

## ANALISIS TERM FREQUENCY INVERSE DOCUMENT FREQUENCY (TF-IDF) DALAM TEMU KEMBALI INFORMASI PADA DOKUMEN TEKS

Dwi Septiani<sup>1)</sup>, Ica Isabela<sup>2)</sup>

<sup>1</sup> Pendidikan Teknik Informatika dan Komputer, Fakultas Teknik, Universitas Negeri Jakarta

<sup>2</sup> Pendidikan Teknik Informatika dan Komputer, Fakultas Teknik, Universitas Negeri Jakarta  
email: [DwiSeptiani\\_1512620002@mhs.unj.ac.id](mailto:DwiSeptiani_1512620002@mhs.unj.ac.id), [IcaIsabela\\_1512620015@mhs.unj.ac.id](mailto:IcaIsabela_1512620015@mhs.unj.ac.id)

### Abstract

*Information retrieval system (information retrieval system) is a system that is used to find information that is relevant to the needs of its users, by implementing this system the problem of searching for text document information can provide relevant results according to user needs. This research method uses the Term Frequency-Inverse Document Frequency (TF-IDF) method, the Term Frequency-Inverse Document Frequency (TF-IDF) method is one of the techniques used in text processing and natural language modeling. The main objective of the TF-IDF method is to evaluate how important a word (term) is in a document in the context of a larger collection of documents.*

**Keywords:** *Information retrieval system and Term Frequency-Inverse Document Frequency (TF-IDF) method*

### Abstrak

Sistem temu kembali informasi (information retrieval system) merupakan sistem yang digunakan untuk menemukan informasi yang relevan dengan kebutuhan dari penggunaannya, dengan menerapkan sistem tersebut permasalahan pencarian informasi dokumen teks bisa memberikan hasil yang relevan sesuai kebutuhan pengguna. Metode penelitian ini menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF), metode Term Frequency-Inverse Document Frequency (TF-IDF) adalah salah satu teknik yang digunakan dalam pengolahan teks dan pemodelan bahasa alami. Tujuan utama dari metode TF-IDF adalah untuk mengevaluasi seberapa penting suatu kata (term) dalam sebuah dokumen dalam konteks koleksi dokumen yang lebih besar.

**Kata Kunci:** Sistem temu kembali informasi dan metode Term Frequency-Inverse Document Frequency (TF-IDF)

### 1. PENDAHULUAN

Saat ini, sistem temu kembali informasi (*Information Retrieval*) pada dokumen teks telah mengalami beberapa perkembangan yang signifikan. Seperti representasi dokumen yang lebih maju, dimana terdapat peningkatan dalam cara dokumen-dokumen teks direpresentasikan untuk memungkinkan temu kembali informasi yang lebih baik. Pendekatan yang umum digunakan adalah pemodelan vektor, di mana dokumen direpresentasikan sebagai vektor dalam ruang multidimensi. Teknik seperti Word2Vec, GloVe, atau BERT telah mengubah cara dokumen-dokumen diinterpretasikan dan memungkinkan sistem temu kembali informasi

untuk memahami konteks dan hubungan antara kata-kata.

Selain itu, temu kembali informasi juga meningkat dalam pencocokan query, algoritma pencocokan query terus ditingkatkan untuk memberikan hasil yang lebih relevan. Pencocokan query tidak lagi hanya berdasarkan kata kunci yang tepat, tetapi juga mempertimbangkan konteks, sinonim, dan relasi antar kata. Model seperti BM25 (Okapi Best Matching 25), Language Models, dan Transformer-based models telah memperbaiki kemampuan sistem dalam memahami maksud pengguna dan menghasilkan hasil yang lebih akurat. Dengan perkembangan ini, sistem temu

kembali informasi pada dokumen teks telah menjadi lebih canggih dalam memahami maksud pengguna, menghasilkan hasil yang lebih relevan, dan memberikan pengalaman pencarian yang lebih personalisasi.

Berkaitan dengan hal tersebut, sistem temu kembali informasi bertujuan menghasilkan dokumen yang paling relevan berdasarkan *keyword* pada *query* yang diberikan pengguna. Sistem temu kembali informasi ini digunakan untuk mengurangi informasi yang terlalu banyak sehingga sulit untuk dikelola.

