1. What Are the Different Types of Machine Learning?

There are three types of machine learning:

Supervised Learning

In supervised machine learning, a model makes predictions or decisions based on past or labeled data. Labeled data refers to sets of data that are given tags or labels, and thus made more meaningful.

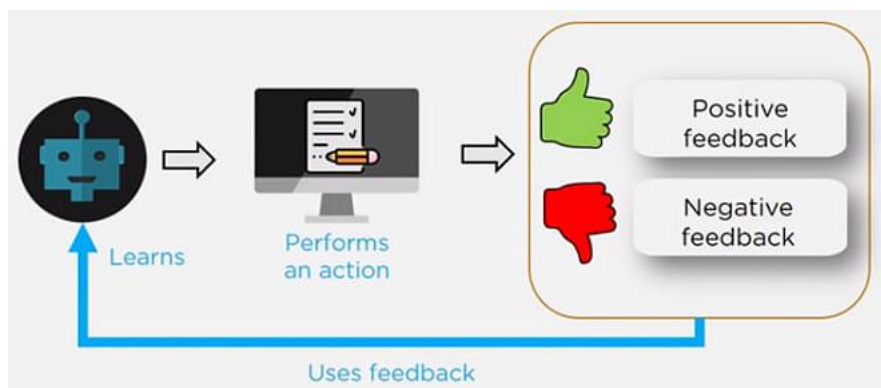Past data | New data | Predicted output

Unsupervised Learning

In unsupervised learning, we don't have labeled data. A model can identify patterns, anomalies, and relationships in the input data.
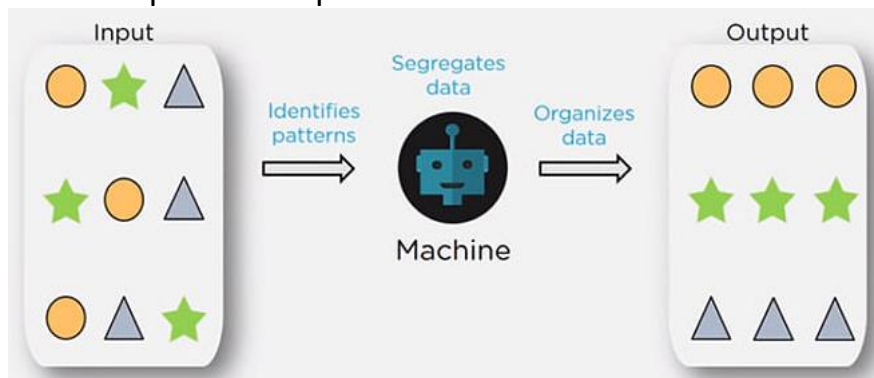
Reinforcement Learning

Using reinforcement learning, the model can learn based on the rewards it received for its previous action.

Consider an environment where an agent is working. The agent is given a target to achieve. Every time the agent takes some action toward the target, it is given positive feedback. And, if the action taken is going away from the goal, the agent is given negative feedback.

2. What is Overfitting, and How Can You Avoid It?

The Overfitting is a situation that occurs when a model learns the training set too well, taking up random fluctuations in the training data as concepts. These impact the model's ability to generalize and don't apply to new data.

When a model is given the training data, it shows 100 percent accuracy—technically a slight loss. But, when we use the test data, there may be an error and low efficiency. This condition is known as overfitting.

There are multiple ways of avoiding overfitting, such as:

- Regularization. It involves a cost term for the features involved with the objective function
- Making a simple model. With lesser variables and parameters, the variance can be reduced
- Cross-validation methods like k-folds can also be used
- If some model parameters are likely to cause overfitting, techniques for regularization like LASSO can be used that penalize these parameters

3. What is 'training Set' and 'test Set' in a Machine Learning Model? How Much Data Will You Allocate for Your Training, Validation, and Test Sets?

There is a three-step process followed to create a model:

1. Train the model
2. Test the model
3. Deploy the model

Training Set	Test Set
- The training set is examples given to the model to analyze and learn 70% of the total data is typically taken as the training dataset - This is labeled data used to train the model	- The test set is used to test the accuracy of the hypothesis generated by the model - Remaining 30% is taken as testing dataset - We test without labeled data and then verify results with labels

Consider a case where you have labeled data for 1,000 records. One way to train the model is to expose all 1,000 records during the training process. Then you take a small set of the same data to test the model, which would give good results in this case.

