International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

A Review on Machine Learning Algorithm

Dr.Saurabh Saxena

Government Shyam Sundar Agrawal PG College Sihora Jabalpur

drsaurabhsaxena356@gmail.com

ABSTRACT

Machine learning (ML) is a branch of artificial intelligence focused on the development of algorithms and statistical models that enable computer systems to perform specific tasks without explicit programming. ML algorithms are widely used in everyday applications. For instance, when using a web search engine like Google, a learning algorithm plays a crucial role in ranking web pages based on relevance. Beyond search engines, ML is applied in various domains such as data mining, image processing, and predictive analytics. One of its key advantages is the ability to automate tasks once the algorithm has been trained on relevant data. This paper provides a brief review of machine learning applications and explores future prospects in this rapidly evolving field.

INTRODUCTION

Machine learning (ML) is used to teach machines how to handle data more efficiently. In many cases, humans may struggle to detect patterns or extract meaningful insights from complex datasets. This is where machine learning becomes invaluable. With the growing availability of large datasets, the demand for ML techniques is steadily rising. Today, many industries—from medicine to the military—leverage machine learning to extract relevant information and support decision-making.

The term "Machine Learning" was coined by Arthur Samuel, a pioneer in artificial intelligence and computer gaming. He described machine learning as a "field of study that gives computers the ability to learn without being explicitly programmed."

ML can be defined as the process of automating and improving a computer’s learning ability based on data and past experiences, without requiring direct human intervention through programming. The process begins by supplying high-quality data, which is then used to train the machine by creating models using various algorithms. The choice of algorithm depends on the nature of the data and the specific task we aim to automate.

In essence, machine learning is a subset of artificial intelligence focused on developing algorithms that can learn from and make predictions or decisions based on data. These algorithms improve over time as they are exposed to more information, refining their accuracy and performance.

Machine learning is the process of enabling computers to make predictions or decisions based on data—referred to as *training data*—without being explicitly programmed for specific tasks. These algorithms are employed across a variety of domains, including medicine, email filtering, and computer vision, where manually developing algorithms to perform the desired tasks is often challenging or infeasible.

Not all machine learning approaches fall under the umbrella of statistical learning. A subset of machine learning is closely linked to computational statistics, which emphasizes prediction using computers. The field also draws heavily from mathematical optimization, which offers the theoretical foundation and tools necessary for algorithm development and performance improvement. Furthermore, *data mining*—a related discipline—focuses on unsupervised learning techniques for exploratory data analysis. When machine learning is applied to solve business problems, it is often referred to as *predictive analytics*.

Literature Review

Machine learning involves training computers to learn how to perform tasks without explicitly programming them to do so. This process relies on algorithms that enable computers to learn from data, allowing them to execute tasks autonomously. While basic computational tasks can be hardcoded via traditional algorithms, more complex tasks benefit from a machine learning approach where the system is trained to generate its own algorithm based on data.

In practice, rather than programming every individual step, developers train the system using a dataset containing examples of correct solutions. The system then learns from this training data and refines its internal algorithms to handle similar tasks in the future. A well-known example is the MNIST dataset, which contains images of handwritten digits and is commonly used to train models for digit recognition.

Types of Learning

Supervised Learning

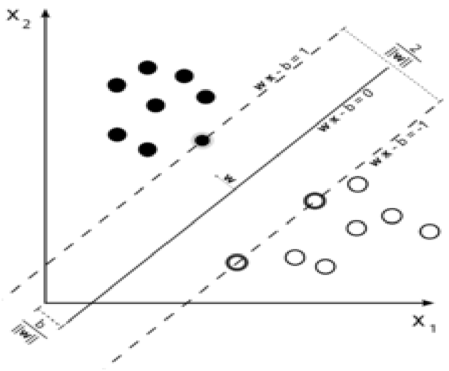
Supervised learning involves training a model using a labeled dataset that includes both input data and the corresponding desired outputs. This training data comprises numerous *training instances*, each consisting of an input (or *feature vector*) and an associated output (or *label*). Mathematically, the entire training set is often represented as a matrix.

