Explore More

Subcription : Premium CDAC NOTES & MATERIAL @99

Contact to Join
Premium Group

Click to Join
Telegram Group

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=llOS-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/IynIouBZvY4?si=CXzQs3BsRkjKhZCw
9	Municipal Corporation system Project	https://youtu.be/cVMx9NVyI4I?si=qX0oQt-GT-LR_5jF
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=HURvAY2VnizIyJlh
15	Employee management system project	https://youtu.be/lD1iE3W_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=fiDvrhhrjb4KSlSm
18	Online Pizza Delivery system project	https://youtu.be/Tp3izreZ458?si=8eWAOzA8SVdNwlyM
19	Online Crime Reporting system Project	https://youtu.be/0UlzReSk9tQ?si=6vN0e70TVY1GOwPO
20	Online Children Adoption Project	https://youtu.be/3T5HC2HKyT4?si=bntP78niYH802I7N

1 . What is Machine learning?

Machine learning is the form of Artificial Intelligence that deals with system programming and automates data analysis to enable computers to learn and act through experiences without being explicitly programmed.

EX : Robots are coded in such a way that they can perform the tasks based on data they collect from sensors. They automatically learn programs from data and improve with experiences.

2 . What Are the Different Types of Machine Learning?

Machine Learning algorithms can be primarily classified depending on the presence/absence of target variables.

Supervised learning : [Target is present]
* The machine learns using labelled data. The model is trained on an existing data set before it starts making decisions with the new data.
* The target variable is continuous: Linear Regression, polynomial Regression, quadratic Regression.
* The target variable is categorical: Logistic regression, Naive Bayes, KNN, SVM, Decision Tree, Gradient Boosting, ADA boosting, Bagging, Random forest etc.

Unsupervised learning : [Target is absent]
* The machine is trained on unlabelled data and without any proper guidance. It automatically infers patterns and relationships in the data by creating clusters. The model * learns through observations and deduced structures in the data.
* Principal component Analysis, Factor analysis, Singular Value Decomposition etc.

Reinforcement Learning :
* The model learns through a trial and error method. This kind of learning involves an agent that will interact with the environment to create actions and then discover errors or rewards of that action.

3 . What are the different types of data used in Machine Learning?

Machine Learning Are Used Two Types of Datas.

* Structured Data
* Unstructured Data.

Structured Data : This type of data is predefined, labeled, and well-formatted before being stored in a data storage. Example: Student Records Table.
Unstructured Data : This Type of data is in native format, and it's not processed until it is used. Example: Text, Audio, Video, Emails, etc.

4 . What do you mean by Reinforcement Learning?

Reinforcement learning is an area of **machine learning** in which the model is trained according to the rewards given to it based on its previous actions in the environment. There is an agent whose task is to give rewards and also to maximize the rewards. If the model performs the task correctly, it gets a +1 reward, but if it does a task wrong, then it gets a -1 reward.

Applications :

* Self-driven cars
* Automatic parking
* puzzle solver, etc,.

5 . What are some of the most commonly used Machine Learning algorithms?

Some of the popular Machine Learning algorithms are :

* SVM
* KNN
* K-Means
* Naive Bayes
* Random Forest
* Linear Regression
* Gradient Boosting algorithms
* Logistic RegressionDecision Tree
* Dimensionality Reduction Algorithms

6 . What is the difference between Data Mining and Machine learning?

Machine Learning is about the study, design, and development of the algorithms that make computers work without being explicitly programmed.

Data Mining is a process wherein the unstructured data tries to extract knowledge or unknown interesting patterns, using Machine Learning algorithms.

7 . What do you understand by ensemble learning?

Ensemble learning is a machine learning technique that uses various base models such as classifiers or experts to produce an optimal predictive model. To solve any computational program, such models are strategically generated and combined. The ensemble is a supervised learning algorithm, as it can be trained and used to make predictions.

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

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 . Explain the difference between Regression and Classification?

Regression : regression is a process of finding the correlation between the dependent and independent variables. It is helpful in the prediction of continuous variables, such as in the prediction of the stock market, house prices, etc. In regression, our task is to find the best suitable line that can predict the output accurately.

Classification : Classification is the process of finding a function that helps in dividing the data into different classes. These are mainly used in discrete data. In classification, our aim is to find the decision boundary which can divide the dataset into different classes.

10 . How do you select important variables while working on a data set?

There are various means to select important variables from a data set that include the following :

* Lasso Regression
* Random Forest and plot variable chart
* Forward, Backward, and Stepwise selection
* The variables could be selected based on 'p' values from Linear Regression
* Identify and discard correlated variables before finalizing on important variables
* Top features can be selected based on information gain for the available set of features.

