

BAB I

PENDAHULUAN

1.1. Latar Belakang

Teknologi *search engine* dalam beberapa tahun terakhir telah berkembang sangat pesat. Namun, kemajuan tersebut tidak selalu merata. Dari banyaknya penelitian dan aplikasi yang berkaitan dengan *search engine*, tidak sedikit di antaranya masih mengandalkan algoritma tradisional tanpa mengintegrasikan algoritma terbaru seperti *deep learning* yang mampu memahami kalimat secara semantik [1][2] sehingga *search engine* masih berbasis pada aturan, pencocokan keyword [3][4] dan leksikal seperti TF-IDF [5][6]. Ada beberapa *search engine* yang masih menggunakan algoritma tradisional seperti Alfanous [7], Search Truth [8], OpenQuran, QuranSearch, Search the Quran, Lafzi [9][10][11], dan Al-Hadees yang berfokus pada Quran dan bidang Islam. Pendekatan atau algoritma yang mereka gunakan seringkali memiliki batasan dalam memahami input pengguna, yang mengakibatkan hasil pencarian terbatas pada aturan tertentu sehingga pengguna tidak dapat leluasa dalam mencari dan menemukan informasi secara akurat [3].

Untuk mengatasi kelemahan tersebut, penelitian di bidang *Information Retrieval* (IR) kemudian beralih ke *deep learning model* yang berbasis *Transformer* yang terbukti mampu meningkatkan performa dari sistem IR secara signifikan [12][13]. Tidak seperti model-model pendahulunya, *deep learning model* dilatih menggunakan jutaan kalimat dan mampu memahami konteks, sinonim, serta makna dibalik bahasa secara semantik. Kemampuan inilah yang membuatnya menjadi *state-of-the-art* dari keterbatasan algoritma-algoritma tradisional. Namun, untuk membuat *deep learning model* bisa beradaptasi dan digunakan pada domain-domain tertentu, model perlu melalui proses yang dinamakan *fine-tuning*. Selain membuat *deep learning model* mampu beradaptasi pada domain tertentu, keunggulan lain dari pendekatan *fine-tuning* adalah kemampuannya untuk memberikan akurasi tinggi dengan membuat model menjadi “*expert*” di bidang tertentu seperti yang telah dibuktikan pada [14][15].

Namun, di balik keunggulannya sebagai pendekatan *state-of-the-art*, ketergantungan pada *fine-tuning* seringkali menyebabkan berbagai masalah praktis yang sering dihadapi oleh peneliti dan pengembang. Proses ini umumnya memerlukan sumber daya komputasi yang mahal dan seringkali melibatkan *trial-and-error* [16], disertai *hyperparameter tuning* yang kompleks untuk menghindari hasil yang tidak memuaskan [17]. Masalah lain yang dihadapi adalah proses pengembangan *dataset* baru berkualitas tinggi yang memakan waktu, namun tetap membawa risiko *catastrophic forgetting*, kondisi di mana kinerja model menurun karena kehilangan pengetahuan umum setelah dilatih pada *dataset* yang terlalu spesifik [18][19][20][21].

Tantangan-tantangan ini terbukti pada studi kasus pada penelitian sistem pencarian Sequran lainnya [22] di mana sistem menunjukkan performa yang kuat saat dievaluasi menggunakan *dataset* publik. Sistem tersebut mampu mencapai nilai *precision/recall @10* sebesar 0.4469/0.4477 pada 'MIRACL' [23] dan 0.089/0.89 pada 'TyDi' [24][25]. Namun, ketika sistem diuji menggunakan *dataset* pertanyaan-jawaban spesifik bertemakan Islam [26], hasil evaluasi menunjukkan bahwa sistem hanya mampu mencapai nilai *precision@10* sebesar 0.167 dan nilai *recall@10* sebesar 0.423. Meskipun nilai *recall* mencapai 0.423, nilai *precision* yang rendah menunjukkan bahwa sistem Sequran hanya dapat menampilkan sebagian dokumen yang benar-benar relevan dari *dataset* evaluasi [26], yang akibatnya, hasil pencarian mengandung banyak *false positive*.

