

Penerapan Algoritma TF-IDF dan *Cosine Similarity* untuk *Query* Pencarian Pada Dataset Destinasi Wisata

Rio Al Rasyid ^{1*}, Dewi Handayani Untari Ningsih ²

^{1,2} Program Studi Teknik Informatika, Fakultas Teknologi Informasi dan Industri, Universitas Stikubank, Kota Semarang, Provinsi Jawa Tengah, Indonesia.

article info

Article history:

Received 21 July 2023

Received in revised form

20 November 2023

Accepted 3 December 2023

Available online January 2024

DOI:

<https://doi.org/10.35870/jtik.v8i1.1416>

Keywords:

TF-IDF; Cosine Similarity;

Search Queries; Datasets;

Tourist.

Kata Kunci:

TF-IDF; Cosine Similarity;

Kueri Pencarian; Dataset;

Pariwisata.

abstract

This research aims to improve the search for tourist destinations in 50 datasets by using search queries to find relevant documents. By optimizing the search process, the goal is to create an accurate list of tourist destinations based on a given query. To achieve this, researchers used the TF-IDF and Cosine Similarity algorithms to retrieve and compare information, measuring similarity scores between search queries and tourist destinations in the dataset. Finally, the list of tourist destinations is ranked based on the similarity score measurement. The methods used are TF-IDF and Cosine Similarity. The fifty datasets containing text content documents were normalized through pre-processing stages, namely Case Folding, Stopword Removal, and Tokenization. Documents that have been normalized are then processed again through TF-IDF weighting. TF-IDF weighting is also applied to search queries. The similarity calculation between the TF-IDF vector from the document and the TF-IDF vector from the search query is carried out using Cosine Similarity to obtain a similarity score for each document based on the search query. Testing was carried out on 5 different queries, and precision testing results were obtained with an average value of 83%.

abstract

Penelitian ini bertujuan untuk meningkatkan pencarian destinasi wisata dalam 50 dataset dengan menggunakan kueri pencarian untuk menemukan dokumen yang relevan. Dengan mengoptimalkan proses pencarian, tujuannya adalah menciptakan daftar akurat destinasi wisata berdasarkan kueri yang diberikan. Untuk mencapai hal ini, peneliti menggunakan algoritma TF-IDF dan Cosine Similarity untuk mengambil dan membandingkan informasi, mengukur skor kesamaan antara kueri pencarian dan destinasi wisata dalam dataset. Akhirnya, daftar destinasi wisata diberi peringkat berdasarkan pengukuran skor kesamaan tersebut. Metode yang digunakan adalah TF-IDF dan Cosine Similarity. Kelima puluh dataset yang berisi dokumen konten teks dinormalisasikan melalui tahapan pra-pemrosesan, yaitu Case Folding, Stopword Removal, dan Tokenization. Dokumen yang telah dinormalisasikan kemudian diproses lagi melalui pembobotan TF-IDF. Pembobotan TF-IDF juga diterapkan pada kueri pencarian. Perhitungan kesamaan antara vektor TF-IDF dari dokumen dan vektor TF-IDF dari kueri pencarian dilakukan menggunakan Cosine Similarity untuk mendapatkan skor kesamaan untuk setiap dokumen berdasarkan kueri pencarian. Pengujian dilakukan pada 5 kueri yang berbeda, dan hasil pengujian presisi diperoleh dengan nilai rata-rata sebesar 83%.

Corresponding Author. Email: rio.rasyid23@gmail.com ^{1}.



ACM Computing Classification System (CCS)



Communication and Mass Media Complete (CMMC)

© E-ISSN: 2580-1643.

