ISSN: 0474-9030

Our Heritage

Vol-68-Issue-15

January-2020

UGC Care Listed Journal

"The Study of Data Mining Process with the help of Data Mining Implementation Process & Techniques"

[1]Miss Dhanashree Shinde , [2]Dr.Sunil Khilari

[1]MCA,[2]MCA,PhD

[1]Assistant Lecturer ,[2] Assistant Professor

[1]Vishwakarma College of arts, [2]Sinhgad Institute of Management Commerce and Science

Abstract:

Data mining is process where we have to collect large amount of data & extract required and useful data from row data, for data mining process we required some data mining techniques and tools. In this paper we discus four important point's i.e. Introduction of Data mining, Data mining implementation process, architecture of data mining process and some tools which are useful for processing on raw data , in data mining implementation process steps involves such as Business Development, Data understanding, Data Preparation, Modeling, Evaluation, Deployment etc. to working on that data we required some techniques and tools, which are Rapid Minor, Orange and weka, these tools are used to sort various types of data and all these are open source and available for everyone. The stages of data mining implementation process are shown in Fig A, Figure B describes Architecture of Data mining process and Figure C describes stages of clustering.

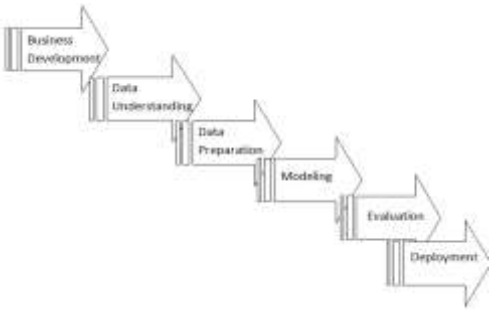
Keywords: *Data Mining Implementation Process, Data Mining Techniques, Data Mining Tools.*

I) INTRODUCTION

There are multiple types of databases like Relational databases Data warehouses Advanced DB etc. It is impossible to find required data from multiple databases because there is a huge amount of data consist in every database and we required only few amount of useful data from entire data. To improve these and finding knowledgeable and require data we need to bifurcate the entire data, it is not an easy process if we do it via human efforts , so we have some techniques which we use to extract useful data from the large amount of data this process is nothing but data mining process. Data mining can be performing using data mining implementation process.

Fig .A

Data mining implementation process



Business Development:

It understands the project targets and prerequisites it is a process of understanding customer requirement as well as what they actual wants to achieve. Reveal significant factors, at the starting, it can impact the result of the project.

Data Understanding:

It examine multiple type of data that means it collect original data from multiple organizations and then get familiar to that data to detect interesting subsets for concealed information hypothesis, it contain following tasks

Tasks:

- Collects initial data
- Describe data
- Explore data
- Verify data quality.

Data Preparation:

It is a process of selecting data from multiple sites and cleaning raw data and transforming raw data into the useful data

Tasks:

- Select data
- Clean data
- Construct data
- Integrate data
- Format data

Modeling:

It is the process of creating a data model for the data to be stored in a Database. This data model is a conceptual representation of Data objects, the associations between different data objects and the rules.

Tasks:

- Select modeling technique
- Generate test design
- Build model
- Access model

Evaluation:

We discover data by mining method or in data selection but all the data which we discover it may not be used data, it is also huge amount data and all data may not be useful so pattern evaluation used to find or sort only needed data which we required.

Tasks:

- Evaluate results
- Review process
- Determine next steps

Deployment:

It includes scoring a database, utilizing results as company guidelines, interactive internet scoring. The information acquired will need to be organized and presented in a way that can be used by the client. However, the deployment phase can be as easy as producing. However, depending on the demands, the deployment phase may be as simple as generating a report or as complicated as applying a repeatable data mining method across the organizations.

Tasks:

- Plan deployment
- Plan monitoring and maintenance
- Produce final report
- Review project

II) DATA MINING ARCITECTURE

The significant components of data mining systems are a data source, data mining engine, data warehouse server, the pattern evaluation module, graphical user interface, and knowledge base.

ISSN: 0474-9030

Our Heritage

Vol-68-Issue-15

January-2020

UGC Care Listed Journal

Fig.B

Here let us see how these modules are work given in architecture.

- World Wide Web: website where from data gets.
- Database : databases store different data types of data
- Data warehouse: process of collecting Multiple data together
- Other data repositories: is a process of getting data from multiple sites
- Data cleaning: Clear data which is not usable
- Integration : multiple data collecting together and collecting in data warehouse
- Server: getting data from database.
- Pattern evaluation: Discover only interested patterns from the entire data.
- Data mining engine :suggesting software for mining data.
- Knowledge base: save previous data information for reuse
- Graphical user interface: is any user who uses system.

