Need For Machine Learning

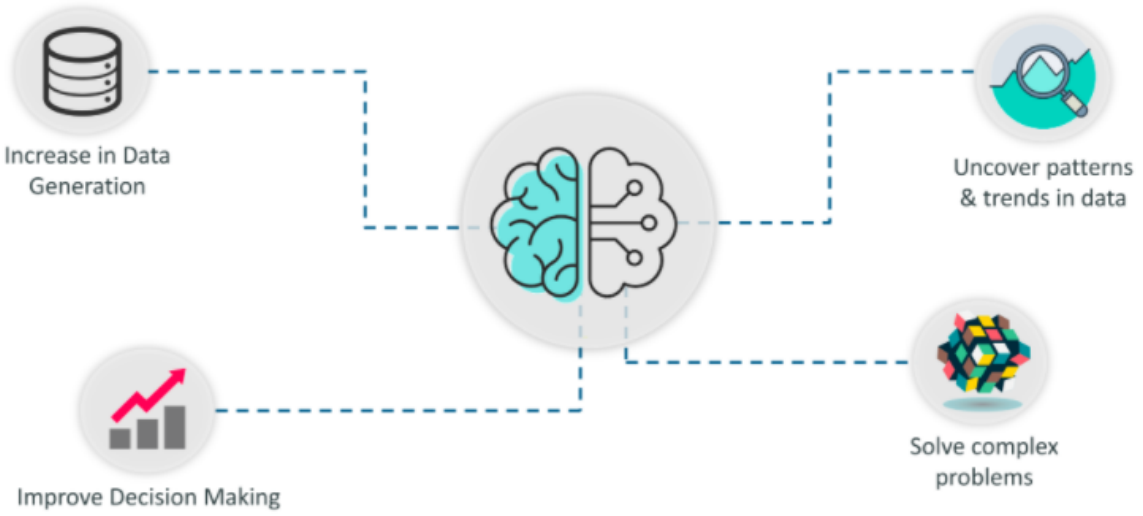
We have been generating an unfathomable amount of data since the technological revolution. According to study, we generate approximately 2.5 quintillion bytes of data per day! It is anticipated that by 2020, every person on the planet would generate 1.7MB of data each second.

With so much data available, it is finally possible to construct predictive models that can study and analyse complex data in order to identify relevant insights and give more accurate outcomes.

Companies like Netflix and Amazon create such Machine Learning models by analysing massive amounts of data in order to uncover profitable possibilities and prevent risky situations.

Here are some of the reasons why Machine Learning is so important:

- Increase in Data Generation: Due to the excessive production of data, we require a strategy for structuring, analysing, and drawing relevant insights from data. This is where Machine Learning enters the picture. It employs data to solve problems and discover solutions to the most difficult tasks that businesses confront.
- Improve Decision Making: Machine Learning can be utilised to create better business decisions by utilising various algorithms. Machine Learning, for example, is used to estimate sales, forecast stock market declines, identify dangers and abnormalities, and so on.



- Discover patterns and trends in data: The most important aspect of Machine Learning is discovering hidden patterns and extracting critical insights from data. Machine Learning, by constructing predictive models and employing statistical approaches, enables you to go under the surface and explore data on a microscale. Manually understanding data and extracting patterns will take days, whereas Machine Learning algorithms can do so in less than a second.

- Machine Learning can be used to address the most complex issues, from finding the genes connected to the deadly ALS disease to developing self-driving cars.

To give you a clearer idea of how vital Machine Learning is, consider the following Machine Learning Applications:

- Netflix's Recommendation Engine: Netflix's notorious recommendation engine is at the heart of the service. Netflix recommends over 75% of what you view, and these recommendations are created using Machine Learning.

- Facebook's Auto-tagging feature: Machine Learning and Neural Networks are at the heart of Facebook's DeepMind face verification engine. DeepMind analyses facial features in images to identify your friends and relatives.

- Amazon's Alexa: Alexa, which is based on Natural Language Processing and Machine Learning, is a high-level Virtual Assistant that does more than just play tunes from your playlist. It can book you an Uber, link to your home's other IoT gadgets, track your health, and so on.

