An Insight to Machine Learning Algorithms and a Case Study

Vijayalakshmi S Hallikeri

Department of Artificial Intelligence and Machine Learning,
BIET.,
Davangere, India

Abstract—Today, user-generated data is everywhere, from social media posts to customer reviews. Machine also generates the data at real time. Data processing is required for businesses to enhance their business plans and get a competitive edge. In machine learning, input data is given to the machine, which is then trained to find patterns in the data, perform a specified task, and produce precise results. In this paper, concise overview of various Machine Learning algorithms and applications is given. A definition of machine learning is presented first, followed by a discussion of the various machine learning paradigms and their applications. A case study involving prediction is presented as final example.

Keywords—Machine Learning, Algorithms, Patterns, Paradigms, Case study, Prediction

I. INTRODUCTION

Machine learning (ML) and Deep learning (DL) are both components of Artificial Intelligence (AI) which is shown in Fig.1. Machine learning was first described as "the field of study that enables computers to learn without explicit programming" by AI pioneer Arthur Samuel in the 1950s. [1]
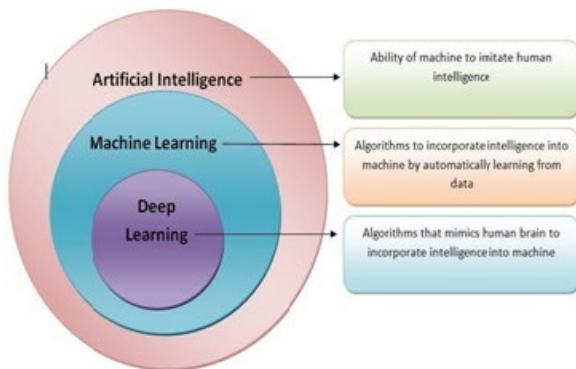

Fig. 1. Machine learning, AI subfield [5]

To develop machine learning algorithms, AI needs a basis of specialized software and hardware. The primary objectives of the learning component of AI programming are the gathering of data and the creation of rules for its transformation into useable knowledge.

Machine learning algorithms are computer programs that evolve and adapt based on the data they analyze to produce predetermined results. Artificial neural networks are used in a subset of machine learning called deep learning, to simulate how the human brain learns. [2]

Artificial intelligence (AI) is a tool that "intelligent" computers use to do independent tasks and imitate human reasoning. Machine learning refers to the method a computer system uses to become intelligent. [3]

One might or might not be aware of technologies, many applications adopt Machine learning, like identification of image, voice search, automated translation, self-driving cars, etc.

II. DIFFERENT APPROACHES IN MACHINE LEARNING

Four major categories shown in Fig.2, corresponding to different learning paradigms are used to categorize machine learning methodologies. [4]
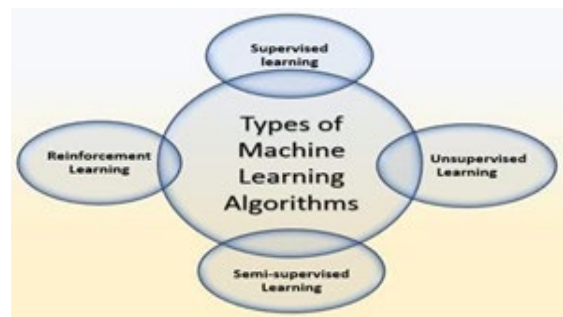

Fig. 2. Main types Machine Learning

A. Supervised Learning

Provided it with examples of inputs and desired outcomes, the objective is to teach the computer a general rule that maps inputs to outputs. It needs a training set of tagged data or data with a known outcome value (such as whether a tumor is malignant or not, or the price of a house). Problems with classification and regression are resolved using supervised learning techniques.

Applications

    Regression
- Risk assessment
- Score prediction

    Classification

- Diagnostics
- Fraud Detection
- Email spam detection
- Image Classification

49

B. Unsupervised Learning

Unsupervised learning doesn't require any labelled input data. Algorithms discover hidden patterns or structure in the data on their own without the aid of a training set. An unsupervised learning strategy can be used to handle clustering problems which have untagged datasets. Without the assistance of humans, these algorithms identify hidden patterns or data clusters.

Applications:

Dimensionality reduction

- Big Data Visualization,
- Recognition of Image,
- Mining of Text,
- Recognition of Face.

