**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 a subset of AI that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values.

Ml applications:

1. Image recognition: 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 with a feature of auto friends tagging suggestions. It is based on the project ”Deep Face” which is responsible for tagging.
2. Speech recognition: search by voice,converting voice instruction into text speech to text. Ex. Google Assistance, Siri, Cortana and Alexa.

**2. Describe the process of human learning:**

i. Under the supervision of experts: Under supervision whereby subject matter experts accelerate the learning process by teaching the technology in real-time

Experts act as an authoritative fig. and transmit knowledge.

Acquisition, retention and modification of experience is described as learning.

ii. With the assistance of experts in an indirect manner: The expert acts as a facilitator, assisting the human in drawing out what he already knows.

iii. Self-education: learning through self experiences. Because of their cognitive potential, newborns can adapt to the world they were born into. human learning (general perspective), while the variation among individuals(differential perspective), occurs as an interaction with the environment and is initialized to adapt personal needs to the external worlds.

For learning to occur, two things are important.

1. The presence of a stimulus in the environment.
2. The innate dispositions like emotional and instinctual disposition.

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

Ans: Types of Machine Learning:

1. Supervised
   1. Housing price prediction
   2. Medical imaging
2. Unsupervised
   1. Customer Segmentation
   2. Market Basket Analysis
3. Semi-Supervised
   1. Text classification
   2. Lane-finding on GPS data
4. Reinforcement
   1. Optimized marketing
   2. Driverless Cars.

**4. Examine the various forms of machine learning.**

Ans: Supervised learning: Common algorithms used during machine learning include neural networks, decision trees, linear regression, svm. This type got its name because the machine is ‘supervised’ while it’s learning, which means that you’re feeding the algorithm information to help it learn. The outcome you provide the machine is labeled data, and the rest of the information you give is used as input features.

Unsupervised learning: doesn’t use the same labeled training sets and data. instead, the machine looks for less obvious patterns in the data. common algorithms include hidden Markov models, k-means, hierarchical clustering, and Gaussian mixture models. Used widely to create predictive models.

Reinforcement Learning: Closest machine learning type to how humans learn. The algorithm or agent used learns by interacting with its environment and getting a positive or negative reward. Common algorithms include temporal difference, deep adversarial networks, and Q-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.**

Ans: A ML problem is well-posed if a solution to its 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.

Any problem can be segregated as well-posed learning problem if it has three traits-

1. Task
2. Performance Measure
3. Experience

A computer program is said to learn from experience E in context to some task T and some measure P, If its performance on T, as was measured by P, upgrades with experience E.

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

Ans: ML has limitations, and is not the answer to all problems.

1. Trusting the data and algorithms more than our judgement has its pros and cons.
2. Machine learning is stochastic, not deterministic
   1. ex. Utilizing a neural network misses the entire physics of the weather system. unless adding physical constraints to nn and other algorithms so that they can be used for such purposes.
3. Lack of data and lack of good data.
4. Misapplication
   1. For deterministic systems, algorithms will succeed, but cannot learn the relationship between the two variables, and will not know when it is violating the physical laws as in the case of a weather system.
   2. For stochastic (random systems), things are little less obvious.The crisis of ML for random systems manifests itself in two ways.
      1. P-hacking: Looking through mountains of data until a correlation showing statistically significant results is found. These are not true correlations and are just responding to the noise in the measurement. This results in individuals ‘fishing’ for statistically significant correlations through large data sets, and masquerading these as true correlations.
      2. Scope of the analysis: Machine learning is inherently exploratory. ML algorithms approaches are best suited for exploratory predictive modeling and classification with massive amounts of data and computationally complex features.As for small data a classic, multivariate statistical methods are much more informative(which are confirmatory in nature).
5. Interpretability: A business manager is more likely to accept the [machine learning method] recommendations if the results are explained in business terms. Interpretability is one of the primary problems with machine learning.

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

Ans: Methods : Regression, Classification, Clustering, Dimensionality reduction, Ensemble methods, Neural nets and Deep learning, Transfer learning, Reinforcement learning. NLP, word embeddings.