- Gmail's Spam Filter: To filter out spam communications, Gmail uses Machine Learning. It analyses emails in real-time and classifies them as spam or non-spam using Machine Learning techniques and Natural Language Processing.

Introduction To Machine Learning

Arthur Samuel popularised the phrase "machine learning" in 1959. Looking back, that was most likely the most momentous year in terms of technological breakthroughs.
If you search the term "machine learning" on the internet, you will find at least 100 different definitions. Tom M. Mitchell, on the other hand, provided the very first formal definition:

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

In layman's words, machine learning is a subset of artificial intelligence (AI) that gives machines the ability to learn automatically and improve based on experience without being expressly programmed to do so. It is, in this sense, the discipline of teaching Machines to solve problems by giving them the ability to reason.

But, wait a minute, can a machine think or make decisions? If you provide a machine enough data, it will learn how to understand, process, and analyse that data using Machine Learning Algorithms to solve real-world problems.

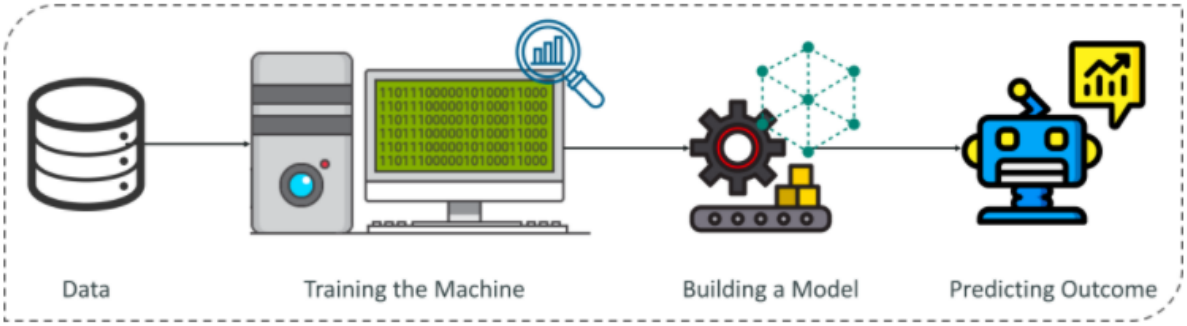
Before we go any further, let's go through some of the most often used Machine Learning terms.

Machine Learning Definitions

Algorithm: A Machine Learning algorithm is a set of rules and statistical techniques used to discover patterns from data and extract meaningful information from it. It is the reasoning that underpins a Machine Learning model. The Linear Regression algorithm is an example of a Machine Learning algorithm.
Model: A model is the most important component of Machine Learning. A Machine Learning Algorithm is used to train a model. In order to produce the proper output, an algorithm maps all of the decisions that a model is expected to make based on the provided input.
Predictor Variable: A data feature(s) that can be utilised to forecast the output.
Response Variable: The response variable is the characteristic or output variable that must be predicted using the predictor variable (s).

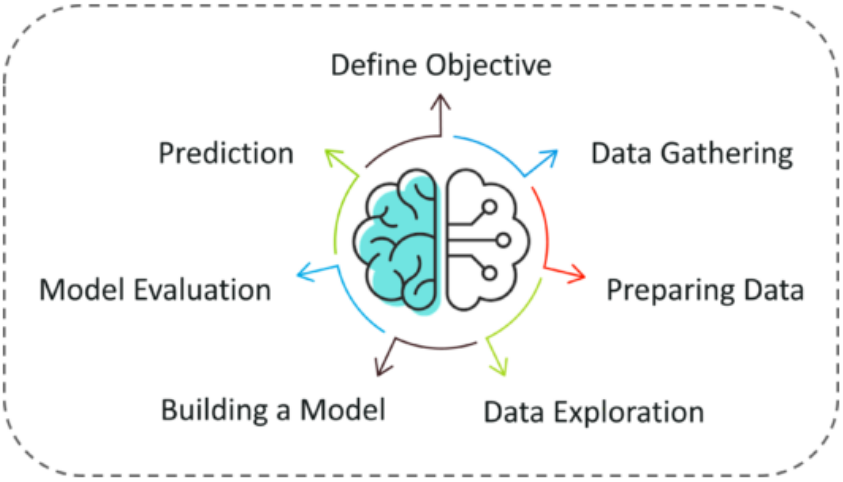
Training Data: The training data is used to build the Machine Learning model. The training data assists the model in identifying key trends and patterns that are required to forecast the output.

