

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Drama Korea merupakan salah satu media penyebaran *Korean wave* dan memiliki jumlah penggemar atau penonton yang banyak di seluruh dunia. Salah satu faktor penentu dalam memilih sebuah drama adalah *rating*. *Rating* tersebut merupakan persentase penonton yang mengukur kinerja sebuah stasiun TV. Setiap stasiun TV berlomba-lomba memproduksi drama yang berbeda untuk mendapatkan performa *rating* yang lebih tinggi dibandingkan TV lainnya. Bagi tim produksi, *rating* merupakan tolak ukur popularitas dari drama yang diproduksi. Perhitungan *rating* dilakukan untuk setiap episode yang tayang. *Rating* yang tinggi atau rendah akan mempengaruhi proses produksi [1].

Di Korea, jarang program televisi diproduksi sepenuhnya terlebih dahulu. Sebagian besar program menjalani pengeditan akhir berdasarkan episode sekitar seminggu sebelum tayang. Saat ini, sekitar \$ 200.000 hingga \$ 500.000 per episode untuk menghasilkan drama serial mini di Korea. Satu judul biasanya terdiri dari 16 hingga 20 episode. Dengan demikian, \$ 3 juta hingga \$ 9,3 juta dari total investasi diperlukan dalam memproduksi satu judul drama, yang sangat besar uang untuk menyiarkan komunitas produksi. Sejauh ini, organisasi penyiaran, perusahaan produksi, penyedia komoditas, dan pengiklan hanya menggunakan *rating* sebagai alat pengukuran kuantitatif untuk mengukur laba atas investasi mereka. Oleh karena itu, tim produksi harus bersemangat menarik perhatian pemirsa untuk

meningkatkan *rating* dramanya guna memperoleh lebih banyak pendapatan iklan [2].

Drama Korea sendiri menjadi semakin populer di seluruh Asia (termasuk Indonesia) dan berkontribusi pada fenomena *Hallyu*. Drama TV Korea pertama kali masuk ke Indonesia pada tahun 2002 dan menjadi terkenal pada tahun 2009 melalui tayangan *Boys Before Flower* (BBF) di TV Indonesia yakni di stasiun televisi Indosiar [3],[4]. Karena perkembangan teknologi yang pesat saat ini, drama Korea juga dapat disaksikan secara *online* selain di stasiun TV. Tidak hanya bisa menonton drama saja, namun juga terdapat berbagai informasi mengenai drama Korea, baik drama yang sudah tayang maupun drama yang akan tayang. Termasuk *review* penulis yang pernah menonton drama Korea. Sehingga *review* tersebut dapat merekomendasikan pecinta drama Korea untuk menonton beberapa drama Korea. Penghitungan *rating* program televisi di Korea Selatan saat ini dilakukan oleh dua organisasi, AGB Nielsen dan TNmS. Proses penghitungan *rating* dilakukan bersamaan dengan penayangan program atau drama. *Rating* penonton dihitung selama penayangan program, dan nilai *rating* akan diberikan kepada setiap stasiun TV yang menyiarkan semua program yang tayang pada hari itu [1].

Di era ini banyak sekali jenis drama Korea, tidak semua orang bisa menikmatinya, tidak semua orang menyukai semua drama Korea, atau ada beberapa jenis (*genre*) drama Korea yang populer di kalangan masyarakat. Oleh karena itu, perlu diketahui dari penilaian penonton drama Korea seperti apa selera drama Korea yang diharapkan penonton drama Korea. Untuk mengetahui *rating* atau penilaian-penilaian dari sebuah drama Korea sebelum drama Korea tersebut dibuat atau

didistribusikan dapat digunakan metode *data mining* untuk melakukan proses prediksi. Pengumpulan data perlu dilakukan terlebih dahulu, biasanya data yang diperoleh bersifat *big data* yang berarti pengumpulan data dari berbagai macam sumber yang relevan [5]. Salah satu metode pembelajaran yang digunakan dalam prediksi adalah *Naïve Bayes* yaitu metode yang memiliki kemampuan perhitungan matematik dasar yang kuat dan efisiensi klasifikasi yang stabil. Namun, kekurangannya adalah perlu mengestimasi parameter model *Naïve Bayes* dan kurang peka terhadap data yang sudah hilang. Dibandingkan dengan algoritma klasifikasi lainnya, model *Naïve Bayes* memiliki tingkat kesalahan yang sangat minimum [6].

