

KLASIFIKASI BERITA HOAKS KAMPANYE PEMILIHAN UMUM (PEMILU) 2024 MENGUNAKAN ALGORITMA NAÏVE BAYES

Samuel Fernando Yeremia, Apriade Voutama, Ade Andri Hendriadi

Sistem Informasi, Universitas Singaperbangsa Karawang
Jl. HS. Ronggo Waluyo, Puseurjaya, Telukjambe Timur Karawang, Indonesia
Samuelyeremia11@gmail.com

ABSTRAK

Internet telah menjadi media informasi yang terkenal di berbagai bidang, termasuk pencarian berita, foto, ulasan produk, layanan masyarakat, film, dan lainnya. Semua ini disajikan melalui sumber berbagi seperti berita, media sosial, dan blog. Blog dan website sering disebut sebagai sumber berita. Berita dapat memiliki perspektif netral, negatif, atau positif. Dalam era informasi digital yang berkembang pesat, berita dan informasi menyebar dengan cepat melalui platform internet. Pada masa pemilihan umum (pemilu), banyak buzzer penyebar berita hoaks yang muncul untuk mengelabui masyarakat demi kepentingan pribadi dan terkadang merugikan beberapa calon legislative ataupun partai politik. Permasalahan ini dapat diselesaikan dengan menggunakan text mining yang mengklasifikasikan judul berita menggunakan algoritma naïve bayes dan metode *Knowledge Discovery in Database* (KDD) dan menggunakan rapidminer sebagai alat pengujian. Algoritma naïve bayes dan metode KDD ini sering digunakan dan dianggap sangat baik dalam mengklasifikasikan dataset dengan jumlah yang besar. Hal ini menimbulkan tantangan baru dalam membedakan berita hoaks dan fakta.. Hasil penelitian ini akan menunjukkan nilai tertinggi dari akurasi sebesar 89.54% , nilai tertinggi presisi sebesar 86.44% dan nilai tertinggi dari recall sebesar 82.00%.

Kata kunci : Klasifikasi, Pemilu 2024, Naïve Bayes, KDD, Hoaks.

1. PENDAHULUAN

Di dalam sejarah nasional Indonesia, pemilihan umum telah diadakan beberapa kali, tetapi pemilihan umum yang dilakukan secara langsung oleh rakyat Indonesia baru pertama kali dilakukan pada era reformasi setelah runtuhnya era orde baru pada tahun 2004 [1] Pada masa pemilihan umum calon presiden 2024 dipastikan ada kampanye yang dilakukan oleh setiap tim sukses dari para calon. Kampanye yang dilakukan melalui koran, media internet dan penyebaran spanduk. Di zaman digital saat ini, mendapatkan informasi secara online menjadi sangat mudah. [2].

Dalam konteks media internet, website dapat dianggap sebagai suatu platform baru untuk melakukan kampanye. Pada masa kampanye pemilihan umum 2024 ini, masyarakat harus memiliki pengetahuan tentang penyebaran berita yang cepat dan mudah diakses di berbagai *platform* media untuk mengatasi penyebaran berita hoaks.

Penyebaran berita hoaks dalam kampanye pemilu seringkali sulit untuk diawasi dan dihentikan. Hukum yang mengatur hoaks mungkin tidak selalu efektif, dan penegakan hukum bisa sulit, terutama ketika pelaku bersembunyi di balik anonimitas internet. Para penyebar hoaks dapat memiliki agenda politik atau agenda pribadi tertentu yang ingin mereka promosikan. Mereka mungkin ingin mendiskreditkan lawan politik atau menguntungkan calon tertentu, sehingga mereka menciptakan atau menyebarkan berita palsu yang mendukung tujuan mereka. Mereka mungkin ingin mendiskreditkan lawan politik atau menguntungkan calon tertentu, sehingga mereka

menciptakan atau menyebarkan berita palsu yang mendukung tujuan mereka.

