Machine Learning For Beginners

Machine learning was defined in 90's by **Arthur Samuel** described as the," **it is a field of study that gives the ability to the computer for self-learn without being explicitly programmed**", that means imbuing knowledge to machines without hard-coding it.

"A computer algorithm/program is said to learn from performance measure P and experience E with some class of tasks T if its performance at tasks in T, as measured by P, improves with experience E." -Tom M. Mitchell.

Machine learning is mainly focused on the development of computer programs which can teach themselves to grow and change when exposed to new data. Machine learning studies algorithms for self-learning to do stuff. It can process massive data faster with the learning algorithm. For instance, it will be interested in learning to complete a task, make accurate predictions, or behave intelligently.

Why we need Machine Learning:-

Data is growing day by day, and it is impossible to understand all of the data with higher speed and higher accuracy. More than 80% of the data is unstructured that is audios, videos, photos, documents, graphs, etc. Finding patterns in data on planet earth is impossible for human brains. The data has been very massive, the time taken to compute would increase, and this is where Machine Learning comes into action, to help people with significant data in minimum time.

Machine Learning is a sub-field of AI. Applying AI, we wanted to build better and intelligent machines. It sounds similar to a new child learning from itself. So in the machine learning, a new capability for computers was developed. And now machine learning is present in so many segments of technology, that we don't even realize it while using it.
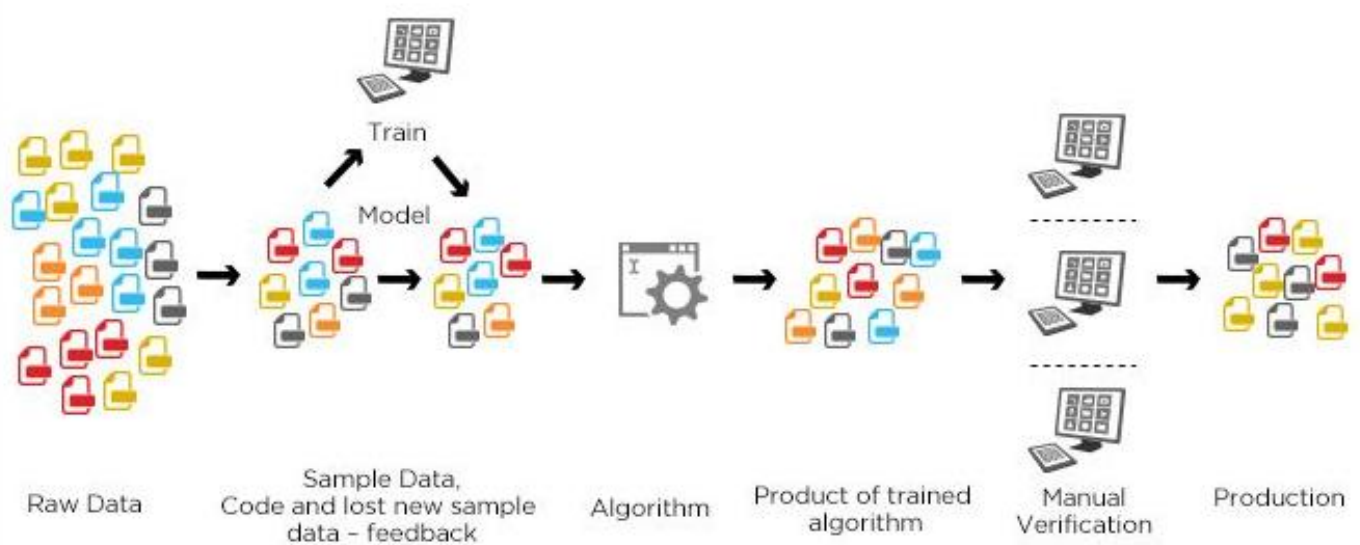
Types of Machine Learning:-

