

IMPLEMENTASI TF-IDF DAN COSINE SIMILARITY UNTUK REKOMENDASI RESEP MASAKAN INDONESIA BERDASARKAN BAHAN MAKANAN

Jeovanni Wong, Yoannita

Informatika, Universitas Multi Data Palembang

Jl Rajawali No.14, 9 Ilir, Kec. Ilir Tim. II, Kota Palembang, Sumatera Selatan

jeovanniwong_2226250039@mhs.mdp.ac.id

ABSTRAK

Perkembangan teknologi informasi membuat ketersediaan resep kuliner semakin melimpah sehingga menimbulkan masalah *information overload* pada pengguna yang ingin mencari resep berdasarkan bahan yang tersedia. Kondisi tersebut menyebabkan proses pencarian menjadi kurang efisien karena pengguna harus menelusuri resep secara manual. Penelitian ini bertujuan mengembangkan sistem rekomendasi resep masakan Indonesia berbasis *content-based filtering* dengan memanfaatkan metode TF-IDF dan *Cosine Similarity* untuk mengukur tingkat kemiripan antara *input* bahan pengguna terhadap kumpulan resep dalam dataset. Proses penelitian dilakukan melalui *preprocessing* teks, normalisasi bahan, deduplikasi, pembentukan vektor TF-IDF, koreksi *typo* berbasis *fuzzy matching*, dan perhitungan kesamaan menggunakan *Cosine Similarity*. Pengujian menggunakan dataset *Indonesian Food Recipes* menunjukkan bahwa sistem mampu memberikan rekomendasi yang relevan dengan nilai *Precision@5* sebesar 0.8454 dan *Recall@5* sebesar 0.5804. Selain itu, integrasi fitur koreksi *typo* dan filter metode memasak terbukti meningkatkan akurasi hasil rekomendasi. Berdasarkan hasil tersebut, sistem ini dapat membantu pengguna menemukan resep secara lebih cepat dan relevan sesuai bahan yang tersedia.

Kata kunci : *content-based filtering, cosine similarity, resep masakan Indonesia, sistem rekomendasi, TF-IDF*

1. PENDAHULUAN

Perkembangan teknologi informasi dalam beberapa tahun terakhir telah mendorong peningkatan signifikan dalam ketersediaan informasi digital, termasuk pada bidang kuliner. Berbagai platform dan situs web kini menyediakan ribuan resep masakan yang dapat diakses dengan mudah oleh masyarakat. Meskipun memberikan kemudahan, jumlah informasi yang sangat besar tersebut menimbulkan masalah baru berupa *information overload*, yaitu kondisi ketika pengguna kesulitan menemukan resep yang sesuai dengan kebutuhan mereka [1]. Situasi ini membuat proses pencarian resep sering kali memakan waktu dan kurang efisien.

Di sisi lain, pencarian resep berdasarkan bahan yang tersedia menjadi kebutuhan penting bagi banyak pengguna, terutama mereka yang ingin memanfaatkan bahan makanan secara optimal. Namun, proses pencarian manual tidak mampu menyaring resep secara tepat berdasarkan kecocokan bahan. Oleh karena itu, dibutuhkan sistem yang mampu memberikan rekomendasi secara otomatis dan relevan. Penelitian–penelitian terbaru menunjukkan bahwa sistem rekomendasi berbasis konten (*content-based filtering*) merupakan pendekatan yang efektif untuk menyelesaikan masalah ini, khususnya ketika data berupa teks seperti daftar bahan masakan [2], [3].

Metode *Term Frequency–Inverse Document Frequency* (TF-IDF) dan *Cosine Similarity* banyak digunakan dalam penelitian sistem rekomendasi berbasis teks karena keduanya mampu mengukur tingkat kemiripan antar dokumen secara efektif. TF-IDF bekerja dengan memberikan bobot pada kata yang dianggap penting dalam dokumen, sementara *Cosine*

Similarity menghitung tingkat kedekatan antara dokumen berdasarkan orientasi vektornya [4]. Berbagai studi dalam lima tahun terakhir juga menunjukkan bahwa kombinasi kedua metode ini mampu menghasilkan rekomendasi yang relevan, baik dalam konteks resep masakan, destinasi wisata, maupun produk lainnya [5], [6].

Berdasarkan kondisi tersebut, penelitian ini berfokus pada penerapan metode TF-IDF dan *Cosine Similarity* untuk menghasilkan rekomendasi resep masakan Indonesia berdasarkan kesesuaian bahan yang dimasukkan oleh pengguna. Penelitian ini bertujuan untuk membantu pengguna menemukan resep yang relevan secara lebih cepat dan akurat, sekaligus mengurangi permasalahan *information overload* dalam pencarian resep. Untuk mencapai tujuan tersebut, penelitian ini memanfaatkan dataset resep masakan Indonesia yang tersedia secara publik, yang kemudian diproses melalui tahapan representasi teks menggunakan TF-IDF dan perhitungan tingkat kemiripan menggunakan *Cosine Similarity*. Melalui pendekatan ini, diharapkan sistem yang dikembangkan mampu memberikan rekomendasi resep yang efektif dan memberikan kontribusi terhadap pengembangan sistem temu kembali informasi berbasis teks dalam bidang kuliner.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Sistem Berbagai penelitian telah mengkaji penerapan *content-based filtering* dalam sistem rekomendasi berbasis teks dengan memanfaatkan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) dan *Cosine Similarity*. Penelitian

oleh V. K. Sari et al. [2] menunjukkan bahwa kombinasi TF-IDF dan *Cosine Similarity* mampu menghasilkan rekomendasi resep masakan yang relevan berdasarkan bahan dan preferensi pengguna. Hasil serupa juga diperoleh oleh Nallapati et al. [4] pada sistem rekomendasi masakan India berbasis bahan, yang mencapai tingkat akurasi tinggi pada dataset berskala besar.