Copyright © 2024 by the authors of this article. Published by Lembaga Otonom Lembaga Informasi dan Riset Indonesia (KITA INFO dan Riset). This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

1. Latar Belakang

Pariwisata merupakan sektor yang sangat bergantung pada penyediaan informasi. Para pelancong yang memutuskan untuk menjelajahi destinasi baru, terutama di negara asing, menghadapi risiko yang cukup besar. Individu-individu ini menginvestasikan waktu dan uang yang signifikan untuk mendapatkan pengalaman liburan yang paling memuaskan. Oleh karena itu, sebelum membuat keputusan akhir mengenai destinasi liburan mereka, mereka melakukan penelitian yang menyeluruh melalui berbagai saluran seperti mesin pencari, media sosial, dan situs web [1]. Teknik optimisasi pencarian memainkan peran penting dalam meningkatkan pengalaman pencarian bagi pengguna. Teknik-teknik ini bertujuan untuk mengoptimalkan proses pencarian dengan menerapkan algoritma dan pendekatan baru untuk meningkatkan efisiensi dan efektivitas pencarian berbasis kata kunci [2]. Metode pencarian berbasis kata kunci tradisional tidak selalu menghasilkan hasil yang akurat dan tepat sasaran, terutama ketika berhadapan dengan data tidak terstruktur seperti deskripsi teks tentang destinasi wisata. Untuk mengatasi keterbatasan ini, diperlukan algoritma dan metode canggih untuk meningkatkan presisi dan efisiensi proses pencarian.

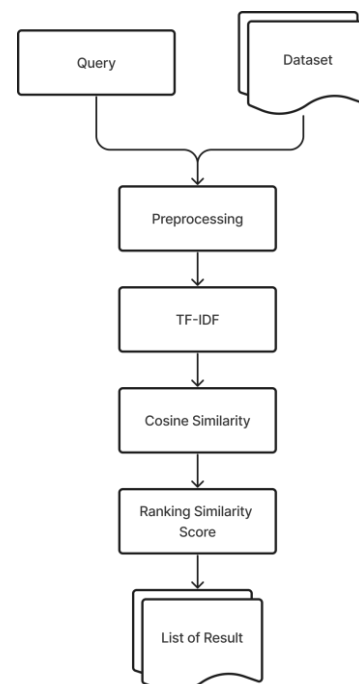
Terdapat banyak algoritma dan metode yang tersedia untuk pengolahan data, tetapi penelitian ini berfokus pada dua yang dikenal dengan nama TF-IDF dan *Cosine Similarity* [3]. TF-IDF adalah ukuran kuantitatif yang menunjukkan signifikansi kata kunci dalam dokumen tertentu. Dengan kata lain, algoritma ini menawarkan kata kunci yang dapat digunakan untuk mengidentifikasi atau mengategorikan dokumen tertentu [3]. Sementara itu, *Cosine Similarity* adalah ukuran yang umum digunakan dalam bidang pengambilan informasi dan area studi serupa. Ukuran ini menggambarkan dokumen teks sebagai vektor yang terdiri dari istilah-istilahnya. Dengan menggunakan model ini, kesamaan antara dua dokumen dapat ditentukan dengan menghitung nilai kosinus antara vektor istilah masing-masing dokumen [4].

Berdasarkan hasil pengujian [5], sistem rekomendasi yang menggunakan kerangka kerja *6 As Tourism Destinations Framework Based Multi-Criteria* pada penelitian sebelumnya terbukti dapat digunakan

secara optimal untuk merekomendasikan daftar destinasi wisata menggunakan 5 tujuan yang dimasukkan dengan skor presisi sebesar 0.622 (62,2%). Studi ini menggunakan model 6AsTD sebagai sarana untuk memperoleh data dan mengevaluasi informasi tentang karakteristik destinasi wisata sebagaimana yang dialami oleh para pengunjung [6]. Penelitian ini memberikan kontribusi bagi bidang pengambilan informasi pariwisata dengan menyajikan pendekatan inovatif untuk mengoptimalkan pencarian destinasi wisata. Metode yang diusulkan meningkatkan presisi dan relevansi hasil pencarian, memberdayakan para pelancong untuk mengakses beragam informasi perjalanan dan membuat keputusan yang berdasarkan informasi. Tinjauan pustaka meliputi pembahasan tentang pra-pemrosesan data teks, Term Frequency-Inverse Document Frequency (TF-IDF), serta *Cosine Similarity*.

2. Metode Penelitian

Studi ini menggunakan beberapa tahap, termasuk: Tahap Pra-pemrosesan, Menghitung Bobot TF-IDF untuk Kueri dan Dataset dokumen, Mengukur skor kesamaan menggunakan *Cosine Similarity*, dan kemudian memberikan peringkat hasil berdasarkan skor *similarity*.