Machine Learning mainly divided into three categories, which are as follows-
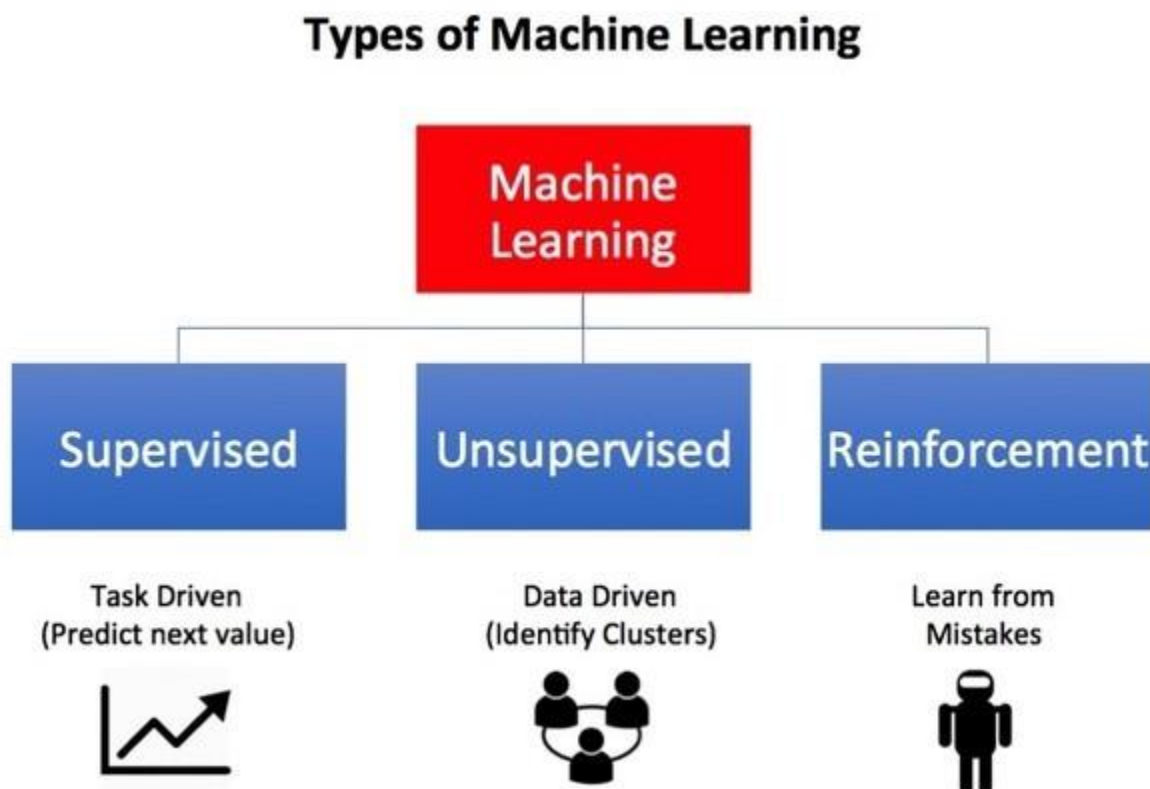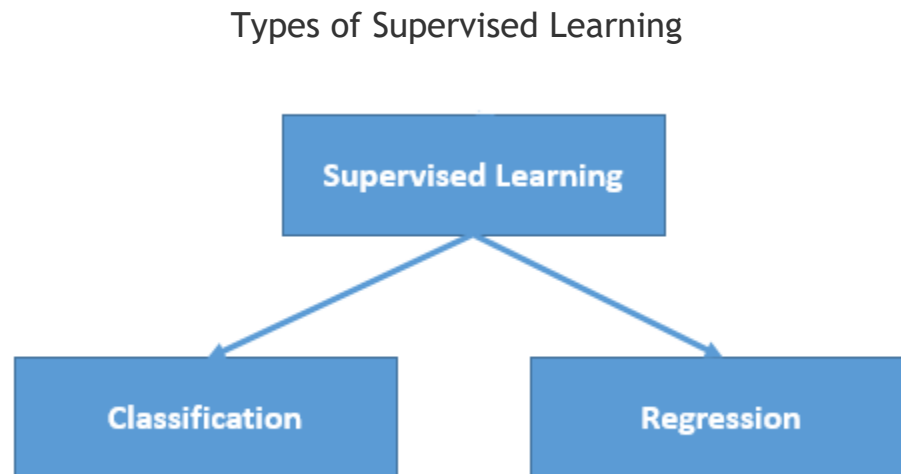
1. Supervised Learning:-

Supervised Learning is the first type of machine learning, in which *labelled* data used to train the algorithms. In supervised learning, algorithms are trained using marked data, where the input and the output are known. We input the data in the learning algorithm as a set of inputs, which is called as Features, denoted by X along with the corresponding outputs, which is indicated by Y, and the algorithm learns by comparing its actual production with correct outputs to find errors. It then modifies the model accordingly. The raw data divided into two parts. The first part is for training the algorithm, and the other region used for test the trained algorithm.

| Raw Data | Sample Data, Code and lost new sample data - feedback | Algorithm | Product of trained algorithm | Manual Verification | Production |

Supervised learning uses the data patterns to predict the values of additional data for the labels. This method will commonly use in applications where historical data predict likely upcoming events. Ex: - It can anticipate when transactions are likely to be fraudulent or which insurance customer is expected to file a claim.

Types of Supervised Learning:-

The Supervised Learning mainly divided into two parts which are as follows-

Types of Supervised Learning

1.1. Regression:-

Regression is the type of Supervised Learning in which labelled data used, and this data is used to make predictions in a continuous form. The output of the input is always ongoing, and the graph is linear. Regression is a form of predictive modelling technique which investigates the relationship between a dependent variable [*Outputs*] and independent variable [*Inputs*]. This technique used for forecasting the weather, time series modelling, process optimization. Ex: - One of the examples of the regression technique is House Price Prediction, where the price of the house will predict from the inputs such as No of rooms, Locality, Ease of transport, Age of house, Area of a home.

Types of Regression Algorithms:-

There are many Regression algorithms are present in machine learning, which will use for different regression applications. Some of the main regression algorithms are as follows-

1.1.1. Simple Linear Regression:-

In simple linear regression, we predict scores on one variable from the ratings on a second variable. The variable we are forecasting is called the criterion variable and referred to as Y. The variable we are basing our predictions on is called the predictor variable and denoted to as X.

1.1.2. Multiple Linear Regression:-

Multiple linear regression is one of the algorithms of regression technique, and it is the most common form of linear regression analysis. As a predictive analysis, the multiple linear regression is used to explain the relationship between one dependent variable with two or more than two independent variables. The independent variables can be continuous or categorical.

1.1.3. Polynomial Regression:-

Polynomial regression is another form of regression in which the maximum power of the independent variable is more than 1. In this regression technique, the best fit line is not a straight line instead it is in the form of a curve.

1.1.4. Support Vector Regression:-

Support Vector Regression can be applied not only to regression problems, but it also used in the case of classification. It contains all the features that characterize maximum margin algorithm. Linear learning machine mapping leans a non-linear function into high dimensional kernel-induced feature space. The system capacity was controlled by parameters that do not depend on the dimensionality of feature space.

1.1.5. Ridge Regression:-

Ridge Regression is one of the algorithms used in Regression technique. It is a technique for analyzing multiple regression data that suffer from multicollinearity. By the addition of a degree of bias to the regression calculates, it reduces the standard errors. The net effect will be to give calculations that are more reliable.

1.1.6. Lasso Regression:-

Lasso regression is a type of linear regression that uses shrinkage. Shrinkage is where data values shrunk towards a central point, like the mean. The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters). This particular type of regression is well-suited for models showing high levels of multicollinearity or when you want to automate certain parts of model selection, like variable selection/parameter elimination.

1.1.7. ElasticNet Regression:-

Elastic net regression combined L1 norms (LASSO) and L2 norms (ridge regression) into a penalized model for generalized linear regression, and it gives it sparsity (L1) and robustness (L2) properties.

1.1.8. Bayesian Regression:-

Bayesian regression allows a reasonably natural mechanism to survive insufficient data or poorly distributed data. It will enable you to put coefficients on the prior and the noise so that the priors can take over in the absence of data. More importantly, you can ask Bayesian regression which parts (if any) of its fit to the data are it confident about, and which parts are very uncertain.

1.1.9. Decision Tree Regression:-

Decision tree builds a form like a tree structure from regression models. It breaks down the data into smaller subsets and while an associated decision tree developed incrementally at the same time. The result is a tree with decision nodes and leaf nodes.

1.1.10. Random Forest Regression:-

Random Forest is also one of the algorithms used in regression technique, and it is very flexible, easy to use machine learning algorithm that produces, even without hyper-parameter tuning. Also, this algorithm widely used because of its simplicity and the fact that it can use for both regression and classification tasks. The forest it builds, is an ensemble of Decision Trees, most of the time trained with the "bagging" method.

1.2. Classification:-

Classification is the type of Supervised Learning in which labelled data can use, and this data is used to make predictions in a non-continuous form. The output of the information is not always continuous, and the graph is non-linear. In the classification technique, the algorithm learns from the data input given to it and then uses this learning to classify new observation. This data set may merely be bi-class, or it may be multi-class too. Ex: - One of the examples of classification problems is to check whether the email is spam or not spam by train the algorithm for different spam words or emails.

Types of Classification Algorithms:-

There are many Classification algorithms are present in machine learning, which used for different classification applications. Some of the main classification algorithms are as follows-

1.2.1. Logistic Regression/Classification:-

Logistic regression falls under the category of supervised learning; it measures the relationship between the dependent variable which is categorical with one or more than one independent variables by estimating probabilities using a logistic/sigmoid function. Logistic regression can generally use where the dependent variable is Binary or Dichotomous. It means that the dependent variable can take only two possible values like "Yes or No", "Living or Dead".

1.2.2. K-Nearest Neighbors:-

KNN algorithm is one of the most straightforward algorithms in classification, and it is one of the most used learning algorithms. A majority vote of an object is classified by its neighbors, with the purpose being assigned to the class most common among its k nearest neighbors. It can also use for regression—output is the value of the object (predicts continuous values). This value is the average (or median) of the benefits of its k nearest neighbors.

1.2.3. Support Vector Machines:-

A Support Vector Machine is a type of Classifier, in which a discriminative classifier formally defined by a separating hyperplane. The algorithm outputs an optimal hyperplane which categorizes new examples. In two dimensional space, this hyperplane is a line dividing a plane into two parts wherein each class lay on either side.

1.2.4. Kernel Support Vector Machines:-

Kernel-SVM algorithm is one the algorithms used in classification technique, and it is mathematical functions set that defined as the kernel. The purpose of the core is to take data as input and transform it into the required form. Different SVM algorithms use different types of kernel functions. These functions can be different types. For example linear and nonlinear functions, polynomial functions, radial basis function, and sigmoid functions.

1.2.5. Naive Bayes:-

Naive Bayes is a type of Classification technique, which based on Bayes' Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular

feature in a class is unrelated to the presence of any other function. Naive Bayes model is accessible to build and particularly useful for extensive datasets.

### 1.2.6. Decision Tree Classification:-

Decision tree makes classification models in the form of a tree structure. An associated decision tree incrementally developed and at the same time it breaks down a large data-set into smaller subsets. The final result is a tree with decision nodes and leaf nodes. A decision node (e.g., Root) has two or more branches. Leaf node represents a classification or decision. The first decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categorical and numerical data.

1.2.7. Random Forest Classification:-

Random Forest is a supervised learning algorithm. It creates a forest and makes it somehow casual. The wood it builds is an ensemble of Decision Trees, it most of the time the decision tree algorithm trained with the "bagging" method, which is a combination of learning models increases the overall result.

2. Unsupervised Learning:-

Unsupervised Learning is the second type of machine learning, in which unlabeled data are used to train the algorithm, which means it used against data that has no historical labels. What is being showing must figure out by the algorithm. The purpose is to explore the data and find some structure within. In unsupervised learning the data is unlabeled, and the input of raw information directly to the algorithm without pre-processing of the data and without knowing the output of the data and the data cannot divide into a train or test data. The algorithm figures out the data and according to the data segments, it makes clusters of data with new labels.

Raw Data   Algorithm   Automated Clusters   Manual Review   Production
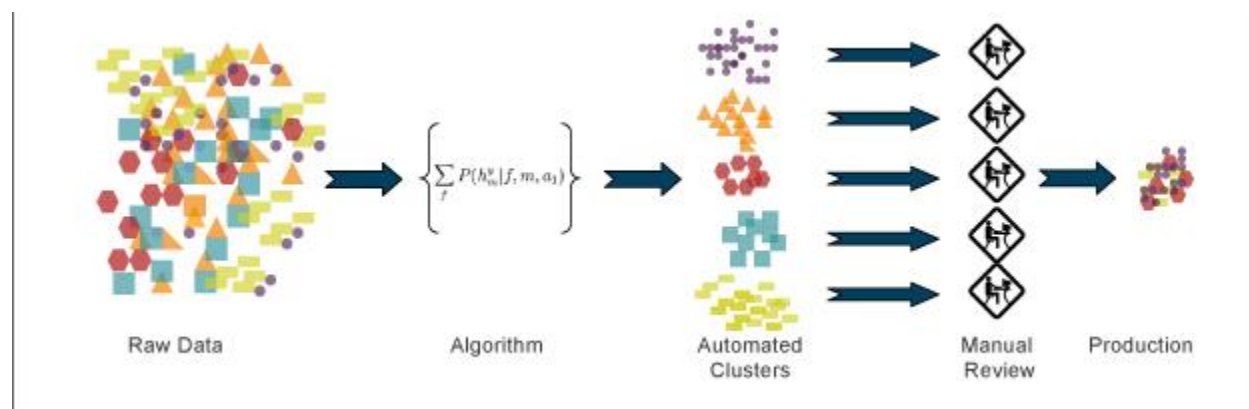
Unsupervised Machine Learning

This learning technique works well on transactional data. For example, it can identify segments of customers with similar attributes who can then be treated similarly in marketing campaigns. Or it can find the primary qualities that separate customer segments from each other. These algorithms are also used to segment text topics, recommend items and identify data outliers.

Types of Unsupervised Learning:-

The Unsupervised Learning mainly divided into two parts which are as follows-
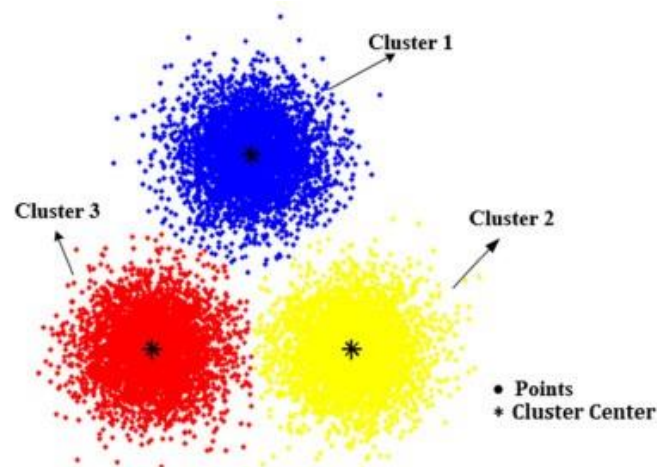
2.1. Clustering:-

Clustering is the type of Unsupervised Learning in which unlabeled data used, and it is the process of grouping similar entities together, and then the grouped data is used to make clusters. The goal of this unsupervised machine learning technique is to find similarities in the data point and group similar data points together and to figures out that new data should belong to which cluster.

Types of Clustering Algorithms:-

There are many Clustering algorithms are present in machine learning, which is used for different clustering applications. Some of the main clustering algorithms are as follows-

2.1.1. K-Means Clustering:-

K-Means clustering is one of the algorithms of Clustering technique, in which similar data grouped in a cluster. K-means is an iterative clustering algorithm that aims to find local maxima in each iteration. It starts with K as the input which is how many groups you want to see. Input k centroids in random locations in your space. Now, with the use of the Euclidean distance method calculate the distance between data points and centroids, and assign data point to the cluster which is close to it. Recalculate the cluster centers as a mean of data points attached to it. Repeat until no further changes occur.