Salah satu hipotesis dari kegagalan ini adalah masalah yang terjadi selama *data preparation* untuk proses *fine-tuning* di mana proses tersebut sangat bergantung pada kualitas dan kuantitas dari data pelatihan yang digunakan. Dalam kasus penelitian Sequran lainnya [22], banyak dokumen sumber melebihi *context window limit* model yang digunakan, sehingga memerlukan proses *semantic chunking*. Proses *semantic chunking* ini cukup berisiko jika tidak dilakukan dengan hati-hati karena dapat menghilangkan informasi penting atau konteks asli dari dokumen tersebut [27], sehingga dapat menyebabkan tugas-tugas menantang lainnya seperti *complex hyperparameter tuning* dan *dynamic hard negative mining*.

Untuk mengatasi tantangan tersebut, penelitian ini mengusulkan arsitektur *refinement* sebagai alternatif dari *fine-tuning*. Tujuan utama dari pendekatan *refinement* ini adalah untuk meningkatkan relevansi hasil pencarian, yang diukur melalui peningkatan *precision* dan *recall*, sehingga mengurangi risiko penyampaian informasi yang tidak akurat atau tidak relevan, terutama pada domain spesifik seperti Quran dan teks-teks Islam lainnya. Arsitektur yang diusulkan menggabungkan *composite scoring* dengan teknik *reranking*, yang dievaluasi secara komprehensif untuk mengukur efektivitas kombinasi tersebut dalam meningkatkan kinerja sistem pencarian.

1.2. Rumusan Masalah

Berdasarkan latar belakang tersebut, maka rumusan masalah pada penelitian ini dapat dirumuskan sebagai berikut:

1. Apa kombinasi efektif dari teknik *composite scoring*, dan sejauh mana kombinasi tersebut meningkatkan relevansi pencarian berdasarkan metrik *precision* dan *recall*?
2. Setelah menerapkan kombinasi efektif dari teknik *composite scoring*, apa dampak dari penambahan teknik *reranking* terhadap relevansi pencarian (*precision* dan *recall*) dan total waktu yang dibutuhkan untuk mengevaluasi *dataset*?

1.3. Tujuan Penelitian

Berdasarkan rumusan masalah tersebut, maka tujuan dari penelitian ini adalah sebagai berikut:

1. Untuk menentukan kombinasi efektif dari teknik *composite scoring* dengan mengevaluasi pengaruh dari komponen-komponennya terhadap relevansi pencarian menggunakan metrik *precision* dan *recall*.
2. Untuk mengevaluasi pengaruh penambahan teknik *reranking* ke teknik *composite scoring* yang dioptimalkan dengan menganalisis *trade-off* antara peningkatan relevansi pencarian (*precision* dan *recall*) dan peningkatan total waktu evaluasi *dataset*.

