

OPTIMIZING NATURAL DISASTER LOCATIONS USING TEXT FILTERING WITH WEB-BASED JARO WINKLER ALGORITHM

Dwiki Jatikusumo^{*1}, Rahmat Rian Hidayat²

¹Informatics Engineering, Computer Science Faculty, Universitas Mercu Buana, Indonesia

²Information and Communication Management, Sekolah Tinggi Multi Media MMTc Yogyakarta, Ministry of Communication and Information Technology, Indonesia

Email: ¹dwiki.jatikusumo@mercubuana.ac.id, ²rianhidayat.r2h@mmtc.ac.id

(Article received: May 25, 2024; Revision: June 23, 2024; published: September 03, 2024)

Abstract

At the end of 2023, several natural disasters were felt, especially in the capital city, namely floods, and the earthquake did occur outside DKI Jakarta but the earthquake was felt. From this incident, it will be made to monitor and optimize the location of earthquakes and floods that occur based on the web. According to the government website, the earthquake also occurred, but there was no flooding. Based on the occurrence of several natural disasters, it is hoped that this research can provide information related to the location of floods, earthquakes, forest fires, and landslides in the region, especially in Indonesia. With the news website source, it is the source of data that will be processed. Furthermore, the percentage level of accuracy obtained by combining from text filtering and Jaro Winkler algorithm.

Keywords: jaro winkler, natural disasters, news, text filtering.

OPTIMASI LOKASI BENCANA ALAM MENGGUNAKAN TEXT FILTERING DENGAN ALGORITMA JARO WINKLER BERBASIS WEB

Abstrak

Pada akhir tahun 2023 terjadi beberapa bencana alam yang terasa khususnya di Ibu Kota, yaitu banjir, dan gempa memang terjadi di luar DKI Jakarta tetapi sampai terasa gempanya. Dari kejadian tersebut akan dibuatkan untuk monitoring dan optimasi lokasi kejadian gempa bumi dan banjir yang terjadi berbasis web. Dari website pemerintah, juga terjadi gempanya saja, tetapi untuk banjir belum ada. Berdasarkan kejadian beberapa bencana alam tersebut, diharapkan penelitian ini dapat memberikan informasi terkait lokasi banjir, gempa bumi, kebakaran hutan, dan tanah longsor di daerah khususnya di daerah Indonesia. Dengan adanya sumber website berita, merupakan sumber data yang akan diproses. Selanjutnya tingkat presentase akurasi yang didapat dalam menggabungkan dari text filtering dan algoritma jaro winkler.

Kata kunci: kata kunci.

1. PENDAHULUAN

Rentetan bencana melanda Indonesia sepanjang 2022. Badan Nasional Penanggulangan Bencana (BNPB) hingga Jumat 16 Desember 2022, mencatat 3.383 kejadian bencana telah terjadi di seluruh daerah [1]. Dari total bencana yang tercatat itu, banjir dengan 1.451 kejadian di berbagai daerah, disusul 1.008 cuaca ekstrem, 620 tanah longsor, 250 kebakaran lahan dan hutan, 27 gempa bumi, dan satu letusan gunung berapi. Provinsi Jawa Barat menjadi wilayah dengan jumlah kejadian paling banyak mencapai 794 kejadian sepanjang 2022. Sedangkan Provinsi Papua menjadi wilayah dengan jumlah kejadian paling sedikit tercatat, yakni delapan kejadian. Keseluruhan

bencana ini telah menelan 832 korban jiwa, 8.718 luka-luka, dan 5.193.417 orang mengungsi.

Kemudian banjir dipengaruhi oleh faktor curah hujan yang tinggi, permukaan tanah lebih rendah dari permukaan laut. Faktor manusia juga ikut berperan, misalnya pemanfaatan lahan yang tidak semestinya seperti bangunan permukiman di daerah bantaran sungai, penggundulan hutan [2].

Selanjutnya, dampak yang ditimbulkan oleh bencana alam seperti tanah longsor, bencana banjir atau gempa bumi, penanggulangan bencana perlu dibuat. Dimana penanggulangan bencana terdiri dari mediasi atau mitigasi, kesiapsiagaan, tanggap darurat, rehabilitasi dan rekonstruksi pada tahap setelah bencana [3].

