

OPTIMASI RANGKING DOKUMEN DENGAN MODIFIKASI TF-IDF BERBASIS WAKTU PUBLIKASI DAN COSINE SIMILARITY

Reni Aurelia¹, Hafiz Irsyad², Abdul Rahman³

^{1,2}Program Studi Informatika, Universitas Multi Data Palembang, Jl Rajawali No.14, Indonesia

³Program Studi Teknik Elektro, Universitas Multi Data Palembang, Jl Rajawali No.14, Indonesia

¹Reniaurelia_2226250028@mhs.mdp.ac.id, ²hafizirsyad@mdp.ac.id, ³arahman@mdp.ac.id

***Nyimas Nisrinaa Kamilah**

Program Studi Informatika, Universitas Multi Data Palembang, Jl Rajawali No.14, Indonesia

Nyimasnisrinaakamilah_2226250064@mhs.mdp.ac.id

Abstrak

Information Retrieval (IR) tradisional belum mempertimbangkan waktu publikasi dalam menentukan relevansi dokumen. Penelitian ini bertujuan untuk meningkatkan relevansi hasil pencarian dengan memodifikasi metode TF-IDF berbasis waktu publikasi. Metode ini menggabungkan bobot TF-IDF dengan *Cosine Similarity* untuk mengukur kesamaan antara kueri dan dokumen. Dalam penelitian ini, dataset dievaluasi menggunakan metode yang diusulkan, dengan pengukuran melalui metrik *precision*, *recall*, dan *F1-score*. Hasil pengujian menunjukkan bahwa pendekatan ini mencapai *precision* 0.87, *recall* 1.00, dan *F1-Score* 0.93. Berdasarkan hasil evaluasi, penambahan aspek temporal pada metode ini terbukti mampu meningkatkan akurasi IR dalam konteks pencarian informasi terkini

Kata kunci : *Cosine Similarity, Information Retrieval, Relevansi, TF-IDF, Waktu Publikasi.*

Abstract

Traditional Information Retrieval (IR) has not considered the time of publication in determining the relevance of documents.. This study aims to increase the relevance of search results by modifying the publication time-based TF-IDF method. This method combines the TF-IDF weight with Cosine Similarity to measure the similarity between the query and the document. In this study, the dataset was evaluated using the proposed method, with measurements through precision metrix, recall, and F1-score. The test results show that this approach achieves a precision of 0.87, a recall of 1.00, and an F1-Score of 0.93. Based on the results of the evaluation, the addition of temporal aspects to this method has been proven to be able to increase IR accuracy in the context of searching for the latest information

Keywords: *Cosine Similarity, Information Retrieval, Publication Time, Relevance, TF-IDF.*

1. PENDAHULUAN

Dalam era digital yang berkembang pesat saat ini, volume data teks yang tersedia secara daring meningkat secara eksponensial [1]. Data teks adalah data satuan lingual yang dimediakan secara tulis atau lisan dengan tata organisasi tertentu untuk mengungkapkan makna secara kontekstual [2]. Data teks ini

mencakup berbagai domain, termasuk berita, artikel ilmiah, dan ulasan produk. Fenomena ini membawa tantangan signifikan dalam menemukan informasi yang relevan dan tepat waktu. *Information Retrieval* (IR) berperan penting dalam membantu pengguna menavigasi dan mengekstraksi informasi dari kumpulan data yang besar dan kompleks [1]. Namun, efektivitas IR sangat bergantung pada

kemampuan untuk mengukur kesamaan antara kueri pengguna dan dokumen dalam koleksi data, yang pada akhirnya menentukan kualitas hasil pencarian [3].

Pada umumnya, IR tradisional menggunakan pendekatan berbasis konten, seperti *Vector Space Model*, yang mengandalkan frekuensi kata untuk menentukan relevansi dokumen [4]. Pendekatan ini, meskipun cukup efektif dalam banyak kasus, sering kali mengabaikan aspek temporal, yaitu waktu publikasi dokumen [5]. Dalam konteks informasi yang sensitif terhadap waktu, seperti berita terkini, tren sosial, atau riset ilmiah yang baru diterbitkan, ketidakhadiran faktor waktu dalam model IR dapat menyebabkan hasil pencarian yang kurang relevan dan tidak memadai. Hal ini menjadi isu penting dalam mengembangkan IR yang lebih efektif, terutama untuk aplikasi yang membutuhkan informasi terbaru dan relevan.

