JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

MACHINE LEARNING ALGORITHM ANALYSIS

[1]APOORV DWIVEDI, [2]SHIVAM AWASTHI
[1]STUDENT, [2]ASSISTANT PROFESSOR
[1]Information Technology Department,
[1]Bachelor of Technology Shri Ramswaroop Memorial College of Engineering and Management, Lucknow, India

Abstract: Artificial intelligence was proposed to improve system performance when it was only a forward-looking concept. If machines can have the same learning ability, same as human beings and the computing power of computers increases gradually, this concept has been placed high hopes among society. Until about 2010, with the explosion of data set and the improvement of computer performance, ML has become a leader in breaking through the bottleneck of artificial intelligence. Research on machine learning in education and teaching, health, weather has attracted much attention. From this research status, we can see this in the current period of the successful development of machine learning, many applications are still not perfect and under development, ordinary education and health issue, weather evaluation is difficult to meet people's requirements, so to gradually improve its efficiency and effectiveness is a significant goal with research significance and practical approach. In this context, according to the theory of machine learning, the effectiveness of different conventional prediction and evaluation methods is analyzed. In this paper, machine learning theory is used to study college students' admission prediction and credit evaluation, medical as well as weather evaluation and comprehensive ability evaluation. The effectiveness of machine theory in teaching, health is analyzed. It is found that machine learning has great advantages in education, teaching evaluation, health issues and weather prediction. It builds models in complex computing environment and the effectiveness of prediction and evaluation is significant.

Index Terms **– Machine Learning Algorithm Analysis**

1. INTRODUCTION

This project is about the ML algorithm accuracies prediction, to analyse the efficiency of algorithm on different scenarios,

So in order to apply a suitable ML algorithm to any problem or any new area, it is important to understand the applicability of various types of ML techniques.
In today's era, we are constantly surrounded by data. Everything around us is connected by data source (i.e., smartphones, social media, personalized advertising, speech and facial recognition, image recognition self-driving cars, genome sequencing, energy-efficient buildings, computer interactive games, language translation
Artificial intelligence is mainly a machine learning only and have grown rapidly in recent years in the context of data analysis and computing that typically allows the applications to function in an intelligent and attractive manner. ML usually provides systems with the ability to learn and enhance from experience automatically without being specifically programmed and is generally referred to as the most popular advanced and latest technologies in the fourth industrial revolution is typically the ongoing automation of conventional manufacturing and industrial practices, including exploratory data processing, using new smart technologies such as machine learning automation. Thus, to intelligently analyse these data and to develop the corresponding real-world applications, machine learning algorithms is the key to do in a better way. The learning algorithms can be categorized mainly into four major types, such as supervised, unsupervised, semi-supervised, and reinforcement learning.

2. RESEARCH METHODOLOGY

The methodology section outline the plan and method that how the study is conducted. This includes Universe of the study, sample of the study, the details are as follows;

The collection of Data Set and its Pre-processing.

Finalisation of algorithms to applied in different Scenarios.
Splitting and Training of Data Set.

