



# A Review on Data Mining and Its Applications

ShikhaVashist, Ayush Gupta  
Student, Dronacharya College of Engineering,  
Gurgaon, Hr., India-123506

**Abstract:** An object oriented framework for data mining operates upon a selected datasource and produces a result file. [4]Certain internal functions are performed by the framework, which interact with the existing function. This separation of internal and extensible functions allows the separation of the specific processing sequence and requirement of a specific data mining operation from the common attribute of all data mining operations. The user may thus define external functions that allow the framework to perform new data mining operations without the framework having the knowledge of the specific processing required by those operations.

**Keywords:** framework, data mining, knowledge, operations

## I. INTRODUCTION

The amount of data being generated and stored is growing exponentially, due in larger part to the continuing advances in computer technology. This presents tremendous chances for those who can unlock the information embedded within this data, but also introduces new challenges. In this we discuss how the modern field of data mining can be used to extract useful knowledge from the data that surround us. Those that can master this technology and its methods can derive great benefits and gain a competitive advantage. In this introduction we begin by discussing what data mining is, why it developed now and what challenges it faces, and what types of problems it can address. In subsequent sections we look at the key data mining tasks: prediction, association rule analysis, [8]cluster analysis, and text, link and usage mining. Before concluding we provide a list of

data mining resources and tools for those who wish further information on the topic.

## II. WHAT IS DATA MINING?

Recently, a new technique called data mining has been developed, which allows a user to find larger databases and to evaluate hidden patterns in thatdata. [2]Data mining is thus the discovery of valuable, non-obvious data from a large collection of data and centers on the automated discovery of new facts and underlying relationships in the data. The term "data mining" comes from an idea that the raw material is the business data, and thedata mining algorithm is the originated, shifting through the vast quantities of raw data looking for the important nuggets of business information.

Because data can be stored in such a wide variety of formats and because these data values can have such a wide variety of

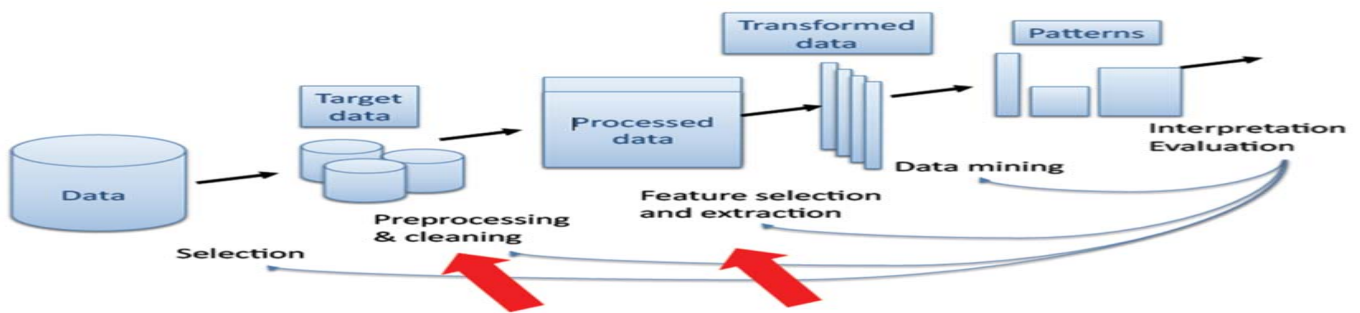


Figure 1.

meanings, data mining applications have inthe past been written to perform specific data mining operations, and there has been a few or no reuse of code between application programs. Thus, each data mining application is written from scratch, makes the development process time taken and expensive. since the nuggets of business information that a data mining application discovers can be quite valuable, they are of very little use if they are costly and untimely discovered. Back to the mining analogy, even if gold is selling for \$900 per ounce, noone is interested in operating a gold mine if it takes two years and \$901 per ounce to get it out of the ground.

[3]The data mining process is an iterative process, although this is not explicitly described in Figure 1. After

the initial run of the process is complete, the user will evaluate the results and decide whether further work is necessary or if the results are adequate. Normally, the initial results are either not acceptable or there is an expectation that further improvements are possible, so the process is repeated after some adjustments are made. These adjustments can be made at any stage of the process. For example, additional data records may be acquired, additional fields (i.e., variables) may be generated from existing information or obtained [7](via purchase or measurement), manual cleaning of the data may be performed, or new data mining algorithms may be selected. At some point the results may become acceptable and the mine knowledge then will be communicated and may be acted upon.