Fokus permasalahan terletak pada bagaimana mengoptimalkan IR untuk menghasilkan hasil pencarian yang lebih relevan dan kontekstual dengan mempertimbangkan faktor temporal. Tanpa mempertimbangkan waktu, dokumen yang diterbitkan beberapa tahun lalu bisa saja mendapatkan peringkat yang lebih tinggi daripada dokumen yang lebih baru, meskipun isinya lebih relevan dengan kueri pengguna. Oleh karena itu, isu yang perlu diselesaikan adalah bagaimana mengintegrasikan waktu publikasi dokumen dalam proses pemeringkatan dokumen agar IR dapat memberikan hasil pencarian yang lebih relevan, terutama untuk domain yang bergantung pada informasi terkini [6].

Beberapa penelitian telah mengusulkan berbagai metode untuk meningkatkan relevansi hasil pencarian dengan mempertimbangkan faktor temporal. Penelitian[7] mengembangkan sistem rekomendasi berita dengan menggunakan metode TF-IDF dan *Cosine Similarity*, yang terbukti meningkatkan akurasi dalam menyajikan berita yang relevan kepada

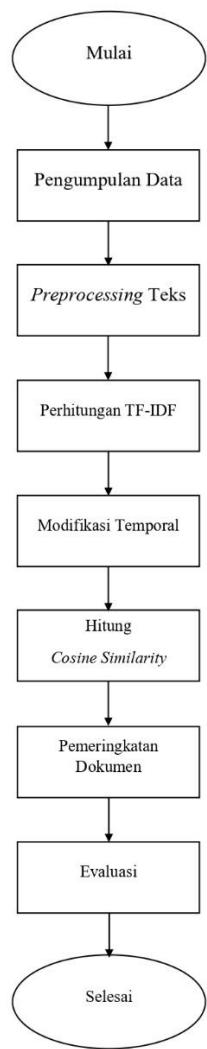
pengguna. Tetapi, penelitian ini masih belum sepenuhnya mengintegrasikan waktu publikasi dalam model pemeringkatan dokumen secara langsung, yang membuatnya kurang efektif dalam konteks informasi yang sensitif terhadap waktu.

Penelitian lainnya, [8] menunjukkan bahwa kombinasi metode TF-IDF dan *Cosine Similarity* dapat memberikan hasil yang baik dalam mengukur kesamaan antara dokumen dan kueri. Walaupun demikian, pendekatan ini juga tidak menangkap faktor temporal, yang sangat penting untuk aplikasi seperti pencarian berita atau dokumen terbaru. Oleh karena itu, meskipun metode ini efektif untuk pengukuran kesamaan konten, ia tidak dapat sepenuhnya diandalkan untuk IR yang memprioritaskan relevansi berdasarkan waktu.

Untuk mengatasi keterbatasan-keterbatasan tersebut, penelitian ini mengusulkan modifikasi pada metode TF-IDF dengan menambahkan faktor waktu publikasi dokumen dalam perhitungan bobot kata. Dengan memasukkan elemen waktu kedalam pembobotan, bobot kata tidak hanya mencerminkan frekuensi dan distribusinya dalam koleksi dokumen, tetapi juga mempertimbangkan relevansi temporal dari setiap dokumen. Pendekatan ini kemudian dipadukan dengan metode *Cosine Similarity* untuk mengukur kesamaan antara kueri dan dokumen berdasarkan bobot yang telah dimodifikasi. Penelitian ini akan meningkatkan relevansi hasil pencarian dalam IR, terutama dalam konteks informasi yang sensitif terhadap waktu, seperti berita atau artikel ilmiah yang baru diterbitkan.

2. METODE

Dalam Penelitian Optimasi Rangking Dokumen dengan Modifikasi TF-IDF Berbasis Waktu Publikasi dan *Cosine Similarity* dilakukan beberapa tahap dalam proses pembuatannya. Alur penelitian [9] dapat dilihat pada Gambar 1.



