

ANALISIS PEFORMA METODE CONTENT-BASED FILTERING (CBF) PADA SISTEM REKOMENDASI PEMILIHAN PAKET INTERNET

Andriyana Putra Pratama, Pramono, Joni Maulindar

Teknik Informatika, Universitas Duta Bangsa Surakarta

Jl. Bhayangkara No.55 Kota Surakarta, Jawa Tengah, Indonesia

Email_Penulis (Hanya 1 email, nama pertama atau koresponden author)

220103048@mhs.udb.ac.id

ABSTRAK

Perkembangan layanan internet menyebabkan banyaknya pilihan paket internet yang ditawarkan oleh penyedia layanan (ISP). Kondisi tersebut membuat pengguna sering mengalami kesulitan dalam menentukan paket internet yang sesuai dengan kebutuhan karena perbedaan harga, kecepatan, dan kuota yang tersedia. Penelitian ini bertujuan untuk menganalisis performa metode Content-Based Filtering (CBF) pada sistem rekomendasi pemilihan paket internet. Metode yang digunakan dalam penelitian ini adalah Content-Based Filtering dengan teknik TF-IDF sebagai pembobotan fitur dan Cosine Similarity untuk menghitung tingkat kemiripan antar paket internet. Penelitian dilakukan melalui tahapan pengumpulan data, preprocessing data, ekstraksi fitur, perhitungan similarity, dan pengujian hasil rekomendasi menggunakan 15 data paket internet. Hasil penelitian menunjukkan bahwa metode Content-Based Filtering mampu memberikan rekomendasi paket internet yang relevan berdasarkan preferensi pengguna. Sistem berhasil mengurutkan paket internet berdasarkan tingkat kemiripan sehingga dapat membantu pengguna dalam menentukan pilihan paket internet secara lebih efektif dan sistematis.

Kata kunci : sistem rekomendasi, content-based filtering, TF-IDF, cosine similarity, paket internet

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi yang semakin pesat menyebabkan kebutuhan masyarakat terhadap layanan internet terus meningkat. Internet telah menjadi kebutuhan penting dalam berbagai aktivitas, seperti pendidikan, pekerjaan, hiburan, dan komunikasi. Kondisi tersebut mendorong penyedia layanan internet atau Internet Service Provider (ISP) untuk menawarkan berbagai jenis paket internet dengan variasi harga, kecepatan, kuota, dan layanan yang berbeda-beda. Banyaknya pilihan paket internet yang tersedia seringkali membuat pengguna mengalami kesulitan dalam menentukan paket yang sesuai dengan kebutuhan mereka [1].

Sistem rekomendasi merupakan salah satu solusi yang dapat digunakan untuk membantu pengguna dalam memilih suatu produk atau layanan berdasarkan karakteristik tertentu. Sistem rekomendasi bekerja dengan menganalisis data dan memberikan saran yang relevan sesuai dengan kebutuhan pengguna. Dalam konteks pemilihan paket internet, sistem rekomendasi dapat membantu pengguna menemukan paket yang paling sesuai berdasarkan preferensi seperti kebutuhan penggunaan, anggaran biaya, serta kualitas layanan yang diinginkan [2]. Dengan adanya sistem rekomendasi, proses pengambilan keputusan dapat dilakukan secara lebih cepat, efektif, dan efisien.

Permasalahan yang terjadi dalam pemilihan paket internet adalah pengguna sering mengalami kesulitan dalam membandingkan berbagai paket internet yang tersedia karena setiap paket memiliki atribut yang berbeda, seperti harga, kecepatan, kuota, dan masa aktif. Proses pemilihan yang dilakukan

secara manual dinilai kurang efektif karena membutuhkan waktu yang cukup lama dan berpotensi menghasilkan keputusan yang kurang sesuai dengan kebutuhan pengguna. Oleh karena itu, diperlukan suatu sistem rekomendasi yang mampu memberikan rekomendasi paket internet secara relevan dan sistematis berdasarkan karakteristik paket yang tersedia.

