

* **Name: Rohit Pravin Sapkal**
* **Course: Data Science For Business**
* **Topic: Machine Learning Algorithms**

**Machine Learning Algorithms**

Table of Contains

• Acknowledgement

• Abstract

• Chapter 1 ---- Introduction

* 1. Introduction
  2. Background of study
  3. Motivation of study
  4. Objective of study

• Chapter 2 ---- Literature Review

• Chapter 3 ---- Methodology

• Chapter 4 ---- Discussion

* 1. Supervised Learning
     1. Linear regression
     2. Logistic regression
     3. Support vector machine (SVM)
     4. K-nearest neighbor
     5. Random forest
  2. Unsupervised Learning
     1. K-means clustering
     2. Hierarchical clustering
     3. Apriori algorithms
     4. Neural networks
  3. Reinforcement Learning

• Chapter 5 ---- Result

• Chapter 6 ---- Conclusion

* 1. Conclusion
  2. Limitation
  3. Recommendations of further Study

• References

* Appendix A: Maxico\_covid19 Analysis and implementation of ML Algorithms
* Sample ML website with Coding

Acknowledgement

I'd want to express my gratitude to the persons listed below for their assistance in completing my dissertation on time.

* Orhan Demirovski
* Mohammad Javaheri

Abstract

Nowadays, everything around us is a machine. Machines have taken on critical roles in our lives. Every human activity is linked to a machine. Machine learning is without a doubt one of the most important and effective technologies in today's society. In machine learning, the computer learns from its previous output and makes judgments on its own, without being explicitly programmed in any particular way. Machine learning is now being used nearly everywhere in today's society. As a result, it's critical to be familiar with the entire machine learning speeds. Machine learning is used in many areas based on different algorithms. We examine machine learning in this study.

**Keyword:** Machine learning, Algorithms, Supervised learning, unsupervised learning

**Chapter 1 - Introduction**

1.1 Introduction

When an artificial intelligence system executes a task, it uses an algorithm to anticipate the output values based on the inputs. Machine learning systems perform two of the most critical tasks: classification and regression. Depending on their application, ML approaches can be classed as either supervised or unattended (Ayyadevara, 2018). Unlike uncontrolled learning algorithms, which deal with unlabeled or unclassified data, supervised learning algorithms have annotated input and output data. An unmonitored algorithm could aggregate unsorted data instead of sorting it based on similarities and differences. Semi-supervised algorithms are utilised more precisely in a variety of master learning systems, including transfer learning and active learning. When a task is accomplished, transfer learning applies what was learned to a new but similar problem, whereas active learning allows an algorithm to search for the user or another source for further information, both of which are advantageous (Bonaccorso, 2017). The author (Bonaccorso, 2017) claims that. Both methods are often used when readily available labelled data is limited. It's a part of AI, according to the theory. Machine learning techniques build a model from a set of sample data known as "training data" that may be used to predict or evaluate without being specifically planned. It is feasible to utilise machine-learning algorithms in a variety of fields such as computer vision and medicine to solve issues that are difficult or impossible to solve using conventional methods.

A type of machine learning called statistical learning is closely linked to computer statistics and attempts to make predictions using computers; however, statistical learning and machine learning are not the same thing at all. By providing tools, theory, and application fields, the study of mathematical optimization contributes to the advancement of machine learning science. Data mining is a discipline of science dealing with the study of exploratory data to solve problems using unsupervised learning methodologies. Some machine learning devices use data and neural networks to simulate the activity of a biological brain. Machine learning is referred to as predictive analytics when it is applied to solve business challenges in a range of industries (Bradley, 1997).

1.2 Background of study

Despite advances in the design of electronic search filters, the authors claim that electronic database scans can result in tens of thousands of articles. This is attributable to a variety of causes, including a rise in overall publications, the complexity of examination problems, which needs a broad range of studies. As a result of these computerized searches, several quotations are frequently produced that are neither relevant nor adequate for full-text inspection (Burrell and Society, 2016).

Literature reviewers devote a significant amount of effort to searching through titles and abstracts in order to locate articles that are relevant to their topic. The lengthy process of screening a large number of search results with a large number of reviewers for a repeating practice that requires careful and attentive expert labour consumes reviewers' time and resources. Examiners do not decide whether or not to include a study at this early stage, the title and abstract phase. This is the cognitive decision that reviewers must make: whether to screen a publication as a full-text publication because it meets the inclusion screening guidelines. At this point in the review process, the decision on whether or not to receive the complete text is made. Because reviewers do not have access to the entire content of the publication, this decision is typically based on informed guesses. At this level, checks are extremely broad, and conflicts among several independent reviewers are rarely addressed. Finally, the primary purpose of this screening method is to ensure that no significant research is missed (Crisci et al., 2012, Domingos and Hulten, 2001). Machine Learning (ML) is a critical component of today's scientific and commercial endeavour. Algorithms and neural network models are used to increase computer system performance. Machine learning methods produce a mathematical model using sample data – generally referred to as "training data" – that may be used to evaluate decisions without explicit training. In the 1950s, Arthur Samuel, an IBM employee, created computer software (Fatima et al., 2017).

Samuel began employing alpha-beta trimming to free up space because the application only had a limited quantity of computer capacity. He was able to build a score formula for his innovation by arranging the items on the board. The scoring function attempts to determine the probability of each win based on the information provided (Gianfrancesco et al., 2018). The software uses the minimax approach, which evolved into the minimax algorithm, to select what to do next in the game. Samuel also invented a number of strategies for adapting and improving his programmed. Samuel's software recorded and memorized all of the locations it had previously encountered, then utilised the rotate learning process to mix this information with the reward function's value. In 1952, Arthur Samuel coined the phrase "machine learning," which translates as "machine learning" (machine learning) (Jordan and Mitchell, 2015).

1.3 Motivation of study

Machine learning tries to identify patterns in your data and then create predictions based on these patterns, which are usually complicated, so that business issues, trends, and problems may be identified and handled. In many areas, including business, machine learning is an advanced data processing method. Automation is used to speed up the process of building data models. First published in Science magazine in 1952, Arthur Samuels' computer training programme was a pioneer in AI and computer games. (Khan and coworkers, 2010) This episode centred on a checker game. The IBM computer will first evaluate which actions are most likely to end in a victory. When Frank Rosenblatt developed the world's first computer neural network in 1957, it changed everything. This was an effective utilisation of the cognitive powers of the human brain. This is where neural networks started, and they've gone a long way since then (Khan et al., 2010).

Finally, a close algorithm is created this year for the first time. It makes it possible for computers to recognize patterns via the use of simple methods. This technique may be used by a visitor to construct a route that starts at a random place and guarantees that the vendor reaches all of the cities on time. A method known as the KNN approach is extensively utilised nowadays in classifying data points according to how they compare to their neighbours. When applied in retail apps that recognize credit card usage patterns, KNN is used to detect trends in credit card usage or to prevent robbery. Gerald DeJong suggested explanatory learning for the first time in 1981. (EBL). The computer analyses the training data and creates an overall rule to eliminate stuff that does not appear to be significant. This type of learning is employed in machine education (Mahesh and . 2020).

Terry Sosnowska's Net Talk, released in 1985, could mimic a baby's speech throughout the language acquisition process (Mohammed et al., 2016). In order to demonstrate how difficult, it is to learn new tasks at the human level, an artificial neural network was created to construct a simplified model of cognitive learning processes. In the 1990s, researchers in the field of machine learning underwent a paradigm change from a knowledge-driven approach to a data-driven one. Researchers and scientists have created computer algorithms capable of analysing massive quantities of data and making decisions based on what they've learned (Xiao et al., 2017).

Microsoft released the Kinect in 2010; it was capable of tracking up to 20 human characteristics at a rate of 30 times per second (Williams et al., 2006). Microsoft Kinect motions and movements allow humans and machines to interact. 2011 was a notable year for advancements in the area of artificial intelligence (AI). IBM's Watson beat human competitors in Jeopardy! among other things. Google Brain is a deep neural network that can be taught to recognise and categorise objects in the real world, according to Google (in particular, cats). Using Google X's machine learning technology, YouTube videos could be searched on their own, and cats could be identified. On Facebook's Deep Face team was created in 2014 a sophisticated software system for detecting and verifying people in photographs that mimics human recognition and verification (Wettschereck, 1994).

This platform was introduced in 2015 and has since made machine learning more accessible to developers while also speeding up software development time. Microsoft has also created a framework for distributed machine learning that makes it simple for programmers to spread machine learning challenges across many workstations in a distributed computer system. While this may be true, more than 3,000 experts in the fields of artificial intelligence and robotics have signed an open letter warning of the dangers of autonomous weapons that can make judgments without human intervention (Watson et al., 2019). In 2016, Google's artificial intelligence systems defeated a professional Go player. Go is regarded as the most challenging board game in the world. AlphaGo, Google's artificial-intelligence program, came out on top in the tournament, winning five of the six games played. (Uddin et al., 2019).

1.4 Objective of Study

Although many people are unaware of it, machine learning has an impact on our daily lives. Google is a major contributor to this saturation. Google has become an integral part of daily life, and machine learning influences individuals whether they are aware of it or not. Take a look at the following two scenarios. Despite processing over 40,000 searches per second on average and a billion searches per day, the Google search engine still returns relevant search results in seconds. These miniatures are incredible in and of themselves. Machine learning has also made its way into your phones; keep this in mind the next time you use the facial recognition tool to unlock your phone, or how the camera recognizes the face when shooting images. I’m a big fan of Facebook's "Friends you know" feature; one day I met someone, and the next day his name showed in a friend's recommendation; this happens all the time, and my friend is constantly concerned about the rubbish from me (Steinkraus et al., 2005).

Now that you understand how machine learning affects your daily life, let's have a look at some of the more popular use cases for which machine learning is employed in various industries around the world. Banks utilise machine learning to detect fraud, approve loans, and manage their investment portfolios (Stallkamp et al., 2012). Thanks to machine learning, hospitals can now detect and diagnose diseases more precisely than traditional doctors. Machine learning is currently being employed for advertising objectives by firms in some client categories. Businesses are now using machine learning to manage their supply chains and inventory control systems. Businesses are currently using chat bots to engage with clients, which are regarded as the next significant technological advance (Snoek et al., 2012). Online businesses can now construct more accurate recommendation systems than ever before thanks to recent advances in machine learning (Singh et al., 2016). Self-Driving cars is the current craze that has captivated the imaginations of companies like Google, Uber, and Tesla, which have made major investments in this future vision using next-generation machine learning technology. These are only a handful of the most prevalent applications, but they are only the top of the iceberg. As I indicated at the opening of this essay, machine learning may be utilised to solve both social and economic challenges, as well as commercial limits, in novel ways. The options are practically limitless (Silver et al., 2013).

**Chapter 2 - Literature Review**

Data science is exploding, and machine learning is a vital part of it. Data mining operations use statistical methods to train algorithms to categorise or predict. These results are then used to influence app and company policy choices, with the goal of increasing key growth KPIs. Because of the growing importance of big data, it will be necessary for data scientists to be in greater demand as they help identify and gather the data needed to solve business problems. Deep machine learning algorithms may benefit from labelled data sets (also known as supervised learning); nevertheless, utilising labelled data sets to be successful has no requirements (Sharma et al., 2017). It's capable of taking raw, unstructured data and automatically identifying the characteristics that set distinct data types apart. Instead of requiring human involvement, machine learning analyses data without it, enabling it to be quantified in more intriguing ways than before. As a result of advancements in deep learning and neural networks in recent years, advances in a number of areas, including computer vision, natural language processing and speech recognition. An ANN node has three layers: an input layer, a hidden layer, and an output layer. Artificial neural systems is another term for networks (Schapire, 2015). Artificial neural networks, often known as neural networks, are made up of node layers that include an input layer, one or more hidden layers, and an output layer (ANNs). Each artificial neuron or node is connected to the others and has a threshold and weight of its own. Each "neuron" or "node" in the network. When a node's output reaches a certain level, the node becomes active and transmits data to the next tier of the network. Otherwise, no data will be transmitted to the network's upper levels. Although "deep" refers to the number of layers in a neural network, it doesn't always mean "deep" in the sense of the network's depth (Ray, 2019). There are inputs and outputs on more than three layers in a neural network or deep learning system. With "deep learning," we mean a neural network that has more than three layers of inputs and outputs. A simple neural network is one that just has two or three layers and two or three neurons. (Ray, 2019).

