

Unit-1

FaceBook

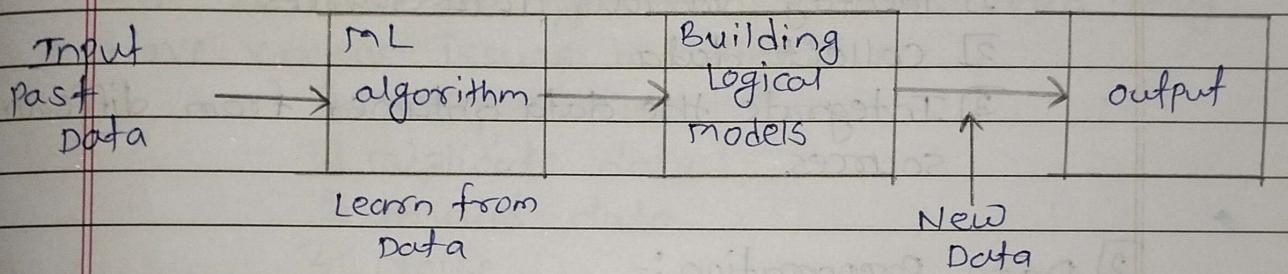
Introduction to Machine Learning

Ques- Describe machine learning life cycle.



Machine Learning :-

Machine Learning (ML) is a subset of artificial intelligence (AI) that involves the development of algorithms and statistical models that enable computers to perform tasks without explicit instructions.



Machine learning life cycle :-

- Machine Learning life cycle involves 7 major steps :-
 - 1) Gathering Data
 - 2) Data preparation
 - 3) Data Wrangling
 - 4) Analyse Data
 - 5) Train the model
 - 6) Test the model
 - 7) Deployment

1) Gathering Data :-

- Data Gathering is the first step of the machine learning life cycle.
- The goal of this step is to identify and obtain all data-related problems.

D	D	M	M	T	T	Y

- In this step, we need to identify the different data sources, as data can be collected from various sources such as files, database, internet or mobile devices.
- The quantity and quality of the collected data will determine the efficiency of the output.
- The more will be the data, the more accurate will be the prediction.
- This step includes the below tasks :-
 - 1] Identify various data sources
 - 2] Collect data
 - 3] Integrate the data obtained from different sources.

2) Data Preparation :-

- ~~After~~ Data preparation is a step where we put our data into a suitable place and prepare it to use in our machine learning training.

- In this step, first we put all data together and then randomize the ordering of data.

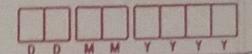
- This step can be further divided into two processes :-

1) Data exploration :-

- It is used to understand the nature of data that we have to work with: we need to understand the characteristics, format and quality of Data.

2) Data Pre-processing :-

- The Pre-processing of data for its analysis.



③ Data wrangling :-

- Data wrangling is the process of cleaning and converting raw data into a useable format.

- It is the process of cleaning the data, selecting the variable to use, and transforming the data in a proper format to make it more suitable for analysis in the next step.

- Cleaning of data is required to address the quality issues. -

- In real-world applications, collected data may have various issues including :-

- 1) Missing values
- 2) Duplicate data
- 3) Invalid data
- 4) Noise

- So, we use various filtering techniques to clean the data.

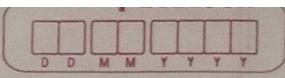
④ Data Analysis :-

- This steps involves :-

- 1) Selection of analytical techniques
- 2) Building models
- 3) Review the result

- The aim of this step is to build a machine learning model to analyze the data using various analytical techniques and review the outcome.

- It starts with the determination of the type of the problems, where we select the machine learning techniques such as classification, regression, cluster analysis, Association etc. then build the model using prepared data and evaluate the model.



Hence, in this step, we take the data and use machine learning algorithms to build the model.

5) Train Model :-

- In this step we train our model to improve its performance for better outcome of the problem.
- We use datasets to train the model using various machine learning algorithms.
- Training a model is required so that it can understand the various patterns, rules and features.

6) Test Model :-

- In this step, we check for the accuracy of our model by providing a test dataset to it.
- Testing the model determines the percentage accuracy of the model as per the requirement of project or problem.

