

Explore More

Subscription : Premium CDAC NOTES & MATERIAL @99



Contact to Join
Premium Group



Click to Join
Telegram Group

<CODEWITHARRAY'S/>

For More E-Notes

Join Our Community to stay Updated

TAP ON THE ICONS TO JOIN!

	codewitharrays.in freelance project available to buy contact on 8007592194	
SR.NO	Project NAME	Technology
1	Online E-Learning Platform Hub	React+Springboot+MySQL
2	PG Mates / RoomSharing / Flat Mates	React+Springboot+MySQL
3	Tour and Travel management System	React+Springboot+MySQL
4	Election commition of India (online Voting System)	React+Springboot+MySQL
5	HomeRental Booking System	React+Springboot+MySQL
6	Event Management System	React+Springboot+MySQL
7	Hotel Management System	React+Springboot+MySQL
8	Agriculture web Project	React+Springboot+MySQL
9	AirLine Reservation System / Flight booking System	React+Springboot+MySQL
10	E-commerce web Project	React+Springboot+MySQL
11	Hospital Management System	React+Springboot+MySQL
12	E-RTO Driving licence portal	React+Springboot+MySQL
13	Transpotation Services portal	React+Springboot+MySQL
14	Courier Services Portal / Courier Management System	React+Springboot+MySQL
15	Online Food Delivery Portal	React+Springboot+MySQL
16	Muncipal Corporation Management	React+Springboot+MySQL
17	Gym Management System	React+Springboot+MySQL
18	Bike/Car ental System Portal	React+Springboot+MySQL
19	CharityDonation web project	React+Springboot+MySQL
20	Movie Booking System	React+Springboot+MySQL

freelance_Project available to buy contact on 8007592194

21	Job Portal web project	React+Springboot+MySql
22	LIC Insurance Portal	React+Springboot+MySql
23	Employee Management System	React+Springboot+MySql
24	Payroll Management System	React+Springboot+MySql
25	RealEstate Property Project	React+Springboot+MySql
26	Marriage Hall Booking Project	React+Springboot+MySql
27	Online Student Management portal	React+Springboot+MySql
28	Resturant management System	React+Springboot+MySql
29	Solar Management Project	React+Springboot+MySql
30	OneStepService LinkLabourContractor	React+Springboot+MySql
31	Vehical Service Center Portal	React+Springboot+MySql
32	E-wallet Banking Project	React+Springboot+MySql
33	Blogg Application Project	React+Springboot+MySql
34	Car Parking booking Project	React+Springboot+MySql
35	OLA Cab Booking Portal	React+NextJs+Springboot+MySql
36	Society management Portal	React+Springboot+MySql
37	E-College Portal	React+Springboot+MySql
38	FoodWaste Management Donate System	React+Springboot+MySql
39	Sports Ground Booking	React+Springboot+MySql
40	BloodBank mangement System	React+Springboot+MySql

41	Bus Tickit Booking Project	React+Springboot+MySQL
42	Fruite Delivery Project	React+Springboot+MySQL
43	Woodworks Bed Shop	React+Springboot+MySQL
44	Online Dairy Product sell Project	React+Springboot+MySQL
45	Online E-Pharma medicine sell Project	React+Springboot+MySQL
46	FarmerMarketplace Web Project	React+Springboot+MySQL
47	Online Cloth Store Project	React+Springboot+MySQL
48	Train Ticket Booking Project	React+Springboot+MySQL
49	Quizz Application Project	JSP+Springboot+MySQL
50	Hotel Room Booking Project	React+Springboot+MySQL
51	Online Crime Reporting Portal Project	React+Springboot+MySQL
52	Online Child Adoption Portal Project	React+Springboot+MySQL
53	online Pizza Delivery System Project	React+Springboot+MySQL
54	Online Social Complaint Portal Project	React+Springboot+MySQL
55	Electric Vehical management system Project	React+Springboot+MySQL
56	Online mess / Tiffin management System Project	React+Springboot+MySQL
57		React+Springboot+MySQL
58		React+Springboot+MySQL
59		React+Springboot+MySQL
60		React+Springboot+MySQL

