Vol. 6 No. 2 Desember 2025 Hal : 322 - 334
E-ISSN: 2723-1089
P-ISSN: 2776-7779

Implementasi TF-IDF dan *Cosine Similarity* untuk Penyaringan Dokumen Berita Program Makan Siang Gratis Pemerintah Indonesia

William Tanuwijaya[1)], Christofer Evan Setiawan[2)*], Hafiz Irsyad[3)] , Abdul Rahman[4)]

[1,2,3)]Prodi Informatika, Fakultas Ilmu Komputer dan Rekayasa,
Universitas Multi Data Palembang, Indonesia
[4)]Prodi Teknik Elektro, Fakultas Ilmu Komputer dan Rekayasa,
Universitas Multi Data Palembang, Indonesia

*Coresponding Email:[1]williamtanuwijaya_2226250012@mhs.mdp.ac.id,
*[2]christoferevansetiawan_2226250090@mhs.mdp.ac.id, [3]hafizirsyad@mdp.ac.id, [4]arahman@mdp.ac.id

Abstrak

Penelitian ini menerapkan metode *Information Retrieval* (IR) dalam menyaring berita yang relevan terkait program makan siang gratis yang diselenggarakan oleh pemerintah Indonesia, sebuah program yang ditujukan untuk meningkatkan gizi pelajar dan mencegah terjadinya *stunting*, namun juga menampilkan data berita dari berbagai media nasional, *preprocessing data* (termasuk *case folding, tokenisasi, stopword removal* dan *stemming*), pembobotan kata menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF), serta menggunakan pengukuran tingkat relevansi menggunakan *Cosine Similarity*. Dataset terdiri dari lima berita dengan topik terkait, yang IR mampu menyaring dokumen secara efektif. Dari lima Berita, empat di antaranya terdeteksi relevan dan satu tidak relevan. Evaluasi model menghasilkan akurasi sebesar 80%, *precision* 100%, *recall* 80% dan *f1-score* 89%. Nilai-nilai ini menunjukkan bahwa sistem dapat mengidentifikasi relevansi konten Berita terhadap topik yang terutama dalam kasus judul Berita yang bersifat *clickbait*. Penelitian ini juga memberikan kontribusi terhadap pengembangan sistem penyaringan informasi yang lebih efisien dan akurat dalam konteks isu publik.

Kata Kunci: *Cosine Similarity*, *Information Retrieval*, Penyaringan Berita, *Text Mining*, TF-IDF.

Abstract

This research applies the Information Retrieval (IR) method in filtering relevant news related to the free lunch program organized by the Indonesian government, a program aimed at improving student nutrition and preventing stunting, but also displays news data from various national media, preprocessing data (including case folding, tokenization, stopword removal and stemming), weighting words using the Term Frequency-Inverse Document Frequency (TF-IDF) method, and using a relevance measurement using Cosine Similarity. The dataset consists of five news stories with related topics, which IR is able to filter documents effectively. Of the five news stories, four were detected as relevant and one as irrelevant. Evaluation of the model resulted in 80% accuracy, 100% precision, 80% recall and 89% f1-score. These values show that the system is sufficient in identifying the relevance of the News content to the topic especially in the case of clickbait News titles. This research also contributes to the development of a more efficient and accurate information filtering system in the context of public issues.

***Keywords:* *Cosine Similarity*, *Information Retrieval*, Penyaringan Berita, *Text Mining*, TF-IDF.**

PENDAHULUAN

Dengan kemajuan teknologi informasi pada masa digital ini, penyebaran informasi bisa tersebar sangat luas dalam waktu singkat terutama melalui media online seperti portal berita. Berita yang menjadi yang sedang menjadi sorotan ramai dibahas belakangan ini adalah berita mengenai salah satu program kebijakan pemerintah Indonesia yaitu berita mengenai Makan Siang Gratis(MSG) untuk para pelajar. Program ini bertujuan memberikan makan siang gratis yang bergizi kepada anak-anak sekolah dan ibu hamil untuk memenuhi kebutuhan nutrisi harian mencegah terjadinya stunting. Dalam implementasinya, program ini malah menimbulkan banyak polemik, terutama isu mengenai anggaran[1].

