

## SISTEM REKOMENDASI BERITA DENGAN METODE CONTENT-BASED FILTERING

Anando Muhammad Rahul Haz<sup>1)</sup>, Arif Nur Rohman<sup>2)</sup>

<sup>1,2</sup> Sistem Informasi, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta,  
Jl. Ring Road Utara, Condongcatur, Sleman, Yogyakarta  
Co Responden Email: anandorahulhz@gmail.com

### Abstract

*The continuous growth of digital content on a daily basis presents challenges in filtering and presenting relevant information to internet users. There is a need for a recommendation system that can assist users in filtering information effectively. This study aims to develop a news article recommendation system based on content-based filtering using the cosine similarity algorithm. The dataset for this research consists of 150 news articles sourced from the Kaggle platform. The data pre-processing stage includes data normalization, data cleaning, removal of duplicate entries, stop word removal, and tokenization. The testing process involves selecting 15 article samples, each producing five recommendations based on the highest similarity scores. The results of this study are beneficial for news portals in providing users with a personalized experience by consistently delivering relevant news recommendations. The data pre-processing stage has proven to play a crucial role in improving the quality of the recommendations.*

### Abstrak

Pertumbuhan jumlah konten digital yang meningkat terus setiap harinya menimbulkan tantangan dalam penyaringan dan penyajian informasi yang relevan kepada pengguna internet. Diperlukan adanya sistem rekomendasi yang mampu membantu pengguna dalam menyaring informasi secara relevan. Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi artikel berita berbasis *content-based filtering* dengan algoritma *cosine similarity*. Dataset penelitian ini terdiri dari 150 artikel berita yang diambil melalui platform Kaggle. Tahap *pre-processing* data mencakup normalisasi data, pembersihan data, penghapusan entri duplikat, penghapusan *stop words*, dan tokenisasi. Pengujian dilakukan dengan mengambil 15 sampel artikel, masing-masing menghasilkan lima rekomendasi berdasarkan skor kemiripan tertinggi. Hasil penelitian ini bermanfaat untuk portal berita dalam memberikan pengalaman kepada penggunanya agar selalu mendapatkan rekomendasi berita yang relevan. Tahap *pre-processing* data terbukti berperan penting dalam meningkatkan kualitas rekomendasi.

### Article history

Received 03 Jan 2025

Revised 28 Mar 2025

Accepted 13 Apr 2025

Available online 30 May 2025

### Keywords

Recommendation system,

Content-based filtering,

Cosine similarity,

Pre-processing,

News article

### Riwayat

Diterima 03 Jan 2025

Revisi 28 Mar 2025

Disetujui 13 Apr 2025

Terbit online 30 Mei 2025

### Kata Kunci

Sistem rekomendasi,

Content-based filtering,

Cosine similarity,

Pre-processing,

Artikel berita

## PENDAHULUAN

Menurut (Sutarsih Tri et al., 2023) yang diterbitkan oleh Badan Pusat Statistik (BPS) pada Agustus 2024, sekitar 76,08 persen penduduk Indonesia usia 5 tahun ke atas mengakses internet dengan tujuan mendapatkan informasi atau berita. Angka ini menunjukkan bahwa konsumsi berita dan informasi merupakan salah satu aktivitas utama dalam penggunaan internet di Indonesia. Namun, seiring dengan tingginya aktivitas tersebut, muncul tantangan dalam penyaringan dan penyajian informasi yang relevan kepada pengguna, terutama di tengah

lonjakan jumlah konten digital yang terus meningkat setiap harinya.

Sebagai solusi atas tantangan ini, pembuatan sistem rekomendasi dapat membantu pengguna untuk menyaring dan menemukan informasi yang relevan sesuai dengan kebutuhan secara efisien. Sistem rekomendasi adalah suatu metode yang digunakan untuk memberikan saran dengan memprediksi nilai suatu item bagi pengguna, lalu menampilkan item dengan nilai prediksi tertinggi (Sari et al., 2024). Salah satu pendekatan untuk membangun sistem rekomendasi adalah pendekatan *Content-*

*based* filtering. Pada penelitian yang pernah dilakukan sebelumnya, sistem rekomendasi artikel berita yang telah dikembangkan menggunakan algoritma TF-IDF Vector Similarity mendapatkan hasil bahwa sistem dapat memberikan daftar rekomendasi dengan baik dan cukup relevan dengan apa yang dibutuhkan pengguna (Huda et al., 2022). Penelitian lain, sistem rekomendasi artikel berita yang dikembangkan menggunakan algoritma K-Nearest Neighbor (Dharma et al., 2021).

Sebagian besar penelitian lebih berfokus pada pengujian algoritma menggunakan dataset simulasi dan peningkatan akurasi rekomendasi. Namun, validasi menggunakan data nyata belum banyak dilakukan padahal hal ini penting untuk memastikan sistem dapat bekerja secara efektif dalam kondisi dunia nyata. Selain itu, kuantitas dataset yang terbatas dapat mempengaruhi performa dan cakupan sistem rekomendasi. Di sisi lain, aspek *pre-processing* data yang belum optimal dapat berpotensi mempengaruhi hasil akhir rekomendasi. Berbagai kondisi ini menunjukkan adanya kebutuhan untuk pengembangan sistem yang tidak hanya akurat, tetapi juga didukung oleh dataset yang lebih besar dan lebih relevan, serta *pre-processing* data yang lebih baik.

Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi artikel atau berita menggunakan pendekatan *content-based filtering* dengan memanfaatkan algoritma *cosine similarity*. Algoritma ini dipilih karena kemampuannya dalam mengukur tingkat kesamaan antar artikel berita, berdasarkan kata kunci pada judul atau konten. Diharapkan, penelitian ini dapat memberikan solusi yang efektif dalam membantu pengguna menemukan informasi yang sesuai di tengah banyaknya konten digital yang tersedia.

## METODE PENELITIAN

Metode penelitian ini dirancang untuk mengembangkan sistem rekomendasi artikel berita yang mampu memberikan rekomendasi relevan dan akurat berdasarkan isi konten berita. Dalam penelitian ini, pendekatan *content-based filtering* dipadukan dengan algoritma *cosine similarity* untuk mengukur tingkat kesamaan antar artikel secara efektif. Tahapan penelitian ini mencakup pengambilan dataset dari sumber yang relevan, *pre-*

*processing* data untuk memastikan kualitas dan konsistensinya, *content-based filtering*, perhitungan *cosine similarity*, serta pengujian dan analisis hasil. Setiap langkah dirancang secara sistematis untuk memastikan hasil penelitian dapat divalidasi dalam pengembangan sistem rekomendasi.

### 1. Pengambilan Dataset

Peneliti melakukan pengambilan dataset yang relevan dari platform Kaggle. Dataset ini mencakup judul, isi konten artikel berita, dan metadata lain yang diperlukan.

### 2. Pre-processing Data

Pada tahap *pre-processing* data, peneliti melakukan pembersihan data mentah agar siap untuk digunakan. *Pre-processing* adalah langkah untuk mentransformasi data yang awalnya tidak terstruktur menjadi data yang terstruktur, sehingga sesuai dengan kebutuhan untuk tahap pengolahan berikutnya (Pratiwi & Qoiriah, 2022). Tujuan dari *pre-processing* data adalah untuk menghilangkan variabilitas atau efek yang tidak diinginkan, sehingga informasi yang relevan dengan properti yang diinginkan dapat digunakan untuk pemodelan yang lebih efisien (Mishra et al., 2020). Langkah *pre-processing data* yang dilakukan adalah sebagai berikut:

- a. Melakukan normalisasi data
- b. Melakukan pembersihan data
- c. Menghapus entri duplikat
- d. *Stop Words Removal*
- e. Tokenisasi

### 3. Content-based Filtering

Peneliti melakukan pencocokan artikel berita berdasarkan kesamaan konten. *Content-based Filtering* dapat menghasilkan rekomendasi dengan menganalisis kesamaan karakteristik atribut item telah dinilai oleh pengguna (Pratiwi & Qoiriah, 2022). *Content-based filtering* menganalisis preferensi berdasarkan perilaku pengguna di masa lalu untuk membangun sebuah model, yang kemudian digunakan untuk mencocokkan karakteristik atribut dari suatu barang akan direkomendasikan (Fajriansyah et al., 2021).

### 4. Perhitungan Cosine Similarity

Pada tahapan ini, peneliti melakukan perhitungan manual tingkat kesamaan antar artikel berita menggunakan perhitungan *cosine*

*similarity*. *Cosine Similarity* merupakan suatu metode yang digunakan untuk menghitung tingkat kesamaan antar dokumen dengan mempertimbangkan ukurannya (Alana & Hartanto, 2024). Dalam sistem rekomendasi dengan pendekatan *content-based filtering*, *Cosine Similarity* diterapkan pada vektor fitur untuk mengidentifikasi item yang memiliki kesamaan (Anugrah, 2021). Semakin tinggi nilai fungsi *similarity*, semakin mirip kedua objek tersebut (Natalia Lindang et al., 2022). Sebaliknya, semakin rendah nilai *similarity*, semakin besar perbedaan antara kedua objek (Natalia Lindang et al., 2022). Perhitungan *cosine similarity* dilakukan menggunakan rumus sebagai berikut:

$$\text{Cosine}(A, B) = \frac{A \cdot B}{|A| \times |B|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

Keterangan:

A = Vektor A

B = Vektor B

A · B = *Dot product* antara vektor A dan B.

|A| = Panjang dari vektor A

|B| = Panjang dari vektor B

## 5. Implementasi dan Pengujian

Pada tahap implementasi dan pengujian, peneliti melakukan implementasi perhitungan rekomendasi menggunakan pengkodean bahasa *python* berdasarkan dataset. Hasil pengujian dianalisis untuk mengetahui efektivitas pendekatan yang digunakan.

## HASIL DAN PEMBAHASAN

### Alur Proses Rekomendasi



Gambar 1. Alur Proses Rekomendasi

Pada Gambar 1. menunjukkan alur proses rekomendasi yang menggambarkan tahapan-tahapan dilakukan mulai dari pemrosesan input data hingga menghasilkan daftar rekomendasi artikel berita berdasarkan kemiripan konten. Adapun alur proses rekomendasi adalah sebagai berikut:

### 1. Input Artikel Berita

Sistem menerima masukan berupa data artikel berita yang telah tersedia dalam bentuk

file dataset. Data ini mencakup informasi seperti judul dan isi artikel.

