DWM Numericals Question Bank

Chapter 1 and 2 and Chapter 3,4,5,6

Refer Syllabus for studying

Numericals :

1. Suppose that a data warehouse consists of the three dimensions time, doctor, and patient, and the two measures count and charge, where charge is the fee that a doctor charges a patient for a visit. Draw a star schema diagram for the above data warehouse.

2. Suppose that a data warehouse for Big-University consists of the following four dimensions: student, course, semester, and instructor, and two measures count and avg grade. When at the lowest conceptual level (e.g., for a given student, course, semester, and instructor combination), the avg grade measure stores the actual course grade of the student. At higher conceptual levels, avg grade stores the average grade for the given combination. Draw a snowflake schema diagram for the data warehouse.

3. Use the methods below to normalize the following group of data: 200, 300, 400, 600, 1000 (a) min-max normalization by setting min = 0 and max = 1 (b) z-score normalization standard deviation =282.2 c)) normalization by decimal scaling

4.

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

Use Naïve Bayesian techniques to Classify a Red Domestic SUV is getting stolen or not.

- Numerical on Data Smoothing by bin method

4. K means numericals for cluster k=2 And k=3

5. Numericals data dispersion and central tendency
6. Create OLAP cube and show olap operations(slice, dice etc in SQL query for a given case study
7. MOLAP ROLAP and HOLAP

Chapter 1:

-Define Data warehouse with features? Explain the architecture of data warehouse with suitable block diagram
-Explain ETL of data warehousing in detail.
- Differentiate between star schema, snowflake schema and fact constellation.
-KDD Process
-OPLAP operations
- differentiate OLAP VS OLTP
- Compare Datawarehouse VS Data Mart
-Factless fact tables
-Numerical to draw Star, Snowflake schema,star Constellation

**Numericals on OLAP .. Assignment no.2 solved in class room

Chapter 2:
-Advantages of Data Mining
-Issues in Data mining
-Data Pre Processing
- Data cleaning and techniques
-Data exploration
-Why Preprocessing?
-Compare Data cleaning and data transformation
-Explain Data Cleaning;
-Explain Data Integration;
-Explain Data Reduction: Attribute subset selection, Histograms, Clustering and Sampling; Data
-Transformation & Data Discretization: Normalization, Binning, Histogram Analysis and
-Concept of hierarchy generation.

-Techniques for data transformation , discretization etc

- nume

Chapter 3
- Classifiers
- Numerical on Describe confusion matrix, accuracy, precision and recall for a classifier and also determine

- Attribute selection methods
- Disadvantages and advantages of all
- Rule Based classifiers
- Naive Bayes Classifier
- Numericals on all the attribute selection methods

the confusion matrix, accuracy, precision and recall for given classifier

Chapter 4
Clustera , clustering , outliers
K means , k mediod , hierarchical , agglomerative , divisive, DBSCAN, Density based , BIRCH
Define outlier and describe its detection methods in detail.
Clustering algorithms in detail with numericals

Chapter 5
- multiple-level association rules and multidimensional association rules
- Apriori al go explain and how to improve performance.
- Numericals for determining strong association rules and the frequent item set.
- Drawback of Apriori
- How it is improved. Different techniques
- FP growth

Chapter 6
Web mining types in detail
Page rank theory and numerical