Banyaknya jumlah informasi yang tersedia menjadi tantangan baru bagi masyarakat, peneliti maupun pengambil kebijakan, dimana menyaring informasi yang relevan menjadi lebih sulit. Information Retrieval (IR) merupakan salah satu pendekatan yang bisa dilakukan untuk mengatasi isu ini. Dimana IR bisa digunakan untuk mencari informasi yang relevan dari dokumen yang hasilnya berdasarkan output pengguna[2]. Penerapan IR dapat membantu menyelesaikan isu ini karena system bisa melakukan pengekstrakan dokumen dengan topik tertentu, misalnya topik terkait berita makan siang gratis.

Penerapan IR dalam pengekstrakan dokumen berita untuk berbagai macam topik, seperti politik, kesehatan, ekonomi telah sering dilakukan dalam beberapa penelitian sebelumnya. Penelitian [3] menggunakan Cosine Similarity untuk membuat sistem rekomendasi berita. Hasil dari penelitian ini menunjukan bahwa Cosine Similarity berhasil secara konsisten memberikan hasil lima rekomendasi yang sama untuk setiap data uji. Sedangkan pada penelitian [4] menggunakan TF-IDF dan K-Means Clustering untuk pengelompokan film trending di youtube menghasilkan nilai Silhouette Score rata-rata hanya 0.066. Dalam penelitian [5], TF-IDF dan Cosine Similarity digunakan untuk chatbot informasi pilkada bekasi 2024 berhasil memberikan jawaban akurat dengan akurasi 93,33%.

Penelitian ini bertujuan untuk mengimplementasikan algoritma IR dengan menggunakan TF-IDF dan Cosine Similarity untuk menyaring dokumen dengan topik berita program MSG pemerintah Indonesia. Dengan menerapkan sistem berbasis IR ini, proses penyaringan media informasi yang akurat dapat dilakukan secara efisien, sehingga mengurangi terjadinya kesalahan informasi yang beredar. Penelitian ini berkontribusi dalam mengembangakan metode penyaringan berita dengan mengimplementasikan algoritma IR menggunakan TF-IDF dan Cosine Similarity dalam konteks kebijakan politik yaitu berita program MSG pemerintah Indonesia.

LANDASAN TEORI

Berita

Berita adalah laporan mengenai fakta atau ide yang dapat dibuktikan kebenaranya, dan menarik bagi sebagian besar khalayak. Berita dapat disebarkan melalui berbagai media, mulai dari media cetak seperti surat kabar, media digital seperti televisi atau radio hingga media online seperti portal berita. [6]

Information Retrieval

Menurut Yasin & Rachman [7], Information Retrieval merupakan kegiatan menyediakan informasi yang relevan kepada pengguna sebagai jawaban dari pertanyaan langsung ataupun apa yang dibutuhkan oleh pengguna.

Term Frequency-Inverse Document Frequency (TF-IDF)

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan proses melakukan transformasi data dari yang sebelumnya tekstual ke bentuk data numerik dengan tujuan untuk dilakukan pembobotan pada tiap kata atau fitur [8]. Menurut [9], metode TF-IDF digunakan untuk menemukan seberapa besar keterhubungan atau korelasi kata (*term*) terhadap dokumen dengan cara memberikan bobot pada setiap kata (*term*).

Cosine Similarity

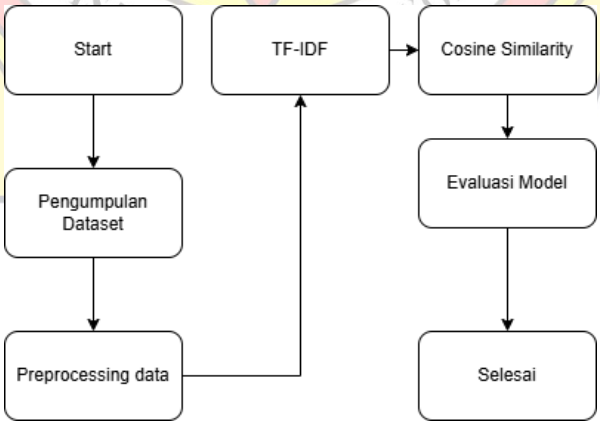
Cosine similarity merupakan salah satu metode dalam *text mining* yang mengklasifikasikan dokumen atau teks menggunakan konsep normalisasi panjang vektor. *Cosine similarity* mengukur tingkat kemiripan dokumen berdasarkan orientasi vektor yang menggambarkan representasi teks. *Cosine similarity* bekerja dengan menghitung kesamaan antar dokumen dengan cara menghitung kesamaan arah vektor [10].