Gambar 1 Alur Penelitian

2.1. Analisis Permasalahan

Pada penelitian ini, fokus utama adalah pada peningkatan relevansi hasil pencarian dalam *Information Retrieval* (IR), dengan memodifikasi metode TF-IDF untuk memasukkan faktor waktu publikasi dokumen. Permasalahan yang dihadapi adalah model IR konvensional, seperti TF-IDF, hanya memperhitungkan frekuensi kata dalam dokumen dan koleksi, tanpa memperhatikan aspek temporal atau waktu publikasi. Hal ini berpotensi mengurangi relevansi hasil pencarian, terutama dalam konteks pencarian informasi yang sensitif terhadap waktu, seperti berita atau data penelitian terbaru.

Oleh karena itu, untuk menyelesaikan masalah ini, penelitian mengusulkan

pendekatan baru dengan memodifikasi pembobotan TF-IDF untuk memasukkan elemen waktu publikasi dokumen. Pendekatan ini bertujuan untuk meningkatkan relevansi hasil pencarian dengan memperhitungkan seberapa terkini atau relevan sebuah dokumen berdasarkan waktu penerbitannya.

2.2 Arsitektur

Arsitektur yang dikembangkan dalam penelitian ini terdiri dari beberapa komponen utama, yang masing-masing memiliki fungsi spesifik dalam memproses dan mengoptimalkan peringkat dokumen dalam IR. Berikut adalah komponen utama dari arsitektur :

2.2.1 Preprocessing Teks

Proses *preprocessing* dilakukan untuk memberikan data teks yang digunakan, proses ini mencakup beberapa tahap [10] yaitu:

1. *Case Folding*: Menstandarkan semua huruf menjadi huruf kecil.
2. *Tokenisasi* : Memecah teks menjadi unit-unit kata atau token.
3. Penghapusan *stopwords* : Menghapus kata-kata yang tidak memiliki makna penting (misalnya: "dan", "atau", "itu")
4. *Stemming* : Mengubah kata menjadi bentuk dasarnya untuk mengurangi variabilitas kata yang serupa (misalnya, "berlari" menjadi "lari").

Preprocessing ini bertujuan untuk mengurangi dimensi data dan meningkatkan akurasi model IR dengan mengurangi *noise* dalam teks.

2.2.2 Pembobotan TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) adalah salah satu metode paling umum digunakan dalam IR untuk mengukur pentingnya suatu kata dalam dokumen dan koleksi dokumen secara keseluruhan [11]. Pembobotan ini dihitung berdasarkan dua faktor [12]:

1. Term Frequency (TF) digunakan untuk mengukur frekuensi kemunculan suatu kata dalam dokumen yang ada.
2. Inverse Document Frequency (IDF) digunakan untuk mengukur seberapa jarang kata tersebut muncul dalam koleksi dokumen yang ada.

Formula TF-IDF untuk setiap kata ω dalam dokumen D adalah

$$TF - IDF(\omega, D) = TF(\omega, D) \times IDF(\omega) \quad (1)$$

Dimana :

$$IDF(\omega) = \log \left(\frac{N}{df(\omega)} \right) \quad \dots \dots \dots \dots \dots \dots \dots \quad (2)$$

N adalah jumlah total dokumen, dan $df(\omega)$ adalah jumlah dokumen yang mengandung kata ω

2.2.3 Modifikasi Pembobotan Temporal

Modifikasi utama dalam penelitian ini adalah memasukkan faktor waktu publikasi dokumen dalam pembobotan TF-IDF [13]. Konsep dasarnya adalah bahwa dokumen yang lebih baru akan memiliki bobot lebih tinggi dalam hal relevansi, terutama jika dokumen tersebut mengandung kata-kata yang relevan dengan kueri pencarian.

Bobot TF-IDF yang dihitung untuk setiap kata dalam dokumen akan dimodifikasi berdasarkan selisih waktu antara publikasi dokumen dan waktu kueri pencarian. Semakin baru dokumen tersebut, semakin tinggi bobotnya.

