Machine Leaning and it's Various Algorithms- A Study

Dr.Chetan Shelke

Associate professor, Alliance university, Banglore, India.

Abstract:

 The scientific study of algorithms and statistical models that computer systems use to perform a specific task without being explicitly programmed is called as Machine learning (ML). There are various machine learning algorithms,these algorithms are used for various purposes like image processing, data mining, predictive analytics, etc. The main advantage of using machine learning is that, once an algorithm learns, it can do its work automatically. In this paper, a brief review about various machine learning algorithm is included.

Keywords: Algorithm, Machine Learning, data mining, predictive analytics.

INTRODUCTION

Machine learning is used to teach machines how to handle the data more efficiently. after viewing the data,sometimes we unable interpret the pattern or extract information from the data. that time, we apply machine learning [11]. When the large amount of datasets are available, the demand for machine learning is in rise. Various industries from military to medicine apply machine learning to extract relevant information. The purpose behind the use of machine learning is to learn from the data. Many studies have been done on how to make machines learn by themselves [12] [13].

One of the fastest growing area of computer science is Machine learning , It called as the automated detection of meaningful patterns in data.. There are various applications for Machine Learning (ML),like image processing, data mining, predictive analytics, etc. the most significant of which is data mining. People are generally make mistakes during analyses or, possibly, when trying to create relationships between multiple features [4]. There has been a huge progress in data mining and machine learning as a result of evolution of smart and Nano technology which makes us curious in finding hidden patterns in data to derive value.

Literature review

1Himanshi Bansal, 2Kapil Sharma," A REVIEW STUDY ON VARIOUS ALGORITHMS OF MACHINE LEARNING"www.jetir.org 2020 JETIR April 2020, Volume 7, Issue 4

Within this paper we offer a short description of what is machine learning and the various forms of algorithms in specific machine learning categories. The methods mentioned can be implemented to different forms of data collection, i.e. medical, financial, etc. This paper provides an introduction to several of the common algorithms for machine learning. Shifting the core problem from how to program a machine to how to enable it to program itself, Machine learning prioritizes the creation of systems that are self-learning and self-reliant and the use of the usable data flow inside the system rather than merely manipulating it.[6]

Isonkobong Christopher Udousoro,” REVIEW Machine Learning: A Review” Semiconductor Science and Information Devices | Volume 02 | Issue 02 | October 2020

This review presented a general research on machine learning, its algorithm and techniques. The paper reviews literatures on the applications of machine learning algorithm and techniques on various fields of life which include medical, agriculture, science and so on. Machine learning is one of the high rising technologies used recently for solutions to various problems. It has various algorithms which include supervised, semi-supervised, unsupervised, reinforcement algorithms and so on.[7]

Ayon Dey,” Machine Learning Algorithms: A Review “(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (3) , 2016, 1174-117

This paper surveys various machine learning algorithms. Today each and every person is using machine learning knowingly or unknowingly. From getting a recommended product in online shopping to updating photos in social networking sites. This paper gives an introduction to most of the popular machine learning algorithms.[8]

IsrealUfumaka,” Comparative Analysis of Machine Learning Algorithms for Heart Disease Prediction” International Journal of Scientific and Research Publications, Volume 11, Issue 1, January 2021

In this research study, six (6) popular machine learning model have been implemented for heart disease diagnosis using data from UCI. Different experiments was conducted on these algorithms using K fold cross validation of 5 and 10 folds, after which standard performance metrics were used to evaluate the performance of these models. The objective of this study was to build different machine learning model and compare their performance of predicting for heart disease. From the results obtained from each fold on the models, the sophisticated SVM came out performing the best even out performing Gradient Boosting ensemble method[9].

Rishi Banerjee, Gabriela Bourla, Steven Chen, Mehal Kashyap, Sonia Purohit, *Jacob Battipaglia,” Comparative Analysis of Machine Learning Algorithms through Credit Card Fraud Detection”

The F-scores of all algorithms under multiple combinations of factors were analyzed as described earlier in this paper, and it was determined that the ideal condition for analysis is the hour1 field. Based on this research, a creditcard company should consider implementing a Support-Vector Machine algorithm that analyzes the purchase time in order to most accurately detect whether a credit-card transaction is fraudulent or not[10].