Spring Boot + React JS + MySQL Project List

Sr.No	Project Name	YouTube Link
1	Online E-Learning Hub Platform Project	https://youtu.be/KMjyBaWmgzg?si=YckHuNzs7eC84-IW
2	PG Mate / Room sharing/Flat sharing	https://youtu.be/4P9clHg3wvk?si=4uEsi0962CG6Xodp
3	Tour and Travel System Project Version 1.0	https://youtu.be/-UHOBywHaP8?si=KHHfE_A0uv725f12
4	Marriage Hall Booking	https://youtu.be/VXz0kZQi5to?si=IiOS-QG3TpAFP5k7
5	Ecommerce Shopping project	https://youtu.be/vJ_C6LkhrZ0?si=YhcBylSErvdn7paq
6	Bike Rental System Project	https://youtu.be/FIzsAmIBCbk?si=7ujQTJqEgkQ8ju2H
7	Multi-Restaurant management system	https://youtu.be/pvV-pM2Jf3s?si=PgvnT-yFc8ktrDxB
8	Hospital management system Project	https://youtu.be/lynLouBZvY4?si=CXzQs3BsRkjKhZCw
9	Municipal Corporation system Project	https://youtu.be/cVMx9NVyl4I?si=qX0oQt-GT-LR_5iF
10	Tour and Travel System Project version 2.0	https://youtu.be/_4u0mB9mHXE?si=gDiAhKBowi2gNUKZ

Sr.No	Project Name	YouTube Link
11	Tour and Travel System Project version 3.0	https://youtu.be/Dm7nOdpasWg?si=P_Lh2gcOFhlyudug
12	Gym Management system Project	https://youtu.be/J8_7Zrkg7ag?si=LcxV51ynfUB7OptX
13	Online Driving License system Project	https://youtu.be/3yRzsMs8TLE?si=JRI_z4FDx4Gmt7fn
14	Online Flight Booking system Project	https://youtu.be/m755rOwdk8U?si=HURvAY2VnizlyJlh
15	Employee management system project	https://youtu.be/ID1iE3W_GRw?si=Y_jv1xV_BljhrD0H
16	Online student school or college portal	https://youtu.be/4A25aEKfei0?si=RoVgZtxMk9TPdQvD
17	Online movie booking system project	https://youtu.be/Lfjv_U74SC4?si=fiDvrhhrjb4KSISm
18	Online Pizza Delivery system project	https://youtu.be/Tp3izreZ458?si=8eWA OzA8SVdNwlyM
19	Online Crime Reporting system Project	https://youtu.be/0UlzReSk9tQ?si=6vN0e70TVY1GOwPO
20	Online Children Adoption Project	https://youtu.be/3T5HC2HKyT4?si=bntP78niYH802i7N

Machine Learning Interview Questions

1) What are the basic differences between Machine Learning and Deep Learning?

Differences between [Machine Learning](#) and Deep Learning are:

	Machine Learning	Deep Learning
Definition	Sub-discipline of AI	A subset of machine learning
Data	Parses the data	Creates an artificial neural network
Accuracy	Requires manual intervention means decreased accuracy	Self-learning capabilities mean higher accuracy
Interpretability	Machine Learning is Faster	10 Times Faster than ML
Output	ML models produce a numerical output	DL algorithms can range from an image to text or even an audio
Data dependencies	High	Low
Hardware dependencies	Can work on low-end machines.	Heavily depend on high-end machines
Future	Effective with image recognition and face recognition in mobiles	Not much effective due to data processing limitations

2) What is the difference between Bias and Variance?

- Bias:** Bias can be defined as a situation where an error has occurred due to the use of assumptions in the learning algorithm.
- Variance:** Variance is an error caused because of the complexity of the algorithm that is been used to analyze the data.

