ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

PENGEMBANGAN *CHATBOT* INFORMASI PARIWISATA DI BALI MENGGUNAKAN METODE *HYBRID RAG*

I Kadek Dwi Pratama✉, Ni Wayan Wisswani, Gde Brahupadhya Subiksa
Jurusan Teknologi Informasi, Politeknik Negeri Bali, Badung, Indonesia
Email: dwipratamaikadek@gmail.com

DOI: https://doi.org/10.46880/jmika.Vol10No2.pp741-751

ABSTRACT

The development of artificial intelligence-based chatbots has enabled the automated delivery of information. However, chatbots still face challenges in generating relevant responses to users' queries. In addition, Retrieval-Augmented Generation (RAG) systems that rely on a single retrieval method may affect the quality of the generated responses. Therefore, this study aims to implement a Hybrid Retrieval-Augmented Generation (RAG) approach by combining BM25, a Bi-Encoder, and a Cross-Encoder. The dataset consists of 200 Bali tourism documents collected through web scraping, Statistics Indonesia (BPS) Bali, and Kaggle. The proposed system employs BM25 as the initial retrieval stage, a Bi-Encoder for semantic retrieval, a Cross-Encoder for reranking, and the Gemini model for response generation. The system was evaluated using Recall@k, Precision@k, and nDCG@k metrics to measure retrieval performance, as well as response time testing to assess system efficiency. The results indicate that the Hybrid RAG approach, integrating BM25, a Bi-Encoder, and a Cross-Encoder, improves document retrieval quality by combining keyword-based retrieval, semantic similarity matching, and reranking. The best-performing configuration achieved an average Recall@k of 33%, Precision@k of 80%, and nDCG@k of 94%. These findings demonstrate that the combination of BM25, a Bi-Encoder, and a Cross-Encoder effectively retrieves and ranks relevant documents, providing higher-quality contextual information for answer generation by the Gemini model.

Keywords: Hybrid Retrieval-Augmented Generation, Gemini, BM25, Bi-Encoder, Cross-Endcoder.

ABSTRAK

Perkembangan chatbot berbasis kecerdasan buatan memungkinkan penyediaan informasi secara otomatis. Namun, chatbot masih memiliki kendala dalam menghasilkan jawaban relevan terhadap pertanyaan pengguna. Selain itu sistem Retrieval-Augmented Generation (RAG) yang hanya mengandalkan satu metode retieval dapat mempengaruhi kualitas jawaban yang dihasilkan. Oleh karena itu penelitian ini bertujuan untuk mengimplementasikan pendekatan Hybrid Retrieval-Augmented Generation (RAG) yang menggabungkan BM25, Bi-Encoder, dan Cross-Encoder. Dataset yang digunakan terdiri dari 200 dokumen pariwisata Bali yang diperoleh dari web scraping, Badan Pusat Statistik Bali, dan Kaggle. Sistem dibangun menggunakan BM25 sebagai tahap awal retrieval, Bi-Encoder untuk semantik retrieval, Cross-Encoder untuk reranking, serta model Gemini untuk generasi jawaban. Evaluasi dilakukan menggunakan matriks Recall@k, Precision@k, dan nDCG@k untuk mengukur kualitas retrieval serta pengujian waktu respond sistem. Hasil menunjukkan bahwa pendekatan Hybrid RAG dengan menggunakan model BM25, Bi-Encoder, dan Cross-Encoder mampu meningkatkan kualitas pencarian dokumen melalui pencarian berdasarkan kata kunci, kesamaan makna, dan proses pengurutan ulang. Konfigurasi terbaik dapat berhasil memperoleh nilai rata-rata Recall@k sebesar 33%, Precision@k sebesar 80%, dan nDCG@k sebesar 94%. Hasil tersebut menunjukkan bahwa kombinasi BM25, Bi-Encoder, dan Cross-Encoder mampu menemukan dan mengurutkan dokumen yang relevan sehingga menghasilkan konteks yang lebih baik untuk generasi jawaban oleh Gemini.

Kata Kunci: Hybrid Retrieval-Augmented Generation, Gemini, BM25, Bi-Encoder, Cross-Endcoder.

PENDAHULUAN

Pariwisata merupakan salah satu sektor penting yang berkontribusi besar terhadap perekonomian daerah, khususnya di Bali. Setiap tahun, wisatawan domestik maupun mancanegara datang untuk menikmati berbagai destinasi wisata, budaya, kuliner, hingga kegiatan rekreasi. Namun, tidak semua wisatawan memiliki informasi yang memadai mengenai lokasi wisata, akses, harga tiket, rekomendasi aktivitas, serta destinasi yang sesuai dengan preferensi mereka (Sanjaya & Winarno, 2024; Situmeang et al., 2023). Kondisi tersebut menimbulkan

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

kesulitan dalam merencanakan perjalanan terutama, bagi wisatawan yang belum mengenal daerah tujuan.

Informasi pariwisata saat ini masih diperoleh melalui media sosial, blog perjalanan, maupun masyarakat setempat. Informasi yang tersedia dari internet maupun media sosial sering tidak terstruktur, tidak selalu benar, dan belum tentu sesuai dengan kebutuhan wisatawan. Di sisi lain masyarakat lokal ataupun pemandu wisata belum tentu memiliki pengetahuan yang lengkap mengenai seluruh destinasi wisata yang ada.

Salah satu pendekatan yang digunakan yaitu RAG *(Retrieval-Augmented Generation)* yang menggabungkan metode *retrieval* dan LLM untuk mengahasilkan jawaban. Penelitan mengenai RAG telah banyak dilakukan untuk meningkatkan hasil jawaban dari kemampuan LLM (*Large Language Model)* dalam menghasilkan jawaban berdasarkan dokumen *ekternal.* Pendekatan ini mampu mengurangi halusinasi pada LLM karena jawaban yang diperoleh dari basis pengetahuan. Namun, *pendekatan* RAG dengan menggunakan satu *retrieval* baik berbasis *lexical* maupun *semantic retrieval* memiliki keterbatasan dalam menemukan seluruh dokumen yang relevan, sehingga dapat mempengaruhi kualitas konteks yang diberikan kepada LLM dan dapat mempengaruhi kualitas jawaban yang dihasilkan (Trung et al., 2024).

Beberapa penelitian telah mengembangkan *Hybrid* RAG dengan mengkombinasikan beberapa metode *retrieval* untuk meningkatkan performa pencarian dokumen (Su et al., 2024). Selain itu, metode *reranking* juga mulai diterapkan untuk menyusun kembali hasil pencarian berdasarkan tingkat relevansi dokumen terhadap pertanyaan pengguna (Helviansyah et al., 2025). Meskipun demikian, penelitian-penelitian tersebut umumnya berfokus pada domain umum atau menggunakan dataset berbahasa Inggris. Implementasi *Hybrid* RAG pada domain pariwisata Indonesia, khususnya pariwisata Bali, masih terbatas. Selain itu, belum banyak penelitian yang *mengintegrasikan* BM25 sebagai *lexical retrieval*, *Bi-Encoder* sebagai *semantic retrieval*, dan *Cross-Encoder* sebagai *reranking* dalam satu alur *Hybrid* RAG, serta mengevaluasi kualitas proses *retrieval* menggunakan metrik *Recall@k*, *Precision@k*, dan nDCG@k sebelum proses generasi jawaban dilakukan.

