Analisis Sentimen Terhadap Ulasan Film Menggunakan Algoritma *Random Forest*

Muhammad Asjad Adna Jihad[1], Adiwijaya[2], Widi Astuti [3]

[1,2,3]Fakultas Informatika, Universitas Telkom, Bandung
[1]adnajihad@students.telkomuniversity.ac.id, [2]adiwijaya@telkomuniversity.ac.id,
[3]astutiwidi@telkomuniversity.ac.id

Abstract
The more public interest in watching the film increases, the more forums that contain discussions and comments/reviews of the audience on the film appeared. The results of the reviews in these forums became important information that could be at risk if the public happens to misinterpret during understanding it. Sentiment analysis techniques are used to prevent this, where sentiment analysis will analyze each text to determine the aspect and level of sentiment (positive or negative). Before starting the processing of the review results that have been collected, there's a step that will be carried out first which is the preprocessing stage so that the data could be easier to proceed to the next stage. Then, feature extraction is performed using Word2Vec, and the Random Forest method is used to classify sentiments. From the research that has been done, it was found that a particular research scenario using the 300-dimensional Word2Vec skip-gram, and then applying Adaptive Boosting to the base model, had the best accuracy value of 75.76%.

Keywords: review imdb, sentiment analysis, preprocessing, Word2Vec, random forest