III) DATA MINING TECHNIQUE

Data mining is a process of mining different kind of data for that process we have to use some mining techniques they are as follows

Data Classification:

Data Classification is also called as supervise learning. In our database we already have train data in which we find class data, for finding that data we need classifier, classifier is a classification model, with the help of that model we are able to classify the given data

ISSN: 0474-9030

Our Heritage

Vol-68-Issue-15

January-2020

UGC Care Listed Journal

The steps we used to classify data is called as Learning step or training step where we construct classification model with the help of training data.

Classification contain given points

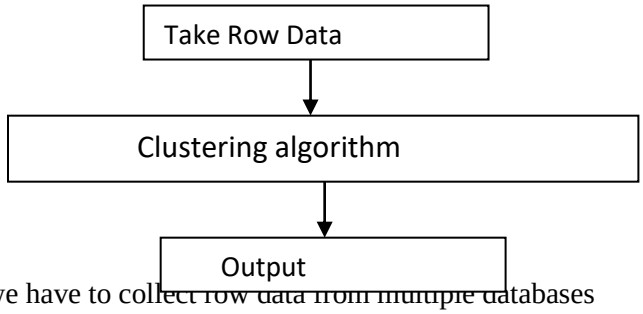
- Classification rule
- Decision tree
- Mathematical formula

Clustering

Clustering is mostly used technique of data mining called as partition of database bases on similarities

Stages of clustering:-

Fig.C



1. Take row data: in this stage we have to collect row data from multiple databases

2. Clustering algorithm: in this stage we have to use clustering algorithm to sort data

3. Output: and then we will get result in the form of output.

Example for clustering Data mining

- Search engine
- Marketing etc.

Association

In data mining association rule is used to predicting customer behavior, this rule helps uncover relationships between non relational data in a relational database or other information related to it

- Classification of association rule.

Single dimensional Association

In single dimensional only one dimension or predicator items refer only one predicate.

Multi dimensional Association

It refers two or more dimensions of predictor for output

Hybrid

In hybrid method predicates or dimensions can be repeater.

III Data mining Tools:

Rapid Minor:

Rapid minor is an open source data mining tool. Programming language used for this tool is java .this tool can be used over a vast range of applications including for business applications, commercial applications, training, education, research, application development, machine learning. It consist three modules

1. Rapid Miner Studio- This module is for workflow design, prototyping, validation etc.
2. Rapid Miner Server- To operate predictive data models created in studio.
3. Rapid Miner Radoop- Executes processes directly in Hadoop cluster to simplify predictive analysis.

Orange:

It is free software. As it is component-based software, the components of orange are called ‘widgets'. These widgets range from data visualization & pre-processing to an evaluation of algorithms and predictive modeling

Weka:

Weka is also called as Waikato Environment it is machine learning software developed at the University of Waikato in New Zealand. Weka supports most of the data mining tasks. Weka is software which is a collection of machine learning algorithm. It works on the assumption that data is available in the form of a flat fill.

Conclusion:

We will understand some basic stages which need to complete data mining implementation process. Data mining architecture gives brief idea about how each and every module is working inside this process. Data mining technique and tools are used to work with data it includes getting data from different places,

organized it and distribute as per the requirement of user or sorted useful data. Tools like weka, orange etc. are used to work on raw data to extract useful data.

References:

1. Bharati M. Ramageri DATA MINING TECHNIQUES AND APPLICATIONS Indian Journal of Computer Science and Engineering Vol. 1 No. 4 301-305.

2. Mengjiao Shang 2017 the characteristics of data mining as a cross discipline [J] Times Finance 263-264.

3. Jingfang Yang 2018 The application of machine learning algorithm in data mining [J] Electronic Technology & Software Engineering 191.

4. HouDi R Groth 2001 Data Mining - Building Competitive Advantages of Enterprises [M] (Xi'an: Xi'an Jiaotong University press).

5. DATA MINING: CONCEPTS, BACKGROUND AND METHODS OF INTEGRATING UNCERTAINTY IN DATA MINING Yihao Li, Southeastern Louisiana University Faculty Advisor: Dr. Theresa Beaubouef, Southeastern Louisiana University.

6. Agrawal, R., & Srikant, R. (2000, May). Privacy-preserving data mining. In ACM Sigmod Record (Vol. 29, No. 2, pp. 439-450). ACM.

7. Agrawal, R., & Srikant, R. (2000, May). Privacy-preserving data mining. In ACM Sigmod Record (Vol. 29, No. 2, pp. 439-450). ACM.

8. Agrawal, R., Gehrke, J., Gunopulos, D., & Raghavan, P. (1998). Automatic subspace

9. clustering of high dimensional data for data mining applications (Vol. 27, No. 2, pp. 94-105). ACM.

10. Agrawal, R., & Srikant, R. (2000, May). Privacy-preserving data mining. In ACM Sigmod

11. Record (Vol. 29, No. 2, pp. 439-450). ACM.

12. erry, M. J., & Linoff, G. (1997). Data mining techniques: for marketing, sales, and customer support. John Wiley & Sons, Inc..