**Chapter 3 – Methodology**

In this study I will explore the types of machine learning algorithms understand each and every algorithm with their mathematical approach and implementation of this algorithm. Workflow for machine learning Algorithms

Collecting Information

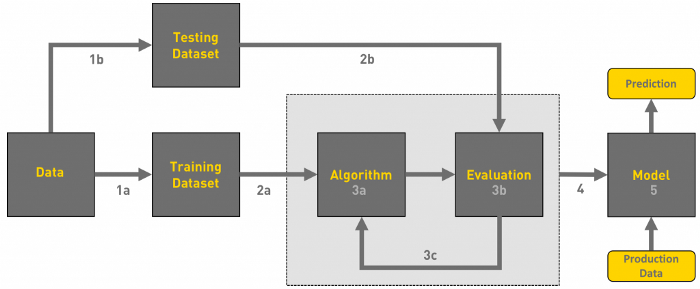
The technique of collecting data is dependent on the sort of project we want to develop; if we want to make an ML project that uses real-time data, then we may design an IoT system utilising various sensors. Numerous different types of data sources may be utilised to gather the data for the analysis process, but the acquired data cannot be used immediately since there may be many missing values or huge amounts of unstructured or noisy text in the data. As a result, Data Preparation has been carried out in order to address this issue. We may also make advantage of publicly available data sets that are freely available on the internet. The most popular repositories for creating Machine Learning algorithm are Kaggle and the UCI Machine Learning Repository.

Pre-processing of data

Figuring out which model will work best with the given set of data

Preparing and putting the model to the test

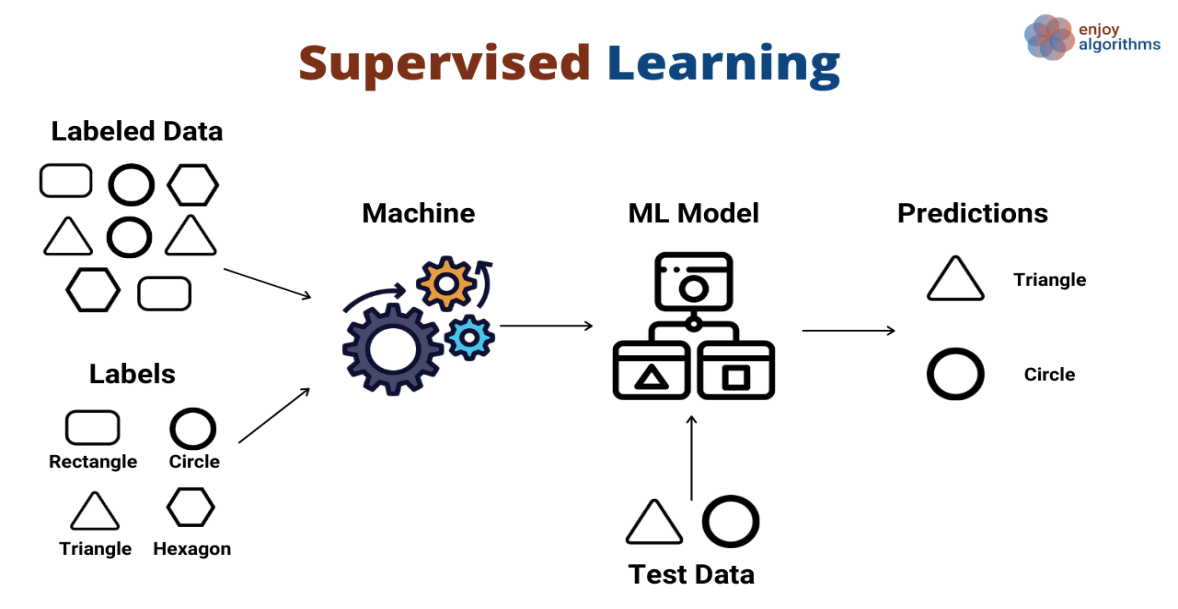
Evaluation



**Chapter 4 - Discussion**

**4.1 Supervised Learning**

To train algorithms that properly identify or predict outcomes via supervised learning (also known as master learning), data sets are labelled and then used to train the algorithms. According to the research conducted by Ossawa et al. Once the model receives new data, the weights are changed to make sure the model is well-suited to the new information.. During the cross-validation step, the model is verified to make sure that it is neither overfitting or underfitting, which is undesirable. Supervised learning in business helps workers react to a broad variety of real-world problems, such as classifying spam and putting it in a separate folder than your e-mail. (SVM) (Ngiam and Khor, 2019).



**4.1.1 Linear regression**

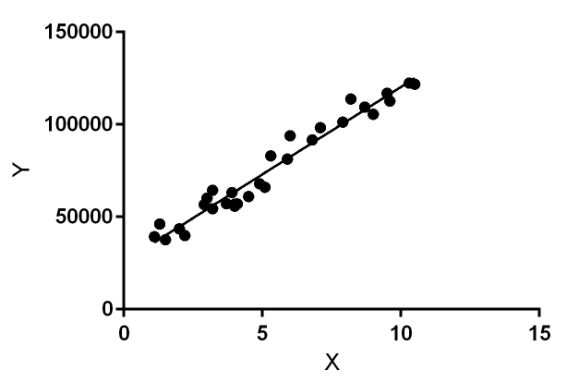
When two variables are compared, one is considered an explanatory variable and the other is considered a dependent variable. To establish a link between people's weights and heights, a modeler, for example, would use a linear regression model (Mohammed et al., 2016).

A model must first establish whether there is a connection between the interest variables in the observed data before evaluating whether there is a link between them. This variable may not necessarily follow the other (for example, better SAT scores may not always indicate higher college rates), but it shows a statistically significant connection between the two variables. . If you want to know how strongly two variables are connected, you may use a dispersion. The effectiveness of a linear regression model is questionable if there seems to be no relationship between the proposed explanatory and dependent variables (i.e. no increasing or decreasing patterns in the scatterplot). This measure shows how closely two variables are connected, with a value ranging from -1 to 1. You may use it to check whether two variables' observed and expected values differ at the same time. Instead than using a single independent variable to explain or forecast the dependent variable's results, multiple linear regression makes use of two or more variables. Regression analysis has many applications outside of finance and investing (Ayyadevara, 2018).

**Working of Linear regression**

It contributes to the creation of a relationship between the variables by determining how much one influences the other. Consider the following scenario: you're looking for a new car and have chosen that the cost per kilometer will be the decisive factor. What strategy would you use to forecast the miles per gallon of upcoming rides. Because you are aware of the car's numerous qualities, one approach of applying it is regression (weight, horsepower, displacement, and so on) (Mahesh and . 2020). After graphing the average MPG of each car, regression techniques can be used to discover the link between MPG and automobile parameters. In the situation where MPG represents Y, X might represent input characteristics such as weight, displacement, horsepower, and so on. $Y = f(X)$ is one way to express the regression function.

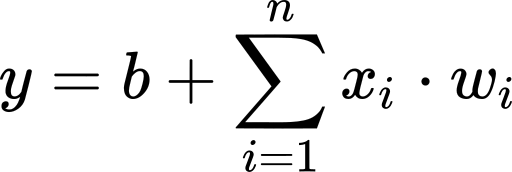
The objective function is $f$, and this curve helps us decide if it is advantageous to acquire. This phenomenon is referred known as "regression." Linear Regression is an approach for forecasting future outcomes based on previous experience. He is responsible for finishing a regression task. Regression is a mathematical model that forecasts an objective result by using independent variables



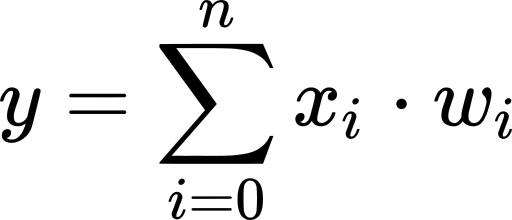
linear regression is a statistical method that uses the value of the independent variable to predict the value of the dependent variable (y) (x). Regression determines the linearity of the connection between x (input) and y (output) (output). As a result, the phrase "linear regression" came into use. The letter X denotes a person's work experience, whereas the letter Y reflects his pay. The regression line is the one that is most similar to our model.

**Mathematics**

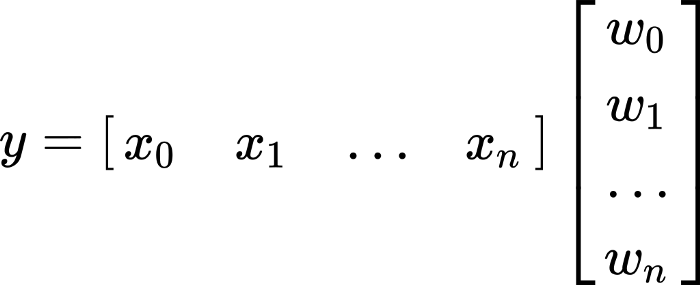
Let's take a mathematical look at how linear regression works. The unknown variable is estimated by constructing a weighted sum of our known variables (input, xi) (noted by y, the output of our model). We then apply a bias factor to the result.

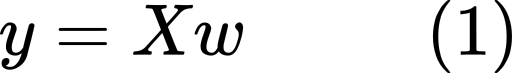


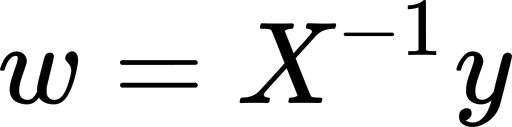
We have n data points available to us. The addition of a bias is equivalent to having an additional input variable that is always 1 and relying solely on the weights to determine the output. We'll utilise this example to make mathematical notation more understandable.

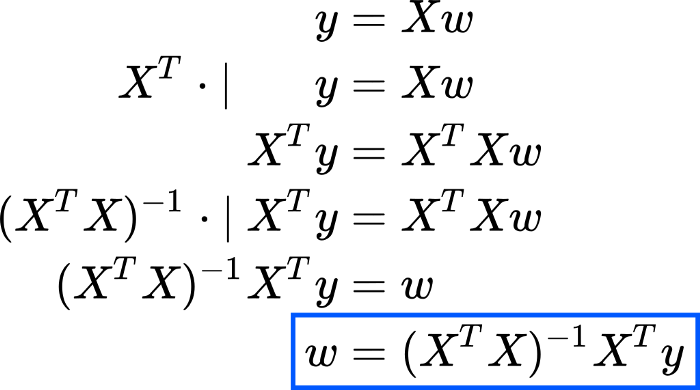


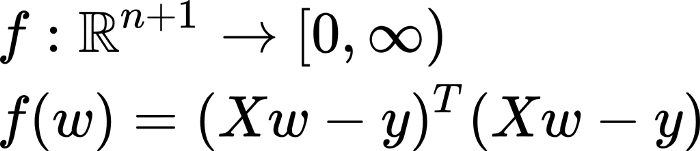
If x0 is always 1, and w0 is the previous b value, we have To make things easier, we'll switch from sum notation to matrix notation. In the previous equation, a row-vector of all input variables is multiplied by a column-vector of all weights to create a weighted sum. In other words.