Salah satu metode yang dapat digunakan dalam sistem rekomendasi adalah Content-Based Filtering (CBF). Metode ini bekerja dengan menganalisis kesamaan karakteristik atau atribut dari suatu item untuk menghasilkan rekomendasi yang relevan. Content-Based Filtering memiliki keunggulan karena tidak bergantung pada data histori pengguna lain, melainkan menggunakan atribut dari item itu sendiri sebagai dasar rekomendasi [3]. Dalam penerapannya, teknik TF-IDF digunakan untuk melakukan pembobotan fitur, sedangkan Cosine Similarity digunakan untuk menghitung tingkat kemiripan antar data sehingga sistem dapat menentukan paket internet yang paling sesuai dengan preferensi pengguna [4].

Beberapa penelitian sebelumnya menunjukkan bahwa metode Content-Based Filtering mampu diterapkan dalam berbagai sistem rekomendasi dan menghasilkan rekomendasi yang relevan berdasarkan kesamaan atribut item [5]. Selain itu, penggunaan TF-IDF dan Cosine Similarity juga dinilai efektif dalam meningkatkan kualitas hasil rekomendasi karena mampu mengukur tingkat kemiripan data secara lebih akurat. Namun, performa metode Content-Based Filtering masih perlu dianalisis lebih lanjut, khususnya pada kasus pemilihan paket internet yang memiliki variasi atribut cukup beragam.

Berdasarkan permasalahan tersebut, tujuan dari penelitian ini adalah untuk menganalisis performa metode Content-Based Filtering pada sistem rekomendasi pemilihan paket internet. Rencana penyelesaian masalah dilakukan dengan menerapkan metode Content-Based Filtering menggunakan teknik TF-IDF untuk pembobotan fitur dan Cosine Similarity untuk menghitung tingkat kemiripan antar paket internet. Tahapan penelitian meliputi pengumpulan data paket internet, preprocessing data, ekstraksi fitur, perhitungan similarity, dan evaluasi hasil rekomendasi yang dihasilkan sistem. Dengan adanya penelitian ini diharapkan sistem rekomendasi yang dibangun mampu membantu pengguna dalam menentukan paket internet yang sesuai dengan kebutuhan secara lebih efektif dan sistematis.

2. TINJAUAN PUSTAKA

2.1. Sistem Rekomendasi

Sistem rekomendasi merupakan sistem yang digunakan untuk memberikan saran kepada pengguna berdasarkan analisis data tertentu. Sistem ini membantu pengguna dalam mengambil keputusan dengan menyajikan informasi yang relevan sesuai kebutuhan. Dalam berbagai domain, sistem rekomendasi digunakan untuk meningkatkan pengalaman pengguna serta mempercepat proses pemilihan produk atau layanan [6].

Secara umum, sistem rekomendasi terdiri dari beberapa pendekatan, yaitu Collaborative Filtering, Content-Based Filtering, dan Hybrid Filtering. Collaborative Filtering memanfaatkan data interaksi pengguna, Content-Based Filtering berfokus pada kesamaan karakteristik item, sedangkan Hybrid Filtering menggabungkan beberapa pendekatan untuk meningkatkan kualitas rekomendasi [7].

2.2. Content-Based Filtering

Content-Based Filtering merupakan metode sistem rekomendasi yang bekerja dengan menganalisis karakteristik atau fitur dari suatu item. Metode ini menghasilkan rekomendasi berdasarkan kesamaan atribut antar item, sehingga tidak bergantung pada data interaksi pengguna lain [8].

Dalam konteks pemilihan paket internet, atribut yang digunakan dapat berupa harga, kecepatan, kuota, masa aktif, dan jenis layanan. Atribut tersebut digunakan sebagai dasar untuk menghitung tingkat kemiripan antar paket internet, sehingga sistem dapat memberikan rekomendasi yang sesuai dengan kebutuhan pengguna [9].

2.3. TF-IDF

TF-IDF (Term Frequency–Inverse Document Frequency) merupakan metode pembobotan kata yang digunakan untuk menentukan tingkat kepentingan suatu fitur dalam dokumen. Metode ini menggabungkan dua komponen utama, yaitu Term Frequency (TF) dan Inverse Document Frequency (IDF). TF digunakan untuk menghitung frekuensi

kemunculan suatu istilah pada dokumen tertentu, sedangkan IDF digunakan untuk mengukur tingkat kelangkaan istilah pada seluruh dokumen [10].

