

BAB I

PENDAHULUAN

1.1 Latar Belakang

Webtoon merupakan salah satu platform komik digital dari Korea Selatan yang diluncurkan oleh salah satu perusahaan teknologi yaitu LINE Corporation dan Naver Corporation [1]. Line merupakan salah satu anak dari perusahaan Naver yang lahir pada tahun 1999 di Korea Selatan [2]. Adapun Naver Corporation merupakan perusahaan terkemuka di Korea Selatan dalam bidang teknologi, informasi, dan komunikasi, mengembangkan berbagai produk termasuk mesin pencarian, aplikasi pengiriman pesan, platform video, dan komik digital. [3]. Webtoon membagi jenis genre berdasarkan usianya. Genre Webtoon ini terdapat beberapa kelompok, yaitu drama, *fantasy*, *slice of life*, *horror*, dan *informative* [4]. Berdasarkan kelompok usia, LINE Webtoon dapat dikategorikan menjadi tiga segmen, yakni remaja, kelompok usia 20-an, dan kelompok usia 30-an. Dengan menggunakan metode klustering, berdasarkan judul, rating, dan genre akan dikelompokkan ke dalam tiga kategori, yaitu kategori tertinggi, kategori sedang, dan kategori rendah.

Saat ini bidang ilmu komputer telah menyediakan berbagai alat pengukuran untuk mengetahui pengelompokan popularitas pada aplikasi Webtoon salah satunya pada metode klustering. Metode klustering merupakan salah satu teknik penerapan data mining yang digunakan untuk menganalisis dan mengevaluasi data dengan tujuan menyelesaikan permasalahan pengelompokan data melibatkan proses pembagian dataset menjadi subset yang lebih teratur. [5]. Metode untuk melakukan klustering dapat dikategorikan menjadi empat metode yaitu: *portioning*, *hierarchical*, *gril-based*, and *model based* [6].

Terdapat beberapa metode dalam klastering seperti *Mixture Modeling*, K-Medoids dan K-Means. *Mixture Modeling* adalah suatu pendekatan yang digunakan untuk memodelkan atau mengelompokkan data dalam suatu dataset menjadi kelompok-kelompok yang tidak terdefinisi sebelumnya. Pendekatan ini bertujuan untuk mengidentifikasi pola atau struktur yang mungkin ada dalam data, dengan memperkenalkan konsep kelompok campuran atau komponen yang membentuk distribusi data secara keseluruhan [7]. Algoritma *Mixture Modeling* memiliki kelebihan mampu untuk memodelkan lebih dari satu *gaussian* untuk sebuah dataset [8]. Namun, disisi lain terdapat kekurangan dalam algoritma *Mixture Modeling* yaitu ketergantungan pada inisiasi awal yang baik, kompleksitas perhitungan, sensitivitas terhadap ukuran dataset, dan asumsi distribusi yang terbatas. Metode K-Medoids adalah salah satu algoritma pengelompokan yang efisien untuk mengatasi dataset dengan ukuran kecil [9]. Algoritma K-Medoid memiliki keunggulan dalam menangani kelemahan yang muncul pada Algoritma K-Means yang cenderung responsif terhadap pencilan (outlier). Hal ini disebabkan oleh kemampuan Algoritma K-Medoid dalam menangani objek yang memiliki nilai yang signifikan jauh dari distribusi data. [10]. Namun, disisi lain Algoritma K-Medoids ini memiliki kekurangan yaitu dalam penentuan atau penginisialisasian kluster awal yang sangat mempengaruhi hasil dari kluster tersebut [9].

Salah satu Algoritma pada metode klastering yang paling banyak digunakan adalah Algoritma K-Means. Algoritma K-Means merupakan suatu teknik klastering *non-hierarki* yang berfokus pada pembagian data ke dalam satu atau beberapa kelompok. Tujuan utama dari algoritma ini adalah untuk menempatkan data yang memiliki karakteristik serupa ke dalam satu kelompok, sementara data dengan karakteristik yang berbeda ditempatkan dalam kelompok terpisah. Metode ini bertujuan untuk mengorganisir data menjadi kelompok-kelompok yang homogen berdasarkan kesamaan karakteristiknya, menciptakan partisi yang memungkinkan untuk analisis lebih lanjut [11]. Adapun kelebihan dalam Algoritma K-Means ini yaitu sederhana, efisien, mudah dipahami dan mudah diterapkan [12]. Di samping keunggulan tersebut, K-Means juga mempunyai kelebihan lain, seperti kemudahan dalam menetapkan kluster awal, terutama pada dataset yang tidak terlalu besar.

Penetapan centroid perlu dilakukan sebelum perhitungan jumlah kluster. Meskipun kluster sebenarnya tidak diketahui menggunakan data yang sama, perlu diingat bahwa jika data dimasukkan dengan cara yang berbeda, besar kemungkinan akan menghasilkan pembentukan kluster yang berbeda juga. Selain itu, setiap atribut memberikan kontribusi dengan bobot yang sama [13]. Namun, K-Means bergantung pada nilai centroid untuk pengelompokan. Pengelompokan dengan nilai centroid yang berbeda akhirnya akan menghasilkan hasil yang berbeda.