7) Deployment :-

- The last step of machine learning life cycle is deployment, where we deploy the model in real-world system.
- If the above-prepared model is producing an accurate result as per our requirement with acceptable speed, then we deploy the model in the real system.
- The deployment phase is similar to making the final report for a project.

Ques- Explain Types of machine Learning with example.



Types of machine Learning :-

- 1) Supervised Machine Learning
- 2) Unsupervised Machine Learning
- 3) Semi-supervised Machine Learning
- 4) Reinforcement learning.

1) Supervised Machine Learning :-

- Supervised Learning is defined as when a model gets trained on a "Labelled Dataset".

- Labelled datasets have both input and output parameters.

- In supervised learning algorithms learn to map points between inputs and correct outputs.

- It has both training and validation datasets labelled.

INPUT

Labeled
Data

Labels
cow camel
Elephant

OUTPUT

Elephant

Camel

Cow

Algorithm

processing

- There are two main categories of supervised learning :-

- 1) Classification
- 2) Regression

D	D	M	M	Y

Example :-

- consider a scenario where you have to build an image classifier to differentiate between cats and dogs.
- If you feed the datasets of dogs and cats labelled images to the algorithm, the machine will learn to classify between a dog or a cat from these labeled images.
- When we input new dog or cat images that it has never seen before, it will use the learned algorithms and predict whether it is a dog or a cat.
- This is how supervised learning works and this is particularly an image classification.

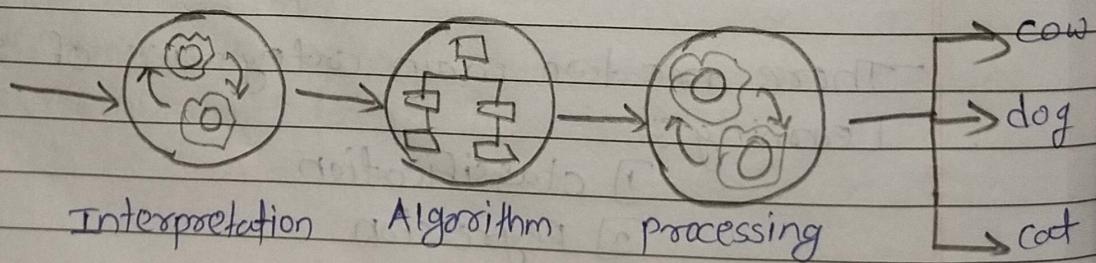
② unsupervised machine learning :-

- unsupervised learning is a type of machine learning technique in which an algorithm discovers patterns and relationships using unlabeled data.
- unsupervised learning doesn't involve providing the algorithm with labeled target outputs.
- The primary goal of unsupervised learning is often to discover hidden patterns, similarities or clusters within the data, which can then be used for various purposes such as data exploration, visualization, dimensionality reduction.

INPUT RAW

DATA

cow,
dog,
cat





There are two main categories of unsupervised

Learning :-

- 1) Clustering
- 2) Association

Example :-

- consider that you have a dataset that contains information about the purchases you made from the shop.
- Through clustering, the algorithm can group the same purchasing behavior among you and other customers, which reveals potential customers without predefined labels.
- This type of information can help businesses get target customers as well as identify outliers.