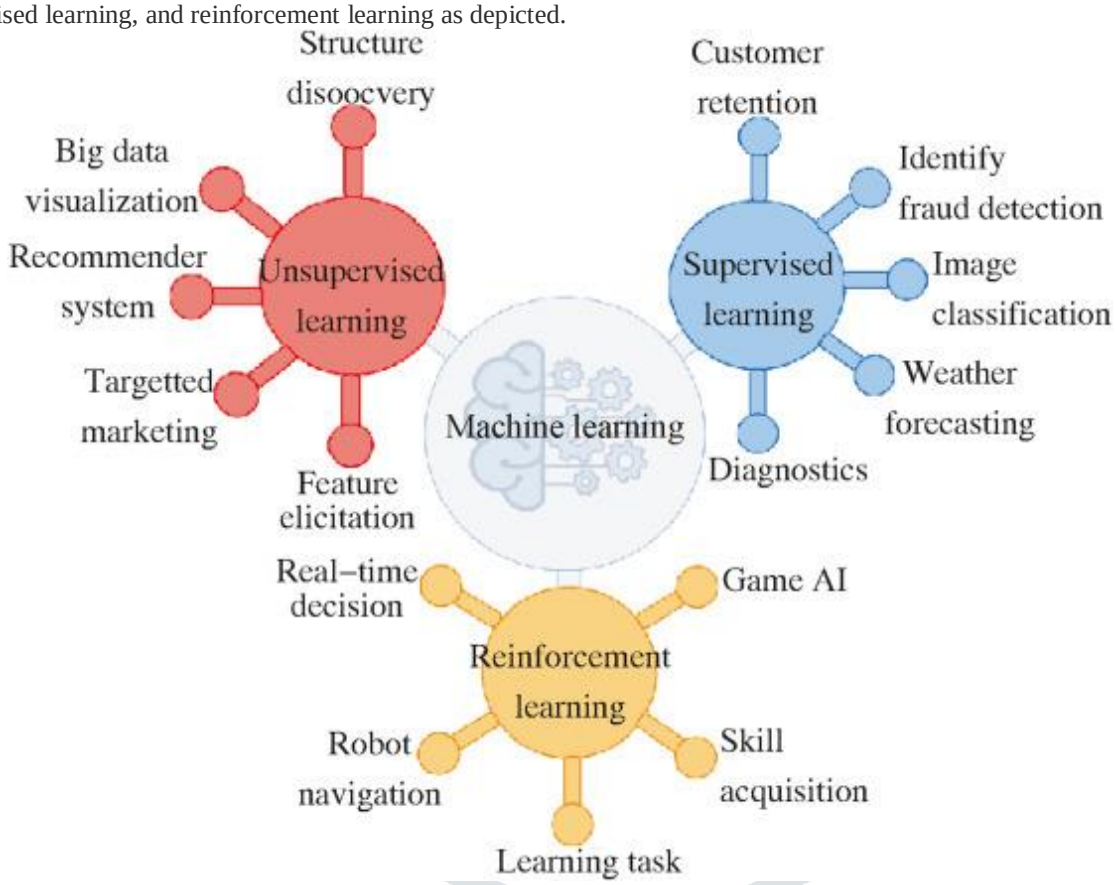
Applying algorithms in Data Set and analysing its Accuracy.

The Result are graphically displayed for comparison of Accuracies

3. LITERATURE REVIEW

Types of machine learning techniques commonly used

Machine Learning involves the development and deployment of algorithms that, rather than being programmed to assign certain outputs (namely results) in response to specific inputs from the environment, analyse the data and its properties, and determine the action by using different statistical tools. Usually, Machine Learning algorithms can dynamic improve or "learn" as more data is introduced. ML algorithms can be majorly classified into four categories: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning as depicted.



Supervised learning

Supervised learning relies on Machine Learning tasks to learn a function that maps an input to an output based on sample input to output pairs. Hence, this learning process is based on comparing the calculated output and predicted output/results, that is learning refers to computing the error and adjusting the error for achieving the expected output. Examples of such algorithms include Naïve Bayes classification, linear and logistic regression, Support Vector Machines (SVMs) and many more. Examples of applied supervised learning are automatic answering of incoming messages, face recognition which is useful as security measures at an ATM, surveillance areas, closed circuit cameras, criminal justice system, and image tagging in social networking sites like Facebook, Instagram.

