

Sistem Rekomendasi Film Menggunakan Content-Based Filtering dan Klasifikasi Convolutional Neural Network (CNN)

Arliyanna Nilla¹, Erwin Budi Setiawan²

^{1,2}Fakultas Informatika, Universitas Telkom, Bandung

arliyannanilla@students.telkomuniversity.ac.id, erwinbudisetiawan@telkomuniversity.ac.id,

Abstrak

Mengelola data dalam jumlah besar merupakan tantangan yang dihadapi oleh pengguna, sehingga dibutuhkan sistem rekomendasi sebagai penyaring informasi untuk memberikan saran item yang relevan. Twitter sering digunakan untuk mencari informasi mengenai review film yang dapat digunakan sebagai dasar pengembangan sistem rekomendasi. Penelitian ini berkontribusi dalam menerapkan content-based filtering dalam konteks Convolutional Neural Network (CNN). Sejauh pengetahuan peneliti, belum ada penelitian yang membahas kombinasi metode dan klasifikasi ini. Fokus utamanya adalah mengevaluasi pengembangan sistem rekomendasi dengan mengintegrasikan dan membandingkan metode identifikasi kemiripan dengan menggunakan pendekatan RoBERTa dan TF-IDF. Pada penelitian ini, RoBERTa dan TF-IDF sebagai metode vectorizer dan klasifikasi diterapkan untuk membentuk sebuah model yang dapat mengenali pola pada data dan menghasilkan prediksi yang akurat berdasarkan fitur-fiturnya. Total data yang digunakan adalah 854 film dan 34086 ulasan film dari 44 akun Twitter. Metode SMOTE diterapkan sebagai teknik untuk mengatasi ketidakseimbangan data. Penelitian dilakukan sebanyak tiga kali dengan hasil akurasi yang semakin meningkat. Percobaan pertama TF-IDF sebagai baseline, SMOTE pada klasifikasi CNN. Eksperimen kedua, menerapkan baseline, SMOTE, embedding pada klasifikasi CNN. Percobaan ketiga, menerapkan baseline, SMOTE, embedding, dan optimizer pada klasifikasi CNN. Hasil percobaan menunjukkan bahwa TF-IDF sebagai baseline, SMOTE, embedding, dan SGD optimizer dengan learning rate terbaik pada klasifikasi CNN dapat memberikan hasil yang optimal dengan tingkat akurasi sebesar 86.41%. Dengan demikian, sistem dapat memberikan rekomendasi film yang relevan dengan akurasi prediksi dan performa yang baik.

Kata kunci : Recommender System, Twitter, *Content based filtering*, Word Embedding, RoBERTa, TFIDF, Classification, Convolutional Neural Network.