**Review Paper 2**

**Big Data**

# Big Data

* A Collection of large and complex datasets which are difficult to store and process using the traditional database and data processing tools is considered as big data.
* Big data is collected from traditional and digital sources which, when refined properly can be used for research and analysis.
* Businesses develop with time, and as a result, the amount of data created by these organizations grows tremendously.

# Where is the Big Data Coming from?

* **Social Media:** Big data corporations such as Facebook and Google collect data from our daily actions. YouTube, Twitter, LinkedIn, Snapchat, WhatsApp, Instagram, Discord, Tik Tok, and more platforms are examples.
* **Public Web:** This incorporates information from Wikipedia, health-care services, the World Bank, the government, weather, and traffic, among other sources.
* **Archives:** This includes medical records, client communication, insurance papers, scanned documents, and other types of data.
* **Doc:** Big data may be found in a variety of formats, including HTML, CSV, PDF, XLS, Word, XML, and others.
* **Media:** Images, video, audio, a live feed, and a podcast.
* **Data Storage:** The different data set and record frameworks which are utilized to store the information fill in as the hotspot for enormous information.
* **Machine Log Data:** Server data, application logs, audit logs, CDR-call detail records, various mobile apps, mobile location, and so on.
* **Sensor Data:** Sensor data from medical gadgets, road cameras, satellites, traffic surveillance devices, video games, home appliances, air conditioning systems, and office buildings.
* **IoT Appliances:** Smart TVs, smart washing machines, smart coffee machines, smart air conditioners, and other internet-connected products generate data for their smart functionality. Machine-produced data is generated by sensors in a variety of devices. Many records are saved as one of the sources of big data in e-commerce transactions, corporate transactions, banking, and the stock market. Payments made by credit card, debit card, or other electronic means are all stored on file as data. GPS in the car aids in the tracking of the vehicle's movement in order to reduce fuel and time consumption by shortening the route to a destination. This technology generates a lot of data on the location and movement of the vehicles.

# 5V’s of Big Data

Big data is defined by three variables: velocity, variety, and volume.

1. **Variety:** It refers to the organized, semi-structured, and unstructured data types. It can also refer to a variety of sources. Variety refers to the influx of data from new sources both inside and outside of an organization. It might be organized, semi-organized, or unorganized.
2. **Volume:** Volume is a gigantic measure of information. To decide the worth of information, size of information assumes an exceptionally urgent part. On the off chance that the volume of information is exceptionally huge, it is really considered as a 'Major Information'.
3. **Velocity:** The term "velocity" refers to the rapid collection of data. Data comes in at a high rate from machines, networks, social media, mobile phones, and other sources in Big Data velocity. A large and constant influx of data exists. This influences the data's potential, or how quickly data is created and processed in order to satisfy needs.
4. **Veracity:** It refers to data inconsistencies and uncertainty, i.e., accessible data can become untidy at times, and quality and accuracy are difficult to control. Because of the numerous data dimensions originating from several distinct data kinds and sources, Big Data is also volatile
5. **Value:** After considering the four V's, there is one more V to consider: Value! The majority of data with no value is useless to the organization until it is converted into something beneficial. Data is of no utility or relevance in and of itself; it must be turned into something useful in order to extract information. As a result, Value might be considered the most essential of the five V's.

# Big Data is Structured, Unstructured and Multi – Structured Data

* **Structured Data:** Structured data is data that has a specified format and is arranged according to a predetermined schema. Structured data may be found in traditional databases and repositories such as mainframes, SQL server, Oracle, DB2, Sybase, Access, Excel, txt, and Teradata. Only this type of data is dealt with by the Relational Database Management system.
* **Unstructured Data:** Unstructured data is data that is not well organized and difficult to comprehend using typical databases or data models. Chatter, text analytics, blogs, Tweets, comments, clicks, tags, and other social media data.
* **Multi – Structured Data:** Multi-structure information are un-demonstrated, it should be coordinated, in spite of the fact that there may be an outline yet it is overlooked. It tends to be gotten from connections among human and machines. This incorporates developing business sector information, online business, and other outsider information like climate, cash change, segment, board and so forth.

# Why is it necessary to process Big Data?

We can have information without data yet we can't have data without information. With such voluminous information comes the intricacy of overseeing it well with procedures that are compelling and human-accommodating as well as convey the ideal outcomes on time.

The meaning of Huge Information doesn't just spin around how much data an association has yet additionally the way in which an association utilizes the assembled data. Every association uses data according to their necessities; the more capably an association uses the data, the more encouraging are the odds of its success.

Huge Information has been a gift from heaven for organizations and businesses especially in the field of showcasing whose essential is to be refreshed with regards to the changing patterns and prospects in an economy occasionally. These days, structures like HADOOP, Multiview Bunching, Anomaly Recognition and Characterization Investigation are being considered over the customary procedures of processing Huge Information.

Big Data has become increasingly important in today's economic world. Looking at the factors given below can help us comprehend this:

* **Cost Savings:** Some Big Data solutions, such as Hadoop and Cloud-Based Analytics, help organizations save money when a large amount of data has to be stored, and these tools can assist in identifying more efficient ways to collaborate.
* **Time Reduction:** Because of the speed with which technologies like Hadoop and in-memory analytics can discover new sources of data, firms can quickly break down information and determine the best choice.
* **Comprehend the economic situations:** Big Data analysis provides a more complete view of the present economic situation. For example, by dissecting customers' purchasing habits, a company may identify the most popular things and distribute them in accordance with this trend. It will be able to outstretch its competitors as a result of this.

