1. What is the concept of supervised learning? What is the significance of the name?

   Supervised learning, also known as supervised machine learning, is a subcategory of machine learning and artificial intelligence. It is defined by its use of labelled datasets to train algorithms that to classify data or predict outcomes accurately. As input data is fed into the model, it adjusts its weights until the model has been fitted appropriately, which occurs as part of the cross-validation process. Supervised learning helps organizations solve for a variety of real-world problems at scale, such as classifying spam in a separate folder from your inbox. In supervised learning, the training data provided to the machines work as the supervisor that teaches the machines to predict the output correctly. It applies the same concept as a student learns in the supervision of the teacher.

2. In the hospital sector, offer an example of supervised learning.

   Patient behavior modification
   Many prevalent diseases are manageable or even avoidable. For example, type 2 diabetes, obesity, and heart diseases in some cases can be avoided by practicing a healthier lifestyle. However, adjusting the lifestyle requires a change in behavior, which is not a one-time effort but one requiring constant reminders and follow-ups. For this purpose, machine learning algorithms can aggregate data incoming from patients' connected health devices and sensors to generate insights into these patients' behavior and guide them during this transformational journey.

   Altering smoking habits with SmokeBeat
   SmokeBeat is an innovative application that passively gathers data on the user's smoking behavior. The application uses an accelerometer on a smartwatch or a smart band to detect hand-to-mouth gestures. SmokeBeat processes this data and offers real-time Cognitive Behavior Therapy incentives. User responses to those incentives are constantly measured and recorded to improve effectiveness. Additionally, SmokeBeat compares users' smoking data with their peers of choice, creating a sort of supportive social network.

3. Give three supervised learning examples.

   Example: House prices
   First, we need data about the houses: square footage, number of rooms, features, whether a house has a garden or not, and so on. We then need to know the prices of these houses, i.e. the corresponding labels. By leveraging data coming from thousands of houses, their features and prices, we can now train a supervised machine learning model to predict a new house's price based on the examples observed by the model.

   Example: Is it a cat or a dog?
   Image classification is a popular problem in the computer vision field. Here, the goal is to predict what class an image belongs to. In this set of problems, we are interested in finding the class label of an image. More precisely: is the image of a car or a plane? A cat or a dog?

   Example: How's the weather today?
   One particularly interesting problem which requires considering a lot of different parameters is predicting weather conditions in a particular location. To make correct

predictions for the weather, we need to take into account various parameters, including historical temperature data, precipitation, wind, humidity, and so on.

4. In supervised learning, what are classification and regression?

   Regression finds correlations between dependent and independent variables. Therefore, regression algorithms help predict continuous variables such as house prices, market trends, weather patterns, oil and gas prices (a critical task these days!), etc.

   The Regression algorithm's task is finding the mapping function so we can map the input variable of "x" to the continuous output variable of "y."

   On the other hand, Classification is an algorithm that finds functions that help divide the dataset into classes based on various parameters. When using a Classification algorithm, a computer program gets taught on the training dataset and categorizes the data into various categories depending on what it learned.

   Classification algorithms find the mapping function to map the "x" input to "y" discrete output. The algorithms estimate discrete values (in other words, binary values such as 0 and 1, yes and no, true or false, based on a particular set of independent variables. To put it another, more straightforward way, classification algorithms predict an event occurrence probability by fitting data to a logit function.

5. Give some popular classification algorithms as examples.

   Logistic Regression
   Naive Bayes
   K-Nearest Neighbors
   Decision Tree
   Support Vector Machines

6. Briefly describe the SVM model.

   A support vector machine (SVM) is a type of deep learning algorithm that performs supervised learning for classification or regression of data groups.
   In AI and machine learning, supervised learning systems provide both input and desired output data, which are labeled for classification. The classification provides a learning basis for future data processing. Support vector machines are used to sort two data groups by like classification. The algorithms draw lines (hyperplanes) to separate the groups according to patterns.

   An SVM builds a learning model that assigns new examples to one group or another. By these functions, SVMs are called a non-probabilistic, binary linear classifier. In probabilistic classification settings, SVMs can use methods such as Platt Scaling.

