

BAB I PENDAHULUAN

I.1 Latar Belakang

Film merupakan sebuah bentuk komunikasi massa elektronik yang berupa media pertunjukkan audio visual yang mampu menampilkan kata-kata, bunyi, citra, dan kombinasinya (Sobur, 2004). Menurut prof. Effendy, *film* merupakan medium komunikasi massa yang bukan hanya untuk hiburan saja, tetapi juga untuk penerangan dan pendidikan. Menurutnya, *film* mempunyai suatu dampak bagi penonton, dampak-dampak tersebut dapat berbagai macam seperti dampak psikologis dan dampak sosial (Effendy, 2003). Salah satu contoh dampak dari psikologis yang diterima oleh masyarakat yaitu berupa sudut pandang dan perspektif terhadap penilaian sebuah *film* yang diberikan oleh masyarakat berupa ulasan. Ulasan yang diberikan pada sebuah *film* tentunya dapat berupa ulasan positif dan ulasan negatif. Didukung dengan perkembangan internet yang sangat pesat menjadi media sarana untuk masyarakat dalam menyalurkan pendapat dan penilaian terhadap *film* itu sendiri. Platform informatif yang berkaitan dengan *film* yang cukup populer adalah *Internet Movie Database* (IMDb). Platform ini merupakan sebuah situs *website* basis data informasi yang berkaitan dengan *film*, acara televisi, video rumahan, dan permainan video. Platform ini juga menyediakan informasi top *film* paling populer berdasarkan hasil rating dan *review* yang diberikan oleh masyarakat.

Berbicara tentang data, data yang akan menjadi objek penelitian kali ini yaitu data 50.000 *movie reviews* yang dirangkum dalam *website* IMDb. Data yang dikelola oleh lakshmi pathi n pada tahun 2019 ini merupakan data yang diolah untuk natural process language atau teks analisis. Karakteristik data ini ditujukan untuk dilakukan proses klasifikasi sentimen biner yang berisikan jumlah data yang sangat besar. Kondisi data dengan jumlah cukup besar serta pemilihan algoritma yang tidak tepat memungkinkan memberi hambatan dalam proses kinerja model dalam melakukan klasifikasi. Untuk itulah perlu dilakukan proses reduksi data. *Feature importance* merupakan cara yang efektif untuk melakukan reduksi data dan menjadi langkah penting yang perlu dilakukan supaya aplikasi *data mining* berhasil dengan baik (Liu, Huan dkk. 2010)

Feature importance merupakan proses pemilihan subset dari fitur/atribut yang

optimal dengan menggunakan kriteria tertentu. *Feature importance* merupakan salah satu dari proses *pre-processing* pada suatu *dataset* yang akan dilakukan proses *data mining*. Dengan melakukan *feature importance* ini mampu untuk mengurangi jumlah *feature* yang tidak relevan, menghilangkan redundansi data, menghilangkan *feature* yang mengandung *noisy* dan akan memberikan efek meningkatkan kecepatan dalam melakukan *data mining*, meningkatkan akurasi *learning*, dan menghasilkan model yang baik (Liu, Huan dkk. 2010). Hal ini, sangat mendukung proses *data mining* pada *dataset* dengan jumlah yang besar seperti *IMDB movie review dataset*. pada penelitian ini akan dilakukan proses *text mining* salah satunya menggunakan algoritma SVM dan pemilihan fitur pada analisis sentimen *IMDB dataset*.

Penelitian yang dilakukan ini bertujuan untuk mengklasifikasi opini positif dan negatif menggunakan SVM. SVM jenis *classifier* atau *supervised learning* merupakan algoritma yang akan digunakan untuk memproses data tersebut. Data yang diperoleh akan diproses melalui *text mining*, hal yang dilakukan pertama yaitu dengan mengklasifikasikan komentar *IMDB* menjadi dua kelas, yaitu positif dan negatif. Klasifikasi yang dilakukan melalui proses *labeling* menggunakan annotator yang kemudian menerapkan algoritma *Support Vector Machine* (SVM) dalam proses *text mining*. Klasifikasi berfungsi memberikan kemudahan kepada pengguna untuk mengetahui opini positif atau negatif. Nilai akurasi yang dihasilkan dengan menggunakan algoritma akan memberikan pengaruh pada hasil klasifikasi. Selain itu, untuk memaksimalkan hasil digunakannya *feature importance* untuk membandingkan hasil akhir pengolahan data dengan menggunakan *feature importance*.