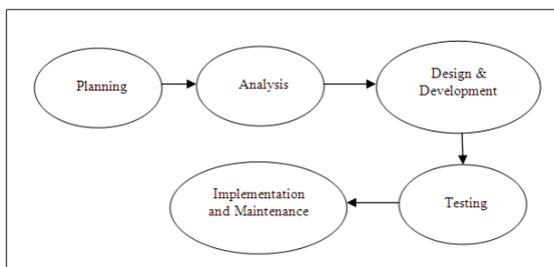
Dari data tersebut, penulis ingin mencoba meneliti mengenai bagaimana cara menentukan lokasi kejadian bencana alam gempa bumi, banjir, kebakaran hutan, dan tanah longsor. Kemudian dengan teknologi telah memudahkan pekerjaan banyak orang, salah satunya dengan kehadiran banyak website berita.

Dari dasar ini, peneliti akan menjadikan isu kejadian bencana alam untuk menjadikan objek suatu penelitian dengan cara mendapatkan lokasi kejadian menggunakan website sebagai sumber data menggunakan text filtering dan algoritma jaro winkler sebagai metode yang digunakan dalam penelitian ini.

Dari hal tersebut, tujuannya dalam penelitian ini mendapatkan lokasi kejadian bencana alam melalui website berita, dan mendapatkan akurasi yang baik dari hasil proses text filtering dan algoritma jaro winkler tersebut lokasi kejadian bencana alam gempa bumi, banjir, kebakaran hutan, dan tanah longsor sekitar 85% lebih dari data sebanyak 5000 sampai 7000 data.

2. METODE PENELITIAN

Metode dalam penelitian ini dari tahapan pengembangan yang digunakan yaitu menggunakan model SDLC (*System Development Life Cycle*) dengan menggunakan model/metode *Web Development Life Cycle* (WDLC) [4]. Model ini dapat dilihat pada gambar 1 berikut.



Gambar 1. Kerangka Kerja Pengembangan Sistem Informasi (WDLC) [4]

Fase planning: Fase pertama dari WDLC adalah perencanaan. Situs Web yang sangat penting seperti itu adalah blok bangunan untuk seluruh situs web. Jika perencanaan salah, maka fase berikutnya dari WDLC juga akan salah jalan. Membuat sebuah keputusan yang baik tentang organisasi situs web dan desain halaman dimulai dengan membuat rencana.

Fase analysis:- Ini adalah serangkaian aktivitas di mana analis mengumpulkan kebutuhan informasi dari pengguna, menganalisisnya secara sistematis dalam bentuk fungsionalitas dari sistem aplikasi, input data kebutuhan dan sumbernya, data output dan kebutuhan presentasi mereka. Analisis sistem mengumpulkan data tentang harapan kinerja pengguna seperti waktu respons yang diharapkan total waktu penyelesaian, dan lain-lain.

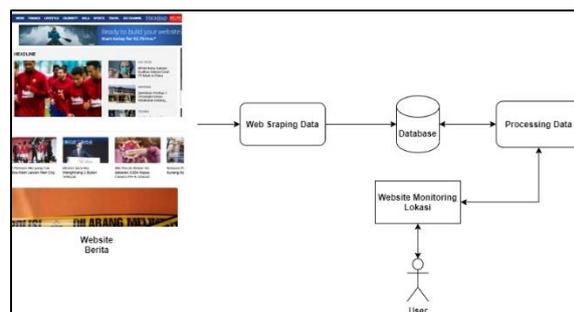
Fase design and development: Hal ini melibatkan mempersiapkan cetak biru dari situs web.

Ini mempersiapkan berbagai diagram representasi dari artefak logis dan fisik untuk dikembangkan selama tahap pengembangan untuk mengikuti. Artefak utama termasuk model data, model proses dan model presentasi. Akhirnya, desain sistem didokumentasikan.

Fase testing: Pada fase ini, Ini untuk menunjukkan kepada tim pengembangan anggota bahwa website bekerja persis untuk memenuhi harapan pengguna persyaratan informasi juga sebagai ekspektasi kinerja. Ini melibatkan perencanaan pengujian, membuat data teks, menjalankan teks berjalan, mencocokkan hasil teks dengan hasil yang diharapkan, menganalisis perbedaan memperbaiki bug dan menguji perbaikan bug berulang kali hingga angka yang memuaskan ketidaksesuaian dihapus.