However, even once the mined knowledge is acted upon the data mining process may not be complete and have to be repeated, since the data distribution may change over time, new data may become available, or new evaluation criteria may be introduced.

### III. GUI GUIDE FOR DATA MINING

A GUI is used to reduce the large task of creating data mining objects into a structured series of smaller steps. Generic headings for the sequence of GUI panels used in developing data mining objects are: Introduction (Welcome) panel; Selection of technique and settings name panel; Selection of a data source (when appropriate); Setting of general parameters (one or more panels); Output/results created and named for executable objects particular to the technique selected (depends on object selected); Naming the settings object and Finish page (completed). [4]Each GUI panel leads the user through a series of low-level decisions. The next panel is chosen by the data input during the existing panel. finally, the user may report the series of GUI panels and selections and entries made to change the resulting mining object. Once created, these objects are graphically arranged to create a sequence object which is then named and saved, thus preserving the arrangements for future use.

### IV. INTEGRATED DATABASE AND DATA MINING SYSTEM

A method and apparatus for mining data relationships from the database and data-mining system are disclosed. [1]A set of 1-itemsets is generated using a group-by query on data transactions. From these frequent 1-itemsets and these transactions, frequent 2-itemsets are evaluated. A candidate set of (n+2)-itemsets are generated from the frequent 2-itemsets, where n=1. Frequent (n+2)-itemsets are determined from candidate set and the transaction table using a query operation. The candidate set and frequent (n+2)-itemset are generated for (n+1) until the candidate set is empty. Rules are then extracted from the union of the determined frequent itemsets.

### V. DATA-CENTRIC AUTOMATIC DATA MINING

A data-centric data mining technique provides greater ease of use and flexibility, yet provides high quality data mining results by providing general methodologies for automatic data mining. A methodology for each major type of mining function is provided, including: supervised modeling (classification and regression), [6]feature selection, and ranking, clustering, outlier detection, projection of the data to lower dimensionality, association discovery, and data source comparison. A method for data-centric data mining comprises invoking a data mining feature to perform data mining on a data source, performing data mining on data from the data source using the data mining feature, wherein the data mining feature uses data mining processes and objects internal to the data mining feature and does not use data mining processes and objects external to the data mining feature, outputting data mining

results from the data mining feature, and removing all data mining processes and objects internal to the data mining feature that were used to process the data from the data source.

### VI. PERSONAL DATA MINING

Personal data mining [5]mechanisms and methods are employed to identify relevant information that otherwise would likely remain undiscovered. Users supply private data that can be analysed in conjunction with data associated with a plurality of other users to provide useful information that can improve business operations and/or quality of life. Personal data can be determined alone or in conjunction with other party data to identify correlations amongst the data and related users. Applications or services can interact with these data and present it to users in a myriad of manners, for instance as notifications of opportunities.

### VII. CONCLUSION

In this paper we briefly reviewed the various data mining trends from its inception to the future. This review report would be helpful to researchers to focus on the various issues of data mining. In future course, we will review the various classification algorithms and significance of evolutionary computing (genetic programming) approach in designing of efficient classification algorithms for data mining.

### VIII. REFERENCES

- [1]. An, X. & Wang, W. (2010). Knowledge management technologies and applications: A literature review. IEEE, 138-141. doi:10.1109/ICAMS.2010.5553046
- [2]. Berson, A., Smith, S.J. & Thearling, K. (1999). Building Data Mining Applications for CRM. New York: McGraw-Hill.
- [3]. Babcock, C. (1994), .Parallel Processing Mines Retail Data, Computer World, 6.
- [4]. Bhandari, I., Colet, E., Parker, J., Pines, Z., Pratap, R., and Ramanujam, K. (1997), .Advanced Scout: Data Mining and Knowledge Discovery in NBA Data, Data Mining and Knowledge Discovery, 1, 121.125.
- [5]. Agrawal, R., Imielinki, T., and Swami, A. (1993). Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM-SIGMOD International Conference on Management of Data, 207-216, Washington, DC.
- [6]. Agrawal, R., and Srikant, R. (1994).Fast algorithms for mining association rules.In Proceedings of the 1994 International Conference on Very Large Databases, 487-499, Santiago, Chile.
- [7]. Heikki, Mannila. 1996. Data mining: machine learning, statistics, and databases, IEEE
- [8]. Fayadd, U., Piatetsky-Shapiro, G., and Smyth, P. 1996. From Data Mining To Knowledge Discovery in Databases, AAAI Press / The MIT Press, Massachusetts Institute Of Technology. ISBN 0-262 56097-6 Fayap.