Unsupervised learning

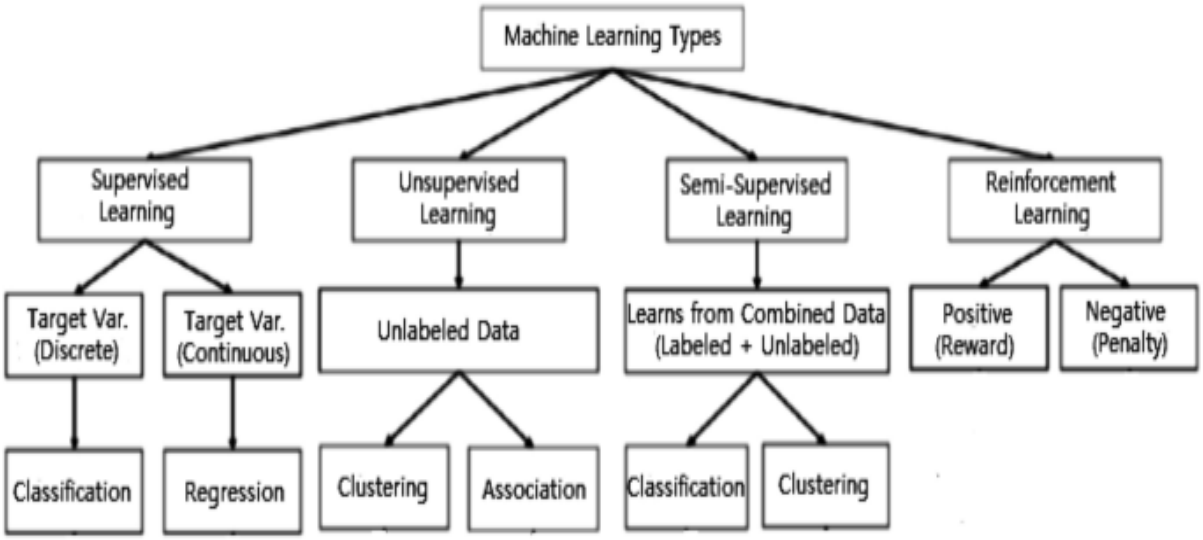
Unsupervised learning analyses the unlabelled datasets without human interference. In unsupervised learning, the algorithm optimally separates the samples into different classes/methods on the basis of the features of the training data alone, without corresponding labels. The unsupervised algorithms are *k*-means clustering, principal component analysis, and auto encoders etc. The most common example of unsupervised learning is the automatic identification of friends for a user in social media channels like Facebook, Google, Instagram or the discerning of the maximum number of mails sent to a particular person and categorized into different collective groups. Furthermore, computational biology (also known as bioinformatics) is developing unsupervised algorithms from biological data to establish relations among various biological systems for better health analysis, collecting lots of useful data about gene sequences, DNA sequences, and a gene expression, thus providing a much better understanding about human genome.

Semi-supervised learning

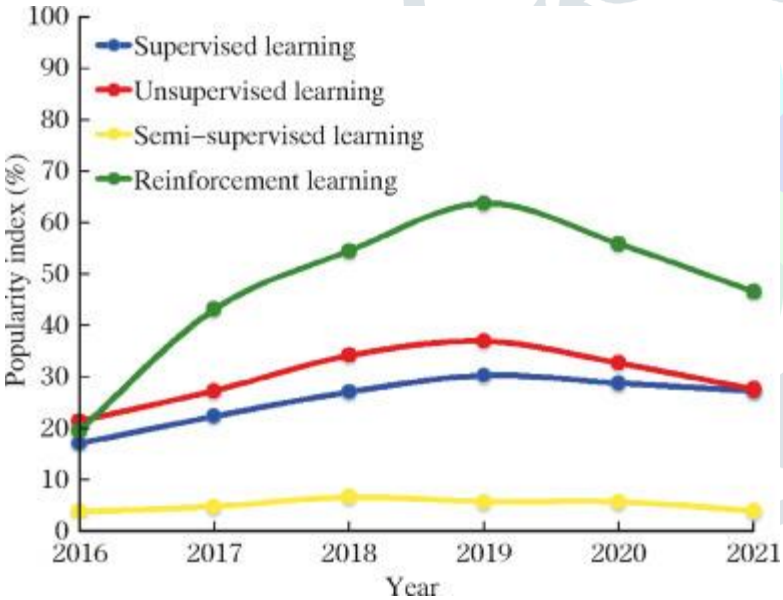
Semi-supervised learning can be defined as a hybridization of the above-mentioned supervised and unsupervised learning methods, as it works on both labelled and unlabeled data. The ultimate goal of a semi-supervised learning model is to provide a better outcome for prediction than that produced using the labelled data alone or from unlabeled data alone, to the model. Such a method is widely used in machine translation, fraud detection, labelling data, image classification, and text classification etc.