Selanjutnya, Verma et al. [7] serta Jaiswal dan Rai [8] membuktikan bahwa metode TF-IDF dan *Cosine Similarity* efektif digunakan dalam sistem rekomendasi resep berbasis *natural language processing* dengan tingkat relevansi yang baik. Dalam konteks Indonesia, N. N. K. Sari et al. [3] mengimplementasikan pendekatan yang sama pada sistem rekomendasi resep berbasis web, namun penelitian tersebut masih terbatas pada jumlah dataset yang relatif kecil.

Selain pada bidang kuliner, TF-IDF dan *Cosine Similarity* juga banyak diterapkan pada bidang lain, seperti pencarian destinasi wisata [1] dan sistem rekomendasi lagu berbasis lirik [9], yang menunjukkan fleksibilitas metode ini dalam pengolahan data teks. Berdasarkan penelitian-penelitian tersebut, dapat disimpulkan bahwa TF-IDF dan *Cosine Similarity* merupakan metode yang efektif dan efisien untuk sistem rekomendasi berbasis teks. Oleh karena itu, penelitian ini mengimplementasikan kedua metode tersebut pada dataset resep masakan Indonesia berskala besar guna menghasilkan rekomendasi resep yang relevan berdasarkan bahan yang dimasukkan pengguna.

2.2. Sistem Rekomendasi

Sistem rekomendasi merupakan teknologi yang digunakan untuk membantu pengguna menemukan item yang relevan dari sejumlah besar data. Secara umum, terdapat tiga pendekatan utama, yaitu *content-based filtering*, *collaborative filtering*, dan *hybrid filtering* [10]. *Content-based filtering* bekerja dengan mencocokkan fitur dari item dalam hal ini daftar bahan masakan dengan preferensi atau kebutuhan pengguna. Pendekatan ini efektif digunakan apabila data yang tersedia merupakan data tekstual.

Perkembangan riset dalam lima tahun terakhir menunjukkan bahwa sistem rekomendasi berbasis teks banyak digunakan untuk berbagai bidang, seperti kuliner, *e-commerce*, pariwisata, dan kesehatan. Widiyanti et al. [11] mencatat bahwa metode *content-based* memiliki keunggulan dalam kebutuhan data yang minimal serta kemampuannya memberikan rekomendasi yang konsisten. Sementara itu, penelitian oleh Sulami et al. [6] menunjukkan bahwa sistem rekomendasi berbasis konten memberikan hasil yang akurat ketika data memiliki deskripsi fitur yang detail dan terstruktur.

2.3. Information Retrieval (IR)

Information Retrieval (IR) adalah bidang yang mempelajari cara menemukan dan menyajikan

informasi yang relevan berdasarkan permintaan pengguna. IR banyak diterapkan dalam pencarian dokumen, pencarian resep, klasifikasi teks, hingga sistem rekomendasi [12]. Komponen utama dalam IR meliputi *preprocessing* teks, ekstraksi fitur, pemeringkatan dokumen, dan evaluasi kinerja.

Model klasik seperti TF-IDF masih banyak digunakan sebagai *baseline* dalam penelitian IR karena kemampuannya memproses data secara efisien [13]. Dalam konteks pencarian resep masakan, IR berperan penting dalam menyeleksi resep berdasarkan kesesuaian bahan, sehingga pengguna tidak perlu melakukan pencarian secara manual.

2.4. Content-Based Filtering

Content-Based Filtering (CBF) merupakan pendekatan sistem rekomendasi yang bekerja dengan mencocokkan fitur suatu item terhadap preferensi atau kebutuhan pengguna sehingga rekomendasi yang dihasilkan bersifat relevan berdasarkan konten item tersebut. Berbeda dengan pendekatan *collaborative filtering* yang membutuhkan data *rating* atau interaksi dari banyak pengguna, metode CBF tetap dapat memberikan rekomendasi meskipun informasi pengguna terbatas karena proses pencarian dilakukan berdasarkan kesamaan konten. Oleh karena itu, pendekatan ini sangat sesuai ketika data berupa teks seperti daftar bahan masakan, karena setiap bahan dapat direpresentasikan sebagai fitur yang dapat dihitung tingkat kemiripannya.

Dalam penelitian ini, CBF digunakan untuk mencocokkan bahan masakan yang dimasukkan pengguna dengan bahan pada setiap resep. Proses komputasi dilakukan dengan pembobotan teks menggunakan TF-IDF dan perhitungan tingkat kesamaan menggunakan *Cosine Similarity*, sehingga sistem mampu menyeleksi resep dengan relevansi tertinggi terhadap *input* pengguna [14]. Selain itu, penelitian terdahulu menunjukkan bahwa CBF efektif digunakan dalam domain konten berbasis teks sehingga metode ini merupakan pendekatan yang tepat untuk sistem rekomendasi resep masakan berbasis bahan makanan [15].

