

Nama : Muh.Ikhsan  
NIM : H071191049

## **Resume 1 – Pengantar Data Mining**

### **A. Data**

Data adalah catatan atas kumpulan fakta. Data merupakan bentuk jamak dari datum, berasal dari bahasa Latin yang berarti "sesuatu yang diberikan". Dalam penggunaan sehari-hari data berarti suatu pernyataan yang diterima secara apa adanya. Pernyataan ini adalah hasil pengukuran atau pengamatan suatu variabel yang bentuknya dapat berupa angka, kata-kata, atau citra. Data yang berupa angka-angka disebut data kuantitatif, sedangkan data yang bukan berupa angka disebut data kualitatif. Berdasarkan cara memperolehnya, data kuantitatif terbagi atas data diskrit dan data kontinu. Data diskrit adalah data yang diperoleh dengan cara menghitung, sedangkan data kontinu adalah data yang diperoleh dengan cara mengukur.

### **B. Definisi Data Mining**

Secara sederhana data mining adalah istilah yang digunakan untuk menjelaskan proses pencarian atau penambangan knowledge dari data yang sangat besar. Sedangkan secara teori data mining adalah proses yang memanfaatkan proses-proses statistik, matematika dan kecerdasan buatan untuk mengekstrak dan mengidentifikasi informasi dan knowledge selanjutnya (atau pola-pola) yang berasal dari sekumpulan data yang sangat besar. Tujuan dari data mining yaitu untuk menggali informasi yang penting terdapat didalam data.

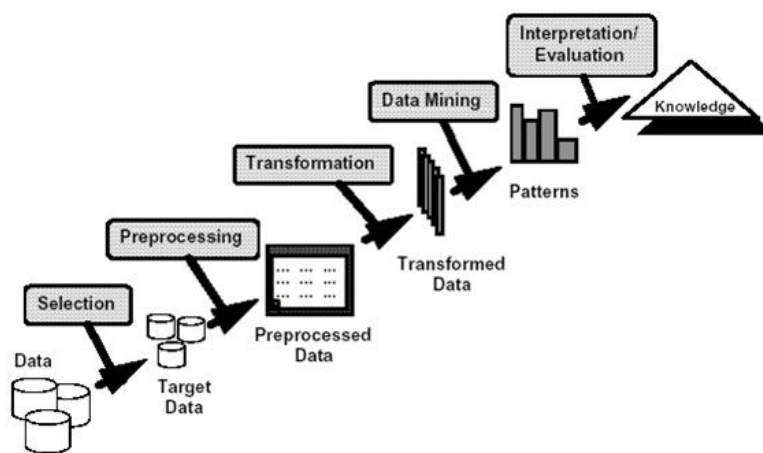
### **C. Manfaat Data Mining**

Data mining sangatlah berguna bagi pengguna baik pada level perusahaan atau individu. Penambangan data adalah proses mencari mencari pola dan atau tren yang tidak dapat ditemukan dengan menggunakan teknik analisis sederhana. Hal tersebut perlu menggunakan algoritma matematika yang kompleks untuk mempelajari data dan kemudian mengevaluasi kemungkinan kejadian di masa depan berdasarkan data historis.

#### D. Kegiatan (Task) dalam Data Mining

1. Klasifikasi (*Classification*) yaitu data yang sudah memiliki label dan biasanya label data berbentuk diskrit, digunakan untuk mengelompokkan data berdasarkan kelas-kelas yang sudah ada.
2. Klastering (*Clustering*) yaitu data yang tidak memiliki label atau kelas, digunakan untuk mengidentifikasi objek data yang mirip satu sama lain.
3. Regresi / Estimasi (*Regression*) yaitu data label berbentuk continue dan datanya linear
4. Asosiasi (*Association*) yaitu data analisis pasar, digunakan pada data market basket analisis. Asosiasi dilakukan dengan menghitung berapa kali dalam suatu set data yang mengandung dua atau beberapa item yang saling berhubungan.

#### E. Tahapan atau Langkah dalam Data Mining



1. Pembersihan data (untuk membuang data yang tidak konsisten dan noise)
2. Integrasi data (penggabungan data dari beberapa sumber)
3. Transformasi data (data diubah menjadi bentuk yang sesuai untuk di-mining)
4. Aplikasi teknik Data Mining, proses ekstraksi pola dari data yang ada
5. Evaluasi pola yang ditemukan (proses interpretasi pola menjadi pengetahuan yang dapat digunakan untuk mendukung pengambilan keputusan)
6. Presentasi pengetahuan (dengan teknik visualisasi)