11 . State the differences between causality and correlation?

Causality applies to situations where one action, say X, causes an outcome, say Y, whereas Correlation is just relating one action (X) to another action(Y) but X does not necessarily cause Y.

12 . What is Supervised Learning?

Supervised learning is a machine learning algorithm of inferring a function from labeled training data. The training data consists of a set of training examples.

Example :

Knowing the height and weight identifying the gender of the person. Below are the popular supervised learning algorithms.

* Support Vector Machines
* Regression
* Naive Bayes
* Decision Trees

* K-nearest Neighbour Algorithm and Neural Networks.

Example :

If you build a T-shirt classifier, the labels will be “this is an S, this is an M and this is L”, based on showing the classifier examples of S, M, and L.

13 . What is Unsupervised Learning?

Unsupervised learning is also a type of machine learning algorithm used to find patterns on the set of data given. In this, we don’t have any dependent variable or label to predict. Unsupervised Learning Algorithms :

* Clustering,
* Anomaly Detection,
* Neural Networks and Latent Variable Models.

Example :

In the same example, a T-shirt clustering will categorize as “collar style and V neck style”, “crew neck style” and “sleeve types”.

14 . When does regularization come into play in Machine Learning?

At times when the model begins to underfit or overfit, regularization becomes necessary. It is a regression that diverts or regularizes the coefficient estimates towards zero. It reduces flexibility and discourages learning in a model to avoid the risk of overfitting. The model complexity is reduced and it becomes better at predicting.

15 . How can we relate standard deviation and variance?

Standard deviation refers to the spread of your data from the mean. Variance is the average degree to which each point differs from the mean i.e. the average of all data points. We can relate Standard deviation and Variance because it is the square root of Variance.

16 . What is the meaning of Overfitting in Machine learning?

Overfitting can be seen in machine learning when a statistical model describes random error or noise instead of the underlying relationship. Overfitting is usually observed when a model is excessively complex. It happens because of having too many parameters concerning the number of training data types. The model displays poor performance, which has been overfitted.

17 . What is the method to avoid overfitting?

Overfitting occurs when we have a small dataset, and a model is trying to learn from it. By using a large amount of data, overfitting can be avoided. But if we have a small database and are forced to build a model based on that, then we can use a technique known as cross-validation. In this method, a model is usually given a dataset of a known data on which training data set is run and dataset of unknown data against which the model is tested. The primary

aim of cross-validation is to define a dataset to "test" the model in the training phase. If there is sufficient data, 'Isotonic Regression' is used to prevent overfitting.

18 . What is the trade-off between bias and variance?

Both bias and variance are errors. Bias is an error due to erroneous or overly simplistic assumptions in the learning algorithm. It can lead to the model under-fitting the data, making it hard to have high predictive accuracy and generalize the knowledge from the training set to the test set.

Variance is an error due to too much complexity in the learning algorithm. It leads to the algorithm being highly sensitive to high degrees of variation in the training data, which can lead the model to overfit the data.

To optimally reduce the number of errors, we will need to tradeoff bias and variance.

19 . What is 'Training set' and 'Test set'?

In various areas of information science like machine learning, a set of data is used to discover the potentially predictive relationship known as 'Training Set'. Training set is an examples given to the learner, while Test set is used to test the accuracy of the hypotheses generated by the learner, and it is the set of example held back from the learner. Training set are distinct from Test set.

20 . What is Genetic Programming?

Genetic programming is one of the two techniques used in machine learning. The model is based on the testing and selecting the best choice among a set of results.

21 . What is Inductive Logic Programming in Machine Learning?

Inductive Logic Programming (ILP) is a subfield of machine learning which uses logical programming representing background knowledge and examples.

22 . Explain false negative, false positive, true negative and true positive with a simple example.

Let's consider a scenario of a fire emergency :

True Positive : If the alarm goes on in case of a fire.
Fire is positive and prediction made by the system is true.

False Positive : If the alarm goes on, and there is no fire.
System predicted fire to be positive which is a wrong prediction, hence the prediction is false.

False Negative : If the alarm does not ring but there was a fire.
System predicted fire to be negative which was false since there was fire.

True Negative : If the alarm does not ring and there was no fire.
The fire is negative and this prediction was true.

23 . 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.

24 . 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.

25 . Differentiate Precision, Recall, Accuracy, and the F1 Score?

Precision is the ratio of correctly predicted positive observation and total predicted positive observation. It shows how precise our model is.