Sistem temu kembali informasi terdiri dari tiga elemen utama, yakni bagian masukan (*input*), pemroses (*processor*), dan bagian keluaran (*output*). Pengguna adalah pemilik kebutuhan akan informasi, yang kemudian menerjemahkan kebutuhannya menjadi sebuah *query*. Untuk dapat diolah oleh pemroses, *input* harus memiliki representasi yang akurat dari setiap dokumen dan *query*. Pemroses bertanggung jawab untuk mengorganisir informasi tersebut dalam format yang sesuai, seperti melalui proses pengindeksan dan klasifikasi, serta melakukan proses temu kembali dengan menerapkan strategi pencarian sebagai respons terhadap *query* yang diberikan. *Output* adalah keluaran yang diberikan oleh pemroses, biasanya berbentuk informasi tentang suatu dokumen, dokumen itu sendiri, dan acuan ke dokumen lain (M fitri, 2013).

Dalam penelitian kali ini penulis menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Metode *Term Frequency-Inverse Document Frequency* (TF-IDF) adalah salah satu teknik yang digunakan dalam pengolahan teks dan pemodelan bahasa alami.

## 2. METODE PENELITIAN

Metode *Term Frequency-Inverse Document Frequency* (TF-IDF) adalah salah satu teknik yang digunakan dalam pengolahan teks dan pemodelan bahasa alami. Tujuan utama dari metode TF-IDF adalah untuk mengevaluasi seberapa penting suatu kata (*term*) dalam sebuah dokumen dalam konteks koleksi dokumen yang lebih besar.

Metode TF-IDF memperhitungkan dua faktor penting:

- a. *Term Frequency* (TF): Mengukur seberapa sering suatu kata muncul dalam sebuah dokumen. Pendekatan umum untuk menghitung TF adalah dengan menghitung jumlah kemunculan kata tersebut dibagi dengan jumlah total kata dalam dokumen. Dalam beberapa kasus, TF dapat diubah dengan menerapkan skema penimbangan yang lebih kompleks.
- b. *Inverse Document Frequency* (IDF): Mengukur seberapa penting suatu kata dalam konteks koleksi dokumen yang lebih besar. Kata-kata yang muncul lebih jarang di seluruh koleksi dokumen cenderung memiliki IDF yang lebih tinggi. IDF dihitung dengan membagi jumlah total dokumen dalam koleksi dengan jumlah dokumen yang mengandung kata tersebut. Hasilnya kemudian diambil logaritma untuk memperhalus skala.

Dalam metode TF-IDF, nilai TF dan IDF dikalikan bersama-sama untuk menghasilkan bobot kata (*term weight*) untuk setiap kata dalam sebuah dokumen. Bobot ini mencerminkan tingkat pentingnya kata dalam dokumen tersebut dibandingkan dengan koleksi dokumen yang lebih besar.

Rumus metode *Term Frequency-Inverse Document Frequency* (TF-IDF):

$$tf = 0,5 + 0,5 \times \frac{tf}{\max(tf)}$$

$$idf_t = \log \left( \frac{D}{df_t} \right)$$

$$W_{d,t} = tf_{d,t} \times idf_{d,t}$$

Keterangan :

D = dokumen ke-d

t = term ke-t dari dokumen

W = bobot ke-d terhadap term ke-t

tf = jumlah kemunculan term i dalam dokumen

idf = *Inversed Document Frequency*

df = banyak dokumen yang mengandung term i

## 3. HASIL DAN PEMBAHASAN

### 3.1 Sistem Temu Kembali Informasi

#### 3.1.1 Pengertian Sistem Temu Kembali Informasi

Sistem Temu Kembali Informasi adalah sistem komputer yang dirancang untuk mencari, mengambil, dan mengorganisir informasi yang relevan dari koleksi dokumen atau sumber informasi lainnya berdasarkan permintaan pengguna. Sistem ini bertujuan untuk memberikan respons yang akurat dan efektif terhadap kebutuhan informasi pengguna.