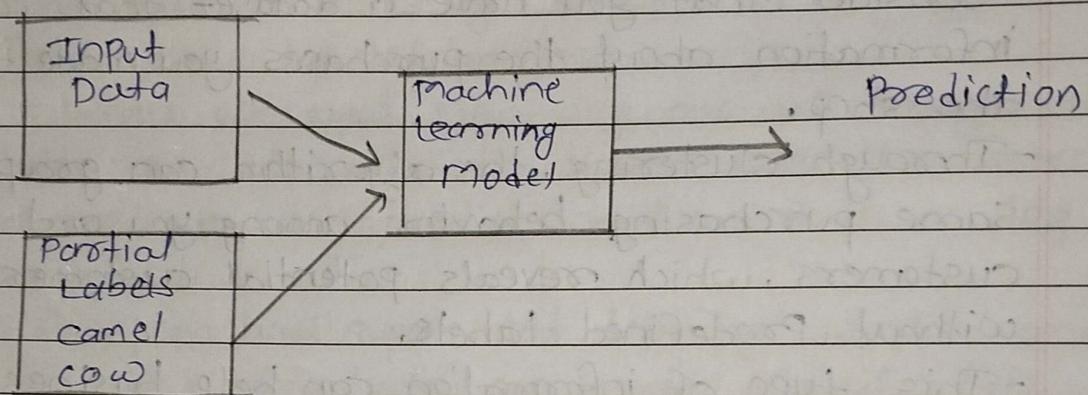
③ Semi-supervised Learning :-

- semi-supervised learning is a machine learning algorithm that works between the supervised and unsupervised learning so it uses both labelled and unlabelled data.
- It's particularly useful when obtaining labeled data is costly, time-consuming or resource-intensive.
- This approach is useful when the dataset is expensive and time-consuming.
- semi-supervised learning is chosen when labeled data requires skills and relevant resources in order to train or learn from it.
- we use these techniques when we are dealing with data that is a little bit labeled and the rest large portion of it is unlabeled.
- We can use the unsupervised techniques to predict

DDMMYY

labels and then feed these labels to supervised techniques.

-this technique is mostly applicable in the case of image data sets where usually all images are not labeled.



Types of semi-supervised learning methods:-

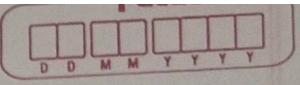
- 1] Graph-based semi-supervised learning
- 2] Label propagation
- 3] Co-training
- 4] Self-training.

Example:-

-consider that we are building a language translation model, having labeled translations for every sentence pair can be resources intensive.

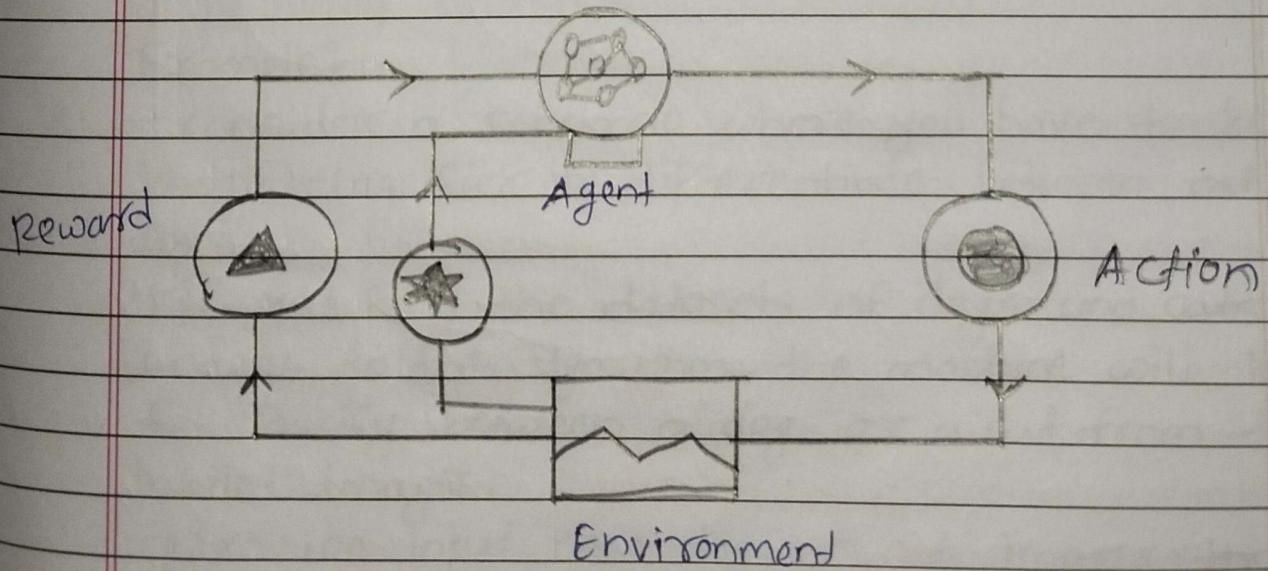
-It allows the models to learn from labeled and unlabeled sentence pairs, making them more accurate.

