100 Machine Learning Interview Questions and Answers

1. Please Explain Machine Learning, Artificial Intelligence, And Deep Learning?

Machine learning is defined as a subset of Artificial Intelligence, and it contains the techniques which enable computers to sort things out from the data and deliver Artificial Intelligence applications. Artificial Intelligence (AI) is a branch of computer science that is mainly focused on building smart machines that can perform certain tasks that mainly require human intelligence. It is the venture to replicate or simulate human intelligence in machines.
Deep learning can be defined as a class of machine learning algorithms in Artificial Intelligence that mainly uses multiple layers to cumulatively extract higher-level features from the given raw input.

2. How Difficult Is Machine Learning?

Machine Learning is huge and comprises a lot of things. Therefore, it will take more than six months to learn Machine Learning if you spend at least 6-7 hours per day. If you have good hands-on mathematical and analytical skills, then six months will be sufficient for you.

3. Can You Explain Kernel Trick In An SVM Algorithm?

A Kernel Trick is a method where the Non-Linear data is projected onto a bigger dimension space in order to make it easy to classify the data where it can be linearly divided by a plane.

4. Can You List Some Of The Popular Cross-Validation Techniques?

1. Holdout Method: This kind of technique works by removing the part of the training data set and sending the same to the model that was trained on the remaining data set to get the required predictions.
2. K-Fold Cross-Validation: Here, the data is divided into k subsets so that every time, one among the k subsets can be used as a validation set, and the other k-1 subsets are used as the training set
3. Stratified K-Fold Cross-Validation: It works on imbalanced data.
4. Leave-P-Out Cross-Validation: Here, we leave p data points out of the training data out of the n data points, then we use the n-p samples to train the model and p points for the validation set.

5. Differences Between The Bagging And Boosting Algorithms?

Bagging	Boosting
It is a method that merges the same type of predictions.	It is a method that merges the different types of predictions.
It decreases the variance, not the bias	It decreases the bias, not the variance.
Each and every model receives equal weight	Models are weighed based on performance.

6. What Are Kernels In SVM? Can You List Some Popular Kernels Used In SVM?

The kernel is basically used to set mathematical functions that are used in the Support Vector Machine by providing the window to manipulate the data. Kernel Function is used to transform the training set of data so that a non-linear decision surface will be transformed to a linear equation in a bigger number of dimension spaces.
Some of the popular kernels used in SVM are:
1. Polynomial kernel
2. Gaussian kernel
3. Gaussian radial basis function (RBF)
4. Laplace RBF kernel
5. Hyperbolic tangent kernel
6. Sigmoid kernel
7. Bessel function of the first kind Kernel
8. ANOVA radial basis kernel

7. Can You Explain The OOB Error?

An out-of-bag error called OBB error, also known as an out-of-bag estimate, is a technique to measure the prediction error of random forests, boosted decision trees. Bagging mainly uses subsampling with replacement to create the training samples for the model to learn from them.

8. Can You Differentiate Between K-Means And KNN Algorithms?

K-Means	KNN algorithms
It is unsupervised machine learning.	It is supervised machine learning.
It is a clustering machine learning algorithm.	It is a classification or regression machine learning algorithm.
Its performance is slow.	It performs much better.
It is an eager learner.	It is a lazy learner.

9. Explain The Term Variance Inflation Factor Mean?

Variance inflation factor known as VIF is a measure of the amount of multicollinearity in the given set of multiple regression variables. The ratio here is calculated for each of the independent variables. A high VIF means that the associated independent variable is mostly collinear with the other variables in the model.

10. Explain SVM (Support Vector Machines) In Machine Learning?

Support Vector Machine, known as SVM, is one of the most commonly used Supervised Learning algorithms that is mainly used for Classification as well as Regression problems. It is primarily used for Classification problems in Machine Learning. The main aim of the SVM algorithm is to create the best decision boundary, which segregates n-dimensional space into classes so that one can easily put the new obtained data point in the correct category in the future.

11. Differentiate Between Supervised And Unsupervised Machine Learning?

Supervised Model	Unsupervised Model
Here, the algorithm learns on a labeled dataset,	Here, it provides unlabeled data.
Here, the models need to find the mapping function that is used to map the input variable (X) with the output variable (Y).	The main aim of unsupervised learning is to find the structure and patterns from the given input data.

12. Explain The Terms Precision And Recall?

 Precision, also known as a positive predictive value, is defined as the fraction of relevant instances among the retrieved instances.
Precision = TP/TP+FP
Where TP is true positive
         FP id False Positive
Recall, also known as sensitivity, is defined as the fraction of relevant instances that were Retrieved.
Recall = TP/TP+FP.
Where TP is true positive
        FP is False positive.

