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?

Machine Learning is a part or sub-domain of AI which emphasis on studying different statistical methodologies by which a chunk of data can be aggregated, trained and evaluated using automated libraries and tools.

Two business applications are: 1. Data Scrapping from an e-commerce website. 2. Movie Recommendation.

Some ethical concerns that machine learning applications could raise are:

i) Data Security and privacy

ii) Accuracy, Bias and fairness

iii) Decision Making

2. Describe the process of human learning:

i. Under the supervision of experts: Supervised Learning

ii. With the assistance of experts in an indirect manner: Reinforced Learning

iii. Self-education: Unsupervised Learning

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

Random Forest, Linear Regression, Logistic Regression, Decision tree, Naive Bias, Support Vector Mechanism, Gradient Boosting algorithms etc.

4. Examine the various forms of machine learning.

Supervised Machine Learning, Unsupervised Machine Learning, Reinforced machine Learning

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.

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.

There could be number of characteristics or sub-category to identify a machine learning problem, but the main characteristics could mainly focus on:

- What is the problem?

- Why do we need to solve the problem?

- How to solve the problem?

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

No, machine learning is not capable of solving all problems. One of the most common limitations is ethics. Machine learning cannot tell us anything about what normative values we should accept, i.e. how we should act in the world in a given situation. Although a ML algorithm can be used to train previous situations to generate the reaction of current situation using deterministic factors. But it doesn’t explain the exception. If one person/human is sad but previously on the same occasion the human was happy. The automated algorithm on training the past features will provide the output as happy. So it cannot tell us the values what we want to accept.

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

Various methods and technology for solving ML problems are:

i) Supervised Learning: In supervised learning, the model learns from labeled data, where the input features are mapped to known output labels. Some popular algorithms for supervised learning include linear regression, logistic regression, decision trees, random forests, support vector machines (SVM), and neural networks.

ii) Un-supervised Learning: Unsupervised learning involves learning from unlabeled data, where the model aims to discover patterns, structures, or relationships within the data. Clustering and dimensionality reduction are common techniques used in unsupervised learning. K-means clustering, hierarchical clustering, and principal component analysis (PCA) are popular algorithms in this category.

iii) Reinforcement Learning: Reinforcement learning involves an agent interacting with an environment and learning to make decisions through trial and error. The agent receives feedback in the form of rewards or penalties, guiding it towards learning optimal behavior. Q-learning and deep Q-networks (DQN) are common algorithms used in reinforcement learning.

iv) Deep Learning: Deep learning is a subset of machine learning that focuses on artificial neural networks, particularly deep neural networks with multiple hidden layers. Deep learning has achieved significant breakthroughs in areas such as image and speech recognition, natural language processing, and generative models. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are commonly used in deep learning.

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

i) Regression: Regression is used when the output variable is continuous or numerical. The goal is to predict a value within a range. For instance, predicting house prices based on features like location, size, and number of rooms is a regression problem. Other examples include predicting stock prices, estimating sales revenue, or forecasting weather conditions.

ii) Classification: Classification is used when the output variable is categorical or discrete. The goal is to assign input data points to predefined classes or categories. For example, email spam detection is a classification problem where the model learns to classify emails as either spam or non-spam based on various features. Other applications include image recognition (e.g., classifying images as cats or dogs), sentiment analysis (classifying text as positive or negative), and fraud detection (classifying transactions as fraudulent or legitimate).

iii) Multiclass Classification: Multiclass classification is an extension of classification where there are more than two classes. The model learns to assign input data points to multiple possible classes. For example, a model might classify handwritten digits into numbers 0-9. Other applications include speech recognition (classifying spoken words or phrases), document categorization (assigning documents to multiple topics), and disease diagnosis (classifying medical conditions into multiple categories).

iv) Sequence Labeling: Sequence labeling is used when the input is a sequence of data points, and the goal is to assign a label to each element in the sequence. Named Entity Recognition (NER) is a popular application of sequence labeling, where the task is to identify and classify named entities (such as person names, organizations, or locations) in text. Part-of-Speech (POS) tagging is another example, where each word in a sentence is labeled with its corresponding grammatical category (e.g., noun, verb, adjective).

v) Time Series Forecasting: Time series forecasting deals with predicting future values based on historical data that is ordered in time. It is used when the input data is sequential and has a temporal dimension. Examples include predicting stock prices, forecasting electricity demand, predicting sales for a retail store, or predicting the number of website visitors over time.

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

Supervised Learning: In supervised learning, the model learns from labeled data, where the input features are mapped to known output labels. Some popular algorithms for supervised learning include linear regression, logistic regression, decision trees, random forests, support vector machines (SVM), and neural networks.

Examples: Identify the class of iris dataset. It is a labelled dataset having multivariate classes as target column, which will determine the species. The Iris Dataset contains four features (length and width of sepals and petals) of 50 samples of three species of Iris (Iris setosa, Iris virginica and Iris versicolor). These measures were used to create a linear discriminant model to classify the species.

