

ABSTRAK

Nama : Muhammad Iqbal Yusril
NIM : 41519010106
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Perbandingan Algoritma Naïve Bayes dan Logistic Regression untuk Analisa Review Film pada Internet Movie Database
Pembimbing : Dr. Afiyati, S.Si., MT.

Film merupakan salah satu bentuk hiburan yang populer di seluruh dunia, termasuk di Indonesia, yang diminati oleh berbagai kalangan, baik anak-anak maupun orang dewasa. Pada tahun 2020, terjadi peningkatan signifikan dalam jumlah penonton film, terutama karena adanya pandemi *COVID-19* yang membuat minat penonton semakin tinggi. Opini masyarakat terhadap film yang mereka tonton dapat bersifat positif atau negatif. Opini-opini ini sering ditemukan dalam bentuk *review* dan dapat dianalisis menggunakan teknik analisis sentimen. Dengan menggunakan analisis sentimen pada kumpulan *review* film, masyarakat dapat memperoleh rekomendasi tentang film-film yang layak ditonton. Salah satu *platform* yang populer untuk membaca ulasan dan *review* film adalah *Internet Movie Database* (IMDb), yang menyediakan informasi lengkap tentang berbagai film beserta ulasan pengguna. Dalam penelitian ini, dilakukan analisis sentimen *review* pengguna terhadap film dengan menggunakan algoritma Naïve Bayes dan Logistic Regression. Dataset yang digunakan dalam penelitian ini diperoleh melalui proses *scraping* data *review* dari *website IMDb*. Selanjutnya, dataset tersebut menjalani tahapan *preprocessing* seperti pembersihan dan pengubahan format, serta dilakukan pelabelan sentimen menjadi *positive* dan *negative* berdasarkan *review* dan jumlah *rating*. Pengujian pada model algoritma dilakukan dengan menggunakan *Confusion Matrix*. Didapatkan bahwa algoritma Naïve Bayes menghasilkan akurasi sebesar 82% dan Logistic Regression sebesar 83%. Hasil ini menunjukkan bahwa algoritma Logistic Regression menghasilkan akurasi yang sedikit lebih baik, karena algoritma tersebut cocok digunakan untuk dapat menangani data yang tidak seimbang.

Kata Kunci : Analisis Sentimen, *Review* Film, Klasifikasi, IMDb, Naïve Bayes, Logistic Regression

ABSTRACT

Name : Muhammad Iqbal Yusril
NIM : 41519010106
Study Program : Informatic Engineering
Title Thesis : Comparison of Naïve Bayes and Logistic Regression Algorithms for Film Review Analysis on Internet Movie Database
Counsellor : Dr. Afiyati, S.Si., MT.

Film is a popular form of entertainment worldwide, including in Indonesia, enjoyed by people of all ages, both children and adults. In 2020, there was a significant increase in the number of film viewers, especially due to the COVID-19 pandemic, which heightened the audience's interest. People's opinions about the films they watch can be positive or negative. These opinions are often found in the form of reviews and can be analyzed using sentiment analysis techniques. By using sentiment analysis on a collection of film reviews, the public can obtain recommendations for films worth watching. One popular platform for reading film reviews is Internet Movie Database (IMDb), which provides comprehensive information about various films along with user reviews. In this study, sentiment analysis of user reviews on films was conducted using the Naive Bayes and Logistic Regression algorithms. The dataset used in this study was obtained through the process of scraping review data from the IMDb website. Subsequently, the dataset underwent preprocessing steps such as cleaning and format conversion, and sentiment labeling was performed, classifying them as positive or negative based on the reviews and rating scores. The performance of the algorithms was evaluated using Confusion Matrix. The results showed that the Naive Bayes algorithm achieved an accuracy of 82%, while Logistic Regression achieved an accuracy of 83%. This indicates that the Logistic Regression algorithm performed slightly better, as it is suitable for handling imbalanced data.

Keywords : Sentiment Analysis, Movie Reviews, Classification, IMDb, Naïve Bayes, Logistic Regression