Formula modifikasi pembobotan dapat dijelaskan sebagai berikut :

$$TF - IDF_{temporal}(\omega, D) = TF - IDF(\omega, D) \times \left(1 + \frac{T_{now} - T_{pub}}{\Delta T} \right) \quad (3)$$

Dimana :

1. T_{now} adalah waktu kueri,
2. T_{pub} adalah waktu publikasi dokumen,
3. T adalah rentang waktu tertentu (misalnya dalam hari) untuk normalisasi.

2.2.4 Perhitungan Cosine Similarity

Setelah pembobotan dokumen sudah selesai, langkah berikutnya yang akan dilakukan adalah menghitung *cosine similarity* antara kueri dan dokumen yang ada. *Cosine Similarity* adalah metode yang mengukur seberapa mirip dua vektor berdasarkan sudut antara vektor tersebut [14]. Formula untuk *Cosine Similarity* antara dua vektor A dan B adalah :

$$\text{Cosine Similarity}(A, B) = \frac{A \cdot B}{\|A\| \|B\|} \quad \dots \dots \dots \dots \quad (4)$$

Dimana :

1. A dan B adalah vektor representasi dokumen dan kueri,
2. $A \cdot B$ adalah hasil perkalian titik antara dua vektor,
3. $\|A\|$ dan $\|B\|$ adalah panjang (norm) masing – masing vector.

2.2.5 Pemeringkatan dan Evaluasi

Dokumen yang dihitung tingkat kesamaannya dengan kueri akan di peringkat berdasarkan nilai *Cosine Similarity*. Dokumen dengan skor *Cosine Similarity* tertinggi akan dianggap paling relevan [15].

Untuk evaluasi, dilakukan perhitungan menggunakan metrix seperti *precision*, *recall*, dan *F1-score* untuk mengukur seberapa baik metode dalam mengambil dokumen relevan dan menghindari dokumen yang tidak relevan [16].

2.3 Implementasi Metode

Metode dikembangkan untuk mengukur kesamaan antara dokumen dan kueri dengan mempertimbangkan konten dan aspek temporal. Proses implementasi mencakup pengumpulan data, *preprocessing* teks, pembobotan menggunakan TF-IDF yang dimodifikasi secara temporal, perhitungan kesamaan dengan *Cosine Similarity*, serta evaluasi menggunakan *precision*, *recall*, dan *F1-score*.

2.3.1 Alat dan Teknologi

Penelitian ini diimplementasikan menggunakan bahasa pemrograman Python, dengan langkah-langkah implementasi sebagai berikut:

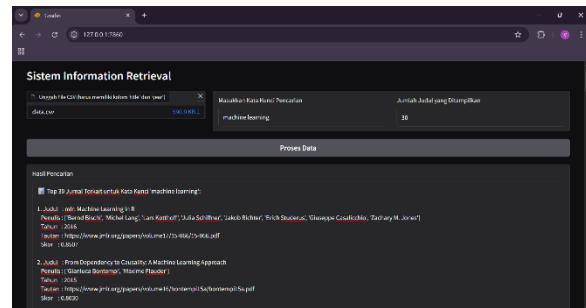
1. Pengumpulan Data, data jurnal yang dipilih mencakup berbagai topik dan rentang waktu publikasi.
2. *Preprocessing*, data teks dibersihkan melalui proses tokenisasi, *stopword removal*, dan *stemming*.
3. Pembobotan, TF-IDF dihitung untuk setiap dokumen dan modifikasi temporal diterapkan pada bobot kata.
4. Peringkat Dokumen, *Cosine Similarity* dihitung antara kueri dan dokumen dan dokumen diperingkat berdasarkan skor tersebut.
5. Evaluasi, hasil peringkatan di evaluasi menggunakan *matrix precision*, *recall*, dan *F1-score* untuk mengukur performa metode .

3. HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan *Information retrieval* (IR) yang memodifikasi metode TF-IDF dengan mempertimbangkan faktor waktu publikasi dokumen untuk meningkatkan relevansi hasil pencarian. Data yang digunakan berasal dari sebuah dataset jurnal dengan berbagai topik dan rentang waktu publikasi. Dari dataset tersebut, dipilih 30 dokumen yang paling relevan berdasarkan skor kemiripan untuk dilakukan pengujian lebih lanjut terhadap performa metode . Metode ini dievaluasi menggunakan *Confusion Matrix*.