Sistem temu kembali informasi bekerja dengan menerima *query* atau kata kunci pencarian dari pengguna dan kemudian mencocokkan *query* tersebut dengan dokumen-dokumen dalam koleksi yang relevan. Proses pencocokan ini melibatkan teknik pengindeksan dokumen, di mana dokumen-dokumen diberi label atau ditandai dengan metadata untuk memfasilitasi pencarian dan pemrosesan lebih lanjut. Selain itu, sistem menggunakan berbagai metode dan algoritma untuk menghitung tingkat relevansi atau kemiripan antara *query* dan dokumen, sehingga dapat memprioritaskan dan mempresentasikan hasil pencarian secara sesuai.

### 3.1.2 Komponen Sistem Temu Kembali Informasi

Sistem temu kembali informasi mempunyai tiga komponen utama, yaitu:

- 1) Masukan (*Input*), komponen ini mencakup *query* atau kata kunci pencarian yang diberikan oleh pengguna. *Query* ini berfungsi sebagai permintaan untuk mencari informasi yang relevan.
- 2) Pemroses (*Processor*), komponen ini bertanggung jawab untuk mengolah *query* atau kata kunci pencarian yang diberikan oleh pengguna. Pemroses akan melakukan pemrosesan lanjutan, seperti pemrosesan bahasa alami untuk memahami maksud dari *query*, pengindeksan dokumen, perhitungan relevansi, dan strategi pencarian yang sesuai.
- 3) Keluaran (*Output*), komponen ini mencakup hasil dari sistem temu kembali informasi. Keluaran dapat berupa daftar dokumen yang relevan

dengan *query*, ringkasan informasi, atau bentuk lainnya yang memudahkan pengguna dalam mengakses informasi yang mereka cari.

### 3.1.3 Jenis Sistem Temu Kembali Informasi

Sistem temu kembali informasi terdapat dalam beberapa jenis yaitu sebagai berikut

- 1) Sistem Temu Kembali Informasi Teks  
Jenis sistem ini fokus pada pencarian dan temu kembali dokumen teks, seperti artikel, laporan, dan dokumen lainnya. Algoritma dan teknik pemrosesan bahasa alami sering digunakan untuk memahami maksud pengguna dan menemukan dokumen yang paling relevan.
- 2) Sistem Temu Kembali Informasi Multimedia  
Jenis sistem ini mencakup pencarian dan temu kembali informasi multimedia seperti gambar, video, dan audio. Teknik pencocokan gambar, pengenalan wajah, dan pengenalan suara digunakan untuk mencari konten yang relevan dengan permintaan pengguna.
- 3) Sistem Temu Kembali Informasi Web  
Jenis sistem ini berfokus pada pencarian dan temu kembali informasi di web. Mesin pencari web seperti Google menggunakan algoritma kompleks untuk mencari dan memeringkat halaman web yang paling relevan dengan *query* pengguna.
- 4) Sistem Temu Kembali Informasi Domain Khusus  
Jenis sistem ini dirancang untuk mencari dan temu kembali informasi dalam domain atau bidang tertentu, seperti bidang kesehatan, hukum, atau keuangan. Pemahaman domain dan penggunaan istilah khusus domain menjadi penting dalam jenis sistem ini.

### 3.1.4 Manfaat Sistem Temu Kembali Informasi

Sistem temu kembali informasi sangat bermanfaat bagi pencarian informasi, manfaat tersebut yaitu:

- 1) Sistem temu kembali informasi dapat membantu pengguna untuk mengakses informasi yang relevan dengan cepat dan efisien.
- 2) Dengan menggunakan mekanisme pemeringkatan, sistem dapat memberikan hasil temu kembali yang paling relevan dan akurat.
- 3) Sistem temu kembali informasi dapat digunakan untuk mengelola koleksi dokumen yang besar dan kompleks.
- 4) Sistem ini dapat membantu pengguna menemukan informasi yang mungkin tidak diketahui sebelumnya atau tidak mudah diakses.