The goal of supervised learning is to develop a function that maps inputs to outputs, allowing the model to predict the output for new, unseen inputs. This is typically achieved by optimizing an objective function through iterative updates during training.

There are two primary types of supervised learning tasks:

- **Classification**: Used when the output is a discrete label. For example, filtering incoming emails into folders is a classification task.
- **Regression**: Used when the output is a continuous value. An example would be predicting the price of a house based on various features.

Related techniques include *active learning*, where the model selectively queries the user to label uncertain data points, enhancing learning efficiency.



Unsupervised Learning

Unsupervised learning methods analyze datasets that contain only inputs, without corresponding outputs or labels. These algorithms aim to detect structure within the data, such as grouping similar data points (clustering), identifying patterns, or reducing dimensionality.

Unlike supervised learning, unsupervised algorithms learn from **unlabeled, unclassified, and uncategorized** data. They discover commonalities in the dataset and make decisions based on the presence or absence of these patterns in new data. Rather than learning from explicit feedback, unsupervised models infer insights directly from the data's structure.

A key application of unsupervised learning is in **density estimation**—a statistical approach used to estimate the probability density function of the data. However, unsupervised learning encompasses a variety of techniques aimed at summarizing and explaining key aspects of the data.

One of the most common unsupervised learning techniques is **cluster analysis**, which involves dividing a set of observations into subsets (called **clusters**). The goal is to ensure that observations within the same cluster are similar according to one or more defined criteria, while observations from different clusters are distinct.

Different clustering methods make different assumptions about the underlying structure of the data. These assumptions are typically based on a **similarity metric** and are evaluated using measures such as:

- **Internal compactness** – the degree of similarity among items within the same cluster.
- **Separation** – the degree of difference between items from different

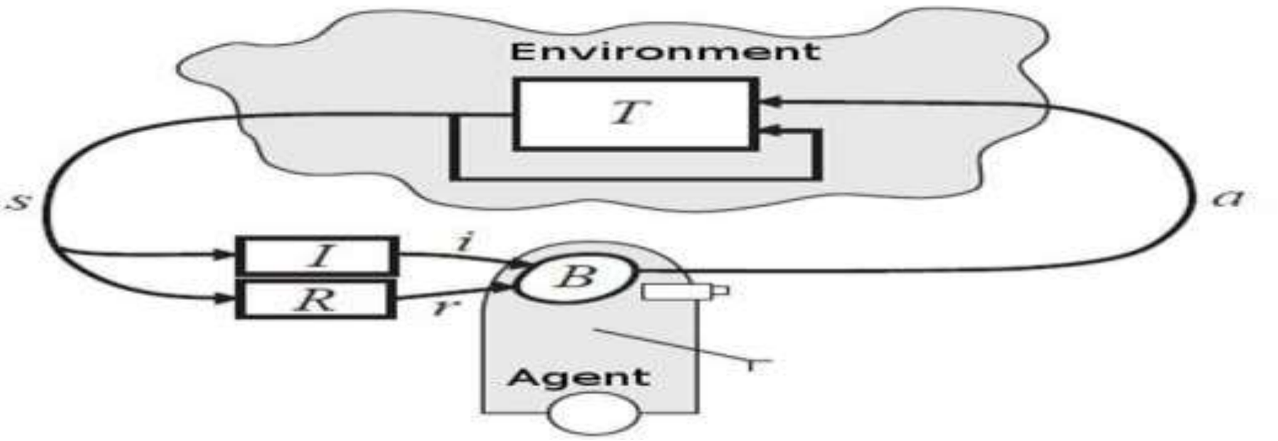
Semi Supervised Learning

Supervised learning relies on fully labeled training data, while unsupervised learning uses data without any labels. Semi-supervised learning lies between these two: although some training examples lack labels, researchers have found that combining a small amount of labeled data with a larger set of unlabeled data can significantly improve learning accuracy. This approach is particularly useful when labeled data is expensive or time-consuming to obtain. In contrast, weakly supervised learning deals with training labels that are noisy, limited, or imprecise—but generally cheaper to acquire—resulting in larger, though less accurate, training datasets.