Text mining

Text mining adalah proses menganalisis data dalam jumlah besar yang digunakan untuk menemukan informasi baru yang belum pernah terungkap [11]. Tujuan dari dilakukanya *text mining* ini adalah untuk menemukan informasi yang sebelumnya tidak diketahui atau belum diketahui dan belum dapat ditulis[12].

METODE PENELITIAN

Penelitian ini bersifat eksperimen dengan mengimplementasikan *Information Retrieval* untuk menyaring berita dengan topik "makan siang gratis" khususnya di Indonesia. Gambar 1 menunjukan tahapan proses yang dilakukan, mulai dari pengumpulan *dataset, preprocessing data,* TF-IDF, *Cosine Similarity.*



Gambar 1. Tahapan Penelitian.

Penjelasan tahapan pada Gambar 1 dijabarkan sebagai berikut:

1. Pengumpulan Dataset

Pengumpulan dataset dilakukan dengan mengambil data dari berbagai media Berita yang ada di Indonesia yaitu CNN Indonesia, Kompas, KabarPali, Detik dan TribunJatim.

2. *Pre-processing*

Tahap pertama yang dilakukan dalam penelitian ini adalah tahap preprocessing, dimana dilakukan proses case folding untuk mengubah semua huruf menjadi lowercase atau huruf kecil. Setelah proses case folding selesai, maka akan dilakukan penghapusan angka dan tanda baca, lalu dilanjutkan dengan proses tokenisasi, stopword removal dan proses stemming. Preprocessing ini dilakukan dengan tujuan untuk membersihkan dokumen dengan menghilangkan kata-kata yang tidak relevan agar lebih mudah untuk dianalisis.

3. *Term Frequency-Inverse Document Frequency* (TF-IDF)

Setelah tahap *preprocessing* selesai, maka dokumen dan *query* tadi akan dilakukan proses TF-IDF dimana dokumen dan *query* tadi akan direpresentasikan dalam bentuk vektor. Proses TF-IDF ini nantinya akan membantu menghitung seberapa pentingnya kata(term) dalam dokumen melalui proses pembobotan. Perhitungan Bobot setiap term di dokumen dilakukan menggunakan persamaan (1).

$$tf(t,d) = 1 + log\ tf \qquad (1)$$

tf menunjukan banyaknya kata per dokumen dan tf(t,d) menunjukan banyaknya kata t pada suatu dokumen. Setelah itu digunakan persamaan (2) untuk mencari nilai *inverse document frequency*:

$$idf(t) = \log(N/df(t)) \qquad (2)$$

N menunjukan jumlah keseluruhan dokumen dan df(t) menunjukan jumlah dokumen yang memuat atau memiliki *term* t di dalamnya. Dengan adanya pembobotan ini, maka TF-IDF dapat membantu untuk mengidentifikasi kata-kata yang sering muncul di dokumen untuk dibandingkan dengan *query* untuk menentukan apakah isi dari judul berita relevan dengan isi berita tersebut.

4. *Cosine Similarity*

Selanjutnya metode *cosine similarity* digunakan untuk mengukur tingkat kesamaan antara dokumen dengan query. *Cosine similarity* ini akan menentukan relevansi dokumen menggunakan nilai kemiripan yang dihasilkan. Jika nilai *cosine similarity* berada di atas batas tertentu akan diklasifikasikan sebagai relevan, sedangkan yang tidak melewati ambang batas dianggap tidak relevan. *Cosine Similarity* dapat dihitung menggunakan persamaan (3):

$$Cosine\ Similarity\ (A,B) = (A.B)/(|(|A|)|.|(|B|)|\ ) \quad (3)$$

A.B menunjukan produk titik vektor A dan B, sedangkan ||A|| dan ||B|| menunjukan panjang dari masing-masing vektor A dan B itu sendiri.