-This technique has led to significant improvement in the quality of machine translation services.



4) Reinforcement Machine Learning :-

- Reinforcement machine learning algorithm is a learning method that interacts with the environment by producing actions and discovering errors.
- Trial, error and delay are the most relevant characteristics of reinforcement learning.
- In this technique, the model keeps on increasing its performance using reward feedback to learn the behavior or pattern.
- These algorithms are specific to a particular problem e.g. Google self driving car, Alpha Go where a bot competes with humans and even itself to get better and better performers in Go Game.
- Each time we feed in data, they learn and add the data to their knowledge which is training data.
- So, the more it learns the better it gets trained and hence experienced.



Example :-

- consider that you are training an AI agent to play a game like chess.
- The agent explores different moves and receives positive or negative feedback based on the outcome.
- Reinforcement Learning also finds applications in which they learn to perform tasks by interacting with their surroundings.

Types of Reinforcement machine Learning :-

- ① Positive reinforcement
- ② Negative reinforcement.

DDMMYY

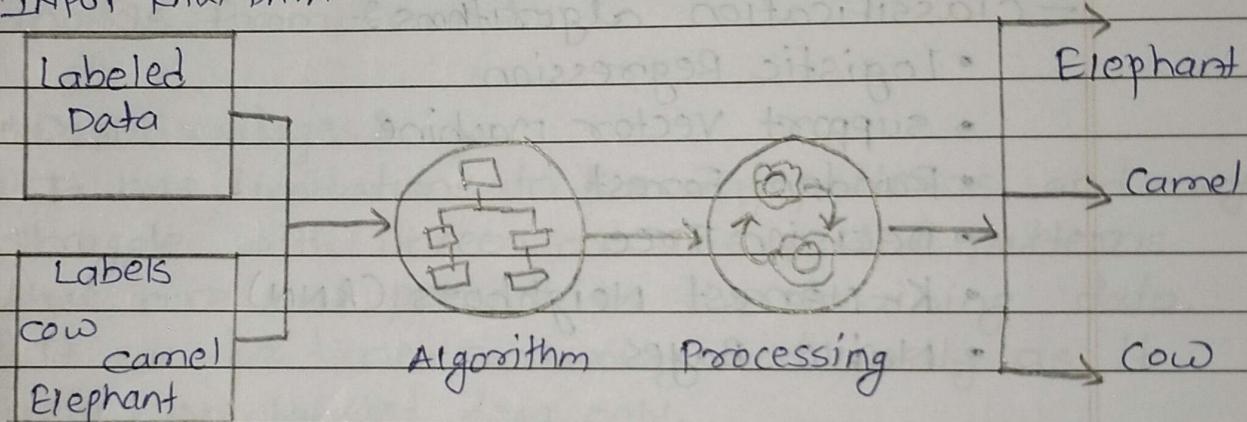
Ques- Explain supervised learning with example.



supervised Machine Learning :-

- supervised learning is defined as when a model gets trained on a "Labelled Dataset".
- Labelled datasets have both input and output parameters.
- In supervised learning algorithms learn to map points between inputs and correct outputs.
- It has both training and validation datasets labelled.

INPUT RAW DATA



Example:-

- consider a scenario where you have to build an image classifier to differentiate between cats and dogs.
- If you feed the datasets of dogs and cats labelled images to the algorithm, the machine will learn to classify between a dog or a cat from these labeled images.
- When we input new dog or cat images that it has never seen before, it will use the learned algorithms and predict whether it is a dog or cat.

DDMMYY

- There are two main categories of supervised Learning :-

1] Classification :-

- classification deals with predicting categorical target variables, which represent discrete classes or labels.
 - For instance, classifying emails as spam or not spam or predicting whether a patient has a high risk of heart disease.
 - classification algorithms learn to map the input features to one of the predefined classes.
- classification algorithms :-
 - logistic Regression
 - support vector machine
 - Random Forest
 - Decision Tree
 - K-Nearest Neighbors (KNN)
 - Naive Bayes

2] Regression :-

- Regression on the other hand, deals with predicting continuous target variables, which represent numerical values.
 - For example, predicting the price of a house based on its size, location and amenities or forecasting the sales of a product.
 - Regression algorithms learn to map the input features to a continuous numerical value.
- regression algorithms :-
 - 1] Linear Regression
 - 2] polynomial Regression

DDMMYY

③ Ridge Regression

④ Lasso Regression

⑤ Decision tree

⑥ Random Forest.