Reinforcement Learning

Reinforcement learning is a branch of machine learning that focuses on how software agents should act in a given environment to maximize cumulative reward. Because of its generality, reinforcement learning is studied across many disciplines, including game theory, control theory, operations research, information theory, simulation-based optimization, multi-agent systems, swarm intelligence, statistics, and genetic algorithms.

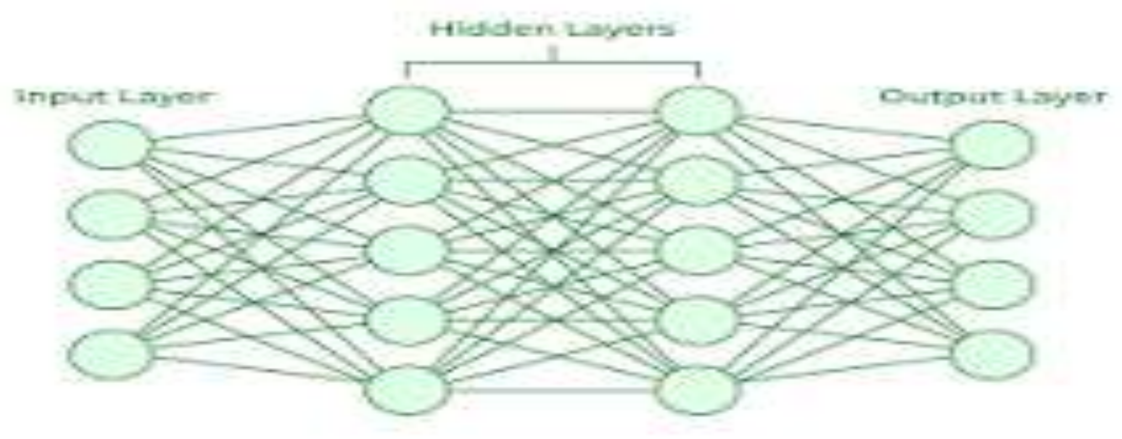
In machine learning, the environment is often modeled as a **Markov Decision Process (MDP)**. Many reinforcement learning systems use dynamic programming techniques; however, when exact mathematical models of the MDP are not feasible, reinforcement learning methods are employed . algorithms are employed in autonomous vehicles and in teaching humans how to play a game.



Artificial Neural Network

Artificial Neural Networks (ANNs), also known as **connectionist systems**, are computing systems inspired by **biological neural networks** found in animal brains. These systems "learn" to perform tasks by analyzing **examples**, rather than following explicitly programmed, task-specific rules.

An ANN is built from a set of **connected units or nodes**, often referred to as **artificial neurons**. These are loosely modeled after the neurons in the human brain. Each connection, much like the **synapses** in a biological brain, can transmit information — or a **signal** — from one artificial neuron to another.



