

BAB I

PENDAHULUAN

1.1 Latar Belakang

Teknologi pengumpulan dan penyimpanan data telah memudahkan organisasi untuk mengumpulkan sejumlah data berukuran besar sehingga menghasilkan gudang data. Ekstraksi informasi yang berguna dari gunung data menjadi pekerjaan yang cukup menantang. Sering kali alat dan teknik analisis data tradisional tidak dapat digunakan dalam mengekstrak informasi dari data berukuran besar. Data mining telah banyak diaplikasikan dalam berbagai bidang, diantaranya dalam bidang bisnis dan kedokteran. Dalam bidang bisnis, teknik data mining digunakan untuk mendukung cakupan yang luas dari aplikasi-aplikasi bisnis intelligent seperti customer profiling, targeted marketing, workflow management, store layout dan fraud detection.

Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. Definisi lain diantaranya adalah pembelajaran berbasis induksi (*induction-based learning*) adalah proses pembentukan definisi-definisi konsep umum yang dilakukan dengan cara mengobservasi contoh-contoh spesifik dari konsep-konsep yang akan dipelajari. Knowledge Discovery in Databases (KDD) adalah penerapan metode saintifik pada data mining (Widayu *et al.*, 2017).

PT. *Cipta Hydropower Abadi* merupakan salah satu pusat pembelanjaan yang bergerak di bidang peralatan hydraulic dan lubrikasi industri alat berat, pertambangan, perkapalan, pertanian dan perkebunan. Di mana PT. *Cipta Hydropower Abadi* ini setiap harinya harus memenuhi kebutuhan klien. Untuk dapat melakukan hal tersebut, maka membutuhkan sumber informasi untuk dapat dianalisis lebih lanjut. Sumber informasi tersebut yaitu pihak penjualan PT. *Cipta Hydropower Abadi*. Pada PT. *Cipta Hydropower Abadi*, terdapat beberapa permasalahan yang kerap muncul mengenai penjualan yaitu sulitnya dalam rekapitulasi dan mengambil kesimpulan dari keseluruhan data penjualan yang ada. Padahal agar dapat memperbaiki penjualan untuk kedepannya, memerlukan informasi dari penjualan pada masa sebelumnya.

Pada penelitian ini peneliti ingin melakukan perbandingan dalam pengolahan data penjualan yang dimiliki oleh PT. *Cipta Hydropower Abadi*, data ini merupakan data penjualan yang telah direkapitulasi oleh pihak perusahaan, perusahaan ingin melihat pengelompokan data atau pola *cluster* terhadap data penjualan yang dimiliki dan melakukan perbandingan antara dua metode *cluster* yaitu Metode *K-Means* Dan *Hierarchical Clustering*.

Adapun algoritma yang digunakan dalam proses integrasi dokumen ini adalah *agglomerative hierarchical clustering* dengan metode *complete linkage*. *Agglomerative hierarchical clustering* adalah suatu metode *hierarchical clustering* yang bersifat *bottom-up* yaitu menggabungkan n buah *cluster* (beberapa dokumen) menjadi satu *cluster tunggal* (sebuah dokumen hasil integrasi). *Agglomerative hierarchical clustering* merupakan metode yang

umum digunakan dalam clustering dokumen dan memiliki beberapa kelebihan, antara lain: tidak memperhitungkan initial centroid sehingga tepat digunakan dalam proses pengelompokan dokumen dan *kinerja information retrieval* berbasis *hierarchical clustering* memiliki hasil yang lebih baik jika dibandingkan dengan metode *partitional clustering*(Pradnyana, 2012).

Sedangkan metode K-Mean merupakan metode Analisa kelompok yang diarahkan pada pemartisian. N Obyek pengamatan kedalam K Kelomok atau disebut sebagai cluster, dimana setiap obyek pengamatan memiliki sebuah kelompok dengan rata-rata atau mean. K-mean merupakan salah satu metode pengelompokkan data sekatan (nonhierarki) yang berusaha mempartisi data yang ada kedalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data yang berkarakter berbeda di kelompokkan ke dalam kelompok yang lain(Nalendra *et al.*, 2020).

