ABSTRACT

Virtual YouTubers (VTubers) have become a rapidly growing form of digital content in Indonesia. Video Engagement level serves as an important indicator for evaluating the success of VTuber content, as it reflects audience interaction through metrics such as views, likes and comments. Understanding the engagement levels is essential for both VTubers and agencies to assess content performance and develop effective channel growth strategies. However, studies that specifically construct predictive models for estimating engagement levels of Indonesian VTuber videos remain limited. This research aims to apply the Random Forest algorithm to predict the engagement level of Indonesian VTuber videos. The *dataset* consists of 37,406 videos from 50 Indonesian VTuber channels, collected over the period 2020-2024 and selected based on overall channel view rankings. Data were gathered through web scraping using the YouTube API, encompassing features such as interaction metrics, video duration, video age, channel age, subscriber count, upload time, livestream status, content category, title length, number of tags and agency affiliation. The research stages include data cleaning, data transformation, one-hot encoding for categorical features, classification of engagement values using quartile boundaries derived from the IQR method, feature selection and train-test splitting. The Random Forest model was trained using the hyperparameters n_estimators = 300, max_depth = 2, min_samples_split = 5, min_samples_leaf = 1, and class_weight = 'balanced'. The model achieved an accuracy of 80.23%, with a macro-average F1-score of 0.80 and a weighted-average F1-score of 0.81. It successfully classified the three engagement classes, where the 'Low' class achieved the highest F1-score (0.84), while the 'High' class showed the lowest performance (0.76). Feature importance analysis identified video_age, duration, subscribers, and channel age months as the most influential features in predicting engagement.

Keywords : Virtual YouTuber, YouTube, *engagement*, *Random Forest*, classification