

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Sistem**

Menurut Gondodiyoto (2007), sistem adalah merupakan suatu kesatuan yang terdiri dari komponen-komponen atau sub sistem yang berorientasi untuk mencapai suatu tujuan tertentu.

Berdasarkan batasan pengertian tersebut, sistem mempunyai karakteristik sebagai berikut:

1. Sistem adalah kumpulan elemen-elemen atau sumberdaya yang saling berkaitan secara terpadu, terintegrasi dalam suatu hubungan hirarkis.
2. Sistem memiliki sasaran yang akan dicapai. Setiap sistem berusaha mencapai satu atau lebih sasaran yang merupakan arah, yang merupakan kekuatan yang memberikan arah suatu sistem.
3. Konstruksi sistem terdiri dari: Masukan-Proses-Keluaran.

Masukan merupakan semua arus berwujud atau tidak berwujud yang masuk ke sistem.

Keluaran merupakan semua arus keluar atau akibat yang dihasilkan. Proses terdiri dari metode yang digunakan untuk mengubah masukan menjadi keluaran.

4. Sistem memiliki pengguna.

Setiap sistem harus mengarahkan sub sistemnya agar dapat mencapai sasaran.

Sasaran sistem sebagai ukuran penentu keberhasilan suatu sistem.

5. Sistem memiliki keterbatasan.

6. Sistem memiliki sub sistem yang membentuk suatu jaringan terpadu.
7. Sistem memerlukan pengendalian.

## 2.2 Informasi

Informasi adalah data yang sudah diolah menjadi bentuk yang lebih berguna dan lebih berarti (bermanfaat) bagi penerimanya, menggambarkan suatu kejadian dan kesatuan nyata yang dapat dipahami dan dapat digunakan untuk pengambilan keputusan, sekarang maupun masa depan (Gondodiyoto, 2007).

Sumber dari informasi adalah data. Data merupakan bentuk jamak dari bentuk tunggal datum atau data item. Data sebagai input perlu diolah oleh suatu sistem pengolahan data agar dapat menjadi output, yaitu informasi yang lebih berguna bagi pemakainya. Dari uraian tersebut dapat dikatakan bahwa :

1. Informasi adalah data yang diolah menjadi bentuk yang lebih berguna, lebih bermanfaat dan lebih berarti bagi penggunanya.
2. Data menggambarkan suatu kejadian-kejadian, data dinyatakan sebagai simbol-simbol, gambar-gambar, kata-kata, angka-angka, atau huruf-huruf yang menunjukkan suatu ide, obyek, kondisi atau situasi tertentu.
3. Informasi digunakan untuk pengambilan keputusan. Bagi manajemen suatu organisasi, informasi berguna untuk membantu dalam pengambilan keputusan yang menentukan keberhasilan atau kesuksesan organisasi pada masa yang akan datang.

### 2.3 Sistem Informasi

Sistem adalah kumpulan dari elemen-elemen yang berinteraksi untuk mencapai tujuan tertentu (Jogiyanto, 1995). Informasi adalah terdiri dari data yang telah diambil kembali dan diolah atau sebaliknya dan digunakan untuk tujuan informatif atau kesimpulan, argumentasi, atau sebagai dasar untuk peramalan atau pengambilan keputusan. Sistem Informasi adalah kombinasi antara prosedur kerja, informasi, orang dan teknologi informasi yang diorganisasikan untuk mencapai tujuan dalam sebuah organisasi.

Sistem informasi adalah sekumpulan komponen pembentuk sistem yang mempunyai keterkaitan antara satu komponen dengan komponen lainnya yang bertujuan menghasilkan suatu informasi dalam suatu bidang tertentu. Dalam sistem informasi diperlukannya klasifikasi alur informasi, hal ini disebabkan keanekaragaman kebutuhan akan suatu informasi oleh pengguna informasi.

