Contents

Unit-1 : Introduction to Machine Learning 1-11

1.1 Overview of Human Learning and Machine Learning 1
1.2 Types of Machine Learning 4
1.3 Applications of Machine Learning 7
1.4 Tools and Technology for Machine Learning 9
■ **Question Bank** 10

Unit-2 : Preparing to model 12-22

2.1 Machine Learning activities 12
2.2 Types of data in Machine Learning 13
2.3 Structures of data 14
2.4 Data quality and remediation 17
2.5 Data Pre-Processing 18
2.5.1 Dimensionality reduction 19
2.5.2 Feature subset selection 20
■ **Question Bank** 21

Unit-3 : Modeling and evaluation 23-37

3.1 Selecting a Model (Predictive/Descriptive) 23
3.2 Training a Model for supervised learning: 24
3.2.1 Holdout method 24
3.2.2 K-fold Cross-validation method 25
3.3 Model representation and interpretability 27
3.4 Evaluating performance of a model 29
3.4.1 Confusion Matrix 29
3.5 Improving Performance of a model 31
■ **Question Bank** 35

Unit-4 : Supervised Learning: Classification and Regression 38-50

4.1 Introduction to supervised learning 38
4.2 Classification Model 38
4.3 Learning Steps 40
4.4 Classification Algorithms 41
4.4.1 k-Nearest Neighbor (kNN) 41
4.4.2 Support Vector Machines 43
4.5 Regression 44
4.5.1 Simple linear regression 44
4.5.2 Multiple linear regression 46
4.5.3 Logistic Regression 47
■ **Question Bank** 49

Unit-5 : Unsupervised Learning 51-62

5.1 Supervised vs. Unsupervised Learning 51
5.2 Applications of unsupervised learning 52
5.3 Clustering 53
5.3.1 K-means Clustering algorithm 54
5.4 Finding pattern using Association Rule 57
5.4.1 Apriori algorithm 57

■ **Question Bank** 61

Unit-6 : Python libraries for Machine Learning 63-73

6.1 Pandas 63
6.2 Numpy 65
6.3 Matplotlib 66
6.4 Scikit-learn 69

■ **Question Bank** 72

• **Lab Manual for Machine Learning** 74-107

PRACTICAL : 1 75
PRACTICAL : 2 81
PRACTICAL : 3 84
PRACTICAL : 4 88
PRACTICAL : 5 90
PRACTICAL : 6 91
PRACTICAL : 7 93
PRACTICAL : 8 95
PRACTICAL : 9 96
PRACTICAL : 10 98
PRACTICAL : 11 101
PRACTICAL : 12 103

• **ABBREVIATION** 108

• **Model Paper for Mid Semester Exam** 109-110

PAPER : 1 109
PAPER : 2 109
PAPER : 3 110

• **Model Paper for Semester Annual Exam** 111-116

PAPER : 1 111
PAPER : 2 113
PAPER : 3 115

❖ ❖ ❖

1 Introduction to Machine Learning

1.1 Overview of Human Learning and Machine Learning

1.1.1 Human Learning :

1.1.2 Types of Human Learning:

1.1.3 Machine Learning:

1.1.4 How do machines learn?

1.1.5 Well-posed learning problem

1.2 Types of Machine Learning

1.2.1 Supervised Learning

1.2.2 Unsupervised Learning

1.2.3 Reinforcement Learning

1.2.4 Difference between Supervised, unsupervised and reinforcement learning

1.3 Applications of Machine Learning

1.4 Tools and Technology for Machine Learning

■ Question Bank

1.1 Overview of Human Learning and Machine Learning

1.1.1 Human Learning :

Human Learning is the process of gaining information through observations.

In our day to day life, we perform several activities. To do the activity in a proper way, we must have proper information related to that activity. Also, we always try to improve our activity by gaining more information or by previous experience related to that one. So, with more learning, we can make our activity more efficient.

1.1.2 Types of Human Learning:

1. Learning under expert guidance

   A kid shows her head when we asked, "Where is your head?" Because, this thing she learns from her parents.

   Guided learning is the process of gaining information from a person who has sufficient knowledge due to the past experience.