But, this is not an accurate way of testing. So, we set aside a portion of that data called the 'test set' before starting the training process. The remaining data is called the 'training set' that we use for training the model. The training set passes through the model multiple times until the accuracy is high, and errors are minimized.
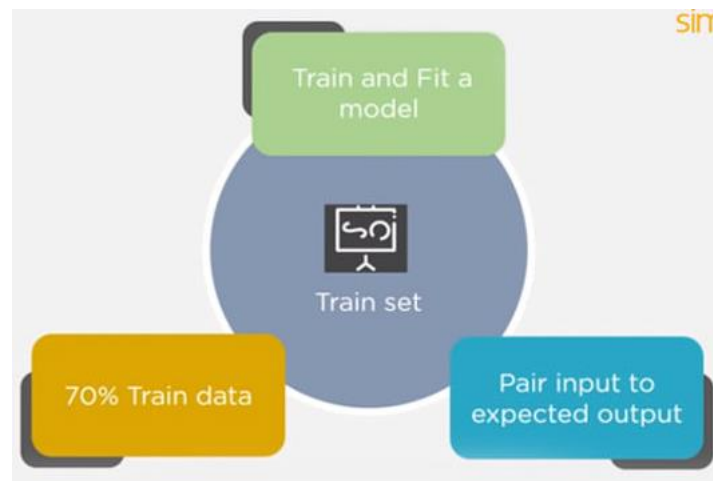
Now, we pass the test data to check if the model can accurately predict the values and determine if training is effective. If you get errors, you either need to change your model or retrain it with more data.
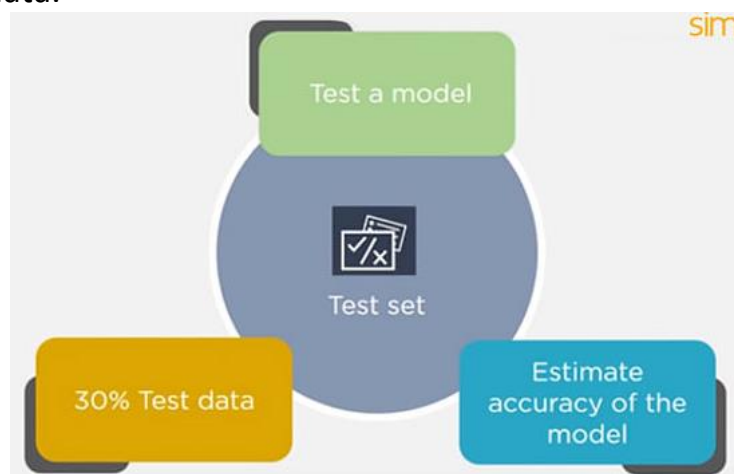

Regarding the question of how to split the data into a training set and test set, there is no fixed rule, and the ratio can vary based on individual preferences.

4. How Do You Handle Missing or Corrupted Data in a Dataset?

One of the easiest ways to handle missing or corrupted data is to drop those rows or columns or replace them entirely with some other value.

There are two useful methods in Pandas:

- IsNull() and dropna() will help to find the columns/rows with missing data and drop them
- Fillna() will replace the wrong values with a placeholder value
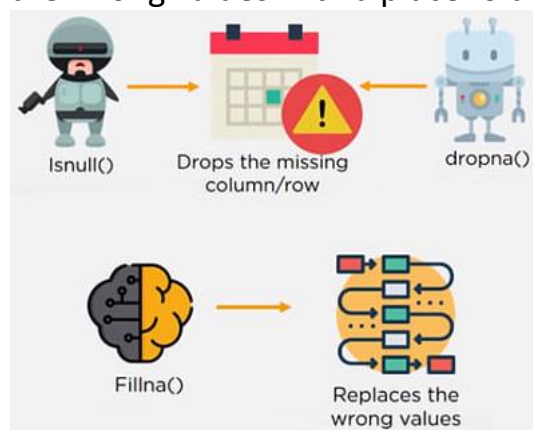
5. How Can You Choose a Classifier Based on a Training Set Data Size?

When the training set is small, a model that has a right bias and low variance seems to work better because they are less likely to overfit.

For example, Naive Bayes works best when the training set is large. Models with low bias and high variance tend to perform better as they work fine with complex relationships.

6. Explain the Confusion Matrix with Respect to Machine Learning Algorithms.

A confusion matrix (or error matrix) is a specific table that is used to measure the performance of an algorithm. It is mostly used in supervised learning; in unsupervised learning, it's called the matching matrix.

The confusion matrix has two parameters:
- Actual
- Predicted

It also has identical sets of features in both of these dimensions.

Consider a confusion matrix (binary matrix) shown below:

Confusion Matrix

Here,

For actual values:

Total Yes = 12+1 = 13

Total No = 3+9 = 12

Similarly, for predicted values:

Total Yes = 12+3 = 15

Total No = 1+9 = 10

For a model to be accurate, the values across the diagonals should be high. The total sum of all the values in the matrix equals the total observations in the test data set.

For the above matrix, total observations = 12+3+1+9 = 25

Now, accuracy = sum of the values across the diagonal/total dataset

= (12+9) / 25

= 21 / 25

= 84%

7. What Is a False Positive and False Negative and How Are They Significant?

False positives are those cases that wrongly get classified as True but are False.

False negatives are those cases that wrongly get classified as False but are True.

In the term 'False Positive,' the word 'Positive' refers to the 'Yes' row of the predicted value in the confusion matrix. The complete term indicates that the system has predicted it as a positive, but the actual value is negative.

