

①

## Assignment - 1

Kampreya Tiwari  
BETNICS21044

### CSL0563 - Data Warehousing and Mining

Q.1. Name the Different Data mining techniques and explain the scope of data mining.

Ans. In data mining we use different techniques. these are -

- 1) Classification:- This technique is used to obtain important and relevant information about data and metadata. This data mining technique helps to classify data in different classes.
- 2) Clustering:- Clustering is a division of information into groups of connected objects. Describing the data by a few clusters mainly loss certain confine details, but accomplishes improvement.
- 3) Regression:- Regression analysis is the data mining process is used to identify and analyze the relationship between variables because of the presence of the other factor.
- 4) Association Rules:- This data mining technique helps to discover a link between two or more items. It finds a hidden pattern in the data set.
- 5) Outlier detection.
- 6) Prediction.
- 7) Sequential patterns.

Scope of data mining → Data created by data mining is used by business to boost their revenues, know about business investment risks, improve customer relationship, etc. Data mining is a crucial part of successful business analysis in organization. Its tools help us to analyze historical as well as real-time data. It is helpful in predicting future trends and allow businesses to become proactive.

Data mining includes both searching and sorting. From given dataset that have sufficient quality and size, data mining is helpful in creating new business opportunities as it provides the following features -

- 1) Prediction of behaviors and trends.
- 2) Discovery of unknown patterns.

Q.2 Differentiate data mining tools and query tools.

Ans Data mining tools and query tools are both used in the field of data management and analysis, but they serve different purposes and have distinct functionalities. Here's differentiation between the two.



## Data mining Tools:-

(2)

1. Purpose → Data mining tools are designed to discover hidden patterns, relationships, trends, and insights within large volumes of data.
2. Functionality → Data mining tools use various techniques such as clustering, classification, regression, association, and anomaly detection to analyze data and uncover meaningful patterns.
3. Example → Popular data mining tools include RapidMiner, IBM SPSS Modeler, Weka, and KNIME. These tools offer a range of algorithms and functionalities for data exploration, feature selection, model building, and evaluation.

## Query Tools:-

1. Purpose → Query tools, also known as database query tools or database management tools, are used to retrieve, manipulate, and manage data stored in ~~datasets~~ databases.
3. Functionalities → Query tools enable users to write and execute SQL queries to retrieve data based on specified criteria. They provide an interface for database management, data extraction, filtering, sorting, and reporting.
4. Example → Common query tools include SQL-based database management systems like MySQL, PostgreSQL, Microsoft SQL Server, and Oracle Database.



Qo3 Given a brief introduction to data mining knowledge discovery?

Ans Data mining knowledge discovery, often simply referred to as knowledge discovery, is a process that involves extracting valuable, ~~previous~~ previously unknown insights and patterns from large datasets.

The process of data mining knowledge discovery typically involves several key steps.

- 1) Data Collection:- The ~~purpose~~ process begins with gathering relevant data from various sources which can include databases, text documents, sensor data, social media and more.
- 2) Data preparation:- Raw data is processed, cleaned, transformed, and integrated to create a unified datasets that is ready for analysis. This steps involves dealing with missing values, standardizing formats, and encoding categorical variables, among other tasks.
- 3) Data exploration:- Exploratory data analysis (EDA) is performed to gain insights into the datasets characteristics, distribution, and potential patterns. Visualization and summary statistics help analysts understand the nature of the data and identify initial trends.



- 4.) Feature Selection:- Relevant feature are selected or engineered to focus the analysis on the most important aspects of the data. This step aims to reduce dimensionality and improve the efficiency and effectiveness of subsequent analysis.
- 5.) Model building:- Various data mining algorithms are applied to the prepared dataset to discover patterns, relationships and insights. These algorithms can include clustering, classification, regression, association rule mining, and anomaly detection.
- 6.) Pattern evaluation:- The discovered patterns and insights are evaluated to determine their significance and usefulness. This step involves assessing the accuracy, validity, and reliability of the patterns and models generated.
- 7.) Knowledge Presentation:- The results of the analysis are presented in a meaningful and understandable way to stakeholders. This could involve visualization, reports, dashboards, or other forms of communication to convey the discovered knowledge effectively.



Q.4. Which kind of data mined in data mining explain with suitable example.

Ans Data mining involves mining various types of data to discover patterns, trends and insights. Here are some common types of data that can be mined in data mining, along with suitable examples for each:-

- 1) Structured data:- Structured data refers to data that is organized into well-defined formats, such as tables with rows and columns. This type of data is commonly found in relational databases, spreadsheets and CSV files.
- 2) Unstructured data:- Unstructured data refers to data that lacks a predefined structure, making it more challenging to analyze. This includes text documents, social media posts, images, audio, and video.
- 3) Temporal data:- Temporal data involves time-related information, such as timestamps, dates, and time intervals. It is valuable for analyzing trends and patterns that change over time.
- 4) Sequential data:- Sequential data represents a series of events or actions that occur in a particular order. It is commonly used in analyzing sequences of actions, behaviors, or processes.

Qs Explain the concept of data mining functionalities.  
Ans The functionality of data mining is →.

- 1) Class / Concept :- Data is associated with classes or concepts so they can be correlated with results. Data class / concept description can be explained for data mining functionalities. Class and concept defined as a-
  - i) Data characterization.
  - ii) Data discrimination.
- 2) Classification :- Classification is probably one of most important data mining functionalities. It uses data models to predict the trends in data.
- 3) Prediction :- Prediction data-mining functionality finds the missing numeric values in the data. It uses regression analysis to find the unavailable data. If the class label is missing, then the prediction is done using classification.
- 4) Association analysis :- Association analysis is a functionality of data mining. It relates two or more attributes of the data. It discovers the relationship between the data and the rules that are binding them.



- 5) Clustering analysis :- Unsupervised classification is called cluster analysis. It is similar to the classification functionality of data mining where the data are grouped. Unlike classification, in cluster analysis, the class label is unknown.
- 6) Outlier analysis :- When data cannot be grouped in any of the class appears, we use outlier analysis. There will be occurrences of data that will have different attributes to any of the other classes or general models.
- 7) Evolution & Deviation analysis :- With evolution analysis being another data mining functionalities in data mining, we get time-related clustering data.