

1. Pendahuluan

Latar Belakang

Di zaman informasi saat ini, internet telah menjadi sarana utama bagi orang-orang untuk memberikan opinimereka, termasuk ulasan tentang film, lewat media sosial. Sebagai bentuk seni visual yang dinamis, film-film mendapat ulasan yang memungkinkan para penonton untuk memberikan penilaian yang detail terhadap suatu karya. Kenaikan jumlah produksi film berbanding lurus dengan bertambahnya jumlah ulasan, dan berbagai pandangan yang diungkapkan dalam ulasan-ulasan tersebut menghadirkan gambaran nyata lewat penggunaan bahasa yang penuh emosi. Karena itu, menjadi penting bagi para penonton untuk secara serius membaca berbagaiulasan untuk memperoleh pemahaman yang luas tentang film tersebut [1].

MyAnimeList atau sering sering juga disingkat MAL, merupakan sebuah situs jejaring sosial dan situs katalogisasi sosial yang menyangkut anime dan manga. Pada *Website MyAnimelist* terdapat anime maupun mangadan dengan lebih dari puluhan ribuan ulasan yang ada. Maka dari itu, dengan banyaknya data yang ada pada *MyAnimeList* dibutuhkan sebuah teknik pengolahan terhadap data ulasan tersebut sehingga akan didapatkan informasi yang berharga [2]. Analisis sentimen merupakan salah satu cara pengolahan data tersebut, dengan melakukan klasifikasi pada kelas negatif dan positif[3]. Analisis sentiment juga kini menjadi salah satu topik penelitian yang populer dalam bidang pemrosesan bahasa alami. Saat ini, platform media sosial seperti Twitter, Facebook, dan YouTube berfungsi sebagai sumber informasi yang dikenal sebagai data social [19].

Penelitian mengenai analisis sentimen sudah sering dilakukan sebelumnya dengan berbagai macam teknik yang berbeda baik pada bagian ekstraksi fitur ataupun metode klasifikasi yang digunakan. Penelitian yang dilakukan pada tahun 2021 mendapatkan hasil akurasi sebesar 82.66% untuk SVM dan 79.33% untuk *Naive Bayes*[4]. Penelitian ini membandingkan metode pembobotan *Term Frequency-Inverse Document Frequency (TF-IDF)* dan *Singular Value Decomposition (SVD)* menggunakan algoritma *Multinomial Naive Bayes*, *Multivariate Bernoulli Naive Bayes*, dan *Support Vector Machine* untuk mengklasifikasikan data dari artikel berita berbahasa Indonesia. Hasil percobaan menunjukkan bahwa kombinasi TF-IDF dengan *Multinomial Naive Bayes* menghasilkan nilai *precision* dan *recall* yang tinggi, yaitu 98,4%, dibandingkan dengan kombinasi TF-IDF dan *Multivariate Bernoulli Naive Bayes* yang mencapai 98,2%. Percobaan dengan SVD tidak memberikan hasil yang baik dan tidak memengaruhi klasifikasi [20]. Lalu penelitian pada tahun 2021 menggunakan model klasifikasi *random forest* dengan ekstraksi fitur TF-IDF menghasilkan akurasi sebesar 85.32% dan F1-Score sebesar 67.35% [5]. Lalu nelitian yang dilakukan oleh E. F. Saraswita, dkk (2021) dengan menggunakan Support Vector Machine(SVM) dan Recursive Feature Elimination (RFE) pada analisis sentimen e-wallet yang mendapatkan akurasi sebesar 81% dan akurasi CART mencapai 96% sedangkan akurasi dari model yang lain sebesar 92% [21].

Berdasarkan pada hasil penelitian [4],[5],[20] dan [21] untuk mengetahui sentimen dari ulasan anime berbahasa inggris dari situs *MyAnimeList* (MAL) dengan menggunakan algoritma yang berfokus pada ekstraksi fitur dari ulasan dengan memperhatikan bobot dari tiap kata ulasan dengan metode TF-IDF, selain itu dilakukan klasifikasi dengan metode klasifikasi CART.

Pemilihan penggunaan metode Decision Tree yang berbentuk CART ini karena metode yang kuat dan serbaguna untuk klasifikasi dan regresi, terutama ketika interpretabilitas dan fleksibilitas adalah prioritas.

Topik dan Batasannya

Topik pada penelitian kali ini membahas tentang pengaruh *preprocessing*, *stemming* dan *hyperparameter tuning* pada data ulasan *MyAnimeList*. Adapun batasan yang ada pada penelitian kali ini dataset yang berasal dari situs *MyAnimeList*. Dataset yang diambil berjumlah 1000 ulasan yang sudah dilabeli menjadi positif dan negatif dengan semua ulasan berbahasa Inggris.

Tujuan

Penelitian ini dilakukan dengan tujuan melakukan analisis sentimen pada ulasan pada *MyAnimeList* dengan pengaruh dari proses *preprocessing*, *stemming* dan *hyperparameter tuning*, dengan penggunaan ekstraksi fitur TF-IDF dan klasifikasi *Decision Tree* CART.