2. Learning guided by knowledge gained from experts

   A kid can group together only those toys which belong to the group of "cars". Here, the kid has the only knowledge regarding which toy is a car. No one taught her to make a group of cars.

   It is some past knowledge shared on some different way which is used as a learning to make decisions.

3. Learning by self

The kids learn to walk through obstacles is the appropriate example for learning by self. Using past experience, observations; one can learn by self.

It is direct learning.

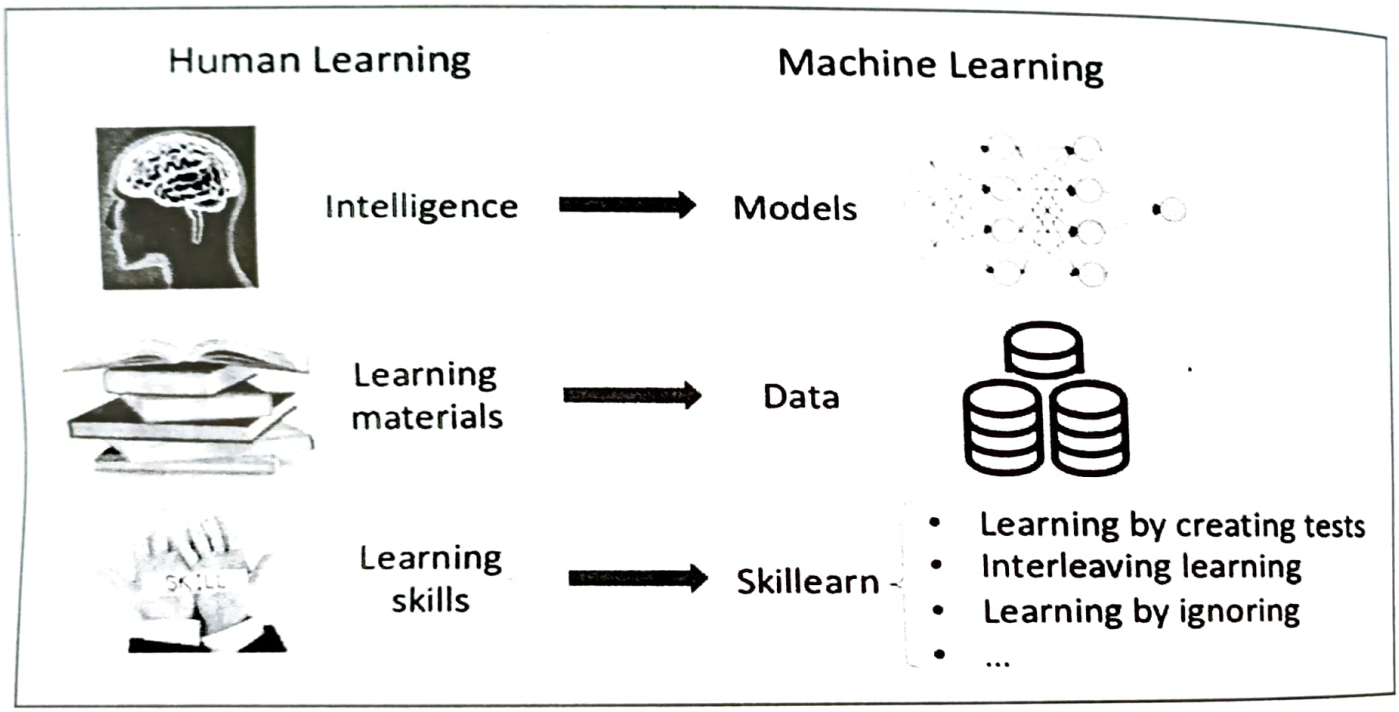
1.1.3 Machine Learning:

Machine Learning (ML) is basically that field of computer science with the help of which computer systems can provide sense to data in much the same way as human beings do.

Machine learning is a subfield of Artificial Intelligence (AI) which extracts patterns out of raw data by using an appropriate algorithm.

Its goal is to enable computers to learn on their own from experience.

A machine's learning algorithm enables it to identify patterns in observed data, build models that explain the world, and predict things without having explicit pre-programmed rules and models.



Machine Learning (ML) is a field of AI consisting of learning algorithms that improve their performance (P) at executing some task (T) over time with experience (E).

