

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kebutuhan akan akses cepat dan akurat terhadap informasi hukum semakin meningkat. Dokumen putusan pengadilan, yang berisi keputusan penting terkait kasus hukum, harus dikelola secara sistematis agar dapat diakses dengan mudah oleh praktisi hukum untuk analisis kasus, peneliti untuk studi akademis, serta masyarakat umum untuk memahami perkembangan hukum yang relevan [1]. Dengan tersedianya dokumen dalam bentuk digital, seharusnya proses pencarian informasi terkait kasus atau putusan dapat dilakukan dengan cepat, akurat, dan relevan. Teknologi temu kembali informasi menawarkan solusi yang memungkinkan pengguna untuk menemukan dokumen yang relevan dengan kata kunci tertentu melalui algoritma pencarian berbasis mesin.

Proses temu kembali informasi pada dokumen putusan pidana umum masih menghadapi tantangan signifikan [2]. Dokumen hukum, khususnya putusan pidana, memiliki struktur yang kompleks dan sering kali panjang, dengan rata-rata mencapai lebih dari 10 halaman. Setiap dokumen memuat berbagai elemen penting, seperti identitas pihak yang terlibat, kronologi peristiwa, serta pertimbangan dan hasil keputusan, yang tersusun dalam narasi panjang dengan struktur yang kompleks [3]. Kompleksitas ini menyulitkan pengguna dalam menemukan informasi spesifik dengan cepat, terlebih jika hanya membutuhkan poin-poin tertentu. Sistem pencarian konvensional sering kali tidak mampu mengatasi tingginya volume informasi dalam dokumen hukum, sehingga proses pencarian menjadi lambat dan tidak efisien. Oleh karena itu, diperlukan pendekatan yang lebih efektif untuk mengekstraksi informasi kunci dari dokumen-dokumen ini, sehingga kecepatan dan efisiensi pencarian dapat ditingkatkan.

Temu kembali informasi adalah proses menemukan dokumen atau data yang relevan dari kumpulan besar informasi berdasarkan permintaan atau kebutuhan spesifik pengguna [4]. Dalam konteks hukum, temu kembali informasi berperan penting untuk memudahkan pencarian dokumen putusan yang relevan dengan cepat dan akurat, terutama ketika informasi yang dibutuhkan tersebar di antara ribuan

dokumen hukum. Teknologi ini diperlukan agar pengguna, seperti praktisi hukum atau masyarakat, dapat menemukan informasi yang tepat dalam waktu singkat [5]. Tanpa penerapan teknologi temu kembali informasi yang optimal, hasil pencarian cenderung kurang relevan, sehingga menyulitkan pengguna dalam memperoleh dokumen yang mereka butuhkan secara efisien.

Penelitian Thakur et al. [6], dalam BEIR Benchmark membahas perbandingan antara metode TF-IDF dan BM25 dalam sistem temu kembali informasi. Kedua metode ini termasuk dalam pendekatan leksikal yang menggunakan frekuensi kemunculan kata dalam dokumen untuk menentukan relevansi terhadap sebuah query. TF-IDF menghitung bobot berdasarkan seberapa sering suatu kata muncul dalam dokumen dan seberapa jarang kata tersebut muncul di seluruh koleksi dokumen. Sementara itu, BM25 merupakan pengembangan dari pendekatan ini dengan menambahkan penyesuaian terhadap panjang dokumen dan memberikan batas terhadap pengaruh frekuensi kata yang terlalu sering muncul. Dalam penelitian ini, keduanya dibandingkan dalam berbagai jenis tugas pencarian pada berbagai domain data, sehingga memberikan gambaran tentang karakteristik dan cara kerja masing-masing metode dalam konteks yang berbeda-beda. Perbandingan ini menjadi penting untuk memahami kelebihan dan kekurangan relatif dari masing-masing metode dalam mendukung sistem temu kembali informasi berbasis teks.