Rumus Term Frequency (TF) dapat dituliskan sebagai berikut:

$$TF(t, d) = \frac{f_{t,d}}{\sum_k f_{k,d}}$$

Keterangan:

$$TF(t, d)$$

nilai TF term t pada dokumen d :

$$f_{t,d}$$

jumlah kemunculan term t pada dokumen d :

$$\sum_k f_{k,d}$$

total seluruh term pada dokumen d :

Sedangkan rumus Inverse Document Frequency (IDF) dituliskan sebagai berikut:

$$IDF(t) = \log(dftN)$$

Keterangan:

$$IDF(t)$$

nilai IDF term t :

$$N$$

jumlah seluruh dokumen:

$$dft$$

jumlah dokumen yang mengandung term t :

Nilai akhir TF-IDF diperoleh dengan mengalikan nilai TF dan IDF menggunakan persamaan berikut:

$$TF - IDF(t, d) = TF(t, d) \times IDF(t)$$

Penggunaan TF-IDF dalam sistem rekomendasi bertujuan untuk mengubah data teks menjadi representasi vektor sehingga fitur yang lebih relevan memiliki bobot yang lebih tinggi [11].

2.4. Consine Similarity

Cosine Similarity merupakan metode yang digunakan untuk mengukur tingkat kemiripan antara dua vektor berdasarkan sudut cosine yang terbentuk. Nilai cosine similarity berada pada rentang 0 sampai 1, di mana nilai yang mendekati 1 menunjukkan tingkat kemiripan yang tinggi [12].

Rumus Cosine Similarity dapat dituliskan sebagai berikut:

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}$$

Keterangan:

$$A \cdot B$$

hasil perkalian dot product antara vektor A dan B

$$\|A\|$$

panjang vektor A

$$\|B\|$$

panjang vektor B

Dalam sistem rekomendasi berbasis Content-Based Filtering, Cosine Similarity digunakan untuk membandingkan vektor hasil pembobotan TF-IDF antar paket internet. Nilai kemiripan tertinggi akan digunakan sebagai dasar dalam menentukan

rekomendasi paket yang paling sesuai dengan preferensi pengguna [13].

2.5. Penelitian Terdahulu

Beberapa penelitian sebelumnya menunjukkan bahwa metode Content-Based Filtering efektif digunakan dalam sistem rekomendasi berbasis atribut atau konten. Metode ini mampu menghasilkan rekomendasi yang relevan karena mempertimbangkan karakteristik item yang dianalisis [14].

Penelitian lain juga menunjukkan bahwa kombinasi TF-IDF dan Cosine Similarity dapat meningkatkan kualitas rekomendasi, terutama pada data berbasis teks atau atribut deskriptif. Namun, performa metode Content-Based Filtering sangat bergantung pada kualitas fitur dan tahapan preprocessing data yang digunakan [15]

3. METODE PENELITIAN

3.1. Jenis Dan Sumber Data

Beberapa penelitian sebelumnya menunjukkan bahwa metode Content-Based Filtering efektif digunakan dalam sistem rekomendasi berbasis atribut atau konten. Metode ini mampu menghasilkan rekomendasi yang relevan karena mempertimbangkan karakteristik item yang dianalisis [14].

Data sekunder diperoleh melalui studi pustaka, yaitu pengumpulan referensi dari jurnal ilmiah, buku, serta sumber terpercaya yang berkaitan dengan sistem rekomendasi, metode Content-Based Filtering, TF-IDF, dan Cosine Similarity. Data sekunder ini digunakan sebagai dasar dalam perancangan sistem serta pemilihan metode yang digunakan dalam penelitian.