*** Precision = TP/TP+FP**

Recall is the ratio of the correct predicted positive observation and the total observation in the class.

*** Recall = TP/TP+FN**

F1-Score is the weighted average of recall and precision.

*** F1-Score = 2*(Recall * Precision) / (Recall + Precision)**

Accuracy is the ratio of correctly predicted positive observations to the total positive observations.

*** Accuracy = TP+TN/TP+TN+FP+FN**

26 . Can you use machine learning for time series analysis?

Yes, it can be used but it depends on the applications. The predictive models based on machine learning have wide applicability across time series projects. These models help in facilitating the predictive distribution of time and resources. The most widely applied machine learning methods for time series forecasting projects are :

* Multi-Layer Perceptron (MLP)

* Recurrent Neural Network (RNN)
* Long Short-Term Memory (LSTM)

27 . What are the full forms of PCA, KPCA, and ICA, and what is their use?

PCA : Principal Components Analysis

KPCA : Kernel-based Principal Component Analysis

ICA : Independent Component Analysis

These are important feature extraction techniques, which are majorly used for dimensionality reduction.

28 . What is a ROC curve?

It is a Receiver Operating Characteristic curve, a fundamental tool for diagnostic test evaluation. ROC curve is a plot of Sensitivity against Specificity for probable cut-off points of a diagnostic test. It is the graphical representation of the contrast between true positive rates and the false positive rate at different thresholds.

29 . What does PCA stand for? Why is it important in ML?

Principal Component Analysis (PCA) is a statistical procedure that uses an orthogonal transformation that converts a set of correlated variables to a set of uncorrelated variables. PCA is the most widely used tool in exploratory data analysis and in machine learning for predictive models. Moreover, PCA is an unsupervised statistical technique used to examine the interrelations among a set of variables. It is also known as a general factor analysis where regression determines a line of best fit.

30 . What's the "kernel trick" and how is it useful?

Kernel trick involves kernel functions which enable higher dimension spaces without actually calculating the coordinates within that dimension. It uses the inner products between the images of all pairs data in a feature space. This allows the calculation of coordinates of higher dimensions with low dimensional data.

31 . What is the dimensionality reduction in Machine Learning?

It is the process of reducing random variables under consideration. Dimensionality reduction can be classified as feature selection and feature extraction.

Feature selection tries to find the subset of input variables, while feature extraction begins from an initial set of measured data and builds derived values.

32 . What is a Boltzmann Machine?

Boltzmann Machines have a simple learning algorithm that helps to discover exciting features in training data. These were among the first neural networks to learn internal representations and are capable of solving severe combinatory problems.

33 . What are the different types of Genetic Programming?

Different types of Genetic Programming are :

* Grammatical Evolution
* Tree-based Genetic Programming
* Stack-based Genetic Programming
* Linear Genetic Programming (LGP)
* Cartesian Genetic Programming (CGP)
* Extended Compact Genetic Programming (ECGP)
* Genetic Improvement of Software for Multiple Objectives (GISMO)
* Probabilistic Incremental Program Evolution (PIPE)
* Strongly Typed Genetic Programming (STGP)

34 . Explain L1 and L2 Regularization?

A regression model that uses L1 Regularization is called Lasso Regression, and the Model which uses L2 Regularization is called Ridge Regression.

* L1 regularization helps in eliminating the features that are not important.
* L1 regularization adds the penalty term in the cost function by adding the absolute value of weight (Wj), while L2 regularization adds the squared value of weights (Wj) in the cost function.
* One more difference between both of them is that L1 regularization tries to estimate the median of the data while L2 regularization tries to estimate the mean of the data.

35 . What is Cross-Validation in Machine Learning?

It is a technique for increasing the model performance by feeding multiple sample data from the dataset. The sampling process is done by breaking the data into smaller parts that have the same number of rows. Out of all the parts, one is randomly selected for the test and another one for train sets.

It consists of the following techniques :

* Holdout method
* k-fold cross-validation
* Stratified k-fold cross-validation
* Leave p-out cross-validation

36 . What is the difference between Gini Impurity and Entropy in a Decision Tree?

* Gini Impurity and Entropy are the metrics used for deciding how to split a Decision Tree.
* Gini measurement is the probability of a random sample being classified correctly if you randomly pick a label according to the distribution in the branch.
* Entropy is a measurement to calculate the lack of information. You calculate the Information Gain (difference in entropies) by making a split. This measure helps to reduce the uncertainty about the output label.

37 . What is the difference between Entropy and Information Gain?