5. *Evaluasi model*

Dalam tahap evaluasi model, dilakukan pengukuran performa model menggunakan metrik akurasi, presisi, *recall, F1-Score*, dan *confusion matrix*. *Confusion matrix* digunakan untuk evaluasi klasifikasi model dalam memperkirakan objek yang benar atau salah. Dalam *confusion matrix* terdapat 4 kondisi, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) seperti yang ditunjukan pada Tabel 2.

Tabel 2 Confusion Matrix

	Actually Positive (1)	**Actually Positive (0)**
Predicted Positive (1)	True Positives(TP)	False Positives(FP)
Predicted Negative (0)	False Negative(FN)	True negatives(TN)

Akurasi digunakan untuk menguji kesesuaian nilai hasil prediksi model dengan nilai sampel data yang dibandingkan. Perhitungan akurasi dapat dilakukan menggunakan persamaan (4).

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{TP} + \boldsymbol{TN}}{\boldsymbol{TP} + \boldsymbol{TN} + \boldsymbol{FP} + \boldsymbol{FN}} \quad (4)$$

Precision digunakan untuk tingkat ketepatan informasi antara yang diminta oleh user dengan jawaban yang dihasilkan model. Perhitungan *precision* dapat dilakukan menggunakan persamaan (5).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5)$$

Recall adalah tingkat banyak sedikitnya informasi sesuai dari hasil model berdasarkan label yang digunakan. Perhitungan *recall* dapat dilakukan menggunakan persamaan (6).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (6)$$

F-1 Score digunakan untuk mengukur keseimbangan atau *harmonic* dari rata-rata antara *precision* dan *recall*. *F-1 Score* dapat dihitung menggunakan persamaan (7).

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

HASIL DAN PEMBAHASAN

Pada bagian ini membahas hasil dari penelitian yang telah dilakukan.

Pre-Processing data

Pada tahap awal dilakukan *preprocessing data* pada *dataset* yang telah didapatkan, pada tahap *preprocessing* ini, hal pertama yang dilakukan adalah *case folding* yaitu membuat semua jadi *lowercase*. Setelah proses *case folding* pada data selesai, maka tahap berikutnya yang dilakukan adalah penghapusan tanda baca karena tanda baca tidak akan digunakan. Selanjutnya akan dilakukan proses tokenisasi pada data dan penghilangan stopword. Yang terakhir akan dilakukan proses stemming atau menghilangkan imbuhan agar *term* menjadi bersih. Tabel 3 menunjukan hasil *Pre-Processing data.*

Tabel 3 Hasil *Pre-Processing data*

No	judul_processed_sr	isi_processed_sr	judul_processed	isi_processed
0	[sekolah, lelah, temuan, ulat, menu, makan, si...	[permasalahan, program, makan, siang, gratis, ...	[sekolah, lelah, temu, ulat, menu, makan, sian...	[masalah, program, makan, siang, gratis, makan...
1	[puluhan, siswa, sdn, simpang, raja, diduga, k...	[puluhan, siswa, sekolah, dasar, negeri, sdn, ...	[puluh, siswa, sdn, simpang, raja, duga, racun...	[puluh, siswa, sekolah, dasar, negeri, sdn, si...
2	[bapanas, program, makan, siang, gratis, digan...	[badan, pangan, nasional, bapanas, informasi, ...	[bapanas, program, makan, siang, gratis, ganti...	[badan, pangan, nasional, bapanas, informasi, ...
3	[makan, siang, bergizi, gratis, papua, butuh, ...	[februari, pelajar, provinsi, papua, pegununga...	[makan, siang, gizi, gratis, papua, butuh, aks...	[februari, ajar, provinsi, papua, gunung, turu...

4	[zulhas, tanggapi, sambat, bu, kantin, omzet, ...	[menteri, koordinator, menko, bidang, pangan, ...	[zulhas, tanggap, sambat, bu, kantin, omzet, t...	[menteri, koordinator, menko, bidang, pangan, ...