# Advantages of Big Data

* Errors inside the company are instantly identified.
* A higher conversion rate means more money in your pocket.
* Your adversary's strategy is quickly discovered.
* Extortion can be identified as soon as it occurs, and legal steps can be taken to mitigate the harm.
* The enhanced speed, capacity, and scalability of storage, as well as having the methods and tools to deal with the data more effectively, are the main attractions of Big Data.

With the advantages come the disadvantages. An end user's primary concern is privacy. As a result, the end user's credentials must be protected at all costs. Users must feel safe and secure in the knowledge that the information they access isn't being shared with third parties.

Another difficulty that Big Data encounters in the real world is timeliness. The larger the data sets, the longer the processing time, resulting in a delay in the decision-making process.

# Disadvantages of Big Data

* Over a period of time, data is gathered from every conceivable source. Because the data is raw and unreliable, it is more susceptible to noise.
* One of the major difficulties that Big Data is currently grappling with, particularly on the social media front, is security.
* The majority of the data a user seeks for analysis and interpretation is locked behind firewalls and private clouds, accessible only to those with the technical knowledge and experience to transform raw data into useful information.

Despite having a thorough understanding of the benefits and drawbacks of Big Data, a number of companies and organizations are eager to take on the task of extracting useful information from this massive volume of data. However, knowledge and experience in emerging tools and technologies do not appear to be sufficient to meet the demands of the end user and provide meaning to data. Here are some of the reasons why Big Data projects fail on a broad scale.

# Why most of the Big Data Projects fail?

* **The way Big Data is perceived by the masses:** Big Data is handled as if it has a set beginning and end point, whereas in reality it is a journey that leads to consistent data analysis and investigation. It may be utilized to deduce trends for future company success. However, in order to locate the answer, you need focus on the process rather than expecting a defined output. Big Data is a continuous investigation aimed at gaining relevant insights rather than reaching conclusions faster than ever before. The core of this data is discovered when it is used in a commercial environment; otherwise, it is just a massive amount of data.
* **Lack of skilled Data Scientist:** The inaccessibility of competent and trained data analysts is the key reason for the lack of adequate research of Big Data endeavors. To get the most out of Big Data, you'll need a lot more knowledge, skill, better adaptability, and longer timescales.
* **Cost cutting and Lack of Budget:** Emerging technologies can only be harnessed with the aid of appropriate tools and well-equipped systems. Businesses aren't always prepared to put a lot of money into enterprises that promise high returns over time. The need to make zillions in a short period of time limits the application of appropriate techniques and technology. As a result, these Big Data ventures are certain to fail.
* **No clarity of thought and poor strategy:** The end user should ask himself about the type of outcome he is concentrating on and the reason for it as the first and most important step in issue resolution. The solution to an issue might be ambiguous, presenting a wide variety of options that may mislead the user and divert him from his primary goal. As a result, in order to get the most out of Big Data, it's critical to create a clear issue description.

There are 800 million websites on the Internet that provide Big Data information. After Cloud Computing, Big Data is the next big thing. Big Data offers numerous opportunities in the fields of health, education, the environment, and business, but managing large volumes of data using standard approaches proves to be incredibly difficult.

**Quantum Computing**

# Quantum Computing

* Quantum computing is the method of accomplishing difficult and huge processes rapidly and effectively utilizing quantum physics. A Quantum computer is used to do Quantum calculations in the same way that classical computers are used to perform classical computations.
* Quantum computations are so difficult to accomplish that they are nearly impossible to do using traditional computers. The term 'quantum' comes from the physics concept of Quantum Mechanics, which defines the physical characteristics of electrons and photons.
* Quantum theory provides the foundation for fully explaining and comprehending nature. As a result, quantum computations have to cope with complexity. An area of Quantum Information Science is Quantum Computing. It explains how to deal with a difficult calculation in the most efficient way possible.
* It uses Qubits to store information. Quantum bits are another name for qubits. A quantum computer can simulate issues or processes that a traditional computer (such as the one we use today) cannot. Even a quantum computer can solve computational tasks more quickly than a traditional computer.
* For example, obtaining the product of (500 \* 187625) using a conventional computer is simple, while obtaining the same answer using a quantum computer is simple and rapid. A conventional computer will return a result in around 5 seconds, but a quantum computer will return a result in 0.005 seconds.

# Quantum Bits

* Quantum bits, or Qbits, are the quantum computer's storage unit. In a quantum computer, all information is stored in the form of qubits. Quantum bits are electrons or photons that make up subatomic particles.
* Qubit generation and management is complicated, and it is a demanding challenge for scientists working in this subject. These are the qubits that have the superposition and entanglement properties.
* It implies qubits can display a variety of 1 and 0 combinations at the same time. As a result, it is superposition.

# History of Quantum Computing

* Paul Benioff presented a quantum mechanical model of the Turing Machine in the early 1980s. Since that time, the notion of quantum computing has emerged. It was later proposed that a quantum computer could imitate things that a conventional computer couldn't. Richard Feynman and Yuri Manin came up with the idea.
* In 1994, Peter Shor created a quantum method for factoring integers. The algorithm was powerful enough to crack RSA-encrypted messages. In the subject of quantum computing, further research is continuously being done.
* On October 23, 2019, Google, in collaboration with NASA in the United States, issued a report claiming to have achieved Quantum Supremacy. Despite the fact that some of them have challenged this assertion, it is still a huge step forward in the field.

