No
Date

2020/ICTS/12

1. Data mining is the process of discovering patterns, trends, correlations, or useful information from large datasets using statical, machine learning, or AI techniques.

Difference :

Data mining	Traditional Data Analysis.
Discovers hidden patterns and trends automatically.	Test specific hypotheses or analyzes known relationships.
Uses machine learning, AI, and complex algorithms.	Uses statistical methods and predefined models.
Works with large, complex and sometimes unstructured data.	Typically works with structured and smaller datasets.

2. • Market Basket Analysis.
Used in retail to analyze customer purchasing habits by finding associations between products. This helps retailers design better product placement, promotions, and cross-selling strategies.
ex: if a customer buys bread, they are likely to also buy butter.

• Fraud Detection
Applied in banking and finance to identify unusual patterns or anomalies in transactions that may indicate fraudulent activity, like unauthorized credit card use.
Data mining helps automate real-time fraud alerts and reduces financial losses significantly.

No.
Date

- Healthcare Diagnosis and Prediction.
  Data mining helps in diagnosing diseases, predicting outbreaks, and recommending treatments by analyzing medical records, lab results, and patient histories.
  It improves clinical decision-making and enables proactive healthcare management.

- Customer Segmentation.
  Businesses use data mining to group customers based on behavior, preferences, or demographics.
  This allows companies to personalize marketing campaigns and improve customer retention.

3. Descriptive
   focuses on analyzing historical data to identify patterns, trends, or relationships. Techniques like clustering, summarizations, and association rule mining are used to describe the underlying structure of the data without making future predictions.

   Predictive.
   Uses existing data to make forecasts or predictions about future events. Techniques such as classification, regression, and time series analysis helps in predicting outcomes like customer behavior or sales trends.

4. Purpose:
   DBMS is designed to store, retrieve and manage data efficiently and securely.

   Data mining is aimed at discovering patterns, correlations, and useful information from large data sets.

No:
Date: 2020/ICTS/12

Data handling:
DBMS handles structured data with predifined schemas and focuses on transaction processing.

Data mining handles large volumes of data (structured, semi-structured, unstructured) and applies algorithms to extract knowledge.

5. Healthcare (disease prediction, patient diagnosis)
Finance (fraud detection, risk management)
Marketing (customer segmentation, recommendation systems)

6. Because data mining must process structured, semi-structured, and unstructed data (text, images, videos, sensor data) which differ in format, scale, and semantics, making integration, analysis, and pattern extraction complex.

7. Preprocessing cleans the data by removing noise, handling missing values, and transforming data into a suitable format, improving accuracy and efficiency of mining algorithms.

8.
- Single point of failure in centralized systems.
- Scalability issues with growing data and users.
- Limited geographic accessibility and slower response times.
- Distributed databases address these by distributing data accross multiple sites, improving availability, scalability, and performance.

9.

Homogeneous	Heterogeneous
All sites use the same DBMS software and schema structure making data integration easier.	Sites use different DBMSs or schema designs, requiring complex integration and translation machanisms.

SAMMANA RULE PAPER (SINGLE RULE) - SIZE A4

No:
Date:

10.	Replication	Fragmentation
	Copies of entire or parts of database are stored at multiple sites for availability and fault tolerance.	Database is divided into fragments (horizontal / vertical) distributed across sites to improve performance and manageability.

11. Horizontal fragmentation:
Dividing a database table into subsets of rows (tuples) based on certain conditions, each stored at different sites.

Vertical fragmentation.
Dividing a table into subsets of columns (attributes), with each fragment containing different attributes of the table.

12. Because queries may require data from multiple distant sites, causing higher communication ~~over~~ overhead delays, and inefficient processing.

13. A database buffer is a reserved area in main memory that temporarily holds data pages read from disk. It reduces disk I/O by keeping frequently accessed data in memory, minimizing slow disk access.

14. Main memory is limited and expensive compared to disk storage, so it's not feasible to store the entire database in memory.

15. By caching recently accessed data blocks in memory, the buffer allows retrieval without disk access, significantly reducing response time.

No:
Date: 2020/ICT/12

16.

Force	No-Force
All modified pages are written to disk at transaction commit.	Modified pages may be written later, not necessarily at commit time.
This ensures durability but slow down the commit process due to increased disk I/O.	It improves performance but requires more complex recovery machanisms to ensure durability.

Steal	No-Steal
Buffer manager can write dirty pages to disk before transaction commits, allowing better buffer usage but requiring undo machanisms.	Dirty pages are not written to disk until commit, simplifying recovery but requiring more buffer space.

17. It loads the block from disk into a free buffer frame. It no frame is available, it uses a replacement strategy to evict an existing block before loading the requested one.

18. It is a policy to decide which buffer block to evict when the buffer is full. It is necessary to efficiently manage limited memory and maintain good performance by keeping frequently or recently used data in memory.