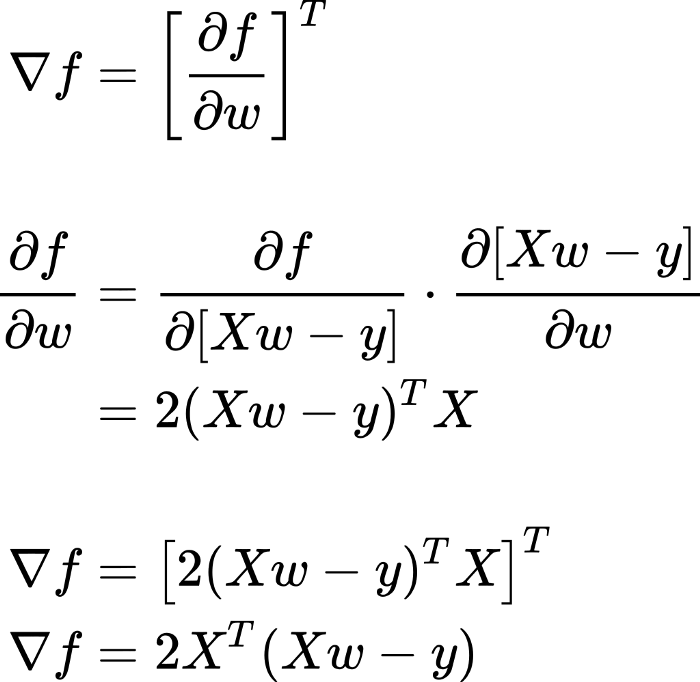


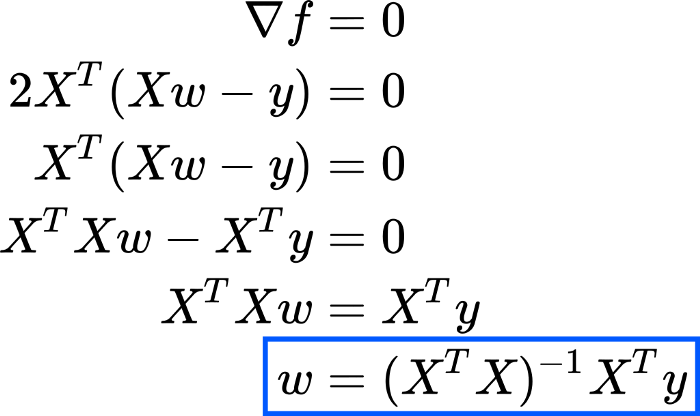


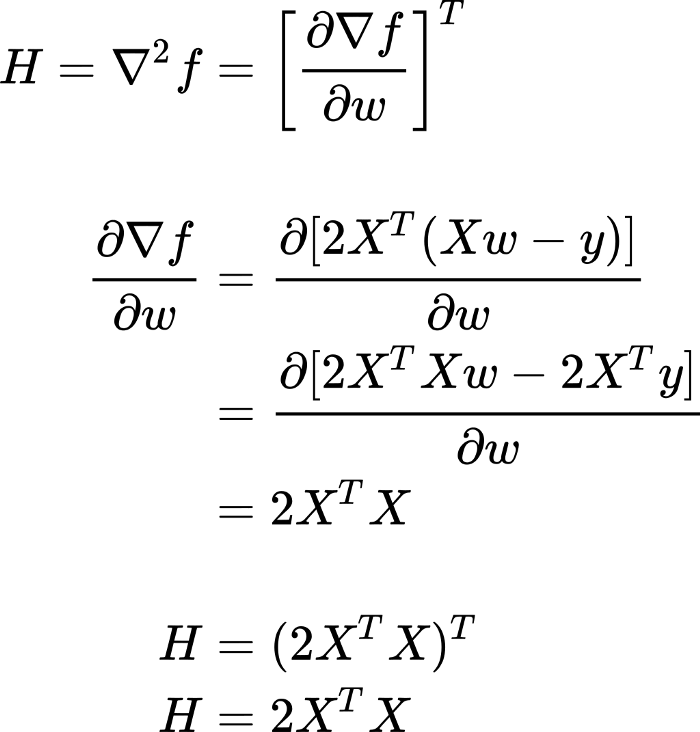


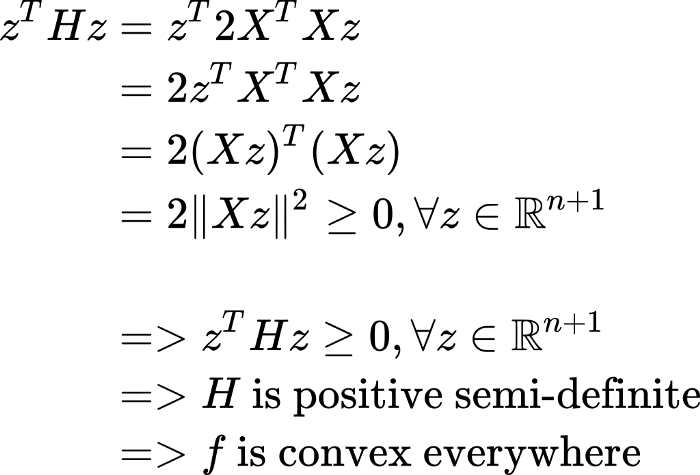












Because f is a convex function, our previously determined solution for w is a minimum point, which is exactly what we desired in this case. As you can see, we got the same answer for w using both the previous linear algebra strategy and the calculus method in this example. When it comes to matrix equations, we can conceive of this as the point that minimizes the sum of square errors or as the solution obtained by substituting y with the projection of y in column X's space. Is this always a realistic option? No. It is, however, more restrictive than the fundamental technique, in which X merely needs to be a square non-singular matrix, but certain requirements must be met. XT X must be invertible and have a full column rank, implying that all of its columns must be linearly independent. This condition is frequently met when our data set has more rows than columns. This condition, however, cannot be maintained if there are fewer data instances than input variables (Bonaccorso, 2017).

In this example, the convexity of f and the criterion that X has entire column rank are entangled. Consider the last section's scant evidence that f is convex. If X has complete column rank, X z cannot be the 0 vector (assuming H is definite positive and f is hence strictly convex). A closed-form solution can exist only in the situation of a strictly convex function, which explains why this is the only example (Gianfrancesco et al., 2018).

**Implementation**

It's time to start learning about the Python linear regression package. To complete the task, make use of the appropriate packages and functions. That is all there is to it. NumPy is a Python scientific tool that enables users to carry out a wide range of high-performance operations on single and multi-dimensional arrays. Scientific Python libraries such as NumPy, which may be used in a wide range of scientific software, use this library. In addition, there are many mathematical methods. It's open source, of course, and fully unrestricted. Khan and his companions (2010)

The official NumPy User Guide and the book Look Ma, No For-Loops: Array Programming with NumPy are excellent resources for learning more about the language. NumPy's performance advantage in your applications may be shown by comparing Pure Python with NumPy, TensorFlow, and TensorFlow vs. NumPy. With NumPy and a few other Python modules, the scikit-learn package creates machine learning models using Python. There are many functionalities available, such as data pretreatment, data reduction and algorithm implementation, such as regression, classification and clustering. Scikit-learn is open-source software, much as NumPy. Our scikit-learn website has additional information on linear models, and the Generalized Linear Models page explains how the package works. (Mahesh and . 2020).

**4.1.2 Logistic regression**

For classification issues, a machine learning method called logistic regression may be used. Probability analysis is a problem-solving and predictive analytical approach built on the notion of probability. (Williams et al., 2006). A logistic regression model is similar to a linear regression model, but its more complex cost function is termed the 'Sigmoid function,' or so-called logistic function, rather than a linear cost function. According to logistic regression, the cost function should have a range of 0 to 1. Because logistic regression does not allow for a value higher than 1 or less than 0, linear functions cannot properly represent it (Sharma et al., 2017).

In a logistic regression model, one or more independent variables are evaluated to determine how they connect to one or more dependent data variables. Whether you want to know if a political candidate will win or lose an election, or if a high school student will be admitted to a certain university, you may use logistic regression. Input criteria for the analytical model created may vary widely. A student's average grades, There may be an algorithm that takes into account a student's SAT score as well as the amount of extracurricular activities they have participated in. It examines new events to see whether they have a high likelihood of falling into a certain group of outcomes. and then compares historical data to previous data using past data and the same input criteria (Uddin et al., 2019).

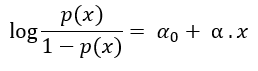
**Working of Logistic regression**

The logistic function is the main function of the logistic regression approach and is so named. Statisticians developed the in environmental studies to describe population growth's features include a fast rise in the capacity for support of the environment using logistic function, also known as sigmoid function. It's a mathematical curve in the form of a S that can convert any real number between 0 and 1, but never exactly within these bounds (Williams et al., 2006).

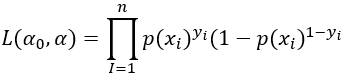
1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 1 / (1 + electronic value) = 1 / (1 + electronic value) = 1 / (1 + electronic value) = 1 / (1 + electronic value) = 1 / (1 + electronic value) = 1 / (1 + electronic value) equivalent to 1 divided by the sum of the electrical value and 1. e is the natural logarithm base (Euler or the EXP() function on your tablet) and value represents the numerical amount be changed, and e denotes the basis of natural logarithms. Here's a list of -5 to 5 values that have been transformed to 0 and 1 using the logistic function: Please refer to the image below (Singh et al., 2016).

**Mathematics**

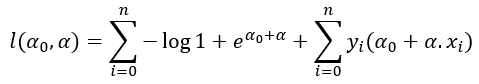
Let's say that p(x) is the linear function for the sake of simplicity. While p represents a probability ranging from 0 to 1, p(x) represents an infinite linear equation, which we do not desire. For this problem, assume log p(x) is a linear x function, and the logit transformation will be used to limit it to the range (0,1). As a result, log p(x)/(1-p(x) is considered. We'll then convert this function to a linear function:











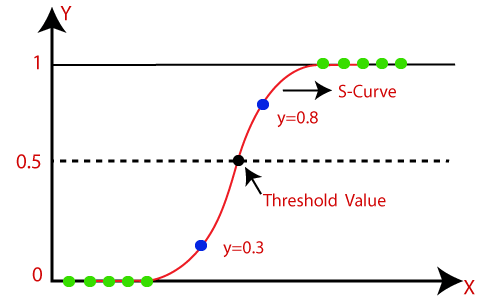
In order to become a linear classification for logistic regression, we might set a certain threshold, like 0.5, for the outcome. Now the misclassification rate can be reduced by predicting y=1 when p is under 0.5 and y=0 when p is over 0.5. The classes are shown as numerals 1 and 0.(Singh et al., 2016)

As logistic regression predicts probabilities, we can adjust it using the distribution method of likelihood. Therefore, the predicted class for each training data point x is represented by the letter y. For y=1, the probability of y is p or 1-p; when y=0, y is p.(Xiao et al., 2017)

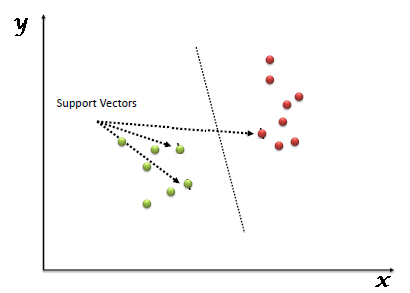
**Implementation**

When using Supervised Learning, a popular method used in machine learning is logistic regression. It uses a collection of categorical independent factors to predict the category dependent variable. Logistic regression can forecast the outcome of a dependent category variable. This means that the conclusion must be either categorical or discrete. It provides probabilistic values from 0 to 1, rather than exact numbers like 0 and 1. A simple yes or no answer is possible, as is a number between 0 and 1. Logistic regression is extremely similar to linear regression, with the exception of the way it is applied. When dealing with regression issues, one should use linear regression; however, when dealing with classification issues, one should use logistic regression (Schapire, 2015).

