**Q1 What is Data Mining, why it is important and what is the types of data mining in Big Data environment**

Simply we can define the Data Mining as the process of discovering useful pattern and trends in large data sets [1] or it is a set of methods used for data analysis, created with the aim to find out specific dependence, relations and rules related to data and making them out in the new higher level quality information [3]

**The significance of the data mining can be summarized in the following paragraph**

Data Mining has great importance in today’s highly competitive business environment. A new concept of Business Intelligence data mining has evolved now, which is widely used by leading corporate houses to stay ahead of their competitors. Business Intelligence (BI) can help in providing latest information and used for competition analysis, market research, economical trends, consume behavior, industry research, geographical information analysis and so on. Business Intelligence Data Mining helps in decision-making. [3]

Data mining applications are widely used in direct marketing, health industry, e-commerce, customer relationship management (CRM), FMCG industry, telecommunication industry and financial sector. Data mining is available in various forms like text mining, web mining, audio & video data mining, pictorial data mining, relational databases, and social networks data mining.

**The types of data mining**

Association Rules

Classification

Clustering

Sequential Patterns

Sequence Similarity

**Q2. Describe the steps involved in data mining when viewed as a process of knowledge discovery. Give THREE examples of real-world knowledge discovery based on literature review**

The steps involved in data mining when viewed as a process of knowledge discovery are as follows: [4]

• **Data cleaning**, a process that removes or transforms noise and inconsistent data

• **Data integration**, where multiple data sources may be combined

• **Data selection**, where data relevant to the analysis task are retrieved from the database

• **Data transformation**, where data are transformed or consolidated into forms appropriate for mining

• **Data mining**, an essential process where intelligent and e±cient methods are applied in order to extract patterns

• **Pattern evaluation**, a process that identifies the truly interesting patterns representing knowledge based on some interestingness measures

• **Knowledge presentation**, where visualization and knowledge representation techniques are used to present the mined knowledge to the user

**Real world examples [5]**

**1.Marketing**

**2.Investment**

**3.Fraud detection**

**4.Manufacturing**

**References**

[1] Discovering Knowledge in Data: An Introduction to Data Mining, 2nd Edition ...

By Daniel T. Larose

[2] Importance of Data mining in today’s business world, Scott Naxton.

[3] Berry, M.J.A., and Linoff, G., "Mastering data mining", The Art and Science of Customer Relationship Management.

[4] General Publishing, San Francisco State University http://cs.sfsu.edu/~huiyang/869-sp2008/hw-sol/sol-01.pdf

[5] From Data Mining to Knowledge Discovery in Databases Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth