

ABSTRAK

Mesin pencari merupakan salah satu sarana penting untuk mengakses informasi di internet, dengan jutaan orang mengandalkan sistem ini untuk menemukan informasi yang dibutuhkan setiap harinya. Keakuratan hasil pencarian sangat bergantung pada algoritma yang digunakan untuk menilai relevansi antara *query* dan dokumen. Dalam penelitian ini dataset diperoleh dari laman <https://www.kaggle.com/datasets/sh1zuka/indonesia-news-dataset-2024> dengan kategori berita politik sebanyak 500 baris. Data uji berjumlah 15 *query* bervariasi dari 2 sampai 3 kata pada kalimat *query*. Penelitian ini bertujuan untuk mengevaluasi perbandingan presisi antara *BM25* dan *Cosine Similarity* dan mengidentifikasi pendekatan yang lebih efektif dalam meningkatkan kualitas hasil pencarian. Uji coba dilakukan menggunakan *precision@k* dengan nilai *k* berkisar dari 5 sampai 30 dengan kelipatan 5. Hasil evaluasi menunjukkan bahwa rata-rata presisi yang diperoleh oleh *BM25* adalah 0.8979, sedangkan *Cosine Similarity* mencapai 0.8554. Presisi tertinggi diperoleh pada *k*-5 sebesar 1 pada *BM25* dan 1 pada *Cosine Similarity*. Hal ini menunjukkan bahwa *BM25* memiliki kinerja yang lebih baik dalam memberikan hasil pencarian yang relevan dibandingkan dengan *Cosine Similarity* pada dataset berita yang digunakan.

Kata Kunci: Presisi, Temu Kembali Informasi, *BM25*, *Cosine Similarity*.

ABSTRACT

Search engines are one of the essential tools for accessing information on the internet, with millions of people relying on these systems to find the information they need every day. The accuracy of search results heavily depends on the algorithms used to assess the relevance between a query and a document. In this study, the dataset was obtained from the website <https://www.kaggle.com/datasets/sh1zuka/indonesia-news-dataset-2024>, which contains 500 rows of political news. The test data consists of 15 queries, varying from 2 to 3 words in the query sentences. This research aims to evaluate and compare the precision between BM25 and Cosine Similarity and identify the more effective approach in improving the quality of search results. The experiment was conducted using $\text{precision}@k$ with values of k ranging from 5 to 30, in increments of 5. The evaluation results showed that the average precision obtained by BM25 is 0.8979, while Cosine Similarity reached 0.8554. The highest precision was obtained at $k=5$ with 1 for BM25 and 1 for Cosine Similarity. This indicates that BM25 performs better in providing relevant search results compared to Cosine Similarity on the political news dataset used.

Keywords: Precision, Information Retrieval, BM25, Cosine Similarity