Instead of a regression line, like with logistics regression, we alter a "S"-shaped logistics function that anticipates two maximum values (0 or 1). As an example of how the logistic function may be used, consider whether or not cells are cancerous or an animal is fat. As a machine learning method for generating probabilities and categorising incoming data, logistic regression is crucial since it can be used to both continuous and discrete datasets and provide reliable results. It's possible to use regression logistics to categorise observations based on several data formats and rapidly find which components are most efficient at categorising observations. Here's a visual representation of the logistics process (Stallkamp et al., 2012).



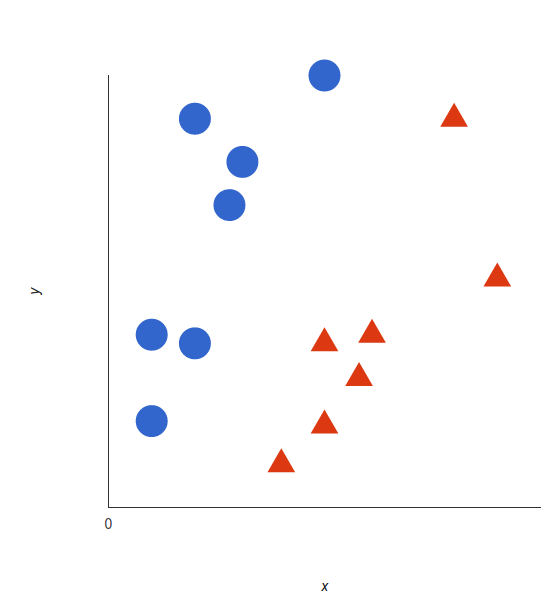
**4.1.3 Support vector machine (SVM)**

The "Support Vector Machine" (SVM) is a supervised machine learning method for classifying and predicting problems. It's most often utilised to address classification issues in real life. Each piece of data is represented as a point in n-dimensional space (n being the number of features), with each feature being a worth a co-ordinate. This method is known as the SVM algorithm. Then we classify by selecting the hyperplane that clearly separates the two groups. 

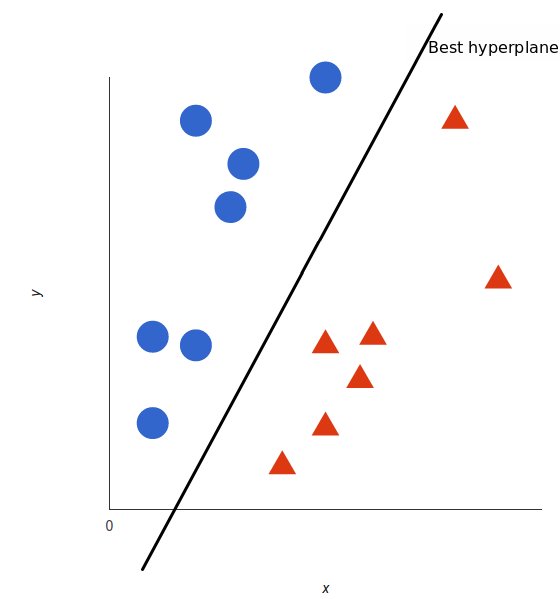
In machine learning, In order to deal with issues involving two data sets, the SVM utilizes classification techniques. New text may be categorised with high accuracy after an SVM model is supplied with sets of training data labelled for each category. Both of these factors set them apart from more recent approaches like artificial neural networks: they are quicker and more accurate when just a limited number of instances are provided (in the thousands). Because only a few thousand classified examples can be accessed in this manner, this technique is particularly well suited to text categorization challenges (Singh et al., 2016).

**Working of support vector machine**

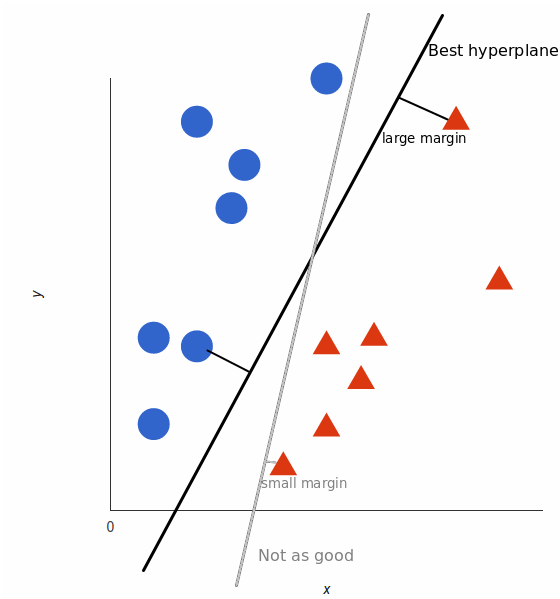
Using a simple example is the most effective method to comprehend the foundations of supporting vector machines and how they work. Consider the following example: (Domingos and Hulten, 2001)We have two tags, red and blue, and two qualities, x and y, in our data. The two tags can then be combined to form a single tag. We're searching for a classifier that can tell whether two (x ,y) coordinates are red, blue, or neither. We track the data that has already been marked on a plane:



When these data points are used, a support vector machine generates the hyperplane (basically a two-dimensional line) that efficiently separates tags. This line symbolizes the moment at which a decision must be taken; everything on one side is blue, while everything on the other is red (Crisci et al., 2012).

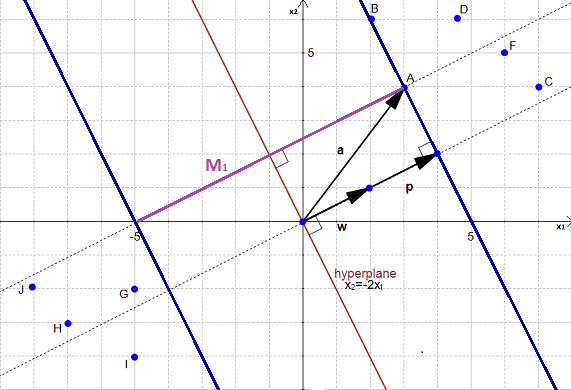


But, specifically, what is the largest hyperplane? In the instance of SVM, it optimizes the margins for both tags. The maximum distance hyperplane is the hyperplane with the greatest distance from the next element in each tag (Jordan and Mitchell, 2015).



**Mathematics**

The distance p between point A and a hyperplane was calculated using the distance formula presented at the end of Part 2. Following that, we estimated a 2p margin. While it did an excellent job of data separation, it was not the best hyperplane for the purpose.(Bonaccorso, 2017)



**Implementation**

Scikit-learn is a Python package used in the programming community to implement computer programmers that use machine learning techniques. SVM is a part of the scikit-learn package and is used in the same way as other libraries (Import library, object creation, fitting model and prediction). The next step is to analyse an actual statement and dataset to see how SVM may be used in the practice of classifying. Dream Housing Finance is a business that specializes in several kinds of mortgages. They are well-represented in all urban, suburban, and rural regions. First, the client applies for a mortgage loan, and then the lender decides whether or not the customer qualifies. The company's goal is to automate (in real time) the credit qualifying process based on information supplied by clients when filling out an online application form. Gender, marital status, educational attainment, number of dependents, income, loan amounts, and credit history are all included in this section. other characteristics. They've established a goal of automating this operation in order to identify the kind of clients who are eligible for loans and target them directly. In this instance, they have only revealed a subset of the data set. To anticipate loan eligibility, utilise the coding window below. Experiment with different hyperparameters to increase the accuracy of the linear SVM (Crisci et al., 2012).

**4.1.4 K-nearest neighbor**

The K-nearest neighbor (KNN) method is simple and most flexible machine learning algorithms. It may help with classification and regression problems. It is easy to construct and understand, but it has a major drawback in that the quantity of data utilised grows much more slowly.

The function computes the distances between a query and all instances in the data, chooses the specified number of examples (K) that are closest to the request, and then votes for the label that occurs the most often (for classification) or averages labels (for averages) (in the case of regression).(Domingos and Hulten, 2001)

In the case of classification and regression, we've seen how the optimal K for our data is determined by trying with several Ks and choosing the best overall.(Crisci et al., 2012)

The KNN approach is founded on the premise that similar things are located close together. In other words, items that are similar to one another are nearby. In the image above, notice how data points that are comparable to each other are often close together. The efficacy of the KNN approach is reliant on the assumption that the algorithm is somewhat accurate. We most likely studied some arithmetic in school that incorporated the concept of similarity (also known as distance, proximity, or closeness). For example, we may recall learning as children how to calculate the distance between two points on a graph. The closest neighbor (KNN) approach is a straightforward and easy supervised machine learning algorithm. It can help with classification and regression problems. It can help with classification and regression problems (Bonaccorso, 2018).

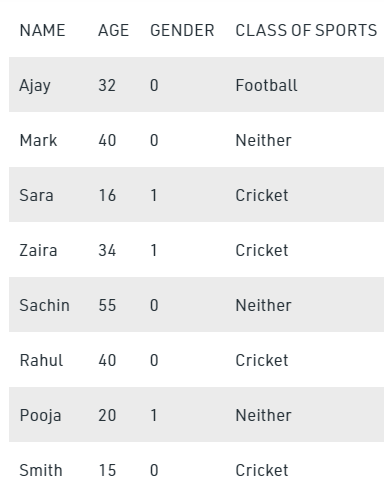
**Working of KNN**

All data science students should be familiar with the K-Nearest Neighbor classifier, which is a basic classifier in data science. Fix and Hodges invented it in 1951 for a pattern categorization assignment, and it was the first time it was used. The KNN classification was given its name because it was intended to be comparable. KNN was developed to aid in pattern recognition (Silver et al., 2013).

K-Nearest Neighbor (commonly known as KNN) is a supervised learning tool that can be used to address regression and classification issues in computer science. It is extensively used in machine learning applications to handle categorization problems. KNN is built on the premise that each data point that is close to another data point belongs to the same class. To put it another way, it categorizes new data points based on how similar they are to old ones. KNN methods select a number k that indicates the data item to be classified's nearest neighbor. If k is 5, the data point's nearest five neighbours will be found. In this scenario, assuming k=4, KNN delivers the names of the four nearest neighbours. The KNN algorithm has determined that all data points near black data points belong to the green class, indicating that this is the sole class to which they belong. The data points in the red class have no connection to the data points in the black class, thus they are ignored. It is possible to forecast the target label using K-classification technology simply by selecting the class that is closest in classification value to the target label in question. The closest class is found by calculating the Euclidean distance between the classification point and the nearest class. (Khan et al., 2010).

**Mathematics**

Because KNN is a "lazy learner," it should only be used if the data set is labelled, noise-free, and constrained in size. Let's look at an example to better understand the KNN algorithm.(Bradley, 1997)



The numeric value 0 is masculine in this example, while the numeric value 1 is feminine. Let's look at which Angelina class she'll be in if her k factor is 3 and she's 5 years old. As a result, we need to calculate the distance.(Khan et al., 2010)

**Implementation**

This method is used to assist the user in resolving issues with the categorization model. The K-nearest neighbor (K-NN) algorithm, often known as the K-NN classification system, is a data classification system that establishes an imaginary border for data classifications. The programmed seeks to maintain fresh data points as near to the boundary line as possible. As a result, increasing the k number results in smoother separation curves and, in general, fewer complex models. Smaller k values, on the other hand, overfit the data and result in more complex models. To avoid data set overlap and failure, it is crucial to use the correct k-value while analysing a dataset. The nearest neighbor method is used to train our model, which allows us to adapt (or train) existing data and forecast the future.(Bonaccorso, 2018)

**4.1.5 Random Forest**

Random forest is a versatile and simple machine learning technique that consistently provides outstanding results even when no hyperparameter modifications are made. Because of its simplicity and wide range of applications, it is also one of the most commonly used algorithms. This page describes the random forest strategy, how it differs from previous approaches, and how it is used.

