

# Supervised and Unsupervised Machine Learning Approaches—A Survey



C. Esther Varma and Puja S. Prasad

**Abstract** Machine learning task is broadly divided into supervised and unsupervised approaches. In supervised learning, output is already known and we have to train the model by giving lot of data called labeled dataset to train our model. The main goal is to predict the outcome. It includes regression and classification problem. In unsupervised learning, no output mapping with input as well as it is independent in nature. The dataset used in unsupervised machine learning is unlabeled. The main focus of this paper is to give detailed understanding of supervised and unsupervised machine learning algorithm with pseudocodes.

**Keywords** Supervised · Unsupervised · Decision tree · K-means · PCA

## 1 Introduction

Machine learning is utilized to show machines how to deal with the information all the more productively. In some cases subsequent to review the information, we cannot decipher the example or concentrate data from the information. All things considered, we apply AI [1]. With the plenitude of datasets accessible, the interest for AI is in ascent. Numerous businesses from medication to military apply AI to remove pertinent data. The motivation behind AI is to gain from the information [2]. Many investigations have been done on the most proficient method to cause machines to learn without help from anyone else [3]. Numerous mathematicians and software engineers apply a few ways to deal with find the arrangement of this issue [3]. Some of them are exhibited in [4, 5]. The different types of machine learning algorithms have been depicted in Fig. 1.

---

C. Esther Varma (✉) · P. S. Prasad  
GCET, Hyderabad, India  
e-mail: [esthervarmait@gmail.com](mailto:esthervarmait@gmail.com)

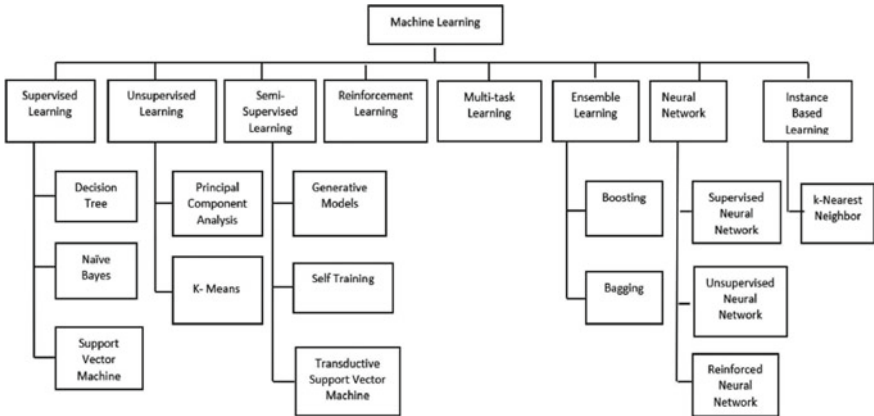


Fig. 1 Machine learning algorithms

## 2 Different Kinds of Learning

The learning directed artificial algorithm is those calculations which need outside help [6]. The info dataset is partitioned into train and test dataset. The train dataset has yield variable which should be anticipated or grouped [2]. All calculations take in some sort of examples from the preparation data set and apply them to the test dataset for expectation or grouping [7, 8].

### 2.1 Supervised Machine Learning

The flowchart of supervised machine learning algorithm has been illustrated in Fig. 2. Three most well-known directed artificial algorithm has been talked about here.

1. Decision tree
2. Naive Bayes

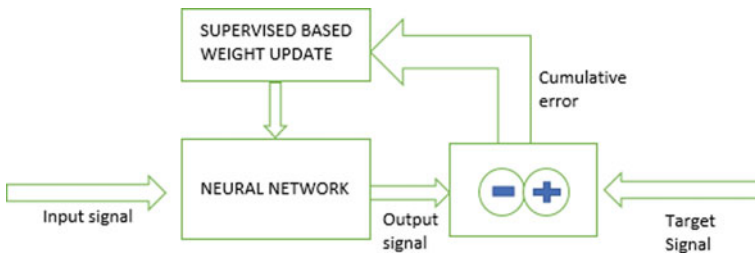
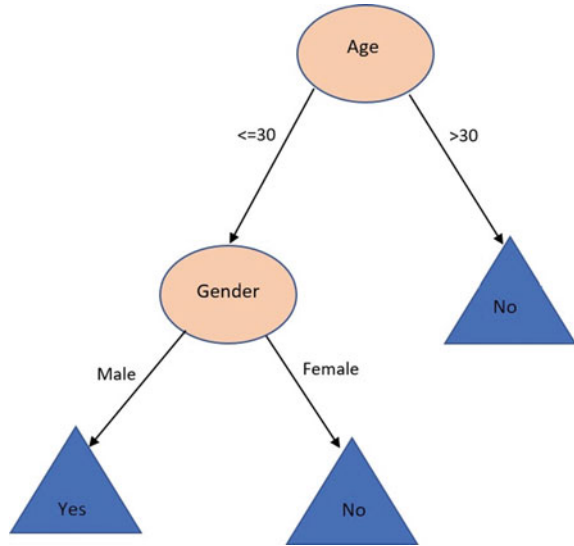


