# SEMINAR REPORT

# ON

# Machine Learning

## AS A PARTIAL REQUIREMENT FOR THE DEGREE

## OF

## BACHELOR OF COMPUTER APPLICATION

## (B.C.A.)

: SUBMITTED BY :

179-5609-vaghani bhautik
mukeshbhai

: GUIDED BY :

Sanskruti A. Ramani

**PROF. V. B. SHAH INSTITUTE OF MANAGEMENT,**
**R. V. PATEL COLLEGE OF COMMERCE (ENG. MED.),**
**V. L. SHAH COLLEGE OF COMMERCE (GUJ. MED.) &**
**SUTEX BANK COLLEGE OF COMPUTER APPLICATIONS & SCIENCE**
Managed by Jivan Jyot Trust, Amroli.

All Colleges are Affiliated to Veer Narmad South Gujarat University, Surat.
Self Financed B.B.A. - B.Com. (Eng. & Guj. Med.) - B.C.A & B.Sc. Data Science Degree Programme
Accredited by National Assessment and Accreditation Council with "B" Grade

**Prin. Dr. Mukesh R. Goyani**
M.Com., M.Ed., M.Phil., GSET, Ph.D.
Senate Member, VNSGU, Surat
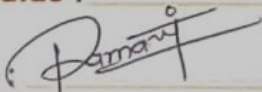
At & Po. : Amroli, Surat,
Pin. 394107, Gujarat (India)
www.amrolicollege.ac.in

(0261) 2494073, 2495643, 2496478
9427113947, 6354626492
principal@amrolicollege.ac.in
Principal_230@vnsgu.ac.in

# Bachelor of Computer Application
## Seminar Certificate
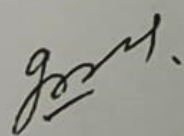
This is to certify that the seminar presentation and report titled **"Machine Learning"** is the bonafide work carried out by **Bhautikbhai Mukeshbhai Vaghani (5609)**, students of TYBCA Sem-VI of Sutex Bank College of Computer Applications and Science, Amroli, (Surat) affiliated to Veer Narmad South Gujarat University. He/she has successfully completed his/her seminar work in partial fulfillment of the requirements for the award of the degree of "Bachelor of Computer Application" during the academic year 2023-24.

**Internal Guide :** Sanskruti A. Ramani

**Signature :**

**Date :**
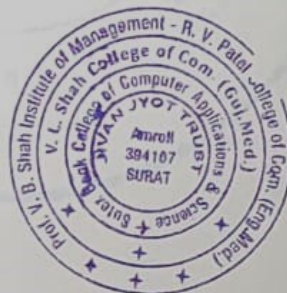
**External Examiner :**

**Dr. Mukesh R. Goyani**
**Principal**

# 1. Introduction of Machine Learning

➢ In the real world, we are surrounded by humans who can learn everything from their experiences with their learning capability, and we have computers or machines which work on our instructions.

➢ But can a machine also learn from experiences or past data like a human does? So here comes the role of **Machine Learning**.
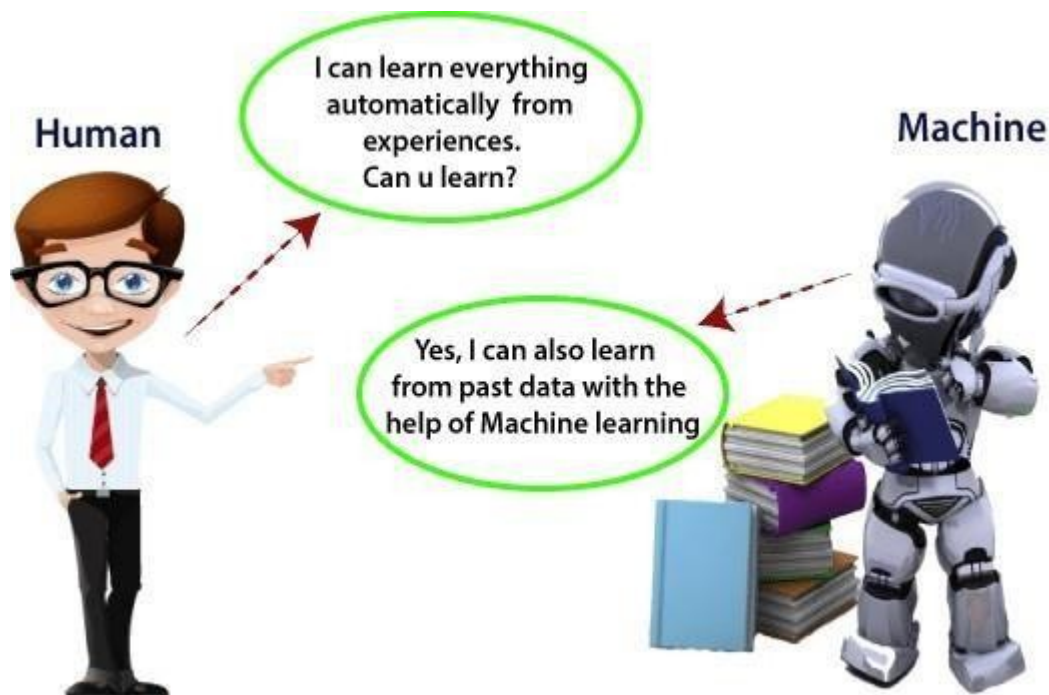


**Introduction to Machine Learning**

➢ A subset of artificial intelligence known as machine learning focuses primarily on the creation of algorithms that enable a computer to independently learn from data and previous experiences.