# Applications of Quantum Computing

* **Cybersecurity:** In today's digital age, personal information is saved on computers. As a result, we require a robust cybersecurity system to prevent data theft. Traditional computers are adequate for cybersecurity, but threats and assaults erode it.
* **Cryptography:** Quantum computers are also assisting in the development of encryption ways to safely transport packets onto the network in the realm of cryptography. Quantum cryptography is the name given to this type of encryption method development.
* **Weather Forecasting:** When utilizing traditional computers to forecast the weather, the analysis process might take too long. Quantum computers, on the other hand, have increased power to analyze, detect trends, and forecast the weather in a short amount of time and with more precision.
* **AI and Machin Learning:** Artificial intelligence (AI) is a rapidly growing area of digitalization. AI and machine learning have been used to create a variety of tools, apps, and features. As time passes, a growing number of apps are being developed. As a result, traditional methods have been challenged to match precision and speed.
* **Drug Design and Development:** Drug Designing and developing pharmaceuticals is a common task. It's because medication research is dependent on the trial and error process, which is both costly and hazardous. It's also a difficult problem for quantum computers. The researchers hope and believe that quantum computing will become a useful tool for understanding medications and their effects on humans.

# Future of Quantum Computing

* Quantum computing's future appears to be bright and promising for global trade. The facts described above indicate that this is only the beginning of the notion, which will undoubtedly become a part of our lives. It hasn't yet entered the mainstream.
* Quantum systems will help industry to address challenges that they previously felt were impossible to handle in the future. Quantum computing is expected to increase rapidly in the future decades, according to studies. Google is demonstrating a lot of attention and effort on quantum computing theory.
* Tensor Flow Quantum is a new version of Tensor Flow that was just released by Google (TFQ). TFQ is a free and open-source software library. It's utilized to make quantum machine learning models prototypes. When it's finished, it'll make it simple for programmers to design hybrid AI algorithms that combine quantum and conventional computing approaches.

# Why do we need Quantum Computers?

* Supercomputers aren't perfect for all issues.
* Supercomputers are used by scientists and engineers to solve challenging issues. These are extremely powerful traditional computers with thousands of CPU and GPU cores. Even supercomputers, however, have difficulty solving some issues.
* If a supercomputer becomes stumped, it's most likely because it was asked to handle a problem with a high level of complexity. When traditional computers fail, it's usually because of their complexity.
* Complex issues are those that include a large number of variables interacting in complex ways. Because of all the various electrons interacting with one another, modelling the behavior of individual atoms in a molecule is a difficult task. It's difficult to figure out the best routes for a few hundred tankers in a worldwide shipping network.

# Where are Quantum Computers used?

* **Chemistry:** In scientific inquiry, actual physical experiments and assessment of the outcomes are successful strategies. It appears that reenacting these tests on an old-style PC and speeding up the cycle without the need for actual trials is still a long way off. One of the first and most exciting applications of quantum computing will be in the realm of chemistry.
* **Meteorology:** Weather forecasts are difficult to produce because there are so many variables to consider. AI employing quantum computers will increase pattern recognition, making it easier to predict extreme climatic events and potentially save thousands of lives each year.
* **Financial Modeling:** The current corporate sectors are the most complex frameworks ever seen. There is no controlled environment in which to conduct the research. Quantum computing has been selected by investors and researchers to solve this problem. One immediate benefit is that quantum computers' inherent "irregularity" is compatible with financial business sectors.
* **Health Care:** Clinicians will be able to integrate many cross-utilitarian data sets into their patient risk factor models because to Quantum's ability to analyze at scale. They will be able to use environmental databases to examine the impact of pollution on a patient's health history, for example.

# Why are Quantum Computers faster?

A supercomputer may be particularly useful for challenging tasks such as sifting through a large database of protein sequences. However, it will have difficulty detecting the small patterns in the data that influence how proteins act. Proteins are lengthy chains of amino acids that fold into complicated forms to create helpful biological machineries.

A traditional supercomputer could try to fold a protein by brute force, using its numerous processors to check every potential method of bending the chemical chain before coming up with a solution. The supercomputer, however, stops as the protein sequences become longer and more complicated.

Quantum algorithms use a novel approach to solving these types of difficult issues by generating multidimensional spaces in which patterns linking individual data points emerge. In the case of proteins, there are already early quantum algorithms that can uncover folding patterns in whole new, more efficient ways, without the time-consuming checks that conventional computers require. As quantum technology improves and these algorithms improve, they may be able to solve protein folding issues that are too difficult for any supercomputer to solve.

**How Complexity stumps Super Computer:** Proteins are lengthy chains of amino acids that fold into complicated forms to create helpful biological machineries. A traditional supercomputer could try to fold a protein by brute force, using its numerous processors to check every potential method of bending the chemical chain before coming up with a solution.

**Quantum Computers are built for complexity:** Quantum algorithms use a novel approach to solving these types of difficult issues by generating multidimensional spaces in which patterns linking individual data points emerge.

# How do Quantum Computers work?

* **Superfluid:** A fan is most usually used to keep your desktop computer cool enough to function. Our quantum processors must be extremely cold, around a tenth of a degree above absolute zero. To do so, we make superconductors using super-cooled superfluid.
* **Superconductor:** Certain materials in our processors demonstrate another significant quantum mechanical feature at such ultra-low temperatures: electrons pass through them without resistance. As a result, they are referred to as "superconductors." When electrons flow through superconductors, they generate "Cooper pairs," which are pairs of electrons that match up.
* **Control:** As superconducting qubits, Josephson junctions are used in our quantum computers. We can regulate the behavior of these qubits and get them to keep, modify, and read out individual units of quantum information by shooting microwave photons at them.
* **Superposition:** A qubit isn't particularly helpful on its own. It can, however, perform a crucial trick by putting the quantum information it contains in a state of superposition, which represents a combination of all conceivable qubit configurations.
* **Entangled:** Entanglement is a quantum mechanical phenomenon in which the behavior of two independent objects is linked. Changes to one qubit have a direct influence on the other when two qubits are entangled. Quantum algorithms take advantage of these connections to solve difficult issues.