### 3.2 Term Frequency Inverse Document Frequency (TF-IDF)

#### 3.2.1 Pengertian Metode TF-IDF

*Term Frequency-Inverse Document Frequency* (TF-IDF) adalah metode statistik yang digunakan dalam pemrosesan bahasa alami dan sistem temu kembali informasi untuk mengukur pentingnya sebuah term (kata kunci) dalam sebuah dokumen dalam konteks koleksi dokumen yang lebih besar.

TF-IDF menghasilkan bobot atau skor yang mewakili tingkat pentingnya sebuah term dalam sebuah dokumen dalam konteks koleksi dokumen yang lebih besar. Bobot ini dapat digunakan untuk memperhitungkan dan membandingkan relevansi dokumen dalam sistem temu kembali informasi. Dokumen yang memiliki skor TF-IDF yang lebih tinggi untuk suatu term cenderung dianggap lebih relevan terhadap query atau permintaan pengguna yang mengandung term tersebut.

TF-IDF adalah salah satu metode yang umum digunakan dalam pemrosesan bahasa alami dan sistem temu kembali informasi untuk meningkatkan akurasi dan relevansi hasil pencarian.

#### 3.2.2 Penggunaan Skor TF-IDF

Skor TF-IDF digunakan dalam sistem temu kembali informasi untuk

memperhitungkan dan membandingkan relevansi dokumen terhadap query atau permintaan pengguna. Berikut adalah beberapa penggunaan skor TF-IDF dalam temu kembali informasi:

- 1) Skor TF-IDF dapat digunakan untuk memeringkat dokumen-dokumen yang paling relevan dengan query pengguna. Dokumen yang memiliki skor TF-IDF yang lebih tinggi untuk term-term dalam query cenderung dianggap lebih relevan dan diberikan peringkat yang lebih tinggi dalam hasil pencarian.
- 2) Skor TF-IDF dapat digunakan untuk menentukan dokumen-dokumen yang paling relevan dengan query pengguna. Dokumen-dokumen yang memiliki skor TF-IDF yang tinggi untuk term-term dalam query dianggap memiliki informasi yang lebih relevan dan kemungkinan besar mengandung jawaban atau informasi yang dicari oleh pengguna.
- 3) Skor TF-IDF dapat digunakan sebagai kriteria pemfilteran untuk menghilangkan dokumen yang tidak relevan dengan query pengguna. Dokumen-dokumen yang memiliki skor TF-IDF yang rendah untuk term-term dalam query dapat dianggap kurang relevan dan dikecualikan dari hasil pencarian.
- 4) Skor TF-IDF juga dapat digunakan dalam proses ekstraksi ringkasan dokumen. Term-term dengan skor TF-IDF yang tinggi dalam sebuah dokumen menunjukkan bahwa term tersebut memiliki bobot yang penting dan mencerminkan inti atau pokok dari dokumen tersebut.

### 3.3 Pencarian Dokumen Teks Menggunakan Metode TF-IDF

#### 3.3.1 Peran TF-IDF dalam Dokumen Teks

Metode *Term Frequency-Inverse Document Frequency* (TF-IDF) memberikan kontribusi yang signifikan pencarian dokumen teks. Metode ini membantu menghasilkan representasi yang efektif

untuk dokumen-dokumen dalam sistem temu kembali informasi. Dengan menghitung bobot term berdasarkan frekuensi kemunculannya dalam dokumen dan inversi frekuensi term tersebut dalam koleksi dokumen, TF-IDF menghasilkan representasi numerik yang menggambarkan pentingnya term dalam dokumen.

Selain itu, metode TF-IDF juga membantu dalam pemilihan term yang relevan dalam proses temu kembali informasi. Dokumen-dokumen yang memiliki skor TF-IDF yang tinggi untuk term-term dalam *query* atau permintaan pengguna cenderung dianggap lebih relevan. Dengan demikian, TF-IDF membantu dalam pemeringkatan dan penyajian hasil pencarian yang lebih relevan bagi pengguna.

