A SURVEY ON MACHINE LEARNING

Kajal Thakur*1, Muskan Joshi*2, Khushboo Pawar*3

*1,2Student, Department of Computer Science and Engineering, Acropolis Institute of Research and Technology, Indore, Madhya Pradesh, India

*3Assistant Professor, Department of Computer Science and Engineering, Acropolis Institute of Research and Technology, Indore, Madhya Pradesh, India

ABSTRACT

Few decades ago, Machine learning always seemed as a something incomprehensible, and was not even proximate to being reality. Machine learning is an artificial intelligence (AI) discipline that is commenced on the principle that systems can learn from data, find patterns, and make judgments with minimal human interaction been impossible or difficult for humans to solve, which is why it has grown dramatically over the years. The pervasiveness of machine learning systems is undeniable. Machine learning has gained its popularity in every field such as game playing, intrusion detection, Information retrieval, marketing, bioinformatics, malware detection, autonomous vehicle and so on. In this paper we have presented an overview of Machine learning, its types along with their various applications and advantages and disadvantages.

Keywords: Machine Learning, Artificial Intelligence, Supervised Machine Learning, Unsupervised Machine Learning,

I. INTRODUCTION

In the year 1959, Arthur Samuel introduced the term "machine learning." and Tom M. Mitchell provided the very first formal definition For Machine learning:

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with

experience E."

Machine learning, in simple terms, is a subset of Artificial Intelligence (AI) that enable machines to learn and improve on their own without being specifically programmed to do so. It is, in a way, the practice of teaching machines to solve problems by teaching them to think. We've been engendering an incalculable quantity of data since the technology advancement. According to statistics, we engender 2.5 quintillion bytes of data per day! The volume of data is presaged to double every two years, which is an astonishing figure. Netflix, Facebook, Amazon and many more companies are utilizing a bunch of data to construct Machine Learning models that avail them to uncover gainful possibilities and avoid dispensable risk. Machine learning has exploded in popularity as a result of its ability to organize, analyze, and extract knowledge from data and use data to solve problems and find solutions to the most difficult difficulties that businesses face. Machine Learning uses a variety of algorithms to help humans make better judgments. Discovering hidden patterns and drawing vital insights from data is the most important component of Machine Learning. By developing prediction models and utilizing statistical methodologies, Machine Learning allows you to go under the surface and explore data on a microscale. Humans can comprehend data and uncover patterns in days, but Machine Learning algorithms can do so in less than a second. Machine Learning is using to solve the most difficult problems, such as identifying genes connected to the fatal ALS disease and constructing driverless vehicles. Supervised machine Learning. Unsupervised machine Learning and Reinforcement Learning are the three types of machine learning.

II. MACHINE LEARNING

We are generally given a training set and a test set in machine learning, neural networks, support vector machines, and evolutionary computation.

The training set will be defined as the sum of the labelled and unlabeled instances available to machine learners. In contrast, the test set contains examples that have never been seen before.

A machine can learn in three ways:1) Supervised Learning
2)Unsupervised Learning
3)Reinforcement Learning

1) Supervised learning: labeled data (X, Y) = {(x1, y1), . . .,(xn, yn)} is use to train algorithm. A generic rule for translating inputs to outputs is learned in supervised learning. This is equivalent to a "teacher" giving a problem and its answer (labelled output data) to a student and teaching the student to solve more, similar problems in the future, such as identifying the mapping from unseen sample features to their right labels or target values. Speech recognition, Object Recognition, Spam detection,
Bioinformatics are some applications of Supervised learning.

Supervised Learning has been broadly classified into 2 types.
1)Regression
2)Classification
1)Regression: In regression algorithm learns from Labeled Datasets and then predicts a continuous-valued output for new data fed to it. When the result is a number, such as weight or income, then we utilize the regression procedure.
2)Classification: Classification is the process of classifying a given collection of data into classes. It may be done with both unstructured and structured data. Making the prediction for data point's class is the first step in the procedure. The classes are also known as label, or categories. Unlike Regression, the outcome will be one of the classes rather than a number.

The Drawbacks of Supervised Learning
i)While working with these algorithms, you may encounter a number of problems and drawbacks.
ii)If the dataset is wrong, your system will learn improperly, resulting in losses.
iii)To train the data, good examples must be utilized.
iv)Supervised Learning requires a significant amount of computation time.
v)Data pre-processing is always a difficulty.
vi)You might easily overfit your algorithm.
vii)Unwanted data may impair accuracy.

2)Unsupervised Learning: In unsupervised learning, the training set comprises exclusively of the unlabeled set X = Xu = x1..., xu. The machine must find solutions on its own (structure, pattern, or understanding in unlabeled data). It's akin to giving a learner a set of patterns and asking him or her to figure out the underlying motifs, that generated the similarity. Computer vision, medical imaging, Anomaly detection, News Section categorizing on basis of same story line, News Sections, Recommendation Engines are some applications where we are using Unsupervised Learning

Unsupervised Learning has been broadly classified into 2 types.
1)Clustering: Clustering is a data mining approach that categorize unlabeled data based on similarities and contrasts. Clustering techniques are used to arrange unclassified raw data objects into groups represented by information structures or patterns. Clustering methods are classified into three types: exclusive, overlapping, and probabilistic, hierarchical clustering techniques.
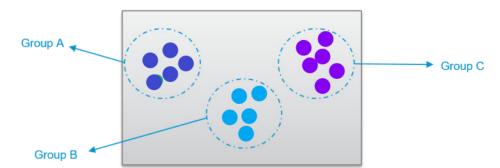
Fig a) Clustering

2)Association: Association is a form of Unsupervised Learning in which you uncover the dependency of one data item on another and map them in order to profit more.