3) What is the difference between supervised and unsupervised machine learning?

Supervised learning is a process where it requires training labeled data. When it comes to Unsupervised learning doesn't require data labeling.

For interesting tech tutorials and career insights, subscribe to

 **MindMajix** YouTube Channel



4) What are the three stages of model building in machine learning?

Following are the three stages of model building:

- **Model Building:** In this stage, we will choose the ideal algorithm for the model, and we will train it based on our requirements.
- **Model Testing:** In this stage, we will check the model's accuracy by using test data.
- **Applying Model:** After testing, we have to make the changes, and then we can use the model for real-time projects.

5) What are the applications of supervised machine learning?

Following are the applications of machine learning:

1. **Fraud Identification:** Supervised learning trains the model for identifying the suspicious patterns; we can identify the feasible fraud instances.
2. **Healthcare:** By giving images about a disease, supervised machine learning can train the model for detecting whether a person is affected by illness or not.
3. **Email spam identification:** We train the model through historical data which contains emails that are classified as spam or not spam. This labeled data is supplied as the input to the model.
4. **Sentiment Analysis:** This relates to the process of using algorithms for mining the documents and determining if they are negative, neutral, positive in sentiment.

Related Article - [Artificial Intelligence Vs Machine Learning](#)

6) What are the techniques of Unsupervised machine learning?

Following are the different techniques of unsupervised machine learning:

1. **Clustering:** It includes the data that must be divided into subsets. These subsets are also known as clusters. Diverse clusters disclose details about objects, unlike regression or classification.
2. **Association:** In the association problem, we can recognize the association patterns between different items and variables. For instance, e-commerce can indicate other items for us to buy according to our previous purchases.

7) What are the different types of Machine Learning?

Following are the different kinds of machine learning:

- **Unsupervised Learning:** In this kind of machine learning, we will not have labeled data. A model can recognize anomalies, relationships, and patterns in the input data.
- **Supervised Learning:** In this kind of machine learning, the model makes decisions or predictions according to the labeled or past data. Labeled data relates to data sets that provide labels or tags.
- **Reinforcement Learning:** In reinforcement learning, a model can learn according to the rewards it obtained from its past actions.

8) What is Deep Learning?

Deep learning is a branch of machine learning which is relevant to neural networks. Deep learning tells us how to use the principles and backpropagation from neuroscience to the large sets of semi-structured or unlabelled data. Deep learning portrays the unsupervised learning algorithm which learns data representation by using neural nets

Explore - [Deep Learning Tools](#) for more information

9) How to Build a Data Pipeline?

Data pipelines are the core of the machine learning engineers, which take data science models and discover methods for scaling and automating them if you are accustomed to the tools for building the platforms and data pipelines where we can host pipelines and models.

10) How is KNN different from k-means clustering?

K-Nearest Neighbours is a supervised algorithm, and k-means clustering is an unsupervised algorithm. For the K-nearest neighbors to work, we require labeled data to classify the unlabeled point. K-means clustering needs only a threshold and a group of unlabeled points: the algorithm takes the unlabeled points and slowly learns how to divide them into groups by calculating the mean of distance between the points.

11) Comparison between Machine Learning and Big Data

Machine Learning Vs Big Data

Feature	Machine Learning	Big Data
Data Use	Technology that helps in reducing human intervention.	Data research, especially when working with huge data.
Operations	Existing data helps to teach machine what can be done further	Design patterns with analytics on existing data in terms of decision making.
Pattern Recognition	Similar to Big Data, existing data helps in pattern recognition.	Sequence and classification analysis helps in pattern recognition.
Data Volume	Best performance, while working with small datasets.	Datasets help in understanding and solving problems associated with large data volumes.
Application	Read existing data to predict future information.	Storing and analyzing patterns within huge data volumes.

12) Explain what is precision and Recall?

