sebagai *data training* karena Dirjen IKP Kementerian Kominfo, InfoPublik dijadikan sebagai kantor dan media berita atas nama pemerintah sehingga menjadi penghubung untuk menyebarkan informasi ke publik [16]. Pemilihan *website* rri.co.id karena RRI adalah satu-satunya radio nasional yang resmi dimiliki oleh pemerintah Indonesia[17]. Kolom yang diperlukan pada penelitian hanya kolom judul sehingga hanya perlu mengambil tag yang menampilkan judul pada laman berita.

Tahap selanjutnya adalah *data understanding*. Pada tahap ini dilakukan pengumpulan data dengan melakukan *scraping website* turnbackhoax.id, kominfo.go.id (cek fakta), infopublik.id (pemilu, politik, ekonomi, sosial dan budaya), rri.co.id, kompas.com, liputan6.com, detiknews.com, antaranews.com, cnnindonesia.com, okezone.com, sindonews.com, kumparan.com, pikiranrakyat.com dan wartatransparansi.com Dari hasil *scraping*, jumlah data yang digunakan untuk pembuatan model sebanyak 1554 data. Pada tabel 1, ditunjukkan sampel data dari hasil *scraping*.

Tabel 1. Sampel Data Hasil *Scraping* [18]

Judul
Pilgub Banten Nihil dari Jalur Perseorangan
Pasangan Calon Jalur Perseorangan Ramaikan Pilkada Kabupaten Tangerang
Wanti-wanti Hilang, Bawaslu Minta Laporan Pengawasan LN Pemilu
Caleg Terpilih Pemilu, Tak Wajib Mundur di Pilkada
KPU-Bawaslu Kota Tangerang Kesulitan Penuhi Kuota Petugas Adhoc
PDIP Godok 11 Nama Balon Pilkada Kota Tangerang

MK Diminta Tolak Tudingan Kecurangan Suara Dapil Jabar  
 KPU Beberkan Alasan Pecat 13 PPD  
 Sidang PHPU Pileg 2024, MK Candai Kekalahan MU  
 Ini Program 100 Hari Pertama Gibran Usai Dilantik

Selanjutnya dilakukan *preprocessing data*. Tahap Ini digunakan untuk memperbaiki data agar komputer dapat membaca bahasa manusia. *Preprocessing data* terdapat berbagai tahapan yaitu *case folding*, *replace texts*, *cleaning*, *filtering*, *stemming*, dan *tokenizing*. Pada tabel 2, ditunjukkan sampel data dari hasil *preprocessing*.

Tabel 2. Sampel Data Hasil *Preprocessing*

Judul Asli	Case Folding	Replace Texts	Cleaning	Filtering	Stemming	Tokenizing
Timnas AMIN	timnas amin	timnas anies	timnas anies	timnas anies	timnas anies	['timnas', 'anies',
Daftarkan	daftarkan	muhammad	muhammad	muhammad	muhammad	'muhammad',
Gugatan	gugatan	daftarkan	daftarkan	daftarkan	muhammad	'daftar',
Hasil Pemilu 2024 ke MK [12]	hasil pemilu 2024 ke mk	gugatan	gugatan	gugatan	daftar gugat	'gugat', 'hasil', 'pemilu', '2024', 'mk']

Setelah data bersih, dilakukan proses TF-IDF untuk menghitung bobot setiap kata. TF-IDF adalah metode statistik yang digunakan untuk mengevaluasi pentingnya sebuah kata dalam teks [19]. Keluaran dari kode ini berupa TF-IDF matrix yang disimpan dalam format csv dan TF-IDF *vectorizer* dalam format.pkl untuk digunakan untuk memproses teks sehingga dapat menerapkan vektorisasi yang sama ke teks baru tanpa harus memfit ulang *vectorizer* [20]. Berikut tahapan perhitungan dengan TF-IDF [21]:

1. Hitung term frequency

$$tf_{t,j} = \frac{n_{i,j}}{\sum_k n_{i,j}} \quad (1)$$

Keterangan:

$tf_{t,j}$  : frekuensi term  
 $n_{i,j}$  : banyak kata i pada dokumen j  
 $k$  : Index dokumen

2. Hitung bobot *inverse document frequency* (idf)

$$idf = \log \frac{N}{df_t} \quad (2)$$

Keterangan :

$idf$  : *Inverse Document Frequency*  
 $N$  : jumlah total dokumen  
 $df_t$  : jumlah frekuensi dokumen yang mengandung *term*

