A Review On Big Data With Machine Learning And Fuzzy Logic For Better Decision Making

J. Betty Jane, Dr.E.N.Ganesh

Abstract:-Big data is playing an important role for many years due to the vast and vide growth of data in various fields and it has become difficult to find the sense from all those large collective data. Since there is an increase in the large amount of data it is difficult to make better decisions .To overcome these challenges, machine learning and fuzzy logic are the two techniques that are used for getting better knowledge from various and such large amount of data sources. This paper discuss about the review on knowledge based decision making that are derived from these two techniques and how efficiently these two techniques provides the decision making in their own way and the steps involved in these techniques for decision making. Machine learning is a technique of artificial intelligence to give better intelligent decisions. Fuzzy logic is a process of giving conclusion or decisions from an uncertain problem.

Keywords: bigdata, fuzzy logic, machine learning

———————————————— ◆ ————————————————

1.INTRODUCTION:

Nowadays the data acquired from various sources is very large that are generated from the modern equipment such as sensors, smart phones and internet [1]. The smart sensors are used to record the activities that are being fixed in the cities and in some other organizations for finding the activities in those areas. The challenging part of the big data is to find the useful information from these large amounts of data for future developments or improvements. Hence the knowledge in the extraction process has to be efficient and it has to be an approachable technique for the real time since the storage of the data is high in many situations. The HACE theorem states that the huge-volume, autonomous sources with distributed and decentralized control and which help in exploring complex and generating relationships among the data. Hence these features made a challenge for extracting new useful information from these large Sources. There are two techniques involved for data extraction, in that machine learning tools are used to develop the domain knowledge through the learning. This tool helps in successful extraction of needed information by the learning process [2]. The machine learning involves various algorithms for discovering hidden features and patterns from recent trends and data. The ML algorithms are used in various areas and in detecting spammed and unspammed mails by introducing it into the ML algorithms.[3] A large change is been achieved with the help of fuzzy logic. After the emergence of fuzzy logic, the work has become easier, saves money, time and energy. LottiZadeh is the one who proposed fuzzy logic in 1965. Before fuzzy logic there were, three ,four valued logic which has helped for the extension of Boolean logic which helps in accepting only values like true or false(0 or1).The Boolean logic gives only two values one is 0 or 1.that is low or high.no other values can be found in this logic. Hence to find the most accurate values in the scenario we go for fuzzy logic which gives the extended view such as low, very low, high, very high also. [4].

————————————————————

- *Author name: J..betty jane (research scholar,vels university)E-mail: author_name@gmail.com*
- *Co-Author name: Dr.E.N.Ganesh (Dean,vels university) E-mail: enganesh50@gmail.com*

2 .BIG DATA PROCESSING:

Big data is the term used for huge and large datasets where the data processing traditional application software's are not able to deal with these data. The term big data emerged in the year 2001 and defined by Doug Laney an analyst [5]. Big data is defined as a huge volume, variety and velocity of data that has challenges such as enhanced and detailed insight and mainly decision making. Big data is classified in to 8V'S such as volume, velocity, veracity, value, variability, volatile, validity. Value means statistical, correlational values. Volume includes extra bytes and zeta bytes. The fast generation of data and speed flow of data is known as velocity. Validity gives correct data and accurate data. The structured and unstructured data comes under variety. Variability is denoted by the variation in the data speed rates. The validation of data comes under volatility. Veracity denotes the trustworthiness in the data.
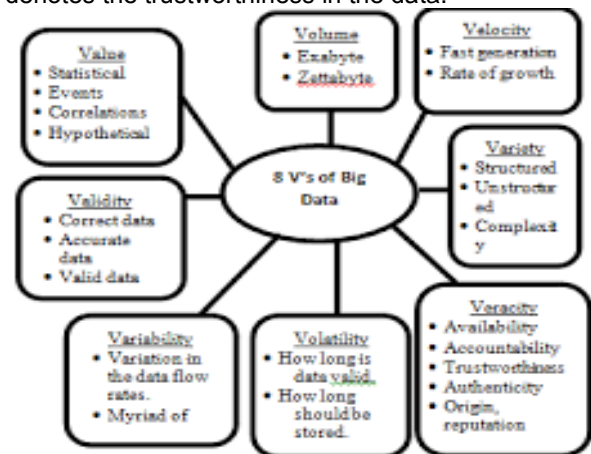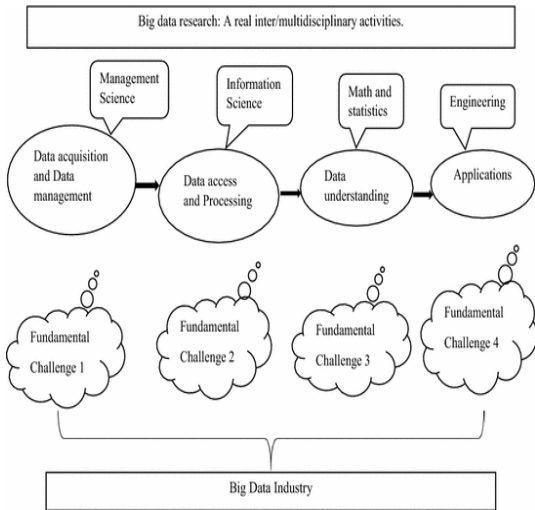

