

ABSTRAK

Vandela, Hensel Revy. 2026. Sistem Rekomendasi Buku Berbasis *Content-Based Filtering* dengan Menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) dan Pendekatan *Cosine Similarity*. Tugas Akhir. Program Sarjana. Program Studi Teknik Informatika. Universitas Muhammadiyah Jember.

Pembimbing : Dr. Ilham Saifudin, S.Pd., M.Si; Dr. Reni Umilasari, S.Pd., M.Si

Kebutuhan masyarakat terhadap informasi dan pengetahuan terus meningkat seiring dengan semakin banyaknya sumber informasi yang tersedia. Banyaknya pilihan bacaan dapat menyebabkan kesulitan dalam menentukan buku yang sesuai dengan minat dan kebutuhan. Oleh karena itu, diperlukan sistem yang dapat membantu proses pencarian dan pemilihan buku yang relevan. Salah satu solusi yang dapat diterapkan adalah sistem rekomendasi dengan pendekatan *Content-Based Filtering*, yaitu pendekatan yang memberikan rekomendasi berdasarkan karakteristik atau kemiripan konten. Penelitian ini bertujuan untuk merancang dan mengimplementasikan sistem rekomendasi buku berbasis *Content-Based Filtering* menggunakan teknik *Term Frequency-Inverse Document Frequency* dan *Cosine Similarity*, serta mengetahui kinerjanya dalam menghasilkan rekomendasi yang relevan.

Data yang digunakan berjumlah 6.810 buku dengan atribut *title*, *authors*, *categories*, dan *description*. Proses penelitian meliputi *preprocessing* teks, pembobotan menggunakan *Term Frequency-Inverse Document Frequency*, dan perhitungan kemiripan menggunakan *Cosine Similarity*. Nilai kemiripan digunakan untuk menghasilkan rekomendasi berdasarkan tingkat kemiripan tertinggi. Sistem kemudian diimplementasikan dalam bentuk website dan dievaluasi menggunakan *Precision*, *Normalized Discounted Cumulative Gain* (NDCG), dan *Running Time* pada Top-3, Top-5, dan Top-10.

Hasil evaluasi menunjukkan nilai *Precision* sebesar 0,523, 0,518, dan 0,508 pada Top-3, Top-5, dan Top-10. Nilai NDCG sebesar 0,666, 0,688, dan 0,706, sedangkan *Running Time* sebesar 0,3051 detik, 0,3117 detik, dan 0,3469 detik. Hasil tersebut menunjukkan bahwa sistem mampu menghasilkan rekomendasi buku yang relevan dengan pengurutan berdasarkan tingkat kemiripan yang baik. Peningkatan nilai NDCG menunjukkan kualitas pengurutan rekomendasi yang tetap baik seiring bertambahnya jumlah rekomendasi. Berdasarkan hasil tersebut, *Content-Based Filtering* menggunakan *Term Frequency-Inverse Document Frequency* dan *Cosine Similarity* berhasil diterapkan pada sistem rekomendasi buku berbasis website.

Kata Kunci: sistem rekomendasi, *Content-Based Filtering*, TF-IDF, *Cosine Similarity*, rekomendasi buku

ABSTRACT

Vandela, Hensel Revy. 2026. *Book Recommendation System Based on Content-Based Filtering Using Term Frequency–Inverse Document Frequency (TF-IDF) and Cosine Similarity Approach*. Undergraduate Thesis. Undergraduate Program. Informatics Engineering Study Program. University of Muhammadiyah Jember.

Advisors: Dr. Ilham Saifudin, S.Pd., M.Si; Dr. Reni Umilasari, S.Pd., M.Si

As the availability of information sources continues to grow, people have access to an increasing number of books and other reading materials. However, the wide range of available books can make it difficult to find materials that match individual interests and needs. Therefore, a system is needed to assist in finding and selecting relevant books. One approach that can be used is Content-Based Filtering, which provides recommendations based on the characteristics and content similarity of items. This study aims to design and implement a book recommendation system using Content-Based Filtering with Term Frequency-Inverse Document Frequency and Cosine Similarity, and to evaluate its performance in generating relevant book recommendations.

The dataset consists of 6,810 books containing the attributes title, authors, categories, and description. The research process includes text preprocessing, text weighting using Term Frequency-Inverse Document Frequency, and similarity calculation using Cosine Similarity. The resulting similarity scores are used to generate book recommendations based on the highest similarity scores. The system is implemented as a web-based application and evaluated using Precision, Normalized Discounted Cumulative Gain (NDCG), and Running Time for Top-3, Top-5, and Top-10 recommendations.

The evaluation results show Precision values of 0.523, 0.518, and 0.508 for Top-3, Top-5, and Top-10, respectively. The NDCG values are 0.666, 0.688, and 0.706, while the Running Time values are 0.3051, 0.3117, and 0.3469 seconds, respectively. These results indicate that the system can generate relevant book recommendations and rank them effectively based on their similarity scores. The increase in NDCG values indicates that the ranking quality remains good as the number of recommendations increases. Based on these results, Content-Based Filtering using Term Frequency-Inverse Document Frequency and Cosine Similarity was successfully implemented in a web-based book recommendation system.

Keywords: recommendation system, Content-Based Filtering, TF-IDF, Cosine Similarity, book recommendation