2.5. Term Frequency – Inverse Document Frequency (TF-IDF)

TF-IDF adalah metode pembobotan istilah yang digunakan untuk menentukan tingkat kepentingan suatu kata dalam sebuah dokumen dibandingkan seluruh dokumen dalam korpus [16]. TF-IDF terdiri dari dua komponen perhitungan, yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF).

a. Term Frequency (TF)

TF mengukur seberapa sering suatu istilah t muncul dalam dokumen d . Nilainya dihitung menggunakan Persamaan (1).

$$TF(t, d) = \frac{f_{t,d}}{\sum_k f_{k,d}} \quad (1)$$

Keterangan :

$f_{t,d}$ = Jumlah kemunculan kata t pada dokumen d .
 $\sum_k f_{k,d}$ = Total seluruh kata pada dokumen tersebut.

b. *Inverse Document Frequency* (IDF)

IDF digunakan untuk mengukur tingkat kepentingan suatu kata dalam keseluruhan dokumen. Perhitungannya ditunjukkan pada Persamaan (2).

$$IDF(t) = \log \left(\frac{N}{1 + n_t} \right) \quad (2)$$

Keterangan :

N = Jumlah total dokumen

n_t = Jumlah dokumen yang mengandung kata t

c. Pembobotan TF-IDF

Nilai bobot akhir kata dihitung dengan mengalikan TF dan IDF seperti ditunjukkan pada Persamaan (3).

$$TF - IDF(t, d) = TF(t, d) \times IDF(t) \quad (3)$$

Metode TF-IDF terbukti efektif dalam mengekstraksi informasi penting dari teks, khususnya pada domain klasifikasi dokumen, pencarian informasi, dan sistem rekomendasi [17].

2.6. Cosine Similarity

Cosine Similarity adalah metode populer untuk menghitung tingkat kemiripan antar dua vektor teks. Keunggulannya adalah tidak terpengaruh oleh panjang dokumen dan mampu memberikan nilai kemiripan yang stabil [18]. Perhitungan *cosine similarity* dirumuskan dalam Persamaan (4).

$$\text{cosine_similarity}(a, b) = \frac{\sum_{i=1}^n a_i b_i}{\sqrt{\sum_{i=1}^n a_i^2} \sqrt{\sum_{i=1}^n b_i^2}} \quad (4)$$

Keterangan :

A dan B = Vektor representasi dokumen.

A_i, B_i = Bobot kata ke- i dalam dokumen.

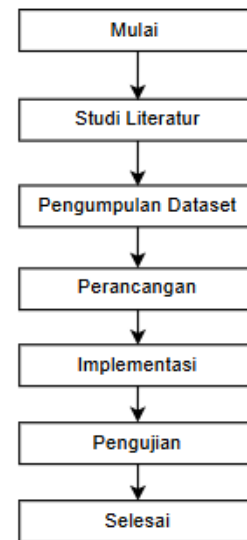
Penelitian oleh Nallapati et al. [4], Chomchaiya et al. [5], dan Rasyid & Ningsih [1] menunjukkan bahwa metode ini dapat meningkatkan kualitas rekomendasi berbasis teks secara signifikan.

2.7. Fuzzy Matching

Fuzzy Matching merupakan teknik pencocokan teks yang digunakan untuk mengukur tingkat kemiripan antar *string* meskipun terdapat perbedaan penulisan seperti *typo* atau variasi ejaan. Teknik ini biasanya menggunakan perhitungan jarak edit seperti *Levenshtein Distance* untuk menentukan seberapa dekat suatu kata dengan kata lain sehingga sistem dapat mengenali *input* yang tidak ditulis dengan benar

[19]. *Fuzzy Matching* banyak digunakan dalam sistem pencarian dan *information retrieval* karena mampu menangani data yang tidak terstruktur dan kesalahan *input* pengguna [20]. Beberapa penelitian juga menyatakan bahwa integrasi *Fuzzy Matching* dengan metode TF-IDF dapat meningkatkan akurasi pencarian teks karena toleran terhadap kesalahan ejaan sekaligus mampu menilai bobot penting suatu kata [21].

3. METODE PENELITIAN



Gambar 1. Tahapan metode penelitian

Metode penelitian yang digunakan dalam studi ini mengikuti alur tahapan yang ditunjukkan pada Gambar 1. Tahapan tersebut meliputi studi literatur, pengumpulan dataset, perancangan sistem, implementasi, dan pengujian. Seluruh tahapan dirancang untuk memastikan bahwa proses pembangunan sistem rekomendasi dilakukan secara terstruktur dan menghasilkan model yang valid serta dapat dievaluasi secara kuantitatif.

3.1. Studi Literatur

Tahap studi literatur dilakukan untuk mengkaji teori dan penelitian terdahulu yang berkaitan dengan sistem rekomendasi, *Information Retrieval*, pembobotan TF-IDF, serta perhitungan *Cosine Similarity*. Sumber referensi diperoleh dari jurnal ilmiah, repositori penelitian, *prosiding*, dan buku yang relevan. Studi literatur ini berfungsi sebagai dasar dalam penyusunan metode dan perancangan sistem.

