A Review : Machine Learning Algorithms

Aditya Vishwakarma Shiv Kumar Tiwari Pinkal Jain

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

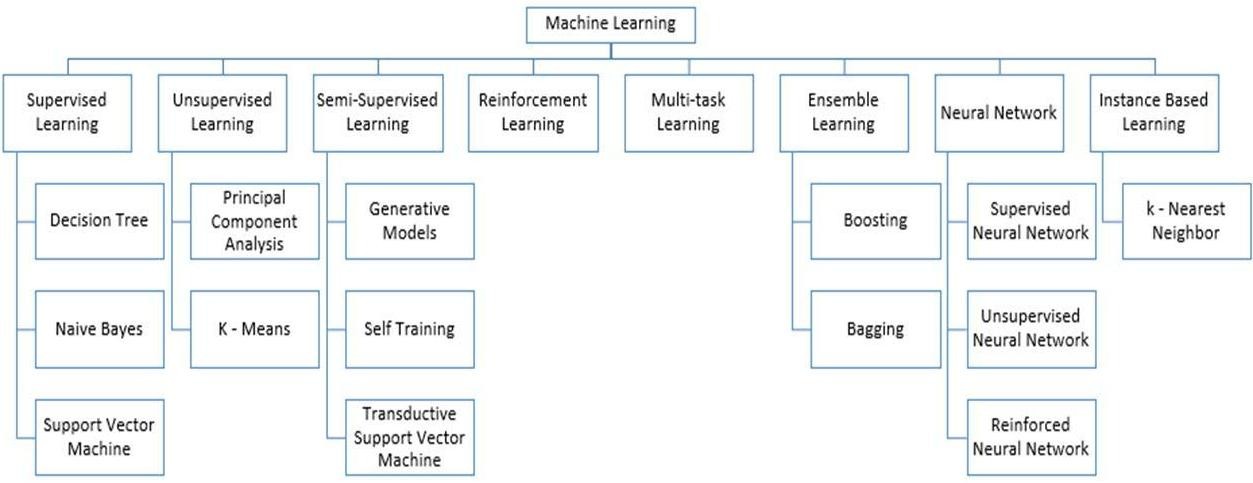
**Abstract: *In this lessons, various machine learning technique have been discussed. Several jobs, such as data mining, image processing, predictive analytics, etc. The main advantage of mechanism learning is the ability of an algorithm to function independently once it has learned the utilization of 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 seeks to get information from data. Several experiments have taught robots to learn on their own [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.

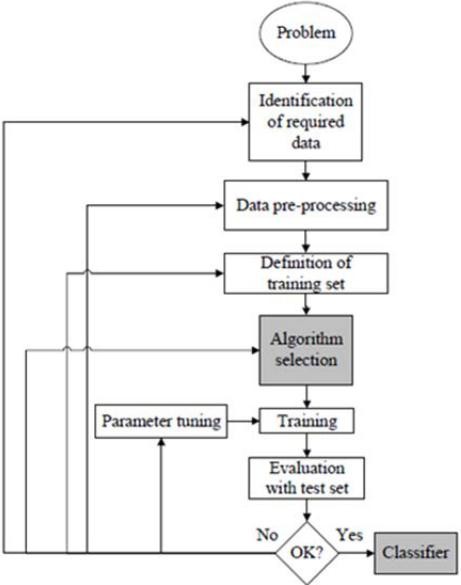
**Figure 1:** Types of Learning [2] [3]

## Supervised Machine Learning

The algorithms that call for outside assistance are those for supervised machine learning. For the dataset, there are train and test datasets. The output variable from the train dataset has to be predicted or classified. To predict or classify the test dataset, every algorithm uses a certain type of pattern from the training dataset [4]. Fig. 2 depicts the process for supervised machine learning algorithms. This article examines three of the most popular supervised machine learning methods. As the name implies, supervision is the cornerstone of supervised machine learning. The "labelled" dataset is used to train the computers using supervised learning, and after training, the computer predicts the outcome.

One can evaluate which inputs have been changed into which outputs by using the data provided here. Simply said, before asking it to predict the outcomes using test datasets, we train the computer with input and its corresponding output. Let's explain supervised learning with an example. Keep in mind that the dataset we are using as our input includes both cats and dogs. After training, we enter a picture of a cat and ask the computer to identify it and predict the result.