Gambar 1. Alur Tahapan Penelitian

Preprocessing

Sebuah koleksi konten/deskripsi yang terdapat dalam dokumen akan disiapkan dan diproses sehingga membentuk kelompok kata/*term* [7]. Proses ini terdiri dari tahap *case folding*, tokenisasi, penyaringan dengan menghapus kata-kata penghenti (*stopwords*), dan stemming.

1) Case Folding

Setiap dokumen atau dataset yang berisi paragraf konten/deskripsi akan mengalami proses case folding. Case folding merujuk pada prosedur mengubah semua teks menjadi huruf kecil dan menghilangkan karakter yang berada di luar rentang 'a' hingga 'z'. Tanda baca, angka, dan simbol juga dikecualikan selama proses ini [8].

2) Penyaringan

Selama tahap penyaringan, ada langkah yang disebut penghapusan *stopwords*. *Stopwords* adalah kata-kata yang kurang memiliki makna penting atau memiliki bobot kurang signifikan, dan sering muncul dalam kumpulan kata-kata [9].

3) Tokenisasi

Tokenisasi adalah prosedur yang mengubah kalimat menjadi kata-kata individual. Ini melibatkan memisahkan kalimat menjadi unit-unit terpisah berdasarkan pemisah yang membentuknya, yang dalam hal ini adalah spasi (" ") [10].

Term Frequency (TF)

Metode TF-IDF, atau Term Frequency Inverse Document Frequency, adalah teknik yang digunakan untuk menentukan bobot suatu istilah terkait dengan sebuah dokumen. Pendekatan ini menggabungkan dua prinsip untuk menghitung bobot, yaitu frekuensi kemunculan suatu istilah dalam dokumen tertentu dan frekuensi invers dari dokumen yang mengandung istilah tersebut. Menilai frekuensi kemunculan kata dalam suatu dokumen memberikan wawasan tentang signifikansi istilah dalam dokumen tersebut [11]. TF-IDF adalah istilah gabungan yang terdiri dari dua kata yang berbeda: Term Frequency dan Inverse Document Frequency. TF digunakan untuk mengukur frekuensi kemunculan suatu istilah dalam dokumen [12]. Misalkan kita memiliki dokumen "D1" yang terdiri dari 5000 kata dan mengandung kata "Magnificent" sebanyak sepuluh kali. Sebagai kenyataan yang telah dikenal, panjang dokumen dapat bervariasi dari pendek hingga sangat panjang, oleh karena itu, ada kemungkinan bahwa

suatu istilah muncul lebih sering dalam dokumen yang lebih besar daripada dalam dokumen yang lebih kecil. Untuk mengatasi masalah ini, frekuensi kemunculan suatu istilah dalam suatu dokumen dibagi dengan jumlah total kata dalam dokumen tersebut untuk menentukan frekuensi istilah. Dalam skenario yang telah disebutkan sebelumnya, frekuensi istilah "Magnificent" dalam dokumen "D1" akan menjadi: $TF = 10/5000 = 0.002$ dengan menggunakan rumus dibawah ini.

$$tf(t, d) = \frac{f_{t,d}}{\sum_t i}$$

$$tf(t, d) = \frac{f_{t,d}}{\sum t^i \in d f_{t^i,d}}$$

Inverse Document Frequency (IDF)

Ketika Term Frequency dalam suatu dokumen dihitung, dapat diamati bahwa algoritma memperlakukan semua kata kunci dengan sama, tanpa memperhatikan apakah itu adalah stop word seperti "of", yang merupakan hal yang tidak benar [3]. Inverse document frequency (IDF) memberikan bobot yang lebih rendah untuk kata-kata yang sering muncul dan memberikan bobot yang lebih tinggi untuk kata-kata yang jarang muncul. Penggunaan stop word adalah praktik umum dalam pemrosesan dokumen. Namun, beberapa stop word seperti "of" dapat memiliki makna terbatas atau bahkan tidak digunakan sama sekali dalam konteks tertentu. Di sinilah konsep IDF, atau inverse document frequency, berperan. IDF memberikan bobot untuk kata-kata berdasarkan frekuensi kemunculannya dalam satu set dokumen. Kata-kata yang sering muncul diberi bobot yang lebih rendah sementara kata-kata yang jarang muncul diberi bobot yang lebih tinggi. Untuk mengilustrasikan ini, pertimbangkan sebuah set dari 10 dokumen, di mana istilah "technology" muncul dalam 5 di antaranya. Dengan menerapkan rumus $IDF = \log_e (10/5) = 0.3010$, kita dapat menghitung inverse document frequency untuk istilah "technology". Berikut adalah formula dari IDF.

