Machine Learning Techniques and Testing

Ashish Verma[1], Kandy Arora[1], Aarush Sharma[1], Neelu Chaudhary[2]
[1]UG Students, [2]Professor
[1,2]Department of Computer Science Engineering, Manav Rachna University, Faridabad, Haryana, India
Email:*akv0116@gmail.com*
DOI: *http://doi.org/10.5281/zenodo.2615118*

Abstract
In this paper, we have discussed about various algorithms of the machine learning. These algorithms are used in various processes like image automated medical diagnostics, online advertising, robot incomotion etc.

Keywords: *Machine Learning, Decision Tree, K-means.*

INTRODUCTION

Machine Learning is a new era branch of Computer Science Engineering which is used in analytics and prediction. It is used to train computer in so that they can use their own intelligence to solve problem related to various tasks like recommendation, classification and identification, generally machine learning can be classified in two broad categories particularly Supervised and Un-Supervised learning.

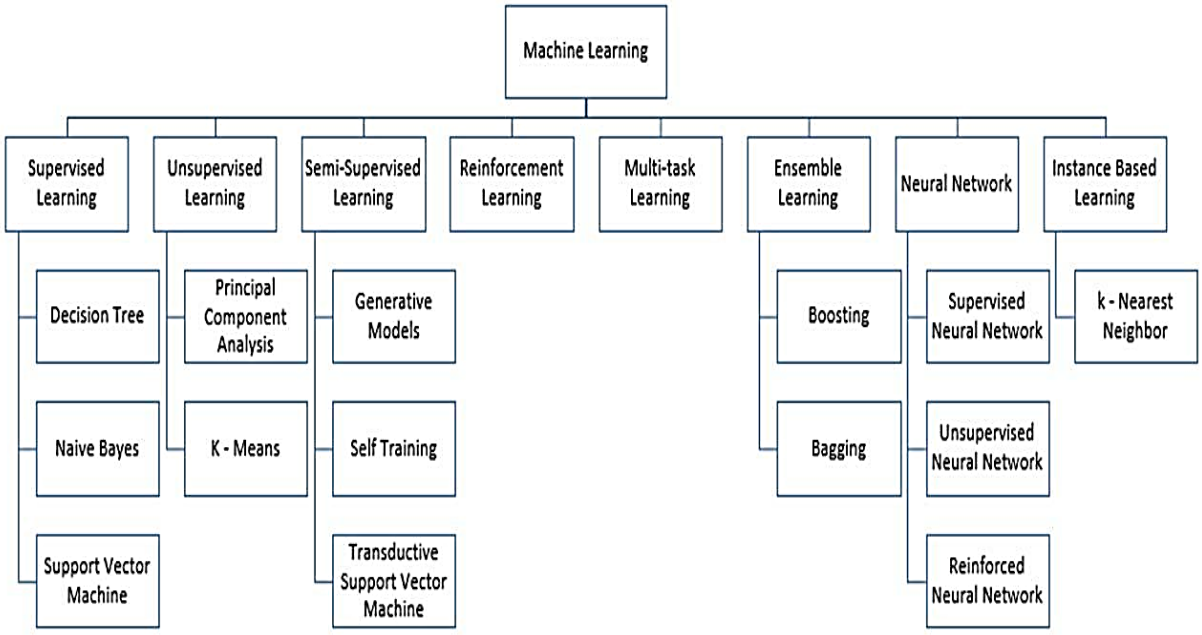


Figure 1: *Showing ML Categories.*

Background/Literature Review

Machine Learning algorithms are

Supervised Learning

The supervised machine learning algorithms are the algorithms in which given dataset is divided into train & test dataset. These algorithms need external assistance. In this, the train dataset has the output variable which is to be predicted or classified. All algorithms learn different types of patterns from the train dataset & then we apply them to test dataset for prediction. Its work flow is given in diagram:

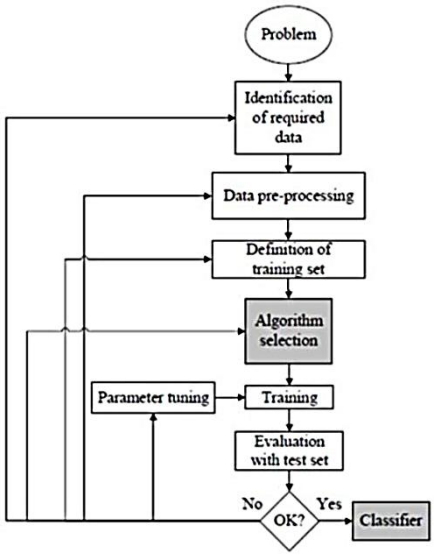


Figure 2: *Working of a Supervised Model*

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 $(v_1,\ldots,v_n)$ gains the best splitting metric.
6: **if** best splitting metric $\geq$ treshold **then**
7: Label the root node in T as $f(A)$
8: **for** each outcome v_i of f(A) **do**
9: $Subtree_i = TreeGrowing(\sigma_{f(A)=v_i}S, A, y)$.
10: Connect the root node of T to $Subtree_i$ with an edge that is labelled 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=Ø
7: Return T

Figure 3: *Work Flow.*

Decision Tree

Decision tree are those type of trees which sort the data on the basis of their values gor grouping attributes. It is mainly used for classification purpose. Each tree contains nodes & branches.

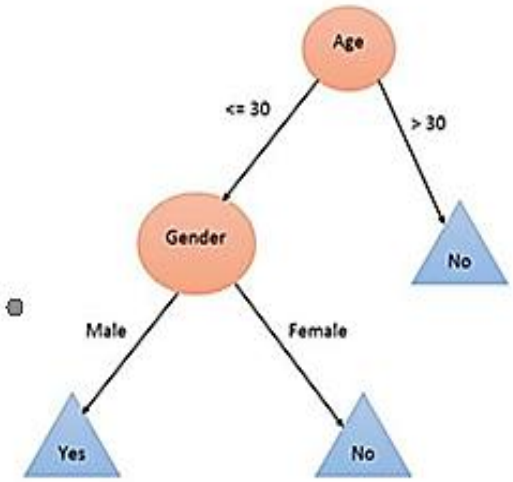


Figure 4: *Decision Tree.*

Naïve Bayes

Naïve Bayes is mainly used in text classification industry. It is mainly used for classification & clustering purpose. It is basically used to find the probability of event A when event B is already occurred.

P (A/B) = [P (A)*P (B/A)]/P (B)

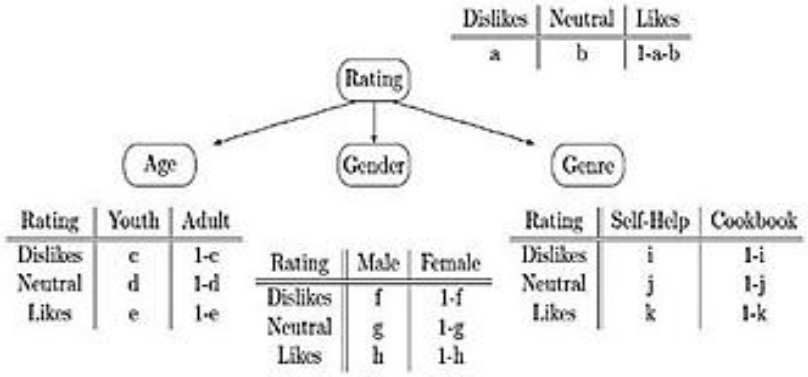


Figure 5: *Naïve Bayes.*

Support Vector Machine

SVM is mainly used for classification & it works on the principle of margin calculation. In this, margins are drawn in such a way so that the distance between the classes and the margins is maximum. It is done to minimize the classification error. An example of working & pseudo code of SVM.

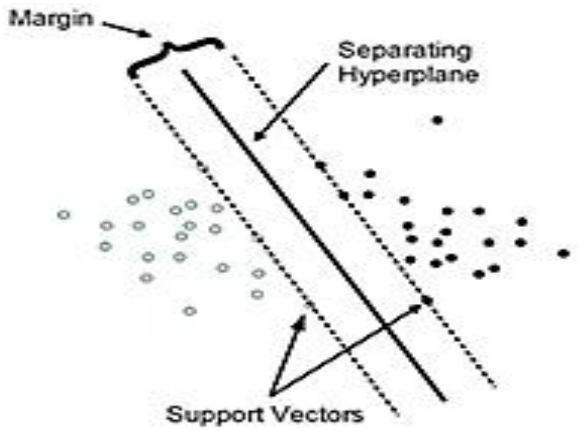


Figure 6: *Support Vectors.*

Unsupervised Learning

The unsupervised learning algorithms learns all the things from the data. When new data is introduced, it uses the previously learned features to recognize the class of data.

