

1. Pendahuluan

Latar Belakang

Review tentang film merupakan kebutuhan bagi semua orang untuk mendapatkan informasi mengenai sebuah film sehingga dapat digunakan untuk membantu mendapatkan informasi tentang isi film yang akan ditonton. Informasi yang bisa didapat melalui sebuah review film adalah mengenai jalan cerita, aktor sampai dengan konflik yang terjadi di dalamnya serta kelebihan dan kekurangan sebuah film. Informasi-informasi hasil review yang dibuat kemudian digunakan sebagai bahan pertimbangan dalam menentukan kualitas dari sebuah film sehingga pecinta film dapat mengetahui sejauh mana film tersebut layak atau tidak layak di tonton. Penilaian mengenai isi dari review film dapat disebut dengan sentiment analysis. Sentiment analysis adalah proses penerapan natural language processing (NLP) dan analisis teks untuk mengidentifikasi dan melakukan ekstrak informasi subjektif dari sebuah teks [1]. Sentiment analysis dapat diaplikasikan menggunakan sebuah metode klasifikasi untuk mempermudah dalam pengelompokan data berupa data positif atau data negatif yaitu dengan menggunakan metode Naïve Bayes. Metode Naïve Bayes digunakan pada proses klasifikasi dalam sebuah penelitian karena bekerja dengan cepat dan efisien sebagai metode pengklasifikasian teks. Penerapan klasifikasi sentiment analysis menjadi kalimat positif maupun negatif dapat dilakukan setelah pemangkasan pada data subjek yang digunakan untuk mengurangi fitur sehingga menghindari banyaknya dimensi yang digunakan pada saat proses klasifikasi [2]. Review film dapat mempunyai ukuran dataset yang cukup besar baik itu pada data training maupun data testing. Dimensi dan fitur yang berlebihan akan meningkatkan ruang pencarian semakin tinggi sehingga akan menyebabkan kesulitan dalam memproses data dan akan menurunkan kinerja serta membuat data tidak konsisten. Analisis dan mining dalam data juga membutuhkan waktu yang lama dalam pemrosesan data. Pengurangan dimensi dapat diterapkan untuk mengurangi dimensi dari data, dimana nantinya akan meningkatkan kinerja dari teknik machine learning dengan menghilangkan fitur yang tidak perlu digunakan. Penyelesaian dalam permasalahan penelitian ini menggunakan metode Tf-Idf yang akan mengukur pembobotan berbasis kata sebagai fitur hasil ringkasan

Ada beberapa algoritma klasifikasi yang banyak digunakan untuk analisis review sentimen antara lain Support Vector Machine, Naïve Bayes, dan K-Nearest Neighbor [3]. Beberapa penelitian yang telah dilakukan dalam klasifikasi sentimen ulasan online diantaranya, Comparative machine learning untuk klasifikasi sentimen analisis ulasan film [4]. Analisis sentimen pada opini review film menggunakan algoritma NB, KNN, dan Random Forest [5]. Analisis sentimen pada opini review film menggunakan Features and Opinion Words Extraction [6]. Analisis sentimen pada opini review film menggunakan algoritma NB, KNN, dan Decision Tree [7]. Analisis sentimen pada opini review film menggunakan algoritma NB dan fitur Gini Index [8]. Analisis sentimen pada opini review film menggunakan algoritma NB dan fitur Chi Square [9]. Analisis sentimen pada You tube movie trailer comments menggunakan algoritma NB dan fitur Tf-Idf [10].

Topik dan Batasannya

Tugas akhir ini difokuskan pada analisis sentimen dari review film yang berpotensi menghasilkan review positif dan review negatif. Banyak situs memberikan review suatu produk yang dapat mencerminkan opini pengguna. Salah satu contohnya adalah situs Internet Movie Database (IMDb). IMDb adalah situs web yang berhubungan dengan produksi film dan film. Penelitian ini mengusulkan untuk menggunakan metode Naïve Bayes. Pemilihan metode didasarkan pada beberapa penelitian terdahulu sebagai referensi.

Pada tugas akhir ini, topik yang akan dibahas adalah bagaimana performansi metode Naïve Bayes dan Fitur Tf-Idf dalam analisis sentimen dari review film yang berasal dari data IMDb. Batasan masalah pada tugas akhir ini adalah sebagai berikut: Pertama, jumlah sample atau data sebanyak 50000. Kedua, data yang berupa Bahasa Inggris. Ketiga, penelitian analisis sentimen menggunakan metode Naïve Bayes Classifier dan fitur Tf-Idf sebagai pengklasifikasi data review film.

Tujuan

Tujuan dari penelitian ini adalah untuk mengidentifikasi akurasi dari Naïve Bayes dengan adanya indikasi review positif dan review negatif pada review film IMDb. Dengan hasil Naïve Bayes dibantu dengan fitur Tf-Idf ini diharapkan dapat membantu bagi semua orang untuk mendapatkan informasi mengenai sebuah film sehingga dapat digunakan untuk membantu mendapatkan informasi tentang isi film yang akan ditonton.

Organisasi Tulisan

Struktur penulisan dari tugas akhir ini disusun sebagai berikut: Bagian pertama berisi pendahuluan terkait tugas akhir ini. Bagian kedua menjelaskan studi yang terkait dengan tugas akhir ini. Bagian ketiga akan menjelaskan pemodelan dan performansi dari sistem yang dibangun. Bagian keempat menjelaskan hasil dan evaluasi hasil pengujian yang telah dilakukan pada bagian ketiga. Kemudian, pada bagian terakhir menjelaskan kesimpulan dan saran berdasarkan hasil pengujian yang dilakukan pada tugas akhir ini.