• Advantages :-

- ① supervised learning models can have high accuracy as they are trained on labelled data.
- ② The process of decision-making is supervised learning models is often interpretable.
- ③ It can often be used in pre-trained models which saves time and resources when developing new models from scratch.

• Disadvantages :-

- ① It has limitations in knowing patterns and may struggle with unseen or unexpected patterns that are not present in the training data.
- ② It can be time-consuming and costly as it relies on labeled data only.
- ③ It may lead to poor generalizations based on new data.

• Applications :-

- ① Image classification
- ② speech recognition
- ③ medical diagnosis
- ④ Fraud detection
- ⑤ Email spam detection
- ⑥ Weather Forecasting
- ⑦ Sports analytics
- ⑧ Gaming.

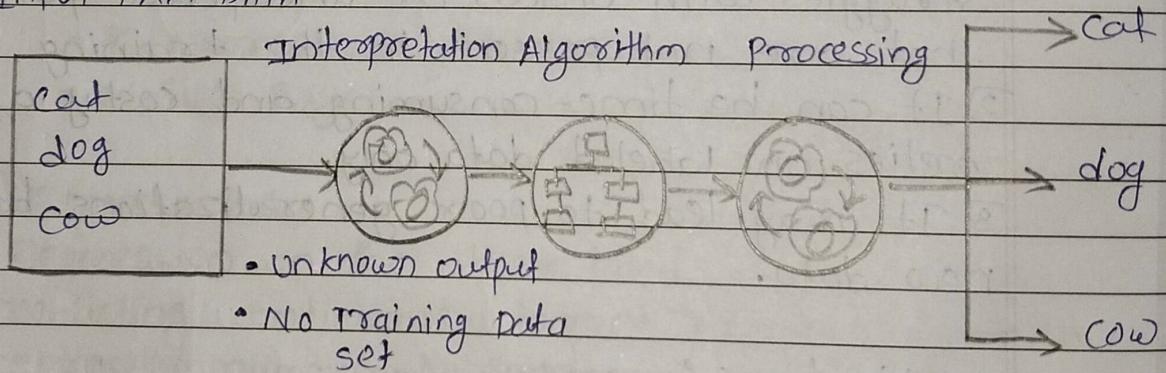
Que- Explain Unsupervised Machine Learning with example.



Unsupervised Machine Learning :-

- unsupervised learning is a type of machine learning technique in which an algorithm discovers patterns and relationships using unlabeled data.
- unsupervised learning doesn't involve providing the algorithm with labeled target outputs.
- The primary goal of unsupervised learning is often to discover hidden patterns, similarities or clusters within the data, which can then be used for various purposes, such as data exploration, visualization and dimensionality reduction.

INPUT RAW DATA



Example :-

- consider that you have a dataset that contains information about the purchases you made from the shop.
- Through clustering, the algorithm can group the same purchasing behavior among you and other customers, which reveals potential customers without predefined labels.

- This type of information can help businesses get target customers as well as identify outliers.

- There are two main categories of unsupervised learning :-

1) Clustering :-

- clustering is the process of grouping data points into clusters based on their similarity.

- The technique is useful for identifying patterns and relationships in data without the need for labeled examples.

- clustering algorithms :-

1) k-means clustering algorithm

2) mean-shift algorithm

3) Principal Component Analysis (PCA)

4) Independent component Analysis (ICA)

2) Association :-

- Association rule learning is a technique for discovering relationships between items in a datasets.

- It identifies rules that indicate the presence of one item implies the presence of another item with a specific probability.

- Association rule learning algorithms :-

1) Apriori Algorithm

2) Eclat

3) FP-growth algorithm.

- Advantages :-

- ① It helps to discover hidden patterns and various relationships between the data.
- ② used for tasks such as customer segmentation, anomaly detection and data exploration.
- ③ It does not require labeled data and reduces the effort of data labeling.

