

LAPORAN PROYEK
SISTEM TEMU KEMBALI INFORMASI

APLIKASI REKOMENDASI MAKANAN LUAR NEGERI



Disusun Oleh:
Ijlal Fachry Attallah Handoko (A11.2023.15170)

TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS DIAN NUSWANTORO
SEMARANG

2025

PENDAHULUAN

Perkembangan teknologi informasi telah memberikan kemudahan bagi pengguna dalam mencari informasi, termasuk dalam bidang kuliner. Banyaknya ragam makanan dari berbagai negara membuat pengguna kesulitan mencari makanan yang sesuai dengan selera atau kebutuhan mereka. Oleh karena itu, diperlukan sebuah sistem informasi berbasis *search engine* yang mampu memberikan rekomendasi makanan asing secara cepat, relevan, dan akurat berdasarkan deskripsi atau kata kunci yang diberikan pengguna.

Proyek ini berfokus pada pembangunan sebuah Aplikasi Rekomendasi Makanan Luar Negeri menggunakan pendekatan *Information Retrieval* (IR) klasik, yaitu Boolean Retrieval Model dan Vector Space Model (VSM). Sistem memanfaatkan proses *text preprocessing*, pembobotan TF-IDF, serta perhitungan Cosine Similarity untuk menentukan tingkat relevansi antara query dan dokumen.

Ruang lingkup proyek meliputi: pemrosesan dataset makanan internasional, pengembangan modul preprocessing, implementasi model pencarian Boolean dan VSM, pembangunan arsitektur search engine sederhana, hingga evaluasi performa sistem. Proyek ini memberikan kontribusi langsung terhadap capaian pembelajaran (Sub-CPMK), antara lain:

1. memahami konsep IR klasik
2. mengimplementasikan proses preprocessing teks
3. menerapkan model pencarian Boolean & VSM
4. melakukan evaluasi dan analisis performa sistem pencarian.

Dengan demikian, aplikasi ini tidak hanya membantu pengguna menemukan makanan asing sesuai preferensi, tetapi juga memberikan pemahaman praktis mengenai implementasi IR dalam konteks dunia nyata.

DATA & PREPROCESSING

Dataset yang digunakan dalam proyek ini berisi informasi mengenai berbagai jenis makanan luar negeri, mencakup:

- Food_ID
- Name (Nama makanan)
- C_Type (Kategori makanan)
- Veg_Non (Vegetarian / Non-vegetarian)
- Describe (Deskripsi bahan yang digunakan)
- User_ID
- Rating

Dataset asli memiliki beragam permasalahan, seperti inkonsistensi format teks, huruf besar–kecil, tanda baca tidak rapi, hingga teks bercampur bahasa. Oleh karena itu, dilakukan tahapan preprocessing berikut:

Tahapan Preprocessing:

1. **Lowercasing** – seluruh teks diubah menjadi huruf kecil.
2. **Cleaning** – menghapus angka, tanda baca berlebih, karakter khusus, dan *whitespace*.
3. **Tokenization** – memecah kalimat menjadi token.
4. **Stopword Removal** – menghapus kata umum yang tidak membawa makna penting.
5. **Stemming / Lemmatization** – menyederhanakan kata ke bentuk dasar.
6. **Normalisasi teks** – memperbaiki kata yang rusak (typo minor) tanpa mengubah kata penting.

Sebelum:

“A sweet-and-sour *Japanese* dish, made with CHICKEN and served with rice!!”

Sesudah:

“sweet sour japan dish make chicken serve rice”

Hasil akhir preprocessing kemudian digabungkan menjadi kolom baru bernama **text**, yang digunakan sebagai dokumen input untuk model IR.

METODE INFORMATION RETRIEVAL

A. Boolean Retrieval Model

Boolean IR memandang dokumen sebagai himpunan kata. Pencarian dilakukan menggunakan logika *AND*, *OR*, dan *NOT*.

Contoh:

Query: “*spicy AND korea*”

- menunjukkan semua dokumen yang mengandung kedua kata tersebut.

Boolean retrieval bekerja cepat tetapi memiliki kelemahan: sensitivitas tinggi terhadap keberadaan kata, tanpa mempertimbangkan tingkat relevansi.

B. Vector Space Model (VSM)

Model VSM merepresentasikan dokumen sebagai vektor numerik berdasarkan frekuensi kata. Setiap kata dihitung bobotnya menggunakan **TF-IDF**.

