

ABSTRACT

The rapid growth of e-commerce has made it easier for consumers to access a wide range of products, yet it also presents challenges in identifying items that match individual preferences. To address this issue, this study develops a hybrid recommendation model that integrates sentiment analysis using the RoBERTa model and a User-Based Collaborative Filtering (UBCF) approach. Sentiment analysis is applied to product reviews in the Amazon All Beauty dataset, with sentiment classification (positive or negative) performed using a fine-tuned RoBERTa-base model. The model demonstrates strong performance, achieving an Accuracy of 97.19% and an F1-score of 97.19%. Rating prediction is conducted through cosine similarity calculations between users in the UBCF approach. Final recommendation scores are generated through a proportional integration of the predicted ratings and sentiment scores, evaluated using the Root Mean Square Error (RMSE) metric. The optimal result is obtained with an integration weight of $\lambda = 0.3$, yielding an RMSE of 0.8355—significantly outperforming the baseline model without sentiment (RMSE 1.4695) and previous approaches such as GloVe + BiLSTM. The hybrid model is capable of generating more relevant and personalized Top-N recommendations, achieving an accuracy improvement of 43.15% compared to the baseline model and 68.24% compared to the previous approach (GloVe + BiLSTM), by leveraging user opinions expressed in product reviews.

Keywords: *Cosine Similarity, Hybrid Recommender Model, RoBERTa, Sentiment Analysis, User Based Collaborative Filtering*

ABSTRAK

Pertumbuhan *e-commerce* yang pesat telah memudahkan konsumen dalam mengakses berbagai produk, namun juga menimbulkan tantangan dalam mengidentifikasi produk yang sesuai dengan preferensi individu. Untuk mengatasi permasalahan tersebut, penelitian ini mengembangkan model rekomendasi *hybrid* yang mengintegrasikan analisis sentimen menggunakan model *RoBERTa* dan pendekatan *User-Based Collaborative Filtering* (UBCF). Analisis sentimen diterapkan pada ulasan produk dalam dataset *Amazon* kategori *All Beauty*, dengan klasifikasi sentimen (positif atau negatif) dilakukan menggunakan model *RoBERTa-base* yang telah di-*fine-tune*. Model ini menunjukkan performa yang kuat dengan *Accuracy* sebesar 97,19% dan *F1-Score* sebesar 97,19%. Prediksi *rating* dilakukan melalui perhitungan *cosine similarity* antar pengguna dalam pendekatan UBCF. Skor rekomendasi akhir dihasilkan melalui integrasi proporsional antara *rating* yang diprediksi dan skor sentimen, yang dievaluasi menggunakan metrik *Root Mean Square Error* (RMSE). Hasil optimal diperoleh pada bobot integrasi $\lambda = 0,3$ dengan RMSE sebesar 0,8355—secara signifikan lebih baik dibandingkan model dasar tanpa sentimen (RMSE 1,4695) dan pendekatan sebelumnya seperti *GloVe + BiLSTM*. Model *hybrid* ini mampu menghasilkan rekomendasi Top-N yang lebih relevan dan personal, dengan peningkatan akurasi sebesar 43,15% dibandingkan model dasar dan 68,24% dibandingkan pendekatan sebelumnya (*GloVe + BiLSTM*), melalui pemanfaatan opini pengguna dalam ulasan produk.

Kata Kunci: Analisis Sentimen, *Cosine Similarity*, Model Rekomendasi *Hybrid*, *RoBERTa*, *User-Based Collaborative Filtering*