13. Differentiate Between L1 And L2 Regularization?

L1 Regularization	L2 Regularization
A regression model that makes use of the L1 regularization process is called Lasso Regression.	A regression model that makes use of the L1 regularization process is called Ridge Regression.
Lasso Regression adds the absolute value of the magnitude of coefficient as a penalty term to the loss function.	Ridge regression adds the squared magnitude of coefficient as a penalty term to the loss function.
It tries to estimate the median of the data.	It tries to estimate the mean of the data.

14. Explain Fourier Transform?

The Fourier transform is a way to split something up into a bunch of sine waves. In terms of mathematics, The Fourier Transform is a process that can transform a signal into its respective constituent components and frequencies. Fourier transform is used not only in signal, radio, acoustic, etc.

15. What Is The F1 Score? How To Use It?

The F1-score combines both the precision and recall of a classifier into one single metric by taking the harmonic mean. It is used to compare the performances of two classifiers. For example, classifier X has a higher recall, and classifier Y has higher precision. Now the F1-scores calculated for both the classifiers will be used to predict which one produces the better results.
The F1 score can be calculated as
2(P*R)/(P+R)
Where P is the precision.
        R is the Recall of the classification model.
Machine Learning Interview Questions And Answers

16. Differentiate Between Type I And Type II Error?

Type I Error	Type II Error
It is equivalent to a False positive.	It is equivalent to a False negative
It refers to non-acceptance of hypothesis	It refers to the acceptance of the hypothesis
There can be a rejection even with an authorized match.	There can be an acceptance even with an unauthorized match.

17. Can You Explain How A ROC Curve Works?

The ROC curve is represented graphically by plotting the true positive rate (TPR) against the FPR (False Positive rates). Where
1. The true positive rate can be defined as the proportion of observations that are predicted to be positive out of all the given positive observations (TP/(TP + FN))
2. The false-positive rate is defined as the proportion of observations that are predicted wrongly to be positive out of all the given negative observations.
(FP/(TN + FP))

18. Differentiate Between Deep Learning And Machine Learning?

Deep Learning	Machine Learning
It is a subset of Machine Learning	It is a superset of Deep Learning.
It solves complex issues.	It is used to learn new things.
It is an evolution to Machine Learning.	It is an evolution of AI.
Here, algorithms are largely self-depicted on the data analysis	Algorithms are detected by the data analysts.

19. Can You Name The Different Machine Learning Algorithms?

Different machine learning algorithms are listed below:
1. Decision trees,
2. Naive Bayes,
3. Random forest
4. Support vector machine
5. K-nearest neighbor,
6. K-means clustering,
7. Gaussian mixture model,
8. Hidden Markov model etc.

20. What Is AI?

AI (Artificial intelligence) refers to the simulation of human intelligence in machines that are programmed to reflect like humans and imitate their actions.
Examples: Face Detection and Recognition, Google Maps, and
Ride-Hailing Applications, E-Payments.

21. How To Select Important Variables While Working On A Data Set?

1. You have to remove the correlated variables before selecting important variables.
2. Make use of linear regression and select the variables based on their p values.
3. Use Forward Selection, Stepwise Selection, and Backward Selection.
4. Use Random Forest, Xgboost, and plot variable importance chart
5. Use the Lasso Regression
6. You have to select top n features by measuring the information gain for the available set of features.

22. Differentiate Between Causality And Correlation?

The Causality explicitly applies to the cases where action A causes the outcome of action B. Correlation can simply be defined as a relationship. Where the actions of A can relate to the actions of B, but here it is not necessary for one event to cause the other event to happen.

23. What Is Overfitting?

Overfitting is a type of modeling error that results in the failure to predict or guess the future observations effectively or fit additional data in the model that already exists.

24. Explain The Terms Standard Deviation And Variance?

A standard deviation is defined as the number that specifies how spread out the values are. A low standard deviation represents that most of the numbers are close to the mean value. The higher standard deviation means that the values are spread out over, the wider range. Variance in Machine Learning is a type of error that occurs due to the model's sensitivity to small fluctuations in the given training set.

25. Explain Multilayer Perceptron And Boltzmann Machine?

A Multilayer Perceptron (MLP) is defined as a class of artificial neural networks that can generate a set of outputs from the set of given inputs. An MLP consists of several layers of input nodes that are connected as a directed graph between input and output layers.
The main purpose of the Boltzmann Machine is to optimize the solution to a given problem. It is mainly used to optimize the weights and quantity related to that specified problem.
Machine Learning Interview Questions And Answers

26. Explain The Term Bias?

Data bias in machine learning is defined as a type of error where certain elements of a given dataset are weighted more heavily than others. A biased dataset will not accurately represent the model's use case, and it results in low accuracy levels and analytical errors.