Salah satu cara untuk menyelesaikan masalah ini adalah dengan menggunakan Text Mining yaitu teknik klasifikasi yang melibatkan analisis otomatis teks berita untuk menentukan apakah berita itu hoaks atau tidak. Text mining adalah jenis Data Mining yang menganalisis teks. Ini adalah proses menganalisis data berupa teks, yang biasanya diambil dari dokumen, dan bertujuan untuk menemukan kata-kata yang dapat mewakili isi dokumen. Dengan menggunakan algoritma text mining, kita dapat mengidentifikasi kata-kata kunci atau frasa yang sering muncul dalam sekelompok dokumen. Hal ini dapat membantu kita memahami topik atau tema yang dominan dalam dokumen tersebut [3]

Proses teks mining menggunakan berbagai algoritma, termasuk Decision Tree, K-Means Clustering, Naïve Bayes Classifier, K-Nearest Neighbor (KNN), dan Support Vector Machine (SVM). [4]

Algoritme klasifikasi yang populer seperti Naïve Bayes Classifiers (NBC) dan Support Vector Machine (SVM) banyak digunakan oleh peneliti untuk deteksi berita hoaks [5]). Tujuan dari klasifikasi adalah untuk melakukan prediksi kelas dari obyek yang belum diketahui kelas dan karakteristik tipe datanya.

Penulis akan mengumpulkan dan menganalisis dataset berita yang berasal dari berbagai jenis dan sumber, kemudian menggunakan metode Naive Bayes untuk membangun model klasifikasi serta menentukan apakah teks berita itu hoaks atau valid. Tujuan dari penelitian ini adalah untuk mengetahui potensi dan keterbatasan metode ini dalam mengklasifikasikan

teks berita sebagai berita hoaks atau fakta. Akurasi, presisi, Recall, dan TF-IDF akan digunakan untuk mengevaluasi kinerja model ini. Oleh karena itu judul dari penelitian ini adalah Klasifikasi Berita Hoax Kampanye Pemilihan Umum (Pemilu) 2024 Menggunakan Algoritma Naïve Bayes.

2. TINJAUAN PUSTAKA

Setelah penulis melakukan telaah terhadap beberapa penelitian, ada beberapa penelitian yang memiliki keterkaitan dengan penelitian yang penulis lakukan.

Studi sebelumnya oleh [6] "Eksperimen Naive Bayes Pada Deteksi Berita Hoax Berbahasa Indonesia" menemukan bahwa metode klasifikasi Naive Bayes dapat digunakan sebagai algoritma untuk mengklasifikasikan berita di internet, dan metode ini memiliki nilai akurasi yang tinggi karena memiliki kinerja yang baik..

Selanjutnya, studi "Klasifikasi Berita Hoax Dengan Menggunakan Metode Naive Bayes" [7] menemukan bahwa sistem klasifikasi berita dapat menggunakan metode Naive Bayes dengan memasukkan teks. Proses preprocessing mencakup tokenization, stopword, parsing, dan pembobotan kata, yang kemudian digunakan untuk klasifikasi. Pada tahap pengukuran divalidasi melalui sepuluh kali uji cross-validation. Hasil penelitian menunjukkan nilai keakuratan terbaik hasil uji keenam dengan nilai keakuratan 85,28%. Dokumen relevan diklasifikasikan menjadi 307 dan dokumen tidak relevan, masing-masing dengan tingkat kesalahan 14,72%. Nilai rata-rata untuk berita hoax dan benar adalah 0,896, dan nilai Recall adalah 0.853.

Penelitian yang dilakukan [8] menyatakan bahwa metode Naive Bayes, yang digunakan dengan 400 data dan empat kategori, dapat digunakan untuk mengkategorikan judul berita di internet. Model yang dibuat di Rapid Miner menghasilkan akurasi akhir sebesar 78,75%, ketepatan sebesar 80,56%, dan Recall sebesar 78,75%.

2.1. Berita Hoaks

Hoax adalah informasi atau berita yang berisi hal-hal yang belum dapat dipastikan atau benar-benar bukan fakta yang terjadi. Penyebaran hoax memiliki berbagai tujuan, seperti untuk bahan lelucon, iseng, promosi, atau sebagai alat untuk menjatuhkan pesaing. Sebagai akibatnya, penerima berita hoax mudah terpancing dan menyebarkannya kepada rekan-rekannya, sehingga berita hoax tersebut menyebar dengan cepat.[9]