Pada penelitian sebelumnya, terkait dengan Metode *naïve bayes*. *Paper* yang ditulis oleh Rizki Wijayatun Pratiwi dan Yusuf Sulisty Nugroho (2016) dengan judul Prediksi *Rating* Film Menggunakan Metode *Naïve Bayes* menghasilkan nilai *accuracy* 55,80%, *precision* 32,41%, serta *recall* 46,70% menunjukkan bahwa mayoritas prediksi *rating* film rendah [7]. Adapun *paper* yang ditulis oleh Guruh Ihda Alfi Ath Thaariq, Dina Ratnaningtyas dan Shoima Fitra Kumala (2020) dengan judul Klasifikasi *Rating* Kegiatan Kampus Menggunakan Metode Algoritma *Naïve Bayes Classifier*. Hasil penelitian membuktikan bahwa model *Naïve Bayes* dapat menentukan kegiatan mana yang baik, favorit, kurang baik, dan tidak kurang menarik bukan hanya dari satu aspek tetapi dari beberapa aspek terkait [8].

Berdasarkan latar belakang tersebut, penulis akan menggunakan metode *Naïve Bayes* untuk memprediksi rating drama korea dengan tujuan untuk membangun sistem yang dapat memprediksi *rating* yang dinyatakan dengan status

*rating* (tinggi atau rendah) dengan menggunakan Algoritma *Naïve Bayes*, sehingga dapat mempermudah industri untuk melakukan prediksi keuntungan *rating* yang akan didapatkan jika membuat drama dengan variabel prediksi antara lain stasiun, *genre*, *writer*, *director*, dan *actor*. Diharapkan dengan variabel tersebut kemudian menggunakan metode yang sudah dibuat sedemikian rupa akan muncul prediksi *rating* yang akan menjadi acuan industri dalam membuat drama agar ketika ditayangkan tidak mengecewakan pihak industri maupun penikmat drama sendiri.

### 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan, maka permasalahan yang dapat dirumuskan adalah:

1. Bagaimana cara menerapkan metode *Naïve Bayes* dalam memprediksi *rating* drama Korea?
2. Bagaimana hasil akurasi dari penerapan metode *Naïve Bayes* untuk memprediksi *rating* drama Korea?

### 1.3 Batasan Masalah

Adapun batasan masalah yang digunakan peneliti agar pembahasan dalam penelitian ini tidak menyimpang dari pokok pembahasan, maka digunakan masalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini adalah data sampel dari *website* koreandrama dan asianwiki dengan data drama drama yang tayang mulai 2005 sampai dengan Desember 2020.
2. Semua proses perhitungan yang melalui teknik *data mining* dengan metode *Naïve Bayes*.

3. Teknik pengujian yang digunakan adalah *k-fold cross validation* dengan nilai *k* diatur dari 2 sampai 10.
4. Output yang dihasilkan dari sistem ini adalah berupa keterangan tinggi atau rendah yang didapatkan dari hasil prediksi.

#### **1.4 Maksud dan Tujuan Penelitian**

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengimplementasikan metode *Naïve Bayes* untuk memprediksi *rating* drama Korea.
2. Mengetahui berapa hasil akurasi prediksi *rating* drama Korea dengan menggunakan metode *Naïve Bayes*.

#### **1.5 Manfaat Penelitian**

Manfaat dari penelitian ini adalah sebagai berikut:

1. Mampu menghasilkan sistem yang dapat memprediksi nilai *rating* dari drama baru yang akan tayang mendapatkan *rating* tinggi atau rendah, sehingga dapat mempermudah industri untuk melakukan prediksi keuntungan *rating* yang akan didapatkan.
2. Penulis dapat menambah ilmu dan wawasan baru mengenai analisis prediksi *rating* drama Korea.

#### **1.6 Metode Penelitian**

Metode penelitian adalah langkah yang dimiliki dan dilakukan oleh peneliti dalam rangka untuk mengumpulkan informasi atau data. Metode yang dipergunakan untuk memperoleh data yang digunakan dalam penulisan skripsi ini adalah:

### **1.6.1 Identifikasi Perumusan Masalah**

Tahapan penelitian di mana merumuskan masalah yang akan diteliti, dalam kasus ini perumusan masalah adalah bagaimana cara menerapkan metode *Naïve Bayes* dalam memprediksi *rating* drama Korea dan bagaimana hasil akurasi dari penerapan metode *Naïve Bayes* untuk memprediksi *rating* drama Korea.

### **1.6.2 Pengumpulan Data**

Tahapan penelitian dimana data yang dikumpulkan dilakukan dengan cara studi dokumentasi yaitu teknik pengumpulan data yang tidak langsung pada subjek penelitian, melainkan pada dokumen-dokumen tertentu. Jenis data yang digunakan dalam penelitian ini merupakan data sekunder. Data tersebut merupakan drama yang tayang mulai 2005 sampai dengan Desember 2020. Metode yang dipergunakan dalam pengumpulan data untuk merancang dan menganalisis adalah sebagai berikut:

#### **1.6.2.1 Melakukan Studi Pustaka**

Tahapan penelitian dimana data dan informasi yang didapat dilakukan dari studi literatur tentang drama Korea, *data mining*, *Naïve Bayes*, dan literatur yang mendasari masalah, informasi, dan pemikiran-pemikiran yang relevan dalam melakukan penelitian ini, serta bersumber dari buku, jurnal, dan skripsi yang dipublikasikan di dalam WWW (*World Wide Web*).