27. Name The Types Of Machine Learning?

The types of machine learning are listed below:
1. Supervised Learning
2. Unsupervised Learning
3. Reinforcement Learning

28. Differentiate Between Classification And Regression?

Classification	Regression
It is about predicting a label	It is about predicting a quantity
Here, the data is labeled in one or multiple classes.	Here, you need to predict the quantity continuously.
It may predict a continuous value.	It may predict a discrete value.
It can be evaluated using accuracy.	It can be evaluated using root mean squared error.

29. What Is A Confusion Matrix?

In the field of machine learning, a confusion matrix also called an error matrix, is defined as a specific table layout that allows the user to visualize the performance of an algorithm, mainly a supervised learning one.

30. When Your Dataset Is Suffering From High Variance, How Would You Handle It?

For datasets with high variance, we can make use of the bagging algorithm. The bagging algorithm splits the data into different subgroups with sampling replicated from random data. Once the data is split, using a training algorithm, the random data can be used to create rules. Then we make use of the polling technique to gather all the predicted outcomes of the model.

31. Differentiate Between Inductive And Deductive Learning?

Inductive Learning	Deductive Learning

It aims at developing a theory.	It aims at testing an existing theory.
It moves from the specific observations to the broad generalizations	If there is no theory, you cannot conduct deductive research.
It consists of three stages,ObservationObserve a patternDevelop a theory	It consists of four stages:Start with an existing theoryFormulate a hypothesis based on existing theoryCollect data to test the hypothesisAnalyze the results

32. Explain The Handling Of Corrupted Values In The Given Dataset?

The below are the ways to handle missing data?
1. Remove the rows with missing values.
2. Build another predictive model so that you can predict the missing values.
3. Use a model in such a way that it can incorporate missing data.
4. You need to replace the missing data with the aggregated values.
5. You can predict the missing values.
6. create an unknown category

33. Which Among These Is More Important Model Accuracy Or Model Performance?

Model accuracy is considered as the important characteristic of a Machine Language /AI model. Whenever we discuss the performance of the model, we first clarify whether it is the model scoring performance or Model training performance.
Model performance is improved by using distributed computing and parallelizing over the given scored assets, but we need to carefully build the accuracy during the model training process.

34. What Is A Time Series?

The time series in Machine learning is defined as a set of random variables that are ordered with respect to time. Time series are studied to interpret a phenomenon, identify the components of a trend, cyclicity, and predict its future values.

35. Differentiate Between Entropy And Information Gain?

The Information Gain is defined as the amount of information gained about a signal or random variable from observing another random variable.
Entropy can be defined as the average rate at which information is produced by the stochastic source of data, Or it can be defined as a measure of the uncertainty that is associated with a random variable.

36. Differentiate Between Stochastic Gradient Descent (SGD) And Gradient Descent (GD)?

Batch Gradient Descent is involved in calculations over the full training set of each step, which results in a very slow process on very large training data. Hence, it becomes very expensive to do Batch GD. However, It is great for relatively smooth error manifolds. Also, it scales well with the number of features.
Stochastic Gradient Descent tries to solve the primary problem in Batch Gradient descent that is the usage of entire training data to calculate the gradients as each step. SGD is stochastic in nature means it picks up some "random" instances of training data at each and every step, and then it computes the gradient making it faster as there are very little data to manipulate at one shot,

Batch Gradient Descent	Stochastic Gradient Descent
It computes the gradient using the entire Training sample.	It computes gradient using a single Training sample.
It can't be suggested for huge training samples.	It can be suggested for large training samples.
It is deterministic in nature.	It is sophisticated in nature.

37. Differentiate Between Gini Impurity And Entropy In A Decision Tree?

Gini	Entropy
It has values inside the interval [0, 0.5]	It has values inside the interval [0, 1]
It is more complex.	It is not complex.
Its measurement is the probability of a random sample that is being classified correctly.	It is a measurement to calculate the lack of information,

38. Mention Some Of The Advantages And Disadvantages Of Decision Trees?

Advantages of the decision tree:
1. Decision trees require less effort for data preparation during the pre-processing when compared with other algorithms.
2. A decision tree doesn't require the normalization of data.
3. It does not require scaling of data.
4. Missing values in the data do not affect the process of building a decision tree.
5. A Decision tree model is very easy to explain to technical teams and stakeholders.

