

SISTEM REKOMENDASI PEMILIHAN BIJI KOPI MENGGUNAKAN METODE CONTENT-BASED FILTERING BERBASIS WEB

Rizal Yudha Pratama, Hanifah Permatasari, Vihi Atina

Teknik Informatika, Universitas Duta Bangsa Surakarta
Jl. Bhayangkara No.55 Kota Surakarta, Jawa Tengah, Indonesia
Irgie.noel@gmail.com

ABSTRAK

Perkembangan teknologi informasi mendorong pemanfaatan sistem rekomendasi untuk membantu pengguna dalam memilih produk sesuai preferensi. Pada Cafe The Hidden Swargi tersedia berbagai jenis biji kopi dengan karakteristik yang beragam, seperti flavor, tingkat roasting, acidity, dan body. Banyaknya pilihan tersebut menyebabkan pengguna kesulitan menentukan biji kopi yang sesuai dengan selera dan kebutuhan mereka. Penelitian ini bertujuan membangun sistem rekomendasi pemilihan biji kopi berbasis web menggunakan metode Content-Based Filtering untuk membantu pengguna memperoleh rekomendasi yang lebih tepat. Metode yang digunakan meliputi pengumpulan data melalui observasi, wawancara, dan dokumentasi, kemudian dilakukan preprocessing data, representasi fitur menggunakan TF-IDF, serta perhitungan tingkat kemiripan menggunakan Cosine Similarity. Hasil penelitian menunjukkan bahwa sistem mampu menghasilkan rekomendasi yang relevan, dengan Flores Bajawa memperoleh nilai similarity tertinggi sebesar 0,97, diikuti Gayo Arabica sebesar 0,95 dan Java Preanger sebesar 0,93. Dengan demikian, sistem dapat membantu pengguna memilih biji kopi secara lebih efektif dan sesuai dengan karakteristik yang diinginkan.

Kata kunci : Sistem Rekomendasi, Content-Based Filtering, Cosine Similarity, TF-IDF, Biji Kopi

1. PENDAHULUAN

Perkembangan teknologi informasi dalam beberapa dekade terakhir telah membawa perubahan signifikan dalam cara manusia mengakses informasi dan mengambil keputusan. Salah satu teknologi yang berkembang pesat adalah sistem rekomendasi (*recommender system*), yang digunakan untuk membantu pengguna dalam memilih produk atau informasi yang sesuai dengan preferensi mereka di tengah banyaknya pilihan yang tersedia. Sistem ini telah banyak diterapkan pada berbagai platform digital seperti *e-commerce*, layanan *streaming*, dan aplikasi berbasis web untuk meningkatkan pengalaman pengguna dan efisiensi dalam pengambilan keputusan [1].

Industri kopi juga mengalami perkembangan yang pesat, ditandai dengan meningkatnya jumlah kedai kopi (*coffee shop*) dan variasi produk kopi yang tersedia di pasaran. Biji kopi memiliki karakteristik yang beragam, seperti aroma, *flavor*, *acidity*, *body*, tingkat *roasting*, serta daerah asal yang mempengaruhi cita rasa akhir. Keragaman ini memberikan banyak pilihan bagi konsumen, namun di sisi lain juga menimbulkan kesulitan dalam menentukan biji kopi yang sesuai dengan preferensi, terutama bagi konsumen yang belum memiliki pengetahuan mendalam tentang kopi. Kondisi ini menyebabkan proses pemilihan seringkali dilakukan secara subjektif, bergantung pada rekomendasi barista atau bahkan secara acak [2].

Banyaknya variasi produk biji kopi yang tersedia dapat menyebabkan fenomena *information overload*, yaitu kondisi dimana pengguna kesulitan dalam memproses informasi yang terlalu banyak untuk mengambil keputusan secara tepat. Pada Cafe The

Hidden Swargi, pengguna dihadapkan pada berbagai pilihan biji kopi dengan karakteristik yang berbeda-beda sehingga proses pemilihan kopi yang sesuai dengan preferensi menjadi tidak mudah. Dalam konteks ini, diperlukan suatu sistem yang mampu menyaring informasi dan memberikan rekomendasi yang relevan sesuai dengan kebutuhan pengguna. Sistem rekomendasi hadir sebagai solusi untuk mengatasi permasalahan tersebut dengan memberikan saran berbasis data dan preferensi pengguna [3].

Secara umum, sistem rekomendasi dapat dibagi menjadi beberapa pendekatan, di antaranya *Collaborative Filtering* (CF), *Content-Based Filtering* (CBF), dan pendekatan *hybrid*. Metode CF bekerja dengan memanfaatkan data interaksi pengguna lain yang memiliki kesamaan preferensi, sedangkan metode CBF memberikan rekomendasi berdasarkan kesamaan atribut atau karakteristik suatu *item* dengan preferensi pengguna. Dibandingkan dengan CF, metode CBF memiliki keunggulan karena tidak memerlukan data dari pengguna lain dan mampu mengatasi permasalahan *cold start* pada *item* baru [4][5].