### 2. Pembersihan dan Persiapan Teks

Artikel berita yang masuk akan diproses untuk menghilangkan elemen-elemen yang tidak penting dan tidak relevan, seperti karakter khusus, angka, serta kata-kata umum (*stop words*). Proses ini juga mencakup normalisasi teks dan pemisahan kata (*tokenisasi*), sehingga teks siap untuk direpresentasikan secara numerik.

### 3. Representasi Teks dalam Bentuk Vektor

Setelah teks dibersihkan, setiap artikel berita diubah ke dalam bentuk vektor berdasarkan kata-kata yang menyusunnya. Representasi ini akan menjadi dasar dalam pengukuran tingkat kemiripan antar artikel berita.

### 4. Penghitungan Skor Kemiripan

Setiap artikel berita dibandingkan satu sama lain menggunakan rumus perhitungan *cosine similarity*. Skor kemiripan dihitung untuk menunjukkan sejauh mana dua artikel berita memiliki kemiripan konten.

### 5. Pemilihan Artikel Acuan dan Rekomendasi

Sistem kemudian mengambil artikel berita yang dijadikan sebagai acuan (*target*) dan dari hasil perhitungan skor kemiripan, sistem akan menghasilkan daftar rekomendasi yang terdiri dari lima artikel berita dengan nilai skor kemiripan tertinggi.

### 6. Output Rekomendasi

Berdasarkan skor kemiripan yang telah dihitung, sistem menyusun peringkat artikel berita lain dari yang paling mirip hingga yang paling rendah. Lima artikel berita dengan skor tertinggi dipilih dan ditampilkan sebagai rekomendasi utama.

### 1. Pengumpulan Dataset

Pada penelitian ini menggunakan dataset yang relevan dan diperoleh dari <https://www.kaggle.com/datasets/rizkia14/berita-indo/data>. Dataset ini cukup besar dengan jumlah data sebanyak 117.352 data, namun pada penelitian ini akan menggunakan sampel 150 data saja. Dataset yang sudah diperoleh masih berupa data mentah terdiri



### 3. Perhitungan Cosine Similarity

Contoh perhitungan manual yang dilakukan adalah membandingkan 3 sampel data content yang diambil dari dataset. Langkah pertama dalam perhitungan *cosine similarity* adalah menentukan frekuensi kata-kata yang sama di antara konten. Berikut merupakan nilai representasi vektor dari kata-kata yang muncul:

Tabel. 1 Representasi Vektor

Kata	Content 1	Content 2	Content 3
liga	2	0	0
pemain	3	0	2
pandemi	2	0	1
klub	1	0	2
gol	0	3	0
Kemenangan	0	1	0

Setiap konten kemudian direpresentasikan dalam bentuk vektor sebagai berikut:

- Content 1: (2, 3, 2, 1, 0, 0)
- Content 2: (0, 0, 0, 0, 3, 1)
- Content 3: (0, 2, 1, 2, 0, 0)

Selanjutnya, dot product antara dua vektor dihitung dengan menjumlahkan hasil perkalian nilai elemen yang bersesuaian dari kedua vektor:

$$\text{Dot Product (Content 1, Content 2):}$$

$$(2 \times 0) + (3 \times 0) + (2 \times 0) + (1 \times 0) + (0 \times 3) + (0 \times 1) = 0$$

$$\text{Dot Product (Content 1, Content 3):}$$

$$(2 \times 0) + (3 \times 2) + (2 \times 1) + (1 \times 2) + (0 \times 0) + (0 \times 0) = 10$$

$$\text{Dot Product (Content 2, Content 3):}$$

$$(0 \times 0) + (0 \times 2) + (0 \times 1) + (0 \times 2) + (3 \times 0) + (1 \times 0) = 0$$

Panjang vektor dihitung menggunakan rumus:

$$\|A\| = \sqrt{\sum_{i=1}^n (A_i)^2}$$

$$\text{Panjang Vektor Content 1:}$$

$$\|A\| = \sqrt{(2)^2 + (3)^2 + (2)^2 + (1)^2 + (0)^2 + (0)^2} = \sqrt{18}$$

$$\text{Panjang Vektor Content 2:}$$

$$\|B\| = \sqrt{(0)^2 + (0)^2 + (0)^2 + (0)^2 + (3)^2 + (1)^2} = \sqrt{10}$$

$$\text{Panjang Vektor Content 3:}$$

$$\|C\| = \sqrt{(0)^2 + (2)^2 + (1)^2 + (2)^2 + (0)^2 + (0)^2} = \sqrt{9}$$

Hasil *cosine similarity* diperoleh dengan membagi dot product dengan hasil perkalian panjang vektor.