Fase Implementation and Maintenance: Pada fase ini melibatkan penginstalan situs web di sistem komputer yang melakukan penyiapan data (*Server*, *DBMS*, dan lain-lain.) paralel berjalan dan ditayangkan sebagai kegiatan inti. Ini adalah tahap dimana website pertama datang dalam kontak dengan pengguna, dan pengguna mendapatkan kesempatan untuk bekerja di atasnya untuk pertama kalinya. Selain itu, ini melibatkan langkah terpenting bagi pengguna pengujian penerimaan, yang menandai teknis dan kejadian penting WDLC (*Web Development Life Cycle*).

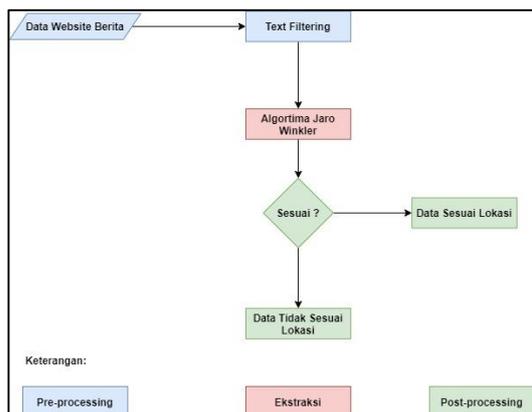
Kemudian pengembangan system menggunakan pada gambar 2 di bawah ini, yang dibutuhkan pada pendeteksi lokasi kejadian bencana alam adalah data dari website berita, kita ambil dari internet kemudian disimpan di dalam database dan akan diperlihatkan peta dari lokasi kejadian bencana alam.



Gambar 2. Arsitektur pendeteksi lokasi kejadian bencana alam

Dalam pengesktrasian data dari data mentah menjadi data yang diolah dengan menentukan lokasi adalah seperti pada gambar 3.

Text filtering, proses pemfilteran meningkatkan efisiensi algoritme jaro winkler dengan secara rasional mengurangi ukuran *file* teks yang diimpor. Ini menghilangkan kata-kata berulang yang tidak mengubah arti kalimat dan tidak memiliki nilai apa pun (misalnya, kata depan dan kata berhenti). Selain itu, menghapus semua *hyperlink*, gambar, rekaman audio, dan video. Kemudian diproses menggunakan algoritma jaro winkler, dan jika tidak sesuai akan dilakukan sampai data sesuai dengan lokasi.



Gambar 3. Alur processing pendeteksi lokasi kejadian bencana alam

3. HASIL DAN PEMBAHASAN

Dari desain sistem yang sudah disiapkan sebelumnya, berikut proses *preprocessing* yang dilakukan. Dalam pengesktrasian data dari data mentah menjadi data yang diolah dengan menentukan lokasi adalah sebagai berikut:

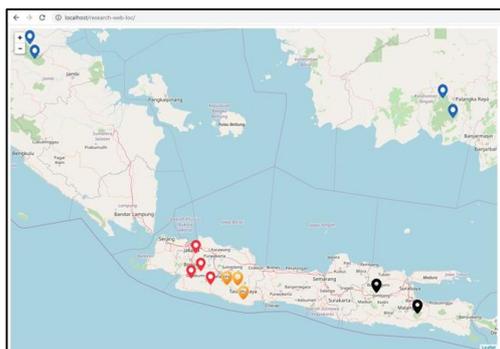
1. *Text filtering*
2. Tidak menggunakan algoritma jaro winkler
3. Menggunakan algoritma jaro winkler

Setelah *text filtering*, selanjutnya menggunakan algoritma jaro winkler yang dilakukan untuk mencari kata-kata lokasi yang sesuai dengan hasil pencarian data dari website berita.

Berikut contoh hasil dari pemetaan dengan menggunakan Openstreetmap API (*Application Programming Interface*) di *browser*. Jadi dari sini mendapatkan letak posisi kejadian bencana alam gempa bumi, banjir, kebakaran hutan, dan tanah longsor dari data website berita yang ada dari bulan Januari sampai April 2023.

