

# A Circumstantial study on Supervised Machine Learning Algorithms

Rednam s s Jyothi <sup>[1]</sup>, Eswar Patnala <sup>[2]</sup>, V.Vikas <sup>[3]</sup>

Asst. Professor <sup>[1] & [2]</sup>, Student <sup>[3]</sup>

Dept. of IT, JNTUK-UCEV, Viziangaram

India

## ABSTRACT

Machine learning is one emerging research area in current research trend; machine learning can be combined with various other technologies for efficient outcome of results, in order to implement machine learning mechanism successfully. We should use various learning methods like, supervised, unsupervised and reinforcement. By using these algorithms we classify the data

**Keywords :**— supervised, unsupervised, naïve bayes, random forest, decision tree, support vector machine.

## I. BACKGROUND OF MACHINE LEARNING

Before Invention of Machine Learning, Artificial Intelligence plays a crucial role in our Computer Era. In artificial Intelligence, Researcher develops the human's perception of thinking artificially and converts them into algorithm. So, in Artificial Intelligence, Computer algorithms will plays vital role[1].

In Artificial Intelligence, we develop the algorithm for situations, if a particular situations occurs then the respective algorithm will be executed without human intervention. Suppose if there is a slight change in the situation, then the AI algorithm won't work, as it doesn't match with the condition. For efficient implementation of Artificial Intelligence, we need algorithms to act upon the situations. So, for this purpose, machine learning concept is introduced.

In machine learning, we design several algorithms , that can communicate with the humans, automated cars, stock reports, weather reports and finding terrorists suspects and various other aspects, So, every researcher should have idea regarding the machine learning[1]

Lets see how machine learning, how machine learning has been evolved from 1950 to the current generation[1]

- a) In 1950,Alan Turing has been developed a turing testing , mechanism whether the computer has real intelligence compared to human.

- b) In 1952,Arthur Samuel developed a computer program as one of the user in a game.So,this program will play against the human, and this program is composite of all winning strategies.
- c) In 1957, Nural network concept was Introduced and developed a first neural network by frank Rosenblatt,which analyses the human brain perception.
- d) In 1967,nearest neighbour,algorithm was Introduced which analyses shortest path from source to destination ,By using this concept Travelling salesmen Problem has been solved.
- e) In 1979,An automated vehicle was built which is named as Stanford Cart ,designed by students of Stanford University ,It can navigate through various Obstacles by taking pictures.
- f) Later in 1981,Explanation based learning has been Introduced ,where computer analyses the training data,before the actual data gets executed.
- g) During 1985,NetTalk app was invented by Terry sejnowski which pronounces as it is ,how an human pronounce.
- h) Later from 1990's to 2011, various applications has been design which beaten against human ,like in chess games etc.
- i) Various different concepts like Deep Learning also Introduced to Indentify objects and text in images and videos.
- j) Later 2011 to 2015 , google,amazon and facebook launches its own machine learning platforms related to the Google's X-labs,facbooks,Deepface etc

k) In 2016,AlphaGo algorithm developed by google to beat a human in Chinese chess board game[1]  
This is how the AI and ML technologies has emerged into the computer technology[1]

## II. IMPORTANCE OF DATA IN MACHINE LEARNING

Data, it is one of the most important artifact of any technology. data is a unprocessed fact, which can be any picture ,text,sound, video, which was not analyzed or Interpreted.For technologies like ,Bigdata,ML & AI data is very Imortant for their analyses Purpose.Many big entriprise companies spent a huge amount of money for the collection.This is why ,facebook owns the whatsapp to access the users Informations.and the data can be categorized into two ways[2]

- 1) **Information:** The data which is interpreted and analysed and in a meaningfull format is known as information[2]
- 2) **Knowledge :** The Information which is drawn from various inferences,learniongns and experiences is called knowledge and it will be an organized format. [2]

