

Analisis Sentimen terhadap Ulasan Film Berbahasa Inggris Menggunakan Metode *Support Vector Machine* dengan Feature Selection *Information Gain*

**Nauffan Muti Hibattullah¹, Said Al Faraby²,
Mahendra Dwifebri Purbolaksono³**

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹nauffanmufti@students.telkomuniversity.ac.id,

²saidalfaraby@telkomuniversity.ac.id,³mahendradp@telkomuniversity.ac.id

Abstract

Sentiment analysis is a field that analyzes the opinions, attitudes, and emotions of many people towards a product, service or other entity, this field of research is quite popular. Review text is a text whose contents are in the form of reviews, reviews or ratings on a work such as films, dramas, and books. The review text has the function of assessing, weighing, and criticizing the works or events being reviewed. English is the most widely spoken language in the world. There are 400 million native speakers and 2 billion people learn it as a second language. Support Vector Machine (SVM) is one of the methods in supervised learning that's usually used for classification (such as Support Vector Classification) and regression (Support Vector Regression). The Information Gain method is a method that uses a scoring technique to weight a feature by using maximum entropy. The selected feature is a feature with an Information Gain value that is greater than or equal to a certain threshold value. This study shows the process by using a combination of Stopword and Stemming, the resulting accuracy will be maximized, because the preprocessing process is more complete at 86.12%. In addition, it was found that the Information Gain (IG) feature selection in this study made the accuracy low, but it could be a good enough solution to overcome the overfitting problem in this sentiment analysis test. And the classification of movie review sentiment analysis is very precise using the Linear kernel Support Vector Machine (SVM) algorithm. Because Linear focuses on features that contain binary values obtained from the best hyperplane by maximizing the distance between classes.

Keywords: Sentiment Analysis, Reviews, Film, English Language, Support Vector Machine, Information Gain