- **Recall:** It is known as a true positive rate. The number of positives that your model has claimed compared to the actual defined number of positives available throughout the data.
- **Precision:** It is also known as a positive predicted value. This is more based on the prediction. It is a measure of the number of accurate positives that the model claims when compared to the number of positives it actually

claims.

13) What is your favorite algorithm and also explain the algorithm briefly in a minute?

This type of question is very common and asked by the interviewers to understand the candidate's skills and assess how well he can communicate complex theories in the simplest language.

This one is a tough question and usually, individuals are not at all prepared for this situation so please be prepared and have a choice of algorithms and make sure you practice a lot before going into any sort of interviews.

Related Article - Machine Learning Applications

14) What is the difference between Type1 and Type2 errors?

Type 1 error is classified as a false positive. I.e. This error claims that something has happened but the fact is nothing has happened. It is like a false fire alarm. The alarm rings but there is no fire.

Type 2 error is classified as a false negative. I.e. This error claims that nothing has happened but the fact is that actually, something happened at the instance.

The best way to differentiate a type 1 vs type 2 error is:

- Calling a man to be pregnant- This is a Type 1 example
- Calling pregnant women and telling them that she isn't carrying any baby- This is a type 2 example

15) Define what is Fourier Transform in a single sentence?

A process of decomposing generic functions into a superposition of symmetric functions is considered to be a Fourier Transform.

16) What is deep learning?

Deep learning is a process where it is considered to be a subset of the machine learning process.

17) What is the F1 score?

The F1 score is defined as a measure of a model's performance.

18) How is the F1 score is used?

The average Precision and Recall of a model is nothing but an F1 score measure. Based on the results, the F1 score is 1 then it is classified as best and 0 being the worst.

19) How can you ensure that you are not overfitting with a particular model?

In Machine Learning concepts, there are three main methods or processes to avoid overfitting:

Firstly, keep the model simple

Must and should use cross-validation techniques

It is mandatory to use regularization techniques, for example, LASSO.

20) How to handle or missing data in a dataset?

An individual can easily find missing or corrupted data in a data set either by dropping the rows or columns. On contrary, they can decide to replace the data with another value.

In Pandas there are two ways to identify the missing data, these two methods are very useful.

isnull() and dropna().

21) Do you have any relevant experience on Spark or any of the big data tools that are used for Machine Learning?

Well, this sort of question is tricky to answer and the best way to respond back is, to be honest. Make sure you are familiar with Big data is and the different tools that are available. If you know about Spark then it is always good to talk about it and if you are unsure then it is best, to be honest, and let the interviewer know about it.

So for this, you have to prepare what is Spark and it's good to prepare other available Big data tools that are used for Machine learning.

Related Article - [Machine Learning with Python](#)

22) Pick an algorithm and write a Pseudocode for the same?

This question depicts your understanding of the algorithm. This is something that one has to be very creative and also should have in-depth knowledge about the algorithms and first and foremost the individual should have a good understanding of the algorithms. The best way to answer this question would be to start off with Web Sequence Diagrams.

23) What is the difference between an array and a Linked list?

An array is an ordered fashion of collection of objects while a linked list is a series of objects that are processed in sequential order.

24): Define a hash table?

They are generally used for database indexing. A hash table is nothing but a data structure that produces an associative array.

25) Mention any one of the data visualization tools that you are familiar with?

This is another question where one has to be completely honest and also giving out your personal experience with these type of tools are really important. Some of the data visualization tools are Tableau, Plot.ly, and matplotlib.

26) What is your opinion on our current data process?

This type of question is asked and the individuals have to carefully listen to their use case and at the same time, the reply should be in a constructive and insightful manner. Based on your responses, the interviewer will have a chance to review and understand whether you are a value add to their team or not.

27) Please let us know what was your last read book or learning paper on Machine Learning?

This type of question is asked to see whether the individual has a keen interest in learning and also he is up to the latest market standards. This is something that every candidate should be looking out for and it is vital for individuals to read through the latest publishings.

28) What is your favorite use case for machine learning models?

