

IDENTIFIKASI BERITA HOAX MENGGUNAKAN KOMBINASI METODE K-NEAREST NEIGHBOUR (KNN) DAN TF-IDF BERBASIS WEB DENGAN MENGGUNAKAN *FRAMEWORK CODEIGNITER*

Danu Nur Irwanto¹

¹ Mahasiswa (Teknik Informatika, Universitas Wijaya Putra)

¹ Email:

ABSTRAK

Internet berkembang menjadi media informasi. Sangat disayangkan apabila informasi yang disampaikan tidak akurat atau bohong (hoax). Yakni berita yang memuat informasi palsu. Saat ini teknologi pendeteksi hoax masih jarang ditemukan. Untuk mengidentifikasi semua berita hoax secara manual adalah hal yang sangat sulit. Oleh karena itu diperlukan pendekatan secara artificial intelligence (AI) atau kecerdasan buatan. Dimana sebuah mesin dilatih terlebih dahulu (machine learning) untuk menyelesaikan masalah tertentu. Dengan adanya kecerdasan buatan ini diharapkan dapat mengidentifikasi berita hoax lebih cepat dan lebih akurat. Dalam ilmu komputer, untuk mengidentifikasi kemungkinan berita yang mengandung hoax atau tidak dilakukan dengan menggunakan pemodelan bahasa. Salah satunya adalah bahasa pemodelan dengan algoritma TF-IDF. Kemudian Metode K-Nearest Neighbor (KNN), dipakai untuk mengklasifikasikan berita yang sudah diidentifikasi sebagai berita palsu. Aplikasi ini dapat mengurangi penyebaran berita hoax. Sistem dapat mengidentifikasi berita hoax. Selama kategori berita ada dalam dataset. Untuk akurasi sistem identifikasi berita hoax dipengaruhi jumlah dataset yang ada dan pembahasan yang sesuai dengan topiknya. Dapat dilakukan penambahan dataset berita hoax secara otomatis dari sumber berita yang sudah teridentifikasi, dari situs-situs resmi berita hoax dan valid yang sudah ada. Agar proses identifikasi menjadi lebih akurat.

Kata Kunci: TF-IDF; K-NN; BERITA; HOAKS; CODEIGNITER

PENDAHULUAN

Pesatnya peningkatan inovasi teknologi mempermudah kita untuk mendapatkan informasi dengan cepat dimana saja dan kapan saja. Sejujurnya, tidak ada informasi yang tidak dapat diperoleh dengan pandangan kita, ini menunjukkan bahwa semakin pesat laju peningkatan teknologi semakin cepat informasi yang dapat kita peroleh [1]

. Kemajuan teknologi saat ini memberikan dampak yang positif, namun juga berdampak negatif. Penyampaian informasi menjadi sangat cepat, dimana setiap orang dengan mudahnya membuat informasi tersebut melalui internet.

Menurut [2] Internet telah berkembang menjadi media informasi yang terkenal tentang berbagai hal seperti pencarian berita, lagu, gambar, produk, film dan lain sebagainya. Semua disajikan dalam berbagai sumber seperti artikel, media sosial, dan blog. Sangat disayangkan jika informasi yang disampaikan merupakan informasi yang tidak akurat. Apalagi informasi tersebut merupakan informasi palsu (hoax) yaitu berita yang mengandung informasi palsu dan tidak bertanggung jawab serta bertujuan untuk

meyakinkan netizen agar tertipu dan percaya pada berita hoax tersebut.

Menurut [3] Semakin maraknya trend hoax maka semakin banyak pula bermunculan ide-ide untuk mewaspadai penyebaran berita hoax tersebut. Banyak tips bermunculan agar tidak terjebak pemberitaan hoax, dan banyak juga media sosial yang memberikan layanan tambahan untuk mengeluhkan konten yang diduga mengandung unsur hoax. Untuk saat ini teknologi pencegahan hoax yang diterapkan dalam sistem deteksi hoax masih jarang ditemukan. Beberapa aplikasi penangkal hoax hanyalah situs web pelaporan berita hoax.

Menurut [4] dapat di lihat untuk mengidentifikasi berita hoax ini secara manual adalah hal yang sangat sulit dan merepotkan. Oleh karena itu dibutuhkan pemanfaatan teknologi seperti *Artificial Intelligent (AI)*. *AI* merupakan mesin (komputer) yang cerdas, dimana sebuah mesin dilatih lebih dahulu (*machine learning*) untuk menyelesaikan masalah tertentu. Kecerdasan buatan ini diharapkan dapat bekerja lebih cepat dan lebih akurat.