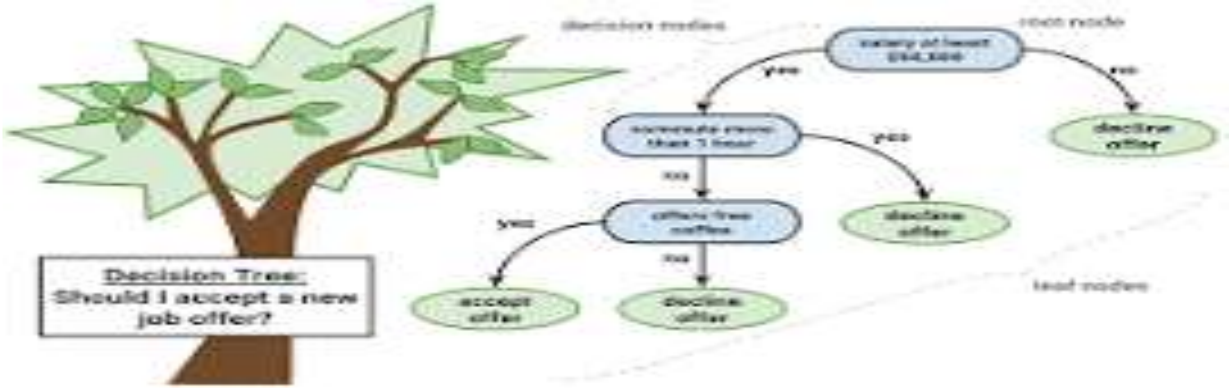
Decision Tree

A **Decision Tree** is a predictive model used to map observations about an item (represented in the **branches**) to conclusions about the item's target value (represented in the **leaves**). It is widely used in **statistics**, **data mining**, and **machine learning** as a powerful and interpretable method for **predictive modeling**.

There are two primary types of decision trees:

1. **Classification Trees**: These are used when the target variable is **categorical** (i.e., it can take on a discrete set of values).
   - **Leaves** represent **class labels** (the outcomes or predictions).
   - **Branches** represent combinations of **features** or **decisions** that lead to those labels.

Decision trees work by recursively splitting the dataset based on feature values to create a tree structure, optimizing for criteria such as **information gain** or **Gini impurity** (in classification) or **variance reduction** (in regression).



Support Vector Machines (SVMs)

Support Vector Machines, also known as **support-vector networks**, are a group of supervised learning algorithms used for **classification** and **regression** tasks.

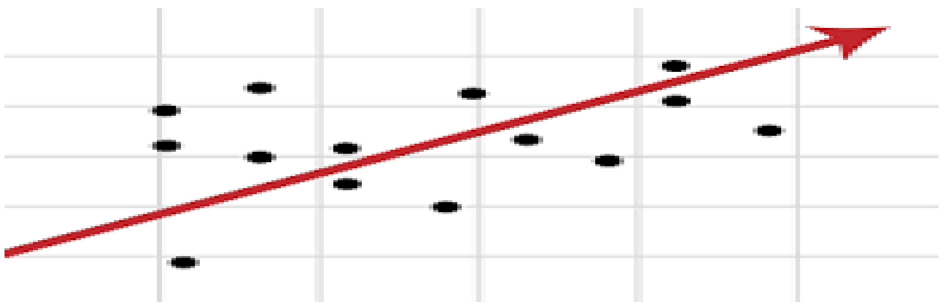
An SVM training algorithm builds a model that predicts whether a new example belongs to one of two categories, based on a set of training examples labeled as belonging to those categories. While methods such as **Platt scaling** can be used to apply SVMs in a probabilistic classification context, the core SVM algorithm is a **non-probabilistic**, **binary**, **linear classifier**.

However, SVMs can also handle **non-linear classification** tasks. This is achieved through the **kernel trick**, which implicitly maps input data into high-dimensional feature spaces, enabling the classifier to find a linear separation in that transformed

Regression Analysis

Regression analysis refers to a range of statistical techniques used to estimate the relationships between input variables (independent variables) and their associated outcomes (dependent variables).

The most common form is **linear regression**, where a straight line is generated to best fit the data. This is typically done using mathematical criteria such as **ordinary least squares (OLS)**, which minimizes the sum of the squared differences between observed and predicted values.