# Difference between Classical Computing and Quantum Computing

|  |  |
| --- | --- |
| **Classical Computing** | **Quantum Computing** |
| For classical computing, classic computers are utilized. | The quantum computing method is used in quantum computers. |
| It works with binary digits to accomplish computations. | It makes computations based on the likelihood of the object. |
| The physical state, which is generally binary, is used to carry out logical operations. | The quantum state, also known as qubits, is used for logical processes. |
| Java, C, and C++ are examples of standardized programming languages. | It is not dependent on any particular programming language. |
| It's made up of CPUs and other processors. | It runs on a set of qubits and has a basic architecture. |

**Edge Computing**

# Edge Computing

* Edge computing is a distributed IT architecture that brings computing resources from clouds and data centers as near to the source as feasible. Edge computing's major purpose is to minimize latency while processing data and lowering network expenses.
* The router, ISP, routing switches, integrated access devices (IADs), multiplexers, and other equipment at the edge are examples. The most important aspect of this network edge is that it should be located near the device.

# How does Edge Computing Work?

* Data is created in a typical environment on a user's PC or any other client application. The data is then sent to the server via channels such as the internet, intranet, LAN, and so on, where it is stored and processed. This is still a tried-and-true method of client-server computing.
* Traditional data center infrastructures, on the other hand, have struggled to keep up with the exponential development in the volume of data created and the number of devices connected to the internet. According to a Gartner report, by 2025, 75% of company data would be generated outside of centralized data centers. This volume of data places a tremendous pressure on the internet, causing congestion and interruption.
* The idea behind edge computing is straightforward: rather than bringing data closer to the data center, the data center is moved closer to the data. The data center's storage and processing resources are installed as close as feasible to where the data is created (preferably in the same area).

|  |  |
| --- | --- |
| **Early Computing** | Only one PC is used to execute the applications. |
| **Personal Computing** | Local applications operate on the user's device or in a data center. |
| **Cloud Computing** | Data centers host applications that are processed through the cloud. |
| **Edge Computing** | Applications run near to the user's device or on the network edge. |

# Why Edge Computing?

* It enables smart applications and devices to react to data practically simultaneously, which is critical in business and self-driving automobiles.
* It can process data without placing it on a public cloud, which assures complete security.
* While over an extended network, data may become corrupt, compromising the data's dependability for companies to utilize.
* The utilization of cloud computing is limited by edge data computation.

# Benefits of Edge Computing

* **Eliminates Latency:** The time it takes to send data between two places on a network is referred to as latency. Delays can be caused by large physical distances between these two places, as well as network congestion. Latency difficulties are essentially non-existent thanks to edge computing, which brings the points closer together.
* **Saves Bandwidth:** The pace at which data is transported over a network is referred to as bandwidth. Because all networks have a finite amount of bandwidth, the amount of data that can be sent and the number of devices that can handle it are also constrained. Edge computing allows multiple devices to run over a much lower and more efficient bandwidth by putting data servers at the places where data is created.
* **Reduces Congestion:** Despite the fact that the Internet has changed through time, the sheer volume of data created every day by billions of devices can cause significant congestion. Local storage is available in edge computing, and local servers can execute critical edge analytics in the case of a network failure.
* **Improved reaction time.**
* **Compliance and security.**
* **Solution that is cost-effective.**
* **With Intermittent Connectivity, Reliable Operation**

# Disadvantages of Edge Computing

* **Implementation Cost:** Implementing an edge infrastructure in a company may be both complicated and costly. Before deployment, it requires a clear scope and goal, as well as extra equipment and resources.
* **Incomplete Data:** Edge computing can only process subsets of data, which should be determined in advance of implementation. Companies may lose crucial data and information as a result of this.
* **Security:** Because edge computing is a distributed system, it might be difficult to provide proper security. Processing data outside of the network's edge has several hazards. The inclusion of new IoT devices might also make it easier for attackers to get access to the device.

# Edge Cloud Computing Services:

* Internet of Things (IoT)
* Gaming
* Health Care
* Enterprise Security in Smart Cities
* Intelligent Transportation

# Application of Edge Computing in Various Industries

* **Healthcare:** Fitness trackers, heart monitoring smartwatches, glucose monitors, and other wearable IoT devices are used in the healthcare business. Every second, these gadgets capture data, which is subsequently examined to gain insights. However, if the data analysis for this real-time data is sluggish, it is meaningless. Edge Computing is critical in healthcare because it allows data to be examined and interpreted quickly. GE Healthcare is an example of this, since the business employs NVIDIA processors in its medical equipment to improve data processing through edge computing.
* **Transportation:** In the transportation industry, edge computing has several uses, notably in self-driving automobiles. To ensure that these self-driving cars perform properly, they need a variety of sensors, including 360-degree cameras, motion sensors, radar-based systems, GPS, and so on.
* **Retail:** Many retail shops are becoming more tech-savvy these days! Customers may use their phone app or a QR code to swipe into the store and begin selecting what they wish to buy. Customers may then just depart the store, and the cost of anything they purchased will be taken from their account automatically. Stores can achieve this by analyzing what all consumers are buying using a mix of motion sensors and in-store cameras. However, this necessitates Edge Computing since too much time lag in data analysis might lead to consumers just picking up their belongings and departing for free.