$$idf(r, D) = \log \frac{N}{|\{d \in D: r \in d\}|}$$

Cosine Similarity

Similarity atau relevansi dihitung antara kueri dan dokumen. Skor similarity diperoleh dengan mengukur skor similarity antara dua vektor, yaitu vektor kueri

dan vektor dokumen. Semakin besar nilai relevansi, semakin mirip/relevan kueri dan dokumen tersebut. Ukuran similarity digunakan untuk menentukan tingkat similarity antara titik data [13]. Selain itu, ukuran ini digunakan untuk mengukur kesamaan antara himpunan berdasarkan irisan kedua himpunan tersebut. Selain itu, ukuran kesamaan diakui sebagai fungsi yang mengevaluasi tingkat kemiripan antara sepasang objek teks. Secara ringkas, tingkat similarity mencerminkan intensitas hubungan antara dua titik data [14]. *Cosine Similarity* adalah ukuran kesamaan yang sering digunakan. Ini digunakan sebagai dasar untuk menghitung nilai relevansi antara kueri dan dokumen, serta nilai relevansi antara dokumen yang berbeda. *Cosine Similarity* ditentukan sebagai kosinus sudut θ yang terbentuk antara vektor-vektor [15]. Formula *Cosine Similarity* memiliki rumus sebagai berikut:

$$\begin{aligned} \text{cosine similarity} &= S_c(A, B) := \cos(\theta) \\ &= \frac{A \cdot B}{\|A\| \|B\|} \\ &= \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \end{aligned}$$

Keterangan:

A = Query TF-IDF Weight Vector

B = Document TF-IDF Weight Vector.

3. Hasil dan Pembahasan

Hasil Preprocessing

Desain dataset yang digunakan dalam penelitian ini terdiri dari data tentang nama-nama destinasi wisata dan deskripsi konten dari tempat-tempat tersebut. Dataset ini dibagi menjadi beberapa kategori pariwisata, masing-masing berisi bidang seperti nama destinasi wisata, URL Google Maps, dan deskripsi/konten destinasi wisata. Bidang-bidang ini difilter terlebih dahulu untuk memastikan bahwa proses pelatihan dapat dilakukan menggunakan bidang deskripsi/konten. Tahap Preprocessing dilakukan melalui tahap Case Folding, Stopword Removal, & Tokenisasi. Berikut ini contoh satu kueri masukan dan contoh dokumen target pada tahap preprocessing:

Query : “Wisata Alam Goa Kreo Kota Semarang”
Document :

Tabel 1. Dokumen Target Sampel

Nama Destinasi	Konten
GOA	Goa Kreo merupakan salah satu daya tarik wisata alam di Kota Semarang yang mahsyur dengan kisah Sunan Kalijaga mencari kayu jati untuk membangun Masjid Agung Demak. Sunan Kalijaga pun sampai di sebuah goa dan didatangi empat ekor kera. Kemudian, Sunan Kalijaga meminta kera-kera tersebut untuk tinggal atau menjaga di sekitar Goa, maka jadilah Goa Kreo. Kata “Kreo” berasal dari Mangreho yang berarti penjaga.
KREO	

Result Query : “wisata alam goa kreo kota semarang”
Result Document :

Tabel 2. Hasil Preprocessing Dokumen

Nama Destinasi	Konten
GOA	"goa" "kreo" "salah" "daya" "tarik"
KREO	"wisata" "alam" "kota" "semarang" "mahsyur" "kisah" "sunan" "kalijaga" "mencari" "kayu" "jati" "membangun" "masjid" "agung" "demak" "sunan" "kalijaga" "goa" "didatangi" "ekor" "kera" "sunan" "kalijaga" "kerakera" "tinggal" "menjaga" "goa" "goa" "kreo" "kreo" "berasal" "mangreho" "penjaga"