Clustering

- City Planning
- Targeted Marketing
- Biology

Association analysis

- Market Basket analysis,
- Mining of Web usage,
- Continuous production.

Hidden Markov models

- Computer aided finance.
- Analysis of speed.
- Speech recognition & Synthesis.
- Tagging of speech segments..
- Document separation in scanning solutions.
- Automatic translation.
- Recognition of handwriting

C. Semi-supervised Learning

Amid supervised and unsupervised learning, semi-supervised learning is a subclass of machine learning. It is a technique for training models that combines a sizable amount of unlabeled data with very little tagged data. Similar to supervised learning, the aim of semi-supervised learning is to build a function that can precisely predict the output variable from the input variables. The method is trained on a dataset that includes both tagged and untagged data, in contrast to supervised learning. In order to organize the data and produce predictions, the model needs to learn its structure.

Applications:

- Where high accuracy needed

With only half of the training data, semi-supervised algorithms can achieve very high accuracy (90%–98%). The best accurate prediction is made by the K-Nearest Neighbor (KNN) model for supervised learning and by logistic regression for semi-supervised learning.

D. Reinforcement Learning

A dynamic environment and a computer program interact. It performs a task in that setting that requires it to do a certain activity (such as driving a car or playing in a game). As the program advances through its problem space, feedback is given that is analogous to rewards or penalties in the environment, which it then attempts to maximize or minimize.[5] This type of process is used in self driving cars and robot control.

Applications

- Gaming
- Finance Sector
- Manufacturing
- Management of Inventory
- Robot Navigation

III. MACHINE LEARNING WORKFLOW

The various stages [5] involved in the workflow of machine learning are as follows:

1. Data Gathering
2. Preprocessing of data
3. Selecting Learning Algorithm
4. Training Model
5. Evaluation of the Model
6. Predictions

Data may be collected from several premises, including files, databases, and so forth. The obtained data's quantity and quality directly affect the targeted system's accuracy. To clean up the raw data, data preparation is carried out. Clean datasets are created from real-world data that has been modified. Raw data could include instances of duplicates, inconsistent or missing numbers, etc. [5]

Then, the best learning algorithm is investigated. Depending on the sort of data we have and the problem that has to be solved. Training dataset and testing dataset are created from the dataset. The split between training and testing is approximately 80/20 or 70/30. [12] The dataset size is another factor.

For training purposes, training datasets are utilized. The testing purpose makes use of the testing dataset. The model is assessed to see how effective it is. To evaluate the performance, metrics including recall, recall accuracy, and precision are used. By adjusting the hyper parameters, the accuracy may be raised still higher.

Finally, the created system is put to use in the real world. Here, machine learning's true value is realized.

IV. HOW MACHINE LEARNING ALGORITHM WORKS

Fig.3 shows the different processes involved in machine learning, starting from data inputting to model building and output generation. [6]
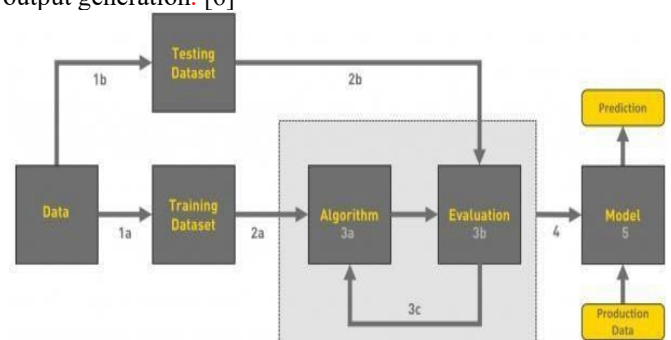

Fig. 3. Different processes in machine learning

50

A typical goal in machine learning is researching and creating algorithms that can learn from data and make predictions based on it. The data input required to construct the model is divided into multiple data sets.

At various points during the model-building process, the training, validation, and test sets are commonly used. Particularly, the training, validation, and test sets are frequently utilized during different stages of model development.

Vectors (or scalars) representing input pairs and the associated output vectors are frequently seen in data sets, while the response key is the target (or label).

The present model is run on the data set of training and the outputs are compared to the target after each input vector produces results. In light of the results of comparison and the particular learning method under use, the variables of the model are modified. Estimates of the parameters and the choice of variables may be part of the model fitting procedure.

In response to the second data set's observations, also known as the validation data set, are incrementally predicted using the fitted model. While the model's hyper parameters are changed, (for instance, the number of hidden units, layers, and layer widths in a neural network), the validation data set offers an objective assessment of a model's fit on the training data set..