Berdasarkan *research* gap tersebut, penelitian ini mengembangkan chatbot infomasi pariwisata di Bali menggunakan pendekatan Hybrid RAG *(Retreival-Augmented Generation)* dengan mengintegrasikan BM25 sebagai tahapan awal *lexical retrieval, Bi-Encoder* untuk *semantic retrieval,* dan *Cross-Encoder* untuk *reranking* sebelum dokumen digunakan pada model LLM Gemini. Pendekatan ini diharapkan mampu meningkatkan kualitas dokumen yang digunakan sebagai konteks, sehingga menghasilkan jawaban yang relevan terhadap pertanyaan pengguna.

Kontribusi penelitian ini meliputi tiga aspek utama. Pertama, mengimplementasikan arsitektur *Hybrid* RAG yang mengombinasikan BM25, *Bi-Encoder*, dan *Cross-Encoder* pada *chatbot* informasi pariwisata Bali. Kedua, membangun basis pengetahuan yang terdiri atas 200 dokumen pariwisata Bali yang diperoleh dari hasil *web scraping*, Badan Pusat Statistik Bali, dan Kaggle. Ketiga, mengevaluasi kualitas proses *retrieval* menggunakan metrik *Recall@k, Precision@k, dan nDCG@k,* serta mengukur waktu respons sistem untuk mengetahui efektivitas pendekatan yang diusulkan dalam mendukung proses generasi jawaban.

KAJIAN LITERATUR DAN PENGEMBANGAN HIPOTESIS

Penelitian mengenai sistem informasi *retrieval* dan *reranking* telah banyak dilakukan oleh penelitian sebelumnya dengan berbagai pendekatan, dan metode yang digunakan. Beberapa penelitian sebelumnya :

Penelitian yang dilakukan oleh Muhammad Adrinta Abdurrazzaq, Edwin Lesmana Tjiong, Aulia Fasya, Michelle Hiu, dan Joses Tanuwidjaya. Dalam pengembangan *chatbot* untuk diagnosa penyakit. Penelitian ini mengintegrasikan *RAG (Retrieval Agrument Generator)* dan GPT OSS – 20B yang dirancang untuk memberikan informasi medis dalam bahasa Indonesia, cepat relevan, dan terstruktur(Abdurrazzaq et al., 2025). Hasil penelitian menunjukkan bahwa *chatbot* berbasis RAG memiliki potensi dalam meningkatkan akses informasi medis berbahasa Indonesia. Namun, penelitian ini masih terdapat keterbatasan, di mana belum menjelaskan mekanisme kerja *Hybrid* RAG khususnya pada alur *retrieval* dan *reranking* sebelum diberikan kepada LLM. Sehingga kontribusi setiap tahapan *retrieval* terhadap kualitas jawaban belum dapat dianalisis secara mendalam

Penelitian yang dilakukan oleh Rifky Ady Sanjaya dan Edy Winarno dalam pengembangan sistem informasi pariwisata berbasis *chatbot* dengan metode NLP *(Natural Languge Procesing)*. Dengan menggunakan *platfrom Diagflow* membuat chatbot jauh lebih cepat dan mudah(Sanjaya & Winarno, 2024). Hasil penelitian menunjukkan penggunaan *framework Diagflow* dapat meningkatkan akurasi jawaban. Namun terdapat keterbatasan, di mana *Diagflow* tidak dapat

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

memproses pertanyaan yang kompleks, serta belum mengimplementasikan mekanisme *retrieval* berbasis *lexsikal* dan *semanatic* maupun proses *reranking* untuk meningkatkan relevansi jawaban.

Penelitian yang dilakukan Tasya Agiyola, Indriati, dan Bayu Rahyodi dalam pengembangan evaluasi pencarian dokumen berbasis relevasi, dan perluasan *query* menggunakan teknik BM25 dan *feedback* relevasi. Yang di mana meningkatkan pencarian *e – book* berbahasa Indonesia. Hasil dari penelitian ini menunjukkan bahwa teknik ini dapat meningkatkan performa pencarian, meskipun efektivitasnya sangat bergantung pada relevansi kata yang ditambahkan dan jumlah dokumen yang diambil(Agiyola & Indriati, 2019). Namun pada penelitian ini terdapat keterbatasan karena hanya mengandalkan satu model *retrieval* berbasis leksikal yaitu BM25, sehingga pada kondisi *query* yang diberikan panjang atau kompleks, sistem cenderung memberikan jawaban yang kurang relevan dengan topik pertanyaan pengguna.

Penelitian yang dilakukan oleh Pajri Aprilio, Michael, Putu Surya Nugraha, Hasanul Fahmi mengkaji pendekatan *information retrieval* dengan menggabungkan leksikal dan *sementic retrieval.* Penelitian ini menggabungkan TF – IDF sebagai representasi leksikal dan BERT sebagai representasi *semantic.* Hasil penelitian ini menunjukan hasil dari kombinasi TF – IDF dan BERT memberikan hasil terbaik dengan hasil pengujian tertinggi(Aprilio et al., 2025). Meskipun penelitian ini membuktikan pendekatan *hybrid* dengan mengambungkan metode leksikal dan *semantic,* mampu meningkatkan kinerja informasi *retrieval,* namun penelitian tersebut masih berfokus pada perbandingan setiap model dan belum membahas mekanisme *hybrid* secara bertahap, dan juga penelitian ini belum menerapkan tahap *reranking* untuk meningkatkan relevansi dokumen hasil dari *retrieval.*

Penelitian yang dilakukan oleh Maulidya Prastita Syah, Ajeng Puspa Wardani, Mohammad Idhom, Trimono yang membahas mengenai perbandingan *information retrieval* leksikal dengan *semantic retrieval*, yang dimana memanfaatkan TF – IDF sebagai representase leksikal dan BERT sebagai representase *semantic.* Hasil penelitian menunjukan bahwa pendekatan berbasis BERT memberikan jawaban yang lebih baik dibandingkan TF – IDF, terutama dalam memahami variasi konteks dan sinonim kata (Syah et al., 2025). Namun, penelitian ini hanya berfokus pada perbandingan performa antara metode leksikal dan *semantic* secara terpisah, tanpa menggabungkan metode tersebut dalam satu mekanisme *retrieval* bertahap maupun tahapan *reranking* untuk meningkatkan relevansi hasil pencarian *retrieval.*

Oleh karena itu, penelitian ini mengimplementasikan pendekatan *Hybrid* RAG *(Retrieval-Augmented Generation)* yang mengintegrasikan BM25 sebagai *lexical,* Bi-Encoder sebagai *semantic,* dan *Cross-Encoder* sebagai *reranking* sebelum informasi diberikan kepada model Gemini. Berbeda dengan penelitian sebelumnya, penelitian ini tidak penelitian ini tidak hanya mengimplementasikan pipeline *Hybrid* RAG secara bertahap, tetapi juga mengevaluasi kualitas *retrieval* menggunakan metrik *Recall@k, Precision@k*, dan nDCG@k berdasarkan *ground truth* yang telah divalidasi oleh pakar. Selain itu, penelitian ini mengkaji kontribusi setiap tahapan *retrieval* dan *reranking* terhadap relevansi dokumen pada *chatbot* informasi pariwisata Bali.

