

## PENERAPAN ALGORITMA COSINE SIMILARITY UNTUK MENINGKATKAN AKURASI TEMU BALIK INFORMASI PADA SISTEM Pencarian DOKUMEN BENCANA ALAM

Mardia Holfiana<sup>1</sup>, Safrizal<sup>2</sup>, Adinda Permata Sari<sup>3</sup>, Dwi Anjani<sup>4</sup>

Sistem Informasi, STMIK Kaputama<sup>1,3,4</sup>, Universitas Muhammadiyah Asahan<sup>2</sup>, Sumatera Utara  
Email: \*[mardiahkholfi@email.com](mailto:mardiahkholfi@email.com)<sup>1</sup>, [rizalsyl75@gmail.com](mailto:rizalsyl75@gmail.com)<sup>2</sup>, [adindapermatasari0506@gmail.com](mailto:adindapermatasari0506@gmail.com)<sup>3</sup>, [dwianjanii18@gmail.com](mailto:dwianjanii18@gmail.com)<sup>4</sup>.

### ABSTRAK

Ketersediaan dokumen bencana alam dalam bentuk digital, seperti berita, laporan, dan arsip resmi, terus mengalami peningkatan sehingga diperlukan sistem pencarian yang mampu menyajikan informasi secara relevan dan akurat. Permasalahan utama pada sistem pencarian konvensional adalah ketidaksesuaian hasil pencarian dengan kebutuhan pengguna karena pencocokan kata kunci yang bersifat sederhana. Penelitian ini bertujuan untuk menerapkan algoritma cosine similarity dalam sistem temu balik informasi guna meningkatkan akurasi pencarian dokumen bencana alam. Dokumen dan kueri direpresentasikan dalam bentuk vektor menggunakan pembobotan Term Frequency-Inverse Document Frequency (TF-IDF), kemudian tingkat kemiripan dihitung menggunakan metode cosine similarity. Proses penelitian meliputi tahap preprocessing teks yang mencakup case folding, tokenizing, dan stopword removal. Hasil pengujian menunjukkan bahwa penerapan algoritma cosine similarity mampu menghasilkan peringkat dokumen yang lebih relevan sesuai dengan kueri pencarian dibandingkan metode pencarian berbasis kata kunci sederhana. Dengan demikian, algoritma cosine similarity dapat menjadi solusi efektif dalam meningkatkan akurasi temu balik informasi pada sistem pencarian dokumen bencana alam.

### Kata Kunci

Temu Balik Informasi, Cosine Similarity, TF-IDF, Dokumen Bencana Alam

### ABSTRACT

*The availability of digital documents related to natural disasters, such as news, reports, and official archives, continues to increase, necessitating a search system capable of presenting relevant and accurate information. A major problem with conventional search systems is the inconsistency of search results with user needs due to simple keyword matching. This study aims to apply the cosine similarity algorithm to an information retrieval system to improve the accuracy of natural disaster document searches. Documents and queries are represented in vector form using Term Frequency-Inverse Document Frequency (TF-IDF) weighting, and then the similarity level is calculated using the cosine similarity method. The research process includes text preprocessing stages, including case folding, tokenizing, and stopword removal. Test results show that the application of the cosine similarity algorithm produces more relevant document rankings according to search queries compared to simple keyword-based search methods. Thus, the cosine similarity algorithm can be an effective solution for improving information retrieval accuracy in natural disaster document retrieval systems.*

### Keywords

*Information Retrieval, Cosine Similarity, TF-IDF, Disaster Documents*

## 1. PENDAHULUAN

Perkembangan teknologi informasi menyebabkan meningkatnya jumlah dokumen digital yang tersedia dan dapat diakses oleh pengguna, termasuk dokumen yang berkaitan dengan bencana alam seperti laporan kejadian, berita daring, dan arsip kebencanaan. Ketersediaan dokumen dalam jumlah besar tersebut menuntut adanya sistem pencarian informasi yang mampu membantu pengguna memperoleh informasi secara cepat dan relevan sesuai dengan kebutuhan pencarian. Tanpa dukungan metode yang tepat, proses pencarian informasi sering kali menghasilkan dokumen yang kurang sesuai dengan konteks yang diinginkan.

