A Review : Machine Learning Algorithms

Aditya Vishwakarma Shiv Kumar Tiwari Pinkal Jain

M.tech, SRIT,Jabalpur ASST. Prof. GGCT, Jbp. ASST. Prof. GGCT, Jbp.

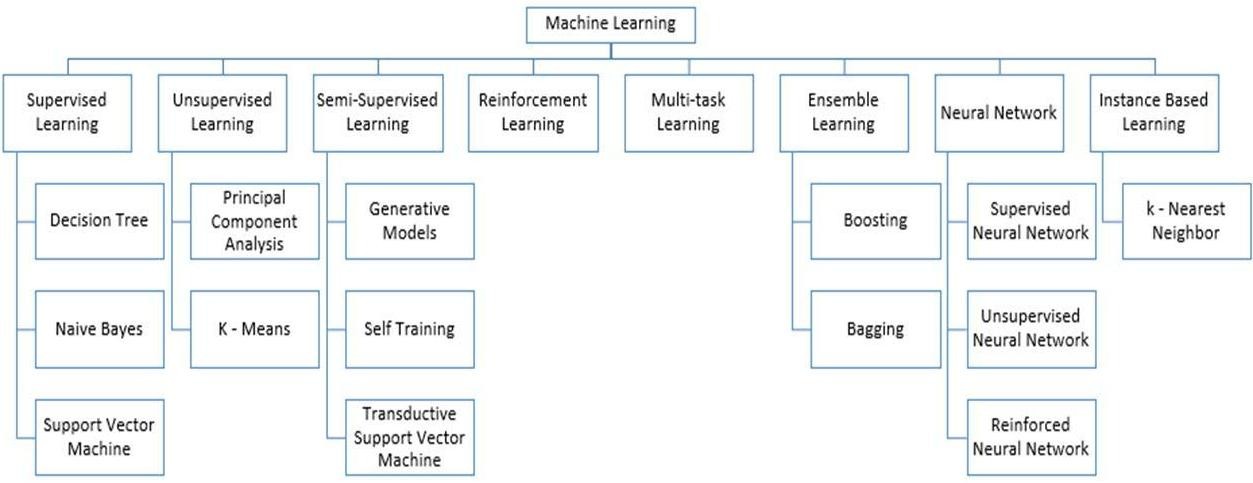
**Abstract:** In this study, a variety of machine learning approaches have been addressed. Numerous jobs, such as data mining, image processing, predictive analytics, etc., use these methods. The main advantage of machine learning is the ability of an algorithm to function independently once it has learned the utilisation of data.

**Keywords:** Machine learning, algorithms,SVM.

# Introduction

Machine learning is used to teach machines how to handle data more effectively. Even after viewing the data, there are times when we are unable to spot a pattern or make inferences from it. We use machine learning in that situation [1]. The large number of datasets has increased demand for machine learning. To obtain pertinent data, machine learning is used in numerous industries, including the military and healthcare. Machine learning aims to learn from the data. Trying to educate robots to learn on their own has been the topic of various studies [2] [3]. Many mathematicians and programmers employ a range of methods to address this issue. Several of them are demonstrated in Fig. 1.Section 2 details each machine learning technique.

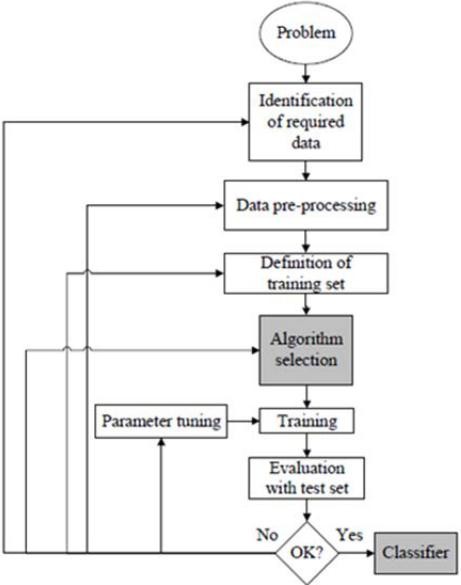
1. **Learning Methods**

Trees that aggregate qualities by sorting them according to their values are known as decision trees. Decision trees are mostly used for classification. In every tree, there are nodes and branches. Each node in a graph represents an attribute.