Cosine Similarity (Content 1, Content 2):

$$\text{Cosine} = \frac{0}{\sqrt{18} \times \sqrt{10}} = 0$$

Cosine Similarity (Content 1, Content 3):

$$\text{Cosine} = \frac{10}{\sqrt{18} \times \sqrt{9}} = \frac{10}{\sqrt{162}} = 0.785$$

Cosine Similarity (Content 2, Content 3):

$$\text{Cosine} = \frac{0}{\sqrt{10} \times \sqrt{9}} = 0$$

Berdasarkan hasil perhitungan manual *cosine similarity* didapatkan bahwa:

a. Content 1 dan Content 2 memiliki nilai *cosine similarity* sebesar 0, menunjukkan tidak ada kesamaan antara kedua konten.

b. Content 1 dan Content 3 memiliki nilai *cosine similarity* sebesar 0.785, yang menunjukkan adanya kesamaan yang cukup signifikan. Hal ini disebabkan oleh kemunculan kata-kata seperti pemain, pandemi, dan klub pada kedua konten.

c. Content 2 dan Content 3 juga memiliki nilai *cosine similarity* sebesar 0, yang menunjukkan kedua konten tidak memiliki kesamaan.

Hasil ini memperlihatkan bahwa *cosine similarity* dapat digunakan untuk mengukur tingkat kemiripan antar konten secara numerik dan memberikan hasil yang dapat diinterpretasikan secara jelas.

### 4. Implementasi Rekomendasi

Tahap terakhir yaitu peneliti melakukan implementasi rekomendasi dengan pendekatan *cosine similarity* menggunakan pengkodean bahasa pemrograman *python*. Pada tahap ini, algoritma *cosine similarity* diterapkan untuk menghitung tingkat kemiripan antar konten berdasarkan dataset yang telah dilakukan *pre-processing* sebelumnya. Dataset tersebut yang terdiri dari 150 data, dimana sebanyak 15 data dipilih secara acak sebagai sampel untuk dilakukan pengujian. Pengujian dilakukan sebanyak empat kali dengan tujuan untuk menguji konsistensi sistem dalam memberikan rekomendasi. Setiap pengujian menghasilkan lima rekomendasi berdasarkan perhitungan skor kemiripan tertinggi. Rekomendasi dihitung dengan membandingkan data sampel

dengan seluruh dataset. Skor kemiripan kemudian diurutkan secara menurun, sehingga lima data dengan nilai skor tertinggi dipilih sebagai rekomendasi untuk data sampel. Hasil pengujian ditunjukkan sebagai berikut.

Content Index	Rekomendasi 1	Rekomendasi 2	Rekomendasi 3	Rekomendasi 4	Rekomendasi 5
0	25 Index 125 (Skor: 0.2587)	Index 102 (Skor: 0.2457)	Index 16 (Skor: 0.1583)	Index 89 (Skor: 0.1552)	Index 22 (Skor: 0.1479)
1	118 Index 66 (Skor: 0.2326)	Index 40 (Skor: 0.2419)	Index 91 (Skor: 0.1586)	Index 133 (Skor: 0.1523)	Index 19 (Skor: 0.1274)
2	107 Index 112 (Skor: 0.2072)	Index 114 (Skor: 0.2396)	Index 134 (Skor: 0.2297)	Index 145 (Skor: 0.1796)	Index 44 (Skor: 0.0924)
3	61 Index 8 (Skor: 0.2096)	Index 41 (Skor: 0.1619)	Index 23 (Skor: 0.1618)	Index 17 (Skor: 0.1578)	Index 123 (Skor: 0.1398)
4	39 Index 77 (Skor: 0.6866)	Index 62 (Skor: 0.0658)	Index 102 (Skor: 0.0588)	Index 28 (Skor: 0.0571)	Index 33 (Skor: 0.0363)
5	23 Index 43 (Skor: 0.2412)	Index 139 (Skor: 0.2343)	Index 17 (Skor: 0.2298)	Index 127 (Skor: 0.2266)	Index 22 (Skor: 0.2106)
6	17 Index 23 (Skor: 0.2298)	Index 43 (Skor: 0.1864)	Index 19 (Skor: 0.1674)	Index 61 (Skor: 0.1578)	Index 41 (Skor: 0.1491)
7	122 Index 3 (Skor: 0.2024)	Index 70 (Skor: 0.2040)	Index 19 (Skor: 0.1962)	Index 29 (Skor: 0.1909)	Index 133 (Skor: 0.1843)
8	47 Index 26 (Skor: 0.2637)	Index 149 (Skor: 0.2359)	Index 148 (Skor: 0.2097)	Index 73 (Skor: 0.1895)	Index 84 (Skor: 0.1445)
9	129 Index 34 (Skor: 0.2095)	Index 80 (Skor: 0.2024)	Index 133 (Skor: 0.1860)	Index 82 (Skor: 0.1827)	Index 27 (Skor: 0.1668)
10	92 Index 113 (Skor: 0.5922)	Index 131 (Skor: 0.2407)	Index 62 (Skor: 0.2311)	Index 81 (Skor: 0.2158)	Index 94 (Skor: 0.2149)
11	104 Index 13 (Skor: 0.5922)	Index 101 (Skor: 0.3464)	Index 126 (Skor: 0.1092)	Index 3 (Skor: 0.0968)	Index 57 (Skor: 0.0962)
12	111 Index 0 (Skor: 0.0000)	Index 1 (Skor: 0.0000)	Index 2 (Skor: 0.0000)	Index 3 (Skor: 0.0000)	Index 4 (Skor: 0.0000)
13	20 Index 123 (Skor: 0.5025)	Index 95 (Skor: 0.1863)	Index 61 (Skor: 0.1224)	Index 61 (Skor: 0.1216)	Index 22 (Skor: 0.1122)
14	35 Index 0 (Skor: 0.3394)	Index 63 (Skor: 0.2977)	Index 128 (Skor: 0.2898)	Index 116 (Skor: 0.2794)	Index 120 (Skor: 0.2395)