Sistem pencarian dokumen pada umumnya masih banyak menerapkan pencocokan kata kunci sederhana. Pendekatan ini memiliki keterbatasan karena hanya mempertimbangkan kesesuaian kata tanpa memperhitungkan bobot dan tingkat kemiripan antar dokumen secara menyeluruh. Akibatnya, hasil pencarian yang ditampilkan sering kali belum mampu memenuhi kebutuhan pengguna secara optimal, khususnya pada pencarian dokumen berbasis teks dalam jumlah besar.

Penerapan metode cosine similarity pada sistem temu balik informasi menjadi salah satu pendekatan yang dapat digunakan untuk meningkatkan relevansi hasil pencarian dokumen. Metode ini bekerja dengan mengukur tingkat kemiripan antara dokumen dan kueri berdasarkan representasi vektor teks, sehingga dokumen yang memiliki tingkat kesamaan lebih tinggi dapat ditampilkan terlebih dahulu. Penggunaan pembobotan kata menggunakan Term Frequency-Inverse Document Frequency (TF-IDF) juga membantu dalam merepresentasikan tingkat kepentingan kata dalam dokumen, sehingga hasil pencarian menjadi lebih akurat (Al Rasyid and Ningsih, 2024).

Dalam konteks pencarian dokumen bencana alam, ketepatan informasi yang diperoleh pengguna memiliki peranan penting karena informasi yang dicari bersifat spesifik dan berkaitan dengan kondisi tertentu. Oleh karena itu, penelitian ini berfokus pada penerapan algoritma cosine similarity dalam sistem pencarian dokumen bencana alam guna meningkatkan akurasi temu balik informasi. Diharapkan sistem yang dikembangkan mampu menghasilkan dokumen yang lebih relevan dan sesuai dengan kebutuhan pengguna.

Penerapan algoritma cosine similarity dalam sistem temu balik informasi telah banyak digunakan untuk meningkatkan ketepatan pencarian dokumen berbasis teks. Dengan mengukur sudut antar vektor dokumen, cosine similarity mampu menilai tingkat kemiripan dokumen secara lebih objektif dibandingkan pencocokan kata kunci biasa, sehingga dokumen yang memiliki kesamaan konteks dapat diprioritaskan dalam hasil pencarian (Wahyuni, Prastiyanto and Suprpto, 2017).

Pendekatan pembobotan TF-IDF juga berperan penting dalam merepresentasikan dokumen ke dalam bentuk numerik. TF-IDF menekankan kata-kata yang bersifat khas dalam dokumen dan mengurangi pengaruh kata umum, sehingga sistem temu balik informasi dapat menghasilkan hasil pencarian yang lebih relevan dan terarah (Septiani and Isabela, 2022). Penggunaan metode TF-IDF dan cosine similarity tidak hanya diterapkan pada dokumen akademik, tetapi juga pada dokumen berita dan arsip digital. Penerapan metode ini terbukti mampu menyaring dokumen berdasarkan tingkat relevansi terhadap kueri pengguna secara lebih efektif (Tanuwijaya *et al.*, 2025). Selain itu, cosine similarity juga banyak dimanfaatkan pada

sistem klasifikasi dan pengelompokan dokumen berbasis web. Dengan perhitungan kemiripan vektor, sistem dapat mengelompokkan dokumen secara otomatis sesuai dengan kesamaan isi dokumen (Anugrah, 2023).

## **2. METODE PENELITIAN**

Penelitian ini menggunakan pendekatan temu balik informasi (Information Retrieval) dengan menerapkan algoritma Cosine Similarity untuk mengukur tingkat kemiripan antara query pengguna dan dokumen bencana alam. Metode ini digunakan untuk meningkatkan akurasi hasil pencarian dokumen berdasarkan kemiripan konten teks. Tahapan penelitian dilakukan secara sistematis mulai dari pengumpulan data hingga evaluasi hasil pencarian. Representasi dokumen dalam bentuk vektor merupakan pendekatan umum dalam sistem information retrieval. Metode Vector Space Model memungkinkan setiap dokumen direpresentasikan sebagai kumpulan nilai numerik sehingga hubungan antar dokumen dapat dianalisis secara matematis (Fauzi and Ginabila, 2019).