Hasil implementasi TF-IDF berhasil menampilkan daftar jurnal yang relevan berdasarkan kata kunci yang dimasukkan. Sebagai contoh, pencarian dengan kata kunci "machine learning" menghasilkan jurnal berjudul "mlr: Machine Learning in R" sebagai hasil teratas dengan skor relevansi 0.8507. Informasi yang ditampilkan meliputi judul, penulis, tahun,

tautan, dan skor kesamaan, yang menunjukkan bahwa sistem mampu menyajikan informasi ilmiah secara akurat dan terurut berdasarkan tingkat relevansi. Hasil dari aplikasi ditampilkan pada Gambar 1.



Gambar 1. Hasil *Output* Aplikasi

Hasil confusion matrix dari proses evaluasi algoritma dalam sistem optimasi rangking dokumen dengan modifikasi TF-IDF berbasis waktu publikasi dan cosine similarity ditampilkan pada Tabel 1. Tabel tersebut menunjukkan performa model dalam mengklasifikasikan relevansi dokumen terhadap kata kunci pencarian. Berdasarkan hasil, terdapat 26 dokumen yang diklasifikasikan dengan benar sebagai relevan (True Positive), tidak ada dokumen relevan yang salah diklasifikasikan sebagai tidak relevan (False Negative), 4 dokumen yang tidak relevan namun diklasifikasikan sebagai relevan (False Positive), dan tidak ada dokumen yang diklasifikasikan sebagai tidak relevan dengan benar (True Negative). Hasil ini menunjukkan bahwa sistem cukup efektif dalam mengenali dokumen relevan, namun masih menghasilkan kesalahan dalam mengklasifikasikan dokumen yang tidak relevan.

Tabel 1. Hasil *Confusion Matrix*

	Positive	Negative
Positive	26 (TP)	0 (FN)
Negative	4 (FP)	0 (TN)

Hasil evaluasi kinerja metode berdasarkan confusion matrix disajikan pada Tabel 2.

Tabel ini menunjukkan metrik evaluasi untuk masing-masing kelas. Pada label 1 (dokumen relevan), diperoleh nilai precision sebesar 0.87 yang berarti 87% dari dokumen yang diprediksi relevan memang benar relevan. Nilai recall sebesar 1.00 menunjukkan bahwa semua dokumen relevan berhasil dikenali oleh sistem. F1-score sebesar 0.93 merupakan rata-rata harmonis dari precision dan recall, yang menunjukkan keseimbangan kinerja model dalam mengenali dokumen relevan.

Sebaliknya, pada label 0 (dokumen tidak relevan), nilai precision, recall, dan F1-score semuanya 0.00, yang mengindikasikan bahwa model tidak mampu mengenali dokumen yang tidak relevan dengan benar. Hal ini berkaitan dengan jumlah data tidak relevan yang sangat sedikit (support = 4), sehingga model lebih condong untuk memprediksi semua dokumen sebagai relevan.

Tabel 2. Hasil Evaluasi Kinerja Metode Berdasarkan *Confusion Matrix*

Label	Precision	Recall	F1-Score	Support
0	0.00	0.00	0.00	4
1	0.87	1.00	0.93	26

Ringkasan metrik evaluasi keseluruhan ditampilkan pada Tabel 3. Nilai akurasi sebesar 0.87 menunjukkan bahwa secara keseluruhan, 87% prediksi sistem adalah benar dari total 30 data uji. Macro average menghasilkan nilai precision 0.43, recall 0.50, dan F1-score 0.46, yang merupakan rata-rata sederhana dari kedua label tanpa mempertimbangkan jumlah data masing-masing label. Sedangkan weighted average menghasilkan precision 0.75, recall 0.87, dan F1-score 0.80, yang mempertimbangkan jumlah data pada tiap label, sehingga lebih mencerminkan performa keseluruhan model pada distribusi data yang tidak seimbang.