Fig 5. 8 V's of Big data
***Fig a:8v's** of big data*

A. Big data applications:

In this era, the organizations produce large amount of data in the form of Exabyte and zettabytes. various fields such as

1. Healthcare
2. Education
3. Media
4. Information technology

1121

Fig b:*big data processing*

B. Big data challenges:
There are challenges in big data and these challenges are
1. Dealing with the growth of data.
2. Validating data.
3. Securing large volume of data.
4. Highly cost effective.
5. Massive data sets integration.

C. Big data analysis tools:
The data is not useful until it gives Data that is unstructured for the management to do decision making. To enhance the decision making by stages we use some software data tools for storing and analyzing the data.   Some of the tools are Apache Hadoop, cloudera, Cassandra, mango DB etc.
1. Apache Hadoop- This framework is used for clustered file system. Hadoop consists of two main components: The HDFS for the storage big data, and Map Reduce for big data analytics [6]
2. Cloudera –collection and processing of unlimited data.
3. Cassandra– managing high volumes of data.
4. Mongo DB- MongoDB is written in C, C++, and JavaScript and used for document-oriented database.

3.MACHINE LEARNING FOR DECISION MAKING:
The paradigm of learning or understanding the past experiences and with that data, trying to improve the future performance is known as machine learning.[7].

D. Why we go for machine learning method for decision making:
The machine learning provides alternative solutions from the large amount of vast data by developing algorithms for processing real time data and gives the accurate results and analysis.The machine learning aims on the computer program development that can access the data and they can learn on their own.
E.ML Algorithms for decision making:
The main advantage of ML algorithms is, once the algorithm it finds or understands how to analyze the data, it will do its work automatically. In case of large date volume, human learning will not be that much effective. Hence we

go for machine learning where machine learning is nothing but teaching the machines how to handle the data.[8]
F. Types of machine learning:
The types of machine learning are:
i.   Supervised learning, ii. Unsupervised learning iii.reinforcement learning
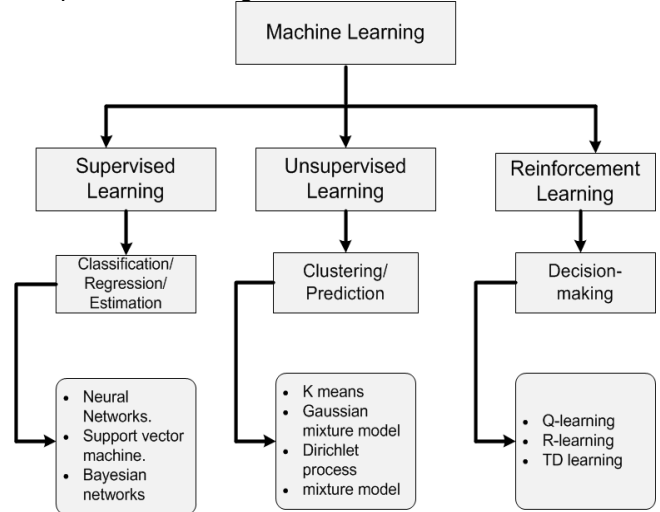i. supervised learning:

Fig c: Types of Machine learning

The algorithms that need external support are called supervised learning algorithm. The given input dataset is splitted as training data and test data. The trained dataset with output variable needs to be classified and predicted. All these learning algorithms find some pattern from the training dataset and then apply it for the classification or prediction process.
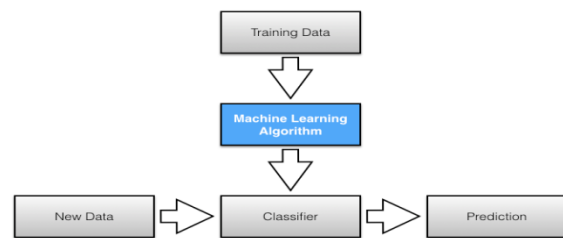