Sistem Cerdas/AI

Artificial Intelegent atau kecerdasan buatan merupakan salah satu ilmu komputer yang berfokus pada pengembangan sistem yang mampu meniru kemampuan berpikir manusia. Sistem ini dirancang dapat melakukan berbagai tugas, seperti: pengambilan keputusan, pemecahan masalah, dan pembelajaran dari data(Lubis, 2021). Dengan memanfaatkan algoritma dan data yang tersedia, sistem cerdas mampu menganalisis informasi secara efisien dan menghasilkan keputusan yang optimal.

Penerapan sistem cerdas telah digunakan di beberapa bidang untuk membantu meningkatkan efisiensi dan efektivitas kerja manusia. Sistem cerdas digunakan karena kemampuan dalam mengelola data dengan jumlah yang besar, menganalisis, menghasilkan keputusan, serta dapat mengurangi kesalahan manusia *(human error).*

Sistem cerdas bekerja dengan menerima *inputan*, kemudian memproses data tersebut menggunakan algoritma tertentu, dan menghasilkan *output* yang digarapkan. Pada proses ini biasanya melalui beberapa tahapan seperti, pengumpulan data, pemrosesan data, analisis, serta pengambilan keputusan.

NLP

NLP *(Natural Language Processing)* merupakan bidang studi yang berfokus interaksi bahasa manusia dengan komputer. NLP memungkinkan komputer untuk memahami, mengelola, dan menghasilkan bahasa manusia dalam bentuk teks(Pais, SebastiãoJoão Cordeiro, 2022). Teknologi ini dikembangkan untuk mempermudah pengguna dalam

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

berinteraksi dengan komputer, tanpa harus menggunakan *programming*, sehingga komunikasi dapat dilakukan secara alami dan intuitif.

RAG

RAG *(Retrieval-Augmented Generation)* merupakan suatu pendekatan dalam pemrosesan alami dengan cara menggabungkan IR *(Information Retrieval)* dengan kemampuan generatif dari LLM *(Large Language Model)*. Pendekatan ini, model tidak hanya mengandalkan pengetahuan yang telah dipelajari sebelumnya, tetapi juga memanfaatkan informasi eksternal yang diambil dari basis data atau dokumen (Cuconasu et al., 2024). Dengan diminikan RAG dapat memperkaya *prompt* awal dengan potongan teks atau dokumen yang relevan. Penggunaan RAG dapat memberikan beberapa keunggulan, yang di mana RAG meningkatkan akurasi jawaban dengan dukungan informasi eksternal, dan dapat mengurangi *hallucination* pada LLM.

Hybrid RAG

Hybrid RAG *(Retrieval – Augmented Generator)* merupakan metode yang digunakan dalam pengembangan sistem *chatbot* untuk menghasilkan jawaban yang relevan. Metode bekerja dengan menggabungkan lebih dari satu metode *retrieval,* yaitu leksikal *retrieval* dan *semantic retrieval*, sebelum informasi diberikan ke model *reranking*(Thakur et al., 2021).

Tujuan dari penerapan *Hybrid* RAG *(Retrieval - Augmented Generator)* untuk mengatasi keterbatasan dari *retrieval* tunggal, khususnya dalam memahami makna dari pertanyaan pengguna (Mikael et al., 2025). Dimulai dengan proses leksikal dengan metode BM25 untuk memperoleh kandidat yang relevan, kemudian dilanjutkan dengan *semantic retrieval* menggunakan *Bi – Encoder* untuk menangkap kesamaan makna antara pertanyaan dan dokumen. Hasil akan diberikan ke tahap *reranking* untuk meningkatkan urutan relevansi dokumen.

Retrieval

Retrieval merupakan proses mencari atau pengambilan dokumen secara relevan dari pertanyaan pengguna dalam sebuah data set (Mikael et al., 2025). Dengan tujuan untuk menyaring dokumen awal yang berpotensi mengandung informasi sesuai dengan pertanyaan yang dicari. Proses ini dilakukan pada tahapan awal sistem pencarian informasi dan dijalankan *backend* yang mengelola kumpulan dokumen. *Retrieval* dilakukan dengan menggabungkan dua buah model leksikal dan *semantic* yaitu BM25 dan *Bi – Encoder* yang bekerja dengan menghitung tingkat relevansi antara pertanyaan pengguna dengan dokumen berdasarkan kata dan makna.

Reranking

Reranking merupakan metode yang digunakan untuk mengurutkan kembali hasil dokumen yang diperoleh dari tahap *retrieval* untuk memperoleh dokumen yang lebih relevan. *Reranking* berfungsi sebagai tahap penyaringan ke dua, dimana dokumen yang diperolah *retrieval* akan diberikan skor ulang menggunakan metode yang lebih kompleks (Luan et al., 2021). Dokument dengan skor tertinggi akan ditempatkan di urutan teratas sebelum diteruskan ke LLM *(Large Language Model)* untuk menghasilkan jawaban.

Leksikal dan Semantik

Leksikal merupakan metode dari *retrieval* yang di mana mencocokkan pertanyaan pengguna dengan dokumen berdasarkan kesamaan kata. Metode ini bekerja dengan menghitung bobot setiap istilah pada dokumen, kemudian menghitung skor relevansi antar pertanyaan pengguna dengan dokumen berdasarkan tingkat kepentingan kata tersebut (Chen et al., 2024). Semakin tinggi bobot suatu istilah, maka semakin besar kontribusinya terhadap skor relevansi dokumen.

Semantik merupakan metode dari *retrieval* yang memahami makna dari pertanyaan pengguna dengan dokumen (Fathuddin et al., 2025). Metode ini bekerja dengan mengubah teks menjadi sebuah *vector* angka *(embedding),* sehingga kesamaan antara pertanyaan pengguna dan dokumen dapat dihitung dengan kemiripan *vector* tersebut. Dengan pendekatan ini yang memiliki makna serupa dengan pertanyaan pengguna tetap dianggap relevan meski tidak memiliki kecocokan kata secara langsung.