Penelitian terkait pengembangan fungsi pemeringkatan dalam sistem temu kembali informasi berbasis teks juga dilakukan oleh de Almeida et al. [7] melalui pendekatan *Combined Component Approach* (CCA) berbasis *Genetic Programming*. Dalam penelitian tersebut, mereka menggabungkan komponen-komponen utama dari berbagai metode pemeringkatan populer, termasuk *Term Frequency-Inverse Document Frequency* (TF-IDF) dan Okapi BM25, untuk menghasilkan fungsi pemeringkatan yang lebih adaptif terhadap karakteristik koleksi dokumen tertentu. Komponen seperti logaritma *term frequency* dari TF-IDF dan faktor normalisasi panjang dokumen dari BM25 digunakan bersama dalam satu formula, berdasarkan asumsi bahwa kedua metode memiliki elemen dasar yang serupa, yaitu *term frequency* (TF), *inverse document frequency* (IDF), dan normalisasi. Oleh karena itu, pendekatan hybrid ini menjadi relevan dan didukung

secara teoritis maupun empiris dalam pengembangan sistem temu kembali informasi yang lebih efektif.

Penelitian sebelumnya, seperti yang dilaporkan dalam BEIR Benchmark [6], menunjukkan bahwa kinerja TF-IDF dan BM25 sangat dipengaruhi oleh karakteristik kueri dan koleksi dokumen yang digunakan. Selain itu, pendekatan penggabungan metode, seperti yang dilakukan dalam penelitian de Almeida et al. [7] melalui Combined Component Approach, menunjukkan bahwa elemen-elemen dari TF-IDF dan BM25 dapat saling melengkapi dalam meningkatkan kualitas pemeringkatan dokumen. Hal ini membuka peluang untuk mengevaluasi efektivitas metode hybrid TF-IDF+BM25 pada domain dokumen hukum berbahasa Indonesia.

Dalam konteks pencarian dokumen putusan pidana, variasi panjang dan bentuk *query* merupakan kondisi yang tidak dapat dihindari. Pengguna dapat memasukkan *query* berupa istilah tunggal yang merepresentasikan delik inti, frasa pendek yang menggambarkan jenis tindak pidana tertentu, maupun *query* panjang yang menyerupai rumusan dakwaan atau potongan narasi perkara. Perbedaan panjang *query* ini secara langsung memengaruhi distribusi istilah dan mekanisme pembobotan relevansi pada metode TF-IDF dan BM25 [8]. Oleh karena itu, diperlukan skenario pengujian yang secara eksplisit mempertimbangkan variasi panjang tokenisasi *query* untuk mengevaluasi sensitivitas dan stabilitas masing-masing metode dalam konteks dokumen hukum.

Pulau Madura yang terdiri dari empat kabupaten, yaitu Bangkalan, Sampang, Pamekasan, dan Sumenep, merupakan wilayah yang termasuk dalam sepuluh besar daerah dengan presentase tingkat kemiskinan tertinggi di Provinsi Jawa Timur [9]. Tingginya angka kemiskinan di wilayah ini menjadi permasalahan multidimensional yang berdampak tidak hanya pada aspek sosial dan ekonomi, tetapi juga pada stabilitas keamanan dan tingkat kriminalitas masyarakat setempat. Secara teoritis, kemiskinan seringkali dikaitkan dengan meningkatnya potensi tindak pidana di suatu daerah. Keterbatasan ekonomi dapat mendorong individu atau kelompok masyarakat untuk melakukan tindakan yang melanggar hukum sebagai bentuk upaya memenuhi kebutuhan hidup. Oleh karena itu, daerah dengan

tingkat kemiskinan yang tinggi cenderung memiliki angka kriminalitas yang relatif tinggi pula [10].