Hasil TF-IDF

Setelah tahap preprocessing diterapkan pada dokumen, langkah selanjutnya adalah menghitung bobot Vektor TF-IDF untuk dokumen berdasarkan kata kunci kueri, yang dalam kasus ini adalah "wisata alam kota semarang". Berikut ini adalah hasil pembobotan TF-IDF pada contoh dokumen:

Tabel 3. Hasil TF-IDF Vector Dokumen

Dokumen	Pembobotan Dokumen (TF-IDF)			
	Query	TF	IDF	TF-IDF
GOA KREO	Wisata	0.026	1.478	0.038
	Alam	0.026	2.514	0.066
	Goa	0.102	4.218	0.444

Kreo	0.076	4.218	0.333
Kota	0.026	1.653	0.043
Semarang	0.026	1.150	0.030

Untuk memenuhi persyaratan pengukuran kesamaan, metode yang sama diterapkan, yaitu menghitung Vektor TF-IDF untuk kueri. Jika perhitungan Vektor TF-IDF sebelumnya dilakukan antara kata kunci kueri dan dokumen contoh, kali ini pembobotan Vektor TF-IDF dilakukan pada kata kunci kueri terhadap kueri itu sendiri.

Tabel 4. Hasil TF-IDF Vector Query

Query	Query Keyword	TF-IDF
wisata alam goa kreo kota semarang	Wisata	0.2463
	Alam	0.4190
	Goa	0.7031
	Kreo	0.7031
	Kota	0.2756
	Semarang	0.1918

Hasil Cosine Similarity

Tahap akhir adalah menghitung *Cosine Similarity* untuk mendapatkan skor untuk setiap dokumen, yang akan digunakan untuk menentukan peringkat kesamaan dokumen dengan kueri masukan. Dari kueri masukan "wisata alam goa kreo kota semarang," hasil *Cosine Similarity* berikut diperoleh dalam tabel di bawah ini.

Tabel 5. Hasil *Cosine Similarity*

Query	Document	Similarity
Wisata alam goa kreo kota semarang	GOA KREO	0.9239

Hasil Perangkingan

Berdasarkan kueri masukan "wisata alam goa kreo kota semarang," satu dari 50 dokumen ditemukan memiliki kesamaan dan dimasukkan dalam daftar peringkat. Hasil perhitungan peringkat untuk daftar destinasi wisata berdasarkan skor peringkat menggunakan *Cosine Similarity* dapat dilihat dalam tabel di bawah ini.

Tabel 6. Hasil Perangkingan Dokumen

Top-N	Nama Destinasi	Similarity Score
Document-28	GOA KREO	0.9239

Dapat dilihat bahwa Dokumen 28, yaitu GOA KREO, adalah satu-satunya dokumen yang memiliki peringkat skor kesamaan dibandingkan dengan dokumen lainnya, dengan skor kesamaan sebesar 0,9239 (92%). Dapat disimpulkan bahwa Dokumen 28, GOA KREO, memiliki kesamaan dengan kueri "wisata alam goa kreo kota semarang," sehingga muncul di bagian atas daftar peringkat dengan peringkat kesamaan tertinggi dari 1 dokumen yang terdaftar dalam peringkat.

Hasil Pengujian

Pengukuran relevansi dan kualitas dari penelitian ini dilakukan menggunakan metode Pengujian Presisi. Rumus untuk menghitung perhitungan Presisi dapat dilihat sebagai berikut.

Precisionn

$$= \frac{\text{Relevant document found}}{\text{All document total found on ranking list}}$$

Tabel di bawah ini menunjukkan hasil daftar peringkat berdasarkan skor kesamaan dan skor presisi untuk masing-masing kueri yang berbeda. Pengujian dilakukan pada 5 jenis kueri yang berbeda sebagai berikut:

Tabel 7. Query Sample

Query
Q1 wisata alam goa kreo kota semarang
Q2 pantai
Q3 taman bermain
Q4 pemandangan alam
Q5 desa wisata

Selain itu, setiap kueri juga menampilkan peringkat yang berbeda dari dokumen-dokumen relevan yang ditemukan dan jumlah total dokumen yang dimasukkan dalam daftar peringkat.