Unsupervised Learning: Unsupervised learning involves learning from unlabeled data, where the model aims to discover patterns, structures, or relationships within the data. Clustering and dimensionality reduction are common techniques used in unsupervised learning. K-means clustering, hierarchical clustering, and principal component analysis (PCA) are popular algorithms in this category.

Examples: Identifying the customer for potential increase in sale or providing additional discounts of an e-commerce platform. The e-commerce data can provide a number of independent features which are unlabelled. Form a clustering point of view one can identify a cluster which are eligible for discounts on next shopping or identify those who are highly susceptible to buy more products in the upcoming season.

10. Describe the machine learning process in depth.

Machine Learning is a part or sub-domain of AI which emphasis on studying different statistical methodologies by which a chunk of data can be aggregated, trained and evaluated using automated libraries and tools. Now in a ML process we will consider the following processes accordingly:

i) Data collection/ingestion: Data is collected from database where data is being recorded. Using proper logging we will retrieve the data from database.

ii) Data transformation: Once the data is collected it is then transformed as per model. The process of model transformation may include tasks like check for null values, check for cat and num columns, encoding the categorical values etc.

iii) Feature engineering: This is also a process of data transformation where features get split into train and test data. And these are passed through encoding techniques, scaling etc.

iv) Model Training: The data set is now ready for model training. Since it is a classification problem we will consider different classifier like logistic regression, random forest classifier etc. Model is trained on number of classifier to get best accuracy.

v) Evaluation: Once the model training is done, the same model is passed through prediction and test data is brought into to find the accuracy of the model. The best accuracy score model will be used for future prediction.

vi) Deployment: Now this model is available for deployment to any cloud server like AWS,GCP or Azure

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

i) MATLAB is one of the most widely used programming languages.

ii. Deep learning applications in healthcare: Deep learning applications had been evolved in the healthcare industry rapidly. Let’s take an example of diabetic or non diabetic prediction of a patient. The independent feature has a number of parameters like BMI, insulin, pregnancy, etc and the output column is a binomial class target output, classifying diabetic and non-diabetic.

The proposed work on data: A multilayer feed-forward perceptron based model was used which also facilitates the properties of ANN and trained with stochastic gradient descent using back-propagation. The network is a collection of four layers emulating nodes and neurons, directed in uni-direction (one-way connection). Each node is connected to the next node in a single way connection and contains two hidden layers where each node trains a copy of global model parameters by applying its local data. Further, it uses multiple threads to process the model and apply the averaging for contributing to the model access across the whole network. The learning model uses stochastic gradient descent training using backpropagation and hidden layer’s neurons which enable more advance features like tanh, rectifier and maxout activation, learning rate, rate annealing.

iii. Study of the market basket

iv. Linear regression (simple): Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

This form of analysis estimates the coefficients of the linear equation, involving one or more independent variables that best predict the value of the dependent variable. Linear regression fits a straight line or surface that minimizes the discrepancies between predicted and actual output values. There are simple linear regression calculators that use a “least squares” method to discover the best-fit line for a set of paired data. We then estimate the value of X (dependent variable) from Y (independent variable).

11. Make a comparison between:-

1. Generalization and abstraction: Generalization focuses on the performance of a model on unseen data, while abstraction deals with simplifying complex information to capture essential features or concepts. Generalization ensures that the model can make accurate predictions on new data, while abstraction facilitates efficient processing and understanding of the underlying domain or problem. Both concepts are essential for developing robust and effective machine learning models.

2. Learning that is guided and unsupervised: it is the simple case of supervised learning and unsupervised learning. Supervised learning need guided approach where a part of labelled data is trained and then predicted on the basis of the previous training. First a data was split into train and test split, the independent and dependent features are trained first. Next the independent test data features will be used to predict the output. Once the prediction is done, accuracy score is measured by comparing the original or true test output data with that of predicted one. Whereas unsupervised consists of unlabelled data, the data set will not contain any target column here. Unsupervised learning involves learning from unlabeled data, where the model aims to discover patterns, structures, or relationships within the data. Clustering and dimensionality reduction are common techniques used in unsupervised learning. K-means clustering, hierarchical clustering, and principal component analysis (PCA) are popular algorithms in this category.

3. Regression and classification: Regression is used when the output variable is continuous or numerical. The goal is to predict a value within a range. For instance, predicting house prices based on features like location, size, and number of rooms is a regression problem. Other examples include predicting stock prices, estimating sales revenue, or forecasting weather conditions. Classification is used when the output variable is categorical or discrete. The goal is to assign input data points to predefined classes or categories. For example, email spam detection is a classification problem where the model learns to classify emails as either spam or non-spam based on various features. Other applications include image recognition (e.g., classifying images as cats or dogs), sentiment analysis (classifying text as positive or negative), and fraud detection (classifying transactions as fraudulent or legitimate).