Review work

Machine Learning is a subset of artificial intelligence which focuses mainly on machine learning from their experience and making predictions based on its experience. No machine algorithm gives 100 percent accuracy. what is machine learning is different from traditional program machine learning means if you have data and you have an output then you must create model and you must test that model and then you must check that model. machine learning is iterative process till result is not achieved. Machine learning is branch of AI used such a high scale now a days because its uses data and identify pattern and predict the output.ML focuses on data and once data increases prediction of program will change. ML helps to find hidden insight of data

Example: Classification of mail:

Machine try to identify the mail is spam or normal mail

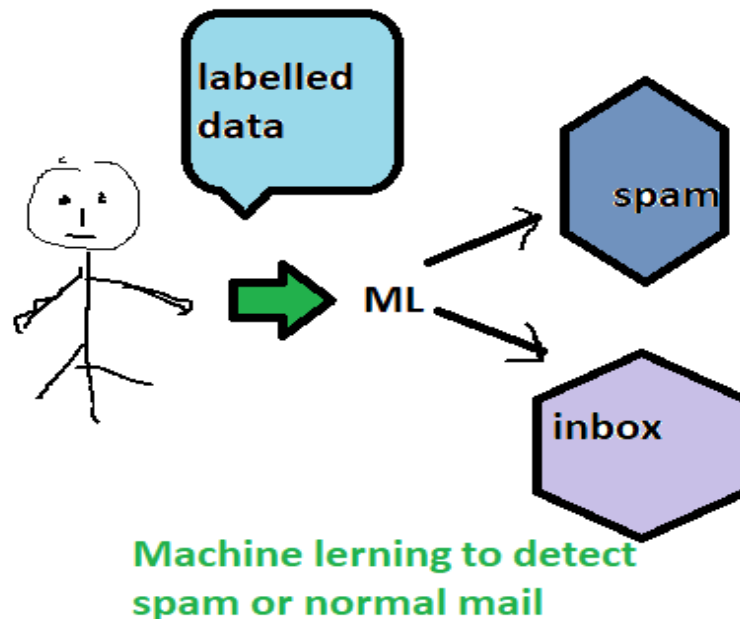Machine try to identify the mail is spam or normal mail

Fig1:Example of machine learning to detect incoming mail is spam or not

Steps of Machine Learning algorithm:

1)Collection of data:

Data can be collected from various sources like pdfs, social media like YouTube, Facebook,

2)Data Wrangling:

To transform data (Data cleansing), you need to filter the data, Data must be in particular format, Data are not consistent like null values are there so we need to filter the data, result will come only when data is proper.

3)Data Analysis:

Data analysis is defined as a process of cleaning, transforming, and modeling data to discover useful information for business decision-making. The purpose of Data Analysis is to extract useful information from data and taking the decision based upon the data analysis.
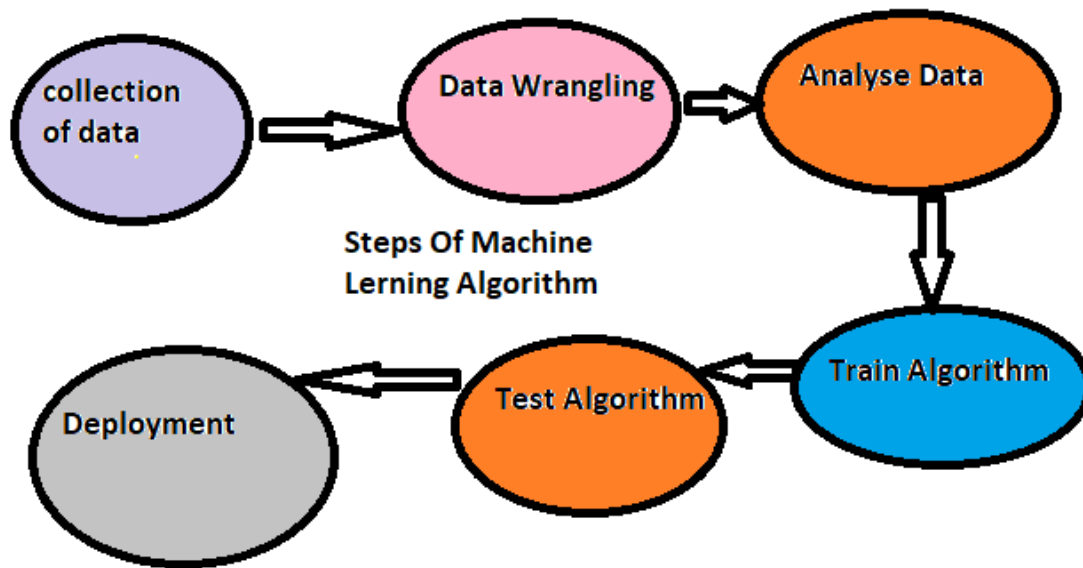
Fig. 2: **Steps involve in Machine Learning algorithm**
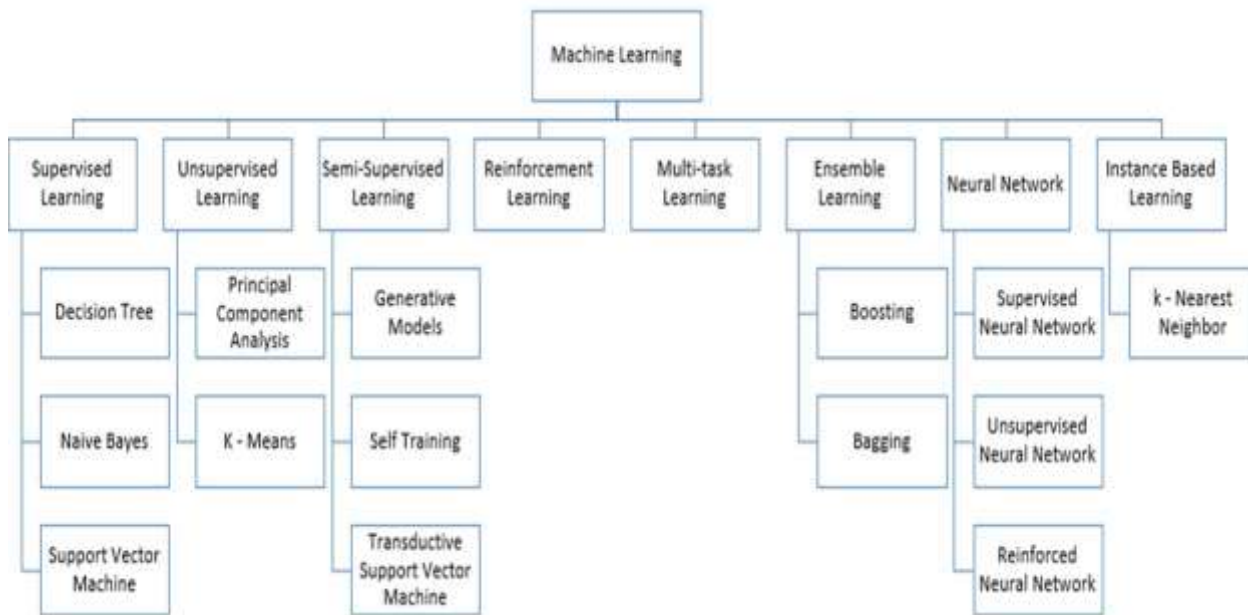
Type of algorithm

Fig. 3: Types of learning[8]

A.Supervised Learning

The supervised Algoriths are the machine learning algorithms which needs external assistance. In this type of algorithm The input dataset is divided into train and test dataset. The train dataset has output variable which has need to be predicted or classified. Each Algorithm learn some kind of patterns from the training dataset and apply them to the test dataset for prediction or classification

1) **Decision Tree:** Decision trees are comes into supervised learning, these are the type of trees which groups attributes by sorting them based on their values. Decision tree is mainly used for classification purpose. Each tree has nodes and branches. Each nodes represents attributes in a group to be classified and each branch represents a value that the node can take.[8]

2) Naive Bayesian (NB) Networks: Naïve Bayes is type of supervised learning, It is basicly used for purpose of clustering and classification [15]. The architecture of Naïve Bayes depends on the conditional probability. It creates trees based on their happening probability. These trees are also called as Bayesian Network[8].These are very simple Bayesian networks which are made of directed acyclic graphs with one parent and several children having a strong assumption of independence among child nodes in the context of their parent **[7].**

3**) Support Vector Machines (SVMs):** Support Vector Machineis an Another most widely used state-of-the-art machine learning technique .This is mainly used for classification.SVM works on the margin calculation principle. It draw margins between the classes. The margins are drawn in such a way that the distance between the margin and the classes is maximum and hence, the classification error minimized.

B.Unsupervised Learning

The Unsupervised Learning is main type if Learning .These algorithms learns some features from the data. When new data is introduced, it uses the features that learned previously to recognize the class of the data. It mainly used for feature reduction and clustering

1) K-Means Clustering: Clustering or grouping is unsupervised learning technique, It is a technique when It initiates, It creates groups automatically. The items which has similar characteristics are put in the same cluster. This algorithm is refer as k-means algorithm because it creates k distinct clusters. The mean of the values in perticuler cluster is center of that cluster [16].[8] the K-means is the algorithm that solve the well-known clustering problem.