Gambar 7. Hasil Pengujian Rekomendasi 1

Pada Gambar 7. menunjukkan hasil implementasi sistem rekomendasi untuk 15 data sampling pada pengujian pertama. Setiap baris tabel merepresentasikan satu data sampel yang diberi nama kolom content. Sistem menghasilkan lima rekomendasi berdasarkan skor kemiripan tertinggi. Sistem rekomendasi ini hanya mencari kemiripan content berita dan tidak memeriksa data tag berita. Rekomendasi diurutkan berdasarkan skor kemiripan dari tinggi ke rendah.

Content Index	Rekomendasi 1	Rekomendasi 2	Rekomendasi 3	Rekomendasi 4	Rekomendasi 5
0	119 Index 4 (Skor: 0.3120)	Index 14 (Skor: 0.1417)	Index 115 (Skor: 0.1253)	Index 146 (Skor: 0.1216)	Index 149 (Skor: 0.1199)
1	70 Index 58 (Skor: 0.3778)	Index 83 (Skor: 0.3648)	Index 29 (Skor: 0.3307)	Index 3 (Skor: 0.2786)	Index 16 (Skor: 0.2516)
2	98 Index 29 (Skor: 0.2715)	Index 3 (Skor: 0.2536)	Index 100 (Skor: 0.2441)	Index 83 (Skor: 0.2036)	Index 70 (Skor: 0.1889)
3	94 Index 93 (Skor: 0.8854)	Index 142 (Skor: 0.4725)	Index 43 (Skor: 0.2481)	Index 80 (Skor: 0.2165)	Index 22 (Skor: 0.1928)
4	52 Index 125 (Skor: 0.7422)	Index 102 (Skor: 0.6782)	Index 109 (Skor: 0.6721)	Index 69 (Skor: 0.2975)	Index 16 (Skor: 0.2416)
5	28 Index 125 (Skor: 0.2967)	Index 102 (Skor: 0.2457)	Index 16 (Skor: 0.1583)	Index 89 (Skor: 0.1552)	Index 22 (Skor: 0.1479)
6	10 Index 0 (Skor: 0.0000)	Index 1 (Skor: 0.0000)	Index 2 (Skor: 0.0000)	Index 3 (Skor: 0.0000)	Index 4 (Skor: 0.0000)
7	80 Index 132 (Skor: 0.2766)	Index 94 (Skor: 0.2165)	Index 62 (Skor: 0.2144)	Index 129 (Skor: 0.2024)	Index 4 (Skor: 0.1540)
8	130 Index 34 (Skor: 0.1227)	Index 6 (Skor: 0.1114)	Index 62 (Skor: 0.1100)	Index 30 (Skor: 0.0859)	Index 127 (Skor: 0.0848)
9	35 Index 0 (Skor: 0.3394)	Index 63 (Skor: 0.2977)	Index 128 (Skor: 0.2898)	Index 116 (Skor: 0.2794)	Index 120 (Skor: 0.2395)
10	100 Index 3 (Skor: 0.3316)	Index 29 (Skor: 0.3293)	Index 98 (Skor: 0.2641)	Index 133 (Skor: 0.2434)	Index 70 (Skor: 0.2310)
11	103 Index 19 (Skor: 0.2036)	Index 41 (Skor: 0.1808)	Index 90 (Skor: 0.1785)	Index 29 (Skor: 0.1673)	Index 45 (Skor: 0.1651)
12	40 Index 66 (Skor: 0.4966)	Index 69 (Skor: 0.3332)	Index 14 (Skor: 0.3314)	Index 125 (Skor: 0.2726)	Index 91 (Skor: 0.2492)
13	101 Index 13 (Skor: 0.3783)	Index 104 (Skor: 0.3464)	Index 58 (Skor: 0.1888)	Index 53 (Skor: 0.1643)	Index 83 (Skor: 0.1455)
14	146 Index 4 (Skor: 0.2678)	Index 115 (Skor: 0.2212)	Index 63 (Skor: 0.2127)	Index 127 (Skor: 0.1860)	Index 99 (Skor: 0.1752)

Gambar 8. Hasil Pengujian Rekomendasi 2

Pada Gambar 8. menunjukkan hasil implementasi sistem rekomendasi untuk 15 data sampling pada pengujian kedua. Dapat dilihat bahwa terdapat data dengan indeks yang sama pada hasil pengujian pertama yaitu content dengan indeks 28 dan content dengan indeks 35. Hasil rekomendasi untuk indeks 28 pada pengujian kedua sama dengan pada pengujian pertama yaitu rekomendasi 1 dengan skor 0.2587, rekomendasi 2 dengan skor 0.2457, rekomendasi 3 dengan skor 0.1583, rekomendasi 4 dengan skor 0.1552, dan rekomendasi 5 dengan skor 0,1479. Kemudian untuk hasil rekomendasi untuk indeks 35 pada pengujian kedua juga sama dengan pengujian pertama yaitu rekomendasi

1 dengan skor 0.3394, rekomendasi 2 dengan skor 0.2988, rekomendasi 3 dengan skor 0.2898, rekomendasi 4 dengan skor 0.2794, dan rekomendasi 5 dengan skor 0.2395.