BM25

BM25 *(Best Matching 25)* merupakan salah satu metode yang dimiliki oleh *retrieval* yang bekerja dengan melakukan pencocokan kata (*retrieval* leksikal) antara pertanyaan pengguna dengan dokumen. BM25 menghitung nilai relevansi dari setiap dokumen terhadap pertanyaan pengguna berdasarkan kemunculan kata, keunikan kata, serta panjang dokumen (Bintana, 2012). Dalam penerapannya, BM25 digunakan setelah tahap *prepocesing teks*. Berdasarkan kemampuannya, BM25 mampu memberikan performa yang stabil, sehingga sistem dapat menyaring kandidat dokumen awal sebelum diproses lebih lanjut menggunakan metode *semantic retrieval, reranking,* dan LLM *(Large Language Model)*

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Bi – Encoder

Bi – Encoder merupakan model *semantic retrieval* berbasis *embedding* yang bertugas untuk mengukur kesamaan makna antara pertanyaan pengguna dengan dokumen(Kasmaee, 2025). Pada pendekatan *Bi – encoder,* pertanyaan pengguna dengan dokumen diproses dengan terpisah untuk menghasilkan representasi *vector*, sehingga pencarian bisa dapat dilakukan secara efisien. *Bi – Encoder* digunakan setelah tahap leksikal untuk meningkatkan relevansi hasil pencarian dengan mempertimbangkan *semantic* sebelum melakukan *reranking.*

Cross – Encoder

Cross Encoder atau sering disebut *reranking* merupakan rangkaian model pemeringkatan tingkatan lanjut yang dikembangkan di BAAI *(Beijing Academy of Artificial Intelliegence)* dan tahap ini digunakan pada *reranking* dalam system *retrieval.* Model ini bekerja dengan memproses *query* dan dokumen secara bersamaan untuk menghasilkan skor relevansi langsung (Lu et al., 2025). Dengan mekanisme tersebut, model ini efektif dalam melakukan penyaringan dan pengurutan ulang dokumen berdasarkan tingkat relevansinya sebelum diberikan ke model LLM.

LLM

LLM *(Large Language Model)* merupakan model kecerdasan buatan yang dirancang untuk memahami dan membuat teks dengan bahasa alami. LLM memiliki kapasitas parameter yang besar, sehingga mampu memberikan jawaban yang interaktif dan *respondsive* yang menyerupai bahasa manusia (Helviansyah et al., 2025). Model ini banyak digunakan dalam pengembangan sistem, contohnya *chatbot* untuk memberikan jawaban yang informatif dan *respondsive*.

Evaluasi Model

Evaluasi model yang digunakan pada penelitian ini menggunakan tiga metrik evaluasi yaitu *Recall@k, Precision@k, dan nDCG@k*. Ketiga metrik ini sama – sama digunakan untuk membantu evaluasi sistem.

Recall@k merupakan metrik evaluasi yang digunakan dalam informasi *retrieval* untuk mengukur sejauh mana sistem berhasil menemukan dokumen atau item yang relevan terhadap suatu pertanyaan(Andina et al., 2025). Metrik ini menghitung hasil dokumen yang muncul di *top – k* yang dikembalikan oleh sistem *retrieval.*

Recall@k digunakan pada tahap evaluasi sistem *retrieval* dan *reranking,* khususnya setelah sistem berhasil menghasilkan daftar dokumen hasil pencarian berdasarkan pertanyaan pengguna. *Recall@k* digunakan untuk mengetahui kemampuan sistem *retrieval* dan *reranking* dalam menemukan dokumen relevan terhadap pertanyaan pengguna.

Recall@k dihitung dengan membandingkan jumlah dokumen relevan yang berhasil ditemukan pada *top – k* hasil pencarian, dengan jumlah dokumen relevan yang tersedia yang telah diberikan lebel secara manual *(ground truth)*.

Secara matematis, *Recall@k* dirumuskan sebagai berikut:

$$Recall@k = \frac{|Rel \cap Top_k|}{|Rel|} \quad \text{.....(1)}$$

dimana:

Rel = Seluruh dokumen yang relevan sebenarnya *(ground truth)*

Top_k = k dokumen teratas hasil sistem

$Rel \cap Top_k$ = Dokumen relevan yang berhasil ditemukan *top-k*

Nilai dari *Recall@k* berada pada rentang 0 sampai 1. Semakin tinggi nilai *ricall@k* maka semakin baik kemampuan sistem dalam menemukan dokumen relevan.

Precision@k merupakan evaluasi metrik untuk mengukur ketepatan hasil *retrieval,* yang menunjukkan sejauh mana sistem mampu menampilkan dokumen relevan pada hasil pencarian teratas(Andina et al., 2025). Metrik ini digunakan pada tahap evaluasi *retrieval* dan *reranking*, khususnya pada saat setelah proses pengambilan *top – k* dokumen.

Precision@k sangat relevan ketika sistem hanya menampilkan sedikit hasil terbaik, karena metrik ini berfokus pada kualitas hasil yang ditampilkan pada posisi teratas oleh sistem. Oleh karena itu, *Precision@k* digunkan untuk mengukur ketepatan hasil *retrieval* dan *reranking* yang ditampilkan.

Precision@k biasanya digunakan bersamaan dengan *Recall@k* untuk mendapatkan evaluasi yang seimbang. *Precision@k* bekerja dengan cara membandingkan jumlah dokumen relevan yang berhasil ditemukan dalam *top – k* ,dengan jumlah total dokumen yang ditampilkan, yaitu sebanyak k dokumen.

Secara matematis, *Precision@k* dirumuskan sebagai berikut:

$$Precision@k = \frac{|Rel \cap Top_k|}{k} \quad \text{.....(2)}$$

dimana:

Rel = Seluruh dokumen yang relevan sebenernya *(ground truth)*

Top_k = k dokumen teratas hasil sistem

$Rel \cap Top_k$ = Dokumen relevan yang berhasil ditemukan *top-k*

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

k = Jumlah total dokumen yang diambil

Nilai dari *Precision@k* berada pada rentang 0 sampai 1. Semakin tinggi nilai *precision@k* maka semakain baik ketepatan sistem dalam menampilkan dokumen yang relevan.

nDCG@k *(Normalized Discounted Cumulative Gain)* merupakan metrik evaluasi kualitas sistem pada kemampuan untuk mengurutkan item berdasarkan relevansi(Evidently AI, 2026). Metrik ini adalah matrik kualitas pemeringkatan, karena tidak hanya memperhatikan apakah dokumen relevan ditemukan, tetapi juga metrik ini mempertimbangkan peringkat posisi dokumen dalam hasil pencarian.

nDCG@k digunakan pada evaluasi hasil retrieval dan *reranking,* khususnya setelah sistem menghasilkan daftar *top – k* dokumen. Metrik ini digunakan untuk mengevaluasi kualitas hasil pemeringkatan dokumen setelah tahap *reranking.*

nDCG@k digunakan karena mampu mengukur kualitas urutan dokumen. Metrik ini memberikan bobot lebih besar pada dokumen yang muncul pada posisi teratas, sehingga sesuai untuk mengevaluasi sistem *retrieval* dan *reranking* berbasis pemeringkatan.