Random forest constructs a "forest" out of a set of decision-making trees, which are often taught via the "bagging" method. The bagging method is based on the notion that mixing several learning models enhances the final result. As the name implies, random forests are a training a large number of decision trees as a group learning method for classification, regression, and other issues. The random forest's output is the classification issue's class with the most selected trees. Individual predictions for tree regression tasks are returned as medium or average predictions. Random forests are used to offset decision trees' proclivity to outperform their training (Sharma et al., 2017).

In terms of overall performance, random woods surpass trees, but their precision is worse than that of gradient boosted trees. On the other hand, the quality of the data can have an effect on their performance. Tin Kam Ho created the Random Sub-Space Means which is a method of implementing Eugene Kleinberg's classification strategy for "stochastic discriminating," in Ho's words. Tin Kam Ho created the first random decision method for the forest in 1995 using the random subspace approach. Leo Bierman and Adele Cutler created a technique adaption dubbed "Random Forests" in 2006, which was registered as "Random Forests." (As of 2019, Minitab, Inc. owns it.) A collection of controlled variance decision-making trees has been produced by combining Bierman’s 'bagging' approach with a random selection of features provided initially by Ho and then independently by Amit and German.(Williams et al., 2006)

**Working of Random Forest**

The random forest algorithm generates a forest in the form of a group of decision trees, with each tree contributing unpredictability to the forest. During node division, the technique looks for the most important qualities in a random selection of features, resulting in a more diverse model due to the algorithm's increased diversity. As a result, only a random selection of features is inspected when a node is divided. Instead of attempting to find the best possible threshold, random thresholds for each attribute can be used to generate more random trees than the best possible threshold. One of the most enticing features of random forests is the ease with which the relative worth of each parameter in the forecast can be determined. The more characteristics there are in a model, the more probable it is too overfit.(Wettschereck, 1994)

By assessing the relevance of a feature, one can decide which features do not contribute to the forecasting process and should thus be removed. Another notable feature of the algorithm is its versatility. Random forest approaches, depending on the task at hand, can be used for both regression and classification applications. Furthermore, the algorithm's assignment of relative priority to the various input functions is obvious. The Random Forest algorithm's simplicity is also an advantage, as it usually gives great prediction results even when using the default hyperparameters. Aside from that, because there aren't many of them, the hyperparameters are straightforward to grasp. More trees are needed to get a more accurate forecast. More trees, on the other hand, slow down the model. The problem of the random technique for forests is that it is not deterministic. Despite the fact that they may be trained fast, these algorithms are exceedingly slow at forecasting. Because of the high number of trees, forecasting real-time results becomes slow and futile. There are times and situations when runtime performance is more critical than total performance, and the Random Forest Algorithm is used to identify alternate approaches in these cases and scenarios. Random forests are also more of a predicting tool than a descriptive tool. As a result, various approaches are chosen over the random forest algorithm when describing the relationships between data points.(Stallkamp et al., 2012)

**Mathematics**

Random-forest employs a Decision Tree as a starting point for row and column sampling. Because of the sampling of columns, the numbers 1, 2, 3, and 4 change from the models obtained by just bagging.(Snoek et al., 2012)



As the number of basic pupils (k) in the sample increases, the variance decreases. The variance grows as the value of k decreases. There is a consistent bias throughout the process. To identify the value of k, cross-validation can be utilised. A random forest consists of the following components: DT (Base Student) + bagging (Row sampling by replacement) + column sampling + aggregation (Medium/Median, majority vote) In this instance, we want our basic learner to have a low preference and a high variance. As a result, train DT to the greatest extent possible. We don't bother about depth, and we let it increase because total variance diminishes at the end of the day. Datasets (D- D's) that were not used in the modelling phase are eliminated from model h1 bag datasets. This model was integrated with the h1 model for cross-validation.(Schapire, 2015)

**Implementation**

Random forest is a supervised machine learning method based on the concept of overall learning (embedded learning). This is a type of learning in which several algorithms are integrated or the same method is used repeatedly to build a more accurate prediction model. The random wood algorithm mixes many equivalent algorithms (i.e., different decision trees) to build a tree forest, hence the name "Random Forest." The random forest technique can be particularly beneficial when dealing with regression and classification difficulties. As with any algorithm, there are benefits and drawbacks to using this technique.(Watson et al., 2019)

In the following two parts, we'll look at the advantages and disadvantages of random forests for tasks like classification and regression. This algorithm is stable at a high level. Random forest technology is an appropriate solution if you have both numerical and category variables. The random forest strategy can be utilised if the data has missing values or is not properly scaled (although we have performed feature scaling in this article just for the purpose of demonstration).(Wettschereck, 1994)

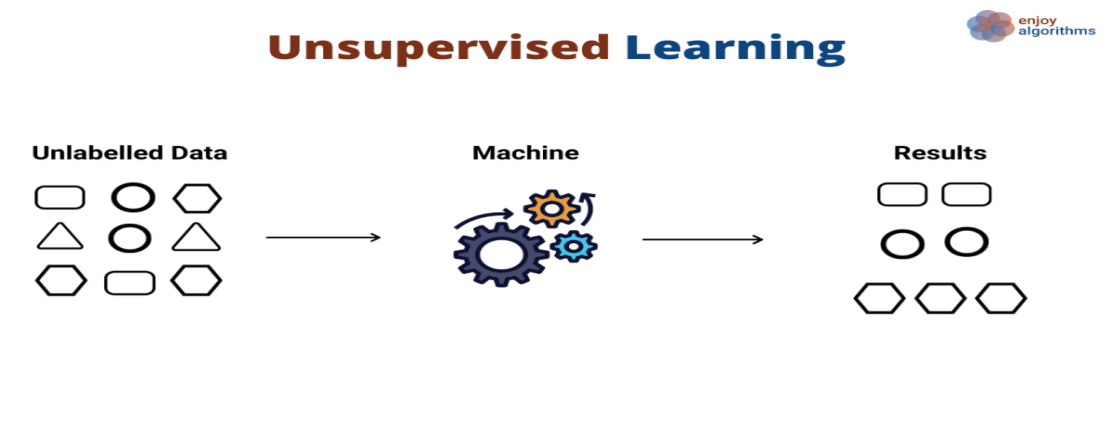
**How all of the above solve the problem and which algorithms are more effective to grow the business.**

Start with linear regression to learn more about machine learning. For example, linear regression predicts continuous output, such as inventory prices or salaries, and is a kind of regression model. Regression analysis uses linear equations to find the linear solution to a given issue. The most efficient method for learning classification strategies is logistic regression, which is also known as linear regression. Classification, not regression, despite the word "regression" being used. It creates a logistical model for a binary output model. Using the logistic return, we can predict whether the probability (0x1) will result in a binary 0 or 1, depending on whether x=0.5. It's possible to use the K-nearest neighbours technique for any kind of data gathering since it's non-parametric. It's one of the most basic methods of machine learning. It's a form of lazy learning that takes a regional approach. One of the greatest algorithms for growing a company is logistic regression. When analysing and predicting a particular set of occurrences, you may find that it is often beneficial, but not always required (Jordan and Mitchell, 2015).

For instance, you may wish to predict the success or failure of a new product or the probability of customer retention or loss. Data from one or more independent variables are used in regression analysis to forecast the result of a categorical dependent variable (also known as independent variables). This is known as Logistic Regression Analysis (LRA) (LRA). A dichotomous variable, which only accepts one of two discrete values, is frequently used to quantify the result of a logistic analysis. The Logistic Regression method examines the connection between the categorical dependent variable and the continuous independent variables by transforming the categorical dependent variable into probability scores. (see Figure 1). In contrast to linear regression models, which predict a continuous result, logistic regression models (LRAs) utilise dichotomous categorical outcome variables to make predictions. Business analysts often use logistic regression modelling. Logistic analysis may be used by an app to predict customer behaviour. You may do research to determine whether or not a potential customer will purchase a product. Instead than using binomial or binary logistic regression to predict just two potential outcomes, multinomial logistic regression is used to predict three or more alternative outcomes, such as yes/no/maybe scenarios for buying goods. (Watson et al., 2019)

**4.2 Unsupervised Learning**

Uncontrolled learning is a machine learning-based technique for analysing and processing unlabeled data that is also known as uncontrolled machine learning. These algorithms can uncover hidden patterns or data groupings without the requirement for human involvement. Its ability to uncover data similarities and differences makes it a fantastic tool for exploratory data analysis, cross-selling techniques, consumer segmentation, and image identification, among other things.. An unsupervised learning strategy for discovering patterns in data sets with unlabeled or uncategorized data points. In this process, artificial intelligence (AI) algorithms are applied. As a result, without the requirement for external monitoring, the algorithms can categories, label, and/or group the data points inside the data sets. To put it another way, unconstrained learning enables the system to discover patterns in data sets without the requirement for human supervision. When an artificial intelligence system organizes data based on similarities and differences even when no categories are provided, this is referred to as uncontrolled learning. Uncontrolled learning algorithms, as opposed to supervised learning systems, can be used to carry out a more sophisticated operation. Artificial intelligence can also be assessed by implementing an uncontrolled learning system.(Uddin et al., 2019)



**4.2.1 K-means clustering**

The K-means method is an iterative approach for dividing a dataset into defined, unique, non-overlapping subgroups (clusters) that only belong to one cluster (clusters). It aims to make data points inside a cluster as similar as feasible while still differentiating the clusters to the greatest extent possible (and therefore as far as possible). It allocates data points to clusters based on their total squared distance from the cluster centroid being equal to or less than a specific threshold. The more homogenous (similar) the data points are, the lower the variation they have inside a cluster. (Sharma and colleagues, 2017)

The K mean algorithm works as follows. K clusters must be provided. To begin, combine the dataset and randomly pick K data points for the centroids, all while preserving the original dataset's data points as much as possible. Continue iterating until the centroids are no longer being updated to reflect your changes. To put it another way, clusters are still being given data points on a regular basis. Add up the square distances between each data point and each centroid in the collection of data. Decide which cluster each piece of data belongs to (centroid). It is possible to compute the cluster centroids by averaging out all of the data from each cluster and then multiplying that result by the number of clusters.(Williams et al., 2006)

**Working of K-means clustering**

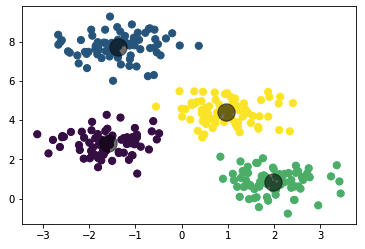
This technique uses k-means to divide an anonymous data set into a predetermined number (k) of clusters (one in which the class identities are unknown). At the start, a random number of centroids are chosen (also known as centroids). Centroids, whether manufactured or natural, are data points located at the center of a cluster. With Pratt, each centroid is a pre-existing data point from the randomly chosen supply data set, ensuring that all centroids are distinct (i.e., all centroids of Ci and Ci > all centroids of Ji). Each of these centres will be used in the training of the kNN classifier. To generate an initial set of randomly gathered clusters that are carried out using the classifier, the data must be categorised (with k = 1) and reviewed (with k = 2). The arithmetic means of the clusters defined by a centroid are determined for each centroid in the collection of clusters. Once the centroid has stabilized, the classification and correction process are repeated until the centroid values have stabilized. The final classification/clustering of the input data will be generated after the final centres.(Burrell and Society, 2016)

**Mathematics**

K-Means clustering, one of the simplest unmonitored clustering techniques can be utilized to partition our data into a total of K groups. The data points are assigned to one of the K clusters progressively. Based on their proximity to the algorithm-determined core cluster. The K-means algorithm produces the following outcome. (Uddin et al., 2019)

1.K denotes the number of cluster centroids in a given cluster.

