

## BAB I

### PENDAHULUAN

#### 1.1 Latar Belakang

Anime merupakan salah satu bentuk hiburan visual yang sangat populer dan berkembang pesat dalam industri kreatif dengan basis penggemar yang luas di berbagai negara. Penelitian ini menggunakan dataset anime yang besar dan beragam, mencakup 19.311 entri anime dari berbagai *genre* dan seri. Analisis preferensi *genre* anime menjadi sangat menarik untuk dipelajari secara kuantitatif menggunakan pendekatan *data-driven*, khususnya dengan metode *clustering* yang mampu mengelompokkan anime berdasarkan kemiripan karakteristik *genre*, *rating*, dan popularitas. Dalam penelitian ini, dua algoritma *clustering* yang umum digunakan, yaitu *K-Means* dan *DBSCAN*, diterapkan dan dibandingkan untuk mengidentifikasi pola preferensi *genre* anime yang beragam dari dataset besar tersebut.

Penelitian mengenai penerapan metode *K-Means* dalam analisis data anime telah banyak dilakukan dalam beberapa tahun terakhir. Sebagai contoh, Zalzabila dan Prathiwi (2025) melakukan analisis preferensi penonton anime berdasarkan *genre* film menggunakan metode *K-Means*. Penelitian ini mengelompokkan 100 judul anime populer berdasarkan *rating*, jumlah suara, dan *genre*, menghasilkan empat *cluster* dengan karakteristik unik yang menunjukkan perbedaan dalam popularitas dan preferensi *genre*. [1]

Penelitian oleh Parlaungan et al. (2022) mengembangkan sistem rekomendasi film berbasis web menggunakan metode *K-Means Clustering*. Penelitian ini bertujuan untuk meningkatkan akurasi rekomendasi film dengan cara mengelompokkan data berdasarkan atribut yang memiliki kemiripan. Sistem yang dibangun memungkinkan pengguna mencari film berdasarkan kombinasi kriteria seperti *rating*, popularitas, dan *genre*, serta memberikan fleksibilitas dalam memilih satu atau lebih *genre* dan rentang tahun rilis sesuai preferensi. [2]

Sementara itu, Widiani dan Lestari (2024) menerapkan algoritma *K-Means Clustering* pada big data untuk sistem rekomendasi film. Dengan memanfaatkan dataset berukuran besar, penelitian ini berhasil mengelompokkan film berdasarkan kesamaan karakteristik seperti *genre*, popularitas, dan rating. Hasil dari proses clustering ini digunakan untuk memberikan rekomendasi film yang lebih relevan kepada pengguna. Penelitian ini menunjukkan bahwa penerapan *K-Means* dalam skala data besar dapat meningkatkan efisiensi dan kualitas sistem rekomendasi berbasis konten. [3]

Studi oleh Cahyani et al. (2024) mengenai pengelompokan film *Netflix* berdasarkan *genre* dan *rating* juga membuktikan bahwa metode ini efektif dalam mengelompokkan data besar secara efisien. [4] Penelitian oleh Fitrianti et al. (2024) menerapkan algoritma *K-Means* untuk mengelompokkan film populer di *Netflix* berdasarkan durasi, rating, dan jumlah *votes*. Hasil clustering menghasilkan tiga kelompok film dengan karakteristik berbeda yang dapat membantu pengguna dalam menemukan film sesuai preferensi mereka. Selain itu, hasil analisis tersebut juga memberikan wawasan yang dapat dimanfaatkan oleh pengembang sistem rekomendasi dalam merancang strategi personalisasi konten secara lebih akurat. [5]

Penelitian Mutiah et al. (2024) membandingkan *K-Means* dan *DBSCAN* dalam mengelompokkan data sosial ekonomi rumah tangga, dan menemukan *DBSCAN* lebih efektif menangani klaster dengan bentuk tidak teratur serta outlier. *DBSCAN* tidak perlu menentukan jumlah klaster terlebih dahulu dan mampu mengelola noise dengan baik, sehingga cocok untuk data kompleks dan heterogen. Meskipun *K-Means* menghasilkan klaster yang lebih terstruktur, *DBSCAN* lebih mampu menangkap variasi nyata dalam data yang tidak homogen, menjadikannya metode yang tepat sebagai pelengkap *K-Means* untuk analisis preferensi *genre* anime yang beragam. [6]

Meskipun metode *K-Means* telah banyak digunakan dalam analisis *clustering* data anime, sebagian besar penelitian terbatas pada dataset dengan ukuran kecil hingga menengah dan belum banyak yang membandingkan performa

*K-Means* dengan algoritma berbasis kepadatan seperti *DBSCAN* pada dataset anime berskala besar. Hal ini menimbulkan kebutuhan untuk melakukan analisis perbandingan kedua metode tersebut guna memperoleh pemahaman yang lebih komprehensif mengenai pola preferensi genre anime dengan cakupan data yang lebih luas dan heterogen.