* Entropy is an indicator of how messy your data is. It decreases as you reach closer to the leaf node.

* The Information Gain is based on the decrease in entropy after a dataset is split on an attribute. It keeps on increasing as you reach closer to the leaf node.

38 . What is inductive learning in machine learning?

From the perspective of **inductive learning**, we are given input samples (x) and output samples (f(x)) and the problem is to estimate the function (f). Specifically, the problem is to generalize from the samples and the mapping to be useful to estimate the output for new samples in the future.

In practice it is almost always too hard to estimate the function, so we are looking for very good approximations of the function.

Some practical examples of induction are :

Credit risk assessment.
* The x is the properties of the customer.
* The f(x) is credit approved or not.

Disease diagnosis.
* The x are the properties of the patient.
* The f(x) is the disease they suffer from.

Face recognition.
* The x are bitmaps of peoples faces.
* The f(x) is to assign a name to the face.

Automatic steering.
* The x are bitmap images from a camera in front of the car.
* The f(x) is the degree the steering wheel should be turned.

39 . When Should You Use Inductive Learning?

There are problems where inductive learning is not a good idea. It is important when to use and when not to use supervised machine learning.

4 problems where inductive learning might be a good idea :

Problems where there is no human expert. If people do not know the answer they cannot write a program to solve it. These are areas of true discovery.

Humans can perform the task but no one can describe how to do it. There are problems where humans can do things that computer cannot do or do well. Examples include riding a bike or driving a car.

Problems where the desired function changes frequently. Humans could describe it and they could write a program to do it, but the problem changes too often. It is not cost effective. Examples include the stock market.

Problems where each user needs a custom function. It is not cost effective to write a custom program for each user. Example is recommendations of movies or books on Netflix or Amazon.

40 . What is deductive learning in machine learning?

Deductive learning is a subclass of machine learning that studies algorithms for learning provably correct knowledge. Typically such methods are used to speedup problem solvers by adding knowledge to them that is deductively entailed by existing knowledge, but that may result in faster solutions.

41 . What is a Box-Cox transformation?

Box-Cox transformation is a power transform which transforms non-normal dependent variables into normal variables as normality is the most common assumption made while using many statistical techniques. It has a lambda parameter which when set to 0 implies that this transform is equivalent to log-transform. It is used for variance stabilization and also to normalize the distribution.

“KickStart your Artificial Intelligence Journey with Great Learning which offers high-rated Artificial Intelligence courses with world-class training by industry leaders. Whether you’re interested in machine learning, data mining, or data analysis, Great Learning has a course for you!”

42 . Explain the differences between Random Forest and Gradient Boosting machines.

Random forests are a significant number of decision trees pooled using averages or majority rules at the end. Gradient boosting machines also combine decision trees but at the beginning of the process unlike Random forests. Random forest creates each tree independent of the others while gradient boosting develops one tree at a time. Gradient boosting yields better outcomes than random forests if parameters are carefully tuned but it’s not a good option if the data set contains a lot of outliers/anomalies/noise as it can result in overfitting of the model.Random forests perform well for multiclass object detection. Gradient Boosting performs well when there is data which is not balanced such as in real time risk assessment.

43 . What is a confusion matrix and why do you need it?

Confusion matrix (also called the error matrix) is a table that is frequently used to illustrate the performance of a classification model i.e. classifier on a set of test data for which the true values are well-known.

It allows us to visualize the performance of an algorithm/model. It allows us to easily identify the confusion between different classes. It is used as a performance measure of a model/algorithm.

A confusion matrix is known as a summary of predictions on a classification model. The number of right and wrong predictions were summarized with count values and broken down by each class label. It gives us information about the errors made through the classifier and also the types of errors made by a classifier.

44 . What do you understand by ILP?

ILP stands for Inductive Logic Programming. It is a part of machine learning which uses logic programming. It aims at searching patterns in data which can be used to build predictive models. In this process, the logic programs are assumed as a hypothesis.

45 . 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).

46 . What is the difference between regularization and normalisation?

Normalisation adjusts the data; regularisation adjusts the prediction function. If your data is on very different scales (especially low to high), you would want to normalise the data. Alter each column to have compatible basic statistics. This can be helpful to make sure there is no loss of accuracy. One of the goals of model training is to identify the signal and ignore the noise if the model is given free rein to minimize error, there is a possibility of suffering from overfitting. Regularization imposes some control on this by providing simpler fitting functions over complex ones.

47 . List all assumptions for data to be met before starting with linear regression.