nDCG@k bekerja dengan menghitung nilai DCG@k *(Discounted Cumulative Gain)* dari hasil pemeringkatan sistem, dimana DCG memberikan bobot nilai pada dokument teratas dan dokumen yang berada pada posisi lebih bawah akan diberikan pengurangan nilai, kemudian membandingkan nilai IDCG@k *(Ideal Discounted Cumulative Gain),* yaitu nilai yang diberikan lebel dari hasil *reranking.*

Secara matematis, nDCG@k dirumuskan sebagai berikut:

$$\text{nDCG@k} = \frac{DCG@K}{IDCG@K} \qquad \text{.....(3)}$$

dimana:

nDCG@k = Seberapa dekat urutan sistem ke urutan ideal (0-1)

DCG@k = Hasil urutan sistem

IDCG@K = DCG ideal berdasarkan label

METODE PENELITIAN

Tahap Penelitian

Jenis penelitian yang digunakan dalam penelitian ini adalah penelitian terapan. penelitian terapan ini berfokus pada penerapan dan pengujian mekanisme kerja model *hybrid Retrieval–Augmented Generation* (RAG) dalam sistem chatbot informasi pariwisata di Bali. Penelitian terapan ini dipilih karena bertujuan untuk menerapkan model *retrieval* dan *reranking* serta mengenalisis alur kerja model *retrieval* dan *reranking* dalam menghasilkan informasi yang relevan.

Pendekatan penelitian yang digunakan pada penelitian ini adalah pendekatan kuantitatif dengan pendekatan pembelajaran mesin sebagai pendukung.

Pendekatan pembelajaran mesin digunakan untuk menerapkan dan mengkaji mekanisme kerja model *hybrid* RAG *(Retrieval Agumented Generation)*, yang meliputi proses *retrieval* BM25 dan *Bi–Encoder*, serta proses *reranking* menggunakan *Cross Encoder*.

Sementara itu, Pendekatan kuantitatif digunakan untuk mengukur kinerja model secara objektif melalui nilai numerik yang dihasilkan dari proses evaluasi model, seperti skor evaluasi model dan hasil pengukuran *precision@k* dan *recall@k.*

Objek Penelitian

Objek penelitian ini adalah sistem chatbot informasi pariwisata Bali berbasis Hybrid Retrieval-Augmented Generation (Hybrid RAG). Sistem dirancang untuk memberikan informasi mengenai destinasi wisata, budaya, akomodasi, dan informasi pendukung pariwisata di Bali berdasarkan data yang diperoleh dari berbagai sumber.

Bahan dan Alat Utama

Perangkat lunak yang digunakan meliputi Python sebagai bahasa pemrograman, FastAPI sebagai framework backend, React sebagai framework frontend, Gemini sebagai Large Language Model (LLM), Model BM25, Bi-Encoder, dan Cross-Encoder

Pengumpulan Data

Pengumpulan data dilakukan melalui studi literatur dan pengumpulan data set. Dataset penelitian terdiri dari 200 dokumen informasi pariwisata di Bali yang diperoleh dari tiga sumber yaitu: hasil *web scraping* pada situs-situs informasi pariwisata di Bali, dataset dari Kaggle yang memuat informasi destinasi wisata di Bali, dan publikasi resmi dari Badan Pusat Statistik Provinsi Bali mengenai sektor pariwisata. Seluruh data dari ketiga sumber tersebut digabungkan menjadi satu basis pengetahuan yang digunakan pada sistem *Hybrid* RAG.

Preprocesing Data

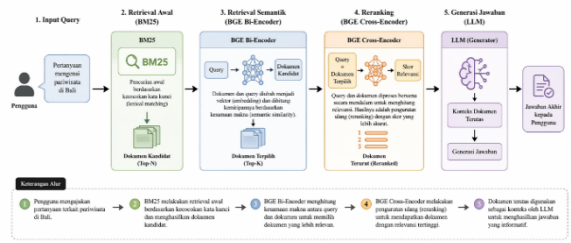
Sebelum digunakan pada sistem, seluruh dokumen melalui tahapa *preprocessing* untuk meningkatkan kualitas data, tahap *preprocessing* meliputi: membersihkan karakteristik khusus HTML dan simbol yang tidak diperlukan, melakukan normalisasi teks agar memiliki fromat seragam, dan memotong beberapa teks menjadi beberapa token.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

Tahap *perprocessing* digunakan agar proses retrieval dapat dilakukan secara lebih efektif serta menghasilkan dokumen yang memiliki kualitas lebih baik.

Pipline Hybrid RAG



Gambar 1. Alur Pipline Hybrid RAG

Pada penelitian ini, menggunakan pendekatan *Hybrid* RAG untuk menghasilkan jawaban berdasarkan pertanyaan pengguna. Alur kerja sistem ini dimulai pada saat pengguna memasukkan pertanyaan terkait pariwisata di Bali. Pertanyaan tersebut kemudian diproses melalui beberapa tahap, yaitu:

1. **Query Input**
   Pengguna memasukan pertanyaan dalam bentuk teks, terkait destinasi wisata di Bali.
2. **Leksikal Retrieval menggunakan BM25**
   Query yang diberikan akan diproses pertama kali menggunakan model BM25 untuk melakukan pencarian awal terhadap dokumen.
   Pada tahap ini, pencarian dilakukan berdasarkan kecocokan kata kunci *(keyword machine)* sehingga menghasilkan sekumpulan kandidat dokumen yang relevan secara leksikal.
3. **Semantic Retrieval menggunakan Bi-Encoder**
   Selanjutnya, hasil dari dokumen bm25 akan diproses menggunakan *Bi-Encoder.* Pada tahap ini, query dan dokumen akan diubah menjadi *veketor (embedding)*, kemudian akan dihitung kemiripannya berdasarkan kesamaan makna *(semantic similarity)* yang di mana perhitungan yang *query* dengan dokumen paling dekat maka itu akan berada di top@k. Hasilnya adalah dokumen yang lebih relevan secara konteks.
4. **Reranking menggunakan Cross-Encoder**
   Hasil dari *Semantic* akan diproses ke *reranking* menggunakan *Cross-Encoder*. Pada tahap ini, *query* dan dokumen akan diproses secara bersamaan dalam satu *input*, sehingga model akan memahami konteks lebih mendalam.
   Hasil dari tahap ini adalah urutan dokumen yang telah diurutkan ulang berdasarkan skor kemiripan yang lebih akurat.
5. **Respond Generator**
   Dokumen dengan skor tertinggi akan dipilih sebagai content, kemudian akan digunakan oleh model LLM untuk menghasilkan jawaban yang lebih informatif dan sesuai dengan pertanyaan pengguna.