Pengembangan sistem temu balik informasi berbasis teks membutuhkan tahapan preprocessing yang tepat agar hasil perhitungan kemiripan menjadi optimal. Tahapan seperti tokenizing, stopword removal, dan stemming bertujuan untuk mengurangi noise serta meningkatkan konsistensi data teks yang diproses (Hapsari, Andoko and Rahmad, 2015). Beberapa penelitian juga mengembangkan sistem temu balik informasi dengan penambahan query expansion untuk meningkatkan cakupan hasil pencarian. Teknik ini bertujuan memperluas makna kueri sehingga sistem dapat menemukan dokumen relevan yang tidak memiliki kesamaan kata secara langsung (Dwi Laxmi and Ali Fauzi, 2019).

Dalam proses perhitungan kemiripan, cosine similarity dipilih karena tidak dipengaruhi oleh panjang dokumen. Hal ini menjadikan metode tersebut sesuai digunakan pada koleksi dokumen dengan ukuran teks yang bervariasi (Harjanto, Nur and Bahtiar, 2012). Perhitungan cosine similarity dilakukan dengan membandingkan nilai dot product antara vektor dokumen dan vektor kueri terhadap panjang masing-masing vektor. Pendekatan ini banyak digunakan karena sederhana namun mampu memberikan hasil perankingan dokumen yang akurat (Akhsa, Ikhwan Burhan and Munandar, 2025). Metode TF-IDF dan cosine similarity juga telah diterapkan dalam berbagai sistem rekomendasi dan pencarian berbasis teks. Hasil penelitian menunjukkan bahwa kombinasi kedua metode tersebut mampu meningkatkan kualitas hasil pencarian dibandingkan metode pencarian konvensional (JATI, 2025).

### **2.1 Jenis dan Pendekatan Penelitian**

Penelitian ini menggunakan jenis penelitian eksperimen dengan pendekatan temu balik informasi (*Information Retrieval*). Pendekatan ini bertujuan untuk menemukan kembali dokumen yang relevan berdasarkan kueri yang diberikan oleh pengguna. Dalam sistem temu balik informasi, dokumen dan kueri diproses serta direpresentasikan dalam bentuk tertentu agar tingkat kesesuaian antara keduanya dapat diukur secara sistematis. Pendekatan ini banyak digunakan dalam penelitian pencarian dokumen berbasis teks karena mampu menghasilkan peringkat dokumen berdasarkan tingkat relevansi (Tanuwijaya et al., 2025).

## 2.2 Dataset Penelitian

Dataset yang digunakan dalam penelitian ini berupa dokumen bencana alam yang berkaitan dengan kejadian kebencanaan di Indonesia. Dokumen terdiri dari laporan kejadian bencana, artikel berita bencana alam, dan arsip kebencanaan. Seluruh dokumen disimpan dalam format teks dan digunakan sebagai basis data dalam sistem pencarian. Dataset ini dipilih untuk menguji kemampuan sistem dalam menemukan dokumen yang relevan terhadap kueri yang bersifat spesifik.

## 2.3 Tahapan Penelitian

Tahapan penelitian dalam sistem temu balik informasi ini terdiri dari beberapa langkah sebagai berikut:

### a. Pengumpulan Data

Pada tahap ini dilakukan pengumpulan dokumen bencana alam yang relevan dengan topik penelitian. Dokumen kemudian diseleksi untuk memastikan kesesuaian isi dengan konteks kebencanaan.

### b. Teks *Preprocessing*

*Preprocessing* teks merupakan tahapan awal dalam pengolahan data teks yang bertujuan untuk membersihkan dan menyiapkan dokumen sebelum dilakukan proses pembobotan dan perhitungan kemiripan. *Preprocessing* diperlukan untuk mengurangi noise serta menyamakan bentuk kata sehingga hasil analisis menjadi lebih akurat. Tahapan *preprocessing* yang dilakukan dalam penelitian ini meliputi *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Proses *preprocessing* terbukti mampu meningkatkan kualitas representasi dokumen dalam sistem temu balik informasi karena kata-kata yang tidak relevan dapat dihilangkan sejak tahap awal (Tanuwijaya et al., 2025).

### c. *Case Folding*

*Case folding* dilakukan dengan mengubah seluruh huruf pada dokumen menjadi huruf kecil. Tahapan ini bertujuan untuk menghindari perbedaan makna akibat penggunaan huruf kapital dan nonkapital dalam teks.

### d. *Tokenizing*

*Tokenizing* merupakan proses pemecahan teks menjadi unit-unit kata atau token. Token hasil proses ini akan digunakan sebagai dasar dalam pembobotan kata pada tahap selanjutnya.