The decision tree is one of my favorite use cases for machine learning models.

29) Is rotation necessary in PCA?

Yes, rotation is definitely necessary because it maximizes the differences between the variance captured by the components.

30) What happens if the components are not rotated in PCA?

It is a straight effect. If the components are not rotated then it will diminish eventually and one has to use a lot of various components to explain the data set variance.

31) Explain why Naive Bayes is so Naive?

It is based on an assumption that all of the features in the data set are important, equal, and independent.

32) How Recall and True positive rates are related?

The relation is True Positive Rate = Recall.

33) Assume that you are working on a data set, explain how would you select important variables?

The following are a few methods that can be used to select important variables:

1. Use of Lasso Regression method.
2. Using Random Forest, plot variable importance chart.
3. Using Linear regression.

34) Explain how we can capture the correlation between continuous and categorical variables?

Yes, it is possible by using the ANCOVA technique. It stands for Analysis of Covariance. It is used to calculate the association between continuous and categorical variables.

35) Explain the concept of machine learning and assume that you are explaining this to a 5-year-old baby?

Yes, the question itself is the answer.

Machine learning is exactly the same way how babies do their day-to-day activities, the way they walk or sleep, etc. It is a common fact that babies cannot walk straight away and they fall and then they get up again and then try. This is the same thing when it comes to machine learning, it is all about how the algorithm is working and at the same time redefining every time to make sure the end result is as perfect as possible.

One has to take real-time examples while explaining these questions.

36) What is the difference between Machine learning and Data Mining?

Data mining is about working on unstructured data and then extract it to a level where interesting and unknown patterns are identified. Machine learning is a process or a study whether it closely relates to the design, development of the algorithms that provide an ability to the machines to capacity to learn.

37) What is inductive machine learning?

Inductive machine learning is all about a process of learning by live examples.

38) Please state a few popular Machine Learning algorithms?

Few popular [Machine Learning algorithms](#) are:

1. Nearest Neighbour
2. Neural Networks
3. Decision Trees etc
4. Support vector machines

39) What are the different types of algorithm techniques are available in machine learning?

Some of them are :

1. Supervised learning
2. Unsupervised learning
3. Semi-supervised learning
4. Transduction

5. Learning to learn

40) What are the three stages to build the model in machine learning?

The three stages to build the model in machine learning is:

1. Model building
2. Model testing
3. Applying the model

41) Explain how the ROC curve works?

A ROC curve (receiver operating characteristic) is a graph that shows the performance of a classification model at all classification thresholds. It plots two parameters -

- True positive rate
- False-positive rate

True Positive Rate (TPR) is defined as follows:

$$TPR = TP/(TP+FN)$$

False Positive Rate (FPR) is defined as follows:

$$FPR = FP/(FP+TN)$$

42) What is the difference between L1 and L2 regularization?

Regularization is a process of introducing some information in order to prevent overfitting.

L1 Regularization	L2 Regularization
It is more binary/sparse	Tends to spread error among all the terms
L1 regularization corresponds to setting a Laplacean prior on the terms	It corresponds to a Gaussian prior

43) What is a type 1 and type 2 error?

- Type 1 Error - Type 1 error also called false positive, is asserting something true when it is actually false.
- Type 2 Error - Type 2 error also called false negative, is a test result indicating that a condition is failed but in actuality it is successful.

44) What is the difference between machine learning and deep learning?

Deep learning is a subset of machine learning and is called so because it makes use of deep neural networks. Let's find out machine learning Vs Deep learning

	Machine learning	Deep Learning
Data dependencies	Performs better on small and medium datasets	Works better for big datasets
Hardware dependencies	Work on low-end machines	Requires powerful machine, preferably with GPU
Interpretability	Algorithms are easy to interpret	Difficult to interpret
Execution time	From a few minutes to hours	It May take up to a week
Feature Engineering	Need to understand the features that represent the data	No need to understand the best feature that represents the data

45) What is Bayes Theorem and how it is used in machine learning?