### Types of data in machine learning[2]:

- 1) Training data: The data which is used to train our model is known as training data. The main aim of training data is to make the design more familiar of what, which it uses frequently.
- 2) Validation data: The data which is used to evaluate the model frequently is known as validation data. This type data is mainly used , whether the trained data set has been working correctly or not.
- 3) Testing data: After model has been trained using trained data ,the mode will be tested wit set of inputs and outputs is known as testing data and then we evaluate ,whether the model has been correctly trained or not using training dataset [2]

### PROPERTIES OF DATA

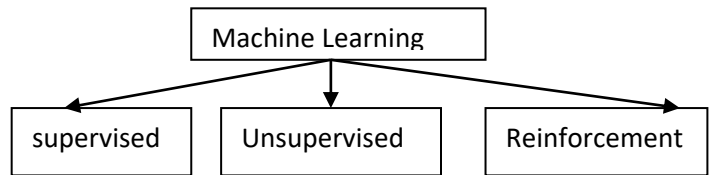
The term data can be effectively described by following properties[2]

- 1.Volume
- 2.Veracity
- 3.velocity
- 4.Value

## III. MACHINE LEARNING CLASSIFICATION

The machine learning algorithms has been classified into 3 categories[3]

- 1) Supervised
- 2) Unsupervised
- 3) Reinforcement



- 1) **Supervised Learning:** By the name itself , we can say that learning process is done under some supervision.we train the machine with some trained data set

For example, suppose we take the basket full of fruits of all kinds and train the machine to identify all kinds of fruits in the basket by analyzing shape and colour. after successful training ,we give Input as one of the fruit ,which is already present in basket .So, now the machine depicts the shape and colour of the given Input fruit and then displays the result as fruit name[3].

Supervised Learning is classified into 2 types they are

- a) Classification
- b) Regression

- 2) **Unsupervised Learning:** It is one of the Machine learning technique,Here we do not train the machine with training data set, Instead the machine has to identify and classify the data by their similarities and various patterns[3]

For example, we have given an image as an input ,which consists of the cricket player.As it is unsupervised learning , the machine was unaware of this information present in the Image.So,Now identify the players and classify them based on their dressing and their characteristics

Unsupervised learning has been classified into 2 types

- 1) Clustering
- 2) Association

### III. SUPERVISED MACHINE LEARNING ALGORITHMS

#### Naïve bayes classifier:

It is one of the machine learning algorithm which comes under supervised machine learning under classification. While dealing with naïve bayes classifier the core part in the algorithm is posterior probability which is also called Bayes theorem[9]

$$P(E|F) = [P(F|E) P(E)] / P(F)$$

The above formula explains how often ‘E’ happen given that ‘F’ happens which was represented by P(E|F) and how often ‘F’ happens given that ‘E’ happen[9].

Where P(E|F) Is Probability of E Given F

P(F|E) Is Probability of F Given E

P(E) Is Probability of E

P(F) Is Probability of F

In this formula F is acts as a evidence and E is treated as hypothesis[9]

#### Example:

TABLE I[9]

S. No	Current weather	Hot/ mild/ Cold	Humidity rate	Speed of the wind	Play Game
1	Rainy	Hot	High	False	No
2	Rainy	Hot	High	True	No
3	Overcast	Hot	High	False	Yes
4	Sunny	Mild	High	False	Yes
5	Sunny	Cool	Normal	False	Yes
6	Sunny	Cool	Normal	True	No
7	Overcast	Cool	Normal	True	Yes
8	Rainy	Mild	High	False	No
9	Rainy	Cool	Normal	False	Yes

10	Sunny	Mild	Normal	False	Yes
----	-------	------	--------	-------	-----

Example[9]:

From the above table we can notice detailed weather report for days. Let us assume that on 11<sup>th</sup> day, the weather conditions are like, sunny, Mild, High, and true for respective fields. Now we predict the weather based on the above measurements and using given data, we take the decision to play game or not. [4][5][6]

