

Analisis Perbandingan Metode *Information Retrieval* (IR) Pada Implementasi Pencarian Ayat Al-Qur'an

M. Fajratul Ikhsan^{1*}, Shofwatul 'Uyun²

^{1,2} Magister Informatika, Universitas Islam Negeri Sunan Kalijaga Yogyakarta, Indonesia

¹ ikhsanm288@gmail.com, ² shofwatul.uyun@uin-suka.ac.id

Diajukan: 3 Januari 2026 | Direvisi: 26 Januari 2026 | Diterima: 27 Januari 2026 | Diterbitkan: 30 Januari 2026

Abstrak

Perkembangan kecerdasan buatan (AI) yang pesat telah mendorong transformasi layanan digital di berbagai bidang, termasuk layanan keagamaan. Penelitian ini mengevaluasi metode *information retrieval* pada sistem pencarian ayat Al-Qur'an dengan membandingkan pendekatan leksikal menggunakan BM25, pendekatan semantik melalui dense retrieval model embedding intfloat/multilingual-e5-base untuk membangkitkan representasi vektor kueri dan dokumen, serta penggabungan keduanya menggunakan metode Reciprocal Rank Fusion (RRF). Eksperimen dilakukan pada korpus ayat Al-Qur'an berformat JSON dengan skenario pencarian top-k ($k=20$). Kinerja sistem diukur menggunakan metrik MRR, nDCG, Recall, dan MAP. Hasil eksperimen menunjukkan bahwa BM25 memiliki cakupan relevansi cukup tinggi dengan Recall sebesar 98%. Sementara itu, dense retrieval memberikan performa yang lebih rendah dengan MRR 60%, nDCG 64%, dan Recall 81%, sehingga berada di bawah BM25 pada sebagian besar metrik. Penggabungan menggunakan RRF menghasilkan kualitas pemeringkatan terbaik dengan MRR 81%, nDCG 85%, MAP 81%, serta Recall tetap tinggi sebesar 97%. Temuan ini menunjukkan bahwa pendekatan penggabungan/RRF memberikan performa terbaik secara keseluruhan, sekaligus menegaskan bahwa kombinasi metode leksikal dan semantik bersifat saling melengkapi untuk meningkatkan relevansi dan kualitas pemeringkatan hasil pencarian ayat Al-Qur'an.

Kata Kunci: Al-Qur'an, BM25, Dense Retrieval, *Information Retrieval*, Pencarian

Abstract

The rapid development of artificial intelligence (AI) has driven the transformation of digital services in various fields, including religious services. This study deploys information retrieval methods in the Quranic verse search system by comparing the lexical approach using BM25, the semantic approach through a dense retrieval model embedded intfloat/multilingual-e5-base to generate query vector representations and documents, and connecting the two using the Reciprocal Rank Fusion (RRF) method. Experiments were conducted on a corpus of Quranic verses in JSON format with a top-k search scenario ($k=20$). System performance was measured using MRR, nDCG, Recall, and MAP metrics. The experimental results show that BM25 has a fairly high relevance coverage with a Recall of 98%. Meanwhile, dense retrieval provides lower performance with an MRR of 60%, nDCG of 64%, and Recall of 81%, thus being below BM25 in most metrics. Merging using RRF yielded the best ranking quality with an MRR of 81%, nDCG of 85%, MAP of 81%, and a consistently high Recall of 97%. These findings demonstrate that the merging/RRF approach provides the best overall performance, while also confirming that the combination of lexical and semantic methods complements each other in improving the relevance and ranking quality of Quranic verse search results.

Keywords: Al-Qur'an, BM25, Dense Retrieval, *Information Retrieval*, search



This work is an open access article and licensed under a Creative Commons Attribution-NonCommercial ShareAlike 4.0 International (CC BY-NC-SA 4.0). Copyright (C) Author's.

1. PENDAHULUAN

Perkembangan teknologi informasi saat ini terjadi dengan sangat cepat dan mencakup hampir seluruh aspek kehidupan, termasuk pendidikan dan layanan keagamaan. Kemajuan ini menciptakan kesempatan untuk menyajikan layanan digital yang lebih efisien, salah satunya adalah sistem pencarian ayat Al-Qur'an yang dapat membantu pengguna menemukan referensi ayat dengan lebih cepat dan praktis dibandingkan dengan metode tradisional. Transformasi digital semacam ini menjadi krusial karena tuntutan akses informasi religius kini tidak hanya terbatas

pada lingkungan kajian resmi, tetapi juga dalam kegiatan belajar mandiri dan pencarian tematik sehari-hari [1].