Confusion Matrix

So, looking at the confusion matrix, we get:

False-positive = 3

True positive = 12

Similarly, in the term 'False Negative,' the word 'Negative' refers to the 'No' row of the predicted value in the confusion matrix. And the complete term indicates that the system has predicted it as negative, but the actual value is positive.

So, looking at the confusion matrix, we get:

False Negative = 1

True Negative = 9

8. What Are the Three Stages of Building a Model in Machine Learning?

The three stages of building a machine learning model are:

- Model Building

  Choose a suitable algorithm for the model and train it according to the requirement

- Model Testing

  Check the accuracy of the model through the test data

- Applying the Model

  Make the required changes after testing and use the final model for real-time projects

Here, it's important to remember that once in a while, the model needs to be checked to make sure it's working correctly. It should be modified to make sure that it is up-to-date.

9. What is Deep Learning?

The Deep learning is a subset of machine learning that involves systems that think and learn like humans using artificial neural networks. The term 'deep' comes from the fact that you can have several layers of neural networks.

One of the primary differences between machine learning and deep learning is that feature engineering is done manually in machine learning. In the case of deep learning, the model consisting of neural networks will automatically determine which features to use (and which not to use).

This is a commonly asked question asked in both Machine Learning Interviews as well as Deep Learning Interview Questions.

10. What Are the Differences Between Machine Learning and Deep Learning?

Machine Learning	Deep Learning
• Enables machines to take decisions on their own, based on past data • It needs only a small amount of data for training • Works well on the low-end system, so you don't need large machines • Most features need to be identified in advance and manually coded • The problem is divided into two parts and solved individually and then combined	• Enables machines to take decisions with the help of artificial neural networks • It needs a large amount of training data • Needs high-end machines because it requires a lot of computing power • The machine learns the features from the data it is provided • The problem is solved in an end-to-end manner

11. What Are the Applications of Supervised Machine Learning in Modern Businesses?
Applications of supervised machine learning include:
- Email Spam Detection
  Here we train the model using historical data that consists of emails categorized as spam or not spam. This labeled information is fed as input to the model.
- Healthcare Diagnosis
  By providing images regarding a disease, a model can be trained to detect if a person is suffering from the disease or not.
- Sentiment Analysis
  This refers to the process of using algorithms to mine documents and determine whether they're positive, neutral, or negative in sentiment.
- Fraud Detection
  By training the model to identify suspicious patterns, we can detect instances of possible fraud.

12. What is Semi-supervised Machine Learning?
Supervised learning uses data that is completely labeled, whereas unsupervised learning uses no training data.
In the case of semi-supervised learning, the training data contains a small amount of labeled data and a large amount of unlabeled data.
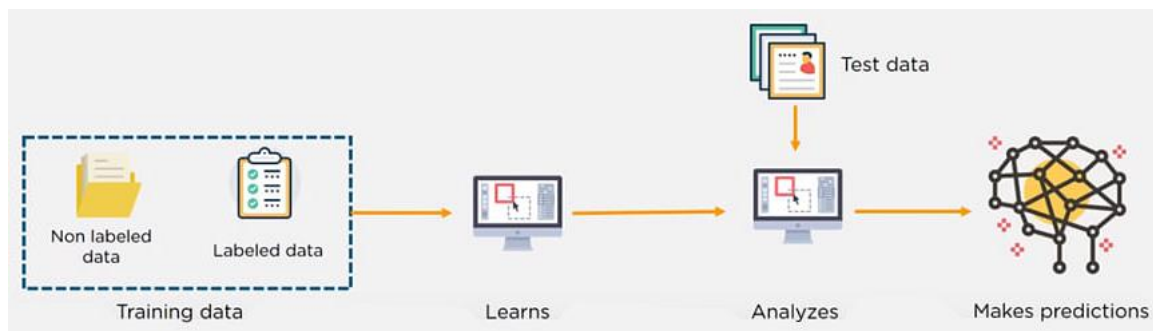
13. What Are Unsupervised Machine Learning Techniques?

There are two techniques used in unsupervised learning: clustering and association.

Clustering

Clustering problems involve data to be divided into subsets. These subsets, also called clusters, contain data that are similar to each other. Different clusters reveal different details about the objects, unlike classification or regression.
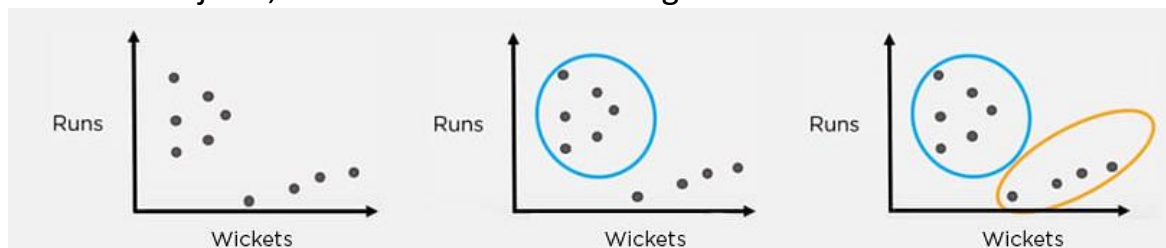
Association

In an association problem, we identify patterns of associations between different variables or items.