➢ Arthur Samuel first used the term "machine learning" in 1959.

➢ "**Without being explicitly programmed, machine learning enables a machine to automatically learn from data, improve performance from experiences, and predict things**".

➢ Machine learning algorithms create a mathematical model that, without being explicitly programmed, aids in making predictions or decisions with the assistance of sample historical data, or training data.

➢ For the purpose of developing predictive models, machine learning brings together statistics and computer science.

➢ Algorithms that learn from historical data are either constructed or utilized in machine learning.

➢ The performance will rise in proportion to the quantity of information we provide.

# 2. Types of Machine Learning

➢ **Machine learning is a subset of AI, which enables the machine to automatically learn from data, improve performance from past experiences, and make predictions**.

➢ Machine learning contains a set of algorithms that work on a huge amount of data.

➢ Data is fed to these algorithms to train them, and on the basis of training, they build the model & perform a specific task.



Types of Machine Learning

➢ These ML algorithms help to solve different business problems like Regression, Classification, Forecasting, Clustering, and Associations, etc.

➢ Based on the methods and way of learning, machine learning is divided into mainly four types, which are:

1. Supervised Machine Learning

2. Unsupervised Machine Learning

3. Semi-Supervised Machine Learning

4. Reinforcement Learning
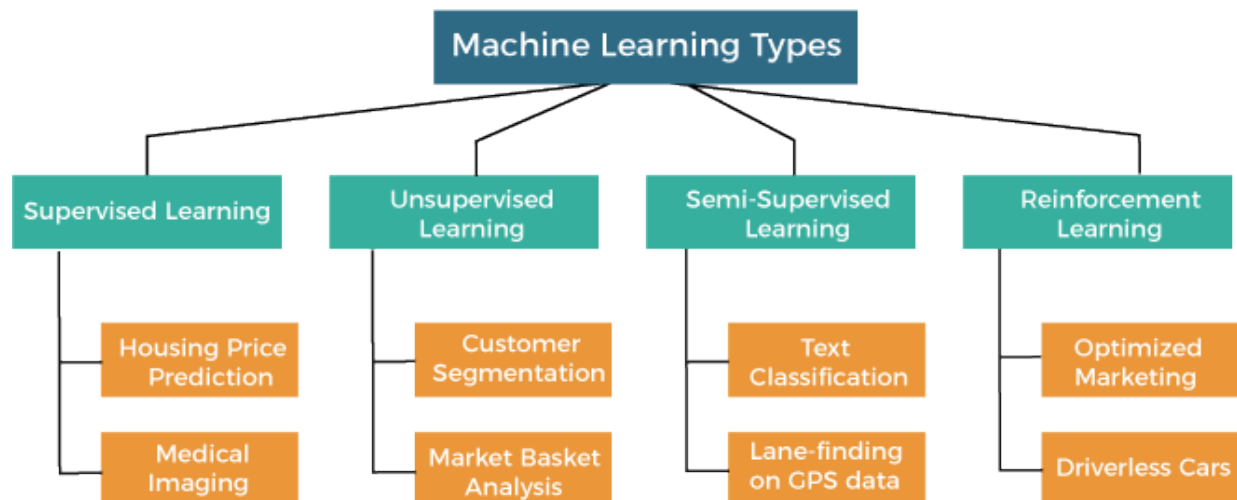
In this topic, we will provide a detailed description of the types of Machine Learning along with their respective algorithms:

# 1. Supervised Machine Learning

- ➤ As its name suggests, Supervised machine learning is based on supervision.
- ➤ It means in the supervised learning technique, we train the machines using the "labelled" dataset, and based on the training, the machine predicts the output.
- ➤ Here, the labelled data specifies that some of the inputs are already mapped to the output. More preciously, we can say; first, we train the machine with the input and corresponding output, and then we ask the machine to predict the output using the test dataset.
- ➤ **The main goal of the supervised learning technique is to map the input variable(x) with the output variable(y).**
- ➤ Some real-world applications of supervised learning are **Risk Assessment, Fraud Detection, Spam filtering,** etc.
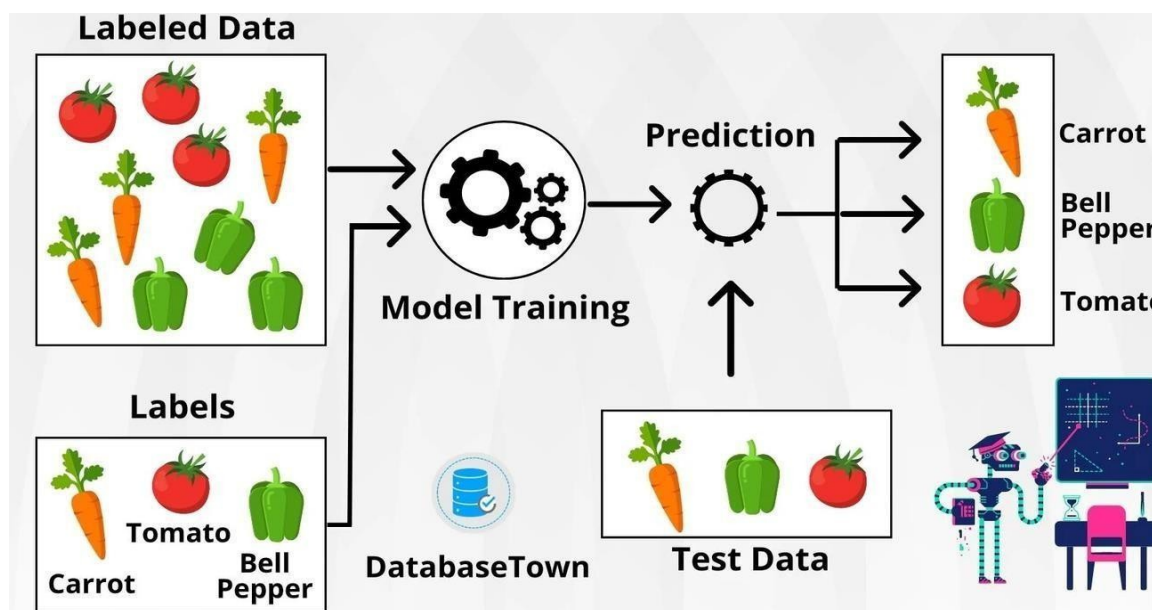
## Advantages and Disadvantages of Unsupervised Learning Algorithm

**Advantages:**

- o These algorithms can be used for complicated tasks compared to the supervised ones because these algorithms work on the unlabeled dataset.
- o Unsupervised algorithms are preferable for various tasks as getting the unlabeled dataset is easier as compared to the labelled dataset.

**Disadvantages:**

- o The output of an unsupervised algorithm can be less accurate as the dataset is not labelled, and algorithms are not trained with the exact output in prior.
- o Working with Unsupervised learning is more difficult as it works with the unlabelled dataset that does not map with the output.

# 2. Unsupervised Machine Learning

- ➤ Unsupervised learning is different from the Supervised learning technique; as its name suggests, there is no need for supervision.
- ➤ It means, in unsupervised machine learning, the machine is trained using the unlabeled dataset, and the machine predicts the output without any supervision.
- ➤ In unsupervised learning, the models are trained with the data that is neither classified nor labelled, and the model acts on that data without any supervision.
- ➤ **The main aim of the unsupervised learning algorithm is to group or categories the unsorted dataset according to the similarities, patterns, and differences.**
- ➤ Machines are instructed to find the hidden patterns from the input dataset.
- ➤ So, now the machine will discover its patterns and differences, such as colour difference, shape difference, and predict the output when it is tested with the test dataset.

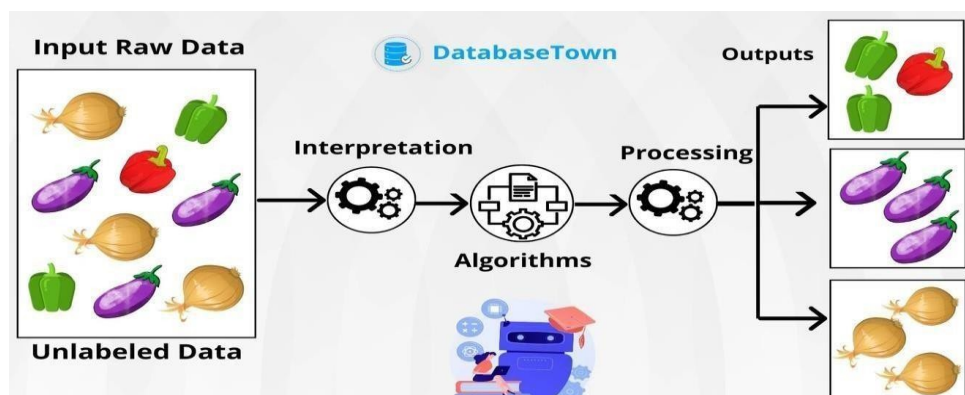## Advantages and Disadvantages of Unsupervised Learning Algorithm

**Advantages:**

o   These algorithms can be used for complicated tasks compared to the supervised ones because these algorithms work on the unlabeled dataset.

o   Unsupervised algorithms are preferable for various tasks as getting the unlabeled dataset is easier as compared to the labelled dataset.

**Disadvantages:**

o   The output of an unsupervised algorithm can be less accurate as the dataset is not labelled, and algorithms are not trained with the exact output in prior.

o   Working with Unsupervised learning is more difficult as it works with the unlabelled dataset that does not map with the output.

# 3. Semi-Supervised Learning

➢   **Semi-Supervised learning is a type of Machine Learning algorithm that lies between Supervised and Unsupervised machine learning**.

