A Review on Data Mining & Big Data, Machine Learning Techniques

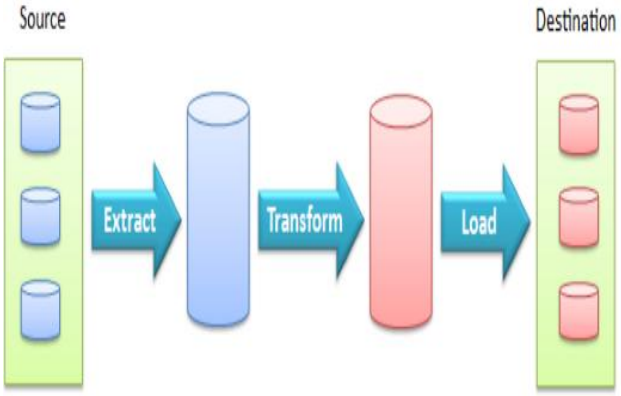
J.Nageswara Rao, M.Ramesh

Abstract: ***Now a day's large amount data stored from different data sources, which is increased based on KDD from various data sources . in the direction of acquire necessary along with useful facts from data sources, some of the techniques and tools to combine the vast amount of data sets. the main aspire of data mining is to mine necessary information from huge amount of data and to retrieval information. in data mining classification and clustering main techniques to organize and come together clear-cut data in a big set of data into essential collection sets of group labels. we present a complete analysis of various clustering and categorization approaches in data mining for capable of information recovery, which includes NN, BN and decision trees.***

key words: Clustering, Supervised, Semi Supervised, KDD, Categorical Data

I. INTRODUCTION

Information Retrieval(IR) is the group of required data assets related to an information need from a group of data resources. IR is the to explore text or content based indexing. It is the knowledge of search for data in a article, and also finding for metadata that describe data, and for databases of images or sounds, texts. AIR systems are used for to reduce information overloaded. for the period of data retrieval, information to mined among subsequent two ML Technique's .the general procedure data extraction of data recovery is shown in Fig.1



As exposed in Fig.1. procedure for a collecting data from different data sources, there are performing various levels of operations such like data selection, pre-processing, data transformation. and information recovery based on data interpretations on user facts. in this article ,we give analysis of various DM and ML techniques.

Revised Manuscript Received, 15 April, 2019

Mr. J.Nageswara Rao, Research Scholar, Acharya Nagarjuna **University,** Andhra Pradesh, India.

Dr. M.Ramesh, Associate Professor, Dept. of IT, RVR&JC COE, Guntur, Andhra Pradesh, India

II. BACKGROUND

In data mining various Authors defined data collecting from various data sources and also discuss uncompromising data explanation with various applications in real time . now a day's traditional approaches using **Data Mining And Big Data ,Machine Learning Techniques**.

DM is the process of discovery and analysis ,with regular or semi automatic means, of vast of quantities of information**[1]**. It is used for exploring and analyzing large amount of data to find patterns for Large Data. In recent years all companies and Marketers collect all the information of business deals stored in a Large Data Base. it requires the functioning of new approaches ,new technology and governance that are collectively being referred to as Large Data**[2].** The main aim of Data Mining is either prediction or classification, clustering. in prediction , to forecast a value Example. Business persons or Atmosphere forecasting. In classification is to exactly estimate the objective class used for every case in the information. for e.g. Top student, Average Student ,Below Student. In clustering grouping the data.

3. Various Techniques Used In DM , Big Data, ML and Deep Learning

A. Data Mining & Big Data

I. Classification tree

it is popular method that is used to categorize a dependent qualified variable based on dimensions of one or more forecaster variables. The outcome is a tree with nodes and associates between the nodes that can be read to form if-then rules.