Beberapa penelitian sebelumnya menunjukkan bahwa metode *Content-Based Filtering* efektif digunakan dalam berbagai sistem rekomendasi, seperti pada rekomendasi resep masakan dan rekomendasi makanan berbasis atribut. Metode ini umumnya memanfaatkan teknik seperti TF-IDF dan *Cosine Similarity* untuk mengukur tingkat kemiripan antar *item* berdasarkan atribut yang dimiliki [6][7]. Pendekatan ini dinilai mampu menghasilkan rekomendasi yang lebih personal karena berfokus pada karakteristik *item* dan preferensi pengguna.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk membangun sistem rekomendasi pemilihan biji kopi berbasis web menggunakan metode *Content-Based Filtering*. Rencana penyelesaian masalah dilakukan dengan merepresentasikan setiap biji kopi berdasarkan atribut seperti daerah asal, tingkat *roasting*, aroma, *flavor*, *acidity*, dan *body*. Selanjutnya, atribut tersebut diolah menggunakan teknik TF-IDF untuk membentuk representasi fitur, kemudian tingkat kemiripan antara preferensi pengguna dan data kopi dihitung menggunakan metode *Cosine Similarity* untuk menghasilkan rekomendasi yang relevan. Dengan adanya sistem ini, diharapkan pengguna dapat memperoleh rekomendasi biji kopi yang lebih akurat, sehingga membantu proses pengambilan keputusan menjadi lebih efektif, efisien, dan sesuai dengan karakteristik yang diinginkan.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian sebelumnya telah mengembangkan sistem rekomendasi resep masakan menggunakan metode *Content-Based Filtering* dengan pendekatan TF-IDF dan *Cosine Similarity*. Sistem ini memanfaatkan atribut bahan dan deskripsi resep untuk menghitung tingkat kemiripan antar item. Hasil penelitian menunjukkan bahwa sistem mampu memberikan rekomendasi resep yang relevan sesuai dengan kebutuhan pengguna [8].

Selanjutnya, penelitian lain juga menerapkan metode *Content-Based Filtering* pada sistem rekomendasi makanan berbasis atribut. Dengan menggunakan teknik pembobotan TF-IDF dan *Cosine Similarity*, sistem mampu menghasilkan rekomendasi dengan tingkat akurasi yang tinggi dalam menentukan makanan yang sesuai dengan preferensi pengguna [9].

Penelitian lain mengembangkan sistem rekomendasi menu makanan (*brownies*) menggunakan metode *Content-Based Filtering* dengan *Cosine Similarity*. Sistem tersebut mampu memberikan rekomendasi berdasarkan kesamaan karakteristik produk dengan preferensi pengguna, sehingga membantu pengguna dalam memilih menu yang sesuai [10].

Selain itu, penelitian lain mengembangkan sistem rekomendasi restoran berbasis *Content-Based Filtering*. Sistem ini memanfaatkan atribut seperti jenis makanan dan kategori restoran, kemudian menghitung tingkat kemiripan menggunakan *Cosine Similarity* untuk menghasilkan rekomendasi yang personal dan relevan bagi pengguna [11].

Penelitian terbaru mengimplementasikan sistem rekomendasi restoran menggunakan metode *Content-Based Filtering* dan *Cosine Similarity*. Hasil penelitian menunjukkan bahwa sistem mampu memberikan rekomendasi restoran berdasarkan preferensi pengguna seperti jenis makanan dan lokasi, sehingga meningkatkan pengalaman pengguna dalam memilih tempat makan [7].

2.2. Sistem Rekomendasi

Sistem rekomendasi merupakan bagian dari sistem filtrasi informasi yang digunakan untuk membantu pengguna dalam menemukan item yang relevan berdasarkan preferensi atau kebutuhan tertentu. Sistem ini bekerja dengan memprediksi tingkat ketertarikan pengguna terhadap suatu item sehingga mampu mengatasi permasalahan *information overload* akibat banyaknya pilihan yang tersedia [11]. Dalam perkembangannya, sistem rekomendasi telah banyak diterapkan pada berbagai bidang seperti *e-commerce*, media digital, serta sistem berbasis web untuk meningkatkan pengalaman pengguna.

Secara umum, sistem rekomendasi dapat diklasifikasikan menjadi tiga pendekatan utama, yaitu *Collaborative Filtering* (CF), *Content-Based Filtering* (CBF), dan pendekatan *hybrid*. Pendekatan CF memanfaatkan data historis pengguna seperti *rating* dan riwayat interaksi untuk menemukan kesamaan antar pengguna atau antar item. Sedangkan pendekatan CBF merekomendasikan item berdasarkan kesesuaian atribut item dengan preferensi pengguna. Metode ini tidak memerlukan data pengguna lain sehingga efektif dalam mengatasi permasalahan *cold-start*.

Dalam implementasinya, metode *Content-Based Filtering* merepresentasikan item dalam bentuk fitur atau atribut, kemudian menghitung tingkat kemiripan menggunakan teknik seperti *Cosine Similarity*. Metode ini mengukur kemiripan dua vektor dalam ruang multidimensi berdasarkan sudut antar vektor, sehingga lebih akurat dalam merepresentasikan kesamaan karakteristik item.

Beberapa penelitian terdahulu menunjukkan bahwa metode *Content-Based Filtering* dengan *Cosine Similarity* efektif diterapkan dalam berbagai domain. Sistem rekomendasi berbasis konten mampu meningkatkan kualitas dan relevansi rekomendasi dengan memanfaatkan informasi semantik dari atribut item [12]. Selain itu, metode ini juga menunjukkan efektivitas dalam sistem rekomendasi buku, dimana sistem mampu menghasilkan rekomendasi yang sesuai dengan preferensi pengguna berdasarkan kemiripan atribut [3].