Penyusunan Ground Truth

Data pengujian terdiri atas lima *query* yang disusun oleh peneliti berdasarkan kebutuhan informasi pariwisata di Bali. Setiap *query* mewakili topik pencarian yang tersedia pada *knowledge base* dan digunakan sebagai data uji dalam proses evaluasi sistem. Untuk setiap *query*, ditentukan dokumen-dokumen yang dianggap relevan sebagai ground truth, yang selanjutnya digunakan sebagai acuan dalam menghitung nilai *Recall@k*, *Precision@k*, dan nDCG@k. Untuk setiap query, seluruh dokumen pada basis data diperiksa secara manual oleh peneliti untuk menemukan dokumen yang relevan. Dokumen yang memenuhi kebutuhan informasi dari pertanyaan uji diberi label sebagai *ground truth* dan digunakan sebagai acuan dalam proses *retrieval*.

Validasi Ground Truth

Ground truth yang telah disusun kemudian divalidasi oleh seorang pakar di bidang pariwisata. Proses validasi dilakukan menggunakan lembar validasi yang berisi daftar *query* beserta dokumen yang telah ditentukan oleh peneliti.

Konfigurasi Eksperimen

Konfigurasi eksperimen digunakan untuk mengetahui pengaruh jumlah dokumen pada setiap tahap *retrieval* terhadap hasil pencarian. *Pipline Hybrid* RAG dijalankan secara bertahap menggunakan konfigurasi yang berbeda.

Tahap pertama menggunakan BM25 untuk mengambil kandidat dokumen (BM25 Top-k). Selanjutnya dokumen tersebut diproses menggunakan Bi-Encoder untuk memperoleh dokumen dengan kesamaan semantik tertinggi (Bi-Encoder Top-k). Hasil semantic retrieval kemudian diproses kembali menggunakan Cross-Encoder sehingga diperoleh dokumen dengan skor relevansi tertinggi (Final Top-k) yang digunakan sebagai konteks oleh model Gemini.

Selain parameter Top-k, penelitian ini juga menguji beberapa nilai temperature pada model Gemini untuk mengetahui pengaruhnya terhadap proses generasi jawaban.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

Metode Evalusi Model

1. Skenario Eksperimen

Setup skenario Eksperimen yang akan digunakan dalam penelitian ini. Setup eksperimen terdiri dari jumlah dokumen *top_k* yang digunakan dari setiap model, dan temperatur dari generasi jawaban. Variasi tersebut bertujuan untuk menganalisis pengaruh konfigurasi *retrieval* dan generasi jawaban terhadap performa sistem. Hasil dari setiap eksperimen akan dievaluasi menggunakan metrik *Recall@k*, *Precision@k*, dan nDCG@k.

Tabel 1. Skenario Eksperimen

Eksperimen	BM25_top_k	Bi_Encoder_top_k	Final_top_k	Temperature
EXP_001	20	10	5	0.2
EXP_002	30	10	5	0.5
EXP_003	20	15	5	0.2

2. Prosedur Evaluasi Model

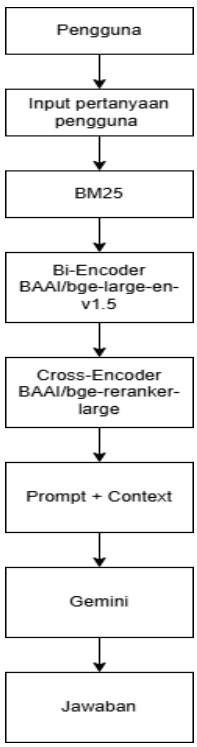
Setiap konfigurasi eksperimen menghasilkan sejumlah dokumen hasil *retrieval.* Dokumen tersebut kemudian dibandingkan dengan *ground truth* yang telah divalidasi oleh pakar.

Evaluasi dilakukan menggunakan tiga metrik, yaitu: Recall@k, Precission@k, dan nDCG@k. Selain itu dilakukan juka pengukuran waktu respons sistem untuk mengetahui efisiensi proses *retrieval* hingga generasi jawaban.

Konfigurasi yang memproleh nilai *Recall@k, Precission@k*, dan nDCG@k terbaik dipilih sebagai konfigurasi optimal pada sistem *Hybrid* RAG

HASIL DAN PEMBAHASAN

Hasil Implementasi Sistem



Gambar 2. Diagram Alur Pipline Hybrid RAG

Gambar 2 menunjukkan implementasi sistem *pipline Hyrbid* RAG pada chatbot informasi pariwisata di Bali. Sistem menerima pertanyaan pengguna kemudian memprosesnya secara bertahap mulai BM25, *Bi-Encoder, Cross-Encoder*, dan model Gemini sebagai LLM untuk menghasilkan jawaban.

BM25 yang diimplementasikan melalui *LangChain* sebagai metode *retrieval* berbasis kata kunci untuk memperoleh 10 dokumen awal yang paling relevan terhadap *query* pengguna. Dokumen hasil *retrieval* kemudian diproses menggunakan model *Bi-Encoder* BAAI/bge-large-en-v1.5 dengan mengubah *query* dan dokumen ke dalam representasi *embedding*. Tingkat kemiripan semantik dihitung sehingga diperoleh lima dokumen dengan nilai kemiripan tertinggi. Selanjutnya, lima dokumen tersebut *direranking* menggunakan model *Cross-Encoder* BAAI/*bge-reranker-large* dengan memasangkan setiap dokumen terhadap *query* pengguna. Proses *reranking* menghasilkan tiga dokumen dengan tingkat relevansi tertinggi yang kemudian digunakan sebagai konteks bagi model LLM dalam menghasilkan jawaban.

Hasil Evaluasi Model

Tabel 2. Hasil Konfigurasi Seluruh Eksperimen

Experiment_id	Recall@k	Precissio n@k	nDCG@k
EXP_001	0.33	0.8	0.94
EXP_002	0.17	0.49	0.99
EXP_003	0.38	0.65	0.91

Berdasarkan tabel 3, konfigurasi exp_001 memperoleh nilai Recall@k sebesar 0,33, Precison@k sebesar 0,8, dan nDCG@k sebesar 0,94. Nilai *Recall@k* menunjukkan sistem mampu memperoleh dokumen relevan sekitar 33% yang terdapat pada *ground truth*. Meskipun nilai *Recall@k* masih relatif rendah, nilai *Precisison@k* yang mencapai 0,80 menunjukkan bahwa sebagian besar dokumen yang

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ditampilkan pada akhir merupakan dokumen yang relevan. Kondisi tersebut penting pada sistem RAG, karena kualitas dokumen yang diberikan pada LLM lebih menentukan kualitas jawaban dibandingkan jumlah dokumen yang berhasil ditemukan.

Nilai nDCG@k sebesar 0,94 menunjukkan bahwa dokumen yang paling relevan telah ditempatkan pada urutan teratas hasil *retrieval*. Hal ini mengindikasikan bahwa proses *reranking* menggunakan *Cross-Encoder* mampu mengurutkan kembali dokumen berdasarkan tingkat relevansi terhadap pertanyaan pengguna.

