18124011600l
Vishwas Acharya

Page No.: 1
Date: / /

22/03/2021

DWDM
Mid-Sem Exam

Q-1] MCQ's

1) A

2) B

3) B

4) D

5) B

6) B

7) A

8) A

9) D

10) B

18124011 6001
Vishwas Acharya

Q2] Descriptive Questions

1] Difference between OLAP & OLTP

Ans - Here,

OLAP = On-Line Analytical Processing
OLTP = On-Line Transaction Processing.

⊙ Functionality difference

Functionality	OLTP	OLAP
• characteristic	Operational Processing informational Processing	Transactional Analysis
• Orientation	Transaction	Analysis
• Users	Clerk, DBA, database professional	knowledge worker (eg., manager, executive, analyst)
• Function	day-to-day operations	long-term information requirements, decision support.
• DB design	ER based, application-oriented	Star/snowflake, subject-oriented
• Data	Current; guaranteed up-to-date	Historical; accuracy maintained over time.

18124011600l
Vishwas Acharya

Functionality	OLTP	OLAP
• Summarization	Primitive, highly detailed	Summarized, consolidated
• View	Detailed, flat relational	Summarized, multidimensional
• Unit of work	short, simple transaction	Complex query
• Access	Read/write	Mostly Read

12) Data Mining? Why it is called data mining rather knowledge mining.

Ans • The overall goal of the data mining process is to extract information from a large data sets and transform it into an understandable structure for further use.

- Data mining is looking for hidden, valid and potentially useful patterns in huge data sets.

- Data Mining is all about discovering unsuspected / previously unknown relationships amongst the data.

- Data mining is also called as knowledge discovery, knowledge extraction, data/pattern analysis, information harvesting, etc.

- This is an analysis based on historic data. We may do statistical calculation and arrive into some information.

- So knowledge is not the only output from a data mining process and it make sense to call it as data mining as a broad umbrella term.

Reason why data mining is not known as knowledge mining

13) Explain KDD Process.

Ans - Here,

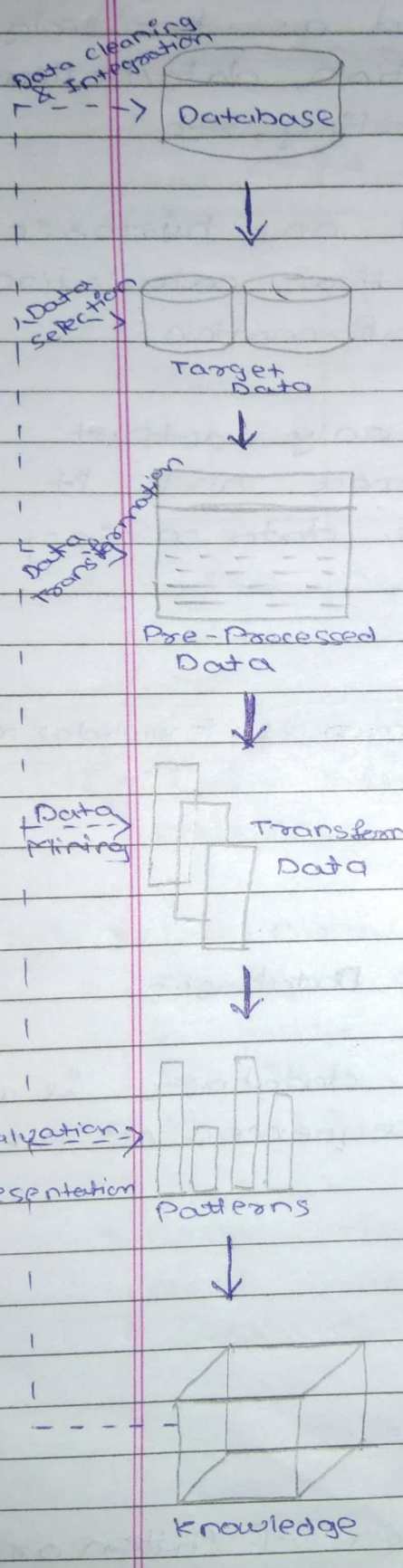
KDD = Knowledge Discovery Databases

- Now, knowledge discovery in databases is a process of an iterative sequence of the following steps:

1. selection
2. Preprocessing
3. Transformation
4. Data Mining
5. Pattern Evaluation
6. User Interface (visualization of pattern or knowledge)

18124011B001
Vishwas Acharya

Page No.: 5
Date: / /



- **Data Selection**:- Where data relevant to the analysis task are retrived from the database
- **Data Cleaning**:- To remove noise and inconsistent data
- **Data Integration**:- Where multiple data sources may be combined
- **Data Transformation**:- Where data are transformed or consolidated into appropriate forms for mining by performing summary or aggregation operation
- **Data Mining**:- An essential process where intelligent methods are applied in order to extract data patterns
- **Pattern Evaluation**:- To identify the truly interesting patterns representing knowledge based on some interestingness measures
- **Knowledge Presentation**:- Where visualization and knowledge representation techniques are uses to present the mined knowledge to the user.

1812401160001
Vishwas Acharya

14) Describe various methods of handling missing data values with example.

Ans - To handle missing values these are few steps :-

1) Ignore the tuple
2) Fill missing value manually
3) Fill in the missing value automatically
4) Use a global constant to fill in the missing value.

=> Now, lets dive deep a little bit

- Ignore the tuple (record/row):

Usually done when class label is missing

- Example :-

↳ The task is to distinguish between two types of emails,
  ↳ Spam
  ↳ Non-Spam

↳ Spam & Non-Spam are called as class label

↳ If an email comes to you, in which class label is missing then it is discarded

- Fill missing value manually:

- Use the average to fill in the missing value.

- Use the average for all samples belonging to the same class.

- Fill in the missing value automatically

- Predict the missing value by using a learning algorithm:

  ↳ Consider the attribute with the missing value as a dependent variable and run a learning algorithm to predict the missing value.

- Use a global constant to fill in the missing value

- Replace all missing attribute values by the same constant such as a label like "unknown".

15) Explain Data Mining Issues

Ans - This issues are classified into five categories:

1. Mining Methodology

2. User Interaction
3. Efficiency and scalability
4. Diversity of Database Types
5. Data Mining and Society

-> Now, let's see each in brief:

- Mining Methodology

- Mining various and new kinds of knowledge
  ↳ Data mining covers a wide spectrum of data analysis and knowledge discovery tasks, so these tasks may use the same database in different ways and require the development of numerous data mining techniques.

- Mining knowledge in multidimensional space
  ↳ When searching for knowledge in large data sets, we can explore the data in multidimensional space.

- Data Mining - an interdisciplinary effort

- Handling uncertainty, noise, or incompleteness of data

- User Interaction

- Interactive mining

↳ The data mining process should be highly interactive.

- Incorporation of background knowledge
  ↳ Background knowledge, constraints, rules, and other information regarding the domain under study should be incorporated into the knowledge discovery process

- Presentation and visualization of data mining results

• Efficiency & Scalability

- Efficiency & scalability of data mining algorithms
  ↳ The running time must be short, predictable, and acceptable by applications.

- Parallel, distributed, and incremental mining algorithms

• Diversity of Database Types

- Handling complex types of data.
  ↳ Data mining is how to uncover knowledge from stream, time-series, sequence, graph, social network, and multirelational & data

- Mining dynamic, networked, and global data repositories.
  ↳ The discovery of knowledge from different sources of structured, semi-structured, or unstructured challengable

- **Data Mining and Society**

- Social impacts of data mining

- Privacy - Preserving data mining
  ↳ It poses the risk of disclosing an individual's personal information

- Invisible data-mining
  ↳ we cannot expect everyone in society to learn and master in data mining techniques.