

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

¹arliyannanilla@students.telkomuniversity.ac.id, ²erwinbudisetiawan@telkomuniversity.ac.id,

Abstract

Managing large amounts of data is a challenge faced by users, so a recommendation system is needed as an information filter to provide relevant item suggestions. Twitter is often used to find information about movie reviews that can be used as a basis for developing recommendation systems. This research contributes to applying content-based filtering in the context of Convolutional Neural Network (CNN). To the best of the researcher's knowledge, there has been no research addressing this combination of method and classification. The main focus is to evaluate the development of a recommendation system by integrating and comparing similarity identification methods using the RoBERTa and TF-IDF approaches. In this research, Roberta and TF-IDF as vectorizer and classification methods are applied to form a model that can recognize patterns in data and produce accurate predictions based on its features. The total data used is 854 movies and 34086 film reviews from 44 Twitter accounts. The SMOTE method was applied as a technique to overcome data imbalance. The research was conducted three times with increasing accuracy results. The first experiment TF-IDF as baseline, SMOTE on CNN classification. The second experiment, applying baseline, SMOTE, embedding on CNN classification. The third experiment applied baseline, SMOTE, embedding, and optimizer to CNN classification. The experimental results show that TF-IDF as baseline, SMOTE, embedding and SGD optimizer with the best learning rate on CNN classification can provide optimal results with an accuracy rate of 86.41%. Thus, the system can provide relevant movie recommendations with good prediction accuracy and performance.

Keywords: Recommender System, Twitter, Content based filtering, Word Embedding, RoBERTa, TFIDE, Classification, Convolutional Neural Network.