Accuracy menunjukkan tingkat keseluruhan prediksi yang benar oleh sistem, yaitu sebesar 87% dari total 30 data uji. Macro Average menghitung rata-rata precision, recall, dan F1-score dari semua label tanpa mempertimbangkan jumlah data per label. Nilai yang diperoleh relatif rendah (precision 0.43, recall 0.50, F1-score 0.46), karena kinerja untuk label minoritas (kelas 0) sangat rendah. Weighted Average menghitung rata-rata yang mempertimbangkan jumlah data pada tiap label (support), sehingga lebih mencerminkan kinerja sistem pada data tidak seimbang. Hasil yang diperoleh cukup baik, dengan precision 0.75, recall 0.87, dan F1-score 0.80.

Tabel ini menunjukkan bahwa meskipun sistem berkinerja tinggi pada kelas mayoritas, masih perlu perbaikan dalam menangani kelas minoritas agar hasil lebih seimbang. Ringkasan Metrik Evaluasi keseluruhan kinerja dapat dilihat pada Tabel 3.

Tabel 3. Ringkasan Metrik Evaluasi Keseluruhan Kinerja Metode

Metrik	Precision	Recall	F1-Score	Support
Accuracy	-	-	0.87	30
Macro Avg	0.43	0.50	0.46	30
Weighted Avg	0.75	0.87	0.80	30

Berdasarkan hasil pengujian evaluasi kinerja sistem menggunakan confusion matrix dan metrik evaluasi (Tabel 2 dan Tabel 3), dapat disimpulkan bahwa metode yang diusulkan berhasil mengidentifikasi dokumen yang relevan dengan tingkat akurasi yang tinggi. Nilai recall sebesar 1.00 pada label dokumen relevan menunjukkan bahwa seluruh dokumen relevan berhasil ditemukan oleh sistem. Selain itu, nilai precision sebesar 0.87 mengindikasikan bahwa sebagian besar dokumen yang diprediksi relevan memang benar-benar sesuai dengan kueri. Hasil ini

memperlihatkan bahwa sistem efektif dalam menampilkan dokumen yang relevan berdasarkan modifikasi TF-IDF berbasis waktu publikasi dan cosine similarity.

Namun, skor nol pada kelas 0 (irrelevant) mengindikasikan bahwa metode kurang mampu mengklasifikasikan dokumen yang tidak relevan secara cepat. Hal ini kemungkinan disebabkan oleh distribusi data yang tidak seimbang (26 relevan dan 4 tidak relevan). Maka dari itu, perlu dilakukan pengujian tambahan dengan dataset yang lebih seimbang dan beragam.

4. KESIMPULAN

Berdasarkan implementasi metode penelitian dan hasil pengujian, dapat disimpulkan bahwa modifikasi metode TF-IDF dengan mempertimbangkan waktu publikasi dokumen dapat meningkatkan relevansi hasil pencarian dalam *Information Retrieval*. Penggunaan bobot temporal dan *Cosine Similarity* mampu mengarahkan hasil pencarian ke dokumen yang lebih sesuai dan terkini.

Evaluasi metode menghasilkan nilai *precision* sebesar 0.87, *recall* 1.00, dan *F1-score* 0.93, yang menandakan performa metode sangat baik dalam menemukan dokumen relevan. Meskipun masih terdapat kelemahan dalam mengenali dokumen tidak relevan, pendekatan ini terbukti efektif untuk domain informasi yang sensitif terhadap waktu.

Dengan demikian, penelitian ini memberikan kontribusi berupa pendekatan sederhana namun bermanfaat dalam peningkatan IR berbasis teks dengan mempertimbangkan dimensi waktu.