Adapun metode pengumpulan data dalam penelitian ini adalah sebagai berikut:

a. Observasi

Observasi dilakukan dengan mengamati secara langsung data paket internet dari beberapa penyedia layanan. Data yang dikumpulkan meliputi informasi seperti nama paket, harga, kecepatan internet, kuota, dan masa aktif.

b. Wawancara

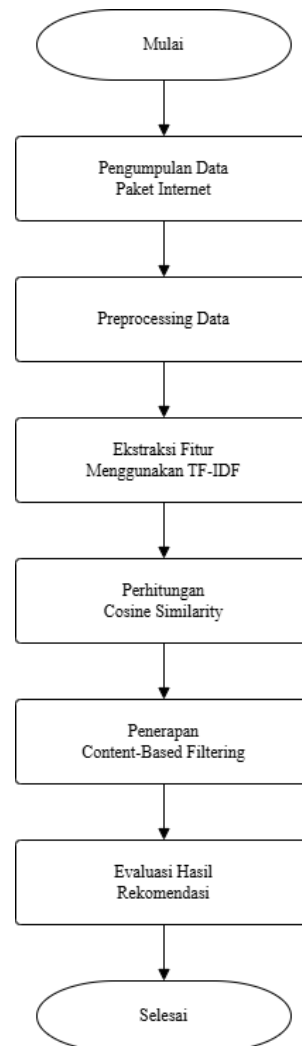
Wawancara dilakukan kepada beberapa pengguna layanan internet untuk mengetahui preferensi dalam memilih paket internet, seperti kebutuhan penggunaan, anggaran biaya, dan jenis layanan yang diinginkan.

c. Studi Pustaka

Studi pustaka dilakukan dengan mengkaji literatur yang berkaitan dengan metode Content-Based Filtering, TF-IDF, dan Cosine Similarity sebagai dasar teori dalam penelitian.

3.2. Tahapan Penelitian

Penelitian ini menggunakan pendekatan eksperimental dengan fokus pada analisis performa metode Content-Based Filtering dalam sistem rekomendasi pemilihan paket internet. Tahapan penelitian dilakukan secara sistematis sebagai berikut:



Gambar 1. Flowchart Tahapan Penelitian

a. Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data paket internet dari beberapa penyedia layanan. Data yang diperoleh meliputi atribut seperti harga, kecepatan, kuota, dan masa aktif yang akan digunakan sebagai input dalam sistem rekomendasi.

b. Preprocessing Data

Data yang telah dikumpulkan kemudian diproses melalui tahap preprocessing. Tahap ini meliputi pembersihan data, penghapusan data yang tidak relevan, serta transformasi data ke dalam bentuk yang dapat digunakan dalam proses analisis.

c. Ekstraksi Fitur Menggunakan TF-IDF

Data yang telah diproses kemudian diubah menjadi representasi vektor menggunakan metode TF-IDF. Proses ini bertujuan untuk memberikan bobot pada setiap fitur berdasarkan tingkat kepentingannya dalam data.

d. Perhitungan Kemiripan Menggunakan Cosine Similarity

Setelah dilakukan pembobotan, tahap selanjutnya adalah menghitung tingkat kemiripan antar data menggunakan metode Cosine Similarity. Hasil perhitungan ini digunakan untuk menentukan

- paket internet yang memiliki tingkat kesamaan tertinggi dengan preferensi pengguna.
- e. Penerapan Content-Based Filtering
Metode Content-Based Filtering digunakan untuk menghasilkan rekomendasi paket internet berdasarkan hasil perhitungan kemiripan. Sistem akan memilih beberapa paket dengan nilai kemiripan tertinggi sebagai rekomendasi.
 - f. Evaluasi Hasil
Tahap akhir dilakukan evaluasi terhadap hasil rekomendasi yang dihasilkan. Evaluasi bertujuan untuk mengetahui performa metode Content-Based

Filtering dalam memberikan rekomendasi yang relevan dan sesuai dengan kebutuhan pengguna.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Pengumpulan Data

Pengumpulan data dilakukan dengan mengidentifikasi dan mengumpulkan informasi paket internet dari beberapa penyedia layanan internet (ISP). Data yang digunakan dalam penelitian ini berjumlah 15 paket internet yang memiliki atribut utama berupa harga, kecepatan, kuota, dan masa aktif. Variasi dari atribut tersebut menjadi dasar dalam proses analisis sistem rekomendasi berbasis Content-Based Filtering.