3. Hitung nilai bobot TF-IDF

$$W_{i,j} = tf_{t,j} \times idf \quad (3)$$

Keterangan:

$W_{i,j}$  : bobot TF-IDF  
 $tf_{t,j}$  : frekuensi term  
 $idf$  : *Inverse Document Frequency*

Hasil dari TF-IDF yang telah dibuat, selanjutnya diproses untuk pembuatan model menggunakan algoritma KNN yang akan digunakan dalam memprediksi berita yang berpotensi hoaks. Langkah-langkah untuk menghitung metode KNN antara lain [22]:

1. Menentukan parameter K
2. Menghitung jarak antara *data training* dan *data testing*

Cara menghitung jarak ketetanggaan dapat menggunakan algoritma euclidean seperti yang ditunjukkan pada persamaan (4).

$$euclidean = \sqrt{((a_1 - b_1)^2 + \dots + (a_n - b_n)^2)} \tag{4}$$

Keterangan:

$a_1 \dots a_n$  : komponen – komponen dari vektor a

$b_1 \dots b_n$  : komponen - komponen dari vektor b

Pada persamaan (4),  $(a_1 - b_1)^2$  adalah kuadrat selisih antara koordinat pertama dari titik pertama dan titik kedua. proses ini diulangi untuk setiap dimensi, dan hasilnya dijumlahkan, akar kuadrat dari jumlah kuadrat selisih ini memberikan jarak Euclidean antara dua titik dalam ruang n dimensi.

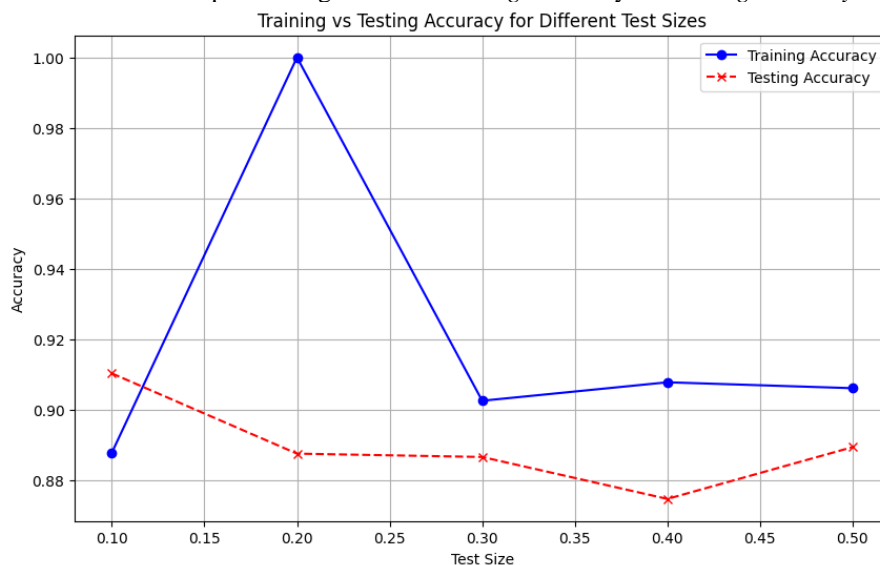
3. Mengurutkan jarak yang terbentuk
4. Menentukan jarak terdekat sampai urutan K
5. Memasangkan kelas yang bersesuaian
6. Mencari jumlah kelas dari tetangga yang terdekat dan tetapkan kelas tersebut sebagai kelas data yang akan dievaluasi.

Langkah pertama adalah menentukan nilai k dengan melakukan perulangan *cross validation* dengan *test size* 0.1, 0.2, 0.3, 0.4, 0.5 dan perulangan k=1 sampai k=20 serta penggunaan *metric Euclidean* [23]. Tabel 3 merupakan hasil terbaik dan selisih sedikit *testing accuracy* dan *training accuracy* dari setiap split data dan penentuan k.