Tabel 8. Hasil Pengujian Q1

Query	Ranking List		
	Document	Nama Destinasi	Similarity Score
Q1 (wisata alam goa kreo kota semarang)	Document-28	GOA KREO	0.9239

Tabel 9. Hasil Pengujian Q2

Query	Ranking List		
	Document	Nama Destinasi	Similarity Score
Q2 (pantai)	Document -13	PANTAI MARINA	1.00
	Document -21	MANGROVE MAEROKOCO	1.00
	Document -39	HUTAN MANGROVE TAPAK	1.00
	Document -40	PANTAI BARUNA	1.00

Tabel 10. Hasil Pengujian Q3

Query	Ranking List		
	Rank List	Nama Destinasi	Similarity Score
Q3 (taman bermain)	Document t-8	DESA WISATA KANDRI	0.8409
	Document t-14	TAMAN LELE	0.5411
	Document t-15	KOTA LAMA SEMARANG	0.5411
	Document t-16	GRAND MAEROKOCO	0.5411
	Document t-17	TUGU MUDA	0.5411
	Document t-20	SEMARANG ZOO	0.5411
	Document t-27	TAMAN BUDAYA RADEN SALEH	0.5411
	Document t-31	KAWASAN KOTA TUA SEMARANG	0.5411
	Document t-50	TAMAN SRIGUNTING	0.5411

Tabel 11. Hasil Pengujian Q4

Query	Ranking List		
	Rank List	Nama Destinasi	Similarity Score
Q4 (pemandangan alam)	Document-32	UMBUL SONGO KOPENG	1.00
	Document-48	WISATA LERENG KELIR	1.00
	Document-41	GUNUNG TELOMOYO	0.8809
	Document-26	GOMBEL GOLF SEMARANG	0.7628
	Document-34	BROWN CANYON	0.7628
	Document-46	CANDI GEDONG SONGO	0.7628
	Document-28	GOA KREO	0.6465
	Document-37	UMBUL SIDOMUKTI	0.6465
	Document-38	WANA PENGGARON UNGARAN	0.6465
	Document-8	DESA WISATA KANDRI	0.6465
	Document-25	WADUK JATIBARANG	0.6465
	Document-33	WATU GUNUNG UNGARAN	0.6465
	Document-36	CURUG LAWE & CURUG BENOWO	0.6465

Tabel 12. Hasil Pengujian Q5

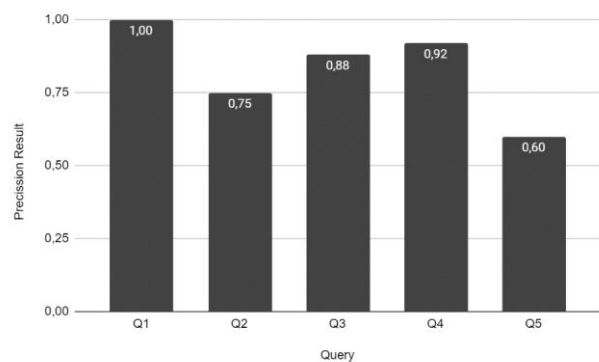
Query	Ranking List		
	Document	Nama Destinasi	Similarity Score
Q5 (desa wisata)	Document -8	DESA WISATA KANDRI	1.00
	Document -7	DESA WISATA WONOLOPO	0.9780
	Document -49	DESA WISATA KALI PANCUR	0.9101
	Document -36	CURUG LAWE & CURUG BENOWO	0.7459
	Document -38	WANA PENGGARON UNGARAN	0.7037

Tabel 13. Hasil Pengujian Presisi

No	Query	Relevant Document Found	Total Document Found	Precision
1	wisata alam goa kreo kota semarang	1	1	1.0
2	pantai	3	4	0.75
3	taman bermain	8	9	0.88
4	pemandangan alam	12	13	0.92
5	desa wisata	3	5	0.6
Total				4.15
Average				0.83
%				83%

