

**PENGEMBANGAN SISTEM PENCARIAN RESEP MASAKAN
INDONESIA MENGGUNAKAN METODE BM25 DAN FASTTEXT**



Disusun Oleh :

NIM : A11.2022.13980
Nama : Annisa Himatul Chasanah
Kelompok : A11.4701

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS DIAN NUSWANTORO**

2024

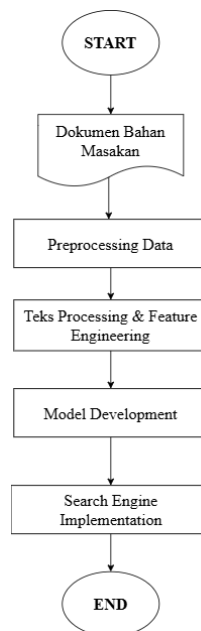
RINGKASAN DAN PERMASALAHAN PROJECT

Project ini bertujuan untuk mengembangkan sebuah search engine yang dapat membantu pengguna menemukan resep masakan Indonesia berdasarkan bahan yang pengguna inginkan. Permasalahan yang dihadapi adalah bagaimana menyajikan hasil pencarian yang relevan dan akurat dari berbagai resep yang ada.

Tujuan yang akan dicapai:

- Mengembangkan sistem pencarian resep yang akurat dan responsif.
- Mengimplementasikan metode BM25 untuk meningkatkan relevansi hasil pencarian dan peringkat dokumen.
- Menggunakan FastText untuk pemahaman semantik yang lebih baik.
- Teknik pemrosesan bahasa alami (NLP) untuk membersihkan dan menyiapkan data resep.
- Menyajikan informasi resep secara terstruktur dan mudah dipahami.

Alur penyelesaian sebagai berikut :



1. Mulai

Dokumen bahan masakan dimuat untuk diproses.

2. Preprocessing Data

- Data Cleaning, membersihkan data dengan langkah-langkah seperti: Menghapus kolom tertentu (contoh: URL) dan menghapus baris yang mengandung nilai kosong (NA).
 - Text Normalization, Mengubah teks menjadi format yang konsisten (lowercase).
3. Text Processing & Feature Engineering
- Tokenization, memisahkan teks menjadi kata-kata atau unit kecil.
 - Stop Words Removal, menghapus kata-kata umum yang tidak relevan (contoh: "dan", "atau").
 - Punctuation Removal, menghapus tanda baca.
 - Feature Extraction, membuat representasi fitur dari teks (contoh: menghitung jumlah kata, frekuensi istilah).
4. Model Development
- FastText Model Training, melatih model berbasis FastText untuk representasi teks.
 - BM25 Implementation, menggunakan algoritma BM25 untuk pencocokan query dan dokumen.
 - Recipe Embedding, membuat embedding (representasi vektor) untuk setiap resep.
5. Search Engine Implementation
- User Input, pengguna memasukkan query atau kata kunci pencarian.
 - Query Processing, memproses query agar sesuai dengan data (contoh: normalisasi teks).
 - Recipe Ranking, mengurutkan hasil berdasarkan kesesuaian dengan query.
 - Result Display, menampilkan hasil pencarian kepada pengguna.
6. End
- Proses selesai setelah hasil pencarian ditampilkan.

PENJELASAN DATASET, EDA, dan PROSES FEATURES DATASET

Dataset yang digunakan terdiri dari beberapa file CSV yang berisi resep masakan dari berbagai bahan seperti ayam, kambing, telur, dan udang. Setiap resep memiliki informasi seperti Title (judul), Ingredients (bahan-bahan), Steps (langkah-langkah pembuatan), dan Loves (jumlah likes). Dalam proses EDA, dilakukan analisis distribusi likes dan jumlah resep per kategori.

Proses feature engineering meliputi:

- Data diimpor dari beberapa file CSV, dan kolom yang tidak relevan seperti URL dihapus.
- Penghapusan kolom yang tidak relevan seperti URL dan penanganan nilai yang hilang (NaN).
- Fitur yang dihasilkan dari dataset meliputi kolom 'clean_ingredients' dan 'clean_instructions', yang merupakan hasil dari pemrosesan teks untuk menghapus stopwords, angka, dan karakter yang tidak relevan.
- Selain itu, frekuensi bahan juga dihitung untuk mendapatkan wawasan tentang bahan yang paling umum dan jarang digunakan.
- Penyusunan Vektor Kata
Menggunakan FastText untuk membangun vektor kata dari setiap instruksi dan bahan-bahan untuk pemodelan lebih lanjut.

PROSES LEARNING / MODELLING

Model yang digunakan dalam proyek ini adalah BM25 untuk pencarian dan FastText untuk menghasilkan embedding dari instruksi resep. BM25 digunakan untuk menghitung relevansi resep berdasarkan kata kunci yang dimasukkan oleh pengguna. Sementara itu, FastText dilatih menggunakan instruksi resep yang telah dibersihkan untuk menghasilkan representasi vektor dari setiap resep. Proses pelatihan melibatkan penggabungan semua instruksi menjadi satu list per resep dan kemudian melatih model FastText untuk mendapatkan embedding yang dapat digunakan dalam pencarian resep yang mirip.

PERFORMA MODEL

Evaluasi sistem menggunakan beberapa metrik:

- Mean Reciprocal Rank (MRR) untuk mengukur kualitas ranking
- Top-K Accuracy untuk mengukur presisi rekomendasi
- NDCG untuk mengukur kualitas pengurutan hasil
- Precision, Recall, dan F1-Score untuk mengukur akurasi sistem

DISKUSI HASIL DAN KESIMPULAN

Proyek ini berhasil mengembangkan sebuah search engine untuk resep masakan Indonesia yang dapat membantu pengguna menemukan resep berdasarkan bahan yang mereka inginkan. Dengan menggunakan teknik pemrosesan bahasa alami dan model BM25, sistem ini mampu memberikan rekomendasi resep yang relevan dan akurat. Hasil evaluasi menunjukkan bahwa model memiliki performa yang baik dalam hal relevansi hasil pencarian, dengan metrik yang menunjukkan bahwa banyak resep yang relevan dapat ditemukan dalam hasil pencarian teratas.

Selain itu, penggunaan FastText untuk menghasilkan embedding resep memberikan keuntungan dalam menemukan resep yang mirip, sehingga meningkatkan pengalaman pengguna. Secara keseluruhan, proyek ini menunjukkan potensi besar dalam memanfaatkan teknologi pencarian dan pemrosesan bahasa alami untuk aplikasi kuliner, dan dapat dikembangkan lebih lanjut dengan menambahkan fitur-fitur seperti rekomendasi berbasis preferensi pengguna atau integrasi dengan platform media sosial untuk berbagi resep.

Berikut adalah tautan terkait:

1. **Tautan Aplikasi Deployment**, aplikasi ini telah berhasil di-deploy dan dapat diakses melalui tautan berikut:
<https://stki-a11202213980-uas-fhzhw92q4xkrybfffmqvrpd.streamlit.app/>
2. **Tautan Repository Tugas**, seluruh kode sumber dan dokumentasi tugas tersedia di repository GitHub berikut
<https://github.com/AnnisaHimatulChasanah/stki24>