➢   It represents the intermediate ground between Supervised (With Labelled training data) and Unsupervised learning (with no labelled training data) algorithms and uses the combination of labelled and unlabeled datasets during the training period.

➢   **A**lthough Semi-supervised learning is the middle ground between supervised and unsupervised learning and operates on the data that consists of a few labels, it mostly consists of unlabeled data.

➢   As labels are costly, but for corporate purposes, they may have few labels.

➢   It is completely different from supervised and unsupervised learning as they are based on the presence & absence of labels.

## Advantages and disadvantages of Semi-supervised Learning

**Advantages:**

o   It is simple and easy to understand the algorithm.

o   It is highly efficient.

o   It is used to solve drawbacks of Supervised and Unsupervised Learning algorithms.

**Disadvantages:**

- Iterations results may not be stable.
- We cannot apply these algorithms to network-level data.
- Accuracy is low.

# 3. Key Components of Machine Learning

➢ **Data**: Data is the foundation of machine learning. It can be structured, semi-structured, or unstructured. The quality, quantity, and relevance of data significantly impact the performance of machine learning models.

➢ **Algorithms**: Machine learning algorithms are the mathematical models or techniques used to learn patterns and relationships from data and make predictions or decisions. Common types of algorithms include:

1. Supervised learning algorithms (e.g., regression, classification)
2. Unsupervised learning algorithms (e.g., clustering, dimensionality reduction)
3. Semi-supervised learning algorithms
4. Reinforcement learning algorithms

➢ **Model Evaluation**: Evaluating the performance of machine learning models is crucial to ensure their effectiveness and reliability. Common evaluation metrics include accuracy, precision, recall, F1-score, ROC-AUC, etc.

➢ **Training**: Training involves using labeled data to teach the machine learning model to recognize patterns or make predictions. During training, the model adjusts its parameters to minimize the difference between predicted and actual outcomes.

➢ **Validation**: Validation is the process of assessing how well the trained model generalizes to new, unseen data. Techniques such as cross-validation are used to estimate the model's performance on independent datasets.

➢ **Optimization**: Optimization involves fine-tuning the model and its hyperparameters to improve performance. Techniques such as grid search, random search, and Bayesian optimization are commonly used for optimization.

➢ **Deployment**: Once a model is trained and validated, it needs to be deployed into production environments where it can make predictions or decisions on new data. Deployment involves integrating the model into software systems and ensuring its scalability, reliability, and security.

# 4. Advantages of Machine Learning

### 1. Automation

Machine Learning is one of the **driving forces** behind automation, and it is cutting down time and human workload. Automation can now be seen everywhere, and the complex algorithm does the hard work for the user. Automation is more reliable, efficient, and quick. With the help of machine learning, now advanced computers are being designed. Now this advanced computer can handle several machine-learning models and complex algorithms. However, automation is spreading faster in the industry but, a lot of research and innovation are required in this field.

### 2. Scope of Improvement

Machine Learning is a field where things keep evolving. It gives many opportunities for improvement and can become the leading technology in the future. A lot of research and innovation is happening in this technology, which helps improve software and hardware.

### 3. Enhanced Experience in Online Shopping and Quality Education

Machine Learning is going to be used in the education sector extensively, and it will be going to enhance the quality of education and student experience. It has emerged in China; machine learning has improved student focus. In the e-commerce field, Machine Learning studies your search **feed and give suggestion** based on them. Depending upon search and browsing history, it pushes targeted advertisements and notifications to users.

### 4. Wide Range of Applicability

This technology has a very wide range of applications. Machine learning plays a role in almost every field, **like hospitality, ed-tech, medicine, science, banking, and business**. It creates more opportunities.

# 5. Disadvantages of Machine Learning

### 1. Data Acquisition

The whole concept of machine learning is about identifying useful data. The outcome will be incorrect if a credible data source is not provided. The quality of the data is also significant. If the user or institution needs more quality data, wait for it. It will cause delays in providing the output. So, machine learning significantly depends on the data and its quality.

### 2. Time and Resources

The data that machines process remains huge in quantity and differs greatly. Machines require time so that their algorithm can adjust to the environment and learn it. Trials runs are held to check the accuracy and reliability of the machine. It requires massive and expensive resources and high-quality expertise to set up that quality of infrastructure. Trials runs are costly as they would cost in terms of time and expenses.

### 3. Results Interpretations

One of the biggest advantages of Machine learning is that interpreted data that we get from the cannot be hundred percent accurate. It will have some degree of inaccuracy. For a high degree of accuracy, algorithms should be developed so that they give reliable results.

### 4. High Error Chances

The error committed during the initial stages is huge, and if not corrected at that time, it creates havoc. Biasness and wrongness have to be dealt with separately; they are not interconnected. Machine learning depends on two factors, **i.e., data and algorithm**. All the errors are dependent on the two variables. Any incorrectness in any variables would have huge repercussions on the output.

# 6. Essential Prerequisites and Steps for Learning Machine Learning

1. Understand the Fundamentals:

   Start with introductory materials such as online courses, tutorials, or textbooks to grasp fundamental concepts and terminologies in machine learning.

