International Journal on Emerging Technologies (Special Issue NCETST-2017) **8**(1): 155-157(2017)
(Published by Research Trend, Website: www.researchtrend.net*)*

ISSN No. (Print) : 0975-8364
ISSN No. (Online) : 2249-3255

Data Mining and Data Warehouse

Richa Pandey, Lalit Mohan, Sanjeev Bisht and Janmejay Pant
Department of Computer Science Graphic Era Hill University, (U.K.), India

ABSTRACT: Data mining is the computational process of discovering patterns in large data which involve methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. It is an important subfield of computer science. Data warehouse is build by collecting data from multiple heterogeneous sources that support analytical reporting and decision making. Data warehousing contains data cleaning, data integration and data consolidations. In this paper the concept of data mining and data warehouse is explained with example.

Keywords: Data Mining, Data Warehouse, OLAP, OLTP,

I. INTRODUCTION

A popular architect W.H. Inmon, build data warehouse systems. A data warehouse is a subject-oriented, integrated, time-variant and nonvolatile collection of data in support of management's decision making process. Data warehouse contains on-line analytical processing (OLAP), the functional or performance requirements which are different from those of the on-line transaction processing (OLTP) applications traditionally supported by the operational databases.
Data Mining helps in extraction of meaningful data from large amount of databases. Data Mining studies algorithms and computational paradigms that allow computers to discover structure in databases, perform prediction and forecasting, and generally improve their performance through interaction with data [1].
Machine learning builds computer systems which have the ability to increase the performance in a specified domain through experience.

II. DATA MINING

Data mining is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both [2]. Data mining have software which have one of the number of analytical tools for data analysis. Users can now analyze data from many dimensions or angles, categorize it, and summarize the relationships identified.
Data mining process have many steps involved:

- Data Cleaning
- Data Integration
- Data Selection
- Data Transformation
- Data Mining
- Pattern Evaluation
- Knowledge Presentation

1. **Data cleaning**: Data cleaning is use to remove the noisy data and correct the inconsistencies of data. Data cleaning involves transformations to correct the wrong data. It is data preprocessing step while preparing the data for a data warehouse.
2. **Data Integration**: Data Integration is a data preprocessing process that combine the data from multiple heterogeneous data sources into data store. Data integration involve inconsistent data and therefore needs data cleaning.
3. **Data Selection**: Data Selection is the technique in which data relevant to the analysis task are retrieved from the database. Data transformation and consolidation are performed before the data selection process.
4. **Data Transformation**: In this step of data transformation data is transformed or consolidated into forms which are ideal for mining by performing summary or aggregation operations.

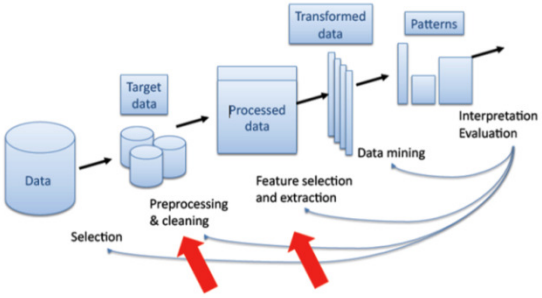

Fig. 1. Data mining process.

5. **Data Mining**: Data mining is defined as extracting information from a large set of data. Data mining is mining the knowledge from data from large amount of database.
6. **Pattern Evaluation**: In this step of pattern evaluation data patterns are evaluated.
7. **Knowledge Presentation**: In this step of knowledge presentation knowledge is presented.

II. DATA WAREHOUSING

A data warehouse is build by integrating data from multiple heterogeneous sources. It supports analytical reporting structured and/or ad hoc queries and decision making [3].
A data warehouses provides us generalized and consolidated view of data. Data warehouses also provides us Online Analytical Processing (OLAP) tools.
These tools help us in interactive and effective analysis of data in a multidimensional space[3]. This analysis results in data generalization and data mining from large amount of data.