Dalam praktiknya, mencari ayat secara manual melalui mushaf atau indeks tematik sering kali memakan waktu yang cukup lama, terutama ketika pengguna tidak tahu kata kunci yang tepat, nomor surah, atau konteks ayat. Situasi ini mendorong kebutuhan akan sistem pencarian ayat berbasis komputer yang dapat mengindeks konten Al-Qur'an dan memberikan hasil yang paling relevan berdasarkan kueri yang dimasukkan oleh pengguna. Dengan cara ini, sistem temu kembali informasi (*Information Retrieval/IR*) dapat berfungsi sebagai jembatan untuk mempercepat proses pencarian ayat sesuai dengan kebutuhan pengguna [2]. Namun, pencarian ayat tidak selalu mudah, terutama karena masalah ketidaksesuaian kosakata, yaitu ketika kata yang digunakan oleh pengguna tidak persis sama dengan kata yang muncul dalam teks ayat atau terjemahan. Variasi dalam kosakata, perbedaan bentuk kata, sinonim, serta perbedaan gaya terjemahan dapat menyebabkan sistem berbasis kata kunci melewati ayat yang sebenarnya relevan. Isu ini juga banyak dibahas dalam penelitian pencarian Al-Qur'an, seperti pendekatan perluasan kueri dan pemanfaatan sumber semantik untuk mengurangi ketidaksesuaian kosakata [3].

Pendekatan klasik yang masih banyak digunakan dalam IR adalah pencarian leksikal/*sparse retrieval*, yaitu mencari berdasarkan kecocokan kata antara kueri dan dokumen. Salah satu metode leksikal yang paling terkenal adalah BM25, yang menghitung *relevance* dokumen berdasarkan frekuensi *term*, kelangkaan *term* (IDF), dan normalisasi panjang dokumen. BM25 telah lama menjadi baseline yang kuat dalam berbagai evaluasi IR, termasuk dalam sistem Okapi pada rangkaian evaluasi TREC [4]. Di sisi lain, kemajuan dalam NLP modern memperkenalkan semantik/*dense retrieval* yang berupaya mencocokkan makna, bukan hanya kata. Dalam pendekatan ini, kueri dan dokumen direpresentasikan sebagai vektor *embedding*, dan kemiripan dihitung menggunakan fungsi seperti *dot-product* atau *cosine similarity*. Kerangka *dense retrieval* seperti Dense Passage Retrieval (DPR) menunjukkan bahwa representasi *dense* dapat digunakan untuk *retrieval* kandidat secara efektif, terutama ketika terdapat perbedaan kata tetapi memiliki maksud yang sama [5].

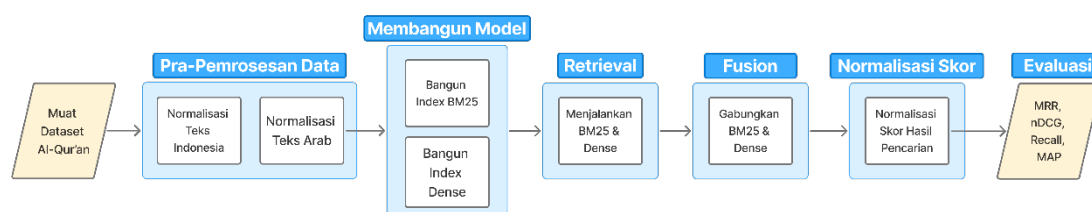
Meskipun demikian, setiap pendekatan memiliki karakteristik yang saling melengkapi. BM25 biasanya lebih unggul ketika kata kueri cocok dengan teks ayat, sementara *dense retrieval* cenderung lebih efektif ketika pengguna menggunakan kueri yang lebih bebas atau melakukan *paraphrase*. Oleh karena itu, strategi penggabungan (*fusion*) sering diterapkan untuk meningkatkan kualitas peringkat dengan memanfaatkan kekuatan kedua sinyal tersebut. Salah satu metode penggabungan yang sederhana namun efektif adalah Reciprocal Rank Fusion (RRF), yang mengkombinasikan peringkat dari beberapa sistem dan terbukti secara konsisten meningkatkan hasil dibandingkan dengan sistem tunggal dalam berbagai eksperimen [6].

Berdasarkan latar belakang tersebut, penelitian ini melakukan eksperimen perbandingan tiga skenario pencarian ayat Al-Qur'an, yaitu BM25 (Leksikal), Dense Retrieval (semantik berbasis *embedding*), serta penggabungan BM25+Dense Retrieval menggunakan Reciprocal Rank Fusion (RRF). BM25 dipilih karena merupakan baseline yang kuat dan paling umum digunakan dalam sistem temu kembali informasi berbasis kata kunci, efektif ketika kueri pengguna memiliki kecocokan kata yang jelas dengan teks ayat, serta relatif stabil dan efisien dalam membangun indeks terbalik (*inverted index*) dan melakukan pemeringkatan. Sementara itu, Dense Retrieval dipilih untuk merepresentasikan perkembangan pemrosesan bahasa alami modern yang berfokus pada pencocokan makna melalui *embedding*, sehingga berpotensi lebih baik ketika terjadi ketidaksesuaian kosakata (misalnya perbedaan kata, sinonim, atau *paraphrase*) antara kueri dan ayat yang relevan. Penggabungan keduanya dengan RRF diuji karena kedua pendekatan ini cenderung saling melengkapi, yaitu BM25 unggul pada kecocokan kata, sedangkan Dense Retrieval unggul pada kedekatan makna. Kebaruan penelitian ini terletak pada

perancangan evaluasi komparatif terkontrol yang membandingkan BM25, Dense Retrieval, dan RRF dalam satu alur pemrosesan yang seragam pada implementasi pencarian ayat Al-Qur'an, mengingat kajian yang membandingkan ketiganya secara langsung pada domain ini masih terbatas. Selanjutnya, evaluasi dilakukan menggunakan metrik pemeringkatan yang umum dalam sistem temu kembali informasi, yaitu MRR, nDCG, Recall, dan MAP untuk menilai kemampuan sistem dalam menemukan ayat relevan sekaligus menempatkannya pada peringkat teratas [7]. Hasil penelitian ini diharapkan memberikan kontribusi pada pengembangan sistem pencarian ayat Al-Qur'an yang lebih akurat dan efektif untuk kebutuhan penelusuran tematik pengguna.