1. Regression: They help to predict or explain a particular numerical value based on a set of prior data, for example predicting the price of a property based on previous pricing data for similar properties. where we use the mathematical equation of line (y=m\*x+b) to model a dataset. We train a linear regression model with many data point pairs(x,y) by calculating the position and slope of the line that minimizes the total distance between all the data points and line. In other words we calculate slope(m) and y-intercept (b) for a line that approximates the observations in the data. Regression techniques run the gamut from simple to complex (like regularized linear regression, polynomial regression, decision trees and random forest regressions, neural nets, among others)
2. Classification: predicts tor explains the class value. The simplest is logistic regression.Logistic regression method estimates the probability of an occurrence of an event based on one or more inputs. Ex. input of two exam scores for a student in order to estimate the probability that the student will get admitted to a particular college. Because the estimate is a probability, the output is a number between 1 and 0, where 1 represents complete certainty. For a student , if the estimated probability is greater than 0.5, then we predict that he or she will be admitted. if the estimated probability is less than 0.5 we predict he or she will be refused.

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

Ans: forms of supervised meaning:

1. Regression:
   1. Linear regression:A linear regression model is used to depict a relationship between variables that are proportional to each other. Meaning, that the dependent variable increases/decreases with the independent variable. For example, as the age of a person increases, the level of glucose in their body increases as well.
   2. Non-linear regression:In the non-linear regression model, the graph doesn’t show a linear progression. Depending on how the response variable reacts to the input variable, the line will rise or fall showing the height or depth of the effect of the response variable. For example, a patient’s response to treatment can be good or bad depending on their body’s tendency and willpower.
   3. Multiple regression model: A multiple regression model is used when there is more than one independent variable affecting a dependent variable. While predicting the outcome variable, it is important to measure how each of the independent variables moves in their environment and how their changes will affect the output or target variable. For example, the chances of a student failing their test can be dependent on various input variables like hard work, family issues, health issues, etc.
2. Classification:
   1. Logistic regression:It is a kind of linear regression. But is used when the dependent variable is not a number but something else (e.g. a “yes/no” response). It’s called a regression but performs classification based on the regression and it classifies the dependent variable into either of the classes.used for binary classification. for example, if a credit card company builds a model to decide whether or not to issue a credit card to a customer, it will model for whether the customer is going to “default” or “not default” on their card.Firstly, linear regression is performed on the relationship between variables to get the model. The threshold for the classification line is assumed to be at 0.5.

supervised machine logistic sigmoid function. Logistic function is applied to the regression to get the probabilities of it belonging in either class.

It gives the log of the probability of the event occurring to the log of the probability of it not occurring. In the end, it classifies the variable based on the higher probability of either class. i.e. log of odds.

* 1. K-nearest neighbors(K-NN):K-NN algorithm is one of the simplest classification algorithms and it is used to identify the data points that are separated into several classes to predict the classification of a new sample point. K-NN is a non-parametric, lazy learning algorithm. It classifies new cases based on a similarity measure (i.e., distance functions).K-NN works well with a small number of input variables (*p*), but struggles when the number of inputs is very large.

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

Ans: Difference

| Supervised ML | Unsupervised ML |
| --- | --- |
| uses labelled Input as well as labelled output training data. so we can have an exact idea about the classes of objects | trained on unlabelled dataset, the algorithms predicts the output without any supervision |
| aim is to find mapping function to map the input variable(x) with the output variable(y) | The aim is to group or categories the unsorted dataset according to the similarities, patterns and differences. |
| used for Risk assessment, Image classification, Fraud detection, Spam filtering etc | Network analysis, recommendations systems, Anomaly detection, singular value decomposition |
| These algorithms are helpful in predicting the output on the basis of prior experience | are preferable for various tasks as getting unlabelled dataset is easier |
| Not able to solve complex tasks | can be used for complicated tasks |
| it may predict wrong output if the test data is different from training data | output can be less accurate as the dataset is not labeled |
| requires lots of computational time to train the algorithm | working with, is more difficult because does not map with the output |

**There are some very practical applications of supervised learning algorithms in real life, including:**

Text categorization

Face Detection

Signature recognition

Customer discovery

Spam detection

Weather forecasting

Predicting housing prices based on the prevailing market price

Stock price predictions, among others

**Some practical applications of unsupervised learning algorithms include**:

Fraud detection

Malware detection

Identification of human errors during data entry

Conducting accurate basket analysis, etc.