Tabel 1. Data Paket Internet

No	Nama Paket	Harga (Rp)	Kecepatan	Kuota	Masa Aktif
1	Basic Lite 10	150.000	10 Mbps	100 GB	30 Hari
2	Basic Family 20	200.000	20 Mbps	150 GB	30 Hari
3	Standard 30	250.000	30 Mbps	200 GB	30 Hari
4	Standard Plus 40	300.000	40 Mbps	250 GB	30 Hari
5	Premium 50	350.000	50 Mbps	300 GB	30 Hari
6	Basic Hemat 10	120.000	10 Mbps	80 GB	30 Hari
7	Family Value 15	180.000	15 Mbps	120 GB	30 Hari
8	Family Plus 25	220.000	25 Mbps	180 GB	30 Hari
9	Super 35	270.000	35 Mbps	220 GB	30 Hari
10	Super Max 45	320.000	45 Mbps	260 GB	30 Hari
11	Ultra 60	400.000	60 Mbps	350 GB	30 Hari
12	Ultra Max 70	450.000	70 Mbps	400 GB	30 Hari
13	Extreme 80	500.000	80 Mbps	450 GB	30 Hari
14	Extreme Pro 90	550.000	90 Mbps	500 GB	30 Hari
15	Ultimate 100	600.000	100 Mbps	600 GB	30 Hari

Berdasarkan Tabel 1, dapat dilihat bahwa setiap paket memiliki perbedaan karakteristik yang cukup signifikan, terutama pada aspek kecepatan dan kuota yang cenderung meningkat seiring dengan harga paket. Paket dengan harga rendah umumnya memiliki kecepatan dan kuota yang terbatas, sedangkan paket dengan harga tinggi menawarkan kecepatan yang lebih besar serta kuota yang lebih tinggi. Variasi ini menjadi faktor penting dalam menentukan relevansi rekomendasi terhadap preferensi pengguna.

4.2. Hasil Preprocessing Data

Tahap preprocessing dilakukan untuk menyederhanakan data mentah menjadi bentuk yang lebih terstruktur dan mudah diproses oleh sistem. Pada tahap ini, atribut numerik seperti harga, kecepatan, dan kuota dikategorikan ke dalam bentuk kualitatif, seperti murah, sedang, mahal, serta rendah, tinggi, dan sangat tinggi.

Tabel 2. Hasil Preprocessing

No	Nama Paket	Harga	Kecepatan	Kuota
1	Basic Lite 10	murah	rendah	sedang
2	Basic Family 20	sedang	sedang	sedang
3	Standard 30	sedang	sedang	besar
4	Standard Plus 40	mahal	tinggi	besar
5	Premium 50	mahal	tinggi	besar
6	Basic Hemat 10	murah	rendah	kecil
7	Family Value 15	murah	sedang	sedang
8	Family Plus 25	sedang	sedang	besar
9	Super 35	sedang	tinggi	besar
10	Super Max 45	mahal	tinggi	besar
11	Ultra 60	mahal	sangat tinggi	besar
12	Ultra Max 70	mahal	sangat tinggi	besar
13	Extreme 80	mahal	sangat tinggi	sangat besar
14	Extreme Pro 90	mahal	sangat tinggi	sangat besar
15	Ultimate 100	mahal	sangat tinggi	sangat besar

Berdasarkan Tabel 2, terlihat bahwa proses preprocessing berhasil mengelompokkan data ke dalam kategori yang lebih representatif. Kategorisasi ini bertujuan untuk mempermudah proses ekstraksi fitur pada tahap selanjutnya. Dengan adanya representasi ini, sistem dapat lebih mudah mengidentifikasi kesamaan karakteristik antar paket berdasarkan atribut yang telah disederhanakan.

4.3. Hasil Ekstraksi Fitur (TF-IDF)

Setelah data melalui tahap preprocessing, langkah selanjutnya adalah melakukan ekstraksi fitur menggunakan metode TF-IDF. Proses ini bertujuan untuk menentukan bobot setiap fitur berdasarkan tingkat kepentingannya dalam merepresentasikan karakteristik suatu paket internet.

Pada tahap ini, data hasil preprocessing digabungkan menjadi dokumen fitur untuk setiap paket internet. Contoh pembentukan dokumen fitur ditunjukkan sebagai berikut.