2) Principal Component Analysis : The Principle Component analysis is another type of unsupervised learning,In Principal Component Analysis, the data's dimension is reduced to make the faster and Easier computations. [8]

C.Semi - Supervised Learning

Semi – supervised learning algorithms is a technique which combines the power of supervised as well as Unsupervised learning. It can be usefull in those areas of machine learning and data mining in which the unlabeled data is already present and getting the labeled data is a difficult process [17]. There are some categories of semi-supervised learning [18].

1) Generative Models: Generative models are one of the semi-supervised learning method,It is the oldest learning method, lets assumes a structure like $p(x,y) = p(y)p(x|y)$ where, $p(x|y)$ is a mixed distribution e.g. Gaussian mixture models., the mixed components can be identifiable ,Within the unlabeled data.

2) Self-Training: The self learning is also the semi-supervised type of machine learning,In self-training, a classifier is trained with a labeled data portion . then,the classifier is fed with unlabeled data. The predicted labels and the unlabeled points are added together in the training set. This procedure is repeated further. The classifier is learning itself, so the name self-training[8].

3) Transductive SVM: Transductive Learning is also the type of semi supervised learning. Transductive support vector machine or TSVM is an extension of SVM. Both,the labeled and unlabeled data are considered In TSVM. It is used to label the unlabeled data in a way that there is maximum margin between the labeled and unlabeled data. Finding an exact solution using TSVM is a NP-hard problem[8].

D. Reinforcement Learning

Reinforcement learning is the a type of learning which makes decisions based on which actions to take such that the outcome is more positive. In this type of learning,The learner has no knowledge which actions to take until it's been given a situation. The action taken by the learner may affect situations.

E. Multitask Learning

Multitask learning has a simple aim of helping other learners to perform better. On a task,when multitask learning algorithms are applied, it remembers the procedure how it reaches to the particular conclusion or how it solved the problem. Then,the algorithm uses these steps to find the solution of other similar problem.

This helping of one to another algorithm can also be called as inductive transfer mechanism. If the learners tell or share their

experience with each other, the learners can learn concurrently and can be much faster [19].

F. Ensemble Learning

When different individual learners are combined to form only one learner then that particular type of learning is termed as ensemble learning. The individual learner may be decision tree, neural network, Naïve Bayes, etc. Ensemble learning is a hot topicfrom 1990s. It has been observed that, a combination of learners is almost always better for doing a particular job rather than individual learners [20].

1) Boosting: Boosting is one of the type of ensemble Learning technique, This technique is used to decrease bias and variance. Boosting creates a collection of weak learners and convert

them into one strong learner. A weak learner is hardly correlated with true classification. And the a strong learner is strongly correlated with true classification [21].

2) Bagging: Bagging is used where the accuracy and stability of a machine learning algorithm needs to be increased. It is applicable in classification and regression. Bagging decreases variance and helps in handling overfitting [22].

G. Neural Network Learning :

The neural network is also called as artificial neural network or ANN.The Nural network is derived from the biological concept of neurons. A neuron is a cell like structure in a brain. To understand neural network, one must understand the working of neuron. A neuron is made up of mainly four parts, which are dendrites, nucleus, soma and axon. The 1st part dendrites receive electrical signals. The next part Soma processes the electrical signal. The output of the process is carried by the next part axon to the dendrite terminals where the output is transferred to next neuron. The heart of the neuron is nucleus. The inter-connection of neuron is called as neural network where electrical impulses travel around the brain.

1) Supervised Neural Network: In the supervised neural network is the type of Neural network learning, In this type of learning,the output of the input is known already. The predicted output of the neural network is then compared with actual output. the parameters are changed, Based on the error, and then converted into the neural network again.

2) Unsupervised Neural Network: This is also the type of Neural network learning. In this type, the neural network has no priction about the output. The main job of the network is to classify the data according their similarities. The neural network then checks the correlation between various inputs and groups them.

3) Reinforced Neural Network: This is also the type of neural network,In reinforced neural network, the network behaves like, if a human communicates with an environment. a feedback has been provided to the network , From the environment, acknowledging the fact ,whether the decision taken by the network is right or wrong. If the decision taken by network is right, the connections points to that particular output is strengthened. If the decision is wrong, the connections are weakened. Here,The network has no previous information about the output.

H. Instance-Based Learning

In this type, the learner learns a particular pattern type. It tries to use the same pattern to the newly fed data. so the name instance-based. It is a type of lazy learner ,It waits for the arrival of test data and then act on it together with training data. In this type,The complexity of the learning algorithm increases with the size of the data [26].