**10. Describe the machine learning process in depth.**

Ans: MAjor steps in machine learning are:

1. Collecting data: Machines initially learn from the data that you give them. It is of the utmost importance to collect reliable data so that your machine learning model can find the correct patterns. The quality of the data that you feed to the machine will determine how accurate your model is. If you have incorrect or outdated data, you will have wrong outcomes or predictions which are not relevant. Make sure you use data from a reliable source, as it will directly affect the outcome of your model. Good data is relevant, contains very few missing and repeated values, and has a good representation of the various subcategories/classes present.
2. Preparing the Data: after you have your data, you have to prepare it. you can do this by;
   1. Putting together all the data you have and randomizing it. This helps make sure that data is evenly distributed, and the ordering does not affect the learning process.
   2. Cleaning the data to remove unwanted data, missing values, rows, and columns, duplicate values, data type conversion, etc. You might even have to restructure the dataset and change the rows and columns or index of rows and columns.
   3. Visualize the data to understand how it is structured and understand the relationship between various variables and classes present.
   4. Splitting the cleaned data into two sets - a training set and a testing set. The training set is the set your model learns from. A testing set is used to check the accuracy of your model after training.
3. Choosing a Model: A machine learning model determines the output you get after running a machine learning algorithm on the collected data. It is important to choose a model which is relevant to the task at hand. Over the years, scientists and engineers developed various models suited for different tasks like speech recognition, image recognition, prediction, etc. Apart from this, you also have to see if your model is suited for numerical or categorical data and choose accordingly.
4. Training the Model: Training is the most important step in machine learning. In training, you pass the prepared data to your machine learning model to find patterns and make predictions. It results in the model learning from the data so that it can accomplish the task set. Over time, with training, the model gets better at predicting.
5. Evaluating the Model:After training your model, you have to check to see how it’s performing. This is done by testing the performance of the model on previously unseen data. The unseen data used is the testing set that you split our data into earlier. If testing was done on the same data which is used for training, you will not get an accurate measure, as the model is already used to the data, and finds the same patterns in it, as it previously did. This will give you disproportionately high accuracy. When used on testing data, you get an accurate measure of how your model will perform and its speed.
6. Parameter Tuning: Once you have created and evaluated your model, see if its accuracy can be improved in any way. This is done by tuning the parameters present in your model. Parameters are the variables in the model that the programmer generally decides. At a particular value of your parameter, the accuracy will be the maximum. Parameter tuning refers to finding these values.
7. Making Predictions: In the end, you can use your model on unseen data to make predictions accurately.

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**

AI and machine learning have gained a lot of popularity and acceptance in recent years. With the onset of the Covid-19 pandemic, the situation changed even more. During the crisis, we witnessed a rapid digital transformation and the adoption of disruptive technology across different industries. Healthcare was one of the potential sectors that gained many benefits from deploying disruptive technologies. AI, machine learning, and deep learning have become an imperative part of the sector. Deep learning in healthcare has a huge impact and it has enabled the sector to improve patient monitoring and diagnostics. Here are the top pathbreaking applications of deep learning in healthcare.

Drug Discovery

The role of deep learning in identifying drug combinations is significant. During the pandemic, vaccine and drug development were funded by disruptive technologies like AI, machine learning, and deep learning. Since drug discovery is a complex task, deep learning can make it faster, cost-effective, and easier. Deep learning algorithms can predict the drug properties, drug-target interaction prediction, and in generating a compound with desired properties. Deep learning algorithms can easily process genomic, clinical, and population data and various toolkits can be used to detect patterns between the data. By leveraging machine learning and deep learning, researchers are now able to perform faster molecular modeling and predictive analytics in defining protein structures.

