

# BAB 1

## PENDAHULUAN

### 1.1 Latar belakang

Film merupakan salah satu media hiburan yang memiliki pengaruh besar dan merupakan sarana untuk menyampaikan pesan penting kepada masyarakat atau penonton dengan cara yang menghibur dan menyenangkan[1]. Kini banyak tersedia platform gratis hingga yang berbayar yaitu salah satunya adalah *platform* netflix yang memimpin konsumsi video premium[2]. Netflix banyak menyediakan *genre* mulai dari genre romantis, *action*, horor, komedi, *thriller* dan banyak lagi yang bisa dinikmati dimana saja. Akses untuk film ataupun tv series semakin banyak bervariasi yang tidak melulu korea drama dan hollywood, kita sekarang juga bisa menonton film dari negara manapun yang pastinya memiliki ciri khas tertentu. Netflix mempunyai 200 juta pengguna dan tersebar di seluruh dunia[3]. Untuk membantu pengguna menemukan film dan acara TV yang sesuai dengan usia atau preferensi pengguna, Netflix menggunakan metode klasifikasi teks menggunakan algoritma *Naive Bayes*.

*Naive bayes Classifier* termasuk algoritma yang mudah di implementasi, cepat dan akurat. Data yang digunakan adalah "Netflix Tv Shows and Movies" yaitu data film yang berisi informasi judul, aktor, genre, umur, negara produksi dan lain-lain yang belum terstruktur yang diambil dari Kaggle Victor Soeiro (2022). Metode Naive Bayes ini adalah salah satu algoritma klasifikasi teks yang paling umum digunakan. Algoritma ini didasarkan pada teorema *Bayes*, yang menyatakan bahwa probabilitas suatu kejadian bergantung pada probabilitas kejadian yang terkait dengannya. Ada banyak metode *Naive Bayes* yang digunakan dalam *text mining*, salah satu kelebihan *Naive Bayes* adalah algoritmanya sederhana, namun memiliki nilai akurasi yang cukup tinggi[4]. Dalam konteks klasifikasi teks, algoritma *Naive Bayes* digunakan untuk memprediksi label kelas yang sesuai untuk sebuah dokumen teks berdasarkan kemunculan kata tertentu dalam dokumen tersebut.

Pada penelitian ini penerapan metode *Naive Bayes* dalam mengklasifikasikan film dan acara TV di Netflix dapat membantu lembaga sensor film (LSF) untuk bertanggung jawab terhadap konten yang mereka produksi dan juga dapat mempertimbangkan efek dari konten

yang mereka produksi serta dampaknya terhadap penonton. Dalam penelitian ini dipilih atribut yang berpengaruh terhadap proses klasifikasi, atribut-atributnya adalah judul, deskripsi, genre, *type*, dan *age certification*. Di *dataset* Netflix dan Tv series ini, setiap film atau program televisi diberi label kategori yang sesuai dengan genre atau jenis konten tertentu, seperti Drama, Komedi, Horor atau Dokumenter. Data ini kemudian diproses menggunakan algoritma *Naive Bayes* untuk memprediksi label kelas yang sesuai untuk setiap dokumen teks (deskripsi film atau program TV). Untuk mengetahui tingkat akurasi hasil klasifikasi teks menggunakan algoritma *Naive Bayes*, dilakukan evaluasi kinerja algoritma tersebut dengan model evaluasi *classification report*. Ini dapat dilakukan dengan membandingkan hasil klasifikasi yang dihasilkan oleh algoritma dengan label kategori sebenarnya dari setiap dokumen teks. Informasi akurat tentang performa algoritma *Naive Bayes* dalam mengklasifikasikan film dan acara TV Netflix dapat diperoleh menggunakan metrik evaluasi seperti presisi, akurasi, *recall*, dan *F1-score*.

### **1.2 Rumusan masalah**

1. Bagaimana penerapan metode *naive bayes* dalam pengklasifikasian data film dan tv series di netflix?
2. Berapa nilai akurasi yang dihasilkan dari pengklasifikasian teks menggunakan algoritma *naive bayes*?

### **1.3 Batasan masalah**

Adapun Batasan masalah pada penelitian ini adalah:

- a) *Dataset* yang digunakan dalam penelitian ini adalah Netflix TV Shows and Movies dari website *kaggle*.
- b) Menggunakan *tools google colab*.
- c) Menggunakan algoritma *Naive Bayes*

#### 1.4 Tujuan Penelitian

Tujuan yang dicapai dari penelitian ini adalah

- a) Mengetahui hasil penerapan metode *naive bayes* dalam pengklasifikasian data film dan tv series di netflix
- b) Mengetahui nilai akurasi dari hasil pengklasifikasian teks menggunakan algoritma *naive bayes*.

#### 1.5 Manfaat penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

- 1 Penelitian ini dapat menjadi acuan atau bahan referensi untuk penelitian selanjutnya. Dengan adanya penelitian yang dilakukan, peneliti lain dapat melakukan penelitian lebih lanjut mengingat faktor-faktor yang tidak dipertimbangkan dalam penelitian ini. Hasil penelitian selanjutnya diharapkan dapat memberikan manfaat yang lebih besar lagi.
- 2 Penelitian ini dapat meningkatkan pemahaman masyarakat umum tentang metode klasifikasi teks dan algoritma *naive bayes*. Pembaca dapat memahami bagaimana algoritma *Naive Bayes* digunakan untuk memproses teks dan membuat klasifikasi yang akurat. Pemahaman yang ditingkatkan ini dapat membantu masyarakat membuat keputusan dalam kehidupan sehari-hari, terutama dalam hal mengklasifikasikan teks.

#### 1.6 Sistematika Penulisan

##### BAB I Pendahuluan

Bab ini merupakan pendahuluan yang meliputi latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

##### BAB II Tinjauan Pustaka

Bab ini berisi mengenai landasan teori yang mendukung dengan penelitian yang akan dilakukan. Sumber yang berisi referensi yang relevan dan asli.

### BAB III Metode Penelitian

Bab ini berisi tentang uraian langkah-langkah penelitian yang sistematis yang akan dilakukan dalam penelitian ini.

### BAB IV Hasil dan Pembahasan

Bab ini akan dijelaskan proses dan tahapan implementasi dari model klasifikasi yang dibuat.

### BAB V Penutup

Bab ini akan dijelaskan kesimpulan dari hasil pengujian yang telah dilakukan dan saran untuk penelitian selanjutnya.