### e. *Stopword Removal*

*Stopword removal* dilakukan dengan menghapus kata-kata umum yang sering muncul tetapi tidak memiliki makna penting dalam pencarian, seperti "dan", "yang", dan "di". Penghapusan *stopword* bertujuan untuk meningkatkan efisiensi dan akurasi sistem temu balik informasi.

### f. *Stemming*

*Stemming* merupakan proses mengubah kata ke bentuk dasarnya. Tahapan ini bertujuan untuk menyatukan berbagai variasi kata yang memiliki makna sama sehingga dapat meningkatkan tingkat kesesuaian antara dokumen dan kueri.

### g. Pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF)

Pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) digunakan untuk memberikan bobot pada setiap kata dalam dokumen berdasarkan tingkat kemunculannya. *Term Frequency* (TF) menunjukkan seberapa sering suatu kata muncul dalam dokumen, sedangkan *Inverse Document Frequency* (IDF)

menunjukkan tingkat kepentingan kata tersebut di seluruh koleksi dokumen. Metode TF-IDF mampu menurunkan bobot kata-kata yang sering muncul di banyak dokumen dan meningkatkan bobot kata yang bersifat spesifik, sehingga efektif digunakan dalam sistem temu balik informasi.

#### h. Representasi Dokumen dalam Bentuk Vektor

Setelah dilakukan pembobotan TF-IDF, setiap dokumen dan kueri direpresentasikan dalam bentuk vektor numerik. Representasi vektor memungkinkan proses perhitungan kemiripan antara dokumen dan kueri dilakukan secara matematis. Pendekatan ini umum digunakan dalam sistem pencarian dokumen karena mampu menggambarkan karakteristik dokumen secara kuantitatif (Nico, Budiyo and Fatimah, 2022). Dalam sistem temu balik informasi, proses transformasi dokumen teks ke dalam bentuk vektor bertujuan untuk menyederhanakan data tekstual agar dapat diolah secara komputasional. Setiap dokumen direpresentasikan sebagai kumpulan nilai numerik yang merefleksikan bobot term tertentu, sehingga sistem dapat melakukan perbandingan antar dokumen secara lebih terukur. Pendekatan ini memudahkan proses analisis kesamaan dokumen dan mendukung efisiensi pemrosesan data pada koleksi dokumen berukuran besar (Wijaya, 2013). Proses interpretasi nilai kemiripan ini bertujuan untuk membantu sistem dalam mengelompokkan dokumen berdasarkan tingkat relevansi secara bertahap. Dengan demikian, dokumen yang paling sesuai dapat ditampilkan terlebih dahulu kepada pengguna, diikuti oleh dokumen dengan tingkat kemiripan yang lebih rendah (Utomo et al., 2022).

#### i. *Cosine Similarity*

Cosine Similarity merupakan metode yang digunakan untuk mengukur tingkat kemiripan antara dua vektor dalam ruang multidimensi. Metode ini mengukur sudut antara vektor dokumen dan vektor kueri tanpa dipengaruhi oleh panjang dokumen. Nilai Cosine Similarity berada pada rentang 0 hingga 1, di mana nilai yang mendekati 1 menunjukkan tingkat kemiripan yang semakin tinggi. Oleh karena itu, Cosine Similarity sangat sesuai digunakan dalam pengukuran kesamaan dokumen teks dalam sistem temu balik informasi (Supiyanto and Sriyono, 2023). Metode cosine similarity digunakan untuk menentukan tingkat kedekatan antara dokumen dan kueri dengan membandingkan arah vektor yang terbentuk dari hasil pembobotan term. Pendekatan ini memungkinkan sistem pencarian menilai kesesuaian dokumen berdasarkan kemiripan representasi vektornya, sehingga dokumen yang memiliki keterkaitan konteks lebih tinggi dapat diprioritaskan dalam hasil pencarian. Penggunaan cosine similarity juga membantu menghasilkan urutan dokumen yang lebih relevan pada sistem pencarian berbasis teks (Tirtana, Zulkarnain and Listio, 2019).