K-Means Clustering showing 3 clusters

2.1.2. Hierarchical Clustering:-

Hierarchical clustering is one of the algorithms of Clustering technique, in which similar data grouped in a cluster. It is an algorithm that builds the hierarchy of clusters. This algorithm starts with all the data points assigned to a bunch of their own. Then two nearest groups are merged into the same cluster. In the end, this algorithm terminates when there is only a single cluster left. Start by assign each data point to its bunch. Now find the closest pair of the group using Euclidean distance and merge them into the single cluster. Then calculate the distance between two nearest clusters and combine until all items clustered into a single cluster.
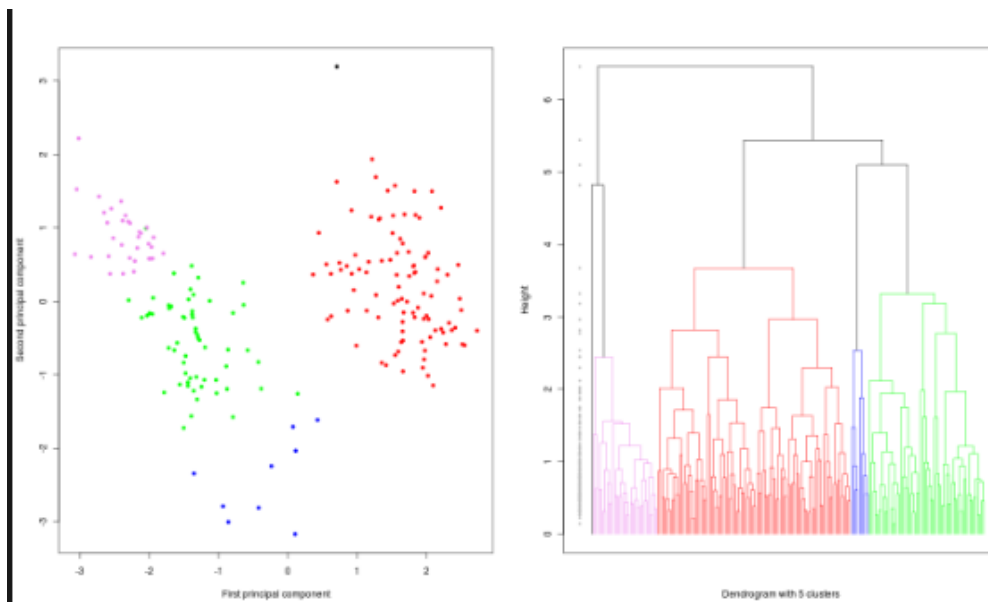
2.2. Dimensionality Reduction:-

Dimensionality Reduction is the type of Unsupervised Learning, in which the dimensions of the data is reduced to remove the unwanted data from the input. This technique is used to remove the undesirable features of the data. It relates to the process of converting a set of data having large dimensions into data with carries same data and small sizes. These techniques used while solving machine learning problems to obtain better features.

Types of Dimensionality Reduction Algorithms:-

There are many Dimensionality reduction algorithms are present in machine learning, which applied for different dimensionality reduction applications. Some of the main dimensionality reduction algorithms are as follows-

2.2.1. Principal Component Analysis:-

Principal Component Analysis is one of the algorithms of Dimensionality Reduction, in this technique, it transformed into a new set of variables from old variables, which are the linear combination of real variables. Specific new set of variables are known as principal components. As a result of the transformation, the first primary component has the most significant possible variance, and each following element has the highest potential difference under the constraint that it is orthogonal to the above ingredients. Keeping only the first $m < n$ components reduces the data dimensionality while retaining most of the data information,

2.2.2. Linear Discriminant Analysis:-

The linear discriminant analysis is one of the algorithms of Dimensionality Reduction in which it also creates linear combinations of your original features. However, unlike PCA, LDA doesn't maximize explained variance. Instead, it optimizes the reparability between classes. LDA can improve the predictive performance of the extracted features. Furthermore, LDA offers variations to tackle specific roadblocks.

2.2.3. Kernel Principal Component Analysis:-

Kernel Principal Component Analysis is one of the algorithms of Dimensionality Reduction, and the variables which are transformed into variables of the new set, which are the non-linear combination of original variables means the nonlinear version of PCA, called as Kernel Principal Component Analysis (KPCA).

It is capable of capturing part of the high order statistics, thus provides more information from the original dataset.

Applications of Machine Learning:-

There are many uses of Machine Learning in various fields, some of the areas are Medical, Defense, Technology, Finance, Security, etc. These fields' areas different applications of Supervised, Unsupervised and Reinforcement learning. Some of the areas where these ML algorithms used are as follows-