



BAB II

TINJAUAN PUSTAKA

2.1. Data Mining

Data mining adalah suatu metode pengolahan data untuk menemukan pola yang tersembunyi dari data tersebut. Hasil dari pengolahan data dengan metode data mining ini dapat digunakan untuk mengambil keputusan di masa depan. Data mining ini juga dikenal dengan istilah pattern recognition (Santosa, 2007).

Data mining merupakan metode pengolahan data berskala besar oleh karena itu data mining ini memiliki peranan penting dalam bidang industri, keuangan, cuaca, ilmu dan teknologi. Secara umum kajian data mining membahas metode-metode seperti, clustering, klasifikasi, regresi, seleksi variable, dan market basket analisis (Santosa, 2007).

Data mining merupakan sebuah analisa dari observasi data dalam jumlah besar untuk menemukan hubungan yang tidak diketahu sebelumnya dan metode baru untuk meringkas data agar mudah dipahami serta kegunaannya untuk pemilik data (David Hand et al, 2001)

2.2 Clustering

Clustering merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (similarity) antara satu data dengan data yang lain. Clustering merupakan salah satu metode data mining yang bersifat tanpa arahan (unsupervised), maksudnya metode ini diterapkan tanpa adanya latihan (training) dan tanpa ada guru (teacher) serta tidak memerlukan target output. Dalam data mining ada dua jenis metode clustering yang digunakan dalam pengelompokan data, yaitu hierarchical clustering dan non-hierarchical clustering (Santosa, 2007).

Hierarchical clustering adalah suatu metode pengelompokan data yang dimulai dengan mengelompokkan dua atau lebih objek yang memiliki kesamaan paling dekat. Kemudian proses diteruskan ke objek lain yang memiliki kedekatan kedua. Demikian seterusnya sehingga cluster akan membentuk semacam pohon dimana ada hierarki (tingkatan) yang jelas antar objek, dari yang paling mirip



sampai yang paling tidak mirip. Secara logika semua objek pada akhirnya hanya akan membentuk sebuah cluster. Dendogram biasanya digunakan untuk membantu memperjelas proses hierarki tersebut (Santoso, 2010).

Berbeda dengan metode hierarchical clustering, metode non-hierarchical clustering justru dimulai dengan menentukan terlebih dahulu jumlah cluster yang diinginkan (dua cluster, tiga cluster, atau lain sebagainya). Setelah jumlah cluster diketahui, baru proses cluster dilakukan tanpa mengikuti proses hierarki. Metode ini biasa disebut dengan K-Means Clustering (Santoso, 2010).

2.3. Algoritma K-Means

K-means clustering merupakan salah satu metode data clustering non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih cluster/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu cluster/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan cluster/kelompok yang lain sehingga data yang berada dalam satu cluster/kelompok memiliki tingkat variasi yang kecil (Agusta, 2007).

Menurut Santosa (2007), langkah-langkah melakukan clustering dengan metode K-Means adalah sebagai berikut:

- a. Pilih jumlah cluster k .
- b. Inisialisasi k pusat cluster ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara using lower and upper bounds. Pusat-pusat cluster diberiduberi nilai awal dengan angka-angka terendah dan tertinggi
- c. Alokasikan semua data/ objek ke cluster terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke cluster tertentu ditentukan jarak antara data dengan pusat cluster. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat cluster. Jarak paling antara satu data dengan satu cluster tertentu akan menentukan suatu data masuk dalam cluster mana. Untuk menghitung jarak semua data ke setiap titik pusat cluster dapat menggunakan teori jarak Euclidean yang dirumuskan sebagai berikut:



$$D(i,j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \quad (1)$$

Dimana:

$D(i,j)$ = Jarak data ke i ke pusat cluster j

X_{ki} = Data ke i pada atribut data ke k

X_{kj} = Titik pusat ke j pada atribut ke k

- d. Hitung kembali pusat cluster dengan keanggotaan cluster yang sekarang. Pusat cluster adalah rata-rata dari semua data/ objek dalam cluster tertentu. Jika dikehendaki bisa juga menggunakan median dari cluster tersebut. Jadi rata-rata (mean) bukan satu-satunya ukuran yang bisa dipakai.
- e. Tugaskan lagi setiap objek memakai pusat cluster yang baru. Jika pusat cluster tidak berubah lagi maka proses clustering selesai. Atau, kembali ke langkah nomor 3 sampai pusat cluster tidak berubah lagi.

