

Tugas Pertemuan 4

1. Peran utama data mining:

- Estimation
Digunakan jika dataset atributnya numerik dan labelnya numerik. Estimasi adalah perkiraan atau prediksi sehingga estimasi hampir sama dengan klasifikasi yang membedakan hanya pada jenis dataset yang digunakan.
- Forecasting
Digunakan jika atributnya numerik dan labelnya numerik serta terdapat time series pada atributnya. Peran data mining prediksi ini secara umum hampir sama dengan klasifikasi.
- Classification
Digunakan jika atributnya bisa berupa nominal atau numerik tetapi labelnya harus berupa nominal. Pada klasifikasi dilakukan pengelompokan berdasarkan hubungan antara variabel kriteria dengan variabel target.
- Clustering
Digunakan jika dataset tidak memiliki label dan atributnya numerik. Pengklasteran merupakan pengelompokan data yang memiliki kemiripan nilai.
- Association
Digunakan jika dataset ingin diketahui seberapa dekat hubungan antar atributnya. Dalam data mining proses asosiasi adalah pencarian atribut yang muncul atau sering muncul dalam waktu yang bersamaan.

2. Algoritma data mining sesuai peran:

- Estimation
Linear Regression, Neural Network, Support Vector Machine.
- Forecasting
Linear Regression, Neural Network, Support Vector Machine.
- Classification
Naive Bayes, K-Nearest Neighbor, C4.5, ID3, CART, Linear Discriminant Analysis.
- Clustering
K-Means, K-Medoids, Self-Organizing Map (SOM), Fuzzy C-Means.
- Association
A priori algorithm, FP-Growth algorithm, GRI algorithm.

3. Perbedaan estimation dan forecasting

Estimation digunakan untuk menentukan nilai variabel berdasarkan data yang tersedia, sedangkan forecasting digunakan untuk memprediksi nilai atau peristiwa di masa depan berdasarkan data masa lalu atau tren.

4. Perbedaan estimation dan classification

Estimation digunakan untuk menentukan nilai variabel berdasarkan data yang tersedia, sedangkan classification digunakan untuk mengelompokkan data ke dalam kategori atau kelas tertentu berdasarkan atribut atau fitur tertentu.

5. Perbedaan classification dan clustering

Classification digunakan untuk mengklasifikasikan data ke dalam kelas tertentu berdasarkan atribut atau fitur tertentu yang telah diberi label sebelumnya, sedangkan clustering digunakan untuk mengelompokkan data ke dalam kelompok-kelompok berdasarkan kesamaan atau perbedaan antar data, tanpa label atau kategori sebelumnya.

6. Perbedaan clustering dan prediction

Clustering digunakan untuk mengelompokkan data ke dalam kelompok-kelompok berdasarkan kesamaan atau perbedaan atribut tertentu, sementara prediction digunakan untuk memprediksi nilai variabel target atau variabel dependen berdasarkan variabel independen atau atribut lainnya.

7. Perbedaan supervised dan unsupervised learning

Supervised learning menggunakan data yang diberi label atau kategori untuk memprediksi variabel target atau variabel dependen, sementara unsupervised learning menggunakan data tanpa label untuk mengidentifikasi pola dan struktur dalam data.

8. Tahapan utama proses data mining

- 1) Data cleaning
- 2) Data integration
- 3) Transformation
- 4) Data mining
- 5) Pattern evaluation