Like other supervised learning machines, an SVM requires labeled data to be trained. Groups of materials are labeled for classification. Training materials for SVMs are classified separately in different points in space and organized into clearly separated groups. After processing numerous training examples, SVMs can perform unsupervised learning. The algorithms will try to achieve the best separation of data with the boundary around the hyperplane being maximized and even between both sides.

7. In SVM, what is the cost of misclassification?

   Misclassification costs are basically weights applied to specific outcomes. These weights are factored into the model and may actually change the prediction (as a way of protecting against costly mistakes).

   With the exception of C5.0 models, misclassification costs are not applied when scoring a model and are not taken into account when ranking or comparing models using an Auto Classifier node, evaluation chart, or Analysis node. A model that includes costs may not produce fewer errors than one that doesn't and may not rank any higher in terms of overall accuracy, but it is likely to perform better in practical terms because it has a built-in bias in favor of less expensive errors.

8. In the SVM model, define Support Vectors.

   Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane. Using these support vectors, we maximize the margin of the classifier. Deleting the support vectors will change the position of the hyperplane

9. In the SVM model, define the kernel.

   “Kernel” is used due to a set of mathematical functions used in Support Vector Machine providing the window to manipulate the data. So, Kernel Function generally transforms the training set of data so that a non-linear decision surface is able to transform to a linear equation in a higher number of dimension spaces.

10. What are the factors that influence SVM's effectiveness?

    The effectiveness of SVM depends on the selection of kernel, kernel's parameters and soft margin parameter C. . Each pair of parameters is checked using cross validation, and the parameters with best cross validation accuracy are picked.

11. What are the benefits of using the SVM model?

    The advantages of SVM and support vector regression include that they can be used to avoid the difficulties of using linear functions in the high-dimensional feature space, and the optimization problem is transformed into dual convex quadratic programs.

12. What are the drawbacks of using the SVM model?

    It does not execute very well when the data set has more sound i.e. target classes are overlapping. In cases where the number of properties for each data point outstrips the number of training data specimens, the support vector machine will underperform.

13. What are some of the benefits of the kNN algorithm?

    Quick calculation time.
    Simple algorithm – to interpret.
    Versatile – useful for regression and classification.
    High accuracy – you do not need to compare with better-supervised learning models.

14. What are some of the kNN algorithm's drawbacks?

    Accuracy depends on the quality of the data.
    With large data, the prediction stage might be slow.
    Sensitive to the scale of the data and irrelevant features.
    Require high memory – need to store all of the training data.
    Given that it stores all of the training, it can be computationally expensive.

15. Explain the decision tree algorithm in a few words.

    A decision tree is a non-parametric supervised learning algorithm, which is utilized for both classification and regression tasks. It has a hierarchical, tree structure, which consists of a root node, branches, internal nodes and leaf nodes.

16. What is the difference between a node and a leaf in a decision tree?

    The final result is a tree with decision nodes and leaf nodes. A decision node (e.g., Outlook) has two or more branches (e.g., Sunny, Overcast and Rainy). Leaf node (e.g., Play) represents a classification or decision.

17. What is a decision tree's entropy?

    In the context of Decision Trees, entropy is a measure of disorder or impurity in a node. Thus, a node with more variable composition, such as 2Pass and 2 Fail would be considered to have higher Entropy than a node which has only pass or only fail.

18. In a decision tree, define knowledge gain.

    information gain is a synonym for Kullback–Leibler divergence; the amount of information gained about a random variable or signal from observing another random variable. However, in the context of decision trees, the term is sometimes used synonymously with mutual information, which is the conditional expected value of the Kullback–Leibler divergence of the univariate probability distribution of one variable from the conditional distribution of this variable given the other one.

19. Choose three advantages of the decision tree approach and write them down.

    It's Great for Making Decisions.
    It is an All-Inclusive Algorithm.
    Simple to Understand for Coders.
    Missing Values Aren't an Issue.
    They are Inexpensive.

20. Make a list of three flaws in the decision tree process.

    Decision Trees are very much prone to Over-fitting of data.
    Performance wise its very weak.
    Can’t learn complex relationships.
    There are no assumptions of distribution of data.

21. Briefly describe the random forest model.

    Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.