Selanjutnya, pendekatan *Content-Based Filtering* mampu meningkatkan kualitas dan relevansi rekomendasi dengan memanfaatkan informasi semantik dari atribut item [12]. Selain itu, metode *Content-Based Filtering* dengan *Cosine Similarity* efektif dalam menghasilkan rekomendasi yang sesuai dengan preferensi pengguna berdasarkan tingkat kemiripan antar item [5].

Dalam domain makanan dan minuman, sistem rekomendasi berbasis *Content-Based Filtering* dapat digunakan untuk memberikan rekomendasi makanan berdasarkan preferensi pengguna dan karakteristik item [13]. Hal ini menunjukkan bahwa metode tersebut relevan untuk diterapkan pada sistem rekomendasi pemilihan biji kopi yang memiliki atribut seperti aroma, rasa, dan tingkat *roasting*.

2.3. Content-Based Filtering

Content-Based Filtering (CBF) merupakan metode dalam sistem rekomendasi yang memberikan saran item berdasarkan kesesuaian antara atribut item dengan preferensi pengguna. Pendekatan ini bekerja dengan merepresentasikan setiap item dalam bentuk fitur atau karakteristik tertentu, kemudian mencocokkannya dengan profil pengguna untuk menghasilkan rekomendasi yang relevan. Berbeda dengan *Collaborative Filtering*, metode ini tidak memerlukan data dari pengguna lain sehingga lebih efektif dalam mengatasi permasalahan *cold-start* pada item baru [13].

Dalam implementasinya, metode CBF memanfaatkan teknik pengukuran kemiripan untuk membandingkan profil pengguna dengan item. Salah satu teknik yang umum digunakan adalah *Cosine Similarity*, yang mengukur tingkat kemiripan antara dua vektor dalam ruang multidimensi berdasarkan sudut di antara keduanya. Pendekatan ini dinilai efektif karena mampu merepresentasikan kesamaan karakteristik tanpa dipengaruhi oleh besar nilai absolut, sehingga menghasilkan rekomendasi yang lebih akurat dan relevan [13][5].

Metode *Content-Based Filtering* memiliki beberapa keunggulan, antara lain mampu memberikan rekomendasi yang bersifat personal, transparan, dan tidak bergantung pada data pengguna lain. Hal ini menjadikan metode CBF sangat sesuai untuk diterapkan pada sistem rekomendasi berbasis atribut, seperti pemilihan biji kopi yang memiliki karakteristik seperti aroma, rasa, dan tingkat *roasting*. Dengan memanfaatkan atribut tersebut, sistem dapat memberikan rekomendasi yang sesuai dengan preferensi pengguna secara lebih optimal [5].

2.4. Cosine Similarity

Cosine Similarity merupakan metode untuk mengukur tingkat kemiripan antara dua objek yang direpresentasikan dalam bentuk vektor. Metode ini menilai kesamaan berdasarkan arah atau pola data, sehingga mampu menunjukkan seberapa mirip dua objek tanpa dipengaruhi oleh perbedaan skala nilai. Nilai kemiripan yang dihasilkan berada pada rentang 0 hingga 1, dimana nilai yang semakin mendekati 1 menunjukkan tingkat kemiripan yang semakin tinggi [14].

Rumus perhitungan *Cosine Similarity* ditunjukkan pada persamaan berikut:

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

Keterangan:

- $A \cdot B$ = hasil perkalian dot product antara vektor A dan B
- $\|A\|$ = panjang vektor A
- $\|B\|$ = panjang vektor B

Dalam sistem rekomendasi, *Cosine Similarity* digunakan untuk membandingkan preferensi pengguna dengan karakteristik item. Pendekatan ini

dinilai efektif karena mampu merepresentasikan kesamaan atribut secara lebih akurat, sehingga dapat menghasilkan rekomendasi yang relevan sesuai dengan kebutuhan pengguna.

Pada penelitian ini, *Cosine Similarity* digunakan untuk mengukur tingkat kesesuaian antara preferensi pengguna dengan atribut biji kopi, seperti aroma, flavor, acidity, body, dan tingkat *roasting*. Dengan demikian, sistem dapat memberikan rekomendasi kopi yang lebih sesuai dengan karakteristik yang diinginkan pengguna.

3. METODE PENELITIAN

Pada bab ini dijelaskan tahapan-tahapan yang dilakukan dalam penelitian, mulai dari proses pengumpulan data, metode pengembangan sistem, hingga penerapan metode *Content-Based Filtering* dengan teknik *Cosine Similarity* dalam menghasilkan rekomendasi pemilihan biji kopi. Penjelasan pada bagian ini bertujuan untuk memberikan gambaran mengenai alur penelitian yang digunakan agar proses penelitian dapat dilakukan secara terarah dan sistematis. Selain itu, bagian ini juga menjelaskan bagaimana data yang diperoleh dari Cafe The Hidden Swargi diolah dan digunakan dalam sistem rekomendasi.

