Vol 6 Issue 2 March 2016

ISSN No : 2230-7850

International Multidisciplinary Research Journal

Indian Streams Research Journal

Welcome to ISRJ

RNI MAHMUL/2011/38595 **ISSN No.2230-7850**

Indian Streams Research Journal is a multidisciplinary research journal, published monthly in English, Hindi & Marathi Language. All research papers submitted to the journal will be double - blind peer reviewed referred by members of the editorial board.Readers will include investigator in universities, research institutes government and industry with research interest in the general subjects.

Address:-Ashok Yakkaldevi 258/34, Raviwar Peth, Solapur - 413 005 Maharashtra, India
Cell : 9595 359 435, Ph No: 02172372010 Email: ayisrj@yahoo.in Website: www.isrj.org

Indian Streams Research Journal
International Recognized Multidisciplinary Research Journal
ISSN: 2230-7850 Impact Factor : 4.1625(UIF) Volume - 6 | Issue - 2 | March - 2016

KNOWLEDGE DISCOVERY IN DATABASE THROUGH DATA MINING

Sathish S.N.[1] and Dr Thippeswamy G[2]
[1]Research Scholar, Himalayan University,Arunachal Pradesh, India.
[2]Research Guide, Himalayan University,Arunachal Pradesh, India.

ABSTRACT:
This paper aims to aims to explain the concept of Knowledge Discovery In Database through Data Mining

KEYWORDS : Data Mining , SEMMA Process, Data Mining, KDD

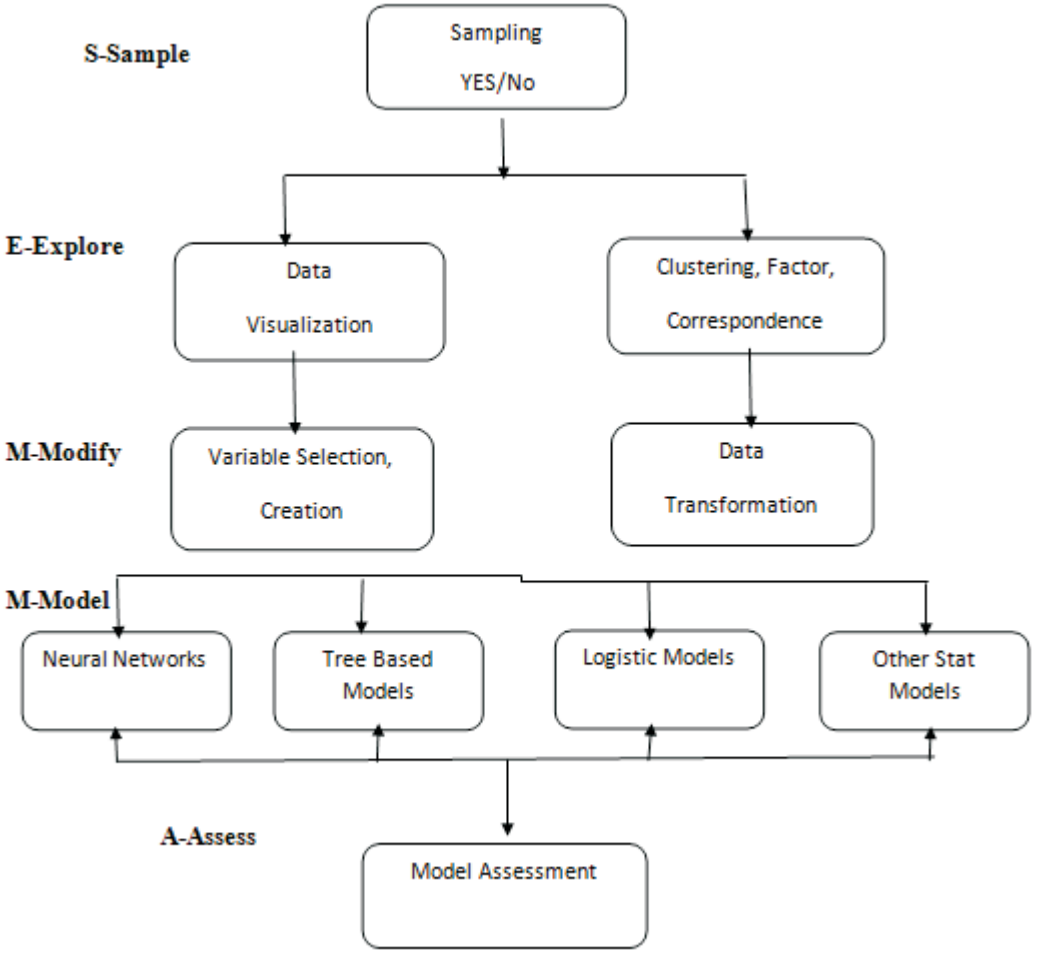
INTRODUCTION:

Objective of Data Mining:

The concept behind Data Mining can be understoond through the following simple analogy. The growth of Data Warehousing has created a huge amount of data which is a valuable resource to the Enterprise. Retrieving required Data Set from these huge Data Warehouses and Data Marts is a challenge and a scope to Data Mining. Data Mining is a traditional data analysis methodology updated with the most advanced analysis techniques applied to discovering previously unknown patterns. It

can be defined as the process of selecting, exploring, and modeling large amounts of data to uncover previously unknown patterns for a business advantage.