The Drawbacks of Unsupervised Learning:

i)The basis on which data was grouped was not transparent.

ii)There is a greater chance of getting erroneous results.

iii)Due to the large amount of training data, there is a significant level of computational complexity.

iv)Validation of output variables require human intervention.

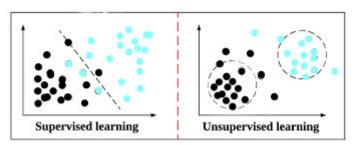v)The basis on which data was grouped was not transparent.

Fig b) Supervised vs unsupervised learning

3) Reinforcement Machine learning: Machine learning includes reinforcement learning as a subfield. It is about engaging in suitable behavior to maximize advantage in a given situation. Various programmed and computers utilize it to identify the best probable action or course of action to take in a particular circumstance. Reinforcement learning differs from supervised learning in that the training data contains the solution key, allowing the model to be trained with the correct answer; however, there is no answer in reinforcement learning, and the reinforcement agent chooses what to do to accomplish the given task. In the absence of a training dataset, it is forced to learn from its own experience.

Reinforcement Learning is commonly used in sophisticated Machine Learning applications like self-driving vehicles.

The Drawbacks of Supervised Learning

i)Too much reward can lead to state overload, which can reduce outcomes.

Reinforcement Learning Practical Applications –

i)Reinforcement Learning may be used in robotics to automate industrial processes.

ii)Machine learning and data processing may both benefit from Reinforcement Learning.

III. ADVANTAGES OF MACHINE LEARNING

1)Find trends and patterns easily:
Machine Learning can evaluate enormous volumes of data and find particular trends and patterns. For example, it can be used to identify the browsing patterns and buy history of its customers for an e-commerce website like Myntra, Flipkart to help people respond to the appropriate items, promotions and reminders that are relevant. It uses the results to disclose the corresponding advertising.
2)There is no need for human interaction:
With ML, you don't always have to monitor your project every step of the way. It enables machines to create predictions because they can learn by itself, we only need to provide data.
3)Consistent enhancement:
As ML algorithms acquire expertise, accuracy and efficiency continue to be improved. This allows them to decide better. You must develop a model for the weather forecast. As the magnitude of data that you have perpetuates to increment, your algorithms learn to predict more accurately.

IV.DISADVANTAGES OF MACHINE LEARNING

1) quality of data
A consequential quantity of data is being utilized in the training and cognition process of machine learning. So, this utilization of data should be of good quality, impartial. During the process of machine learning with avail of software development accommodations, there are additionally moments when we require to wait. In that period of time incipient data is being engendered and can be utilized for further process.
2)Interpretation
When algorithms avail in all of these steps and provide a result so this given output must be examined for mistakes and the corrective process in order to get better and accurate result
3)Although the results and predictions of ML algorithms are powerful, but they do have their own constraints or falls. The most important thing is that black box ML models are fundamentally opaque or untransparent.
 So, internal logic or innerworkings are hidden to human or user in black box models which leads to a serious disadvantage as because it prevents human, expert from being able to interpret and understand the reasoning of the system. and how specific decision are made. The tendency in health care, malefactor equity and other regulated fields has increased to use ML for high-stakes forecasting applications that have profounds effects on human lives. As for high-stakes decisions, the verbalized dilemma is further compounded as the confidence of electronics in a system which cannot and cannot be explained by human beings provides visible risks to significant decisions. Explainable Artificial Insight (XAI) has been developed as a field of research to solve this question and to move to a more transparent AI. In this discipline, research on machine learning interpretability is focused on. The major objective is to construct a series of models and techniques which can be interpreted more clearly while keeping the high levels of prediction.

V.   CONCLUSION

In this paper, we have seen a gentle introduction to the Machine learning. It first presented what is a Machine learning algorithm, its types supervised learning, Unsupervised learning and Reinforcement learning along with their applications and drawback. Supervised machine learning is different from unsupervised machine learning in many ways. Humans can comprehend data and uncover patterns in days, but Machine Learning algorithms can do so in less than a second. Machine Learning is using to solve the most difficult problems, but with that machine learning have disadvantages too.

VI. ACKNOWLEDGEMENTS

VII. REFERENCES

[1] "Machine learning: Trends, perspectives, and prospects" Author M. I. Jordan, T. M. Mitchell.

[2] Goldberg, D. E. (1989). Genetic algorithms in search, optimization, and machine learning, Reading, MA: Addison-Wesley.

[3] " Machine Learning Interpretability: A Survey on Methods and Metrics" Author Diogo V. Carvalho, Eduardo M. Pereira and Jaime S. Cardoso

[4] "Machine Learning that Matters" Kiri L. Wagstaf Journal reference: Proceedings of the Twenty-Ninth International Conference on Machine Learning (ICML), p. 529-536

[5] Carbonell, Jaime. Machine learning: A maturing field. Machine Learning, 9:5–7, 1992.

[6]Book: A Matrix Algebra Approach to Artificial Intelligence pp 223-440| Cite as Machine Learning Author: Xian-Da Zhang   Department of Automation Tsinghua University Beijing China

[7] Frank, A. and Asuncion, A. UCI machine learning repository, 2010.

URL http://archive.ics.uci.edu/m

[8] https://www.ibm.com/cloud/learn/unsupervised-learning

[9] https://www.ibm.com/cloud/learn/supervised-learning

[10] https://www.edureka.co/blog/introduction-to-machine-learning/