Practical-1

1)What is data mining?

Basically data mining is way to find valid, potentially from the large collection of the data.

In other word we can say we filter the data from the large set of data and that data should be used into developing the specific information.

Now that data will be assembled and stored into the data warehouse, now various analysis perform to improve the product.

We also say a data mining is knowledge discovery of data, knowledge extraction ,pattern/data analysis.

Data Mining is a process used by organizations to extract specific data from huge databases to solve business problems. It primarily turns raw data into useful information.

There are various type of data mining is here text mining , pictorial mining , audio and video mining , web mining etc.

2)application of data mining

- Healthcare
- Praud detection
- CRM
- Manufacturing engineering
- Education
- Financial Banking
- Lie Detection
- Market Basket Analysis
- Corporate surveillance

Healthcare:

Data mining in healthcare has excellent potential to improve the health system. It uses data and analytics for better insights and to identify best practices that will enhance health care services and reduce costs. Analysts use data mining approaches such as Machine learning, Multi-dimensional database, Data visualization, Soft computing, and statistics. Data Mining can be used to forecast patients in each category. The procedures ensure that the patients get intensive care at the right place and at the right time. Data mining also enables healthcare insurers to recognize fraud and abuse.

Market Basket analysis:

Market basket analysis is a modeling method based on a hypothesis. If you buy a specific group of products, then you are more likely to buy another group of products. This technique may enable the retailer to understand the purchase behavior of a buyer. This data may assist the retailer in understanding the requirements of the buyer and altering the store's layout accordingly. Using a different analytical comparison of results between various stores, between customers in different demographic groups can be done.

Education:

Education data mining is a newly emerging field, concerned with developing techniques that explore knowledge from the data generated from educational Environments. EDM objectives are recognized as affirming student's future learning behavior, studying the impact of educational support, and promoting learning science. An organization can use data mining to make precise decisions

and also to predict the results of the student. With the results, the institution can concentrate on what to teach and how to teach.

Manufacturing Engineering:

Knowledge is the best asset possessed by a manufacturing company. Data mining tools can be beneficial to find patterns in a complex manufacturing process. Data mining can be used in system-level designing to obtain the relationships between product architecture, product portfolio, and data needs of the customers. It can also be used to forecast the product development period, cost, and expectations among the other tasks.

CRM (Customer relationship management):

Customer Relationship Management (CRM) is all about obtaining and holding Customers, also enhancing customer loyalty and implementing customer-oriented strategies. To get a decent relationship with the customer, a business organization needs to collect data and analyze the data. With data mining technologies, the collected data can be used for analytics.

3)How data mining is differ from the data analysis

Data mining	Data analysis
It is the process of extracting a specific pattern from large datasets.	It is the process of ordering and organizing raw data in order to determine useful insights and decisions.
It involves the intersection of machine learning, statistics, and databases.	t requires the knowledge of computer science, statistics, mathematics, subject knowledge, AI/Machine Learning
It is also known as Knowledge discovery in databases.	Data Analysis is of several types – exploratory, descriptive, text analytics, predictive analysis, data mining etc.
Is responsible for extracting and discovering meaningful patterns and structure in the data.	Is responsible for developing models, explanations, testing and proposing hypotheses using analytical methods.
The output of a data mining task is a data pattern.	The output of Data Analysis is a verified hypothesis or insight on the data.

4)what is Knowledge?

Basically knowledge is collecting the raw data from the database and perform the various task to where we get the accurate information which is helpful in analysis.

For example ,remove the redundant data from the database,, selecting the specific data which are related to the own improvement .

Now that data will be properly define pattern so easily helpful in the growth of the business.

5)Introduction to knowledge discovery from database (KDD)process.

Knowledge is sequence of the several step. Where original data collected and redundant as well unnecessary data will be removed and all other procedure will be added so we find a pattern.

1. Data cleaning
2. Data integration
3. Data Selection
4. Data transformation
5. Data mining
6. Pattern evaluation
7. Knowledge presentation

Data cleaning: unnecessary data will be removed also inconsistent data removed.

Data integration: in the data integration various data will be collected from the various database which are helpful later in analysis.

Data selection: now the specific data will be selected from the given data so building the pattern easily on specific thing. For example, one cloth company monitoring and analysis that which type of trend purchase by the customer according to that company can improve the production of that product.

Data transformation: now collected data was form in row like table, column etc. then it is necessary to convert because it help in data mining so data will be transformed in pattern.

Data mining: now there are various method used by the data mining to which type of patterns have that data.

pattern evacuation: now the data patterns are evacuated.

Knowledge presentation: now we finally find the knowledge.

6)introduction to the warehouse

The data warehouse is where various data will be stored which are collected from the various different databases.

That is vital component of the business intelligence.

That store data which use for analytical purpose.

Data warehousing is used to provide greater insight into the performance of a company by comparing data consolidated from multiple heterogeneous sources. A data warehouse is designed to run query and analysis on historical data derived from transactional sources.

Once the data has been incorporated into the warehouse, it does not change and cannot be altered since a data warehouse runs analytics on events that have already occurred by focusing on the changes in data over time. Warehoused data must be stored in a manner that is secure, reliable, easy to retrieve and easy to manage.