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 strategy is to create a map between the input variable (x) and the output variable (y). Risk evaluation, fraud detection, and spam filtering are supervised learning applications used in the real world. 

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

### 1.Supervised Machine Learning Categories

### The following lists two categories of supervised machine learning problems:

### Classification

### Regression

### Classification - Classification strategies are used to address the issue when the output variable is categorical, such as "Yes" or "No," Male or Female, Red or Blue, etc. The classification algorithms forecast the categories that are present in the dataset. There are currently systems in use for classifying content, such as spam detection and email filtering.

### The list of popular classification algorithms is as follows:

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

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

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

1. **Regression**

Regression methods are used to address problems when the input and output variables are linearly related. They are used, among other things, to anticipate variables with predictable results, such market trends and weather predictions.

The popular regression algorithms on the list below include:

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

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

* The training data will make it easier for you to comprehend the lessons.
* You have little trouble grasping the principles of supervised learning. In the case of unsupervised learning, it is difficult for us to comprehend the inner workings of the computer, how it learns, etc.
  + The specific number of classes may be determined before supplying the data for training.
  + You may train the classifier to have a flawless decision boundary that allows it to differentiate between different classes with great accuracy, enabling you to be extremely specific when describing the classes.
  + Once the entire programme is through, you don't necessarily need to keep the training data in your memory. The mathematical expression for the decision boundary can still be used.

**Disadvantages:**

### • Because of its limitations, supervised learning is unable to complete some of the most difficult machine learning tasks.

### • Supervised learning cannot extract unknown information from training data, in contrast to unsupervised learning.

### • Unlike unsupervised learning, it is unable to cluster or categorise data by figuring out its qualities on its own.

### • If the input does not come from one of the classes in the training data, the classification output can contain a bogus class label. Consider utilising data on dogs and cats to train an image classifier. Thus, if you provide the incorrect image of a giraffe, the result may be either a cat or a dog.

### Applications of Supervised Learning

## Here are a few typical examples of supervised learning applications:

## • Image segmentation: Algorithms based on guided learning are utilised for this purpose. Using pre-established labels, this method classifies images from a range of picture data.

## Supervised algorithms are commonly used for diagnostic purposes in the field of medicine, which provides the underlying diagnosis. It is carried out using historical data that includes labels for depictions of illnesses and information on medications. This technique may be used by the machine to detect a disease in brand-new patients.

## The identification of fraudulent consumers, transactions, etc. is made possible by the use of supervised learning algorithms.

## Historical data is utilised to identify trends that may indicate fraud.

## Spam detection - Spam detection and filtering make use of classification techniques.

## 2. Unsupervised Machine Learning

### Unsupervised learning is different from supervised learning since it doesn't require monitoring. Unsupervised machine learning is the method through which a system picks up new information from an unlabeled dataset and then makes predictions without the help of a human. Unsupervised learning gives the models the freedom to operate on data that hasn't been tagged or categorised after they've been trained on data.

The primary goal of the unsupervised learning technique is to classify or group the unsorted dataset into groups based on patterns, similarities, and differences. The computers' task is to mine the input dataset for hidden patterns.

**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 consequence, the computer will now recognise its patterns and distinctions, such as colour differences and shape differences, and anticipate the outcome when tested with the test dataset.**

**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 identify the logical groupings in the data. It is a method of grouping things such that those who are most similar to one another stay together and are barely related to those in other groupings. It is described how to categorise customers depending on how frequently they make purchases. The list below includes examples of common clustering techniques.

**K-Means Clustering algorithm**

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

### 2) Association

Association rule learning, an unsupervised learning technique, reveals astonishing correlations between variables in a huge sample. Finding correlations between the data points and then mapping the variables for optimum gain are the major goals of this learning approach. There are several applications for this technique, including continuous production, web usage mining, market basket research, and more.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 could be less accurate because the dataset is not tagged and the algorithms are not trained using the precise output in advance.

• Why Since unsupervised learning employs a dataset that is unlabeled and does not match the result, 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: This common use of unsupervised learning finds out-of-the-ordinary data items in a collection. To find fraudulent transactions, it is employed.

## • Singular Value Decomposition, or SVD, is a technique used to extract specific data from databases. obtaining information on each user who is at a specific place, for instance.

