ISSN 2455-4863 (Online) **www.ijisset.org** *Volume: 5 Issue: 7 | 2019*

An Overview on Machine Learning

Ramya.S.K[1], Jyothi Lakshmi G Kava[2], M P Nayana[3]

[1]*Assistant Professor of Computer Science, SDM MMK MMV, Mysuru, Karnataka, India*

[2]*Assistant Professor of Computer Science, SDM MMK MMV, Mysuru, Karnataka, India*

[3]*Assistant Professor of Computer Science, SDM MMK MMV, Mysuru, Karnataka, India*

Abstract: *The Various developments in Research area of Computer Science has led to one more branch of learning that is the Machine Learning(ML) which is the Emerging Technology. The Computers or Systems behave as independent learners in Machine Learning. The Methodology involves a Machine becoming more intelligent which is capable of handling problems by improving its programming capability based on the requirement of the application using the present/previous data without intervention of Humans. The Machine automates its intelligence using a wide range of Machine Learning Algorithms. Machine Learning can be defined as the self learning of Machines for producing better results. This paper focuses on Algorithms of ML.*

Keywords: *Machine Learning(ML), Independent Learning, Intelligent Learning, Classification, Clustering, Supervised Learning(SL), Unsupervised Learning(UL).*

1 INTRODUCTION

Machine Learning is the field inherited by Artificial Intelligence where in the Machine is made to learn by itself depending on the data sets available in the domain of interest. The decisions in Machine Learning are driven by data. Machine Learning came into existence in the early 90s and it relates to the field of data science. Humans learn from past experience whereas the Computers/Machines follow the instructions given to it by the user. Now the research is made wherein the Machines learn from the previous data and can act independently when exposed to new data without waiting for the users instructions frequently by developing its own programs. This exposure to new data is analyzed by Machine and tries to improve its computing power and make predictions accordingly. Machine Learning makes use of Probability and Statistical concepts and many more concepts to make itself intelligent and this intelligence level are coping up to solve the present real world application problems. In this area, the basic features/basic data of a particular instance will be given as input to a Machine and is trained with large training data sets, tested with large testing data sets. In Machine Learning, a Machine changes its behavior depending on the data and act accordingly and is capable of handling big data which is a challenging task. Over the years, data is increasing day by day in WWW and to handle such big data problem, machine is made intelligent. The data may rise up to 44 Zeta Bytes by 2020. ML is closely related to data mining and it gets associated with many more branches of Computer Science such as text interpretation, Pattern Recognition. This branch of Computer Science has emerged from smaller end applications to higher end applications.

The rest of this paper is organized as follows. Section II will be the Literature Survey. Section III familiarizes the languages used in ML. Section IV focuses on Methodology of Machine Learning Section V gives the conclusion.

2 LITERATURE SURVEY

This section reveals a brief knowledge about the research papers on Machine Learning.

[1] In the paper "A Survey on Supervised Classification Techniques in Machine Learning", the authors have noticed the Brute force method to handle and analyze data. The literature review of the paper focuses on maintaining the mining quality, avoiding irrelevant and redundant features. The task of quality maintenance of data during data mining improves the efficiency of data mining, which in turn improves the efficiency of ML algorithms. The paper has focused light on logic based algorithms such as the decision trees(DT), Perceptron based techniques, statistical learning algorithms, instance based learning, Support Vector Machines(SVM). The ML algorithms such as the decision trees, neural networks, naïve bayes , kNN, SVM, Rule learners are compared in the paper based on the issues such as Accuracy of the algorithm, Fastness of the algorithm, Tolerance level to irrelevant data, Tolerance to noise, Attempts of individual learning and many more criterions.

[2] In the paper "Machine Learning Wearable device information in Parkinson unwellness health watching", the authors have made known about the Wearable sensors in the medical field to recognize the symptoms of Parkinson disease. Parkinson disease is a disorder of the nervous system in Humans. Nerve cells will be dead in the brain in a Parkinson diseased patient. It is a neurodegenerative disorder. The Accelerometer , Gyroscope ,Inertial sensors and microphones are used as sensors to monitor the patients behavior. The wearable accelerometer sensor are tri-axial in nature.

The Zigbee protocol is used for the purpose of data collection by patients through sensors at the rate of 62.5 characters per second(cps). The ML Algorithms used in this paper are Artificial Neural Network(ANN), Induction decision tree version 3(ID3),Call tree induction(c4.5/j48) which uses divide and conquer approach, Classification and Regression Tree(CART) and the DT Algorithm. The Paper has compared the accuracy and precision, and also the error rate of all the algorithms where DT is showing optimum results in the paper. Based on the predications of these algorithms, the intensity of nerve cell damage in the human brain is identified and immediate action can be taken by doctors to avoid further intensifying of the disease.