# Examples and Use Cases

* Smart home gadgets are one of the finest places to start with edge computing. A multitude of IoT gadgets collect data from around the house in smart homes. After then, the information is transferred to a distant server, where it is stored and processed. In the case of a network outage, this architecture might bring a slew of issues. Edge computing can cut backhaul costs and latency by bringing data storage and processing centers closer to the smart home.
* The cloud gaming sector is another application of edge computing. Companies that provide cloud gaming services want to put their servers as near to the players as feasible. This will cut down on lags and provide you a more immersive gaming experience.
* **Autonomous Vehicles:** Intel, a GE Digital partner, forecasts that autonomous cars would create 40 TB of data for every eight hours of driving due to hundreds of on-board sensors. As a result, wheels—edge computing plays a key role. Sending all of your data to the cloud is both risky and impracticable.
* **Fleet Management:** Let's take an example. In the case of a trucking firm, the major purpose is to aggregate and transfer data from different operational data points to the cloud, such as wheels, brakes, and batteries. The cloud analyses the major operational components of health.

**Machine Learning**

# What is Machine Learning?

Machine learning is a branch of computer science that enables computers to learn without being explicitly programmed. One of the most intriguing technologies that one has ever encountered is machine learning.

As the name suggests, it provides the computer the ability to learn, which makes it more human-like. Machine learning is currently in use, maybe in many more locations than one might imagine.

# Introduction to Machine Learning

* Arthur Samuel, an American pioneer in the fields of computer games and artificial intelligence, invented the phrase Machine Learning in 1959, stating that "it offers computers the ability to learn without being expressly taught."
* The term "machine learning" is the most recent craze. It is well-deserved, as it is one of the most fascinating areas of computer science.
* Because many real-world situations are so complicated, it's difficult, if not impossible, to come up with specific algorithms that will handle them properly every time. "Is this cancer?" "Will this person enjoy this movie?" "Will this person like this movie?" are some examples of machine learning challenges. Machine Learning is a perfect fit for such challenges, and it has been used to solve them successfully in the past.

# Classification of Machine Learning

1. On the Basis of the nature of the learning “signal” or “feedback” available to a learning system.
   * 1. **Supervised Learning:** A "teacher" presents the computer with sample inputs and desired outputs, with the purpose of learning a general rule that maps inputs to outputs. The model is trained until it reaches the appropriate degree of accuracy on the training data. Some real – life examples are:
        1. I**mage Clarification:** You practice using pictures and labels. Then you offer a fresh image in the future, expecting the computer to recognize the new thing.
        2. **Market Prediction/Regression:** You use previous market data to educate the computer and then ask it to forecast the new price in the future. You practice using pictures and labels. Then you offer a fresh image in the future, expecting the computer to recognize the new thing.
     2. **Unsupervised Learning:** The learning algorithm is given no labels and is left to identify structure in its data on its own. It's utilized to divide people into separate categories. Learning without supervision may be a goal in and of itself (discovering hidden patterns in data).
        + **Clustering:** You instruct the computer to group comparable data into clusters, which is critical in science and research.
        + **High Dimension Visualization:** Utilize the computer to aid in the visualization of high-dimensional data.
        + **Generative Models:** A model will be able to create more data if it has captured the probability distribution of your input data. This can help you improve the robustness of your classifier.
     3. **Semi – Supervised Learning:** Semi-supervised learning issues are those in which you have a significant quantity of input data but only portion of it is labelled. These issues fall between supervised and unsupervised learning. Consider a photo repository with only a few photographs tagged (e.g. dog, cat, and person) and the bulk of the images unlabeled.
     4. **Reinforcement Learning:** A computer programme interacts with a dynamic environment in order to accomplish a certain task (such as driving a vehicle or playing a game against an opponent). As it navigates its issue space, the software receives feedback in the form of incentives and penalties.
2. On the basis of “output” desired from a machine learned system
   * 1. **Clarification:** The learner must create a model that assigns unseen inputs to one or more (multi-label categorization) of these classes once the inputs are separated into two or more classes. This is usually done under the supervision of a professional. Spam filtering is an example of classification, with email (or other) messages as inputs and the classifications "spam" and "not spam" as classes.
     2. **Regression:** It's a supervised learning issue with continuous rather than discrete outputs. Predicting stock prices, for example, using past data.
     3. **Clustering:** A collection of inputs is separated into groups in this case. Because the groupings aren't known ahead of time, unlike classification, this is usually an unsupervised task.
     4. **Density Estimation:** The goal is to determine the distribution of inputs in a given area.
     5. **Dimensional Reduction:** It reduces the number of variables by mapping them onto a lower-dimensional space. A related challenge is topic modelling, in which a software is given a series of human language documents and asked to determine whether documents cover comparable themes.

We have a variety of algorithms that are utilized to complete these machine learning tasks/problems based on these machine learning tasks/problems. Linear Regression, Logistic Regression, Decision Tree, SVM (Support vector machines), Naive Bayes, KNN (K closest neighbors), K-Means, Random Forest, and others are some of the most often used machine learning methods.

# Terminologies OF MACHINE LEARNING

* **Model:** A model is a particular representation that is learned from data using a machine learning technique. Hypothesis is another term for a model.
* **Feature:** A feature is a quantifiable aspect of our data that is unique to it. A feature vector is a useful way to define a set of numeric features. The model is fed with feature vectors as input. Color, smell, and taste, for example, may be used to forecast the appearance of a fruit.
* **Target (Label):** A target variable, often known as a label, is the value that our model is supposed to predict. The label for each piece of input in the fruit example stated in the features section would be the name of the fruit, such as apple, orange, or banana.
* **Training:** The objective is to provide a collection of inputs (features) and predicted outputs (labels), such that after training, we will have a model (hypothesis) that will map incoming data to one of the trained categories.
* **Prediction:** Once our model is complete, we can feed it a set of inputs and get a projected result (label).