2.4 Usecase Diagram

Use Case mendeskripsikan sebuah interaksi antara satu atau lebih aktor dengan sistem informasi yang akan dibuat (Rosa dan Shalahuddin, 2013). Simbol-simbol *Use Case Diagram* dapat dilihat pada Tabel 2.1

Tabel 2.1 Simbol Use Case Diagram

No	Simbol	Nama	Keterangan
1.		Aktor/ <i>Actor</i>	Aktor adalah pengguna sistem. aktor tidak terbatas hanya manusia saja, jika sebuah sistem berkomunikasi dengan aplikasi lain dan membutuhkan input atau memberikan output, maka aplikasi tersebut juga bisa dianggap sebagai aktor.



2.		<i>Use case</i>	<i>Use case</i> digambarkan sebagai lingkaran elips dengan nama <i>use case</i> dituliskan didalam elips tersebut, digunakan sebagai unit-unit yang saling bertukar pesan antar unit / aktor.
3.		<i>Association /</i> Asosiasi	Asosiasi digunakan untuk menghubungkan <i>actor</i> dengan <i>use case</i> . Asosiasi digambarkan dengan sebuah garis yang menghubungkan antara <i>Actor</i> dengan <i>Use Case</i> .
4.		<i>Generalization /</i> Generalisasi	Hubungan dimana objek anak (<i>descendent</i>) berbagi perilaku dan struktur data dari objek yang ada di atasnya objek induk (<i>ancestor</i>).
5.		<i>Include</i>	Menspesifikasikan bahwa <i>use case</i> sumber secara <i>eksplisit</i> .
6.		<i>Extend /</i> Ekstensi	Menspesifikasikan bahwa <i>use case</i> target memperluas perilaku dari <i>use case</i> sumber pada suatu titik yang diberikan

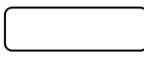
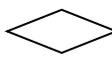
(Sumber: Sukamto dan Shalahuddin dalam Aprianti (2016 : 23))

2.5 Activity Diagram

Menurut Sukamto dan Shalahuddin dalam Aprianti (2016 : 23) *activity diagram* menggambarkan *workflow* (aliran kerja) atau aktivitas dari sebuah sistem atau proses bisnis. Diagram aktivitas menggambarkan aktivitas –aktivitas sistem bukan apa yang dilakukan aktor.



Tabel 2.2 Simbol Activity Diagram

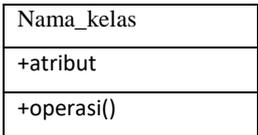
No	Simbol	Nama	Keterangan
1.		<i>Start state</i>	Titik awal atau permulaan
2.		<i>End state</i>	Titik akhir atau akhir dari aktivitas
3.		<i>Activity</i>	<i>Activity</i> atau aktivitas yang dilakukan oleh actor
4.		<i>Decision</i>	Pilihan untuk mengambil keputusan
5.		<i>Join</i>	Asosiasi penggabungan dimana lebih dari satu aktivitas digabungkan menjadi satu.

Sumber: Sukamto dan Shalahuddin dalam Aprianti (2016 : 23)

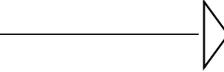
2.6 Class Diagram

Class diagram menggambarkan struktur sistem dari segi pendefinisian kelas-kelas yang akan di buat untuk membangun sistem. Kelas memiliki apa yang disebut atribut dan metode atau operasi. Simbol-simbol yang ada pada *Class* diagram ditunjukkan oleh Tabel 2.3

Tabel 2.3 Simbol Class Diagram

Simbol	Nama	Keterangan
	<i>Class/</i> Kelas	Kelas pada struktur system
	Antarmuka <i>/interface</i>	Sama dengan konsep <i>interface</i> dalam pemrograman berorientasi objek.