S.No	Current weather	Hot/ mild/ Cold	Humidity rate	Speed of the wind	Play Game
11	Sunny	Mild	High	True	No

TABLE 2[9]

#### Support vector machine:

SVvMs are used to identify decision boundaries this can be done by the concept of planes. Decision planes are used to identify and separate the set of objects which belongs to different class of memberships[4][5][6]

For example suppose we have set of circle objects in combination of two colors they are yellow and blue. By using decision plane concept we divide them into two groups by placing all yellow circles on left side and all the blue circles on right by drawing a single line between then. Now if any new object of different colour (for example brown) other than yellow and blue is came into the boundary then if it falls left side it is called as yellow or if it falls on right it is called as blue though it is brown color[4][5][6]

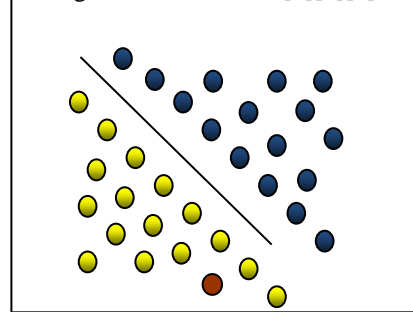


Fig 1 Decision boundaries sample image[4][5][6]

In order to place the object in correct boundary we introduce a concept called hyper plane which will be a curve shape. So in order to distinguish between objects of different classes hyper plane classifies plays major role[4][5][6].

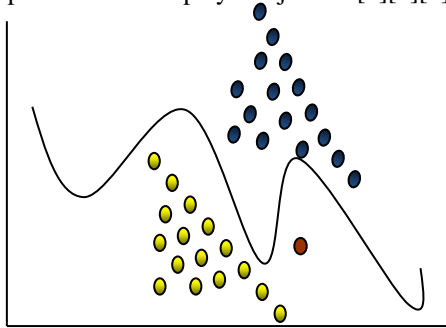


Fig 2: SVM Hyper Plane[4][5][6]

Hence the re arrangement of object using mathematical functions are known as kernels and the entire process is known as mapping or transformation

**Mapping Diagram**

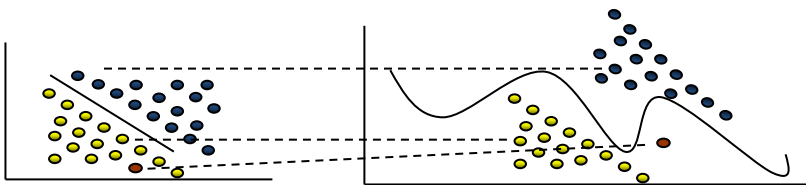


Fig 3: Mapping sample image

**Technical notes:**

So, support vector machine is one of the important classifier by using hyper plane concept in multi dimensional space that consists of labels of different classes. By using SVM we can deal with both continuous and categorical variables which belongs for regression and classification. While using categorical variables we create a dummy variable which can be either 0 or 1[4][5][6].

**Support Vector machine for classification:**

Using SVM in classification is one of the supervised learning mechanism. Here classification can be done by using training and testing data. After the model has been successfully trained, then whenever an input arrive based on the feature extraction we correctly place the object in certain group[4][5][6]

**Support Vector machine for regression:**

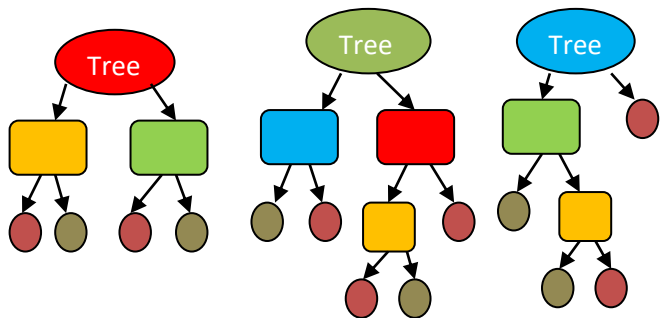
In the regression technique by using SVM we will assign a particular object to the respective group by introducing an alternative loss function by using this function we can assign them to a particular group[4][5][6].

