

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi telah mendorong lahirnya berbagai penelitian di bidang *information retrieval*, salah satunya mengenai efektivitas algoritma dalam mengolah dan menganalisis teks. Dalam konteks tersebut, algoritma *Cosine Similarity* dan *Boyer–Moore* memiliki peran penting karena masing-masing menawarkan kemampuan yang mendukung proses pencarian dokumen digital. [1], [2]

Algoritma *Cosine Similarity* merupakan algoritma yang menggunakan pendekatan berbasis vektor untuk mengukur tingkat kemiripan antar dokumen. Prinsip kerjanya adalah menghitung sudut kosinus antara dua vektor yang mewakili dokumen, sehingga nilai kesamaan dapat diperoleh tanpa dipengaruhi oleh panjang dokumen. [3], [4]

Berbagai penelitian menunjukkan bahwa algoritma *Cosine Similarity* mampu menghasilkan tingkat akurasi yang tinggi dalam deteksi plagiarisme, klasifikasi dokumen, dan pengukuran semantik antar teks. [5], [6] Misalnya, penelitian pada pengukuran kesamaan judul skripsi menunjukkan bahwa algoritma ini dapat mendeteksi kemiripan secara konsisten, bahkan dengan variasi panjang teks yang berbeda. [6]

Sementara itu, algoritma *Boyer–Moore* merupakan algoritma *string matching* yang dikenal efisien dalam menemukan pola teks di dalam dokumen. Algoritma ini melakukan pencarian dari kanan ke kiri dan menggunakan dua heuristik utama, yaitu *bad character shift* dan *good suffix shift*, untuk mempercepat proses pencarian. [7] Beberapa penelitian membuktikan bahwa algoritma *Boyer–Moore* mampu menurunkan waktu pencarian secara signifikan dibandingkan dengan metode pencarian teks lainnya seperti *Brute Force* atau *Knuth–Morris–Pratt*. [8], [9]. Dalam implementasinya, *Boyer–Moore* efektif digunakan untuk menangani pencocokan teks dalam dokumen berukuran sedang hingga besar dengan stabilitas performa yang baik. [2]

Berdasarkan hal tersebut, efektivitas algoritma dalam sistem pencarian dokumen digital dapat dilihat dari dua aspek utama, yaitu akurasi hasil pencarian dan efisiensi waktu proses. Algoritma *Cosine Similarity* dinilai efektif dari segi kemampuan mengukur tingkat kemiripan antar dokumen secara semantik. [1], [10] Sedangkan algoritma *Boyer–Moore* dinilai efektif dari segi kecepatan dalam mengeksekusi proses pencarian teks. [7], [8] Dengan kata lain, *Cosine Similarity* lebih tepat digunakan untuk menganalisis

kemiripan isi dokumen, sedangkan algoritma *Boyer-Moore* lebih sesuai jika kebutuhan utamanya adalah efisiensi waktu pencarian.

Dengan demikian, penelitian ini difokuskan untuk menganalisis efektivitas penerapan kedua algoritma tersebut dalam konteks sistem pencarian dokumen digital, sesuai dengan karakteristik dan tujuan masing-masing algoritma. Urgensi penelitian ini terletak pada dua aspek, yaitu:

1. Aspek praktis, hasil penelitian diharapkan dapat menjadi acuan dalam penerapan algoritma yang sesuai dengan kebutuhan sistem pencarian dokumen digital, baik dari sisi ketepatan hasil (*Cosine Similarity*) maupun efisiensi waktu (*Boyer-Moore*).
2. Aspek teoritis, penelitian ini memberikan kontribusi dalam pengembangan kajian efektivitas algoritma di bidang *information retrieval*, khususnya dalam konteks dokumen digital berskala kecil hingga menengah. [4], [5]

Dengan demikian, penelitian ini diarahkan untuk memberikan gambaran yang lebih jelas mengenai analisis efektivitas kedua algoritma. Berdasarkan uraian tersebut, judul penelitian yang diangkat adalah “Analisis Efektivitas Algoritma *Cosine Similarity* dan algoritma *Boyer-Moore* dalam Pencarian Dokumen Digital Dinas Kesehatan.”.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, rumusan masalah dalam penelitian ini mencakup tiga aspek utama, yaitu bagaimana efektivitas algoritma *Cosine Similarity* dalam mengukur tingkat kemiripan dokumen digital, bagaimana efektivitas algoritma *Boyer-Moore* dalam melakukan pencocokan string pada dokumen digital, serta bagaimana analisis efektivitas antara algoritma *Cosine Similarity* dan *Boyer-Moore*.

1.3 Batasan Masalah

Agar penelitian ini dapat berjalan dengan fokus dan terarah, maka peneliti memberikan batasan masalah sebagai berikut:

- a) Penelitian ini menganalisis dan membandingkan efektivitas Algoritma *Cosine Similarity* dan Algoritma *Boyer-Moore* saja.
- b) Dataset terbatas pada 56 dokumen dari Dinas Kesehatan Kabupaten Bengkalis dan 55 query uji.
- c) Evaluasi menggunakan metrik klasifikasi (Precision, Recall, F1-Score, Accuracy), statistik (untuk *Cosine*), dan waktu eksekusi, dengan ground truth semi-otomatis berdasarkan metadata.

- d) Penelitian dilakukan pada sistem SIPANDOK sebagai media uji.
- e) Tidak membahas algoritme pencarian lain, teknik indexing, atau optimasi performa menyeluruh.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dipaparkan, maka penelitian ini bertujuan untuk:

- a) Menganalisis efektivitas algoritma *Cosine Similarity* dalam mengukur kemiripan antar dokumen digital berdasarkan skor kesamaan.
- b) Menganalisis efektivitas algoritma *Boyer-Moore* dalam melakukan pencocokan *string* berdasarkan efisiensi waktu eksekusi.
- c) Membandingkan hasil efektivitas kedua algoritma untuk mengetahui keunggulan dan kelemahan masing-masing dalam konteks sistem pencarian dokumen digital pada dataset terbatas.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

- a. Manfaat Teoretis
 - 1) Menambah literatur ilmiah mengenai analisis komparatif algoritma *Cosine Similarity* dan algoritma *Boyer-Moore* dalam bidang *information retrieval*.
 - 2) Memberikan pemahaman lebih mendalam tentang perbedaan prinsip kerja dan efektivitas kedua algoritma.
- b. Manfaat Praktis
 - 1) Memberikan rekomendasi pemilihan algoritma yang sesuai berdasarkan kebutuhan: apakah lebih menekankan kecepatan pencarian (*Boyer-Moore*) atau akurasi kesamaan makna (*Cosine Similarity*).
 - 2) Menjadi acuan bagi mahasiswa, peneliti, maupun praktisi yang ingin mengembangkan penelitian serupa dengan fokus pada analisis algoritma pencarian dan pengukuran kemiripan teks.