	Asosiasi/ <i>Association</i>	Relasi antarkelas dengan makna umum, asosiasi biasanya disertai dengan <i>multiplicity</i> .
	Asosiasi berarah/ <i>directed association</i>	Relasi antarkelas dengan makna kelas yang satu digunakan oleh kelas yang lain, asosiasi biasanya juga disertai dengan <i>multiplicity</i> .
	Generalisasi	Relasi antar kelas dengan makna generalisasi- spesialisasi (umum khusus).
	Kebergantungan/ <i>Dependency</i>	Relasi antar kelas dengan makna kebergantungan antar kelas.
	Agregasi/ <i>aggregation</i>	Relasi antar kelas dengan makna semua-bagian.

(Sumber: Sukamto dan Shalahuddin dalam Aprianti (2016 : 22))

2.7 Sequence Diagram

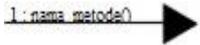
Menurut Pratama (2013:48), “*Sequence diagram* menggambarkan *sequence* (aliran) pengiriman pesan (*message*) yang terjadi di aplikasi, sebagai bentuk interaksi dengan pengguna (*user*)”.

Adapun simbol-simbol yang digunakan dalam *sequence diagram* adalah sebagai berikut:

Tabel 2.4 Simbol *Sequence Diagram*

No.	Gambar	Nama	Keterangan
1		Aktor	Orang, proses, atau sistem lain yang berinteraksi dengan sistem informasi yang akan dibuat di luar sistem informasi yang akan dibuat itu sendiri.



2		Lifeline	Menyatakan kehidupan suatu objek.
3		Waktu Aktif	Menyatakan objek dalam keadaan aktif dan berinteraksi, semua yang terhubung dengan waktu aktif ini adalah sebuah tahapan yang dilakukan di dalamnya.
4		Pesan tipe call	Menyatakan suatu objek memanggil operasi / metode yang ada pada objek lain atau dirinya sendiri.

Sumber: Shalahuddin et al. (2013:162)

2.8 Penelitian Terdahulu

Menurut penelitian Johan (2013) dalam jurnal yang berjudul Implementasi Algoritma K-Means Clustering untuk Strategi Marketing President University bahwa dengan menggunakan metode Algoritma K-Means, data- data yang telah didapatkan dapat dikelompokkan kedalam beberapa cluster berdasarkan kemiripan dari data-data tersebut, sehingga data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu cluster dan yang memiliki karakteristik yang berbeda dikelompokkan dalam cluster yang lain yang memiliki karakteristik yang sama.

Menurut Muningsih dan Kiswati (2015) dalam jurnal yang berjudul Penerapan Metode K-Means untuk Clustering Produk Online Shop dalam Penentuan Stok Barang bahwa penelitian tersebut bisa dimanfaatkan untuk pelaku bisnis lainnya bukan hanya online shop yang membutuhkan informasi penentuan stok barang secara lebih cepat dan akurat.

Menurut Nasari dan Sianturi (2016) luasnya wilayah kabupaten langkat memungkinkan perlunya sebuah pengelompokan wilayah penyebaran diare, pengelompokan wilayah penyebaran diare akan menghasilkan titik-titik pusat penyebaran diare yang merata di hampir seluruh kawasan indonesia, salah satunya di kabupaten langkat.

Menurut Asroni dan Adrian (2015) dalam jurnal yang berjudul Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik



Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang bahwa dengan menerapkan metode K-Means ini dapat memberikan rekomendasi mahasiswa yang layak maju sebagai peserta event Cyber jawara.

Dalam penelitian yang dibuat oleh Tulus dan Hendry (2014) dalam jurnal yang berjudul Perancangan Clustering Data Menggunakan Algoritma K-Means Berbasis Heatmap clustering yang dilakukan akan ditampilkan dalam sebuah aplikasi web. Daerah yang berpotensi untuk menghasilkan hasil tambang yang banyak akan ditampilkan Heatmap dengan warna merah dan yang kurang akan dengan warna biru.