   Familiarize yourself with different types of machine learning problems and algorithms.

2. Math Review and Refresh:

   Review key mathematical concepts required for machine learning such as linear algebra, calculus, and probability/statistics.

   Practice solving mathematical problems related to machine learning to strengthen your understanding.

3. Programming Skills Development:

   Learn Python programming language if you haven't already, as it's widely used in the machine learning community.

   Practice coding machine learning algorithms from scratch to gain a deeper understanding of how they work.

4. Explore Libraries and Frameworks:

   Get hands-on experience with popular machine learning libraries and frameworks such as Scikit-learn, TensorFlow, and PyTorch.

   Experiment with example datasets and work through tutorials to understand how to apply machine learning techniques using these libraries.

5. Work on Projects:

   Start working on small machine learning projects to apply what you've learned and gain practical experience.

   Choose projects that align with your interests and allow you to explore different types of machine learning problems and techniques.

6. Experiment and Iterate:

   Experiment with different algorithms, hyperparameters, and feature engineering techniques to understand their impact on model performance.

   Iterate on your projects by continuously refining your models based on feedback and evaluation results.

# 7. Modules for machine learning

## Best Python libraries for Machine Learning

- ➤ Machine learning is a science of programming the computer by which they can learn from different types of data.
- ➤ According to machine learning's definition of **Arthur Samuel** - "Field of study that gives computers the ability to learn without being explicitly programmed".
- ➤ The concept of machine learning is basically used for solving different types of life problems.



- ➤ In previous days, the users used to perform tasks of machine learning by manually coding all the algorithms and using mathematical and statistical formulas.
- ➤ This process was time-consuming, inefficient, and tiresome compared to Python libraries, frameworks, and modules.
- ➤ But in today's world, users can use the Python language, which is the most popular and productive language for machine learning.
- ➤ Python has replaced many languages as it is a vast collection of libraries, and it makes work easier and simpler.

  - o NumPy
  - o Scikit-learn
  - o Theano
  - o TensorFlow
  - o Pandas
  - o Matplotlib

# 8. NumPy Library



➤ NumPy is the most popular library in Python.
➤ This library is used for processing large multi-dimensional array and matrix formation by using a large collection of high-level mathematical functions and formulas.
➤ It is mainly used for the computation of fundamental science in machine learning.
➤ It is widely used for linear algebra, Fourier transformation, and random number capabilities.
➤ There are other High-end libraries such as TensorFlow, which user NumPy as internal functioning for manipulation of tensors.

## Install Python NumPy

➤ Numpy can be installed for Mac and Linux users via the following pip command:
➤ pip install numpy
➤ Windows does not have any package manager analogous to that in Linux or Mac. Please download the pre-built Windows installer for NumPy from here (according to your system configuration and Python version). And then install the packages manually.

## Arrays in NumPy

➤ NumPy's main object is the homogeneous multidimensional array.
➤ It is a table of elements (usually numbers), all of the same type, indexed by a tuple of positive integers.
➤ In NumPy, dimensions are called axes. The number of axes is rank.
➤ NumPy's array class is called ndarray. It is also known by the alias array.

# 8. pandas Library



Pandas is a Python library that is mainly used for data analysis.

The users have to prepare the dataset before using it for training the machine learning.

Pandas make it easy for the developers as it is developed specifically for data extraction.

It has a wide variety of tools for analysing data in detail, providing high-level data structures.

## What is Python Pandas used for?

➤ The Pandas library is generally used for data science, but have you wondered why? This is because the Pandas library is used in conjunction with other libraries that are used for data science.

➤ It is built on top of the NumPy library which means that a lot of the structures of NumPy are used or replicated in Pandas.

➤ The data produced by Pandas is often used as input for plotting functions in Matplotlib, statistical analysis in SciPy, and machine learning algorithms in Scikit-learn.

➤ You must be wondering, Why should you use the Pandas Library. Python's Pandas library is the best tool to analyze, clean, and manipulate data.

➤ Here is a list of things that we can do using Pandas.

➤ Data set cleaning, merging, and joining.

➤ Easy handling of missing data (represented as NaN) in floating point as well as non-floating point data.

➤ Columns can be inserted and deleted from DataFrame and higher-dimensional objects.

➤ Powerful group by functionality for performing split-apply-combine operations on data sets.