As per Tom M. Mitchell, Machine Learning is defined as , "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."

For example, **Robot Driving in a maze.**

- Class of task: To reach at the end point
- Performance Measure: Time taken to reach at the end point
- Source of experience: Navigating the maze from start to end by the robot

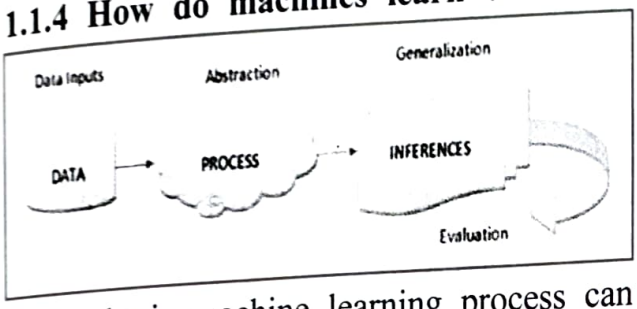
Remember that,

Task (T): It is a real world problem. Classification, clustering, regression are the examples of ML based task.

Experience (E): It is the knowledge gained from data points provided to the algorithm or model. The experience gained by out ML model or algorithm will be used to solve the task T.

Performance (P): The measure which tells whether ML algorithm is performing as per expectation or not is its performance. An ML algorithm is supposed to perform task and gain experience with the defined time period.

1.1.4 How do machines learn ?



The basic machine learning process can be divided in four parts:

1. **Data Inputs**:

   Past data is used as a base for future prediction.

   The vast pool of knowledge is available from the input data. The machine can use the past data for decision-making process.

2. **Abstraction**:

   The input data is represented in a broader way using appropriate algorithms.

Abstraction helps to derive a model which is a summarized knowledge representation of the raw data. The model may be a mathematical equation, rule-based computational blocks, logical combinations of same observations or well-defined data structures.

The decision related to choice of a model is based on the type of a problem, nature of input data, domain of the problem, and many more.

In this part, we get the finite set of data which is known as **training data** set.

Once model is chosen then the next step is to fit the model based on the input received.

3. **Generalization**:

   The abstracted representation is generalized to create a framework for making decisions.

Generalization is the most critical part in ML process.

In this part, the abstracted knowledge will be tuned up to a form which can be used to take decisions.

We want to apply the model to take decision on a set of unknown data which is called **test data** set. For decision-making, approximate/heuristic approach has to be taken.

4. **Evaluation:**

   Evaluation is the process of giving feedback to the user to measure the utility of the learned knowledge. This feedback is then utilized to effect improvements in the whole learning process.

1.1.5 Well-posed learning problem

A (machine learning) problem is well-posed if a solution to it exists, if that solution is unique, and if that solution depends on the data / experience but it is not sensitive to (reasonably small) changes in the data / experience.

To have a well-defined learning problem, three features needs to be identified:

1. The class of tasks (T)
2. The measure of performance to be improved (P)

3. The source of experience (E)

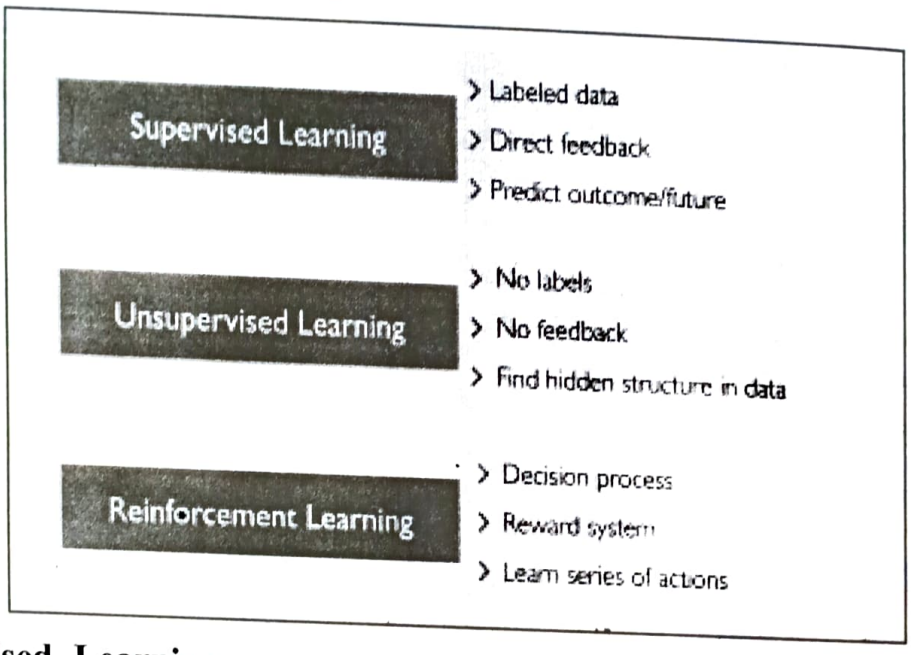
Example, **Face Recognition Problem**

- Class of Task: predicting different types of faces
- Performance Measure: able to predict maximum types of faces
- Source of Experience: training machine with maximum amount of datasets of different face images

To solve well-posed problem using Machine Learning, a simple framework this involves three questions to answer:

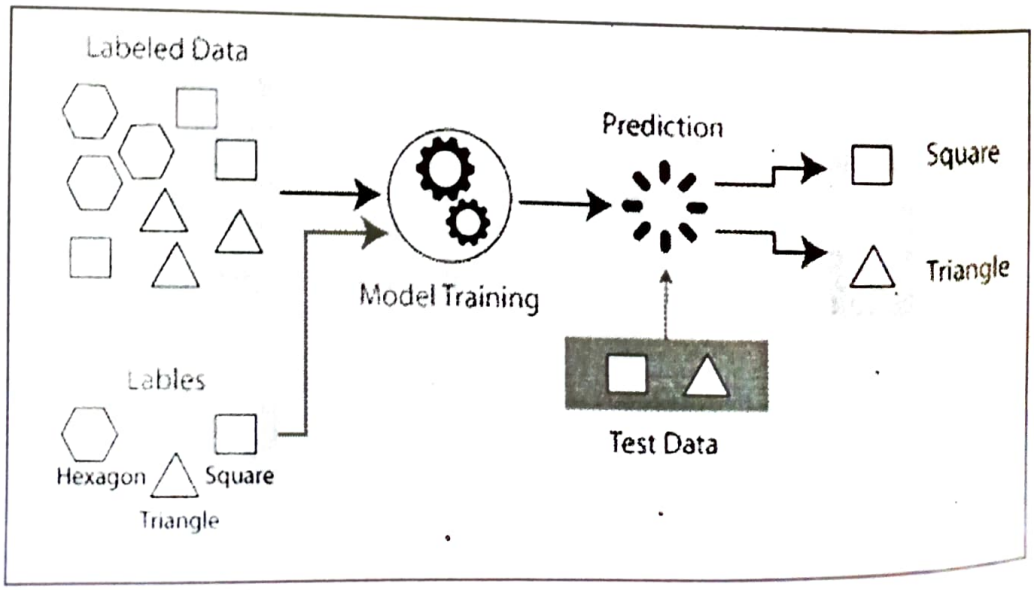
1. What is the problem? (Formalism the problem based on Task, Experience and Performance)
2. Why does the problem need to be solved? (Motivation and benefits to solve the problem)
3. How to solve the problem? (Creating the model/algorithm using knowledge)

1.2 Types of Machine Learning



1.2.1 Supervised Learning

Supervised learning is the types of machine learning in which machines are trained using well "labelled" training data, and on basis of that data, machines predict the output.



The labelled data means some input data is already tagged with the correct output.

In supervised learning, models are trained using labelled dataset, where the model learns about each type of data. Once the training process is completed, the model is tested on the basis of test data and then it predicts the output.

Supervised learning is a process of providing input data as well as correct output data to the machine learning model. The aim of a supervised learning algorithm is to find a mapping function to map the input variable(x) with the output variable(y).

