

1. Pendahuluan

Dikarenakan berkembang pesatnya teknologi informasi pada saat ini, menjadikan beberapa media dapat lebih mudah diakses oleh orang – orang dalam mencari suatu informasi dan mengungkapkan opini mengenai film yang telah dilihatnya [1]. Website review film merupakan salah satu yang terkena dampak dari pesatnya perkembangan teknologi informasi yang ada pada saat ini. Salah satu yang terdampak adalah Rotten Tomatoes yang merupakan sebuah situs yang berisi opini yang berupa ulasan dari berbagai individu atau golongan tentang pendapat mereka mengenai berbagai film.

Orang – orang dapat dengan mudah mengakses ulasan yang telah diberikan oleh orang lain dari berbagai macam film yang diinginkan dan juga dapat menuliskan ulasan yang merupakan opini mengenai kepuasan mereka mengenai film yang telah mereka tonton [2]. Sentiment Analysis juga dikenal sebagai metode untuk menganalisis opini dan mengklasifikasikan teks opini yang didapat berdasarkan tipe nya menjadi beberapa kelas yang berbeda sehingga membuat pemrosesan data menjadi lebih mudah [3]. Setiap harinya jutaan data mengenai ulasan seseorang mengenai film yang telah mereka tonton dikirim melalui website review film. Data yang telah dikumpulkan kemudian dikategorikan menjadi beberapa kelas seperti positif, netral dan negative [3].

Data yang diambil merupakan data yang menggunakan berbahasa inggris dari website Rotten Tomatoes. Ulasan yang masuk kedalam klasifikasi positif menandakan bahwa penonton puas dengan apa yang disajikan dalam sebuah film, klasifikasi netral menandakan bahwa ulasan tidak mengandung makna negatif ataupun positif, sedangkan ulasan negatif menandakan bahwa penonton memiliki berbagai keluhan atau bahkan kecewa terhadap apa yang disajikan dalam film yang telah mereka tonton. Berdasarkan klasifikasi tersebut, dapat dilihat bahwa cara pengungkapan opini masing – masing orang berbeda [4]. Metode yang ditawarkan untuk analisis sentiment ini adalah metode klasifikasi Support Vector Machine dengan menggunakan fitur ekstraksi Word2Vec.

Penelitian ini [5] menjelaskan bahwa fitur ekstraksi Word2Vec sebagai fitur ekstraksi sangat bagus digunakan karena fitur ekstraksi ini mewakili setiap kata dengan sebuah vektor, selain itu Word2Vec mempertahankan arti sintaksis dari kata – kata dan mengaturnya dengan kesamaan sintaksis tiap katanya, sehingga dalam ruang vektor Word2Vec, kata – kata yang struktur sintaksis nya mirip akan tetap berdekatan tetapi kata – kata dari polaritas sentiment yang berlawanan juga dapat tetap lebih dekat yang dapat menyebabkan ambiguitas pada hasil analisis sentiment. Jadi, dalam fitur ekstraksi ini, polaritas skor setiap kata memiliki peran penting terhadap hasil analisis sentimen. Metode Support Vector Machine dipilih karena akurasi yang didapat relatif tinggi dan tingkat generalisasi pada metode ini tidak dipengaruhi banyaknya data latih [6].

Dalam tugas akhir ini akan dilakukan analisis sentimen pada website Rotten Tomatoes yang berkaitan dengan review personal mengenai berbagai macam film. Tujuan dari penelitian ini adalah untuk mengetahui model terbaik yang dapat digunakan untuk mengolah dataset menggunakan kernel dalam metode klasifikasi SVM dengan fitur ekstraksi Word2Vec. Karena terdapat beberapa kernel pada metode SVM, maka hasil dari penelitian ini akan disimpulkan berdasarkan model terbaik dari penggunaan kernel dan kombinasi dari penggunaan teknik pengolahan data yang tepat. Penelitian ini diharapkan akan mendapatkan hasil berupa data yang informatif karena berisi tentang analisis sentimen penonton mengenai sebuah karya yang berbentuk film. Data hasil penelitian ini diharapkan dapat membantu pihak – pihak yang berkaitan dengan pembuatan film dalam hal evaluasi karya yang telah dibuat. Selain itu, penelitian ini diharapkan dapat meningkatkan akurasi dari metode SVM dengan adanya implementasi fitur ekstraksi Word2Vec.