Term Frequency (TF)

$$TF(t, d) = \frac{\text{jumlah kemunculan } t \text{ di } d}{\text{total kata di } d}$$

Inverse Document Frequency (IDF)

$$IDF(t) = \log \left(\frac{N}{df(t)} \right)$$

TF-IDF

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

Setelah dokumen dan query dikonversi menjadi vektor, dihitung **Cosine Similarity** untuk menentukan tingkat kesamaan:

$$\cos(\theta) = \frac{\vec{d} \cdot \vec{q}}{\|\vec{d}\| \|\vec{q}\|}$$

Semakin tinggi nilainya, semakin relevan dokumen tersebut terhadap query. Model VSM lebih fleksibel dibanding Boolean dan mampu memberikan ranking hasil pencarian.

ARSITEKTUR SEARCH ENGINE

1. Data Loader

Memuat dataset makanan dan mempersiapkan struktur penyimpanan dokumen.
2. Preprocessor Module

Melakukan seluruh tahapan preprocessing dan menghasilkan “clean text”.
3. Indexing Module
 - Untuk Boolean: membuat inverted index sederhana.
 - Untuk VSM: menghitung dan menyimpan TF-IDF matrix.
4. Query Processor

Membersihkan query menggunakan langkah preprocessing yang sama.
5. Retrieval Engine
 - BooleanIR: mencocokkan kata berdasarkan logika boolean.
 - VSM: menghitung cosine similarity dan mengembalikan ranking dokumen terbaik.
6. Result Filter

Opsional: mengurutkan berdasarkan **rating** untuk rekomendasi makanan terbaik.

EKSPERIMEN & EVALUASI

Evaluasi dilakukan untuk menguji kinerja model Boolean dan VSM dalam skenario pencarian makanan asing berdasarkan kata kunci.

Skenario Uji

- Query 1: "spicy korean noodle"
- Query 2: "sweet japanese dessert"
- Query 3: "fried seafood thailand"

(Setiap query diuji menggunakan kedua model.)

Metode Evaluasi

Model dianalisis menggunakan:

1. Precision
2. Recall
3. F1-Score
4. Relevansi subjektif berdasarkan interpretasi manual deskripsi makanan.
5. Kecepatan pencarian (runtime ringan).

Hasil Ringkas

Model	Kelebihan	Kekurangan	Nilai Rata-rata Evaluasi
Boolean IR	Cepat, sederhana	Tidak mempertimbangkan ranking	Precision tinggi, recall rendah
VSM (TF-IDF)	Ranking relevan, akurat	Lebih berat komputasi	Precision & recall seimbang, hasil terbaik

Berdasarkan analisis, VSM memberikan hasil paling relevan untuk rekomendasi makanan karena mampu mengukur kedekatan semantik query–dokumen.

DISKUSI

Proyek ini berhasil menunjukkan bahwa metode IR klasik masih sangat efektif digunakan untuk aplikasi sederhana seperti pencarian rekomendasi makanan luar negeri. Boolean IR cocok untuk pencarian langsung berbasis kata kunci eksak, sementara VSM unggul dalam memberikan ranking relevansi yang lebih konsisten dengan kebutuhan pengguna.

1. Kelebihan Sistem

- Arsitektur sederhana dan mudah dikembangkan.
- Menggunakan model IR yang transparan dan dapat dijelaskan.
- Hasil pencarian relevan terutama menggunakan VSM.
- Dapat diintegrasikan dengan rating untuk sistem rekomendasi hybrid.

2. Keterbatasan Sistem

- Tidak menggunakan model embedding modern seperti Word2Vec/BERT.
- Preprocessing masih berbasis aturan, belum menangani sinonim kompleks.
- Dataset terbatas sehingga variasi makanan masih kurang.

3. Saran Pengembangan

- Mengintegrasikan semantic search berbasis embedding.
- Menambahkan rekomendasi personalisasi berdasarkan preferensi pengguna.
- Memperluas dataset dari API kuliner internasional.
- Menerapkan model machine learning untuk ranking optimization.

Dengan pengembangan lebih lanjut, aplikasi ini berpotensi menjadi platform pencarian kuliner internasional yang sangat membantu pengguna dalam eksplorasi makanan dari seluruh dunia.