## 1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang ini, rumusan masalah yang akan dibahas adalah:

1. Bagaimana algoritma *K-Means* dan *DBSCAN* dapat mengelompokan anime berdasarkan preferensi *genre* dengan dataset yang berisi 19.311 entri?
2. Berapa jumlah cluster optimal pada metode *K-Means* berdasarkan evaluasi *Elbow Method*, *Silhouette Score*, dan *Davies-Bouldin Index*?
3. Apa karakteristik utama dan pola preferensi *genre* yang ditemukan dari hasil *clustering* kedua algoritma tersebut?

## 1.3 Batasan Masalah

Batasan masalah ini ditetapkan agar ruang lingkup penelitian tetap fokus dan sesuai dengan tujuan yang ingin dicapai, serta untuk menghindari pembahasan yang terlalu luas. Berikut adalah beberapa batasan masalah dari penelitian ini:

1. Dataset yang digunakan mencakup 19.311 entri anime dengan atribut seperti *genre*, *rating*, jumlah episode, dan popularitas.
2. Penelitian ini menggunakan algoritma *K-Means* dan *DBSCAN*, serta membandingkan hasil *clustering* keduanya.
3. Evaluasi jumlah cluster optimal hanya dilakukan untuk *K-Means* menggunakan *Elbow*, *Silhouette*, dan *Davies-Bouldin Index*, sedangkan *DBSCAN* dievaluasi menggunakan *Silhouette Score* dan Analisis Noise.

## **1.4 Tujuan Penelitian**

Adapun tujuan penelitian adalah sebagai berikut:

1. Melakukan pengelompokan anime berdasarkan preferensi *genre* menggunakan algoritma *K-Means* dan *DBSCAN*.
2. Menentukan jumlah *cluster* optimal pada *K-Means* dengan metode evaluasi yang valid dan terpercaya.
3. Menganalisis dan membandingkan hasil *clustering* kedua algoritma untuk memahami pola preferensi *genre* anime secara lebih mendalam.

## **1.5 Manfaat Penelitian**

Penelitian ini diharapkan dapat memberikan kontribusi baik dari segi teori maupun penerapan teknis. Manfaat penelitian dapat dijabarkan menjadi 3, yaitu:

1. Manfaat Teoritis dan Teknikal
  - a. Menambah wawasan tentang penerapan metode *clustering K-Means* dan *DBSCAN* dalam pengelompokan data besar dan kompleks, khususnya pada data hiburan seperti anime.
  - b. Memberikan dasar teoritis dalam pengembangan algoritma *clustering* yang lebih adaptif untuk data *multi-genre* dan heterogen yang memiliki karakteristik tidak seragam.
  - c. Memperkaya literatur tentang evaluasi metode *clustering* pada dataset dengan jumlah data besar dan variasi atribut yang luas.
2. Manfaat Praktis
  - a. Membantu platform streaming dan komunitas penggemar anime memahami pola preferensi *genre* yang beragam dan tren popularitas anime secara lebih menyeluruh.
  - b. Memberikan *insight* strategis bagi industry kreatif dan penyedia konten dalam merancang sistem rekomendasi yang lebih akurat dan personalisasi konten berdasarkan hasil *clustering*.

### 3. Manfaat bagi Peneliti Selanjutnya

- a. Menjadi referensi metodologis bagi peneliti yang ingin membandingkan atau mengembangkan algoritma *clustering* pada data hiburan atau dataset besar lainnya.
- b. Memberikan gambaran praktis terkait pengolahan dan analisis data anime berskala besar yang dapat digunakan untuk penelitian lebih lanjut dalam bidang sistem rekomendasi dan *data mining*.

### 1.6 Sistematika Penulisan

Sistematika penulisan ini dibuat dengan tujuan untuk mempermudah pemahaman penelitian. Laporan skripsi ini disusun secara sistematis, adapun bagian-bagiannya sebagai berikut:

Contoh :

**BAB I PENDAHULUAN**, berisi Latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

**BAB II TINJAUAN PUSTAKA**, berisi uraian teori-teori yang mendasari penelitian serta studi literatur atau jurnal penelitian terdahulu yang relevan.

**BAB III METODE PENELITIAN**, menjelaskan tentang objek penelitian, alur penelitian, metode pengumpulan dan pengolahan data, serta tahapan-tahapan yang digunakan dalam proses analisis data menggunakan metode *K-Means* dan *DBSCAN*.

**BAB IV HASIL DAN PEMBAHASAN**, menyajikan hasil penelitian berdasarkan pengolahan data yang dilakukan, serta pembahasan terhadap hasil tersebut dengan mengaitkannya pada teori dan temuan dari penelitian sebelumnya.

**BAB V PENUTUP**, berisi kesimpulan dari hasil penelitian serta saran yang dapat diberikan peneliti untuk pengembangan penelitian selanjutnya.