Dalam pengembangan sistem temu balik informasi, nilai cosine similarity sering digunakan sebagai dasar dalam proses perankingan dokumen. Dengan memanfaatkan nilai kemiripan yang dihasilkan, sistem dapat mengurutkan dokumen berdasarkan tingkat relevansinya terhadap kueri pencarian. Optimalisasi pada tahap perankingan, khususnya yang dikombinasikan dengan pembobotan TF-IDF, dapat membantu meningkatkan ketepatan hasil pencarian dan menampilkan dokumen yang lebih sesuai dengan kebutuhan pengguna (Aurelia, Irsyad and Rahman, 2025)

#### j. Perangkingan Dokumen

Hasil perhitungan Cosine Similarity digunakan sebagai dasar dalam proses perangkangan dokumen. Dokumen diurutkan berdasarkan nilai kemiripan tertinggi hingga terendah. Dokumen dengan nilai Cosine Similarity tertinggi dianggap sebagai dokumen yang paling relevan terhadap kueri pencarian yang diberikan oleh pengguna.

### 3. HASIL DAN PEMBAHASAN

Bab ini membahas hasil implementasi metode TF-IDF dan Cosine Similarity pada sistem temu balik informasi dokumen bencana alam.

#### 3.1 Deskripsi Kasus

Pengujian dilakukan untuk mengetahui tingkat kemiripan antara kueri pencarian dengan dokumen bencana alam menggunakan metode TF-IDF dan Cosine Similarity. Dataset yang digunakan terdiri dari 20 dokumen bencana alam yang relevan dengan konteks kebencanaan di Indonesia. Kueri yang digunakan dalam pengujian ini adalah:

Querri (Q) : "banjir bandang sumatera"

#### 3.2 Data Dokumen

Dokumen yang digunakan berupa teks singkat yang merepresentasikan isi atau judul dokumen bencana alam. Daftar dokumen yang digunakan dapat dilihat pada Tabel 3.1.

**Tabel 1. Data Dokumen**

| <b>Kode</b> | <b>Isi Dokumen</b>                        |
|-------------|-------------------------------------------|
| D1          | banjir bandang sumatera barat             |
| D2          | banjir bandang sumatera utara             |
| D3          | banjir sumatera selatan                   |
| D4          | banjir bandang aceh                       |
| D5          | banjir kota medan                         |
| D6          | banjir bandang padang                     |
| D7          | banjir bandang sumatera barat hujan       |
| D8          | banjir bandang sumatera utara rumah rusak |
| D9          | banjir sumatera barat                     |
| D10         | banjir bandang                            |
| D11         | gempa bumi sumatera barat                 |
| D12         | gempa bumi aceh                           |
| D13         | gempa bumi sumatera utara                 |
| D14         | longsor sumatera barat                    |
| D15         | longsor hujan deras                       |
| D16         | tsunami aceh                              |
| D17         | tsunami sumatera barat                    |
| D18         | kebakaran hutan sumatera                  |
| D19         | angin puting beliung jawa                 |
| D20         | erupsi gunung Sinabung                    |

### 3.3 Penentuan Term

Berdasarkan hasil preprocessing terhadap kueri dan seluruh dokumen, diperoleh term utama yang digunakan dalam proses perhitungan, yaitu:

$T = \{\text{banjir, bandang, sumatera}\}$ .

### 3.4 Term Frequency (TF)

Nilai Term Frequency (TF) menunjukkan jumlah kemunculan suatu term dalam dokumen.

TF Query

**Tabel 2. Data Term Frequency TF**

| Term     | Q |
|----------|---|
| banjir   | 1 |
| bandang  | 1 |
| sumatera | 1 |

TF Seluruh Dokumen

**Tabel 3. TF dokumen mengikuti data perhitungan sebelumnya**

| Dok | banjir | bandang | sumatera |
|-----|--------|---------|----------|
| D1  | 1      | 1       | 1        |
| D2  | 1      | 1       | 1        |
| D3  | 1      | 0       | 1        |
| D4  | 1      | 1       | 0        |
| D5  | 1      | 0       | 0        |
| D6  | 1      | 1       | 0        |
| D7  | 1      | 1       | 1        |
| D8  | 1      | 1       | 1        |
| D9  | 1      | 0       | 1        |
| D10 | 1      | 1       | 0        |
| D11 | 0      | 0       | 1        |
| D12 | 0      | 0       | 0        |
| D13 | 0      | 0       | 1        |
| D14 | 0      | 0       | 1        |
| D15 | 0      | 0       | 0        |
| D16 | 0      | 0       | 0        |
| D17 | 0      | 0       | 1        |
| D18 | 0      | 0       | 1        |
| D19 | 0      | 0       | 0        |

