Getting to Know Your Data & Knowledge Discovery in Database Process (KDD) in Data Mining

Pada artikel sebelumnya telah dibahas mengenai konsep dasar data mining. Sedikit mengingat kembali bahwa data mining adalah proses pencarian pengetahuan dan informasi penting dari sejumlah data yang besar dengan menggunakan pendekatan statistika, matematika, sampai memanfaaatkan teknologi artificial intelligence. Dikesempatan kali ini kita akan mengenal data dan beberapa istilah yaitu salah satunya adalah knowledge discovery in database (KDD).

* Getting to Know Your Data
* Data Object

Objek data terdiri dari kumpulan data. Objek data mewakili atau merepresentasikan sebuah entitas. Entitas dapat disebut sebagai objek yang mempresentasikan data, terdiri dari row (baris) dan column (kolom). Misalny, database penjualan memiliki memiliki tabel atau entitas pelanggan, barang, dan penjualan. Contoh lain database medis memiliki entitas pasien dan perawat. Objek data tersebut dideskripsikan oleh attribut yang ada pada entitas. Setiap entitas tersusun atas rows (data object) dan column (attributes).

* Attributes

Atribut atau dimensi/fitur/variabel merupakan kumpulan data yang mewakili karakteristik atau fitur dari objek data. Atribut data memiliki beberapa tipe/jenis, yaitu :

* Nominal adalah jenis atribut yang hanya membedakkan antara satu objek dan objek lain dengan nilai yang cukup. Atribut ini hanya melakukan perbandingan sama atau tidaknya suatu data. Contoh : nama orang, nomor KTP, warna mata dan lain-lain.
* Binary merupakan jenis atribut yang hanya berisi 2 kemungkinan mutlak, seperti 0 dan 1 atau biasa dikenal dengan data boolean.
* Numeric merupakan atribut yang bernilai bilangan bulat atau bilangan rill. Atribut numeric terdiri dari interval dan rasio.
* Ordinal adalah jenis atribut yang memberikan suatu informasi mengenai suatu tingkatan yang menjadi pembeda dari objek. Nilai ordinal berupa rangking-an yang dapat diurut berdasarkan tingkatan data. Contoh : nilai mahasiswa, sedang, rendah, tinggi.
* Knowledge Discovery in Database Process (KDD) adalah salah satu teknik yang digunakan di data mining. Adapun definisi lain dari KDD adalah proses menggunakan metode data mining untuk mencari informasi-informasi yang penting, proses menemukan pola, yang melibatkan algoritma dalam pengidentifikasiannya. Selanjutnya proses KDD dimulai beberapa step/tahap, yaitu data cleaning and integration, data selection and transformation, data mining, evaluation and presentation.
* Data Cleaning and Integration

Pembersihan data bertujuan untuk menghilangkan noise dan menghindari inkonsisten data. Proses ini disebut dengan data cleaning. Sedangkan data integration (penggabungan data) bertujuan untuk menggabungkan sekumpulan data dari berbagai sumber.

* Data Selection and Transformation

Data selection bertujuan untuk menyeleksi data relevan dari database yang akan digunakan oleh analisis data. Data transformation (transformasi data) memiliki tujuan mengubah data ke dalam bentuk yang mudah dibaca.

* Data Mining

Data mining merupakan proses penting dalam KDD untuk mengekstrak pola data.

* Evaluation and Presentation

Pattern Evaluation atau evaluasi pola berfungsi untuk mengidentifikasi pola berdasarkan karakteristik tertentu. Sedangkan knowledge presentation atau presentasi pengetahuan bertujuan untuk memvisualisasikan pengetahuan kepada pengguna.

Demikian arikel mengenai data mining yang dapat penulis sampaikan. Semoga menambah pengetahuan teman-teman mengenai topik diatas. Sekian terima kasih semoga bermanfaat.