# 9. Matplotlib Library
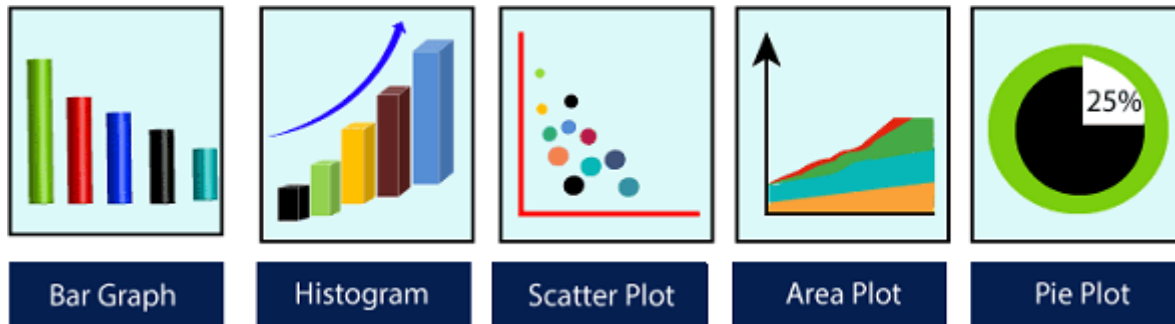


- ➢ Matplotlib is a Python library that is used for data visualization.
- ➢ It is used by developers when they want to visualize the data and its patterns.
- ➢ It is a 2-D plotting library that is used to create 2-D graphs and plots.
- ➢ It has a module pyplot which is used for plotting graphs, and it provides different features for control line styles, font properties, formatting axes and many more.
- ➢ Matplotlib provides different types of graphs and plots such as histograms, error charts, bar charts and many more.

## Types of Matplotlib

- ➢ Matplotlib comes with a wide variety of plots. Plots help to understand trends, and patterns, and to make correlations. They're typically instruments for reasoning about quantitative information. Some of the sample plots are covered here.
    1. Matplotlib Line Plot
    2. Matplotlib Bar Plot
    3. Matplotlib Histograms Plot
    4. Matplotlib Scatter Plot
    5. Matplotlib Pie Charts
    6. Matplotlib Area Plot

# 10. Applications of Machine learning

Machine learning is a buzzword for today's technology, and it is growing very rapidly day by day.

We are using machine learning in our daily life even without knowing it such as Google Maps, Google assistant, Alexa, etc. Below are some most trending real-world applications of Machine Learning:
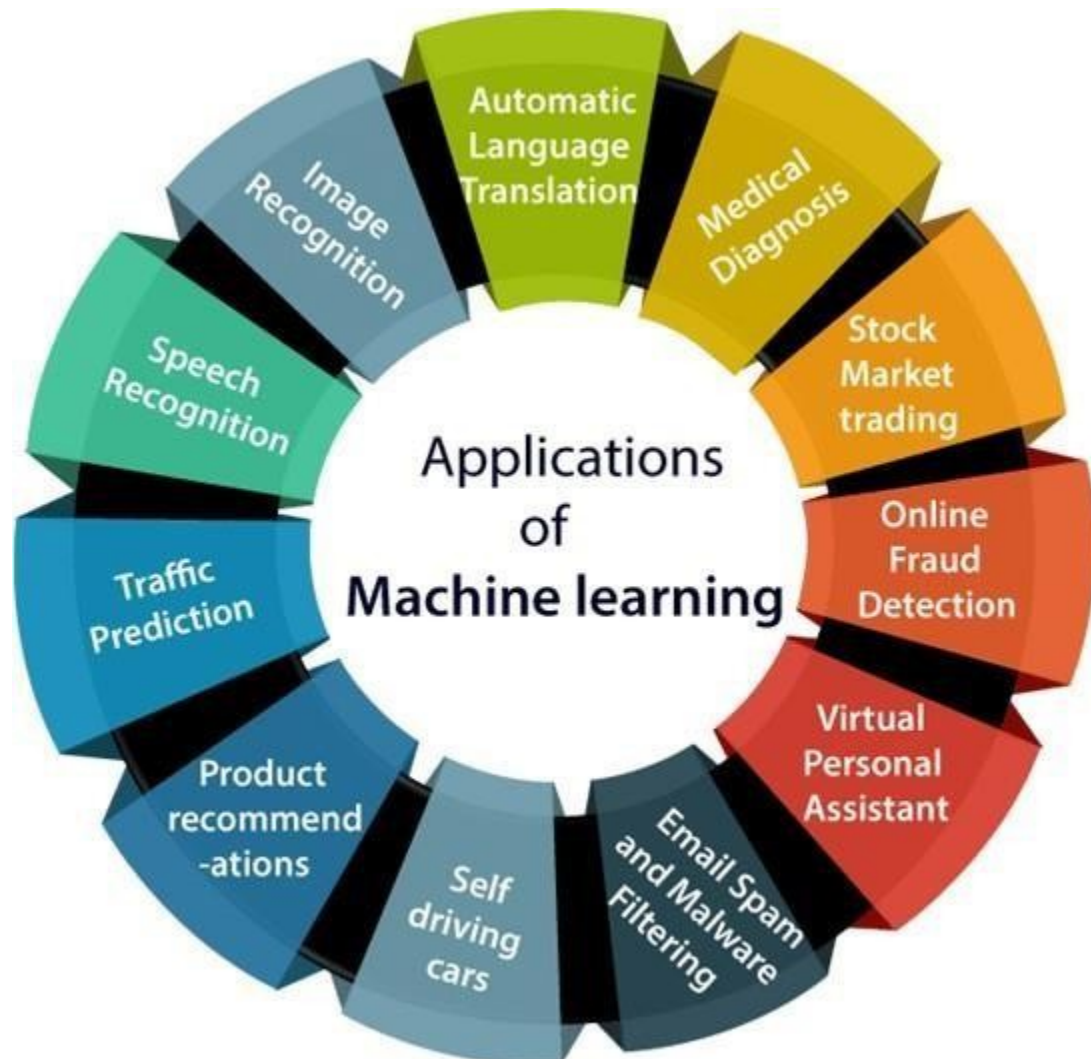


## 1. Image Recognition:

➢ Image recognition is one of the most common applications of machine learning. It is used to identify objects, persons, places, digital images, etc.
➢ The popular use case of image recognition and face detection is, **Automatic friend tagging suggestion**:
➢ Facebook provides us a feature of auto friend tagging suggestion.