### 2.4 Sistem Pendukung Keputusan

Sistem pendukung keputusan mulai dikembangkan pada tahun 1960-an, tetapi istilah Sistem Pendukung Keputusan itu sendiri baru muncul pada tahun 1971, yang diciptakan oleh G. Anthony Gorry dan Michael S. Scott Morton, keduanya adalah profesor di MIT. Sistem pendukung keputusan pasti tidak terlepas dari proses pengambilan keputusan itu sendiri. Pada dasarnya, proses pengambilan keputusan terdiri dari 3 fase proses: *intelligence*, *design*, dan *choice*. *Intelligence* yaitu pencarian kondisi-kondisi yang dapat menghasilkan keputusan. *Design* yaitu menemukan, mengembangkan, dan menganalisis materi-

materi yang mungkin untuk dikerjakan, sedangkan *choice* yaitu pemilihan dari materi-materi yang tersedia, mana yang akan dikerjakan.

Proses-proses yang terjadi pada kerangka kerja *Decision Support* dibedakan atas:

1. Terstruktur, mengacu pada permasalahan rutin dan berulang untuk solusi standar yang ada.
2. Tak terstruktur, adalah “fuzzy”, permasalahan kompleks dimana tak ada solusi serta merta. Masalah yang tak terstruktur adalah tak adanya 3 fase proses yang terstruktur.
3. Semi terstruktur, terdapat beberapa keputusan terstruktur, tetapi tak semuanya dari fase-fase yang ada.

Secara umum, sistem pendukung keputusan (SPK) adalah sistem berbasis komputer yang interaktif, yang membantu pengambil keputusan memanfaatkan data dan model untuk menyelesaikan masalah-masalah yang tak terstruktur dan semi terstruktur. SPK ini mendayagunakan *resources* individu-individu secara intelek dengan kemampuan komputer untuk meningkatkan kualitas keputusan. Sebenarnya definisi awalnya, SPK adalah sistem berbasis model yang terdiri dari prosedur-prosedur dalam pemrosesan data dan pertimbangannya untuk membantu manajer dalam mengambil keputusan. Agar berhasil mencapai tujuannya maka sistem tersebut harus sederhana, mudah untuk dikontrol, mudah beradaptasi, lengkap pada hal-hal penting, dan mudah berkomunikasi dengannya.

Ada juga definisi yang menyatakan bahwa SPK adalah sistem berbasis komputer yang terdiri 3 komponen interaktif: (1) sistem bahasa-mekanisme yang

menyediakan komunikasi diantara user dan pelbagai komponen dalam SPK, (2) *knowledge* sistem-penyimpanan *knowledge* domain permasalahan yang ditanamkan dalam SPK, baik sebagai data ataupun prosedur, dan (3) sistem pemrosesan permasalahan-link diantara dua komponen, mengandung satu atau lebih kemampuan memanipulasi masalah yang dibutuhkan untuk pengambilan keputusan. Definisi terakhir, SPK mengacu pada situasi dimana sistem final dapat dikembangkan hanya melalui *adaptive process* pembelajaran dan evolusi.

## 2.5 Data Mining

Faktor penentu bagi bentuk usaha atau bisnis apapun pada masa sekarang ini adalah kemampuan untuk menggunakan informasi secara seefektif mungkin. Penggunaan data yang strategis ini bisa diakibatkan oleh kesempatan-kesempatan yang dihasilkan atau ditimbulkan karena penemuan fakta-fakta sangat berharga yang cukup sering, tersembunyi dan tidak terdeteksi sebelumnya mengenai konsumen, *retailer* dan *supplier*, tren – tren bisnis, dan faktor – faktor petunjuk yang lain (Berson, 1997). Menurut Kamber (2007) secara sederhana data mining mengacu kepada mengekstrak atau “menambang” pengetahuan dari sekumpulan besar data. Menambang dalam hal ini bukan diibaratkan sebagai menambang emas atau menambang pasir, tetapi lebih diibaratkan sebagai “*knowledge mining from data*” atau lebih ringkasnya menambang pengetahuan. Pengertian lain data mining juga dapat berarti prose untuk mempekerjakan satu atau lebih teknik pembelajaran terkomputerisasi untuk mengotomasi analisa dan mengekstrak pengetahuan dari data didalam database (Roger and Geatz, 2003).

