

Sistem Pencarian Resep Makanan Berbasis Semantic Embedding Dan Sentence-Bert

Imanda Salsabila Aprilian¹, Izshauma Pahawana Sindura², Firdaus Syafiq³, Yosi Sofyan Pangestu⁴, Nailu Nur Rohmah⁵, Arya Fahmi Nugroho⁶

^{1,2,3,4,5,6}Informatika, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Purwokerto, Indonesia
surel:¹imandasalsabila04@gmail.com,²Izshauma.ps@gmail.com,³doz2876i@gmail.com,⁴yosisofyan5@gmail.com,
⁵nailunrohmah05@gmail.com,⁶aryafahminugroho@gmail.com

Info Artikel

Sejarah artikel:

Diterima 26-01-2026

Revisi 12-02-2026

Diterima 02-03-2026

Kata kunci:

Pencarian Resep

Semantic Search

Sentence-BERT

FAISS

Mean Average Precision (MAP)

ABSTRAK

Tujuan dari proyek ini adalah untuk membuat sistem pencarian resep makanan berbasis pencarian semantik menggunakan model Sentence-BERT (SBERT) dan embedding semantik. Masalah utama dengan pencarian resep adalah bahwa sistem pencarian tradisional tidak dapat memahami makna kata yang digunakan dalam pencarian, yang sering mengakibatkan hasil pencarian yang tidak berguna. Untuk memungkinkan pencocokan berbasis makna daripada hanya pencocokan kata kunci, studi ini menyarankan penggunaan SBERT untuk membuat representasi semantik dari teks resep. Metodologi eksperimental studi ini terdiri dari dua fase utama: pencarian internet dan pengindeksan *offline*. SBERT digunakan untuk memproses dan mengubah data resep makanan Indonesia menjadi vektor, yang kemudian disimpan dalam indeks FAISS untuk pencarian yang efisien. Dengan menggunakan 390 data uji, statistik *Mean Average Precision* (MAP) digunakan untuk menilai kinerja sistem. Menurut hasil pengujian, peningkatan relevansi hasil pencarian paling baik dicapai dengan nilai ambang 0,50. Berdasarkan makna dari kueri pengguna, sistem yang dikembangkan dapat menghasilkan pencarian yang lebih relevan, menurut temuan penelitian. Selain menawarkan peluang untuk pertumbuhan di masa depan dengan fitur personalisasi dan integrasi data multimodal, penelitian ini mendorong pengembangan sistem pencarian berbasis semantik.

Penulis yang sesuai:

Firdaus Syafiq

Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Purwokerto

Email: doz2876i@gmail.com

1. PENDAHULUAN

Informasi digital kini jauh lebih mudah diakses, bahkan dalam industri kuliner, berkat kemajuan teknologi informasi dan internet. Saat ini, berbagai resep makanan dapat diakses secara *online* dalam format teks dan multimedia melalui situs web, blog, dan aplikasi. Namun, menemukan resep yang cocok bagi orang-orang tidak selalu mudah meskipun banyaknya informasi yang tersedia. Sebagian besar mesin pencari resep populer saat ini masih menggunakan metode berbasis kata kunci, yang membatasi kemampuan mereka untuk memahami konteks dan makna dari masukan pengguna. Hasil pencarian seringkali tidak relevan atau tidak akurat karena perbedaan kosakata, sinonim, dan variasi ejaan dari berbagai jenis makanan [1].



Teknik pengambilan informasi berbasis teks telah digunakan dalam sejumlah penelitian sebelumnya untuk mencoba menyelesaikan masalah ini. Dengan tingkat akurasi 80% pada nilai K tertentu, Nuryaman dkk. membuat sistem pengambilan resep tradisional Indonesia berbasis bahan dengan memanfaatkan algoritma K-Nearest Neighbor (KNN) dengan representasi teks TF-IDF dan pengukuran kemiripan menggunakan cosine similarity [2]. Dalam studi yang berbeda, Siregar dkk menggunakan algoritma BM25 dengan langkah-langkah persiapan teks untuk mencari resep makanan dan menemukan bahwa hasil pencarian lebih relevan dibandingkan dengan yang diperoleh menggunakan teknik pencocokan kata dasar [3]. Meskipun demikian, pendekatan-pendekatan yang disebutkan sebelumnya masih bersifat leksikal dan tidak mampu memahami hubungan semantik antara kata-kata, sehingga rentan terhadap sinonim, konteks, dan variasi bahasa.