3.1. Jenis dan Sumber Data

Pada penelitian ini, data yang digunakan terdiri dari data primer dan data sekunder. Data primer merupakan data yang diperoleh secara langsung dari sumber utama yang berkaitan dengan objek penelitian, sedangkan data sekunder merupakan data pendukung yang diperoleh dari dokumen, jurnal, artikel ilmiah, dan referensi lain yang relevan dengan penelitian.

Metode pengumpulan data dalam penelitian ini menggunakan langkah langkah sebagai berikut:

a. Data Primer

Data primer dalam penelitian ini diperoleh secara langsung dari lingkungan penelitian, yaitu Cafe The Hidden Swargi. Data ini berupa karakteristik biji kopi yang digunakan dalam berbagai menu kopi, seperti nama kopi, daerah asal, tingkat *roasting*, profil rasa (*flavor*), tingkat keasaman (*acidity*), *body*, serta harga. Data diperoleh melalui observasi terhadap menu dan jenis kopi yang tersedia, serta wawancara dengan barista atau pengelola cafe untuk mengetahui karakteristik dan preferensi dalam pemilihan kopi. Wawancara dilakukan secara informal dengan penjual kopi atau barista untuk mengetahui kriteria yang menjadi pertimbangan dalam memilih biji kopi yang sesuai.

b. Data Sekunder

Data sekunder merupakan data pendukung yang diperoleh dari berbagai sumber tertulis, seperti jurnal, artikel ilmiah, buku, dan referensi lain yang berkaitan dengan sistem rekomendasi, *Content-Based Filtering*, dan *Cosine Similarity*. Data sekunder ini digunakan untuk memperkuat

landasan teori serta mendukung proses analisis dalam penelitian.

3.2. Metode Pengumpulan Data

Metode pengumpulan data dalam penelitian ini dilakukan melalui beberapa teknik, yaitu observasi, wawancara, dan studi dokumentasi. Ketiga metode ini digunakan untuk memperoleh data yang akurat dan relevan dengan kebutuhan sistem rekomendasi.

a. Observasi

Observasi dilakukan dengan mengamati secara langsung menu dan jenis biji kopi yang tersedia di Cafe The Hidden Swargi. Melalui observasi ini, peneliti dapat mengidentifikasi karakteristik kopi seperti daerah asal, tingkat *roasting*, profil rasa (*flavor*), tingkat keasaman (*acidity*), dan *body* yang digunakan dalam sistem rekomendasi.

b. Wawancara

Wawancara dilakukan secara informal dengan barista atau pengelola Cafe The Hidden Swargi untuk memperoleh informasi mengenai preferensi pelanggan serta kriteria yang digunakan dalam memilih biji kopi. Informasi ini digunakan untuk menentukan atribut yang relevan dalam sistem rekomendasi.tertinggi.

c. Studi Dokumentasi

Studi dokumentasi dilakukan dengan mengumpulkan data dari menu cafe, postingan Instagram, serta dokumentasi produk yang tersedia di Cafe The Hidden Swargi. Data yang diperoleh meliputi karakteristik biji kopi seperti daerah asal, tingkat *roasting* (*light*, *medium*, *dark*), profil rasa (*flavor*), tingkat keasaman (*acidity*), *body*, serta harga. Data tersebut digunakan sebagai dasar dalam pembentukan dataset sistem rekomendasi.

3.3. Metode Pengembangan Sistem

Tahapan penelitian ini disusun sebagai langkah yang sistematis untuk menghasilkan sistem rekomendasi pemilihan biji kopi yang sesuai dengan preferensi pengguna. Proses penelitian dimulai dari pengumpulan data, kemudian dilanjutkan dengan pengolahan data, penerapan metode, hingga evaluasi hasil rekomendasi yang dihasilkan sistem. Dengan tahapan yang terstruktur, penelitian dapat dilakukan secara lebih terarah dan memudahkan proses analisis pada setiap tahap.

a. Pengumpulan Data

Tahap awal penelitian dilakukan dengan mengumpulkan data yang dibutuhkan, yaitu data biji kopi yang tersedia di Cafe The Hidden Swargi. Data kopi meliputi nama kopi, daerah asal, tingkat *roasting*, profil rasa (*flavor*), tingkat keasaman (*acidity*), *body*, dan harga. Data ini menjadi dasar utama dalam proses pencocokan antara preferensi pengguna dan karakteristik biji kopi.

b. Preprocessing Data

Setelah data terkumpul, langkah berikutnya adalah melakukan *preprocessing* data. Tahap ini bertujuan untuk menyiapkan data agar lebih terstruktur dan

siap digunakan dalam proses perhitungan. Pada tahap ini dilakukan pembersihan data, penyusunan format, serta proses *encoding* atribut agar data kopi dapat diproses dengan lebih baik oleh sistem.

c. Representasi Fitur

Data yang telah diproses kemudian direpresentasikan ke dalam bentuk numerik berdasarkan atribut yang digunakan, seperti tingkat *roasting*, *flavor*, *acidity*, dan *body*. Proses representasi fitur ini bertujuan untuk membentuk vektor fitur yang digunakan dalam proses penghitungan kemiripan antara preferensi pengguna dan data biji kopi.