TF-IDF

Pada tahap ini dilakukan penghitungan bobot setiap kata (*term*) menggunakan TF-IDF, bobot diberikan pada setiap *term* dengan melihat seberapa sering kata muncul dalam satu dokumen dan seberapa sering kata tersebut muncul dalam keseluruhan dokumen. Tabel 4 menunjukan hasil TF-IDF pada setiap *term* (kata).

Tabel 4 Tabel hasil TF-IDF

	ac	ajar	akibat	akses	alami	...	aman	anak	anakanak
0	0.088373	0.132559	0.044186	0	0	...	0	0	0
1	0	0	0	0	0.081513	...	0.081513	0.093170	0.093170
2	0	0	0	0	0	...	0.081159	0	0
3	0	0.224654	0.022465	0.089861	0.019655	...	0.019655	0.022465	0.067396
4	0	0	0	0	0.103870	...	0	0	0

Proses perhitungan TF-IDF dilakukan menggunakan persamaan (2). Nilai pada tabel menunjukan bobot seberapa penting *term* dalam dokumen semakin besar nilai. Semakin besar nilai dalam tabel, maka semakin penting pula term tersebut dalam dokumen. Seperti terlihat pada Tabel 4, nilai suatu *term* bisa berbeda- beda setiap dataset. Misalnya pada *term* "aman" pada dataset 1 dan dataset 5 memiliki nilai 0, artinya term aman tidak penting atau tidak memiliki keterkaitan dengan dataset 1 dan 5. Tetapi, pada dataset 2, 3 dan 4 *term* "aman" memiliki hasil yang berbeda-beda dimana dalam dataset 4 term aman hanya menghasilkan nilai 0.019655 yang artinya *term* "aman" ini tidak terlalu penting dalam dataset 4 dibandingkan dengan dataset 2 dan 3 yang menghasilkan nilai di atas 0.08 yang

artinya *term* "aman" ini jauh lebih penting di dataset 2 dan 3 dibandingkan dataset 4.

Cosine Similarity

Dari hasil pembobotan TF-IDF, maka dilakukan perhitungan *cosine similarity* yang digunakan untuk menemukan dokumen mana yang paling relevan. Perhitungan cosine similarity dapat dilakukan dengan menggunakan persamaan (3).

Hasil Analisis

Tabel 5. Tabel hasil *Cosine Similarity* yang menunjukan kerelevanan berita

	Judul	Isi	*cosine_similarity*	relevan
0	Sekolah Lelah Banyak Temuan Ulat di Menu Makan Siang Gratis, Kini Minta Berhenti, Wakepsek: Butu...	Permasalahan program makan siang gratis atau Makan Bergizi Gratis (MBG) di Kota Yogyakarta membu...	0.398150	Tidak
1	Puluhan Siswa SDN 28 Simpang Raja Diduga Keracunan Usai Santap Makan Siang Gratis	Puluhan siswa Sekolah Dasar Negeri (SDN) 28 Simpang Raja, yang terletak di Kelurahan Handayani M...	0.501533	Ya
2	Bapanas Ungkap Program Makan Siang Gratis Diganti Sarapan Gratis	Badan Pangan Nasional (Bapanas) memberikan informasi terbaru terkait rencana program makan bergi...	0.566153	Ya
3	Bukan Makan Siang Bergizi Gratis, Papua Lebih Butuh Akses Pendidikan	Pada 17 Februari 2025, sejumlah pelajar di Provinsi Papua Pegunungan turun ke jalan melakukan ak...	0.504767	Ya
4	Zulhas Tanggapi Sambat Bu Kantin soal Omzet Turun gegara Makan Bergizi Gratis	Menteri Koordinator (Menko) Bidang Pangan, Zulkifli Hasan, merespons keluhan ibu kantin dengan h...	0.506594	Ya

Pada tabel 5 didapat bahwa hasil *cosine similarity* dari kelima berita yang dianalisis, satu berita yaitu dataset 1 memiliki judul dengan isi yang kurang relevan dengan topik berita yaitu makan siang gratis. Dataset 1 hanya berhasil menghasilkan nilai sebesar 0.398150 yang berada di bawah ambang batas relevan yaitu 0,5. Sedangkan dataset 2,3,4, dan 5 memiliki isi dan judul yang relevan karena berhasil mendapatkan nilai di atas ambang batas 0,5.