IV. WHAT IS A DATA WAREHOUSE?

- A data warehouse is a database, which is separate from the organization's operational database.
- There are no frequent updates.
- It consolidate historical data, which support the organization to analyze its business.
- A data warehouse helps executives to make strategy as it have organized data.
- Data warehouse is integrated.

V. DATA WAREHOUSE FEATURES

The key features of a data warehouse are as follows:

- **Subject Oriented** - Data warehouse is subject oriented as it provides information for subject moreover of the organization's operations. The subjects can be product, customers etc.
- **Integrated** - A data warehouse is made of integrated data from heterogeneous data sources. This integration helps in analysis of data.
- **Time Variant** - The data collected for data warehouse is identified with a particular time period. The data in data warehouse provides information from the historical records.
- **Non-volatile** - Non-volatile means the previous data is not deleted or changed when new data is added to it.

VI. DIFFERENCE BETWEEN DATA MINING AND DATA WAREHOUSE[4]

If you've ever used a credit card, then you may know that credit card companies will alert you when they think that your credit card is being fraudulently used by someone other than you. This is a perfect example of data mining – credit card companies have a history of your purchases from the past and know geographically where those purchases have been made. If all of a sudden some purchases are made in a city far from where you live, the credit card companies are put on alert to a possible fraud since their data mining shows that you don't normally make purchases in that city. Then, the credit card company can disable your card for that transaction or just put a flag on your card for suspicious activity.
Whereas, data warehouse have A great example of data warehousing that everyone can relate to is what Facebook does. Facebook basically gathers all of your data – your friends, your likes, who you stalk, etc – and then stores that data into one central repository. Even though Facebook most likely stores your friends, your likes, etc, in separate databases, they do want to take the most relevant and important information and put it into one central aggregated database. Why would they want to do this? For many reasons – they want to make sure that you see the most relevant ads that you're most likely to click on, they want to make sure that the friends that they suggest are the most relevant to you, etc – keep in mind that this is the data mining phase, in which meaningful data and patterns are extracted from the aggregated data. But, underlying all these motives is the main motive: to make more money – after all, Facebook is a business.

VII. CONCLUSION

In this paper I have explained the concept of data mining and data warehouse with examples.As now a days data is in abundant form and need for data mining and better understanding of data warehouse.

REFERENCES

[1]. An Introduction to the WEKA Data Mining System Zdravko Markov Central Connecticut State University New Britain, CT, USA 01-860-832-2712
[2]. Data mining and knowledge discovery: making sense out of data.
[3]. https://www.tutorialspoint.com/dwh/pdf/dwh_overview.pdf
[4]. http://www.programmerinterview.com/index.php/database-sql/data-mining-vs-warehousing/
[5]. Kumar, A., Kumar, D., Russell, I., "Non-Traditional Projects in the Undergraduate AI Course", Proceedings of the Thirty-Seventh SIGCSE Technical Symposium on Computer

Science Education, ACM Press, New York, NY, February 2006.
[6]. Markov, Z., Russell, I., Neller, T. Proceedings of the Thirty-Fifth Annual Frontiers in Education Conference, IEEE Press, October 2005.
[7]. Mitchell, T., Does Machine Learning Really Work, *AI Magazine*, Vol. **18**, No. 3, AAAI Press, Fall 1997.
[8]. Neller, T., Presser, C., Russell, I., Markov, Z., "Pedagogical Possibilities for the Dice Game Pig", *The Journal of Computing Sciences in Colleges*, **21**(5), May 2006.
[9]. Neller, T., Markov, Z., Russell, I., "Clue Deduction: Professor Plum Teaches Logic", *Proceedings of the International FLAIRS Conference,* AAAI Press, May 2006.
[10]. Russell, I., Markov, Z., Neller, T., "Unifying an Introduction to Artificial Intelligence Course through Machine Learning Laboratory Experiences", *Proceedings of the 2005 Annual American Society for Engineering Education Conference*, June 2005.