

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan industri bidang perfilman Indonesia telah mengalami pertumbuhan yang pesat dalam beberapa tahun terakhir. Data dari Badan Perfilman Indonesia (BPI) mencatat bahwa jumlah film nasional yang diproduksi mengalami peningkatan dari 36 film pada tahun 2021 menjadi 129 pada tahun 2023^[1]. Peningkatan ini tidak hanya terjadi pada bioskop konvensional, tetapi juga didukung oleh pesatnya perkembangan platform *streaming* seperti Netflix, Vidio, dan Youtube yang dapat diakses kapan saja. Hal ini menjadi tantangan untuk mengorganisir dan menemukan film sesuai dengan preferensi genre bagi penonton. Genre film sendiri berfungsi sebagai alat kategorisasi yang menjadi penanda konten dan komponen kunci dalam sistem rekomendasi.

Proses klasifikasi genre film secara tradisional sering kali dilakukan secara manual lewat para kritikus film, penyedia platform, maupun konten moderator. Yang membuat metode ini rentan akan subjektivitas dan ketidakkonsistenan. Sebuah sinopsis yang kompleks dapat ditafsirkan secara berbeda oleh setiap individu, berakibat adanya perbedaan dalam pemberian label genre pada film. Permasalahan semakin bertambah dengan bertambahnya jumlah film baru yang terus bertambah, membuat klasifikasi manual menjadi tidak efisien dan tidak dapat diskalakan. Selain itu, dominasi genre tertentu (seperti horor) dalam produksi film Indonesia menambah kompleksitas kebutuhan untuk sistem klasifikasi yang akurat dan konsisten.

Berdasarkan permasalahan tersebut, salah satu solusi yang dapat dilakukan adalah menggunakan *machine learning*. *Machine learning* merupakan salah satu disiplin ilmu yang biasanya dilakukan untuk pemrosesan komputasi secara otomatis. Salah satu metode yang dapat dimanfaatkan dari *machine learning* adalah klasifikasi. Klasifikasi sendiri adalah metode yang mempelajari pola-pola dari

training data untuk memprediksikan objek yang baru ke masing-masing kelasnya^[2] yang salah satu teknikny adalah Algoritma *Naive Bayes Classifier (NBC)*.

Algoritma *Naive Bayes Classifier (NBC)* dipilih sebagai solusi karena kemampuannya dalam menganalisis pola teks sinopsis secara cepat, konsisten, dan akurat, sekaligus mengatasi keterbatasan klasifikasi manual. Melalui penelitian ini, diharapkan tercipta sistem klasifikasi otomatis yang dapat berkontribusi terhadap pengembangan database film, peningkatan akurasi sistem rekomendasi pada platform streaming, serta memudahkan penonton dalam menemukan konten sesuai preferensi genre mereka.

1.2 Rumusan Masalah

Rumusan masalah dari penulisan skripsi ini adalah :

1. Bagaimana mengklasifikasikan genre pada film berdasarkan teks sinopsis?
2. Bagaimana performa algoritma Naive Bayes dalam mengklasifikasikan genre film berdasarkan teks sinopsis?
3. Faktor apa yang berpengaruh terhadap akurasi klasifikasi genre film menggunakan Naive Bayes?

1.3 Batasan Masalah

Guna memperlancar penulisan skripsi ini, maka diperlukan batasan masalah sebagai berikut:

1. Genre film yang dianalisa dibatasi pada genre primer dan memiliki 50 sampel data.
2. Sinopsis dalam bentuk teks dan Bahasa Inggris.
3. Menggunakan Bahasa pemrograman Python.
4. Metode klasifikasi terbatas pada algoritma Naive Bayes

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dijabarkan, tujuan dari skripsi ini adalah:

1. Dapat mengategorikan genre berdasarkan sinopsis.

2. Mengalisis performa algoritma Naive Bayes Classifier dalam mengklasifikasikan genre film berdasarkan sinopsis.
3. Mengidentifikasi faktor-faktor yang mempengaruhi hasil klasifikasi.

1.5 Manfaat penelitian

Berdasarkan penelitian yang dilakukan, manfaat yang dapat diperoleh adalah sebagai berikut:

1. Diharapkan dapat memberikan kontribusi dalam pengembangan metode *Natural Language Processing*.
2. Dapat mengevaluasi Algoritma *Naive Bayes* dan menjadi referensi mengenai performa maupun karakteristik algoritma *Naive Bayes* dalam menangani data teks genre film.
3. Memberikan insight tentang kata-kata kunci yang menjadi pembeda antar genre film berdasarkan analisis bobot TF-IDF.
4. Dapat mengembangkan pendekatan *computational* untuk analisis konten film secara objektif

1.6 Sistematika Penulisan

Pembuatan skripsi ini dilakukan dengan sistematika penulisan sebagai berikut:

1. Bagian awal
Memuat komponen administratif yang meliputi: sampul depan, halaman judul, halaman persetujuan pembimbing, halaman pengesahan, halaman pernyataan orisinalitas, kata pengantar, daftar isi, daftar tabel, daftar gambar, daftar lampiran, intisari dalam Bahasa Indonesia, dan abstract dalam Bahasa Inggris.
2. Bab I Pendahuluan
Memuat latar belakang, rumusan masalah, tujuan penelitian, Batasan masalah, manfaat penelitian dan keaslian penelitian dengan menunjukkan parameter- parameter penelitian yang terukur. Pendahuluan ditulis dengan bahasa yang jelas dan mudah dipahami dan dikenali oleh pembaca.
3. Bab II Tinjauan Pustaka

Bab ini terdiri atas studi literatur yang menganalisis penelitian terdahulu dan keaslian penelitian, serta dasar teori yang meliputi konsep genre film, pemrosesan bahasa alami, algoritma *Naïve Bayes*, *TF-IDF*, dan evaluasi model *machine learning*.

4. Bab III Metode Penelitian

Bab ini menjelaskan metodologi penelitian yang mencakup objek penelitian (dataset film), alur penelitian dari *preprocessing* hingga evaluasi, serta alat dan bahan yang digunakan dalam implementasi sistem.

5. Bab IV Hasil dan Pembahasan

Bab ini menyajikan hasil eksperimen klasifikasi genre film, analisis performa model *Naïve Bayes* berdasarkan metrik evaluasi, serta pembahasan terhadap faktor-faktor yang memengaruhi hasil klasifikasi.

6. Bab V Penutup

Bab terakhir berisi kesimpulan dari temuan penelitian dan saran konstruktif untuk pengembangan penelitian selanjutnya.

7. Bagian Akhir

Bagian akhir memuat daftar pustaka yang berisi referensi yang digunakan dan lampiran-lampiran pendukung penelitian.