39. Can You Explain The Ensemble Learning Technique In Machine Learning?

Ensemble methods are the techniques used to create multiple models and combine them to produce enhanced results. Ensemble methods usually produce more precise solutions than a single model would.
In Ensemble Learning, we divide the training data set into multiple subsets, where each subset is then used to build a separate model. Once the models are trained, they are then combined to predict an outcome in such a way that there is a reduction in the variance of the output.
Machine Learning Interview Questions And Answers

40. Explain The Terms Collinearity And Multicollinearity?

Multicollinearity occurs when multiple independent variables are highly correlated with each other in a regression model, which means that an independent variable can be predicted from another independent variable inside a regression model.
Collinearity mainly occurs when two predictor variables in a multiple regression have some correlation.

41. Differentiate Between Random Forest And Gradient Boosting Machines?

Like random forests, gradient boosting is also a set of decision trees. The two primary differences are:
1. How trees are built: Each tree in the random forest is built independently, whereas gradient boosting builds only one tree at a time.
2. Combining results: random forests combine results at the end of the process by averaging. Whereas gradient boosting combines results along the path.

42. Explain The Terms Eigenvectors And Eigenvalues?

Eigenvectors are unit vectors, meaning their length or magnitude is equal to 1.0. They are referred to as right vectors, which means a column vector.
Eigenvalues are coefficients that are applied to eigenvectors that, in turn, give the vectors their length or magnitude.

43. Can You Explain Associative Rule Mining (ARM)?

Association rule mining (ARM) aims to find out the association rules that will satisfy the predefined minimum support and confidence from a database. AMO is mainly used to reduce the number of association rules with the new fitness functions that can incorporate frequent rules.

44. What Is A/B Testing?

A/B testing is defined as a basic randomized control experiment. It is used to compare two versions of a variable to find out which one among them performs better in a controlled environment.
A/B Testing can be best used to compare two models to check which one is the best-recommended product to a customer.

45. Explain Marginalisation And Its Process?

Marginalization is a method that requires the summing of the possible values of one variable to determine the marginal contribution of another variable.
$P(X=x) = \sum Y P(X=x,Y)$

46. What Is Cluster Sampling?

Cluster sampling is defined as a type of sampling method. With cluster sampling, the researchers usually divide the population into separate groups or sets, known as clusters. Then, a random sample of clusters is picked from the population. Then the researcher conducts their analysis on the data from the collected sampled clusters.

47. Explain The Term"Curse Of Dimensionality"?

The curse of dimensionality basically refers to the increase in the error with the increase in the number of features. It can be referred to the fact that algorithms are vigorous to design in high dimensions, and they often have a running time exponential in the dimensions.

48. Can You Name A Few Libraries In Python Used For Data Analysis And Scientific Computations?

1. NumPy
2. SciPy
3. Pandas
4. SciKit
5. Matplotlib
6. Seaborn
7. Bokeh

49. What Are Outliers? Mention The Methods To Deal With Outliers?

An outlier can be defined as an object that deviates significantly from other objects. They can be caused by execution errors.
The three main methods to deal with outliers are as follows:
1. Univariate method

2. Multivariate method
3. Minkowski error

50. List Some Popular Distribution Curves Along With Scenarios Where You Will Use Them In An Algorithm?

The most popular distribution curves are:
Uniform distribution can be defined as a probability distribution that has a constant probability. Example: Rolling a single dice since it has multiple outcomes.
The binomial distribution is defined as a probability with two possible outcomes only. Example: a coin toss. The result will either be heads or tails.
Normal distribution specifies how the values of a variable are distributed. Example: The height of students in a classroom.
Poisson distribution helps to predict the probability of specific events that are happening when you know how often that event has occurred.
The exponential distribution is mainly concerned with the amount of time until the specific event occurs. Example: how long a car battery could last, in months.

51. Can You List The Assumptions For Data To Be Met Before Starting With Linear Regression?

The assumptions to be met are:
1. Linear relationship
2. Multivariate normality
3. No or little multicollinearity
4. No auto-correlation
5. Homoscedasticity

52. Explain The Term Variance Inflation Factor Mean?

Variance inflation factor that is VIF is defined as a measure of the amount of multicollinearity in a given set of multiple regression variables.
Mathematically, the Variance inflation factor for a regression model variable is equal to the ratio of the final model variance to the variance of a model that comprises that single independent variable.
This ratio is calculated for each of the independent variables. A high VIF represents that the associated independent variable is hugely collinear with the other variables in the model.

53. Can You Tell Us When The Linear Regression Line Stops Rotating Or Finds An Optimal Spot Where It Is Fitted On Data?

The place where the highest RSquared value is found is where the line comes to rest. RSquared usually represents the amount of variance that is captured by the virtual linear regression line w.r.t the total variance captured by the dataset.