2.2. Teks Mining

Text mining merupakan salah satu jenis Data Mining yang fokus pada analisis teks. Proses ini melibatkan analisis data dalam bentuk teks yang biasanya diambil dari dokumen, dengan tujuan untuk menemukan kata-kata yang dapat merepresentasikan isi dokumen tersebut. Text mining merupakan bidang

ilmu yang dapat mengatasi masalah ini dengan cara mengekstraksi kata-kata unik guna mendapatkan keunikan dari setiap dokumen.[10]

2.3. Klasifikasi

Pada penelitian ini, dilakukan pengklasifikasian data pada dataset berita hoax berdasarkan atribut yang dimiliki oleh dataset tersebut. Metode klasifikasi yang digunakan adalah Naïve Bayes. Naïve Bayes adalah metode klasifikasi yang berdasarkan pada teorema Bayes.[11]

Untuk menggambarkan sebuah kelas dokumen, terdapat karakteristik petunjuk yang diperlukan untuk melakukan klasifikasi. Karakteristik petunjuk ini dapat berupa kata-kata kunci, topik, atau atribut lain yang relevan dengan kelas dokumen tersebut. Dalam penelitian ini, atribut-atribut yang digunakan untuk mengklasifikasikan berita hoax adalah atribut-atribut yang terdapat dalam dataset berita hoax.

Pada metode klasifikasi Naïve Bayes, terdapat asumsi bahwa keberadaan atau ketiadaan ciri tertentu dalam suatu kelas tidak memiliki hubungan dengan ciri dari kelas lainnya. Dengan kata lain, atribut-atribut yang digunakan untuk mengklasifikasikan berita hoax tidak saling bergantung satu sama lain.

3. METODE PENELITIAN

Naïve Bayes adalah algoritma yang sering digunakan dalam bidang machine learning dan Data Mining untuk melakukan klasifikasi pada data yang tidak terstruktur. Algoritma ini menggunakan Teorema Bayes dengan menggunakan teknik probabilitas dan statistik untuk memperkirakan atau memprediksi peluang kejadian berdasarkan pengalaman sebelumnya. Dalam Naïve Bayes Classifier, kita mengasumsikan bahwa setiap fitur atau atribut dalam data tidak bergantung satu sama lain. Meskipun asumsi ini seringkali tidak realistis dalam dunia nyata, namun Naïve Bayes Classifier tetap memberikan hasil yang cukup akurat dalam banyak kasus [12]

Naive Bayes merupakan klasifikasi paling sederhana dan paling umum digunakan dalam menghitung probabilitas kelas berdasarkan distribusi kata dalam data. Teori klasifikasi Bayes adalah pendekatan statistik yang mendasar dalam memprediksi peluang pada data. [13]

Metodologi *Knowledge Discovery in Database* (KDD) adalah metode penelitian yang digunakan penulis. Tahapan KDD pada penelitian ini adalah sebagai berikut:

- a. Data Selection. (Pemilihan Data)
- b. Data Preprocessing. (Pemrosesan Data)
- c. Transformation. (Transformasi Data)
- d. Data Mining. (Penggalian Data)
- e. Interpretation/Evaluation.

Beberapa proses tersebut dapat dijelaskan sebagai berikut :



Gambar 1. Proses KDD

KDD (*Knowledge Discovery Database*) adalah teknik penambangan data yang dapat digunakan proses penambangan data untuk menemukan pola dan informasi penting dalam data.

4. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk melakukan klasifikasi berita hoaks dalam pemilihan umum (pemilu) tahun 2024. Metodologi yang digunakan dalam penelitian ini adalah Knowledge Discovery in Database (KDD) yang melibatkan beberapa tahapan, yaitu Data Selection, Data Preprocessing, Transformasi, Data Mining, dan Evaluasi.

4.1. Pemilihan Data

Untuk mengumpulkan data, peneliti menggunakan teknik Scrapping data dari website Google News dan Turnbackhoaks.id. Total data berita yang berhasil dikumpulkan sebanyak 3255 data dalam rentang waktu satu tahun, mulai dari Januari 2023 hingga Januari 2024 dan data tersebut di konversi menjadi file excel untuk selanjutnya dilakukan tahapan *preprocessing* data.