For example, an e-commerce website can suggest other items for you to buy, based on the prior purchases that you have made, spending habits, items in your wishlist, other customers' purchase habits, and so on.

14. What is the Difference Between Supervised and Unsupervised Machine Learning?

- Supervised learning - This model learns from the labeled data and makes a future prediction as output
- Unsupervised learning - This model uses unlabeled input data and allows the algorithm to act on that information without guidance.

15. What is the Difference Between Inductive Machine Learning and Deductive Machine Learning?

Inductive Learning	Deductive Learning
- It observes instances based on defined principles to draw a conclusion - Example: Explaining to a child to keep away from the fire by showing a video where fire causes damage	- It concludes experiences - Example: Allow the child to play with fire. If he or she gets burned, they will learn that it is dangerous and will refrain from making the same mistake again

16. Compare K-means and KNN Algorithms.

K-means	KNN
- K-Means is unsupervised - K-Means is a clustering algorithm - The points in each cluster are similar to each other, and each cluster is different from its neighboring clusters	- KNN is supervised in nature - KNN is a classification algorithm - It classifies an unlabeled observation based on its K (can be any number) surrounding neighbors

17. What Is 'naive' in the Naive Bayes Classifier?

The classifier is called 'naive' because it makes assumptions that may or may not turn out to be correct.

The algorithm assumes that the presence of one feature of a class is not related to the presence of any other feature (absolute independence of features), given the class variable. For instance, a fruit may be considered to be a cherry if it is red in color and round in shape, regardless of other features. This assumption may or may not be right (as an apple also matches the description).

18. Explain How a System Can Play a Game of Chess Using Reinforcement Learning.

Reinforcement learning has an environment and an agent. The agent performs some actions to achieve a specific goal. Every time the agent performs a task that is taking it towards the goal, it is rewarded. And, every time it takes a step that goes against that goal or in the reverse direction, it is penalized.

Earlier, chess programs had to determine the best moves after much research on numerous factors. Building a machine designed to play such games would require many rules to be specified.

With reinforced learning, we don't have to deal with this problem as the learning agent learns by playing the game. It will make a move (decision), check if it's the right move

(feedback), and keep the outcomes in memory for the next step it takes (learning). There is a reward for every correct decision the system takes and punishment for the wrong one.
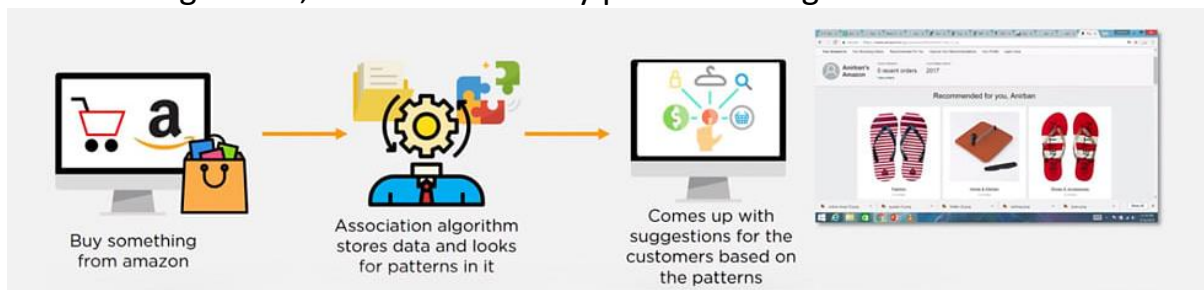
19. How Will You Know Which Machine Learning Algorithm to Choose for Your Classification Problem?

While there is no fixed rule to choose an algorithm for a classification problem, you can follow these guidelines:

- If accuracy is a concern, test different algorithms and cross-validate them
- If the training dataset is small, use models that have low variance and high bias
- If the training dataset is large, use models that have high variance and little bias

20. How is Amazon Able to Recommend Other Things to Buy? How Does the Recommendation Engine Work?

Once a user buys something from Amazon, Amazon stores that purchase data for future reference and finds products that are most likely also to be bought, it is possible because of the Association algorithm, which can identify patterns in a given dataset.

Buy something from amazon → Association algorithm stores data and looks for patterns in it → Comes up with suggestions for the customers based on the patterns

21. When Will You Use Classification over Regression?

Classification is used when your target is categorical, while regression is used when your target variable is continuous. Both classification and regression belong to the category of supervised machine learning algorithms.

Examples of classification problems include:

- Predicting yes or no
- Estimating gender
- Breed of an animal
- Type of color

Examples of regression problems include:

- Estimating sales and price of a product
- Predicting the score of a team
- Predicting the amount of rainfall

22. How Do You Design an Email Spam Filter?

Building a spam filter involves the following process:

- The email spam filter will be fed with thousands of emails
- Each of these emails already has a label: 'spam' or 'not spam.'
- The supervised machine learning algorithm will then determine which type of emails are being marked as spam based on spam words like the lottery, free offer, no money, full refund, etc.