2. METODE PENELITIAN/ALGORITMA

Penelitian ini dilaksanakan melalui serangkaian tahapan yang sistematis dan terstruktur guna mencapai tujuan utama seperti yang ditunjukkan pada gambar 1 dibawah ini :



Gambar 1. Alur Penelitian

2.1. Muat Dataset

Dataset yang digunakan dalam penelitian ini terdiri dari korpus ayat Al-Qur'an yang diperoleh dari Quran API (repositori GitHub fawazahmed0/quran-api), yang kemudian diubah ke dalam format JSON. Dataset ini memuat 6.236 ayat dari 114 surah beserta teks arab dan terjemahan bahasa Indonesia, setiap entri ayat memiliki atribut utama berupa nomor surah, nama surah, nomor ayat, teks arab, terjemahan indonesia. Struktur data yang digunakan pada penelitian ini konsisten yaitu surah, surah_name, ayat, text_ar dan text_id.

Data tersebut disusun secara sistematis agar mudah diproses untuk keperluan pengambilan informasi. Setiap entri ayat mencakup atribut-atribut penting seperti ID surah/ayat, nama surah, nomor ayat, teks dalam bahasa Arab, serta terjemahan dalam bahasa Indonesia, sehingga memungkinkan dilakukan proses pra-pemrosesan. Representasi terstruktur seperti ini umum digunakan dalam penelitian pencarian ayat atau terjemahan Al-Qur'an karena memudahkan pemetaan kueri pengguna terhadap koleksi dokumen dan memungkinkan evaluasi pencarian dilakukan dengan cara yang terukur [8].

2.2. Pra-Pemrosesan Data

Sebelum data diproses, sistem mempersiapkan konten agar seragam dan siap digunakan untuk dua jenis pencarian yaitu lexikal dan semantik. Normalisasi Teks Bahasa Indonesia berarti merapikan teks yang diterjemahkan (contohnya memperbaiki jarak antar kata, mengubah huruf agar seragam, menghapus karakter yang tidak perlu, dan/atau melakukan tokenisasi sederhana). Tujuan langkah ini adalah supaya pencarian berbasis kata (BM25) menjadi lebih konsisten dan tidak terpengaruh oleh perbedaan cara penulisan. Normalisasi teks Bahasa Arab berarti merapikan teks arab agar lebih seragam (menyelaraskan variasi penulisan serta membersihkan simbol-simbol tertentu). Ini membantu dalam menjaga konsistensi penyimpanan dan tampilan hasil, serta mendukung jika teks arab digunakan sebagai bagian dari representasi dokumen [9].

2.3. Membangun Model

Setelah tahap pra-pemrosesan data, sistem menciptakan dua jenis indeks yang terpisah, yaitu membentuk Indeks BM25. Sistem ini membuat indeks inverted atau berbasis istilah dari

korpus (dari teks berbahasa Indonesia), kemudian menyiapkan parameter BM25 sehingga saat pertanyaan diterima, sistem dapat menghitung nilai relevansi berdasarkan frekuensi kata serta penimbangan kata [10].

Sistem Dense Retrieval menggunakan model embedding intfloat/multilingual-e5-base untuk membangkitkan representasi vektor kueri dan dokumen (terjemahan Indonesia/text_id). *Embedding* yang dihasilkan berdimensi 768 dan dinormalisasi (L2). Kemiripan dihitung menggunakan cosine similarity (diimplementasikan sebagai inner product pada vektor ternormalisasi), dan pencarian dilakukan menggunakan indeks FAISS IndexFlatIP untuk mengambil top-k hasil [11]. Tujuan dari metode ini adalah agar pencarian bisa dilakukan berdasarkan kesesuaian makna, bukan hanya sekadar kesesuaian kata [12]. Pada skenario Dense Retrieval, sistem melakukan pencarian semantik dengan merepresentasikan kueri dan dokumen ke dalam ruang vektor menggunakan model embedding (bi-encoder). Berbeda dari BM25 yang mengandalkan kecocokan kata, pendekatan ini menilai relevansi berdasarkan kedekatan makna antara kueri dan ayat.

Setiap dokumen ayat diberi identitas konsisten doc_id = surah:ayat. Pada tahap *offline*, sistem menghasilkan *embedding* dokumen untuk seluruh ayat, jika diperlukan dilakukan *pooling* (misalnya rata-rata token) dan normalisasi vektor agar perhitungan kemiripan lebih stabil. Pada waktu pencarian, kueri pengguna diproses dengan *encoder* yang sama untuk menghasilkan *embedding* kueri. Sistem kemudian menghitung skor kesamaan (misalnya *cosine similarity*) antara *embedding* kueri dan *embedding* dokumen, lalu mengambil top-k dokumen dengan skor tertinggi dari indeks vektor sebagai hasil peringkat Dense Retrieval. Untuk mendukung replikasi, penelitian ini memaparkan konfigurasi Dense Retrieval secara eksplisit, meliputi nama model *embedding*, dimensi *embedding*, teks yang di-embed, metode *pooling*/normalisasi, fungsi kemiripan, jenis indeks vektor (FAISS), parameter utama indeks, serta nilai top-k yang digunakan pada evaluasi [13].