Teknik semantik berbasis embedding semakin banyak digunakan untuk meningkatkan kualitas pengambilan informasi seiring dengan kemajuan di bidang pemrosesan bahasa alami [4]. Sentence-BERT (SBERT), sebuah model yang dibuat oleh Reimers dan Gurevych, dirancang untuk menghasilkan representasi vektor kalimat yang lebih efektif menangkap makna semantik dibandingkan BERT konvensional [5]. Telah ditunjukkan bahwa model ini dapat meningkatkan kinerja pada sejumlah tugas pencarian semantik dan penilaian kesamaan teks. Fathuddin et al. juga telah menerapkan SBERT pada pencarian semantik dokumen, menunjukkan bahwa kombinasi SBERT dan kesamaan kosinus dapat sangat meningkatkan relevansi hasil pencarian dokumen akademik. [6].

Dalam bidang memasak dan saran makanan, strategi serupa juga mulai digunakan. RecipeBERT, sebuah sistem rekomendasi resep berbasis BERT dengan kesadaran semantik terhadap preferensi dan konteks pengguna, diusulkan oleh Mereddy dan Reddy [7]. Selain itu, penelitian Morales dkk. menunjukkan bahwa, dibandingkan dengan pendekatan berbasis kata kunci, penerapan embedding kontekstual pada data resep dapat meningkatkan kualitas pencarian makanan dan saran [8]. Penelitian-penelitian tersebut membuktikan bahwa pendekatan semantik memiliki potensi besar dalam meningkatkan akurasi dan relevansi sistem pencarian resep.

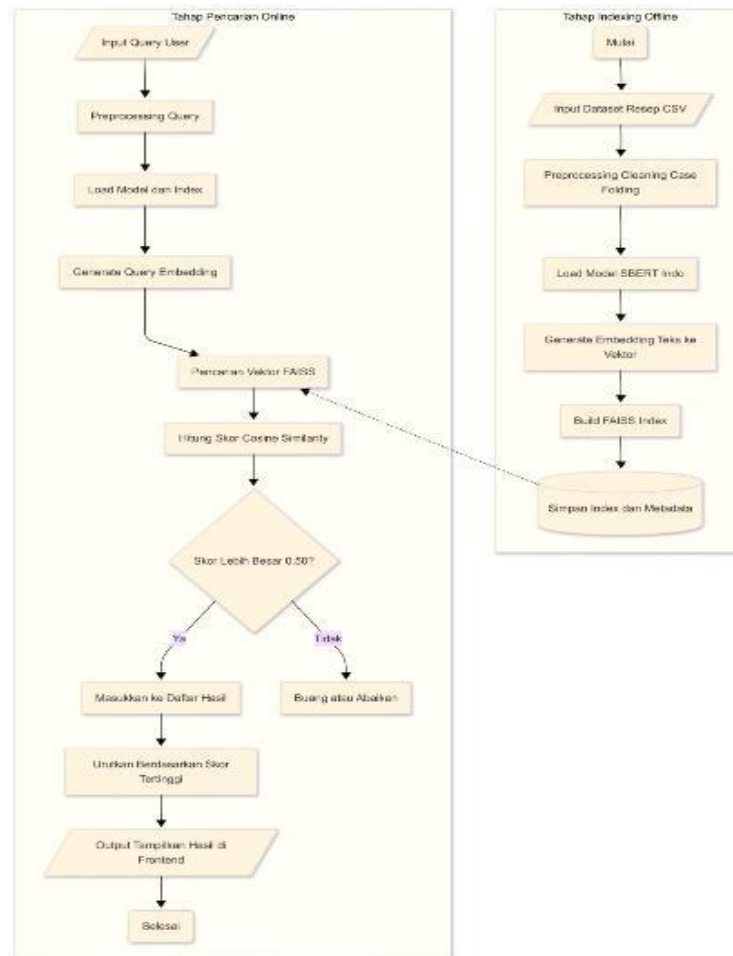
Penelitian ini mengusulkan penggunaan embedding semantik dan model Sentence-BERT untuk membangun sistem pencarian resep makanan berbasis pencarian semantik sebagai respons terhadap masalah dan keterbatasan dari penelitian sebelumnya. Dengan menggunakan koleksi resep masakan Indonesia, sistem ini memungkinkan pengguna untuk mencari berdasarkan bahan atau nama sambil memahami konten kueri. Sistem ini mengintegrasikan Facebook AI Similarity Search (FAISS) sebagai mesin pencarian vektor, yang telah terbukti sukses untuk pencarian kesamaan vektor skala besar, guna memungkinkan pencarian yang efektif pada dataset besar [9].