Testing Data: Once the model has been trained, it must be tested to see how well it predicts outcomes. The testing data set accomplishes this.



Machine Learning Process

The Machine Learning technique entails creating a Predictive model that can be utilised to solve a Problem Statement. Assume you've been given an issue that needs to be solved using Machine Learning to better grasp the process.



The challenge is to use Machine Learning to anticipate the occurrence of rain in your local area.

In a Machine Learning process, the following stages are taken:

Step 1: Define the Problem Statement's goal.

At this point, we must determine exactly what needs to be projected. In our situation, the goal is to forecast the possibility of rain by researching meteorological patterns. At this point, it is also critical to make mental notes on what type of data may be used to address the problem or the route that must be taken to arrive at a solution.

Step 2: Data Collection

At this point, you should be asking questions like,

What type of information is required to address this problem?

Is the information available?

How can I obtain the information?

Once you understand the types of data that are required, you must understand how to obtain this data. Data can be collected manually or by web scraping. However, if you are a newbie trying to learn Machine Learning, you do not need to worry about obtaining data. There are thousands of data resources available on the internet; simply download the data set and get started.

Returning to the original issue, the data required for weather forecasting comprises measurements such as humidity level, temperature, pressure, locale, whether or not you live in a hill station, and so on. Such information must be gathered and preserved in order to be analysed.

Step 3: Gathering Data

The information you gathered is nearly never in the correct format. There will be many irregularities in the data set, such as missing values, redundant variables, duplicate values, and so on. It is critical to eliminate such inconsistencies since they can lead to incorrect computations and forecasts. As a result, at this point, you check the data set for discrepancies and correct them right away.

Step 4: Analyze Exploratory Data

Put on your detective hat because this stage is all about delving deep into data and uncovering all the hidden data riddles. EDA, or Exploratory Data Analysis, is the Machine Learning

brainstorming stage. Understanding the patterns and trends in data is what data exploration entails. At this point, all of the useful insights have been derived, and the relationships between the variables are understood.

For example, in the instance of predicting rainfall, we know that if the temperature has dropped below a certain threshold, there is a good chance of rain. At this point, such relationships must be understood and mapped.

Step 5: Construct a Machine Learning Model

The Machine Learning Model is built using all of the insights and patterns discovered during Data Exploration. This stage always starts with dividing the data set into two parts: training data and testing data. The model will be built and analysed using the training data. The model's reasoning is based on the Machine Learning Algorithm that is being used.

In the case of predicting rainfall, we can use a Classification Algorithm such as Logistic Regression because the result will be in the form of True (if it will rain tomorrow) or False (if it will not rain tomorrow). Choosing the proper method is dependent on the sort of problem you're attempting to answer, the data set, and the problem's level of complexity. In the following sections, we will go over the many sorts of problems that Machine Learning can handle.

Step 6: Model Evaluation and Optimization

It is finally time to put the model to the test once it has been built using the training data set. The testing data set is used to assess the model's efficiency and accuracy in predicting the outcome. Any further improvements to the model can be introduced once the accuracy has been calculated. To increase the model's performance, methods such as parameter adjustment and cross-validation can be applied.

Step 7: Predictions

The model is finally utilised to make predictions when it has been evaluated and modified. The final result can be a categorical variable (for example, True or False) or a continuous quantity (eg. the predicted value of a stock).

In our situation, the result will be a categorical variable for predicting the occurrence of rainfall.

That was the end of the Machine Learning process. It is now time to study about the various methods in which Machines can learn.