[3] In the paper "Machine Learning Algorithms : A Review", the author has made known about the ML algorithms such as Supervised Learning, Unsupervised Learning, Semi Supervised Learning, Reinforcement Learning, Multi-task learning, Ensemble Learning, Neural Network, Instance based learning and the classification of these algorithms and the sub classifications. The paper consists of the pseudo codes of several ML Algorithms which is a step way for the researchers to build additional code upon those pseudo codes to improve the performance of the algorithms further.

[4] In the paper "Supervised Machine Learning Approaches: A Survey", the paper has detailed about Supervised Learning Algorithm,Logical based algorihms, Statistical learning algorithms, instance based learning algorithms, SVMs and Deep learning. The authors have made known about the Algorithm selection based on the Accuracy which is calculated using the formula given below

A1=Number of correct classification

A2=Total number of test cases

Accuracy=A1/A2

In addition to Accuracy of the algorithm, Cross validation(CV) or Rotation Estimation approaches are used for better usage of the sample. Under Logic based algorithms, the authors have described about DT and Learning set of rules. DT was designed to handle classification problems. The merits of DT algorithm are it is capable of producing better results, holds organized knowledgeable structure and easy to understand. In the rule based method, Disjunctive Normal Form (DNF) is used to represent cluster concept. One of the well known statistical learning algorithm is the Bayesian Network. Bayesian networks make use of Directed Acyclic Graph(DAG). Instance based learning works depending on Nearest Neighbor Algorithm. Statistical Methods are used not only for classification but it also provides a gateway for instances to belong to a particular class. The paper has mentioned about Linear discriminate Analysis(LDA) and the Fishers Linear Discriminate which are used in statistics and Machine Learning to classify the objects in the sample. The paper discusses about SVMs which are used for classification, Regression and outliers detection. Deep Learning is a subset of Machine Learning. This learning creeps in depth to a neural network . The network consists of several nodes and links. The links are capable of transferring data from the source node to the destination node with high level of security and the network makes use of Advance Machine learning concepts.

[5] In the paper " A Survey of Machine Learning Approaches to Analysis of Large Corpora", the authors are concentrating on Natural language processing. The paper is accomplishing three different types of classification namely linguistic level of analysis, secondly a new approach to machine learning which is applicable for the linguistic annotation of corpora. Corpora or text corpus are large set of texts in an organized manner. Finally the paper has discussed the most challenging levels of linguistic annotations. The paper briefs about Tokenization of text which involves splitting of text into tokens/words. The space character is considered for the purpose of tokenization. Next phase discussed in this paper is about part of speech tagging which involves recognizing noun, verb, adjective, adverb. PoS-taggers are used to identify grammatical feature such as singular, plural, number, tense, gender. The process of parsing is also described in the paper. In the parsing process a parse tree data structure is generated. This tree is capable to parse a new sentence by finding optimal combination of subtrees. In Semantic Analysis process, the text makes use of thesaurus class for the purpose of documentation classification and Management. Discourse analysis deals with learning of sentences by machine beyond sentence boundary i.e naturally occurring language. Discourse analysis aims at predicting Psychological characteristic of a person rather than text structure. In Machine learning techniques for linguistic annotation of corpora, the following approaches such as N-gram and Markov Models, Neural Networks, Transformation based learning, decision tree classification, vector based clustering are used. The paper is providing a frame work for the development of new algorithms.

[6] In the paper "A Survey on Machine Learning: Concept, Algorithms and Applications", the authors have defined Machine Learning as the intersection of statistics and computer science. The paper also has specified about the study of human and animal brain in Neuroscience, psychology and many more fields. Though the complexity of Machine learning is high, researchers are working more in this field. The researchers are trying to link different ML algorithms to improve the efficiency of the systems. The paper has mentioned about the Never Ending Language Learner(NELL) which has a hold on the web pages . This Algorithm tracks the web activities every hour.

The Paper has measured and compared the performances of Naïve Bayes(Gaussian), SVM and DT Algorithms and the accuracy is 0.692,0.6565,0.69 respectively. The authors have specified about some languages to be learnt for ML Algorithms such as Phython, Scikit-learn and the R. Finally, the goal of working with ML Algorithms is that the researchers have to concentrate on Time and space complexity of the algorithm.

