

## 1. Pendahuluan

### Latar Belakang

Menonton film merupakan suatu aktivitas yang dilakukan untuk mengisi rasa bosan atau menghilangkan stres yang dilakukan oleh seorang individu atau sekelompok orang. Ulasan mengenai suatu film sangat penting dikarenakan informasi tersebut dapat dijadikan referensi untuk menonton suatu film. Dikarenakan data mengenai ulasan film berjumlah sangat banyak, memerlukan banyak waktu untuk membaca ulasan tersebut satu per satu [1]. Berdasarkan kondisi tersebut, analisis sentimen dalam ulasan film merupakan topik yang menarik untuk diselesaikan dengan menggunakan machine learning, karena secara otomatis akan melakukan proses klasifikasi dan mempersingkat waktu.

Terdapat berbagai metode machine learning yang digunakan untuk melakukan proses klasifikasi, salah satunya menggunakan metode *K-Nearest Neighbor* (KNN). Metode KNN memiliki beberapa keunggulan jika dibandingkan dengan metode lainnya seperti efektif terhadap data latih yang ukurannya besar dan memiliki banyak *noise*, tetapi metode KNN memiliki performa hasil klasifikasi yang kurang baik apabila fitur yang digunakan tidak ditentukan dengan baik [2]. Untuk menentukan fitur yang akan digunakan, penggunaan seleksi fitur dan ekstraksi fitur perlu dilakukan.

Dalam penelitian ini, metode *machine learning* yang digunakan adalah KNN dengan menggunakan seleksi fitur *chi-square* dan ekstraksi fitur *Term Frequency-Inverse Document Frequency* (TF-IDF). Berdasarkan penelitian yang dilakukan oleh Luqman Rizky Dharmawan [3], penggunaan seleksi fitur *chi-square* dan metode KNN pada dataset yang diambil dari twitter tentang layanan Sistem Informasi Akademik Mahasiswa Universitas Brawijaya (SIAM UB) menghasilkan nilai akurasi sebesar 86%. Untuk ekstraksi fitur, TF-IDF dipilih dikarenakan mampu mengasosiasikan setiap kata dalam dokumen dalam bentuk angka yang mewakili seberapa relevan setiap kata tersebut dalam suatu dokumen [4].

### Topik dan Batasannya

Topik pada penelitian tugas akhir ini adalah melakukan pengujian kombinasi antara seleksi fitur *chi-square*, ekstraksi fitur TF-IDF, dan metode KNN terhadap *dataset* ulasan film. Untuk batasannya, *dataset* yang digunakan merupakan ulasan film dalam Bahasa Inggris dari situs IMDb dengan jumlah data yang digunakan sebanyak 2001 data, terdiri dari 1000 data ulasan positif dan 1001 data ulasan negatif.

### Tujuan

Tujuan yang ingin dicapai pada penelitian tugas akhir ini adalah menganalisis hasil performansi kombinasi ekstraksi fitur TF-IDF dan metode KNN terhadap penggunaan stemming, *threshold* seleksi fitur *chi-square* yang digunakan, dan pengaruh perubahan nilai K metode KNN pada *dataset* ulasan film dalam Bahasa Inggris.

### Organisasi Tulisan

Pada bab 2 dibahas mengenai studi terkait penelitian yang dilakukan, bab 3 membahas teori dan perancangan sistem penelitian, bab 4 membahas evaluasi model, dan bab 5 membahas kesimpulan dari penelitian yang digunakan.