In the real-world, supervised learning can be used for **Risk Assessment, image classification, Fraud Detection, spam filtering, etc.**

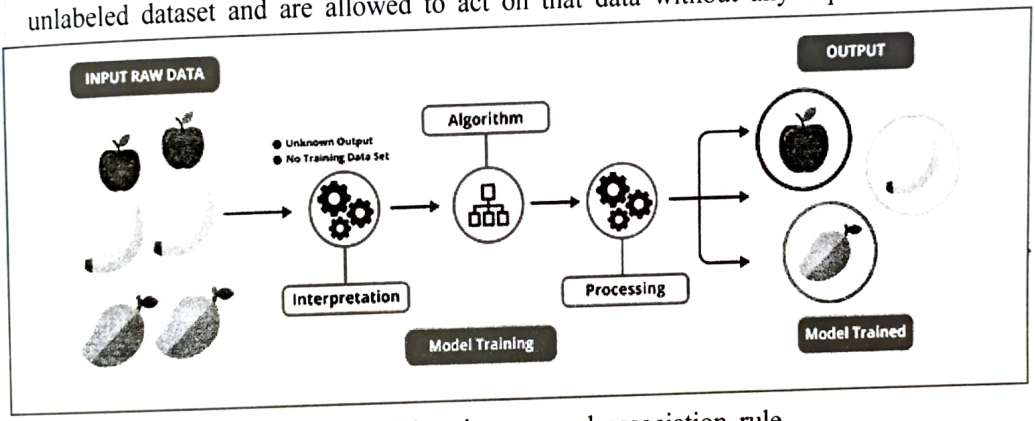
Regression and classification are two main areas of Supervised Learning.

Regression algorithms are used if there is a relationship between the input variable and the output variable. It is used for the prediction of continuous variables, such as Weather forecasting, Market Trends, etc.

Classification algorithms are used when the output variable is categorical, which means there are two classes such as Yes-No, Male-Female, True-false, etc.

1.2.2 Unsupervised Learning

Unsupervised learning is a type of machine learning in which models are trained using unlabeled dataset and are allowed to act 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.

In the real world, unsupervised learning can be used for pattern recognition, inventory management, data exploration, recommender system, etc.

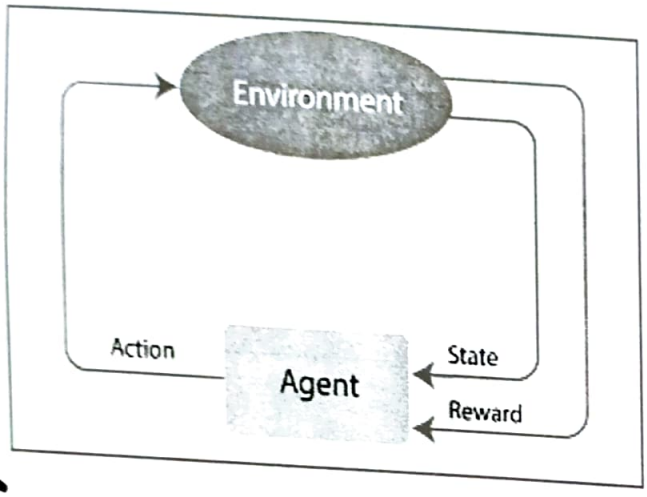
Unsupervised learning is classified into two categories of algorithms : Clustering and association rule.

- **Clustering**: A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior.
- **Association**: An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

1.2.3 Reinforcement Learning :

Reinforcement learning is a type of machine learning method where an intelligent agent (computer program) interacts with the environment and learns to act within that.

In general, a reinforcement learning agent is able to perceive and interpret its environment, take actions and learn through trial and error.



The Reinforcement Learning problem involves an agent exploring an unknown environment to achieve a goal. It is based on the hypothesis that all goals can be described by the maximization of expected cumulative reward. The agent must learn to sense and perturb the state of the environment using its actions to derive maximal reward.

There are four elements of reinforcement learning: policy, reward function, value function and model of the environment.

1. **Policy :** A set of rules or strategies that the agent uses to decide what action to take in a particular state. The policy is updated based on the rewards received by the agent.
2. **Reward Function :** A reward function is a function that provides a numerical score based on the state of the environment. If it makes the right decision, it gets a reward. If it makes mistakes, it receives a penalty. A reward is a positive reinforcement that the agent receives for taking an action that brings it closer to its goal. A penalty is a negative reinforcement that the agent receives for taking an action that moves it further away from its goal.
3. **Value function :** A function that estimates the expected cumulative reward that the agent will receive if it follows a particular policy from a particular state. The value function is used to evaluate different policies and select the best one.
4. **Model of the environment :** In reinforcement learning, the environment is modeled as a Markov Decision Process (MDP) which includes states and actions.