Bayes theorem is a way of calculating conditional probability ie. finding the probability of an event occurring based on the given probability of other events that have already occurred. Mathematically, it is stated as -

$$P(A|B) = \{P(B|A).P(A)\}/P(B)$$

Bayes theorem has become a very useful tool in applied machine learning. It provides a way of thinking about the relationship shared by data and the models.

A machine learning model is a specific way of thinking about the structured relationship in the data such as relationships shared by input (x) and output (y).

If we have some prior domain knowledge about the hypothesis, Then the Bayes theorem can help in solving machine learning problems.

46) What is cross-validation techniques would you be using on a time series dataset?

Cross-validation is used for tuning the hyperparameters and producing measurements of model performance. With the time series data, we can't use the traditional cross-validation technique due to two main reasons which are as follows -

- Temporal dependencies
- Arbitrary Choice of Test Set

For time-series data, we use nested cross-validation that provides an almost unbiased estimate of the true error. A nested CV consists of an inner loop for parameter tuning and an outer loop for error estimation.

47) What are the Discriminative and generative models?

To understand these terms better, let us consider an example. Suppose a person has two kids - kid A and kid B. Kid A learns and understands everything in depth whereas Kid B can only learn the differences between what he sees. One day, that person took them to the zoo where they saw a deer and a lion. After coming from the zoo, the person showed them an animal and asked them what it was. Kid A drew the images of both the animals he saw in the zoo. He compared the images and answered "the animal is deer" based on the closest match of the image. As Kid B learns things based on only differences, therefore, he easily answered: "the animal is deer.".

In ML, we call Kid A a Generative Model and Kid B a Discriminative Model. To make it more clear, the Generative Model learns the joint probability distribution $p(x,y)$. It predicts the conditional probability using Bayes Theorem. Whereas a Discriminative model predicts the conditional probability distribution $p(y|x)$. Both of these models are used in supervised learning problems.

48) How is a decision tree pruned?

In Machine learning, pruning means simplifying and optimizing a decision tree by cutting nodes of the tree that causes overfitting. The pruning process can be divided into two types -

- **Bottom-up pruning** - procedure starts at the last node
- **Top-down pruning** - procedure starts at the root node

Pruning is done to increase the predictive accuracy of a decision tree model.

49) What is more important - model accuracy or model performance?

Accuracy is more important in machine learning models. We can improve model performance by using distributed computing and parallelizing over the scored assets. But accuracy should be built during the model training process.

50) How would you handle an imbalanced dataset?

Imbalanced data set is a classification problem where the number of observations per class is not distributed equally. For some classes, there will be a large number of observations whereas for others fewer observations are present. We can fix this issue by -

- Collecting more data to even the imbalances in the dataset.
- Resample the dataset to correct for imbalances.
- Try a different algorithm altogether on your dataset.

51) When should you use classification over regression?

In supervised learning, we have datasets and a list of outcomes. Types of outcomes that we have helped us categorize the problem into classification and regression. For regression problems, the outcomes are typically in real numbers whereas for classification problems outcomes are classes or categories. Therefore, we can say that we would use

regression if the outputs are in real numbers and we would go with classifications if the outputs are in the form of classes or categories.

52) Tell me a situation where ensemble techniques might be useful?

Ensemble learning combines several models into one predictive model to decrease the variance and improve results. The ensemble method is divided into two groups - the sequential method and the parallel method.

Sequential method - base learners are generated sequentially

Parallel method - base learners are generated parallelly

Ensembles techniques are -

- Bagging
- Stacking
- Boosting

Scenario - suppose you want to buy a new pair of headphones. What will you do? Being an aware consumer, first, you will do research on which company offers the best headphones and also take some suggestions from your friends. In short, you will be making informed decisions after thoroughly researching work.

53) What are the data types supported by JSON?

Here, the interviewer wants to test your knowledge of JSON. There are six basic data types supported by JSON: strings, numbers, objects, arrays, booleans, and null values.