- Whenever we upload a photo with our Facebook friends, then we automatically get a tagging suggestion with name, and the technology behind this is machine learning's **face detection** and **recognition algorithm**.
- It is based on the Facebook project named "**Deep Face**," which is responsible for face recognition and person identification in the picture.

## 2. Speech Recognition

- While using Google, we get an option of "**Search by voice**," it comes under speech recognition, and it's a popular application of machine learning.
- Speech recognition is a process of converting voice instructions into text, and it is also known as "**Speech to text**", or "**Computer speech recognition**." At present, machine learning algorithms are widely used by various applications of speech recognition.
- **Google assistant**, **Siri**, **Cortana**, and **Alexa** are using speech recognition technology to follow the voice instructions.

## 3. Traffic prediction:

- If we want to visit a new place, we take help of Google Maps, which shows us the correct path with the shortest route and predicts the traffic conditions.
- It predicts the traffic conditions such as whether traffic is cleared, slow-moving, or heavily congested with the help of two ways:

    - **Real Time location** of the vehicle form Google Map app and sensors
    - **Average time has taken** on past days at the same time.

- Everyone who is using Google Map is helping this app to make it better. It takes information from the user and sends back to its database to improve the performance.

## 4. Product recommendations:

- Machine learning is widely used by various e-commerce and entertainment companies such as **Amazon**, **Netflix**, etc., for product recommendation to the user. Whenever we search for some product on Amazon, then we started getting an advertisement for the same product while internet surfing on the same browser and this is because of machine learning.
- Google understands the user interest using various machine learning algorithms and suggests the product as per customer interest.
- As similar, when we use Netflix, we find some recommendations for entertainment series, movies, etc., and this is also done with the help of machine learning.

## 5. Self-driving cars:

➤ One of the most exciting applications of machine learning is self-driving cars.
➤ Machine learning plays a significant role in self-driving cars.
➤ Tesla, the most popular car manufacturing company is working on self-driving car.
➤ It is using unsupervised learning method to train the car models to detect people and objects while driving.

## 6. Email Spam and Malware Filtering:

➤ Whenever we receive a new email, it is filtered automatically as important, normal, and spam.
➤ We always receive an important mail in our inbox with the important symbol and spam emails in our spam box, and the technology behind this is Machine learning.
➤ Below are some spam filters used by Gmail:

   o   Content Filter

   o   Header filter

   o   General blacklists filter

   o   Rules-based filters

   o   Permission filters

➤ Some machine learning algorithms such as **Multi-Layer Perceptron**, **Decision tree**, and **Naïve Bayes classifier** are used for email spam filtering and malware detection.

## 7. Virtual Personal Assistant:

➤ We have various virtual personal assistants such as **Google assistant**, **Alexa**, **Cortana**, **Siri**.
➤ As the name suggests, they help us in finding the information using our voice instruction.
➤ These assistants can help us in various ways just by our voice instructions such as Play music, call someone, Open an email, Scheduling an appointment, etc.
➤ These virtual assistants use machine learning algorithms as an important part.
➤ These assistant record our voice instructions, send it over the server on a cloud, and decode it using ML algorithms and act accordingly.

## 8. Online Fraud Detection:

➤ Machine learning is making our online transaction safe and secure by detecting fraud transaction.
➤ Whenever we perform some online transaction, there may be various ways that a fraudulent transaction can take place such as **fake accounts**, **fake ids**, and **steal money** in the middle of a transaction.
➤ So to detect this, **Feed Forward Neural network** helps us by checking whether it is a genuine transaction or a fraud transaction.
➤ For each genuine transaction, the output is converted into some hash values, and these values become the input for the next round.
➤ For each genuine transaction, there is a specific pattern which gets change for the fraud transaction hence, it detects it and makes our online transactions more secure.

## 9. Stock Market trading:

➤ Machine learning is widely used in stock market trading.
➤ In the stock market, there is always a risk of up and downs in shares, so for this machine learning's **long short term memory neural network** is used for the prediction of stock market trends.

## 10. Medical Diagnosis:

➤ In medical science, machine learning is used for diseases diagnoses. With this, medical technology is growing very fast and able to build 3D models that can predict the exact position of lesions in the brain
➤ It helps in finding brain tumors and other brain-related diseases easily.