Reinforcement learning

Reinforcement learning lies on a family of algorithms that typically operate sequentially to automatically evaluate the optimal behavior in a particular environment to improve its efficiency and accuracy, i.e. an environment-driven approach. At each step, a reinforced algorithm, also referred as "agent", that acts and predicts the features at a future step on the basis of past and present features, and a reward or penalty is assigned on the basis of prediction. Therefore, it is a powerful tool for training AI models that could optimize the operational efficiency of sophisticated systems, such as robotics, autonomous driving tasks, manufacturing, and supply chains etc. The reinforced algorithms are with function of approximation, gradient temporal difference learning, and least-squares method. An outstanding example of an application of reinforcement learning is the algorithm that can automatically generate the appropriate tension and optimal direction for a given cutting trajectory of either a laparoscopic surgeon or an automated cutting instrument.



Comparison of all four types of Machine Learning



Analysis of commonly used Algorithms for Machine Learning

Algorithms Used

1. **Linear Regression**
2. **Decision Tree**
3. **Random Forest Repression**

REASON FOR SELECTING THESE ALGORITHMS

On a previous theoretical research analysis we found that these three algorithms have a greater impact over numerical data and we have to work on Numerical Data Analysis so rather to work on all algorithms one by one it is time consuming also, we decided to do analysis of these three Algorithms and analyze its impact.

Linear Regression

Linear regression analysys is used for prediction, the value of a variable is based on the value of another variable. The variable value you want to predict is called the Dependent variable. The variable you are using for predicting the other variable's value is called the independent variable.

This form of analysis signifies that the coefficients of the linear equation, involving one or more independent variables that best predicts the value of the dependent variable. Linear regression fits a straight line or surface that minimizes the discrepancies between

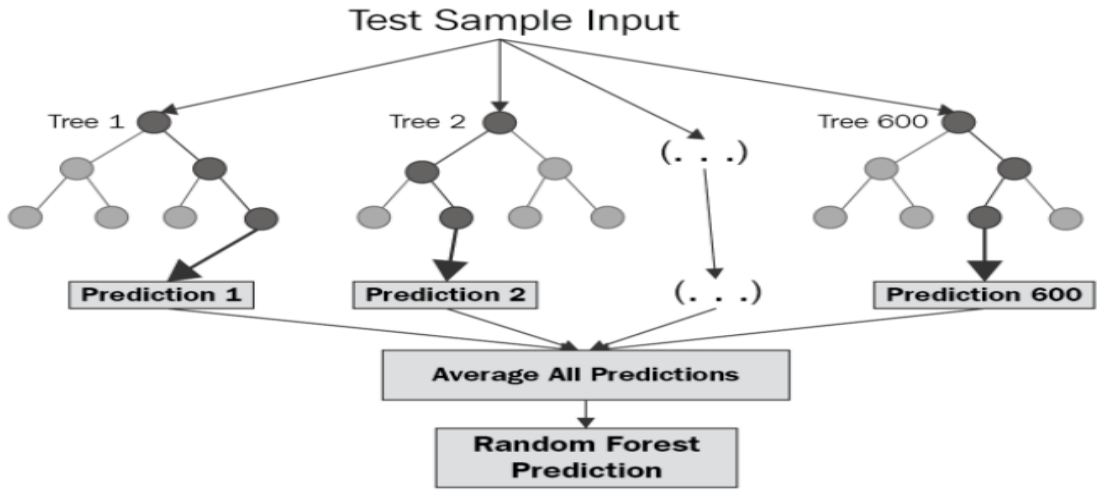
predicted and actual output/resulting values. There are simple linear regression calculators that use a 'least squares' method to discover the best-fit line for a set of paired data. You then estimate the value of X i.e. dependent variable from Y i.e. independent variable.

Decision Tree

Decision tree builds regression or classification models like in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets of itself, while at the same time an associated decision tree is incrementally developed. The final result is a tree with **decision nodes** and **tail nodes**. A decision node has two or more branches, each representing the values for the attribute tested. Leaf node or tail node represents a decision on the numerical target. The top-most decision node in a tree which corresponds to the best predictor called **Root node**. Decision trees can handle both categorical and numerical form of data.