**Figure 1:** Types of Learning [2] [3]

## Supervised Machine Learning

Supervised Machine Learning refers to algorithms that require assistance from outside sources. The dataset is separated into two parts: train and test. The output variable of the train dataset must be predicted or categorised. Every algorithm predicts or categorizes the test dataset using some form of pattern identified from the training dataset [4]. Figure 2 depicts the approach for supervised machine learning algorithms. This article examines three of the most popular supervised machine learning approaches. Supervised machine learning, as the name suggests, is based on supervision. The "labelled" dataset is used to train the machines using the supervised learning approach, and the computer predicts the output after training. The indicated data shows which inputs have previously been mapped to which outputs. To put it another way, we ask the machine to predict the results using test datasets after training it with input and related output. Let's look at an example of supervised learning. Assume our input is a collection of photographs of cats and dogs. As a result, we will first educate the computer to interpret the visuals, such as the size and shape of a dog's tail and the shape, colour, and height of a cat's eyes (cats are shorter than dogs, for example). Following training, we feed the computer a picture of a cat and ask it to identify it and predict the outcome. After gaining knowledge, the machine will examine every feature of the object, including height, form, colour, eyes, ears, tail, and so on, and decide that it is a cat.

This is the method used by the computer in supervised learning to recognise the things.

**Figure 2:** Workflow of supervised machine learning algorithm [4]

The basic goal of the supervised learning technique is to create a map between the input variable (x) and the output variable (y). Supervised learning has real-world applications such as risk analysis, fraud detection, and spam filtering.



**Figure 3:** Decision Tree [5]

### 1.Supervised Machine Learning Categories

### There are two types of supervised machine learning problems, which are outlined below:

### Classification

### Regression

1. **Classification**

### Classification techniques are used to solve issues with categorical output variables, such as "Yes" or "No," Male or Female, Red or Blue, and so on. The categorization algorithms predict the categories found in the dataset. Spam detection, email filtering, and other classification methods are now in use. The following are some popular classification algorithms:

### ****Decision trees,****

### ****Random Forest algorithm****

### ****Logistic Regression algorithm****

* **Regression**

Regression techniques are used to deal with regression problems where the input and output variables are linearly related. They are used to forecast variables with consistent outcomes, such as market trends and weather forecasts.The following list of popular regression algorithms includes:

* Decision Tree Algorithm
* Multivariate Regression Algorithm
* Simple Linear Regression Algorithm
* Lashing Regression

**The benefits and drawbacks of supervised learning:**

• Because supervised learning employs a labelled dataset, we can precisely identify the object classes; and

• These techniques are effective for forecasting the outcome based on prior performance.

**Disadvantages:**

### • These algorithms are incapable of dealing with difficult challenges.

### • If the test data differs from the training data, it may predict the wrong outcome.

### • Training the algorithm takes a significant amount of computer time.

### Applications of Supervised Learning

## The following are some examples of common Supervised Learning applications:

## • Image Segmentation: For image segmentation, algorithms based on Supervised Learning are used. Image classification is performed on various image data in this procedure using pre-established labels.

## • The medical examination

## In the medical field, supervised algorithms are frequently utilised for diagnostic purposes. It employs historical data, which contains labels for medical images and descriptions of ailments. The machine may employ such a procedure to diagnose a disease in new patients.

## • Detection of Fraud

## The use of supervised learning algorithms enables the detection of fraudulent clients, transactions, and so on. Historical data is used to uncover trends that may indicate fraud.

