



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

UJIAN TENGAH SEMESTER GENAP TA 2021/2022

Mata Kuliah : DATA MINING
Program Studi/Semester : Teknik Informatika / VI (Enam)
Hari/Tanggal : Sabtu, 23 April 2022
Nama Dosen : Karsito, S.Kom., M.Kom.
Kelas : TI.19.D.5 Sifat : Online

SOAL :

1. Jelaskan untuk hal berikut :
 - a. Apa itu data mining?
 - b. Apakah pada masa sekarang ini perlu adanya data mining, mengapa?
2. Jelaskan tentang kualitas data :
 - a. Noise, Missing Value, Outliers, Duplicate
 - b. Bagaimana langkah untuk menghilangkan Noise, Missing Value, Outliers, Duplicate?
3. Apa itu data warehouse ? Jelaskan kriteria Data Warehouse
4. Tentukan data test dengan Algoritma KNN

Thickness	Diameter	Class
550	5	Good
770	6	Good
440	5	NG
525	5	Good
520	5	Good
680	6	NG
710	6	Good
700	6	NG
705	6	?
518	5	?

----- Selamat mengerjakan -----

=RANK(B19;\$B\$19:\$B\$24;1)

Diverifikasi oleh :

Program Studi

Wahyu Hadikristanto, S.Kom., M.Kom

LEMBAR JAWABAN

Nama Lengkap : Juan Fakhri

NIM : 311910361

Kelas : TI.19.D5

1. A. 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. 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. A. - 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
B.
- Noise
 - Binning
Pertama urutkan data dan partisi ke tempat sampah (frekuensi sama)
 - Regression
Dengan memasang data ke dalam function regresi
 - Clustering
Mendeteksi dan menghapus Outlier
- 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)
- Outliers
 - Mengganti nilai outliers menggunakan nilai lower bound dan upper bound.
 - Capping menggunakan nilai median
 - Capping menggunakan nilai mean
 - Capping menggunakan nilai modus
- Duplicate
 - Menghapus data yang sama

3. 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 pemrisesan 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