**IOSR Journal of Engineering (IOSR JEN)** **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 33-36

# Survey On Data Mining

Ashmita Shetty[1], Komal Gothwal[2]
[1](*Information Technology Department, Atharva College of Engineering)*
[2](*Information Technology Department, Atharva College of Engineering)*

***Abstract:*** *In this paper, data mining concepts are summarized. Data mining is the process of analyzing hidden patterns of data according to different perspectives for categorization into useful information, which is collected and assembled in common areas, such as data warehouses, for efficient analysis, data mining algorithms, facilitating business decision making and other information requirements to ultimately cut costs and increase revenue.*

## I. Introduction

Data mining alludes to extricating or mining the learning from vast measure of information. The term information mining is suitably named as 'Knowledge mining from data' or "Knowledge mining". Data mining is otherwise called information disclosure and learning revelation. Data mining is also known as data discovery and knowledge discovery. [4] Data mining is the procedure of investigation and examination, via programmed or self-loader implies, of expansive amounts of information so as to find significant examples and tenets.

## II. Types Of Data

Data mining can be performed on following types of data

- Relational databases
- Data warehouses
- Advanced DB and information repositories
- Object-oriented and object-relational databases
- Transactional and Spatial databases
- Heterogeneous and legacy databases
- Multimedia and streaming database
- Text databases
- Text mining and Web mining

## III. Data Mining Processes

Data mining processes can be classified into two types:

1. data preparation or data preprocessing
2. data mining.

Four processes, that are data cleaning, data integration, data selection and data transformation, are grouped as data preparation processes. Last 3 processes includes data mining, pattern evaluation and knowledge representation i.e. called as data mining.

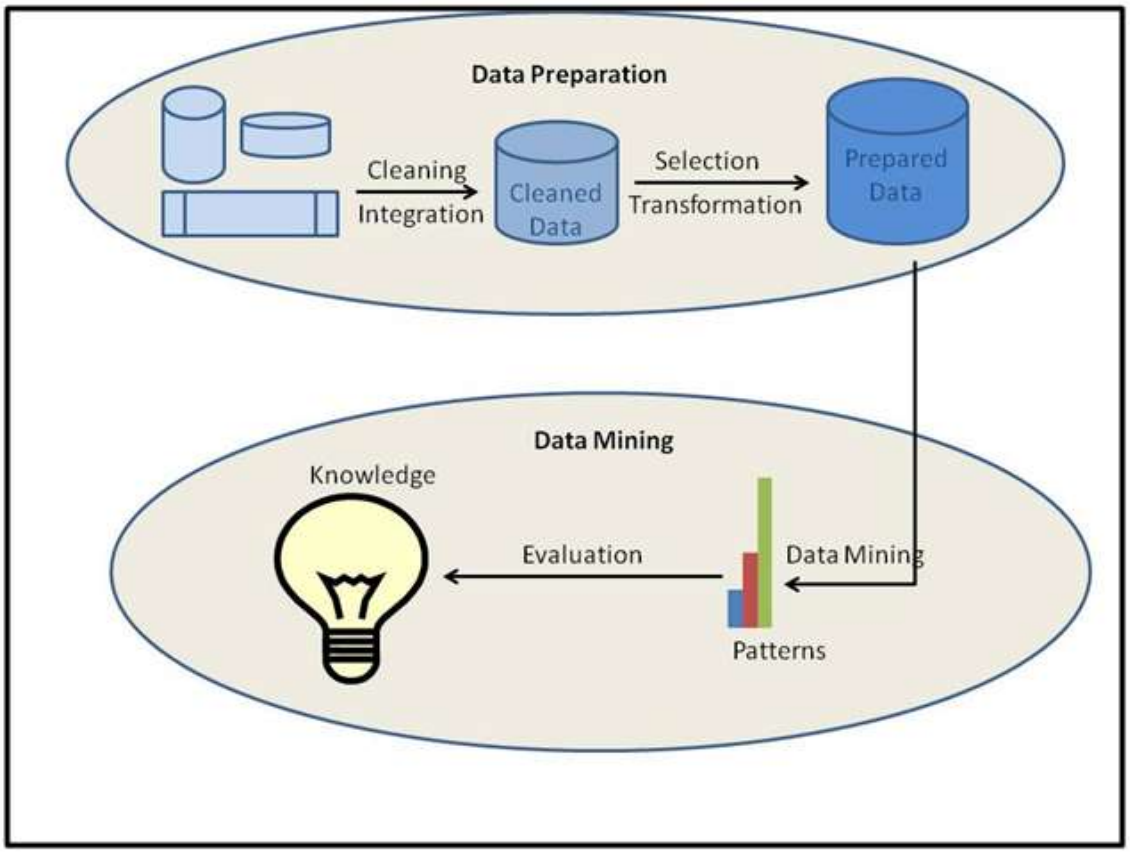


**Fig.1: Data Mining Process**

**A. Data Cleaning**

Data cleaning is where the data gets cleaned. Data in reality is ordinarily inadequate, boisterous and conflicting. The data accessible in data sources may need property estimations, data of intrigue and so forth. For instance, you need the statistic data of clients and consider the possibility that the accessible data does exclude properties for the sexual orientation or age of the clients. At that point the data is obviously deficient. [3] At times the data may contain blunders or anomalies. A model is an age property with esteem 200. Clearly the age esteem isn't right for this situation. The data could likewise be conflicting. For instance, the name of a worker may be put away diversely in various data tables or archives. Here, the data is conflicting. On the off chance that the data isn't perfect, the data mining results would be neither solid nor precise.

