

BAB 1

PENDAHULUAN

Bab ini membahas enam hal, yaitu latar belakang masalah, rumusan masalah, ruang lingkup, tujuan penelitian, manfaat penelitian, dan sistematika penulisan. Keenam hal tersebut dibahas pada subbab berikut ini.

1.1 Latar Belakang Masalah

Klasterisasi merupakan proses pengelompokan yang tidak melakukan klasifikasi, estimasi, atau memprediksi nilai dari variabel acuan, tetapi membagi keseluruhan data menjadi kelompok yang memiliki kemiripan. Kemiripan data dalam satu kelompok akan bernilai maksimum sedangkan kemiripan data antar kelompok akan bernilai minimum. Klasterisasi dapat digunakan untuk mengetahui struktur dalam data yang dapat dipakai lebih lanjut dalam berbagai aplikasi secara luas (Yohannes 2016).

Yudi Agusta (2007) dalam jurnalnya menjelaskan klastering merupakan salah satu metode *data mining* yang bersifat tanpa arahan (*unsupervised*). Ada dua jenis data klastering yang sering dipergunakan dalam proses pengelompokan data yaitu *hierarchical* (hirarki) data *clustering* dan *non-hierarchical* (non hirarki) data *clustering*. Algoritma non hirarki merupakan algoritma yang mengharuskan pemakai menentukan jumlah k partisi yang diinginkan lalu setiap data dites untuk dimasukkan pada salah satu partisi sehingga tidak ada data yang *overlap* dan satu data hanya memiliki satu kluster. Contohnya adalah algoritma $K - Means$, Fuzzy $C - Means$.

Rivani (2010) mengaplikasikan K – Means klastering terhadap data produksi tanaman pangan utama seperti padi, jagung, kedelai, dan kacang hijau masing-masing provinsi, sehingga bisa diketahui pengelompokan provinsi berdasarkan data tersebut. Setiap pengelompokan juga dapat dilihat masing – masing karakteristiknya, sehingga diketahui kelompok mana yang memproduksi keempat tanaman pangan utama tersebut dalam jumlah yang paling banyak dan sedikit.

Harliana dan Azhari (2012) melakukan klastering terhadap desa – desa yang masuk dalam kelompok rawan pangan dan tahan pangan serta faktor penyebabnya dengan menggunakan Fuzzy C - Means (FCM).

Oleh karena itu akan dilakukan penelitian tentang perbandingan analisis klaster dengan metode K – Means dan Fuzzy C – Means pada data yang sama. Data yang digunakan merupakan data *online retail*. Analisis klaster adalah mengelompokkan objek – objek berdasarkan kesamaan karakteristik di antara objek – objek tersebut. Kualitas suatu klaster dapat dilihat dari nilai yang dijadikan parameter. Seperti nilai SSE dan rasio Sw/Sb .

1.2 Rumusan Masalah

Rumusan masalah dibuat agar penelitian menjadi lebih terarah. Peneliti merumuskan masalah “Bagaimana perbandingan hasil kinerja metode K – Means dan Fuzzy C – Means untuk klastering data *Online Retail* berdasarkan nilai SSE dan rasio Sw/Sb ”.

1.3 Ruang Lingkup

Ruang lingkup yang akan dibahas dalam penelitian ini adalah

1. Metode klastering yang digunakan adalah K – Means dan Fuzzy C – Means.
2. Analisis yang dilakukan mengenai kualitas hasil klastering.
3. Jumlah klaster proses klastering dapat diubah – ubah.
4. Kualitas klastering ditentukan dengan nilai SSE dan rasio Sw/Sb.
5. Dataset yang digunakan adalah data *online retail* dari UCI Machine Learning Repository dengan 8 atribut, yaitu *InvoiceNo*, *StockCode*, *Description*, *Quantity*, *InvoiceDate*, *UnitPrice*, *CustomerID*, dan *Country*.
6. Atribut yang digunakan untuk klastering adalah *Quantity*, *Unit Price* dan *Country*.
7. Dataset *online retail* memiliki 541909 data, dan dipilih 1482 data sebagai sampel proses *clustering*.
8. Aplikasi yang dibuat berbasis desktop menggunakan bahasa pemrograman java.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah membandingkan hasil klastering metode K – Means dan Fuzzy C – Means berupa nilai SSE dan nilai rasio Sw/Sb serta menentukan metode mana yang lebih baik untuk klastering data *Online Retail*.

1.5 Manfaat Penelitian

- a. Penelitian dapat menambah wawasan dan pengetahuan tentang analisis kinerja metode K – Means dan Fuzzy C – Means.
- b. Penelitian dapat dijadikan sebagai bahan pertimbangan atau dikembangkan lebih lanjut, serta referensi terhadap penelitian yang sejenis.

1.6 Sistematika Penulisan

Bab 1 Pendahuluan, bab ini mencakup latar belakang masalah dilakukan analisis klaster dengan metode K – Means dan Fuzzy C – Means, rumusan masalah dan ruang lingkup perbandingan metode, tujuan penelitian perbandingan metode, manfaat yang diperoleh dari hasil penelitian, dan sistematika penulisan pada penelitian perbandingan metode.

Bab 2 Tinjauan Pustaka dan Dasar Teori, bab ini menjelaskan tinjauan pustaka dan dasar teori yang dibutuhkan untuk mendukung dalam perbandingan metode K – Means dan Fuzzy C – Means.

Bab 3 Metode Penelitian, bab ini berisi tentang penjelasan bahan yang dibutuhkan, peralatan yang digunakan, serta analisis sistem dan perancangan sistem untuk perbandingan metode K – Means dan Fuzzy C – Means guna memberikan kemudahan dalam memenuhi kebutuhan penelitian.

Bab 4 Implementasi dan Pembahasan, bab ini berisi implementasi dan uji coba aplikasi perbandingan metode K – Means dan Fuzzy C – Means dan pembahasan tentang hasil pengujian aplikasi.

Bab 5 Kesimpulan dan Saran, bab ini berisi kesimpulan jawaban dari rumusan masalah dan saran yang berisikan simpulan yang perlu dilanjutkan atau ditambahkan dalam perbandingan metode K – Means dan Fuzzy C – Means.