Before starting linear regression, the assumptions to be met are as follow :

* Linear relationship
* Multivariate normality
* No or little multicollinearity
* No auto-correlation
* Homoscedasticity

48 . What is SVM in machine learning? What are the classification methods that SVM can handle?

SVM stands for **Support Vector Machine**. SVM are supervised learning models with an associated learning algorithm which analyze the data used for classification and regression analysis.

The classification methods that SVM can handle are :

* Combining binary classifiers
* Modifying binary to incorporate multiclass learning

49 . What are Different Kernels in SVM?

There are six types of kernels in SVM :

Linear kernel : used when data is linearly separable.
Polynomial kernel : When you have discrete data that has no natural notion of smoothness.
Radial basis kernel : Create a decision boundary able to do a much better job of separating two classes than the linear kernel.
Sigmoid kernel : used as an activation function for neural networks.

50 . What is entropy in Machine Learning?

Entropy in Machine Learning measures the randomness in the data that needs to be processed. The more entropy in the given data, the more difficult it becomes to draw any useful conclusion from the data. For example, let's take the incident of flipping a coin. The result of this is random as it does not favor heads or tails. Here, the result for any

number of tosses cannot be predicted easily as there is no definite relationship between the action of flipping and the possible outcomes.

51 . What is epoch in Machine Learning?

Epoch in Machine Learning is used to indicate the count of passes in a given training dataset where the Machine Learning algorithm has done its job. Generally, when there is a huge chunk of data, it is grouped into several batches. Here, each of these batches goes through the given model, and this process is referred to as iteration. Now, if the batch size comprises the complete training dataset, then the count of iterations is the same as that of epochs.

In case there is more than one batch, **d*e=i*b** is the formula used, wherein **'d' is the dataset**, **'e' is the number of epochs**, **'i' is the number of iterations**, and **'b' is the batch size**.

52 . Explain Logistic Regression.

Logistic regression is the proper regression analysis used when the dependent variable is categorical or binary. Like all regression analyses, logistic regression is a technique for predictive analysis. Logistic regression is used to explain data and the relationship between one dependent binary variable and one or more independent variables. Also, it is employed to predict the probability of a categorical dependent variable.

We can use logistic regression in the following scenarios :

* To predict whether a citizen is a Senior Citizen (1) or not (0)
* To check whether a person is having a disease (Yes) or not (No)

There are three types of logistic regression:

* **Binary Logistic Regression :** In this, there are only two outcomes possible.
  Ex : To predict whether it will rain (1) or not (0)

* **Multinomial Logistic Regression:** In this, the output consists of three or more unordered categories.
  Ex : Prediction on the regional languages (Kannada, Telugu, Marathi, etc.)

* **Ordinal Logistic Regression:** In ordinal logistic regression, the output consists of three or more ordered categories.
  Ex : Rating an Android application from 1 to 5 stars.

53 . List the advantages and limitations of the Temporal Difference Learning Method.

Temporal Difference Learning Method is a mix of Monte Carlo method and Dynamic programming method. Some of the advantages of this method include :

* It can learn in every step online or offline.
* It can learn from a sequence which is not complete as well.
* It can work in continuous environments.
* It has lower variance compared to MC method and is more efficient than MC method.

Limitations of TD method are :

* It is a biased estimation.
* It is more sensitive to initialization.

54 . How would you handle an imbalanced dataset?

Sampling Techniques can help with an imbalanced dataset. There are two ways to perform sampling, Under Sample or Over Sampling.

In Under Sampling, we reduce the size of the majority class to match minority class thus help by improving performance w.r.t storage and run-time execution, but it potentially discards useful information.

For Over Sampling, we upsample the Minority class and thus solve the problem of information loss, however, we get into the trouble of having Overfitting.

There are other techniques as well :

Cluster-Based Over Sampling : In this case, the K-means clustering algorithm is independently applied to minority and majority class instances. This is to identify clusters in the dataset. Subsequently, each cluster is oversampled such that all clusters of the same class have an equal number of instances and all classes have the same size

Synthetic Minority Over-sampling Technique (SMOTE) : A subset of data is taken from the minority class as an example and then new synthetic similar instances are created which are then added to the original dataset. This technique is good for Numerical data points.

55 . Mention some of the EDA Techniques?

Exploratory Data Analysis (EDA) helps analysts to understand the data better and forms the foundation of better models.