# How Machine learning works

* Obtaining historical data in any format that can be processed. The higher the data quality, the better it will be for modelling.
* Data Preparation — Data is sometimes obtained in its raw form and must be pre-processed. For example, certain tuples may have missing values for specific attributes, which must be replaced with appropriate values before machine learning or any other type of data mining can be performed.
* Missing values for numerical variables like the home price can be replaced with the attribute's mean value, while missing values for categorical characteristics can be replaced with the attribute having the highest mode. This is always contingent on the filters we employ. Converting data from text or pictures to numerical representation, whether in the form of a list, array, or matrix, is necessary. To put it another way, data must be made more relevant and consistent. It will be transformed into a machine-readable format.
* Divide the data into three sets: training, cross-validation, and test. The proportions between the two sets must equal 6:2:2.
* On the training set, construct models using appropriate methods and methodologies.
* Testing our proposed model with data that was not provided to it during training and assessing its performance using measures like F1 score, accuracy, and recall.

# Properties of Data in machine learning

* **Volume:** Data Dimensions. Huge amounts of data are created every millisecond as the world's population grows and technology becomes more accessible.
* **Variety:** Healthcare, pictures, videos, and audio snippets are examples of different types of data.
* **Velocity:** The rate at which data is streamed and generated.
* **Value:** Data's usefulness in terms of the information that researchers may glean from it.
* **Veracity:** We need certainty and accuracy in the data we're dealing with.

# Applications of Machine learning

* **Web Search Engine:** Because the system has learned how to rank pages using a complicated learning algorithm, search engines like Google, Bing, and others operate so well.
* **Photo Tagging Applications:** The option to tag pals makes any photo tagging programme, whether it's Facebook or another, even more fun. All of this is made possible by a facial recognition system that works behind the scenes of the app.
* **Spam Detector:** Our mail agent, such as Gmail or Hotmail, performs a lot of the heavy lifting for us when it comes to categorizing emails and transferring junk to the spam folder. This is accomplished once again via a spam classifier operating in the mail application's backend.

Machine Learning is now being used by businesses to enhance business choices, increase productivity, identify disease, forecast weather, and many other tasks. We need better tools to comprehend the data we have now, but we also need to prepare for the data we will have in the future, thanks to the exponential rise of technology. To do this, we must create intelligent machines. To accomplish simple tasks, we can develop a programme. However, hardwiring intelligence into it is sometimes challenging. The best way to achieve it is to devise a method for machines to self-learn.

# Examples of Machine Learning

**Database Mining for growth of automation:** Web-click data for improved UX (User Experience), medical records for greater automation in healthcare, biological data, and many more applications are examples.

**Application that cannot be programmed:** Some jobs can't be coded since the machines we use aren't designed that way. Autonomous driving, recognition tasks from unstructured data (Face Recognition/Handwriting Recognition), natural language processing, and computer vision are only a few examples.

**Understanding Human Learning:** This is the closest we've come to comprehending and simulating the human brain. It is the beginning of a new revolution, the true AI revolution. Let us now go on to a more formal definition of Machine Learning after a quick overview.

# Difference between Artificial Learning and Machine learning

|  |  |
| --- | --- |
| **Artificial Learning** | **Machine Learning** |
| Artificial intelligence (AI) is described as the ability to learn and apply information, while intelligence is defined as the acquisition of knowledge. | Machine Learning (ML) is described as the acquisition of knowledge or expertise by a computer. |
| The goal is to enhance the likelihood of success rather than precision. | Its goal is to improve accuracy, but it is unconcerned with success. |
| The idea is to tackle complicated problems by simulating natural intelligence. | The idea is to learn from data on a certain task in order to improve performance. |
| It is working on a system that solves issues in the same way as humans do. | It entails the development of self-learning algorithms. |
| The AI will seek for the best option. | ML will choose the best solution, regardless of whether it actually the best. |

**Deep** **Learning**

# What is deep learning?

* Deep learning is a subset of artificial intelligence that is built on the branch of machine learning. Deep learning will suffice since neural networks mimic the human brain. Nothing is explicitly coded in deep learning. Essentially, it is a machine teaching class that does feature extraction and modification using a large number of nonlinear processing units. Each of the subsequent levels uses the output from the previous layer as input.
* Deep learning models are capable of focusing on accurate features on their own, with just a little input from the programmer, and are highly useful in resolving the dimensionality problem. When there are a large number of inputs and outputs, deep learning techniques are applied.

# Architectures of deep learning

1. **Deep Neural Networks:** It's a neural network that integrates a specific amount of complexity, which implies there are multiple hidden layers in between the input and output layers. They have a strong understanding of how to model and handle non-linear relationships.
2. **Deep Belief Networks:** A multi-layer belief network makes up a deep belief network, which is a type of Deep Neural Network.
   1. **Steps to perform DBM:**
      * 1. A layer of features is learnt from perceptible units using the Contrastive Divergence method.
        2. The previously taught features are then considered as visible units that perform feature learning.
        3. Finally, after the final hidden layer has been learned, the DBN as a whole is trained.
3. **Recurrent Neural Networks:** It can perform both parallel and sequential computations and is quite comparable to the human brain (large feedback network of connected neurons). They are more exact because they are capable of recalling all of the critical aspects of the input they have received.