States : The environment can be in one of a finite number of states. A state is a snapshot of the environment at a given time, and it contains all the information the agent needs to make a decision.

Actions : At each state, the agent can take one of a finite number of actions. An action is a decision made by the agent that causes the environment to transition to a new state.

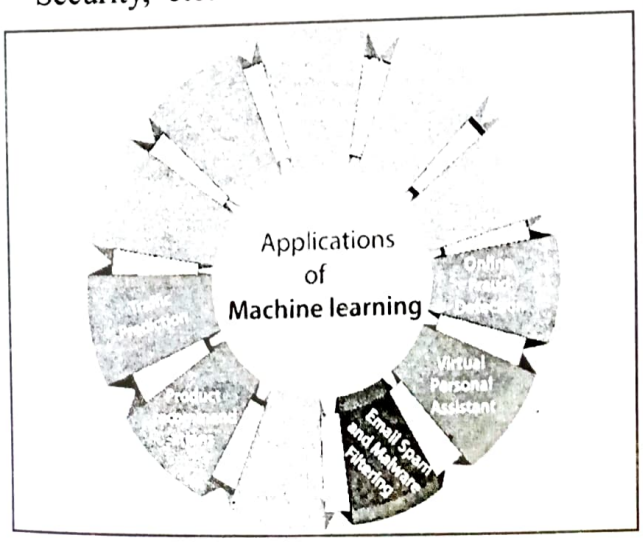
In the real world, reinforcement learning can be used for self driving cars, gaming, robotics, etc.

1.2.4 Difference between Supervised Learning, Unsupervised Learning and Reinforcement Learning

Criteria	Supervised ML	Unsupervised ML	Reinforcement ML
Definition	Machine Learns by using labelled data	Machine is trained using unlabelled data without any guidance	Agent interacts with the environment by performing action. Learns by errors and rewards
Type of data	Labelled data	Unlabelled data	No – predefined data
Type of problems	Regression and classification	Association and Clustering	Reward and error based
Supervision	External supervision	No supervision	No supervision
Algorithms	Linear Regression, Logistic Regression, Naive Byes Decision trees	K – Means clustering, KNN (K-nearest neighbours) Principle Component Analysis Neural Networks	Monte Carlo, Q-Learning, SARSA
Aim	Calculate outcomes	Discover underlying patterns	Learn a series of action
Approach	Maps labelled inputs to the known outputs	Understands patterns & discover the output	Follow the trial and error method
Application	Risk Evaluation, Forecast Sales	Recommendation System, Anomaly Detection	Self-Driving Cars, Gaming, Healthcare

1.3 Applications of Machine Learning

Machine learning approach is adopted in multiple forms in every business domain. The major areas are Banking and finance, Insurance, Healthcare, e-Commerce, Marketing and sales, Search engines, Social Media, Virtual assistant, Transport, Security, etc.



- **Healthcare :** Machine learning is used in healthcare to analyze medical images, predict patient outcomes, and develop personalized treatment plans. It is also used in drug discovery and genomics research.

  Machine learning is used in medical imaging to analyze images from X-rays, CT scans, and MRIs. Algorithms can detect patterns and anomalies in the images, which can help with diagnosis and treatment planning.

  Machine learning can be used to analyze patient data and develop personalized treatment plans based on factors such as genetics, lifestyle, and medical history.

- **Banking and Finance :** Machine learning is used in banking and finance for fraud detection, risk assessment, credit scoring, and investment analysis.

Machine learning is used to analyze large datasets to identify fraudulent transactions. Algorithms can detect patterns and anomalies in transaction data, which can help prevent financial crimes such as credit card fraud and money laundering.

Machine learning is used to assess credit risk when evaluating loan applications. Algorithms can analyze a range of data points, such as credit history, income, and employment status, to determine the likelihood of default.