2.4. Retrieval (Pengambilan Informasi)

Pada tahap *retrieval*, ketika pengguna memasukkan kueri, sistem menjalankan dua pencarian secara paralel. Retrieval BM25 menghasilkan daftar kandidat ayat berdasarkan kecocokan kata kunci (leksikal *match*) beserta skor BM25. Dense Retrieval menghasilkan daftar kandidat ayat berdasarkan kemiripan *embedding* (semantik *match*) beserta skor *similarity* (misalnya *cosine/dot-product*). Tahap ini menghasilkan dua daftar peringkat (ranking) terpisah untuk kueri yang sama.

2.5. Fusion (Penggabungan)

Pada tahap ini penggabungan/RRF atau menggabungkan dua hasil retrieval (BM25 dan Dense Retrieval) agar sistem memanfaatkan kelebihan keduanya. Pada sistem ini, penggabungan dapat dilakukan dengan Reciprocal Rank Fusion (RRF) yaitu dokumen/ayat yang muncul di peringkat tinggi pada salah satu atau kedua metode akan memperoleh skor gabungan yang lebih baik. Dengan penggabungan/RRF, ayat yang “kuat secara kata” (BM25) dan ayat yang “kuat secara makna” (Dense) Retrieval dapat sama-sama terdorong ke posisi atas, sehingga kualitas ranking akhir cenderung meningkat.

2.6. Normalisasi Skor Hasil Pencarian

Karena skor BM25 dan skor Dense Retrieval berada pada skala yang berbeda (BM25 bisa tak terbatas, sementara *similarity* Dense Retrieval biasanya berada pada rentang tertentu), sistem melakukan normalisasi agar skor menjadi lebih sebanding, proses penggabungan/analisis lebih stabil, dan interpretasi hasil lebih mudah. Dalam implementasinya, normalisasi dilakukan per kueri (misalnya *min-max scaling*). Untuk Dense Retrieval, skor *similarity* sering diubah dulu ke

rentang 0–1, sedangkan BM25/RRF dinormalisasi langsung per-kueri agar tidak bias akibat variasi rentang skor antar kueri [14].

2.7. Evaluasi

Tahap akhir yaitu evaluasi model atau tahap pengujian untuk menilai performa sebuah model pembelajaran mesin atau statistik dengan menggunakan data tertentu dan metrik yang sesuai. Evaluasi membantu menentukan kualitas model, mengidentifikasi kekuatan dan kelemahan, serta menilai kemampuan model dalam menangani data baru atau situasi nyata [15]. Evaluasi yang digunakan yaitu MRR, nDCG, Recall, dan MAP, untuk menilai kemampuan sistem menemukan ayat relevan sekaligus menempatkannya pada urutan teratas, penggunaan nDCG juga relevan untuk menilai kualitas urutan peringkat dengan mempertimbangkan bobot posisi.

Recall yaitu mengukur cakupan kemampuan sistem dalam menemukan semua item relevan yang tersedia untuk pengguna dalam daftar top-K. Pada sistem ini diartikan sebagai seberapa sering dokumen/ayat yang relevan berhasil muncul di 20 teratas. Semakin tinggi maka semakin “tidak ketinggalan” relevansi.

MRR (Mean Reciprocal Rank) yaitu mengukur seberapa awal sistem berhasil menampilkan item relevan dalam daftar rekomendasi [16]. Pada sistem ini fokus ke seberapa cepat hasil relevan pertama muncul. Semakin tinggi skor maka ayat relevan muncul di peringkat atas (rank 1–3), bukan di bawah.

nDCG atau Discounted Cumulative Gain (DCG) merupakan Teknik lain yang digunakan untuk mengukur kualitas dari ranking rekomendasi *Top-N*. DCG merupakan jumlah bobot dari tingkat relevansi film yang diberi ranking. Normalized Discounted Cumulative Gain (NDCG) adalah normalisasi dari DCG [17]. NDCG pada sistem ini diartikan sebagai ukuran relevansi hasil yang dihasilkan oleh model yang dibangun seberapa bagus urutan ranking hingga 20 hasil (relevan yang muncul lebih awal nilainya lebih besar).

Mean Average Precision (MAP) adalah ukuran kinerja yang umum digunakan dalam sistem temu kembali informasi yang dihitung untuk setiap kueri dan memberikan satu angka yang mewakili performa keseluruhan dari sistem temu kembali informasi. Presisi mengukur ketepatan sistem dalam menghasilkan citra yang relevan, dan didefinisikan sebagai rasio antara jumlah citra yang relevan yang ditemukan dengan jumlah total citra yang ditemukan [18]. Pada penelitian ini berarti rata-rata kualitas presisi berdasarkan posisi dokumen relevan.