### 3.5 Perhitungan IDF

Jumlah dokumen ( $N$ ) = 20. Nilai IDF dihitung menggunakan rumus:

$IDF = \log(N/df)$

**Tabel 3. Perhitungan IDF**

| Term     | df | IDF                   |
|----------|----|-----------------------|
| banjir   | 10 | $\log(20/10) = 0,301$ |
| bandang  | 7  | $\log(20/7) = 0,456$  |
| sumatera | 9  | $\log(20/9) = 0,346$  |

### 3.6 TF-IDF Query

**Tabel 4. TF-IDF Query**

| Term     | TF | IDF   | TF-IDF |
|----------|----|-------|--------|
| banjir   | 1  | 0,301 | 0,301  |
| bandang  | 1  | 0,456 | 0,456  |
| sumatera | 1  | 0,346 | 0,346  |

Nilai TF-IDF diperoleh dari hasil perkalian TF dan IDF. Vektor TF-IDF untuk query adalah:

Vector Q = (0,301 ; 0,456 ; 0,346)

Panjang vektor kueri  $|Q| = \sqrt{0,301^2 + 0,456^2 + 0,346^2} = 0,646$

### 3.7 TF-IDF Seluruh Dokumen

**Tabel 5. TF-IDF Seluruh Dokumen**

| Dok | banjir | bandang | sumatera |
|-----|--------|---------|----------|
| D1  | 0,301  | 0,456   | 0,346    |
| D2  | 0,301  | 0,456   | 0,346    |
| D3  | 0,301  | 0       | 0,346    |
| D4  | 0,301  | 0,456   | 0        |
| D5  | 0,301  | 0       | 0        |
| D6  | 0,301  | 0,456   | 0        |
| D7  | 0,301  | 0,456   | 0,346    |
| D8  | 0,301  | 0,456   | 0,346    |
| D9  | 0,301  | 0       | 0,346    |
| D10 | 0,301  | 0,456   | 0        |
| D11 | 0      | 0       | 0,346    |
| D12 | 0      | 0       | 0        |
| D13 | 0      | 0       | 0,346    |
| D14 | 0      | 0       | 0,346    |
| D15 | 0      | 0       | 0        |
| D16 | 0      | 0       | 0        |
| D17 | 0      | 0       | 0,346    |
| D18 | 0      | 0       | 0,346    |
| D19 | 0      | 0       | 0        |
| D20 | 0      | 0       | 0        |

### 3.8 Dot Product (Q.D)

Dot product dilakukan antara vektor kueri dan vektor dokumen untuk mengetahui tingkat kesamaan awal antar vektor.

(Hasil dot product mengikuti data perhitungan sebelumnya)

**Tabel 6. Dot Product (Q.D)**

| Dok | Perhitungan Q · D               | Nilai |
|-----|---------------------------------|-------|
| D1  | $(0,301^2 + 0,456^2 + 0,346^2)$ | 0,417 |
| D2  | $(0,301^2 + 0,456^2 + 0,346^2)$ | 0,417 |
| D3  | $(0,301^2 + 0 + 0,346^2)$       | 0,211 |
| D4  | $(0,301^2 + 0,456^2 + 0)$       | 0,299 |
| D5  | $(0,301^2)$                     | 0,091 |
| D6  | $(0,301^2 + 0,456^2)$           | 0,299 |

|     |                                 |       |
|-----|---------------------------------|-------|
| D7  | $(0,301^2 + 0,456^2 + 0,346^2)$ | 0,417 |
| D8  | $(0,301^2 + 0,456^2 + 0,346^2)$ | 0,417 |
| D9  | $(0,301^2 + 0 + 0,346^2)$       | 0,211 |
| D10 | $(0,301^2 + 0,456^2)$           | 0,299 |
| D11 | $(0,346^2)$                     | 0,120 |
| D12 | 0                               | 0     |
| D13 | $(0,346^2)$                     | 0,120 |
| D14 | $(0,346^2)$                     | 0,120 |
| D15 | 0                               | 0     |
| D16 | 0                               | 0     |
| D17 | $(0,346^2)$                     | 0,120 |
| D18 | $(0,346^2)$                     | 0,120 |
| D19 | 0                               | 0     |
| D20 | 0                               | 0     |

### 3.9 Perhitungan Cosine Similarity (Hasil Akhir)

Cosine Similarity digunakan untuk menghitung tingkat kemiripan antara kueri dan dokumen menggunakan rumus:

$$\text{CosSim} = (Q \cdot D) / (|Q| \times |D|)$$

Hasil perhitungan Cosine Similarity menunjukkan bahwa dokumen yang memiliki nilai tertinggi merupakan dokumen yang paling relevan dengan kueri pencarian.