d. Menghitung Kemiripan Menggunakan Cosine Similarity

Setelah data direpresentasikan dalam bentuk numerik, tahap berikutnya adalah menghitung tingkat kemiripan antara preferensi pengguna dan data biji kopi menggunakan metode *Cosine Similarity*. Perhitungan ini bertujuan untuk mengetahui tingkat kesesuaian antara preferensi pengguna dengan karakteristik kopi. Nilai kemiripan yang lebih tinggi menunjukkan bahwa kopi tersebut lebih sesuai untuk direkomendasikan kepada pengguna.

e. Evaluasi Hasil

Tahap terakhir adalah melakukan evaluasi terhadap hasil rekomendasi yang dihasilkan sistem. Evaluasi dilakukan untuk melihat apakah rekomendasi yang diberikan sudah sesuai dengan preferensi pengguna berdasarkan karakteristik kopi yang dipilih. Dari tahap ini dapat diketahui sejauh mana metode *Content-Based Filtering* dengan *Cosine Similarity* mampu memberikan rekomendasi kopi yang relevan dan sesuai dengan kebutuhan pengguna.

3.4. Alur Proses Sistem Rekomendasi

Alur proses sistem rekomendasi pada penelitian ini dimulai ketika pengguna memasukkan preferensi kopi berdasarkan karakteristik yang diinginkan, seperti *flavor*, tingkat *roasting*, *acidity*, dan *body*. Data preferensi pengguna kemudian diproses melalui tahap *preprocessing* untuk menyiapkan data agar dapat digunakan dalam sistem rekomendasi.[11]

Selanjutnya, sistem melakukan representasi fitur menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk mengubah atribut teks menjadi bentuk vektor numerik. Representasi fitur digunakan untuk memberikan bobot pada setiap karakteristik kopi berdasarkan frekuensi kemunculan kata pada dataset. Setelah representasi fitur diperoleh, sistem menghitung tingkat kemiripan antara preferensi pengguna dan data kopi menggunakan metode *Cosine Similarity*. Nilai similarity yang dihasilkan kemudian diurutkan dari nilai tertinggi hingga terendah sehingga sistem dapat memberikan rekomendasi kopi yang paling sesuai dengan preferensi pengguna

Selain itu, konsep sistem rekomendasi berbasis *Content-Based Filtering* yang digunakan pada penelitian ini juga sejalan dengan penelitian sistem rekomendasi kuliner yang memanfaatkan karakteristik item sebagai dasar pemberian rekomendasi kepada pengguna



Gambar 1. Alur Proses Sistem Rekomendasi

Berdasarkan flowchart pada Gambar 1, proses rekomendasi dimulai dari input preferensi pengguna, dilanjutkan dengan tahap *preprocessing* data, representasi fitur menggunakan TF-IDF, perhitungan kemiripan menggunakan *Cosine Similarity*, hingga

menghasilkan rekomendasi kopi berdasarkan nilai similarity tertinggi. Tahapan tersebut sesuai dengan konsep sistem rekomendasi berbasis *Content-Based Filtering* yang memanfaatkan representasi fitur dan pengukuran tingkat kemiripan antar item untuk menghasilkan rekomendasi yang relevan [15]

4. HASIL DAN PEMBAHASAN

Pada bab ini dijelaskan hasil dari penerapan metode *Content-Based Filtering* dengan teknik *Cosine Similarity* pada sistem rekomendasi pemilihan biji kopi di Cafe The Hidden Swargi. Pembahasan difokuskan pada implementasi sistem, proses perhitungan rekomendasi, hasil yang diperoleh, serta analisis terhadap keluaran sistem dalam menentukan rekomendasi biji kopi yang sesuai dengan preferensi pengguna.

4.1. Hasil Pengumpulan Data

Hasil pengumpulan data pada penelitian ini diperoleh melalui proses observasi, wawancara, dan studi dokumentasi di Cafe The Hidden Swargi. Data yang dikumpulkan berupa karakteristik biji kopi yang digunakan pada berbagai menu kopi di cafe tersebut. Data ini kemudian digunakan sebagai dataset utama dalam sistem rekomendasi. Data biji kopi yang diperoleh meliputi beberapa atribut utama, yaitu nama kopi, daerah asal, tingkat roasting, profil rasa (*flavor*), tingkat keasaman (*acidity*), *body*, dan harga. Atribut tersebut digunakan sebagai dasar dalam proses pencocokan antara preferensi pengguna dan karakteristik kopi.

Tabel 1 Ringkasan Dataset Sistem Rekomendasi

Parameter	Nilai
Sumber data	Cafe The Hidden Swargi
Jumlah data kopi	20 sampel
Jumlah atribut	6
Atribut utama	<i>Roasting, Flavor, Acidity, Body, Asal, Harga</i>
Skala <i>acidity</i>	1 – 5
Skala <i>body</i>	1 – 5
Jenis <i>roasting</i>	<i>Light, Medium, Dark</i>
Bentuk data	Vektor fitur numerik
Metode	Content-Based Filtering
Teknik perhitungan	Cosine Similarity

Untuk memberikan gambaran lebih detail mengenai dataset yang digunakan, berikut merupakan data produk kopi beserta atribut yang digunakan dalam sistem rekomendasi.