Algoritma SVM menghasilkan nilai akurasi tinggi seperti menguji sentimen terhadap wacana politik pada media sosial *online*, analisis sebuah sentimen komentar mahasiswa pada sistem pembelajaran di perguruan tinggi, analisis pada *tweets* di *Twitter* yang mengeluarkan opini tentang produk mobil otomatis dan produk Apple, dan rata-rata tingkat akurasi yang didapatkan adalah 50 %-90 % (Hidayat A.N, 2015).

Maka dari itu, pada tugas akhir ini penulis mencoba menggunakan Algoritma SVM dan *feature importance* untuk pengolahan data dengan topik permasalahan yang akan diangkat yaitu mengenai analisis sentimen *IMDb dataset* dengan objek

data *movie reviews*. Hasil ulasan *film* yang diberikan oleh masyarakat terhadap suatu *film* ini sangat menarik untuk dianalisis terkait berbagai opini dan fenomena. Penelitian ini bermaksud untuk menganalisis kecenderungan penilaian terhadap suatu topik dalam hal ini adalah mengenai *film* pada *platform* IMDB serta menganalisis nilai akurasi yang dihasilkan untuk mengukur kinerja model.

I.2 Perumusan Masalah

Rumusan masalah yang mendasari penelitian ini adalah:

- a. Bagaimana analisis sentimen berdasarkan ulasan yang diberikan oleh masyarakat atau penonton terhadap *film-film* yang ada pada *website* IMDB ?
- b. Bagaimana klasifikasi ulasan *film* berdasarkan ulasan yang ada pada IMDB *review* dengan proses seleksi *Feature Importance* ?
- c. Bagaimana performansi model dan akurasi yang dihasilkan berdasarkan proses seleksi *Feature importance* ?
- d. Bagaimana kecenderungan kata dan opini yang dihasilkan berdasarkan hasil proses *Text mining* ?

I.3 Tujuan Penelitian

Penelitian ini bertujuan untuk:

- a. Menganalisis sentimen berdasarkan ulasan yang diberikan oleh masyarakat atau penonton terhadap *film-film* yang ada pada *website* IMDB
- b. Menganalisis klasifikasi ulasan *film* berdasarkan ulasan yang ada pada IMDB *review* dengan proses seleksi *Feature Importance*
- c. Menganalisis performansi model dan akurasi yang dihasilkan berdasarkan proses seleksi *Feature importance*
- d. Menganalisis kecenderungan kata dan opini yang dihasilkan berdasarkan hasil proses *Text mining*

I.4 Batasan Penelitian

Adapun batasan dari penelitian ini yaitu mencakup pembahasan bagaimana implementasi proses *text mining* menggunakan algoritma *Support Vector Machine* (SVM) terhadap objek *dataset* IMDB *dataset* yang didukung oleh *feature*

importance untuk memperoleh *output* evaluasi dan informasi data yang dihasilkan. Dalam implementasi tugas akhir ini akan dibatasi oleh beberapa hal yaitu:

1. *Dataset* yang digunakan merupakan data *review rating film* yang diambil dari *Kaggle open source data*.
2. Proses implementasi *text mining dataset* menggunakan algoritma SVM dan *Feature importance*.
3. Menganalisis ketepatan prediksi klasifikasi opini yang dilihat dari performa model terhadap nilai akurasi yang dihasilkan.

I.5 Manfaat Penelitian

Manfaat penelitian ini:

1. Bagi peneliti dapat mengimplementasikan ilmu dan teori yang telah diperoleh pada masa perkuliahan baik materi dalam perkuliahan maupun materi dari luar perkuliahan.
2. Bagi peneliti lain yang bergerak dalam sistem informasi pendidikan tinggi, penelitian ini bermanfaat dalam menjelaskan pendekatan dan kaskas yang paling tepat dalam memberikan upaya klasifikasi *sentimen* secara baik dalam objek apapun dan salah satu kasusnya ada pada algoritma *Support Vector Machine* (SVM).