Content Index	Rekomendasi 1	Rekomendasi 2	Rekomendasi 3	Rekomendasi 4	Rekomendasi 5
0	57 Index 95 (Skor: 0.1076)	Index 26 (Skor: 0.0974)	Index 63 (Skor: 0.0964)	Index 53 (Skor: 0.0798)	Index 6 (Skor: 0.0686)
1	133 Index 19 (Skor: 0.3095)	Index 90 (Skor: 0.2968)	Index 100 (Skor: 0.2434)	Index 29 (Skor: 0.2348)	Index 32 (Skor: 0.2183)
2	78 Index 62 (Skor: 0.1852)	Index 131 (Skor: 0.1798)	Index 141 (Skor: 0.1231)	Index 92 (Skor: 0.1203)	Index 105 (Skor: 0.1087)
3	90 Index 32 (Skor: 0.4557)	Index 19 (Skor: 0.3228)	Index 133 (Skor: 0.2968)	Index 29 (Skor: 0.2348)	Index 12 (Skor: 0.1994)
4	103 Index 19 (Skor: 0.2036)	Index 41 (Skor: 0.1808)	Index 90 (Skor: 0.1785)	Index 29 (Skor: 0.1673)	Index 45 (Skor: 0.1651)
5	13 Index 104 (Skor: 0.5922)	Index 101 (Skor: 0.3733)	Index 97 (Skor: 0.1800)	Index 91 (Skor: 0.1412)	Index 81 (Skor: 0.1390)
6	134 Index 112 (Skor: 0.4024)	Index 114 (Skor: 0.3766)	Index 145 (Skor: 0.2939)	Index 107 (Skor: 0.2297)	Index 43 (Skor: 0.1928)
7	22 Index 74 (Skor: 0.6390)	Index 58 (Skor: 0.2729)	Index 83 (Skor: 0.2728)	Index 70 (Skor: 0.1929)	Index 43 (Skor: 0.1928)
8	101 Index 13 (Skor: 0.3733)	Index 104 (Skor: 0.3464)	Index 58 (Skor: 0.1888)	Index 53 (Skor: 0.1643)	Index 83 (Skor: 0.1455)
9	44 Index 114 (Skor: 0.2244)	Index 145 (Skor: 0.1796)	Index 8 (Skor: 0.0792)	Index 136 (Skor: 0.1551)	Index 144 (Skor: 0.1506)
10	84 Index 47 (Skor: 0.1445)	Index 26 (Skor: 0.0829)	Index 127 (Skor: 0.0792)	Index 79 (Skor: 0.0783)	Index 52 (Skor: 0.0548)
11	65 Index 135 (Skor: 0.1928)	Index 53 (Skor: 0.1226)	Index 9 (Skor: 0.1225)	Index 126 (Skor: 0.1215)	Index 118 (Skor: 0.1206)
12	120 Index 63 (Skor: 0.5855)	Index 128 (Skor: 0.2270)	Index 7 (Skor: 0.2962)	Index 9 (Skor: 0.2559)	Index 71 (Skor: 0.2442)
13	77 Index 83 (Skor: 0.2451)	Index 82 (Skor: 0.2765)	Index 53 (Skor: 0.2073)	Index 5 (Skor: 0.1863)	Index 70 (Skor: 0.1752)
14	2 Index 140 (Skor: 0.2290)	Index 63 (Skor: 0.1497)	Index 25 (Skor: 0.1417)	Index 8 (Skor: 0.1392)	Index 128 (Skor: 0.1345)

Gambar 9. Hasil Pengujian Rekomendasi 3

Pada Gambar 9. menunjukkan hasil implementasi sistem rekomendasi untuk 15 data sampling pada pengujian ketiga. Dapat dilihat bahwa terdapat data dengan indeks yang sama pada hasil pengujian ketiga yaitu content dengan indeks 101 dan content dengan indeks 103. Hasil rekomendasi untuk indeks 101 pada pengujian ketiga sama dengan pada pengujian kedua yaitu rekomendasi 1 dengan skor 0.3733, rekomendasi 2 dengan skor 0.3464, rekomendasi 3 dengan skor 0.1888, rekomendasi 4 dengan skor 0.1643, dan rekomendasi 5 dengan skor 0,1465. Kemudian untuk hasil rekomendasi untuk indeks 103 pada pengujian ketiga juga sama dengan pengujian kedua yaitu rekomendasi 1 dengan skor 0.2036, rekomendasi 2 dengan skor 0.1808, rekomendasi 3 dengan skor 0.1785, rekomendasi 4 dengan skor 0.1673, dan rekomendasi 5 dengan skor 0.1661.