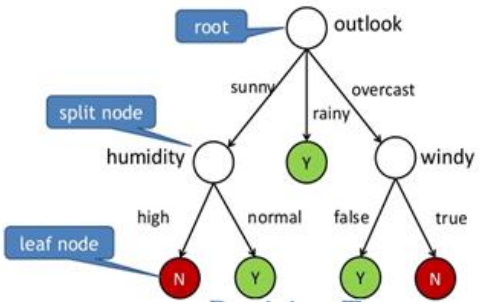


Fig.2Classification tree

III. NEURAL NETWORK

It is a set of approaches and comparable architecture of various animal ,human brains. The system contains Input , Hidden, Output layers. Each layer assigned a particular weight. information is given to the input node, and by the network of trial and error, the method adjust the weights until it reach a certain ending criteria.

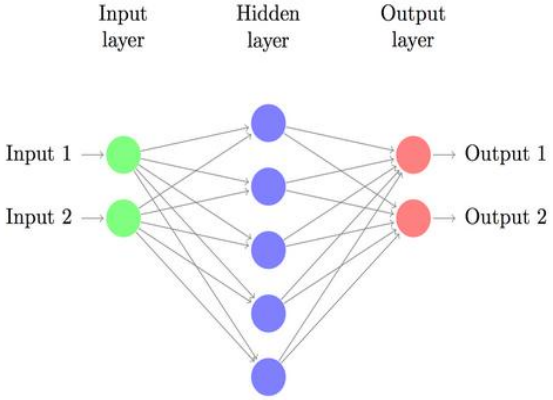


Fig.3.Neural network

IV. CLUSTERING METHODS

It is a technique grouping the similar data .in clustering K-NN technique is find the distance records and points in the historical data.

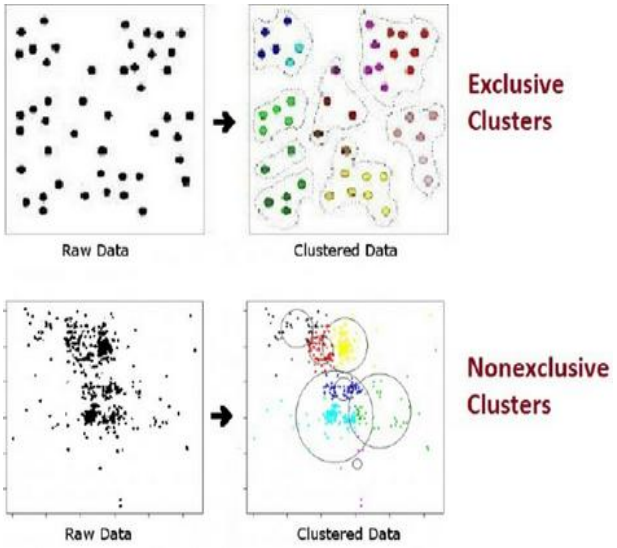


Fig.4.Clustering

B. Machine Learning & Deep learning

It is branch of the Artificial Intelligence. the main goal of Machine Learning to focus on various real time direct experience ,in charge to appear for patterns in information and create enhanced decisions in the prospect based on the examples. Basically ML Classify 4 different ways. i. Supervised ,ii. Unsupervised, iii. Semi-supervised iv. Reinforcement.

i.a.Supervised Learning again divide into **2** categories of algorithms

- **Classification:** it is when the output of a variable is a group, such as "Male" or "Female" and "Human" or "Animal".
- **Regression:** it is when the outcome variable is real value, such as "dollars" or "weight".