- Disadvantages :-

- ① Without using labels, it may be difficult to predict the quality of the model's output.
- ② Cluster Interpretability may not be clear and may not have meaningful interpretations.
- ③ Complexity and computation
- ④ Uncertainty in outputs.

Application :-

- ① Market basket analysis
- ② Image segmentation
- ③ Genomic data analysis
- ④ Data preprocessing
- ⑤ Customer behavior analysis
- ⑥ Exploratory data analysis (EDA)
- ⑦ Content recommendation
- ⑧ Recommendation systems
- ⑨ Anomaly detection.

Que - Describe Reinforcement Learning. State why it is needed.

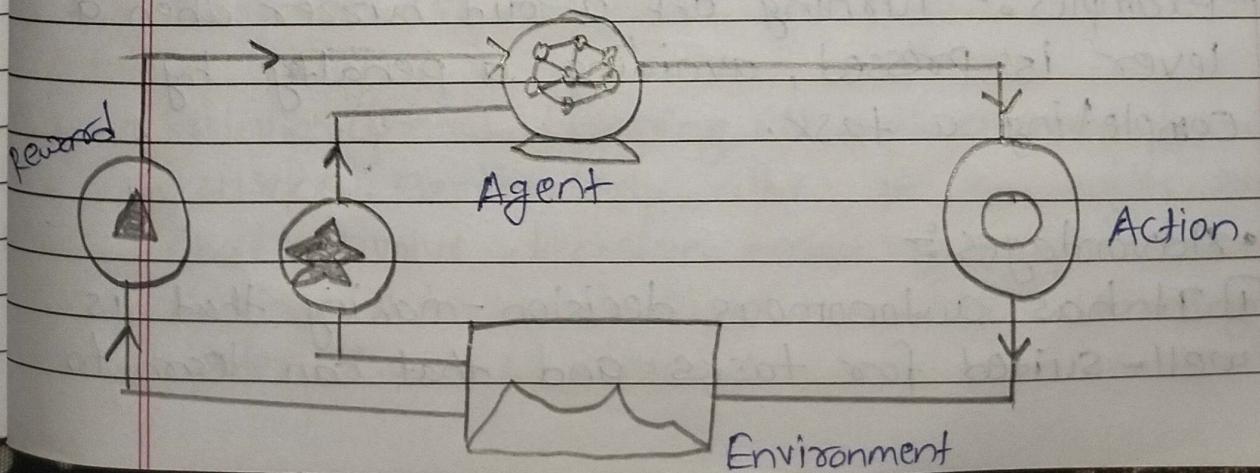
→ Reinforcement Learning :-

- Reinforcement machine learning algorithm is a learning method that interacts with the environment by producing actions and discovering errors.
- Trial, error and delay are the most relevant characteristics of reinforcement learning.
- In this technique, the model keeps on increasing its performance using Reward Feedback to learn the behavior or pattern.
- These algorithms are specific to a particular problem e.g. Google self driving car, AlphaGo where a bot competes with humans and even itself to get better and better performers in Go Game.
- Each time we feed in data, they learn and add the data to their knowledge which is training data. So, the more it learns the better it gets trained and hence experienced.
- Reinforcement Learning algorithm :-

1) Q-Learning

2) SARSA (State - Action - Reward - State - Action)

3) Deep Q-Learning



Example :-

- consider that you are training an AI agent to play a game like chess.
- The agent explores different moves and receives positive or negative feedback based on the outcome.
- Reinforcement learning also finds applications in which they learn to perform tasks by interacting with their surroundings.

Types of Reinforcement Learning :-

1) Positive reinforcement :-

- Rewards the agent for taking a desired action.
- Encourages the agent to repeat the behavior.
- Example - Giving a treat to a dog for sitting, Providing a point in a game for a correct answer.

2) Negative reinforcement :-

- Removes an undesirable stimulus to encourage a desired behavior.
- Discourages the agent from repeating the behavior.
- Examples :- Turning off a loud buzzer when a lever is pressed, avoiding a penalty by completing a task.