Penentuan nilai *centroid* menjadi masalah penting dalam klustering dimana ketika salah dalam penentuan nilai *centroid* akan menyebabkan K-Means jatuh ke dalam titik minimum lokal yang buruk. Titik minimum lokal adalah kondisi dimana suatu titik dalam ruang solusi yang merupakan lokal terendah atau terbaik dalam kaitannya dengan fungsi objektif yang sedang dioptimalkan. Algoritma yang terjebak pada titik minimum lokal yang buruk akan gagal menemukan solusi yang lebih baik secara global [14]. Untuk mengoptimalkan nilai *centroid*, maka membutuhkan optimasi untuk menghindari terjebak pada titik lokal yang buruk.

Terdapat beberapa metode yang dapat digunakan untuk mengatasi K-Means tersebut salah satunya algoritma *metaheuristic*. Algoritma *metaheuristic* adalah suatu pendekatan komputasional yang diterapkan untuk menangani masalah optimasi yang bersifat sulit atau kompleks. Pendekatan ini seringkali digunakan ketika metode-metode optimasi konvensional tidak dapat memberikan solusi yang efisien atau memadai untuk permasalahan tertentu. Algoritma ini dirancang untuk mencari solusi mendekati optimal melalui eksplorasi dan eksploitasi ruang solusi dengan cara yang efisien dan efektif [15]. Berbeda dengan metode optimasi konvensional, pendekatan *metaheuristic* tidak mengandalkan informasi yang eksplisit dari masalah, seperti turunan atau struktur matematik yang khusus. Salah satu *metaheuristic* yang bisa digunakan untuk proses pengelompokan data, atau klustering adalah algoritma genetika. Algoritma Genetika (GA) adalah algoritma berdasarkan prinsip evolusi alami untuk melakukan pencarian dan pengoptimalan hasil [16]. Berdasarkan beberapa penelitian sebelumnya, Algoritma Genetika diterapkan dalam upaya mengoptimalkan K-Means dan berbagai metode klustering lainnya untuk melakukan pengelompokan dataset. Dari sisi positif, Algoritma

Genetika memiliki sejumlah keunggulan, antara lain kemampuannya dalam menghasilkan solusi yang lebih optimal. Algoritma ini menggunakan berbagai operator genetik, termasuk seleksi, mutasi, dan crossover, untuk melakukan peningkatan dan variasi dalam populasi solusi, yang berkontribusi pada penemuan solusi yang lebih baik. [17].

Berdasarkan kelebihan dari Genetika Algoritma maka penelitian ini akan mengkombinasikan K-Means yang memiliki kelemahan dalam menentukan nilai centroid dengan algoritma genetika sehingga kluster menjadi lebih optimal. Diharapkan dengan kombinasi kedua algoritma tersebut dapat menentukan kelompok genre dalam aplikasi Webtoon dan juga pemilihan metode yang akan digunakan nantinya.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas maka didapatkan rumusan masalah yaitu Penentuan centroid menjadi masalah penting dalam klastering dimana ketika salah dalam penentuan nilai *centroid* akan menyebabkan K-Means jatuh ke dalam titik minimum lokal yang buruk. Sehingga mengakibatkan hasil klastering yang tidak optimal atau tidak mewakili struktur sebenarnya.

1.3 Pertanyaan Penelitian

1. Berapa jumlah kluster optimal yang dihasilkan algoritma K-Means dalam pengelompokan popularitas dalam aplikasi Webtoon?
2. Bagaimana pengaruh Algoritma Genetika dalam menentukan *centroid* yang optimal pada K-Means?
3. Berapa hasil perbandingan kinerja antara K-Means dengan K-Means yang menggunakan Algoritma Genetika?

1.4 Batasan Masalah

1. Data yang sudah diperoleh melalui *website* Kaggle
2. Penelitian ini akan melibatkan penggunaan algoritma K-Means untuk menghasilkan hasil yang lebih mudah diinterpretasikan.

3. Penelitian ini akan mengelompokkan popularitas dalam aplikasi Webtoon menggunakan algoritma K-Means.

1.5 Tujuan Penelitian

1. Mengelompokkan popularitas dalam aplikasi Webtoon menggunakan algoritma K-Means
2. Menganalisis pengaruh Algoritma Genetika dalam penentuan *centroid* yang optimal dalam algoritma K-Means
3. Menganalisis perbandingan kinerja K-Means dengan K-Means yang menggunakan Algoritma Genetika.

1.6 Manfaat Penelitian

1. Menganalisis beberapa kelompok atau kluster popularitas yang terdapat dalam aplikasi Webtoon setelah dilakukannya klastering menggunakan K-Means dan Algoritma Genetika.
2. Menganalisis bagaimana cara pengolahan dalam mendapatkan nilai *centroid* menggunakan K-Means dan Algoritma Genetika.
3. Mendapatkan hasil perbandingan kinerja K-Means dengan K-Means yang menggunakan Algoritma Genetika.