A Review : Machine Learning Algorithms

Aditya Vishwakarma Shiv Kumar Tiwari Pinkal Jain

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

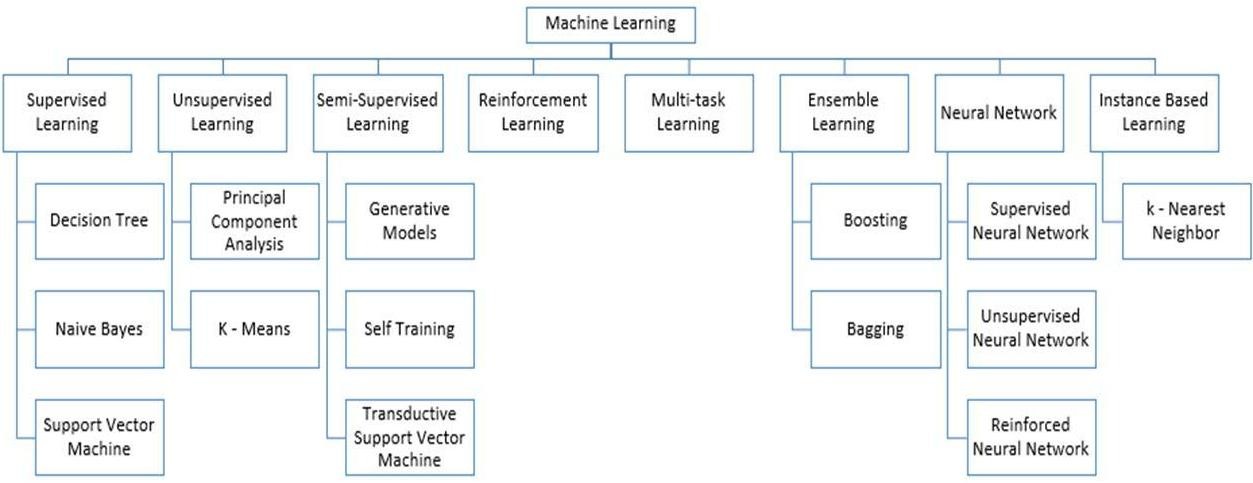
**Abstract: *In this study, various machine learning technique have been discussed. several jobs, such as data mining, image processing, predictive analytics, etc.The main advantage of machine learning is the ability of an algorithm to function independently once it has learned the utilization of data.***

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

# Introduction

To educate machines how to handle data more efficiently, machine learning is utilized. There are instances when we are unable to detect a pattern or draw conclusions from the data, even after viewing it. We use machine learning in that situation [1]. The need for machine learning has expanded due to the abundance of datasets. To obtain pertinent data, machine learning is used in numerous industries, including the military and healthcare. Machine learning aims to learn from data. Robots have been taught to learn on their own in a number of experiments [2][3]. Numerous mathematicians and programmers approach this issue in various ways. Fig. 1 depicts a number of them in use. Each machine learning approach is explained in Section 2.

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.

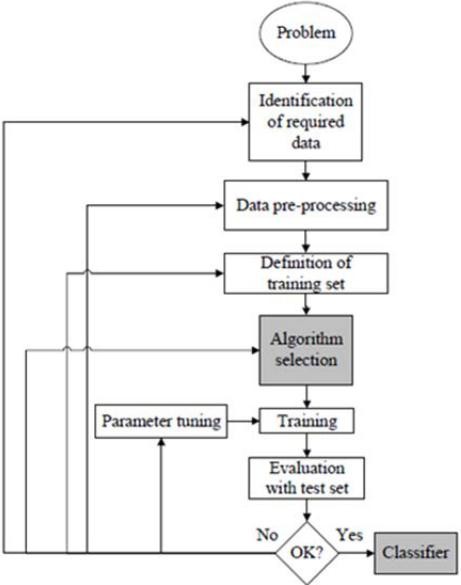
**Figure 1:** Types of Learning [2] [3]

## Supervised Machine Learning

Algorithms for supervised machine learning are ones that require outside help. There are train and test datasets for the dataset. The train dataset's output variable has to be predicted or categorised. Every algorithm employs a type of pattern from the training dataset to predict or categorise the test dataset [4]. The procedure for supervised machine learning algorithms is shown in Fig. 2. Three of the most well-liked supervised machine learning techniques are examined in this article. The foundation of supervised machine learning is supervision, as the name suggests. Using supervised learning, the "labelled" dataset is used to train the computers, and after training, the computer predicts the result.