Guna untuk mengetahui hasil *cluster* dari perbandingan metode *K-Means* Dan *Hierarchical Clustering*, peneliti menggunakan alat bantu *rapidminer*, yang dimana didalam alat bantu tersebut sudah ada operator data *K-Means* Dan *Hierarchical Clustering*, sehingga dengan *tools* tersebut akan didapat hasil dari perbandingan dari penggunaan *K-Means* Dan *Hierarchical Clustering*.

Dari latar belakang diatas, maka akan dibahas lebih lanjut dalam tugas akhir dengan judul **“PERBANDINGAN METODE K-MEANS DAN HIERARCHICAL CLUSTERING PT. CIPTA HYDROPOWER ABADI DATA PENJUALAN DARI 2017-2019”**

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang dapat dirumuskan masalah yang akan dibahas sebagai berikut:

1. Bagaimana menerapkan metode *K-Means* Dan *Hierarchical Clustering* dalam melakukan cluster data penjualan perusahaan PT. Cipta Hydropower?
2. Bagaimana melakukan perbandingan metode *K-Means* Dan *Hierarchical Clustering* untuk melakukan *cluster* dengan bantuan *tools Rrapidminer*?

1.3 Batasan Masalah

Agar pembahasan terfokus pada tujuan yang ingin dicapai maka masalah-masalah yang dibahas akan dibatasi pada hal-hal berikut:

1. Data dan informasi yang ditampilkan hanya berupa data-data jumlah penjualan perusahaan dari 2017-2019.
2. Membuat suatu pengolahan data untuk mengelompokan data penjualan dengan metode *K-Means* Dan *Hierarchical Clustering*.
3. Melakukan perbandingan hasil cluster antara metode *K-Means* Dan *Hierarchical Clustering*
4. Alat bantu yang digunakan dalam menerapkan metode *K-Means* Dan *Hierarchical Clustering* adalah *rapidminer*.

1.4 Tujuan Penelitian

Berikut merupakan tujuan dari penelitian yang akan dijalankan oleh peneliti, sebagai berikut ini:

1. Melakukan perbandingan metode *K-Means* Dan *Hierarchical Clustering* untuk *cluster data* penjualan perusahaan PT. Cipta Hydropower.
2. Melakukan pengelompokan data penjualan yang terjadi antara 2017-2019 pada perusahaan PT. Cipta Hydropower.

1.5 Manfaat Penelitian

Berikut merupakan manfaat dari penelitian yang akan dijalankan oleh peneliti, sebagai berikut ini:

1. Diharapkan dapat menentukan metode yang lebih cocok untuk data penjualan *Pt Cipta Hydropower Abadi*.
2. Diharapkan dapat membantu pihak *Pt Cipta Hydropower Abadi* dalam perbandingan metode *K-Means* Dan *Hierarchical Clustering* untuk *cluster data* penjualan.

1.6 Sistematika Penulisan

Sistematika ini secara garis besar dapat memberikan gambaran isi, yang berupa susunan bab dari hasil penelitian.

BAB I PENDAHULUAN

Dalam bab ini penulis menguraikan Latar Belakang penelitian, batasan masalah, perumusan masalah penelitian, tujuan dan manfaat penelitian skripsi, metodologi penelitian, serta sistematika penelitian.

BAB II LANDASAN TEORI

Membahas tentang teori yang mendukung terhadap masalah yang dibahas dalam penyusunan penelitian skripsi ini yang terdiri dari pengertian penerapan KDD, sampai teori metode *K-Means* Dan *Hierarchical Clustering*.

BAB III ANALIS DAN PERANCANGAN

Pada bab ini menjelaskan mengenai Metode pengumpulan data atau Metode penelitian, Lokasi, Waktu, Teknik Analisis serta Pemodelan sistem dalam membangun sistem dan perancangan sistem yang berguna sebagai acuan dalam pembangunan sebuah sistem

BAB IV HASIL DAN PEMBAHASAN

Bab ini menguraikan tentang implementasi sistem secara detail. Sesuai dengan rancangan dan berdasarkan komponen atau *tools* serta bahasa pemrograman yang dipakai.

BAB V KESIMPULAN DAN SARAN

Bab ini menguraikan tentang beberapa kesimpulan dari permasalahan pada bab-bab sebelumnya serta memberikan pengembangan saran untuk pengembangan sistem selanjutnya