## Semi-Supervised Learning:

### **A machine learning method that falls somewhere between supervised and unsupervised learning is called semi-supervised learning. It uses a combination of labelled and unlabeled datasets during the training phase, falling 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.**

### **To overcome the drawbacks of supervised learning and unsupervised learning techniques, the concept of semi-supervised learning is put forth. In contrast to supervised learning, which mostly relies on tagged data, semi-supervised learning's main objective is to use every piece of accessible data as extensively as is practical. Prior to labelling the unlabeled data, similar data is first clustered using an unsupervised learning approach. This is due to the fact that tagged data costs more to purchase 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. A reinforcement learning agent's goal is to maximise rewards since it receives rewards for every successful action and punishments for every unsuccessful one. Reinforcement learning is entirely reliant on the experiences of the agents, as opposed to supervised learning.**

## **Reinforcement learning works in a manner akin to how people learn; for example, a young toddler learns new information via encounters in his everyday life. Playing a game where an agent's activities result in states at each step and the environment serves as the game's environment is a good example of reinforcement learning.**

## Categories of Reinforcement Learning

### In reinforcement learning, there are essentially two types of methods or algorithms used:

### • Positive Reinforcement Learning: The practise of adding something to the required conduct to increase the likelihood that it will happen again is known as positive reinforcement learning. It improves and favourably affects the agent's conduct.

### • Negative Reinforcement Learning: Constructive RL is in direct conflict with this approach of teaching. Making the unfavourable situation less likely increases the likelihood that the particular activity would occur again.

### Real-world Use cases of Reinforcement Learning

**• Video games: Gaming apps frequently employ real-time learning strategies. It facilitates superhuman performance. Two well-known RL algorithms are the video games AlphaGO and AlphaGO Zero.**

**• Resource Management: The "Resource Management with Deep Reinforcement Learning" research shown how to automate training and organising resources to wait for various workloads, decreasing the average job slowness, using RL in computers.**

**• Robotics: Applications of robotics typically use RL. Robot performance is enhanced through reinforcement learning in the manufacturing and industrial sectors. 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**

• 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.

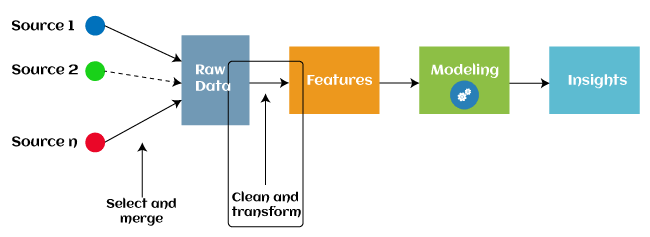
**Disadvantage**

# 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

**Feature engineering is a pre-processing stage in machine learning to extract features from raw data that may be used to build a prediction model using either statistical modelling or 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 feature engineering stage of machine learning involves taking the raw data and extracting its features. It aids in the clarification of a significant issue for predictive models, increasing the model's accuracy for unobserved data. The feature engineering technique is used to build the model, which comprises of the most practical predictor variables and an outcome variable.**

Since 2016, a number of machine learning programmes have used automated feature engineering to help with the automatic extraction of features from raw data.

The majority of feature engineering in machine learning is composed of four operations: feature generation, feature transformations, feature extraction, and feature selection. As follows is a description of these procedures:

1. The process of generating features for a predictive model is known as feature generation. The procedure is random and subject to human meddling and invention. Because the new features were built using addition, subtraction, and ration operations, they are highly adaptable.

2. Transformations: To improve the efficacy and accuracy of the model, the predictor variable is modified during the feature engineering transformation stage. By making sure that all the variables are on the same scale and that the model is flexible enough to receive input from many sources, for instance, it makes the model easier 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. Cluster analysis, text analytics, edge detection algorithms, and principle components analysis are a few examples of feature extraction techniques (PCA).