Data Mining involves a five step **SEMMA** process which can be abbreviated as



Knowledge Discovery in Database:

Data Mining Definition - 1: In general, data mining can be defined as the process of multi-dimensional analysis of data in order to consolidate it into improvised information which can be used to increase profit margin of the business, cost reduction, or both. It is a software used for data analysis. The major purpose of this technology is to derive a relationship between multiple dimensions of data. In technical terms, data mining is the process of finding bits and patterns of processed data in a given large relational databases.

An automatic Data Miner without Human Intervention

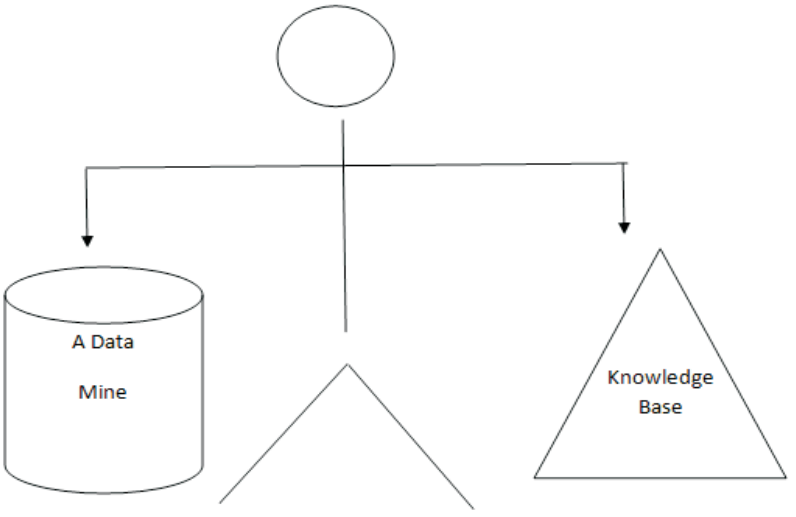


Fig: Automated Data Mining represetnation

Data Mining Defnition-2: Data Mining is an automated activity of extracting hidden patterns and relationship between those patterns from huge databases. In short, with less manual involvement analyzing data and dividing it as processed information. In this view, data mining is knowledge discovery in databases, or at least it is automated knowledge discovery in databases. The difference between automated discovery of patterns and manual discover of patterns are, the hypothesis patterns get limited by the skill and experience of humans, while the data mining tools are independent of manual involvement and mostly dependent on pattern-matching algorithms.

Data Mining Defnition-3: Data Mining is the first and foremost initiative in knowledge discovery in databases, that inputs predominantly processed patterns of data, which analyses the data in an automated way with more accuracy, and outputs those patterns and relationship between the patterns to the exploration or enhancement of the KDD process. This definition clearly implies that what data mining (in this view) discovers is hypotheses about patterns and relationships. Those patterns and relationships are then subject to interpretation and evaluation before they can be called knowledge.