Validation datasets can be used to achieve regularization by stopping training while the error on the validation data set grows—an indication that the model is becoming overfit to the training set of data. [6]

This basic technique is more difficult to use in practical since the inaccuracy of validation dataset may change during training, leading to several local minima. Due of this complexity, numerous ad hoc rules have been developed to determine when over-fitting has actually started.

The test data set, which is the last but not least, is a set of information used to unbiasedly evaluate how well the training data set was fit by the final model. The data in the test data set that has never been used in training, may also be referred to as a "holdout data set." (e.g. in cross-validation). [12]

In certain publications, the term "test data" is interchanged with the word "validation data" [12] (for example, if the initial data set was split into only two subgroups, the test data can be referred to as the validation data.)

The task and the information at hand heavily influence the sizes and division methods for training, test, and validation sets.

Every machine learning algorithm has three main components, [12] which are as follows:

- Representation: How knowledge is represented; how the model appears.
- Evaluation: How to distinguish between good models; how to assess programs.
- Optimization: the method for identifying effective models; the method for producing programs.

The final step of the optimization process, which involves continually training the model, is the evaluation of the maximum and minimum function. This is one of the most important phenomena in machine learning for enhancing results.

Optimization reduces the degree of error. Iteration by iteration, it improves accuracy. Statistical or probabilistic techniques are used for optimization in machine learning. [7]

V.   DIFFERENT MACHINE LEARNING ALGORITHMS

The different machine learning algorithms with the categories they belong to are given in Fig. 4.
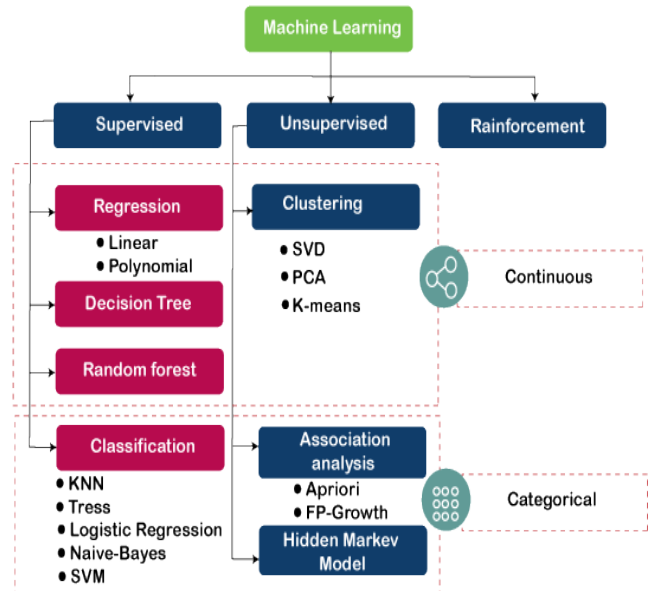

Fig. 4. Different Machine Learning Algorithms

A. Regression

Regression analysis is a statistical technique that uses one or more independent variables to explain the relationship between dependent (target) and independent (predictor) variables. It predicts real, continuous values like temperature, age, salary, and cost, among others. The value of one variable can be predicted using linear regression analysis, based on the value of another variable. [5]

Regression techniques like polynomial regression use a linear model to represent a non-linear dataset. The logistic regression model serves as an example of supervised learning. It is used to determine or forecast the likelihood that a binary (yes/no) event will occur. Predictive analytics and classification frequently use logistic regression.

The main applications of regression are forecasting, prediction, time series modelling, and establishing the relationship of casual impact between variables.

Advantages:

- It is simpler to implement, comprehend, and train.
- It tackles overfitting reasonably well utilizing dimensionality reduction approaches, regularization, and cross-validation.

51

Disadvantages

- Linear Regression assumes the dependent variable's relationship to the independent variables is linear. The data is rarely linearly separable in the real world,
- It is prone to noise and sensitive to outliers.

Applications:

- Determining Market trends,
- Using temperature and other factors, Prediction of rain
- Prediction of road accidents due to rash driving.

B. Decision Tree

Decision trees in machine learning presents the issue and all potential solutions, it provides an efficient way to make decisions. Categorical variable and continuous variable decision trees are the two categories into which decision trees may be classified.

Classification trees determine whether or not an event occurred. This often has a "yes" or "no" resolution. On the other hand, regression trees provide predictions about continuous values based on prior information or data.