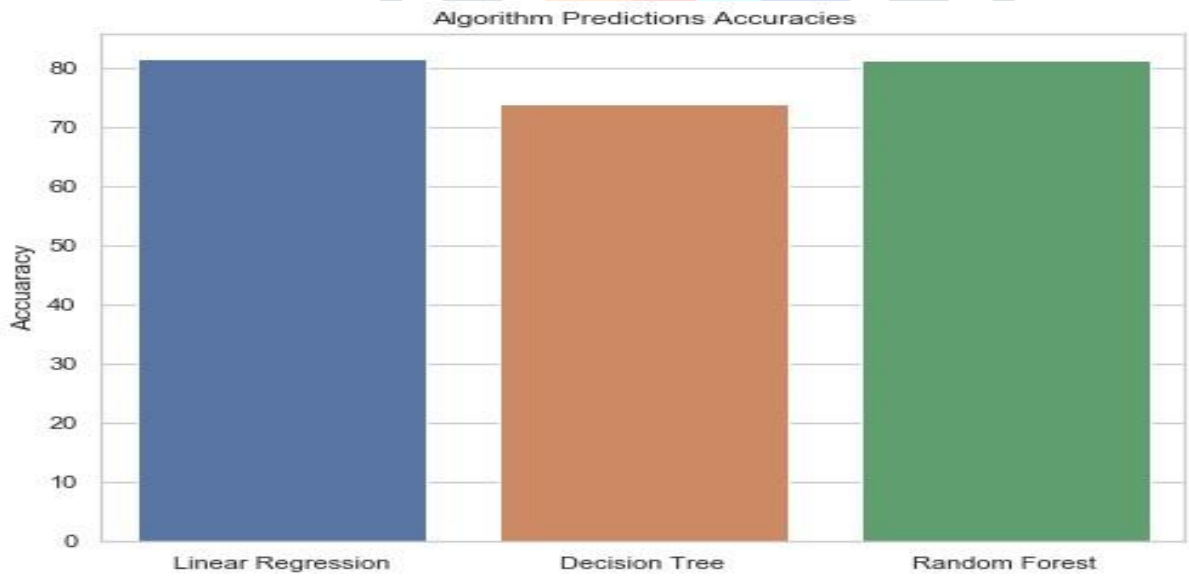
Random Forest Repressor

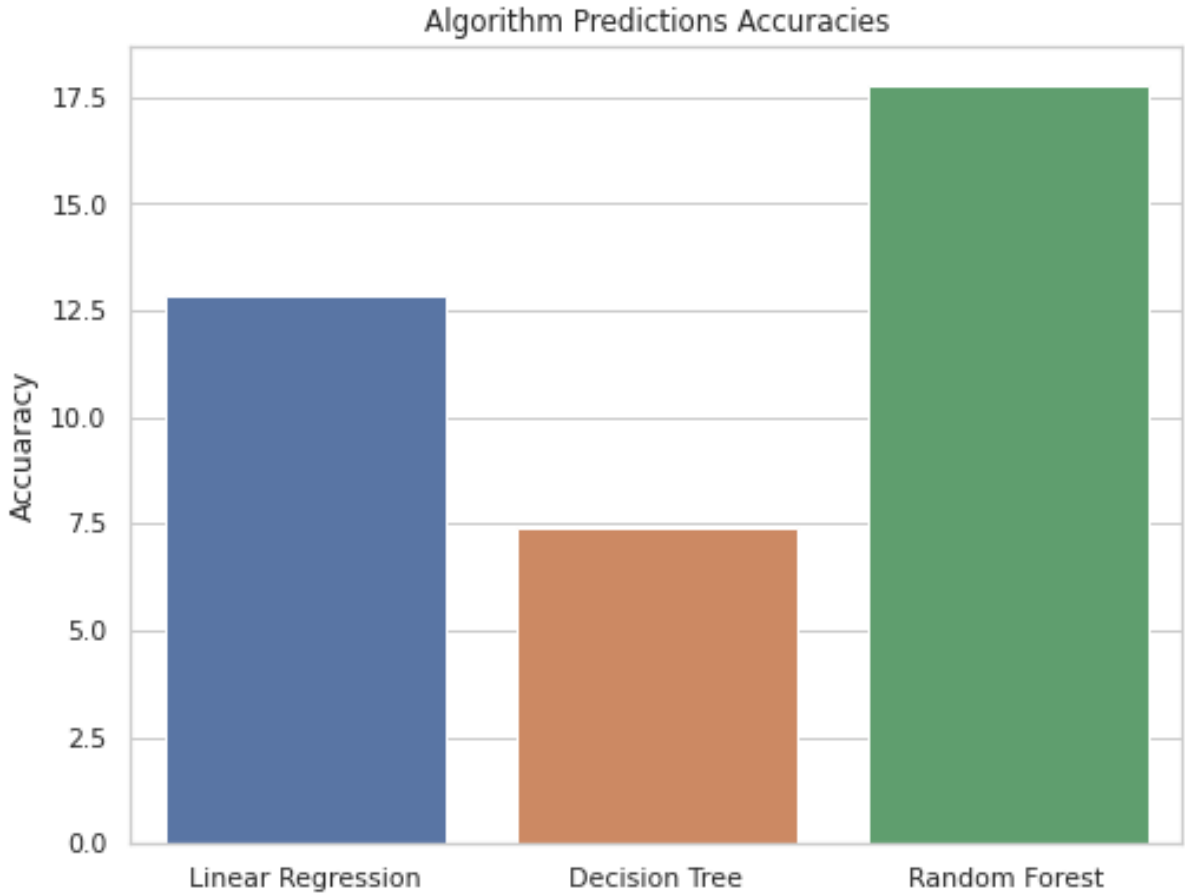
Random Forest Regression is a supervised Machine learning algorithm that uses **ensemble learning** method for regression. The learning method is a technique that combines predictions from multiple machine learning algorithms to make a more accurate and efficient prediction than a single model.



4. RESULTS AND DISCUSSION

Comparison of these three Algorithms on basis of accuracies and Complexities





5. FUTURE SCOPE

The scope of this research is not only limited to the activities of comparison only. which includes will improving long term care, enhance access and care in the Estate and surrounding communities and ensuring best major columns, the use of a computer based management system or a machine learning technology is for improving the efficiency of prediction is needed and it is an essential part of any modern continuously evolving society.

By these analysis it will be helpful in future to develop a model/software on basis of these analysis in field of--;

1 Education

2 Medical

3 Weather

6. CONCLUSION

We would like to conclude by our analysis that we performed on these three algorithms

1. Linear regression
2. Decision tree
3. Random forest