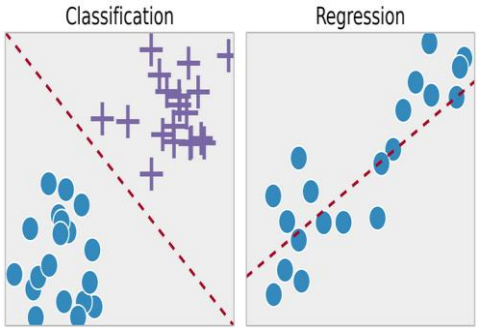


Fig.5.Classification & Regression

i.b. Unsupervised learning classify into **2** categories of algorithms

- **Clustering**: it is nothing but grouping the similar data. For example, in the above example each customer is put into one group out of the 10 groups.
- **Association:** it is "relationship between people that buy X also tend to buy Y "

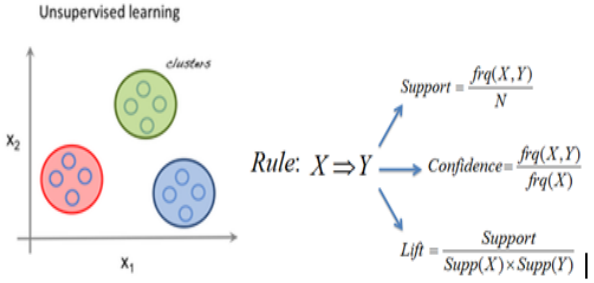


Fig.6.Clustering &Association

i. c. Semi-supervised learning:

It is mainly considers the difficulty of when only a very small subset of the remarks have corresponding class labels .it uses a small amount of labelled bolstering a larger set of unlabelled data.

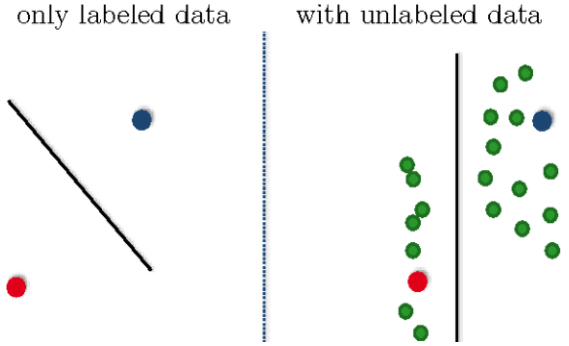


Fig.7.Semi-supervised learning

i. d. Reinforcement learning

it is creation decision consecutively. in easy style we can say that O/P depends on the state of the present I/P, and next I/P depends on the O/P of the earlier input. for example: "Chess Game"

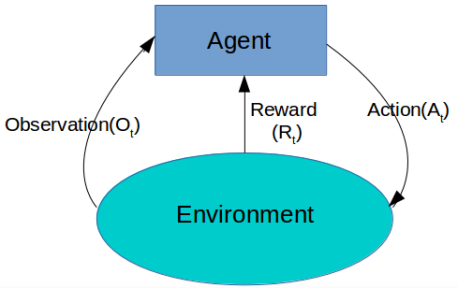


Fig.8.Reinforcement learning

4.Main issues , challenges of Data mining & Big Data

Big data is stored a huge amount of information. the exploring a large amount of information , testing useful information from various data sets. it is also unstructured data, large size and it is not easy to handle.

S.no	The main issues	Major challenges
1	very poor data quality. for example noisy data, dirty data	Big Data Mining Platform
2	Redundant data is uploaded from various sources	Information Sharing and Data Privacy
3	Security, privacy of the companies	Domain and Application Knowledge
4	Higher cost, less flexibility	Big Data Mining Algorithm

Table.1

V. CONCULSION

Now a day's all Researchers and IT sector Professionals , Engineers working on Big Data, ML and Deep Learning many persons are proposed a various system models, approaches for big data. The high performance computing hypothesis is necessary for data mining to solve the problem of big data. We conclude that there are still probability to improve various algorithms and techniques for data mining . in this paper review on Data Mining & Big Data ,Machine Learning Techniques.

REFERENCES

1. Xingquan Zhu, Ian Davidson, "Knowledge Discovery and Data Mining: Challenges and Realities", ISBN 978-1-59904-252, Hershey, New York, 2007.
2. Chen M, Mao S, Liu Y. Big data: a survey. Mobile Networks and Applications. 2014; 19(2):171–209.