## • Spam detection - Classification techniques are used in spam detection and filtering.

## 2. Unsupervised Machine Learning

### Unsupervised learning varies from supervised learning in that it does not require supervision. This is the definition of unsupervised machine learning, which occurs when the system predicts output without any human intervention after being trained on an unlabeled dataset.Unsupervised learning involves training models on data that has not been tagged or classified, and then allowing them to act autonomously on that data.

### The unsupervised learning method's major goal is to categorise or group the unsorted dataset based on similarities, differences, and patterns. The machines are tasked with discovering hidden patterns in the input dataset.

**Let me provide an example to help you grasp it better. Assume we feed the machine learning model photos of a fruit basket. The model's objective is to recognise patterns and groups of objects without any prior knowledge of the images.**

**As a result, when tested with the test dataset, the machine will learn its patterns and distinctions, such as colour differences and shape differences, and will predict the outcome.**

**Unsupervised Machine Learning Categories**

### Unsupervised learning may be further divided into the following two categories:

### Clustering

### Association

### 1) Clustering

* We use the clustering approach to locate natural categories in the data. It is a method of categorising goods so that those that are most similar to one another stay together and have little to no resemblance to those in other groupings. The clustering method is demonstrated by categorising clients based on their purchasing behaviour. Among the popular clustering approaches are those listed below.
* **K-Means Clustering algorithm**
* **Mean-shift algorithm**
* **DBSCAN Algorithm**
* **Principal Component Analysis**
* **Independent Component Analysis**

### 2) Association

An unsupervised learning technique known as association rule learning reveals surprising correlations between variables in a huge dataset. The primary goal of this learning strategy is to discover the relationships between data points and then map the variables to maximum profit. This technology's uses include continuous manufacturing, web usage mining, market basket analysis, and so on.

Apriori, Eclat, and FP-growth are three well-known algorithms for learning association rules.

**The benefits and drawbacks of using an unsupervised learning algorithm**

**Advantages:**

* Since they operate on unlabeled datasets, these algorithms, as opposed to supervised ones, can be applied to more difficult issues.
* For many applications, unsupervised methods are chosen since obtaining the unlabeled dataset is easier than obtaining the labelled dataset.

**Disadvantages:**

## • Because the dataset is not labelled and the algorithms are not trained using the exact output in advance, the output of an unsupervised method may be less precise.

## • Why Working with unsupervised learning is more complex since it uses an unlabeled dataset that does not match to the output.

## Applications of Unsupervised Learning

## • Network System: Unsupervised learning is used in document network analysis of text data for scholarly papers to detect plagiarism and copyright.

## • suggestion Systems: Recommendation systems commonly use unsupervised learning approaches to build suggestion apps for various web applications and e-commerce websites.

## • Anomaly Detection: A frequent application of unsupervised learning, anomaly detection can identify data points that are out of the ordinary in a dataset. It is used to detect fraudulent transactions.

## • Singular Value Decomposition (SVD): This approach is used to extract specific data from databases. Taking data about each user who is present in a certain location, for example.

## Semi-Supervised Learning:

### **Semi-supervised learning is a machine learning strategy that sits somewhere between supervised and unsupervised learning. During the training phase, it employs a mix of labelled and unlabeled datasets and falls somewhere between supervised learning (with labelled training data) and unsupervised learning (without labelled training data).**

### **Despite being a middle ground between supervised and unsupervised learning, the majority of the data utilised in semi-supervised learning is unlabeled. It works with data that contains a few labels. Despite the fact that labels are costly, firms may only require a minimal number of labels. It is fundamentally distinct from supervised and unsupervised learning, which are dependent on the presence or absence of labels.**

### **To solve the drawbacks of supervised and unsupervised learning approaches, the concept of semi-supervised learning is introduced. Instead than relying mostly on labelled data, like in supervised learning, the primary purpose of semi-supervised learning is to use all available data as thoroughly as possible. Comparable data are initially clustered using an unsupervised learning technique, assisting in the conversion of unlabeled data to labelled data. This is due to the fact that tagged data is more expensive to purchase than unlabeled data.**

