

BAB 1 PENDAHULUAN

1.1 Latar Belakang Masalah

Website telah menjadi salah satu media utama dalam penyediaan informasi di berbagai bidang, termasuk kuliner. Salah satu implementasinya adalah *website* penyedia resep masakan yang membantu pengguna mencari resep dan ide memasak [1]. Seiring meningkatnya jumlah informasi yang tersedia, kualitas sistem pencarian menjadi aspek penting agar pengguna dapat menemukan informasi yang sesuai dengan kebutuhannya [2]. Namun, sebagian besar *website* resep masih menggunakan pendekatan *keyword search* berbasis *Lexical Matching*. Metode ini bekerja dengan cara membandingkan kesamaan karakter teks secara literal antara kueri pengguna dengan data yang ada di *database* [3]. Pendekatan tersebut terbatas pada pencocokan kata secara literal, sehingga tidak mampu memahami konteks kueri pengguna [4]. Kondisi ini dikenal dengan istilah *lexical gap* atau *vocabulary mismatch*. Akibatnya, sistem pencarian dapat gagal menampilkan data yang relevan, meskipun data yang dicari sebenarnya tersedia di dalam *database* [5]. Kondisi tersebut menyebabkan pengguna kesulitan menemukan resep yang sesuai dengan kebutuhannya.

Keterbatasan pencarian berbasis kata kunci (*keyword search*) atau *lexical matching* dapat ditemui pada berbagai platform penyedia resep masakan Indonesia seperti cookpad.com, dapurumami.com, dan masakapahariini.com. Pada platform tersebut, sistem pencarian masih bergantung pada kecocokan karakter teks secara literal antara kueri pengguna dengan judul resep masakan yang tersimpan di *database*. Kondisi ini menimbulkan hambatan ketika pengguna melakukan pencarian secara natural dengan menggunakan kueri deskriptif yang tidak hanya merujuk pada nama masakan, tetapi juga mencakup aspek seperti rasa, tekstur, atau suasana tertentu. Akibatnya, sistem berbasis *keyword search* sering kali gagal memberikan hasil yang relevan karena tidak mampu menangkap konteks antara kueri pengguna dengan data yang tersimpan dalam *database* [6].

Penelitian sebelumnya yang berjudul “Pengembangan Sistem Pencarian Resep Makanan dengan Implementasi Text Preprocessing dan BM25” mengembangkan sistem pencarian resep makanan berbasis web menggunakan metode BM25. Hasil penelitian menunjukkan bahwa metode BM25 mampu

menghasilkan hasil pencarian resep yang relevan pada sistem pencarian berbasis teks [7]. Meskipun demikian, sistem tersebut masih menggunakan pendekatan *keyword search*, sehingga pencarian bergantung pada kecocokan kata secara literal dan belum mampu menangkap konteks kueri pengguna. Sementara itu, penelitian yang berjudul “VectorSearch: Enhancing Document Retrieval with Semantic Embeddings and Optimized Search” mengembangkan sistem pencarian dokumen berbasis *vector search* menggunakan *semantic embedding*. Hasil penelitian menunjukkan bahwa pendekatan tersebut mampu meningkatkan akurasi *retrieval* serta menghasilkan nilai *precision* dan *recall* yang tinggi [8]. Kedua penelitian sebelumnya menunjukkan bahwa pendekatan *keyword search* maupun *vector search* diimplementasikan pada domain yang berbeda. Namun, belum terdapat penelitian yang mengimplementasikan *vector search* pada sistem pencarian resep masakan Indonesia berbasis web menggunakan model Multilingual-E5-Large.

Penelitian ini menerapkan metode *vector search* berbasis konteks pada resep masakan Indonesia dengan menggunakan model *embedding* Multilingual-E5-Large. Model ini dipilih karena dirancang khusus untuk tugas *retrieval* dan mampu menghasilkan representasi teks untuk mendukung proses pencarian berbasis vektor [9]. Selanjutnya, model mengubah teks menjadi representasi angka dalam bentuk vektor berdimensi tinggi (*dense vector*). Dengan pendekatan ini, sistem dapat menghitung kedekatan posisi antar vektor dalam ruang matematis yang merepresentasikan kemiripan makna antara kueri pengguna dan resep. Untuk mendukung penyimpanan dan pencarian vektor, penelitian ini menggunakan Qdrant sebagai *vector database* dengan algoritma HNSW yang memungkinkan sistem melakukan *vector search* pada resep masakan Indonesia.