# Types of deep Learning

1. **Feed Forward Neural Network:** A feed-forward neural network is a type of Artificial Neural Network that prevents nodes from forming a loop. All perceptions are grouped inside layers in this type of neural network, with the input layer taking the input and the output layer producing the output. The term "hidden layers" refers to the fact that the hidden layers are not connected to the outer world. Each node in the succeeding layer is related with each of the perceptions included in the previous layer. All of the nodes are fully linked, as can be seen. It has no visible or concealed connections between nodes within the same layer.
   * 1. Application:

Data Compression

Pattern Recognition

Computer Vision

1. **Recurrent Neural Network:** Another type of feed-forward network is recurrent neural networks. In this case, each of the neurons in the hidden layers receives an input with a particular time delay. The recurrent neural network mostly uses previous information from previous iterations. To guess the next word in a phrase, for example, one must be familiar with the words that came before it. It not only analyses the data, but it also distributes the length and weights throughout time. It prevents the model's size from growing in lockstep with the input size.
   * 1. Application:

Robot Control

Machine Translation

Speech Recognition

1. **Convolutional Neural Network:** Convolutional Neural Networks are a type of neural network that is used to classify pictures, cluster them, and recognize objects. DNNs allow for the creation of hierarchical picture representations without the need for supervision. Deep convolutional neural networks are chosen over other neural networks for achieving the greatest accuracy.
   * 1. Application:

Video Analysis

NLP

Drug Discovery

1. **Restricted Boltzmann Machine:** RBMs are a different type of Boltzmann Machine. In this case, the neurons in the input layer and the hidden layer have symmetric connections between them. Within the relevant layer, however, there is no internal relationship. Boltzmann machines, unlike RBM, do have internal connections within the hidden layer. These constraints in BMs allow the model to train more effectively.
2. Application:

Risk Detection

Filtering

Feature Learning

1. **Auto encoders:** Another type of unsupervised machine learning technique is an auto encoder neural network. The number of concealed cells is simply less than the number of input cells in this case. The number of input cells, on the other hand, is equal to the number of output cells. To force AEs to locate common patterns and generalize the data, an auto encoder network is trained to show the output similarly to the fed input.
   * 1. Application:

Classification

Clustering

Feature compression

* + 1. **Encoder:** Reduce the dimensionality of the supplied data.
    2. **Decoder:** Reassemble the data that has been compressed.

# Deep Learning applications

* **Self-Driving Cars:** Self-driving automobiles can record the sights around them by digesting a large quantity of data, and then select which actions to include, such as turning left or right or stopping. As a result, it will determine what activities it should take in order to minimize the number of incidents that occur each year.
* **Voice Controlled Assistance:** Siri is the first thing that springs to mind when we think about voice control help. As a result, you can tell Siri to do whatever you want, and it will search for it and show it foryou.
* **Automatic Image Caption Generator:** The programme will produce captions for every image you supply. If you say blue colored eye, the image will be of a blue-hued eye with a caption at the bottom.
* **Automatic Machine Translation:** With the use of deep learning, we can transform one language into another via automatic machine translation.

# Limitations

* It is only able to learn through observation.
* It has concerns with prejudices.

# Advantages

* The necessity for feature engineering is reduced.
* It eliminates any unnecessary expenditures.
* It quickly detects even the most challenging flaws.
* It yields the best-in-class problem-solving solutions.

# Disadvantages

* It necessitates a large volume of data.
* The cost of training is fairly high.
* It lacks a solid theoretical foundation.

# Real Life examples

1. What distinguishes a square from other shapes?

a) Take a look at the four lines!

b) Is it a closed figure?

c) Are the two sides perpendicular to one another?

d) Is it true that all sides are equal?

As a result, Deep Learning is a hard process of detecting the form that may be split down into smaller tasks.

1. Recognizing an Animal! (Is it a Dog or a Cat?)

Defining crucial face traits for categorization, which the algorithm will subsequently recognize automatically. (Machine Learning, on the other hand, will manually assign such characteristics for categorization.)

**Image Processing**

# What is an Image?

* Before we go into image processing, it's important to first grasp what an image is. The dimensions (height and breadth) of an image are determined by the number of pixels. For example, if an image's dimensions are 500 × 400 (width x height), the image's total number of pixels is 200000.

This pixel is a spot on the image that is shaded, opacified, or colored in a certain way. It is often shown in one of the following ways:

* **Gray Scale:** A pixel is a number that ranges from 0 to 255. (0 is completely black and 255 is completely white).
* **RGB:** Pixels are made up of three numbers ranging from 0 to 255. (The integers represent the intensity of red, green, and blue).
* **RGBA:** It's an extension of RGB with an additional alpha field that reflects the image's opacity.

# Types of an image

**Binary Image:** The binary picture, as the name implies, has only two pixel elements: 0 and 1, where 0 denotes black and 1 denotes white. Monochrome is another name for this picture.

**Black and White Image:** BLACK AND WHITE Picture is an image made entirely of black and white pixels.

**8 bit Color Format:** It is by far the most well-known picture format. Grayscale Image is a type of image that contains 256 different shades of color. In this format, 0 represents black, 255 represents white, and 127 represents grey.

**16 bit Color Format:** – It's a format for color images. It contains 65,536 distinct colors. High Color Format is another name for it. The color distribution in this format differs from that of a grayscale picture.

# Image Processing

Image processing is the process of converting a physical image to a digital representation and then conducting operations on it to extract valuable information. When implementing specific specified signal processing algorithms, the image processing system normally interprets all pictures as 2D signals.

# Image Processing Mainly Include Following steps:

1. Importing the image using image acquisition software;

2. Analyzing and editing the image;

3. Output, which can be a changed image or a report based on the image analysis;

# Division Of iMAGE pROCESSING:

* **Analog Image Processing:** Analog signals are used in analogue image processing. It involves two-dimensional analogue signal processing. The visuals are modified by electrical methods by altering the electrical signal in this form of processing. The picture on television is a frequent example.
* **Digital Image Processing:** The development of a digital system that conducts operations on a digital picture is the subject of digital image processing. With the passage of time, digital image processing has surpassed analogue image processing due to its broader range of uses.