Penggunaan Sentence-BERT untuk membuat representasi semantik dari kueri pengguna dan data resep dalam konteks resep masakan Indonesia, bersama dengan penerapan FAISS untuk meningkatkan relevansi dan efisiensi pencarian, merupakan aspek inovatif dari penelitian ini. Selain itu, sinonim dan variasi linguistik yang sering muncul pada makanan daerah juga ditangani oleh sistem ini. Oleh karena itu, diharapkan penelitian ini akan memberikan kontribusi ilmiah bagi kemajuan sistem pengambilan informasi berbasis semantik, khususnya di bidang kuliner.

2. METODE

Penelitian ini menggunakan metodologi eksperimental dan kuantitatif. Karena penelitian ini berfokus pada penggunaan kriteria evaluasi numerik untuk menilai efektivitas sistem pencarian resep makanan berbasis pencarian semantik, pendekatan kuantitatif dipilih. Model Sentence-BERT (SBERT) dan embedding semantik diuji secara eksperimental untuk melihat seberapa baik keduanya meningkatkan relevansi hasil pencarian resep dibandingkan dengan metode pencocokan berbasis kata kunci. Penelitian tentang sistem pencarian semantik dan pengambilan informasi sering menggunakan metode ini. [10], [11].





Gambar 1. Alur Penelitian

2.1 Desain dan Alur Penelitian

Alur sistem yang digambarkan pada Gambar 1 menjadi dasar bagi pengaturan kronologis dari desain penelitian. Tahap pengindeksan offline dan tahap pencarian online adalah dua fase utama dari sistem. Dengan melaksanakan proses komputasi yang berat hanya sekali pada tahap pertama dan memungkinkan tahap pencarian berjalan dengan cepat saat sistem digunakan, pemisahan kedua tahap ini bertujuan untuk meningkatkan efisiensi sistem [2].

2.2 Tahap Indexing Offline

Sebelum konsumen menggunakan sistem, tahap pengindeksan offline diselesaikan sekali. Langkah pertama dalam proses ini adalah memasukkan dataset CSV resep makanan Indonesia. Nama resep, daftar bahan, dan instruksi memasak semuanya termasuk dalam dataset. Setelah itu, dilakukan prosedur pra-pemrosesan, yang melibatkan penyeragaman huruf menjadi huruf kecil dan penghapusan karakter yang tidak diperlukan melalui pembersihan teks dan normalisasi huruf. Tahap ini sangat penting untuk meningkatkan kualitas representasi teks dan mengurangi kebisingan data [3].