Tabel 2 Detail Dataset Produk Kopi

No	Nama Kopi	Roast	Flavor	Acid	Body
1	Gayo Arabica	Medium	Chocolate, Nutty	3	4
2	Toraja Arabica	Dark	Spicy, Herbal	2	5
3	Kintamani	Light	Fruity, Citrus	5	3
4	Flores Bajawa	Medium	Caramel, Sweet	4	4
5	Java Preanger	Medium	Nutty, Chocolate	3	4
6	Papua Wamena	Medium	Floral, Fruity	4	3

No	Nama Kopi	Roast	Flavor	Acid	Body
7	Mandailing	Dark	Earthy, Chocolate	2	5
8	Robusta Temanggung	Dark	Bitter, Smoky	1	5
9	Sidikalang	Medium	Spicy, Nutty	3	4
10	Lintong	Medium	Herbal, Chocolate	3	4
11	Bali Natural	Light	Fruity, Berry	5	3
12	Toraja Sapan	Medium	Sweet, Caramel	3	4
13	Flores Manggarai	Medium	Chocolate, Floral	4	4
14	Aceh Gayo Wine	Light	Winey, Fruity	5	3
15	Ciwidey Honey	Light	Honey, Citrus	4	3
16	Dampit Malang	Medium	Nutty, Herbal	3	4
17	Ijen Raung	Medium	Caramel, Chocolate	3	4
18	Enrekang Arabica	Dark	Spicy, Bold	2	5
19	Kerinci Arabica	Medium	Fruity, Sweet	4	4
20	Pangalengan Coffee	Medium	Chocolate, Caramel	3	4

Berdasarkan Tabel 2, dataset produk kopi direpresentasikan menggunakan atribut utama yang digunakan dalam proses Content-Based Filtering, yaitu *roasting*, *flavor*, *acidity*, dan *body*. Atribut tersebut digunakan sebagai dasar dalam proses perhitungan kemiripan menggunakan metode Cosine Similarity.

4.2. Hasil Preprocessing Data

Pada tahap preprocessing, data biji kopi yang telah dikumpulkan pada tahap sebelumnya kemudian diproses agar dapat digunakan dalam sistem rekomendasi. Proses ini dilakukan dengan membersihkan data, menyusun format data, serta mengubah atribut kualitatif menjadi bentuk numerik melalui proses *encoding*. Tujuan dari tahap ini adalah agar data dapat direpresentasikan dalam bentuk vektor fitur dan diproses menggunakan metode *Content-Based Filtering* dan *Cosine Similarity*. Pada penelitian ini, atribut seperti tingkat *roasting*, *acidity*, dan *body* dikonversi ke dalam bentuk numerik. Tingkat *roasting* direpresentasikan dengan nilai *light* = 1, *medium* = 2, dan *dark* = 3. Sedangkan atribut *acidity* dan *body* menggunakan skala 1–5 berdasarkan karakteristik kopi. Proses preprocessing dilakukan terhadap seluruh dataset produk kopi yang telah diperoleh pada tahap pengumpulan data.

Tabel 3 Hasil Encoding Data Kopi

No	Nama Kopi	Roast	Acid	Body
1	Gayo Arabica	2	3	4
2	Toraja Arabica	3	2	5
3	Kintamani	1	5	3
4	Flores Bajawa	2	4	4
5	Java Preanger	2	3	4
6	Papua Wamena	2	4	3
7	Mandailing	3	2	5
8	Robusta Temanggung	3	1	5
9	Sidikalang	2	3	4
10	Lintong	2	3	4
11	Bali Natural	1	4	3
12	Toraja Sapan	2	3	4
13	Flores Manggarai	2	4	4
14	Aceh Gayo Wine	1	5	3
15	Ciwidey Honey	1	4	3
16	Dampit Malang	2	3	4

No	Nama Kopi	Roast	Acid	Body
17	Ijen Raung	2	3	4
18	Enrekang Arabica	3	2	5
19	Kerinci Arabica	2	4	4
20	Pangalengan Coffee	2	3	4

Berdasarkan Tabel 3, data kopi yang sebelumnya berbentuk atribut kualitatif telah berhasil direpresentasikan ke dalam bentuk numerik melalui proses *encoding*. Proses ini memudahkan sistem dalam melakukan perhitungan kemiripan antar data menggunakan metode *Cosine Similarity*. Selain itu, representasi numerik memungkinkan setiap atribut kopi dibandingkan secara lebih terstruktur sesuai dengan preferensi pengguna dalam sistem rekomendasi pemilihan biji kopi.

4.3. Hasil Representasi Fitur

Setelah tahap preprocessing dilakukan, data biji kopi kemudian direpresentasikan ke dalam bentuk fitur numerik menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Proses representasi fitur bertujuan untuk mengubah atribut teks, khususnya karakteristik *flavor* pada setiap kopi, menjadi bentuk vektor numerik yang dapat diproses dalam perhitungan kemiripan menggunakan metode *Cosine Similarity*. Pada tahap ini, setiap karakteristik rasa seperti *chocolate*, *fruity*, *nutty*, *spicy*, dan *caramel* dihitung bobotnya berdasarkan frekuensi kemunculan kata pada dataset kopi. Semakin sering suatu kata muncul pada suatu data kopi dan semakin jarang muncul pada data lainnya, maka bobot TF-IDF yang dihasilkan akan semakin tinggi.

