

Nama : Novia Riska Aulia

NIM : 19.01.013.080

Kelas : Sains Data 3

### Tugas Individu

#### 1. Jelaskan apa yang dimaksud dengan data mining ?

Menurut saya, data mining merupakan suatu proses pengumpulan informasi dan data yang penting dalam jumlah yang besar atau sering disebut dengan big data.

#### 2. Konsep alur proses data mining

- pembersihan data (pembuangan data yang tidak konsisten atau noise)
- Integrasi data (penggabungan data dari beberapa sumber)
- Transformasi data (data diubah menjadi bentuk sesuai di-mining)
- Aplikasi teknik data mining merupakan proses ekstraksi pola dari data yang ada
- Evaluasi pola yang ditemukan (proses interpretasi pola menjadi pengetahuan yang dapat digunakan untuk mendukung pengambilan keputusan)
- Presentasi Pengetahuan (dengan menggunakan teknik Visualisasi)

#### 3. 5 peran utama data mining

- Description
- Classification
- Prediction
- Estimation
- clustering
- Association

#### 4. perbedaan estimation dan forecasting

Estimation (estimasi) merupakan untuk menerka sebuah nilai yang belum diketahui. misalnya, menerka penghasilan seseorang ketika informasi mengenai orang tersebut diketahui. Sedangkan Forecasting (prediksi), untuk menerka nilai masa mendatang, misalnya, memprediksi stok barang satu tahun ke depan.

#### 5. Perbedaan Forecasting dan klasifikasi

Forecasting (prediksi), untuk menerka nilai masa depan. Sedangkan

klasifikasi merupakan metode penganalisa data yang sering di masukkan sebagai salah satu metode data mining.

#### 6. Perbedaan Klasifikasi dan Klustering

Klasifikasi adalah pemrosesan untuk menemukan sebuah model atau fungsi yang menjelaskan dan mendeskripsikan konsep atau kelas data untuk kepentingan tertentu. Sedangkan klustering merupakan pengelompokan data berdasarkan kemiripan pada objek data dan sebaliknya meminimalkan kemiripan terhadap kluster yang lain.

#### 7. Perbedaan Klustering dan Association

Klustering merupakan pengelompokan data berdasarkan kemiripan objek data. Sedangkan Association merupakan metode data mining yang digunakan secara luas dan mengarah ke penemuan yang menarik berhubungan antara variabel - variabel.

#### 8. Perbedaan Estimasi dan Klasifikasi

Estimasi, untuk menerka sebuah nilai yang belum diketahui. Sedangkan klasifikasi adalah metode penganalisa data yang sering dimasukkan sebagai salah satu metode data mining.

#### 9. Perbedaan Estimasi dan Klustering

Estimasi, untuk menerka sebuah nilai yang belum diketahui. Sedangkan klustering, merupakan pengelompokan data berdasarkan kemiripan objek data.

#### 10. Perbedaan Supervised dan Unsupervised

Supervised, mengenali data dari label khusus yang telah diberikan sebelumnya. Sedangkan

Unsupervised, mengenali data secara real-time begitu data disajikan.

Tujuan supervised, untuk memperkirakan atau memprediksi fungsi pada bidang pemetaan sehingga ketika ada variabel input ( $x$ ) kita pun dapat memprediksi output variabel ( $y$ ). Sedangkan

Tujuan Unsupervised, untuk memodelkan struktur data yang mana yang dapat mempelajari data - data tersebut lebih lanjut.

#### 11. Tahapan utama proses data mining

- Masukan ( Dataset )
- Metode ( Algoritma Data Mining )
- Keluaran / pola / Model / Pengetahuan
- Evaluasi ( Akurasi , Error , dll )