The offered data here allows one to examine which inputs have previously been converted into which outputs. Simply said, we train the computer with input and associated output before asking it to predict the outcomes using test datasets. Let's use an illustration to clarify supervised learning. Take into account that both cats and dogs are represented in the dataset we are utilising as our input. Following training, we enter a cat image and ask the computer to recognise it and forecast the outcome.

The machine will conclude that the object is a cat after learning more about it and analysing its height, shape, colour, eyes, ears, tail, and other traits.

During supervised learning, the computer use this technique to identify the objects.

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

The main goal of the supervised learning approach is to create a map between the input variable (x) and the output variable (y). Risk assessment, fraud detection, and spam filtering are supervised learning applications that are used in the real world. 

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

### 1.Supervised Machine Learning Categories

### There are two sorts of supervised machine learning issues, which are outlined below:

### Classification

### Regression

### Classification - When the output variable is categorical, such as "Yes" or "No," Male or Female, Red or Blue, etc., classification techniques are employed to solve the problem. The categories that are present in the dataset are predicted by the categorization algorithms. Systems for categorising content are now in use, including spam detection and email filtering. The list of popular classification algorithms is as follows:

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

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

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

1. **Regression**

When there is a linear connection between the input and output variables, regression techniques are utilised to solve the problem. They are used to forecast variables that have predictable outcomes, such as market trends and weather forecasts, among other things.

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:**

* The training data will give you a clear sense of the classes.
* You can easily comprehend the process of supervised learning. Unsupervised learning makes it difficult for us to comprehend the inner workings of the computer, how it learns, etc.
* Before providing the data for training, you can determine the precise number of classes.
* You may train the classifier in a way that has a perfect decision boundary to precisely discriminate between distinct classes, allowing you to be very exact about the description of the classes.
* You don't necessary need to retain the training data in your memory once the full programme is through. You can continue to use the decision boundary's mathematical formula as an alternative.

**Disadvantages:**

### Supervised learning is constrained in a number of ways, making it unable to tackle some of the most challenging machine learning tasks.

### Unlike unsupervised learning, supervised learning is unable to extract unknown information from training data.

### In contrast to unsupervised learning, it is unable to cluster or categorise data by independently determining its characteristics.

### In the case of classification, the output could be a false class label if the input is not from one of the classes in the training data. Consider training an image classifier using data on cats and dogs. The outcome may thus be either a cat or a dog if you supply the image of a giraffe, which is incorrect.

### Applications of Supervised Learning

## Following are a few typical uses of supervised learning:

## • Image Segmentation: Supervised learning-based algorithms are used for this task. This approach does image classification on a variety of picture data using pre-established labels.

## The underlying diagnosis in the realm of medicine, supervised algorithms are frequently employed for diagnostic reasons. It is performed with historical information that contains labels for illustrations of ailments and details of medicine. The machine may employ this method to identify a sickness in brand-new patients.

## The identification of fraudulent consumers, transactions, etc. is made possible by the use of supervised learning algorithms. To find patterns that can point to possible fraud, historical data is used.

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

## 2. Unsupervised Machine Learning

### Because it doesn't need monitoring, unsupervised learning differs from the supervised learning technique. Unsupervised machine learning is the process by which a system learns from an unlabeled dataset and subsequently generates predictions without the assistance of a human. After being trained on data, unsupervised learning enables the models to work freely on data that hasn't been classified or labelled.

### The classification or grouping of the unsorted dataset into categories based on patterns, similarities, and differences is the main objective of the unsupervised learning technique. The machines are charged with finding hidden patterns in the input dataset.

**To make things easier for you to comprehend, let's use an illustration. Let's say we provide the machine learning model with photos of a fruit basket. The model's job is to recognise groups of items and patterns in an image without any prior knowledge of the images.**

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

**Unsupervised Machine Learning Categories**

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

### Clustering

### Association

### 1) Clustering

To find the logical groupings in the data, we employ the clustering technique. It is a strategy for organising items into groups so that those that are most similar to one another remain together and have little to no similarity to those in other groups. It is explained how to classify clients based on how they often make purchases. The popular clustering methods are included in the list below.

**K-Means Clustering algorithm**

* **Mean-shift algorithm**
* **DBSCAN Algorithm**
* **Principal Component Analysis**
* **Independent Component Analysis**

### 2) Association

In a huge dataset, association rule learning, an unsupervised learning technique, finds amazing correlations between variables. The main objectives of this learning technique are to identify the correlations between the data points and then map the variables for maximum profit. Continuous manufacturing, web use mining, market basket research, and more uses for this method are available.

Apriority, Éclat, and FP-growth are a few 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 labeled dataset.