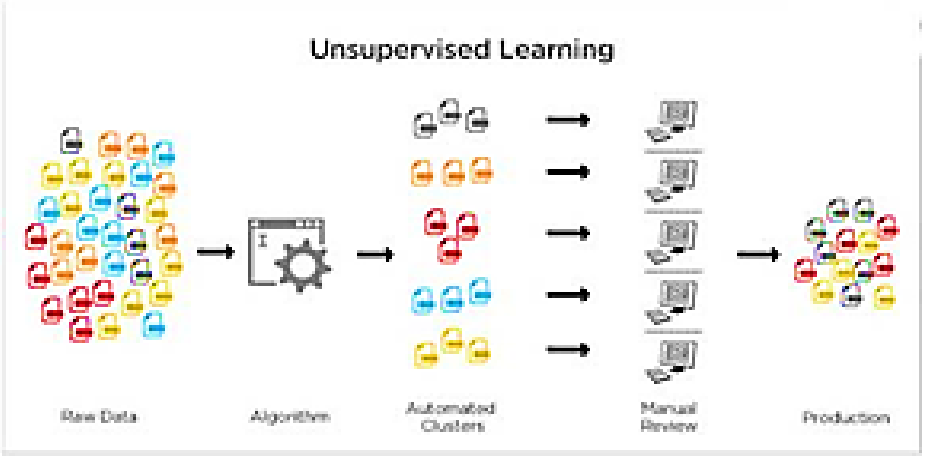


Figure 7: *Working of Un-Supervised Model.*

These algorithms used for dimensionally reduction & clustering techniques.

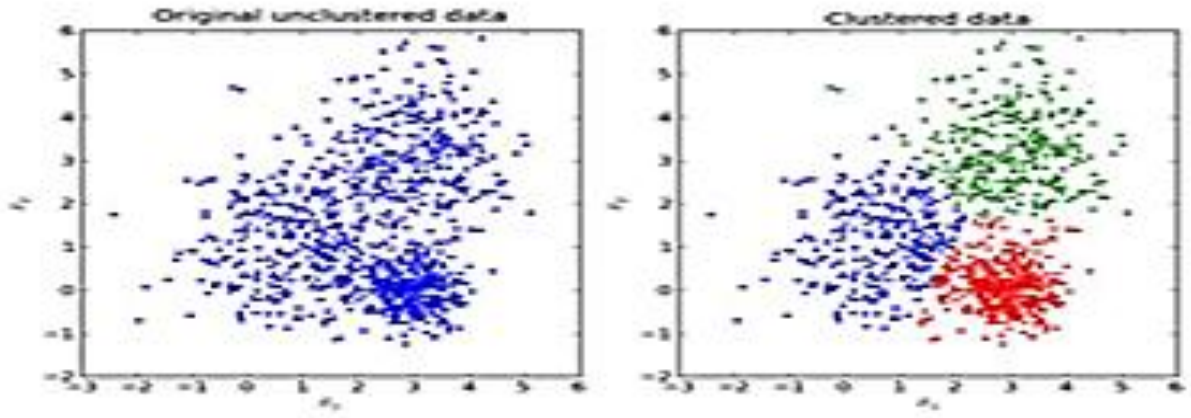
Its types are:

Principal Component Analysis

In this dimension of the data is reduced to make computations faster and easier. Lets take an example of 2D data to explain how principal component analysis works. When the data is placed in graph then it takes two axis. When we apply PCA to it, then the data will be in 1D.

K-means clustering

K-means clustering is type of unsupervised learning.This algorithm categorizes the items into k groups of similarity. We have to find euclidean distance as measurement to calculate the similarity.