Content Index	Rekomendasi 1	Rekomendasi 2	Rekomendasi 3	Rekomendasi 4	Rekomendasi 5
0	116 Index 0 (Skor: 0.4567)	Index 35 (Skor: 0.2794)	Index 7 (Skor: 0.2555)	Index 138 (Skor: 0.2099)	Index 9 (Skor: 0.1910)
1	145 Index 114 (Skor: 0.3766)	Index 112 (Skor: 0.2235)	Index 134 (Skor: 0.2099)	Index 107 (Skor: 0.1796)	Index 44 (Skor: 0.1726)
2	49 Index 99 (Skor: 0.1099)	Index 0 (Skor: 0.0972)	Index 7 (Skor: 0.0915)	Index 128 (Skor: 0.0870)	Index 25 (Skor: 0.0796)
3	142 Index 94 (Skor: 0.4725)	Index 93 (Skor: 0.4381)	Index 43 (Skor: 0.2341)	Index 22 (Skor: 0.1874)	Index 127 (Skor: 0.1595)
4	21 Index 51 (Skor: 0.1910)	Index 132 (Skor: 0.1318)	Index 64 (Skor: 0.1276)	Index 23 (Skor: 0.1232)	Index 117 (Skor: 0.1192)
5	43 Index 139 (Skor: 0.3942)	Index 127 (Skor: 0.2592)	Index 94 (Skor: 0.2481)	Index 123 (Skor: 0.2412)	Index 93 (Skor: 0.2359)
6	146 Index 4 (Skor: 0.2678)	Index 115 (Skor: 0.2212)	Index 63 (Skor: 0.2127)	Index 127 (Skor: 0.1860)	Index 59 (Skor: 0.1752)
7	71 Index 120 (Skor: 0.2442)	Index 1 (Skor: 0.2015)	Index 63 (Skor: 0.2007)	Index 128 (Skor: 0.1931)	Index 3 (Skor: 0.1806)
8	143 Index 13 (Skor: 0.1341)	Index 76 (Skor: 0.1286)	Index 3 (Skor: 0.1249)	Index 101 (Skor: 0.1247)	Index 135 (Skor: 0.1213)
9	61 Index 8 (Skor: 0.2090)	Index 41 (Skor: 0.1619)	Index 23 (Skor: 0.1616)	Index 121 (Skor: 0.1578)	Index 123 (Skor: 0.1396)
10	22 Index 74 (Skor: 0.6390)	Index 58 (Skor: 0.2729)	Index 83 (Skor: 0.2728)	Index 70 (Skor: 0.1929)	Index 43 (Skor: 0.1928)
11	88 Index 116 (Skor: 0.0892)	Index 50 (Skor: 0.0872)	Index 121 (Skor: 0.0824)	Index 92 (Skor: 0.0816)	Index 78 (Skor: 0.0797)
12	128 Index 63 (Skor: 0.3603)	Index 120 (Skor: 0.3270)	Index 0 (Skor: 0.3091)	Index 35 (Skor: 0.2898)	Index 25 (Skor: 0.2693)
13	99 Index 140 (Skor: 0.4017)	Index 75 (Skor: 0.3436)	Index 0 (Skor: 0.3217)	Index 7 (Skor: 0.2930)	Index 128 (Skor: 0.2789)
14	31 Index 43 (Skor: 0.1389)	Index 127 (Skor: 0.1325)	Index 133 (Skor: 0.1152)	Index 82 (Skor: 0.1146)	Index 19 (Skor: 0.1110)

Gambar 10. Hasil Pengujian Rekomendasi 4

Pada Gambar 10. menunjukkan hasil implementasi sistem rekomendasi untuk 15 data sampling pada pengujian keempat. Dapat dilihat bahwa terdapat data dengan indeks yang sama pada hasil pengujian keempat dengan pengujian kedua yaitu content dengan indeks 146 dan data dengan indeks yang sama pada hasil pengujian keempat dengan pengujian ketiga yaitu content dengan indeks

22. Hasil rekomendasi untuk indeks 146 pada pengujian keempat sama dengan pada pengujian kedua yaitu rekomendasi 1 dengan skor 0.2673, rekomendasi 2 dengan skor 0.2212, rekomendasi 3 dengan skor 0.2127, rekomendasi 4 dengan skor 0.1860, dan rekomendasi 5 dengan skor 0,1752. Selanjutnya, untuk hasil rekomendasi untuk indeks 22 pada pengujian keempat juga sama dengan pengujian ketiga yaitu rekomendasi 1 dengan skor 0.6390, rekomendasi 2 dengan skor 0.2729, rekomendasi 3 dengan skor 0.2728, rekomendasi 4 dengan skor 0.1929, dan rekomendasi 5 dengan skor 0.1928.