**Disadvantages:**

## The output of an unsupervised algorithm may be less precise because the dataset is not labelled and the algorithms are not trained using the exact output in advance.

## Why Because unsupervised learning employs a dataset that is unlabeled and does not correspond to the output, working with it is more difficult.

## Applications of Unsupervised Learning

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

## • Recommendation Systems: Using unsupervised learning approaches, recommendation systems typically create suggestion applications for a variety of online apps and e-commerce websites.

## • Anomaly Detection: A popular use of unsupervised learning, anomaly detection identifies unusual data points within a collection. It is used to track down fraudulent transactions.

## • Singular Value Decomposition (often referred to as SVD): This method is employed to extract certain data from databases. capturing data about each user who is present in a certain location, for example.

## Semi-Supervised Learning:

### **A machine learning method that falls in between supervised and unsupervised learning is called semi-supervised learning. It employs a combination of labelled and unlabeled datasets during the training phase, falling in between supervised learning (with labelled training data) and unsupervised learning (without labelled training data) approaches.**

### **Semi-supervised learning, which falls between supervised and unsupervised learning, mostly employs unlabeled data. It functions with data that has a few labels. Businesses could only need a few labels, despite the fact that they are costly. It is completely distinct from supervised and unsupervised learning, which depend on the presence or absence of labels.**

### **The idea of semi-supervised learning is proposed to address the shortcomings of supervised learning and unsupervised learning methods. Instead than depending mostly on tagged data as in supervised learning, the major goal of semi-supervised learning is to employ all available data as thoroughly as is practicable. An unsupervised learning technique is used to first cluster comparable data, converting the unlabeled data into labelled data. This is because tagged data is more expensive to buy than unlabeled data.**

### **An example will help us visualise these algorithms. When a student engages in supervised learning at home or at school, a teacher is watching them closely. When a student undertakes independent research on a subject without the teacher's direction, it is known as unsupervised learning.**

### 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

## **Reinforcement learning is used by a software component to explore its surroundings independently. Action is taken, mistakes are made, and with practise, it gets stronger. Feedback is the cornerstone of reinforcement learning. Because a reinforcement learning agent is rewarded for every successful action and penalised for every failed one, its objective is to maximise rewards. In contrast to supervised learning, reinforcement learning is totally based on the experiences of the agents.**

## **The process of reinforcement learning is similar to how humans learn; for instance, a young child picks up new information from experiences in his daily life. Reinforcement learning is exemplified 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 essentially two types of methods or algorithms used:

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

### • Negative Reinforcement Learning: This method of teaching is in direct opposition to constructive RL. By avoiding the undesirable circumstance, it makes it more likely that the specific behaviour would happen again.

### Real-world Use cases of Reinforcement Learning

**• Video games: Real-time learning techniques are commonly used in gaming applications. It helps people perform at a superhuman level. The video games AlphaGO and AlphaGO Zero are two examples of well-known RL algorithms.**

**• Resource Management: The "Resource Management with Deep Reinforcement Learning" research showed how to use RL in computers to automatically train and organise resources to wait for different workloads, reducing the average job slowness.**

**• Robotics: RL is frequently utilised in robotics applications. Reinforcement learning is used to improve robot performance in the manufacturing and industrial sectors. The goal of developing intelligent robots utilising AI and machine learning technologies is shared by many sectors.**

* **Data Mining: One of the great uses of NLP is text mining.**

### Advantages and Disadvantages of Reinforcement Learning

**Advantages**

• The most accurate findings could be generated as a result of the parallels between the RL learning model and human learning.

• It helps resolve difficult-to-handle real-world problems that are complicated and require innovative solutions.

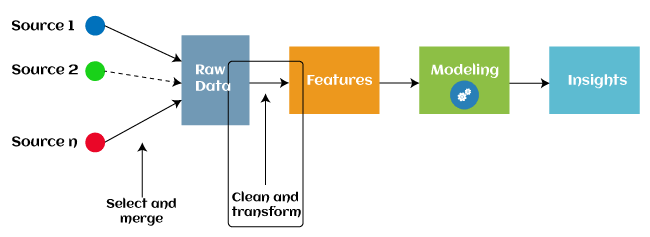
• Helps provide long-lasting results.

**Disadvantage**

# RL algorithms are not ideal for simple tasks, they use a lot of computer and data resources, and they may produce too many states as a result of too aggressive reinforcement learning.