Untuk membuat representasi semantik dari setiap teks resep, sistem memuat model Sentence-BERT dalam bahasa Indonesia setelah melakukan pra-pemrosesan. Teks resep diubah menjadi vektor numerik (embedding) yang secara kontekstual menunjukkan makna kata-kata menggunakan model SBERT [8]. Setelah itu, embeddings yang dibuat disimpan dalam indeks FAISS tipe IndexFlatIP. FAISS dipilih karena kemampuannya untuk melakukan pencarian kesamaan vektor skala besar dengan efisiensi [8]. Selain indeks vektor, metadata resep disimpan dalam bentuk berkas terpisah untuk mendukung penampilan informasi hasil pencarian.

2.3 Tahap Pencarian Online

Ketika seorang pengguna memasukkan kueri dalam bentuk nama atau bahan makanan, fase pencarian online dimulai. Untuk menjaga konsistensi representasi teks, kueri awalnya melewati prosedur persiapan yang mirip dengan

tahap pengindeksan, yang mencakup pembersihan dan penyesuaian huruf besar-kecil. Sistem kemudian memuat indeks FAISS dan model SBERT yang telah dibangun pada fase sebelumnya.

Model SBERT kemudian digunakan untuk mengubah kueri yang telah diproses menjadi embeddings. Prosedur pencarian vektor kemudian digunakan untuk membandingkan embeddings kueri ini dengan embeddings resep yang disimpan dalam database FAISS. Sistem menentukan tingkat kesamaan antara setiap resep dan kueri dengan menghitung skor kesamaan menggunakan cosine similarity [9]. Hasil pencarian berupa daftar kandidat resep beserta skor kemiripannya [12].

2.4 Mekanisme Seleksi dan Penyajian Hasil

Sistem menggunakan metode seleksi berbasis ambang untuk meningkatkan relevansi hasil pencarian [13]. Resep akan muncul dalam daftar hasil jika skor kemiripan kosinus lebih tinggi dari 0,50. Resep yang mendapatkan skor lebih rendah dari ini akan diabaikan. Setelah melewati proses seleksi, resep-resep tersebut diurutkan berdasarkan skor kemiripan tertinggi dan disajikan kepada pengguna melalui antarmuka depan. Ketepatan sistem meningkat dan hasil yang tidak relevan berkurang dengan menggunakan metode berbasis ambang ini [14].

2.5 Metode Pengumpulan Data

Dataset resep kuliner Indonesia dari repositori publik Kaggle digunakan sebagai metode pengumpulan data arsip dalam penelitian ini. Semua data resep dalam dataset termasuk dalam populasi penelitian. Algoritma Slovin dengan margin kesalahan 5% digunakan untuk men-sampling data uji dalam evaluasi sistem, menghasilkan 390 sampel data. Penggunaan metode ini menjamin bahwa sampel secara akurat mencerminkan populasi [2].

2.6 Metode Pengujian dan Analisis Data

Metode Mean Average Precision (MAP) digunakan untuk mengukur kinerja relevansi hasil pencarian guna menguji sistem. MAP sering digunakan dalam penelitian pengambilan informasi karena dapat menilai secara menyeluruh kualitas urutan hasil pencarian [15]. Metode Mean Average Precision (MAP) digunakan untuk mengukur kinerja relevansi hasil pencarian guna menguji sistem. MAP sering digunakan dalam penelitian pengambilan informasi karena dapat menilai secara menyeluruh kualitas urutan hasil pencarian.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Implementasi Sistem

Penelitian ini menghasilkan sistem pencarian resep makanan berbasis pencarian semantik yang memanfaatkan model Sentence-BERT (SBERT) dan embedding semantik. Dua tahap utama dalam implementasi efektif sistem ini adalah pengindeksan offline dan pencarian online. Semua data resep makanan Indonesia yang telah dipra-proses berhasil diubah menjadi representasi vektor dan dimasukkan ke dalam indeks FAISS selama tahap pengindeksan. Selama tahap pencarian, prosedur ini memungkinkan sistem untuk melakukan pencarian berbasis kesamaan semantik secara efektif.