The algorithm of decision tree is mainly used for building a training / regression / classification model in tree structure form (root, branch, and leaf) with the aid of decision trees or decision rules, as to predict classes/categorize future/new data target variables. [8]

Random Forest has a decision trees group and the outcomes of all are combined into a single output. They are effective models because they are capable of reducing overfitting without significantly raising error related to bias.

Advantages:

- Easy to Interpret,
- Less Data Preparation,
- Non-Parametric i.e. for both quantitative and Qualitative,
- Versatility i.e. for both categorical and numeric data.
- Can effectively handle non-linear data sets.

Disadvantages:

- Unstable in nature & Overfitting.
- Feature Reduction & Data Resampling,
- Expensive process,
- Unsuitable for continuous variables.

Applications:

- In engineering, civil, law and business

C. Support Vector Machine (SVM)

Support vector machines are a collection of supervised learning methods for data classification, regression analysis, and outlier detection. Given a set of training examples that have all been classified as belonging to one of two categories, an SVM training algorithm builds a model that predicts whether a new example will fall into one of the two categories.

SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

Converting data into a feature space with high dimensions for analysis, SVM categorizes even when data points are not linearly separable. Once a line has been drawn between the categories, the data are modified to enable the hyperplane representation of the separator. [6]

Advantages:

- SVM performs reasonably well when classes have large gap,
- High dimensional spaces lend themselves better to SVM,
- SVM works well in situations where there are more dimensions than samples,
- SVM uses relatively little memory.

Disadvantages:

SVM algorithm does perform well in following situations:

- When compared to number of training samples, if number of features for each data is greater,
- When the target groups overlap and the data set has more noise,
- There is no probabilistic explanation for the classification because the support vector classifier places data points above and below the classifying hyper plane,
- When data sets are large.

Applications:

- Gene, web pages and email classification,
- Intrusion detection,
- Detection of Face,
- Recognition of handwriting.

D. K Nearest Neighbour (KNN) Algorithm

It's a simple technique that stores all available data and classifies any new data by taking a majority approval of its k neighbors. The data is then placed in the class with which it has matching characteristics. A distance function (Euclidean, Minkowski, Manhattan, or Hamming distance) performs this measurement.

Advantages:

- Simpler to interpret,
- Short time of calculation,
- Accuracy is higher,
- Good for both classification and regression.

Disadvantages:

- KNN requires expensive computation.
- Variables should be normalized to prevent algorithm bias from upper range variables.
- Preprocessing of the data is still required.
- Determining the K (number of nearest neighbors) parameter value is necessary.

52

Applications:

- Both classification and regression issues can benefit from its use.

E. K means Algorithm

K-means is a statistical technique that identifies the centroid by averaging the dataset, where K is the number of clusters. K-means classifies the datasets into K different clusters based on similarities and differences. Because the algorithm is centroid-based, each cluster has a corresponding centroid.

The goal is to reduce the distance inside a cluster in between data points and their centroids. The positions of the clustered centroids are improved repeatedly by beginning with a cluster of centroids that is randomly selected. [8]

K-means cluster formation process:

- For each cluster, the K-means algorithm chooses k centroids, or points,
- Each element of data forms a cluster with its nearest centroids, or K clusters,
- Based on the cluster members already present, it now generates new centroids,
- Using updated centroids, each data point's closest distance is calculated. This process is done until the centroids don't change.

Advantages:

- Mass data sets can be handled,
- Simpler for implementation,
- Can adapt to new examples,
- Ensures convergence,
- Generalizes to other clusters of different sizes and shapes, including elliptical clusters,
- Can warm-start centroids' positions.

Disadvantages:

- k value has to be chosen manually,
- Relies on initial points,
- Clustering data of varied sizes and density.
- Clustering outliers,
- Scaling over a variety of dimensions.

Applications:

- Spam detection and filtering,
- Identification of fake news, etc.

F. Principle Component Analysis (PCA) Algorithm

Dimensionality reduction is accomplished by the unsupervised learning method of Principle Component Analysis. [8] Because it contains many highly connected features, it helps to reduce the dataset's dimensionality. [6]

The goal is to create a clear, accurate representation of the data that reduces or eliminates statistically unnecessary components. [6] [10]

It is a statistical procedure that turns a set of linearly uncorrelated data into correlated features observations. PCA considers each attribute variance since it lowers the dimensionality and a large variance denotes a good separation between groups.