- **Insurance :** Machine learning (ML) is increasingly being used in the insurance industry to improve risk assessment, automate underwriting, and enhance the customer experience.

  Machine learning is used to automate claims processing. Algorithms can analyze claims data and make decisions about coverage, processing claims more quickly and accurately than human adjusters.

- **Marketing :** Machine learning is used in marketing to personalize advertising, improve customer segmentation, and predict customer behavior.

  Machine learning algorithms can analyze customer data to identify patterns and segments, allowing marketers to create targeted campaigns that resonate with specific customer groups. It can analyze customer data to create personalized recommendations and content. This can help improve customer engagement and loyalty.

- **Manufacturing :** Machine learning is used in manufacturing to optimize production processes, predict equipment failure, and improve quality control. It is the heart of Industry 4.0.

  Machine learning algorithms can analyze data from sensors and other sources to predict when maintenance is needed. This can help reduce downtime and improve the reliability of manufacturing equipment. ML is critical for the development of advanced robotics and automation systems. These systems can perform complex tasks and adapt to changes in the manufacturing environment.

- **Transportation :** Machine learning is used in transportation for route optimization, demand forecasting, and autonomous vehicles.

  It can analyze data on traffic patterns, road conditions, and weather to optimize routes for transportation systems. This can help reduce fuel consumption and emissions, as well as improve delivery times.

- **Energy :** Machine learning is used in the energy sector for predictive maintenance, grid optimization, and renewable energy forecasting.

  Machine learning can be used to predict energy demand based on historical data, weather patterns, and other factors. This can help energy companies optimize their production and distribution processes, reduce costs, and improve sustainability. It can analyze data from solar panels, wind turbines, and other renewable energy sources to optimize energy production. This can help improve the efficiency of renewable energy systems and reduce the cost of energy production.

- **Education :** Machine learning is used in education to personalize learning, improve

student engagem
performance.
Machine learnin
student data to
weaknesses, an
This informatio
personalized lea
tailored to the
can be used to
systems that
feedback and
These systems
pace of learnin
support when

- **Agriculture :**
in agriculture
disease detecti

1.4 Tools and T Learning

1. Python

Python is a hi
bject-oriented
huge n
signifi
The most
which is
calculations; S
computations;
mining
consist
which
machine

2. R Progra

R is a procedu
created by s
specifically fo
language for st

Machine learning is used to analyze large datasets to identify fraudulent transactions. Algorithms can detect patterns and anomalies in transaction data, which can help prevent financial crimes such as credit card fraud and money laundering.

Machine learning is used to assess credit risk when evaluating loan applications. Algorithms can analyze a range of data points, such as credit history, income, and employment status, to determine the likelihood of default.

- **Insurance :** Machine learning (ML) is increasingly being used in the insurance industry to improve risk assessment, automate underwriting, and enhance the customer experience.

  Machine learning is used to automate claims processing. Algorithms can analyze claims data and make decisions about coverage, processing claims more quickly and accurately than human adjusters.

- **Marketing :** Machine learning is used in marketing to personalize advertising, improve customer segmentation, and predict customer behavior.

  Machine learning algorithms can analyze customer data to identify patterns and segments, allowing marketers to create targeted campaigns that resonate with specific customer groups. It can analyze customer data to create personalized recommendations and content. This can help improve customer engagement and loyalty.

- **Manufacturing :** Machine learning is used in manufacturing to optimize production processes, predict equipment failure, and improve quality control. It is the heart of Industry 4.0.

  Machine learning algorithms can analyze data from sensors and other sources to predict when maintenance is needed. This can help reduce downtime and improve the reliability of manufacturing equipment. ML is critical for the development of advanced robotics and automation systems. These systems can perform complex tasks and adapt to changes in the manufacturing environment.

- **Transportation :** Machine learning is used in transportation for route optimization, demand forecasting, and autonomous vehicles.

  It can analyze data on traffic patterns, road conditions, and weather to optimize routes for transportation systems. This can help reduce fuel consumption and emissions, as well as improve delivery times.

- **Energy :** Machine learning is used in the energy sector for predictive maintenance, grid optimization, and renewable energy forecasting.