### **We can visualise these algorithms with an example. An instructor keeps an eye on a student while they are engaged in supervised learning, both at home and at school. Unsupervised learning occurs when a student analyses a subject on their own without the assistance of an instructor.**

### Advantages and disadvantages of Semi-supervised Learning

## ****The method is straightforward and simple to grasp, yet it is also quite effective.****

## **It is used to address issues with algorithms for supervised and unsupervised learning.**

## **The outcomes of iterations may not be stable.**

## **We are unable to use these techniques on data at the network level.**

## **Low accuracy**.****

## Reinforcement Learning

## **A software component explores its environment on its own via reinforcement learning. It takes action, learns from its mistakes, and improves with practise. The foundation of reinforcement learning is feedback. The goal of a reinforcement learning agent is to maximise rewards because it is rewarded for every positive behaviour and penalised for every negative one. In contrast to supervised learning, reinforcement learning is entirely dependent on the agents' experiences.The reinforcement learning approach is similar to that of a human being; for example, a child learns new things through everyday experiences. Reinforcement learning is demonstrated by playing a game in which an agent's actions produce states at each step and the environment acts as the game's environment.**

## Categories of Reinforcement Learning

### In reinforcement learning, there are basically two types of techniques or algorithms:

### • Positive Reinforcement Learning: Positive reinforcement learning is the practise of adding something to the needed behaviour to increase the likelihood that it will occur again. It improves the agent's behaviour and has a favourable effect on it.

### • Negative Reinforcement Learning: This method of learning is diametrically opposed to constructive RL. By avoiding the undesirable situation, the likelihood of the specific behaviour occurring again increases.

### Real-world Use cases of Reinforcement Learning

### ****Video Games:** Applications for games frequently utilise real-time learning techniques. It's used to achieve superhuman performance. Examples of well-known RL algorithms are the video games AlphaGO and AlphaGO Zero.**

### ****Management of Resources:** In order to reduce the average job slowness, the "Resource Management with Deep Reinforcement Learning" research demonstrated how to apply RL in computers to automatically train and organise resources to wait for various workloads.**

* **Robotics : In robotics applications, RL is commonly used. Robots are utilised in the manufacturing and industrial sectors, and reinforcement learning is used to boost their performance. Many industries share the objective of creating intelligent robots using AI and machine learning technology.**
* **Data Mining: One of the great uses of NLP is text mining.**

### Advantages and Disadvantages of Reinforcement Learning

**Advantages**

• Due to the similarities between the RL learning model and human learning, the most accurate results may be produced.

• It assists in the resolution of complex real-world challenges that are difficult to handle using traditional approaches.

• Assists in achieving lasting results.

**Disadvantage**

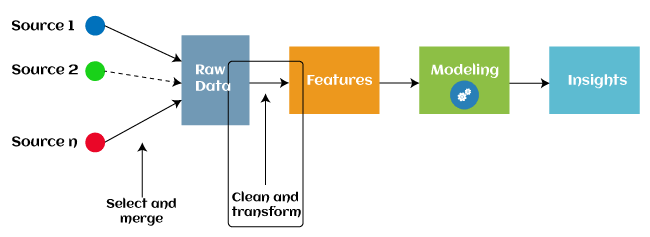
# Simple tasks are not appropriate for RL algorithms.

# RL algorithms demand enormous computer and data resources.

# An overabundance of states brought on by excessive reinforcement learning may undermine the outcomes.

# Feature Engineering for Machine Learning

**Feature engineering, a machine learning pre-processing step, is used to extract features from raw data that can then be utilised to build a prediction model using either machine learning or statistical modelling. Machine learning elements are designed to improve model performance. This topic will teach us a lot about feature engineering in machine learning. So, before we go into the specifics, let's first establish the characteristics. So, why is feature engineering required?**

**Feature engineering, the pre-processing stage of machine learning, extracts features from raw data. It helps predictive models better communicate a basic issue, enhancing the model's accuracy for unseen data. The feature engineering approach selects the best predictor variables for the model, which is made up of predictor variables and an outcome variable.**