Selain itu, metode TF-IDF juga memberikan keuntungan dalam skalabilitas. Metode ini dapat diterapkan pada koleksi dokumen yang besar dan heterogen, dengan menggunakan teknik pengindeksan dan perhitungan yang efisien. Ini memungkinkan sistem temu kembali informasi untuk mengelola dan mencari informasi dalam jumlah data yang besar dengan kinerja yang baik

Selain itu, TF-IDF juga memiliki fleksibilitas yang memungkinkan penggunaannya dalam berbagai tugas pemrosesan bahasa alami. Metode ini dapat digunakan dalam ekstraksi fitur, klasifikasi teks, analisis sentimen, dan tugas-tugas lainnya untuk menggambarkan pentingnya term dalam mewakili dan memprediksi atribut atau label tertentu.

### 3.3.2 Pembuatan Indeks dalam Dokumen Teks

Pembuatan indeks *Term Frequency-Inverse Document Frequency* (TF-IDF) dalam dokumen teks melibatkan beberapa langkah penting. Berikut adalah tahapan-tahapan utama dalam pembuatan indeks TF-IDF:

#### 1) Pra-Pemrosesan Dokumen

Langkah ini melibatkan penghapusan karakter khusus, tokenisasi (memisahkan teks menjadi unit-unit

yang lebih kecil seperti kata-kata), mengubah teks menjadi huruf kecil, dan pembersihan data lainnya.

#### 2) Perhitungan *Term Frequency* (TF)

Setelah pra-pemrosesan, langkah selanjutnya adalah menghitung frekuensi kemunculan setiap term dalam setiap dokumen. Frekuensi kemunculan term dalam dokumen dihitung untuk memperoleh skor *Term Frequency* (TF). Skor TF dapat dihitung dengan menghitung jumlah kemunculan term dalam dokumen atau dengan menggunakan metode normalisasi seperti TF Logarithmic atau TF Augmented.

#### 3) Perhitungan Inverse Document Frequency (IDF)

Setelah mendapatkan skor TF untuk setiap term dalam dokumen, langkah selanjutnya adalah menghitung *Inverse Document Frequency* (IDF). IDF menggambarkan pentingnya suatu term dalam koleksi dokumen secara keseluruhan. IDF dihitung dengan membagi total jumlah dokumen dalam koleksi dengan jumlah dokumen yang mengandung term tersebut. Kemudian, hasilnya di-logaritman atau dinormalisasi untuk mendapatkan skor IDF yang akhir.

#### 4) Perhitungan TF-IDF

Langkah berikutnya adalah mengalikan skor TF dengan skor IDF untuk setiap term dalam setiap dokumen. Hasil perkalian ini menghasilkan skor TF-IDF untuk setiap term dalam setiap dokumen. Skor TF-IDF menggambarkan tingkat pentingnya term tersebut dalam dokumen dan koleksi secara keseluruhan.

#### 5) Pembuatan Indeks

Setelah menghitung skor TF-IDF untuk setiap term dalam dokumen, langkah terakhir adalah membangun indeks TF-IDF. Indeks ini mencatat informasi term-term yang ada dalam dokumen, termasuk skor TF-IDF untuk setiap term dalam setiap dokumen. Indeks ini memungkinkan akses cepat dan efisien

ke dokumen-dokumen yang relevan berdasarkan term-term yang dicari dalam proses temu kembali informasi.

### 3.3.3 Keunggulan metode TF-IDF dalam Pencarian Dokumen Teks

Metode *Term Frequency-Inverse Document Frequency* (TF-IDF) memiliki keunggulan dalam pencarian dokumen teks. Salah satu keunggulannya adalah kemampuannya untuk memberikan bobot yang tepat pada term-term dalam dokumen. Dengan menggabungkan skor *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF), metode TF-IDF mampu mengidentifikasi term-term yang paling relevan dalam dokumen. Hal ini memungkinkan sistem untuk memilih dan memperingkat dokumen berdasarkan tingkat relevansinya dengan query pengguna.

Keunggulan lain dari metode TF-IDF adalah kemampuannya untuk menangani term-term umum. Dengan memberikan bobot yang rendah pada term-term umum melalui perhitungan IDF, metode ini membantu mengurangi pengaruh term-term yang tidak informatif dalam proses pemeringkatan dokumen. Sebagai contoh, kata-kata seperti "dan", "atau", atau "yang" yang sering muncul dalam dokumen teks umumnya memiliki bobot yang rendah dalam metode TF-IDF.

