

## 1. Pendahuluan

Masyarakat Indonesia saat ini sangat suka menonton film, sehingga akhir-akhir ini produksi film di Indonesia semakin meningkat. Informasi berupa opini-opini yang berkaitan dengan film dapat ditemui pada situs Internet Movie Database (IMDb). Situs ini merupakan salah satu tempat yang menyediakan informasi terkait film di seluruh dunia dan memuat ulasan tentang film di setiap kategorinya. Ulasan yang terdapat didalamnya berisi pendapat-pendapat penonton yang beragam baik pendapat positif maupun negatif terkait film tersebut [1]. Dan ulasan film juga dapat dimanfaatkan atau digunakan oleh penikmat film untuk memutuskan film apa yang mau ditonton dan para produser film dapat memanfaatkan opini tersebut untuk mengetahui penilaian yang diberikan oleh para penikmat film tentang film-film yang paling banyak dikritik [2]. Untuk itu analisis sentimen mengenai ulasan film sangat menarik untuk dijadikan obyek penelitian.

Analisis sentimen yaitu suatu proses memahami, mengekstrak, dan mengolah data tekstual secara otomatis, atau merupakan studi komputasi pendapat, perasaan dan emosi yang dinyakan dalam bentuk teks. Ada beberapa algoritma atau metode yang di gunakan untuk analisis sentimen, antara lain *Naïve Bayes*(NB), *Support Vector Machine*(SVM), dan clustering K-Means [3].

Proses pelabelan manual yang dilakukan oleh manusia memang menghasilkan hasil yang akurat namun, akan menjadi kendala jika data yang harus diberi label berjumlah sangat banyak. Belum lagi banyaknya *fake reviews* yang membuat validasi semakin bertambah lama. Contoh saja pada data *review* Yelp.com, 75% *spammers* membuat 6 *review* atau bahkan dalam sehari. Dan dari 40 ribu data, ada sekitar 8 ribu *spammers*. Yang berarti 20% dari *reviewer* merupakan *spammers* [4]. Tentu ini akan memakan waktu lama ketika dilakukan secara manual. Oleh karena itu, pelabelan secara manual akan sangat tidak efektif mengingat kecepatan pertambahannya data. Sehingga akan lebih baik jika ada sistem untuk melabelkan data tersebut secara otomatis dengan memanfaatkan data pelabelan manual sebagai referensinya [5].

Berdasarkan penjelasan di atas, peneliti tertarik untuk memanfaatkan data yang ada dalam pembuatan sistem pelabelan otomatis yang dapat mengkategorikan *review* film berbahasa inggris ke dalam dua kategori sentimen yaitu positif dan negatif menggunakan metode *Support Vector Machines*. Untuk setiap data yang telah dilabelkan secara manual akan dijadikan train data sehingga memaksimalkan efektifitas proses pelabelan. Dengan pertimbangan karena pada beberapa penelitian menunjukkan efektifitas dari pelabelan otomatis sendiri [5][6][7]. Selain menghemat waktu, akurasi sendiri cukup baik untuk diimplementasikan. Bahkan pada salah satu penelitian, didapat tingkat akurasi hingga lebih dari 90% yang berarti tingkat kesalahannya bisa dikurangi seiring dengan bertambahnya dataset [8].

SVM dimanfaatkan dalam penelitian ini karena merupakan teknik paling umum digunakan untuk klasifikasi data [6]. Selain itu, menurut Bang et al dalam penelitiannya untuk membandingkan beberapa *classifier*, didapatkan tingkat akurasi SVM merupakan yang tertinggi dalam pengklasifikasian data *review* [7]. Selain itu, penelitian ini akan membandingkan hasil uji antar kernel SVM (linear, RBF dan polynomial). Ini karena dalam penelitian sebelumnya, belum ada yang melakukan perbandingan antara masing-masing kernel untuk data berupa ulasan. Paper-paper terkait sentimen analisis yang memanfaatkan metode SVM umumnya hanya menggunakan salah satu dari kernel tersebut. Sebagian besar menggunakan default (linear), dan ada beberapa yang memanfaatkan RBF juga. Harapan perbandingan ini nantinya dapat dijadikan referensi terkait tingkat akurasi dari masing-masing kernel. Penelitian ini nantinya hanya akan terbatas pada *movie review* dengan bahasa inggris yang menggunakan satuan rating 1-10 dan berfokus pada 3 kernel SVM (Linier, RBF, polynomial) dengan menggunakan TF-IDF sebagai *Extraction Feature*nya.

Tujuan yang ingin dicapai dalam penelitian ini yang pertama adalah mengetahui implementasi SVM dalam sistem pelabelan otomatis, yang kedua Mengetahui performa penggunaan pelabelan otomatis terhadap data ulasan film, dan yang ketiga Mengetahui akurasi dari setiap kernel yang ada pada SVM. Pada bab 2 dibahas mengenai studi terkait tentang analisis sentimen, tujuan dari analisis sentimen itu sendiri. Selain itu juga menjelaskan tentang *review* film, perbedaan *review* film dengan *review* lainnya, penelitian serupa, baik dalam hal algoritma, bentuk atau pun jenis data, *extraction feature*, dan proses pelabelannya. Bab 3 membahas teori dan sistem penelitian, bab 4 membahas evaluasi hasil pengujian, dan bab 5 membahas kesimpulan dari penelitian yang telah dilakukan.