Fig d: *machine learning algorithm processing*

There are two types of supervised learning they are
i. regression
ii.Classification
Under regression there are three decision making algorithms:
They are
a. Decision tree:
Decision trees are trees that form a group by sorting based on their given values. For the purpose of classification purpose it is used. The decision tree consists of branches and nodes. The branches represent the node value and the nodes represent the attributes to be classified.
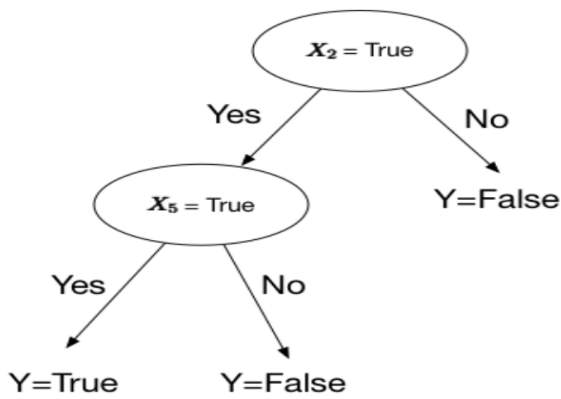
Fig e: Decision tree

b. Linear regression:

A machine learning linear regression algorithm is a supervised learning model. The performance of this method is doing the task of regression.Regression provides a target prediction values on independent variables. It is used to build and finding out the relationship between prediction forecasting and variables.
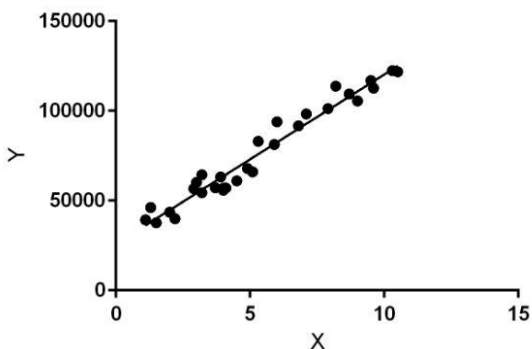

Fig f: Linear regression

c. Logistic regression: Logistic regression is a supervised classification algorithm. It is a predictive analyzer. It transforms the output values using the function called sigmoid and return the value of probability.
Under classification, there are three decision algorithms:
They are

d. Naive bayes algorithm:
Bayesian classification algorithm, is a Bayesian theory and it is also called as the statistical algorithm. Bayesian classifier is developed from a training dataset with labels of classes. Assuming n is an attributes A1, A2,…, An, an instance E is a vector representation , where ai is the Ai value, C is used as a class variable representation, c is the value of C, and c(E) represents the class label E where it belongs to. [9].
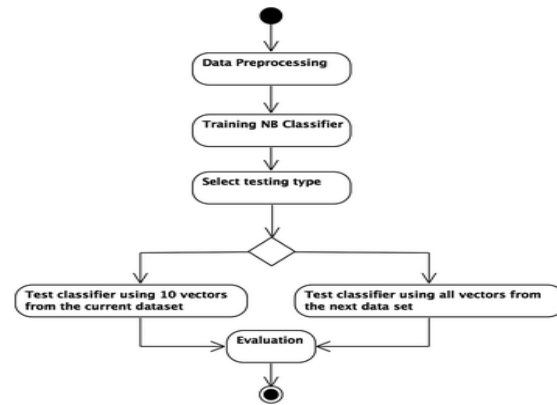

Fig g: Naïve bayes algorithm

e. support vector machine:

Support vector machine is a linear division method. The division is done as two categorical points that divide their points into two-dimensional space, SVM is used to map the low-dimensional points in the curve to the high dimensional space and hence they are separable linearly.[10]
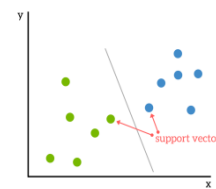

Fig h: support vector machine

f. k-nearest neighbor algorithm:
The KNN algorithm is an efficient classification algorithm. [11]They mainly adapt three rules:
They are:
Rule: 1 move the k-nearest code vectors, if the code vector matches the sample training set, then the code vector matches the training vector.
Rule: 2Only the k'th and the k + l'th nearest vector codes are moved
Rule: 3The k + 1nearest code vectors are moved.