2.The data points were used to form clusters.



**Implementation**

Before we begin, let's describe the type of problem we're attempting to address. As a result, we have a dataset called Mall Clients that contains information about clients that visit and spend money at the mall. The variables in the data collection are client ID, gender, age, annual revenue ($), and spending value (which is the calculated value of how much a customer has spent in the mall, the more the value, the more he has spent). We need to calculate certain patterns from this dataset because it is unattended and we don't know what to calculate. The following measures must be followed to put the strategy into action:(Steinkraus et al., 2005)

* Preparation and pre-processing of data
* Use the elbow method for optimal number of clusters identification.
* The K-means algorithm is trained using the training dataset.
* to produce a graphical representation of the groupings

**4.2.2 Hierarchical clustering**

Hierarchical clustering, also known as hierarchical cluster analysis, divides components into clusters that are comparable in some way. Hierarchical clustering is used to evaluate data. The endpoint is made up of unique clusters, but the items in each cluster are structurally similar. Hierarchical cluster analysis (also known as HCA) is a non-controlled clustering approach that entails creating groups with dominating order from top to bottom from the start of the procedure to the finish of the algorithm.(Uddin et al., 2019)

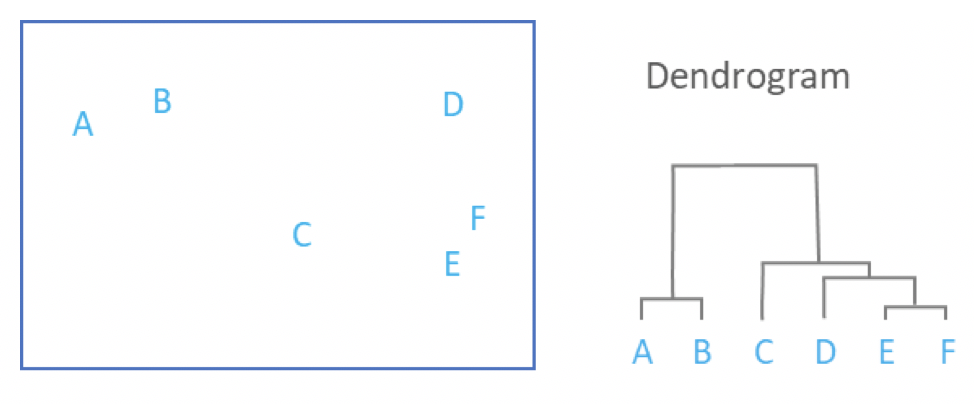
For example, all of the files and folders on our hard drive are organised in a hierarchical sequence. Clusters are formed by grouping comparable elements together. The endpoint is made up of a sequence of clusters or groupings, each distinct from the others, with architecturally identical objects within each cluster.(Ray, 2019)

**Working of Hierarchical clustering**

At the start of the hierarchical clustering technique, each observation is handled as a separate cluster. Then the next two stages are repeated: (1) determine the two most similar clusters, and (2) join the two most similar clusters. This iterative procedure is performed until all clusters are merged into a single measurement unit. The diagrams below show how this is accomplished.(Steinkraus et al., 2005)



Hierarchical Clustering generates a dendrogram, which displays the clusters' hierarchical relationship



**Mathematics**

With hierarchical clustering, there is no mathematical aim to pursue. True, there are a number of limitations to any method for calculating cluster similarity. Hierarchical clustering is difficult in terms of space and time. As a result, this clustering approach is useless when dealing with vast amounts of data.(Silver et al., 2013)

**Implementation**

The following procedures must be performed to ensure the success of agglomeration

* At start, treat each data point as an independent cluster. As a result, at the start of the simulation, the number of clusters is K, where K is an integer denoting the number of data points.
* K-1 clusters are formed by joining the two nearest data points, while K-2 clusters are formed by linking the two nearest clusters.
* The three preceding procedures will be repeated until a huge cluster forms.

Dendrograms are then used, depending on the nature of the problem, to further divide the data into several clusters. In the following section, we'll explore more into the concept of the dendrogram.

Many methods are used to calculate the distance between clusters. The distance between two points can be measured in either Euclidean or Manhattan units. Here are a few other methods for measuring distance between two clusters:(Schapire, 2015)

* Multiply the distance between two clusters' closest points by two.
* Determine the distance between the points of the two most distant clusters.
* To calculate the distance between the two clusters, multiply the distance between the centres by two.
* To get the mean, divide the distance between the two clusters by the number of points within each cluster.

**4.2.3 Apriori algorithms**

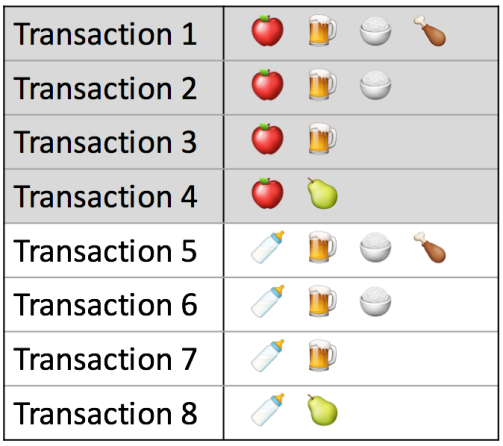
The Apriori approach is intended for use with transaction-based databases and generates association rules using frequently occurring item sets. It is feasible to use association rules to determine whether two items are strongly or weakly connected. Using a first broad search and a Hash Tree, this strategy constructs the object set relationships as rapidly as possible. This algorithm was developed in 1994 by R. Agrawal and Srikant and is still in use today. It is mostly used for market basket analysis, which aids in the finding of products that are more cost-effective when purchased together. The technique can be used in the medical field to detect dangerous medication reactions in patients.(Williams et al., 2006)

**Working of Apriori algorithms**

The algorithm is divided into various steps, the most essential of which are as follows: The technique for creating candidate sets is known as the candidate set creation procedure. Reduce the number of candidates in a pool by deleting those who obtained less than the specified level of support. Form sets of size k+1 by combining frequently occurring item sets and repeating the process until there are no more frequently occurring item sets. This happens when the support in the produced set(s) is less than the planned support.(Uddin et al., 2019)

**Mathematics**

The proportion of transactions in which the item is shown in relation to the overall number of transactions is a popular item. In the table below, the term "apple" receives 4 out of 8 votes, indicating 50% approval. A set of goods can include a large number of products. Apple, beer, and rice, for example, were given to two out of every eight ballots, or 25%.(Fatima et al., 2017)



If you discover that sales of goods over a certain percentage of your entire sales have a significant impact on your revenues, you might want to consider using this percentage as the supporting level for your product. Objects with support values higher than this threshold can also be identified as important items.(Silver et al., 2013)

**Implementation**

This is a sort of algorithm used to obtain insight into the structural relationships between the things being investigated. The most prevalent practical application in the real world is product recommendations based on things currently in the customer's shopping basket. Walmart, in example, used the algorithm extensively to promote goods to its customers.

Step 1: Load and assess the information.

Step 2: Add the libraries to the system.

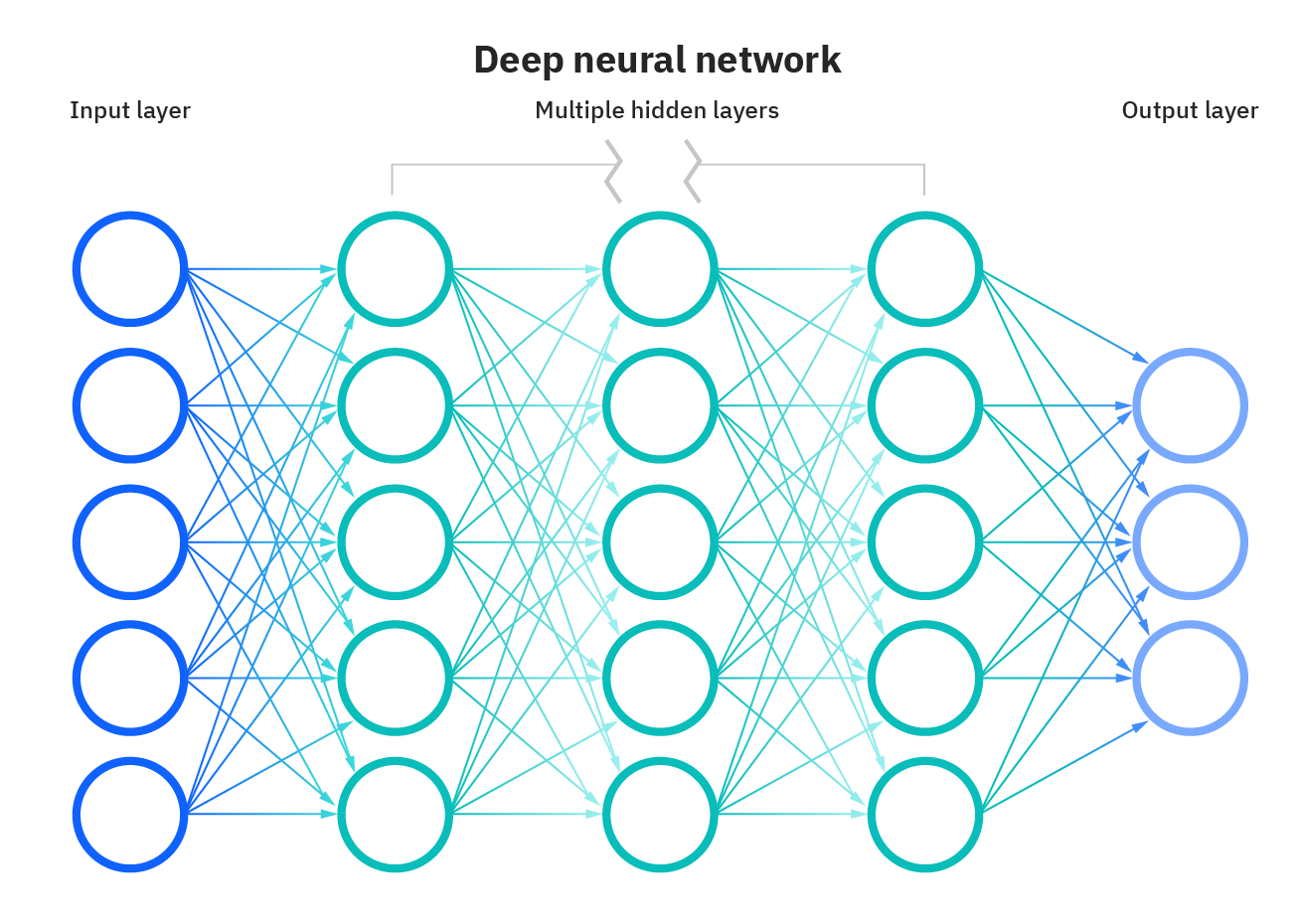
Step 3: Sort through the data.

The data is separated into sections based on the geographic location of the transaction. Data that has been hot encoded is the fifth step.

Step 6: Create a model and evaluate the simulation results.