Pengujian dilakukan menggunakan $N = 500$ kueri. Kueri dibangun oleh penulis yang dibangkitkan secara otomatis dengan memilih 500 ayat, mengambil potongan 7 kata pertama terjemahan Indonesia (*text_id*), kemudian membentuk kueri sintetis menggunakan templat pertanyaan. Ground truth relevansi (*qrels*) ditetapkan secara otomatis (*silver standard*), yaitu ayat asal pembentuk kueri dijadikan dokumen relevan (*doc_id* = *surah:ayat*, *label* = 1). Berdasarkan skema ini, evaluasi merefleksikan skenario *single-relevant/known-item retrieval*. Untuk mendukung replikasi, parameter pembangkitan *qrels* (*MAX_VERSES*, panjang potongan, templat, dan *random seed* bila digunakan) serta berkas *qrels* yang digunakan pada pengujian disertakan/dipublikasikan [19].

3. HASIL PENELITIAN DAN PEMBAHASAN

3.1. Hasil Muat Dataset

Sebelum melakukan perbandingan antara metode BM25 dan Dense Retrieval yang selanjutnya akan dibuat juga model penggabungan/RRF atau gabungan metode BM25 dan Dense Retrieval, Dataset yang telah diperoleh tersebut dilakukan pra-pemrosesan data untuk

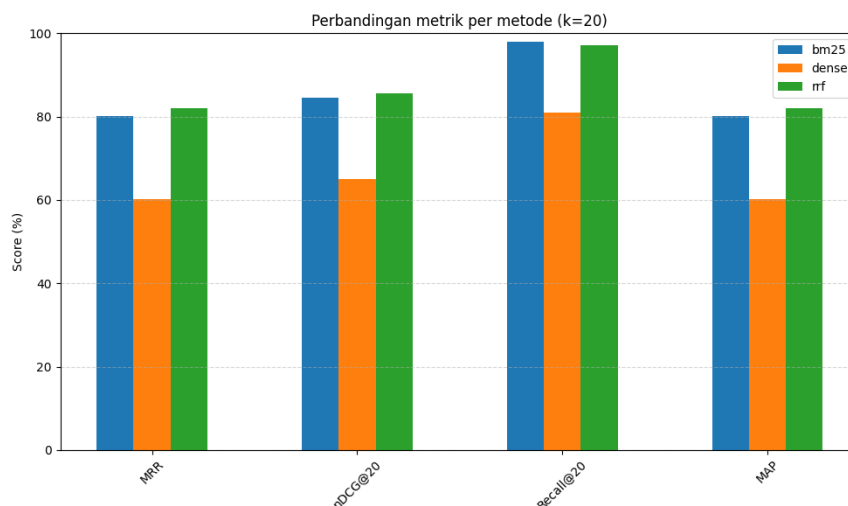
pembersihan dan normalisasi. Dataset ini berisi 6236 ayat, 114 surah, dan terjemahan. Dataset yang telah siap dapat ditunjukkan 5 teratas pada table 1 dibawah ini :

Tabel 1. Dataset

No Surah	Nama Surah	Ayat	Text Arab	Text Indonesia
1	Pembukaan	1	بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ	Dengan menyebut nama Allah Yang Maha Pemurah lagi Maha Penyayang.
1	Pembukaan	2	الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ	Segala puji bagi Allah, Tuhan semesta alam.
1	Pembukaan	3	الرَّحْمَنِ الرَّحِيمِ	maha pemurah lagi maha penyayang
1	Pembukaan	4	مَلِكِ يَوْمِ الدِّينِ	yang menguasai di hari pembalasan
1	Pembukaan	5	إِيَّاكَ نَعْبُدُ وَإِيَّاكَ نَسْتَعِينُ	hanya engkau yang kami sembah dan hanya kepada engkau kami meminta pertolongan

3.2. Hasil Kinerja Model

Berdasarkan implementasi sistem temu kembali informasi pada pencarian ayat Al-Qur'an yang dibangun, penelitian ini menguji tiga pendekatan IR, yaitu BM25 (leksikal), Dense Retrieval (semantik/dense), dan penggabungan/RRF, dengan masing-masing metode menghasilkan 20 ayat teratas (k=20) untuk setiap kueri.



Gambar 2. Diagram Performa BM25, Dense Retrieval, penggabungan/RRF

Berdasarkan Diagram batang pada gambar 2 diatas, performa ketiga metode kinerja metode terlihat jelas pada empat metrik evaluasi sistem temu kembali informasi, yaitu MRR, nDCG, Recall, MAP. Warna biru digunakan untuk menggambarkan metode BM25, warna jingga menggambarkan Dense Retrieval, dan warna hijau menggambarkan metode penggabungan/RRF. Dari grafik tersebut menunjukkan bahwa metode penggabungan/RRF atau penggabungan BM25 dan Dense Retrieval memberikan performa yang dapat mengungguli BM25 dan Dense Retrieval dalam tiga metrik evaluasi, Namun BM25 juga menunjukkan performa yang sedikit dibawah

penggabungan/RRF, ditunjukkan pada satu metrik evaluasi yaitu Recall BM25 mengungguli penggabungan/RRF.

