

1. Pendahuluan

Latar Belakang

Secara umum film dapat dinilai dalam berbagai aspek, dimana yang akan membuat film tersebut disukai. Contohnya dalam penyajian cerita, efek visual, dan lain sebagainya. Penilaian dalam aspek-aspek tersebut dapat diuraikan dalam suatu ulasan atau opini. Dengan adanya opini masyarakat dalam ulasan film tersebut, akan menjadi hal penting bagi kita untuk memperhatikan sebelum membuat keputusan. Dalam mengolah informasi tersebut, dapat dilakukan dengan cara analisis sentimen. Analisis sentimen sendiri adalah sebuah proses yang berfungsi untuk mengklasifikasi isi dari *dataset* apakah tergolong teks bersifat positif, negatif, atau netral [1,2]. Dalam melakukan klasifikasi, terdapat beberapa metode *machine learning* yang bisa digunakan salah satunya adalah metode *Support Vector Machine* (SVM). Penelitian ini menggunakan seleksi fitur *chi square* dan TF-IDF sebagai ekstraksi fiturnya. Seleksi fitur *chi square* [3] bertujuan untuk menghilangkan fitur yang tidak diperlukan dalam klasifikasi. Sedangkan untuk ekstraksi fitur [4], TF-IDF adalah salah satu metode ekstraksi fitur yang paling populer dan banyak digunakan.

Pada penelitian sebelumnya [5] yang membahas tentang *naïve bayes* dan SVM untuk analisis sentimen review film, dijelaskan bahwa dengan menggunakan metode algoritma SVM memiliki tingkat akurasi mencapai 90.00%, sedangkan dengan metode algoritma *naïve bayes* memiliki tingkat akurasi yang lebih rendah dibandingkan dengan metode algoritma SVM yaitu dengan 84.50%. Penelitian selanjutnya [6] membahas tentang analisis sentimen pada review film dengan menggunakan algoritma klasifikasi NB berdasarkan Term Objects Keywords memiliki tingkat akurasi sebesar 28%. Pada penelitian selanjutnya [7] yang membahas sentimen analisis pada review film dengan mengekstraksi data menggunakan *Dov2Vec* dengan metode SVM menghasilkan nilai *f1-measure* sebesar 54,19%, dan pada penelitian [6,7] menggunakan *dataset* dari website berbahasa Inggris.

Pada penelitian ini akan menggunakan metode klasifikasi berbasis *support vector machine* dengan menerapkan seleksi fitur menggunakan metode *Chi-Square*. Karena, dapat dilihat dari studi literatur atau penelitian sebelumnya ketika menggunakan metode klasifikasi *Support Vector Machine* dengan menerapkan seleksi fitur *Chi-square* mendapat tingkat akurasi yang lebih baik dibanding tanpa menerapkan seleksi fitur *chi-square*. Sehingga dari hal tersebut penelitian ini akan menggunakan metode klasifikasi *support vector machine* dan menerapkan seleksi fitur berbasis *chi-square* yang di bandingkan dengan *support vector machine* tanpa menerapkan seleksi fitur *chi-square*.

Perumusan Masalah

Rumusan masalah yang akan diangkat berdasarkan permasalahan yang sudah dijelaskan pada latar belakang pada penelitian ini adalah:

- a. Bagaimana cara membangun sistem klasifikasi analisis sentimen menggunakan *feature selection chi square* dengan metode *Support Vector Machine*?
- b. Bagaimana performansi dari seleksi fitur yang akan dilakukan menggunakan *feature selection chi square* dan tanpa menggunakan *feature chi square* dalam klasifikasi *Support Vector Machine*.

Tujuan

Pada penelitian ini memiliki tujuan yang berdasarkan dari latar belakang dan rumusan masalah yang sudah dijelaskan. Tujuan pada penelitian ini adalah:

- a. Mengetahui cara membangun sistem klasifikasi analisis sentimen menggunakan *feature selection chi square* dengan metode *Support Vector Machine*.
- b. Menganalisis performansi tingkat akurasi yang akan dilakukan menggunakan *feature selection chi square* dan tanpa menggunakan *feature chi square* dalam klasifikasi *Support Vector Machine*.

Batasan Masalah

Sedangkan batasan masalah yang dipakai pada penelitian ini adalah:

- a. *Dataset* yang digunakan merupakan ulasan dari situs <http://movfreak.blogspot.com> menggunakan Bahasa Indonesia.
Klasifikasi analisis sentimen berupa kelas positif dan negatif