Machine Learning Types

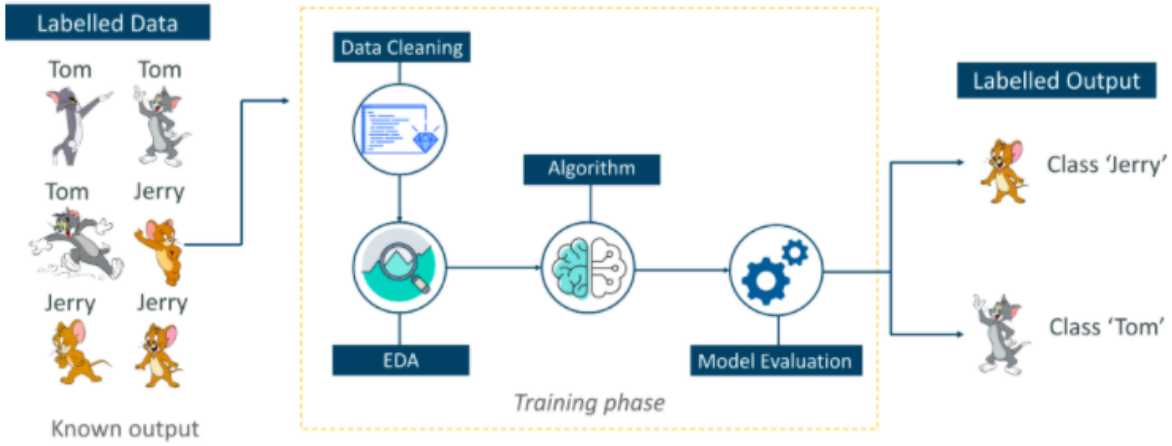
A machine can learn to solve a problem by using any of the three ways listed below.

1. Supervised Learning
2. Unsupervised Learning
3. Reinforcement Learning

Supervised Learning

Supervised learning is a technique in which we instruct or train the machine using labelled data.

Consider the following analogy to better understand Supervised Learning. When we were kids, we all required help with math issues. Our teachers assisted us in understanding what addition is and how it is performed. Similarly, consider supervised learning to be a subset of Machine Learning that includes a guide. The labelled data set is the teacher who will teach you how to recognise patterns in data. The labelled data set consists entirely of the training data set.

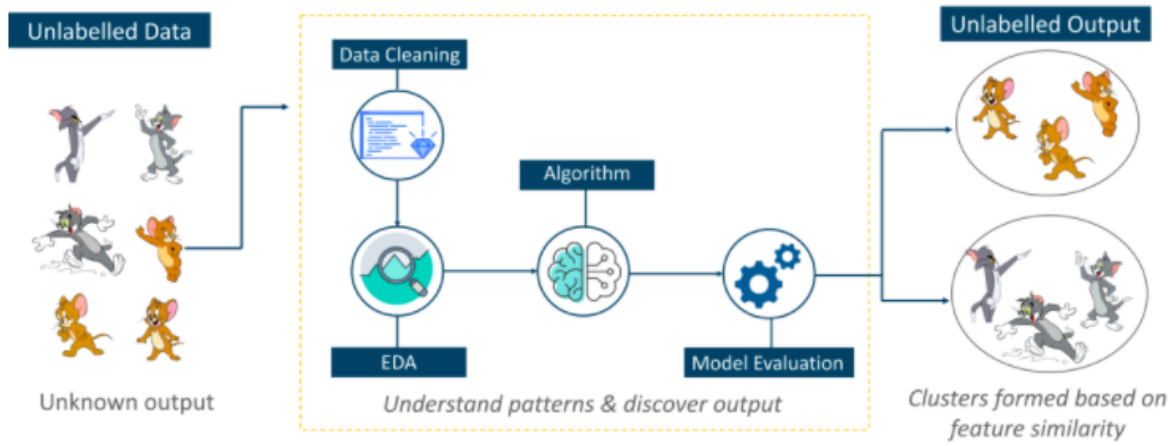


Consider the diagram above. We're providing photographs of Tom and Jerry to the computer, and the goal is for the machine to detect and classify the images into two groups (Tom images and Jerry images). The training data set that is provided to the model is labelled, as in, we tell the machine, 'this is how Tom looks, and this is Jerry.' You are training the system by using labelled data in this manner. There is a well-defined training phase in Supervised Learning that is carried out with the use of labelled data.