4.**Feature Selection**: Only a tiny portion of the dataset's variables can be used to create a machine learning model; the rest are unnecessary or useless. The overall effectiveness and accuracy of the model may decrease if all of these pointless and redundant pieces of data are added to the dataset. The feature selection method of machine learning is used to find and choose the most important qualities from the data in order to remove irrelevant or pointless information. 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 utilised to evaluate a classifier's performance, with the resulting probability value falling between**

**• A good binary classification model should have a log loss value that is near to 0.**

**• The value of log loss increases if the predicted value and the actual value are different.**

**• The accuracy of the model increases with decreasing log loss.**

**• How can the cross-entropy for binary classification be calculated?**

**• (log(p)+(1?y)log(1?p))**

**• Where y = actual output and p = predicted output**

**2.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**

**3.AUC-ROC curve:**

**The terms "area under the curve" (AUC) and "receiver operating characteristics curve" (ROC) are interchangeable.**

**• A graph is used to show how well the classification model performs at various thresholds.**

**• The performance of the multi-class classification model is demonstrated using the AUC-ROC Curve.**

**• The ROC curve is drawn using the TPR and FPR, with the 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 one of the most often used machine learning algorithms in conjunction with the supervised learning approach. It is used to forecast the categorical dependent variable using a specified set of independent variables.
* • A categorical dependent variable's results are predicted via logistic regression. The outcome must thus be a discrete or categorical value. It offers the probabilistic values that lie between 0 and 1 rather than the exact values between 0 and 1. Possible results include True, False, 0 or 1, as well as Yes or No.
* • The method of application is the primary distinction between logistic regression and linear regression. Regression problems are resolved using logistic regression, whereas classification problems are resolved using linear regression.
* 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

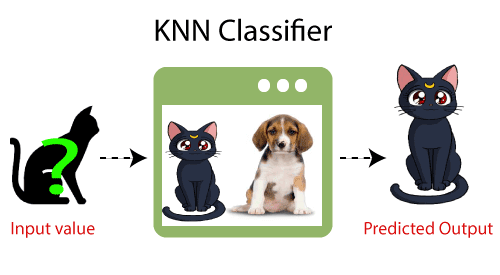
• Machine Learning K-Nearest Neighbour (KNN) Algorithm One of the most fundamental supervised learning-based machine learning algorithms is K-Nearest Neighbours.

The K-NN technique, which assumes that the new case/data and the previous instances are comparable, places the new example in the category that is most similar to the other categories that are available. The K-NN approach classifies new data based on similarity while maintaining all of the previous data. This demonstrates the speed and accuracy with which the K-NN algorithm can classify brand-new data.

• The K-NN technique may be used to address classification and regression difficulties, however it is most frequently employed to address classification concerns.

• K-NN uses a non-parametric approach, thus it doesn't make any assumptions about the underlying data.

• During the training phase, the KNN algorithm simply saves the dataset and then classifies new data into a category that is quite similar to the new data.

• When a creature that resembles both a dog and a cat is spotted and its identification is unclear. The KNN approach, however, may be used for this identification because it is based on a similarity measure. Using our KNN model, we will look for similarities in the new data.

**How does K-NN function?**

• The algorithm used by the K-NN is as follows:

• In the first stage, select the Kith neighbor's phone number.

• Measure the Euclidean distance between K neighbours in step two.

Step 3: Based on the determined Euclidean distance, choose the K neighbours that are closest to you.

Step 4: Count the number of data points in each category among these k neighbours.

• Step 5: Distribute the new data points to the group with the greatest number of neighbours.

• The model is complete after step six.

**Conclusion**

# Several different machine learning algorithms are examined in this paper. Nowadays, machine learning is used by everyone, whether they are aware of it or not. updating your profile image on social networking sites or receiving product recommendations when you purchase online. The great bulk of the well-known machine learning techniques are introduced in this work.

# 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,

[26] [Tiwari, S.K.](https://www.scopus.com/authid/detail.uri?authorId=58140299900), [Neogi, S.G.](https://www.scopus.com/authid/detail.uri?authorId=35759211300)Design and Implementation of Enhanced Security Algorithm for Hybrid Cloud using Kerberos[SN Computer Science](https://www.scopus.com/authid/detail.uri?authorId=58140299900#disabled) , 2023, 4(5), 430.

[27] [Tiwari, S.K.](https://www.scopus.com/authid/detail.uri?authorId=58140299900), [Neogi, S.G.](https://www.scopus.com/authid/detail.uri?authorId=35759211300), [Mishra, A.](https://www.scopus.com/authid/detail.uri?authorId=58280425700)Design and Implementation of Security Enhancement for Trusted Cloud Computing[Cognitive Science and Technology](https://www.scopus.com/authid/detail.uri?authorId=58140299900#disabled)  , 2023, pp. 1047–1057