

BAB 1

PENDAHULUAN

1.1 Latar belakang

Data merupakan salah satu sumber yang dapat digunakan untuk memperoleh informasi. Akan tetapi, tidak jarang kumpulan data tersebut dibiarkan begitu saja seakan-akan menjadi kuburan data, sehingga diperlukan suatu metode yang dapat dipakai untuk menggali informasi sebanyak mungkin dari data tersebut. *Data Mining* sebagai salah satu ilmu di bidang teknologi informasi, dapat digunakan untuk mengekstraksi informasi berharga yang sebelumnya tidak diketahui dari suatu *database*. Sebagai contoh adalah data tagihan Pelanggan PT Telkom yang akan digunakan pada Tugas Akhir ini.

Salah satu informasi yang dapat digali dari data tersebut adalah pengelompokan pelanggan. Hal ini dilakukan untuk mendukung strategi manajemen yang bisa jadi berbeda untuk tiap kelompoknya. Data pelanggan ini terdiri dari beberapa atribut dengan jumlah *record* yang banyak sehingga diperlukan suatu proses *data mining* yang dapat mengelompokkan data tersebut, yaitu *clustering*. Dengan menggunakan *clustering* diharapkan dapat memberikan prediksi pengelompokan pelanggan tersebut.

Salah satu metode *clustering* yang dapat digunakan untuk mengelompokkan data adalah *Agglomerative Hierarchical Clustering (AHC)*. *Agglomerative Hierarchical Clustering (AHC)* merupakan suatu pengelompokan hirarki yang bersifat *bottom up* dimana keberadaan setiap titik data dalam klaster ditentukan oleh *proximity* antar titik tersebut. Metode *Agglomerative Hierarchical Clustering (AHC)* yang akan digunakan dalam Tugas Akhir ini ialah *Single linkage* (jarak terkecil), *complete linkage* (jarak terjauh) dan *average linkage* (jarak rata-rata). Metode ini berawal dari objek-objek individual yang paling mirip dikelompokkan dan kelompok-kelompok awal ini digabungkan sesuai dengan kemiripannya, berulang hingga menjadi satu *cluster* tunggal.

Dengan metode ini, data pelanggan akan direpresentasikan ke dalam bentuk hirarki klaster yang selanjutnya akan dikelompokkan ke dalam kelompok-kelompok yang berbeda. Selain itu, akan dihitung juga *cophenetic correlation coefficient* untuk mengukur seberapa baik sebuah *hierarchical clustering* memenuhi kesesuaian data. Kemudian dilakukan analisa dari hasil pengelompokan yang menggunakan metode *single linkage*, *complete linkage* dan *average linkage* untuk mengetahui hirarki yang terbaik.

1.2 Perumusan masalah

Berdasarkan latar belakang masalah, maka permasalahan yang akan diangkat dalam Tugas Akhir ini, yaitu :

1. Bagaimana mengimplementasikan *Agglomerative Hierarchical Clustering (AHC)* dengan pendekatan *single linkage*, *complete linkage* dan *average linkage* untuk mengelompokkan suatu data.
2. Bagaimana menentukan jarak antar setiap titik data.
3. Bagaimana mengukur kesesuaian data hasil *hierarchical clustering* dengan metode *cophenetic distance* untuk memperoleh hirarki yang terbaik.

Dalam Tugas Akhir ini ada beberapa batasan masalah yaitu :

1. Data yang akan digunakan sebagai studi kasus adalah data tagihan pelanggan layanan PT.Telkom.
2. Data yang akan digunakan dalam format MS.Excel dengan tipe *.csv.

1.3 Tujuan

Secara umum tujuan yang ingin dicapai dalam Tugas Akhir ini adalah :

1. Mengimplementasikan metode *Agglomerative Hierarchical Clustering (AHC)* untuk pengelompokan data dalam sebuah perangkat lunak.
2. Menerapkan metode *data mining*, *Agglomerative Hierarchical Clustering (AHC)* untuk membentuk hirarki dari data .
3. Memberikan hasil pengelompokan data menggunakan metode *Agglomerative Hierarchical Clustering (AHC)* dengan pendekatan *single linkage*, *complete linkage* dan *average linkage* serta analisis hasil hirarkinya dengan *cophenetic distance*.

1.4 Metodologi Penyelesaian Masalah

Metode penyelesaian masalah yang dilakukan dalam Tugas Akhir ini mencakup hal-hal berikut :

1. Mencari dan mengumpulkan bahan-bahan literatur yang berhubungan dengan permasalahan ini, meliputi : *Data Mining*, *Clustering*, *Agglomerative Hierarchical Clustering (AHC)*, *single linkage*, *complete linkage*, *average linkage*, *cophenetic distance* dan pengukuran evaluasi.
2. Studi literature tentang *Data Mining*, *Clustering*, *Agglomerative Hierarchical Clustering (AHC)*, *single linkage*, *complete linkage*, *average linkage*, *cophenetic distance* dan hal-hal lain yang mendukung pendalaman materi.
3. Melakukan pencarian data yang akan dikelompokkan.
4. Merancang aplikasi untuk melakukan pengelompokan data dan mengimplementasikannya ke dalam perangkat lunak.
5. Melakukan pengujian sistem dengan menggunakan data yang diperoleh.
6. Melakukan analisis hasil pengelompokan data.
7. Menentukan kesimpulan dari hasil implementasi dan analisis.
8. Penyusunan laporan Tugas Akhir.

1.5 Sistematika Penulisan

Penulisan Tugas Akhir ini dibagi dalam lima bab, yang terdiri atas :

- Bab 1 Pendahuluan
Menjelaskan mengenai latar belakang dari pembuatan Tugas Akhir ini, rumusan masalah yang akan dianalisa, batasan dari masalah yang timbul, tujuan yang ingin dicapai dan penentuan metodologi penyelesaian masalah dari sistem yang akan dibuat serta sistematika pembahasan.
- Bab 2 Landasan Teori
Mengemukakan berbagai teori dasar yang mendukung Tugas Akhir ini, antara lain mengenai *data mining*, *clustering*, *agglomerative hierarchical clustering*, dan *cophenetic distance*.
- Bab 3 Analisa dan Perancangan Sistem
Membahas tentang analisis dan perancangan awal sistem yang akan dibangun dengan tujuan untuk memahami secara jelas proses yang dilakukan pada sistem dalam bentuk Diagram Aliran Data (DAD).
- Bab 4 Implementasi dan Pengujian
Membahas kebutuhan perangkat lunak dan perangkat keras yang digunakan untuk merealisasikan sistem, membahas scenario pengujian perangkat lunak, hasil uji coba dan analisa dari hasil yang diperoleh.
- Bab 5 Kesimpulan dan Saran
Berisi kesimpulan dan saran terhadap pengembangan dari penelitian Tugas Akhir ini selanjutnya.