**Random Forest:**

Random forest algorithm is one of the most popular classification algorithms under supervised machine learning. Technically it is called as random forest classifier which can used for both classification and regression methods.

**What is random forest algorithm:**

By the name itself we can tell that algorithm creates a forest with more number of trees. If we have more trees then the forest look like more robust. In order to produce high accuracy of results we are creating more no of trees. Now each tree is constructed by using an decision tree classifier so, by using this algorithm we build trees for each data set respectively[7][8].



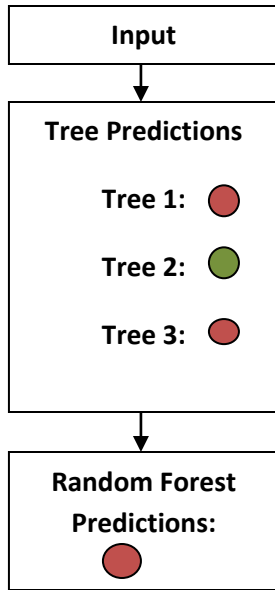


Fig 4: Random Forest predictions

### Random forest real life example:

Suppose we have person name Prince some how he got two weeks leave and now he want to go for the vacation for the best place he may like it. So, he asked his best friend to suggest a best holiday spot that prince may like it. Now his best friend started asking question about the last trips of the prince. So that based on the past trip information his best friend may suggest him a best place [7][8].

But his best friend want to suggest best place to prince so he decide to ask few more friends to recommend the places he may like it. So his best friend asks some random question to all the other friends and the place which got high no. of votes will be suggested to prince.

So in this above example the conversation between prince and his best friend is used to develop a decision tree by using a decision tree classifier and the conversation between his best friend and his few more friends asking random questions done by using random forest algorithm [7][8].

### APPLICATION OF RANDOM FOREST [7][8]

#### Banks:

In banking sector by using random forest we can identify the customer who is loyal and who are fraud.

#### Medicine:

In medical field, By identifying the random combinations of the various components of medicine to particular disease.

#### Stock Market :

In this area, Random Forest is used analyse the behaviour of various products and estimates the loss or profits

#### E-Commerce:

In this area random forest is used to suggest the product to the customers based on their previous transaction and searching history

#### Decision tree:

It is one of the simplest and most important algorithm in machine learning. This algorithm can be solved for both regression and classification tasks. By using this algorithm we create a decision tree which is also called as training model. In order to predict a particular class with the help of decision trees which are inferred from the training data. Here problems are solved by using tree representation. An attribute represents the internal node and a class label represents a leaf node [7][8].

#### Steps to construct decision tree [7][8]:

Step1: Identify the best attribute and make that attribute as the root of the tree

Step2: Split the training set into subsets. subsets should be made in such a way that each subset contains data with some value for an attribute.

Step3: Till you find leaf node in all branches, we should repeat step 1, 2

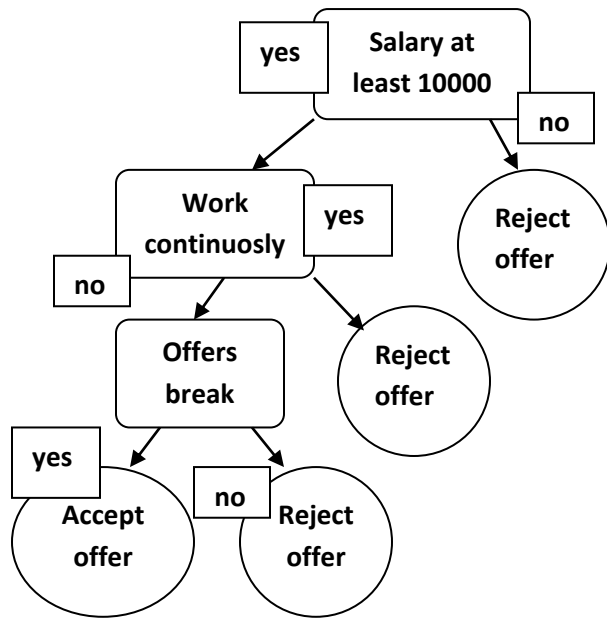


Fig 5: Steps for the construction of DT[7][8].