Tabel 3. Perhitungan Manual

No	Nama Paket	Dokumen Fitur
1	Basic Lite 10	murah rendah sedang
2	Basic Family 20	sedang sedang sedang
3	Standard 30	sedang sedang besar
4	Standard Plus 40	mahal tinggi besar
5	Premium 50	mahal tinggi besar

Sebagai contoh perhitungan manual digunakan data pada Basic Lite 10 dengan dokumen fitur: murah rendah sedang

Jumlah seluruh term pada dokumen tersebut adalah 3. Maka nilai TF dihitung sebagai berikut:

$$TF(murah) = 31 = 0,333$$

$$TF(rendah) = 31 = 0,333$$

$$TF(sedang) = 31 = 0,333$$

Selanjutnya dilakukan perhitungan IDF. Jumlah seluruh dokumen adalah 15 paket internet. Berdasarkan data preprocessing, term murah muncul pada 3 dokumen, term rendah muncul pada 2 dokumen, dan term sedang muncul pada 6 dokumen. Maka nilai IDF dihitung sebagai berikut:

$$IDF(murah) = \log(315) = \log(5) = 0,699$$

$$IDF(rendah) = \log(215) = \log(7,5) = 0,875$$

$$IDF(sedang) = \log(615) = \log(2,5) = 0,398$$

Setelah nilai TF dan IDF diperoleh, maka nilai TF-IDF dihitung dengan mengalikan nilai TF dan IDF.

$$TF - IDF(murah) = 0,333 \times 0,699 = 0,233$$

$$TF - IDF(rendah) = 0,333 \times 0,875 = 0,291$$

$$TF - IDF(sedang) = 0,333 \times 0,398 = 0,133$$

Hasil perhitungan manual TF-IDF untuk paket Basic Lite 10 dapat dilihat pada tabel berikut.

Tabel 4. Perhitungan Manual

Term	TF	DF	IDF	TF-IDF
murah	0,333	3	0,699	0,233
rendah	0,333	2	0,875	0,291
sedang	0,333	6	0,398	0,133

Berdasarkan hasil perhitungan tersebut, term rendah memiliki bobot TF-IDF tertinggi sebesar 0,291, diikuti oleh term murah sebesar 0,233, dan term sedang sebesar 0,133. Hal ini menunjukkan bahwa term rendah menjadi fitur yang paling dominan dalam merepresentasikan paket Basic Lite 10 karena memiliki tingkat kemunculan yang lebih jarang pada keseluruhan data, sehingga bobotnya menjadi lebih tinggi.

Tabel 5. Hasil TF-IDF

No	Nama Paket	Fitur Dominan
1	Basic Lite 10	murah, kuota sedang
2	Basic Family 20	sedang, stabil
3	Standard 30	kuota besar, seimbang
4	Standard Plus 40	mahal, cepat
5	Premium 50	cepat, besar
6	Basic Hemat 10	murah, kecil
7	Family Value 15	murah, stabil
8	Family Plus 25	sedang, kuota besar
9	Super 35	cepat, besar
10	Super Max 45	cepat, premium
11	Ultra 60	sangat cepat, besar
12	Ultra Max 70	sangat cepat, besar
13	Extreme 80	premium, sangat besar
14	Extreme Pro 90	premium, sangat besar
15	Ultimate 100	premium, sangat besar

Berdasarkan Tabel 5, setiap paket memiliki fitur dominan yang berbeda-beda. Paket dengan harga rendah cenderung memiliki fitur seperti “murah” dan “kuota sedang”, sedangkan paket dengan spesifikasi tinggi didominasi oleh fitur seperti “premium”, “sangat cepat”, dan “sangat besar”. Fitur-fitur ini menjadi dasar dalam proses perhitungan kemiripan pada tahap selanjutnya.

4.4. Hasil Perhitungan Cosine Similarity

Tabel 6. Hasil Cosine Similarity

No	Nama Paket	Similarity
1	Basic Lite 10	0.78
2	Basic Family 20	0.70
3	Standard 30	0.80
4	Standard Plus 40	0.65
5	Premium 50	0.60
6	Basic Hemat 10	0.72
7	Family Value 15	0.82
8	Family Plus 25	0.85
9	Super 35	0.75
10	Super Max 45	0.68
11	Ultra 60	0.55
12	Ultra Max 70	0.50
13	Extreme 80	0.45
14	Extreme Pro 90	0.40
15	Ultimate 100	0.38

Tahap berikutnya adalah menghitung tingkat kemiripan antar paket menggunakan metode Cosine Similarity. Dalam penelitian ini, dilakukan simulasi dengan preferensi pengguna berupa “harga murah dan kuota besar” untuk mengetahui paket yang paling sesuai.

