International Journal of Trend in Research and Development, Volume 4(2), ISSN: 2394-9333
www.ijtrd.com

Basics of Data Mining: A Survey Paper

Charmi Mehta
Computer Engineering Department, Atmiya Institute of Technology and Science, Rajkot, Gujarat, India

Abstract **–** Data mining is a technique for examining large pre-existing databases in order to generate new information which helps us to determine future trends. It also helps to find a unique pattern and important knowledge from the existing database.

Keywords: *Data mining, Knowledge Discovery In Database, Information Forecast*

I. INTRODUCTION

There is a huge amount of data available in the Information Industry; these data are of no use until they are converted into useful information. From the hidden data, we can derive relevant and useful knowledge. Such type of knowledge may lead us to a rich resource for knowledge discovery. In this process of mining, we manipulate previously stored data for further forecast and prediction.

II. DATA MINING PROCESS

Implicit and previously unknown information can be derived with using data mining. Data mining is the process which provides a technique to find some important values or convert the information into knowledge. Data mining refers to extracting or mining the knowledge from large amount of data.[1]

Data mining process is also known as KDD (knowledge discovery in data base) process. Knowledge extraction or discovery is done in seven following steps used in data mining [2]:

1. **Data cleaning:** In this step, we remove noise and inconsistent data from the raw data.
2. **Data integration:** Various data is combined here into single data to target the data.
3. **Data Selection**: At this step, we can retrieve task related data which we used in further process.
4. **Data Transformation**: Where data are transformed appropriate forms for mining by performing summary or aggregation operation.
5. **Data Mining**: Here, different techniques and tools are used for extract knowledge.
6. **Pattern evaluation**: In this step, identify the patterns from the data mining values.
7. **Knowledge representation**: visualization and knowledge representation is done in this step with an aim to help the user to understand mined knowledge.