Tabel 3. Hasil *Testing Accuracy* dan *Training Accuracy*

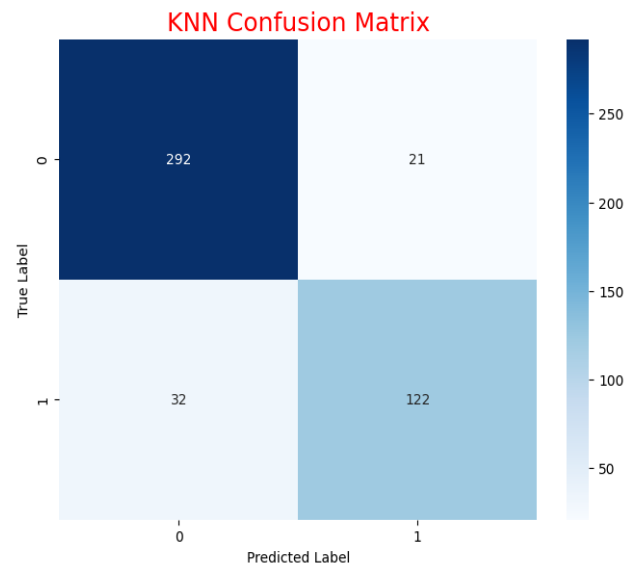
<i>Test size</i>	k	<i>Train accuracy</i>	<i>Test accuracy</i>
0.1	19	0.8876967095851216	0.9102564102564102
0.2	1	1.0	0.8876967095851216
0.3	19	0.9024839006439742	0.8865096359743041
0.4	19	0.907725321888412	0.8745980707395499
0.5	17	0.9060489060489061	0.8893178893178894

Dari hasil tabel diatas *train accuracy* terbaik berada di *test size* 0.2 dengan k=1 yang menghasilkan akurasi sebesar 1.0, tetapi memiliki nilai perbandingan yang jauh dengan *test accuracy*. Hal tersebut dapat menyebabkan model menjadi *overfitting* karena model telah belajar terlalu spesifik pada data pelatihan dan tidak dapat menggeneralisasi dengan baik ke data yang belum pernah dilihat sebelumnya [24]. Gambar 2 adalah visualisasi untuk melihat perbandingan antara *training accuracy* dan *testing accuracy*.



Gambar 2. Grafik Perbandingan *Training Accuracy* dan *Testing Accuracy*

Dari visualisasi gambar 2 dipilihlah *test size* 0.3 dengan  $k=19$  untuk model yang akan digunakan dalam mengklasifikasi berita hoaks. Model disimpan dalam format *pickel* (pkl) agar dapat digunakan pada tahap *deployment* [12]. Selanjutnya model dievaluasi untuk mengetahui performa model dalam memprediksi data [25]. Pada penelitian ini, evaluasi model menggunakan *Confussion Matrix* [20]. Pada gambar 3 adalah *confusion matrix* dari model yang telah dibuat.



Gambar 3. *Confusion Matrix*

Struktur *Confusion Matrix* [21]

1. *True Label* (Label Sebenarnya): Baris pada matriks menunjukkan label sebenarnya dari data.0 (Berpotensi Fakta) dan 1 (Berpotensi Hoaks).
2. *Predicted Label* (Label Prediksi): Kolom pada matriks menunjukkan label yang diprediksi oleh model.0 (Berpotensi Fakta) dan 1 (Berpotensi Hoaks).
3. *True Positive* (TP): Model memprediksi "berpotensi fakta" dan benar-benar adalah hoaks. Terdapat 292 kasus seperti ini.
4. *True Negative* (TN): Model memprediksi "berpotensi hoaks" dan benar-benar adalah fakta. Terdapat 122 kasus seperti ini.
5. *False Positive* (FP): Model memprediksi "berpotensi hoaks" tetapi sebenarnya adalah fakta. Terdapat 21 kasus seperti ini.
6. *False Negative* (FN): Model memprediksi "berpotensi fakta" tetapi sebenarnya adalah hoaks. Terdapat 32 kasus seperti ini.

Selanjutnya tahap *deployment*, yaitu dilakukan penerapan model pada berita yang belum terklasifikasi berpotensi hoaks atau tidak. TF-IDF *vectorizer* dan model yang sebelumnya telah disimpan pada format pkl kemudian diterapkan pada berita yang ada di laman *kompas.com*, *liputan6.com*, *detik.com*, *antaranews.com*, *cnnindonesia.com*, *okezone.com*, *sindonews.com*, *pikiran rakyat.com*, *kumparan.com*, dan *wartatransparansi.com*. Proses implementasi mencakup langkah-langkah mulai dari membaca data mentah, melakukan preprocessing, hingga melakukan prediksi menggunakan model yang telah dilatih [26]

Pada tabel 4 adalah hasil rekap dari 10 *website* yang sudah diklasifikasi. Secara keseluruhan, situs-situs berita seperti *detik.com*, *okezone.com*, *liputan6.com*, dan *cnnindonesia.com* memiliki potensi fakta lebih tinggi dalam berita yang disajikan. Sementara itu, situs-situs seperti *kompas.com*, *sindonews.com*, *kumparan.com*, *pikiranrakyat.com*, dan *wartatransparansi.com* menunjukkan bahwa masih ada ruang untuk peningkatan dalam hal penyajian berita yang lebih akurat. Hasil klasifikasi ini bisa menjadi panduan untuk memilih sumber berita yang lebih dapat diandalkan.

Tabel 4. Hasil Implementasi Model

Situs Berita	Berpotensi Fakta	Berpotensi Hoaks
Kompas	73%	27%
Liputan6	79%	21%

Detik	87%	13%
CNN Indonesia	76%	24%
Okezone	81%	19%
Sindo News	58%	42%
Kumparan	60%	40%
Pikiran Rakyat	71%	29%
Wartatransparansi	68%	32%

**KESIMPULAN** Katadata, "Jaga Ruang Digital, Menkominfo: Kami Tangani 203 Isu Hoaks Pemilu 2024le," *Kominfo.Go.Id.*, 2024

Penggunaan model KNN menyajikan probabilitas potensi fakta dan hoaks dari situs-situs web berita dengan *accuracy* sebesar 0,90 pada *training accuracy* dan 0,88 pada *testing accuracy*. Hasil implementasi model menghasilkan persentase potensi fakta 87% sampai 68% dan potensi hoaks sebesar 42% sampai 13%. Hasil dari implementasi model diperoleh situs web yang memiliki tingkat potensi fakta yang lebih tinggi dibanding situs web lainnya. Situs web yang memiliki potensi fakta yang tinggi yaitu detik.com, okezone.com, liputan6.com, dan cnnindonesia.com.

#### DAFTAR PUSTAKA

- [1] D. A. Nugroho and R. M. Sukmariningsih, "PERANAN KOMISI PEMILIHAN UMUM DALAM MEWUJUDKAN PEMILU YANG DEMOKRATIS," *Jurnal JURISTIC*, vol. 1, no. 01, p. 22, Mar. 2020, doi: 10.35973/jrs.v1i01.1449.
- [2] B. Imran, M. Karim, and N. Ningsih, "KLASIFIKASI BERITA HOAX TERKAIT PEMILIHAN UMUM PRESIDEN REPUBLIK INDONESIA TAHUN 2024 MENGGUNAKAN NAÏVE BAYES DAN SVM," vol. 20, pp. 1–9, Jan. 2024, doi: 10.20884/s2v0sq74.
- [3] Google Trends, "Analisis Kueri Pemilu," <https://trends.google.co.id/>.
- [4] Biro Humas Kementerian Kominfo e-mail: [humas@mail.kominfo.go.id](mailto:humas@mail.kominfo.go.id), "Jaga Ruang Digital, Menkominfo: Kami Tangani 203 Isu Hoaks Pemilu 2024," <https://www.komdigi.go.id/berita/siaran-pers/detail/siaran-pers-no-03-hm-kominfo-01-2024-tentang-jaga-ruang-digital-menkominfo-kami-tangani-203-isu-hoaks-pemilu-2024>.
- [5] Fajri, N., "Hoaks Merajalela? Jangan Sampai Kamu Jadi Korbannya," [djkn.kemenkeu.go.id](http://djkn.kemenkeu.go.id).
- [6] S. Zhang and J. Li, "KNN Classification With One-Step Computation," *IEEE Trans Knowl Data Eng*, vol. 35, no. 3, pp. 2711–2723, 2023, doi: 10.1109/TKDE.2021.3119140.
- [7] K. Taha, P. D. Yoo, C. Yeun, and A. Taha, "Text Classification Techniques: A Holistic Review, Observational Analysis, and Experimental Investigation," *Big Data Mining and Analytics*, vol. 8, no. 3, pp. 624–660, Jun. 2025, doi: 10.26599/BDMA.2024.9020092.
- [8] A. P. Farhan and F. Arif, "Pendekatan Data Mining Pada Metodologi Six Sigma untuk Perbaikan Kualitas Produk," *Itenas*, 2022.
- [9] M. Djufri, "PENERAPAN TEKNIK WEB SCRAPING UNTUK PENGGALIAN POTENSI PAJAK (STUDI KASUS PADA ONLINE MARKET PLACE TOKOPEDIA, SHOPEE DAN BUKALAPAK)," *Jurnal BPPK : Badan Pendidikan dan Pelatihan Keuangan*, vol. 13, pp. 65–75, Dec. 2020, doi: 10.48108/jurnalbppk.v13i2.636.
- [10] N. Nofiyani and W. Wulandari, "Implementasi Electronic Data Processing Untuk meningkatkan Efektifitas dan Efisiensi Pada Text Mining," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 3, p. 1621, Jul. 2022, doi: 10.30865/mib.v6i3.4332.
- [11] E. P. Sutrisno and S. Amini, "IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOR PADA ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI DIGITAL KORLANTAS POLRI," Jakarta, 2023. Accessed: May 19, 2025. [Online]. Available: <https://senafiti.budiluhur.ac.id/senafiti/article/view/836/543>
- [12] R. A. S. MUHAMAD FADLI, "KLASIFIKASI DAN EVALUASI PERFORMA MODEL RANDOM FOREST UNTUK PREDIKSI STROKE," *Jurnal Teknik FT UMT*, vol. 12, no. 2, pp. 72–80, Oct. 2023.
- [13] Kominfo, "TurnBackHoax, Komunitas Online Anti Hoax di Indonesia," <https://www.komdigi.go.id/>.
- [14] Kominfo, "Dirjen IKP inginkan Portal InfoPublik dapat menyuarakan Pemerintah dan aspirasi Publik," <https://www.komdigi.go.id/>.
- [15] RRI, "Mengenal Sejarah Berdirinya RRI," [rri.co.id](http://rri.co.id).
- [16] Rini Hairani, "Timnas AMIN Daftarkan Gugatan Hasil Pemilu 2024 ke MK," <https://www.rri.co.id/pemilu/600861/timnas-amin-daftarkan-gugatan-hasil-pemilu-2024-ke-mk>.
- [17] H. Zhou, "Research of Text Classification Based on TF-IDF and CNN-LSTM," *J Phys Conf Ser*, vol. 2171, no. 1, p. 012021, Jan. 2022, doi: 10.1088/1742-6596/2171/1/012021.