Data yang dihimpun dari beberapa pengadilan negeri di wilayah Madura menunjukkan jumlah tindak pidana yang signifikan. Di Pengadilan Negeri Bangkalan tercatat sebanyak 2.971 kasus pidana umum, Pengadilan Negeri Sampang sebanyak 2.365 kasus pidana umum, Pengadilan Negeri Pamekasan sebanyak 1.974 kasus pidana umum, dan Pengadilan Negeri Sumenep sebanyak 3.197 kasus pidana umum. Data tersebut tercatat per tanggal 11 Juni 2025. Jumlah perkara pidana umum yang ditangani oleh masing-masing pengadilan ini mengindikasikan adanya kecenderungan maraknya tindak pidana di wilayah tersebut. Berdasarkan fenomena tersebut, penelitian ini memfokuskan perhatian pada pengkajian lebih lanjut terhadap dokumen-dokumen putusan pidana yang diputus oleh pengadilan di wilayah Madura. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem pencarian informasi hukum dan dapat menjadi acuan awal untuk pengembangan sistem serupa di daerah lain.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk membandingkan dan menggabungkan metode temu kembali informasi menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF), *Best Match 25* (BM25), dan TF-IDF+BM25 pada dokumen putusan hukum berbahasa Indonesia. Dataset yang digunakan berasal dari situs resmi Mahkamah Agung (MA) dan difokuskan pada putusan pidana umum di wilayah Madura. Dengan pendekatan ini, diharapkan dapat memberikan kontribusi dalam pemilihan metode pencarian yang paling sesuai untuk konteks dokumen hukum, serta membantu praktisi hukum, peneliti, dan masyarakat umum dalam mengakses informasi hukum secara lebih tepat, efisien, dan mudah, khususnya pada dokumen putusan pidana umum.

1.2 Rumusan Masalah

1.2.1 Permasalahan

Dalam temu kembali informasi pada dokumen putusan pidana umum, terdapat ketidaksesuaian antara karakteristik dokumen yang panjang dan memiliki struktur naratif yang kompleks dengan kinerja metode pemeringkatan berbasis leksikal dalam merespons variasi panjang *query* pencarian. Metode seperti TF-IDF

dan BM25 sangat dipengaruhi oleh distribusi frekuensi istilah serta mekanisme pembobotan relevansi yang berbeda, sehingga perbedaan panjang dan kelengkapan informasi dalam *query* berpotensi menghasilkan variasi kinerja pemeringkatan dokumen. Tanpa adanya pengujian yang secara sistematis mempertimbangkan variasi karakteristik *query* dalam konteks dokumen putusan pidana umum, sulit untuk menentukan metode temu kembali informasi yang paling stabil dan efektif dalam menghasilkan dokumen yang relevan.

1.2.2 Solusi Permasalahan

Sebagai solusi, penelitian ini membandingkan metode pemeringkatan TF-IDF, BM25, serta metode hybrid TF-IDF+BM25 untuk mengevaluasi respons masing-masing pendekatan terhadap variasi panjang tokenisasi *query* pada dokumen putusan pidana umum. Metode hybrid digunakan untuk menguji apakah integrasi bobot term berbasis kelangkaan dari TF-IDF dan mekanisme normalisasi panjang dokumen dari BM25 dapat menghasilkan pemeringkatan yang lebih stabil dibandingkan metode tunggal pada kondisi *query* yang berbeda.

1.2.3 Pertanyaan Penelitian

Bagaimana perbandingan kinerja metode *Hybrid* TF-IDF dan BM25 terhadap metode TF-IDF dan BM25 secara terpisah dalam temu kembali dokumen putusan pidana umum Pengadilan Negeri di wilayah Madura, berdasarkan variasi panjang *query*?

1.3 Tujuan Dan Manfaat

1.3.1 Tujuan

Menganalisis dan membandingkan kinerja metode *Hybrid* TF-IDF dan BM25 dengan metode TF-IDF dan BM25 secara terpisah dalam temu kembali dokumen putusan pidana umum Pengadilan Negeri di wilayah Madura, berdasarkan variasi panjang *query*.

