

Nama : Dimas Wahyu Damarjati
NIM : A11.2018.10871
KEL : A11.4621
Tugas Data Mining 3

1. Sebutkan 5 peran utama data mining!

Jawab :

- Data mining dapat digunakan dalam melakukan prediksi trend dan pola untuk kegiatan bisnis.
- Data mining juga dapat menentukan pola kelompok data yang nantinya di olah menjadi suatu informasi.
- Data mining juga dapat menjadi acuan atau pengambilan keputusan secara kritis.
- Selain itu data mining juga dapat digunakan untuk mendiagnosis ataupun mendeteksi adanya suatu masalah.
- Terakhir data mining menjadi sumber informasi dalam meningkatkan pengetahuan dari penggunaanya.

2. Algoritma apa saja yang dapat digunakan untuk 5 peran utama data mining di atas?

Jawab :

- Algoritma Estimasi (Liniear Regression, Neural Network, SVM)
- Algoritma Forecasting (Linier Regression, Neural Network, Support Vector Machine/SVM)
- Algoritma Klasifikasi (Naïve Bayes, K-Nearest Neighbor/KNN)
- Algoritma Klastering (K-Means, K-Medoids, Fuzzy C-Means)
- Algoritma Asosiasi (FP-Growth, A Priori, Coefficient of Correlation)

3. Jelaskan perbedaan estimasi dan prediksi!

Jawab :

- **Estimasi**

Algoritma Estimasi atau Estimation merupakan sebuah algoritma yang digunakan untuk menerka atau memperkirakan sebuah nilai yang belum diketahui. Algoritma Estimasi ini seperti algoritma klasifikasi, akan tetapi variable targetnya berupa bilangan numeric dan bukan kategorial (nominal atau diskrit). Estimasi nilai dari variable target ditentukan berdasarkan nilai dari variable predicator (atribut).

Contoh Algoritma estimasi yang biasa digunakan antara lain : Liniear Regression, Neural Network, Support Vector Machine.

- **Prediksi**

Algoritma Prediksi atau disebut forecasting merupakan proses pengestimasian nilai berdasarkan pola-pola yang terdapat dalam sekumpulan data yang menggunakan beberapa atribut data untuk melakukan prediksi terhadap nilai nilai variable yang belum diketahui.

Algoritma prediksi ini sama dengan algoritma estimasi dimana label/target/class yang

bertipe numeric, hanya saja bedanya data yang digunakan merupakan rentetan waktu (time series).

Contoh Algoritma prediksi yang biasa digunakan seperti prediksi saham, dll

4. Jelaskan perbedaan estimasi dan klasifikasi!

Jawab :

- **Klasifikasi**

Algoritma klasifikasi merupakan algoritma yang menggunakan data dengan target/class/label yang berupa nilai kategorial atau nomilan. Jadi apabila target/class/label merupakan pendapat, maka bisa digunakan nilai nominal (kategorial) seperti pendapatan besar, menengah, maupun kecil.

- **Perbedaan antara algoritma Klasifikasi dan Estimasi**

Jika pada algoritma klasifikasi estimasi merupakan sebuah algoritma untuk memperkirakan sedangkan klasifikasi lebih digunakan untuk mengelompokan.

5. Jelaskan perbedaan klasifikasi dan klastering!

Jawab :

- **Klastering**

Algoritma klastering merupakan algoritma pengelompokan data yang berasal dari hasil observasi dan kasus ke dalam class yang mirip dengan suatu klaster/cluster yang merupakan koleksi data mirip antara satu dengan yang lain, dan memiliki perbedaan bila di bandingkan dari klaster lain. Klastering sering digunakan sebagai tahap awal dalam unsupervised learning.

- **Perbedaan Klastering dan Klasifikasi**

Perbedaan utama algoritma klastering dengan klasifikasi yaitu jika klasifikasi memiliki target/class/label sedangkan klastering tidak memiliki target/class/label.

6. Jelaskan perbedaan klastering dan prediksi!

Jawab :

- **Prediksi**

Algoritma Prediksi merupakan proses estimasi nilai berdasarkan pola dalam sekumpulan data yang menggunakan beberapa atribut dalam melakukan prediksi nilai variable yang belum di ketahui

- **Perbedaan Prediksi dan Klastering**

Perbedaan utama antara algoritma Prediksi dan Klastering yaitu metode klastering tidak memiliki target/class/label. Selain itu pada algoritma prediksi menggunakan data yang merupakan rentet waktu data time series.

7. Jelaskan perbedaan supervised dan unsupervised learning!

Jawab :

- **Supervised**

Supervised learning atau pembelajaran dengan guru pada dasarnya merupakan persamaan dari klasifikasi. Target dalam pembelajaran yang berasal dari contoh berlabel dalam kumpulan data set. Jadi misalnya dalam masalah pengenalan kode yang diterjemahkan dengan mesin yang digunakan sebagai contoh pelatihan. Sebagian besar algoritma data mining seperti estimation, prediction/forecasting, classification adalah supervised learning 2.

- **Unsupervised Learning**

Unsupervised learning atau pembelajaran tanpa guru pada dasarnya merupakan persamaan dari klastering. Jadi dalam proses pembelajarannya tidak diawasi misalnya input tidak diberikan label kelas. Kita menggunakan klastering untuk menemukan kelas dalam data. Misal menemukan 10 kelompok data, klaster ini dapat sesuai dengan masing masing 10 digit 0 sampai 9. Namun dikarenakan data training tidak diberikan label sebagai model yang dipelajari sehingga tidak dapat memberi tau makna dari klaster yang ditemukan.

8. Sebutkan tahapan utama proses data mining!

Jawab :

Terdapat beberapa tahapan utama yang ada pada proses data mining diantaranya :

- **Pembersihan data**

Hal ini dilakukan untuk membuang data yang bersifat noise ataupun data yang tidak konsisten.

- **Integrasi data**

Merupakan penggabungan data dari beberapa sumber.

- **Transformasi data**

Dimana data diubah menjadi bentuk sesuai kebutuhan untuk data mining.

- **Aplikasi teknik data mining**

Yaitu proses ekstraksi pola dari data yang sudah ada.

- **Evaluasi pola yang ditemukan**

Dimana proses interpretasi pola menjadi pengetahuan yang dapat digunakan untuk mendukung pengambilan keputusan.

- **Presentasi pengetahuan**

Yaitu dengan menggunakan teknik yang disebut visualisasi.