* Advantages :-

- ### 1) It has autonomous decision-making that is well-suited for tasks and that can learn to

make a sequence of decisions, like robotics and game-playing.

- ② This technique is preferred to achieve long-term results that are very difficult to achieve.
- ③ It is used to solve complex problems that cannot be solved by conventional techniques.

• Disadvantages:-

- ① Training Reinforcement Learning agents can be computationally expensive and time-consuming.
- ② Reinforcement learning is not preferable to solving simple problems.
- ③ It needs a lot of data and a lot of computation which makes it impractical and costly.

• Applications:-

- ① Game playing
- ② Robotics
- ③ Autonomous Vehicles
- ④ Finance and Trading
- ⑤ Game AI
- ⑥ Education
- ⑦ Industrial control
- ⑧ Recommendation systems.

Why it is needed?



- Reinforcement Learning (RL) is needed for several reasons, particularly when dealing with problems that involve decision-making, dynamic environments, and tasks where learning from interaction is crucial.

- Some reasons why RL is important :-

① dynamic Environments :-

- RL is well-suited for environments that change over time or where the agent's actions can affect the environment.

- It allows the agent to learn optimal behaviors by interacting with the environment and receiving feedback in the form of rewards or penalties.

② sequential Decision Making :-

- many real-world problems require making a series of decisions over time, where each decision can influence future decisions.

- RL provides a framework for learning strategies that optimize long-term rewards rather than just immediate gains.

③ complexity and uncertainty :-

- RL allows agents to learn from experience and adapt to uncertainty by continuously improving their policies based on observed outcomes.

Q. Describe libraries used for machine Learning.

→ Python is one of the most popular programming languages for this task and it has replaced many languages in the industry, one of the reasons is its vast collection of libraries.

- Python libraries that are used in Machine Learning are:-

- 1) Numpy
- 2) Scipy
- 3) Pandas
- 4) Matplotlib
- 5) scikit-learn
- 6) TensorFlow
- 7) Keras
- 8) PyTorch

1) Numpy :-

- Numpy is a very popular python library for large multi-dimensional array and matrix processing, with the help of a large collection of high-level mathematical functions.
- It is very useful for fundamental scientific computations in Machine Learning.
- It is particularly useful for linear algebra, Fourier transform and random number capabilities.
- High-end libraries like TensorFlow uses Numpy internally for manipulation of Tensors.
- Importing numpy :- import numpy as np.



② Scipy :-

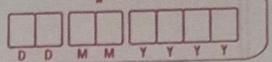
- Scipy is a very popular library among machine learning enthusiasts as it contains different modules for optimization, linear algebra, integration and statistics.
- There is a difference between the scipy library and the scipy stack.
- The Scipy is one of the core packages that make up the scipy stack.
- The Scipy is also very useful for image manipulation.

③ Pandas :-

- Pandas is a popular python library for Data analysis.
- It is not directly related to machine Learning.
- As we know that the dataset must be prepared before training.
- In this case, pandas comes handy as it was developed specifically for data extraction and preparation.
- It provides high-level data structures and wide variety tools for data analysis.
- It provides many inbuilt methods for grouping, combining and filtering data.

④ matplotlib :-

- Matplotlib is a very popular python library for data visualization.
- It is not directly related to machine learning.
- It particularly comes in handy when a



programmer wants to visualize the patterns in the data.

- It is a 2D plotting library used for creating 2D graphs and plots.
- A module named pyplot makes it easy for programmers for plotting as it provides features to control line styles, font properties, formatting axes etc.
- It provides various kinds of graphs and plots for data visualization viz histogram, error charts, bar charts etc.

5) Scikit - Learn :-

- scikit-learn is one of the most popular ML libraries for classical ML algorithms.
- It is built on top of two basic python Libraries viz. numpy and scipy.
- scikit-learn can also be used for data-mining and data analysis which makes it a great tool who is starting out with ML.

6) Tensorflow :-

- Tensorflow is a very popular open-source library for high performance numerical computation developed by the Google Brain team in Google.
- Tensorflow is a framework that involves defining and running computations involving tensors.
- It can train and run deep neural networks that can be used to develop several AI applications.

-Tensorflow is widely used in the field of deep learning research and application.