54. Can You Tell Us Which Machine Learning Algorithm Is Known As The Lazy Learner And Why It Is Called So?

KNN Machine Learning algorithm is called a lazy learner. K-NN is defined as a lazy learner because it will not learn any machine-learned values or variables from the given training data, but dynamically it calculates the distance every time it wants to classify. Hence it memorizes the training dataset instead.

55. Can You Tell Us What Could Be The Problem When The Beta Value For A Specific Variable Varies Too Much In Each Subset When Regression Is Run On Various Subsets Of The Dataset?

The variations in the beta values in every subset suggest that the dataset is heterogeneous. To overcome this problem, we use a different model for each of the clustered subsets of the given dataset, or we use a non-parametric model like decision trees.

56. How To Choose A Classifier Based On A Training Set Data Size?

If the training set is small in size, high bias or low variance models, for example, Naive Bayes tends to perform better as they are less likely to overfit.
If the training set is large in size, low bias or high variance models, for example, Logistic Regression, tend to perform better as they can reflect more complicated relationships.

57. Differentiate Between Training Set And Test Set In A Machine Learning Model?

Training set	Test set
70% of the total data is taken as the training dataset.	The remaining 30% is taken as a testing dataset.
It is implemented to build up a model.	It is used to validate the model built.
It is a labeled data used to train the model.	We usually test without labeled data and then verify the results with labels.

58. Explain A False Positive And False Negative And How Are They Significant?

A false positive is a concept where you receive a positive result for a given test when you should have actually received a negative result. It's also called a "false alarm" or "false positive error." It is basically used in the medical field, but it can also apply to software testing.
 Examples of False positive:

1. A pregnancy test is positive, where in fact, you are not pregnant.
2. A cancer screening test is positive, but you do not have the disease.
3. Prenatal tests are positive for Down's Syndrome when your fetus does not have any disorder.
4. Virus software on your system incorrectly identifies a harmless program as the malicious one.

A false negative is defined where a negative test result is wrong. In simple words, you get a negative test result, where you should have got a positive test result.

For example, consider taking a pregnancy test, and you test as negative (not pregnant). But in fact, you are pregnant.

The false negative pregnancy test results due to taking the test too early, using the diluted urine, or checking the results very soon. Just about every medical test has the risk of a false negative.

59. Explain The Term Semi-Supervised Machine Learning?

Semi-supervised learning is defined as an approach to machine learning that combines a less amount of labeled data with a huge amount of unlabeled data during the training process. It falls between unsupervised learning and supervised learning.

60. Can You Tell Us The Applications Of Supervised Machine Learning In Modern Businesses?

1. Healthcare Diagnosis
2. Fraud detection
3. Email spam detection
4. Sentimental analysis

61. Can You Differentiate Between Inductive Machine Learning And Deductive Machine Learning?

Inductive Machine Learning	Deductive Machine Learning
$A \wedge B \vdash A \rightarrow B$ (Induction)	$A \wedge (A \rightarrow B) \vdash B$ (Deduction)
It observes and learns from the set of instances, and then it draws the conclusion.	It derives the conclusion first, and then it works on it based on the previous decision.
It is a Statistical machine learning like KNN or SVM,	Machine learning algorithm to deductive reasoning using the decision tree.

62. What Is Random Forest In Machine Learning?

The random forest can be defined as a supervised learning algorithm that is used for classifications and regression. Similarly, the random forest algorithm creates decision trees on the data samples, and then it gets the prediction from each of the samples and finally selects the best one by means of voting.

63. Explain The Trade-Off Between Bias And Variance?

Bias can be defined as the assumptions made by the model to make the target function easy to approximate.
Variance is defined as the amount that the estimate of the target function will change given the different training data.
The trade-off is defined as the tension between the error introduced by bias and variance.

64. Explain Pruning In Decision Trees, And How Is It Done?

Pruning is a data compression process in machine learning and search algorithms that can reduce the size of the decision trees by removing certain sections of the tree that are non-critical and unnecessary to classify instances. A tree that is too huge risks overfitting the training data and is poorly generalizing to the new samples.
Pruning can take place as follows.
1. Top-down fashion (It will travel the nodes and trim subtrees starting at the root)
2. Bottom-up fashion (It will start at the leaf nodes)
We have reduced the error algorithm for the pruning of decision trees.

65. How Reduced Error Algorithms Work For Pruning In Decision Trees?

The reduced error algorithm works as follows:
1. It considers each node for pruning.
2. Pruning = removing the subtree at that node, then make it a leaf and assign the major common class at that node.
3. A node is removed from the tree if the resulting tree performs worse than the original.
4. Nodes are removed iteratively by choosing the node in such a way that whose removal mostly increases the accuracy of the decision tree on the graph.
5. Pruning continues to perform until further pruning is harmful.
6. It uses training, test sets, and validations. It is an effective approach if a vast amount of data is available.