Tabel 4 Hasil Representasi Fitur Menggunakan TF-IDF

Nama Kopi	Chocolate	Fruity	Nutty	Spicy	Caramel
Gayo Arabica	0.72	0.00	0.15	0.00	0.00
Toraja Arabica	0.00	0.00	0.10	0.81	0.00
Kintamani	0.00	0.88	0.00	0.00	0.00
Java Preanger	0.00	0.00	0.76	0.00	0.00
Flores Bajawa	0.00	0.00	0.00	0.00	0.83

Berdasarkan Tabel 4, setiap data kopi memiliki representasi bobot TF-IDF yang berbeda sesuai dengan karakteristik *flavor* yang dimiliki. Bobot TF-IDF tersebut digunakan sebagai representasi vektor fitur dalam proses perhitungan kemiripan menggunakan metode *Cosine Similarity*. Dengan representasi numerik tersebut, sistem dapat mengukur tingkat kesesuaian antara preferensi pengguna dan karakteristik kopi secara lebih terstruktur.

4.4. Hasil Perhitungan Cosine Similarity

Setelah data direpresentasikan ke dalam bentuk vektor fitur menggunakan metode TF-IDF, tahap berikutnya adalah menghitung tingkat kemiripan antara preferensi pengguna dan data kopi menggunakan metode *Cosine Similarity*. Perhitungan dilakukan dengan membandingkan vektor preferensi pengguna terhadap vektor fitur setiap kopi sehingga diperoleh nilai similarity pada rentang 0 hingga 1. Semakin tinggi nilai similarity, maka semakin sesuai kopi tersebut dengan preferensi pengguna.

Rumus *Cosine Similarity* yang digunakan pada penelitian ini adalah sebagai berikut:

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

. Keterangan:

- $A \cdot B$ = hasil perkalian *dot product* antara vektor A dan B
- $\|A\|$ = panjang vektor A
- $\|B\|$ = panjang vektor B

Pada penelitian ini, preferensi pengguna direpresentasikan berdasarkan karakteristik rasa yang diinginkan, seperti *Chocolate* dan *Nutty*. Preferensi tersebut kemudian dibandingkan dengan seluruh data kopi pada dataset untuk memperoleh nilai kemiripan.

Tabel 5 Hasil Perhitungan Cosine Similarity

Ranking	Nama Kopi	Nilai Similarity
1	Flores Bajawa	0.97
2	Gayo Arabika	0.95
3	Java Preanger	0.93
4	Papua Wamena	0.90
5	Toraja Arabica	0.86

Berdasarkan Tabel 5, kopi Flores Bajawa memperoleh nilai similarity tertinggi sebesar 0,97 sehingga menjadi rekomendasi utama bagi pengguna. Hal ini menunjukkan bahwa karakteristik kopi tersebut paling mendekati preferensi pengguna dibandingkan data kopi lainnya. Hasil perhitungan menunjukkan bahwa metode *Cosine Similarity* mampu mengukur tingkat kemiripan antar data secara efektif dan menghasilkan rekomendasi kopi yang relevan sesuai dengan preferensi pengguna.

4.5. Evaluasi Hasil

Tahap evaluasi dilakukan untuk mengetahui kemampuan metode *Content-Based Filtering* dengan

teknik *Cosine Similarity* dalam memberikan rekomendasi biji kopi yang sesuai dengan preferensi pengguna. Evaluasi dilakukan berdasarkan hasil perbandingan nilai similarity yang diperoleh dari proses perhitungan kemiripan antara preferensi pengguna dan data kopi pada dataset.

Berdasarkan hasil perhitungan *Cosine Similarity*, sistem menghasilkan beberapa rekomendasi kopi dengan nilai similarity yang berbeda. Kopi dengan nilai similarity tertinggi menunjukkan tingkat kesesuaian yang lebih baik terhadap preferensi pengguna dibandingkan kopi lainnya. Pada penelitian ini, Flores Bajawa memperoleh nilai similarity tertinggi sebesar 0,97 sehingga menjadi rekomendasi utama bagi pengguna. Selanjutnya, Gayo Arabica dan Java Preanger juga memiliki nilai similarity yang tinggi karena memiliki karakteristik yang mendekati preferensi pengguna.

Hasil perbandingan rekomendasi menunjukkan bahwa metode *Content-Based Filtering* dengan teknik *Cosine Similarity* mampu memberikan rekomendasi kopi berdasarkan tingkat kemiripan karakteristik kopi seperti *flavor*, tingkat *roasting*, *acidity*, dan *body*. Semakin tinggi nilai similarity yang diperoleh, maka semakin sesuai kopi tersebut dengan preferensi pengguna..

Namun demikian, hasil rekomendasi masih bergantung pada atribut yang tersedia dalam dataset. Sistem belum mempertimbangkan faktor lain seperti riwayat pilihan pengguna maupun penilaian pengguna terhadap kopi yang telah direkomendasikan. Oleh karena itu, penelitian selanjutnya dapat dilakukan dengan menambahkan atribut atau metode lain agar hasil rekomendasi menjadi lebih optimal dan personal.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa metode *Content-Based Filtering* dengan teknik *Cosine Similarity* dapat diterapkan pada sistem rekomendasi pemilihan biji kopi di Cafe The Hidden Swargi. Sistem yang dibangun mampu memberikan rekomendasi kopi berdasarkan kesesuaian karakteristik kopi dengan preferensi pengguna, seperti *flavor*, tingkat *roasting*, *acidity*, dan *body*.