3.2. Pengumpulan Dataset

Dataset yang digunakan merupakan data sekunder dari Kaggle, yaitu *Indonesian Food Recipes Dataset*, yang berisi 14.945 resep masakan Indonesia. Dataset mencakup judul resep, daftar bahan, langkah memasak, kategori, dan data hasil *pre-cleaning*. Dataset ini menjadi sumber data utama dalam pembentukan model rekomendasi.

3.3. Perancangan Sistem

Tahap perancangan sistem dilakukan untuk menyusun alur kerja model rekomendasi. Proses dimulai dengan *preprocessing* data. Pada tahap ini, data bahan dan tahapan cara memasak dibersihkan dari elemen yang tidak diperlukan, yaitu dengan proses tokenisasi (proses memecah teks menjadi unit-unit yang lebih kecil berupa kata), konversi seluruh teks menjadi huruf *lowercase*, penghapusan tanda baca (., ! ? ; () / \ + - *), serta penghilangan karakter khusus seperti simbol (@, #, %, &, +), angka yang tidak relevan, serta emoji untuk memastikan data berada dalam format yang seragam. Setelah itu, dilakukan normalisasi bahan untuk menyeragamkan istilah atau sinonim yang memiliki makna sama, sehingga variasi penulisan tidak mempengaruhi hasil perhitungan kemiripan. Data yang memiliki kemiripan tinggi atau duplikasi dihapus melalui proses deduplikasi. Tahap selanjutnya adalah pembentukan *vocabulary* sebagai kumpulan kata unik dari seluruh bahan masakan dalam dataset, yang kemudian digunakan sebagai dasar pembentukan representasi fitur. Selain itu, dilakukan ekstraksi metode memasak seperti menggoreng, merebus, atau mengukus untuk menambah konteks relevansi dalam proses rekomendasi. Seluruh data yang telah diproses kemudian direpresentasikan ke dalam bentuk vektor menggunakan metode TF-IDF, sehingga setiap resep memiliki nilai bobot yang dapat dibandingkan. Terakhir, dirancang modul *Cosine Similarity* untuk menghitung tingkat kemiripan antara *input* bahan dari pengguna dengan setiap resep dalam *database*. Perancangan ini memastikan alur sistem berjalan secara sistematis dan menghasilkan rekomendasi yang akurat serta relevan.

3.4. Implementasi

Tahap implementasi dilakukan untuk merealisasikan seluruh rancangan sistem yang telah disusun pada tahap sebelumnya. Pada tahap ini, data bahan masakan yang telah melalui proses *preprocessing*, normalisasi, dan deduplikasi diubah menjadi representasi numerik menggunakan metode TF-IDF dengan konfigurasi n-gram 1–3. Representasi ini kemudian digunakan sebagai dasar perhitungan kemiripan antar resep menggunakan *Cosine Similarity*, sehingga resep dengan tingkat kesamaan tertinggi dapat direkomendasikan kepada pengguna. Selain itu, modul koreksi *typo* berbasis *fuzzy matching* diintegrasikan untuk memastikan sistem tetap dapat memberikan rekomendasi meskipun pengguna melakukan kesalahan pengetikan pada *input* bahan. Implementasi juga mencakup penambahan fitur filter metode memasak agar pengguna dapat menyesuaikan rekomendasi berdasarkan preferensi teknik memasak tertentu. Tahap ini diakhiri dengan pembangunan antarmuka *input-output* yang memungkinkan

pengguna memasukkan daftar bahan masakan dan memperoleh hasil rekomendasi secara langsung dalam bentuk daftar resep dengan skor kemiripan tertinggi.

3.5. Pengujian

Tahap pengujian dilakukan untuk menilai performa dan keandalan sistem rekomendasi yang telah dibangun. Dataset dibagi menjadi data latih dan data uji dengan proporsi 80:20 untuk memastikan model dapat dievaluasi secara objektif. Selanjutnya, beberapa metrik evaluasi digunakan untuk mengukur kualitas rekomendasi, yaitu *Precision@K*, *Recall@K*, *F1-Score@K*, *Accuracy*, serta nilai rata-rata *Cosine Similarity*. Pengujian dilakukan pada berbagai skenario, meliputi pencarian resep berdasarkan satu bahan, kombinasi beberapa bahan, dan *input* bahan yang mengandung kesalahan penulisan. Melalui skenario tersebut, sistem diuji untuk melihat tingkat ketepatan dan relevansi rekomendasi yang dihasilkan pada kondisi yang berbeda-beda. Hasil pengujian menunjukkan bahwa sistem mampu memberikan rekomendasi yang sesuai dengan bahan yang dimasukkan pengguna serta toleran terhadap variasi *input*, termasuk ketika terjadi *typo* pada kata tertentu.