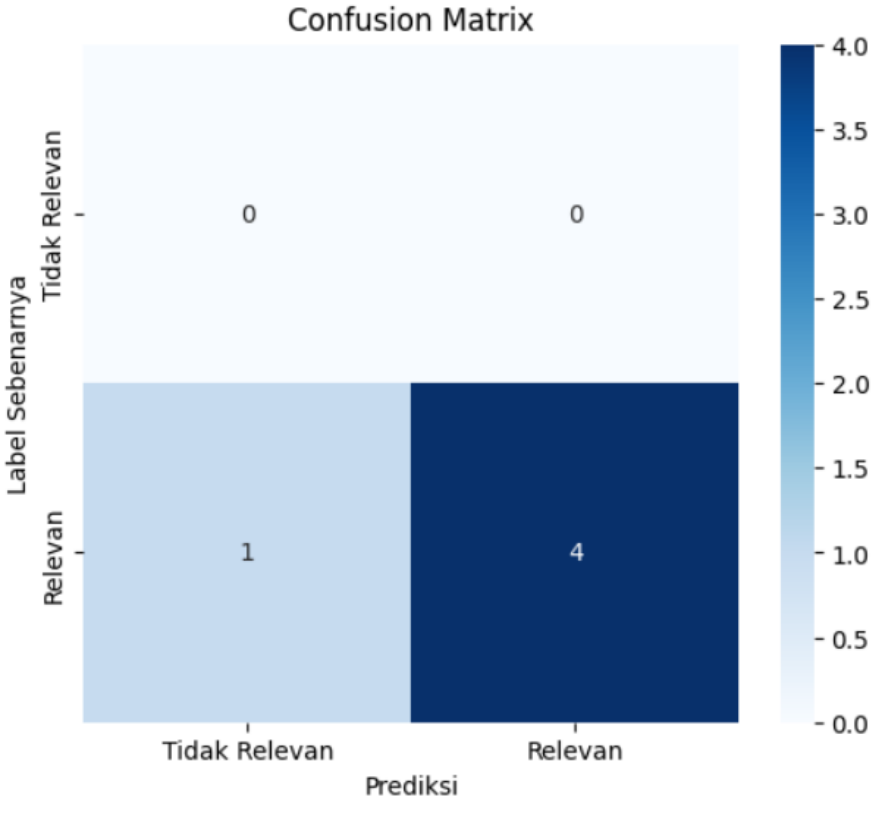
Evaluasi

Berdasarkan hasil evaluasi, model relevansi berita memiliki performa yang cukup baik dalam menentukan relevansi antara judul dan isi berita. Hasil metrik evaluasi seperti yang tertera pada tabel 6.

Tabel 6 Tabel hasil metrik evaluasi model

Metrik	**Nilai**
Accuracy	80%
Precision	100%
Recall	80%
F1-score	89%

Berdasarkan Tabel 11 hasil metrik evaluasi model relevansi berita berhasil mendapatkan *accuracy* sebesar 80%, hasil ini menunjukan bahwa model bekerja dengan baik untuk mengklasifikasikan data dengan benar untuk sebagian besar kasus. Nilai *precision* 100% menunjukan bahwa hasil prediksi model ini sangat akurat. Nilai *recall* 80% menunjukan bahwa model mampu menangkap sebagian besar dari kasus berita yang relevan. Sedangkan hasil *f1-score* 89% menandakan nilai rata-rata antara *precision* dan *recall* lumayan imbang. Hasil *Confusion Matrix* seperti tertera pada Gambar 2.



