

Presentation on Data Mining

Submitted to:

Dr.K.P.Singh

Deptt. Of Library & Info. Sc.

University of Delhi

Delhi 110007

Presented by:

Nisbat Ali

M.L.I.Sc.

M-18

Introduction

Data mining is a technique of drawing out the most useful knowledge item from a data base. It is the process of extracting hidden patterns from data. As more data is gathered, with the amount of data doubling every three years. data mining is becoming an increasingly important tool to transform this data into information.

Data mining can be used to uncover hidden patterns in data samples that have been "mined", it is important to be aware that the use of a sample of the data may produce results that are not indicative of the domain. Data mining will not uncover patterns that are present in the domain, but not in the sample.

Cont...

- Data Mining is Commonly use in the corporate sectors to arrange and exploit and serve knowledge. Philosopher an poets Businessman and educators. All has the struggled with the destination between data, information, knowledge, wisdom.

Definitions of Data Mining

- Process of semi-automatically analyzing large databases to find patterns that are:
 - novel:** non-obvious to the system
 - useful:** should be possible to act on the item
 - understandable:** humans should be able to interpret the pattern
 - valid:** hold on new data with some certainty.

Cont...

- Also known as Knowledge Discovery in Databases (KDD).
- Data mining identifies trends within data that go beyond simple data analysis. Through the use of sophisticated algorithms, non-statistician users have the opportunity to identify key attributes of processes and target opportunities.

Cont...

- Data mining is sorting through data to identify patterns and establish relationships.
- Data mining techniques are used in a many research areas, including mathematics, cybernetics, genetics and marketing.
- Web mining a type of data mining used in customer relationship management.(CRM).
- takes advantage of the huge amount of information gathered by a Web site to look for patterns in user behavior.

Data mining

- Data mining commonly involves four classes of task:
 - 1-) Classification
 - 2-) Clustering
 - 3-) Regression
 - 4-) Association rule learning

- **Classification**

Arranges the data into predefined groups. For example an email program might attempt to classify an email as legitimate or spam. Common algorithms include.

- **Clustering:-**
- Is like classification but the groups are not predefined, so the algorithm will try to group similar items together.

● **Regression:-**

To find a function which models the data with the least error. A common method is to use genetic programme.

- **Association rule learning:-**

Searches for relationships between variables. For example a supermarket might gather data of what each customer buys. Using association rule learning, the supermarket can work out what products are frequently bought together, which is useful for marketing purposes. This is sometimes referred to as "market basket analysis".

Steps in Data Mining

- According to Han and Campher there are seven steps in Data Mining.
- **1-) Data Cleaning**
- **2-) Data Intigration**
- **3-) Data Selection**
- **4-) Data Transformation**
- **5-) Data Mining**
- **6-) Patron Evaluation**
- **7-) Data Representation**

1.Data Cleaning

Data Cleaning is use to remove noise and inconsistent data.

2. Data Integration

Data Integration use to to combine data from various sources.

- **3. Data Selection**

Data Selection use to separate data from analytical purpose.

4 .Data Transformation

Data Transformation use to apply method for abstracting knowledge from the data.

- **5-) Data Mining:-**

Data Mining use to make data suitable for analysis purpose.

- **6-) Patron Evaluation:-**

Patron Evaluation use to identify interesting representing knowledge.

- **7-) Data Representation:-**

Data Representation use to present the mind information to the user.

Conclusion:

- Thus the Data mining require an intermigrations of various techniques that have come from multidisciplinary areas such as statistics visualization and machine learning and intelligence .
Data Mining is a valuable information is large volume of data.