Berikut website berita yang diambil, dari data XML (*Extensible Markup Language*) dan disimpan ke dalam database internal aplikasi web peta ini:

1. <https://www.antaranews.com/rss/terkini.xml>
2. <https://www.antaranews.com/rss/top-news.xml>
3. <https://www.antaranews.com/rss/metro.xml>
4. <https://www.suara.com/rss/news>
5. <https://www.suara.com/rss/mostpopular>
6. https://www.bbc.com/indonesia/berita_indonesia/index.xml



Gambar 4. Hasil setelah perhitungan penentuan lokasi dan ditampilkan ke peta

Data diproses menggunakan *preprocessing* seperti tahap pada metodologi riset ini, data yang diolah sebanyak 7336 data data website berita sebelumnya sudah disimpan di dalam database untuk keperluan riset ini, dari bulan Januari sampai April 2023 hasilnya adalah 5279 yang terdapat lokasi kejadian berupa kata lokasi yang dicari.

Selanjutnya adalah perbandingan yang merupakan kejadian atau tidak dengan ujicoba tanpa dan menggunakan *text filtering* dan algoritma jaro winkler. Di sini membedakan dari kata dan kalimat berikut “gempa bumi”, “banjir”, “kebakaran hutan”, “tanah longsor”. Dari beberapa referensi untuk mengetahui akurasi [5], [6], [7], [8].

A. Tidak menggunakan *text filtering* dan algoritma jaro winkler

Tabel 1. Tanpa *text filtering* dan algoritma jaro winkler

No.	Kata dan Kalimat	Kejadian	Tidak Kejadian	Total	Persentase Akurasi
1.	“gempa bumi”	1103	323	1426	77,35%
2.	“kebakaran hutan”	1082	384	1466	73,81%
3.	“banjir”	1181	331	1512	78,12%
4.	“tanah longsor”	1238	359	1597	77,52%

B. Menggunakan *text filtering* dan algoritma jaro winkler

Tabel 2. Menggunakan *text filtering* dan algoritma jaro winkler

No	Kata dan Kalimat	Kejadian	Tidak Kejadian	Total	Persentase Akurasi
1.	“gempa bumi”	1292	134	1426	90,6%
2.	“kebakaran hutan”	1321	145	1466	90,11%
3.	“banjir”	1249	263	1512	82,61%
4.	“tanah longsor”	1383	214	1597	86,6%

4. DISKUSI

Berdasarkan untuk *text filtering* itu sendiri adalah proses pemfilteran meningkatkan efisiensi pencarian data dengan secara rasional mengurangi ukuran file teks yang diimpor. Ini menghilangkan kata-kata berulang yang tidak mengubah arti kalimat dan tidak memiliki nilai apa pun (misalnya, kata depan dan kata berhenti). Selain itu, menghapus semua *hyperlink*, gambar, rekaman audio, dan video [9].

Selanjutnya, algoritma Jaro-Winkler Distance memiliki tiga bagian dasar [10], [11], [12], [13], yaitu:

1. Hitung panjang string.
2. Menentukan karakter yang sama dari kedua string.
3. Tentukan jumlah transposisi.

Algoritma Jaro-Winkler Distance menggunakan rumus untuk menghitung jarak (d_j) antara dua string, yaitu s_1 dan s_2 , yang ditunjukkan oleh persamaan (1).

$$d_j = \frac{1}{3} x \left(\frac{m}{s_1} + \frac{m}{s_2} + \frac{m-t}{m} \right) \quad (1)$$

Catatan:

dj: Jumlah karakter yang sama di kedua string

s1 : Panjang string 1

s2 : Panjang string 2

t : Jumlah transposisi.

4.1. Analisis Hasil Pengujian dari algoritma lain

Berikut adalah algoritma Jaro Winkler [14], Winnowing [15], Naive Bayes [16], KNN (*K-Nearest Neighbors*) [17], SVM (*Support Vector Machine*) [18], berupa tabel 3 sebagai berikut.

Tabel 3. Perbandingan

No.	Nama Algoritma	Akurasi
1.	Algoritma Jaro Winkler	85%
2.	Algoritma Winnowing	85%
3.	Algoritma Naive Bayes	67,64%
4.	Algoritma KNN	80,92%
5.	Algoritma SVM	88%

Dengan adanya data di atas, walaupun algoritma SVM sekitar 88%, dari sini penulis ingin menggunakan algoritma jaro winkler sebagai pengujianya, karena sangat jarang menggunakan algoritma ini sebagai pengujian berdasarkan hasil pencarian referensi terkait.