Fig. 2 Supervised machine learning algorithm

**Fig. 3** Decision Tree

### 3. Support vector machine

**Decision Tree:** Decision trees are those kind of trees which gather credits by arranging them dependent on their qualities [9]. Choice tree is utilized primarily for grouping reason. Each tree comprises hubs and branches [10]. Every hub addresses credits in a gathering that will be arranged and each branch addresses a worth that the hub can take [11]. A model of decision tree has been shown in Fig. 3. There are two types of decision tree and are based totally at the type of goal variable we have. It may be of two sorts:

**Categorical Variable Decision Tree:** Decision tree which has a specific target variable is called a categorical variable decision tree [12].

**Continuous Variable Decision Tree:** Decision tree which has a continuous target variable is called continuous variable decision tree.

Decision bushes classify the examples through sorting them down the tree from the basis to some leaf node, with the leaf node supplying the class to the instance. Each node in the tree acts as a test case for a few attribute, and every edge descending from that node corresponds to one of the feasible answers to the check case. This process is recursive in nature and is repeated for each subtree rooted at the brand new nodes [13]. Decision tree is simple to apprehend, interpret and visualize. Decision trees implicitly perform variable screening or function choice. It can manage both numerical and express information. It can also take care of multi-output problems. Decision trees require incredibly little attempt from customers for statistics coaching. Nonlinear relationships among parameters do no longer have an effect on tree performance. The pseudocode for decision tree is portrayed below, where  $S$ ,  $A$  and  $y$  are preparing set, input quality and target characteristic separately [14].

## Pseudocode for Decision Tree

```

procedure DTInducer(S, A, y)
  1: T = TreeGrowing(S, A, y)
  2: Return TreePruning(S, T)
procedure TreeGrowing(S, A, y)
  1: Create a tree T
  2: if one of the Stopping Criteria is fulfilled then
  3:   Mark the root node in T as a leaf with the most common
     value of y in S as the class.
  4: else
  5:   Find a discrete function  $f(A)$  of the input attributes values
     such that splitting S
     according to  $f(A)$ 's outcomes  $(v_1, \dots, v_n)$  gains the best split-
     ting metric.
  6:   if best splitting metric  $\geq$  threshold then
  7:     Label the root node in T as  $f(A)$ 
  8:     for each outcome  $v_i$  of  $f(A)$  do
  9:       Subtreei = TreeGrowing( $\sigma_{f(A)=v_i} S, A, y$ ).
  10:      Connect the root node of T to Subtreei, with an
     edge that is labeled as  $v_i$ 
  11:    end for
  12:   else
  13:     Mark the root node in T as a leaf with the most
     Common value of y in S as the class.
  14:   end if
  15: end if
  16: Return T
procedure TreePruning(S, T, y)
  1: repeat
  2:   Select a node t in T such that pruning it maximally
     improve some evaluation criteria
  3:   if  $t \neq \emptyset$  then
  4:     T = pruned(T, t)
  5:   end if
  6: until  $t = \emptyset$ 
  7: Return T

```

Naive Bayes: Two important styles of Naive Bayes algorithms are:

Gaussian Naive Bayes: Gaussian Naive Bayes is a variation of Naive Bayes that follows Gaussian normal distribution and supports continuous facts. Naive Bayes is a collection of supervised gadget getting to know category algorithms based on the Bayes theorem. It is a simple class approach, but has excessive functionality.

Multinomial Naive Bayes: The Gaussian assumption just described is in no way the most effective easy assumption that might be used to specify the generative distribution for every label. Another useful instance is multinomial naive Bayes, where the features are assumed to be generated from a simple multinomial distribution [15]. The multinomial distribution describes the possibility of watching counts among some of categories, and accordingly multinomial Naive Bayes is most appropriate for features that represent counts or count number rates. Mostly focuses on the text order industry [16]. It is primarily utilized for bunching and order reason [17]. The fundamental engineering of Bayes relies upon the restrictive likelihood. It makes trees dependent on their likelihood of occurring. These trees are otherwise called Bayesian network.