Gambar 2. Implementasi Sistem

Langkah pencarian online ditunjukkan pada Gambar 2, di mana pengguna memasukkan kueri dalam bentuk nama atau item makanan. Model SBERT yang sama yang digunakan pada tahap pengindeksan digunakan untuk memproses kueri dan mengubahnya menjadi embeddings. Sistem kemudian menggunakan FAISS untuk melakukan pencarian vektor dan menghasilkan daftar resep potensial beserta skor kesamaan (*cosine similarity*). Setelah itu, hasil

pencarian diurutkan sesuai dengan skor tertinggi, disaring menggunakan nilai ambang yang telah ditentukan, dan ditampilkan kepada pengguna melalui antarmuka depan. Implementasi ini menunjukkan bahwa sistem dapat beroperasi sesuai dengan alur yang ditunjukkan pada Gambar 1.

3.2 Hasil Evaluasi Menggunakan *Mean Average Precision* (MAP)

Metrik *Mean Average Precision* (MAP) digunakan untuk menilai kinerja sistem. Sebanyak 390 data uji, dihitung menggunakan rumus Slovin dengan tingkat kesalahan 5%, digunakan untuk pengujian. Dengan memfokuskan pada urutan penyajian hasil pencarian kepada pengguna, MAP digunakan untuk menilai kualitas relevansi keseluruhan dari hasil pencarian.

Tabel 1 Evaluasi *Mean Average Precision*

NO	THRESHOLD	NILAI MAP
1	0.15	0.62
2	0.30	0.71
3	0.50	0.79
4	0.70	0.74

Untuk mendapatkan batas skor kesamaan yang ideal, pengujian dilakukan dengan mencoba berbagai nilai ambang. Jika dibandingkan dengan nilai cutoff alternatif, hasil uji pada Tabel 1 menunjukkan nilai ambang 0,50 menghasilkan nilai MAP yang optimal. Hal ini menunjukkan bahwa jumlah hasil yang ditampilkan dan relevansinya dapat diseimbangkan dengan nilai ini. Nilai ambang yang lebih tinggi menghasilkan terlalu sedikit hasil, sedangkan nilai ambang yang lebih rendah biasanya menghasilkan lebih banyak hasil pencarian dengan relevansi yang lebih rendah.

3.3 Analisis Relevansi Hasil Pencarian

Sistem dapat menghasilkan resep yang secara semantik terkait dengan pertanyaan pengguna, sesuai dengan temuan uji. Sistem mempertimbangkan konteks dan makna dari kueri selain pencocokan kata secara literal. Misalnya, karena model SBERT menggunakan vektor semantik untuk menggambarkan teks, pencarian dengan nama bahan kuliner dengan ejaan yang berbeda tetap dapat memberikan resep yang relevan.

Penyematan semantik mengurangi masalah yang sering terjadi pada mesin pencari berbasis kata kunci, seperti tidak menemukan hasil karena sinonim atau perbedaan istilah, dengan memungkinkan algoritma memahami hubungan antar kata dan konteks kalimat. Selain itu, telah terbukti bahwa penggunaan FAISS mempercepat prosedur pencarian vektor, memastikan sistem tetap responsif bahkan ketika terdapat banyak data resep.

Secara keseluruhan, temuan studi ini menunjukkan bahwa penggunaan Sentence-BERT dan embedding semantik pada sistem pencarian resep makanan dapat menghasilkan hasil yang relevan tergantung pada kueri pengguna. Efektivitas sistem ini dalam mengelola pencarian pada jumlah data yang besar didukung oleh penggunaan FAISS sebagai mesin pencarian vektor. Akibatnya, sistem yang dibuat dapat secara efektif mengatasi masalah pencarian resep, yang tidak dapat ditangani dengan baik oleh sistem berbasis kata kunci..