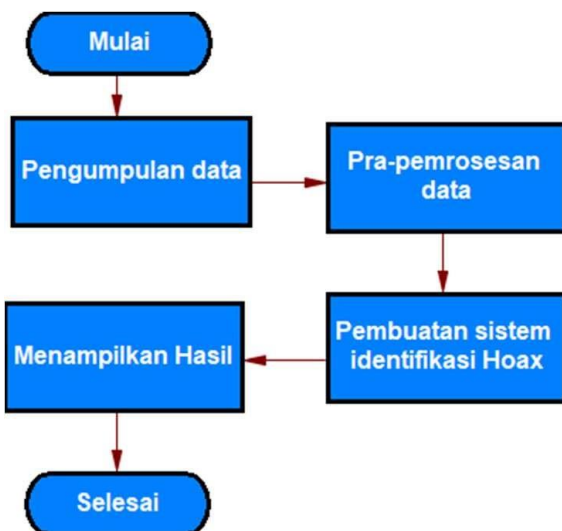
Dalam ilmu komputer, untuk mengidentifikasi kemungkinan berita yang mengandung hoax atau tidak dilakukan dengan menggunakan pemodelan bahasa [4]. Salah satunya adalah bahasa pemodelan dengan algoritma TF-IDF. Algoritma TF-IDF akan digunakan untuk mengolah setiap kata, memisahkan dan membandingkannya dengan kata-kata yang ada. Nilai kesamaannya akan digunakan untuk mengidentifikasi apakah suatu berita mengandung hoax.

Metode K-Nearest Neighbor (KNN), dipakai untuk mengklasifikasikan berita yang sudah diidentifikasi sebagai berita palsu. K-Nearest Neighbor (KNN) akan menghitung jumlah frekuensi sebuah kata muncul dalam sebuah kalimat. Nilai frekuensi yang tertinggi merupakan klasifikasi kalimat tersebut.

METODE PENELITIAN

Dalam pembuatan tugas akhir ini akan dikembangkan sistem yang mengimplementasikan konsep identifikasi berita hoax dengan kombinasi metode yaitu K-Nearest Neighbor (KNN) dan TF-IDF. Dalam sistem tersebut terdapat beberapa proses yang dimulai dari pengumpulan data, pre-processing data, K-Nearest Neighbor (KNN) dan TF-IDF, dan output berupa identifikasi berita hoax. Data penelitian ini diperoleh langsung dari dataset yang sudah tersedia di situs web "<https://data.mendeley.com/datasets/p3hfgr5j3m/1>".

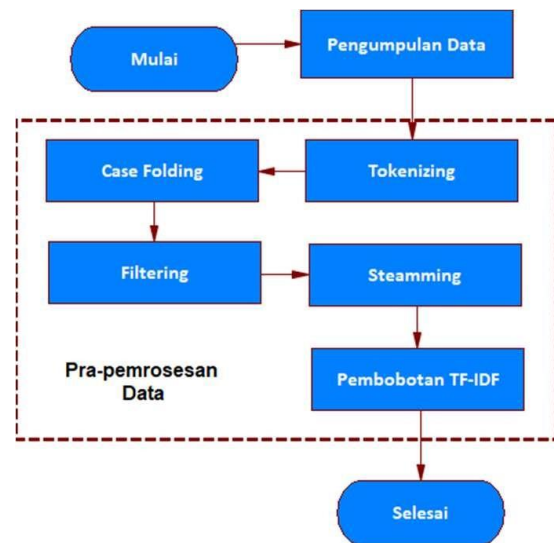
Berikut ini adalah tahapan alir perancangan sistem :



Gambar 1 Tahap Alir Perancangan Sistem

Dataset tersebut dikumpulkan secara manual melalui mesin pencarian google. Peneliti mengumpulkan dataset di portal-portal berita secara manual melalui internet, dimana menggunakan Google sebagai mesin pencari berita. Caranya adalah peneliti memasukkan kata kunci berita pada Google. Kemudian Google akan mencari berita yang sama dengan kata kunci yang dimasukkan. Pakarjuga harus mengetahui apakah berita yang diberikan oleh Google terinfeksi sebagai berita atau penilaian umum dengan cara identifikasi. Berdasarkan penelitian sebelumnya [5] terdapat situs berita yang divalidasi, diantaranya adalah merdeka.com, kompas.com, viva.co.id, metrotvnews.com, republika.co.id, tempo.co, dan tribunnews.com.