**4.2.4 Neural networks**

Neural networks, sometimes referred to as neural artificial networks or neural simulated networks (NNSs), are a subset of machine learning that serve as the foundation for strong algorithms. They are also known as artificial neural networks (ANNs) or neural networks that imitate (SNNs). They got their name and structure from the human brain, and they're built to communicate like genuine neurons. Artificial neural networks (ANNs) consist of numerous layers of nodes, each with its own input, hidden layer, and output layer. It's a network of artificial neurons, each with its own weight and threshold. If the output of a node reaches a predetermined threshold, the node becomes active and data is transmitted to the next network. layer from there. If this is not the case, no data is transmitted to the next network hierarchy layer.(Bonaccorso, 2017)



**Working of Neural networks**

Neural networks are a machine learning technology that uses samples of prior tasks to educate a computer how to finish a task. The examples were usually manually marked in front of the class. When an object recognition system gets thousands of label photographs of autos, houses, coffee cups, and other similar items over time, visual patterns in images that are regularly associated with certain labels are searched for. This is referred to as object recognition. A neural network is a computer network that is loosely modelled on humans and is made up of tens of thousands, if not millions, of densely connected fundamental processing nodes. Layer links connect one node to multiple nodes in the underlying layer which it uses to gather data, and a number of nodes in the layer above it to which it sends it. It's called "weight" to identify which node is receiving which connections to that connection. When the network is operational, each node receives a new data item — a specific number — and multiplies the weight of that data item. The results are then added together to generate a single numerical value. If this number goes below a specific threshold, the node does not transfer data to the next level. When the number reaches a certain threshold, the "fires" node is activated, which usually means that all outgoing connections to all of its neighbors are delivered to the total amount of weighted inputs in today's neural networks. All of the weights and thresholds in a neural network are set to random values when it is first trained. To begin, the input layer receives training data, which is multiplied and added in a variety of ways until it reaches the output layer, where it is profoundly modified. Weights and thresholds are continually changed during training until the outputs produced by training data with the same labels are identical to the actual data.(Burrell and Society, 2016)

**Mathematics**

Machines have always been available to us since the beginning of the Industrial Revolution. They have made some development throughout millennia, and they have suffered a wide spectrum of aberrations since their emergence on the scene. However, one component has remained consistent, notably the explicit reliance on human minds in applying the standards. For the last few years, researchers have been investigating the possibility of implanting intelligence into robots. As a result, artificial intelligence and machine learning have developed as new fields of research. My earlier blog post, "A Brief Introduction to Term Machine Learning," takes an in-depth look at the subject.(Domingos and Hulten, 2001)

Neural networks have been an important aspect of Artificial Intelligence in recent years. The blog post "Neural Networks: An Art to Imitate Human Brain" delves into the definition and function of neural networks in great detail. Throughout this section, I will explain the mathematics underlying various network designs in detail. I feel it is vital to comprehend exactly what these networks are and how they function, which needs a thorough understanding of mathematics. It will surely allow you to experiment with numerous aspects in order to reach the desired results.(Crisci et al., 2012)



**Implementation**

It is based on the biological network of brain connections that ANNs were developed. Neural networks aid in problem solving by eliminating the need to programmed problem-specific rules or conditions. They are generic models that conduct the most difficult mathematical operations in a Blackbox environment. Among the many types of neural networks available are neural network development, recurrent neuro-networks, feedforward neural networks, multilayer perceptron’s, and many others. The first section of this chapter provides an overview of neural network implementation. Neural networks consist of three basic layers. There are hidden layers in the data layer. In accordance with the findings of Crisci et al.

As a result, the input layer, which contains neurons for inputting functions, has a higher output layer than the subsequent layers. Additionally, a bias was added to the input layer as a means of rebalancing the system. As a result, if n features are present, the input layer will have n+1 neurons.

There are levels that are hidden beneath the surface. Circuit input and output layers are separated by hidden layers. You can have as many hidden levels as you like because there are no limits. There are several layers to deep neural networks (DNNs), therefore they can manage a large quantity of data. Layer 2 neurons get input from Layer 1 and generate output for Layer 2. Layer 3 neurons receive input from Layer 2 neurons and generate output for Layer 3.

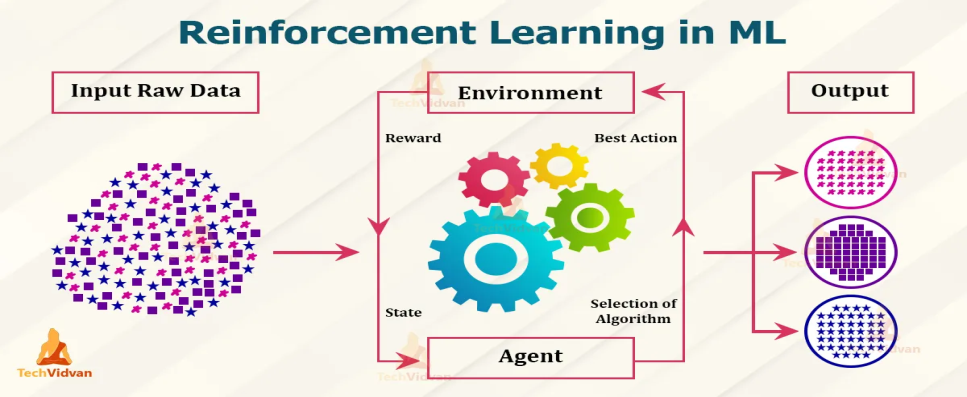
The output layer, which has the same number of neurons as the number of output classes, is the third layer. There are as many neurons in the network as there are classes if the task is a classification problem with multiple classes. A single neuron is all that's within, and it's responsible for binary categorization. In the year 2020, Mahash and others will have published a book.

**How all of the above solve the problem and which algorithms are more effective to grow the business.**

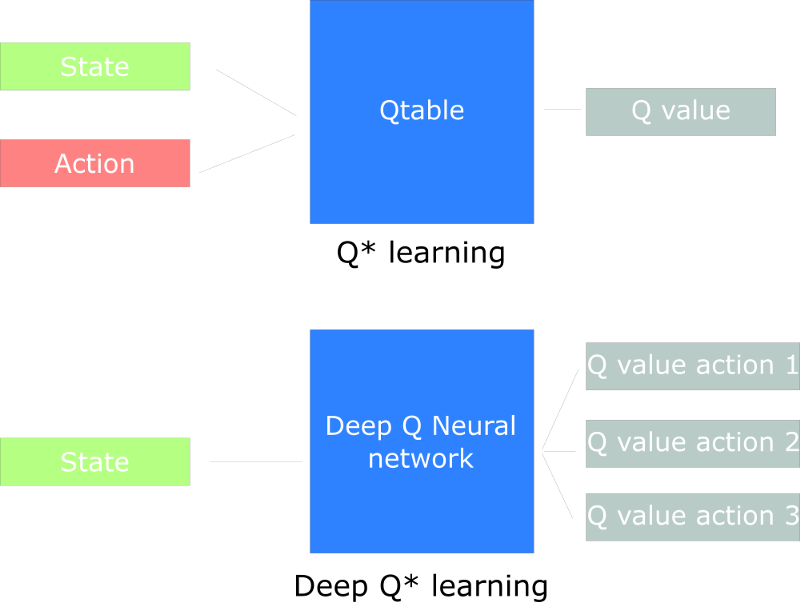
K-means is one of the most basic unattended learning algorithms accessible, and it is frequently used to tackle the well-known clustering problem. The method employs a straightforward and simple method of classifying a batch of data by employing a preset number of clusters (we assume k clusters). The fundamental idea is to choose k centres, one for each cluster in the data set. Because different locations produce varied results, it is vital that these centres be strategically located. As a result, they should be kept as isolated from one another as feasible. The next stage is to link up all of the data collection points with the nearest hub. No comments mean the first stage is complete and an early grouping is finished. We now have to recalculate k new centroids to act as barycenter’s for the clusters that formed earlier. Once we have these k new centres, we'll need to establish a new connection between the existing data sets and the new center that's closest to them. A loop was consequently created. The k centroids' positions change incrementally until the loop undergoes further adjustments. Therefore, centroids have ceased to move. Clustering mechanically splits the dataset into groups based on the similarities between them. Anomaly detection can uncover unexpected data objects in your collection. It's a fantastic thing when you uncover a scam. Association mining identifies recurring groups of things in your dataset. Models with latent variables are often used for data preparation.. (Gianfrancesco et al., 2018).

**4.3 Reinforcement Learning**

It is possible to think of an agent as a reinforcement learning algorithm.In exchange for good performance, the agent earns incentives and incurs penalties for poor performance. By maximizing its reward and reducing its punishment, the agent learns on its own, without the assistance of a human teacher. A form of dynamic programming, it uses a system of rewards and penalties to teach algorithms. Convolutional networks can be used in reinforcement learning to detect an agent's state when the input is visual, such as Mario's screen or the landscape in front of a drone. That is to say, they go about their normal business of image recognition. This machine learning model is being trained to make a series of judgments. The models try to accomplish uncertain, potentially complex environment. To solve the challenge trial and error was utilized by the computer. The artificial intelligence will be rewarded if the machine learning algorithm works as intended by the programmer or if this algorithm is not working properly then this AI gives the penalties as well for the actions it performs. Its depends upon the model How to conduct the problem statement assignment to maximize the reward, starting randomly, and advancing to intricate tactics, as well as the development of superhuman talents. A person who uses reinforcement learning learns from an interactive environment through trial and error, based on feedback from their own actions and experiences. These algorithms learn from the consequences of its actions this reinforcement learning agent receive signal in a numerical rewards. Moreover, Reinforcement learning, after supervised and unsupervised learning, is the third major machine learning type. Agent, environment, condition, action, and reward are all key parts of it. Reinforcement learning objective is to maximize reward while minimizing risk by utilizing the environment. Periodically improving the Reinforcement learning algorithm (called the agent) by investigating the environment and passing through the various potential states is a goal of Reinforcement learning. So that the agents can function at their best, they'll automatically decide what the optimal conduct is.



Q learning is a value-based approach to delivering data that helps agents decide what action to do next.

****

**Chapter 5 - Result**

Result

Learning algorithms based on prior experience can predict data trends. Machine Learning is a branch of science that includes both robotics and data analysis (the equivalent of thinking on the computer). These algorithms discover predictable, recurrent patterns that can be used to a wide range of sectors, including e-commerce, data management, and upcoming technologies such as self-driving cars. Machine Learning is only now beginning to have a significant impact, and it has the potential to significantly change both the way products are manufactured and how people live.

Machine Learning Algorithms are used to train enormous amounts of data, allowing the "robot" to learn and predict problems and patterns in future data sets. The Curiosity rover on Mars uses machine learning to navigate the Martian surface, and NASA thinks the same approach will be employed in driverless cars in the future. In the business realm, trend forecasting and analytics are used to forecast changes in purchasing behaviour. Machine learning algorithms are utilised to generate significantly better projections than existed before to the creation of the algorithm.

**Chapter 6 - Conclusion**

6.1 Conclusion

This concludes the machine learning series. Machine learning is a rapidly growing area of study in computer science. Machine learning has applications in almost every other field of research and is now widely employed because it can solve problems that would otherwise be too complex or time consuming to solve. The application of a range of models to interpret data patterns and create accurate predictions based on observable patterns is referred to as machine learning.

I began by discussing the concepts of generalization and overfitting. These two subjects are connected to supervised learning, a sort of training that makes use of training data. When a machine learning model can accurately anticipate data findings that it has never seen before, this is referred to as generalization. Overfitting occurs when a model is unable to generalize the learnt knowledge and learns the training data too well. Failure can come from supervised learning, which is the inverse of overfitting. Because of the model's inefficiency, accurate predictions using both training and new data are hard to make. This is referred to as overfitting. After that, I went on to datasets. Data is divided into three sections when utilising supervised learning: the train dataset, the development dataset, and the test dataset. The train data set is used to train the model, which then forecasts future events. The dev dataset is used to test the model during the model's development, but it is not used during training. After building the model, the test dataset must be used to see how it reacts to data it has never seen before. I also talked about how to select the most relevant fields from a dataset. For whatever reason, information is occasionally irrelevant and should not be included in a dataset. Following that, I examined artificial neural networks, which were the first model examined in this series of blog postings. To connect the layers, vectors are employed. Neural networks were among the first learning models to be devised, and other forms of neural networks have been explored throughout history. Deep neural networks are the next task on my agenda. Deep neural networks differ from artificial neural networks in that they include numerous hidden layers, whereas artificial neural networks only have one. Deep neural networks outperform low neural networks for certain tasks, owing in part to the complexity of contributing to the model via numerous hidden layers. However, as their complexity grows, they become increasingly difficult to train.