Data cleaning includes various strategies incorporating filling in the missing qualities physically, joined PC and human examination, and so on. The yield of data cleaning process is satisfactorily cleaned data.

**B. Data Integration**

Data integration is where data from various data sources are incorporated into one. Data lies in various organizations in various areas. Data could be put away in databases, content records, spreadsheets, archives, data shapes, Internet, etc. Data integration is an extremely perplexing and dubious assignment since data from various sources does not coordinate ordinarily.[5] It is extremely hard to guarantee that whether both these substances allude to a similar esteem or not. Metadata can be utilized successfully to decrease blunders in the data integration process. Another issue confronted is data excess. Similar data may be accessible in various tables in a similar database or even in various data sources. Data integration endeavors to diminish excess to the most extreme conceivable dimension without influencing the unwavering quality of data.

**C. Data Selection**

Data mining process requires huge volumes of verifiable information for examination. In this way, more often than not the information store with incorporated information contains considerably more information than really required. From the accessible information, data of intrigue should be chosen and put away.[2] Data selection is where the information significant to the investigation is recovered from the database.

**D. Data Transformation**

Data transformation is the way toward changing and solidifying the data into various structures that are appropriate for mining. Data transformation ordinarily includes standardization, conglomeration, speculation and so forth. For instance, a data set accessible as "- 5, 37, 100, 89, 78" can be changed as "- 0.05, 0.37, 1.00, 0.89, 0.78". Here data turns out to be increasingly appropriate for data mining. After data incorporation, the accessible data is prepared for data mining.

**E. Data Mining**

Data mining is the center procedure where various perplexing and wise techniques are connected to remove designs from data. Data mining process incorporates various errands, for example, affiliation, characterization, forecast, bunching, time arrangement examination, etc.

**F. Pattern Evaluation**

The pattern evaluation distinguishes the genuinely fascinating patterns speaking to learning dependent on various sorts of intriguing quality measures. A pattern is viewed as intriguing in the event that it is conceivably helpful, effectively reasonable by people, approves some theory that somebody needs to affirm or substantial on new information with some level of sureness.

**G. Knowledge Representation**

The data mined from the information should be exhibited to the user in an engaging way. Distinctive learning representation and visualization strategies are connected to give the yield of information mining to the clients.

## IV. Data Mining Techniques


**Fig 2 : Data Mining Techniques**

**1. Classification:**

This investigation is utilized to recover imperative and pertinent data about information, and metadata. This information mining technique characterizes information in various classes.

**2. Clustering:**

Clustering analysis is an information mining method to distinguish information that resemble one another. This procedure comprehends the distinctions and likenesses between the information.

**3. Regression:**
Regression analysis is the information mining technique for distinguishing and breaking down the connection between factors. It is utilized to distinguish the probability of a particular variable, given the nearness of different factors.

**4. Association Rules:**
This information mining method finds the association between at least two Items. It finds a concealed example in the informational collection.

**5. Outlier Detection:**
This sort of data mining strategy alludes to perception of information things in the dataset which don't coordinate a normal example or anticipated conduct. This system can be utilized in an assortment of spaces, for example, interruption, identification, extortion or blame discovery, and so on. uter detection is also called Outlier Analysis or Outlier mining.

**6. Sequential Patterns:**
This data mining procedure finds or recognize comparative examples or patterns in exchange information for certain period.

**7. Prediction:**
Prediction has utilized a blend of different data mining strategies like patterns, successive examples, classification, clustering, and so on. It dissects past occasions or occurrences in a correct arrangement for foreseeing a future occasion.

## V. Conclusion

Data mining is a choice help process in which we look for patterns of data in information. This strategy can be utilized in numerous kinds of information. Current data mining is done principally on basic numeric and clear cut information. Later on, information mining will incorporate increasingly complex information types. Moreover, for any model that has been structured, further refinement is conceivable by looking at different factors and their connections. Research in data mining will result in new techniques to decide the most intriguing qualities with regards to the information. As models are created and actualized, they can be utilized as an instrument in enlistment the executives.

## References

[1]. Charmi Mehta,"Basics of Data Mining: A Survey Paper", International Journal of Trend in Research and Development, Volume 4(2), ISSN: 2394-9333
[2]. Nikita Jain, Vishal Srivastava, "DATA MINING TECHNIQUES: A SURVEY PAPER", IJRET: International Journal of Research in Enginee ring and Technology eISSN: 2319-1163
[3]. https://www.techopedia.com/definition/1181/data-mining
[4]. https://www.wideskills.com/data-mining-tutorial/data-mining-processes
[5]. http://www.zentut.com/data-mining/data-mining-processes/
[6]. https://www.guru99.com/data-mining-tutorial.html#2