4. HASIL DAN PEMBAHASAN

```
Mounted at /content/drive
Membaca file dari: '/content/drive/MyDrive/PTA Jao/dataset/Indonesian_Food_Recipes.xlsx' ...

Memproses data...
✓ Metode masak diekstrak dari kolom 'Steps'
✓ Kategori bahan ditambahkan ke kolom 'Ingredient Categories'

✓ Preprocessing selesai
Menghapus duplikat: 4 baris dihapus. Ukuran: 14945 -> 14941

Total data: 14941

✓ Membangun kamus bahan...
✓ Kamus bahan dibangun: 3061 kata unik
```

Gambar 1. Proses *preprocessing* data resep

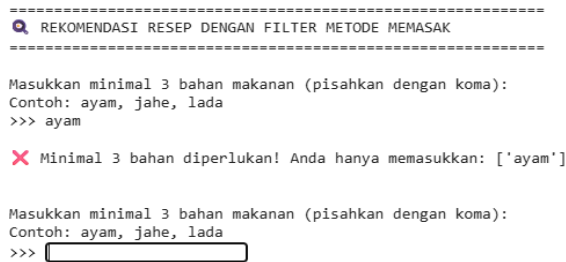
Gambar 1 menunjukkan proses *preprocessing* data resep yang meliputi pembersihan teks, normalisasi istilah bahan, penghapusan karakter khusus, serta deduplikasi data. Tahapan ini bertujuan untuk menghasilkan data yang konsisten sebelum dilakukan pembobotan TF-IDF dan perhitungan *Cosine Similarity*.

```
=====
REKOMENDASI RESEP DENGAN FILTER METODE MEMASAK
=====

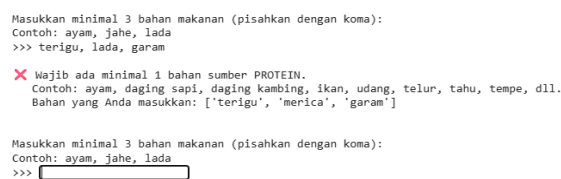
Masukkan minimal 3 bahan makanan (pisahkan dengan koma):
Contoh: ayam, jahe, lada
>>> 
```

Gambar 2. Tampilan *input* bahan pada sistem rekomendasi resep

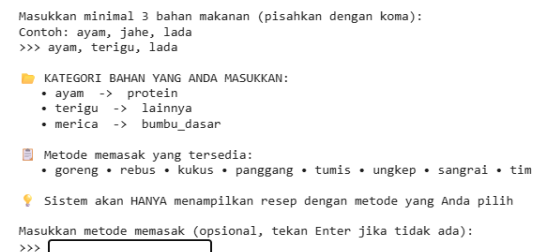
Gambar 2. menunjukkan tampilan *input* bahan makanan yang digunakan pengguna sebagai masukan awal dalam sistem rekomendasi resep.

Gambar 3. Kesalahan *input* bahan

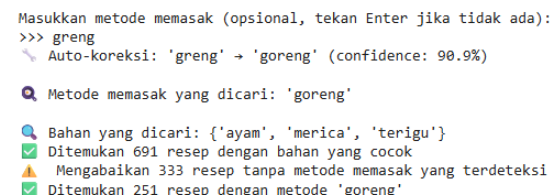
Gambar 3 menampilkan proses *input* bahan oleh pengguna, di mana pengguna diwajibkan memasukkan minimal tiga bahan sebagai syarat pencarian resep. Validasi dilakukan untuk memastikan input sesuai dengan ketentuan sistem.

Gambar 4. Kesalahan *input* bahan apabila *input* tidak memiliki protein

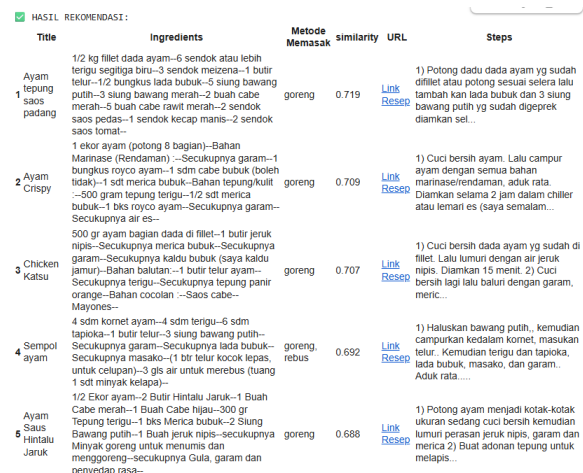
Gambar 4 memperlihatkan kondisi ketika *input* pengguna tidak mengandung bahan sumber protein. Sistem memberikan peringatan karena keberadaan minimal satu bahan protein menjadi syarat utama agar rekomendasi yang dihasilkan lebih relevan.

Gambar 5. *Input* resep pengguna

Gambar 5 menunjukkan proses validasi bahan dan pemilihan metode memasak sebagai tahap awal sebelum sistem menampilkan rekomendasi resep.

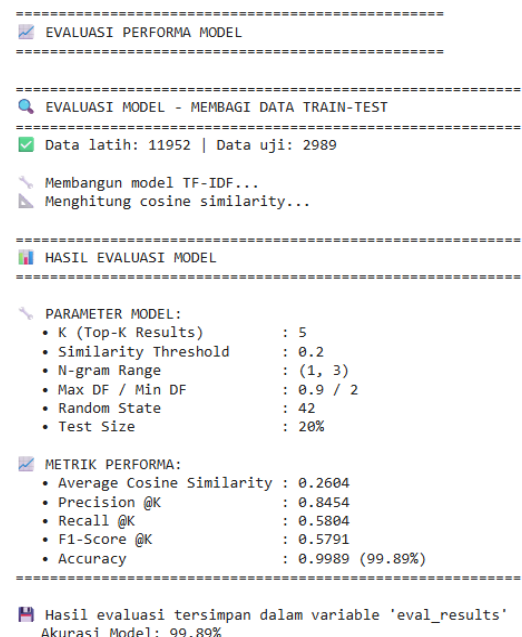
Gambar 6. Proses koreksi *typo* menggunakan *fuzzy matching*

Pada Gambar 6. memperlihatkan proses penentuan dan pemfilteran metode memasak menggunakan *fuzzy matching* untuk meningkatkan relevansi hasil rekomendasi resep.