1)K-Nearest Neighbor: It is a type of Instance-Based Learning, In k-nearest neighbor or KNN , In this technique,the well-labeled training data is fed into the learner. When the test data is introduced to the learner, it compares both the data. k most correlated data is taken from training set. The majority of k is taken which then serves as the new class for the test data [25].

. **CONCLUSION**

This paper studies the concept of Machine Learning and various machine learning algorithm. knowingly or unknowingly, Today each and every person is using machine learning.. This paper gives an introduction to most of the popular machine learning algorithms.This paper also explains the steps of machine learning algorithms which are –Collection of data,Data Wraggeling,Analyse data,Deployment,Test algorithm,and train algorithm.

References

[1] Batta Mahesh,"Machine Learning Algorithms - A Review"International Journal of Science and Research (IJSR)

[2] OsisanwoF.Y. ,Akinsola J.E.T., Awodele O. , Hinmikaiye J. O. , Olakanmi O. ,Akinjobi J. ,"Supervised Machine Learning Algorithms: Classification and Comparison"International Journal of Computer Trends and Technology (IJCTT) – Volume 48 Number 3 June 2017

[3] Shai Shalev-Shwartz and Shai Ben-David (2014). Understanding Machine Learning From Theory to Algorithms.

[4] Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques. Informatica 31 (2007). Pp. 249 – 268. Retrieved from IJS website: http://wen.ijs.si/ojs2.4.3/index.php/informatica/article/download/148/140.

[5]Taiwo, O. A. (2010). Types of Machine Learning Algorithms, New Advances in Machine Learning, Yagang Zhang (Ed.), ISBN: 978-953-307-034-6, InTech, University of Portsmouth United Kingdom. Pp 3 – 31. Available at InTech open website: http://www.intechopen.com/books/new-advances-inmachine-learning/types-of-machine-learning-algorithms

[6]1Himanshi Bansal, 2Kapil Sharma,"A REVIEW STUDY ON VARIOUS ALGORITHMS OF MACHINE LEARNING"www.jetir.org2020 JETIR April 2020, Volume 7, Issue 4

[7]Isonkobong Christopher Udousoro,"REVIEW Machine Learning: A Review"Semiconductor Science and Information Devices | Volume 02 | Issue 02 | October 2020

[8] Ayon Dey," Machine Learning Algorithms: A Review "(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (3) , 2016, 1174-117

[9] IsrealUfumaka," Comparative Analysis of Machine Learning Algorithms for Heart Disease Prediction" International Journal of Scientific and Research Publications, Volume 11, Issue 1, January 2021.

[10]Rishi Banerjee, Gabriela Bourla, Steven Chen, Mehal Kashyap, Sonia Purohit, *Jacob Battipaglia," Comparative Analysis of Machine Learning Algorithms through Credit Card Fraud Detection"

[11] W. Richert, L. P. Coelho, "Building Machine Learning Systems with Python", Packt Publishing Ltd., ISBN 978-1-78216-140-0
[12] M. Welling, "A First Encounter with Machine Learning"
[13] M. Bowles, "Machine Learning in Python: Essential Techniques for Predictive Analytics", John Wiley & Sons Inc., ISBN: 978-1-118-
96174-2

[14] Vansh Jatana," Machine Learning Algorithms" Ayon Dey / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (3) , 2016, 1174-1179

[15] D. Lowd, P. Domingos, "Naïve Bayes Models for Probability Estimation"

[16] S. S. Shwartz, Y. Singer, N. Srebro, "Pegasos: Primal Estimated sub - Gradient Solver for SVM", Proceedings of the 24th International Conference on Machine Learning, Corvallis, OR, 2007

[17] X. Zhu, A. B. Goldberg, "Introduction to Semi – Supervised Learning", Synthesis Lectures on Artificial Intelligence and Machine

Learning, 2009, Vol. 3, No. 1, Pages 1-130

[18] X. Zhu, "Semi-Supervised Learning Literature Survey", Computer Sciences, University of Wisconsin-Madison, No. 1530, 2005.

[19]R. Caruana, "Multitask Learning", Machine Learning, 28, 41-75, Kluwer Academic Publishers, 1997

[20] D. Opitz, R. Maclin, "Popular Ensemble Methods: An Empirical Study", Journal of Artificial Intelligence Research, 11, Pages 169-

198, 1999

[21] Z. H. Zhou, "Ensemble Learning", National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China

[23] https://en.wikipedia.org/wiki/Bootstrap_aggregating

[24] https://en.wikipedia.org/wiki/Instance-based_learning

[25] P. Harrington, "Machine Learning in Action", Manning Publications Co., Shelter Island, New York, ISBN 9781617290183, 2012