

ABSTRAK

Irdianto, Airlangga Zabaniyah Putro. 2026. Sistem Rekomendasi Produk *Fitness* dan Olahraga Berbasis *Content-Based Filtering* Menggunakan *Doc2Vec* dan *Cosine Similarity*. Skripsi. Program Studi Sarjana Teknik Informatika, Universitas Muhammadiyah Jember.

Pembimbing: Dr. Ilham Saifudin, S.Pd., M.Si; Taufiq Timur W. S.Kom., M.Kom.

Sistem rekomendasi merupakan salah satu teknologi penyaringan informasi yang membantu pengguna menemukan produk sesuai kebutuhan di tengah banyaknya pilihan pada platform digital. Peningkatan jumlah dan variasi produk *fitness* dan olahraga dapat menyebabkan *information overload*. Kondisi tersebut membuat pengguna kesulitan menemukan produk yang relevan secara cepat. Pendekatan *content-based filtering* dapat digunakan untuk menghasilkan rekomendasi berdasarkan kemiripan karakteristik konten produk. Namun, pencocokan kata kunci secara langsung belum sepenuhnya mampu memahami hubungan semantik antara kebutuhan pengguna dan informasi produk.

Penelitian ini bertujuan membangun dan mengevaluasi sistem rekomendasi produk *fitness* dan olahraga berbasis *content-based filtering* menggunakan metode *Doc2Vec* dan *cosine similarity*. Penelitian menggunakan metode eksperimen kuantitatif dengan data yang bersumber dari *Amazon Reviews'23 Dataset*. Atribut yang digunakan meliputi *Product_name*, *category*, dan *clean_text*, sedangkan *Product_id* digunakan sebagai identitas produk. Data melalui tahapan seleksi, pembersihan teks, *case folding*, tokenisasi, dan penghapusan *stopword*. Selanjutnya, data dibagi menjadi 80% data latih dan 20% data uji menggunakan metode *hold-out validation*.

Model *Doc2Vec* menggunakan arsitektur *Paragraph Vector Distributed Bag of Words* atau PV-DBOW dengan ukuran vektor 100 dimensi, *window* 5, *min_count* 2, *epochs* 20, *workers* 1, dan *seed* 42. Vektor *query* pengguna dibentuk menggunakan fungsi *infer_vector*, kemudian dibandingkan dengan seluruh vektor produk pada data uji menggunakan *cosine similarity*. Produk diurutkan berdasarkan nilai kemiripan tertinggi dan disajikan dalam bentuk rekomendasi

Top-N. Evaluasi dilakukan menggunakan empat *query* pada skenario *Top-5*, *Top-10*, dan *Top-30* dengan metrik *Precision*, *Recall*, dan *Normalized Discounted Cumulative Gain* atau NDCG.

Hasil penelitian menunjukkan bahwa pada skenario *Top-5*, sistem memperoleh nilai rata-rata *Precision* sebesar 0,9500, *Recall* sebesar 0,1100, dan NDCG sebesar 0,9732. Pada skenario *Top-10*, sistem memperoleh *Precision* sebesar 0,9000, *Recall* sebesar 0,2078, dan NDCG sebesar 0,9584. Sementara itu, pada skenario *Top-30*, sistem memperoleh *Precision* sebesar 0,8750, *Recall* sebesar 0,6082, dan NDCG sebesar 0,9361. Hasil tersebut menunjukkan bahwa jumlah rekomendasi yang lebih sedikit menghasilkan tingkat ketepatan dan kualitas pemeringkatan yang lebih tinggi. Sebaliknya, jumlah rekomendasi yang lebih banyak meningkatkan kemampuan sistem dalam menemukan produk relevan.

Sistem selanjutnya diimplementasikan dalam bentuk *website* bernama Bolang menggunakan *Flask*. Berdasarkan hasil penelitian, kombinasi *Doc2Vec* dan *cosine similarity* dapat digunakan untuk menghasilkan rekomendasi produk *fitness* dan olahraga berdasarkan kemiripan semantik informasi tekstual produk.

Kata kunci: sistem rekomendasi, *content-based filtering*, *Doc2Vec*, *cosine similarity*, produk *fitness* dan olahraga.

ABSTRACT

Irdianto, Airlangga Zabaniyah Putro. 2026. *Fitness and Sports Product Recommendation System Based on Content-Based Filtering Using Doc2Vec and Cosine Similarity*. Undergraduate Thesis. Bachelor's Program in Informatics Engineering, Universitas Muhammadiyah Jember. Advisors: Dr. Ilham Saifudin, S.Pd., M.Si; Taufiq Timur W. S.Kom., M.Kom.

Recommendation systems are information-filtering technologies that help users find products suited to their needs among the numerous options available on digital platforms. The increasing number and variety of fitness and sports products may lead to information overload, making it difficult for users to identify relevant products efficiently. A content-based filtering approach can generate recommendations based on similarities in product content characteristics. However, direct keyword matching has a limited ability to understand semantic relationships between user needs and product information.

Therefore, this study aims to develop and evaluate a fitness and sports product recommendation system based on content-based filtering using Doc2Vec and cosine similarity. This study employed a quantitative experimental method using data obtained from the Amazon Reviews'23 Dataset. The attributes used in the recommendation process were Product_name, category, and clean_text, while Product_id was used as the product identifier. The data underwent selection, text cleaning, case folding, tokenization, and stopword removal. The dataset was then divided into 80% training data and 20% test data using the hold-out validation method.

The Doc2Vec model applied the Paragraph Vector Distributed Bag of Words architecture, or PV-DBOW, with a vector size of 100 dimensions, a window size of 5, a minimum word count of 2, 20 training epochs, one worker, and a random seed of 42. User query vectors were generated using the infer_vector function and compared with all product vectors in the test data using cosine similarity. Products were ranked according to the highest similarity values and presented as Top-N recommendations. The system was evaluated using four test queries under

Top-5, Top-10, and Top-30 scenarios with Precision, Recall, and Normalized Discounted Cumulative Gain, or NDCG, as evaluation metrics.

The results show that the Top-5 scenario achieved an average Precision of 0.9500, Recall of 0.1100, and NDCG of 0.9732. The Top-10 scenario achieved a Precision of 0.9000, Recall of 0.2078, and NDCG of 0.9584. Meanwhile, the Top-30 scenario achieved a Precision of 0.8750, Recall of 0.6082, and NDCG of 0.9361. These results indicate that a smaller number of recommendations produces higher accuracy and ranking quality, while a larger number of recommendations improves the system's ability to retrieve more relevant products.

The recommendation system was implemented as a Flask-based website named Bolang, allowing users to enter product-related queries and obtain recommendations directly. The findings demonstrate that the combination of Doc2Vec and cosine similarity can generate fitness and sports product recommendations based on the semantic similarity of textual product information.

Keywords: *recommendation system, content-based filtering, Doc2Vec, cosine similarity, fitness and sports products.*