Sebaliknya exp_002 menghasilkan nilai *Precision@k* sebesar 0,49, dan nilai nDCG@k sebesar 0,99 menunjukkan bahwa dokumen yang relevan berhasil ditempatkan pada urutan awal, namun jumlah dokumen yang relevan lebih sedikit dibandingkan dengan exp_001. Sementara itu, exp_003 memperoleh *Recall@k* sebesar 0,38 yang menunjukkan kemampuan sistem menemukan lebih banyak dokumen relevan, tetapi *Precision@k* dan nDCG@k mengalami penurunan sehingga kualitas dokumen yang diberikan kepada LLM menjadi kurang konsisten.

Berdasarkan tiga metrik tersebut, exp_001 dipilih sebagai konfigurasi terbaik karena memiliki keseimbangan antara kemampuan menemukan dokumen relevan, ketepatan dokumen yang ditampilkan, serta pengurutan dokumen sebelum digunakan sebagai konteks oleh model Gemini.

Analisis Waktu Respons

Tabel 3. Analisis Pengujian Waktu Respons

Query	Waktu Respond (detik)
Q1	76.89
Q2	56.2
Q3	54.4
Q4	87.96
Q5	56.4
Rata - rata	66.37

Berdasarkan pengujian terhadap lima query, sistem memperoleh rata-rata waktu respons sebesar 66,37 detik. Waktu tercepat diperoleh pada query Q3 sebesar 54,40 detik, sedangkan waktu terlama terdapat pada query Q4 sebesar 87,96 detik. Variasi waktu respons dipengaruhi oleh proses retrieval menggunakan BM25, perhitungan kemiripan semantik menggunakan Bi-Encoder, proses reranking menggunakan Cross-Encoder, serta generasi jawaban oleh model Gemini. Hasil ini menunjukkan bahwa sistem mampu menghasilkan jawaban secara berurutan melalui seluruh tahapan pipeline Hybrid RAG, meskipun waktu respons masih relatif tinggi dan dapat menjadi fokus optimasi pada penelitian selanjutnya.

Analisis Pipline Hybrid RAG

Untuk memahami kontribusi setiap tahapan pada pipeline Hybrid RAG, dilakukan analisis terhadap salah satu query uji, yaitu *"Tempat wisata di Bali yang vibenya hijau"*. Analisis dilakukan dengan mengamati perubahan hasil retrieval pada setiap tahap, mulai dari BM25, Bi-Encoder, hingga Cross-Encoder.

Tabel 4. Hasil Top-k retrieval B25

Rank	BM25 Doc_id	Title	BM25 Skor
1	59	Mahapraja Bangli	10.48163
2	6	Desa Wisata Penarungan	10.30645
3	102	Pantai Karang Sewu	10.27559
4	62	Air Terjun Desa Kutuh Bangli	10.23002
5	58	Tibumana Waterfall	10.19029

BM25 menghasilkan lima dokumen kandidat yang memiliki keterkaitan kata kunci dengan *query*. Dokumen seperti Mahapraja Bangli dan Desa Wisata Penarungan memperoleh skor tertinggi karena mengandung istilah yang memiliki kemiripan dengan kata-kata pada *query*. Namun, hasil *retrieval* masih didasarkan pada kecocokan kata sehingga belum sepenuhnya merepresentasikan makna dari pertanyaan pengguna. Akibatnya dokumen yang memiliki karakteristik alam atau suasana hijau, tetapi tidak menggunakan istilah yang sama dengan *query* belum tentu berada pada posisi teratas. Oleh karena itu, BM25 masih memerlukan pemrosesan *semantic retrieval* untuk meningkatkan relevansi dokumen yang akan digunakan sebagai konteks.

Tabel 5. Hasil Top-k retrieval Bi-Encoder

Rank	Bi – Encoder Doc_id	Title	Bi – Encoder Skor
1	77	Desa Wisata Pule	0.775484
2	67	Pelisan Waterfalls	0.742001

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

3	60	Langgahan Waterfalls	0.742001
4	57	Pengibul Waterfall	0.739367
5	55	Desa Adat Panglipuran	0.73497

Setelah dilakukan pemeringkatan menggunakan Bi-Encoder, terjadi perubahan komposisi dokumen. Desa Wisata Pule berpindah ke peringkat pertama, sedangkan beberapa dokumen yang sebelumnya berada pada posisi atas pada tahap BM25 tidak lagi muncul dalam lima besar. Perubahan ini menunjukkan bahwa Bi-Encoder tidak hanya mempertimbangkan kesamaan kata, tetapi juga menghitung kemiripan makna melalui representasi *embedding.* Dengan demikian dokumen yang memiliki suasana alam, vegetasi, atau lingkungan hijau dapat memperoleh skor yang lebih tinggi meskipun tidak menngandung frasa yang sama. Hasil ini menunjukkan bahwa *semantic* retrieval mampu meningkatkan kualitas kandidat dokumen sebelum dilakukan proses *rerangking.*

Tabel 6. Hasil Top-k Reranking Cross-Encoder

Rank	**Cross – Encoder Doc_id**	**Title**	**Cross – Encoder Skor**
1	67	Pelisan Waterfalls	1.829226
2	60	Langgahan Waterfalls	1.829226
3	55	Desa Adat Panglipuran	-0.0953
4	57	Pengibul Waterfall	-0.30586
5	77	Desa Wisata Pule	-0.94669

Pada tahap Cross-Encoder terjadi perubahan urutan dokumen kembali. Pelisan Waterfalls menempati peringkat pertama, sedangkan Desa Wisata Pule turun ke posisi terakhir. Perubahan tersebut menunjukkan bahwa Cross-Encoder melakukan evaluasi yang lebih mendalam terhadap pasangan query dan dokumen sehingga menghasilkan pemeringkatan yang lebih akurat dibandingkan tahap sebelumnya. Proses reranking ini menghasilkan urutan dokumen yang lebih sesuai dengan maksud pertanyaan pengguna. Sehingga dokumen yang digunakan sebagai konteks model Gemini memiliki tingkat relevansi yang lebih tinggi. Dengan demikian, proses reranking berperan dalam meningkatkan kualitas konteks yang diberikan kepada LLM sebelum menghasilkan jawaban

KESIMPULAN

Penelitian ini berhasil mengimplementasikan chatbot informasi pariwisata di Bali *berbasis* Hybrid RAG yang mengintegrasikan BM25 sebagai *lexical retrieval,* Bi-Encoder sebagai *semantic retrival*, *Cross-Encoder* sebagai *reranking*, dan model Gemini sebagai LLM (*Large Language Model*) untuk menghasilkan jawaban.

Berdasarkan hasil tersebut tiga konfigurasi tersebut, konfigurasi terbaik memperoleh nilai rata-rata metrik *Recall@k* sebesar 33%, *Precision@k* sebesar 80%, dan nDCG@k sebesar 94%. Nilai *Precision@k* menunjukkan bahwa sebagian besar dokumen yang digunakan sebagai konteks merupakan dokumen yang relevan, sedangkan nilai nDCG@k menunjukkan bahwa dokumen dengan tingkat relevansi tinggi berada pada posisi teratas. Nilai *Recall@k* menunjukkan bahwa sistem masih belum mampu menemukan seluruh dokumen relevan yang terdapat pada *ground truth*, sehingga masih terdapat peluang untuk meningkatkan cakupan *retrieval* pada penelitian selanjutnya.