Gambar 2 *Confusion Matrix*

SIMPULAN

Berdasarkan hasil penelitian dan pembahasan kesimpulan dari implementasi *Information Retrieval* untuk menyaring dokumen berita dengan topik program MSG Pemerintah Indonesia adalah sebagai berikut:

1. Penggunaan metode *Information Retrieval* dalam penyaringan berita dengan topik program makan siang gratis Indonesia, berjalan dengan baik dimana dengan menggunakan metode *Information Retrieval* yaitu TF-IDF dan *Cosine similarity*, model berhasil untuk menentukan berita mana yang memiliki kesesuaian antara isi dengan judul dengan akurasi 80 persen.
2. Dengan menggunakan metode *Information Retrieval* ini, dapat mengurangi terjadinya kesalahan informasi yang disampaikan di judul atau *clickbait*.

DAFTAR PUSTAKA

[1] D. C. Imelda Dwi Putri Nainggolan, "Analisis Framing Portal Berita Kompas.Com dan Mediaindonesia.Com atas Pemberitaan Program Makan Siang Gratis oleh Paslon 02 (Prabowo-Gibran) Pada Periode 28 November 2023 - 10 Februari 2024," vol. 6, pp. 2266–2282, 2024, doi: 10.47476/reslaj.v6i10.3038.

[2] N. D. S. Muhammad Irfan Luthfi, "Pengembangan Mesin Tanya Jawab Artikel Ilmiah Berbahasa Indonesia," no. September, 2024.

[3] A. Muhammad, R. Haz, A. N. Rohman, S. Informasi, F. I. Komputer, and U. A. Yogyakarta, "SISTEM REKOMENDASI BERITA DENGAN METODE," vol. 9, no. 2, pp. 143–150, 2025.

[4] B. D. R. Dwi Remawati, Hendro Wijayanto, Yustina Retno Wahyu Utami, "Pengelompokkan Film Trending di Youtube Menggunakan TF-IDF dan," vol. 4, pp. 65–74, 2025.

[5] A. Muhammad, R. Haz, A. N. Rohman, S. Informasi, F. I. Komputer, and U. A. Yogyakarta, "SISTEM REKOMENDASI BERITA DENGAN METODE," vol. 9, no. 2, pp. 143–150, 2025.

[6] E. P. Bangun, F. V. I. A Koagouw, and J. S. Kalangi, "Analisis Isi Unsur Kelengkapan Berita Pada Media Online Manadopostonline.com," Acta Diurna Komun., vol. 1, no. 3, pp. 4–13, 2019, [Online]. Available: https://ejournal.unsrat.ac.id/index.php/actadiurnakomunikasi/article/view/25560

[7] S. A. Putri, Y. Winoto, and R. Rohanda, "Pemetaan penelitian information retrieval system menggunakan VOSviewer," Informatio J. Libr. Inf. Sci., vol. 3, no. 2, p. 93, 2023, doi: 10.24198/inf.v3i2.46646.

[8] N. Silalahi and Guidio Leonarde Ginting, "Rekomendasi Berita Berkaitan dengan Menerapkan Algoritma Text Mining dan TF-IDF," Bull. Comput. Sci. Res., vol. 3, no. 4, pp. 276–282, 2023, doi: 10.47065/bulletincsr.v3i4.266.

[9] D. Andreswari, D. Suranti, and D. A. Trianggara, "Application Of Text Mining In Grouping Thesis Topics Using TF-IDF Method Based On Thesis Abstract Penerapan Text Mining Dalam Pengelompokkan Topik Skripsi Menggunakan Metode TF-IDF Berdasarkan Abstrak Skripsi," vol. 3, no. 2, pp. 69–78, 2024.

[10] A. PUTRI, "Sistem Rekomendasi Skincare Dengan Metode Keyword Extraction Dan Cosine Similarity," Repository.Unissula.Ac.Id, vol. 13, no. 1, pp. 104–116, 2023.

[11] Z. Alhaq and J. D. S. , Ali Mustopa , Sri Mulyatun, "PENERAPAN METODE SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER," Media Inform., vol. 20, no. 2, pp. 97–108, 2021, doi: 10.37595/mediainfo.v20i2.59.

[12] S. M. Fani, R. Santoso, and S. Suparti, "Penerapan Text Mining Untuk Melakukan Clustering Data Tweet Akun Blibli Pada Media Sosial Twitter Menggunakan K-Means Clustering," J. Gaussian, vol. 10, no. 4, pp. 583–593, 2021, doi: 10.14710/j.gauss.v10i4.30409.