Gambar 7. Tampilan hasil rekomendasi resep

Hasil akhir rekomendasi resep ditampilkan pada Gambar 7, di mana resep disusun berdasarkan nilai kemiripan tertinggi yang diperoleh dari perhitungan *Cosine Similarity* terhadap vektor TF-IDF.



Gambar 8. Hasil evaluasi sistem penelitian

Evaluasi performa sistem ditunjukkan pada Gambar 8 menggunakan metrik *Precision*, *Recall*, *F1-Score*, *Accuracy*, dan rata-rata *Cosine Similarity*. Hasil evaluasi menunjukkan bahwa sistem mampu memberikan rekomendasi resep dengan tingkat ketepatan yang baik.

Pada tahap *preprocessing*, dilakukan proses deduplikasi untuk menghilangkan data resep yang memiliki kesamaan tinggi pada judul dan daftar bahan setelah melalui proses normalisasi teks. Proses deduplikasi ini bertujuan untuk mencegah terjadinya bias pada tahap pembobotan TF-IDF dan perhitungan nilai kemiripan akibat adanya resep yang identik atau sangat mirip. Ringkasan hasil proses deduplikasi ditunjukkan pada Tabel 1.

Tabel 1. Hasil deduplikasi data

Keterangan Data	Jumlah Resep
Jumlah data awal	14.945
Data duplikat terdeteksi	4
Data setelah deduplikasi	14.941

Berdasarkan Tabel 1, terlihat bahwa proses deduplikasi berhasil mengurangi jumlah data yang berulang dalam dataset. Dengan dataset yang telah terbebas dari duplikasi, sistem rekomendasi dapat melakukan proses ekstraksi fitur dan perhitungan kemiripan secara lebih akurat, sehingga hasil rekomendasi menjadi lebih representatif terhadap masukan pengguna.

Setelah proses deduplikasi, sistem melakukan koreksi kesalahan penulisan bahan menggunakan metode *fuzzy matching*. Metode ini digunakan untuk menangani variasi penulisan bahan yang tidak konsisten dari pengguna. Hasil pengujian koreksi *typo* ditunjukkan pada Tabel 2.

Tabel 2. Hasil koreksi *typo* menggunakan *fuzzy matching*

Input Pengguna	Hasil Koreksi	Confidence (%)	Status
Aym	Ayam	85.71	Berhasil dikoreksi
Bwang	Bawang	90.91	Berhasil dikoreksi
Mericaa	Merica	92.31	Berhasil dikoreksi
Kkecap	Kecap	90.91	Berhasil dikoreksi
Gaaram	Garam	90.91	Berhasil dikoreksi

Berdasarkan data pada Tabel 2, metode *fuzzy matching* mampu memperbaiki variasi kesalahan penulisan bahan. Misalnya, beberapa *input* seperti “Aym” atau “Mericaa” berhasil dikoreksi menjadi “Ayam” dan “Merica” dengan tingkat *confidence* di atas 85%. Hal ini menunjukkan bahwa *fuzzy matching* dapat menangani variasi *input* yang tidak konsisten sehingga sistem tetap mampu menafsirkan bahan yang dimaksud pengguna.

Selanjutnya, sistem menyediakan fitur *filter* metode memasak untuk menghasilkan rekomendasi yang lebih sesuai dengan preferensi pengguna. Hasil pengujian terhadap *filter* dengan metode memasak “goreng” ditunjukkan pada Tabel 3.

Tabel 3. Contoh hasil *filter* “goreng”

Judul Resep	Metode Memasak	Status Filter
Ayam Goreng Lengkuas	goreng	Dipertahankan
Tempe Goreng Tepung	goreng	Dipertahankan
Tumis Buncis Pedas	tumis	Dikecualikan
Sup Ayam Jahe	rebus	Dikecualikan
Ikan Kukus Serai	kukus	Dikecualikan

Data pada Tabel 3 menunjukkan bahwa proses *filter* mampu mempertahankan resep dengan metode

memasak goreng sekaligus mengecualikan resep yang tidak sesuai dengan kriteria filter pengguna. Metode memasak pada setiap resep diperoleh melalui proses ekstraksi dari langkah-langkah memasak (*steps*) dengan mengidentifikasi kata kunci dan pola kalimat yang menunjukkan teknik pengolahan utama, seperti “digoreng” atau “goreng hingga matang”. Pada pengujian menggunakan filter “goreng”, sistem dapat menyeleksi resep yang benar-benar menggunakan metode tersebut. Resep seperti “Ayam Goreng Lengkuas” dan “Tempe Goreng Tepung” dipertahankan, sedangkan resep dengan metode tumis, rebus, atau kukus dikeluarkan dari hasil rekomendasi. Dengan demikian, proses seleksi resep tidak hanya berdasarkan kecocokan bahan, tetapi juga preferensi teknik memasak yang diinginkan.

