

Nama : Fatihurroyyan
Nim : 211001003

Tugas Individu

Sains Data Kelas C

1) Data mining ialah suatu proses pengolahan dari kumpulan data menjadi sebuah pengetahuan baru untuk mengekstrak pengetahuan & menemukan pola dari suatu kumpulan data yang sangat besar sehingga dapat menghasilkan pola, rumus, aturan, ataupun model.

2) Konsep alur proses data mining

1. Himpunan data (untuk memahami dan mempersiapkan data / preprocessing)
2. Metode data mining (memilih metode yang sesuai dengan karakter data / modeling)
3. Pengetahuan (untuk memahami model dan pengetahuan yang sesuai sehingga dapat memilih model)

3) 5 peran utama data mining

1. Estimasi (Linear regression, neural network, support vector machine, dll)
2. Forecasting (prediksi / peramalan):
3. Klasifikasi (naive bayes, K-Nearest Neighbor, C4.5, ID3, CART, Random Forest, dll)
4. Clustering (K-Means, K-Medoids, Self-Organizing Map (SOM), dll)
5. Asosiasi (FP-Growth, Apriori, dll)

4) Perbedaan estimasi dan forecasting

Algoritma prediksi / forecasting sama dengan estimasi dimana label / target / class bertipe numerik, bedanya adalah data yang digunakan merupakan data rentet waktu (data time series)

5) Perbedaan forecasting dan klasifikasi

Forecasting itu sebuah metode untuk memperkirakan nilai dari suatu populasi yang menggunakan nilai dari sampel. Sedangkan klasifikasi itu metode menganalisa data mining.

6) Perbedaan klasifikasi dan clustering

Perbedaan utama klasifikasi dan clustering yaitu tidak memiliki target / class / label jadi termasuk ke dalam unsupervised learning

7) Perbedaan clustering dan asosiasi

Klastering adalah pengelompokan data, hasil observasi dan kasus ke dalam class yang mirip. Sedangkan asosiasi merupakan suatu bentuk algoritma yang memberikan informasi tentang hubungan antar item data di dalam database.

⑧ Perbedaan estimasi dan klasifikasi!

Estimasi ialah metode dimana kita dapat memperkirakan nilai dari suatu populasi dengan menggunakan nilai dari sampel sedangkan klasifikasi dan estimasi hampir mirip, tetapi variabel targetnya berupa bilangan numerik (kontinyu) dan bukan kategorikal (nominal / diskrit)

⑨ Perbedaan estimasi dan klastering!

Estimasi metode untuk menyelesaikan yang berkaitan dengan memperkirakan seberapa banyak hasil produksi sedangkan klastering merupakan salah satu aktifitas analisis data untuk mengelompokkan data ke dalam beberapa kategori / kluster.

⑩ Perbedaan supervised dan unsupervised learning!

Supervised learning ialah sebuah pendekatan yang sudah terdapat data training dan terdapat variable yang ditargetkan untuk mengelompokkan suatu data ke data yang sudah ada.

Sedangkan, untuk unsupervised learning tidak memiliki data training, sehingga dari data yang ada akan dikelompokkan menjadi 2 bagian atau 3 bagian data dan seterusnya.

⑪ Tahapan Utama Proses data mining

Pertama yaitu pembersihan data, lalu integrasi data, transformasi data, lalu proses ekstraksi pola, evaluasi pola, dan presentasi pengetahuan