Advantages:

- PCA is based on linear algebra, calculations can be computerized easily. So, computations are easier,
- Machine learning algorithms are made run faster,
- Addresses the problems of high-dimensional data.

Disadvantages:

- Principle components' poor interpretability,
- Difficult to interpret even when there is linear relationship between principle components and the features from the original data,
- The trade-off between dimensionality reduction and information loss.

Applications:

- It is most frequently used tool for exploratory data analysis and predictive modelling,
- In the real world, PCA can be applied to tasks like movie recommendation systems, image processing and power distribution optimization over a range of communication channels.

G. Singular Value Decomposition

Singular Value Decomposition (SVD) of a matrix finding is the process of finding three different matrices as factors. SVD is used in calculation of Pseudo-inverse, the rank of matrix and to minimize the least square error which are adopted in digital signal processing and image processing.

Advantages:

- Data is made simpler,
- Noise is eliminated,
- Algorithm outcomes may improve.

Disadvantages:

- It could be difficult to understand transformed data,
- Not unique because for each Eigen space, different orthogonal basis are chosen.

Applications:

- Matrix inverse calculation,
- As data reduction method,
- Used in least squares linear regression,
- Removing noise from data,
- Image compression.

H. Naïve Bayes Algorithm

The Nave Bayes classifier is a supervised learning system that uses object probability to make predictions about the future. The nave assumption that says that variables are independent of one another is followed by the algorithm known as Nave Bayes, which is based on the Bayes theorem.

The conditional probability, which describes the probability that event (A) will take place provided event (B) has already taken place, is the foundation of the Bayes

53

theorem.[8] Following shows the The Bayes theorem's is in equation 1:

$$P(A/B) = (P(B/A) * P(A)) / P(B) \qquad (1)$$

The Nave Bayes classifier is among the top classifiers that offer a good solution for a particular issue. A Naive Bayesian model can be easily constructed and is well suited for the vast amount of dataset.

Advantages:

- Simple to build,
- No requirement of training data,
- Can handle both discrete and continuous data,
- Being faster, it is used to make real-time predictions.
- It is largely scalable since it can accommodate large number of predictors and data points.

Disadvantages:

- The assumption that all predictors (or attributes) are independent is extremely uncommon in reality.
- The 'problem of zero-frequency' is a situation in which an algorithm assigns zero probability to a categorical variable whose group was absent in the training dataset but was included in the test data set.

Applications:

- It is used for text classification

I. Machine Learning Algorithms' Comparison

The accuracy, precision, and recall outcomes of the used machine learning methods are displayed in Table 1. [11]

TABLE I.   ACCURACY, PRECISION AND RECALL

Algorithm	Accuracy	Precision	Recall
Logistisc Regression	91.30%	87.50%	85.40%
Random Forest	88.70%	81.60%	84.20%
Decision Tree Classifier	84.40%	80.60%	75.70%
K-nearest Neighbors	81.20%	76.40%	71.20%
Support Vector Machine	77.30%	59.80%	68.30%

The best metric for selecting the most effective machine learning algorithms is accuracy. [12] Better decisions can be made with the aid of a highly accurate machine learning algorithm. Accuracy aids in speeding up the machine learning algorithm selection process.

The accuracy mathematical formula [11] is:

$$Accuracy = (TP+TN) / (TP+FP+FN+TN) \qquad (2)$$

To calculate precision, divide the total number of true positives by the sum of true positives and false positives.

The precision mathematical formula [11] is:

$$Precision = (TP) / (TP + FP) \qquad (3)$$

Recall is referred to as the indicator of accurately detecting true positives. By using the pertinent data, recall can be used to determine how accurate the machine learning model is.

The recall mathematical formula [11] is:

$$Recall = (TP) / (TP + FN) \qquad (4)$$

TP (True Positive): The positive cases are predicted to be positive by true positive.

TN (True Negative): The negative events are predicted to be positive by True Negative.

FP (False Positive): Negative cases are predicted as positive by False Positive.

FN (False Negative): The positive cases are predicted as negative by False Negative.