Bayesian Network

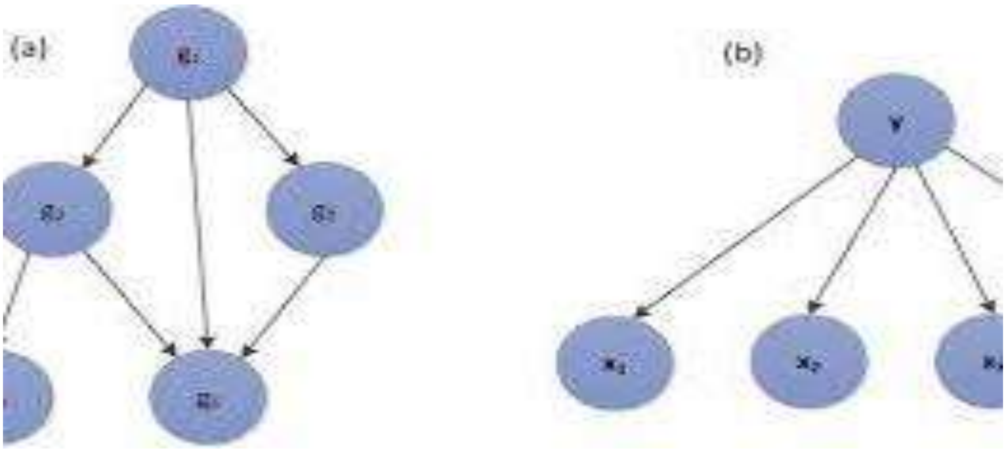
A **Bayesian network**, also known as a **belief network** or a **directed acyclic graphical model (DAG)**, is a **probabilistic graphical model**. It uses a **directed acyclic graph** to represent a set of **random variables** and their **conditional dependencies**.

Each node in the graph represents a variable, and the directed edges represent conditional dependencies between these variables. The absence of an edge implies conditional independence.

For example, a Bayesian network can model the **probabilistic relationships between diseases and symptoms**. Given a set of observed symptoms, the network can be used to calculate the **likelihood of various diseases** being present.

Bayesian networks support **efficient inference and learning algorithms**, making them useful for decision-making under uncertainty.

A special case, called a **Dynamic Bayesian Network (DBN)**, extends this framework to model **temporal sequences** of variables, such as **speech signals**, **biological sequences**, or **financial time series**.



Conclusion

The study explores various machine learning algorithms and emphasizes their growing role in everyday life. It highlights how machine learning is integrated into common activities, such as product recommendations during online shopping and image uploads on social media. Overall, the document serves as an overview of the most commonly used machine learning algorithms, illustrating their widespread and often unnoticed presence in modern digital interactions.

REFERENCES

[1]. R. Praba, “A Study on Data Science Basics with Python Concepts”, Volume-4, Issue-14, ISSN: 2582-3930.

[2]. W. Richert, L. P. Coelho, “Building Machine Learning Systems with Python”, Packt Publishing Ltd., ISBN 978-1-78216-140-0

[3]. M. Welling, “A First Encounter with MachineLearning”

[4]. M. Bowles, “Machine Learning in Python: Essential Techniques for Predictive Analytics”, John Wiley & Sons Inc., ISBN: 978-1-118- 96174-2

[5]. S.B. Kotsiantis, “Supervised Machine Learning:AReviewofClassificationTechniques”, Informatica31(2007)249-268

[6]. L.Rokach,O.Maimon,“Top–Down Induction of Decision Trees Classifiers – A Survey”, IEEE Transactions on

Systems,
[7]. D.Lowd,P.Domingos,“NaïveBayesModels for Probability Estimation”
[8]. https://webdocs.cs.ualberta.ca/~greiner/C- 651/Homework2_Fall2008.html
[9]. D. Meyer, “Support Vector Machines – The Interface to libsvm in package e1071”, August 2015
[10]. S. S. Shwartz, Y. Singer, N. Srebro, “Pegasos: Primal Estimated sub - Gradient Solver for SVM”, Proceedings of the 24th InternationalConferenceonMachine Learning,
Corvallis,OR, 2007
[11]. http://www.simplilearn.com/what-is-machine- learning-and-why-itmatters- article
[12].P. Harrington, “Machine Learning in action”, Manning Publications Co., Shelter Island, New York, 2012

Citethisarticleas

Praba. R,Darshan.G,Roshanraj.K.T, Surya Prakash. B, "Study On Machine Learning Algorithms", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume7 Issue 4, pp. 67-72, July-August 2021. Available at doi: https://doi.org/10.32628/CSEIT2173105
JournalURL:https://ijsrcseit.com/CSEIT2173105