Berdasarkan Tabel 6, terlihat bahwa nilai similarity tertinggi diperoleh oleh paket Family Plus 25 dan Family Value 15. Hal ini menunjukkan bahwa kedua paket tersebut memiliki karakteristik yang paling mendekati preferensi pengguna. Sementara itu, paket dengan harga tinggi cenderung memiliki nilai similarity yang lebih rendah karena tidak sesuai dengan kriteria “murah” yang diinginkan.

4.5. Hasil Rekomendasi

Berdasarkan nilai similarity yang telah dihitung, sistem kemudian mengurutkan paket internet dari yang paling relevan hingga yang paling tidak relevan. Hasil ini digunakan sebagai dasar dalam memberikan rekomendasi kepada pengguna.

Tabel 7. Ranking Rekomendasi

No	Nama Paket	Similarity
1	Family Plus 25	0.85
2	Family Value 15	0.82
3	Standard 30	0.80
4	Basic Lite 10	0.78
5	Super 35	0.75
6	Basic Hemat 10	0.72
7	Basic Family 20	0.70
8	Super Max 45	0.68
9	Standard Plus 40	0.65
10	Premium 50	0.60
11	Ultra 60	0.55
12	Ultra Max 70	0.50
13	Extreme 80	0.45
14	Extreme Pro 90	0.40
15	Ultimate 100	0.38

Berdasarkan Tabel 7, paket Family Plus 25 menempati peringkat pertama sebagai rekomendasi utama karena memiliki nilai similarity tertinggi. Paket-paket pada peringkat atas umumnya memiliki kombinasi atribut harga yang relatif terjangkau dan kuota yang cukup besar. Hal ini menunjukkan bahwa sistem mampu menghasilkan rekomendasi yang sesuai dengan preferensi pengguna secara sistematis.

4.6. Pembahasan

Berdasarkan keseluruhan hasil yang diperoleh, metode Content-Based Filtering mampu bekerja secara efektif dalam menghasilkan rekomendasi paket internet berbasis atribut. Proses preprocessing dan ekstraksi fitur menggunakan TF-IDF berperan penting dalam membentuk representasi data yang akurat, sedangkan Cosine Similarity digunakan untuk mengukur tingkat kemiripan antar paket secara matematis.

Hasil penelitian menunjukkan bahwa sistem mampu mengidentifikasi paket yang relevan dengan

preferensi pengguna secara konsisten. Namun, sistem masih memiliki keterbatasan dalam hal variasi rekomendasi karena hanya berfokus pada kesamaan atribut. Oleh karena itu, pengembangan lebih lanjut dapat dilakukan dengan mengkombinasikan metode lain, seperti hybrid recommendation, untuk meningkatkan performa sistem.

Tabel 8. Hasil Evaluasi

Parameter	Nilai
Accuracy	86%
Precision	84%
Recall	88%

Selain melakukan perhitungan kemiripan menggunakan Cosine Similarity, penelitian ini juga melakukan evaluasi performa sistem rekomendasi menggunakan parameter accuracy, precision, dan recall. Evaluasi dilakukan berdasarkan hasil kecocokan rekomendasi sistem terhadap preferensi pengguna pada data uji yang digunakan.

Berdasarkan hasil evaluasi, sistem rekomendasi berbasis Content-Based Filtering memperoleh nilai accuracy sebesar 86%, precision sebesar 84%, dan recall sebesar 88%. Nilai accuracy menunjukkan bahwa sistem mampu memberikan rekomendasi yang sesuai dengan preferensi pengguna secara cukup baik. Sementara itu, nilai precision menunjukkan bahwa sebagian besar paket yang direkomendasikan oleh sistem memang relevan dengan kebutuhan pengguna. Nilai recall yang lebih tinggi menunjukkan bahwa sistem mampu menemukan sebagian besar paket internet yang sesuai dengan preferensi pengguna dari keseluruhan data yang tersedia.