Berikut ini adalah diagram alir pra-pemrosesan sistem :



Gambar 2 Diagram Alir Pra-pemrosesan Data

Pra-pemrosesan data secara umum memiliki beberapa proses utama yaitu Tokenizing, Case Folding, Filtering, Steaming dan yang terakhir adalah Pembobotan TF-TDF. Tahapan Pra-pemrosesan ini akan menghasilkan kata-kata berita hoax terseleksi dan disimpan dalam dokumen dengan format csv. Dokumen tersebut nantinya akan digunakan sebagai dataset dalam klasifikasi K-Nearest Neighbor (KNN), dengan mencari jarak terdekat antara berita hoax yang sudah di proses dengan berita hoax yang diuji.

HASIL DAN PEMBAHASAN

Tahap ini merupakan tahap terakhir dalam implementasi system identifikasi berita hoax. Pada tahap ini menampilkan hasil identifikasi dari data uji berita. Hasil identifikasi didapatkan melalui keterangan validasi berita terdekat. Ada 3 hasil identifikasi yaitu HOAX, VALID, UNIDENTIFIED. Berikut ini tabel hasil validasi berita :

Tabel 1 Validasi Berita Dataset

NO	DATASET BERITA	VALIDASI BERITA
1	DATASET 2	HOAX
2	DATASET 3	VALID
3	DATASET 4	HOAX
4	DATASET 1	VALID
5	DATASET 5	VALID

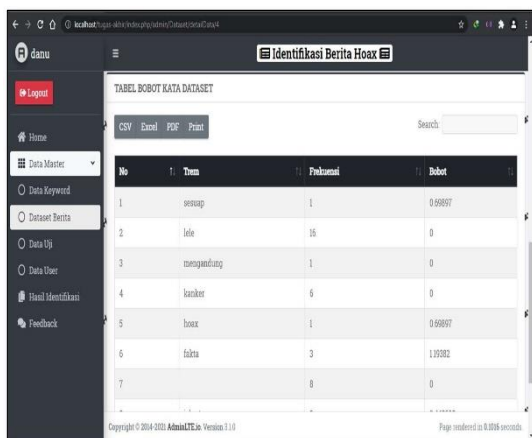
Hasil indentifikasi berita dapat dilihat dari tabel berikut ini :

Tabel 2 Hasil Identifikasi Berita

N O	DATA UJI BERITA	HASIL IDENTIFIKASI
1	DATA UJI	HOAX

Implementasi Identifikasi Hoax

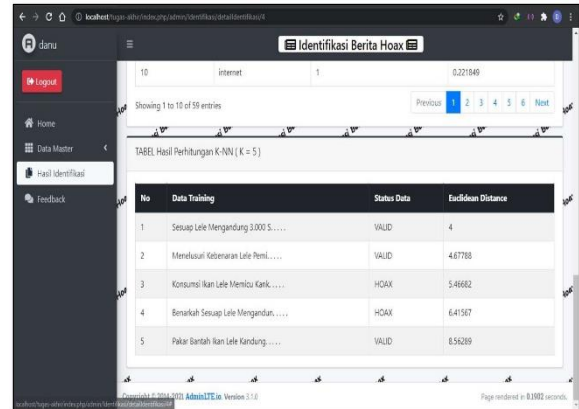
Pada gambar 3 merupakan tampilan input berita untuk identifikasi. Di sini admin bisa menginputkan berita yang mau di identifikasi untuk mengetahui berita tersebut merupakan berita hoax atau bukan. Admin dapat menginputkan berita tersebut pada kolom input, setelah di input admin bisa menekan tombol Identifikasi untuk mengetahui hasilnya.



No	Term	Frekuensi	Bobot
1	sesuap	1	0.69897
2	lele	15	0
3	mengandung	1	0
4	kaniker	6	0
5	hoax	1	0.69897
6	faits	3	1.03202
7		8	0

Gambar 3 Tampilan Hasil Bobot TF-IDF

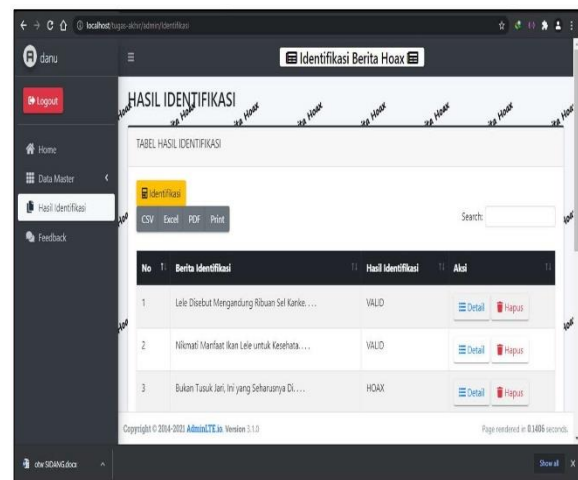
Pada gambar 5 merupakan tampilan halaman hasil proses TF-IDF. Disini setiap berita yang sudah diinputkan akan dilakukan pra-pemrosesan data. Yaitu Tokenizing, Case folding, Filtering, dan Steaming. Kemudian dilanjutkan proses pembobotan TF-IDF Hal ini dilakukan untuk inialisasi berita dengan cara menghitung jumlah kata yang sama (*Trem Frekuensi*), dan dihitung bobotnya. Tampilan Hasil KNN