DAFTAR PUSTAKA

- [1] Z. Zheng, K. Hui, B. He, X. Han, L. Sun, and A. Yates, “BERT-QE: Contextualized query expansion for document re-ranking,” *Findings of the Association for Computational Linguistics Findings of ACL: EMNLP 2020*, pp. 4718–4728, 2020, doi: 10.18653/v1/2020.findings-emnlp.424.
- [2] U. Mansyuri, “Kompresi Data Teks Dengan Metode Run Length Encoding,” *Jurnal Ilmiah Sistem Informasi*, vol. 1, no. 2, pp. 102–109, 2021, doi: 10.46306/sm.v1i2.13.
- [3] K. Juvekar and A. Purwar, “COS-Mix: Cosine Similarity and Distance Fusion for Improved Information Retrieval,” *Arxiv.Org*, pp. 1–10, 2024.
- [4] C. Lyu, S. M. Cavaletto, C. H. Keitel, and Z. Harman, “Interrogating the Temporal Coherence of EUV Frequency Combs with Highly Charged Ions,” *Physical Review Letters*, vol. 125, no. 9, 2020, doi: 10.1103/PhysRevLett.125.093201.
- [5] C. Xu and J. McAuley, “A Survey on Dynamic Neural Networks for Natural Language Processing,” *EACL 2023 - 17th Conference of the European Chapter of the Association for Computational Linguistics, Findings of EACL 2023*, pp. 2325–2336, 2023, doi: 10.18653/v1/2023.findings-eacl.180.
- [6] A. Abdallah, B. Piryani, J. Wallat, A. Anand, and A. Jatowt, *TempRetriever: Fusion-based Temporal Dense Passage Retrieval for Time-Sensitive Questions*, vol. 1, no. 1. Association for Computing Machinery, 2025.
- [7] G. Yunanda, D. Nurjanah, and S. Meliana, “Recommendation System from Microsoft News Data using TF-IDF and Cosine Similarity Methods,” *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 1, pp. 277–284, 2022, doi: 10.47065/bits.v4i1.1670.
- [8] A. Widianto, E. Pebriyanto, Fitriyanti, and Marna, “Document Similarity using Term Frequency-Inverse Document Frequency Representation and Cosine Similarity Adi,” vol. 4, no. 2, pp. 149–153, 2024.
- [9] T. A. R. Widyastuti *et al.*, *METODOLOGI PENELITIAN*:

- Panduan Lengkap Penulisan Karya Ilmiah. 2024.
- [10] W. O. W. Makmun, I. P. Ningrum, and A. M. Sajiah, "Penerapan Vector Space Model (Vsm) Pada Sistem Pencarian Artikel Arkeologi," *semanTIK*, vol. 8, no. 1, p. 69, 2022, doi: 10.55679/semantik.v8i1.15346.
- [11] R. Al Rasyid and D. H. U. Ningsih, "Penerapan Algoritma TF-IDF dan Cosine Similarity untuk Query Pencarian Pada Dataset Destinasi Wisata," *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 8, no. 1, pp. 170–178, 2024, doi: 10.35870/jtik.v8i1.1416.
- [12] Nico, U. Budiyanto, and T. Fatimah, "Implementasi Algoritma Pembobotan TF-IDF dan Cosine Similarity untuk Penetapan Kategori Artikel pada Website Universitas Budi Luhur," *Jurnal TICOM: Technology of Information and Communication*, vol. 10, no. 3, pp. 218–222, 2022.
- [13] D. Marwah and J. Beel, "Term-Recency for {TF}-{IDF}, {BM}25 and {USE} Term Weighting," *Proceedings of the 8th International Workshop on Mining Scientific Publications*, pp. 36–41, 2020.
- [14] F. A. Nugroho, F. Septian, D. A. Pungkastyo, and J. Riyanto, "Penerapan Algoritma Cosine Similarity untuk Deteksi Kesamaan Konten pada Sistem Informasi Penelitian dan Pengabdian Kepada Masyarakat," vol. 5, no. 4, pp. 529–536, 2021.
- [15] E. L. Amalia, A. J. Jumadi, I. A. Mashudi, and D. W. Wibowo, "ANALISIS METODE COSINE SIMILARITY PADA APLIKASI UJIAN ONLINE ESAI OTOMATIS (STUDI KASUS JTI POLINEMA) COSINE SIMILARITY METHOD ANALYSIS ON AUTOMATIC ESAI ONLINE TEST," vol. 8, no. 2, pp. 343–348, 2021, doi: 10.25126/jtiik.202184356.
- [16] R. Fauzan, A. V. Vitianingsih, and D. Cahyono, "Application of Classification Algorithms in Machine Learning for Phishing Detection Penerapan Algoritma Klasifikasi pada Machine Learning untuk Deteksi Phishing," vol. 5, no. April, pp. 531–540, 2025.