## 11. Automatic Language Translation:

➤ Nowadays, if we visit a new place and we are not aware of the language then it is not a problem at all, as for this also machine learning helps us by converting the text into our known languages.
➤ Google's GNMT (Google Neural Machine Translation) provide this feature, which is a Neural Machine Learning that translates the text into our familiar language, and it called as automatic translation.
➤ The technology behind the automatic translation is a sequence to sequence learning algorithm, which is used with image recognition and translates the text from one language to another language.
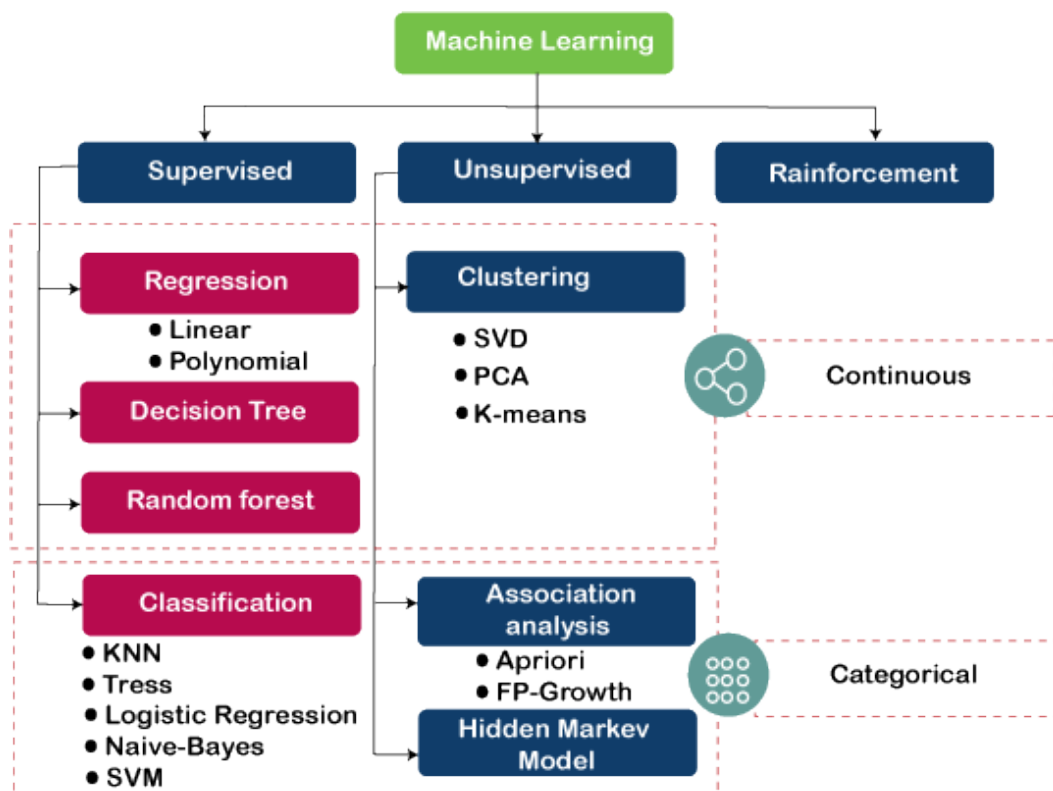
# 10. Machine Learning Algorithms

➤ Machine Learning algorithms are the programs that can learn the hidden patterns from the data, predict the output, and improve the performance from experiences on their own.

➤ Different algorithms can be used in machine learning for different tasks, such as simple linear regression that can be used **for prediction problem**s like **stock market prediction,** and **the KNN algorithm can be used for classification problems**.

➤ In this topic, we will see the overview of some popular and most commonly used machine learning algorithms along with their use cases and categories.

## Types of Machine Learning Algorithms

Machine Learning Algorithm can be broadly classified into three types:

1. **Supervised Learning Algorithms**

2. **Unsupervised Learning Algorithms**

3. **Reinforcement Learning algorithm**

The below diagram illustrates the different ML algorithm, along with the categories:

# 1) Supervised Learning Algorithm

➤ Supervised learning is a type of Machine learning in which the machine needs external supervision to learn.

➤ The supervised learning models are trained using the labeled dataset. Once the training and processing are done, the model is tested by providing a sample test data to check whether it predicts the correct output.

➤ The goal of supervised learning is to map input data with the output data. Supervised learning is based on supervision, and it is the same as when a student learns things in the teacher's supervision. The example of supervised learning is **spam filtering**.

➤ Supervised learning can be divided further into two categories of problem:

  o [Classification](#)

  o [Regression](#)

# 2) Unsupervised Learning Algorithm

➤ It is a type of machine learning in which the machine does not need any external supervision to learn from the data, hence called unsupervised learning.

➤ The unsupervised models can be trained using the unlabelled dataset that is not classified, nor categorized, and the algorithm needs to act on that data without any supervision.

➤ In unsupervised learning, the model doesn't have a predefined output, and it tries to find useful insights from the huge amount of data. These are used to solve the Association and Clustering problems. **Hence further, it can be classified into two types:**

  o [Clustering](#)

  o Association

# 3) Reinforcement Learning

➤ In Reinforcement learning, an agent interacts with its environment by producing actions, and learn with the help of feedback.

➤ The feedback is given to the agent in the form of rewards, such as for each good action, he gets a positive reward, and for each bad action, he gets a negative reward.