ABSTRAK

Industri perfilman terus berkembang dengan jumlah film yang semakin meningkat setiap tahunnya. Klasifikasi genre film menjadi aspek penting dalam pengelompokan dan rekomendasi film kepada penonton. Selain poster film sebagai elemen visual utama dalam pemasaran, informasi tekstual seperti plot film juga dapat digunakan untuk meningkatkan akurasi prediksi genre secara otomatis. Penelitian ini membandingkan performa *Convolutional Neural Network* (CNN) dan *Contrastive Language-Image Pretraining* (CLIP) dalam tugas klasifikasi genre film berbasis analisis poster dan plot film. Dataset diperoleh dari IMDb dan OMDb, yang kemudian melalui tahap preprocessing untuk memastikan kualitas data. Model CNN menggunakan arsitektur BiT-ResNet50, sementara CLIP menggunakan ViT-B/16, ViT-L/14, dan RN50x16 untuk analisis poster, serta BERT untuk analisis plot film. Eksperimen dilakukan dengan berbagai kombinasi hyperparameter, termasuk variasi batch size, learning rate, dan optimizer. Hasil penelitian menunjukkan bahwa CLIP memiliki performa lebih unggul dibandingkan CNN dalam tugas klasifikasi genre film. Model CLIP dengan arsitektur ViT-L/14 mencapai akurasi tertinggi sebesar 83,2% dengan nilai Hamming Loss 0,1679, sedangkan CNN mencapai akurasi 77,8%. Selain itu, penambahan analisis plot film dengan model BERT meningkatkan performa klasifikasi multi-label secara signifikan, dengan peningkatan akurasi sekitar 5% dibandingkan metode berbasis poster saja. Kesimpulan dari penelitian ini menunjukkan bahwa kombinasi metode berbasis vision-language model (CLIP) dan analisis teks (BERT) lebih efektif dibandingkan metode konvensional berbasis CNN dalam tugas klasifikasi genre film.

Kata kunci: klasifikasi genre film, CNN, CLIP, *deep learning*, poster film, *multi-label classification.*

ABSTRAC

The film industry continues to grow, with an increasing number of films produced each year. Film genre classification plays a crucial role in categorizing and recommending films to audiences. In addition to movie posters as the primary visual element in marketing, textual information such as film plots can also enhance the accuracy of automatic genre prediction. This study compares the performance of Convolutional Neural Network (CNN) and Contrastive Language-Image Pretraining (CLIP) in film genre classification based on poster and plot analysis. The dataset was obtained from IMDb and OMDb, followed by a preprocessing stage to ensure data quality. The CNN model utilizes the BiT-ResNet50 architecture, while CLIP employs ViT-B/16, ViT-L/14, and RN50x16 for poster analysis and BERT for plot analysis. Experiments were conducted with various hyperparameter combinations, including batch size, learning rate, and optimizer variations. The results indicate that CLIP outperforms CNN in film genre classification tasks. The CLIP model with the ViT-L/14 architecture achieved the highest accuracy of 83.2% with a Hamming Loss of 0.1679, while CNN reached an accuracy of 77.8%. Furthermore, incorporating film plot analysis using the BERT model significantly improved multi-label classification performance, with an approximately 5% accuracy increase compared to poster-based methods alone. This study concludes that the combination of vision-language models (CLIP) and text analysis (BERT) is more effective than conventional CNN-based methods for film genre classification.

Keywords: film genre classification, CNN, CLIP, deep learning, movie posters, multi-label classification.