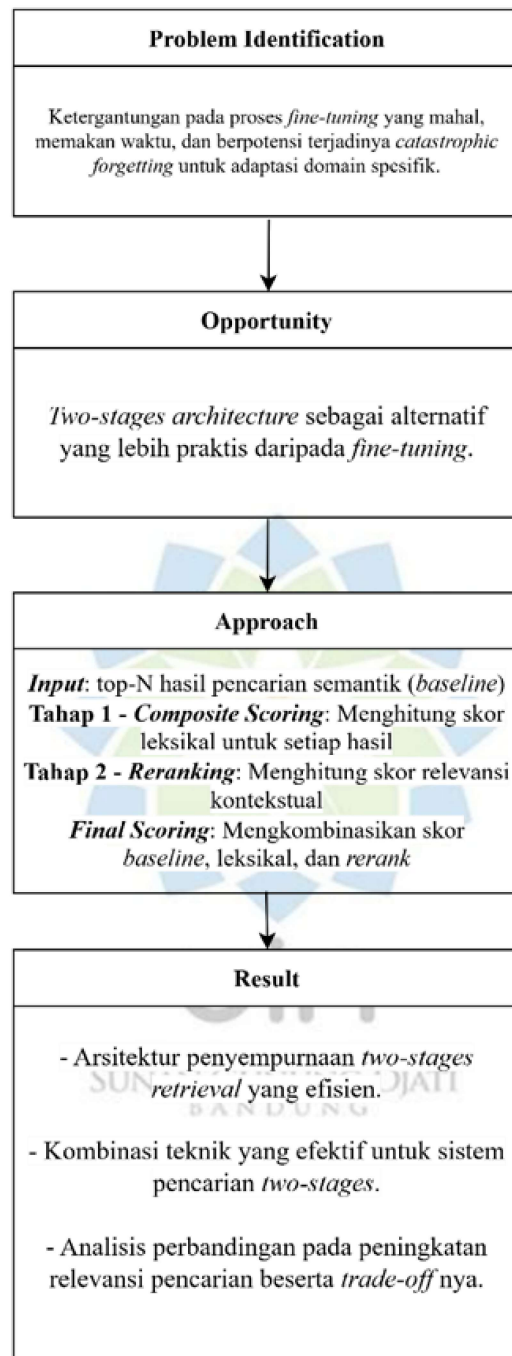
1.4. Batasan Masalah

Batasan masalah dimaksudkan untuk memberikan ruang lingkup yang jelas sehingga penelitian menjadi lebih fokus dan terukur. Adapun batasan masalah pada penelitian ini adalah sebagai berikut:

1. Penelitian ini berfokus pada perancangan dan evaluasi algoritma, bukan pada pengembangan aplikasi secara utuh.
2. *Dataset* evaluasi yang digunakan hanya berjumlah 27 data pertanyaan-jawaban bertemakan Islam dan berbahasa Indonesia (Melibatkan kurator dan ulama pada proses pembuatannya) dan tidak menggunakan *benchmark* IR pada umumnya.
3. Kategori *intent* (maksud dari sebuah teks atau kalimat) yang digunakan dalam sistem dibatasi hanya berjumlah 10 kategori yang telah disesuaikan untuk domain Quran.
4. Kinerja sistem diukur menggunakan metrik *Precision* dan *Recall*, dengan fokus pada *set-based relevance*.
5. Seluruh proses implementasi dan pengujian algoritma dilakukan pada spesifikasi hardware yang telah ditentukan, yaitu komputer dengan sistem operasi Windows 11, prosesor Intel Core i7-11800H, dan GPU NVIDIA RTX 3060. Performa latensi yang dilaporkan spesifik untuk konfigurasi *hardware* ini.

1.5. Kerangka Pemikiran

Berikut adalah kerangka pemikiran yang menjadi dasar pada penelitian ini. Kerangka ini memvisualisasikan alur logis yang dilakukan pada penelitian ini, mulai dari identifikasi masalah, peluang, hingga hasil yang menjadi target dari penelitian ini.



Gambar 1.1 Kerangka Pemikiran

1.6. Sistematika Penulisan

Untuk memberikan gambaran yang jelas terkait alur penelitian, keseluruhan penulisan disusun menjadi lima bab yang saling terkait satu sama lain, dengan rincian sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisikan latar belakang yang memotivasi penelitian, rumusan masalah, tujuan, batasan, kerangka pemikiran, hingga sistematika penulisan dari penelitian ini.

BAB II KAJIAN LITERATUR

Bab ini berisikan kerangka teoritis dan tinjauan penelitian terkait. Pembahasannya mencakup konsep-konsep dasar dalam *Information Retrieval* (IR), arsitektur multi-tahap, serta analisis terhadap penelitian-penelitian terdahulu untuk menegaskan *research gap* yang ingin diisi.

BAB III METODOLOGI PENELITIAN

Bab ini menguraikan *blueprint* dan langkah-langkah yang perlu dilakukan pada penelitian secara rinci guna memudahkan peneliti lain dalam mereproduksi dan menguji hasil dari penelitian ini.

BAB IV HASIL DAN PEMBAHASAN

Bab ini memaparkan hasil, temuan, dan pengujian dari implementasi metode yang telah dilakukan. Hasil pengujian akan disajikan secara objektif dalam bentuk tabel yang kemudian akan dianalisis serta diinterpretasikan secara rinci. Pembahasan berfokus pada evaluasi kinerja, perbandingan metode yang diuji, dan signifikansi dari hasil yang diperoleh.

BAB V PENUTUP

Bab ini berisi rangkuman dari temuan-temuan utama, jawaban dari setiap rumusan masalah yang diajukan pada BAB I, serta mengidentifikasi keterbatasan penelitian. Saran untuk penelitian lebih lanjut didasarkan pada kesimpulan dan keterbatasan yang telah dijelaskan sebelumnya.

DAFTAR PUSTAKA