4.2. Preprocessing Data

Setelah data terkumpul, langkah selanjutnya adalah tahap Preprocessing untuk membersihkan data berita dari kata-kata atau karakter yang tidak relevan dalam proses klasifikasi. Selain itu, kata-kata juga disesuaikan menjadi huruf kecil (*lowercase*). Setelah melalui tahap pembersihan data manual dan menggunakan tools rapidminer, jumlah data yang digunakan dalam penelitian ini menjadi 1031 data.



Gambar 2. Pre-processing data

Pada Tahap Preprocessing data terdapat beberapa tahapan yaitu :

- Tokenize : untuk melakukan pemecahan kalimat menjadi kata per kata
- Case folding : untuk melakukan perubahan kata kata menjadi huruf kecil
- Stopwords : untuk melakukan penghapusan karakter, link, kata kata yang tidak digunakan dalam penelitian, dan emoticon.

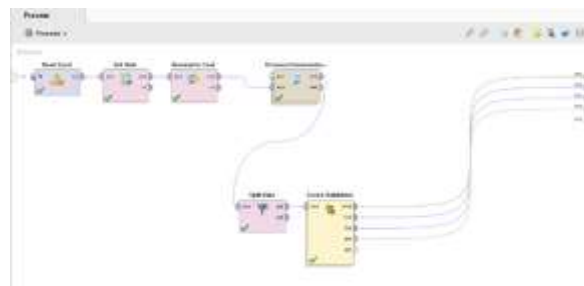
- Stemming : untuk melakukan penghapusan kata imbuhan dari kata tersebut.

4.3. Transformasi data

Tahap berikutnya adalah transformasi data, di mana kata-kata atau term dalam dokumen diberi bobot. Proses ini menggunakan metode TF-IDF dengan menggunakan operator stopwords removal dan operator stemming menggunakan kamus yang diperoleh dari github genediazjr dan github sastrawi.

4.4. Data Mining

Setelah pembobotan kata dilakukan, tahap selanjutnya adalah pembagian data menggunakan operator split data. Data dibagi menjadi data testing dan data training secara otomatis dengan menggunakan empat skenario, yaitu 90:10, 80:20, 70:30, dan 60:40. Tools yang digunakan untuk mengklasifikasikan berita hoaks dalam masa pemilihan umum (pemilu) 2024 adalah RapidMiner.



Gambar 3. Proses Datamining

4.5. Interpretasi dan Evaluasi

Selanjutnya dilakukan pengujian performa menggunakan operator cross validation dengan menggunakan confusion matrix. Selain itu, dilakukan pengujian menggunakan operator apply model dan operator performance untuk mendapatkan nilai akurasi, presisi, dan recall dari klasifikasi menggunakan algoritma naïve bayes.

Tabel 1. Hasil Pengujian

Skenario	Accuracy	Precision	Recall
90:10	89.54%	86.44%	82.00%
80:20	88.71%	84.58%	81.62%
70:30	88.63%	84.18%	81.97%
60:40	88.83%	84.85%	81.60%

Setelah menjalankan operator, dalam pengujian akurasi, skenario 70 : 30 memiliki akurasi terendah sebesar 88.63% dan skenario 90 : 10 memiliki akurasi tertinggi sebesar 89.54%. Pada pengujian presisi, skenario 70 : 30 memiliki presisi terendah sebesar 84.18% dan skenario 90 : 10 memiliki presisi tertinggi sebesar 86.44%. Pada pengujian recall, skenario 90 : 10 memiliki recall tertinggi sebesar 86.44% dan skenario 70 : 30 memiliki recall terendah sebesar 84.18%.

5. KESIMPULAN DAN SARAN

Penerapan klasifikasi naïve bayes untuk mencegah penyebaran berita hoaks pada pemilihan umum (pemilu) 2024 dengan menggunakan algoritma naïve bayes dan metode Knowledge Discovery in Database (KDD) serta menggunakan sumber data dari website google news dan website turnbackhoax.id menampilkan hasil akurasi, presisi dan recall dengan pengujian 4 skenario memiliki hasil valid yang baik. Jumlah presentase berita hoaks pada masa kampanye pemilihan umum 2024 sebanyak 25% atau 255 data. Penulis menyarankan untuk Dataset yang digunakan pada penelitian selanjutnya dapat lebih besar untuk meningkatkan nilai akurasi, presisi dan recall. Setelah itu lakukan penelitian pada sumber berita, metode klasifikasi dan tools yang berbeda, agar dapat dibandingkan serta mengedukasi Masyarakat sumber website berita mana saja yang dapat di percaya serta memiliki berita yang valid atau tidak. Lakukan validasi hasil klasifikasi dengan sumber data lain atau metode lain untuk memastikan keakuratannya.