1.3.2 Manfaat

Penelitian ini bertujuan untuk mengevaluasi kinerja metode temu kembali informasi pada dokumen putusan pidana umum Pengadilan Negeri di wilayah Madura dengan membandingkan metode TF-IDF, BM25, dan metode hybrid TF-IDF+BM25. Evaluasi dilakukan untuk menilai efektivitas dan stabilitas masing-

masing metode dalam menghasilkan dokumen yang relevan pada berbagai variasi karakteristik *query* pencarian, menggunakan metrik *precision*, *recall*, dan *f1-score*. Hasil penelitian diharapkan dapat memberikan dasar dalam pemilihan metode temu kembali informasi yang paling sesuai untuk pencarian dokumen putusan pidana umum serta mendukung pengembangan sistem pencarian informasi hukum yang lebih efektif.

1.4 Batasan Masalah

1. Data berasal dari hasil *scraping* website resmi Mahkamah Agung (MA), <https://putusan3.mahkamahagung.go.id/>.
2. Data yang digunakan sebagai pencarian adalah dokumen putusan pidana umum hasil ekstraksi entitas.
3. Data yang digunakan adalah 749 dokumen putusan hukum dengan klasifikasi pidana umum di pengadilan negeri wilayah Madura.

1.5 Sistematika Proposal

BAB I : PENDAHULUAN

Pada bab ini, menjelaskan tentang pemberian solusi dari permasalahan yang berada di sistem temu kembali informasi dokumen pidana umum Pengadilan Negeri di Wilayah Madura. Solusi yang diberikan adalah membandingkan dan menggabungkan metode pencarian TF-IDF, BM25, dan TF-IDF+BM25 yang memanfaatkan ekstraksi entitas dari dokumen putusan pidana umum. Kemudian menjelaskan metode TF-IDF dan BM25 berdasarkan penelitian sebelumnya. Serta, berisi Perumusan Masalah, Tujuan, Manfaat, dan Jadwal Penelitian.

BAB II : KAJIAN PUSTAKA

Pada bab ini, berisi tinjauan literatur tentang Temu Kembali Informasi, Metode *Term Frequency-Inverse Document Frequency* (TF-IDF), *Best Match 25* (BM25), dan penelitian terkait tentang dokumen putusan hukum.

BAB III : METODOLOGI PENELITIAN

Pada bab ini menjelaskan tahapan-tahapan penelitian secara rinci, mulai dari pengumpulan dan pengolahan data, hingga implementasi dan evaluasi metode TF-

IDF, BM25, dan TF-IDF+BM25. Evaluasi dilakukan berdasarkan metrik seperti *precision*, *recall*, dan *f1-score*.

BAB IV : PEMBAHASAN

Bab ini menyajikan hasil implementasi dan pengujian sistem temu kembali informasi yang telah dibangun. Pengujian dilakukan melalui tiga skenario pencarian berdasarkan variasi panjang query, yaitu query satu kata, dua kata, dan query panjang. Hasil pengujian masing-masing metode TF-IDF, BM25, dan TF-IDF+BM25 dianalisis berdasarkan nilai *precision*, *recall*, dan *f1-score*. Selanjutnya, dilakukan pembahasan terhadap hasil pengujian tersebut untuk menjelaskan perbedaan kinerja setiap metode, serta faktor-faktor yang memengaruhi hasil pencarian pada setiap skenario pengujian.

BAB V : PENUTUP

Pada bab ini berisi kesimpulan yang diperoleh berdasarkan hasil pengujian dan pembahasan pada bab sebelumnya. Kesimpulan disusun dengan menekankan perbandingan kinerja metode TF-IDF, BM25, dan TF-IDF+BM25 pada berbagai skenario pencarian, serta kelebihan dan keterbatasan masing-masing metode. Selain itu, bab ini juga memuat saran yang ditujukan untuk pengembangan penelitian selanjutnya, khususnya terkait peningkatan kualitas pencarian dokumen putusan pidana umum.