Medical Imaging and Diagnostics

Deep learning models can interpret medical images like X-ray, MRI scan, CT scan, etc., to perform diagnosis. The algorithms can detect any risk and flag anomalies in the medical images. Deep learning is extensively used in detecting cancer. The recent innovation of computer vision was enabled by machine learning and deep learning. With a faster diagnosis through medical imaging, it becomes easier to treat diseases.

Simplifying Clinical Trials

Clinical trials are complicated and expensive. Machine learning and deep learning can be leveraged to perform predictive analytics to identify potential candidates for clinical trials and enable scientists to pool in people from different data points and sources. Deep learning will also enable continuous monitoring of these trials with minimum errors and human intervention.

Personalized Treatment

With deep learning models, it becomes easier to analyze patient’s health data, medical history, vital symptoms, medical test results, and others. Hence, this enables healthcare providers to understand each patient and provide personalized treatment for them. These disruptive technologies enable the detection of suitable and multiple treatment options for different patients. With real-time data collection through connected devices, machine learning models can use deep neural networks to predict upcoming health conditions or risks and provide specific medicines or treatments.

Improved Health Records and Patient Monitoring

Deep learning and machine learning models can process and analyze various medical and healthcare data, both structured and unstructured. Document classification and maintaining up-to-date health records might become manually difficult. Thus, machine learning and its subset deep learning can be used to maintain smart health records. With the advent of telemedicine, wearables, and remote patient monitoring, there is now abundant real-time data on health and deep learning can help in intelligently monitoring the patients and predict risks.

Health Insurance and Fraud Detection

Deep learning can efficiently identify insurance frauds and predict future risks. Health insurance providers are also an advantage if they use deep learning because the models can predict the future trends and behavior to suggest smart insurance policies to their clients.

Deep Learning and NLP

Natural language processing (NLP) leverages deep learning algorithms for classification and identification. These two technologies can be used in identifying and classifying health data and can also be leveraged to develop chatbots and voice bots. In the current scenario of telehealth, chatbots play a pivotal role. It makes the interaction with patients easier and faster. These chatbots were also used to spread the word about Covid-19 and answer primary queries.

**iii. Study of the market basket:**

*In retail sector:*

Market basket analysis is a data mining technique used by retailers to increase sales by better understanding customers. Identifying product groups and items that are most likely to be bought together, includes evaluating big data sets, such as purchase history.

1. Purpose of Market Basket Analysis

Finding items that buyers desire to buy is the major goal of market basket analysis. Market basket analysis may help sales and marketing teams develop more effective product placement, pricing, cross-sell, and up-sell tactics.

1. Types Of Market Basket Analysis

● Predictive Market Basket Analysis

This kind employs supervised learning methods like regression and classification. In essence, it seeks to imitate the market to examine what factors influence events. In essence, it determines cross-selling by taking into account things bought in a particular order.

● Differential Market Basket Analysis

For competition analysis, this kind of analysis is useful. To identify intriguing patterns in consumer behavior, it compares purchase histories across brands, periods, seasons, days of the week, etc.

1. Algorithms Associated With Market Basket Analysis

The market study definition is based on Association Mining rules, as was already explained. Association mining is a technique used by the AIS, SETM, and Apriori algorithms. The Apriori Algorithm is the MBA algorithm that is used the most frequently.

1. How Does Market Basket Analysis Work?

The IF, THEN construct is used in association rule mining to replicate market basket analysis. When a customer buys bread, he is likely to also buy butter. Examples of association rules include the following: "Bread" -> "Butter"

1. Definitions to better understand market basket analysis:

● Antecedent

The entities or "itemsets" produced from the data are called antecedents. To put it another way, it's the IF element on the left. In the situation before, bread serves as the antecedent.

● Consequent

The term "consequent" refers to an item or group of items that are encountered along with the antecedent. The THEN part of the sentence is displayed on the right-hand side. The result in the aforementioned case is butter.

1. Metrics For Market Basket Analysis In Data Mining

You can put a lot of interesting controls on your association rules. These consist of

* Support
* Confidence
* Lift

Consider the following scenario: A well-known e-commerce site handled 4000 transactions. They are trying to determine how many transactions, how much lift, trust, and support there is for the two things, a phone, and a phone cover, out of 5000. The phone has 500 transactions, the phone case has 800 transactions, and the two together have 1000 transactions.