Fig 5 shows plot of Accuracy, Precision and Recall with respect to different ML algorithms given in Table 1.
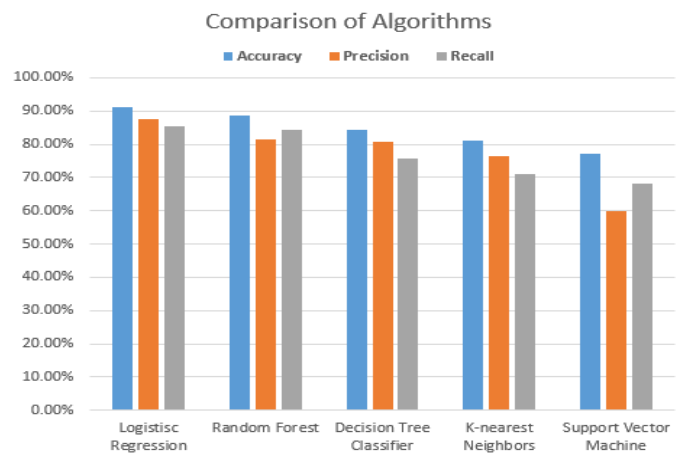

Fig. 5. Comparison of Algorithms

Results of Logistic Regression had an accuracy of 91.3%, Random Forest was second with 88.7%, Decision Tree Classifier was third with 84.4, K Nearest Neighbors was fourth with 81.2%, and Support Vector Machine was seventh with 77.3%.

VI.   CASE STUDY

Consider the problem statement of predicting house price as per size of the house adopting linear regression ML algorithm. House size is independent variable (x) and House price (h) is dependent variable. Assuming both are linearly dependent, we can create a hypothesis that looks like the equation 5 of a straight line (y=mx+c):

$$h(x) = \theta_0 + \theta_1 \qquad (5)$$

Here (θ_0) and (θ_1) are also referred to as regression coefficients.

Table II shows sample dataset and hypothesis are displayed in the graph shown in Fig 6 below:

TABLE II.   SAMPLE DATASET

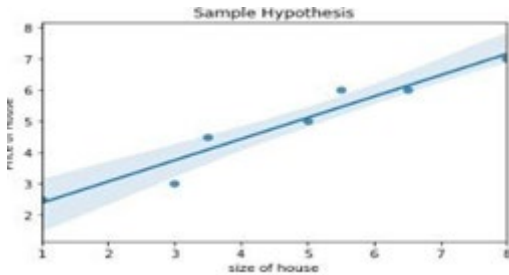| Size of house (10 sq mt)	Price of house (10L)
1	2.5
3	3
3.5	4.5
5	5
5.5	6
6.5	6
8	7

54

Fig. 6. Hypothesis

Different values of θ_0 and θ_1 give different models (straight lines) as in Fig 8.
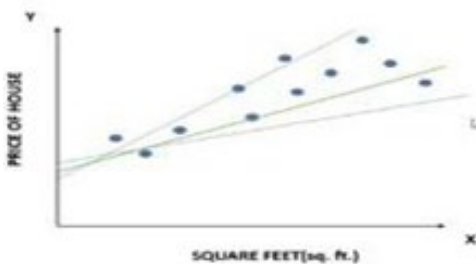

Fig. 7. Different models

It is required to choose the value of θ_0 and θ_1 correctly as to get best fit model (straight line) that assures least cost. Cost function [14] that reflects the error/difference between our hypothetical model's predictions and the real values is is given in equation 6:

$$C(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{i=m} (h_0(x^i) - y^i)^2$$

(6)

The following graph shown in Fig 9 compares changes in $C(\theta_1)$ with changes in (θ_1).
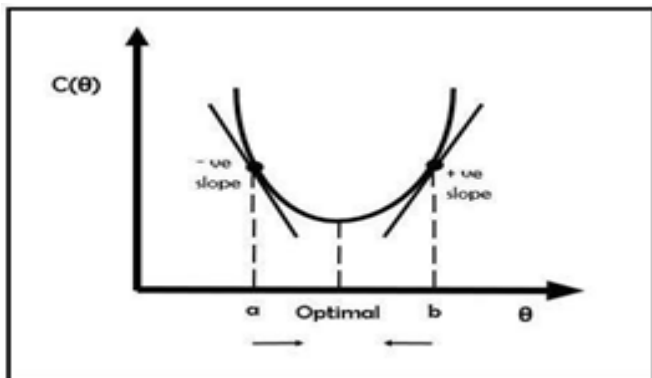

Fig. 8. Cost function Vs θ

The cost function's gradient, which indicates the direction of the sharpest ascent, is the derivative with respect to the model's parameters.