Since 2016, some machine learning programmes that aid in automatically extracting features from raw data have also adopted automated feature engineering. Four operations make up the majority of feature engineering in machine learning: feature creation, transformations, feature extraction, and feature selection. These processes are described as below:

1. **Feature creation:** Finding the most beneficial variables to include in a predictive model is known as feature creation. The procedure needs human ingenuity and intervention and is subjective. The addition, subtraction, and ration operations used to construct the new features provide them a tremendous deal of versatility.
2. **Transformations:** The feature engineering transformation stage entails modifying the predictor variable to raise the model's precision and effectiveness. By ensuring that all the variables are on the same scale and that the model is flexible enough to accept input from a range of data, for instance, it makes the model simpler to comprehend.
3. **Feature Extraction:** A feature engineering procedure that automatically creates new variables by removing existing ones from raw data is called feature extraction. This step's major goal is to decrease the amount of data so that it can be utilised and handled for data modelling more simply. Cluster analysis, text analytics, edge detection algorithms, and principle components analysis are examples of feature extraction techniques (PCA).
4. **Feature Selection:** Only a small subset of the dataset's variables may be used to create a machine learning model; the remainder are either redundant or useless. The overall performance and accuracy of the model may suffer if all these redundant and pointless information are included in the dataset.

In order to remove the unnecessary or less significant features from the data, it is crucial to discover and choose the most appropriate features from the data, which is accomplished with the aid of feature selection in machine learning. By eliminating the duplicate, irrelevant, or noisy characteristics from the original feature set, feature selection is a technique for choosing the subset of the most pertinent features. Below are some benefits of using feature selection in machine learning:

* It aids in avoiding the dimensionality curse.
* It aids in the model's simplicity so that researchers may quickly comprehend it.
* It cuts down on training time.
* It improves generalisation by reducing over fitting.

## Types of ML Classification Algorithms:

## The following two categories can be used to further split classification algorithms:

## model linear

## logarithmic regression

## Support Vector Machines

## Non-Linear Models

## K-Nearest Neighbours

## Kernel SVM

## Naive Bayes

## Decision Tree Classification

## Random Forest Classification

## are some examples of machine learning techniques.

## Evaluating a Classification model:

When our model is finished, we must assess its performance to determine if it is a classification model or a regression model. Thus, we have the following options for assessing a classification model:

**1. Cross-Entropy Loss or Log Loss**

* **It is used to assess the effectiveness of a classifier, the result of which is a probability value between**
* **The value of log loss for a decent binary classification model should be close to 0.**
* **If the anticipated value differs from the actual value, the value of log loss rises.**
* **The lower the log loss, the more accurate the model is.**
* **Cross-entropy for binary classification may be computed as:?**
* **(ylog(p)+(1?y)log(1?p))**
* **Where p = projected output and y = actual output**
  + 1. **Confusion Matrix:**

**The confusion matrix describes the performance of the model and gives us a matrix or table as an output.**

* **The error matrix is another name for it.**
* **The matrix includes the results of the forecasts in a summary manner, including the total number of accurate and inaccurate predictions.**

**Accuracy = (TP+TN)/Total Population**

* + 1. **AUC-ROC curve:**

**AUC stands for the area under the curve, while ROC stands for the receiver operating characteristics curve.**

* **It is a graph that displays the classification model's performance at various thresholds.**
* **The AUC-ROC Curve is used to show how well the multi-class classification model is doing.**
* **The TPR and FPR are used to draw the ROC curve, with the True Positive Rate (TPR) on the Y-axis and the FPR (False Positive Rate) on the X-axis.**

**Use cases of Classification Algorithms**

# Several situations call for the usage of classification methods. These are a few frequent applications for classification algorithms:

# Voice Recognition o Spam Email Detection

# Tumour cell identifications in cancer.

# Classification of drugs, biometric identification, etc

# Logistic Regression in Machine Learning

* • Logistic regression is a well-known Machine Learning algorithm that falls under the Supervised Learning method. It is used to predict the categorical dependent variable using a specified set of independent factors.
* • The results of a dependent variable with a categorical component are forecasted using logistic regression. As a result, the outcome must be discrete or categorical. Rather of supplying exact values of 0 and 1, it delivers probabilistic values between 0 and 1. It can be Yes or No, 0 or 1, true or false, and so on.
* • The fundamental distinction between linear regression and logistic regression is in their application. Unlike logistic regression, which is used to solve regression problems, linear regression is used to solve classification problems.
* In logistic regression, we fit a "S" shaped logistic function, which predicts two maximum values, rather than a regression line (0 or 1).
* The logistic function's curve shows the possibility of several things, like whether or not the cells are malignant, whether or not a mouse is fat depending on its weight, etc.
* Since it can classify fresh data using both continuous and discrete datasets, logistic regression is a key machine learning approach.
* Logistic regression may be used to categories observations using a variety of data formats and can quickly identify the factors that will work best for the classification.

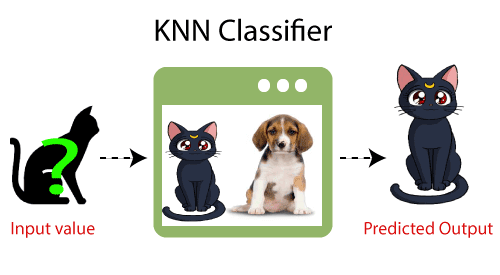
# K-Nearest Neighbor(KNN) Algorithm for Machine Learning

* K-Nearest Neighbor(KNN) Machine Learning Algorithms• K-Nearest Neighbour is a simple Machine Learning algorithm based on the Supervised Learning method.
* Assuming that the new case/data and the existing cases are similar, the K-NN algorithm places the new example in the category that is most similar to the available categories.
* The K-NN algorithm maintains all available information and categorises fresh data based on similarity. This means that new data can be swiftly and reliably classified into a suitable category using the K-NN approach.
* The K-NN algorithm can be used for both classification and regression problems, but it is most frequently used for classification issues.
* Since K-NN is a non-parametric method, it makes no assumptions about the underlying data.

• It is also known as a lazy learner algorithm since it saves the training dataset rather than learning from it immediately. Instead, it uses the dataset to execute an action when classifying data.

• The KNN algorithm simply saves the dataset during the training phase and subsequently classifies fresh data into a category that is quite similar to the new data.

• If we have a visual of a creature that resembles both a cat and a dog, but we are unsure of its identity. However, since the KNN method is based on a similarity metric, we may utilize it for this identification. Our KNN model will identify the shared characteristics in the fresh data.

• If we have a visual of a creature that resembles both a cat and a dog, but we are unsure of its identity. However, since the KNN method is based on a similarity metric, we may utilize it for this identification. Our KNN model will identify the shared characteristics in the fresh data.

**How does K-NN function?**

• The K-NN can be described using the following algorithm:

• Step 1: Determine the number of the Kth neighbour.

• In step two, compute the Euclidean distance between K neighbours.

• Step 3: Determine the K nearest neighbours based on the Euclidean distance.

• Step 4: Count how many data points are in each category among these k neighbours.

• Step 5: Assign the new data points to the category with the highest neighbour count.

• Step 6: Our model is finished.

**Conclusion**

Several machine learning algorithms are examined in this paper. Everyone today, whether purposefully or unintentionally, employs machine learning. Everything from changing photographs on social networking sites to receiving product recommendations when shopping online is possible. This book introduces the vast majority of well-known machine learning methods.