Visualization :
* Univariate visualization
* Bivariate visualization
* Multivariate visualization

Missing Value Treatment : Replace missing values with Either Mean/Median

Outlier Detection : Use Boxplot to identify the distribution of Outliers, then Apply IQR to set the boundary for IQR

Transformation : Based on the distribution, apply a transformation on the features

Scaling the Dataset : Apply MinMax, Standard Scaler or Z Score Scaling mechanism to scale the data.

Feature Engineering : Need of the domain, and SME knowledge helps Analyst find derivative fields which can fetch more information about the nature of the data

Dimensionality reduction : Helps in reducing the volume of data without losing much information

56 . 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.

57 . 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.

58 . 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.

59 . 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.

60 . 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 :

* Decision Tree
* Naive Bayes Classifier
* K-Nearest Neighbors
* Support Vector Machines
* Artificial Neural Networks

61 . 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 :

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

62 . What is Data augmentation?

Data augmentation is a machine learning strategy that enables the users to increase the data diversity for training models remarkably from internal and external sources within an enterprise. This does not require any new data collection.

Modification in images is one of the most helpful examples of data augmentation. We can easily perform the following activities on an image and modify it :

* Deforming
* Adding noise
* Modifying colors
* Resizing the image
* Flipping it horizontally or vertically

63 . Python or R â€“ Which is the best for machine learning?

In machine learning projects, both R and Python come with their own advantages. However, Python is more useful in data manipulation and repetitive tasks, making it the right choice if you plan to build a digital product based on machine learning. Moreover, to develop a tool for ad-hoc analysis at an early stage of the project, R is more suitable.

64 . Why rotation is required in PCA? What will happen if you don't rotate the components?

Rotation is a significant step in PCA as it maximizes the separation within the variance obtained by components. Due to this, the interpretation of components becomes easier.

The motive behind doing PCA is to choose fewer components that can explain the greatest variance in a dataset. When rotation is performed, the original coordinates of the points get changed. However, there is no change in the relative position of the components.

If the components are not rotated, then we need more extended components to describe the variance.

65 . What is Rescaling of data and how is it done?

In real-world scenarios, the attributes present in data will be in a varying pattern. So, rescaling of the characteristics to a common scale gives benefit to algorithms to process the data efficiently.

We can rescale the data using Scikit-learn. The code for rescaling the data using MinMaxScaler is as follows:

```
#Rescaling data
import pandas
```

```
import scipy
import numpy
from sklearn.preprocessing import MinMaxScaler
names = ['Ramu', 'Ramana', 'Mounika', 'Sathya', 'raj', 'mani', 'samu', 'venu', 'sam']
Dataframe = pandas.read_csv(url, names=names)
Array = dataframe.values
# Splitting the array into input and output
X = array[:,0:8]
Y = array[:,8]
Scaler = MinMaxScaler(feature_range=(0, 1))
rescaledX = scaler.fit_transform(X)
# Summarizing the modified data
numpy.set_printoptions(precision=3)
print(rescaledX[0:5,:])
```

66 . Why do we need to convert categorical variables into factor? Which functions are used to perform the conversion?

Most Machine learning algorithms require number as input. That is why we convert categorical values into factors to get numerical values. We also don't have to deal with dummy variables.

The functions **factor()** and **as.factor()** are used to convert variables into factors.

67 . What is a good metric for measuring the level of multicollinearity?

VIF or **1/tolerance** is a good measure of measuring multicollinearity in models. VIF is the percentage of the variance of a predictor which remains unaffected by other predictors. So higher the VIF value, greater is the multicollinearity amongst the predictors.

A rule of thumb for interpreting the variance inflation factor :

* 1 = not correlated.
* Between 1 and 5 = moderately correlated.
* Greater than 5 = highly correlated.

68 . What is a pipeline?

A pipeline is a sophisticated way of writing software such that each intended action while building a model can be serialized and the process calls the individual functions for the individual tasks. The tasks are carried out in sequence for a given sequence of data points and the entire process can be run onto n threads by use of composite estimators in scikit learn.

69 . What are the advantages of SVM algorithms?

SVM algorithms have basically advantages in terms of complexity. First I would like to clear that both Logistic regression as well as SVM can form non linear decision surfaces and can be coupled with the kernel trick. If Logistic regression can be coupled with kernel then why use SVM?

â—Ž SVM is found to have better performance practically in most cases.

â—Ž SVM is computationally cheaper O(N^2*K) where K is no of support vectors (support vectors are those points that lie on the class margin) where as logistic regression is O(N^3)

â—Ž Classifier in SVM depends only on a subset of points . Since we need to maximize distance between closest points of two classes (aka margin) we need to care about only a subset of points unlike logistic regression.