DAFTAR PUSTAKA

- [1] I. Kurniawan and A. Susanto, "Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019," *Eksplora Inform.*, vol. 9, no. 1, pp. 1–10, Sep. 2019, doi: 10.30864/eksplora.v9i1.237.
- [2] V. R. PASARIBU, A. Abdiansah, and A. S. Utami, "Penerapan Algoritme Naive Bayes Classifiers Pada Sistem Pendeteksi Berita Hoax Berbahasa Indonesia," 2021. [Online]. Available: https://repository.unsri.ac.id/49533/%0Ahttps://repository.unsri.ac.id/49533/11/RAMA_55201_09021281722044_0001108401_0022127804_01_front_ref.pdf
- [3] N. Silalahi and G. Leonarde Ginting, "Rekomendasi Berita Berkaitan dengan Menerapkan Algoritma Text Mining dan TF-IDF," *Bull. Comput. Sci. Res.*, vol. 3, no. 4, pp. 276–282, 2023, doi: 10.47065/bulletincsr.v3i4.266.
- [4] R. R. Sani, Y. A. Pratiwi, S. Winarno, E. D. Udayanti, and F. Alzami, "Analisis Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Berita Hoax pada Berita Online Indonesia," 2022. doi: 10.14710/jmasif.13.2.47983.
- [5] S. Soleman, "Pemanfaatan Metode Klasifikasi Naïve Bayes Untuk Pendeteksi Berita Hoax Pada Artikel Berbahasa Indonesia," *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 7, no. 2, p. 83, Dec. 2021, doi: 10.24014/coreit.v7i2.14290.
- [6] F. Rahutomo, I. Y. R. Pratiwi, and D. M. Ramadhani, "Eksperimen Naïve Bayes Pada Deteksi Berita Hoax Berbahasa Indonesia," *J. Penelit. Komun. Dan Opini Publik*, vol. 23, no. 1, 2019, doi: 10.33299/jpkop.23.1.1805.
- [7] H. Mustofa and A. A. Mahfudh, "Klasifikasi Berita Hoax Dengan Menggunakan Metode Naive Bayes," *Walisono J. Inf. Technol.*, vol. 1, no. 1, p. 1, Nov. 2019, doi: 10.21580/wjit.2019.1.1.3915.
- [8] M. Sholih 'afif, M. Muzakir, M. I. Al, and G. Al Awalaen, "TEXT MINING UNTUK MENGLASIFIKASI JUDUL BERITA ONLINE STUDI KASUS RADAR BANJARMASIN MENGGUNAKAN METODE NAÏVE BAYES," *Kumpul. J. Ilmu Komput.*, vol. 08, no. 2, 2021.
- [9] C. S. Sriyano and E. B. Setiawan, "Pendeteksian Berita Hoax Menggunakanfile:///D:/SEMESTER_8/ref/113-345-1-PB.pdfan Naive Bayes Multinomial Pada Twitter dengan Fitur Pembobotan TF-IDF," 2021.
- [10] R. Nanda, E. Haerani, S. K. Gusti, and S. Ramadhani, "Klasifikasi Berita Menggunakan Metode Support Vector Machine," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 2, pp. 269–278, 2022, doi: 10.32672/jnkti.v5i2.4193.
- [11] A. Yonathan, H. Saujaini, E. Esyudha Pratama, J. H. Nawawi, and K. Barat, "Perbandingan Algoritma Klasifikasi dalam Pendeteksian Hoax pada Media Sosial", doi: 10.26418/juara.v1i1.53126.
- [12] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 437, 2020, doi: 10.30865/mib.v4i2.2080.
- [13] D. Novianti, "Implementasi Algoritma Naïve Bayes Pada Data Set Hepatitis Menggunakan Rapid Miner," *Paradig. - J. Komput. dan Inform.*, vol. 21, no. 1, pp. 49–54, 2019, doi: 10.31294/p.v21i1.4979.