- The next time an email is about to hit your inbox, the spam filter will use statistical analysis and algorithms like Decision Trees and SVM to determine how likely the email is spam
- If the likelihood is high, it will label it as spam, and the email won't hit your inbox
- Based on the accuracy of each model, we will use the algorithm with the highest accuracy after testing all the models

23. What is a Random Forest?

A 'random forest' is a supervised machine learning algorithm that is generally used for classification problems. It operates by constructing multiple decision trees during the training phase. The random forest chooses the decision of the majority of the trees as the final decision.
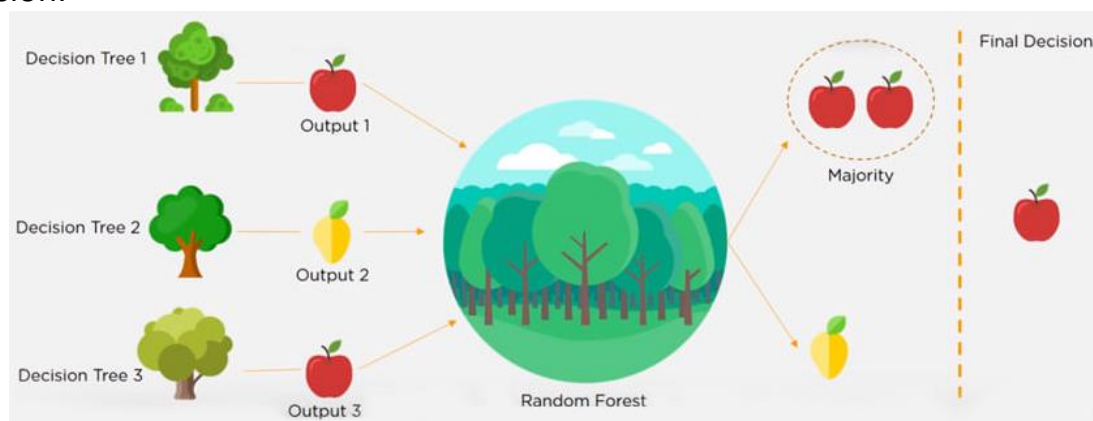

24. Considering a Long List of Machine Learning Algorithms, given a Data Set, How Do You Decide Which One to Use?

There is no master algorithm for all situations. Choosing an algorithm depends on the following questions:

- How much data do you have, and is it continuous or categorical?
- Is the problem related to classification, association, clustering, or regression?
- Predefined variables (labeled), unlabeled, or mix?
- What is the goal?

Based on the above questions, the following algorithms can be used:
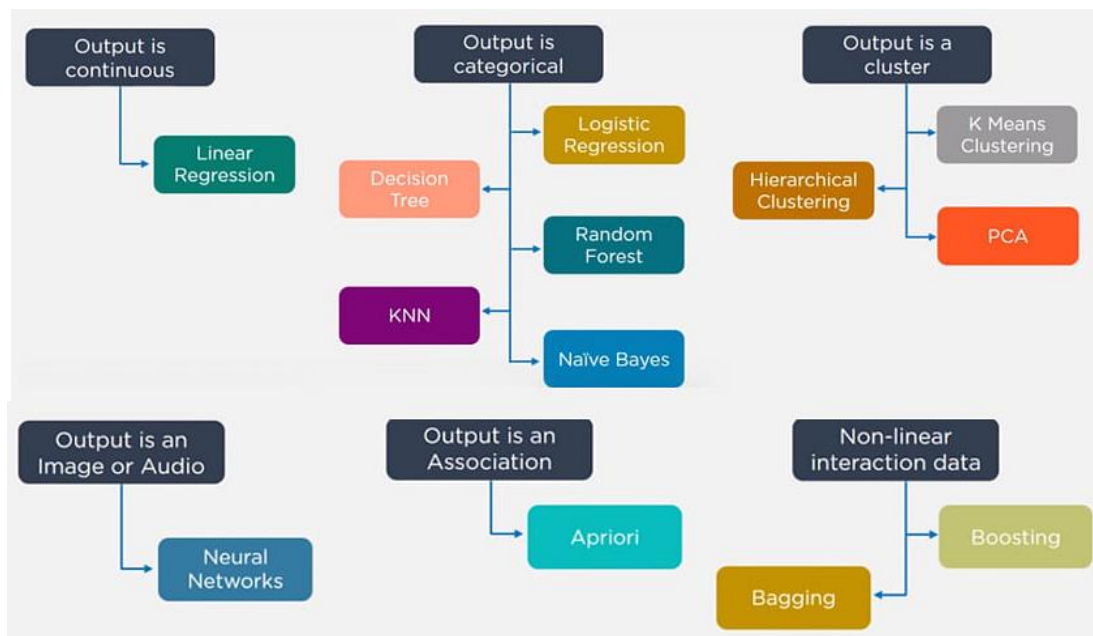
25. What is Bias and Variance in a Machine Learning Model?