4. KESIMPULAN

Dengan menggunakan model Sentence-BERT (SBERT) dan embedding semantik, studi ini secara efektif menciptakan mesin pencari resep makanan yang berbasis pencarian semantik. Sistem ini dapat mencapai tujuan penelitian yang disebutkan dalam pendahuluan dengan mengambil input berupa nama makanan atau bahan, dan menghasilkan resep yang relevan berdasarkan makna dari pencarian tersebut. Nilai ambang 0,50 memberikan performa optimal dalam membatasi hasil pencarian yang relevan, menurut hasil evaluasi menggunakan Mean Average Precision (MAP). Efektivitas sistem dalam memindai perpustakaan resep yang luas semakin didukung oleh penggunaan FAISS sebagai mesin pencari vektor. Untuk mencapai analisis yang lebih menyeluruh, studi ini dapat dikembangkan lebih lanjut dengan menyertakan fitur personalisasi pencarian, evaluasi berbasis pengguna, dan perbandingan dengan metode pencarian berbasis kata kunci sebagai baseline.

REFERENSI

- [1] A. Shiri, "Introduction to Modern Information Retrieval (2nd edition)," *Libr. Rev.*, vol. 53, no. 9, pp. 462–463, 2004.
- [2] Y. Nuryaman, A. Asistyasari, B. Sudarsono, and U. Faddilah, "Sistem Temu Kembali Masakan Tradisional Indonesia Berdasarkan Bahan Berbahasa Inggris Menggunakan Algoritma KNN," *J. JUPITER*, vol. 13, no. 1, pp. 60–65, 2021.
- [3] S. Rosetya Wardhana, T. N. Audrey, H. Nugroho, T. Informatika, T. Adhi, and T. Surabaya, "Sistem Rekomendasi Tempat Makan Berbasis Konten Berdasarkan Metode Best Match 25 Lucene (BM25L)," pp. 1–8, 2024.
- [4] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using siamese BERT-networks," *EMNLP-IJCNLP 2019 - 2019*



-
- Conf. Empir. Methods Nat. Lang. Process. 9th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf.*, pp. 3982–3992, 2019.
- [5] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *Naacl-Hlt 2019*, no. Mlm, pp. 4171–4186, 2018.
- [6] M. A. H. Fathuddin, E. P. Mandyartha, and A. L. Nurlaili, “Penerapan Sentence-Bert dan Cosine Similarity untuk Pencarian Semantik Dokumen Skripsi dalam Format PDF,” *Ranah Res. J. Multidiscip. Res. Dev.*, vol. 8, no. 1, pp. 322–337, 2025.
- [7] D. Mereddy and J. S. R. Beedareddy, “Enabling Next-Generation Smart Homes Through Bert Personalized Food Recommendations - RecipeBERT,” in *2024 IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, 2024, pp. 796–803.
- [8] A. Morales-Garzón, J. Gómez-Romero, and M. J. Martín-Bautista, “Contextual Sentence Embeddings for Obtaining Food Recipe Versions,” in *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, 2022, pp. 306–316.
- [9] J. Johnson, M. Douze, and H. Jegou, “Billion-Scale Similarity Search with GPUs,” *IEEE Trans. Big Data*, vol. 7, no. 3, pp. 535–547, 2021.
- [10] C. D. Manning, P. Raghavan, and H. Schütze, “Chapter 9: Relevance feedback and query expansion - L15,” *Introd. to Inf. Retr.*, no. c, pp. 177–194, 2008.
- [11] A. I. S. Arif Rachman, Yochanan, *Dan R & D*. 2024.
- [12] J. Kehrli, “Introduction to Information Retrieval Resume of the MSE lecture Resume of the Data management lecture,” 2010.
- [13] T. Saracevic, “Evaluation of evaluation in information retrieval,” *SIGIR Forum (ACM Spec. Interes. Gr. Inf. Retrieval)*, no. July 1995, pp. 137–146, 1995.
- [14] H. Wang *et al.*, “Cross-Modal Food Retrieval: Learning a Joint Embedding of Food Images and Recipes With Semantic Consistency and Attention Mechanism,” *IEEE Trans. Multimed.*, vol. 24, pp. 2515–2525, 2022.
- [15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.