Pengujian yang dilakukan terhadap sistem menggunakan berbagai kueri telah menghasilkan hasil yang menarik. Setiap kueri menghasilkan daftar peringkat yang berbeda-beda, dengan setidaknya satu dokumen relevan ditemukan dalam setiap daftar. Presisi dari dokumen yang ditemukan bervariasi dari satu kueri ke kueri lainnya, dengan skor rata-rata keseluruhan mencapai 0.83 atau 83%. Ketika melihat

secara spesifik, kueri dengan presisi tertinggi adalah kueri nomor 1, "wisata alam goa kreo kota semarang", yang mencapai skor presisi sempurna, yaitu 1.0 atau 100%. Namun, ada juga kueri lain yang memiliki presisi yang cukup tinggi, seperti kueri "taman bermain" dan "pemandangan alam", yang masing-masing memiliki presisi 0.88 dan 0.92. Kemudian, untuk membandingkan kinerja sistem ini dengan sistem sebelumnya yang menggunakan 6AsTD Framework, dapat dilihat bahwa sistem yang menggunakan TF-IDF & Cosine Similarity telah mengalami peningkatan signifikan dalam presisi, dengan skor presisi sebesar 0.83 dibandingkan dengan 0.622 pada 6AsTD Framework. Hal ini menunjukkan bahwa penggunaan metode TF-IDF & Cosine Similarity dapat memberikan hasil yang lebih baik dalam mengidentifikasi dokumen relevan dalam pencarian informasi wisata.



Gambar 2. Grafik Perbandingan Presisi Setiap Query

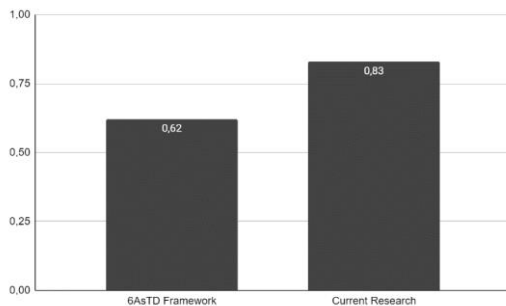
Pengujian dengan kueri yang berbeda menghasilkan jumlah daftar peringkat yang berbeda juga. Setiap kueri menghasilkan daftar dokumen dengan setidaknya 1 dokumen relevan ditemukan. Pengujian presisi terhadap dokumen yang dihasilkan dalam daftar ranking menghasilkan skor yang berbeda-beda untuk setiap query. Hasil pengujian presisi keseluruhan untuk setiap query kemudian dihitung, sehingga menghasilkan skor rata-rata sebesar 0,83 (83%). Query dengan skor presisi tertinggi dapat dilihat pada query nomor 1 yang mencapai skor presisi 1,0 (100%). Perbandingan dengan sistem sebelumnya, yang menggunakan 6AsTD Framework, menunjukkan peningkatan yang signifikan dalam presisi dengan menggunakan metode TF-IDF & Cosine Similarity. Hal ini mengindikasikan bahwa metode yang digunakan dalam penelitian saat ini lebih efektif dalam menemukan dokumen relevan dalam

konteks pencarian informasi wisata. Hasil perbandingan perbandingan antara penelitian ini dengan penelitian yang menggunakan sistem 6AsTD Framework sebelumnya dapat dilihat pada tabel dibawah.

Tabel 14. Perbandingan Presisi

Sistem	Presisi
6AsTD Framework	0.622
TF-IDF & Cosine Similarity	0.83

Grafik Perbandingan Presisi Sistem dapat dilihat pada Gambar 3. Penambahan metode TF-IDF & Cosine Similarity dalam penelitian ini menghasilkan peningkatan signifikan dalam presisi sistem, memperbaiki kemampuan sistem dalam menemukan informasi yang relevan dengan kueri pengguna dalam konteks wisata.



Gambar 3. Grafik Perbandingan Presisi Sistem

4. Kesimpulan

Penelitian ini berhasil mengoptimalkan pencarian kueri pada dataset destinasi wisata dengan menggabungkan dua algoritma, yaitu TF-IDF dan Cosine Similarity. Sistem Peringkat dievaluasi menggunakan pengujian presisi. Berdasarkan pengujian dari 5 kueri yang berbeda, Sistem Peringkat berhasil mencapai hasil presisi optimal dengan nilai rata-rata 0.83. Dibandingkan dengan penelitian sebelumnya yang menggunakan Kerangka Kerja 6AsTD dan memperoleh hasil presisi sebesar 0.622 (62,2%), penelitian ini berhasil mencapai peningkatan presisi sebesar 20,8%, mencapai nilai 0.83 (83%).