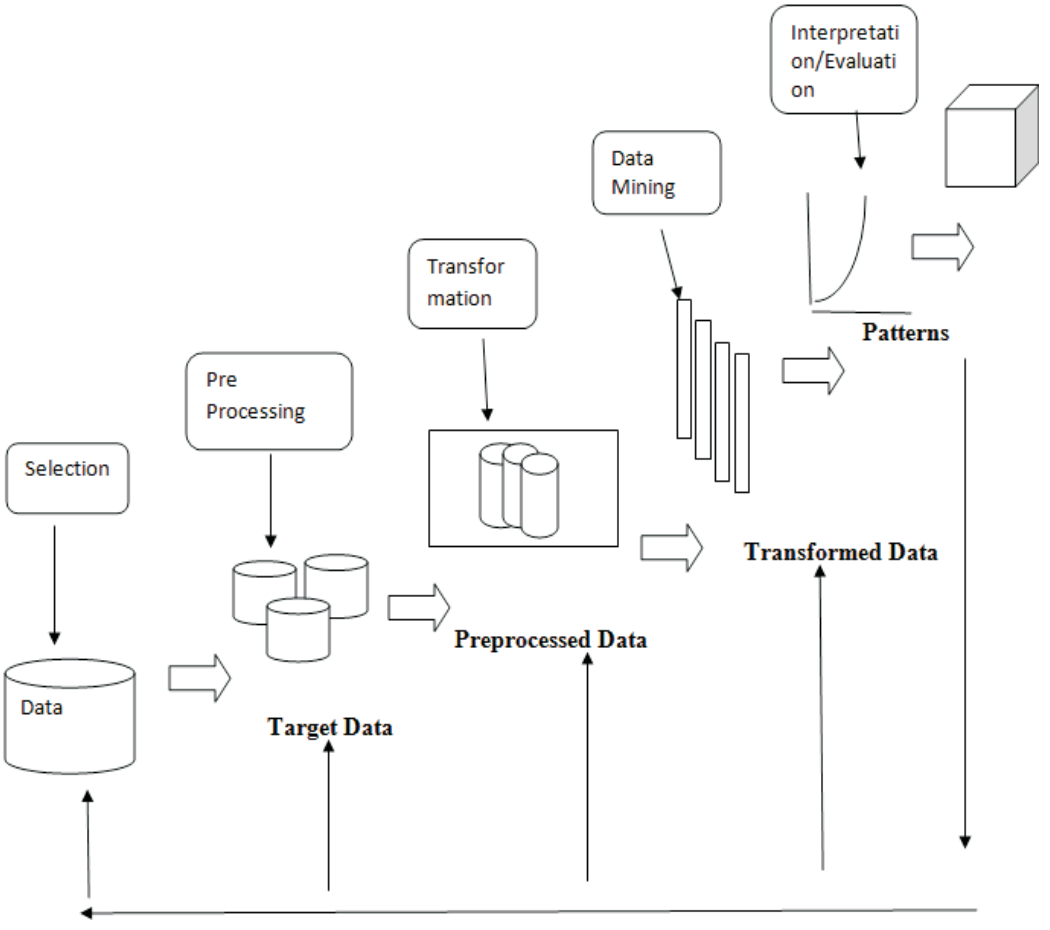


Fig: Data Mining in KDD

The Idea of knowledge Discovery in Databases (KDD):

There are three distinct definitions for KDD.

KDD Definition 1: Knowledge discovery in databases is a process that requires a thorough analysis whether that model is relevant or not, and model testing is a kind of specification test which will test the proposition whether the model is relevant or not, and down the line all the existing data models are necessary to bring to a successful conclusion. This kind of analyzing data includes both exploratory, and confirmatory data analysis.

KDD Definition 2: An advanced knowledge discovery in databases can be formed when there exists an automated process of data analysis with less or no manual intervention. Of Course, this process doesn't show much difference between Data Mining and KDD. Business with no loss can be assured only when we implement such an automated technology, because Data Mining doesn't just give a rough idea, it will provide knowledge in Data Analysis.

KDD Definition 3: "Knowledge discovery in databases is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data. Further, this process includes five steps: data selection, data preprocessing, data transformation, data mining, and interpreting and evaluating mined patterns and relationships.

How Data Mining relates to KDD:

As there are three different definitions for Data Mining and KDD, there are three different approaches to explain the relationship between Data Mining and KDD.

Traditional Data Mining:

• This approach can be explained as one of the foremost and oldest methodology of Data Mining.
• In this approach there is no much difference between Data Mining and KDD. As this approach follows statistical methods such as SEMMA Data Mining process.
• A few tools such as Neural Networks and Tree based models may be different but the approach used to investigate the patterns and validation of those consolidated patterns are the same.

Automated Data Mining:

• This approach is a mixture of both manual process and also the automated process.
• It is manual in a way that the analysis of data, data patterns and deriving relationship between these patterns can be done by all manual processes.
• Automation comes into picture to select a specific pattern which will empirically fit to a particular product, because products may vary in the extent of scope of their validation criteria.
• Apart from the fitting patterns, major challenge is out of several models of existing data mining models, which is well suited for your Data Mine?

CONCLUSION:

• Data mining serves as a part of KDD. This is an approach which attempts to distinguish Data Mining and KDD with fully automated process.
• The difference is in the area of exploratory data analysis where the experts decide which of the automated method to be used to decide the pattern.
• But the key point that exists is the cyclic process used to prepare data mining and implement it follows the careful investigation of traditional analysis.

REFERENCE(S):

1. www.ask.com ,
2. www.wikipedia.com
3. www.gartner.com
4. www.wikibon.org
5. www.computerworld.com
6. www.mckinsey.com
7. https://www.ibm.com
8. www.google.com
9. http://www.cs.uic.edu
10. Introduction to data Mining with Case Studies by G.K. Gupta
11. Data Mining Theory and Practice by K.P. Soman

Publish Research Article
International Level Multidisciplinary Research Journal
For All Subjects

Dear Sir/Mam,

We invite unpublished Research Paper,Summary of Research Project,Theses,Books and Book Review for publication,you will be pleased to know that our journals are

Associated and Indexed,India

- International Scientific Journal Consortium
- OPEN J-GATE

Associated and Indexed,USA

- Google Scholar
- EBSCO
- DOAJ
- Index Copernicus
- Publication Index
- Academic Journal Database
- Contemporary Research Index
- Academic Paper Databse
- Digital Journals Database
- Current Index to Scholarly Journals
- Elite Scientific Journal Archive
- Directory Of Academic Resources
- Scholar Journal Index
- Recent Science Index
- Scientific Resources Database
- Directory Of Research Journal Indexing

Indian Streams Research Journal
258/34 Raviwar Peth Solapur-413005,Maharashtra
Contact-9595359435
E-Mail-ayisrj@yahoo.in/ayisrj2011@gmail.com
Website : www.isrj.org