66. Explain The Term Decision Tree Classification?

A decision tree builds classification models as a tree structure, with datasets broken up into smaller subsets while developing the decision tree; basically, it is a tree-like way with branches and nodes defined. Decision trees handle both categorical and numerical data.

67. Explain Logistic Regression?

Logistic regression analysis is a technique used to examine the association of independent variables with one dichotomous dependent variable. This is in contrast to the linear regression analysis, where the dependent variable is a continuous variable.
Every time the output of logistic regression is 0 or 1 with a threshold value of 0.5. Any value above 0.5 is taken as 1, and any point below 0.5 is taken as 0.

68. Name Some Methods Of Reducing Dimensionality?

Some of the methods of reducing dimensionality are given below:
1. By combining features with feature engineering
2. Removing collinear features
3. using algorithmic dimensionality reduction.

69. What Is A Recommendation System?

Recommendation systems mainly collect the customer data and auto analyze this data to generate the customized recommendations for the customers. These systems mainly rely on implicit data like browsing history and recent purchases and explicit data like ratings provided by the customer.

70. Explain The K Nearest Neighbor Algorithm?

K-Nearest Neighbour is the simplest Machine Learning algorithm that is based on the Supervised Learning technique. It assumes the similarity between the new case or data and the available cases, and it puts the new case into a category that is similar to that of the available categories.
For example, we have an image of a creature that looks similar to that of a cat and a dog, but we want to know whether it is a cat or a dog. For this identification, we can make use of the KNN algorithm, as it works on a similarity basis. The KNN model will find the similarities of the new data set to that of the cats and dogs images, and that is based on the similar features; it will put it in either a cat or a dog category.

71. Considering A Given Long List Of Machine Learning Algorithms, Given A Data Set, How Do The Spam Filters Of The Email Will Be Fed With Hundreds Of Emails You Decide Which One To Use?

Choosing an algorithm depends on the below-mentioned questions:
1. How much data you have, and is that continuous or categorical?
2. Is the problem related to classification, clustering, association, or regression?

3. Is it a Predefined variable (labeled), unlabeled, or a mix of both?
4. What is the primary purpose?

Based on the above questions, one has to choose the right algorithm that suits their requirement.

72. Can You Tell Us How To Design An Email Spam Filter?

1. The spam filter of the email will be fed with hundreds of emails.
2. Each of these emails has a label: 'spam' or 'not spam.'
3. The supervised machine learning algorithm will then identify which type of emails are being marked as spam based on spam keywords like the lottery, no money, full refund, etc.
4. The next time an email hits the inbox, the spam filter will use statistical analysis and algorithms like Decision Trees and SVM to identify how likely the email is spam.
5. If the probability is high, then it will be labeled as spam, and the email will not hit your inbox.
6. Based on the accuracy of each of the models, we use the algorithm with the highest reliability after testing all the given models.

73. How Can You Avoid Overfitting?

Overfitting is avoided by following the steps:
1. Cross-validation: The idea here is to use the initial training data to generate various small train test spills. Where these test spills are used to tune the model.
2. Train with more data: Training with a lot of data can help the algorithms to detect the signals better.
3. Remove feature: You can manually remove some of the features.
4. Early stopping: It refers to stopping the training process before the learner passes the specified point.
5. Regularization: It refers to a broad range of techniques for artificially forcing the model to be simple.
6. Ensembling: These are machine learning algorithms that combine predictions from multiple separate models.

74. Explain The Term Selection Bias In Machine Learning?

Selection bias takes place if a data set's examples are chosen in such a way that it is not reflective of their real-world distribution. Selection bias can take many various forms.
1. Coverage bias: Data here is not selected in a representative manner.

Example: A model is trained in such a way to predict the future sales of a new product based on the phone surveys conducted with the sample of customers who bought the product. Consumers who instead opted for buying a competing product were not surveyed, and as a result, this set of people were not represented in the training data.

2. Non-response bias: Data here ends up being unrepresentative due to the participation gaps in the collection of data processes.

Example: A model is trained in such a way to predict the future sales of a new product based on the phone surveys conducted with a sample of customers who bought the product and with a sample of customers who bought the competing product. Customers who bought the competing product were 80% more expected to refuse to complete the survey, and their data were underrepresented in the sample.

3. Sampling bias: Here, proper randomization is not used during the data collection process.

Example: A model that is trained to predict the future sales of a new product based on the phone surveys conducted with a sample of customers who bought the product and with a sample of customers who bought a competing product. Instead of randomly targeting customers, the surveyor chose the first 200 consumers that responded to their email, who might have been more eager about the product than the average purchasers.