  Machine learning can be used to predict energy demand based on historical data, weather patterns, and other factors. This can help energy companies optimize their production and distribution processes, reduce costs, and improve sustainability. It can analyze data from solar panels, wind turbines, and other renewable energy sources to optimize energy production. This can help improve the efficiency of renewable energy systems and reduce the cost of energy production.

- **Education :** Machine learning is used in education to personalize learning, improve

student engagement, and assess student performance.

Machine learning algorithms can analyze student data to identify their strengths, weaknesses, and learning preferences. This information can be used to create personalized learning experiences that are tailored to the needs of each student. It can be used to create intelligent tutoring systems that provide personalized feedback and guidance to students. These systems can adapt to the student's pace of learning and provide additional support when needed.

- **Agriculture :** Machine learning is used in agriculture for crop yield prediction, disease detection, and soil analysis.

1.4 Tools and Technology for Machine Learning

1. Python

Python is a high-level, general-purpose, object-oriented programming language. **It has a huge number of libraries and frameworks :** The Python language comes with many libraries and frameworks that make coding easy. This also saves a significant amount of time.

The most popular libraries are NumPy, which is used for scientific calculations; SciPy for more advanced computations; and scikit, for learning data mining and data analysis. Python is consistent and is anchored on simplicity, which makes it most appropriate for machine learning.

2. R Programming

R is a procedural programming language created by statisticians for statistics, specifically for working with data. It is a language for statistical computing and data visualizations used widely by business analysts, data analysts, data scientists, and scientists.

The R language has become popular because it lets researchers easily combine different machine learning techniques into a single program.

3. MATLAB

MATLAB (matrix laboratory) makes machine learning easy. With tools and functions for handling big data, as well as apps to make machine learning accessible, MATLAB is an ideal environment for applying machine learning to your data analytics.

In MATLAB it takes fewer lines of code and builds a machine learning or deep learning model, without needing to be a specialist in the techniques. MATLAB provides the ideal environment for machine learning, through model training and deployment.

4. SAS

SAS (Statistical Analysis System) is an enterprise-grade statistical software suite developed by SAS Institute. It is used for data management, advanced analytics, and predictive analytics, and has been one of the leading analytics platforms for a long time.

The SAS software contains all the necessary features required to design, develop, train, and evaluate machine learning models.

5. Julia

Julia is a general purpose programming language well suited for numerical analysis and computational science.

It is an open-source. It has ability to implement high performance machine learning algorithms.

Question Bank

■ Short-Answer Questions (3 or 4 marks)

1. Explain human learning in brief.
2. Define Machine Learning. Also, compare it with Human Learning.
3. Explain classification and regression in brief.
4. Explain clustering and association in brief.
5. Give the difference between supervised learning and unsupervised learning.
6. Explain the concept of penalty and rewards in reinforcement learning.
7. Write a short note on machine learning in finance and banking.
8. Give the difference between Python and R.
9. Identify main three features for following well-posed problem:
   a. Fruit prediction Problem
   b. Handwriting recognition Problem

■ Long-Answer Questions (7 marks)

1. Define well-posed problem. Explain important features that are required to well-define a learning problem.
2. Compare the different types of machine learning.
3. Explain types of machine learning in detail.
4. Define machine learning. Explain any two business applications of Machine Learning in detail.
5. Explain different tools and technology used in Machine Learning.
6. Describe basic concept of Machine Learning and its application.

■ Check your knowledge :

1. What is Machine Learning ?
   a. A field of computer science b. A method of data analysis
   c. A type of algorithm **d. All of the above**
2. A type of learning where the model learns by interacting with an environment and receiving rewards or penalties is called ____.
   a. Supervised Learning b. Unsupervised Learning
   c. Reinforcement Learning d. Machine Learning

3. Which type of machine learning involves no target variable and the model learns by finding patterns in the data ?
   a. Supervised Learning **b. Unsupervised Learning**
   c. Reinforcement Learning d. Machine Learning
4. Which type of machine learning involves a target variable that is used to train the model ?
   a. Supervised Learning b. Unsupervised Learning
   c. Reinforcement Learning d. Machine Learning
5. Which of the following is an application of machine learning in transportation?
   a. Face recognition **b. Predictive maintenance**
   c. Sentiment analysis d. Language translation

■