# References

1. W. Richert, L. P. Coelho, “*Building Machine Learning Systems with Python*”, Packt Publishing Ltd., ISBN 978-1-78216-140-0
2. M. Welling, “*A First Encounter with Machine Learning*”
3. M. Bowles, “*Machine Learning in Python: Essential Techniques for Predictive Analytics*”, John Wiley & Sons Inc., ISBN: 978-1-118-96174-2
4. S.B. Kotsiantis, “*Supervised Machine Learning: A Review of Classification Techniques*”, Informatica 31 (2007) 249-268
5. L. Rokach, O. Maimon, “*Top – Down Induction of Decision Trees Classifiers – A Survey*”, IEEE Transactions on Systems, D. Lowd, P. Domingos, “Naïve Bayes Models for Probability Estimation” https://webdocs.cs.ualberta.ca/~greiner/C- 651/Homework2\_Fall2008.html
6. D. Meyer, “Support Vector Machines – The Interface to libsvm in package e1071”, August 2015
7. 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.

<http://www.simplilearn.com/what-is-machine-learning-> and-why-it-matters-article

1. P. Harrington, “*Machine Learning in action*”, Manning Publications Co., Shelter Island, New York, 2012 <http://pypr.sourceforge.net/kmeans.html>
2. K. Alsabati, S. Ranaka, V. Singh, “*An efficient k- means clustering algorithm*”, Electrical Engineering and Computer Science, 1997
3. M. Andrecut, “*Parallel GPU Implementation of Iterative PCA Algorithms*”, Institute of Biocomplexity and Informatics, University of Calgary, Canada, 2008
4. 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
5. X. Zhu, “*Semi-Supervised Learning Literature Survey*”, Computer Sciences, University of Wisconsin-

Madison, No. 1530, 2005

1. R. S. Sutton, “*Introduction: The Challenge of Reinforcement Learning*”, Machine Learning, 8, Page 225-227, Kluwer Academic Publishers, Boston, 1992
2. L. P. Kaelbing, M. L. Littman, A. W. Moore, “*Reinforcement Learning: A Survey*”, Journal of Artificial Intelligence Research, 4, Page 237-285, 1996
3. R. Caruana, “*Multitask Learning*”, Machine Learning, 28, 41-75, Kluwer Academic Publishers, 1997
4. D. Opitz, R. Maclin, “*Popular Ensemble Methods: An Empirical Study*”, Journal of Artificial Intelligence Research, 11, Pages 169-198, 1999
5. Z. H. Zhou, “*Ensemble Learning*”, National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China https://en.wikipedia.org/wiki/Boosting\_(machine\_learn ing) https://en.wikipedia.org/wiki/Bootstrap\_aggregating
6. V. Sharma, S. Rai, A. Dev, “*A Comprehensive Study of Artificial Neural Networks*”, International Journal of Advanced Research in Computer Science and Software Engineering, ISSN 2277128X, Volume 2, Issue 10,

October 2012

1. S. B. Hiregoudar, K. Manjunath, K. S. Patil, “*A Survey: Research Summary on Neural Networks*”, International Journal of Research in Engineering and Technology, ISSN: 2319 1163, Volume 03, Special Issue 03, pages 385-389, May, 2014 https://en.wikipedia.org/wiki/Instance-based\_learning
2. P. Harrington,“ Machine Learning in Action”, Manning Publications Co., Shelter Island, New York,

ISBN 9781617290183, 2012

1. J. M. Keller, M. R. Gray, J. A. Givens Jr., “*A Fuzzy K- Nearest Neighbor Algorithm*”, IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-15, No. 4, August 1985.

[24] Tiwari, S.K., Neogi, S.G., Mishra, A. (2023). Design and Implementation of Security Enhancement for Trusted Cloud Computing. In: Kumar, A., Mozar, S., Haase, J. (eds) Advances in Cognitive Science and Communications. ICCCE 2022. Cognitive Science and Technology. Springer, Singapore.

[25 Tiwari,S.K. Sao,N., Kushwaha, A.,Book Title “[Fundamental of Machine Learning](https://scholar.google.com/citations?view_op=view_citation&hl=en&user=gpoUlO4AAAAJ&citation_for_view=gpoUlO4AAAAJ:eQOLeE2rZwMC)”,lambert Publication, ISBN: 9789391265335, Oct. 2022,