## KESIMPULAN

Pengembangan sistem rekomendasi artikel berita dengan pendekatan *content-based filtering* menggunakan algoritma *cosine similarity* dapat memberikan output rekomendasi untuk membantu pengguna dalam menemukan konten yang relevan. Hasil pengujian menunjukkan bahwa sistem dapat menghasilkan daftar rekomendasi yang diurutkan berdasarkan skor kemiripan konten artikel atau berita tertinggi. Dalam empat kali pengujian dengan sampel 15 data tiap pengujian dari 150 dataset, sistem secara konsisten memberikan lima rekomendasi yang sama untuk setiap data uji. Hal ini menunjukkan keakuratan algoritma *cosine similarity* dalam mengukur kesamaan antar konten berita. Selain itu, metode *pre-processing* yang diterapkan bisa meningkatkan kualitas data dan mendukung performa sistem. Namun, penelitian ini juga memiliki beberapa keterbatasan, seperti cakupan dataset yang masih dapat diperluas untuk mencakup lebih banyak jenis artikel dan pengujian dalam skenario yang lebih kompleks. Untuk penelitian selanjutnya, dapat dilakukan validasi sistem menggunakan dataset yang lebih besar dan menyempurnakan lagi *pre-processing* datanya seperti menambahkan *stemming* untuk memastikan konsistensi bentuk kata.

## REFERENSI

Alana, R., & Hartanto, A. (2024). Sistemasi: Jurnal Sistem Informasi Implementasi Algoritma Content Based Filtering dalam Sistem Rekomendasi Komik. *Sistemasi: Jurnal Sistem Informasi*, 13(4), 1344–1355. <http://sistemasi.ftik.unisi.ac.id>

Anugrah, I. G. (2021). Penerapan Metode N-Gram dan Cosine Similarity Dalam Pencarian Pada Repositori Artikel Jurnal Publikasi. *Building of Informatics, Technology and Science (BITS)*, 3(3), 275–284.

<https://doi.org/10.47065/bits.v3i3.1058>

Budiman, A. E., & Widjaja, A. (2020). Analisis Pengaruh Teks Preprocessing Terhadap Deteksi Plagiarisme Pada Dokumen Tugas Akhir. *Jurnal Teknik Informatika Dan Sistem Informasi*, 6(3). <https://doi.org/10.28932/jutisi.v6i3.2892>

Christopher D. Manning, Prabhakar Raghavan, & Hinrich Schütze. (2009). *An Introduction to Information Retrieval*.

Dharma, A. S., Basadena, R. B., Hutasoit, A., & Pangaribuan, R. R. (2021). *Arie Satia Dharma: Sistem Rekomendasi Menggunakan Item-based Collaborative Filtering pada Konten Artikel Berita*.

Fajriansyah, M., Adikara, P. P., & Widodo, A. W. (2021). Sistem Rekomendasi Film Menggunakan Content Based Filtering. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 5(6), 2188–2199. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/9163>

Gilbert, Syariful Alam, & M. Imam Sulisty. (2023). ANALISIS SENTIMEN BERDASARKAN ULASAN PENGGUNA APLIKASI MYPERTAMINA PADA GOOGLE PLAYSTORE MENGGUNAKAN METODE NAÏVE BAYES. *STORAGE: Jurnal Ilmiah Teknik Dan Ilmu Komputer*, 2(3), 100–108. <https://doi.org/10.55123/storage.v2i3.2333>

Huda, A. A., Fajarudin, R., & Hadinegoro, A. (2022). Sistem Rekomendasi Content-based Filtering Menggunakan TF-IDF Vector Similarity Untuk Rekomendasi Artikel Berita. *Building of Informatics, Technology and Science (BITS)*, 4(3). <https://doi.org/10.47065/bits.v4i3.2511>

Jain, S., Shukla, S., & Wadhvani, R. (2018). Dynamic selection of normalization techniques using data complexity measures. *Expert Systems with*

- Applications*, 106, 252–262.  
<https://doi.org/10.1016/j.eswa.2018.04.008>
- Mishra, P., Biancolillo, A., Roger, J. M., Marini, F., & Rutledge, D. N. (2020). New data preprocessing trends based on ensemble of multiple preprocessing techniques. *TrAC Trends in Analytical Chemistry*, 132, 116045. <https://doi.org/10.1016/j.trac.2020.116045>
- Natalia Lindang, D., Yulia Muniar, A., Halid, A., Muhajirin, & Amiruddin, A. (2022). Sistem Penentuan Kemiripan Antar Skripsi Menggunakan Metode Cosine Similarity Pada Perpustakaan. *Prosiding Seminar Nasional Teknik Elektro Dan Informatika (SNTEI)*, 8(1), 321–324.
- Pratiwi, D. A., & Qoiriah, A. (2022). Sistem Rekomendasi Wedding Organizer Menggunakan Metode Content-Based Filtering Dengan Algoritma Random Forest Regression. *Journal of Informatics and Computer Science*, 03, 231–239.
- Rohman, A. N., Fauzy, M. N., & Sa'di, A. (2024). Sistem Rekomendasi Buku Menggunakan Algoritma Rabin-Karp. *Jurnal Eksplora Informatika*, 12(1), 86–94. <https://doi.org/10.30864/eksplora.v12i1.1074>
- Sari, K. R., Suharso, W., & Azhar, Y. (2024). Pembuatan Sistem Rekomendasi Film dengan Menggunakan Metode Item Based Collaborative Filtering pada Apache Mahout. *Jurnal Repositor*, 2(6). <https://doi.org/10.22219/repositor.v2i6.30715>
- Sutarsih Tri, Sari Eka, Syakilah Adriyani, & Maharani Karmila. (2023). *Statistik Telekomunikasi Indonesia 2023* (Vol. 12).