## Pseudocode of Naive Bayes

```

INPUT: training set  $T$ , hold-out set  $H$ , initial number of components  $k_0$ ,
and convergence threshold  $\delta_{EM}$  and  $\delta_{Add}$ 

Initialize  $M$  with one component.
 $k \leftarrow k_0$ 
repeat
  Add  $k$  new mixture components to  $M$ , initialized using  $k$ 
  random examples from  $T$ .
  Remove the  $k$  initialization examples from  $T$ .
  repeat
    E-step: Fractionally assign examples in  $T$  to mixture components, using  $M$ .
    M-step: Compute maximum likelihood parameters for  $M$ ,
    using the filled in data.
    If  $\log P(H | M)$  is best so far, save  $M$  in  $M_{beat}$ .
    Every 5 cycles, prune low-weight components of  $M$ .
  until  $\log P(H | M)$  fails to improve by ratio  $\delta_{Add}$ .
  Execute E-step and M-step twice more on  $M_{beat}$  using examples from both  $H$  and  $T$ .
Return  $M_{beat}$ .

```

**Support Vector Machine:** Another most generally utilized best in class AI procedure is support vector machine (SVM). It is mostly utilized for characterization. SVM deals with the rule of edge computation [18]. It fundamentally draws edges between the classes. The edges are attracted such a design that the distance between the margin and the classes is maximum and hence minimizing the classification error [14]. The SVM kernel is a feature that takes low-dimensional enter space and transforms it into better-dimensional area, i.e., it converts no longer separable trouble to separable trouble. It is generally beneficial in nonlinear separation issues. Simply placed the kernel, it does some extraordinarily complicated data alterations then finds out the procedure to separate the facts primarily based on the labels or outputs described. Support vector machine has several advantages—very effective in excessive dimensional cases. Its reminiscence green because it makes use of a subset of training points in the decision function referred to as guide vectors. Different kernel capabilities may be specified for the selection capabilities and its possible to specify custom kernels.

## 2.2 Unsupervised Machine Learning Algorithm

It is also sometimes called unaided learning. In this algorithm learns not many components from the information [19]. At the point when new information is presented, it utilizes the recently scholarly components to perceive the class of the information. It is mostly utilized for bunching and size reduction [20]. Two main clustering algorithms are:

1. K-means clustering bunching or gathering is a sort of solo learning method that when starts, makes bunches naturally. The things which has comparable attributes are placed in a similar group [18]. This calculation is called k-implies on the grounds that it makes k particular groups. The following are the drawbacks of the algorithm-
  - (a) The learning set of rules calls for a priori specification of the number of cluster facilities.
  - (b) The use of exclusive assignment—If there are two incredibly overlapping information, then okay means will no longer be capable of solve that there are two clusters.
  - (c) The studying algorithm is not invariant to nonlinear ameliorations, i.e., with distinctive illustration of records, we get exclusive outcomes (information represented in form of Cartesian coordinates and polar coordinates will supply special effects).
  - (d) Euclidean distance measures can unequally weight underlying elements.
  - (e) The gaining knowledge of set of rules offers the neighborhood optima of the squared errors feature.
  - (f) Randomly choosing of the cluster center cannot lead us to the fruitful end result. Pl. Refer Fig.
  - (g) Applicable best when imply is described, i.e., fails for express information.
  - (h) Unable to address noisy data and outliers.
  - (i) Algorithm fails for nonlinear facts set. The mean of the qualities in a specific group is the focal point of that group [21].

**Pseudocode of k-means Clustering**

```

function Direct-k-means ()
    Initialize k prototypes (w1, ..., wk) such that wj=il, j ∈ {1, ..., k},
    l ∈ {1, ..., n}
    Each cluster Cj is associated with a prototype wj
    Repeat
        for each input vector il , where l ∈ {1, ..., n},
            do
                Assign il to the cluster Cj with nearest prototype wj.
                (i.e, |il - wj| ≤ |il - wj|, j ∈ {1, ..., k})

        for each cluster Cj, where , j ∈ {1, ..., k}, do
            Update the proto-
            type wj to be the centroid of all samples currently
            in Cj so that wj = ∑ il ∈ Cj il / |Cj|
    Compute the error function:
    
```

$$E = \sum_{j=1}^k \sum_{i_l \in C_j} |i_l - w_j|^2$$

Until E does not change significantly or cluster membership no longer changes

## 2. Principal Component Analysis(PCA)

In principal component analysis or PCA, the element of the information is decreased to make the calculations quicker and simpler. To see how PCA functions, we should take an illustration of 2D information. Principal element analysis of a records matrix extracts the dominant styles in the matrix in phrases of a complementary set of rating and loading plots. It is the obligation of the facts analyst to formulate the clinical issue handy in terms of PC projections, PLS regressions and so forth. Ask yourself, or the investigator, why the facts matrix became accrued, and for what purpose the experiments and measurements had been made. Specify before the analysis what varieties of patterns you would count on and what you will find interesting. At the point when the information is being plot in a chart, it will take up two tomahawks [17]. PCA is applied on the information, the information then, at that point, will be 1D.