70 . Why does XGBoost perform better than SVM?

First reason is that **XGBoos** is an ensemble method that uses many trees to make a decision so it gains power by repeating itself.

SVM is a linear separator, when data is not linearly separable SVM needs a Kernel to project the data into a space where it can separate it, there lies its greatest strength and weakness, by being able to project data into a high dimensional space SVM can find a linear separation for almost any data but at the same time it needs to use a Kernel and we can argue that there's not a perfect kernel for every dataset.

71 . What is Finds Algorithm in Machine Learning?

We will implement and demonstrate the FIND-S algorithm for finding the most specific hypothesis based on a given set of training data samples.

In finds algorithm , we initialize hypothesis as an array of phi then in the first step we replace it with the first positive row of our dataset which is most specific hypothesis.

In next step, we will traverse the dataset and check if the target value of dataset is positive or not, we will only consider positive value. if the value is positive we will traverse that row from start to end and check if any element matches with our respective hypothesis. if the element does not matches with the hypothesis, we will generalize the hypothesis and we will replace element in hypothesis with the dataset element .

Discussion of algorithm :
* FIND-S Algorithm starts from the most specific hypothesis and generalizes it by considering only positive examples.

* FIND-S algorithm ignores negative examples. – As long as the hypothesis space contains a hypothesis that describes the true target concept, and the training data contains no errors, ignoring negative examples does not cause any problem.

72 . What is F1 score? How would you use it?

Let's have a look at this table before directly jumping into the F1 score.

Prediction	Predicted Yes	Predicted No
Actual Yes	True Positive (TP)	False Negative (FN)
Actual No	False Positive (FP)	True Negative (TN)

In binary classification we consider the **F1** score to be a measure of the model's accuracy. The **F1** score is a weighted average of precision and recall scores.

F1 = 2TP/2TP + FP + FN

We see scores for F1 between 0 and 1, where 0 is the worst score and 1 is the best score.

The F1 score is typically used in information retrieval to see how well a model retrieves relevant results and our model is performing.

73 . How to Tackle Overfitting and Underfitting?

Overfitting means the model fitted to training data too well, in this case, we need to resample the data and estimate the model accuracy using techniques like k-fold cross-validation.

Whereas for the Underfitting case we are not able to understand or capture the patterns from the data, in this case, we need to change the algorithms, or we need to feed more data points to the model.

74 . How can you select K for K-means Clustering?

There are two kinds of methods that include direct methods and statistical testing methods :

* **Direct methods :** It contains elbow and silhouette
* **Statistical testing methods :** It has gap statistics.

The silhouette is the most frequently used while determining the optimal value of k.

75 . How do check the Normality of a dataset?

Visually, we can use plots. A few of the normality checks are as follows :

* Shapiro-Wilk Test
* Anderson-Darling Test
* Martinez-Iglewicz Test
* Kolmogorov-Smirnov Test
* D'Agostino Skewness Test

76 . What is a model selection in Machine Learning?

The process of choosing models among diverse mathematical models, which are used to define the same data is known as **Model Selection**. Model learning is applied to the fields of **statistics**, **data mining**, and **machine learning**.

77 . What are the necessary steps involved in Machine Learning Project?

There are several essential steps we must follow to achieve a good working model while doing a Machine Learning Project.

Those steps may include **parameter tuning**, **data preparation**, **data collection**, **training the model**, **model evaluation**, and **prediction**, etc.

78 . What are the similarities and differences between bagging and boosting in Machine Learning?

Similarities of Bagging and Boosting :

* Both are the ensemble methods to get N learns from 1 learner.
* Both generate several training data sets with random sampling.
* Both generate the final result by taking the average of N learners.
* Both reduce variance and provide higher scalability.

Differences between Bagging and Boosting :

* Although they are built independently, but for Bagging, Boosting tries to add new models which perform well where previous models fail.
* Only Boosting determines the weight for the data to tip the scales in favor of the most challenging cases.
* Only Boosting tries to reduce bias. Instead, Bagging may solve the problem of over-fitting while boosting can increase it.

79 . Why instance-based learning algorithm sometimes referred to as Lazy learning algorithm?

In **machine learning**, **lazy learning** can be described as a method where induction and generalization processes are delayed until classification is performed. Because of the same property, an instance-based learning algorithm is sometimes called lazy learning algorithm.

80 . What is Regularization? What kind of problems does regularization solve?

A **regularization** is a form of regression, which constrains/ **regularizes** or **shrinks** the coefficient estimates towards zero. In other words, it discourages learning a more complex or flexible model to avoid the risk of overfitting. It reduces the variance of the model, without a substantial increase in its bias.