No	Data Training	Status Data	Euclidean Distance
1	Sesup Lele Mengandung 3.000 S...	VALID	4
2	Meluruslri Kebenaran Lele Peml...	VALID	4.67768
3	Konsumsi Ikan Lele Mencu Kank...	HOAX	5.46682
4	Benark Sesup Lele Mengandung...	HOAX	6.41567
5	Pakar Bantah Ikan Lele Kandung...	VALID	6.56269

Gambar 5 Tampilan Hasil KNN

Pada gambar 6 merupakan tampilan halaman hasil KNN. Di sini akan diambil 5 buah dataset untuk klasifikasi hasil akhir identifikasi atau $K = 5$. Proses pengambilan dataset tersebut ditentukan melalui hasil perhitungan KNN yaitu *Euclidean Distance*. Perhitungan tersebut digunakan untuk menentukan jarak terdekat yaitu hasil nilai terkecil. Proses ini nanti akan digunakan untuk klasifikasi KNN. Tampilan Hasil Klasifikasi Berikut adalah tampilan hasil Klasifikasi KNN :



No	Berita Identifikasi	Hasil Identifikasi	Aksi
1	Lele Disuab Mengandung Ribuan Sel Kank...	VALID	Detail Hapus
2	Nikmati Manfaat Ikan Lele untuk Kesehata...	VALID	Detail Hapus
3	Bukan Tausuk Iari, Ini yang Seharanya Di...	HOAX	Detail Hapus

Gambar 6 Tampilan Hasil Klasifikasi

Hasil pengujian metode TF-IDF dapat di lihat pada tabel 3.

Tabel 3 Tabel Uji Coba TF-IDF

Uji Coba	Hasil Perbandingan antara hasil pada sistem dan perhitungan manual
Mengambil kata penting	Hasil Sesuai
Menghitung jumlah kata yang sama	Hasil Sesuai
Pembobotan kata	Hasil Sesuai
Pembobotan dokumen	Hasil Sesuai

Hasil pengujian metode KNN dapat di lihat pada tabel 4.

Tabel 4 Tabel Uji Coba Metode KNN

Uji Coba	Hasil Perbandingan antara hasil pada sistem dan perhitungan manual
Menghitung jarak euclidean (KNN)	Hasil Sesuai
Menentukan jarak terdekat	Hasil Sesuai
Klasifikasi KNN	Hasil Sesuai

KESIMPULAN

Aplikasi ini dapat mengurangi penyebaran berita hoax. Sistem dapat mengidentifikasi berita hoax. Selama kategori berita ada dalam dataset. Untuk akurasi sistem identifikasi berita hoax dipengaruhi jumlah dataset yang ada dan pembahasan yang sesuai dengan topiknya.

DAFTAR PUSTAKA

- [1] Hakim, I. (2018). FRAMING KOMPAS.COM TERHADAP BERITA KASUS PERNYATAANHOAX RATNA SARUMPAET. 1.
- [2] Faisal Rahutomo, Ingrid Yanuar Risca Pratiwi, & Diana Mayangsari Ramadhani. (2019).

EKSPERIMEN NAÏVE BAYES PADA DETEKSI BERITA HOAX BERBAHASA INDONESIA. Jurnal Penelitian Komunikasi dan Opini Publik, 1-15.

- [3] Weddiningrum, F. G. (2018). DETEKSI KONTEN HOAX BERBAHASA INDONESIA PADA MEDIA SOSIAL MENGGUNAKAN METODE LEVENSHTAIN DISTANCE
- [4] Jesri Hotman Tridayana Purba. (2019). IDENTIFIKASI DAN KLASIFIKASI SMS HOAX DENGAN KOMBINASI METODE N-GRAM DAN NAÏVE BAYES.
- [5] Faisal Rahutomo, Ingrid Yanuar Risca Pratiwi, & Diana Mayangsari Ramadhani. (2019). EKSPERIMEN NAÏVE BAYES PADA DETEKSI BERITA HOAX BERBAHASA INDONESIA. Jurnal Penelitian Komunikasi dan Opini Publik, 1-15.