Tabel 2. Evaluasi Kinerja Model

Metode	MRR	nDCG	Recall	MAP
BM25	80%	84%	98%	80%
Dense Retrieval	60%	64%	81%	60%
RRF (<i>Fusion</i>)	81%	85%	97%	81%

Analisis Perbandingan evaluasi pada tabel 3 diatas bahwa BM25 mencapai Recall sebesar 98% dengan MRR 80% dan nDCG 84%, sedangkan Dense Retrieval berada lebih rendah MRR 60%, nDCG 65%, Recall 81%. Metode penggabungan/RRF menunjukkan kualitas pemeringkatan tertinggi MRR 82% dan nDCG 86% dengan Recall tetap tinggi 97%. Pola ini mengindikasikan bahwa keuntungan utama RRF pada eksperimen ini adalah peningkatan kualitas urutan hasil teratas (MRR/nDCG), sementara penurunan cakupan pada top-20 relatif kecil dibanding BM25.

Keunggulan BM25 dalam Recall dapat dipahami melalui sifat dari pendekatan leksikal yang sangat efisien ketika kueri memiliki kecocokan kata dengan teks dokumen. Dalam konteks pencarian ayat, banyak kueri dari pengguna sering kali menggunakan kata kunci atau frasa yang secara eksplisit ada dalam terjemahan, sehingga sinyal dari kecocokan kata cukup kuat untuk menarik ayat-ayat relevan ke dalam daftar teratas dengan konsisten. Oleh karena itu, BM25 berfungsi sebagai metode yang handal untuk memastikan bahwa ayat target dapat ditemukan dalam daftar kandidat awal, terutama saat terdapat kesamaan yang jelas antara kata pada kueri dan dokumen.

Di sisi lain, Model Dense Retrieval yang lebih rendah menunjukkan bahwa representasi *embedding* dalam konfigurasi yang digunakan belum mencapai tingkat optimal untuk menempatkan ayat yang dicari di posisi yang lebih tinggi. Situasi ini dapat terjadi jika model *embedding* belum sepenuhnya kompatibel dengan domain dan gaya bahasa dari terjemahan Al-Qur'an, atau ketika kueri yang digunakan terlampaui pendek sehingga detail semantik yang ditangkap menjadi kurang tajam. Sebagai hasilnya, Dense Retrieval mungkin saja mengembalikan ayat yang masih "satu tema" tetapi tidak selalu sejajar dengan ayat yang dicari, sehingga nilai MRR dan nDCG cenderung menurun karena ayat yang dicari tidak berada di peringkat teratas.

Tabel 3. Perbandingan Hasil Top-5 BM25 dan Dense Retrieval pada kueri “riba”

Rank	BM25 (Surah: Ayat)	Topik	Dense Retrieval (Surah: Ayat)	Topik
1	2:275	riba	107:6	riya (tidak relevan)
2	2:279	riba	3:130	riba
3	30:39	riba	2:278	riba
4	2:276	riba	2:279	riba
5	3:130	riba	2:275	riba

Berdasarkan contoh perbandingan hasil pada Tabel 3 pada query bertema riba, BM25 menempatkan ayat 2:275 pada peringkat pertama karena adanya kecocokan kata kunci “riba” secara eksplisit. Sebaliknya, Dense Retrieval menempatkan ayat 107:6 (tema riya) pada peringkat pertama sehingga menunjukkan adanya pergeseran fokus relevansi pada hasil teratas. Meskipun Dense Retrieval kemudian memunculkan beberapa ayat riba pada peringkat berikutnya, kesalahan

pada posisi teratas berdampak pada metrik yang menekankan kualitas peringkat awal seperti MRR dan nDCG. Pada kueri yang sangat singkat seperti “riba”, Dense Retrieval berpotensi mengalami kekeliruan fokus topik karena embedding kueri memiliki konteks yang terbatas. Akibatnya, model dapat memprioritaskan ayat yang dekat secara makna moral umum (misalnya perilaku tercela seperti riya) meskipun tidak spesifik membahas riba. Sementara itu, BM25 lebih stabil untuk kueri satu kata yang spesifik karena mengandalkan kecocokan istilah secara eksplisit pada teks.

Penggabungan/RRF memperoleh MRR dan nDCG tertinggi karena menggabungkan dua sinyal peringkat yang berbeda, kecocokan kata dari BM25 dan kedekatan makna dari Dense Retrieval, sehingga pada sebagian kueri, ayat target yang tidak menonjol pada salah satu metode dapat terdorong naik ketika kedua daftar peringkat digabung. Namun, Recall RRF sedikit lebih rendah dari BM25 yaitu antara 97% dan 98%, yang dapat dijelaskan oleh efek batas top-k, penggabungan peringkat dapat menggeser ayat target turun tipis melewati posisi ke-20 pada sebagian kecil kueri. Oleh karena itu, interpretasi yang tepat adalah RRF meningkatkan kualitas ranking awal dengan *trade-off* kecil pada cakupan top-20. Secara praktis, temuan ini menyiratkan bahwa apabila tujuan utama sistem adalah menampilkan ayat paling relevan pada urutan teratas, maka penggabungan/RRF lebih menguntungkan karena meningkatkan kualitas pemeringkatan awal. Sebaliknya, jika prioritasnya adalah memaksimalkan peluang ayat target muncul dalam top-k, BM25 cukup kompetitif.