**Tabel 7. Perhitungan Consine Similarity**

| Dok | Q·D   | D     | Cosine Similarity |
|-----|-------|-------|-------------------|
| D1  | 0,417 | 0,646 | 1,000             |
| D2  | 0,417 | 0,646 | 1,000             |
| D3  | 0,211 | 0,459 | 0,712             |
| D4  | 0,299 | 0,545 | 0,849             |
| D5  | 0,091 | 0,301 | 0,469             |
| D6  | 0,299 | 0,545 | 0,849             |
| D7  | 0,417 | 0,646 | 1,000             |
| D8  | 0,417 | 0,646 | 1,000             |
| D9  | 0,211 | 0,459 | 0,712             |
| D10 | 0,299 | 0,545 | 0,849             |
| D11 | 0,120 | 0,346 | 0,536             |
| D12 | 0     | 0     | 0,000             |
| D13 | 0,120 | 0,346 | 0,536             |
| D14 | 0,120 | 0,346 | 0,536             |
| D15 | 0     | 0     | 0,000             |
| D16 | 0     | 0     | 0,000             |
| D17 | 0,120 | 0,346 | 0,536             |
| D18 | 0,120 | 0,346 | 0,536             |
| D19 | 0     | 0     | 0,000             |
| D20 | 0     | 0     | 0,000             |

### 3.10 Hasil Akhir Perangkingan

1. D1, D2, D7, D8 → Sangat relevan
2. D4, D6, D10 → Relevan
3. D3, D9 → Cukup relevan
4. D11, D13, D14, D17, D18 → Relevansi rendah
5. D12, D15, D16, D19, D20 → Tidak relevan

## 4. KESIMPULAN

Penelitian ini bertujuan untuk menerapkan algoritma Cosine Similarity dalam sistem temu balik informasi guna meningkatkan akurasi pencarian dokumen bencana alam. Dokumen dan kueri direpresentasikan dalam bentuk vektor menggunakan metode pembobotan TF-IDF, kemudian tingkat kemiripan dihitung untuk menentukan peringkat dokumen yang paling relevan terhadap kueri pencarian.

Berdasarkan hasil pengujian terhadap 20 dokumen bencana alam dengan kueri “banjir bandang sumatera”, diperoleh hasil bahwa dokumen yang mengandung kesesuaian term tertinggi memiliki nilai Cosine Similarity yang lebih besar dan menempati peringkat teratas. Dokumen D1, D2, D7, dan D8 menunjukkan tingkat kemiripan tertinggi dan dikategorikan sebagai dokumen sangat relevan, sedangkan dokumen yang tidak memiliki kesamaan term memperoleh nilai kemiripan nol dan dinyatakan tidak relevan.

Hasil penelitian menunjukkan bahwa penerapan algoritma Cosine Similarity yang dikombinasikan dengan pembobotan TF-IDF mampu meningkatkan relevansi hasil pencarian dibandingkan metode pencarian berbasis pencocokan kata kunci sederhana. Dengan demikian, metode ini efektif digunakan dalam sistem pencarian dokumen bencana alam.

Secara keseluruhan, penerapan pembobotan TF-IDF dan cosine similarity mampu membantu sistem dalam mengidentifikasi serta mengurutkan dokumen bencana alam berdasarkan tingkat kemiripannya terhadap kueri pencarian. Pendekatan ini menghasilkan hasil pencarian yang lebih terarah dan relevan, sehingga memudahkan pengguna dalam memperoleh informasi kebencanaan yang sesuai dengan kebutuhan.