54) According to you, what is the most valuable data in our business?

Through this question, the interviewer tries to test you on two dimensions: your knowledge and understanding about business models and how you correlated data and apply that thinking about the company. To answer this question, you'll have to research the business model, learn their business problems, and solve most with their data.

55) Tell us about machine learning papers you've read lately?

To answer this question, you need to keep yourself updated with the latest scientific literature on machine learning to demonstrate your interest in a machine learning position.

56) What GPU/hardware do you use and what models do you train for?

This question tests if you have handled [machine learning projects](#) outside of a corporate role and understand how to resource projects and allocate GPU time efficiently. These kinds of questions are usually asked by hiring managers as they want to know what you've done independently.

There are some general questions that an interviewer may ask you depending upon your working experiences and awareness. Some of them are as follows -

1. Do you have research experience in machine learning?
2. What use cases do you like the most in machine learning?
3. What are your views on GPT-3 and OpenAI's model?

57) What is a Fourier Transform?

Fourier Transform is a general method for decomposing the general functions into asymmetric functions superposition. The Fourier transform discovers the cycle set amplitudes, phases, and speeds for matching any time signal. Fourier transform converts the signal from the time to frequency domain.

58) What is the use of the Kernel trick?

Kernel trick includes kernel functions that can allow higher-dimension spaces without externally computing the dimension's points coordinates. Kernel functions calculate the inner products among the images of all the data pairs in the feature space. This enables them to attribute computing coordinates of higher dimensions when the computation of the said coordinates' external calculation is cheaper.

59) Why Naive Bayes is Naive?

Regardless of its practical applications, particularly in text mining, we consider Naive Bayes naive because it makes superposition which is practically impossible to see in real-time data. We calculate the conditional probability in the form of the product of the separate probabilities of the components.

60) Define Overfitting? How do we assure that we are not overfitting a model?

Overfitting happens when the model researches the training data to affect the model performance on the latest data significantly. This indicates that we record the disruption in the training data, and we learn the concepts by model. The problem is that concepts that do not employ the testing data negatively affect the ability of the model for classifying the new data; therefore, it decreases the testing data accuracy.

To avoid Overfitting, we have to apply the following methods:

- We collect more data so that we can train the model with diverse samples.
- We can avoid overfitting by using the ensembling methods, like Random Forest. According to the bagging idea, we use them to minimize the change in the projections by joining the result of the multiple decision trees on various samples of the data set.
- By Selecting the correct algorithm, we can avoid overfitting.

61) Explain Cluster sampling?

Cluster sampling is a process of arbitrarily choosing the integral algorithms inside a specified population, and distributing the same characteristics. Cluster sampling is the likelihood sample where a single sampling unit is a cluster or collection of the elements. For instance, if we are clustering the cumulative number of managers in a group of companies, in such a case, managers will depict employees and companies will depict the clusters.

62) What are the methods available to screen the Outliers?

We can use the following methods to screen the outliers:

Linear models: Linear models like logistic regression can be trained to screen the outliers. In this way, the model collects the subsequent outlier it meets.

Boxplot: The box plot depicts the allocation of the data and its changeability. Box plot includes lower and upper quartiles; therefore, the box fundamentally stretches the Inter-Quartile Range(IQR). The main reason for using the box plot is to identify the outliers in the data.

Proximity-based models: K-means clustering is the example of this kind of model, where data points form various or "k" clusters based on the features like distance or similarity.

Probabilistic and Statistical models: We can use statistical models like exponential distribution and normal distribution for identifying the variations in the allocation of the data points. If we found any data point outside the distribution scope, then we can render it an outlier.



<https://www.youtube.com/@codewitharrays>



<https://www.instagram.com/codewitharrays/>



<https://t.me/codewitharrays> Group Link: <https://t.me/ccee2025notes>



[+91 8007592194 +91 9284926333](#)



codewitharrays@gmail.com



<https://codewitharrays.in/project>