75. Explain The Types Of Supervised Learning?

Supervised learning is of two types, namely,

1. Regression: It is a kind of Supervised Learning that learns from the given Labelled Datasets, and then it is able to predict the continuous-valued output for the new data that is given to the algorithm. It is used in cases where an output requirement is a number like money or height etc. Some popular Supervised Learning algorithms are Linear Regression, Logistic Regression.

2. Classification: It is a kind of learning where the algorithm needs to be mapped to the new data that is obtained from any one of the two classes that we have in the dataset. The classes have to be mapped to either 1 or 0, which in real-life translates to the 'Yes' or 'No.' The output will have to be either one of the classes, and it should not be a number as it was in the case of Regression. Some of the most well-known algorithms are Decision trees, Naive Bayes Classifier, Support vector Algorithms.

76. What Vanishing Gradient Descent?

In Machine Learning, we encounter the Vanishing Gradient Problem while training the Neural Networks with gradient-based methods like Back Propagation. This problem makes it hard to tune and learn the parameters of the earlier layers in the given network.
The vanishing gradients problem can be taken as one example of the unstable behavior that we may encounter when training the deep neural network.

It describes a situation where the deep multilayer feed-forward network or the recurrent neural network is not able to propagate the useful gradient information from the given output end of the model back to the layers close to the input end of the model.

77. Can You Name The Proposed Methods To Overcome The Vanishing Gradient Problem?

The methods proposed to overcome the vanishing gradient problems are:
1. Multi-level hierarchy
2. The long short – term memory
3. Faster hardware
4. Residual neural networks (ResNets)
5. ReLU

78. Differentiate Between Data Mining And Machine Learning?

Data Mining	Machine Learning
It extracts useful information from a large amount of data.	It introduces algorithms from data as well as from past experience.
It is used to understand the flow of data.	It teaches the computers to learn and understand from the data flow.
It has huge databases with unstructured data.	It has existing data as well as algorithms.
It requires human interference in it.	No need for the human effort required after design
Models are developed using data mining technique	machine-learning algorithm can be used in the decision tree, neural networks, and some other parts of artificial intelligence
It is more of research using methods like machine learning.	It is self-learned and trains the system to do intelligent tasks.

79. Name The Different Algorithm Techniques In Machine Learning?

The different algorithm techniques in machines learning are listed below:
1. Unsupervised Learning
2. Semi-supervised Learning
3. Transduction

4. Reinforcement Learning
5. Learning to Learn
6. Supervised Learning

80. Explain The Function Of 'Unsupervised Learning?

1. It has to find clusters of the data.
2. Find the low-dimensional representations of the data
3. To find interesting directions in data
4. To calculate interesting coordinates and correlations.
5. Find novel observations or database cleaning.

81. Explain The Term Classifier In Machine Learning?

A classifier in machine learning is defined as an algorithm that automatically categorizes the data into one or more of a group of "classes." One of the common examples is an email classifier that can scan the emails to filter them by the given class labels: Spam or Not Spam. We have five types of classification algorithms, namely,

1. Decision Tree
2. Naive Bayes Classifier
3. K-Nearest Neighbors
4. Support Vector Machines
5. Artificial Neural Networks

82. What Are Genetic Algorithms ?

Genetic algorithms are defined as stochastic search algorithms which can act on a population of possible solutions. Genetic algorithms are mainly used in artificial intelligence to search a space of potential solutions to find one who can solve the problem.

83. Can You Name The Area Where Pattern Recognition Can Be Used?

1. Speech Recognition
2. Statistics
3. Informal Retrieval
4. Bioinformatics
5. Data Mining
6. Computer Vision

84. Explain The Term Perceptron In Machine Learning?

A Perceptron is defined as an algorithm for supervised learning of binary classifiers. This algorithm enables the neurons to learn and processes the elements in the given training set one at a time. There are two types of Perceptrons, namely.
1. Single-layer
2. Multilayer.

85. What Is Isotonic Regression?

Isotonic regression is used iteratively to fit ideal distances to protect the relative dissimilarity order. Isotonic regression is also used in the probabilistic classification to balance the predicted probabilities of the supervised machine learning models.

86. What Are Bayesian Networks?

A Bayesian network can be defined as a probabilistic graphical model that presents a set of variables and their conditional dependencies through a DAG (directed acyclic graph).
For example, a Bayesian network would represent the probabilistic relationships between the diseases and their symptoms. Given the specific symptoms, the network can be used to compute the possibilities of the presence of different diseases.

87. Can You Explain The Two Components Of The Bayesian Logic Program?

The bayesian logic program mainly comprises two components.
1. The first component is the logical one: it comprises a set of Bayesian Clauses that captures the qualitative structure of the domain.
2. The second component is quantitative: it encodes the quantitative information about the domain.