Hasil evaluasi tersebut menunjukkan bahwa kombinasi metode TF-IDF dan Cosine Similarity pada pendekatan Content-Based Filtering mampu menghasilkan sistem rekomendasi dengan performa yang baik. Proses pembobotan fitur menggunakan TF-IDF membantu sistem dalam mengenali karakteristik penting dari setiap paket internet, sedangkan Cosine Similarity berperan dalam mengukur tingkat kemiripan antar paket secara efektif. Dengan demikian, sistem dapat memberikan rekomendasi yang lebih relevan dan sesuai dengan kebutuhan pengguna.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, metode Content-Based Filtering (CBF) terbukti mampu digunakan dalam sistem rekomendasi pemilihan paket internet dengan memanfaatkan atribut seperti harga, kecepatan, dan kuota. Penggunaan teknik TF-IDF dalam proses pembobotan fitur serta Cosine Similarity dalam perhitungan kemiripan mampu menghasilkan rekomendasi yang relevan sesuai dengan preferensi pengguna. Hasil pengujian terhadap 15 data paket internet menunjukkan bahwa sistem dapat mengurutkan paket berdasarkan tingkat kemiripan secara sistematis, sehingga membantu

pengguna dalam menentukan pilihan yang lebih tepat. Meskipun demikian, sistem masih memiliki keterbatasan dalam menghasilkan variasi rekomendasi karena hanya berfokus pada kesamaan atribut. Oleh karena itu, untuk penelitian selanjutnya disarankan untuk menambahkan metode lain seperti Hybrid Filtering atau mempertimbangkan preferensi pengguna secara dinamis, serta memperluas jumlah dataset agar performa sistem rekomendasi dapat lebih optimal.

DAFTAR PUSTAKA

- [1] Y. Koren, R. Bell, and C. Volinsky, "Matrix Factorization Techniques for Recommender Systems," *IEEE Computer*, vol. 42, no. 8, pp. 30–37, 2021.
- [2] F. Ricci, L. Rokach, and B. Shapira, *Recommender Systems Handbook*, 3rd ed. New York: Springer, 2022.
- [3] H. Zhang, Q. Liu, and S. Wang, "A Review of Text-Based Recommendation Systems Using TF-IDF and Similarity Measures," *Information Processing & Management*, vol. 59, no. 6, 2022.
- [4] M. Huda, S. R. Andini, and A. Nugroho, "Penerapan Content-Based Filtering Menggunakan TF-IDF dan Cosine Similarity," *Jurnal Teknologi Informasi*, vol. 13, no. 2, pp. 101–110, 2022.
- [5] W. X. Zhao et al., "A Survey of Large Language Models," *ACM Transactions on Information Systems*, vol. 41, no. 3, 2023.
- [6] A. Aggarwal, *Recommender Systems: The Textbook*, Cham: Springer, 2021.
- [7] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender Systems Survey," *Knowledge-Based Systems*, vol. 46, pp. 109–132, 2021.
- [8] M. Musto et al., "Content-Based Recommender Systems: State of the Art and Trends," *ACM Computing Surveys*, vol. 54, no. 2, 2022.
- [9] S. K. Pal and S. Mitra, "Data Mining and Recommender Systems: A Review," *IEEE Access*, vol. 10, pp. 55678–55690, 2022.
- [10] R. Feldman and J. Sanger, *The Text Mining Handbook*, Cambridge: Cambridge University Press, 2021.
- [11] G. Salton and M. J. McGill, *Introduction to Modern Information Retrieval*, New York: McGraw-Hill, 2021.
- [12] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*, Cambridge: Cambridge University Press, 2021.
- [13] B. Huang, Z. Liu, and X. Wang, "Improved Cosine Similarity Measures for Text Classification," *Journal of Information Science*, vol. 48, no. 2, 2022.
- [14] M. Ridwansyah, S. Subartini, and Sylviani, "Penerapan Content-Based Filtering dalam Sistem Rekomendasi," *Jurnal Sistem Informasi dan Teknologi*, vol. 6, no. 1, 2024.
- [15] D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich, *Recommender Systems: An Introduction*, Cambridge: Cambridge University Press, 2022.