4.2. Analisis Hasil Pengujian dari algoritma jaro winkler

Dari hasil pengujian yang sudah dilakukan menggunakan tabel di bawah ini:

Tabel 4. Perbandingan Hasil

No.	Kata dan Kalimat	Tanpa <i>text filtering</i> dan algoritma jaro winkler	Menggunakan <i>text filtering</i> dan algoritma jaro winkler
1	"gempa bumi"	77,35%	90,6%
2	"kebakaran hutan"	73,81%	90,11%
3	"banjir"	78,12%	82,61%
4	"tanah longsor"	77,52%	86,6%

Setelah analisa yang dilakukan dengan pengujian yang tidak menggunakan *text filtering* dan algoritma jaro winkler terdapat besaran dengan kata "gempa bumi" sebesar 77,35% dan yang menggunakan sebesar 90,6%. Selanjutnya, kata "kebakaran hutan" yang tidak menggunakan *text filtering* dan algoritma jaro winkler terdapat besaran sebesar 73,81% dan yang menggunakan 90,11%. Selanjutnya, kata "banjir" yang tidak menggunakan *text filtering* dan algoritma jaro winkler terdapat besaran sebesar 78,12% dan yang menggunakan 82,61%. Selanjutnya, kata "tanah longsor" yang tidak menggunakan *text filtering* dan algoritma jaro winkler terdapat besaran sebesar 77,52% dan yang menggunakan 86,6%.

5. KESIMPULAN

Kesimpulan dari penelitian yang telah dilakukan, dari pengolahan data gempa bumi, kebakaran hutan, banjir, dan tanah longsor didapat lokasi kejadiannya yang tidak menggunakan *text filtering* dan algoritma jaro winkler terdapat besaran akurasi rata-rata semua kejadian 76,7%. Kemudian untuk data yang diproses menggunakan *text filtering* dan algoritma jaro winkler rata-rata semua kejadian 87,48%. Dalam perhitungan yang sudah dilakukan bisa dikatakan setidaknya lebih besar menggunakan *text filtering* dan algoritma jaro winkler dibandingkan tidak menggunakan algoritma dan teknik tersebut. Saran untuk pengembangan penelitian selanjutnya dengan kombinasi dari *text filtering* dan algoritma jaro winkler tersebut, memungkinkan untuk membuat algoritma terbaru khususnya untuk algoritma pendeteksi lokasi dari website berita.

UCAPAN TERIMA KASIH

Ucapan terimakasih kepada Lembaga Penelitian dan Pengabdian Masyarakat Universitas Mercu Buana, dengan nomor kontrak: 02-5/117/B-SPK/IV/2023 yang telah memberikan Dana Penelitian ini.

DAFTAR PUSTAKA

- [1] CNN Indonesia, "Rentetan Bencana sepanjang 2022, Gempa Bertubi-tubi di Akhir Tahun." [Online]. Available: <https://www.cnnindonesia.com/nasional/20221226054204-20-891947/rentetan-bencana-sepanjang-2022-gempa-bertubi-tubi-di-akhir-tahun>
- [2] Yoga Pranata, Enrico Setya Damaputra, Pangilinan Gunawan, and Anita Ratnasari, "SISTEM PENDUKUNG KEPUTUSAN PENERIMAAN BANTUAN DANA KORBAN BENCANA ALAM BANJIR," *JSiI (Jurnal Sistem Informasi)*, vol. 9, no. 2, pp. 207–214, Sep. 2022, doi: 10.30656/jsii.v9i2.5110.
- [3] S. Dwiasnati and Y. Devianto, "Classification of Flood Disaster Predictions using the C5.0 and SVM Algorithms based on Flood Disaster Prone Areas," *International Journal of Computer Trends and Technology*, vol. 67, 2019, [Online]. Available: <http://www.ijctjournal.org>
- [4] R. Kamatchia, J. Iyer, and S. Singh, "Software Engineering: Web Development Life cycle," *International Journal of Engineering Research & Technology (IJERT)*, vol. 2, no. 3, pp. 1–4, 2013.
- [5] M. Z. Alksasbeh, B. A. Y. Alqaralleh, T. Abukhalil, A. Abukaraki, T. Al Rawashdeh, and M. Al-Jaafreh, "Smart detection of offensive words in social media using the