### Pseudocode of PCA

```

R ← X
for ( $k = 0, \dots, K-1$ ) do
  {
     $\lambda = 0$ 
    T(k) ← R(k)
    for ( $j=0, \dots, J$ ) do
      {
        P(k) ← R(k) T(k)
        P(k) ← P(k) || P(k) ||-1
        T(k) ← R P(k)
         $\lambda' \leftarrow || \mathbf{T}^{(k)} ||$ 
        if (  $|\lambda' - \lambda| \leq \varepsilon$  ) then break
         $\lambda \leftarrow \lambda'$ 
      }
    R ← R - T(k) (P(k))T
  }
Return T, P, R

```

## 3 Conclusion

In this paper, we have discussed about different machine learning algorithm. Decision tree, SVM and Naive Bayes are supervised machine learning Algorithm. In machine learning, the input is result as well as data and the output is rules contrary to

traditional programming languages. This paper gives idea about supervised as well as unsupervised machine learning algorithm as well as their types. Decision tree is classifier and it can be used for classification as well as regression purposes both. But mostly, it is used for classification purposes. SVM is support vector machine and its main aim is separate the classes using hyper-plane. In unsupervised machine learning, machine only looks for the pattern as data has no labels. Training starts with huge data that form a feature vectors which using an algorithm is converted into predictive model that is tested with new set of data supervised machine learning is less complex, conducts offline analysis and gives comparatively more accurate result than unsupervised learning that is more complex and performs real-time analysis.

## References

1. Bonaccorso G (2017) Machine learning algorithms. Packt Publishing Ltd.
2. Goodfellow I, Bengio Y, Courville A (2016) Machine learning basics. *Deep Learn* 1(7):98–164
3. Dietterich TG (1997) Machine-learning research. *AI magazine* 18(4):97–97
4. El Naqa I, Murphy MJ (2015) What is machine learning? In: *Machine learning in radiation oncology*. Springer, pp 3–11
5. Kording KP, Tenenbaum JB (2001) Supervised and unsupervised learning with two sites of synaptic integration. *J Comput Neurosci* 11(3):207–215
6. Arunraj NS, Hable R, Fernandes M, Leidl K, Heigl M (2017) Comparison of supervised, semi-supervised and unsupervised learning methods in network intrusion detection system (nids) application. *Anwendungen und Konzepte der Wirtschaftsinformatik* 6
7. Chen L, Zhai Y, He Q, Wang W, Deng M (2020) Integrating deep supervised, self-supervised and unsupervised learning for single-cell RNA-seq clustering and annotation. *Genes* 11(7):792
8. Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A (2018) Machine learning for molecular and materials science. *Nature* 559(7715):547–555
9. Liu W, Chawla S, Cieslak DA, Chawla NV (2010) A robust decision tree algorithm for imbalanced data sets. In: *Proceedings of the 2010 SIAM international conference on data mining*. SIAM, pp 766–777
10. Jordan MI, Mitchell TM (2015) Machine learning: trends, perspectives, and prospects. *Science* 349(6245):255–260
11. Manwani N, Sastry PS (2011) Geometric decision tree. *IEEE Trans Syst Man Cybern Part B Cybern* 42(1):181–192
12. Ayodele TO (2010) Types of machine learning algorithms. *New Adv Mach Learn* 3:19–48
13. Wei J, Chu X, Sun X-Y, Kun Xu, Deng H-X, Chen J, Wei Z, Lei M (2019) Machine learning in materials science. *InfoMat* 1(3):338–358
14. Lee K, Caverlee J, Webb S (2010) Uncovering social spammers: social honeypots+ machine learning. In: *Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval*, pp 435–442
15. Witten IH, Frank E, Hall MA, Pal CJ (2005) Mining data: Practical machine learning tools and techniques. In: *Data Mining* 2, p 4
16. Tom M Mitchell. Does machine learning really work? *AI magazine*, 18(3):11–11, 1997.
17. Mohri M, Rostamizadeh A, Talwalkar A (2018) Foundations of machine learning. MIT press
18. Raschka S (2015) Python machine learning. Packt publishing Ltd.
19. Zhou Z-H (2016) Learnware: on the future of machine learning. *Front Comput Sci* 10(4):589–590



20. Hilar CS, Mastorocostas PA (2008) An application of supervised and unsupervised learning approaches to telecommunications fraud detection. *Knowl Based Syst* 21(7):721–726
21. Oral M, Oral EL, Aydın A (2012) Supervised versus unsupervised learning for construction crew productivity prediction. *Autom Constr* 22:271–276