#### **1.6.2.2 Dokumentasi Data**

Pengumpulan data dari penelitian ini yaitu data dari website koreandrama dan AsianWiki dengan melakukan ekstraksi data dan informasi dari suatu *website*

kemudian menyimpannya dalam format tertentu. Pada penelitian ini melakukan pengumpulan data dengan python menggunakan *Selenium*.

### **1.6.3 Analisis Kebutuhan**

Analisis kebutuhan merupakan proses dalam memperoleh informasi, model, dan spesifikasi tentang perangkat lunak yang dibutuhkan oleh pengguna. Informasi yang diperoleh pengguna akan menjadi acuan dalam perancangan perangkat lunak.

#### **1.6.3.1 Metode Analisis Data**

Metode analisis data dalam penelitian ini adalah metode *Naïve Bayes Classifier* yang digunakan untuk melakukan klasifikasi terhadap *rating* pada drama Korea menjadi 2 kelas yaitu, *rating* tinggi dan *rating* rendah.

### **1.6.4 Metode Perancangan**

Metode perancangan menggunakan UML (*Unified Modeling Language*).

### **1.6.5 Metode Pengembangan Sistem**

Sistem ini menggunakan model *waterfall* dengan metode pengembangan sistem SDLC (*System Development Life Cycle*). Disebut *waterfall* karena harus menunggu langkah sebelumnya untuk menyelesaikan langkah demi langkah.

### **1.6.6 Metode Testing**

Metode testing dapat dilakukan dengan 4 cara, yaitu:

1. *White Box Testing* adalah metode pengujian aplikasi atau perangkat lunak dengan mempelajari dan menganalisis apakah kode program yang dibuat

terdapat kesalahan atau tidak. Jika bentuk keluaran dari model yang dihasilkan tidak memenuhi harapan, kode akan dikompilasi ulang dan diperiksa ulang hingga memenuhi harapan.

2. *Black Box Testing* memainkan peran penting dalam pengujian perangkat lunak yang membantu dalam validasi fungsionalitas keseluruhan sistem. Perangkat lunak tersebut akan dieksekusi kemudian berusaha dites apakah telah memenuhi kebutuhan pengguna yang didefinisikan pada saat awal tanpa harus membongkar *listing* programnya. Pengujian *black box* cenderung diaplikasikan selama tahap akhir pengujian.
3. *K-fold Cross Validation* merupakan sebuah teknik verifikasi validasi model yang digunakan untuk mengevaluasi bagaimana hasil analisis statistik akan digeneralisasikan ke kumpulan data independen. Teknik ini terutama digunakan untuk membuat prediksi model dan memperkirakan seberapa akurat sebuah model prediktif ketika dijalankan dalam praktiknya.
4. *Confusion matrix* merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja metode klasifikasi. Pada umumnya *confusion matrix* berisi informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil yang seharusnya diklasifikasikan.

### 1.6.7 Penarikan Kesimpulan

Langkah terakhir dalam penelitian ini yaitu penarikan kesimpulan, tahap ini berisikan tentang rangkuman penelitian dan hasil yang didapatkan dalam prediksi *rating* drama Korea. Penarikan kesimpulan dilakukan dengan memperhatikan hasil



dari klasifikasi penelitian ini dan perolehan performa dari analisis data.

## **1.7 Sistematika Penulisan**

Sesuai dengan petunjuk penulisan laporan skripsi yang berlaku di Universitas Amikom Yogyakarta, sistematika penulisan laporan ini adalah sebagai berikut:

### **BAB I PENDAHULUAN**

Berisi latar belakang, rumusan masalah, batasan masalah, maksud dan tujuan penelitian, metode penelitian, dan sistematika penulisan.

### **BAB II LANDASAN TEORI**

Bab ini berisi tentang studi pustaka mengenai *data mining*, *machine learning*, *Naïve Bayes*, dan literatur yang mendasari masalah, informasi, dan pemikiran-pemikiran yang relevan dalam melakukan penelitian ini.

### **BAB III METODE PENELITIAN**

Bab ini berisi mengenai alur penelitian, penjelasan alur penelitian, analisis kebutuhan, uml (*unified modelling language*), perancangan sistem, dan perhitungan manual.

### **BAB IV HASIL DAN PEMBAHASAN**

Menjelaskan mengenai tahapan penelitian berupa hasil penelitian secara menyeluruh serta hasil pengujian.

### **BAB V PENUTUP**

Penutup berupa kesimpulan dan saran yang didapatkan dari hasil penelitian.

### **DAFTAR PUSTAKA**

Bab ini berisikan daftar referensi-referensi yang telah digunakan dalam penulisan dan pengembangan.