# Relationship between Digital image and a Signal

Signal:

* In the physical world, any quantity that can be measured through time, space, or a higher dimension can be considered a signal. A signal is a mathematical function that communicates data.
* A signal can be one-dimensional, two-dimensional, or even three-dimensional. A signal that is measured throughout time is referred to as a one-dimensional signal. A speech signal is a frequent example.
* Signals that are measured across other physical qualities are referred to as two-dimensional signals. A digital picture is an example of a two-dimensional signal. In the following session, we'll look at how one-dimensional or two-dimensional signals, as well as higher signals, are created and evaluated in greater depth.

Relationship:

* Since everything that transmits information or broadcasts a message between two observers in the physical world is a signal. As a signal, this involves speech or (human voice) or a picture. Since our voice is changed to a sound wave/signal when we talk, it is transformed with regard to the time to the person to whom we are speaking.
* Not only that, but the way a digital camera works, since getting a picture from one portion of the system to the other requires the transmission of a signal.

# How a Digital image is formed

* Because taking a picture with a camera is a physical procedure. As an energy source, sunlight is employed. The picture is captured with the help of a sensor array. When sunlight shines on an item, the quantity of light reflected by that object is sensed by the sensors, and the amount of sensed data generates a continuous voltage signal.
* We must transfer this data into digital form in order to build a digital picture. This necessitates the use of sampling and quantification. (They will be mentioned later.) After sampling and quantization, a two-dimensional array or matrix of integers is created, which is nothing more than a digital picture.

# Phases of Image processing

* **Acquisition:** The initial stage in image processing is picture acquisition. In image processing, this stage is also known as preprocessing. It entails getting the picture from a source, which is generally hardware.
* **Image Enhancement:** The technique of bringing out and highlighting particular areas of interest in a picture that has been concealed is known as image enhancement. This can include adjusting the brightness, contrast, and other settings.
* **Image Restoration:** The technique of enhancing the look of a picture is known as image restoration. Picture restoration, unlike image augmentation, is done using mathematical or probabilistic models.
* **Color Image Processing:** In the digital world, color image processing encompasses a variety of color modelling approaches. Because of the widespread usage of digital photos on the internet, this step has acquired popularity.
* **Wavelets and Multi- Resolution Processing:** Wavelets are a type of visual representation that comes in a variety of resolutions. For data compression and pyramidal representation, the pictures are separated into wavelets or smaller sections.
* **Image Compression:** Compression is a technique for reducing the amount of storage or bandwidth needed to preserve or transmit a picture. This is especially true when the photograph will be used on the Internet.
* **Morphological Processing:** Morphological processing is a collection of techniques used to modify pictures based on their forms.
* **Segmentation Procedure:** One of the most difficult aspects of image processing is segmentation. It entails breaking down a picture into its component elements or objects.
* **Representation and Description:** Each region of an image is represented and described in a manner appropriate for further computer processing once it is divided into regions in the segmentation process. The qualities and geographical aspects of an image are dealt with in representation. The task of description is to extract quantitative information that may be used to distinguish one class of things from another.
* **Objection and Rejection:** Based on its description, recognition assigns a label to an item.

# Overlapping Field

* **Machine/Computer Vision:** The development of a system in which the input is an image and the output is some information is known as machine vision or computer vision. Consider developing a technology that scans a person's face and unlocks any type of lock. This is how the system would look like.
* **Computer Graphics:** Rather of the picture being collected by any instrument, computer graphics deals with the generation of images from object models. Consider the case of object rendering. Creating a picture from a model of an object. This is how such a system might seem.
* **Artificial Intelligence:** Artificial intelligence is essentially the science of imbuing robots with human intellect. In image processing, artificial intelligence has a wide range of applications. For instance, creating computer-assisted diagnostic systems that assist clinicians in evaluating X-ray, MRI, and other pictures and then marking salient sections to be reviewed by the doctor.
* **Signal Processing:** Image processing falls under the umbrella of signal processing. The quantity of light reflected by an item in the physical world (3d world) passes through the camera's lens and is converted to a 2d signal, resulting in picture generation. This image is subsequently converted to a digital format utilizing signal processing techniques, and then altered via digital image processing.

# Application of Image Processing

* **Medical Image Retrieval:** Medical research has made substantial use of image processing, which has resulted in more efficient and precise treatment strategies. It may, for example, be utilized in breast scans to detect breast cancer early using a sophisticated nodule identification algorithm.
* **Traffic Sensing Technologies:** We employ a video image processing system, or VIPS, for traffic sensors. This is made up of three parts: a) an image capture system, b) a communications system, and c) an image processing system. When collecting video, a VIPS has many detecting zones that emit an "on" signal whenever a vehicle enters the zone and an "off" signal when the vehicle leaves the zone
* **Image Reconstruction:** Picture processing may be used to restore and fill in the gaps in an image that are missing or corrupted. This entails creating newer copies of old and damaged photographs using image processing systems that have been thoroughly trained using existing photo datasets.
* **Face Detection:** Face detection is one of the most prevalent applications of image processing that we employ today. It uses deep learning methods to educate the computer on certain aspects of human faces, such as the shape of the face, the distance between the eyes, and so on. Face detection is an important element in security, biometrics, and even the filters on most social networking platforms these days.

# Advantages of Image Processing

The use of image processing algorithms has had a significant influence on numerous tech companies. Here are a few of the most important advantages of image processing, regardless of the application:

* Any format of the digital image can be made available (improved image, X-Ray, photo negative, etc.)
* It aids in the enhancement of pictures for human understanding.
* Images may be analyzed and information retrieved for machine interpretation.
* The density and contrast of the pixels in the image may be adjusted to any desired level.
* Images can be simply stored and accessed.
* It makes it simple to send photos to third-party suppliers through the internet.

# Disadvantages of Image processing

* Depending on the system, the initial cost is significant.
* The picture will be lost if the system is destroyed.