Untuk mengevaluasi kualitas sistem pencarian, diperlukan metrik yang mampu mengukur relevansi hasil pencarian. Pengguna cenderung memperhatikan hasil pencarian pada peringkat teratas [10]. Oleh karena itu, evaluasi sistem pencarian tidak hanya mempertimbangkan tingkat relevansi, namun juga posisi atau urutan hasil dalam daftar peringkat [11]. Berdasarkan karakteristik sistem pencarian tersebut, penelitian ini menggunakan metrik *Normalized Discounted Cumulative Gain* karena mampu mengevaluasi tingkat relevansi sekaligus mempertimbangkan posisi hasil pencarian dalam daftar peringkat. Selanjutnya nilai K atau jumlah evaluasi ditetapkan sebesar 30 karena evaluasi difokuskan pada 30 hasil pencarian teratas, yang merupakan indikator penting untuk mengevaluasi performa sistem *retrieval* dan kepuasan pengguna [12].

1.2 Rumusan Masalah

Berdasarkan latar belakang permasalahan yang telah dijelaskan, rumusan masalah dalam penelitian sebagai berikut

1. Bagaimana mengimplementasikan metode *vector search* menggunakan model Multilingual-E5-Large?
2. Bagaimana tingkat akurasi sistem pencarian resep masakan Indonesia berbasis *website* menggunakan metode *vector search* dan model Multilingual-E5-Large dengan menggunakan metrik *NDCG@30*?

1.3 Batasan Permasalahan

Penelitian ini memiliki beberapa batasan masalah sebagai berikut.

- a. Dataset yang digunakan adalah resep masakan Indonesia yang diperoleh melalui platform Kaggle yang terdiri dari kategori bahan utama (ayam, ikan, kambing, sapi, tahu, telur, tempe, udang) dengan total sampel data tertentu.
- b. Model *embedding* yang digunakan terbatas pada Multilingual-E5-Large.
- c. Sistem hanya menerima *input* berupa teks dan tidak memproses pencarian berupa gambar, suara, atau file dokumen lainnya.
- d. *Ground truth* pada penelitian ini disusun melalui penilaian relevansi secara manual.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah sebagai berikut.

1. Mengimplementasikan metode *vector search* dengan menggunakan model Multilingual-E5-Large
2. Mengevaluasi tingkat akurasi sistem pencarian resep masakan Indonesia berbasis *website* menggunakan metode *vector search* dan model Multilingual-E5-Large dengan menggunakan metrik *NDCG@30*.

1.5 Manfaat Penelitian

Manfaat pada penelitian ini adalah membantu pengguna dalam menemukan resep masakan Indonesia secara kontekstual melalui sistem pencarian berbasis vektor dan memberikan gambaran implementasi serta akurasi sistem pencarian berbasis vektor menggunakan model Multilingual-E5-Large yang dapat dijadikan referensi bagi peneliti selanjutnya.

1.6 Sistematika Penulisan

Sistematika penulisan laporan penelitian ini adalah sebagai berikut

a. Bab 1 PENDAHULUAN

Bab ini terdiri atas latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

b. Bab 2 LANDASAN TEORI

Bab ini menjelaskan penelitian terdahulu, landasan teori yang mendukung penelitian mengenai *vector search*, Multilingual-E5-Large, algoritma HNSW, dan *vector database* Qdrant beserta *cosine similarity*.

c. Bab 3 METODOLOGI PENELITIAN

Bab ini membahas metodologi penelitian, analisis sistem, dan perancangan sistem yang dikembangkan terdiri dari *data flow diagram*, *flowchart*, skema *database*, struktur tabel, serta *wireframe*.

d. Bab 4 HASIL DAN DISKUSI

Bab ini membahas implementasi sistem dan algoritma, pengujian algoritma, pengujian sistem, serta hasil pengujian dan evaluasi yang dilakukan.

e. Bab 5 SIMPULAN DAN SARAN

Bab ini berisi kesimpulan yang diperoleh dari hasil penelitian serta saran untuk pengembangan penelitian selanjutnya.