Pada tahap berikutnya, tingkat kemiripan dihitung menggunakan metode *Cosine Similarity* untuk menentukan urutan rekomendasi berdasarkan nilai *similarity* tertinggi. Hasil rekomendasi ditunjukkan pada Tabel 4.

Tabel 4. Lima rekomendasi teratas berdasarkan *cosine similarity*

Judul Resep	Metode Memasak	Nilai Similarity
Ayam Geprek Pedas	Goreng	0.708
Chicken Katsu	Goreng	0.687
Chicken Crispy Wings Saus Barbeque Pedas	Goreng, rebus	0.679
Ayam Goreng Korea (Korean Fried Chicken)	Goreng	0.673
Ayam Geprek	Goreng	0.654

Hasil pada Tabel 4 memperlihatkan bahwa masakan berbasis ayam dengan teknik pemasakan goreng memiliki tingkat *similarity* tertinggi terhadap *input* pengguna. Hal ini mengindikasikan bahwa sistem mampu mengidentifikasi kombinasi bahan yang beririsan secara signifikan terhadap masukan pengguna.

Evaluasi performa sistem dilakukan menggunakan beberapa metrik sebagai alat ukur keberhasilan rekomendasi. Hasil evaluasi performa ditunjukkan pada Tabel 5.

Tabel 5. Hasil evaluasi performa model

Metrik Evaluasi	Nilai
Average Cosine Similarity	0.2604
Precision@5	0.8454
Recall@5	0.5804
F1-Score	0.5791
Accuracy	99.89%

Berdasarkan Tabel 5, dapat diketahui bahwa nilai *Precision@5* yang mencapai 0.8454 mengindikasikan bahwa sebagian besar rekomendasi yang diberikan relevan terhadap *input* pengguna, sedangkan nilai *Recall@5* sebesar 0.5804 menunjukkan bahwa sistem telah mampu menyaring sebagian besar resep yang relevan meskipun masih terdapat kemungkinan

peningkatan cakupan rekomendasi. Meskipun nilai *Average Cosine Similarity* relatif rendah akibat tingginya variasi bahan dalam dataset, nilai *F1-score* dan akurasi menunjukkan bahwa model memiliki performa yang baik dalam memberikan rekomendasi resep masakan secara otomatis.

Secara keseluruhan, hasil penelitian menunjukkan bahwa sistem rekomendasi yang dikembangkan mampu menghasilkan rekomendasi relevan dengan tingkat *precision* tinggi, didukung kemampuan koreksi *typo* dan *filter* metode memasak yang bekerja secara efektif. Meskipun rata-rata *similarity* masih dapat ditingkatkan, performa sistem secara umum sudah sangat baik dan dapat diterapkan dalam skenario penggunaan nyata.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengimplementasikan metode TF-IDF dan *Cosine Similarity* pada sistem rekomendasi resep masakan Indonesia berbasis *content-based filtering* yang mampu memberikan rekomendasi secara relevan berdasarkan kecocokan bahan yang dimasukkan pengguna. Hasil evaluasi menunjukkan bahwa sistem memperoleh nilai *Precision@5* sebesar 0,8454 dan *Recall@5* sebesar 0,5804, serta didukung fitur *fuzzy matching* dan *filter* metode memasak yang terbukti meningkatkan ketepatan hasil rekomendasi sehingga dapat membantu mengatasi permasalahan *information overload* pada pencarian resep. Meskipun demikian, rata-rata nilai *similarity* masih dipengaruhi tingginya variasi bahan pada dataset sehingga pada penelitian berikutnya disarankan untuk menerapkan teknik representasi teks yang lebih baik, memperluas struktur data bahan, serta menambahkan fitur preferensi pengguna agar kualitas rekomendasi semakin personal, dan akurat.