# Feature Engineering for Machine Learning

**To extract features from raw data that may be used to create a prediction model using either machine learning or statistical modelling, feature engineering is a pre-processing phase in machine learning. The creation of machine learning components improves model performance. In this topic, we'll discover a lot about feature engineering in machine learning. So before we dive into the details, let's first define the qualities. Therefore, why is feature engineering required?**

**The pre-processing phase of machine learning, called feature engineering, extracts features from raw data. It helps to clarify a key problem for predictive models, improving the model's accuracy for unseen data. The model, which consists of predictor variables and an outcome variable, is built using the feature engineering approach, which selects the most useful predictor variables.**

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. As follows is a description of these procedures:

1. Finding the most useful variables to use in a predictive model is referred to as feature generation. The process is arbitrary and dependent on human inventiveness and interference. They are incredibly versatile thanks to the addition, subtraction, and ration operations that were employed to build the additional features.

2.Transformations: During the feature engineering transformation step, the predictor variable is changed to increase the model's efficacy and accuracy. By ensuring that all the variables are on the same scale and that the model is flexible enough to receive input from a number of sources, for instance, it makes the model simpler to grasp.

3. Engineering approach known as "feature extraction" that substitutes new variables for the ones that are currently present. Feature Extraction: A feature engineering technique known as feature extraction automatically adds new variables by eliminating the ones that already exist from raw data. The main objective of this stage is to reduce the amount of data in order to make it easier to use and manage for data modelling. Examples of feature extraction methods (PCA) include cluster analysis, text analytics, edge detection algorithms, and principal components analysis.

4.Feature Selection: Only a tiny portion of the dataset's variables can be used to create a machine learning model; the rest are either redundant or useless. The overall effectiveness and accuracy of the model may suffer if all of these pointless and redundant pieces of data are included to the dataset. It is essential to identify and pick the most appropriate features from the data in order to remove the superfluous or less important information, which is performed with the help of feature selection in machine learning. Feature selection is a method for selecting the subset of the most important features by removing the redundant, unnecessary, or noisy qualities from the initial feature set. The advantages of machine feature selection are listed below.

## 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 Bays

## 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:?**
* **(log(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

* One of the most popular machine learning algorithms utilised with the supervised learning method is logistic regression. Using a predetermined set of independent factors, it is used to forecast the categorical dependent variable.
* Logistic regression predicts the results of a dependent variable having a categorical component. Consequently, the result must be a discrete or categorical value. Instead of the precise values between 0 and 1, it provides the probabilistic values that fall between those numbers. Either True or False, 0 or 1, or Yes or No, are possible outcomes.
* The main difference between logistic regression and linear regression is the application technique. Logistic regression is used to solve regression issues, while linear regression is used to address classification difficulties.
* Logistic regression may be used.
* 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 Neighbour (KNN) Algorithm for Machine Learning K-Nearest Neighbours is one of the most basic supervised learning-based machine learning algorithms.

• The K-NN approach places the new example in the category that is most similar to the other categories that are available, assuming that the new case/data and the prior examples are comparable.

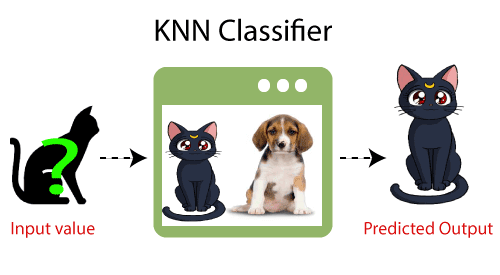
• The K-NN method maintains all of the old data while classifying new data based on similarity. This shows that the K-NN method can rapidly and accurately categorise fresh data.

• While the K-NN approach may be used to solve classification and regression issues, it is most typically applied to classification problems.

• K-NN doesn't make any assumptions about the underlying data because it employs a non-parametric method.

• 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 see a creature that resembles both a dog and a cat but we are unsure of its identity. However, since the KNN technique is based on a similarity measure, it may be utilised for this identification. We will find the similarities in the new data using our KNN model.



**How does K-NN function?**

* The K-NN operates according to the following algorithm:
* Choose the Kith neighbor's phone number in step one.
* In step two, get the Euclidean distance between K neighbors.
* Step 3: Select the K closest neighbors based on the calculated Euclidean distance.
* Step 4: Among these k neighbors, count the number of data points in each category.
* Step 5: Assign the fresh data points to the group with the highest neighbor count.
* Step 6: The model is finished.

**Conclusion**

This study surveys a wide range of machine learning algorithms. Nowadays, everyone employs machine learning, whether consciously or unconsciously. By changing pictures on social networking sites to getting product suggestions when shopping online. This publication introduces the vast majority of the well-known machine learning techniques.

# 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,