ii. Unsupervised learning:
Unsupervised learning is an artificial intelligence algorithm in which the input sets are classified without any proper guidance. Under unsupervised learning there are two classifications clustering and reduction. Under clustering, the variables are grouped into clusters and they may be either user defined or model based one.

a. Cluster analysis:
Cluster analysis is the process of grouping the clusters with similar data. There are different types of clustering such as c-means, hierarchical, iterative clustering.[12].
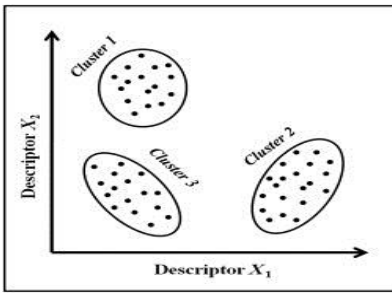
1123

Fig i:cluster analysis

b. Dimensionality reduction:
The main aim of dimension reduction algorithms is to reduce the noise or irrelevant information of the data.[13] There are algorithms for dimensionality reduction they are as follows:
Principal Component Analysis (PCA),
Non-negative Matrix Factorization (NMF),
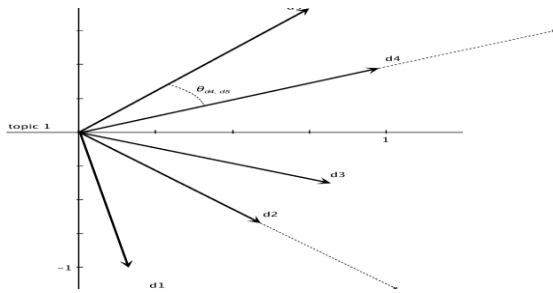Random projection (RP) and auto-encoder (AE)

Fig j: Dimensionality reduction

iii. Reinforcement learning:
Reinforcement learning is an artificial intelligence branch in which allows automatically determining the behavior with in a small context to the machines.

IV. FUZZY LOGIC FOR DECISION MAKING:

Fuzzy logic is a "degrees of truth" approach rather than the "true or false" (1 or 0). The idea of fuzzy logic was first introduced by Dr. LotfiZadeh from University of California in the 1960s.Fuzzy logic gives the decisions as same as that of human perception and reasoning. Fuzzy logic has been proved to be applied in well expertise systems. The creation of fuzzy sets help in determining whether the element belongs to the set or not ranging from 0 to 1.It is used for making the decisions in the terms of uncertainty.[14,15]

a. Fuzzy sets:
Fuzzy sets are crisp sets where the characteristic function are transformed to the membership function A: $X \rightarrow [0, 1]$

b. properties of fuzzy sets:

Operation	Crisp	Fuzzy
Addition	$A+B$	$\tilde{A}+\tilde{B}=\left[A_1^{(\alpha)}+B_1^{(\alpha)}, A_3^{(\alpha)}+B_3^{(\alpha)}\right]$
Subtraction	$A-B$	$\tilde{A}-\tilde{B}=\left[A_1^{(\alpha)}-B_3^{(\alpha)}, A_3^{(\alpha)}-B_1^{(\alpha)}\right]$
Multiplication	$A \cdot B$	$\tilde{A} \cdot \tilde{B}=\left[A_1^{(\alpha)} \cdot B_1^{(\alpha)}, A_3^{(\alpha)} \cdot B_3^{(\alpha)}\right]$
Division	$A \div B$	$\tilde{A} \div \tilde{B}=\left[A_1^{(\alpha)} \div B_3^{(\alpha)}, A_3^{(\alpha)} \div B_1^{(\alpha)}\right], \text{if } 0 \notin \left[B_1^{(\alpha)}, B_3^{(\alpha)}\right]$

c. Algorithm for fuzzy decision making:
The steps involved in fuzzy decision making are as follows:
i. In the first step is identification of variables and the alternatives is done.
ii. The fuzzification process involves converting the linguistic ones from real variables.
iii. The variables that need to be taken in the knowledge base are selected by the user.
iv. The membership function expressed in mathematical function is called membership function
v. The next step is to give the if- then condition rule. Each variable represents one rule.
vi. The next step is getting output variable by converting the fuzzy value.
vii. The last phase of fuzzy process is the implementation of alternative in practice. if the implementation is Successful it will improve the system operation with respect to the aim of the process.
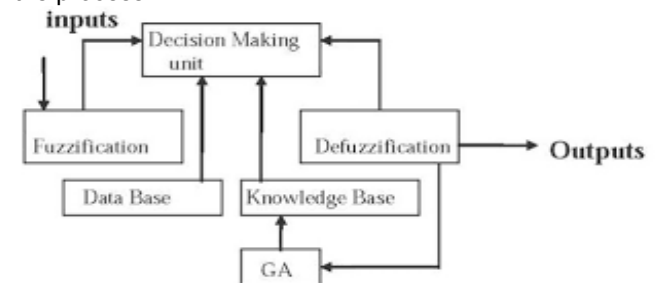