Data Mining mempunyai langkah-langkah yang penting dalam proses *knowledge discovery*. Prosesnya disusun berurutan sebagai berikut :

1. Tahap seleksi (*Selection*). Tahap memilih atau memilah-milah data berdasarkan kepada sejumlah kriteria.
2. Tahap proses awal (*Preprocessing*). Tahap pembersihan data dimana informasi-informasi tertentu dipindahkan karena dianggap tidak penting dan bisa memperlambat *queries*.
3. Tahap transformasi (*Transformation*). Dalam tahap ini data tidak hanya dipindahkan saja tetapi diubah sehingga dapat ditambahkan lapisan-lapisan seperti contoh lapisan demografis yang biasanya digunakan dalam penelitian.
4. Tahap data mining (*Data Mining*). Tahap ini mengenai ekstraksi pola dari data.
5. Tahap implementasi dan evaluasi (*Interpretation and evaluation*). Pola-pola yang diidentifikasi oleh sistem yang ada diinterpretasikan ke dalam pengetahuan yang kemudian bisa digunakan untuk mendukung pengambilan keputusan.

## 2.6 Clustering

*Cluster* adalah suatu kumpulan dari entitas yang hampir sama (Everit, 1993). Pengertian lain menurut Kamber (2007), *cluster* adalah kumpulan dari objek yang mirip dengan objek lainnya dan berada pada kelompok yang sama. Sedangkan proses untuk mengelompokkan data baik itu bersifat fisik atau abstrak kedalam suatu kelompok atau kelas yang memiliki kesamaan sifat disebut *clustering*.

*Clustering* dikategorikan kedalam teknik *Undirect Knowledge* atau *Unsupervised Learning* karena tidak membutuhkan proses pelatihan untuk klasifikasi awal data dalam masing-masing kelompok atau *cluster*. Tujuan utama *clustering* adalah untuk menemukan atau mencari pola yang bermanfaat atau berguna pada suatu database, kemudian merangkumnya dan membuat lebih mudah untuk dipahami.

Dalam melakukan proses analisa terhadap *cluster-cluster* yang telah terbentuk dan pencarian pengetahuan dengan metode tertentu disebut *cluster analyse* (Kamber, 2007).

## 2.7 *K-Means Clustering*

Metode *K-Means* adalah metode pengelompokan data dengan mengambil parameter sejumlah  $k$  *cluster*, dan mempartisi data kedalam *cluster* tersebut, dengan berpatokan pada kemiripan antar data dalam satu *cluster* dan ketidakmiripan di antar *cluster* yang berbeda, pusat dari *cluster* adalah rata-rata dari nilai anggota *cluster* yang disebut *centroid* atau *center of gravity* (Kamber, 2007). Selain itu *K-Means* melakukan pengelompokan dengan meminimalkan jumlah kuadrat dari jarak (*distance*) antara data dengan *centroid cluster* yang cocok (Teknomo, 2006).

Algoritma *K-Means* adalah algoritma *partitional (Non Hierarchical) clustering* yang paling umum digunakan, hal ini disebabkan karena kesederhanaan dan keefisienan algoritma tersebut. Algoritma *K-Means* mempartisi atau membagi sekumpulan data ke dalam sejumlah *cluster*. Algoritma *pseudocode* dari perhitungan *k-means* (Kamber, 2007) yaitu :