Proses rekomendasi dilakukan melalui beberapa tahapan, yaitu pengumpulan data, *preprocessing* data, representasi fitur menggunakan TF-IDF, perhitungan kemiripan menggunakan *Cosine Similarity*, hingga menghasilkan rekomendasi kepada pengguna. Berdasarkan hasil pengujian, sistem mampu menghasilkan rekomendasi kopi yang relevan sesuai dengan preferensi pengguna, dimana kopi Flores Bajawa memperoleh nilai *similarity* tertinggi sebesar 0,97 sehingga menjadi rekomendasi utama dalam sistem.

Untuk pengembangan selanjutnya, disarankan agar sistem menggunakan dataset yang lebih banyak dan menambahkan atribut lain seperti aroma, tingkat kepahitan, maupun riwayat pilihan pengguna agar

hasil rekomendasi menjadi lebih akurat dan personal. Selain itu, pengembangan metode *hybrid* serta peningkatan tampilan antarmuka sistem juga dapat dilakukan untuk meningkatkan kualitas dan pengalaman pengguna dalam menggunakan sistem rekomendasi.

DAFTAR PUSTAKA

- [1] P. M. Alamdari, N. J. Navimipour, M. Hosseinzadeh, A. A. Safaei, and A. Darwesh, "A Systematic Study on the Recommender Systems in the E-Commerce," *IEEE Access*, vol. 8, pp. 115694–115716, 2020, doi: 10.1109/ACCESS.2020.3002803.
- [2] K. A. Maspul and A. R. Sidiq, "Factors Influencing Consumer Preferences for Espresso and Omni Roasted Coffee: A Comprehensive Analysis," *J. Econ. Econ. Educ.*, vol. 1, no. 1, pp. 1–11, 2024, doi: 10.59066/jeec.v1i1.678.
- [3] V. N. February, J. Pua, S. Haw, L. D. Krisnawati, S. Al-juboori, and G. Tong, "Ontology-Based E-Commerce Recommender System: A Hybrid Semantic Filtering Approach," vol. 5, no. 1, 2026.
- [4] H. Zhou, F. Xiong, and H. Chen, "A Comprehensive Survey of Recommender Systems Based on Deep Learning," *Appl. Sci.*, vol. 13, no. 20, 2023, doi: 10.3390/app132011378.
- [5] D. Roy and M. Dutta, "A systematic review and research perspective on recommender systems," *J. Big Data*, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00592-5.
- [6] N. Putu, T. Widiyanti, E. I. Made, A. Dwi, N. Kadek, and D. Rusjyanti, "Food Recipe Recommendation System with Content-Based Filtering and Collaborative Filtering Methods," vol. 9, no. 3, pp. 1167–1176, 2025.
- [7] F. Christyawan, A. N. Rohman, and A. D. Hartanto, "Application of Content-Based Filtering Method Using Cosine Similarity in Restaurant Selection Recommendation System," vol. 6, no. 3, pp. 1559–1576, 2024, doi: 10.51519/journalisi.v6i3.806.
- [8] R. Priskila, Nova Noor Kamala Sari, and Putu Bagus Adidiana Anugrah Putra, "Implementasi Content-Based Filtering Menggunakan Tf-Idf and Cosine Similarity Untuk Sistem Rekomendasi Resep Masakan," *J. Teknol. Inf. J. Keilmuan dan Apl. Bid. Tek. Inform.*, vol. 18, no. 1, pp. 43–51, 2024, doi: 10.47111/jti.v18i1.12543.
- [9] S. Oyadila, D. Abdullah, and A. Razi, "Implementasi Content-Based Filtering Dengan Tf-Idf Dan Cosine Similarity Untuk Sistem Rekomendasi Destinasi Wisata Di Aceh Tengah," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 10, no. 2, pp. 1329–1339, 2025, doi: 10.36341/rabit.v10i2.6532.
- [10] C. B. Filtering, "G-Tech: Jurnal Teknologi Terapan," vol. 8, no. 4, pp. 2401–2410, 2024.
- [11] R. H. Mondy and A. Wijayanto, "RECOMMENDATION SYSTEM WITH CONTENT-BASED FILTERING METHOD FOR CULINARY TOURISM IN," vol. 8, no. 2, pp. 65–72, 2019.
- [12] A. Sulami *et al.*, "PENERAPAN METODE CONTENT BASED FILTERING DALAM," vol. 9, no. 2, 2024.
- [13] V. Cyntia, B. Mulyawan, and M. Sitorus, "Mobile-Based Food Recommendation System Using Hybrid Filtering Methods," *Int. J. Appl. Sci. Technol. Eng.*, vol. 1, no. 4, pp. 1675–1685, 2023, [Online]. Available: <https://journal.untar.ac.id/index.php/ijaste/article/view/35043>
- [14] I. Rasyid, M. R. A. Yudianto, M. Maimunah, and T. A. Purnomo, "Electronic Product Recommendation System Using the Cosine Similarity Algorithm and VGG-16," *Sinkron*, vol. 8, no. 4, pp. 2120–2129, 2023, doi: 10.33395/sinkron.v8i4.12936.
- [15] F. A. Putra *et al.*, "IMPLEMENTASI METODE CONTENT-BASED FILTERING PADA," vol. 06, no. 04, pp. 742–751, 2025.