Bias

Bias in a machine learning model occurs when the predicted values are further from the actual values. Low bias indicates a model where the prediction values are very close to the actual ones.

Underfitting: High bias can cause an algorithm to miss the relevant relations between features and target outputs.

Variance

Variance refers to the amount the target model will change when trained with different training data. For a good model, the variance should be minimized.

Overfitting: High variance can cause an algorithm to model the random noise in the training data rather than the intended outputs.

26. What is the Trade-off Between Bias and Variance?

The bias-variance decomposition essentially decomposes the learning error from any algorithm by adding the bias, variance, and a bit of irreducible error due to noise in the underlying dataset.

Necessarily, if you make the model more complex and add more variables, you'll lose bias but gain variance. To get the optimally-reduced amount of error, you'll have to trade off bias and variance. Neither high bias nor high variance is desired.

High bias and low variance algorithms train models that are consistent, but inaccurate on average.

High variance and low bias algorithms train models that are accurate but inconsistent.

27. Define Precision and Recall.

Precision

Precision is the ratio of several events you can correctly recall to the total number of events you recall (mix of correct and wrong recalls).

Precision = (True Positive) / (True Positive + False Positive)

Recall

A recall is the ratio of the number of events you can recall the number of total events.

Recall = (True Positive) / (True Positive + False Negative)

28. What is a Decision Tree Classification?

A decision tree builds classification (or regression) models as a tree structure, with datasets broken up into ever-smaller subsets while developing the decision tree, literally in a tree-like way with branches and nodes. Decision trees can handle both categorical and numerical data.

29. What is Pruning in Decision Trees, and How Is It Done?

Pruning is a technique in machine learning that reduces the size of decision trees. It reduces the complexity of the final classifier, and hence improves predictive accuracy by the reduction of overfitting.

Pruning can occur in:

- Top-down fashion. It will traverse nodes and trim subtrees starting at the root
- Bottom-up fashion. It will begin at the leaf nodes

There is a popular pruning algorithm called reduced error pruning, in which:

- Starting at the leaves, each node is replaced with its most popular class
- If the prediction accuracy is not affected, the change is kept
- There is an advantage of simplicity and speed

30. Briefly Explain Logistic Regression.

Logistic regression is a classification algorithm used to predict a binary outcome for a given set of independent variables.

The output of logistic regression is either a 0 or 1 with a threshold value of generally 0.5. Any value above 0.5 is considered as 1, and any point below 0.5 is considered as 0.

Regression model created based on the performance of past participants
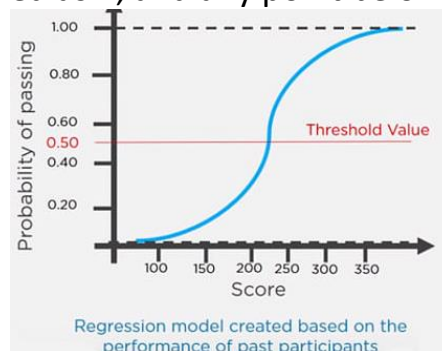
31. Explain the K Nearest Neighbor Algorithm.

K nearest neighbor algorithm is a classification algorithm that works in a way that a new data point is assigned to a neighboring group to which it is most similar.

In K nearest neighbors, K can be an integer greater than 1. So, for every new data point, we want to classify, we compute to which neighboring group it is closest.
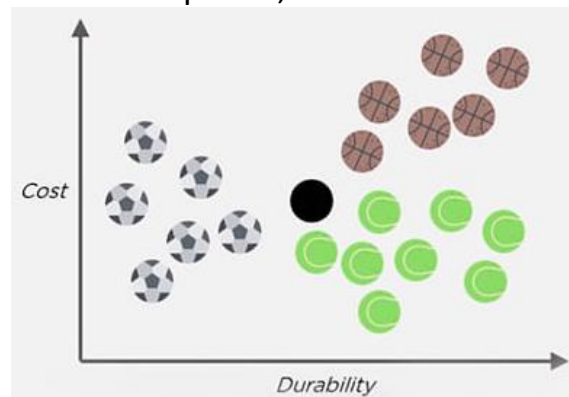
Let us classify an object using the following example. Consider there are three clusters:

- Football
- Basketball
- Tennis ball
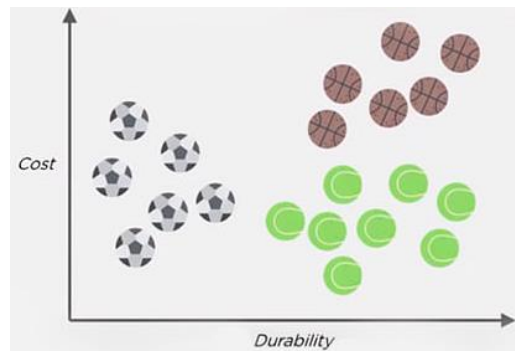
Let the new data point to be classified is a black ball. We use KNN to classify it. Assume K = 5 (initially).