Selain itu, TF-IDF juga mampu mengatasi masalah term-term yang jarang muncul dalam koleksi dokumen. Dalam metode ini, term-term yang jarang muncul dalam dokumen akan mendapatkan bobot yang lebih tinggi, karena term-term tersebut dianggap memiliki informasi yang lebih spesifik. Dengan demikian, metode TF-IDF memungkinkan identifikasi dokumen-dokumen yang mungkin kurang umum tetapi relevan dengan query pengguna.

Terakhir, metode TF-IDF memiliki keunggulan dalam hal implementasi yang sederhana dan mudah dipahami. Perhitungan skor TF dan IDF relatif sederhana, dan konsep dasar dari metode ini dapat dipahami dengan mudah. Selain itu, metode TF-IDF dapat disesuaikan dan dikustomisasi sesuai

dengan kebutuhan aplikasi tertentu, seperti dengan memodifikasi rumus perhitungan atau memperkenalkan teknik penghalusan tambahan.

### 3.3.4 Keterbatasan metode TF-IDF dalam Pencarian Dokumen Teks

Metode *Term Frequency-Inverse Document Frequency* (TF-IDF), juga memiliki beberapa keterbatasan dalam konteks sistem temu kembali informasi dokumen teks. Metode ini tidak memperhatikan konteks makna atau struktur bahasa dalam pemilihan dan pemeringkatan dokumen. Hal ini dapat menyebabkan ketidakakuratan dalam menemukan dokumen yang sesuai dengan kebutuhan pengguna jika terdapat term-term dengan makna ganda atau jika konteks perlu dipertimbangkan.

Metode TF-IDF memiliki kelemahan dalam mengatasi dokumen-dokumen dengan panjang yang berbeda. Metode ini hanya menghitung frekuensi term dalam dokumen tanpa mempertimbangkan panjangnya. Sebagai hasilnya, dokumen yang lebih panjang cenderung memiliki bobot yang lebih tinggi hanya karena jumlah term yang lebih banyak, meskipun term-term tersebut mungkin tidak memberikan informasi yang relevan secara signifikan.

Selain itu, metode TF-IDF tidak memperhitungkan relasi antar term dalam dokumen. Metode ini memperlakukan setiap term secara independen dan mengabaikan hubungan antar term. Padahal, informasi penting sering kali terkait dengan relasi semantik atau hubungan yang terdefinisi antar term dalam dokumen.

Terakhir, metode TF-IDF tidak secara eksplisit menangani sinonim atau polisemi, yaitu istilah dengan makna yang sama atau istilah dengan makna ganda. Hal ini dapat menyebabkan ketidakakuratan dalam hasil temu balik jika terdapat variasi istilah yang digunakan dalam dokumen, sehingga menghilangkan beberapa informasi yang relevan.

#### 4. PENUTUP

##### Kesimpulan

- 1) Sistem Temu Kembali Informasi adalah sistem komputer yang dirancang untuk mencari, mengambil, dan mengorganisir informasi yang relevan dari koleksi dokumen atau sumber informasi lainnya berdasarkan permintaan pengguna. Tujuannya adalah memberikan respons yang akurat dan efektif terhadap kebutuhan informasi pengguna.
- 2) Komponen utama dalam sistem temu kembali informasi meliputi masukan (*query*), pemroses, dan keluaran. Masukan berupa *query* atau kata kunci pencarian yang diberikan oleh pengguna, pemroses bertanggung jawab untuk mengolah *query* tersebut, dan keluaran mencakup hasil atau respons dari sistem.
- 3) Jenis-jenis sistem temu kembali informasi meliputi sistem temu kembali informasi teks, web, dan domain khusus. Setiap jenis sistem memiliki fokus dan tujuan tertentu, tergantung pada jenis informasi yang dicari oleh pengguna.
- 4) Manfaat dari sistem temu kembali informasi meliputi efisiensi dalam mengakses informasi, akurasi dalam memberikan hasil yang relevan, skalabilitas dalam mengelola koleksi dokumen yang besar, dan kemampuan untuk menemukan informasi baru yang mungkin tidak diketahui sebelumnya.
- 5) Metode *Term Frequency-Inverse Document Frequency* (TF-IDF) adalah metode statistik yang digunakan dalam sistem temu kembali informasi untuk mengukur pentingnya sebuah term dalam sebuah dokumen dalam konteks koleksi dokumen yang lebih besar. Metode ini menggabungkan skor *term frequency* (TF) dengan skor *inverse document frequency* (IDF).
- 6) Skor TF-IDF digunakan dalam pemeringkatan dokumen, pencarian informasi relevan, pemfilteran dokumen, dan ekstraksi ringkasan. Metode ini membantu memperhitungkan dan membandingkan relevansi dokumen dengan *query* atau permintaan pengguna.

