

ABSTRAK

Ulasan film merupakan salah satu bentuk data yang memiliki potensi besar untuk dianalisis guna memahami sentimen penonton dan preferensi mereka terhadap sebuah film. Kaskus adalah sebuah situs forum online yang menawarkan ruang untuk mencari informasi dan berbagi minat. Salah satu bagian dari forum tersebut adalah forum *Movies*, yang berisi berbagai film dan ulasan terkait film yang telah ditonton. Klasifikasi teks menjadi langkah penting untuk mengelompokkan ulasan tersebut ke dalam kategori positif dan negatif. Penelitian ini bertujuan untuk mengklasifikasikan teks ulasan film berbahasa Indonesia ke dalam kategori sentimen positif dan negatif menggunakan metode *Naive Bayes* dengan pendekatan *Lexicon-Based*. Penelitian ini menggunakan data ulasan film yang dikumpulkan dari situs Kaskus.com. Data yang sudah di kumpulkan kemudian melalui tahapan *preprocessing* data meliputi *case folding*, *cleaning*, *tokenisasi*, *normalisasi*, *stopwords removal*, dan *stemming*. Selanjutnya, data dibagi menjadi 80:20 data latih dan data uji, dimana setiap data terdiri dari dua kelas yaitu positif dan negatif. Algoritma *Naive Bayes*, diterapkan untuk membangun model *klasifikasi*. Evaluasi performa dilakukan menggunakan *confusion matrix* yang mencakup *accuracy*, *precision*, *recall*, dan *F1-score*. Berdasarkan hasil klasifikasi, model menghasilkan nilai *accuracy* sebesar 0,84 dengan nilai *precision* 0,84 *recall* 0,84 dan *F1-score* 0,84. Penelitian ini diharapkan dapat memberikan wawasan mengenai efektivitas pendekatan *Lexicon-Based* dalam meningkatkan akurasi klasifikasi teks ulasan film serta menjadi referensi bagi pengembang sistem dalam pemrosesan bahasa alami.

Kata Kunci : *Kaskus, Klasifikasi teks, Lexicon-Based, Naive Bayes, Ulasan film*

ABSTRACT

Movie reviews are a valuable source of data that can be analyzed to understand audience sentiment and preferences toward a film. Kaskus is an online forum that provides a space for users to share information and interests. One of its sections, the Movies forum, contains various discussions and reviews of films watched by users. Text classification plays an essential role in categorizing these reviews into positive and negative sentiments. This study aims to classify Indonesian-language movie review texts into positive and negative sentiment categories using the Naive Bayes algorithm with a Lexicon-Based approach. The dataset used in this study was collected from the website Kaskus.com. The collected data underwent several preprocessing stages, including case folding, cleaning, tokenization, normalization, stopword removal, and stemming. The dataset was then divided into 80% training data and 20% testing data, each consisting of two sentiment classes: positive and negative. The Naive Bayes algorithm was applied to build the classification model. The model's performance was evaluated using a confusion matrix that includes accuracy, precision, recall, and F1-score. The classification results showed that the model achieved an accuracy of 0.84, with precision, recall, and F1-score values all reaching 0.84. This study is expected to provide insights into the effectiveness of the Lexicon-Based approach in improving text classification accuracy for movie reviews and to serve as a reference for system developers in the field of natural language processing.

Keywords: *Kaskus, Text Classification, Lexicon-Based, Naive Bayes, Movie Reviews*