

## BAB V PENUTUP

### 5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma *K-Means* dan *DBSCAN* dapat digunakan secara efektif untuk mengelompokkan anime berdasarkan preferensi *genre*. Berikut ini merupakan kesimpulan yang disusun berdasarkan rumusan masalah penelitian:

1. *K-Means* mengelompokkan data berdasarkan jarak ke *centroid*, sedangkan *DBSCAN* menggunakan kepadatan data sebagai dasar pembentukan cluster. Dalam penelitian ini, *K-Means* diuji pada rentang jumlah cluster  $K = 2$  hingga 11, dan *DBSCAN* menggunakan parameter  $\epsilon = 70$  dan  $\min\_samples = 5$ . Kedua algoritma mampu mengelompokkan lebih dari 19.311 entri anime ke dalam beberapa cluster, dengan pendekatan numerik berbasis skor, popularitas, dan jumlah episode. *DBSCAN* memiliki keunggulan dalam memisahkan outlier dan menangani distribusi data yang tidak merata, kelebihan yang tidak dimiliki oleh *K-Means*.
2. Evaluasi jumlah cluster optimal dilakukan dengan menguji nilai  $k$  pada rentang 2 hingga 11 menggunakan *Elbow Method*, *Silhouette Score*, dan *Davies-Bouldin Index*. Hasil terbaik ditemukan pada  $k = 2$ , dengan nilai *SSE* sebesar 82.157.487.071, *Silhouette Score* sebesar 0,6550, dan *Davies-Bouldin Index* sebesar 0,4586. Nilai *Silhouette Score* yang cukup tinggi menunjukkan bahwa data dalam satu cluster sangat mirip dan terpisah jelas dari cluster lain, sedangkan nilai *Davies-Bouldin Index* yang rendah mengindikasikan bahwa antar cluster memiliki jarak yang baik dan bentuk yang kompak. Dalam *K-Means*, setiap cluster memiliki *centroid*, yaitu titik pusat rata-rata dari data dalam cluster tersebut. Pada  $k = 2$ , dua centroid yang terbentuk mampu merepresentasikan distribusi data dengan baik, sehingga jumlah cluster tersebut dinilai paling optimal berdasarkan hasil evaluasi.
3. Hasil clustering menggunakan algoritma *K-Means* dan *DBSCAN*

menunjukkan adanya perbedaan karakteristik *genre* pada setiap *cluster* yang terbentuk, terlihat seperti penjelasan berikut ini:

- a. Pada algoritma *K-Means*, *Cluster 0* terdiri dari 7.484 anime dengan *genre* dominan *comedy* (2.208), *kids* (1.999), dan *music* (1.463), yang menggambarkan kecenderungan terhadap anime yang ditujukan untuk audiens berusia muda. Sementara itu, *Cluster 1* terdiri dari 7.703 anime dengan *genre* yang lebih bervariasi, yaitu *comedy* (3.614), *action* (2.727), *drama* (1.872), *fantasy* (1.764), dan *adventure* (1.703), yang menunjukkan preferensi dari audiens yang lebih luas, termasuk remaja dan dewasa.
- b. Pada algoritma *DBSCAN*, *Cluster 0* mencakup sebagian besar data, yaitu sebanyak 15.081 anime, dengan *genre* dominan seperti *comedy*, *action*, *fantasy*, *adventure*, dan *drama*, yang merupakan *genre* populer dalam industri anime. Sementara itu, *Cluster 1* hanya terdiri dari 6 anime yang memiliki kombinasi *genre* seperti *action*, *adventure*, *shounen*, dan *fantasy*, yang tidak umum ditemukan dalam pola distribusi utama, sehingga dikelompokkan secara terpisah. Selain itu, terdapat 100 data yang diklasifikasikan sebagai *noise* atau *outlier* karena tidak memenuhi kriteria kepadatan minimum untuk dimasukkan ke dalam *cluster* manapun.

## 5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, menggunakan metode *K-Means* dan *DBSCAN* untuk mengelompokkan anime berdasarkan *genre*, terdapat beberapa saran yang dapat memperbaiki dan mengembangkan penelitian lebih lanjut, seperti:

1. Meskipun *K-Means* dan *DBSCAN* memberikan hasil yang berguna, penelitian ini masih dapat lebih diperluas dengan menggunakan algoritma *clustering* lain, seperti *Agglomerative Clustering* atau *Gaussian Mixture Models (GMM)*. Metode ini dapat memberikan hasil yang lebih stabil atau lebih sesuai untuk dataset dengan kepadatan yang bervariasi. Oleh karena

itu, melakukan eksperimen dengan algoritma lain dapat memperkaya perbandingan hasil *clustering*.

2. *DBSCAN* sangat bergantung pada pengaturan *epsilon* dan *min\_samples*. Oleh karena itu, eksperimen lebih lanjut dengan mencoba berbagai nilai *epsilon* dan *min\_samples* dapat membantu menentukan parameter terbaik untuk menghasilkan *cluster* yang lebih representatif. Menggunakan teknik seperti *Grid Search* atau *Cross-Validation* untuk mencari nilai parameter yang optimal dapat meningkatkan kualitas hasil *DBSCAN*.
3. Penelitian ini menggunakan *Silhouette Score*, *Davies-Bouldin Index*, dan *SSE* untuk evaluasi, namun masih banyak metrik lain yang dapat digunakan untuk analisis lebih mendalam. Salah satunya adalah *Adjusted Rand Index (ARI)* atau *V-Measure*, yang dapat memberikan gambaran lebih detail tentang kualitas pemisahan *cluster*. Dengan menggunakan metrik ini, evaluasi hasil *clustering* dapat menjadi lebih komprehensif.
4. Hasil *clustering* yang diperoleh dari *K-Means* dan *DBSCAN* dapat digunakan untuk membangun sistem rekomendasi anime berbasis preferensi *genre*. Dengan mengetahui *genre* dominan di setiap *cluster*, sistem rekomendasi dapat memberikan rekomendasi yang lebih relevan dan akurat sesuai dengan preferensi pengguna. Penambahan fitur seperti *rating* pengguna atau riwayat tontonan dapat lebih menyempurnakan dan meningkatkan akurasi sistem ini.