Next, we find the K (five) nearest data points, as shown.

Observe that all five selected points do not belong to the same cluster. There are three tennis balls and one each of basketball and football.

When multiple classes are involved, we prefer the majority. Here the majority is with the tennis ball, so the new data point is assigned to this cluster.

32. What is a Recommendation System?

Anyone who has used Spotify or shopped at Amazon will recognize a recommendation system: It's an information filtering system that predicts what a user might want to hear or see based on choice patterns provided by the user.

33. What is Kernel SVM?

Kernel SVM is the abbreviated version of the kernel support vector machine. Kernel methods are a class of algorithms for pattern analysis, and the most common one is the kernel SVM.

34. What Are Some Methods of Reducing Dimensionality?

You can reduce dimensionality by combining features with feature engineering, removing collinear features, or using algorithmic dimensionality reduction.

Now that you have gone through these machine learning interview questions, you must have got an idea of your strengths and weaknesses in this domain.

35. What is Principal Component Analysis?

Principal Component Analysis or PCA is a multivariate statistical technique that is used for analyzing quantitative data. The objective of PCA is to reduce higher dimensional data to lower dimensions, remove noise, and extract crucial information such as features and attributes from large amounts of data.

36. What do you understand by the F1 score?
The F1 score is a metric that combines both Precision and Recall. It is also the weighted average of precision and recall.
The F1 score can be calculated using the below formula:
F1 = 2 * (P * R) / (P + R)
The F1 score is one when both Precision and Recall scores are one.

37. What do you understand by Type I vs Type II error?
Type I Error: Type I error occurs when the null hypothesis is true and we reject it.
Type II Error: Type II error occurs when the null hypothesis is false and we accept it.

		reality	
		H_0 = True	H_0 = False
Conclusion	H_0 is not rejected	OK	Type II error
	H_0 is rejected	Type I error	OK

38. Explain Correlation and Covariance?
Correlation: Correlation tells us how strongly two random variables are related to each other. It takes values between -1 to +1.
Formula to calculate Correlation:

$$Correlation = \frac{Cov\ (x,\ y)}{6x * 6y}$$

Covariance: Covariance tells us the direction of the linear relationship between two random variables. It can take any value between - ∞ and + ∞.
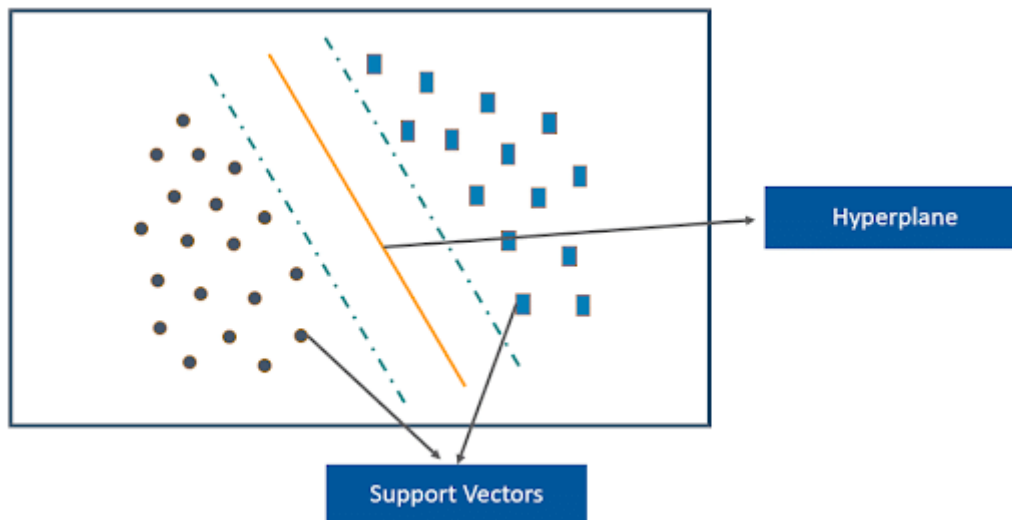Formula to calculate Covariance:

$$Cov(x,\ y) = \frac{\Sigma\ (x_i - x') * (y_i - y')}{N}$$

39. What are Support Vectors in SVM?
Support Vectors are data points that are nearest to the hyperplane. It influences the position and orientation of the hyperplane. Removing the support vectors will alter the

position of the hyperplane. The support vectors help us build our support vector machine model.

40. What is Ensemble learning?

Ensemble learning is a combination of the results obtained from multiple machine learning models to increase the accuracy for improved decision-making.