Analisis pada pipline Hybrid RAG menunjukkan bahwa BM25 berperan dalam menghasilkan kandidat dokumen berdasarkan kecocokan kata kunci, *Bi-Encoder* penyaringan berdasarkan kesamaan makna, dan *Cross-Encoder* mengurutkan kembali urutan dokumen berdasarkan relevansi terhadap *query.*

Dengan demikian, pendekatan *Hybrid* RAG berhasil diterapkan pada *chatbot* informasi pariwisata di Bali dan mampu menghasilkan proses *retrieval* serta reranking yang menyediakan konteks relevansi bagi model Gemini dalam menghasilkan jawaban berdasarkan basis pengetahuan yang digunakan.

DAFTAR PUSTAKA

Abdurrazzaq, M. A., Tjiong, E. L., Fasya, A., Hiu, M., & Tanuwidjaya, J. (2025). An Indonesian Chatbot for Disease Diagnosis Using Retrieval-Augmented Generation. *Journal of Innovation and Technology Polbeng Series on Informatics (INOVTEK Polbeng-Seri Informatika), 10*(3), 1877-1887. https://doi.org/10.35314/9nnkn955

Agiyola, T., & Indriati, B. R. (2019). Relevance Feedback Pada Sistem Temu Kembali Informasi Dokumen E-Book Berbahasa Indonesia Menggunakan Metode BM25. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer E-ISSN*, *2548*(5), 964X.

Andina, S. P., Pramukantoro, E. S., & Yahya, W. (2025). Studi Kinerja TF-IDF dan IndoBERT

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

dalam Rekomendasi Resep Berbasis Mobile dengan Arsitektur Microservice. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 9*(10).

Aprilio, P., Felix, M., Nugraha, P. S., & Fahmi, H. (2025). Hybrid Feature Combination of TF-IDF and BERT for Enhanced Information Retrieval Accuracy. *JISA(Jurnal Informatika Dan Sains)*, *8*(1), 8–15. https://doi.org/10.31326/jisa.v8i1.2179

Bintana, R. R. (2012). Penerapan Model Okapi Bm25 Pada Sistem Temu Kembali Informasi. *Seminar Nasional Teknologi Informasi Komunikasi Dan Industri (SNTIKI) 4*, 273–279.

Chen, J., Xiao, S., Zhang, P., Luo, K., Lian, D., & Liu, Z. (2024). M3-Embedding: Multi-Linguality, Multi-Functionality, Multi-Granularity Text Embeddings Through Self-Knowledge Distillation. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 2318–2335. https://doi.org/10.18653/v1/2024.findings-acl.137

Cuconasu, F., Trappolini, G., Siciliano, F., & Filice, S. (2024). The Power of Noise: Redefining Retrieval for RAG Systems. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '24), July 14â•fi18, 2024, Washington, DC, USA* (Vol. 1, Number 1). Association for Computing Machinery. https://doi.org/10.1145/3626772.3657834

Evidently AI. (2026). *Normalized Discounted Cumulative Gain (NDCG) explained*. https://www.evidentlyai.com/

Fathuddin, M. A. H., Mandyartha, E. P., & Nurlaili, A. L. (2025). Penerapan Sentence-Bert dan Cosine Similarity untuk Pencarian Semantik Dokumen Skripsi dalam Format PDF. *Ranah Research: Journal of Multidisciplinary Research and Development*, *8*(1), 322–337. https://doi.org/10.38035/rrj.v8i1.1865

Helviansyah, T., Harahap, N. S., Irsyad, M., & Negara, B. S. (2025). Sistem Tanya Jawab Berbasis Chatbot Website Menggunakan Gemini Ai Pada Data Fiqih Kontemporer. *Journal of Information System Management (JOISM)*, *7*(1), 38–47. https://doi.org/10.24076/joism.2025v7i1.2082

Kasmaee, A. S. (2025). *Domain-Specific Text Embedding Models for Information Retrieval*.

Lu, M., Chen, C., & Eickhoff, C. (2025). *Cross-Encoder Rediscovers a Semantic Variant of BM25*.

Luan, Y., Eisenstein, J., Toutanova, K., & Collins, M. (2021). Sparse, dense, and attentional representations for text retrieval. *Transactions of the Association for Computational Linguistics*, *9*, 329–345. https://doi.org/10.1162/tacl_a_00369

Lubis, M. S. Y. (2021, August). Implementasi artificial intelligence pada system manufaktur terpadu. In *Prosiding Seminar Nasional Teknik UISU (SEMNASTEK)* (Vol. 4, No. 1, pp. 1-7).

Mikael, K., Oz, C., Rashid, T. A., & Nariman, G. S. (2025). A Hybrid Chatbot Model for Enhancing Administrative Support in Education: Comparative Analysis, Integration, and Optimization. *IEEE Access*, *13*(February), 50741–50760. https://doi.org/10.1109/ACCESS.2025.3552501

Pais, S., Cordeiro, J., & Jamil, M. L. (2022). *NLP-based platform as a service a brief* (pp. 1–26).

Sanjaya, R. A., & Winarno, E. (2024). Pengembangan Chatbot Informasi Pariwisata di Kabupaten Pati Menggunakan Metode Natural Language Processing Berbasis Dialogflow. *Jutisi : Jurnal Ilmiah Teknik Informatika Dan Sistem Informasi*, *13*(1), 368. https://doi.org/10.35889/jutisi.v13i1.1828

Situmeang, R. N., Jamaluddin, Purba, E. N., & Gunawati Harianja, E. J. (2023). Designing Chatbot About Tourism Recommendations In North Sumatra. *2023 International Conference of Computer Science and Information Technology (ICOSNIKOM)*, 1–4. https://doi.org/10.1109/ICoSNIKOM60230.2023.10364534

Su, C., Wen, J., Kang, J., Wang, Y., Pan, H., & Hossain, M. S. (2024). Hybrid rag-empowered multi-modal llm for secure healthcare data management: A diffusion-based contract theory approach. *arXiv preprint arXiv:2407.00978*.

Syah, M. P., Wardani, A. P., Idhom, M., & Trimono, T. (2025). Perbandingan representasi teks TF-IDF dan BERT terhadap akurasi cosine similarity dalam penilaian otomatis jawaban berbasis teks. *Data Sciences Indonesia (DSI), 5*(1), 47-59. https://doi.org/10.47709/dsi.v5i1.6021

Thakur, N., Reimers, N., Rücklé, A., Srivastava, A., & Gurevych, I. (2021). BEIR: A Heterogeneous Benchmark for Zero-shot Evaluation of Information Retrieval Models. *Advances in Neural Information Processing Systems*. https://doi.org/10.48550/arXiv.2104.08663

Trung, Q. H., Hoang, L. T., & Phuc, N. V. H. (2024). *Optimizing Multi-Stage Language Models for Effective Text Retrieval*.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.