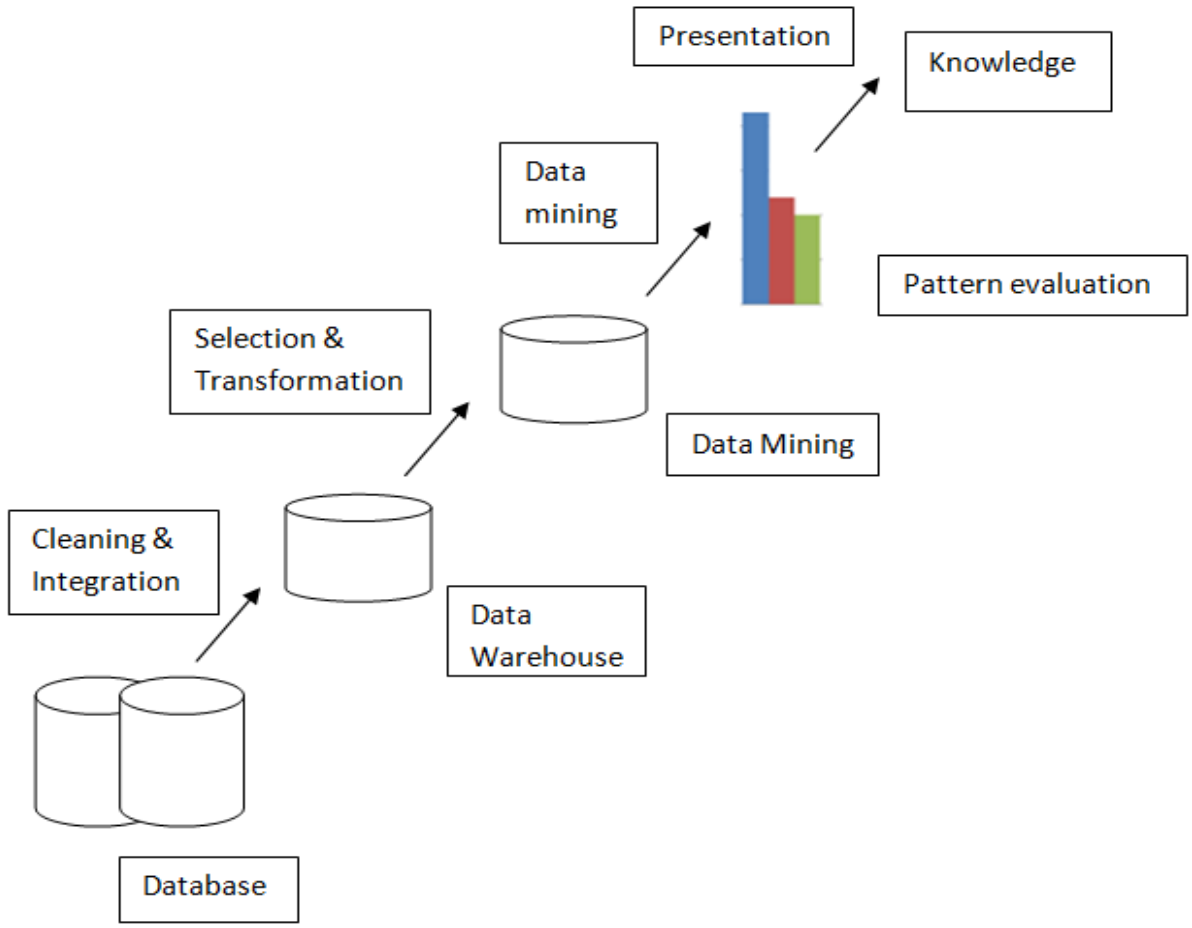

Figure 1: Data mining process

The goal of knowledge discovery and data mining process is to find the patterns that are hidden among the huge set of data and interpret useful knowledge and information.

III. DATA MINING TECHNIQUES [3]

There are various major data mining techniques that have been developed and used in data mining projects recently including association, rule classification, clustering ,prediction and Evaluation pattern etc., are used for knowledge discovery from database.

1. Concept/class description:

- Generalize, summarize, and contrast data characteristics

2. Mining frequent pattern- Association and Correlation

- It is one of the most popular data mining techniques. In this technique we mine frequent patterns lead to discovery of interesting association and correlations within data.
- Association technique is used in marketing analysis to identify items which are frequently purchased within the same transactions.
- For example, Bread butter Jam – we often purchase Bread butter and jam together.

3. Classification and Prediction

- Finding models (functions) that describe and distinguish classes or concepts for future prediction
  E.g., classify countries based on climate, or classify cars based on gas mileage.
- Classification is further divided into three types:
  If then else rule
  Decision tree method
  Neural network
- **Prediction**: Predict some unknown or missing numerical values

4. Cluster analysis

- Clustering based on the principle: maximizing the intra-class similarity and minimizing the interclass similarity.
- A cluster is a collection of objects which are "similar" among them and are "dissimilar" to the objects belonging to other clusters.

5. Outlier analysis

- Outlier: a data class object that does not comply with the general behavior of the data.
- It can be considered as noise or exception but is quite useful in fraud detection, rare events analysis.

6. Trend and evolution analysis

- Trend and deviation: regression analysis
- Sequential pattern mining, periodicity analysis
- Similarity-based analysis

IV. APPLICATION OF DATA MINING

1. Market analysis

- Using different techniques, data mining gives us the total scenario of the market.
- data mining can tell you what types of customers buy what products
- it also identify customer's requirement

2. Corporate analysis

- It is used for finance planning & resource planning
- Set the prize into today's tough competition
- Develops the strategy according to completion

3. Future Healthcare

- Data mining is used to predict the volume of patients in future and for preventing from disease we will aware about it.

4. Telecommunication industry

- Data mining in telecommunication industry helps in identifying the telecommunication patterns, catch fraudulent activities, make better use of resource, and improve quality of service.

5. Educational analysis

- The goals of educational analysis are prediction about students' future learning behavior, studying the effects of educational support and other education related matter.

6. Fraud detection

- Data mining is used in fraud detection in following sector:
  Auto insurance
  Money laundering
  Medical insurance
  Telephone fraud

CONCLUSION

In present, we have tremendous amounts of data stored in databases, data warehouses and other information repositories. So 'Data explosion problem' is there. From data mining we can resolve it. In data mining, We are drowning in data, but starving for knowledge. Data mining is very useful and efficient procedure for finding interesting knowledge i.e. rules, pattern, graph, chart, constraints from the raw data. Therefore data mining is very useful in present and future environment.

References

(JOURNAL PAPAERS)

[1] Jain, Nikita, and Vishal Srivastava. "Data mining techniques: a survey paper." *IJRET: International Journal of Research in Engineering and Technology* 2.11 (2013): 2319-1163.

[2] Sharma, Aarti, et al. "Application of Data Mining–A Survey Paper." *International Journal of Computer Science and Information Technologies* 5.2 (2014): 2023-2025.

[3] Kalyani et al., International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X ,Volume 2, Issue 10,October 2012 .

(BOOKS)

[4] Han and Kamber, "Data Mining and Concepts".