Fig k: fuzzy logic decision making

V.COMPARITIVE REVIEW ON FUZZY LOGIC AND MACHINE LEARNING FOR DECISION MAKING:

Machine learning algorithms aims at extracting knowledge from large amount of data and gives traditional methods for classification and clustering process. It handles multi variety of data and they can be used in large environments. Moreover, learning time is needed more for the algorithms to perform the progress in accuracy and relevantly. Fuzzy logic measures the certainty of the problem, and the algorithms are robust and they adapt easily to the changing environments.

VI.CONCLUSION:

The machine learning techniques for decision making produce good results by handling large data environments and give good ideas to the experts in different fields to improve the future enhancements in the fields they are involved. Fuzzy logic also equally helps in finding out the uncertainties in the problem and they adapt themselves to the changing environments and also help in decision making. In this paper, the review gives the understanding that machine learning is an emerging field and its

performance is well compared to fuzzy logic. Hence to improve the fuzzy logic performance we can enhance the fuzzy extensions on ML algorithms.

REFERENCES:

[1] Vınod Jha, K.Parvathi "Complete Forward and Reverse Text Transcription of Odia Braille and Hindi Braille", : Progress in Advanced Computing and Intelligent Engineering, Springer, sequence no.12,201

[2]].Machine Learning for Big Data Processing:A Literature Review:Bharat Kumar Padhi1 , Dr. S SNayak2 , and Dr. B N Biswal3,December 2018 | IJIRT | Volume 5 Issue 7 | ISSN: 2349-6002.

[3] A Very Brief Introduction to Machine LearningWith Applications to Communication Systems,King's College London, United Kingdom ,Osvaldo Simeone,IEEE.

[4]].Machine learning and its applications: A review,shhenaangra,sachinahuja,@2017

[5] Application of fuzzy logic: A literature review, RenuMakkar,International Journal of Statistics and Applied Mathematics 2018; 3(1): 357-359

[6]].A Review on Big Data theoretical and Application approach:Past and Present: RabelGuharoy, SumitKumar, Upasak Pal;©2017 IEEE.

[7] Big Data Analytics: A Literature Review PaperNadaElgendy and Ahmed Elragal,Department of Business Informatics & Operations,ICDM 2014, LNAI 8557, pp. 214–227, 2014.© Springer International Publishing Switzerland 2014

[8] A Survey on Machine Learning: Concept,Algorithms and Applications,Kajaree Das1Rabi Narayan BeheraInternational Journal of Innovative Research in Computerand Communication Engineering,Vol. 5, Issue 2, February 2017.

[9] Machine Learning Algorithms: A Review AyonDey; International Journal of Computer Science and Information Technologies, Vol. 7 (3) , 2016, 1174-1179.

[10] A novel Naive Bayes model: Packaged HiddenNaive Bayes,YaguangJi,Songnian Yu,Yafeng Zhang; ©2011 IEEE

[11] THE RESEARCH OF THE FAST SVM CLASSIFIER METHOD;YUJUN YANG 1,2,3, JIANPING LI 1, YIMEI YANG ©2015 IEEE;

[12] Classification with Learning k-Nearest Neighbors;JormaLaaksonen and ErkkiOja ,1996 IEEE

[13] A Research Study on Unsupervised Machine Learning Algorithms for Early Fault Detection in Predictive Maintenance,NagdevAmruthnath,Tarun Gupta,2018 5th International Conference on Industrial Engineering and Applications.

[14] Dimension Reduction with Extreme LearningMachine LiyanaarachchiLekamalageChamaraKasun, Yan Yang, Guang-Bin Huang and Zhengyou Zhang .IEEE

[15] The Role of Fuzzy Logic in Decision Making Process;KatarínaValášková1,a,*,TomášKlieštik2,band Mária Mišánková;2nd International Conference on Management Innovation and Business Innovation.

[16] Fuzzy Based Machine Learning: A Promising Approach; Sujamol S., Sreeja Ashok and U Krishna Kumar, CSI COMMUN I C A T I O N S | N O V E M BER 2017