The gradient tells us the direction of the steepest ascent, and by moving in the opposite direction, we can find the direction of the steepest descent.

As we can see, the learning rate (α) is a tuning parameter that determines how fast (θ_0) and (θ_1) change. It decides the length of the steps. [15]

The issue is how to select an appropriate learning rate (α) for the model.

• If the value of (α) is not high enough, our model will take longer to run and converge slowly.

• Our (θ_0) and (θ_1) may exceed the ideal value if the value of (α) is too high, which will reduce the model's accuracy.

• It's also possible that (θ_0) and (θ_1) will keep bouncing between two values in the high value of (α) and never achieve the ideal value.

Starting from random value of θ and to reach minimum value of C takes many iterations.

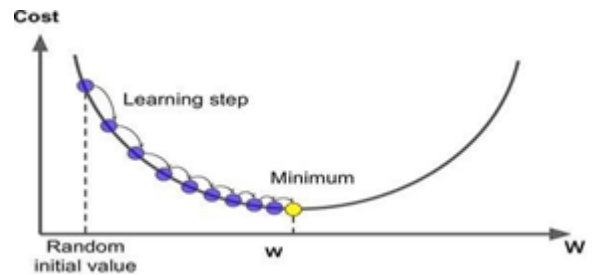Following graph i.e. Fig 10 shows Cost getting reduced as θ is varied iteration by iteration:

Fig. 9. Cost function value Vs θ

The technique's objective is to minimize the cost function by starting with a baseline set of parameters and updating them incrementally. It takes many iterations to find optimal vale of θ.

To improve the model, we can make 2 modifications.

• Increase the amount of features you learn with (use Multi-linear regression).
• Calculate (θ_0) and (θ_1) without utilizing an iterative approach like gradient descent algorithm

Gradient Descent Algorithm:

The algorithm starts with an initial set of parameters (θ) and updates them in small steps to minimize the cost function (C).

$$\text{repeat until convergence } \{$$
$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$
$$(\text{for } j = 1 \text{ and } j = 0)$$
$$\}$$

Every time (θ) is updated till gradient of the function is 0 and hence cost (error) is minimum. Then, it indicates that the model has reached the optimal set of parameters i.e. intercept (θ_0) and slope or gradient (θ_1) in equation $h(x) = \theta_0 + \theta_1$.

VII.   IMPLEMENTATION AND RESULTS

Problem Statement: Build a machine learning model using the data below to estimate home prices based on area.

Inputs: homeprices.csv

ML model: Linear regression with one variable.

area is independent variable, price is dependent variable

TABLE III.   HOME DATASET

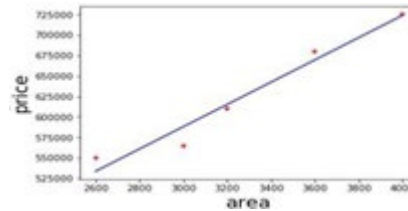| A	B
area	price
2600	550000
3000	565000
3200	610000
3600	680000
4000	725000

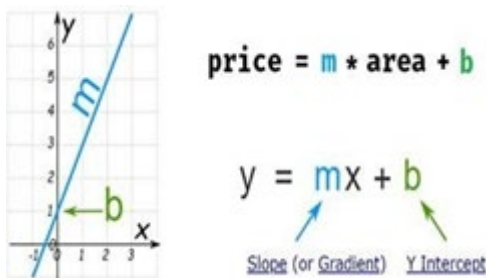Fig. 10 shows the plot of a straight line where m is the slope and b is the intercept.

Fig. 10. y=mx+b plot

The use of linear regression presupposes that the dependent variable (price) and independent variable (area) have a linear relationship. [6]
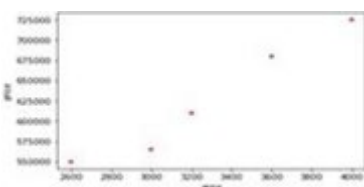
Programming language: Python,

Libraries: pandas, numpy, sklearn, matplot.lib

#Code snippet: Importing libraries

```
import pandas as pd

import numpy as np

from sklearn import linear_model

import matplotlib.pyplot as plt
```

Plot obtained:


```
#Create linear regression object

reg= linear_model.LinearRegression()

reg.fit(new_df,price)
```

#new_def has area values, price has only price values ML model generated (Straight line)

m*x+ b (m is coefficient and b is intercept)

3300*135.78767123 + 180616.43835616432

#Output

628715.7534151643

#Predict home price with area = 5000 sqr ft

#predicting reg.predict([[5000]])

#Output

array([859554.79452055])

Price of house of 5000sq.ft area is predicted to be Rs.859554.79452055.