- soundex algorithm and permuterm index,” *International Journal of Electrical and Computer Engineering*, vol. 11, no. 5, pp. 4431–4438, 2021, doi: 10.11591/ijece.v11i5.pp4431-4438.
- [6] J. W. Laiya *et al.*, “PENTINGNYA AKURASI DATA DALAM MEMPERTAHANKAN KINERJA PERUSAHAAN PADA PT. MASSINDO SOLARIS NUSANTARA,” 2022.
- [7] M. Sitanggang and E. Simamora, “Increasing Accuracy of Classification in C4.5 Algorithm by Applying Principle Component Analysis for Diabetes Diagnosis,” vol. 6, no. 2, 2022, doi: 10.25217/numerical.v6i2.
- [8] Saruni Dwiasnati and Yudo Devianto, “Optimasi Prediksi Bencana Banjir menggunakan Algoritma SVM untuk penentuan Daerah Rawan Bencana Banjir”.
- [9] T. Tohti, J. Huang, A. Hamdulla, and X. Tan, “Text filtering through multi-pattern matching: A case study of Wu-Manber-Uy on the language of Uyghur,” *Information (Switzerland)*, vol. 10, no. 8, 2019, doi: 10.3390/info10080246.
- [10] T. Efriyanto and M. Hayaty, “Jaro winkler algorithm for measuring similarity online news,” *Jurnal Teknik Informatika (JUTIF)*, vol. 3, no. 4, pp. 975–982, 2022.
- [11] J. Frando, I. Ruslianto, R. Hidayati, J. Rekayasa Sistem Komputer, and J. H. Hadari Nawawi, “PENERAPAN JARO WINKLER DISTANCE DALAM APLIKASI PENGOREKSI KESALAHAN PENULISAN BAHASA INDONESIA BERBASIS WEB [1],” 2019.
- [12] H. Nur Hanani, H. Jayadianti, H. Cahya Rustamaji, and U. Pembangunan Nasional Veteran Yogyakarta, “Fuzzy String Matching untuk Semi-Otomatisasi Pencocokan Kata dengan Algoritma Jaro Winkler Distance pada Dokumen Microsoft Word,” *SEMNASIF*, pp. 145–160, 2021, [Online]. Available: www.myvocabulary.com
- [13] M. A. Yulianto and N. Nurhasanah, “The Hybrid of Jaro-Winkler and Rabin-Karp Algorithm in Detecting Indonesian Text Similarity,” *Jurnal Online Informatika*, vol. 6, no. 1, p. 88, Jun. 2021, doi: 10.15575/join.v6i1.640.
- [14] A. Sanjaya, “OPTIMASI PENCARIAN DATA MENGGUNAKAN TEXT FILTERING DAN ALGORITMA JARO WINKLER,” *Jurnal Ilmiah NERO*, vol. 5, no. 1, pp. 24–29, 2020.
- [15] H. T. Nugroho, “Pengaruh Algoritma Stemming Nazief-Adriani Terhadap Kinerja Algoritma Winnowing Untuk Mendeteksi Plagiarisme Bahasa Indonesia,” *Jurnal ULTIMA Computing*, vol. 9, no. 1, pp. 36–40, 2017, doi: 10.31937/sk.v9i1.572.
- [16] K. K. Kiilu, G. Okeyo, R. Rimiru, and K. Ogada, “Using Naïve Bayes Algorithm in detection of Hate Tweets,” *International Journal of Scientific and Research Publications (IJSRP)*, vol. 8, no. 3, pp. 99–107, Mar. 2018, doi: 10.29322/ijsrp.8.3.2018.p7517.
- [17] N. Alinda Rahmi and R. Wulan Dari, “IMPLEMENTATION OF NATURAL LANGUAGE PROCESSING (NLP) IN CONSUMER SENTIMENT ANALYSIS OF PRODUCT COMMENTS ON THE MARKETPLACE,” *Jurnal Teknik Informatika (JUTIF)*, vol. 5, no. 3, pp. 693–701, 2024, doi: 10.52436/1.jutif.2024.5.3.1666.
- [18] W. Trisnawati and A. Wibowo, “SENTIMENT ANALYSIS OF ICT SERVICE USER USING NAIVE BAYES CLASSIFIER AND SVM METHODS WITH TF-IDF TEXT WEIGHTING,” *Jurnal Teknik Informatika (JUTIF)*, vol. 5, no. 3, pp. 709–719, 2024, doi: 10.52436/1.jutif.2024.5.3.1784.