*Input* :  $k$  : jumlah *cluster* yang akan di bentuk

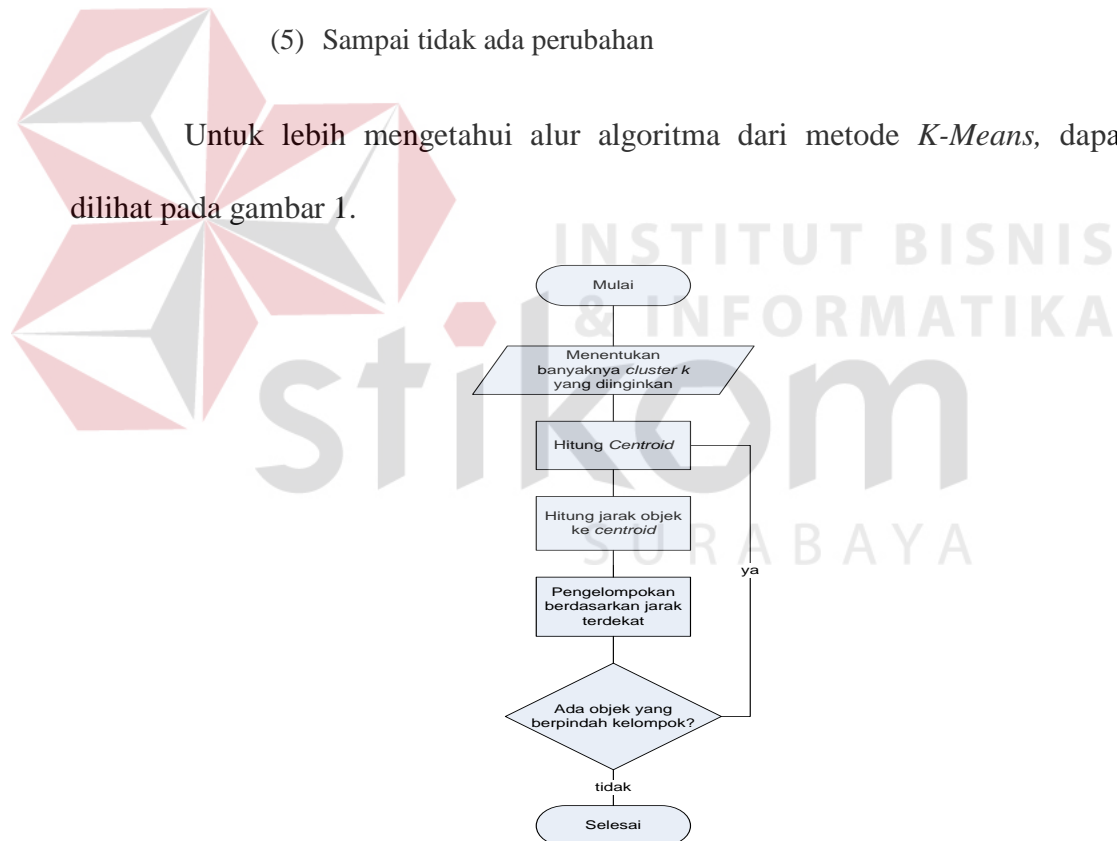
$D$  : *dataset* yang berisikan  $n$  objek

*Output* : *cluster* sebanyak  $k$

*Method* :

- (1) Pilih secara random objek  $k$  dari  $D$  sebagai *centroid* awal
- (2) Ulangi
- (3) Masukkan objek ke cluster yang beranggotakan objek yang mirip, sesuai nilai rata-rata dari objek yang ada pada cluster.
- (4) Perbaharui nilai rata-rata cluster.
- (5) Sampai tidak ada perubahan

Untuk lebih mengetahui alur algoritma dari metode *K-Means*, dapat dilihat pada gambar 1.



Gambar 2.1 Flowchart Algoritma *Clustering K-Means* (Teknomo, 2006)

Pada gambar diatas, pertama user menentukan banyaknya jumlah *cluster* yang diinginkan, kemudian menghitung *centroid*. Pada iterasi awal, pemilihan *centroid* dilakukan secara acak pada objek. Langkah selanjutnya adalah



menghitung jarak setiap anggota *cluster* ke semua *centroid*. Lalu dilakukan pengelompokan berdasarkan jarak terdekat yang telah dihitung sebelumnya. Selanjutnya adalah menghitung kembali *centroid* yang baru. Titik *centroid* baru didapat dari rata-rata semua titik anggota *cluster*. Langkah-langkah diatas terus dilakukan hingga tidak ada satupun anggota *cluster* yang berpindah *cluster*.

## 2.8 Centroid

*Centroid* merupakan rata-rata dari nilai anggota *cluster* yang disebut juga *center of gravity* (Kamber, 2007). Sedangkan menurut Roger and Geatz, 2003, *centroid* merupakan pusat dari sebuah *cluster*. Di dalam perhitungan *k-means* akan terbentuk *cluster* atau kelompok-kelompok data sejumlah *k*. Nilai *k* merupakan nilai parameter yang telah di tentukan sebelumnya. Nilai *centroid* adalah penjumlahan dari titik koordinat anggota *cluster* dibagi jumlah anggota. Misalnya terdapat 2 data anggota pada suatu *cluster* yaitu  $A = ($