[7] In the paper "A Survey on Machine Learning Assisted Big Data Analysis for Health Care Domain", focuses on big data analysis and ML methods related to healthcare to improve patients health. This analysed data is helpful for the medical practioners. Big data analytics helps in lowering the costs in any organization and also finds solution of hidden problems. Big data challenges to mention a few are extracting knowledge from unstructured data set, efficient handling of large volumes of data. Methods ued in this paper are Multi Layer Perception (MLP),Support Vector Regression Machine(SVR), Generalized Regression Neural Network(GRNN), K Nearest Neighbor Regression(KNNR). Big data can be divided into two types Batch Oriented Computing and Real Time Oriented Computing(Stream Computing). Apache Hadoop(an open source) provides better performance for large set of data management and analysis.

[8] In the paper "Machine Learning for Computer Security", the authors have briefed about Dynamic Markov Compression(DMC) and Prediction by Partial Matching(PPM) Algorithms for the purpose of security. The paper has focused on identifying spam mails and has specified Hidden Markov Model (HMM) to handle Network traffic.

3 SOME LANGUAGES USED IN ML

With the emerging innovations in Machine Learning, many languages are used by researchers such as Python, R Language. Python was developed in the year 1991. It is an object oriented programming language with several other paradigms such as imperative, functional, procedural and reflective. Imperative feature has the capability to change the program's state. Functional and procedural features make use of subroutines which is a modular approach in solving problems. Reflective property of a programming language modifies its own structure and behavior during run time. R and Phython belong to GNU package which is an operating system belongs to Unix OS family. The Paradigms of R are similar to Phython with Array concepts. Phython doesn't have a native array data structure but makes use of list.

4 METHODOLOGY

Figure 1 depicts that Machine Learning depends on several other areas of Computer Science such as Data Bases, Statistics, Pattern Recognition, Data Mining, Neuro-Computing, (Knowledge Discovery in Data Bases) KDD and Artificial Intelligence.

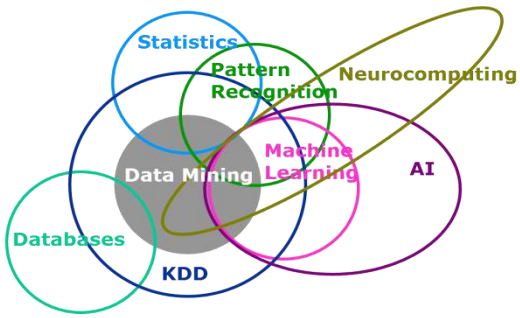

Figure 1 – *Overview of ML*

The main classification of Machine Learning involves Supervised and Unsupervised Learning. Figure 2 shows the classification of Supervised and Unsupervised Algorithms and also the sub classifications and the methods involved in ML.

Supervised Learning depends on labeled training data whereas Unsupervised learning depends on Unlabelled training data.

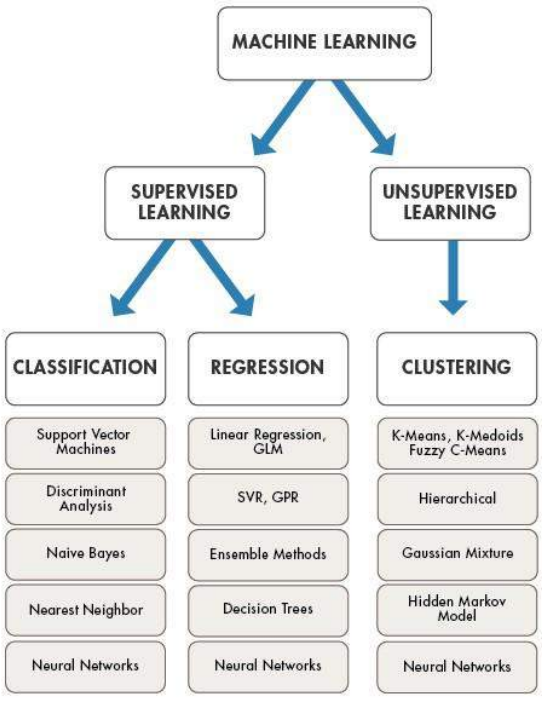

Figure 2 – *Block Diagram of ML*

The steps involved in Supervised learning are determining the training data set, gathering a training data set, determining learned function and to prepare the learning algorithm and evaluating the accuracy of the algorithm by undergoing several test cases. The Unsupervised learning works on abnormal data patterns which is a complex task.

Classification are of two types – Binary and Multi Classification. If the predication of classification Algorithm results in two target classes, then it is a Binary Classification. If the predicate values results in more than two, then it is a Multi Classification as shown in Figure 3.