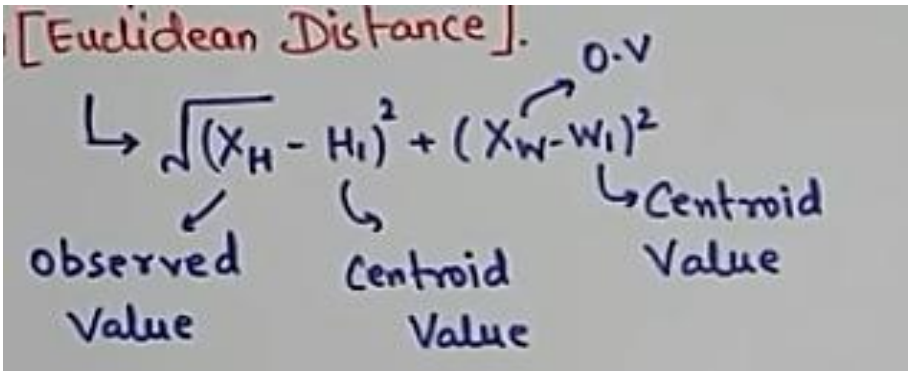


a) b)

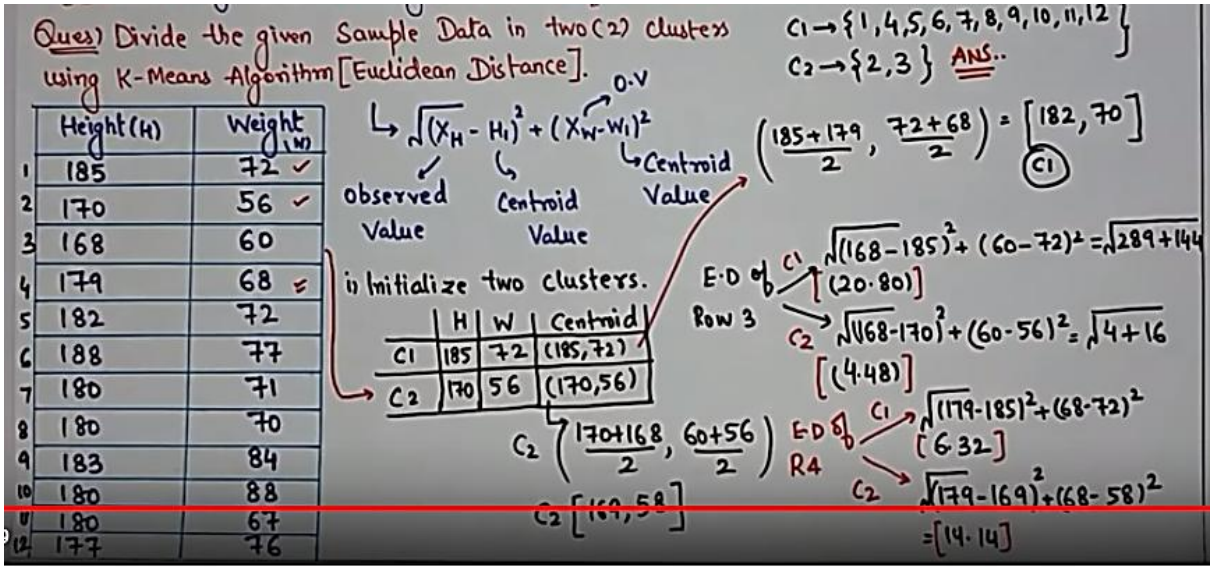
Figure 8 (a, b): *K-means Clustering.*

The algorithm works as follows:

1. First we initialize k points, called means, randomly.
2. We categorize each item to its closest mean and we update the mean's coordinates, which are the averages of the items categorized in that mean so far.
3. We repeat the process for a given number of iterations and at the end, we have our clusters.



a).



b)

Figure 9 (a, b): *Algorithm Work.*

Applications

1 Automated theorem proving
2 Affective computing
3 Classifying DNA sequence
4 Computer vision, including object recognition
5 Automated medical diagnosis
6 Internet fraud detection
7 Marketing
8 Online advertising
9 Sequence mining
10 Software Engineering

CONCLUSION

This paper is the survey of the various machine learning algorithms. Now a days, all of us are using machine learning knowingly or unknowingly. From buying a product from online shopping websites to updating photos and stories on social networking sites. This paper gives an introduction & overview about the algorithms of machine learning which are used in daily life.

REFERENCES

1. https://en.wikipedia.org/wiki/Machine_learning
2. https://www.geeksforgeeks.org/machine-learning/
3. https://ijcsit.com/docs/Volume%207/vol7issue3/ijcsit2016070332.pdf

Cite this article as: *Ashish Verma, Kandy Arora, Aarush Sharma, & Neelu Chaudhary. (2019). Machine Learning Techniques and Testing. Journal of Android and IOS Applications and Testing, 4(1), 19–23. http://doi.org/10.5281/zenodo.2615118*