DAFTAR PUSTAKA

- [1] R. Al Rasyid and D. H. U. Ningsih, "Penerapan algoritma TF-IDF dan cosine similarity untuk query pencarian pada dataset destinasi wisata," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 8, no. 1, p. 2024, 2024, doi: 10.35870/jti.
- [2] V. K. Sari, D. Hartanti, and E. Purwanto, "Sistem rekomendasi resep masakan menggunakan metode content based filtering berdasarkan preferensi pengguna," *Infotek: Jurnal Informatika dan Teknologi*, vol. 8, no. 2, pp. 345–355, 2025, doi: 10.29408/jit.v8i2.30625.
- [3] N. N. K. Sari, R. Priskila, and P. A. A. P. Bagus, "Implementasi content-based filtering menggunakan TF-IDF and cosine similarity untuk sistem rekomendasi resep masakan," *Jurnal Teknologi Informasi*, vol. 18, no. 1, pp. 43–51, 2024, doi: 10.47111/JTI.
- [4] D. Nallapati, S. R. Thatiparthi, A. Aitha, L. Chinthala, and S. Lekhi, "Ingredients-based indian cuisine recommendation system," in *Proceedings of the KILBY 100 7th International Conference on Computing Sciences 2023 (ICCS 2023)*, 2023. [Online]. Available: <https://ssrn.com/abstract=4485294>
- [5] S. Chomchaiya, B. Yoosooka, C. Rodmee, D. Naowaratphunsina, and W. Kaichid, "Web-based cooking recipe recommender system based on stocked groceries," *Progress in Applied Science and Technology*, vol. 14, no. 1, pp. 11–19, 2024, doi: 10.60101/past.2024.251684.
- [6] A. Sulami, V. Atina, and Nuralitasari, "Penerapan metode content based filtering dalam sistem rekomendasi pemilihan produk skincare," *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 9, no. 2, pp. 172–181, 2024.
- [7] V. P. Verma, S. Kashyap, and F. Senger, "Recipes recommendation system using machine learning," *Int J Res Appl Sci Eng Technol*, vol. 13, no. 5, pp. 3583–3587, 2025, doi: 10.22214/ijraset.2025.71001.
- [8] K. Jaiswal and R. Rai, "Time crunch cuisine-Express recipe recommendation system," *International Journal of Advances in Engineering and Management (IJAEM)*, vol. 6, no. 01, pp. 41–45, 2024, doi: 10.35629/5252-06014145.
- [9] A. Y. Timur and A. N. Rohman, "Sistem rekomendasi lagu indonesia menggunakan metode content-based filtering dan cosine similarity," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 1, 2025, doi: 10.23960/jitet.v13i1.5949.
- [10] H. Zhou, F. Xiong, and H. Chen, "A comprehensive survey of recommender systems based on deep learning," *Applied Sciences*, vol. 13, no. 20, 2023, doi: 10.3390/app132011378.
- [11] N. P. T. Widiyanti, I. M. A. D. Suarjaya, and N. K. D. Rusjyanthi, "Food recipe recommendation system with content-based filtering and collaborative filtering methods," *Sinkron: Jurnal dan Penelitian Teknik Informatika*, vol. 9, no. 3, pp. 1167–1176, 2025, doi: 10.33395/sinkron.v9i3.14778.
- [12] W. Hersh, "Search still matters: Information retrieval in the era of generative AI," *Journal of the American Medical Informatics Association*, vol. 31, no. 9, pp. 2159–2161, 2024, doi: 10.1093/jamia/ocae014.
- [13] Y. Zhu et al., "Large language models for information retrieval: A survey," *ACM Trans Inf Syst*, 2025, doi: 10.1145/3748304.
- [14] W. F. A. M. Irwan, B. Santoso, and S. Kacung, "Implementasi metode content based filtering menggunakan synopsis similarity untuk pemilihan anime," *Jurnal Informatika Teknologi dan Sains*, vol. 7, no. 2, pp. 947–955, 2025, Accessed: Dec. 10, 2025. [Online]. Available: <https://jurnal.uts.ac.id/index.php/JINTEKS/article/view/5815/2720>
- [15] A. Pranata and F. Sulianta, "Implementasi dan Analisis Algoritma Content-based filtering pada

- sistem rekomendasi produk tas pada basis data mysql,” *Jurnal Ilmiah Global Education*, vol. 6, no. 3, pp. 1419–1444, 2025, doi: 10.55681/jige.v6i3.4017.
- [16] M. Das, S. Kamalanathan, and P. Alphonse, “A comparative study on TF-IDF feature weighting method and its analysis using unstructured dataset,” in *COLINS-2021: 5th International Conference on Computational Linguistics and Intelligent Systems*, 2021. doi: 10.48550/arXiv.2308.04037.
- [17] P. Guleria, J. Frnda, and P. N. Srinivasu, “NLP based text classification using TF-IDF enabled fine-tuned long short-term memory: An empirical analysis,” *Array*, vol. 27, 2025, doi: 10.1016/j.array.2025.100467.
- [18] P. Jiang and X. Cai, “A survey of text-matching techniques,” *Information*, vol. 15, no. 6, 2024, doi: 10.3390/info15060332.
- [19] N. B. Siahaan and M. Alda, “Implementasi rule-based chatbot dan fuzzy string matching dalam sistem informasi layanan haji dan umrah di pt alfata wafiqah wisata,” *RABIT : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 10, no. 2, pp. 991–1008, 2025, doi: 10.36341/rabit.v10i2.6413.
- [20] L. Q. Zian, N. Z. Zulkarnain, and Y. J. Kumar, “Fuzzy matching: An alternative technique for merging extracted web data,” *Journal of Advanced Computing Technology and Application*, vol. 6, no. 1, pp. 1–13, 2024, Accessed: Dec. 10, 2025. [Online]. Available: <https://jacta.utem.edu.my/jacta/article/view/5289>
- [21] M. Fauzan Azima, A. Nur Listanto, Fitria, and Chairani, “Kombinasi algoritma TF-IDF dan fuzzy matching untuk deteksi kemiripan judul skripsi,” *JURNAL TEKNIKA*, vol. 19, no. 1, pp. 1–11, 2024, Accessed: Dec. 10, 2025. [Online]. Available: <https://jurnal.polsri.ac.id/index.php/teknika/article/view/8871>
- [22] A. Nathaniel, “Indonesian Food Recipes Dataset,” Kaggle, 2024. [Online]. Available: <https://www.kaggle.com/datasets/albertnathaniel12/food-recipes-dataset>