- [18] K. M. Suryaningrum, "Comparison of the TF-IDF Method with the Count Vectorizer to Classify Hate Speech," *Engineering, MATHematics and Computer Science (EMACS) Journal*, vol. 5, no. 2, pp. 79–83, May 2023, doi: 10.21512/emacsjournal.v5i2.9978.
- [19] I. W. S. Supriana, M. A. Raharja, and I. M. S. Bimantara, "Pengembangan Sistem Prediksi Bantuan Program Keluarga Harapan (PKH) Berbasis Machine Learning," *SINTECH (Science and Information Technology) Journal*, vol. 6, no. 1, pp. 26–36, Apr. 2023, doi: 10.31598/sintechjournal.v6i1.1297.
- [20] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," 2021.
- [21] Muhammad Ferian Rizky Akbari, Bayu Rahayudi, and Lailil Muflikhah, "Implementasi Deep Learning menggunakan Algoritma EfficientDet untuk Sistem Deteksi Kelayakan Penerima Bantuan Langsung Tunai berdasarkan Citra Rumahdi Wilayah Kabupaten Kediri," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 4, pp. 1817–1825, Apr. 2023, Accessed: May 19, 2025. [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12596/5720>
- [22] D. Cahyanti, A. Rahmayani, and S. A. Husniar, "Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara," *Indonesian Journal of Data and Science*, vol. 1, no. 2, pp. 39–43, Jul. 2020, doi: 10.33096/ijodas.v1i2.13.
- [23] H. Hafid, "Penerapan K-Fold Cross Validation untuk Menganalisis Kinerja Algoritma K-Nearest Neighbor pada Data Kasus Covid-19 di Indonesia." [Online]. Available: <http://www.ojs.unm.ac.id/jmathcos>
- [24] C. N. Syahputri and M. S. Hasibuan, "OPTIMASI KLASIFIKASI DECISION TREE DENGAN TEKNIK PRUNING UNTUK MENGURANGI OVERFITTING," *JSii (Jurnal Sistem Informasi)*, vol. 11, no. 2, pp. 87–96, Sep. 2024, doi: 10.30656/jsii.v11i2.9161.
- [25] I. W. S. Supriana, M. A. Raharja, and I. M. S. Bimantara, "Pengembangan Sistem Prediksi Bantuan Program Keluarga Harapan (PKH) Berbasis Machine Learning," *SINTECH (Science and Information Technology) Journal*, vol. 6, no. 1, pp. 26–36, Apr. 2023, doi: 10.31598/sintechjournal.v6i1.1297.
- [26] P. Rahayu *et al.*, *BUKU AJAR DATA MINING*. 2024.