Admission prediction – Random forest and linear regression equally predicted the results.
Heart disease prediction – Random forest works best among all three.
Weather prediction -- Random forest works best among all three.
Finally among all three algorithms Random forest works well in all three condition.
It's not a final conclusion of prediction as we will increase the dataset and provide Varity of Data set, the prediction will become more precise.

7. REFERENCES

1) Smith, Kevin D.; Jewett, Jim J.; Montanaro, Skip; Baxter, Anthony (2 September 2004). "PEP 318 – Decorators for Functions and Methods".
2 Schemenauer, Neil; Peters, Tim; Hetland, Magnus Lie (18 May 2001). "PEP 255 – Simple Generators". Python Enhancement Proposals. Python Software Foundation. Retrieved 9 February 2012.
3) Lundh, Fredrik. "Call By Object". Effbot.org. Retrieved 21 November 2017
4) Simionato, Michele. "The Python 2.3 Method Resolution Order". Python Software Foundation.

5) Learn Python the Hard Way by Zed A. Shaw (3rd Edition)
6) Python Programming by John Zelle
7) Learning with Python by Allen Downey, Jeff Elkner, and Chris Meyers
8) A Byte of Python by C.H. Swaroop
9) The Elements of statistical learning by Robert
10) Machine Learning by Ton M Mitchel
11) https://www.w3school.com
12) https://www.geeksforgeeks.com.