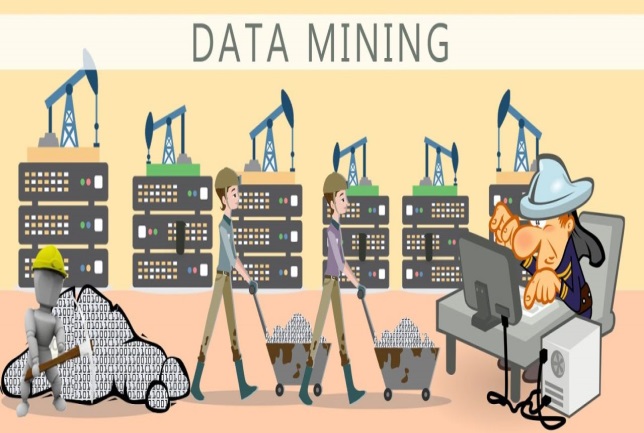
**Nama : Diaz Adha Asri Prakoso**

**NIM : 0102518007**

**DATA MINING**

**WHAT IS DATA MINING?**

Proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan algoritma untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai volume data yang besar. Salah satu ciri dari data minig adalah datanya sangat banyak dan tidak terstruktur.

**FUNCTION?**

Data mining mempunyai empat fungsi dasar yaitu:

* **Fungsi Prediksi (prediction)**

Proses untuk menemukan pola dari data dengan menggunakan beberapa variabel untuk memprediksikan variabel lain yang tidak diketahui jenis atau nilainya.

* **Fungsi Deskripsi (description)**

Proses untuk menemukan suatu karakteristik penting dari data dalam suatu basis data.

* **Fungsi Klasifikasi (classification)**

Klasifikasi merupakan suatu proses untuk menemukan model atau fungsi untuk menggambarkan class atau konsep dari suatu data. Proses yang digunakan untuk mendeskripsikan data yang penting serta dapat meramalkan kecenderungan data pada masa depan.

* **Fungsi Asosiasi (association)**

Proses ini digunakan untuk menemukan suatu hubungan yang terdapat pada nilai atribut dari sekumpulan data.

**EXAMPLE IN EVERYDAY LIFE?**

* **Bidang bisnis**

1. Mengetahui hilangnya pelanggan dikarenakan adanya pesaing.
2. Mengetahui item suatu produk yang memiliki kesamaan karakteristik.
3. Mengidentifikasi produk-produk yang sudah terjual dengan produk lainnya.
4. Memprediksi dari tingkat penjualan.
5. Menilai tingkat resiko dalam menentukan jumlah produksi pada suatu item.
6. Memprediksi perilaku bisnis dimasa depan.

* **Bidang pemasaran**

1. Mengidentifikasikan pembelian para konsumen, yang dapat mencari dan menemukan hubungan karakteristik demografi pelanggan
2. Memperkirakan tanggapan penawaran yang diberikan melalui surat.

* **Bidang perbankan**

1. Mendeteksi pola penyalahgunaan kartu kredit dan mengidentifikasi tingkat kesetiaan dari pelanggan dari instansi

* **Bidang asuransi**

1. Analisis klaim dalam memperkirakan pelanggan yang akan membeli produk terbarunya.

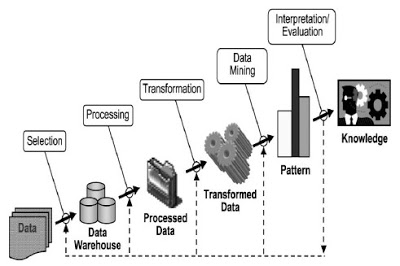
* **Bidang kesehatan**

1. Pendekatan yang biasa di lakukan dalam bidang kesehatan yaitu database multi dimensi, mesin pembelajaran, soft computing, statistik dan visualisasi data.
2. Memprediksi volume pasien dalam setiap kategori.
3. Menganalisis untuk mengidentifikasikan praktik yang terbaik untuk meningkatkan perawatan yang diberikan

* **Bidang EDM (Educational Data Mining)**

1. Mengidentifikasi dan memprediksi perilaku belajar di masa depan siswa,
2. Memajukan pengetahuan ilmiah tentang pembelajaran.
3. Mengambil keputusan yang akurat dan juga memprediksi perilaku siswa.

**STAGE OF DATA MINING**



* **Data selection**

Data hasil seleksi yang digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

* **Pre-processing / cleaning**

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data.

* **Transformation**

Coding/Transformasi adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses coding merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

* **Data mining**

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses secara keseluruhan.

* **Interpretation / evalution**

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses yang disebut interpretation. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

**PROCESS OF DATA MINING**

* **Deskripsi**

Mengidentifikasi pola yang muncul secara berulang pada suatu data dan mengubah pola tersebut menjadi aturan dan kriteria yang dapat mudah dimengerti oleh para ahli pada domain aplikasinya. Aturan yang dihasilkan harus mudah dimengerti agar dapat dengan efektif meningkatkan tingkat pengetahuan (knowledge) pada sistem.

* **Prediksi**

Data diklasifikasikan berdasarkan perilaku atau nilai yang diperkirakan pada masa yang akan datang. Contoh dari tugas prediksi misalnya untuk memprediksikan adanya pengurangan jumlah pelanggan dalam waktu dekat dan prediksi harga saham dalam tiga bulan yang akan datang.

* **Estimasi**

Variabel target estimasi lebih ke arah numerik dari pada ke arah kategori. Model dibangun menggunakan record lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Sebagai contoh, akan dilakukan estimasi tekanan darah sistolik pada pasien rumah sakit berdasarkan umur pasien, jenis kelamin, berat badan, dan level sodium darah. Hubungan antara tekanan darah sistolik dan nilai variabel prediksi dalam proses pembelajaran akan menghasilkan model estimasi.

* **Klasifikasi**

Proses menemukan sebuah model atau fungsi yang mendeskripsikan dan membedakan data ke dalam kelas-kelas. Klasifikasi melibatkan proses pemeriksaan karakteristik dari objek dan memasukkan objek ke dalam salah satu kelas yang sudah didefinisikan sebelumnya.

* **Clustering**

Pengelompokan data tanpa berdasarkan kelas data tertentu ke dalam kelas objek yang sama. Sebuah kluster adalah kumpulan record yang memiliki kemiripan sesuatu dengan yang lainnya dan memiliki ketidakmiripan dengan record dalam kluster lain. Tujuannya adalah untuk menghasilkan pengelompokan objek yang mirip satu sama lain dalam kelompok-kelompok. Semakin besar kemiripan objek dalam suatu cluster dan semakin besar perbedaan tiap cluster maka kualitas analisis cluster semakin baik.

* **Asosiasi**

Menemukan atribut yang muncul dalam suatu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja (market basket analisys). Tugas asosiasi berusaha untuk mengungkap aturan untuk mengukur hubungan antara dua atau lebih atribut.

**REFERENSI :**

* <https://www.kajianpustaka.com/2017/09/data-mining.html>
* <https://garudacyber.co.id/artikel/898-pengertian-data-mining-dan-contohnya>
* <http://artikel-az.com/pengertian-data-mining/>