

Analisis Sentimen Pada Ulasan Film menggunakan Kombinasi Seleksi Fitur Information Gain dan Document Frequency

FeyDevastioPutraNaldi¹, Adiwijaya², WidiAstuti³

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

¹feydevastio@students.telkomuniversity.ac.id, ²adiwijaya@telkomuniversity.ac.id,

³widiastuti@telkomuniversity.ac.id

Abstrak

Ulasan film merupakan salah cara untuk menentukan sebuah film layak ditonton atau tidak. Ulasan dapat dikategorikan sebagai positif atau negatif. Untuk menangani ulasan film ini, salah satu metode machine learning yang dapat digunakan adalah analisis sentimen. Pada penelitian ini akan diterapkan metode analisis sentimen pada ulasan film dengan menerapkan kombinasi Information Gain (IG) dan Document Frequency (DF) serta Negation Handling yang akan diterapkan pada classifier Multinomial Naïve Bayes untuk mengetahui apakah metode tersebut dapat mempengaruhi performa dari classifier atau tidak. Penelitian ini dibagi menjadi empat skenario untuk mendapatkan model terbaik, pertama adalah skenario dengan nilai minimal IG *threshold* 0,5, kedua adalah penurunan nilai minimal IG *threshold* hingga 0,1, ketiga adalah mengambil jumlah fitur dengan jumlah yang ditentukan dan keempat perbandingan performa dari model terhadap penerapan Negation Handling dengan tanpa penerapan Negation Handling. Skenario yang mendapatkan performa paling optimal adalah skenario pengambilan jumlah fitur dengan penerapan Negation Handling yang mampu meningkatkan performa hingga nilai maksimal yang didapatkan adalah 73,05% dan baris data yang tidak memiliki fitur hingga rata-rata 0,3 baris data saja.

Kata kunci: Ulasan Film, Information Gain, Document Frequency, Multinomial Naïve Bayes

Abstract

Movie reviews are one way to determine whether a film is worth watching or not. Reviews can be categorized as positive or negative. One of the machine learning methods that can be used with movie review dataset is sentiment analysis. This research will apply the sentiment analysis method to film reviews by applying a combination of Information Gain (IG) and Document Frequency (DF) and Negation Handling which will be applied to the Multinomial Naïve Bayes classifier to see whether this method can affect the performance of the classifier or not. This study is divided into four scenarios to get the best model, first is a scenario with a minimum IG threshold value of 0.5, second is a decrease in the minimum IG threshold value to 0.1, third is taking the number of features with a specified number and the fourth is comparison of performance from model for the application of negation handling and without. The scenario that gets the most optimal performance is third scenario with the application of Negation Handling which can increase performance up to 73.05% and data rows that do not have features up to an average of only 0.3 rows of data.

Keywords: Movie Reviews, Information Gain, Document Frequency, Multinomial Naïve Bayes

1. Pendahuluan

Latar Belakang

Pada era digital seperti saat ini, terdapat banyak hiburan yang dapat dinikmati dan salah satu hiburan tersebut adalah film. Untuk menentukan film apa yang layak ditonton atau tidak, salah satu caranya adalah dengan melihat ulasan mengenai film tersebut. Ulasan dapat berasal dari pakar atau penonton film yang berisikan opini personal terhadap film yang mereka tonton. Secara umum, ulasan dapat dikategorikan sebagai ulasan positif atau negatif. Untuk menangani data seperti ulasan film ini, salah satu metode machine learning yang dapat digunakan adalah analisis sentimen. Analisis sentimen atau *Opinion Mining* merupakan sebuah bidang studi yang menganalisis sebuah pendapat tentang suatu hal dan mengklasifikasikannya ke dalam beberapa kategori sentimen. Analisis sentimen dapat dibagi menjadi tiga jenis tingkatan [1], pertama adalah *document level* yang melakukan proses klasifikasi pada satu dokumen penuh. Kedua adalah *sentence level* yang melakukan proses klasifikasi pada kalimat yang ada pada dokumen. Ketiga adalah *aspect level* yang berfokus pada melakukan proses klasifikasi pada tiap aspek yang ada pada dokumen tersebut.

Pada penerapan analisis sentimen di dataset ulasan film terdapat karakteristik yang membedakan antara