1. Benefits Of Market Basket Analysis

Gaining market share: Once a business reaches its peak growth, finding new ways to do so might be difficult. Market basket analysis may be used to integrate gentrification and demographic data to locate the sites of new businesses or geo-targeted marketing.

Campaigns and promotions: MBA is used to identify the goods that work well together as well as the products that serve as the cornerstones of their product range.

Behavior analysis: A fundamental tenet of marketing is comprehending consumer behavior patterns. MBA may be used for anything, including UI/UX and basic catalog designs.

Optimization of in-store activities: MBA is useful in deciding what goes on the shelves as well as at the back of the shop. Because geographic patterns are a major factor in determining the strength or popularity of particular products, MBA is increasingly used to manage inventory for each store or warehouse.

1. Examples Of Market Basket Analysis

Retail

The most well-known case study using market basket analysis is probably Amazon.com. As soon as you visit Amazon to look at a product, the product description will suggest "Items purchased together frequently." It is the clearest and most straightforward example of Market Basket Analysis cross-selling tactics.

Along with e-commerce methods, consumer in-store retailers also greatly benefit from BA. For grocery stores, visual merchandising and shelf optimization is crucial. For instance, shower gel is almost usually kept close to one another at the grocery store.

IBFS

Examining credit or debit card history is a highly advantageous MBA opportunity for IBFS companies. For instance, Citibank frequently sends sales representatives to large malls to tempt potential customers with enticing on-the-go discounts.

Additionally, they collaborate with services like Swiggy and Zomato to provide customers with a selection of offers that they may use their credit cards to redeem.

Telecom

Due to the intense competition in the telecom sector, businesses are paying close attention to the advantages that customers frequently utilize. For instance, telecom has started to combine TV and Internet bundles with other affordable internet platforms to reduce migration.

iv. Linear regression (simple)

**11. Make a comparison between:-**

1. Generalization and abstraction:

Abstraction aims at simplifying the description of an entity while generalization looks for common properties among these abstractions.

**Generalization**

"The concept of generalization in OOP means that an object encapsulates common state and behavior for a category of objects."

So for example, if you apply generalisation to shapes, then the common properties for all types of shape are area and perimeter.

, if you were working in the domain of jet aircraft, you could have a Jet as a generalisation, which would have a wingspan property. A specialisation of a Jet could be a FighterJet, which would inherit the wingspan property and would have its own property unique to fighter jets e.g. NumberOfMissiles.

**Abstraction**

"the process of identifying common patterns that have systematic variations; an abstraction represents the common pattern and provides a means for specifying which variation to use" (Richard Gabriel)"

In the domain of programming:

An abstract class is a parent class that allows inheritance but can never be instantiated.

Hence in the example given in the Generalization section above, a Shape is abstract as:

In the real world, you never calculate the area or perimeter of a generic shape, you must know what kind of geometric shape you have because each shape (eg. square, circle, rectangle, etc.) has its own area and perimeter formulas.

However, as well as being abstract a shape is also a generalisation (because it "encapsulates common state and behavior for a category of objects" where in this case the objects are shapes).

Going back to the example I given about Jets and FighterJets, a Jet is not abstract as a concrete instance of a Jet is feasible, as one can exist in the real world, unlike a shape i.e. in the real world you cant hold a shape you hold an instance of a shape e.g. a cube. So in the aircraft example, a Jet is not abstract, it is a generalisation as it is possible to have a "concrete" instance of a jet.

2. Learning that is guided and unsupervised

A learning model is a mathematical representation of a tangible process. If the learning model is supervised, the inputs of the algorithm are labeled. The algorithm receives a paired dataset that includes the sample and the label for that sample. These inputs are also called observations, examples, or instances. If the datasets are unlabeled, the algorithm is unsupervised and must categorize, compute, and deliver outputs on its own with no predefined parameters.

3. Regression and classification

The main difference between Regression and Classification algorithms that Regression algorithms are used to predict the continuous values such as price, salary, age, etc. and Classification algorithms are used to predict/Classify the discrete values such as Male or Female, True or False, Spam or Not Spam, etc.