1.Explain the term machine learning, and how does it work? Explain two machine learning applications in the business world. What are some of the ethical concerns that machine learning applications could raise?

**Ans:** Machine learning is augmented AI. With machine learning we can develop a kind of system which solves problems with the help of machines as well as human beings.

Machine learning is a subset of artificial intelligence. It teaches machines to think like humans.

It works by exploring data and identifying patterns, and involves minimal human intervention.

Real world examples of machine learning:

1. Online fraud detection - PayPal is using ML tools to distinguish between legitimate or illegitimate transactions taking place between the buyers and sellers.

2.Self-Driving Cars and Automated Transportation - Boeing 777 pilot uses ML tools for automation. The navigation issues are solved by use of google maps.

3. Home Security and Smart Homes - AI-powered alarms and cameras are used for home security.

4. Product Recommendations – Users can get recommendations of movies, books and other products with the help of ML tools.

Ethical concerns that machine learning applications could raise:

1.Joblessness

2.Humanity

3.Bias

4.Trust

2. Describe the process of human learning:

i. Under the supervision of experts

**Ans:** Teachers teach concepts to students. Students perform experiments in laboratories.

ii. With the assistance of experts in an indirect manner

**Ans:** Watching interactive videos to solve problems.

iii. Self-education

Ans: Reading books or articles on google and performing experiments.

Getting the results validated by someone.

3. Provide a few examples of various types of machine learning.

**Ans: 1. Supervised learning**

Predicting the price of houses, classifying mails into spam and not spam

**2. Unsupervised learning**

Customer segmentation, grouping user logs

4. Examine the various forms of machine learning.

**Ans:**Following are the forms of machine learning:

1]Supervised learning-Training is done in the presence of a teacher/guide

The dataset used is a labelled dataset

Mainly the problems of classification and regression fall under

Supervised learning.

Ex:linear regression,SVN,KNN,Decision trees

2] Unsupervised learning-Training is under by self learning

The dataset is unlabelled

It is data driven

Rerecommended system, grouping user logs

3] Reinforcement learning-We need an agent and an environment

Agent interacts with the environment and gets penalty or

reward for his action

Ex:video games,industrial simulations

5. Can you explain what a well-posed learning problem is? Explain the main characteristics that must be present to identify a learning problem properly.

**Ans:** A problem is said to be well posed id there exists a solution for that problem and it is a unique solution. The solution must be dependent on data but it should not be sensitive to changes in the data.

The main characteristics of a well posed learning problem are:

1.Task

2.Performane measure

3.Experience

6. Is machine learning capable of solving all problems? Give a detailed explanation of your answer.

**Ans:** Machine learning is not always able to solve our problems. Sometimes our senses can give us solution. For example, we follow GPS instructions to reach a path. But it can navigate us to wrong path, but our senses and knowledge can lead us our destination.

The machine learning does not understand the physical constraints like density cannot be negative.

7. What are the various methods and technologies for solving machine learning problems? Any two of them should be defined in detail.

**Ans**: Following are the methods of machine learning:

1.Classification

2.Regression

3.Random Forest

4.Decision Trees

5.Principal Component Analysis

6. Support Vector Machines

7. Recommender System

8.Reinforcement learning

**1.Classification:** Classification is the supervised learning technique of machine learning.

It uses the labelled dataset for training. It classifies or applies labels to data. The classification may be binary or multiclass. What are the various methods and technologies for solving machine learning problems.

The output variable of Classification is a category, not a value.

* Linear Models
  + Logistic Regression
  + Support Vector Machines
* Non-linear Models
  + K-Nearest Neighbours
  + Kernel SVM
  + Naïve Bayes
  + Decision Tree Classification
  + Random Forest Classification

**2. Clustering:** Clustering is a method of grouping the objects into clusters such that objects with most similarities remains into a group and has less or no similarities with the objects of another group. Cluster analysis finds the commonalities between the data objects and categorizes them as per the presence and absence of those commonalities.

Following are methods of clustering:

1.Principal Component Analysis

2.K-Means clustering

3.Apriori algorithm

8. Can you explain the various forms of supervised learning? Explain each one with an example application.

**Ans:** Following are the various forms of supervised machine learning algorithm:

1. **Classification:** The emails are classified as spam or no spam.
2. **Regression:** The weight of a person can be predicted by using his height and weight.

9. What is the difference between supervised and unsupervised learning? With a sample application in each region, explain the differences.

**Ans:** Supervised machine learning uses labelled data and unsupervised learning uses non labelled data.

Supervised learning partitions the data according the labels of the data, unsupervised data gain insights from the data and groups them according to their similarities.

Supervised learning: The weight of a person can be predicted by using his height and weight.

Unsupervised learning: Classifying the fruit using the attributes shape, taste, size, smell

Reinforcement learning: Video games

10. Describe the machine learning process in depth.

**Ans:** Following are steps in the machine learning process:

1.Data Gathering: Data comes from various sources and it could be in any format. CSV, XML.JSON

2. Data Pre-processing: The data may be structured or unstructured. And in any format.It may contain duplicates, null values. So, by pre-processing the data it is made suitable for feeding into the machine learning algorithm.

3.EDA:To understand the relationship between the various attributes in the dataset,to find important features EDA is performed.

4. Model Building: According the problem statement ,the machine learning algorithm is selected.The data is divided into training and testing dataset.The model is trained using training data set.On testing data set the results are tested.

5. Model Evaluation:

**Evaluating** the Regression **Model**.

* 1. Sum of Squared Error (SSE)
  2. Mean Squared Error (MSE)
  3. Root Mean Squared Error (RMSE)
  4. Mean Absolute Error (MAE)
  5. Coefficient of Determination (R2)
  6. Adjusted R2

**Evaluating** Classification **Model**.

* 1. Confusion Matrix.
  2. Accuracy Score.
  3. AUC and ROC.|

6.Deployment:  Integration of the finalized model into a production environment and getting results to make business decisions.

11.Make brief notes on any two of the following:

I. MATLAB is one of the most widely used programming languages.

**Ans:** MATLAB language is dedicated to mathematical and technical computing.Function names are memorable and easy is understand. It is matrix-based language.

ii. Deep learning applications in healthcare

**Ans:** Drug discovery**,** medical imaging,genome,Insurance fraud

iii. Study of the market basket

iv. Linear regression (simple)

12. Make a comparison between: -

1. Generalization and abstraction

**Ans:** Abstraction is proving only necessary details to user and hiding other details. Generalization is writing a code that can be reused in other applications.

2. Learning that is guided and unsupervised

**Ans:**Guided learning is done with labelled data.We know what is going to be the output.Unsupervised learning uses unlabelled data.It finds properties of data and accordingly performs grouping of data based on similarities.

3.Regression and classification

**Ans:** The target variable in classification is categorical. In regression it is continuous.