In decision trees, in order to find the leaf node which is appropriate for our situation is to be start from the root node. First, we check the root node attribute and decided

the answer and then go according to the next internal node. We process the each and every internal node based on our choices then we will reach the ultimate leaf node which is our final decision

The above diagram explains the clearly the concept of decision tree.

**Linear regression:**

This is one of the supervised machine learning algorithms.in which we predict the value of the dependent variable based on the independent variable[9].

The linear regression line of the form  $y_1=m_1x_1+c_1$

Where  $y_1$  is dependent variable or scalar variable  
 $x_1$  is independent variable or predictor variable or explanatory variable.

$m_1$  is called slope of the regression line.

$c_1$  is  $y_1$  intercept of regression line .it is constant it shows the value  $y_1$  when the value of  $x_1=0$ .

Example:

We predict the value of the dependent variable based on more than one independent variables is called multiple regression[9].

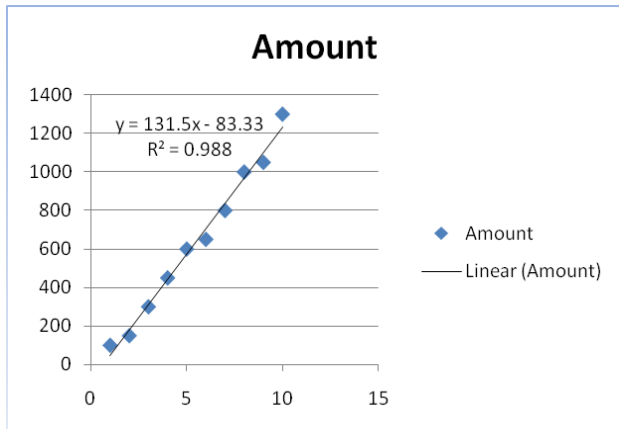
**Applications of the regression :**

- Profit estimation of the company based on purchasing the items.
- Weather forecasting
- In business to find the profit based on the sales, advertisement of the product
- In business to find the sale of product based on the advertisement of the product[9].

Example: the following diagram illustrates as number of items increases, amount also will be increased simultaneously and graph represents the graph of amount and items relation[9].

TABLE 2[9]

No of items order	Amount
1	100
2	150
3	300
4	450
5	600
6	650
7	800
8	1000
9	1050
10	1300



#### IV. CONCLUSION

In this paper we mainly discuss about evolution machine learning, how Artificial Intelligence is evolved into machine learning and how it has been utilized for current research solutions. and discussed about various supervised algorithms. How these algorithms will work in various applications are also represented and provided with some of the examples

#### REFERENCES

- [1] <https://www.forbes.com/sites/bernardmarr/2016/02/19/a-short-history-of-machine-learning-every-manager-should-read/#8a0628215e78>
- [2] <https://www.geeksforgeeks.org/ml-introduction-data-machine-learning/>
- [3] <https://www.geeksforgeeks.org/introduction-machine-learning/>
- [4] <http://www.statsoft.com/textbook/support-vector-machines>.
- [5] <https://monkeylearn.com/blog/introduction-to-support-vector-machines-svm/>
- [6] [http://saedsayad.com/support\\_vector\\_machine.htm](http://saedsayad.com/support_vector_machine.htm)
- [7] <https://dataaspirant.com/2017/05/22/random-forest-algorithm-machine-learning/>
- [8] <https://dataaspirant.com/2017/05/22/random-forest-algorithm-machine-learning/>
- [9] <https://www.geeksforgeeks.org/naive-bayes-classifiers/>