Given a set of area values, set of predicted prices can also be generated. To consider more than one independent variable, multiple variable regression has to be used.

VIII.   CONCLUSION

In this paper, the concept of machine learning, various machine learning algorithms and their applications in various field along with advantages and disadvantages are discussed. A look into the machine learning algorithms comparison is taken as well. In the case study, the mathematical aspects of linear regression and its application are verified, as well as the use of these theoretical features, including how they may be used in machine learning techniques, for solving real-world computational problems. To best use Machine Learning, in Artificial Intelligence, large data set, correct choice of Machine Learning algorithm, mathematics involved and programming skills are needed.

REFERENCES

[1]A Brief Survey of Machine Learning Methods and their Sensor and IoT Applications, Uday Shankar Shanthamallu, Andreas Spanias, Cihan Tepedelenlioglu, and Mike Stanley* SenSIP Center, School of ECEE, Arizona State University, NXP Semiconductors*, Tempe, AZ 85287, USA
[2] Sindhu V, Nivedha S, Prakash M (February 2020). "An Empirical Science Research on Bioinformatics in Machine Learning". Journal of Mechanics of Continua and Mathematical Sciences https://jmcms.s3.amazonaws.com/wp-content/uploads /2020 /02/24093934/6-AN-EMPIRICAL-SCIENCE-RESEARCH.pdf
[3] "Machine Learning: Algorithms, Real-World Applications and Research Directions", Iqbal H. Sarker, 2021, https://doi.org/10.1007/s42979-021-00592-x
[4]"Machine Learning Algorithms and Its Applications: A Survey" by Raja Tawseef Ahmad Mir, International Journal for Research in Applied Science & Engineering Technology,

56

Volume 10 Issue VI June 2022
DOI:https://doi.org/10.22214/ijraset.2022.44672
[5] "Machine Learning Techniques and Algorithms: A Survey" by Renuka R Patil, Bharathi S Shetty, Suresha, International Journal of Innovative Research in Science, Engineering and Technology, Volume 10, Issue 2, February 2021

[6] Machine Learning Algorithm Validation From Essentials to Advanced Applications and Implications for Regulatory Certification and Deployment Farhad Maleki,
[7] "Elements of Statistical learning, Data Mining, Inference, and Prediction" Second Edition, Springer
https://drive.google.com/file/d/ 15VJTmF8IfOcDi3zuva0_q-mnaSCuSx_L/view
DOI:10.1162/neco.1997.9.7.1493
[8] "Machine Learning Algorithms - A Review", Batta Mahesh, International Journal of Science and Research (IJSR) ISSN: 2319-7064 ResearchGate Impact Factor (2018): 0.28 | SJIF (2018): 7.426
[9] Cortes, Corinna; Vapnik, Vladimir N. (1995). "Support-vector networks". Machine Learning. 20 (3) : 273–297. doi:10.1007/BF00994018
[10] "Dimension Reduction by Local Principal Component Analysis", Nanda Kambhatla IBM, Todd Leen Georgetown University, October 1997, Neural computation 9(7):1493-1516
[11] "Comparison of Machine Learning algorithms on detecting the confusion of students while watching MOOCs" by Ganesh Bhumireddy Venkata Ajay Surendra Manikanta Anala
DOI:10.15680/IJIRSET.2021.1002016
[12] "Machine Learning Algorithm Validation From Essentials to Advanced Applications and Implications for Regulatory Certification and Deployment",Farhad Maleki, Nikesh Muthukrishnan, MEnga, Katie Ovens, Caroline Reinhold, Reza Forghani.
[13] "Research on Machine Learning and Its Algorithms and Development", Wei Jin, Northwestern Polytechnical University Ming De College,Xi'an, Shaanxi, China, IOP publishing limited
doi:10.1088/1742-6596/1544/012033
[14] https://www.simplilearn.com/tutorials/machine-learning-tutorial/cost-function-in-machine-learning
[15] https://www.javatpoint.com/gradient-descent-in-machine-learning
[16] Ron Kohavi; Foster Provost (1998). "Glossary ofterms ". Machine Learning. 30: 271–274.
doi:10.1023/A:1007411609915.