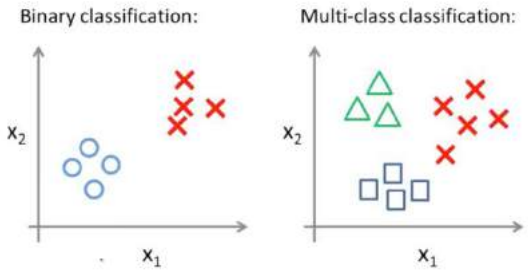

Figure 3 - *Classification*

Classification forecasts target class/classes whereas Regression forecasts a value. Figure 4 shows Weight along X axis and Height along Y axis. Given any unknown weight value in between the range 20 to 30, height value can be found with the help of Regression algorithm which targets a value. From the figure, the height value is within the range 30 to 40.

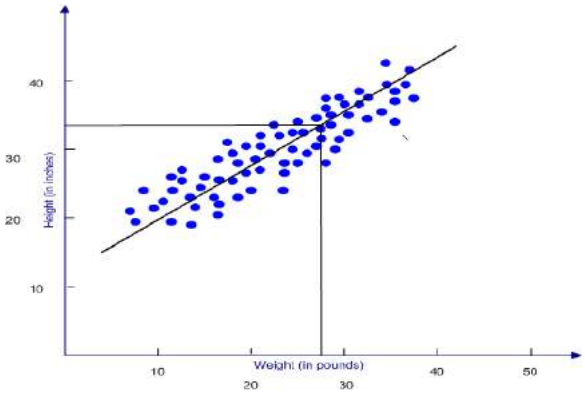

Figure 4 - Regression

Clustering is grouping of similar type of objects as shown in Figure 5. It works based on distance measures and clustering Algorithms.

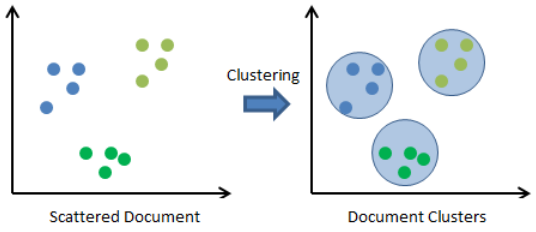

Figure 5 - Clustering

	Supervised Learning	Unsupervised Learning
Discrete	classification or categorization	clustering
Continuous	regression	dimensionality reduction

Table 1 – *Continuous/Discrete versus SL/UL*

Table 1 gives a clear understanding about Continuous and Discrete Supervised and Unsupervised Learning. Classification, Regression and Clustering are the three main areas of concern in the field of ML. The Figure also has the dimensionality reduction which considers random variables and tries to reduce the dimensionality of the inputs without any loss of data to improvise the efficiency of the ML Algorihms.

5 CONCLUSION

In this paper, we have made an attempt to familiarize about Classification, Regression, Clustering which are the primary concerns in understanding ML Algorithms. The paper also briefs about the Phython and R language and the block diagram of ML Algorithms and its Overview.

REFERENCES

[1] Aparna Gullapelly, " A Survey on Supervised Classification Techniques in Machine Learning", Internation Journal of Computer and Mathematical gorithm,Science(IJCMS), ISSN 2347-8527, Volume 6, Issue 11, November 2017.

[2] Bala Brahmeswara Kadaru,Siva Chintaiah Narni, Dr.B.Raja Srinivasa Reddy, " Machine Learning wearable device information in Parkinson unwellness Health watching", International Journal of Engineering Technology Science and Research(IJETSR),www.ijetsr.com, ISSN 2394-3386, Volume 4, Issue 9, September 2017.

[3] Ayon Dey, "Machine Learning Algorithms: A Review", International Journal of Computer Science and Information Technologies, Vol.7(3),2016,1174-1179, ISSN:0975-9646

[4] Iqbal Muhammad and Zhu Yan, "Supervised Machine Learning Approaches : A Survey", DOI:10.21917/ijsc.2015.0133.

[5] Xunlei Rose Hu and Eric Atwell, "A survey of machine learning approaches to analysis of large corpora", School of Computing, University of Leeds, U.K. LS2 9JT.

[6] Kajaree Das, Rabi Narayan Behera, " A Survey on Machine Learning: Concept, Algorithms and Applications",International Journal of Innovative Research in Computer and Communication

Enginering(ijircce),ISSN(online):2320-9801, ISSN(Print): 2320-9798, Vol 5, Issue 2, February 2017.

[7] Raol Priyanka Ajaysinh, Hinal Somani," A Survey on Machine Learning Assisted Big Data Analysis for Health Care Domain", © 2016 IJEDR, Volume 4, Issue 4, ISSN: 2321-9939.

[8] Philip K Chan, Richard P Lippmann, "Machine Learning for Computer Security",Journal of Machine Learning Research 7(2006) 2669-2672.