Regularization is used to address overfitting problems as it penalizes the loss function by adding a multiple of an **L1 (LASSO)** or an **L2 (Ridge)** norm of weights vector w.

81 . Can you explain how dimensionality reduction works?

Dimensionality reduction is a technique for reducing the number of features in a dataset while still retaining the important information. This can be useful in many areas, such as image and speech recognition, natural language processing, and even in stock market analysis.

The basic idea behind dimensionality reduction is to project the high-dimensional data onto a lower-dimensional space while preserving the structure and relationships between the data points. There are many techniques for doing this, including:

Principal Component Analysis (PCA) : This is a linear dimensionality reduction technique that transforms the data to a lower-dimensional space by finding the directions of maximum variance in the data. The transformed data is then projected onto these directions, effectively removing the redundancy in the data.

Singular Value Decomposition (SVD) : This is a linear dimensionality reduction technique that uses matrix decomposition to reduce the dimensionality of the data.

t-SNE (t-distributed Stochastic Neighbor Embedding) : This is a non-linear dimensionality reduction technique that maps high-dimensional data to a lower-dimensional space while preserving the local structure of the data.

Auto-Encoder : This is a deep learning-based approach to dimensionality reduction that involves training a neural network to reconstruct the original data from a lower-dimensional representation.

In each of these techniques, the goal is to find a lower-dimensional representation of the data that retains the important information. The reduced representation can then be used for further analysis or as input to a machine learning algorithm.

82 . Can you discuss some popular techniques for feature engineering?

Feature engineering is the process of creating new features or transforming existing ones in order to improve the performance of machine learning models. Some popular techniques for feature engineering include:

Binning : Binning is the process of converting a continuous feature into a categorical feature by dividing the range of the feature into bins. This can be useful for dealing with non-linear relationships in the data.

One-hot encoding : One-hot encoding is a technique for converting categorical features into a numerical representation that can be used by machine learning algorithms. The idea is to convert each unique category into a new binary feature, with a value of 1 indicating the presence of that category and a value of 0 indicating its absence.

Polynomial features : Polynomial features are new features generated by raising the original features to a power and combining them in a polynomial expression. This can be useful for capturing non-linear relationships in the data.

Interaction features : Interaction features are new features generated by combining two or more existing features. This can be useful for capturing the combined effect of multiple features on the target variable.

Logarithmic transformation : Logarithmic transformation is a technique for transforming a feature by taking the logarithm of its values. This can be useful for reducing the skew in the distribution of the feature and improving the linearity of the relationship between the feature and the target variable.

Scaling and normalization : Scaling and normalization are techniques for transforming features so that they have similar ranges and distributions. This can be important for ensuring that the features have similar importance in the machine learning algorithms.

Aggregation : Aggregation is the process of summarizing data by aggregating it into a smaller number of features. This can be useful for reducing the noise in the data and improving the interpretability of the features.

These are just a few of the many techniques that can be used for feature engineering. The choice of technique will depend on the specific problem and the characteristics of the data. The goal of feature engineering is to create new features that capture the relationships and patterns in the data in a way that will improve the performance of the machine learning models.

83 . Can you explain the difference between reinforcement learning and supervised learning?

Supervised learning is a machine learning paradigm where the algorithm is trained on labeled data, where the desired output is known for each input. The goal of supervised learning is to learn the mapping between inputs and outputs, and to make predictions about the output for new, unseen inputs. Supervised learning algorithms are trained by minimizing the difference between the predicted outputs and the actual outputs.

Reinforcement learning, on the other hand, is a type of machine learning where the algorithm learns to make

decisions by taking actions in an environment to maximize a reward signal. In reinforcement learning, the algorithm does not receive labeled data, but instead receives feedback in the form of rewards or penalties based on the actions it takes. The goal of reinforcement learning is to learn a policy, which is a mapping from states to actions, that maximizes the cumulative reward over time.

* One key difference between the two paradigms is that in ***supervised learning the desired output is known***, while in ***reinforcement learning the desired outcome is not known***, and the algorithm must learn it through trial and error.

* Another difference is that in supervised learning the algorithm is trained on a fixed dataset, while in reinforcement learning the algorithm must continually interact with its environment to receive feedback and improve its policy.

In summary, supervised learning is best suited for problems where the desired output is known for each input, and the goal is to learn the mapping between inputs and outputs. Reinforcement learning is best suited for problems where the goal is to learn a policy that maximizes a reward signal in an environment through trial and error.

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