6.2 Limitation

There are several potential advantages to employing machine learning techniques, as well as numerous prospective uses. However, there are also numerous dangers and obstacles that will influence whether and how broadly the advantages are realized. These aren't problems with the algorithm itself; rather, they're with how it's being used and what it's supposed to accomplish.

Considering assumptions is essential for achieving better results. Various machine learning techniques might arrive at vastly different results, even when given excellent quality, impartial data. Variations in the methods used to create a model can have a major impact on whether or not a meaningful connection is found. However, regardless of the machine learning technique employed, data analysis will always necessitate some challenging subjective judgments regarding the most appropriate models or variables to use.

**Collaboration between computers and people**

Algorithms for machine learning are rarely programmed to provide a rationale for their decisions or outputs. There is a general apprehension about machine learning as a decision-making tool because of this notion. Physicians and politicians, for example, need to know exactly why a choice was made; simply relying on the algorithm's purported excellence will not do. Particularly if the system is vulnerable to mistakes or if it is unclear why a certain model was chosen. Humans obviously have a higher level of confidence in one another than in computers, but this aversion to trusting sophisticated learning systems poses a significant barrier to their true ability.

**Capabilities and knowledge**

Using these approaches effectively requires training personnel at all levels who are well-versed in algorithmic processes and who can appropriately deploy and examine their results. The 'Crowdsourcing Analytics' experiment shows how important it is to have humans who can comprehend the machine learning process and will not just accept its results as true. In many situations, a different sort of data scientist will be required, one who does not have the fundamental technical competence to create code but has enough of a broad knowledge of what can and cannot be done using machine learning techniques to properly assess its outputs. This 'type II' data scientist may not require a deep grasp of the code, but he or she may be in charge of a group of data scientists and must be able to communicate between the business or policy challenge and the technological environment.

**Change and regulation occur at lightning speed.**

With the rapid advancements in machine learning, it will be difficult to maintain a proper grasp of capabilities and limitations over the long term. It will become increasingly challenging to ensure that regulation can keep pace with these advancements without impeding progress.

**Lack of proper data**

To begin producing effective results, many machine learning algorithms need a significant quantity of training data. A neural network is an excellent illustration of this. Neural networks are data-hungry devices that want a colossal amount of input data. The more complex the design, the more data is required to generate usable results. Reusing data is a terrible idea, and adding additional data might be beneficial in some cases, but having more data is always the better option.

6.3 Recommendations for further study

These methods are quite useful and will grow in importance over time. Despite their prowess, these systems can't function without human input. To properly assess the machine learning algorithm's results, you'll need individuals with the appropriate degree of expertise and understanding. Better user-oriented design and the ability to probe the model underlying the outputs might be enormously useful to prevent these systems from being seen as a black box. Having a comprehensive grasp of what these systems can and cannot accomplish will become increasingly crucial as their acceptance and fast developments rise. As a world leader in applying these approaches effectively to public policy, the UK government has a chance to leverage the vast local knowledge. Understanding how to use these systems successfully for policy would be a great advantage, especially in identifying where regulation might be required and what it might look like. One of the most essential initial tasks is to figure out where regulation is needed, how new technologies will affect society and how to deal with the exponential increase in capabilities. Many of the potential problems and perils of employing machine learning algorithms are well-known in applications such as autonomous vehicles and high-frequency trading, but developing frameworks to address these issues is not simple. It will need much deliberation and experimentation. In a broader sense, concerns like software behavior verification or responsibility may necessitate regulation, although the relevance of regulation in many cases is uncertain. Foresight exercises and more focused talks between academics, industry, and government would be a useful first step in dealing with these problems. They may be utilized to solve a wide range of issues and provide insightful results when done correctly. Unfortunately, machine learning techniques are sometimes viewed as a magic bullet that can fix all of a company's issues overnight. Machine learning is a collection of algorithms and data processing, not some sort of magical artifact. As these methods become more commonly used, a dependence on technology or a misunderstanding of its capabilities might have catastrophic repercussions. To help individuals grow better at what they do, these computational systems are built to be a tool, not a replacement. As long as technology and knowledgeable humans work together, we'll be OK.

**References**

AYYADEVARA, V. K. J. A. B., CA, USA 2018. Pro machine learning algorithms.

BONACCORSO, G. 2017. *Machine learning algorithms*, Packt Publishing Ltd.

BONACCORSO, G. 2018. *Machine Learning Algorithms: Popular algorithms for data science and machine learning*, Packt Publishing Ltd.

BRADLEY, A. P. J. P. R. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. 30**,** 1145-1159.

BURRELL, J. J. B. D. & SOCIETY 2016. How the machine ‘thinks’: Understanding opacity in machine learning algorithms. 3**,** 2053951715622512.

CRISCI, C., GHATTAS, B. & PERERA, G. J. E. M. 2012. A review of supervised machine learning algorithms and their applications to ecological data. 240**,** 113-122.

DOMINGOS, P. & HULTEN, G. A general method for scaling up machine learning algorithms and its application to clustering. In proceedings of the eighteenth international conference on machine learning, 2001. Citeseer.

FATIMA, M., PASHA, M. J. J. O. I. L. S. & APPLICATIONS 2017. Survey of machine learning algorithms for disease diagnostic. 9**,** 1.

GIANFRANCESCO, M. A., TAMANG, S., YAZDANY, J. & SCHMAJUK, G. J. J. I. M. 2018. Potential biases in machine learning algorithms using electronic health record data. 178**,** 1544-1547.

JORDAN, M. I. & MITCHELL, T. M. J. S. 2015. Machine learning: Trends, perspectives, and prospects. 349**,** 255-260.

KHAN, A., BAHARUDIN, B., LEE, L. H. & KHAN, K. J. J. O. A. I. I. T. 2010. A review of machine learning algorithms for text-documents classification. 1**,** 4-20.

MAHESH, B. J. I. J. O. S. & ., R. 2020. Machine Learning Algorithms-A Review. 9**,** 381-386.

MOHAMMED, M., KHAN, M. B. & BASHIER, E. B. M. 2016. *Machine learning: algorithms and applications*, Crc Press.

NGIAM, K. Y. & KHOR, W. J. T. L. O. 2019. Big data and machine learning algorithms for health-care delivery. 20**,** e262-e273.

RAY, S. A quick review of machine learning algorithms. 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon), 2019. IEEE, 35-39.

SCHAPIRE, R. J. P. U. 2015. Machine learning algorithms for classification. 10.

SHARMA, D., KUMAR, N. J. I. J. O. A. R. I. C. E. & TECHNOLOGY 2017. A review on machine learning algorithms, tasks and applications. 6**,** 1548-1552.

SILVER, D. L., YANG, Q. & LI, L. Lifelong machine learning systems: Beyond learning algorithms. 2013 AAAI spring symposium series, 2013.

SINGH, A., THAKUR, N. & SHARMA, A. A review of supervised machine learning algorithms. 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), 2016. Ieee, 1310-1315.

SNOEK, J., LAROCHELLE, H. & ADAMS, R. P. J. A. I. N. I. P. S. 2012. Practical bayesian optimization of machine learning algorithms. 25.

STALLKAMP, J., SCHLIPSING, M., SALMEN, J. & IGEL, C. J. N. N. 2012. Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. 32**,** 323-332.

STEINKRAUS, D., BUCK, I. & SIMARD, P. Using GPUs for machine learning algorithms. Eighth International Conference on Document Analysis and Recognition (ICDAR'05), 2005. IEEE, 1115-1120.

UDDIN, S., KHAN, A., HOSSAIN, M. E., MONI, M. A. J. B. M. I. & MAKING, D. 2019. Comparing different supervised machine learning algorithms for disease prediction. 19**,** 1-16.

WATSON, D. S., KRUTZINNA, J., BRUCE, I. N., GRIFFITHS, C. E., MCINNES, I. B., BARNES, M. R. & FLORIDI, L. J. B. 2019. Clinical applications of machine learning algorithms: beyond the black box. 364.

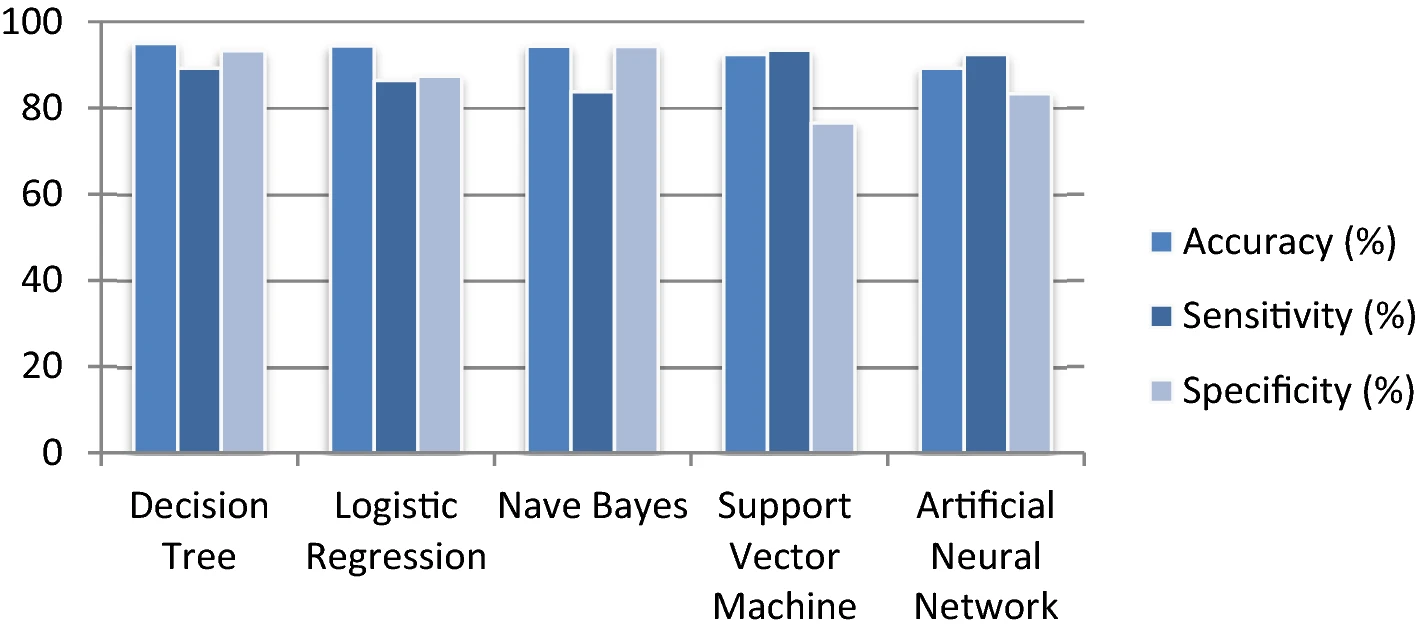
WETTSCHERECK, D. 1994. A study of distance-based machine learning algorithms.

WILLIAMS, N., ZANDER, S. & ARMITAGE, G. J. A. S. C. C. R. 2006. A preliminary performance comparison of five machine learning algorithms for practical IP traffic flow classification. 36**,** 5-16.

XIAO, H., RASUL, K. & VOLLGRAF, R. J. A. P. A. 2017. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms.

Appendix A: Maxico\_covid19 Analysis and implementation of Machine Learning Algorithms

Dataset: <https://www.kaggle.com/marianarfranklin/mexico-covid19-clinical-data/metadata>



**EDA**



**Sample ML website with Coding**

Web sample overview

****

**Code**

****