5. Daftar Pustaka

- [1] Sipayung, E.M., Fiarni, C. and Febrian, M., 2021, October. Implementation of Search Engine Optimization (SEO) in Wellness and Beauty Tourism Industry. In *2021 8th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)* (pp. 397-402). IEEE. DOI: <https://doi.org/10.23919/EECSI53397.2021.9624309>.
- [2] Ravichandiran, A., Vijayan, A. and Ravikumar, K., 2015. Memory Optimization Using Genetic Algorithm of Relational Keyword Search Techniques. *Memory*, 3(4).
- [3] Qaiser, S. and Ali, R., 2018. Text mining: use of TF-IDF to examine the relevance of words to documents. *International Journal of Computer Applications*, 181(1), pp.25-29.
- [4] Salton, G. and Buckley, C., 1988. Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5), pp.513-523. DOI: [https://doi.org/10.1016/0306-4573\(88\)90021-0](https://doi.org/10.1016/0306-4573(88)90021-0).
- [5] Arif, Y.M., Putra, D.D. and Khan, N., 2023. Selecting tourism site using 6 as tourism destinations framework based multi-criteria recommender system. *Applied Information System and Management (AISM)*, 6(1), pp.7-12.
- [6] Arif, Y.M., Nurhayati, H., Harini, S., Nugroho, S.M.S. and Hariadi, M., 2020, February. Decentralized tourism destinations rating system using 6AsTD framework and blockchain. In *2020 international conference on smart technology and applications (ICoSTA)* (pp. 1-6). IEEE. DOI: <https://doi.org/10.1109/ICoSTA48221.2020.1570614662>.

- [7] Liu, G., Lee, K.Y. and Jordan, H.F., 1997. TDM and TWDM de Bruijn networks and shufflenets for optical communications. *IEEE Transactions on Computers*, 46(6), pp.695-701. DOI: <https://doi.org/10.1109/12.600827>.
- [8] Gunawansyah, R. and Nurwathi, S., 2020. Automated essay scoring using natural language processing and text mining method. In *Proceeding of 14th International Conference on Telecommunication Systems, Services, and Applications*.
- [9] Pathak, P., Raghav, S., Jain, S. and Jalal, S., 2021, October. Essay Rating System Using Machine Learning. In *2021 5th International Conference on Information Systems and Computer Networks (ISCON)* (pp. 1-6). IEEE. DOI: <https://doi.org/10.1109/ISCON52037.2021.9702504>.
- [10] Ratna, A.A.P., Santiar, L., Ibrahim, I., Purnamasari, P.D., Luhurkinanti, D.L. and Larasati, A., 2019, October. Latent semantic analysis and winnowing algorithm based automatic Japanese short essay answer grading system comparative performance. In *2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST)* (pp. 1-7). IEEE. DOI: <https://doi.org/10.1109/ICAwST.2019.8923226>.
- [11] Robertson, S., 2004. Understanding inverse document frequency: on theoretical arguments for IDF. *Journal of documentation*, 60(5), pp.503-520.
- [12] Hakim, A.A., Erwin, A., Eng, K.I., Galinium, M. and Muliady, W., 2014, October. Automated document classification for news article in Bahasa Indonesia based on term frequency inverse document frequency (TF-IDF) approach. In *2014 6th international conference on information technology and electrical engineering (ICITEE)* (pp. 1-4). IEEE. DOI: <https://doi.org/10.1109/ICITEED.2014.7007894>.
- [13] Ayob, A., 2019. Comparison between conventional and digital essay writing assessment system: Consumer concept and user friendly. *Research in World Economy*, 10(2), pp.96-101.
- [14] Yang, Y., Xia, L. and Zhao, Q., 2019. An automated grader for Chinese essay combining shallow and deep semantic attributes. *IEEE Access*, 7, pp.176306-176316. DOI: <https://doi.org/10.1109/ACCESS.2019.2957582>.
- [15] Yulita, W., Untoro, M.C., Praseptiawan, M., Ashari, I.F., Afriansyah, A. and Pee, A.N.B.C., 2023. Automatic Scoring Using Term Frequency Inverse Document Frequency Document Frequency and Cosine Similarity. *Scientific Journal of Informatics*, 10(2), pp.93-104. DOI: <https://doi.org/10.15294/sji.v10i2.42209>.