- 7) Kelebihan metode TF-IDF meliputi representasi term yang sederhana, perhitungan signifikansi term, pengurangan bobot term yang umum, dan skalabilitas. Namun, metode ini juga memiliki kelemahan, seperti tidak memperhatikan urutan kata, sensitivitas terhadap term yang jarang muncul, tidak memodelkan hubungan semantik, dan rentan terhadap manipulasi.

##### Saran

Berdasarkan pada hasil yang sudah penulis lakukan, maka yang penulis sarankan adalah dilakukannya pengembangan sistem temu kembali informasi dengan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF), seperti memperhitungkan urutan kata dan memodelkan hubungan semantik antara term-term dalam dokumen. Selain itu, penelitian dapat fokus pada integrasi metode TF-IDF dengan teknologi terkini seperti kecerdasan buatan (AI) dan analisis sentimen untuk meningkatkan kinerja sistem temu kembali informasi dalam menangani data yang besar (*big data*) dan pencarian informasi dalam bahasa alami.

#### 5. REFERENSI

- [1] Fitri, M, *Perancangan Information Retrieval System* dengan Metode Pembobotan Kombinasi TF-IDF untuk Pencarian Dokumen Berbahasa Indonesia. Jurnal Tanjungpura, 2013.
- [2] Wahyuni, Rizki T., Prastiyanto, Dhidik., Suprpto, *et al*., Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi. Jurnal Teknik Elektro Vol. 9 No. 1, 2017.
- [3] Husni, Sistem Temu Kembali Informasi. Universitas Trunojoyo, 2010.
- [4] Ma'arif, Abdul Aziz, Penerapan Algoritma Tf-Idf Untuk Pencarian Karya Ilmiah. Jurnal. Semarang: Jurusan Teknik Informatika, Universitas Dian Nuswantoro, 2015.
- [5] Okfalisa and A. H. Harahap, "Implementasi Metode *Terms Frequency-Inverse Document Frequency* (TF-IDF) dan *Maximum Marginal Relevance* untuk Monitoring

- Diskusi Online," Jurnal Sains, Teknologi dan Industri, vol. 13, no. 2, pp. 151-159, 2016.
- [6] I. Mulyana, S. Ramadona, and Herfina, "Penerapan *Terms Frequency-Inverse Document Frequency* Pada Sistem Ringkasan NTeks Otomatis Dokumen Tunggal Berbahasa Indonesia," Jurnal Konferensi Nasional Teknologi Informasi dan Komunikasi, 2012.
- [7] N. P. A. Yuliani, P. Aliandu and P. N. Nani, "SISTEM TEMU BALIK DOKUMEN TEKS MENGGUNAKAN METODE BOOLEAN DAN TERM WIGHT TF.IDF," Seminar Nasional dan Konferensi Sistem Informasi, Informatika dan Komunikasi, pp. 499-504, 2021.
- [8] D. S. Harjanto, S. N. Endah, and N. Bahtiar, "Sistem Temu Kembali Informasi pada Dokumen Teks Menggunakan Metode Term Frequency Inverse Document Frequency (TF-IDF)," JURNAL SAINS DAN MATEMATIKA, vol. 20, no. 3, pp. 64-70, 2015.
- [9] Irmawati, "SISTEM TEMU KEMBALI INFORMASI PADA DOKUMEN DENGAN," Jurnal Ilmiah FIFO, vol. 9, no. 1, 2017