Example: A Random Forest with 100 trees can provide much better results than using just one decision tree.
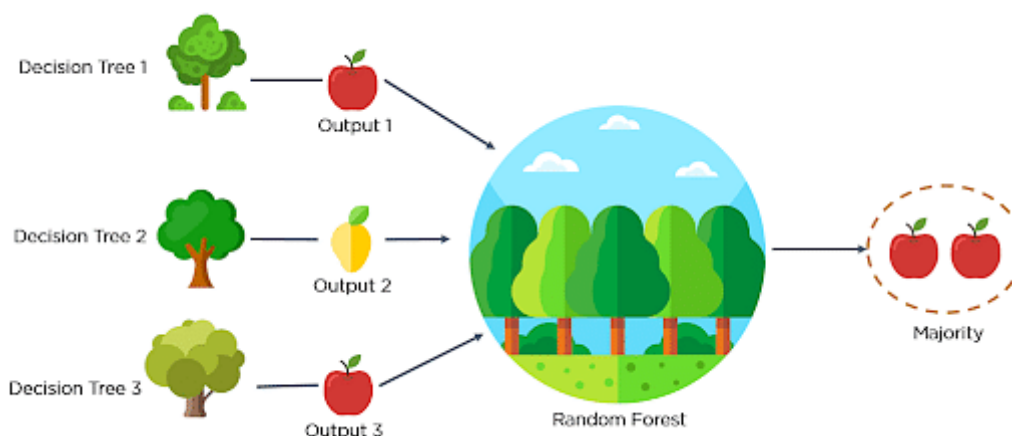
41. What is Cross-Validation?

Cross-Validation in Machine Learning is a statistical resampling technique that uses different parts of the dataset to train and test a machine learning algorithm on different iterations. The aim of cross-validation is to test the model's ability to predict a new set of data that was not used to train the model. Cross-validation avoids the overfitting of data.

K-Fold Cross Validation is the most popular resampling technique that divides the whole dataset into K sets of equal sizes.

42. What are the different methods to split a tree in a decision tree algorithm?

Variance: Splitting the nodes of a decision tree using the variance is done when the target variable is continuous.

$$\text{Variance} = \frac{\sum (X-\bar{X})^2}{N}$$

Information Gain: Splitting the nodes of a decision tree using Information Gain is preferred when the target variable is categorical.

$$IG = 1 - \text{Entropy}$$

$$\text{Entropy} = -\sum p_i \log_2 p_i$$

Gini Impurity: Splitting the nodes of a decision tree using Gini Impurity is followed when the target variable is categorical.

$$I_G(n) = 1 - \sum_{i=1}^{n} (p_i)^2$$

43. How does the Support Vector Machine algorithm handle self-learning?

The SVM algorithm has a learning rate and expansion rate which takes care of self-learning. The learning rate compensates or penalizes the hyperplanes for making all the incorrect moves while the expansion rate handles finding the maximum separation area between different classes.

44. What are the assumptions you need to take before starting with linear regression?

There are primarily 5 assumptions for a Linear Regression model:
- Multivariate normality
- No auto-correlation
- Homoscedasticity
- Linear relationship
- No or little multicollinearity

45. What is the difference between Lasso and Ridge regression?

Lasso(also known as L1) and Ridge(also known as L2) regression are two popular regularization techniques that are used to avoid overfitting of data. These methods are used to penalize the coefficients to find the optimum solution and reduce complexity. The Lasso regression works by penalizing the sum of the absolute values of the coefficients. In

Ridge or L2 regression, the penalty function is determined by the sum of the squares of the coefficients.

Ways to Reduce the reduce Variance in Machine Learning:
- **Cross-validation:** By splitting the data into training and testing sets multiple times, cross-validation can help identify if a model is overfitting or underfitting and can be used to tune hyperparameters to reduce variance.
- **Feature selection:** By choosing the only relevant feature will decrease the model's complexity. and it can reduce the variance error.
- **Regularization:** We can use L1 or L2 regularization to reduce variance in machine learning models
- **Ensemble methods:** It will combine multiple models to improve generalization performance. Bagging, boosting, and stacking are common ensemble methods that can help reduce variance and improve generalization performance.
- **Simplifying the model:** Reducing the complexity of the model, such as decreasing the number of parameters or layers in a neural network, can also help reduce variance and improve generalization performance.
- **Early stopping:** Early stopping is a technique used to prevent overfitting by stopping the training of the deep learning model when the performance on the validation set stops improving.

Different Combinations of Bias-Variance

Machine Learning Algorithm	Bias	Variance
Linear Regression	High Bias	Less Variance
Decision Tree	Low Bias	High Variance
Random Forest	Low Bias	High Variance
Bagging	Low Bias	High Variance

What are some of the hyperparameters of the random forest regressor which help to avoid overfitting?

The most important hyperparameters of a Random Forest are:
- **max_depth:** Sometimes the larger depth of the tree can create overfitting. To overcome it, the depth should be limited.
- **n-estimator:** It is the number of decision trees we want in our forest.
- **min_sample_split:** It is the minimum number of samples an internal node must hold in order to split into further nodes.
- **max_leaf_nodes:** It helps the model to control the splitting of the nodes and in turn, the depth of the model is also restricted.