## 5. DAFTAR PUSTAKA

- Akhsa, A.T.P.D., Ikhwan Burhan, M. and Munandar, A. (2025) ‘Integrasi OCR dan TF-IDF untuk Metadata Otomatis pada Pencarian Dokumen Digital’, *Jurnal Fasilkom*, 15(2), pp. 304–311.
- Anugrah, R.R. (2023) ‘PENERAPAN COSINE SIMILARITY DAN PEMBOBOTAN TF-IDF UNTUK KLASIFIKASI PENGADUAN MASYARAKAT BERBASIS WEB (Studi Kasus: BAGWASSIDIK DITRESKRIMUM POLDA KALBAR)’, *Coding Jurnal Komputer dan Aplikasi*, 11(1), p. 100. Available at: <https://doi.org/10.26418/coding.v11i1.55598>.
- Aurelia, R., Irsyad, H. and Rahman, A. (2025) ‘Optimasi Rangking Dokumen Dengan Modifikasi Tf-Idf Berbasis Waktu Publikasi Dan Cosine Similarity’, *Jiscom*, 3.
- Dwi Laxmi, M. and Ali Fauzi, M. (2019) ‘Query Expansion Pada Sistem Temu Kembali

- Informasi Berbahasa Indonesia Dengan Metode Pembobotan TF-IDF Dan Algoritme Cosine Similarity Berbasis Wordnet', *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(1), pp. 2548-964.
- Fauzi, A. and Ginabila, G. (2019) 'Information Retrieval System Pada File Pencarian Dokumen Tesis Berbasis Text Menggunakan Metode Vector Space Model', *Jurnal Pilar Nusa Mandiri*, 15(1), pp. 41-46. Available at: <https://doi.org/10.33480/pilar.v15i1.61>.
- Hapsari, I.T., Andoko, B.S. and Rahmad, C. (2015) 'Aplikasi Information Retrieval Untuk Pencarian Dokumen Laporan Penelitian', *Jurnal Informatika Polinema*, 1(3), p. 23. Available at: <https://doi.org/10.33795/jip.v1i3.109>.
- Harjanto, D.S., Nur, E.S. and Bahtiar, N. (2012) 'Sistem Temu Kembali Informasi Pada Dokumen Teks Mengutamakan Metode Term Frequency Inverse Document Frequency (TF-IDF)', *Jurnal Sains dan Matematika*, 20, pp. 64-70.
- Nico, Budiyanto, U. and Fatimah, T. (2022) 'Implementasi Algoritma Pembobotan TF-IDF dan Cosine Similarity untuk Penetapan Kategori Artikel pada Website Universitas Budi Luhur', *Jurnal TICOM: Technology of Information and Communication*, 10(3), pp. 218-222.
- Al Rasyid, R. and Ningsih, D.H.U. (2024) 'Penerapan Algoritma TF-IDF dan Cosine Similarity untuk Query Pencarian Pada Dataset Destinasi Wisata', *Jurnal JTik (Jurnal Teknologi Informasi dan Komunikasi)*, 8(1), pp. 170-178. Available at: <https://doi.org/10.35870/jtik.v8i1.1416>.
- Septiani, D. and Isabela, I. (2022) 'Analisis Term Frequency Inverse Document Frequency (Tf-Idf) Dalam Temu Kembali Informasi Pada Dokumen Teks', *Sintesia*, 1, pp. 81-88.
- Supiyanto and Sriyono (2023) 'Cosine Similarity Method To Detect Similarities In Text Documents', *Jurnal MIPA dan Pengajarannya(SAINS)*, 1(1), pp. 1-7.
- Tanuwijaya, W. *et al.* (2025) 'Implementasi TF-IDF dan Cosine Similarity untuk Penyaringan Dokumen Berita Program Makan Siang Gratis Pemerintah Indonesia', 6(2), pp. 322-334.
- Tirtana, A., Zulkarnain, A. and Listio, Y.D. (2019) 'Pembuatan Sistem Pencarian Pekerjaan Menggunakan TF-IDF', *Jurnal Ilmiah Teknologi Informasi Asia*, 13(2), pp. 91-100. Available at: <https://doi.org/10.32815/jitika.v13i2.389>.
- Wahyuni, R.T., Prastiyanto, D. and Suprpto, E. (2017) 'Application of Cosine Similarity Algorithm and TF-IDF Weighting in Thesis Document Classification System', *Jurnal Teknik Elektro Universitas Negeri Semarang*, 9(1), pp. 18-23.
- Wijaya, E.T. (2013) 'Perancangan Information Retrieval (IR) Berbasis Term Frequency-Inverse Document Frequency (TF-IDF) Untuk Peringkasan Teks Tugas Khusus Berbahasa Indonesia', *Jurnal Ilmiah Teknologi Informasi Asia*, 7(1), pp. 79-93.