4. KESIMPULAN

Hasil eksperimen perbandingan tiga metode sistem temu kembali informasi pada sistem pencarian ayat Al-Qur'an menunjukkan hasil evaluasi pada k=20, pendekatan penggabungan/RRF memberikan performa terbaik secara keseluruhan dengan menghasilkan kualitas ranking tertinggi (ditunjukkan oleh peningkatan metrik seperti MRR/nDCG/MAP) dan tetap mempertahankan recall yang sangat tinggi, yang mengindikasikan bahwa penggabungan metode leksikal dan semantik bersifat saling melengkapi serta lebih efektif untuk meningkatkan relevansi dan kualitas pemeringkatan hasil pencarian ayat.

Sementara itu, penelitian ini juga menyimpulkan bahwa BM25 menunjukkan keunggulan utama pada aspek recall, yaitu paling konsisten dalam menemukan ayat relevan pada 20 hasil teratas, sehingga menegaskan efektivitas pendekatan leksikal untuk pencarian ayat berbasis kata kunci. Sedangkan Dense Retrieval pada konfigurasi yang digunakan masih belum mampu menyamai performa BM25, baik dari sisi cakupan relevansi Recall maupun kualitas urutan hasil awal yang tercermin pada metrik MRR dan nDCG, oleh karena itu peneliti menyarankan penelitian selanjutnya untuk berfokus pada peningkatan performa pendekatan semantik mengingat hasil saat ini menunjukkan Dense Retrieval masih tertinggal dibanding BM25. Salah satu arah utama adalah melakukan optimasi atau fine-tuning model Dense Retrieval menggunakan data domain-spesifik (misalnya pasangan kueri dan ayat relevan atau qrels yang lebih kaya) agar representasi embedding lebih sesuai dengan karakteristik bahasa dan konteks pencarian ayat. Dengan rangkaian pengembangan tersebut, diharapkan sistem mampu mempertahankan keunggulan metode penggabungan/RRF dan BM25, sekaligus meningkatkan kualitas pemeringkatan melalui metode Dense Retrieval yang lebih optimal.

DAFTAR PUSTAKA

- [1] Amrullah, “Transformasi Digital Dalam Pendidikan Agama Islam : Kajian Implementasi Metaverse Sebagai Media Pembelajaran Interaktif,” *J. Pendidik. Dan Kebud.*, Vol. 6, No. November, Pp. 57–66, 2024, Doi: [Jurnalstithualhikmah.Ac.Id](https://doi.org/10.24252/shift.v6i1.272).
- [2] R. M. Poerbaningtyas, Evy, “Penerapan Metode Cosine Similarity Pada Sistem Informasi Retrival,” *J-Intech (Journal Inf. Technol.*, No. 204, 2023, Doi:

<https://creativecommons.org/licenses/by-sa/4.0/>.

