

IMPLEMENTASI CONTENT-BASED FILTERING MENGGUNAKAN WORD EMBEDDINGS PADA SISTEM PENCARIAN FILM

Andani Chacha Cahya Dewi¹, Deni Arifianto², Nanda Kurnia Wardati³

¹²³Program Studi Sistem Informasi, Fakultas Teknik,
Universitas Muhammadiyah Jember

Email: andanichachacahyadewi@gmail.com¹,
deniarifianto@unmuhjember.ac.id², nandakurniawardati@unmuhjember.ac.id³

ABSTRAK

Pertumbuhan industri film dan layanan *streaming* digital memunculkan masalah kelebihan informasi (*information overload*) bagi pengguna. Sistem temu kembali informasi saat ini umumnya masih bergantung pada pencocokan kata berbasis frekuensi (seperti TF-IDF) yang kaku secara leksikal, mengabaikan makna semantik alur cerita, serta seringkali masih mensyaratkan judul film acuan (*seed movie*). Penelitian ini bertujuan untuk mengimplementasikan metode *Content-Based Filtering* menggunakan *Word Embeddings* (Word2Vec dengan arsitektur *Skip-gram*) dan *Cosine Similarity* pada sistem pencarian film. Pendekatan ini memungkinkan sistem mengekstraksi kedalaman makna dari teks sinopsis dan mencocokkannya secara langsung dengan masukan kata kunci (*keyword*) bebas dari pengguna. Penelitian ini menggunakan 5.000 data film berbahasa Inggris dari IMDb yang diperoleh melalui teknik *web scraping*. Pengujian kinerja algoritma difokuskan pada jendela pencarian tiga besar (*Top-3 Retrieval*), yang dievaluasi berdasarkan nilai kebenaran acuan (*ground truth*) dari validasi pakar menggunakan metode *Inter-Annotator Agreement* (IAA). Hasil pengujian menunjukkan bahwa sistem berkinerja sangat optimal dengan pencapaian nilai *Precision@3* sebesar 53,33%, *Recall@3* sebesar 78,00%, dan *Mean Reciprocal Rank* (MRR) sebesar 0,7300. Tingginya nilai MRR membuktikan bahwa algoritma berhasil memposisikan film yang relevan pada peringkat pertama atau kedua teratas. Kesimpulannya, integrasi Word2Vec terbukti tangguh dalam memahami variasi konteks semantik kueri pengguna untuk menghasilkan daftar pencarian film lintas genre yang relevan secara presisi.

Kata Kunci: *Content-Based Filtering*, *Word Embeddings*, *Word2Vec*, *Cosine Similarity*, *Sistem Temu Kembali Informasi*, *Web Scraping*.