88. What Is An Incremental Learning Algorithm In An Ensemble?

The incremental learning method is defined as the ability of an algorithm to learn from new data that is available after the classifier has already been generated from the already available dataset.

89. Name The Components Of Relational Evaluation Techniques?

The components of the relational evaluation technique are listed below:
1. Data Acquisition
2. Ground Truth Acquisition
3. Cross-Validation Technique

4. Query Type
5. Scoring Metric
6. Significance Test

90. Can You Explain The Bias-Variance Decomposition Of Classification Error In The Ensemble Method?

The expected error of the learning algorithm can be divided into bias and variance. A bias term is a measure of how closely the average classifier produced by the learning algorithm matches with the target function. The variance term is a measure of how much the learning algorithm's prediction fluctuates for various training sets.

91. Name The Different Methods For Sequential Supervised Learning?

The different methods for sequential supervised learning are given below:
1. Recurrent sliding windows
2. Hidden Markow models
3. Maximum entropy Markow models
4. Conditional random fields
5. Graph transformer networks
6. Sliding-window methods

92. What Is Batch Statistical Learning?

A training dataset is divided into one or more batches. When all the training samples are used in the creation of one batch, then that learning algorithm is known as batch gradient descent. When the given batch is the size of one sample, then the learning algorithm is called stochastic gradient descent.

93. Can You Name The Areas In Robotics And Information Processing Where Sequential Prediction Problem Arises?

The areas in robotics and information processing where sequential prediction problem arises are given below
1. Structured prediction
2. Model-based reinforcement learning
3. Imitation Learning

94. Name The Different Categories You Can Categorize The Sequence Learning Process?

The different categories where you can categorize the sequence learning process are listed below:

1. Sequence generation
2. Sequence recognition
3. Sequential decision
4. Sequence prediction

95. What Is Sequence Prediction?

Sequence prediction aims to predict elements of the sequence on the basis of the preceding elements.
A prediction model is trained with the set of training sequences. On training, the model is used to perform sequence predictions. A prediction comprises predicting the next items of a sequence. This task has a number of applications like web page prefetching, weather forecasting, consumer product recommendation, and stock market prediction.
Examples of sequence prediction problems include:

1. Weather Forecasting. Given a sequence of observations about the particular weather over a period of time, it predicts the expected tomorrow's weather.
2. Stock Market Prediction. Given a sequence of movements of the security over a period of time, it predicts the next movement of the security.
3. Product Recommendation. Given a sequence of the last purchases of a customer, it predicts the next purchase of a customer.

96. Explain PAC Learning?

Probably approximately correct, i.e., PAC learning is defined as a theoretical framework used for analyzing the generalization error of the learning algorithm in terms of its error on a given training set and some measures of the complexity. The main goal here is to typically show that an algorithm can achieve low generalization error with high probability.

97. What Are PCA, KPCA, And ICA, And What Are They Used For?

Principal Components Analysis(PCA): It linearly transforms the original inputs into the new uncorrelated features.
Kernel-based Principal Component Analysis(KCPA): It is a nonlinear PCA developed by using the kernel method.
Independent Component Analysis(ICA): In ICA, the original inputs are linearly transformed into certain features that are mutually statistically independent.

98. Explain The Three Stages Of Building A Model In Machine Learning?

The three stages are:

1. Model Building
2. Model Testing
3. Applying the model

99. Explain The Term Hypothesis In ML?

Machine Learning, especially supervised learning, can be specified as the desire to use the available data to learn a function that best maps the inputs to outputs.
Technically, this problem is called function approximation, where we are approximating an unknown target function that we assume as it exists that can best map the given inputs to outputs on all possible considerations from the problem domain.
An example of the model that approximates the target function and performs the mappings of inputs to the outputs is known as the hypothesis in machine learning.
The choice of algorithm and the configuration of the algorithm define the space of possible hypotheses that the model may constitute.

100. Explain The Terms Eepoch, Eentropy, Bbias, And Vvariance In Machine Learning?

Epoch is a term widely used in machine learning that indicates the number of passes of the whole training dataset that the machine learning algorithm has completed. If the batch size is the entire training dataset, then the number of epochs is defined as the number of iterations.
Entropy in Machine learning can be defined as the measure of disorder or uncertainty. The main goal of machine learning models and Data Scientists, in general, is to decrease uncertainty.
Data bias is a type of error in which certain elements of a dataset are more heavily weighted than others.
Variance is defined as the amount that the estimate of the target function will change if a different training data set was used. The target function is usually estimated from the training data by the machine learning algorithm.