Unsupervised Learning

Unsupervised learning is training with unlabeled input and then letting the model to act on that knowledge without supervision.

Consider unsupervised learning to be a smart child who learns without supervision. The model in this sort of Machine Learning is not fed with labelled data, thus the model has no idea that 'this image is Tom and this is Jerry,' instead it works out patterns and differences between Tom and Jerry on its own by ingesting massive amounts of data.



For example, it recognises prominent traits of Tom, such as pointy ears, larger stature, and so on, in order to determine that this image is of type 1. Similarly, it detects such characteristics in Jerry and determines that this image is of type 2. As a result, it divides the photographs into two categories without knowing who Tom or Jerry is.

Reinforcement Learning

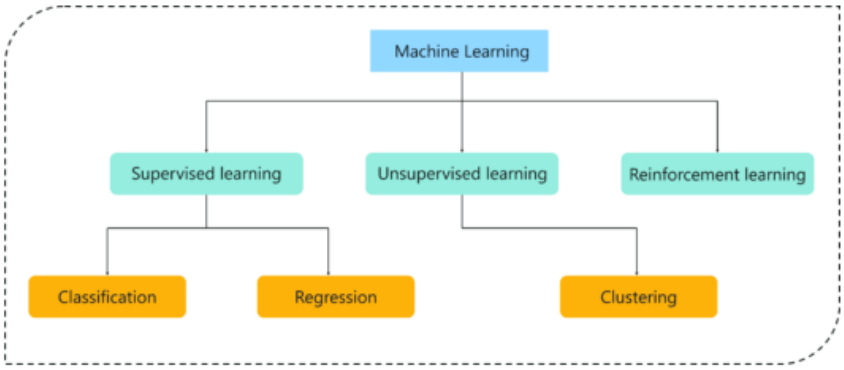
Reinforcement Learning is a subset of machine learning in which an agent is placed in an environment and learns how to behave in that environment by executing certain actions and watching the rewards that result from those actions.

This type of Machine Learning is somewhat unique. Assume you've been left off on a remote island! What would you do in this situation?

Panic? Of course, at first, we all would. However, as time goes on, you will learn how to live on the island. You will investigate the ecosystem, learn about the climate, the types of food that grow there, the island's threats, and so on. This is exactly how Reinforcement Learning works; an Agent (you, trapped on the island) is placed in an unknown environment (island), where he must learn by observing and performing activities that result in rewards.

Reinforcement Learning is most commonly employed in advanced Machine Learning applications such as self-driving cars, AlphaGo, and so on.

Type Of Problems In Machine Learning



Consider the diagram above: there are three major categories of problems that Machine Learning can solve:

1. **Regression:** The outcome in this sort of problem is a continuous quantity. For example, predicting the speed of an automobile given the distance is a Regression problem. Supervised Learning algorithms such as Linear Regression can be used to tackle regression problems.

2. **Classification:** The output of this type is a category value. Classifying emails into spam and non-spam categories is a classification problem that can be tackled using Supervised Learning classification techniques such as Support Vector Machines, Naive Bayes, Logistic Regression, K Nearest Neighbor, and so on.

3. **Clustering:** In this sort of challenge, the input is divided into two or more clusters based on feature similarity. Unsupervised Learning methods such as K-Means Clustering, for example, can be used to classify viewers into comparable groups based on their hobbies, age, region, and so on.

Here's a table that sums up the difference between Regression, Classification, and Clustering.

Regression	Classification	Clustering
• Supervised Learning • Output is a continuous quantity • Main aim is to forecast or predict • Eg: Predict stock market price • Algorithm: Linear Regression	• Supervised Learning • Output is a categorical quantity • Main aim is to compute the category of the data • Eg: Classify emails as spam or non-spam • Algorithm: Logistic Regression	• Unsupervised Learning • Assigns data points into clusters • Main aim is to group similar items clusters • Eg: Find all transactions which are fraudulent in nature • Algorithm: K-means