- [3] M. A. M. Y. Nuhu Yusuf, N. M. N. Norfaradilla Wahid, And N. A. Noor Azah Samsudin, "Query Expansion Method For Quran Search Using Semantic Search And Lucene Ranking," *J. Eng. Sci. Technol.*, Vol. 15, No. 1, Pp. 675–692, 2020, Doi: <https://Jestec.Taylors.Edu.My/>.
- [4] A. W. W. Puteri Aulia Indrasti, Indriati, "Temu Kembali Informasi Berita Berbahasa Indonesia Menggunakan," *J. Pengemb. Teknol. Inf. Dan Ilmu Komput.*, Vol. 4, No. 9, Pp. 3005–3013, 2020, Doi: <http://J-Ptiik.Ub.Ac.Id>.
- [5] T. R. K. Teguh Ikhlas Ramadhan, Agus Supriatman, "Passage Retrieval Untuk Question Answering Bahasa Indonesia," *J. Algoritma*, 2024, Doi: 10.33364/Algoritma/V.21-2.2100.
- [6] E. P. M. Muhammad Abdul Hafizh Fathuddin And A. L. Nurlaili, "Penerapan Sentence-Bert Dan Cosine Similarity Untuk Pencarian Semantik Dokumen Skripsi Dalam Format Pdf," *J. Multidiscip. Res. Dev.*, Vol. 8, No. 1, Pp. 322–337, 2025, Doi: <https://Jurnal.Ranahresearch.Com/Index.Php/R2j>.
- [7] A. Rizkiashuripratama, Yunussafi'i, Maulidhanadynugraha And N. Satussobihah, "Perbandingan User-Based Dan Item-Based Pada Sistem Rekomendasi Film Kombinasi Teknik Reduksi Dimensi Dan Clustering," *J. Tekno Insentif*, Vol. 19, No. 1, 2025, Doi: <https://doi.org/10.36787/Jti.V19i1.1662>.
- [8] F. I. R. Gempar Perkasa Tahir, Emil Agusalam Habi Talib, "Implementasi Metode Hybrid Fuzzy Jaro Winkler Dan Cosine Similarity Pada Sistem Pencarian Ayat Al-Quran Berbasis Transliterasi Latin," *J. Inform. Prog.*, Vol. 17, No. 2, Pp. 105–115, 2025, Doi: <https://Jurnal.Stmikprofesional.Ac.Id/Index.Php/Progress/Article/Download/482/218>.
- [9] A. S. B. Afrit Solihin, Fandy Setyo Utomo, "Evaluasi Pengaruh Varian Daftar Stopword Terhadap Kinerja Klasifikasi Teks Al- Qur ' An Dengan Support Vector Machine Dan Backpropagation Neural Network Program Studi Teknologi Informasi , Fakultas Ilmu Komputer , Universitas Amikom Purwokerto , Indonesia," *J. Pendidik. Dan Teknol. Indones.*, Vol. 5, No. 7, Pp. 1867–1880, 2025, Doi: <https://doi.org/10.52436/1.Jpti.875>.
- [10] I. M. A. W. P. I Gede Nyoman Agung Jayarana, I Gede Wira Darma, I Wayan Ady Juliantera, "Study Literatur Information Retrieval Model : Teknik Dan Aplikasi," *J. Sutasoma*, Vol. 03, No. 01, Pp. 61–69, 2025, Doi: <https://doi.org/10.58878/Sutasoma.V3i2.392>.
- [11] I. Taemoon Jung, "An Intelligent Docent System With A Small Large Language Model (Sllm) Based On Retrieval-Augmented Generation (Rag)," *Appl. Sci.*, Pp. 1–30, 2025, Doi: <https://www.mdpi.com/2076-3417/15/17/9398>.
- [12] S. A. Anggun Tri Utami Br. Lubis, Nazruddin Safaat Harahap And I. A. Muhammad Irsyad, "Question Answering System On Telegram Chatbot Using Large Language Models (Llm) And Langchain (Case Study : Health Law) Question Answering System Pada Chatbot Telegram Menggunakan Large Language Models (Llm) Dan Langchain (Studi Kasus Uu Kesehatan," *Indones. J. Mach. Learn. Comput. Sci.*, Vol. 4, No. July, Pp. 955–964, 2024, Doi: <https://Journal.Irpi.Or.Id/Index.Php/Malcom>.
- [13] W. Y. Vladimir Karpukhin, Barlas Oguz, Sewon Min, Patrick Lewis Ledell Wu, Sergey Edunov, Danqi Chen, "Dense Passage Retrieval For Open-Domain Question Answering," *Assoc. Comput. Linguist.*, Vol. Proceeding, Pp. 6769–6781, 2020, Doi:

10.18653/V1/2020.Emnlp-Main.550.

- [14] R. V. H. G. Fadhel Akhmad Hizham, “Pengembangan Metode Information Retrieval Dan Haversine Formula Untuk Rekomendasi Penentuan Klinik Di Kabupaten Jember,” *J. Informatics Dev.*, Vol. 1, Pp. 13–25, 2022, Doi: <https://Ejournal.Itbwigalumajang.Ac.Id/Index.Php/Jid>.
- [15] A. W. Anggitta Ratu, Leni Kusneti, Thomas Filikano, Andronikus G., “Implementasi Dan Evaluasi Sistem Pencarian Informasi Ulasan Restoran India Menggunakan Algoritma Vsm,” *J. Informatics Business*, Vol. 02, No. 04, Pp. 529–538, 2025, Doi: <https://Jurnal.Ittc.Web.Id/Index.Php/Jibs/Index>.
- [16] D. P. K. Assyfa Rasida Hanum, Eko Sakti Pramukantoro, “Studi Perbandingan Kinerja Tf-Idf Dan Indobert Untuk Rekomendasi Resep Berdasarkan Ketersediaan Bahan Makanan Berbasis Website,” *J. Pengemb. Teknol. Inf. Dan Ilmu Komput.*, Vol. 9, No. 10, Pp. 1–8, 2025, Doi: <http://J-Ptiik.Ub.Ac.Id>.
- [17] N. I. Fadetul Fitriyeh, Nuskhatul Haqqi, Layla Mufah Choiriyah, “Dampak Pembobotan Pada Metode Hybrid User-Based Dan Item-Based Untuk Sistem Rekomendasi Film,” *J. Comput. Sci. Inf. Technol.*, Vol. 5, No. 3, Pp. 516–525, 2024, Doi: <http://Ejournal.Umri.Ac.Id/Index.Php/Coscitech/Index>.
- [18] N. N. Mutmainnah Muchtar, Noorhasanah Zainuddin, Adha Mashur Sajiah And Y. P. Pasrun, “Perbandingan Jarak Euclidean, Cityblock, Minkowski, Canberra, Dan Chebyshev Dalam Sistem Temu Kembali Citra Batik,” *Jitet (Jurnal Inform. Dan Tek. Elektro Ter.)*, Vol. 12, No. 3, 2024, Doi: <http://Dx.Doi.Org/10.23960/Jitet.V12i3s1.5324>.
- [19] I. S. Nick Craswell, Bhaskar Mitra, Emine Yilmaz, Daniel Campos, Ellen M. Voorhees, *Trec Deep Learning Track : Reusable Test Collections In The Large Data Regime*, Vol. 1, No. 1. Association For Computing Machinery, 2021. Doi: <https://Dl.Acm.Org/Doi/Abs/10.1145/3404835.3463249>.