



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS PELITA BANGSA**

**UJIAN AKHIR SEMESTER GENAP TA 2021/2022**

Mata Kuliah : **DATA MINING**  
Program Studi/Semester : Teknik Informatika / 6  
Hari/Tanggal : Sabtu, 23-04-2022  
Nama Dosen : Karsito, S.Kom., M.Kom  
Kelas : TI.19.D5

**LEMBAR JAWABAN**

Nama Lengkap : Putri Nabila Amir  
NIM : 311910077

**1. Jelaskan untuk hal berikut :**

**a. Apa itu data mining?**

Data mining adalah disiplin ilmu mengolah himpunan data yang sangat besar menjadi suatu pengetahuan, rumus, pola sehingga nanti dapat dimanfaatkan manusia untuk memprediksi kejadian kedepan dari kehidupan manusia.

**b. Apakah pada masa sekarang ini perlu adanya data mining, mengapa?**

Pada masa ini tentu perlu adanya data mining dikarenakan manusia sehari-hari memproduksi data yang sangat besar sebagai contoh dalam bisnis, kedokteran, ekonomi, cuaca dan bahkan olahraga. Ini menyebabkan data yang sangat besar yang akan menyebabkan data over load oleh karena itu peran data mining sangat membantu untuk mengolah data over load tersebut.

**2. Jelaskan tentang kualitas data :**

**a. Noise, Missing Value, Outliers, Duplicate**

Noise : Data yang tidak diinginkan dan salah.

Missing Value : Hilangnya beberapa data yang telah diperoleh.

Outliers : Data yang berada di luar jangkauan data yang diharapkan

Duplicate : data-set mungkin terdiri dari objek data yang ganda (duplikat), atau hamper selalu terjadi duplikasi antara satu dengan yang lainnya.

**b. Bagaimana langkah untuk menghilangkan Noise, Missing Value, Outliers, Duplicate?**

**a. Noise**

• Binning

Pertama urutkan data dan partisi ke tempat sampah (frekuensi sama)

• Regression

Ddengan memasang data ke dalam function regresi

• Clustering

Mendeteksi dan menghapus Outlier

**b. Missing Value**

• Menigisi nilai-nilai yang hilang secara otomatis (Metode ini cara yang baik untuk menangani Missing Value )

• Menigisi nilai-nilai yang hilang secara manual otomatis (Metode ini effort besar untuk menangani Missing Value )

- Mengabaikan tuple atau record (Metode ini alternatif terakhir untuk menangani Missing Value )
- c. Outliers
- Mengganti nilai outliers menggunakan nilai lower bound dan upper bound.
  - Capping menggunakan nilai median
  - Capping menggunakan nilai mean
  - Capping menggunakan nilai modus
- d. Duplicate
- Menghapus data yang sama

### 3. Apa itu data warehouse ? Jelaskan kriteria Data Warehouse

Data warehouse adalah koleksi dari data yang subject-oriented, terintegrasi, time-varian, dan nonvolatile, dalam mendukung proses pembuatan keputusan.

kriteria Data Warehouse:

#### • SUBJECT ORIENTED

Diorganisasikan seputar subjek utama seperti customer, produk, sales. Fokus pada pemodelan data dan analisis data untuk pembuatan keputusan

#### • NON-VOLATILE

Penyimpanan data yang terpisah secara fisik yang ditransformasikan dari lingkungan operasional. Data warehouse tidak memerlukan oemrisesan transaksi

#### • TERINTIGRASI

Mengintegrasikan banyak sumber data yang heterogen seperti: relational database, flatfile, olap. Teknik yang digunakan adalah data cleaning dan integrasi data

#### • TIME-VARIANT

Data disimpan untuk menyediakan informasi 5-10 tahun yang lalu. Data mengandung elemen waktu baik eksplisit maupun implisit

### 4. Tentukan data test dengan Algoritma KNN

Thickess	Diameter	Class	Rangking	Distance	Class	K3
550	5	Good	5	155,0032258	Good	
770	6	Good	4	65	Good	
440	5	NG	8	265,0018868	NG	
525	5	Good	6	180,0027778	Good	
520	5	Good	7	185,0027027	Good	
680	6	NG	3	25	NG	NG
710	6	Good	1	5	Good	Good
700	6	NG	1	5	NG	NG
705	6	?				
518	5	?				

  

Rangking	Distance	Class	K3
3	32	Good	Good
8	252,0019841	Good	
4	78	NG	
2	7	Good	Good
1	2	Good	Good
5	162,0030864	NG	
7	192,0026041	Good	
6	182,0027472	NG	

Jadi Data Dibawah ini dari ketiga data terdekat yang sudah saya tandai dengan warna Kuning terdapat dua kelas NG dan satu kelas Good Sehingga dapat disimpulkan bahwa kelas dari

Thickess	Diameter	Class
<u>705</u>	<u>6</u>	NG

Jadi Data Dibawah ini dari ketiga data terdekat yang sudah saya tandai dengan warna Kuning terdapat 3 kelas Good dan tidak ada kelas NG Sehingga dapat disimpulkan bahwa kelas dari data baru adalah Good

Thickess	Diameter	Class
518	<u>5</u>	Good