7) Keras :-

- It provides many inbuilt methods for grouping, combining and filtering data.
- Keras is a very popular machine learning library for python.
- It is a high-level neural networks API capable of running on top of TensorFlow, CNTK or Theano.
- It can run seamlessly on both CPU & GPU.
- Keras makes it really for ML beginners to build and design a Neural Network.
- One of the best things about Keras is that it allows for easy and fast prototyping.

8) py torch :-

- PyTorch is a popular open-source Machine Learning library for Python based on Torch, which is an open-source Machine Learning library that is implemented in C with a wrapper in Lua.
- It has an extensive choice of tools and libraries that support Computer Vision, Natural Language Processing (NLP) and many more ML programs.
- It allows developers to perform computations on Tensors with GPU acceleration and also helps in creating computational graphs.

DD MM YY YY

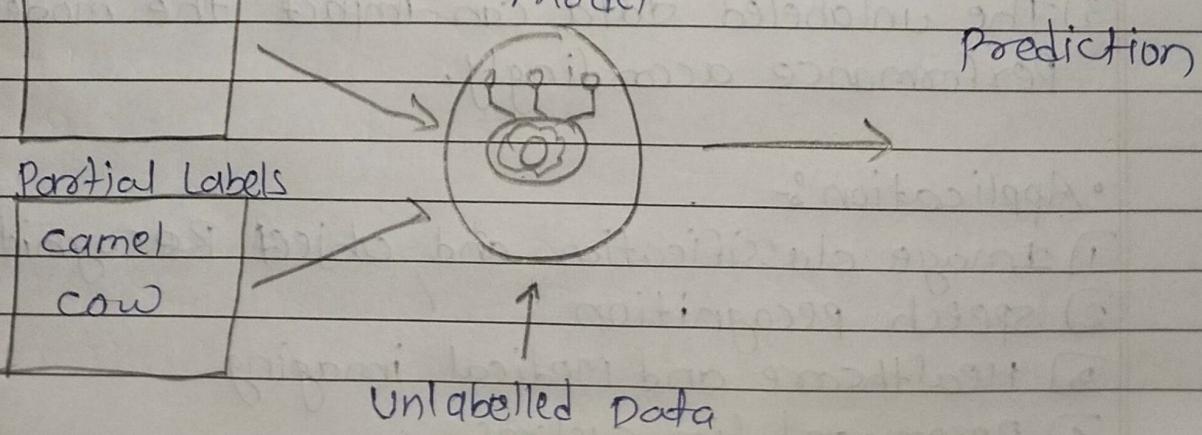
Ques-
→

Explain semi-supervised Learning with example

Semi-supervised Learning :-

- semi-supervised learning is a machine learning algorithm that works between the supervised and unsupervised learning so it uses both labelled and unlabelled data.
- It's particularly useful when obtaining labeled data is costly, time-consuming or resource-intensive.
- This approach is useful when the dataset is expensive and time-consuming.
- semi-supervised learning is chosen when labeled data requires skills and relevant resources in order to train or learn from it.
- We use these techniques when we are dealing with data that is a little bit labeled and the rest large portion of it is unlabeled.
- We can use the unsupervised techniques to predict labels and then feed these labels to supervised techniques.
- This technique is mostly applicable in the case of image datasets where usually all images are not labeled.

INPUT Data Machine Learning model



Example :-

- consider that we are building a language translation model, having labeled translations for every sentence pair can be resources intensive.
- It allows the models to learn from labeled and unlabeled sentence pairs, making them more accurate.
- This technique has led to significant improvements in the quality of machine translation services.

• Advantages :-

- 1) It Leads to better generalization as compared to supervised learning, as it takes both labeled and unlabeled data.
- 2) can be applied to a wide range of data.

• Disadvantages :-

- 1) Semi-supervised methods can be more complex to implement compared to other approaches.
- 2) It still requires some labeled data that might not always be available or easy to obtain.
- 3) The unlabeled data can impact the model performance accordingly.

• Application :-

- 1) Image classification and object Recognition.
- 2) speech recognition
- 3) Healthcare and medical imaging
- 4) Recommendation systems.